Narodna in univerzitetna knjižnica
v Ljubljani

Č 278998

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MARIJAN BLEJEC

STATISTIČNE METODE ZA EKONOMISTE

Druga, predelana in razširjena izdaja

Prvi ponatis

LJUBLJANA 1976

UNIVERZA V LJUBLJANI

EKONOMSKA FAKULTETA

MARIJAN BLEJEC

STATISTIČNE METODE ZA EKONOMISTE

Druga, predelana in razširjena izdaja

Prvi ponatis

LJUBLJANA 1976

278998

278998



24-1-1977 01/2-21

a

PREDGOVOR K PRVI IZDAJI

Učbenik je napisan po predavanjih iz Splošne statistike na Ekonomski fakulteti v Ljubljani. Kljub temu, da predavanja iz Splošne statistike na Ekonomski fakulteti in učbenik obravnavajo splošne statistične metode, je upravičen naslov "Statistične metode za ekonomiste", ker je snov podajana iz aspekta uporabe v ekonomiji, kar je razvidno iz razmerja med posameznimi poglavji, pa tudi iz primerov, s katerimi so ilustrirane posamezne metode. Tako med drugim npr. poglavja: indeksi in časovne vrste obravnavajo metode, ki jih uporabljamo skoraj izključno pri analizi socialno-ekonomskih pojavov.

Učbenik je razdeljen v 17 poglavij, od katerih vsako obravnava bodisi posamezno fazo statističnega proučevanja ali statistične metode analize. V dodatku je dana zbirka oznak in obrazcev iz učbenika in osnovne statistične tabele. Zbirka oznak in obrazcev in statistične tabele so priložene ločeno, da jih moremo po potrebi uporabljati pri praktičnem delu samostojno.

Primeri, ki ilustrirajo posamezne postopke so vzeti predvsem iz publikacij Zveznega zavoda za statistiko in Zavoda za statistiko LRS. Viri so pri posameznih tabelah, podatkih in grafikonih označeni največkrat s kraticami publikacij oziroma ustanov, ki so podatke dale. Uporabili smo tele kratice:

SG = Statistični godišnjak; Zvezni zavod za statistiko (številka zraven kratice pomeni letnik (npr.: SG-57)

SB = Statistički bilten: Zvezni zavod za statistiko (številka zraven kratice pomeni številko biltena)

I = Indeks, Zvezni zavod za statistiko

SL = Statistični letopis LRS; Zavod za statistiko LRS (številka zraven kratice pomeni letnik)

SB-LRS = Statistični Bilten LRS

MSP = Mesečni statistični pregled; Zavod za statistiko LRS.

V zbirki oznak in obrazcev so obrazci urejeni in označeni enako kot v ustreznih poglavjih, tako da je dana neposredna zveza med obrazci v zbirki in tolmačenjem v ustreznih poglavjih.

Ljubljana, februar, 1959

M. Blejec

***

PREDGOVOR K DRUGI IZDAJI

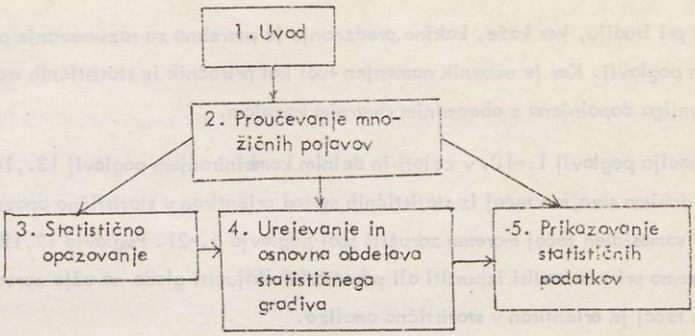
Prva izdaja učbenika "Statistične metode za ekonomiste", ki je izšla leta 1959 je bila v času od 1959-1972 štirikrat ponatisnjena v skupno čez 14.000 izvodih. To kaže na dosti veliko zanimanje tudi izven kroga slušateljev ekonomske fakultete, za katere je bil učbenik obvezna literatura pri predmetu: Statistične metode. Vse večja uporaba statističnih metod tudi pri nas in razširitev predmeta: Statistične metode na ekonomski fakulteti iz enoletnega tečaja v dva enoletna tečaja je narekovala potrebo po predelavi in razširitvi obstoječega učbenika.

Predelana in razširjena izdaja obsega 21 poglavij. Od poglavij, ki so bila vključena že v prvi izdaji, so v bistveni meri spremenjena in dopolnjena poglavja: 16- Preskušanje domnev, 18- Proučevanje odvisnosti med pojavi in 21- Proučevanje dinamike pojavov. Nanovo pa so vključena poglavja: 11- Mere koncentracije, 17- Statistična analiza odločitev, 19- Statistično preverjanje kakovosti in 20-Statistično načrtovanje poskusov. Spremenjen vrstni red poglavij je pogojen s sodobnim vzorčnim konceptom posameznih področij (npr. poglavje 18).

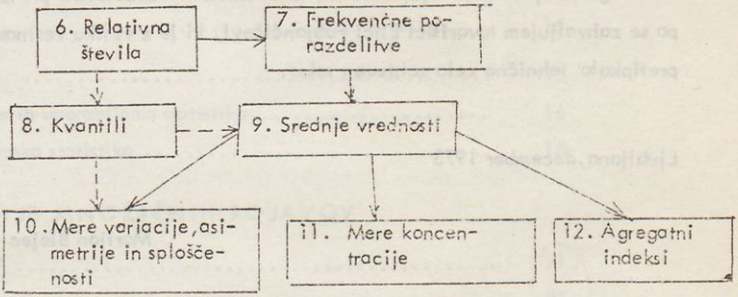
Glede na obsežno problematiko, ki jo obravnava učbenik, podajamo v shemi pregled poglavij s tesnejšo ——→ in rahlejšo povezavo - - -> med poglavji. V shemi so poglavja 1.-5. združena v skupino I.-Statistično opazovanje, poglavja 6.-12. v skupino II.- Opisne metode statistične analize in poglavja 13.-31. v skupino III.-Vzorčne in posebne metode statistične analize. Ta shema more služiti kot pri-

SHEMATIČEN PREGLED POGLAVIJ

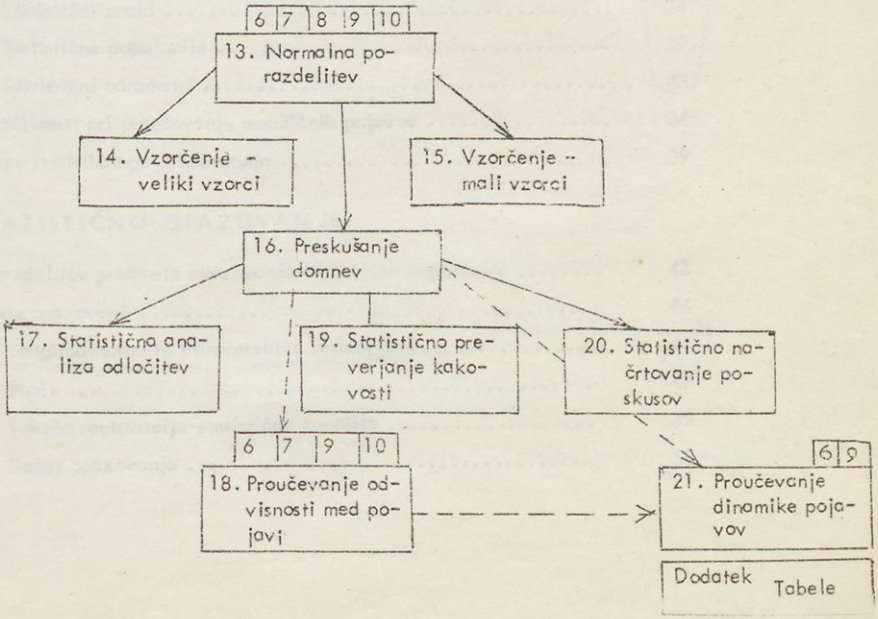
I. STATISTIČNO OPAZOVANJE



II. OPISNE METODE ANALIZE



III. VZORČNE IN POSEBNE METODE ANALIZE



pomoček pri študiju, ker kaže, kakšno predznanje je potrebno za razumevanje posameznih poglavij. Ker je učbenik namenjen tudi kot priročnik iz statističnih metod, je knjiga dopolnjena z abecednim stvarnim kazalom.

S kombinacijo poglavij 1.-10. v celoti in delnim kombiniranjem poglavij 13., 14, 18. in 21. dobimo skrajšan tečaj iz statističnih metod orientiran v statistično opazovanje. V samostojen tečaj moremo združiti tudi poglavja 6.-21. Poglavja 17., 19. in 20. moremo pri tem bodisi izpustiti ali posamezna vključiti glede na ožjo opredelitev. Ta tečaj je orientiran v statistično analizo.

Ob drugi izdaji se zahvaljujem vsem, ki so kakorkoli sodelovali pri izidu. Posebej pa se zahvaljujem tovarišici Elici Fabjančičevi, ki je z veliko vestnostjo in skrbjo pretipkala tehnično zelo zahteven tekst.

Ljubljana, december 1973

Marijan Blejec

VSEBINA

Str.

1. UVOD 15

Kaj je statistika 15

Področja, v katerih uporabljamo statistiko 16

Socialno-ekonomska statistika 18

2. PROUČEVANJE MNOŽIČNIH POJAVOV

Množični pojavi 21

Statistične enote 23

Statistični znaki 24

Statistične populacije 29

Statistični parametri 33

Značilnosti pri proučevanju množičnih pojavov 34

Etape statističnega proučevanja 39

3. STATISTIČNO OPAZOVANJE

Opredelitev predmeta opazovanja-statistične populacije 43

Vrste opazovanj 44

Longitudinalno in transverzalno opazovanje 45

Popis 47

Tekoča registracija-statistična poročila 49

Delna opazovanja 50

Str.

Viri podatkov 52
Načini posrednega opazovanja 54
Znaki opazovanja 57
Sredstva opazovanja 58
Kraj opazovanja 66
Organi statističnega opazovanja 67
Napake in kontrola pri statističnem opazovanju 69

4. UREJEVANJE IN OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

Grupiranje vrednosti znakov 75
Šifriranje osnovnega gradiva 87
Osnovna obdelava statističnega gradiva 88

5. PRIKAZOVANJE STATISTIČNIH PODATKOV

Statistične vrste 100
Statistične tabele 104
Grafično prikazovanje 110
Prvine grafičnega prikazovanja 111
Skale - lestvice 113
Mreže 116
Vrste grafikonov 118
Stolpci 118
Linijski grafikoni 122
Figure 128
Kartogrami 130

6. RELATIVNA ŠTEVILA

Strukture ali razčlenitvena števila 138
Enostavne strukture 138
Večkratne strukture 140

Splošne zveze za dvojne strukture 142
Grafično prikazovanje struktur 144
Statistični koeficienti in gostote 157
Statistični koeficienti in gostote 157
Grafično prikazovanje 168
Enostavni indeksi 174
Stvarni in krajevni indeksi 176
Časovni indeksi 177

7. FREKVENČNE PORAZDELITVE

Sestavljanje frekvenčnih porazdelitev 185
Frekvenčne porazdelitve z neenakimi razredi 188
Grafično prikazovanje frekvenčnih porazdelitev 191
Oblike frekvenčnih porazdelitev 195
Kumulativna frekvenčna porazdelitev 200

8. KVANTILI 203

Ranžirna vrsta, rang 203
Kvantilni rang 205
Kvantili 206
Rang grafikon 211
Izračun kvantilnih rangov in kvantilov iz frekvenčnih porazdelitev 213

9. SREDNJE VREDNOSTI

Vrste srednjih vrednosti 220
Mediana 220
Modus 221
Aritmetična sredina 224
Izračun aritmetične sredine iz frekvenčnih porazdelitev 228
Aritmetična sredina aritmetičnih sredin 233

Str.

Harmonična sredina 234
Poprečja iz relativnih števil 237
Standardizirani pokazovalci 240
Geometrijska sredina 242
Odnosi med različnimi vrstami srednjih vrednosti 246

10. MERE VARIACIJE, ASIMETRIJE IN SPLOŠČENOSTI

Mere variacije 249
Vrste mer variacije 249
Variacijski razmik 250
Kvartilni odklon 251
Poprečen absolutni odklon 252
Varianca. Standardni odklon 254
Izračun iz negrupiranih podatkov 254
Izračun iz grupiranih podatkov 256
Skupna varianca 260
Zveza standardnega odklona z normalno porazdelitvijo 262
Poprečna razlika Δ_1 263
Razmerje med Q, AD in SD za normalno porazdelitev 265
Relativne mere variacije 266
Relativni odkloni 270
Mere variacije za relativna števila 271
Mere asimetrije in sploščenosti 276
Mere asimetrije in sploščenosti po legi 276
Mere asimetrije in sploščenosti, izračunane iz momentov 277

11. MERE KONCENTRACIJE

Lorenzov grafikon 283
Ginijev koeficient koncentracije 287

Hirschmanov koeficient koncentracije 290
Mera koncentracije na osnovi AD_{me} 291

12. AGREGATNI INDEKSI

Relativne spremembe agregatov 295
Laspeyeresov in Paachejev agregatni indeks 297
Fisherjev idealni obrazec 299
Agregatni verižni indeksi 299
Preskusi z agregatnimi indeksi 300
Skupinski indeksi 301
Indeksi cen 303
Agregatni indeksi količin - obsega 310
Indeksi recipročnih pokazovalcev 312

13. NORMALNA PORAZDELITEV

Pomen proučevanja teoretičnih porazdelitev 315
Normalna porazdelitev 316
Opis 317
Standardizirana normalna porazdelitev 319
Prilagoditev normalne porazdelitve stvarni frekvenčni porazdelitvi 325
Verjetnosti grafikoni 327
Normalna porazdelitev kot verjetnostna porazdelitev 333

14. VZORČENJE - VELIKI VZORCI

Osnove vzorčenja 339
Prednosti in pomanjkljivosti metode vzorčenja 340
Uporaba vzorčenja pri statističnem opazovanju 342
Osnovna populacija. Vzorec. Populacija vseh vzorcev 343
Enostavno slučajnostno vzorčenje 348

Str.

Ocenjevanje aritmetične sredine 348

Točkovna ocena. Odklon zaupanja. Intervalna ocena. Razmik zaupanja 350

Ocenjevanje parametrov z enostavnim slučajnostnim vzorčenjem na splošno 361

Ocenjevanje agregata Y, strukturnega deleža P% in števila enot z dano značilnostjo N_a 361

Tehnika izbora enostavnega slučajnostnega vzorca 365

Določanje velikosti vzorca pri zahtevani natančnosti ocene 376

Ocenjevanje po metodi razmerij 380

Vzorčenja z omejitvami 383

Stratificirano vzorčenje 383

Vzorčenje v skupinicah 395

Sistematično vzorčenje 400

Vzorčenje v dveh ali več stopnjah 403

Vzorčenje v dveh fazah 404

Kvotno vzorčenje 406

Časovno vzorčenje 409

Sklep 415

15. VZORČENJE - MALI VZORCI

Teoretične verjetnostne porazdelitve 417

Zvezne verjetnostne porazdelitve 417

Standardizirana normalna porazdelitev 417

χ^2 - porazdelitev 418

Studentova t- porazdelitev 421

F-porazdelitev 422

Zveze med teoretičnimi slučajnostnimi spremenljivkami, ki so zasnovane na normalni porazdelitvi 423

Str.

Razmične ocene z malimi vzorci 425
Ocenjevanje aritmetičnih sredin 426
Ocenjevanje razlik med aritmetičnima sredinama 430
Ocenjevanje variance 434
Ocenjevanje razmerja med variancama 436
Nezvezne porazdelitve 437
Hipergeometrična porazdelitev 437
Binomska-Bernoullijeva porazdelitev 438
Poissonova porazdelitev 443
Rekurzijski obrazci za nezvezne verjetnostne porazdelitve... 448

16. PRESKUŠANJE DOMNEV

Osnovna in ničelna domneva 454
Enostavne in sestavljene domneve 455
Preskušanje enostavnih domnev 455
Napaka I. in II. vrste 456
Značilnost in neznačilnost vzorčnih izrazov 459
Preskušanje sestavljenih domnev 464
Operativna karakteristična krivulja OKK 466
Posplošenje preskušanja domnev 471
Preskušanje domnev z velikimi vzorci 472
Preskus domnev o parametrih 472
Preskus domnev o razlikah med parametri 474
Preskušanje domnev z malimi vzorci 475
Preskušanje domnev o aritmetičnih sredinah 476
Preskus o popečni razliki 477
Preskus domneve o razliki med aritmetičnima sredinama za neodvisna vzorca 479
Preskušanje domnev o varianci 481
Preskušanje domnev o razmerju med variancama 482

Str.

Preskus domnev o frekvenčnih porazdelitvah 483
Analiza variance 485
Neparametrične metode za preskušanje domnev 491
Neparametrično preskušanje domnev o srednjih vrednostih 492
Preskus s predznaki - preskus z mediano 493
Preskus s predznačenimi rangi (Wicoxonov test predznačenih rangov) 494
Preskus z vsoto rangov (Mann-Withneyev U test)..... 495
Analiza variance z rangi (Kruskal-Wallisov test) 497
Preskusi o slučajnosti razporeda enot 498
Preskus s sekvencami 499
Preskus o slučajnosti razporeda z variancami - von Neumannov preskus 502

17. STATISTIČNE METODE ODLOČITEV

Prvine 505
Matrika posledic 505
Priložnostna izguba 506
Indeks koristnosti 507
Odločitve pri gotovosti 508
Odločitve pri tveganju 508
Odločitve pri negotovosti 509
Pravilo maksimin 510
Pravilo maksimaks 510
Hurwiczovo pravilo 511
Minimaks sodilo 512
Bayesovo sodilo - priorna analiza 512
Linearna funkcija dohodka in točka preloma 513

Str.

Bayesovo sodilo z vzorčenjem - posteriorna analiza 517
Preposteriorna analiza 519

18. PROUČEVANJE ODVISNOSTI MED MNOŽIČNIMI POJAVI

Funkcijske odvisnosti 521
Korelacijske odvisnosti 524
Prikazovanje korelacijskih odvisnosti 524
Regresijska krivulja 531
Metode za določanje regresijske krivulje 533
Mere jakosti odvisnosti 539
Standardna napaka ocene 542
Linearna odvisnost 543
Pokazovalci 543
Izračun pokazovalcev za linearno odvisnost 545
Linearna regresija 547
Linearna korelacija 555
Parametri 555
Verjetnostna porazdelitev ocen r_{xy} 558
Preskušanje domnev o korelacijskih koeficientih 559
Krivuljčne odvisnosti 566
Poseben tip regresijskega modela 569
Regresijski modeli tipa $g(y) = \alpha_o + \beta_1 f(x) + \epsilon$ 571
Večparametrski regresijski model 576
Korelacijsko razmerje 580
Korelacija ranga 584
Asociacija in kontingenca 587
Kontingenčna tabela 588
Merila stopnje kontingence 591

Str.

Preskušanje domnev o odvisnosti s χ^2 preskusom 598
Homogenost za skupino strukturnih deležev 598
Asociacija 601
Pokazovalci asociacije 601
Preskušanje domnev o asociaciji 605
Preskus z mediano 606
Neparametrično preskušanje domnev za vezana vzorca 608
Neparametrični preskus za dva neodvisna vzorca 609
Multipla regresija in korelacija 610
Linearna multipla regresija 611
Koeficient multiple korelacije 615
Parcialna linearna korelacija 617
Partkorelacijski koeficient 620
Multipla krivuljčna regresija in korelacija 624
Krivuljčna parabolična regresija drugega reda 624
Coob-Douglasova funkcija 624

19. STATISTIČNO PREVERJANJE KAKOVOSTI

Osnova 629
Statistično preverjanje procesov SPP 630
Kontrolne karte $\bar{y}$-s karta 632
y-me-R karta 635
Kontrolne karte za atributivne znake 636
Normalna in poostrena kontrola 637
Preverjanje deleža neustreznih izdelkov P% z $\bar{y}$-s^2 kontrolno karto 638
Statistično preverjanje kakovosti skupin SPS 645
Popolna in vzorčna kontrola pri prevzemu 646
Karakteristike načrtov za kontrolo pri prevzemu na osnovi izhodne kakovosti 647

Str.

Enojni, dvojni, večkratni in sekvencialni načrti 652

20. STATISTIČNO NAČRTOVANJE POSKUSOV

Prvine načrtovanja poskusov 658
Čisto slučajnostni poskus 662
Primerjave med aritmetičnimi sredinami, dobljenimi z analizo variance 666
Slučajnostni bloki 670
Latinski kvadrati 676
Faktorski poskusi 683
Drugi načrti poskusov 688

21. PROUČEVANJE DINAMIKE POJAVOV

Oblike časovnih vrst 692
Trenutne in razmične časovne vrste 692
Izvedene časovne vrste 693
Kumulativna časovna vrsta 693
Vrsta sredin 694
Vrsta drsečih vsot 696
Časovna vrsta drsečih sredin 698
Grafično prikazovanje časovnih vrst 702
Pollogaritemski grafikon 703
Polarni grafikon 708
Z-diagram 710
Brunsmanov grafikon 712
Ganttov grafikon 713
Analiza časovnih vrst 718
Primerljivost podatkov v časovni vrsti 728
Enostavni pokazovalci dinamike 723

Str.

Sestavine dinamike v časovnih vrstah 726
Osnovni modeli časovnih vrst 729
Vloga poprečij pri proučitvi časovnih vrst 731
Trend 732
Preskus o značilnosti trenda 736
Metode za določanje trenda 738
Prostoročna metoda 738
Metoda drsečih sredin 740
Metoda sredin med najnižjimi in najvišjimi točkami ... 742
Analitične metode za določanje trenda 743
Transformacija časa 743
Metoda izbranih točk 744
Metoda delnih sredin 748
Metoda delnih vsot 751
Metoda najmanjših kvadratov 759
Linearni trend 760
Določanje trenda z ortogonalnimi polinomi 763
Eksponentni trend 773
Periodične variacije 777
Metode za določanje sezonske variacije 778
Metoda vsot 779
Metoda vsot s popravkom za smer razvoja 782
Metoda koeficientov dinamike 786
Metoda kvocientov na vrsto drsečih sredin 791
Dinamična sezonska komponenta 796
Ciklični vplivi 803
Preskus o značilnosti ciklične komponente 803
Merjenje ciklov 805
Rezidualna metoda 805
Izločitev trenda in sezonskih vplivov 805
Izločanje iregularnih vplivov 808
Primerjava ciklov - Standardizacija 813

Str.

Proučevanje odvisnosti v dinamiki pojavov 816

Korelacija med časovnima vrstama po izločitvi trenda 817

Korelacija z odlogom 822

Avtokorelacija 826

Multipla korelacija pri dinamiki pojavov 828

Durbin-Watsonov preskus 828

DODATEK: T A B E L E

Tabela A Normalna porazdelitev - ploščine H(z) 831

Tabela B Normalna porazdelitev - ordinate $\varphi(z)$ 832

Tabela C t-porazdelitev - Kritične vrednosti t_p 833

Tabela Č χ^2-porazdelitev - Kritične vrednosti χ^2_p 834

Tabela D F-porazdelitev - Kritične vrednosti F_p

a) P = 0.05 835

b) P = 0.01 836

c) P = 0.001 837

Tabela E Studentiziran razmik $q = (\bar{Y}_{max} - \bar{Y}_{min})/s_y$ ... 838

Tabela F Preskus z mediano - Kritične vrednosti za n_s^+ ... 839

Tabela G Preskus s predznačenimi rangi Wilcoxonov preskus

Spodnja - T_S in zgornja T_Z kritična vrednost za vzorce z n = 6(1) 20 enotami 840

Tabela H Preskus z vsoto rangov - Man Withneyev preskus

Spodnja T'_s in zgornja T'_z kritična vrednost za vsote rangov T' 841

Tabela I Preskus s sekvencami

Spodnja r_s in zgornja r_z kritična vrednost za število sekvenc r za α=P = 0.5 in α=2P=0.10 842

Tabela J Fisherjeva transformacija $Z = \frac{1}{2} \ln \frac{1+r}{1-r}$ 843

a) Z = Z(r)

b) r = r(Z)

Str.

Tabela K Ortogonalni polinomi binomskih funkcij 844

Matrike d_{kh} I D_k za N = 2(1)30

Tabela L Slučajnostna števila 847

Tabela M Logaritmi tromestnih števil 848

STVARNO KAZALO 849

P R V O P O G L A V J E

U V O D

K A J J E S T A T I S T I K A ?

1.1 Z imenom statistika razumemo več stvari. Statistike so številčni p o d a t k i , s katerimi opisujemo pojave iz najrazličnejših socialno-ekonomskih področij. Tako so statistike zbrane v najrazličnejših publikacijah, statističnih letopisih, specialnih statističnih publikacijah, npr. o industrijski statistiki, kmetijski statistiki, statistiki prebivalstva itd.

Kot statistiko razumemo tudi delo pri z b i r a n j u s t a t i s t i č n i h p o d a t k o v. Tako imamo statistiko zunanje trgovine; v njej zbiramo podatke iz področja zunanje trgovine; industrijsko statistiko, ki zbira podatke iz industrijske dejavnosti; kmetijsko statistiko, ki se ukvarja z zbiranjem podatkov iz kmetijstva, itd.

Z imenom statistika obeležujemo tudi o r g a n e , ki zbirajo statistične podatke. Tako pomeni državna statistika mrežo vseh organov, ki zbirajo podatke iz najrazličnejših socialno-ekonomskih področij, da prikaže številčno sliko dogajanja, ki rabi državnim in drugim organom za njihovo delo in analize.

Statistika kot z n a n o s t pa pomeni teorijo in metode statističnega proučevanja. Statistiko kot znanost o p r e d e l i m o takole: statistika je veda, ki s številčnim proučevanjem množičnih pojavov, z metodami, ki so njej lastne, odkriva zakonitosti množičnega pojavljanja in podaja kakovostno analizo pojavov.

Iz te opredelitve spoznamo, da ima statistika svoje področje -množične pojave - in svojo metodo, da s kvantitativnim proučevanjem analizira kvalitativne odnose množičnih pojavov. Tako na primer količina proizvodnje, proizvedena v enoti časa, pokaže produktivnost dela, višina hektarskega donosa uspeh agrotehničnih mer v kmetijstvu, koeficient umrljivosti zdravstveni standard itd.

Ker je v statističnem proučevanju pojavov vedno poudarek na metodi proučevanja, je v uporabi teh metod statistična metoda v ozki povezavi z znanostjo v katero spada proučevan pojav. Zato bi mogli imeti statistiko le za eno izmed metod za proučevanje pojavov. Vendar statistična metoda, razen v nekaterih posebnih primerih, ni bistveno vezana na določeno področje, marveč veljajo njene metode in zakonitosti neodvisno od predmeta za vse množične pojave. Tako matematična statistika, ki razvija metode proučevanja množičnih pojavov in katere osnova so postavke verjetnostnega računa, razvija svoje metode neodvisno od področja uporabe. Zato je statistika kot znanost v ozki zvezi z matematiko, saj se tudi ukvarja s številčnimi odnosi, vendar pa ne s kvantitativnim proučevanjem množičnih pojavov.

Statistiko moramo ločiti od evidence, ker bi mogla v nekaterih področjih nastati zmeda v razmejitvi med statistiko in evidenco. Kakor knjigovodstvo ni statistika, tako tudi evidenca na splošno ni statistika, čeprav se knjigovodstvo in evidenca ukvarjata z množičnimi pojavi. Knjigovodstvo in evidenca sistematično registrirata pojave, ki so množični. To pa je edina stična točka s statistiko. Namen zbiranja podatkov v knjigovodstvu in evidenci je predvsem registrirati posamezne dogodke ali stvari, brez težnje, da bi iz teh podatkov dali sliko ali analizo celote. V statistiki pa je registracija samo sredstvo, ki omogoča, da s statističnimi metodami analiziramo pojav in množico podatkov kot celoto. Posamezen pojav je v statistiki zanimiv samo toliko, kolikor prispeva k tej splošni sliki ali analizi pojava. Knjigovodstvu in evidenci je torej registracija namen, medtem ko je statistiki le sredstvo.

PODROČJA, V KATERIH UPORABLJAMO STATISTIKO X

1.2 Beseda statistika je izšla iz latinske besede "status"-država. Ta izraz je za proučevanje množičnih pojavov zgodovinsko upravičen. Razvojno je statistika res rabila najprej za opisovanje ekonomskih in socialnih razmer v razvitih državah starega in sred-

njega veka. Vendar je ta pojem že dolgo ime za nekaj drugega kakor za metodo opisovanja socialnih in ekonomskih razmer v državi. Statistika je razširila in razvila svojo metodologijo na splošno proučevanje množičnih pojavov, od katerih je mnogo takih, ki jih ne uporabljamo za opis in analizo ekonomskih in socialnih razmer v državi. Razen tega pa je statistika tako razširila področje uporabe, da so socialno-ekonomski pojavi samo eden, čeprav ne najmanjši sektor v uporabi statističnih metod.

Ker so statistiko razvojno dolgo časa uporabljali izključno v socialno-ekonomskih znanostih, je prevladovalo dolgo časa mišljenje, da s statistiko proučujemo le socialno-ekonomske pojave in da se zato vtaplja v socialno-ekonomskih znanostih.

Vendar je statistika osvajala vedno nova področja uporabe. Na prelomu med devetnajstim in dvajsetim stoletjem opazimo velikansko povečanje uporabe statistike v vseh področjih in silen razvoj v statistični metodologiji. Statistika je postala ena izmed osnovnih metod za proučevanje v najrazličnejših znanostih.

Statistika je še vedno ostala ena osnovnih metod za proučevanje v socialno-ekonomskih znanostih. Poleg splošnih metod statistike so razvili tudi take, ki jih uporabljamo izključno za proučevanje socialnih ali ekonomskih pojavov. Prav tako je statistika veliko pripomogla k razvoju in uporabi ekonometričnih metod.

Statistika je tako osnovna metoda v demografskih proučevanjih, da so demografijo celo istovetili z demografsko statistiko.

Metode statistične analize z velikim pridom uporabljamo tudi v biologiji. Problematika biologije je bila v mnogih primerih povod za izdelavo nekaterih statističnih metod, ki so jih v začetku uporabljali samo v biologiji, kasneje pa so se kot splošne metode proučevanja množičnih pojavov razširile tudi na druga področja. Skoraj vsa metodološka plat biometrike je statistika.

Enako je statistika ena izmed metod merjenja v psihologiji, ki so združene v psihometriki. Testiranje, ki je osnova marsikaterega psihološkega proučevanja, je zasnovano na statistiki.

Statistiko z velikim pridom uporabljamo tudi v meteorologiji.

Agronomija je tudi tipično področje za uporabo statističnih metod pri načrtovanju poskusov.

Statistika ima svoje mesto tudi v moderni fiziki. Kinetična teorija plinov, mehanika in radioaktivnost, vse to so področja, v katerih veljajo zakonitosti množičnega pojavljanja in zanje statistika pomaga odkrivati in tolmačiti najrazličnejše zakonitosti.

Statistika si je priborila posebno mesto v sodobni industrijski proizvodnji. Množičnost in tempo proizvodnje je pri kontroli proizvodnega procesa in proizvodnje zahtevala nove metode. Na osnovi vzorčenja, ki ga bomo kasneje podrobneje proučevali, so izdelali specifične in učinkovite metode za kontrolo proizvodnje in proizvodnih procesov.

Statistiko so začeli uporabljati tudi v področjih, za katera bi na prvi pogled bila njena uporaba neutemeljena in nemogoča. S statističnimi metodami analizirajo celo literarna dela, študirajo stile posameznih pisateljev in dob itd.

Čeprav imamo statistiko za samostojno vedo, je v uporabi tesno povezana s predmetom, ki ga s statističnimi metodami proučujemo. Ta zveza se kaže v tem, da pri določenem proučevanju sodelujejo v skupinskem delu strokovnjaki iz področja, iz katerega je problem in strokovnjaki-statistiki. Brez te zveze bi se proučevanje izrodilo v neuporabno preračunavanje številk, ne glede na to, da bi uporabljali razmeroma visoke statistične metode.

SOCIALNO-EKONOMSKA STATISTIKA X

1.3 Kakor smo že navedli, so statistiko začeli uporabljati najprej za proučevanje socialno-ekonomskih pojavov. To je izzvala razmeroma velika potreba po teh podatkih in razmeroma jasno opredeljen predmet proučevanja. Medtem ko ima v drugih znanostih statistika za predmet množične pojave, ki so bolj ali manj umišljeni in je njihovo število neomejeno, proučujemo v socialno-ekonomskih področjih množice realnih pojavov, ki so v času in prostoru v končnem številu.

Splošne metode statističnega proučevanja moramo uporabiti v vseh področjih, kjer naletimo na množične pojave. Vendar so se v socialno-ekonomskih znanostih razen splošnih metod zaradi specifičnih lastnosti pojavov v teh področjih razvile metode, ki so tipične za proučevanje socialno-ekonomskih pojavov in jih v splošnem ne uporabljamo v drugih znanostih. Medtem ko so nekatere metode statistike, ki jih uporabljamo v socialno-ekonomskih področjih, splošnega značaja (srednje vrednosti, mere variacije, korelacija itd.), so druge take, da jih uporabljamo samo ali pretežno za proučevanje socialno-ekonomskih pojavov. Med temi so na primer analiza časovnih vrst, teorija indeksov in konjunkturna statistika.

V tej knjigi obravnavamo splošne metode statističnega proučevanja socialno-ekonomskih pojavov. Posamezne metode obravnavamo na zgledih iz socialno-ekonomskih področij, vendar samo ilustrativno. Te metode na splošno uporabljamo za proučevanje katerega koli socialno-ekonomskega pojava. Tako z istimi statističnimi metodami za proučevanje časovnih vrst proučujemo dinamiko pojavov v demografiji, industriji, kmetijstvu, prometu itd. Enako je s srednjimi vrednostmi, merami variacije, korelacijo in drugimi statističnimi metodami.

1.4 Vendar so kljub nekim splošnim načelom in metodam analize v posameznih socialno-ekonomskih področjih posebnosti, ki se jim mora statistično proučevanje prilagoditi. Zato imamo razen splošne statistike tudi metode statističnega proučevanja, ki so ozko povezane z vsebino posameznih socialno-ekonomskih vej.

Tako imamo posebej:

Demografsko statistiko; ta obravnava posebnosti proučevanja pojavov prebivalstva s statističnimi metodami.

Kmetijska statistika se ukvarja s posebnimi problemi statističnega proučevanja v kmetijstvu.

Industrijska statistika ima tudi svoje posebne prijeme in probleme, ki jih rešujemo s specifičnimi metodami.

Analogno imamo nadalje gozdarsko statistiko, statistiko obrti, grad-

beništva, prometa, trgovine, statistiko gostinstva, turizma, financ, statistiko cen, kulturno-prosvetno statistiko, statistiko zdravstva, sodno statistiko itd.

Vsaka izmed teh statistik se ukvarja z vsebinsko problematiko iz svojega področja in je z njimi v ozki povezanosti. Navedene posebne statistike se pečajo s specifično problematiko opredelitve, zbiranja in analize pojavov na svojih področjih.

DRUGO POGLAVJE

PROUČEVANJE MNOŽIČNIH POJAVOV

MNOŽIČNI POJAVI

2.1 V naravi in družbi pojavi ne nastopajo posamič, marveč v velikem številu, ne izolirano, marveč povezani med seboj. Če opazujemo te pojave individualno in izolirano, se zdi njihovo pojavljanje brez reda. V resnici pa veljajo zanje določene zakonitosti. Te odkrijemo šele, če opazujemo ne samo posamezen pojav, temveč skupnosti pojavov. Omenili smo že, da statistika opazuje in analizira množične pojave. Množičen pa je vsak pojav, ki se v času in prostoru pojavlja v velikem številu. Na take pojave naletimo v različnih področjih, med drugimi tudi v socialno-ekonomskih znanostih. Tako je na primer množičen pojav industrijsko podjetje, kupoprodaja, oseba, predmet, ki ga proizvajamo v množični proizvodnji itd.

Če analiziramo pojave, ki množično nastopajo, odkrijemo, da ti pojavi niso med seboj enaki, marveč se v svojih značilnostih razlikujejo. Podjetja imajo različno število delavcev, različno mesečno proizvodnjo, različno porabo surovin, so iz različnih strok itd. Kupoprodaje se med seboj razlikujejo po času, prodajalcu, kupcu, ceni, količini, kakovosti prodanega blaga itd. Osebe so različnega spola, starosti, stanu, zaposlitve, imajo različno šolsko izobrazbo, mesečne prejemke, število otrok itd. Izdelki so različnih dimenzij, imajo različno kakovost, uporabnost itd., čeprav se zde na oko med seboj enaki.

2.2 Če natančneje proučimo možnosti za analiziranje množičnih pojavov, spoznamo, da je treba področje proučevanja opredeliti. Če na primer opazujemo prebivalstvo, ne moremo hkrati opazovati vsega človeštva v vseh časih, ker je to neizvedljivo, niti ni v posebnem v dani raziskavi zanimivo. Zato se omejimo samo na del prebivalstva tako, da opredelimo, katerim pogojem morajo ustrezati osebe, ki so predmet konkretne raziskave.

Z o p r e d e l j u j o č i m i p o g o j i razmejimo pojave, ki jih proučujemo, od pojavov, ki v konkretnem primeru niso predmet proučevanja. Skupnost pojavov, ki jih opredelimo zato, da jih proučimo, imenujemo statistično množico ali populacijo. Čeprav je beseda populacija privzeta iz pojma prebivalstva, uporabljamo ta izraz za vse vrste statističnih množic. Populacija v statističnem smislu ni samo prebivalstvo, temveč tudi skupnost industrijskih podjetij v Jugoslaviji na določen datum, skupnost izdelkov, proizvedenih na določenem stroju v določenem času itd.

Vsak posamezen pojav populacije imenujemo statistična e n o t a . Tako je v zgornjih populacijah enota posamezen prebivalec, posamezno industrijsko podjetje, posamezen artikel.

Vsaka enota populacije ima veliko značilnosti. Izmed teh pa so samo nekatere predmet konkretnega proučevanja. Značilnosti, ki so v posameznem primeru predmet proučevanja, imenujemo s t a t i s t i č n e z n a k e , ali kratko , z n a k e . Vsak znak ima za posamezne enote različne v r e d n o s t i . Tako niso vsi ljudje enako stari, vsa podjetja nimajo enako število delavstva, vsi izdelki niso enako uporabni itd.

Medtem, ko je spol znak za posamezno enoto, je število oseb v populaciji, ki so moškega spola, značilnost za celotno populacijo, čeprav je ta podatek odvisen od tega, kakšnega spola so posamezni prebivalci v populaciji. Enako je odstotek neuporabnih izdelkov v proizvodnji za določen dan značilnost o celotni proizvodnji, ne pa značilnost za posamezen izdelek, čeprav smo ta podatek izvedli iz podatkov o uporabnosti posameznih artiklov. Značilnosti populacije, kot so npr. število moških v prebivalstvu, odstotek neuporabnih izdelkov v dnevni proizvodnji, poprečen dohodek delavcev v neki stroki, poprečen hektarski pridelek pšenice v socialističnih kmetijskih obratih, itd., imenujemo

parametre populacije.

Enote sestavljajo populacije. Značilnosti enot imenujemo znake, značilnosti populacij pa parametre.

Statistične enote

2.3 Po zgornjem uvodu opredelimo statistično enoto takole: statistična enota je vsak pojav, ki v času in prostoru množično nastopa in je predmet statističnega proučevanja. Po tej opredelitvi more biti enota v statističnem smislu:

a) oseba (prebivalec, delavec, študent),
b) žival (konj, krava, prašič itd.),
c) stvar (avto, motor, izdelek),
d) pravna tvorba (zadruga, društvo),
d) administrativna enota (občina, krajevna skupnost),
f) gospodarska tvorba (industrijsko podjetje, trgovina, kmetijsko gospodarstvo),
g) dogodek (kupoprodaja, smrt, nesreča itd.),
h) poskus (poskusna obdelava parcele, cepljenje živali),
i) trenutek (v katerem opazujemo pojav),

Formalno je važna delitev enot na: a) realne enote, b) dogodke in c) dogajanja. Realne enote so na primer ljudje, podjetja, živina itd., ker v času in prostoru obstajajo. Dogodki so pojavi, ki se v času dogode. Med dogodke štejemo npr. smrt, rojstvo, nesrečo, kupoprodajo. Dogodek se zgodi praktično vzeto v trenutku ali v zelo kratkem času. Vmesna stopnja med realnimi enotami, ki obstajajo in dogodki, ki se dogode trenutno, je dogajanje, ki traja dalj časa. Dogajanje je na primer gradnja hiše, proizvodnja izdelka itd. Zgornja delitev enot je posebno važna pri opredelitvi populacij. Od tega, ali proučujemo populacijo realnih enot, dogodkov ali dogajanj, je namreč odvisna časovna opredelitev populacije.

Statistične enote so bodisi enostavne enote ali pa agregati - skupinice enostavnih enot. Tako štejemo posameznega človeka za enostavno enoto.

Družina, občina, društvo itd. pa so agregati, ker so sestavljeni iz enostavnih enot: članov družine, prebivalcev občine, članov društva itd. Združevanje enostavnih enot v enote višje stopnje-agregate je za statistično proučevanje vsebinsko in tehnično izredno pomembno.

Statistični znaki

2.4 Statistični znaki dajo enotam vsebino. Znaki so tiste značilnosti statističnih enot, ki so predmet proučevanja. Statistične enote oziroma množični pojavi na splošno imajo namreč veliko najrazličnejših značilnosti. V konkretni raziskavi pa podobno kakor iz množice pojavov izberemo populacijo, iz množice vseh možnih značilnosti izberemo tiste, ki so važne za proučevanje in te imenujemo znake. Katere značilnosti so v posebnem primeru znaki, je odvisno od namena proučevanja. Tako je pri splošnem popisu prebivalstva neumestno , da vzamemo za znak število čevljev, zbiramo pa podatke o zaposlitvi, številu otrok itd. Če pa anketiramo prebivalstvo z namenom, da zberemo podatke o velikosti nog za potrebe čevljarske industrije, vzamemo za osnoven znak velikost noge; nobenega smisla pa nima v tem primeru spraševati za poklic, po številu otrok anketirancev itd., ker te značilnosti nimajo zveze s predmetom proučevanja.

2.5 V r s t e s t a t i s t i č n i h z n a k o v . Vsebinsko delimo znake na:

a) krajevne ali geografske, b) časovne in c) stvarne. Krajevni ali geografski znak je lahko kraj, ki je v zvezi z enoto proučevanja ali z dogodkom, ki je značilen za proučevano enoto. Tako je krajevni znak kraj rojstva, kraj stalnega bivališča, kraj zaposlitve, sedež podjetja, kraj nesreče itd.

Č a s o v n i so vsi znaki, ki so v zvezi s časom, v katerem se je zgodil kak dogodek, ki je v zvezi s proučevano enoto. Časovni znak je npr. čas rojstva novorojenčka, čas prodaje izdelka, čas ustanovitve podjetja itd.

Vsi drugi znaki so s t v a r n i . Stvarnih znakov je največ in z njimi označujemo vse značilnosti pojavov, ki niso krajevne ali časovne.

Glede na to, kako izražamo stvarne znake, delimo stvarne znake na: a) atributivne in b) numerične.

A t r i b u t i v n i so znaki, za katere vrednosti izražamo opisno, z besedami. Tako je na primer atributiven znak spol, ker so njegove vrednosti: moški, ženski, izražene z besedami. Iz istega vzroka je atributiven znak tudi zaposlitev, panoga dejavnosti, vrsta zgradbe, vzrok smrti itd.

N u m e r i č n i so znaki, katerih vrednosti izražamo številčno. Po tem, katere vrednosti morejo zavzeti, delimo numerične znake v: a) n e z v e z n e - diskontinuirne in b) z v e z n e - kontinuirne. N e z v e z e n z n a k je na primer število članov v gospodinjstvu, število otrok, ki jih ima mati, število delavcev itd. Nezvezni znaki morejo imeti samo neke, običajno cele vrednosti na določenem razmaku. Tako ne more imeti gospodinjstvo 3 1/2 članov, mati 1 1/4 otrok, podjetje 86 1/2 zaposlenih itd. Z v e z n i z n a k i so tisti numerični znaki, ki morejo teoretično imeti vsako izmed vrednosti v določenem razmaku. Zvezni znaki so na primer starost, teža izdelka, premer profila.

2.6 Vrednosti časovnih in numeričnih znakov se dajo urediti po velikosti po nedvoumnem vrstnem redu. Te lastnosti krajevni in atributivni znaki v splošnem nimajo. Možnost, da lahko uredimo vrednosti znakov po velikosti, je velika prednost časovnih in numeričnih znakov; omogoča namreč poglobljeno analizo, ki je za geografske in atributivne znake ne moremo izvesti. Možnost analize atributivnih znakov v primerjavi z numeričnimi je torej okrnjena. Medtem ko moremo vse metode za analizo atributivnih znakov uporabiti tudi za numerične znake, obratno niso vse metode za numerične znake uporabne za atributivne.

2.7 Iz vsebinskih razlogov, ki so pogoj za uporabo različnih metod statistične analize, je pomemben z n a č a j z n a k o v . Tako govorimo o nominalnosti, ordinalnosti, intervalnosti in razmernosti znakov.

N o m i n a l n o s t (ATRIBUTIVNI) je lastnost znakov, da moremo posamezne vrednosti znakov med seboj razlikovati. Nominalen značaj znakov je torej izražen z $y_1 \neq y_2$. Samo nomi-

nalen značaj imajo stvarno atributivni znaki, po katerih moremo vrednosti znakov samo razlikovati.

O r d i n a l n o s t znakov je v tem, da moremo vrednosti znaka ali enot urediti oziroma razvrstiti po logičnem zaporedju, običajno po velikosti. Ordinalnost znaka je izražena z $y_1 \leqslant y_2$. Z ordinalnimi znaki pogosto tipološko izražamo stopnje kakovosti določenega pojma. Tipičen znak z ordinalnim značajem je npr. kakovost izdelka, ta je izražena z : odličen, uporaben, neuporaben popravljiv, neuporaben nepopravljiv. Enako je z ocenami: odličen, prav dober, dober, zadosten, nezadosten. Po znakih z ordinalnim značajem moremo razvrstiti enote populacije po vrstnem redu, čeprav vrednosti znaka za posamezne enote ni mogoče izraziti numerično. Tako skupino delavcev razvrstimo po prizadevnosti ipd. V tem primeru se izraža ordinalnost v tem, da moremo enote populacije glede na dano enoto razdeliti v enote, ki so manjše, slabše, manj uspešne in v skupino enot, ki so boljše, večje, bolj uspešne kot opazovana enota. Na znake z ordinalnim značajem naletimo pri najrazličnejših raziskavah. Ocenjevanje, preskušanje izdelkov, proučevanje okusa potrošnika, psihološke raziskave ipd. so navezane v veliki meri na opazovanje znakov, ki so ordinalnega značaja.

Vrednosti znakov z i n t e r v a l n i m - r a z m i č n i m z n a č a j e m izražamo numerično. Vendar je zanje tipično, da ti znaki nimajo nekega enoličnega izhodišča.

Tipičen primer intervalnosti znakov je temperatura. Njo merimo z različnimi skalami, ki imajo različna izhodišča. Niz znakov s področja psihologije porabnika, vsi znaki, ki imajo značaj točkovanja, so primeri intervalnih znakov.

Intervalnost znakov se izraža v tem, da moremo za znake z intervalnim značajem izračunavati razlike med vrednostmi znakov $y_2 - y_1 = d$.

Z razmerji med dvema vrednostima znaka je možna primerjava le za znake, ki imajo r a z m e r n o s t e n značaj. Vrednosti takih znakov primerjamo z razmerjem $y_2 / y_1 = R$. Razmernosten značaj imajo znaki, ki imajo enolično izhodišče. Taki znaki so na primer količina, vrednost proizvodnje, velikost posestva ipd. Numeričnih znakov z razmernostnim značajem je v socialno-ekonomskih raziskavah veliko. Vsi pojavi, ki po svojem zna-

čaju ne morejo biti negativni, imajo enolično izhodišče in so zato razmernostni.

Kot je razvidno iz opredelitve značaja znakov imajo razmernostni znaki tudi intervalen, ordinalen in nominalen značaj, intervalnost vključuje ordinalnost in nominalnost, ordinalnost pa nominalnost. Tako imajo vsi znaki nominalen značaj, saj moremo vrednosti za katerikoli znak grupirati v grupe. Če proučimo znake na splošno, spoznamo, da ima večina stvarno atributivnih znakov samo nominalen značaj, da pa ima večina stvarno numeričnih znakov razmernosten značaj, ker moremo iz njih izračunavati razen razlik tudi razmerja.

Najbolj razvite so metode statistične analize za razmernostne znake. Zanje tudi najgloblje prodremo v zakonitosti množičnih pojavov. Čim nižji je značaj znakov, tem bornejša je možna statistična analiza. Že pri merah variacije se izkaže, da so relativne mere variacije smiselne le za razmernostne, ne pa za znake, ki imajo le intervalen značaj.

Čeprav je obseg metod in možnost analize za ordinalne in nominalne znake okrnjena, pa te metode, ki so na splošno vključene v neparametrične metode, dostikrat s pridom uporabljamo tudi pri proučevanju intervalnih in razmernostnih znakov. Neparametrične metode namreč niso vezane na niz predpostavk, katerim mora običajno zadoščati populacija, da moremo uporabljati t.i. parametrične metode. Razen tega so običajno neparametrične metode tudi tehnično računsko manj zahtevne in zato racionalnejše. Gotovo pa je, da se vrednost dobljenih informacij z uporabo neparametričnih metod zmanjša.

2.8 Numerične znake delimo po drugem merilu na: a) ekstenzivne in b) intenzivne. E k s t e n z i v n i znaki nakazujejo količino, i n t e n z i v n i pa kakovost. Tako so ekstenzivni znaki npr. vrednost proizvodnje podjetja v določenem razdobju, število delavcev po podjetjih, količina določenega blaga na trgu itd. Intenzivni znaki pa so na primer cena izdelka, produktivnost dela, merjena z vrednostjo proizvoda na enoto časa, hektarski pridelek itd.

faktorialni - količina
rezultativni = kakovost

2.9 Pomembna je delitev znakov po vlogi, ki jo imajo pri proučevanju množičnih pojavov. V tej zvezi razdelimo znake na: a) f a k t o r i a l n e in b) r e z u l t a t i v n e. Faktorialni znaki so izraz faktorjev-dejavnikov, ki vplivajo na pojav. Rezultativni znaki pa so rezultat vpliva dejavnikov, ki nanje vplivajo. Tako je na primer količina uporabljenega umetnega gnojila faktorialen, hektarski donos pa rezultativen znak. Enako je število delavcev v podjetju faktorialen, vrednost proizvodnje pa rezultativen znak. Količina blaga na trgu je faktorialen, cena blaga pa je rezultativen znak. Zveze med faktorialnimi in rezultativnimi znaki kažejo, kako se spreminja vrednost rezultativnega znaka, če se spreminjajo vrednosti faktorialnih znakov. Proučevanje teh zvez je ena izmed važnih metod v statistični analizi.

2.10 V a r i i r a n j e s t a t i s t i č n i h z n a k o v. Vsak znak ima določeno število vrednosti znakov. Število vseh teoretično možnih vrednosti je najmanj dve, največ neomejeno. Spol ima na primer dve vrednosti: moški, ženski; stan štiri vrednosti: samski, poročen, razvezan, vdovec; poklic veliko, vendar končno število različnih možnih vrednosti, človekova starost pa teoretično neomejeno število vrednosti.

Značilno za znake je, da imajo posamezne enote različne vrednosti znakov. Temu pojavu pravimo v a r i i r a n j e statističnih znakov. Statistični znaki so torej variabilni. Da imajo nekatere enote tudi iste vrednosti enega in istega znaka, ni v nasprotju z opredelitvijo variabilnosti. Vsaka oseba v dani populaciji namreč ne more biti različnega spola, če ima znak spol samo dve vrednosti: moški ali ženski.

Variabilnost statističnih znakov je ena izmed temeljnih lastnosti množičnih pojavov. Statistika se predvsem in v najrazličnejših oblikah ukvarja z variabilnostjo pojavov. Če množični pojavi ne bi bili variabilni, pojavov sploh ne bi proučevali s statističnimi metodami. Zato smo tudi navedli, da ima statistični znak najmanj dve vrednosti, saj značilnost z eno samo vrednostjo ne more biti statistični znak, ker ne variira.

Posebno važna je variabilnost numeričnih znakov. Kakor bomo videli kasneje, je variabilnost numeričnih znakov osnova za najpodrobnejšo proučevanje numeričnih pojavov.

Vrednost znaka variira v populaciji od enote do enote. Vrednost znaka pa se spreminja tudi časovno za isto enoto. Tako se starost spreminja zvezno, kraj zaposlitve pa skokoma, ker je za neko razdobje isti, se pa teoretično ob menjanju zaposlitve v momentu spremeni. Imamo pa tudi znake, katerih vrednost ostane za enote stalna. Tak znak je na primer kraj rojstva, leto ustanovitve podjetja itd.

Statistične populacije

2.11 Opredelitev statistične populacije. Statistična populacija je množica vseh istovrstnih pojavov - statističnih enot, ki izpolnjujejo opredeljujoče pogoje, s katerimi je opredeljena. Statistično populacijo opredelimo tako, da navedemo, kakšne vrednosti znakov, ki so opredeljujoči pogoji, morajo imeti enote, da so enote populacije. V znakih, ki opredeljujejo populacijo, je torej variacija okrnjena. Če proučujemo razmere zaposlenih poročenih žensk na področju Slovenije po stanju 31.12.1972, je s tem že opredeljena populacija, ki jo proučujemo. Proučujemo samo poročene, zaposlene ženske. V stanu, zaposlenosti in spolu populacija ne variira. Enako je z variacijo v časovnem znaku, ker je populacija opredeljena s teoretično trenutnim stanjem konec leta 1972. V krajevnem znaku pa je variabilnost populacije okrnjena, ker proučujemo žene v Sloveniji.

Za populacijo ne velja samo množica vseh statističnih enot v gornjem smislu, marveč je populacija tudi množica podatkov za neki znak. Tako je populacija v tem smislu tudi množica vseh podatkov o starosti poročenih zaposlenih žena v SRS po stanju 31.12. 1972.

2.12 Kakor vidimo iz gornjega zgleda, mora biti populacija nedvoumno opredeljena: a) časovno, b) krajevno in c) stvarno.

Gornjo populacijo smo časovno opredelili tako, da smo navedli: enote populacije so samo tiste ženske, ki zadoščajo vsem opredeljujočim pogojem konec leta 1972. Krajevno je zgornja populacija opredeljena s tem, da navedemo: enote populacije so samo

ženske, ki zadoščajo drugim opredeljujočim pogojem in so v Sloveniji. Stvarno je populacija opredeljena s tremi znaki: s stanom, s spolom in z zaposlenostjo.

V splošnem krajevna opredelitev ne dela posebnih težav. Običajno pri krajevni opredelitvi uporabimo obstoječo upravno razdelitev teritorija. To olajša krajevno opredelitev. Razen tega pa imajo večjo operativno vrednost podatki, ki so dani za območja, skladna z upravno razdelitvijo.

Časovna opredelitev je bolj problematična in je odvisna od narave populacije, ki jo proučujemo.

Zgornjo populacijo žensk smo opredelili s trenutkom, ker je populacija realnih enot. Zato populacije realnih enot včasih tudi imenujemo momentne ali trenutne populacije. Momentnih populacij imamo veliko. Vse časovno enolično opredelimo z momentom, v katerem morajo posamezne enote populacije zadoščati vsem drugim pogojem, da so enote populacije. Če se je katera izmed zaposlenih žena poročila v januarju, ne sodi v populacijo. Enako tudi ni enota populacije poročena žena, ki je bila še zaposlena v decembru 1972, je pa v decembru prekinila delovno razmerje.

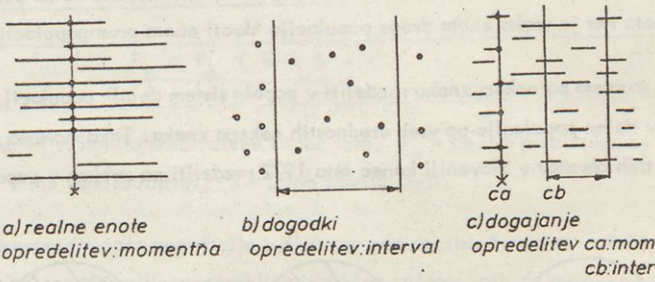
Momentna opredelitev populacij realnih enot je potrebna zaradi nedvoumne razmejitve pojavov v pojave, ki so enote populacije, od pojavov, ki niso enote populacije. Če bi populacijo stvarnih enot skušali opredeliti z razdobjem, te nedvoumnosti ne bi dosegli. Če bi obravnavano populacijo zaposlenih poročenih žena opredelili časovno z mesecem decembrom 1972, bi bilo dvomljivo, katere žene pridejo v populacijo in katere ne. Problematični bi bilo vsi primeri žena, ki so se v decembru bodisi zaposlile ali prekinile delovno razmerje, umrle ipd.

Časovna opredelitev populacij dogodkov z momentom ni smiselna. V določenem momentu se more zgoditi in se zgodi samo omejeno in slučajno število dogodkov. Tako nima smisla iskati populacijo rojstev, nesreč, smrti, proizvodnje itd., ob določenem momentu. Populacijo dogodkov opredelimo z razdobjem in vanjo štejemo vse dogodke, ki so se zgodili v določenem razdobju in zadoščajo vsem drugim pogojem. Te populacije imenujemo intervalne ali razmične populacije, ker so opredeljene s ča-

sovnimi razmiki. Intervalne populacije so npr. proizvodnja, rojstva, nesreče, prodaja itd.

Dogajanja imajo intervalen in momenten značaj. Zato moremo opredeliti populacijo gradenj momentno, če iščemo populacijo poslopij, ki so v gradnji in intervalno, če iščemo populacijo poslopij, ki so jih v določenem razdobju gradili.

Stvarno je obravnavana populacija žensk opredeljena s spolom; ženske, zaposlenostjo: zaposlene in stanom: poročene. Za to populacijo je stvarna opredelitev enostavna. V splošnem pa je stvarno opredelitev populacij najtežja in zahteva dobro poznavanje in proučitev pojava, ki ga raziskujemo. Iz navedenega spoznamo, da v opredeljujočih znakih enote ne varirajo (spol: ženske) ali pa je variiranje okrnjeno (kraji Slovenija).



a) realne enote opredelitev: momentna

b) dogodki opredelitev: interval

c) dogajanje opredelitev ca: moment cb: interval

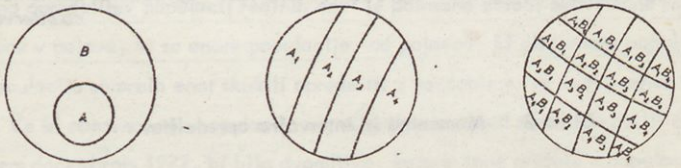
Slika 2.1 Momentna in intervalna opredelitev

2.13 Homogene in heterogene populacije. Po sorodnosti enot so populacije homogene - istorodne in heterogene - raznorodne. V homogenih populacijah enote med seboj niso preveč različne, medtem ko so heterogene populacije sestavljene iz raznorodnih enot. Homogenost populacije dosežemo, če jo opredelimo po čim več faktorialnih znakih. Delitev populacij na homogene in heterogene ni ostra in imamo bolj ali manj homogene in bolj ali manj heterogene populacije. Tako je populacija obrt-

nih podjetij za določeno stroko, v katerih obrtnik razen sebe zaposluje enega samega delavca, bolj homogena kot populacija obrtnih podjetij za isto stroko ne glede na število zaposlenih. Čeprav druga populacija ni heterogena, je vendar manj homogena kot prva. Homogenost populacij je sicer pogojena z opredeljujočimi pogoji, ni pa istovetna. Opredelitev v faktorialnih znakih ima za posledico manjše variiranje oziroma večjo sorodnost ne samo v opredeljujočih temveč tudi v rezultativnih znakih, ki niso opredeljujoči. Krajevna opredelitev kmetijskih gospodarstev pogojuje homogeno populacijo glede na strukturo proizvodnje, način gospodarjenja ipd.

2.14 D e l n e p o p u l a c i j e . Če v zgornji populaciji zaposlenih poročenih žena v Sloveniji konec leta 1972 dodamo pogoj, da je zaposlena kot delavka, iz prve populacije izločimo novo populacijo. To populacijo imenujemo delno populacijo prvotne, zato ker je vsaka enota druge populacije hkrati enota prve populacije.

Populacijo moremo po nekem znaku razdeliti v popoln sistem delnih populacij, če jo razdelimo v delne populacije po vseh vrednostih nekega znaka. Tako moremo populacijo vseh obrtnih obratov v Sloveniji konec leta 1972 razdeliti po strokah v popoln sistem



a) delna populacija A b) razdelitev v delne populacije po A c) dvojna razdelitev v delne populacije po A in B

Slika 2.2 Delne populacije

delnih populacij. Delne populacije so bolj homogene kot osnovna populacija, ker se enote delnih populacij ujemajo še v dodatnem pogoju. Če populacijo razdelimo v popoln sistem delnih populacij po nadaljnjih faktorialnih znakih, heterogeno populacijo, razdelimo v sistem homogenih populacij. To metodo večkrat uporabljamo pri analizi množičnih pojavov.

V sliki 2.2 je nakazana osnovna populacija in delna populacija, razdelitev populacije v popoln sistem delnih populacij po enem in dveh znakih.

2.15 Korespondirujoče populacije. Med nekaterimi momentnimi in intervalnimi populacijami je zveza, ki je nakazana z naslednjim zgledom. Če vzamemo momentno populacijo-prebivalstvo v začetku leta 1968 v SRS, smrt vsakega izmed teh prebivalcev vpliva na momentno populacijo - prebivalstvo konec leta 1968 tako, da se populacija za eno enoto zmanjša. Vsako rojstvo pa ima obraten učinek in se populacija prebivalcev za eno enoto zveča. Vsa rojstva v letu 1968 so intervalna populacija rojstev, vse smrti v tem letu pa intervalna populacija smrti. Iz zgleda vidimo, da sta dve momentni populaciji (prebivalstvo v začetku in prebivalstvo na koncu leta) odvisni od dveh intervalnih populacij (populacije smrti in populacije rojstev v letu). Število enot v posameznih populacijah pa je v tejle zvezi

$$P_1 = P_o + R - S \qquad (2.1)$$

Pri tem pomeni: P_1 = število prebivalcev konec leta; P_o = število prebivalcev v začetku leta; R = letno število rojstev; S = letno število smrti.

Podobno so korespondirujoče populacije tudi druge populacije. Enako se vežejo na primer : začetna zaloga-nabavljeno-porabljeno-končna zaloga; ali: delavstvo na začetku meseca: na novo prišli-odšli-delavstvo na koncu meseca itd.

Statistični parametri

2.16 S statističnimi parametri opisujemo in proučujemo množične pojave. Z njimi številčno izražamo številčne in kakovostne značilnosti in odnose populacij. Nekatere parametre populacij dobimo s preprostim preštevanjem in seštevanjem podatkov za vse enote populacije. Tako je statistični parameter za populacijo kmetijskih gospodarstev število kmetijskih gospodarstev N, ki ga dobimo, če preštejemo vse enote populacije. Parameter "skupna vrednost proizvodnje vseh kmetijskih gospodarstev" dobimo, če se-

štejemo vrednost proizvodnje vseh kmetijskih gospodarstev v populaciji. $Y = \sum y_i$.
Parameter: vsota podatkov za populacijo imenujemo agregat. Parametre z analitično vrednostjo dobimo že s preštevanjem in seštevanjem v delnih populacijah. Tako dobimo na primer s preštevanjem po delnih populacijah število gospodarstev po vrsti proizvodnje, velikosti gospodarstev itd., s seštevanjem podatkov pa vrednost proizvodnje v vseh kmetijskih gospodarstvih po vrsti proizvodnje, velikosti gospodarstev itd.

Te parametre moremo shematično nakazati s simboli:

število enot	N_1	N_2	N_k	N
agregat	Y_1	Y_2	Y_k	Y
	X_1	X_2	X_k	X

pri čemer indeksi nakazujejo, da se število enot ali agregati nanašajo na prvo 1, drugo 2, ... k- to delno populacijo.

Drugi parametri, kot so relativna števila, srednje vrednosti, mere variacije, mere koncentracije, mere sploščenosti in mere asimetrije, so parametri, s katerimi opisujemo in analiziramo predvsem posamezne populacije. S korelacijsko analizo pa dobimo parametre, s katerimi analiziramo odnose med več različnimi statističnimi populacijami.

ZNAČILNOSTI PRI PROUČEVANJU MNOŽIČNIH POJAVOV

2.17 Če primerjamo proučevanje pojavov v fiziki, kemiji ali agronomiji s proučevanjem socialno-ekonomskih pojavov, opazimo med njimi nekatere podobnosti, a tudi razlike.

V fiziki, kemiji in v agronomiji proučujemo pojave večinoma s poskusi. Poskuse v f i z i k i in k e m i j i delamo laboratorijsko v natančno določenih pogojih, da čimbolj odstranimo vse vplive, ki bi mogli poskus motiti. Če bi mogli odstraniti vse dodatne vplive in rezultate meriti z idealno točnimi merilnimi instrumenti, bi zadoščal en sam poskus, ker bi dobili pri vsakem poskusu ob enakih pogojih isti rezultat. Vendar v prak-

tičnih primerih rezultati pri ponovitvah niso povsem enaki. Pokažejo se majhne razlike, ker je nemogoče poskus ponoviti natančno enako. Zaradi dodatnih vplivov, ki jih z obstoječimi poskusnimi sredstvi ne moremo odpraviti, so rezultati med seboj različni. Te razlike imenujemo slučajne razlike, faktorje, ki jih povzroče, pa slučajne faktorje. Vendar se te razlike v velikem številu poskusov pojavljajo v nekih zakonitostih. Zaradi slučajnih faktorjev se rezultati posameznih poskusov odklanjajo od prave vrednosti navzgor in navzdol. Po zakonu o velikih številih pa se slučajnostni odkloni v vsoti rezultatov iz velikega števila poskusov izravnavajo. Zato je poprečen rezultat bliže rezultatu, ki bi ga dobili, če ne bi bilo subjektivnih momentov eksperimentatorja, nemožnosti popolne določitve poskusnih pogojev in napak v meritvah. Da dobimo objektivnejše poprečje, je dobro, če vsak poskus izvedejo različni eksperimentatorji z različnimi merilnimi instrumenti. Tako merilne priprave kot eksperimentator lahko povzročijo sistematične napake. Napaka v instrumentu je lahko take narave, da sistematično daje vedno premajhne rezultate. Isto more biti tudi z eksperimentatorjem. Če ponavljamo poskuse z istimi merilnimi instrumenti ali eksperimentatorji, se sistematične napake ne izravnajo in poprečen rezultat se ne približa dejanski vrednosti - je pristranski.

Podobno je pri poskusih v drugih znanostih. Za razliko od laboratorijskih fizikalnih poskusov je na primer pri poskusih v a g r o n o m i j i kompleks slučajnih faktorjev močnejši kot pri fizikalnih poskusih. Mikrosestav zemlje, razlike v semenu, lega parcele itd. so faktorji, ki jih pri posameznih poskusih ne moremo natančno izenačiti. Pri poskusih v agronomiji moramo na primer določen gnojilni poskus izvesti v več ponovitvah, da zmanjšamo vpliv slučajnih faktorjev, katerih iz objektivnih razlogov ne moremo odstraniti pri posamičnih poskusih.

Kakor pri fizikalnih poskusih tako smo tudi pri poskusih v agronomiji ugotovili isto važno dejstvo. Zakonitost, ki jo preizkušamo s poskusom, v posamičnem primeru motijo slučajni vplivi. V veliki množici istovrstnih poskusov pa se učinek slučajnih vplivov v poprečju izravna in se zaradi delovanja zakona o velikih številih pokaže tipičnost pojava. Ta postopek je povsem specifičen. Točnejših rezultatov pri poskusih ne dosegamo z izboljšavo laboratorijske tehnike, temveč s ponavljanjem poskusov ob danih pogojih iz rezultatov skupine poskusov dobimo natančnejši rezultat. Taka pot je v dosti primerih

ustrezna, ker je običajno ceneje ponavljati poskus z danimi sredstvi, kakor pa z dragimi aparaturami neposredno odstranjevati slučajne vplive. V primeru, da dosežemo meje možnosti pri opredelitvi pogojev, so pa slučajni vplivi kljub temu močni, kakor je to npr. pri poskusih v agronomiji, edino s ponovnimi poskusi oziroma poprečji dosežemo natančnejše rezultate.

V fiziki, kemiji in agronomiji preskušamo učinek nekega faktorja tako, da spremenimo določilne pogoje in raziskujemo učinek spremenjenih pogojev na rezultat. Če proučujemo pri prostem padu odvisnost med časom padanja in potjo, spreminjamo višino, s katere spuščamo kroglico in proučujemo čas padanja kroglice iz različnih višin. Prav tako pri poskusih v agronomiji preskušamo učinkovitost določenih agrotehničnih mer na hektarski pridelek tako da izvedemo vrste poskusov ob spremenjenih pogojih in proučujemo razlike v hektarskem pridelku zaradi spremenjenih pogojev.

Iz gornjih zgledov sklepamo, da je z množičnim opazovanjem pojavov možno preseči meje natančnosti, ki jih dosežemo s posamičnim opazovanjem.

2.18 Če skušamo postopke, ki smo jih navedli kot primerne za proučevanje v fiziki in agronomiji, prenesti v s o c i a l n o - e k o n o m s k e z n a n o s t i, spoznamo neke bistvene razlike v možnosti opazovanja. V fiziki in agronomiji smo za proučevanje določene zakonitosti izvedli poskuse v laboratoriju ali na polju. Predmet opazovanja v socialno-ekonomskih znanostih pa po pravilu ne moremo ustvariti s poskusom. Če proučujemo, kako je struktura porabe odvisna od skupnih dohodkov pó gospodinjstvih, ne moremo zaradi te raziskave sestaviti poskusnih gospodinjstev z natančno določeno sestavo članov, jim odrediti skupne dohodke in opazovati razlike pri različnih skupnih dohodkih. Enako ne moremo zaradi raziskave o višini proizvodnje v podjetjih kovinske stroke z 51 do 100 zaposlenimi delavci zaradi raziskave ustanoviti niz podjetij kovinske stroke, ki imajo 51 do 100 zaposlenih delavcev in opazovati njihovo proizvodnjo. V socialno-ekonomskih znanostih v splošnem enot opazovanja ne moremo (razen v nekaterih posebnih primerih) dobiti s poskusi kakor v fiziki in agronomiji. Tu uporabljamo druge metode. Življenje samo ustvarja pojave v najrazličnejših oblikah. Da proučimo, kako je odvisna sestava porabe od višine dohodkov, ne sestavimo poskusnih družin, temveč iz množice

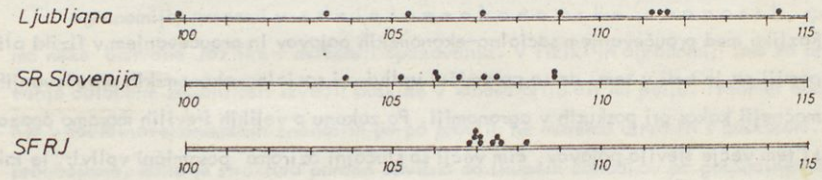
obstoječih gospodinjstev poiščemo gospodinjstva, ki ustrezajo pogojem za naše raziskovanje. Če proučujemo višino proizvodnje podjetij v kovinski stroki, ki imajo zaposlenih od 51 do 100 delavcev, ne ustanavljamo podjetij zaradi proučevanja, temveč iz obstoječih industrijskih podjetij poiščemo podjetja iz kovinske stroke, ki imajo zaposlenih od 51 do 100 delavcev in zanje proučimo proizvodnjo. To je ena izmed temeljnih razlik med proučevanjem s poskusi in med proučevanjem, ki je običajno v socialno-ekonomskih raziskavah.

Vendar razmejitev ni tako ostra in moremo v določenih primerih in za določene potrebe tudi v socialno-ekonomskih proučevanjih uporabiti poskusno metodo. Če proučujemo odvisnost prodaje nekega izdelka od embalaže in načina prodaje, moremo to problem raziskati s poskusom v pravem pomenu besede. V določeni veleblagovnici organiziramo prodajo tega izdelka po natančno določenem načrtu: koliko časa, kdaj in kje prodajamo določeni izdelek na en ali drug način. Ta razpored more biti sestavljen kot čisti poskus prav tako, kakor načrtujemo poskuse v agronomiji ali drugje. Takih zgledov bi mogli našteti veliko. Vsi ti poskusi pa so večinoma manjšega obsega in pomena. S poskusi more na primer podjetje zase proučevati učinke različnih pogojev na produktivnost dela, probleme notranjega transporta, učinek odmora na delovno storilnost itd.

Razlika med proučevanjem socialno-ekonomskih pojavov in proučevanjem v fiziki ali agronomiji pa je tudi v tem, da so posamični vplivi pri socialno-ekonomskih proučevanjih še močnejši kakor pri poskusih v agronomiji. Po zakonu o velikih številih moremo opazovati tem večje število pojavov, čim večji so slučajni oziroma posamični vplivi; le tako se rezultati posamičnih vplivov izravnajo in se pokaže tipičnost pojava. Izraz slučajni vplivi pri proučevanju socialno-ekonomskih pojavov raje zamenjujemo z izrazom posamični vplivi, ker ima ta izraz širši pomen. Medtem ko so slučajni vplivi tisti, ki jih ne moremo kontrolirati oziroma določiti, so posamični vplivi vsi tisti vplivi, katerih učinki se od enote do enote populacije spreminjajo. Zato moramo pri proučevanju socialno-ekonomskih pojavov opazovati razmeroma velike populacije, če hočemo, da se po zakonu velikih števil pokaže tisto, kar je za pojav tipično, netipično pa odstrani oziroma izravna.

2.19 Vzemimo za ponazoritev zakona o velikih številih proučevanje spolnega indeksa za novorojenčke. Spolni indeks je razmerje med številom rojenih dečkov in deklic v določenem času na določenem ozemlju. To razmerje je demografska konstanta, ki se ne spreminja niti časovno niti krajevno. Jasno je, da je razmerje med rojenimi dečki in deklicami za posamezno družino zelo različno. So družine, v katerih so vsi otroci dečki in družine s samimi deklicami. Število rojstev po družinah je premajhno, da bi se pokazala zakonitost v spolnem sestavu novorojenčkov. Da proučimo stabilnost spolnega indeksa v odvisnosti od velikosti populacije, vzemimo spolni indeks za mesto Ljubljano, Slovenijo in Jugoslavijo v letih 1947-1955.

Leto	1947	1948	1949	1950	1951	1952	1953	1954	1955
Ljubljana	111.3	114.4	107.1	103.3	97.9	109.0	111.6	105.3	111.2
Slovenija	107.5	106.0	105.9	107.2	105.2	106.6	108.8	103.8	108.8
Jugoslavija	108.2	107.3	106.7	107.5	107.4	106.8	107.0	106.8	107.0



Slika 2.3 Spolni indeks novorojenčkov

Iz podatkov in iz slike je razvidno, kako raste stabilnost podatkov, čim večje so populacije, ki jih proučujemo. Medtem ko je število rojstev v Ljubljani po letih ca 2500 letno, je število rojstev v Sloveniji poprečno 33 tisoč, v Jugoslaviji pa 470 tisoč na leto. Skladno z velikostjo populacij je za Ljubljano razlika med največjim in najmanjšim letnim vitalnim indeksom $R = 114,4 - 97.9 = 16,5$, za Slovenijo $R = 108,8 - 103,8 = 5,0$, za Jugoslavijo pa $R = 108,2 - 106,7 = 1,5$.

Če izračunamo poprečni spolni indeks iz podatkov za Jugoslavijo v vseh devetih letih 1947-1955, dobimo, da je spolni indeks, izračunan iz 4 milijonov 230 tisoč rojstev, enak 107,2. V razdobju devetih let 1931-1939 pred vojno je bil spolni indeks za Jugoslavijo tudi 107,2.

Iz zgleda lepo vidimo veljavnost zakona o velikih številih. Zakonitosti se pokažejo v množičnem opazovanju, medtem ko v posamičnih primerih oziroma v malem številu primerov prevladujejo posamični vplivi, ki splošno zakonitost zabrišejo.

Iz zgleda pa je razvidno to, kako se le za velike populacije pokaže zakonitost v spolnem indeksu. To je vselej, kadar so posamični vplivi zelo močni.

Vendar dostikrat nimamo možnosti opazovati tako velikega števila primerov kakor v zgornjem zgledu, ko smo analizirali spolni indeks novorojenčkov. Zato se za male populacije omejimo navadno le na statistično opisovanje množičnega pojava, ne iščemo pa zakonitosti. Statistično je proučevanje socialno-ekonomskih pojavov včasih omejeno le na opisovanje populacij.

ETAPE STATISTIČNEGA PROUČEVANJA

2.20 Statistično proučevanje socialno-ekonomskih pojavov je razmeroma kompleksen in obsežen posel in sestoji iz več etap. Te so po značaju in vlogi bistveno različne, vse so pa med seboj tesno povezane. Izvajanje vsake izmed njih je odvisno od metode in izvedbe v prejšnjih stopnjah. Enako je izvedba posamezne etape odvisna od tega, kako bomo obdelovali statistične podatke v naslednjih stopnjah. Zato je potreben za vsako statistično proučevanje dobro premišljen splošni načrt, ki določi osnovne metode in smernice dela v posameznih etapah.

Statistično proučevanje moremo v glavnem razdeliti v tele velike, po funkciji različne etape:

a) določitev vsebine in namena statističnega proučevanja z analizo pojava, ki ga proučujemo,

b) izdelava splošnega načrta statističnega proučevanja,
c) opazovanje v ožjem smislu,
d) urejevanje in osnovna obdelava,
e) analitična obdelava in analiza podatkov.

2.21 Za vsako statistično proučevanje je potrebno najprej določiti namen in vsebino proučevanja. Natančno moramo opredeliti pojav, ki ga proučujemo in določiti znake, ki jih bomo proučevali. Če hočemo proučevati delovno silo v kmetijstvu, je treba natančno določiti obseg proučevanja. Lahko se omejimo samo na delovno silo v privatnem sektorju, na delovno silo, ki je zaposlena izključno v kmetijstvu, itd. Enako je od namena proučevanja odvisno, ali se bomo omejili pri proučevanju samo na številčno stanje delovne sile v kmetijstvu, ali nas zanima stanje po velikostnih skupinah za različne vrste kmetijskih gospodarstev, ali so za proučevanje važne tudi življenjske razmere itd. Le če je natančno postavljen namen statističnega proučevanja, moremo predmet opredeliti in določiti posamezne pojme.

Glede na namen in potrebe proučevanja je treba že v začetku določiti, ali je zadostno, da dobimo za posamezne proučevane pojave le ocene.

Ko določimo vsebino pojava in namen statističnega proučevanja, postavimo splošen načrt, ki povezuje vse nadaljnje stopnje. Pri tem nastopijo že v tej stopnji predvsem problemi z opredelitvijo pojmov; ti pa so osnovnega pomena tako za tehnično izvajanje kakor za analizo rezultatov. Ni brez vsega razumljivo, kaj je kmetijsko gospodarstvo, koga imamo za zaposlenega v kmetijstvu itd. Zato je treba vsak pojav in pojem, ki ga v proučevanju uporabljamo, natančno opredeliti. To je potrebno, da bodo opredelitve nedvoumno razumljive vsem, ki sodelujejo pri izvajanju opazovanja in vsem, ki podatke kasneje uporabljajo. Nič ne koristijo s še tolikšnim trudom in stroški zbrani podatki, če ne vemo, kaj ti podatki pravzaprav pomenijo. V prvi stopnji je potrebno podati tudi vsebino proučevanja. Statistično to pomeni, da moramo postaviti osnovna vprašanja, ki naj jih obravnava statistično proučevanje, če naj izpolni svoj namen.

Ko postavljamo namen in vsebino statističnega proučevanja, sodelujejo pri tem predvsem

porabniki statističnih podatkov in strokovnjaki iz področja, iz katerega je problem, ki ga nameravamo analizirati.

2.22 Splošni načrt statističnega proučevanja. Glede na postavljeni namen in vsebino proučevanja pojava v splošnem načrtu določimo osnovne smernice dela v vseh nadaljnjih stopnjah. Brez predhodnega splošnega načrta bi bilo delo v posameznih stopnjah neučinkovito. Od tega, kako bomo podatke obdelovali, je odvisno že statistično opazovanje. Enako je obdelava statističnih podatkov odvisna od tega, kakšna bo analitična obdelava podatkov itd. Naloga splošnega plana je, da glede na namen opazovanja in razpoložljiva sredstva določi metodo dela v posameznih stopnjah, tako da dobimo čim bolj zadovoljiv odgovor na problem, ki ga proučujemo.

Zahteve, ki jih postavljamo splošnemu načrtu, ni lahko izpolniti. Zato je za sestavljanje splošnega načrta potrebno poznavanje teoretičnih in tehničnih prijemov v vseh stopnjah statističnega proučevanja, tako v opazovanju kakor pri obdelavi in analizi.

V splošnem načrtu moramo glede na namen in razpoložljiva materialna in finančna sredstva določiti metode opazovanja, način osnovne in analitične obdelave in način prikazovanja oziroma publiciranja podatkov. Postavke splošnega načrta moramo upoštevati pri sestavljanju operativnega načrta v posameznih stopnjah.

2.23 Statistično opazovanje v ožjem smislu. Statistično opazovanje v ožjem pomenu besede je zbiranje statističnih podatkov o statističnih enotah. S statističnim opazovanjem zberemo podatke o populaciji v obliki, ki je primerna za osnovno obdelavo. Ločimo več vrst statističnih opazovanj. Vsako izmed njih ima svoje tipičnosti glede na metodo dela in glede na to, kakšna je vrednost in zanesljivost rezultatov, ki jih z njim dobimo. Izvajanje statističnega opazovanja je bolj tehničnega kot pa vsebinskega značaja. V zvezi z njim moramo rešiti niz tehničnih problemov, katerih teža je odvisna od populacije, ki jo proučujemo in vrste opazovanja, ki ga uporabljamo.

2.24 Osnovna obdelava. Osnovna obdelava sestoji iz urejevanja, prešteva-
nja in seštevanja podatkov, ki smo jih zbrali s statističnim opazovanjem po da-
nih shemah grupiranja. Ta stopnja zahteva za velike populacije razmeroma veliko časa,
kar gre v škodo uporabnosti podatkov. Z mehanizacijo čas obdelave znatno skrčimo. Če
pomislimo, da pri popisu prebivalstva obdelujemo na milijone popisnih obrazcev, da
pri obdelavi sodeluje na stotine ljudi, da obdelava traja tudi pri mehanizirani obdelavi
več let, si moremo predstaviti razsežnost dela pri osnovni obdelavi za velike populaci-
je.

2.25 Analitična obdelava. V osnovni obdelavi dobimo podatke urejene, pre-
štete in seštete. Vendar ti podatki še ne zadostujejo za analizo pojava. Mno-
žične pojave analiziramo z obdelavo osnovnih absolutnih podatkov s posebnimi metoda-
mi, ki jih bomo v nadaljevanju podrobneje obravnavali. Iz osnovnih podatkov namreč
dalje izračunavamo relativna števila, srednje vrednosti, mere variacije, mere korela-
cije, pokazovalce dinamike pojavov ipd. s katerimi prikažemo in analiziramo zakoni-
tosti množičnih pojavov.

Iz navedenega je bolj razumljivo, kako potrebna je povezava vseh stopenj proučeva-
nja; samo tak postopek zagotavlja uspešno delo in sklepe statističnega proučevanja.

TRETJE POGLAVJE

STATISTIČNO OPAZOVANJE

3.1 Statistično opazovanje, ki obsega zbiranje podatkov o proučevani populaciji, je osnova statističnega proučevanja. Od tega, ali je ta osnova dobra ali slaba, je odvisna kakovost celotnega proučevanja ne glede na to, kako je opravljeno delo v nadaljnjih stopnjah. Iz slabih osnovnih podatkov je s še tako dobrimi metodami obdelave nemogoče proučevani pojav pravilno analizirati.

Problemi pri izvedbi statističnega opazovanja izvirajo iz tega, ker je delo v tej stopnji pri proučevanjih večjega obsega navezano na veliko ljudi, ki nimajo posebne statistične izobrazbe; razen tega so krajevno razkropljeni; pri svojem delu, ki ga je treba hitro in odgovorno izvesti, so v večini primerov navezani le sami nase in na pisana splošna navodila. Ker se opazovane populacije običajno hitro spreminjajo, je po izvedenem opazovanju težko znova iskati podatke, če se izkaže, da so napačni. Še nevarneje pa je, da veliko napak v opazovanju ni mogoče odkriti. Statistično opazovanje je zaradi obširne organizacije in udeležbe mnogih sodelavcev običajno tudi razmeroma drago. Vse to govori za to, da je treba predhodno dobro premisliti, za kakšno obliko opazovanja se odločimo in kako ga izvedemo.

OPREDELITEV PREDMETA OPAZOVANJA - STATISTIČNE POPULACIJE

3.2 Že v programu statističnega proučevanja v načelu opredelimo enoto proučevanja oziroma populacijo, ki jo opazujemo. Vendar moramo pred izvedbo statističnega opazovanja populacijo in enote opazovanja opredeliti tako, da je nedvoumno in

preprosto razumljivo, kateri pojavi so predmet opazovanja in kateri ne. Populacijo moramo, kakor smo že nakazali, opredeliti: stvarno, krajevno in časovno.

Stvarna opredelitev predmeta opazovanja je najtežja. Dostikrat je težko v nekaj stavkih določiti enoto opazovanja tako, da je v vsakem primeru nedvoumno jasno, ali določen pojav spada v populacijo ali ne. Predmet opazovanja je neredko opredeljen tako, da moramo poznati več podatkov, da ugotovimo, ali pojav ustreza opredeljujočim pogojem ali ne. Tako npr. je težko opredeliti, kaj je kmetijsko gospodarstvo, kaj industrijsko podjetje, še teže pa, koga imamo za srednjega ali velikega kmeta. Opredelitve vseh teh pojavov so sestavljene iz več značilnosti pojava.

Včasih se zaradi enostavnosti in nedvoumnosti, kaj zbrani podatki pomenijo, omejimo na en sam znak in z njim opredelimo enoto opazovanja. Tako smo v letu 1946 vključili v popis industrijskih podjetij kot industrijska podjetja vsa proizvodna podjetja z več ko petimi delavci. Leta 1947 pa smo popisali kot kmetijsko gospodarstvo vsako posest z nad 500 m^2 zemlje. Čeprav zgornji definiciji nista natančni, smo z njima dosegli vsaj enoličnost in nedvoumnost tako pri opazovanju kakor pri tolmačenju podatkov. Meji pet delavcev za popis industrije in 500 m^2 za popis kmetijskih gospodarstev imenujemo c e n s u s n o r m o . S census normo enostavno z enim samim znakom vsebinsko opredelimo pojav. Čeprav ta način teoretično ni najboljši, ima take tehnične prednosti, da ga često uporabljamo.

Krajevna opredelitev je običajno brez težav, ker se opremo na obstoječe občinske, republiške ali državne meje.

Časovna opredelitev je tudi brez posebnih težav in pri popisih opredelimo populacijo časovno s teoretičnim momentom - k r i t i č n i m m o m e n t o m, pri tekočih registracijah pa navadno s koledarskim razdobjem.

VRSTE OPAZOVANJ

3.3 Glede na proučevano populacijo uporabljamo različne vrste statističnega opazovanja. Medtem ko populacije realnih enot, kot so prebivalstvo, stavbe, gospodarstva,

živina itd., opazujemo s p o p i s i , populacije dogodkov, kot so smrti, rojstva, nesreče, itd., opazujemo s t e k o č o r e g i s t r a c i j o . Glede na potrebe, materialne in finančne možnosti pa opazujemo v s e e n o t e p o p u l a c i j e , če potrebujemo popolno in podrobno sliko o populaciji, ali samo d e l p o p u l a c i j e , če se zadovoljimo z ocenami.

3.4 L o n g i t u d i n a l n o i n t r a n s v e r z a l n o o p a z o v a n j e

Uvodoma smo navedli, da dobimo z opazovanjem določenega pojava v različnih časovnih trenutkih sliko o časovnem spreminjanju oziroma gibanju pojava. Taka t.i. transverzalna opazovanja samo delno prikažejo gibanje pojava. Populacija realnih enot se časovno spreminja zaradi različnih dogodkov, ki se dogode v vmesnem razdobju med dvema opazovanjema in spreminjajo značilnosti enot oziroma populacij. Te spremembe pa se morejo v določenem razdobju delno ali v celoti izravnati.

Transverzalna opazovanja v določenih časovnih razmikih sicer pokažejo spremenjeno sliko, ki je projekcija sprememb populacije v določenem trenutku, vendar je ta projekcija samo delno slika sprememb oziroma dogodkov, ki so vplivali na populacijo. Vzemimo shematičen zgled populacije, ki jo opazujemo po treh vrednostih določenega znaka: A, B in C. Možni dogodki, ki spremene sestavo populacije realnih enot po tem znaku so: enota ne preide v drug razred ali enota preide v enega izmed drugih razredov.

Če sestavimo matriko teh dogodkov, dobimo:

Stanje 1	dogodek med 1 in 2		
A	AvA	AvB	AvC
B	BvA	BvB	BvC
C	CvA	CvB	CvC
Stanje 2	A	B	C

V numeričnem zgledu vidimo kakšno sliko moremo dobiti v praksi.

I

stanje 1		1v2		
A	20	20	0	0
B	30	0	30	0
C	50	0	0	50
	100	20	30	50
stanje 2		A	B	C

II

stanje 1		1v2		
A	20	15	3	2
B	30	2	15	13
C	50	3	12	35
	100	20	30	50
stanje 2		A	B	C

Transverzalna opazovanje v obeh trenutkih pokaže enako stanje in sestavo populacije. S tem pa še ni rečeno, da v vmesnem razdobju ni bilo sprememb. Slika I pokaže situacijo, v kateri ni bilo sprememb, slika II pa kaže situacijo, v kateri se je populacija spreminjala, kljub temu pa smo dobili enako sestavo.

Opazovanje, v katerem proučujemo dinamiko populacij prek sprememb posameznih enot v določenem razdobju, imenujemo longitudinalna opazovanja. Prikazani zgledi so samo shematični primeri longitudinalnega opazovanja. V vmesnem razdobju more na posamezno enoto vplivati ne en sam dogodek temveč več dogodkov, ki spremene značilnosti enot. V takem primeru je seveda shema in opazovanje znatno bolj zamotano.

Enota	1	AAAAAAAAABBBBBBBBBB
	2	AAAAAAAAAAAAAAAAAAAA
	3	BBBBBBBBBAAAAAAAAAA
	4	AAAAAABBBBBBBBBBBBB
	5	BBBBAAAAAAABBBBBBBB
	6	AAAAAABBBBBBBBBBBAA

Če proučimo sestavov začetku, dobimo: 4xA in 2xB. Ob koncu pa je stanje: 3xA in 3xB. To sta rezultata transverzalnih opazovanj v začetku in na koncu razdobja. Če pa proučimo celotno razdobje, pa vidimo, da sta se v transverzalnem opazovanju enota 1 in 3 kompenzirali. Longitudinalno proučevanje med začetkom in koncem pa da naslednji rezultat:

1			
A	4	2	2
B	2	1	1
	6	3	3
	2	A	B

Vendar iz longitudinalnega pregleda med začetkom in koncem izpade iz enote 5 vmesen prehod iz B v A in v B nazaj in iz enote 6 vmesen prehod iz A v B in iz B v A.

Popis

3.5 Če potrebujemo popolno in podrobno sliko pojava, opazujemo populacije stvarnih enot s popisom. Popis je zgodovinsko najstarejša in še danes pogosta oblika statističnega opazovanja.

Osnovna značilnost popisov je, da popišemo vse enote populacije po stanju v določenem trenutku - kritičnem momentu. Popisati vse enote populacije je ideal, ki se mu bolj ali manj približamo. Odstotek, ki pove, kolik del populacije smo uspeli popisati, meri stopnjo zajetja. Odstotek zajetja je odvisen od proučevane populacije in uspešnosti organizacije opazovanja. Populacije, ki se gibljejo, je teže popolno popisati kakor populacije, ki se malo gibljejo ali pa mirujejo. Popolnost zajetja pri popisu prebivalstva je zaradi gibanja prebivalstva večji problem kakor pri popisih zgradb, popisih kmetijskih gospodarstev itd.

3.6 S popisom dobimo podobno kakor s fotografskim posnetkom trenutno sliko populacije. Nemogoče je npr. s statističnim popisom obseči prebivalstvo v mesecu ali letu, ker je stanje prebivalstva vsak hip drugačno.

Zato pri vsakem popisu določimo trenutek, na katerega se nanašajo zbrani podatki. Ta trenutek imenujemo k r i t i č n i t r e n u t e k . Čeprav populacije ne moremo opazovati natančno v kritičnem trenutku in popisujemo enote po kritičnem trenutku, jo vendar

skušamo popisati po stanju, kakršno je bilo v kritičnem trenutku. Kako dolg naj bo čas popisovanja, ki se ne ujema s kritičnim trenutkom, je odvisno od populacije in tehničnih možnosti popisovanja. Posebno za populacije, ki se gibljejo, čas popisovanja ne sme biti preveč oddaljen od kritičnega datuma, niti predolg, ker drugače težko ugotovimo stanje ob kritičnem trenutku. Čas popisovanja pri popisu prebivalstva mora biti razmeroma kratek in morajo popisovalci popisati prebivalstvo v nekaj dneh po kritičnem trenutku. Pri popisu zgradb pa npr. ni nujno, da je tako kratek, ker je populacija zgradb znatno manj dinamična kakor prebivalstvo.

Kritični trenutek je odvisen od vsebine pojava, je pa odvisen tudi od tehničnih možnosti popisovanja. Iz vsebinskega stališča naj popis zajame populacijo v hipu, ko je ta v normalnem stanju ali v stanju, ki je za proučevanje važno in značilno. Tako je npr. za živino značilno maksimalno stanje živine. Zato so v Jugoslaviji popisi živine v mesecu januarju, v katerem je stanje večine vrst živine maksimalno.

Iz tehničnega vidika je najprikladnejši čas za popis takrat, ko je populacija v takem stanju, da zanjo najlaže zberemo podatke. Za živino je neprikladen čas popisovanja poletje, ko je velik del živine izven kmetijskih gospodarstev na paši v planinah. Čeprav bi za popis prebivalstva iz vsebinskih razlogov najbolj ustrezal kritični trenutek 31. december, prebivalstva ne popisujemo v tem času, ker bi verjetno za ta datum težko izvedli popis. Iz tehničnih in vsebinskih razlogov ne ustreza tudi poletje, ko je čas dopustov. Niti regionalna porazdelitev prebivalstva niti tehnična možnost nista v tem času za popisovanje prebivalstva najugodnejši. Zato kljub temu, da je 31. marec iz vsebinskih razlogov neprimeren, vzamemo ta datum za kritični trenutek popisa, ker je na ta datum prebivalstvo v večini primerov v krajih svojega stalnega bivališča in ni takih sezonskih premikov kakor poleti ali neugodnih okoliščin popisovanja, kakršne so okrog Novega leta. Izkaže se, da je enostavneje podatke po stanju na dan 31. marca po potrebi s podatki o naravnem in mehanskem gibanju prebivalstva prevesti na stanje v začetku ali v sredini leta, kakor pa popisovati ob neprimernem času.

V vsakem popisu je pri določitvi kritičnega trenutka potrebno najti ravnotežje med vsebinskimi in tehničnimi argumenti.

3.7 S popisom dobimo trenutno sliko pojava. Vendar je za proučevanje populacij dostikrat zelo važna d i n a m i k a . Te pa enkratni popis ne da. Z več zaporednimi popisi v določenih časovnih razmikih, podobno kakor v kinu iz slik v časovnem zaporedju, dobimo vtis o gibanju. Če hočemo, da dajo rezultati zaporednih popisov osnovo za proučevanje dinamike, moramo upoštevati tale tri načela:

Za populacije, ki se časovno hitro spreminjajo, mora biti čas med dvema zaporednima popisoma krajši kot pri populacijah, ki se spreminjajo počasi. To je eden izmed razlogov, da imamo npr. popise prebivalstva vsakih pet ali deset let, popise živine pa vsako leto.

Razdobja med zaporednimi popisi morajo biti enaka, če hočemo iz rezultatov popisov neposredno sklepati na jakost sprememb. To načelo je razumljivo. Razumljivo je, da se pri isti intenziteti sprememb populacija v daljšem razdobju bolj spremeni kot v krajšem razdobju. Zato so popisi prebivalstva v enakih razmikih vsakih pet ali deset let, ne pa enkrat po enem letu, drugič po petih letih, tretjič po desetih letih itd.

Če je pojav, ki ga proučujemo, sezonski, kar pomeni, da je odvisen od letnega časa, moramo popisovati vsako leto na isti datum, če proučujemo dinamiko pojava na daljša razdobja. Če tega ne storimo, sezonski vplivi motijo dinamiko razvoja in ne dobimo pravilne slike o časovnih spremembah pojava. Zato popisujemo npr. živino vsako leto na isti datum (15. januar). Drugače je, če z vrsto popisov proučujemo sezonsko dinamiko stanja živine. Takrat je nujno, da popišemo živino v določenem letu ali zaporedju več let na krajša razdobja, npr. vsak mesec.

Tekoča registracija - Statistična poročila

3.8 Medtem ko je za popis bistvena istočasnost obstoja vseh enot populacije v kritičnem trenutku, je za tekočo registracijo bistveno, da registrira vsak dogodek v različnem času, v trenutku, ko se dogodi ali neposredno po dogodku.

Tudi za tekočo registracijo je značilno popolno zajetje vseh enot populacije.

Redkeje ko pri popisih je namen tekoče registracije zbiranje podatkov samo za statistično

proučevanje pojava. Matična služba, register prebivalstva, operativne evidence v podjetjih, knjigovodski podatki itd. so samo posredno vir statističnih podatkov. Te podatke zbiramo s t a t i s t i č n i m i p o r o č i l i , ki podajajo splošno sliko za dogodke v določenem razdobju (dnevu, tednu, mesecu, letu). Statistična poročila so ena izmed običajnih oblik registracije socialno-ekonomskih pojavov. Tako imamo npr. statistična poročila iz statistike cen, gostinstva, turizma, gradbeništva, industrije, matične službe itd.

D e l n a o p a z o v a n j a

3.9 Značilno za popise in tekoče registracije je, da z njimi opazujemo vse enote populacije. Z njimi dobimo popolno in podrobno sliko o proučevanem pojavu. Slaba stran popolnega opazovanja pa je, da zahteva veliko materialnih sredstev in časa.

Kadar za proučitev pojava ni potrebna podrobna slika opazovane populacije, ali če nimamo na razpolago zadosti sredstev, podatke o populaciji ocenjujemo. Najpopularnejše so metode ocenjevanja, pri katerih skušamo sklepati in ocenjevati podatke o populaciji, če poznamo razmeroma majhen del enot populacije.

Najobjektivnejša metoda pri ocenjevanju statističnih podatkov je v z o r č e n j e . Osnovna značilnost vzorčenja je, da so enote, ki jih opazujemo, da iz njih ocenimo celoto, izbrane slučajnostno. Metoda slučajnostnega izbora ima niz prednosti pred drugimi metodami ocenjevanja. Ocene, ki jih dobimo z vzorčenjem, so objektivne, na kakovost ocen moremo enostavno vplivati, zanesljivost ocen pa moremo objektivno določiti. Ker je ta metoda ocenjevanja v socialno-ekonomskih proučevanjih zelo pomembna, za njeno razumevanje pa je potrebno poznavanje osnov statistične analize, na tem mestu vzorčenje samo omenjamo, podrobneje pa ga bomo obravnavali v posebnem poglavju kasneje.

3.10 Druge metode delnega opazovanja, od katerih pa nobena nima vseh kvalitet, ki jih ima vzorčenje, so med drugimi: metoda izbora tipičnih enot in monografija.

Preden se je uveljavila v praksi metoda vzorčenja, ki jo uporabljamo za socialno-ekonomska proučevanja šele nekaj desetletij, so metodo izbora tipičnih enot imeli za eno izmed osnovnih metod delnega opazovanja. Metoda izbora tipičnih enot na prvi pogled nudi nekaj prednosti pred metodo vzorčenja. Skupina enot, ki so pazljivo izbrane, tako da čim bolje predstavlja celoto, da navidezno zanesljivejše ocene kot skupnost enot, ki so iz populacije izbrane slučajnostno. Vendar pri metodi izbora tipičnih enot nastopi težava, ki je slučajni izbor nima. Če hočemo izbrati iz populacije enote, ki predstavljajo celoto, moramo proučevano populacijo podrobno poznati. To pri vzorčenju ni potrebno. Tudi odločitev, kaj je tipično za populacijo, je subjektivna. Za ocene, ki jih dobimo po metodi izbora tipičnih enot, ne moremo oceniti zanesljivost podatkov niti ne moremo zavestno vplivati na kakovost ocen, čeprav verjetno z opazovanjem z majhnim številom enot dobimo z izborom tipičnih enot boljše ocene kakor z vzorčenjem.

3.11 M o n o g r a f i j o uporabljamo, kadar hočemo zelo podrobno osvetliti določen problem na eni ali nekaj tipičnih enotah populacije. Namen monografije ni posploševanje na celoto, temveč podrobna osvetlitev posameznih primerov. Znana so podrobna monografska proučevanja o produktivnosti dela, o delovnih razmerah delavcev itd. Monografska proučevanja so zelo uspešna tudi pri odkrivanju novih kvalitet pojavov.

3.12 P a n e l a n k e t e . Proučevanje dinamike pojavov z vrsto neodvisnih vzorcev je problematično. Ker so v vsak vzorec izbrane druge enote: osebe, gospodinjstva, gospodarstva ipd., je vzorčni pogrešek razmeroma velik. Četudi se ne bi stanje v populaciji spremenilo, bi verjetno v dveh zaporednih vzorcih dobili zelo različne rezultate, kar bi bila posledica tega, da so v posameznih vzorcih druge enote. To hibo odpravimo tako, da vzorec istih enot uporabimo za niz opazovanj in opazujemo spremembe na posameznih enotah tega vzorca v sukcesivnih opazovanjih. Taka anketa, ki jo imenujemo panel anketa, ima za osnovo longitudinalno opazovanje in je odlično sredstvo za proučevanje dinamike pojavov. Zato panel ankete uporabljamo v primerih, kadar gre za proučevanje časovnih sprememb, ki so rezultat najrazličnejših vplivov. S panel anketami proučujemo spremembe v obnašanju gospodinjstev pod vplivom najrazličnejših

dejavnikov, ki vplivajo na spremembe v standardu prebivalstva. S panel anketami moremo meriti reakcijo na oglaševanje, na cene, na konkurenco, na spreminjanje javnega mnenja ipd. Medtem ko je v primerih, ko gre za dolgoročnejše raziskave, en vzorec dalj časa osnova za panel anketo, more za določene raziskave, npr. vpliva na oglaševanje rabiti panel anketa s samo dvema opazovanjema: pred in po reklamiranju. S panel anketami oziroma longitudinalnim opazovanjem izločimo predvsem razlike med enotami opazovanja.

Seveda pa ima panel anketa tudi niz pomanjkljivosti. Problemi nastopijo zaradi izpadov posameznih enot iz panela zaradi različnih vzrokov: smrti, preselitve, nezainteresiranosti ipd. Slaba reprezentativnost panel anket gre v dosti primerih tudi na račun tega, ker je težko sestaviti reprezentativen panel vzorec, ker predstavljajo oni, ki so voljni sodelovati v panel anketi, dostikrat nereprezentativen del celote.

Daljše sodelovanje v panel anketi izzove tudi drug način obnašanja. Gospodinjstvo gospodari drugače, če beleži svoje dohodke in izdatke kot gospodinjstvo, ki tega ne dela. Izhod v takih primerih je postopna zamenjava enot panel vzorca z drugimi enotami. Če od časa do časa zamenjamo po en del enot (npr. četrtino), se pri štirikratni zamenjavi izmenja celoten vzorec. Kljub temu pa je ohranjena zveznost panel ankete, ker je v dveh zaporednih razdobjih vedno vključenih tri četrtine enot, ki časovno vežejo podatke.

VIRI PODATKOV

3.13 Pri organiziranju statističnega opazovanja je važno predhodno proučiti statistične vire, ki se nanašajo na pojav, ki ga nameravamo proučevati. Predhodna analiza virov mnogokrat poenostavi opazovanje, včasih pa celo uspemo, da pridemo do podatkov o proučevanem pojavu brez statističnega opazovanja. Zavedati se moramo namreč, da je statistično opazovanje najdražji in razmeroma zamotan način, da pridemo do statističnih podatkov.

Zato predhodno pregledamo vse publicirane in nepublicirane podatke o problemu, ki ga proučujemo. Včasih dobimo odgovor na postavljeni problem že nepo-

sredno iz podatkov, ki so jih zbrali in obdelali drugi. Tako uspemo, da analiziramo pojav s podatki iz enega ali več različnih virov brez posebnega statističnega opazovanja in osnovne obdelave. Večina podatkov, ki jih objavlja uradna statistika, so osnovni podatki. Njih namen je, da jih dalje prouči porabnik za svoje potrebe in po svojih vidikih.

3.14 Podatki že izvedenih statističnih opazovanj niso vedno obdelani tako, da bi mogli rezultate neposredno uporabiti za analizo problema, ki ga proučujemo. Preden se odločimo, da izvedemo samostojno opazovanje, proučimo še osnovno gradivo že izvedenih statističnih akcij. Dostikrat se izkaže, da sicer obdelani rezultati ne dajo odgovora na proučevan problem, moremo pa do njega priti s posebno obdelavo že zbranega gradiva po ustreznih vidikih in kombinacijah. Če nam to uspe, ni treba zbirati osnovnih podatkov. Tako izkoriščamo osnovno gradivo najrazličnejših popisov, kot so popisi prebivalstva, industrije, obrti itd. in prihranimo stroške zbiranja podatkov.

3.15 Za osnovo pri proučevanju uporabljamo tudi dosedanje zapise - evidence, ki jih zbira ena ali druga organizacija ali enota v nestatistične namene. Ti zapisi - imenujemo jih tudi sekundaren vir statističnih podatkov - so osnova za podatke iz vseh gospodarskih statistik, statistike industrije, trgovine, gostinstva itd. Prav tako je sekundaren vir statističnih podatkov matična služba, ki zbira podatke iz naravnega gibanja prebivalstva in register prebivalstva, ki beleži mehansko gibanje prebivalstva.

Obstoječi zapisi olajšajo statistično opazovanje, ker so podatki, ki jih iščemo, že zbrani. Zato je zbiranje osnovnih podatkov omejeno le na prepisovanje. Ker so ti zapiski običajno uradni dokumenti, so ti podatki dosti zanesljivi.

3.16 Dostikrat pa ne razpolagamo za proučevan pojav niti s takimi zapisi. Takrat smo primorani, da podatke o enotah populacije zberemo drugače. Eden izmed takih načinov je posredno opazovanje po osebah, ki enoto opazovanja poznajo in mo-

rejo dati o njej zahtevane podatke. Posredno opazovanje je zelo razširjeno. Tako dajejo gospodarji kmetijskih gospodarstev podatke o svojih gospodarstvih, gospodinje o porabi v svojih gospodinjstvih, starešine gospodinjstev podatke o članih, obrtniki o svojih podjetjih itd.

V nekaterih primerih, posebno če gre za poskuse je vir osnovnih statističnih podatkov neposredno opazovanje. Neposredno opazujemo pojave, če sami izvedemo poskus oziroma merimo enote opazovanja.

3.17 Posamezni viri so v zgornjem pregledu navedeni po vrstnem redu prednosti. Če ni publiciranih podatkov, ki bi dali odgovor na naš problem, iščemo ali je zbrano osnovno gradivo o tem problemu. V teh dveh primerih populacije statistično v ožjem smislu sploh ne opazujemo. Če tudi osnovnega gradiva o proučevanem problemu ni, iščemo sekundarne vire - obstoječe zapise. Čeprav v teh primerih izvršimo opazovanje, ker so ti podatki verjetno teritorialno in organizacijsko raztreseni, se izognemo izpraševanju ali merjenju. Če ni niti zapisov o proučevanem pojavu, smo prisiljeni, da opazujemo populacijo v pravem smislu besede in sicer tako, da iz posrednih virov dobimo podatke o enotah populacije. Če pa odpovedo še posredni viri, moramo meriti, to se pravi: neposredno opazovati.

Zgled za posredno in neposredno opazovanje je določanje pridelka po kmetijskih gospodarstvih. Pri posrednem opazovanju se zanesemo na izjave gospodarja kmetijskega gospodarstva, pri neposrednem opazovanju pa pridelek v posameznih gospodarstvih izmerimo. Iz tega zgleda vidimo tudi prednosti in pomanjkljivosti posrednega in neposrednega opazovanja. Gospodarjevo napoved dobimo zlahka, podatek pa more biti nezanesljiv; merjenje pridelka pa da zanesljivejše podatke, je pa bolj zamotano.

NAČINI POSREDNEGA OPAZOVANJA

3.18 Med osebami, ki podatke dajejo in organi statističnega opazovanja je več vrstnikov. Običajno popisovalec ali anketar o b i š č e osebo, ki more dati o proučevani enoti podatke. Tako pri popisih prebivalstva popisovalci obiščejo posamezna

gospodinjstva in osebe v svojem popisnem okolišu. Enako pri vzorčenju obiščejo anketarji izbrane enote, da jih popišejo.

Pri popisovalčevem obisku ločimo dve tehniki popisovanja. Pri samoregistraciji popisovalec ob prvem obisku razdeli obrazce, razloži tehniko in namen popisa in prosi popisovano osebo, da do določenega roka izpolni obrazce; pri ponovnem obisku pa izpolnjene obrazce pobere. Samoregistracija ima prednosti in hibe. Prednost je v tem, da oseba, ki podatke daje, lahko v miru in premišljeno izpolni obrazec, kar je v prid kakovosti, posebno če popisana oseba vpisuje podatke iz dokumentov. Prednost samoregistracije je tudi ta, da popisovalec ali anketar pri prvem obisku obrazce in navodila lahko pusti pri sosedih ali sorodnikih popisovane osebe, če je ni doma in naroči, da izpolni obrazce do predpisanega roka. V tem primeru sploh ni potreben neposreden stik med popisovalcem in osebo, ki podatke daje. Pomanjkljivost samoregistracije pa je v tem, da je uporabna le tedaj, če moremo predpostavljati, da so osebe zmožne, da same pravilno in vestno izpolnijo obrazce.

3.19 Pri ekspedicijskem načinu ali metodi intervjuva vnaša odgovore, ki jih dobiva od popisanih oseb,v obrazec popisovalec. Ta postopek je v nekaterih točkah boljši, v drugih pa slabši kakor samoregistracija. Ekspedicijski način zahteva mnogo več časa za popisovanje kakor samoregistracija. Čeprav je pri ekspedicijskem načinu teoretično potreben en sam obisk, je treba osebe, ki jih obiskovalec pri prvem obisku ne najde doma, ponovno obiskati. Posebne težave so pri ekspedicijskem načinu z osebami, ki so čez dan odsotne. Za te primere je zelo problematično, kdaj je najbolj primeren čas za obisk, da dobi popisovalec popisano osebo doma. Ekspedicijski način je vsekakor primernejši za populacijo z nizko kulturno ravnijo in pri zamotanejših opazovanjih. Tudi odgovori so običajno pri ekspedicijskem načinu popolnejši, boljši, pravilnejši in enotnejši, ker izpolnjuje obrazce popisovalec, ki je napravil poseben tečaj za popisovanje, ima obširnejša navodila, prakso v popisovanju itd. Popisovalec ali anketar pri ekspedicijskem načinu tudi hitreje presodi, ali so odgovori pravilni ali ne in more z dodatnimi kontrolnimi vprašanji priti do pravilnih podatkov.

3.20 O prijavnem načinu govorimo, če se morajo popisane enote prijaviti in dati podatke v zato določenih prijavnih pisarnah. Ta postopek je za popisovalca enostavnejši kot obiski, ker odpade obhod terena. Po prijavnem načinu so v Sloveniji izvedli popis kmetijskih gospodarstev v letu 1947. Vsak lastnik zemljiške posesti je bil obvezan, da na prijavnem uradu dá podatke o svoji zemljiški posesti. Po tem načinu, kolikor je lagoden, običajno zajamemo mnogo manjši odstotek populacije kot pri drugih metodah, razen če je uradno predpisana prijava in določene sankcije, če se oseba ne prijavi. Pri imenovanem popisu zemljiških gospodarstev je bila prednost prijavnega načina tudi ta, da so v prijavni pisarni imeli podatke katastra, s katerimi so prijavne komisije sproti kontrolirale navedbe zemljiških posestnikov. Po prijavnem načinu so organizirane različne službe npr. matična služba, registracija prebivalstva itd.

3.21 Pri poštnem načinu statistično mrežo popisovalcev na terenu zamenja pošta. Poštni način je znatno cenejši kot opazovanje z obiski ali prijavni način, ker odpadejo vsi vmesni organi med osebo ali organizacijo, ki podatke daje in centralnim organom, ki zbira podatke. Strošek stika z osebo ali organizacijo, ki podatke daje, se skrči na poštne stroške. Pri poštnem načinu so stroški znatno manjši, organizacija pa enostavnejša kot pri drugih metodah. Enako je tudi čas izvedbe opazovanja pri disciplininih enotah pri poštni metodi zelo kratek. Hiba poštnega načina pa je v tem, da moremo pri poštnem načinu uporabiti le samoregistracijo in so osebe, ki dajejo podatke, navezane na pisana navodila brez osebnega stika s statističnimi organi. Zaradi tega je pri poštnem načinu kakovost podatkov slaba, delež neodgovorov pa v splošnem še večji kot pri samoregistraciji. Poštni način uporabljamo le, če vemo, da so osebe, ki podatke dajejo, disciplinirane in zmožne, da same brez pomoči popisovalcev izpolnijo in pošljejo obrazce organizaciji, ki podatke zbira. Zgledi za uspelo opazovanje po poštnem načinu so tekoče statistične službe v industriji, trgovini, gradbeništvu itd. Podjetja so po službeni dolžnosti zavezana, da v roku pošljejo podatke o svoji dejavnosti, zagotovljena pa je tudi strokovnost izpolnjevanja.

Če opazujemo s poštnim načinom privatne osebe, obrtne obrate, kmetijska gospodarstva itd. običajno kombiniramo poštni način z metodo obiskov tako, da skušamo dobiti po pošti odgovore za čim večji del populacije, iz enot, za katere nismo prejeli odgovorov

pa izberemo vzorec. Izbrane enote obiščejo anketarji, da dobimo od njih podatke z osebnim stikom.

3.22 Podobne narave kot poštni način so tudi ankete, ki jih razpisujejo različni časopisi ali organizacije, tako da s pozivi v časopisju prosijo bralce ali člane, da pošljejo odgovore na zastavljena vprašanja pismeno na uredništvo lista ali na sedež organizacije. Pri tem načinu pričakujemo, da je število neodgovorov še večje. Število odgovorov skušamo zvečati z žrebanjem različnih nagrad za tiste, ki pošljejo odgovore, ali jih nagradimo kako drugače.

3.23 O korespodentnem načinu govorimo, če imamo za proučevanje določenega pojava na terenu stalno mrežo korespodentov, ki občasno pošiljajo podatke o dejavnosti, za katero so določeni. Korespodenti po stroki niso statistiki in imajo drugo glavno zaposlitev. Običajno pa so korespodenti zaposleni v stroki, iz katere je pojav, ki ga proučujemo. Tako imamo v kmetijski statistiki stalno mrežo korespodentov, ki so izbrani tako, da so razmeščeni po celem ozemlju, ki ga proučujemo. Kmetijski korespodenti so običajno agronomi ali razgledani privatni kmetovalci, ki so sposobni, da ocenjujejo pojave iz kmetijstva in občasno pošiljajo podatke v centralni urad, kjer podatke obdelujejo. Korespodenti pošiljajo podatke o stanju posevkov, o vegetaciji, o tem, kako kaže pridelek itd.

ZNAKI OPAZOVANJA

3.24 K opredelitvi opazovanja ne sodi samo opredelitev populacije, temveč tudi opredelitev vsebine opazovanja. Vsebino dajo opazovanju znaki opazovanja. Od dobre in smiselne izbire znakov je odvisen uspeh opazovanja in proučevanja nasploh.

Pri izbiri znakov statističnega proučevanja moramo upoštevati nekaj splošnih načel, ki pomagajo pri izbiri znakov. V opazovanje je treba vključiti vse znake, ki so v neposredni zvezi s proučevanim pojavom, vendar samo tiste, ki so nujno potrebni. Vključevanje znakov, ki nimajo zveze s proučevanim pojavom, ali odvečna vprašanja, organizacijsko in vsebinsko bremene opazovanje. Organizacijsko ga bremene v tem smislu, da je opazovanje za več znakov dražje in bolj zapleteno. Vsebinsko pa je kakovost odgovorov na

splošno slabša, če iščemo podatke za več znakov. Zato se omejimo samo na bistvena vprašanja in le na tiste znake, ki jih nameravamo obdelati.

Ne iščemo podatkov za vprašanja, za katera že v naprej vemo, da iz katerih koli objektivnih ali subjektivnih razlogov ne bomo dobili zanesljivih odgovorov. Zato se npr. izogibamo kočljivih vprašanj iz osebnega življenja oseb, katere popisujemo. Enako ne dobimo zanesljivih odgovorov na vprašanja iz preteklosti, če se opirajo ti odgovori na spomin. Tako npr. ne moremo zahtevati, da se anketirana oseba spomni, koliko alkohola je popila v preteklem letu. To vprašanje je dvakrat problematično. Anketirana oseba verjetno teži k temu, da pove manjšo količino od resnične. Razen tega pa se je objektivno težko spomniti in dati zadovoljiv odgovor na to vprašanje za celo leto nazaj. Prav tako se izogibamo znakov, ki zahtevajo zapletena preračunavanja, ki jih ne zmorejo osebe, ki te podatke dajejo. Enako ne zahtevamo pretirano natančnost odgovorov, ker je nestvarna.

3.25 Razen vsebinskih znakov včasih postavljamo še k o n t r o l n e z n a k e . Kontrolni znaki so v zvezi z vsebinskimi znaki tako, da moremo iz primerjave podatkov s kontrolnimi znaki sklepati na pravilnost odgovorov na osnovna vsebinska vprašanja.

I d e n t i f i k a c i j s k i z n a k i niso v zvezi z vsebino pojava, ampak je njihova funkcija v tem, da z njimi po potrebi ugotovimo za katero enoto veljajo dani podatki. Identifikacijski znaki so npr. ime in priimek, naslov poročevalske enote itd. Ti podatki so potrebni, če iščemo za enoto opazovanja dodatne podatke, popravke itd.

SREDSTVA OPAZOVANJA

3.26 Rezultate vnašamo v statistične o b r a z c e , p o p i s n i c e ali v p r a š a l n e p o l e . Na statističnih obrazcih zbiramo podatke o posameznih enotah opazovanja. Po končanem opazovanju izpolnjeni statistični obrazci zamenjajo opazovane enote, ker vsebuje posamezen obrazec vse informacije, ki jih iščemo o enoti.

Statistični obrazec je papir, na katerem so nanizani vsi znaki statističnega opazovanja. Prirejen je tako, da more popisana oseba, popisovalec ali anketar vanj vpisovati značil-

nosti za popisano enoto.

Obrazec mora biti tako sestavljen, da z njim čim laže dobimo popolne in pravilne odgovore na vprašanja, ki so v obrazcu zastavljena. Razen tega mora biti obrazec prilagojen načrtovani obdelavi zbranih podatkov. Obrazec mora biti torej prilagojen načinu opazovanja in obdelave.

3.27 Vrste obrazcev. Glede na to, ali je obrazec namenjen za popis ene same ali več enot opazovanja, imamo individualne in kolektivne obrazce. Z individualnim obrazcem popisujemo eno samo enoto, s kolektivnim pa več enot. Enote na kolektivnem obrazcu morejo biti med seboj vsebinsko povezane, ni pa to nujno. Pri popisu prebivalstva kolektiven obrazec za gospodinjstvo združuje podatke vseh članov enega gospodinjstva. V tem primeru je med enotami, ki so popisane na kolektivnem obrazcu, vsebinska zveza. Večkrat pa je kolektiven obrazec namenjen za popis vseh enot enega popisnega okoliša in med enotami v kolektivnem obrazcu razen regionalne ni globlje vsebinske zveze.

Izbira med individualnim ali kolektivnim obrazcem je odvisna od več vzrokov. Kolektiven obrazec moremo uporabiti le pri ekspedicijskem načinu, za samoregistracijo pa ni uporaben. Medtem ko so individualni obrazci primerni za vse vrste obdelave, moremo s kolektivnimi obrazci izvesti obdelavo samo po nekaterih metodah. Če so enote v kolektivnem obrazcu združene po kakšnem vsebinskem načelu, je kolektiven obrazec primeren za ročno obdelavo, ker moremo na njem neposredno izvesti osnovne operacije obdelave, to je seštevanje podatkov v podatke za skupinice.

V splošnem je opazovanje s kolektivnimi obrazci cenejše kakor opazovanje z individualnimi obrazci. To pa zato, ker en kolektiven obrazec rabi za popis več enot opazovanja, medtem ko potrebujemo po en individualni obrazec za popis vsake enote. Vendar je uporaba kolektivnih obrazcev zaradi navedenih vzrokov omejena.

3.28 Sistem obrazcev. Pri statističnem opazovanju navadno ne uporabljamo samo enega obrazca, temveč cel sistem obrazcev. Vsi obrazci pa imajo en sam namen: zbrati čim popolnejše in pravilnejše odgovore na zastavljena vprašanja.

Po vlogi, ki jo imajo, delimo obrazce v glavne, pomožne in kontrolne.

G l a v n i o b r a z e c je običajno v enem opazovanju en sam. Ta obrazec vsebuje listo znakov oziroma vprašanj, na katera mora dati popisana enota odgovor. Glavni obrazec je namenjen za nadaljnjo obdelavo.

P o m o ž n i o b r a z c i pomagajo k čim boljšim in popolnejšim odgovorom na vprašanja v glavnem obrazcu. Pri popisu prebivalstva je npr. poseben problem dobiti zanesljive podatke na vprašanja, kot so: poklic, položaj v poklicu, panoga dejavnosti itd. Zato pri popisih prebivalstva praviloma uvedemo pomožen obrazec; tega za vsako zaposleno osebo pred popisom izpolni podjetje oziroma ustanova, kjer je popisana oseba zaposlena. Popisovalec ali popisana oseba vnese odgovore, ki so zanesljivejši kakor odgovori, ki jih da posamezna popisana oseba sama, iz pomožnega obrazca v glavni obrazec.

Naloga k o n t r o l n i h o b r a z c e v je predvsem kontrola polnoštevilnosti izpolnjenih obrazcev. Včasih pa rabi kontrolni obrazec tudi kot kolektiven obrazec za osnovne podatke o popisovani populaciji.

3.29 E l e m e n t i o b r a z c a . Statistični obrazci so izdelani po nekih splošnih načelih, vendar so različni glede na populacijo, način in organizacijo opazovanja in obdelave. Kakor vidimo iz zgledov v slikah 3.1 do 3.3 imajo vsi trije neke skupne črte.

Statistični obrazci imajo razen vsebinskih vprašanj in prostora za odgovore tudi druge elemente, ki so organizacijskega značaja in je njih namen olajšati izvedbo opazovanja in obdelave.

Vsak obrazec ima svoj n a s l o v : ta v kratkem poda vsebino opazovanja in vlogo obrazca, če ima opazovanje več obrazcev.

Obrazec označimo razen z naslovom še s kratko oznako - k r a t i c o . Tako ima prikazani obrazec popisa prebivalstva oznako "Obrazec PS-1", obrazec kmetijske ankete "Obrazec PA-4", obrazec za mesečno industrijsko poročilo pa "IND-1-1958".

V glavi obrazca je tudi n a z i v o r g a n a , ki je obrazec izdal (npr. Zvezni za-

vod za statistiko), k d o j e o b r a z e c o d o b r i l, če ga izdaja organizacija, ki ni statistična, in n a z i v i n n a s l o v e n o t e, ki daje podatke.

V obrazcu je naveden tudi č a s o p a z o v a n j a (popis prebivalstva 31.3.1953, kmetijska anketa 15.1.1958, mesečno industrijsko poročilo: za mesec ..... 1958), r o k i z p o l n i t v e (mesečno industrijsko poročilo: rok 6. v mesecu), š t e v i l o i z v o d o v (mesečno industrijsko poročilo: v 4 izvodih) in k o m u p o š l j e poročevalska enota obrazce pri samoregistraciji (mesečno industrijsko poročilo: odda Zavodu za statistiko občine, v kateri je sedež uprave podjetja).

Obrazec ima običajno tudi mesto za p o d p i s o d g o v o r n e o s e b e ali osebe, ki je podatke dala.

3.30 P o s t a v l j a n j e v p r a š a n j . Najbolj kočljiva pri sestavljanju statističnih obrazcev so vsebinska vprašanja. To ni samo tehnično vprašanje, marveč je, posebno pri samoregistraciji, treba upoštevati tudi psihološke momente, ki nastopijo pri enem ali drugem načinu spraševanja. Zato je potrebno dobro preštudirati možnosti in zmožnosti oseb, ki bodo izpolnjevale obrazce, in pripraviti vprašanja tako, da kar najbolj verjetno dobimo pravilne odgovore. Odgovori morejo biti napačni iz dveh vzrokov. Popisana oseba bodisi ne zna ali noče na vprašanja pravilno odgovoriti. Težave prve vrste laže odpravimo kot težave druge vrste. Jasna in enostavna vprašanja in izčrpna navodila mnogo pripomorejo k razumevanju in odpravljanju napak, ki nastanejo zaradi nerazumevanja. Če sodimo, da je celoten obrazec pretežak za samoregistracijo, popišemo enote ekspedicijsko. Hujše so napake, ki se pojavijo, ker popisane osebe nočejo pravilno odgovoriti. Vprašanja, za katera pričakujemo, da popisane osebe ne bodo hotele odgovoriti, skušamo postaviti neposredno tako, da se izognemo vzroku, zaradi katerega popisana oseba noče pravilno odgovoriti. Če take rešitve ne najdemo, taka vprašanja raje opustimo.

3.31 Najmanj oseben je obrazec, v katerem n a n i z a m o z n a k e, popisane enote pa vnašajo odgovore v ustrezen prazen prostor poleg označitve znaka. Ta tehnika je npr. uporabljena običajno pri popisih prebivalstva. Ta metoda je zelo uporabna, ker je enostavna in pregledna. Posebno je priporočljiva pri ekspedicijskem načinu opazovanja.

Federativna ljudska republika Jugoslavija

ZVEZNI ZAVOD ZA STATISTIKO IN EVIDENCO

Obrazec PS - 1

Popis prebivalstva 31. marca 1953

Popisnica

Izpolni popisovalec		Izpolni zavod za statistiko in evidenco
Številka popisnega okoliša		
Zaporedna številka kontrolnika		
Okraj - mesto		
Občina		
Naselje (mesto, vas)		
Sestavni del naselja		
Ulica in hišna številka		
Ali je popisana oseba	trajno prisotna (*da, ne*)	10
	začasno odsotna (*da, ne*)	
	začasno prisotna (*da, ne*)	

Pred izpolnitvijo prečitajte vso popisnico!
Preden napišete odgovor na posamezna vprašanja, ponovno prečitajte navodila, ki so tiskana pri teh vprašanjih!
Če vam ni kaj jasno, vprašajte popisovalca za pojasnilo!

Popisnica se izpolni za vsako osebo, ki živi opolnoči med 31. marcem in 1. aprilom 1953.

Oseba, ki je ob času popisa v kraju, kjer stalno stanuje, se popiše v tem kraju (v svojem gospodinjstvu). V svojem gospodinjstvu se popiše (kot trajno prisotna) tudi vsaka oseba, ki je ob popisu pri vojakih, na vojaških vajah ali v zaporu.

Oseba, ki ob popisu ni v kraju, kjer stalno stanuje (ker je na potovanju, v bolnici in pod.), se popiše v kraju, kjer stalno stanuje (kot začasno odsotna), in tudi v kraju, kjer je ob popisu (kot začasno prisotna).

Ako je oseba začasno odsotna, navedemo naslov, kjer je ob popisu; če je začasno prisotna, je treba navesti naslov v kraju, kjer stalno stanuje	Ljudska republika	
	Okraj - mesto	
	Občina	
	Naselje (mesto, vas)	
	Sestavni del naselja	
	Ulica in hišna številka	

1. PRIIMEK, OČETOVO IME (ZA POROČENE ŽENE MOŽEVO IME) IN IME _____

2. SPOL (odgovor: *moški — ženski*) . . . . . _____

3. ROJSTNI DATUM dan _____ mesec _____ leto _____

4. ROJSTNI KRAJ a) naselje (mesto, vas) _____ b) okraj _____ 20

c) ljudska republika (osebe, rojene v inozemstvu, navedejo državo) . . . . . _____

5. DRUŽINSKI STAN [Pri osebah, rojenih po 31. marcu 1939, naredimo črtico (—); pri osebah, ki so se rodile pred tem datumom, vpišemo enega izmed odgovorov: *samski-samska, poročen-poročena, vdovec-vdova, razvezan-razvezana.*] . . . . . _____

6. KATERI ZAKON JE PO ŠTEVILU [Na to vprašanje odgovorijo samo osebe, ki so ob popisu poročene, in napišejo z besedami, n. pr.: *prvi, drugi, tretji* itd.] . . . . . _____

7. DOPOLNJENA LETA STAROSTI OB SKLENITVI PRVEGA ZAKONA [Osebe, ki so se poročile do dneva popisa, napišejo leta s številkami, n. pr.: *18, 23, 29* itd.; osebe, ki se do popisa niso poročile, naredijo črtico (—).] . . . . . _____

8. ŠTEVILO ŽIVOROJENIH OTROK [Na to vprašanje odgovori vsaka ženska, ki se je rodila pred 31. marcem 1939; pri tem naj upošteva vse svoje zakonske in nezakonske otroke.] . . . . . _____

9. KOLIKO OD TEH OTROK SEDAJ ŽIVI . . . . . _____ 30

10. DRŽAVLJANSTVO [Državljani FLRJ vpišejo: *FLRJ;* tuji državljani navedejo državljanstvo, ki ga imajo.] . . . . . _____

11. NARODNOST [Vsaka oseba vpiše svojo narodnost, n. pr.: *Srb, Hrvat, Slovenec, Makedonec, Črnogorec, Madžar, Šiptar, Nemec, Italijan, Čeh, Slovak, Turek, Cigan* itd. Osebe jugoslovanskega rodu, ki niso narodnostno opredeljene, vpišejo: *Jugoslovan-neopredeljen,* medtem ko ostale narodnostno neopredeljene osebe vpišejo: *narodnostno neopredeljen.* Za otroke pod 10 leti je odločilna izjava staršev.] _____

12. MATERIN JEZIK [Vpišemo jezik, katerega oseba največ uporablja v svojem gospodinjstvu, oz. jezik, ki ga ima oseba za svoj materin jezik. Za otroke pod 10 leti je odločilna izjava staršev.] . . . . . _____

13. ODNOS DO VERE [Osebe, ki imajo določeno versko prepričanje, napišejo veroizpoved, kateri pripadajo; osebe brez verskega prepričanja napišejo: *brez vere.* Za otroke pod 14 leti je odločilna izjava staršev.] . . . . . _____

Slika 3.1. Prva stran popisnice pri popisu prebivalstva 31.3.1953 v SFR Jugoslaviji.

Bolj osebno je, da damo namesto nazivov znakov v p r a š a n j a . Namesto, da vpišemo npr. poklic in pričakujemo, da bo popisana oseba v ustrezni prostor vpisala svoj poklic, raje zastavimo vprašanje: Kakšen je vaš poklic? Na tako vprašanje moremo pričakovati večje razumevanje, kaj mora popisana oseba storiti, kot pa če postavimo suhoparno: Poklic. Vprašanja uporabljamo posebno takrat, kadar med osebami, ki podatke dajejo in statističnimi organi ni osebnega stika. Na razumljivo in vljudno postavljena vprašanja moremo pričakovati več in pravilnejše odgovore - kot pa na suhoparno listo znakov, za katere naj popisana oseba dá odgovore.

3.32 Izkazalo se je, da so zelo primerna vprašanja, na katera oseba, ki podatke daje, odgovori z d a ali n e . Ta vprašanja so najbolj neposredna in zahtevajo od osebe, ki podatke daje, najmanj napora. Vendar mora biti tako vprašanje zastavljeno tako, da se čimbolj izognemo pristranosti odgovorov.

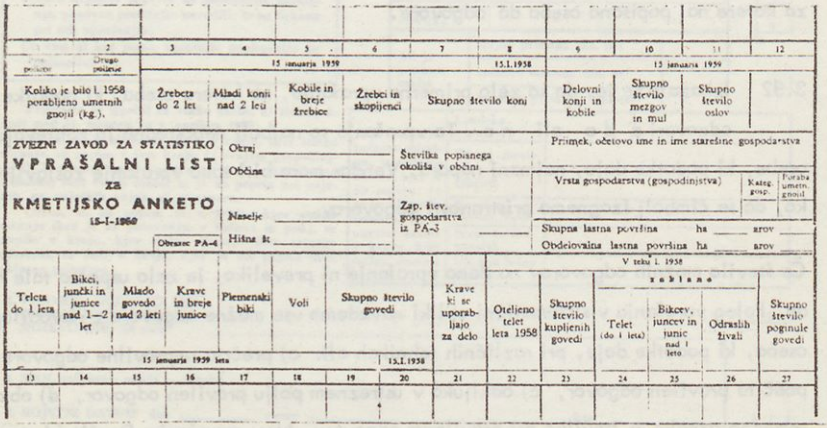
Če število možnih odgovorov na dano vprašanje ni preveliko, je zelo uspešna tale metoda. Poleg vprašanja v sistematični obliki navedemo vse možne odgovore z navodilom, da oseba, ki podatke daje, pri različnih tehnikah ali: a) prečrta nepravilne odgovore, b) podčrta pravilen odgovor, c) odkljuka v ustreznem polju pravilen odgovor, d) obkroži ustrezno oznako - številko pri pravilnem odgovoru. Npr. vprašanje: Družinski stan, moremo po posameznih sistemih postaviti vprašanja in dobiti odgovore takole:

a) samski, poročen, razvezan, vdovec
b) samski, poročen, razvezan, vdovec

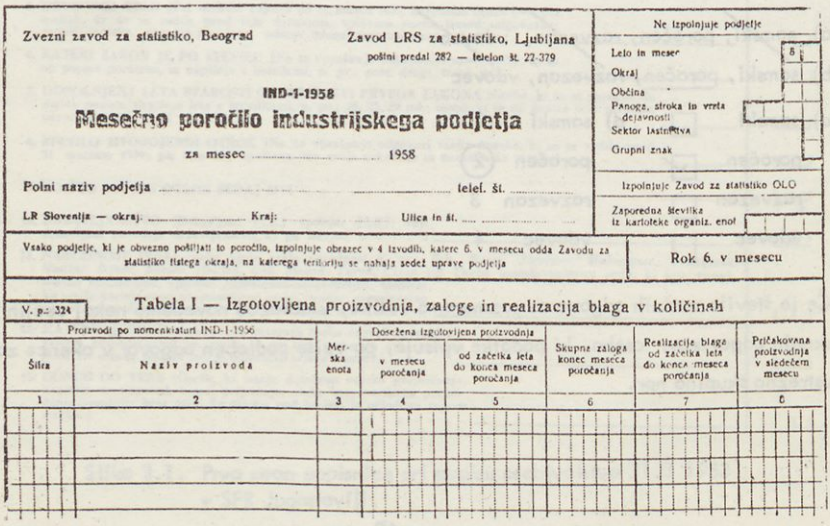
c)		d)	
samski	☐	samski	1
poročen	☑	poročen	(2)
razvezan	☐	razvezan	3
vdovec	☐	vdovec	4

Če je število možnih odgovorov razmeroma veliko, v obrazcu navedemo nekaj večjih skupin odgovorov; oseba, ki podatke vpisuje, pa vpiše podroben odgovor v okence za ustrezno skupino npr.

Šolska izobrazba	Število razredov
brez	
osnovna šola	
srednja šola	3 razredi
visoka šola	



Slika 3.2 Prva stran vprašalnega lista za kmetijsko anketo 15.1.1958.



Slika 3.3. Glava mesečnega poročila industrijskega podjetja v letu 1958.

Če hočemo, da bodo vprašanja ustrezna, moramo preštudirati različne primere. Katero izmed možnih variant uporabimo v popisu ali anketi, se odločimo šele po temeljitem premisleku in preskusu na terenu, saj se bo tam najbolj pokazalo, kateri način je najprikladnejši.

3.33 N a v o d i l a . Obrazec bi bilo običajno nemogoče pravilno izpolniti, če bi vseboval le listo znakov ali spisek vprašanj. Mnoga vprašanja niso takoj jasna in moramo zanje dati pojasnila v n a v o d i l i h z a i z p o l n j e v a n j e o b r a z c a .

Če opazujemo ekspedicijsko, dobe popisovalci navodila za izpolnjevanje obrazcev na tečajih za popisovanje in v posebnih tiskanih navodilih. Pri ekspedicijskem načinu navodila za izpolnjevanje obrazca na samem obrazcu niso potrebna in samo zamotajo osnovni obrazec.

Če pa je za opazovanje predvidena samoregistracija, moramo vsako osebo, ki izpolnjuje obrazec, poučiti, kako je treba pravilno izpolniti obrazec. Zanje v splošnem ne moremo prirediti posebnih tečajev za izpolnjevanje obrazca. Morejo pa v tem primeru dobro rabiti članki v časopisu in oddaje v radiu ali televiziji; tako opozarjamo osebe, ki izpolnjujejo obrazce, na najbolj kočljive točke v izpolnjevanju obrazcev. Pri samoregistraciji pa so najbolj uspešna tiskana navodila na samem obrazcu. Splošna navodila damo na popisnem obrazcu tik pod naslovom, pojasnila k posameznim vprašanjem pa pri ustreznem vprašanju. Nikakor niso priporočljiva navodila na zadnji strani obrazcev ali celo navodila, ki so ločena od obrazca. Dostikrat osvetli izpolnjevanje izmišljen primer, ki nazorno pokaže, kako je treba obrazec izpolniti.

3.34 Posebnost prikazanega obrazca za popis prebivalstva je v tem, da ima ob desni strani prostor, določen za vnašanje številčnih šifer - oznak za posamezne vrednosti znakov. Šifre so potrebne za strojno obdelavo podatkov. Pri obrazcu za kmetijsko anketo v sliki 3.2 pa vidimo, da je prostor za vnašanje podatkov o številu živine sistematično nanesen ob robu obrazca in da ima obrazec nenavadno-podolgovato obliko. Ta posebna oblika in prostor za podatke sta določena z načrtom ročne obdelave podatkov. Seštevanje podatkov je tako znatno poenostavljeno. Obrazce, za katere moramo podat-

ke seštevati, naložimo drug čez drugega. Tako zlahka seštevamo podatke brez prepisovanja v obdelovalne tabele.

Podobnih prijemov, ki jih uporabljamo pri sestavi obrazcev, skladno s predvideno obdelavo, je še več in jih je vredno upoštevati. Razmeroma neznatna sprememba popisnega obrazca poenostavi popisovanje ali obdelavo podatkov.

KRAJ OPAZOVANJA

3.35 Regionalna razmestitev populacije je važna iz vsebinskega in tehničnega vidika. Proučevanje aglomeracije prebivalstva, lokacije industrije, trgovinske mreže, transporta itd. je vezano na podatke o regionalni razmestitvi posameznih pojavov. Zato večina opazovanj skuša zbrati podatke tako, da je možno proučevati regionalno razmestitev pojava.

Razen tega pa je regionalna razmestitev zelo važna za organizacijo opazovanja. Za populacije, za katere so enote razmeščene po vsem teritoriju, je organizacija opazovanja bistveno drugačna kakor za populacije, za katere so enote nakopičene v enem ali nekaj središčih.

3.36 Pri statističnih opazovanjih z velikim številom enot, ki so teritorialno raztresene, običajno celotno področje opazovanja razdelimo v p o p i s n e o k o l i š e. Popisni okoliši so najmanjše teritorialne enote, sestavljene za potrebe opazovanja. Popisni okoliš je po pravilu področje, v katerem popisuje en sam popisovalec. Zato so popisni okoliši glede na razmere terena zelo različni. Popisni okoliš more biti naselje ali del naselja, mestna ulica, stanovanjski blok, samotna kmetija itd. Popisni okoliši morajo biti po pravilu taki, da enolično zajamejo vse področje opazovanja. To pomeni, da mora biti vsaka enota v enem samem popisnem okolišu. Nepravilna bi bila razdelitev teritorija v popisne okoliše, če bi neka hiša pri popisu prebivalstva spadala v dva popisna okoliša, neka samotna kmetija pa v nobenega.

Razdelitev teritorija popisa v popisne okoliše mora biti izvedena tako, da moremo z njimi sestaviti najmanjše politično-upravne teritorialne enote, t.j. občine. To načelo je

razen drugega zelo koristno, če potrebujemo podatke po upravno-teritorialnih enotah. Če ga upoštevamo, moremo iz popisnih okolišev brez težav sestaviti podatke za katero koli upravno-teritorialno enoto.

3.37 Kraj opazovane enote more biti v načelu različen od k r a j a p o p i s a . Pri posrednem opazovanju daje podatke oseba, ki ni nujno v istem kraju kot enota opazovanja. Tako daje centrala podatke za svoje podružnice, ki niso v istem kraju, kmetijski posestnik podatke o svojem imetju (živini, zemljiščih itd.), ki ni v istem kraju kakor sedež gospodarstva.

Razmestitev opazovane populacije ni nujno ista kakor razmestitev enot poročanja - to je oseb, organizacij, ustanov itd., ki dajejo podatke o opazovanih enotah. Regionalna razmestitev enot populacije je važna iz vsebinskih razlogov, ker da sliko o razmestitvi opazovanega pojava; razmestitev poročevalskih enot pa je pomembna za organizacijo opazovanja, ker od njih dobimo podatke o enotah populacije.

ORGANI STATISTIČNEGA OPAZOVANJA

3.38 Pri statističnem opazovanju večjega obsega sodeluje veliko ljudi z različnimi nalogami.

Program statističnega opazovanja sestavljajo s t a t i s t i č n i s t r o k o v n j a k i v tesnem sodelovanju s porabniki statističnih podatkov in s t r o k o v n j a k i i z p o d r o č j a , kamor spada konkretno statistično opazovanje. Pri sestavljanju programa sodeluje razmeroma malo toda visoko kvalificiranih strokovnjakov.

Enako je potrebno tesno sodelovanje statističnih strokovnjakov in strokovnjakov s področja, kamor spada opazovanje, tudi pri sestavljanju obrazcev, splošnem načrtu in organizaciji opazovanja. Vendar je pri tem situacija obrnjena. Program opazovanja je predvsem delo strokovnjakov s področja, v katerega spada opazovanje po vsebini; statistični strokovnjaki samo pomagajo pri izdelavi programa, da je formulacija programa taka, da zadošča statističnim načelom. Organizacija opazovanja pa je predvsem delo statističnih strokovnjakov in jim strokovnjaki s področja, v katerega spada pojav, samo pomagajo. Ta pomoč je

neobhodna pri sestavljanju obrazcev, vsebinskih navodil itd.

Važni sodelavci pri statističnem opazovanju so p o p i s o v a l c i oziroma a n k e t a r j i. Popisovalci neposredno izvajajo statistično opazovanje na terenu. Njihov poklic običajno ni statistika. Za to delo so izbrani pri vsakem popisu posebej. Čeprav to niso statistiki, si pri vsakem popisu prizadevamo, da izberemo za popisovalce osebe, ki imajo tako zaposlitev, da jim predmet ni tuj. Tako npr. pri anketah iz kmetijstva izberemo za popisovalce agronome, pri anketah iz gozdarstva gozdarje ali logarje, pri popisih iz šolstva in prosvete učitelje ali profesorje itd. To znatno pripomore h kakovostni izvedbi opazovanja. Popisovalce usposobimo za njihovo delo na posebnih tečajih ali s podrobnimi navodili za izvajanje statističnega opazovanja, pri katerem sodelujejo.

Razen kontrolorjev, ki so bolje poučeni in izurjeni za opazovanje kakor popisovalci in so običajno, če le mogoče, statistični organi, sestavimo pri velikih popisih p o p i s n e k o m i s i j e. Popisne komisije in kontrolorji imajo nalogo pomagati drugim organom pri opazovanju. Obenem popisne komisije zbirajo gradivo od popisovalcev in skrbe za pravilno izvajanje opazovanja na svojem področju.

Razen popisovalcev so organi opazovanja še o s e b e, k i d a j e j o p o d a t k e o opazovanih enotah. To so v vsakem primeru osebe, ki predmet opazovanja najbolje poznajo. Tako dajejo pri popisih prebivalstva starši razen zase še podatke za svoje otroke, gospodarji kmetijskih gospodarstev za svoje gospodarstvo, statističarji v industrijskih podjetjih za svoje podjetje itd.

Kakovost izvedenega opazovanja je odvisna predvsem od oseb, ki podatke dajejo. Kakovost njihovih izjav pa je odvisna od organizacije opazovanja, popisnih obrazcev, navodil za izpolnjevanje, dela popisovalcev itd. Le razumevanje in skupen napor vseh organov opazovanja, od vodstva do tistih, ki podatke dajejo, more zagotoviti uspešno izvedbo statističnega opazovanja.

V posameznih opazovanjih more ta ali oni organ izpasti. Tako pri poštnem načinu odpadejo popisovalci in popisne komisije, pri opazovanjih manjšega obsega popisne komisije itd. Vrsta in število organov pri posameznem statističnem opazovanju je odvisno od proučevane populacije in organizacije opazovanja.

3.39 V z r o k i n a p a k . Pri statističnem opazovanju sodeluje veliko ljudi, ki imajo najrazličnejše težnje, so na različni kulturni ravni, so bolje ali slabše poučeni o vlogi opazovanja in tehniki izpolnjevanja obrazcev itd. Zato organizatorji opazovanja vnaprej vedo, da zbrani podatki ne bodo niti popolni niti popolnoma pravilni. Vendar skušamo opazovanje izvesti in organizirati tako, da je čim manj napak v opazovanju, gradivo pa čim popolnejše. Vnaprejšnja analiza napak, ki morejo nastati v statističnem opazovanju, more dati smernice, kako posamezne vrste napak odpravimo ali pa vsaj kar najbolj omejimo.

V z r o k o v za napake v statističnem opazovanju je več. Pomanjkljiva navodila, nerazumljive opredelitve pojmov, slabo zastavljena vprašanja v popisnem obrazcu, vprašanja, ki preveč posegajo v osebno življenje, slaba organizacija itd. so vzroki napak, ki jih zagreši vodstvo opazovanja. Vzrok napak more biti tudi napačno tolmačenje in slabo delo popisovalcev. Vendar moremo kljub temu, da so v opazovanju vprašanja dobro zastavljena in delo popisovalcev dobro, dobiti slabe podatke, če ljudje namerno navajajo napačne podatke. To je pogost pojav pri podatkih o premoženjskih razmerah, dohodkih, potrošnji itd. Popisana oseba se boji, da bi zbrane podatke izkoriščali v nestatistične namene, čeprav je ta bojazen neupravičena. Z zakonom je namreč zagotovljena tajnost posamičnih statističnih podatkov. Namen zbiranja statističnih podatkov je izključno statistična obdelava in slika celote. Ti podatki se ne smejo izkoriščati v nobene druge namene. Zaupanje ljudi, da se zbrani podatki ne uporabijo za posamične sklepe, včasih utrdimo tako, da izvedemo anketo anonimno, čeprav je v tem primeru otežkočena kontrola zbranih podatkov.

3.40 V r s t e n a p a k . Napake v statističnem opazovanju morejo biti slučajne ali sistematične. S l u č a j n e n a p a k e izvirajo iz malomarnega opazovanja in izpolnjevanja obrazcev. Učinek slučajnih napak na rezultate opazovanja ni prevelik, ker se po zakonu o velikih številih slučajne napake izravnavajo.

Hujši je učinek sistematičnih napak. S i s t e m a t i č n e n a p a k e so napake, ki se pojavljajo iz nekega določenega vzroka s sistematično enakim učinkom pri posameznih enotah

opazovanja. Učinek sistematične napake se v velikem številu enot ne izravna, temveč kopiči. Zato so sistematične napake nevarnejše kakor slučajne. Vir sistematičnih napak more biti vsak izmed vzrokov, ki smo jih našteli zgoraj. Tako more biti vzrok sistematične napake nejasno opredeljen pojem ali slabo zastavljeno vprašanje. Če npr. iščemo po podjetjih vrednost proizvodnje v določenem razdobju, a ne določimo natančno, kaj pomeni vrednost industrijske proizvodnje, more biti to vir sistematične napake, ker podjetja v dobri veri vključujejo ali izključujejo neke od svojih dejavnosti v vrednost proizvodnje.

Vir sistematičnih napak more biti tudi zaokroževanje podatkov. Tipičen primer je navajanje starosti pri popisih prebivalstva. Stari ljudje radi zaokrožujejo svojo starost na okrogla leta in npr. navajajo, da so stari 60, 65, 70, 75, 80 let, čeprav niso stari 60 let, temveč okrog 60 let, ne 65, marveč okrog 65 let itd. Zato imamo v rezultatih števila prebivalcev po starosti značilna kopičenja prebivalstva za starosti 60, 65, 70 itd. let. Ta sestava ni stvarna, marveč je rezultat sistematične napake zaradi zaokroževanja let. To napako moremo pri opazovanju odpraviti, če od popisovalca zahtevamo, da starost preveri z rojstnim listom ali drugim dokumentom. [vprašaj za datum rojstva

Vir sistematičnih napak more biti tudi namerno dajanje napačnih podatkov. Kakor smo že navedli, opažamo sistematično navajanje premajhnih podatkov o premoženju, dohodkih itd.

3.41 K o n t r o l a n a p a k . Da bi dobili čim boljše podatke o opazovanem pojavu, je potrebna dobra in učinkovita kontrola zbranega gradiva.

Kontrola med opazovanjem obstoji iz terenske kontrole o delu popisovalcev in izpolnjevanja obrazcev. Namen te kontrole , ki ne more biti celotna, je predvsem odkrivanje tipičnih napak v organizaciji in napak pri izpolnjevanju obrazcev. Z dodatnimi splošnimi navodili in priporočili odstranimo težave in napake v opazovanju, ki jih pri načrtu opazovanja nismo poznali in jih je odkrila šele kontrola v toku opazovanja.

Kontrola zbranega gradiva je popoln pregled gradiva opazovanja. Kontrola zbranega gradiva obsega tele stopnje: kontrolo polnoštevilnosti zajetja enot, kontrolo polnoštevilnosti odgovorov in kontrolo pravilnosti odgovorov.

Kontrola polnoštevilnosti zajetja je omejena na pregled gradiva in številčno primerjavo zbranih obrazcev s kontrolniki, spiski oziroma registri enot. Ta kontrola odkrije enote, ki niso bile popisane oziroma enote, popisane večkrat.

Kontrola o polnoštevilnosti odgovorov je preprosta. V tej stopnji kontrole obrazce pregledamo, ali so na vsa zastavljena vprašanja vpisani odgovori.

Kontrola pravilnosti odgovorov je najtežja. Kontrola pravilnosti odgovorov je trojna: stvarna, logična in računska.

Stvarna kontrola odkrije napake, če je podatek sam zase neverjeten. Stvarna kontrola odkrije, če nekdo navede, da je star 105 let, da ima njegovo privatno gospodarstvo 200 ha obdelovalne površine itd. Te napake so običajno napake v pravem smislu, ker se je tisti, ki je podatke vpisoval, verjetno zmotil, brez namena, da bi potvarjal podatke.

Logična kontrola kontrolira posamezne podatke v medsebojni odvisnosti. Podatek more biti zase verjeten, v primerjavi z drugim pa je nelogičen. Tako stvarna kontrola ne odkrije napake, če je nekdo napisal, da je star dve leti, a da je po poklicu inženir. Eno in drugo je samo zase možno, v medsebojni zvezi pa je nelogično. Z logično kontrolo odkrijemo tudi manj nelogične odgovore.

Računska kontrola odkriva napake v numeričnih podatkih, ki so v medsebojni zvezi. Tako mora biti vsota izdatkov po vrstah stroškov enaka skupnim izdatkom, produkt cene in količine enak vrednosti itd.

Čeprav se izogibamo odvečnih podatkov, posebno, če moremo iz njih sami izračunati dodatne podatke, dostikrat namenoma postavljamo kontrolna vprašanja, da z računsko ali logično kontrolo odkrijemo napake, ki jih brez kontrolnih vprašanj ne bi odkrili.

Z zgornjimi vrstami kontrol pa ne moremo odkriti vseh napak. V popisnem gradivu so namreč lahko še napake, ki jih z nobeno od zgornjih kontrol ni moč odkriti. Določen podatek je verjeten v zvezi z vsemi drugimi podatki, pa je kljub temu napačen. Te vrste napak so najhujše in tudi najpogostejše.

Podatke, ki so verjetni, vendar nepravilni, more odkriti le dober poznavalec krajevnih

razmer ali oseba, ki popisano enoto pozna. Te napake odkrije tudi ponoven popis, ki ga izvedemo bolj natančno z dokumenti, pregledom itd. Jasno je, da taka kontrola ne pride v poštev, ker je pravzaprav ponovna izvedba opazovanja s strožjimi merili.

Da ugotovimo skupen učinek napak te vrste, dostikrat izvedemo kontrolni slučajni vzorec in za izbrane enote izvedemo ponoven popis. Pri tem vzorcu pa uporabimo vsa razpoložljiva sredstva, da dobimo popolne podatke. Primerjava podatkov, ki jih dobimo s kontrolnim vzorcem, s podatki, ki jih dobimo z osnovnim opazovanjem za iste enote, pokaže skupen učinek napak. Z njim moremo popraviti popisne podatke. Če je namreč vzorčna kontrola pri popisu živine pokazala, da smo s kontrolo v vzorčnih gospodarstvih ugotovili za 5% več živine kakor za ista gospodarstva pri popisu, moremo to rezultat posplošiti na celoto in skupno število živine te vrste po popisu povečati za 5%.

3.42 Podatke kontrolirajo organi opazovanja v vseh stopnjah. Kontrola podatkov na vsaki stopnji ima svoje pomanjkljivosti in prednosti. Prvi, ki pri samoregistraciji kontrolira podatke, je popisovalec, ko prevzema izpolnjene obrazce. Prednost te kontrole je v tem, da je razmeroma najcenejša, hitra in tudi učinkovita. Tako je najceneje, če popisovalec pri prevzemu kontrolira polnoštevilnost odgovorov, ker takoj dobi odgovore na neizpolnjena vprašanja. Enako more popisovalec, če pozna enoto ali razmere v svojem popisnem okolišu, odkriti veliko več napak v odgovorih, ki se zde na prvi pogled pravilni, kakor kontrolor, ki ne pozna razmer.

Kontrola na prvi stopnji pa ima svoje hibe. Je neenotna in ne more biti sistematična. Glede tega je boljša kontrola, ki jo izvedejo popisne komisije ali drugi vmesni organi med popisovalcem in središčem, ki opazovanje organizira. Kontrola v središču je sicer enotna, sistematična in kvalitetna, vendar le za napake, ki jih more odkriti običajna stvarna, logična ali računska kontrola. Napake, ki so take narave, da so podatki verjetni, kljub temu pa napačni, pa odkrije kontrola tem teže, čim bolj je odmaknjena od terena. Razen tega je popravljanje napak tem teže in tem bolj zapleteno, čim višji organ napako odkrije. Medtem ko napako, ki jo odkrije popisovalec, popravi popisovalec mimogrede pri prevzemu popisanega gradiva, je pri napakah, ki jih odkrije šele višji organ, potrebno dopisovanje, ponoven obhod terena itd.

ČETRTO POGLAVJE

UREJEVANJE IN OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

4.1 Ko z opazovanjem zberemo podatke o populaciji, ki jo proučujemo, je populacija dana v nepregledni množici posamičnih obrazcev. Ta množica obrazcev in podatkov nadomesti populacijo v obliki, ki je primerna za obdelavo. Zbrani obrazci so surovina, ki v nadaljnji obdelavi dà končni proizvod - statistične tabele, grafikone, preglede in parametre, s katerimi analiziramo pojav.

Osnovna obdelava statističnih podatkov je u r e j e v a n j e . Urejevanje sestoji iz grupiranja, prešteva nja in seštevanja statističnih enot in podatkov. Namen urejevanja so absolutni podatki o statistični populaciji, takó, da jih je z analitičnimi metodami možno obdelati in analizirati. Pri popisu prebivalstva npr. z urejevanjem ugotovimo sestavo prebivalstva po spolu, stanu, starosti, šolski izobrazbi, poklicu itd., skratka po vseh znakih in kombinacijah znakov, ki smo jih zbrali s popisom. Rezultate urejevanja prikazujemo s statističnimi vrstami v statističnih tabelah. Statistična vrsta je osnovni način prikazovanja statističnih podatkov, tabela pa eden izmed načinov za prikazovanje statističnih vrst. Osnovni podatki, ki jih dobimo z urejevanjem, so osnova za preračunavanja in analizo.

Urejevanje statističnega gradiva sestoji iz več različnih stopenj, ki so med seboj vezane časovno in vsebinsko. Urejevanje statističnega gradiva je za velika opazovanja tehnični posel, ki traja razmeroma dolgo. Ves napor za izboljšanje tehnike urejevanja je us-

merjen v to, da skrajšamo čas urejevanja. Vrednost podatkov določenega opazovanja je tem večja, čim prej so podatki po izvedenem opazovanju na razpolago. Obdelavo pospešimo, če zaposlimo več obdelovalcev. Vendar je število obdelovalcev omejeno, ker so potrebne za urejevanje kvalificirane osebe. Zaradi vseh naštetih razlogov na splošno uvajamo moderna sredstva elektronske obdelave podatkov. Ta je pri opazovanjih večjega obsega že skoraj izpodrinila ročno obdelavo. Vendar so tudi pri urejevanju določena dela, ki se ne dajo mehanizirati.

4.2 Po kontroli statističnega gradiva sestoji urejevanje statističnega gradiva iz različnih stopenj dela. Z načrtom urejevanja določimo in rešimo vprašanja za izvedbo naslednje stopnje urejevanja:

a) grupiranje vrednosti znakov.
b) šifriranje oziroma signiranje zbranih statističnih podatkov.
c) načrt za osnovno obdelavo.
d) način obdelave podatkov.

Kakor smo že nakazali pri obravnavanju znakov, imajo nekateri znaki razmeroma veliko, včasih tudi neomejeno število vrednosti. Naloga g r u p i r a n j a je za znake z velikim številom vrednosti združiti sorodne vrednosti npr. sorodne poklice, sorodne artikle itd., v grupe tako, da število grup ni preveliko. Ti grupni znaki dajo sicer bolj grobo sliko, vendar so zaradi manjšega števila grup bolj prikladni za obdelavo in dado preglednejšo sliko o pojavu. Grupiranje znaka je ozko povezano z vsebino in nameni proučevanja.

Čisto tehničnega značaja pa je š i f r i r a n j e oziroma signiranje podatkov. Da je nadaljnja obdelava podatkov čim bolj avtomatična, posamezne podatke v gradivu zaznamujemo s kratkimi oznakami - šiframi, pri obdelavi pa se oziramo le na šifre in ne več na osnovne podatke.

N a č r t o b d e l a v e določa, po katerih kombinacijah znakov ali grupah znakov bomo obdelali gradivo, da bomo dobili odgovor za dano proučevanje. Pri popisu prebivalstva ne iščemo sestave prebivalstva samo ločeno po spolu in starosti, temveč je pomembna sestava, ki razdeli prebivalstvo po spolu in starosti hkrati. Tako dobimo vpogled ne le v

spolno in starostno sestavo ločeno, ampak v spolno-starostno sestavo prebivalstva. Kombinirana obdelava dveh ali več znakov hkrati je ena izmed osnovnih metod obdelave in osnova za analizo odvisnosti med pojavi. Z načrtom obdelave glede na namen proučevanja določimo, katere kombinacije znakov so za proučevanje pomembne in ali nameravamo po teh kombinacijah znake preštevati ali seštevati.

V načrtu osnovne obdelave je treba tudi določiti, kako bomo obdelovali statistične podatke. Že pri sestavljanju obrazcev (glej primere obrazcev) posebno pa pri šifriranju podatkov, je treba upoštevati, ali nameravamo obdelovati podatke ročno ali strojno.

Rezultati urejevanja so prikazani v obdelovalnih tabelah, ki so prikrojene načinu obdelave. Šele iz obdelovalnih tabel sestavimo končne publikacijske ali analitične tabele, s katerimi prikažemo populacijo ali damo osnovo za analizo pojava.

GRUPIRANJE VREDNOSTI ZNAKOV

4.3 Grupne značilnosti uporabljamo že v vsakdanjem izražanju. Tako starost zaokrožujemo v starost v letih. Če zaokrožujemo starost na izpolnjena leta, vzamemo za 26 let stare vse, ki so stari več kot 26 in manj kot 27 let. Če pa zaokrožujemo starost na najbližje leto, štejemo za 26 let stare vse, ki so stari od 25 in pol do 26 in pol let. Vse, katerih starost je v navedenem razmiku, štejemo da so stari enako - 26 let. Prav tako tudi osebne dohodke ne navajamo v parah, marveč dinarjih.

Enako v pojmu delavec ali uslužbenec združujemo vse delavske ali uslužbenske poklice. Če za nekoga pravimo, da je delavec, zanj sicer ne vemo natančno, kakšen je njegov poklic; vemo pa, v kateri skupini poklicev je. Vsi delavski poklici so združeni pod skupnim pojmom delavec.

Tudi časovno v vsakdanjem govoru uporabljamo grupe in pravimo: ta in ta nesreča se je zgodila v nedeljo 25.1.1968. Tako je čas dogodka za dane potrebe zadosti določen, čeprav se je mogla po teh podatkih nesreča dogoditi v razmiku 24 ur. Včasih zadoščajo še širše časovne grupe. Na vprašanje, kdaj se je neka oseba poročila, ta odgovori: v letu 1956. Ta podatek, čeprav je dan z razmikom enega leta, povsem zadošča za opredelitev časa poroke.

Enako ravnamo tudi s krajevnimi opredelitvami. Včasih na vprašanje odkod smo, odgovorimo: iz Ljubljane, ne navedemo pa ulice, hišne številke, nadstropja in stopnišča. Ko rečemo, da smo iz Ljubljane, povemo, da je naše stanovanje na območju Ljubljane. Da so iz Ljubljane, odgovorijo na vprašanje, odkod so, vsi, ki so doma na območju Ljubljane.

Iz gornjih zgledov vidimo, da v vsakdanjem življenju in govoru ne zaokrožujemo samo numeričnih ampak na splošno vse znake. V kako velikih grupah se izražamo, je odvisno od potreb in razmer, v katerih smo. Včasih ni dovolj, če kdo reče, da je uslužbenec. Zato je treba poklic natančneje določiti. Enako je s krajem stalnega bivališča. Če vas v Ameriki nekdo vpraša, odkod ste, je dovolj, da rečete, da ste iz Jugoslavije, čeprav je to razmeroma širok pojem. Če vas pa nekdo, ki vé, da ste Ljubljančan, v Ljubljani vpraša, kje ste doma, morate navesti ulico, torej niti opredelitev, da ste Ljubljančan, ni zadostna.

4.4 V bistvu iz istega razloga in po istih načelih kakor v vsakdanjem življenju grupiramo znake tudi v **statistične namene**. Natančne vrednosti dostikrat niso potrebne. Število vseh možnih vrednosti znaka je za nekatere znake veliko ali celo neomejeno. Zato je običajno prikladneje, da sorodne vrednosti združimo v grupne vrednosti. Tako sicer izgubimo pri natančnosti, pridobimo pa pri preglednosti. Še en razlog, ki ga pri grupiranju v vsakdanjem življenju ni, je odločilen za grupiranje znakov v statistiki. Šele pri grupnih znakih se za večino znakov pokaže množičnost pojavljanja, ki je ena izmed bistvenih lastnosti množičnih pojavov. To bomo opazovali pri podatkih o porabi lesa v kmetijskih gospodarstvih, ko bomo obravnavali frekvenčne porazdelitve.

4.5 Preden bomo obravnavali grupiranje za posamezne vrste znakov, navedimo nekaj splošnih načel, ki veljajo za grupiranje vseh znakov.

Grupiranje mora biti izvedeno pri vseh znakih opazovanja enolično. Vse vrednosti znaka grupiramo v grupni znak tako, da vsaka osnovna vrednost znaka spada v eno in samo eno grupo. Napačno bi bilo grupiranje, pri katerem bi bila določena vrednost v dveh ali več grupah istočasno ali pa v nobeni grupi. To načelo zahteva, da smo posebno pazljivi pri mejnih vrednostih. Tako je pri časovnih grupah za dneve vprašanje, kam spada trenutek

ob koncu enega dne in v začetku drugega dne. Pri geografskih grupah more biti sporno, kam spadajo točke, ki so teoretično na meji med dvema geografskima območjema.

Grupe vseh znakov imajo tudi tole lastnost; grupe znakov moremo grupirati v grupe višje vrste enako, kakor grupiramo osnovne vrednosti znakov. Tako moremo občine, ki so grupe prve stopnje, grupirati v okraje, ki so grupe druge stopnje, te v republike kot grupe tretje stopnje itd. Enako moremo ure združevati v dneve, te v mesece, mesece v leta, leta v petletja. Podobno je tudi s stvarnimi znaki. Posamezne poklice grupiramo v najožje grupe poklicev, te grupe v širše grupe itd. Ta postopek ponavljamo, dokler grupiranje ne dovede do najširših grup oziroma do grupe, ki obseže vse vrednosti. Vsaka podrobnejša grupacija v tem stopnjevanju grup podrobneje določa poklice. Enako grupiramo predmete porabe, vzroke smrti itd.

Vrednosti enega in istega znaka moremo grupirati po različnih načelih, tako da za isti znak dobimo več različnih grupacij. Ker posamezne vrednosti združujemo v grupe po sorodnosti, daje načelo grupiranja merilo o sorodnosti. Načelo grupiranja je v tesni zvezi s predmetom opazovanja, namenom in potrebami proučevanja pojava. Artikle, ki so predmet zunanje trgovine, grupiramo v grupe po načelu surovine, iz katere je predmet pretežno izdelan ali po načelu uporabe. Enako sestavljamo geografske grupe po upravno-političnem načelu in so v posameznih grupah kraji, ki so v istih občinah, itd. Moremo pa geografske grupe sestaviti tudi glede na kmetijsko proizvodnjo; po tem načelu pa združujemo v grupe vse kraje, v katerih je kmetijska proizvodnja pretežno iste vrste. Tako dobimo kmetijske rajone. Ti so sestavljeni po sorodnosti krajev glede na kmetijsko proizvodnjo in se ne ozirajo na upravno-politično razdelitev. Grupiranje po upravno-političnem načelu pa se spet ozira samo na to, ali so posamezni kraji v isti občini itd.

Razen tega pri grupiranju vsaka grupa običajno dobi neko ime-novo vrednost, ki pojmovno združuje vse vrednosti znakov za ustrezno grupo. Grupe dobe posebna imena: geografske grupe npr. imena občin, republik itd., časovne grupe datume, mesece, leta itd., stvarne grupe nazive grup poklicev itd.

Razen skupnih načel in lastnosti grupiranja in grup, ki smo jih navedli, imajo posamezne vrste znakov svoje posebne probleme. Te bomo obravnavali ločeno za vsako vrsto posebej.

4.6 Krajevne znake običajno grupiramo v področja po upravno političnem vidiku. Ta grupacija je upravičena iz dveh razlogov. Porabnik statističnih podatkov, če ne gre za neke posebne raziskave, potrebuje za svojo uporabo podatke po upravno-politični razdelitvi. Razen tega so upravno politične enote prikladne grupe, ker so že izdelane v druge namene. Za potrebe posebnega proučevanja se samo naslonimo na to razdelitev terena. Hiba upravno-politične razdelitve pa je v tem, da se ta razdelitev časovno menja. Zato je krajevno proučevanje dinamike pojavov otežkočeno ter zahteva razmeroma težavna preračunavanja. Hiba pri grupiranju podatkov po upravno-politični teritorialni razdelitvi je tudi, da je načelo tega grupiranja za večino proučevanj formalno in ima zato omejeno analitično uporabnost. Marsikatera značilnost se namreč pri tem zabriše.

Za proučevanje pojava je vsekakor bolje grupiranje po načelu, ki je v vsebinski zvezi s pojavom, ki ga proučujemo. Če upoštevamo to načelo, razdelimo teritorij v vsebinske geografske grupe - r a j o n e . Geografske rajone izdelujemo po različnih merilih in za različne potrebe. Rajonizacijo pa spremljajo tehnične težave. Izvesti rajonizacijo po nekem vsebinskem načelu je običajno zelo težko, ker moramo teritorij in razmestitev pojava, ki ga proučujemo, zelo dobro poznati. Čeprav je za določeno proučevanje taka razdelitev najboljša, so težavę v tem, da je treba tako rekoč za vsako proučevanje posebej podrobno razdeliti teritorij na vsebinske grupe. Te grupe v splošnem sekajo upravno-teritorialno razdelitev. Običajno vzamemo kompromisno rešitev, da manjše administrativne enote npr. občine ne grupiramo dalje v republike, marveč po vsebinskem načelu po pretežnosti v vsebinske grupe - rajone. Ta način, čeprav ni najbolj natančen, omogoča rajonizacijo brez posebnih dodatnih težav. Razen tega pa moremo po potrebi grupirati iste podatke po obeh načelih: upravno-teritorialnem in vsebinskem.

Najmanjše krajevne grupe, ki so sestavljene v statistične namene, so s t a t i s t i č n i o k o l i š i . Ti so sestavljeni tako, da ne sekajo upravno-teritorialnih enot.

4.7 Za č a s o v n e z n a k e je sorodnost momentov običajno dana s časovnim razmikom med momenti. Po tem načelu dobimo naravne enote: minute, ure, dneve, tedne, mesece, leta. Izmed teh grup so najbolj problematični meseci, ker so različno dolgi (od 28 do 31 dne). Razlika med najkrajšim in najdaljšim mesecem je tri dni ali približno deset odstotkov. Še večje so razlike, če namesto koledarskih dni upoštevamo

samo delovne dni. S posebnimi prijemi reduciramo podatke tako, da lahko mesečne podatke primerjamo med seboj.

Za proučevanje nekaterih pojavov uporabljamo drugačna načela grupiranja časa. Za proučevanje sezonskih vplivov so si sorodnejši vsi januarji, vsi februarji, vsi marci za vsa leta v določenem razdobju kakor pa januar, februar, marec itd. istega leta, čeprav so prvi časovno bolj oddaljeni. Če proučujemo na primer turizem, gradbeništvo, poljedelstvo, je takoj razumljivo, da je res tako. Okoliščine za vsako od teh dejavnosti so v istih mesecih v zaporednih letih bolj sorodne kot v različnih mesecih istega leta. Enak problem je z grupiranjem dni v tednu, če proučujemo promet, prometne nesreče, rabo pijač itd. V tem primeru tvorimo grupe vseh ponedeljkov, vseh torkov itd. Enako sestavljamo grupe iz določenih ur iz zaporednih dni, če proučujemo nihanje v dnevni rabi električne energije ali vode, število prevozov v lokalnem prometu itd.

4.8 Grupiranje stvarnih znakov je bistveno različno za atributivne in za numerične znake. Grupiranje atributivnih znakov je najtežje. Čeprav postavimo za grupiranje atributivnih znakov npr. poklice, vzroke smrti, dejavnost, izdelke v zunanji trgovini, surovine v industriji itd. neko načelo grupiranja, s tem grupiranje še ni izvedeno kakor je npr. izvedeno krajevno grupiranje, če se odločimo za upravno-teritorialno načelo ali če določimo načelo grupiranja za časovne znake. Za stvarno-atributivne znake so z načelom postavljene samo smernice za grupiranje. Grupiranje posameznih poklicev, vzrokov smrti, dejavnosti itd. pa je obširno delo, ki ga morejo opraviti le strokovnjaki. Ti po načelih grupiranja sistematično klasificirajo vse mogoče posamezne poklice, vzroke smrti, dejavnosti itd. in izdelajo podrobne klasifikacije - nomenklature, v katerih je za vsak poklic, vzrok smrti, artikel itd. navedeno, v katero grupo spada. Klasifikacije in nomenklature so torej sistematično po grupah in podgrupah urejeni seznami poklicev, vzrokov smrti, artiklov itd.

4.9 Atributivne znake z velikim številom vrednosti običajno tehnično grupiramo po decimalni klasifikaciji. Osnova decimalne klasifikacije je v tem, da največ po deset osnovnih pojmov združimo v grupe prve stopnje, največ po deset grup prve stopnje v grupe druge stopnje itd. To grupiranje v grupe višjih stopenj ponavljamo toliko

časa, da imamo največ deset grup zadnje stopnje. To načelo ni vsebinsko temveč tehnično. Vsak pojem moremo namreč po decimalni klasifikaciji označiti s številom, ki ima toliko mest, kolikor stopenj ima grupiranje.

To je zelo prikladno. Prva številka označbe po decimalni klasifikaciji pove, v kateri najširši grupi je določen pojem, prvi dve mesti opredelita grupo po naslednji razdelitvi in tako dalje. Čim večmestno število vzamemo, tem bolj je pojem določen.

Za zgled podajamo prvo razdelitev blaga v statistiki zunanje trgovine (Vir: Nomenklatura statistike spoljne trgovine SFR Jugoslavije, ki jo je v letu 1957 izdal Zvezni zavod za statistiko):

0 Prehrambeni proizvodi.
1 Pijače in tobak.
2 Surovine, ki niso prehrambene (razen goriv).
3 Mineralna goriva, maziva in sorodni proizvodi.
4 Živalska in rastlinska olja in maščobe.
5 Kmetijski proizvodi.
6 Izdelki, klasificirani pretežno po materialu.
7 Stroji in transportne priprave.
8 Razni gotovi izdelki.
9 Razne transakcije in nikjer omenjeno blago.

Izsek iz te nomenklature za izdelke iz lesa in plute pa je takle:

63 Izdelki iz lesa in plute (razen pohištva)
631 Furnir, vezane plošče, deske, umetno ali rekonstruiran in drugi les, obdelan neomejeno
631-01 Furnir
631-01-11 Furnir bukov normalen m^3/kg
631-01-12 Furnir bukov za stole "
631-01-20 Orehov furnir "
631-01-30 Hrastov furnir "
631-01-40 Furnir slepi "
631-01-90 Furnir iz drugih listavcev "

Iz zgleda vidimo smisel in pomen decimalne klasifikacije. Prva številka v oznaki 631-01-11, s katero je zaznamovan normalen bukov furnir (6), pove, da normalen bukov furnir spada po klasifikaciji v grupo izdelkov, klasificiranih pretežno po materialu. Prvi dve številki (63) pomenita, da spada ta izdelek v podgrupo izdelkov iz lesa in plute(brez pohištva). Prve tri (631) številke oznake pomenijo, da spada navedeni izdelek v nadaljnjo podgrupo furnirjev, vezanih plošč, desk itd.

Decimalna klasifikacija je zelo prikladna, ker številčne oznake nazorno pokažejo v katero grupo po kateri koli grupaciji sodi posamezen pojav.

4.10 Pri grupiranju stvarno atributivnih znakov včasih vrednosti, ki jih ne moremo grupirati v grupe ali pa za proučevanje niso bistvene, združimo v grupo "drugo". Seveda moramo grupo "ostalo" tvoriti tako, da ni preobširna. Včasih tvorimo tudi grupo "n e z n a n o", v katero zaradi popolnosti pregledov vključimo vse enote, za katere iz katerega koli vzroka nimamo podatkov.

4.11 Grupiranje za vrednosti s t v a r n o - n u m e r i č n i h z n a k o v mora ustrezati vsebinskim in tehničnim pogojem. Ker je analiza numeričnih znakov najbolj razvita, se pri grupiranju numeričnih znakov oziramo na formalno-tehnična načela grupiranja, če le taka grupacija ni nedopustna iz vsebinskih razlogov.

Za zgled grupiranja zveznega numeričnega znaka vzemimo skupno površino privatnih kmetijskih gospodarstev. V"statističnem godišnjaku 1972" imamo naslednje grupe: -2 ha; nad 2 ha - 3 ha; nad 3 ha - 5 ha; nad 5ha - 8 ha in nad 8 ha in več.

4.12 Grupe za numerične znake, ki jih imenujemo r a z r e d e , so določene s s p o d n j o m e j o $y_{k,min}$ in z g o r n j o m e j o $y_{k,max}$. Nobena vrednost v razredu ni manjša od spodnje meje razreda in nobena ni večja od zgornje meje razreda. V posamezen razred spadajo vse vrednosti med spodnjo in zgornjo mejo razreda. Vsak razred ima svojo š i r i n o r a z r e d a i_k, ki je razlika med zgornjo in spodnjo mejo razreda

$$i_k = y_{k,max} - y_{k,min} \qquad (4.1)$$

in s r e d i n o r a z r e d a y_k, ki je polovica vsote mej razreda

$$\chi_i \; y_k = \frac{y_{k,min} + y_{k,max}}{2} \qquad (4.2)$$

Kakor smo navedli, ima na splošno vsaka grupa neko svojo vrednost. Tako vzamemo, da je pri numeričnih znakih sredina razreda reprezentant vseh vrednosti v razredu - torej grupna vrednost.

Iz zgornjega zgleda vidimo, da niso vsi razredi omejeni s spodnjo in zgornjo mejo. Zadnji razred v našem zgledu ima le spodnjo mejo (8 ha). Take razrede, ki so omejeni samo navzdol ali samo navzgor, imenujemo o d p r t e r a z r e d e . Odprte razrede tvorimo, če je število enot, ki ima vrednost nad ali pod neko mejo, majhno, podatki pa so med seboj zelo različni.

Za zgornjo grupacijo površin so širine razredov po vrsti: 2 ha, 1 ha, 2 ha, 3 ha. Za zadnji razred ne moremo izračunati širine, ker je odprt. Sredine razredov pa so tele: 1 ha, 2.5 ha, 4 ha, 6,5 ha. Za zadnji odprt razred tudi sredine razreda ne moremo izračunati.

Zgornji razredi ustrezajo načelu enoličnosti in popolnosti grupacije, ker vsaka vrednost spada v en in samo en razred. Pri zgornjem grupiranju štejemo vrednosti, ki so na meji razredov, v spodnji razred, ker je spodnja meja razredov označena z nad, kar pomeni, da spodnja meja ni vključena v razred.

Če hočemo meje razredov vključiti v zgornji razred, pa opišemo razrede takole:

0 ha	do pod 2 ha	0 do pod 2
2 ha	do pod 3 ha	2 - pod 3
3 ha	do pod 5 ha	3 - pod 5
5 ha	do pod 8 ha	5 - pod 8
8 ha	in več	8 - --

Obe zgornji grupaciji sta enolični in kompletni. Grupacija - 2 ha; 2 ha - 3 ha; 3 ha - 5 ha; 5 ha - 8 ha; 8 ha - pa ni enolična, ker mejne vrednosti spadajo v dva razreda.

4.13 Površine gospodarstev ne merimo natančno do mm^2, marveč jih običajno zaokrožujemo na are. Kakor smo že omenili, pa je zaokroževanje numeričnih znakov dvojno: zaokrožujemo jih na n a j b l i ž j o celo vrednost ali pa na n a j v e č j o celo vrednost v enoti, v kateri navajamo podatek. Tako gospodarstvo, ki ima 3 ha 27 a 58 m^2, po prvem načinu zaokrožimo na 3 ha 28 a, ker je stvarna površina bliže 3 ha 28 a kakor površini 3 ha 27 a, po drugem načinu pa je zaokrožena vrednost 3 ha 27 a, ker je 3 ha 27 a največja površina, navedena v arih, ki je manjša kakor stvarna površina gospodarstva.

Če upoštevamo zaokroževanje podatkov na are, pišemo zgornjo grupacijo površin takole:

0 ha - 1,99 ha
2,00 ha - 2,99 ha
3,00 ha - 4,99 ha
5,00 ha - 7,99 ha
8,00 ha -

Na prvi pogled se zdi, da spredna grupacija ni popolna, ker je npr. zgornja meja prvega razreda označena z 1,99 ha, spodnja meja naslednjega razreda pa z 2,00. Če pa upoštevamo, da so to zaokroženi podatki, ni nobenega dvoma o kompletnosti te grupacije. Vendar je od tega, kako zaokrožujemo, odvisno, katere vrednosti so meje razredov. Če zaokrožujemo na najbližjo celo vrednost v najmanjši enoti, zaokrožena vrednost 1,99 pomeni vse površine od 1,985 do 1,995 ha, zaokrožena vrednost 2,00 ha pa vse površine od 1.995 do 2.005 ha. Meja med prvim in drugim razredom je torej 1.995 ha. Če pa zaokrožujemo na največjo celo vrednost v najmanjši enoti, pa vrednost 1,99 pomeni vse površine od vključno 1,99 do pod 2,00, vrednost 2,00 ha pa vse površine od vključno 2,00 do pod 2,01 itd. Meja razreda pri tem zaokroževanju je 2,00 ha. Iz zgleda vidimo, da način zaokroževanja vpliva na meje razredov in po njih na sredine razredov. V splošnem da naravnejše razrede zaokroževanje na največjo celo vrednost, kar vidimo že iz zgleda. Pri prvem načinu zaokroževanja dobimo meje razredov 1,995 ha, 2,995 ha, 4,995 ha itd., pri drugem pa 2,00 ha, 3,00 ha 5,00 ha itd.

V zgornjem zgledu smo iz vsebinskih razlogov vzeli različne širine razredov.

Iz tehničnih razlogov primernosti grupacije za nadaljnjo statistično obdelavo pa težimo, da so razredi enako široki in da grupacije nimajo odprtih razredov.

Tako na primer mesečne osebne dohodke grupiramo v grupe

600 - 799 din
800 - 999 "
1000 - 1199 "
1200 - 1399 "
1400 - 1599 "
1600 - 1799 "
1800 - 1999 "
2000 - 2199 "
2200 - 2399 "
2400 - 2599 "
2600 - 2799 "
2800 - 2999 "
3000 - "

Širina razreda je v tej grupaciji enaka 200 N din. Zato se tudi sredine razredov 900, 1100, 1300, 1500, 1700, 1900, večajo v aritmetični postopici.

4.13 Grupiranje nezveznih znakov je podobno kakor grupiranje zveznih znakov, le da vprašanje mej za nezvezne znake v praksi ni tako pereče kakor pri zveznih znakih. Če grupiramo število zaposlenih v trgovinskih podjetjih v grupe po tri zaposlene, dobimo grupacijo nezveznega znaka po številu zaposlenih: 1-3, 4-6, 7-9, 10-12, 13-15, 16-18. Pri nezveznih znakih moramo paziti, katere vrednosti so meje razredov. Zgornja grupacija kaže, kot da ni popolna, ker se en razred konča s 3, drugi pa začne s 4. Vendar to ni res, ker more biti število zaposlenih samo celo število in med 3 in 4 ni vrednosti. Tako navajanje razredov pa moti pri izračunavanju širine razredov. Razlike med tri in ena je dva, v razredu so pa tri vrednosti: 1,2,3. Širina razreda je tri, ne pa dva.

Vsem zmedam se izognemo, če obravnavamo tudi nezvezne znake kot zvezne in vsaki vrednosti priredimo enotin razmik tako, da je ustrezna vrednost sredi tega razmika. Vrednosti 1 priredimo razmik od 0,5 do 1,5, vrednosti 2 razmik od 1,5 do 2,5 itd. Tako sta meji prvega razreda ne 1 in 3, ampak 0,50 in 3,5. Razlika teh mej pa je resnična širina razreda (3,5 - 0,5 = 3,0). Tako grupiranje in obravnavanje nezveznih znakov izenačimo z zveznimi. Pri zveznih znakih razmikom pripišemo kot reprezentanta sredino razreda, pri nezveznih pa posamezni vrednosti priredimo enotin razmik.

4.14 Sorodnost dveh numeričnih vrednosti merimo z razliko med vrednostima. Po tem merilu sta dohodka 800 in 850 din enako različna kot dohodka 2000 in 2050 din. Na prvi pogled pa spoznamo, da je razlika med dohodkoma 800 N din in 850 N din večja kot med dohodkoma 2000 in 2050 dinarjev, čeprav je v absolutnem razlika v obeh primerih 50 din. Za merilo razlik in sorodnosti ni vedno ugodna absolutna razlika, temveč je boljša relativna razlika. Kadar so možne velike razlike med vrednostmi znaka, so za merilo o sorodnosti vrednosti primernejše relativne kakor pa absolutne razlike. Za take primere sestavljamo razrede, za katere niso razlike, ampak kvocienti mej, konstante. Meje razredov za tako grupacijo niso aritmetična, ampak g e o m e t r i č n o z a p o r e d j e.

Zgled za tako grupacijo v naši statistiki je grupiranje industrijskih podjetij po številu delavcev.

Podjetja so po številu delavcev grupirana v tele grupe:

število delavcev	kvocient za meje
do 15	
	2.00
16 do 30	
	2.00
31 do 60	
	2.08
61 do 125	
	2.00
126 do 250	
	2.00
251 do 500	
	2.00
501 do 1000	
	2.00
1001 do 2000	
	-
2001 -	

Kvocienti med mejami so stalni (2,00), razen v razredu 61 do 125, v katerem je 2,08, da

so meje naslednjih razredov prešle na naravnejše meje.

Celoten cikel mej razredov s stalnim razmerjem 2 je v okviru tromestnega števila naslednji:

1 2 4 8 15 30 60 125 250 500 1000

Naprej gre na 1000 2000 4000 ... torej na ista števila kot zgoraj samo tisočkrat večja. V zgornji vrsti mej razredov sta izjemi le prehoda iz 8 na 15, pri katerem je razmerje 15/8 = 1, 88 in prehod iz 60 na 125, pri katerem je razmerje 125/60 = 2,08. Ta odstopa sta nebistvena, uspemo pa z njima zaokrožiti meje.

Cikel mej razredov v približnem geometrijskem zaporedju s kvocientom 2,5 je:

10 25 60 150 400 1000

Ta cikel se ponavlja v okviru dveh decimalnih mest. Le odstopanje v razmerju med zaporednima mejama je 60/25 = 2,4 in 400/150 = 2,67.

Cikel mej razredov v približnem geometrijskem zaporedju 3 je:

1 3 10

Ta cikel se zaključi v okviru enega decimalnega mesta. Odstop od razmerja 3 pa je 10/3 = 3,33

Cikel zaporedja s približnim razmerjem 4 pa je

1 4 16 60 250 1000

Od razmerja 4 odstopata 60/16 = 3,75 in 250/60 = 4,17

Zgornje sisteme z mejami v geometrijskih zaporedjih po potrebi priredimo v problemu ustrezne meje oziroma razrede, če jih pomnožimo z 10^k ali 5.10^k. Podobno moremo cikle mej ekstrapolirati v levo ali desno poljubno mnogokrat. Tako moremo iz cikla 1 3 10 izdelati meje za širšo grupacijo, ki začne npr. z 0,5. Tako dobimo meje

0,5 1,5 5 15 50 150 500 1500 5000 15000 50000

To načelo je uveljavljeno v dosti primerih v publikacijah Zveznega zavoda za statistiko SFRJ za pojave, ki so med seboj zelo različni.

ŠIFRIRANJE OSNOVNEGA GRADIVA

4.15 V osnovni statistični obdelavi podatkov preštevamo in seštevamo podatke po grupacijah in klasifikacijah, ki jih predpisuje načrt obdelave. Če sestavimo grupe za posamezne znake, delo še ni opravljeno. Sestavljene grupacije so samo navodilo, kako moramo grupirati. V nadaljnji obdelavi zbrano gradivo po izdelanih grupacijah sortiramo in dalje obdelujemo.

To delo si znatno olajšamo, če predhodno celotno gradivo pregledamo in pregledno označimo, v katero grupo po posameznem znaku spada posamezna enota opazovanja. Temu delu pravimo šifriranje ali signiranje gradiva. Oznakam za posamezne grupe pravimo šifre, seznam grup z ustreznimi šiframi pa je šifrant. Če gradiva ne šifriramo, moramo vsakokrat, ko pride v obdelavo določen znak, ponovno ugotavljati, v katero izmed grup posamezna enota po tem znaku spada.

Šifre morejo biti različne. Zelo pogoste so številčne šifre. Za šifre pa uporabljamo tudi črke ali kakšne druge oznake.

Pri številčnih šifrah z eno izmed zaporednih številk označimo posamezno vrednost znaka ali pri grupiranem znaku posamezno grupo. Tako morejo biti številčne šifre za stan: samski 1, poročen 2, razvezan 3, vdovec 4.; za spol: moški 1, ženski 2.

Številčne oznake pri decimalni klasifikaciji so idealne šifre. Glede na to, ali potrebujemo podrobno ali grobo grupacijo, vzamemo za šifro več ali manj-mestno oznako v decimalni klasifikaciji. Tako pri šifriranju blaga v zunanji trgovini po vrsti blaga vzamemo samo prvo številko, če obdelava predpisuje razdelitev blaga po najširših skupinah ali večmestne šifre, če obdelava predvideva podrobnejšo obdelavo.

Tudi črkovne šifre so prikladne, posebno, če je število grup manjše. V tem primeru je dobro vzeti za šifre prve črke oznak za vrednosti ali grupe. Tako vzamemo za

stan šifre: samski - S , poročen - P, razvezan - R in vdovec - V; ali za spol: moški - M, ženski - Ž. Za take primere so črkovne šifre primernejše od številčnih, ker si jih lažje zapomnimo.

Druge oznake uporabljamo redkeje. V praksi pa za znake, ki imajo samo dve gru-pi, včasih namesto številčnih ali črkovnih šifer uporabljamo oznaki + in -.

Medtem ko je pri načrtovani ročni obdelavi šifriranje priporočljivo in moremo izbirati med številčnimi in črkovnimi šiframi, moramo pri strojni obdelavi obvezno šifrirati podatke s številčnimi šiframi, ki so sestavljene po načelu decimalne klasifikacije.

OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

4.16 Z osnovno obdelavo dobimo absolutne podatke o statistični populaciji. Če hočemo dobiti vpogled v starostno strukturo prebivalstva, razdelimo popisno gradivo o popisu prebivalstva v grupe po starosti in preštejemo, koliko oseb oziroma popisnih obrazcev je v posameznih grupah. Do pregleda o sestavu prebivalstva pridemo z dvema stopnjama obdelave. Najprej gradivo razdelimo - r a z s o r t i r a m o po starostnih skupinah prebivalstva, nato pa p r e š t e j e m o , koliko oseb oziroma obrazcev je v posamezni grupi.

Če želimo dobiti vpogled v razdelitev zemlje po posameznih velikostnih skupinah privatnih gospodarstev, najprej razdelimo popisno gradivo po velikostnih skupinah, nato pa za vsako skupino seštejemo površino vseh gospodarstev v posameznih skupinah. Enako pridemo do podatkov o osebnih dohodkih po grupah delavcev itd. Tudi pri tej obdelavi delo opravimo v dveh stopnjah: najprej gradivo razdelimo - r a z s o r t i r a m o v grupe, po grupah pa s e š t e j e m o podatke o velikosti gospodarstev, podatke o dohodkih delavcev itd. Glavne stopnje osnovne obdelave so torej: sortiranje gradiva v grupe, preštevanje in seštevanje podatkov.

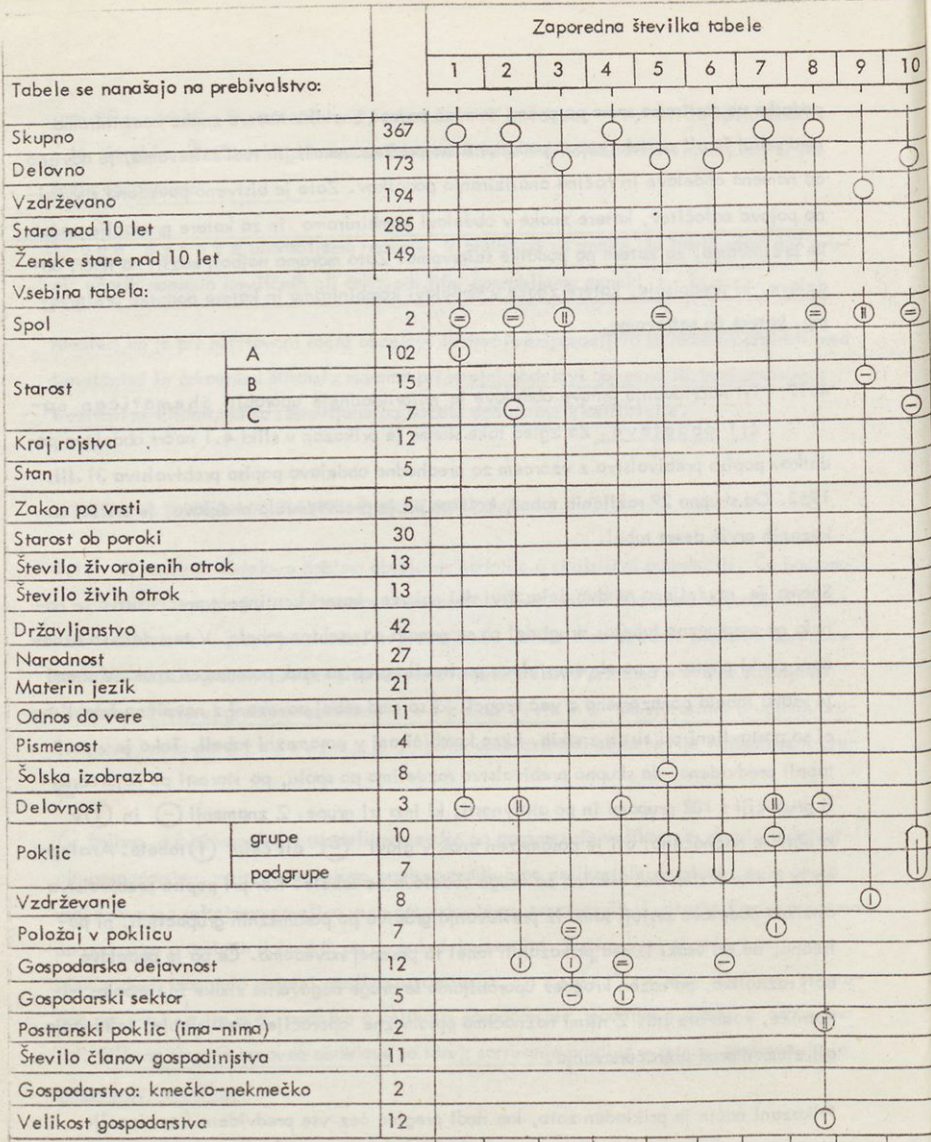
Obdelava statističnega gradiva je v praksi bolj zamotana. Navadno ne iščemo sestave populacij po enem, temveč po več znakih, ki so med seboj povezani, hkrati. V teh primerih

podatke ne sortiramo samo po enem, temveč po več znakih. Katere znake kombiniramo med seboj in ali po teh grupah podatke samo preštevamo ali jih tudi seštevamo, je odvisno od namena obdelave in načina analiziranja podatkov. Zato je bistveno povezana z vsebino pojava odločitev, katere znake v obdelavi kombiniramo in za katere grupacije podatke preštevamo, za katere pa podatke seštevamo. Zato moramo najbolj paziti na načrt obdelave, ki predpisuje, katere znake v obdelavi kombiniramo in katere podatke preštevamo, katere pa seštevamo.

4.17 Pri načrtovanju osnove obdelave je najprikladneje uporabiti shematičen načrt obdelave. Za zgled take sheme je prikazan v sliki 4.1 načrt obdelave podatkov popisa prebivalstva z vzorcem za predhodno obdelavo popisa prebivalstva 31.III. 1953. Od skupno 29 različnih tabel, kolikor jih je predvidevala obdelava, je v sliki nakazanih prvih deset tabel.

Shema je razdeljena na dva dela: Prvi del nakaže, kateri kontingent prebivalstva se nanaša na posamezno tabelo, drugi del pa se nanaša na vsebino tabele. V tem delu so označeni znaki popisa, v posebnem stolpcu pa število grup za vsak posamezen znak. V shemi je vsaka tabela ponazorjena z več krogci, ki so med seboj povezani z navpično črto. Krogci so postavljeni pri tistih znakih, ki so kombinirani v posamezni tabeli. Tako je v prvi tabeli predvideno, da skupno prebivalstvo razdelimo po spolu, po starosti po najdrobnejši grupaciji s 102 grupami in po aktivnosti, ki ima tri grupe. Z znamenji ⊖ in ⊕ v krogih je naznačeno, ali je posamezen znak v glavi ⊖ ali čelu ⊕ tabele. Analogno iz sheme razberemo vsebino za druge obdelovalne tabele. Ker pri popisu prebivalstva osnovna obdelava sestoji samo iz preštevanja gradiva po posameznih grupacijah, ni potrebno, da pri vsaki izmed prikazanih tabel to posebej navedemo. Če pa je obdelava bolj raznolika, pa razen krogcev uporabljamo še druge dogovorne znake in simbole: trikotnike, kvadrate itd. Z njimi naznačimo posamezne operacije: preštevanje, seštevanje ali elementarna preračunavanja.

Prikazeni način je prikladen zato, ker nudi pregled čez vse predvidene kombinacije hkrati in moremo zlahka ugotoviti, ali smo kateri znak ali kombinacijo iz obdelave izpustili ali pa določeno kombinacijo dvojno predvideli.



Slika 4.1 Shema načrta obdelave vzorca predhodnih rezultatov popisa prebivalstva 21.3.1953 (Vir: Popis stanovništva 1953 knjiga XI)

Shema načrta obdelave pa v končni obliki ne kaže samo obdelave gradiva, marveč tudi kakšne so obdelovalne tabele, v katere vnašamo podatke obdelave. Skladno s shemo, v kateri je pri tabeli 1 vodoravna črtica postavljena v krogcih pri spolu z dvema grupama in pri aktivnosti s tremi grupami, navpična črtica pa v krogcu pri starosti z 102 grupami, sta v obdelovalni tabeli znaka spol in aktivnost v glavi, znak starost pa v čelu tabele.

4.18 Seštevanje statističnih podatkov ni brez zapletljajev. Neposredno moremo seštevati le istovrstne podatke. Če seštevamo dnevno proizvodnjo za določeno vrsto cigaret za določeno podjetje, dobimo mesečno proizvodnjo cigaret te vrste. Če seštevamo mesečne osebne dohodke za posamezne uslužbence, dobimo skupen fond za osebne dohodke vseh uslužbencev v določeni delovni enoti itd. V navedenih zgledih je seštevanje podatkov smiselno in dobimo s seštevanjem smiselno količino. Iz dnevne proizvodnje dobimo mesečno proizvodnjo, iz osebnih dohodkov posameznih uslužbencev pa fond osebnih dohodkov za ustanovo.

Problem pa se zamota že pri seštevanju istovrstnih količin. Vzemimo število kilometrov, ki jih prepotuje v določenem razdobju posamezen potnik. Če seštejemo število kilometrov, ki jih je prepotoval posamezen potnik, za vse potnike, dobimo vsoto, katere pomen ni več tako jasen kakor v zgornjih primerih. S seštevanjem poti posameznih potnikov dobimo skupno število kilometrov, ki so jih prepotovali vsi potniki. Ta vsota pa se zdi, da je brez pomena. Razumljivo je, da je potnik, ki je prepotoval prvi dan 100 kilometrov, drugi dan 130 km, tretji dan 150 km, skupno prepotoval 380 kilometrov, manj jasna pa je vsota, če seštevamo poti različnih potnikov. Vendar imajo te vrste podatkov v prometni statistiki velik pomen. Da se izognemo napačnemu pojmovanju tega podatka, damo vsoti poti vseh potnikov novo enoto mere - potniški kilometri. Če je prvi potnik prepotoval 30 kilometrov, drugi 65 kilometrov, tretji pa 34 kilometrov, je podjetje, ki jih je prevozilo, opravilo 30 + 65 + 34 = 129 potniških kilometrov. Po enakem sklepu dobimo v prometni statistiki vlakovne kilometre, če seštevamo dolžino poti posameznih vlakov; če seštevamo prevožene kilometre, ki so jih prevozili posamezni železniški vozovi, dobimo vozovne kilometre itd. Tudi v industrijski statistiki naletimo na podobne probleme. Število ur, kolikor je v določenem mesecu delal posamezen delavec, seštevamo v skupno število delavec - ur ali skupno število opravljenih delovnih ur v podjetju. Isti problem imamo, če seštevamo čas dela pri strojih: tu dobimo s seštevanjem ur dela za posamezen stroj strojne ure za podjetje itd.

Medtem ko imajo zgoraj navedene vsote vsaj posreden smisel, seštevanje nekaterih istovrstnih podatkov nima smisla. Tako je sama zase brez smisla vsota starosti posameznikov za določeno populacijo, vsota cen za določeno vrsto blaga za različne trge itd.

4.19 Poseben problem je v statistiki seštevanje raznovrstnih količin. Dostikrat seštevamo dobesedno hruške in jabolka. Seštevanje proizvodnje premoga za različne rudnike se zdi na prvi pogled brez posebnih težav. Vendar je ta vsota smiselna le, če jo proučujemo kot skupen delovni učinek vseh rudnikov. Če pa pomislimo, da je kakovost premoga za različne premogovnike od lignita mimo različnih vrst premoga do črnega premoga med seboj bistveno različna, je upravičen pomislek, ali smemo proizvodnjo različnih premogovnikov seštevati. Enako je upravičen pomislek, ali je vselej smiselno preštevati npr. konje v določeni občini. Pri preštevanju štejemo kot istovrstno enoto žrebička, starega manj kot leto dni in težkega delovnega konja. Enako je vprašanje ali je prav v zunanje trgovinski statistiki seštevati težo posameznih pošiljk, ker je v enem primeru malo vredna ruda, v drugem pa precizna in draga aparatura.

Odgovor, ali ima seštevanje teh količin smisel ali ne, je odvisen od tega, za kakšne potrebe seštevamo in zbiramo podatke. Če zbiramo podatke o proizvodnji premoga zato, da ugotovimo skupni delovni učinek premogovnikov, ima vsota raznovrstnih podatkov smisel. Ta vsota proizvodnje pa nima smisla, če hočemo z njo izraziti ekonomski pomen te proizvodnje. Določena količina črnega premoga ima povsem drugačen pomen kakor enaka količina lignita.

4.20 Raznovrstne količine, ki imajo različne enote mere, pa sploh ne moremo seštevati. Podjetje, ki izdeluje različne izdelke, svoje mesečne proizvodnje ne more sešteti in izraziti v enem podatku. Enako trgovina ne more izraziti svojega skupnega prometa v naturalnih enotah mere, ker prodaja popolnoma različno blago.

Na skupno enoto mere privedemo količine v naturalnih enotah mere tudi tako, da količine izrazimo v vrednosti. Vrednost proizvodnje za vsak posamezen proizvod je produkt proizvedene količine in cene, skupna vrednost proizvodnje pa je vsota vrednosti proizvodnje v dinarjih za vse proizvode. Z obrazcem moremo to izraziti

$$V = \sum_{i=1}^{N} p_i q_i \qquad (4.3)$$

Pri tem pomeni: V = skupna vrednost, p_i = cena za posamezen artikel, q_i = ustrezna količina $\sum$ = znak za seštevanje.

4.21 Načelo, da posamezne k o l i č i n e t e h t a m o glede na njihov pomen in vrednost, kakor ga uporabljamo pri sestavljanju skupne vrednosti, pri kateri količine tehtamo s ceno, v ekonomski statistiki razširimo.

Za premog moremo uporabiti za težo ali ponder ceno posameznih vrst premoga, ker je cena skladna s kvaliteto premoga. Vendar je objektivnejše merilo kakovosti kalorična vrednost premoga. Po tem načelu vzamemo za enoto tono premoga z določeno kalorično vrednostjo. Ta premog, ki ga označimo z 1, imenujemo pogojno enoto. Premog, za katerega je kalorična vrednost samo tri četrtine kakovosti za pogojno enoto, označimo s koeficientom 0,75, premog, ki ima za 20% višjo kalorično vrednost, pa s koeficientom 1,20 itd. Podobno kakor smo dobili skupno vrednost, če smo produkte količin s cenami sešteli, izrazimo proizvodnjo premoga v p o g o j n i h e n o t a h premoga določene kakovosti tako, da proizvodnjo za posamezno vrsto premoga pomnožimo z ustreznim koeficientom, te produkte pa seštejemo. Dobljena vsota pomeni, kolikim tonam pogojne kakovosti ustreza stvarna proizvodnja premoga. Ta podatek ima večji ekonomski pomen kakor pa navadna vsota proizvodnje za vse vrste premoga ne glede na kakovost.

Na podoben problem tudi naletimo, če preštevamo motorna vozila, traktorje itd. Jakost raznih traktorjev je različna. Pri enostavnem preštevanju pa jih izenačujemo. Do pravilnejše slike o jakosti traktorskega parka pridemo, če podobno, kakor smo to naredili pri premogu, traktor z določeno močjo vzamemo za pogojni traktor. Za druge vrste in tipe traktorjev pa glede na jakost določimo ustrezni koeficient, ki je manjši od 1, če je slabši in večji od 1, če je močnejši. Z enakim postopkom, da vsak posamezen traktor ne štejemo za 1, temveč z ustreznim koeficientom, skupna vsota prikaže, koliko je skupno število traktorjev, če jih izrazimo v pogojnih traktorjih. Ta podatek je sicer fiktiven, podaja pa boljšo stvarno moč traktorskega parka kakor število traktorjev.

Tudi v kmetijski statistiki upoštevamo enako načelo. V statistiki živinoreje vzamemo za pogojno žival 400 kg težko govedo. Vse druge živali izražamo s koeficienti v razmerju s to pogojno-normalno živaljo. Tako ponderiramo posamezne vrste živali glede na velikost in proizvodno moč. Ta način omogoča, da seštevamo govejo živino, konje, prašiče, ovce itd. v skupno število normalnih živali. Če vzamemo za zgled konje, imamo za

posamezne skupine tele koeficiente: Žrebeta do enega leta 0,6; žrebeta od enega leta do treh let starosti 1,0; kobile nad tri leta 1,4; žrebci 1,3; konji nad tri leta 1,5. Če je na nekem področju 100 žrebet do enega leta, 200 žrebet od enega leta do treh let, 250 kobil nad tri leta, 50 žrebcev in 400 konj nad tri leta, je skupno: 100.0,6+ 200.1,0+ 250.1,4+ 50.1,3+ 400.1,5 = 1275 normalnih živali. Čeprav je ta podatek fiktiven, ga moremo s pridom uporabljati za primerjave med območji, ker z enim podatkom izrazimo skupnost vseh živali na določenem območju, v določenem velikostnem razredu itd.

Osnovna obdelava podatkov

Ročna obdelava. Pri ročni obdelavi statističnega gradiva ločimo v glavnem dva osnovna načina obdelave: a) odlaganje obrazcev in b) črtkanje.

4.22 Pri odlaganju obrazcev popisne obrazce razdelimo v grupe tako, da posamezne popisne obrazce odložimo na obdelovalni mizi na mesto, ki je določeno za posamezno grupo. Če moramo določeno število popisnic razdeliti po spolu, popisnice na obdelovalni mizi odlagamo na dva kupa. Vsak obrazec posebej pregledamo in ko ugotovimo, kakšen je spol osebe, na katero se obrazec nanaša, obrazec odložimo na ustrezno mesto, kjer zbiramo na enem kupu popisnice za moške, na drugem pa popisnice za ženske. Če načrt obdelave predvideva kombinirano obdelavo po spolu in stanu, popisnice za moške podobno dalje z odlaganjem razdelimo na štiri grupe - kupe: moški-samski, moški-poročen, moški-razvezan in moški-vdovec. Če storimo isto s popisnicami za ženske, dobimo podobne grupe: ženske-samske, ženske-poročene, ženske-razvezane in ženske-vdove. Tako z dvojnim odlaganjem obrazcev dobimo osem grup za kombinacijo znakov spol x stan. Po enakem načelu z odlaganjem razdelimo populacijo na grupe za poljubno kombinacijo znakov. Temu postopku pravimo s o r t i r a n j e osnovnega gradiva po grupah.

Številčen pregled o razdelitvi populacije v grupe po predpisani kombinaciji znakov dobimo, če p r e š t e j e m o po grupah število enot oziroma obrazcev in rezultate vpišemo v obdelovalno tabelo. Če je v načrtu obdelave predvideno seštevanje določenega podatka po grupah, s e š t e j e m o ustrezen podatek za vse enote v posameznih gru-

pah. Tako seštevamo površino, donos, število posameznih vrst živine v obrazcih, ki smo jih prej razsortirali po velikostnih skupinah, po skupni površini. Pri tem ne štejemo, da je obdelava strojna, če podatke seštevamo z običajnimi računskimi stroji.

Metoda odlaganja je zelo uporabna v vseh primerih. Kontrola sortiranja je razmeroma lahka in morebitne napake pri odlaganju kasneje zlahka odkrijemo. Obdelava z odlaganjem ima praktično neomejeno uporabo. S postopnim sortiranjem po posameznih znakih populacijo razdelimo z odlaganjem v poljubno število kombinacij znakov. Obseg populacije obdelave ne zapleta, temveč jo samo poveča.

Vendar moremo odlaganje uporabiti samo pri opazovanjih s posamičnimi obrazci, ni pa uporabno za kolektivne obrazce. Če so obrazci kolektivni, navadno prepišemo šifre podatkov na posamične obdelovalne listke, te pa potem po znanem postopku odlaganja samostojno obdelujemo.

4.23 Č r t k a n j e uporabljamo pri manjših obdelavah. Če hočemo dobiti za dano populacijo prebivalstva sestavo po spolu, urejujemo podatke s črtkanjem enostavno takole: Sestavimo obdelovalno tabelo, ki ima za posamezno grupo primerno veliko polja. Če razdeljujemo populacijo po spolu, ima obdelovalna tabela samo dve polji: eno za moške, drugo za ženske. Ko po vrsti pregledujemo v katero grupo spadajo posamezne enote, naznačimo to za vsako enoto posebej v ustreznem polju v obdelovalni tabeli s črtico. Če je prva popisnica za moškega, napravimo črtico v polju za moške. Če je druga popisnica tudi za moškega, napravimo zraven prve črtice drugo črtico. Če je tretja popisnica za žensko, napravimo črtico v polju za ženske. Ta postopek ponavljamo za vsako popisnico, dokler populacije ne izčrpamo. Če preštejemo črtice v polju za moške, dobimo število moških v populaciji. Analogno dobimo število žensk, če preštejemo črtice v polju za ženske. Skupno število vseh črtic pa je enako številu enot v populaciji.

Da si olajšamo štetje črtic po končani obdelavi, včasih že med obdelavo sestavljamo skupine po pet črtic, tako da s peto črtico v vsaki skupini po pet črtic prve štiri prečrtamo: 𝍸

Drug način, po katerem združujemo po deset enot je: ⁞X⁞ s štirimi oglišči, štirimi sta-

nicami in dvema diagonalama kvadrata registriramo deset vrednosti.

Vzemimo populacijo 25 oseb. Zanje je spol označen z M za moške in z Ž za ženske. Osnovni podatki o enotah so: M Ž M M Ž M M Ž Ž Ž M Ž M Ž M M Ž Ž Ž Ž M M Ž Ž Ž Ž . Po metodi črtkanja dobimo takle rezultat obdelave:

	Prvi način		Drugi način	
Moški	𝍸 𝍸 /	11	⊠	11
Ženske	𝍸 𝍸 ////	14	⊠	14
		25 = N		25 = N

4.24 Črtkanje uporabljamo, kadar obdelava narekuje preštevanje gradiva. Če moramo glede na obdelavo po grupah seštevati podatke za nek numeričen znak, moremo podobno kakor pri črtkanju sestaviti obdelovalno tabelo; v ustrezna polja pa ne rišemo črtice, temveč vpisujemo ustrezne vrednosti za numerični znak za posamezno enoto.

Vzemimo, da je populacija sestavljena iz desetih enot, od katerih so neke popisnice za delavce - D, druge pa za uslužbence - U. Imamo pa še podatke o dohodkih za teh deset enot (v din). Osnovni podatki za posamezne enote so: D-690, U-930, D-810, D-910, U-1150, U-950, D-1230, U-960, U-1170, U-940. Da dobimo skupni fond dohodkov po grupah za delavce in uslužbence, po zgornjem postopku obdelamo podatke takole:

							število	fond v din
Delavci	690,	810	910,	1230,			4	3640
Uslužbenci	930,	1150,	950,	960,	1170,	940	6	6100
							10	9740

Razen vsot dobimo število enot v posamezni grupi, če vpisane podatke preštejemo.

Črtkanje je uporabno vselej: pri individualnih in kolektivnih obrazcih ali če so podatki navedeni za vse enote na enem listu papirja kakor je to v naših zgledih. To je njegova prednost. Črtkanje pa ima svoje hibe, ki omejujejo uporabnost. Črtkanje je neprikladno za obdelave, v katerih kombiniramo več znakov. Za take obdelave je iskanje ustreznega polja v obdelovalni tabeli zamudno. Razen tega je za populacije velikega obsega

črtkanje neuporabno, ker nimamo kontrole. Napake moremo naknadno odkriti le s ponovno obdelavo. Pri črtkanju si pomagamo tako, da osnovno populacijo razstavimo v več obdelovalnih - delovnih populacij. Vendar tudi ta rešitev ni vselej ustrezna. Zato črtkanje uporabljamo le za manjše populacije, ne pa za zamotane obdelave.

4.25 Avtomatizirana obdelava podatkov. Ročna obdelava podatkov, pa četudi pri tem uporabljamo klasične računske stroje, ni kos večjim statističnim proučevanjem. Z razvojem elektronskih računalnikov so se posebej statističnim proučevanjem odprla vrata na stežaj. Možnost shranjevanja podatkov, hitrost in kompleksnost avtomatizirane obdelave podatkov je revolucionarno spremenila statistično obdelavo. Čas obdelave se je skrčil na minimum, obseg dela in metodologija pridobivanja statističnih informacij pa se od generacije do generacije računalnikov spreminja tako po hitrosti kot po enostavnosti upravljanja.

Sistem elektronskega računalnika ima pet glavnih elementov: centralno enoto, pomnilnik, enoto za računanje in vhodne in izhodne elemente. Centralna enota koordinira delo celotnega sistema za obdelavo podatkov. V pomnilnik shranjujemo osnovne podatke, ki jih obdelujemo, programe in vmesne in končne rezultate. Najpogosteje so nosilci podatkov v pomnilniku magnetna jedra v obliki diskov ali magnetnih trakov. Računski elementi posredujejo računske operacije, ki jih izvaja računalnik. Z avtomatizirano transformacijo iz dekadnega številčnega sistema v binarni sistem in obratno je omogočeno delo računalnika.

Vhodni elementi podatke prek določenih nosilcev informacij čitajo in prenašajo do pomnilnika. Vhodni elementi morejo biti: čitalci luknjanih kartic, čitalci magnetnih trakov, čitalci luknjanih papirnatih trakov, optični čitalci, sprejemniki akustičnih (govornih) elementov in sprejemniki raznih fizikalnih količin. Izhodni elementi računalnika so: hitri pisalni stroj, akustični izhodni elementi, ekrani, izhodi na mikrofilm, izhodi na luknjane kartice, luknjane papirnate trakove ali magnetne trakove in priprave za risanje. Najširšo uporabo ima vsekakor hitri pisalni stroj.

S terminali povezujemo vhodne in izhodne enote s centralnim računalnikom na velike razdalje. S to dislocirano konfiguracijo računalnika se poveča možnost uporabe velikega računalnika za več uporabnikov hkrati.

Pri koriščenju računalnika je poleg tehnične opremljenosti osnovnega pomena tudi intelektualni del v zvezi z zbiranjem in obdelavo. Tega sestavlja niz programov in podprogramov. V praksi je vse večji poudarek tudi na tem delu. Uporabnost računalnika se z izdelavo standardnih paketov programov, od katerih so nekateri že vgrajeni v računalnik, bistveno poveča.

Z avtomatizirano obdelavo podatkov ne izvajamo samo osnovnih obdelav statističnih podatkov ampak tudi in predvsem analitično obdelavo podatkov. AOP je omogočila praktično uporabo najrazličnejših metod statistične analize, katerih brez nje zaradi obsežnosti in zapletenosti ne bi mogli izvesti.

PETO POGLAVJE

PRIKAZOVANJE STATISTIČNIH PODATKOV

5.1 Statistične podatke redkokdaj navajamo posamič, ker je bistvo statističnega proučevanja pojavov v primerjavi med podatki. Sorodne statistične podatke združujemo v statistične vrste, te pa prikazujemo v tabelah in grafikonih. Tabelarični in grafični prikaz sta izvirna načina prikazovanja statističnih podatkov. Tabelarično in grafično prikazovanje statističnih podatkov se po svojih sredstvih med seboj razlikujeta. Imata pa isti namen: čim nazorneje in pregledneje prikazati proučevano populacijo, tako da je analiza pojava čim lažja.

Sredstva, s katerimi to dosežeta pa so različna. Zato grafično prikazovanje ni enakovredno tabelaričnemu prikazovanju statističnih podatkov ali narobe. Vsak način ima svoje prednosti in pomanjkljivosti. Neredko se tabelarični in grafični prikaz istih podatkov medsebojno dopolnjujeta in skupno pripomoreta k čim boljši sliki in analizi proučevanega pojava.

Prednosti tabelaričnega prikaza statističnih podatkov so predvsem: a) v tabeli moremo prikazati po potrebi razmeroma veliko podatkov; b) v tabeli prikažemo podatke s poljubno natančnostjo; c) tabelarični način je enostavnejši kakor grafični, ker so načela za sestavljanje dobrega tabelaričnega prikaza enotnejša kakor pri grafičnem prikazovanju, pri katerem za različne statistične vrste in v različne namene uporabljamo različne metode grafičnega prikazovanja.

Prednost grafičnega prikazovanja v primerjavi s tabelaričnim prikazom pa je

predvsem v tem: a) v grafikonu nazorno odkrijemo zveze in odnose za več podatkov hkrati, to v tabelaričnem prikazu ni možno; b) grafikon, s katerim hočemo popularizirati določen pojav, je bolj privlačen in neposreden kakor tabelaričen prikaz; c) z analitičnimi grafikoni statističnih podatkov uspemo včasih analizirati pojav neposredno brez dodatne analitične obdelave. Dostikrat pa da grafikon smernice za podrobno statistično obdelavo, kot je to npr. pri proučevanju odvisnosti med pojavi, proučevanju dinamike ekonomskih pojavov itd.

STATISTIČNE VRSTE

5.2 Statistična vrsta je niz sorodnih statističnih podatkov, od katerih se vsak nanaša na eno izmed vrednosti ali na grupo določenega znaka.

Glede na to, po kakšnem znaku se razlikujejo med seboj posamezni členi statistične vrste, razlikujemo: krajevne, časovne in stvarne statistične vrste. Stvarne statistične vrste pa dalje delimo v atributivne in numerične vrste, glede na to, ali statistična vrsta prikazuje podatke porazdeljene po atributivnem ali numeričnem znaku. Iz opredelitve statistične vrste sledi, da se posamezni členi vrste nanašajo na vrednosti osnovnega znaka.

5.3 V tabeli 5.1 je prikazana kot zgled k r a j e v n e a l i g e o g r a f s k e v r s t e porazdelitev narodnega dohodka v SFRJ v letu 1965 po socialističnih republikah.

Tabela 5.1 Narodni dohodek v SFRJ v letu 1965 po socialističnih republikah v mlj din (vir: SGJ - 67)

SR	Narodni dohodek v mlj din
SFRJ	73.570
BiH	9.022
Črna gora	1.244
Hrvatska	19.422
Makedonija	3.907
Slovenija	11.263
Srbija	28.712

5.4 Časovna vrsta je npr. proizvodnja električne energije v SFRJ v razdobju 1952-1971 v tabeli 5.2, ki prikazuje dinamiko v proizvodnji električne energije v tem razdobju.

Tabela 5.2. Proizvodnja električne energije v SFRJ v razdobju 1952-1971 v mlj kWh (Vir: SGJ-72)

Leto	1952	1953	1954	1955	1956	1957	1958	1959	1960
Proizvodnja	2700	2982	3440	4340	5048	6252	7356	8106	8928

Leto	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
Proizvodnja	9924	11275	13535	14189	15523	17174	18702	20641	23375	26023	29509

Statistična vrsta proizvodnje električne energije v tabeli 5.2 je intervalna ali razmična časovna vrsta, ker prikazuje intervalne oziroma razmične podatke. Trenutna časovna vrsta pa je npr. število zaposlenih v družbenem sektorju na dan 30. septembra v razdobju 1957-1971 v SFRJ v tabeli 5.3.

Tabela 5.3 Število zaposlenih dne 30. septembra v letu v razdobju 1957-1971 v SFRJ (Vir: SGJ-72)

Leto	1957	1958	1959	1960	1961	1962	1963
Število zaposlenih v tisočih	2480	2634	2829	3072	3366	3442	3532

Leto	1964	1965	1966	1967	1968	1969	1970	1971
Število zaposlenih v tisočih	3743	3690	3646	3606	3666	3806	3957	4133

Statistična vrsta v tabeli 5.3 je trenutna časovna vrsta, ker prikazuje število zaposlenih v določenem trenutku v vsakem letu.

5.5 Stvarno atributivna statistična vrsta je npr. število zaposlenih v družbenem in privatnem sektorju dne 30. septembra 1966 po gospodarskih dejavnostih, ki je prikazana v tabeli 5.4.

Tabela 5.4. Število zaposlenih v družbenem in privatnem sektorju v gospodarstvu v SFRJ na dan 30.septembra 1966 po panogah dejavnosti (Vir SB 515).

Panoga dejavnosti	Število zaposlenih v tisočih
Industrija in rudarstvo	1372
Kmetijstvo	313
Gozdarstvo	77
Gradbeništvo	328
Promet	247
Trgovina in gostinstvo	361
Obrt	241
Komunala	102
Kulturna in socialna dejavnost	428
Družbene in državne službe	166

5.6 Zgled za zvezno numerično vrsto so poprečni skupni izdatki za hrano na eno gospodinjstvo za štiričlanska delavska gospodinjstva z enim zaposlenim po višini skupnih razpoložljivih sredstev.

Tabela 5.5 Poprečni skupni denarni izdatki za hrano na eno gospodarstvo v SFRJ za gospodinjstva s štirimi člani, z enim zaposlenim po višini skupnih razpoložljivih sredstev v letu 1963 (Vir SB 429).

Skupna letna razpoložljiva sredstva v tisoč din	Poprečni skupni letni izdatki za hrano v tisoč din
do 199	100.9
200 - 299	129.0
300 - 399	163.2
400 - 499	203.0
500 - 599	236.1
600 - 699	267.2
700 - 799	299.6
800 - 899	331.4
900 -	369.1

5.7 Nezvezna numerična statistična vrsta pa je število anketiranih kmetijskih gospodarstev po številu članov v kmetijski anketi v letu 1956 v Sloveniji.

Tabela 5.6. Število gospodarstev v kmetijski anketi v Sloveniji v letu 1956 po številu članov v gospodinjstvu (Vir SB 113).

Število članov	Število gospodarstev
Skupno	317
1	8
2	28
3	43
4	58
5	74
6	52
7	33
8	12
9 in 10	6
11 in 12	3

Prikazane statistične vrste so po svojem značaju med seboj različne tudi po drugih lastnostih in ne samo po tem, da imajo za osnovo različne vrste znakov. Tabela 5.1 prikazuje s e s t a v o skupnega narodnega dohodka v SFRJ po republikah. Enak značaj ima tudi statistična vrsta v tabeli 5.4, ki prikazuje sestavo števila zaposlenih po panogah dejavnosti.

Prav poseben pomen imajo statistične vrste kakršna je npr. število kmetijskih gospodarstev po številu članov v tabeli 5.6. Ta statistična vrsta prikazuje variabilnost števila članov za anketirana gospodarstva. Take statistične vrste imenujemo f r e k v e n č n e p o r a z d e l i t v e. Ker so za analizo populacij osnovne važnosti, jih bomo podrobneje proučevali v posebnem poglavju. Podobne narave je tudi statistična vrsta o številu zaposlenih po panogah dejavnosti v tabeli 5.4. Tudi ta statistična vrsta pokaže variabilnost znaka: panoga dejavnosti. Vendar so metode proučevanja statističnih vrst, ki kažejo variabilnost, za numerične znake mnogo bolj razvite kakor za nenumerične.

Tudi analiza časovnih vrst je specifična. S časovnimi vrstami proučujemo d i n a m i k o socialno-ekonomskih pojavov. Ker je analiza časovnih vrst specifična, tudi časovne vrste proučujemo v posebnem, za ekonomista važnem poglavju o analizi časovnih

vrst. Dinamika proizvodnje električne energije v SFRJ je nakazana v tabeli 5.2., dinamika števila zaposlenih v družbenem sektorju pa v tabeli 5.3.

Statistična vrsta o poprečnih skupnih izdatkih za prehrano po grupah denarnih dohodkov v tabeli 5.5 prikazuje, kako so poprečni mesečni izdatki za prehrano odvisni od skupnih denarnih dohodkov v kmetijskem gospodarstvu. Ta vrsta osvetljuje važen in splošen problem socialno-ekonomskih pojavov - o d v i s n o s t med pojavi. Tudi o odvisnosti med socialno-ekonomskimi pojavi je govora v posebnem poglavju.

Iz navedenega smo spoznali, da imamo najrazličnejše statistične vrste. Vsaka od njih prikaže določeno značilnost ali lastnost proučevanih pojavov.

STATISTIČNE TABELE

5.8 Statistične vrste v zgledih iz odstavka 5.7 so prikazane v tabelah. Statistične podatke najpogosteje prikazujemo v statističnih tabelah ali razpredelnicah, ker je tabela najpreglednejši in najracionalnejši prikaz podatkov.

Tabela je s sistemom v r s t in s t o l p c e v razdeljena v p o l j a , v katera vpisujemo po določenem sistemu statistične podatke. Dogovorno velja, da imajo vsa polja v isti vrsti isto vsebinsko oznako, ki je naznačena za vsako vrsto posebej v posebnem tekstualnem delu - č e l u tabele. Enako imajo vsa polja v istem stolpcu enako vsebinsko oznako, ki je ustrezno naznačena v tekstualnem delu - g l a v i tabele. Pomen podatka v določenem polju dobimo, če križamo pomen ustrezne vrste s pomenom ustreznega stolpca, v katerem je polje oziroma podatek. Vsi ti pojmi so shematično nakazani v tabeli 5.7. Polje A v tabeli 5.7 je polje za podatek o vzdrževanih ženskah v starosti izpolnjenih treh let.

Tabela 5.7. Tabela 1 za obdelavo vzorca predhodnih rezultatov za popis prebivalstva 31. III.1953 v SFRJ

Delavnost	Delovni		Osebe z osebnimi dohodki		Vzdrževani	
starost \ spol	Moški	Ženske	Moški	Ženske	Moški	Ženske
0						
1						
2		V r s t a			s	
3					+	A
4			polje		o	
5					—	
6					d	
Čelo					e	
					c	

Tabela 5.8 Osnovni podatki o gospodarskih organizacijah družbenega sektorja v SFRJ v letu 1965 po panogah. (Vir SGJ 67)

Panoga	Število organizacij	Število zaposlenih (letno povprečje	Osnovna sredstva 31. XII sed.vred milj. din	Poprečna obratna sredstva milj. din	Neto produkt milj din
	1	2	3	4	5
Skupno	14210	3001272	112221.6	54823.9	70803.4
Industrija in rudarstvo	2463	1370649	48773.5	28996.9	31822.4
Kmetijstvo in ribištvo	2741	331348	11397.7	6993.8	6955.4
Gozdarstvo	148	73642	8820.0	345.1	1113.6
Gradbeništvo	706	290929	2140.0	1680.6	5393.1
Promet in zveze	397	314228	30477.7	1788.5	6248.7
Trgovina, gostinstvo in turizem	3635	349499	6439.8	13272.2	14421.0
Obrt	2729	165055	1056.9	1243.6	2942.1
Komunalna dej.	249	26557	2353.3	298.6	652.2
Storitvene obrti	364	12796	43.8	21.7	185.6
Drugo	800	45099	719.0	182.8	1069.2

5.9 Tabela je e n o s t a v n a , če v njej prikažemo eno samo statistično vrsto. Enostavne tabele so torej vse tabele od 5.1 do 5.6 v prejšnjem odstavku, ker je v njih prikazana ena sama statistična vrsta. Vse enostavne tabele imajo bodisi en sam stolpec (tabele 5.1, 5.4, 5.5 in 5.6), ali eno samo vrsto (tabeli 5.2, 5.3).

5.10 Dostikrat imamo dve ali več različnih statističnih vrst, ki imajo za osnovo isti znak in so med seboj v vsebinski zvezi. Take statistične vrste združimo v eno samo statistično tabelo - s e s t a v l j e n o t a b e l o . V sestavljeni tabeli je združenih več enostavnih tabel. Sestavljena tabela je tabela 5.8, ki prikazuje število organizacij, poprečno število zaposlenega osebja, osnovna in obratna sredstva ter neto produkt v letu 1965 v SFRJ po panogah.

V tabeli 5.8 prikazani podatki so združeni iz petih enostavnih tabel.

5.11 Če razčlenimo populacijo samo po enem znaku, dobimo za rezultat enostavne statistične vrste, ki jih prikazujemo v enostavnih statističnih tabelah. Dostikrat pa populacijo razdelimo po dveh, treh ali celo več znakih hkrati. Ta obdelava da podroben vpogled v odvisnosti pri množičnih pojavih in je posebno pomembna, če proučujemo zvezo med faktorialnimi in rezultativnimi znaki. Rezultate take večkratne istočasne obdelave prikazujemo v k o m b i n a c i j s k i h t a b e l a h . V kombinacijski tabeli križamo po dva ali več znakov. Kombinacijska tabela je npr. razdelitev privatnih kmetijskih gospodarstvih v letu 1956 v tabeli 5.9. V njej so razčlenjena kmetijska gospodarstva po velikostnih skupinah in dohodkih od kmetijskega gospodarstva za 2236 anketiranih gospodarstev.

Tabela 5.9. Število privatnih kmetijskih gospodarstev v kmetijski anketi v SFRJ po velikosti gospodarstev in denarnih dohodkih od kmetijstva v letu 1956 (Vir SB 113).

Skupna površina v ha	Denarni dohodki od kmetijstva v tisočih din							Skupno
	-100	101-200	201-300	301-400	401-500	501-600	601-	
do 4 ha	942	203	46	19	7	5	2	1224
nad 4 - 8 ha	337	172	91	35	14	2	4	655
nad 8 - 12 ha	72	72	40	30	9	4	7	234
nad 12 - 16 ha	21	20	14	7	2	2	1	67
nad 16 - 20 ha	6	5	7	3	1	5	2	29
nad 20 ha	2	7	8	2	2	2	4	27
S k u p a j	1380	479	206	96	35	20	20	2236

V kombinacijski tabeli 5.9 so razen običajnih stolpcev in vrst še seštevki stolpcev in vrst. To sta zbirna vrsta in zbirni stolpec. Zbirno vrsto in zbirni stolpec postavimo na konec tabele, kakor je to v našem zgledu, včasih pa na začetek tabele.

Vrste tabel po vlogi v statističnem proučevanju

5.12 Glede na vlogo, ki jo imajo v statističnem proučevanju, delimo tabele na: obdelovalne, publikacijske in analitične.

Osnovni namen obdelovalnih tabel je sistematično zapisovanje rezultatov osnovne obdelave statističnega gradiva, medtem ko je osnovni namen publikacijskih tabel sistematično prikazovanje končnih rezultatov. Z analitičnimi tabelami pa prikažemo podatke tako, da je iz njih neposredno možna analiza statističnih podatkov.

5.13 Oblika obdelovalnih tabel je prilagojena obdelavi podatkov. Zato se obdelovalne tabele pri ročni obdelavi razlikujejo od obdelovalnih tabel pri elektronski obdelavi. Oblika in velikost obdelovalnih tabel je drugotnega pomena in pri sestavljanju dajemo prednost momentom obdelave. Zato obdelovalne tabele pri nekaterih statističnih obdelavah združujejo kombinacije razmeroma velikega števila znakov. Tako vsebujejo obdelovalne tabele pri popisih prebivalstva kombinacije tudi po šest in več znakov hkrati in so sestavljene iz več delnih tabel.

5.14 V publikacijskih tabelah je še poudarek na tem, da damo strnjeno čim več statističnih podatkov. Vendar je pri njih odločilen vsebinski moment. V publikacijskih tabelah so podatki prikazani tako, da so podatki, ki so med seboj v zvezi, čim bolj primerljivi. Publikacijske tabele so v glavnem vse tabele, ki so objavljene v statističnih publikacijah statističnih organov, kot so: letopisi, bilteni, rezultati popisov itd. Čeprav v publikacijskih tabelah že močno čutimo, da so sestavljene tako, da moremo prikazane podatke čim laže proučiti, je njihov osnovni namen še vedno dati običajne absolutne podatke o populaciji, ki jo prikazujejo, pojava pa ne analizirajo. Zato osnovne statistične publikacije redko vsebujejo več kakor absolutne podatke o statistični populaciji.

5.15 Namen analitičnih tabel pa je prikazati statistične podatke tako, da je iz njih možno neposredno analizirati določeno značilnost pojava. Zato mora biti v analitični tabeli primerljivost podatkov čim boljša, osnovne značilnosti pojava pa čim bolj vidne. Ker je razmeroma težko, včasih pa celo nemogoče, istočasno primerjati veliko podatkov hkrati, so analitične tabele po obsegu manjše in vsebujejo samo podatke, ki so v neposredni zvezi s problemom, ki ga hočemo z analitično tabelo prikazati in analizirati. Zato v analitičnih tabelah običajno kombiniramo po dva, največ tri znake hkrati. Kombinacija več znakov analizo podatkov znatno zamoti. V analitičnih tabelah ne prikazujemo samo absolutnih števil, kakor v obdelovalnih tabelah in v veliki večini publikacijskih tabel. Vanje vključujemo vse vrste statističnih parametrov: relativna števila, srednje vrednosti, mere variacije itd. Analitične tabele so navadno sestavni del pismene analize proučevanega pojava.

Tehnična načela za sestavljanje tabel.

5.16 Če hočemo, da je tabela razumljiva in pregledna, mora biti sestavljena po nekih enotnih tehničnih načelih. Tabela naj ima samostojen naslov, iz katerega je v kratkem razvidno, kaj prikazuje, območje, ki ga obsega, čas, na katerega se podatki v tabeli nanašajo in kako so podatki v tabeli razdeljeni. Tako je npr. iz naslova tabele 5.1 takoj vidno, kaj prikazuje: narodni dohodek; območje: SFRJ; čas: leto 1965; razdeljen: po republikah.

Pojasnila k tabeli dajemo včasih v podnaslovu ali, če so daljša, pod tabelo; to naznačimo v naslovu z ustreznim znakom (zvezdico, križcem ali črko).

V nadaljevanju naslova navedemo vir, iz katerega podatki izvirajo. To je potrebno, ker iz vira sklepamo na kakovost podatkov, po potrebi pa moremo iz osnovnega vira dobiti podrobne dodatne podatke.

V tabeli je treba naznačiti enoto mere, ki jo vnesemo nad tabelo ali v podnaslov, če je enota mere enotna za vse podatke v tabeli (npr. tabela 5.9). Če so enote mere za posamezne podatke različne (npr. tabela 5.8), jih vnesemo v ustrezna polja v glavi oziroma čelu tabele. Opredelitve v opisnem delu tabele - glava in

čelo tabele - morajo biti kratko in jasno označene. Če zaradi pomanjkanja prostora to v sami tabeli ni mogoče, damo dodatna pojasnila o posameznih pojmih v pripombah pod tabelo.

V analitičnih tabelah se izogibamo podatkom z velikim številom mest, zato podatke ustrezno zaokrožujemo. V glavnem so pri večmestnih podatkih vsebinsko pomembna tri mesta. Vendar zaokrožujemo podatke na naravna ali dogovorna merila, (npr. vrednosti na tisoče, milijone ali milijarde dinarjev, ne pa na desettisoče, deset milijone itd.).

Če je število stolpcev ali vrst veliko, jih oštevilčimo z zaporednimi številkami.

5.17 Grupe za posamezne znake razvrstimo po različnih načelih. Časovne grupe in razrede za numerične znake razporedimo po naravnem vrstnem redu po velikosti v čelu tabele od zgoraj navzdol, v glavi tabele pa od leve na desno (glej tabele 5.2, 5.3, 5.5 in 5.6).

Za druge vrste znakov vrstni red ni vnaprej določen. Zanje uporabljamo različne načine. Če je število grup znaka veliko in ni nobenega merila, po katerem bi združevali grupe nižjega reda v grupe višjega reda, uredimo vrednosti po abecednem redu.

Včasih je vrstni red ustaljen po kakem drugem sodilu. Tako je npr. ustaljen vrstni red za panoge dejavnosti v tabeli 5.4.

Če prikazujemo eno samo statistično vrsto, je včasih ugodno, da navedemo grupe po velikosti podatkov v statistični vrsti. Če imamo sestavljeno tabelo, pa včasih uredimo člene po velikosti podatka, ki je v tabeli vodilen.

Za nekatere atributivne znake moremo grupe urediti po enoličnem vrstnem redu po velikosti, čeprav znak ni numeričen.

Tak znak je npr. šolska izobrazba, za katero razvrstimo grupe po stopnji od grupe: brez šolske izobrazbe do končne, najvišje stopnje: dokončana visoka šola. Enako stopnje kvalifikacije navajamo po vrstnem redu: visoko kvalificiran, kvalificiran, priučen in nekvalificiran ali v obrnjenem vrstnem redu.

5.18 Polja v tabeli so razmeroma majhna. Zato p o j a s n i l a k posameznim podatkom ali poljem ne moremo dati v sami tabeli. Da skrajšamo in poenostavimo pojasnila k posameznim okencem, uporabljamo d o g o v o r j e n e k r a t i c e oziroma oznake, ki enostavno opisujejo določene značilnosti v tabeli. V uradnih statističnih publikacijah statističnih zavodov v SFRJ so ustaljeni tile dogovorjeni znaki.

—	pojava ni,
...	ne razpolagamo s podatkom,
0	podatek je manjši kakor je 0,5 od enote mere,
0,0	podatek je manjši kakor je 0,05 od enote mere.
Ø	poprečje
1)	oznaka za opombo pod tabelo
( )	nepopoln oziroma nezadostno preverjen podatek
*	popravljen podatek
↑→	obseženo v podatke v smeri puščice.

5.19 V tabeli ne včrtavamo preveliko š t e v i l o č r t . Običajno vse osnovne črte izpuščamo in včrtavamo v tabelo le črte, ki delijo grupe podatkov. V obširnejših tabelah zvečamo preglednost z različno širokimi presledki med posameznimi vrstami oziroma stolpci. Če je število istovrstnih grup veliko, če so podatki prikazani npr. po občinah, vrstah bolezni ali poklicih, je primerno, da po tri ali pet vrst ločimo s širšimi presledki. To je veliko pregledneje, kakor pa če prenatrpamo tabelo s črtami. S črtami običajno delimo glavo in čelo tabele od številčnega dela - telesa tabele in posamezne pojme v glavi tabele, v številčnem delu pa se omejimo samo na najnujnejše črte.

GRAFIČNO PRIKAZOVANJE

5.20 V uvodu v prikazovanje statističnih podatkov smo nakazali, da statistične podatke prikazujemo razen s tabelami tudi z g r a f i k o n i . Na kratko smo navedli tudi prednosti in pomanjkljivosti enega in drugega načina.

Grafično prikazani podatki dajo enostavno in nazorno sliko o statističnih vrstah. Čeprav

more biti tabela teoretično poljubno velika, moremo v njej istočasno primerjati zelo omejeno število podatkov. Število istočasno primerjanih podatkov je v grafičnem prikazu znatno večje. Grafikon ne prikaže natančnih podatkov. Vendar to ni bistveno, ker je njegov namen, da dá grobo sliko in vtis o pojavu. Zato običajno grafikon dopolnjuje tabela, če niso podatki številčno vpisani že v grafikon.

Grafikoni s svojo nazornostjo niso važni le za približanje statističnih podatkov in sklepov nestrokovnjaku, temveč so koristno in uporabno sredstvo za analizo tudi za strokovnjaka. Z grafikonom laže kakor iz tabele odkrijemo določene težnje in značilnosti v množičnih pojavih. Razen tega dosti statističnih metod vsaj v zasnovi analize sloni na grafičnem reševanju problemov.

Zato moremo grafikone deliti po svoji vlogi v dve vrsti: enostavne grafikone; z njimi ponazarjamo osnovne statistične podatke tako, da so razumljivi čim večjemu številu porabnikov; analitične grafikone, ki so analitično sredstvo raziskovalca.

Ker je večina analitičnih grafikonov zvezana z vsebino pojava in specifičnimi metodami analize, ki jih za sedaj še ne poznamo, so v tem odstavku nakazane splošne osnove grafičnega prikazovanja in obdelani le najenostavnejši grafikoni. Posebne grafikone pa obravnavamo pri posameznih metodah analize posebej. Tako pri poglavju o relativnih številih obravnavamo specifične načine za grafično prikazovanje sestave populacij in statističnih koeficientov, pri poglavju o frekvenčnih porazdelitvah grafično prikazovanje frekvenčnih porazdelitev, pri korelaciji grafične metode primerne za analizo odvisnosti med pojavi, pri časovnih vrstah pa metode prikazovanja, ki so specifične za analizo časovnih vrst.

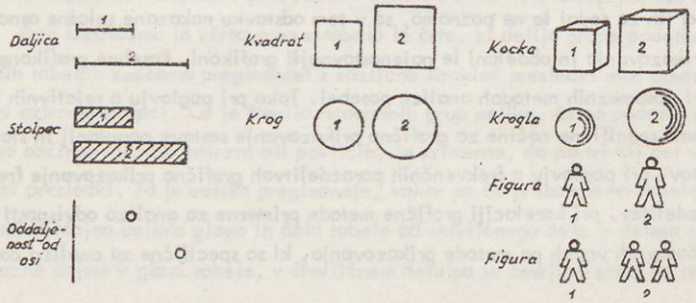
Prvine grafičnega prikazovanja

5.21 Vsaka statistična vrsta je sestavljena iz niza številčnih podatkov, od katerih se vsak nanaša na posamezno vrednost ali grupo vrednosti za osnovni znak v statistični vrsti. Ker so statistični podatki dani številčno, uporabljamo za njegov prikaz najrazličnejše geometrijske elemente, ki imajo svoje mere. Ustrezna povezava teh elementov večinoma dá nazorno sliko pojava. Za prikazovanje statističnih podatkov med drugim

uporabljamo tele geometrijske elemente:

a) daljico, b) pravokotnike s stalno širino, ki jih imenujemo stolpce c) oddaljenost točke od dane osi, d) kvadrate, e) kroge, f) kocke g) krogle itd.

Vsak izmed teh elementov ima določeno mero: za daljico in stolpec je značilna dolžina. Enako z dolžinsko mero merimo oddaljenost točke od osi. Kvadrat in kocka sta določena s stranico, krog in krogla s polmerom. Lastnost, da moremo te elemente meriti, izkoriščamo za prikazovanje statističnih podatkov. Če z daljico določene dolžine ponazorimo dani podatek, narišemo za dvakrat večji podatek dvakrat daljšo daljico ali stolpec. Enako s točko, ki je dvakrat bolj oddaljena od osi kot prva, prikazujemo dvakrat večji podatek kot s prvo točko. Če ponazorimo statistične podatke s kvadrati ali krogi, je dvakrat večji podatek ponazorjen s kvadratom ali krogom, ki ima dvakrat večjo ploščino. Če ga ponazorimo s kocko ali kroglo, pa je volumen kocke ali krogle dvakrat večji.



Slika 5.1 Prvine grafičnega prikazovanja

Če primerjamo daljico in stolpec s kvadratom in krogom ali kocko in kroglo, spoznamo, da je najlažja primerjava različno velikih količin pri daljicah in stolpcih: elementi pa so tem slabše primerljivi, čim višja je razsežnost geometrijskega elementa. To je lepo razvidno iz slike 5.1, v kateri so narisani dvakrat večje posamezne zgoraj naštete prvine.

Razen gornjih elementov, od katerih je vsak določen samo z enim podatkom, uporabljamo za prikazovanje tudi druge elemente, ki imajo več različnih razsežnosti. Pravokot-

nik, ki ima svojo višini in dolžino, uporabljamo za prikazovanje dveh statističnih podatkov hkrati; zaradi zveze med širino, dolžino in ploščino pa celo treh. Enako je krogov izsek določen z dvema podatkoma: radijem in kotom, ki ga izsek oklepa itd. Te elemente uporabljamo v kompleksnejših primerih grafičnega prikazovanja.

V grafikonih, ki so namenjeni popularizaciji podatkov, dostikrat uporabljamo za element grafičnega prikaza idealizirane figure, ki so vezane z vsebino pojava. Podatke o prebivalstvu prikažemo npr. z idealizirano figuro moškega ali ženske (glej sliko 5.1) železniški promet z idealizirano lokomotivo, promet po morju z idealiziranim prikazom parnika itd. Tudi pri figurah uporabljamo dve metodi: Po prvi je volumen prikazane figure sorazmeren s podatkom, ki ga prikazujemo, po drugem pa velikost podatka ponazorimo z ustreznim številom enako velikih figur.

Skale- lestvice.

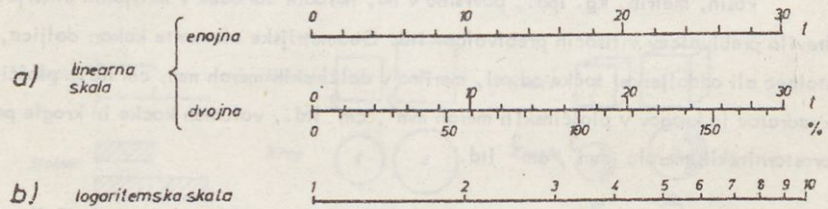
5.22 Statistični podatki imajo različne enote mere. Vrednost proizvodnje merimo v kosih, metrih, kg. ipd., površino v ha, narodni dohodek v milijonih dinarjev, število prebivalcev v tisočih prebivalcev itd. Geometrijske elemente kakor: daljica, stolpec ali oddaljenost točke od osi, merimo v dolžinskih merah mm, cm itd., ploščine kvadratov in krogov v ploščinskih merah mm^2, cm^2 itd., volumen kocke in krogle pa v prostorninskih merah; mm^3, cm^3 itd.

Če hočemo nakazane geometrijske elemente uporabiti za prikazovanje statističnih podatkov, moramo določiti razmerje med enoto mere za statistični podatek in enoto mere za geometrijski element, s katerim prikazujemo podatek. Če vzamemo, da 1 cm pomeni 10 t proizvodnje, je s tem razmerjem dana dolžina daljice za katero koli vrednosti proizvodnje. Najprikladneje pa je, da sestavimo skalo, ki grafično podaja razmerje med dolžino stolpcev ali daljice in statističnimi podatki. Ko se odločimo za določeno razmerje in sestavimo skalo, prvotno geometrijsko merilo sploh odpade in s skalo merimo proučevani pojav neposredno.

5.23 Uporabljamo različne skale ali lestvice. Najobičajnejše so aritmetične- linearne skale. Na aritmetičnih skalah so daljice med dvema vredno-

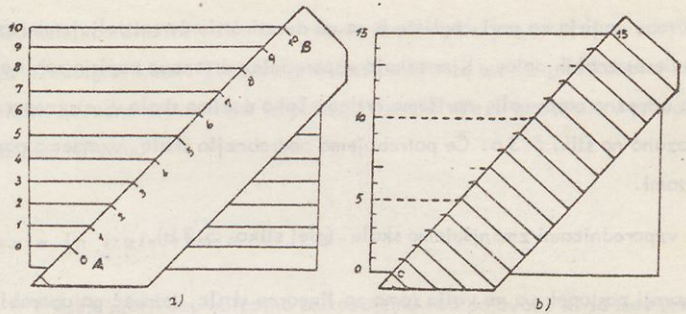
stima v sorazmerju z razliko vrednosti. Aritmetična skala je npr. skala za vrednost proizvodnje v sliki 5.2 in skale v večini grafikonov v tem odstavku (slike 5.7, 5.8, 5.10, 5.12, 5.15).

5.24 Razen aritmetičnih - linearnih skal imamo še druge vrste skal, v katerih odnos med podatkom in geometrijskim elementom ni linearen. Od teh je pri proučevanju dinamike pojavov s časovnimi vrstami posebno uporabna logaritemska skala. Logaritemska skala je skala b na sliki 5.2 Na logaritemski skali so daljice med dvema vrednostima v sorazmerju z logaritmi podatkov, ki jih skala meri. Logaritemska skala je za proučevanje statističnih podatkov priporočljiva, ker z njo ne merimo razlik med podatki, kakor je to primer pri aritmetični skali, temveč razmerja, oziroma kvociente. Razmerja pa so v statistični analizi mnogo važnejša in učinkovitejša za primerjavo kakor razlike. Podrobno o uporabi ligaritemskih skal in o logaritemskih grafikonih bo govora pri časovnih vrstah, ker so logaritemski grafikoni posebno priporočljivi za analizo časovnih vrst.

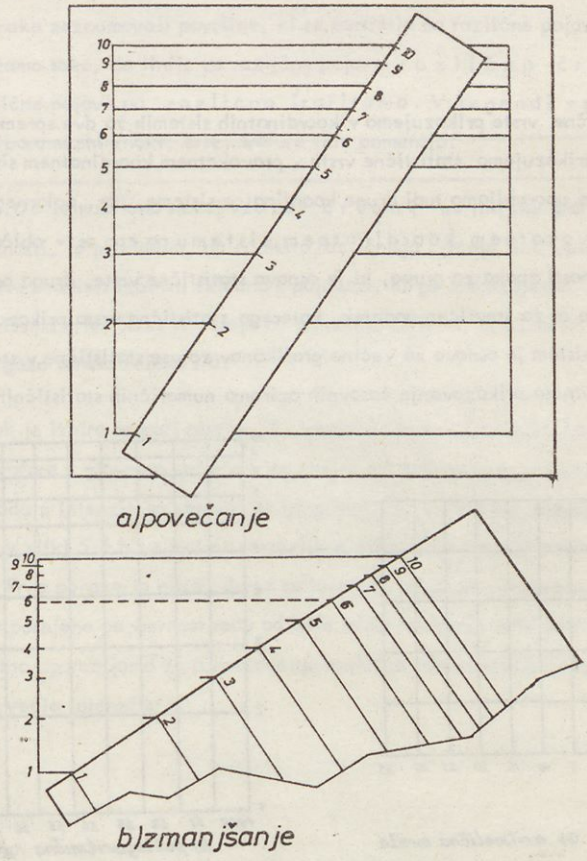


Slika 5.2 Skale ali lestvice

5.25 Risanje skal v ustreznem merilu. Pri risanju grafikonov je vsakdanji problem, kako narisati skalo v merilu, ki je v skladu z velikostjo grafikona. Pri tem uporabljamo enostavno metodo, ki pomaga risati skale v poljubnem merilu. Vzemimo, da moramo narisati tako linearno skalo na daljici 7,8 cm, da bo celotna daljica merila proizvodnjo 10 t. Daljico 7,8 cm moramo torej razdeliti na deset ali pri natančnejši skali na 20 enakih delov. To skalo dobimo takole: na robu papirnatega traku odmerimo predpisano dolžino skale 7.8 cm. Eno oglišče zaznamujemo z A, drugo pa z B. Rob traku primaknemo k listu črtanega ali milimeterskega papirja tako, da je oglišče A



Slika 5.3 .Risanje linearnih skal v poljubnem merilu



Slika 5.4. Povečevanje in zmanjševanje logaritemskih skal.

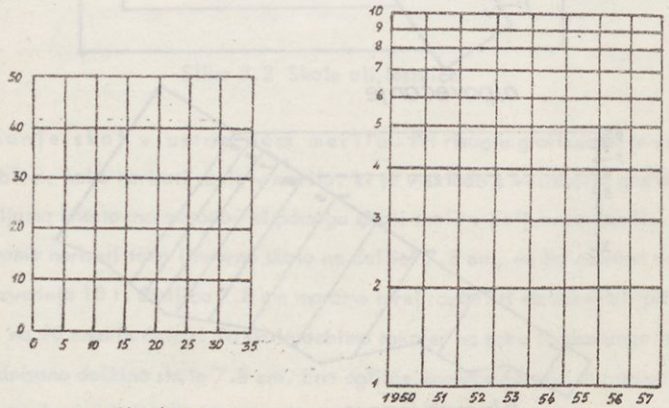
daljice na robu papirja na prvi, oglišče B pa na deseti črti, če razdeljujemo razdaljo 7,8 cm na deset enakih delov. Kjer sekajo vzporednice črtanega papirja rob traku, katerega smo ustrezno naravnali, narišemo črtice. Tako dobimo skalo v ustreznem merilu. To je prikazano na sliki 5.3 a. Če potrebujemo podrobnejšo skalo, vzamemo papir z gostejšimi črtami.

Podobno z vzporednicami zmanjšujemo skale (glej sliko 5.3 b).

5.26 Zgornji postopek pa ne velja samo za linearne skale, temveč ga uporabljamo tudi za druge vrste skal. Na sliki 5.4 je nakazano kako poljubno povečujemo ali zmanjšujemo logaritemske skale.

Mreže

5.27 Statistične vrste prikazujemo v koordinatnih sistemih za dve spremenljivki. Običajno prikazujemo statistične vrste v pravokotnem koordinatnem sistemu. Le za posebne prikaze uporabljamo tudi druge koordinatne sisteme, npr. polarnega, trikotniškega itd. V pravokotnem koordinatnem sistemu na eno os - običajno absciso - nanašamo vrednosti znaka za grupo, ki je osnova statistične vrste, druga os - običajno ordinata - pa je os za številčen podatek, katerega statistična vrsta prikazuje. Pravokotni koordinatni sistem je osnova za večino grafikonov za vse statistične vrste. Vendar je najbolj upravičen za prikazovanje časovnih oziroma numeričnih statističnih vrst. Če v



a) aritmetična mreža b) polilogaritmična mreža

Slika 5.5. Mreži grafikona v pravokotnem koordinatnem sistemu.

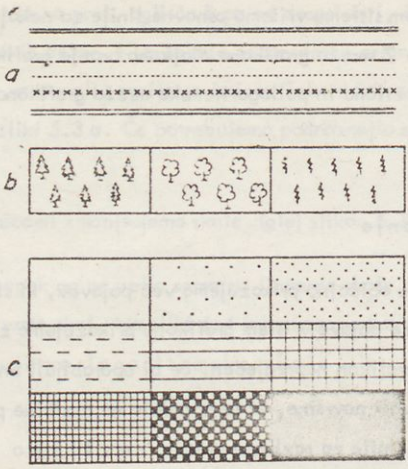
pravokotnem koordinatnem sistemu vrišemo osnovne linije za nekaj važnejših vrednosti, dobimo mrežo grafikona. Z mrežo grafikona olajšamo branje grafikona. Na sliki 5.5 sta vrisani navadna aritmetična in pollogaritemska mreža grafikona v pravokotnem koordinatnem sistemu.

Črtanje in šrafriranje

5.28 V istem grafikonu običajno prikazujemo več pojavov, ki so med seboj vsebinsko povezani. Ker vse pojave v istem grafikonu prikazujemo z daljicami, stolpci, površinami itd., bi bil grafikon nepregleden, če bi uporabljali enake črte za različne pojave, enako zaznamovali površine, ki se nanašajo na različne pojave itd. Preglednost povečamo tako, da linije za različne pojave r a z l i č n o č r t a m o, površine za različne pojave pa r a z l i č n o š r a f i r a m o. V l e g e n d i - pojasnilu pa označimo, kaj posamezni znaki, črte, šrafure itd. pomenijo.

Na sliki 5.6 a je načrtano nekaj vzorcev č r t a n j za linijske grafikone. Če so tehnične možnosti, je prikladno, da namesto različnega črtanja linij uporabljamo različne barve. Barve moremo izbirati skladno s pojavom, ki ga prikazujemo. Tako v kmetijski statistiki rišemo črte, ki se nanašajo na travnike, zeleno, črte, ki se nanašajo na njive, rjavo, za gozd temno zeleno itd.

Pri šrafurah je izbira še bolj pisana. Tu imamo možnost spreminjati barve in vzorce v šrafuri. Za šrafure v glavnem veljajo trije načini: a) šrafure brez zveze s pojavom; b) šrafure v skladu z intenziteto pojava, ki ga prikazuje; c) šrafure asociativno povezane s pojavom. V sliki 5.6 b so šrafure sestavljene tako, da podajajo pojav, ki ga prikazuje površina. Prva ponazarja gozd, druga sadovnjak, tretja pa vinograd. V sliki 5.6 c pa so šrafure razporejene po vrstnem redu od bele mimo različnih odtenkov do črne. Take šrafure koristno uporabljamo za prikazovanje različne intenzitete pojava in pomeni temnejša šrafura večjo intenziteto.



Slika 5.6. Črte in šrafure

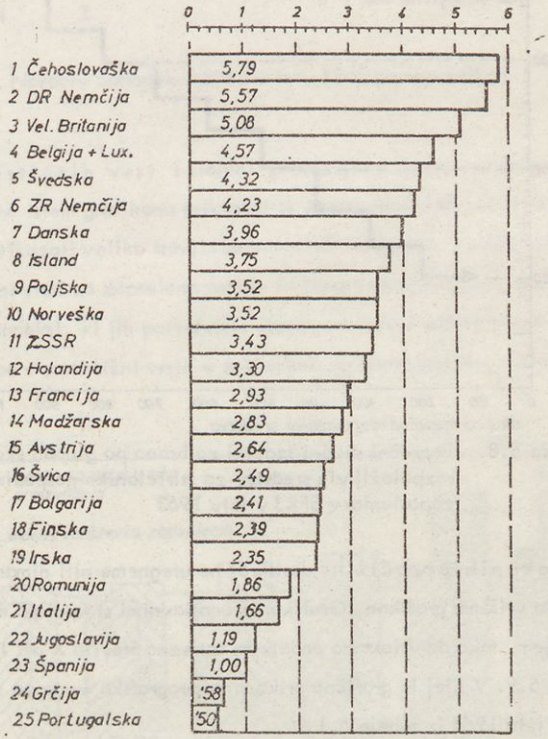
Vrste grafikonov

5.29 Običajni grafikoni so: a) s t o l p c i , b) l i n i j s k i g r a f i k o n i c) g r a f i k o n i s f i g u r a m i , d) k a r t o g r a m i . Vsak izmed teh načinov ima svoje prednosti in pomanjkljivosti, ker ima vsak grafikon posebne značilnosti, ki jih drug nima. Stolpce moremo uporabljati za prikazovanje vseh statističnih vrst, vendar niso v vseh primerih najprikladnejši. Linijski diagrami imajo niz prednosti, ki jih drugi načini grafičnega prikazovanja nimajo. Zato so analitični grafikoni običajno linijski. Figure uporabljamo za popularizacijo pojavov, ki jih z njimi prikazujejo, nimajo pa posebnega študijskega značaja. Regionalno razmestitev pojava pa lepo prikažemo s kartogrami. Pri teh navadno mrežo v koordinatnem sistemu zamenja geografska karta.

S t o l p c i

5.30 Z e n o s t a v n i m i s t o l p c i prikazujemo grafično vse statistične vrste. To je prednost stolpcev. Te lastnosti linijski grafikoni npr. nimajo, čeprav so zaradi drugih lastnosti bolj čislani. S stolpci prikazujemo predvsem tiste statistične vrste, ki jih ne moremo prikazovati z linijskimi grafikoni. To so: geografske vrste in stvarno atributivne vrste.

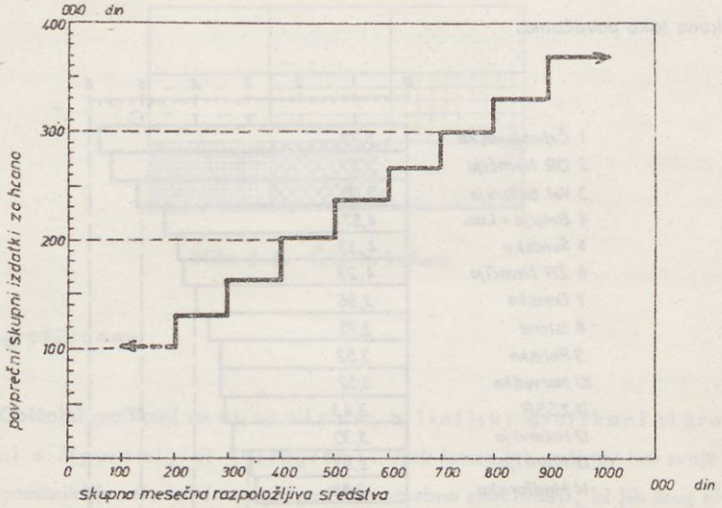
Stolpce v grafikonu večkrat razporedimo po velikosti členov, da podatke laže primerjamo med seboj. V sliki 5.7 je s stolpci prikazana poraba skupne energije v tonah pogojnega premoga na prebivalca v letu 1964 v evropskih državah. V grafikonu so stolpci nanizani po velikosti podatkov v ranžirni vrsti. Stolpci so postavljeni vodoravno, da moremo zlahka odbrati državo, na katero se stolpec nanaša. Običajno v stolpce, če le moremo, vnesemo številčne podatke. Skala je v sliki 5.7 nanešena vodoravno, mreža pa je v grafikon včrtana tako, da se zdi, kot da so stolpci položeni prek mreže. Optični učinek grafikona tako povečamo.



Slika 5.7. Poraba skupne energije na prebivalca za evropske države v tonah pogojnega premoga v letu 1964 (Vir: SGJ-67).

Stolpce uporabljamo včasih tudi za prikazovanje numeričnih vrst, čeprav so linijski grafikoni zanje priporočljivejši. V sliki 5.8 imamo prikazane povprečne denarne izdat-

ke za hrano za gospodinjstvo po grupah skupnih razpoložljivih sredstev za štiričlansko gospodinjstvo z enim zaposlenim v SFRJ v letu 1963 iz tabele 5.5. Posebnost tega grafikona je v tem, da so vrisani samo obrisi stolpcev. Tako povečamo nazornost grafikona. Enako je mreža včrtana podobno kot v gornjem primeru samo do stolpcev. Širina stolpcev je v razmerju s širino razredov. Le tako dobimo pravilen vtis o spremembah med posameznimi razredi denarnih dohodkov. Slika zelo nazorno pokaže odvisnost izdatkov od dohodkov v kmetijskih gospodarstvih.



Slika 5.8. Poprečni skupni izdatki za hrano po grupah skupnih mesečnih razpoložljivih sredstev za štiričlanska gospodinjstva z enim zaposlenim v SFRJ v letu 1963

5.31 V tipkanih poročilih dostikrat ne utegnemo niti nimamo tehničnih možnosti, da vrišemo grafikon. Grafikon z enostavnimi stolpci pa moremo vrisati kar s pisalnim strojem, tako da vtipkamo podatkom ustrezno število X ali I. Tak primer je nakazan v sliki 5.9. V njej je grafično prikazana geografska vrsta za narodni dohodek po republikah v letu 1965 iz tabele 5.1.

Narodni dohodek

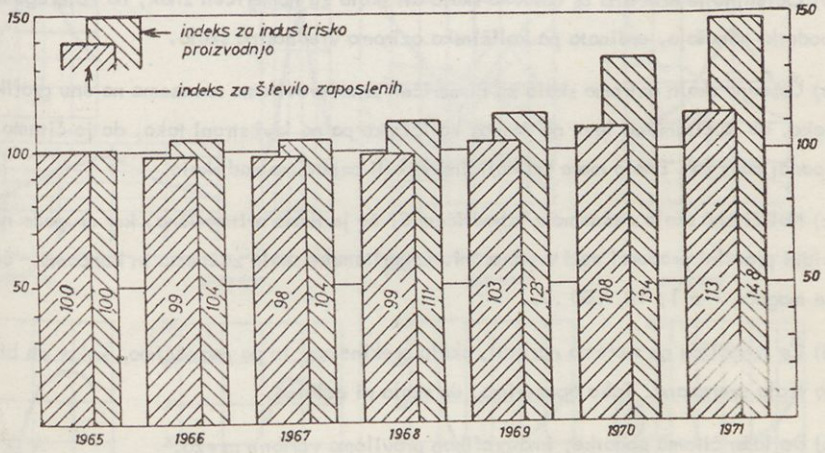
	v mlj din	1 2 3 12345678901234567890123456789 0
Srbija	28.712	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXI
Hrvatska	19.422	XXXXXXXXXXXXXXXXXXI
Slovenija	11.263	XXXXXXXXXXI
BiH	9.022	XXXXXXXXX
Makedonija	3.907	XXXX
Črna gora	1.244	X

X = 1 milijarda din

I = 500 milijonov din

Slika 5.9. Narodni dohodek v SFRJ v letu 1965 po republikah

5.32 Več statističnih vrst s stolpci prikazujemo poredko, ne glede na to, da moremo na istem grafikonu prikazati le istovrstne količine. Razen tega je tudi tehnično težko prikazati veliko število raznovrstnih stolpcev. V sliki 5.10 sta s stolpci prikazani indeksni vrsti za zaposleno osebje in industrijsko proizvodnjo v SFRJ v razdobju 1965 - 71. Indeksi, ki jih podrobneje obravnavamo pri relativnih številih, kažejo razmerja podatkov v statistični vrsti v primerjavi na določen člen - osnovo primerja-



Slika 5.10. Indeksi za število zaposlenih in za skupno industrijsko proizvodnjo v SFRJ v razdobju 1965-1971 (1965=100) (Vir: SG - 71).

ve. V naši sliki je osnova primerjave leto 1965. Iz stolpcev nazorno vidimo, da proizvodnost dela po letih raste, ker se stolpci za indekse proizvodnje hitreje večajo kot stolpci za indekse zaposlenih.

Stolpci so v sliki 5.10 risani tako, kakor da bi bili prekriti. To poveča učinek grafikona. Z različnima šrafurama poudarimo, da grafikon prikazuje dve različni časovni vrsti.

Linijski grafikoni

5.33 Največjo analitično vrednost imajo linijski grafikoni vseh vrst. Z njimi prikazujemo stvarno numerične, predvsem pa časovne vrste. Ker posebne oblike linijskih grafikonov časovnih vrst obravnavamo v poglavju o časovnih vrstah, se v tem odstavku omejimo samo na splošne oblike in pravila za risanje linijskih grafikonov.

Bistvo linijskih grafikonov je v tem, da podatke vnašamo s točkami, ki so od abscisne osi oddaljene v razmerju z velikostjo podatkov, te točke pa zvežemo z daljicami.

Za risanje linijskih grafikonov veljajo ustaljena pravila. Z njimi dosežemo, da je grafikon pregleden in nazoren in da nedvoumno prikaže proučeni pojav. Ta p r a v i l a moremo v kratkem združiti v naslednje točke:

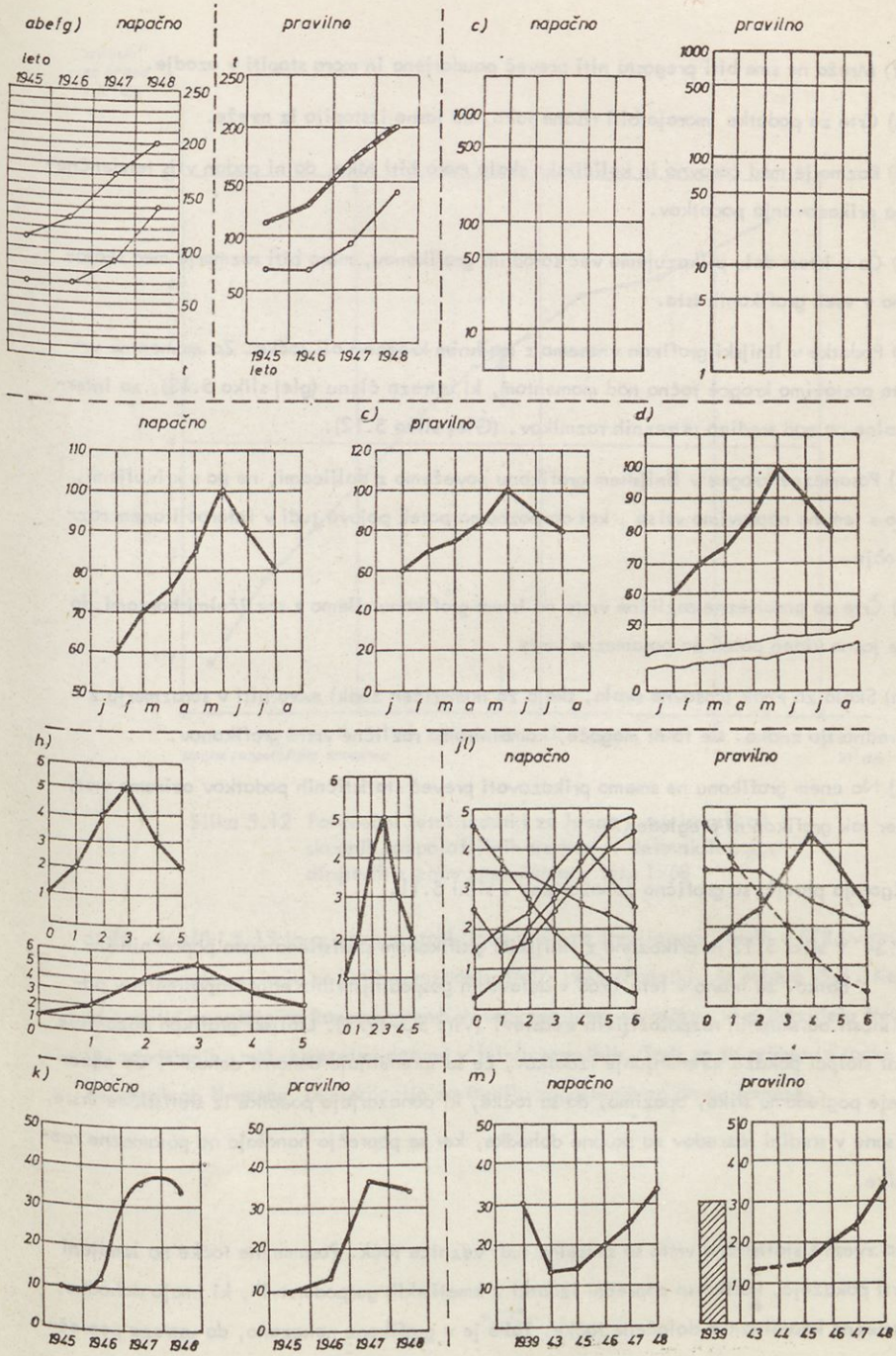
a) Običajno je abscisna os časovna skala ali skala za numeričen znak, na katerega se podatki nanašajo, ordinata pa količinska oziroma vrednostna skala.

b) Časovno skalo oziroma skalo za numeričen znak praviloma nanesemo na dnu grafikona tako, da jo čitamo od leve na desno, količinsko pa na levi strani tako, da jo čitamo od spodaj navzgor. Enoto mere pri količinski skali označimo nad skalo.

c) Količinsko skalo začnemo z izhodiščem 0, če je skala aritmetična, ker drugače niso vidni pravilni proporci med vrednostmi. Logaritemske skale začnemo in končamo - če je le mogoče - z 1, 10, 100 ....

d) Če izhodišča ne moremo narisati, skalo prekinemo; to pa naznačimo, kakor da bi bila skala prerezana. Tako opozorimo, da skala ni celotna.

e) Da laže čitamo podatke, ima grafikon praviloma včrtano mrežo.



Slika 5.11 Pravila za risanje linijskih grafikonov

f) Mreža ne sme biti pregosta niti preveč poudarjena in mora stopiti v ozadje.

g) Črte za podatke morajo biti risane tako, da jasno izstopijo iz mreže.

h) Razmerje med časovno in količinsko skalo mora biti tako, da ni podan vtis tendenčnega prikazovanja podatkov.

i) Če v istem delu prikazujemo več sorodnih grafikonov, mora biti razmerje med skalama v vseh grafikonih isto.

j) Podatke v linijski grafikon vnesemo z majhnim krogcem ali točko. Za momentne vrste postavimo krogce točno nad momentom, ki ustreza členu (glej sliko 5.13), za intervalne pa nad sredino ustreznih razmikov. (Glej sliko 5.12).

k) Posamezne krogce v linijskem grafikonu povežemo z daljicami, ne pa s krivuljami, da s tem ne napravimo vtisa , kot da poznamo potek pojava tudi v interpoliranem razdobju.

l) Črte za posamezne različne vrste na istem grafikonu rišemo z različnimi barvami,da je jasno viden potek za posamezne vrste.

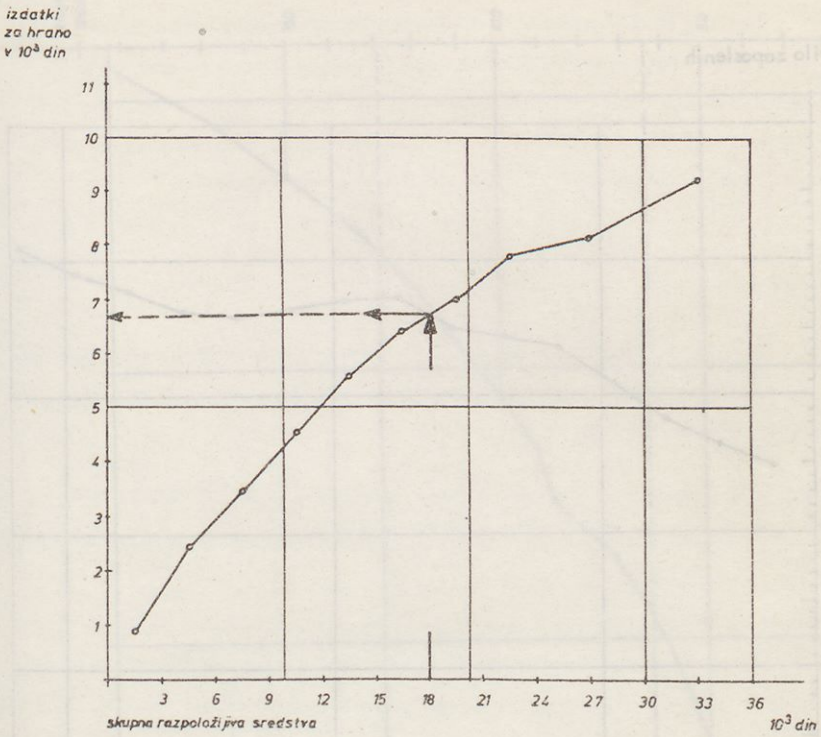
m) Skala za znak (časovna skala, skala za numeričen znak) mora biti v sorazmerju z vrednostjo znaka. Če to ni mogoče, kombiniramo različne vrste grafikonov.

n) Na enem grafikonu ne smemo prikazovati preveč statističnih podatkov oziroma vrst, ker tak grafikon ni pregleden.

Zgornja pravila so grafično ponazorjena v sliki 5.11.

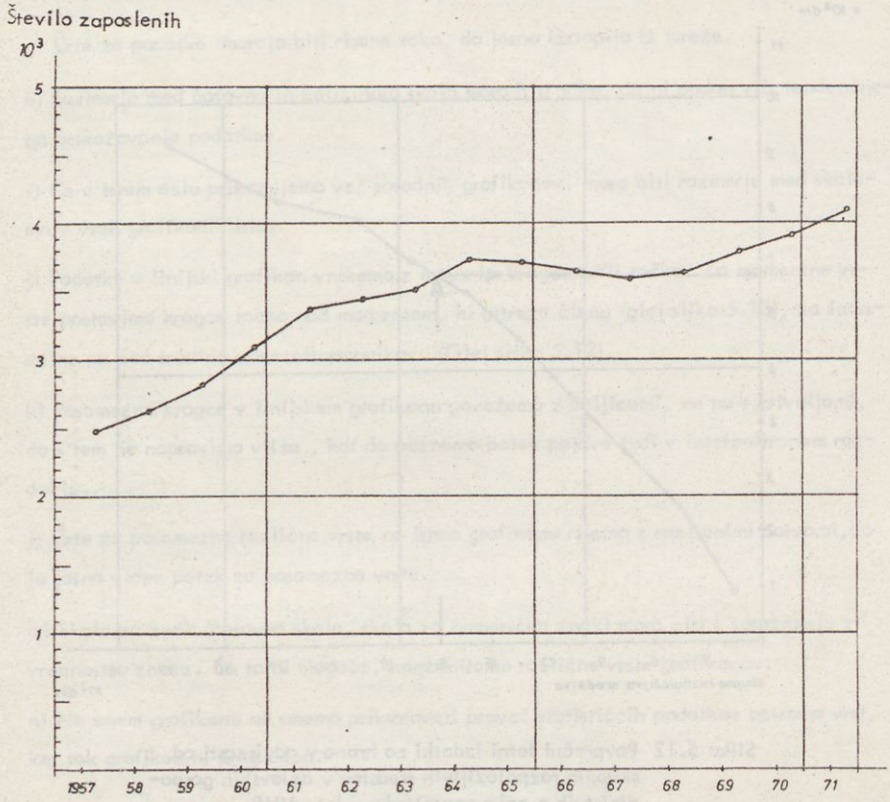
5.34 V sliki 5.12 je prikazana z linijskim grafikonom statistična vrsta poprečnih izdatkov za hrano v letu 1968 v delavskih gospodinjstvih z enim zaposlenim v odvisnosti od skupnih razpoložljivih sredstev. (Vir: SG-1971). Linijski grafikon nazorneje kot stolpci pokaže spreminjanje izdatkov, če se spreminjajo denarni dohodki. Če pozorneje pogledamo sliko, opazimo, da so točke, ki ponazarjajo podatke iz statistične vrste, risane v sredini razredov za skupne dohodke, ker se poprečja nanašajo na posamezne razmike.

Za zvezne statistične vrste so smiselne tudi veznice točk. Posamezne točke na lomljeni črti pokažejo, koliko so poprečni izdatki v kmetijskih gospodarstvih, ki imajo dohodke, ustrezne koordinatam določene točke. Tako je v grafikonu nakazano, da ustreza poprečnim dohodkom 18000 dinarjev 6700 dinarjev denarnih izdatkov za hrano.



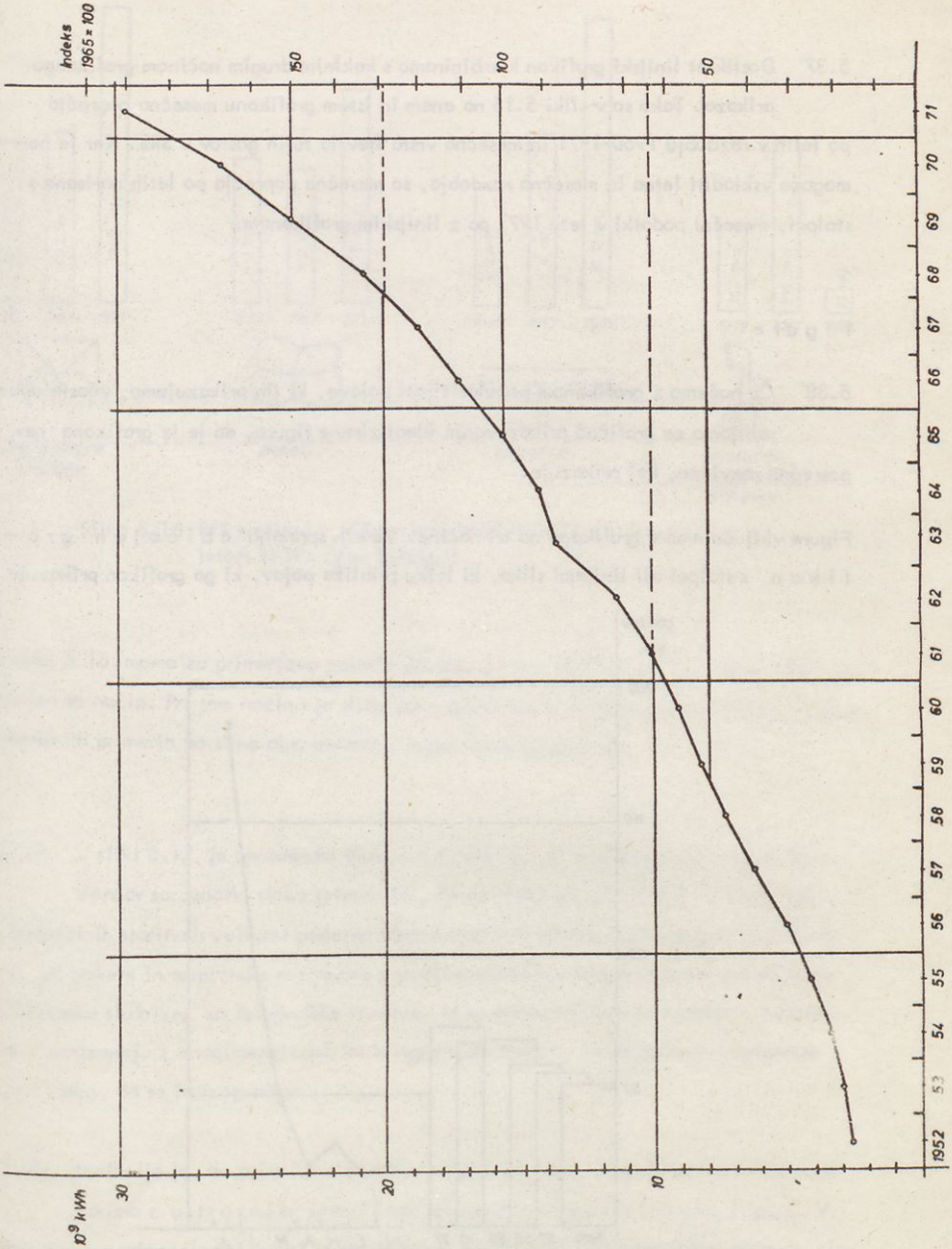
Slika 5.12 Povprečni letni izdatki za hrano v odvisnosti od skupnih razpoložljivih sredstev v delavskih gospodinjstvih z enim zaposlenim v letu 1968

5.25 V sliki 5.13 je prikazana statistična vrsta za število zaposlenih v SFRJ v družbenem sektorju po letih v razdobju 1957-1971. (Podatki iz tabele 5.3.) Ker se podatki nanašajo na konec septembra vsakega leta, so točke, ki prikazujejo število zaposlenih, nad ustreznimi datumi v letnih razmikih. Tudi za ta primer so točke na veznicah linearne interpolacije za število zaposlenih za druge datume.



Slika 5.13. Število zaposlenih v družbenem sektorju v SFRJ v razdobju 1957-1971 (Vir: Tabela 5.3)

5.36 Slika 5.14, ki prikazuje proizvodnjo električne energije, je zanimiva, ker ima dvojno skalo. Na levi je skala za količino električne energije, na desno pa je skala indeksov s stalno osnovo leto 1965. Da se skali ne mešata, sta narisani le do krivulje, vsaka s svoje strani. Ta grafikon prikazuje intervalno časovno vrsto. Posamezni vrisani podatki se torej nanašajo na posamezna leta. Vendar imajo tudi točke na lomljeni črti grafikona svoj smisel. Posamezne točke na liniji pomenijo linearno interpolacijo za proizvodnjo električne energije v letu, za katerega je čas, ki ustreza tej točki, sredi leta.



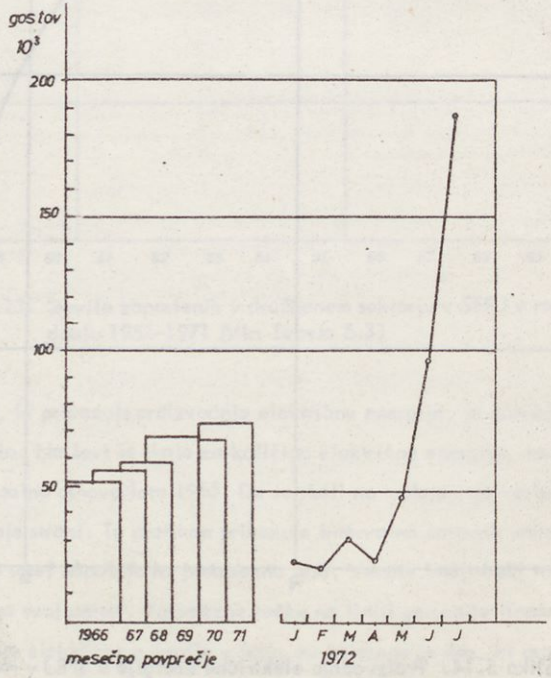
Slika 5.14. Proizvodnja električne energije v SFRJ v razdobju 1952-1971. (Vir: Tabela 5.2)

5.37 Dostikrat linijski grafikon kombiniramo s kakšnim drugim načinom grafičnega prikaza. Tako so v sliki 5.15 na enem in istem grafikonu mesečna poprečja po letih v razdobju 1960-1971 in mesečna vrsta števila tujih gostov v SRS. Ker je nemogoče vskladiti letna in mesečna razdobja, so mesečna poprečja po letih narisana s stolpci, mesečni podatki v letu 1971 pa z linijskim grafikonom.

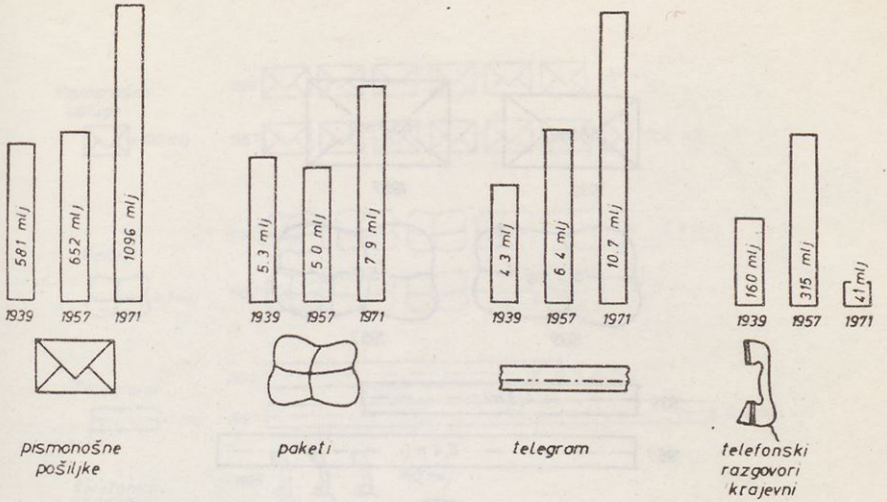
F i g u r e

5.38 Če hočemo z grafikonom popularizirati pojave, ki jih prikazujemo, včasih uporabljamo za grafično prikazovanje idealizirane figure, da je iz grafikona neposredno razvidno, kaj prikazuje.

Figure vključujemo v grafikone na tri načine. Včasih spremlja o b i č a j e n g r a f i k o n s stolpci ali linijami slika, ki laiku približa pojav, ki ga grafikon prikazuje.



Slika 5.15. Število tujih gostov v SRS v razdobju 1966-1972

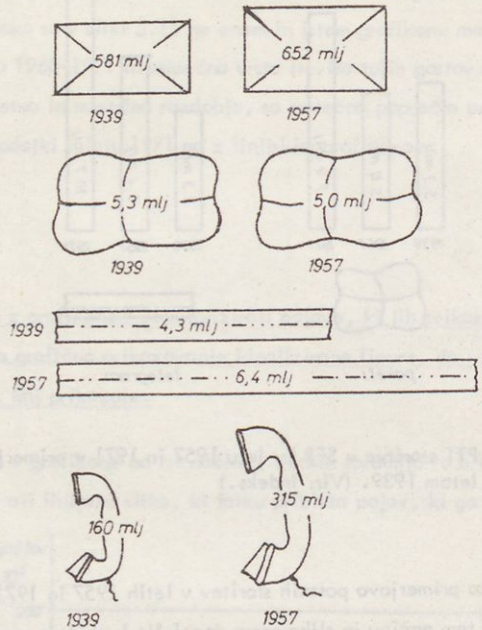


Slika 5.16 PTT storitve v SFRJ v letu 1957 in 1971 v primerjavi z letom 1939. (Vir: Indeks.)

V sliki 5.16 imamo za primerjavo poštnih storitev v letih 1957 in 1971 z letom 1939 uporabljen ta način. Pri tem načinu je slika samo dopolnilo k običajnemu grafikonu. V obeh nadaljnjih primerih pa slika obenem kaže, kako velik je podatek.

5.39 V sliki 5.17 je posamezna slika v razmerju z velikostjo podatka. Vendar so podatki slabo primerljivi, če so risani po tej metodi. Tako je pri pismonošnih storitvah velikost podatka sorazmerna z dvodimenzionalno površino kuverte; za pakete in telefonske razgovore s trodimenzionalno prostornino paketa oziroma telefonske slušalke; za telegrafske storitve, ki so ponazorjene s telegrafskim trakom, pa v sorazmerju z enodimenzionalnim telegrafskim trakom. Ta metoda je v splošnem tako slaba, da se je izogibamo.

5.40 Najbolje je, da prikažemo podatke s figurami tako, da velikost podatka ponazorimo z ustreznim številom enakih, idealiziranih figur. V tem primeru odpadejo vsi dvomi in nevšečnosti zaradi različne razsežnosti pojavov, sla-

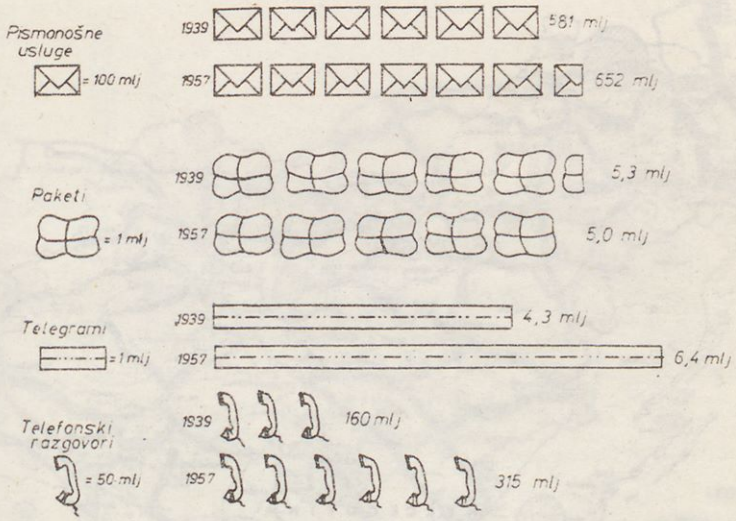


Slika 5.17. PTT storitve v SFRJ v letu 1957 v primerjavi z letom 1939.

be primerljivosti prostornin itd. Po tem načinu so prikazani podatki o PTT storitvah v sliki 5.18. Vsaka slikica danega elementa predstavlja določeno velikost pojava. Tako pri pismonošnih storitvah ena kuverta ponazarja 100 milijonov pismonošnih storitev, en paket 1 milijon paketov, 1 cm telegrafskega traku 1 milijon telegrafskih storitev in ena telefonska slušalka 50 milijonov telefonskih pogovorov. Kakor vidimo iz slike 5.18, moremo z delom osnovne slikice nakazati podatek, ki je manjši od podatka, ki ga prikazuje cela figura.

Kartogrami

5.41 Geografsko razmestitev pojavov najlepše prikažemo s kartogrami. Osnova kartogramov je geografska karta, v katero vrisujemo podatke na mesto, ki ustreza stvarni razmestitvi pojava. Mreža kartograma je torej geografska karta. Vse grafiko-



Slika 5.18. PTT storitve v SFRJ v letu 1957 v primerjavi z letom 1939.

ne, ki kakor koli prikažejo razmestitev določenega pojava v geografski karti, imenujemo kartograme v širšem smislu.

5.42 V diagramski karti rišemo običajno grafikone (stolpce, kroge, linijske diagrame, figure), v tisto območje, na katerega se nanaša. Tako dobimo nazoren pregled o regionalni razmestitvi in primerjavi pojava, ki ga proučujemo. Tega ne dobimo, če grafikone, ki smo jih vnesli v diagramsko karto, narišemo v vrsti stolpcev, krogov itd.

Za primer diagramske karte vzemimo kartogram zaposlenih žena po okrajih v SFRJ. Število zaposlenih žena je prikazano s krogi, za katere je ploščina v sorazmerju s številom žena. Odstotek zaposlenih žena od skupnega števila žena v okraju pa je nakazan s šrafuro kroga. Krogi so narisani v ustreznih okrajih. Iz grafikona vidimo, da je število in odstotek zaposlenih žena veliko predvsem v razvitejših predelih države.



Slika 5.19. Diagramska karta o številu in odstotku zaposlenih žena po okrajih v SFRJ. (Vir: ZZS: Predlog za anketo o zaposlenih ženah 1955)

5.43 Lep primer diagramske karte je tudi kartogram prometne obremenitve cestne mreže v Sloveniji v letu 1952.



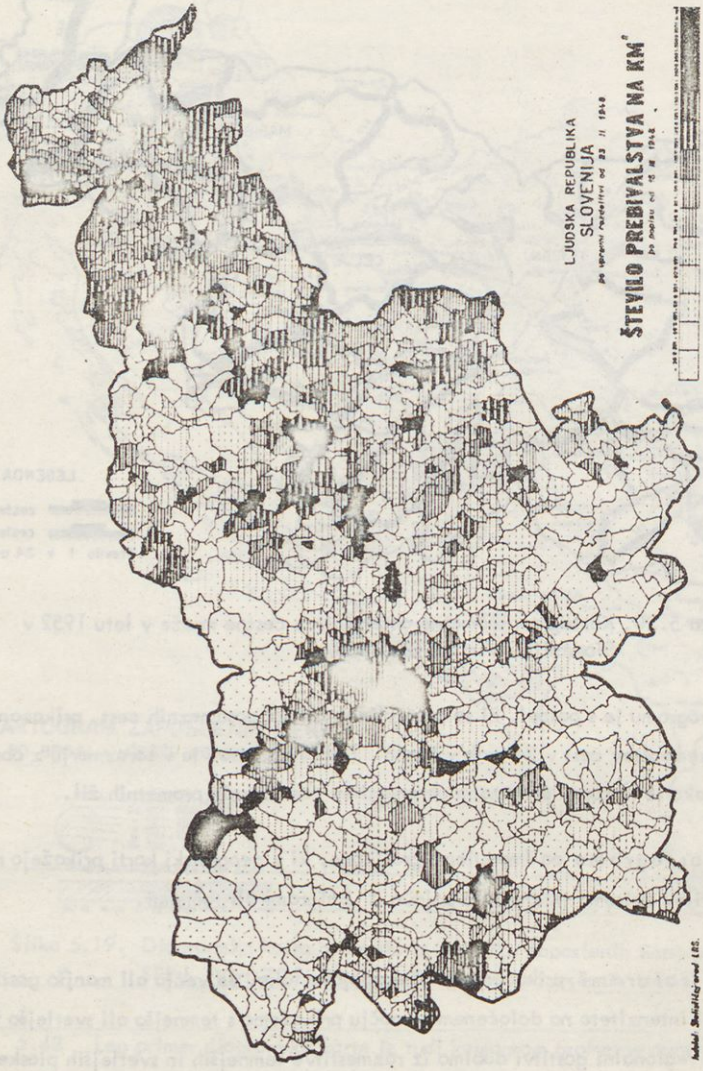
Slika 5.20. Kartogram prometne obremenitve cestne mreže v letu 1952 v Sloveniji. (Vir: SL LRS 1953).

V tem kartogramu je s pasovi, ki so postavljeni vzdolž posameznih cest, prikazana poprečna obremenitev cest v 24 urah v tonah. Debelina pasov je v sorazmerju z obremenitvijo. Tako je nazorno prikazana obremenitev posameznih prometnih žil.

Prave kartograme pa imenujemo grafikone, ki v geografski karti prikažejo regionalno gostitev oziroma intenziteto pojava s šrafurami ali točkami.

5.44 S šrafurami prikažemo gostitev pojava tako, da večjo ali manjšo gostitev ali intenziteto na določenem območju prikažemo s temnejšo ali svetlejšo šrafuro. Vtis o regionalni gostitvi dobimo iz razmestitve temnejših in svetlejših ploskev v geografski karti.

V sliki 5.21 je s šrafurami prikazana gostota prebivalstva po popisu 15.3.1948. V legendi je dana lestvica šrafur, ki je skladna z gostoto prebivalstva. Prva - najsvetlejša šrafura pomeni gostoto do 30 prebivalcev na km^2, druga-temnejša, gostoto nad 30 do



Slika 5.21. Kartogram gostote prebivalstva v Sloveniji po popisu 15.3.1948. (Vir: Statistični atlas Slovenije.)

40 prebivalcev na km^2, tretja - še temnejša, gostoto nad 40 do 50 prebivalcev na km^2 itd. Kartogram pokaže veliko gostoto prebivalstva v vzhodnih krajih Slovenije, veliko gostoto prebivalstva v savski dolini in majhno gostoto prebivalstva v hribovitih predelih.

V kartogramih s šrafurami pa ne prikazujemo samo gostote, temveč na splošno vsa relativna števila: strukturne deleže, koeficiente, gostote in indekse.

Ko sestavljamo kartogram s šrafurami, moramo predhodno proučiti variabilnost podatkov po geografskih enotah, ki so osnova kartograma. Glede na variabilnost podatkov se odločimo za razrede tako, da število razredov ni preveliko (6-10). Premajhno število razredov ne pokaže bistva regionalnih razlik za pojav. Če pa vzamemo preveč razredov, težko sestavimo tako skalo šrafur, da moremo iz nje nedvoumno razbrati rang po intenziteti.

5.45 Zelo nazorno prikažemo gostitev pojava v kartogramu s t o č k a m i. Medtem ko v kartogramih s šrafurami prikazujemo relativna števila, v kartogramu s točkami prikazujemo absolutne podatke. Ko predhodno proučimo velikost pojava, se odločimo, koliko enot osnovnega pojava prikažemo z eno točko oziroma krogcem. V vsako najmanjšo regionalno enoto, za katere imamo dane podatke, ki jih prikazujemo v kartogramu, narišemo toliko točk, kolikor jih ustreza velikosti pojava v tem najmanjšem regionalnem delu. Prednost te metode je tudi v tem, da moremo razmestiti točke ustrezno regionalni razmestitvi tudi znotraj najmanjših geografskih delov. Tako dobimo naravno in podrobno sliko o regionalni razmestitvi pojava. Za zgled je v sliki 5.22 prikazana s kartogramom s točkami razmestitev vinske trte v Sloveniji po popisu 1. 2. 1949. Vsaka točka pomeni pet tisoč trt. Iz grafikona nazorno vidimo vinorodne kraje v Sloveniji in regionalno gostoto vinske trte.



Slika 5.22. Kartogram s točkami za vinsko trto v Sloveniji po popisu 1.2.1949. (Vir: Statistični atlas Slovenije).

ŠESTO POGLAVJE

RELATIVNA ŠTEVILA

6.1 Že v poglavju o prikazovanju statističnih podatkov smo opozorili, da en sam statistični podatek, čeprav pomemben, dobi pravo vrednost šele tedaj, ko ga primerjamo z drugimi sorodnimi ali vsebinsko povezanimi podatki. Zaradi tega že v tabelah prikazujemo podatke tako, da je primerjava čim boljša.

Vendar za proučevanje statističnih podatkov ni zadostna le površna primerjava, kakršno dobimo iz statistične tabele z osnovnimi podatki. Primerjavo zelo poglobimo in olajšamo z relativnimi števili. Veliko boljšo predstavo o številu prebivalstva v Sloveniji v primerjavi s številom prebivalstva Jugoslavije dobimo, če rečemo, da je v Sloveniji 8,58% od skupnega števila prebivalstva v Jugoslaviji, kakor pa če navedemo, da je bilo v Sloveniji po popisu 31.3.1961 1,591.523 prebivalcev od skupnega števila 18.549.291 prebivalcev v Jugoslaviji. Enako dobimo najboljši pregled o prirastku za število prebivalstva v Sloveniji, če ga izrazimo relativno in rečemo, da se je število prebivalstva v razdobju med popisoma 31.3.1953 in 31.3.1961 povečalo v Sloveniji za 8,53%. Tako določenega in hitrega vtisa o povečanju ne dobimo niti, če povemo, da je bilo število prebivalstva 31.3.1953 1,466.425, 31.3.1961 pa 1,591.523 prebivalcev, niti če povemo, da se je število prebivalstva v osmih letih med popisoma povečalo za 125.098 prebivalcev.

Relativno število iz dveh podatkov, ki ju primerjamo, je v splošnem kvocient med primerjanima podatkoma. Razumljivo je, da enako kakor med seboj primerjamo le

podatke, za katere je primerjava smiselna, izračunavamo tudi relativna števila le iz podatkov, ki so v medsebojni vsebinski povezavi. Relativna števila izračunamo v oblikah, ki so prikladnejše za prikazovanje in analizo (npr. v odstotkih, indeksih ipd.).

6.2 Po tem, v kakšni zvezi sta si podatka, iz katerih izračunamo relativno število, razlikujemo:

a) strukture ali razčlenitvena števila, kadar primerjamo del s celoto (npr. od skupnega števila prebivalstva v Jugoslaviji je bilo v letu 1961 8,56% v Sloveniji).

b) koeficiente in gostote, kadar primerjamo raznovrstne podatke, ki so v medsebojni zvezi (npr. proizvodnja jekla na 1 prebivalca je bila v letu 1966 v SFRJ 94,6 kg/prebivalca);

c) indekse, kadar primerjamo prirejene istovrstne podatke (npr. indeks spremembe za število prebivalstva v Sloveniji od popisa 1953 do popisa 1961 je 108,5 če vzamemo, da je število prebivalstva v letu 1953 enako 100,0).

STRUKTURE ALI RAZČLENITVENA ŠTEVILA.

Enostavne strukture.

6.3 Volilne rezultate običajno ne objavljamo z absolutnim številom glasov, marveč z odstotkom glasov, ki jih je kandidat dobil v primerjavi s skupnim številom oddanih glasov. Šele odstotek dobljenih glasov namreč pokaže stvaren uspeh kandidata v določenem volilnem okolišu ali občini.

Število oddanih glasov za nekega kandidata ne pokaže uspeha zato, ker je bistveno odvisno od skupnega števila oddanih glasov. Enako število glasov, oddano za določenega kandidata, pomeni uspeh, če je volilni okoliš majhen in neuspeh, če je volilni okoliš velik. Z odstotki dobljenih glasov pa velikosti obeh okolišev računsko izenačimo. Odstotki namreč pokažejo, koliko od skupno sto oddanih glasov je v poprečju dobil kandidat N.N. v prvem in drugem okolišu. Zaradi tega le odstotek glasov meri uspeh kandidata in ne absolutno število oddanih glasov.

Ta način na splošno uporabljamo vselej, kadar hočemo pregledno in primerljivo prikazati sestavo statističnih populacij. Največkrat ne izračunamo samo enega strukturnega pokazovalca, temveč cele vrste struktur, ki nazorno in primerljivo prikažejo sestave populacij.

Strukture izražamo v delih od celote ali v odstotkih. Če pa so deli izjemno majhni, izražamo strukture tudi v promilih. Strukturni koeficient Y_1^o dobimo, da delni podatek delimo s celoto. Strukturni odstotek dobimo tako, da delni podatek delimo s skupnim, kvocient pa pomnožimo s 100. Če izražamo strukturni delež v promilih, pa kvocient pomnožimo s 1000. Računski postopek za izračunanje strukturnih pokazovalcev moremo nakazati tudi z obrazci:

$$Y_1^o = \frac{Y_1}{Y} \quad ; \quad Y_1\,\% = 100\,\frac{Y_1}{Y} \quad ; \quad Y_1\,\text{‰} = 1000\,\frac{Y_1}{Y} \qquad (6.1)$$

Pri tem je Y_1 podatek za del populacije, Y podatek za populacijo, Y_1^o strukturni koeficient, $Y_1\,\%$ strukturni odstotek, $Y_1\,\text{‰}$ pa strukturni delež, izražen v promilih.

6.4 Uporabnost enostavnih strukturnih vrst za statistično analizo prikažimo s sestavo izvoza iz SFRJ po stopnji obdelave izvoženih proizvodov v razdobju 1946–1971.

Spremembe v sestavi izvoza po stopnji obdelave za izvožene proizvode iz absolutnih podatkov ne vidimo neposredno, ker je skupen obseg izvoza od leta do leta različen. Če primerjamo absolutne vrednosti izvoza po posameznih stopnjah obdelave, opazimo, da se vse tri komponente v absolutnem časovno večajo. Šele časovne vrste struktur izvoza odkrijejo spremembe v strukturi izvoza. Analiza struktur pokaže, da je delež visoko obdelanih proizvodov v našem izvozu od leta do leta večji, kar je rezultat industrializacije.

Tabela 6.1. Vrednost izvoza iz SFRJ po stopnji obdelave izvoženih proizvodov v razdobju 1946-1971. (Vir: Jugoslavija 1946-1966, SG-72)

Stopnja obdelave	Vrednost v mlj. din						struktura v %					
	1946	1951	1956	1961	1966	1971	1946	1951	1956	1961	1966	1971
Skupno	162	536	970	1707	15251	27217	100	100	100	100	100	100
Neobdelani proizvodi	65	183	355	421	2050	3113	40	34	37	25	13	11
Proizvodi srednje obdelave	56	309	415	655	5064	8622	34	58	43	38	33	32
Proizvodi visoke obdelave	42	44	199	631	8137	15482	26	8	20	37	54	57

Večkratne strukture.

6.5 Če je populacija razdeljena po več znakih hkrati, moremo zanjo izračunati več vrst struktur. Vsaka od njih po svoje osvetli prikazano populacijo. Vzemimo kot zgled razdelitev družbenega produkta v SFRJ v letu 1966 po elementih in gospodarskih sektorjih

Tabela 6.2. Družbeni produkt v SFRJ v letu 1966 po elementih in sektorjih v mlj. dinarjev (Vir: SGJ. 67)

Gospodarski sektor	neto osebni dohodki	akumulacija in fondi	amortizacija	skupno
Družbeni	19944	37095	5412	62451
Privatni	13057	3477	530	17064
Skupno	33001	40572	5942	79515

V tej tabeli moremo vzeti kot celoto skupen družbeni produkt (79515 mlj. dinarjev) in izračunati v odstotkih, kakšen delež te celote je vsak izmed podatkov. Neto osebni dohodki v družbenem sektorju so po tem načelu

$$100 \frac{19944}{79515} = 25,1\ \%$$

od skupnega družbenega produkta.

Struktura, izračunana na tej osnovi, je dana v tabeli 6.3 .

Tabela 6.3. Sestava družbenega produkta v SFRJ v letu 1966 po elementih in sektorjih

Sektor	Neto osebni dohodki	Akumulacija in fondi	Amortizacija	S k u p n o
Splošno družbeni	25,1	46,6	6,8	78,5
Privatni	16,4	4,4	0,7	21,5
Skupno	41,5	51,0	7,5	100,0

Ta tabela struktur po svoje osvetli notranjo sestavo družbenega produkta. Z njo je možno primerjati sestavo družbenega produkta za druga leta, posamezne republike itd.

Iz tabele 6.2 o sestavi družbenega produkta pa moremo izračunati tudi druge vrste struktur. Dohodki v splošno družbenem sektorju so del skupnega družbenega produkta, so pa istočasno del družbenega produkta v splošno družbenem sektorju (62.451 mlj din) in del skupnih osebnih dohodkov (33.001 mlj din). Zaradi tega moremo razen gornje tabele o sestavu izračunati še dve tabeli struktur.

Tabela 6.4. Struktura družbenega produkta sektorjev po elementih.

Gospodarski sektor	Neto osebni dohodki	Akumulacija in fondi	Amortizacija	S k u p n o
Družbeni	31,9	59,4	8,7	100,0
Privatni	76,5	20,4	3,1	100,0
Skupno	41,5	51,0	7,5	100,0

Tabela 6.5 Struktura elementov družbenega produkta po gospodarskih sektorjih

Gospodarski sektor	Neto osebni dohodki	Akumulacija in fondi	Amortizacija	S k u p n o
Družbeni	60,4	91,4	91,1	78,5
Privatni	39,6	8,6	8,9	21,5
Skupaj	100,0	100,0	100,0	100,0

V tabeli 6.4 vidimo velike razlike v sestavi družbenega produkta po sektorjih. Medtem ko so na primer v družbenem sektorju neto osebni dohodki samo 31,9% družbenega produkta, pomenijo neto osebni dohodki v privatnem sektorju 76,5% od skupnega družbenega produkta. Podobno so velike razlike tudi v drugih sestavinah.

Tudi tabela 6.5 je zelo poučna. Udeležba sektorjev v vsakem izmed treh elementov je zelo različna in po svoje osvetli sestavo družbenega produkta. Medtem ko je udeležba posameznega sektorja v skupni akumulaciji, fondih in amortizaciji razmeroma podobna, je udeležba v skupnih osebnih dohodkih bistveno različna.

Splošne zveze za dvojne strukture

6.6 Simbolično moremo vse možne deleže dvojnih struktur nakazati takole: Če z N_{AB} zaznamujemo kombinacijsko tabelo za število enot in ustrezne robne vsote $N_A = \sum_B N_{AB}$: $N_B = \sum_A N_{AB}$: $N = \sum_A N_A = \sum_B N_B$ zapišemo splošno kombinacijsko tabelo z vsotami takole

$$\begin{array}{c|c} N_{AB} & N_A \\ \hline N_B & N \end{array} \qquad (6.2)$$

Ustrezni strukturni deleži na celotno populacijo so

$$\begin{array}{c|c} \frac{N_{AB}}{N} = P_{AB} & \frac{N_A}{N} = P_A \\ \hline \frac{N_B}{N} = P_B & \frac{N}{N} = 1 \end{array} \qquad (6.3)$$

Vrstične ali stolpične strukturne vrste zapišimo takole

$$\begin{array}{c|c} \frac{N_{AB}}{N_B} = P_{A/B} & \frac{N_A}{N} = P_A \\ \hline \frac{N_B}{N_B} = 1 & \frac{N}{N} = 1 \end{array} \qquad (6.4)$$

$$\frac{N_{AB}}{N_A} = P_{B/A} \qquad \frac{N_A}{N_A} = 1$$

$$\frac{N_B}{N} = P_B \qquad \frac{N}{N} = 1 \tag{6.5}$$

$P_{A/B}$ pomeni strukturne vrste po znaku A pri pogoju, da se podatki nanašajo na posamezno vrednost B in

$P_{B/A}$ strukturne vrste po znaku B za posamezne delne populacije po znaku A.

Iz gornjih tabel dobimo naslednje zveze:

$$\frac{N_{AB}}{N} = \frac{N_B}{N} \cdot \frac{N_{AB}}{N_B} = \frac{N_A}{N} \cdot \frac{N_{AB}}{N_A} \quad \text{ali} \tag{6.6}$$

$$P_{AB} = P_B \cdot P_{A/B} = P_A \cdot P_{B/A}$$

iz te zveze pa dobimo dalje obrazec

$$P_{B/A} = \frac{P_{AB}}{P_A} = \frac{P_B \cdot P_{A/B}}{\sum P_B \cdot P_{A/B}} \tag{6.7}$$

S pomočjo teh obrazcev dobimo strukturne vrste P_{AB} in vrste $P_{B/A}$ posredno iz strukturnih vrst $P_{A/B}$, če poznamo sumarno strukturno vrsto P_B, brez osnovnih podatkov N_{AB}

6.7 Vzemimo za zgled, da poznamo za štiri oddelke (0_1, 0_2, 0_3, 0_4) strukturo skupne proizvodnje za oddelke P_B in strukturo izdelkov po kakovosti (I. II. III) $P_{A/B}$

Tabela 6.6 Struktura kakovosti $P_{AB} = P_B P_{A/B}$

$P_{A/B}$	I.	II.	III.	$\sum$	P_B	P_{AB}	I.	II.	III.	P_B
0_1	·50	·30	·20	1·00	·40	0_1	·20	·12	·08	·40
0_2	·30	·40	·30	1·00	·20	0_2	·06	·08	·06	·20
0_3	·10	·20	·70	1·00	·30	0_3	·03	·06	·21	·30
0_4	·80	·10	·10	1·00	·10	0_4	·08	·01	·01	·10
						P_A	·37	·27	·36	1·00

in tabela pogojnih strukturnih deležev $P_{B/A} = \frac{P_{AB}}{P_A} = \frac{P_B \cdot P_{A/B}}{\sum_B P_B \cdot P_{A/B}}$

Tabela 6.7 Struktura kakovosti po oddelkih

$P_{B/A}$	I	II	III	P_B
0_1	·54	·44	·22	·40
0_2	·16	·30	·17	·20
0_3	·08	·22	·58	·30
0_4	·22	·04	·03	·10
	1.00	1.00	1.00	1.00

Iz tabele ugotavljamo, kako so izdelki za posamezno kakovost porazdeljeni po oddelkih ·54 ali 54 % vseh izdelkov kakovosti I izhaja iz prvega 0_1 oddelka ipd.

6.8 Če štejemo relativne deleže za priorne verjetnosti, nakazane zveze med strukturnimi deleži veljajo za verjetnosti $Pr(A \cap B)$, $Pr(A)$, $Pr(B)$, $Pr(A/B)$ in $Pr(B/A)$ za dogodke A in B , ki so analogne ustreznim strukturnim deležem.

Razen stavka o množenju verjetnosti za odvisne dogodke

$$Pr(A \cap B) = Pr(B) \cdot Pr(A/B) = Pr(A) \cdot Pr(B/A) \qquad (6.8)$$

dobimo pomemben Bayesov obrazec za inverzno verjetnost

$$Pr(B/A) = \frac{Pr(B) \cdot Pr(A/B)}{\sum_B Pr(B) \cdot Pr(A/B)} \qquad (6.9)$$

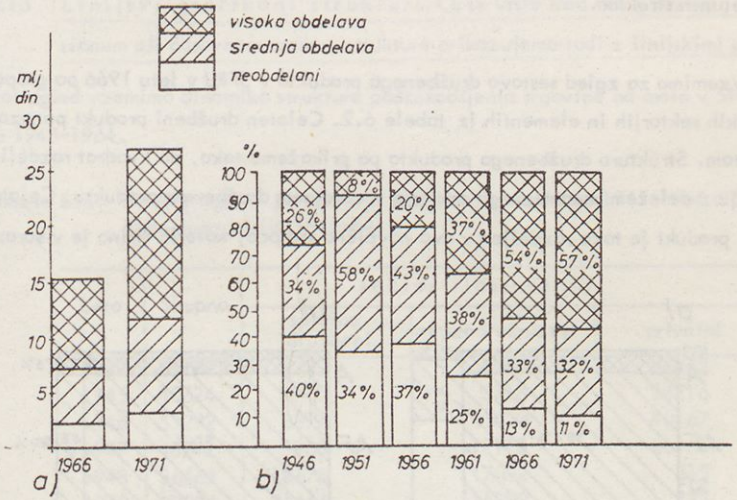
s katerim bomo v kasnejših poglavjih izvajali pomembne sklepe.

Grafično prikazovanje struktur

6.9 Vpogled v sestavo populacije dobimo zelo nazorno z grafikoni. Strukture grafično prikazujemo z vsemi vrstami grafikonov: s stolpci, krogi, linijami, figurami itd.

6.10 Strukturni stolpci. Z navadnim stolpcem prikažemo en sam podatek. Višina stolpca je v sorazmerju z velikostjo podatka, ki ga stolpec ponazarja. Če pa stolpec, ki predstavlja celoto, razdelimo v sorazmerju s sestavo populacije, dobimo strukturni stolpec, ki prikaže sestavo populacije. Več sorodnih strukturnih stolpcev pa nazorno pokaže spremembe oziroma razlike v sestavi.

Vzemimo za zgled vrednost izvoza iz Jugoslavije po stopnji obdelave v razdobju 1946-1971 iz tabele 6.1. Strukturni stolpci iz absolutnih podatkov v tej tabeli so prikazani v sliki 6.1.



Slika 6.1. Struktura vrednosti izvoza iz SFRJ po stopnji obdelave izvoženih proizvodov

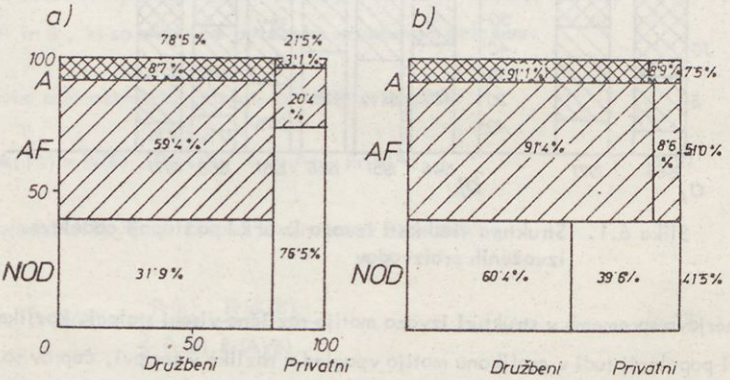
Primerjavo sprememb v strukturi izvoza motijo različno visoki stolpci. Razlike v velikosti populacij tudi v grafikonu motijo vpogled v razlike v sestavi, čeprav so iz grafikona spremembe v strukturi le bolj vidne kakor iz tabele absolutnih podatkov.

Veliko bolj neposredno je sprememba strukture izvoza vidna iz slike 6.1b, v kateri je struktura izvoza prikazana z enako visokimi stolpci, ki so razdeljeni glede na strukturne deleže. Pri tem sicer izgubimo pregled o spremembah v skupni vrednosti izvoza, ven-

dar boljši pregled o spremembah v sestavu izvoza odtehta to pomanjkljivost. To pomanjkljivost včasih odpravimo tako, da v isti sliki z linijskim grafikonom ponazorimo gibanje skupne vrednosti.

6.11 Strukturni stolpci za večkratne strukture. Širine strukturnih stolpcev iz prejšnjega zgleda so enake za vsa leta. Širine stolpcev pa moremo risati v sorazmerju z obsegom populacije. Vrsta takih strukturnih stolpcev pokaže tudi razlike v obsegu populacije. Ker pa so vsi strukturni stolpci enako visoki, je iz njih neposredno vidna sprememba v strukturi. Ta postopek s pridom izkoriščamo predvsem za prikazovanje večkratnih struktur.

6.12 Vzemimo za zgled sestavo družbenega produkta v SFRJ v letu 1966 po gospodarskih sektorjih in elementih iz tabele 6.2. Celoten družbeni produkt ponazorimo s kvadratom. Strukturo družbenega produkta pa prikažemo tako, da kvadrat razdelimo v sorazmerju z deležem posameznega sektorja v celotnem družbenem produktu. Celoten družbeni produkt je tako razdeljen v dva navpična stolpca, katerih širina je v sorazmer-



Slika 6.2 Struktura družbenega produkta v SFRJ v letu 1966 po elementih in gospodarskih sektorjih.

ju z deležem posameznega sektorja. Če vsakega izmed teh stolpcev razdelimo v navpični smeri v sorazmerju s sestavo družbenega produkta za posamezni sektor po elementih, dobimo sliko 6.2a. Ta slika je fotografija tabele 6.4. Iz grafikona je zelo nazorno vid-

na struktura družbenega produkta po sektorjih, predvsem pa razlike v sestavi med sektorji. Poleg tega pa je ploščina posameznega pravokotnika v kvadratu sorazmerna z absolutno velikostjo določenega elementa oziroma s strukturnimi odstotki iz tabele 6.4.

Če kvadrat, ki je okvir grafikona, razdelimo najprej v sorazmerju s sestavo skupnega družbenega produkta po elementih, dobljene delne stolpce p- nodobno kakor v prejšnjem primeru, v sorazmerju s sestavo posameznega elementa po sektorjih, dobimo sliko 6.2b: ta grafično pove isto kakor tabela 6.5.

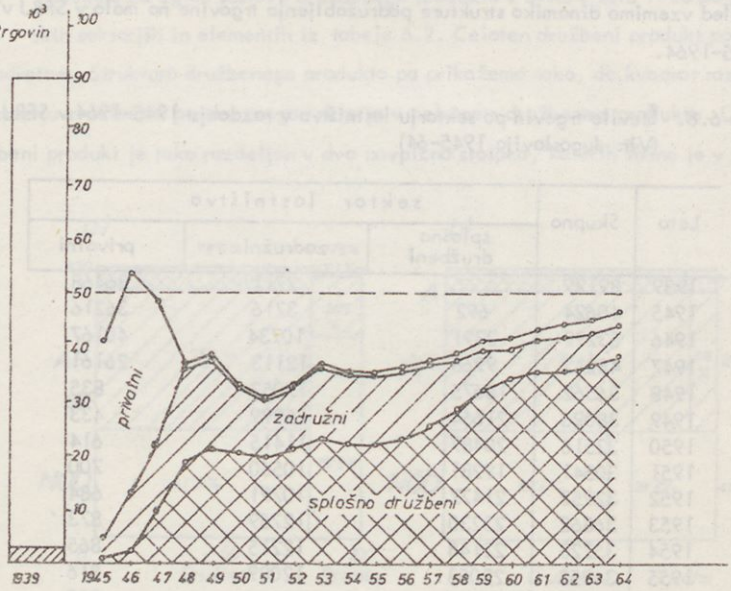
6.13 Linijski grafikoni struktur. Če se vrste med seboj razlikujejo v numeričnem ali časovnem znaku, strukture prikazujemo tudi z linijskimi grafikoni.

Kot zgled vzemimo dinamiko strukture podružabljenja trgovine na malo v SFRJ v razdobju 1945-1964.

Tabela 6.8. Število trgovin po sektorju lastništva v razdobju 1945-1964 v SFRJ (Vir: Jugoslavija 1945-64)

Leto	Skupno	sektor lastništva		
		splošno družbeni	zadružni	privatni
1939	89179	.	2711	86468
1945	40624	692	3716	36216
1946	53292	2391	10734	40167
1947	48242	9968	12113	26161
1948	36562	18675	17052	835
1949	38096	21664	15999	433
1950	32516	20487	11415	614
1951	30541	19281	10560	700
1952	32952	21477	10791	684
1953	36862	23230	12759	873
1954	35228	22148	12215	865
1955	34958	22053	12089	816
1956	36257	23156	12243	858
1957	37651	25482	11169	1000
1958	38608	28385	8923	1300
1959	40980	32361	6919	1700
1960	41286	33897	5289	2100
1961	42915	35199	5416	2300
1962	43866	35125	6191	2550
1963	44432	36218	5954	2260
1964	46043	37864	5929	2250

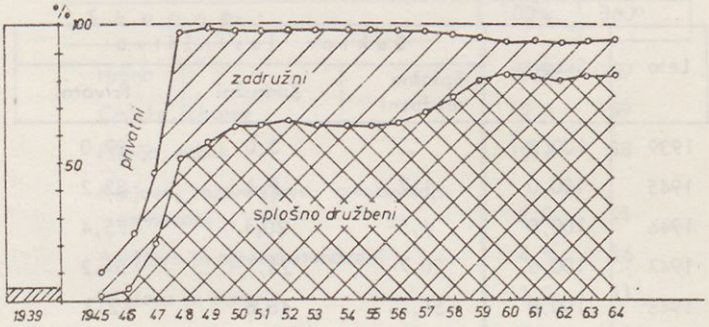
Podatki iz tabele 6.8. so v sliki 6.3. za posamezne sektorje nanešeni drug na drugega. Za leto 1947 npr. s točkami vnesemo podatek 9968, nadalje 9968 + 12113 = = 22081 in končno 22081 + 26161 = 48244. Če enako vrišemo strukturo za druga leta in zvežemo ustrezne točke, dobimo linijski grafikon, v katerem posamezni pasovi ponazarjajo dinamiko števila trgovin v posameznem sektorju lastništva. Ker je leto 1939 od zveznega razdobja 1947-1964 ločeno, je struktura za 1939 vrisana v stolpcu. Iz grafikona je nazorneje kot iz tabele vidno podružabljenje v trgovini in dinamika v strukturi v družbenem sektorju. Po velikih spremembah v strukturi v povojnih letih 1945-1948 se struktura v nadaljnjih letih malo spreminja s težnjo k povečanju števila trgovin v splošno družbenem sektorju, nazadovanju zadružnega sektorja in rahlim povečevanjem v privatnem sektorju.



Slika 6.3. Struktura števila trgovin na malo v SFRJ v razdobju 1944-1966 po sektorjih lastništva

Enako kakor smo prikazali strukturo z linijskim grafikonom absolutnih podatkov, moremo prikazati tudi vrste strukturnih odstotkov. Ta slika nazorneje pokaže dinamiko v

strukturi kot slika absolutnih podatkov, izgubimo pa sliko o spremembah skupnega števila trgovin.



Slika 6.4 Struktura števila trgovin na malo v SFRJ v razdobju 1945-1964 po sektorju lastništva.

V sliki 6.4 je z linijskim grafikonom prikazana struktura števila trgovin na malo iz tabele 6.9.

6.14 S t r u k t u r n i k r o g i . Za prikazovanje struktur se izkaže kot primeren tudi krog, ker nazorno prikazuje celoto. Strukturne deleže ponazorimo s k r o k o v i m i i z s e k i , kot za posamezen izsek pa je v sorazmerju z velikostjo strukturnega deleža. Ker četrtina kroga ponazarja 25%, polovica kroga pa 50%, s pogledom enostavno ocenimo velikost strukturnih deležev.

Ker imamo običajno kotomere z ločnimi stopinjami, moramo strukturne deleže Y % pred risanjem preračunati po obrazcu

$$Y_1^{st} = 3{,}6\ Y_1\ \% \qquad (6.10)$$

v ločne stopinje Y_1^{st}

Tabela 6.9. Struktura števila trgovin na malo v SFRJ po sektorju lastništva v razdobju 1945-1964

Leto	Skupno	Sektor lastništva		
		Splošno družbeni	Zadružni	Privatni
1939	100,0	.	3,0	97,0
1945	100,0	1,7	9,1	82,2
1946	100,0	4,5	20,1	75,4
1947	100,0	20,7	25,1	54,2
1948	100,0	51,1	46,6	2,3
1949	100,0	56,9	42,0	1,1
1950	100,0	63,0	35,1	1,9
1951	100,0	63,1	34,6	2,3
1952	100,0	65,2	32,7	2,1
1953	100,0	63,0	34,6	2,4
1954	100,0	62,9	34,7	2,4
1955	100,0	63,1	34,6	2,3
1956	100,0	63,8	33,8	2,4
1957	100,0	67,7	29,6	2,7
1958	100,0	73,5	23,1	3,4
1959	100,0	79,0	16,9	4,1
1960	100,0	82,1	12,8	5,1
1961	100,0	82,0	12,6	5,4
1962	100,0	80,1	14,1	5,8
1963	100,0	81,5	13,4	5,1
1964	100,0	82,2	12,9	4,9

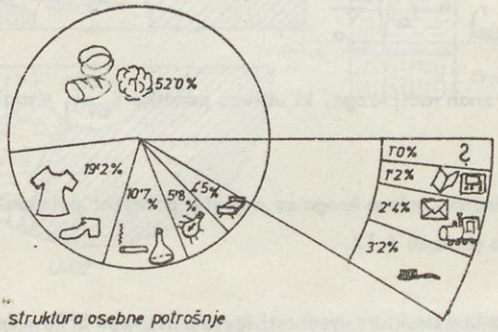
6.15 Kot zgled bomo prikazali strukturo za osebno porabo materialnih dobrin in proizvodnih stroškov za prebivalstvo v SFRJ v letu 1956.

Tabela 6.10 Struktura za osebno porabo materialnih dobrin in proizvodnih uslug prebivalstva v SFRJ v letu 1956 (Vir: SG 58)

S k u p n o	100%	360°
Hrana	52,0	187
Obleka, obutev	19,2	69
Pijača, tobak	10,7	38
Kurjava, razsvetljava, vzdrževanje stanovanj	5,8	21
Pohištvo in oprema stanovanj	4,5	16
Higiena in zdravje	3,2	11
Promet, pošta	2,4	9
Knjige, tisk, radio itd.	1,2	5
Nerazporejeno	1,0	4

Ločne stopinje smo izračunali: $3,6.52,0\% = 187^{\circ}$ itd.

Posamezne postavke so zaradi preglednosti v tabeli in grafikonu nanizane po velikosti. To olajša in izboljša pregled. V sliki so v ustreznih izsekih vpisani odstotki in shematično ponazorjena vsebina podatka, kar olajša čitanje grafikona. Posebnost grafikona je tudi v tem, da so izseki kroga za manjše postavke povečani. Tako izboljšamo preglednost teh podatkov.



Slika 6.5 Struktura za osebno porabo materialnih dobrin in proizvodnih storitev za prebivalstvo SFRJ v letu 1956

6.16 Če med seboj primerjamo več strukturnih vrst, strukturni krogi običajno niso prikladni. Primerljivost med strukturami, če so prikazane s krogi, je namreč veliko slabša kakor če uporabimo stolpce ali linije. Kljub temu včasih kroge uporabimo za prikazovanje več struktur. Uporabni so npr. za prikazovanje regionalnih sprememb struktur. Strukturne kroge v tem primeru vnašamo v ustrezna območja v mreži geografske karte. Ta način je razmeroma ugoden zaradi primerne oblike kroga.

6.17 Sestavi dveh populacij, ki sta si v zvezi tako kakor npr. uvoz in izvoz, nabavljeno in prodano, dohodki in izdatki, dostikrat prikažemo s polkrogi. Posamezno populacijo prikažemo s polovico kroga. Strukturne odstotke $Y_1\%$ v tem primeru preračunamo v ločne stopinje Y_1^{st} po obrazcu

$$Y_1^{st} = 1{,}8\, Y_1\% \qquad (6.11)$$

ker celoto predstavlja le 180^o.

Če ne upoštevamo absolutne velikosti pojava, imata obe polovici kroga isti radij. Z velikostjo radija kroga pa nakažemo velikost pojava. Ker mora biti ploščina, ne pa radij kroga ali polkroga, sorazmerna velikosti pojava, radija krogov, za katere želimo, da so njihove ploščine v sorazmerju z velikostjo pojava, izračunamo po obrazcu

$$r_1 = r_o\sqrt{\frac{Y_1}{Y_o}} \qquad (6.12)$$

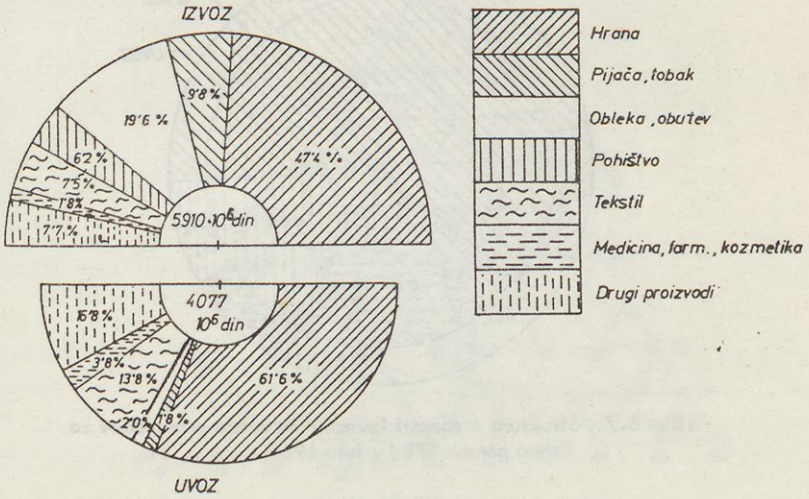
Pri tem pomeni: r_o = znan radij kroga, ki ustreza podatku Y_o; r_1 = radij kroga za podatek Y_1.

Običajno vnaprej določimo polmer kroga za največji podatek, polmere vseh drugih krogov pa izračunamo po obrazcu 6.12.

6.18 Za zgled vzemimo strukturo vrednosti izvoza in uvoza SFRJ v letu 1966.

Tabela 6.11 Struktura vrednosti izvoza in uvoza proizvodov za široko porabo v SFRJ v letu 1966

Skupina predmetov	Struktura v %		Stopinje v polkrogih	
	Izvoz	Uvoz	Izvoz	Uvoz
Skupno v mlj.din	5910 %	4077 = %		
	100,0	100,0	180°	180°
Hrana	47,4	61,6	85	111
Pijača in tobak	9,8	1,8	18	3
Obleka, obutev	19,6	2,0	35	4
Pohištvo	6,2	0,2	11	0
Tekstil (brez obleke)	7,5	13,8	14	25
Med.farmac.kozm. preparati	1,8	3,8	3	7
Drugi proizvodi	7,7	16,8	14	30



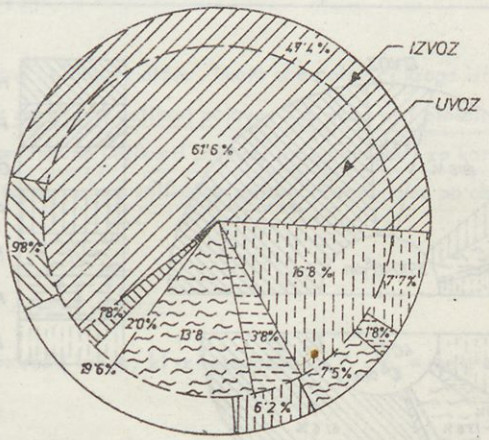
Slika 6.6. Struktura vrednosti izvoza in uvoza SFRJ v letu 1966.

Če se odločimo, da je radij večjega polkroga (za izvoz) enak 4 cm, izračunamo radij polkroga za uvoz po obrazcu

$$r_U = r_I \sqrt{\frac{U}{I}} = 4 \cdot \sqrt{\frac{4077}{5910}} = 3{,}32 \text{ cm}$$

Ker sta ploščini polkrogov sorazmerni z absolutnima vrednostima uvoza in izvoza, je iz grafikona razen strukturvidno razmerje med uvozom in izvozom. Vseeno pa absolutni znesek vpišemo sredi polkrogov. Strukturi izvoza in uvoza pa v tem grafikonu nista neposredno primerljivi.

6.19 Primerjava strukture izvoza s strukturo uvoza pa je veliko bolj neposredna v sliki 6.7. V tej sliki si mislimo, da je strukturni krog za uvoz (manjši krog) položen koncentrično na strukturni krog za izvoz (večji krog). Ker so izseki kroga za ustrezne skupine proizvodov blizu skupaj, je primerjava struktur veliko boljša. Da še izboljšamo primerjavo ustreznih krogovih izsekov, je krožnica manjšega kroga na mestih, kjer se stikata ustrezni skupini, nepoudarjena.

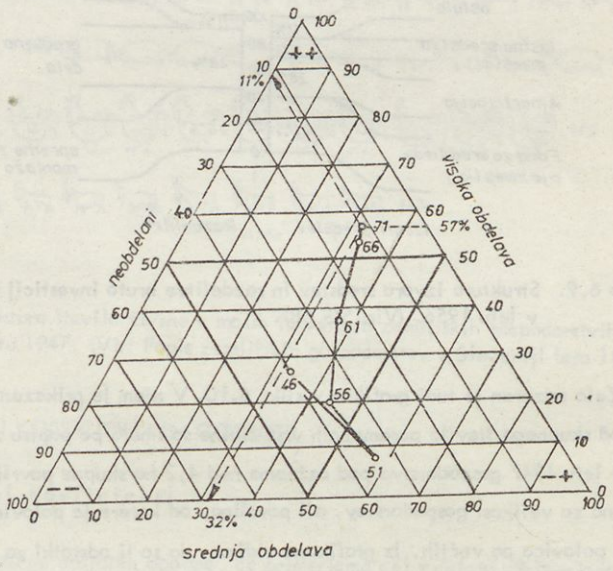


Slika 6.7. Struktura vrednosti izvoza in uvoza proizvodov za široko porabo SFRJ v letu 1966.

6.2o Strukture v trikotniku. Ker je vsota strukturnih odstotkov za vsako populacijo enaka 100%, moremo strukture populacij, ki so razdeljene v tri dele: $Y_1\% + Y_2\% + Y_3\% = 100$, prikazati s točko v tako imenovani trikotniški mreži oziroma koordinatnem sistemu. Ta način je za strukture s tremi členi zelo prikladen. Več

sorodnih strukturnih vrst s tremi členi v njej enostavno prikažemo s tolikimi točkami, kolikor imamo strukturnih vrst. Te točke glede na značaj vrst povezujemo v linijske grafikone, vektorje itd.

6.21 V sliki 6.8 je prikazana struktura vrednosti izvoza SFRJ po stopnji obdelave v razdobju 1946-1971 iz tabele 6.1. Iz grafikona vidimo, da je na vsaki stranici enakostraničnega trikotnika skala odstotkov za eno izmed treh stopenj obdelave. Trikotnik je razdeljen z mrežo vzporednic k posameznim osem. V sliki 6.8 je za leto 1971 nakazano, kako odčitamo na skalah odstotke za posamezno stopnjo obdelave. Črte za 50% so odebeljene. Tako je trikotnik razdeljen v štiri manjše enakostranične trikotnike. Točka v notranjem trikotniku nakaže strukturo, v kateri noben izmed treh odstotkov ni večji kot 50%, točke v drugih trikotnikih pa nakazujejo strukture, v katerih je ustrezen odstotek večji kot 50%. Tako dobimo neposredno grobo orientacijo o strukturi že iz razdelitve na te štiri dele.



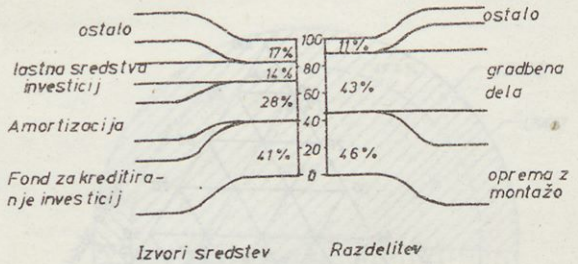
Slika 6.8. Struktura za vrednost izvoza SFRJ po stopnji obdelave v razdobju 1946-71.

Iz slike 6.8 nazorno vidimo posebno velike in značilne spremembe v strukturi kakovosti izvoza. Za leto 1951 je struktura izvoza v primerjavi 1946 doživela velike spremembe

v zvezi z izdelki srednje in visoke obdelave. Od leta 1956 naprej pa je vidna trajna težnja v spremembi strukture v smeri zmanjšanja izvoza neobdelanih proizvodov v dobro obdelanih proizvodov.

Skale za neobdelano (NO) srednjo obdelavo (SO) in za visoko obdelavo (VO) so namenoma postavljene tako, da je levi trikotnik za neobdelano, v desni smeri se veča srednja obdelava, navzgor pa visoka obdelava.

6.22 Druge metode za prikazovanje struktur. Ker si moremo predstavljati bilančne podatke kot tok, ki na eni strani doteka, na drugi pa odteka, strukturo bilanc nazorno prikažemo po metodi, ki je nakazana v sliki 6.9. Grafikon ponazarja izvor in razdelitev bruto investicij v letu 1956 v SFRJ. Posamezni deli izvora in razdelitve so risani shematično kot pretoki, kar vzbuja vtis o dotoku in odtoku.

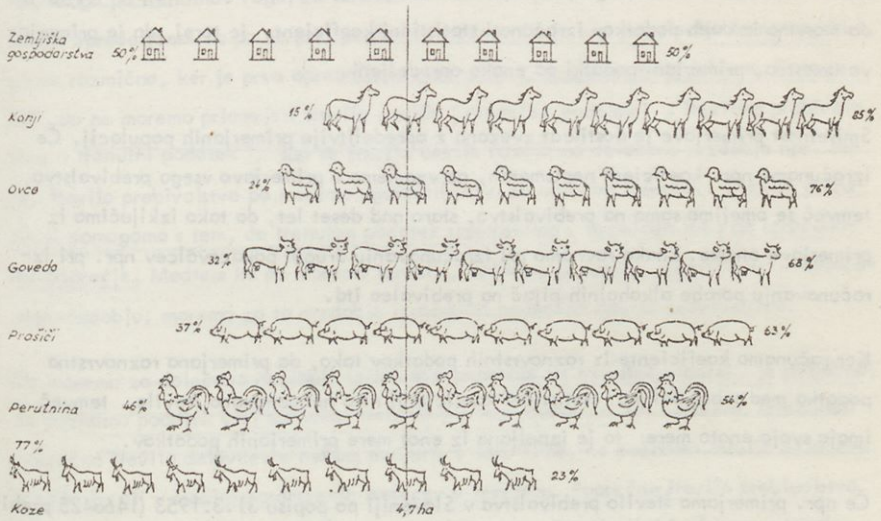


Slika 6.9. Struktura izvora sredstev in razdelitve bruto investicij v SFRJ v letu 1956. (Vir: SG 58).

6.23 Zelo nazoren je tudi grafikon v sliki 6.10. V njem je prikazano, kolikšen del od skupnega števila posameznih vrst živine so imela po popisu zemljiških gospodarstev v letu 1947 gospodarstva pod oziroma nad 4,7 ha skupne površine. Ta površina je mediana za velikost gospodarstev, ali površina, od katere je polovica gospodarstev manjših, polovica pa večjih. Iz grafikona vidimo, da so ti odstotki za posamezne vrste živine zelo različni. 50% malih gospodarstev je imelo v letu 1947 v Sloveniji samo 15% od skupnega števila konj, 24% ovac, 32% govedi, 37% prašičev, 46% perutnine in 77% koz. Iz tega moremo sklepati, da imajo konje predvsem večja gospodarstva, da pa je koza tipično domača žival malih gospodarstev, da perutnino precej enako rede

mala in velika gospodarstva itd.

Struktura je prikazana za vsako vrsto živine z desetimi shematičnimi figurami, ki so pomaknjene glede na strukturo na levo ali na desno od črte, ki razmeji grupo manjših gospodarstev od grupe večjih gospodarstev.



Slika 6.10. Struktura števila živine v malih in velikih zemljiških gospodarstvih v Sloveniji v letu 1947. (Vir: Popis zemljiških gospodarstev v Sloveniji leta 1947).

STATISTIČNI KOEFICIENTI IN GOSTOTE

Statistični koeficienti

6.24 Strukturne odstotke dobimo, če primerjamo del s celoto. Primerjana podatka sta torej istovrstna, rezultat, ki ga dobimo, pa je neimenovano število; običajno ga izražamo v odstotkih. Smiselna pa je tudi primerjava raznovrstnih podatkov, če zadoščajo nekim osnovnim pogojem primerljivosti. Primerjana podatka morata biti namreč v vse-

binski zvezi in enako opredeljena. Če primerjamo število prebivalstva s skupno površino, pokaže ta primerjava gostoto prebivalstva. Primerjava proizvodnje s številom ur, ki so bile potrebne za izdelavo te proizvodnje, pokaže produktivnost dela itd. Nima pa smisla primerjati npr. vrednosti proizvodnje na nekem območju s številom tifusnih obolenj na istem območju, ker ta dva pojava nista v vsebinski zvezi. Promet v trgovini na drobno v Sloveniji v decembru 1968 pa je smiselno primerjati le s številom prebivalstva v SR Sloveniji in ne v SR Hrvatski in le s srednjim številom prebivalstva v decembru 1968 in ne za kako drugo razdobje. Le tako je koeficient smiseln. Osnovni pogoj, da moremo iz dveh podatkov izračunati statistični koeficient, je torej, da je primerjava umestna, primerjani podatki pa enako opredeljeni.

Smiselnost primerjave je dostikrat zvezana z opredelitvijo primerjanih populacij. Če izračunamo npr. koeficient nepismenih, ne vzamemo v primerjavo vsega prebivalstva, temveč se omejimo samo na prebivalstvo, staro nad deset let, da tako izključimo iz primerjave otroke. Enako ravnamo pri izračunavanju drugih pokazovalcev npr. pri izračunavanju porabe alkoholnih pijač na prebivalca itd.

Ker računamo koeficiente iz raznovrstnih podatkov tako, da primerjana raznovrstna podatka med seboj delimo, statistični koeficienti niso neimenovana števila, temveč imajo svojo enoto mere; ta je izpeljana iz enot mere primerjanih podatkov.

Če npr. primerjamo število prebivalstva v Sloveniji po popisu 31.3.1953 (1466425 prebivalcev) s površino Slovenije ob istem popisu (19992 km^2), dobimo gostoto prebivalstva v Sloveniji 31.3.1958

$$\text{gostota prebivalstva} = \frac{1466425 \text{ prebivalcev}}{19992 \text{ km}^2} = 73{,}4 \text{ prebivalcev}/\text{km}^2$$

Gostota prebivalstva v Sloveniji je torej bila 73,4 prebivalcev na kvadratni kilometer. Ta podatek je računska fikcija in pomeni, da bi bilo ob popisu na vsak kvadratni kilometer 73,4 prebivalcev, če bi bilo prebivalstvo enakomerno razporejeno po teritoriju Slovenije. Kljub temu pa ima ta koeficient svojo analitično vrednost, ker pokaže večjo ali manjšo naseljenost na določenem teritoriju.

Ob istem popisu je bila gostota prebivalstva v drugih republikah: v Srbiji 79,0, Hrvat-

ski 69,7, BiH 55,7, Makedoniji 50,7, Črni gori 30,4 prebivalcev/km^2. Vidimo, da je gostota prebivalstva po republikah zelo različna in da je bila najmanjša v Črni gori (30,4 preb./km^2) največja pa v Srbiji (79,0 preb./km^2).

6.25 Koeficient iz razmičnega in trenutnega podatka. Ker izhajata primerjana podatka, iz katerih izračunavamo koeficiente, iz različnih populacij, naletimo na težavo pri opredelitvi, če je ena od primerjanih populacij razmična, druga pa trenutna. Pogoj za izračun koeficienta je namreč, da sta oba podatka enako opredeljena. To pa na prvi pogled ni možno, če je ena populacija trenutna, druga pa razmična, ker je prva opredeljena s trenutkom, druga pa z razdobjem. Zato se zdi, da ne moremo primerjati števila umrlih (razmičen podatek) s številom prebivalstva (trenutni podatek), ker se število umrlih nanaša na določeno razdobje npr. leto, število prebivalstva pa moremo ugotoviti samo za določen trenutek. V takih primerih si pomagamo s tem, da trenuten podatek spremenimo v razmičen tako da izračunamo poprečje. Medtem ko ne moremo izraziti, kolikšno je število prebivalstva v določenem razdobju, moremo za to razdobje izračunati poprečno število prebivalstva.

Da moremo za določeno razdobje izračunati poprečje za trenutni podatek, je potrebno, da poznamo podatek vsaj za nekaj trenutkov v tem razdobju. Tako moremo izračunati poprečno število delavcev v nekem podjetju v letu 1968, če poznamo število delavcev v začetku ali v sredini posameznih mesecev v tem letu. Poprečno število prebivalstva v petletnem razdobju moremo izračunati, če poznamo število prebivalstva v začetku vsakega leta ali v sredini leta za vsa leta petletja. Poprečno zalogo v letu pa dobimo, če poznamo stanje zalog ob koncu vsakega četrtletja itd.

6.26 Poprečja iz trenutnih podatkov računamo glede na to, s kakšnimi podatki razpolagamo, na dva načina.

a) Če imamo razdobje, za katerega računamo poprečje, razdeljeno na r enakih razmikov (leto na 12 mesecev, četrtletje na 3 mesece, mesec na 3 dekade, petletje na 5 let itd.) za vsak tak delen razmik pa dan trenuten podatek za sredino razmika (sredine mesecev, sredine dekad, sredine let itd.), izračunamo poprečje po obrazcu

$$\bar{X} = \frac{1}{r}(X_1 + X_2 + \ldots + X_r) \qquad (6.13)$$

tako, da seštejemo vse podatke, vsoto pa delimo s številom osnovnih razmikov r .

Po tem obrazcu izračunamo npr. poprečno število delavcev, če poznamo število delavstva sredi mesecev, poprečno število prebivalstva v petletki, če poznamo število prebivalstva sredi posameznih let itd.

b) Če pa je razdobje, za katerega računamo poprečje, razdeljeno na r enakih razmikov, razpolagamo pa s podatki za začetke oziroma konce teh razmikov, izračunamo poprečje po obrazcu

$$\bar{X} = \frac{1}{r}\left(\frac{1}{2} X_o + X_1 + \ldots + X_{r-1} + \frac{1}{2} X_r\right) \qquad (6.14)$$

tako da vsoto polovičnih vrednosti podatkov na začetku in koncu celotnega razmika (X_o in X_r) in celih vrednosti ostalih podatkov na začetkih ali koncih delnih razmikov delimo s številom osnovnih razmikov r.

Če imamo npr. število delavcev na začetku oziroma na koncu posameznih mesecev, izračunati pa je treba poprečno število delavstva v letu, vzamemo polovico števila delavcev 1. januarja, število delavstva 1. februarja, 1. marca, 1. aprila itd. do 1. decembra in polovično število delavstva 1. januarja naslednjega leta. Vsoto vseh teh vrednosti pa delimo z 12.

Če vzamemo, da je razdobje, za katerega hočemo izračunati poprečje, sestavljeno iz enega samega osnovnega razmika (r = 1), dobimo po obrazcu 6.13, da je $\bar{X} = X_1$ Iz tega sledi, da moremo poprečne nadomestiti s stanjem v sredini razdobja. Tako npr. nadomeščamo poprečno letno število prebivalstva s "srednjim številom prebivalstva", to je s številom prebivalstva sredi leta (30. junija).

Če je r = 1, obrazec 6.14 degenerira v

$$\bar{X} = \left(\frac{1}{2} X_o + \frac{1}{2} X_1\right) = \frac{1}{2}(X_o + X_1)$$

To pomeni, da moremo poprečje nadomestiti s polovično vrednostjo med začetnim in končnim stanjem v osnovnem razdobju.

6.27 Ko imamo izračunano poprečje iz trenutnih podatkov, izračunamo koeficient po obrazcu

$$K = \frac{Y \cdot E}{\bar{X} \cdot i} \qquad (6.15)$$

pri čemer pomeni: K = koeficient, Y = razmični podatek, $\bar{X}$ = poprečje za trenutni podatek, i = dolžina časovnega razdobja, za katerega računamo koeficient, E = 100, 1000, 10000, odvisno od tega, na koliko enot trenutnega podatka se nanaša koeficient.

6.28 Ker je npr. koeficient natalitete število rojenih v enem letu na 1000 prebivalcev, izračunamo koeficiente natalitete za SFRJ v razdobju 1951-1955 takole:

Poprečno število prebivalstva v tem razdobju je $\bar{X}$ = 17067000, število rojenih Y = 2395816, razdobje i = 5 let, koeficient pa je računan na E = 1000 prebivalcev. Iz teh podatkov dobimo:

$$K = \frac{2395816.1000}{17067000.5} = 28,1 \text{ rojstev na leto na 1000 prebivalcev.}$$

Podobno računamo vse druge koeficiente, ki imajo za osnovo primerjavo razmičnega podatka s trenutnim.

Po obrazcu 6.15 računamo koeficiente tudi za razdobja, ki so krajša od enotinega razdobja. Tako npr. koeficient mortalitete, ki je po definiciji število umrlih v letu na 1000 prebivalcev, za Slovenijo v januarju 1958 izračunano takole: Ker je bilo srednje število prebivalstva v januarju 1958 $\bar{X}$ = 1556578, število umrlih v januarju 1958 Y = 1464, E = 1000, i = 31/365, je koeficient mortalitete enak

$$K = \frac{1464 \cdot 1000}{1556578.31/365} = 11,1 \text{ umrlih na leto na 1000 prebivalcev.}$$

To pomeni: Če bi bila v Sloveniji leto dni umrljivost taka kakor v januarju 1958, bi v enem letu umrlo na 1000 prebivalcev 11,1 oseb.

6.29 Koeficient obračanja zalog pove, kolikokrat se zaloga blaga ali surovin obrne v enoti časa. Računamo ga tako, da primerjamo promet določenega blaga ali

porabo surovin z ustreznimi poprečnimi zalogami.

Iz mesečne industrijske službe ZS SRS imamo za podjetje A iz nekovinske stroke podatke o porabi in zalogah kremenovega peska v prvem polletju 1958 po mesecih.

Tabela 6.12. Mesečna zaloga in poraba kremenovega peska v podjetju A iz nekovinske stroke v prvem polletju 1958 (Vir: ZS SRS)

Mesec		Zaloga konec meseca v t	Poraba v t
1957	D	664	
1958	J	774	494
	F	888	489
	M	789	562
	A	546	468
	M	568	465
	J	726	544

Izračunati je treba koeficient obračanja zalog za kremenov pesek v prvem polletju 1958, izražen s številom obratov zalog v enem mesecu.

Po postopku za računanje koeficientov iz podatkov v tabeli 6.12 izračunamo skupno porabo kremenovega peska v prvem polletju 1958 in poprečno zalogo v tem razdobju. Skupna poraba v polletju je vsota mesečne porabe za ustreznih šest mesecev.

$$Y = 494 + 489 + 562 + 468 + 465 + 544 = 3022 \text{ t}$$

Poprečno zalogo $\bar{X}$ pa izračunamo po obrazcu 6.14, ker imamo podatke za konce mesecev.

$$\bar{X} = \frac{\frac{664}{2} + 774 + 888 + 789 + 546 + 568 + \frac{726}{2}}{6} = 710 \text{ t}$$

Poprečna zaloga je torej 710 t.

Ker je koeficient merjen s številom obratov v enem mesecu, naše razdobje pa obsega šest mesecev, je $i = 6$. Iz teh podatkov izračunamo, da je koeficient obračanja za-

log za kremenov pesek v prvem polletju 1958 enak

$$K = \frac{Y \cdot E}{\bar{X} \cdot i} = \frac{3022 \cdot 1}{710 \cdot 6} = 0,71$$

Koeficient obračanja zalog K = 0,71. To pomeni, da zaloge kremenovega peska v podjetju A napravijo mesečno 0,71 obrata.

S podobnim računom smo dobili koeficiente obračanja zalog za isto podjetje za rjavi premog K = 12,0, pepeliko K = 0,50, kalcinirano sodo K = 2,78. Kakor vidimo, so koeficienti obračanja zalog, ki so tudi operativno zelo pomemben pokazovalec za posamezne surovine in goriva, zelo različni.

Obračanje zalog pa moremo izraziti tudi z dolžino enega obrata. Ta koeficient, ki je recipročni pokazovalec za gornji pokazovalec, izračunamo po obrazcu

$$K = \frac{\bar{X} \cdot i}{Y} \qquad (6.8)$$

Če dolžino enega obrata izrazimo v delovnih dneh, je za gornji zgled i = 181, koeficient obračanja zalog za kremenov pesek v podjetju A, izražen z dolžino enega obrata pa je enak

$$K = \frac{710 \cdot 181}{3022} = 42,5 \text{ dni}$$

6.30 Po navedenem postopku računamo še niz drugih koeficientov iz ekonomskih statistik. Tako je npr. proizvodnja na enega delavca pokazovalec o produktivnosti dela, proizvodnja določenih proizvodov na prebivalca npr. jekla, pokazovalec industrializacije, poraba določenega proizvoda na prebivalca pokazovalec standarda, razmerje med prometom in zalogami pokazovalec obračanja zalog, število prodanih knjig na leto na 1000 prebivalcev pokazovalec za kulturno raven itd.

6.31 Kot zgled za kompleksno analizo s koeficienti vzemimo trgovino na drobno v letu 1957 po republikah. V tabeli 6.11 imamo osnovne podatke, iz katerih bomo izračunali koeficiente.

Tabela 6.11 Trgovina na drobno v SFRJ v letu 1957 po republikah (Vir: SG-59)

Pokazovalec	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Število prebivalstva 30.VI.1957 v 10^3	18234	7448	4109	1582	3168	1454	473
Število trgovin na drobno:							
31.XII.1956	36257	14017	9296	4662	4634	2795	853
31.XII.1957	37651	14601	9798	4722	4822	2810	898
Število zaposlenih v trgovini na drobno							
31.XII.1956	112511	45954	27589	13689	14184	8523	2572
31.XII.1957	116119	46187	29292	14021	14566	9141	2912
Promet v letu 1957 v 10^6							
Skupno	750776	291478	198339	110259	90123	44331	16246
od tega:							
Živila	232637	75604	64414	37519	32744	15348	7008
Industrijski proizvodi	483608	202410	125065	68917	51662	27133	8421

Iz tabele 6.11 vidimo, da so prvi trije podatki trenutni, drugi pa razmični. Ker imamo srednje število prebivalstva, ta podatek vzamemo za poprečno število prebivalstva. Za število trgovin in število zaposlenih pa imamo stanje na začetku in na koncu leta 1957. Zaradi tega najprej izračunamo poprečje, da bi mogli število trgovin in število zaposlenih primerjati s prometom, ki je razmičnega značaja.

Za število trgovin v SFRJ je npr. iz gornjih podatkov poprečno število trgovin enako polovici vsote števila trgovin po stanju 31.XII.1956 in 31.XII.1957

$$\bar{X} = \frac{1}{2}(36257 + 37651) = 36954$$

Podobno izračunamo vsa druga poprečja. Tako dobimo vrsti za poprečno število trgovin in za zaposleno osebje v trgovini na drobno po republikah.

Tabela 6.12. Poprečno število trgovin in zaposlenega osebja v trgovini na drobno v letu 1957 po republikah

Poprečno število	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
trgovin	36954	14309	9547	4692	4728	2802	876
zaposlenih	114315	46071	28440	13855	14375	8832	2742

Iz podatkov v tabeli 6.11 in 6.12 izračunamo niz koeficientov, ki osvetljujejo trgovino na drobno v SFRJ. Tako moremo s številom prebivalstva primerjati vse druge podatke. Število prebivalstva na eno trgovino meri gostoto trgovinske mreže, število prebivalstva na enega zaposlenega preskrbljenost s trgovskim kadrom itd. Ne da bi se spuščali v globljo analizo prometa na prebivalca, ki je bistveno odvisen razen od kupne moči tudi od drugih faktorjev, npr. od socialne strukture prebivalstva, merimo s temi koeficienti kupno moč prebivalstva.

Primerjava števila trgovin s številom zaposlenih ali s skupnim prometom prikaže velikost trgovinskih obratov. Primerjava skupnega števila zaposlenih s skupnim prometom pokaže produktivnost dela v trgovini na drobno, razmerje med prometom z industrijskimi proizvodi in živili pa da vpogled v strukturo prometa, ki je odvisna od socialne sestave prebivalstva, življenjske ravni itd.

Vprašanje je, ali je možno vse te koeficiente izračunati iz gornjih podatkov.

Primerjava števila trgovin in zaposlenega osebja s številom prebivalstva iz osnovnih podatkov ni možna, ker imamo število prebivalstva na dan 30.VI.1957, podatke o številu trgovin in številu zaposlenih pa na dan 31.XII.1956 in 31.XII.1957. Primerjani trenutni populaciji nista časovno enako opredeljeni. Ker pa moremo imeti srednje število prebivalstva za poprečje v letu 1957, smo upravičeni, da primerjamo srednje število prebivalstva s poprečnim številom trgovin in zaposlenih v letu 1957, ker se vsi ti podatki nanašajo na leto 1957. Število zaposlenih moremo primerjati s številom trgovin po stanju 31.XII.1956 ali 31.XII.1957, ker imamo za te datume podatke o številu trgovin in številu zaposlenih. Moremo pa primerjati tudi poprečno število zaposlenih s poprečnim številom trgovin, ker veljata oba podatka za isto razdobje. V tabeli 6.13 so izračunani vsi nakazani koeficienti.

Tabela 6.13 Koeficienti trgovine na malo v SFRJ v letu 1957 po republikah (Vir: SG-58).

Pokazovalec	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Število prebivalcev na eno trgovino	493	520	431	337	670	519	539
Število prebivalcev na enega zaposlenega	159	165	144	114	221	165	172
Skupni promet na enega prebivalca v tisoč din	41,2	39,1	48,3	69,7	28,4	30,5	34,2
Promet z živili na prebivalca v tisoč din	12,8	10,2	15,7	23,7	10,3	10,5	14,8
Promet z industrijskimi proizvodi na prebivalca v tisoč din	26,5	27,1	30,4	43,6	16,3	18,6	17,8
Število zaposlenih na eno trgovino (poprečno)	3,09	3,22	2,98	2,95	3,04	3,15	3,11
Skupni promet na eno trgovino v mlj. din	20,3	20,4	20,8	23,5	19,0	15,8	18,5
Skupni promet na enega zaposlenega v mlj. din	6570	6320	6970	7970	6260	5020	5930
Promet z industrijskimi proizvodi na sto din prometa z živili	207	267	194	184	158	176	120

Tabela koeficientov v trgovini na drobno ilustrativno pokaže razlike v trgovini na drobno po republikah. Najmanj prebivalcev pride na eno trgovino v Sloveniji (337) , največ pa v BiH (670). Enako velja za število prebivalstva, ki pride na enega zaposlenega (Slovenija 114, BiH 221). Ti podatki kažejo gostoto trgovinske mreže. Tudi skupni promet na enega prebivalca kaže velike razlike med republikami (od 28,4 tisoč dinarjev

za BiH do 69,7 za Slovenijo). Vendar ta koeficient nima polne analitične veljave, ker je njegova vrednost odvisna od velikega števila faktorjev (socialne strukture prebivalstva, kupne moči itd.). Promet z živili na prebivalca je precej odvisen od odstotka nekmečkega prebivalstva, promet z industrijskimi proizvodi na prebivalca pa moremo imeti bolj za pokazovalec o kupni moči in standardu prebivalstva. Tudi ta koeficient kaže velike razlike med republikami. Medtem ko je za BiH, Črno goro in Makedonijo izredno nizek, je za druge republike, posebno za Slovenijo, znatno večji.

Velikost trgovinskih obratov, ki je razvidna iz poprečnega števila zaposlenih na eno trgovino, med republikami malo variira. Iz tega sklepamo, da je tip trgovin po republikah precej enoten. Enako težnjo kaže tudi skupen promet na eno trgovino in so znatnejše razlike edino le za Slovenijo in Makedonijo. Promet na enega zaposlenega, ki ga štejemo za pokazovalec produktivnosti dela v trgovini, kaže znatno odklonitev navzgor za Slovenijo in znatno odklonitev navzdol za Makedonijo. Razmerje med prometom z industrijskimi proizvodi in prometom z živili kaže med republikami velike razlike. Največji koeficient dobimo za Srbijo (267 din na 100 dinarjev prometa z živili), najmanjši pa za Črno goro (120 din na 100 din prometa z živili). Razlike izvirajo iz različne sestave kmečkega in nekmečkega prebivalstva in različne kupne moči prebivalstva. Zato koeficient nima polne analitične vrednosti.

Iz zgleda iz trgovinske statistike vidimo, kako so statistični koeficienti uporabni za kompleksno analizo statističnih podatkov. Podobno moremo analizirati podatke za najrazličnejša druga socialno-ekonomska področja.

6.32 Recipročni koeficienti. Medtem ko računamo strukturne odstotke vedno tako, da del primerjamo s celoto, ne pa obratno, so smiselni tudi recipročni statistični koeficienti. Tako moremo pokazovalec za gostoto trgovinske mreže, ki je dan s številom prebivalstva na eno trgovino, izraziti tudi z recipročnim koeficientom številom trgovin na 1000 prebivalcev. Medtem ko pri prvotnem koeficientu primerjamo število prebivalstva s številom trgovin, z drugim primerjamo število trgovin s številom prebivalstva. Koeficient 493 prebivalcev na eno trgovino v SFRJ moremo torej izraziti tudi z 2,15 trgovin na 1000 prebivalcev. Kateri izmed obeh možnih koeficientov, ki jih moremo izračunati v vsakem primeru, je boljši, je odvisno od namena

izračuna in predstave koeficienta. V mnogo primerih sta oba koeficienta enako uporabljiva in vsak po svoje pokaže isto značilnost pojava. Tako merimo produktivnost dela s količino, proizvedeno v enoti časa, ali s časom, v katerem je bila proizvedena enota proizvodnje. Obračanje zalog merimo s številom obratov zalog v enoti časa ali s časom, v katerem so se zaloge enkrat obrnile. Obremenjenost učiteljstva merimo s številom učencev na enega učitelja ali s številom učiteljev na 100 ali 1000 učencev itd.

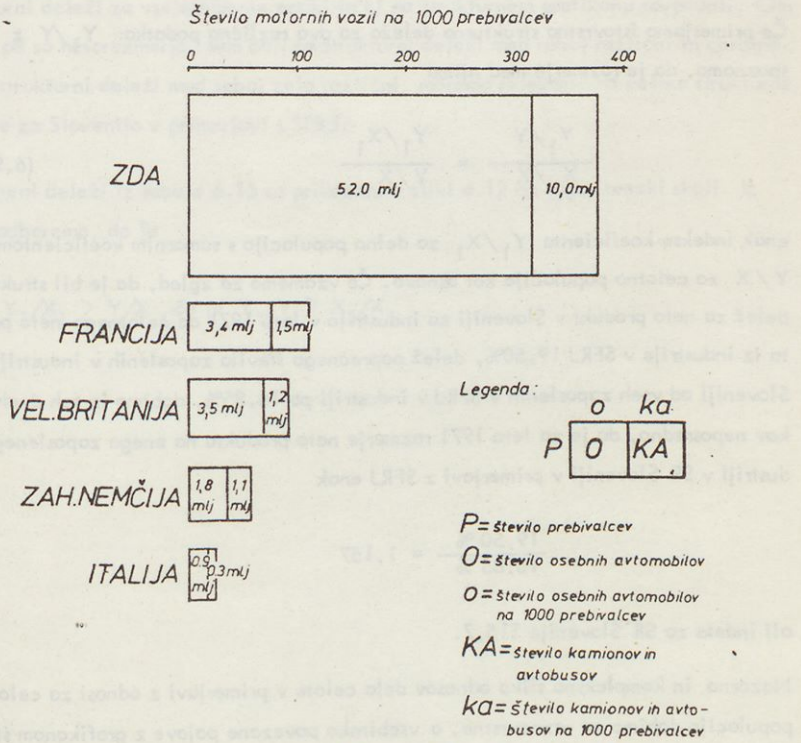
Grafično prikazovanje statističnih koeficientov

6.33 Pravokotniki. Za koeficiente nimamo nekih posebnih metod za grafično prikazovanje kakor za strukture. Vrste koeficientov prikazujemo z običajnimi metodami grafičnega prikazovanja: s stolpci, linijskimi grafikoni, figurami itd. Vendar imamo nekaj specifičnih metod, ki jih moremo s pridom uporabiti ravno za prikazovanje koeficientov. Tako so za grafično prikazovanje koeficientov posebno primerni pravokotniki, ker imata stranici in ploščina pravokotnika podobno algebrajsko zvezo kakor koeficient in oba podatka, iz katerih je koeficient izračunan. Za pravokotnik velja $p = a.b$. Pri tem pomeni p ploščino, a in b pa stranici, med K, X in Y pa velja podobna zveza $K = Y/X$ ali $Y = K.X$. Stranici pravokotnika moremo uporabiti za prikaz koeficienta in enega absolutnega podatka, ploščina pravokotnika pa je zaradi lastnosti pravokotnika sorazmerna absolutni velikosti drugega podatka - Y.

6.34 Uporabnost teh zvez bomo videli na praktičnem zgledu. Za pet zapadnih držav imamo v tabeli 6.14 podatke o številu prebivalstva in številu osebnih in tovornih avtomobilov in avtobusov. Iz teh podatkov moremo za vsako državo posebej izračunati koeficiente o številu motornih vozil na 1000 prebivalcev - koeficiente motorizacije. Ti podatki so po zgornjem načinu prikazani v sliki 6.11.

Tabela 6.14. Število prebivalstva in število motornih vozil za pet zapadnih držav v letu 1955. (Vir: SG-58)

Država	Število prebivalstva v tisočih	Število motornih vozil v tisočih			Število motornih vozil na tisoč prebivalcev		
		Skupno	osebnih	kamionov avtobusov	Skupno	osebnih	kamionov avtobusov
ZDA	165271	62020	51989	10031	375	315	60
Francija	43274	4820	3350	1470	111	77	34
Vel.Britanija	51215	4732	3550	1182	92	69	23
Zap.Nemčija	49995	2909	1813	1096	58	36	22
Italija	48016	1210	879	341	25	18	7



Slika 6.11. Motorizacija petih zapadnih držav v letu 1955.

V grafikonu nazorno primerjamo tri vrste podatkov hkrati. Ker je najvažnejša medržavna primerjava o številu motornih vozil na prebivalca, so pravokotniki postavljeni tako, da je primerjava teh podatkov najboljša. Razen tega je iz slike razvidno število prebivalstva in čeprav s ploščino tudi skupno število motornih vozil. Pravokotniki so razdeljeni še v sorazmerju s strukturo po vrsti motornega vozila.

Ta način moremo uporabiti za prikazovanje statističnih vrst za koeficiente iz najrazličnejših podatkov.

6.35 Grafikoni strukturnih deležev. Posredno moremo koeficiente odbrati tudi iz grafikona strukturnih vrst. Ta metoda dá kompleksno sliko odnosov med pojavi, ki jih proučujemo. Zaradi tega je tak grafikon predvsem sredstvo za prikazovanje relativnih odnosov med raznovrstnimi podatki, čeprav prikazuje v osnovi strukture.

Če primerjamo istovrstna strukturna deleža za dva različna podatka: Y_1/Y z X_1/X, spoznamo, da je razmerje med njima

$$\frac{Y_1/Y}{X_1/X} = \frac{Y_1/X_1}{Y/X} \qquad (6.9)$$

enak indeksu koeficienta Y_1/X_1 za delno populacijo s sumarnim koeficientom Y/X za celotno populacijo kot osnovo. Če vzamemo za zgled, da je bil strukturni delež za neto produkt v Sloveniji za industrijo v letu 1971 od celotnega neto produkta iz industrije v SFRJ 19,50%, delež poprečnega števila zaposlenih v industriji v Sloveniji od vseh zaposlenih v SFRJ v industriji pa 16,85%, dobimo iz teh dveh podatkov neposredno, da je za leto 1971 razmerje neto produkta na enega zaposlenega v industriji v SR Sloveniji v primerjavi z SFRJ enak

$$\frac{19,50\,\%}{16,85\,\%} = 1,157$$

ali indeks za SR Slovenijo 115,7.

Nazorno in kompleksno sliko odnosov dela celote v primerjavi z odnosi za celotno populacijo dobimo za raznovrstne, a vsebinsko povezane pojave z grafikonom strukturnih deležev. Čeprav so podatki po pravilu raznovrstni, jih v tem grafikonu narišemo brez težav v isti grafikon, ker so prikazani z neimenovanimi števili - strukturni-

mi deleži. Da so vidna relativna razmerja, uporabimo za njihov prikaz logaritemske skale. Vrsto strukturnih deležev za isto podpopulacijo moremo risati bodisi z vrstami stolpcev, ali še nazorneje, v grafikonu, v katerem ob logaritmični skali strukturnih deležev na ustreznih mestih vnesemo strukturne deleže s črtico, puščico itd.

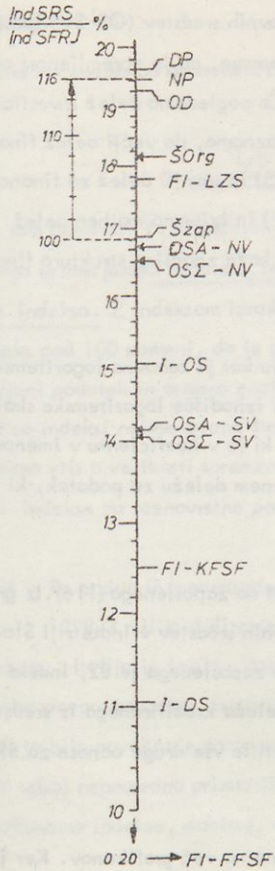
6.36 Za zgled vzemimo osnovne gospodarske pokazovalce za industrijo v Sloveniji v primerjavi s SFRJ v letu 1970. V tabeli 6.15 so po elementih podani za SR Slovenijo in SFRJ glavni absolutni podatki, v zadnjem stolpcu pa strukturni odstotki za SR Slovenijo. Če bi bila razmerja med posameznimi elementi npr. med neto produktom in številom zaposlenih, ki pokaže produktivnost, med sedanjo in nabavno vrednostjo osnovnih sredstev, ki pokaže izrabljenost osnovnih sredstev, med osebnimi dohodki in številom zaposlenih, ki pokaže raven poprečnih osebnih dohodkov; med viri finansiranja in skupnimi investicijami za SR Slovenijo, enaka kot za SFRJ, bi bili strukturni deleži za vse elemente enaki in bi na strukturnem grafikonu sovpadali. Čim večja pa so nesorazmerja, tem bolj so strukturni deleži med seboj različni in obratno, če so strukturni deleži med seboj zelo različni, moremo sklepati na velike strukturne razlike za Slovenijo v primerjavi s SFRJ.

Strukturni deleži iz tabele 6.15 so prikazani v sliki 6.12 ob logaritemski skali. Iz slike odberemo, da je

$$Y_1/X_1 > Y/X \text{ , če je } Y_1/Y > X_1/X$$

Tabela 6.15 Primerjava strukturnih deležev za industrijo SR Slovenije v industriji SFRJ za leto 1970. (Vir: SG-1972)

Pokazovalec	v milj. din		SRS / SFRJ %
	SRS	SFRJ	
Število industrijskih organizacij	426	2374	18,15
Število zaposlenih	244636	1451745	16,85
Osnovna sredstva 31.12.1970			
Nabavna vrednost	18150	110450	16,43
Sedanja vrednost	9437	67136	14,06
Aktivna osnovna sredstva 31.12.1970			
Nabavna vrednost	17091	102385	16,69
Sedanja vrednost	8521	60192	14,16
Družbeni proizvod	10848	55285	19,62
Neto produkt	9388	48147	19,50
Osebni dohodki (neto)	4076	21154	19,27
Investicije v osnovna sredstva	2254	14985	15,04
Investicije v družbeni standard	82	741	11,05
Vir financiranja investicij			
Lastna in združena sredstva	1314	7396	17,77
Kreditirano iz sredstev in fondov federacije	938	7496	12,51
Financirano iz sredstev in fondov federacije	2	93	2,15



DP	= družbeni produkt
NP	= neto produkt
OD	= osebni dohodek (neto)
ŠOrg	= število organizacij
Šzap	= število zaposlenih
OS	= osnovna sredstva - skupno
OSA	= osnovna sredstva - aktivna
NV	= nabavna vrednost
SV	= sedanja vrednost
I-OS	= investicije v osnovna sredstva
I-DS	= investicije v družbeni standard
FI-LZS	= financiranje investicij: iz lastnih in združenih sredstev
FI-KFSF	= financiranje investicij s krediti sredstev in fondi federacije
FI-FFSF	= financiranje investicij: financirano s sredstvi in fondi federacije

Slika 6.12. Strukturni grafikon za odnose v industriji SR Slovenije v letu 1970

Iz te zveze moremo iz grafikona neposredno odčitati: družbeni produkt (DP) in neto produkt (NP) na osebni dohodek (OD) je v SR Sloveniji malenkostno večji kot v poprečju za SFRJ. Obenem pa je osebni dohodek (OD) na zaposlenega (Z) v industriji precej večji v Sloveniji kot je jugoslovansko poprečje. Zaradi prvih dveh zvez je to izraz višje produktivnosti dela. Ker je strukturni delež za skupna (OS Σ -SV) in aktivna osnovna sredstva (OSA-SV) po sedanji vrednosti znatno nižji kot po nabavni vrednosti (OS Σ -NV) (OSA-NV), moremo računati, da je izrabljenost osnovnih sredstev v Sloveniji znatno

večja kot v poprečju za vso državo. Ker so deleži osnovnih sredstev (OS-SV) (posebno sedanja vrednost) nižja od številnih zaposlenih (Z) sklepamo, da je opremljenost oziroma sestav v Sloveniji pod jugoslovanskim poprečjem. Če pogledamo delež investicij v osnovna sredstva (I-OS) in strukturo financiranja, spoznamo, da večji delež financiranja investicij iz lastnih in združenih sredstev (FI-LZS) in manjši delež za financiranje iz kreditov sredstev in fondov federacije (FI-KFSF) in izjemno majhen delež za financiranje iz federalnih sredstev in fondov (FI-FFSF) kaže značilno strukturo financiranja investicij v industriji v SR Sloveniji.

S premakljivo logaritemsko skalo, ki je v enakem merilu kot je osnovna logaritemska skala, bi mogli nakazane odnose izraziti z indeksi. Če izhodišče logaritemske skale (1 ali 100) naravnamo v grafikonu na strukturni delež, ki je v koeficientu v imenovalcu, odberemo indeks ustreznega koeficienta ob strukturnem deležu za podatek, ki je v koeficientu v številcu.

V grafikonu je za zgled nakazan indeks za neto produkt na zaposlenega (116). Iz grafikona odberemo podobno, da je indeks izrabljenosti osnovnih sredstev v industriji Slovenije 118 v primerjavi s SFRJ. Indeks osnovnih sredstev na zaposlenega je 82, indeks deleža financiranja investicij iz lastnih sredstev je 118, deleža kreditiranega iz sredstev in fondov federacije pa 83. Podobno bi mogli analizirati še vse druge odnose za SR Slovenijo v primerjavi s SFRJ.

Uporabnost grafikona moremo razširiti iz enega grafikona na več grafikonov. Ker je en grafikon za en del narisan ob eni premici, moremo nazorno prikazati odnose za vse republike hkrati za eno panogo ali za različne panoge po republikah. Časovna vrsta strukturnih grafikonov pokaže spreminjanje v indeksih odnosov ipd.

ENOSTAVNI INDEKSI

6.37 O indeksih govorimo, kadar z relativnimi števili primerjamo istovrstne, prirejene podatke. Strukture in koeficiente računamo iz absolutnih podatkov. Indekse pa računamo iz vseh vrst statističnih podatkov, ne pa samo iz absolutnih. Tako moremo z indeksi primerjati med seboj različne koeficiente, strukturne odstotke in druge izvede-

ne pokazovalce.

Indeks računamo po osnovnem obrazcu

$$I_{1/0} = 100 \cdot Y_1/Y_o \qquad (6.10)$$

Pri tem pomeni: Y_1 = podatek, ki ga primerjamo s podatkom Y_o. Y_o = podatek, na katerega primerjamo. Podatek, na katerega primerjamo, imenujemo b a z o ali o s n o v o indeksa. Z indeksom izražamo tekoči podatek v stotinkah ali poenih od osnove Y_o. Indeks pod 100 pomeni, da je pojav manjši od osnove, indeks 100 pomeni, da sta primerjani podatek in osnova enaka. Indeks pa je nad 100, če je podatek večji od osnove. Ker so indeksi neimenovana števila in imajo enoten merski sistem, iz njih zelo nazorno dobimo vtis o velikosti sprememb oziroma razlik. Razen tega moremo med seboj primerjati indekse za raznovrstne podatke, ki drugače ne bi bili neposredno primerljivi.

6.38 Po podatkih mednarodnega pregleda v SG 1958 je imela Jugoslavija v letu 1955 1298,3 milijard dinarjev narodnega dogodka, v letu 1956 pa 1444,1 milijard dinarjev. Italija je imela v letu 1955 10814 milijard lir, v letu 1956 pa 11504 milijard lir narodnega dohodka. Iz absolutnih podatkov si ne moremo ustvariti slike, v kateri državi je bilo povečanje narodnega dohodka večje. Podatki za Italijo in Jugoslavijo niso med seboj neposredno primerljivi, ker so eni dani v dinarjih, drugi pa v lirah. Če pa izračunamo indekse, dobimo, da je za Jugoslavijo indeks enak:

$$I_{56/55} = 100 \cdot Y_{56}/Y_{55} = \frac{100.1444,1}{1298,3} = 111,2$$

za Italijo pa

$$I_{56/55} = \frac{100.11504}{10814} = 106,4$$

Indeksa pa sta primerljiva in sicer kažeta, da je bil porast narodnega dohodka relativno večji v Jugoslaviji kakor v Italiji.

Če so spremembe, ki jih izražamo z indeksi, velike, indekse računamo na cela števila, na eno decimalko pa jih računamo le, če prikazujemo majhne spremembe. Praviloma pa ne računamo indeksov na več decimalk, ker tako indeks izgubi svojo osnovno kvaliteto - nazornost.

Stvarni in krajevni indeksi

6.39 Ker se podatka, ki smo ju primerjali med seboj v zgornjem zgledu, razlikujeta v času (leto 1955 in leto 1956), imenujemo take indekse časovne indekse. Primerjana podatka pa se moreta razlikovati tudi v krajevnem ali stvarnem znaku. V tem primeru računamo krajevne in stvarne indekse. Običajno računamo indekse za cele vrste podatkov. V takih primerih vzamemo običajno za bazo vseh indeksov en in isti člen.

6.40 Kot zgled za vrsto krajevnih indeksov vzemimo poprečno pogodbeno prodajno ceno za m^2 stanovanjske površine za glavna mesta republik in pokrajin v letu 1971.

Tabela 6.16. Poprečna pogodbena prodajna cena za 1 m^2 (Vir: SG-72)

Mesto	Poprečna pogodbena prodajna cena za 1 m^2 v letu 1971	
	din	indeks
Beograd	2903	84
Ljubljana	3474	100
Sarajevo	3145	91
Skopje	2050	59
Titograd	2422	70
Zagreb	3877	112

Za osnovo vzamemo člen, za katerega najbolje poznamo razmere. Ker kot Slovenci najbolje poznamo razmere v Ljubljani, zato vzamemo za osnovo primerjave Ljubljano in nanjo izračunamo vse indekse. Iz tabele 6.16 vidimo, da je stanovanjska površina le v Zagrebu višja od Ljubljane, da pa je indeks za Skopje najnižji (59). Pri proučitvi teh podatkov pa moramo upoštevati ne le razlike v stroških temveč tudi v stanovanjskem standardu.

6.41 Za zgled indeksov iz vrste za stvarni znak vzemimo poprečne mesečne osebne dohodke v gospodarskih in negospodarskih dejavnostih v letu 1969 po stopnji strokovnosti.

Tabela 6.17 Poprečni neto osebni dohodki po stopnji strokovnosti v gospodarstvu in negospodarstvu (Ø 1969) (Vir: SG-72)

Stopnja strokovnosti	Gospodarstvo	Negospodarstvo	Indeks Negospodarstvo gospodarstvo
Skupno	995	1211	121,7
Visoka strokovna izobrazba	1943	1963	101,0
Višja strokovna izobrazba	1506	1269	84,3
Srednja strokovna izobrazba	1122	1106	98,6
Nižja strokovna izobrazba	881	856	97,2
Visokokvalificiran delavec	1278	1246	97,5
Kvalificiran delavec	930	998	107,3
Polkvalificiran delavec	784	703	89,7
Nekvalificiran delavec	720	641	89,0

Indeksi nazorneje kakor pa absolutni podatki v enem številu pokažejo razmerje med gospodarstvom in negospodarstvom. Indeks za kvalificiranega delavca pokaže za 7.3 poena višjo raven v negospodarstvu v razmerju z gospodarstvom. Enako je indeks večji od 100 za zaposlene z visoko strokovno izobrazbo, kar je razumljivo. Za vse druge vrste delavcev pa so indeksi manjši kot 100 in sicer so največje razlike pri zaposlenih z višjo strokovno izobrazbo. Zaradi različne sestave skupno zaposlenih v gospodarstvu v primerjavi z negospodarstvom pa je indeks izredno velik 121,7.

Časovni indeksi

6.42 Indeksi s stalno osnovo. Največkrat računamo indekse za časovne vrste in z njimi proučujemo dinamiko pojavov. Časovna vrsta osnovnih podatkov

sicer že sama prikazuje dinamiko, vendar ji razmeroma težko sledimo, ker so posamezni pojavi različno veliki, imajo različne enote mere itd. Preračunanje take vrste v indeksno vrsto s pravilno izbrano bazo pa pokaže spremembe od osnove, ki je v vsakem primeru 100, v stalnem merilu - stotinkah.

6.43 Za zgled časovnih indeksov vzemimo časovne vrste v razdobju 1965-71 za nekaj pomembnih podatkov za SFRJ.

Tabela 6.18 Razvoj za nekaj pojavov za SFRJ v razdobju 1965-71 (Vir: SG-72).

Leto	ŠP	ZDS	ND	EE	PR	NF	S
1965	19434	3583	84476	15523	29957	2063	1769
1966	19644	3491	91733	17174	29292	2222	1867
1967	19840	3466	94014	18702	26467	2374	1832
1968	20029	3487	97692	20641	26732	2494	1924
1969	20209	3622	107856	23375	26496	2699	2047
1970	20371	3765	114269	26023	28422	2854	2078
1971	20554	3944	123598	29509	30902	2961	2206

ŠP = število prebivalcev v 10^3

ZDS = zaposleni v družbenem sektorju v 10^3

ND = narodni dohodek v 10^6 po stalnih cenah v letu 1966

EE = skupna proizvodnja elektro-energije v 10^6 KWh

PR = skupna proizvodnja premoga v 10^3 ton

NF = proizvodnja surove nafte v 10^3 ton

S = surovo jeklo SM in EL v 10^3 ton

Tako enote mere kot ravni, na katerih so posamezni pojavi, otežujejo, če že ne onemogočajo primerjavo v dinamiki med posameznimi pojavi, zato izračunajmo vrste časovnih indeksov in vzemimo leto 1965 (prvo leto gospodarske reforme) za osnovo.

V tabeli 6.19 so iz absolutnih podatkov izračunane vrste indeksov. Ti indeksi so med seboj primerljivi, ker so vsi izračunani na isto osnovo in so neimenovana števila. Iz njih sklepamo, kako je npr. rastel narodni dohodek znatno hitreje kakor število prebivalstva, da je od indeksov proizvodnje energetskih sredstev proizvodnja premoga stag-

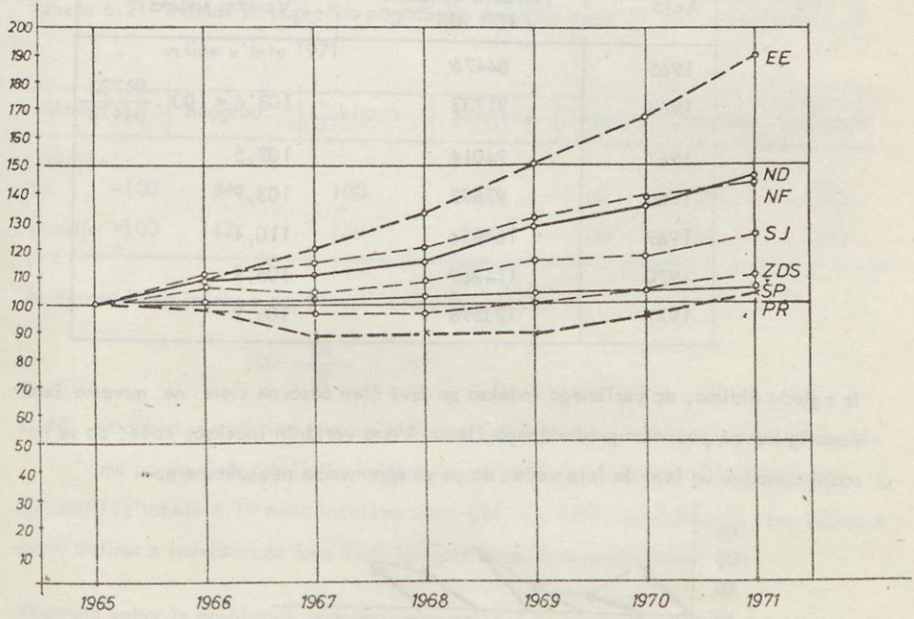
Tabela 6.19. Indeksi razvoja za nekaj pojavov za SFRJ v razdobju 1965-71

Leto	ŠP	ZDS	ND	EE	PR	NF	SJ
1965	100	100	100	100	100	100	100
1966	101	97	109	111	98	108	106
1967	102	97	111	120	88	115	104
1968	103	97	116	133	89	121	109
1969	104	101	128	151	88	131	116
1970	105	105	135	168	95	138	117
1971	106	110	146	190	103	144	125

nirala, medtem ko kažejo indeksi za elektroenergijo močan vzpon.

Indeksne vrste iz tabele 6.19 so še nazorneje prikazane v sliki 6.13.

6.44 Indeksi s premično osnovo. V gornjem primeru smo za posamezne pojave primerjali podatek za vsako leto s podatkom za leto leta 1965. Take indeksne vrste imenujemo indeksne vrste s stalno osnovo ali bazo.



Slika 6.13 Indeksi razvoja za nekaj pojavov iz gospodarstva za SFRJ v razdobju 1965-71

Večkrat pa računamo indekse tudi tako, da v isti časovni vrsti menjamo osnovo ali bazo primerjave. Take indeksne vrste imenujemo indeksne vrste s premično bazo. Izmed indeksnih vrst s premično bazo najpogosteje uporabljamo verižne indekse. Za dano časovno vrsto računamo vrsto verižnih indeksov tako, da za vsak podatek, za katerega računamo verižni indeks, vzamemo za osnovo predhodni člen. Tako dobimo vrsto verižnih indeksov, ki pokažejo relativne spremembe od člena do člena. Verižne indekse torej računamo po splošnem obrazcu

$$I_k = 100 \cdot Y_k / Y_{k-1} \qquad (6.11)$$

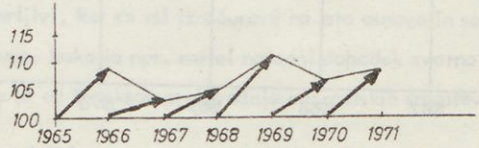
Pri tem pomeni: Y_k = tekoči podatek; Y_{k-1} = podatek za predhodni člen; I_k = verižni indeks.

6.45 Če vzamemo iz tabele 6.18 podatke o narodnem dohodku v SFRJ v letih 1965-1971, je vrsta verižnih indeksov takale:

Tabela 6.20 Verižni indeksi za narodni dohodek v SFRJ v razdobju 1965-1971

Leto	Narodni dohodek v 10^6 din	Verižni indeks
1965	84476	
1966	91733	$108,6 = 100 \cdot \frac{91733}{84476}$
1967	94014	102,5
1968	97692	103,9
1969	107856	110,4
1970	114269	105,9
1971	123598	108,2

Iz zgleda vidimo, da verižnega indeksa za prvi člen časovne vrste ne moremo izračunati, ker ne poznamo predhodnega člena. Vrsta verižnih indeksov kaže, da se narodni dohodek od leta do leta veča, da pa so spremembe neenakomerne.



Slika 6.14 Verižni indeksi za ND v SFRJ

6.46 Preračunavanje indeksov na drugo osnovo. V navedenih primerih smo izračunali indekse iz vrst osnovnih podatkov. Dostikrat pa imamo indeksne vrste, iz katerih želimo izračunati indeksne vrste z novo osnovo, nimamo pa osnovnih podatkov. V takih primerih preračunamo indeksno vrsto v indeksno vrsto z novo bazo 1 tako kakor iz osnovne vrste

$$I_{2/1} = 100 \frac{I_{2/0}}{I_{1/0}} \qquad (6.12)$$

Velja namreč

$$100 \cdot \frac{I_{2/0}}{I_{1/0}} = 100 \cdot \frac{100.Y_2/Y_0}{100.Y_1/Y_0} = 100.Y_2/Y_1 = I_{2/1} .$$

6.47 Vzemimo za zgled iz tabele 6.16 indeksno vrsto za poprečno pogodbeno prodajno ceno za 1 m^2. Osnova te indeksne vrste je Ljubljana. Želimo pa izračunati novo indeksno vrsto, v kateri je osnova mesto z najnižjim indeksom (Skoplje). Po obrazcu 6.12 indeks za vsako mesto primerjamo z bazičnim indeksom (Skoplje 59).

Tabela 6.21 Indeksi za poprečno pogodbeno prodajno ceno za 1 m^2 stanovanjske površine v letu 1971

Mesto	Beograd	Ljubljana	Sarajevo	Skoplje	Titograd	Zagreb
Ljubljana =100	84	100	91	59	70	112
Skoplje =100	142	169	154	100	119	192

Za Beograd smo indeks z novo bazo izračunali takole:

$$100 \frac{84}{59} = 142$$

6.48 Podobno preračunamo na novo osnovo tudi vrste časovnih indeksov. Tako dobimo po gornjem pravilu iz indeksne vrste za proizvodnjo električne energije v Jugoslaviji iz tabele 6.19 novo indeksno vrsto z osnovo 1971, tako da vsak člen indeksne vrste delimo z indeksom za leto 1971 (190) in kvociente pomnožimo s 100.

Podobno kakor iz absolutnih podatkov moremo izračunati iz časovnih indeksnih vrst tudi vrste verižnih indeksov. Iz vrste verižnih indeksov pa dobimo indeksne vrste s poljubno

stalno osnovo s postopnim množenjem oziroma deljenjem verižnih indeksov.

Tabela 6.22 Indeksna vrsta za proizvodnjo električne energije v Jugoslaviji v razdobju 1965-1971.

Leto	1965	1966	1967	1968	1969	1970	1971
1965 = 100	100	111	120	133	151	168	190
1971 = 100	53	58	63	70	79	88	100
verižni indeks	.	111	108	111	114	111	113

6.49 Izbira baze ali osnove. Formalno moremo vzeti za osnovo pri računanju indeksov kateri koli člen v vrsti, za katero računamo indekse, ali tudi vrednost, ki je izven nje. Vsebinsko pa je odločitev o osnovi primerjave odvisna od namena, primernosti in smiselnosti primerjave. Zaradi tega je za izbiro baze nemogoče dati enotno pravilo, marveč samo nekaj splošnih načel, ki pomagajo pri pravilni izbiri baze.

Pri stvarnih in geografskih indeksih vzamemo za osnovo po pravilu pojav oziroma območje, ki si ga najbolje predstavljamo oziroma ga najbolje poznamo. Tako v medrepubliški primerjavi vzamemo npr. za osnovo SR Slovenijo, v meddržavni primerjavi SFRJ itd. Če primerjamo z indeksi relativna števila ali druge izvedene pokazovalce za posamezne dele populacije, je najprimerneje in najenostavneje, da vzamemo za bazo sumarni pokazovalec za celoto. Taki indeksi pokažejo odklone od nekega povprečnega stanja.

Poseben problem je izbira osnove pri časovnih indeksih. Za časovne indekse velja splošno pravilo, da vzamemo za osnovo čas, ko je pojav normalen in ustaljen. Kdaj pa je pojav normalen, je težko določiti. Vendar moremo iz tega pravila vsaj sklepati, katere člene ne smemo vzeti za osnovo, ker je laže ugotoviti, kdaj pojav ni normalen. Zaradi tega za primerjavo večine pojavov ne vzamemo vojna leta, leta gospodarskih kriz in tako dalje. Enako v zdravstveni statistiki ne vzamemo za osnovo leta epidemij. Večja verjetnost je, da je pojav normalen v daljšem razdobju kot v krajšem. Zaradi tega običajno ne jemljemo za osnovo kratka razdobja, npr. mesece pri indeksih proizvodnje, marveč vzamemo za osnovo leto. Težko je tudi npr. reči v kmetijstvu, katero leto je normalno, ali je to leto ugodne ali neugodne vegetacije.

Zato v kmetijstvu dostikrat vzamemo za osnovo večletno (petletno ali desetletno) poprečje.

Povojni razvoj običajno primerjamo s stanjem pred vojno. Za osnovo primerjave vzamemo čim kasnejše leto pred vojno, vendar tako, da nanj še ne vplivajo priprave na vojno. V SFRJ vzamemo za primerjavo s predvojnim stanjem za bazo leto 1939, OZN pa leto 1937. Vendar moramo paziti, kdaj so primerjave pred in povojnega stanja vsebinsko utemeljene in smiselne. Za zelo dolga razdobja in v primeru velikih sprememb primerjava na določeno staro stanje nima smisla.

Če je bila nekaj let po drugi svetovni vojni primerjava s predvojnim stanjem upravičena, ker smo tako približno dobili vtis o pojavih in spremembah, je primerjava s predvojnim stanjem sedaj nepomembna, ker nas zanima povojni razvoj, ne pa primerjava s predvojnim stanjem, ki je že razmeroma zelo odmaknjeno. Vendar ne vzamemo za osnovo povojnega razvoja prvo povojno leto, ko so bile razmere še neustaljene, temveč kasnejše razdobje normaliziranega stanja.

Včasih vzamemo za osnovo indeksov tudi zadnje - tekoče leto. Taka indeksna vrsta primerja preteklo stanje s podatkom, ki je časovno najbližji in tudi najzanimivejši, ker predstavlja trenutno stanje.

Velikih sprememb tudi ne kaže prikazovati z indeksi. Tako nima smisla npr. računati indeksa proizvodnje valjanih aluminijevih proizvodov na bazo 1939, ko je bila proizvodnja 15 ton, s proizvodnjo v letu 1957, ko smo proizvedli 4527 ton valjanih aluminijevih proizvodov. Indeks, ki ga izračunamo iz zgornjih podatkov, je 30180 in nenazoren. Kljub temu, da indeksi v splošnem izboljšajo primerjavo, ta indeks nima smisla, ker je nepredstavljiv.

SEDMO POGLAVJE

7. FREKVENČNE PORAZDELITVE

SESTAVLJANJE FREKVENČNE PORAZDELITVE

7.1 Med statističnimi vrstami imajo posebno mesto vrste, ki prikazujejo razporeditev vrednosti za numerične znake. Take vrste imenujemo **frekvenčne porazdelitve ali pogostnostne porazdelitve**. Za dano populacijo dobimo frekvenčno porazdelitev, če za posamezne razrede za proučevani znak poiščemo, koliko enot ima vrednost znaka v ustreznem razredu.

Vzemimo za zgled porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v takratnem okraju Novo mesto. Ta gospodarstva so bila anketirana v anketi o porabi lesa leta 1953. Osnovni podatki o porabi lesa v posameznih anketiranih gospodarstvih so navedeni v tabeli 7.1.

Tabela 7.1 Poraba lesa v m^3 v 149 kmetijskih gospodarstvih v letu 1953 v okraju Novo mesto. (Vir: Anketa o porabi lesa v letu 1953).

11,6	8,0	8,5	12,2	11,6	11,8	10,2	15,1	2,1	18,5	7,2	10,1
12,7	18,0	14,4	16,2	20,5	20,1	14,6	15,7	23,5	12,2	9,0	12,7
19,4	6,5	6,0	9,8	8,5	5,0	3,6	10,1	10,3	8,6	8,4	5,5
6,0	13,9	12,8	12,4	10,9	15,7	13,6	13,4	12,8	19,7	14,3	7,5
16,0	9,0	9,4	14,2	11,0	10,5	5,9	29,4	14,0	17,5	10,0	19,5
24,4	12,4	17,3	11,3	11,8	26,7	16,5	16,9	12,1	17,0	26,7	24,4
20,1	10,9	13,9	10,9	10,3	7,6	10,3	21,7	17,7	17,4	10,8	15,1
15,2	23,2	21,8	16,2	19,2	22,7	19,0	13,7	14,3	24,9	23,8	19,9
12,1	31,6	29,9	26,5	6,7	8,2	4,9	7,9	9,4	9,3	16,2	19,6
12,9	32,4	18,3	15,2	11,3	8,3	4,7	15,6	13,3	17,3	27,5	15,1
18,4	14,8	17,1	10,5	8,4	7,5	12,2	9,7	18,1	16,9	14,0	12,1
10,5	6,2	22,7	14,3	15,7	14,6	12,0	18,9	16,1	13,2	28,6	6,2
12,2	19,8	23,4	33,5	15,8							

Uredimo zgornjo nepregledno množico podatkov v frekvenčno porazdelitev, v kateri vzemimo po 1 m^3 široke razrede! V tabeli 7.2 pomeni npr. 3 porabo od 3,0 m^3 do 3,9 m^3.

Tabela 7.2 Frekvenčna porazdelitev 149 kmetijskih gospodarstev v Novem mestu po porabi lesa v letu 1953.

Poraba lesa m^3	y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Število gospodarstev	f	-	-	1	1	2	3	6	5	8	7	14	7	14	7	10

y	15	15	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
f	10	8	7	6	8	3	3	2	4	3	-	3	1	1	2	-	1	1	1

Število enot v posameznem razredu imenujemo frekvenca. Po pravilu zaznamujemo frekvence s črko f.

Frekvenčna porazdelitev v tabeli 7.2 kaže veliko pregledneje porabo lesa v gospodarstvih kakor posamični podatki v tabeli 7.1. Vendar frekvenčna porazdelitev ne da natančnih vrednosti. Iz nje npr. vidimo, da je eno gospodarstvo porabilo od 2,0 do 2,9 m^3 lesa, ne vemo pa, koliko so natančno porabili lesa v tem gospodarstvu. Enako je iz frekvenčne porazdelitve razvidno, da je 14 gospodarstev porabilo 10,0 do 10,9 m^3, ne vemo pa natančne porabe lesa. S frekvenčno porazdelitvijo se je preglednost povečala, natančnost informacij pa zmanjšala.

Iz frekvenčne porazdelitve dobimo nekatere zelo koristne ugotovitve. Predvsem je frekvenčna porazdelitev slika o variiranju znaka. Iz tabele 7.2 vidimo, da je bila npr. poraba lesa v kmetijskih gospodarstvih od 2 m^3 do 34 m^3. Razen tega si iz frekvenčne porazdelitve zlahka ustvarimo sliko o porabi lesa znotraj teh meja. Bežen pogled pokaže, da je gospodarstev z majhno porabo malo, medtem ko se število gospodarstev v posameznih razredih do neke porabe v glavnem veča, za večjo porabo pa so frekvence vedno manjše. Frekvenčna porazdelitev nakazuje zakonitost, ki jo opazimo pri večini populacij. V dani populaciji se vrednosti najbolj goste okrog nekega središča, od katerega so odkloni tem redkejši, čim večji so. Vendar se ta zakonitost v frekvenčni porazdelitvi v tabeli 7.2 ne kaže popolnoma. Imamo več odstopanj od tega pravila. Frekvence v sosednih razredih so včasih večje, včasih manjše. Število enot v posameznih razredih je namreč razmeroma majhno, zato prevladujejo še slučajni vplivi. Če bi bilo število enot v posameznih razredih večje, slučajni vplivi ne bi bili tako močni in bi se gornja zakonitost izražala močneje. Frekvence v razredih moremo pove-

čati na dva načina. Če bi namesto 149 gospodarstev imeli desetkrat večjo populacijo s 1490 gospodarstvi, bi bile tako frekvence večje. Zakonitost bi se izražala močneje. Vendar tako večanje frekvenc običajno ne pride v poštev. Frekvence pa povečamo tudi, če vzamemo širše razrede. S tem se sicer zmanjša natančnost prikaza, zakonitost variiranja vrednosti pa je prikazana bolje, ker so slučajni vplivi manjši.

Če v tabeli 7.2 združimo po štiri razrede, dobimo novo frekvenčno porazdelitev; ta je prikazana v tabeli 7.3.

Tabela 7.3 Frekvenčna porazdelitev za porabo lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto.

Poraba v m^3	f
0,0 - 3,9	2
4,0 - 7,9	16
8,0 - 11,9	36
12,0 - 15,9	41
16,0 - 19,9	29
20,0 - 23,9	12
24,0 - 27,9	7
28,0 - 31,9	4
32,0 - 35,9	2
	149 = N

Frekvenčna porazdelitev v tabeli 7.3 je preglednejša, ker je število razredov manjše. Razen tega pa je zaradi razmeroma velikih frekvenc dobro vidna zakonitost gostitve. Frekvence se sprva večajo, dokler ne dosežejo največjo gostitev v razredu 12,0 - 15,9 m^3, od tu pa stalno padajo.

Seveda pa ne smejo biti razredi preširoki. S širjenjem razredov se natančnost bolj in bolj manjša. Tudi zakonitost gostitve je vedno manj vidna. Če v gornji frekvenčni porazdelitvi združimo po dva razreda, dobimo skrčeno porazdelitev, ki je prikazana v tabeli 7.4.

Tabela 7.4 Frekvenčna porazdelitev za porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

Poraba v m^3	f
0,0 - 7,9	18
8,0 - 15,9	77
16,0 - 23,9	41
24,0 - 31,9	11
32,0 - 39,9	2
	149 = N

Ta frekvenčna porazdelitev sicer še vedno nakazuje osnovno zakonitost, vendar daje pregrobo sliko, ker so v istem razredu gospodarstva, ki imajo do 8 m^3 razlik v porabi lesa.

Zato je pri sestavljanju frekvenčnih porazdelitev važno število in velikost razredov. Če vzamemo premajhne razrede, na frekvence preveč vplivajo slučajni vplivi in zabrišejo osnovno zakonitost gostitve, preveliki razredi pa dajo pregrobo sliko. Nimamo splošnega pravila, kakšne in koliko razredov naj ima frekvenčna porazdelitev, pač pa je praksa pokazala tole: razredi, ki razdele razmik med najmanjšo in največjo vrednostjo v populaciji na osem do šestnajst razredov, dajo običajno dosti dobro sliko gostitve. Število razredov se ravna po velikosti populacije. Za manjše populacije vzamemo manj, za večje pa več razredov. V našem primeru smo resnično dobili s štiridesetimi razredi nepregledno, s petimi razredi pregrobo sliko o porabi lesa, medtem ko je frekvenčna porazdelitev z desetimi razredi pokazala vse značilnosti gostitve.

7.2 Frekvenčne porazdelitve z neenakimi razredi. Čeprav je iz tehničnih razlogov najbolje, da so vsi razredi v frekvenčni porazdelitvi enako široki, včasih iz vsebinskih razlogov sestavljamo frekvenčne porazdelitve, v katerih so razredi različno široki. Vzemimo za zgled industrijska podjetja! Če ima eno podjetje 10, drugo pa 50 zaposlenih, je med njima večja vsebinska razlika kakor pa med podjetjema s 1000 in 1040 delavci, čeprav je v obeh primerih razlika 40 delavcev. Zato industrijska podjetja po številu delavstva prikazujemo v frekvenčni porazdelitvi z različnimi širinami razredov.

Tabela 7.5 Industrijska podjetja v SFRJ v letu 1970 po poprečnem številu zaposlenih (Vir: SG 72)

Število zaposlenih	f
- 15	79
16 - 29	61
30 - 60	150
61 - 125	378
126 - 250	522
251 - 500	475
501 - 1000	347
1001 - 2000	221
2001 -	141
	2374 = N

Iz tabele 7.5 vidimo, da se meje razredov vrste v približnem geometrijskem zaporedju z razmerjem k = 2; to pa je v skladu z relativno primerljivostjo števila delavstva po podjetjih.

7.3 V frekvenčnih porazdelitvah z enakimi razredi so razlike med frekvencami predvsem izraz različne stopnje gostitve v posameznih razredih. V frekvenčnih porazdelitvah z neenakimi razredi pa je frekvenca v danem razredu odvisna razen od gostitve tudi od širine razreda. Pri isti stopnji gostitve imajo širši razredi večjo, ožji razredi pa manjšo frekvenco. Da iz frekvence odstranimo vpliv različne širine razredov in prikažemo samo stopnjo gostitve, izračunavamo po obrazcu

$$g_k = f_k / i_k \qquad (7.1)$$

gostoto frekvence g_k; ta pokaže, koliko frekvence v razredu odpade na enotin razmik.

Za vsako frekvenčno porazdelitev moremo izračunati strukturne deleže; ti pokažejo, koliki del celotne populacije je v posameznem razredu. Strukturne deleže, ki jih dobimo, če frekvence f_k delimo z obsegom populacije N, imenujemo relativne frekvence f_k^o

$$f_k^o = f_k / N \qquad (7.2)$$

Kakor iz frekvenc računamo gostoto frekvence, tako moremo po obrazcu

$$\varphi_k = g_k^o = f_k^o / i_k = f_k / N.i_k \qquad (7.3)$$

izračunati tudi gostoto relativne frekvence φ_k, če relativno frekvenco delimo s širino razreda.

Iz obrazca 7.3 zlahka dobimo, da je

$$f_k = N.i_k \cdot \varphi_k \qquad (7.4)$$

Frekvenca v danem razredu je torej premo sorazmerna z obsegom populacije N, s širino razreda i_k in z gostoto relativne frekvence φ_k, ki je povezana z vsebino pojava.

7.4 Za frekvenčno porazdelitev industrijskih podjetij v SFRJ v letu 1970 so nakazani pokazovalci dani v tabeli 7.6.

Tabela 7.6. Industrijska podjetja v SFRJ v letu 1970 po poprečnem številu zaposlenih

Število zaposlenih	f_k	i_k	g_k	f_k^o	φ_k
8 - 15	79	8	9.9	·033	·00416
16 - 29	61	14	4.4	·026	184
30 - 60	150	31	4.8	·063	204
61 - 125	378	65	5.8	·159	245
126 - 250	522	125	4.2	·220	176
251 - 500	475	250	1.9	·200	80
501 - 1000	347	500	·7	·146	29
1001 - 2000	221	1000	·2	·093	9
2001 -	141	-	-	·059	-
Skupaj	2374 = N			·999	

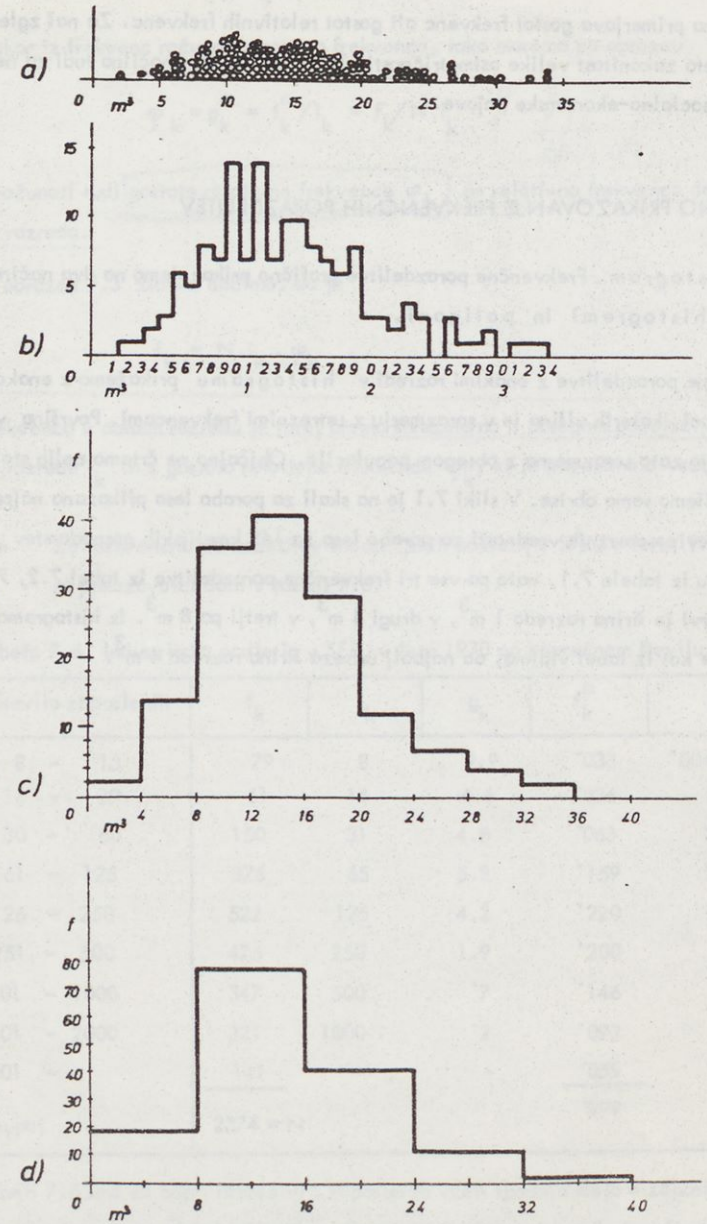
V tabeli 7.6 smo za odprt razred -15 zaposlenih vzeli spodnjo mejo 8 zaposlenih, ker predpostavljamo, da ni industrijskih podjetij, ki imajo manj kot 8 zaposlenih. Zadnji razred je odprt navzgor in zanj ne poznamo širine razreda. Zato zanj nismo mogli izračunati ustreznih gostot. Absolutnih frekvenc med seboj ni mogoče primerjati. Pač pa

je smiselna primerjava gostot frekvenc ali gostot relativnih frekvenc. Za naš zgled ti vrsti kažeta zakonitost velike asimetričnosti. Ta zakonitost je značilna tudi za nekatere druge socialno-ekonomske pojave.

GRAFIČNO PRIKAZOVANJE FREKVENČNIH PORAZDELITEV

7.5 Histogram. Frekvenčne porazdelitve grafično prikazujemo na dva načina: s histogrami in poligoni.

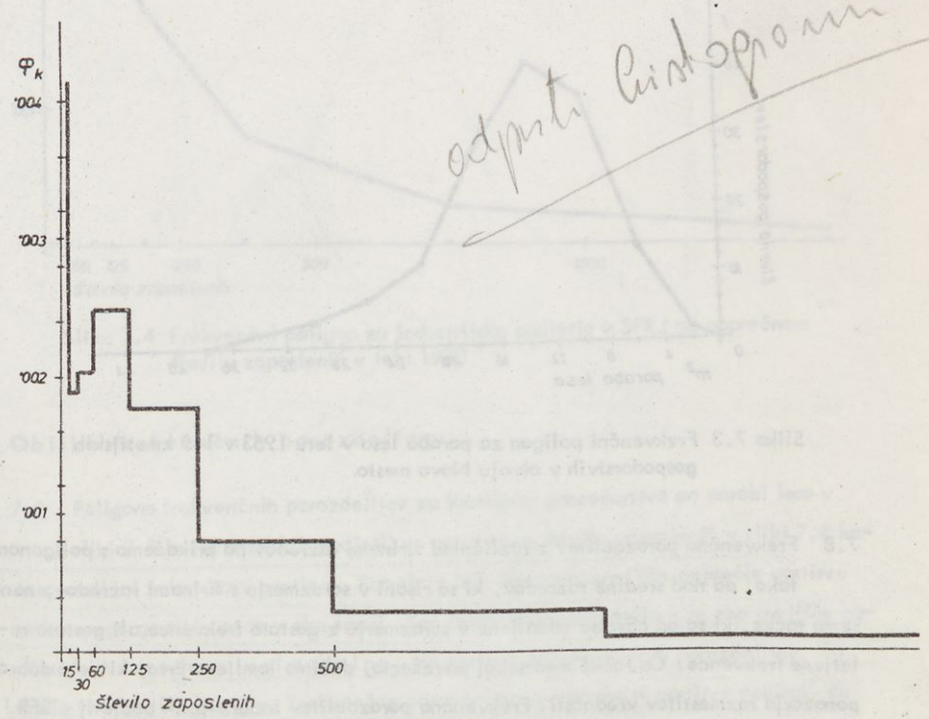
Frekvenčne porazdelitve z enakimi razredi v histogramu prikažemo z enako širokimi stolpci, katerih višina je v sorazmerju z ustreznimi frekvencami. Površina vseh stolpcev je zato sorazmerna z obsegom populacije. Običajno ne črtamo celih stolpcev, temveč rišemo samo obrise. V sliki 7.1 je na skali za porabo lesa prikazana najprej razmestitev posameznih vrednosti za porabo lesa za 149 kmetijskih gospodarstev v Novem mestu iz tabele 7.1, nato pa vse tri frekvenčne porazdelitve iz tabel 7.2, 7.3, 7.4. V prvi je širina razreda 1 m^3, v drugi 4 m^3, v tretji pa 8 m^3. Iz histogramov še nazorneje kot iz tabel vidimo, da najbolj ustreza širina razreda 4 m^3.



Slika 7.1 Histogrami za porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

7.6 Če prikazujemo s histogramom frekvenčne porazdelitve z različnimi širinami razredov, rišemo gostoto frekvence g_k ali gostote relativnih frekvenc φ_k. Ker je po obrazcu 7.1 frekvenca zmnožek širine razreda in gostote frekvence, so ploščine stolpcev v sorazmerju z ustrezno frekvenco f_k, ploščina vseh stolpcev pa v sorazmerju z obsegom populacije N. Če pa rišemo v histogramu gostote relativnih frekvenc, so ploščine sorazmerne z relativnimi frekvencami.

Edino tako dobimo pravilno sliko o razmestitvi pojava. V sliki 7.2 je s histogramom prikazana frekvenčna porazdelitev industrijskih podjetij v SFRJ v letu 1970 po poprečnem številu delavcev iz tabele 7.6.

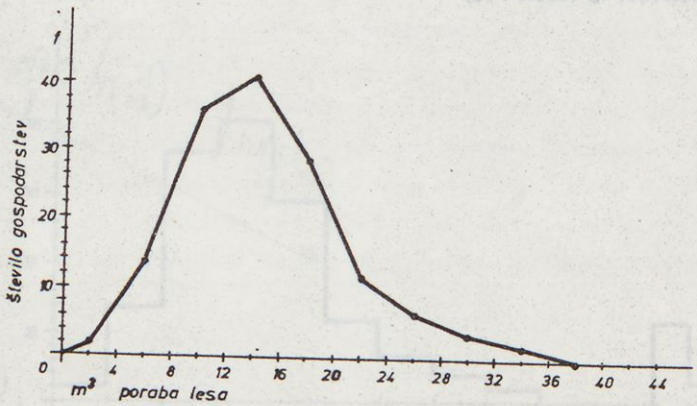


Slika 7.2 Histogram industrijskih podjetij po poprečnem številu zaposlenih v SFRJ v letu 1970

7.7 Poligon. Frekvenčni poligon za vrste z enakimi razredi narišemo tako, da nad sredine razmikov za posamezne razrede nanesemo točke, ki so od abscisne osi oddaljene v sorazmerju z velikostjo frekvence. Dobljene točke po vrsti zvežemo z dalji-

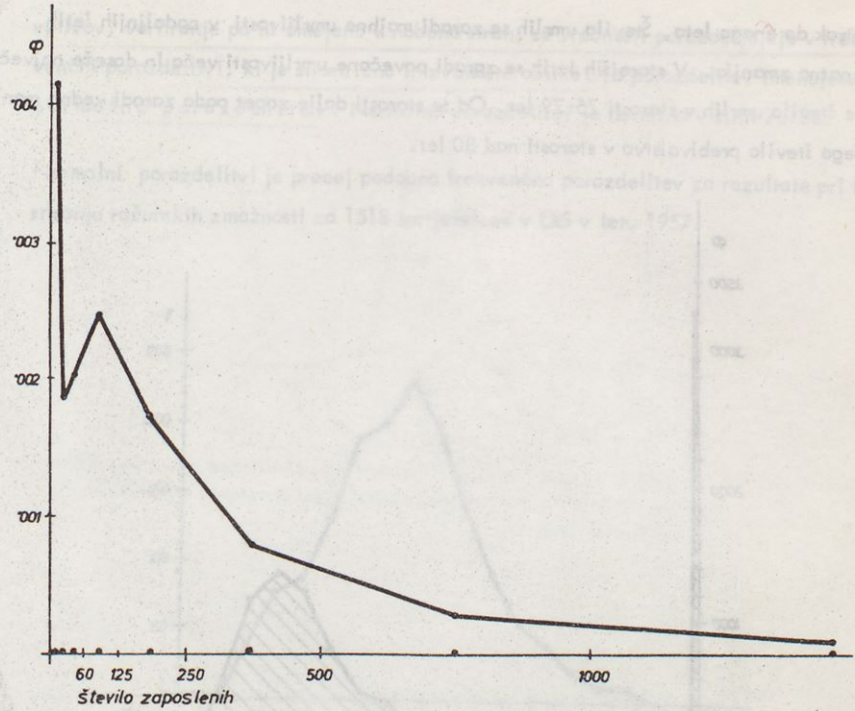
cami. Če vzamemo, da je frekvenca v razredu pred prvim razredom in po zadnjem razredu enaka nič in da ustrezni točki ležita na abscisni osi, je poligon sklenjen. Poligon bolje prikazuje stvarno razmestitev vrednosti, ker z njim nakažemo verjetnejšo razmestitev vrednosti znotraj razredov kot s histogramom, ki nakazuje enakomerno razmestitev znotraj razredov. Logično je namreč, da je znotraj razredov na straneh, ki teže proti največji gostitvi, gostota večja kakor na zunanjih straneh razredov.

V sliki 7.3 je s poligonom prikazana frekvenčna porazdelitev o porabi lesa iz tabele 7.3. Poligon zelo lepo pokaže razmestitev gospodarstev po porabi lesa.



Slika 7.3 Frekvenčni poligon za porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

7.8 Frekvenčno porazdelitev z različnimi širinami razredov pa prikažemo s poligonom tako, da nad sredine razredov, ki so risani v sorazmerju s širinami razredov, nanesemo točke, ki so od abscise oddaljene v sorazmerju z gostoto frekvence ali gostoto relativne frekvence. Če točke med seboj povežemo, dobimo lomljeno črto, ki zelo dobro ponazarja razmestitev vrednosti. Frekvenčna porazdelitev industrijskih podjetij v SFRJ po poprečnem številu zaposlenih je s poligonom narisana v sliki 7.4. Poligon še nazorneje kakor histogram kaže zakonitost gostitve.

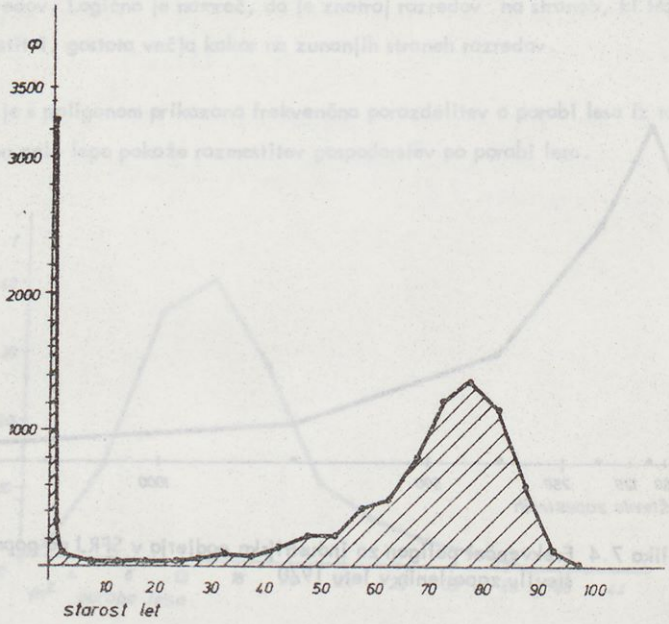


Slika 7.4 Frekvenčni poligon za industrijska podjetja v SFRJ po poprečnem številu zaposlenih v letu 1970

Oblike frekvenčnih porazdelitev

7.9 Poligona frekvenčnih porazdelitev za kmetijska gospodarstva po porabi lesa v sliki 7.3 in industrijskih podjetij po poprečnem številu zaposlenih v sliki 7.4 kažeta različni zakonitosti gostitve. Potrošnja lesa ima neko središče največje gostitve od katerega gostitev pada v obe smeri. Ker imajo take porazdelitve le eno središče gostitve, jih imenujemo unimodalne, za razliko od frekvenčnih porazdelitev, ki imajo več središč gostitve. Frekvenčne porazdelitve z več mesti gostitve dobimo, če je populacija heterogena, torej sestavljena iz več homogenih populacij. Te vrste porazdelitev imenujemo bimodalne, če imajo dva vrha, in polimodalne, če imajo več vrhov. V sliki 7.5 je prikazana frekvenčna porazdelitev umrlih žensk po starosti v letu 1956 v Sloveniji. V grafikonu opazimo izredno veliko število umrlih

otrok do enega leta. Število umrlih se zaradi majhne umrljivosti v nadaljnjih letih znatno zmanjša. V starejših letih se zaradi povečane umrljivosti veča in doseže največje število umrlih v starosti 75-79 let. Od te starosti dalje zopet pada zaradi vedno manjšega števila prebivalstva v starosti nad 80 let.



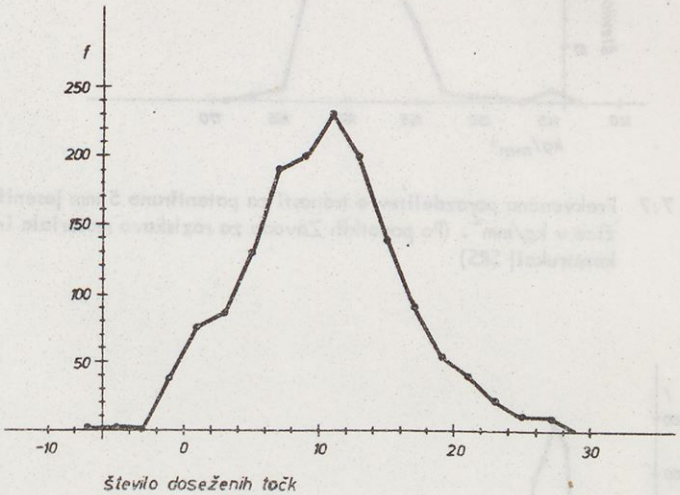
Slika 7.5 Število umrlih žensk v letu 1956 v Sloveniji po starosti ob smrti

7.10 Poligon za frekvenčno porazdelitev o porabi lesa v kmetijskih gospodarstvih kaže bolj ali manj enako upadanje gostitve od središča gostitve. Frekvenčne porazdelitve, ki kažejo enako upadanje frekvenc na levo in desno od središča gostitve, imenujemo simetrične porazdelitve, za razliko od asimetričnih porazdelitev, pri katerih je upadanje gostitve na eno ali drugo stran hitrejše. Če je upadanje gostitve počasnejše na levo od središča gostitve, pravimo, da je frekvenčna porazdelitev asimetrična v levo, če pa je upadanje gostitve počasnejše desno od središča, govorimo o asimetriji v desno. Simetrično porazdelitev dobimo navadno za homogene populacije.

Če je populacija popolnoma homogena in so razlike za enote rezultat samo slučajnih

vplivov, variiranje pa ni omejeno v nobeno stran, se vrednosti porazdeljujejo v frekvenčni porazdelitvi, ki je simetrična in zvonaste oblike. To porazdelitev imenujemo normalno porazdelitev. Normalna porazdelitev je narisana v sliki 7.10a.

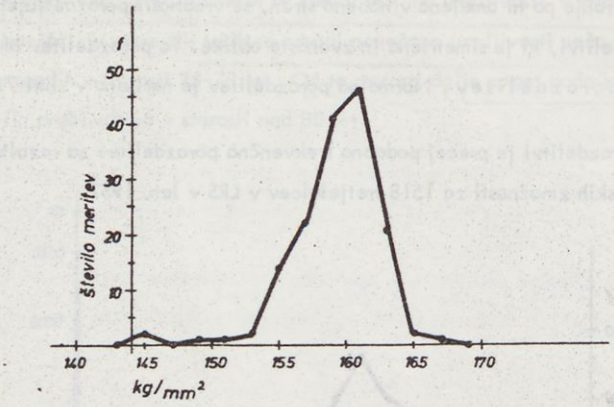
Normalni porazdelitvi je precej podobna frekvenčna porazdelitev za rezultate pri testiranju računskih zmožnosti za 1518 tretješolcev v LRS v letu 1957.



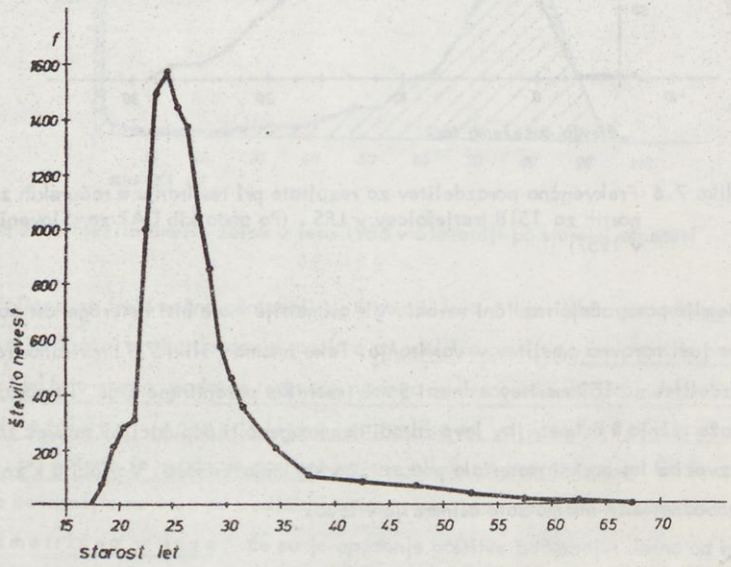
Slika 7.6 Frekvenčna porazdelitev za rezultate pri testiranju o računskih zmožnostih za 1518 tretješolcev v LRS. (Po podatkih DAT za Slovenijo v 1957)

Asimetrijo povzročajo različni vzroki. Vir asimetrije more biti heterogenost populacije ali pa tudi naravna omejitev v variiranju. Tako imamo v sliki 7.7 narisano frekvenčno porazdelitev za 153 meritev trdnosti 5 mm jeseniške patentirane žice. Ta porazdelitev ne kaže težnjo k asimetriji v levo zaradi nehomogenosti populacije, temveč zato, ker je v zvezi z lastnostmi materiala podana zgornja meja trdnosti. Variacija k znižanju je svobodnejša in imamo zato asimetrijo v levo.

Asimetrija na desno pa je vidna iz porazdelitve starosti nevest v letu 1956 v Sloveniji.

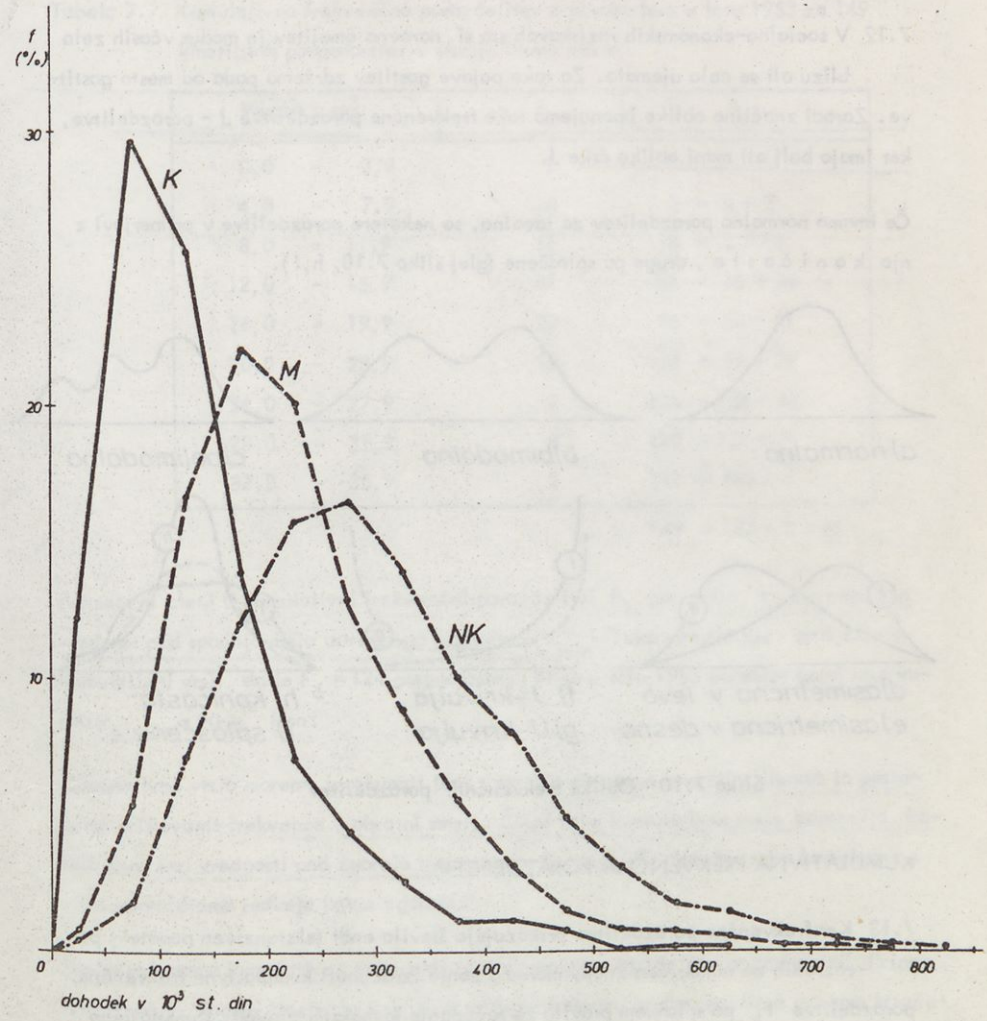


Slika 7.7 Frekvenčna porazdelitev o trdnosti za patentirano 5 mm jeseniško žico v kg/mm^2. (Po podatkih Zavoda za raziskavo materiala in konstrukcij SRS)



Slika 7.8 Frekvenčna porazdelitev za starost nevest v letu 1956 v Sloveniji. (Vir: SL LRS 57)

7.11 Kako je stopnja asimetrije odvisna od tega, kako blizu je jedro gostitve od naravne meje, je nazorno vidno na sliki 7.9. V njej so prikazane frekvenčne po-



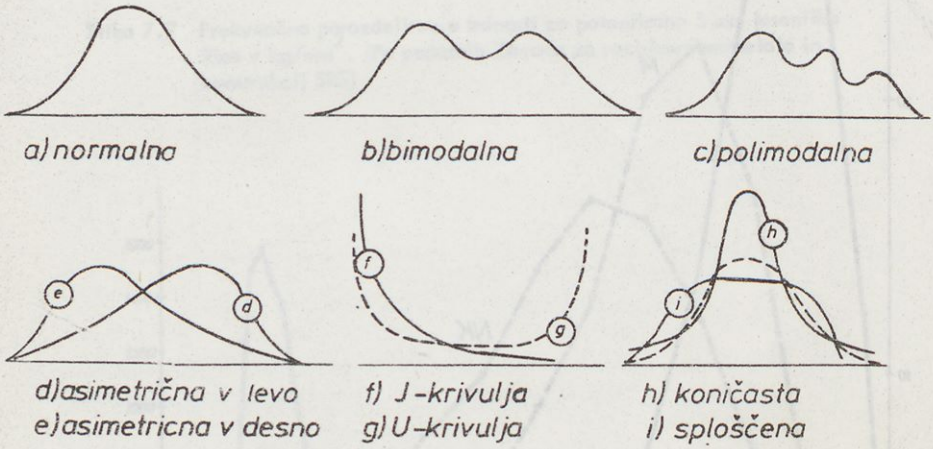
Slika 7.9 Frekvenčne porazdelitve K = kmečkih, M = mešanih in NK = nekmečkih gospodinjstev v SR Sloveniji v letu 1960 po letnem dohodku

razdelitve gospodinjstev v SR Sloveniji v letu 1960 po letnem dohodku. Gospodinjstva so razdeljena v tri skupine: K = kmečka, M = mešana in NK = nekmečka. Ker je za

kmečka gospodinjstva jedro gostitve - modus najbližje naravni omejitvi dohodka (dohodek o), je asimetrija v desno za to skupino največja. Za nekmečka gospodinjstva je iz obratnega razloga stopnja asimetrije znatno manjša.

7.12 V socialno-ekonomskih raziskavah sta si naravna omejitev in modus včasih zelo blizu ali se celo ujemata. Za take pojave gostitev zdržema pada od mesta gostitve. Zaradi značilne oblike imenujemo take frekvenčne porazdelitve J - porazdelitve, ker imajo bolj ali manj obliko črke J.

Če imamo normalno porazdelitev za idealno, so nekatere porazdelitve v primerjavi z njo k o n i č a s t e , druge pa sploščene (glej sliko 7.10, h,i).



Slika 7.10 Oblike frekvenčnih porazdelitev

KUMULATIVNA FREKVENČNA PORAZDELITEV

7.13 Ker frekvenčne porazdelitve prikazujejo število enot (ekstenziven podatek) po razredih za numeričen znak, moremo zanje izračunati kumulativne frekvenčne porazdelitve F_k po splošnem pravilu za računanje kumulativnih vrst. Kumulativno frekvenčno porazdelitev F_k dobimo, če postopoma prištevamo frekvence f_k v frekvenčni porazdelitvi. Člene v kumulativni frekvenčni porazdelitvi dobimo torej po obrazcu

$$F_{k+1} = F_k + f_k \qquad F_{i+1} = F_i + f_i \qquad (7.5)$$

7.14 Če za frekvenčno porazdelitev o porabi lesa iz tabele 7.3 izračunamo kumulativno porazdelitev, dobimo naslednjo kumulativno vrsto:

Tabela 7.7 Kumulativna frekvenčna porazdelitev o porabi lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m^3	f_k	F_k
0,0 - 3,9	2	0
4,0 - 7,9	16	2 = 0 + 2
8,0 - 11,9	36	18 = 2 + 16
12,0 - 15,9	41	54 = 18 + 36
16,0 - 19,9	29	95 = 54 + 41
20,0 - 23,9	12	124 = 95 + 29
24,0 - 27,9	7	136 = 124 + 12
28,0 - 31,9	4	143 = 136 + 7
32,0 - 35,9	2	147 = 143 + 4
		149 = 147 + 2 = N

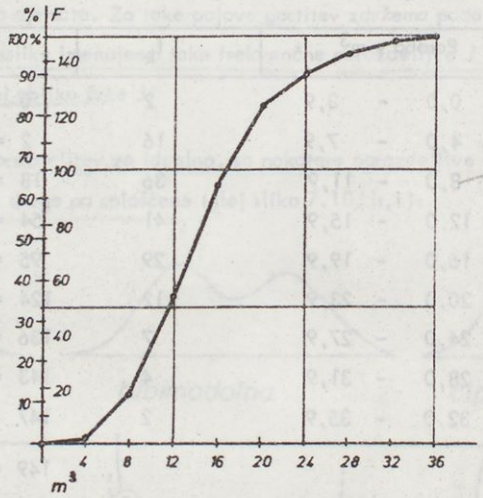
Posamezni členi v kumulativni frekvenčni porazdelitvi F_k pomenijo, koliko enot ima vrednost pod spodnjo mejo ustreznega razreda $y_{k,min}$. Tako pomeni npr. šesti člen v kumulativni vrsti, da je $F_6 = 124$ gospodarstev, ki so v letu 1953 porabila manj lesa kakor $y_{6,min} = 20\ m^3$ lesa.

Kumulativno vrsto moremo izračunati tudi tako, da začnemo z večjim členom in postopoma prištevamo frekvence v obratni smeri. Členi take kumulativne vrste pomenijo, koliko enot ima vrednosti nad zgornjo mejo ustreznih razredov. Vendar tako kumulativno vrsto uporabljamo redkeje kakor zgornjo.

Enako računamo kumulativne vrste tudi za frekvenčne porazdelitve z različnimi širinami razredov. Pomen členov je isti. Kumulativne frekvenčne porazdelitve moremo izračunati tudi iz vrst relativnih frekvenc.

7.13 Če je frekvenčna porazdelitev unimodalna, ima grafični prikaz za kumulativno frekvenčno porazdelitev značilno obliko črke S. O tem se moremo prepričati tudi

v grafikonu za kumulativno frekvenčno porazdelitev o porabi lesa v kmetijskih gospodarstvih v sliki 7.11, ki je načrtana po podatkih iz tabele 7.7. Opomniti moramo, da vrednosti kumulativnih frekvenc v grafikonu nanašamo nad spodnjo mejo ustreznih razredov na abscisi.



Slika 7.11 Kumulativna frekvenčna porazdelitev o porabi lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

V grafikonu imamo na ordinatni osi dve skali: skalo kumulativnih absolutnih in relativnih frekvenc.

OSMO POGLAVJE

KVANTILI
RANŽIRNA VRSTA, RANG

8.1 Osnovni statistični podatki, ki jih zberemo s statističnim opazovanjem, so dani v neurejeni množici podatkov. Te podatke običajno uredimo v frekvenčno porazdelitev. Moremo pa jih pregledno prikazati tudi v ranžirni vrsti, v kateri so urejeni po velikosti od najmanjšega do največjega ali obratno. V tabeli 8.1 so prikazani osnovni podatki za poprečen promet v trgovini na drobno na prebivalca v SR Sloveniji v letu 1971 po občinah.

Tabela 8.1 Promet v trgovini na malo na prebivalca v SR Sloveniji v letu 1971 po občinah v din (Vir: SG-72)

Ajdovščina	AJD	6357	Maribor	MAR	8884
Brežice	BRE	6509	Metlika	MET	6453
Celje	CEL	12983	Mozirje	MOZ	5077
Cerknica	CER	4718	Murska Sobota	MSO	5367
Črnomelj	ČRN	4068	Nova gorica	NGO	12039
Domžale	DOM	6205	Novo mesto	NME	6445
Dravograd	DRA	4826	Ormož	ORM	2752
Gornja Radgona	GRA	4359	Piran	PIR	9678
Grosuplje	GRO	3457	Postojna	POS	8107
Hrastnik	HRA	5082	Ptuj	PTU	5406
Idrija	IDR	5900	Radlje ob Dravi	RJE	4051
Ilirska Bistrica	IBI	5504	Radovljica	RAD	9132

Izola	IZO	8666	Ravne na Koroškem	RAV	5163
Jesenice	JES	8541	Ribnica	RIB	5539
Kamnik	KAM	5728	Sevnica	SEV	3978
Kočevje	KOČ	7089	Sežana	SEŽ	15607
Koper	KOP	14398	Slov. Bistrica	SBI	4135
Kranj	KRA	8538	Slov. Konjice	SKO	4222
Krško	KRŠ	4400	Slov. Gradec	SLG	6332
Laško	LAŠ	4453	Šentjur pri Celju	ŠTJ	2244
Lenart	LRT	2377	Škofja Loka	ŠLO	6667
Lendava	LVA	3973	Šmarje pri Jelšah	ŠMA	3557
Litija	LIT	3743	Tolmin	TOL	8021
Logatec	LOG	5350	Trbovlje	TRB	7714
Ljubljana-Bežigrad	LJB	11293	Trebnje	TRE	4029
Ljubljana-Center	LJC	49433	Tržič	TRŽ	6966
Ljubljana-Moste-Polje	LJM	4499	Velenje	VEL	7161
Ljubljana-Šiška	LJŠ	8598	Vrhnika	VRH	5801
Ljubljana-Vič-Rudnik	LJV	3988	Zagorje ob Savi	ZAG	4907
Ljutomer	LJU	4974	Žalec	ŽAL	4492

Podatki v tabeli 8.1 so urejeni po abecednem redu občin. Za potrebe evidence je tak pregled primeren, statistično pa dobimo veliko boljši pregled, če podatke uredimo po velikosti in jih podamo v ranžirni vrsti. Podatki o prometu v trgovini na drobno na prebivalca po občinah v SRS v letu 1971 so podani v ranžirni vrsti v tabeli 8.2.

Iz ranžirne vrste pa moremo napraviti različne sklepe o prikazanem pojavu. Iz nje podobno kakor iz frekvenčne porazdelitve neposredno vidimo, v kakšnih mejah variira populacija, ker kaže podatek (2244 din), ki je prvi po rangu, najmanjši promet, oni, ki je zadnji po rangu, pa največji promet na malo (49433 din).

Razen tega je v ranžirni vrsti vsaki enoti dodana vrednost novega znaka-ranga. Rang daje o posameznih enotah dodatno informacijo, katere osnovna vrednost znaka - v našem zgledu poprečen promet v občini, ne daje. Če npr. za občino Nova Gorica vemo, da je bil poprečen promet na drobno v letu 1971 12039 din, iz tega podatka ne moremo sklepati, ali je to malo ali veliko. Če pa dodatno povemo, da je v ranžirni vrsti

Tabela 8.2 Ranžirna vrsta za občine v SR Sloveniji za promet na drobno na prebivalca v letu 1971

R	01	02	03	04	05	06	07	08	09	10
Vrednost	2244	2372	2752	3457	3557	3743	3973	3982	3988	4029
Oznaka	ŠTJ	LRT	ORM	GRO	ŠMA	LIT	LEN	SEV	LJ-V	TRE
R	11	12	13	14	15	16	17	18	19	20
Vrednost	4051	4068	4135	4222	4358	4400	4453	4492	4499	4718
Oznaka	RJE	ČRN	SBI	SKO	GRA	KRŠ	LAŠ	ŽAL	LJ-M	CER
R	21	22	23	24	25	26	27	28	29	30
Vrednost	4826	4904	4947	5077	5082	5163	5350	5367	5406	5504
Oznaka	DRA	ZAG	LJU	MOZ	HRA	RAV	LOG	MSO	PTU	IBI
R	31	32	33	34	35	36	37	38	39	40
Vrednost	5539	5728	5804	5900	6205	6332	6357	6445	6453	6509
Oznaka	RIB	KAM	VRH	IDR	DOM	SLG	AJD	NME	MET	BRE
R	41	42	43	44	45	46	47	48	49	50
Vrednost	6667	6966	7089	7161	7714	8021	8107	8538	8541	8598
Oznaka	ŠLO	TRŽ	KOČ	VEL	TRB	TOL	POS	KRA	JES	LJ-Š
R	50	52	53	54	55	56	57	58	59	60
Vrednost	8666	8884	9132	9687	11293	12039	12083	14398	15607	49433
Oznaka	IZO	MAR	RAD	PIR	LJB	NGO	CEL	KOP	SEŽ	LJ-C

60 občin, občina Nova Gorica pa 56- po rangu, sklepamo, da je ta poprečen promet za Slovenijo razmeroma velik, ker imajo od 60 občin samo 4 občine večji promet na prebivalca, 55 občin pa manjšega.

KVANTILNI RANG

8.2 Hiba ranga je v tem, da moramo poleg ranga navesti še skupno število enot populacije, če hočemo, da rang pokaže mesto enote v populaciji. Občina, ki je 56. po rangu, je npr. za populacijo s 60 enotami občina z razmeroma velikim prometom. Če pa je neka občina 56. po rangu v ranžirni vrsti za vseh 500 občin v SFRJ, pa

bi bil ta podatek razmeroma majhen, ker bi imelo 444 občin večji promet. Rang R pokaže torej mesto enote v populaciji šele, če ga primerjamo z obsegom populacije N. Zato je primerneje, da izračunamo mesto enote v populaciji namesto z rangom R s kvantilnim rangom P. Tega dobimo, če primerjamo rang R z obsegom populacije N. Kvantilni rang v relativnem številu pove, na katerem delu celotnega ranžirnega razmika leži določena enota oziroma koliki del celote ima manjše vrednosti kakor je dana vrednost.

8.3 Teoretično vzamemo, da je ranžirni razmik zvezen in vsakemu rangu pripišemo razmik polovico enote na levo in desno. Po tej predpostavki se ranžirni razmik začne z 0,5 (spodnja meja razmika, ki ustreza rangu 1) in konča z N + 0,5 (zgornja meja razmika, ki ustreza rangu N). Ker se skala kvantilnih rangov P začne z 0 in konča z 1, je zveza med rangom R in kvantilnim rangom P dana z obrazcem

$$R = NP + 0,5 \qquad (8.1)$$

Če je P = 0, je R = 0,5, če pa je P = 1, je R = N + 0,5

Iz gornje zveze dobimo, da velja tudi

$$P = \frac{R - 0,5}{N} \qquad (8.2)$$

Če vzamemo naš zgled o prometu na drobno v Novi Gorici z rangom R = 56 in N = 60, dobimo po obrazcu

$$P = \frac{56 - 0,5}{60} = 0,93$$

Kvantilni rang P = 0,93 pove, da ima 0,93 del celotne populacije občin manjši poprečni promet na prebivalca kot občina Nova Gorica. Iz zgleda vidimo, da kvantilni rang P sam zase, ne da bi navajali obseg populacije, nazorno prikaže mesto določene enote v populaciji.

KVANTILI

8.4 Z obrazcem 8.1 in 8.2 moremo reševati dva različna problema.

Za posamezno enoto populacije moremo določiti mesto enote v populaciji, če

izračunamo vrednosti y ustrezen kvantilni rang P . Kvantilni rangi so torej značilnosti posameznih enot.

Moremo pa analogno reševati tudi drug problem. Če se vprašamo, kakšna vrednost y_P ustreza npr. kvantilnemu rangu P = 0,50, vrednost, ki jo dobimo, ne označuje posamezne enote, marveč je značilnost za populacijo. y_P, ki ustreza danemu kvantilnemu rangu P, imenujemo kvantil. Kvantil $y_{0,50}$ je npr. vrednost, od katere ima polovica enot populacije manjše, polovica pa večje vrednosti. Ta vrednost je vsekakor važen parameter populacije in jo imenujemo mediana .

$$Me = y_{P=0.50} \qquad (8.3)$$

Kvantili $y_{0,25}$, $y_{0,50}$ in $y_{0,75}$ so vrednosti, ki razdele populacijo v štiri dele tako, da je pod $y_{0,25}$, med $y_{0,25}$ in $y_{0,50}$, med $y_{0.50}$ in $y_{0,75}$ in nad $y_{0,75}$ po četrtina po velikosti urejenih vrednosti populacije. Te vrednosti, ki razdele populacijo v štiri po obsegu enake dele, imenujemo kvartile in zaznamujemo z

$$Q_1 = y_{P=0.25}; \quad Q_2 = y_{P=0.50}; \quad Q_3 = y_{P=0.75} \qquad (8.4)$$

Podobno z decili

$$D_1 = y_{0,10}, \; D_2 = y_{0,20} \cdots\cdots\cdots D_9 = y_{0,90} \qquad (8.5)$$

razdelimo celotno populacijo v deset po obsegu enakih delov, s centili:

$$C_1 = y_{0,01} \quad C_2 = y_{0,02} \cdots\cdots C_{98} = y_{0,98} \text{ in } C_{99} = y_{0,99} \qquad (8.6)$$

pa v sto po obsegu enakih delov.

Z vrsto kvantilov prikažemo gostitev vrednosti populacije vendar drugače kot s frekvenčno porazdelitvijo. V frekvenčni porazdelitvi z enakimi razredi se glede na gostitev pojava menja frekvenca v razredih, s kvantili pa razmejimo razrede, v katerih so frekvence enake: N/4 pri kvantilih, N/10 pri decilih in N/100 pri centilih, od stopnje gostitve pojava pa je odvisna širina razredov oziroma razmiki med zaporednimi kvantili.

Preračunanje kvantilov iz negrupiranih podatkov

8.5 Ker predpostavljamo, da je rang zvezna količina, moremo določiti range za vsako vrednost med najmanjšo in največjo vrednostjo v populaciji in ne samo za podatke, navedene v ranžirni vrsti. Za vmesne vrednosti uporabimo linearno interpolacijo. Tako npr. menimo, da ustreza v našem zgledu rangu 12,5 kot kvantil sredina med vrednostjo za rang 12 in 13: (4068 + 4135) : 2 = 4101,5 din

Če predpostavljamo zveznost rangov, izračunamo iz ranžirne vrste za poljubno vrednost y med najmanjšo y_{min} in največjo vrednostjo populacije y_{max} ustrezni kvantilni rang P_y po naslednjem postopku:

a) Imamo ranžirno vrsto vrednosti enot populacije.

b) V ranžirni vrsti poiščemo, med kateri vrednosti y_0 in y_1 pade vrednost y, za katero iščemo P_y, tako da velja: $y_0 \leq y < y_1$. Vrednosti y_0 naj ustreza rang R_0.

c) Rang R_y, ki ustreza vrednosti y, dobimo z linearno interpolacijo po obrazcu

$$R_y = R_0 + \frac{y - y_0}{y_1 - y_0} \qquad (8.7)$$

d) Kvantilni rang P_y izračunamo iz ranga R_y in obsega populacije po obrazcu

$$P_y = \frac{R_y - 0,5}{N} \qquad (8.8)$$

Če iščemo, kateri kvantilni rang ustreza npr. prometu na drobno na prebivalca x = = 5000 din, dobimo po gornjem postopku: y = 5000 din leži v ranžirni vrsti med vrednostima $y_0 = 4947$, za katero je rang $R_o = 23$ in vrednostjo $y_1 = 5077$.

Iz teh podatkov izračunamo rang $R_{y=5000}$ po obrazcu 8.7

$$R = 23 + \frac{5000 - 4947}{5077 - 4947} = 23,41$$

Po obrazcu 8.8 pa dobimo dalje: $P_y = \frac{23,41 - 0,50}{60} = 0,382$

Kvantilni rang za občino, s prometom na drobno na prebivalca $y = 5000$ din, je $P_{y=5000} = 0.382$.

8.6 Obratni problem, da k danemu kvantilnemu rangu P_y poiščemo ustrezen kvantil y_p, pa rešimo takole:

a) Za populacijo imamo ranžirno vrsto.

b) Iz danega P po obrazcu

$$R_p = NP + 0,5 \qquad (8.9)$$

izračunamo ustrezni rang R.

c) V ranžirni vrsti poiščemo, med katera cela ranga pade izračunani R_p, tako da velja: $R_o \leq R_p < R_1$. Rangoma R_o in R_1 ustrezata vrednosti y_o in y_1.

d) Iz teh podatkov izračunamo z linearno interpolacijo kvantil y_p po obrazcu:

$$y_p = y_o + (y_1 - y_o) \cdot (R_p - R_o) \qquad (8.10)$$

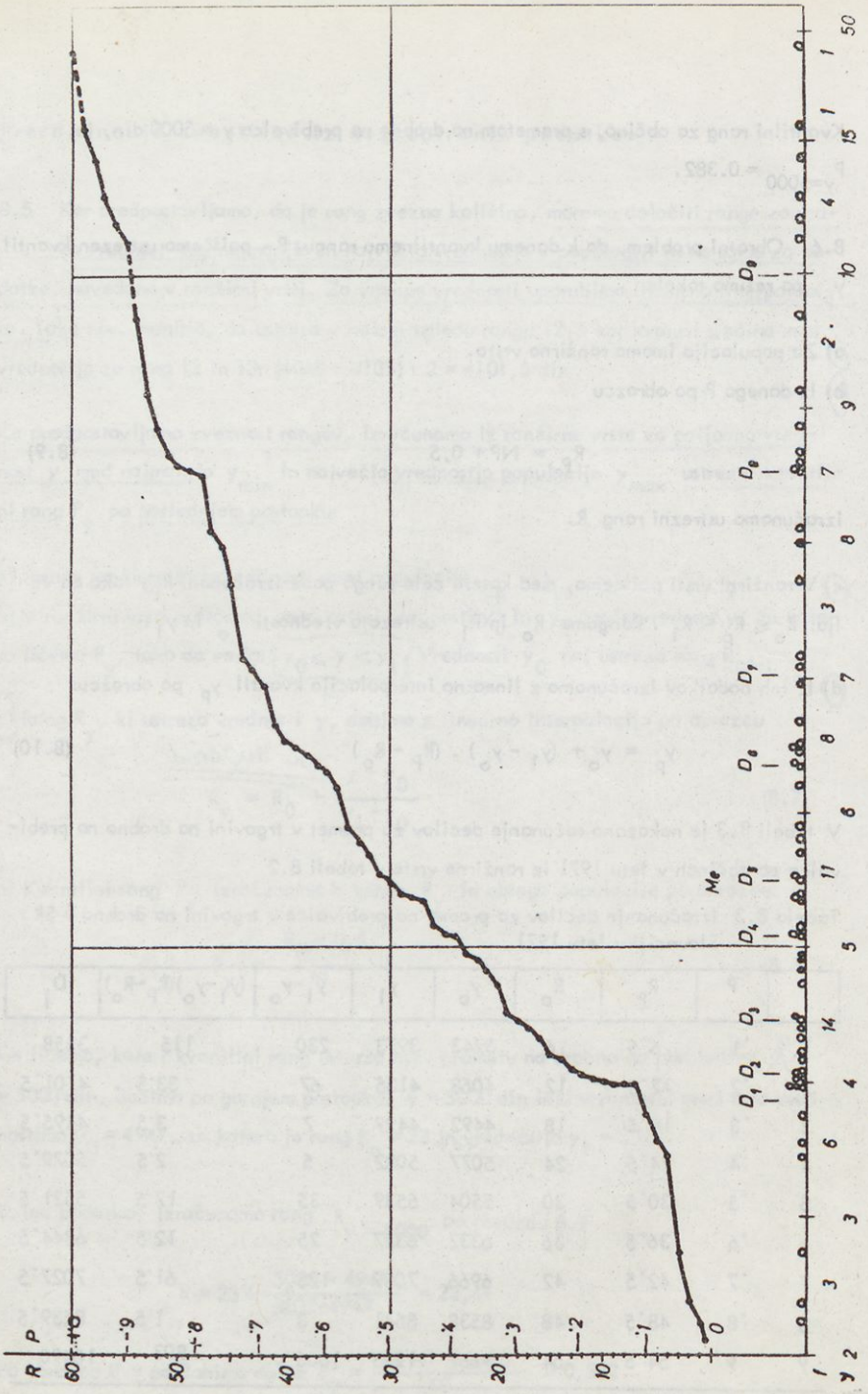
2 močim

V tabeli 8.3 je nakazano računanje decilov za promet v trgovini na drobno na prebivalca po občinah v letu 1971 iz ranžirne vrste v tabeli 8.2

Tabela 8.3 Izračunanje decilov za promet na prebivalca v trgovini na drobno v SR Sloveniji v letu 1971

i	P	R_p	R_o	y_o	y_1	$y_1 - y_o$	$(y_1 - y_o)(R_p - R_o)$	D_i
1	·1	6·5	6	3743	3973	230	115	3858
2	·2	12·5	12	4068	4135	67	33·5	4101·5
3	·3	18·5	18	4492	4499	7	3·5	4495·5
4	·4	24·5	24	5077	5082	5	2·5	5079·5
5	·5	30·5	30	5504	5539	35	17·5	5521·5
6	·6	36·5	36	6332	6357	25	12·5	6344·5
7	·7	42·5	42	6966	7089	123	61·5	7027·5
8	·8	48·5	48	8538	8541	3	1·5	8539·5
9	·9	54·5	54	9687	11293	1606	803	10490

R_p dobimo po obrazcu: $R_p = 60.0,1 \cdot i + 0,5 = 6.i + 0,5$

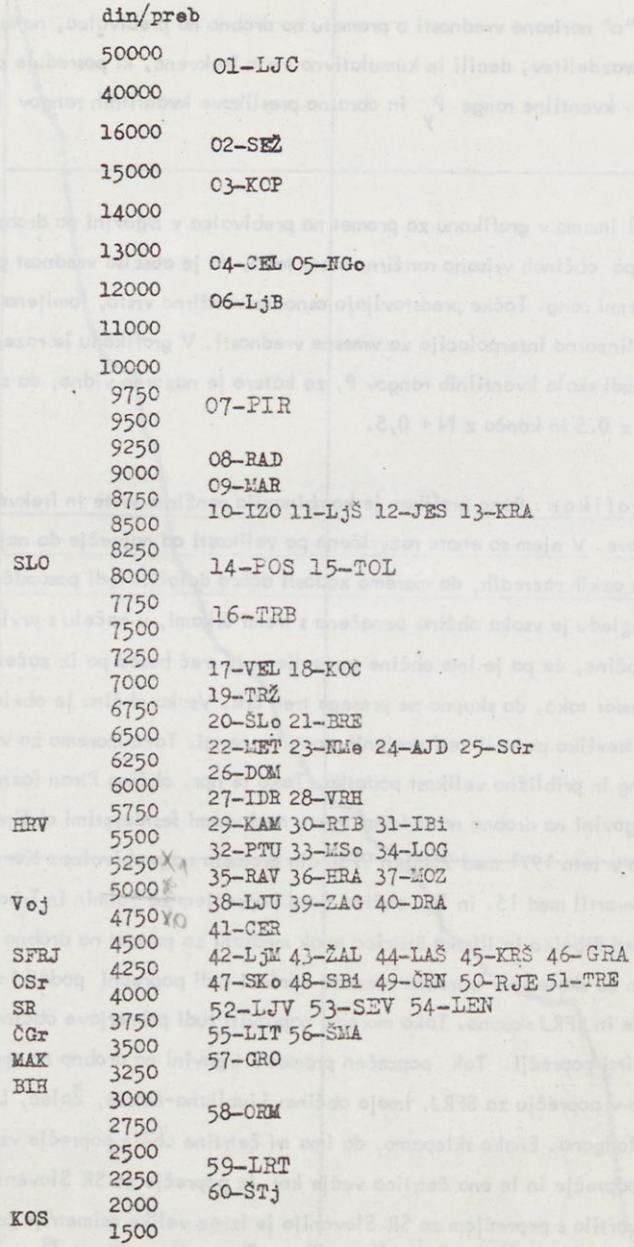


Slika 8.1 Ranžirna vrsta za promet na prebivalca v trgovini na drobno v SR Sloveniji v letu 1971 po občinah

V sliki 8.1 so z "o" narisane vrednosti o prometu na drobno na prebivalca, nakazana je frekvenčna porazdelitev, decili in kumulativna vrsta frekvenc, ki posreduje preslikave vrednosti y v kvantilne range P_y in obratno preslikave kvantilnih rangov P v kvantile y_P.

8.7 V sliki 8.1 imamo v grafikonu za promet na prebivalca v trgovini na drobno v SR Sloveniji po občinah vrisano ranžirno vrsto tako, da je abscisa vrednost podatka, ordinata pa ustrezni rang. Točke predstavljajo osnovno ranžirno vrsto, lomljena črta, ki veže točke, pa linearno interpolacijo za vmesne vrednosti. V grafikonu je razen skale rangov vrisana tudi skala kvantilnih rangov P, za katero je nazorno vidno, da se na skali rangov začne z 0.5 in konča z N + 0,5.

8.8 Rang grafikon. Rang grafikon je kombinacija ranžirne vrste in frekvenčne porazdelitve. V njem so enote razvrščene po velikosti od največje do najmanjše vrednosti v tako ozkih razredih, da moremo zadosti dobro določiti tudi posamično vrednost. V našem zgledu je vsaka občina označena s tremi črkami, v načelu s prvimi tremi črkami imena občine, če pa je ime občine sestavljeno iz več besed pa iz začetnic vseh besed imena vendar tako, da skupno ne presega treh črk. Vsaka občina je obeležena tudi z zaporedno številko po velikosti urejenih enot - z rangi. Tako moremo za vsako občino odbrati rang in približno velikost podatka. Tako je npr. občina Piran (oznaka PIR) po prometu v trgovini na drobno na sedmem mestu med vsemi šestdesetimi občinami v SR Sloveniji in ima v letu 1971 med 9500 in 9750 din prometa na prebivalca. Ker je N = 60 občin, je prvi kvartil med 15. in 16. občino (med poprečjem za Tolmin in Trbovlje), dalje je promet med Ribnico in Ilirsko Bistrico enak mediani za promet na drobno na prebivalca ipd. Levo od skale so v ustrezne razrede vnešeni tudi poprečni podatki za posamezne republike in SFRJ skupno. Tako moremo napraviti tudi primerjave občinskih podatkov z republiškimi poprečji. Tak poprečen promet v trgovini na drobno na prebivalca, kot je približno v poprečju za SFRJ, imajo občine: Ljubljana-Moste, Žalec, Laško, Krško in Gornja Radgona. Enako sklepamo, da ima tri četrtine občin poprečje večje kot je jugoslovansko poprečje in le ena četrtina večje kot je poprečje za SR Slovenijo. Skladnost tretjega kvartila s poprečjem za SR Slovenijo je izraz velike asimetrije proučevane porazdelitve. Ker je informacija za vsako občino na ustreznem mestu dana z istomestno oznako: dvomesten rang, črtica, tromestna oznaka občine, daje rang grafikon kot celota



Slika 8.2 Rang grafikon prometa na prebivalca v trgovini na drobno v SR Sloveniji po občinah v letu 1971

vtis histograma. Razen nakazanih informacij o poprečnem prometu moremo iz rang grafikonov dobiti še druge informacije, ki so kombinacija naziva, ranga in vrednosti podatka za posamezno enoto.

Računanje kvantilnih rangov in kvantilov iz frekvenčnih porazdelitev

8.9 Če je populacija, ki jo proučujemo, obsežna, je razvrščanje vrednosti v ranžirno vrsto neprikladno. Za obsežne populacije pa moremo izračunati približne vrednosti kvantilnih rangov in kvantilov iz kumulativnih frekvenčnih porazdelitev. Če vemo, kaj pomenijo členi v kumulativni vrsti, spoznamo, v kakšni zvezi je kumulativna frekvenčna porazdelitev z rangi.

Proučimo frekvenčno porazdelitev o porabi lesa v 149 kmetijskih gospodarstvih v Novem mestu iz tabele 7.3. Iz nje vidimo, da ima frekvenčna porazdelitev nekatere lastnosti ranžirne vrste. V ranžirni vrsti so vrednosti, ki leže levo od dane vrednosti, manjše, desno pa večje. V frekvenčni porazdelitvi so vse vrednosti enot v danem razredu večje od vrednosti v vseh spodnjih razredih in manjše od vrednosti v vseh zgornjih razredih. Tudi v frekvenčni porazdelitvi so torej vrednosti razporejene po velikosti, le da je ranžiranje izvedeno med razredi, ne pa znotraj razredov.

Iz kumulativne frekvenčne porazdelitve moremo celo približno ugotoviti, katere vrednosti ustrezajo določenim rangom. Če preučimo kumulativno frekvenčno porazdelitev v tabeli 8.4, moremo iz nje sklepati tole:

Tabela 8.4 Kumulativna frekvenčna porazdelitev o porabi lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m^3	f_k	F_k
0,0 - 3,9	2	0
4,0 - 7,9	16	2
8,0 - 11,9	36	18
12,0 - 15,9	41	54
16,0 - 19,9	29	95
20,0 - 23,9	12	124
24,0 - 27,9	7	136
28,0 - 31,9	4	143
32,0 - 35,9	2	147
		149 = N

Dvoje kmetijskih gospodarstev je porabilo manj kot 4,0 m^3 lesa. Če vzamemo, da ima gospodarstvo z največjo porabo v tem razredu v približku porabo enako tej meji, je gospodarstvo, ki ima rang 2, porabilo 4,0 m^3. Iz kumulative dalje izvemo, da je 18 gospodarstev porabilo pod 8,0 m^3. Če vzamemo, da je največja poraba v teh 18 gospodarstvih enaka tej meji, dobimo, da ima gospodarstvo, ki je porabilo 8,0 m^3 lesa, rang 18 itd. Iz frekvenčne porazdelitve sklepamo, da imajo gospodarstva, ki so porabila toliko kot so spodnje meje razredov, range enake ustreznim vrednostim v kumulativni vrsti. Če to upoštevamo, moremo sestaviti tabelo, ki je zelo podobna ranžirni vrsti, le da so rangi dani samo za meje razredov.

Tabela 8.5 Okrnjena ranžirna vrsta za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto. (Dobljena iz kumulativne vrste v tabeli 8.4)

Rang R F_k	2	18	54	95	124	136	143	147	149
y_R y_{min}	4,0	8,0	12,0	16,0	20,0	24,0	28,0	32,0	36,0

Ker za druge enote v ranžirni vrsti ne poznamo vrednosti, predpostavljamo, da so vrednosti znotraj razredov razmeščene enakomerno. Tako dobimo vrednosti drugih enot v ranžirni vrsti z linearno interpolacijo. Grafikon kumulativne frekvenčne porazdelitve v sliki 8.3. ponazarja približen odnos med rangom R in vrednostjo znaka y. Ker tudi pri grafičnem prikazu predpostavljamo enakomerno razporeditev vrednosti znotraj razredov, so točke zvezane z daljicami.

Ker je kumulativna frekvenčna porazdelitev nadomestek za ranžirno vrsto, računamo iz nje kvantilne range in kvantile podobno kakor iz ranžirne vrste.

8.10 Oceno kvantilnega ranga P_y izračunamo po naslednjem postopku:

(a) Iz frekvenčne porazdelitve izračunamo kumulativno frekvenčno porazdelitev F_k.

(b) Poiščemo, v kateri razred pade vrednost y za katero iščemo kvantilni rang P_y. Ta razred imenujemo kvantilni razred in ga zaznamujmo z o . Zanj izpišemo iz frekvenčne porazdelitve ustrezne vrednosti:

$$y_{o,min}, i_o, f_o, F_o.$$

(c) Iz gornjih vrednosti izračunamo vrednosti y ustrezni rang R_y po obrazcu:

$$R_y = F_o + f_o \frac{y - y_{o.min}}{i_o} \qquad (8.11)$$

(d) Iz dobljenega ranga R_y pa izračunamo kvantilni rang P_y po obrazcu:

$$P_y = \frac{R_y - 0{,}50}{N} \qquad (8.12)$$

Če je obseg populacije N velik, iz obrazca 8.12 običajno izpuščamo 0,5, ker je ta količina za velike populacije nebistvena.

Postopek velja tudi za frekvenčne porazdelitve z različnimi širinami razredov.

8.11 Če za populacijo o porabi lesa iz tabele 8.4 iščemo kvantilni rang za gospodarstvo, ki je v letu 1953 porabilo $y = 13{,}8\ m^3$ lesa, izračunamo P_y po zgornjem postopku takole: Poraba lesa $y = 13{,}8\ m^3$ pade v razred 12,0–15,9. Iz tabele 8.4 odberemo:

$$y_{o,min} = 12{,}0;\ i_o = 4;\ f_o = 41;\ F_o = 54;$$

Če vstavimo te podatke v obrazec 8.11, dobimo:

$$R_y = 54 + 41 \frac{13{,}8 - 12{,}0}{4} = 72{,}4$$

in dalje po obrazcu 8.12:

$$P_{y=13{,}8} = \frac{72{,}4 - 0{,}5}{149} = 0{,}482$$

Gospodarstvo s porabo $y = 13{,}8\ m^3$ lesa je v 48. centilu.

8.12 Podobno računamo iz frekvenčne porazdelitve tudi k v a n t i l e . Postopek je naslednji:

(a) Iz frekvenčne porazdelitve izračunamo kumulativno frekvenčno porazdelitev F_k.

(b) Iz danega P izračunamo ustrezni rang R_p po obrazcu

$$R_p = NP + 0{,}5 \qquad (8.12)$$

Če je populacija velika, v obrazcu 8.13 izpustimo 0,5.

c) V kumulativni frekvenčni porazdelitvi F_k poiščemo, med kateri vrednosti kumulativne vrste pade R, tako da je: $F_o \leq R_p < F_1$. F_o ustrezen razred je kvantilni razred. Zanj poiščemo v frekvenčni porazdelitvi količine:

$$y_{o,min}, i_o, f_o \text{ in } F_o .$$

d) Iz teh količin izračunamo ustrezni kvantil y_p po obrazcu:

$$y_p = y_{o,min} + i_o \frac{R_p - F_o}{f_o} \qquad (8.13)$$

Enako kot za kvantilne range velja nakazani postopek tudi za frekvenčne porazdelitve z različnimi širinami razredov.

8.13 Da za naš zgled izračunamo kvartile, je najbolje, da sistematično izračunamo v računski tabeli vse tri kvartile hkrati.

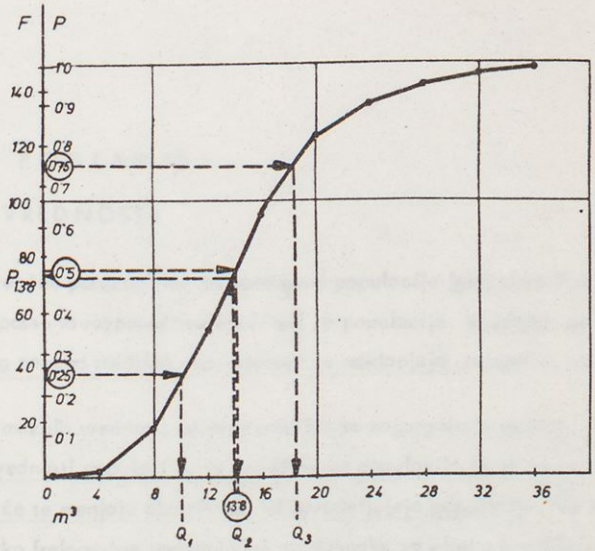
Tabela 8.6 Izračunavanje kvartilov za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto.

i	P	R_p	$y_{0,min}$	i_o	f_o	F_o	$R_p - F_o$	$i_o \frac{R_p - F_o}{f_o}$	Q_i
1	0,25	37,75	8,0	4	36	18	19.75	2,19	$10,19 = Q_1$
2	0,50	75,00	12,0	4	41	54	21.00	2,05	$14,05 = Q_2$
3	0,75	112,25	16,0	4	29	95	17,25	2,38	$18,38 = Q_3$

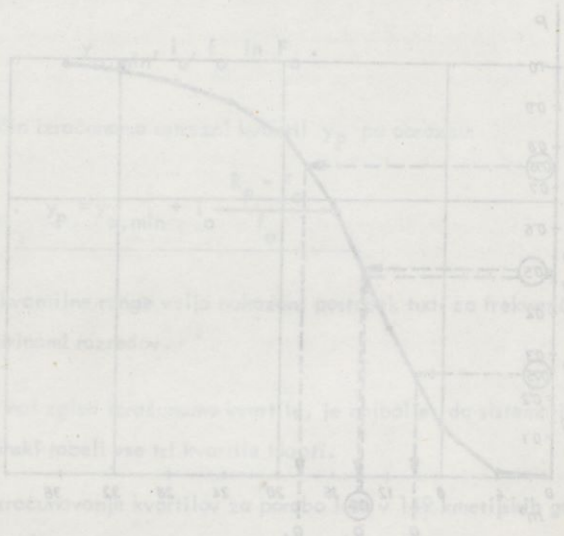
8.14 Podobno kakor v sliki 8.1 za negrupirane podatke, moremo grafično oceniti vse zgornje količine iz grafikona kumulativne vrste. V sliki 8.3 je nakazano, kako dani vrednosti $y = 13,8\ m^3$ poiščemo ustrezni kvantilni rang in obratno: kako moremo grafično oceniti kvartile.

8.15 V praksi večkrat izdelujemo decilne ali centilne norme, s katerimi dobimo neposredno zvezo vrednosti znaka z decili oziroma pri podrobnejših normnih tablicah

s centili. Centilne norme s pridom uporabljamo v psihologiji, moremo pa jih izdelati tudi za določene socialno-ekonomske probleme. Tako tablica decilnih ali centilnih norm o produktivnosti dela pomaga presoditi, kakšna je produktivnost dela za posameznega delavca v razmerju s celotnim kolektivom itd.



Slika 8.3 Grafično ocenjevanje kvantilnih rangov in kvantilov iz slike kumulativnih frekvenčnih porazdelitev za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu v letu 1953



DEVETO POGLAVJE

SREDNJE VREDNOSTI

9.1 Frekvenčne porazdelitve za homogene populacije (glej slike 7.3, 7.6, 7.7) kažejo posebno razporeditev vrednosti za populacije. Vrednosti za te populacije se goste okrog nekega središča, od katerega se odklanjajo navzgor in navzdol.

Odkloni posameznih vrednosti od tega središča so najpogosteje majhni. To središče, ki predstavlja vrednosti populacije, je značilno za populacijo in ne za posamezne enote in se menja, če se menjajo okoliščine, ki opredeljujejo populacijo. Če primerjamo za določeno stroko frekvenčne porazdelitve za dohodke za visoko kvalificirane, kvalificirane in proučene delavce, opazimo, da so središča - **srednje vrednosti** za posamezne populacije različna. Dohodki za visoko kvalificirane delavce se goste najvišje, dohodki kvalificiranih delavcev nižje in dohodki priučenih delavcev najnižje. Ta središča so torej za posamezne populacije določene z opredeljujočimi pogoji populacije, ti pa so za vse enote enaki. Če bi na dohodek za posameznega delavca vplivala samo kvalifikacija, bi imeli vsi delavci z enako kvalifikacijo enake dohodke. Zaradi drugih - posamičnih vplivov, ki so za vsakega delavca različni, pa se dohodki odklanjajo navzgor in navzdol od dohodkov, ki so pogojeni s kvalifikacijo.

Čim manjši so posamični vplivi, tem bolje srednja vrednost predstavlja vrednosti populacije in obratno: čim večji so posamični vplivi, tem slabše srednja vrednost predstavlja vrednosti populacije. Zato je srednja vrednost predstavnik vrednosti v populaciji le, če je populacija enovita.

9.2 Parametrov, ki pokažejo osrednjo težnjo vrednosti populacije, poznamo več. Od teh so v socialno-ekonomski statistiki pomembne naslednje srednje vrednosti:

a) m e d i a n a, - Me
b) m o d u s , - Mo
c) a r i t m e t i č n a s r e d i n a, - M
d) h a r m o n i č n a s r e d i n a , - H
e) g e o m e t r i j s k a s r e d i n a , - G

Medtem ko sta mediana in modus dana z lego vrednosti, štejemo aritmetično, harmonično in geometrijsko sredino med izračunane srednje vrednosti.

MEDIANA

9.3 Za srednjo vrednost more dobro rabiti mediana, ki jo poznamo že iz poglavja o kvantilih. Mediana je vrednost, ki ustreza kvantilnemu rangu P = 0,50,

$$Me = y_{P=0,50}$$

Mediana je uporabna srednja vrednost, ker nujno leži sredi posamičnih vrednosti. Po opredelitvi ima namreč polovica enot manjše, polovica pa večje vrednosti kakor je mediana.

Določanje mediane ne bomo obravnavali posebej, ker se sklada z določanjem kvantilov na splošno. Te pa smo obravnavali že v odstavku o kvantilih.

9.4 L a s t n o s t i m e d i a n e . Prednost mediane je predvsem v tem, da je lahko razumljiva srednja vrednost in zato kot opisni parameter priporočljiva. Velika prednost mediane pred izračunanimi sredinami je tudi ta, da za določanje mediane ni nujno, da poznamo vrednosti za vse enote populacije. Zadosti je, da poznamo vrednosti za enote, ki leže okoli sredine v ranžirni vrsti. Ta lastnost mediane pride posebno prav, če iščemo srednje vrednosti za frekvenčne porazdelitve, ki imajo odprte razrede.

Za odprte razrede namreč dostikrat niti približno ne vemo, kakšne vrednosti vsebujejo, ker so omejeni samo navzgor ali samo navzdol. Mediana je primerna srednja vrednost tudi takrat, če so skrajne vrednosti take, da sumimo, ali sploh sodijo v homogeno populacijo, ker so mogoče izraz nekih drugih kakovosti.

Pomanjkljivost mediane pa je v tem, da je le preveč neobčutljiva za spremembe vrednosti in se njena vrednost ne spremeni vse dotlej, dokler so spremembe take, da vrednosti ne preidejo iz ene polovice v drugo.

Pomembna lastnost mediane je tudi, da je vsota absolutnih odklonov posameznih vrednosti najmanjša, če odklone računamo od mediane. Mediana je torej sorazmeroma dober predstavnik posameznih vrednosti

$$\sum_{i=1}^{N} |y_i - A| = \min \; ; \quad \text{če je } A = Me \tag{9.1}$$

MODUS

9.5 Mediana pa ni vedno dober predstavnik vrednosti populacije. Za asimetrične ali polimodalne porazdelitve je mediana vrednost, ki je različna od večine vrednosti v populaciji. Če naj bo srednja vrednost predstavnik populacije, je v takih primerih veliko primerneje, da vzamemo za sredino vrednost, okrog katere se vrednosti populacije najbolj goste. Če pogledamo poligone frekvenčnih porazdelitev, moremo za vsako frekvenčno porazdelitev zlahka ugotoviti približno mesto največje gostitve. Vrednost, okrog katere se vrednosti populacije najbolj goste, imenujemo m o d u s ali n a j p o g o s t e j š o v r e d n o s t .

Modus moremo ugotoviti samo za razmeroma obsežne populacije, ki so grupirane v frekvenčni porazdelitvi. Iz posamičnih podatkov je modus neposredno nemogoče ugotoviti.

Če imamo frekvenčno porazdelitev dano s frekvenčno krivuljo, poiščemo modus enostavno. V tem primeru je modus abscisa tiste točke na frekvenčni krivulji, za katero je ordinata (gostota frekvence) največja.

9.6 Računanje modusa iz frekvenčne porazdelitve. Za populacije običajno nimamo frekvenčnih krivulj, temveč frekvenčne porazdelitve, ki jih mo-

remo narisati v histogramu. Ker histogram nakaže približno obliko frekvenčne krivulje, moremo tudi iz frekvenčne porazdelitve oziroma histograma sklepati, kje je gostota frekvence največja. Iz histograma za frekvenčno porazdelitev z enakimi razdeli sklepamo, da je modus v razredu, v katerem je frekvenca največja. Za porabo lesa v 149 kmetijskih gospodarstvih v letu 1953 v Novem mestu iz tabele 7.3 sklepamo, da je najpogostejša vrednost v razredu od 12,0 do 15,9 m^3 lesa. Kot prvi približek modusa vzamemo kar sredino razreda z največjo frekvenco - sredino modalnega razreda. V našem zgledu je prva ocena modusa za porabo lesa $Mo = 14\ m^3$. Sredina razreda pa je prava vrednost modusa le, če je porazdelitev simetrična. Drugače leži modus nad sredino razreda ali pod njo, odvisno od oblike frekvenčne porazdelitve.

Natančnejšo oceno za modus dobimo, če upoštevamo tudi frekvence v sosednih razredih. Iz histograma o porabi lesa v sliki 7.1 sklepamo, da je modus za porabo lesa verjetno pod sredino modalnega razreda s sredino 14 m^3. Ker je frekvenca pred modalnim razredom večja kot frekvenca za modalnim razredom, bi bila frekvenčna krivulja, ki bi jo načrtali med stolpce histograma, nagnjena proti razredu z višjo frekvenco. Predpostavimo, da je v okolici modusa frekvenčna krivulja v približku parabola druge stopnje, ki gre skozi točke s koordinatama $(y_{-1} \cdot f_{-1})$, $(y_o \cdot f_o)$ in $(y_{+1} \cdot f_{+1})$. Pri tem pomenijo y_{-1}, y_o, y_{+1} sredine modalnega razreda in njegovih sosedov, f_{-1}, f_o in f_{+1} pa ustrezne frekvence. Pri tej predpostavki izračunamo modus po naslednjem postopku:

a) Podatke imamo grupirane v frekvenčni porazdelitvi z razredi z enako širino i. Razredi morajo biti tako veliki, da frekvenčna porazdelitev izraža zakonitost gostitve frekvenc.

b) V frekvenčni porazdelitvi poiščemo razred z največjo frekvenco $f_{-1} < f_o > f_{+1}$. Razred z največjo frekvenco (o) imenujmo m o d a l n i r a z r e d .

c) Modus izračunamo po obrazcu

$$M_o = y_{o.min} + i \frac{f_o - f_{-1}}{2f_o - f_{-1} - f_{+1}} \qquad (9.2)$$

Pri tem pomeni razen že navedenih izrazov: $y_{0,min}$ = spodnja meja za modalni razred.

9.7 Za zgled o porabi lesa v 149 kmetijskih gospodarstvih v letu 1953 v Novem mestu je od treh porazdelitev, ki smo jih sestavili, za izračunavanje modusa prikladna

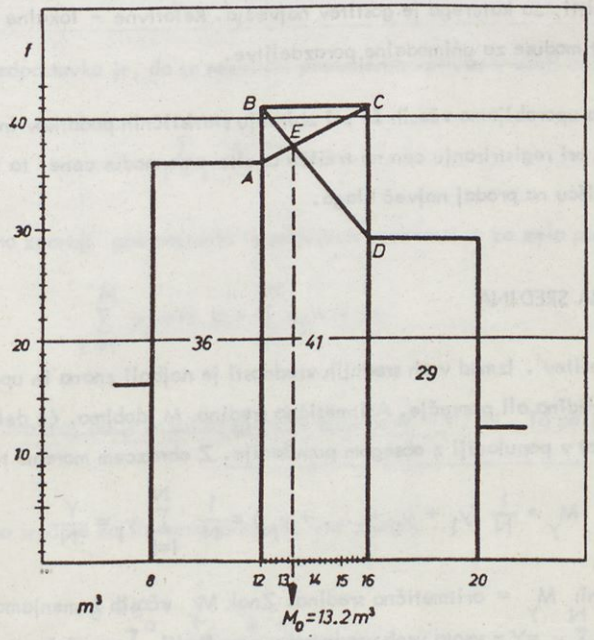
porazdelitev, v kateri je razredna širina i = 4 m^3. V porazdelitvi z i = 1 m^3 se zakonitost gostitve še ne kaže, v porazdelitvi z i = 8 m^3 pa je zaradi prevelike širine razredov zakonitost gostitve že zabrisana.

Iz frekvenčne porazdelitve v tabeli 7.3 sklepamo, da je modalni razred razred 12,0 - 15,9 m^3, ker ima največjo frekvenco (f_o = 41). Ker je za to porazdelitev f_{-1} = 36, f_{+1} = 29; i = 4; $y_{o.min}$ = 12,0, je modus po obrazcu 9.2

$$Mo = 12,0 + 4 \cdot \frac{41 - 36}{2.41 - 36 - 29} = 13,18\ m^3$$

Kot smo pričakovali, je druga ocena za modus Mo = 13,18 m^3 pod prvo oceno 14,0 m^3.

9.8 Modus pa moremo ob istih predpostavkah zelo enostavno oceniti iz histograma za frekvenčno porazdelitev tudi grafično. Kakor kaže slika 9.1, zvežemo v histogramu



Slika 9.1 Grafična določitev modusa iz frekvenčne porazdelitve za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu

oglišče A s C, oglišče B z D. Projekcija sečišča obeh veznic E na abscisi je ocena modusa Mo. Slika pokaže, da grafičen način dá isti rezultat kakor račun.

9.9 Lastnosti modusa. Modus je srednja vrednost, ki dobro predstavlja vrednosti za populacijo, ker je že po opredelitvi vrednost, ki se v populaciji najpogosteje pojavlja. Enako kakor mediana je modus srednja vrednost, ki je dana z lego vrednosti populacije. Zato je tudi modus neobčutljiv za spremembe vrednosti posameznih enot vse dotlej, dokler gostitev na nekem drugem mestu ne prekorači stopnje gostitve v modusu. To moremo šteti za dobro lastnost modusa, ker ni odvisen od vrednosti, ki za populacijo niso tipične, more pa biti tudi hiba, ker je modus le premalo odvisen od posamičnih vrednosti.

Nehomogena populacija more imeti tudi več mest gostitve. Bimodalne porazdelitve imajo dva, polimodalne porazdelitve pa celo več središč gostitve. Take porazdelitve imajo torej dva ali več relativnih - lokalnih modusov. Od teh pa je absolutni modus tisti, za katerega je gostitev največja. Relativne - lokalne moduse ocenjujemo enako kot moduse za unimodalne porazdelitve.

Pojem modusa uporabljamo včasih že pri zbiranju statističnih podatkov in ne samo pri obdelavi. Tako pri registriranju cen na tržišču upoštevamo modus cene, to je ceno, po kateri je na tržišču na prodaj največ blaga.

ARITMETIČNA SREDINA

9.10 Opredelitev. Izmed vseh srednjih vrednosti je najbolj znana in uporabljana aritmetična sredina ali poprečje. Aritmetično sredino M dobimo, če delimo vsoto vrednosti vseh enot v populaciji z obsegom populacije. Z obrazcem moremo to izraziti

$$M_y = \frac{1}{N}(y_1 + y_2 + \ldots + y_N) = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{Y}{N} \qquad (9.3)$$

Pri tem pomeni: M_y = aritmetična sredina. Znak M_y včasih zamenjamo z znakom $\bar{y}$ (y prečna); $\sum_{i=1}^{N} y_i = Y$ = vsota vseh vrednosti v populaciji; $\sum$ = splošen znak za seštevanje izraza, ki stoji za njim; y_i = posamične vrednosti. Nakazane oznake uporabljamo na splošno.

9.11 Za zgled vzemimo podatke o številu izdelkov, ki jih je skupina desetih delavcev izdelala v eni izmeni. Če so posamični podatki:

38, 35, 42, 47, 39, 48, 32, 42, 44, 50, je poprečno število izdelkov po obrazcu 9.3

$$M = \frac{38 + 35 + \ldots\ldots\ldots\ldots\ldots\ldots + 44 + 50}{10} = 41.7 \text{ kosov}$$

Poprečno število M = 41,7 pove, da bi moral vsak delavec izdelati po 41,7 kosov, če bi hoteli, da bi bila skupna proizvodnja enaka stvarni.

9.12 Lastnosti aritmetične sredine

a) Aritmetična sredina je izpeljana ob predpostavki, da je vrednost y za posamezno enoto vsota rezultatov splošnih (M) in posamičnih vplivov (e)

$$y_i = M + e_i \qquad (9.4)$$

Nadaljnja predpostavka je, da se rezultati posamičnih vplivov v vsoti uničijo

$$\sum_{i=1}^{N} e_i = o.$$

Če upoštevamo zgornji predpostavki in seštejemo vrednosti y_i za celo populacijo, dobimo

$$\sum_{i=1}^{N} y_i = N.M + \sum_{i=1}^{N} e_i = N.M$$

Iz te enačbe dobimo, da je rezultat splošnih vplivov $M = Y/N$. To pa je aritmetična sredina.

b) Aritmetična sredina za linearno zvezo iz več znakov

$$u = a_o + \sum_{k=1}^{r} a_k y_k \qquad (9.5)$$

je enaka linearni zvezi iz aritmetičnih sredin

$$u = a_o + \sum_{k=1}^{r} a_k y_k \; ; \quad M_u = a_o + \sum_{k=1}^{r} a_k M_k \qquad (9.6)$$

Dokaz je dokaj enostaven. Če enačbo

$$u = \sum_{i=1}^{N} u_i = \sum_i (a_o + \sum_k a_k y_{ki}) = a_o N + \sum_k a_k \sum_i y_{ki} = a_o N + \sum_k a_k Y_k$$

ki jo dobimo, če seštejemo linearne enačbe 9.5 za vse enote populacije, delimo z obsegom populacije N, dobimo zvezo 9.6. Kot poseben primer zveze 9.6 je

$$M_a = a \qquad (9.7)$$

Poprečje konstante je enako konstanti in

$$M_{by} = b.M_y \qquad (9.8)$$

poprečje znaka, pomnoženega s konstanto, je enako produktu konstante s poprečjem znaka.

Vzemimo za zgled, da je skupni dohodek D delavca sestavljen iz osnovnega dela 0, dohodka iz rednega delovnega časa R, za katere je urni dohodek r, in izrednega dela, ki ga je opravljal P ur po urnem dohodku p

$$D = 0 + R \cdot r + P \cdot p$$

Po gornjem stavku je poprečni dohodek na delavca

$$M_D = M_o + R M_r + P M_p$$

c) Vsota odklonov posamičnih vrednosti y_i od aritmetične sredine M_y je nič

$$\sum_{i=1}^{N} (y_i - M_y) = 0 \qquad (9.9)$$

Dokaz je preprost. Ker je vsota razlik enaka razliki vsot, spoznamo, da je

$$\sum y_i - N \cdot M_y = 0$$

v skladu z opredelitvijo aritmetične sredine

d) Vsota kvadratov odklonov posamičnih vrednosti y_i od neke konstante A je najmanjša, če je A enak aritmetični sredini M_y

$$SK = \sum_{i=1}^{N} (y_i - A)^2 : \quad \text{če je: } A = M_y \tag{9.10}$$

Dokaz: Če naj bo izraz SK minimalen, mora biti $\frac{dSK}{dA} = 0$

$$\frac{dSK}{dA} = -2 \sum_i (y_i - A) = 0; \quad \sum (y_i - A) = \sum y_i - NA = 0; \quad A = \frac{1}{N} \sum y_i = M_y$$

Ta lastnost za aritmetično sredino je zelo važna za nadaljnje proučevanje populacij. Iz vsote kvadratov odklonov od aritmetične sredine je namreč izpeljan poseben parameter, v a r i a n c a , ki je najvažnejša mera jakosti za posamične vplive.

e) Sumarno aritmetično sredino populacije M izračunamo iz aritmetičnih sredin M_k delnih populacij z obsegi N_k po obrazcu

$$M = \frac{N_1M_1 + N_2M_2 + \ldots + N_rM_r}{N_1 + N_2 + \ldots + N_r} = \frac{\sum_{k=1}^{r} N_kM_k}{\sum_{k=1}^{r} N_k} = \frac{1}{N} \sum_k N_kM_k \tag{9.11}$$

Ta način za računanje aritmetične sredine imenujemo t e h t a n o r a č u n a n j e za razliko od enostavnega računanja po obrazcu 9.3. Dokaz: Skupna vsota Y je enaka vsoti delnih vsot Y_k za y

$$Y = Y_1 + Y_2 + \ldots + Y_r = \sum_k Y_k$$

Ker velja za skupno populacijo in delne populacije $Y = NM_y$, je dalje:

$$NM = N_1M_1 + N_2M_2 + \ldots + N_rM_r = \sum_k N_kM_k$$

Če delimo to enačbo z N dobimo obrazec 9.11. Splošnejše moremo tehtan način za izračunanje aritmetične sredine pisati

$$M = \frac{\sum_k w_k M_k}{\sum_k w_k} \qquad (9.12)$$

Pri tem pomeni: w_k = teža - ponder

Računanje aritmetične sredine iz frekvenčnih porazdelitev

9.13 Neposredno računanje aritmetične sredine po osnovnem obrazcu 9.3 je za velike populacije zaradi velikega števila sumandov zamudno. Ker daje frekvenčna porazdelitev približno sliko vseh vrednosti populacije, moremo aritmetično sredino oceniti iz frekvenčne porazdelitve.

Če predpostavljamo, da sredina razreda y_k v frekvenčni porazdelitvi predstavlja vrednosti v posameznem razredu, dobimo oceno za vsoto vrednosti v tem razredu, če sredino razreda pomnožimo s frekvenco f_k. Oceno za vsoto vrednosti za celo populacijo pa dobimo, če produkte $f_k y_k$ za vse razrede seštejemo. Če to uporabimo, računamo oceno aritmetične sredine iz frekvenčne porazdelitve po obrazcu

$$M_y = \frac{f_1 y_1 + f_2 y_2 + \ldots + f_r y_r}{f_1 + f_2 + \ldots + f_r} = \frac{\sum f_k y_k}{\sum f_k} = \frac{1}{N} \sum_{k=1}^{r} f_k y_k \qquad (9.13)$$

Ta obrazec je samo posebna oblika splošnega obrazca 9.12 za tehtan način računanja aritmetične sredine. Pri tem so frekvence t e ž e ali p o n d e r i .

Po obrazcu 9.13 dobljena ocena je tem boljša, čim ožji so razredi, vendar dobimo tudi pri razmeroma širokih razredih še vedno zadovoljive rezultate, če je frekvenčna porazdelitev unimodalna in ne preveč asimetrična. Za zelo asimetrične porazdelitve, posebno za porazdelitve tipa J, pa daje tako računanje sistematično napačne ocene. Če je porazdelitev asimetrična v desno, dobimo sistematično prevelike ocene, pri asimetriji v levo pa sistematično premajhne ocene.

Računanje aritmetične sredine po obrazcu 9.13 velja za porazdelitve z enakimi in z neenakimi razredi. Ne moremo pa po tem obrazcu ocenjevati aritmetične sredine za frekvenčne porazdelitve z odprtimi razredi, ker zanje ne moremo izračunati sredine razredov. Če za odprt razred poznamo razen frekvence f_r še vsoto vrednosti v odprtem razre-

du Y_r, izračunamo oceno za aritmetično sredino po obrazcu

$$M_y = \frac{1}{N}(f_1y_1 + f_2y_2 + \ldots + f_{r-1}y_{r-1} + Y_r) \qquad (9.14)$$

9.14 Računanje ocene aritmetične sredine po neposredni metodi je v tabeli 9.1 nakazano za frekvenčno razdelitev za porabo lesa v letu 1953 za kmetijska gospodarstva v Novem mestu.

Tabela 9.1 Računanje ocene aritmetične sredine za potrošnjo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto v letu 1953 po neposredni metodi.

Poraba v m^3	f_k	y_k	$f_k y_k$
0,0 - 3,9	2	2	4
4,0 - 7,9	16	6	96
8,0 - 11,9	36	10	360
12,0 - 15,9	41	14	574
16,0 - 19,9	29	18	522
20,0 - 23,9	12	22	264
24,0 - 27,9	7	26	182
28,0 - 31,9	4	30	120
32,0 - 35,9	2	34	68
	$N = 149$	$\sum f_k y_k$	$= 2190$

Po obrazcu 9.13 je $M_y = 2190/149 = 14{,}70\ m^3$.

Če primerjamo dobljeni rezultat s pravo aritmetično sredino $M_y = 2182{,}1/149 = 14{,}64\ m^3$, izračunano iz osnovnih podatkov v tabeli 7.1, vidimo, da je razlika malenkostna.

9.15 Pomožni znak u. Za frekvenčne porazdelitve, ki imajo enake razrede, računanje aritmetične sredine poenostavimo, če vpeljemo namesto sredin razredov y_k pomožni znak u_k, ki je z osnovnim znakom y_k v naslednji linearni zvezi:

$$y_k = y_o + i \cdot u_k \qquad (9.15)$$

Pri tem pomeni: y_o = sredina razreda, ki je približno v sredini frekvenčne porazdelitve

ali blizu razreda z največjo frekvenco, i = širina razreda.

Če natančneje proučimo vrednosti znaka u, spoznamo, da sredinam razredov v posameznih razredih ustrezajo u ... - 3, -2, -1, 0, +1, +2, +3... Pri tem je vrednost u = o v razredu, ki mu ustreza y_o.
Zaradi obrazca 9.13 in 9.6 pa velja

$$M_y = y_o + iM_u \qquad (9.16)$$

Ta postopek je prikladnejši kakor neposredna metoda, ker poenostavi množenje. Po tem postopku izračunamo aritmetično sredino po tehle točkah.

a) V dani frekvenčni porazdelitvi izberemo nekje v sredini ali v razredu, okoli katerega so frekvence največje, izhodišče za pomožni znak u. Glede na to poljubno izhodišče postavimo v posamezne razrede ustrezne vrednosti pomožnega znaka u

$$\ldots -3,\ -2,\ -1, 0, +1, +2, +3 \ldots..$$

b) Pomnožimo frekvence f_k z ustreznimi vrednostmi u_k. Tako dobimo produkte $f_k u_k$.

c) Seštejemo dobljene produkte in vsoto $\sum f_k u_k$ vnesemo v obrazec

$$M_y = y_o + i \cdot \frac{1}{N} \sum f_k u_k \qquad (9.17)$$

9.16 Za porabo lesa v Novem mestu je izračunanje aritmetične sredine s pomožnim znakom u nakazano v tabeli 9.2.

Tabela 9.2 Računanje ocene za aritmetično sredino za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi pomožnega znaka u.

Poraba v m^3 y	f_k	u_k	$f_k u_k$
0,0 - 3,9	2	- 3	- 6
4,0 - 7,9	16	- 2	- 32
8,0 - 11,9	36	- 1	- 36
12,0 - 15,9	41	0	0
16,0 - 19,9	29	+ 1	+ 29
20,0 - 23,9	12	+ 2	+ 24
24,0 - 27,9	7	+ 3	+ 21
28,0 - 31,9	4	+ 4	+ 16
32,0 - 35,9	2	+ 5	+ 10
	N = 149		$\sum fu$ = + 26

Ker je $y_o = 14,0$, $i = 4$, $N = 149$ in $\sum fu = +26$, dobimo po obrazcu 9.17

$$M_y = 14,0 + 4 \cdot \frac{1}{149} (+ 26) = 14,70 \text{ m}^3$$

Rezultat se sklada z rezultatom, ki smo ga dobili po neposredni metodi.

9.17 Metoda kumulativ. Če so podatki grupirani v frekvenčni porazdelitvi z enakimi razredi, aritmetično sredino enostavno izračunamo s kumulativami. Pri tej metodi odpade vmesno množenje. Razen tega pa imamo kot postranski rezultat izračunano kumulativno frekvenčno porazdelitev; ta pa je važna sama zase ali za računanje kvantilov.

Po metodi kumulativ izračunamo aritmetično sredino po naslednjem postopku:

a) Iz frekvenčne porazdelitve izračunamo iz f_k kumulativno frekvenčno porazdelitev F_k.

b) Seštejemo člene v kumulativni vrsti, razen zadnjega, ki leži pod črto, ki pomeni obseg populacije N. Vsoto členov iz kumulativne vrste zaznamujemo s C.

c) Aritmetično sredino M_y izračunamo iz dobljenih podatkov po obrazcu

$$M_y = y_o - i \cdot \frac{C}{N} \qquad (9.17)$$

Pri tem pomeni: y_o = sredina zadnjega razreda v frekvenčni porazdelitvi; i = širina razreda; C = vsota členov v kumulativni vrsti; N = obseg populacije.

9.18 Aritmetična sredina za porabo lesa v letu 1953 v Novem mestu je izračunana po metodi kumulativ v tabeli 9.3.

Tabela 9.3 Izračunanje aritmetične sredine za porabo lesa v 149 kmetijskih gospodarstvih v letu 1953 po metodi kumulativ.

Poraba v m^3	f	F
0,0 - 3.9	2	0
4,0 - 7,9	16	2
8,0 - 11,9	36	18
12,0 - 15,9	41	54
16,0 - 19,9	29	95
20,0 - 23,9	12	124
24,0 - 27,9	7	136
28,0 - 31,9	4	143
32,0 - 35,9	2	147
	N = 149	2+18+54+...+143+147= = 719 = C

Ker je sredina zadnjega razreda $y_o = 34$, dobimo po obrazcu 9.17

$$M_y = 34,0 - 4.719/149 = 14,70\ m^3$$

Dobljeni rezultat se sklada z rezultatom, ki smo ga dobili po prejšnjih metodah.

Modificirana aritmetična sredina

9.19 Aritmetična sredina je odvisna od vseh vrednosti populacije. Zato nanjo vplivajo tudi skrajne vrednosti, ki so včasih rezultat izjemnih okoliščin in jih ne moremo šteti, da sodijo v prikazano populacijo. Te vrednosti pačijo sliko populacije in tudi aritmetično sredino populacije. Zato za take primere izračunamo sredino, ki je nekak kompromis med aritmetično sredino in mediano.

Iz ranžirne vrste dobimo mediano, če postopoma paroma izpuščamo po en spodnji in en zgornji člen v ranžirni vrsti. Na koncu tega postopka ostane samo en člen - mediana, ali dva člena, katerih sredino štejemo za mediano. Kompromisna rešitev pa je v tem, da izpuščamo po en zgornji in en spodnji člen le toliko časa, dokler ne izključimo vseh netipičnih vrednosti, iz ostanka pa izračunamo aritmetično sredino. Tako odstranimo vpliv skrajnih - izjemnih vrednosti, ki včasih občutno vplivajo na aritmetično sredino. Kljub

vsemu pa je ta sredina izračunana iz večine vrednosti populacije. Modificirano aritmetično sredino dobimo torej tako, da iz ranžirne vrste posamičnih vrednosti odstranimo po en, dva ali več parov skrajnih - izjemnih vrednosti, iz ostalih pa izračunamo poprečje. Ta postopek uporabljamo pogosto pri analizi časovnih vrst. Zaradi izjemnih razmer v določenih razdobjih nekateri podatki pačijo tipičnost aritmetične sredine. Če pa teh skrajnih vrednosti ne upoštevamo in uporabimo modificirano poprečje, dobimo realnejšo sliko.

Aritmetična sredina aritmetičnih sredin

9.20 Če poznamo aritmetične sredine M_k in obsege N_k za delne populacije, ki sestavljajo populacijo, iz teh podatkov izračunamo aritmetično sredino M za populacijo po obrazcu

$$M = \frac{\sum N_k M_k}{\sum N_k} \qquad (9.18)$$

Ta obrazec smo dokazali v odstavku o lastnostih aritmetičnih sredin. Lastnosti, da moremo izračunati skupno srednjo vrednost, če poznamo ustrezne srednje vrednosti za delne populacije, nimata niti mediana niti modus. Pri obeh je treba iz delnih populacij sestaviti skupno populacijo in iz nje poiskati mediano ali modus.

Zgornje računanje za aritmetično sredino imenujemo t e h t a n ali p o n d e r i r a n n a č i n, ker je iz delnih sredin izračunana skupna sredina tako, da upoštevamo velikost- t e ž o ali p o n d e r za posamezno delno aritmetično sredino. Obrazec 9,18 je poseben primer splošnega obrazca za računanje tehtane aritmetične sredine

$$M_R = \frac{\sum w_k R_k}{\sum w_k} \qquad (9.19)$$

Pri tem pomeni: M_R = tehtana aritmetična sredina količin R_k, w_k = ponder-teža za posamezne vrednosti R_k.

9.21 V tabeli 9.4 je nakazano računanje skupnih poprečnih mesečnih prejemkov delavcev v elektrogospodarstvu, če poznamo poprečne mesečne prejemke in skupno število delavstva po vrstah podjetij.

Tabela 9.4. Računanje skupnih poprečnih mesečnih prejemkov zaposlenega osebja v proizvodnji, prenosu in distribuciji električne energije.

Stroka	Število zaposlenih N_k	Ø dohodki IX - 1970 M_k	$N_k M_k$
Hidroelektrarne	895	2349	2102355
Termoelektrarne	1292	1736	2242912
Podjetje za prenos elektroenergije	170	2211	375870
Podjetje za razdelitev elektroenergije	3740	1738	6500120
Skupno	$\sum N_k = 6097$	1840	$\sum M_k N_k = 11221257$

Poprečni dohodek za stroko je

$$M = \frac{\sum N_k M_k}{\sum N_k} = \frac{11221257}{6097} = 1840 \text{ din}$$

Popolnoma nepravilno bi bilo, če bi izračunali navadno aritmetično sredino iz podatkov o mesečnih prejemkih po kvalifikaciji. Rezultat, ki bi ga dobili,

$M_{nav} = 1/4\ (2349 + 1736 + 2211 + 1738) = 2008.5$ je seveda različen od zgornjega in brez logičnega smisla, medtem, ko pravo sumarno poprečje pove, kakšni bi bili prejemki delavcev v elektrogospodarstvu, če bi imeli vsi enake prejemke pri istem fondu dohodkov.

Sumarna poprečja pa moramo uporabljati zelo previdno. V splošnem sumarne aritmetične sredine ne predstavljajo vrednosti v populaciji. Sumarno poprečje more biti celo vrednost, ki jo ima malo število enot. Vendar kljub hibam dostikrat izračunavamo sumarna poprečja, ki so, če jih pravilno tolmačimo, dobro sredstvo za analizo.

HARMONIČNA SREDINA

9.22 Med izračunane srednje vrednosti štejemo tudi harmonično sredino, ki je po definiciji recipročna vrednost aritmetične sredine reciprokov iz osnovnih podatkov. Po

tej opredelitvi je harmonična sredina H enaka

$$H = \frac{N}{\frac{1}{y_1} + \frac{1}{y_2} \ldots + \frac{1}{y_N}} = \frac{N}{\sum \frac{1}{y_i}} \qquad (9.20)$$

Tehtana harmonična sredina pa podobno

$$H = \frac{w_1 + w_2 + \ldots + w_r}{\frac{w_1}{y_1} + \frac{w_2}{y_2} + \ldots + \frac{w_r}{y_r}} = \frac{\sum w_k}{\sum \frac{w_k}{y_k}} \qquad (9.21)$$

9.23 Če iz istih podatkov izračunamo aritmetično sredino in harmonično sredino, dobimo različne rezultate. Za podatke 1,2,4,7,9, je aritmetična sredina enaka

$$M = \frac{1 + 2 + 4 + 7 + 9}{5} = 4,6,$$

harmonična sredina pa

$$H = \frac{5}{\frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{7} + \frac{1}{9}} = 2,5$$

Iz shematičnega zgleda vidimo, da so razlike znatne.

Harmonično sredino uporabljamo redkeje kakor aritmetično. Kadar se vrednosti populacije porazdeljujejo v asimetrični porazdelitvi, a tako, da je porazdelitev recipročnih vrednosti simetrično, pa harmonična sredina bolje kaže osrednjo težnjo kakor aritmetična sredina. Zato v takih primerih dajemo prednost harmonični sredini.

9.24 S harmonično sredino računamo včasih tudi poprečja iz relativnih števil.

Vzemimo, da je pet delavcev delalo po eno uro. Pri tem so dosegli naslednjo proizvodnost dela: 4,5,2,6,3 minute za en izdelek. Proizvodnost dela merimo s časom, ki so ga potrebovali za izdelavo enega izdelka.

Sumarni pokazovalec o produktivnosti dela za vseh pet delavcev dobimo, če skupno po-

rabljen čas delimo s številom proizvedenih izdelkov. Vseh pet delavcev je delalo skupno 5.60 = 300 minut. Število izdelkov, ki jih je izdelal posamezen delavec, dobimo, če za vsakega delavca čas (60 minut) delimo s časom, ki ga je potreboval za izdelavo enega izdelka (pokazovalec produktivnosti dela). Za posamezne delavce dobimo po vrsti:

$60/4 = 15$; $60/5 = 12$; $60/2 = 30$; $60/6 = 10$; $60/3 = 20$.

Skupno število izdelkov pa dobimo, če te podatke seštejemo. Poprečna produktivnost dela je torej

$$\frac{5.60}{60/4 + 60/5 + 60/2 + 60/6 + 60/3} = \frac{300}{87} = 3{,}45 \text{ minut/izdelek}$$

Iz računa vidimo, da smo izračunali poprečno proizvodnost dela s harmonično sredino. Če v gornjem računu krajšamo s ponderom 60, dobimo

$$\frac{5}{1/4 + 1/5 + 1/2 + 1/6 + 1/3} = 3{,}45 \text{ minut/izdelek}$$

Ker so ponderi med seboj enaki, dobimo isti rezultat, če vzamemo tehtano ali netehtano obliko.

Če iz gornjih podatkov izračunamo aritmetično sredino, pa dobimo

$$M = \frac{1}{5}(4 + 5 + 2 + 6 + 3) = 4 .$$

Preizkus pokaže, da harmonična sredina da pravi prezultat. Vseh pet delavcev, ki so delali skupno 300 minut, je izdelalo 87 izdelkov. Če bi vsi delavci delali s poprečno proizvodnostjo dela 3,45 minut za izdelek, bi napravili enako število 300/3,45 = 87 izdelkov. To pa je v skladu z opredelitvijo poprečja. Če bi vseh pet delavcev delalo z enotno produktivnostjo dela 4 minute za kos, ki smo jo dobili z aritmetično sredino, pa bi v skupno 300 minutah izdelali le 75 artiklov in ne 87, kolikor so jih v resnici.

Podrobneje o uporabi harmonične sredine pri računanju poprečnih relativnih števil bomo zvedeli v naslednjem odstavku.

9.25 Problem računanja poprečij iz relativnih števil je za socialno-ekonomsko statistiko tako osnoven in pomemben, da mu bomo posvetili poseben odstavek.

Relativna števila so v splošnem kvocienti primerjanih količin. Če vzamemo, da je grupno relativno število R_k razmerje dveh absolutnih podatkov Y_k in X_k

$$R_k = Y_k/X_k \tag{9.22}$$

moremo ta obrazec pisati v več oblikah. Iz njega dobimo, da je

$$Y_k = X_k R_k \tag{9.23}$$

in

$$X_k = Y_k/R_k \tag{9.24}$$

Če upoštevamo te zveze, računamo sumarno relativno število R na tri načine:

a) Sumarno relativno število je razmerje med vsotami absolutnih grupnih vrednosti za X_k in Y_k.

Po tej opredelitvi je sumarno relativno število R enako

$$R = \frac{Y}{X} = \frac{\sum Y_k}{\sum X_k} \tag{9.25}$$

S tem obrazcem računamo sumarno relativno število, če imamo po grupah absolutne podatke za Y_k in X_k.

b) Če upoštevamo obrazec 9.23, je sumarno relativno število

$$R = \frac{\sum X_k R_k}{\sum X_k} \tag{9.26}$$

tehtana aritmetična sredina grupnih relativnih števil R_k. Pri tem so vrednosti, ki so v imenovalcu relativnega števila, ponderi. Tako računamo poprečno relativno števi-

lo, če razpolagamo z grupnimi relativnimi števili in vrednostmi, ki nastopajo v imenovalcu relativnih števil, oziroma če je smiseln produkt pondera z relativnim številom $X_k \cdot R_k$.

c) Če upoštevamo obrazec 9.24, pa sumarno relativno število izračunamo po obrazcu

$$R = \frac{\sum Y_k}{\sum Y_k / R_k} \qquad (9.27)$$

To pa je tehtana harmonična sredina grupnih relativnih števil R_k. Pri tem so ponderi količine, ki nastopajo v relativnem številu v številu. Po tem obrazcu računamo sumarno relativno število, če razpolagamo z grupnimi relativnimi vrednostmi R_k in grupnimi količinami Y_k, ki nastopajo v relativnem številu v številu, oziroma če je smiseln kvocient pondera in relativnega števila Y_k/R_k.

Obrazci 9.25, 9.26 in 9.27 oziroma gornja pravila ne veljajo samo za relativna števila v ožjem smislu, temveč tudi za računanje sumarnih aritmetičnih sredin, saj so aritmetične sredine tudi relativna števila, ker so razmerja med vsoto vrednosti in številom enot.

9.26 Po zgornjih pravilih ugotovimo, kdaj uporabimo za računanje sumarnih relativnih števil tehtano aritmetično ali tehtano harmonično sredino.

Po teh pravilih ugotovimo, da je treba v primeru produktivnosti dela, ki smo ga navedli v odstavku 9.24, računati harmonično sredino. Produktivnost dela je merjena z relativnim številom "čas, porabljen za izdelavo enega izdelka", ki ga dobimo, če porabljen čas delimo s številom izdelkov. Ker razpolagamo s časom, koliko so posamezni delavci delali (60 minut), so ponderi količine, ki v relativnem številu nastopajo v številu. Po gornjih pravilih je treba za sumarno relativno število uporabiti harmonično sredino.

Če poznamo površine in poprečno gostoto prebivalstva po republikah, izračunamo gostoto prebivalstva za SFRJ kot tehtano aritmetično sredino, ker so ponderi površine v koeficientu "gostota prebivalstva" v imenovalcu.

Če imamo po podjetjih dano število zaposlenih žensk in odstotek zaposlenih žensk od skupnega števila zaposlenih, izračunamo poprečen odstotek zaposlenih žensk za celo

stroko po obrazcu za računanje harmonične sredine. Relativno število "odstotek zaposlenih žensk" je kvocient med številom zaposlenih žensk in številom skupno zaposlenih v podjetju. Ponder - število zaposlenih žensk - je v tem primeru v števcu relativnega števila.

Za podjetje imamo koeficiente o obračanju zalog posameznih vrst surovin (merjeno s časom enega obrata za posamezno vrsto surovine) in ustrezne vrednosti porabljenih surovin. Sumarni koeficient obračanja zalog izračunamo po obrazcu za izračunanje tehtane aritmetične sredine. Pokazovalec o obračanju zalog "čas enega obrata" je namreč kvocient med poprečnimi zalogami in vrednostjo porabljenih surovin. Ponder - vrednost porabljenih surovin - pa je v relativnem številu v imenovalcu.

Če imamo po občinah podatke o površini in skupnem pridelku pšenice v določenem letu, izračunamo poprečen hektarski pridelek pšenice po obrazcu 9.25 tako, da vsoto pridelkov po občinah delimo z vsoto površin.

Odvisnost sumarnega relativnega števila od sestave ponderov

9.27 Če preuredimo obrazec 9.26, dobimo

$$R = \frac{\sum X_k R_k}{X} = \sum \frac{X_k}{X} \cdot R_k = \sum X_k^o R_k \qquad (9.28)$$

Podobno dobimo iz obrazca 9.27

$$R = \frac{Y}{\sum Y_k / R_k} = \frac{1}{\sum \frac{Y_k}{Y} / R_k} = \frac{1}{\sum Y_k^o / R_k} \qquad (9.29)$$

Pri tem pomeni: $X_k^o = X_k / X$ strukturni delež za grupni podatek X_k; $Y_k^o = Y_k / Y$ strukturni delež za grupni podatek Y_k.

Iz teh dveh obrazcev sklepamo, da sumarno relativno število ni odvisno samo od grupnih relativnih števil, R_k, temveč tudi od sestave ustreznih ponderov X_k^o ali Y_k^o.

To je ena izmed osnovnih hib sumarnih relativnih števil, zaradi katere imamo včasih zelo

resne pomisleke o uporabnosti sumarnih pokazovalcev. Za populaciji, ki imata enaka grupna relativna števila, sta sumarni relativni števili med seboj različni, če sta sestavi ponderov različni. Še več. Razlike v sestavi ponderov morejo povzročiti, da je sumarno relativno število za populacijo A večje kot za populacijo B, čeprav so vsa grupna relativna števila za populacijo A manjša kot za populacijo B.

9.28 Vzemimo za zgled poprečne mesečne prejemke delavcev v kmetijstvu in gradbeništvu po kvalifikaciji.

Tabela 9.5. Poprečni prejemki delavcev v kmetijstvu in gradbeništvu po kvalifikaciji v letu 1957 v SFRJ (Vir SB 114).

Kvalifikacija	Poprečni mesečni prejemki v dinarjih		Sestava o številu delavcev	
	Kmetijstvo	Gradbeništvo	Kmetijstvo	Gradbeništvo
Visoko kvalificirani	15310	15200	0,020	0,078
Kvalificirani	12430	11920	0,218	0,298
Priučeni	9530	8950	0,157	0,218
Nekvalificirani	8030	6770	0,605	0,406
S k u p n o	8910	9690	1,000	1,000

Iz tabele 9.5 vidimo, da so poprečni prejemki delavcev v kmetijstvu za vse kvalifikacije višji kot v gradbeništvu. Kljub temu pa so sumarni poprečni prejemki v kmetijstvu nižji kot v gradbeništvu. Vzrok je različna sestava za število delavstva v kmetijstvu in gradbeništvu. Iz zadnjih dveh stolpcev v tabeli 9.5. vidimo, da je sestava delavstva po kvalifikaciji v poljedelstvu znatno nižja kot v gradbeništvu. Ta razlika v sestavi delavstva po kvalifikaciji ima tako močan vpliv na sumarno poprečje, da so poprečni prejemki delavstva v celoti v kmetijstvu nižji kot v gradbeništvu, čeprav so za posamezne kvalifikacije prejemki obratni.

S t a n d a r d i z i r a n i p o k a z o v a l c i

9.29 Omenjena lastnost sumarnih pokazovalcev je tako značilna, da se vprašamo, ali imajo sumarni pokazovalci sploh analitičen pomen in smisel, ker je primerljivost

sumarnih pokazovalcev zelo dvomljiva in meglena, če ne vemo, ali izvira razlika iz razlik v grupnih pokazovalcih ali iz razlik v sestavi ponderov.

Splošen koeficient mortalitete je po tem odvisen od umrljivosti po posameznih starostnih skupinah za ženske in moške in spolno-starostne sestave prebivalstva. Prebivalstvo z razmeroma nizko stopnjo umrljivosti po posameznih starostnih skupinah more imeti visok splošen koeficient umrljivosti, če so v populaciji predvsem stari ljudje. Prebivalstvo z razmeroma visoko stopnjo umrljivosti po posameznih starostnih skupinah pa more imeti nizek splošen koeficient umrljivosti, če sestoji predvsem iz mladih ljudi.

Podjetje izkaže visok sumarni koeficient za produktivnost dela, če izdeluje predvsem izdelke, za katere je splošna produktivnost dela visoka, čeprav je produktivnost dela za posamezen izdelek v podjetju razmeroma nizka in obratno.

Primerljivost tako izračunanih sumarnih pokazovalcev motijo razlike v sestavi. Da izločimo vpliv razlik v sestavi, izračunavamo standardizirane pokazovalce tako, da vzamemo za vse sumarne pokazovalce, ki jih primerjamo, stalno- standardno sestavo za ponderje. Tako je standardiziran koeficient umrljivosti tehtana sredina specifičnih koeficientov umrljivosti po starosti, pri tem pa vzamemo enotno standardno starostno sestavo za vse države, za katere primerjamo podatke. Ti koeficienti imajo večjo analitično vrednost kot navadni koeficienti umrljivosti, ker razlike med standardiziranimi koeficienti kažejo samo razlike v umrljivosti, ne pa v starostni in spolni strukturi. Standardna struktura, ki jo vzamemo za osnovo za izračunanje sumarnih koeficientov, mora biti seveda taka, da čim bolje ustreza stvarnim sestavam za posamezne populacije, ki jih med seboj primerjamo. Standardna sestava je zato včasih poprečna sestava iz vseh populacij, ki jih primerjamo, včasih pa idealna sestava, ki je pogojena z analizo pojava in podobno.

9.30 Izračunajmo za prejemke delavcev po kvalifikaciji za kmetijstvo in gradbeništvo iz tabele 9.5 standardizirane poprečne prejemke delavcev za obe panogi. Za standardno sestavo vzemimo povprečno sestavo delavstva v SFRJ po kvalifikaciji v vseh panogah.

Tabela 9.6 Izračunanje standardiziranih mesečnih prejemkov v kmetijstvu in gradbeništvu v letu 1957 v SFRJ

Kvalifikacija	Poprečni mesečni prejemki v din		Standardna sestava	(1).(3)	(2).(3)
	kmetijstvo	gradbeništvo			
Visoko kvalificirani	15310	15200	0,075	1148.25	1140.00
Kvalificirani	12430	11920	0,365	4536.95	4350,80
Priučeni	9530	8950	0,263	2506.39	2353.85
Nekvalificirani	8030	6770	0,297	2384,91	2010.69
			1.000	10576.50	9855.34

Za izračunanje sumarnih poprečnih prejemkov vzamemo za standardno sestavo sestavo delavstva po kvalifikaciji v SFRJ. Tako dobimo, da so v kmetijstvu poprečni mesečni prejemki večji (10576 din) kakor v gradbeništvu (9855 din). Standardizirana poprečja resnično pokažejo, da je raven dohodkov v kmetijstvu višja kot v gradbeništvu, ker razlike v sumarnih dohodkih izvirajo samo iz razlik v ravni dohodkov po kvalifikacijah, ne pa tudi iz razlik v sestavi zaposlenih. Seveda so standardizirana poprečja različna od poprečij, ki smo jih dobili v tabeli 9.5. Te razlike so rezultat tega, da je stvarna sestava v posamezni panogi različna od poprečne sestave za vse panoge.

GEOMETRIJSKA SREDINA

9.31 Geometrijska sredina G iz N vrednosti $y_1, y_2 \cdots y_N$ je N-ti koren iz produkta vseh N vrednosti

$$G = \sqrt[N]{y_1 \cdot y_2 \cdots y_N} \qquad (9.30)$$

Iz te opredelitve sklepamo, da ima smisel računati geometrijsko sredino le tedaj, če noben izmed členov ni negativen ali nič.

9.32 Tehtana geometrijska sredina je za razliko od navadne geometrijske sredine, ki je dana v obrazcu 9.30, enaka

$$G = \sqrt[\sum w_k]{y_1^{w1} \cdot y_2^{w2} \ldots y_r^{w_r}} \qquad (9.31)$$

9.33 Neposredno računanje geometrijske sredine po zgornjih osnovnih obrazcih je, razen za najenostavnejše primere, neizvedljivo. Pač pa moremo geometrijsko sredino razmeroma enostavno izračunati z logaritmi. Če namreč obrazca 9.30 in 9.31 logaritmiramo, dobimo

$$\log G = \frac{1}{N} \sum \log y_i \qquad (9.32)$$

in

$$\log G = \frac{1}{\sum w_k} \sum w_k \log y_k \qquad (9.33)$$

Logaritem navadne geometrijske sredine je enak navadni aritmetični sredini logaritmov iz osnovnih podatkov, logaritem za tehtano geometrijsko sredino pa je enak tehtani aritmetični sredini iz logaritmov osnovnih podatkov. S tema obrazcema posredno računamo geometrijske sredine z logaritmiranjem.

9.34 Zaradi razmeroma zamotanega računanja in težjega tolmačenja geometrijsko sredino računamo redkeje kot druge vrste sredin. Vendar so problemi, pri katerih je upravičeno edino računanje geometrijske sredine. Tako računamo geometrijsko sredino v nekaterih posebnih problemih z indeksi oziroma relativnimi števili na splošno.

Povečanju cene za 100% (indeks 200) smiselno ne ustreza znižanje cene za 100%, ker bi v tem primeru bila cena nič, temveč znižanje za 50% (indeks 50). Izravnava raznosmernih učinkov se pokaže pri geometrijski sredini, ker je geometrijska sredina iz indeksa 200 in 50

$$\sqrt{200.50} = 100$$

Ne pokaže pa se na primer pri aritmetični sredini, ker je aritmetična sredina med 200 in 50 enaka 125.

Zato je za izračunanje časovnih sredin iz posamičnih indeksov bolj upravičena geometrijska kot aritmetična sredina.

9.35 Geometrijsko sredino uporabljamo tudi za računanje srednjega koeficienta dinamike ali verižnega indeksa poprečne stopnje rasti.

Zaznamujmo z $Y_o, Y_1, Y_2, \ldots Y_N$ posamezne člene v časovni vrsti, s $k_1 = Y_1/Y_o$, $k_2 = Y_2/Y_1, \ldots\ldots k_N = Y_N/Y_{N-1}$ pa posamezne koeficiente dinamike. k naj bo srednji koeficient dinamike, s katerim bi se moral pojav v času od o do N spreminjati, da bi dosegli isti končni člen Y_N.

Med temi količinami so naslednje zveze

$$Y_o.k_1.k_2 \ldots\ldots k_N = Y_N = Y_o k.k \ldots\ldots k = Y_o k^N \qquad (9.34)$$

Če namreč Y_o postopoma množimo s k_1, $k_2 \ldots k_N$, dobimo zaradi pomena koeficientov dinamike postopoma vrednosti členov v časovni vrsti in končno zadnji člen Y_N. Glede na opredelitev srednjega koeficienta dinamike k pa dobimo Y_N tudi, če začetno vrednost Y_o N-krat pomnožimo s k, ali če jo pomnožimo s k^N.

Iz zveze v obrazcu 9.34 moremo izpeljati dva obrazca

$$k = \sqrt[N]{k_1.k_2 \ldots\ldots k_N} \qquad (9.35)$$

in

$$k = \sqrt[N]{Y_N/Y_o} \qquad (9.36)$$

ali splošneje

$$k = \sqrt[t_1-t_o]{Y_{t_1}/Y_{t_o}} \qquad (9.37)$$

Srednji koeficient dinamike je po obrazcu 9.35 geometrijska sredina iz posameznih koeficientov dinamike.

Po drugem obrazcu pa računamo srednji koeficient dinamike, če poznamo za razmik, za katerega računamo srednji koeficient dinamike, začetno Y_{t_o} in končno vrednost pojava Y_{t_1}. Srednji koeficient dinamike je v tem primeru koren iz kvocienta iz Y_{t_1} in Y_{t_o}. Stopnja korena pa je določena s časovnim razmikom, za katerega iščemo srednji koeficient dinamike. Zato ni nujno, da je N celo število, ampak more biti (t_1-t_o) kateri koli

časovni razmik, kar je določneje vidno iz obrazca 9.37.

9.36 Pri obravnavanju verižnih indeksov v odstavku o enostavnih indeksih smo izračunali vrsto verižnih indeksov za proizvodnjo električne energije v SFRJ v letih 1951-1956. Verižni indeksi v teh letih so: 105,9 ; 110,4 ; 115,4 ; 126,1 ; 117,1.

Srednji verižni indeks v tem razdobju je po obrazcu 9.35 enak geometrijski sredini iz posameznih verižnih indeksov

$$\bar{I} = \sqrt[5]{105,9.110,4.115,4.126,1.117,1}$$

Z logaritmi izračunamo srednji verižni indeks po obrazcu 9.32

$$\log \bar{I} = \frac{1}{5} (\log 105,9 + \log 110,4 + \log 115,4 + \log 126,1 + \log 117,1) = \frac{1}{5} (2,02490 + 2,04297 + 2,06221 + 2,101106 + 2,06856) = 2,0599.$$

Z antilogaritmiranjem dobimo, da je srednji verižni indeks $\bar{I} = 114,8$, poprečna stopnja rasti $T = \bar{I} - 100 = 14,8\%$

Da bi bilo zvečanje proizvodnje električne energije v razdobju 1951-1956 enako stvarnemu, bi se morala proizvodnja električne energije večati s srednjim verižnim indeksom 114,8.

Isti rezultat dobimo, če računamo srednji verižni indeks po obrazcu 9.36 iz začetne in končne proizvodnje električne energije.

9.37 Po republikah imamo podatke o družbenem produktu, po cenah iz leta 1966 za leti 1963 in 1971 po republikah. Ker je od 1963 do 1971 razdobje osmih let, je poprečen koeficient dinamike osmi koren iz kvocientov Y_{71} in Y_{63}

Tabela 9.8 Poprečni koeficienti dinamike za družbeni produkt v razdobju 1963-1971, po republikah in pokrajinah (Vir:SG-72)

Republika-pokrajina	1963	1971	poprečen koeficient dinamike	poprečen verižni indeks	poprečna stopnja rasti %
SFRJ	80577	134234	1,065	106,5	+ 6.5
BiH	9890	15749	1,060	106,0	+ 6.0
Črna gora	1433	2463	1,070	107,0	+ 7.0
Hrvatska	21478	35680	1,066	106,6	+ 6.6
Makedonija	4059	7403	1,078	107,8	+ 7.8
Slovenija	12220	20814	1,069	106,9	+ 6.9
Srbija - skupno	31497	52125	1,065	106,5	+ 6.5
Srbija - ožje področje	20565	33942	1,065	106,5	+ 6.5
Vojvodina	9471	15429	1,063	106,3	+ 6.3
Kosovo	1461	2754	1,082	108,2	+ 8.2

Za posamezne republike in pokrajine so osemletni poprečni koeficienti dinamike v razmiku od 1,060 za BiH do 1,082 za Kosovo.

ODNOSI MED RAZLIČNIMI VRSTAMI SREDNJIH VREDNOSTI

9.38 Če vzamemo ilustrativen zgled in iz 1 in 4 izračunamo harmonično H, geometrijsko G in aritmetično M sredino, dobimo, da je

$$H = \frac{2}{1/1 + 1/4} = 1,6; \quad G = \sqrt{1.4} = 2; \quad M = \frac{1+4}{2} = 2,5$$

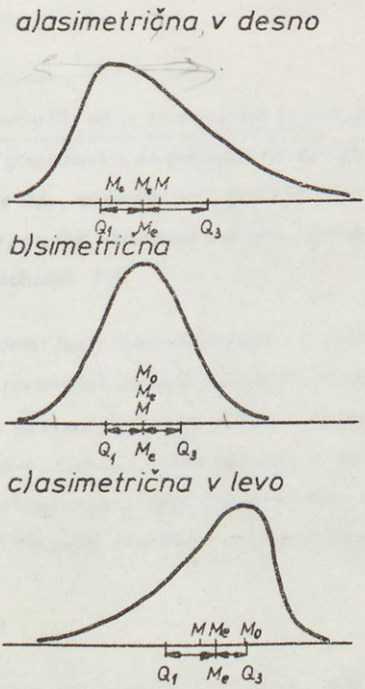
Iz teh rezultatov vidimo, da je harmonična sredina, če jo izračunamo iz istih podatkov, najmanjša, aritmetična sredina največja, geometrijska sredina pa leži med obema. To ne velja samo za ta primer, ampak velja splošno pravilo; harmonična sredina je manjša kakor geometrijska, ta pa manjša kakor aritmetična sredina, če jih izračunamo iz istih podatkov.

$$H \leq G \leq M \qquad (9.34)$$

9.39 Tudi med vrednostmi modusa Mo, mediane Me in aritmetične sredine M opazimo neke stalne odnose. Za unimodalne simetrične porazdelitve so vrednosti modusa, mediane in aritmetične sredine enake ($M_o = M_e = M$). Za unimodalne porazdelitve, ki so asimetrične v desno, je modus manjši, aritmetična sredina pa večja kot mediana ($Mo < Me < M$). Za unimodalne porazdelitve, ki so asimetrične levo, pa se vrstni red zamenja in je aritmetična sredina manjša, modus pa večji kot mediana ($M < Me < Mo$).

Za zvezne, ne preveč asimetrične porazdelitve velja, da je razlika med aritmetično sredino in modusom približno trikrat večja kot razlika med aritmetično sredino in mediano.

$$M - Mo = 3 \cdot (M - Me) \qquad (9.35)$$



Slika 9.2. Odnosi med Mo, M in Me za različne tipe unimodalnih porazdelitev

Ta obrazec moremo uporabiti za ne preveč asimetrične unimodalne porazdelitve, da z njim ocenimo modus, čeprav nimamo podatke grupirane v frekvenčni porazdelitvi. Ker

moremo M in Me izračunati tudi iz negrupiranih podatkov, ocenimo vrednost modusa z obrazcem

$$Mo = M - 3.\ (M - Me) \qquad (9.36)$$

Odnosi med modusom, mediano in aritmetično sredino za posamezne vrste porazdelitev so nakazani v sliki 9.2.

DESETO POGLAVJE

MERE VARIACIJE, ASIMETRIJE IN SPLOŠČENOSTI

MERE VARIACIJE

10.1 Že večkrat smo poudarili, da je za množične pojave značilna variabilnost. Tako se od kmetijskega gospodarstva do gospodarstva spreminja velikost, pridelek, število članov, število živine itd., od podjetja do podjetja poraba surovin, število delavcev, vrednost proizvodnje, poraba električne energije, goriva itd., od človeka do človeka starost, spol, stan, dohodek itd.

Medtem ko je srednja vrednost izraz splošnih pogojev, ki so za vso populacijo enaki - nespremenjeni, se zaradi posamičnih vplivov vrednosti za posamezne enote od nje odklanjajo navzgor in navzdol. Čim manjši so posamični vplivi, tem manjši so odkloni in tem manjša je variabilnost pojava in obratno; čim močnejši so posamični vplivi, tem večji so odkloni oziroma variabilnost pojava. Variabilnost je izraz delovanja posamičnih vplivov, ki morejo biti manjši ali večji. Variabilnost moremo torej meriti.

Vrste mer variacije

10.2 Variabilnost merimo z več parametri. Vsak izmed njih ima drugo osnovo, vendar vsi - vsak po svoje, kaže isto lastnost - variabilnost pojava.

Obravnavali bomo naslednje mere variacije:

a) variacijski razmik R,

b) kvartilni odklon Q,

c) poprečni absolutni odklon AD,

d) standardni odklon SD- σ , oziroma varianco σ^2 , iz katere je SD izpeljan,

e) poprečna razlika Δ_1 .

Od teh mer variacije sta variacijski razmik R in kvartilni odklon Q dana z lego nekaterih členov v populaciji, poprečni absolutni odklon AD, standardni odklon SD in poprečna razlika Δ_1 pa so izračunani iz vseh vrednosti populacije in so zaradi tega solidnejše mere variabilnosti. Zaradi svojih lastnosti je izmed vseh mer variacije najvažnejši standardni odklon SD.

Variacijski razmik- R

10.3 Najenostavnejša, a tudi najbolj problematična mera variacije je variacijski razmik

$$R = y_{max} - y_{min} \qquad (10.1)$$

ki je po definiciji razlika med največjo in najmanjšo vrednostjo v populaciji.

Najmanjšo in največjo vrednost zlahka najdemo, če so podatki urejeni v ranžirni vrsti. Če vzamemo, da je bil v letu 1969 v SRS najnižji blagovni promet na prebivalca v občini Šentjur (1.497) in najvišji v občini Ljubljana-Center (30.011 din), je variacijski razmik za blagovni promet na prebivalca

$$R = 30.011 - 1.497 = 28.514 \text{ din}$$

Včasih variacijski razmik posredno izražamo tudi tako, da navedemo najnižjo in najvišjo vrednost.

Variacijski razmik je nestabilna mera variacije, ker je izračunan iz obeh skrajnih vrednosti populacije, ki moreta biti rezultat nekih izjemnih - netipičnih vplivov. V našem primeru je to gotovo s podatkom za Ljubljano-Center. R ni odvisen od razmestitve vrednosti znotraj variacijskega razmika. Ta more biti najrazličnejša. Vse vrednosti morejo biti v variacijskem razmiku razmeščene enakomerno ali pa se goste okrog nekega središča. V prvem primeru je variabilnost znatno večja kot v drugem, variacijski razmik pa tega

ne pokaže, ker sta skrajni vrednosti v obeh primerih isti. Kljub tem hibam pa variacijski razmik uporabljamo vselej, kadar hočemo jakost variacije na hitro orientacijsko oceniti in ni časa, niti potrebe za izračun drugih, natančnejših mer variacije. Eden izmed takih primerov je tekoče preverjanje kakovosti dela strojev. Variacijski razmik, ki ga določimo iz meritev razmeroma majhnega števila izdelkov, rabi za orientacijo jakosti variacije, ki je eden izmed osnovnih pokazovalcev kakovosti proizvodnje.

Kvartilni odklon - Q

10.4 Variacijski razmik je slabo merilo variabilnosti zato, ker je odvisen samo od skrajnih vrednosti. To hibo odpravimo tako, da merimo variacijo z razmikom, v katerem je samo del populacije. Tako s prvim in devetim decilom omejimo razmik, v katerem je 80% vseh vrednosti, ker smo izločili po 10% najmanjših in največjih vrednosti. Ta razmik

$$D_9 - D_1$$

imenujemo decilni razmik.

Podobno s prvim in tretjim kvartilom omejimo razmik, v katerem je 50% vrednosti in iz njega izključimo 25% najmanjših in 25% največjih vrednosti.

Običajno pa namesto kvartilnega razmika

$$Q_3 - Q_1$$

vzamemo za mero variacije kvartilni odklon

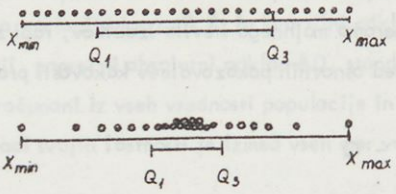
$$Q = \frac{1}{2} (Q_3 - Q_1) \qquad (10.2)$$

ki je polovica kvartilnega razmika.

Gornji meri variacije pravilneje pokažeta jakost variacije kakor variacijski razmik, ker nanju ne vplivajo skrajne - netipične vrednosti. Vendar pogosteje uporabljamo kvartilni kot decilni razmik. V sliki 10.1 je shematično prikazano, kako je variacijski razmik neobčutljiv za razlike v razmestitvi vrednosti znotraj variacijskega razmika, in kako se raz-

lična razmestitev pokaže na kvartilnem razmiku.

Določanje gornjih mer variacije ne dela težav, ker smo v odstavku o kvantilih nakazali, kako izračunamo decile in kvartile.



Slika 10.1 Variacijski in kvartilni razmik pri različni razmestitvi vrednosti

10.5 Za blagovni promet na prebivalca med občinami v SRS v letu 1969, za katerega smo izračunali že variacijski razmik, je kvartilni odklon

$$Q = \frac{1}{2}(Q_3 - Q_1) = \frac{5102,5 - 2632,5}{2} = 1235 \text{ din}$$

Poprečni absolutni odklon - AD

10.6 Uvodoma smo naznačili, da z odkloni posamičnih vrednosti od sredine merimo jakost posamičnih vplivov na posamezno enoto. Sumarno merilo jakosti teh vplivov bi moglo biti poprečje posamičnih odklonov. To pa je, kakor vemo, enako nič, če odklone računamo od aritmetične sredine, ker so odkloni pozitivni in negativni. Zato vzamemo za merilo variabilnosti poprečje absolutnih vrednosti posamičnih odklonov od sredine. Po tej definiciji je poprečni absolutni odklon od aritmetične sredine

$$AD_M = \frac{1}{N}\sum|y - M| \qquad (10.3)$$

Ker je poprečni absolutni odklon najmanjši, če računamo odklone od mediane, računamo AD tudi iz absolutnih odklonov od mediane

$$AD_{Me} = \frac{1}{N}\sum|y - Me| \qquad (10.4)$$

AD_{Me} je vsebinsko zaradi te lastnosti mediane bolj upravičen kot AD_M.

10.7 Če z $Y_s = \sum_{y < M_e} y_i$ zaznamujemo vsoto posamičnih podatkov, katerih vrednost je manjša kot mediana, z $Y_z = \sum_{y > M_e} y_i$ pa vsoto podatkov, ki so večji kakor mediana, izračunamo AD_{Me} po obrazcu

$$AD_{Me} = \frac{1}{N} (Y_z - Y_s) \qquad (10.5)$$

AD_{Me} pa moremo računati tudi po obrazcu

$$AD_{Me} = \frac{1}{2} (M_z - M_s) \qquad (10.6)$$

pri čemer je M_s poprečje iz členov pod, M_z pa poprečje iz členov nad mediano. Ta zveza je po sestavu podobna kvartilnemu odklonu, le da namesto kvartilov, ki sta mediani za prvo in drugo polovico populacije, vzamemo ustrezne aritmetične sredine.

Za blagovni promet na prebivalca v letu 1969 je:

$Y_s = 79347$, $Y_z = 183078$, število občin $N = 60$

Po obrazcu 10.5 je

$$AD_{Me} = \frac{1}{60} (183978 - 79347) = 1743{,}85 \text{ din}$$

AD_M pa dobimo po obrazcu

$$AD_M = 2 \cdot P_z \cdot P_s (M_z - M_s) \qquad (10.7)$$

pri tem pomeni: $P_s = N_s/N$ in $P_z = N_z/N$ strukturna deleža števila enot pod in nad aritmetično sredino M, $M_s = Y_s/N_s$ in $M_z = Y_z/N_z$ pa poprečji členov, ki so pod oziroma nad skupnim poprečjem M.

Za blagovni promet na prebivalca v SRS po občinah je $N_s = 42$, $N_z = 18$, $Y_s = 126249$, $Y_z = 183978$. Iz teh podatkov dobimo, da je: $P_s = 42/60 = 0{,}70$; $P_z = 18/60 = 0{,}30$;

$M_s = 126249/42 = 3005,93$ in $M_z = 183978/18 = 7615,33$.

Po obrazcu 10,7 pa je

$$AD_M = 2.0,30 \cdot 0,70 \cdot (7615,33 - 3005,93) = 1935,95 \text{ din}$$

Varianca. Standardni odklon - σ^2, σ - SD

10.8 Podobno osnovo kakor poprečen absolutni odklon ima tudi varianca σ^2, ki je po definiciji poprečje kvadratov odklonov od aritmetične sredine

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - M)^2 \qquad (10.8)$$

Pri AD odpravimo predznake odklonov tako, da vzamemo absolutne odklone, pri σ^2 pa tako, da odklone kvadriramo.

Podobno kot AD je tudi σ^2 izračunana mera variacije in je odvisna od vsake izmed vrednosti populacije. Vendar so druge lastnosti variance osnovne važnosti za statistično analizo. Zaradi tega je, enako kakor je aritmetična sredina osnovna srednja vrednost, varianca osnovna mera variacije, čeprav je njeno izračunanje razmeroma zamudno.

10.9 Varianca je poprečje kvadratov odklonov od aritmetične sredine. Zato je izražena v neprikladnih enotah mere - kvadratu enote mere znaka. Temu se izognemo tako, da izračunamo kvadratni koren iz variance. Dobljena mera variacije, ki jo imenujemo standardni odklon, zaznamujemo pa z σ ali SD pa je merjena v isti enoti mere kot osnovni znak. Varianca in standardni odklon sta v naslednji enostavni zvezi

$$SD = \sigma = \sqrt{\sigma^2} \qquad (10.9)$$

Računanje iz negrupiranih podatkov

10.10 Računanje variance po osnovnem obrazcu 10.8 se izkaže za zelo neprikladno. Po njem bi morali najprej izračunati aritmetično sredino za proučevan znak M, izračunati odklone posameznih vredbosti od aritmetične sredine $(y_i - M)$, vsak posamezen odklon kvadrirati $(y_i - M)^2$, dobljene kvadrate sešteti

in deliti z obsegom populacije. Zato varianco iz posamičnih podatkov raje računamo po enem izmed obrazcev

$$\sigma_y^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{N}}{N} = \frac{N\sum y_2^2 - (\sum y_i)^2}{N^2} \qquad (10.10)$$

$$\sigma_y^2 = \frac{Ky}{N} \; ; \; Ky = Q_i - Q \; ; \; Q_i = \sum y_i^2 \quad Q = \frac{(\sum y)^2}{N} \qquad (10.11)$$

Prvi obrazec dobimo iz osnovnega obrazca 10.8 z enostavnim računom

$$\sigma_y^2 = \frac{1}{N}\sum (y_i - M)^2 = \frac{1}{N}\left[\sum y^2 - 2M\sum y + NM^2\right] = \frac{1}{N}\left[\sum y_i^2 - \frac{(\sum y)^2}{N}\right] \qquad (10.12)$$

Nakazani način je še posebno prikladen, če računamo varianco z računskim strojem, ki istočasno zbira vsote podatkov in vsote kvadratov.

10.11 Če iz podatkov za 20 največjih mest v SFRJ izračunamo varianco poprečnih cen na drobno (v starih din) v letu 1970 za kokošja jajca, dobimo iz osnovnih podatkov

71	71	68	71	74	73	63	78	73	66	60	78	74
71	66	62	56	59	58	66						

$$\sum y_i = Y = 71 + 71 + 68 + \ldots + 58 + 66 = 1358$$

$$\sum y_i^2 = 71^2 + 71^2 + 68^2 + \ldots + 58^2 + 66^2 = 93028$$

$$\sigma_y^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{N}}{N} = \frac{93028 - \frac{1358^2}{20}}{20} = 40,99$$

standardni odklon pa

$$\sigma_y = \sqrt{40,99} = 6,4 \text{ din}$$

10.12 Ker ostane varianca enaka, če vsem osnovnim podatkom prištejemo ali odštejemo isto konstanto, je včasih prikladno, da od osnovnih podatkov pred računanjem od-

štejemo osnovnim podatkom ustrezno konstanto y_o, tako da so reducirane vrednosti u čim manjše

$$u_i = y_i - y_o$$

Varianco osnovnega znaka σ_y^2 dobimo, če računamo varianco iz reduciranih podatkov u po obrazcu

$$\sigma_y^2 = \sigma_u^2 = K_u/N \quad ; \quad K_u = \sum u^2 - \frac{(\sum u)^2}{N} \tag{10.13}$$

10.13 Če v našem zgledu za ceno jajc odštejemo od osnovnih podatkov $y_o = 68$, dobimo reducirane vrednosti:

3 3 0 3 6 5 -5 10 5 -2 -8 10 6
3 -2 -6 -12 -9 -10 -2

$$U = \sum u_i; = 3 + 3 + 0 + \ldots + (-9) + (-10) + (-2) = -2$$

$$\sum u_i^2 = 3^2 + 3^2 + 0^2 + \ldots + (-9)^2 + (-10)^2 + (-2)^2 = 820$$

$$\sigma^2 = \frac{\sum u^2 - \frac{U^2}{N}}{N} = \frac{820 - (-2)^2/20}{20} = 40,99$$

Dobili smo isti rezultat kot po obrazcu 10.10

Računanje iz grupiranih podatkov

10.14 Neposredna metoda. Za velike populacije je računanje variance iz negrupiranih podatkov še zamudneje kakor računanje aritmetične sredine. Zaradi tega za velike populacije tudi variance ocenjujemo iz podatkov, ki so grupirani v frekvenčne porazdelitve. Ocena variance iz frekvenčne porazdelitve je

$$\sigma_y^2 = \frac{1}{N}\sum f_k (y_k - M_y)^2 \tag{10.14}$$

tehtana aritmetična sredina kvadratov odklonov sredin razredov y_k od aritmetične sredine M_y. Vendar se ta postopek izkaže neprikladen v primerjavi z drugima postopkoma,

s katerima moremo tudi izračunati varianco iz frekvenčnih porazdelitev. Zato zanj ne navajamo primera, ker v splošnem te metode ne uporabljamo.

10.15 Metoda pomožnega znaka u. Metoda pomožnega znaka u je samo razširitev metode pomožnega znaka u za računanje aritmetične sredine iz grupiranih podatkov.

Po tej metodi računamo varianco tako, da:

a) Enako kakor za aritmetično sredino izberemo razred, ki je nekje sredi frekvenčne porazdelitve. Vanj postavimo izhodišče znaka $u = 0$, v druge razrede pa navzdol in navzgor od izhodišča ustrezne vrednosti pomožnega znaka u:

... -3, -2, -1, 0, +1, +2, +3, ...

b) Enako kakor za aritmetično sredino, izračunamo produkte frekvenc f_k z ustreznimi vrednostmi znaka u_k, da dobimo $f_k u_k$

c) Produkte $f_h u_k$ ponovno pomnožimo z ustreznimi vrednostmi znaka u_k: tako dobimo vrednosti $f_k u_k^2$.

d) Seštejemo frekvence f_k, produkte $f_k u_k$ in produkte $f_k u_k^2$.

e) Iz teh količin izračunamo varianco po obrazcih

$$\sigma_y^2 = \frac{i^2}{N} K_u \quad : \quad K_u = \sum f_k u_k^2 - \frac{(\sum f_k u_k)^2}{N} \tag{10.15}$$

Pri tem je: i = širina razreda.

10.16 Za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto je prikazano računanje variance po tej metodi v tabeli 10.1.

Tabela 10.1 Izračun variance σ^2 in standardnega odklona σ za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi pomožnega znaka u

Poraba v m^3	f	u	fu	fu^2
0,0 - 3,9	2	-3	- 6	18
4,0 - 7,9	16	-2	- 32	64
8,0 - 11,9	36	-1	- 36	36
12,0 - 15,9	41	0	0	0
16,0 - 19,9	29	+1	+ 29	29
20,0 - 23,9	12	+ 2	+ 24	48
24,0 - 27,9	7	+ 3	+ 21	63
28,0 - 31,9	4	+ 4	+ 16	64
32,0 - 35,9	2	+ 5	+ 10	50
	149		+ 26	372
	N		U	$\sum fu^2$

Po obrazcih 10.11 dobimo dalje

$$K_u = 372 - (+26)^2/149 = 367,4631$$

$$\sigma_y^2 = 4^2 \cdot 367,4631/149 = 39,4591$$

$$\sigma_y = \sqrt{39,4591} = 6,28\ m^3$$

10.17 Metoda kumulativ. Tudi za računanje variance moremo koristno uporabiti lastnosti kumulativnih vrst. Ta metoda je enako kakor metoda pomožnih znakov u samo razširitev iste metode pri računanju aritmetične sredine.

Po metodi kumulativ računamo varianco tako, da:

a) Enako kakor pri računanju aritmetične sredine iz frekvenčne porazdelitve iz f izračunamo kumulativno vrsto F. Zadnji člen kumulativne vrste (pod črto) je obseg populacije N.

b) Po enakem postopku kumuliranja izračunamo iz prve kumulativne vrste F drugo kumulativno vrsto FF. Zadnji člen druge kumulativne vrste FF (pod črto) je C_1.

c) Seštejemo vrednosti členov druge kumulativne vrste (brez člena pod črto). To vsoto zaznamujmo s C_2.

d) Iz teh količin izračunamo varianco po obrazcih

$$K_u = 2C_2 + C_1 - C_1^2/N \; ; \quad \sigma_y^2 = \frac{i^2}{N} K_u \qquad (10.16)$$

e) Kumulativne vrste moremo računati od zgoraj navzdol ali obratno od spodaj navzgor. Za asimetrične porazdelitve je prikladneje začeti izračunavati kumulative na strani asimetrije.

10.18 V tabeli 10.2 je postopek nakazan za porabo lesa v Novem mestu.

Tabela 10.2. Računanje variance za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi kumulativ.

Poraba v m^3	f	F	FF
0,0 - 3,9	2	0	
4,0 - 7,9	16	2	0
8,0 - 11,9	36	18	2
12,0 - 15,9	41	54	20
16,9 - 19,9	29	95	74
20,0 - 23,9	12	124	169
24,0 - 27,9	7	136	293
28,0 - 31,9	4	143	429
32,0 - 35,9	2	147	572
		149	719
		N	C_1

1559=2+20+...
+429+572 = C_2

$K_u = 2.1559 + 719 - 719^2/149 = 367,4631$, torej enako kot po metodi pomožnega znaka.

10.19 Sheppardov popravek. Varianca, ki jo izračunamo po zgornjih postopkih, je samo ocena za pravo vrednost variance, ki bi jo dobili, če bi jo izraču-

nali iz posamičnih podatkov. Analiza te ocene pa pokaže, da dobimo za unimodalne, ne preveč asimetrične porazdelitve za zvezne znake po tej metodi sistematično prevelike vrednosti za varianco. Ta napaka je tem večja, čim širši so razredi. Varianco, izračunano po prejšnjih metodah, moremo popraviti po obrazcu

$$\sigma^2_{y,pop} = \sigma^2_y - i^2/12 \qquad (10.17)$$

tako, da od prvotne ocene variance, ki jo dobimo po metodi znaka u ali po metodi kumulativ, odštejemo dvanajstino kvadrata širine razreda.

Ta popravek imenujemo Sheppardov popravek.

Za frekvenčno porazdelitev o porabi lesa v Novem mestu v letu 1953 je popravljena varianca enaka

$$\sigma^2_{y,pop} = 39,4591 - 4^2/12 = 38,1258$$

$$\sigma_{y,pop} = \sqrt{38,1258} = 6,17\ m^3$$

Skupna varianca

10.20 Enako kakor sumarno aritmetično sredino moremo tudi skupno varianco izračunati iz grupnih podatkov. Če poznamo po grupah število enot N_k, aritmetične sredine M_k in grupne variance σ^2_k, izračunamo skupno varianco za celo populacijo po obrazcu

$$\sigma^2 = M_{\sigma^2} + \sigma^2_M \qquad (10.18)$$

Pri tem je

$$M_{\sigma^2} = \frac{1}{N} \sum N_k \sigma^2_k \qquad (10.19)$$

tehtana aritmetična sredina grupnih varianc,

$$\sigma^2_M = \frac{1}{N} \sum N_k (M_k - M)^2 \qquad (10.20)$$

pa tehtana varianca med grupnimi aritmetičnimi sredinami M_k.

To zelo pomembno zvezo moremo dokazati enostavno. Iz znanih zvez velja $N_k \sigma_k^2 = \sum_i (y_{ki} - A)^2 - N_k (M_k - A)^2$. Ta zveza velja za poljuben A. Če postavimo A = M in seštejemo po grupah, dobimo $\sum_k N_k \sigma_k^2 = \sum_k \sum_i (y_{ki} - M)^2 - \sum_k N_k (M_k - M)^2$. Če to zvezo delimo z N, dobimo $M_{\sigma^2} = \sigma^2 - \sigma_M^2$.

10.21 Vzemimo za zgled stroške za kulturno in družbeno življenje v novembru 1957 za 36 delavskih družin v Mariboru. V tabeli 10.3 so navedeni N_k, M_k in σ_k^2 po višini dohodkov.

Tabela 10.3 Podatki za N_k, M_k in σ_k^2 za stroške za kulturno in družbeno življenje za 36 delavskih družin v novembru 1957 v Mariboru po skupinah dohodkov.

Višina dohodkov v tisoč din	N_k	M_k	σ_k^2
15 - 19	12	479	28358
20 - 24	10	670	35820
25 - 29	7	949	48212
30 - 34	4	1317	83469
35 -	3	1703	8089
	36=N	819=M	

Iz teh podatkov dobimo

$$\sigma_M^2 = \frac{1}{36}\left[12 \cdot (479-819)^2 + \ldots + 3 \cdot (1703-819)^2\right] = 140709$$

$$M_{\sigma^2} = \frac{1}{36}(12 \cdot 28358 + \ldots + 3.8089) = 38726$$

$$\sigma^2 = M_{\sigma^2} + \sigma_M^2 = 38726 + 140709 = 179435$$

10.22 Pomembnejše kakor to, da moremo izračunati skupno varianco, če poznamo grupne podatke, pa je dejstvo, da moremo z zgornjim postopkom skupno varianco razstaviti po grupnem znaku v dva dela. Skupna varianca namreč meri vpliv vseh posamičnih faktorjev, medtem ko meri σ_M^2 vpliv grupnega znaka, M_{σ^2} pa vpliv drugih, posamičnih faktorjev na variabilnost pojava.

Po zgornjem postopku moremo torej varianco analizirati v tem smislu, da ugotovimo, koliko od skupne variance je rezultat dejavnika, ki ga vzamemo za grupni znak.

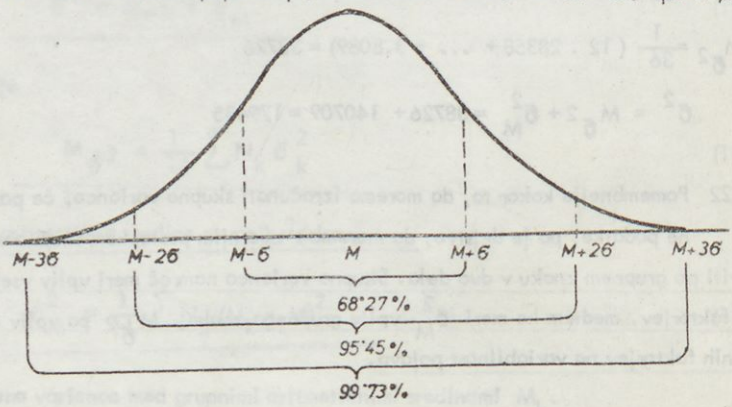
Za stroške za kulturno in družbeno življenje za 36 družin v novembru 1967 v Mariboru vidimo, da je velik del (78%) od skupne variance stroškov za kulturno in družbeno življenje pojasnjen z višino dohodkov. Višina dohodkov je torej eden izmed bistvenih faktorjev, ki vplivajo na stroške za kulturno in družbeno življenje.

10.23 Zveza standardnega odklona z normalno porazdelitvijo

Variacijski razmik ali kvartilni razmik imata enostavno tolmačenje in pomen. Enako je tudi poprečni absolutni odklon razmeroma razumljiv. Teže pa je s standardnim odklonom, dokler ga ne obravnavamo v zvezi z nekaterimi lastnostmi porazdelitev. Standardni odklon namreč dobimo z razmeroma zapletenim računskim postopkom, ki zamegli predstavo o smislu tega parametra.

Čeprav normalno porazdelitev podrobneje obravnavamo kasneje, je prav, da navedemo nekaj njenih lastnosti v zvezi s standardnim odklonom. Kakor že vemo, je normalna porazdelitev ena izmed osnovnih teoretičnih porazdelitev. Pomembna je po svojih teoretičnih lastnostih in praktičnem pomenu. V stvarnosti se ob določenih pogojih pogosto pojavljajo porazdelitve, ki so po svojih značilnostih normalni več ali manj podobne, ker so simetrične, unimodalne in zvonaste oblike.

Za vse normalne porazdelitve na splošno velja, da leži v razmiku $M - \sigma$ do $M + \sigma$ 68,27% ali okroglo 2/3 vseh vrednosti, v razmiku $M - 2\sigma$ do $M + 2\sigma$ 95,45% ali približno 95% vseh vrednosti populacije, v razmiku $M - 3\sigma$ do $M + 3\sigma$ pa 99,73%



Slika 10.2. Deleži za normalno porazdelitev v razmikih $M \pm \sigma$, $M \pm 2\sigma$ in $M \pm 3\sigma$.

ali praktično vse vrednosti. Odkloni od aritmetične sredine, ki so absolutno večji od σ, se v splošnem pojavijo v tretjini vseh primerov, odkloni, ki so večji kot 2σ, samo v 5% vseh primerov, odkloni, ki so večji kot 3σ, pa so zelo zelo redki. Ti odnosi veljajo strogo za normalno porazdelitev. Približno pa veljajo tudi za druge unimodalne, simetrične in zvonaste porazdelitve.

Koliko veljajo gornje zakonitosti za unimodalne zvonaste porazdelitve na splošno, si moremo ustvariti sliko iz porazdelitve o porabi lesa v Novem mestu. Zanjo dobimo, da leži v razmiku $M - \sigma$ do $M + \sigma$, to je od 8,53 m^3 do 20,87 m^3 70% vseh primerov, v razmiku $M - 2\sigma$ do $M + 2\sigma$, to je od 2,36 m^3 do 27,04 m^3 94% vseh primerov, v razmiku $M - 3\sigma$ do $M + 3\sigma$, to je od 0 do 33,21 m^3 pa 99,3% vseh primerov. Razlike od deležev pri normalni porazdelitvi so torej neznatne.

Poprečna razlika - Δ_1

10.24 Poprečni absolutni odklon in varianca imata za osnovo odklone od neke srednje vrednosti (Me ali M). Mera variacije, ki ni zasnovana na odklonih od neke srednje vrednosti, je **poprečna razlika** Δ_1. Poprečna razlika je poprečje iz vseh možnih pozitivnih razlik vseh vrednosti v populaciji. Če je število enot enako N, je vseh možnih pozitivnih razlik $\binom{N}{2}$. Za ranžirno vrsto posamičnih vrednosti je poprečna razlika Δ_1 dana z obrazcem

$$\Delta_1 = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N} \sum_{j<i} (y_i - y_j); \quad y_1 \leq y_2 < \cdots \leq y_i \leq \cdots < y_N \qquad (10.21)$$

Iz podatkov, ki so urejeni v ranžirni vrsti, izračunamo Δ_1 s kumulativami ker je

$$\sum_{i=1}^{N} \sum_{j<i} (y_i - y_j) = \sum_{i=1}^{N} \left[(i-1)\, y_i - (N-i) y_i \right] = \sum_{i=1}^{N} \left[2\, i y_i - (N+1) y_i \right] = (N-1)C_o - 2C_1$$

Iz kumulative ranžirne vrste dobimo pomožne vrednosti C_o in C_1. Za kumulativne vrste vemo, da je $C_o = \sum_{i=1}^{N} y_i$ in $C_1 = \sum_{i=1}^{N} (N-i) y_i$ iz česar je $\sum_{i=1}^{N} i\, y_i = N C_o - C_1$, Δ_1 pa

$$\Delta_1 = \frac{2\left[(N-1)C_o - 2C_1\right]}{N(N-1)} \qquad (10.22)$$

10.25 Za zgled izračunajmo Δ_1 za naslednje izmišljene podatke: 7,9,13,3,15,8,5. V ranžirno vrsto urejeni podatki so naslednji:

y_i	3	5	7	8	9	13	15	
Fy_i	0	3	8	15	23	32	45	$60 = C_o$

$$0 + 3 + 8 + \ldots + 45 = 126 = C_1$$

Po obrazcu 10.22 je

$$\Delta_1 = \frac{2\left[(7-1)\ 60 - 2.126\right]}{7 \cdot (7-1)} = 5{,}14$$

Neposredno računanje je takole: Iz ranžirne vrste izračunane vse možne pozitivne razlike so

y_i	3	5	7	8	9	13	15
$y_i - y_1$		2	4	5	6	10	12
$y_i - y_2$			2	3	4	8	10
$y_i - y_3$				1	2	6	8
$y_i - y_4$					1	5	7
$y_i - y_5$						4	6
$y_i - y_6$							2

$$\Delta_1 = \frac{\sum d_i}{\binom{N}{2}} = \frac{108}{21} = 5{,}14$$

10.26 Če so podatki grupirani v frekvenčni porazdelitvi, dobimo približek poprečne razlike po obrazcu

$$\Delta_1 = \frac{2\left[\sum Y_k (F_k + F_{k+1}) - NY\right]}{N(N-1)} \qquad (10.23)$$

pri čemer so: Y_k = grupna vsota razredu k

F_k = kumulativna frekvenca v razredu k

$Y = \sum Y_k$

N = število enot

10.27 Za zgled vzemimo industrijska podjetja v SFRJ po številu zaposlenih v letu 1963. Postopek računa za Δ_1 je nakazan v tabeli 10.4.

Tabela 10.4 Izračunanje poprečne razlike Δ_1 za število zaposlenih v industrijskih podjetjih v SFRJ v letu 1963 (Vir: SG-65,str.76)

Število zaposlenih	Število podjetij			Število zaposlenih Y_k	$(F_k+F_{k+1})Y_k$
	f_k	F_k	$F_k + F_{k+1}$		
- 15	51	0	51	507	25857
16 - 29	82	51	184	1831	336904
30 - 60	202	133	468	9236	4322448
61 - 125	464	335	1134	42038	47671092
126 - 250	581	799	2179	104848	228463792
251 - 500	520	1380	3280	184033	603628240
501 - 1000	310	1900	4110	219033	900225630
1001 - 2000	195	2210	4615	268391	1238624465
2001 -	102	2405	4912	386879	1900349648
		N = 2507		Y =1216796	4923648076

$$\Delta_1 = \frac{2\left[4923648076 - 2507 \cdot 1216796\right]}{2507\,(2507 - 1)} = 596$$

Razmerje med Q, AD in SD za normalno porazdelitev

10.27 Za normalne porazdelitve so med Q, AD in SD stalni odnosi, ki jih navajamo zaradi tega, ker iz njih lahko sklepamo na približne odnose za druge simetrične, zvonaste in unimodalne porazdelitve.

Za normalno porazdelitev velja:

$$Q = 0.6745\ \sigma \doteq 2/3\ \sigma \qquad (10.24)$$

$$AD = 0.7979\ \sigma \doteq 4/5\ \sigma \qquad (10.25)$$

Relativne mere variacije

10.28 Absolutne mere variacije, ki smo jih spoznali v prejšnjih odstavkih, moremo med seboj le redko primerjati. Nemogoče je primerjati absolutne mere variabilnosti za raznovrstne pojave, čeprav so v vsebinski zvezi. Ne moremo, recimo, primerjati variabilnost porabe alkoholnih pijač na prebivalca z variabilnostjo za porabo cigaret na prebivalca ali z variabilnostjo za porabo kruha. Mere variacije vsakega izmed teh pojavov imajo različno enoto mere in zato med seboj niso primerljive. Absolutnih mer variacij pa ne moremo primerjati tudi pri istovrstnih pojavih. 15 para razlike v ceni jajc je mnogo pomembnejše kakor na primer 8 din v ceni blaga za moške plašče. Če vzamemo, da je poprečna cena jajc npr. '75 din, je '15 din 20% povečanje; če pa vzamemo, da je poprečna cena blaga za moške suknje npr. 400 din, je 8 din razlike le 2 %. Čeprav v obeh primerih opazujemo razlike v dinarjih, teh razlik ne moremo primerjati, ker se nanašajo na artikle z zelo različnimi višinami cen. Zato absolutne mere variacije lahko primerjamo za istovrstne pojave, pa še za te le tedaj, če je raven, ki je izražena s srednjo vrednostjo, med primerjanima pojavoma ista. Te omejitve v uporabi absolutnih mer variacije odpravimo, če izražamo mere variacije relativno v razmerju do neke srednje vrednosti. Kvocient med mero variacije in ustrezno srednjo vrednostjo je neimenovano število, ki ga običajno izražamo z odstotkom. Zato moremo relativne mere variacije primerjati med seboj tudi za raznovrstne pojave, ki so vsebinsko odvisni. Ker pa meri relativna mera variacije absolutno variabilnost v razmerju s srednjo vrednostjo, moremo med seboj primerjati tudi variabilnosti podatkov na različnih ravneh.

10.29 Relativne mere variacije moremo izračunati iz vseh vrst absolutnih mer variacije: R , Q, AD, SD in Δ_1. Vsako izmed teh primerjamo z ustrezno srednjo vrednostjo. Variacijski razmik R primerjamo običajno z aritmetično sredino med skrajnima vrednostima $1/2\ (y_{min} + y_{max})$, kvartilni odklon Q z mediano ali aritmetično sredino med Q_1 in Q_3, poprečni absolutni odklon AD z mediano ali aritmetično sredino,

standardni odklon SD z aritmetično sredino M, poprečno razliko Δ_1 z aritmetično sredino M.

Tako dobimo naslednje relativne mere variacije:

Na osnovi variacijskega razmika

1) $$\frac{2(y_{max} - y_{min})}{y_{min} + y_{max}} \qquad (10.26)$$

Na osnovi kvartilnega razmika $Q_3 - Q_1$

2) $$\frac{Q_3 - Q_1}{M_e} \quad \text{ali} \quad \frac{2(Q_3 - Q_1)}{Q_1 + Q_3} \qquad (10.27)$$

Na osnovi poprečnega absolutnega odklona AD

3) $$\frac{AD_{Me}}{M} \; , \; \frac{AD_M}{M} \qquad (10.28)$$

Na osnovi standardnega odklona σ

4) $$KV = \frac{\sigma}{M} \qquad (10.29)$$

Na osnovi poprečne razlike Δ_1

5) $$\frac{\Delta_1}{M} \qquad (10.30)$$

Koeficiente variacije izražamo kakor je nakazano v gornjih izrazih ali v odstotkih tako, da zgornja razmerja pomnožimo s 100. Pobliže se bomo seznanili z relativnima merama variacije, ki sta zasnovani na poprečnem absolutnem odklonu in na standardnem odklonu.

10.30 Če proučimo relativno mero variacije zasnovano na AD_{Me} dobimo iz obrazca 10.6

$$\frac{AD_{Me}}{M} = Y_z^o - Y_s^o = 2Y_z^o - 1 = 1 - 2Y_s^o \qquad (10.31)$$

Pri tem sta Y_s^o in Y_z^o strukturna deleža za agregat Y pod oziroma nad mediano.

Za relativno mero variacije, zasnovano na AD_M pa je

$$\frac{AD_M}{M} = (Y_z^o - Y_s^o) - (N_z^o - N_s^o) = 2\,(Y_z^o - N_z^o) = 2\,(N_s^o - Y_s^o) \qquad (10.32)$$

Pri AD_M/M pomeni Y_s^o in Y_z^o strukturni delež za agregat Y pod oziroma nad poprečjem M, N_s^o oziroma N_z^o pa delež enot pod oziroma nad poprečjem M.

10.31 Za zgled izračunajmo AD_{Me}/M za blagovni promet na prebivalca iz 8.1. Ker je

$$Y_z^o = Y_z/Y = 183978/263325 = 0,69867,$$

je ustrezen koeficient variacije

$$2\,Y_z^o - 1 = 2.0,69867 - 1 = 0,39734$$

AD_M/M ustreza $N_z^o = 18/60 = 0,30$ in $Y_z^o = Y_z/Y =$

$= 137076/263325 = 0,52056$. Iz obrazca 10.32 sledi

$$AD_M/M = 2\,(\cdot 52056 - \cdot 30) = \cdot 44112$$

10.32 Koeficient variacije, izračunan na osnovi standardnega odklona ima operativno obliko

$$KV = \sqrt{\frac{Q_i}{Q} - 1} \quad ; \quad Q_i = \sum y_i^2 \quad ; \quad Q = \frac{Y^2}{N} \qquad (10.33)$$

Standardni odklon za ceno jabolk, ki pokaže variabilnost cene jabolk med 35 izbranimi kraji v Jugoslaviji, je bil v septembru 1957 SD = 13,2 din, v januarju 1958 pa SD = 27,4 din. Absolutna mera variacije pokaže, da je bila regionalna variacija cen jabolk v januarju večja kakor v septembru. Če pa računamo, da je bila poprečna cena jabolk v septembru 1957 40,4 din, v januarju 1958 pa 113,6 din, sklepamo, da je bila variabilnost v ceni jabolk v septembru pomembnejša kakor v januarju, ker je koeficient variacije cen KV v septembru 1957

$$KV\% = 100 \cdot \frac{13,2}{40,4} = 32,7\ \%$$

v januarju 1958 pa

$$KV\% = 100 \cdot \frac{27,4}{113,6} = 24,1\ \%$$

V septembru je standardni odklon približno tretjino od poprečja, v januarju pa samo približno četrtino.

Še bolj izrazito vidimo pomen relativne mere variacije v regionalnem variiranju cen za več različnih predmetov.

Tabela 10.5. Koeficienti variacije za pet predmetov široke porabe med 35 izbranimi kraji v Jugoslaviji v januarju 1958 in za 20 izbranih krajev za Ø 1970. (Vir: Po podatkih v SB 126)

Naziv artikla		januar 1958			Ø 1970
		v din		KV%	KV%
		M	SD		
Enoten kruh	kg	47,3	2,5	5,3	5,1
Svinjska mast uvožena	kg	325,1	9,8	3,0	2,6
Jabolka	kg	113,6	27,4	24,1	17,0
Jajca	kos	17,8	3,8	21,3	9,4
Prava kava	kg	2086,0	54,2	2,6	3,6

V tabeli 10.5 vidimo, da je standardni odklon za ceno prave kave SD = 54,2 din; ta je od vseh navedenih predmetov največji, stvarno pa najmanj pomemben, ker znaša komaj KV% = 2,6% od poprečne cene. Iz tabele vidimo, da sta koeficienta variacije za jabolka in jajca razmeroma izredno velika. Vzrok temu je tipično krajevni značaj teh predmetov, če jih primerjamo z drugimi tremi. Čeprav je v tabeli 10.5 za januar 1958 vključenih 35 izbranih krajev, v letu 1970 pa samo 20, moremo vseeno izvesti tudi časovno primerjavo koeficientov variacije. Medtem ko sta se koeficienta variacije za kruh in mast relativno malo spremenila, se izkaže, da je za poprečne cene leta 1970 koeficient variacije za jabolka in jajca znatno manjši kot v januarju 1958. Ne glede na časovni razmik dvajset let je ta razlika rezultat sezonskega značaja obeh predmetov. Izraz razlik v času pa je po vsej verjetnosti povečanje koeficienta variaci-

je za pravo kavo.

10.33 Relativni odkloni. Pri proučevanju posamičnih vrednosti populacije merimo z razliko posamičnih vrednosti od aritmetične sredine

$$d_i = y_i - M \tag{10.34}$$

jakost vpliva posamičnih faktorjev na i-to enoto.

Pogosto ta odklon merimo relativno tako, da ga primerjamo z aritmetično sredino M

$$v_i = \frac{y_i - M}{M} \tag{10.35}$$

Relativni odkloni v_i so neimenovana števila in so zato med seboj primerljivi tudi za raznovrstne podatke.

Ker je relativen odklon v_i tudi statistični znak, moremo zanj izračunati vse tiste parametre, ki jih računamo za osnovni znak y. Če izračunamo iz v_i aritmetično sredino, spoznamo, da je $M_v = 0$.

Če poiščemo za v_i standardni odklon, je

$$\sigma_v^2 = \frac{\sigma^2}{M^2} \tag{10.36}$$

Standardni odklon relativnega odklona v_i

$$\sigma_v = \sqrt{\frac{\sigma_y^2}{M^2}} = \frac{\sigma_y}{M} = KV_y \tag{10.37}$$

pa je koeficient variacije za osnovni znak.

10.34 Standardiziran z- odklon. Razen z M primerjamo oziroma merimo absolutne odklone od aritmetične sredine $d_i = y_i - M$ tudi s standardnim odklonom. Tako dobimo standardiziran odklon

$$z_i = \frac{y_i - M}{\sigma} \tag{10.38}$$

Za normalno porazdelitev je za približno 2/3 odklonov absolutna vrednost za z manjša kot ena, približno 21/22 manjših kot 2 in približno 369/370 manjših kot 3. Ti odnosi veljajo približno za porazdelitve, ki so normalni podobne. S standardiziranim z-odklonom moremo za posamezne enote za porazdelitve, ki so normalne ali so normalni podobne, določiti mesto enote v populaciji. Predznak pri z pomeni, da je osnovna vrednost nad ali pod poprečjem, absolutna vrednost za z pa pokaže, za koliko je posamična vrednost odklonjena od M. Če izračunamo za z aritmetično sredino in standardni odklon, dobimo

$$M_z = \frac{1}{N}\sum z = \frac{1}{N}\sum \frac{y_i - M}{\sigma} = 0 \qquad (10.39)$$

$$\sigma_z^2 = \frac{1}{N}\sum (z - M_z)^2 = \frac{1}{N}\sum \left(\frac{y_i - M}{\sigma}\right)^2 = \frac{1}{\sigma^2} \cdot \frac{1}{N}\sum (y_i - M)^2 = 1 \qquad (10.40)$$

Mere variacije za relativna števila

10.35 Prav tako kot so poseben problem poprečja iz relativnih števil, so problematične tudi mere variacije iz relativnih števil

Variacijski razmik in kvartilni odklon moremo brez težav določiti tudi za relativna števila. Tako moremo ugotoviti variacijski razmik in kvartilni odklon za poprečne hektarske pridelke med občinami, za odstotek izkoriščanja kapacitet med podjetji, za produktivnost dela med delavci ipd.

Pri računanju variance in iz nje izvedenih pokazovalcev - standardnem odklonu in koeficientu variacije pa smo pogosto v dvomu, kako računati ustrezne parametre: tehtano ali netehtano.

Medtem ko računamo netehtano varianco iz relativnih števil R_k po obrazcu

$$\sigma_R^2 = \frac{1}{N}\sum_k (R_k - \bar{R})^2 \; ; \qquad \bar{R} = \frac{1}{N}\sum_k R_k = \frac{1}{N}\sum \frac{Y_k}{X_k} \qquad (10.41)$$

je obrazec za računanje variance s tehtanjem X_k

$$\sigma_R^2 = \frac{1}{X}\sum X_k (R_k - R)^2 \; ; \qquad R = \frac{\sum Y_k}{\sum X_k} \qquad (10.42)$$

Pri odločitvi, kateri izmed gornjih načinov je ustrezen, je problem podoben kot pri računanju poprečnega relativnega števila, ki ga moremo izračunati netehtano ali tehtano.

10.36 Če npr. proučujemo produktivnost dela za skupino delavcev, računamo poprečno produktivnost dela in varianco v produktivnosti dela netehtano. V tem primeru je enota opazovanja delavec in proučujemo produktivnost dela delavca ne glede na to ali je izračunana iz daljšega ali krajšega časovnega razmika. Drugače pa je z izračunanjem poprečne produktivnosti in variance za skupino podjetij, koeficientov izkoriščenosti kapacitet med panogami ali republikami ipd. Za take vrste pojavov je vsebinsko bolj umestno računati tehtano poprečje in tehtano varianco, ker majhna enota manj prispeva k poprečju kot velika in je zanjo odklon od poprečja manj pomemben kot za velike.

10.37 Če upoštevano, da je

$$R_k = \frac{Y_k}{X_k} \quad ; \qquad R = \frac{Y}{X} \tag{10.43}$$

dobimo s preračunom iz obrazca 10.42 operativni obrazec

$$\sigma_R^2 = \frac{1}{X}\sum X_k (R_k - R)^2 = \frac{1}{X}\left[\sum \frac{Y_k^2}{X_k} - \frac{Y^2}{X}\right] = \frac{1}{X}\left[Q_k - Q\right] \tag{10.44}$$

Pri tem sta $Q_k = Y_k^2/X_k$ in $Q = Y^2/X$ izraza, ki se pogosto pojavljata pri statistični analizi.

Ker pri analizah variabilnosti v socialno-ekonomskih pojavih pogosto analiziramo variabilnost večih pojavov hkrati, je zelo pomemben koeficient variacije iz relativnih števil. Tega dobimo po obrazcu

$$KV_R = \frac{\sigma_R}{R} = \frac{X}{Y}\sqrt{\frac{1}{X}(Q_K - Q)} = \sqrt{\frac{Q_K}{Q} - 1} \tag{10.45}$$

10.38 Za zgled računanja koeficienta variacije iz relativnih števil vzemimo KV_R za koeficient izrabljenosti osnovnih sredstev v industriji in rudarstvu v letu 1963 med

strokami v SFRJ. Koeficient izrabljenosti je razmerje med sedanjo in nabavno vrednostjo.

Tabela 10.6. Izračun KV_R za koeficient izrabljenosti osnovnih sredstev med strokami industrije in rudarstva v SFRJ v letu 1963. (Vir: SG 1965)

Stroka	Osnovna sredstva vrednost v 10^9din		$R_k = Y_k/X_k$	$\frac{Y_k^2}{X_k}$
	nabavna X_k	sedanja Y_k		
Elektroenergija	861,2	624,8	73	453,3
Premog in koks	304,3	183,5	60	110,7
Nafta	122,9	66,4	54	35,9
Črna metalurgija	313,2	211,3	67	142,6
Barvasta metalurgija	266,5	170,1	64	108,6
Nekovine	89,0	57,9	65	37,7
Kovinska	449,0	260,9	58	151,6
Ladjedelništvo	62,9	35,4	56	19,9
Elektro	95,4	62,0	65	40,3
Kemična	242,7	157,9	65	102,7
Gradbeni material	146,6	87,0	59	51,6
Lesna	159,4	100,2	63	63,0
Papirna	110,8	78,4	71	55,5
Tekstilna	296,8	185,1	62	115,4
Usnje in obutev	41,6	26,1	63	16,4
Guma	23,4	12,0	51	6,1
Živilska	219,4	146,3	67	97,6
Grafična	50,7	27,8	55	15,2
Tobak	31,7	20,5	65	13,3
Filmska	4,7	3,0	64	1,9
Rudarske raziskave	11,0	5,7	51	3,0
Razna	6,5	4,1	62	2,6
Skupno	3909,5	2526,4	65	$1644,9 = Q_k$
	X	Y	R	$\sum Y_k^2/X_k$
				$1632,6 = Q$
				Y^2/X

Iz teh podatkov dobimo

$$KV_R = \sqrt{\frac{Q_K}{Q} - 1} = \sqrt{\frac{1644.9}{1632,6} - 1} = \cdot 087$$

ali v odstotkih $KV_R\,\% = 8,7\,\%$.

10.39 Analitično uporabnost KV_R iz relativnih števil osvetlimo še s koeficienti variacije za pet osnovnih razmerij ekonomskih pokazovalcev SFRJ v letu 1962 in 1969 med republikami v industriji.

Tabela 10.7. Koeficienti variacije za pet osnovnih razmerij v industriji SFRJ med republikami (Vir SG-70)

Sedanja vrednost osnovnih sredstev = OS(SV)

Nabavna vrednost osnovnih sredstev = OS(NV)

Družbeni proizvod = DP

Neto proizvod = NP

Neto osebni dohodki = NOD

Koeficient	Koeficient variacije $KV_R\,\%$	
	1962	1969
OS(SV) : OS(NV)	10,3	18,3
DP : OS (SV)	27,3	23,0
NP : OS(SV)	29,1	24,4
NP : DP	2,1	1,7
NOD : NP	4,0	6,2

Primerjava nakazanih koeficientov variacije kaže na povečanje razlik v izrabljenosti osnovnih sredstev med republikami in na razmeroma veliko variabilnost družbenega produkta in neto produkta na enoto osnovnih sredstev.

10.40 Podobni problemi kot so pri računanju mer variacije iz relativnih števil, so tudi pri računanju mer variacije iz grupnih poprečij. Če upoštevamo število enot N_k kot uteži, veljajo za tehtano varianco iz grupnih sredin naslednji obrazci

$$\sigma_M^2 = \frac{1}{N}\sum N_K(M_K - M)^2 = \frac{1}{N}\left[\sum \frac{Y_k^2}{N_k} - \frac{Y^2}{N}\right] = \frac{1}{N}\left[Q_K - Q\right] \qquad (10.46)$$

10.41 Za zgled izračunajmo tehtano varianco iz grupnih sredin za poprečni neto osebni dohodek po stopnji strokovne izobrazbe na delovnem mestu v SFRJ (poprečje 1967).

Tabela 10.8. Izračunanje variance za poprečen neto osebni dohodek po stopnji strokovne izobrazbe (Vir SG 70, str. 266)

Strokovna izobrazba	Število zaposlenih v 000 din N_k	Poprečni osebni dohodek v din M_k
Visoka strokovna izobrazba	224	1559
Višja strokovna izobrazba	159	1133
Srednja strokovna izobrazba	533	899
Nižja strokovna izobrazba	174	703
Visokokvalificirani delavec	328	1002
Kvalificiran delavec	1059	735
Polkvalificiran delavec	547	610
Nekvalificiran delavec	506	555

$$M = \frac{\sum N_K M_K}{\sum N_k} = \frac{224.1559 + \ldots + 506.555}{224 + \ldots + 506} = \frac{2852373}{3530} = 808$$

$$Q_K = \sum N_k M_k^2 = 224.1559^2 + \ldots + 506.555^2 = 2526108731$$

$$Q = NM^2 = Y^2/N = 2852373^2/3530 = 2304824853$$

$$K_M = \sum N_k M_k^2 - N.M^2 = 2526108731 - 2304824853 = 221283878$$

$$\sigma_M^2 = K_M/N = 221283878/3530 = 62686,65$$

$$\sigma_M = \sqrt{\sigma_M^2} = \sqrt{62686,65} = 250,37 \text{ din}$$

$$KV\% = 100.250,37/808 = 31\ \%$$

Mere asimetrije in sploščenosti

10.42 Za frekvenčne porazdelitve sta razen centralne tendence in variabilnosti pomembni tudi asimetrija in sploščenost. Porazdelitve so simetrične, asimetrične v levo in asimetrične v desno. Stopnja asimetrije more biti večja ali manjša. Razen tega porazdelitve z isto srednjo vrednostjo, isto mero variabilnosti in stopnjo asimetrije niso nujno med seboj enake, marveč so ene bolj druge pa manj sploščene. Sploščenost normalne porazdelitve štejemo za idealno. Z njo primerjamo sploščenost drugih porazdelitev.

Mere asimetrije in sploščenosti po legi

10.43 **Mere asimetrije.** Že pri srednjih vrednostih smo navedli, da so za simetrične, unimodalne porazdelitve vrednosti modusa, mediane in aritmetične sredine enake ($Mo=Me=M$). Za unimodalne porazdelitve, ki so asimetrične v desno, je modus manjši, aritmetična sredina pa večja kakor mediana ($Mo<Me<M$). Za unimodalne porazdelitve, ki so asimetrične v levo, pa se vrstni red zamenja in je asimetrična sredina manjša, modus pa večji kot mediana ($M>Me>Mo$).

10.44 Za zvezne, ne preveč asimetrične porazdelitve velja, da je razlika med aritmetično sredino in modusom približno trikrat večja kakor razlika med aritmetično sredino in mediano.

$$M - M_o = 3\,(M - M_e) \qquad (10.47)$$

Pri isti variabilnosti so razlike med M, Me in Mo tem večje, čim večja je stopnja asimetrije. To lastnost frekvenčnih porazdelitev izkoriščamo in merimo asimetrijo s koeficientom asimetrije, ki ima za osnovo razliko med aritmetično sredino in modusom

$$KA_{Mo} = \frac{M - M_o}{\sigma} \qquad (10.48)$$

ali s koeficientom asimetrije, ki ima za osnovo razliko med aritmetično sredino in mediano

$$KA_{Me} = \frac{3(M-Me)}{\sigma} \qquad (10.49)$$

10.45 Za simetrične porazdelitve sta tretji in prvi kvartil enako oddaljena od mediane. Pri asimetričnih porazdelitvah pa je tretji kvartil bolj oddaljen od mediane, če je porazdelitev asimetrična v desno, in prvi kvartil bolj oddaljen od mediane, če je porazdelitev asimetrična v levo. Pri isti variabilnosti so razlike tem večje, čim večja je stopnja asimetrije. To lastnost porazdelitev izkorišča koeficient asimetrije, ki ima za osnovo razliko v oddaljenosti tretjega in prvega kvartila od mediane

$$KA_Q = \frac{(Q_3 - M_e) - (M_e - Q_1)}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1} \qquad (10.50)$$

Vsi trije koeficienti KA_{Mo}, KA_{Me} in KA_Q so pozitivni, če je porazdelitev asimetrična v desno, enaki nič za simetrične porazdelitve, in negativni, če je porazdelitev asimetrična v levo.

Absolutno pa so koeficienti tem večji, čim večja je stopnja asimetrije.

10.46 Mera sploščenosti. Za mero sploščenosti vzamemo razmerje med kvartilnim in decilnim razmikom

$$KS = 1.9 \frac{Q_3 - Q_1}{D_9 - D_1} \qquad (10.51)$$

Koeficient sploščenosti KS je enak ena, če je porazdelitev normalna, za koničaste porazdelitve je $KS < 1$, za sploščene pa je $KS > 1$.

Mere asimetrije in sploščenosti, izračunane iz momentov

10.47 Kot osnovo za računanje mere asimetrije in sploščenosti uporabljamo v analitične namene tudi momente.

Centralni moment stopnje r

$$\mu_r = \frac{1}{N} \sum (y - M)^r \qquad (10.52)$$

je poprečna r-ta potenca odklonov posamičnih vrednosti od aritmetične sredine.

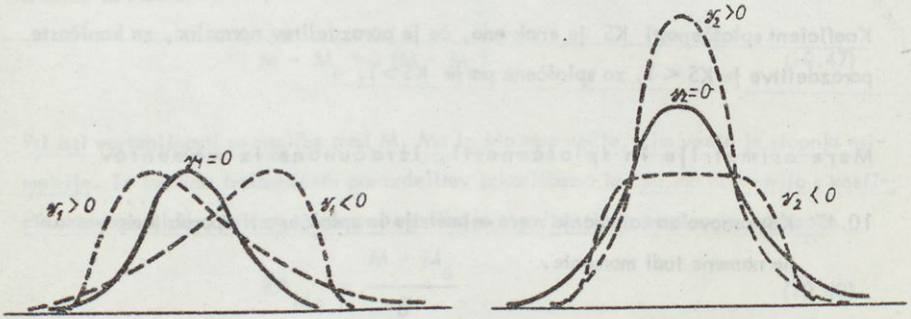
Iz te splošne opredelitve dobimo, da je: $\mu_o = 1$ $\mu_1 = 0$, $\mu_2 = \sigma^2$

Analiza tretjega in četrtega centralnega momenta pa pokaže, da moremo vzeti tretji centralni moment za osnovo za mero asimetrije, ker je μ_3 nič, če je porazdelitev simetrična, pozitiven, če je porazdelitev asimetrična v levo in negativen, če je porazdelitev asimetrična v desno. Podobno je vrednost četrtega centralnega momenta pri nespremenjenih drugih pogojih večja, če je porazdelitev koničasta in manjša, če je porazdelitev sploščena. μ_3 in μ_4 pa sta v bistveni meri odvisna tudi od variabilnosti pojava. Iz njih izločimo vpliv variabilnosti, če jih izračunamo iz standardiziranih odklonov. Tako dobimo, da je na splošno

$$\alpha_r = \frac{1}{N}\sum z^r = \frac{1}{N}\sum\left(\frac{\gamma - M}{\sigma}\right)^r = \frac{\mu_r}{\sigma^r} = \frac{\mu_r}{\sqrt{\mu_2^r}} \qquad (10.53)$$

ali $\alpha_o = 1$, $\alpha_1 = 0$, $\alpha_2 = 1$, α_3 pa je primerljivo merilo asimetrije za različne pojave, ker je neimenovano število in ni odvisen od variabilnosti pojava.

Za normalno porazdelitev je $\alpha_4 = 3$. Za porazdelitve, ki so v primerjavi z normalno porazdelitvijo koničaste, se izkaže, da je $\alpha_4 > 3$, za porazdelitve, ki so v primerjavi z normalno sploščene, pa je $\alpha_4 < 3$.



Slika 10.3. Asimetrija in sploščenost porazdelitev

Iz α_3 in α_4 izpeljemo koeficienta, ki imata vrednost 0, če sta po simetričnosti in sploščenosti enaka normalni, če vpeljemo parametra

$$\gamma_1 = \alpha_3 \; ; \quad \gamma_2 = \alpha_4 - 3 \qquad (10.54)$$

10.48 Računanje centralnih momentov oziroma koeficientov asimetrije in sploščenosti si olajšamo s pomožnimi momenti

$$\nu_r = \frac{1}{N} \sum (y - y_o)^r = \frac{1}{N} \sum u^r \qquad (10.55)$$

ki so podobni centralnim momentom, le da odklone posamičnih vrednosti ne računamo od aritmetične sredine, temveč od poljubne vrednosti y_o.

Če upoštevamo binomski stavek, dobimo, da veljajo zveze

$$\begin{aligned} \mu_2 &= \nu_2 - \nu_1^2 \\ \mu_3 &= \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 \\ \mu_4 &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4 \end{aligned} \qquad (10.56)$$

Te zveze omogočajo računanje centralnih momentov iz pomožnih momentov.

10.49 Za primer vzemimo urne učinke za skupino 10 delavcev. Shematično računanje je prikazano v tabeli 10.9.

Tabela 10.9 Računanje γ_1 in γ_2 prek pomožnih momentov

y	$u=y-39$	u^2	u^3	u^4
40	+1	1	1	1
32	-7	49	-343	2401
41	+2	4	8	16
35	-4	16	-64	256
40	+1	1	1	1
36	-3	9	-27	81
39	0	0	0	0
40	+1	1	1	1
39	0	0	0	0
38	-1	1	-1	1
$\sum u^r$	-10	82	-425	2758
ν_r	- 1,0	8,2	- 42,5	275,8

$$\mu_2 = 8{,}2 - (-1{,}0)^2 = 7{,}2$$
$$\mu_3 = -42{,}5 - 3.8{,}2 \cdot (-1{,}0) + 2(-1{,}0)^3 = -19{,}9$$

$$\mu_4 = 275,8 - 4 \cdot (-42,5)(-1,0) + 6.8,2 \cdot (-1,0)^2 - 3 \cdot (-1)^4 = 152,0$$

$$\alpha_3 = \frac{-19,9}{7,2^3} = -1,03$$

$$\alpha_4 = \frac{152,0}{7,2^2} = 2,93$$

Dalje je $\gamma_1 = \alpha_3 = -1,03$, $\gamma_2 = \alpha_4 - 3 = 2,93 - 3 = -0,07$

Vrednosti se porazdeljujejo asimetrično v levo, ker je γ_1 negativen, sploščenost pa je podobna kot pri normalni porazdelitvi, ker se $\gamma_2 = 0,07$ zelo malo razlikuje od 0.

10.50 Pokazovalce asimetrije in sploščenosti pa pogosteje računamo za večje populacije, za katere so podatki grupirani v frekvenčnih porazdelitvah.

Računanje pomožnih količin $\sum f u^r$ za pokazovalce asimetrije in sploščenosti je v tem primeru naravno nadaljevanje računanja $\sum f u$ in $\sum f u^2$ za aritmetično sredino in varianco, ki je po definiciji centralni moment druge stopnje.

Podobno kot za aritmetično sredino in varianco moremo pri računanju momentov uporabiti metodo pomožnega znaka u ali metodo kumulativ.

V shemi 10.11 je nakazan postopek računanja parametrov γ_1 in γ_2 po obeh metodah. Če uporabljamo metodo pomožnega znaka $u_k = (y_k - y_o)/i$ izračunamo najprej po znanem postopku količine $\sum f u^r$, iz teh ν_r, če vrednosti iz stolpca (3) delimo z N. Iz ν_r izračunamo μ'_r po obrazcih, ki so v shemi nakazani v stolpcu (5). V stolpcu (6) so dani obrazci za Sheppardov popravek momentov μ'_r. Iz popravljenih količin $\mu'_{r,pop}$ izračunamo po enakih obrazcih kot za negrupirane podatke α_r in iz teh dalje γ_1 in γ_2.

Če računamo pomožne momente s kumulativami, iz frekvenčne porazdelitve najprej izračunamo kumulativne vsote C_r, iz teh pa $\sum f u^r$ po obrazcih v stolpcu (3). Nadaljnji postopek pa je enak kot pri prvi metodi.

10.51 Za zgled vzemimo frekvenčno porazdelitev trdnosti konic svedrov, ki je merjena v urah zdržljivosti.

Tabela 10.10. Izračun pomožnih količin $\sum f u^r$ za porazdelitev trdnosti konic svedrov

Trdnost v urah	f	u	fu	fu^2	fu^3	fu^4	(u+1)	$(u+1)^4$	$f(u+1)^4$
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
27,5 - 29,5	3	-4	-12	48	-192	768	-3	81	243
29,5 - 31,5	9	-3	-27	81	-243	729	-2	16	144
31,5 - 33,5	23	-2	-46	92	-184	368	-1	1	23
33,5 - 35,5	33	-1	-33	33	- 33	33	0	0	0
35,5 - 37,5	38	0	0	0	0	0	1	1	38
37,5 - 39,5	34	1	34	34	34	34	2	16	544
39,5 - 41,5	21	2	42	84	168	336	3	81	1701
41,5 - 43,5	8	3	24	42	216	648	4	256	2048
43,5 - 45,5	1	4	4	16	64	256	5	625	625
	170		- 14	460	-170	3172			5366

Kompleksen preskus razmeroma zamotanega računanja količin dobimo s Charlierovim preskusom. Z uporabo četrte potence binoma $(u+1)^4$ sledi

$$\sum f(u+1)^4 = \sum fu^2 + 4\sum fu^3 + 6\sum fu^2 + 4\sum fu + N$$

$$5366 = 3172 + 4 \cdot (-170) + 6.460 + 4 \cdot (-14) + 170$$

Naslednjo shemo moremo uporabiti tudi za računanje γ_1 in γ_2 iz negrupiranih podatkov. V tem primeru velja, da je vrsta 7 enaka 9, ker ne pride v poštev Sheppardova korektura, ki je nakazana v vrsti 8 za grupirane podatke.

Tabela 10.11 Shema za izračunanje centralnih momentov in mer asimetrije in sploščenosti z zgledom za trdnost konic svedrov

r	1	0	1	2	3	4
C_r	2	Co=N	C_1	C_2	C_3	C_4
$\sum_r a_r C_r = \sum f u^r$	3	Co	C_1 -14	$C_1 + 2C_2$ $= 460$	$C_1 + 6C_2 + 6C_3 =$ $= -170$	$C_1 + 14C_2 + 36C_3 + 24C_4 =$ $= 3172$
ν_1^r	4	1	$\nu_1 = -0,08235$	$\nu_1^2 = +0,00678$	$\nu_1^3 = -0,00056$	$\nu_1^4 = 0,00005$
$\nu_r = \frac{1}{N}\sum f u^r$	5 $= \frac{3}{N}$	1	$\nu_1 = -0,08235$	$\nu_2 = 2,70588$	$\nu_3 = -1,00000$	$\nu_4 = 18,65882$
	6			$-\nu_1^2 = 0,00678$	$-3\nu_2\nu_1 = +0,66849$ $+2\nu_1^3 = 0,00112$	$-4\nu_3\nu_1 = 0,32940$ $+6\nu_2\nu_1^2 = +0,11008$ $-3\nu_1^4 = -0,00015$
μ_r	7=5+6 8		0	$\mu_2 = 2,69910$ $-\frac{1}{12} = -0,08333$	$\mu_3 = -0,33263$ 0	$\mu_4 = 19,09815$ $-\frac{1}{2}\mu_2 = 1,34955$ $+\frac{7}{240} = +0,02917$
$\mu_{r,pop}$	9=7+8			$\mu_{2,pop} = 2,61577$	$\mu_{3,pop} = 0,33263$	$\mu_{4,pop} = 17,77777$
$(\mu_{2,pop})^{r/2} = \sigma^r$	10		$\sigma = 1,61733$	$\sigma^2 = 2,61577$	$\sigma^3 = 4,23056$	$\sigma^4 = 6,842253$
$\alpha_r = \frac{\mu_{r,pop}}{\sigma^r}$	11 $= \frac{9}{10}$		$\alpha_1 = 0$	$\alpha_2 = 1$	$\alpha_3 = -0,0786$	$\alpha_4 = 2,5982$
γ_{r-2}	12				$\gamma_1 = \alpha_3 =$ $= 0,0786$	$\gamma_2 = \alpha_4 - 3$ $= -0,4018$

ENAJSTO POGLAVJE

11 MERE KONCENTRACIJE

11.0 Agregati morejo biti porazdeljeni po enotah populacije različno. Tako dobe posamezni delavci različen del skupnega fonda osebnih dohodkov. Skupna kmetijska površina je neenakomerno razdeljena med posamične kmetijske obdelovalce in imajo neki obdelovalci velike površine, drugi pa majhne. Podobno imajo od skupnega števila zaposlenih neka podjetja malo, druga pa veliko zaposlenih. Čeprav gornja lastnost agregatov pride do izraza v variiranju, vpeljemo za nakazano lastnost agregatov nov pojem - koncentracijo pojava. Če je pojav enakomerno porazdeljen po enotah, pravimo da koncentracije ni, če je pojav porazdeljen po enotah populacije tako, da ima nekaj enot razmeroma velik del agregata, večina pa le majhen del, govorimo o veliki koncentraciji pojava.

Tako opredeljen pojem koncentracije v statističnem smislu je premosorazmeren z variacijo pojava. Če pojav ne variira, koncentracije ni, velika koncentracija pa se izraža v velikem variiranju.

11.1 Lorenzov grafikon. Ker frekvenčna porazdelitev kaže variabilnost pojava, je iz nje posredno mogoče sklepati tudi na koncentriranost. Lepše kot iz grafičnega prikaza za frekvenčno porazdelitev pa je koncentracija pojava vidna iz Lorenzovega grafikona. Vzemimo, da imamo v frekvenčni porazdelitvi razen frekvenc f_k dane še podatke o delu agregata, ki odpade na posamezne razrede Y_k. Lorenzov grafikon narišemo tako, da v kvadrat s stranicama za kumulative relativnih frekvenc F% (abscisa) in kumulative relativnih vsot $\emptyset$ % (ordinata) vnesemo točke s koordinatama (F_k%, $\emptyset_k$%). Če povežemo po vrsti vrisane točke, dobimo lomljeno črto, ki v loku povezuje spodnje levo oglišče kvadrata (0,0) z zgornjim desnim ogliščem (100,100). V primeru, da na vsako enoto odpade

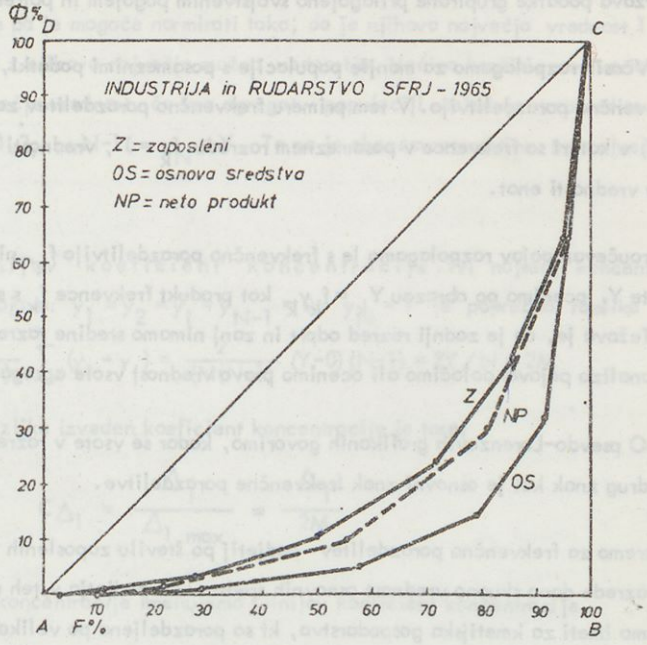
enak del agregata, pojav ni koncentriran, Lorenzova krivulja pa poteka v diagonali. Če je koncentriranost pojava velika je pri večanju koncentracije Lorenzova krivulja bolj in bolj ukrivljena in pri največji koncentraciji preide v lomljeno črto $\overline{AB} - \overline{BC}$. Za stvarne primere Lorenzova krivulja poteka med diagonalo $\overline{AC}$ in stranicama kvadrata $\overline{AB} - \overline{BC}$, odvisno od stopnje koncentracije.

11.2 Za zgled Lorenzovega grafikona vzemimo industrijska podjetja v letu 1965 v SFRJ po številu zaposlenih. Za Lorenzov grafikon potrebne količine so izračunane v tabeli 11.1.

Tabela 11.1 Izračunanje količin za Lorenzov grafikon za število zaposlenih v letu 1965 v SFRJ

število zaposlenih	f_k	F_k%	Y_k	φ_k%
- 15	55	0	516	0
16 - 29	56	2,23	1214	0,04
30 - 60	171	4,50	7979	0,13
61 - 125	407	11,44	37457	0,71
126 - 250	547	27,94	99062	3,45
251 - 500	542	50,12	192701	10,68
501 - 1000	349	72,10	246873	24,77
1001 - 2000	216	86,25	303859	42,80
2001 -	123	95,01	478814	65,01
	2466	100,00	1368475	100,00

V grafikonu 11.1 sta razen za število zaposlenih v industriji in rudarstvu narisani Lorenzovi krivulji še za osnovna sredstva in neto produkt za SFRJ v letu 1965. Ker vrisujemo v Lorenzov grafikon relativna števila, moremo v isti grafikon vrisati Lorenzove krivulje za raznovrstne pojave, ki pa morajo biti seveda v vsebinski zvezi. Iz grafikona sklepamo, da je koncentracija podjetij glede na neto produkt malenkostno višja kot za zaposlene, da pa je koncentracija osnovnih sredstev po podjetjih v industriji znatno večja kot je koncentracija zaposlenih.



Grafikon 11.1 Lorenzov grafikon za število zaposlenih, vrednost osnovnih sredstev in vrednost neto produkta v industriji in rudarstvu v SFRJ v letu 1965.

11.3 V Lorenzovem grafikonu z daljicami, ki vežejo posamezne točke, podajamo linearno interpolacijo poteka za Lorenzovo krivuljo. Zato moremo iz njega odbrati raznovrstne odnose med kumulativo relativnih vsot in frekvenc. Tako moremo iz njega oceniti nekaj pomembnih odnosov. Med drugim so pomembni odnosi, kolik del agregata ima polovica manjših enot in kolik del največjih enot v populaciji ima polovico agregata. Iz grafikona ocenimo, da ima polovica podjetij (majhnih), zaposlenih 11% od vseh zaposlenih v industriji in rudarstvu, 8,5% neto produkta in 4% vseh osnovnih sredstev. Obratno pa odberemo, da ima 11% velikih podjetij polovico vseh zaposlenih, da ima 10,5% velikih podjetij polovico neto produkta in 6% velikih podjetij polovico vseh osnovnih sredstev.

Lorenzov grafikon ima razen tega, da moremo na njem narisati Lorenzove krivulje za raznovrstne pojave tudi to prednost, da moremo na isti grafikon risati Lorenzove krivulje iz različno grupiranih podatkov. To posebno olajša mednarodne primerjave, ker ima

vsaka država podatke grupirane prilagojeno svojstvenim pogojem in potrebam.

11.4 Včasih razpolagamo za manjše populacije s posameznimi podatki, ne pa s frekvenčno porazdelitvijo. V tem primeru frekvenčno porazdelitev zamenja ranžirna vrsta, v kateri so frekvence v posameznem razredu $f_k = 1$, vrednosti $Y_k = y_k$ pa posamezne vrednosti enot.

Če za proučevan pojav razpolagamo le s frekvenčno porazdelitvijo f_k, nimamo pa vsot Y_k, vsote Y_k ocenimo po obrazcu $Y_k = f_k y_k$ kot produkt frekvence f_k s sredino razreda y_k. Težava je, če je zadnji razred odprt in zanj nimamo sredine razreda. V tem primeru z analizo pojava določimo ali ocenimo pravo vrednost vsote agregata.

11.5 O psevdo-Lorenzovih grafikonih govorimo, kadar se vsote v razredih nanašajo na drug znak kot je osnovni znak frekvenčne porazdelitve.

Tako moremo za frekvenčno porazdelitev podjetij po številu zaposlenih imeti za posamezne razrede dano skupno vrednost osnovnih sredstev za podjetja v teh razredih. Podobno moremo imeti za kmetijska gospodarstva, ki so porazdeljena po velikosti površine, število živine po posameznih razredih. Formalno moremo v vseh teh primerih računati kumulative relativnih frekvenc in kumulative relativnih vsot in vrisati te podatke v grafikon, ki je podoben Lorenzovemu grafikonu, ni pa Lorenzov grafikon v pravem smislu, ker se frekvenčna porazdelitev nanaša na drug znak kot pa vsota. Zato imenujemo tovrstne grafikone psevdo-Lorenzove grafikone. Medtem ko more biti krivulja v pravem Lorenzovem grafikonu le v trikotniku $\triangle ABC$, more biti krivulja v psevdo-Lorenzovem grafikonu na celem kvadratu $\square ABCD$. Vsekakor ima psevdo-Lorenzov grafikon svojo analitično vrednost, vendar je treba pri analizi upoštevati njegove posebnosti.

Koeficienti koncentracije

11.6 Od različnih mer koncentracije v praksi pogosto uporabljamo Gini-jev koeficient koncentracije in Hirschmanov koeficient koncentracije. Ugotovili smo, da je stopnja koncentracije premosorazmerna meram variacije. Koncentracije pojava ni, če imajo vse enote enako vrednost - če ne variirajo. Obratno je pri veliki stopnji koncentracije variabilnost velika. Hiba mer variacije za merjenje koncentracije je le ta, da tako absolutne kot relativne mere variacije nimajo ustrezne zgornje meje.

Mere variacije pa je mogoče normirati tako, da je njihova največja vrednost 1, če jih reduciramo z njihovo največjo možno vrednostjo. Večina koeficientov variacije ima svojo največjo vrednost, če ima agregat v populaciji naslednjo razporeditev: $y_i = 0$; $i = 1,2,\ldots,N-1$; $y_N = Y$. To pa je obenem razmestitev pri največji koncentraciji.

11.7 Ginijev koeficient koncentracije. Pri največji koncentraciji pojava: $y_1 = y_2 = y_i = y_{N-1} = 0$; $y_N = Y$ je poprečna razlika

$$\Delta_1 = \frac{2}{N(N-1)} \sum (y_i - y_j) = \frac{2}{(N(N-1)} (Y-0)(N-1) = 2Y/N = 2M$$

Iz poprečne razlike izveden koeficient koncentracije je torej

$$C_{\Delta_1} = \frac{\Delta_1}{\Delta_{1,max}} = \frac{\Delta_1}{2M} \qquad (11.1)$$

Ta koeficient koncentracije imenujemo Ginijev koeficient koncentracije.

11.7 Iz negrupiranih podatkov računamo Ginijev koeficient koncentracije prek obrazca za izračunanje poprečne razlike iz negrupiranih podatkov (obrazec 10.22).

Če upoštevamo ta obrazec, dobimo, da je Ginijev koeficient koncentracije enak

$$C_{\Delta_1} = 1 - \frac{2C_1}{(N-1)C_o} \qquad (11.2)$$

11.8 Za zgled koeficienta koncentracije iz negrupiranih podatkov naj služijo podatki o izvozu SFRJ v evropske države v prvem polletju 1965.

Tabela 11.3 Računanje Ginijevega koeficienta koncentracije za izvoz SFRJ v evropske države v I. polletju 1965. (Vir: Stat. spoljne trgovine, 1965)

Država	y_R	Kum (y_R)
8 držav (ostala Evropa)	68	0
Malta	72	68
Norveška	78	140
Finska	131	218
Španija	217	349
Albanija	250	566
Danska	593	816
Belgija	703	1409
Holandija	1444	2112
Švedska	1455	3556
Romunija	2005	5011
Francija	2229	7016
Švica	2593	9245
Grčija	2656	11838
Avstrija	3161	14494
Bolgarija	3238	17655
Madžarska	4026	20893
Velika Britanija	5510	24919
Čehoslovaška	8259	30429
Poljska	9019	38688
DR Nemčija	11392	47707
ZR Nemčija	11675	59099
Italija	19724	70774
ZSSR	20843	90498

$C_o = 111341$ $C_1 = 457500$

Če je v ranžirni vrsti n enot, za katere nimamo posameznih vrednosti, temveč le njihovo vsoto podatkov Y_1, C_1 korigiramo tako, da mu prištejemo $(n-1)Y_1$. Za

tak primer izračunamo Ginijev koeficient koncentracije po obrazcu

$$C\Delta_1 = 1 - \frac{2C_1 + (n-1)Y_1}{(N-1)C_o} \qquad (11.3)$$

Tako dobimo za naš zgled:

$$C\Delta_1 = 1 - \frac{2 \cdot 457500 + (8-1) \cdot 68}{(31-1).111341} = 0,726$$

11.9 Za podatke, ki so grupirani v frekvenčni porazdelitvi, moremo kot v vseh primerih, ko gre za grupirane podatke, izračunati le oceno za Ginijev koeficient koncentracije. Če ga računamo prek relativne mere variacije iz poprečne razlike, dobimo:

$$C\Delta_1 = \frac{\sum (F_k + F_{k+1}) Y_k}{N \cdot Y} - 1 \qquad (11.4)$$

pri čemer so F_k, Y_k . N in Y k obrazcu 10.23 pojasnjene količine.

Če vzamemo kot zgled število zaposlenih v industrijskih podjetjih v letu 1963 v SFRJ, je iz delnih rezultatov tega zgleda

$$C\Delta_1 = \frac{\Delta_1}{2M} = \frac{595}{2.486} = 0,613$$

11.10 Če namesto z absolutnimi podatki razpolagamo z relativnimi frekvencami $f_k\%$ in sestavo grupnih vsot $Y_k\%$, dobi obrazec 11.4 obliko

$$C\Delta_1 = \frac{\sum (F_k\% + F_{k+1}\%) Y_k\%}{10000} - 1 \qquad (11.5)$$

11.11 Vzemimo za zgled izračunanje Ginijevega koeficienta koncentracije ob gornjih pogojih za mesečne osebne dohodke v januarju 1966 v gospodarstvu v SFRJ .

Tabela 11.4. Izračunavanje Ginijevega koeficienta koncentracije za osebne dohodke zaposlenih v gospodarstvu SFRJ v januarju 1966

osebni dohodek Din	y_k	f_k%	F_k%	F_k%+F_k%	$Y_k = f_k Y_k$	$(F_k\% + F_{k+1}\%)Y_k$
- 300	150	6,9	0	6,9	1035	
301 - 400	350	17,5	6,9	31,3	6125	
401 - 500	450	20,1	24,4	68,9	9045	
501 - 600	550	16,2	44,5	105,2	8910	
601 - 700	650	11,7	60,7	133,1	7605	
701 - 800	750	8,1	72,4	152,9	6075	
801 - 1000	900	9,1	80,5	170,1	8190	
1001 - 1200	1100	4,7	89,6	183,9	5170	
1201 - 1400	1300	2,4	94,3	191,0	3120	
1401 -	1600	3,3	96,7	196,7	5280	
		100,0	100,0		60550	7678857,5
					Y	$\sum (F_k\% + F_{k+1}\%) Y_k$

$$C\Delta_1 = \frac{\sum (F_k\% + F_{k+1}\%)\, Y_k}{100 \cdot Y} - 1 = \frac{7678857{,}5}{100 \cdot 60550} - 1 = 0{,}268$$

Za september 1969 je koeficient koncentracije za osebne dohodke zaposlenih v gospodarstvu $C\Delta_1 = 0{,}231$.

11.12 Če opazujemo ploščino lika, ki ga oklepa Lorenzova krivulja in diagonala $\overline{AC}$ v Lorenzovem grafikonu, sklepamo, da je ta ploščina nič, če koncentracije ni in se Lorenzova krivulja sklada z diagonalo in je enaka ploščini trikotnika ABC, če je stopnja koncentracije maksimalna. Izkaže se, da je razmerje med ploščino Lorenzova krivulja - diagonala $\overline{AC}$ in ploščino trikotnika ΔABC enaka Ginijevem koeficientu koncentracije, izračunanim prek poprečne razlike.

11.13. Hirschmanov koeficient koncentracije. Standardni odklon je osnova za Hirschmanov koeficient koncentracije. Standardni odklon pri največji koncentraciji je

$$\sigma_{max} = \sqrt{\frac{1}{N}\sum (y_i - M)^2} = \sqrt{\frac{1}{N}\left[(N-1)M^2 + (Y - M)^2\right]} = M\sqrt{N-1} \qquad (11.6)$$

Iz njega dobimo koeficient koncentracije

$$C_\sigma = \frac{\sigma}{\sigma_{max}} = \frac{\sigma}{M\sqrt{N-1}} = \frac{KV}{\sqrt{N-1}} = \sqrt{\frac{N\sum (y_i/Y)^2 - 1}{N-1}} = \sqrt{\frac{Q_1/Q - 1}{N-1}} \qquad (11.7)$$

Pri velikem N dobimo za Hirschmanov koeficient koncentracije približek z naslednjim obrazcem

$$C_\sigma = \sqrt{\sum p_i^2} \qquad (11.8)$$

$p_i = y_i/Y$ = strukturni delež vrednosti posamezne enote v celotnem agregatu.

11.14 Za zgled izračunajmo Hirschmanov koeficient koncentracije za podatke o izvozu SFRJ v evropske države v prvem polletju 1965 iz tabele 11.3.

Za 8 držav "ostala Evropa" imamo samo delni agregat $Y_1 = 68$. Če predpostavljamo, da je za teh 8 držav izvoz enak, je prispevek teh držav k Q_1 enak $Y_1^2/f_1 = 68^2/8 = 578$

$$Q_1 = \frac{68^2}{8} + 72^2 + 78^2 + \ldots + 20843^2 = 1334101963$$

$$Q = 11341^2/31 = 399897364$$

$$Q_1/Q = \frac{1334101963}{399897364} = 3{,}336$$

$$C_\sigma = \sqrt{\frac{3{,}336-1}{31-1}} = 0{,}279$$

11.15. Mera koncentracije na osnovi AD_{Me}. Zbirko koeficientov koncentracije dopolnimo še s koeficientom, ki ima za osnovo AD_{Me}. Pri maksimalni koncentraciji je agregat razporejen: $y_1 = y_2 = y_3 = \ldots y_{N-1} = 0$; $y_N = Y$

Po obrazcu 10.5 je $AD_{Me} = \frac{1}{N}(Y_z - Y_s)$

$$AD_{Me,max} = \frac{1}{N}(Y - 0) \qquad (11.9)$$

Če sestavimo koeficient koncentracije kot kvocient med izračunanim AD_{Me} in $AD_{Me,max}$ dobimo, da je

$$C_{AD} = \frac{AD_{Me}}{AD_{Me,max}} = \frac{Y_z - Y_s}{Y} = Y_z^o - Y_s^o = 1 - 2\,Y_s^o \qquad (11.10)$$

11.16 Za zgled izračunajmo C_{AD} za izvoz savinjskega in bačkega hmelja iz SFRJ v letu 1969. Oba koeficienta izračunamo zaradi primerjave. Pri obračunu bomo vzeli v poštev samo države, v katere smo izvažali eno ali drugo vrsto hmelja. Osnovni podatki o količini izvoza hmelja so naslednji.

Tabela 11.5. Izvoz savinjskega in bačkega hmelja iz SFRJ v letu 1969.
Vir: Statistika spoljne trgovine SFRJ za 1969. god.

Država	savinjski t	bački t	SFRJ t
Avstrija	95	2 < Me	97
Belgija	39 = Me	22	61
Čehoslovaška	-	53	53
Danska	80	3 < Me	83
DR Nemčija	-	30	30 < Me
Francija	217	10 < Me	227
Grčija	14 < Me	-	14 < Me
Holandija	90	24	114
Italija	3 < Me	-	3 < Me
Madžarska	136	-	136
Romunija	-	390	390
ZR Nemčija	369	237	606
Švica	29 < Me	27	56
Švedska	13 < Me	18	31 < Me
Velika Britanija	866	14 < Me	880
Indija	-	5 < Me	5 < Me
Irak	1 < Me	-	1 < Me
Izrael	128	2 < Me	130

Država	savinjski t	bački t	SFRJ t
Japonska	-	53	53
Jordanija	3 < Me	-	3 < Me
Dahomej	5 < Me	-	5 < Me
Gornja Volta	5 < Me	-	5 < Me
Kamerun	10 < Me	-	10 < Me
Sirija	-	2 < Me	2 < Me
Slonokoščena obala	20 < Me	-	20 < Me
Egipt	-	10 < Me	10 < Me
Senegal	5 < Me	-	5 < Me
Kostarika	149	-	149
Kanada	-	2 < Me	2 < Me
ZDA	596	941	1537
Brazilija	43	5 < Me	48 = Me
Kolumbija	228	-	228
Ost. Oceanija	7 < Me	-	7 < Me
N	25	20	33
Y_s	115	55	153
Y_s/Y	0.0365	0.0298	0.0306
C_{AD}	. 927	. 940	. 939

Izračunani koeficienti koncentracije ne pokažejo posebnih razlik v koncentraciji izvoza med savinjskim in bačkim hmeljem. Bački hmelj kaže za malenkost večjo koncentracijo kot savinjski.

DVANAJSTO POGLAVJE

12 AGREGATNI INDEKSI

12.1 Z indeksi merimo razlike oziroma spremembe najrazličnejših statističnih parametrov, od absolutnih števil - agregatov, do relativnih števil in srednjih vrednosti. Analitična vrednost indeksov sumarnih parametrov pa je enako vprašljiva kot primerljivost sumarnih relativnih števil ali srednjih vrednosti. Agregati, sumarna relativna števila in sumarne srednje vrednosti so rezultat različnih prvin sestave populacije in velikosti skupinskih parametrov. Indeksi iz takih parametrov kažejo torej hkrati rezultat spremembe oziroma razlike zaradi več vzrokov. Zato je njihov analitični pomen problematičen. Podobno kot s standardiziranimi pokazovalci izločamo vpliv spremembe sestave na sumarno relativno število, poskušamo tudi pri indeksih osamiti vpliv dejavnikov, ki vplivajo na agregat.

12.2 Relativne spremembe agregatov. Izhajajmo iz osnovne zveze med agregati in relativnimi števili

$$Y = X \cdot R = \sum X_k R_k = \sum X \cdot X_k^0 R_k = X \cdot \sum X_k^0 R_k \qquad (12.1)$$

Agregat Y je torej odvisen od absolutne velikosti agregata X, sestave agregata $X_k^0 = X_k / X$ in skupinskih relativnih števil $R_k = Y_k / X_k$. V obrazcih zaradi poenostavitve izpustimo k kot oznako za grupne vrednosti in z indeksom 0 zaznamujemo vrednosti za osnovo oziroma bazo, z 1 pa tekoče vrednosti. Če upoštevamo zveze

12.1, dobimo naslednje zveze indeksov

$$I_Y = \frac{Y_1}{Y_o} = \frac{X_1 R_1}{X_o R_o} = I_X \cdot I_R \tag{12.2}$$

$$I_Y = \frac{Y_1}{Y_o} = \frac{\sum X_1 R_1}{\sum X_o R_o} \tag{12.3}$$

$$I_Y = \frac{Y_1}{Y_o} = \frac{X_1 \sum X_1^o R_1}{X_o \sum X_o^o R_o} = I_X \cdot \frac{\sum X_1^o R_1}{\sum X_o^o R_o} \tag{12.4}$$

Iz obrazca 12.2 in 12.4 sledi, da je

$$I_R = \frac{\sum X_1^o R_1}{\sum X_o^o R_o} \tag{12.5}$$

Iz obrazca 12.3 sledi, da je sprememba agregata Y odvisna od spremembe grupnih vrednosti agregatov X_k in grupnih relativnih števil R_k ali natančneje od spremembe celotnega agregata X, spremembe sestava X_k^o in spremembe grupnih relativnih števil R_k, kot sledi iz obrazca 12.4. Obrazec 12.5 nakaže, da je sprememba sumarnega relativnega števila R odvisna od spremembe sestava X_k^o in spremembe grupnih relativnih števil R_k.

Ker je problematika v vseh treh primerih podobna, podrobneje proučimo zvezo 12.3.

Če v tem obrazcu števec in imenovalec pomnožimo z $\sum X_1 R_o$, dobimo

$$I_Y = \frac{\sum X_1 R_1}{\sum X_o R_o} = \frac{\sum X_1 R_1}{\sum X_1 R_o} \cdot \frac{\sum X_1 R_o}{\sum X_o R_o} = I_{R/X_1} \cdot I_{X/R_o} \tag{12.6}$$

Če ga pomnožimo z $\sum X_o R_1$ pa dobimo:

$$I_Y = \frac{\sum X_1 R_1}{\sum X_o R_o} = \frac{\sum X_1 R_1}{\sum X_o R_1} \cdot \frac{\sum X_o R_1}{\sum X_o R_o} = I_{X/R_1} \cdot I_{R/X_o} \tag{12.7}$$

Prvi faktor

$$I_{R/X_1} = \frac{\sum X_1 R_1}{\sum X_1 R_o} \tag{12.8}$$

v obrazcu 12.6 kaže sumarno spremembo agregata Y zaradi spremembe grupnih relativnih števil R_k pri nespremenjenih tekočih ponderih X_1, drugi

$$I_{X/R_o} = \frac{\sum X_1 R_o}{\sum X_o R_o} \qquad (12.9)$$

pa podobno kaže spremembo agregata Y zaradi spremembe grupnih vrednosti X pri nespremenjenih baznih vrednostih R_o. Podoben pomen imajo indeksi I_{X/R_1} in I_{R/X_o} v obrazcu 12.7.

12.3 Laspeyeresov in Paaschejev agregatni indeksi. Zgornji sumarni indeksi merijo spremembe agregatov Y kot rezultat sprememb sestavin X in R. Zato jih imenujemo agregatne indekse sprememb X in R.

Vsi agregatni indeksi imajo splošno obliko

$$I_{z/w} = \frac{\sum wz_1}{\sum wz_o} \qquad (12.10)$$

V tem splošnem obrazcu za računanje agregatnega indeksa teže ali ponderi w niso vezani na tekoče (1) ali bazne vrednosti (0). Ponderi w morejo biti bodisi tekoče ali bazične vrednosti ali kake druge vrednosti. V posebnem primeru, ko so ponderi bazne količine w_o, govorimo o Laspeyeresovem agregatnem indeksu

$$I_{z/w_o} = L_{z/w} = \frac{\sum w_o z_1}{\sum w_o z_o} \qquad (12.11)$$

v posebnem primeru, če so ponderi tekoče količine w_1, pa o Paaschejevem agregatnem indeksu

$$I_{z/w_1} = P_{z/w} = \frac{\sum w_1 z_1}{\sum w_1 z_o} \qquad (12.12)$$

Če primerjamo prikazane indekse, spoznamo, da je

$$I_{R/X_1} = P_{R/X} = \frac{\sum X_1 R_1}{\sum X_1 R_o} \quad ; \quad I_{X/R_o} = L_{X/R} = \frac{\sum X_1 R_o}{\sum X_o R_o} \tag{12.13}$$

I_{R/X_1} je Paaschejev indeks za R, I_{X/R_o} pa Laspeyeresov indeks za X.

Iz obrazca 12.6 sledi, da je

$$I_Y = I_{R/X_1} \cdot I_{X/R_o} = P_{R/X} \cdot L_{X/R} \tag{12.14}$$

iz obrazca 12.7 pa

$$I_Y = I_{X/R_1} \cdot I_{R/X_o} = L_{R/X} \cdot P_{X/R} \tag{12.15}$$

12.4 Laspeyeresov in Paaschejev agregatni indeksi kot tehtane sredine individualnih indeksov.

Iz identitet

$$P_{R/X} = \frac{\sum X_1 R_1}{\sum X_1 R_o} = \frac{\sum X_1 R_1}{\sum X_1 R_1 / \frac{R_1}{R_o}} = \frac{\sum Y_1}{\sum \frac{Y_1}{I_R}} = H_{I_R}(Y_1)$$

$$P_{X/R} = \frac{\sum X_1 R_1}{\sum X_o R_1} = \frac{\sum X_1 R_1}{\sum X_1 R_1 / \frac{X_1}{X_o}} = \frac{\sum Y_1}{\sum \frac{Y_1}{I_X}} = H_{I_X}(Y_1)$$

$$L_{R/X} = \frac{\sum X_o R_1}{\sum X_o R_o} = \frac{\sum X_o R_o \cdot \frac{R_1}{R_o}}{\sum X_o R_o} = \frac{\sum Y_o I_R}{\sum Y_o} = M_{I_R}(Y_o) \tag{12.16}$$

$$L_{X/R} = \frac{\sum X_1 R_o}{\sum X_o R_o} = \frac{\sum X_o R_o \cdot \frac{X_1}{X_o}}{\sum X_o R_o} = \frac{\sum Y_o I_X}{\sum Y_o} = M_{I_X}(Y_o)$$

spoznamo, da je Paaschejev agregatni indeks tehtana harmonična sredina iz individualnih indeksov, pri čemer so ponderi tekoče vrednosti Y_1. Laspeyeresov indeks pa je tehtana aritmetična sredina z bazičnimi vrednostmi Y_o kot ponderi.

12.5 Če primerjamo Paaschejev obrazec z Laspeyeresovim obrazcem

$$P_{z/w} = \frac{\sum w_1 z_1}{\sum w_1 z_o} \; ; \; L_{z/w} = \frac{\sum w_o z_1}{\sum w_o z_o} \qquad (12.17)$$

spoznamo, da dasta obrazca različne vrednosti, čeprav oba merita spremembo nivoja zaradi sprememb z. Vzrok za razlike je v tem, da so teže pri $P_{z/w}$ tekoče vrednosti w_1, pri $L_{z/w}$ pa bazne vrednosti w_o. Zato dobimo v enem primeru prevelike, pri drugem pa premale vrednosti indeksov, odvisno od vsebinske problematike proučevanih količin. Da bi dobili realno sliko in izravnali raznosmeren učinek obeh sestav, včasih izračunamo iz $P_{z/w}$ in $L_{z/w}$ geometrijsko sredino

$$F_{z/w} = \sqrt{P_{z/w} \cdot L_{z/w}} = \sqrt{\frac{\sum w_1 z_1}{\sum w_1 z_o} \cdot \frac{\sum w_o z_1}{\sum w_o z_o}} \qquad (12.18)$$

Ta obrazec, imenovan "Fisherjev" idealni obrazec, je samo računska operacija, ki temelji na simetriji ponderov in količin iz bazičnega in tekočega razdobja. Ima pa razen izravnavanja učinkov $P_{z/w}$ in $L_{z/w}$ tudi druge pomembne lastnosti, ki jih bomo nakazali v posebnem odstavku.

12.6 Agregatni verižni indeksi. Podobno kot pri enostavnih indeksih, izračunamo tudi verižne agregatne indekse

$$\rho_{z.k} = \frac{\sum w_{k-1} z_k}{\sum w_{k-1} z_{k-1}} \qquad (12.19)$$

po Laspeyeresovem

$$P_{z.k} = \frac{\sum w_k z_k}{\sum w_k z_{k-1}} \qquad (12.20)$$

po Paaschejevem obrazcu.

Po logiki, ki velja za enostavne verižne indekse

$$I_{k/o} = \frac{Y_1}{Y_o} \cdot \frac{Y_2}{Y_1} \cdots \frac{Y_k}{Y_{k-1}} = i_1 \cdot i_2 \cdots i_k \qquad (12.21)$$

naj bi dobili agregatni indeks s stalno bazo 0

$$I_{k/o} = I_1 \cdot I_2 \cdots I_k = \frac{\sum w_o z_1}{\sum w_o z_o} \cdot \frac{\sum w_1 z_2}{\sum w_1 z_1} \cdots \frac{\sum w_{k-1} z_k}{\sum w_{k-1} z_{k-1}} \qquad (12.22)$$

Če vzamemo za osnovo Laspeyersov obrazec in

$$I_{k/o} = p_1 p_2 \cdots p_k = \frac{\sum w_1 z_1}{\sum w_1 z_o} \cdot \frac{\sum w_2 z_2}{\sum w_2 z_1} \cdots \frac{\sum w_k z_k}{\sum w_k z_{k-1}} \qquad (12.23)$$

če vzamemo za osnovo Paaschejev obrazec.

To tehnika računanja indeksov s stalno bazo 0 ima svoje prednosti, a tudi svoje pomanjkljivosti. Predvsem omogoča računanje vrst agregatnih indeksov s stalno osnovo tudi v primeru, če se ponderacijski sistem menja, posebej še, če nekateri elementi agregata v dolgi vrsti izpadejo, pojavijo pa se novi. To je posebno pogosto pri računanju indeksov cen, ko se količine hitro menjajo.

Indeks s stalno osnovo, izračunan prek verižnih indeksov, je odvisen od ponderacije v vsem razdobju med 0 in k.

Preskusi z agregatnimi indeksi

12.7 Preskus o zamenljivosti osnove. Če v enostavnem indeksu zamenjamo osnovo 0 s tekočo vrednostjo 1, dobimo, da je

$$I_{1/0} \cdot I_{0/1} = \frac{z_1}{z_o} \cdot \frac{z_o}{z_1} = 1$$

Od nakazanih agregatnih indeksov ima to lastnost le agregatni indeks, izračunan po

$$F_{1/o}\, F_{o/1} = \sqrt{\frac{\sum X_o R_1}{\sum X_o R_o} \cdot \frac{\sum X_1 R_1}{\sum X_1 R_o}} \cdot \sqrt{\frac{\sum X_o R_o}{\sum X_o R_1} \frac{\sum X_1 R_o}{\sum X_1 R_1}} = 1 \qquad (12.24)$$

Preskus o zamenljivosti sestavin

Ker je $Y = XR$, velja za enostavne indekse

$$\frac{Y_1}{Y_o} = \frac{X_1 \cdot R_1}{X_o \cdot R_o} \; ; \qquad I_Y = I_X \cdot I_R \qquad (12.25)$$

Te lastnosti nima niti Laspeyeresov niti Paaschejev indeks. Velja namreč le:
$I_Y = L_X \cdot P_R = P_X \cdot L_R$, zaradi česar je

$$I_Y \neq L_X \cdot L_R \quad \text{ali} \quad I_Y \neq P_X \cdot P_R \qquad (12.26)$$

Pač pa velja preskus o zamenljivosti sestavin za Fischerjev indeksni obrazec, kjer je

$$F_R \cdot F_X = \sqrt{L_R \cdot P_R} \cdot \sqrt{L_X \cdot P_X} = \sqrt{L_R P_X \cdot P_R L_X} = \sqrt{I_Y \cdot I_Y} = I_Y \qquad (12.27)$$

Med drugim si je Fischerjev idealni obrazec pridobil prilastek "idealni" zaradi teh dveh lastnosti, ki enačijo zveze z agregatnimi indeksi z zvezami med enostavnimi indeksi.

12.8 Skupinski indeksi. Pogosto računamo agregatne indekse prek skupinskih indeksov. Sestava agregatov omogoča tako računanje, ki je zaradi tehničnih, a tudi vsebinskih ozirov, nujno. Pri proučitvi gibanja agregatnih indeksov cen ni pomemben samo skupni agregatni indeks cen, temveč so prav tako, če ne še bolj, pomembni skupinski indeksi za posamezne skupine izdelkov. Ti podrobneje osvetlijo gibanje ravni cen kakor pa sumaren agregatni indeks.

Računanje reprezentativnih indeksov cen pa je edino možno posredno prek skupinskih indeksov.

Računanje skupinskih indeksov in prek njih skupni indeks nakažimo z izpeljavo in primerom.

Če z I_{ki} zaznamujemo indeks i-tega elementa v skupini k (npr. indęks cene za izdelek i iz skupine izdelkov k, z w_{ki} indeksu I_{ki} ustrezni ponder (npr. vrednost proizvodnje), izračunamo skupinski indeks I_k skupine k po obrazcu:

$$I_k = \frac{\sum_i I_{ki} w_{ki}}{\sum_i w_{ki}} = \sum_i I_{ki} w^o_{ki} \qquad (12.28)$$

w^o_{ki} = ponder, izražen kot delež v skupini

Skupni agregatni indeks I pa podobno izračunamo dalje po obrazcu

$$I = \frac{\sum_k I_k w_k}{\sum_k w_k} = \sum_k I_k w^o_k \qquad (12.29)$$

pri čemer je $w_k = \sum_i w_{ki}$ $\quad w^o_k = w_k / \sum_k w_k$

Tabela 12.1. Shematično izračunanje skupinskih indeksov

k_i	w_{ki}	w^o_{ki}	I_{ki}	$I_{ki} \cdot w^o_{ki}$	I_k	w^o_k	$I_k w^o_k$
11	10	·100	150	15,00			
12	30	·300	167	50,10			
13	60	·600	100	60,00			
1	100	1.000		125,1 = I_1		·357	44,66
21	20	·333	90	29,97			
22	40	·667	120	80.04			
2	60	1.000		110,0 = I_2		·214	23,54
31	60	·500	150	75,00			
32	60	·500	140	70,00			
3	120	1.000		145,0 = I_3		·429	62,20
						1.000	130,4 = I

Izračunanje skupnega indeksa prek skupinskih indeksov pa moremo izvršiti tudi prek več sistemov.

INDEKSI CEN

Poprečni indeks cen

12.9 Problem o primerjanju cen je precej enostaven, če proučujemo spremembe v cenah za posamezen artikel. Vendar enostavni indeksi v praksi ne zadovoljijo vseh potreb po proučevanju sprememb pri ekonomskih pojavih. Merjenje in proučevanje sprememb cen za posamezne artikle se izkaže za zelo neprikladno, ker je njih število veliko in je proučevanje sprememb posamičnih cen zaradi velikega števila razdrobljenih podatkov nepregledno in neučinkovito. V praktičnih primerih je zadosti, če poznamo, kako se spreminja raven cen, ne pa, kakšne so spremembe v ravni cen za skupino artiklov s poprečnim indeksom, ki ga izračunamo iz poprečnih indeksov cen za vsak artikel. Povprečni indeks iz poprečnih indeksov moremo izraziti z obrazcem

$$\bar{i} = \frac{1}{N} \sum \frac{p_1}{p_o} \qquad (12.30)$$

Pri tem pomeni:

p_o = cena v bazičnem trenutku, p_1 = cena v tekočem trenutku,
N = število artiklov, $\bar{i}$ = poprečni indeks.

Poprečni indeks, izračunan po obrazcu 12.30, pa je razmeroma slabo merilo sprememb. Njegova hiba je v tem, da v njem ne upoštevamo važnosti posameznega artikla. Važnost za posamezen artikel je dana s proizvodnjo, če gre za indeks cen pri proizvajalcu, ali s prodajo, če gre za cene v trgovini na malo. Iz vsakdanje prakse vemo, da so spremembe v ceni za artikle, ki jih množično uporabljamo (npr. moka, kruh, mleko, meso itd.), za potrošnika veliko pomembnejše, kot spremembe za cene artiklov, za katere je poraba majhna (npr. televizijski aparati, avtomobili, različni specialni električni aparati itd.). Zato poprečne indekse po obrazcu 12.30 razmeroma redko računamo, pač pa namesto njih uporabljamo agregatne indekse, ki upoštevajo različne teže za posamezne artikle.

Izbira ponderov pri računanju agregatnega indeksa cen

12.10 Stalne količine q v obrazcu za računanje agregatnega indeksa cen morejo biti teoretično katere koli količine. Vsebinsko pa je jasno, da indeks cen tem bolje prikaže resnične odnose, čim bolj količine, ki jih vzamemo za pondere, ustrezajo stvarnosti. Zato po možnosti vzamemo za računanje agregatnega indeksa cen količine, ki so čim bolj podobne stvarni porabi ali proizvodnji v preučevanem razdobju. Pri tem pa imamo različne možnosti.

12.11 Laspeyresov obrazec. Če zaznamujemo s q_o količine za bazični čas, s p_o in p_1 pa času ustrezno ceno, je Laspeyresov obrazec za računanje agregatnega indeksa cen

$$L_p = \frac{\sum p_1 q_o}{\sum p_o q_o} \qquad (12.31)$$

Ta obrazec ima prednost, da so ponderi za katero koli razdobje isti. Ta prednost, ki je nekatere druge metode nimajo, daje Laspeyresovi metodi pri praktičnem računanju agregatnih indeksov posebno ceno.

Če uporabimo za računanje indeksa cen Laspeyresov obrazec, da indeks pristrane rezultate. Ta pristranost izvira iz tega, ker sta cena in prodana ali porabljena količina med seboj odvisni. Če vzamemo, da je cena nekega artikla v tekočem razdobju večja kakor v bazičnem $p_1 > p_o$, ima ta podražitev za rezultat zmanjšanje porabe tega artikla: $q_1 < q_o$. Obratno znižana cena $p_1 < p_o$ navadno poveča porabo $q_1 > q_o$. Ker v Laspeyresovem obrazcu vzamemo za pondere količine iz bazičnega razdobja, so ponderi pri artiklih, ki so se podražili, preveliki, pri artiklih, ki so se pocenili, pa premajhni. Skupen učinek teh razlik je sistematična pristranost navzgor, ker so ponderi za artikle, ki se podražijo, preveliki, ponderi za artikle, ki se pocenijo, pa premajhni.

12.12 Vzemimo kot zgled za računanje agregatnega indeksa cen po Laspeyresovem obrazcu indeks cen mesa v razdobju 1964-1957 v SRS. V obračun vzemimo artikle, ki jih v skupini "surovo in prekajeno meso" vsebuje splošni indeks cen na drob-

no za Slovenijo. Izključili smo le kokoš, ker je zanjo ponder minimalen. Podatki so vzeti iz censke in trgovinske statistike ZS SR Slovenije.

Tabela 12.2. Izračunanje agregatnega indeksa cen za surovo in prekajeno meso v splošnem indeksu cen na drobno v SR Sloveniji po Laspeyresovem obrazcu (cene v dinarjih, količine v tisočih tonah)

Artikel	q_p	1954		1955		1956		1957	
		p_0	p_0q_0	p_1	p_1q_0	p_2	p_2q_0	p_3	p_3q_0
Surovo meso s kostmi:									
- goveje	11,6	223	2586,8	250	2900,0	243	2818,8	251	2911,6
- telečje	2,9	271	785,9	304	881,6	296	858,4	305	884,5
- svinjsko	3,7	249	921,3	292	1080,4	291	1076,7	351	1298,7
- ovčje	0,2	166	33,2	192	38,4	180	36,0	202	40,4
Prekajeno svinjsko meso	2,1	320	672,0	364	764,4	375	787,5	402	844,2
Prekajena slanina	1,2	323	387,6	351	421,2	374	448,8	396	475,2
			5386,8		6086,0		6026,2		6454,6
			$=\sum p_0q_0$		$=\sum p_1q_0$		$=\sum p_2q_0$		$=\sum p_3q_0$
	Indeks		100,0		$L_{1/0}=\frac{6086,0}{5386,8}=$		$L_{2/0}=\frac{6026,2}{5386,8}$		$L_{3/0}=\frac{6454,6}{5386,8}$
					$=113,0$		$=111,9$		$=119,8$

12.13 Paaschejev obrazec. Po Paascheievi metodi pa vzamemo za pondere pri indeksu cen tekoče količine q_1. Po tem osnutku računamo indeks cen po obrazcu

$$P_P = \frac{\sum p_1q_1}{\sum p_0q_1} \qquad (12.32)$$

Če pogledamo obrazec 12.32, vidimo, da je treba po njem od primera do primera računati tako števce kot imenovalce indeksa, medtem ko pri Laspeyresovi metodi imenovalec

$\sum p_o q_o$ ostane pri vseh indeksih isti. Razen tega moramo pri Paaschejevi metodi poznati pondere - količine za vsa tekoča razdobja, za katera računamo indekse. S praktičnega vidika je torej računanje po Paaschejevi metodi bolj zamotano kakor po Laspeyresovi. Prednost Paaschejeve metode pa je v tem, da je indeks bližji stvarnim razmeram kot indeks, izračunan po Laspeyresovi metodi. Vendar imamo pri indeksih cen po Paaschejevem obrazcu iz istega razloga kot pri Laspeyresovem obrazcu pristranost z obratnim učinkom, ker stoji umetno sestavljena količina ($\sum p_o q_1$) v imenovalcu indeksa. Zato pričakujemo, da dá Paaschejeva metoda premajhne rezultate.

12.14 Če izračunamo za skupino surovo in prekajeno meso indeks cen še po Paaschejevi metodi, dobimo rezultate, ki so dani v tabeli 12.3.

Primerjava indeksov cen, ki smo jih dobili po Laspeyresu, z ustreznimi indeksi cen, ki smo jih dobili po Paascheju, kaže, da so resnično indeksi cen, izračunani z Laspeyresovim obrazcem višji kot indeksi cen, izračunani s Paaschejevim obrazcem. To pa je skladno z učinkom, ki izvira iz odvisnosti potrošene količine od cene.

12.15 Fischerjev idealni obrazec. Da omilimo pri indeksih cen učinke, ki izvirajo iz tega, da pri Laspeyresovi formuli vzamemo za pondere $q = q_o$ bazične, pri Paaschejevi pa $q = q_1$ tekoče količine, vzamemo včasih za pondere poprečje iz tekočih in bazičnih količin $q = 1/2\ (q_o + q_1)$, včasih pa poprečne količine v celotnem razdobju, za katerega proučujemo gibanje cen.

Zaradi kakovosti, ki smo jih že nakazali, pa računamo Fischerjev idealni indeks

$$F_p = \sqrt{L_p \cdot P_p} = \sqrt{\frac{\sum p_1 q_o}{\sum p_o q_o} \cdot \frac{\sum p_1 q_1}{\sum p_o q_1}} \qquad (12.33)$$

Indeks, izračunan s Fischerjevim idealnim obrazcem, nima istega logičnega smisla kot Laspeyresov ali Paaschejev, je pa nekak kompromis med obema. V tabeli 12.4 so prikazane indeksne vrste za primer skupinskega indeksa za surovo in prekajeno meso, izračunane po vseh treh obrazcih: Laspeyresovem, Paaschejevem in Fischerjevem. Iz tabele vidimo, da je učinek ponderacije tem večji, čim večje je razdobje med bazičnim in tekočim razdobjem.

Tabela 12.3. Izračun agregatnega indeksa cen za surovo in prekajeno meso v letih 1954–1957 po Paaschejevi metodi (cene v dinarjih, količine v tisoč tonah)

Artikel	1954	1955				1956				1957			
	p_0	p_1	q_1	p_0q_1	p_0q_1	p_2	q_2	p_0q_2	p_2q_2	p_3	q_3	p_0q_3	p_3q_3
Surovo meso s kostmi:													
- goveje	223	250	12,2	2720,6	3050,0	243	13,4	2988,2	3256,2	251	14,5	3233,5	3639,5
- telečje	271	304	3,8	1029,8	1155,2	293	5,2	1409,2	1539,2	305	5,8	1571,8	1769,0
- svinjsko	249	292	4,0	996,0	1168,0	291	3,7	921,3	1076,7	351	3,7	921,3	1298,7
- ovčje	166	192	0,2	33,2	38,4	180	0,2	33,2	36,0	202	0,2	33,2	40,4
Prekajeno svinjsko meso	320	364	2,6	832,0	946,4	375	2,5	800,0	937,5	402	2,6	832,0	1045,2
Prekajena slanina	323	351	1,5	484,5	526,5	374	1,5	484,5	561,0	396	1,5	484,5	594,0
				6096,1	6884,5			6636,4	7406,6			7076,3	8386,8
				$\sum p_0q_1$	$\sum p_1q_1$			$\sum p_0q_2$	$\sum p_2q_2$			$\sum p_0q_3$	$\sum p_3q_3$
		$P_1 = 100 \frac{6884,5}{6096,1} = 112,9$				$P_2 = 100 \frac{7406,6}{6633,4} = 111,6$				$P_3 = 100 \frac{8386,8}{7076,3} = 118,5$			

12.16 V tabeli 12.4 so prikazane indeksne vrste za skupinski indeks cen za surovo in prekajeno meso v razdobju 1954-1957 v SR Sloveniji.

Iz tabele 12.4 vidimo, da je Laspwyersov indeks sistematično nad Paaschejevim, kar je v skladu s pričakovanjem. Indeksna vrsta, ki smo jo dobili z metodo verižnih indeksov, pa se najbolj prilagaja Paaschejevim indeksom. V splošnem razlike niso velike, in sicer zato, ker je štiriletno razdobje razmeroma kratko in strukturne razlike v porabi niso velike.

Tabela 12.4. Primerjalna tabela indeksnih vrst za skupinski indeks cen za surovo in prekajeno meso, izračunane po različnih postopkih

Vrsta indeksa	1954	1955	1956	1957
Laspeyresov indeks	100,0	113,0	111,9	119,8
Paaschejev indeks	100,0	112,9	111,6	118,5
Fischerjev idealni indeks	100,0	112,9	111,8	119,1
Verižni indeks	100,0	113,0	111,5	118,7

Reprezentativni indeksi cen

12.17 Agregatni indeks cen, v katerega vključimo vse elemente določenega pojava, je na primer agregatni indeks cen pri proizvajalcu za določeno podjetje, ki ne izdeluje veliko različnih artiklov.

V splošnem pa je takih primerov malo. Največ pojavov, ki jih obravnavamo z agregatnimi indeksi, sestoji iz toliko posamičnih elementov, da jih je tehnično nemogoče v celoti zajeti. V splošni indeks cen na drobno ne moremo vključiti vseh artiklov, ki so v prodaji na drobno, ker je število teh artiklov veliko. Dinamiko cen za take primere ne opazujemo z agregatnim indeksom, ki je sestavljen iz vseh artiklov v prodaji na drobno, ampak z reprezentativnim indeksom cen v trgovini na drobno. Ta pa je sestavljen iz podatkov o cenah za nekatere artikle. Izkaže se namreč, da moremo s takim reprezentativnim indeksom zadovoljivo spremljati dinamiko cen za cele skupine predmetov. Gibanje cen za skupine sorodnih artiklov kaže namreč isti razvoj. Sprememba cene

artiklov v eni skupini ima običajno nek skupen vzrok. Sprememba v ceni surovine ali v osebnih dohodkih delavcev v določeni stroki povzroči spremembo v ceni za vse artikle, ki jih proizvajajo v tej stroki itd.

Pri reprezentativnem indeksu cen vzamemo iz vsake grupe samo določeno število predmetov, ki so reprezentanti te skupine. Gibanje cen za te predmete imamo značilno za gibanje cen vseh artiklov v grupi.

Da je agregatni indeks, v katerega ne vključimo vseh elementov določenega pojava, zares reprezentativen indeks za celo grupo, moramo paziti na naslednje:

a) Če hočemo sestaviti reprezentativni agregatni indeks cen za raznovrstno populacijo predmetov, predhodno združimo predmete v več grup sorodnih predmetov.

b) Iz vsakega homogenega dela izberemo nekaj predmetov. Gibanje cen za te izbrane reprezentante skupine imamo za gibanje ravni cen za vse artikle skupine. Za reprezentante skupine običajno izberemo take predmete, katerih proizvodnja ali poraba pomeni velik del proizvodnje ali porabe za skupino. Včasih izberemo tudi najznačilnejše predmete, ki so po svoji naravi taki, da njihove cene izražajo gibanje cen v skupini. Če izberemo iz posamezne skupine veliko artiklov, jih običajno izberemo po metodi slučajnega izbora, da indeks zadosti predpisom vzorčenja.

c) Iz podatkov o cenah izbranih artiklov, ki so reprezentanti skupine, izračunamo skupinski indeks; ta more biti bodisi agregatni indeks ali navadni povprečni indeks. Če vzamemo za osnovo pri računanju indeksa grupe agregatni indeks, vzamemo za pondere po možnosti količine prometa ali proizvodnje za vse artikle grupe, ki jo izbrani artikel reprezentira in ne samo promet ali proizvodnjo za ta artikel.

d) Iz skupinskih indeksov sestavimo skupni indeks kot tehtani agregatni indeks. Pri tem pa vzamemo za ponder za posamezno grupo vrednost cele grupe in ne samo vrednost artiklov, ki so reprezentanti grupe.

Po teh načelih so na splošno sestavljeni vsi agregatni indeksi cen.

12.18 Indeks življenjskih stroškov. Posebna oblika indeksa cen je indeks življenjskih stroškov. Ta indeks ni splošni indeks cen v trgovini na drobno, mar-

več je po svojem sestavu in namenu posebna oblika indeksa cen na drobno. Indeks življenjskih stroškov ima namen spremljati gibanje življenjskih stroškov v posameznih skupinah prebivalstva. Nemogoče je spremljati življenjske stroške vseh družin ali gospodinjstev za določeno socialno skupino. Statistično proučujemo probleme iz tega področja tako, da anketiramo določene družine in zanje proučujemo spremembe v količini in sestavi porabe. Take ankete dajejo stvarno sliko o porabi in stvarne življenjske stroške. Drug instrument za proučevanje standarda pa je indeks življenjskih stroškov. Pri indeksu življenjskih stroškov sestavimo skupnost predmetov široke porabe, v količini, za katero mislimo, da je za določen tip gospodinjstev pri določenem standardu potrebna. To skupnost predmetov imenujemo p o t r o š n i k o š . Ta tako imenovana teoretska lista je sestavljena skladno s stvarnimi potrebami. Čeprav je teoretična, vendar vsebuje vse osnovne artikle, ki so potrebni in običajni za življenje na določeni stopnji standarda.

Teoretska lista je osnova za računanje indeksa življenjskih stroškov. Od meseca do meseca ali od leta do leta računamo na osnovi stalne teoretske liste, kako se menjajo življenjski stroški, če ostane standard oziroma potrošni koš stalen. Če izračunamo, kolikšni so stroški za preživljanje, ako hočemo zadostiti porabi po teoretski listi, dobimo absoluten iznos: ta je fiktiven in ga ne uporabljamo za analizo. Indeksi, ki jih izračunamo iz teh absolutnih - fiktivnih stroškov, pa zelo dobro kažejo, kako se giblje raven stvarnih življenjskih stroškov, če je potrošni koš po svoji sestavi in izbiri artiklov primeren.

Agregatni indeksi količin - obsega

12.19 Doslej smo podrobneje obravnavali le agregatne indekse cen. Vendar enako - le da pomen cen in količin zamenjamo - izračunavamo tudi agregatne indekse količin, tako da spreminjamo količine in vzamemo cene za stalne. Tako dobimo, podobno kakor za indeks cen, indeks količin po Laspeyresovem obrazcu, če vzamemo v agregatnem indeksu

$$L_q = \frac{\sum p_o q_1}{\sum p_o q_o} \qquad (12.34)$$

za pondere cene v osnovnem razdobju in Paaschejev obrazec za računanje indeksa ko-

ličin, če v obrazcu vzamemo

$$P_q = \frac{\sum p_1 q_1}{\sum p_1 q_0} \qquad (12.35)$$

za pondere cene iz tekočega razdobja.

Enako je geometrijska sredina med Laspeyresovim in Paaschejevim obrazcem Fischerjev idealni obrazec za računanje agregatnega indeksa količin

$$F_q = \sqrt{L_q \cdot P_q} = \sqrt{\frac{\sum p_o q_1}{\sum p_o q_o} \cdot \frac{\sum p_1 q_1}{\sum p_1 q_o}} \qquad (12.36)$$

Z agregatnimi indeksi za količine v praksi računamo fizični obseg proizvodnje v industrijskih podjetjih in računamo vrednost proizvodnje po stalnih cenah. Kot so pri računanju indeksa cen sčasoma stalne količine, ki jih vzamemo za pondere, problematične, so pri indeksih obsega proizvodnje za daljša razdobja problematični ponderi tudi stare, stalne cene. V bistvu se pri računanju indeksov količin borimo z enakimi težavami kot pri agregatnih indeksih cen.

12.20 Če izračunamo za zgled indeks obsega za porabo mesa v Sloveniji v tabelah 12.2 in 12.3, je indeks obsega porabe v letu 1957, če vzamemo za osnovo leto 1954

po Laspeyresovem obrazcu $L_q = \frac{\sum p_o q_3}{\sum p_o q_o} = \frac{7076,3}{5386,8} = 131,3$

po Paaschejevem obrazcu $P_q = \frac{\sum p_3 q_3}{\sum p_3 q_o} = \frac{8386,8}{6454,6} = 129,9$

po Fischerjevem obrazcu $F_q = \sqrt{L_q \cdot P_q} = \sqrt{131,3 \cdot 129,9} = 131,1$

12.21 Ko računamo indeks obsega proizvodnje ali obseg prometa v trgovini na drobno ali na debelo, naletimo dostikrat na hude težave. Računanje reprezentativnih indeksov obsega nima istega logičnega ozadja kakor reprezentativni indeksi cen. Zato smo navezani le na kompletno računanje in preračunavanje vrednosti po stalnih cenah.

To pa je sila težavno, če agregat sestavlja veliko število različnih predmetov. Dostikrat si pomagamo pri ocenjevanju vrednosti proizvodnje, prometa ali porabe po enotnih - stalnih cenah z enostavnim sklepom. Če je stvarna vrednost proizvodnje, porabe ali prometa v tekočem razdobju $V_1 = \sum p_1 q_1$, moremo vrednost proizvodnje reducirati na cene v bazičnem razdobju $V_1' = \sum p_o q_1$, če jo delimo z ustreznim reprezentativnim indeksom cen

$$V_1' = \sum p_o q_1 \approx \frac{V_1}{P_p'} \quad ; \quad P_p' = \frac{\sum p_1' q_1'}{\sum p_o' q_1'} \tag{12.37}$$

Pri cenah in količinah so napravljene črtice, ker so to cene in količine za reprezentativne artikle te skupine, ne pa za vse artikle skupine, kakor je to v izrazih $\sum p_1 q_1$ in $\sum p_o q_1$.

Če je dana vrednost po tekočih cenah, moremo torej oceniti vrednost po bazičnih cenah, če stvarno vrednost $V_1 = \sum p_1 q_1$ delimo z ustreznim indeksom cen P_p', vrednost prometa v trgovini na drobno z indeksom cen v trgovini na drobno, promet v trgovini na debelo z indeksom cen v trgovini na debelo itd. Če ocenjujemo vrednost samo za določeno skupino, reduciramo vrednost z ustreznim grupnim indeksom cen itd.

Indeksi recipročnih pokazovalcev

12.22 Za pokazovalce, za katere imajo smisel tudi recipročni pokazovalci, je treba paziti, da so agregatni indeksi računani tako, da kažejo pravilno smer sprememb proučevanega pojava. Tako npr. produktivnost dela merimo s količino na enoto časa $q = \frac{Q}{T}$ ali s časom, potrebnim za izdelavo enote izdelka $t = \frac{T}{Q}$. Ker je pokazovalec $q = \frac{Q}{T}$ premosorazmeren s produktivnostjo dela, ker je proizvedena količina na enoto časa večja pri večji produktivnosti, računamo agregatni indeks za ta primer po obrazcu

$$L_{q/T} = \frac{\sum T_o q_1}{\sum T_o q_o} \quad \text{ali} \quad P_{q/T} = \frac{\sum T_1 q_1}{\sum T_1 q_o} \tag{12.38}$$

alternativno

$$L_{v/T} = \frac{\sum T_o v_1}{\sum T_o v_o} \quad \text{ali} \quad P_{v/T} = \frac{\sum T_1 v_1}{\sum T_1 v_o} \tag{12.39}$$

če gre za različne artikle in vzamemo vrednosti namesto količin.

Pokazovalec $t = \frac{T}{Q}$ pa je pri večji produktivnosti dela manjši, ker je čas za izdelavo enote pri večji proizvodnosti dela manjši. Zato računamo agregatne indekse produktivnosti po obrazcih

$$L_{t/Q} = \frac{\sum Q_o t_o}{\sum Q_o t_1} \quad ; \quad P_{t/Q} = \frac{\sum Q_1 t_o}{\sum Q_1 t_1} \tag{12,40}$$

da so indeksi premosorazmerni s proizvodnostjo dela. Podobno je z vsemi drugimi podobnimi pokazovalci.

$L_{q/T}$, $P_{q/T}$, $L_{v/T}$ in $P_{v/T}$ so indeksi produktivnosti dela, izračunani pri stalni sestavi časa, $L_{t/Q}$ in $P_{t/Q}$ pa indeksa produktivnosti dela pri stalni sestavi količin.

12.23 Nekaj socialno-ekonomskih količin, za katere moremo analizirati spremembe z agregatnimi indeksi.

Splošni odnosi	sredine	$Y = N.M_y$	agregat Y = število enot N x poprečje M
	deleži	$N_a = N.P_a$	število enot N_a = število enot N x delež P
	koeficienti	$Y = X.R$	agregat Y = agregat X x razmerje R
Posebni primeri		$V = Q.p$	vrednost V = količina Q x cena p
cene, obseg		$Q = V . q$	količina Q = vrednost V x kupna moč q
Proizvodnost dela		$V = T . v$	vrednost V = čas T x proizvodnost dela v
		$T = Q . t$	čas T = količina Q x proizvodnost dela t
Obračanje zalog		$P = Z . p$	promet na enoto časa P = zaloga Z x število obratov p
		$Z = P . z$	zaloga Z = promet na enoto časa P x dolžina obrata z
Pokazovalci na prebivalca		$P = Pr . p$	poraba na enoto časa P = prebivalci Pr x poraba na enega prebivalca p
		$Pr = P . pr$	št. prebivalcev Pr = poraba P x prebivalcev na enoto P pr

Pokazovalci	$D = P \cdot d$	izmet D = proizvodnja P x delež izmeta d
na proizvodnjo	$S = P \cdot s$	poraba surovin S = proizvodnja P x poraba surovin na enoto proizvodnje s
	$G = P \cdot g$	poraba goriva G = proizvodnja P x gorivo na enoto proizvodnje g
Demografski pokazovalci	$R = P_r\, r$	število rojenih R = prebivalstvo P_r x rojeni na prebivalca r
	$U = P_r\, u$	število umrlih U = prebivalstvo P_r x umrli na prebivalca u

12.24 V dosedanjih obrazcih smo obravnavali indekse s stališča proučevanja dinamike pojavov. Pri tem smo vzeli oznako O za bazično razdobje, 1 pa za tekoče razdobje. Vendar moremo vse, kar smo povedali o agregatnih indeksih, brez težav uporabiti tudi za primerjave, ki niso časovne. Tako moremo z agregatnim indeksom o produktivnosti dela primerjati produktivnost dela med podjetji iste stroke. Pri tem z O označimo podatke za prvo, z 1 pa podatke za drugo podjetje, ki ju primerjamo. Enako moremo agregatne indekse uporabiti pri računanju indeksov o izpolnjenju plana. Pri tem z 1 zaznamujemo stvarne, z O pa planske podatke. Prav tako računamo indekse prihrankov v materialu. Za take primere v indeksu pomenijo q_o normirane količine porabljenih surovin, q_1 pa stvarne količine porabljenih surovin.

Seveda v analizah te vrste odpadejo določene metode, ki so umestne le pri časovnih agregatnih indeksih. Ena izmed njih je npr. metoda verižnih indeksov. V glavnem pa celotno teorijo indeksov, ki smo jo nakazali v tem poglavju, uporabljamo v različnih primerih in ni omejena le na proučevanje dinamike pojavov.

TRINAJSTO POGLAVJE

NORMALNA PORAZDELITEV

Pomen proučevanja teoretičnih porazdelitev

13.1 V dosedanjih poglavjih smo proučevali predvsem stvarne porazdelitve, ki so slika opazovanih populacij.

Vendar smo na nekaj mestih nakazali, da imamo razen stvarnih frekvenčnih porazdelitev tudi frekvenčne porazdelitve, ki jih dobimo s teoretičnim razglabljanjem. Temeljna teoretična porazdelitev je normalna porazdelitev. Zanjo smo že nakazali nekaj lastnosti v zvezi s standardnim odklonom. Razen normalne porazdelitve imamo tudi druge teoretične porazdelitve. Od teh bomo obravnavali: binomsko, Poissonovo, hipergeometrično, t-porazdelitev, χ^2-porazdelitev in F-porazdelitev. Do teoretičnih porazdelitev pridemo s proučevanjem slučajnostnih spremenljivk, ki zadoščajo določenim osnovnim pogojem.

Proučevanje teoretičnih porazdelitev je pomembno z več vidikov:

a) Če stvarna populacija zadošča istim pogojem kot teoretična porazdelitev, je stvarna frekvenčna porazdelitev podobna, če že ne enaka teoretični. Če primerjamo stvarno frekvenčno porazdelitev s teoretično, ki jo dobimo v določenih pogojih, moremo sklepati, ali stvarna porazdelitev zadošča tem pogojem ali ne. Če na primer na določen zvezen znak vplivajo samo slučajni vplivi, variacija pa v nobeno smer ni okrnjena, se vrednosti te populacije porazdeljujejo v normalni porazdelitvi. Če je stvarna frekvenčna porazdelitev za določen rezultativen znak skladna z ustrezno teoretično - normalno porazdelitvijo, sklepamo, da so na populacijo vplivali samo slučajni vplivi brez omejitev v variiranju.

b) Včasih ne poznamo stvarne frekvenčne porazdelitve. Za populacijo pa poznamo določene parametre. Če predpostavljamo, da vrednosti populacije zadoščajo danim pogojem, moremo teoretično ugotoviti, kakšna je približno stvarna porazdelitev. Če npr. za proizvodnjo določenih profilov predpostavimo, da so odkloni od poprečne proizvodnje za posamezne izdelke slučajni, poznamo pa aritmetično sredino in standardni odklon za proizvodnjo profilov, moremo določiti, kako se predvidoma porazdeljujejo premeri profilov, ne da bi poznali individualne vrednosti.

c) Teoretične frekvenčne porazdelitve pa kot verjetnostne porazdelitve nastopajo v vzorčenju, ki je, kakor smo že omenili, najpomembnejša metoda s katero ocenjujemo in sklepamo v statistiki. Če vzorec zadošča osnovnim pogojem, se posamezne slučajnostne spremenljivke, ki nastopajo v vzorčenju, porazdeljujejo v določenih teoretičnih porazdelitvah. Razen za nepomembno majhne populacije, teh ne moremo neposredno sestaviti. Z verjetnostnim računom pa moremo te porazdelitve razviti teoretično. Tako npr. s teoretičnim študijem odkrijemo zakon o velikih številih, ki smo ga uvodoma nakazali in ilustrirali s praktičnim primerom.

Teorija vzorčenja je zgrajena ob predpostavki, da so enote, ki jih izbiramo v vzorec, izbrane slučajnostno. To pomeni, da mora imeti vsaka enota populacije enako možnost, da jo izberemo v vzorec. Ob tej predpostavki je z uporabo verjetnostnega računa izdelana teoretična osnova vzorčenja. Razumljivo je, da teorija velja le, če je ta osnovni pogoj izpolnjen. Zato se v vzorčenju na vso moč trudimo, da zadostimo temu osnovnemu pogoju.

Teoretične porazdelitve proučujemo torej s treh vidikov. Z njimi ugotavljamo, ali stvarne porazdelitve zadoščajo določenim predpostavkam. Če poznamo zakonitost variiranja v populaciji, sklepamo, kakšna je stvarna porazdelitev. Teoretične porazdelitve pa so tudi osnova za vzorčenje, ki je pomembna metoda ocenjevanja in sklepanja.

NORMALNA PORAZDELITEV

13.2 Pri proučevanjih množičnih pojavov imamo veliko stvarnih populacij, ki vsaj v grobem zadoščajo pogojem, v katerih se določen znak porazdeljuje normalno. Normalna porazdelitev je osnovna verjetnostna porazdelitev v vzorčenju. Mnogo slučajnost-

nih spremenljivk, ki jih sestavljamo v vzorčenju, se porazdeljuje v normalni porazdelitvi, če ne neposredno pa vsaj asimptotično. Normalna porazdelitev je tudi osnova za druge teoretične porazdelitve, kot so: t-porazdelitev, χ^2-porazdelitev, F-porazdelitev. V normalno porazdelitev preide asimptotično večina teoretičnih porazdelitev, med njimi tudi: binomska, Poissonova, hipergeometrična, t-porazdelitev, χ^2-porazdelitev, F-porazdelitev.

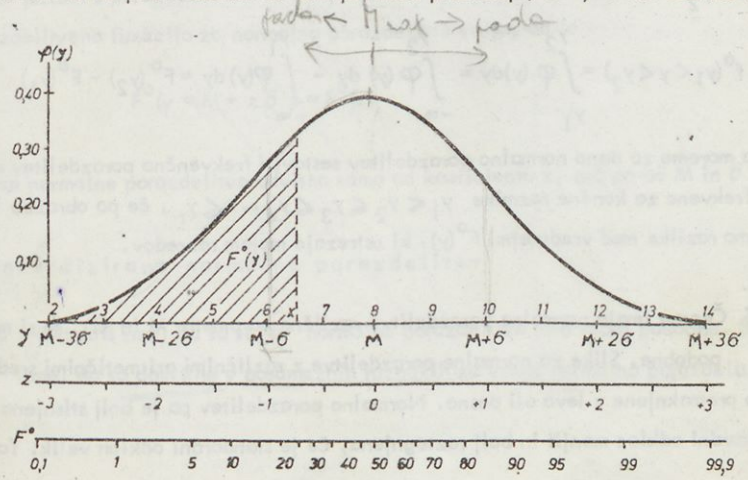
O p i s

13.3 Normalna porazdelitev je določena z dvema parametroma: z aritmetično sredino M, ki je rezultat splošnih vplivov in s standardnim odklonom σ, ki je rezultat posamičnih oziroma slučajnih vpiivov. Gostota relativne frekvence za normalno porazdelitev je dana s funkcijo

$$\varphi(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-M)^2}{2\sigma^2}} \qquad (13.1)$$

Ta enačba pokaže, v kakšni odvisnosti je relativna gostota frekvence $\varphi(y)$ z vrednostjo y.

Če vzamemo vrednost znaka y za absciso, ustrezno vrednost relativne gostote frekvence $\varphi(y)$ pa za ordinato, dobimo sliko normalne porazdelitve (glej sliko 13.1). Iz enačbe (13.1) in slike spoznamo, da je normalna porazdelitev unimodalna, simetrična in zvonasta, zaradi česar je zanjo M = Me = Mo. Največja gostota relativne frekvence je za y = M, gostota pa vztrajno pada, čim bolj je y oddaljen od aritmetične sredine. Normalna krivulja se abscisni osi asimptotično približuje, če y raste čez vse meje.



Slika 13.1. Normalna porazdelitev za M = 8 in SD = 2

13.4 Določen integral

$$F^{o}(y) = \int_{-\infty}^{y} \varphi(y)\,dy \qquad (13.2)$$

je funkcija zgornje meje in ga imenujemo na splošno porazdelitvena funkcija. Porazdelitvena funkcija $F^{o}(y)$ kaže relativno frekvenco v razmiku $-\infty$ do y. $F^{o}(y)$ je funkcija zgornje meje v razmiku od $-\infty$ do y in pomeni grafično ploščino, ki je v sliki 13.1 omejena z delom krivulje nad razmikom od $-\infty$ do y, z ustreznim delom razmika na abscisni osi in vrednosti y ustrezno ordinato za normalno krivuljo. Če narišemo porazdelitveno funkcijo za normalno porazdelitev v odvisnosti od y, dobimo značilno S krivuljo, ki smo jo srečali že pri stvarnih frekvenčnih porazdelitvah. S-krivuljo za normalno porazdelitev vzamemo za idealno.

V sliki 13.2 so narisane tri normalne porazdelitve, pod njo pa ustrezne porazdelitvene funkcije.

Iz opredelitve relativne frekvence, ki jo poznamo že iz poglavja o frekvenčnih porazdelitvah, sklepamo dalje, da je

$$F^{o}_{(\infty)} = \int_{-\infty}^{+\infty} \varphi(y)\,dy = 1 \qquad (13.3)$$

Če poznamo funkcijo $F^{o}(y)$, dobimo relativno frekvenco v poljubnem končnem razmiku y_1 do y_2 enostavno

$$f^{o}(y_1 < y < y_2) = \int_{y_1}^{y_2} \varphi(y)\,dy = \int_{-\infty}^{y_2} \varphi(y)\,dy - \int_{-\infty}^{y_1} \varphi(y)\,dy = F^{o}(y_2) - F^{o}(y_1) \qquad (13.4)$$

Tako moremo za dano normalno porazdelitev sestaviti frekvenčno porazdelitev relativnih frekvenc za končne razmike $y_1 \leq y_2 \leq y_3 \leq y_4 \cdots \leq y_r$, če po obrazcu 13.4 poiščemo razlike med vrednostmi $F^{o}(y)$, ki ustrezajo mejam razredov.

13.5 Čeprav imajo normalne porazdelitve različne parametre M in SD, so si med seboj podobne. Slike za normalne porazdelitve z različnimi aritmetičnimi sredinami so samo premaknjene v levo ali desno. Normalna porazdelitev pa je bolj stisnjena, če je standardni odklon manjši in bolj raztegnjena, če je standardni odklon velik. To je nazor-

no razvidno iz slike 13.2, v kateri imata porazdelitvi A in B enaki aritmetični sredini, a različna standardna odklona, porazdelitvi A in C različni aritmetični sredini in enaka standardna odklona, porazdelitvi B in C pa različni aritmetični sredini in različna standardna odklona. Pod njima so prikazane ustrezne porazdelitvene funkcije. Iz slike 13.2 vidimo, da sta S-krivulji za porazdelitvi, ki imata enak standardni odklon, a različni aritmetični sredini, enake oblike, le da sta premaknjeni vzdolž abscisne osi. S-krivulja za normalno porazdelitev, za katero je standardni odklon manjši, pa je strmejša.

Za vsak poseben primer normalne porazdelitve bi bilo zelo neprikladno neposredno računati gostoto relativne frekvence $\varphi(y)$, vrednosti za porazdelitveno funkcijo $F^{\circ}(y)$ itd. Ker so si vse normalne porazdelitve podobne, velja zanje nekaj splošnih lastnosti, ki to nalogo poenostavijo.

Iz slike 13.1 sklepamo, da je $F^{\circ}(y=M) = 0,50$ za vsako porazdelitev, ne glede na to, kolik je M. V vsaki normalni porazdelitvi je zaradi simetrije glede na aritmetično sredino polovica vrednosti pod, polovica pa nad aritmetično sredino M. Celo več! Za vsako normalno porazdelitev, ne glede na to, kakšen je M in σ, je v razmiku od $M-\sigma$ do $M+\sigma$ relativna frekvenca enaka 0,6827. Podobno je v razmiku $M-2\sigma$ do $M+2\sigma$ relativna frekvenca enaka 0,9545, v razmiku $M-3\sigma$ do $M+3\sigma$ 0,9973 itd. V splošnem moremo reči, da je za normalno porazdelitev relativna frekvenca v razmiku $M - z\sigma$ do $M + z\sigma$ odvisna samo od koeficienta z, nič pa od M in σ.

Enako je tudi s porazdelitveno funkcijo oziroma ploščino v razmiku $-\infty$ do $M+z\sigma$. Za porazdelitveno funkcijo za normalno porazdelitev velja, da je

$$F^{\circ}(y = M + z\sigma) = F^{\circ}(z)$$

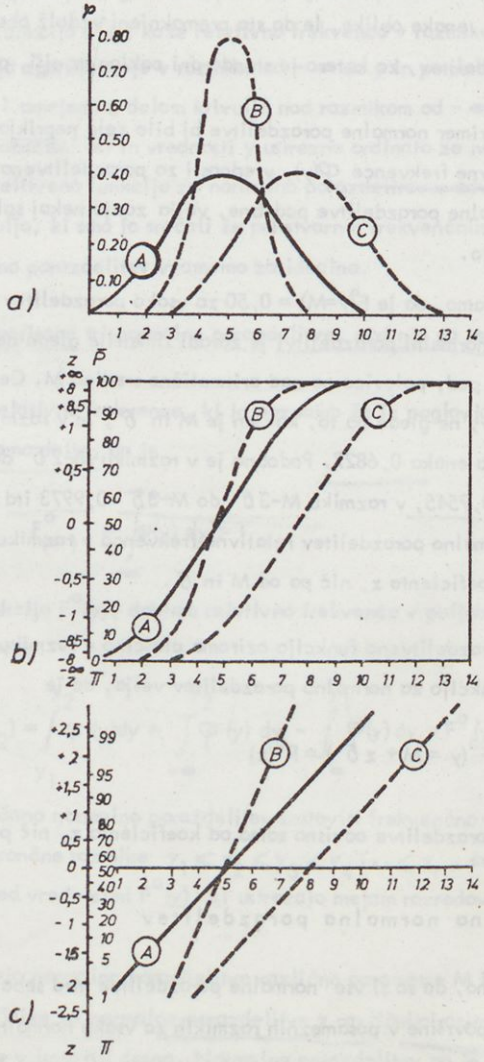
za vse normalne porazdelitve odvisna samo od koeficienta z, nič pa od M in σ.

Standardizirana normalna porazdelitev

13.6 Nakazali smo, da so si vse normalne porazdelitve med seboj podobne. Spoznali smo, da so površine v posameznih razmikih za vsako normalno porazdelitev odvis-

ne le od standardiziranih odklonov z, ne pa od M in σ. Ker je standardiziran odklon z osnovnim znakom y v linearni zvezi, se tudi z porazdeljuje v normalni porazdelitvi.

Iz lastnosti standardiziranega odklona spoznamo, da se z za normalno populacijo po-



Slika 13.2. Različne normalne porazdelitve in njihove porazdelitvene funkcije

razdeljuje normalno z $M_z = 0$ in $\sigma_z = 1$. To porazdelitev imenujemo standardizirano normalno porazdelitev. Če vstavimo te posebne vrednosti v obrazec za gostoto relativne frekvence v obrazcu 13.1, dobimo, da je gostota relativne frekvence za standardizirano normalno porazdelitev

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{13.5}$$

porazdelitvena funkcija $F^{\circ}(z)$ pa

$$F^{\circ}(z) = \int_{-\infty}^{z} \varphi(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz \tag{13.6}$$

V tabelah v dodatku imamo tabelirane ordinate $\varphi(z)$ in površine $H^{\circ}(z)$ za standardizirano normalno porazdelitev za razmik od 0 do z

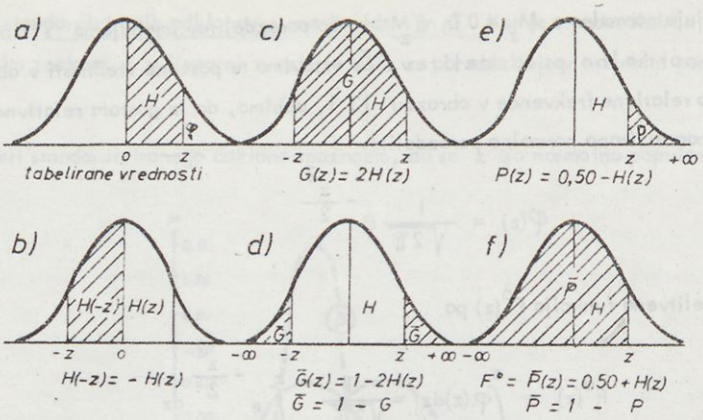
$$H(z) = \int_{0}^{z} \varphi(z)\, dz \tag{13.7}$$

V tabelah so dane ordinate za pozitivne vrednosti za z. Ker je normalna porazdelitev simetrična, velja

$$H(-z) = -H(z) \tag{13.8}$$

Za normalno porazdelitev potrebujemo večkrat različno opredeljene površine, tabelirane pa imamo samo površine H v razmiku od 0 do z. Zato so v sliki 13.3 nakazani odnosi med posameznimi površinami, ki pridejo v praksi v poštev v različnih primerih.

$$z = \frac{x - \Lambda}{\sigma}$$



Slika 13.3. Različne površine pod standardizirano normalno porazdelitvijo

13.7 Tabele za standardizirano normalno porazdelitev uporabljamo za vse normalne porazdelitve. Gostoto frekvence g(y) za normalno porazdelitev, ki ima obseg N, aritmetično sredino M in standardni odklon σ za poljuben y dobimo iz zveze

$$g(y) = N\,\varphi(y) = N \cdot \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-M}{\sigma}\right)^2} = \frac{N}{\sigma} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} = \frac{N}{\sigma}\varphi\left(z = \frac{y-M}{\sigma}\right) \tag{13.9}$$

tako, da poiščemo najprej z, ki ustreza danemu y, nato v tabelah za standardizirano normalno porazdelitev poiščeno ustrezno gostoto relativne frekvence $\varphi(z)$. To pa pomnožimo s kvocientom N/σ.

Vzemimo za zgled normalno porazdelitev o produktivnosti dela za kolektiv z N = 400 delavci s poprečno produktivnostjo dela M = 340 kosov na uro in σ = 60 kosov. Ordinato za y = 390 dobimo po nakazanem postopku. Najprej izračunamo y = 390 ustrezajoči z.

$$z = \frac{390 - 340}{60} = 0,83$$

V tabelah dobimo, da je ustrezna ordinata za standardizirano normalno porazdelitev $\varphi(z = 83) = 0,2827$. Gostota frekvence na iskani porazdelitvi pa je za y = 390 po obrazcu 13.9 enaka

$$g(y = 390) = \frac{400}{60} \cdot 0{,}2827 = 1{,}89$$

Absolutne frekvence za normalno populacijo z obsegom N pa dobimo tako, da ustrezne relativne frekvence iz tabele za standardizirano porazdelitev pomnožimo z obsegom populacije N.

Predpostavimo, da je porazdelitev o produktivnosti dela normalna. Da ugotovimo, koliko delavcev presega produktivnost dela y = 390 kosov na uro, poiščemo z ustrezno vrednost

$P(z) = 0{,}50 - H(z) = 0{,}50 - 0{,}2967 = 0{,}2033$

Če to relativno frekvenco pomnožimo z obsegom populacije N = 400, dobimo $400 \times 0{,}2033 = 81$ delavcev.

13.8 Čeprav imamo podrobno tabelo površin, navajamo v tabeli 13.1 nekaj najznačilnejših odnosov med standardiziranim odklonom z in ustreznimi površinami H, G, $\bar{G}$, P in $\bar{P}$.

Tabela 13.1 Odnosi med standardiziranim odklonom z in ustreznimi relativnimi frekvencami H, G, $\bar{G}$, P in $\bar{P}$.

	v o d s t o t k i h				
z	H	G	$\bar{G}$	P	$\bar{P} = F^{o}$
0,6745	25,00	50,00	50,00	25,00	75,00
1,000	34,13	68,26	31,74	15,87	84,13
1,645	45,00	90,00	10,00	5,00	95,00
1,960	47,50	95,00	5,00	2,50	97,50
2,000	47,73	95,46	4,54	2,27	97,73
2,326	49,00	98,00	2,00	1,00	99,00
2,576	49,50	99,00	1,00	0,50	99,50
3,000	49,87	99,74	0,26	0,13	99,87
3,090	49,90	99,80	0,20	0,10	99,90
3,291	49,95	99,90	0,10	0,05	99,95

13.9 Če predpostavljamo normalnost populacij, s tabelami za normalno porazdelitev rešujemo najrazličnejše probleme. Za ilustracijo vzemimo tole nalogo. Stroj izdeluje profile s poprečnim premerom M = 49,3 mm in standardnim odklonom σ = 0,8mm. Izdelki pa so uporabni, če je premer profila $M = 50 \pm 1{,}5$ mm. Ugotoviti je treba:

a) kolik je poprečen odstotek izmeta na stroju,

b) koliko moremo proizvodnjo popraviti, če moremo aritmetično sredino M regulirati in naravnati na optimalno stanje,

c) kolik bi moral biti pri optimalnem M standardni odklon, da bi bil izmet samo 5%.

Standardizirana odklona za meji uporabnosti sta:

$$z_s = \frac{y_s - M}{\sigma} = \frac{48{,}5 - 49{,}3}{0{,}8} = -1{,}00; \quad z_z = \frac{y_z - M}{\sigma} = \frac{51{,}5 - 49{,}3}{0{,}8} = +2{,}75$$

Če poiščemo v tabeli za normalno porazdelitev tem standardiziranim odklonom ustrezne relativne frekvence, dobimo:

H(-1,00) = -H(1,0) = -0,3413; H(+2,75) = 0,4970. Pod z = -1,0 je 0,50 - 0,3413 = = 0,1587, nad z = + 2,75 pa 0,50 - 0,4970 = 0,0030 relativne frekvence. Skupen odstotek izmeta je torej 15,87 + 0,30 = 16,17 %.

Kakor vidimo, poprečje za profil, ki ga proizvajamo, ni sredi razmika, v katerem so izdelki uporabni. Zato je delež premajhnih profilov velik. Glede na obliko normalne porazdelitve bi bilo najboljše, če bi se M in sredina razmika ujemala. Ker moremo spreminjati povprečje profila, ga naravnajmo na sredino tolerančnega razmika (M = 50).

Glede na predpis je torej izdelek še uporaben, če se od novega poprečja ne odklanja za več kot 1,50 mm. Temu odklonu ustrezni standardiziran odklon je

$$z = \frac{1{,}50}{0{,}8} = 1{,}875$$

Zanj je po tabeli H(z = 1,875) = 0,4697. Ker je $\bar{G} = 1 - 2H = 1{,}0 - 2 \cdot 0{,}4697 =$ = 0,0606, vidimo, da smo uspeli izmet skrčiti za 10,11%, ker je novi odstotek izmeta le 6,06 %.

Da bi rešili še tretji problem, moramo postopati v obratnem vrstnem redu. Da bi do-

segli cilj, da bi bil izmet 5%, moramo najprej poiskati, kakšen standardiziran odklon ustreza temu $\bar{G}(z) = 1-2H(z) = 0,05$. Vrednosti $H(z) = 0,475$ ustreza $z = 1,96$, standardizirani tolerančni odklon pa je 1,96, ker je

$$\frac{1,5}{\sigma} = 1,96$$

oziroma $\sigma = \frac{1,5}{1,96} = ,765$

Prilagoditev normalne porazdelitve stvarni frekvenčni porazdelitvi

13.10 Kakor smo že omenili v uvodu, s teoretičnimi porazdelitvami rešujemo tudi tale problem. Imamo stvarno porazdelitev, za katero predpostavljamo, da je nastala ob določenih pogojih. Vzemimo, da vemo, katera teoretična porazdelitev ustreza tem pogojem. Če primerjamo obe porazdelitvi med seboj, nadalje ugotovimo, ali je stvarna porazdelitev taka, da moremo sklepati, če zadošča tem pogojem ali ne.

Pri reševanju takih in podobnih problemov pogosto stvarni porazdelitvi prilagodimo normalno. Za prilagojeno štejemo normalno porazdelitev, ki ima iste parametre kakor stvarna porazdelitev, torej isti obseg N, isto aritmetično sredino M in isti standardni odklon σ. Ker imamo tabelirane vrednosti za standardizirano normalno porazdelitev, si pri takih nalogah pomagamo z njo. V splošnem moremo prilagoditi normalno porazdelitev z dvema postopkoma: z metodo ordinat in metodo površin. Obravnavali bomo le metodo površin, ker daje natančnejše rezultate. Po metodi površin prilagodimo normalno porazdelitev stvarni porazdelitvi takole:

a) Za stvarno frekvenčno porazdelitev, ki ji hočemo prilagoditi normalno porazdelitev, izračunamo aritmetično sredino M in standardni odklon σ.

b) Za meje razredov $y_{k,min}$ izračunamo standardizirane odklone $z_{k,min}$.

c) Za dobljene $z_{k,min}$ poiščemo ustrezne vrednosti porazdelitvene funkcije $F^{o}(z)$ za normalno porazdelitev. Te dobimo iz tabel za standardizirano normalno porazdelitev v tabeli v dodatku. Ker imamo tabelirane H, ne pa F^{o}, izračunamo kumulativne relativne frekvence po znanem obrazcu $F^{o} = 0,5 + H$. Pri tem moramo paziti na negativne predznake pri z in H.

d) Iz kumulative relativnih frekvenc F° dobimo kumulativo absolutnih frekvenc F', če posamezne člene kumulativne vrste pomnožimo z obsegom populacije N.

e) Absolutne frekvence za prilagojeno normalno porazdelitev f'_k dobimo, če poiščemo razlike med zaporednimi členi v kumulativni vrsti frekvenc za normalno porazdelitev.

13.11 Za zgled prilagoditve normalne porazdelitve stvarni porazdelitvi vzemimo premere glav N = 500 zakovic, proizvedenih na določenem stroju. Po nakazanih stopnjah je prilagoditev dana v tabeli.

Tabela 13.2. Prilagoditev normalne porazdelitve populaciji premerov za N = 500 zakovic

y v mm	f_k	$y_{k,min}$	$\frac{y_{k,min}-M}{\delta} = z_{k,min}$	$H(z_{k,min})$	$H(z_{k,min}) + 0.5 = F^{\circ}(z_{k,min})$	$N F^{\circ}(z) = F'(z_{k,min})$	$\Delta F' = f'_k$	$f_k - f'_k = \Delta f$
13,05-13,14	5	13,045	- 3,31	- ·4995	·0005	·3	·3	
13,15-13,24	22	13,145	- 2,44	- ·4927	·0073	3,7	3,4	+ 1,3
13,25-13,34	94	13,245	- 1,57	- ·4418	·0582	29,1	25,4	- 3,4
13,35-13,44	165	13,345	- 0,70	- ·2580	·2420	121,0	91,9	+ 2,1
13,45-13,54	140	13,445	+ 0,17	+ ·0675	·5675	283,8	162,8	+ 2,7
13,55-13,64	59	13,545	+ 1,03	+ ·3485	·8485	424,3	140,5	- 0,5
13,65-13,74	14	13,645	+ 1,90	+ ·4713	·9713	485,7	61,4	- 2,4
13,75-13,84	1	13,745	+ 2,77	+ ·4972	·9972	498,6	12,9	+ 1,1
		13,845	+ 3,64	+ ·4999	·9999	500,0	1,4	- 0,4

$N = 500$ $N = 500,0$ ô

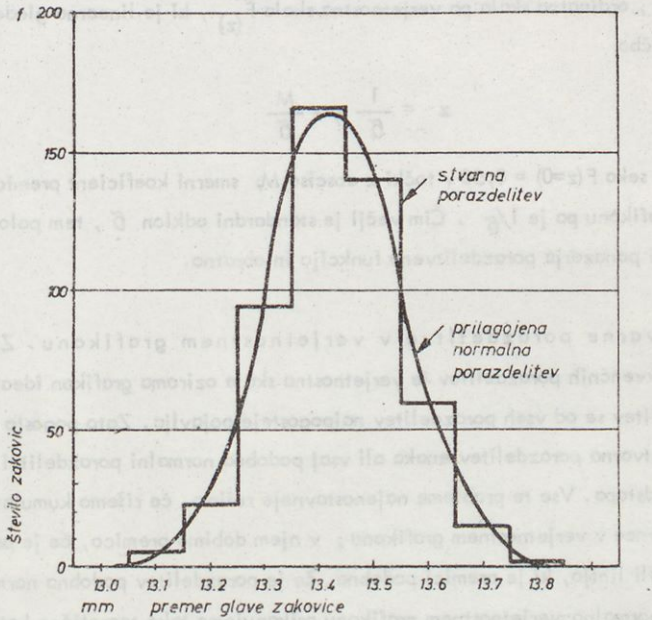
Normalna porazdelitev je prilagojena na osnovi naslednjih podatkov, ki so bili izračunani iz stvarne porazdelitve, za katero je: M = 13,426 mm; δ = 0,115 mm, i = 0,10 mm; N = 500. Prva vrednost za $z_{1,min}$ je izračunana neposredno

$$z_{1,min} = \frac{13,045 - 13,426}{0,115} = -3,31$$

za naslednje meje pa dobimo po obrazcu

$$z_{k+1,\min} = z_{k,\min} + \frac{i}{\sigma}\ ;(\text{npr. } z_{2,\min} = -3{,}31 + \frac{10}{115} = -2{,}44 \text{ itd.})$$

iz razlik v zadnjem stolpcu tabele spoznamo, da odstopanja stvarne porazdelitve od prilagojene normalne niso velika. V sliki 13.4 sta zaradi primerjave prikazani obe porazdelitvi: stvarna porazdelitev s histogramom, prilagojena normalna porazdelitev pa s krivuljo.



Slika 13.4. Stvarna frekvenčna porazdelitev premerov glav zakovic in prilagojena normalna porazdelitev.

Verjetnostni grafikoni

13.12 Verjetnostna skala. Porazdelitvena funkcija za normalno porazdelitev ima značilno obliko črke S. Kumulative za večino unimodalnih porazdelitev imajo tudi obliko črke S, čeprav niso normalne. Oblika seveda ni tako pravilna kot pri normalni porazdelitvi, vendar je tudi težko presoditi, ali ustreza normalni porazdelitvi ali ne.

Pri določeni transformaciji skale za kumulativo relativnih frekvenc pa uspemo, da je na takem grafikonu kumulativa relativnih frekvenc za vsako normalno porazdelitev premica. To dosežemo, če namesto linearne skale za kumulative relativnih frekvenc uvedemo tako imenovano v e r j e t n o s t n o s k a l o. Verjetnostno skalo dobimo, če linearni z-skali priredimo ustrezno skalo. Verjetnostna skala je vrisana kot tretja skala v sliki 13.1.

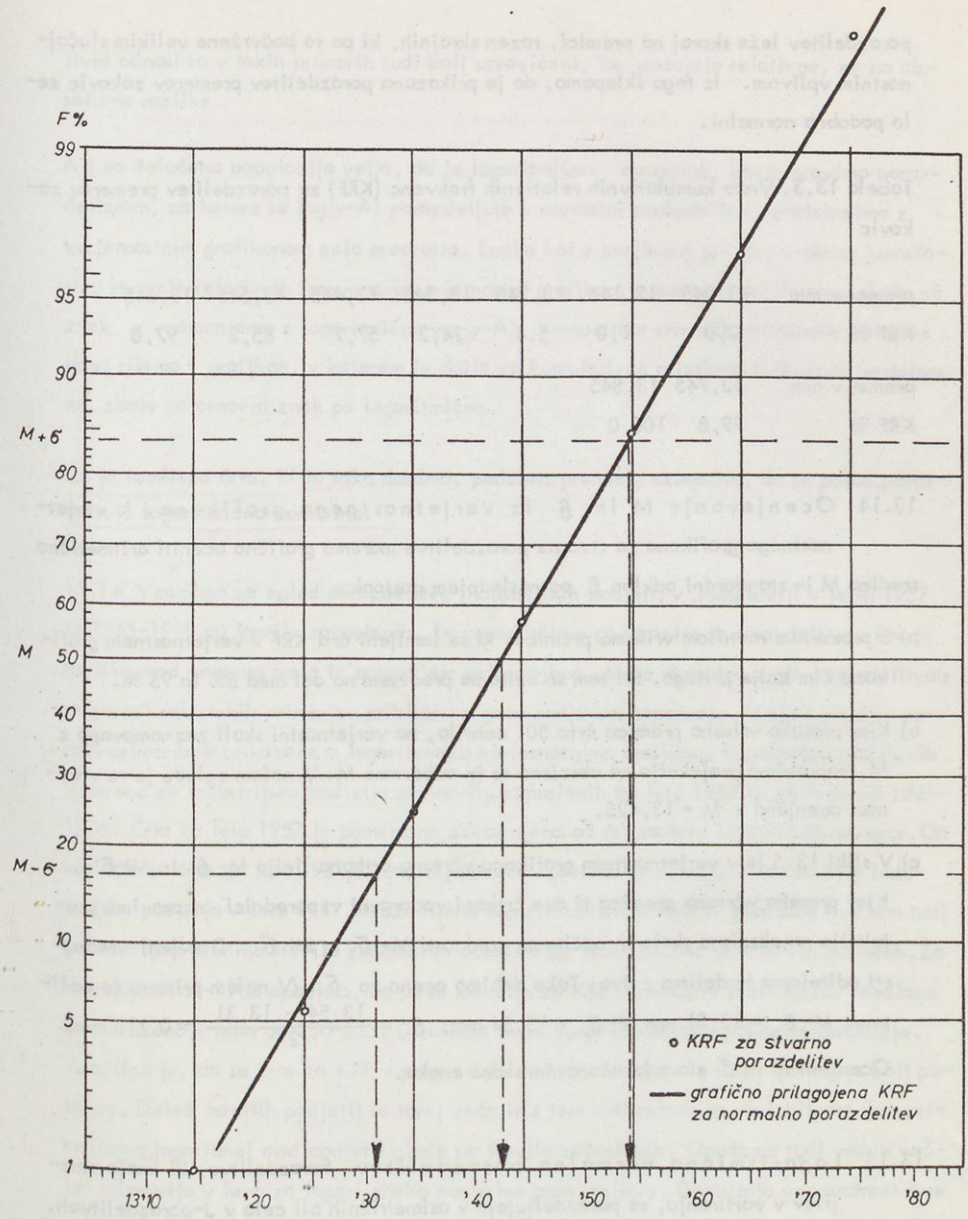
Ker je $F^{o}_{(z)}$ skala prirejena linearni skali za standardiziran znak, je porazdelitvena funkcija za normalno porazdelitev v grafikonu, v katerem je abscisna os linearna skala za znak y, ordinatna skala pa verjetnostna skala $F^{o}_{(z)}$, ki je linearna glede na z, premica z enačbo

$$z = \frac{1}{\sigma} y - \frac{M}{\sigma}$$

Ta premica seka $F(z=0) = 0,50$ v točki z absciso M; smerni koeficient premice v verjetnostnem grafikonu pa je $1/\sigma$. Čim večji je standardni odklon σ, tem položnejša je premica, ki ponazarja porazdelitveno funkcijo in obratno.

13.13. S t v a r n e p o r a z d e l i t v e v v e r j e t n o s t n e m g r a f i k o n u. Za študij frekvenčnih porazdelitev je verjetnostna skala oziroma grafikon idealna. Normalna porazdelitev se od vseh porazdelitev najpogosteje pojavlja. Zato pogosto ugotavljamo, ali je stvarna porazdelitev enaka ali vsaj podobna normalni porazdelitvi, ali in kako od nje odstopa. Vse te probleme najenostavneje rešimo, če rišemo kumulativne relativne frekvence v verjetnostnem grafikonu; v njem dobimo premico, če je porazdelitev normalna, ali linijo, ki je premici podobna, če je porazdelitev podobna normalni porazdelitvi. V normalno-verjetnostnem grafikonu prikazujemo tako teoretične kot stvarne porazdelitve. Če je kumulativa relativnih frekvenc za stvarno porazdelitev v verjetnostnem grafikonu lomljena črta, ki ima linearno smer, lahko sklepamo na normalnost porazdelitve. V takem grafikonu moremo vrisati premico, ki se lomljeni črti najbolje prilega. Ta premica je slika kumulative relativnih frekvenc za prilagojeno normalno porazdelitev.

V sliki 13.5 je na verjetnostnem grafikonu narisana kumulativa relativnih frekvenc za porazdelitev premerov glav zakovic iz zgleda 13.11. Točke, ki med seboj niso povezane, predstavljajo kumulativo relativnih frekvenc za stvarno porazdelitev, premica pa je slika porazdelitvene funkcije za prilagojeno normalno porazdelitev. Točke za stvarno



$$6 = \frac{13,54 - 13,31}{2} = 115$$

Slika 13.5 Verjetnostni grafikon za porazdelitev premerov glav N = 500 zakovic iz zgleda 13.11.

porazdelitev leže skoraj na premici, razen skrajnih, ki pa so podvržene velikim slučajnostnim vplivom. Iz tega sklepamo, da je prikazana porazdelitev premerov zakovic zelo podobna normalni.

Tabela 13.3. Vrsta kumulativnih relativnih frekvenc (KRF) za porazdelitev premerov zakovic

premer v mm	13,045	13,145	13,245	13,345	13,445	13,545	13,645
KRF %	0,0	1,0	5,4	24,2	57,2	85,2	97,0
premer v mm	13,745	13,845					
KRF %	99,8	100,0					

13.14 Ocenjevanje M in σ iz verjetnostnega grafikona. Iz verjetnostnega grafikona za stvarne porazdelitve moremo grafično oceniti aritmetično sredino M in standardni odklon σ po naslednjem postopku:

a) S prozornim ravnilom vrišemo premico, ki se lomljeni črti KRF v verjetnostnem grafikonu čim bolje prilega. Pri tem se oziramo predvsem na del med 5% in 95 %.

b) Kjer preseka vrisana premica črto 50. centila, na verjetnostni skali zaznamovana z M, napravimo projekcijo na abscisno os in odberemo M. V našem zgledu je ta vrednost ocenjena z $M = 13{,}425$.

c) V sliki 13.5 je v verjetnostnem grafikonu včrtana črtkana linija $M-\sigma$ in $M+\sigma$. Kjer preseka včrtana premica ti dve črtkani vodoravni vzporednici, napravimo projekcijo na abscisno skalo in odčitamo vrednosti $M-\sigma$ in $M+\sigma$. Dobljeni vrednosti odštejemo in delimo z dva. Tako dobimo oceno za σ. (V našem primeru je odčitano: $M-\sigma = 13{,}31$ mm $M+\sigma = 13{,}54$ mm; $\sigma = \frac{13{,}54 - 13{,}31}{2} = 0{,}115$ mm). Ocenjeni M in σ sta z izračunanim skoro enaka.

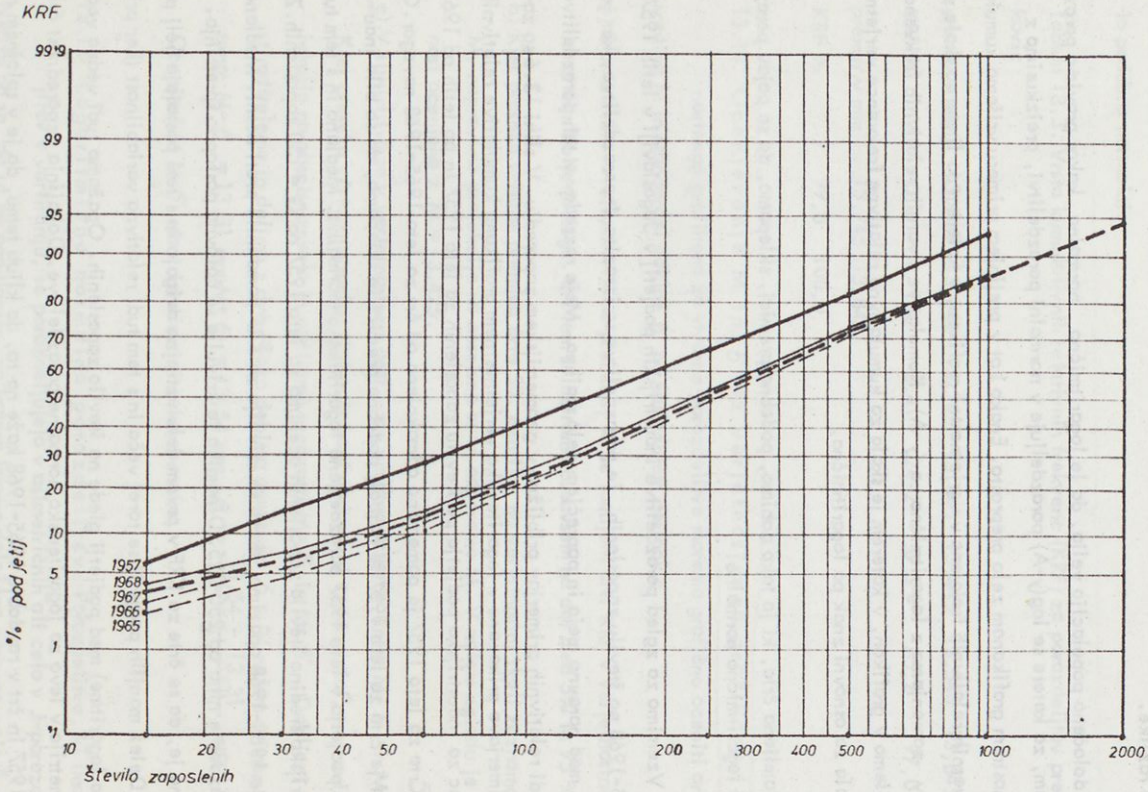
13.15 Logaritmično normalna porazdelitev. Porazdelitve, ki imajo omejitev v variiranju, se porazdeljujejo v asimetričnih ali celo v J-porazdelitvah. Izkaže pa se, da se za populacije, za katere imamo določeno naravno mejo A v variiranju, normalno porazdeljuje log (y-A). Ker je za veliko ekonomskih količin naravna meja A=0, se zanje porazdeljujejo v normalni porazdelitvi log y. Logaritmi oziroma rela-

tivni odnosi so v takih primerih tudi bolj upravičeni, ker podajajo relativne, ne pa absolutne razlike.

Ali za določeno populacijo velja, da je logaritmično normalna, kakor pravimo porazdelitvam, za katere se log(y-A) porazdeljuje v normalni porazdelitvi, preizkusimo z verjetnostnim grafikonom zelo preprosto. Enako kot v prejšnjem primeru vrišemo kumulativo stvarnih relativnih frekvenc v verjetnostni grafikon, v katerem pa linearno skalo za znak y zamenjamo z logaritmično za (y-A). Kumulative stvarnih relativnih frekvenc torej rišemo v grafikon, v katerem je skala za kumulativne relativne frekvence verjetnostna, skala za osnovni znak pa logaritmična.

Če je lomljena črta, ki jo tako dobimo, podobna premici, sklepamo, da se pojav porazdeljuje logaritmično normalno.

13.16 Vzemimo za zgled porazdelitve industrijskih podjetij v Jugoslaviji v letih 1957 in 1965-1968 po številu zaposlenih. Te porazdelitve so izrazite J-porazdelitve, ker je razlika med naravno mejo in poprečjem zelo majhna. Meje razredov v teh porazdelitvah so zaradi relativnih primerjav približno v geometrijskem zaporedju. V sliki 13.6 so zaradi primerjave prikazane v logaritemsko verjetnostnem grafikonu kumulative relativnih frekvenc za industrijska podjetja po številu zaposlenih za leto 1957 in po letih od 1965-1968. Črta za leto 1957 je pomembno odmaknjena od črt za leta 1965-1968 navzgor. Od vseh črt je črta za leto 1957 še najbolj podobna premici in sklepamo, da so bila industrijska podjetja v letu 1957 porazdeljena logaritemsko-normalno. Mediano in s tem tudi geometrijsko sredino števila zaposlenih ocenimo za leto 1957 na G = 130 zaposlenih. Za razdobje 1965-1968 opazimo, da so se krivulje za KRF pomaknile proti desni. Mediana za to razdobje niha od 250-225. Dinamika te krivulje se časovno postopno spreminja. Značilno je, da se črte za KRF v prvem delu vztrajno dvigajo in s tem postajajo bolj položne. Delež manjših podjetij se torej veča in s tem tudi relativna variabilnost (ker primerjamo logaritme) med podjetji glede na število zaposlenih. Opažamo tudi vedno večjo asimetrijo v levo za logaritemsko normalne porazdelitve. Dokajšnja vzporednost črte v letu 1957 in črt v razdobju 1965-1968 kaže na to, da kljub temu, da je v splošnem naraslo poprečno število zaposlenih, relativne razlike v številu zaposlenih niso bistveno različne.



Slika 13.6 Logaritemsko-verjetnostni grafikon industrijskih podjetij v SFRJ v letih 1957 in 1965-68 po številu zaposlenih

Tabela 13.4. Kumulativne relativne frekvence (KRF) za porazdelitev industrijskih podjetij po številu zaposlenih v letu 1957

Število delavcev	- 15	- 30	- 60	- 125	- 250	- 500	- 1000
KRF	6,1	14,7	28,6	48,8	67,5	82,4	93,3

Ti podatki so vrisani v logaritmično-verjetnostni grafikon v sliki 13.6.

13.18 Normalna porazdelitev kot verjetnostna porazdelitev. Najpomembnejša uporaba normalne porazdelitve pa je vsekakor v vzorčenju. Normalno porazdelitev vzamemo v vzorčenju bodisi kot predpostavko za določeno populacijo, ki jo proučujemo ali kot zakonitost porazdelitve za količine, ki so važne v vzorčenju. Normalna porazdelitev je tudi osnova za druge teoretične porazdelitve. Razen tega večina teoretičnih porazdelitev pri vzorčenju preide v limiti v normalno porazdelitev, če vzorec oziroma del populacije, iz katere z vzorčenjem ocenjujemo parametre, večamo.

Iz tega in iz nadaljnjih poglavij o vzorčenju spoznamo, da se normalna porazdelitev prepleta prek vseh delov in problemov vzorčenja, tako pri velikih vzorcih kot pri malih, pri ocenjevanju kot pri preskušanju domnev.

13.19 Vsa teorija vzorčenja je zgrajena na verjetnostnem računu. Zato morajo biti osnovne postavke, ob katerih veljajo zakonitosti, ki smo jih odkrili za vzorčenje, strogo izpolnjene.

Če predpostavimo, da poznamo osnove verjetnostnega računa, se omejimo na osnovno črto, ki je važna v vzorčenju. Verjetnostni račun se ukvarja z računanjem in preračunavanjem verjetnosti za katere koli verjetne dogodke. Ti verjetni dogodki dobe v vzorčenju določeno obliko. V vzorčenju je dejanje, ki ga opredelimo kot poskus, izbiranje enot iz osnovne populacije. Dogodek v smislu verjetnostnega računa pa je vrednost določenega znaka ali značilnosti več znakov za izbrane enote. Pojem aposteriorne verjetnosti, ki je opredeljena kot limita kvocienta števila ugodnih dogodkov in števila poskusov v enakih pogojih, če število poskusov večamo čez vse meje, predpostavlja, da so neki osnovni pogoji poskusov enaki. Pojem apriorne verjetnosti pa je zasnovan na predpostavki, da more-

mo skupnost vseh možnih dogodkov razstaviti v e lementarne dogodke, ki so enakomožni - simetrični. V vzorčenju je elementaren dogodek izbor posamezne enote iz populacije. Če hočemo doseči, da so ti dogodki enakomožni, moramo tehniko izbora enot prilagoditi temu pogoju. To dosežemo tako, da izbiramo enote populacije na slepo, na srečo ali slučajnostno, kakor pravimo izboru, ki zagotavlja enako možnost izbora za vsako enoto. Tehnične prijeme slučajnostnega izbora bomo obravnavali kasneje. Če dosežemo slučajnost izbora, se pojem apriorne verjetnosti sklada s pojmom relativne frekvence v statistiki. Če je v populaciji relativna frekvenca za neko karakteristiko, npr. 0,30, pomeni, da je verjetnost, da iz te populacije izberemo enoto, ki ima to značilnost, enaka 0,30, če je izbor enote slučajnosten. Če imamo npr. populacijo izdelkov, v kateri je relativna frekvenca za neuporabne artikle 0,05, je verjetnost, da iz te populacije s slučajnostnim izborom izberemo izdelek, ki je neuporaben, enaka 0,05.

13.20 Zaradi zakona o velikih številih velja, da se za slučajnostni izbor enot aposteriorna in apriorna verjetnost skladata. To pomeni, da se kvocient med številom slučajnostno izbranih enot, ki imajo dano značilnost in skupnim številom vseh slučajnostno izbranih enot približuje relativni frekvenci za to značilnost, če število izbranih enot večamo.

Iz gornjega sledi, da moremo vzeti porazdelitev relativnih frekvenc za verjetnostno porazdelitev, če je izbor iz te populacije slučajnosten. Pri tem se relativna frekvenca za določeno vrednost ali razmik vrednosti sklada z apriorno verjetnostjo, da iz populacije slučajnostno izberemo enoto, ki ima ustrezno vrednost ali vrednosti iz ustreznega razmika. To velja za nezvezne in zvezne porazdelitve. Pri nezveznih porazdelitvah pomeni relativna frekvenca za določeno vrednost verjetnost, da slučajnostno izberemo enoto, ki ima ustrezno vrednost. Pri zveznih porazdelitvah pa relativna frekvenca za določen razmik pove, kolikšna je verjetnost, da iz zvezne populacije slučajnostno izberemo enoto, za katero je vrednost v ustreznem razmiku. Pojem slučajnostne spremenljivke pa se sklada s pojmom znaka.

Porazdelitev relativnih frekvenc za normalno porazdelitev moremo torej vzeti za verjetnostno porazdelitev za populacijo, ki je normalno porazdeljena. Tabela površin za normalno porazdelitev je obenem tabela verjetnosti. Normalna verjetnostna porazdelitev je iz razlogov, ki smo jih navedli uvodoma, posebno pomembna za vzorčenje.

Če vzamemo populacijo, ki je normalno porazdeljena, je npr. verjetnost, da iz nje slučajnostno izberemo enoto, ki je v mejah $M - \delta$ do $M + \delta$, enaka relativni frekvenci za ta razmik. Ta pa je po tabeli 13,1 enaka 0,6826. Če slučajnostni izbor iz populacije ponavljamo, pričakujemo, da v poprečju približno 2/3 vseh slučajnostno izbranih vrednosti leži v gornjem razmiku. Podobno dobimo ustrezne verjetnosti za poljuben razmik.

13.21 Pojem tveganja. V vsakdanjem življenju vzamemo dogodke, ki so zelo verjetni, praktično tako, kakor da se bodo za gotovo zgodili. Zato računamo, da nas pri sprehodu ne bo povozil avto, ker je velika verjetnost, da nas resnično ne bo. Ko kupujemo konzervo, računamo, da ni pokvarjena, ker je zelo verjetno, da res ni, itd. Pri vseh teh sklepih pa je določena stopnja tveganja. Lahko se zgodi, da nas povozi avto. Lahko se zgodi, da je konzerva pokvarjena, čeprav računamo, da ni. To tveganje je tem večje, čim večja je verjetnost, da se dogodek, ki ga pričakujemo, ne zgodi. Večje je tveganje je, da konzerva iz lanske proizvodnje ni pokvarjena, kakor da ni pokvarjena konzerva iz letošnje proizvodnje. Tveganje je torej verjetnost, da se dogodek, ki ga pričakujemo, ne zgodi.

Pojmu tveganja v vsakdanjem življenju popolnoma ustreza pojem tveganja v verjetnostnem računu oziroma statistiki.

Vrednosti standardizirano normalne porazdelitve teoretično leže na vsem razmiku od $-\infty$ do $+\infty$. Vendar praktično vzamemo, da vse vrednosti leže v razmiku med -3,29 do +3,29. Če trdimo, da je vrednost enote, ki jo slučajnostno izberemo iz standardizirano normalno porazdeljene populacije v razmiku od -3,29 do +3,29, je tveganje te izjave 0,001 (glej tabelo 13,1). To pomeni: le z verjetnostjo 0,001 moremo dobiti pri slučajnostnem izboru enoto, za katero je vrednost izven navedenega razmika. To tveganje pa moremo tolmačiti tudi drugače. Glede na aposteriorno verjetnost moremo sklepati: če hoče določen raziskovalec preizkusiti trditev, da slučajnostno izbrana enota leži med -3,29 in +3,29, in ponavlja izbor enote večkrat, v velikem številu ponavljanj za 0,001 ali 1 ‰ enot ugotovi, da njihove vrednosti ne leže v imenovanem razmiku.

Stopnjo tveganja moremo poljubno spreminjati. Če vzamemo širši razmik, je tveganje

manjše, če vzamemo ožji, je stopnja tveganja večja. Za vsak poseben primer moremo stopnjo tveganja ugotoviti tako, da izračunamo verjetnost, da se napovedani dogodek ne zgodi.

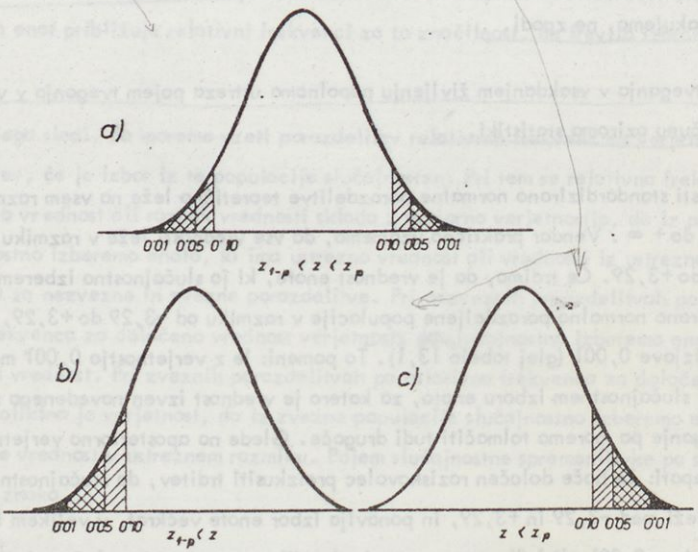
13.22 Za normalno porazdelitev moremo v splošnem z določenim tveganjem napovedati, kakšna je vrednost enot, ki jo slučajnostno izberemo iz normalne populacije, na tri različne načine:

a) nakazati moremo razmik, v k a t e r e m je z določeno stopnjo tveganja vrednost, ki jo ima slučajnostno izbrana enota;

b) nakazati moremo vrednost, n a d k a t e r o je z določeno stopnjo tveganja nost, ki jo ima slučajnostno izbrana enota;

c) nakazati moremo mejo, p o d k a t e r o je z določeno stopnjo tveganja za enoto, ki jo slučajnostno izberemo iz normal ne populacije.

Prva je dvostranska, drugi dve pa sta enostranski trditvi.



Slika 13. 7 Kritični razmiki in kritične meje za normalno porazdelitev

Stopnja tveganja, s katero dajemo izjave, ki smo jih pravkar navedli, more biti poljubna. Običajne stopnje tveganja so: 0,10, 0,05, 0,01 in 0,001. Stopnja tveganja 0,10 je zelo velika in pomeni, da moremo pričakovati, da se v poprečju v enem od desetih poskusov izkaže, da trditev, ki jo postavljamo, ne drži. Zato to stopnjo tveganja redko uporabljamo. Najpogosteje sklepamo s tveganjem 0,05. To pomeni, da moremo pri takem sklepanju pričakovati, da se v enem izmed dvajsetih primerov v poprečju izkaže, da naša trditev ne drži. Veliko bolj zanesljive so trditve s tveganjem 0,01 ali 0,001. Pri teh se izkaže, da trditev ne velja poprečno v enem od sto oziroma od tisoč preskusov.

Razmik, v katerem je z določeno stopnjo tveganja vrednost za slučajnostno izbrano enoto, imenujemo kritični razmik, meje pa kritične meje.

V sliki 13.7 so prikazani kritični razmiki in kritične meje za vse tri vrste trditev.

Glede na pomen P (glej sliko 13.3) na splošno velja za slučajnostno izbrano enoto iz standardizirano normalne porazdelitve:

	Razmik	Tveganje
s tveganjem 2P	$-1,64 < z < +1,64$	0,10
	$-1,96 < z < +1,96$	0,05
$z_{1-P} < z < z_P$	$-2,58 < z < +2,58$	0,01
	$-3,29 < z < +3,29$	0,001
s tveganjem P	$-1,28 < z$	0,10
	$-1,64 < z$	0,05
$z_{1-P} < z$	$-2,32 < z$	0,01
	$-3,09 < z$	0,001
s tveganjem P	$z < +1,28$	0,10
	$z < +1,64$	0,05
$z < z_P$	$z < +2,32$	0,01
	$z < +3,09$	0,001

13.23 Vse, kar smo povedali o tveganju v zvezi z normalno porazdelitvijo, velja za katero koli drugo verjetnostno porazdelitev. Vse to, skupno s kritičnimi razmiki in kritičnimi mejami, moremo prenesti na binomialno, Poissonovo in hipergeometrijsko porazdelitev, t-, χ^2 - in F-porazdelitev. Vse to velja za vse te porazdelitve, čeprav so prve tri nezvezne. Zanje po znanem postopku za nezvezne vrste vzamemo, da

se relativna frekvenca - verjetnost nanaša na enotin razmik okrog dane vrednosti. Tako izenačimo zvezne verjetnosti z nezveznimi.

ŠTIRINAJSTO POGLAVJE

VZORČENJE - VELIKI VZORCI

OSNOVE VZORČENJA

14.1 Vzporedno s hitrim ekonomskim in družbenim razvojem se je v zadnjih desetletjih vedno bolj večala potreba po statističnih podatkih v najrazličnejših področjih.

Stare metode zbiranja in obdelave podatkov niso mogle zadostiti novim zahtevam, ker so bile drage, počasne, zvezane z angažiranjem velikega števila ljudi itd. Razen tega so se te metode zaradi posebnih pogojev izkazale za neuporabne za nova nerazvita območja.

Zato smo morali odstopiti od osnovnega načina opazovanja statističnih populacij, to je popisovanja v ožjem smislu in v marsikaterem področju začeti ocenjevati statistične podatke. Statistične podatke ocenjujemo na več načinov. Nekatere od njih smo našteli v prvem delu. Od teh so nekatere ocene izdelane na osnovi poznavanja predmeta, bolj ali manj brez uporabe statistike. Druge, npr. metoda izbora tipičnih enot, monografija, metoda opazovanja osnovne mase itd., pa so izdelane na poznavanju dela populacije. Vendar imajo vse te metode ocenjevanja večinoma iste pomanjkljivosti. Vse so subjektivne, za sestavljanje ocen pa moramo razmeroma zelo podrobno poznati osnovno populacijo: razen tega pri nobeni od njih ne vemo, kolikšna je razlika ocene od prave vrednosti, ki bi jo dobili, če bi opazovali populacijo, oziroma koliko so te ocene zanesljive.

Vseh teh pomanjkljivosti, ki jih imajo ocene gornjih tipov, pa nima metoda vzorčenja. Po tej metodi ocenjujemo parametre populacije iz dela enot populacije - vzor-

ca. Enote vzorca izberemo iz populacije podobno kot številke pri loteriji po metodi slučajnega izbora tako, da ima vsaka enota populacije enako možnost, da je izbrana. Če je vzorec, kakor pravimo, slučajnosten, ima ta ocena kakovosti, ki jih nobena drugačna ocena nima.

Za slučajnosten vzorec namreč veljajo zakonitosti, ki bistveno pripomorejo h kakovosti ocene. Te pa moremo odkriti z uporabo verjetnostnega računa. Zaradi tega je osnova metode vzorčenja verjetnostni račun. Zakonitosti vzorčenja pa veljajo le, če so enote izbrane slučajnostno.

14.2 Metoda vzorčenja je v svojih posebnih oblikah v nekaterih znanostih že dolgo znana kot raziskovalna metoda. Tako je vzorčenje že razmeroma dolgo znano v agronomiji, biologiji, medicini, psihologiji itd. Vzrok, da se je metoda vzorčenja uveljavila najprej v teh znanostih, ni slučaj. Raziskovalec v teh področjih je bil že od vsega začetka nujno navezan na opazovanje samo dela populacije, ker celotne populacije ne obstajajo. Tako je za agronoma trideset preskusnih parcelic samo vzorec iz populacije neomejenega števila parcelic, ki bi jih obravnaval enako kakor stvarnih trideset parcelic. Prav tako je v poskusih v biologiji ali medicini trideset ali štirideset poskusov na miših samo vzorec iz neomejenega števila poskusov ob enakih pogojih. Enako vzamemo določeno število izdelkov, ki jih proizvedemo na določenem stroju, za vzorec iz neomejene populacije izdelkov, ki bi jih stroj proizvedel ob enakih pogojih.

V teh znanostih se je vzorčenje uveljavilo razmeroma zgodaj, ker s popolnim opazovanjem niso mogli zajeti neomejenih - umišljenih populacij. Ravno to pa je vzrok, da se je vzorčenje razmeroma pozno začelo uveljavljati tudi v socialno-ekonomskih znanostih. Te znanosti namreč večinoma obravnavajo populacije, ki imajo končno število enot. Te pa je možno opazovati s klasičnimi metodami opazovanja, s popisi, s tekočimi službami itd. Vendar je potreba po statističnih podatkih prerasla možnosti, ki jih dajejo klasične metode. Zato so izdelali specifične metode vzorčenja, ki omogočajo objektivno in kakovostno ocenjevanje podatkov tudi za to področje. Te metode so sedaj splošno orodje statistike in v mnogih področjih izpodrivajo klasične metode opazovanja.

Prednosti in pomanjkljivosti metode vzorčenja

14.3 Prednosti vzorčenja pred drugimi metodami ocenjevanja, a v nekaterih primerih tudi pred metodami popolnega opazovanja, so tele:

a) Število enot v vzorcu je znatno manjše kakor pri popolnem opazovanju.
Pri velikih populacijah je število enot v vzorcu komaj nekaj promilov ali odstotkov od skupnega obsega populacije.
Manjše število enot pri vzorčenju pa ima niz prednosti.

b) Stroški vzorčenja so znatno manjši kakor stroški za popolno opazovanje.

c) Kakovost zbranih podatkov je v marsikaterem primeru celo boljša kot pri popisu, ker je za izvedbo potrebno manj ljudi. Ti pa so zaradi lažje izbire boljši, bolje poučeni itd.

d) Zaradi izbire kadrov in možnosti kontrole so napake pri popisovanju kar najbolj skrčene.

e) Čas, v katerem dobimo rezultate, je znatno krajši kot pri popolnem opazovanju: to pa poveča operativno vrednost podatkov.

f) Pogrešek, ki izvira iz tega, da nismo opazovali celotne populacije, temveč samo del, moremo pri vzorčenju objektivno izmeriti in uravnavati. Te lastnosti nima nobena druga metoda ocenjevanja. Ta pogrešek, ki ga moremo oceniti, je dostikrat celo manjši kakor pa je rezultat popisnih napak pri celotnem opazovanju. Popisnih napak je predvsem kriv velik obseg populacije, sodelovanje velikega števila popisovalcev, ki jih ne moremo uspešno poučiti in kontrolirati, itd.

g) V raziskovalnem delu je metoda vzorčenja edino možna, ker umišljenih - hipotečnih populacij v nobenem primeru ne moremo zajeti v celoti.

P o m a n j k l j i v o s t i vzorčenja pa so predvsem v tem:

a) Vzorčenje moremo uporabiti le za populacije z velikim številom enot.

b) Zato ne moremo dobiti z njim zadovoljivih ocen za manjša, temveč le za večja območja.

c) Vzorčenje ne da zanesljivih podrobnih rezultatov. Zato z njim ne moremo obdelovati kombinacije več znakov.

d) Zanesljivost ocen za posamezne podatke je različna.

e) Načrt vzorčenja je bolj zapleten kakor načrt popolnega opazovanja. Enako je računanje ocen in pokazovalcev zanesljivosti ocen razmeroma zamotano.

Uporaba vzorčenja pri statističnem opazovanju

14.4 Vzorčenje moremo uporabiti v vseh stopnjah statističnega opazovanja.

Pri izvedbi statistične akcije more vzorec nadomestiti popolno opazovanje ali pa ga uporabimo kot dopolnilo k popolnemu opazovanju. Tako npr. z vzorcem v nekaterih letih v Jugoslaviji nadomestimo popis živine. Kot dopolnilo pa vzorčenje uporabimo, če podatke, ki jih potrebujemo zelo podrobno tudi za manjša območja, zberemo s popisom. Podatke, ki jih potrebujemo samo za večja območja pa istočasno opazujemo z vzorcem. Zato prve podatke zberemo od vseh enot populacije, druge pa samo od vnaprej naključno izbranih enot.

Pri obdelavi statističnega gradiva moremo uporabiti vzorčenje za dva namena. Obdelava velikih popisov traja razmeroma dolgo. V tem času podatki velikokrat že izgube aktualnost. Z vzorcem predhodnih rezultatov pa moremo dobiti ocene podatkov znatno pred koncem popolne obdelave. V tem primeru je popolna obdelava tudi potrebna, ker dá vzorec samo globalne, ne pa podrobnih rezultatov.

Moremo pa vzorčenje uporabiti tudi za dopolnilno obdelavo. Čeprav s statističnim opazovanjem zberemo popolne podatke za vsa vprašanja, včasih popolno obdelamo le nekatera od teh vprašanj. Druge pa obdelamo z vzorcem v dopolnilni obdelavi, ker sicer bodisi zaradi pomanjkanja sredstev ali drugih vzrokov ne bi teh podatkov sploh obdelali.

Z vzorcem moremo tudi kontrolirati izvedbo statistične akcije v vseh stopnjah. Z njim ocenimo delo popisovalcev ali obdelavo statističnih podatkov v celoti. Vzorčenje uporabljamo celo za popravo podatkov, ki smo jih zbrali s popisom. Pri popisu živine, pa tudi pri drugih statističnih akcijah, neposredno po popisu izvedemo kontrolni vzorec. Kakovostno dobri, zanesljivi in posebej za to poučeni kontrolorji z vzorcem ugotove, koliko so zbrani podatki različni od stvarnih. Podatke zberejo kontrolorji z drugačnimi sredstvi, bolj vestno in natančno kakor popisovalci. Tako ugotovimo učinek napak, ki jih z drugimi sredstvi kontrole ne moremo ugotoviti. Iz podatkov kontrolnega vzorca pa ocenimo za vsako vrsto živine oziroma podatek posebej, koliko je stvaren oziroma za koliko je treba v skupnem ta podatek popraviti. Vzorčna kontrola ima namen ugotoviti skupni učinek napak, ne pa napake posameznika.

Čeprav je v problemih, ki smo jih nakazali v prejšnjem odstavku, število enot v vzorcu razmeroma majhen del celotne populacije, je število enot vseeno razmeroma veliko, saj je imel npr. 2% vzorec popisa prebivalstva v Sloveniji v letu 1953 ca 30.000 enot.

Vzorčenje se je razvilo v eno izmed osnovnih metod pri proučevanju javnega mnenja, analize tržišča, kontrole proizvodnje in proizvodnega procesa itd. Pri teh problemih so vzorci razmeroma majhni in v dosti primerih ne presegajo nekaj deset enot populacije. Seveda rezultati takih vzorcev niso tako zanesljivi kakor rezultati vzorcev z več tisoči enot. Vendar je njihova zanesljivost, glede na problem, ki ga s takimi vzorci rešujemo, večinoma ustrezna.

Osnovna populacija. Vzorec. Populacija vseh vzorcev.

14.5 Osnovna populacija. Populacijo, za katero skušamo oceniti parametre z vzorčenjem, imenujemo osnovno populacijo. Osnovna populacija je npr. populacija vseh kmetijskih gospodarstev v SR Sloveniji na določen dan, če z vzorčenjem ocenjujemo podatke za skupnost kmetijskih gospodarstev v Sloveniji na ta dan. Za osnovno populacijo štejemo tudi skupnost podatkov o številu goveje živine za posamezna gospodarstva v Sloveniji na določen dan. Osnovna populacija je tudi npr. populacija industrijskih podjetij ali pa populacija vrednosti proizvodnje posameznih podjetij, populacija števila delavcev posameznih podjetij itd. Te populacije imenujemo končne osnovne populacije, ker imajo končno število enot.

Z vzorčenjem pa obravnavamo tudi osnovne populacije, ki imajo neomejeno število enot. Te populacije sicer stvarno ne obstajajo, vendar jih moremo miselno dojeti in razumeti. Te vrste populacij obravnavamo predvsem pri raziskovalnem delu. Umišljena ali hipotetična populacija, kakor imenujemo take vrste populacij, je npr. skupnost vseh ponovitev določenega poskusa ob enakih pogojih.

Enako imamo za hipotetično populacijo skupnost vseh izdelkov, ki bi jih proizvedel nek stroj ob enakih pogojih, če bi delal neomejeno. Tudi iz tega primera vidimo, da ima taka umišljena populacija neomejeno število enot, da pa obstaja samo miselno, ne pa stvarno. Praktično je namreč nemogoče ustvariti enake pogoje za razmeroma veliko - kaj šele za neomejeno proizvodnjo izdelkov, ne glede na to, da ne moremo take populacije nikdar izčrpati.

Osnovne populacije z neomejenim številom enot moremo izpeljati tudi iz zveznih populacij. Parcelo z obsegom enega hektarja moremo razdeliti v populacijo s končnim številom površinic po 1 m^2. Če pa vzamemo drugo osnovo in imamo za enoto populacije vsako parcelico kvadratne oblike s površino 1 m^2, ne glede na to, kako je položena, in dopustimo, da se med seboj prekrivajo, dobimo osnovno populacijo z neomejenim številom enot. Podobno populacijo dobimo iz osnovne zvezne populacije velikega kosa tkanine, zvezne populacije proizvodnje cementa v določenem dnevu itd.

Enote osnovne populacije so posamezno gospodarstvo, posamezno industrijsko podjetje, poskus ob enakih pogojih, parcelica po 1 m^2, 10 dkg cementa itd. Znaki teh enot so vse značilnosti, ki jih proučujemo. Parametri osnovnih populacij pa so vsota podatkov, srednje vrednosti, strukturni odstotki, mere variacije, mere korelacije itd.

14.6 Vzorec. Iz osnovne populacije izberemo delno populacijo, ki jo imenujemo vzorec. Enote vzorca morajo biti izbrane tako, da ima vsaka enota enako možnost, da je izbrana v vzorec. To dosežemo s slučajnostnim izborom. Vzorec je sam zase tudi populacija, ker ima svoje enote (enote osnovne populacije, ki smo jih izbrali), te pa znake, ki so isti kot znaki v osnovni populaciji. Kakor za vsako drugo populacijo moremo iz podatkov v vzorcu izračunati sredine, proporce, mere variacije in korelacije itd.

Ker je glede na število enot teorija vzorčenja različna, ločimo male vzorce od velikih. Stroge meje med malimi in velikimi vzorci ni. Običajno imamo vzorce, ki imajo nekaj deset enot, za male, če obseg vzorca ne prekorači sto enot. Število enot v velikih vzorcih je običajno znatno večje. Včasih znaša tudi nekaj tisoč ali celo deset tisoč enot.

Medtem ko uporabljamo male vzorce predvsem v raziskovalnem delu pri preskušanju domnev, z velikimi vzorci običajno nadomeščamo popolna opazovanja.

Glede na način izbora ločimo vzorce s ponavljanjem od vzorcev brez ponavljanja. Ta razlika izvira iz načina, kako izbiramo enote. Pri vzorčenju s po-

navljanjem moremo enoto, ki smo jo enkrat izbrali v vzorec, izbrati ponovno. Glede na sistem izbiranja enot ima izbrana enota to možnost. Pri vzorcu brez ponavljanja pa vsako izbrano enoto izključimo iz nadaljnjega izbora. Ta razlika je teoretično važna. Praktično pa se različen sistem izbora pokaže le, če je vzorec razmeroma velik del osnovne populacije. Teoretično obravnavanje vzorcev brez ponavljanja je nekoliko bolj zamotano kakor obravnavanje vzorcev s ponavljanjem. Ta razlika pa ne pride v poštev pri hipotetičnih in zveznih populacijah in se njen pomen tudi pri končnih populacijah tem bolj izgublja, čim večja je osnovna populacija. Običajno v praksi za končne populacije jemljemo vzorce brez ponavljanja, ker so bolj logični, razen tega pa so ocene vzorcev brez ponavljanja zanesljivejše kakor ocene iz vzorcev s ponavljanjem.

14.7 Populacija vseh vzorcev. Vzorec, ki smo ga po postopku slučajnega izbora izbrali iz osnovne populacije, je samo eden izmed vseh možnih vzorcev, ki jih moremo dobiti iz osnovne populacije, če enote vzorca izbiramo po istem postopku. Če izbiranje slučajnega vzorca ponovimo, prav gotovo ne bomo izbrali v vzorec vse iste enote, kakor v prvem vzorcu. Če se vprašamo, kolikšno je število vseh možnih različnih vzorcev, ki jih moremo iz osnovne populacije izbrati, spoznamo, da je to število že za razmeroma majhne populacije in vzorce izredno veliko. Za večje populacije in vzorce pa je število vseh možnih vzorcev praktično neomejeno.

Če proučimo skupnost vseh možnih vzorcev, se prepričamo, da sestavljajo vsi možni vzorci populacijo v statističnem smislu. Vsak vzorec moremo imeti za enoto populacije vseh možnih vzorcev. Ta populacija ima vse lastnosti statističnih populacij. Sestavljena je iz istovrstnih enot - vzorcev, ki so izbrani v enakih pogojih. Enote te populacije - vzorci imajo svoje znake, ki so v tem primeru poprečja, proporci, mere variacije in korelacije itd., ki jih izračunamo iz podatkov za posamezni vzorec. Ne da bi navajali določene primere, sklepamo, da ti znaki variirajo, ker so izračunani za vsak vzorec iz podatkov za posamezni vzorec. Skupnost vseh vzorcev je torej populacija, v kateri so posamezni vzorci enote, poprečja, proporci, mere variacije in korelacije itd., pa znaki teh enot.

14.8 Kakor smo že navedli, je število vseh možnih vzorcev za dano populacijo razmeroma veliko že za ne preveč velike osnovne populacije in vzorce. Če poznamo število enot v osnovni populaciji N in število enot v vzorcu n moremo število vseh možnih vzorcev izračunati. Število vseh možnih vzorcev brez ponavljanja je enako številu vseh možnih kombinacij brez ponavljanja po n enot iz N enot. To moremo računsko izraziti s simbolom

$$\binom{N}{n} = \frac{N(N-1)\ldots(N-n+2)(N-n+1)}{1\cdot 2 \quad \ldots (n-1)\cdot n} \qquad (14.1)$$

Število vseh možnih vzorcev s ponavljanjem pa je enako številu kombinacij s ponavljanjem po n enot iz N enot. To število pa je

$$\binom{N+n-1}{n} = \frac{N\cdot(N+1)\ldots(N+n-2)(N+n-1)}{1\cdot 2 \qquad\qquad (n-1)\cdot n} \qquad (14.2)$$

Če vzamemo za zgled razmeroma majhno osnovno populacijo z N = 75 enotami in vzorec z n = 25 enotami, pokaže račun, da je število vseh možnih vzorcev brez ponavljanja 52,59 trilijonov, število vseh možnih vzorcev s ponavljanjem pa 181890 trilijonov. Primer pokaže, da je že za tako majhne populacije in vzorce obseg populacije vseh možnih vzorcev velikanski in veliko večji za vzorce s ponavljanjem kakor za vzorce brez ponavljanja.

Podajamo pregled osnovnih značilnosti za osnovno populacijo (OP) vzorec (V) in populacijo vseh vzorcev (PVMV):

Populacija	Število enot - N		Enota	Znak	Zgledi za populacijo	Parameter
	končen	neskončen				
OP	N	∞	gospodarstvo, časovni moment itd.	y, A	vsa gospodarstva, časovni razmik	M_y, σ_y^2, P_A
V	n	n	gospodarstvo, časovni moment itd.	y, A	vzorec	$\bar{y}, s_y^2, P_A$
PVMV	$\binom{N}{n}$ $\binom{N+n-1}{n}$	∞	vzorec	$\bar{y}, s_y^2, P_A$	skupnost vseh možnih vzorcev	$E\bar{y}, E\, s_y^2, E\, P_A$, $Var\, \bar{y}, Var\, P_A$

Zakonitosti v populaciji vseh vzorcev

14.9 Ker je število vseh možnih vzorcev v praktičnih primerih izredno veliko, ne moremo neposredno proučevati populacije vseh možnih vzorcev, razen če ima populacija malo enot. Take populacije pa so brez praktičnega pomena. Ne moremo si misliti, da bi npr. frekvenčno porazdelitev vzorčnih poprečij sestavili tako, da bi iz populacije izbrali vse možne vzorce, za vsak vzorec izračunali aritmetično sredino $\bar{y}$, iz vseh dobljenih poprečij pa sestavili frekvenčno porazdelitev. Prav tako bi bilo težko neposredno računati aritmetično sredino poprečij vzorcev, varianco vseh poprečij itd. Po tej poti je torej težko ali pa celo nemogoče proučevati populacije vseh vzorcev.

Vendar s teoretičnim razglabljanjem pridemo do zvez, ki veljajo med osnovno populacijo in populacijo vseh možnih vzorcev. Izkaže se, da moremo za dosti primerov iz podatkov za osnovno populacijo sklepati, kakšne so porazdelitve in parametri v populaciji vseh vzorcev. K uspešnemu odkrivanju zakonitosti zvez med osnovno populacijo in populacijo vseh vzorcev bistveno pripomore verjetnostni račun; ta je, kot smo že omenili, osnova vzorčenja. Te zakonitosti, ki so za nekatere probleme znane natančno, za druge pa samo približno, izkoristimo za praktično uporabo pri ocenjevanju

in sklepanju iz vzorca na osnovno populacijo.

Za posamezne statistične količine so zakonitosti več ali manj različne in odvisne od podatka, ki ga proučujemo, od osnovne populacije in od tipa vzorca. Zato bomo za aritmetično sredino, ki je eden izmed najvažnejših in najpogostejših statističnih parametrov, razložili splošna načela vzorčenja, v nadaljevanju pa posebej obdelali še druge parametre in probleme, ki so važni v socialno-ekonomskih proučevanjih.

Enostavno slučajnostno vzorčenje

Ocenjevanje aritmetične sredine

14.10 Vzorčenje, pri katerem ima vsaka enota populacije enako verjetnost, da je izbrana v vzorec, imenujemo enostavno slučajnostno vzorčenje.

Za enostavne slučajnostne vzorce, izbrane iz neomejene populacije, in za enostavne slučajnostne vzorce, ki so izbrani iz končnih populacij z izborom s ponavljanjem, veljajo za aritmetične sredine vzorcev $\bar{y}$ naslednje zakonitosti:

a) Aritmetična sredina iz poprečij $\bar{y}$ za vse možne vzorce je enaka aritmetični sredini My v osnovni populaciji. V skladu s pojmi iz verjetnostnega računa se ta zakonitost glasi tudi takole: Matematično upanje za slučajnostno spremenljivko $\bar{y}$ slučajnostnih vzorcev je enako aritmetični sredini v populaciji My

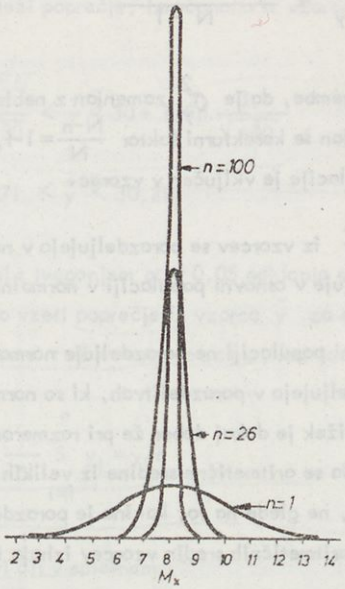
$$M(\bar{y}) = E(\bar{y}) = My \tag{14.3}$$

b) Varianca aritmetičnih sredin slučajnostnih vzorcev s ponavljanjem $Var(\bar{y})$ je n krat manjša kot varianca znaka y v osnovni populaciji

$$Var\,\bar{y} = \frac{\sigma_y^2}{n} \tag{14.4}$$

Če iz variance aritmetičnih sredin vzorcev izračunamo kvadratni koren, dobimo standardni odklon za aritmetične sredine vzorcev. Standardni odklon za vzorčne ocene parametra na splošno imenujemo **standardni pogrešek ocene**, zaznamujemo pa ga s SE . Za ocene aritmetične sredine velja

$$SE(\bar{y}) = \sqrt{Var\ \bar{y}} = \frac{\sigma_y}{\sqrt{n}} \qquad (14.5)$$



Slika 14.1 Porazdelitev aritmetičnih sredin vseh vzorcev za h = 25 in n = 100 enot in za osnovno populacijo n = 1.

c) V praksi običajno proučujemo končne populacije z vzorci brez ponavljanja. Zanje pa veljajo za aritmetične sredine vseh vzorcev nekoliko drugačne zakonitosti.

Matematično upanje za aritmetične sredine vzorcev brez ponavljanja je še vedno enako aritmetični sredini v osnovni populaciji:

$$M(\bar{y}) = E(\bar{y}) = My \qquad (14.6)$$

varianca za aritmetične sredine vzorcev brez ponavljanja pa je enaka

$$Var(\bar{y}) = \frac{S_y^2}{n} \cdot \frac{N-n}{N} = \frac{S_y^2}{n} \cdot (1-f) \qquad (14.7)$$

Pri tem je S_y^2 srednji kvadratični odklon za znak y, ki je izračunan po obrazcu

$$S_y^2 = \frac{\sum_{i=1}^{N} (y_i - M_y)^2}{N - 1} \qquad (14.8)$$

Razen neznatne spremembe, da je σ_y^2 zamenjan z nebistveno različnim izrazom S_y^2, je v obrazcu dodan še korekturni faktor $\frac{N-n}{N} = 1-f$, ki je odvisen od tega, koliki del osnovne populacije je vključen v vzorec.

d) Aritmetične sredine $\bar{y}$ iz vzorcev se porazdeljujejo v normalni porazdelitvi, če se znak y_i porazdeljuje v osnovni populaciji v normalni porazdelitvi.

Če se pa y_i v osnovni populaciji ne porazdeljuje normalno, se aritmetične sredine $\bar{y}$ kljub temu porazdeljujejo v porazdelitvah, ki so normalni tem bolj podobne, čim večji je vzorec. Ta približek je dokaj dober že pri razmeroma majhnih vzorcih, tako da praktično vzamemo, da se aritmetične sredine iz velikih vzorcev porazdeljujejo v normalni porazdelitvi, ne glede na to, kakšna je porazdelitev znaka v osnovni populaciji. Ta zakonitost aritmetičnih sredin vzorcev izhaja iz centralnega limitnega teorema.

Iz zakonitosti od a) - d) povzemamo, da poznamo porazdelitev povprečij $\bar{y}$ v populaciji vseh možnih vzorcev, če poznamo parametre M, σ^2 in S^2 v osnovni populaciji. Tako dobimo posredno porazdelitev poprečij $\bar{y}$, ki je za teorijo in prakso vzorčenja osnovega pomena.

14.11 Točkovna ocena. Odklon zaupanja. Intervalna ali razmična ocena. Razmik zaupanja. Glede na zakonitosti v populaciji sredin vzorcev in glede na lastnosti normalne porazdelitve v zvezi s stopnjo tveganja, na splošno postavimo trditev, da je aritmetična sredina za enostaven slučajnostni vzorec $\bar{y}$ s tveganjem $\alpha = 0,05$ v razmiku

$$My - 1,96\, SE(\bar{y}) < \bar{y} < My + 1.96\, SE\ (\bar{y}) \qquad (14.9)$$

Aritmetične sredine iz vzorcev se tem bolj goste okrog prave aritmetične sredine, čim

manjši je standardni pogrešek sredine. Ta pa je tem manjši, čim večji je vzorec.

Če iz populacije, za katero je My = 30 in $\sigma = 3$ izberemo enostaven slučajnostni vzorec z n = 400 enotami, leži poprečje, izračunano iz vzorca, s tveganjem $\alpha = 0,05$, v razmiku

$$30 - 1,96 \frac{3}{\sqrt{400}} < \bar{y} < 30 + 1.96 \frac{3}{\sqrt{400}}$$

$$29,71 < \bar{y} < 30,29$$

Poprečje iz vzorca se torej s tveganjem $\alpha = 0,05$ odklanja od pravega poprečja za največ 0,29. Zato moremo vzeti poprečje iz vzorca $\bar{y}$ za oceno pravega poprečja My. To oceno $\bar{y}$ imenujemo točkovno oceno, ker je dana z eno samo vrednostjo.

$$\bar{y}_{sl} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = y/n \qquad (14.10)$$

Količine $D(\bar{y}) = 1,96\, SE(\bar{y})$ ali v splošnem

$$D_{\alpha=2P}(\bar{y}) = z_{\alpha=2P} \cdot SE(\bar{y}) \qquad (14.11)$$

imenujemo odklon zaupanja. Ta pove, za koliko se z določenim tveganjem točkovna ocena za aritmetično sredino največ odklanja od prave sredine za populacijo. V našem primeru je s tveganjem $\alpha = 0,05$ odklon zaupanja $D_{.05}(\bar{y}) = 0,29$. S tveganjem $\alpha = 0,05$ ocena z vzorcem ni različna od prave vrednosti za več kot $D_{.05}(\bar{y}) = 0,29$.

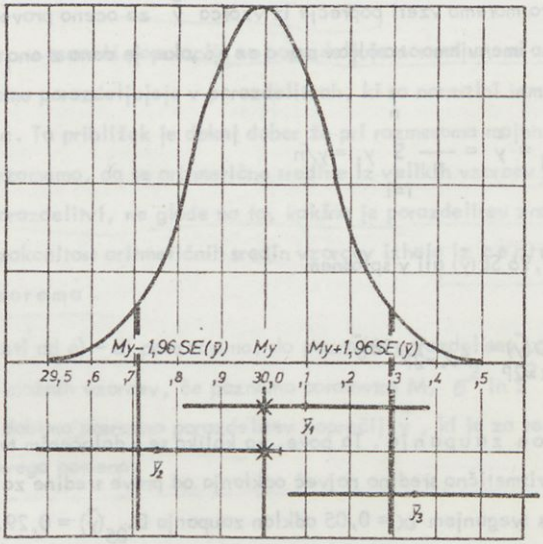
Če s tveganjem α ocena iz vzorca $\bar{y}$ ni od prave vrednosti različna za več kot $D(\bar{y})$, tudi prava vrednost poprečja My ni od ocene $\bar{y}$ s tveganjem α različna za več kot $D(\bar{y})$. Iz tega sledi, da je slučajnostno izbrani vzorec s tveganjem $\alpha = 0,05$ tak, da je prava sredina v razmiku

$$\bar{y} - 1.96 \cdot SE(\bar{y}) < My < \bar{y} + 1.96 \cdot SE(\bar{y}) \qquad (14.12)$$

To razmik imenujemo razmik zaupanja, obe meji pa spodnjo in zgornjo mejo zaupanja ocene. To oceno prave vrednosti poprečja imenujemo intervalno oceno, ker je dana z razmikom zaupanja, v katerem je z določenim tveganjem pravo poprečje.

V sliki 14.2 je nakazana porazdelitev sredin vzorcev za gornji zgled.

Iz slike vidimo, da za vzorce, za katere leže sredine v razmiku od 29,71 do 30,29 · (glej $\bar{y}_1$ in $\bar{y}_2$), prava vrednost My leži v razmikih zaupanja okrog ocen $\bar{y}$. Za ocene, ki leže izven tega razmika, pa razmik zaupanja ne vključuje prave vrednosti My (glej $\bar{y}_3$).



Slika 14.2 Porazdelitev sredin vzorcev s ponavljanjem z n = 400 enotami iz populacije premerov z My = 30 in $\sigma = 3$

14.12 Nepristranska ocena variance in poprečnega kvadratičnega odklona. Vendar obrazec 14.12 za praktično ocenjevanje poprečja z razmikom zaupanja nima posebnega pomena. Da bi izračunali odklon zaupanja, moramo poznati parameter σ_y^2 oziroma S_y^2 v osnovni populaciji. Ker pa teh parametrov ne poznamo, si pomagamo z oceno varianc iz podatkov v izbranem slučajnostnem vzorcu.

Varianco σ_y^2 iz obrazca za vzorce s ponavljanjem in poprečni kvadratičen odklon S_y^2 iz obrazca za vzorce brez ponavljanja ocenimo z enotnim obrazcem

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - y^2/n}{n-1} \qquad (14.13)$$

Ta ocena je nepristranska ocena za varianco σ_y^2 za vzorce s ponavljanjem oziroma nepristranska ocena poprečnega kvadratičnega odklona S_y^2 za vzorce brez ponavljanja, ker je za vzorec s ponavljanjem

$$E(s_y^2) = \sigma_y^2 \qquad (14.14a)$$

in za vzorec brez ponavljanja

$$E(s_y^2) = S_y^2 \qquad (14.14b)$$

Dan izraz, ki ga izračunamo iz vzorca, imenujemo namreč nepristransko oceno nekega parametra iz osnovne populacije, če je matematično upanje tega izraza enako parametru.

Zato je tudi aritmetična sredina, izračunana iz vzorca, zaradi obrazcev 14.3 in 14.6 nepristranska ocena aritmetične sredine v osnovni populaciji.

Če v obrazcih 14.4 in 14.7 nadomestimo prave vrednosti za σ_y^2 in S_y^2 z oceno s_y^2, dobimo oceni varianc za oceno poprečja.

Za vzorce s ponavljanjem je ocena variance za aritmetično sredino enaka

$$\mathrm{var}(\bar{y}) = \frac{s_y^2}{n} \qquad (14.15)$$

za vzorce brez ponavljanja pa

$$\mathrm{var}(\bar{y}) = \frac{s_y^2}{n} \cdot \frac{N-n}{N} = \frac{s_y^2}{n} \cdot (1-f) \qquad (14.16)$$

14.13 Preskus zakonitosti vzorčenja na shematični populaciji.

Gornje zakonitosti za aritmetično sredino preskusimo na shematični populaciji z $N = 5$ enotami, iz katere izbiramo vzorce z $n = 3$ enotami! Osnovni podatki za to populacijo so: $y_1 = 1$, $y_2 = 2$, $y_3 = 3$, $y_4 = 4$, $y_5 = 5$. Zanjo je:

$$M_y = \frac{1}{5}(1 + 2 + 3 + 4 + 5) = 3$$

$$\sigma_y^2 = \frac{1}{5}\left[(1 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 + (5 - 3)^2\right] = 2$$

$$s_y^2 = \frac{N}{N-1}\sigma^2 = \frac{5}{5-1} \cdot 2 = \frac{5}{2}$$

Najprej proučimo populacijo vzorcev s ponavljanjem in sestavimo sliko vseh možnih vzorcev s ponavljanjem s tremi enotami. Glede na obrazec 14.2 je število vseh možnih vzorcev s ponavljanjem enako številu kombinacij s tremi elementi s ponavljanjem iz populacije petih elementov, torej:

$$\binom{N + n - 1}{n} = \binom{5 + 3 - 1}{3} = \frac{5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3} = 35$$

Vseh teh 35 različnih vzorcev pa se ne ponavlja z enako verjetnostjo. Medtem ko je samo ena možnost, da izberemo vzorec, v katerem v prvem, drugem in tretjem izvlačenju izberemo prvo enoto (111), so npr. tri različne možnosti, da izberemo vzorec, v katerem sta dvakrat prva, enkrat pa druga enota (112 121 211) in šest različnih možnosti za vzorec, v katerem je prva, druga in tretja enota (123 132 213 231 312 321). Ker je vsaka izmed teh permutacij enakomožna, vseh možnosti pa je v celoti 125, je verjetnost, da izberemo v vzorec trikrat prvo enoto 1/125, da izberemo dvakrat prvo in enkrat drugo enoto 3/125 in verjetnost, da izberemo v vzorec prvo, drugo in tretjo enoto, enaka 6/125. Podobno je tudi za druge podatke in kombinacije.

V tabeli 14.1 so nanizani osnovni podatki o vsakem izmed petintridesetih različnih vzorcev s ponavljanjem z ustrezno verjetnostjo P, ocena za aritmetično sredino $\bar{y}$ in ocena variance s^2.

Za vzorec, v katerega smo izbrali enote 1 1 3, je npr.

$$\bar{y} = \frac{1}{3}(1 + 1 + 3) = 5/3 \; ; \; s^2 = \frac{1^2 + 1^2 + 3^2 - 5^2/3}{3 - 1} = 8/6 = 4/3$$

Iz gornjih podatkov za posamezne vzorce sestavimo najprej verjetnostno porazdelitev za aritmetično sredino vzorcev $\bar{y}$ kot slučajnostno spremenljivko. To verjetnostno porazdelitev dobimo tako, da seštejemo verjetnosti za vse vzorce, ki imajo enako aritmetično sredino. Tako dobimo tabelo 14.2.

Iz dobljenih rezultatov spoznamo, da je res $E(\bar{y}) = 3 = My$.

Če pa računamo iz podatkov za populacijo po obrazcu 14.4 še

$$\text{Var}(\bar{y}) = \frac{\sigma^2}{n} = \frac{2}{3}$$

spoznamo, da je tudi ta rezultat skladen z rezultatom, ki ga dobimo neposredno iz verjetnostne porazdelitve. Ostane še preskus stavka, da je ocena variance iz vzorca s^2 nepristranska ocena za varianco v osnovni populaciji σ^2. Če podobno kot za poprečja iz vzorcev sestavimo verjetnostno porazdelitev še za oceno variance s^2, dobimo tabelo 14.3.

Rezultat pokaže skladnost s teorijo. Matematično upanje za oceno variance je enako varianci v osnovni populaciji:

$$E(s^2) = \sigma^2 = 2.$$

Tabela 14.1 Populacija vseh možnih vzorcev z n = 3 enotami s ponavljanjem iz populacije z N = 5 enotami. Osnovna populacija: 1 2 3 4 5

Osnovni podatki za vzorce	Pr	$\bar{y}$	s^2	Osnovni podatki za vzorce	Pr	$\bar{y}$	s^2
1 1 1	1/125	3/3	0	2 3 3	3/125	8/3	2/6
1 1 2	3/125	4/3	2/6	2 3 4	6/125	9/3	6/6
1 1 3	3/125	5/3	8/6	2 3 5	6/125	10/3	14/6
1 1 4	3/125	6/3	18/6	2 4 4	3/125	10/3	8/6
1 1 5	3/125	7/3	32/6	2 4 5	6/125	11/3	14/6
1 2 2	3/125	5/3	2/6	2 5 5	3/125	12/3	18/6
1 2 3	6/125	6/3	6/6	3 3 3	1/125	9/3	0
1 2 4	6/125	7/3	14/6	3 3 4	3/125	10/3	2/6
1 2 5	6/125	8/3	26/6	3 3 5	3/125	11/3	8/6
1 3 3	3/125	7/3	8/6	3 4 4	3/125	11/3	2/6
1 3 4	6/125	8/3	14/6	3 4 5	6/125	12/3	6/6
1 3 5	6/125	9/3	24/6	3 5 5	3/125	13/3	8/6
1 4 4	3/125	9/3	18/6	4 4 4	1/125	12/3	0
1 4 5	6/125	10/3	26/6	4 4 5	3/125	13/3	2/6
1 5 5	3/125	11/3	32/6	4 5 5	3/125	14/3	2/6
2 2 2	1/125	6/3	0	5 5 5	1/125	15/3	0
2 2 3	3/125	7/3	2/6				
2 2 4	3/125	8/3	8/6				
2 2 5	3/125	9/3	18/6				

Tabela 14.2 Verjetnostna porazdelitev za aritmetično sredino vzorcev iz populacije $y = 1, 2, 3, 4, 5$ za vzorce z $n = 3$ s ponavljanjem

$\bar{y}$	$Pr(\bar{y})$	$Pr(\bar{y}).\bar{y}$	$\bar{y}-E(\bar{y})$	$Pr(\bar{y}).(\bar{y}-E\bar{y})^2$
3/3	1/125	3/375	-6/3	36/1125
4/3	3/125	12/375	-5/3	75/1125
5/3	6/125	30/375	-4/3	96/1125
6/3	10/125	60/375	-3/3	90/1125
7/3	15/125	105/375	-2/3	60/1125
8/3	18/125	144/375	-1/3	18/1125
9/3	19/125	171/375	0	0
10/3	18/125	180/375	+1/3	18/1125
11/3	15/125	165/375	+2/3	60/1125
12/3	10/125	120/375	+3/3	90/1125
13/3	6/125	78/375	+4/3	96/1125
14/3	3/125	42/375	+5/3	75/1125
15/3	1/125	15/375	+6/3	36/1125
	125/125	1125/375		750/1125

$$\frac{125}{125} = 1 = \sum Pr(\bar{y}) \; ; \qquad \frac{1125}{375} = 3 = E(\bar{y}) = \sum Pr(\bar{y}) \cdot \bar{y} \; ;$$

$$\frac{750}{1125} = \frac{2}{3} = Var(\bar{y}) = \sum Pr(\bar{y}) \cdot (y - E\bar{y})^2$$

Tabela 14.3 Verjetnostna porazdelitev za oceno variance s^2 iz vzorcev

s^2	$Pr(s^2)$	$Pr(s^2).s^2$
0	5/125	0
2/6	24/125	48/750
6/6	18/125	108/750
8/6	18/125	144/750
14/6	24/125	336/750
18/6	12/125	216/750
24/6	6/125	144/750
26/6	12/125	312/750
32/6	6/125	192/750
	125/125	1500/750

$$\sum Pr(s^2) = \frac{125}{125} = 1 \; ; \sum Pr(s^2).s^2 = \frac{1500}{750} = 2 = E(s^2)$$

14.14 Napravimo za isto populacijo y (1 2 3 4 5) podoben preskus še za vzorce brez ponavljanja.

Število različnih vzorcev brez ponavljanja z n = 3 enotami iz populacije z N=5 enotami je po obrazcu 14.1 enako $\binom{N}{n} = \binom{5}{3} = \frac{5.4.3}{1.2.3} = 10$. Če proučimo, kakšna je verjetnost za izbor posameznega vzorca brez ponavljanja, spoznamo, da moremo v vzorec brez ponavljanja izbrati iste enote na toliko različnih načinov, kolikor je permutacij iz n = 3 elementov, v našem primeru torej na 3! = 1.2.3 = 6 različnih načinov: 123 132 213 231 213 321. Ker je skupno 5.4.3 = 60 različnih možnosti za vse vzorce, je verjetnost za vsak vzorec P = 6/60=1/10. Vsak vzorec je pri vzorčenju brez ponavljanja enako verjeten. Vsi možni vzorci za naš zgled so nakazani v tabeli 14.4.

Tabela 14.4 Vzorci brez ponavljanja iz populacije (1 2 3 4 5)

Osnovni podatki za vzorec	Pr	$\bar{y}$	s^2
1 2 3	1/10	6/3	6/6
1 2 4	1/10	7/3	14/6
1 2 5	1/10	8/3	26/6
1 3 4	1/10	8/3	14/6
1 3 5	1/10	9/3	24/6
1 4 5	1/10	10/3	26/6
2 3 4	1/10	9/3	6/6
2 3 5	1/10	10/3	14/6
2 4 5	1/10	11/3	14/6
3 4 5	1/10	12/3	6/6

Enako kot za vzorec s ponavljanjem sestavimo tudi za vzorce brez ponavljanja verjetnostno porazdelitev za aritmetično sredino vzorcev in izračunajmo matematično upanje in varianco za aritmetično sredino vzorcev.

Tabela 14.5 Verjetnostna porazdelitev sredine $\bar{y}$ vzorcev brez ponavljanja

$\bar{y}$	$Pr(\bar{y})$	$Pr(\bar{y}) \cdot \bar{y}$	$\bar{y}-E(\bar{y})$	$Pr\left[\bar{y}-E(\bar{y})\right]^2$
6/3	1/10	6/30	-3/3	9/90
7/3	1/10	7/30	-2/3	4/90
8/3	2/10	16/30	-1/3	2/90
9/3	2/10	18/30	0	0
10/3	2/10	20/30	+1/3	3/90
11/3	1/10	11/30	+2/3	4/90
12/3	1/10	12/30	+3/3	9/90
	10/10	90/30		30/90

$$\sum Pr(\bar{y}).\bar{y} = 90/30 = 3 = E(\bar{y}) \qquad \sum Pr(\bar{y})\left[\bar{y}-E(\bar{y})\right]^2 = 30/90 = 1/3 = Var(\bar{y})$$

Če izračunamo $E(\bar{y})$ in $Var(\bar{y})$ posredno iz parametrov za osnovno populacijo, dobimo po obrazcih 14.6 in 14.7:

$$E(\bar{y}) = My = 3$$

$$Var(\bar{y}) = \frac{S^2}{n} \cdot \frac{N-n}{N} = \frac{5/2}{3}\,\frac{5-3}{5} = \frac{1}{3}$$

Dobljena rezultata sta v skladu z rezultati, ki jih dobimo neposredno iz verjetnostne porazdelitve aritmetičnih sredin in vzorcev.

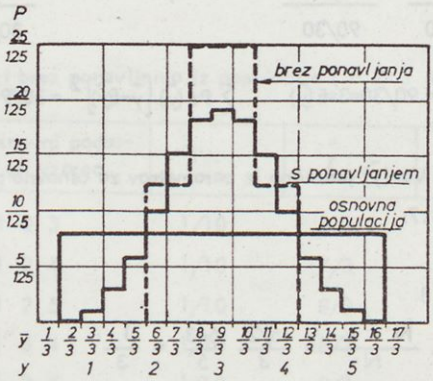
Preskusimo še stavek o nepristranosti ocene poprečnega kvadratičnega odklona s^2. Najprej sestavimo verjetnostno porazdelitev za s^2, tako da grupiramo rezultate iz stolpca za s^2 iz tabele 14.4.

Če matematično upanje za oceno variance s^2 izračunamo posredno iz parametrov za populacijo, dobimo po obrazcu 14.13b $E(s^2) = S^2 = \frac{5}{2}$. Tudi ta rezultata sta skladna.

Tabela 14.6 Verjetnostna porazdelitev varianc s^2 v vzorcih

s^2	$Pr(s^2)$	$Pr(s^2) \cdot s^2$
6/6	3/10	18/60
14/6	4/10	56/60
24/6	1/10	24/60
26/6	2/10	52/60
	10/10	150/60

$$\sum Pr(s^2) = \frac{10}{10} = 1 \qquad \sum Pr(s^2) \cdot s^2 = \frac{150}{60} = \frac{5}{2} = E(s^2)$$



Slika 14.3 Verjetnostni porazdelitvi za poprečje iz vzorcev z in brez ponavljanja

Iz verjetnostne porazdelitve aritmetične sredine vzorcev s ponavljanjem v sliki 14.3 vidimo, da ta teži k normalni porazdelitvi, ki smo jo navedli kot zakonitost. Čeprav porazdelitev osnovnih podatkov ni normalna in gre za izjemno majhne vzorce, se aritmetične sredine vzorcev porazdeljujejo v porazdelitvi, ki je simetrična, zvonasta, torej v porazdelitvi, ki ima iste značilnosti kot normalna. Ta težnja je manj vidna pri verjetnostni porazdelitvi sredine vzorcev brez ponavljanja, ker je vsega le deset možnih različnih vzorcev.

14.15 Ocenjevanje parametrov z enostavnim slučajnostnim vzorčenjem na splošno. Gornji sklepi ne veljajo samo za ocenjevanje aritmetične sredine, ampak za ocenjevanje za katerikoli drug parameter z enostavnimi velikimi vzorci. V približku se namreč za katerikoli parameter G ocene g porazdeljujejo normalno s sredino G in standardnim odklonom SE(g).

Če z G zaznamujemo pravo vrednost ocenjevanega parametra, z g pa točkovno oceno tega parametra iz velikega vzorca, je odklon zaupanja za oceno s tveganjem $\alpha = 2P$

$$D_p(g) = z_p \, SE(g) \tag{14.17}$$

razmik zaupanja pa

$$g - D_p(g) < G < g + D_p(g) \tag{14.18}$$

s tveganjem, ki ustreza koeficientu z. Ker v praksi ne poznamo parametrov populacije, v odklonu zaupanja in razmiku zaupanja nadomestimo pravo vrednost za standardni pogrešek SE(g) z oceno iz vzorca se(g). Glede na to je ocena odklona zaupanja

$$d_p(g) = z_p \cdot se(g) \tag{14.19}$$

ocena za razmik zaupanja pa

$$g - z_p \cdot se(g) < G < g + z_p \cdot se(g) \tag{14.20}$$

Ocenjevanje agregata Y, strukturnega deleža P% in števila enot z dano značilnostjo N_a

14.16 Razen aritmetične sredine pogosto ocenjujemo vsoto vrednosti za dan znak, npr. za skupno proizvodnjo v obrtnih podjetjih, za skupno obdelovalno površino kmetijskih gospodarstev v privatnem sektorju ipd.

Med aritmetično sredino za populacijo My in vsoto ali agregatom $Y = \sum y_i$ je enostavna zveza

$$Y = \sum y_i = N \cdot My \tag{14.21}$$

Oceno za agregat Y dobimo, če v obrazcu 14.21 nadomestimo pravo vrednost My z oceno iz vzorca $\bar{y}$.

Oceno agregata z enostavnim slučajnostnim vzorcem Y_{sl} izračunamo torej po obrazcu

$$Y_{sl} = N\bar{y} = \frac{N}{n} \sum_{i=1}^{n} y_i = \frac{N}{n} y \tag{14.22}$$

pri čemer pomeni: y = vsota podatkov iz enostavnega slučajnostnega vzorca.

Ker je po znanih stavkih o varianci $Var(ax) = a^2 Var(x)$, velja, da je

$$Var\, Y_{sl} = N^2 Var\bar{y} = N(N - n)\frac{S_y^2}{n} \tag{14.23}$$

Ker agregat običajno ocenjujemo z vzorčenjem brez ponavljanja, navajamo samo obrazec za ta tip vzorčenja.

Iz podatkov vzorca pa ocenjujemo varianco agregata $var(Y_{sl})$ po obrazcu, ki je prav tak kot obrazec 14.23, samo da je S_y^2 zamenjan z oceno s_y^2

$$var Y_{sl} = N(N - n)\frac{s_y^2}{n} \tag{14.24}$$

14.17 Pogosto ocenjujemo tudi **strukturni delež** P, ki ga dobimo, če število enot z dano značilnostjo N_a v populaciji delimo z obsegom populacije N

$$P = \frac{N_a}{N} \tag{14.25}$$

Oceno strukturnega deleža p dobimo iz podatkov enostavnega slučajnostnega vzorca podobno tako, da število enot z dano značilnostjo n_a iz vzorca delimo s skupnim številom enot v vzorcu n

$$P_{sl} = p = \frac{n_a}{n} \tag{14.26}$$

Varianca za oceno strukturnega deleža iz neomejene populacije ali iz vzorca s ponavljanjem je enaka

$$Var(p) = \frac{\sigma_a^2}{n} \quad : \quad \sigma_a^2 = P(1 - P) \tag{14.27}$$

Za vzorce brez ponavljanja iz končnih populacij pa je

$$Var(p) = \frac{s_a^2}{n} \frac{N - n}{N} \quad : \quad s_a^2 = \frac{N_a (N - N_a)}{N(N - 1)} \tag{14.28}$$

Podobno je ocena variance za oceno strukturnih deležev vzorcev iz neomejenih populacij in vzorcev s ponavljanjem iz končnih populacij

$$var(p) = \frac{s_a^2}{n} \tag{14.29}$$

za vzorce brez ponavljanja iz končnih populacij pa

$$var(p) = \frac{s_a^2}{n} \frac{N - n}{N} \tag{14.30}$$

Pri tem je izraz

$$s_a^2 = \frac{n_a(n - n_a)}{n\,(n - 1)} \tag{14.31}$$

analogen poprečnemu kvadratnemu odklonu S_a^2, le da je izračunan iz podatkov vzorca.

Opomba: Če strukturne deleže izražamo v odstotkih, je P% = 100.P in p% = 100.p. Ker je Var (p%) = Var(100.p) = 100^2Var(p), moramo v tem primeru vse obrazce za računanje variance pomnožiti s 100^2, če ocenjujemo strukturne deleže v odstotkih.

14.18 Število enot z dano značilnostjo N_a v populaciji dobimo, če

$$N_a = N\frac{N_a}{N} = NP \qquad (14.32)$$

strukturni delež P pomnožimo s skupnim številom enot v populaciji N. Analogno kot pri oceni za vsoto, dobimo oceno skupnega števila enot z dano značilnostjo $N_{a,st}$, če v obrazcu 14.32 pravo vrednost strukturnega deleža zamenjamo z oceno p

$$N_{a,st} = Np = \frac{N}{n} n_a \qquad (14.33)$$

Podobno kot pri agregatih je varianca za oceno skupnega števila enot z dano značilnostjo za vzorce brez ponavljanja enaka

$$Var(N_{a,st}) = N(N-n)\frac{S_a^2}{n} \; ; \quad S_a^2 = \frac{N_a(N-N_a)}{N(N-1)} \qquad (14.34)$$

$$var(N_{a,st}) = N(N-n)\frac{s_a^2}{n} \; ; \; s_a^2 = \frac{n_a(n-n_a)}{n(n-1)} \qquad (14.35)$$

14.19 Simbolika in obrazci za ocene parametrov My, Y, P in N_a so sistematično nakazani v naslednjem pregledu:

	Populacija	Vzorec
Znak za seštevanje	Σ	S
Število enot	N	n
Število enot z dano značilnostjo	N_a	n_a
Strukturni delež	P	p
Strukturni odstotek	P%	p%
Posamična vrednost	Y_i	y_i
Vsota za znak y	Y	y
Aritmetična sredina	My	$\bar{y}$
Poprečni kvadratni odklon	S_y^2	s_y^2
Varianca	σ_y^2	s_y^2

	Populacija	Vzorec
Poprečni kvadratni odklon za strukturni delež	S_a^2	s_a^2
Varianca strukturnega deleža	σ_a^2	s_a^2

	Prava vrednost	Ocena
Varianca	Var	var
Standardni pogrešek	SE	se
Odklon zaupanja	D	d

Parametri: Aritmetična sredina $My = \frac{1}{N} Y$; $\bar{y} = \frac{1}{n} Sy_i$

Agregat $Y = Y_i$ $Y_{sl} = \frac{N}{n} y$

strukturni delež $P = N_a/N$ $p = n_a/n$

število enot z dano značilnostjo N_a $Na_{sl} = \frac{N}{n} n_a$

Varianca ocen parametrov z vzorcem s ponavljanjem

$$Var(\bar{y}) = \frac{\sigma_y^2}{n} \qquad var(\bar{y}) = \frac{s_y^2}{n}$$

$$Var(p) = \frac{\sigma_a^2}{n} \qquad var(p) = \frac{s_a^2}{n}$$

Varianca ocen parametrov z vzorcem brez ponavljanja

$$Var(\bar{y}) = \frac{S_y^2}{n} \cdot \frac{N-n}{N} \qquad var(\bar{y}) = \frac{s_y^2}{n} \cdot \frac{N-n}{N}$$

$$Var(Y_{sl}) = N(N-n)\frac{S_y^2}{n} \qquad var(Y_{sl}) = N(N-n)\frac{s_y^2}{n}$$

$$Var(p) = \frac{S_a^2}{n} \cdot \frac{N-n}{N} \qquad var(p) = \frac{s_a^2}{n} \cdot \frac{N-n}{N}$$

$$Var(N_{a_{sl}}) = N(N-n)\frac{S_a^2}{n} \qquad var(N_{a,sl}) = N(N-n)\frac{s_a^2}{n}$$

Tehnika izbora enostavnega slučajnostnega vzorca

14.20 Prvi in glavni problem pri izvedbi enostavnega slučajnostnega vzorčenja je, kako izbrati slučajnostni vzorec, ki bi zadostil osnovnemu pogoju, da ima vsaka enota v populaciji enako možnost, da je vključena v enostavni vzorec.

Pri vsakem sistemu vzorčenja ima posamezna enota populacije določeno vnaprej dano verjetnost, da je vključena v vzorec. Način izbora enot mora biti tak, da zadosti te-

mu osnovnemu pogoju, ker le v tem primeru veljajo zakonitosti vzorčenja, ki so izpeljane iz verjetnostnega računa. Pri enostavnem vzorčenju ima vsaka enota populacije enako verjetnost oziroma možnost, da je vključena v vzorec. Izbor takega vzorca imenujemo izbor brez omejitve.

14.21 Okvir vzorčenja. Tehnično izberemo enostaven slučajnostni vzorec po naslednjem postopku. Osnova slučajnostnega izbora je v vsakem primeru razvid vseh enot v populaciji. V tem razvidu, ki ga imenujemo okvir vzorčenja, je vsaka enota obvezno tako označena, da jo moremo identificirati. Razvid enot populacije, ki sestavlja okvir vzorčenja, more biti podan različno. Navadno je to spisek enot, ki so oštevilčene z zaporednimi številkami. Zelo prikladen okvir je tudi oštevilčena kartoteka enot populacije. Tudi geografska karta, v katero so vnešene enote populacije glede na geografsko lego in zaznamovane z zaporednimi številkami, more biti nazoren okvir populacije. Če so enote geografska območja (občine, katastrske občine, statistični okoliši), je geografska karta teh območij, v kateri je vsako območje označeno z zaporedno številko, tudi primeren okvir vzorčenja. Kot okvir vzorčenja moremo uporabiti tudi vsako drugo shemo oziroma način, iz katerega je možno identificirati enote osnovne populacije.

14.22 Loterijski način. Eden izmed načinov slučajnostnega izbora je loterijski način. Pri loterijskem načinu imamo v žari listke z zaporednimi številkami enot. V žari je torej toliko listkov, kolikor je enot v populaciji.

Enako možnost za izbor vsake enote ustvarimo tako, da oštevilčene listke dobro premešamo in iz žare "na slepo" potegnemo listek. Enoto, ki ima izžrebano zaporedno številko, vzamemo kot enoto slučajnostnega vzorca. Postopek žrebanja enot vzorca ponavljamo vse dotlej, dokler nimamo izžrebanega ustreznega števila enot n. Pri tem posamezne izžrebane listke pred izborom nove enote vračamo v žaro, če gre za izbor s ponavljanjem. Tako dosežemo, da ima ista izbrana enota možnost, da je ponovno vključena v izbor. Če pa izbiramo vzorec brez ponavljanja, izžrebani listek izločimo, da ga ne moremo ponovno izbrati.

Loterijski način slučajnostnega izbora pa je posebno za večje populacije zamuden in okorel.

14.23 Isti cilj dosežemo na drug način, brez listkov za vsako enoto.

Vzemimo žaro, v kateri je deset enakih listkov, kroglic ali kock s številkami od 0 do 9. Če imamo populacijo z N = 999 enotami, dobimo slučajnostno zaporedno številko tako, da s trikratnim žrebanjem sestavimo tromestno slučajnostno številko. Izbiramo seveda s ponavljanjem. Če dobimo s prvim žrebom številko 3, z drugim 1, s tretjim žrebom pa številko 3, ta trojni žreb ustreza izžrebani zaporedni številki 313 po prejšnjem načinu. V vzorec vključimo torej enoto, ki ima v okviru vzorčenja zaporedno številko 313. Prednost tega načina v primerjavi s prvim je v tem, da ni treba sestavljati obširnih žar, pomanjkljiv pa je v tem, da je treba za izbor ene same enote toliko žrebanj, kolikor mestno je število enot v populaciji N. To pa je zamuden posel.

14.24 Tablice slučajnostnih številk. Hibe loterijskega izbora odpravimo s tablicami slučajnostnih številk. Številke od 0 do 9, ki jih enkrat izberemo, zabeležimo. Te slučajnostne številke moremo uporabiti za več različnih vzorcev. Tako dobimo tablice slučajnostnih števil, v katerih so zapisane številke od 0 do 9, kot so bile po vrsti slučajnostno izbrane iz žare z desetimi številkami. En del iz tablic slučajnostnih številk iz knjige: Fisher and Yates: Statistical Tables for Biological, agricultural and medical research je dana v tabelah v dodatku.

Večmestna slučajnostna števila iz teh tablic dobimo, če združimo več slučajnostnih številk v skupine. Vzemimo prvo vrsto iz tablic slučajnostnih številk iz tabele 14.7: 03 47 43 73 86 36 96 47 36 61 46 98 63 71 itd. Zaporedne številke enot, ki jih iz populacije z N = 8324 izberemo v slučajnostni vzorec, določimo tako, da vzamemo po vrsti skupine po štiri slučajnostne številke. Prva izbrana enota ima zaporedno številko 0347. Skupine zaporednih štirimestnih slučajnostnih številk so: 0347 4373 8636 9647 3661 4698 6371.

Prva izbrana enota ima torej zaporedno številko 347 (0 na prvem mestu odpade). Naslednja enota ima zaporedno številko 4373. Slučajnostni številki 8636 in 9647 ne prideta v poštev, ker sta večji kot je obseg populacije in ima zadnja enota v populaciji zaporedno številko 8324. Pač pa vključimo v vzorec še enote z zaporednimi številkami 3661, 4698 in 6371.

Skupine slučajnostnih zaporednih številk pa se morejo med seboj tudi delno prekrivati. Tako znatno povečamo obseg tablic slučajnostnih številk. Če za gornji primer tvorimo štirimestne skupine slučajnostnih številk tako, da slučajnostne številke premikamo po eno mesto, dobimo tele rezultate: Prva skupina 0347 ostane. Naslednjo skupino pa dobimo, če se v vrsti slučajnostnih številk premaknemo le za eno mesto. Druga skupina je torej: 3474. Naprej dobimo po istem postopku nadaljnje skupine: 4743 7437 · 4373 3738 itd. Namesto sedmih skupin po štiri ločene slučajnostne številke, izmed katerih jih moremo le pet uporabiti kot zaporedne številke izbranih enot, dobimo 25 skupin, od katerih je 21 manjših kot 8324. Namesto 5 enot iz istega dela tablic slučajnostnih številk izberemo 21 enot.

Po potrebi obseg tablic slučajnostnih številk povečamo tudi tako, da slučajnostne številke v tablici beremo v obratni smeri, od zgoraj navzdol ali od spodaj navzgor. V vsakem primeru dobimo nove skupine slučajnostnih številk.

14.25 V navedenem zgledu nekatere skupine štirimestnih slučajnostnih številk niso uporabne, ker presegajo obseg populacije. Število skupin, ki zaradi tega odpadejo, more biti zelo veliko. Če je npr. obseg populacije N = 1832, od 10000 različnih štirimestnih skupin odpade 10000 - 1832 = 8168 slučajnostnih števil. To pa zelo okrni uporabnost tablic.

S preprostim postopkom obseg uporabnih skupin povečamo. Od vseh štirimestnih skupin pridejo po zgornjem v konkurenco le one, ki imajo prvo številko 0 ali 1. Vse druge izpadejo. Če pa postavimo pravilo, da štejemo kot 0 na prvem mestu štirimestnega števila vsako sodo /0, 2, 4, 6, 8/, kot 1 pa vsako liho /1, 3, 5, 7, 9/ slučajnostno številko, se uporabnost tablic znatno poveča. Namesto 8168 neuporabnih skupin se število neuporabnih skupin od 8168 zniža na 840. Ker je npr. prva številka v skupini 5674 liha, ustreza tej slučajnostni številki enota z zaporedno številko 1674. Po prvem načinu pa je to slučajnostno število neuporabno.

Podobno reduciramo prva mesta skupin slučajnostnih številk za populacije, katerih obseg N se začenja s številko 2. Številke 0, 3, 6 štejemo kot 0, slučajnostne številke 1, 4, 7 kot 1, slučajnostne številke 2, 5, 8 pa kot 2. Po tem sistemu je ustrezna prva

številka ostanek slučajnostne številke, če prvo slučajnostno številko od 0 do 8 delimo s tri. Opozoriti moramo, da slučajnostne skupine, ki se začenjajo z 9, izpuščamo.

Analogno je pri populacijah, za katere se obseg začne s številko 3, odločilen ostanek, če delimo prvo slučajnostno številko s štiri, ne upoštevamo pa 8 in 9.

Če pa se število enot N začne s 4, je prva številka zaporednega števila ostanek, če delimo prvo slučajnostno številko s 5.

14.26 Uporabne slučajnostne številke moremo enostavno dobiti tudi z računalnikom.

Če vzamemo poljuben večmesten decimalen ulomek npr. S_o in operativno konstanto C, dobimo S_1 po naslednjem obrazcu:

$$S_1 = C.S_o \text{ /modulo 1/} \qquad (14.36)$$

tako, da od produkta $C.S_o$ vzamemo samo decimalni ostanek. Zaporedje prvih štirih ali petih številk tega ostanka predstavlja pet slučajnostnih številk R_1. Če postopek ponovimo z S_1, dobimo na enak način

$$S_2 = C.S_1 \text{ /modulo 1/} \qquad (14.37)$$

in naslednjih 5 slučajnostnih številk R_2. Če ta postopek ciklično ponavljamo, dobimo iz S_i skupino slučajnostnih številk R_{i+1} tako, da od

$$S_{i+1} = C.S_i \text{ /modulo 1/} \qquad (14.38)$$

vzamemo kot R_{i+1} prvih pet številk iz decimalnega ulomka S_{i+1}.

Za zgled vzemimo, da je C = 10011 in S_o = ·53952704. Če postopamo po gornjem pravilu, je

S_1 = 10011.0,53952704/mod 1/=5401,20519744/mod 1/ = ·20519744
in R_1 = 20519

S_2 = 10011.0,20519744/mod 1/= 2054,23157184/mod 1/= ·23157184
in R_2 = 23157

S_3 = 10011.0,23157184/mod 1/= ·26569024 R_3 = 26569

S_4 = 10011.0,26569024/mod 1/= ·82499264 R_4 = 82499

Če potrebujemo slučajnostna števila za populacijo z N enotami, moremo iz slučajnostnih decimalnih ulomkov S_i dobiti ustrezne slučajnostne številke tako, da od produkta /N + 1 /.S_i = P_i upoštevamo le celo število. Če je npr. obseg populacije N = 38540, so gornjim štirim S_i ustrezna slučajnostna števila naslednja:

/38540+1/.0,20519744 = 7908
/38540+1/.0,23157184 = 8925
/38540+1/.0,26569024 =10239
/38540+1/.0,82499264 =31796

14.27 Zgled za ocenjevanje z enostavnim slučajnostnim vzorcem.

Za prikaz različnih vrst vzorčenja, med njimi tudi za enostavno slučajnostno vzorčenje, bomo uporabili shematično preskusno kvadratno ploskev s površino 57,6 ha. Ta je razdeljena na N = 24 x 24 = 576 osnovnih površinic kvadratne oblike po 10 a. Kljub temu, da je zgled shematičen, je populacija zadosti velika, da bomo mogli na različnih primerih oceniti kakovost posameznih vrst vzorčenja. V skici ploskve v sliki 14.4 so nakazane osnovne površinice in vanje vpisani podatki o pridelkih na posameznih površinicah. Te podatke v nadaljevanju uporabljamo za to, da z njimi dobimo podatke o izbranih parcelah. V stvarnosti dobimo za izbrane enote te podatke z opazovanjem na terenu.

Prave vrednosti parametrov, katere bomo v nadaljevanju ocenjevali, so: število parcel, za katere je pridelek večji kot 30: N_a = 320. Odstotek parcel, na katerih je pridelek večji kot 30: P% = 55,6%. Skupen pridelek: Y = 17439. Poprečen pridelek na osnovno parcelo: My = 30,3. Te parametre navajamo zaradi primerjave z ocenami.

23	21	22	20	29	33	32	31	31	27	28	32	36	29	36	30	35	43	36	37	36	37	33	36
19	26	18	22	33	24	27	28	23	34	34	29	31	32	27	38	32	35	34	33	43	34	41	43
23	20	25	21	32	29	30	25	32	28	27	34	28	36	33	30	32	37	36	37	36	41	38	33
22	22	20	22	23	23	24	28	30	24	32	30	27	31	32	29	35	33	35	35	41	35	41	38
21	22	21	20	24	33	28	29	22	32	30	30	35	30	31	32	39	34	34	34	33	34	33	38
23	18	21	22	22	31	20	25	34	26	28	35	29	36	30	31	38	37	35	37	36	35	41	36
22	21	23	21	20	28	26	24	28	24	33	31	33	31	32	39	34	33	37	33	35	37	35	41
23	20	21	23	24	23	25	29	27	30	38	25	31	37	35	32	35	36	34	36	38	34	44	33
20	25	18	25	20	29	32	21	30	26	32	27	28	34	34	39	34	37	36	33	35	37	35	42
22	18	24	22	22	30	23	30	36	25	30	38	32	33	32	36	36	33	34	39	42	35	33	38
23	23	20	23	25	32	27	33	25	26	27	30	37	35	37	33	34	33	35	34	34	34	42	35
24	21	23	18	23	22	23	25	26	24	34	37	33	33	36	35	36	35	39	33	37	35	34	40
20	23	24	22	21	20	22	22	33	34	28	29	34	35	31	36	34	40	33	37	34	36	40	35
22	19	23	25	20	21	24	33	27	27	31	37	35	34	36	35	39	36	36	35	39	40	33	38
23	21	21	23	22	19	21	26	25	31	34	34	33	37	33	34	36	33	37	39	34	34	36	34
22	19	22	20	24	22	20	34	30	34	29	34	35	32	35	35	34	37	34	35	40	35	40	38
20	23	21	21	18	25	28	26	22	27	30	35	36	35	33	39	37	35	36	34	36	34	36	33
23	20	24	22	21	20	27	33	26	26	29	31	33	34	34	34	35	33	37	36	38	40	35	38
22	21	22	19	26	22	33	26	33	30	29	33	37	33	36	39	34	39	34	36	33	34	38	35
24	18	25	21	22	20	34	36	24	27	30	29	28	36	34	24	35	34	37	35	39	38	34	36
22	22	20	22	18	21	22	28	26	24	31	29	35	35	30	34	35	35	34	37	35	33	42	33
24	22	23	21	20	23	24	32	33	31	28	31	29	31	35	35	32	37	36	33	38	38	35	38
20	21	19	26	22	21	19	26	21	24	26	29	34	28	29	37	36	39	34	35	40	35	34	35
23	22	23	21	21	24	22	20	24	21	33	27	29	34	31	30	34	35	34	40	34	45	38	35

Slika 14.4 Skica poskusne ploskve z N = 576 osnovnimi parcelicami s podatki o pridelkih na osnovnih ploskvah

14.28 Z enostavnim slučajnostnim vzorcem brez ponavljanja n = 72 osnovnih parcel na poskusni ploskvi ocenimo:

a) popečen pridelek na osnovno površino: $\bar{y}$,

b) skupen pridelek: Y_{sl} ,

c) odstotek osnovnih parcelic, ki imajo pridelek večji kot 30: p%

d) število parcelic na katerih je pridelek večji kot 30: $N_{a,sl}$

Vzorčna enota je v našem primeru posamezna parcelica izmed N = 576 osnovnih parcelic. Osnovna populacija pa je sestavljena iz vseh N = 576 osnovnih parcelic.

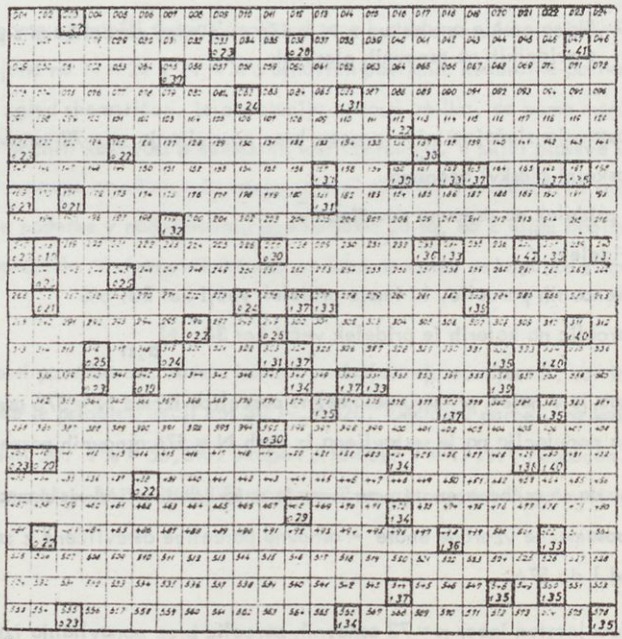
Glede na to, da so osnovne enote površine, je najprikladnejši okvir shematična karta, v kateri so posamezne vzorčne enote - osnovne parcelice oštevilčene z zaporednimi številkami od 1 do 576.

Enostaven slučajnosten izbor n = 72 osnovnih parcelic brez ponavljanja izvedemo ta-

ko, da iz tablic slučajnostnih številk vzamemo kot zaporedne številke izbranih enot vsa tromestna slučajnostna števila od 001 do 576. Tromestne slučajnostne številke sestavimo po določenem pravilu.

V našem zgledu vzemimo po vrstnem redu tromestna števila iz zaporednih stolpcev v tablici slučajnostnih števil .

Najprej vzemimo kot tromestna slučajnostna števila prvo skupino tromestnih številk (prva, druga, tretja) iz vsake vrste v tablici slučajnostnih številk v dodatku, nato drugo skupino (četrte, pete, šeste slučajnostne številke) iz vsake vrste, nato tretjo skupino itd. vse dotlej, dokler ne dobimo n = 72 slučajnostnih tromestnih številk brez ponavljanja. n = 72 tako ugotovljenih tromestnih slučajnostnih številk brez ponavljanja je nanizanih v tabeli 14.7. Iz te tabele je razvidno, da se je v izboru pet tromestnih številk ponovilo /dv/. Izbrane parcele so v okviru vzorca v sliki 14.5 uokvirjene. V tabeli 14.7, kakor tudi v sliki 14.5 so za izbrane parcele navedeni tudi pridelki, katere smo ugotovili iz skice 14.4. Z 0 ali 1 pa je zaznamovano, da pridelek ni /0/ oziroma da je /1/ večji kot 30.



Slika 14.5 Okvir enostavnega slučajnostnega vzorca n = 72 parcelic z vrisanimi slučajnostno izbranimi parcelicami in vpisanimi pridelki

Tabela 14.7 Osnovni podatki o enostavnem slučajnostnem izboru za vzorec brez ponavljanja z n = 72 parcelicami

Sluč. štev.	pridelek		Sluč. štev.	pridelek		Sluč. štev.	pridelek	
	y_i	$y_i > 30$		y_i	$y_i > 30$		y_i	$y_i > 30$
034	28	0	277	33	1	324	dv	dv
167	35	1	217	22	0	137	38	1
125	22	0	163	37	1	296	22	0
555	23	0	112	dv	dv	126	dv	dv
162	33	1	086	31	1	242	23	0
332	40	1	238	35	1	438	22	0
576	35	1	240	38	1	157	33	1
181	31	1	218	19	0	409	23	0
266	21	0	430	40	1	482	22	0
234	33	1	237	42	1	550	35	1
323	31	1	498	36	1	283	39	1
378	37	1	082	24	0	340	23	0
566	34	1	410	20	0	233	36	1
160	39	1	316	25	0	350	37	1
311	40	1	276	37	1	055	30	0
274	24	0	356	39	1	348	dv	dv
003	22	0	342	19	0	166	37	1
299	28	0	324	37	1	468	29	0
169	23	0	199	32	1	033	23	0
112	32	1	351	33	1	502	33	1
352	34	1	121	23	0	548	35	1
382	35	1	373	35	1	544	37	1
319	24	0	227	30	0	245	25	0
424	34	1	047	41	1	472	34	1
						171	21	0
						160	dv	dv
						395	30	0
						332	35	1
						429	38	1

Iz posamičnih podatkov za $n = 72$ izbranih parcel dobimo naslednje pomožne rezultate: število parcel, ki imajo pridelek večji kot 30, $n_a = 43$, vsoto pridelkov na izbranih parcelicah $y = S\, y_i = 2226$, vsoto kvadratov pridelkov na izbranih parcelicah $S\, y_i^2 = 71902$.

Če upoštevamo teoretične osnove za enostavni slučajnostni vzorec, dobimo naslednje rezultate:

Točkovna ocena za poprečni pridelek na eno parcelo je po obrazcu 14.10

$$\bar{y} = y/n = 2226/72 = 30,92$$

Iz pomožnih rezultatov dobimo dalje, da je ocena variance za osnovni znak y

$$s_y^2 = \frac{S\, y_i^2 - y^2/n}{n-1} = \frac{71902 - 2226^2/72}{72-1} = 43,40$$

Ocena variance za aritmetično sredino z vzorcem brez ponavljanja je po obrazcu 14.16

$$\text{var}(\bar{y}) = \frac{s_y^2}{n}\,\frac{N-n}{N} = \frac{43,40}{72}\cdot\frac{576-72}{576} = 0,527431$$

$$se(\bar{y}) = \sqrt{\text{var}(\bar{y})} = \sqrt{0,527431} = 0,726$$

$$d(\bar{y}) = z_p . se(\bar{y}) = 1,96 . 0,726 = 1,42$$

$$M_y = 30,92 \pm 1,42$$

Prava vrednost aritmetične sredine pridelka na eno parcelo leži s tveganjem $\alpha = 2P = 0,05$ v razmiku od 29,50 do 32,34.

Ocena aritmetične sredine pridelka torej s tveganjem $\alpha = 0,05$ od prave aritmetične sredine M_y ni različna za več kot

$$d\%(\bar{y}) = 100\, d(\bar{y})/\bar{y} = 100 . 1,45/30,92 = 4,7\ \%$$

Ocena, ki smo jo dobili z vzorcem, je razmeroma dobra.

Podobno ocenimo tudi skupen pridelek. Točkovna ocena skupnega pridelka Y_{sl} je

$$Y_{sl} = \frac{N}{n} y = \frac{576}{72} \cdot 2226 = 17808$$

Ocena variance za skupen pridelek pa je

$$\text{var}(Y_{sl}) = N(N-n)\frac{s_y^2}{n} = 576(576-72)\frac{43,40}{72} = 174988,80$$

$$\text{se}(Y_{sl}) = \sqrt{\text{var}(Y_{sl})} = \sqrt{174988,80} = 418,32$$

$$d(Y_{sl}) = z_p \cdot \text{se}(Y_{sl}) = 1,96.418,32 = 820$$

$$Y = 17808 \pm 820$$

Pravi skupni pridelek sestoja je s tveganjem $\alpha = 0,05$ med 16988 in 18628.

Če primerjamo dobljeni rezultat ocene z enostavnim vzorčenjem brez ponavljanja s pravim pridelkom v raziskovanem sestoju Y = 17439, vidimo, da je rezultat v skladu s teorijo. Prava vrednost pridelka resnično leži v razmiku zaupanja. Dobljena točkovna ocena za pridelek Y_{sl} = 17808 je razmeroma dobra ocena skupnega pridelka Y = 17439. Razlika ocene od prave vrednosti je 17808-17439 =+ 369 ali v odstotkih $100 \frac{369}{17439} = 2,1\ \%$.

Ocena strukturnega deleža parcel s pridelkom nad 30 je $p = n_a/n = 43/72 = 0,597$ ali v odstotkih p% = 100.0,597 = 59,7%.

Da bi ocenili varianco te ocene najprej iz podatkov vzorca, izračunamo oceno s_a^2

$$s_{a.}^2 = \frac{n_a(n - n_a)}{n(n - 1)} = \frac{43(72-43)}{72(72-1)} = 0,243936$$

Ocena variance za p je

$$\text{var}(p) = \frac{s_a^2}{n} \cdot \frac{N - n}{N} = \frac{0,243936}{72} \cdot \frac{576-72}{576} = 0,00296450$$

$$se(p) = \sqrt{var(p)} = \sqrt{0,00296450} = 0,0544$$

$$se(p\%) = 100.0,0544 = 5,44\%$$

$$d(p\%) = z_p \cdot se(p\%) = 1,96 \cdot 5,44 = 10,7\%$$

Odstotek parcel, ki imajo pridelek večji kot 30, je P% = 59,7% $\pm$ 10,7%. Pravi odstotek s tveganjem $\alpha = 0,05$ ni manjši kot 49% in ne večji kot 70,4%.

Če ocenimo še skupno število parcel, na katerih je pridelek večji kot 30 m^3, dobimo

$$N_{a,sl} = \frac{N}{n} n_a = \frac{576}{72} \cdot 43 = 344$$

$$var\, N_{a,sl} = N(N-n) \frac{s_a^2}{n} = 576.(576-72) \frac{0,243936}{72} = 983,55$$

$$se(N_{a,sl}) = \sqrt{var(N_{a,sl})} = \sqrt{983,55} = 31,4$$

$$d(N_{a,sl}) = z_p \cdot se(N_{a,sl}) = 1,96.31,4 = 61,5$$

$$N_{a,sl} = 344 \pm 62 \text{ parcel}$$

Skupno število parcel, na katerih je pridelek večji kot 30, je s tveganjem $\alpha = 0,05$ v mejah zaupanja med 282 in 406. Če za potrdilo teorije in za preskus stvarne zanesljivosti rezultata primerjamo dobljeno oceno s pravim številom parcel, ki smo ga dobili s pregledom vseh parcel, dobimo, da je pravo število parcel $N_a = 320$ v mejah zaupanja in da se ocena $N_{a,sl} = 344$ od pravega števila razlikuje za 344 - 320 = 24 ali $100 \frac{24}{320} = 7,5\%$.

Določanje velikosti vzorca pri zahtevani natančnosti ocene

14.29 **Določanje velikosti vzorca za oceno poprečja.** Pri dosedanjem obravnavanju ocenjevanja z vzorčenjem je bila velikost vzorca vnaprej znana in smo pri dani velikosti vzorca ocenjevali parametre in njihovo zanesljivost.

Pri načrtovanju vzorcev pa običajno nastopi problem določiti velikost vzorca tako, da bo

dal predpisano zanesljivost. Če ne pazimo na velikost vzorca, se utegne zgoditi, da so ocene, ki jih dobimo, neuporabne ali pa imajo omejeno vrednost, ker so premalo zanesljive, če je vzorec premajhen. Obratno pa je ocena glede na stvarne potrebe lahko predobra in smo po nepotrebnem vzeli prevelik vzorec.

Zanesljivost ocene z enostavnim vzorcem je dana s standardnim pogreškom ocene SE ali pa z odklonom zaupanja $D = z.SE$, ki pomeni največji odklon ocene od prave vrednosti pri danem tveganju α. Kot je znano, je od stopnje tveganja odvisen koeficient z.

Najprej skušajmo rešiti problem o velikosti vzorca za oceno poprečja z vzorcem s ponavljanjem. Vprašanje je, koliko enot moramo vzeti v vzorec, da se ocena poprečja z danim tveganjem $\alpha = 2P$ ne odklanja od prave vrednosti za več kot $D(\bar{y})$.

Iz obrazca 14.5 in obrazca 14.6 povzemamo, da je za enostaven slučajnostni izbor s ponavljanjem

$$D(\bar{y}) = z_p \cdot SE(\bar{y}) = z_p \cdot \frac{\sigma_y}{\sqrt{n}} \qquad (14.39)$$

Iz tega obrazca dobimo, da je za vzorec s ponavljanjem

$$n = \left(\frac{z_p \cdot \sigma_y}{D(\bar{y})} \right)^2 \qquad (14.40)$$

Vzemimo našo poskusno ploskev z $N = 576$ parcelicami in skušajmo določiti, kako velik vzorec s ponavljanjem moramo vzeti, da se ocena poprečja od pravega poprečnega pridelka na eno parcelo s tveganjem $\alpha = 0,05$ ne bo odklanjala za več kot $D(\bar{y}) = 1$.

Za tveganje $\alpha = 2P = 0,05$ je $z_p = 1,96$. Standardni odklon za pridelek po parcelicah je $\sigma_y = 6,29$. Iz teh podatkov dobimo po obrazcu 14.40, da je

$$n = \left(\frac{1,96 \cdot 6,29}{1} \right)^2 = 152$$

Da z enostavnim vzorčenjem s ponavljanjem ocenjeni poprečni pridelek na eno parcelo od pravega poprečja ne bo različen za več kot $D = 1$, je treba v slučajnosten vzorec vključiti $n = 152$ enot.

14.30 Med številom enot v vzorcu s ponavljanjem n in številom enot v vzorcu brez ponavljanja n´ pri isti natančnosti ocene je zveza

$$n' = \frac{N \cdot n}{N+n-1} \doteq \frac{n}{1+f} \qquad (14.41)$$

V našem zgledu dobimo

$$n' = \frac{576 \cdot 152}{576+152-1} = 120$$

Če ocenjujemo poprečje z vzorcem brez ponavljanja, dosežemo predpisano zanesljivost ocene že z vzorcem, v katerega vključimo le n´= 120 parcelic.

14.31 Velikost vzorca pri predpisani relativni natančnosti.

Zanesljivost ocene večkrat predpisujemo tudi relativno. Namesto absolutnega dopustnega odklona zaupanja izrazimo odklon relativno v odstotku od parametra, v našem primeru v odstotku od aritmetične sredine.

Če obrazec 14.40 ustrezno preuredimo, dobimo

$$n = \left(\frac{z \cdot KV\%}{D\%(\bar{y})} \right)^2 \qquad (14.42)$$

V obrazcu 14.42 smo števec in imenovalec pomnožili s 100/My. Tako dobimo v števcu namesto standardnega odklona σ koeficient variacije KV%, v imenovalcu pa namesto absolutnega odklona zaupanja $D(\bar{y})$ relativen odklon $D\%(\bar{y})$. Iz tega obrazca za število enot v vzorcu s ponavljanjem pa dobimo število enot v vzorcu brez ponavljanja po obrazcu 14.41.

14.32 Skušajmo določiti, s kako velikim vzorcem brez ponavljanja bi se ocena poprečnega pridelka na parcelo v naši preskusni parceli s tveganjem $\alpha = 0,01$ ne odklanjala od pravega poprečja za več kot $D\%(\bar{y}) = 5\%$.

Tveganju $\alpha = 0,01$ ustreza koeficient $z_p = 2,576$. Ker je koeficient variacije v našem zgledu KV% = 20,8%, dobimo po obrazcu 14.42

$$n = \left(\frac{2{,}576.20{,}8}{5}\right)^2 = 115$$

Potrebno število enot v vzorcu brez ponavljanja pa je po obrazcu 14.41

$$n' = \frac{576.115}{576+115-1} = 96$$

Določanje števila enot v vzorcu za ocene parametrov My, Y, P%, N_a

14.33 Podobno dobimo obrazce za določanje števila enot v vzorcu tudi za ocene parametrov: Y, P% in N_a. Obrazci za vse štiri osnovne parametre so dani v tabeli 14.8.

Tabela 14.8 Obrazci za določanje potrebnega števila enot pri dani zanesljivosti

Ocenjevani parameter		Število enot v vzorcu s ponavljanjem, če je maksimalni odklon predpisan	
		absolutno D	relativno D%
Aritmetična sredina	My	$n = \left(\frac{z \cdot \sigma}{D(\bar{y})}\right)^2$	$n = \left(\frac{z \cdot KV\%}{D\%}\right)^2$
Vsota podatkov	Y	$n = \left(\frac{z \cdot N \cdot \sigma}{D(Ysl)}\right)^2$	
Strukturni delež	P%	$n = \frac{z^2 \cdot P\%\,(100 - P\%)}{D(p\%)^2}$	$n = \left(\frac{100 \cdot z}{D\%}\right)^2 \cdot \frac{100-P\%}{P\%}$
Število enot z dano značilnostjo	N_a	$n = \frac{z^2 \cdot N_a \cdot (N-N_a)}{D(N_{a,sl})^2}$	
Število enot v vzorcu brez ponavljanja		$n' = \frac{N \cdot n}{N+n-1} \doteq \frac{n}{1+f}$	

14.34 Hiba navedenega postopka je v tem, da moramo poznati določene parametre iz osnovne populacije, če hočemo določiti potrebno število enot za načrtovani vzorec. Teh pa običajno nimamo. Zato si pri načrtovanju vzorcev pomagamo z ocenami za potrebne parametre. Te dobimo s predhodno analizo pojava, ki ga proučujemo ali iz razpoložljivih podatkov že izvršenih sorodnih raziskav. Včasih vrednosti teh parametrov ocenimo z manjšim vzorcem, katerega kasneje vključimo v glavni vzorec.

Drug problem, ki se pojavi v zvezi z določanjem velikosti vzorca, pa je, da običajno z enim vzorcem ne opazujemo enega samega znaka in ne ocenjujemo enega samega parametra. Vsak izmed teh znakov ima drugo variabilnost, ocena pa drugo zanesljivost. Zato pri načrtovanju velikosti vzorca dobimo za oceno vsakega parametra drugačno potrebno velikost vzorca. Težko se je izmed različnih "n" odločiti za velikost vzorca, ki naj bo osnova za vzorec, s katerim dobimo podatke za ocene vseh parametrov. V tem primeru se odločimo za kompromisno srednjo velikost vzorca. Pri tem se oziramo predvsem na važnejše parametre in skušamo vzeti vzorec tako velik, da zanesljivost ocen ustreza predvsem zanje.

Ocenjevanje po metodi razmerij

14.35 Vzemimo, da ocenjujemo skupno vrednost proizvodnje v obrtnih podjetjih za določeno stroko. Iz populacije vseh obrtnih podjetij te stroke izberemo na slučajnosten način vzorec podjetij. Po osnovnem obrazcu za enostavno ocenjevanje agregata v tabeli 14.1 izračunamo iz podatkov vzorca oceno poprečne proizvodnje na eno podjetje, to poprečje pa pomnožimo s skupnim številom podjetij. Zanesljivost te ocene je verjetno majhna, ker je vrednost proizvodnje po podjetjih zaradi različne velikosti obrtnih obratov zelo variabilna. Predpostavljajmo, da proizvodnost dela, merjena z vrednostjo proizvodnje na enega delavca po podjetjih, ni preveč različna. Če ta predpostavka velja, dobimo zanesljivejšo oceno skupne proizvodnje, če ocenimo poprečno proizvodnjo na delavca, to pa pomnožimo s skupnim številom delavcev v tej stroki.

Oceno skupne vrednosti proizvodnje Y po metodi razmerij dobimo z obrazcem

$$Y_{r,s} = X \frac{Sy}{Sx} = X \cdot r \qquad (14.43)$$

tako, da vsoto vrednosti proizvodnje v izbranih podjetjih Sy delimo z vsoto števila zaposlenih Sx v teh podjetjih, dobljeno oceno razmerja med vrednostjo proizvodnje in številom delavstva r pa pomnožimo s skupnim številom delavstva X. Ta metoda da tem boljše ocene agregatov, čim večja je korelacija med znakoma x in y. Če bi bila proizvodnja na enega delavca po podjetjih enaka, bi bila odvisnost med številom zaposlenih in proizvodnjo funkcijska. r bi bil v tem ekstremnem primeru prava vrednost razmerja med vrednostjo proizvodnje in številom delavstva, pa četudi bi ga izračunali iz podatkov za eno samo podjetje.

Čeprav je ocena po metodi razmerja samo dosledna, ne pa nepristranska ocena agregata, jo z velikim pridom uporabljamo vselej, kadar imamo za podatek, ki je v korelaciji s podatkom, za katerega ocenjujemo agregat, skupno vrednost za populacijo. To pa se v praksi pogosto primeri. V že izvršenih popisih in evidencah je veliko podatkov, ki jih moremo izkoriščati za ocenjevanje po metodi razmerij. Če smo npr. v preteklem letu popisali živino po kmetijskih obratih, moremo podatke lanskega popisa uporabiti za osnovo za ocenjevanje po metodi razmerja za vzorec v letošnjem letu. Podatki o številu živine v lanskem in letošnjem letu po posameznih kmetijskih gospodarstvih so v razmeroma visoki korelacijski odvisnosti. Po metodi razmerij pomnožimo lansko število živine z razmerjem skupnega števila živine za izbrana gospodarstva v letošnjem in lanskem letu.

Prav tako metodo razmerij s pridom uporabljamo pri vzorčni popravi popisnih podatkov. Neposredno po popisu ponovno skrbneje popišemo enote, ki jih s slučajnim vzorčenjem izberemo v kontrolni vzorec. Popisne podatke popravimo tako, da po metodi razmerij pomnožimo popisni podatek z razmerjem med vsoto podatkov, ki smo jih dobili v kontrolnem vzorcu in vsoto popisnih podatkov za enote v vzorcu. Upravičeno pričakujemo, da bo ocena po metodi razmerij dobra, ker je zelo verjetno, da so popisni podatki in podatki kontrole med seboj zelo odvisni.

Metoda razmerij se obnese tudi pri različnih drugih problemih. Če imamo za večje število pošiljk deklarirane teže (npr. carina), popravimo skupno deklarirano težo vseh pošiljk po metodi razmerij tako, da izmed vseh pošiljk izberemo slučajnostni vzorec in izbrane pošiljke stehtamo. Če vsoto pravih tež za izbrane pošiljke delimo s skupno vsoto deklariranih tež za iste pošiljke, dobljeno razmerje pa pomnožimo s skupno deklarirano težo vseh pošiljk, dobimo popravljen podatek o teži. Podobno popravljamo

ocene na oko itd.

Varianca ocene po metodi razmerij je enaka

$$\text{Var } Y_{r.sl} = N(N-n)\frac{s_r^2}{n} \; ; \quad s_r^2 = \frac{\sum (y - xR)^2}{N-1} \; ; \quad R = \frac{\sum y}{\sum x}$$

$$\text{var } Y_{r,sl} = N(N-n)\frac{s_r^2}{n} \; ; \quad s_r^2 = \frac{S(y-xr)^2}{n-1} = \frac{Sy^2 - 2r\,Sxy + r^2 Sx^2}{n-1} \; ; r = \frac{Sy}{Sx} \qquad (14.44)$$

Iz tega obrazca vidimo, da je standardni pogrešek ocene po metodi razmerij tem manjši, čim večja je korelacija med x in y.

Za občino z N = 2185 gospodarstvi z vzorcem n = 655 gospodarstev ocenjujemo letno proizvodnjo mleka v tej občini. Ker razpolagamo s podatkom o skupnem številu krav X = 4376, uporabimo metodo razmerij.

V n = 655 gospodarstvih, ki smo jih izbrali s slučajnim vzorcem, smo dobili naslednje podatke: skupno število krav v anketiranih gospodarstvih je bilo Sx = 1291 krav, skupna količina proizvedenega mleka v teh gospodarstvih pa je bila Sy = 2053 tisoč litrov. Po metodi razmerij je ocena skupne proizvodnje mleka

$$Y_{r,sl} = 4376 \cdot \frac{2053}{1291} = 6959 \text{ tisoč litrov}$$

Če primerjamo dobljeno oceno s pravim rezultatom Y = 6945, vidimo, da se rezultata presenetljivo skladata. Vendar menimo, da je ta skladnost bolj ali manj slučajna. Z enostavno oceno dobimo, da je ocenjena količina mleka Y_{sl} = 6850.

Najboljši pokazovalec uspešnosti metode razmerij v primerjavi z enostavno oceno je razmerje kvadratov standardnih pogreškov po eni in drugi metodi. To razmerje namreč pokaže, kolikokrat večji vzorec moramo vzeti, če hočemo z enostavnim načinom ocenjevanja dobiti tako zanesljivo oceno, kakor smo jo dobili z metodo razmerij. Ker je v našem zgledu za enostavno oceno Var Y_{sl} = 11300, za oceno po metodi razmerij pa Var $Y_{r,sl}$ = 2230, dobimo, da je

$$U_{sl\,/rsl} = \frac{Var\,Y_{sl}}{Var\,Y_{r,sl}} = \frac{11300}{2230} = 5,07$$

Da bi z enostavno oceno dobili tako zanesljivo oceno, kakor smo jo dobili po metodi razmerij, bi morali vzeti v vzorec s ponavljanjem petkratno število enot. Iz primera vidimo, da je izboljšava lahko znatna, opozoriti moramo pa na to, da je ocena le dosledna.

VZORČENJA Z OMEJITVAMI

14.36 Osnova dosedanjih ocen so bili enostavni vzorci. Vzorec je enostaven, če ima vsaka enota osnovne populacije enako možnost, da jo izberemo v vzorec. Enostaven vzorec imenujemo zaradi tega tudi vzorec brez omejitev.

Vendar ta način v praksi dostikrat ni najboljši niti po tehnični možnosti izvedbe niti po zanesljivosti ocen. Zato večkrat uporabljamo tudi druge načine vzorčenja. Od teh bomo obravnavali stratificirano vzorčenje, vzorčenje v skupinicah, vzorčenje v več stopnjah, vzorčenje v več fazah, sistematično vzorčenje in kvotna vzorčenja. Vsak izmed teh postopkov odpravi to ali ono pomanjkljivost enostavnega vzorčenja.

Stratificirano vzorčenje

14.37 Osnova in lastnosti. Obrazci o varianci ocen z enostavnim vzorčenjem pokažejo, da je zanesljivost ocene razen od velikosti vzorca v bistveni meri odvisna od variabilnosti proučevane populacije. Za homogene populacije, za katere je variabilnost majhna, je ocena boljša, za heterogene populacije oziroma za populacije z veliko variabilnostjo, pa so ocene pri isti velikosti vzorca manj zanesljive. Medtem ko moremo velikost vzorca spreminjati in tako vplivati na kakovost ocene, na variabilnost neposredno ne moremo vplivati, ker je lastnost populacije.

Vendar kljub temu pri vzorčenju izkoriščamo zakonitost, da so ocene za homogene populacije zanesljivejše kot za heterogene. Če namreč uspemo heterogeno populacijo razdeliti v homogene populacije, je variabilnost v posameznih delnih populacijah, ki jih imenujemo - s t r a t u m e , manjša. S samostojnimi enostavnimi vzorci v posamez-

nih stratumih dobimo ocene za posamezne stratume, iz teh ocen pa sestavimo ocene za populacijo. Tak način ocenjevanja je v primerjavi z enostavnim vzorcem tem uspešnejši, čim manjša je variabilnost pojava znotraj stratumov, neodvisna pa je od variabilnosti med stratumi. Zato skušamo pri stratificiranem vzorčenju razdeliti populacijo v stratume po onem znaku oziroma znakih, ki bistveno vplivajo na proučevani pojav. Najpopularnejša je regionalna stratifikacija, pri kateri populacijo razdelimo v homogene grupe glede na geografske rajone. Tako je uspešna razdelitev gospodarstev po kmetijskih rajonih, prebivalstva po republikah ipd. Kot načelo za razdelitev na stratume pa more služiti vsak drug kriterij, ki prispeva k temu, da so razmere v stratumih čim bolj izenačene. Tako moremo prebivalstvo razdeliti v stratume po spolu, dohodkovnih skupinah ipd., trgovine po vrstah trgovin ipd. Pogosto pa za razdelitev populacije v stratume ni vodilo en sam znak, ampak več znakov hkrati.

Hiba stratificiranega vzorčenja pa je predvsem v tem, da moramo populacijo razmeroma dobro poznati, če hočemo izvesti stratifikacijo. Medtem ko moramo pri enostavnem vzorčenju poznati le okvir vzorčenja, tj. spisek enot, moramo pri stratificiranem vzorčenju populacije znatno bolje spoznati, če jo hočemo razdeliti v homogene dele. Vsako enoto v osnovni populaciji je treba namreč pred izborom vzorca vključiti v ustrezno grupo - stratum.

Običajno pa ne proučujemo popolnoma neznanih pojavov. Zato izvedemo stratifikacijo na osnovi prejšnjih pregledov, statistik ali analiz.

Ker je stratifikacija po pravilu izvedena po nekem načelu, ki je v vsebinski zvezi s proučevanim pojavom, dobimo s stratificiranim vzorčenjem kot postranski rezultat tudi rezultate po grupah - stratumih. Zavedati pa se moramo, da so rezultati za posamezne stratume nezanesljivi, če je število enot vzorcev v posameznih stratumih majhno.

14.38 Ocenjevanje agregata s stratificiranim vzorcem. Nakazali smo že, da pri stratiriciranem vzorčenju iz delnih ocen, ki jih dobimo iz samostojnih enostavnih vzorcev po stratumih, sestavimo skupno oceno za celotno populacijo.

Ocena agregata s stratificiranim vzorčenjem Y_{str} je vsota ocen agregatov v posameznih stratumih $Y_{k,sl}$

$$Y_{str} = \sum_{k=1}^{l} Y_{k,sl} = \sum_{k} \frac{N_k}{n_k} S_{y_{ki}} \qquad (14.45)$$

Varianca za oceno agregata s stratificiranim vzorčenjem Var (Y_{str}) pa je vsota varianc ocen agregatov po stratumih

$$Var(Y_{str}) = \sum_{k=1}^{l} Var(Y_{k.sl}) = \sum_{k} N_k (N_k - n_k) \frac{S_k^2}{n_k} \qquad (14.46)$$

Iz teh dveh obrazcev moremo z upoštevanjem zvez med M_y, Y, P in N_a razviti podobne obrazce za prave vrednosti ali ocene parametrov ali varianc za kateregakoli izmed navedenih štirih parametrov.

14.39 Razmestitev enot vzorčenja po stratumih. Pri stratificiranem vzorčenju je ena izmed osnovnih nalog določiti, koliko enot od skupnega števila enot v vzorcu vzamemo v posamezen stratum. Ta razmestitev namreč ni dana vnaprej. Pri podrobnejšem študiju stratificiranega vzorčenja opazimo, da je od tega, koliko enot od skupnega števila enot v celotnem vzorcu je v posameznem stratumu, v bistveni meri odvisna zanesljivost ocene. Zato je ta problem toliko pomembnejši.

14.40 Proporcionalna razmestitev. Najenostavnejša razmestitev, ki pa je tudi najpogostejša, je proporcionalna razmestitev. Pri proporcionalni razmestitvi je velikost vzorcev v posameznih stratumih sorazmerna skupnemu številu enot v stratumih. Če z N_k zaznamujemo skupno število enot, z n_k pa število enot v vzorcu v stratumu k, določimo n_k pri proporcionalni razmestitvi po obrazcu

$$n_k = a N_k \; ; \quad a = \frac{n}{N} = f \qquad (14.47)$$

14.41 Vzemimo naš kompleksen zgled sestoja z N = 576 parcelicami, ki so razdeljene v tri stratume 1, 2 in 3. Osnovna razdelitev na stratume za to populacijo je nakazana v okviru na sliki 14.6 na strani 395.

V tabeli 14.9 je prikazano, kako skupno n = 72 enot vzorca razmestimo po stratumih pri proporcionalni razmestitvi. Razen tega je v isti tabeli dana varianca za oceno skupnega pridelka pri proporcionalni razmestitvi. Konstanta, s katero pomnožimo N_k, da dobimo n_k, je po obrazcu 14.47 a = 72/576 = 0,125. Enotni vzorčni delež f je torej f = 0,125.

Tabela 14.9 Izračunanje velikosti vzorcev v stratumih in varianca za oceno agregata za poskusno populacijo pri proporcionalni razmestitvi

Stratum	N_k	n_k	S_k^2	$Var(Y_{k,sl})$
1	144	18	3,2552	3281
2	192	24	14,5720	19585
3	240	30	6,4508	10837
	N = 576	n = 72		33703 = $Var(Y_{str,p})$

Dalje je: $SE(Y_{str,p}) = \sqrt{Var(Y_{str,p})} = \sqrt{33703} = 183,6$

$$SE\%(Y_{str,p}) = 100.SE(Y_{str,p})/Y = 100.183,6/17439 = 1,05\%$$

Stratificirano vzorčenje da v našem zgledu torej znatno boljše rezultate kot enostavno vzorčenje. Medtem ko je relativni pogrešek za oceno agregata Y z enostavnim vzorčenjem enak $SE\%(Y_{sl}) = 2,29\%$, je relativni pogrešek ocene pri stratificiranem vzorčenju s proporcionalno razmestitvijo skrčen na $SE\%(Y_{str,p}) = 1,05\%$.

Če izračunamo koeficient uspešnosti stratificiranega vzorca s proporcionalno razmestitvijo v primerjavi z enostavnim slučajnostnim vzorcem dobimo

$$U_{sl/str,p} = \frac{Var(Y_{sl})}{Var(Y_{str,p})} = \frac{160180}{33703} = 4,75$$

Če hočemo z enostavnim vzorcem doseči tako kakovost za oceno pridelka kot jo dosežemo s stratificiranim vzorcem s proporcionalno razmestitvijo enot s skupno n = 72 osnovnimi površinicami, moramo v enostaven slučajen vzorec vzeti približno 4,75 krat toliko enot kot pri stratificiranem vzorčenju. Stratifikacija je v tem primeru zelo uspešna.

Varianco ocene agregatov v posameznih stratumih smo izračunali po znanem obrazcu 14.23 za izračunavanje variance za ocene agregatov z enostavnimi slučajnostnimi vzorci.

14.42 Optimalna razmestitev. Kljub uspešnosti proporcionalne razmestitve pri stratificiranem vzorčenju pa proporcionalna razmestitev ni najboljša. Ocena v posameznih stratumih je tem zanesljivejša, čim manjša je variabilnost v stratumu in obratno. V stratumih, ki so zelo homogeni, je potreben za razmeroma dobro oceno sorazmerno manjši vzorec kot v variabilnejših stratumih. Iz tega sklepamo, da dobimo ugodnejšo razmestitev enot v stratumih, če razen števila enot v posameznih stratumih pri določanju velikosti vzorcev upoštevamo tudi variabilnost. Račun pokaže, da je optimalna razmestitev enot v stratumih, če upoštevamo število enot in variabilnost, dana z obrazcem

$$n_k = a \cdot N_k S_k \; ; \qquad a = n / \sum N_k S_k \tag{14.48}$$

Po tem obrazcu dobimo dano število enot v vzorcu optimalno razmeščeno po stratumih, če vzamemo število enot v posameznem stratumu proporcionalno številu enot N_k in standardnemu odklonu S_k v ustreznem stratumu.

14.43 Za naš zgled je izračunanje optimalne razmestitve enot in variance ocene za tak stratificiran vzorec nakazan za oceno agregata v tabeli 14.10.

Tabela 14.10 Izračunanje velikosti vzorcev v stratumih in variance ocene agregata za poskusno populacijo pri optimalni razmestitvi

Stratum	N_k	S_k	$N_k S_k$	n_k	$Var(Y_{k,sl})$
1	144	1,80	259,20	12	5156
2	192	3,82	733,44	33	13480
3	240	2,54	609,60	27	12214
	576		1602,24	72	30850
	N		$\sum N_k S_k$	n	$Var(Y_{str,0})$

Število enot po stratumih dobimo po obrazcu 14.48, pri čemer je $n/\sum N_k S_k = 72/1602,24 = 0,044937$.

Osnovne podatke o N_k in S_k smo dobili iz tabele 14.9. Dalje je:

$$SE(Y_{str,o}) = \sqrt{Var(Y_{str,o})} = \sqrt{30850} = 175,6$$

relativen standardni pogrešek pa $SE\%\ (Y_{str,o}) = 100.SE(Y_{str,o})/Y = 100.175,6/17439 = 1,01\%$.

Primerjava variance za stratificirano vzorčenje z optimalno razmestitvijo s stratificiranim vzorčenjem s proporcionalno razmestitvijo da

$$U_{str,p/str,o} = Var\ Y_{str,p}/Var\ Y_{str,o} = 33703/30850 = 1,092.$$

Da bi s proporcionalno razmestitvijo dobili enako zanesljivost kot z optimalno razmestitvijo n = 72 enot vzorca, bi morali vzeti približno za 9% večji vzorec.

14.44 Optimalna razmestitev enot glede na stroške. Optimalna razmestitev, ki smo jo obravnavali v gornjem odstavku, je optimalna glede na število enot. V praktičnem delu pa so pomembnejši element kot število enot pri načrtovanju vzorca stroški opazovanja. Zato skušamo dalje iskati optimalno razmestitev enot

glede na skupne stroške opazovanja. Ta razmestitev je enaka kot pri optimalni razmestitvi glede na število enot, če so stroški za opazovanje ene enote v vseh stratumih enaki. Če pa so poprečni stroški za opazovanje vzorčnih enot po stratumih različni, optimalna razmestitev glede na stroške odstopa od optimalne razmestitve glede na število enot. Glede na stroške je optimalno število enot v posameznih stratumih tisto, ki da pri istih stroških najmanjšo skupno varianco in s tem najzanesljivejšo oceno. Vzemimo enostavno funkcijo stroškov, ki upošteva samo konstantne stroške in stroške, ki so v zvezi z opazovanjem posamezne enote vzorčenja. Konstantni stroški so oni, ki niso odvisni od tega, kakšno število enot vzamemo v vzorec. To so stroški za pripravo ankete, za sestavo okvirja itd. V variabilne stroške pa vzemimo stroške, ki so v zvezi z opazovanjem posamezne enote. Pri enostavni funkciji stroškov predpostavljamo da so stroški po posameznih stratumih proporcionalni številu enot v vzorcih po stratumih. V tem primeru so skupni stroški dani s funkcijo

$$C = C_o + C_v = C_o + \sum \bar{c}_k . n_k \qquad (14.49)$$

Pri tem pomeni: C = skupni stroški, C_o = konstantni stroški, C_v = variabilni stroški, $\bar{c}_k$ = poprečni stroški za opazovanje ene enote v stratumu, n_k = število enot v stratumu k.

Pri optimalni razmestitvi enot glede na stroške določimo število enot po posameznih stratumih po obrazcu

$$n_k = a \frac{N_k S_k}{\sqrt{\bar{c}_k}} \quad ; \quad a = C_v / \sum N_k S_k \sqrt{\bar{c}_k} \qquad (14.50)$$

14.45 Za naš zgled vzemimo, da so poprečni stroški za opazovanje ene osnovne parcele (potni stroški, stroški merjenja itd.) po stratumih naslednji:

$$\bar{c}_1 = 3600 \text{ din}, \quad \bar{c}_2 = 2500 \text{ din}, \quad \bar{c}_3 = 1600 \text{ din}$$

Stroški so različni zaradi različnega terena, velikosti teritorija itd. Če upoštevamo zgornje stroške na enoto po posameznih stratumih, znašajo spremenljivi stroški glede

na število enot po stratumih pri optimalni razmestitvi iz tabele 14.10.

$$C_v = \bar{c}_1 n_1 + \bar{c}_2 n_2 + \bar{c}_3 n_3 = 3600.12 + 2500.33 + 1600.27 = 168900 \text{ din}$$

Po nakazanem postopku določimo optimalno razmestitev enot po stratumih tako, da dobimo pri enakih stroških 168900, ki so odvisni od števila enot, čim manjšo varianco. Postopek za izračunanje optimalne razmestitve enot po stratumih glede na stroške je nakazan v tabeli 14.11:

Tabela 14.11 Določitev optimalne razmestitve enot po stratumih glede na stroške za enote po stratumih

Stratum	N_k	S_k	$\bar{c}_k$	$N_k S_k \sqrt{\bar{c}_k}$	$\frac{N_k S_k}{\sqrt{\bar{c}_k}}$	n_k	S_k^2	$Var(Y_{k,si})$
1	144	1,80	3600	15552	4,32	10	3,2552	6300
2	192	3,82	2500	36672	14,67	32	14,5720	13989
3	240	2,54	1600	24384	15,24	33	6,4508	9712
	576			76608		75		30001
	N			$N_k S_k \sqrt{\bar{c}_k}$		n		$Var(Y_{str,s})$

Konstanta a, s katero pomnožimo izraze $N_k S_k / \sqrt{\bar{c}_k}$, da dobimo število enot v posameznih stratumih, je

$$a = C_v / \sum N_k S_k \sqrt{\bar{c}_k} = 168900/76608 = 2,2047$$

Iz variance za oceno agregata s stratificiranim vzorčenjem pri novi razmestitvi spoznamo, da smo za isti denar kot pri optimalni razmestitvi glede na število enot, dobili večji vzorec in boljše rezultate. Primerjava varianc za obe razmestitvi

$$U_{str,o/str,c} = Var(Y_{str,o}) / Var(Y_{str,c}) = 30850/30001 = 1,028$$

namreč pokaže, da bi morali v stratificiran vzorec pri optimalni razmestitvi, če se ne oziramo na stroške, potrošiti za 2,8% več sredstev, če bi hoteli z njim dobiti enako kakovostno oceno kot s stratificiranim vzorčenjem, pri katerem upoštevamo razlike v stroških opazovanja po stratumih.

14.46 Zgled za ocenjevanje s stratificiranim vzorčenjem. Omenili smo že, da pri stratificiranem vzorčenju v vsakem stratumu izvedemo samostojen, od drugih stratumov neodvisen vzorec, da ocenimo iskani parameter za vsak stratum posebej, iz ocen za posamezne stratume pa sestavimo oceno za populacijo.

Prva naloga pri izvedbi stratificiranega vzorčenja je izdelava okvira vzorčenja. Pri stratificiranem vzorčenju je okvir sestavljen iz samostojnih seznamov enot po stratumih. V seznamu enot po posameznih stratumih ima vsak stratum svojo posebno numeracijo, ki omogoča, da pri izboru identificiramo na slučajnosten način izbrane enote. Tudi pri stratificiranem vzorčenju morejo biti osnova okvira seznami, kartoteke, geografske karte, skice in podobno, le da so okviri sestavljeni po stratumih. Vsak stratum štejemo za samostojno populacijo s samostojnim okvirom in numeracijo.

V sliki 14.6 je nakazan okvir vzorčenja za poskusno populacijo. Ploskev je glede na kakovost tal razdeljena v tri statume: 1, 2 in 3. Vsak stratum ima v okviru samostojno numeracijo parcelic.

Okvir vzorčenja je v našem zgledu shema ploskve z vpisanimi zaporednimi številkami parcelic v posameznem stratumu.

Zaporedne številke izbranih parcel dobimo iz tablic slučajnostnih številk v dodatku. Ker je skupno število osnovnih parcelic v prvem stratumu $N_1 = 144$, v drugem $N_2 = 192$ in v tretjem $N_3 = 240$, iz tablic slučajnostnih številk dobimo zaporedne številke izbranih parcelic po naslednjem pravilu: Od začetka tablic slučajnostnih številk sestavljamo drseče trimestne slučajnostne številke tako, da tromestno skupino pomikamo za eno slučajnostno številko. Ker je število enot v prvem stratumu pod 200, dobimo iz tromestne slučajnostne številke, ki je večja kot 200, zaporedno številko izbrane parcele tako, da od nje odštevamo ustrezen mnogokratnik od 200. Enako postopamo v drugem stratumu, ker tudi v njem število enot ni večje kot 200. V tretjem stratumu je število enot $N_3 = 240$. Zanj od tromestnih slučajnostnih številk odštevamo po znanem postopku 300. Vendar slučajnostne številke, ki so večje kot 900, v tretjem stratumu ne pridejo v poštev, ker bi sicer ne imela vsaka enota v stratumu enake možnosti, da jo izberemo in zato ocena ne bi bila nepristranska. V tabeli 14.12 je prikazan izbor vzorcev v posameznih stratumih. Poleg slučajnostne številke je v dru-

gem stolpcu vpisana zaporedna številka parcele, ki ustreza posamezni slučajnostni številki. Iz tabele je razvidno, da nekatere slučajnostne številke ne pridejo v poštev, ker so večje kot je skupno število enot po stratumih.

V prvem stratumu dobimo na primer zaporedno številko izbrane parcele, če ostanek tromestne slučajnostne številke, ki smo jo delili z 200 ni večji kot $N_1 = 144$. Po tem načelu ustreza prvi tromestni slučajnostni številki 034 parcela z zaporedno številko 34, slučajnostna številka 347 ne pride v poštev, ker je ostanek 347-200 = 147 večji kot $N_1 = 144$. Slučajnostni številki 474 ustreza parcela z zaporedno številko 474-400 = 74 itd. Podobno postopamo pri drugem in pri tretjem stratumu.

V tretjem stolpcu je vpisan pridelek na izbranih parcelah (vzet iz slike 14.4). V četrtem stolpcu je z 0 naznačeno, če je pridelek na izbrani parceli manjši ali enak 30, z 1 pa so označene parcele, na katerih je pridelek večji kot 30.

Izbrane parcele z vpisanimi osnovnimi podatki o pridelkih so vrisane v okviru v sliki 14.6.

Da ocenimo skupen pridelek in standardni pogrešek oziroma meje zaupanja, najprej izračunamo iz osnovnih podatkov o vzorcih za vsak stratum posebej pomožne količine y_k in s^2_{yk}.

Iz teh količin dalje po obrazcih iz pregleda obrazcev v odstavku 14.19 izračunamo ocene agregata - pridelka $Y_{k,sl}$ in ocene varianc $var(Y_{k,sl})$ za posamezen stratum. Vsota ocen pridelkov po stratumih je stratificirana ocena pridelka na celi ploskvi, vsota ocen varianc pa ocena variance pridelka s stratificiranim vzorcem. Analogno za oceno parcel s pridelkom nad 30 najprej preštejemo za posamezne stratume število parcel z nad 30 pridelka na_k in izračunamo s^2_{ak}. Iz teh količin pa dalje po obrazcih iz pregleda 14.19 izračunamo oceno števila enot z nad 30 $Na_{k,sl}$ in ocene variance tega števila $var(Na_{k,sl})$ za posamezen stratum. Enako kot pri agregatu je vsota ocen po stratumih enaka stratificirani oceni za celotno ploskev. Pomožne količine in osnovno izračunavanje stratificiranih ocen je nakazano v tabeli 14.13.

Tabela 14.12 Osnovni podatki o stratificiranem vzorcu na poskusni ploskvi

Stratum 1				Stratum 2				Stratum 3			
sl. št.	z. št.	y_i	$y_i > 30$	sl. št.	z. št.	y_i	$y_i > 30$	sl. št.	z. št.	y_i	$y_i > 30$
034	34	20	0	614	14	33	1	742	142	37	1
347	-	-	-	146	146	33	1	424	124	34	1
474	74	21	0	469	69	35	1	246	-	-	-
743	143	24	0	698	98	27	0	467	167	36	1
437	37	24	0	986	186	37	1	676	76	33	1
373	-	-	-	863	63	31	1	762	162	35	1
738	138	21	0	637	37	39	1	624	24	35	1
386	-	-	-	371	171	35	1	242	-	-	-
863	63	21	0	716	116	25	0	428	128	36	1
636	36	23	0	162	162	30	0	281	-	-	-
369	-	-	-	623	23	32	1	811	211	37	1
696	96	22	0	233	33	27	0	114	114	36	1
964	-	-	-	332	132	25	0	145	145	34	1
647	47	22	0	326	126	33	1	457	157	40	1
473	73	23	0	261	61	31	1	572	-	-	-
736	136	22	0	616	16	21	0	720	120	36	1
366	-	-	-	168	168	31	1	204	204	38	1
661	61	24	0	680	80	31	1	042	42	35	1
				804	4	31	1	425	125	36	1
	12	267	0	045	45	32	1	253	-	-	-
	n_1	y_1	na_1	456	56	32	1	533	233	34	1
				560	160	24	0	332	32	34	1
				601	1	29	0	323	23	35	1
				011	11	36	1	237	237	34	1
				111	111	25	0	373	73	37	1
				114	114	30	0	732	132	40	1
				141	141	22	0	322	22	33	1
				410	10	29	0	227	227	34	1
				109	109	27	0	270	-	-	-
				095	95	30	0	707	107	37	1
				959	159	28	0	073	73	dv	dv
				597	-	-	-	736	136	33	1
				977	177	29	0	360	60	36	1
				774	174	31	1	607	7	36	1
					33	991	17		27	961	27
					n_2	y_2	na_2		n_3	y_3	na_3

Tabela 14.13 Ocenitev skupnega pridelka in števila parcelic s pridelkom nad 30 na poskusni ploskvi s stratificiranim vzorčenjem

Stratum	N_k	n_k	y_k	$s^2_{y,k}$	$Y_{k,sl}$	$var(Y_{k,sl})$	na_k	$s^2_{a,k}$	$Na_{k,sl}$	$var\ Na_{k,sl}$
1	144	12	267	1,84	3204	2915	0	0	0	0
2	192	33	991	17,22	5766	15930	17	0,2576	99	238,27
3	240	27	961	3,48	8542	6589	27	0	240	0
	576	72			17512	25434			339	238,27
	N	n			Y_{str}	$var(Y_{str})$			Na_{str}	$var(Na_{str})$

Iz varianc ocen izračunamo standardne pogreške in meje zaupanja

$$se(Y_{str}) = \sqrt{var(Y_{str})} = \sqrt{25434} = 159,5$$

$$d(Y_{str}) = z_p \cdot se(Y_{str}) = 1,96 . 159,5 = 313$$

Prava vrednost pridelka s tveganjem $\alpha = 0,05$ ni manjša kot

$$Y_{min} = 17512 - 313 = 17199 \text{ in ne večja kot}$$

$$Y_{max} = 17512 + 313 = 17825.$$

Če primerjamo dobljeni rezultat s pravo vrednostjo $Y = 17439$ vidimo, da ta vrednost leži v mejah zaupanja. Razlika ocene od prave vrednosti $Y_{str} - Y = 17512 - 17439 = +73$ je komaj 0,42% od skupnega podatka. Dobljena ocena je znatno boljša kot pa smo jo dobili z enostavnim vzorcem.

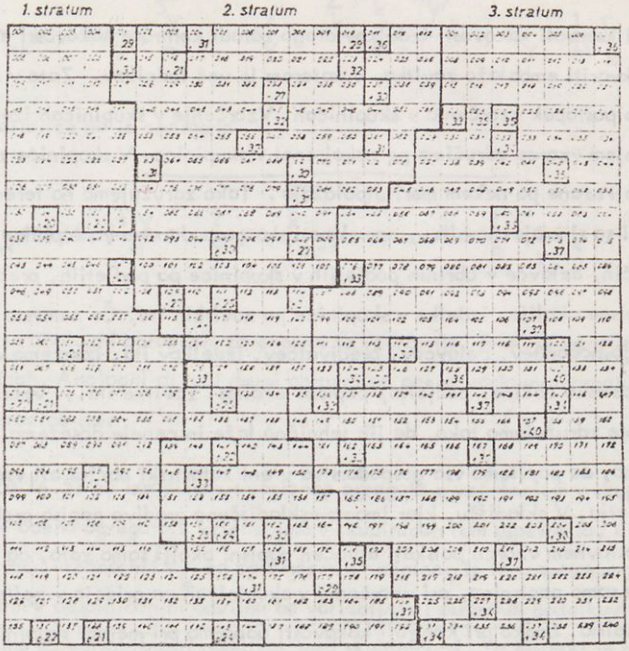
Za število parcel s pridelkom nad 30 pa je:

$$se(Na_{str}) = \sqrt{var(Na_{str})} = \sqrt{238,27} = 15,44$$

$$d(Na_{str}) = z_p \cdot se(Na_{str}) = 1,96 . 15,44 = 30$$

Pravo število parcel z nad 30 je s tveganjem $\alpha = 0,05$ večje kot $Na_{min} = Na_{str} - d(Na_{str}) = 339 - 30 = 309$, in manjše kot $Na_{max} = Na_{str} + d(Na_{str}) = 339 + 30 = 369$.

Stratifikacija je v našem zgledu zelo uspešna, saj se ocena $Na_{str} = 339$ od pravega



Slika 14.6 Okvir za stratificirano vzorčenje z vrisanimi slučajnostno izbranimi parcelami po stratumih pri optimalni razmestitvi

števila parcel s pridelkom nad 30 (Na = 344) razlikuje za 5.

Vzorčenje v skupinicah

14.47 Vendar stratifikacija ne odstrani eno izmed osnovnih tehničnih slabosti pri enostavnem vzorčenju. Če iz osnovne populacije izberemo na slučajen način vzorec, so te enote posamično razmeščene po vsej populaciji. To posebno nazorno opazimo, če so enote populacije razmeščene po terenu in je anketiranje teh enot razmeroma obsežno terensko delo. V posameznih primerih moramo iti v oddaljen kraj, da anketiramo eno samo enoto. Kljub temu, da je taka razmestitev za kakovost ocene ugodna, je iz tehničnih ozirov težko izvedljiva in draga.

Na podobne težave naletimo tudi, če so enote zbrane na enem in istem mestu. Iskanje popisnih listov, ki ustrezajo slučajnim številkam vzorca, v skladišču za popis prebivalstva, kjer je vskladiščenih več milijonov popisnih obrazcev, je tudi razmeroma težavno.

Enako je pri kontroli kakovosti izdelkov drago zaradi kontrole kakovosti enega izdelka odpirati in kvariti embalažo zavitka, v katerem je več izdelkov. Zato v dosti primerih uporabljamo postopek vzorčenja v skupinicah. Vzorčenje v skupinicah izvedeno tako, da pred izbiranjem vzorca združimo po nekaj enot v skupinice. V skupiniče združujemo enote, ki so vezane po nekem merilu položaja. Tako združujemo na terenu v skupinice enote, ki so si krajevno blizu, npr. kmetijska gospodarstva v skupinice v okviru katastrskih občin, delavce v obrtnih podjetjih v skupinice po podjetjih, prebivalstvo v družine ali popisne okoliše, izdelke v skupinice po zavitkih itd. Če namesto vzorca osnovnih enot - gospodarstev, delavcev, prebivalcev, izdelkov itd. izberemo v vzorec skupinice; katastrske občine, podjetja, popisne okoliše, zavitke itd., v vzorcu skupinic osnovne enote združujemo tako, da jih tehnično laže in ceneje dosežemo. V določen kraj potujemo, da popišemo vsa gospodarstva v eni katastrski občini ali vse zaposlene v enem podjetju. V skladišču, kjer imamo vskladiščeno gradivo popisa prebivalstva, moramo pri enostavnem vzorcu poiskati določen popisni okoliš samo zato, da iz njega izvlečemo eno samo popisnico, pri vzorčenju v skupinicah pa pridejo v poštev vse popisnice tega okoliša. Enako pri kontroli kakovosti odpremo pri metodi skupinic zavitek, da pregledamo vse izdelke, ne pa samo enega itd.

14.48 Z y_{ki} zaznamujemo vrednost osnovne enote i v skupinici k. Vrednost agregata v skupinici k je vsota vseh vrednosti osnovnih enot

$$Y_k = \sum_{i=1}^{N_k} y_{ki} \qquad (14.51)$$

Če je v populaciji M skupinic in s slučajnostnim vzorcem izberemo m skupinic, je po znanih stavkih za ocenjevanje agregata z enostavnim vzorčenjem ocena agregata z vzorčenjem v skupinicah

$$Y_{sk} = \frac{M}{m} \sum_{k=1}^{m} Y_k \quad ; \qquad (14.52)$$

varianca za oceno agregata z vzorčenjem v skupinicah pa je

$$\mathrm{Var}(Y_{sk}) = M(M-m)\frac{s_Y^2}{m} \; ; \quad s_Y^2 = \frac{\sum_k (Y_k - M_Y)^2}{M-1} \; ; \quad M_Y = \frac{1}{M}\sum_k Y_k \qquad (14.53)$$

Ocena variance pa je analogno

$$\mathrm{var}\, Y_{sk} = M(M-m)\frac{s_Y^2}{m} \; ; \quad s_Y^2 = \frac{\sum_k (Y_k - \bar{Y})^2}{m-1} \; ; \quad \bar{Y} = \frac{1}{m}\sum_k Y_k \qquad (14.54)$$

Obrazci so analogni obrazcem za enostavno vzorčenje, le da se nanašajo na podatke o skupinicah. Analogni so tudi drugi obrazci za ocene varianc agregata, za ocene aritmetične sredine itd.

Iz obrazca 14.53 za izračunanje variance agregata z vzorčenjem v skupinicah sklepamo, da je ocena agregata z vzorčenjem v skupinicah tem uspešnejša, čim manjša je varianca med vrednostmi Y_k za posamezne skupinice. Iz tega spoznamo, da ocena z vzorčenjem v skupinicah ni odvisna od variabilnosti osnovnih podatkov v skupinicah, ampak samo od variabilnosti podatkov med skupinicami. Ta lastnost je ravno obratna kot pri stratificiranem vzorčenju, pri katerem je zanesljivost ocene odvisna samo od variabilnosti znotraj skupin - stratumov. Zato pri stratificiranem vzorčenju težimo za tem, da so razlike med stratumi čim večje, variabilnost znotraj stratumov pa čim manjša. Pri vzorčenju v skupinicah pa je obratno ocena tem boljša, čim manjša je variabilnost med skupinicami.

14.49 Zanesljivost ocen, rentabilnost oziroma upravičenost vzorčenja v skupinicah je v bistveni meri odvisna od intraklasne korelacije skupinic.

Intraklasna korelacija raziskuje kakšna je korelacija med členi v posameznih skupinah. V zasnovo običajnega načina proučevanja korelacije to načelo prenesemo tako, da tvorimo pare podatkov in za vse skupine izračunamo korelacijski koeficient. Vzemimo za shematičen primer $m = 3$ skupinice, v katerih so po $\bar{N} = 3$ osnovne enote:

$a_1a_2a_3 \quad b_1b_2b_3 \quad c_1c_2c_3$.

Za naš zgled dobimo naslednje dvojice:

a_1a_2 a_1a_3 a_2a_3 b_1b_2 b_1b_3 b_2b_3 c_1c_2 c_1c_3 c_2c_3 .

Izkaže se, da je intraklasna korelacija v ozki povezavi s korelacijskim razmerjem η^2_{xy} in velja, da je koeficient intraklasne korelacije r_I

$$r_I = \frac{\bar{N}\eta^2_{xy} - 1}{\bar{N} - 1} \qquad (14.55)$$

η^2_I pove, kolik del skupne variabilnosti je pojasnjen z grupnimi aritmetičnimi sredinami. Če je vsa variabilnost pojasnjena z grupnimi aritmetičnimi sredinemi, so vrednosti v grupah med seboj enake. Korelacijsko razmerje je v tem primeru 1. Enako je po gornjem obrazcu tudi koeficient intraklasne korelacije r_I enak 1, kar je razumljivo, ker obstaja med členi funkcijska odvisnost, ker so med seboj enaki.

Če so grupne aritmetične sredine med seboj enake, je korelacijsko razmerje enako nič, koeficient intraklasne korelacije pa je po gornjem obrazcu negativen in ima vrednost $r_I = -1/(\bar{N} - 1)$.

14.50 Za varianco ocene za aritmetično sredino pri vzorčenju v skupinicah velja obrazec

$$\mathrm{Var}\,\bar{y}_{sk} = \frac{\sigma^2_y}{\bar{N}.m} \cdot \left[1 + (\bar{N} - 1)\, r_I\right] \qquad (14.56)$$

analogno pa za strukturni delež

$$\mathrm{Var}\,p_{sk} = \frac{P \cdot (1-P)}{\bar{N} \cdot m} \left[1 + (\bar{N} - 1)\, r_I\right] \qquad (14.57)$$

Varianca za oceno aritmetične sredine in strukturnega deleža je razen od komponent, ki vplivajo na zanesljivost ocen pri enostavnem vzorčenju v bistveni meri odvisna tudi od koeficienta intraklasne korelacije r_I. V naslednji tabeli so prikazani odnosi med korelacijskim razmerjem η^2_{Ixy}, koeficientom intraklasne korelacije r_I in variancama za

aritmetično sredino in strukturni delež

η^2_{yx}	r_I	Var $\bar{y}$	Var p
0	$-1/(\bar{N}-1)$	0	0
$1/\bar{N}$	0	$\sigma^2/\bar{N}.m$	$P(1-P)/\bar{N}.m$
1	1	σ^2/m	$P(1-P)/m$

V skrajnem primeru, ko je korelacijsko razmerje enako nič, (v tem primeru aritmetične sredine ali strukturni deleži po skupinicah niso različni) je koeficient intraklasne korelacije enak $-1/(\bar{N}-1)$, varianci ocen pa nič. To pomeni, da dobimo popolno informacijo o aritmetični sredini ali strukturnem deležu že z opazovanjem ene same skupinice.

Če je korelacijsko razmerje enako $1/\bar{N}$, je koeficient intraklasne korelacije 0, varianca aritmetične sredine pa Var $\bar{y} = \sigma^2/m.\bar{N}$, torej enaka kot bi vzeli enostaven slučajnostni vzorec z enakim številom enot druge stopnje $\bar{N}.m$ kot smo jih vzeli pri vzorčenju v skupinicah. Če se ne oziramo na stroške, dá vzorčenje v skupinicah enako kakovostne ocene kot slučajnostno vzorčenje z enakim številom enot $\bar{N}.m$.

Če je korelacijsko razmerje enako 1, pomeni, da imajo v grupah vse enote enake vrednosti, ker ni variabilnosti. Vzorčenje v skupinicah dá z opazovanjem $\bar{N}.m$ enot enako kakovostne rezultate kot enostavno slučajnostno vzorčenje z m enotami. Ker so v tem primeru vse enote v grupi med seboj enake, je razumljivo, da je tako vzorčenje v skupinicah enakovredno enostavnemu slučajnostnemu vzorcu z enakim številom enot kot je število izbranih skupinic.

Iz gornjih zvez sklepamo, da dá vzorčenje v skupinicah boljše ocene kot enostavno slučajnostno vzorčenje, če je intraklasna korelacija negativna. Če je intraklasna korelacija enaka nič, je vzorčenje v skupinicah enakovredno enostavnemu slučajnostnemu vzorčenju z enakim številom enot. Če pa je intraklasna korelacija pozitivna, pa vzorčenje v skupinicah daje manj zanesljive ocene kot enostavno slučajnostno vzorčenje.

14.51 Tehnika izbora enot vzorca s tablicami slučajnostnih števil je posebno za velike populacije neprikladna. Zato jo pogosto zamenjamo s sistematičnim izborom. Ta dá v določenih okoliščinah zadovoljive, včasih pa celo boljše rezultate kakor čisto slučajni vzorec. Pri sistematičnem izboru vzamemo v vzorec npr. vsako deseto enoto po spisku, če je vzorec desetodstotni. Ta izbor je sila preprost. Zanj je potreben samo utrjen spisek enot, ki je okvir vzorčenja. Prvo enoto izberemo slučajnostno izmed slučajnih števil, ki so manjša od stopinje izbora. Stopinja izbora je število enot od ene izbrane enote do naslednje enote, ki jo izberemo v vzorec.

Stopinjo izbora s določimo po obrazcu

$$s = 1/f \tag{14.55}$$

če poznamo vzorčni delež f ali po obrazcu

$$s = N/n \tag{14.56}$$

če imamo dano število enot v vzorcu in populaciji.

Običajno s ni celo število. V teh primerih vzamemo za stopinjo najbližje celo število. Na koncu sistematičnega izbora pa s ponovnim sistematičnim ali slučajnim izborom izključimo odvečne ali dopolnimo manjkajoče število enot. Če hočemo npr. v našem zgledu občine z N = 2185 gospodarstvi izbrati v vzorec n = 655 gospodarstev, je stopinja ustreznega sistematičnega vzorca

$$s = 2185/655 = 3,3$$

Če vključimo v vzorec vsako tretjo enoto po spisku, dobimo v vzorec n = 728 enot namesto 655 enot. Ker smo dobili v vzorec 73 enot preveč, moremo iz teh 728 izbranih enot s sistematičnim izborom s stopinjo

$$s = 728/73 = 10$$

izločiti odvečne enote.

14.52 Če je populacija sestavljena iz več delnih populacij oziroma spiskov enot, izberemo s slučajnim izborom samo prvo enoto v prvi delni populaciji oziroma spisku; prve enote v vseh naslednjih populacijah pa so določene z ostankom iz predhodne - delne populacije oziroma spiska. Če izbiramo npr. vzorec s stopinjo 10 in v prvem spisku določimo zadnjo enoto vzorca, ostanejo pa v tem spisku še štiri enote, v naslednjem spisku začnemo s šesto enoto, tako da je stopinja med zadnjo izbrano enoto v enem spisku in prvo izbrano enoto v naslednjem spisku zopet 10.

Za površine je sistematičen izbor določen s parcelami, ki so enako oddaljene druga od druge. Ta postopek dostikrat uporabljamo v kmetijski statistiki, v gozdarstvu itd.

14.53 Shematičen prikaz regionalnega sistematičnega vzorca je dan v sliki 14.10 .

Sistematičen izbor je posebno prikladen pri decentraliziranem izboru vzorca, ker je lahko razumljiv in izvedljiv.

Vendar ima sistematičen izbor svoje hibe, zaradi katerih moramo v vsakem posebnem primeru podrobno analizirati zaporedje enot, preden se odločimo za sistematičen izbor.

Zavedati se namreč moramo, da sistematičen izbor ni slučajen, zato moremo z njim nadomestiti slučajen izbor samo v določenih okoliščinah.

Zaradi ilustracije nakažimo shematičen zgled, v katerem pretiramo učinek, ki more nastati pri sistematičnem izboru.

Vzemimo, da smo v popisu prebivalstva v kolektiven obrazec vpisovali podatke o posameznih prebivalcih po vrsti po družinah. Vse popisane družine so petčlanske v temle sestavu: oče, mati in trije otroci. Mislimo si, da so družinski člani navedeni v spisku po vrsti: oče, mati in otroci po starosti. V spisku imamo v daljši vrsti prebivalce razporejene v naslednjem vrstnem redu:

MŽoooMŽoooMŽoooMŽoooMŽoooMŽoooMŽoooMŽoooMŽoooMŽoooMŽoooMŽooo

Če iz tega spiska izberemo sistematičen vzorec s stopinjo s = 10 in začnemo npr. s sedmo enoto po spisku, dobimo izbrane enote, ki so v gornji vrsti podčrtane.

Izbrali smo samo ženske oziroma matere. Seveda tak izbor ni reprezentativen in ga ne

moremo vzeti za osnovo za ocenjevanje podatkov o prebivalstvu. Sistematičen izbor je dal neuporabne podatke zato, ker je v zaporedju enot v spisku periodično ponavljanje. Perioda ponavljanja pa je ubrana s stopinjo izbora. V praktičnih primerih učinek sistematičnega izbora ni tako drastičen, ker nimamo toke periodičnosti v stvarnih spiskih, vendar more biti ta pojav, čeprav v milejši obliki, vir velike pristranosti v rezultatih ocenjevanja. Zato je treba pred uporabo sistematičnega izbora temeljito proučiti okvir vzorčenja in v primeru, če se pokaže gornji učinek, opustiti sistematičen izbor in ohraniti slučajni izbor enot.

Omenjeni učinek more nastati tudi pri regionalnem vzorčenju, ker morejo tudi na terenu nastopiti periodični pasovi boljših ali slabših razmer v vegetaciji in tako vplivati na kakovost ocene.

Sistematično vzorčenje pa dá ocene, ki so enakovredne rezultatom pri slučajnostnem izboru, če so enote v okviru razmeščene slučajnostno. V tem primeru je sistematičen izbor enakovreden slučajnostnemu izboru. S sistematičnim izborom pa moremo dobiti celo ocene, ki so boljše od ocen s slučajnostnim izborom. Če je okvir vzorčenja tak, da so v njem enote grupirane v neke smiselne homogene dele, dá sistematičen izbor rezultate, ki so po kakovosti enakovredni ocenam s stratificiranim vzorčenjem s proporcionalno razmestitvijo. Če so enote v okviru vzorčenja grupirane po občinah, s sistematičnim vzorcem s stopinjo izbora npr. s = 10 iz vsake občine avtomatično izberemo vzorec po 10% enot, kar je enakovredno stratificiranemu vzorčenju občin s proporcionalno razmestitvijo.

Ocena agregata s sistematičnim vzorčenjem je

$$Y_{sis} = \frac{N}{n} \cdot Sy = s.Sy \qquad (14.57)$$

Ker moremo po dve zaporedni vrednosti y_{2k+1} in y_{2k+2} $(k = 0,1,\ldots,\frac{n-1}{2})$ vzeti kot vzorec z $n_k = 2$ vrednosti v zaporednih $\frac{n}{2}$ mikrostratumih z $N = 2s$ enotami, dobimo oceno variance s sistematičnim vzorcem po obrazcu

$$\operatorname{var} Y_{sis} = s(s-1) \sum_{k=0}^{\frac{n-1}{2}} (y_{2k+2} - y_{2k+1})^2 \qquad (14.58)$$

Vzorčenje v dveh ali več stopnjah

14.54 Če v vzorcu kmetijskih gospodarstev v SFRJ v skupinicah popišemo vsa kmetijska gospodarstva v izbranih občinah, je število enot v posameznih občinah toliko, da imamo po metodi skupinic razmeroma veliko dela s popolnim popisom v izbranih občinah. Zato ocenimo agregat za posamezne izbrane občine s samostojnimi enostavnimi vzorci gospodarstev. Ta način vzorčenja imenujemo vzorčenje v dveh stopnjah. Ime je metoda dobila po tem, ker izbiramo dva vzorca. V prvi stopnji izberemo vzorec občin, v drugi stopnji pa v izbranih občinah vzorce kmetijskih gospodarstev. Obrazec za ocenjevanje agregata z vzorcem v skupinicah se za vzorčenje v dveh stopnjah toliko spremeni, da prave vrednosti agregatov Y_k za skupinice zamenjamo z ocenami $Y_{k,sl}$, ki jih dobimo z vzorci v drugi stopnji v enotah prve stopnje. Obrazci za ocenjevanje z vzorčenjem v dveh stopnjah so naslednji:

$$Y_{2.st} = \frac{M}{m} \sum_{k=1}^{m} \frac{N_k}{n_k} \sum_{i=1}^{n_k} y_{ki} \qquad (14.59)$$

$$Var\, Y_{2st} = Var\, Y_{sk} + \frac{M}{m} \sum_{k=1}^{M} Var\, Y_{k,sl} \; ; \quad var\, Y_{2st} = var\, Y_{sk} + \frac{M^2}{m^2} \sum_{k=1}^{m} var\, Y_{k,si} \qquad (14.60)$$

14.55 Vzorčenje v dveh ali več stopnjah zelo pogosto uporabljamo. Čeprav je pri vzorčenju v več stopnjah v skupnem številu potrebno večje število osnovnih enot kot pri enostavnem vzorčenju za isto zanesljivost rezultatov, običajno dobiček v stroških popisa odtehta vse teoretične hibe vzorčenja v več stopnjah.

Prednost vzorčenja v več stopnjah pred enostavnim vzorčenjem se pokaže že pri izbiranju enot. Pri enostavnem vzorčenju potrebujemo za gornji primer spisek vseh gospodarstev v SFRJ, ki rabi za okvir vzorčenja. Pri vzorčenju v dveh stopnjah pa potrebujemo spisek vseh občin, spiske kmetijskih gospodarstev pa samo za občine, ki so bile izbrane v prvi stopnji.

Pri načrtovanju vzorcev v dveh stopnjah je posebej pomembno razmerje med številom enot v vzorcu prve stopnje m in številom enot vzorcev v drugi stopnji n_k. Izkaže se,

da je optimalno razmerje med številom enot v vzorcu prve stopnje m in poprečnim številom enot $\bar{n}$ v vzorcih druge stopnje odvisno od razmerja stroškov, ki odpade na izvedbo vzorca v enoti prve stopnje in poprečnih stroškov, ki so potrebni za anketiranje enote v vzorcih druge stopnje in od intraklasne korelacije r_i. Praktični primeri so pokazali, da je varianca predvsem odvisna od m in manj od $\bar{n}$, zato je pri različnih anketah (anketa javnega mnenja, ankete o živini ipd.) optimalno poprečno število enot druge stopnje $\bar{n}$ med tri in osem.

Podatke ocenjujemo tudi z vzorčenjem v več stopnjah. Vzorčenje v štirih stopnjah moremo uporabiti za ocenjevanje pridelka za pšenico. V prvi stopnji izberemo enote prve stopnje - občine, v teh pa v drugi stopnji gospodarstva, ki goje pšenico. V izbranih gospodarstvih izberemo enote tretje stopnje - parcele, zasejane s pšenico, v teh pa enote četrte stopnje - kvadrate površin po 1 m^2. Oceno sestavimo po stopnjah podobno kakor za vzorčenje v dveh stopnjah.

Tudi odstotek pepela v premogu moremo za določen rudnik oceniti z vzorčenjem v več stopnjah. Iz tekoče proizvodnje premoga izberemo na slučajen način vagončke - enote prve stopnje. Iz vsakega izbranega vagončka v drugi stopnji izberemo določeno število kosov premoga. Skupnost vseh teh kosov razdrobimo in dobro premešamo, da dobimo homogeno populacijo. Iz te mešanice izbiramo v tretji stopnji nove enote - npr. po 10 dkg premogovega prahu. Ta vzorec predstavlja populacijo celotne proizvodnje premoga. Odstotek pepela, ki ga dobimo iz tega vzorca, je ocena za odstotek pepela v celotni proizvodnji.

Vzorčenje v dveh fazah

14.56 Vzemimo, da za določeno področje po metodi razmerij ocenjujemo skupno letno proizvodnjo v obrti. Pri metodi razmerij moramo poznati vrednost agregata za neki podatek, ki je z ocenjevanim v čim večji korelaciji. Vprašanje je, ali moremo idejo metode razmerij ohraniti, če v našem primeru skupnega števila zaposlenih v obrti ne poznamo. Če analiziramo podatka: število zaposlenih in letna proizvodnja v obrtnem podjetju, moremo presoditi, da je za posamezno obrtno podjetje veliko laže dobiti podatke o številu zaposlenih kakor o proizvodnji. Zato se izkaže, da je pametno, da

proizvodnjo po metodi razmerij ocenimo z vzorčenjem v dveh fazah po obrazcu

$$Y_{2f.r} = X_{sl.1} \cdot r_2 = \frac{N}{n_1} S_{x_1} \times \frac{S_{y_2}}{S_{x_2}} \qquad (14.61$$

Skupno število zaposlenih $X_{sl.1}$, ki ga ne poznamo, ocenimo z razmeroma velikim vzorcem (prva faza) n_1, ker je zanesljivost te ocene zaradi variabilnosti zaposlenih po obratih pri majhnem vzorcu majhna, zbiranje podatkov o številu zaposlenih pa ni zvezano s posebnimi težavami. Razmerje med proizvodnjo in številom zaposlenih r_2 pa cenimo po metodi razmerij z manjšim vzorcem obratov n_2, ker razmerje zanesljivo ocenimo tudi z majhnim vzorcem. Ta gospodarstva iz vzorca v prvi fazi izberemo v drugi fazi.

14.57 Idejo o vzorčenju v dveh fazah uporabimo tudi pri s t r a t i f i c i r a n e m vzorčenju. Stratifikacija zahteva, da popolnoma poznamo populacijo po znaku, po katerem stratificiramo, da moremo opredeliti vsako enoto osnovne populacije v enega izmed stratumov. To pa je včasih zelo obsežno in težavno delo. Zato izdelamo stratificirano oceno z vzorčenjem v dveh fazah. Da ni treba celotne osnovne populacije razdeljevati v stratume, vzamemo v prvi fazi razmeroma velik vzorec enot, za katerega pa zberemo samo podatke o znaku, po katerem populacijo stratificiramo. Iz podatkov vzorca v prvi fazi ocenimo po obrazcu za enostavno ocenjevanje število enot z značilnostjo posameznega stratuma $N_{k,sl}$. Iz vzorca v prvi fazi, ki je dal sestavo po stratumih, izberemo v posameznih stratumih samostojne vzorce v drugi fazi. Za enote druge faze pa šele zberemo podatke za katere ocenjujemo agregat. Iz vzorcev v drugi fazi ocenimo za vsak stratum oceno aritmetične sredine $\bar{y}$", skupno oceno pa sestavimo po obrazcu

$$Y_{str.2f} = \sum_k N_{k.sl.1} \bar{y}_{k.sl.2} = \frac{N}{n_1} \sum_k \frac{n_{k1}}{n_{k2}} \cdot S_{y_2} \qquad (14.62)$$

14.58 To metodo dostikrat uporabljamo pri anketiranju, da popravimo rezultate zaradi n e o d g o v o r o v . Pri večini anket določeno število enot, ki so izbrane v vzorec, iz katerega koli vzroka ne odgovori na anketo. Odstotek neodgovorov je bist-

veno odvisen od izvedbe ankete. Če npr. anketirane osebe prosimo, da po pošti pošljejo odgovore na zastavljena vprašanja, je odstotek oseb, ki na anketo ne odgovore, znatno večji, kakor če anketiramo z anketarji, ki z osebnim stikom pridejo do podatkov. Enako je odstotek neodgovorov odvisen od vsebine ankete. Na vprašanja o intimnih ali premoženjskih razmerah imamo v splošnem velik odstotek neodgovorov.

Popolnoma napačno ravnamo, če sklepamo takole: Ker računamo, da npr. v anketi, ki naj zajame 1000 oseb, 20% oseb ne bo poslalo odgovora, vzamemo v vzorec 1250 oseb in dobimo odgovore za želeno število oseb. Populacija oseb, ki ne odgovori na anketo, je po svojih lastnostih verjetno bistveno različna od populacije, ki odgovori. Če delamo na gornji način, dobimo oceno, ki ne velja za celo populacijo, marveč za tisti del prebivalstva, ki je voljan odgovoriti na anketna vprašanja. Ker pa hočemo dobiti oceno za celo populacijo, to je tudi za tisti del, ki npr. ni voljan odgovarjati na poštne ankete, je pravilen postopek takle. Izvedemo vzorec z n_1 enotami. Te enote anketiramo in od njih za n_1^+ enot dobimo odgovore, za n_1^- pa ne. Tako razen podatkov za n_1^+ enot, za katere je vsota podatkov S_{1y}^+, dobimo še stratifikacijo tako, da imamo v prvi grupi osebe, ki odgovarjajo na anketna vprašanja, v drugi grupi pa osebe, ki ne odgovarjajo na poštne ankete. Iz vzorca tistih n_1^-, ki niso odgovorili, izberemo v drugi fazi vzorec z n_2^- enotami. Zanje pa z drugimi sredstvi kakor pri prvi anketi skušamo zbrati podatke. Tako se na te osebe ne obračamo po pošti, marveč osebno poskušamo priti do podatkov s poizvedbami itd. Vsoto podatkov za teh n_2^- enot zaznamujemo z S_{2y}^-. Skupno oceno za agregate iz tega dvofaznega vzorca dobimo:

$$Y_{str.2f} = \frac{N}{n_1}\left(S_{1y}^+ + S_{2y}^-/f_2^-\right) \qquad (14.63)$$

Pri tem je razen znanih izrazov $f_2^- = n_2^-/n_1^-$ vzorčni delež enot v drugi fazi v populaciji neodgovorov.

14.59 Kvotno vzorčenje. Kljub najrazličnejšim izboljšavam mora biti vzorec razmeroma velik, da dosežemo zadovoljive rezultate. V vzorcu predhodnih rezultatov pri popisu prebivalstva v Sloveniji smo z vzorcem s približno 30000 enot dosegli rezultate s standardnim pogreškom okoli 1%.

Dostikrat pa nimamo možnosti, da bi ocenjevali populacijo s tako velikimi vzorci. Veliki vzorci zahtevajo, čeprav so le majhen del populacije, dosti časa in stroškov. S slučajnimi vzorci nekaj sto enot pa so ocene tako nezanesljive, da so v večini primerov neuporabne za namene, za katere jih sestavljamo. Tipičen primer so ankete javnega mnenja, ankete analize tržišča itd. Za reševanje teh problemov vzorčenje ne sme biti preobširno, rezultati pa morajo biti zadosti zanesljivi, da jih moremo analizirati.

V takih primerih uporabljamo za ocenjevanje metodo kvotnega vzorca. Kvotni vzorec odstopa od predpostavk slučajnega vzorčenja. Zato kvotni vzorec nima kvalitet, ki jih ima slučajen vzorec. Ima pa druge kvalitete. Za rezultate, ki jih dobimo s kvotnim vzorcem, ne moremo sicer določiti standardnega pogreška ocene. Glede na to, kako je ocena sestavljena, pa moremo sklepati, da je dobljena ocena boljša kot bi jo dobili s slučajnim vzorcem, ki bi bil enako velik. Princip kvotnega vzorca je v tem, da enote vzorca zavestno izberemo tako, da je sestava vzorca, ki rabi za ocenjevanje, za določene ključne znake enaka kot v celotni populaciji. Kvotni vzorec mora biti po sestavi enak populaciji v tistih znakih, ki so bistveni faktorji za podatke, ki jih ocenjujemo. Tako vzamemo pri anketah o javnem mnenju za osnovo kvotnega vzorca sestavo po starosti, poklicu, kraju itd. Iz sestave za populacijo moremo glede na skupno število enot vzorca določiti, koliko oseb, ki jih anketiramo (kvota), mora imeti določeno starost, poklic in biti iz določenega kraja. Po tej opredelitvi običajno anketar ne izbira posamezne osebe slučajnostno, marveč pustimo, da svobodno izbira osebe, ki danim pogojem zadoščajo. Ta vzorec ima v osnovi precej elementov stratifikacije s proporcionalno razmestitvijo.

14.60 Kot poskus, kako je kvotni vzorec uporaben, so v Zavodu za statistiko SRS po principu kvotnega vzorca iz gradiva popisa živine na dan 15.1.1953 izbrali n = 700 kmetijskih gospodarstev. Gospodarstva so bila izbrana tako, da je sestava vzorca po okrajih in velikostnih skupinah ustrezala sestavi v celotni populaciji. Ta dva znaka smo šteli za dva izmed bistvenih faktorjev, ki vplivajo na število živine. V tabeli 14.14 je prikazana razmestitev izbranih gospodarstev po tedanjih okrajih in po velikostnih skupinah. Ta razmestitev je bila glede na sestavo v celotni populaciji postavljena vnaprej. Iz gradiva o popisu živine pa so bila izbrana gospodarstva, ki so ustrezala tej sestavi.

Tabela 14.14 Razmestitev n = 700 gospodarstev v poskusni anketi po okrajih in velikostnih skupinah

Okraj	Skupno	Skupna površina v ha:										
		a	0,5	-1	1-2	2-3	3-5	5-8	8-10	10-15	15-20	20-
SR Slovenija	700	16	10	51	76	73	120	112	53	88	45	56
Celje-mesto	2	-	-	-	-	-	1	1	-	-	-	-
Celje-okolica	72	1	1	5	9	8	12	12	6	9	4	5
Črnomelj	18	-	-	1	1	1	1	3	2	5	2	2
Gorica	44	1	1	4	7	5	9	7	3	4	1	2
Kočevje	19	1	-	1	1	1	2	3	2	4	2	2
Kranj	24	-	-	1	1	2	3	3	2	4	3	5
Krško	48	1	-	4	5	5	9	9	4	7	2	2
Ljubljana-mesto	8	-	-	1	1	1	1	1	1	1	1	-
Ljubljana-okolica	61	1	1	4	4	4	8	9	5	10	6	9
Ljutomer	36	5	-	4	5	5	7	4	1	2	1	1
Maribor-mesto	3	-	-	-	1	1	1	-	-	-	-	-
Maribor-okolica	55	4	1	4	7	6	10	8	3	5	3	4
Mur.Sobota	79	1	1	5	12	14	23	15	3	4	1	-
Novo mesto	47	-	-	4	5	4	7	7	5	8	4	3
Postojna	26	-	-	1	1	1	3	4	3	5	4	4
Ptuj	61	1	3	9	10	7	10	9	4	5	2	1
Radovljica	10	-	-	-	-	1	1	2	1	2	1	2
Sežana	27	-	-	2	2	2	5	6	3	4	2	1
Slovenj Gradec	14	-	1	-	1	1	1	1	1	2	1	5
Šoštanj	12	-	-	-	1	1	1	2	1	2	1	3
Tolmin	25	-	1	1	2	2	4	5	2	3	2	2
Trbovlje	10	-	-	-	-	1	1	1	1	2	2	2

a = gospodarstva kmetijskih delavcev

V tabeli 14.15 so prikazani rezultati kvotnega vzorca za sestavo gospodarstev po številu konj, govedi in prašičev v primerjavi s podatki popolne obdelave popisa. Glede na to, da je vzorec majhen, so rezultati zelo dobri.

Tabela 14.15 Primerjalna slika med rezultati ankete po kvotnem vzorcu /A/ in rezultati popisa /P/ o številu gospodarstev po številu konj, govedi in prašičev /v tisočih/

Število konj	Število gospodinjstev		Število govedi	Število gospodinjstev		Število prašičev	Število gospodinjstev	
	P	A		P	A		P	A
Skupaj	146,6	146,6	Skupaj	146,6	146,7	Skupaj	146,6	146,5
0	110,8	108,9	0	15,0	16,8	0	27,7	29,7
1	17,8	16,8	1	23,7	23,5	1	32,9	32,7
2	14,2	16,1	2	33,7	33,7	2	32,6	32,3
3	3,2	4,2	3	28,1	25,1	3	19,9	19,7
4-	0,7	0,6	4	19,9	22,6	4	12,8	11,5
			5	11,3	10,3	5	7,6	7,5
			6-	15,9	14,7	6-7	7,5	5,2
						8-10	3,9	4,6
						11-	1,8	3,4

Časovno vzorčenje

14.61 Posebna uporaba vzorčenja, ki se je razvila skoraj v samostojno disciplino, je časovno vzorčenje. Časovno vzorčenje uporabljamo za analiziranje časovnih procesov. Pri proučevanju sestave o izrabi delovnega časa in najrazličnejših drugih časovnih študijah je velikega pomena sestava porabljenega časa v krajšem ali daljšem časovnem razmiku. Sestava izrabe delovnega časa v najrazličnejših organizacijah (proizvodnih podjetjih, administraciji, bolnicah, šolah ipd.) je osnovnega pomena pri reorganizacijah, izboljšavah poslovanja ipd. Če hočemo v enostavnejšem primeru priti do podatkov o produktivnem delovnem času v določenem razdobju, mora delavec ali kontrolor tekoče spremljati delo in registrirati, koliko časa delavec dela eno ali drugo fazo dela. Ker pa zahteva tak način registracije stalno spremljanje delovnega procesa, je drag in zamuden. Še zamudnejše in težavneje pa je opazovanje, če gre za celo ustanovo. Do rešitve problema s področja časovnih procesov z vzorčenjem pridemo z naslednjim razmišljanjem. Vzemimo proučevano časovno razdobje za zvezno populacijo, ki

sestoji iz neomejenega števila trenutkov. Če je y (t) funkcija časa, je

$$Y = \int_0^T y(t)\,dt \qquad (14.64)$$

vrednost agregata Y v proučevanem razdobju (0,T). Funkcija y (t) more biti v posebnih primerih zelo različna. Če opazujemo enega samega delavca in njegovo delo v proizvodnem procesu, postavimo, da je y(t) = 0 takrat, ko delavec ne dela proučevano fazo, y(t) = 1 pa za trenutke oziroma čas, v katerem delavec dela proučevano fazo. Y, dobljen po obrazcu 14,64 npr. pove, koliko časa je delavec v proučevanem razdobju T delal določeno fazo. Če proučujemo izkoriščeni čas za skupino delavcev, funkcija y(t) pomeni število delavcev, ki v trenutku t delajo, agregat Y pa skupno izkoriščen čas vseh delavcev. Podobno pomeni $M_y = Y/T$ v prvem primeru strukturni delež produktivno uporabljenega časa v razdobju T, v drugem primeru pa poprečno število delavcev, ki so v razdobju T delali. Podoben pomen imajo zgornji parametri tudi pri drugih problemih.

Če v proučevanem razdobju na slučajnosten način izberemo n trenutkov, teh n trenutkov štejemo kot slučajnosten vzorec iz zvezne populacije proučevanega časovnega razdobja T (populacija neomejenega števila trenutkov) z vsemi dobrimi in slabimi lastnostmi slučajnostnih vzorcev. Zato je

$$Y_{sI} = T\,\frac{\sum_i y(t_i)}{n} \qquad (14.65)$$

nepristranska ocena za časovni agregat Y

$$\bar{y}_{sI} = \frac{1}{n}\sum_i y(t_i) \qquad (14.66)$$

pa nepristranska ocena časovnega poprečja M_y.

Zvezna populacija razmika T more biti seveda glede na proučevan problem sestavljena iz več ločenih časovnih razmikov. Tako je pri proučevanju uporabe delovnega časa T npr. sestavljen iz osemurnih delavnikov za eno- ali večmesečno razdobje. Pri študi-

jah časa more biti T sestavljen iz razmikov, ko delavec dela za strojem, to je iz niza razmikov v eni izmeni ipd.

Pri proučevanju procesov najpogosteje naletimo na probleme, pri katerih proučevani pojav nastopa z daljšimi ali krajšimi prekinitvami. Najenostavnejši primer te vrste je produktivni delovni čas, če proučujemo enega samega delavca ali splošneje operatorja. Kot operatorja definiramo vsako proučevano enoto v časovnem procesu. Pri kompleksnejših raziskavah ima podoben značaj podatek, ko v skupini več delavcev delajo npr. trije delavci produktivno, ali ko pri študiju repov čaka pred okencem ali na bencinski črpalki npr. 5 strank ipd. Kot smo že nakazali, ima za take primere časovna funkcija procesa y(t) vrednosti 0 in 1. Če je število na slučajnosten način izbranih momentov zadosti veliko, vemo, da se strukturni delež $p = y/n$ porazdeljuje asimptotično normalno z vsemi zakonitostmi, ki veljajo za ocenjevanje relativnih frekvenc s slučajnostnim vzorčenjem.

Medtem ko je točkovna ocena za strukturni delež $p\% = 100\frac{y}{n}$, je intervalna ocena

$$P\% = p\% \pm 1.96\sqrt{\frac{p\%(100 - p\%)}{n - 1}} \qquad (14.67)$$

Če je y(t) kakšna druga zvezna ali nezvezna funkcija, ocenimo agregat s točkovno oceno

$$Y_{sI} = T \cdot \bar{y}$$

ali z intervalno oceno

$$Y = Y_{sI} \pm 1,96 \cdot T \frac{s_y}{\sqrt{n}} \qquad (14.69)$$

pri čemer je s_y ocena standardnega odklona σ_y, izračunana po obrazcu

$$s_y^2 = \frac{Sy_i^2 - \frac{(Sy_i)^2}{n}}{n - 1} \qquad (14.70)$$

14.62 Na shematičnem zgledu ponazorimo časovno vzorčenje deležev s časovnim trakom. za celotno proučevano razdobje T. V traku je s šrafiranimi odseki nakazano, kdaj se je v procesu dogajala proučevana faza, z nešrafiranimi pa čas, v katerem se ta faza ni vršila. Zaradi poenostavitve vzemimo, da proučevano razdobje obsega čas ene izmene od 6^h do 14^h; zaradi konkretnosti pa naj bo proučevana značilnost produktivni čas. V sliki 14.7 je nakazana popolna slika delovnega procesa (s trakom), pod njim pa je nakazano, kaj od celotnega procesa registriramo s slučajnostnim vzorcem n = 10 trenutkov. V sliki 14.7b je nakazano časovno vzorčenje, če gre za skupino operatorjev, v sliki c pa za sestavo števila enot v repih.

Čeprav je gornji zgled zaradi premajhnega razdobja in premajhnega vzorca nerealen, daje nazorno sliko o bistvu časovnega vzorčenja.

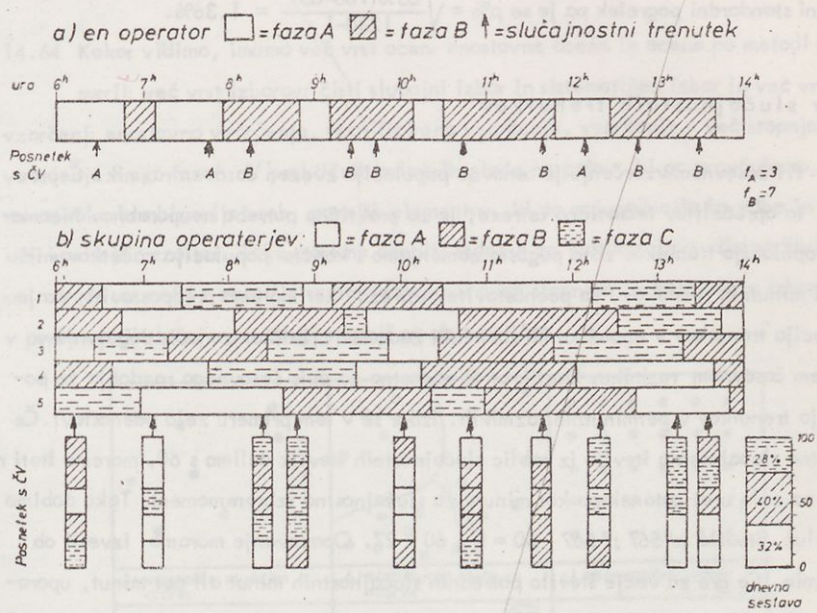
Navedeni način ocenjevanja uporabljamo pri vseh problemih, kjer imamo opravka z atributivnimi znaki dveh vrednosti, ki se s časom spreminjajo, npr.: dela, ne dela; je na zalogi, ni na zalogi.

To ne izključuje ocenjevanja in proučevanja sestave po več vidikih. Z enako tehniko in obrazci za ocenjevanje obravnavamo npr. sestavo izrabe delovnega časa za proces, ki sestoji iz več faz dela.

Za zgled vzemimo, da smo s časovnim vzorčenjem iz n = 1600 opazovanih trenutkov dobili za določenega operatorja naslednje rezultate:

Tabela 14.16 Rezultati časovnega vzorčenja za enega operatorja

	Delal produktivno				Ni delal produktivno			Skupno
	Faza A	Faza B	Faza C	Skupno	Vzrok A	Vzrok B	Skupno	
y	800	350	80	1230	170	200	370	1600
p%	50%	21,9%	5%	76,9%	10,6%	12,5%	23,1%	100,0%
se%	1,25%	1,03%	0,54%	1,05%	0,77%	0,83%	1,05%	-
p%	65,0%	28,5%	6,5%	100,0%	45,9%	54,1%	100,0%	-
se%	1,36%	1,29%	0,70%	-	2,60%	2,60%	-	-



Slika 14.7 Shematičen prikaz časovnega vzorčenja

Medtem ko je izračunani odstotek $50\% = 100 \cdot \frac{800}{1600}$ in ustrezni standardni pogrešek se $p\% = \sqrt{\frac{50 \cdot (100-50)}{1600-1}} = 1{,}25\%$, je izračunan odstotek za fazo $65{,}0\% = 100\,\frac{800}{1230}$, ustrezni standardni pogrešek pa je se $p\% = \sqrt{\frac{65\%(100-65)}{1230-1}} = 1{,}36\%$.

Izbor slučajnostnih trenutkov

14.63 Pri časovnem vzorčenju je osnovna populacija zvezen časoven razmik. Čeprav ta opredelitev teoretično ustreza, je za praktično potrebo neuporabna. Neomejeno populacijo trenutkov zato pogosto zamenjamo s končno populacijo začetkov minut ali minutnih razmikov. Ta poenostavitev ustreza, ker moremo predpostaviti, da je populacija trenutkov v enominutnih razmikih zadosten približek za opis dogajanja v zveznem časovnem razmiku. Dostikrat je zadostna za opis časovnega razdobja že populacija trenutkov v petminutnih razmikih. Izbor se v tem primeru zelo poenostavi. Če tromestno slučajnostno število iz tablic slučajnostnih številk delimo s 60, moremo šteti rezultat za cele ure, ostanek pa kot minute za slučajnostno izbran moment. Tako dobimo npr.: sluč. število je 567 ; $567 : 60 = 9 \times 60 + 27$. Opazovanje moramo izvesti ob 9^h 27 min. Če gre za večje število potrebnih slučajnostnih minut ali pet minut, uporabimo tablice slučajnostnih minut ali petminutnih razmakov, ki jih najdemo v ustrezni literaturi.

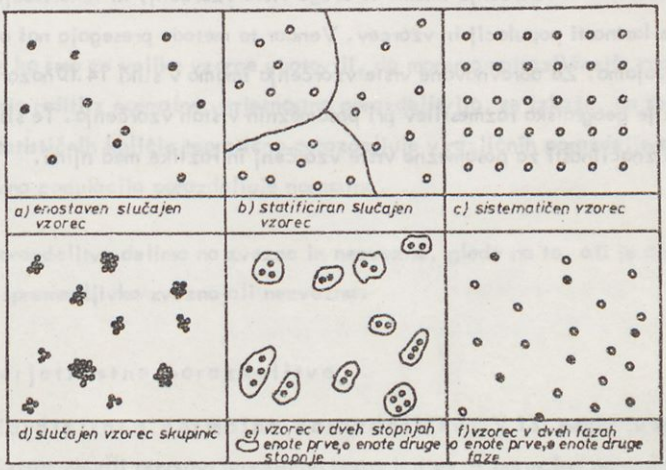
Pri časovnem vzorčenju moremo uporabiti vse metode, ki jih poznamo iz splošne teorije vzorčenja. Tako morejo stratume predstavljati razmiki po en dan, če gre za večdnevno opazovanje. Če vemo, da je proces različen glede na posamezne ure v izmeni, moremo kot stratume vzeti prve, druge ... zadnje delovne ure za vse dni raziskovanja.

Poseben pomen imajo pri časovnem vzorčenju skupinice. V praktičnih primerih namreč ob istem trenutku ne opazujemo le delo enega operatorja, temveč vse operatorje, ki delajo v istem prostoru. S tega vidika predstavljajo vsi delavci v enem prostoru ob istem trenutku skupinico. Tako časovno vzorčenje je treba šteti kot vzorčenje v skupinicah in ustrezno načrtovati in obračunavati zanesljivost dobljenih rezultatov.
V tej zvezi je bistvenega pomena intraklasna korelacija. Za določene primere je ta pri časovnem vzorčenju zelo velika, v drugih pa majhna oziroma celo negativna. Pri splošnem obravnavanju vzorčenja v skupinicah smo namreč spoznali, da intraklasna

korelacija bistveno vpliva na zanesljivost ocene.

S K L E P

14.64 Kakor vidimo, imamo več vrst ocen: enostavne ocene in ocene po metodi razmerij: več vrst izborov: čisti slučajni izbor in sistematičen izbor in več vrst vzorčenj: enostavno vzorčenje, stratificirano vzorčenje, vzorčenje v več stopnjah in vzorčenje v več fazah. V praktičnih primerih glede na pojav, ki ga proučujemo, vzamemo tako kombinacijo vseh gornjih elementov, ki da najugodnejšo izvedbo in najučinkovitejši rezultat. Tako moremo oceniti izdatke za kulturno in družbeno življenje po metodi razmerij s stratificiranim vzorcem v dveh stopnjah, v katerem je izbor enot v prvi stopnji slučajen, v drugi stopnji pa sistematičen.



Slika 14.10 Regionalna razmestitev enot vzorca pri različnih vrstah vzorčenja

Po tem načrtu moremo oceniti, koliko trošijo npr. delavci za kulturno in družbeno življenje. Stratuma sta npr.: podjetja v mestu in podjetja na deželi. V prvi stopnji v posameznem stratumu s slučajnim vzorcem izberemo podjetja, v teh pa s sistematičnim vzorcem vsakega desetega delavca. Za osnovo ocene po metodi razmerij uporabimo osebni dohodek, ker predpostavljamo, da so stroški za kulturno in družbeno življenje odvisni od osebnega dohodka, razpolagamo pa s skupnim fondom dohodkov za podjet-

ja. Obrazec za to oceno je

$$Y_{oc} = \sum_k \frac{M_k}{m_k} \sum_j X_{Kj} \frac{\sum_i y_{kji}}{\sum_i x_{kji}} \qquad (14.71)$$

Pri tem pomeni: y_{kji} = strošek za kulturno življenje, x_{kji} = dohodek delavca i v podjetju j in stratumu k ; $\sum_i$ = seštevanje podatkov iz vzorcev v posameznem podjetju ; x_{kj} = fond dohodkov izbranega podjetja j v stratumu k ; $\sum_j$ = seštevanje podatkov za vzorce podjetij ; M_k = število podjetij v stratumu k ; m_k = število podjetij v vzorcu v stratumu k in $\sum_k$ = seštevanje podatkov stratumov.

Razen navedenih vrst vzorčenjo imamo še druge vrste vzorčenj, ki izkoriščajo še različne druge lastnosti populacij in vzorcev. Vendar te metode presegajo naš okvir in jih zato ne navajamo. Za obravnavane vrste vzorčenja imamo v sliki 14.10 nazorno prikazano, kakšna je geografska razmestitev pri posameznih vrstah vzorčenja. Te slike nakažejo bistvene značilnosti za posamezne vrste vzorčenj in razlike med njimi.

PETNAJSTO POGLAVJE

VZORČENJE - MALI VZORCI

TEORETIČNE VERJETNOSTNE PORAZDELITVE

15.0 Medtem ko smo za velike vzorce ugotovili, da moremo najrazličnejše probleme vzorčenja rešiti z normalno verjetnostno porazdelitvijo, se izkaže, da se za male vzorce niz statističnih količin teoretično porazdeljuje v različnih porazdelitvah, čeprav se osnovna populacija porazdeljuje normalno.

Teoretične porazdelitve delimo na zvezne in nezvezne, glede na to, ali je osnovna slučajnostna spremenljivka zvezna ali nezvezna.

Zvezne verjetnostne porazdelitve

15.1 Standardizirana normalna porazdelitev. V 13. poglavju smo podrobno obravnavali osnovno teoretično porazdelitev - normalno porazdelitev. Normalna porazdelitev je osnovna teoretična porazdelitev glede na to, da so normalno porazdeljene slučajnostne spremenljivke osnova za vse druge in je večina teoretičnih porazdelitev izvedena kot funkcija normalno porazdeljenih slučajnostnih spremenljivk. Razen tega pa se večina teoretičnih porazdelitev asiptomatično porazdeljuje normalno.

Temeljna normalno porazdeljena slučajnostna spremenljivka je standardizirana slučajnostna spremenljivka z, ki je z osnovno normalno porazdeljeno spremenljivko y v zvezi

$$z = \frac{y - M_y}{\sigma_y} \qquad (15.1)$$

Gostota verjetnosti za $z =: N(0,1)$ je

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad (15.2)$$

porazdelitvena funkcija pa je funkcija zgornje meje

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} \qquad (15.3)$$

Ker je bila N (0,1) obravnavana v poglavju 13, je na tem mestu ne bomo podrobneje obravnavali.

15.2 χ^2-porazdelitev.

Vsota kvadratov m neodvisnih standardizirano normalnih slučajnostnih spremenljivk z_i

$$S z_i^2 = z_1^2 + z_2^2 + \ldots + z_m^2 =: \chi^2(m) \qquad (15.4)$$

se porazdeljuje v porazdelitvi, ki jo imenujemo χ^2 (hi-kvadrat) porazdelitev. Gostota verjetnosti za χ^2-porazdelitev je dana s funkcijo

$$\varphi(\chi^2) = C_{\chi^2} \cdot e^{\frac{\chi^2}{z}} (\chi^2)^{\frac{m}{2}-1} \qquad (15.5)$$

To je posebna oblika Γ-porazdelitve s $\varphi(x) = C_\Gamma e^{-x} x^{t-1}$, v kateri je $x = \chi^2/2$ in $t = \frac{m}{2}$

Porazdelitvena funkcija za χ^2-porazdelitev je

$$F(\chi^2) = \int_0^{\chi^2} \varphi(\chi^2)\, d\chi^2 \qquad (15.6)$$

Iz obrazca za gostoto verjetnosti je razvidno, da je gostota verjetnosti razen od slučajnostne spremenljivke χ^2 odvisna tudi od parametra m, ki ga imenujemo **stopi-**

nja prostosti. Stopinja prostosti m je število neodvisnih spremenljivk, iz katerih je sestavljen χ^2.

Iz definicije χ^2-porazdelitve sledi, da je $E(\chi^2) = m$; $Var(\chi^2) = 2m$. χ^2-porazdelitev je asimetrična v desno, stopnja asimetrije pa se manjša z večanjem števila stopinj prostosti m. Medtem, ko je χ^2-porazdelitev za $m = 1$ tipična J porazdelitev, je za $m = 10$ χ^2-porazdelitev že skoraj simetrična in zelo podobna normalni porazdelitvi.

Če je število stopinj prostosti $m > 30$, se izraz

$$\frac{1}{2}(z + \sqrt{2m-1})^2 =: \chi^2(m) \qquad (15.7)$$

asimptotično porazdeljuje v χ^2-porazdelitvi z m stopinjami prostosti. Pri tem velja, da je $z =: N(0,1)$.

χ^2-porazdelitev je odvisna od števila stopinj prostosti m, ki jo bolj ali manj določa. Zaradi tega za χ^2-porazdelitev niso možne tablice v obliki kot za normalno porazdelitev. So pa tudi nepotrebne. Izkaže se namreč, da potrebujemo pri praktičnem delu le kritične vrednosti oziroma kvantile za določene karakteristične vrednosti P.
Za χ^2-porazdelitev imamo v tabelah kritične vredbosti χ^2_p za

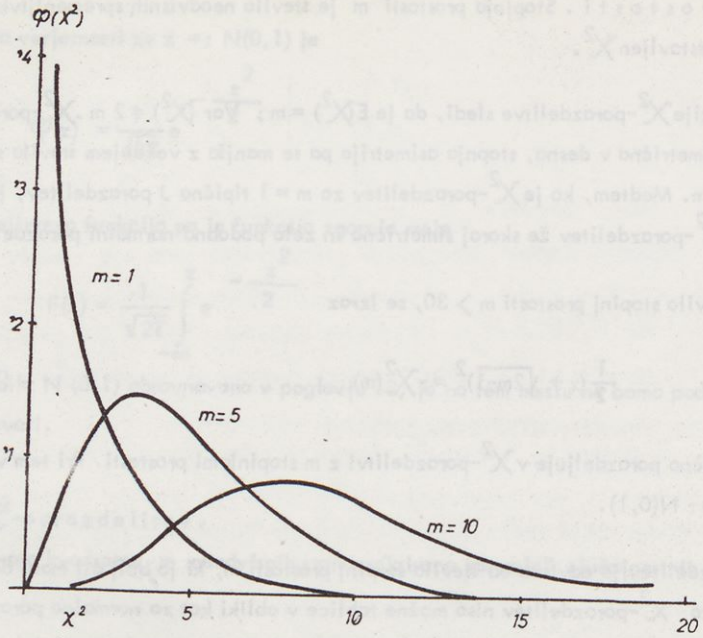
P = 0,995, 0.99, 0,975, 0.95, 0.90, 0,50, 0,25, 0,10, 0,05, 0,025, 0,01 in 0,005, 0,001.

Iz prej omenjenega razloga so tablice izdelane le do $m = 30$. Za druge vrednosti pa $\chi^2_p(m)$ izračunamo iz obrazca 15.7 tako, da vanj vstavimo ustrezen P in m.
Iz tablic moremo npr. odbrati, da $P = 0,10$ za $m = 15$ ustreza kritična vrednost $\chi^2_{P=0,10}(15) = 22,31$.

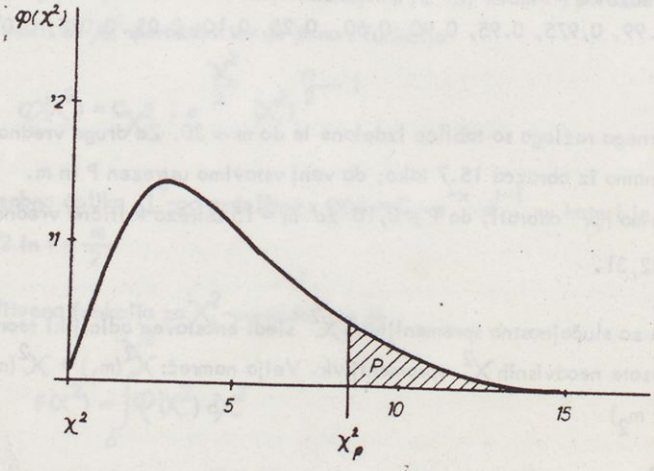
Iz opredelitve za slučajnostno spremenljivko χ^2 sledi enostaven adicijski teorem o porazdelitvi vsote neodvisnih χ^2 spremenljivk. Velja namreč: $\chi^2(m_1) + \chi^2(m_2) = \chi^2(m_1 + m_2)$

ali na splošno

$$\sum_{k=1}^{r} \chi^2(m_k) =: \chi^2\left(\sum_{k=1}^{r} m_k\right) \qquad (15.8)$$



Slika 15.1 χ^2-porazdelitev



Slika 15.2 Kritična vrednost za χ^2- porazdelitev

15.3 Studentova t- porazdelitev. Če se z porazdeljuje standardizirano normalno, χ^2 v hi- kvadrat porazdelitvi z m stopinjamo prostosti, z in χ^2 pa sta neodvisna, se slučajnostna spremenljivka,

$$\frac{z}{\sqrt{\chi^2/m}} =: t(m) \qquad (15.9)$$

ki je funkcija slučajnostnih spremenljivk z in χ^2, porazdeljuje v t- porazdelitvi z m stopinjami prostosti.

Za t- porazdelitev je gostota verjetnosti $\varphi(t)$ dana s funkcijo

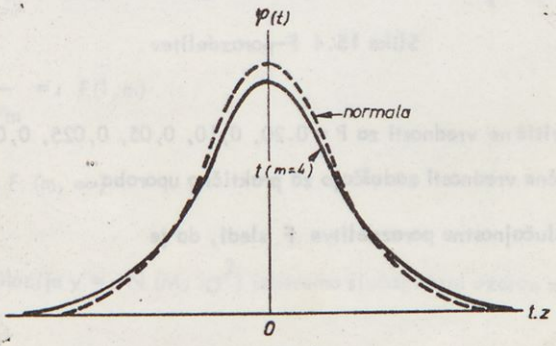
$$\varphi(t) = C_t \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \qquad (15.10)$$

pri čemer je t slučajnostna spremenljivka, C_t je

konstanta, ki je odvisna od števila stopinj prostosti, m pa število stopinj prostosti. Iz obrazca 15.10 in iz slike 15.3 je razvidno, da je t-porazdelitev simetrična porazdelitev glede na t=o, ki je obenem modalna vrednost. $\varphi(t)$ pa se asimptotično približuje abscisni osi, če se absolutna vrednost za t veča čez vse meje. Iz gornjega izraza in iz slike vidimo, da se t-porazdelitve asimptotično približujejo N(0,1), če m raste.

Iz istega razloga kot za χ^2-porazdelitev tudi za t-porazdelitev tabeliramo samo kritične vrednosti $t_p(m)$ in sicer so tabelirane kritične vrednosti za P=0,25, 0,05, 0,025, 0,01, 0,005, 0,001 in 0,0005.

Iz tablice za t-porazdelitev opazimo, kako se z večanjem m kritične vrednosti bolj in bolj približujejo kritičnim vrednostim za N (0,1).



Slika 15.3 t- porazdelitev

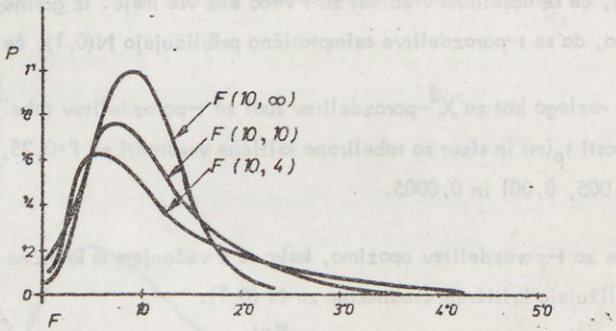
15.4 F-p o r a z d e l i t e v . Če sta $\chi^2(m_1)$ in $\chi^2(m_2)$ neodvisni slučajnostni spremenljivki, ki se porazdeljujeta v χ^2-porazdelitvi, prva z m_1 in druga z m_2 stopinjami prostosti, se

$$\frac{\chi^2(m_1)/m_1}{\chi^2(m_2)/m_2} =: F(m_1 \cdot m_2) \tag{15.11}$$

porazdeljuje v F porazdelitvi z m_1 in m_2 stopinjami prostosti. Gostota verjetnosti za F-porazdelitev je dana s

$$\varphi(F) = C_F \cdot F^{\frac{m_1-2}{2}} \cdot \left(1 + \frac{m_1}{m_2} F\right)^{-\frac{m_1+m_2}{2}} \tag{15.12}$$

F-porazdelitev je odvisna od dveh stopinj prostosti: m_1 ustreza χ^2 v števcu, m_2 pa χ^2 v imenovalcu. Iz analize funkcije in iz slike 15.4 izhaja, da je F-porazdelitev asimetrična v desno in da se stopnja asimetrije z večanjem m_1 in m_2 manjša.



Slika 15.4 F-porazdelitev

Tabelirane so kritične vrednosti za P = 0.20, 0,10, 0,05, 0,025, 0,01, 0,005 in 0,001. Te kritične vrednosti zadoščajo za praktično uporabo.

Iz opredelitve slučajnostne porazdelitve F sledi, da je

$$F_{1-P}(m_2, m_1) = \frac{1}{F_P(m_1, m_2)} \qquad (15.13)$$

S to zvezo iz tabeliranih kritičnih vrednosti zlahka dobimo kritične vrednosti za P = 0.80, 0,90, 0,95, 0,975, 0,09, 0,99 in 0,999, tako da vzamemo reciproke ustreznih kritičnih vrednosti pri zamenjanih stopinjah prostosti.

15.5 Zveze med teoretičnimi slučajnostnimi spremenljivkami, ki so zasnovane na normalni porazdelitvi. V nadaljevanju dajemo pregled zvez med slučajnostnimi spremenljivkami, ki so zasnovane na normalni porazdelitvi. Razen povzetka že znanih zvez podajamo še nekaj novih, ki slede iz opredelitev porazdelitev:

$$y_i =: N(My_i;\ \sigma_i^2);\quad u = \sum a_i y_i =: N\left(M_u = \sum a_i My_i;\ \sigma_u^2 = \sum a_i^2 \sigma_i^2\right) \qquad (15.14)$$

y_i neodvisni

$$z_i =: N(0,1) \qquad \sum_{i=1}^{n} z_i =: N(0,n) \qquad (15.15)$$

$$\sum_{1}^{m} z_i^2 =: \chi^2(m) \quad ; \quad z^2 =: \chi^2(1) \qquad \sqrt{\chi^2(1)} =: z \qquad (15.16)$$

$$\frac{\left(\sum_i z_i\right)^2}{n} =: \chi^2(1) \qquad (15.17)$$

$$\frac{z}{\sqrt{\chi^2/m}} =: t(m) \quad ; \qquad t(\infty) =: z \qquad (15.18)$$

$$t^2(m) = \frac{z^2}{\chi^2/m} =: F(1,m) \qquad (15.19)$$

$$\chi^2(m)/m =: F(m, \infty) \qquad (15.10)$$

15.6 Če iz populacije $y =: N(M, \sigma^2)$ izberemo slučajnostni vzorec z n enotami, velja:

$$\frac{\bar{y} - M}{\sigma} \cdot \sqrt{n} =: z \qquad (15.21)$$

$$\frac{\bar{y} - M}{s} \cdot \sqrt{n} =: t(m = n-1) \qquad (15.22)$$

$$\frac{(n-1)\, s^2}{\sigma^2} =: \chi^2 \; (m = n-1) \qquad (15.23)$$

Pri tem sta $\bar{y}$ in s^2 izračunana iz vzorca po obrazcih

$$\bar{y} = \frac{1}{n} Sy \quad , \quad s^2 = \frac{S(y - \bar{y})^2}{n-1} \qquad (15.24)$$

Iz gornjih treh obrazcev je lepo vidna zveza:

Ker je $z/\sqrt{\chi^2/m} =: t(m)$, dobimo iz 15.21 in 15.23

$$\frac{z}{\sqrt{\chi^2(m)/m}} = \frac{\frac{\bar{y} - M}{\sigma} n}{\sqrt{\frac{(n-1)s^2}{\sigma^2} /(n-1)}} = \frac{\bar{y} - M}{s} \cdot \sqrt{n} =: t(m = n-1) \qquad (15.25)$$

15.7 Če iz dveh populacij, ki imata enaki varianci

$y_1 =: N(M_1; \sigma^2)$, $y_2 =: N(M_2; \sigma^2)$ izberemo vzorca iz prve z n_1 iz druge z n_2 enotami in so:

$$\bar{y}_1 = \frac{1}{n_1} Sy_1 : \bar{y}_2 = \frac{1}{n_2} Sy_2 \quad : \quad s_1^2 = \frac{S(y_1 - \bar{y}_1)^2}{n_1 - 1} : s_2^2 = \frac{S(y_2 - \bar{y}_2)^2}{n_2 - 1} \qquad (15.26)$$

velja

$$\frac{\bar{y}_2 - \bar{y}_1 - (M_2 - M_1)}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} =: t(m = n_1 + n_2 - 2) \qquad (15.27)$$

$$s^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

15.8 Če iz dveh normalno porazdeljenih populacij z σ_1^2 in σ_2^2 izberemo vzorca z n_1 in n_2 enotami, velja

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = : F(m_1 ; m_2) \qquad (15.28)$$

Vse te zveze slede iz definicij za porazdelitve χ^2, t in F.

Razmične ocene z malimi vzorci

15.9 Če poznamo zakonitosti porazdeljevanja vzorčnih rezultatov, ki smo jih nakazali v prejšnjih odstavkih, moremo zlahka določiti meje zaupanja za parametre, ki jih ocenjujemo z malimi vzorci.

Ker je standardizirana normalna porazdelitev simetrična, zanjo velja $z_{1-P} = -z_P$. Zato moremo trditi: Če iz podatkov slučajnostnega vzorca izračunamo vzorčni izraz, ki se porazdeljuje standardizirano normalno, ta s tveganjem $\alpha = 2P$ leži v mejah med $-z_P$ in $+z_P$. S tveganjem $\alpha = P$ pa ni večji kot $+z_P$. Z enakim tveganjem ni manjši kot $-z_P$.

Podobno velja za izraze, ki se porazdeljujejo v t-porazdelitvi. Ker je t-porazdelitev simetrična, tudi za kritične vrednosti t velja $t_{1-P} = -t_P$. Zato moremo enako trditi: Če iz slučajnostnega vzorca izračunamo vzorčni izraz, ki se porazdeljuje v t-porazdelitvi, ta izraz s tveganjem $\alpha = 2P$ leži v mejah med $-t_P$ in $+t_P$. S tveganjem $\alpha = P$ pa ni večji kot $+t_P$ oziroma z enakim tveganjem $\alpha = P$ ni manjši kot $-t_P$.

Za izraze, ki se porazdeljujejo v χ^2-porazdelitvah, pa moremo trditi: Če iz slučajnostnega vzorca izračunamo vzorčni izraz, ki se porazdeljuje v χ^2-porazdelitvi, ta izraz s tveganjem $\alpha = 2P$ leži v mejah med χ^2_{1-P} in χ^2_P. S tveganjem $\alpha = P$ pa ni večji kot χ^2_P.

Za izraze, ki se porazdeljujejo v F-porazdelitvi, trdimo podobno: Če se vzorčni izraz porazdeljuje v F-porazdelitvi, je vrednost tega izraza s tveganjem $\alpha = P$ manjša kot je kritična vrednost F_P. Če to upoštevamo, moremo za vse parametre, ki nastopajo v vzorčnih izrazih za male vzorce, izračunati razmične ocene. Te razmične ocene dobimo tako, da ustrezne vzorčne izraze preuredimo, tako da ločimo ocenjevani parameter od znanih količin.

Ocenjevanje aritmetičnih sredin

15.10 Ocenjevanje aritmetične sredine, če poznamo σ.

Če je y porazdeljen normalno z znanim σ, ocenimo z razmično oceno aritmetično sredino po obrazcu

$$\bar{y} - Z_p \frac{\sigma}{\sqrt{n}} < M < \bar{y} + Z_p \frac{\sigma}{\sqrt{n}} \qquad (15.29)$$

15.11 Po tem postopku moremo rešiti npr. tale problem. Za stroj, za katerega vemo, da je standardni odklon za profile $\sigma = 2$ mm, ocenjujemo aritmetično sredino z vzorcem z $n = 50$ enotami. Iz podatkov vzorca dobimo, da je $\bar{y} = 28$ mm. Določiti je treba meji zaupanja s tveganjem $\alpha = 2P = 0,05$ za aritmetično sredino. Po obrazcu 15.29 dobimo

$$28 - 1,96 \frac{2}{\sqrt{50}} < M < 28 + 1,96 \cdot \frac{2}{\sqrt{50}}$$

$$27,45 \text{ mm} < M < 28,55 \text{ mm}$$

Včasih izrazimo razmično oceno tudi v obliki: $M = 28 \pm 0,55$ mm s tveganjem $\alpha = 0.05$.

15.12 Ocenjevanje aritmetične sredine, če σ ne poznamo.

Če se y porazdeljuje normalno, ne poznamo pa prave vrednosti za standardni odklon σ, ocenimo meji zaupanja za aritmetično sredino po obrazcu

$$\bar{y} - t_p(m = n-1) \frac{s}{\sqrt{n}} < M < \bar{y} + t_p(m = n-1) \frac{s}{\sqrt{n}} \qquad (15.30)$$

ki je izpeljan iz obrazca 15.22

Ker se t_p asimptotično približuje z_p, drugače pa je po pravilu večji, so razmiki zaupanja nekoliko širši, kar je razumljivo. V ocenah za meji zaupanja je namreč razen $\bar{y}$ tudi s_y ocenjen iz podatkov vzorca.

15.13 Vzemimo zgled 15.11 s to razliko, da smo standardni odklon σ ocenili iz podatkov vzorca, iz katerega smo ocenili tudi M. Iz vzorca n = 50 izdelkov smo dobili $\bar{y}$ = 28 mm in s_y = 2,1 mm. Ker je število stopinj prostosti m = n-1 = 50-1 = 49, najdemo v tabeli za t-porazdelitev $t_{P=0,025}\ (m = 49) = 2,01$. Če to upoštevamo, dobimo, da sta meji zaupanja za aritmetično sredino po obrazcu 15.30

$$28 - 2,01 \frac{2,1}{\sqrt{50}} < M < 28 + 2,01 \frac{2,1}{\sqrt{50}}$$

$$27,40 \text{ mm} < M < 28,60 \text{ mm}$$

15.14 Subjektivna ocena za aritmetično sredino. Določen parameter npr. aritmetično sredino moremo pogosto oceniti na osnovi subjektivnega poznavanja proučevanega pojava brez statističnega opazovanja (popisa ali vzorca). Tako moremo iz poznavanja o porabi v določenem času subjektivno oceniti, da je poprečna poraba za dani izdelek "okrog" m = 10 kg. Iz subjektivnega poznavanja problema presodimo, da je ocena "bolj ali manj verjetna", "bolj ali manj zanesljiva". Če štejemo, da je ocena manj verjetna, dopuščamo, da se prava vrednost M bolj odklanja od subjektivne ocene m kot bi se v primeru, če smatramo oceno m kot bolj zanesljivo. To zanesljivost subjektivne ocene merimo z $Var(M) = \sigma_M^2$. Varianco subjektivnih ocen določamo tudi subjektivno glede na to, kolikšno zanesljivost pripisujemo subjektivnim ocenam. Ker sta stopnja zanesljivosti in σ_M^2 obratno sorazmerna, vzamemo kot merilo zanesljivosti ali informiranosti $1/\sigma_M^2 \cdot \sigma_M^2$ ni v nikaki zvezi z variabilnostjo znaka v populaciji σ_y^2, temveč je le merilo za stopnjo zanesljivosti subjektivnih ocen. Če privzamemo, da se subjektivne ocene porazdeljujejo normalno

$$M =: N(m; \sigma_M^2) \qquad (15.31)$$

okrog prave vrednosti, moremo razen točkovne ocene m za aritmetično sredino podati tudi subjektivno razmično oceno

$$m - z_p \cdot \sigma_M < M < m + z_p \sigma_M \qquad (15.32)$$

ki velja s tveganjem $\alpha = 2P$

15.15 Vzemimo, da za določeno dobrino cenimo, da je poprečna poraba M okrog m = 10 kg. Dalje cenimo, da je standardni odklon subjektivne ocene σ_M=0,5 kg. Medtem ko je m = 10 kg subjektivna točkovna ocena za poprečno porabo, je razmična subjektivna ocena s tveganjem $\alpha = 2P = 0,05$ po obrazcu 15.32 :

$$10 - 1,96 \cdot 0,5 < M < 10 + 1,96 \cdot 0,5$$

$$9,02 \text{ kg} < M < 10,98 \text{ kg}$$

Če bi subjektivno oceno m = 10 kg šteli za manj zanesljivo, bi to izrazili z večjim σ_M. Če vzamemo, da je naša vera v dano oceno znatno manjša kot v prvem zgledu, moramo vzeti večji σ_M (npr. $\sigma_M = 2$). Po enakem postopku dobimo za ta zgled, da s tveganjem $\alpha = 0,05$ velja $6 < M < 14$.

15.16 Bayesov pristop za oceno aritmetične sredine. Iz gornjega razmišljanja spoznamo, da je logika subjektivnega ocenjevanja podobna ocenjevanju z vzorčenjem, le da pri vzorčenju subjektivno oceno zamenjamo z oceno iz vzorca, zanesljivost ocene pa merimo z varianco ocen $\text{Var}\,\bar{y} = \sigma_y^2/n$, ki je odvisna od variabilnosti v osnovni populaciji in od velikosti vzorca n. Iz istega razloga kot pri subjektivni oceni merimo stopnjo zanesljivosti oziroma informiranost z $1/\text{Var}\,\bar{y} = n/\sigma_y^2$. Iz ideje, da bi mogli subjektivno oceno m popraviti z rezultati, ki jih dobimo iz vzorca, dobimo kombinirano posteriorno-Bayesovo oceno m_B, ki je tehtano poprečje med subjektivno oceno m in oceno iz vzorca $\bar{y}$. Kot pondere uporabimo stopnje zanesljivosti ali informacije za obe oceni: $1/\sigma_M^2$ in n/σ_y^2

$$m_B = \frac{m \cdot \frac{1}{\sigma_M^2} + \bar{y} \cdot \frac{n}{\sigma_y^2}}{\frac{1}{\sigma_M^2} + \frac{n}{\sigma_y^2}} = \frac{m\sigma_y^2 + \bar{y} \cdot n\sigma_M^2}{\sigma_y^2 + n\sigma_M^2} \qquad (15.33)$$

Varianca za Bayesovo oceno $\text{Var}\,m_B$ pa je

$$\text{Var}(m_B) = \frac{1}{\frac{1}{\sigma_M^2} + \frac{n}{\sigma_y^2}} = \frac{\sigma_y^2 \cdot \sigma_M^2}{\sigma_y^2 + n\sigma_M^2} \qquad (15.34)$$

Če predpostavljamo normalnost in nepristranost za oceni m in $\bar{y}$, veljajo iste lastnosti tudi za Bayesovo oceno m_B

$$m_B =: N(M, \text{Var } m_B) \qquad (15.37)$$

Zakonitosti za subjektivno oceno poprečja m in za vzorčno oceno sta posebna primera Bayesove ocene. Če ocenjujemo brez vzorca, je n=o in ocena čisto subjektivna. Iz splošnih obrazcev 15.33 in 15.34 dobimo, da je za ta primer $m_B = m$ in $\text{Var } m_B = \sigma_M^2$. Obratno pa je, če subjektivni oceni ne pripisujemo nobene teže. Stopnja zanesljivosti subjektivne ocene $1/\sigma_M^2 = 0$. Zato je v primeru, če je ocena zasnovana samo na vzorčnih informacijah, $m_B = \bar{y}$ in $\text{Var } m_B = \frac{\sigma_y^2}{n}$.

Iz zakonitosti, ki je za Bayesovo oceno nakazana v 15.37, dobimo razmično oceno za aritmetično sredino po Bayesovem načelu

$$m_B - z_P \cdot \sqrt{\text{Var } m_B} < M < m_B + z_P \cdot \sqrt{\text{Var } m_B} \qquad (15.38)$$

15.17 Vzemimo, da smo za porabo razen subjektivne ocene m = 10 kg in $\sigma_M = 0,5$ kg zbrali informacije tudi z vzorcem n =48 gospodinjstev. Ocena iz vzorca je $\bar{y} = 11,0$ kg, predpostavljamo pa, da je $\sigma_y = 2$ kg. Stopnja informacije za subjektivno oceno je $1/\sigma_M^2 = 1/0,5^2 = 4$, stopnja informacije z vzorcem pa $n/\sigma_y^2 = \frac{48}{2^2} = 12$. Iz teh delnih rezultatov spoznamo, da je stopnja informacij iz vzorca znatno večja kot iz subjektivne ocene. Iz obrazca 15.34 spoznamo, da je stopnja informacije za kombinirano Bayesovo oceno vsota stopenj informacij za obe oceni

$$\frac{1}{\sigma_{mB}^2} = \frac{1}{\sigma_M^2} + \frac{n}{\sigma_y^2} . \qquad (15.39)$$

V našem zgledu je

$$1/\sigma_{m_B}^2 = 4 + 12 = 16$$

iz česar je $\sigma_{m_B}^2 = 1/16$ oziroma $\sigma_{m_B} = \sqrt{1/16} = 0.25$ kg

Standardni pogrešek za Bayesovo oceno je praviloma manjši kot za subjektivno ali vzorčno oceno.

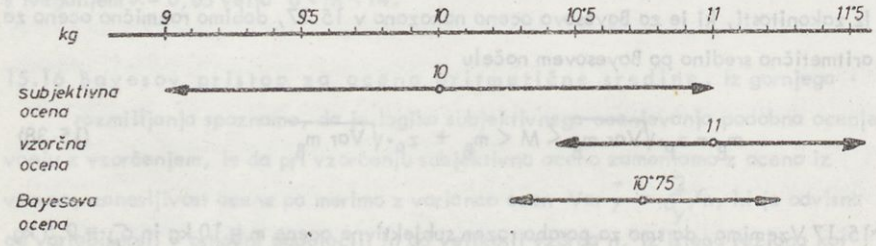
Po obrazcu 15.33 dobimo, da je Bayesova ocena za poprečno porabo

$$m_B = \frac{10.4 + 11.12}{4 + 12} = 10{,}75 \text{ kg}$$

Razmična Bayesova ocena za poprečno porabo pa je

$$10{,}75 - 1{,}96 . 0{,}25 < M < 10{,}75 + 1{,}96 . 0{,}25$$

$$10{,}26 \text{ kg} < M < 11{,}24 \text{ kg}$$



Slika 15.5 Razmična subjektivna, vzorčna in Bayesova ocena za poprečno porabo

15.18 Nakazani način za izračunanje Bayesovih ocen ne velja samo za aritmetično sredino, ampak moremo uporabiti enako sklepanje tudi za ocenjevanje drugih parametrov.

15.19 Ocenjevanje razlik med aritmetičnima sredinama- Neodvisna vzorca. Pri proučevanju množičnih pojavov je primerjava eno izmed osnovnih analitičnih sredstev. Dva parametra med seboj primerjamo tako, da poiščemo razliko ali razmerje med njima. Enako velja tudi za ocene. Vzemimo dve normalno porazdeljeni populaciji z različnima aritmetičnima sredinama M_1 in M_2 in z enako varianco σ^2. Če iz prve izberemo vzorec z n_1 enotami in iz druge vzorec z n_2 enotami, moremo iz podatkov obeh neodvisnih vzorcev oceniti razliko med aritmetičnima sredinama po obrazcu

$$\bar{y}_2 - \bar{y}_1 - t_p \cdot s \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} < M_2 - M_1 < \bar{y}_2 - \bar{y}_1 + t_p \cdot s \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \qquad (15.40)$$

$$t_p = t_p\,(m = n_1 + n_2 - 2): \quad s^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

ki je izpeljan iz zakonitosti, podane v 15.27 .

Po tem postopku moremo oceniti razliko v produktivnosti dela med spoloma, razlike v navadah v porabi za različne skupine prebivalstva, razlike med agrotehničnima ukrepoma, razlike v trpežnosti tkanin, izdelanih po različnih tehnoloških postopkih, razlike v prodaji pri različnem načinu embaliranja itd. Dostikrat vzamemo eno populacijo kot kontrolno oziroma en vzorec kot kontrolni in na rezultate kontrolnega vzorca primerjamo učinek pri raziskovanih pogojih.

15.20 Vzemimo za zgled podatke iz ankete o delavskih družinah v novembru 1957, ki jo je izvedel Zavod SR Slovenije za statistiko. Iz majhnega vzorca $n_1 = 12$ družin, ki imajo 15 do 20 tisoč starih dinarjev mesečnih dohodkov in iz vzorca $n_2 = 10$ družin, ki imajo 20 do 25 tisoč starih dinarjev dohodkov, ocenimo koliko s tveganjem $\alpha = P = 0{,}05$ najmanj znaša razlika v poprečnih izdatkih za kulturno in družbeno življenje med tema dvema skupinama.

Vzorca sta dala naslednje osnovne podatke:

Skupina 15-20 tisoč st.din : $n_1 = 12$

izdatki y_{1i} : 260 330 490 310 580 560 330 480 860 590 330 630

Skupina 20-25 tisoč st.din $n = 10$

izdatki y_{2i} : 650 1100 760 460 700 540 750 370 630 740

Da bi izračunali varianco s^2, moramo poznati oceni varianc s_1^2 in s_2^2 da bi izračunali točkovno oceno razlike med poprečjem pa $\bar{y}_1$ in $\bar{y}_2$.

Iz gornjih posamičnih podatkov smo izračunali vsoti podatkov in vsoti kvadratov.

$$\sum_i y_{1i} = 5750; \quad \sum_i y_{1i}^2 = 3095500; \quad \sum_i y_{2i} = 6700; \quad \sum_i y_{2i}^2 = 4847200$$

Iz njih dobimo dalje:

$\bar{y}_1 = 5750/12 = 479$ din $\qquad s_1^2 = \frac{3095500 - 5750^2/12}{12 - 1} = 30936$

$\bar{y}_2 = 6700/10 = 670$ din $\qquad s_2^2 = \frac{4847200 - 6700^2/10}{10 - 1} = 39800$

$$s^2 = \frac{(12-1)30936 + (10-1)\,39800}{12 + 10 - 2} = 34925$$

$$s = \sqrt{34925} = 187 \text{ din}$$

Ker se vprašamo, koliko najmanj znaša stvarna razlika v poprečnih izdatkih za kulturno in družbeno življenje med proučevanima skupinama, ocenimo razmik zaupanja enostransko in poiščemo tisto vrednost, nad katero je prava razlika med sredinama s tveganjem $\alpha = P = 0,05$. Ta pa je

$$\bar{y}_2 - \bar{y}_1 - t_P \cdot s \cdot \sqrt{\frac{n_1 + n_2}{n_1 n_2}} < M_2 - M_1 \qquad (15.41)$$

Točkovna ocena za razliko je $\bar{y}_2 - \bar{y}_1 = 670 - 479 = 191$ din

Ker je $t_{P=0,05}\ (m = n_1+n_2-2=12+10-2=20) = 1,72$, je

razmična ocena po obrazcu 15.41

$$191 - 1,72 \cdot 187 \cdot \sqrt{\frac{12 + 10}{12 \cdot 10}} = 53 \text{ din} < M_2 - M_1$$

S tveganjem $\alpha = P = 0,05$ je razlika v poprečni porabi najmanj 53 st.dinarjev.

15.21 Ocenjevanje razlik med sredinama: odvisna vzorca. Pri proučevanju odvisnosti izdatkov od višine dohodkov sta bila preskusna vzorca različna oziroma neodvisna. Učinek spremembe določenega dejavnika pa moremo pogosto ugotavljati na istih poskusnih enotah. Tako moremo pri ugotavljanju uspešnosti novega postopka dobiti potrebne informacije na istem vzorcu delavcev, za katere primerjamo učinek dela, če delajo po starem načinu, z učinkom, če delajo po novem načinu. Enako moremo preskušati učinek reklamiranja blaga na TV na vzorcu istih trgovin pred in po reklamira-

nju. Razen tega, da je pri tem načinu potreben en sam vzorec poskusnih enot, je prednost tudi v tem, da se učinek posamičnih dejavnikov, ki vplivajo na poskusne enote, znatno omilili, zaradi česar dá taka metoda zanesljivejše ocene. Varovati se moramo le, da se ne prikrade v rezultate sistematična napaka, ki izvira iz poznavanja, privajenosti itd. Če delajo isti delavci najprej po metodi A, si morejo pridobiti pri tem določene izkušnje, ki vplivajo na učinek pri delovni metodi B. Vsebinsko enak, vendar v obratni smeri, bi bil učinek utrujenosti ipd.

Pri nakazanem postopku izračunamo razmično oceno po tehle stopnjah.

a) Izberemo vzorec n enot.

b) Za vsako enoto merimo proučevan podatek pri pogoju 1

y_{1i} in pri pogoju 2 y_{2i} $\quad i = 1,2 \ldots n$

c) Za vsako enoto izračunamo ustrezne razlike

$$d_i = y_{2i} - y_{1i} \tag{15.42}$$

d) Iz d_i ocenimo poprečno razliko učinka

$$\bar{d} = \frac{1}{n} S\, d_i = \bar{y}_2 - \bar{y}_1 \tag{15.43}$$

in varianco razlik $s_d^2 = \dfrac{S(d - \bar{d})^2}{n - 1}$ (15.44)

e) Če računamo, da se y_{2i} in y_{1i} porazdeljujeta normalno, se tudi d_i porazdeljuje normalno. Zato moremo za $\bar{d}$ uporabiti obrazec za računanje razmičnih ocen za poprečje

$$\bar{d} - t_p \frac{s_d}{\sqrt{n}} < \Delta M = M_2 - M_1 < \bar{d} + t_p \frac{s_d}{\sqrt{n}} \quad ; \quad \alpha = 2P \tag{15.45}$$

pri čemer je kritična vrednost $t_p(m=n-1)$ določena z $m = n-1$ stopinjami prostosti.

15.22 Za preskus učinka določenega reklamiranja izdelka smo izbrali $n = 10$ trgovin enakega tipa. Dnevni promet za izdelek pred in po reklamiranju po trgovinah je

Trgovina	A	B	C	D	E	F	G	H	I	J
Pred reklamiranjem	80	103	95	76	115	130	72	90	95	77
Po reklamiranju	100	127	116	90	138	161	97	101	121	90
d_i	+ 20	+ 24	+ 21	+ 14	+ 23	+ 31	+ 25	+ 11	+ 26	+ 13

Iz dobljenih razlik d_i dobimo: $n = 10$, $Sd_i = 208$, $Sd_i^2 = 4694$

$$\bar{d} = \frac{1}{n} Sd_i = \frac{+208}{10} = 20,8$$

$$s_d^2 = \frac{Sd_i^2 - \frac{(Sd)^2}{n}}{n-1} = \frac{4694 - \frac{208^2}{10}}{10-1} = 40,8444 \; ; \; s_d = \sqrt{40,8444} = 6,39$$

Točkovna ocena poprečnega povečanja prometa je $\bar{d} = 20,8$.

Intervalna pa je po obrazcu 15.45

$$20,8 - 2,26 \frac{6.39}{\sqrt{10}} < \Delta M = M_2 - M_1 < 20,8 + 2,26 \frac{6,39}{\sqrt{10}}$$

$$16,2 < \Delta M < 25,4$$

15.23 Ocenjevanje variance . Če se osnovna populacija porazdeljuje normalno, je razmična ocena za varianco

$$\frac{(n-1)s^2}{\chi^2_{\frac{(n-1)}{1-P}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{P(n-1)}} \qquad (15.47)$$

izpeljana iz obrazca 15.23

15.24 Vzemimo, da za pošiljko jabolk ugotavljamo, kolikšna je stopnja sortiranosti sadja glede na težo sadežev. Pošiljka je boljša, čim manjše so razlike v teži za posamezna jabolka. Kot pokazovalec za stopnjo enovitosti vzemimo varianco za težo. V ta namen smo slučajnostno izbrali iz pošiljke $n = 25$ jabolk. Zanje ugotovimo, da je ocena za varianco $s^2 = 6,2$. Kakšni sta meji zaupanja s tveganjem $\alpha = 2P = 0,10$?

Ker je $n = 25$, je število stopinj prostosti $m = n-1 = 25-1 = 24$. V tabeli za χ^2-porazdelitev dobimo, da je $\chi^2_{.95}(m=24) = 13,85$ in $\chi^2_{.05}(m=24) = 36,42$. Če te podatke vnesemo v obrazec 15.47, dobimo

$$\frac{(25-1)\cdot 6,2}{36,42} < \sigma^2 < \frac{(25-1)\cdot 6,2}{13,85}$$

$$4,09 < \sigma^2 < 10,73$$

Meji zaupanja za standardni odklon dobimo, če poiščemo kvadratna korena iz mej zaupanja za varianco.

$$2,02 \text{ dkg} < \sigma < 3,30 \text{ dkg}$$

15.25 Ker menimo, da je natančnost za dobljeno oceno premajhna, vzemimo večji vzorec z npr. $n = 75$ sadeži. Za ta vzorec dobimo, da je ocena variance $s^2 = 7,6$. Meji zaupanja za to oceno izračunamo enako kot v gornjem zgledu po obrazcu 15.47. Ker je število stopinj prostosti $m = n-1 = 75-1 = 74$, zahtevamo pa rezultate z isto stopnjo tveganja $\alpha = 2P = 0,10$, moramo poiskati $\chi^2_{.95}(m=74)$ in $\chi^2_{.05}(m=74)$. Teh količin v tabeli za kritične vrednosti za χ^2 porazdelitev ni. Pač pa moremo za $m > 30$ izračunati kritične vrednosti za χ^2 prek standardiziranega odklona z po obrazcu 15.7. Za naš zgled dobimo, da je

$$\chi^2_{.95}(m=74) = \frac{1}{2}\left(\sqrt{2.74-1} - 1,645\right)^2 = 54,9$$

$$\chi^2_{.05}(m=74) = \frac{1}{2}\left(\sqrt{2.74-1} + 1,645\right)^2 = 94,8$$

ker je $z_{0,95} = -1,645$ in $z_{0,05} = +1,645$

Meji zaupanja sta dalje po obrazcu 15.47

$$\frac{(75-1)\cdot 7,6}{94,8} < \sigma^2 < \frac{(75-1)\cdot 7,6}{54,9}$$

$$5,93 < \sigma^2 < 10,24$$

$$2,44 \text{ dkg} < \sigma < 3,20 \text{ dkg}$$

Razmik zaupanja se je skrčil. Ocena je zanesljivejša.

15.26 Ocena razmerja med variancama. Z obrazcem 15.28 ocenimo razmerje med variancama dveh populacij, ki se normalno porazdeljujeta, po naslednjem postopku. Če je ocena variance z vzorcem z n_1 enotami za prvo populacijo s_1^2, za vzorec z n_2 enotami iz druge populacije pa s_2^2, velja za dvostranski razmik zaupanja

$$P_r\left[F_{1-P}(m_1=n_1-1, m_2=n_2-1) < \frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2} < F_P(m_1=n_1-1,\ m_2=n_2-1)\right] = 1-2P \qquad (15.48)$$

Če ta obrazec preuredimo in upoštevamo 15.13, dobimo, da velja

$$\frac{s_1^2/s_2^2}{F_P(m_1=n_1-1,\ m_2=n_2-1)} < \sigma_1^2/\sigma_2^2 < s_1^2/s_2^2 . F_P(m_1=n_2-1,\ m_2=n_1-1) \qquad (15.49)$$

s tveganjem $\alpha = 2P$

Paziti moramo, da sta stopinji prostosti na desni strani zaradi 15.13 zamenjani.

15.27 Stroj, za katerega smo z vzorcem z $n_1 = 50$ izdelki ugotovili, da je za profil ocena standardnega odklona $s_1 = 1,5$ mm, smo dali v popravilo. S preskusnim vzorcem $n_2 = 25$ izdelkov smo dobili, da je ocena za standardni odklon za izdelke po popravilu $s_2 = 0.5$ mm. Določiti je treba meji zaupanja za razmerje standardnih odklonov σ_1/σ_2 pred in po popravilu z $\alpha = 2P = 0.10$.

Meji zaupanja za razmerje med standardnima odklonoma σ_1/σ_2 dobimo iz obrazca 15.49 enostavno tako, da izračunamo za posamezne dele kvadratne korene. Tako dobimo za naš zgled

$$\frac{1,5/0,5}{\sqrt{F_{0.05}(m_1=49, m_2=24)}} < \frac{\sigma_1}{\sigma_2} < 1,5/0,5 \cdot \sqrt{F_{0,05}(m_1=24, m_2=49)};\quad \alpha = 2P = 0.10$$

$$\frac{1,5/0.5}{\sqrt{1.86}} < \frac{\sigma_1}{\sigma_2} < 1,5/0.5 \cdot \sqrt{1,74}$$

$$2.20 < \sigma_1 / \sigma_2 < 3,96$$

Nezvezne porazdelitve.

15.28 Nezvezne porazdelitve so porazdelitve verjetnosti za nezvezne slučajnostne spremenljivke. Nezvezna slučajnostna spremenljivka je npr. število neustreznih izdelkov v vzorcu n izdelkov. Enako je nezvezna slučajnostna spremenljivka strukturni delež neustreznih izdelkov v vzorcu $p = \frac{x}{n}$ ipd. Od nezveznih verjetnostnih porazdelitev bomo obravnavali: hipergeometrično, binomsko in Poissonovo. Te namreč najpogosteje pridejo v poštev pri statističnem proučevanju.

Hipergeometrična porazdelitev.

15.29 Pri slučajnostnem izboru enot brez ponavljanja iz končne populacije z obsegom N, enkrat izbrana enota nima več možnosti, da bi bila ponovno izbrana. Verjetnost, da s slučajnostnim izborom izberemo enoto z dano značilnostjo, se od izbora do izbora spreminja. Zato število enot x z dano značilnostjo, pri izboru n enot brez ponavljanja sledi zakonitosti hipergeometrične verjetnostne porazdelitve za x, ki je dana s

$$H(x/N, N_a, n) = \frac{\binom{N_a}{x}\binom{N - N_a}{n - x}}{\binom{N}{n}} \qquad (15.50)$$

Poleg že znanih oznak je H verjetnost, N_a = število enot z dano značilnostjo v osnovni populaciji.

Za hipergeometrijsko porazdelitev velja, da je

$$E(p=\frac{x}{n}) = \frac{N_a}{N} = P; \quad Var(p=\frac{x}{n}) = \frac{P.(1-P)}{n} \cdot \frac{N-n}{N-1} = \frac{N_a}{N}\frac{N-N_a}{N-1}\cdot\frac{1}{n}\cdot\frac{N-n}{N} = \frac{S_a^2}{n}\cdot\frac{N-n}{N}$$

(15.51)

Varianca za $p = \frac{x}{n}$ se v primerjavi z binomsko porazdelitvijo razlikuje za popravni faktor za končnost $\frac{N-n}{N-1}$

Pri populacijah z velikim obsegom N se verjetnost pri izboru od enote do enote relativno počasi spreminja. Zato se v primeru, da je obseg vzorca n v primerjavi z obsegom populacije N majhen, $p = \frac{x}{n}$ porazdeljuje približno v binomski porazdelitvi, če pa je n zadosti velik pa asimptotično v normalni porazdelitvi

$$p = \frac{x}{n} =: N\left[E(p) = \frac{N_a}{N};\ \mathrm{Var}\, p = \frac{N_a(N - N_a)}{N(N-1)} \cdot \frac{1}{n} \cdot \frac{N-n}{N}\right] \qquad (15.52)$$

15.30 Binomska-Bernoullijeva porazdelitev.

Proučujemo skupino n neodvisnih poskusov. Verjetnost, da se pri posameznem poskusu zgodi dogodek A naj bo za vse poskuse enaka P. Verjetnost sestavljenega dogodka, da se v skupini n poskusov zgodi dogodek A x-krat, je dana z obrazcem

$$B(x/n;P) = \binom{n}{x} P^x (1 - P)^{n-x} \qquad (15.53)$$

Porazdelitev se imenuje binomska, ker imajo verjetnosti formalno enako obliko kot členi pri razvitju binoma

$$1 = \left[P + (1 - P)\right]^n = \sum_{x=0}^{n} \binom{n}{x} P^x (1 - P)^{n-x} \qquad (15.54)$$

S tem je tudi pokazano, da je vsota verjetnosti za $x = 0, 1, 2 \ldots, n$ enaka 1.

Največja verjetnost je za celo vrednost x_{Mo}

$$(n+1).P - 1 < X_{Mo} < (n+1).P \qquad (15.55)$$

Modus v verjetnostni porazdelitvi je za $M_o = X_{Mo}$

Za slučajnostno spremenljivko x, ki se porazdeljuje v binomski porazdelitvi, veljajo parametri

slučajnostna spremenljivka x	slučajnostna spremenljivka $p = \frac{x}{n}$
$M(x) = E(x) = nP$	$M(p = \frac{x}{n}) = E(p = \frac{x}{n}) = P$
$Var(x) = nP \cdot Q$	$Var\ (p = \frac{x}{n}) = \frac{P \cdot Q}{n}$
$\gamma_1(x) = \frac{(Q - P)}{(nPQ)^{1/2}}$	$\gamma_1\ (p = \frac{x}{n}) = \frac{(Q - P)}{(nPQ)^{1/2}}$
$\gamma_2(x) = \frac{1 - 6PQ}{nPQ}$	$\gamma_2\ (p = \frac{x}{n}) = \frac{1 - 6PQ}{nPQ}$

(15.56)

$Q = 1 - P$

15.31 Iz gornjih parametrov spoznamo, da je $p = \frac{x}{n}$ nepristranska ocena za P in da se porazdelitev asimptotično približuje normalni. Iz γ_1 spoznamo, da je porazdelitev simetrična, če je P = 0,5 oziroma 50%, asimptotično simetrična pa, če je n velik. Prav tako je sploščenost za binomsko porazdelitev asimptotično enaka kot za normalno porazdelitev. Zato moremo vzeti, da pri velikem vzorcu n velja

$$x =: N(nP;\ nPQ)\ ;\ p = \frac{x}{n} =: N(P;\ \frac{PQ}{n}) \qquad (15.57)$$

Nakazana asimptotična zakonitost, ki povezuje binomsko porazdelitev z normalno, velja že za ne preveč velike vzorce, če P ni preveč različen od P = 0,50. Iz 15.57 moremo dalje povzeti, da za te primere velja

$$\frac{x - nP}{\sqrt{nPQ}} =: z\ ;\ \frac{\frac{x}{n} - P}{\sqrt{\frac{PQ}{n}}} =: z \qquad (15.58)$$

kar je v bistvu zakonitost 15.57, pisana v obliki standardizirane normalne porazdelitve.

Iz zvez 15.58 izvedemo, da je pri zgornjih omejitvah porazdelitvena funkcija za binomsko porazdelitev

$$\sum_{x=0}^{x=x_1} B(x/n;\ P) = 0{,}50 + H\left(z_1 = \frac{x_1 + 0.5 - nP}{\sqrt{nPQ}}\right) \qquad (15.59)$$

dana s porazdelitveno funkcijo za normalno porazdelitev.

0.5, ki se v obrazcu 15.59 pojavi v oklepaju na desni strani enačbe, izvira iz vskladitve nezveznosti binomske porazdelitve z zveznostjo pri normalni porazdelitvi.

15.32 Za zgled vzemimo binomsko porazdelitev z $n = 15$ in $P = 0,50$ in izračunajmo prek obrazca 15.59 verjetnost $\sum_{x=o}^{x=8} B(x/n=15;\ P=0,50)$

Po obrazcu 15.59 je

$$\sum_{x=o}^{x=8} B(x/n=15;\ P=0,50) = 0,50 + H\left(z = \frac{8+0,5-15.0,50}{\sqrt{15.0,50.0,50}}\right) = 0,5164$$

Z linearno interpolacijo tabeliranih vrednosti dobimo:

$H\ (z = 0,51) = 0,1950$

$\Delta H = 0,0035 \quad \Delta H \,.\, 0,64 = 0,0022$

$H\ (z = 0,52) = 0,1985$

$$H\ (z = 0,5164) = H(z = 0,51) + \Delta H \,.\, 0,64 = 0,1950 + 0,0022 = 0,1972$$

in dalje

$$\sum_{x=o}^{x=8} B(x/n=15;\ P = 0,50) = 0,5000 + 0,1972 = 0,6972$$

Za preskus skladnosti primerjamo dobljeni rezultat s pravo vrednostjo porazdelitvene funkcije iz tabele 15.1. Iz nje dobimo, da je $\sum_{x=o}^{8} B(x/n=15;\ P=0,50) = 0,6964$. Razlika $0,6972-0,6964=0,0008$ je zanemarljiva.

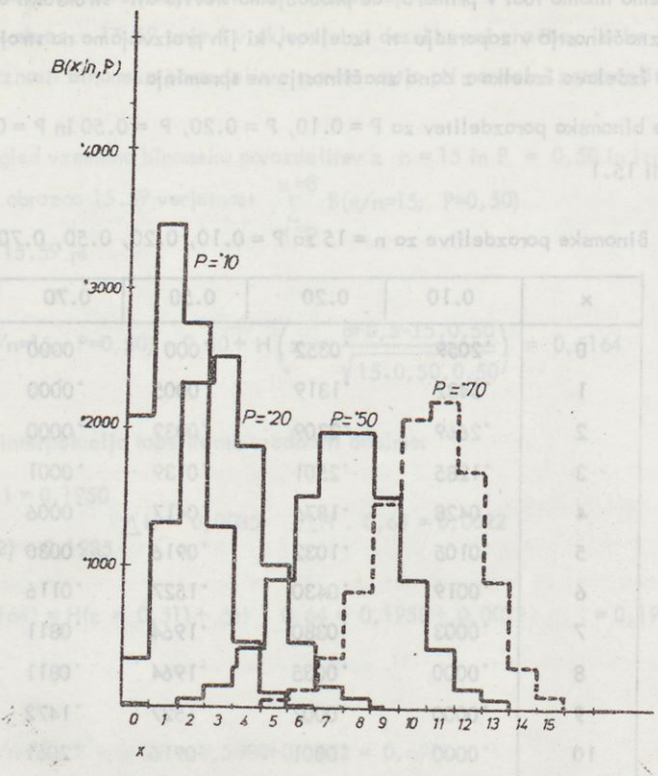
15.33 Zgledi, v katerih se slučajnostna spremenljivka porazdeljuje v binomski porazdelitvi. Osnovni pogoj, da se x oziroma $p = \frac{x}{n}$ iz sestavljenega dogodka porazdeljuje v binomski porazdelitvi, je, da se od poskusa do poskusa verjetnost dogodka ne spremeni. Če to upoštevamo, se v binomski porazdelitvi porazdeljuje število enot z dano značilnostjo, ki smo jih izbrali z vzorcem s ponavljanjem z obsegom n. Če gre za vzorec s ponavljanjem, pri vsakem poskusu izbrano enoto ponovno vključimo v populacijo in s tem odnose v populaciji ne spremenimo.

Binomsko shemo imamo tudi v primeru, če proučujemo število ali strukturni delež izdelkov z dano značilnostjo v zaporedju n izdelkov, ki jih proizvajamo na stroju, za katerega se P za izdelavo izdelka z dano značilnostjo ne spreminja.

Za n = 15 je binomska porazdelitev za P = 0.10, P = 0.20, P = 0.50 in P = 0.70 nakazana v tabeli 15.1

Tabela 15.1 Binomske porazdelitve za n = 15 za P = 0.10, 0.20, 0.50, 0.70

x	0.10	0.20	0.50	0.70
0	·2059	·0352	·000	·0000
1	·3432	·1319	·0005	·0000
2	·2669	·2309	·0032	·0000
3	·1285	·2501	·0139	·0001
4	·0428	·1876	·0417	·0006
5	·0105	·1032	·0916	·0030
6	·0019	·0430	·1527	·0116
7	·0003	·0380	·1964	·0811
8	·0000	·0035	·1964	·0811
9	·0000	·0007	·1527	·1472
10	·0000	·0001	·0916	·2061
11	·0000	·0000	·0417	·2186
12	·0000	·0000	·0139	·1700
13	·0000	·0000	·0032	·0916
14	·0000	·0000	·0005	·0305
15	·0000	·0000	·0000	·0047



Slika 15.6 Binomske porazdelitve za P = .10, .20, .50, .70

15.34 Razmična ocena za strukturni delež. Če je vzorec majhen, bi morali računati meje zaupanja za oceno strukturnega deleža prek binomske porazdelitve. Vendar je neposredno določanje mej zaupanja zamudno. Izkaže pa se, da moremo najti meje zaupanja za strukturne deleže iz majhnih vzorcev prek F-porazdelitve.

Če iz slučajnostnega vzorca z n enotami dobimo x enot z dano značilnostjo, določimo spodnjo in zgornjo mejo zaupanja $P_s\%$ in $P_z\%$ za strukturni odstotek P% po obrazcih

$$P_s\% = \frac{100x}{x+(n-x+1)F_s} \quad ; \quad P_z\% = \frac{100(x+1)F_z}{(x+1)F_z+(n-x)} \qquad (15.60)$$

Pri tem pomeni: $F_s = F_P\left[m_1=2(n-x+1),\ m_2=2x\right]$; $F_z=F_P\left[m_1=2(x+1),\ m_2=2(n-x)\right]$

n = velikost vzorca in x = število enot z dano značilnostjo v vzorcu. Prava vrednost strukturnega deleža P% je v razmiku

$$P_s\% < P\% < P_z\% \qquad (15.61)$$

s tveganjem $\alpha = 2P$

15.35 Vzemimo, da smo iz populacije delavcev v nekem podjetju na slepo izbrali n = 30 delavcev. Od teh je x = 19 izpolnilo določeno nalogo v predpisanem času. Kateri sta meji zaupanja za odstotek delavcev, ki so sposobni, da rešijo proučevani problem v predpisanem času? Stopnja tveganja naj bo $\alpha = 2P = 0.10$. Po nakazanem postopku moramo najprej v tabeli za F-porazdelitev poiskati kritični vrednosti

$$F_s = F_{0.05}\left[m_1 = 2\,(30-19+1) = 24\,;\ m_2 = 2 \cdot 19 = 38\right] = 1{,}81$$
$$F_z = F_{0.05}\left[m_1 = 2(19+1) = 40\,;\ m_2 = 2\,(30-19) = 22\right] = 1{,}94$$

Te vrednosti dobimo z linearno interpolacijo, ki v prvem približku zadostuje. Če nakazane vrednosti vnesemo v obrazca 15.60, dobimo

$$P_S\% = \frac{100.19}{19+(30-19+1).1{,}81} = 47\%\,;\quad P_z\% = \frac{100.(19+1)\,.\,1{,}94}{(19+1).1{,}94+\,(30-19)} = 78\%$$

Prava vrednost za odstotek delavcev je s tveganjem $\alpha = 10$ med 47% in 78%.

Poissonova porazdelitev.

15.36 Če je v Bernoullijevem procesu P zelo majhen, n pa zelo velik, in gre v limiti $P \to 0$ in $n \to \infty$ produkt nP pa proti konstanti λ

$$\lim_{n\to\infty,\,P\to 0} nP = \lambda \qquad (15.62)$$

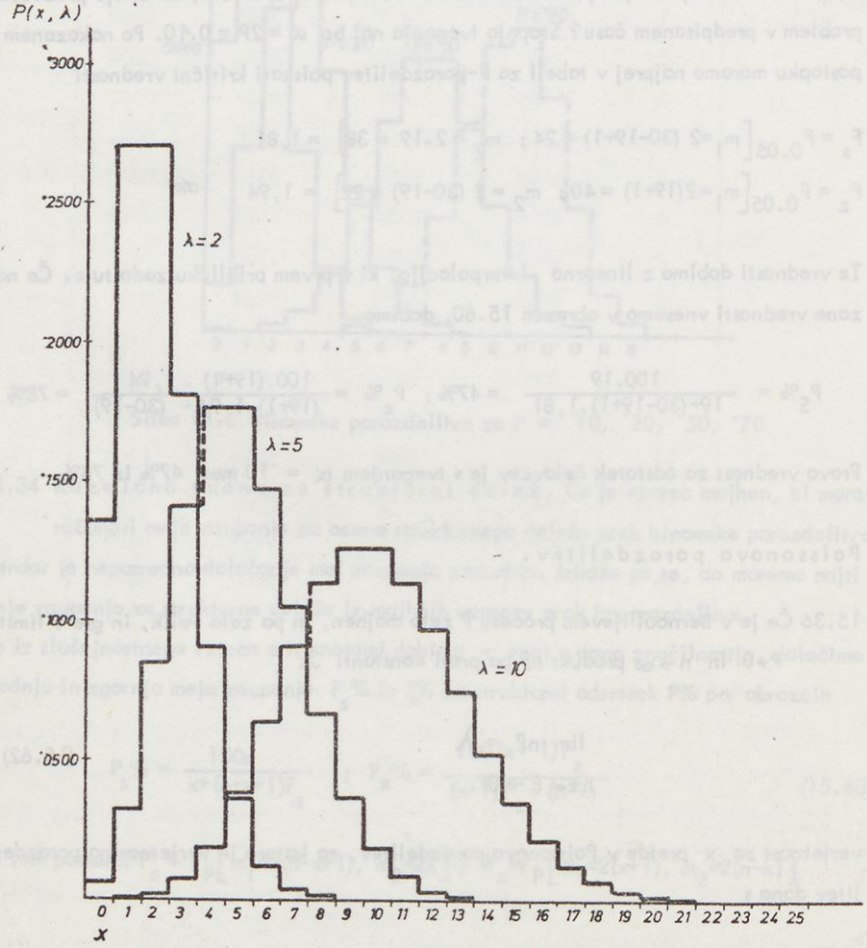
verjetnost za x preide v Poissonovo porazdelitev, za katero je verjetnostna porazdelitev dana s

$$P(x;\lambda) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \qquad (15.63)$$

Ob teh pogojih preidejo parametri binomske porazdelitve za Poissonovo porazdelitev v naslednjo obliko

$$M_x = E(x) = \lambda \; ; \; \mathrm{Var}\, x = \lambda \; ; \; \gamma_1 = \frac{1}{\sqrt{\lambda}} \; ; \; \gamma_2 = \frac{1}{\lambda} \qquad (15.64)$$

Tudi za Poissonovo porazdelitev sta γ_1 in γ_2 majhna če je λ velik, zaradi česar velja, da je Poissonova porazdelitev asimptotično normalna.



Slika 15.7 Poissonova porazdelitev za $\lambda = 2$, $\lambda = 5$ in $\lambda = 10$

Za $\lambda = 2$, $\lambda = 5$ in $\lambda = 10$ so Poissonove porazdelitve prikazane v tabeli 15.2

Tabela 15.2 Poissonove porazdelitve za $\lambda = 2$, $\lambda = 5$, $\lambda = 10$

x	$\lambda = 2$	$\lambda = 5$	$\lambda = 10$	x	$\lambda = 10$
0	·1353	·0067	·0000	16	·0217
1	·2707	·0337	·0005	17	·0128
2	·2707	·0842	·0023	18	·0071
3	·1804	·1404	·0076	19	·0037
4	·0902	·1755	·0189	20	·0019
5	·0361	·1755	·0378	21	·0009
6	·0120	·1462	·0631	22	·0004
7	·0034	·1044	·0901	23	·0002
8	·0009	·0653	·1126	24	·0001
9	·0002	·0363	·1251		
10	·0000	·0181	·1251		
11		·0082	·1137		
12		·0034	·0948		
13		·0013	·0729		
14		·0005	·0521		
15		·0002	·0347		

Tako tabela kot slika kažeta, da je z večanjem λ Poissonova porazdelitev bolj in bolj simetrična in zvonasta - skratka podobna normalni.

15.37 Praktičnih zgledov za slučajnostne spremenljivke s Poissonovo verjetnostno porazdelitvijo najdemo veliko na različnih področjih. Število drobnih napak na sestavljenih izdelkih je praktično majhno (0, 1,2,3,...), teoretično pa je možnih veliko napak. Če sestavljene izdelke proizvajamo ob enakih pogojih, se število napak porazdeljuje v Poissonovi porazdelitvi. λ pri tem pomeni poprečno število napak na sestavljenih izdelkih.

Število tipkarskih napak na posameznih straneh tipkanega teksta je majhno, čeprav je

možnih toliko napak, kolikor je avtorskih znakov na eni strani teksta. Zato se število napak na straneh porazdeljuje v Poissonovi porazdelitvi, če je tipkanje izvedeno ob enakih pogojih (ista tipkarica, enaki delovni pogoji ipd.). λ v tem zgledu pomeni poprečno število napak na eno stran.

Pri enaki gostoti prometa se število avtomobilov, ki se ustavijo na bencinski črpalki v določenem stalnem razdobju, porazdeljuje v Poissonovi porazdelitvi z λ, ki je enak poprečnemu številu avtomobilov v določenem časovnem razmiku.

Število napak v tkanju je na 1 m^2 tkanine se porazdeljuje v Poissonovi porazdelitvi, če je bila proučevana tkanina proizvedena ob enakih pogojih. λ pomeni poprečno število napak na 1 m^2.

5.38 Zaradi lastnosti, da se za velike λ Poissonova porazdelitev približuje normalni in iz zveze, da je $M(x) = \lambda$ in $Var(x) = \lambda$, moremo vzeti iz vzorca dobljen x kot točkovno oceno za λ in $\sqrt{x}$ kot oceno za standardni pogrešek za oceno X.
Tako dobimo, da velja asimptotično

$$Pr\,(x - z_p\sqrt{x} < \lambda < x + z_p\sqrt{x}) = 1 - 2P = 1 - \alpha \qquad (15.65)$$

Da bi bil x zadosti velik, da moremo uporabiti normalno aproksimacijo, moremo zaradi adicijskega teorema za Poissonovo porazdelitev postopati pri določanju ocene tako, da vzamemo v račun več vzorcev. Če smo za k enot ugotovili skupno x napak, je poprečje napak na enoto enako $\frac{x}{k}$, verjetnostna enačba pa preide v tem primeru v

$$Pr\left\{\frac{x}{k} - z_p\frac{\sqrt{x}}{k} < \lambda < \frac{x}{k} + z_p\frac{\sqrt{x}}{k}\right\} = 1 - 2P = 1 - \alpha \qquad (15.66)$$

$\frac{x}{k}$ je točkovna ocena za λ (število napak na enoto).

15.39 Če smo npr. za k = 20 m^2 ugotovili skupno x = 70 napak, moremo izračunati meji zaupanja po obrazcu 15.66

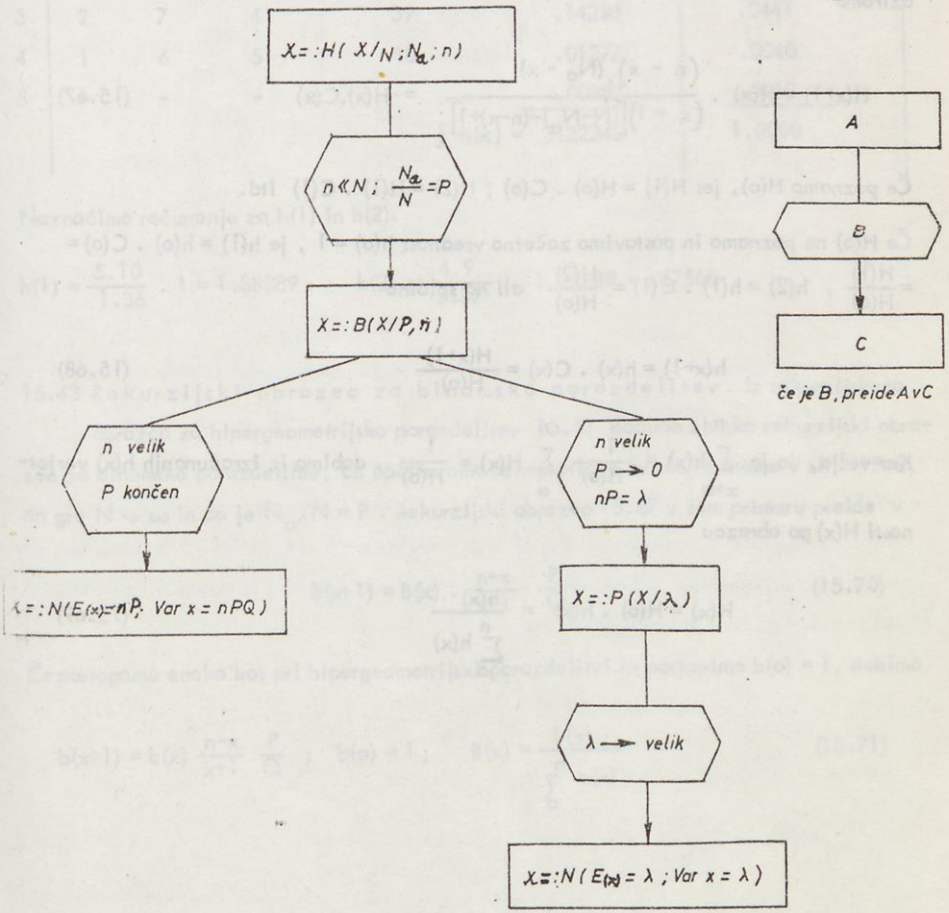
$$\frac{70}{20} - 1{,}96\ \frac{\sqrt{70}}{20} < \lambda < \frac{70}{20} + 1{,}96\ \frac{\sqrt{70}}{20}$$

$$2{,}68 < \lambda < 4{,}32$$

Poprečno število napak je z $\alpha = 0{,}05$ med 2,68 in 4,32 na m^2.

15.40 Zveze med nezveznimi porazdelitvami

Zveze med nezveznimi porazdelitvami moremo strniti v shemo 15.8.



Slika 15.8 Shema asimptotičnih zvez med nezveznimi porazdelitvami

15.41 Rekurzijski obrazci za nezvezne verjetnostne porazdelitve

Rekurzijski obrazec za hipergeometrijsko porazdelitev

Iz opredelitve za hipergeometrijsko porazdelitev v obrazcu 15.50 sledi

$$H(x+1) = \frac{\binom{N_a}{x+1}\cdot\binom{N-N_a}{n-x-1}}{\binom{N}{n}} = \frac{\binom{N_a}{x}\cdot\binom{N-N_a}{n-x}}{\binom{N}{n}} \cdot \frac{n-x}{x+1} \cdot \frac{N_a - x}{(N-N_a)-(n-x)+1}$$

oziroma

$$H(x+1) = H(x) \cdot \frac{(n-x)\cdot(N_a - x)}{(x+1)\left[(N-N_a)-(n-x)+1\right]} = H(x).C(x) \qquad (15.67)$$

Če poznamo H(o), je: $H(1) = H(o) \cdot C(o)$; $H(2) = H(1) \cdot C(1)$ itd.

Če H(o) ne poznamo in postavimo začetno vrednost $h(o) = 1$, je $h(1) = h(o) \cdot C(o) =$ $= \frac{H(1)}{H(o)}$, $h(2) = h(1) \cdot C(1) = \frac{H(2)}{H(o)}$ ali na splošno

$$h(x+1) = h(x) \cdot C(x) = \frac{H(x+1)}{H(o)} \qquad (15.68)$$

Ker velja, da je $\sum_{x=o}^{n} h(x) = \frac{1}{H(o)} \sum_{o}^{n} H(x) = \frac{1}{H(o)}$, dobimo iz izračunanih h(x) verjetnosti H(x) po obrazcu

$$H(x) = H(o) \cdot h(x) = \frac{h(x)}{\sum_{o}^{n} h(x)} \qquad (15.69)$$

15.42 Vzemimo za zgled hipergeometrijsko porazdelitev za vzorec brez ponavljanja z $n = 5$ iz populacije z $N_a = 10$ in $N = 50$

Tabela 15.3 Izračunanje verjetnosti za hipergeometrijsko porazdelitev $H(x/N_a = 10;\ N = 50;\ n = 5)$

x	n-x / 5-x	N_a-x / 10-x	x+1 / x+1	N-N_a-(n-x)+1 / 36+x	h(x)	H(x) = H(x)/$\sum$ h(x)
0	5	10	1	36	1.	·3102
1	4	9	2	37	1.38889	·4309
2	3	8	3	38	.67568	·2096
3	2	7	4	39	.14225	.0441
4	1	6	5	40	.01277	·.0040
5	-	-	-	-	.00383	.0012
					$\sum h(x) =$ 3.22342	1.0000

Naznačimo računanje za h(1) in h(2):

$$h(1) = \frac{5.10}{1.36} \cdot 1 = 1.38889; \qquad h(2) = \frac{4.9}{2.37} \cdot 1{,}38889 = .67568$$

15.43 Rekurzijski obrazec za binomsko porazdelitev. Iz rekurzijskega obrazca za hipergeometrijsko porazdelitev 15.67 dobimo zlahka rekurzijski obrazec za binomsko porazdelitev, če upoštevamo asimptotične zveze iz sheme v sliki 15.8, da gre $N \to \infty$ in da je $N_a/N = P$. Rekurzijski obrazec 15.67 v tem primeru preide v

$$B(x+1) = B(x) \cdot \frac{n-x}{x+1} \frac{P}{Q} \qquad (15.70)$$

Če postopamo enako kot pri hipergeometrijski porazdelitvi in postavimo $b(o) = 1$, dobimo

$$b(x+1) = b(x) \frac{n-x}{x+1} \frac{P}{Q}\ ;\quad b(o) = 1\ ;\qquad B(x) = \frac{b(x)}{\sum_{o}^{n} b(x)} \qquad (15.71)$$

15.44 Za zgled izračunajmo verjetnosti za x za binomsko porazdelitev z n=15 in P=0.10

Tabela 15.4 Računanje verjetnosti za binomsko porazdelitev z n = 15 in P = 0.10

x	n-x	x+1	P/Q	b(x)	B(x) = b(x)/$\sum$ b(x)
x	15-x	x+1	1/9		
0	15	1	1/9	1.	.2059
1	14	2	1/9	1.66667	.3432
2	13	3	1/9	1.29630	.2669
3	12	4	1/9	.62414	.1285
4	11	5	1/9	.20805	.0428
5	10	6	1/9	.05086	.0105
6	9	7	1/9	.00942	.0019
7	8	8	1/9	.00135	.0003
8	7	9	1/9	.00015	.0000
9	6	10	1/9	.00001	.0000
			$\sum_0^n$ b(x)	= 4.85695	1.0000

Dobljene vrednosti so skladne z verjetnostmi za binomsko porazdelitev za n = 15 in P = 0.10 iz tabele 15.1

15.45 Rekurzijski obrazec za Poissonovo porazdelitev. Če vzamemo, da gre $P \to 0$, $n \to \infty$ in $\lim nP = \lambda$, preide po shemi v sliki 15.8 binomska porazdelitev v Poissonovo porazdelitev. Z istim procesom dobimo iz zveze 15.70 rekurzijski obrazec za Poissonovo porazdelitev

$$P(x+1) = P(x)\frac{\lambda}{x+1} \tag{15.72}$$

oziroma

$$p(x+1) = p(x)\frac{\lambda}{x+1}\ ; \quad p(0) = 1\ ; \quad P(x) = \frac{p(x)}{\sum_0^n p(x)} \tag{15.73}$$

15.46 Za zgled izračunajmo verjetnosti za x za Poissonovo porazdelitev z $\lambda = 2$

Tabela 15.5 Računanje verjetnosti za Poissonovo porazdelitev z $\lambda = 2$

x	$\frac{\lambda/(x+1)}{2/(x+1)}$	p(x)	$P_i(x) = p(x)/\Sigma p(x)$
0	2/1	1.	.1353
1	2/2	2.	.2707
2	2/3	2.	.2707
3	2/4	1.33333	.1804
4	2/5	.66667	.0902
5	2/6	.26667	.0361
6	2/7	.08889	.0120
7	2/8	.02540	.0034
8	2/9	.00635	.0009
9	2/10	.00141	.0002
10	2/11	.00028	.0000
11	2/12	.00005	.0000
12	2/13	.00001	.0000
13	2/14	.00000	
		$\sum p(x) = 7.38906$	1.0000

Rezultat je skladen s podatki iz tabele 15.2

ŠESTNAJSTO POGLAVJE

16. PRESKUŠANJE DOMNEV

16.1 Informacije o populacijah smo dobili bodisi iz podatkov celotne populacije ali ocenjevali z vzorci. Razen nakazanih načinov kvantitativnega proučevanja pa v praksi pogosto postopamo tudi drugače. O proučevanem pojavu postavimo določene domneve , z opazovanjem pa postavljene domneve preskušamo, ali veljajo ali ne.

Domnevo , da velja vremenska napoved za določen dan, preverimo z opazovanjem vremena v napovedanem dnevu. Če se napoved sklada z dejanskim vremenom, je domneva potrjena, v nasprotnem primeru pa zavrnjena.

Če o določeni skupini izdelkov postavimo domnevo, da je odstotek izmeta manjši kot P % = 5%, s pregledom vseh izdelkov preverimo, ali ta domneva velja ali ne in domnevo sprejmemo ali zavrnemo. Domnevo, da je mesečna poprečna poraba za določeno potrošno blago v štiričlanskih gospodinjstvih v SRS večja kot M = 4 kg, preskusimo tako,da popišemo vsa štiričlanska gospodinjstva in izračunamo poprečje. S primerjavo dobljenega poprečja z domnevnim poprečjem ugotovimo, ali domneva velja ali ne.

Izkaže pa se, da so v večini primerov za preskus domneve zadostne informacije, ki jih dobimo z delnim opazovanjem populacij. Še celo več. Za uspešen preskus domnev so v splošnem zadostni znatno manjši vzorci kot pa za ocenjevanje parametrov. Kot pri ocenjevanju statističnih parametrov tudi pri preskušanju domnev vsi sklepi veljajo z večjim ali manjšim tveganjem.

Uporabnost sklepov s preskušanjem domnev je tolikšna, da je velik del statistične analize zasnovan na preskušanju domnev. Ti izsledki so bodisi končni rezultati za odločitev ali opora za nadaljnje proučevanje.

16.2 Osnovna in ničelna domneva. Domneve ima smisel preskušati le, če je možnih več, najmanj pa dvoje stanj. Če je možen en sam izid, preskušanje te domneve nima smisla, ker vnaprej vemo, kakšen bo rezultat.

Zato domnevi, ki jo preskušamo, - imenujmo jo osnovno domnevo H_1, priredimo nasprotno domnevo, - imenujmo jo ničelno domnevo H_o. Zavrnitev ene naj bi pomenila sprejem druge. Zavrnitev ničelne domneve H_o ima npr. za posledico, da sprejmemo osnovno domnevo H_1. Naštejmo nekaj primerov osnovnih in ničelnih domnev.

Osnovni domnevi $H_1 : P = P_1 = 5\ \%$, da je delež slabih izdelkov v proizvodnji enak 5%, postavimo alternativno ničelno domnevo $H_o : P = P_o = 3\ \%$, da je delež slabih izdelkov enak 3%.

Osnovno domnevo $H_1 : M > M_o = 10$, da je poprečna produktivnost po modernizaciji proizvodnega postopka večja kot pred modernizacijo, preskušamo proti ničelni domnevi $H_o : M = M_o = 10$, da je poprečna produktivnost enaka $M = 10$, tj. produktivnosti pred modernizacijo.

Osnovno domnevo H_1, da način prodaje vpliva na višino prodaje, preskušamo z ničelno domnevo H_o, da višina prodaje ni odvisna od načina prodaje.

Osnovni domnevi $H_1 : \sigma_1^2 < \sigma_2^2$, da se je varianca za določeno značilnost izdelkov na stroju, ki smo ga popravili, po popravilu zmanjšala, priredimo ničelno domnevo $H_o : \sigma_1^2 = \sigma_2^2$, da se varianca s popravilom ni spremenila.

Osnovni domnevi, da dejavnik A vpliva na proizvodnjo, ustreza ničelna domneva, da A ne vpliva na proizvodnjo.

Osnovni domnevi, da je poprečna poraba določene življenjske dobrine odvisna od socialno-ekonomske skupine $H_1 : M_1 \neq M_2 \neq M_3 \neq M_4$, preskušamo proti ničelni domnevi $H_o : M_1 = M_2 = M_3 = M_4$, da je poraba za vse skupine enaka. Zavrnitev ni-

čeine domneve pomeni sprejem osnovne domneve.

16.3 Enostavne in sestavljene domneve. Statistično preskušamo domneve o vrednostih in odnosih med parametri za eno ali več populacij. Glede na to, ali domneva vsebuje eno samo ali več vrednosti, govorimo o enostavnih in sestavljenih domnevah. V prvem zgledu o deležu slabih izdelkov sta tako osnovna kot ničelna domneva enostavni, ker obe (osnovna H_1: $P = P_1 = 5\%$ kot ničelna H_0: $P = P_0 = 3\%$) domnevi vsebujeta po eno samo vrednost. Pri zgledu o produktivnosti dela pa je enostavna le ničelna domneva H_0: $M = M_0 = 10$, medtem ko je osnovna domneva H_1: $M > M_0 = 10$ sestavljena, ker vsebuje vse vrednosti parametra M na razmiku $M > M_0 = 10$. Podobno ugotovimo ali so domneve enostavne ali sestavljene za vse druge zglede v 16.2.

16.4 Preskušanje enostavnih domnev. Prvine preskušanja domnev ponazorimo na zgledu, v katerem sta H_0 in H_1 enostavni domnevi o aritmetičnih sredinah.

Preskušamo domnevo H_1, da je pošiljka izdelkov izdelana iz surovine S_1, pri kateri je poprečna trdnost H_1: $M = M_1 = 21,5$ proti alternativni ničelni domnevi, da so izdelki izdelani iz surovine S_0, pri kateri je poprečna trdnost H_0: $M = M_0 = 20$.
Predpostavljamo, da se trdnost y pri izdelkih iz katerekoli surovine porazdeljuje normalno s $\sigma = 2$.
Vzemimo, da dani domnevi preskušamo s preskusnim vzorcem z $n = 16$ izdelki in poglejmo, kakšne sklepe moremo napraviti iz njega o domnevah H_0 in H_1.

Za aritmetične sredine vzorcev iz normalnih populacij vemo, da se porazdeljujejo normalno

$$\bar{y} =: N\left(M; \frac{\sigma^2}{n}\right) \qquad (16.1)$$

Če velja ničelna domneva H_0: $M = M_0 = 20$, se v našem zgledu porazdeljujejo aritmetične sredine $\bar{y}$ iz preskusnih vzorcev po 16.1

$$\bar{y} =: N\left[M = M_0 = 20;\ \mathrm{Var}\,\bar{y} = \frac{2^2}{16} = 0.25\right]$$

Iz teorije vzorčenja vemo, da v tem primeru velja

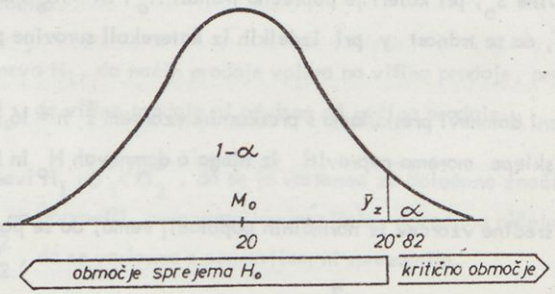
$$P_r(\bar{y} < \bar{y}_z = M_o + z_p \frac{\sigma}{\sqrt{n}}) = 1 - P = 1 - \alpha \qquad (16.2)$$

Če vzamemo, da je v našem zgledu $\alpha = 0,05$, velja

$$P_r(\bar{y} < \bar{y}_z = 20 + 1,645 \frac{2}{\sqrt{16}} = 20,822) = 1 - 0,05 = 0.95$$

Glede na to sklepamo, da je v primeru, če velja H_o, velika verjetnost ($1 - \alpha = 0,95$), da je poprečje $\bar{y}$ iz preskusnega vzorca manjše kot $\bar{y}_z = 20,822$ in majhna verjetnost ($\alpha = 0,05$), da je $\bar{y}$ iz preskusnega vzorca večji kot $\bar{y}_z = 20.822$. Zato je umestno, da ničelno domnevo sprejmemo, če je $\bar{y} \leq \bar{y}_z$ in zavrnemo, če je poprečje preskusnega vzorca $\bar{y} > \bar{y}_z$, ker je malo verjetno, da bi v tem primeru veljala H_o.
Razmik $\bar{y} \leq \bar{y}_z$, v katerem sprejmemo ničelno domnevo H_o, imenujemo **območje sprejema** za ničelno domnevo, razmik $\bar{y} > \bar{y}_z$, ki opredeljuje razmik za $\bar{y}$, v katerem zavrnemo ničelno domnevo, pa **kritično območje**. Vrednost $\bar{y}_z$, ki razmejuje območje sprejema od kritičnega območja, imenujemo **kritično mejo**.

Opisana situacija je nazorno razvidna iz grafikona za primer, da velja ničelna domneva H_o.



Slika 16.1 Situacija v primeru, če velja $H_o : M = M_o = 20$

16.5 **Napaka I. vrste in napaka II. vrste**. Iz gornjega zgleda spoznamo, da je možno, da je poprečje iz preskusnega vzorca $\bar{y}$ večje od kritične meje

$\bar{y}_z$ = 20,822, kljub temu, da velja H_o. Po nakazanem pravilu v tem primeru ničelno domnevo zavrnemo, čeprav velja. Sklep je torej napačen. Napako, da ničelno domnevo zavrnemo, kljub temu, da velja H_o, imenujemo napako I. vrste. Verjetnost, da napravimo napako prve vrste, je v našem zgledu 0,05. Dogovorno zaznamujemo verjetnost za napako I. vrste z α. Iz zgleda je razvidno, da kritično mejo določamo glede na vnaprej predpisano velikost α.

Proučimo še situacijo, če velja osnovna domneva $H_1 : M = M_1 = 21.5$. Če ohranimo pravilo za preskus ničelne domneve oziroma kritično mejo $\bar{y}_z$ = 20,822, je verjetnost, da je poprečje iz preskusnega vzorca $\bar{y}$ v kritičnem območju, enaka

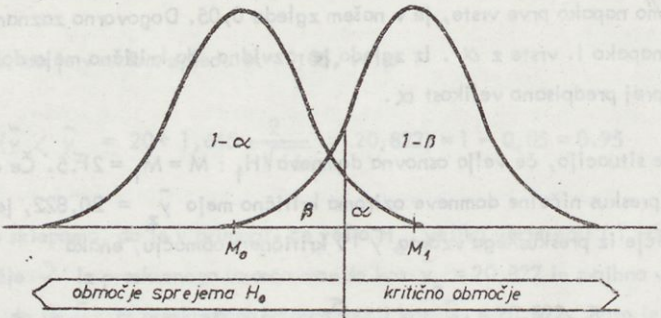
$$Pr\left(\bar{y} > \bar{y}_z = M_1 - z_{1-\beta}\frac{\sigma_y}{\sqrt{n}}\right) = 1 - \beta \qquad (16.3)$$

V verjetnostni neenačbi 16.3 je neznan edino $z_{1-\beta}$ oziroma β. Zato ga moremo izračunati iz zveze

$$z_{1-\beta} = \frac{\bar{y}_z - M_1}{\sigma_y}\sqrt{n} = \frac{20,822 - 21,5}{2}\sqrt{16} = -1,356$$

Iz $z_{1-\beta} = -1,356$ dobimo iz tablic za standardizirano normalno porazdelitev $(1-\beta)_{z=-1,356}$ = 0,912. Verjetnost, da je poprečje iz preskusnega vzorca v kritičnem območju, če velja $H_1 : M = M_1 = 21,5$ je torej enaka $1 - \beta = 0,912$. Ker je razmeroma velika verjetnost, da je $\bar{y}$ v kritičnem območju, če velja H_1 in razmeroma majhna, če velja H_o, upravičeno zavrnemo ničelno domnevo, če je $\bar{y}$ v kritičnem območju in sprejmemo osnovno domnevo H_1. Pri proučitvi odnosov pa spoznamo, da je kljub temu, da velja domneva H_1, možno, da je $\bar{y}$ v območju sprejema ničelne domneve. Sklep, da velja v tem primeru ničelna domneva H_o in ne H_1, je torej napačen. Napako, da zavrnemo osnovno domnevo H_1, čeprav ta velja, imenujemo napako II. vrste. Verjetnost, da napravimo napako II. vrste, dogovorno zaznamujemo z β. Medtem ko verjetnost za napako I. vrste določimo vnaprej, verjetnost β za napako II. vrste dobimo iz podatkov o preskusu. V našem zgledu je verjetnost β odvisna od vrednosti M_1 pri H_1, velikosti vzorca n in prek kritične meje $\bar{y}_z$ od M_o pri H_o in verjetnosti za napako I.

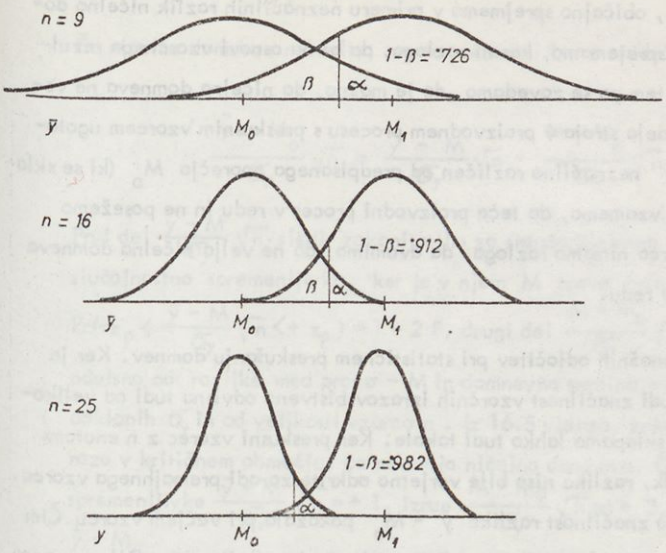
vrste α. Verjetnost $1-\beta$, da sprejmemo osnovno domnevo H_1, če ta velja, imenujemo moč preskusa. Čim večja je moč preskusa, tem večja je verjetnost, da sprejmemo na osnovi vzorca domnevo H_1.



Rezultat preskusnega vzorca	$\bar{y} \leq \bar{y}_z$	$\bar{y} > \bar{y}_z$
Sklep / Dejansko stanje	sprejmemo H_0	sprejmemo H_1
$M = M_0$	sklep pravilen $1-\alpha$	sklep napačen α
$M = M_1$	sklep napačen β	sklep pravilen $1-\beta$

Slika 16.2 Primerjalni pregled možnih situacij

Situacije v posameznih primerih najlaže proučimo iz shematičnega prikaza 16.2. V njem so narisane porazdelitve za poprečja vzorcev v primeru, če velja H_0 ali H_1, ustrezna kritična vrednost $\bar{y}_z$, območje sprejema H_0 in kritično območje, možni sklepi in verjetnosti za pravilne oziroma napačne sklepe. Kako sta pri danih domnevah H_0 in H_1 in predpisani verjetnosti za napako I. vrste α verjetnosti za napako II. vrste β oziroma moč preskusa $1-\beta$ odvisni od velikosti vzorca n, ponazorimo na našem zgledu pri različnih velikostih preskusnih vzorcev: $n = 9$, $n = 16$ in $n = 25$. Iz slike 16.3 je nazorno razvidno, kako z večanjem velikosti preskusnega vzorca pada verjetnost za napako II. vrste β in kako narašča moč preskusa.



n	1-ß	ß
9	·726	·274
16	·912	·098
25	·982	·018

Slika 16.3 Odvisnost verjetnosti za napako II. vrste in moči preskusa od velikosti vzorca.

16.6 Značilnost in neznačilnost vzorčnih izrazov.

V prikazanem zgledu o preskusu domnev o izvoru surovin smo izračunali za aritmetične sredine iz vzorcev $\bar{y}$ kritično mejo, ki razmeji območje za $\bar{y}$ v a): kritično območje in b) območje sprejema za ničelno domnevo. Če $\bar{y}$ pade v kritično območje, sklepamo, da z razmeroma majhnim tveganjem zavrnemo ničelno domnevo H_0 oziroma sprejmemo osnovno domnevo H_1. Ta sklep strnemo v ugotovitev, da je preskusni $\bar{y}$ **značilno različen** od vrednosti M_0 za ničelno domnevo H_0. Značilne razlike pomenijo, da je vzorec s tveganjem, ki ni večje kakor α, odkril, da velja osnovna domneva oziroma da je z majhnim tveganjem zavrnjena ničelna domneva.

Če pa pade $\bar{y}$ v območje sprejema ničelne domneve, pravimo, da je $\bar{y}$ od M_0 **neznačilno različen**. Če so razlike neznačilne, po gornjem načelu sprejmemo ničelno domnevo. Ker pa ne poznamo verjetnosti β za napako druge vrste, vemo

pa, da more biti velika, običajno sprejmemo v primeru neznačilnih razlik ničelno domnevo z razlago, da jo sprejemamo, ker ni razloga, da bi na osnovi vzorčnega rezultata dvomili o njej. Pri tem pa se zavedamo, da je možno, da ničelna domneva ne velja. Če pri preverjanju dela stroja v proizvodnem procesu s preskusnim vzorcem ugotovimo, da je preskusni $\bar{y}$ neznačilno različen od predpisanega poprečja M_o (ki se sklada z ničelno domnevo), vzamemo, da teče proizvodni proces v redu in ne posežemo vanj, ker na osnovi vzorca nimamo razloga, da dvomimo, da ne velja ničelna domneva in da proces ne poteka v redu.

Ta sklep je eden izmed možnih odločitev pri statističnem preskušanju domnev. Ker je kritična meja in s tem tudi značilnost vzorčnih izrazov bistveno odvisna tudi od velikosti preskusnega vzorca, sklepamo lahko tudi takole. Ker preskusni vzorec z n enotami ni odkril značilnih razlik, razlike niso bile verjetno odkrite zaradi premajhnega vzorca. Če $M - M_o \neq 0$, se bo značilnost razlike $\bar{y} - M_o$ pokazala pri večjem vzorcu. Čim manjša je razlika $M - M_o$, tem večji vzorec je potreben, da odkrijemo značilnost razlik.

16.7 Stopnja značilnosti. Ker je kritično območje odvisno od tveganja α, se more zgoditi, da je $\bar{y}$ v kritičnem območju, če je $\alpha = 0.05$, če pa je $\alpha = 0.01$ pa v območju sprejema ničelne domneve. V tem primeru vzamemo, da je $\bar{y}$ značilno različen s tveganjem $\alpha = 0.05$. Podobno govorimo o značilnosti razlik na stopnji $\alpha = 0.01$, če so razlike značilne na stopnji $\alpha = 0.01$, ne pa na stopnji $\alpha = 0.001$ in so značilne razlike na stopnji $\alpha = 0.001$, če je $\alpha < 0.001$.

16.8 Ničelno domnevo $H_o: M = M_o$ o poprečjih enostavneje preskušamo prek vzorčnega izraza

$$\frac{\bar{y} - M_o}{\sigma_y}\sqrt{n} = z \qquad (16.4)$$

ker se ta izraz v primeru, če velja ničelna domneva $H_o: M = M_o$, porazdeljuje standardizirano normalno. Poenostavitev je v tem, da so kritične meje neposredno enake kritičnim vrednostim za standardizirano normalno porazdelitev. Tako razlike štejemo za neznačilne, če je izračunani $|z| < 1,96$, značilne na stopnji $\alpha = 0.05$, če $1,96 \leq |z| < 2.58$, značilne na stopnji $\alpha = 0.01$, če je $2,58 \leq |z| < 3,29$ in značilne

na stopnji $\alpha = 0.001$, če je $3{,}29 \leqslant |z|$

Če ničelna domneva ne velja, moremo vzorčni izraz razdeliti v dva dela.

$$\frac{\bar{y} - M_o}{\sigma_y}\sqrt{n} = \frac{\bar{y} - M}{\sigma_y}\sqrt{n} + \frac{M - M_o}{\sigma_y}\sqrt{n} \qquad (16.5)$$

Prvi del $\frac{\bar{y} - M}{\sigma_y}\sqrt{n}$ sledi zakonitostim za standardizirano normalno porazdeljeno slučajnostno spremenljivko, ker je v njem M prava aritmetična sredina $Pr(-z_p < \frac{\bar{y} - M}{\sigma_y}\sqrt{n} < +z_p) = 1 - 2P$, drugi del $\frac{M - M_o}{\sigma_y}\sqrt{n}$ pa je konstanta, ki je odvisna od: razlike med pravo - M in domnevno sredino - M_o, merjeno v standardnih odklonih σ_y in od velikosti vzorca n . Iz 16.5 vidimo, zakaj je vrednost vzorčnega izraza v kritičnem območju, če ne velja ničelna domneva. Če je vrednost slučajnostne spremenljivke $\frac{\bar{y} - M}{\sigma_y}\sqrt{n} = +1$, izraz $\frac{M - M_o}{\sigma_y}\sqrt{n}$ pa $+3$, je vrednost vzorčnega izraza $\frac{\bar{y} - M_o}{\sigma}\sqrt{n} = +4$, kar pokaže pravilno sliko značilnih razlik.

16.9 Verjetnosti α in $1-\alpha$ nastopajo v primeru, če velja ničelna domneva $M=M_o$, vrejetnosti β oziroma $1-\beta$ pa v primeru, če velja osnovna domneva $M=M_1$. Te verjetnosti so torej pogojne verjetnosti.

$$\alpha = Pr\,(zav\ H_o/H_o) \qquad 1-\alpha = Pr\,(spr\ H_o/H_o)$$

$$\beta = Pr\,(zav\ H_1/H_1) \qquad 1-\beta = Pr\,(spr\ H_1/H_1)$$

Te verjetnosti so pogojno vezane na dejansko stanje, ki je v skladu bodisi s H_o ali s H_1. Ker pa dejanskega stanja ne poznamo, so seveda te verjetnosti samo posredno pomembne. Iz njih npr. ne moremo ugotoviti, kakšna je verjetnost, da smo napravili pravilen sklep (ta nikakor ni enaka $[(1-\alpha) + (1-\beta)]$), niti kakšna je verjetnost, če velja H_o, ali kakšna je verjetnost, če velja H_1, če je rezultat iz preskusnega vzorca v območju sprejema ničelne domneve oziroma v kritičnem območju. Te verjetnosti moremo izračunati le, če poznamo verjetnosti, s katerimi je dejansko stanje v skladu s H_o oziroma s H_1. Te verjetnosti moremo zapisati kot $Pr\,(M_o)$ oziroma $Pr(M_1)$. Verjetnosti $Pr\,(M_b)$ oziroma $Pr(M_1)$, za kateri velja

$$Pr(M_o \cup M_1) = Pr(M_o) + Pr(M_1) = 1 \qquad (16.6)$$

ker imamo samo dve možnosti, običajno ne poznamo. Vendar jih moremo za potrebe sklepanja na osnovi izkušenj in poznavanja problema dostikrat oceniti.

Na osnovi izkušenj, ki jih imamo z določenim proizvajalcem, moremo oceniti kakšna je verjetnost, da je kakovost dobavljenega blaga v skladu s predpisom oziroma da ni v skladu s predpisom. Te verjetnosti morejo biti glede na sloves od proizvajalca do proizvajalca različne. Čeprav so ocenjene verjetnosti subjektivne narave in mogoče niso v skladu z dejanskim stanjem, so vendar boljša ali slabša orientacija za prave odnose. Boljše poznavanje in širša informiranost daje boljše ocene. Na vsak način pa je tudi nepopolna informacija boljša kot nič.

16.10 Za zgled vzemimo preskušanje domnev o surovinah. Da bi izračunali ustrezne verjetnosti, bi morali poznati verjetnost $Pr(S_o)$, da je pošiljka iz surovine S_o in verjetnost $Pr(S_1)$, da je iz surovine S_1. Te verjetnosti moremo izračunati iz podatkov o uporabljenih surovinah v proizvodnji, moremo jih pa tudi na osnovi poznavanja subjektivno oceniti. Če vemo, da opuščajo uporabo surovin S_o, ki daje slabše izdelke in ocenimo verjetnost, da je pošiljka iz surovine S_o $Pr(S_o) = 0,20$ verjetnost, da je pošiljka iz surovine S_1 pa štirikrat večja, $Pr(S_1) = 0.80$, dobimo iz sheme 16.2 takele rezultate:

Dejansko stanje	Preskusni vzorec		$Pr(M=M_1)$
	$\bar{y} \leq \bar{y}_z$ sprejmemo H_o	$\bar{y} > \bar{y}_z$ sprejmemo H_1	
$M = M_o = 20.0$	sklep: pravilen $Pr(H_o/M_o) = 1-\alpha = 0.95$	sklep: napačen $Pr(H_1/M_o) = \alpha = 0.05$	$Pr(M=M_o) = 0.20$
$M = M_1 = 21.5$	sklep: napačen $Pr(H_o/M_1) = \beta = 0.088$	sklep: pravilen $Pr(H_1/M_1) = 1-\beta = 0.912$	$Pr(M=M_1) = 0.80$

Po stavku za množenje verjetnosti dobimo:

$Pr(H_1 \cap M_0) = Pr(M_0) \cdot Pr(H_1/M_0) = 0.20 \cdot 0.05 = 0.01$

$Pr(H_0 \cap M_0) = Pr(M_0) \cdot Pr(H_0/M_0) = 0.20 \cdot 0.95 = 0.19$

$Pr(H_1 \cap M_1) = Pr(M_1) \cdot Pr(H_1/M_1) = 0.80 \cdot 0.912 = 0.73$

$Pr(H_0 \cap M_1) = Pr(M_1) \cdot Pr(H_0/M_1) = 0.80 \cdot 0.088 = 0.07$

Iz teh rezultatov moremo sestaviti tabelo

Dejansko stanje	Sprejmemo H_0	Sprejmemo H_1	$Pr(M_1)$
M_0	$Pr(H_0 \cap M_0) = \cdot 19$	$Pr(H_1 \cap M_0) = \cdot 01$	$Pr(M_0) = \cdot 20$
M_1	$Pr(H_0 \cap M_1) = \cdot 07$	$Pr(H_1 \cap M_1) = \cdot 73$	$Pr(M_1) = \cdot 80$
$Pr(H_j)$	$Pr(H_0) = \cdot 26$	$Pr(H_1) = \cdot 74$	1.00

Pri veljavi postavljenih verjetnosti $Pr(M_0)$ in $Pr(M_1)$ dobimo, da je

$$Pr(H_0) = Pr(H_0 \cap M_0) + Pr(H_0 \cap M_1) = \cdot 19 + \cdot 07 = \cdot 26$$

$$Pr(H_1) = Pr(H_1 \cap M_0) + Pr(H_1 \cap M_1) = \cdot 01 + \cdot 73 = \cdot 74$$

Verjetnost, da se odločimo, da sprejmemo H_1 ne glede na to, ali je sklep pravilen ali napačen, je $Pr(H_1) = \cdot 74$

Verjetnost, da napravimo pravilen sklep, pa je

$$Pr(H_0 \cap M_0) + Pr(H_1 \cap M_1) = \cdot 19 + \cdot 73 = \cdot 92$$

in obratno verjetnost, da je sklep napačen

$$Pr(H_1 \cap M_0) + Pr(H_0 \cap M_1) = \cdot 01 + \cdot 07 = \cdot 08$$

Če nas zanima, kakšna je pogojna verjetnost, da je sklep pravilen, če smo sprejeli domnevo H_1, je ta enaka

$$Pr(M_1/H_1) = \frac{Pr(H_1 \cap M_1)}{Pr(H_1)} = \frac{\cdot 7296}{\cdot 7396} = \cdot 9865$$

Analogno dobimo tudi druge ustrezne verjetnosti

Dejansko stanje	s p r e j m e m o	
	H_o	H_1
M_o	·7296	0·0135
M_1	·2704	·9865
	1·00	1·00

Verjetnosti $Pr(M_k/H_i)$ so izračunane po splošnem Bayesovem obrazcu

$$Pr(M_k/H_i) = \frac{Pr(M_k) \cdot Pr(H_i/M_k)}{\sum_i Pr(M_i)\, Pr(H_i/M_i)} \qquad (16.7)$$

za računanje t.i. inverzne verjetnosti.

16.11 Preskušanje sestavljenih domnev. Pri znanem σ postavimo osnovno domnevo, da poprečna trdnost ni $M_o = 20$. $H_1 : M \neq M_o = 20$ in $H_o: M = M_o = 20$. $H_o : M = M_o = 20$ je enostavna domneva, ker sestoji iz ene same vrednosti $M = M_o = 20$, $H_1 : M \neq M_o = 20$ pa je sestavljena domneva, ker vsebuje več vrednosti (vse vrednosti na razmiku $o < M_1 < \infty$ razen $M_1 = M_o = 20$).

Če pa sta domnevi določeni s

$$H_o : M \geqslant M_o = 20 \; ; \; H_1 : M < M_o = 20$$

sta ničelna in alternativna domneva sestavljeni domnevi, ker tako H_o kot H_1 sestojita iz več vrednosti.

Najprej določimo območja značilnosti razlik za poprečje preskusnega vzorca $\bar{y}$ od M_o za domnevi: $H_o : M = M_o = 20$; $H_1 : M \neq M_o = 20$

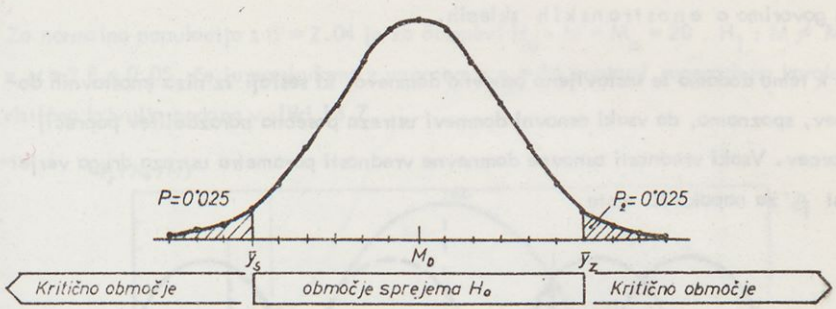
V tem primeru območje za sprejem ničelne domneve obsega razmik

$$\bar{y}_s = M_o - z_p \frac{\sigma}{\sqrt{n}} < \bar{y} < M_o + z_p \frac{\sigma}{\sqrt{n}} = \bar{y}_z$$

$$\bar{y}_s = 20 - 1{,}96 \frac{2}{\sqrt{16}} < \bar{y} < 20 + 1{,}96 \frac{2}{\sqrt{16}} = \bar{y}_z$$

$$\bar{y}_s = 19{,}02 < \bar{y} < 20{,}98 = \bar{y}_z$$

s tveganjem $\alpha = 2P = 0.05$



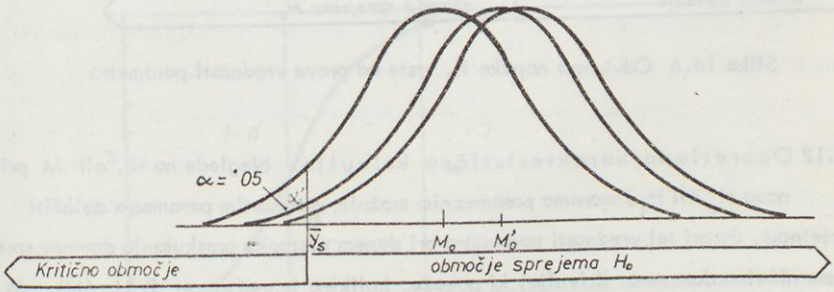
Slika 16.4 Dvostranski preskus domnev $H_0 : M = M_0$; $H_1 : M \neq M_0$

Ker v tem primeru kritično območje določeno z dvema razmikoma, govorimo o dvostranskem sklepu.

Če je tudi ničelna domneva sestavljena, kot je to npr. primer za $H_0 : M \geq M_0 = 20$
$H_1 : M < M_0 = 20$, je s kritično mejo

$$\bar{y}_s = M_0 - z_P \frac{\sigma}{\sqrt{n}} = 20 - 1{,}645 \frac{2}{\sqrt{16}} = 19{,}178$$

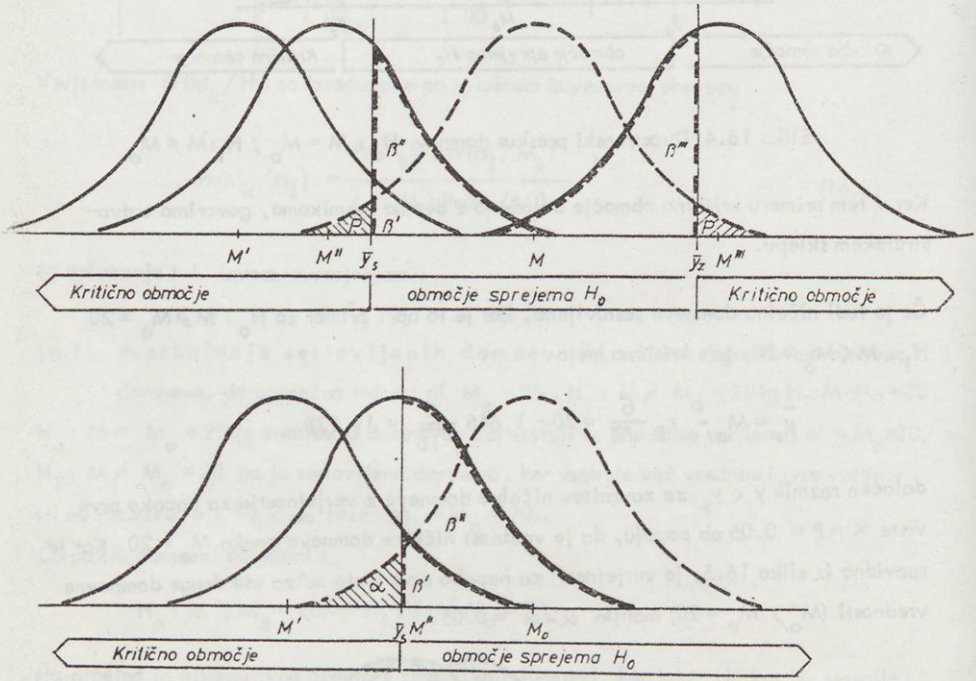
določen razmik $\bar{y} < \bar{y}_s$ za zavrnitev ničelne domneve z verjetnostjo za napako prve vrste $\alpha = P = 0.05$ ob pogoju, da je vrednost ničelne domneve enaka $M_0 = 20$. Kot je razvidno iz slike 16.5, je verjetnost za napako prve vrste α' za vse druge domnevne vrednosti ($M_0' > M_0 = 20$) manjša $\alpha' < \alpha = 0.05$.



Slika 16.5 Odvisnost napake I. vrste od vrednosti ničelne domneve

V gornjem zgledu je kritično območje določeno z eno samo vrednostjo. V takih primerih govorimo o enostranskih sklepih.

Če k temu dodamo še sestavljeno osnovno domnevo, ki sestoji iz niza enostavnih domnev, spoznamo, da vsaki osnovni domnevi ustreza posebna porazdelitev poprečij vzorcev. Vsaki vrednosti osnovne domnevne vrednosti parametra ustreza druga verjetnost β' za napako II. vrste.

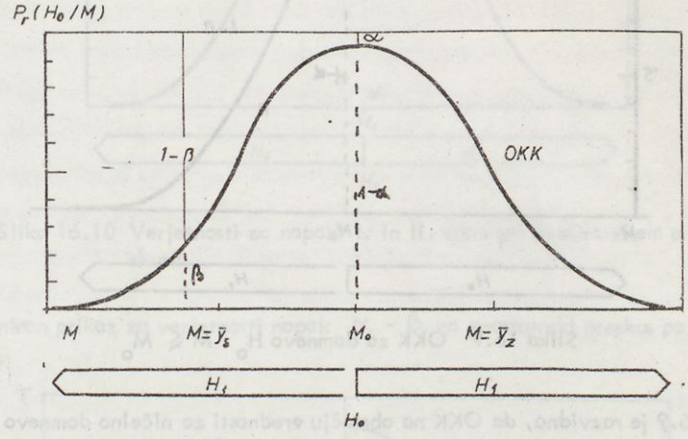


Slika 16.6 Odvisnost napake II. vrste od prave vrednosti parametra

16.12 Operativna karakteristična krivulja. Neglede na to, ali M pripada H_0 ali H_1, moremo posameznim možnim vrednostim parametra določiti verjetnost, da pri tej vrednosti parametra pri danem postopku preskušanja domnev sprejmemo ničelno domnevo. Krivulja, ki pokaže, kolikšna je verjetnost $Pr(H_0/M)$, da sprejmemo ničelno domnevo v odvisnosti od vrednosti parametra M, imenujemo opera-

tivno karakteristično krivuljo (OKK).

Za normalno populacijo s $\sigma = 2.04$ je za domnevi $H_o : M = M_o = 20$, $H_1 : M \neq M_o$ z $\alpha = 2\,P = 0.05$, če ju preskušamo z vzorcem z $n = 16$ enotami, operativna karakteristična krivulja podana v sliki 16.7

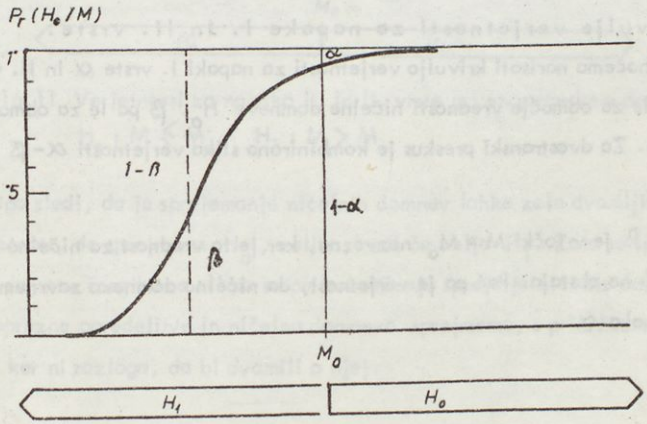


Slika 16.7 Operativna karakteristična krivulja OKK za dvostranski preskus domneve

Za enostranski preskus domnev

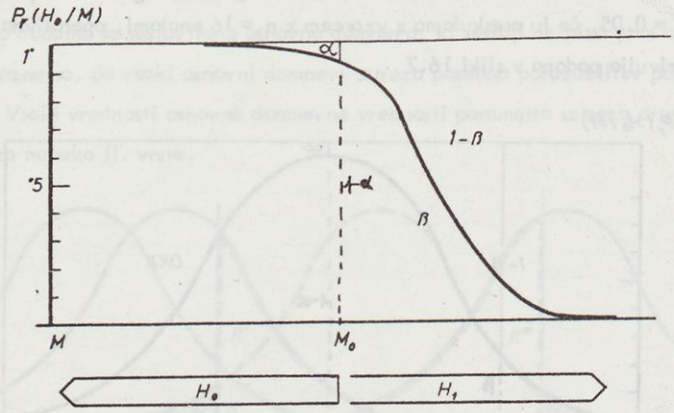
$$H_o : M \geq M_o = 20 \quad H_1 : M < M_o$$

je OKK dana v sliki 16.8.



Slika 16.8 OKK za domnevo $H_o : M \geq M_o$

Za enostranski preskus domnev: $H_o : M \leq M_o$; $H_1 : M > M_o$ pa je potek OKK pokazan v sliki 16.9.



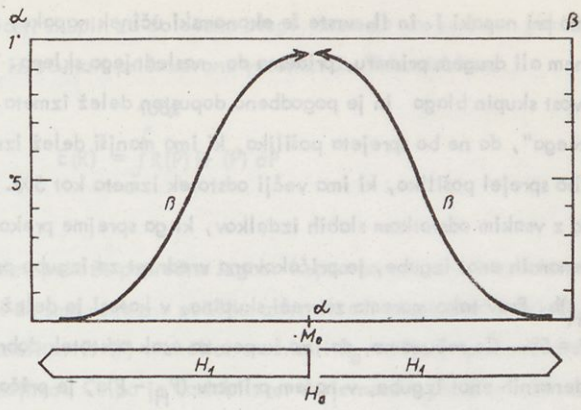
Slika 16.9 OKK za domnevo $H_o : M \leq M_o$

Iz slike 16.9 je razvidno, da OKK na območju vrednosti za ničelno domnevo H_o kaže verjetnosti za napake I. vrste α oziroma verjetnosti za sprejem ničelne domneve $1 - \alpha$. Na območju vrednosti domneve H_1 pa kaže OKK verjetnosti za napako II. vrste β oziroma za moč preskusa $1 - \beta$. Kot vidimo, ima OKK kakovostno drugačen pomen na območju ničelnih H_o in alternativnih H_1 domnevnih vrednosti parametra.

16.13 Krivulje verjetnosti za napake I. in II. vrste.

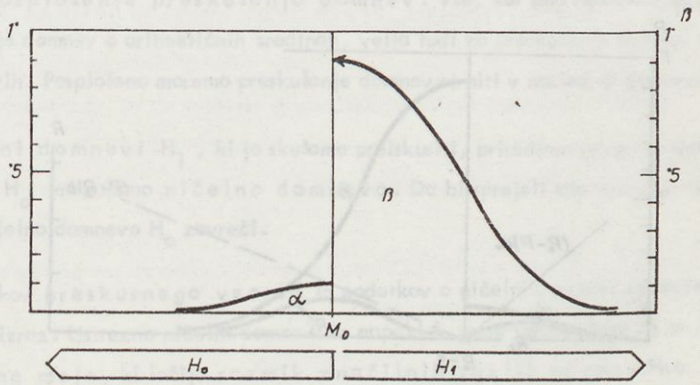
Če hočemo narisati krivuljo verjetnosti za napaki I. vrste α in II. vrste β, nastopa α le za območje vrednosti ničelne domneve H_o, β pa le za območje osnovne domneve. Za dvostranski preskus je kombinirana slika verjetnosti $\alpha - \beta$ dana v sliki 16.10.

Krivulja za β je v točki $M = M_o$ nezvezna, ker je to vrednost za ničelno domnevo, za katero β ne obstaja. Pač pa je verjetnost, da ničelno domnevo zavrnemo, če je $M = M_o$, enaka α.



Slika 16.10 Verjetnosti za napaki I. in II. vrste pri dvostranskem preskusu.

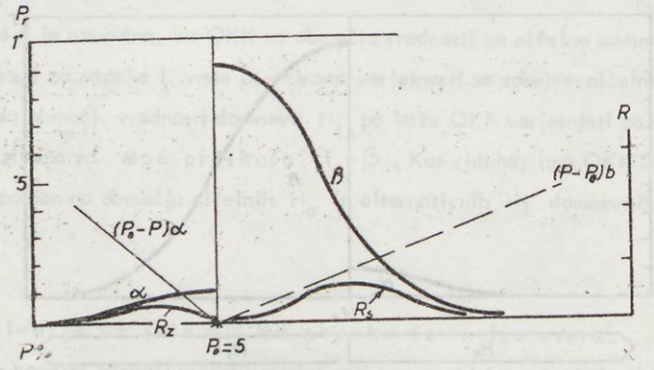
Kombiniran prikaz za verjetnosti napak $\alpha - \beta$ za enostranski preskus pa je



Slika 16.11 Verjetnosti za napako I. in II. vrste pri enostranskem preskusu:
$H_0 : M \leq M_0$; $H_1 : M > M_0$

Iz gornjega sledi, da je sprejemanje ničelnih domnev lahko zelo dvomljivo, ker more biti verjetnost, da sprejmemo H_0, velika, tudi če velja H_1. Zato se sprejemanja ničelnih domnev izogibamo ali pa se pri tolmačenju sprejetja ničelne domneve poslužujemo oprezne opredelitve in ničelno domnevo sprejmemo, s pridržkom, da jo sprejemamo, ker ni razloga, da bi dvomili o njej.

16.14 Če ocenimo pri napaki I. in II. vrste še ekonomski učinek napake, ki jo napravimo v enem ali drugem primeru, pridemo do naslednjega sklepa. Vzemimo, da proučujemo kakovost skupin blaga in je pogodbeno dopusten delež izmeta 5%. Tako kot prodajalec "tvega", da ne bo sprejeta pošiljka, ki ima manjši delež izmeta kot 5%, tvega kupec, da bo sprejel pošiljko, ki ima večji odstotek izmeta kot 5%. Če računamo, da ima kupec z vsakim odstotkom slabih izdelkov, ki ga sprejme preko pogodbeno določenega b denarnih enot izgube, je pričakovana vrednost za izgubo pri določenem P enako: $\beta(P - P_H)b$. Prav tako moremo zavreči skupino, v kateri je delež slabih izdelkov manjši kot $P = 5\%$. Če računamo, da ima kupec za vsak odstotek dobrih izdelkov, ki ga zavrne, a denarnih enot izgube, v našem primeru $(P_H - P)a$, je pričakovana vrednost te izgube kot funkcija deleža izmeta $\propto (P_H - P)a$. Če narišemo dobljeno pričakovano vrednost izgube namesto tveganja, dobimo realnejšo sliko o dejanskem tveganju kupca, ki je izraženo s poprečno izgubo pri sprejetju neustrezne ali zavrnitvi blaga z ustrezno kakovostjo.



Slika 16.12 Pričakovane izgube R v odvisnosti od prave vrednosti parametra

Iz izvajanja in iz slike 16.12 je razvidno, da je pričakovana izguba v okolici $P = 5\%$ razmeroma majhna, čeprav je tveganje v tem odseku največje. Vzrok za to je iskati v tem, da je v tem delu sicer verjetnost za napačen sklep razmeroma velika, vendar pa so posledice zaradi majhnega odklona od pogodbenega stanja nepomembne. V primeru, da poznamo verjetnostno porazdelitev pojavljanja skupin z določeno kakovostjo Pr(P) (to porazdelitev moremo oceniti subjektivno ali iz dolgoročnejšega sprem-

ljanja kakovosti skupin za določeno blago, moremo kot sumaren pokazovalec pričakovane izgube izračunati pričakovano vrednost R. Tako dobimo

$$E(R) = \int_{0}^{100\%} R(P)\, Pr\,(P)\, dP \qquad (16.8)$$

Ta pokazovalec pokaže poprečno izgubo kupca na eno skupino, ki smo jo napačno opredelili (bodisi da smo ustrezno pošiljko zavrnili ali neustrezno sprejeli). Če je verjetnostna porazdelitev Pr (P) v okolici vrednosti $P_o = 5\%$, je pričakovana vrednost E(R) razmeroma majhna. Če pa je variabilnost verjetnostne porazdelitve razmeroma velika, (skupine heterogene) ali če je pričakovana vrednost E(P) odmaknjena od $P_o = 5\%$, pa je pričakovana izguba razmeroma velika.

16.15 Posplošenje preskušanja domnev. Vse, kar smo nakazali za preskušanje domnev o aritmetičnih sredinah, velja tudi za preskušanje domnev o drugih parametrih. Posplošeno moremo preskušanje domnev strniti v naslednji postopek:

Osnovni domnevi H_1, ki jo skušamo preiskusiti, priredimo ustrezno protidomnevo H_o. H_o imenujemo ničelno domnevo. Da bi sprejeli osnovno domnevo H_1, skušamo ničelno domnevo H_o zavreči.

Iz podatkov preskusnega vzorca in podatkov o ničelni domnevi sestavimo ustrezen vzorčni izraz. Ustrezno ničelni domnevi in napaki I. vrste izračunamo za ta izraz kritične meje, ki ločijo razmik značilnih razlik od razmika neznačilnih razlik vzorčnega izraza od ničelne domneve. Razmik značilnih razlik imenujemo kritično območje.

Če pade vrednost izraza, ki ga izračunamo iz podatkov preskusnega vzorca, v kritično območje, ničelno domnevo zavrnemo in sprejmemo osnovno domnevo. Ta sklep napravimo s tveganjem α, ki je v skladu z napako I. vrste. Bistvo napake I. vrste je v tem, da ničelno domnevo zavrnemo, čeprav velja. Če pa pade vrednost izračunanega vzorčnega izraza v razmik, v katerem so razlike neznačilne, sklepamo

ustrezno problemu, ki ga raziskujemo: a) ker ni razloga za zavrnitev H_o, H_o sprejmemo, b) se vzdržimo sklepa z izjavo, da so razlike neznačilne, c) povečamo vzorec, da se pokažejo razlike, če obstajajo.

Če sprejmemo ničelno domnevo, ako izračunani izraz pade v razmik neznačilnih razlik, je lahko tveganje veliko. Napravimo lahko napako II. vrste, ki je v tem, da ničelno domnevo sprejmemo, čeprav ne velja. Ker je verjetnost β za napako II. vrste odvisna od parametrov v populaciji, teh pa ne poznamo, v splošnem ne vemo, kolikšna je napaka II. vrste. Vemo le to, da je tem večja, čim manj se dejanske razmere v populaciji razlikujejo od domnevnih in obratno tem manjša, čim večje so razlike med dejanskim in domnevnim stanjem. Če je β verjetnost, da ničelno domnevo sprejmemo, čeprav ne velja, je $(1 - \beta)$ verjetnost, da ničelno domnevo zavrnemo, če ta ne velja. To verjetnost imenujemo moč preskusa ali moč testa. Krivuljo, ki grafično pokaže, kolikšna je verjetnost, da sprejmemo ničelno domnevo, v odvisnosti od prave vrednosti parametra, imenujemo operativno karakteristično krivuljo - OKK.

PREIZKUŠANJE DOMNEV Z VELIKIMI VZORCI

16.16 Preizkušanje domnev z velikimi vzorci je precej enotno za vse parametre, ker vemo, da se za velike vzorce izraz

$$\frac{g - G}{SE_g} =: z \qquad (16.9)$$

porazdeljuje standardizirano normalno. Pri tem pomeni: G = prava vrednost parametra, g = ocena parametra s preskusnim vzorcem, SE_g = standardni pogrešek za oceno.

Če preizkušamo osnovno domnevo H_1, da je prava vrednost parametra G različna od domnevne vrednosti G_o ($H_1 : G \neq G_o$), postavimo ničelno domnevo H_o, da je prava vrednost enaka domnevni ($H_o : G = G_o$). Če velja ničelna domneva, moremo v obrazcu 16.9 nadomestiti pravo vrednost G s hipotetično G_o, ker sta enaki. Tako dobimo

$$z = \frac{g - G_o}{SE_g} \qquad (16.10)$$

Ta izraz se porazdeljuje standardizirano normalno le, če velja ničelna domneva $H_o : G = G_o$. V tem primeru je majhna verjetnost $\alpha = 2P$, da dobimo za konkreten preizkus z izven razmika $-z_p$ do $+z_p$. Če pade izračunani z v kritično območje, so razlike med g in G_o značilne. Zato ovržemo ničelno domnevo $H_o : G = G_o$ in sklepamo, da je dejanska vrednost parametra G različna od domnevne vrednosti G_o. Če pa pade izračunani z izven kritičnega območja, vzamemo, da so razlike neznačilne.

Podobno sklepamo tudi pri enostransko postavljenih domnevah, le da kritični razmik omejimo enostransko v levo ali v desno.

16.17 Preskušanje domnev o strukturnih deležih

Vzemimo, da iz velike populacije z vzorcem $n = 400$ izdelkov preizkušamo, ali je delež prvorazrednih izdelkov večji kot $P_o = 20\%$ ($H_1 : P > P_o = 20\%$). Ustrezna ničelna domneva je, da je odstotek prvovrstnih izdelkov enak ali manjši kot $P_o = 20\%$ ($H_o : P \leq P_o = 20\ \%$). S preizkusnim vzorcem $n = 400$ izdelkov dobimo $p = 25\%$ prvovrstnih izdelkov.

Po obrazcu 16.10 sestavimo izraz

$$z = \frac{p - P_o}{\sqrt{\frac{p(100-p)}{n}}} \qquad (16.11)$$

Iz njega dobimo

$$z = \frac{25 - 20}{\sqrt{\frac{25(100-25)}{400}}} = +2.31$$

Glede na zgornje pogoje sklepamo, da je odstotek prvorazrednih izdelkov iz preizkusnega vzorca značilno večji kot je $P_o = 20\%$ na stopnji $\alpha = 0.05$, tako dobimo, da je po obrazcu 16.11 izračunani z večji kot $z_{0.05} = +1,64$, na stopnji $\alpha = 0,01$,

če je večji kot $z_{0.01} = +2,32$ in na stopnji $\alpha = 0,001$, če je večji kot $z_{0.001} = +3,09$.

Ker je izračunani $z = +2,31$, sklepamo, da je delež prvovrstnih izdelkov značilno večji na stopnji $\alpha = 0,05$, ker je $z = +2,31$ večji kot $z_{0.05} = 1.64$ in manjši kot $z_{0.01} = 2,32$. Ker je $z = 2,31$ skoro enak $z_{0.01} = 2.32$ smatramo, da je značilnost praktično na stopnji $\alpha = 0.01$

16.18 Preskus domnev o razlikah med parametri. Podobno kot značilnost razlik g od domnevne vrednosti G_o preskušamo tudi domneve o razlikah parametrov med dvema populacijama s splošnim vzorčnim izrazom

$$z = \frac{(g_2 - g_1) - (G_{o2} - G_{o1})}{\sqrt{\text{var } g_1 + \text{var } g_2}} \qquad (16.12)$$

Ta izraz se v primeru, če velja ničelna domneva $H_o : G_2 - G_1 = G_{02} - G_{01}$, da je razlika med parametroma $G_2 - G_1$ enaka domnevni razliki $G_{o2} - G_{o1}$, porazdeljuje standardizirano normalno.

16.19 Za zgled vzemimo proučitev, ali je poprečna poraba določenega blaga ned socioekonomskima skupinama 2 in 1 značilno različna. Gre za dvostransko preskušanje razlik med poprečjema z ničelno domnevo $H_o : M_2 - M_1 = 0$, da ni razlike v porabi: Z anketo $n_1 = 500$ gospodinjstev iz socioekonomske skupine 1 smo dobili, da je $\bar{y}_1 = 400$; $s_1^2 = 10000$, z anketo $n_2 = 500$ gospodinjstev za drugo socioekonomsko skupino pa $\bar{y}_2 = 450$ in $s_2^2 = 16000$. Če vstavimo podatke vzorca in ničelne domneve v splošen izraz 16.12, prirejen za ta zgled, dobimo:

$$z = \frac{\bar{y}_2 - \bar{y}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (16.13)$$

$$z = \frac{450 - 400}{\sqrt{\frac{10000}{500} + \frac{16000}{500}}} = +6.93$$

Ker je izračunani $|z| = 6,93$ večji kot $z_{\alpha=2P=0.001} = 3,29$ sklepamo, da so razlike v poprečni porabi zelo značilne.

16.20 Da bi proučili, ali je število obratnih nesreč v popoldanski izmeni večje kot v dopoldanski, so v podjetju A evidentirali obratne nesreče po izmenah. Rezultat polletnega evidentiranja je: $h_{dop} = 120$; $h_{pop} = 150$. Vprašanju ustreza ničelna domneva $H_o : G_{dop} \geqslant G_{pop}$ (G = gostota nesreč).

Glede na to, da je število možnih nesreč veliko, sklepamo, da se število nesreč v polletju porazdeljuje v Poissonovi porazdelitvi, ker pa je h razmeroma velik, moremo vzeti približek, da se h porazdeljujeta v normalnih porazdelitvah. Glede na to je ustrezen vzorčni izraz po obrazcu 16.12 prirejen za ta problem

$$z = \frac{h_2 - h_1}{\sqrt{h_2 + h_1}} \qquad (16.14)$$

$$z = \frac{150 - 120}{\sqrt{150 + 120}} = +1,826$$

Ker je preskus enostranski (iščemo, ali je v drugi izmeni več nesreč) je $z = 1,826$ med kritičnima vrednostima $z_{\alpha=P=0.05} = +1,645 < z = +1,826 < z_{\alpha=P=0.01} = +2.58$

Sklep je, da je število obratnih nesreč v popoldanski izmeni značilno večje na stopnji tveganja $\alpha = 0.05$.

Preskušanje domnev z malimi vzorci

16.21 Posebnost preskušanja domnev z malimi vzorci je v tem, da se izrazi, ustrezni posameznim ničelnim domnevam, praviloma ne porazdeljujejo v standardizirano normalnih porazdelitvah, ampak tudi v različnih drugih porazdelitvah. Zato velja vse, kar smo nakazali za preskušanje domnev z velikimi vzorci, tudi za male vzorce, s to razliko, da izračunane vzorčne izraze primerjamo s kritičnimi vrednostmi za ustrezne

porazdelitve, v katerih se vzorčni izraz porazdeljuje, če velja ničelna domneva. Te so predvsem: z , $t_\alpha(m)$, $\chi^2_\alpha(m)$ in $F_\alpha(m_1, m_2)$. Spomnimo se, da so kritične vrednosti za te porazdelitve odvisne razen od α tudi od stopinj prostosti m.

16.22 Preskušanje domnev o aritmetičnih sredinah. Če velja $y =: N(M; \sigma^2)$, dobimo ničelni domnevi $H_o : M = M_o$ ustrezen vzorčni izraz

$$z = \frac{\bar{y} - M_o}{\sigma}\sqrt{n} \qquad (16.15)$$

Da bi mogli uporabiti ta obrazec, moramo poznati σ. Če ga ne poznamo in v izrazu 16.15 nadomestimo pravo vrednost σ z oceno iz preskusnega vzorca s, se za male vzorce izraz 16.15 ne porazdeljuje standardizirano normalno, temveč v t porazdelitvi. Tako dobimo vzorčni izraz

$$t = \frac{\bar{y} - M_o}{s}\cdot\sqrt{n} \qquad (16.16)$$

ki ga preskušamo s $t_\alpha(m = n-1)$.

Pogodbeno je določeno, da poprečna teža sadežev ne sme biti manjša kot $M_o = 20$ dkg. Temu problemu ustrezni domnevi sta $H_1 : M \geqslant M_o = 20$ dkg; $H_o : M < M_o = 20$ dkg. Da bi preskusili, ali pošiljka ustreza pogoju, so na slučajnosten način izbrali n = 16 sadežev in s tehtanjem dobili: $\bar{y} = 21$ dkg, $s = 2$ dkg. Če predpostavljamo, da se teža sadežev porazdeljuje normalno, uporabimo t preskus po obrazcu 16.16

$$t = \frac{21 - 20}{2}\sqrt{16} = +2$$

Ker je preskus enostranski, je ničelni domnevi ustrezna kritična vrednost $t_{\alpha=P=0.05}(m = n-1=16-1=15) = 1.753$; $t_{\alpha=0.01}(m=15) = +2,603$.

Ker je izračunani t med obema kritičnima vrednostima, sklepamo, da je poprečna teža sadežev v vzorcu značilno večja od $M_o = 20$ dkg na stopnji $\alpha = 0.05$ in pošiljko sprejmemo.

16.23 Preskus o poprečni razliki. Po obrazcu 16.16 za določene probleme posredno rešujemo probleme o razlikah med poprečji. Vzemimo, da proučujemo, ali je produktivnost dela po postopku B večja kot po postopku A. Preskus načrtujmo tako, da slučajnostno izberemo n delavcev in predpišemo, da vsak delavec dela po obeh načinih. Za vsakega delavca dobimo dva podatka: y_{1i} in y_{2i} za produktivnost po prvem in za produktivnost po drugem načinu. Če vzamemo, da je posamezna produktivnost dela

$$y_{1i} = M_1 + A_i + e_{1i}$$

$$y_{2i} = M_2 + A_i + e_{2i}$$

vsota učinka postopka (M_1 ali M_2), osebne sposobnosti A_i in slučajnostnih vplivov e_{1i} ali e_{2i}, dobimo, da je razlika

$$d_i = y_{2i} - y_{1i} = (M_2 - M_1) + e_{2i} - e_{1i} = D + e_i$$

le izraz razlike v produktivnosti v odvisnosti od postopka $(M_2 - M_1)$ in slučajnostne komponente $(e_i = e_{2i} - e_{1i})$. Odstranjen je osebnostni vpliv A oziroma heterogenost, ki izvira iz razlik v sposobnosti med delavci. Ker imamo za vsakega delavca le en podatek, ki je odvisen od razlik v postopku in slučajnostne komponente, za proučitev ničelne domneve $H_o : D = D_o$ priredimo obrazec 16.16 in dobimo

$$t = \frac{\bar{d} - D_o}{s_d}\sqrt{n} \quad ; \quad t_\alpha (m = n - 1) \qquad (16.17)$$

Pri tem je: $D_o = M_{o2} - M_{o1}$: $\bar{d} = \frac{1}{n} Sdi$; $s_d^2 = \frac{S(d_i - \bar{d})^2}{n - 1}$

Po obrazcu

$$t = \frac{\bar{d}}{s_d}\sqrt{n} \qquad t_\alpha (m = n - 1)$$

preskušamo ničelno domnevo $H_o : D_o = 0$, da razlik med aritmetičnima sredinama ni.

16.24 Če vzamemo shematičen zgled, dobimo

Delavec	Način dela 1	2		
i	y_{1i}	y_{2i}	d_i	d_i^2
1	12	11	- 1	1
2	7	10	+ 3	9
3	9	14	+ 5	25
4	12	15	+ 3	9
5	15	12	- 3	9
6	9	12	+ 3	9
7	8	7	- 1	1
8	13	13	0	0
9	12	11	- 1	1
10	7	9	+ 2	4
			$Sd_i = +10$	$68 = Sd_i^2$

$$\bar{d} = \frac{10}{10} = +1 \quad : \quad s_d^2 = \frac{68 - \frac{10^2}{10}}{10-1} = 6,4444 ; \quad s_d = 2,54$$

Po obrazcu 16.17 je

$$t = \frac{\bar{d}}{s}\sqrt{n} = \frac{+1}{2,54}\sqrt{10} = +1,24$$

Ker je izračunani $t = 1,24$ manjši kot $t_{\alpha=P=0.05}(m = n-1=8) = 1,833$, sklepamo, da produktivnost po drugem postopku ni značilno večja.

16.25 Nakazani postopek s pridom uporabljamo vselej, kadar lahko izvedemo oba postopka na eni in isti enoti oziroma na sorodnih parih.

Če proučujemo učinek določenega proizvodnega dejavnika na kakovost izdelkov, za vsak poskus vzamemo določeno količino surovine, jo razpolovimo, na eni polovici uporabimo način A, na drugi pa način B. Tako izločimo moteč vpliv heterogenosti surovin.

Če proučujemo vpliv načina prodaje na blagovni promet, izberemo določeno število trgovin in v vsaki trgovini prodajamo po možnosti istočasno na oba načina. Če izračunamo za vsako trgovino indekse, ki jih dobimo, če primerjamo $y_{2i}/y_{1i} = K_i$ prodajo po obeh načinih, dobimo podatke iz katerih so izločene razlike, ki izvirajo iz razlik v velikosti zaledja, geografskih pogojev ipd. in pridejo do izraza predvsem razlike v načinu prodaje.

Če proučujemo časovne spremembe v strukturi porabe za štiričlanska gospodinjstva, moremo ob času 2 opazovati ista gospodinjstva kot ob času 1 in tako izločimo variabilnost zaradi sestave gospodinjstev.

Zgledov, pri katerih moremo uporabiti nakazani način, da na isti enoti ali sorodnem paru izvedemo dvoje opazovanj ob spremenjenih pogojih, je mnogo. Ti ti. "odvisni vzorci", odvisni zato, ker je drugo opazovanje izvedeno na istih ali sorodnih enotah, imajo zaradi zmanjšanja slučajnostne komponente veliko uporabnost in prednost pred drugimi metodami. Vendar je v vseh primerih proučevanja razlik med poprečji ne moremo uporabiti.

16.26 Preskus domneve o razliki med aritmetičnima sredinama za neodvisna vzorca. Pri proučevanju odvisnosti produktivnosti dela od spola je uporaba nakazane metode že težja, čeprav bi v pare mogli združevati delavca in delavko, ki sta približno enake starosti, ki imata enak staž ipd. Enak problem je v primeru, če preskušamo značilnost razlik v porabi v odvisnosti od socioekonomske skupine.

Za take primere preskušamo razlike med aritmetičnimi sredinami prek vzorčnega izraza

$$t = \frac{\bar{y}_2 - \bar{y}_1 - \Delta M_o}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \; ; \; t_p(m=n_1+n_2-2) \; ; \; s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} \qquad (16.18)$$

V primeru, da razen predpostavk, da sta obe populaciji normalni s $\sigma_1^2 = \sigma_2^2$ velja, da je hipotetična razlika med aritmetičnima sredinama enaka dejanski razliki ($\Delta M_o = \Delta M$) se t porazdeljuje v t-porazdelitvi z $m = n_1 + n_2 - 2$ stopinjami pro-

stosti. Vzorčni izraz uporabljamo za preskušanje ničelne domneve $H_o : \Delta M_o = \Delta M$. Za preskušanje ničelne domneve $H_o : \Delta M_o = 0$ pa velja

$$t = \frac{\bar{y}_2 - \bar{y}_1}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \qquad (16.19)$$

16.27 Vzemimo za zgled proučevanje izdatkov za obleko in obutev na enega člana v letu 1957 v kmetijskih gospodarstvih s površino 3-8 ha v Sloveniji. Domneva, ki jo moramo preizkusiti, je, da je poprečje v izdatkih za obleko in obutev na enega člana v kmetijskih gospodarstvih v primorski regiji (Gorica, Koper) različna od poprečnih izdatkov za obleko in obutev na enega člana v kmetijskih gospodarstvih v štajerski regiji (Maribor, Ptuj, Murska Sobota). Domneva je dvostranska, ker iščemo samo značilnost razlik, ne predpostavljamo pa, v katerem predelu je poprečje večje. Iz podatkov ankete o življenju kmetov ZS SRS dobimo za slučajnostno izbrana gospodarstva tele osnovne podatke (v tisoč dinarjev na člana):

Primorska: $n_1 = 7$	y_{1i}	8,3	5,2	14,2	8,8	6,2	10,4	11,4
Štajerska: $n_2 = 14$	y_{2i}	3,6	4,5	2,7	2,3	1,0	3,2	3,1
		3,6	9,1	1,3	11,0	2,1	8,6	4,4

Iz teh podatkov dobimo dalje:

$$y_1 = 64,5 \quad Sy_{1i}^2 = 651,57; \quad \bar{y}_1 = 64,5/7 = 9,21; \quad s_1^2 = \frac{651,57 - \frac{64,5^2}{7}}{7-1} = 9,54$$

$$y_2 = 60,5 \quad Sy_{2i}^2 = 382,83 \quad \bar{y}_2 = 60,5/14 = 4,32; \quad s_2^2 = \frac{382,83 - 60,5^2/14}{14-1} = 9,34$$

Po obrazcu iz 16.18 je

$$s^2 = \frac{(7-1)\,9,54 + (14-1)\,9,34}{7+14-2} = 9,40; \quad s = \sqrt{9,40} = 3,07$$

Če te podatke vstavimo v obrazec 16.19, dobimo

$$t = \frac{4,32 - 9,21}{3,07} \sqrt{\frac{7 \cdot 14}{7+14}} = -3,44$$

Ker je število stopinj prostosti $m = n_1 + n_2 - 2 = 7 + 14 - 2 = 19$, sta za tveganje $\alpha = 2P = 0{,}01$ kritični meji za $t = -2{,}86$ in $+2{,}86$, za tveganje $\alpha = 2P = 0.001$ pa $-3{,}88$ in $+3{,}88$

Iz tega sklepamo, da so razlike med obema predeloma značilne na stopnji $\alpha = 0.01$, kar je razmeroma visoka značilnost.

16.27 Preizkušanje domnev o varianci. Z izrazom

$$\chi^2 = \frac{(n-1)s^2}{\sigma_o^2} \quad ; \quad \chi^2_p (m = n-1) \qquad (16.20)$$

preizkušamo domneve o varianci, če predpostavljamo, da se y porazdeljuje normalno.

16.28 Podjetje, ki je dobavilo drugemu podjetju stroj za proizvodnjo vijakov, je navedlo kot značilnost kakovosti dela stroja standardni odklon za dolžino vijakov $\sigma = 1$ mm. Podjetje, ki je stroj kupilo, skuša preizkusiti to značilnost in postavi domnevo, da je dejanski standardni odklon večji kot 1 mm ($H_1 : \sigma > \sigma_o = 1$). Tej domnevi ustreza ničelna domneva, da je dejanski standardni odklon enak 1 mm ($H_o : \sigma = \sigma_o = 1$). To domnevo preizkušamo z majhnim vzorcem $n = 30$ vijakov, ki jih proizvedemo v preizkusni proizvodnji. Iz izdelkov v poskusni proizvodnji ocenjeni standardni odklon je $s = 1{,}1$ mm. Preizkusiti je treba značilnost razlik med dejansko in domnevno variabilnostjo. Ker potrebujemo v obrazcu variance, ne pa standardne odklone, izračunamo: $\sigma_o^2 = 1^2 = 1$; $s^2 = 1{,}1^2 = 1{,}21$.

Če te podatke vstavimo v obrazec 16.20, dobimo, da je

$$\chi^2 = \frac{(30-1) \cdot 1{,}21}{1} = 35{,}09$$

Ustrezna kritična vrednost je $\chi^2_{p=0.05}(m = 30 - 1 = 29) = 42{,}56$

Ker je izračunana vrednost za χ^2 manjša kot kritična vrednost za tveganje $\alpha = 0{,}05$, sklepamo, da razlike niso značilne. To pomeni, da preizkusni vzorec ni odkril, da stvarna variabilnost ni v skladu z naznačeno kakovostjo.

16.29 Preskušanje domnev o razmerju med variancama. Značilnost razlik med variancama za dva neodvisna vzorca preskušamo z vzorčnim izrazom

$$F = \frac{s_1^2 / s_2^2}{\left[\sigma_1^2 / \sigma_2^2\right]_o} \quad ; \quad F_p(m_1 = n_1 - 1 \; ; \; m_2 = n_2 - 1) \qquad (16.21)$$

ki se porazdeljuje v F-porazdelitvi z $m_1 = n_1 - 1$ in $m_2 = n_2 - 1$ stopinjami prostosti, če je domnevno razmerje enako dejanskemu razmerju in se obe populaciji porazdeljujeta normalno.

16.30 Vzemimo za zgled naslednji problem. Podjetje namerava dati izrabljen stroj v popravilo. Pričakujejo, da bo varianca po popravilu najmanj za četrtino manjša kot pred popravilom. $H_1 (\sigma_2^2 \leq 0.75 \, \sigma_1^2)$. Da bi preskusili to domnevo, so vzorcem $n_1 = 25$ izdelkov, izdelanih pred popravilom, ocenili $s_1^2 = 200$, in iz vzorca $n_2 = 25$ izdelkov, izdelanih po popravilu pa $s_2^2 = 125$. Ustrezen vzorčni izraz je po 16.21,

$$F = \frac{200/125}{1/0.75} = 1.2$$

Ker je $F_{\alpha = P = 0.05}(m_1 = 25-1=24 \; ; \; m_2 = 25-1 = 24) = 1,98$ večji kot izračunani F, sklepamo, da je F neznačilen. Zato je neupravičen sklep, da je variabilnost po popravilu za četrtino manjša.

16.31 Pogosto preskušamo ničelno domnevo $H_o : \sigma_1^2 = \sigma_2^2$ da sta varianci za primerjani populaciji enaki. V tem primeru dobi vzorčni izraz 16.21 posebno obliko,

$$F = \frac{s_1^2}{s_2^2} \quad ; \quad F_p (m_1 = n_1 - 1 \; ; \; m_2 = n_2 - 1) \qquad (16.22)$$

16.32 Če preizkusimo značilnost razlik med variancama za izdatke za obleko in obutev na enega člana v kmetijskih gospodarstvih na Primorskem in Štajerskem, do-

bimo

$$F = s_1^2 / s_2^2 = 9,54/9,34 = 1,02$$

Če primerjamo izračunani $F = 1,02$ s kritično vrednostjo za F porazdelitev za $m_1 = n_1 - 1 = 7-1 = 6$ in $m_2 = 14-1 = 13$ stopinjami prostosti $F_{0.05}(m_1 = 6, m_2 = 13) = 2,9$, spoznamo, da razlike niso značilne, ker je izračunani F manjši kot kritična vrednost na stopnji $\alpha = 0,05$.

16.33 Preskus domnev o frekvenčnih porazdelitvah. Če je stvarna frekvenčna porazdelitev f_i rezultat vzorca iz populacije z verjetnostmi P_i, velja

$$\sum_{i=1}^{k} \frac{(f_i - nP_i)^2}{nP_i} =: \chi^2 \quad (m = k - 1) \tag{16.23}$$

Če v obrazcu 16.23 zamenjamo nP_i z ocenami f_i' za teoretične frekvence, ki jih dobimo ob določenih predpostavkah iz vzorca, se

$$\sum_{i=1}^{k} \frac{(f_i - f_i')^2}{f_i'} =: \chi^2 \quad (m = k - r) \tag{16.24}$$

porazdeljuje tudi v χ^2 porazdelitvi s stopinjami prostosti, ki so enake številu razredov k, zmanjšane za število zvez r med stvarnimi f_i in teoretičnimi f_i' frekvencami.

Ker gornja zveza velja pri ničelni domnevi o osnovni populaciji, jo moremo uporabiti za preskušanje te domneve po ustaljenem načinu za preskušanje domnev. Če je izračunani χ^2 manjši kakor kritična vrednost $\chi^2_{\alpha=0.05}$ sklepamo, da se frekvenčna porazdelitev neznačilno odklanja od teoretične, če pa je $\chi^2 > \chi^2_{\alpha=0.05}$, pa sklepamo, da je stvarna frekvenčna porazdelitev značilno različna od teoretične, kar pomeni, da osnovna porazdelitev ne ustreza pogojem pri ničelni domnevi.

16.34 Vzemimo za zgled porazdelitev tež svitkov preje v eni izmed tekstilnih tovarn v Sloveniji.

Preizkušamo osnovno domnevo, da je registracija tež svitkov pristranska, ker imajo delavci, ki sami registrirajo težo svitkov, vzrok, da teže prikrojijo.

Tej osnovni domnevi postavimo ustrezno ničelno domnevo, da registracija tež ni pristranska. V primeru, da so razlike v teži zgolj slučajne, je porazdelitev tež svitkov normalna.

Dejansko frekvenčno porazdelitev svitkov vzemimo za vzorec iz hipotetične populacije svitkov, ki bi bili tehtani ob enakih pogojih. Po znanih načelih izračunamo teoretično normalno porazdelitev z enakim obsegom, enako aritmetično sredino in standardnim odklonom kot stvarno porazdelitev. V tabeli 16.1 je prikazano izračunanje za ta zgled.

Tabela 16.1 Izračunanje χ^2 za frekvenčno porazdelitev n = 561 svitkov po registrirani teži.

Teža kg	f_i		f'_i		$f_i - f'_i$	$(f_i - f'_i)^2$	$(f_i - f'_i)^2 / f'_i$
18,55	3						
18,60	-		0,3				
18,65	14	28	2,6	17,2	+ 10,8	116,64	6,78
18,70	11		14,3				
18,75	23		49,6		- 26,6	707,56	14,27
18,80	71		108,3		- 37,3	1391,29	12,85
18,85	229		149,9		+ 79,1	6256,81	41,74
18,90	129		129,7		- 0,7	0,49	0,00
18,95	50		73,6		- 23,6	556,96	7,57
19,00	11		26,0		- 15,0	225,0	8,65
19,05	18		5,8				
19,10	2	20	0,8	6,7	+ 13,3	176,89	26,4
19,15	-		0,1				
	561		561,0		0		χ^2 = 118,26

Izraz 16.24 se porazdeljuje v χ^2-porazdelitvi le, če so teoretične frekvence zadosti velike, praktično, če so večje kot 5, zato smo v našem zgledu skrajne frekvence, ki so manjše kot 5, združili s frekvencami v sosednih razredih. Seveda smo morali združiti tudi frekvence v ustreznih razredih v stvarni frekvenčni porazdelitvi.

Vprašanje, ki ga moramo rešiti, je število stopinj prostosti. Po združitvi razredov z majhnimi frekvencami imamo $k = 8$ grup. Variiranje v frekvencah obeh porazdelitev pa je omejeno s tremi pogoji: z enakim skupnim številom enot n, z enako aritmetično sredino $\bar{y}$ in z enako varianco s^2. Število stopinj prostosti se v tem sistemu zniža za število omejitev, torej za 3. Skupno število stopinj prostosti je torej $m = k - r = 8 - 3 = 5$, ustrezna kritična vrednost $\chi^2_{P=0.001}(m=5) = 20,52$. Izračunana vrednost za χ^2 je veliko večja kot kritična vrednost z $\alpha = 0,001$. Iz tega sklepamo, da so razlike med stvarno porazdelitvijo svitkov f_i in hipotetično normalno porazdelitvijo f'_i zelo značilne in da je pravilna osnovna domneva, da je registracija teže pristranska. Če podrobneje proučimo razlike med vzorčno in teoretično porazdelitvijo, opazimo izredno velike razlike v razredu 18,85 kg, v katerem dobimo kopičenje v vzorcu ($f - f' = +79,1$). Ker je 18,85 kg predpisana teža za svitek, je to kopičenje psihološko utemeljeno. Enako opazimo kopičenje, čeprav ni tako intenzivno, tudi za teži 18,65 kg in 19,05 kg. Tudi to kopičenja je utemeljeno, ker sta ti teži meji za uporabnost svitkov.

ANALIZA VARIANCE

16.35 Analiza variance, za katero je osnovno tehnično orodje preizkušanje domnev o variancah, se je razvilo v samostojno in zelo pomembno orodje statistične analize. Vse statistično načrtovanje poskusov temelji na analizi variance v najrazličnejših oblikah.

V tem odstavku bomo nakazali osnovo analize variance in le najenostavnejši primer uporabe. Različne variante te metode pa bomo spoznali pri načrtovanju poskusov.

Vzemimo, da merimo proizvodnost dela (čas, ki ga porabijo za izdelavo enega izdel-

ka) za štiri delavce. Za vsakega delavca imamo podatke o porabljenem času za po pet izdelkov. Proučujmo variabilnost teh podatkov! Ti podatki so shematično nakazani v tabeli 16.2, v kateri so v vsaki vrsti nanizani podatki za posameznega delavca.

Tabela 16.2 Shema podatkov o produktivnosti dela za štiri delavce, ki so proizvedli po 5 izdelkov.

Delavec 1	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	Y_1	$\bar{y}_1$
Delavec 2	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	Y_2	$\bar{y}_2$
Delavec 3	y_{31}	y_{32}	y_{33}	y_{34}	y_{35}	Y_3	$\bar{y}_3$
Delavec 4	y_{41}	y_{42}	y_{43}	y_{44}	y_{45}	Y_4	$\bar{y}_4$
						Y	$\bar{y}$

Predpostavljajmo, da je poprečna proizvodnost dela vseh štirih delavcev enaka! Seveda zaradi slučajnih vplivov čas izdelave za posamezne izdelke ni enak, marveč variira.

Če ni razlik v proizvodnosti dela med delavci, moremo varianco v produktivnosti dela σ^2 oceniti na več načinov.

Eno oceno variance dobimo, če izračunamo po znanem obrazcu vsoto kvadratov odklonov od skupne aritmetične sredine $\bar{y}$, dobljeno vsoto pa delimo s številom stopinj prostosti $m = 4.5 - 1 = 19$. To oceno moremo zapisati

$$s^2 = \frac{\sum_k \sum_i (y_{ki} - \bar{y})^2}{4.5 - 1} = \frac{Q_{ki} - Q}{4.5 - 1} = \frac{K}{4.5-1} \qquad (16.25)$$

Varianco pa moremo oceniti tudi iz petih podatkov za posameznega delavca posebej: s_1^2; s_2^2; s_3^2; s_4^2;

$$s_K^2 = \frac{\sum_i (y_{ki} - \bar{y}_k)^2}{5 - 1}$$

Tako dobimo štiri neodvisne ocene variance, vsako z m=5-1 stopinjami prostosti. Iz teh štirih grupnih varianc moremo izračunati poprečje. Če bi bilo število podatkov za po-

samezne grupe različno, bi morali seveda izračunoti tehtano poprečje in uporabiti za pondere število stopinj prostosti. Ker pa je v našem zgledu število podatkov v grupah enako, dobimo isti rezultat, če izračunamo navadno poprečje.

Tako dobimo, da je

$$s_e^2 = \overline{s^2} = \frac{1}{4} \sum_k s_k^2 = \frac{\sum_k \sum_i (y_{ki} - \bar{y}_k)^2}{5 - 1} = \frac{Q_{ki} - Q_k}{4(5-1)} = \frac{K_e}{4(5-1)} \qquad (16.26)$$

Ta ocena je izračunana z $m = 4(5-1)$ stopinjami prostosti, ker smo jo izračunali iz štirih grup z $m_k = 5 - 1 = 4$ stopinjami prostosti.

Razen teh dveh ocen pa dobimo tretjo oceno iz grupnih aritmetičnih sredin. Če variirajo posamične vrednosti, variirajo tudi sredine, izračunane iz teh vrednosti. Vendar je variacija med sredinami manjša kot med posamičnimi vrednostmi. Iz zakonitosti o variiranju sredin vemo, da je varianca sredin $\sigma_{\bar{y}}^2 = \sigma_y^2/n$, ali v našem primeru $\sigma_{\bar{y}}^2 = \sigma^2/5$. Varianca med grupnimi poprečji

$$s_{\bar{y}}^2 = \frac{\sum_k (\bar{y}_k - \bar{y})^2}{4 - 1}$$

torej ni ocena za σ_y^2, marveč ocena za $\sigma_y^2/5$. Če hočemo, da bo to ocena za σ_y^2, moramo $s_{\bar{y}}^2$ pomnožiti s 5. Tretja ocena za varianco o produktivnosti dela σ_y^2 je torej

$$s_k^2 = 5 s_{\bar{y}}^2 = \frac{5 \sum_k (\bar{y}_k - \bar{y})^2}{4 - 1} = \frac{Q_k - Q}{4 - 1} = \frac{K_k}{4-1} \qquad (16.27)$$

Ta ocena je izračunana z $m = 4-1 = 3$ stopinjami prostosti, ker smo jo izračunali iz štirih grupnih sredin, ki so vezane na skupno sredino.

Če pozorneje pogledamo obrazce za posamezno vrsto ocen za σ_y^2, opazimo, da je med njimi določena zveza.

$$K = \sum_k \sum_i y_{ki}^2 - Y^2/20 = \left(\sum_k \sum_i y_{ki}^2 - \sum_k y_k^2/5\right) + \left(\sum_k y_k^2/5 - Y^2/20\right) =$$

$$= (Q_{ki} - Q_k) + (Q_k - Q) = K_e + K_k$$

Enaka zveza je tudi med stopinjami prostosti

$$m = 4.5-1 = 4.5-4+4-1 = 4(5-1) + 4-1 = m_e + m_k$$

Ker sta s_k^2 in s_e^2 oceni za isto varianco, čeprav je s_k^2 ocena iz variabilnosti v grupah, izraz

$$s_k^2/s_e^2 =: F\,(m_1 = k-1\,;\quad m_2 = n-k) \qquad (16.29)$$

ustreza zakonitosti za F porazdelitev z $m_1 = k-1 = 3$ in $m_2 = n-k = 16$ stopinjami prostosti. To pomeni, da ta izraz s tveganjem 0,05 ni večji kakor kritična vrednost $F_{0.05}(3,16) = 3{,}24$.

Gornji postopek je izdelan ob predpostavki, da poprečna produktivnost za posamezne delavce ni različna ($M_1 = M_2 = M_3 = M_4$). Vprašanje je, kaj se zgodi s posameznimi ocenami za varianco, če je produktivnost dela med delavci različna. Če pogledamo s_e^2, spoznamo, da različna raven produktivnosti dela za posamezne delavce nanjo ne vpliva, ker je s_e^2 odvisna samo od variabilnosti v produktivnosti dela za posameznega delavca. Pač pa ima različna produktivnost dela za posameznega delavca bistven vpliv na s_k^2. Ker je ta ocena izračunana iz ocen grupnih sredin, je s_k^2 tem večji, čim večje so stvarne razlike v produktivnosti dela med delavci. Ker je s_k^2 v izrazu s_k^2/s_e^2 v števcu, je izračunani F tem večji, čim večje so razlike v produktivnosti dela med delavci.

To sklepanje moremo izkoristiti za preizkušanje domneve o razlikah med aritmetičnimi sredinami za več delavcev. Osnovna domneva je, da je produktivnost dela med delavci različna ($H_1 : M_1 \neq M_2 \neq M_3 \neq M_4$). Ustrezna ničelna domneva je, da ni razlik v proizvodnosti dela med delavci ($H_o : M_1 = M_2 = M_3 = M_4$).
Kot vzorčni izraz za preskus te domneve rabi

$$F = s_k^2/s_e^2 \quad \text{s kritičnimi vrednostmi } F_{\alpha=p}(m_1=k-1\,;\ m_2 = n-k) \qquad (16.30)$$

Če je izračunani F večji kot ustrezna kritična vrednost F_p sklepamo, da so med aritmetičnimi sredinami v produktivnosti dela med delavci značilne razlike. Z napako prve vrste $\alpha = 0,05$ sklepamo, da je proizvodnost med delavci različna.

16.36 Vrednost postopka analize variance, ki smo ga nakazali za primer o produktivnosti dela in analize variance nasploh, je v tem, da moremo preizkusiti značilnost razlik med aritmetičnimi sredinami za več grupnih sredin hkrati, medtem ko moremo s preizkusom o značilnosti razlik med poprečjema za dve populaciji iz odstavka 16.26 primerjati le dvoje poprečij naenkrat. Pogoj za uporabo analize variance je predpostavka, da je variabilnost znotraj grup za vse grupe enaka in da je populacija normalno porazdeljena: $y_{ki} =: N(M_k ; \sigma^2)$

Z analizo variance moremo reševati najrazličnejše probleme. Preizkušati moremo razlike v porabi za več socialnih skupin hkrati, preverjati enakomernost dela za skupine strojev, proučevati razlike v pridelkih na parcelah, ki so bile obravnavane z različnimi agrotehničnimi postopki, preskušati razlike v prodaji v odvisnosti od različnih načinov oglaševanja za določen izdelek itd.

16.37 Postopek za analizo variance za en dejavnik v splošnejšem primeru, če je število enot v grupah različno, moremo nakazati v tehle točkah:

a) Iz vsake izmed k grup, katerih razlike proučujemo, izberemo vzorec z n_k enotami in zanje poiščemo vrednosti za znak y_{ki}, ki ga proučujemo. Indeks k pri y_{ki} pomeni grupo, i pa tekočo številko enote znotraj grupe.

b) Za vsako grupo izračunamo iz posamičnih vrednosti grupne vsote $Y_k = \sum_i y_{ki}$, iz teh pa skupno vsoto $Y = \sum_k Y_k$

c) Izračunamo vsoto kvadratov posamičnih vrednosti $\sum_k \sum_i y_{ki}^2 = Q_{ki}$, iz grupnih vsot y_k izraz $\sum_k Y_k^2/n_k = Q_k$, iz skupne vsote pa izraz $Y^2/n = Q$; $(\sum_k n_k = n)$

d) Iz izrazov c obračunamo analizo variance po standardni shemi

Tabela 16.3 Shema za obračun analize variance

Vir variacije	Vsota kvadratov odklonov SK	Stopinje prostosti m	$s^2 = K/m$	F
med grupami	$Q_k - Q = K_k$	$k - 1 = m_k$	$K_k/m_k = s^2_k$	$s^2_k/s^2_e = F$
znotraj grup	$Q_{ki} - Q_k = K_e$	$n - k = m_e$	$K_e/m_e = s^2_e$	1
Skupaj	$Q_{ki} - Q = K$	$n - 1$		

Razlike med grupnimi sredinami so značilne na stopnji α, če je izračunani F večji kot kritična vrednost F_α ($m_1 = k-1$; $m_2 = n - k$).

16.38 Za platišča za kamione, ki jih proizvaja ena izmed tovarn v Sloveniji, preizkušamo, ali je kakovost platišč odvisna od različnih žarilnih temperatur. V ta namen smo vzeli za vsako žarilno temperaturo po šest platišč in vsako izmed teh obdelovali pri različnih žarilnih temperaturah: 450°, 550° in 650°. Značilnost, ki je izraz trpežnosti platišč, je število pulzacij do uničenja. Osnovne podatke imamo v tabeli 16.4.

Tabela 16.4 Osnovni podatki preiskave, ali je trpežnost platišč odvisna od žarilne temperature (y = število pulzacij).

Žarilna temperatura	y_{ki}						Y_k
450°	335	394	314	282	211	253	1789
550°	906	634	401	560	390	513	3404
650°	253	336	264	311	133	247	1544

$6737 = Y$

$n_k = \bar{n} = 6, \quad k = 3, \quad n = 18$

$$Q_{ki} = \underset{k}{S}\,\underset{i}{S}\, y^2_{ki} = 335^2 + 394^2 + \ldots\ldots + 247^2 = 3088593$$

$$Q_k = \sum_k Y_k^2 / n_k = \frac{1789^2}{6} + \frac{3404^2}{6} + \frac{1544^2}{6} = 2861945$$

$$Q = Y^2/n = 6737^2/18 = 2521509$$

Tabela 16.5 Analiza variance za žarilno temperaturo

Vir variacije	SK	m	s^2	F
Žarilna temperatura	$K_k = 2861945-2521509 = 340436$	$m_k = 3-1=2$	$s_k^2 = 170218$	11,3
Slučajni vplivi	$K_e = 3068593-2861945 = 226648$	$m_e = 18-3=15$	$s_e^2 = 15110$	1
Skupno	$K = 567084$	17		

Razlike so zelo značilne, ker je izračunani F = 11,3 večji kakor $F_{0.01}(2.15) = 6,36$. To pomeni, da žarilna temperatura vpliva na kakovost platišč.

Neparametrične metode za preskušanje domnev

16.39 Pri preskušanju domnev o osnovnih parametrih s t-preskusom, χ^2-preskusi in F-preskusi predpostavljamo, da so proučevane populacije normalne, variance pa enake. Razen tega so ti preskusi razmeroma zamudni, ker so vezani na tehnično dokaj zapletena računanja. V primerih, v katerih niso izpolnjeni gornji pogoji, je uporaba teh preskusov, če že ne onemogočena, vsaj problematična. Dostikrat pa niti ne vemo, ali osnovne populacije ustrezajo pogojem. Zato v statistični praksi bolj in bolj uporabljamo ti. neparametrične preizkuse. Ti so zasnovani na ordinalnosti ali na nominalnosti znakov. Njihova uporaba ni vezana na določeno porazdelitev osnovnega znaka niti na zveze med parametri v osnovnih populacijah, odtod skupen naziv - neparametrični preskusi. Čeprav je moč neparametričnih preskusov manjša kot pri parametričnih, saj le delno izkoriščajo podatke iz vzorcev, so, zaradi svoje splošnejše uporabnosti in zaradi razmeroma enostavnega in hitrega postopka, ved-

no bolj čislani. To se pozna v vedno večjem številu neparametričnih preskusov in v vedno večji uporabi v najrazličnejših področjih. Manjšo moč preskusa odtehta enostavnejša tehnika in nevezanost na osnovne pogoje. Razen tega se z neparametričnimi preskusi razširi uporaba preskušanja domnev tudi na znake, ki imajo le ordinalen ali v določenih primerih tudi na znake, ki imajo le nominalen značaj.

16.40 Neparametrično preskušanje domnev o srednjih vrednostih.

Do sedaj obravnavani preskusi domnev o sredinah imajo za osnovo t- preskus. Te preskuse pa moremo uporabiti, če so proučevane populacije normalne, pri primerjavi aritmetičnih sredin dveh vzorcev pa morata imeti obe proučevani populaciji tudi enaki varianci. To pa so precejšnje omejitve za uporabo t- preskusa. Razen tega zahteva t- preskus razmeroma zapletena računanja, ki posebno pri operativnem delu zavirajo redne in hitre odločitve.

Preskus s predznaki

Primerjavo dveh smiselno povezanih znakov x in y za enote proučevane populacije moremo na splošno izraziti z linearnim izrazom

$$u = a + bx + cy \qquad (16.31)$$

Če zamenjamo obča števila a, b, c s posebnimi števili, dobimo nekaj vsebinsko pomembnih zvez:

$$\begin{array}{llll} a = -Me, & b = 1, & c = 0 & u = x - Me \\ a = 0, & b = -1, & c = 1 & u = y - x \\ a = 0, & b = -k & c = 1 & u = y - kx \\ a = -a & b = -1, & c = 1 & u = y - x - a \end{array} \qquad (16.32)$$

Če postavimo na splošno, da je Me(u) = 0, pomeni to v prvem primeru, da je Me(x-Me) = 0, v drugem, da je Me(y-x)=0, v tretjem Me(y-kx) = 0, v četrtem pa Me(y-x-a) = 0.

16.41 Preskus z mediano . Z ničelno domnevo H_o : $Me(u) = 0$ preskušamo dvostransko, z ničelno domnevo H_o : $Me(u) > 0$ ali H_o : $Me(u) < 0$ pa enostranske domneve o različnih problemih. Preskus je izredno enostaven. V primeru ničelne domneve velja zaradi opredelitve mediane, da je verjetnost, da izberemo enoto s pozitivnim u, enaka $P = 0,5$. V vzorcu iz hipotetične populacije se n_+ (oziroma n_-) porazdeljuje v binomski porazdelitvi z $E(n^+) = n/2$ in $Var(n^+) = n/4$. Za velike vzorce moremo vzeti kar preskusni izraz

$$z = \frac{n^+ - E(n^+)}{\sqrt{Var n^+}} = \frac{n^+ - n/2}{\sqrt{n/4}} = \frac{2n^+ - n}{\sqrt{n}} \qquad (16.33)$$

ali spodnja kritična meja za n^+

$n^+_s = \frac{1}{2}(n - z_p\sqrt{n})$ ali s korekturo za nezveznost

$n^+_s = \frac{1}{2}(n - 1 - z_p\sqrt{n+1})$

Za male vzorce (do $n = 100$) so v zbirki tabel dane kritične vrednosti za n^+ oziroma n^- za dogovorjene stopnje tveganja. n^+ pomeni število pozitivnih, n^- pa število negativnih vrednosti za u.

16.42 Z izrazom $u = x - Me_o$ preskušamo domneve o velikosti mediane. Pri prevzemu blaga preskuša kupec domnevo, da je mediana za težo jabolk pod $Me_o = 15$ dkg.

Ustrezna ničelna domneva je H_o : $Me < Me_o = 15$ dkg. S tehtanjem ugotovimo (moremo tehtati tako, da posamezna jabolka damo na tehtnico, na kateri je utež za 15 dkg in ugotavljamo samo ali je jabolko težje ali lažje od 15 dkg), da je $n^+ = 12$ težjih, $n^- = 38$ pa lažjih kot $Me_o = 15$ dkg. Za $n = 50$ so kritične vrednosti za enostranski preskus.

P	0,005	0,01	0,025	0,05	0,10
n^+ ali n^-	15	16	17	18	19

Z verjetnostjo, ki je večja kot $1 - 0,005 = 0,995$ sklepamo, da je mediana teže značilno manjša od domnevne. ($n^+ = 12 < 15$).

16.43 Da bi preskusili, ali da obdelava z novim postopkom boljše izdelke, je 40 delavcev izdelalo po en izdelek po standardnem S in en izdelek po novem N postopku. Kakovost izdelka je dana s hrapavostjo površine izdelkov.

Rezultat preskusa je: za $n^- = 10$ delavcev so ugotovili, da je kakovost izdelka, izdelanega po starem postopku, boljša in da so izdelki izdelani po standardnem postopku bolj gladki, za no = 5 delavcev se ni moglo ugotoviti, kateri izdelek je kvalitetnejši, za $n^+ = 25$ delavcev pa so ugotovili, da so izdelki, izdelani po novi metodi, kvalitetnejši. Čeprav se hrapavost ne da izraziti številčno, ima ta znak ordinalen značaj in za pare izdelkov ugotovimo, kateri je bolj kakovosten. Zato moremo zanj uporabiti metodo predznakov. V našem primeru pride v poštev varianta $u = y - x$. Posebnost našega primera je pet delavcev, za katere nismo mogli ugotoviti razlike v kakovosti in zato tudi ne predznaka. Teh pet primerov izpustimo iz analize in imamo $n^- + n^+ = n$; $10 + 25 = 35$. Za $n = 35$ dobimo v tablicah

P	0,005	0,01	0,025	0,05	0,10
n^-	9	10	11	12	13

Ker je $n^- = 10$ enak kritični vrednosti za $P = 0.01$, sklepamo, da je novi postopek s tveganjem $\alpha = 0,01$ boljši od standardnega.

16.44 Preskus s predznačenimi rangi. (Wilcoxonov test predznačenih rangov). Preskus s predznaki je vsekakor hiter in enostaven postopek, vendar je njegova hiba v tem, da upošteva le predznake in je moč takega preskusa seveda manjša. Kompromis med preskusi, ki so zasnovani na numeričnih vrednostih za u in preskusom s predznaki, so metode, ki upoštevajo range. Ta neparametrična metoda se zato imenuje metoda s predznačenimi rangi.

Po tej metodi:

a) glede na tip preskusa izračunamo vrednosti u

b) vrednostim u priredimo range po velikosti absolutnih vrednosti u .

c) seštejemo range za pozitivne u. Tako dobimo vrednost T.

d) iz tabele za kritične vrednosti presodimo značilnost odklona me(u) od 0.

16.45 Da bi preskusili, ali televizijsko reklamiranje učinkuje na prodajo, so za $n = 10$ trgovin primerjali dnevni promet pred in po reklamiranju in dobili so tele podatke:

promet pred x	po y	$u = y - x$	rang	rang s predznakom
87	92	+ 5	4	4
111	131	+ 20	10	10
99	96	- 3	2	- 2
78	88	+ 10	7	7
118	110	- 8	6	- 6
77	88	+ 11	8	8
67	60	- 7	5	- 5
89	93	+ 4	3	3
64	62	- 2	1	- 1
88	101	+ 13	9	9
				41 = T

Če je $n = 10$ so za enostranski preskus kritične meje za

α	·10	·05	·01
T	41	45	50

Za $T = 41$ je $\alpha = 0{,}10$ in obstaja sum na značilnost

16.46 Preskus z vsoto rangov (Mann-Withney U-preskus)

Če niso izpolnjeni pogoji za uporabo t-preskusa, s preskusom z vsoto rangov preskušamo, ali dva vzorca izhajata iz identičnih populacij. Preskus je zasnovan na ideji, da se v primeru, če izhajata dva vzorca iz identičnih populacij, poprečna ranga iz enega in drugega vzorca združene ranžirne vrste razlikujeta v mejah slučajnosti.

Preskus izvedemo po naslednjih točkah:

a) Numerične podatke preskusnih vzorcev z n_1 in n_2 enotami ($n_1 \leq n_2$) za primerjani populaciji združimo. Ko jih uredimo po velikosti, jim po velikosti priredimo ustrezne range.

b) Seštejemo range za vzorec z n_1 enotami $(n_1 \leq n_2)$

c) Dobljeno vsoto rangov T' primerjamo s kritičnimi vrednostmi v tabeli.[H] Kritične vrednosti v tabeli v zbirki tabel so dane za štiri ravni za enostranske preskuse $\alpha = P = 0.10,\ 0,05,\ 0,01$ in $0,001$.

Če sta n_1 in n_2 večja kot 10, za preskušanje domneve o enaki porazdelitvi dveh populacij uporabimo zakonitost, da se T' porazdeljuje približno normalno

$$T' =: N\left[E(T') = \frac{n_1 (n_1 + n_2 + 1)}{2}\ ;\ \mathrm{Var}\ T' = \frac{n_2}{6} E(T)\right] \qquad (16.34)$$

Moč preskusa z vsotami rangov je približno enaka kot moč preskusa za t-preskus. Če po populaciji nista normalni, more biti moč preskusa z vsotami rangov večja kot s t- preskusom, neglede na to, da za take primere ne moremo uporabiti t- preskusa, ker osnovne populacije ne zadoščajo pogojem za uporabo.

16.47 Da bi ugotovili, ali daje novi postopek v proizvodnji boljše rezultate v primerjavi s starim postopkom, smo iz proizvodnje po starem načinu (S) izbrali $n_2 = 6$ izdelkov, iz proizvodnje po novem načinu (N) pa $n_1 = 5$ izdelkov.

Preskusiti je treba, ali je nov način uspešnejši.

Produktivnost	S:	42	38	47	39	30	33					
	N:	50	49	45	54	40						
Združena ranžirna vrsta	S	S	S	S	N	S	N	S	N	N	N	
	30	33	38	39	40	42	45	47	49	50	54	
Rang S	1	2	3	4		6		8				
N					5		7		9	10	11	42=T'

V tabeli najdemo za $(n_1, n_2) = (5, 6)$, da so kritične vrednosti 38 40 43 -. Novi postopek je torej značilno boljši na stopnji $\alpha = 0,05$, ker je $40 < 42 < 43$

16.48 Analiza variance z rangi (Kruskal-Wallisov test). Podobno kot z osnovnimi vrednostmi tudi z analizo variance rangov preskušamo domnevo, da je k vzorcev z n_1, n_2 .. n_k enotami slučajnostno izbranih iz k populacij z enakimi porazdelitvami. Analizo variance rangov izvedemo po tehle stopnjah.

a) Imamo podatke k vzorcev z n_j ($j = 1,2 \ldots k$) enotami. Skupno imajo vsi vzorci $\sum n_j = n$ enot.

b) Vrednosti vseh vzorcev razvrstimo v skupno ranžirno vrsto in enotam v ranžirni vrsti pripišemo range 1, 2..... n v združeni ranžirni vrsti.

c) Seštejemo range za vzorce iz posameznih populacij v vsote R_j

d) Izračunamo vzorčni izraz

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1) \quad ; \quad \chi_p^2 \ (m=k-1) \qquad (16.35)$$

Če je domneva o enakosti porazdelitev pravilna, se vzorčni izraz H porazdeljuje približno v χ^2 - porazdelitvi z $m = k-1$ stopinjami prostosti. Ta približek je praktično uporaben že, če imajo posamezni vzorci več kot 5 enot.

16.49 Vzemimo za zgled opazovanje o porabi za neko živilo v štiričlanskih gospodinjstvih v treh različnih dohodkovnih skupinah: nizki dohodki N, srednji dohodki S in visoki dohodki V. Iz vsake skupine izberemo po $n_1 = n_2 = n_3 = 6$ gospodinjstev. Dobili smo tele rezultate.

N	24	28	27	19	25	14
S	22	29	26	27	30	33
V	36	31	30	36	39	40

Če razporedimo vrednosti vseh vzorcev v združeno ranžirno vrsto, dobimo:

		Rangi		
		N	S	V
14	N	1.		
19	N	2		
22	S		3	
24	N	4		
25	N	5		
26	S		6	
27	N	7.5		
27	S		7.5	
28	N	9		
29	S		10	
30	S		11.5	
30	V			11.5
31	V			13
33	S		14	
36	V			15
36	V			16
39	V			17
40	V			18
		28.5	52	90'5
		R_1	R_2	R_3

171

R

$$H = \frac{12}{18(18+1)} \cdot \left[\frac{28{\cdot}5^2}{6} + \frac{52^2}{6} + \frac{90{\cdot}5^2}{6} \right] - 3(18+1) = 11{,}46$$

Iz tablice za χ^2 dobimo kritično vrednost $\chi^2_{0.005}(m=k-1=2) = 10{,}6$

Ker je izračunani vzorčni izraz H večji kot $\chi^2_{0.005}(2) = 10{,}6$, sklepamo, da so razlike v uporabi po dohodkovnih skupinah značilne s tveganjem $\alpha = 0.005$

Kontrola vsote $\sum R_i = R = \frac{n(n+1)}{2} = \frac{18(18+1)}{2} = 171$

16.50 Preskusi o slučajnosti razporeda enot. Za niz problemov je študij vrstnega reda enot pomemben in kaže na določene zakonitosti oziroma značilnosti populacije oziroma procesov.

Potreba po proučevanju zaporedij enot se pojavlja na različnih področjih, posebno pri proučevanju časovnih procesov. Značilnosti v zaporedju enot se pojavljajo povsod tam, kjer se delovanje bistvenih faktorjev v zaporedju enot spreminja, kar se količinsko kaže na vrednostih zaporednih členov v vrsti podatkov.

16.51 Vzemimo za zgled proizvodnjo na določenem stroju. Pogoji dela naj se spreminjajo tako, da stroj proizvaja izdelke, za katere je premer vedno večji in večji. Če za proizvedene izdelke, razvrščene po vrstnem redu kakor so bili časovno proizvedeni, izmerimo premere, opazimo večanje premerov. Velikost premera dveh zaporedno proizvedenih izdelkov je med seboj odvisna, ker se proizvodni dejavniki časovno spreminjajo. V drugem primeru iz določenih vzrokov pojav ciklično spreminja svoje značilnosti. Podoben zgled zakonitosti v razporedu enot opažamo tudi v repih, kjer se osebe razporede po določenih značilnostih in vrstni red odstopa od slučajnostnega razporeda, ker obstaja med zaporednimi enotama v repu odvisnost, interes povezovanja ipd. Na podobne probleme naletimo pri proučevanju zaporedja členov v analizi časovnih vrst, v katerih se časovno spreminjanje delovanja dejavnikov kaže v odvisnosti med zaporednimi členi v časovni vrsti. Slučajnost razporeda enot oziroma elementov preskušamo v glavnem z dvema metodama: s preskusom sekvenc in s preskusom varianc.

16.52 Preskus s sekvencami. Vzemimo, da n_1 enot vzorca ali populacije ima, n_2 pa nima določene značilnosti. Če so enote po določenem vodilu razvrščene po vrstnem redu, dobimo zaporedje. Enote, ki imajo proučevano značilnost, zaznamujemo v zaporedju s +, enote, ki te značilnosti nimajo pa z - . Zaporedje znakov +- je v skladu z zakonitostmi enot z opazovano značilnostjo.

Vzemimo kot zgled tri zaporedja po n = 15 enot

+ + + + + + + + - - - - - - -

+ - + - + - + - + - + - + - +

+ + - + + + - - - + + + - + -

Na prvi pogled sklepamo, da prvo zaporedje ni slučajnostno, ker se najprej razvrste sami +, nato pa sami - . Podobno bi težko sklepali na slučajnost za drugo zaporedje, ker se + in - sistematično menjata. Rezultat slučajnostnega razporeda bi moglo biti tretje zaporedje, kjer se + in - v zaporedju pojavljata nepredvidoma.

Če vzamemo za "sekvenco" strnjeno zaporedje enako predznačenih enot, je število

sekvenc parameter za zaporedje. V prvem zaporedju je število sekvenc r = 2, v drugem r = 15, v tretjem r = 8. Že iz tega zgleda spoznamo, da je izredno majhno ali izredno veliko število sekvenc izraz neslučajnostnega razporeda enot.

Zaznamujemo z n_1 število pozitivnih, z n_2 število negativnih enot oziroma elementov, z r pa število sekvenc v zaporedju! V tabeli sta dani kritični meji za r za zaporedja z $n_1 \leq 20$ in $n_2 \leq 20$. Zaporedje štejemo s tveganjem $\alpha = 0,05$ za neslučajnostno, če je r enak ali manjši kot spodnja ali enak ali večji kot zgornja kritična meja.

Če je $n_1 > 20$ ali $n_2 > 20$, moremo uporabiti približek z normalno porazdelitvijo. Število sekvenc r se za take primere v približku porazdeljuje normalno

$$r =: N\left[E(r) = \frac{2 n_1 n_2}{n_1 + n_2} + 1;\ Var(r) = \frac{(E(r) - 1)(E(r) - 2)}{n_1 + n_2 - 1}\right] \qquad (16.36)$$

število sekvenc pa preskušamo z z - preskusom

$$z = \frac{r - E(r)}{\sqrt{Var(r)}} \qquad (16.37)$$

16.53 Uporaba preskusa s sekvencami je mnogostranska. Z njo odkrivamo morebitne periodične ali ciklične motnje v proizvodnem procesu. V takih primerih se na določen čas ali na določeno število izdelkov zaporedoma proizvedenih izdelkov pojavi krajša ali daljša sekvenca izdelkov, ki so bodisi neuporabni ali imajo neko drugo značilnost. V tem primeru je število sekvenc manjše kot bi pričakovali pri slučajnostnem razporedu. Študij sekvenc pomaga pri analizi dinamike nasploh.

Študij sekvenc po določenem znaku osvetljuje tudi razmere v repih ipd. Posebna prilagoditev preskusa s sekvencami je primerjava z mediano. Za zaporedje numeričnih podatkov poiščemo mediano in posameznim vrednostim priredimo znak +, če je vrednost nad in znak -, če je vrednost pod mediano. Preskus s številom sekvenc pokaže ali so vrednosti razporejene slučajnostno ali je v velikosti podana določena značil-

nost tendence porasta ali pada. Izraz 16.37 se znatno poenostavi, če je $n_1 = n_2$

$$z = \left[r - (n_1 + 1) \right] / \sqrt{\frac{n_1 (n_1 - 1)}{2 n_1 - 1}} \qquad (16.38)$$

16.54 V obratu, kjer mora biti glede na potrebe proizvodnega procesa ustaljena temperatura, vsake četrt ure merijo temperaturo. Za razdobje štirih ur so dobili naslednje rezultate:

18	20	17	16	15	17	19	18	16	17	15	14	16	17	18	19
+	+	-	-	-	-	+	+	-	-	-	-	-	-	+	+
	1		2		3		4							5	

Mediana je $M_e = 17$. Za $n_1 = 10$ meritev je $y < M_e$ in za $n_2 = 6$ meritev je $y > M_e$. Število sekvenc je enako $r = 5$.

Iz tabele povzamemo, da je kritično število sekvenc $r_s = 4$. Število sekvenc je torej nad spodnjo kritično in značilno mejo in iz tega preskusa ne moremo sklepati, da se je temperatura v času spreminjala sistematično.

16.55 S preskusom sekvenc preskušamo tudi značilnost razlik v porazdelitvi dveh populacij. Vzemimo, da smo v desetih trgovinah prodajali v dveh mesecih določeno blago na enak način, v petih pa smo poskusili blago prodajati po novem načinu. Za vseh petnajst trgovin smo izračunali indekse prodaje med obema mesecema in dobili tele rezultate:

Stari način : 122 105 120 105 122 121 107 111 103 115

Novi način : 128 127 130 125 130

Če obe vrsti podatkov razporedimo v združeno ranžirno vrsto, dobimo:

103	105	105	107	111	112	115	120	121	122	125	127	128	130	130
S	S	S	S	S	S	S	S	S	S	N	N	N	N	N

Ker je $n_1 = 10$, $n_2 = 5$, $r = 2$, iz tablice pa povzamemo, da je $r_s = 3$, sklepamo, da da novi način prodaje značilno večji promet.

16.56 Preskus o slučajnosti razporeda z variancami (von Neumannov preskus). Pri preskusu slučajnosti razporeda z variancami primerjamo poprečen kvadrat prvih razlik zaporednih členov v proučevani vrsti

$$\delta^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2 \quad \text{z varianco} \quad s^2 = \frac{1}{n} \sum_{1}^{n} (y_i - \bar{y})^2$$

Že če je število enot $n > 10$, populacija normalna, zaporedje pa slučajnostno, velja, da se izraz

$$z = \left(1 - \frac{1}{2}\eta\right)\sqrt{\frac{n^2 - 1}{n - 2}} \; ; \; \eta = \frac{\delta^2}{s^2} \qquad (16.39)$$

porazdeljuje standardizirano normalno. Značilno večji z je izraz velike pozitivne povezave med zaporednimi členi, značilno manjši z pa odraz negativne povezave med zaporednimi členi.

16.57 V proizvodnji smo na vsakih 10 minut izbrali po en izdelek in zanje dobili zaporedoma naslednje premere.

y_i 110 110 109 111 112 115 114 117 116 118 120 119 118 117 116

y_i 121 115 117 120 118

$$\bar{y} = 115{,}65$$

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{(110 - 115{,}65)^2 + (110 - 115{,}65)^2 + \ldots + (118 - 115{,}65)^2}{20} = 12{,}3275$$

$$\delta^2 = \frac{\sum (y_{i+1} - y_i)^2}{n-1} = \frac{(110-110)^2 + (109-110)^2 + \ldots + (118-120)^2}{20-1} = 6.1053$$

$$\eta = \delta^2 / s^2 = 6{,}1053/12{,}3275 = 0{,}4953$$

$$z = (1 - \frac{1}{2}\eta) \cdot \sqrt{\frac{n^2-1}{n-2}} = (1 - \frac{0{,}4953}{2}) \cdot \sqrt{\frac{20^2-1}{20-2}} = +3{,}5422$$

Ker je izračunani $z = 3{,}54$ večji kot $z_{0.001} = 3.29$, vzamemo, da so razlike v razporedu značilne na stopnji $\alpha = 2P = 0.001$. Členi v časovnem zaporedju so značilno odvisni.

SEDEMNAJSTO POGLAVJE

STATISTIČNA ANALIZA ODLOČITEV

Prvine

17.1 Matrika posledic. Ugotavljanje stanja za določen pojav s popisom, vzorci, subjektivnimi ocenami ali preskušanjem domnev ni samo sebi namen, temveč je osnova za odločitve, ki jih sprejema raziskovalec, podjetnik, kupec ipd.

Če vzamemo poenostavljen zgled, je od s t a n j a na trgu, ki je: ugodno ali neugodno za prodajo, odvisna o d l o č i t e v podjetja o u k r e p u , ali kupiti ali ne kupiti nov stroj, ki bo proizvedel določeno število izdelkov.

Rezultat določenega ukrepa ob nekem stanju merimo običajno denarno. Rezultat je pozitiven, če gre za dohodek in negativen, če gre za izgubo. Razen dohodka rabi kot merilo tudi: strošek, skladi itd. Če vzamemo, da bo dohodek $D_{11} = 100$ d.e, če bodo razmere ugodne in bomo nabavili nov stroj in $D_{21} = -50$ d.e izgube, če bodo razmere neugodne in bomo nabavili nov stroj in da ne bo ekonomskega učinka, če ne bomo nabavili nov stroj neglede na ugodno ali neugodno stanje $D_{12} = D_{22} = 0$, moremo dohodke ob posameznih priložnostih strniti v dohodkovno matriko.

Tabela 17.1 Matrika dohodkov za shematičen zgled

Stanje na trgu	ukrep: nov stroj	
	nabavimo A_1	ne nabavimo A_2
Ugodno S_1	$D_{11} = 100$	$D_{12} = 0$
Neugodno S_2	$D_{21} = -50$	$D_{22} = 0$

Iz tega zgleda moremo posplošiti matriko posledic na poljubno število stanj in poljubno število ukrepov.

Tabela 17.2 Splošna matrika dohodkov ${}_S D_A$

S_j \ A_i	A_1	A_2 --------	A_g
S_1	D_{11}	D_{12} - - - - -	D_{1g}
S_2	D_{21}	D_{22} - - - - -	D_{2g}
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
S_k	D_{k1}	D_{k2} - - - - -	D_{kg}

17.2 Priložnostna izguba. Odločitve niso odvisne od absolutnih vrednosti dohodka ali stroškov, temveč le od razlik pri posameznih ukrepih. Zato na odločitve ne vplivajo fiksni stroški, ki so neodvisni od ukrepov ipd. Da poenostavimo račune, pogosto namesto matrike dohodkov ali stroškov odnose med ukrepi izrazimo enostavneje z matriko ti. priložnostnih izgub. Če se pri danem stanju ne odločimo za najugodnejši ukrep, imamo določeno izgubo, ker smo zamudili priložnost, ki bi jo izkoristili, če bi se zanj odločili. Izguba zaradi neizkoriščene priložnosti Z za posamezen ukrep je enaka razliki med največjim uspehom in uspehom pri

določenem ukrepu

$$s^{Z}{}_{A} = s^{D}{}_{max} - s^{D}{}_{A} \qquad (17.1)$$

Izguba zaradi neizkoriščene priložnosti enostavno podaja odnose med uspehi pri posameznih ukrepih, le da je obratno sorazmerna z dohodkom: večjemu uspehu ustreza manjša priložnostna izguba in obratno.

Za naš zgled je matrika izgub zaradi neizkoriščene priložnosti takale:

Tabela 7.3 Matrika priložnostnih izgub $s^{Z}{}_{A}$

	$s^{D}A$		$s^{D}{}_{max}$	$s^{Z}A$	
	A_1	A_2		A_1	A_2
S_1	100	0	100	0	100
S_2	- 50	0	0	50	0

17.3 Indeks koristnosti. Dohodek ali izguba, izražena denarno, nista vselej ustrezno merilo za poslovni uspeh. Pri posameznih ukrepih nastopajo pogosto tudi drugi elementi. Podjetnik je pripravljen več tvegati, samo da si pridobi ali ohrani svoj sloves. Položaj na trgu narekuje, da prodajamo blago, pa čeprav za manjšo ceno. Konkurenca nas prisili, da ukrepamo drugače, kot bi ukrepali, če konkurence ne bi bilo. Teh dodatnih prvin ni možno vselej izraziti denarno. Pač pa jim moremo dati kvantitativen izraz z indeksom koristnosti, ki vsebuje vse: denarne in nedenarne posledice posameznih ukrepov. Indeks koristnosti je v linearni zvezi z dohodkom, če teh dodatnih elementov ni in je nelinearna funkcija dohodka, če te dodatne elemente upoštevamo. Zato pri poslovnih odločitvah denarno izražene dohodke ali izgube pogosto zamenjamo z ustreznimi indeksi koristnosti.

17.4 Pri odločanju o najugodnejšem ukrepu se v praksi znajdemo v različnih situacijah.

a) Stanje pojava poznamo. b) Poznamo verjetnosti $Pr(S_k)$ za nastop posameznega

stanja. c) Nimamo informacije o stanju.

Pri določanju o tem, kateri ukrep bomo realizovali, govorimo: v primeru a) o odločitvah pri gotovosti; v primeru b) o odločitvah pri tveganju in v primeru c) o odločitvah pri negotovosti.

17.5 Odločitve pri gotovosti. Za odločitve pri gotovosti je verjetnost za določeno stanje $Pr(S_1) = 1$, za vsa druga stanja pa 0. V tem primeru izberemo kot ustrezno odločitev ono, ki pri danem znanem stanju daje največji dohodek.

Če za naš zgled vemo, da je stanje na trgu ugodno, se bomo odločili za ukrep A_1 in kupili nov stroj, ker je $D_{11} = 100 > D_{12} = 0$.

17.6 Odločitve pri tveganju. Če ne poznamo stanja, ki pogojuje dohodek, poznamo pa verjetnosti $Pr(S_S)$, s katerimi pričakujemo posamezna stanja, se je primerno odločiti za ukrep, pri katerem je pričakovana vrednost (matematično upanje) za dohodek največja

$$E(D_A) = \sum_S Pr(S_S) D_{SA} = \max \qquad (17.2)$$

ali pričakovana vrednost za izgubo najmanjša

$$E(Z_A) = \sum_S Pr(S_S)\, Z_{SA} = \text{Min} \qquad (17.3)$$

Če v našem zgledu poznamo verjetnost, da je stanje na trgu ugodno $P_r(S_1) = 0.70$ in verjetnost, da je stanje neugodno $P_r(S_2) = 0.30$, dobimo

Tabela 17.4 Izračunanje najugodnejšega ukrepa pri tveganju iz matrike dohodkov

$S^D A$	A_1	A_2	$Pr(S_s)$	$Pr(S_s).D_{s1}$	$Pr(S_s)D_{s2}$
S_1	$D_{11} = 100$	$D_{12}=0$	$P_r(S_1) = 0.70$	$\cdot 70.100 = 70$	$\cdot 70.0 = 0$
S_2	$D_{21} = -50$	$D_{22}=0$	$P_r(S_2) = 0.30$	$+ \cdot 30(-50) = -15$	$\cdot 30.0 = 0$
				$E(D_{A_1}) = 55$	$E(D_{A2}) = 0$

Ker je pričakovana vrednost za dohodek pri ukrepu 1: $E(D_{A_1}) = 55$ večji od pričakovane vrednosti pri ukrepu 2 : $E(DA_2) = 0$, se odločimo za ukrep A_1 in stroj nabavimo.

Če isti problem rešimo z matriko priložnostnih izgub, dobimo:

Tabela 17.5 Izračunanje najugodnejšega ukrepa pri tveganju iz matrike priložnostnih izgub

$S^Z A$	A_1	A_2	$Pr(S_s)$	$Pr(S_s) \cdot Z_{S1}$	$P_r(S_s).Z_{S2}$
S_1	0	100	0.70	0	70
S_2	50	0	0.30	15	0

$E(Z_{A1}) = 15$ $E(Z_{A2}) = 70$

Odločimo se za ukrep z najmanjšo priložnostno izgubo, to je A_1 in stroj nabavimo. Vidimo, da je odločitev na osnovi priložnostne izgube skladna z odločitvijo na osnovi dohodka.

17.7 Odločitve pri negotovosti. Če ne poznamo, kakšno je stanje, ob katerem odločamo o ustreznem ukrepu, niti ne vemo, kolikšna je verjetnost, da nastopi določeno stanje, govorimo o odločitvah pri negotovosti. V tem primeru moremo odločati o optimalnem ukrepu po več različnih načinih.

17.8 Pravilo maksimin . Po maksimin metodi iz matrike dohodka ${}_SD_A$ za vsak ukrep A poiščemo najmanjši možen dohodek ${}_AD_{min}$ in se odločimo za tisti ukrep, pri katerem je ${}_AD_{min}$ največji. To sodilo je najopreznejše in ne izkorišča možnosti velikih dohodkov. Za naš zgled po tem sodilu najustreznejši ukrep dobimo takole:

Tabela 17.6 Določanje najugodnejšega ukrepa po sodilu maximin

Stanje na trgu	nabavimo A_1	ne nabavimo A_2	
Ugodno S_1	${}_1D_1 = 100$	${}_2D_1 = 0$	
Neugodno S_2	${}_1D_2 = -50$	${}_2D_2 = 0$	
${}_AD_{min}$	${}_1D_{min} = -50$	${}_2D_{min} = 0$	$\max({}_AD_{min}) = 0 = {}_2D_{min} \Rightarrow A_2$

Ker je ${}_2D_{min}$ maksimalen minimalni dohodek (odtod tudi ime), se odločimo za A_2 in stroja ne nabavimo.

17.9 Pravilo Maximax . Po tej metodi za vsak ukrep A poiščemo največji dohodek ${}_AD_{max}$ in se odločimo za ukrep, pri katerem je ${}_AD_{max}$ največji. To pravilo je najsmelejše in se usmerja na največje možne dohodke. Za naš zgled določimo po maximax metodi najustreznejši ukrep tako, da poiščemo za posamezne ukrepe največje možne dohodke: ${}_AD_{max}$ (${}_1D_{max} = 100$; ${}_2D_{max} = 0$) in se odločimo za ukrep A_1, ker je ${}_1D_{max} = 100$ ukrep z največjim dohodkom.

Tabela 17.7 Določitev najugodnejšega ukrepa po sodilu maximax

Stanje	Ukrep: nabavimo A_1	Ukrep: ne nabavimo A_2
ugodno S_1	${}_1D_1 = 100$	${}_2D_1 = 0$
neugodno S_2	${}_1D_2 = -50$	${}_2D_2 = 0$
${}_AD_{max}$	${}_1D_{max} = 100$	${}_2D_{max} = 0$

$$\max({}_AD_{max}) = 100 = {}_1D_{max} \Rightarrow A_1$$

17.10 Hurwiczovo pravilo. Po sodilu maximax se odločimo za A_1 in stroj nabavimo. Kombinacija maximin in maximax sodila je Hurwiczovo pravilo, po katerem ni osnova za izbor najustreznejšega ukrepa niti najmanjši ${}_AD_{min}$ niti največji ${}_AD_{max}$ dohodek pri posameznih ukrepih, temveč poprečje

$$A\bar{D} = {}_AD_{min} \cdot w_{min} + {}_AD_{max} \cdot w_{max} \qquad (17.4)$$

med obema. Da je ${}_A\bar{D}$ poprečje, mora veljati, da je $w_{min} + w_{max} = 1$

Kot najustreznejši ukrep vzamemo tistega, za katerega je ${}_A\bar{D}$ največji. To sodilo imenujemo Hurwiczovo pravilo.

Različni odnosi ponderov dajejo večji ali manjši povdarek maximin oziroma maximax sodilu. V skrajnosti preide pri $w_{min} = 1$; $w_{max} = 0$ Hurwiczovo pravilo v maximin sodilo, pri $w_{min} = 0$; $w_{max} = 1$ pa v maximax sodilo.

Vzemimo, da v našem zgledu damo sodilu maximax dvakrat večjo težo kot sodilu maximin. Ponderi so torej: $w_{min} = 1/3$; $w_{max} = 2/3$

Če po tem razmerju določimo najustreznejši ukrep, dobimo

Tabela 17.8 Določitev najustreznejšega ukrepa po Hurwiczovem pravilu

	Ukrep			
	A_1	A_2	A_1	A_2
U S_1	100	0		
NU S_2	- 50	0	${}_AD_m \cdot w_m$	
$w_{min} = 1/3\ {}_AD_{min}$	- 50	0	$-50 \cdot \frac{1}{3}$	0
$w_{max} = 2/3\ {}_AD_{max}$	100	0	$+100 \cdot \frac{2}{3}$	0
			$A\bar{D}$ 50	0

$\max({}_A\bar{D}) = 50 = {}_1\bar{D} \Rightarrow A_1$

Po tem Hurwiczovem pravilu je največji popredčen dohodek ${}_1\bar{D} = 50$ in se zato odločimo za ukrep A_1 in stroj nabavimo.

17.11 Minimax sodilo. Minimax pravilo izhaja iz matrike priložnostnih izgub ${}_AZ_S$. Kot najustreznejšo odločitev vzamemo po sodilu minimax ukrep z najmanjšo maksimalno izgubo.

V našem zgledu je situacija naslednja:

Tabela 17.8 Določitev najustreznejšega ukrepa po sodilu minimax

${}_AZ_S$	A_1	A_2	
S_1	0	100	
S_2	50	0	
${}_AZ_{max}$	${}_1Z_{max} = 50$	${}_2Z_{max} = 100$	$\min({}_AZ_{max}) = 50 = {}_1Z_{max} \Rightarrow A_1$

Po minimax sodilu je najustreznejši ukrep A_1, da stroj nabavimo.

17.12 Bayesovo sodilo. Pri odločitvah pri tveganju verjetnosti Pr (S) za posamezna stanja poznamo. Pri negotovosti pa verjetnosti Pr(S) ne poznamo. Pač pa jih moremo, kot že v drugih primerih, nadomestiti s subjektivnimi verjetnostmi pr(S), do katerih pridemo s subjektivnim poznavanjem problema, izkušnjami ipd. Ko določimo subjektivne verjetnosti, jih moremo uporabiti za osnovo pri določanju pričakovane vrednosti dohodka pri posameznih ukrepih, s tem da verjetnosti Pr(S), s katerimi računamo v primeru odločitev pri tveganju, zamenjamo s subjektivnimi verjetnostmi. Na osnovi tako izračunanih pričakovanih vrednosti $E({}_AD)$ se odločimo za ukrep, pri katerem je $E({}_AD)$ največji. Ta način odločanja je Bayesovo sodilo o najustreznejši odločitvi.

Vrnimo se na naš zgled. Če nimamo nikakršnih informacij o stanju na tržišču, ne moremo dati prednosti nobenemu stanju. V tem primeru vzamemo za vsa stanja enake subjektivne verjetnosti. V našem zgledu bi v tem primeru vzeli $pr(S_1) = \cdot 50$; $pr(S_2) = \cdot 50$, ker imamo samo dvoje stanj. Če pa se nagibamo k mnenju, da je dosti

večja verjetnost, da so razmere na trgu ugodne kot pa, da so neugodne, vzemimo npr. $pr(S_1) = '80$; $pr(S_2) = '20$.

S temi subjektivnimi verjetnostmi izračunamo pričakovane vrednosti $E({}_AZ)$ in z njihovo pomočjo najustreznejši ukrep po naslednjem postopku.

Pogosto dohodek ${}_AD_S$ zamenjamo pri Bayesovem sklepanju z izgubami oziroma priložnostnimi izgubami ${}_AZ_S$ in vzamemo za najustreznejši ukrep z najmanjšo pričakovano izgubo.

Tabela 17.9 Določitev najustreznejšega ukrepa po Bayesovem sodilu.

S	pr(S)	${}_AZ_S$		$pr(S) \cdot {}_AZ_S$	
		A_1	A_2	$pr(S) \cdot {}_1Z_S$	$pr(S) \cdot {}_2Z_S$
S_1	0.80	0	100	0.80.0= 0	0.80.100 = 80
S_2	0.20	50	0	0.20.50=10	0.20.0 = 0
	1.00			$E(A_1Z) = 10$	$E(A_2Z) = 80$

$$\min |E({}_AZ) = 10 = E({}_{A_1}Z) \Rightarrow A_1$$

Ker je najmanjša priložnostna izguba $E({}_{A_1}Z) = 10$, se odločimo za ukrep A_1 in stroj nabavimo.

17.13 Linearna funkcija dohodka in točka preloma.

V našem zgledu smo imeli zaradi enostavnosti razlaganja le dve stanji - ugodne in neugodne razmere na trgu. V stvarnih raziskavah število možnih stanj ni omejeno na dve stanji, temveč jih more biti več, celo neomejeno.

Stanje na trgu moremo podati s strukturnim deležem gospodinjstev, p, ki so pripravljena nabaviti določen mali aparat. Ta strukturni delež more biti vsako število med nič in ena oziroma med 0 in 100%. Za tak primer je dohodek pri posameznem ukrepu funkcija strukturnega deleža p ali na splošno parametra, ki pogojuje stanje

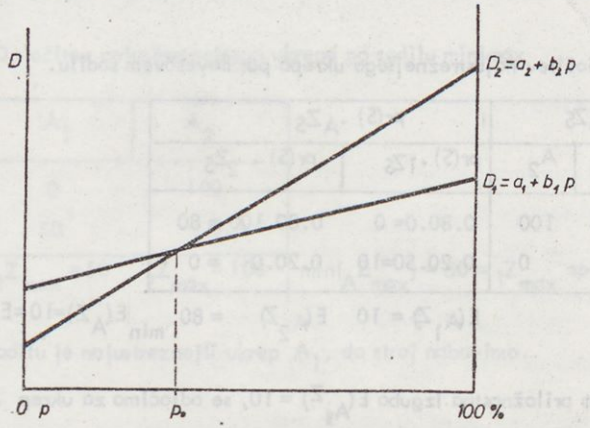
$$D_A = f_A(p) \qquad (17.5)$$

V poenostavljenem primeru, ki pa je praktično pogost, je funkcija dohodka linearna

funkcija parametra p

$$D_A = a_A + b_A \cdot p \qquad (17.6)$$

Vsakemu ukrepu ustreza posebna funkcija-premica za odvisnost dohodka od parametra. Če narišemo shematično primer za odločitev z dvema ukrepoma A_1 in A_2, dobimo za premici dohodka



Slika 17.1 Linearni funkciji dohodka za dva ukrepa

Iz slike 17.1 vidimo, da je do presečišča obeh premic pri gotovosti ustreznejši ukrep A_1, ker je za $p < p_o$ dohodek pri A_1 višji, za vrednosti $p > p_o$ pa je ugodnejši ukrep A_2, ker je v tem primeru dohodek pri A_2 večji kot pri A_1.

Točko preloma pri $p = p_o$, pri katerem sta oba ukrepa enakovredna, dobimo iz enačbe

$$a_1 + b_1 p_o = a_2 + b_2 p_o$$

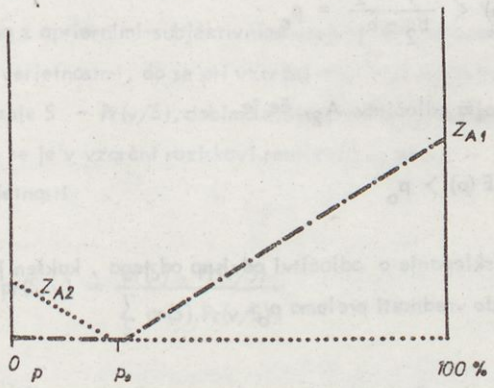
Iz nje sledi, da je vrednost p_o za točko preloma enaka

$$p_o = \frac{a_1 - a_2}{b_2 - b_1} \qquad (17.7)$$

Predpostavljamo, da je $b_2 > b_1$ in $a_1 > a_2$. Tako imamo pravilo o odločitvi o ukrepu A_1 ali A_2 neposredno iz deleža p

delež	dohodek	ukrep
$p < p_o$	$D_1 > D_2$	A_1
$p = p_o$	$D_1 = D_2$	neopredeljeno
$p > p_o$	$D_1 < D_2$	A_2

Iz slike o linearnih funkcijah dohodka moremo zlahka dobiti sliko o priložnostnih izgubah kot funkcijo deleža p. Ker je za $p < p_o$ D_1 večji kot D_2, je od nič do p_1 priložnostna izguba pri ukrepu A_1 enaka nič, za $p > p_o$ pa je priložnostna izguba pri ukrepu A_1 enaka razliki $D_2 - D_1$. Obratno pa je za ukrep A_2 priložnostna izguba za $p < p_o$ enaka $D_1 - D_2$, za $p > p_o$ pa enaka nič, ker je dohodek pri ukrepu A_2 največji.



Slika 17.2 Priložnostne izgube pri linearnih funkcijah dohodka

17.14 Pri odločitvah s tveganjem poznamo verjetnosti za posamezno stanje. Če je stanje dano z zvezno spremenljivko npr. deležem p, predpostavljamo, da pri odločitvah s tveganjem poznamo verjetnostno porazdelitev za p. Ta je bodisi normalna, binomska ipd.

Pri odločitvah s tveganjem je sodilo o uspešnosti pričakovana vrednost za dohodek

$$E({}_{A}D) \tag{17.8}$$

Če je funkcija dohodka linearna, sledi iz obrazca 17.6, da je

$$E({}_{A}D) = E(a_A + b_A p) = a_A + b_A E(p) \tag{17.9}$$

Pri odločitvah s tveganjem vzamemo, da je ukrep A_1 ustreznejši kot A_2, če je

$$E({}_{A_1}D) > E({}_{A_2}D) \tag{17.10}$$

iz česar sledi dalje po obrazcu 17.9 in 17.10

$$a_1 + b_1 E(p) > a_2 + b_2 E(p) \tag{17.11}$$

ali

$$E(p) < \frac{a_1 - a_2}{b_2 - b_1} = p_o \tag{17.12}$$

Analogno je ustreznejša odločitev A_2, če je

$$E(p) > p_o \tag{17.13}$$

Iz tega sledi, da je sklepanje o odločitvi odvisno od tega, kakšen je odnos pričakovane vrednosti $E(p)$ do vrednosti preloma p_o.

17.15 Bayesovo sodilo. Pri sklepanju pri negotovosti z Bayesovim sodilom je situacija analogna, le da pravo porazdelitev verjetnosti zamenjamo z apriorno-subjektivno verjetnostno porazdelitvijo in je pričakovana vrednost izračunana iz subjektivnih verjetnosti. Odločamo pa v primeru odločitev pri tveganju in pri odločitvah pri negotovosti po Bayesovem sodilu enako:

če je: $E(p) < p_o$, se odločimo za ukrep A_1

če je: $E(p) = p_o$, je odločitev neopredeljena

če je: $E(p) > p_o$, se odločimo za ukrep A_2

17.16 Bayesovo sodilo z vzorčenjem - posteriorna analiza.

Od sodil pri negotovosti je sklepanju s tveganjem najbolj podobno Bayesovo pravilo. Boljša je namreč subjektivna ocena verjetnosti za posamezna stanja, kot pa neupoštevanje verjetnosti posameznih stanj in sklepanje na osnovi skrajnih možnosti. Subjektivne verjetnosti določamo po poznavanju in izkušnjah brez konkretnih informacij o stanju. Če že ne moremo dobiti informacije o točnih objektivnih verjetnosti za nastop posameznega stanja, pa moremo subjektivne verjetnosti popraviti z rezultati vzorčenja. Nadaljnji postopek je enak kot pri Bayesovem pravilu z apriornimi subjektivnimi verjetnostmi: Kot najustreznejši ukrep vzamemo onega, za katerega je pričakovani dohodek največji ali pričakovana izguba najmanjša.

Če razpolagamo z apriornimi-subjektivnimi verjetnostmi za posamezna stanja pr(S) in s pogojnimi verjetnostmi, da se pri vzorčni raziskavi realizira vzorec z značilnostjo v, če je stanje S - Pr(v/S), dobimo subjektivno pogojno verjetnost pr(S/v), da je stanje S, če se je v vzorčni raziskavi realiziral vzorec z v po Bayesovem stavku o inverzni verjetnosti

$$pr(S/v) = \frac{pr(S) \cdot Pr(v/S)}{\sum_S pr(S).Pr(v/S)}$$

pr(S/v) so popravljene subjektivne verjetnosti za stanje S, če je rezultat vzorca v.

Po Bayesovi posteriorni analizi dalje določimo pričakovane vrednosti za izgubo pri posameznih ukrepih

$$E({}_AZ) = \sum_S {}_AZ_S \cdot pr(S/v)$$

in kot optimalen ukrep vzamemo onega, za katerega je $E({}_AZ)$ najmanjši.

17.17 Vzemimo, da v našem zgledu štejemo kot stanje na trgu za ugodno, če je $S_1 : P_1 = 25\%$ gospodinjstev pripravljenih kupiti nov aparat in kot neugodno, če je le $S_2 : P_2 = 10\%$ gospodinjstev pripravljenih kupiti nov aparat. Izvedli smo majhno anketo $n = 20$ gospodinjstev, od katerih so $n_a = 4$ gospodinjstva izrazila pripravljenost, da kupijo nov aparat.

Pri deležu P je verjetnost, da z vzorcem n izberemo n_a enot z dano značilnostjo, podano z binomskim obrazcem

$$P_r(n_a/n\ ;\ P) = \binom{n}{n_a} P^{n_a} (1-P)^{n-n_a}$$

ali v našem primeru s

$$P_r(4/20\ ;\ P) = \binom{20}{4} P^4 (1-P)^{20-4}$$

ali pri $S_1 : P_1 = 25\%$

$$P_r(n_a = 4/n=20;\ P_1 = 0.25) = \binom{20}{4} \cdot {}^{\cdot}25^4 \cdot {}^{\cdot}75^{16} = {}^{\cdot}1897$$

in pri $S_2 : P_1 = 10\%$

$$P_r(n_a=4/n = 20;\ P_2 = 0.10) = \binom{20}{4} \cdot {}^{\cdot}10^4 \cdot {}^{\cdot}90^{16} = {}^{\cdot}0898$$

Te verjetnosti uporabimo dalje v smislu gornjih pravil. Rezultati so podani v tabeli 17.10.

Tabela 17.10. Odločitev o najustreznejšem ukrepu s posteriorno analizo

S	$P_r(4/20;S)$	$P_r(S)$	$P_r(4 \cap S) = P_r(S).P_r(4/20;S)$	$P_r(S/4;\ 20) = \frac{P_r(4 \cap S)}{P_r(4)}$	s^{Z_A}		$P_r(S/4,20).s^{Z_A}$	
					A_1	A_2	A_1	A_2
$P_1=0.25$	·1897	·80	·80.·1897=·15176	$\frac{·15176}{·16972} = ·89$	0	100	0	89
$P_2=0.10$	·0898	·20	·20.·0898=·01796	$\frac{·01796}{·16972} = ·11$	50	0	5.5	0

$P_r(4) = ·16972$

$E(_1Z) = 5.5; E(_2Z) = 89$

$\min E(_AZ) = 5.5 = E(_1Z) \Rightarrow A_1$

Z vzorčnimi podatki smo popravili subjektivne verjetnosti v tem smislu, da je po izvršenem vzorcu še večja verjetnost ($p_r(S_1/4\ ;\ 20) = \cdot 89$), da je stanje na trgu ugodno. Razlika med pričakovano izgubo pri optimalnem in alternativnem ukrepu po posteriorni analizi je 89 - 55 = 83.5 za razliko od 80 - 10 = 70 pri priorni analizi.

Poskusimo, kako bi delovali v našem zgledu izsledki vzorca $n = 20$ enot z $n_a = 4$ pozitivnimi odgovori v primeru, da se ne moremo odločiti ali je večja verjetnost za ugodno ali neugodno stanje. V tem primeru vzamemo, da so subjektivne priorne verjetnosti za vsa stanja enake, v našem zgledu $p_r(S_1) = p_r(S_2) = \cdot 50$. Posteriorna analiza je v tem primeru nakazana v tabeli 17.11.

Tabela 17.11 Odločitev o najustreznejšem ukrepu s posteriorno analizo v primeru enakih priornih verjetnosti

S	$Pr(4/20,S)$	$p_r(S)$	$Pr(4 \cap S)$	$pr(S/4,20)$	S_{ZA}		$p_r(S/4,20) \cdot S_{ZA}$	
					A_1	A_2	A_1	A_2
$P_1 = \cdot 25$	$\cdot 1897$	$\cdot 50$	$\cdot 09485$	$\cdot 68$	0	100	0	68
$P_2 = \cdot 10$	$\cdot 0898$	$\cdot 50$	$\cdot 04490$	$\cdot 32$	50	0	16	0
		$P(4) = \cdot 13975$		$1 \cdot 00$			$E(z_{A_1}) = 16$	$E(Z_{A_2}) = 68$

Tudi v tem primeru se odločimo za A_1 in stroj nabavimo.

17.18 Preposteriorna analiza. Pri posteriorni analizi oziroma odločitvah smo vpeljali dodatne informacije, ki smo jih dobili z vzorčenjem. Pri dosti velikem vzorcu začne pričakovana vrednost za priložnostno izgubo za optimalni ukrep padati, vendar se te razlike z velikostjo vzorca manjšajo. Če ne upoštevamo, da imamo z vzorčenjem stroške, nekega optimalnega vzorca, pri katerem so pričakovane izgube manjše, ni, ker da večji vzorec manjšo pričakovano izgubo. Če pa upoštevamo stroške vzorčenja, ki se z večanjem vzorca večajo, pa skupna vsota pričakovanih izgub pri optimalnem ukrepu in stroških vzorčenja ne pada konstantno, temveč nekaj časa pada (dodatni stroški za večji vzorec so manjši kot je zmanjšanje pričakovane izgube), po nekem minimumu pa narašča (dodatni stroški za večji vzorec so večji kot pa je zmanjšanje

za pričakovano izgubo). Tako dobimo s preposteriorno analizo optimalen vzorec, pri katerem so prikazane izgube, pri katerih so všteti tudi stroški vzorčenja, najmanjše.

OSEMNAJSTO POGLAVJE

PROUČEVANJE ODVISNOSTI MED MNOŽIČNIMI POJAVI

18.1 V prejšnjih poglavjih smo proučevali posamezne znake samostojno, brez zveze z drugimi znaki. Te metode analize statističnih podatkov zelo podrobno obravnavajo značilnosti za posamezne znake. Vendar obsegajo samo del analize statističnih podatkov. Ne zajemajo namreč enega izmed najvažnejših problemov v analizi socialno-ekonomskih pojavov in množičnih pojavov na sploh, to je njihovo medsebojno odvisnost in povezanost. Pri proučevanju socialno-ekonomskih pojavov namreč često naletimo na problem povezanosti in odvisnosti med pojavi. Obseg proizvodnje kmetijskega obrata je odvisen od velikosti obrata, strukture osnovnih sredstev, delovne sile, klimatskih faktorjev itd. Vrednost proizvodnje industrijskega obrata je odvisna od števila delavcev, mehanizacije, produktivnosti dela, vrste proizvodnje itd. Cena je odvisna od količine blaga, ki je na trgu, osebni dohodek od kvalifikacije, službene dobe itd. V medsebojni povezavi sta tudi nepismenost moških in nepismenost žensk po občinah, starost ženina in neveste, itd. Podobnih primerov moremo našteti veliko. Vendar niso vsi pojavi med seboj povezani, čeprav bi po vsebini mogli odvisnost pričakovati. Imamo pa tudi primere nesmiselne povezave. Nesmiselno je npr. proučevati odvisnost med številom porok in množino padavin po letih, cene kmetijskih pridelkov od obolelosti za tuberkulozo itd.

FUNKCIJSKE ODVISNOSTI

18.2 O funkcijski odvisnosti med x in y v matematičnem smislu govorimo, če je dano neko pravilo zveze med neodvisno spremenljivko x in odvisno spremen-

ljivko y , po katerem določeni vrednosti neodvisne spremenljivke x ustreza ena ali več natančno določenih vrednosti odvisne spremenljivke y. S simbolom pišemo:

$$y = f(x) \qquad (18.1)$$

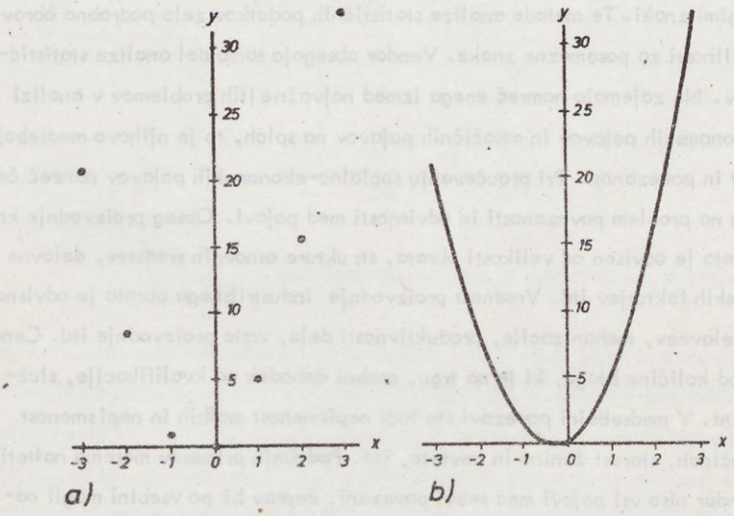
in pravimo: y je funkcija od x.

Funkcijska zveza med y in x je običajno nakazana z enačbo funkcije, iz katere moremo za vsak x izračunati ustrezno vrednost y. Tako je npr. za funkcijo

$$y = 2x + 3x^2 \qquad (18.2)$$

x = 2 ustrezna vrednost

$y = 2.2 + 3.2^2 = 16.$



Slika 18.1 Slika funkcijske odvisnosti $y = 2x + 3x^2$

Razen z obrazcem moremo funkcijsko odvisnost med x in y nakazati tudi z nizom dvojic ustreznih vrednosti x in y. Tako je v tabeli 18.1 dana funkcijska zveza iz

gornje enačbe za osem parov vrednosti x in y.

Tabela 18.1 Pari vrednosti za x in y

x	y
- 3	+ 21
- 2	+ 8
- 1	+ 1
0	+ 0
+ 1	+ 5
+ 2	+ 16
+ 3	+ 33
+ 4	+ 56

S sistemom dvojic pa za gornji zgled ne moremo podati funkcijske zveze med x in y za vse vrednosti x. To moremo prikazati edino z enačbo med x in y.

Niz dvojic vrednosti x in y moremo prikazati tudi grafično v pravokotnem koordinatnem sistemu. Posamezno dvojico vrednosti x, y po znanem načinu prikažemo s točko v pravokotnem koordinatnem sistemu. Kolikor dvojic ustreznih vrednosti x in y imamo, toliko imamo točk v koordinatnem sistemu. Medtem ko moremo niz dvojic vrednosti grafično prikazati s sistemom točk, prikažemo funkcijo, dano z obrazcem, s krivuljo. V določenem razmiku ustreza vsaki vrednosti x funkcijska vrednost y. Slika niza dvojic iz tabele 18.1 je prikazana v sliki 18.1a, slika funkcijske odvisnosti iz obrazca 18.2 pa v sliki 18.1b.

Funkcijske odvisnosti moremo torej prikazati na tri načine:

a) z enačbo: $y = f(x)$,

b) v tabeli z nizom ustreznih vrednosti x in y,

c) grafično s sistemom točk ali s krivuljo v pravokotnem koordinatnem sistemu.

18.3 Če prenesemo pojem funkcijske odvisnosti na množične pojave, bi imela v primeru funkcijske odvisnosti med površino in proizvodnjo kmetijskega gospodarstva vsa gospodarstva z enako površino enako proizvodnjo. Vendar to ni tako. Čeprav sta površina gospodarstva in proizvodnja med seboj odvisni, imata gospodarstvi z enakima površinama le redko enako proizvodnjo. Še več. Čeprav sodimo, da ima večje gospodarstvo večjo proizvodnjo, velja ta odvisnost samo na splošno, v posameznih primerih pa more imeti tudi večje gospodarstvo manjšo proizvodnjo. Do tega pride zato, ker proizvodnja ni odvisna samo od površine, temveč še od mnogo drugih faktorjev. Poizkus, da bi odstranili vse dodatne faktorje in tako prišli do funkcijske povezave med dvema znakoma, se ne bi posrečil. Vedno ostanejo neki faktorji, katerih vpliv ne moremo odstraniti in kontrolirati in jih štejemo med slučajne faktorje. Pri množičnih pojavih moremo torej opazovati le splošno tendenco odvisnosti, v posameznih primerih pa zakonitost zaradi delovanja dodatnih - posamičnih vplivov ni nujno vidna. Zato imenujemo za razliko od funkcijskih odvisnosti te vrste odvisnosti korelacijske odvisnosti.

Proučevanje korelacijskih odvisnosti je različno od proučevanja funkcijskih odvisnosti, čeprav imata obe vrsti proučevanja svoje stične točke.

Prikazovanje korelacijskih odvisnosti

Korelacijske odvisnosti prikazujemo na enake tri načine kakor funkcijske odvisnosti:
a) v tabeli z nizom dvojic vrednosti koreliranih podatkov za vsako enoto populacije, ali v korelacijski tabeli,
b) s točkami v korelacijskem grafikonu,
c) v funkcijski obliki z regresijsko funkcijo ali črto.

18.4 Niz dvojic podatkov. Z nizom dvojic vrednosti koreliranih podatkov na splošno prikazujemo osnovne podatke pri proučevanju korelacije med pojavi. Čeprav je ta način nepregleden in iz njega še ne dobimo vtisa o zakonitosti povezave

med dvema pojavoma, ga pogosto uporabljamo, ker je osnova za vsa nadaljnja proučevanja.

Ker je vir podatkov o množičnih pojavih populacija, posamezna dvojica koreliranih podatkov velja za posamezne enote proučevane populacije.

V tabeli 18.2 so prikazani osnovni podatki za proučevanje korelacije med odstotkom nepismenega prebivalstva (x) in odstotkom nepismenih žensk (y) v SFRJ v letu 1961 po okrajih.

Tabela 18.2 Odstotki nepismenega prebivalstva (x) in odstotki nepismenih žensk (y) v SFRJ v letu 1961 po okrajih

BiH		Črna Gora		Hrvatska		Makedonija		Slovenija		Srbija	
x	y	x	y	x	y	x	y	x	y	x	y
35,3	50,7	21,7	33,2	10,6	14,0	20,9	28,8	1,9	1,9	9,2	14,5
37,8	53,6			19,1	28,4	30,6	44,4	3,1	3,6	23,2	36,9
32,5	48,3			11,0	15,2	25,0	36,2	1,2	1,2	25,5	40,6
32,7	47,7			13,1	17,5	20,6	28,8	1,9	2,1	33,1	49,4
26,8	39,9			5,9	8,1	25,1	34,1			24,9	38,1
34,6	51,7			17,0	25,2	33,8	48,5			24,4	38,3
				18,1	26,4	19,7	27,7			26,0	41,5
				7,4	10,2					26,7	43,3
				7,6	10,7					25,8	38,5
										9,8	14,1
										12,1	16,8
										11,6	16,6
										8,6	11,6
										12,1	17,2
										41,1	56,0

V tabeli 18.3 so prikazani dohodki in stroški za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957.

Tabela 18.3 Dohodki v tisočih dinarjev (x) in stroški za kulturno in družbeno življenje v dinarjih (y) za 36 delavskih družin v Mariboru v novembru 1957.

(Vir: Anketa ZS SRS o življenju delavcev in nameščencev)

x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7	16,5
y	260	330	490	310	580	560	330	480	590	330	630
x	16,9	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4
y	630	650	1100	760	460	700	540	750	370	630	740
x	28,4	25,8	30,5	29,8	27,5	26,5	26,7	26,7	32,9	37,5	30,8
y	1080	630	990	1210	1100	870	640	1110	1610	1590	1070
x	31,8	38,4	35,9								
y	1600	1710	1810								

V tabeli 18.4 so podatki o odstotkih travniških in pašniških površin za okraje v SRS po stanju v letu 1952.

Tabela 18.4 Odstotki travniških in pašniških površin po okrajih v SRS po stanju leta 1952.

(Vir: Statistični bilten SRS)

Okraj	Odstotek površin pod	
	travniki	pašniki
Celje - mesto	22	8
Celje - okolica	3	28
Črnomelj	9	20
Gorica	5	13
Kočevje	6	15
Kranj	7	7
Krško	16	9
Ljubljana - mesto	28	3

Okraj	Odstotek površin pod	
	travniki	pašniki
Ljubljana - okolica	18	11
Ljutomer	18	6
Maribor - mesto	15	6
Maribor - okolica	15	8
Murska Sobota	21	3
Novo mesto	8	9
Postojna	10	19
Ptuj	18	10
Radovljica	5	16
Sežana	1	37
Slovenj Gradec	7	10
Šoštanj	8	9
Tolmin	5	20
Trbovlje	9	8

13.5 Iz gornjih treh primerov sklepamo še na določeno vsebinsko razliko med korelacijskimi odvisnostmi. V prvih dveh primerih je jasno, da je delež nepismenih odvisen od republike, stroški za kulturne in družbene potrebe pa od dohodkov, ne pa obratno. To so v z r o č n e povezave oziroma odvisnosti, ker je en pojav vzrok, drugi pa posledica ali statistično, en znak faktorialen, drug pa rezultativen. Vzročne povezave pa ne zasledimo v primeru odstotka površin, ker ne moremo reči, da je visok odstotek travniških površin vzrok za nizek odstotek pašniških površin ali obratno. Kljub temu, da med njima ni vzročne odvisnosti, pa sta ta dva podatka med seboj povezana, ker na oba vplivajo isti pogoji, ki imajo v našem primeru za posledico, da je travnikov mnogo, pašnikov pa malo in obratno. Dva taka odločilna skupna faktorja sta npr. nadmorska višina in vrsta tal.

Na podobno nevzročno č i s t o k o r e l a c i j s k o p o v e z a v o naletimo npr. tudi pri proučevanju korelacije med odstotkom nepismenih žensk in moških po okrajih. Čeprav med nepismenostjo moških in žensk ni vzročne povezanosti, sta med seboj v korelaciji, ker na oba deluje isti kompleks faktorjev, od katerih je odvisna nepismenost. To so npr. število šol v okraju, kulturno prosvetna dejavnost v preteklosti itd.

Medtem ko pri vzročnih odvisnostih po pravilu proučujemo samo odvisnost posledice od vzroka, moremo pri čistih korelacijskih odvisnostih proučevati odvisnost znaka x od y in obratno, odvisnost znaka y od x.

18.6 Korelacijska tabela. Že pri majhnem številu enot je prikazovanje posamičnih dvojic koreliranih podatkov v gorniji obliki zelo obširno in nepregledno. Zato za večje populacije prikazujemo korelirane podatke v kombinacijski tabeli, ki jo imenujemo korelacijska tabela. Kakor dobimo frekvenčno porazdelitev, če podatke uredimo v razrede po enem znaku, dobimo korelacijsko tabelo, če podatke uredimo v razrede po obeh koreliranih znakih hkrati. Položaj frekvenc v korelacijski tabeli zelo nazorno pokaže smer povezave med koreliranima znakoma.

Kot zgled je v tabeli 18.5 prikazana korelacijska tabela med poprečnim neto osebnim dohodkom vseh zaposlenih (x) in poprečnim osebnim dohodkom direktorjev v letu 1963 za 205 izbranih podjetij iz ankete o neto osebnih dohodkih v letu 1963.

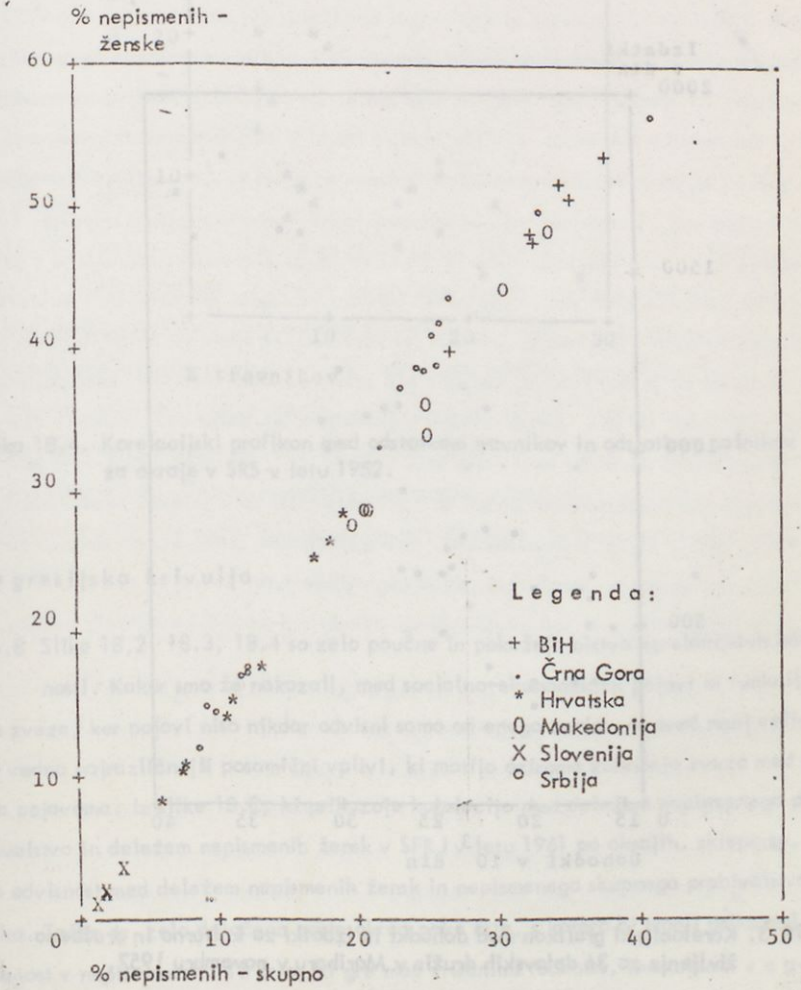
Tabela 18.5. Korelacijska tabela med poprečnim mesečnim dohodkom zaposlenih in poprečnim mesečnim dohodkom direktorjev za 205 v vzorec izbranih obrtnih podjetij v letu 1963 v SFRJ (v tisoč starih dinarjih).
Vir: SZS SB 369, Beograd, 1965.

Ø dohodek direktorja v tisoč din y \ Ø dohodek zaposlenih v tisoč din x	-10	10-20	20-30	30-40	40-50	50-60	Skupno
20 - 40	3	18	16	3	1	-	41
40 - 60	-	19	66	17	1	-	103
60 - 80	-	-	12	20	10	7	49
80 - 100	-	-	-	4	2	3	9
100 - 120	-	-	-	1	1	1	3
Skupno	3	37	94	45	15	11	205

Podatki v tabeli 18.5 pomenijo število enot, ki ima vrednost ustreznemu kombiniranemu razredu. Tako je npr. bilo 19 podjetij, v katerih je bil poprečen osebni dohodek zaposle-

nega od 10-20 tisoč din, poprečen dohodek direktorja pa med 40 in 60 tisoč din. Korelacijska tabela je torej frekvenčna porazdelitev, v kateri je populacija razdeljena v razrede po dveh znakih hkrati. Iz korelacijske tabele je jasno razvidna korelacijska odvisnost med x in y.

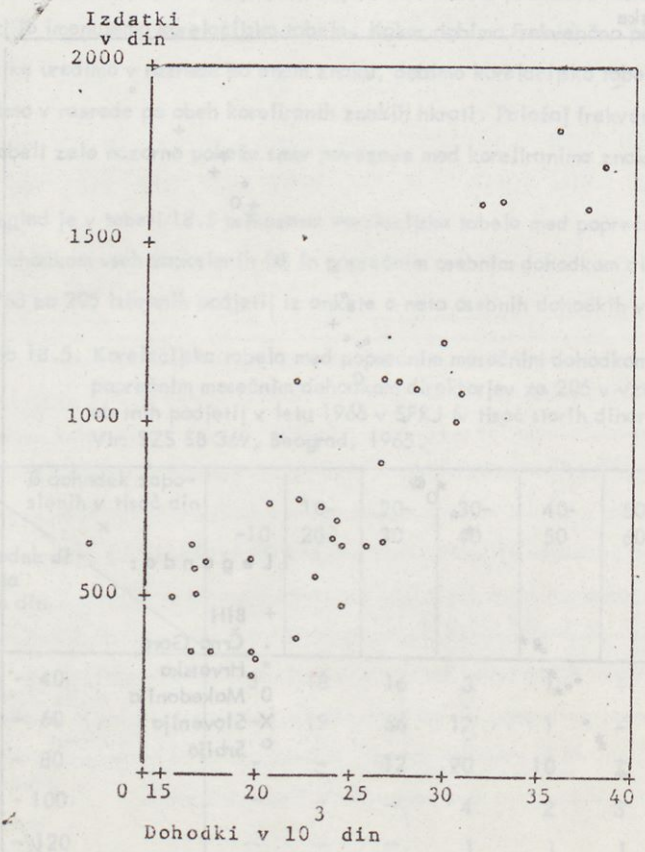
18.7 Korelacijski grafikon. Nazorno prikažemo korelacijo med dvema znakoma v korelacijskem grafikonu. V njem je vsaka dvojica vrednosti prikazana



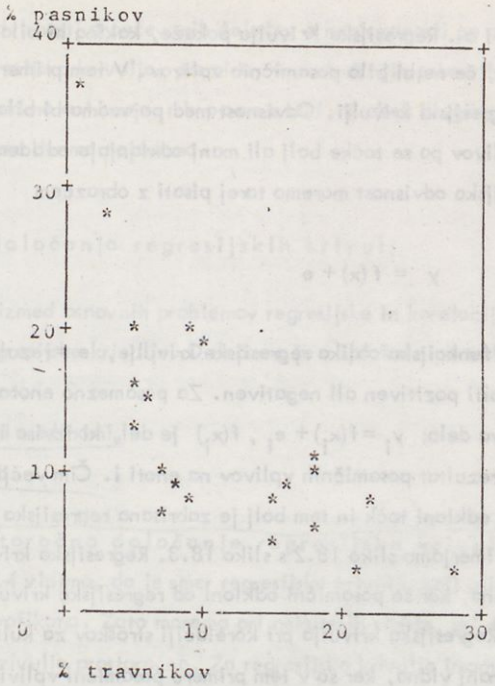
Slika 18.2. Korelacijski grafikon med odstotkom nepismenega prebivalstva in odstotkom nepismenih žensk v SFRJ v letu 1961 po okrajih.

s točko v pravokotnem koordinatnem sistemu, v katerem je na abscisi skala za znak x, na ordinati pa skala za znak y. V korelacijskem grafikonu je torej toliko točk, kolikor je enot populacije. Ker s pogledom zajamemo vse točke hkrati, je korelacija med dvema znakoma v korelacijskem grafikonu vidna neposredno.

V slikah 18.2, 18.3 in 18.4 so prikazani korelacijski grafikoni podatkov iz tabel 18.2, 18.3 in 18.4.



Slika 18.3. Korelacijski grafikon med dohodki in izdatki za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957.



Slika 18.4. Korelacijski grafikon med odstotkom travnikov in odstotkom pašnikov za okraje v SRS v letu 1952.

Regresijska krivulja

18.8 Slike 18,2, 18.3, 18.4 so zelo poučne in pokažejo bistvo korelacijskih odvisnosti. Kakor smo že nakazali, med socialno-ekonomskimi pojavi ni funkcijske zveze, ker pojavi niso nikdar odvisni samo od enega znaka, temveč nanj vplivajo vedno najrazličnejši posamični vplivi, ki motijo oziroma zabrišejo zvezo med dvema pojavoma. Iz slike 18.2, ki prikazuje korelacijo med deležem nepismenega prebivalstva in deležem nepismenih žensk v SFRJ v letu 1961 po okrajih, sklepamo, da je odvisnost med deležem nepismenih žensk in nepismenega skupnega prebivalstva velika. Točke se zelo določeno goste okrog neke črte, s katero bi mogli prikazati odvisnost v nepismenosti. To črto, ki gre med vrisanimi točkami, imenujemo r e g r e -

s i j s k o k r i v u l j o . Regresijska krivulja pokaže, kakšna bi bila zveza med koreliranima znakoma, če ne bi bilo posamičnih vplivov. V tem primeru bi bile točke za vse enote na regresijski krivulji. Odvisnost med pojavoma bi bila funkcijska. Zaradi posamičnih vplivov pa se točke bolj ali manj odklanjajo od idealne regresijske krivulje. Korelacijsko odvisnost moremo torej pisati z obrazcem

$$y = f(x) + e \qquad (18.3)$$

Pri tem pomeni: $f(x)$ = funkcijska oblika regresijske krivulje, e = rezultat posamičnih vplivov, ki more biti pozitiven ali negativen. Za posamezno enoto moremo vrednost y razstaviti v dva dela: $y_i = f(x_i) + e_i$, $f(x_i)$ je del, ki izvira iz povezanosti med y in x, e_i pa je rezultat posamičnih vplivov na enoti i. Čim večji so posamični vplivi e, tem večji so odkloni točk in tem bolj je zabrisana regresijska krivulja. To nazorno vidimo, če primerjamo sliko 18.2 s sliko 18.3. Regresijska krivulja je v prvem primeru dobro vidna, ker so posamični odkloni od regresijske krivulje v primeru nepismenosti majhni. Regresijska krivulja pri korelaciji stroškov za kulturno in družbeno življenje pa je manj vidna, ker so v tem primeru posamični vplivi zelo veliki. Iz tega sklepamo, da je v prvem primeru korelacijska odvisnost večja kakor v drugem. Korelacijska odvisnost more torej biti v e č j a a l i m a n j š a . Največja je v skrajnem primeru funkcijske odvisnosti ($e = o$), najmanjša pa v primeru neodvisnosti ($f(x) = C$). Če sta dva pojava neodvisna, na korelacijskem grafikonu ne moremo začrtati črte, za katero bi mogli reči, da je regresijska črta. Takrat se točke brez reda goste okrog točke, ki ima za koordinati aritmetični sredini koreliranih znakov.

V prvih dveh zgledih opazimo, da je smer regresijske krivulje taka, da se veča vrednost enega znaka, če se veča vrednost drugega. Takim vrstam povezav pravimo p o z i t i v n e p o v e z a v e . Če pa se z večanjem enega znaka vrednost drugega manjša, pa pravimo, da je p o v e z a v a n e g a t i v n a . Negativna povezanost je med odstotkoma travniške in pašniške površine po okrajih, prikazana v sliki 18.4. Iz gostitve točk moremo sklepati, da je za okraje z večjim odstotkom travnikov odstotek pašnikov v splošnem manjši in obratno.

V sliki o odvisnosti proučevanih deležev o nepismenosti po okrajih, je črta, ki ponazarja regresijsko krivuljo, premica, v obeh drugih primerih pa krivulja. V prvem primeru govorimo o linearni povezavi, v obeh drugih pa o krivuljčni povezavi med dvema pojavoma.

Metode določanja regresijskih krivulj

18.9 Eden izmed osnovnih problemov regresijske in korelacijske analize je določitev regresijske krivulje. Določamo jo v splošnem po treh metodah:

a) prostoročno,

b) z grupnimi sredinami,

c) analitično.

18.10 Prostoročno določanje regresijske krivulje. Iz slik 18.2, 18.3 in 18.4 vidimo, da je smer regresijske krivulje bolj ali manj vidna že iz korelacijskega grafikona. Zato moremo pri nekaterih slabše, pri drugih pa bolje, včrtati regresijsko krivuljo prostoročno. Za regresijsko krivuljo imamo krivuljo, ki jo včrtamo med točke tako, da se meglici točk najbolje prilega. Seveda ne sme biti vrisana črta preveč komplicirana, marveč čimbolj izglajena. V naših primerih regresijsko krivuljo najlaže včrtamo za pojava, prikazana v sliki 18.2, za katera je povezava največja. Zato je regresijska krivulja najbolj vidna. Težja je odločitev v drugih dveh zgledih.

Prostoročno včrtovanje regresijskih črt je prikladno, ker je izredno preprosto. Ima pa to napako, da je subjektivno. Regresijsko črto rišemo prostoročno kot osnovo za druge načine določanja regresije. V dosti primerih pa je kljub subjektivnosti ravno zaradi svoje hitrosti v primerjavi z drugimi metodami zelo koristna.

18.11 Metoda grupnih sredin. Če izhajamo iz obrazca

$$y = f(x) + e \qquad (18.4)$$

za določanje regresijske krivulje s pridom uporabimo grupne sredine. Če za enote,

ki imajo iste ali ne preveč različne vrednosti x, izračunamo poprečje iz vrednosti y, dobimo:

$$\bar{y} = \bar{f}(x) + \bar{e} \doteq f(\bar{x}) \qquad (18.5)$$

Ker štejemo, da je x konstanten, je konstantna tudi vrednost f(x), poprečje konstante pa konstanta. Po znanih stavkih o sredinah pa se v poprečju vpliv posamičnih ali slučajnih faktorjev sicer ne uniči, omili se pa vsekakor. Z vrsto poprečij po grupah x dobimo tako niz točk, ki so vsaj v bližini regresijske krivulje, če že niso na njej. Če te točke vrišemo v korelacijski grafikon in zvežemo z daljicami, dobimo lomljeno črto, ki je zaradi gornjih lastnosti dober približek poteka regresijske krivulje.

Po metodi grupnih sredin določamo regresijsko krivuljo torej takole:

a) Dvojice vrednosti x, y grupiramo v razrede po vrednosti x. Grupe morajo biti take, da razredi za x niso preširoki, da pa število enot v njih ni premajhno.

b) Za vsako grupo izračunamo aritmetični sredini za vrednosti x in y. Tako dobimo vrsto dvojic grupnih aritmetičnih sredin $\bar{x}_k$ in $\bar{y}_k$.

c) V korelacijskem grafikonu narišemo točke, ki imajo kot koordinati ustrezne grupne aritmetične sredine za x in y $(\bar{x}_k, \bar{y}_k)$. Tako imamo v grafikonu toliko točk, ki leže na regresijski krivulji ali v njeni bližini, kolikor imamo grup.

d) Vrisane točke grupnih sredin med seboj zvežemo z daljicami. Lomljena črta, ki jo dobimo, je približno regresijska krivulja.

18.12 Po tej metodi določimo regresijsko krivuljo za odvisnost stroškov za kulturno in družbeno življenje za 36 delavskih gospodinjstev v Mariboru. Podatki so vzeti iz tabele 18.3.

Ker je med dohodki in stroški vzročna zveza, vzemimo dohodke za neodvisno spremenljivko (x), stroške pa za odvisno spremenljivko (y). Po analizi podatkov glede na velikost dohodkov vzamemo razrede dohodkov po 5.000 din in po njih razdelimo dvojice vrednosti x in y. Tako grupirani podatki so navedeni v tabeli 18.6.

Tabela 18.6. Dohodki in izdatki za kulturno in družbeno življenje iz tabele 18.3, grupirani po višini dohodkov.

(x = dohodki v tisoč dinarjev ; y = izdatki v dinarjih)

15,0-	x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7	16,5	16,9
19,9	y	260	330	490	310	580	560	330	480	860	590	330	630
20,0-	x	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4		
24,9	y	650	1100	760	460	700	540	750	370	630	740		
25,0-	x	28,4	25,8	29,8	27,5	26,5	26,7	26,7					
29,9	y	1080	630	1210	1100	870	640	1110					
30,0-	x	32,9	30,8	31,8	30,5								
34,9	y	1610	1070	1600	990								
35,0-	x	37,5	38,4	35,9									
	y	1590	1710	1810									

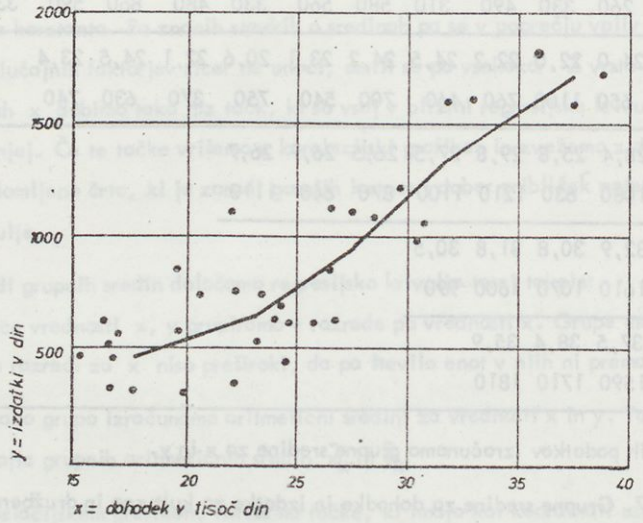
Iz grupiranih podatkov izračunamo grupne sredine za x in y.

Tabela 18.7. Grupne sredine za dohodke in izdatke za kulturno in družbeno življenje za 36 delavskih družin v Mariboru iz tabele 18.6

Dohodki v tisoč din	N_k	X_k	Y_k	$\bar{x}_k$	$\bar{y}_k$
15,0 - 19,9	12	212,2	5750	17,7	479
20,0 - 24,9	10	230,6	6700	23,1	670
25,0 - 29,9	7	191,4	6640	27,3	949
30,0 - 34,5	4	126,0	5270	31,5	1318
35,0 -	3	111,8	5110	37,3	1703

Če grupne sredine vnesemo v korelacijski grafikon, dobimo sliko 18.5. Iz slike vidimo, da se dobljena regresijska črta resnično prilagaja osnovnim vrednostim in da ponazarja smer odvisnosti med dohodki in stroški za kulturno in družbeno življenje.

18.13 Po metodi grupnih sredin moremo določiti regresijsko črto tudi, če so podatki grupirani v korelacijski tabeli. Grupne sredine v tem primeru izračunamo po katerikoli izmed metod za izračunavanje sredin iz frekvenčnih porazdelitev. Te grupne sredine včrtamo nad sredine razrednih razmikov.



Slika 18.5. Regresijska črta med dohodki in izdatki za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957.

Za korelacijo med poprečnimi dohodki zaposlenih in dohodki direktorjev v obrti iz tabele 18.5 je smiselno izračunati obe regresijski vrsti sredin, ker zveza ni vzročna. Ker so v korelacijski tabeli podatki že grupirani, grupiranje odpade.

Če najprej izračunamo poprečja za dohodke direktorjev za posamezne dohodkovne skupine zaposlenih, dobimo regresijsko črto, ki kaže odvisnost dohodkov direktorjev od dohodkov zaposlenih. Ker je v frekvenčni porazdelitvi dohodkov direktorjev za dohodkovno skupino do 10 tisoč din ena sama frekvenca, je ocena poprečja sredina razreda $\bar{y}_1 = 30$. Za skupino 10-20 je $\bar{y}_2 = \frac{18.30+19.50}{37} = 40.27$ tisoč din itd.

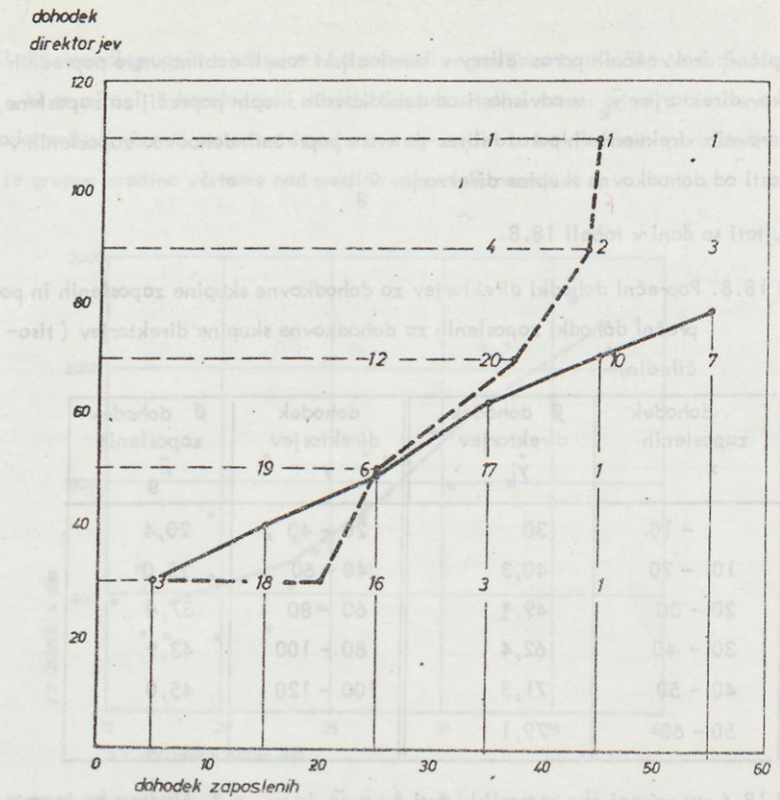
Iz navpičnih frekvenčnih porazdelitev v korelacijski tabeli dobimo vrsto poprečnih dohodkov direktorjev $\bar{y}_k$ v odvisnosti od dohodkovnih skupin poprečij za zaposlene, iz vodoravnih frekvenčnih porazdelitev pa vrsto poprečnih dohodkov zaposlenih v odvisnosti od dohodkovne skupine direktorjev $\bar{x}_g$.

Ti rezultati so dani v tabeli 18.8.

Tabela 18.8. Poprečni dohodki direktorjev za dohodkovne skupine zaposlenih in poprečni dohodki zaposlenih za dohodkovne skupine direktorjev (tisočih din)

dohodek zaposlenih x	Ø dohodek direktorjev $\bar{y}_k$	dohodek direktorjev y	Ø dohodek zaposlenih $\bar{x}_g$
- 10	30	20 - 40	20,4
10 - 20	40,3	40 - 60	25,0
20 - 30	49,1	60 - 80	37,4
30 - 40	62,4	80 - 100	43,9
40 - 50	71,3	100 - 120	45,0
50 - 60	79,1		

V sliki 18.6 sta vrisani obe regresijski črti (x_k, $\bar{y}_k$ in y_g, $\bar{x}_g$). Medtem ko je prva, ki kaže, kako se poprečje dohodkov direktorjev spreminja v odvisnosti od dohodkov zaposlenih, skoraj linearna, je druga, ki kaže obrnjene odnose, krivuljčna. Obe regresijski črti se praviloma sekata. Zaradi nazornosti so v sliki v ustreznih poljih vpisane tudi frekvence.



Slika 18.6 Regresijske črte sredin med dohodki zaposlenih in dohodki direktorjev iz tabele 18.8.

18.14 Analitična metoda. Regresijsko krivuljo moremo določiti tudi analitično. Če po analizi podatkov ugotovimo ustrezen tip funkcije

$$y' = f(x;\ a, b, c...) \qquad (18.6)$$

ki ima več parametrov (a, b, c...), primeren za regresijsko krivuljo, je treba glede na podatke, ki jih proučujemo, poiskati vrednosti parametrov a, b, c..., tako da se regresijska krivulja danim vrednostim najbolj prilega. Kot merilo prilagojeno-

sti krivulje vzamemo vsoto kvadratov odklonov stvarnih vrednosti y od y´

$$\sum_i (y_i - y_i')^2 = F(a, b, c...) \qquad (18.7)$$

Za regresijsko krivuljo vzamemo izmed vseh krivulj istega tipa tisto, za katero je vsota kvadratov odklonov stvarnih vrednosti y od vrednosti y´ na regresijski krivulji F(a,b,c...) najmanjša. Po tej metodi, ki jo imenujemo m e t o d o n a j m a n j š i h k v a d r a t o v , moremo določiti vrednosti parametrov za regresijsko krivuljo in tako regresijsko krivuljo samo. Kadar ugotavljamo regresijsko krivuljo analitično, uporabljamo najpogosteje metodo najmanjših kvadratov.

Mere j a k o s t i o d v i s n o s t i

18.15 Dejavnike, ki vplivajo na določen pojav (y), moremo razdeliti v tri skupine:

a) na splošne vplive, ki so za vse enote isti.

b) na vpliv faktorja x, s katerim je y v korelaciji.

c) na ostale posamične vplive skupno s slučajnostnimi.

Tako je na primer dohodek zaposlenih odvisen od:

a) splošnih vplivov: kategorije zaposlenih, časa, za katerega velja dohodek itd.,

b) let službe, s katerimi je dohodek v korelacijski zvezi,

c) drugih, posamičnih vplivov, kot so: vestnost pri delu, šolska izobrazba, delovno mesto, itd.

Prav tako je na primer proizvodnja kmetijskega obrata odvisna od:

a) splošnih faktorjev: klimatskih razmer, specializiranosti itd.

b) velikosti gospodarstva, s katero je proizvodnja v korelacijski povezanosti in

c) od drugih, posamičnih faktorjev, ki vplivajo na proizvodnjo, kot so: razlike tal, različna produktivnost dela, različno seme itd.

Če poznamo aritmetično sredino $\bar{y}$ in regresijsko krivuljo y´ med x in y, moremo celotno vrednost y za posamezno enoto razdeliti na nakazane tri komponente, od

katerih vsaka prikazuje rezultat vplivov po naštetih treh skupinah.

y je rezultat vseh dejavnikov, ki vplivajo na pojav; splošnih, koreliranega in drugih posamičnih; y' je rezultat splošnih dejavnikov in koreliranega dejavnika x; $\bar{y}$ je rezultat samo splošnih faktorjev. Zato moremo pisati

$$y = \bar{y} + (y' - \bar{y}) + (y - y') \qquad (18.8)$$

Pri tem pomeni:
$\bar{y}$ = rezultat splošnih vplivov; $(y' - \bar{y})$ = rezultat vpliva koreliranega znaka x; $(y - y')$ = rezultat vpliva drugih, posamičnih dejavnikov.

Tako smo uspeli razdeliti celotno vrednost y po značaju dejavnikov v tri ločene sestavine.

Če je vpliv dejavnika x velik, je odklon $(y' - \bar{y})$ velik. Če je vpliv x majhen, so vrednosti $(y' - \bar{y})$ majhne, če pa y od x ni odvisen, je $(y' - \bar{y})$ za vse enote enak nič. Za skupno merilo velikosti vpliva x na y uporabimo varianco odklonov $(y' - \bar{y})$

$$\sigma^2_{y'} = \frac{1}{N} \sum (y' - \bar{y})^2 \qquad (18.9)$$

Enako so odkloni $(y - y')$ veliki, če so močni posamični dejavniki. Če je njihov vpliv majhen, so ti odkloni majhni, če jih ni, je $(y - y')$ enak nič. Povezava je v tem primeru funkcijska. Skupno merilo jakosti posamičnih vplivov je analogno zgornjemu merilu varianca odklonov $(y - y')$

$$\sigma^2_e = \frac{1}{N} \sum (y - y')^2 \qquad (18.10)$$

Po stavku o razstavljanju variance na vsoto varianc po dejavnikih (glej obrazec 10.18) velja, da je skupna varianca σ^2_y vsota dveh varianc: variance zaradi vpliva x na y $\sigma^2_{y'}$ in variance zaradi vpliva posamičnih dejavnikov σ^2_e.

Ker tako pojasnimo, da en del celotne variance izvira iz povezave y z x, imenujemo $\sigma^2_{y'}$ pojasnjeno varianco, za razliko od σ^2_e, ki izvira iz drugih, neznanih

dejavnikov in jo imenujemo nepojasnjeno varianco. Z obrazcem moremo gornje napisati:

$$\sigma_y^2 = \sigma_{y\cdot}^2 + \sigma_e^2 \qquad (18.11)$$

Če to enačbo delimo z σ_y^2 dobimo

$$\frac{\sigma_{y\cdot}^2}{\sigma_y^2} + \frac{\sigma_e^2}{\sigma_y^2} = 1 \qquad (18.12)$$

V tem obrazcu kvocient $\sigma_{y\cdot}^2 / \sigma_y^2$ pove, kolikšen del skupne variance y je pojasnjene s povezanostjo y z x, σ_e^2 / σ_y^2 pa, kolikšen del skupne variance y odpade na druge, nepojasnjene vplive.

Velikost izraza $\sigma_{y\cdot}^2 / \sigma_y^2$ se ravna po tem, kako velik je vpliv x na y. Če y od x ni odvisen, je vrednost tega izraza nič, če pa je odvisnost funkcijska, je ta izraz največji in sicer enak 1. Vrednost kvocienta $\sigma_{y\cdot}^2 / \sigma_y^2$ je torej med 0 in 1 ter pomeni večjo ali manjšo stopnjo vpliva x na y. Zato kvocient $\sigma_{y\cdot}^2 / \sigma_y^2$ uporabljamo za merilo stopnje odvisnosti y od x in ga imenujemo determinacijski koeficient. Čim večji je vpliv x na y, tem večja je namreč vrednost determinacijskega koeficienta.

Čeprav je determinacijski koeficient $\sigma_{y\cdot}^2 / \sigma_y^2$ vsebinsko zelo nazoren, saj je delež pojasnjene variance od skupne variance, dostikrat namesto njega uporabljamo kot merilo povezanosti kvadratni koren tega izraza. To merilo korelacije imenujemo indeks korelacije. Indeks korelacije je torej razmerje med standardnim odklonom za pojasnjeno variabilnost in skupnim standardnim odklonom za y

$$I_{y.x} = \sigma_{y\cdot} / \sigma_y \qquad (18.13)$$

Če upoštevamo enačbi 18.12 in 18.13, moremo indeks korelacije pisati v obliki, ki jo splošno uporabljamo

$$I_{y.x} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}} \qquad (18.14)$$

Standardna napaka ocene

18.16 Če je povezava med x in y tesna, se točke za dejanske vrednosti x in y razmeroma malo odklanjajo od regresijske krivulje. Ako za dano enoto poznamo samo vrednost x, moremo vzeti ustrezno vrednost y' na regresijski črti za oceno dejanske vrednosti y. To ocenjevanje ima veliko praktično vrednost. Če poznamo regresijsko krivuljo med količino padavin in hektarskim pridelkom, moremo iz količine padavin v določenem letu napovedati hektarski pridelek. Če imamo regresijsko krivuljo med oceno in stvarnim hektarskim pridelkom, moremo z njo popraviti vizuelno oceno. Če poznamo regresijsko krivuljo o odvisnosti cene od količine blaga, ki je naprodaj, moremo pri znani količini blaga oceniti ceno itd.

Ta ocena je jasno tem boljša, čim bolj je y odvisen od x. Najboljša bi bila seveda, če bi bila funkcijska odvisnost. Takrat bi se iz regresijske krivulje izračunana vrednost y' skladala s pravo vrednostjo y. Vendar, kakor vemo, takih primerov v stvarnosti ni.

Za vsako posamezno oceno ne vemo, koliko smo pri ocenjevanju pogrešili, ali z drugimi besedami: kolikšna je razlika ocene y' od stvarne vrednosti y. Vendar moremo izračunati standardni odklon teh napak. Napaka pri ocenjevanju izvira iz posamičnih vplivov, ki jih ne poznamo. Standardni odklon napake je torej enak standardnemu odklonu iz nepojasnjene variance. To pa moremo izračunati, če poznamo indeks korelacije $I_{y.x}$ med podatkoma y in x. Iz obrazca 18.14 namreč zlahka dobimo, da je

$$\sigma_e = \sigma_y \sqrt{1 - I_{y.x}^2} \qquad (18.15)$$

Standardni odklon nepojasnjene variance zato imenujemo tudi **standardno napako ocene**. Če predpostavimo, da se odkloni ocen od pravih vrednosti porazde-

ljujejo normalno, moremo sklepati, da 68% ocen ne bo napačnih za več kot σ_e, približno 95% ne več kot $2.\sigma_e$ in da bo le izjemoma ocena od prave vrednosti različna za več kot $3.\sigma_e$. Tako moremo s standardno napako ocene oceniti napako ocene, ki jo dobimo z regresijsko enačbo.

Koeficient $\sqrt{1 - i^2_{y.x}}$ s katerim pomnožimo standardni odklon σ_y, da dobimo standardno napako ocene σ_e, imenujemo na splošno koeficient alienacije ali koeficient nepovezanosti.

LINEARNA ODVISNOST

18.17 Pokazovalci. Dosedanja razglabljanja o regresiji in korelaciji veljajo na splošno, neodvisno od tipa regresijske krivulje. V vsakem primeru izračunamo regresijsko krivuljo, ki pokaže obliko odvisnosti med x in y, indeks korelacije, ki pokaže jakost povezanosti in standardno napako ocene, ki pokaže kvaliteto ocenjevanja z regresijsko krivuljo.

Vse te količine najenostavneje izračunamo, če je odvisnost med x in y linearna, ali drugače povedano, če je regresijska krivulja premica. Linearne odvisnosti so v praksi zelo pogoste. Tudi če je povezava krivuljčna, je na krajših odsekih premica zelo dober približek regresijske krivulje. Zato je linearna regresija za prakso zelo važna in jo najpogosteje uporabljamo.

Če upoštevamo načela metode najmanjših kvadratov, izračunavamo pokazovalce linearne regresije in korelacije po osnovnih obrazcih:

Ker imamo v splošnem dve regresijski krivulji, imamo tudi dve regresijski premici. Prva

$$y' = M_y + b_1 (x - M_x) \tag{18.16}$$

podaja odvisnost y od x, druga

$$x' = M_x + b_2 (y - M_y) \tag{18.17}$$

pa pokaže odvisnost x od y.

V gornjih enačbah pomenita M_x in M_y aritmetični sredini za x in y, b_1 in b_2 pa sta p r v i in drugi regresijski koeficient. Regresijska koeficienta sta smerna koeficienta obeh regresijskih premic.

Prvi regresijski koeficient (b_1) pokaže, za koliko se v poprečju spremeni y, če se x poveča za enoto; drugi regresijski koeficient b_2 pa pokaže, za koliko se v poprečju spremeni x, če se y poveča za enoto. Regresijska koeficienta b_1 in b_2 računamo po obrazcih

$$b_1 = \frac{C_{xy}}{\sigma_x^2} \quad ; \quad b_2 = \frac{C_{xy}}{\sigma_y^2} \qquad (18.18)$$

Pri tem pomenita σ_x^2 in σ_y^2 varianci za znaka x in y. C_{xy} pa je nov izraz, ki ga imenujemo k o v a r i a n c o. Kovarianca je po definiciji poprečen produkt odklonov x in y od ustreznih aritmetičnih sredin M_x in M_y in jo izrazimo z obrazcem

$$C_{xy} = \frac{1}{N} \sum (x - M_x)(y - M_y) \qquad (18.19)$$

Za razliko od variance more biti vrednost kovariance pozitivna ali negativna.

Indeks korelacije pri linearni korelaciji imenujemo k o r e l a c i j s k i k o e f i c i e n t. Zaznamujemo pa ga z ϱ_{xy}.

Korelacijski koeficient ϱ_{xy} izračunamo po osnovnem obrazcu

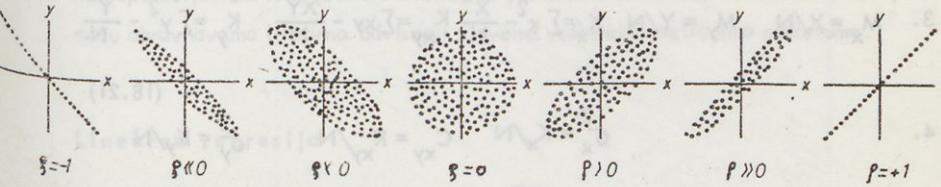
$$\varrho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} \qquad (18.20)$$

Vrednosti korelacijskega koeficienta leže med -1 in +1. Korelacijski koeficient je pozitiven, če je povezava med x in y pozitivna; to pomeni, da se veča tudi vrednost druge spremenljivke, če se veča vrednost prve.

Korelacijski koeficient pa je negativen, če je povezava negativna. Taka pa je povezava, če se z večanjem enega znaka vrednost drugega manjša.

Korelacijski koeficient ima vrednost -1 oziroma +1, če je povezava med x in y funkcijska in linearna, 0 pa, kadar med x in y ni linearne zveze.

Slika 18.7 nazorno kaže velikost korelacijskega koeficienta za nekaj najznačilnejših oblik linearnih odvisnosti med x in y. Te povezave so ponazorjene s korelacijskim grafikonom.



Slika 18.7. Vrednosti korelacijskega koeficienta pri različnih razmestitvah točk v korelacijskem grafikonu.

Kvadrat korelacijskega koeficienta ρ^2_{xy} imenujemo determinacijski koeficient za linearno odvisnost. Determinacijski koeficient sicer nima predznaka, ki bi nakazoval smer povezave, pač pa pove, kakšen del celotne variance je pojasnjen z linearno zvezo med x in y. Zato je analitično pomembnejši kakor korelacijski koeficient. Njegove vrednosti so med 0 in 1.

Računanje pokazovalcev za linearno odvisnost

18.18 Če imamo dane posamične podatke x_i, y_i (i = 1,2..N) za celotno populacijo, izračunamo pokazovalce linearne regresije in korelacije po shemi:

	x	y	x^2	xy	y^2
1.	x_1	y_1	x_1^2	x_1y_1	y_1^2
	x_2	y_2	x_2^2	x_2y_2	y_2^2
	.	.	.	.	.
	.	.	.	.	.
	x_N	y_N	x_N^2	x_Ny_N	y_N^2
2.	$X=\sum x$	$Y=\sum y$	$\sum x^2$	$\sum xy$	$\sum y^2$

3. $M_x = X/N \quad M_y = Y/N \quad K_x = \sum x^2 - \frac{X^2}{N} \quad K_{xy} = \sum xy - \frac{XY}{N} \quad K_y = \sum y^2 - \frac{Y^2}{N}$

(18.21)

4. $\sigma_x^2 = K_x/N \quad C_{xy} = K_{xy}/N \quad \sigma_y^2 = K_y/N$

5. $\sigma_x = \sqrt{\sigma_x^2} \quad \sigma_y = \sqrt{\sigma_y^2}$

6. $b_1 = \frac{C_{xy}}{\sigma_x^2} = \frac{K_{xy}}{K_x}; \quad \varrho_{xy}^2 = b_1 b_2 = \frac{C_{xy}^2}{\sigma_x^2 \sigma_y^2}; \quad b_2 = \frac{C_{xy}}{\sigma_y^2} = \frac{K_{xy}}{K_y}$

7. $\varrho_{xy} = \frac{K_{xy}}{\sqrt{K_x K_y}}$

8. $\sigma_{e.y} = \sigma_y \sqrt{1 - \varrho_{xy}^2}; \quad \sqrt{1 - \varrho_{xy}^2}; \quad \sigma_{e.x} = \sigma_x \sqrt{1 - \varrho_{xy}^2}$

9. $y' = M_y + b_1 (x - M_x) \quad x' = M_x + b_2 (y - M_y)$

V shemi je nazorno nakazan potek za računanje pokazovalcev linearne odvisnosti. Pogosto posamičnih rezultatov iz točke 1 niti ne pišemo, ker na računskem stroju dobimo vsote iz točke 2 direktno

Včasih se izkaže tehnično koristno, če obračunamo pokazovalce linearne odvisnosti iz reduciranih vrednosti $u = x - x_o$ in $v = y - y_o$. Vse količine, razen $M_x = x_o + M_u$; $M_y = y_o + M_v$ so invariantne in velja $\sigma_x^2 = \sigma_u^2$; $\sigma_y^2 = \sigma_v^2$;

$C_{xy} = C_{uv}$; $b_{1xy} = b_{1uv}$; $b_{2xy} = b_{2uv}$; $\varrho_{xy} = \varrho_{uv}$

Zglede za izračunanje bomo podali v naslednjih odstavkih.

18.19 Pogosteje kot iz podatkov za celotno populacijo proučujemo linearno odvisnost oziroma korelacijske odvisnosti nasploh iz populacij, ki so ali resnično slučajnostni vzorci ali pa jih štejemo za slučajnostne vzorce iz neomejene-umišljene populacije. Za take primere dobimo za parametre odvisnosti le ocene. V tem primeru obravnavamo linearno odvisnost z dvema vsebinsko različnima modeloma.

Linearna regresija

18.20 Linearni regresijski model, ki je dostikrat v skladu z dejansko odvisnostjo med dvema pojavoma, je

$$y = \alpha + \beta x + \varepsilon \tag{18.22}$$

V modelu sta α in β parametra regresijske premice

$$y' = \alpha + \beta x \tag{18.23}$$

x pa neodvisna spremenljivka oziroma znak, ki ni slučajnostna spremenljivka, in ga moremo vnaprej določiti. ε je slučajnostna spremenljivka, ki je rezultat slučajnostnih vplivov. Zanjo predpostavljamo, da se porazdeljuje

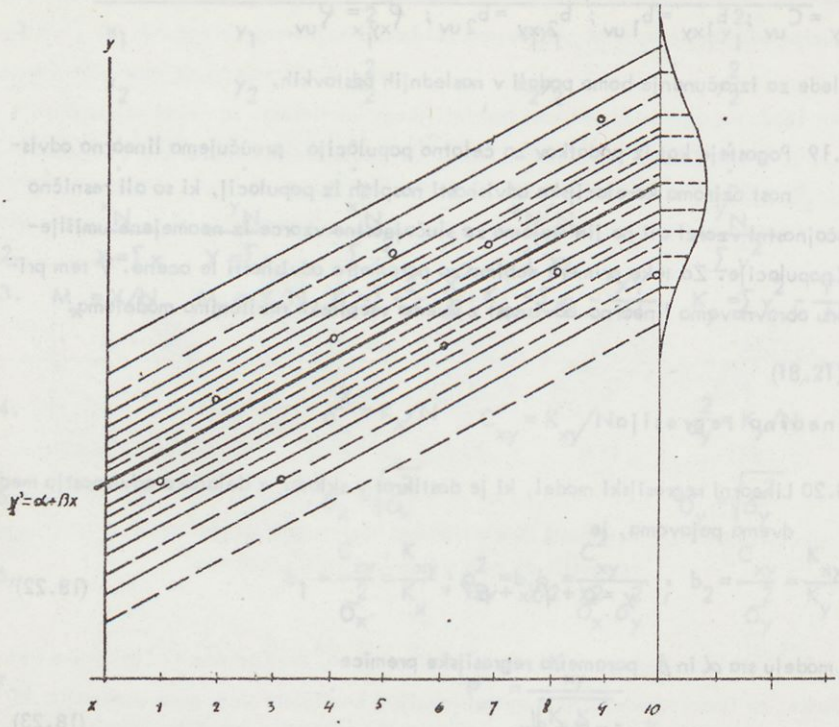
$$\varepsilon =: N(0, \sigma_\varepsilon^2) \tag{18.24}$$

normalno z $E(\varepsilon) = 0$ in $Var\,\varepsilon = \sigma_\varepsilon^2$. V enačbi 18.22 je y slučajnostna spremenljivka, ki se pri danem x porazdeljuje

$$y =: N(E(y) = y' = \alpha + \beta x ; \sigma_\varepsilon^2) \tag{18.25}$$

Najboljšo oceno (ocene z najmanjšimi variancami) dobimo, če ocenimo parametre po metodi najmanjših kvadratov, tako, da je za oceno regresijske premice

$$y^* = a + b x \tag{18.26}$$



Slika 18.8. Linearni regresijski model

vsota kvadratov odklonov dejanskih vrednosti y vzorca od ocen y'' najmanjša

$$F(a,b) = \sum_{i=1}^{n} (y_i - y_i'')^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \min \qquad (18.27)$$

Potreben pogoj za nastop ekstrema je, da sta parcialna odvoda po a in b enaka nič.

$$\frac{\partial F}{\partial a} = -2 \cdot S\,(y - a - bx) = 0$$

$$\frac{\partial F}{\partial b} = -2 \cdot S\,(y - a - bx)x = 0 \qquad (18.28)$$

Če ti enačbi preuredimo, dobimo sistem dveh linearnih enačb

$$\begin{aligned} a\,n + b\,S_x &= S_y \\ a\,S_x + b\,S_x^2 &= S_{xy} \end{aligned} \qquad (18.29)$$

ki ju imenujemo normalni enačbi. Če ti enačbi rešimo, dobimo, da je ocena regresijske premice

$$y'' = \bar{y} + b(x - \bar{x}) \qquad (18.30)$$

Pri tem je: $\bar{x} = \frac{1}{n} S_x$; $\bar{y} = \frac{1}{n} S_y$; $b = \frac{c_{xy}}{s_x^2} = \frac{k_{xy}}{k_x}$

$$k_x = S_x^2 - \frac{(S_x)^2}{n} \quad ; \quad k_{xy} = S_{xy} - \frac{S_x \cdot S_y}{n}$$

$$c_{xy} = \frac{S(x - \bar{x})(y - \bar{y})}{n - 1} = \frac{S_{xy} - \frac{(S_x)(S_y)}{n}}{n - 1} = \frac{k_{xy}}{n - 1} \; ; \; c_{xx} = s_x^2 \; ; \; c_{yy} = s_y^2$$

s_x^2 in s_y^2 sta nepristranski oceni varianc σ_x^2 , σ_y^2 , c_{xy} pa nepristranska ocena kovariance C_{xy}.

Za oceno regresijske premice y´

$$y'' = \bar{y} + b(x - \bar{x})$$

se oceni parametrov a in b porazdeljujeta, kot je nakazano v 18.31.

$$\bar{y} =: N\left(M_y; \frac{\sigma_\varepsilon^2}{n}\right) ; \quad b = \frac{c_{xy}}{s_x^2} =: N\left(\beta; \frac{\sigma_\varepsilon^2}{k_x}\right) \qquad (18.31)$$

Nepristranska ocena za varianco σ_ε^2 je

$$\hat{\sigma}_\varepsilon^2 = s_e^2 = \frac{k_y - \frac{k_{xy}^2}{k_x}}{n - 2} \qquad (18.32)$$

in se porazdeljuje

$$\frac{(n-2)\,s_e^2}{\sigma_\varepsilon^2} =: \chi^2 \quad (m = n-2) \tag{18.33}$$

Ocena regresijske premice

$$y'' = \bar{y} + b\,(x - \bar{x}) \tag{18.34}$$

je tudi slučajnostna spremenljivka, ker vsebuje $\bar{y}$ in b.

Pri danem x se y'' porazdeljuje normalno

$$y''_x = \bar{y} + b(x-\bar{x}) =: N(E\,y''_x = y'_x = M_y + \beta(x-\bar{x});\ \mathrm{Var}\,y''_x = \sigma_\varepsilon^2\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{k_x}\right] \tag{18.35}$$

Ocene za regresijsko premico y'' se zaradi $\bar{y}$ odklanjajo navzgor ali navzdol od $y' = \alpha + \beta\,(x - \bar{x})$ paralelno, ali pa so zaradi različnih b strmejše ali položnejše od $y' = \alpha + \beta\,(x - \bar{x})$.

Če na koncu opazujemo še posamezne vrednosti y v modelu, spoznamo, da se ocene $y_x = \bar{y} + b(x - \bar{x}) + e$ porazdeljujejo tudi normalno z

$$y_x =: N(E\,y_x = y' = M_y + \beta\,(x - \bar{x});\ \mathrm{Var}\,y_x = \sigma_\varepsilon^2\left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{k_x}\right]) \tag{18.36}$$

18.21 Iz gornjih zakonitosti dobimo po znanih postopkih več obrazcev, ki jih uporabljamo bodisi za preskušanje domnev ali za določanje razmikov zaupanja za parametre linearne regresije.

Iz zveze $z/\sqrt{\chi^2/m} = t(m)$ sledi, da je

$$\frac{b-\beta}{s_e}\sqrt{k_x} =: t\ (m = n-2) \tag{18.37}$$

Ta obrazec služi za preskušanje domnev o regresijskem koeficientu ali za računanje razmika zaupanja za β ker velja

$$P_r\left(b - t_p(n-2)\frac{s_e}{\sqrt{k_x}} < \beta < b + t_p(n-2)\frac{s_e}{\sqrt{k_x}}\right) = 1 - \alpha \qquad (18.38)$$

Za oceno regresijske premice $y'' = \bar{y} + b(x - \bar{x})$ pa dobimo ploskev zaupanja, znotraj katere leži z danim tveganjem $\alpha = 2P$ regresijska premica $y' = \alpha + \beta(x - \bar{x})$. Meje te ploskve so

$$\bar{y}+b(x-\bar{x})-t_p(n-2)\,s_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{k_x}} < y' = \alpha+\beta(x-\bar{x}) < \bar{y}+b(x-\bar{x})+t_p(n-2)\,s_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{k_x}} \qquad (18.39)$$

18.22 Vzemimo za zgled odvisnost kakovosti proizvodnje, ki je merjena s številom slabih izdelkov na 1000 kosov (y), od hitrosti dela strojev (x), ki je merjena s številom obratov na enoto časa. Ocene osnovnih parametrov so izračunane po standardnih obrazcih za računanje regresije iz 18.18 in po zakonitostih 18.20. Ker je vzorec majhen, so ocene dokaj nezanesljive.

Tabela 18.9 Izračunanje pokazovalcev linearne regresije za odvisnost kakovosti dela stroja (y) od hitrosti stroja (x)

x	y
164	114
102	70
131	96
158	90
109	57
132	94
81	60
138	92
120	70
108	75
174	123
149	122

$S_x = 1566$ $S_y = 1063$ $Sx^2 = 212876$ $Sxy = 144889$ $Sy^2 = 99899$

$\bar{x} = 130,5$ $\bar{y} = 88,6$ $-X^2/n = -204363$ $- XY/n = -138721$ $-Y^2/n = -94164$

$k_x = 8513$ $k_{xy} = 6168$ $k_y = 5735$

$$b = \frac{k_{xy}}{k_x} = \frac{6168}{8513} = 0,7245 ; y = 88,6 + 0,7245\,(x - 130,5) = -5,95 + 0,7245x$$

$$k_{y.x} = k_y - \frac{k_{xy}^2}{k_x} = 5735 - \frac{6168^2}{8513} = 1266$$

$$\hat{\sigma}_\varepsilon^2 = s_e^2 = \frac{k_{y.x}}{n-2} = \frac{1266}{12-2} = 126,6 \; ; \quad s_e = 11,25$$

$$s_y^2 = \frac{s_e^2}{n} = \frac{126,6}{12} = 10,55$$

$$s_b^2 = \frac{s_e^2}{k_x} = \frac{126,6}{8513} = .014871 \; ; \quad s_b = .122$$

$t\,(n-2=10) = 1,81$
$\alpha = 2P = .10$

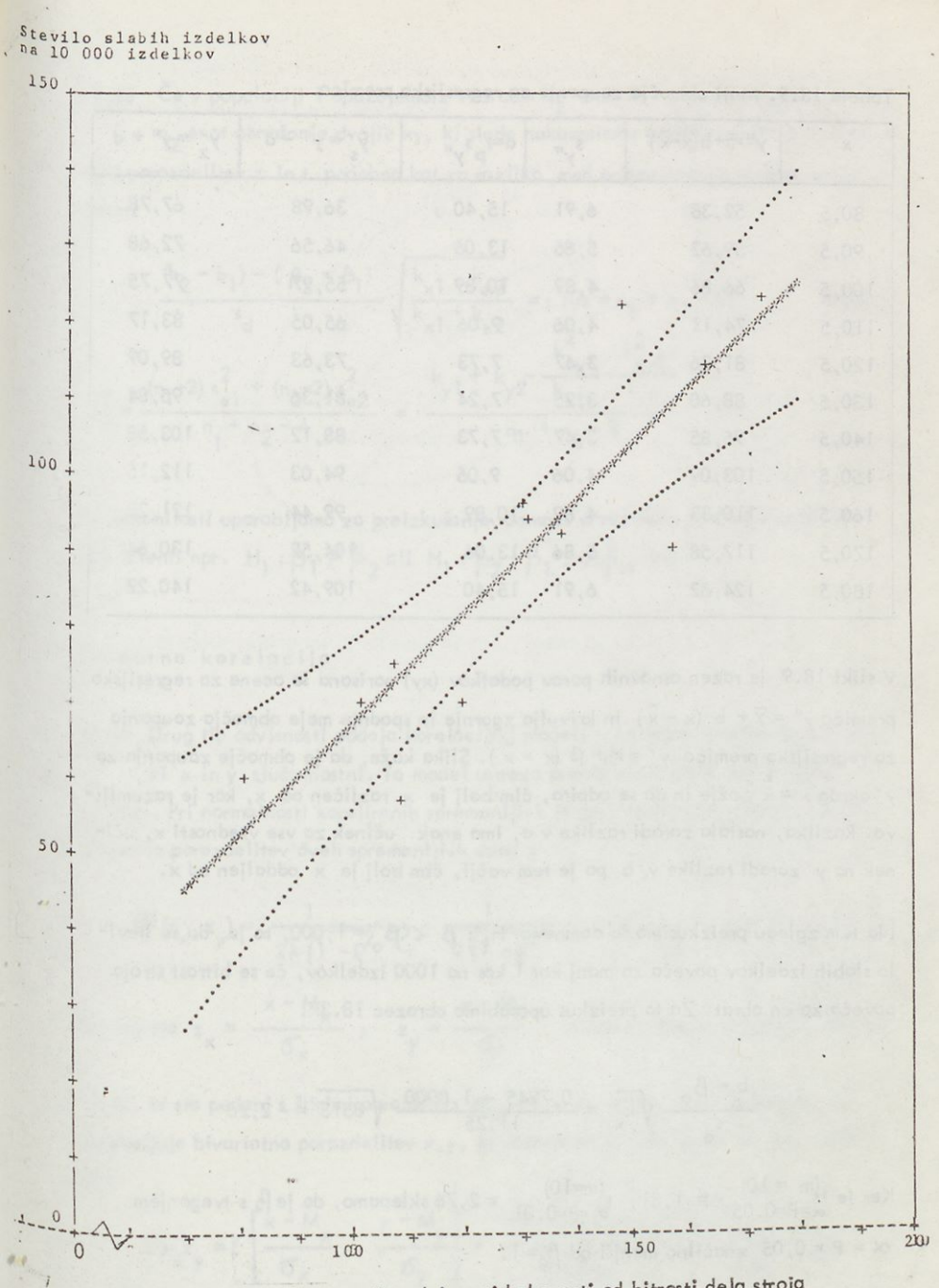
razmik zaupanja za β z $\alpha = 2P = 0,10$ tveganjem

$$b - t_P^{(m)} \, s_b < \beta < b + t_P^{(m)} \, s_b$$

$$.7245 - 1,81 \cdot .122 < \beta < .7245 + 1,81 \cdot .122$$

$$.50 < \beta < .95$$

Če ocenimo še meji zaupanja za regresijsko premico $y' = \alpha + \beta\,(x - \bar{x})$ za vrednosti x v razmikih po 10, v razmiku od 80,5 - 180,5 dobimo



Slika 18.9. Linearna regresija odvisnosti kakovosti od hitrosti dela stroja

Tabela 18.9. Meji območja zaupanja za regresijsko premico

x	$y''=\bar{y}+b(x-\bar{x})$	$s_{y''}$	$d=t_p s_{y''}$	$y'_s = y'' - d$	$y'_z = y'' + d$
80,5	52,38	6,91	15,40	36,98	67,78
90,5	59,62	5,86	13,06	46,56	72,68
100,5	66,86	4,89	10,89	55,97	77,75
110,5	74,11	4,06	9,06	65,05	83,17
120,5	81,36	3,47	7,73	73,63	89,09
130,5	88,60	3,25	7,24	81,36	95,84
140,5	95,85	3,47	7,73	88,12	103,58
150,5	103,09	4,06	9,06	94,03	112,15
160,5	110,33	4,89	10,89	99,44	121,22
170,5	117,58	5,86	13,06	104,52	130,64
180,5	124,82	6,91	15,40	109,42	140,22

V sliki 18.9 je razen osnovnih parov podatkov (xy) narisana še ocena za regresijsko premico $y'' = \bar{y} + b\,(x - \bar{x})$ in krivulja zgornje in spodnje meje območja zaupanja za regresijsko premico $y'_x = M_y + \beta\,(x - \bar{x})$. Slika kaže, da je območje zaupanja za y' okrog $x = \bar{x}$ ožje in da se odpira, čim bolj je x različen od $\bar{x}$, kar je razumljivo. Razlika, nastala zaradi razlike v a, ima enak učinek za vse vrednosti x, učinek na y' zaradi razlike v b pa je tem večji, čim bolj je x oddaljen od $\bar{x}$.

Na tem zgledu preizkusimo še domnevo $H_1 : \beta < \beta_o = 1{,}000$, to je, da se število slabih izdelkov poveča za manj kot 1 kos na 1000 izdelkov, če se hitrost stroja poveča za en obrat. Za ta preizkus uporabimo obrazec 18.37.

$$\frac{b - \beta_o}{s_e}\sqrt{k_x} = \frac{0{,}7245 - 1{,}0000}{11{,}25}\sqrt{8513} = -2{,}26$$

Ker je $t^{(m=10)}_{\alpha=P=0.05} = 1{,}81$, $t^{(m=10)}_{\alpha=P=0.01} = 2{,}76$ sklepamo, da je β s tveganjem $\alpha = P = 0{,}05$ značilno manjši od $\beta_o = 1$.

18.23 Če v populaciji 1 opazujemo z vzorcem n_1 enot in v populaciji 2 z vzorcem n_2 enot obnašanje dvojic xy, ki slede nakazanemu modelu, dobimo iz zakonitosti porazdelitev z in t podobno kot za razliko med aritmetičnima sredinama zakonitosti

$$\frac{(b_2 - b_1) - (\beta_2 - \beta_1)}{s_d} \sqrt{\frac{k_{x1} \cdot k_{x2}}{k_{x1} + k_{x2}}} =: t(m = n_1 + n_2 - 4) \qquad (18.40)$$

$$s_d^2 = \frac{(n_1-2)\, s_{e1}^2 + (n_2-2)\, s_{e2}^2}{n_1 + n_2 - 4} = \frac{k_{y1} + k_{y2} - \frac{k_{xy1}^2}{k_{x1}} - \frac{k_{xy2}^2}{k_{x2}}}{n_1 + n_2 - 4}$$

Te zakonitosti uporabljamo za preizkušanje domnev o razlikah med regresijskimi koeficienti npr. $H_1 : \beta_1 \neq \beta_2$ ali $H_1 : \beta_2 - \beta_1 > \Delta\beta_H$ itd.

Linearna korelacija

18.24 Drug tip odvisnosti podaja korelacijski model, v katerem sta obe spremenljivki x in y slučajnostni. Ta model ustreza proučevanju čiste korelacijske odvisnosti. Pri normalnosti koreliranih spremenljivk je gostota relativne frekvence za normalno porazdelitev dveh spremenljivk dana z

$$\varphi(z_x, z_y) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \cdot e^{-\frac{1}{2\sqrt{1-\varrho^2}}(z_x^2 - 2\varrho z_x z_y + z_y^2)} \qquad (18.41)$$

Pri tem sta $z_x = \frac{x - M_x}{\sigma_x}$; $z_y = \frac{y - M_y}{\sigma_y}$ standardizirano normalni spremenljivki, ki sta podani s štirimi parametri: M_x, M_y, σ_x in σ_y. Peti parameter, ki opredeljuje bivariatno porazdelitev x,y, je parameter ϱ. Parameter ϱ je po definiciji

$$E\, z_x z_y = E\left[\frac{x - M_x}{\sigma_x} \cdot \frac{y - M_y}{\sigma_y}\right] = \frac{E[(x - M_x)(y - M_y)]}{\sigma_x \sigma_y} = \frac{C_{xy}}{\sigma_x \sigma_y} = \varrho_{xy} \qquad (18.42)$$

in je korelacijski koeficient za bivariantno porazdelitev. Korelacijski koeficient meri stopnjo linearne odvisnosti med x in y, ki je dana z regresijskima premicama

$$z_y' = \varrho z_x \; ; \qquad y' = M_y + \varrho \frac{\sigma_y}{\sigma_x} \cdot (x - M_x) \tag{18.43}$$

in

$$z_x' = \varrho z_y \; ; \qquad x' = M_x + \varrho \frac{\sigma_x}{\sigma_y} (y - M_y) \tag{18.44}$$

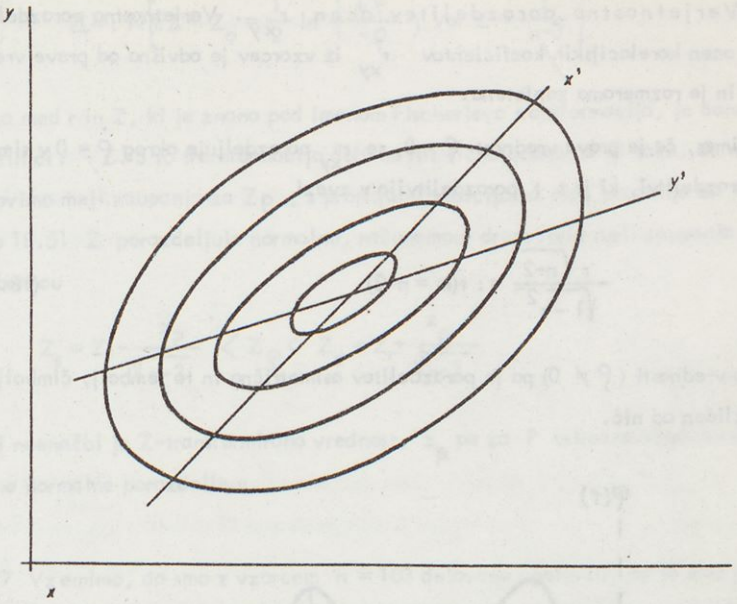
V primeru čiste korelacijske odvisnosti imamo torej dve regresijski premici. Prva kaže linearno odvisnost y od x, druga pa linearno odvisnost x on y. Prvi regresijski koeficient

$$\beta_1 = \varrho \frac{\sigma_y}{\sigma_x} = \frac{C_{xy}}{\sigma_x^2} \tag{18.45}$$

pove, za koliko se v poprečju spremeni y, če se poveča x za enoto, drugi regresijski koeficient

$$\beta_2 = \varrho \frac{\sigma_x}{\sigma_y} = \frac{C_{xy}}{\sigma_y^2} \tag{18.46}$$

pa pove, za koliko se spremeni v poprečju x, če se y poveča za enoto. Slika 18.10 nakazuje krivulje enake verjetnosti za bivariatno porazdelitev dveh odvisnih spremenljivk. Vrednosti korelacijskega koeficienta ϱ leže med -1 in +1. Ta ima drugače iste lastnosti, kot korelacijski koeficient, ki ga računamo v primeru, če predstavlja populacijo N dvojic (xy).



Slika 18.10. Linearni korelacijski model

18.25 Ocene iz vzorcev iz normalne porazdelitve dveh koreliranih spremenljivk. Če iz populacije normalno porazdeljenih koreliranih spremenljivk xy slučajnostno izberemo n enot in zanje merimo vrednosti (xy), ocenimo vse parametre linearne regresije in korelacije za ta primer po istih obrazcih kot smo jih dobili za populacijo N enot

$$y'' = \bar{y} + b_1 (x - \bar{x}) \quad ; \quad b_1 = \frac{c_{xy}}{s_x^2} = \frac{k_{xy}}{k_x} \tag{18.47}$$

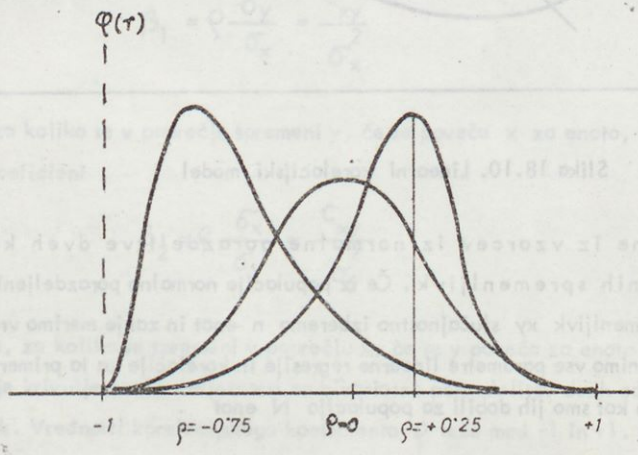
$$x'' = \bar{x} + b_2 (y - \bar{y}) \quad ; \quad b_2 = \frac{c_{xy}}{s_y^2} = \frac{k_{xy}}{k_y} \tag{18.48}$$

18.26 Verjetnostna porazdelitev ocen r_{xy}. Verjetnostna porazdelitev ocen korelacijskih koeficientov r_{xy} iz vzorcev je odvisna od prave vrednosti ϱ in je razmeroma zapletena.

Le za primer, če je prava vrednost $\varrho = 0$, se r_{xy} porazdeljuje okrog $\varrho = 0$ v simetrični porazdelitvi, ki je s t porazdelitvijo v zvezi

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} =: t(m = n-2) \tag{18.49}$$

Za druge vrednosti ($\varrho \neq 0$) pa je porazdelitev asimetrična in to tembolj, čimbolj je ϱ različen od nič.



Slika 18.11. Porazdelitev ocen r_{xy} pri različnih ϱ.

Kot se v primeru, če je $\varrho = 0$, določen izraz, ki vsebuje r, porazdeljuje v t-porazdelitvi, se ne glede na to, kolikšen je ϱ, porazdeljuje transformirana vrednost iz r

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r} \tag{18.50}$$

približno v normalni porazdelitvi

$$Z_r =: N\left[EZ = Z_\rho = \frac{1}{2}\ln\frac{1+\rho}{1-\rho};\ \mathrm{Var}\,Z = \frac{1}{n-3}\right] \qquad (18.51)$$

Zveza med r in Z, ki je znana pod imenom Fischerjeva transformacija, je dana v razpredelnici r - Z. S to transformacijo dobimo razmično oceno za ρ tako, da najprej ugotovimo meji zaupanja za Z_ρ, s protitransformacijo pa meji zaupanja za ρ. Ker se po 18.51 Z porazdeljuje normalno, računamo v drostranski meji zaupanja Z po obrazcu

$$Z_s = Z_r - \frac{z_p}{\sqrt{n-3}} < Z_\rho < Z_z = Z_r + \frac{z_p}{\sqrt{n-3}} \qquad (18.52)$$

V tej neenačbi je Z-transformirana vrednost, z_p pa za P ustrezna vrednost standardizirane normalne porazdelitve.

18.27 Vzemimo, da smo z vzorcem $n = 103$ delavcev ugotovili, da je med porabo predmeta za široko porabo A in predmeta B korelacijski koeficient $r = 0{,}50$. Zanima nas, v kolikšnem razmiku leži s tveganjem $\alpha = 2P = 0{,}05$ prava vrednost ρ. Iz tabele r-Z Fisherjeve transformacije dobimo, da je $Z(r = 0.50) = 0{,}55$. $\alpha = 2P = 0{,}05$ ustrezen $Z_p = 1{,}96$. Po neenačbi 18.52 dobimo:

$$Z_s = 0{,}55 - \frac{1{,}96}{\sqrt{103-3}} < Z_\rho < Z_z = 0{,}55 + \frac{1{,}96}{\sqrt{103-3}}$$

$$Z_s = 0{,}354 < Z_\rho < Z_z = 0{,}746$$

S protitransformacijo dobimo iz tabele r-Z, da je s tveganjem $\alpha = 2P = 0{,}05$.

$$\rho_{Zs} = 0{,}34 < \rho < \rho_z = 0{,}63$$

Preskušanje domnev o korelacijskih koeficientih

18.28 Problemov v zvezi s preskušanjem domnev o korelacijskih koeficientih je več.

Pogosto z r preskušamo domnevo, da obstaja linearna odvisnost med x in y.

Ker je v tem primeru H_o : $\rho = 0$; H_1 : $\rho \neq 0$, preskušamo značilnost odvisnosti z obrazcem 18.49.

18.29 Preskušanje domneve $H_o: \rho = 0; H_1: \rho \neq 0$. Vzemimo, da preskušamo značilnost odvisnosti med kakovostjo dela, ki jo merimo z odstotkom prvovrstnih izdelkov in delovnim stažem. S preskusnim vzorcem n = 50 delavcev smo dobili r = 0,45. Domnevo H_o : $\rho = 0$ preskusimo z izrazom

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{+\,0{,}45\sqrt{50-2}}{\sqrt{1-0{,}45^2}} = +\,3{,}49$$

Ker ne predpostavljamo niti, da se s stažem kakovost dela slabša niti da se boljša, je preskus dvostranski. Za $m = n - 2 = 50 - 2 = 48$ je za $\alpha = 2P = 0{,}01$ $t^{(m=48)}_{P=0{,}005} = 2{,}68$, za $\alpha = 2P = 0{,}001$ pa $t^{(m=48)}_{P=0{,}0005} = 3{,}51$. Iz tega sklepamo, da je med kakovostjo dela in stažem značilna odvisnost s tveganjem $\alpha = 0{,}01$.

18.30 Preskušanje domnev $H_o: \rho = \rho_H$; $H_1: \rho \neq \rho_H$ ali enostranskih domnev $H_o: \rho \leq \rho_H$; $H_1: \rho > \rho_H$. Značilnost razlik ocene korelacijskega koeficienta r iz vzorca od hipotetične vrednosti ρ_H preskušamo s Fisherjevo transformacijo prek izraza

$$(Z - Z_H)\sqrt{n-3} = z \qquad (18.53)$$

ki ga dobimo s standardizacijo Z. Če velja ničelna domneva, se z porazdeljuje standardizirano normalno.

18.31 Z vzorcem n = 30 potrošnikov skušamo potrditi domnevo, da je korelacijski koeficient med porabo izdelka A in porabo izdelka B večji kot $\rho_H = 0{,}50$. Za ta primer velja: $H_o: \rho \leq \rho_H = 0.50$, $H_1: \rho > \rho_H = 0.50$. Iz vzorca dobimo r = 0,67. Če transformiramo $\rho_H = 0{,}50$ in r = 0,67, dobimo $Z_{\rho_H} = 0{,}55$, $Z_r = 0{,}81$. Preskusni izraz 18.53 je za naš zgled

$$z = (0{,}81 - 0{,}55)\sqrt{30-3} = +\,1{,}35$$

Ker je izračunani $z = 1,35$ manjši kot $z_{\alpha=P=0.05} = 1,645$, sklepamo, da vzorčni korelacijski koeficient ni značilno večji od hipotetičnega $\rho_H = 0,50$.

18.32 Preskušanje domnev $H_o: \rho_1 = \rho_2$; $H_1: \rho_1 \neq \rho_2$. Ker se Z porazdeljuje normalno, moremo za preskušanje gornje domneve sestaviti izraz, ki ustreza razliki dveh normalno porazdeljenih spremenljivk

$$\frac{Z_2 - Z_1}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = z \qquad (18.54)$$

pri čemer je Z_1 r_1 ustrezen Z iz vzorca iz populacije 1 z n_1 enotami, Z_2 pa r_2 ustrezen Z iz vzorca iz populacije 2 z n_2 enotami.

18.33 Preskušamo domnevo, da je za dve skupini prebivalstva stopnja odvisnosti izdatkov za tobak in pijačo od dohodkov različna. Za prvo skupino smo z vzorcem z $n_1 = 30$ gospodinjstev dobili $r_1 = 0,48$, za drugo pa z vzorcem z $n_2 = 50$ gospodinjstev $r_2 = 0,69$. $Z(r_1 = 0,48) = 0,52$, $Z(r_2 = 0,69) = 0,85$. Če te podatke vstavimo v obrazec 18.54, dobimo

$$z = \frac{0,85 - 0,52}{\sqrt{\frac{1}{30-3} + \frac{1}{50-3}}} = +1,37$$

Ker je izračunani $z = 1,37$ manjši od kritične vrednosti $z = 1,96$ za $\alpha = 0,05$ sklepamo, da razlika ni značilna.

18.34 Proučitev linearne odvisnosti iz grupiranih podatkov.

Za obsežne populacije ali vzorce proučimo linearno odvisnost iz korelacijske tabele f_{xy}. Običajno izvedemo računanje prek pomožnih znakov $u = (x - x_o)/i_x$ in $v = (y - y_o)/i_y$. Shema za računanje pokazovalcev linearne regresije in korela-

cije iz korelacijske tabele f_{uv} je takole:

a) Shema za izračunanje pomožnih količin n, U, V, $\sum u^2$, $\sum uv$, $\sum v^2$ (v pravokotnikih so zaznamovane tabele, vrste in stolpci, pod njimi in ob njih pa sumarne količine)

y \ x	f_{uv}	f_v	v	vf_v	v^2f_v	$U_v=\sum_u f_{uv}u$	vU_v
	f_u	n		$\sum vf_v = V$	$\sum v^2f_v = Sv^2$	$\sum U_v = U$	$\sum vU_v = Suv$

(18.55)

u	
uf_u	$\sum uf_u = U$
u^2f_u	$\sum u^2f_u = Su^2$
$V_u=\sum_v f_{uv}v$	$\sum V_u = V$
uV_u	$\sum uV_u = Suv$

b) Shema za računanje pokazovalcev linearne regresije in korelacije iz pomožnih količin

	i_x	i_y			
n	U	V			
	$i_x\frac{U}{n}$	$i_y\frac{V}{n}$	Su^2	Suv	Sv^2 (18.56)
			$-U^2/n$	$-UV/n$	$-V^2/n$
	$+x_o$	$+y_o$	$k_u=Su^2-U^2/n$	$k_{uv}=Suv-UV/n$	$k_v=Sv^2-V^2/n$
$\bar{x}=x_o+i_x\frac{U}{n}$		$\bar{y}=y_o+i\frac{V}{n}$	$s_x^2=\frac{i^2k_u}{n-1}$	$c_{xy}=\frac{i_x i_y k_{uv}}{n-1}$	$s_y^2=\frac{i^2k_v}{n-1}$

$$b_1=\frac{i_y}{i_x}\cdot\frac{k_{uv}}{k_u}\;;\quad r_{xy}=\frac{k_{uv}}{\sqrt{k_u\cdot k_v}}\;;\quad b_2=\frac{i_x}{i_y}\cdot\frac{k_{uv}}{k_v}$$

$$y''=\bar{y}+b_1(x-\bar{x})\qquad x''=\bar{x}+b_2(y-\bar{y})$$

18.35 Za zgled vzemimo odvisnost med poprečnim neto osebnim dohodkom zaposlenih v oktobru 1963 in poprečnim neto osebnim dohodkom direktorjev v letu 1963 v obrtnih podjetjih v SFRJ. Odvisnost ocenjujemo iz podatkov za n = 205 slučajnostno izbranih obrtnih podjetij, iz ankete o neto osebnih dohodkih v letu 1963.

V tabelah 18.10 in 18.11 je po shemah 18.55 in 18.56 sistematično nakazano obračunanje linearne regresije in korelacije.

Predhodna analiza grupnih poprečij podatkov v odstavku 18.13 je pokazala, da moremo odvisnost med x in y resnično šteti za linearno.

V tabeli 18.10 je nujno izračunati ali stolpca 7, 8 (U_v, vU_v) ali pa vrsti 7, 8 (V_u, uV_u). Če pa računamo oboje, je to kontrola za U, V, predvsem pa kontrola za razmeroma zapleten izraz $\sum uv$.

V sicer jasnem računu je potrebno mogoče nakazati izračun količine $\sum uv$. To je

$$\sum uv = \sum_v \sum_u f_{uv}\, uv = \sum_v (\sum_u f_{uv} u)v = \sum_v U_v v \; ; \quad U_v = \sum_u f_{uv} u$$

ali

$$\sum uv = \sum_u \sum_v f_{uv}\, vu = \sum_u (\sum_v f_{uv} v)\, u = \sum_u V_u u \; ; \quad V_u = \sum_v f_{uv} v$$

npr. $U_1 = 3.(-3) + 18.(-2) + 16.(-1) + 3.0 + 1.(+1) = -60$

Iz dobljenih pomožnih količin n, U, V, $\sum u^2$, $\sum uv$ in $\sum v^2$ dobimo ocene pokazateljev linearne regresije.

t-preskus

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.6540\sqrt{205-2}}{\sqrt{0,5722}} = 12,317$$

pokaže, da je linearna odvisnost visoko značilna ($t_{\alpha=2P=0.001}^{(m=n-2=203)} = 3,32$)

Tabela 18.10 Izračunanje pomožnih količin za oceno pokazovalcev linearne odvisnosti med poprečnim neto osebnim dohodkom vseh zaposlenih v oktobru 1963 (x) in poprečnim neto osebnim dohodkom v letu 1963 za direktorje(y) za vzorec 205 obrtnih podjetij v SFRJ. (Vir: SZS SB 369, Beograd, 1965)

(v tisoč dinarjev)

y \ x	1	- 10	10- 20	20- 30	30- 40	40- 50	50- 60	f_v	v	$v.f_v$	$v^2 f_v$	U_v	$v.U_v$
1					2			3	4	5	6	7	8
20 - 40		3	18	16	3	1	-	41	-2	-82	164	-60	120
40 - 60		-	19	66	17	1	-	103	-1	-103	103	-103	103
60 - 80		-	-	12	20	10	7	49	0	0	0	12	0
80 -100	2	-	-	-	4	2	3	9	+1	+ 9	9	8	8
100- 120		-	-	-	1	1	1	3	+2	+ 6	12	3	6
f_u	3	3	37	94	45	15	11	205=n		-170	288	-140	237
u	4	-3	-2	-1	0	+1	+2			V	$\sum v^2$	U	$\sum uv$
$u.f_u$	5	-9	-74	-94	0	15	22	-140 = U					
$u^2.f_u$	6	27	148	94	0	15	44	328 = $\sum u^2$					
V_u	7	-6	-55	-98	-17	1	5	-170 = V					
$u.V_u$	8	+18	110	98	0	1	10	237 = $\sum uv$					

Tabela 18.11 Izračunanje pokazovalcev linearne odvisnosti iz pomožnih količin

	$i_x = 10$	$i_y = 20$			
$n = 205$	$U = -140$	$V = -170$	$\sum u^2 = 328$	$\sum uv = 237$	$\sum v^2 = 288$
	$i_x U/n = -6,829$	$i_y V/n = -16,585$	$-U^2/n = -95,610$	$-UV/n = -116,098$	$-V^2/n = -140,976$
	$+ x_o = 35$	$+ y_o = 70$	$k_u = 232,390$	$k_{uv} = 120,902$	$k_v = 147,024$
	$\bar{x} = 28,171$	$\bar{y} = 53,415$		$r^2_{xy} = 0,4278$	
			$b_1 = 1,0406$	$r_{xy} = 0,6540$	$b_2 = 0,41124$
				$1-r^2_{xy} = 0,5722$	

$$y' = 53,415 + 1,0406\,(x - 28,171) = 24,10 + 1,0406\,x$$

$$x' = 28,171 + 0,4112\,(y - 53,414) = 6,207 + 0,4112\,y$$

Korelacijska analiza je nakazala dosti veliko odvisnost med skupnimi poprečnimi dohodki in poprečnimi dohodki direktorjev. $r^2_{xy} = 0,43$, tj. 43% od skupne variabilnosti dohodkov direktorjev je pojasnjeno s skupnimi neto poprečnimi dohodki. Ker je r_{xy} značilen, moremo zanj poiskati prek Fisherjevega Z meji zaupanja s tveganjem $\alpha = 2P = 0,05$. Ker je $Z\,(r = 0,654) = 0,78$, je po obrazcu 18.52

$$Z_s = 0,78 - \frac{1,96}{\sqrt{205-3}} < Z_\varrho < Z_z = 0,78 + \frac{1,96}{\sqrt{205-3}}$$

$$0,642 < Z_\varrho < 0,918$$

$$0,57 < \varrho < 0,72$$

Krivuljčne odvisnosti

18.36 Že uvodoma smo nakazali, da odvisnosti v socialno-ekonomskih pojavih niso vedno linearne, ampak imamo, posebno če proučujemo pojave na širšem variacijskem razmiku, opravka s krivuljčno regresijo in korelacijo. V primeru, če proučujemo odvisnost določenega izdatka v družinskem proračunu, se izdatek pri spreminjanju dohodka ne spreminja linearno, temveč progresivno ali degresivno ipd.

V teh primerih je regresijska črta krivulja in pridejo v poštev za proučevanje zakonitosti krivuljčni in ne linearni modeli. Najsplošnejši model korelacijske odvisnosti je podan simbolično z

$$h\,(y;\ x;\ \varepsilon) = 0 \qquad (18.57)$$

in kaže na zakonitost odvisnosti med x in y. V eksplicitni obliki je regresijska zveza dana z modelom

$$y = f(x;\ \varepsilon) \qquad (18.58)$$

Ta zveza nakazuje, kako je y odvisen od x in posamičnih vplivov ε.

Regresijska funkcija vsebuje običajno več parametrov, za katere je treba vrednosti šele določiti.

Splošen model krivuljčne odvisnosti moremo pisati v nekoliko modificirani obliki

$$y = f(x;\ a,\ b,\ c\ldots) + \varepsilon \tag{18.59}$$

V njem smo predpostavili, da sta komponenta, ki je odvisna od proučevanega neodvisnega znaka, tj. $f(x;\ a,\ b,\ c\ldots)$ in rezultat posamičnih vplivov ε, vezana aditivno.

Ta predpostavka zelo olajša nadaljnje proučevanje regresijskih odvisnosti. Že takoj je jasno, da

$$y' = f(x;\ a,b,c\ldots) \tag{18.60}$$

pomeni regresijsko krivuljo odvisnosti y od x, ki kaže zvezo med x in y, če ne bi bilo posamičnih vplivov. Podobno merimo z

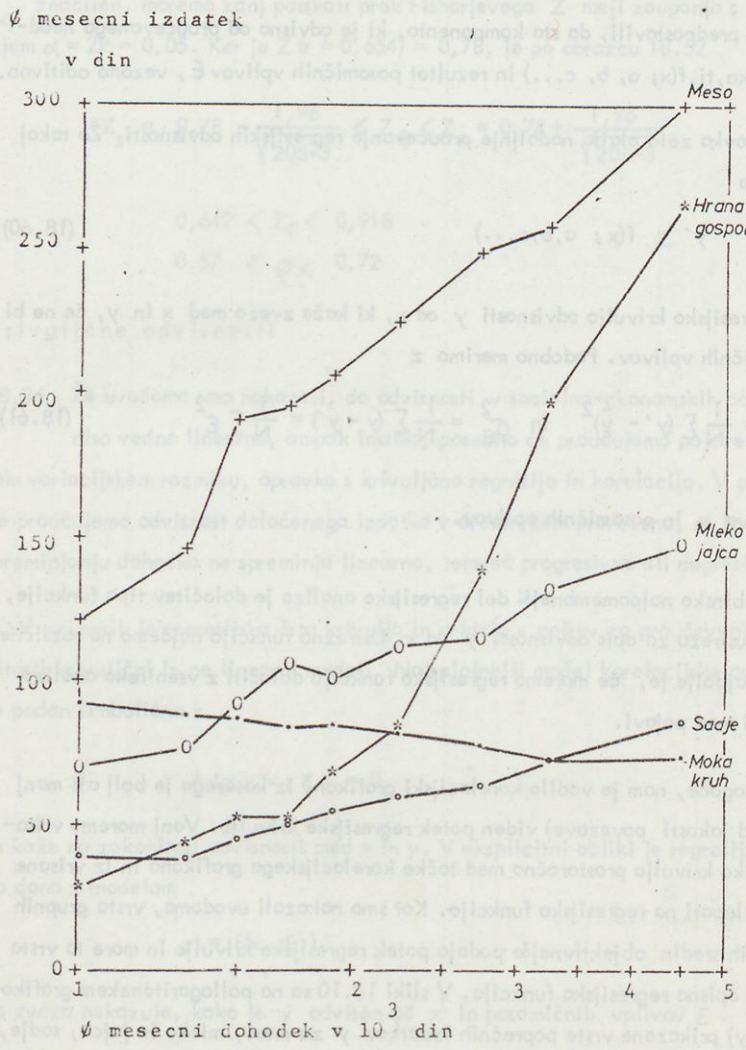
$$\sigma_{y'}^2 = \frac{1}{N}\sum (y' - \bar{y})^2 \quad \text{in} \quad \sigma_{\varepsilon}^2 = \frac{1}{N}\sum (y - y')^2 = \frac{1}{N}\sum \varepsilon^2 \tag{18.61}$$

jakost vpliva x in posamičnih vplivov.

18.37 Vsebinsko najpomembnejši del regresijske analize je določitev tipa funkcije, ki ustreza za opis odvisnosti y od x. Ustrezno funkcijo najdemo na različne načine. Najbolje je, če moremo regresijsko funkcijo določiti z vsebinsko analizo zakonitosti med pojavi.

Če to ni mogoče, nam je vodilo korelacijski grafikon, iz katerega je bolj ali manj (odvisno od jakosti povezave) viden potek regresijske krivulje. Vanj moremo vrisati regresijsko krivuljo prostoročno med točke korelacijskega grafikona in iz vrisane krivulje sklepati na regresijsko funkcijo. Kot smo nakazali uvodoma, vrsta grupnih aritmetičnih sredin objektivnejše podaja potek regresijske krivulje in more ta vrsta služiti kot opisna regresijska funkcija. V sliki 18.10 so na pollogaritemskem grafikonu (log x,y) prikazane vrste poprečnih izdatkov y za meso, mleko in jajca, sadje, moko in izdelke ter hrano izven gospodinjstva v odvisnosti od poprečnih dohodkov za 4-članska delavska gospodinjstva v letu 1970. Iz grafikona vidimo, da so v grafikonu s transformiranimi skalami črte poprečij za meso, mleko in jajca, sadje in

moko ter izdelke potekajo v linearno. Iz tega povzemamo, da je za te skupine prehrane ustrezna regresijska funkcija logaritemska funkcija $y' = a + b \log x$.



Slika 18.10. Popračni mesečni izdatki štiričlanskih delavskih gospodinjstev v letu 1970 v odvisnosti od mesečnih dohodkov. (Vir: SB, št. 685).

18.38 Poseben tip regresijskega modela. Iz razlogov, ki jih narekuje analitično določanje parametrov regresijskih funkcij je najpreprosteje, če moremo predpostaviti, da so v regresijski funkciji parametri v aditivni zvezi. Splošen model krivuljčne regresije, ki je vezan na to omejitev, je

$$g(y) = a_o f_o(x) + a_1 f_1(x) + \ldots\ldots + a_p f_p(x) + \varepsilon \qquad (18.62)$$

$$g(y') = a_o f_o(x) + a_1 f_1(x) + \ldots\ldots + a_p f_p(x) \qquad (18.63)$$

Za ta model moremo določiti parametre a_o, a_1 ... po standardni metodi najmanjših kvadratov, ne za y, temveč za $Y' = g(y')$, ki je funkcija neodvisnega znaka y.

Če na tem modelu uporabimo metodo najmanjših kvadratov, štejemo kot regresijsko funkcijo tisto, za katero velja

$$\sum (Y - Y')^2 = \sum \left[g(y) - g(y')\right]^2 = \min \qquad (18.64)$$

Tako dobimo dalje:

$$\sum_1^N \left[g(y) - g(y')\right]^2 = \sum_{i=1}^N \left[g(y_i) - \sum_{k=o}^p a_k f_k(x_i)\right]^2 = F(a_o, a_1 \ldots a_p) = \min \qquad (18.65)$$

Če funkcijo $F(a_o, a_1 \ldots a_p)$ odvajamo po parametrih (a_i) dobimo potrebne pogoje za ekstrem

$$\frac{\partial F}{\partial a_j} = -2 \sum_i \left[g - \sum_k a_k f_k\right] f_j = 0 \qquad j = 0,1,2..p \qquad (18.66)$$

Ker so drugi parcialni odvodi po a_i vsi pozitivni

$$\frac{\partial^2 F}{\partial a_i^2} = + \sum_1^N f_i^2$$

spoznamo, da je dobljeni ekstrem minimum. Gornji pogoji dajo naslednji sistem $p+1$ linearnih normalnih enačb

$$\sum_{1}^{N} gf_j = \sum_{k=o}^{p} a_k \sum_{1}^{N} f_k f_j \qquad (j = 0,1,2\ldots p) \qquad (h = 0,1,2..p) \qquad (18.67)$$

ki jih moremo pisati v razviti obliki:

$$\begin{aligned}
\sum gf_o &= a_o \sum f_o^2 + a_1 \sum f_1 f_o + a_2 \sum f_2 f_o + \ldots + a_p \sum f_p f_o \\
\sum gf_1 &= a_o \sum f_o f_1 + a_1 \sum f_1^2 + a_2 \sum f_2 f_1 + \ldots + a_p \sum f_p f_1 \\
\sum gf_2 &= a_o \sum f_o f_2 + a_1 \sum f_1 f_2 + a_2 \sum f_2^2 + \ldots + a_p \sum f_p f_2 \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
\sum gf_p &= a_o \sum f_o f_p + a_1 \sum f_1 f_p + a_2 \sum f_2 f_p + \ldots a_p \sum f_p^2
\end{aligned} \qquad (18.68)$$

Če zaznamujemo vektorja funkcije y in parametrov a

$$(g(y_1), g(y_2) \ldots g(y_N)) = g \qquad (18.69)$$

$$(a_o, a_1 \ldots a_p) = a$$

in matriko funkcij f(x)

$$\begin{pmatrix} f_o(x_1) & f_o(x_2) & \ldots\ldots & f_o(x_N) \\ f_1(x_1) & f_1(x_2) & \ldots\ldots & f_1(x_N) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f_p(x_1) & f_p(x_2) & \ldots\ldots & f_p(x_N) \end{pmatrix} = f \qquad (18.70)$$

moremo sistem normalnih enačb pisati v matrični obliki

$$g f' = a f f' \qquad (18.71)$$

Če postavimo, da je $gf' = G$, $ff' = F$, je

$$G = aF \text{ in } a = GF^{-1} \qquad (18.72)$$

Ker je $g(y') = af$, je dalje:

$$\sigma_\varepsilon^2 = \frac{1}{N}(g - af)(g' - f'a') = \frac{1}{N}\left[gg' - afg' - (gf' - aff')a'\right] = \frac{1}{N}\left[gg' - afg'\right] =$$

$$= \frac{1}{N}\left[gg' - GF^{-1}G'\right]$$

oziroma

$$\sigma_{\varepsilon(g)}^2 = \frac{1}{N}\left[\sum_{i=1}^{N} g_i^2 - \sum_{k=o}^{P} a_k \sum_{i=1}^{N} f_{ki}\, g_i\right] \tag{18.73}$$

Iz tega obrazca dobimo dalje

$$I_{g(y),x}^2 = 1 - \frac{\sigma_{\varepsilon(g)}^2}{\sigma_{g(y)}^2} \tag{18.74}$$

Treba je opozoriti, da veljata determinacijski koeficient $I_{g.x}^2$ in indeks korelacije $I_{g.x}$ za odvisnost med $g(y)$ in x in ne med y in x.

18.39 Regresijski modeli tipa $g(y) = \alpha_o + \beta_1 f(x) + \varepsilon$. Poseben primer gornjega splošnejšega modela je regresijski model

$$g(y) = \alpha_o + \beta_1 f(x) + \varepsilon \tag{18.75}$$

ki ga dobimo iz 18.62, če vstavimo

$$f_o(x) = 1 \; ; \quad f_1(x) = f(x) \; ; \quad f_2(x) = f_3(x) \ldots f_p(x) = 0$$

Če vpeljemo za ta primer transformacijo $g(y) = Y$ in $f(x) = X$, je regresijski model v transformiranih znakih linearni model

$$Y = \alpha_o + \beta_1 X + \varepsilon \tag{18.76}$$

Take vrste krivuljične regresije obravnavamo enostavno tako kot linearne odvisnosti po naslednjih stopnjah:

1. Glede na poznavanje zakonitosti povezanosti med x in y ali iz korelacijskega grafikona ugotovimo ustrezni transformaciji $Y = g(y)$, $X = f(x)$.

2. Vrednosti osnovnih znakov $x_i = x_1 \ldots x_N$ in $y_i = y_1 \ldots y_N$ prevedemo v transformirane znake $X_i = f(x_i)$, $Y_i = g(y_i)$, $i = 1;2 \ldots N$.

3. Iz transformiranih vrednosti znakov X_i in Y_i izračunamo pokazovalce za linearno regresijo in korelacijo za X in Y

$$Y' = M_Y + b_1 (X - M_X) \; ; \; b_1 = \frac{C_{XY}}{\sigma_X^2} \; ; \; \rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} \qquad (18.77)$$

4. V regresijski enačbi med X in Y' nadomestimo transformirane spremenljivke z f(x) in g(y')

$$g(y') = M_Y + b_1 \left[f(x) - M_X \right] \qquad (18.78)$$

in y' iz te enačbe izrazimo eksplicitno

$$y' = h (x; M_X, M_Y, b_1) \qquad (18.79)$$

5. Korelacijski koeficient $\rho_{XY} = \rho_{f(x),g(y)}$ je koeficient linearne korelacije med f(x) in g(y) in ne med x in y. Glede na znaka x in y ima $\rho_{f(x),g(y)}$ značaj mere krivuljične korelacije. Prav tak značaj ima za ta primer tudi $\sigma_\varepsilon^2 = \sigma_{g(y).f(x)}^2$.

18.40 Pri regresijski analizi je vsekakor najpomembnejše in najodgovornejše delo ugotoviti ustrezen tip funkcije. Vsebinsko upravičena regresijska funkcija ima niz prednosti pred drugimi funkcijami, ki se mogoče na videz tudi razmeroma dobro prilegajo dejanskim podatkom. Iz take regresijske funkcije je lažje vsebinsko tolmačenje parametrov regresijske krivulje in regresije nasploh. Zanje je možno z večjo zanesljivostjo ekstrapolirati dobljene zakonitosti tudi izven območja, za katerega imamo dane stvarne podatke.

Pri vsebinski preučitvi ustreznega tipa funkcije si pogosto pomagamo z analizo odvodov funkcije, ki nakazujejo in pojasnjujejo način spremembe neodvisne spremenljivke

pri spreminjanju odvisne spremenljivke. Diferenciali posameznih spremenljivk morejo biti bodisi absolutni dx in dy ali pa relativni oziroma logaritemski dx/x in dy/y. Prvi ponazarjajo absolutne, drugi pa relativne spremembe. Iz gornjih diferencialov moremo sestaviti štiri specifične diferencialne kvociente. dy/dx kaže absolutno spremembo y pri absolutni spremembi x za enoto. $\frac{dy/y}{dx}$ kaže relativno spremembo y pri absolutni spremembi x za enoto. $\frac{dy}{dx/x}$ kaže absolutno spremembo y, če se relativna vrednost x spremeni za enoto, in končno $\frac{dy/y}{dx/x}$ kaže relativno spremembo y pri relativni spremembi x za enoto. Medtem ko je $\frac{dy}{dx}$ diferencialni kvocient, je $\frac{dy/y}{dx/x}$ koeficient prožnosti, ker je kvocient logaritemskih diferencialov.

18.41 Pri preučitvi zakonitosti se bomo omejili na nekatere najpomembnejše načine spreminjanja nakazanih diferencialnih kvocientov. V nadaljnji analizi bomo prikazali, pri katerih funkcijah se posamezni diferencialni kvocienti E obnašajo v naslednjih zakonitostih:

1. $E = b$; konstanten,
2. $E = b/x$; E je obratno sorazmeren z neodvisno spremenljivko x,
3. $E = b/y$; E je obratno sorazmeren z odvisno spremenljivko y.

Pri obeh zadnjih primerih b pomeni vrednost E, če je x ali y enak 1,

4. $E = b\,(1 - \frac{y}{y_\infty})$; E = linearno odvisen od y.
5. $E = b(1-x/x_o)$; E linearno odvisen od x.

V primeru 4 je b vrednost za E pri $y = 0$, E pa z večanjem y linearno pada in se vrednosti $E = 0$ asimptotično približuje z večanjem y in doseže nič pri vrednosti $y = y_\infty$. Analogno je v petem tipu, v katerem je $b = E(x=0)$. Z večanjem x E pada vse do $x = x_o$, ko je $E(x=x_o) = 0$. V tem primeru more za vrednosti $x > x_o$ priti tudi do negativnih vrednosti za E.

V tabeli 18.12 so za posamezne vrste diferencialnih kvocientov E in za posamezne vrste zakonitosti spreminjanja E v odvisnosti od x in y podane ustrezne regresijske

funkcije z nakazanimi transformacijami. Za $E = b$, $E = b/x$ in $E = b/y$ so funkcije dvoparametrske, za $E = b(1-y/y_\infty)$ in $E = b(1-x/x_0)$ pa imajo regresijske funkcije tri parametre. V tabeli so za $E=b$, $E=b/x$ in $E=b/y$ funkcije podane s parametri, ki jih dobimo po metodi najmanjših kvadratov. b je v vseh primerih izračunan kot

$$b = \frac{K_{g(y)\, f(x)}}{K_{f(x)}}$$

iz transformiranih znakov, medtem ko z M, H in G označujemo aritmetične, harmonične in geometrijske sredine. Za $E=b(1-\frac{y}{y_\infty})$ je temeljna modificirana eksponentna funkcija in so za druge funkcije z zakonitostjo tega tipa nakazane transformacije, ki jih prevedejo v modificirano eksponentno. Za $E = b(1-\frac{x}{x_0})$ so vse štiri regresijske funkcije podane v obliki:

$$g(y') = a_0 + a_1 f_1(x) + a_2 f_2(x)$$

Zato so pod vsako eksplicitno regresijsko funkcijo dane ustrezne funkcije $g(y)$, $f_1(x)$ in $f_2(x)$. Z njimi moremo po metodi najmanjših kvadratov izvrednotiti vrednosti parametrov. y_∞ in x_0 v teh enačbah so vrednosti za x in y, pri katerih je $E = 0$. Za vse, razen za funkcije z zakonitostjo $E=b\,(1-\frac{y}{y_\infty})$, moremo oceniti parametre po metodi najmanjših kvadratov, če ustrezno transformiramo osnovne znake.

18.42 Za zgled vzemimo odvisnost proizvodnje od števila zaposlenih za N = 55 predilnic in tkalnic v Sloveniji v letu 1961. Proučitev osnovnih podatkov je pokazala, da je regresija med proizvodnjo in številom zaposlenih opisana z modelom

$$p = a\, z^b \cdot \varepsilon$$

Pri tem pomeni p proizvodnjo, z število zaposlenih; a, b sta parametra regresijske funkcije; ε pa sumaren faktorialni vpliv drugih dejavnikov. Z logaritmiranjem dobimo

$$\log p = \log a + b \log z + \log \varepsilon$$

Tabela 18.12 Regresijske funkcije za različne zakonitosti spreminjanja spremenljivk x in y

	$E = b$	$E = \frac{b}{x}$	$E = \frac{b}{y}$	$E = b\left(1 - \frac{y}{y_\infty}\right)$	$E = b\left(1 - \frac{x}{x_o}\right)$
	$y' = f(x)$	$y' = f(x)$	$y' = f(x)$	$y' = f(x)$	$y' = f(x)$
	$g(y)$ | $h(x)$	$g(y)$ | $h(x)$	$g(y)$ | $h(x)$	$g(y)$ | $h(x)$	$g(y)$ | $h_1(x)$ | $h_2(x)$
$E = \frac{dy}{dx}$	$y' = M_y + b(x - M_x)$	$y' = M_y + b \ln \frac{x}{G_x}$	$y = \sqrt{Q_y + 2b(x - M_x)}$	$y' = y_\infty \left(1 - \frac{y_\infty - y_o}{y_\infty} e^{-\frac{bx}{y_\infty}}\right)$	$y' = y_o - \frac{b}{2} \frac{(x - x_o)^2}{x_o}$
	y | x	y | $\ln x$	y^2 | x	y | $e^{-\frac{bx}{y_\infty}}$	y | x | x^2
$E = \frac{dy/y}{dx}$	$y' = G_y \cdot e^{b(x - M_x)}$	$y' = G_y \left(\frac{x}{G_x}\right)^b$	$y' = M_y + b(x - M_x)$	$y' = \frac{y_\infty}{1 + \frac{y_\infty - y_o}{y_o} e^{-bx}}$	$y' = y_o e^{-\frac{b}{2} \frac{(x - x_o)^2}{x_o}}$
	$\ln y$ | x	$\ln y$ | $\ln x$	y | x	$\frac{1}{y}$ | e^{-bx}	$\ln y$ | x | x^2
$E = \frac{dy}{dx/x}$	$y' = M_y + b \ln \frac{x}{G_x}$	$y' = M_y - b\left(\frac{1}{x} - \frac{1}{H_x}\right)$	$y' = \sqrt{G_y + \frac{b}{2} \ln \frac{x}{G_x}}$	$y' = y_\infty \left(1 - \frac{y_\infty - y_1}{y_\infty} x^{-\frac{b}{y_\infty}}\right)$	$y' = y_o + b\left(\ln \frac{x}{x_o} - \frac{x - x_o}{x_o}\right)$
	y | $\ln x$	y | $\frac{1}{x}$	y^2 | $\ln x$	y | x^{-b/y_∞}	y | $\ln x$ | x
$E = \frac{dy/y}{dx/x}$	$y' = Gy \left(\frac{x}{G_x}\right)^b$	$y' = G_y e^{-b\left(\frac{1}{x} - \frac{1}{H}\right)}$	$y' = M_y + b \ln \frac{x}{G_x}$	$y' = \frac{y_\infty \cdot}{1 + \frac{y_\infty - y_1}{y_1} x^{-b}}$	$y' = y_o \left(\frac{x}{x_o}\right)^b \cdot e^{-b \frac{x - x_o}{x_o}}$
	$\ln y$ | $\ln x$	$\ln y$ | $\frac{1}{x}$	y | $\ln x$	$\frac{1}{y}$ | x^{-b}	$\ln y$ | $\ln x$ | x

Če postavimo $\log p = P$; $\log z = Z$; $\log a = A$ in $b = B$, dobimo regresijsko krivuljo

$$P' = A + BZ$$

Model ima obliko, nakazano v 18.75, in zanj najdemo v tabeli 18.12 regresijsko funkcijo z lastnostjo $E = b = \frac{dy/y}{dx/x}$. Iz podatkov $N = 55$ podjetij dobimo:

$N=55$; $SZ=165,7316$; $SP=149,9542$ $\quad SZ^2 = 509,5476 \quad SZP = 463,5277 \quad SP^2 = 425,0757$

$$\bar{Z} = 3,0133 \quad \bar{P} = 2,7264 \quad \frac{(SZ)^2}{N} = 499,3993 \quad \frac{(SZSP)}{n} = 451,8573 \quad \frac{(SP)^2}{n} = 408,8411$$

$$k_Z \quad 10,1483 \quad k_{ZP} = 11,6704 \quad k_P = 16,2346$$

$$s_Z^2 = 0,18793 \quad c_{ZP} = 0,21612 \quad s_P^2 = 0,30064$$

$$b = \frac{c_{ZP}}{s_Z^2} = \frac{\cdot 21612}{\cdot 18793} = 1,1500$$

$$\log p' = \bar{P} + b\,(\log z - \bar{Z}) = 2,7264 + 1,1500\,(\log z - 3,0133) = 0,2611_{-1} + 1,15 \log z$$

Z obratno transformacijo dobimo regresijsko funkcijo

$$p' = 0,1824\, z^{1,15}$$

Glede na pomen parametra $b = 1,15$ sklepamo, da se proizvodnja poprečno relativno spremeni za $b = 1,15$, če se število zaposlenih spremeni relativno za enoto npr. če se število zaposlenih spremeni za 1%, se proizvodnja poprečno spremeni za 1,15%.

18.43 Večparametrski regresijski model. Poseben primer splošnega modela 18.62 je

$$y = \sum_k a_k f_k(x) + \varepsilon \qquad (18.80)$$

V njem je odvisna spremenljivka netransformirana in je regresijska funkcija

$$y' = \sum_k a_k f_k(x) \qquad (18.81)$$

dana eksplicitno.

Če ni vsebinskega razloga za določene funkcije f(x), v gornji model vstavimo pogosto cele potence od x, $f_k(x) = x^k$. Regresijska funkcija je polinom celih potenc

$$y' = \sum_{k=o}^{p} a_k x^k = a_o + a_1 x + a_2 x^2 + \ldots + a_p x^p \qquad (18.82)$$

Sistem normalnih enačb pa je

$$\sum_{i=1}^{N} y_i x_i^j = \sum_{k=o}^{p} a_k \sum_{i=1}^{N} x_i^{k+j} \qquad \begin{matrix} j = 0,1,2 \ldots p \\ k = o,1,2 \ldots p \end{matrix} \qquad (18.83)$$

Vsote produktov $\sum f_k f_j$ so v tem posebnem primeru enake $\sum f_k f_j = \sum x^{k+j}$ vsotam potenc znaka x za vse enote populacije.

18.44 Pogosto so vrednosti x dane v aritmetičnem zaporedju $x_{i+1} = x_i + \Delta$. V tem primeru uvedemo namesto x ustrezno transformacijo

$$u = \frac{x - x_1}{\Delta} - \frac{N-1}{2} \qquad \text{če je N = liho število} \qquad (18.84)$$

$$u = \frac{2(x-x_1)}{\Delta} - (N-1) \qquad \text{če je N = sodo število}$$

Za transformirani u je vsota lihih potenc nič

$$\sum u^{2s+1} = 0 \qquad s = 0, 1, 2 \ldots p-1 \qquad (18.85)$$

medtem ko je vsota sodih potenc

$$\sum u^{2s} \qquad s = 0, 1, 2 \ldots P \qquad (18.86)$$

odvisna samo od N in s.

Sistem normalnih enačb razpade v dva sistema linearnih enačb. Za polinom četrte stopnje sta sistema enačb npr. takale:

$$\begin{aligned}
\sum y &= a_o N + a_2 \sum u^2 + a_4 \sum u^4 \\
\sum yu &= a_1 \sum u^2 + a_3 \sum u^4 \\
\sum yu^2 &= a_o \sum u^2 + a_2 \sum u^4 + a_4 \sum u^6 \qquad (18.87) \\
\sum yu^3 &= a_1 \sum u^4 + a_3 \sum u^6 \\
\sum yu^4 &= a_o \sum u^4 + a_2 \sum u^6 + a_4 \sum u^8
\end{aligned}$$

$$\begin{aligned}
\sum y &= a_o N + a_2 \sum u^2 + a_4 \sum u^4 \\
\sum yu^2 &= a_o \sum u^2 + a_2 \sum u^4 + a_4 \sum u^6 \\
\sum yu^4 &= a_o \sum u^4 + a_2 \sum u^6 + a_4 \sum u^8 \qquad (18.88)
\end{aligned}$$

$$\begin{aligned}
\sum yu &= a_1 \sum u^2 + a_3 \sum u^4 \\
\sum yu^3 &= a_1 \sum u^4 + a_3 \sum u^6
\end{aligned}$$

Prvi velja za parametre a_o, a_2, a_4 in drugi pa za a_1 in a_3.

Iz obrazca 18.73 pa dobimo za obravnavani tip regresijske funkcije

$$\sigma_\varepsilon^2 = \frac{1}{N}\left[\sum_{i=1}^{N} y_i^2 - \sum_{k=o}^{p} a_k \sum_{i=1}^{N} y_i x_i^k\right] \qquad (18.89)$$

18.45 Če predpostavljamo regresijski model 18.62

$$g(y) = \sum_k a_k f_k(x) + \varepsilon \quad ; \quad \varepsilon =: N(0, \sigma_\varepsilon^2) \qquad (18.90)$$

in iz populacije na slučajnostni način izberemo n vrednosti y pri n vrednostih x, so po gornjih obrazcih izračunani parametri a_u le ocene pravih vrednosti .

Nepristransko oceno za σ_ε^2 pa izračunamo po obrazcu

$$s_e^2 = \frac{1}{n-p-1}\left[\sum_i g(y_i)^2 - \sum_k a_k \sum_i f_k\, g(y_i)\right] \qquad (18.91)$$

m=n-p-1 je število stopinj prostosti, ker imamo v modelu p + 1 parameter.

18.46 Iz korelacijskega grafikona med odstotkom travnikov in pašnikov za okraje v SRS v letu 1952 sklepamo, da moremo v tem zgledu za regresijsko krivuljo vzeti del parabole druge stopnje tipa $y = a + bx + cx^2$.

Zanjo velja sistem normalnih enačb 18.83. Zato moramo iz osnovnih podatkov izračunati izraze: $\sum y, \sum xy, \sum x^2y, \sum x, \sum x^2, \sum x^3, \sum x^4$.

Tabela 18.13. Odstotki travnikov in pašnikov za okraje SRS v letu 1952

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
x	22	3	9	5	6	7	16	28	18	18	15	15	21	8	10	18	5	1	7	8	5	9
y	8	28	20	13	15	7	9	3	11	6	6	8	3	9	19	10	16	37	10	9	20	8

Za izračunanje parametrov potrebne količine so:

$\sum x = 254 \quad \sum y = 275 \quad \sum xy = 2324 \quad \sum x^2y = 29956 \quad \sum x^2 = 3996$

$\sum x^3 = 74990 \quad \sum x^4 = 1564476 \quad \sum y^2 = 4859$

Če te izraze vnesemo v sistem normalnih enačb 18.83 dobimo:

$$\begin{aligned} 22a + 254b + 3996c &= 275 \\ 254a + 3996b + 74990c &= 2324 \\ 3996a + 74990b + 1564476c &= 29956 \end{aligned}$$

Iz tega sistema normalnih enačb dobimo parametre:

$a = +29{,}0468 \; ; \quad b = -2{,}3066 \; ; \quad c = +0{,}0555$

Regresijska krivulja pa je

$$y' = +29,0468 - 2,3066x + 0,0555x^2$$

Standardna napaka ocene je po obrazcu 18.89 enaka:

$$\sigma_\varepsilon^2 = \frac{1}{22}\left[(4859-(+29,0468).275-(-2,3066).2324-(+0,0555).29956\right] = 25,8686$$

$$\sigma_\varepsilon = \sqrt{25,8686} = 5,086$$

Ker je

$$\sigma_y^2 = \frac{1}{N}\left[\sum y^2 - \frac{(\sum y)^2}{N}\right] = \frac{1}{22}\left(4859 - \frac{275^2}{22}\right) = 64,6136$$

dobimo, da je indeks korelacije I_{yx} po obrazcu 18.74 enak

$$I_{yx} = \sqrt{1 - \frac{25,8686}{64,6136}} = \sqrt{0,5996} = 0,774$$

Če primerjamo dobljeni indeks korelacije $I_{yx} = 0,774$ s korelacijskima koeficientoma, katera smo dobili z logaritemsko transformacijo $\varrho_{\log x, \log y} = 0,778$ ali s transformacijo s kvadratnimi koreni $\varrho_{\sqrt{x}, \sqrt{y}} = 0,809$, spoznamo, da smo s parabolo druge stopnje tipa $y' = a+bx+cx^2$ dobili manjši indeks korelacije, čeprav ima ta funkcija tri, transformirani funkciji pa samo po dva parametra. Iz tega sklepamo, da transformacija, posebno pa transformacija s kvadratnimi korenî, da boljše rezultate.

KORELACIJSKO RAZMERJE

18.47 Pri obravnavanju regresijskih krivulj smo videli, da moremo kot regresijsko črto vzeti tudi vrsto grupnih sredin. Ta regresijska črta je razmeroma najbolje prilagojena dejanskim vrednostim, ker je aritmetična sredina za vsako grupo najboljši reprezentant vrednosti v svojem razredu. Če za kak pojav dobimo, da so

grupne sredine konstantne vrednosti, sklepamo, da znak, po katerem smo izvedli grupacijo, ne vpliva na znak, za katerega smo izračunali aritmetične sredine. Nasprotno pa pomeni, da je ta vpliv velik, če so razlike med grupnimi aritmetičnimi sredinami v primerjavi s posamičnimi odkloni velike.

V primeru grupnih aritmetičnih sredin imamo varianco med grupnimi aritmetičnimi sredinami za pojasnjen del variance, varianco znotraj grup pa za nepojasnjen del variance. To načelo je istovetno z uvodnim tolmačenjem o pojasnjenem in nepojasnjenem delu variance.

Pri računanju skupne variance iz grupnih varianc smo dobili, da je skupna varianca σ_y^2 vsota variance med aritmetičnimi sredinami σ_M^2 in aritmetične sredine grupnih varianc M_{σ^2}

$$\sigma_y^2 = \sigma_M^2 + M_{\sigma^2} \qquad (18.92)$$

Pri tej osnovi dobimo kot determinacijski koeficient kvocient med varianco med aritmetičnimi sredinami σ_M^2 in skupno varianco σ_y^2. Po tej definiciji računan koeficient imenujemo korelacijsko razmerje, zaznamujemo pa z $\eta_{y.x}^2$ (eta kvadrat). Korelacijsko razmerje pokaže maksimalno stopnjo povezanosti dveh pojavov, ker se aritmetične sredine najbolj prilagajajo stvarnim podatkom.

Po tej opredelitvi je korelacijsko razmerje

$$\eta_{y.x}^2 = \frac{\sigma_M^2}{\sigma_y^2} = \frac{\frac{1}{N}\sum N_k (M_k - M)^2}{\frac{1}{N}\sum (y - M)^2} \qquad (18.93)$$

Razen dobrih lastnosti pa ima korelacijsko razmerje svoje hibe. Težko je namreč presoditi, kdaj so razredi pravilni, ker tako premajhni kot preveliki razredi popačijo rezultate. Sredine iz premajhnih razredov morejo biti rezultat slučajnih vplivov, grupne sredine iz prevelikih razredov pa ne podajajo regresijske smeri. Zato po pravilu razrede s premajhnim številom enot vključujemo v sorodne razrede.

Velika prednost korelacijskega razmerja $\eta_{y.x}^2$ pa je, da ga moremo izračunati ne

glede na značaj znaka x, da je le y numeričen, medtem ko je bilo dosedanje obravnavanje korelacije omejeno le na odvisnosti med numeričnimi znaki. Grupe za znak x moremo namreč sestaviti, čeprav je znak ordinalen ali nominalen. S korelacijskim razmerjem moremo izračunati mero odvisnosti dohodkov od kvalifikacije, pridelka od različnih vrst agrotehnike, kakovosti proizvodnje od izmene, starosti ob sklenitvi zakona od socialne skupine itd.

18.48 Računanje korelacijskega razmerja z osnovnim obrazcem je sicer izvedljivo, vendar neprikladno. Zato tehnično izvedemo računanje korelacijskega razmerja takole:

a) Sestavimo za znak x grupe s primernim številom enot po grupah.

b) Izdelamo tabelo, v kateri so podatki koreliranega znaka y vpisani po grupah za x.

c) Izračunamo grupne vsote Y_k za znak y in vsoto vseh grupnih vsot $Y = \sum Y_k$

d) Izračunamo vsoto kvadratov vseh posamičnih vrednosti za y. Tako dobimo

$$\sum_k \sum_i y_{ki}^2 = Q_{ki}$$

e) Korelacijsko razmerje izračunamo iz dobljenih izrazov po obrazcu

$$\eta_{y.x}^2 = \frac{\sum_k Y_k^2/N_k - Y^2/N}{\sum_k \sum_i y_{ki}^2 - Y^2/N} = \frac{Q_k - Q}{Q_{ki} - Q}$$

Če so podatki o znaku y_{ki} po grupah dani v frekvenčnih porazdelitvah, si pri računanju korelacijskega razmerja $\eta_{y.x}^2$ olajšamo posel po eni izmed metod za računanje aritmetične sredine in variance iz grupiranih podatkov.

18.49 Za zgled izračunajmo korelacijsko razmerje odvisnosti odstotka nepismenosti skupnega prebivalstva po okrajih SFRJ od republik. (Podatki so vzeti iz tabele 18.2). Iz teh podatkov dobimo:

	y_{ki}									Y_k	N_k
BiH	35,3	37,8	32,5	32,7	26,8	34,6				199,7	6
Hrvatska	10,6	19,1	11,0	13,1	5,9	17,0	18,1	7,4	7,6	109,8	9
Makedonija + Črna Gora	20,9	30,6	25,0	20,6	25,1	33,8	19,7	21,7		197,4	8
Slovenija	1,9	3,1	1,2	1,9						8,1	4
Srbija	9,2	23,2	25,5	33,1	24,9	24,4	26,0	26,7	25,8	314,1	15
	9,8	12,1	11,6	8,6	12,1	41,1					
										829,1	42

$$Q_{ki} = \sum_k \sum_i y_{ki}^2 = 35,3^2 + 37,8^2 + \ldots + 12,1^2 + 41,1^2 = 21247,05$$

$$Q_k = \sum_k Y_k^2/N_k = \frac{199,7^2}{6} + \frac{109,8^2}{9} + \frac{197,4^2}{8} + \frac{8,1^2}{4} + \frac{314,1^2}{15} = 19450,74$$

$$Q = Y^2/N = \frac{829,1^2}{42} = 16366,83$$

$$\eta^2_{y.Rep} = \frac{Q_k - Q}{Q_{ki} - Q} = \frac{19450,74 - 16366,83}{21247,05 - 16366,83} = 0,632$$

Iz tega sklepamo, da je 63,2% variabilnosti za odstotek nepismenih po okrajih pojasnjeno z razlikami med republikami. Zanimivo je, da je približno enak delež (62,9%) variabilnosti za odstotek nepismenih žensk po okrajih pojasnjeno z razlikami med republikami.

18.50 Če so podatki po grupah rezultati slučajnostnih vzorcev, moremo značilnost za korelacijsko razmerje preskusiti z izrazom

$$F = \frac{\eta^2_{y.x}}{1 - \eta^2_{y.x}} \cdot \frac{n - k}{k - 1} \qquad (18.94)$$

ki se pri istih predpostavkah kot pri analizi variance porazdeljuje v F porazdelitvi z $m_1 = k-1$ in $m_2 = n-k$ stopinjami prostosti. Izračunani F torej primerjamo s kri-

tičnimi vrednostmi F_α ($m_1 = k-1$; $m_2 = n-k$) pri različnih stopnjah tveganja α.

Z izrazom

$$F = \frac{\eta^2_{y.x} - r^2_{xy}}{1 - \eta^2_{y.x}} \cdot \frac{n-k}{k-2} \qquad (18.95)$$

pa s pomočjo F_α ($m_1 = k-2$; $m_2 = n-k$) preskušamo ničelno domnevo $H_o : y = \alpha + \beta x + \varepsilon$, da je y linearno odvisen od x. Seveda moremo ta preskus izvesti le, če sta x in y numerična, ker drugače ne moremo izračunati korelacijskega koeficienta r_{xy}.

KORELACIJA RANGA

18.51 Rang je v ozki zvezi z vrednostjo znaka. Večji vrednosti ustreza višji rang in obratno. To lastnost rangov s pridom uporabljamo tudi pri proučevanju korelacije med pojavi. Razvrstimo enote po velikosti po obeh koreliranih znakih in enotam pripišemo rang, enkrat po prvem, drugič pa po drugem znaku. Če sta proučevana znaka v ozki pozitivni povezavi, se rangi obeh znakov med seboj precej dobro ujemajo. V skrajnem primeru - če je monotona povezava funkcijska, so celo med seboj identični. Obratno so rangi pri tesni, a negativni odvisnosti med pojavoma močno različni, in nizkim rangom za en znak ustrezajo visoki rangi za drug znak. Če je med pojavoma majhna povezava ali je sploh ni, je tudi med rangi rahla zveza ali ni nobene zveze. Če osnovne podatke x y nadomestimo z rangi, imamo namesto niza dvojic vrednosti x in y niz dvojic rangov za x in y. Če je torej med znakoma x in y povezava tesna, je tesna tudi med rangoma. Če pa je povezava rahla ali je sploh ni, moremo to opaziti tudi na rangih.

To lastnost rangov izkoriščamo in računamo koeficient korelacije iz rangov namesto iz osnovnih podatkov. Če število enot ni preveliko, je računanje korelacije ranga znatno enostavnejše kot računanje korelacijskega koeficienta ϱ_{xy}. Koeficient korelacije ranga računamo po Spearmanovem obrazcu

$$\varrho_S = 1 - \frac{6\sum d_R^2}{N(N^2-1)} \qquad (18.96)$$

Pri tem je: ϱ_S = koeficient korelacije ranga; $\sum d_R^2 = \sum (R_y - R_x)^2$ = vsota kvadratov razlik med rangoma za oba znaka. Ker je koeficient korelacije ranga navaden koeficient korelacije med rangi, je vrednost Spearmanovega koeficienta korelacije ranga med -1 do +1.

Spearmanov koeficient ranga ϱ_S računamo tako, da ugotovimo za posamezne enote najprej range po znaku x (R_x) in po znaku y (R_y). Za vsako enoto izračunamo razliko med rangoma ($R_y - R_x$). Te razlike kvadriramo, kvadrate seštejemo, vsoto kvadratov pa vnesemo v obrazec 18.96.

Če ima več enot iste vrednosti x ali y, upoštevamo vezane range. Vezan rang je poprečje iz rangov za enote, ki imajo enako vrednost. Če ima npr. 5. in 6. enota po vrstnem redu enako vrednost za x, je vezan rang, ki ga upoštevamo za peto in šesto enoto $\frac{5+6}{2} = 5{,}5$. Če število vezanih rangov ni preveliko in zaporedja enakih vrednosti ne predolga, moremo kot približek izračunati ϱ_S po obrazcu 18.96; drugače pa je treba upoštevati ustrezen popravek.

ϱ_{xy} in ϱ_S, izračunana iz istih podatkov, nista enaka, ker sta izračunana različno. Vendar v večini primerov razlike niso velike in moremo ϱ_S vsaj v prvi stopnji proučevanja korelacije s pridom uporabiti kot orientacijski podatek. Prednost korelacije ranga je tudi v tem, da zmanjša učinek skrajnih vrednosti; to je ena izmed hib korelacijskega koeficienta ϱ_{xy}.

S korelacijo ranga razširimo proučevanje korelacije tudi na atributivne znake z ordinalnim značajem. Za take znake sicer ne moremo izračunati korelacijskega koeficienta ϱ_{xy}, ker so vrednosti znaka atributivne narave, moremo pa zanje izračunati korelacijski koeficient ranga ϱ_S. Tako moremo kolektiv delavcev razporediti po delovni vnemi, ki je sicer ne moremo ali zelo težko izrazimo numerično, učence po zanimanju za posamezne predmete, itd.

Korelacija ranga je posebno prikladna, kadar proučujemo odvisnosti več znakov. V teh primerih dobimo razmeroma hitro in enostavno celoten pregled o korelaciji med pojavi.

18.52 Za zgled vzemimo korelacijo med številom delavcev in številom opravljenih delovnih ur za 19 podjetij industrije obutve v SRS v decembru 1952.

Podatki in način računanja so dani v tabeli 18.14. Rangi so bili predhodno določeni z obdelovalnimi listki. Iz obdelovalnih listkov je bila nato sestavljena tabela 18.14.

Tabela 18.14. Računanje korelacijskega koeficienta ranga med številom delavcev (x) in številom opravljenih delovnih ur (y) za 19 podjetij industrije obutve v SRS v decembru 1952

x	y	R_x	R_y	d_R	d_R^2
7	0,8	1	1	0	0
12	2,2	2	2	0	0
15	2,9	3	3	0	0
23	4,1	4	4	0	0
28	4,9	5	5	0	0
37	7,3	6	6	0	0
49	9,9	7	8	-1	1
50	9,5	8	7	+1	1
51	10,1	9	10	-1	1
57	11,0	10	11	-1	1
58	10,0	11	9	+2	4
80	15,1	12	13	-1	1
81	15,0	13	12	+1	1
92	17,9	14	15	-1	1
100	16,7	15	14	+1	1
111	20,7	16	16	0	0
112	25,4	17	17	0	0
150	31,3	18	18	0	0
154	35,3	19	19	0	0

$12 = \sum d_R^2$

$$\rho_S = 1 - \frac{6,12}{19.(19^2 - 1)} = 0,9895$$

Razlika med ρ_S = ·9895 in ustreznim ρ_{xy} = ·988 je neznatna.

18.53 Za zgled uporabnosti Spearmanovega koeficienta korelacije ranga vzemimo še proučitev sorodnosti sestave tiskanih del po 37 vsebinskih področjih. Kot osnova za rangiranje je bila skupna tiraža za področje za posamezne republike. S Spearmanovim koeficientom korelacije ranga je bila sestava tiraže po posameznih republikah primerjana s sestavo v Sloveniji. Za leto 1966 so bili nakazani koeficienti korelacije ranga naslednji

$$\rho_{Slov.BiH} = 0,605; \quad \rho_{Slov.Črna\ G.} = 0,619; \quad \rho_{Slov.Hrv.} = 0,775$$

$$\rho_{Slov.Mak.} = 0,671; \quad \rho_{Slov.Srb.} = 0,880$$

Največja skladnost Slovenije v sestavi po področjih je bila s Srbijo, najmanjša pa z Bosno in Hercegovino.

18.54 Če so proučevani podatki slučajnosten vzorec, značilnost Spearmanovega koeficienta korelacije ranga r_S preskušamo z obrazcem, ki je enak obrazcu 18.49 za preskušanje značilnosti s koeficientom linearne odvisnosti r_{xy}. Za primer ničelne domneve $H_o: \rho_S = 0$ preskušamo iz vzorca izračunani r_S po obrazcu

$$\frac{r_S\sqrt{n-2}}{\sqrt{1-r_S^2}} = t \qquad (18.97)$$

s t-preskusom z $m = n-2$ stopinjami prostosti.

ASOCIACIJA IN KONTINGENCA

18.55 Dosedanje proučevanje odvisnosti je obsegalo predvsem odvisnosti med numeričnimi znaki. Izjema je bilo korelacijsko razmerje $\eta^2_{y.x}$ in Spearmanov koeficient korelacije ranga ρ_S. Pri proučevanju odvisnosti med socialno ekonomskimi pojavi pa spoznamo, da pojem odvisnosti ni omejen samo na numerične znake. Tako je pismenost odvisna od spola; socialni sestav od republike; kakovost izdelkov od dneva v tednu; izbira vrste izdelka od socialnega sestava ipd.

18.56 Kontingenčna tabela. Edino kar moremo na atributivne znake prenesti iz proučevanja odvisnosti med numeričnimi znaki, je tabela dvojic vrednosti in korelacijska tabela. Pri proučevanju odvisnosti med atributivnimi znaki v korelacijski tabeli numerična znaka zamenjamo s proučevanima atributivnima znakoma. Tako tabelo imenujemo kontingenčno tabelo.

Za zgled kontingenčne tabele vzemimo razvrstitev N = 530 izdelkov glede na izmeno, v kateri so bili izdelani in kakovost.

Rezultat evidence je tale kontingenčna tabela.

Tabela 18.15. Kontingenčna tabela odvisnosti kakovosti od izmene (dejanske frekvence f_{kg})

Kakovost	Izmena			Skupaj
	I.	II.	III.	
Dober	130	162	90	382
Slab, popravljiv	13	52	40	105
Slab, nepopravljiv	7	16	20	43
Skupaj	150	230	150	530

Iz absolutnih frekvenc v kontingenčni tabeli je težko presoditi, ali je kakovost odvisna od izmene, razen če primerjamo kakovost v I. izmeni s kakovostjo izdelkov v III. izmeni. Ker je proizvodnja v obeh izmenah enaka, moremo frekvenčne porazdelitve izdelkov po kakovosti med seboj neposredno primerjati. Bežen pregled pokaže, da je kakovost odvisna od izmene, ker sta sestavi po kakovosti v I. in III. izmeni različni in moremo iz njih sklepati, da je kakovost v III. izmeni slabša kot v I. izmeni.

Če je število izdelkov po izmenah različno, odkrijemo med izmenami razlike v kakovosti tako, da izračunamo strukturne vrste.

Za naš zgled je sestava kakovosti po izmenah dana v tabeli 18.16.

Tabela 18.16 Strukturne vrste za kakovost po izmenah

Kakovost	I z m e n a			Skupaj
	I.	II.	III.	
Izdelek, dober	86,7	70,4	60,0	72,1
Slab, popravljiv	8,7	22,6	26,7	19,8
Slab, nepopravljiv	4,6	7,0	13,3	8,1
S k u p a j :	100,0	100,0	100,0	100,0

Iz strukturnih vrst se jasno vidijo razlike v kakovosti v odvisnosti od izmene. S primerjavo strukturnih vrst sklepamo, da je kakovost v prvi izmeni najboljša, se v drugi izmeni poslabša in je najslabša v tretji izmeni.

Iz gornje tabele povzamemo, da je odvisnost kakovosti od izmene tem večja, čim večje so razlike v sestavi kakovosti po izmenah. Najbolj bi bila kakovost odvisna od izmene v primeru, če bi v eni izmeni bili vsi izdelki uporabni, v eni vsi slabi, a popravljivi, a v eni vsi slabi - nepopravljivi. V tej skrajnosti največje odvisnosti bi bila kontingenčna tabela kakor je prikazana v tabeli 18.17.

Tabela 18.17 Kontingenčna tabela pri največji odvisnosti

Kakovost	I.	II.	III.	Skupaj
Izdelek, dober	150	-	-	150
Slab, popravljiv	-	230	-	230
Slab, nepopravljiv	-	-	150	150
Skupaj	150	230	150	530

Iz te tabele spoznamo, da gre v primeru največje odvisnosti za funkcijsko odvisnost kakovosti od izmene, kajti za vsak izdelek vemo, kakšna je njegova kakovost,

če vemo, iz katere izmene izhaja.

Če skušamo ugotoviti, kakšna bi bila kontingenčna tabela, če kakovost ne bi bila odvisna od izmene, sklepamo, da bi v primeru, če kakovost ne bi bila odvisna od izmene, bila sestava kakovosti v vseh izmenah enaka in sicer enaka poprečni sestavi. Iz poprečne strukturne vrste (D = 72,1%, SP = 19,8%, SNP = 8,1%) dobimo število izdelkov posamezne kakovosti po izmenah tako, da poprečno strukturno vrsto po kakovosti pomnožimo z ustreznim številom izdelkov v izmeni. Tako dobimo tabelo 18.18.

Tabela 18.18. Kontingenčna tabela pri neodvisnosti kakovosti od izmene (teoretične frekvence pri neodvisnosti f'_{kg})

Kakovost	I z m e n a			Skupaj
	I.	II.	III.	
Izdelek, dober	108	166	108	382
Slab, popravljiv	30	45	30	105
Slab, nepopravljiv	12	19	12	43
S k u p a j :	150	230	150	530

Prvo frekvenco pri neodvisnosti dobimo, če skupno število izdelkov v prvi izmeni (150) pomnožimo s poprečnim strukturnim deležem ($\frac{382}{530} = 0{,}721$)

$$108 = 150 \cdot \frac{382}{530}$$

Podobno dobimo vse druge. Za skupino "slab - nepopravljiv" v III. izmeni dobimo $12 = 150 \cdot \frac{43}{530}$.

Če s f_{kg} zaznamujemo na splošno frekvenco v stolpcu "k" in vrsti "g", s f_k vsote frekvenc po stolpcih, s f_g vsote frekvenc po vrstah, z N pa skupno število enot, iz gornjega primera presodimo, da dobimo "teoretične frekvence" pri predpostavki,

da pojava nista odvisna, po splošnem obrazcu. Frekvenci f_{kg} ustrezna "teoretična frekvenca f'_{kg}" je

$$f'_{kg} = \frac{f_k \cdot f_g}{N} \qquad (18.98)$$

produkt polju kg ustreznih robnih frekvenc deljen z obsegom populacije.

V tabeli 18.18 so teoretične frekvence zaradi nazornosti zaokrožene na cele vrednosti, vendar to ni nujno niti običajno. V računih teoretične frekvence računamo kot decimalna števila.

Kot povzamemo iz dosedanjih razmišljanj, je razlika med stvarnimi f_{kg} in teoretičnimi frekvencami f'_{kg} tem večja, čim večja je odvisnost med proučevanima pojavoma; če odvisnosti ni, pa so za vsa polja $d_{kg} = f_{kg} - f'_{kg} = 0$.

Za naš zgled so razlike $d_{kg} = f_{kg} - f'_{kg}$ podane v tabeli 18.19

Tabela 18.19 Razlike med stvarnimi in teoretičnimi frekvencami $d_{kg} = f_{kg} - f'_{kg}$ za odvisnost kakovosti od izmene.

Kakovost	I z m e n a			Skupaj
	I.	II.	III.	
Izdelek, dober	+22	-4	-18	0
Slab, popravljiv	-17	+7	+10	0
Slab, nepopravljiv	- 5	-3	+8	0
S k u p a j	0	0	0	0

V tej tabeli je za vsako polje razvidno, za koliko se vsaka frekvenca odklanja od stanja, če bi bila pojava neodvisna.

Merila stopnje kontingence

18.57 V tabeli d_{kg} je k.g razlik, za katere vemo, da so v splošnem absolutno velike, če je odvisnost velika in v splošnem absolutno majhne, če je odvisnost

majhna oziroma nič, če odvisnosti ni. Zato je jasno, da sintetična mera stopnje odvisnosti med znakoma vsebuje razlike d_{kg}. Iz razlogov, ki jih bomo kasneje obravnavali, vzamemo kot sumarno merilo odvisnosti med dvema atributivnima znakoma χ^2

$$\chi^2 = \sum_{kg} \frac{(f_{kg} - f'_{kg})^2}{f'_{kg}} \qquad (18.99)$$

ki ga že poznamo kot merilo prilagojenosti dejanskih frekvenc teoretičnim frekvencam.

Ta obrazec moremo dalje pisati v operativnejših obrazcih

$$\chi^2 = \sum_{kg} \frac{f_{kg}^2}{f'_{kg}} - N = \left(\sum_{kg} \frac{f_{kg}^2}{f_k \cdot f_g} - 1 \right) N = \left(\sum_k \frac{1}{f_k} \sum_g \frac{f_{kg}^2}{f_g} - 1 \right) N \qquad (18.100)$$

18.58 Če za zgled izračunamo χ^2 za primer odvisnosti kakovosti od izmene, dobimo iz tabel 18.18 in 18.19

$$\chi^2 = \frac{(+22)^2}{108} + \frac{(-4)^2}{166} + \frac{(-18)^2}{108} + \ldots + \frac{(-3)^2}{19} + \frac{(+8)^2}{12} = 29{,}52$$

Če χ^2 izračunamo po zadnjem obrazcu v 18.100, po katerem ne potrebujemo neposredno teoretičnih frekvenc f'_{kg}, dobimo iz tabele 18.15

$$\chi^2 = \left(\sum_k \frac{1}{f_k} \sum_g \frac{f_{kg}^2}{f_g} - 1 \right) N = \left[\frac{1}{150} \left(\frac{130^2}{382} + \frac{13^2}{105} + \frac{7^2}{43} \right) + \frac{1}{230} \cdot \right.$$

$$\left. \left(\frac{162^2}{382} + \frac{52^2}{105} + \frac{16^2}{43} \right) + \frac{1}{150} \left(\frac{90^2}{382} + \frac{40^2}{105} + \frac{20^2}{43} \right) - 1 \right] 530 = 29{,}03$$

Dobljeni rezultat je točnejši kot rezultat $\chi^2 = 29{,}52$, ki je izračunan iz zaokroženih podatkov.

18.59 χ^2 pa je slabo primerljivo merilo odvisnosti - kontingence, predvsem zaradi tega, ker je pri isti stopnji kontingence odvisen tudi od velikosti populacije. Vpliv velikosti populacije odpravimo tako, da χ^2 reduciramo in izračunamo poprečno kvadratično kontingenco

$$\phi^2 = \chi^2/N = \sum_{kg} \frac{f_{kg}^2}{f_k f_g} - 1 \qquad (18.101)$$

Vendar tudi ϕ^2 ni idealno merilo kontingence, ker je odvisna od tega, koliko vrst oziroma stolpcev ima kontingenčna tabela. Da odstranimo še ta vpliv, izračunavamo dva nadaljnja koeficienta. S Pearsonovim koeficientom kontingence C

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}} = \sqrt{\frac{\phi^2}{1+\phi^2}} \qquad (18.102)$$

dosežemo, da mera kontingence nima le svoje spodnje meje 0, temveč tudi zgornja meja C v nobenem primeru ni večja kot 1. Vendar je tudi za C zgornja meja še vedno odvisna od števila vrstic oziroma stolpcev. Zato ga bomo z največjo možno vrednostjo normirali tako, da bo njegova zgornja meja vedno 1.

Kratek račun pokaže, da je $\chi^2_{max} = N(k-1)$, pri čemer je k število vrednosti za znak, ki ima manj vrednosti, $k \leq g$.

Če to upoštevamo, je

$$C_{max} = \sqrt{\frac{N(k-1)}{N+N(k-1)}} = \sqrt{\frac{k-1}{k}} \qquad (18.103)$$

Če s to vrednostjo normiramo C, dobimo

$$C_{pop} = \frac{C}{C_{max}} = \sqrt{\frac{k}{k-1} \cdot \frac{\chi^2}{N+\chi^2}} \qquad (18.104)$$

S χ^2_{max} moremo normirati tudi χ^2

$$\chi^2_{pop} = \frac{\chi^2}{N(k-1)} \tag{18.105}$$

Običajno izračunamo kvadratni koren in dobimo

$$r = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{18.106}$$

korelacijski koeficient, ki leži v mejah $0 \leq r \leq 1$.

18.60 Za naš zgled, v katerem je $\chi^2 = 29,03$ in $N = 530$, je

$$\phi^2 = \frac{29,03}{530} = 0,05477$$

$$C = \sqrt{\frac{29,03}{530+29,03}} = 0,2279$$

$$C_{pop} = \sqrt{\frac{3}{(3-1)} \cdot \frac{29,03}{(530+29,03)}} = 0,2791$$

$$r = \sqrt{\frac{29,03}{530\,(3-1)}} = 0,1655$$

18.61 Za zgled proučitve odvisnosti z merami kontingence vzemimo še odvisnost med prejšnjim stanom ženina in neveste in med vrstnim redom zakona ženinov in nevest v razdobju 1955-1968 v SFRJ.

Za leto 1968 sta ustrezni tabeli podani v tabeli 18.20.

Tabela 18.20. Število porok a) po prejšnjem stanu ženina in neveste in b) vrstnem redu zakona ženinov in nevest v SFRJ v letu 1968. (Vir: SG 1970, str. 85.)

a)

Nevesta	Ženin			Skupaj
	samski	vdovec	razvezan	
samska	143809	2557	9043	155409
vdova	704	1956	907	3567
razvezana	4435	1683	5376	11494
Skupaj	148948	6196	15326	170470

b)

Nevesta	Ženin			Skupaj
	prvi	drugi	tretji in več	
prvi	143809	10889	711	155409
drugi	4989	8238	936	14163
tretji in več	150	524	224	898
Skupaj	148948	19651	1871	170470

V tabeli 18.19 so stvarne frekvence f_{kg} primerjane s teoretičnimi f'_{kg} z diferencami $d_{kg} = f_{kg} - f'_{kg}$. Včasih je naravneje, če jih primerjamo z indeksi $I_{kg} = 100 \cdot \frac{f_{kg}}{f'_{kg}}$.

V naših dveh zgledih iz demografske statistike so indeksi I_{kg} dani v tabeli 18.21.

Tabela 18.21. Indeksi I_{kg} za število porok po prejšnjem stanu in vrstnem redu zakona ženinov in nevest

a) Nevesta	Ž e n i n:		
	samski	vdovec	razvezan
samska	106	45	65
vdova	23	1509	283
razvezana	44	403	520

b) Zakon neveste	Zakon ženina		
	prvi	drugi	tretji in več
prvi	106	61	42
drugi	40	505	602
tretji in več	19	506	2273

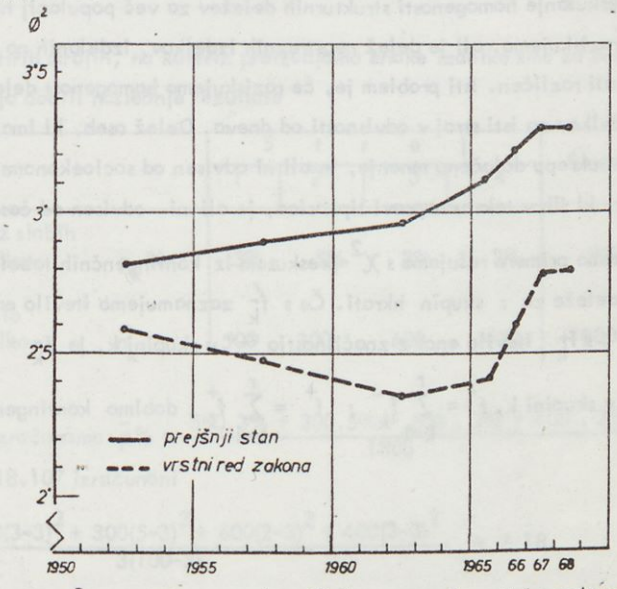
Ti indeksi pokažejo nazorno odvisnosti med proučevanimi znaki. Če bi ne bilo odvisnosti, bi bili vsi indeksi enaki 100. Tako pa opazimo, da je podpoprečno zakonskih parov, v katerih se samski ali samska poroči z vdovo ali razvezano oziroma vdovcem oziroma razvezanim. Nadpoprečno pa je zakonov med vdovami in razvezanimi z vdovci in razvedenci, posebno pa izstopa kombinacija vdovec-vdova z indeksom 1509. Teh zakonov je torej 15-krat toliko kolikor bi jih bilo, če odvisnosti ne bi bilo. Precej podobna slika se pokaže tudi pri proučitvi odvisnosti glede na vrstni red zakona. Enako so kombinacije prvič, drugič ali večkratnim za moške in ženske podpoprečen, nadpoprečni pa večkrat z večkrat. Posebej pride do izraza tretjič in več pri ženinu in nevesti z indeksom 2273.

Da proučimo dinamiko v odvisnosti med prejšnjim stanom in odvisnosti glede na vrstni red zakona, imamo v tabeli 18.22 izračunane ϕ^2 po letih. Ker so vsi ϕ^2 izračunani iz tabel 3x3, so med seboj primerljivi.

Tabela 18.22. ϕ^2 za odvisnost med prejšnjim stanom in vrstnim redom zakona za poroke v SFRJ v razdobju 1950-1968

Razdobje	ϕ^2	
	prejšnji stan	vrstni red zakona
1950 - 1954	0,2817	0,2582
1955 - 1959	0,2907	0,2470
1960 - 1964	0,2956	0,2346
1965	0,3115	0,2409
1966	0,3215	0,2600
1967	0,3296	0,2798
1968	0,3296	0,2803

Dinamika odvisnosti je še lepše vidna iz slike 18.11. ϕ^2 kaže, da je odvisnost glede na prejšnji stan močnejša kot glede na vrstni red zakona. V dinamiki se je odvisnost od prejšnjega stanu povečevala, medtem ko je stopnja odvisnosti glede na vrstni red zakona v proučevanem obdobju najprej padala, stagnirala, v zadnjih letih pa konstantno naraščala.



Slika 18.11. ϕ^2 za odvisnost med prejšnjim stanom in vrstnim redom zakona za poroke v SFRJ v razdobju 1950-1968.

18.62 Preskušanje domnev o odvisnosti s χ^2-preskusom. Pogosto je kontingenčna tabela rezultat vzorca iz umišljene populacije. Če velja ničelna domneva, da sta znaka neodvisna, se koeficient kontingence χ^2 porazdeljuje v χ^2 porazdelitvi. Število stopinj prostosti je lahko določiti. Od skupno k.g podatkov v kontingenčni tabeli je v vsaki vrsti oziroma stolpcu ena frekvenca odvisna od drugih, ker so vsote frekvenc vezane na robne vsote. Prav zato je število stopinj prostosti $m = (k-1)(g-1)$.

Če štejemo, da je $n = 530$ izdelkov vzorec iz populacije z neomejenim številom izdelkov, ki bi jih izdelali ob enakih pogojih, moremo preskusiti, ali je kakovost značilno odvisna od izmene. Ker je $\chi^2_{P=0.001}(m=(3-1).(3-1)=4 = 18,467$, izračunani $\chi^2 = 29,03$ pa večji od njega, sklepamo, da je kakovost zelo visoko ($z\alpha = 0.001$) značilno odvisna od izmene.

Homogenost za skupino strukturnih deležev

18.63 Poseben primer uporabe preskušanja značilnosti kontingence s χ^2-preskusom je preskušanje homogenosti strukturnih deležev za več populacij hkrati. V proizvodnji raziskujemo, ali je delež neustreznih izdelkov, izdelanih na različnih strojih, enak ali različen. Isti problem je, če raziskujemo homogenost deleža neustreznih izdelkov za isti stroj v odvisnosti od dneva. Delež oseb, ki imajo o nekem gospodarskem ukrepu določeno mnenje, je ali ni odvisen od socioekonomske skupine. Delež napak, ki jih v tekstu napravi tipkarica, je ali ni odvisen od časa ipd.

Take in podobne primere rešujemo s χ^2-preskusom iz kontingenčnih tabel $2 \times r$, če proučujemo deleže za r skupin hkrati. Če s f_k^- zaznamujemo število enot z značilnostjo "-", s f_k^+ število enot z značilnostjo "+" v skupini k, je $f_k^- + f_k^+ = n_k$ število enot v skupini k, $f^- = \sum_{k=1}^{r} f_k^-$; $f^+ = \sum_{k=1}^{r} f_k^+$, dobimo kontingenčno tabelo:

	1	2 ..... k ..... r	
−	f_1^-	f_2^- ..... f_k^- ..... f_r^-	f^-
+	f_1^+	f_2^+ ..... f_k^+ ..... f_r^+	f^+
	n_1	n_2 n_k n_r	n

Za kontingenčno tabelo 2 x r izračunamo χ^2 po enem izmed obrazcev

$$\chi^2 = \frac{\sum_k n_k (p_k\% - \bar{p}\%)^2}{\bar{p}\%\,(100 - \bar{p}\%)} \; ; \; p_k\% = 100\,\frac{f_k^+}{n_k} \; ; \; \bar{p}\% = 100\,\frac{f^+}{n} \qquad (18.107)$$

$$\chi^2 = \frac{n^2}{f^- \cdot f^+}\left[\sum_k \frac{(f_k^+)^2}{n_k} - \frac{(f^+)^2}{n}\right] \qquad (18.108)$$

Če štejemo n_k enot kot vzorce, značilnost odvisnosti proučujemo z χ^2-preskusom z $m = (2-1)(r-1) = r-1$ stopnjami prostosti.

18.64 Na štirih strojih, na katerih proizvajamo enake izdelke smo za dnevno proizvodnjo dobili naslednje rezultate

		Stroj				Skupaj
		1	2	3	4	
delež slabih izdelkov	$p_k\%$	3%	5%	2%	3%	3%
število izdelkov	n_k	500	300	600	400	1800

Najprej izračunamo $\bar{p}\% = \frac{500.3\% + 300.5\% + 600 \cdot 2\% + 400 \cdot 3\%}{1800} = 3\%$.

Po obrazcu 18.107 izračunani

$$\chi^2 = \frac{500(3-3)^2 + 300(5-3)^2 + 600(2-3)^2 + 400(3-3)^2}{3(100-3)} = 6{,}18.$$

Ker je izračunani $\chi^2 = 6,18$ manjši kot je kritična vrednost $\chi^2_{0,10}$ (m=4-1=3)=6,25, sklepamo, da so razlike v deležih med stroji neznačilne, kar pomeni, da ni razloga za sklep, da so deleži med stroji različni.

18.65 V anketi je od $n_A = 200$ oseb iz socioekonomske skupine A imelo $f_A^+ = 50$, od $n_B = 300$ oseb iz skupine B $f_B^+ = 80$, od $n_C = 400$ oseb iz skupine C pa $f_C^+ = 160$ oseb o ukrepu U ugodno mišljenje. Preskusiti je treba, ali je delež oseb, ki ima ugodno mišljenje, značilno odvisen od socioekonomske skupine. Glede na to, da imamo dane frekvence, uporabimo obrazec 18.108. Tako dobimo

$$\chi^2 = \frac{900^2}{290 \cdot 610}\left[\frac{50^2}{200} + \frac{80^2}{300} + \frac{160^2}{400} - \frac{290^2}{900}\right] = 20,10$$

Ker je $\chi^2_{P=0.001}$ (m=2) $= 13,82$, sklepamo, da je med socioekonomskimi skupinami značilno različen delež oseb, ki ima o ukrepu dano mišljenje.

18.66 Za pojave, ki redko nastopajo, običajno nimamo podatkov o skupnem številu za skupine. Tako so za nesreče pri delu kot potencialne enote vsi zaposleni delavci ali mera zanj število delovnih ur. Če predpostavimo, da je število zaposlenih v času ene izmene enako, moremo tudi izračunati, da je teoretična frekvenca nesreč v danem razdobju sorazmerna dolžini razdobja. V posebnem primeru, če so vsa ta razdobja enaka, je teoretično število nesreč pri domnevi, da število nesreč ni odvisno od izmene, v vsaki uri enako. V shematičnem zgledu vzemimo, da smo s statističnim opazovanjem zbrali podatke o številu nesreč po urah v izmeni in dobili tele rezultate:

ura v izmeni	1	2	3	4	5	6	7	8	
število nesreč	45	20	12	24	30	41	51	59	282 = n
									37 = $\bar{f}$

Teoretično število pri predpostavki, da je število delavcev po urah v izmeni stalno in pri ničelni domnevi, da število nesreč ni odvisno od ure v izmeni, je:

$\bar{f} = n/r = 282/8 = 37$. Za ta primer prirejen obrazec za χ^2 je

$$\chi^2 = \frac{\sum_k (f_k - \bar{f})^2}{\bar{f}} = \frac{r}{n} \sum_k f_k^2 - n$$

pri čemer so f_k grupne frekvence, r število enakih razdobij, n pa skupno število nesreč v proučevanem razdobju. V primeru, če velja ničelna domneva, se ta izraz porazdeljuje v χ^2 z m = r-1 stopinjami prostosti. Za gornji zgled dobimo:

$$\chi^2 = \frac{8}{282} \left[45^2 + 20^2 + \ldots + 59^2 \right] - 282 = 52{,}98$$

Ker je izračunani $\chi^2 = 52{,}98$ večji kot kritična vrednost $\chi^2_{0,001}(7) = 24{,}32$, sodimo, da so razlike v številu nesreč po urah značilno različne na ravni tveganja $\alpha = 0{,}001$.

ASOCIACIJA

18.67 Pokazovalci asociacije. Če proučujemo kontingenco za znaka z dvema vrednostima, govorimo o asociaciji. Problematika proučevanja asociacije je enaka kot pri kontingenci, le da je asociacija poseben primer kontingence. Zato moremo za merjenje stopnje asociacije uporabiti vse pokazovalce kontingence χ^2, ϕ^2 in C.

Za asociacijo, za katero je kontingenčna tabela 2 x 2, ki jo moremo pisati simbolično

a	b	a + b
c	d	c + d
a+c	b+d	N

(18.109)

je tabela teoretičnih frekvenc

$$a' = \frac{(a+c)(a+b)}{N} \qquad b' = \frac{(b+d)(a+b)}{N} \tag{18.110}$$

$$c' = \frac{(a+c)(c+d)}{N} \qquad d' = \frac{(b+d)(c+d)}{N}$$

Pokazovalci asociacije, prenešeni iz kontingence za 2 x 2 pa so v gornjih simbolih

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = (a-a')^2\left[\frac{1}{a'} + \frac{1}{b'} + \frac{1}{c'} + \frac{1}{d'}\right] \tag{18.111}$$

$$\phi^2 = \frac{(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \tag{18.112}$$

Atributivnim znakom v ožjem pomenu, to je znakom z dvema vrednostima, pripišemo formalne numerične vrednosti: eni 0, drugi pa 1. Za tak primer moremo iz kombinacijske tabele

y \ x	0	1	
0	a	b	a+b
1	c	d	c+d
	a+c	b+d	N

izračunati formalno korelacijski koeficient ρ_{xy}. Izkaže se, da je izračunani koeficient V

$$V = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \tag{18.114}$$

v ozki zvezi s χ^2 oziroma ϕ^2

$$V = \sqrt{\frac{\chi^2}{N}} = \sqrt{\phi^2} \tag{18.115}$$

Razen nakazanih koeficientov asociacije pogosto izračunavamo Yule-ov koeficient asociacije Q

$$Q = \frac{ad - bc}{ad + bc} \qquad (116)$$

Če pregledamo vse nakazane koeficiente, spoznamo, da je v vseh izraz (ad-bc), enak nič, če ni asociacije.

V naslednji razpredelnici so nakazani razmiki, v katerih leže vrednosti za posamezne koeficiente in pri katerih pogojih.

Pogoj za spodnjo mejo	Razmik	Pogoj za zgornjo mejo
$ad = bc$	$0 \leq \chi^2 \leq N$ $0 \leq \phi^2 \leq 1$	$b = c = 0$ ali $a = d = 0$
$a = d = 0$	$-1 \leq V \leq +1$	$b = c = 0$
$a = 0$ ali $d = 0$	$-1 \leq Q \leq +1$	$b = 0$ ali $c = 0$

Predznak pri koeficientu asociacije V in Q pride v poštev le v izjemnih primerih, če imata atributa točno določeno smer npr. negnojeno 0 - gnojeno 1 ; zmanjšanje pridelka 0, povečanje pridelka 1 ; ali stari način 0 , novi način 1 ; produktivnost: manjša 0, produktivnost večja 1 ipd. V splošnem pa uporabimo kot pokazovalce stopnje asociacije absolutne vrednosti.

18.68 Za prvi zgled vzemimo kontingenčno tabelo med pogojema A, B in uspehom neuspešno - uspešno. Za N = 350 preskusov smo dobili tole kombinacijsko tabelo:

		uspešnost		skupaj
		-	+	
Pogoj	A	29	121	150
	B	11	189	200
	Skupaj	40	310	350

$$\chi^2 = \frac{350\,(29.189 - 11.121)^2}{150 \cdot 200 \cdot 40 \cdot 310} = 16,204$$

$$\phi^2 = 16,2040/350 = 0,0463$$

$$V = \sqrt{0,0463} = +0,2151$$

$$Q = \frac{29.189 - 11.121}{29.189 + 11.121} = 0,6092$$

Če za preskus izračunamo χ^2 še po obrazcu, ki vsebuje teoretične frekvence, dobimo

$$a' = \frac{40.150}{350} = 17,14 \quad b' = \frac{310 \cdot 150}{350} = 132,86$$

$$c' = \frac{40.200}{350} = 22,86 \quad d' = \frac{310 \cdot 200}{350} = 177,14$$

$$\chi^2 = (29 - 17,14)^2 \left[\frac{1}{17,14} + \frac{1}{132,86} + \frac{1}{22,86} + \frac{1}{177,14} \right] = 16,212$$

18.69 Za zgled analize asociacije z Yule-ovimi koeficienti Q vzemimo prebivalstvo po spolu in delovnosti po republikah po popisih 1953 in 1961. Koeficienti Q so izračunani iz tabel, od katerih navajamo tabelo za SR Slovenijo, po popisu 1961.

Tabela 18.23. Število prebivalstva po spolu in delovnosti v SRS po popisu 1961 (v tisočih prebivalcev) - Vir: SG-1965)

Spol	Delovnost		Skupaj
	delovni	vzdrževani	
moški	512	249	761
ženske	380	450	830
Skupaj	892	699	1591

$$Q = \frac{512.450 - 380.249}{512.450 + 380.249} = 0,418$$

V tabeli 18.24 so po republikah dani analogni koeficienti po popisih prebivalstva 1953 in 1961.

Tabela 18.24. Koeficienti asociacije Q za odvisnost delovnosti od spola po republikah po popisu prebivalstva 1953 in 1961

Republika	Q_{53}	Q_{61}	Q_{61}/Q_{53}
BiH	0,584	0,577	98,8
Črna Gora	0,593	0,579	97,6
Hrvatska	0,597	0,524	87,8
Makedonija	0,646	0,594	92,0
Slovenija	0,488	0,418	85,7
Srbija	0,583	0,542	93,0

Iz preglednice 18.24 razberemo, da je bila v 1953 najmanjša povezanost med spolom in delovnostjo v SR Sloveniji ($Q_{61} = 0,488$), največja pa v Makedoniji ($Q_{53} = 0,646$), medtem ko je za druge republike precej izenačena (od 0,583 do 0,597). Opozoriti moramo, da je majhen Q rezultat izenačenih razmer za moške in ženske. Če primerjamo Q_{53} s Q_{61}, spoznamo, da se je stopnja asociacije zmanjšala v vseh republikah (proces je šel v smeri zaposlovanja žensk). Največje relativno zmanjšanje je v Sloveniji (indeks 85.7), najmanjše pa v BiH (indeks 98,8).

18.70 Preskušanje domnev o asociaciji. Če so podatki v kontingenčni tabeli 2 x 2 rezultat iz vzorca, značilnost asociacije podobno kot za kontingenco preskušamo z χ^2-preskusom. Ker je ta preskus le približen, izračunamo po obrazcu

$$\chi^2_{pop} = \frac{\left(|ad-bc| - \frac{n}{2}\right)^2 n}{(a+b)(c+d)(a+c)(b+d)} = \left(|a - a'| - 0.5\right)^2 \left[\frac{1}{a'} + \frac{1}{b'} + \frac{1}{c'} + \frac{1}{d'}\right] \qquad (18.117)$$

popravljen χ^2 (Yatesov popravek). Ker je za asociacijo $m = (k-1)(g-1)=(2-1)(2-1)=1$, primerjamo χ^2_{pop} s kritičnimi vrednostmi za posamezne stopnje tveganja pri $m = 1$

α	0,10	0,05	0,01	0,001
$\chi^2_\alpha(1)$	2,71	3,84	6,64	10,83

Popravek uporabljamo, če so teoretične frekvence relativno majhne (najmanjša manjša kot 5).

18.71 Če želimo rezultate, dobljene z n = 350 preskusi v 18.68 posplošiti, štejemo teh 350 preskusov za vzorec in s χ^2-preskusom preskušamo značilnost asociacije. Po obrazcu 18.117 dobimo

$$\chi^2_{pop} = (29-17,14 \quad -0,5)^2\left[\frac{1}{17,14}+\frac{1}{132,86}+\frac{1}{22,86}+\frac{1}{177,14}\right] = 14,87$$

Z Yatesovim popravkom popravljeni χ^2_{pop} je malo manjši od χ^2, ki smo ga izračunali po obrazcu 18.111. Če ga primerjamo z ustreznimi kritičnimi vrednostmi $\chi^2_\alpha(m=1)$, sklepamo, da je asociacija značilna na stopnji $\alpha = 0,001$, ker je

$$\chi^2_{pop} = 14,87 > \chi^2_{\alpha=0.001}(1) = 10,83$$

Preskus z mediano

18.72 Asociacijo vzamemo za osnovo pri nekaterih neparametričnih preskusih. Eden izmed njih je preskus z mediano.

Vzemimo za zgled tekoči trak, na katerem smo v teku ene izmene proizvedli n izdelkov. Če razdelimo proizvodnjo glede na čas proizvodnje v dve, po številu izdelkov enaki polovici in najdemo, da je bilo a neustreznih izdelkov v prvi polovici in b neustreznih izdelkov v drugi polovici proizvodnje, moremo sestaviti tabelo 2 x 2.

	slabih	dobrih	
proizvedeno: - v prvi polovici	a	n/2-a	n/2
- v drugi polovici	b	n/2-b	n/2
	a+b	n-a-b	n

Če za ta primer izračunamo χ^2, dobimo po obrazcu 18.117

$$\chi^2_{pop} = \frac{(|a-b| - 1)^2 . n}{(a+b)(n - a - b)} \qquad (18.118)$$

18.73 Za n=151 zaporedoma proizvedenih izdelkov smo ugotovili, da je od 75 izdelkov v prvi polovici 5, od 75 izdelkov v drugi polovici pa 10 neuporabnih. 76. izdelek po vrstnem redu smo izpustili. Zato vzamemo v račun n=75+75=150 izdelkov. Ali obstaja značilna razlika v proizvodnji glede na čas, ugotovimo s χ^2-preiskusom po obrazcu 18.118

$$\chi^2_{pop} = \frac{(15 - 10| - 1)^2 . 150}{(5 + 10)(150 - 5-10)} = 1,18$$

Ker je izračunani $\chi^2 = 1,18$ manjši kot $\chi^2_{0.10}(1) = 2,71$, sklepamo, da so razlike v kakovosti dela, v odvisnosti od časa, neznačilne.

V tem primeru je bilo vodilo za razdelitev v dve skupini oziroma za zaporedje enot čas. S preskusom z mediano smo zato ugotavljali, ali so razlike v kakovosti glede na čas.

18.74 Preskus z mediano pa moremo uporabiti tudi za preskus o merah centralne tendence. Pri proučevanju značilnosti razlik med sredinama smo imeli opravka z dvema vrstama problemov. Vezana vzorca sta pripeljala na problem preskušanja značilnosti razlik, neodvisna vzorca pa na preskušanje razlik med aritmetičnima sredinama za dva vzorca.

18.75 Neparametrično preskušanje domnev za vezana vzorca.

Neparametrična metoda preskušanja razlik za dva vezana vzorca se s prireditvijo χ^2-preskusa za mediano prevede na preskus s predznaki. Da bi preskusili domnevo $H_o : M_e(y_2) = Me(y_1) + D$, izračunamo iz podatkov $y_{1,i}$ in $y_{2,i}$ za dva odvisna vzorca izraze $u_i = y_{2i} - y_{1i} - D$. Če velja ničelna domneva, je število vrednosti u_i s pozitivnimi predznaki (f^+) in število enot z negativnimi predznaki (f^-) približno enako oziroma v skladu s slučajnostnimi odkloni. V nasprotnem primeru, če ničelna domneva ne velja, pa razlika med f^+ in f^- ni odvisna samo od slučaja. Za ta primer odvisnosti preskušamo ničelno domnevo s χ^2- preskusom števila pozitivnih f^+ in negativnih f^- po obrazcu

$$\chi^2 = \frac{(|f^+ - f^-| - 1)^2}{f^+ + f^-} \qquad (18.119)$$

z m=1 stopinjo prostosti.

Do tega obrazca pridemo iz 18.118, če vzamemo, da je a+b $\leq$ n.

Pri uvajanju novega načina dela preskušamo, ali je po novem postopku produktivnost dela vsaj za D = 10 kosov na uro večja. V ta namen je 100 slučajnostno izbranih delavcav delalo po obeh postopkih. Ugotovili smo, da je $u_i = y_{2i} - y_{1i} - 10$ za $f^+ = 60$ delavcev pozitiven, za $f^- = 35$ negativen in za $f^0 = 5$ delavcev enak nič. V χ^2-preskusu po obrazcu 18.119 upoštevamo le f^+ in f^- in izpustimo delavce, za katere je bil $u_i = 0$. Za naše podatke dobimo:

$$\chi^2 = \frac{(60 - 35 - 1)^2}{60 + 35} = 6,06 ,$$

Ker je preskus enostranski, bi v primeru, če bi bil $f^+ \leq f^-$, ne postopali naprej in produktivnost dela po B ne bi šteli za značilno večjo od A. Ker je v našem primeru $f^+ > f^-$, χ^2 pa 6,06, smatramo, da so razlike v produktivnosti dela značilno večje od D = 10 s tveganjem $\alpha = P = 0,05$, ker je

$$\chi^2_{0.05}(1) = 3.84 < \chi^2 = 6,06 < \chi^2_{0.01}(1) = 6,64$$

Neparametrični preskus za dva neodvisna vzorca

18.76 Če gre za dva neodvisno vzorca z n_1 in n_2 enotami, neparametrično preskušamo ničelno domnevo $H_o : Me_1 = Me_2$ po temle postopku. Pri časovnem proučevanju je bilo vodilo za razvrstitev izdelkov čas proizvodnje za izdelek. V splošnem pa more biti vodilo katerikoli proučevani znak, ki je vsaj ordinalen.

Če bi v prejšnjem zgledu nakazani problem reševali z dvema neodvisnima vzorcema, bi n_1 slučajnostno izbranih delavcev delalo po prvem načinu, n_2 pa po drugem. Če n_1 rezultatov po prvem načinu $y_{1,1} \cdot y_{1,2} \cdot \cdots \cdot y_{1,n1}$ in n_2 rezultatov $y_{2,1} \cdot y_{2,2} \cdot \cdots \cdot y_{2,n_2}$, ki smo jih dobili po drugem načinu razvrstimo po velikosti v skupno ranžirno vrsto z $n=n_1 + n_2$ členi, pričakujemo, da bo več vrednosti y_2 desno od mediane, če je drugi način različen od prvega, in približno enako število y_2 levo in desno od mediane, če sta postopka po kakovosti enaka.

Sestaviti moremo tabelo 2 x 2 glede na način dela in glede na to ali je dosežen uspeh pod ali nad mediano.

	pod mediano	nad mediano	Skupaj
Pogoj 1	$n_1 - a_1$	a_1	n_1
Pogoj 2	$n_2 - a_2$	a_2	n_2
Skupaj	$\frac{n}{2}$	$\frac{n}{2}$	n

Če za ta primer priredimo obrazec 18.117, dobimo

$$\chi^2 = \frac{(|n_1 - 2a_1| - 1)^2 . n}{n_1 (n - n_1)} = \frac{(|n_2 - 2a_2| - 1)^2 . n}{n_2 (n - n_2)} \qquad (18.120)$$

ki ga preskušamo s $\chi^2(1)$ z eno stopinjo prostosti.

Če je $n_1 = n_2 = \bar{n}$, je obrazec 18.120 še enostavnejši

$$\chi^2 = \frac{2(|\bar{n} - 2a_2| - 1)^2}{\bar{n}} \qquad (18.121)$$

18.77 Dve skupini, vsaka po $\bar{n} = 50$ delavcev, sta delali, prva po načinu A, druga po načinu B. Od skupno 100 delavcev je iz skupine B $a_2 = 35$ delavcev doseglo produktivnost višjo od mediane. Ali je produktivnost po drugem postopku značilno višja? Po obrazcu 18.121 dobimo

$$\chi^2 = \frac{2(50 - 2.35 - 1)^2}{50} = 14,44$$

Ker je $\chi^2_{p=0.001}^{(1)} = 10,83$ manjši kot izračunani $\chi^2 = 14,44$, sklepamo, da je drugi način značilno boljši od prvega.

MULTIPLA REGRESIJA IN KORELACIJA

18.78 Proučevanje odvisnosti rezultativnega znaka od enega samega faktorialnega znaka ima hibe iz več razlogov. Pri zelo izoliranih proučevanjih je odvisna spremenljivka odvisna samo od enega dejavnika in slučajnostnih vplivov. Običajno, posebno pri socialno-ekonomskih pojavih vpliva na rezultativni znak cel splet različnih bistvenih dejavnikov. Proučevanje odvisnosti y od enega znaka x je v tem primeru slabo, ker pustimo delovanje drugih dejavnikov vnemar, s tem da njihov vpliv proučujemo skupno s slučajnostnimi vplivi kot posamične vplive. Zaradi tega pa je tudi proučevanje odvisnosti y od x zelo zamegleno, ker so posamični vplivi večji. Zato je velikega pomena proučevanje vpliva več bistvenih dejavnikov na rezultativni znak. To dosežemo z multiplo regresijsko in korelacijsko analizo. Analogno funkcijski odvisnosti y od več neodvisnih spremenljivk zapišemo multiplo regresijsko odvisnost takole:

$$y = h(x_1, x_2 \cdots x_p, \varepsilon) \qquad (18.122)$$

(y = odvisna spremenljivka; $x_1, x_2, \ldots, x_p$ = neodvisne spremenljivke)

ali če povežemo del, ki kaže regresijsko funkcijo in posamične vplive, aditivno

$$y = f(x_1, x_2 \cdots x_p) + \varepsilon \qquad (18.123)$$

$$y' = f(x_1, x_2 \cdots x_p)$$

y' ponazarja odvisnost y od znakov $x_1, x_2 \ldots x_p$, če ne bi bilo posamičnih vplivov. Ta enačba torej geometrijsko predstavlja hiperploskev v $p+1$ dimenzionalnem prostoru. Podobno kot v dosedanji razlagi je

$$\sigma_\varepsilon^2 = \sigma_{y.1,2\ldots p}^2 = \frac{1}{N}\sum(y - y')^2 \qquad (18.124)$$

varianca zaradi posamičnih vplivov.

Linearna multipla regresija

18.79 Linearno odvisnost y od enega znaka x moremo posplošiti na linearno odvisnost od več dejavnikov

$$y = a_o + b_1x_1 + b_2x_2 + \ldots + b_px_p + \varepsilon \qquad (18.125)$$

Regresijsko premico zamenja v tem primeru regresijska hiperravnina

$$y' = a_o + b_1x_1 + b_2x_2 + \ldots + b_px_p \qquad (18.126)$$

Zaradi lažje pisave uporabljajmo tudi za prave vrednosti latinske in ne grške črke. Tudi v primeru multiple regresije poiščemo parametre po metodi najmanjših kvadratov. Preden uporabimo metodo najmanjših kvadratov, kakor smo jo uporabili pri krivuljčni regresiji, v modelu za linearno multiplo regresijo zamenjajmo osnovne znake $x_1x_2\ldots x_p$ s standardiziranimi spremenljivkami

$$z_y = \frac{y - M_y}{\sigma_y}; \quad z_1 = \frac{x_1 - M_1}{\sigma_1}; \quad z_2 = \frac{x_2 - M_2}{\sigma_2}; \quad \ldots\ z_p = \frac{x_p - M_p}{\sigma_p} \qquad (18.127)$$

Kot vemo, je

$$\frac{1}{N}\sum z = M_z = 0; \quad \frac{1}{N}\sum z^2 = \sigma_z^2 = 1; \quad \frac{1}{N}\sum z_iz_j = r_{ij}, \ i \neq j \qquad (18.128)$$

Tako prenarejen regresijski model ima takole obliko

$$z_y = a_o + b_1z_1 + b_2z_2 + \ldots + b_pz_p + \varepsilon_{y.1.2\ldots p} \qquad (18.129)$$

Opozoriti moramo, da so parametri $a_o, b_1, b_2 \ldots b_p$ v obrazcu 18.129 drugi kot v 18.126, čeprav je med njimi enostavna zveza in moremo po potrebi takoj prevesti ene v druge.

Z odvajanjem izraza

$$F(a_o, b_1, b_2 \ldots b_p) = \frac{1}{N} \sum (z_y - a_o - \sum_{k=1}^{p} b_k z_k)^2 \qquad (18.130)$$

dobimo

$$\frac{\partial F}{\partial a_o} = -2\frac{1}{N} \sum (z_y - a_o - \sum b_k z_k) = 0 \qquad (18.131)$$

$$\frac{\partial F}{\partial b_j} = -2\frac{1}{N} \sum (z_y - a_o - \sum b_k z_k) z_j = 0 \qquad j = 1,2 \ldots p \qquad (18.132)$$

sistem normalnih enačb, ki imajo zaradi lastnosti standardiziranih odklonov končno obliko:

$$0 = a_o$$

$$r_{yj} = \sum_k b_k r_{kj} \qquad j = 1,2 \ldots p \qquad r_{kj} = 1, \text{ če je } k=j \qquad (18.133)$$

Če ta sistem pišemo v razviti obliki, so normalne enačbe za parametre b_j naslednje:

$$\begin{aligned} r_{y1} &= b_1 + b_2 r_{21} + \ldots + b_p r_{p1} \\ r_{y2} &= b_1 r_{12} + b_2 + \ldots + b_p r_{p2} \\ &\cdots \\ r_{yp} &= b_1 r_{1p} + b_2 r_{2p} + \ldots + b_p \end{aligned} \qquad (18.134)$$

Če definiramo vektorja r in b in matriko korelacijskih koeficientov R

$$r = \begin{pmatrix} r_{y1} \\ r_{y2} \\ \cdot \\ \cdot \\ \cdot \\ r_{yp} \end{pmatrix} \quad ; \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_p \end{pmatrix} \quad ; \quad R = \begin{pmatrix} 1 & r_{21} & \cdots & r_{p1} \\ r_{12} & 1 & \cdots & r_{p2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ r_{1p} & r_{2p} & \cdots & 1 \end{pmatrix} \tag{18.135}$$

moremo sistem normalnih enačb napisati v matrični obliki

$$r = Rb \tag{18.136}$$

Iz tega je dalje

$$b = R^{-1} r \tag{18.137}$$

Če temu sistemu dodamo še enačbo regresijske ravnine

$$z''_y = z'b \tag{18.138}$$

pri čemer je z vektor $z = \begin{pmatrix} z_1 \\ z_2 \\ \cdot \\ z_p \end{pmatrix}$ neodvisnih spremenljivk , ima homogen sistem enačb

$$\begin{aligned} z''_y - z'b &= 0 \\ r - R\,b &= 0 \end{aligned} \tag{18.139}$$

netrivialno rešitev, če je izpolnjen pogoj

$$\begin{vmatrix} z''_y & z' \\ r & R \end{vmatrix} = 0 \tag{18.140}$$

V razviti obliki moremo to enačbo pisati:

$$\begin{vmatrix} z''_y & z_1 & z_2 & \cdots & z_p \\ r_{y1} & 1 & r_{21} & \cdots & r_{p1} \\ r_{y2} & r_{12} & 1 & \cdots & r_{p2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{yp} & r_{1p} & r_{2p} & \cdots & 1 \end{vmatrix} = 0 \qquad (18.141)$$

Če z R_{yi} označimo minorje ustrezne prvi vrsti v enačbi 18.141 dobimo, da je

$$z''_y R_{yy} + z_1 R_{y1} + z_2 R_{y2} + \cdots + z_p R_{yp} = 0 \qquad (18.142)$$

ali z''_y eksplicitno

$$z''_y = -\frac{R_{y1}}{R_{yy}} z_1 - \frac{R_{y2}}{R_{yy}} z_2 - \cdots - \frac{R_{yp}}{R_{yy}} z_p \qquad (18.143)$$

18.80 Vzemimo najenostavnejši primer linearne multiple odvisnosti, ko je y odvisen od dveh znakov x_1 in x_2.

Enačba regresijske zveze je

$$z'_y = b_1 z_1 + b_2 z_2 \qquad (18.144)$$

ustrezni vektorji b, r in matrika R pa

$$r = \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix} \quad ; \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad ; \quad R = \begin{pmatrix} 1 & r_{21} \\ r_{12} & 1 \end{pmatrix} \qquad (18.145)$$

Da bi izračunali vektor koeficientov b, moramo najprej poiskati R^{-1}

$$R^{-1} = \frac{1}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{21} & 1 \end{pmatrix} \tag{18.146}$$

$$b = R^{-1} r = \frac{1}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{21} & 1 \end{pmatrix} \cdot \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix} = \begin{pmatrix} \frac{r_{y1} - r_{12} r_{y2}}{1 - r_{12}^2} \\ \frac{r_{y2} - r_{21} r_{y1}}{1 - r_{12}^2} \end{pmatrix} \tag{18.147}$$

$$b_1 = \frac{r_{y1} - r_{12} r_{y2}}{1 - r_{12}^2} \; ; \quad b_2 = \frac{r_{y2} - r_{21} r_{y1}}{1 - r_{12}^2} \tag{18.148}$$

Regresijsko funkcijo moremo po 18.141 pisati tudi v obliki determinante:

$$\begin{vmatrix} z''_y & z_1 & z_2 \\ r_{y1} & 1 & r_{21} \\ r_{y2} & r_{12} & 1 \end{vmatrix} = 0 \tag{18.149}$$

Koeficient multiple korelacije

18.81 Z $e_{y.12...p}$ označimo skupen učinek drugih vplivov (brez linearnega dela $x_1, x_2, \ldots, x_p$) na y. Determinacijski koeficient multiple korelacije $R^2_{y.12...p}$ je po definicijski formuli za determinacijski koeficient

$$R^2_{y.12...p} = 1 - \frac{\sigma^2_{y.12..p}}{\sigma^2_y} \tag{18.150}$$

Ker je z_y standardiziran znak, je nepojasnjena varianca

$$\frac{1}{N} \sum e^2_{y.12..p} = \frac{1}{N} \sum (z_y - z''_y)^2 = \frac{\sigma^2_{y.12..p}}{\sigma^2_y} = 1 - R^2_{y.12..p} \tag{18.151}$$

tudi standardizirana.

Iz definicije determinacijskega koeficienta na splošno sledi, da je

$$z''_y = \sum b_k z_k \tag{18.152}$$

pojasnjeni del standardiziranega znaka z_y. Ker je varianca iz pojasnjenega dela z'_y

$$\sigma^2_{z''_y} = \frac{\sigma^2_{y''}}{\sigma^2_y} = R^2_{y.12..p} \tag{18.153}$$

velja, da je $\sigma^2_{z''_y}$ neposredno enaka determinacijskemu koeficientu multiple korelacije.

Če izračunamo

$$\sigma^2_{z''_y} = \frac{1}{N}\sum_1^N z^2_{y''} = \frac{1}{N}\sum_i (\sum_k b_k z_{ki})^2 = \sum_k \sum_j b_k b_j r_{kj} \tag{18.154}$$

$$k = 1,2..p$$

$$j = 1,2..p$$

Ker velja zveza 18.133 ter 18.153, dobimo

$$R^2_{y.12..p} = \sum_{k=1}^{p} b_k r_{yk} \tag{18.155}$$

V matrični obliki moremo pisati

$$R^2_{y.12..p} = r'b = r'R^{-1}r = \frac{-\begin{vmatrix} 0 & r' \\ r & R \end{vmatrix}}{|R|} \tag{18.156}$$

če upoštevamo 18.137

18.82 Če izračunamo po teh obrazcih $R^2_{y.12}$ za p=2, dobimo:

$$R^2_{y.12} = (r_{y1}, r_{y2}) \cdot \frac{1}{1 - r^2_{12}} \begin{pmatrix} 1 & -r_{21} \\ -r_{12} & 1 \end{pmatrix} \cdot \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix} = \frac{r^2_{y1} + r^2_{y2} - 2r_{12} r_{y1} r_{y2}}{1 - r^2_{12}} \tag{18.157}$$

ali

$$R^2_{y.12} = \frac{-\begin{vmatrix} 0 & r_{y1} & r_{y2} \\ r_{1y} & 1 & r_{12} \\ r_{2y} & r_{21} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} \\ r_{21} & 1 \end{vmatrix}} = \frac{r^2_{y1} + r^2_{y2} - 2\, r_{12}\, r_{y1}\, r_{y2}}{1 - r^2_{12}} \qquad (18.158)$$

PARCIALNA LINEARNA KORELACIJA

18.83 Pri proučevanju odvisnosti med socialno-ekonomskimi pojavi pogosto opazimo povezanost med pojavi, ki izvira iz posrednega vpliva drugih dejavnikov. Tako sta npr. nepismenost žensk in nepismenost moških med seboj povezana, vzrok za to povezanost pa je splet skupnih dejavnikov, ki vpliva tako na nepismenost moških kot žensk, tj. npr. splošna prosvetna raven v občini, sestava prebivalstva ipd. S parcialno korelacijo merimo odvisnost med dvema pojavoma, če prej izločimo vpliv enega ali več faktorjev, ki vplivajo na korelirana znaka. Izhajajmo iz modelov multiple linearne regresije za znaka x in y, za katera iščemo koeficient parcialne korelacije, s tem da izločimo linearen vpliv p neodvisnih spremenljivk

$$z_x = \sum_{k=1}^{p} b_{xk} z_k + e_{x.12..p} \; ; \quad z_y = \sum_{k=1}^{p} b_{yk} z_k + e_{y.12..p} \qquad (18.159)$$

Ker sta

$$e_{x.12..p} = z_x - \sum_{k=1}^{p} b_{xk} z_k$$

$$\text{in} \quad e_{y.12..p} = z_y - \sum_{k=1}^{p} b_{yk} z_k \qquad (18.160)$$

ostanka, ki ju dobimo, če izločimo iz z_x in z_y vpliv neodvisnih spremenljivk 1,2...p , je po standardnem obrazcu za računanje korelacijskega koeficienta

$$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$$

koeficient parcialne korelacije po definiciji korelacijski koeficient

$$r_{xy.12..p} = r_{e_{x.12..p} \cdot e_{y.12..p}} = \frac{Ce_{x.12..p} \cdot e_{y.12..p}}{\sigma e_{x.12..p}\, \sigma e_{x.12..p}} \qquad (18.161)$$

med $e_{x.12..p}$ in $e_{y.12..p}$

Kot vemo iz 18.151 je

$$\sigma_{x.12..p} = \sqrt{1 - R^2_{x.12..p}}\,; \quad \sigma_{y.12..p} = \sqrt{1 - R^2_{y.12..p}} \qquad (18.162)$$

enak koeficientom multiple alienacije.

$$Ce_{x.12..p},\, e_{y.12..p} = \frac{1}{N} \sum e_{x.12..p} \cdot e_{y.12..p} \qquad (18.163)$$

ker je poprečje iz $e_{x.12..p}$ in $e_{y.12..p}$ enako nič.

Iz 18.160 in 18.163 sledi

$$Ce_{x.12..p} e_{y.12..p} = \frac{1}{N} \sum (z_x - \sum b_{kx} z_k)(z_y - \sum b_{ky} z_k) =$$

$$= r_{xy} - \sum_{k=1}^{p} b_{xk}\, r_{ky} = r_{xy} - \sum_{k=1}^{p} b_{yk}\, r_{kx} \qquad (18.164)$$

Iz tega sledi, da je koeficient parcialne korelacije enak:

$$r_{xy.12..p} = \frac{r_{xy} - \sum_k b_{xk}\, r_{ky}}{\sqrt{1-R^2_{x.12..p}}\,\sqrt{1-R^2_{y.12..p}}} = \frac{r_{xy} - \sum_k b_{yk}\, r_{kx}}{\sqrt{1-R^2_{x.12..p}}\,\sqrt{1-R^2_{y.12..p}}} \qquad (18.165)$$

Ta obrazec moremo pisati v matrični simboliki takole. Vektor $b^{x,} = (b_{x1}, b_{x2} \ldots b_{xp})$

$r^{x,} = (r_{x1}, r_{x2} \ldots r_{xp})$, R je korelacijska matrika, opredeljena v 18.135

in $\overset{y}{r}_{,} = (r_{y1}, r_{y2} \ldots r_{yp})$. Če to upoštevamo, dobimo

$$r_{xy.12..p} = \frac{r_{xy} - \overset{x,}{r}\, R^{-1}\, \overset{y}{r}}{\sqrt{1 - \overset{x,}{r}\, R^{-1}\, \overset{x}{r}}\ \sqrt{1 - \overset{y,}{r}\, R^{-1}\, \overset{y}{r}}} = \frac{\begin{vmatrix} r_{xy} & \overset{y,}{r} \\ \overset{x}{r} & R \end{vmatrix}}{\sqrt{\begin{vmatrix} 1 & \overset{x,}{r} \\ \overset{x}{r} & R \end{vmatrix} \cdot \begin{vmatrix} 1 & \overset{y,}{r} \\ \overset{y}{r} & R \end{vmatrix}}} \qquad (18.166)$$

18.84 Če poiščemo po 18.163 obrazec za računanje parcialnega korelacijskega koeficienta $r_{xy.1}$ med x in y, pri čemer smo izločili vpliv znaka 1, so ustrezni degenerirani vektorji $\overset{x}{r}$ in $\overset{y}{r}$ in matrika R^{-1}:

$$\overset{x}{r} = (r_{x1}) \; ; \quad \overset{y}{r} = (r_{y1}) \; ; \quad R = (1) \; ; \quad R^{-1} = (1) \, .$$

Iz tega dobimo po obrazcu 18.166

$$r_{xy.1} = \frac{r_{xy} - r_{x1}\, r_{y1}}{\sqrt{1 - r_{x1}^2}\ \sqrt{1 - r_{y1}^2}} \qquad (18.167)$$

18.85 Obrazcu 18.167 podoben, a splošnejši izraz dobimo, če izhajamo iz znakov, iz katerih smo že izločili vpliv enega znaka, npr. 2. Pri tem so ustrezni linearni modeli

$$z_{x.2} = r_{x1.2}\, z_{1.2} + e_{x.12} \; ; \quad z_{y.2} = r_{y1.2}\, z_{1.2} + e_{y.12} \qquad (18.168)$$

Če postopamo analogno kot v prejšnjem primeru, dobimo, da je

$$r_{xy.12} = \frac{r_{xy.2} - r_{x1.2}\, r_{y1.2}}{\sqrt{1 - r_{x\,1.2}^2}\ \sqrt{1 - r_{y1.2}^2}} \qquad (28.169)$$

Obrazec je podoben obrazcu 18.167. V njem je le iz vseh korelacijskih koeficientov izločen znak 2. Z njim dobimo parcialni korelacijski koeficient, z dvema izločenima faktorjema (12) iz parcialnih korelacijskih koeficientov z enim izločenim

faktorjem (2). 2 more pomeniti tudi simbol za množico S faktorjev. Po tem vidimo, da moremo parcialni korelacijski koeficient z s+1 izločenimi faktorji izračunati iz parcialnih korelacijskih koeficientov $r_{xy.S}$, $r_{x1.S}$, $r_{y1.S}$, ki imajo izločenih le s faktorjev.

Partkorelacijski koeficient

18.86 Multiplo linearno regresijsko zvezo moremo v standardiziranih znakih pisati tudi v obliki

$$z_y = A_1 z_1 + A_{2.1} z_{2.1} + \ldots + A_{k.12\ldots k-1} z_{k.12\ldots k-1} + A_{p.12.(p-1)} z_{p.12.(p-1)} + \\ + A_{y.12..p} z_{y.12\ldots p} \qquad (18.170)$$

Ta enačba je zapisana tako, da je iz vsake naslednje neodvisne spremenljivke izločen linearni vpliv predhodnih. Npr. $z_{k.12.(k-1)}$ je zapis standardizirane spremenljivke k, iz katere je izločen 1,2 (k-1). Zato so vsi $z_{k.12.(k-1)}$ med seboj ortogonalni. Iz tega sledi, da je

$$\frac{1}{N} \sum z_{k.12..(k-1)} z_{j,12..(j-1)} = \begin{matrix} 0 & k \neq j \\ 1 & k = j \end{matrix} \qquad (18.171)$$

Sistem normalnih enačb je za ta primer

$$\frac{1}{N} \sum z_y z_1 = r_{y1} = {}_y r_1 = A_1$$

$$\frac{1}{N} \sum z_y z_{2.1} = {}_y r_{2.1} = A_{2.1}$$

$$\frac{1}{N} \sum z_y z_{k.1.2..(k-1)} = {}_y r_{k.12..(k-1)} = A_{k.12..(k-1)}$$

$$\frac{1}{N} \sum z_y z_{p.12..(p-1)} = {}_y r_{p.12..(p-1)} = A_{p.12..(p-1)} \qquad (18.172)$$

Podobno dobimo tudi, da je

$$\frac{1}{N}\sum z_y z_{y.12..p} = A_{y.12..p} \qquad (18.173)$$

Dobljeni parametri $A_{k.12..(k-1)} = {}_{y}r_{k.12..(k-1)}$ so korelacijski koeficienti med y in k-tim znakom, iz katerega je izločen vpliv predhodnih znakov $z_{k.12.(k-1)}$. Koeficiente ${}_{y}r_{k.12..(k-1)}$ imenujemo na splošno partkorelacijske koeficiente za razliko od parcialnih korelacijskih koeficientov $r_{yk.12..(k-1)}$ iz katerih je vpliv znakov 1,2,...k-1 izločen iz obeh koreliranih znakov.

Če pogledamo še pomen parametra $A_{y.12..p}$, spoznamo, da je

$$A_{y.12..p} z_{y.12..p} = e_{y.12..p}$$

Po obrazcu 18.151 je

$$\frac{1}{N}\sum e^2_{y.12..p} = 1 - R^2_{y.12..p}, \text{ in dalje}$$

$$\frac{1}{N}\sum (A_{y.12..p} z_{y.12..p})^2 = A^2_{y.12..p} \cdot \frac{1}{N}\sum z^2_{y.12..p} = A^2_{y.12..p}$$

Iz tega sledi

$$A_{y.12.p} = \sqrt{1 - R^2_{y.12..p}} \qquad (18.174)$$

Glede na 18.170 je

$$z_y - A_1 z_1 - A_{2.1} z_{2.1} - A_{(k-1).12..(k-2)} z_{(k-1).12..(k-2)} = \qquad (18.175)$$
$$= z_{y.12..(k-1)} \sqrt{1 - R^2_{y.12..(k-1)}}$$

Če to enačbo pomnožimo $z_{k.12..(k-1)}$, seštejemo in delimo z N, dobimo

$$\frac{1}{N}\sum z_y z_{k.12..(k-1)} = \frac{1}{N}\sum z_{y.12..(k-1)} z_{k.12..(k-1)} \sqrt{1 - R^2_{y.12..(k-1)}} \qquad (18.176)$$

Zaradi ortogonalnosti vsi drugi izrazi odpadejo. Zaradi 18.172 dobimo iz 18.174

$$r_{y.12..(k-1)} = r_{yk.12..(k-1)} \sqrt{1 - R^2_{y.12..(k-1)}} \qquad (18.177)$$

Sestava multiple korelacije je nazorno podana s partkorelacijskimi koeficienti. Ker je

$$z''_y = A_1 z_1 + A_{2.1} z_{2.1} + \ldots\ldots A_{p.12..(p-1)} z_{p.12..(p-1)} \qquad (18.178)$$

je determinacijski koeficient multiple korelacije

$$R^2_{y.12..p} = \frac{1}{N}\sum z''^2_y = \frac{1}{N}\sum\left[A_1 z_1 + A_{2.1} z_{2.1} + \ldots + A_{p.12..(p-1)} z_{p.12..(p-1)}\right]^2 =$$

$$= A_1^2 + A_{2.1}^2 + \ldots A^2_{p.12..(p-1)} \qquad (18.179)$$

ker so $z_{k.12..(k-1)}$ med seboj ortonormirani. Tako dobimo dalje, da je po 18.172 in 18.179

$$R^2_{y.12..p} = r^2_{y1} + r^2_{y2.1} + r^2_{y3.12} + \ldots + r^2_{yp.12..(p-1)} = \sum_{k=1}^{p} r^2_{yk.12..(k-1)} \qquad (18.180)$$

determinacijski koeficient multiple linearne regresije je vsota kvadratov zaporednih partkorelacijskih koeficientov. Vsak dodan nov znak k prispeva k determinacijskemu koeficientu $r^2_{y.12..(k-1)}$.

Glede na pomen partkorelacijskih koeficientov, iz 18.177 moremo pisati determinacijski koeficient tudi

$$R^2_{y.12..p} = r^2_{y1} + r^2_{y2.1}(1-R^2_{y.1}) + r^2_{y3.12}(1 - R^2_{y.12}) + \ldots + r^2_{yp.12.(p-1)}(1-R^2_{y.12.(p-1)}) =$$

$$= \sum_{k=1}^{p} r^2_{yk.12..(k-1)}(1 - R^2_{y.12..(k-1)}) \qquad (18.181)$$

Pomen obeh obrazcev je velik, ker je z njimi podana sestava determinacijskega koeficienta. Vsak nadaljnji člen pove, koliko je netoprispevek dodanega znaka k pojasnit-

vi variance odvisnega znaka.

18.87 Iz zveze 18.180 dobimo zlahka obrazec

$$y r^2_{1.H} = R^2_{y.H1} - R^2_{y.H} \qquad (18.182)$$

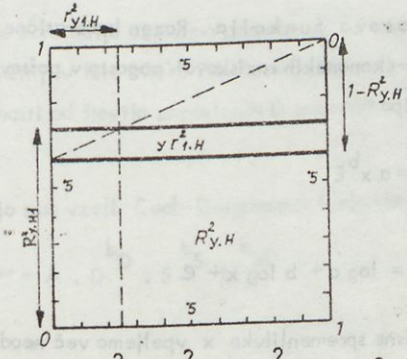
po katerem je partkorelacijski determinacijski koeficient enak neposrednemu povečanju determinacijskega koeficienta, če smo kompleksu neodvisnih dejavnikov H dodali še dejavnik 1.

Iz zveze 18.181 pa dobimo prek

$$r^2_{y1.H}(1 - R^2_{y.H}) = R^2_{y.H1} - R^2_{y.H}$$

$$r^2_{y1.H} = \frac{R^2_{y.H1} - R^2_{y.H}}{1 - R^2_{y.H}} \qquad (18.183)$$

po katerem je determinacijski koeficient parcialne korelacije enak razmerju med dodano pojasnjenostjo na račun dejavnika 1 in preostalim delom variance po izločitvi kompleksa dejavnikov H. Grafično so ponazorjeni odnosi med determinacijskimi koeficienti multiple korelacije, partkorelacije in parcialne korelacije v shemi 18.12. Celotna varianca, ki je v relativnem enaka 1, je ponazorjena s kvadratom s stranico 1. Ta kvadrat je v skladu z odnosi med raziskovanimi pokazovalci razdeljen v ustrezne pravokotnike, katerih ploščina je v sorazmerju z njihovo velikostjo.



Slika 18.12 Odnosi med $R^2_{y.H}$, $R^2_{y.H1}$, $y r^2_{1.H}$ in $r^2_{y1.H}$

18.87 Krivuljčna parabolična regresija drugega reda. Kot smo pri odvisnosti y od enega znaka razširili linearno regresijo na parabolično

$$y = a + b\,x + c\,x^2 + e \tag{18.184}$$

moremo definirati tudi kvadratično multiplo regresijo

$$y = a + \sum_{k=1}^{p} b_k x_k + \sum_{k=1}^{p} \sum_{j \le k} c_{kj}\, x_k x_j + e \tag{18.185}$$

s $\binom{p+2}{2}$ parametri

V posebnem primeru, če je p=2, dobimo

$$y = a + b_1 x_1 + b_2 x_2 + c_{11} x_1^2 + c_{12} x_1 x_2 + c_{22} x_2^2 + e \tag{18.186}$$

Prav tako moremo razširiti npr. tudi regresijo $y = a + b\sqrt{x} + c\,x + e$

$$y = a + b_1\sqrt{x_1} + b_2\sqrt{x_2} + c_{11} x_1 + c_{12}\sqrt{x_1 \cdot x_2} + c_{22} x_2 + e$$

Parametre in koeficient multiple korelacije računamo po analognem postopku kot je nakazano za linearno multiplo regresijo oziroma za krivuljčno odvisnost.

18.88 Coob-Douglasova funkcija. Razen kvadratične multiple regresije pride v socialno-ekonomskih raziskavah pogosto v poštev naravna razširitev parabolične regresije tipa

$$y = a\,x^b E \tag{18.187}$$

oziroma

$$\log y = \log a + b \log x + e \tag{18.188}$$

Če namesto ene neodvisne spremenljivke x vpeljemo več neodvisnih znakov $x_1, x_2 \ldots x_p$, za katere predpostavimo, da so enako vezani z y, dobimo

$$y = a \cdot x_1^{b_1} x_2^{b_2} \ldots x_p^{b_p} \cdot E \tag{18.189}$$

ali v implicitni obliki

$$\log y = \log a + b_1 \log x_1 + b_2 \log x_2 \ldots + b_p \log x_p + e \qquad (18.190)$$

Z logaritemsko transformacijo vseh znakov privedemo nakazano regresijo, ki je znana pod imenom Coob-Douglasova funkcija v multiplo linearno regresijo logaritmov in jo na ta način tudi naprej obravnavamo.

Pomen regresijskih koeficientov b_k postane jasen, če poiščemo parcialni diferencialni kvocient po x_k iz regresijske funkcije:

$$y' = a\, x_1^{b_1}\, x_2^{b_2} \ldots x_k^{b_k} \ldots x_p^{b_p} \qquad (18.191)$$

$$\frac{\partial y'}{\partial x_k} = b_k \cdot a\, x_1^{b_1}\, x_2^{b_2} \ldots x_k^{b_k-1} \ldots x_p^{b_p} = b_k \frac{y'}{x_k} \qquad (18.192)$$

Iz tega spoznamo, da je

$$b_k = \frac{\partial y'/y'}{\partial x_k/x_k} \qquad (18.193)$$

parcialni prožnostni koeficient, ki pove, za koliko se relativno spremeni v poprečju y', če se x_k spremeni relativno za enoto, npr. en odstotek, če so drugi x_k nespremenjeni.

18.89 Za zgled multiple korelacije vzemimo podatke za 55 bombažnih predilnic in tkalnic v SR Sloveniji v letu 1961. Proučevali smo proizvodno funkcijo odvisnosti produkta P v odvisnosti od števila zaposlenih D osnovnih sredstev S in prometnega davka M.

Kot produkcijsko funkcijo smo vzeli Coob-Douglasovo funkcijo, po kateri je

$$P' = A \cdot D^{b_D} \cdot S^{b_S} M^{b_M}$$

To funkcijo moremo privesti v obliko, ki je primerna za proučitev, z logaritmiranjem

$$\log P' = \log A + b_D \log D + b_s \log S + b_M \log M$$

Če v teh enačbah označimo: $\log P' - \overline{\log P} = p'$; $\log D - \overline{\log D} = d$;

$\log S - \overline{\log S} = s$ in $\log M - \overline{\log M} = m$, je

$$p' = b_D d + b_S s + b_M m$$

p' je v linearni zvezi z d, s in m.

Za to zvezo izračunana matrika kovarianc je:

c_{ij}

	p	d	s	m
p	·30064	·21612	·27011	·36050
d	·21612	·18793	·21575	·25780
s	·27011	·21515	·36094	·27980
m	·36050	·25780	·27980	·54890

Iz nje izračunamo po obrazcu, ki je analogen 18.141, regresijsko funkcijo

$$\begin{vmatrix} p' & d & s & m \\ \cdot 21612 & \cdot 18793 & \cdot 21575 & \cdot 25780 \\ \cdot 27011 & \cdot 21515 & \cdot 36094 & \cdot 27980 \\ \cdot 36050 & \cdot 25780 & \cdot 27980 & \cdot 54890 \end{vmatrix} = 0$$

ki je dana v eksplicitni obliki transformiranih spremenljivk

$$p' = \cdot 39243\, d + \cdot 24391\, s + \cdot 34812\, m$$

ali v eksplicitni obliki originalnih spremenljivk

$$P' = A \cdot D^{\cdot 39243} \cdot S^{\cdot 24391} \cdot M^{\cdot 34812}$$

Iz matrike kovarianc c_{ij} dobimo matriko enostavnih korelacijskih koeficientov r_{ij} po obrazcu

$$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii} \cdot c_{jj}}}$$

r_{ij}	p	d	s	m
p	1	.90923	.81997	.88743
d	.90923	1	.82839	.80267
s	.81997	.82839	1	.62861
m	.88743	.80267	.62861	1

Iz gornjih podatkov dobimo, da je po obrazcu 18.156 determinacijski koeficient multiple korelacije, izračunan iz logaritmov enak $R^2_{p.dsm} = .91868$. Glede na to, da je s številom zaposlenih, osnovnimi sredstvi in prometnim davkom pojasnjen velik delež v variabilnosti proizvodnje (92%), sklepamo, da so odnosi med dejavniki in proizvodnjo taki, da je pri enaki relativni spremembi vseh treh dejavnikov relativna sprememba proizvodnje enaka relativni spremembi dejavnikov. Če izračunamo iz regresijske enačbe diferencial

$$dp' = .39243 \cdot dd + .24391 \cdot ds + .34812 \cdot dm$$

se izkaže, da so rezultati skladni s pričakovanjem. Če je $dz = dd = dm = df$, je

$$dp' = (.39243 + .24391 + .34812)\, df = .98446 \cdot df$$

Vsota regresijskih koeficientov je resnično blizu ena, kolikor bi morala biti teoretično vsota, če bi bila relativna sprememba proizvodnje resnično sorazmerna s spremembo dejavnikov.

Če uporabimo zveze 18.182 in 18.183 ter razvrstimo neodvisne znake po velikosti stopnje odvisnosti, dobimo:

$r^2_{pd} = .82670$ $R^2_{p.dm} = .89655$ $R^2_{p.dms} = .91868$

$$p r^2_{m.d} = .89655 - .82670 = .06985 \qquad p r^2_{s.dm} = .91868 - .89655 = .02213$$

$$r^2_{pm.d} = \frac{.06985}{1-.82670} = .40306 \qquad r^2_{ps.dm} = \frac{.02213}{1-.89655} = .21392$$

Iz teh rezultatov sklepamo, da je s številom zaposlenih pojasnjeno 82,7% skupne variabilnosti v proizvodnji. Če vključimo še prometni davek, je z d in m pojasnjeno skupno 89.7% skupne variance ali: z dodanim m se je pojasnjen del povečal za 7.0%, kar je 40,3% nepojasnjenega dela variance. Z d, m in s je pojasnjen 91,9% skupne variabilnosti v p ali: s tem, da smo v regresijo vključili še s, se je pojasnjen del variabilnosti p povečal za 2,2% od skupne variance ali 21,4% nepojasnjenega dela variance. Ti odnosi so v odstotkih izraženi determinacijski koeficienti.

DEVETNAJSTO POGLAVJE

STATISTIČNO PREVERJANJE KAKOVOSTI

19.1 O s n o v a . Neposredna in kompleksna uporaba statističnih metod o preskušanju domnev je statistično preverjanje kakovosti (SPK). Neglede na to, da moremo marsikatero dosedaj obravnavano metodo za preskušanje domnev uporabiti pri problemih v zvezi s preverjanjem kakovosti (preskušanje standardov, razlik v postopkih z analizo variance) se je ožje SPK izkristaliziralo v samostojno disciplino za uporabo statističnih metod pri preverjanju kakovosti.

SPK se je posebej v množični in avtomatizirani proizvodnji pokazalo ne samo kot uporabno, temveč v nekaterih primerih celo za neobhodno. Popolno ali "stoodstotno preverjanje" je pogosto neuporabno iz več razlogov. Dolgotrajnost popolnega preverjanja zavira proizvodni proces in prevzem blaga, Monotonost pri preverjanju ima za posledico prepuste pri preverjanju. Za primere, v katerih je preverjanje povezano s kvarom ali uničenjem izdelka, pa je vzorčno preverjanje edini možni način. SPK zagotavlja željeno kakovost, hitrost izvedbe, racionalnost, možnost ustreznega načrta glede na potrebe in predpise o kakovosti preverjanja, ki so dani s parametri načrta. Zato je SPK v večini primerov izpodrinilo popolno preverjanje kakovosti in druge načine preverjanja kakovosti z neobjektivnimi osnovami.

19.2 V r s t e SPK. SPK vsebinsko in metodološko delimo v dva samostojna dela: p r e v e r j a n j e p r o c e s o v in p r e v e r j a n j e s k u p i n . Ker preverjanje skupin običajno uporabljamo pri prevzemu skupin blaga, ga imenujemo tudi p r e -

verjanje pri prevzemu.

Čeprav je večina postopkov SPK izdelana predvsem za uporabo v množični industrijski proizvodnji, jih s pridom uporabljamo tudi na najrazličnejših drugih področjih: pri preverjanju pisarniškega poslovanja, pri preverjanju kakovosti dela pri mehanografski obdelavi, pri preverjanju kakovosti dela v posameznih fazah pri statistični obdelavi, pri kontroli procesov in prevzemu v blagovnem prometu ipd.

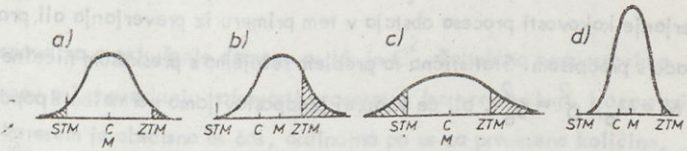
19.3 Statistično preverjanje procesov SPP. S statističnim preverjanjem procesov kontroliramo, ali določen proces poteka po predpisu. Proizvodnja oziroma značilnosti proizvedenih izdelkov variirajo iz najrazličnejših vzrokov. Te variacije pa morejo biti rezultat slučajnostnih ali pa bistvenih dejavnikov. Slučajnostni dejavniki so nehomogenost surovin, neenakomerno delo strojev ali delavcev ipd.

Iz slučajnostnih vzrokov se značilnosti izdelkov odklanjajo od poprečja v nepredvideni smeri in velikosti, vendar v skladu z zakonitostmi, ki so pogojene s kakovostjo dela strojev, predvideno neenovitostjo surovin, dela delavcev ipd.

Razen slučajnostnih vzrokov pa se morejo pojaviti v procesu tudi sistematični dejavniki, ki s svojim trenutnim ali daljšim delovanjem povzroče, da se značilnosti izdelkov odklanjajo od predpisa. Proces štejemo za stabilen, če nanj delujejo samo slučajnostni vplivi. Proces more biti stabilen tudi v primeru, če ne teče po predpisu. Stroj more zaradi bistveno slabše surovine izdelovati izdelke, za katere trdnost bistveno odstopa od predpisa. Podobno more stroj zaradi okvare na stroju začeti izdelovati izdelke, za katere je variabilnost za določen znak bistveno povečana. Iz različnih bistvenih vzrokov se more proces spremeniti naenkrat ali spreminjati zvezno.

19.4 Odnosi med tolerančnim razmikom, porazdelitvijo znaka proizvodnje in deležem neuporabnih izdelkov za numeričen znak. Če predpostavimo, da se numeričen znak za izdelek v procesu porazdeljuje normalno, morejo nastopiti v primerjavi s tolerančnim razmikom, v kate-

rem so izdelki uporabni, različne situacije. Ob vsakem trenutku med proizvodnim procesom si moremo predstavljati umišljen stabilen proces, ki bi potekal ob nespremenjenih pogojih. V odnosu na tolerančni meji morejo nastopiti situacije kakor so prikazane v sliki 19.1.



	a)	b)	c)	d)
Poprečje:	usklajeno	neusklajeno	usklajeno	neusklajeno
variabilnost:	usklajena	usklajena	neusklajena	neusklajena

Slika 19.1 Odnosi med porazdelitvijo in tolerančnim razmikom za znak proizvodnje.

Slike a do d v sliki 19.1 kažejo nekaj značilnih in pogostih situacij, v katerih se more najti proizvodni proces. a) kaže porazdelitev za proces, ki poteka v redu, ker sta tako poprečje - izraz splošnih pogojev in variabilnost - izraz posamičnih pogojev taka, da je izmet v dopustni količini. V situaciji b) je poprečje neusklajeno, ker odstopa od sredine tolerančnega razmika. Posledica je prevelik izmet. Vzrok, da je proces decentriran, je običajno v napačnem naravnanju stroja, ne pa v kakovosti dela stroja in se da taka napaka zlahka odpraviti. Situacija, ki je nakazana v c pa velja za delo stroja, ki je centriran na poprečje, prevelika pa je variabilnost, kar povzroči povečan izmet. Vzrok za povečano variabilnost more razen iz drugih vzrokov izvirati tudi iz izrabljenosti stroja in zahteva večja popravila. V d se decentriranost procesa in variabilnost izravnata, tako da je izmet v dopustnih mejah, vendar je situacija taka, da bi z istim strojem, če bi bilo poprečje usklajeno s sredino tolerančnega razmika, mogli zadostiti ostrejšim zahtevam za proces. V delovnem procesu je kakovost dela določena s porazdelitvijo znaka ob nespremenjenih pogojih. S spremembo procesa pa se spremeni tudi porazdelitev znaka proizvodnje.

19.5 Za procese običajno predpostavljamo, da se porazdeljujejo numerični znaki normalno

$$y =: N(M, \sigma^2) \qquad (19.1)$$

Če proces poteka v skladu s predpisom stabilnosti $\sigma = \sigma_0$ in $M = M_0$, se y porazdeljuje

$$y =: N(M_0; \sigma_0^2) \qquad (19.2)$$

Preverjanje kakovosti procesa obstaja v tem primeru iz preverjanja ali proces poteka v skladu s predpisom. Statistično ta problem rešujemo s preskusom ničelne domneve $H_0 : M = M_0$, $\sigma^2 = \sigma_0^2$, ali če vnaprej predpostavljamo normalnost populacije, s

$$H_0 : y = N(M_0; \sigma_0^2)$$

Ker se moreta M in σ^2 spremeniti v kateremkoli trenutku v toku procesa, bi morali teoretično preskušati gornjo ničelno domnevo zvezno prek vsega proučevanega časovnega razmika, če bi hoteli odkriti spremembe v stabilnosti procesa. Zvezno preskušanje pa ni niti mogoče niti potrebno. Ali proces poteka v skladu s predpisi ali ne, preverjamo z občasnimi vzorčnimi preskusi. Ti so pogostejši, če je verjetnost, da se proces spremeni v nestabilnega, večja, in v primeru, če je škoda, ki je povezana z nepravilnim potekom procesa, velika. Vzorčne preskuse pa vršimo redkeje, če je omenjena nevarnost manj verjetna.

Domnevo $H_0 : y =: N(M_0, \sigma_0^2)$ preskušamo s preskušanjem domnev o centralni tendenci in o variabilnosti.

19.6 Kontrolne karte. $\bar{y}$ - s karta. Domnevi: $H_0 : M = M_0$ in $H_0 : \sigma^2 = \sigma_0^2$ iz vzorca n meritev $y_1, y_2 \ldots y_n$ preskušamo prek preskusnih izrazov

$$z = \frac{\bar{y} - M_0}{\sigma_0} \sqrt{n} \; ; \; \chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \; ; \; z_p \; ; \; \chi^2_p(m=n-1) \qquad (19.3)$$

Če veljata ničelni domnevi o poprečju in varianci, veljata znani verjetnostni neenačbi

$$Pr(\bar{y}_s = M_o - z_p \frac{\sigma_o}{\sqrt{n}} < \bar{y} < \bar{y}_z = M_o + z_p \frac{\sigma_o}{\sqrt{n}}) = 1 - 2P = 1 - \alpha$$

$$Pr(s^2 < s_z^2 = \frac{\sigma_o^2 \chi_p^2}{m-1}) = 1 - P = 1 - \alpha \qquad (19.4)$$

ki služita kot osnova za preskušanje domnev o M in σ^2. Tehnično poenostavimo preskušanje domnev pri preverjanju kakovosti procesov s kontrolnimi kartami. V grafikonu, v katerem je abscisna os čas, ordinatna pa os za preskusne količine, so vrisane kritične meje. Te razdelijo območje neznačilnih razlik od območij značilnih razlik. Preskusne vrednosti npr. $\bar{y}$ ali s vnašamo v grafikon. Če je preskusna količina npr. $\bar{y}$ ali s, izračunana iz preskusnega vzorca, v območju neznačilnih razlik, procesa ne ustavimo, ker ni razloga za sklep, da proces ne teče v skladu s predpisom. Če pa v grafikonu pade preskusna količina v pas značilnih razlik s tveganjem α, ki je v skladu z napako I. vrste, sklepamo, da proces glede na proučevan parameter ni v skladu z ničelno domnevo, da proces teče po predpisu in proces ustavimo.

19.7 Zgled za $\bar{y}$ karto. Vzemimo za zgled, da s preskusnim vzorcem $n = 16$ izdelkov preverjamo delo stroja, za katerega je $H_o: M = M_o = 10$; $\sigma = \sigma_o = 1$. Kritični meji za $\bar{y}$ s preskusnimi vzorci s tveganjem $\alpha = 0{,}05$ sta naslednji:

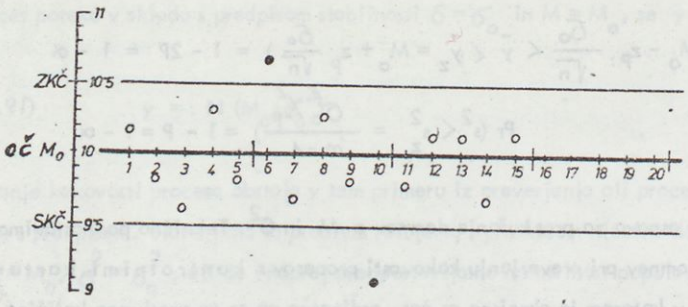
zgornja kontrolna črta ZKČ: $\bar{y}_z = M_o + z_p \frac{\sigma_o}{\sqrt{n}} = 10 + 1{,}96 \frac{1}{\sqrt{16}} = 10{,}49$

osrednja črta OČ: $M_o = 10$

spodnja kontrolna črta SKČ: $\bar{y}_s = M_o - z_p \frac{\sigma_o}{\sqrt{n}} = 10 - 1{,}96 \frac{1}{\sqrt{16}} = 9{,}51$

Kritični meji v kontrolni karti imenujemo kontrolni črti. Če iz teh podatkov sestavimo kontrolno karto, dobimo grafikon 19.2

V kontrolni karti so s krogci "o" vrisani podatki iz kontrolnih vzorcev. Dokler so točke za poprečja iz kontrolnih vzorcev v razmiku neznačilnih razlik med SKČ in ZKČ, sodimo, da poteka proces v redu, če pa pade o izven tega razmika v kri-



Slika 19.2 Kontrolna karta za poprečje

tično območje, moramo proces ustaviti in poiskati vzrok za značilno odstopanje vzorčnega poprečja. Preskusne vzorce sestavljajo skupine izdelkov, ki jih glede na načrt preverjanja izbiramo na daljše ali krajše časovne razmike: npr. vsake pol ure, vsako uro, desetkrat v izmeni ob slučajnostno izbranih trenutkih ipd.

Iz slike 19.2 moremo presoditi, da je kontrolna karta odkrila pri 6. vzorcu preveliko, pri 10. vzorcu pa premajhno aritmetično sredino.

19.8 Zgled za s-karto. Iz dosedanje razlage je razvidno, da je pri preverjanju kakovosti za kompleksno preverjanje procesa potrebno preverjati istočasno poprečje in variabilnost. Tako sodi h kompleksnemu preverjanju procesa razen kontrolne karte za $\bar{y}$ tudi kontrolna karta za s. Ker se moramo s preverjanjem variabilnosti zavarovati predvsem pred povečanjem variabilnosti, običajno pri preverjanju variabilnosti procesa uporabimo enostranski preskus in iščemo samo zgornjo kontrolno črto. Iz obrazca 19.4 najdemo zgornjo kontrolno črto za s karto po obrazcu

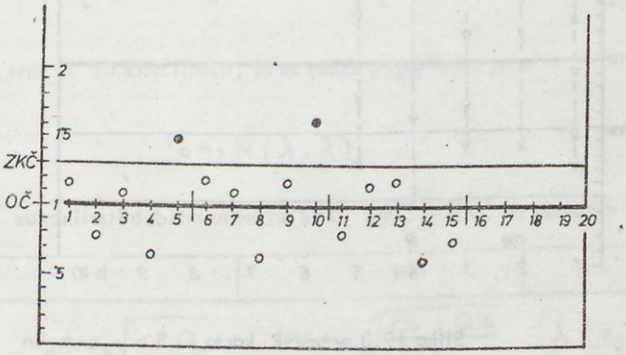
zgornja kontrolna črta ZKČ: $s_z = \sqrt{\dfrac{\sigma_o^2 \chi_p^2 (n-1)}{n-1}}$

osrednja črta OČ: σ_o

Za naš zgled je ZKČ za s s tveganjem $\alpha = 0,05$

$$s_z = \sqrt{\frac{1^2 \cdot 25}{16-1}} = 1{,}29$$

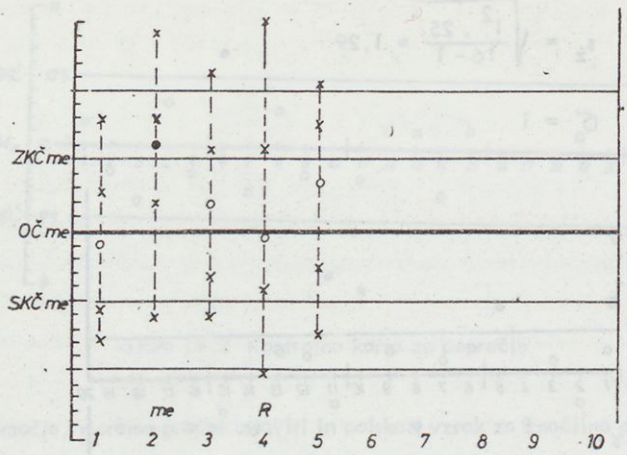
$$\sigma_o = 1$$



Slika 19.3 Kontrolna s- karta za standardni odklon

Iz s karte je razvidno, da je s značilno prevelik za 5. in 10. vzorec. Glede na to, da so vzroki, zaradi katerih se spremeni poprečje, drugi, kakor vzroki, zaradi katerih se spremeni variabilnost, moremo iz kombinirane $\bar{y}$-s karte, v kateri podajamo poprečja in standardne odklone iz istih vzorcev, sklepati tudi na vzrok, zaradi katerega proces ni stabilen.

19.9. y-me-R k a r t a . Ker je računanje $\bar{y}$ in s pri tekočem preverjanju proizvodnega procesa razmeroma zamotano, pogosto $\bar{y}$-s karto zamenjamo z y-me-R k a r t o posamičnih vrednosti, mediane me in variacijskega razmika R. V tej karti je preskusna količina za centralno težnjo mediana me iz vzorca, kot preskusna količina za variabilnost pa variacijski razmik R iz preskusnih vzorcev. Karto me-R uspešno povezujemo s karto posamičnih vrednosti y iz preskusnih vzorcev, ker moremo iz grafikona posamičnih vrednosti neposredno- brez računanja odčitati me in R.



Slika 19.3 y-me-R karta

V y-me-R kontrolni karti je vrisanih po n=5 posamičnih vrednosti (y) iz preskusnih vzorcev. Točka o za srednjo vrednost je me, razlika med skrajnima pa variacijski razmik R. Za me je v kontrolni karti območje neznačilnih razlik razmejeno od območja značilnih razlik s kontrolnima črtama SKČ in ZKČ, medtem ko je variacijski razmik grafično prikazan iz dveh delov. S črticami --- je prikazan neznačilen, s črto ——— pa značilen del variacijskega razmika. Če je ves variacijski razmik črtkan, je za ta vzorec neznačilen, če pa vsebuje tudi sklenjen del, pa je variacijski razmik značilno prevelik. Tako opazimo, da centralna tendenca kaže značilen premik v drugem, variabilnost pa v četrtem vzorcu.

19.10 Kontrolne karte za atributivne znake. V proizvodnem procesu so pogosto predmet kontrole tudi atributivni znaki. To so pravi atributi ali pa izvedeni iz enega ali več različnih znakov. Pri kalibraciji je npr. izdelek uporaben, če je premer v določenem razmiku, v nasprotnem primeru pa je neuporaben. Ker se za majhne vzorce število enot z dano značilnostjo v vzorcih porazdeljuje v binomski ali Poissonovi porazdelitvi, če gre za redke pojave, je preverjanje atributov za majhne vzorce precej neučinkovito.

Za preskusne vzorce z večjim številom enot po vemo, da se število enot z dano značilnostjo d ali strukturni delež $p = d/n$ za P, ki je zadosti velik, porazdeljujeta asimptotično normalno

$$d \doteq : N\left[nP;\ nP(1-P)\right] \quad ; \quad p\% \doteq : N\left[P\%\ ;\ \frac{P\%\ (100-P\%)}{n}\right] \tag{19.5}$$

Število enot c za značilnosti, ki se redko pojavljajo, pa v

$$c =: N(\lambda, \lambda) \tag{19.6}$$

Iz take zakonitosti dobimo ustrezne SKČ, OKČ in ZKČ po obrazcih

	d	p%	c
ZKČ	$nP_o + z_p\sqrt{n P_o Q_o}$	$P\% + z_p\sqrt{\frac{P\% \cdot Q\%}{n}}$	$\lambda + z_p\sqrt{\lambda}$
OČ	nP_o	$P\%$	λ
SKČ	$nP_o - z_p\sqrt{nP_oQ_o}$	$P\% - z_p\sqrt{\frac{P\%\ Q\%}{n}}$	$\lambda - z_p\sqrt{\lambda}$

(19.7)

Prave vrednosti P_o, P% in λ običajno zamenjamo s poprečji iz več vzorcev pri stabilnem procesu.

19.11 Normalna in poostrena kontrola z_p v kontrolnih kartah je odvisen od stopnje tveganja $\alpha = 2P$, ki je v tem, da zavrnemo ničelno domnevo, da proces teče v skladu s predpisom, kljub temu, da je tako. Pri večji stopnji tvegan ja je verjetnost za napako I. vrste večja in z njo tudi verjetnost, da neupravičeno ustavljamo proces in iščemo vzroke, ki jih ni. Pri manjši stopnji tveganja pa se pas neznačilnih razlik razširi in je preverjanje manj občutljivo za razlike. Temu se ognemo tako, da uvedemo več pasov kritičnih območij, v katerem uvedemo različne režime za ukrepe.

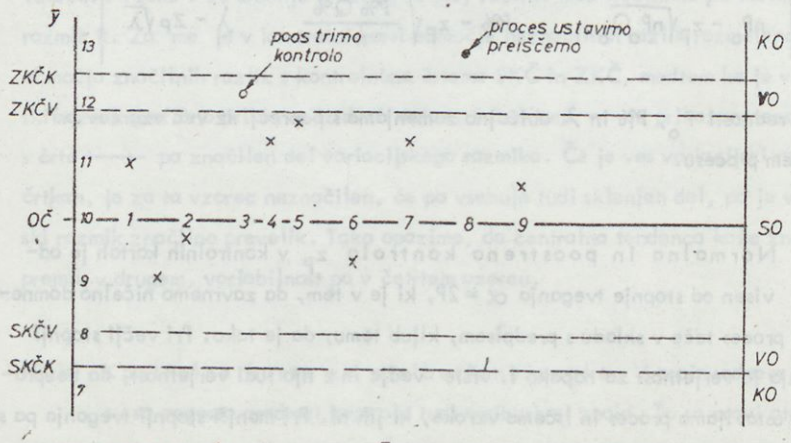
Prvi pas - območje stabilnosti procesa je pas, v katerem so razlike neznačilne tudi

pri večjem tveganju α. V drugem-varnostnem pasu so razlike značilne z večjim tveganjem npr. $\alpha = 0,05$, niso pa značilne na npr. $\alpha = 0,01$. V tretjem-kontrolnem pasu pa so razlike značilne z manjšim tveganjem npr. $\alpha = 0,01$. Ukrepi, ki sledijo, so različni v vsakem izmed naznačenih treh pasov.

območje	značilnost razlik	sklep: ukrep
I. območje stabilnosti	neznačilne	sodimo, da je proces stabilen: proces nadaljujemo
II. varnostno območje	značilne z $\alpha = 0,05$	sum na nestabilnost poostrena kontrola
III. kontrolno območje	značilne z $\alpha = 0,01$	velika verjetnost na nestabilnost: proces ustavimo in poiščemo vzrok

Opomba: $z(\alpha = 2P = 0,05) = 1,96$ v ameriških standardih zamenjamo z 2,

$z(\alpha = 2P = 0,01) = 2.58$ pa s 3.



Slika 19.4 Kontrolna karta za $\bar{y}$ z normalno in poostreno kontrolo

19.12 Preverjanje deleža neustreznih izdelkov P% z $\bar{y} - s^2$ kontrolno karto. Za numerične znake o proizvodnji so pogosto dane tolerančne meje. Glede na določeno značilnost razmejujemo izdelke z ustrezno kakovostjo od izdelkov z neustrezno kakovostjo. Tolerančni razmik je bodisi zaprt (omejen s spodnjo in zgornjo tolerančno mejo), pogosto pa je tolerančni razmik od-

odprt in je meja uporabnosti dana z eno samo tolerančno mejo, nad katero ali pod katero mejo so izdelki neuporabni. Zaprte ali odprte tolerančne razmike imamo glede na lastnost izdelka. Za trdnost je npr. dana le spodnja meja, pod katero ne sme biti trdnost, da je izdelek še uporaben, debelina profila pa je označena s spodnjo in zgornjo tolerančno mejo ipd.

Delež neuporabnih izdelkov oziroma izmet P% je v konkretnem primeru odvisen od frekvenčne porazdelitve za proučevan znak y in od tolerančnih mej. Če predpostavimo, da se y porazdeljuje v normalni porazdelitvi, je P% odvisen od aritmetične sredine M_y in standardnega odklona σ_y. S spreminjanjem teh dveh parametrov se spreminja tudi P%. V naslednjem odstavku bomo proučili, koliko je delež neustreznih izdelkov P% občutljiv za spremembe v aritmetični sredini M in standardnem odklonu σ in kolikšna je moč preskusa za odkrivanje neustreznih izdelkov iz razlik v aritmetični sredini in standardnem odklonu.

Vzemimo, da je y_T tolerančna meja in štejemo, da so izdelki neuporabni, če je $y > y_T$ in uporabni, če je $y \leq y_T$. Pri tem sodimo, da je proces stabilen, če je delež izmeta manjši kot P_o. Če upoštevamo zakonitosti o normalno porazdeljenih količinah, moremo napisati:

$$y_T = M_o - z_{P_o}\sigma$$
$$y_T = M - z_P\sigma \qquad (19.8)$$

Poleg že znane vrednosti y_T pomeni: M_o = aritmetična sredina, ki ustreza deležu neuporabnih izdelkov P_o; z_{P_o} = standardiziran odklon za normalno porazdelitev, ki ustreza dopustnemu deležu izmeta P_o, σ = standardni odklon, za katerega v prvem delu predpostavimo, da se ne spreminja; M = dejanska aritmetična sredina; z_P = standardiziran odklon, ki ustreza dejanskemu deležu neustreznih izdelkov P.

Če analogno sestavimo enačbi, ki ustrezata za aritmetične sredine iz vzorcev, dobimo za kritično mejo $\bar{y}_s$ dve enačbi: prvo za domnevno, drugo pa za dejansko stanje

$$\bar{y}_s = M_o - z_\alpha \sigma/\sqrt{n} \quad ;; \quad \bar{y}_s = M - z_{1-\beta}\sigma/\sqrt{n} \qquad (19.9)$$

Razen že znanih količin je: z_α = standardiziran odklon, ki ustreza tveganju I. vrste α; $z_{1-\beta}$ = standardiziran odklon, ki ustreza moči preskusa $1-\beta$, n pa je število enot v preskusnem vzorcu. Če enačbi 19.8 odštejemo in isto napravimo z enačbama 19.9, dobimo

$$M - M_o = (z_p - z_{p_o})\,\sigma$$
$$M - M_o = (z_{1-\beta} - z_\alpha)\,\sigma/\sqrt{n} \qquad (19.10)$$

Če dobljeni enačbi delimo eno z drugo, dobimo

$$z_p - z_{p_o} = (z_{1-\beta} - z_\alpha)\sqrt{n} \qquad (19.11)$$

Obrazec 19.11 podaja prek z posredno zvezo med deleži neustreznih izdelkov P in P_o, verjetnostjo za napako I. vrste α in močjo preskusa $1-\beta$ za kontrolo aritmetične sredine s poprečjem $\bar{y}$.

Z enačbo 19.11 moremo proučevati različne odnose. Predvsem pa je ta enačba pomembna zato, ker je osnova za OKK za preskus deležev neustreznih izdelkov iz kontrolne karte za aritmetično sredino. Iz enačbe 19.11 namreč dobimo

$$z_{1-\beta} = z_\alpha + (z_p - z_{p_o})\sqrt{n} \qquad (19.12)$$

ali

$$1 - \beta = P_r\left[z = z_\alpha + (z_p - z_{p_o})\sqrt{n}\right] \qquad (19.13)$$

19.13 Vzemimo za zgled kontrolno karto za aritmetično sredino z vzorci z n=25 enotami. Predpišimo verjetnost za napako I. vrste: $\alpha = 0{,}05$. Največji dopusten delež neustreznih izdelkov predpišemo s $P_o = 2\%$.
Iz obrazca 19.13 moremo sestaviti OKK za odstotek neustreznih izdelkov prek $\bar{y}$ - karte, ne da bi poznali parametre procesa M in σ.

Tabela 19.1 OKK za P% iz $\bar{y}$-kontrolne karte

$z_{\alpha=0.05} = 1,6449$ $\qquad$ $z_{P_o = 0,02} = 2,0537$

P%	$z_{P\%}$	$z_{1-\beta} = z_\alpha + (z_P - z_{P_o})\sqrt{n}$	β
2	2,0537	1,6449	,950
3	1,8808	,7804	,782
4	1,7507	,1299	,552
5	1,6449	- ,3991	,345
6	1,5548	- ,8496	,198
7	1,4758	- 1,2446	,107
8	1,4051	- 1,5981	,055
9	1,3408	- 1,9196	,027
10	1,2816	- 2,2156	,013

19.14 Potrebna velikost vzorca za odkrivanje razlik od P% = 2% z $\bar{y}$ - karto. Z enačbo 19.11 rešujemo tudi problem o velikosti vzorcev, s katerimi z določenim tveganjem β odkrijemo povečanje deleža neustreznih izdelkov. Iz enačbe 19.11 dobimo, da je potrebna velikost vzorca

$$n = \left(\frac{z_{1-\beta} - z_\alpha}{z_P - z_{P_o}} \right)^2 \qquad (19.14)$$

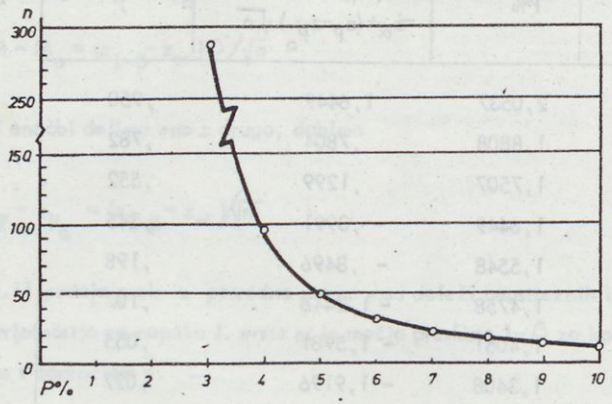
Vzemimo tveganje proizvajalca $\alpha = 0,05$, tveganje kupca pa omejimo na $\beta = 0,10$. Sprejemljiva kakovost je $P_o = 2\%$ izmeta. Za konkreten zgled dobimo iz obrazca 19.14

$$n(P_o) = \left(\frac{-1,2816 - 1,6449}{z_P - 2,0537} \right)^2 = \left(\frac{-2,9265}{z_P - 2,0537} \right)^2$$

Iz tega obrazca moremo sestaviti tabelo za odvisnost n od P

Tabela 19.2 Potrebna velikost vzorca za $\bar{y}$-karto za odkrivanje značilnih razlik od $P_0\% = 2\%$ v odvisnosti od P%.

P%	3	4	5	6	7	8	9	10
n	286	93	51	34	26	20	17	14



Slika 19.5 Potrebna velikost vzorcev za odkrivanje razlik P% od $P_0\%$ s $\bar{y}$-karto.

19.15 S podobnim sklepanjem kot za aritmetično sredino dobimo tudi zvezo med kontrolno karto za standardni odklon ali varianco in deležem izmeta P.

Če upoštevano zakonitosti o porazdeljevanju ocen za varianco, dobimo iz obrazcev za domnevno σ_0^2 in dejansko varianco σ^2

$$\frac{(n-1)s_z^2}{\sigma_0^2} = \chi_\alpha^2 \quad ; \quad \frac{(n-1)s_z^2}{\sigma^2} = \chi_{1-\beta}^2 \qquad (19.15)$$

Razen znanih količin je s_z^2 kritična vrednost za s^2 s tveganjem α. Če predpostavljamo, da se spreminja samo σ^2 ne pa M, dobimo analogno enačbama 19.8

$$y_T = M - z_{P_o}\sigma_o$$
$$y_T = M - z_P\sigma \qquad (19.16)$$

Z enostavno ureditvijo dobimo iz 19.15 in 19.16

$$\frac{\chi^2_{1-\beta}}{\chi^2_{\alpha}} = \frac{z_P^2}{z_{P_o}^2} \qquad (19.17)$$

Količine, ki nastopajo v obrazcu 19.17, so odvisne samo od tveganja α in moči preskusa $1-\beta$ za kontrolno karto za s^2, od dejanskega P% in sprejemljivega deleža izmeta P_o%. Vse druge količine, ki so prvine s^2 karte: kritična meja, domnevna varianca in stvarna varianca, pa so izločene. Podobno kot smo iz $\bar{y}$-karte sestavili OKK za P, s pomočjo obrazca 19.17 tudi iz s^2-karte sestavimo OKK za P%.

$$\chi^2_{1-\beta} = \chi^2_{\alpha}\frac{z_P^2}{z_{P_o}^2} \quad ; \quad 1-\beta = P_r\left(\chi^2 > \chi^2_{\alpha}\frac{z_P^2}{z_{P_o}^2}\right) \qquad (19.18)$$

19.16 Vzemimo za zgled enako velik vzorec n = 25 izdelkov kot za $\bar{y}$-karto, isti $\alpha = 0,05$ in enak sprejemljiv delež izmeta $P_o\% = 2\%$

$\chi^2_{(P=0,05)}(m=n-1=24) = 36,415$ $\qquad z_{(P_o\% = 2\%)} = 2,0537$

V sliki 19.6 so narisane tri OKK za odstotek izmeta v razmiku P%=2% do P%=10%. OKK-$\bar{y}$ nakazuje moč preskusa za odstotek izmeta zaradi premika aritmetične sredine, ki ga ugotavljamo z $\bar{y}$-karto.

OKK-s^2 nazakuje moč preskusa za odstotek izmeta, ki nastane zaradi povečanja variabilnosti. OKK-p pa kaže moč preskusa za odstotek izmeta, če ga kontroliramo s p oziroma c-karto. Ker je za c=1 karto za ta načrt $\alpha = 0,01324$, za c=2 karto pa $\alpha = 0,08864$, je v sliki 19.6 načrtana OKK iz poprečne verjetnosti za oba načrta.

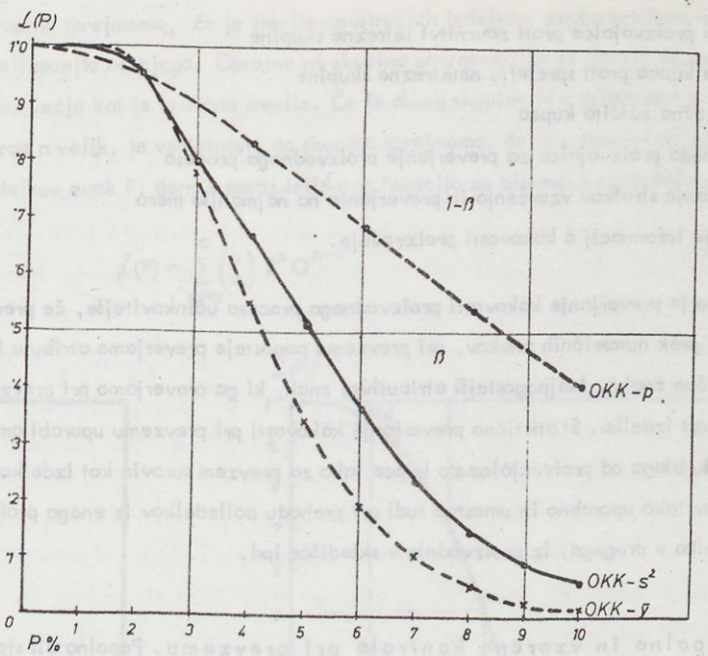
Tabela 19.2 OKK za P% iz s^2-karte

P%	z_p	$1-\beta^2=36,41\frac{z_p^2}{2,053^2}$	β
2	2,0537	36,41	,95
3	1,8808	30,54	,800
4	1,7507	26,47	,672
5	1,6449	23,36	,510
6	1,5548	20,87	,361
7	1,4758	18,80	,242
8	1,4051	17,05	,151
9	1,3408	15,52	,095
10	1,2816	14,18	,060

Tabela 19.3 OKK za P% z $\alpha = 0,05$ iz P-karte

P%	β		
	c=1	c=2	$\emptyset$
2%	.98676	.91136	.949
4%	,92353	,73582	,830
6%	,81289	,55266	,683
8%	,67684	,39473	,536
10%	,53709	,27120	,404

Slika 19.6 je zelo poučna, saj kaže, da je veliko večja verjetnost, da odkrijemo povečanje deleža izmeta s kontrolo z ustreznimi kartami za numeričen znak ($\bar{y}$ in s^2-karto) kakor pa če s tako majhnimi vzorci kontroliramo delež izmeta neposredno s c ali p-karto. Razen tega dobimo iz kombinirane $\bar{y}$-s^2 karte še dodatno informacijo o vzroku za povečanje izmeta. Iz nje namreč dobimo informacijo ali se je izmet povečal zaradi porušene stabilnosti v splošnih faktorjih ali zaradi povečanja vpliva posamičnih dejavnikov.



Slika 19.6 OKK za delež neustreznih izdelkov P% z vzorci n = 25, P_o = 2% in α = 0,05 s kontrolnimi kartami za $\bar{y}$, s^2 in p.

STATISTIČNO PREVERJANJE KAKOVOSTI SKUPIN SPS

19.17 Cilj statističnega preverjanja kakovosti v proizvodnem procesu je odkrivanje in odklanjanje vzrokov nepravilnega poteka procesa med samim procesom.

Cilj statističnega preverjanja pri prevzemu pa je ugotoviti ustreznost kakovosti skupin že izdelanih izdelkov, polizdelkov ipd. Statistična ali vzorčna kontrola pri prevzemu ima nasproti popolnemu preverjanju niz prednosti. Pravilno načrtovana in izvedena vzorčna kontrola v poprečju v velikem številu pregledanih skupin zagotavlja predpisano kakovost po opravljeni kontroli.

Značilnosti dobrega načrta kontrole pri sprejemu so:

1. zaščita proizvajalca proti zavrnitvi ustrezne skupine
2. zaščita kupca proti sprejetju neustrezne skupine
3. dolgoročna zaščita kupca
4. vzpodbuda proizvajalca za preverjanje proizvodnega procesa
5. zmanjšanje stroškov vzorčenja in preverjanja na najmanjšo mero
6. zbiranje informacij o kakovosti proizvodnje.

Medtem ko je preverjanje kakovosti proizvodnega procesa učinkovitejše, če preverjamo proces prek numeričnih znakov, pri prevzemu pogosteje preverjamo atribute kot pa numerične znake. Najpogostejši atributivni znak, ki ga preverjamo pri prevzemu, je ustreznost izdelka. Statistično preverjanje kakovosti pri prevzemu uporabljamo pri prehodu blaga od proizvajalca do kupca tako za prevzem surovin kot izdelkov. SPS je prav tako uporabno in umestno tudi pri prehodu polizdelkov iz enega proizvodnega oddelka v drugega, iz proizvodnje v skladišče ipd.

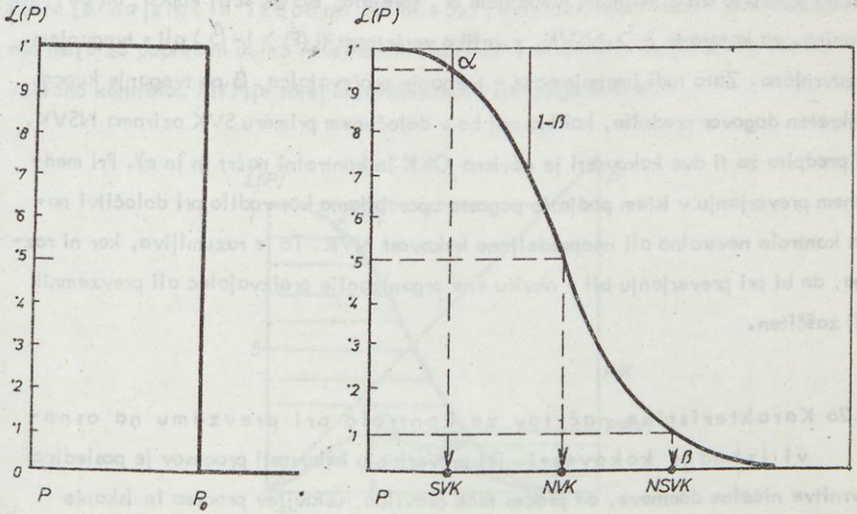
19.18 Popolna in vzorčna kontrola pri prevzemu. Popolna ali stoodstotna kontrola pri prevzemu predvideva pregled vsakega izdelka posebej. Za vsak pregledani izdelek se posebej odločimo ali ga sprejmemo kot ustreznega ali pa ga zavrnemo kot neustreznega. Običajno neustrezen izdelek pri popolni kontroli zamenjamo z ustreznim, tako da po opravljeni kontroli idealno ni neustreznih izdelkov. Če uporabimo popolno preverjanje pri sprejemanju skupin izdelkov, skupino sprejmemo, če je delež neustreznih izdelkov P v skupini enako ali pod dopustno mejo P_o in zavrnemo, če je delež neustreznih izdelkov večji kot je dopustna meja P_o. Medtem ko ustrezno skupino kupec sprejme, je v vsakem konkretnem primeru posebej predpisano, kaj napraviti z neustreznimi skupinami. Odločitev more terjati zamenjavo neustreznih izdelkov z ustreznimi v celoti ali do dopustne meje P_o, popust pri ceni za prevzem skupine itd.

OKK, ki kaže verjetnost, da skupno sprejmemo, če je delež neustreznih izdelkov P, je nakazana za popolno kontrolo v sliki 19.7a.

Pri vzorčni kontroli pa za razliko od popolne kontrole sklepamo o ustreznosti ali neustreznosti skupine izdelkov na osnovi slučajnostnega vzorca n izdelkov. Po na-

črtu skupino sprejmemo, če je število neustreznih izdelkov enako kritičnemu številu c ali manjše od njega. Obratno pa skupino zavrnemo, če je število neustreznih izdelkov večje kot je kritično število. Če je obseg skupine N v primerjavi z velikostjo vzorca n velik, je verjetnost, da skupino sprejmemo, če je v skupini delež neustreznih izdelkov enak P, dana s porazdelitveno funkcijo za binomsko porazdelitev

$$\mathcal{L}(P) = \sum_{x=0}^{c} \binom{n}{x} P^x Q^{n-x} \qquad (19.19)$$



a) popolno preverjanje b) vzorčno preverjanje

Slika 19.7 OKK pri preverjanju kakovosti pri prevzemu

Verjetnost, da z vzorčno kontrolo skupino sprejmemo, če je P=0, je enaka L(P=0)=1, če se P veča, pa L(P) monotono pada in se asimptotično približuje L(P=1) = 0.

19.19 Karakteristike načrtov za kontrolo pri prevzemu na osnovi vhodne kakovosti. Verjetnost L(P) za sprejem skupine v odvisnosti od P je podana v OKK v sliki 19.7b. Na OKK za verjetnost sprejema skupine so naznačene tri značilne točke za vhodno kakovost. Vhodna kakovost je kakovost,

dana z deležem neustreznih izdelkov v skupini pred preverjanjem. Delež P, pri katerem skupino z veliko verjetnostjo $(1-\alpha)$ sprejmemo, imenujemo "sprejemljiva vhodna kakovost" -SVK. Delež P, pri katerem skupino z veliko verjetnostjo $(1-\beta)$ zavrnemo, imenujemo "nesprejemljiva vhodna kakovost" - NSVK. Delež P, pri katerem je enaka verjetnost, da skupino sprejmemo ali zavrnemo imenujemo "nevtralna ali neopredeljena vhodna kakovost - NVK. Te tri vrste vhodne kakovosti so pomembne karakteristike za načrte pri prevzemu. Medtem ko SVK ščiti proizvajalca, da bo skupina, v kateri je $P < SVK$, z veliko verjetnostjo $(L(P) \geq 1-\alpha)$, ali z majhnim tveganjem α, sprejeta, NSVK ščiti kupca, da bo skupnina, za katero je $P > NSVK$, z veliko verjetnost $(L(P) \geq 1-\beta)$ ali s tveganjem β zavrnjena. Zato tudi imenujemo α - tveganje proizvajalca, β pa tveganje kupca. Konkreten dogovor predpiše, kakšna naj bo v določenem primeru SVK oziroma NSVK. Od predpisa za ti dve kakovosti je odvisna OKK in kontrolni načrt (n in c). Pri medfaznem preverjanju v istem podjetju pogosto uporabljamo kot vodilo pri določitvi načrta kontrole nevtralno ali neopredeljeno kakovost NVK. To je razumljivo, ker ni razloga, da bi pri preverjanju bil v okviru ene organizacije proizvajalec ali prevzemnik bolj zaščiten.

19.2o Karakteristike načrtov za kontrolo pri prevzemu na osnovi izhodne kakovosti. Pri preverjanju kakovosti procesov je posledica zavrnitve ničeine domneve, da proces teče pravilno, ustavitev procesa in iskanje vzrokov. Zavrnitev skupine pri preverjanju kakovosti pri prevzemu pa more imeti različne posledice. Zavrnjeno skupino moremo predati prevzemniku z določenim popustom ali drugo ugodnostjo, s katero se izravna ugotovljena slabša kakovost. Običajno pa za zavrnjeno skupino pregledamo vse izdelke v skupini in neustrezne izdelke zamenjamo z ustreznimi. Tako je z vzorčno kontrolo zavrnjena skupina predana prevzemniku po opravljenem preverjanju teoretično brez neustreznih izdelkov.

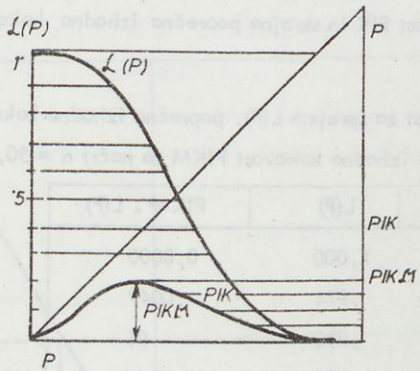
Če je vhodna kakovost VK enaka P, je v poprečju L(P) skupin sprejetih in $1 - L(P)$ zavrnjenih. Izhodna kakovost IK za sprejete skupine je P, izhodna kakovost za zavrnjene skupine pa P=0. Poprečna ali pričakovana izhodna kakovost PIK je enaka

$$PIK = E(P) = P \cdot L(P) + 0 \cdot [1 - L(P)] = P \cdot L(P) \qquad (19.20)$$

Če je vhodna kakovost VK dobra, je pričakovana izhodna kakovost PIK tudi dobra, ker je PIK majhen. V skrajnem, če je P=0, je PIK = 0. Če pa je za vhodno kakovost VK P velik, je veliko skupin zavrnjenih in v njih slabi izdelki zamenjani z dobrimi. Za velike P je PIK majhen, ker je L(P) majhen. Pri določenem P je PIK največji oziroma poprečna izhodna kakovost najslabša.

$$PIK_{max} = PIKM \qquad (19.21)$$

PIKM je najslabša izhodna kakovost. Pri določenem načrtu je PIKM zgornja meja za poprečen delež neustreznih izdelkov v skupinah, če jih preverjamo z vzorčno kontrolo. PIKM je torej karakteristika vzorčnega načrta.



Slika 19.8. Poprečna izhodna kakovost PIK in najslabša poprečna izhodna kakovost PIKM

19.20 Izdelani načrti se ravnajo po velikosti skupin N, vhodni kakovosti VK, ki je dana s poprečnim deležem neustreznih izdelkov in uvedenimi karakteristikami: sprejemljiva vhodna kakovost, SVK, neopredeljena vhodna kakovost NVK, nesprejemljiva vhodna kakovost NSVK in najslabša poprečna izhodna kakovost PIKM. Po teh karakteristikah so izdelane najrazličnejše tablice za vzorčno kontrolo pri prevzemu. Med njimi so najbolj poznane: Dodge-Romig-ove tabele. Military Standard 105 in Philipsove tabele.

19.21 Za zgled vzemimo načrt, po katerem s preskusnim vzorcem pregledamo $n = 50$ izdelkov. Skupino sprejmimo, če v vzorcu ne najdemo nobenega ali en sam neustrezen izdelek: $c = 1$. Glede na to, da je vzorec sorazmerno z velikostjo skupin majhen, moremo privzeti, da se število neustreznih izdelkov d v vzorcih porazdeljuje v binomski porazdelitvi.

Verjetnost $L(P)$, da skupino sprejmemo, če je delež neustreznih izdelkov v skupini P, je glede na vzorčni načrt enaka

$$L(P) = \sum_{d=0}^{c} \binom{n}{d} P^d Q^{n-d} = \sum_{d=0}^{1} \binom{50}{d} P^d Q^{50-d} = \binom{50}{0} P^0 Q^{50-0} + \binom{50}{1} P^1 Q^{50-1} = Q^{50}\left[1+50\frac{P}{Q}\right]$$

Verjetnost $L(P)$, da skupino sprejmemo, je torej funkcija deleža neustreznih izdelkov v skupini. V tabeli 19.5 so nakazane verjetnosti $L(P)$ za nekatere vrednosti P, poprečna izhodna kakovost PIK in skrajna poprečna izhodna kakovost PIKM za načrt $n = 50$, $c = 1$

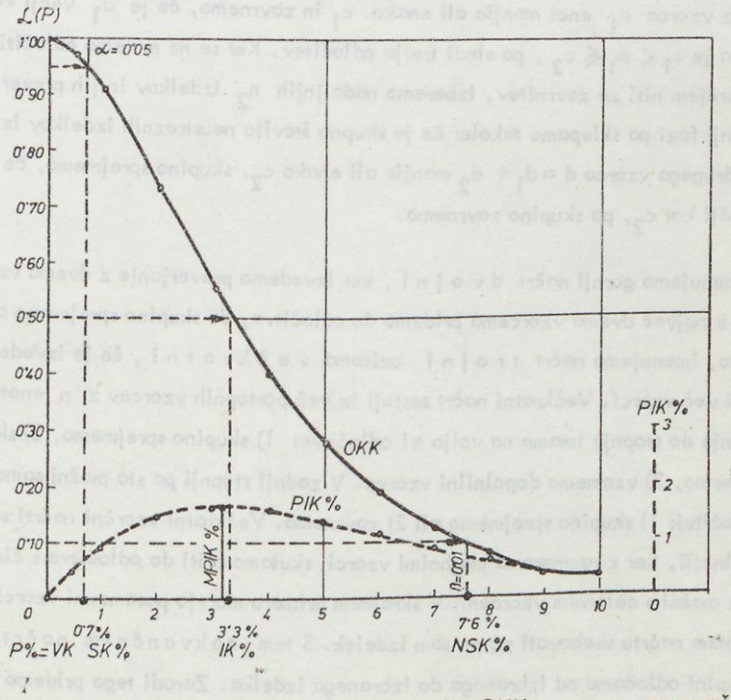
Tabela 19.5 Verjetnost za sprejem $L(P)$, poprečna izhodna kakovost PIK in skrajna poprečna izhodna kakovost PIKM za načrt $n = 50$, $c = 1$.

VK = P	L(P)	PIK=P . L(P)
0,00	1,000	0,0000
0,005	,974	,0049
0,01	,910	91
0,02	,736	147
0,03	,555	167
0,04	,400	160 ← 0,0168 = PIKM, P_{PIKM}=0,0324
0,05	,279	139
0,06	,190	114
0,07	,127	89
0,08	,083	66
0,09	,053	48
0,10	,038	38

Tabeli 19.5 so izračunani PIK za določene vrednosti VK. Za podane vrednosti P je PIK = 0,0167 največja za VK = 0,03 interpolirana PIKM = 0,0168 pa pri VK = 0,0324.

V sliki 19.9 je narisana: operativna karakteristična krivulja OKK = $L(P)$ ki kaže verjetnost za sprejem skupine v odvisnosti od VK = P: krivulja poprečne izhodne kakovosti PIK.

Na sliki 19.9 so nakazane bistvene karakteristike za načrt prevzema, ki so obenem tudi merila za izbiro načrta. Sprejemljiva vhodna kakovost SVK, ki je podana s tistim deležem neustreznih izdelkov pred preverjanjem, ki jo z dovolj veliko verjetnostjo (običajno L(P) = 0,95) oziroma z majhnim tveganjem (običajno α = 0,05) sprejmemo. V našem načrtu je grafično interpolirana vrednost za SVK = 0,007 ali 0,7%. Nasprotno je nesprejemljiva vhodna kakovost tisti delež za vhodno kakovost



Slika 19.9 Operativna karakteristična krivulja OKK za poprečno izhodno kakovost PIK in karakteristične točke za načrt prevzema n = 50, c=1

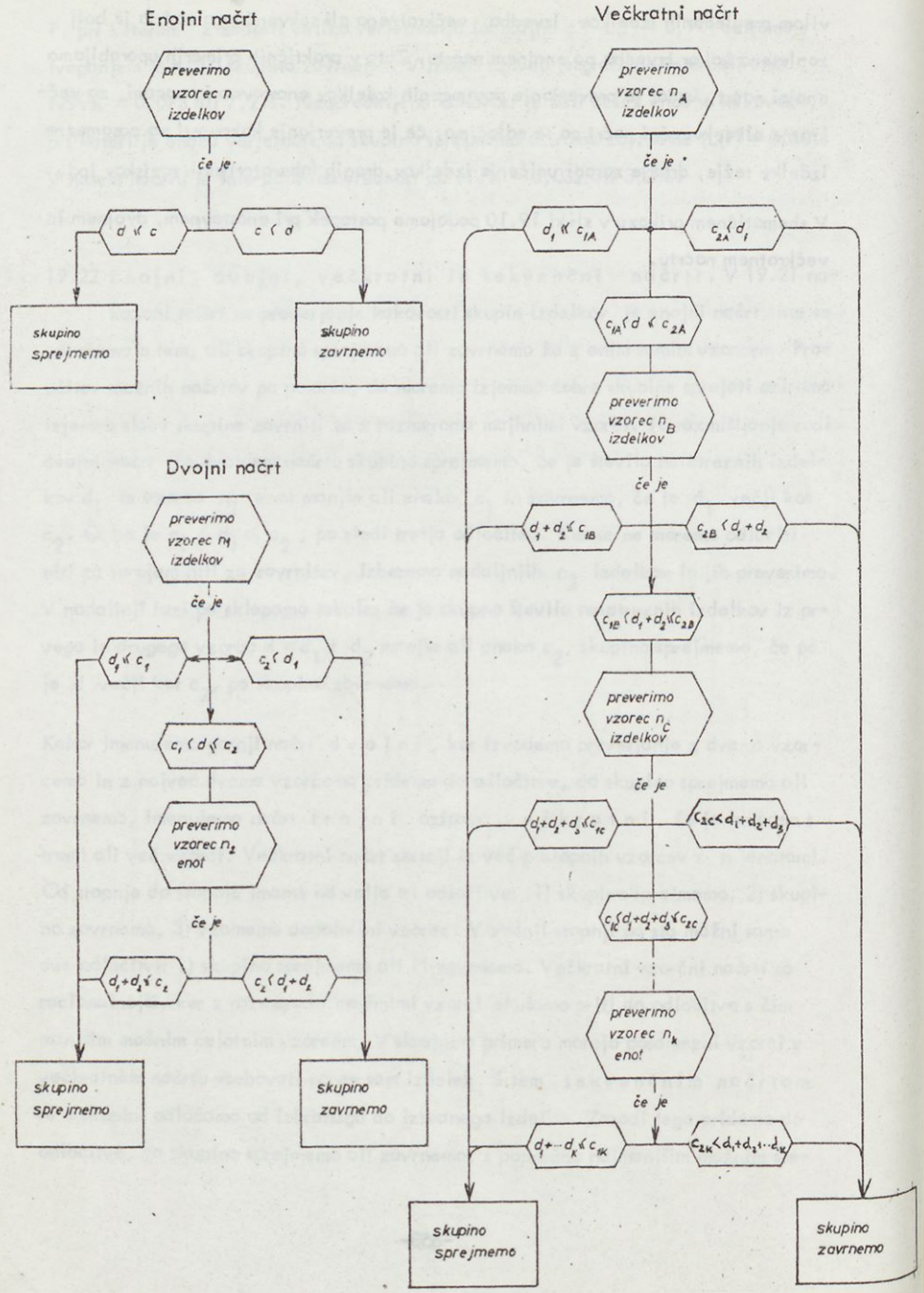
P, pri katerem z zadosti veliko verjetnostjo (običajno z 1-L(P) = 0,90) oziroma s tveganjem $\beta = .10$ skupino zavrnemo. V našem zgledu je grafično interpolirana NSVK = 0,072 ali 7,2%. Neopredeljena kakovost je tisti delež vhodne kakovosti, pri kateri je enaka verjetnost da skupino sprejmemo oziroma zavrnemo (L(P) = 0,50). V našem načrtu je interpolirana vrednost za NVK = 0,032 ali 3,2%.

19.22 Enojni, dvojni, večkratni in sekvenčni načrti. V 19.21 nakazani načrt za preverjanje kakovosti skupin izdelkov je enojni načrt, ker se odločamo o tem, ali skupino sprejmemo ali zavrnemo že z enim samim vzorcem. Proučitev možnih načrtov pa pokaže, da moremo izjemno dobre skupine sprejeti oziroma izjemno slabe skupine zavrniti že z razmeroma majhnimi vzorci. To razmišljanje rodi dvojni načrt. Po dvojnem načrtu skupino sprejmemo, če je število neustreznih izdelkov d_1 iz vzorca n_1 enot manjše ali enako c_1 in zavrnemo, če je d_1 večji kot c_2. Če pa je $c_1 < d_1 \leqslant c_2$, pa sledi tretja odločitev. Ker se ne moremo odločiti niti za sprejem niti za zavrnitev, izberemo nadaljnjih n_2 izdelkov in jih preverimo. V nadaljnji fazi pa sklepamo takole: če je skupno število neustreznih izdelkov iz prvega in drugega vzorca $d = d_1 + d_2$ manjše ali enako c_2, skupino sprejmemo, če pa je d večji kot c_2, pa skupino zavrnemo.

Kakor imenujemo gornji načrt dvojni, ker izvedemo preverjanje z dvema vzorcema in z največ dvema vzorcema pridemo do odločitve, da skupino sprejmemo ali zavrnemo, imenujemo načrt trojni oziroma večkratni, če je izveden s tremi ali več vzorci. Večkratni načrt sestoji iz več postopnih vzorcev z n enotami. Od stopnje do stopnje imamo na voljo tri odločitve: 1) skupino sprejmemo, 2) skupino zavrnemo, 3) vzamemo dopolnilni vzorec. V zadnji stopnji pa sta možni samo dve odločitvi: 1) skupino sprejmemo ali 2) zavrnemo. Večkratni vzorčni načrti so racionalnejši, ker z razmeroma majhnimi vzorci skušamo priti do odločitve s čim manjšim možnim celotnim vzorcem. V skrajnem primeru morejo posamezni vzorci v večkratnem načrtu vsebovati po en sam izdelek. S tem sekvenčnim načrtom se o skupini odločamo od izbranega do izbranega izdelka. Zaradi tega pridemo do odločitve, da skupino sprejmemo ali zavrnemo, s poprečno najmanjšim možnim šte-

vilom pregledanih izdelkov. Izvedba večkratnega ali sekvenčnega načrta je bolj zapletena kakor izvedba po enojnem načrtu. Zato v praktičnih primerih uporabljamo enojni načrt, kadar je preverjanje posameznih izdelkov enostavno in poceni, za večkratne ali sekvenčni načrt pa se odločimo, če je preverjanje kakovosti za posamezne izdelke težje, dražje zaradi uničenja izdelkov, dragih laboratorijskih raziskav ipd.

V shematičnem prikazu v sliki 19.10 podajamo postopek pri enostavnem, dvojnem in večkratnem načrtu.



Slika 19.10 Sheme za potek kontrole pri enojnem, dvojnem in večkratnem vzorčnem načrtu.

19.23 Za zgled stvarnih načrtov vzemimo tabelo iz knjige: Freeman, Friedman. Mostelier Wallis: Sampling Inspection. Iz priročnika je vzeta tabela za kontrolo skupin z N = 500-800 enotami, za primer, če pričakujemo poprečni odstotek neustreznih izdelkov po kontroli PIK od 1,2-2,2%, PIKM pa v mejah 2,5% do 3,5%.

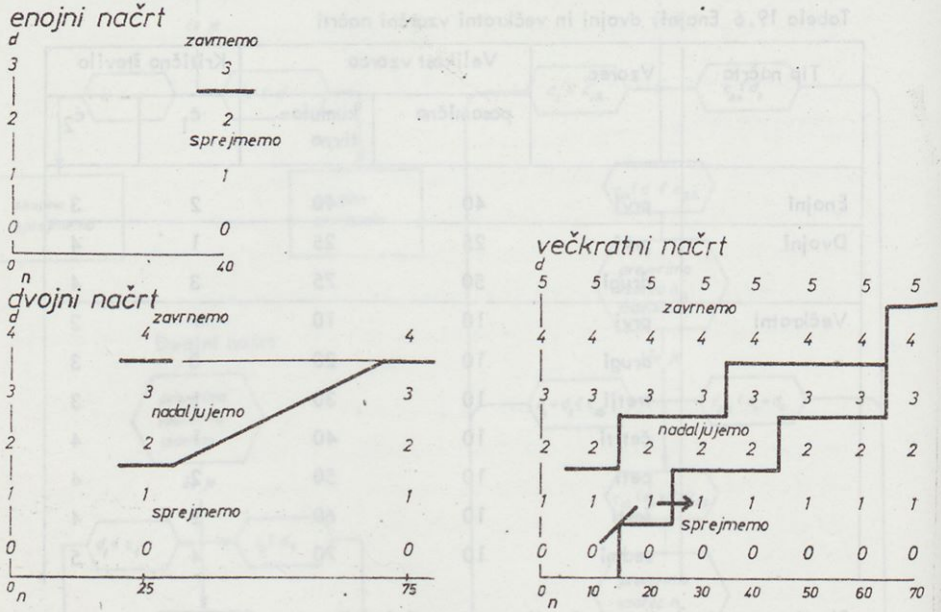
Tabela 19.6 Enojni, dvojni in večkratni vzorčni načrti

Tip načrta	Vzorec	Velikost vzorca		Kritično število	
		posamična	kumulativna	c_1	c_2
Enojni	prvi	40	40	2	3
Dvojni	prvi	25	25	1	4
	drugi	50	75	3	4
Večkratni	prvi	10	10	+	2
	drugi	10	20	0	3
	tretji	10	30	1	3
	četrti	10	40	1	4
	peti	10	50	2	4
	šesti	10	60	2	4
	sedmi	10	70	4	5

Skupino sprejmemo, če je število neustreznih izdelkov v vzorcu $d \leq c_1$ in zavrnemo, če je $d > c_2$ in vzamemo nadaljnji vzorec, če je $c_1 < d \leq c_2$; + pomeni, da pri prvem vzorcu skupine ne moremo sprejeti.

V sliki 19.11 so grafično prikazani načrti iz tabele 19.6. Iz teh grafikonov je še nazorneje vidna razdelitev v tri pasove sprejetja, nadaljevanja in zavrnitve do končnega sklepa o zavrnitvi ali sprejemu. V sliki za večkratni načrt je vrisan potek preverjanja. V prvem vzorcu $n_1 = 10$ enot nismo dobili nobenega slabega izdelka. Ker pade ta vrednost v pas nadaljevanja, smo vzeli dodaten vzorec $n_2 = 10$ enot in dobili v skupnem vzorcu en slab izdelek. Ker je tudi ta vrednost v pasu nadaljevanja, vzamemo tretji vzorec z $n_3 = 10$ enotami in dobimo skupno v celotnem vzorcu

d = 1 slab izdelek. Ker pade ta vrednost v območje sprejema, vzorčenje ustavimo in kontrolirano skupino sprejmemo.



Slika 19.11 Enojni, dvojni in večkratni načrt iz tabele 19.6

DVAJSETO POGLAVJE

20 STATISTIČNO NAČRTOVANJE POSKUSOV

20.1 Pri kvantitativnem proučevanju ekonomskih pojavov je eden izmed pomembnih problemov proučevanje vpliva dejavnikov na najrazličnejše pojave kot npr. na višino proizvodnje, produktivnost dela, kakovost proizvodov, višino prodaje ipd. Tovrstne probleme rešujemo različno. Če proučujemo celotne populacije si pomagamo s tabelarno analizo. Regresijska in korelacijska analiza sta osnovni metodi za proučevanje odvisnosti. Enako z analizo variance preskušamo značilnosti razlik med poprečji pri spremenjenih pogojih in dobimo tako rezultate o vplivu določenega dejavnika. Nakazana analiza variance pa je le zametek metodologije, po kateri moremo s kompleksno analizo variance proučevati vpliv različnih dejavnikov na socialnoekonomske pojave. Z razširitvijo analize variance uspemo varianco razstaviti na dele-sestavine, ki so odvisne od jakosti delovanja dejavnikov, ki jih proučujemo. S tem se odpirajo velike možnosti kompleksnega analiziranja vpliva dejavnikov. Ker imajo ti načini v večini primerov značaj poskusa, vse te metode združujemo v posebno disciplino statističnega proučevanja - načrtovanje poskusov. Predmet statističnega načrtovanja poskusov je sistematično proučevanje delovanja dejavnikov, teoretična osnova pa je analiza variance. Uporaba statističnega načrtovanja poskusov v socialno-ekonomskih raziskavah je mnogostranska. Poskuse izvajamo v razvojnih oddelkih proizvodnih podjetij pri proučevanju vpliva proizvodnih dejavnikov, uporabljamo jih pri statistič-

nem preverjanju kakovosti, pri študijah časa ipd. V posebnih oblikah uporabljamo metode statističnega načrtovanja poskusov pri analizah trga, raziskovanju javnega mnenja, pri proučevanju življenjske ravni ipd.

20.2 Prvine statističnega načrtovanja poskusov. Vzemimo za zgled raziskovanje proizvodnosti dela po dveh načinih. En način zaznamujemo z A_1, drugega pa z A_2.

a) Če raziskujemo proizvodnost dela tako, da samo en delavec dela po načinu A_1, eden pa po načinu A_2, razlika med doseženima rezultatoma ne da osnove za sklep, kateri način je boljši. Število opazovanj je premajhno in pridejo do izraza osebnostni vplivi obeh delavcev. Zato vzamemo v preskus večje skupine - vzorce in preskušamo razlike med poprečji. V poprečju se namreč vpliv posamičnih faktorjev omili.

b) Če vzamemo v poskus npr. 20 delavcev, posamične vplive zmanjšamo tako, da vzamemo skupino delavcev, ki je enovita - homogena glede na vrsto značilnosti, po katerih smo preskusno skupino opredelili. Izberemo npr. samo moške delavce, približno enako stare, s približno enako delovno dobo in če imamo na razpolago podatke o njihovih splošnih delovnih uspehih, vključimo v preskus delavce s približno enako splošno delovno zmogljivostjo.

c) Ko imamo izbrano enovito preskusno skupino delavcev je seveda popolnoma napačno, če v skupino, ki dela po načinu A_1, izberemo sistematično delavce, za katere kljub zgornji opredelitvi pričakujemo boljše delovne rezultate. Na prvi pogled je jasno, da bi ta sistematičnost vplivala na poprečno razliko med obema načinoma.

Razlika med preskusnima skupinama v tem primeru izvira razen od načina dela tudi iz pristranskega izbora delavcev. Da bi se izognili zavestni ali podzavestni pristranosti pri razdeljevanju enot poskusa v skupine, s slučajnostnim izborom določimo, kateri delavci bodo delali po enem, kateri pa po drugem načinu.

d) Pri poskusu se moremo odločiti, da od skupno 20 delavcev, kolikor jih sodeluje

pri poskusu, npr. 5 dela po načinu A_1, 15 pa po načinu A_2. Kratek račun pa pokaže, da ta razdelitev ni najprimernejša. Upravičeno pričakujemo, da bomo najboljše rezultate za oceno razlike dobili, če vzamemo v vsako skupino enako število delavcev, to je po 10 v vsako skupino.

Varianca za razliko med aritmetičnima sredinama dveh populacij z enakima variancama je

$$\mathrm{Var}\,\Delta\bar{y} = \sigma_y^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right] \qquad (20.1)$$

Če je skupno število enot v obeh vzorcih n, je $n_2 = n - n_1$, n_1 pa je optimalen, če je varianca minimalna. Ta pa je minimalna, če je odvod $\mathrm{Var}\,\Delta\bar{y}$ po n_1 enak 0.

$$\frac{d}{dn_1}\mathrm{Var}\,\Delta\bar{y} = \frac{d}{dn_1}\sigma_y^2\left[\frac{1}{n_1} + \frac{1}{n-n_1}\right] = \sigma_y^2\left[-\frac{1}{n_1^2} + \frac{1}{(n-n_1)^2}\right] = 0$$

Iz tega sledi: $\frac{1}{(n-n_1)^2} = \frac{1}{n_1^2}$; $n_1 = \frac{n}{2}$

Spoznamo, da je varianca med aritmetičnima sredinama ekstremna, oziroma minimalna, če je $n_1 = n/2$, oziroma če je $n_1 = n_2$.

c) Kljub slučajnostnemu dodeljevanju načina dela /A_1 in A_2/ pa je lahko osebnostna komponenta delavca še vedno velika. Ta učinek odpravimo, če predpišemo, da vsak delavec dela po obeh načinih, npr.: najprej po načinu A_1, nato pa po načinu A_2. Razlika v proizvodnosti dela $\bar{y}_{A2} - \bar{y}_{A1}$ za vsakega delavca nima več, ali vsaj ne v veliki meri, osebnostnega vpliva. Kajti, če je npr. delavec dober, se to pokaže v višji proizvodnosti dela po enem in drugem načinu.

V tem primeru pa se more pojaviti drug - v določenem primeru negativen učinek. Če vsi delavci delajo najprej po načinu A_1, nato pa po načinu A_2, more imeti tako načrtovana raziskava vpliv na višino proizvodnosti po načinu A_2. Delavci si morejo pri prvem načinu pridobiti izkušnje, ki izzovejo pri delu po načinu A_2 določen sistematičen vpliv, ki izvira iz vrstnega reda izvajanja del. V tem primeru more biti

produktivnost po načinu A_2 oziroma razlika $\bar{y}_{A_2} - \bar{y}_{A_1}$ previsoka zaradi privajenosti poslu, ki so si jo delavci pridobili pri delu po načinu A_1. Da bi postavili oba načina v enak položaj, predpišemo, da deset delavcev dela najprej po načinu A_1, nato pa po načinu A_2, druga polovica pa obratno; najprej po načinu A_2, nato pa po načinu A_1. Popročna razlika $\bar{y}_{A_2} - \bar{y}_{A_1}$ je v tem primeru predvsem izraz učinka različnega načina dela, ne pa tudi zapovrstnega reda. Ni treba posebej ponavljati, da je treba tudi v tem primeru izbrati skupini, ki delata po načrtu A_1A_2 ali A_2A_1 slučajnostno. Tako smo uspeli z načrtom poskusa odstraniti vpliv dveh dejavnikov, ki motita rezultate: osebnostno komponento delavca in privajenost delu.

Na tem enostavnem primeru smo nakazali prvine statičnega načrtovanja poskusov, ki pridejo pri nekih načrtih bolj, pri drugih pa manj do izraza.

S poskusi proučujemo vpliv različnih dejavnikov na proučevani pojav. V našem zgledu preskušamo vpliv načina dela, ki je proučevani dejavnik, na proizvodnost dela, ki jo proučujemo. V ta namen smo izbrali skupino delavcev, osnovno poskusno gradivo, ki je v splošnem sestavljeno iz poskusnih enot, na katerih uporabimo posamezne postopke, v našem primeru: način dela A_1 in način A_2. Skupnost enot poskusa naj bi bila čim bolj homogena. To smo v našem zgledu dosegli z načrtno izbiro delavcev. Katere enote so izbrane v poskus, oziroma kakšno je osnovno poskusno gradivo, je v veliki meri odvisno od cilja poskusa, ker v vsakem primeru proučujemo vpliv dejavnikov pri danih osnovnih pogojih. Drugačen je namreč vpliv činiteljev npr. pri starih ali mladih delavcih. Zato opredelitev homogenega poskusnega vzorca spada k vsebinski plati poskusa.

Važna prvina poskusa je ponavljanje istih postopkov. S ponavljanjem poskusa dosežemo večjo jasnost rezultatov. Poskusni pogrešek poprečij se s ponavljanjem poskusa manjša s kvadratnim korenom iz števila ponavljanj. Slučajnost uporabe postopkov na poskusnih enotah je nadaljnji zelo pomemben element poskusa. S tem morebitne opredeljive, vendar za poskus nepomembne dejavnike vključimo v sklop slučajnostnih vplivov, s čimer se njihov vpliv manjša, obenem pa je mogoče zaradi zakonitosti, ki veljajo za slučajnostne vplive, njihov učinek vrednotiti.

Z opredelitvijo poskusnega gradiva v veliki meri odstranimo moteč vpliv vrste činite-

ljev. Vendar pri velikem številu postopkov in ponovitvah dostikrat težko uspemo dobiti enovito poskusno gradivo. Zato z načrtom poskusa izločimo iz poskusa določene dejavnike, ki jih moremo opredeliti. V našem primeru smo uspeli izločiti osebnostni vpliv tako, da je vsak delavec delal po obeh načinih, vpliv dodatnega motečega dejavnika pa tako, da je polovica delavcev delala po sistemu A_1A_2, polovica po sistemu A_2A_1. Z enakim številom poskusov po enem in drugem načinu smo uspeli poskusni pogrešek zmanjšati na minimum in tako vpliv proučevanega činitelja podati v čim jasnejši sliki.

Nakazane prvine, ki so prvine vsega statističnega načrtovanja poskusov, so strnjene v shemi nastopajočih dejavnikov in njihovo obravnavanje v tabeli 20.1.

Tabela 20.1 Prvine statističnega načrtovanja poskusov

<table>
<tr><td rowspan="2">Dejavniki</td><td colspan="2">Raziskovani</td><td colspan="3">Drugi - moteči</td></tr>
<tr><td colspan="3">Določljivi</td><td>Nedoločljivi</td></tr>
<tr><td></td><td>več nivojev</td><td>en nivo</td><td colspan="2">več nivojev</td><td></td></tr>
<tr><td rowspan="2">Obravnavanje dejavnikov</td><td>načrt poskusa</td><td>opredeljujoči</td><td>načrt poskusa</td><td>slučajnostno dodeljevanje nivojev enotam</td><td>slučajnosten izbor poskusnih enot</td></tr>
<tr><td colspan="2">proučitev vpliva pri opredeljivih pogojih</td><td>vpliv izločen</td><td>značaj slučajnosten
s ponavljanjem se učinek zmanjša</td><td></td></tr>
</table>

20.3 Posamezni statistični načrti poskusov v večji ali manjši meri izkoriščajo v prejšnjem odstavku nakazane prvine. Z dobrim načrtom izločimo vpliv tistih nepomembnih dejavnikov, katerih učinek je največji. S kompleksnejšimi načrti izločamo po več nepomembnih činiteljev hkrati. Tako so iz poskusa izločeni moteči - sistematični vplivi, slučajnostni vplivi pa zmanjšani. S tem dosežemo, da so rezultati uporabni le z majhnim številom ponovitev.

Obravnavali bomo le najenostavnejše statistične načrte poskusov, s katerimi pa moremo vseeno reševati niz problemov iz socialno-ekonomskih področij. Ti statistični

načrti poskusov so:

- Čisto slučajnosten poskus
- Slučajnostni bloki
- Latinski kvadrat
- Dvofaktorski poskus.

Čisto slučajnostni poskus

20.4 Najenostavnejši načrt poskusa, ki pa ni tudi najboljši, je čisto slučajnostni poskus. Od prvin statističnega načrta poskusa uporabimo le homogenost poskusnega gradiva, slučajnost dodeljevanja postopkov poskusim enotam in ponavljanje.

Zgled za čisto slučajnostni poskus je raziskava o razlikah v produktivnosti dela pri treh različnih pripravah dela: P_1, P_2 in P_3. Izberemo izenačeno poskusno skupino npr. $n = \bar{n}.p = 15$ delavcev. Petim izmed na slučajnosten način izbranih delavcev priredimo, po katerem načinu bodo delali. Kot smo že nakazali, je najuspešneje, če pozamezen postopek uporabimo na enakem številu enot. V našem zgledu smo izmed $\bar{n}p = 15$ za preskus določenih delavcev po $\bar{n} = 5$ delavcem predpisali, da delajo po načinu P_1, P_2 in P_3.

Rezultati, merjeni s številom izdelkov, izdelanih v enoti časa, so naslednji:

način P_1	25	28	25	23	26
način P_2	28	27	31	29	21
način P_3	33	35	31	29	36

Če analiziramo dejavnike, ki vplivajo na posamično proizvodnost dela vsakega izmed $\bar{n}p = 15$ delavcev, sklepamo, da je proizvodnost dela y_{Pi} rezultat splošnih pogojev dela M, ki je za vse delavce enak; posebnih pogojev dela, ki izvirajo iz različnih načinov priprave dela in se menjajo od skupine do skupine (P) in posamičnih dejavnikov, ki se menjajo od delavca do delavca e_{Pi}. Če označimo z y_{Pi} produktivnost dela za delavca i, ki je delal po načinu P, moremo sestaviti

linearni model

$$y_{Pi} = M + (P) + e_{Pi} \qquad (20.2)$$

Predpostavimo, da velja ničelna domneva $H_o: P_1 = P_2 = \ldots = P_p$ in da je

$$\sum (P) = 0 \; ; \quad S_P^2 = \frac{1}{P-1} \sum P^2 \; ; \quad e_{Pi} =: N(0, \sigma_e^2) \qquad (20.3)$$

Analizo variance izvedemo po shemi v tabeli 20.2

Tabela 20.2. Analiza variance za čisto slučajnostni poskus

Vir variacije	Vsota kvadratov SK	Stopinje prostosti m	Ocena variance $s^2 = \frac{K}{m}$	Pričakovana vrednost $E(s^2)$	F
Faktor P	$K_P = q_P$	$P - 1$	$s_P^2 = \frac{K_P}{P-1}$	$\sigma_e^2 + \bar{n} S_P^2$	s_P^2 / s_e^2
Poskusni pogrešek e	$K_e = q_{Pi} - q_P$	$P(\bar{n} - 1)$	$s_e^2 = \frac{K_e}{P(\bar{n}-1)}$	σ_e^2	1
Skupaj	$K = q_{Pi}$	$n - 1$			

(20.4)

Pri tem so q_{Pi}, q_P količine, znane iz enostavne analize variance in izračunane iz osnovnih podatkov po shemi :

$$y_{Pi} \mid y_P \mid y$$

$$Q_{Pi} = \sum_{P\,i} S\, y_{Pi}^2 \; ; \quad Q_P = \frac{1}{\bar{n}} \sum y_P^2 \; ; \quad Q = \frac{1}{\bar{n}P}\, y^2$$

$$q_{Pi} = Q_{Pi} - Q \; ; \quad q_P = Q_P - Q \; ; \qquad (20.5)$$

V našem zgledu je obračun za analizo variance takle:

a) osnovni podatki

	y_{Pi}					y_P
P_1	25	28	25	23	26	127
P_2	28	27	31	29	21	136
P_3	33	35	31	29	36	164
						427 = y

b) pomožne količine:

$$Q_{Pi} = \sum_S y_{Pi}^2 = 25^2 + 28^2 \ldots + 29^2 + 36^2 = 12407$$

$$Q_P = \frac{1}{\bar{n}} \sum y_P^2 = \frac{1}{5} (127^2 + 136^2 + 164^2) = 12304{,}20$$

$$Q = \frac{1}{\bar{n} \cdot P} y^2 = \frac{1}{5.3} \cdot 427^2 = 12155{,}27$$

$$q_{Pi} = 12407 - 12155{,}27 = 251{,}73$$

$$q_P = 12304{,}20 - 12155{,}27 = 148{,}93$$

c) analiza variance:

Vir variacije	SK	m	s^2	F
priprava	$K_P = q_P = 148{,}93$	3 - 1 = 2	74,47	$8{,}69^{**}$
poskusni pogrešek	$q_{Pi} = 251{,}73$ $- q_P = -148{,}93$ $K_e = 102{,}80$	3(5-1) = 12	8,57	1
Skupno	$K = q_{Pi} = 251{,}73$	15-1 = 14		

Iz sheme o analizi variance za čisto slučajnostni poskus spoznamo, da je v primeru, če velja ničelna domneva, da ni razlik v proizvodnosti dela pri različnih pripravah ($H_o : P_1 = P_2 = P_3 = 0$), t.j. da različne priprave ne vplivajo na produktivnost dela, s_p^2 in s_e^2 oceni variance zaradi slučajnostnih vplivov σ_e^2, ker je za ta primer $S_p^2 = 0$. Zato moremo s kvocientom s_p^2/s_e^2 s F- preskusom preskušati ničelno domnevo in tako posredno značilnost razlik med učinkom različnih priprav dela na proizvodnost.

Iz sheme o analizi variance je razvidno, da je s_e^2 nepristrana ocena za σ_e^2 s $P(\bar{n} - 1)$ stopinjami prostosti. Zato jo moremo uporabiti za razmične ocene poprečij, razlik med poprečji ipd. Ker je $F_{0,01}\ (2,12) = 6,93$, $F_{0,001}\ (2,12) = 12,97$, sklepamo, da so razlike v poprečni proizvodnosti dela značilno različne na stopnji $\alpha = 0,01$, kar potrjuje osnovno domnevo, da različne priprave vplivajo na produktivnost dela ($H_1 : P_1 \neq P_2 \neq P_3$).

Največji verjeten odklon iz poskusa ocenjevanega poprečja od poprečne proizvodnosti dela ocenimo po obrazcu

$$d(\bar{y}_p) = t_\alpha (m_e) \sqrt{\frac{s_e^2}{n}} \qquad (20.6)$$

Za naš zgled je s tveganjem $\alpha = 5\ \%$

$$d(\bar{y}_p) = 2,179 \cdot \sqrt{\frac{8,57}{5}} = 2,85$$

Ocene poprečij

$$\bar{y}_{P_1} = \frac{127}{5} = 25,4\,; \quad \bar{y}_{P_2} = \frac{136}{5} = 27,2\,; \quad \bar{y}_{P_3} = \frac{164}{5} = 32,8$$

imajo isti največji verjeten odklon $d(\bar{y}) = 2,85$.

Podobno dobimo, da je največji verjeten odklon za oceno razlik med dvema ocenama poprečne produktivnosti dela

$$d(\Delta\bar{y}_p) = t_{\alpha}(m_e)\sqrt{\frac{2.s_e^2}{\bar{n}}} \qquad (20.7)$$

za naš zgled

$$d(\Delta\bar{y}_p) = 2{,}179\sqrt{\frac{2.8{,}57}{5}} = 4{,}03$$

iz česar sklepamo, da značilnost razlik izvira iz tega, da je P_3 značilno ugodnejši od P_1 in P_2.

Primerjave med aritmetičnimi sredinami, dobljenimi z analizo variance

20.5 Vnaprej načrtovane primerjave. Pri proučevanju značilnosti razlik med aritmetičnimi sredinami, ki jih dobimo iz poskusa, razlikujemo dva vsebinsko različna pristopa. V našem zgledu moremo že vnaprej (a priori) načrtovati oziroma predvideti primerjavo med aritmetičnimi sredinami poskusa. Načrtovana primerjava je zasnovana na vnaprej postavljenih domnevah o razlikah med točno določenimi sredinami. V našem zgledu bi mogla biti npr. taka vnaprej postavljena primerjava sestavljena na domnevi, da dá tretji postopek P_3, ki je nov, boljše rezultate od standardnega (prvega P_1) postopka. To domnevo preskusimo s poznanim t-preskusom

$$\frac{\bar{y}_3 - \bar{y}_1}{\sqrt{\frac{s_e^2}{\bar{n}}(1^2 + 1^2)}} = t\,(m = n - P) \qquad (20.8)$$

Ker je tako izračunani $t = \frac{32{,}8 - 25{,}4}{\sqrt{\frac{8{,}57}{5}(1 + 1)}} = +4{,}0$

večji kot je kritična vrednost $t_{P=0.005}^{(m=12)} = 3{,}06$, sklepamo, da je nov način boljši od starega s tveganjem, ki je manjše kot $\alpha = P = 0{,}005$.

Z vnaprejšnjimi primerjavami preskušamo s t-preskusom tudi druge domneve: Če šte-

jemo, da sta postopek P_1 in P_2 stara načina dela, P_3 pa nov način dela, pokaže primerjava

$$P_3 - \frac{P_1 + P_2}{2} = P_3 - \frac{1}{2}P_1 - \frac{1}{2}P_2$$

če je novi način boljši kot poprečen učinek starih načinov. Ta učinek ocenimo z izrazom

$$\bar{y}_3 - 1/2\ \bar{y}_1 - 1/2\ \bar{y}_2$$

Iz te ocene izračunani izraz

$$t = \frac{\bar{y}_3 - \frac{1}{2}\bar{y}_2 - \frac{1}{2}\bar{y}_1}{\sqrt{\frac{s_e^2}{\bar{n}}\left[1^2 + (\pm\frac{1}{2})^2 + (-\frac{1}{2})^2\right]}}$$

se v primeru ničelne domneve porazdeljuje v t-porazdelitvi z $m = m_e = n-P=12$ stopinjami prostosti

$$t = \frac{32,8 - \frac{1}{2}\,25,4 - \frac{1}{2}\,27,2}{\sqrt{\frac{8,57}{5}(1 + \frac{1}{4} + \frac{1}{4})}} = 4,05$$

Če primerjamo dobljeni izraz s kritično vrednostjo $t_{P=0.005}^{(m=12)} = 3,06$, vidimo, da je ta primerjava odkrila, da je s tveganjem, ki je manjše od $\alpha = P = 0,005$, novi način boljši kot poprečje starih dveh načinov.

Ti primerjavi pa sta samo poseben primer preskušanja značilnosti splošnejše primerjave med aritmetičnimi sredinami, po kateri se za linearno zvezo med aritmetičnimi sredinami $\sum c_k \bar{y}_k$ porazdeljuje izraz

$$\frac{\sum c_k \bar{y}_k}{\sqrt{\frac{s_e^2}{\bar{n}} \cdot \sum c_k^2}} = : t\ (m = n - P) \qquad (20.9)$$

v t-porazdelitvi z $m = n-P$ stopinjami prostosti, če razen ničelne domneve $H_o: \sum c_K P_K = 0$ velja še, da je vsota koeficientov $\sum c_k = 0$. To pa velja v našem prvem zgledu (-1 0 1) pri katerem smo primerjali novi postopek P_3 s P_1 in v drugem zgledu, za katerega enako velja, da je

$$(-1/2) + (-1/2) + (1) = 0$$

Linearne zveze $\sum c_k \bar{y}_k$, za katere imajo koeficienti c_k lastnost $\sum c_k = 0$, imenujemo primerjave.

20.6 Po izvršenem poskusu (aposteriorno) izvedene primerjave. Pogosto primerjav med aritmetičnimi sredinami, ki jih imamo s poskusom, ne načrtujemo vnaprej, temveč raziskovalec proučuje razlike med rezultati poskusa po izvršenem poskusu brez vnaprej načrtovanih domnev o primerjavah. Na splošno imamo v primeru, če je v poskusu p postopkov, $\binom{p}{2}$ možnih različnih primerjav med dvema aritmetičnima sredinama.

Na prvi pogled je jasno, da je variacijski razmik med aritmetičnimi sredinami poskusa R odvisen tudi od števila postopkov oziroma s poskusom dobljenih sredin in da je razlika med največjo in najmanjšo sredino tem večja, čim večje je število postopkov. S tem pa se tudi veča verjetnost za napako α. Tudi v primeru, če velja ničelna domneva in so sredine pri različnih postopkih enake, je pričakovati pri večjem številu postopkov med največjo in najmanjšo sredino iz poskusa razmeroma veliko razliko, ki pa je slučajnostna in izraz tega, da se pri slučajnem izboru enot iz populacije variacijski razmik veča, če se veča število postopkov. Zato moramo pri aposteriornih primerjavah o značilnosti razlik med sredinami upoštevati število proučevanih postopkov in je uporaba najmanjše značilne razlike lahko sporna, ker je zanjo α velik.

Poznamo več metod, ki upoštevajo nakazane lastnosti poprečij iz vzorcev. Od teh bomo nakazali le najenostavnejšo, ki je zasnovana na kritičnem razmerju, ki pokaže značilnost razlik med sredinami tako, da je verjetnost za napako prve vrste največ α.

To je Tukeyev preskus, ki je zasnovan na Studentiziranem razmiku q. Ta je dan s

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{s_e} \sqrt{n} \qquad (20.10)$$

Kritične vrednosti za tveganje α za največjo razliko so odvisne od tveganja α, števila stopinj prostosti m_e, iz kolikor je ocenjena s_e in od števila postopkov P v poskusu.

Tabela za $q_{\alpha=0.05}(m,p)$ in $q_{\alpha=0.01}(m,p)$ za P = 2(1) 10 je dana v dodatku.

Po Tukeyevem preskusu je z največjim tveganjem α značilna vsaka razlika med dvema sredinama v poskusu, če je

$$|\bar{y}_2 - \bar{y}_1| > q_\alpha(m,p) \frac{s_e}{\sqrt{\bar{n}}} \qquad (20.11)$$

Za P = 2 je v skladu s teorijo $q_\alpha(m,2) = t_\alpha(m)\sqrt{2}$.

Za naš zgled o produktivnosti dela za P = 3 postopke P_1, P_2 in P_3 velja za $\Delta\bar{y}_P$

	P_1	P_2	P_3	$\bar{y}_P$
$\bar{y}_P$	25,4	27,2	32,8	
$\Delta\bar{y}_P$		1,8	7,4**	25,4
			5,6*	27,2
				32,8

Ker je m = 12 in P = 3 je iz tablice dobljen $q_{0.05}(12,3) = 3,77$; $q_{0.01}(12,3) = 5,05$.

Ker je $\bar{n} = 5$, $s_e = \sqrt{8,57} = 2,93$, je kritična vrednost za razlike sredin na stopnji tveganja $\alpha = 0,05$

$$3,77 \cdot \frac{2,93}{\sqrt{5}} = 4,94$$

na stopnji $\alpha = 0,01$ pa 6,62. Izkaže se, da je razlika $\bar{y}_3 - \bar{y}_2 = 5,6$ značilna na stopnji $\alpha = 0,05$, razlika $\bar{y}_3 - \bar{y}_1 = 7,4$ pa na stopnji $\alpha = 0,01$.

Slučajnostni bloki

20.7 V prejšnjem zgledu čisto slučajnostnega poskusa smo izbrali v poskus 15 delavcev, od katerih je vsak delal po enem izmed postopkov. Naša težnja je bila, da bi bila celotna skupina delavcev, ki je bila vključena v poskus, čim bolj izenačena - homogena. Čim večji pa je obseg poskusa, tem večje število osnovnih poskusnih enot oziroma poskusov je treba. Za take primere je težko zagotoviti zadostno homogenost poskusnih enot. Heterogenost večjega poskusnega gradiva obidemo z metodo blokov.

Vzemimo kot zgled študijo, pri kateri proučujemo čas sestavljanja za štiri različne razmestitve sestavnih delov. Poskus nameravamo izvršiti v desetih ponovitvah. Za poskus s štirimi postopki v desetih ponovitvah je treba štirideset poskusnih enot. Izbrati homogeno skupino 40 sestavljalcev pa je težko. Lažje zagotovimo enovitost manjših skupin sestavljalcev, npr. tolikega števila, kolikor je postopkov v eni ponovitvi. S poznanjem sposobnosti posameznih delavcev, ki pridejo v poštev kot poskusne enote, moremo brez dvoma sestaviti deset dovolj izenačenih skupin po štiri delavce. Enovita skupina poskusnih enot, na katerih izvedemo eno ponovitev poskusa, predstavlja b l o k . Z analizo slučajnostnih blokov dosežemo, da heterogenost med bloki ne vpliva na zanesljivost poskusnih rezultatov. Z metodo blokov uspemo namreč vpliv faktorjev, ki so vir razlik med bloki, izločiti iz poskusa, s tem pa se zanesljivost dobljenih rezultatov zveča.

Če je merilo za združevanje delavcev v bloke delovna doba in poprečno doseganje norme pri podobnih delih, je pri metodi blokov vpliv teh dveh dejavnikov izločen, medtem ko bi njihov vpliv pri čisto slučajnostnem poskusu vključili med slučajnostne činitelje, s tem pa dobili večji poskusni pogrešek.

Ko imamo sestavljene bloke, poskus nadaljujemo podobno kot pri čisto slučajnostnem

Tabela 20.3 Shema o razmestitvah za poskus v slučajnostnih blokih

B_1

P_1	P_3	P_4	P_2
8	23	16	11

B_2

P_1	P_3	P_4	P_2
34	41	39	35

B_3

P_2	P_3	P_1	P_4
52	64	28	50

B_4

P_3	P_2	P_1	P_4
11	14	6	8

B_5

P_3	P_2	P_1	P_4
35	38	24	38

B_6

P_3	P_2	P_4	P_1
20	23	20	16

B_7

P_3	P_2	P_4	P_1
27	26	24	16

B_8

P_2	P_1	P_3	P_4
49	34	48	50

B_9

P_1	P_3	P_2	P_4
34	43	38	39

B_{10}

P_1	P_3	P_2	P_4
26	39	43	43

poskusu. Vsakemu delavcu iz istega bloka na slučajnosten način priredimo en postopek, tako da vsak delavec posameznega bloka dela po enem načinu. Zato je število poskusnih enot v blokih po pravilu enako številu postopkov poskusa P.

V načrtu v tabeli 20.3 so nanizani podatki o časih za sestavljanje za posamezne razmestitve delov (P_1 P_2 P_3 P_4) za bloke po štiri delavce tako, kot je potekal poskus. Zato je tudi vrstni red razmestitev za vsak blok zapisan, kot je bil delavcem zapovrstjo dodeljen s slučajnim razporedom.

V naslednji tabeli so iz prejšnjega načrta prepisani podatki v urejeni obliki, ki jo moremo shematično napisati takole:

y_{BP}	y_B
y_P	y

Tabela 20.4 Rezultati poskusa o razmestitvah

Blok	Razmestitev				y_B
	P_1	P_2	P_3	P_4	
B_1	8	11	23	16	58
B_2	34	35	41	39	149
B_3	28	52	64	50	194
B_4	6	14	11	8	39
B_5	24	38	35	38	135
B_6	16	23	20	20	79
B_7	16	26	27	24	93
B_8	34	49	48	50	181
B_9	34	38	43	39	154
B_{10}	26	43	39	43	151
y_P	226	329	351	327	1233 y
$\bar{y}_P$	22,6	32,9	35,1	32,7	

20.8 Če z y_{BP} označimo posamezen podatek za delavca iz bloka B, ki je delal po razmestitvi P, sumarni rezultat zapišemo v linearnem modelu, razstavljen na učinke posameznih faktorjev

$$y_{BP} = M + (B) + (P) + e_{BP} \qquad (20.12)$$

Pri tem je M rezultat splošnih pogojev, (B) učinek blokov, ki je v posebnih primerih B_1, B_2 ..... B_{10}, računsko pa predpostavljamo, da je $\sum(B) = 0$. (P) je učinek

proučevanega faktorja, v našem zgledu razmestitve, ki ima v našem zgledu vrednosti P_1, P_2, P_3, P_4 in zvezo $\sum(P) = 0$; e_{BP} pa je rezultat slučajnostnih vplivov, za katerega predpostavljamo, da je slučajnostna spremenljivka, ki je normalno porazdeljena z

$$e_{BR} =: N(0, \sigma_e^2) \tag{20.13}$$

Če podobno kot pri čisto slučajnostnem poskusu izračunamo iz tabel

y_{BP}	y_B
y_p	y

$$Q_{BP} = \sum_B \sum_P y_{BP}^2 ; \quad Q_B = \frac{1}{p} \sum_B y_B^2 ; \quad Q_P = \frac{1}{b} \sum_P y_P^2 ; \quad Q = \frac{1}{b.p} y^2 \tag{20.14}$$

$$q_{BP} = Q_{BP} - Q ; \quad q_B = Q_B - Q ; \quad Q_P = Q_P - Q ,$$

pri čemer razen znanih izrazov pomeni b = število blokov; p = število postopkov, dobimo v sistematični obliki shemo za analizo variance za poskus v slučajnostnih blokih v tabeli 20.5.

Tabela 20.5 Shema za analizo variance za poskus v slučajnostnih blokih (20.15)

Vir variacije	SK	m	$s^2 = \frac{SK}{m}$	$E(s^2)$	F
Bloki B	$K_B = q_B$	b - 1	s_B^2	$\sigma_e^2 + p S_B^2$	
Faktor P	$K_P = q_P$	p - 1	s_P^2	$\sigma_e^2 + b S_P^2$	$F = s_P^2 / s_e^2$
Poskusni pogrešek	$K_e = q_{BP} - q_B - q_P$	(b-1)(p-1)	s_e^2	σ_e^2	1
Skupaj	$K = q_{BP}$	bp - 1			

Pri tem je razen znanih izrazov

$$s_B^2 = \frac{1}{b-1} \sum_{B=1}^{b} B_B^2 \; ; \quad s_P^2 = \frac{1}{p-1} \sum_{P} P_P^2 \qquad (20.16)$$

Ničelna domneva, da dejavnik P ne vpliva na proučevan pojav (razmestitev ne vpliva na čas sestavljanja) ($H_o : P_1 = P_2 = P_3 = P_4 = 0$), je podana z $S_P^2 = 0$.

s_p^2 in s_e^2 sta v tem primeru dve neodvisni nepristranski oceni za varianco σ_e^2 za slučajnostno spremenljivko $e_{BP} =: N(0, \sigma_e^2)$. Zato moremo s kvocientom $s_p^2/s_e^2 = F$ izvesti F-preskus $m_1 = p-1$; $m_2 = (b-1)(p-1)$

Če z F-preskusom ugotovimo značilne razlike med aritmetičnimi sredinami, je po standardnem postopku možno analizirati ocene poprečij pri različnih postopkih in razlike med njimi. Za enostavne slučajnostne bloke je

$$d(\bar{y}_p) = t_\alpha \left[m_e = (b-1)(p-1) \right] \sqrt{\frac{s_e^2}{b}} \; ; \quad d(\Delta \bar{y}_p) = t_\alpha \left[(b-1)(p-1) \right] \sqrt{\frac{2 s_e^2}{b}} \qquad (20.17)$$

20.9 Za naš zgled dobimo iz tabele osnovnih podatkov 20.2 tele elemente za analizo variance:

$b = 10$, $p = 4$

$Q_{BP} = 8^2 + 11^2 + \ldots + 39^2 + 43^2 = 45791$; b=10; p=4 ; $q_{BP} = 45791 - 38007{,}2 = 7783{,}8$

$Q_B = \frac{1}{4}(58^2 + 149^2 + \ldots + 151^2) = 44278{,}8$; $q_B = 44278{,}8 - 38007{,}2 = 6271{,}6$

$Q_P = \frac{1}{10}(226^2 + 329^2 + 351^2 + 327^2) = 38944{,}7$; $q_P = 38944{,}7 - 38007{,}2 = 937{,}5$

$Q = \frac{1}{40}(1233)^2 = 38007{,}2$

20.6 Analiza variance za poskus o razmestitvah v čisto slučajnostnih blokih

Vir variacije	SK	m	s^2	F
Bloki	$K_B = q_B = 6271,6$	9	674,62	
Razmestitve	$K_P = q_P = 937,5$	3	312,50	14,67***
Poskusni pogrešek	$q_{BP} = 7783,8$ $-q_B = -6271,6$ $-q_P = -937,5$			
	$K_e = 574,7$	27	21,29	1,00
Skupno	$K = q_{BP} = 7783,8$	39		

Iz izračunanja analize variance spoznamo, da so bloki v tem primeru dali uspešne rezultate. Z združevanjem delavcev v bloke je uspelo pojasniti oziroma odstraniti iz skupne vsote kvadratov $K_B = 6271,6$ z $m_B = 9$ stopinjami prostosti in je ostalo kot rezultat slučajnostnih vplivov le $K_e = 574,7$ z $m_e = 27$ stopinjami prostosti. Razlike v času za sestavljanje so se pokazale kot visoko značilne, saj je izračunani $F = 14,67$ večji kot $F_{0,001}(3,27) = 7,27$, tj. kritične vrednosti z $\alpha = 0,001$. Zato je analiza poprečnih časov za posamezne razmestitve umestna.

Če vzamemo, da nismo napravili vnaprej nikakih domnev o primerjavah med sredinami, bomo značilnost razlik med razmestitvami proučili aposteriorno s Tukeyevim preskusom.

Iz tabele v dodatku dobimo z interpolacijo, da je

$$q_{0,05}(27,4) = 3,88 \qquad q_{0,01}(27,4) = 4,85$$

$$\Delta \bar{y}_{0.05} = q_{0,05}^{(27,4)} \sqrt{\frac{s_e^2}{\bar{n}}} = 3,88 \sqrt{\frac{21,29}{10}} = 5,67$$

$$\Delta \bar{y}_{0,01} = q_{0,01}^{(27,4)} \sqrt{\frac{s_e^2}{\bar{n}}} = 4,85 \sqrt{\frac{21,29}{10}} = 7,07$$

Iz tabele možnih razlik, dobimo

	$\bar{y}_1 = 22,6$	$\bar{y}_4 = 32,7$	$\bar{y}_2 = 32,9$	$\bar{y}_3 = 35,1$
$\bar{y}_1 = 22,6$		10,1**	10,3**	12,5**
$\bar{y}_4 = 32,7$			0,2	2,4
$\bar{y}_2 = 32,9$				2,2

Kot sklep dobimo, da je le $\bar{y}_1$ značilno različna od vseh treh drugih ($\bar{y}_4$ $\bar{y}_2$ $\bar{y}_3$) sredin in sicer najmanj na stopnji tveganja $\alpha = 0,01$.

Studentizirane odklone uporabljamo tudi za računanje razmičnih ocen. S tveganjem $\alpha = 0,05$ je npr. razlika med M_3 in M_1 najmanj

$$M_3 - M_1 = 12,5 \pm 5,67$$

Latinski kvadrati

20.10 Z enostavnimi bloki izboljšamo uspešnost poskusa tako, da izločimo vpliv enega motečega faktorja. Z določenim načrtom pa uspemo iz poskusnega gradiva izločiti hkrati dva faktorja, ki motita poskus in tako še zmanjšati poskusni pogrešek.

Vzemimo prejšnji zgled sestavljanja določenega mehanizma pri različnih razmestitvah. V posamezne bloke smo vključili po štiri delavce z izenačenimi delovnimi sposobnostmi. Vsak delavec iz bloka je delal po enem izmed postopkov. Najbolj izenačene osebnostne pogoje pa dosežemo, če dela en delavec po vseh štirih postopkih. Na razlike v času sestavljanja pri istem delavcu namreč ne vplivajo osebnostni faktorji. V tem primeru je blok posamezen delavec. Kot nov moteč dejavnik pa nastopi vrstni red dela. Privajenost more v bistveni meri vplivati na čas dela. Če bi vsi delavci delali po istem vrstnem redu razmestitev P_1 P_2 P_3 P_4, bi poprečni rezultati kazali skupen učinek razmestitev in privajenosti k delu. Vpliv vrstnega reda na po-

prečje izravnotežimo tako, da vsakemu izmed štirih delavcev predpišemo različen vrstni red dela, tako da v celoti vsaka razmestitev pride v skupini delavcev enkrat kot prva, druga, tretja in četrta. Takih načrtov je več. Eden izmed možnih je npr. tale:

Vrstni red:

	1.	2.	3.	4.
Delavec 1	P_1	P_2	P_3	P_4
Delavec 2	P_2	P_3	P_4	P_1
Delavec 3	P_3	P_4	P_1	P_2
Delavec 4	P_4	P_1	P_2	P_3

(20.20)

V tako pisanem poskusu so razmestitve (P) oziroma postopki proučevanega faktorja razmeščeni tako, da imamo istočasno bloke po dveh dejavnikih (v našem zgledu: delavec, vrstni red dela). To je razvidno iz sheme v tabeli 20.7

Tabela 20.7 Dvojni bloki v latinskem kvadratu

	Vrstni red 1.	2.	3.	4.
Delavec 1	P_1	P_2	P_3	P_4
Delavec 2	P_2	P_3	P_4	P_1
Delavec 3	P_3	P_4	P_1	P_2
Delavec 4	P_4	P_1	P_2	P_3

	Vrstni red 1.	2.	3.	4.
Delavec 1	P_1	P_2	P_3	P_4
Delavec 2	P_2	P_3	P_4	P_1
Delavec 3	P_3	P_4	P_1	P_2
Delavec 4	P_4	P_1	P_2	P_3

Tak razpored omogoča, da po metodi blokov izločimo vpliv dveh faktorjev: vrstni red in delavca hkrati.

Če z y_{VSP} označimo posamezne vrednosti poskusa v latinskem kvadratu (ime izvira iz tega, ker po tradiciji navadno zaznamujemo posamezne postopke z latinskimi črkami A B C D), je model poskusa v latinskem kvadratu

$$y_{VSP} = M + (V) + (S) + (P) + e_{VSP} \qquad (20.19)$$

Pri tem so: (V) in (S) sta dejavnika, katerih učinek iz sklopa slučajnostnih faktorjev izločimo, (P) je dejavnik, katerega proučujemo, $e_{VSP} = :N(0, \sigma_e^2)$ pa rezultat slučajnostnih vplivov, ki se porazdeljuje normalno.

Analiza variance za latinski kvadrat je dana v tabeli 20.8.

Tabela 20.8 Analiza variance za poskus v latinskem kvadratu

Vir variacije	SK	m	s^2	$E(s^2)$	F
(V)	$K_V = q_V$	p - 1	s_V^2	$\sigma_e^2 + p\,S_V^2$	s_V^2 / s_e^2
(S)	$K_S = q_S$	p - 1	s_S^2	$\sigma_e^2 + p\,S_S^2$	s_S^2 / s_e^2
(P)	$K_P = q_P$	p - 1	s_P^2	$\sigma_e^2 + p\,S_P^2$	s_P^2 / s_e^2
pp	$K_e = q_{VSP} - q_V - q_S - q_P$	(p-1)(p-2)	s_e^2	σ_e^2	1
Skupaj	$K = q_{VSP}$	$p^2 - 1$	-	-	-

Podobno kot pri že obravnavanih poskusih pomeni:

$$Q_{VSP} = \sum_V \sum_S y_{VSP}^2;\quad Q_V = \frac{1}{p}\sum_V y_V^2;\quad Q_S = \frac{1}{p}\sum_S y_S^2;\quad Q_P = \frac{1}{p}\sum_P y_P^2;\quad Q = \frac{1}{p^2} y^2$$

$$q_{VSP} = Q_{VSP} - Q;\quad q_V = Q_V - Q;\quad q_S = Q_S - Q;\quad q_P = Q_P - Q \qquad (20.21)$$

20.11 Za zgled, ki smo ga nakazali uvodoma, je številčno obračunanje takole:

Po razporedu, ki je predviden za poskus v latinskem kvadratu, smo izvedli časovno študijo za sestavljanje mehanizma po štirih razmestitvah delov in dobili osnovno gradivo. (Podatki se nanašajo na druge mehanizme)

a) osnovni podatki

Delavec	Vrstni red V_1	V_2	V_3	V_4	y_D
D_1	R_1 75	R_2 61	R_3 101	R_4 57	294
D_2	R_2 104	R_3 142	R_4 91	R_1 65	402
D_3	R_3 164	R_4 111	R_1 89	R_2 96	460
D_4	R_4 94	R_1 68	R_2 69	R_3 94	325
y_V	437	382	350	312	1481 y
	R_1	R_2	R_3	R_4	
y_R	297	330	501	353	

b) pomožne količine

Za analizo variance potrebne količine so:

$$Q_{DVR} = 75^2 + 61^2 \ldots + 69^2 + 94^2 = 149693$$

$$q_{DVR} = 149693 - 137085,06 = 12607,94$$

$$Q_D = 1/4(294^2 + 402^2 + 460^2 + 325^2) = 565265/4 = 141316,25$$

$$q_D = 141316,25 - 137085,06 = 4231,19$$

$$Q_V = 1/4(437^2 + 382^2 + 350^2 + 312^2) = 556737/4 = 139184,25$$

$$q_V = 139184,25 - 137085,06 = 2099,19$$

$$Q_R = 1/4(297^2 + 330^2 + 501^2 + 353^2) = 572719/4 = 143179,75$$

$$q_R = 143179,75 - 137085,06 = 6094,69$$

$$Q = 1481^2/4^2 = 2193361/16 = 137085,06$$

Indeksi D, V, R se nanašajo na: D na delavce, V na vrstni red, R na razmestitve.

c) analiza variance.

Analiza variance je po shemi iz tabele 20.8

Vir variacije	vsota kvadratov SK	stopinje prostosti m	s^2	F
delavci	$K_D = q_D = 4231,19$	3	1410,40	46,27***
vrstni red	$K_V = q_V = 2099,19$	3	699,73	22,96**
razmestitve	$K_R = q_R = 6094,69$	3	2031,56	66,65***
poskusni pogrešek	q_{DVR} 12607,94 $-q_D$ - 4231,19 $-q_V$ - 2099,19 $-q_R$ - 6094,69 K_e = 182,87	6	30,48	1,00
Skupno	K = 12607,94	15		

Gornja analiza variance pokaže, da je čas za sestavljanje mehanizma značilno odvisen od razmestitve in sicer je ta značilnost izredno visoka. Kritična vrednost je za $F_{o,001}(3,6) = 23,70$, izračunani $F = 66,65$ pa je dosti višji kot kritična vrednost za $\alpha = 0,001$.

20.12 Na obravnavanem zgledu pa moremo spoznati še neke pomembne lastnosti latinskih kvadratov in načrtov poskusov na splošno. V zgledu nastopata delavec in vrstni red kot dejavnika, katerih vpliv z načrtom izločimo, da moremo jasneje ugotoviti značilnost tretjega - za raziskavo pomembnega dejavnika - razmestitve. V tem primeru je zanimiv in smiseln F-preskus le za razmestitev. V latinskem kvadratu pa ni bistveno, da sta dva činitelja nepomembna, temveč morejo biti vsi trije proučevani činitelji tudi vsebinsko pomembni. Edina omejitev, ki sledi iz modela poskusa v latinskem kvadratu, je, da so proučevani dejavniki med seboj neodvisni, ni pa omejen značaj posameznih dejavnikov. Zato moremo z latinskimi kvadrati reševati

tri vrste problemov: proučevati en dejavnik, pri čemer izločimo dva nepomembna, a določljiva; proučevati dva neodvisna dejavnika, pri čemer izločimo en nepomemben dejavnik in proučevati tri neodvisne dejavnike brez izložitve kateregakoli drugega dejavnika. Seveda imajo v drugem in tretjem primeru pomembni dejavniki enako vlogo kot nepomembni. Ta je v tem, da se z vključevanjem večjega števila dejavnikov veča pojasnjena varianca, nepojasnjena varianca, ki je izraz drugih dejavnikov pa se s tem manjša, kar vpliva na kakovost sklepov.

V našem zgledu sta delavec in vrstni red dejavnika, ki sta zanimiva tudi sama zase. Gotovo je, ne glede na to, da je to postranski rezultat, zanimiva ugotovitev, ali so razlike med delavci značilne ali ne. Enako, mogoče pa še bolj, je pomembno vprašanje, ali vrstni red in z njim privajenost vpliva na čas sestavljanja. Iz analize variance sledi, da so razlike med delavci visoko značilne ($F = 46,27$), medtem ko so razlike v vrstnem redu značilne na stopnji tveganja $\alpha = 0,01$ oziroma je $F = 22,96$ že zelo blizu stopnje $\alpha = 0,001$.

20.13 Drugo vprašanje pa je značaj znakov. V našem zgledu moremo šteti štiri delavce, če so bili ti izbrani slučajnostno, za slučajnosten dejavnik, če želimo sklepe posplošiti na večjo, lahko tudi hipotetično populacijo delavcev. Vrstni red v našem zgledu ne more nastopiti kot dejavnik s slučajnostnim značajem, ker gre za točno opredeljene štiri vrednosti in sicer: sestavlja mehanizem prvič, drugič, tretjič, četrtič. Ti štirje vrstni redi nikakor niso vzorec iz teoretično tudi neomejenega števila vrstnih redov, ampak so prvi štirje poskusi sestavljanja. Podoben primer je tudi z razmestitvijo. Značaj posameznih znakov se pri analizi variance za latinski kvadrat ne kaže pri izračunavanju in primerjavah, pač pa v simboliki, ki nastopa v $E(s^2)$. Za slučajnostne činitelje je za varianco postavljen σ_D^2, za činitelje s stalnim značajem pa S_V^2 in S_R^2.

Ker se je pokazala pri vseh faktorjih razmeroma visoka značilnost, moremo za ocene časov po različnih faktorjih izračunati točkovne in razmične ocene:

$$se(\bar{y}) = \sqrt{\frac{s_e^2}{p}} = \sqrt{\frac{30,48}{4}} = \sqrt{7,62} = 2,76$$

$$d(\bar{y}) = t_\alpha(m_e)\, se(\bar{y}) = 2,45 \cdot 2,76 = 6,762$$

$$se(\Delta\bar{y}) = \sqrt{\frac{2 s_e^2}{p}} = \sqrt{\frac{2.30,48}{4}} = \sqrt{15,24} = 3,90$$

delavci $\bar{y}_{D_1} = 73,5$; $\bar{y}_{D_2} = 100,5$; $\bar{y}_{D_3} = 115,0$; $\bar{y}_{D_4} = 81,25$

vrstni red $\bar{y}_{V_1} = 109,25$; $\bar{y}_{V_2} = 82,0$; $\bar{y}_{V_3} = 87,5$; $\bar{y}_{V_4} = 78,0$

razmestitev $\bar{y}_{R_1} = 74,25$; $\bar{y}_{R_2} = 82,5$; $\bar{y}_{R_3} = 125,25$; $\bar{y}_{R_4} = 88,25$

Da bi preskusili apriorno domnevo, da je čas za sestavljanje mehanizma pri prvem sestavljanju večji kot pri drugih treh, uporabimo primerjavo

$$\bar{y}_{V_1} = \frac{\bar{y}_{V_2} + \bar{y}_{V_3} + \bar{y}_{V_4}}{3}$$

Zanjo je:

$$t = \frac{\bar{y}_{V_1} - \frac{1}{3}(\bar{y}_{V_2} + \bar{y}_{V_3} + \bar{y}_{V_4})}{\sqrt{\frac{s_e^2}{n}\left[1 + (\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2\right]}} =$$

$$= \frac{109,25 - \frac{1}{3}(82 + 87,5 + 78)}{\sqrt{\frac{30,48}{4} \cdot \frac{12}{9}}} = +8,39$$

Ker je $t_{\alpha=P=0,005}^{(m=6)} = 3,71$, sklepamo, da je poprečni čas pri prvem sestavljanju značilno večji od poprečnega časa pri nadaljnjih sestavljanjih, in sicer najmanj na stopnji $\alpha = 0,005$.

Enako kot slučajnostni bloki so tudi latinski kvadrati enostaven in nadvse uspešen

pripomoček pri načrtovanju poskusov. Tako pri analizi tržišča ko preskušamo različne načine prodaje ali različne vrste danega blaga, vzamemo za dejavnika, ki tvorita ogrodje latinskega kvadrata, npr. trgovine in dneve v tednu, če pri načrtu predpostavimo, da dan v tednu in trgovina vplivata na proučevan pojav.

Faktorski poskusi

20.14 V raziskovalnem delu nasploh, pa tudi pri raziskavah ekonomskih pojavov, pogosto raziskujemo učinek več dejavnikov hkrati. Čeprav v vsem raziskovalnem delu iščemo učinek dejavnikov na proučevani pojav, govorimo pri statističnem načrtovanju poskusov o faktorskih poskusih tedaj, ko proučujemo v poskusu istočasno učinek dveh ali več dejavnikov. Učinek posameznih dejavnikov pa ni vselej enostavno vsota učinkov posameznih dejavnikov, kot smo to nakazali npr. pri latinskem kvadratu, temveč je aditivnost prej izjema kot pravilo. To je posebno pomembno, kadar iščemo sestavo delovanja dejavnikov, ki daje najbolj kakovostne proizvode, največjo proizvodnost, način učenja, ki da najboljše rezultate, način najuspešnejšega reklamiranja ipd.

20.16 Čisti učinek in interakcija. Vzemimo najpreprostejši primer proučevanja dveh dejavnikov (A) in (B). Če je učinek teh dveh dejavnikov na proučevan pojav neodvisen, se njihovi učinki vežejo aditivno

$$M + (A) + (B) \tag{20.22}$$

Učinki pa morejo biti tudi odvisni. Poraba določenega potrošnega blaga je odvisna od števila članov v gospodinjstvu. Če pa proučujemo odvisnost porabe od števila članov in višine dohodkov, pa opazimo, da so zakonitosti porabe v odvisnosti od števila članov bistveno drugačni pri različnih dohodkih. To moremo zapisati v obliki

$$M + (A) + (B) + (AB) \tag{20.23}$$

pri čemer pomeni M poprečno raven porabe, (A) učinek dohodka na porabo pri

poprečnem učinku števila članov, (B) učinek števila članov pri poprečnem dohodku, (AB) pa vzajemni učinek dejavnikov A in B. Če je zveza med A in B neodvisna je (AB) = 0.

20.16 Linearni model za dvofaktorski poskus je torej

$$y_{ABi} = M + (A) + (B) + (AB) + e_{ABi}; \quad e_{ABi} =: N(0, \sigma_e^2) \qquad (20.24)$$

Shema za dvofaktorski poskus dveh dejavnikov A in B z $\bar{n}$ ponovitvami je po stopnjah takale:

a) osnovni podatki:

Če z y_{ABi} zaznamujemo i-to ponovitev poskusa pod kombinacijo faktorjev A in B so potrebne vsote dane po znanih pravilih naslednje:

y_{ABi}	
y_{AB}	y_A
y_B	y

Če je vsak postopek AB ponovljen $\bar{n}$ krat, če ima dejavnik A v poskusu a vrednosti, dejavnik B pa b vrednosti, imamo v celotnem poskusu $n = \bar{n}ab$ posameznih poskusov.

b) pomožne količine

$$Q_{ABi} = \sum_A \sum_B \sum_i y_{ABi}^2 \qquad q_{ABi} = Q_{ABi} - Q$$

$$Q_{AB} = \frac{1}{\bar{n}} \sum_A \sum_B y_{AB}^2 \qquad q_{AB} = Q_{AB} - Q$$

$$Q_A = \frac{1}{\bar{n}b} \sum_A y_A^2 \qquad q_A = Q_A - Q$$

$$Q = \frac{1}{\bar{n}a} \sum_B y_B^2 \qquad q_B = Q_B - Q$$

$$Q = \frac{1}{\bar{n}ab} y^2$$

Reducirane vsote kvadratov dobimo tako, da odštejemo količino Q.

c) analiza variance

Analiza variance za ta tip dvofaktorskega poskusa je nakazana v tabeli 20.9

20.9 Analiza variance dvofaktorskega poskusa s ponovitvami

Vir variacije	K	m	s^2	$E(s^2)$	F
Dejavnik A	$K_A = q_A$	a - 1	s_A^2	$\sigma_e^2 + \bar{n}b S_A^2$	s_A^2/s_e^2
Dejavnik B	$K_B = q_B$	b - 1	s_B^2	$\sigma_e^2 + \bar{n}a S_B^2$	s_B^2/s_e^2
Interakcija AB	$K_{AB} = q_{AB} - q_A - q_B$	(a-1)(b-1)	s_{AB}^2	$\sigma_e^2 + \bar{n} S_{AB}^2$	s_{AB}^2/s_e^2
Poskusni pogrešek e	$K_e = q_{ABi} - q_{AB}$	$ab(\bar{n} - 1)$	s_e^2		1
Skupno	$K = q_{ABi}$	$\bar{n}ab-1$			

(20.25)

Iz sheme za analizo variance spoznamo, da smo uspeli vsoto kvadratov razstaviti na štiri dele, od katerih je v prvem razen slučajnostne komponente navzoč le čisti učinek dejavnika A, v drugem čisti učinek dejavnika B, v tretjem vzajemni učinek dejavnikov A in B ali interakcija, v četrtem pa samo rezultat slučajnostne komponente. Če so izpolnjene predpostavke o enakosti varianc po posameznih kombinacijah dejavnikov in normalnega razporeda slučajnostne komponente, moremo s F-preskusom preskušati značilnost posameznih komponent, ki so osnova za nadaljnjo proučitev vpliva posameznih dejavnikov prek skupinskih poprečij.

20.17 Vzemimo za zgled dvofaktorskega poskusa proučitev vpliva višine mesečnega dohodka in števila članov gospodinjstva na porabo določene skupine porabnega blaga. V ta namen smo določili, da bomo proučevali tri tipe gospodinjstev (dvo, tro in štiričlanska) s štirimi različnimi dohodki (800 din, 1200 din, 1600 din in

2000 din). Da bi preskusili vpliv teh dveh dejavnikov smo za vsako kombinacijo proučevanih dejavnikov (3x4=12 kombinacij) iz drugače homogene populacije izbrali po dvoje gospodinjstev (najmanj dvoje osnovnih enot oziroma poskusov je potrebno, da ocenimo vse količine v poskusu).

Značilno za nakazano raziskavo je, da posamezne "poskuse" nismo izvedli, temveč smo iz populacije vseh gospodinjstev izbrali gospodinjstva, ki ustrezajo poskusnim pogojem. Tak način proučevanja v socialno-ekonomskih pojavih je iz znanih razlogov pogost.

a) osnovni podatki

Osnovni podatki nakazane raziskave so tile (podatki so izmišljeni):

$y_{DČi}$

$y_{DČ}$ | y_D

$y_Č$ | y

Število članov	Dohodek v din				Skupno
	800	1200	1600	2000	$y_Č$
	$y_{DČi}$				
2	27	25	25	21	
	29	29	27	25	
$y_{DČ}$	56	54	52	46	208
3	21	23	34	39	
	23	29	30	33	
$y_{DČ}$	44	52	64	72	232
4	25	30	35	49	
	25	32	41	43	
$y_{DČ}$	50	62	76	92	280
Skupno y_D	150	168	192	210	720 = y

b) pomožne količine

Iz osnovnih podatkov moremo izračunati potrebne pomožne količine Q_u in q_u:

$$Q_{\check{C}Di} = \sum_{\check{C}Di} y^2_{\check{C}Di} = 27^2 + 25^2 + \ldots + 41^2 + 43^2 = 22792$$

$$Q_{\check{C}D} = 1/2 \sum_{\check{C}D} y^2_{\check{C}D} = 1/2\ (56^2 + 54^2 + \ldots + 76^2 + 92^2 = 1/2\ (45376) = 22688$$

$$Q_{\check{C}} = 1/2.4 \sum_{\check{C}} y^2_{\check{C}} = 1/8(208^2+232^2+280^2) = 1/8\ (175488) = 21936$$

$$Q_D = 1/2.3 \sum_{D} y^2_{D} = 1/6\ (150^2+168^2+192^2+210^2) = 1/6\ (131688) = 21948$$

$$Q = 1/2.3.4.\ (y^2) = 1/2.3.4\ (720^2) = 1/2.3.4\ (518400) = 21600$$

Iz Q izračunajmo q:

$$q_{\check{C}Di} = Q_{\check{C}Di} - Q = 22792 - 21600 = 1192$$

$$q_{\check{C}D} = Q_{\check{C}D} - Q = 22688 - 21600 = 1088$$

$$q_{\check{C}} = Q_{\check{C}} - Q = 21936 - 21600 = 336$$

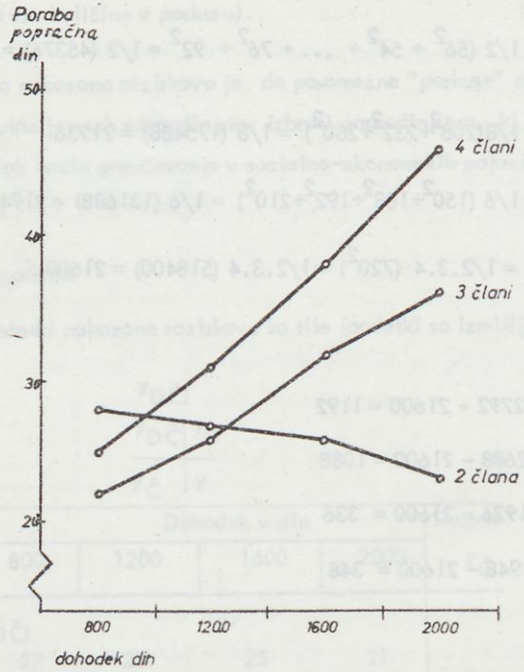
$$q_D = Q_D - Q = 21948 - 21600 = 348$$

c) analiza variance

Obračun analize variance iz pomožnih količin je

Vir variacije	K	m	s^2	F
Število članov	$K_{\check{C}} = q_{\check{C}} = 336$	$m_{\check{C}} = 3 - 1 = 2$	168	$\frac{168}{8,67} = 19,38^{**}$
Dohodek	$K_D = q_D = 348$	$m_D = 4 - 1 = 3$	116	$\frac{116}{8,67} = 13,38^{**}$
Interakcija ČxD	$q_{\check{C}D} = 1088$ $-q_{\check{C}} = -336$ $-q_D = -348$			
	$K_{\check{C}D} = 404$	$m_{\check{C}D} = (3-1)(4-1)=6$	67,33	$\frac{67,33}{8,67} = 7,77^{**}$
Poskusni pogrešek	$q_{\check{C}Di} = 1192$ $-q_{\check{C}D} = -1088$			
	$K_e = 104$	$m_e = (2-1)3.4=12$	8,67	1,00
Skupno	$K = q_{\check{C}Di} = 1192$	$m = 2.3.4-1 = 23$		

$$d_{\bar{y}_{\check{C}D}} = t_{0.05}^{(m=12)} \sqrt{\frac{s_e^2}{\bar{n}}} = 2{,}18 \cdot \sqrt{\frac{8{,}67}{2}} = 4{,}54$$



Slika 20.1 Poraba v odvisnosti od dohodka in števila članov

Ker se je interakcija ČD pokazala značilna, moramo proučevati poprečja $\bar{y}_{\check{C}D}$, ne pa sumarna poprečja $\bar{y}_{\check{C}}$ oli $\bar{y}_{D}$, ker se poraba v odvisnosti od višine dohodka menja različno pri različnem številu članov.

20.18 Drugi načrti poskusov. Prikazani načrti poskusov so samo elementarni poskusi. Nakazane ideje so razvite naprej bodisi glede na število dejavnikov, ki jih proučujemo, bodisi na postopke, s katerimi izločamo iz poskusa vplive, ki posebno motijo. Tako so posebej razviti faktorski poskusi z velikim številom dejavnikov, ki imajo samo po dve vrednosti, tj. 2^p poskusi.

Obsežnost blokov pri faktorskih poskusih z velikim številom postopkov zmanjšamo z delnimi bloki, v katerih so vključeni samo nekateri dejavniki. Frakcionalni poskusi rešujejo isti problem velikega števila možnih kombinacij dejavnikov pri faktorskih poskusih tako, da v poskus vključimo le del vseh možnih kombinacij. Tudi z delnim žrtvovanjem informacij, posebno o interakcijah višjih redov, zmanjšamo obseg blokov, tako, da učinek teh interakcij pomešamo oziroma združimo z razlikami med bloki. Včasih je število postopkov že pri enem dejavniku tako veliko, da ne moremo zagotoviti v enem bloku enovitih pogojev za vse postopke. Z nepopolnimi bloki v pravem pomenu, v katere vključimo le del vseh postopkov, moremo obiti prikazano težavo. S posebnim postopkom - analizo kovariance moremo pri posameznih nakazanih načrtih izločiti moteč vpliv določenega dejavnika ali več dejavnikov hkrati.

Pri vsakem poskusu oziroma načrtu poskusa težimo za tem, da dobimo čim boljše osvetljen vpliv posameznih dejavnikov s čim manjšimi stroški. V tej zvezi je gotovo pomembno določiti obseg vzorca, ki je zadosti velik, da da odgovor na postavljeno vprašanje s predpisano zanesljivostjo.

ENAINDVAJSETO POGLAVJE

PROUČEVANJE DINAMIKE POJAVOV-ČASOVNE VRSTE

21.1 Socialno-ekonomski pojavi niso časovno nespremenljivi. Spremembe so rezultat delovanja najrazličnejših dejavnikov, ki v tej ali oni obliki vplivajo na pojave. Sliko dinamike pojavov dobimo s časovnimi vrstami. Ena slika daje statični prikaz. Niz slik, ki slede na filmskem platnu, pa daje videz gibanja. Enako dá en sam podatek statično sliko pojava, niz istovrstnih podatkov v enakih časovnih razmikih pa daje sliko dinamike pojava.

Časovna vrsta je niz istovrstnih podatkov, ki se nanašajo na zaporedne časovne razmike ali trenutke. Z njimi proučujemo časovni razvoj pojavov, ker prikazujejo spremembe pojavov v odvisnosti od časa. Po namenu in načinu proučevanja se razlikujejo od drugih statističnih vrst. Osnovni namen proučevanja časovnih vrst je, opazovati časovni razvoj socialno-ekonomskih pojavov, iskati njegove zakonitosti in napovedati nadaljnji razvoj. V pogojih planskega gospodarstva je proučevanje časovnih vrst važno tako pri sestavljanju plana kot pri kontroli izvajanja plana. Predvidevanje in napovedovanje razvoja ekonomskih pojavov, če poznamo zakonitosti in sedanje stanje, se je razvilo v posebno disciplino-vedo o konjunkturah, ki s proučevanjem časovnih vrst najvažnejših ekonomskih pokazovalcev skuša sklepati na nadaljnji ekonomski razvoj. Seveda to napovedovanje ne more biti popolnoma zanesljivo, ker je ne-

mogoče vnaprej napovedati in upoštevati vse faktorje, ki vplivajo na ekonomski razvoj. Napoved bi veljala strogo le v primeru, če bi bile izpolnjene predpostavke, pod katerimi je napoved izdelana.

OBLIKE ČASOVNIH VRST

Trenutne in razmične časovne vrste

21.2 Časovne vrste po značaju podatkov, ki jih prikazujejo, delimo na **trenutne in razmične**. Trenutne časovne vrste prikazujejo časovni razvoj za pojave, ki veljajo za posamezne trenutke; razmične časovne vrste pa časovni razvoj pojavov, ki so razmičnega značaja. Trenutna časovna vrsta je npr. vrsta podatkov o številu delavstva konec posameznega meseca, primer razmične časovne vrste pa je proizvodnja v podjetju po mesecih. Enako je v tabeli 21.1 število prebivalstva v SFRJ po letih sredi leta trenutna, naravni prirastek po letih pa razmična časovna vrsta.

Tabela 21.1: Število prebivalstva in naravni prirastek prebivalstva v SFRJ
(Vir: SG - 1971)

Leto	Število prebivalstva sredi leta v tisočih	Naravni prirastek v tisočih
1960	18402	250
1961	18612	255
1962	18819	226
1963	19029	238
1964	19222	220
1965	19434	238
1966	19644	240
1967	19840	216
1968	20029	208
1969	20209	194
1970	20371	179

Izvedene časovne vrste

21.3 Iz osnovnih časovnih vrst moremo izračunati več izvedenih časovnih vrst, ki pripomorejo k proučitvi dinamike pojavov. Izmed teh bomo navedli: kumulativno časovno vrsto, časovno vrsto sredin, časovno vrsto drsečih vsot in časovno vrsto drsečih sredin.

Kumulativna časovna vrsta

21.4 Kumulativno časovno vrsto izračunavamo samo za razmične vrste ekstenzivnih podatkov, ker ima samo zanje kumuliranje oziroma vsota členov smisel. Kumulativno časovno vrsto dobimo s postopnim seštevanjem členov osnovne vrste. Členi kumulativne vrste pomenijo vrednosti v razmiku od začetka proučevanega razdobja do določenega trenutka, ki se od člena do člena menja. Kumulativno časovno vrsto računamo običajno za zaključene časovne razmike kot so: mesečne vrednosti v razdobju enega leta, dnevne vrednosti v razdobju enega meseca, letne vrednosti v razdobju ene petletke itd. Kumulativne vrste se v teh primerih nanašajo na začetek petletke, začetek leta, začetek meseca.

Kumulativno časovno vrsto dobimo tako, da vzamemo vrednost prvega člena kumulativne vrste enako 0, nadaljnje člene kumulativne vrste pa izračunamo po znanem obrazcu

$$K_{k+1} = K_k + Y_k \qquad (21.1)$$

K_k = kumulativa ; Y_k = vrednost v tekočem razdobju k

21.5 Za zgled izračunanja kumulativne časovne vrste vzemimo mesečno vrsto proizvodnje krogličnih ležajev v SFRJ.

Tabela 21.2: Izračunanje kumulativne časovne vrste za proizvodnjo krogličnih ležajev v SFRJ v letu 1957 (Vir: Indeks)

Mesec	Mesečna proizvodnja v tonah	Kumulativa
J	17	0 = 0
F	16	0 + 17 = 17
M	22	17 + 16 = 33
A	21	33 + 22 = 55
M	19	55 + 21 = 76
J	19	76 + 19 = 95
J	20	95 + 19 = 114
A	5	114 + 20 = 134
S	18	134 + 5 = 139
O	14	139 + 18 = 157
N	18	157 + 14 = 171
D	18	171 + 18 = 189
Skupaj		189 + 18 = 207

Šesti člen v kumulativni vrsti 95 ton pome ni, da je bilo od začetka leta 1957 do začetka meseca junija v SFRJ proizvedeno 95 ton krogličnih ležajev. Podoben pomen imajo drugi členi kumulativnih vrst. Zadnji člen v kumulativni časovni vrsti je enak vsoti vseh členov iz osnovne časovne vrste.

Vrsta sredin

21.6 Časovne vrste z velikim številom členov so dostikrat nepregledne in je osnovna dinamika zaradi prevelike dolžine vrste zabrisana. Za take vrste zmanjšamo število členov z združevanjem členov v vsote ali povprečja. Iz časovne vrste o številu tujih gostov po mesecih v razdobju 1956-1967, ki ima 144 členov, dobimo časovno vrsto z 12 členi, če seštejemo število tujih gostov po mesecih v letne vsote. Seveda ta časovna vrsta ne podaja tako natančno gibanje proučevanega pojava kot

osnovna vrsta. S tem zabrišemo marsikatero značilnost pojava.

Medtem ko ima vrsta vsot smisel le, če je vrsta, ki jo prikazujemo, razmična in ekstenzivna, moremo z grupnimi poprečji skrčiti število členov za vsako časovno vrsto ne glede na značaj členov.

Člene v osnovni časovni vrsti združujemo po naravnih časovnih enotah. Tako združujemo mesečne podatke v četrtletne ali letne, dnevne podatke v tedenske ali mesečne, letne podatke pa običajno v petletja, tako da iz njih dobimo zaokrožena petletja ali desetletja, npr. petletja 1900-1904, 1905-1909, 1910-1914...1945-1949,1950-1954.

21.7 Če izračunavamo časovno vrsto poprečij, je treba paziti, na katere trenutke se nanašajo osnovne vrednosti v časovni vrsti. Od tega je namreč odvisno, kako računamo sredine.

Če se nanašajo vrednosti na sredine osnovnih razmikov kot je to primer pri razmičnih časovnih vrstah in pri nekaterih trenutnih, računamo poprečja po obrazcu

$$\bar{Y} = \frac{1}{r}(Y_1 + Y_2 + \ldots + Y_r) \qquad (21.2)$$

Za trenutne časovne vrste, za katere se vrednosti nanašajo na začetek ali konec osnovnih razmikov, pa izračunavamo sredine po obrazcu

$$\bar{Y} = \frac{1}{r}\left(\frac{1}{2}Y_0 + Y_1 + Y_2 + \ldots + Y_{r-1} + \frac{1}{2}Y_r\right) \qquad (21.3)$$

Po obrazcu 21.2 računamo na primer poprečno mesečno proizvodnjo po letih, če imamo mesečne podatke o proizvodnji, ali poprečno število prebivalstva po petletjih, če razpolagamo s številom prebivalstva po letih sredi leta. Obrazec 21.3 pa upoštevamo, če izračunavamo na primer poprečno število delavcev po letih, imamo pa podatke o številu delavcev v začetku vsakega meseca itd. Na to moramo paziti zato, da se poprečja nanašajo na zaokrožena koledarska razdobja: tedne, mesece, leta, pet-

letja in na iste trenutke kot člani v osnovni časovni vrsti.

Vrsta drsečih vsot

21.8 Iz mesečnih podatkov o proizvodnji s seštevanjem dobimo časovno vrsto za letno proizvodnjo. Ti letni podatki se nanašajo na letne razmike koledarskih let od januarja do decembra. Pri proučevanju dinamike pa so poleg teh važni tudi drugi letni razmiki, na primer od februarja danega leta do januarja naslednjega leta, od marca do februarja naslednjega leta itd. Ti letni razmiki se med seboj delno prekrivajo, tako da je vsak razmik, če ga primerjamo s predhodnim, premaknjen za en mesec. Ti letni razmiki tako rekoč drsijo od člena do člena za en mesec. Odtod tudi ime drseče vsote. Časovna vrsta drsečih vsot ima niz lastnosti, ki ji dajo posebno analitično vrednost. Ena izmed teh je, da se v časovni vrsti drsečih vsot izravnavajo slučajni in periodični vplivi, če obsega vsota časovno periodo. O tem bomo podrobneje slišali v naslednjih poglavjih.

Zaradi zveze med zaporednimi člani v časovni vrsti drsečih vsot računamo časovno vrsto drsečih vrednosti razmeroma enostavno. Če imamo izračunan prvi člen časovne vrste drsečih vsot (dobimo ga tako, da seštejemo vrednosti od januarja do decembra za prvo leto v časovni vrsti), dobimo naslednji člen v časovni vrsti drsečih vsot tako, da januarsko vrednost od prve letne vsote odštejemo, prištejemo pa vrednost za januar naslednjega leta. Iz tako dobljenega drugega člena dobimo tretji člen časovne vrste drsečih vsot, če prištejemo februarski člen prihodnjega in odštejemo februarski člen prejšnjega leta iz osnovne časovne vrste. Posamezni členi v vrsti drsečih vsot pomenijo letno proizvodnjo v preteklem letu do začetka januarja, do začetka februarja itd.

Nakazano računanje vrste drsečih vsot ne velja samo za mesečne podatke, marveč na splošno za katero koli vrsto ekstenzivnih razmičnih podatkov. Enako iz dnevnih podatkov računamo časovno vrsto tedenskih drsečih vrednosti o številu prevoženih potnikov. Iz letnih podatkov o številu živorojenih pa moremo sestaviti petletno vrsto drsečih vsot za število živorojenih itd.

Drseče vsote v vseh primerih računamo po splošnem obrazcu:

$$S_{k+1} = S_k + Y_k - Y_{k-r} = S_k + {}_r d_k \qquad (21.4)$$

Pri tem pomeni: S_k in S_{k+1} = ustrezne vrednosti v vrsti drsečih vsot; Y_k = člen k v osnovni vrsti; Y_{k-r} = člen osnovne vrste, ki je za r členov pred Y_k; r = število členov, iz kolikor so računane vsote; ${}_r d_k = Y_k - Y_{k-r}$ = razlika ustreznih vrednosti za Y.

21.9 Za število potnikov v zračnem prometu v SFRJ v razdobju 1955-1958 je izračunanje časovne vrste drsečih vsot nakazano v tabeli 21.3.

Tabela 21.3 Izračunanje časovne vrste drsečih vsot za število potnikov v zračnem prometu po četrtletjih v SFRJ v razdobju 1955-1958 (Vir: Indeks)

Leto	Četrtletje	Y_k	Y_{k-4}	${}_r d_k$	S_k
1955	1	8,2			
	2	23,1			
	3	58,0			
	4	14,7			
1956	1	8,3	8,2	+ 0,1	104,0
	2	19,9	23,1	- 3,2	104,1
	3	50,6	58,0	- 7,4	100,9
	4	15,2	14,7	+ 0,5	93,5
1957	1	9,8	8,3	+ 1,5	94,0
	2	23,6	19,9	+ 3,7	95,5
	3	60,4	50,6	+ 9,8	99,2
	4	17,9	15,2	+ 2,7	109,0
1958	1	10,9	9,8	+ 1,1	111,7
	2	26,0	23,6	+ 2,4	112,8
	3	72,3	60,4	+11,9	115,2
	4	21,0	17,9	+ 3,1	127,1
					130,2

Iz zgleda je jasno razviden postopek:

$S_5=8,2+23,1+38,0+14,7=104,0$; $S_6=104,0+(+0,1)=104,1$; $S_7=104,1+(-3,2)=100,9$ itd.

Časovna vrsta drsečih sredin

21.10 Vrednosti členov v časovni vrsti sredin priredimo sredini razmikov, na katere se poprečja nanašajo (npr. sredini let pri letnih poprečjih, tretjemu letu petletke pri petletnih poprečjih itd.). Členi v časovni vrsti sredin so torej prirejeni samo nekaterim členom v osnovni časovni vrsti. Za analizo pa je dostikrat potrebno, da imamo prirejena poprečja vsem, ne pa samo nekaterim členom v osnovni časovni vrsti. To dosežemo s časovno vrsto drsečih sredin, ki jo izračunavamo podobno kot vrsto drsečih vsot.

Ker se morajo posamezni členi v vrsti drsečih sredin nanašati na iste trenutke kot členi v osnovni časovni vrsti, izračunamo člene vrste drsečih vsot glede na osnovno časovno vrsto na dva načina:

Če imajo razmiki, na katere se nanašajo sredine, liho število osnovnih razmikov: $r = 2i+1$, izračunavamo časovno vrsto drsečih sredin po obrazcu

$$\bar{y}_k = \frac{1}{r}(Y_{k-i}+Y_{k-i+1}+\ldots+Y_k+\ldots+Y_{k+i-1}+Y_{k+i}) = \frac{1}{r}\,S_{k+i+1} \qquad (21.5)$$

Tak primer je računanje tedenskih ali petletnih poprečij.

21.11 Vzemimo za zgled časovno vrsto o številu umrlih na 1000 prebivalcev v Jugoslaviji v letih 1921-1939 in zanjo izračunajmo časovno vrsto petletnih sredin. Ker računamo poprečja iz lihega števila osnovnih razmikov, uporabimo obrazec 21.5.

Iz obrazca 21.5 povzamemo, da je najbolje, če po postopku, ki smo ga že navedli, izračunamo vrsto drsečih petletnih vsot S_k, petletne vsote pa delimo s pet.

Tabela 21.4 Izračun časovne vrste drsečih sredin za število umrlih na 1000 prebivalcev v bivši Jugoslaviji v letih 1921-1939

Leto	Y_k	Y_{k-5}	5^d_k	S_k	$\widetilde{Y}_k$
1921	20,9				
1922	20,8				
1923	20,3				20,2
1924	20,2				19,8
1925	18,7				19,8
1926	18,8	20,9	-2,1	100,9	19,8
1927	20,9	20,8	+0,1	98,8	20,0
1928	20,4	20,3	+0,1	98,9	20,1
1929	21,2	20,2	+1,0	99,0	20,3
1930	19,0	18,7	+0,3	100,0	19,9
1931	19,8	18,8	+1,0	100,3	19,2
1932	19,2	20,9	-1,7	101,3	18,4
1933	16,9	20,4	-3,5	99,6	18,0
1934	17,1	21,2	-4,1	96,1	17,2
1935	16,8	19,0	-2,2	92,0	16,6
1936	16,1	19,8	-3,7	89,8	16,3
1937	15,9	19,2	-3,3	86,1	15,9
1938	15,6	16,9	-1,3	82,8	
1939	14,9	17,1	-2,2	81,5	
				79,3	

21.12 Če pa sredine veljajo za sodo število osnovnih razmikov ($r = 2i$), računamo drseče sredine po obrazcu:

$$\widetilde{Y}_k = \frac{1}{r}\left(\frac{1}{2}Y_{k-i} + Y_{k-i+1} + \ldots + Y_k + \ldots + Y_{k+i-1} + \frac{1}{2}Y_{k+i}\right) \qquad (21.6)$$

Tak primer imamo pri letnih poprečjih iz mesečnih ($r = 12$) ali četrtletnih ($r = 4$) podatkov.

Če je število osnovnih razmikov sodo ($r = 2i$), je najprikladneje obrazec 21.6 preurediti v obliko

$$\bar{Y}_k = \frac{1}{2r}\,(Y_{k-i} + 2Y_{k-i+1} + \ldots + 2Y_k + \ldots + 2Y_{k+i-1} + Y_{k+i}) = \frac{1}{2r}\,S'_k \qquad (21.7)$$

Ker je

$$S'_{k+1} = S_{k+i+1} + S_{k+i+2} = S_{k+i} + {}_r d_{k+i} + S_{k+i+1} + {}_r d_{k+i+1} = S'_k + {}_r d_{k+i} + {}_r d_{k+i+1}$$

je ponderirano vsoto S'_k najprikladneje računati po obrazcu 21.8 tako, da prvo vsoto izračunamo, naslednje pa dobimo s postopnim prištevanjem dveh diferenc ${}_r d_k$.

21.13 Če imamo vrsto mesečnih podatkov, po pravilu računamo vrsto drsečih sredin za letna razdobja. Za ta primer je po gornjih obrazcih: r=2i=12 ; i = 6. Ker začnemo s $S'_{i+1} = S'_{6+1} = S'_7$, se prvi člen drsečih sredin nanaša na 7.mesec (julij) prvega leta. Zanj računamo S'_7 po obrazcu 21.7 :

$$S_7 = (Y_1 + 2Y_2 + \ldots + 2Y_7 + \ldots + 2Y_{12} + Y_{13})$$

Y_{13} = podatek za januar drugega leta. Po obrazcu 21.8 pa dobimo dalje:

$S_8 = S_7 + {}_{12}d_{13} + {}_{12}d_{14}$: $S_9 = S_8 + {}_{12}d_{14} + {}_{12}d_{15}$ itd : ..... ${}_{12}d_{13} = Y_{13} - Y_1$:

${}_{12}d_{14} = Y_{14} - Y_2$ itd.

Kot zgled vzemimo indeks industrije za gradbeni material v SFRJ. (Glej tabelo 21.5!).

Razlike med ustreznimi meseci zaporednih let smo izračunali takole: Prva možna razlika je za mesec januar 1952:

${}_{12}d_{13} = Y_{J,52} - Y_{J,51} = 31 - 38 = -7$; za februar 1952: ${}_{12}d_{14} = Y_{F,52} - Y_{F,51} =$
$= 27 - 37 = -10$ itd.

Prvo vsoto (za julij 1951) smo izračunali v celoti:

$S'_{J,51} = 38 + 2.37 + 2.49 + \ldots\ldots + 2.84 + 2.71 + 2.58 + 31 = 1707.$

Nadaljnje vsote pa so izračunane, kot kaže slika 21.1

Tabela 21.5 Izračunanje časovne vrste drsečih sredin za indeks industrije gradbenega materiala v SFRJ (Ø 1956 = 100)

(Vir: SB 108)

		J	F	M	A	M	J	J	A	S	O	N	D
1951	Y	38	37	49	56	78	95	97	104	90	84	71	58
	d												
S'= 24	$\bar{Y}$							1707	1690	1667	1648	1635	1634
	$\bar{\bar{Y}}$							71,1	70,4	69,5	68,7	68,1	68,1
1952	Y	31	27	36	50	71	101	124	122	112	98	71	56
	d	-7	-10	-13	-6	-7	+6	+27	+18	+22	+14	0	-2
S'= 24	$\bar{Y}$	1667	1712	1752	1788	1802	1800	1801	1810	1828	1858	1897	1936
	$\bar{\bar{Y}}$	69,5	71,3	73,0	74,5	75,1	75,0	75,0	75,4	76,2	77,4	79,0	80,7
1953	Y	34	33	48	68	92	119	125	131	127	118	86	74
	d	+3	+6	+12	+18	+21	+18	+1	+9	+15	+20	+15	+18
S'= 24	$\bar{Y}$	1955	1965	1989	2024	2059	2092	2108	2095	2084	2081	2071	2062
	$\bar{\bar{Y}}$	81,5	81,9	82,9	84,3	85,8	87,2	87,8	87,3	86,8	86,7	86,3	85,9
1954	Y	32	22	48	65	85	117	134	143	145	122	90	83
	d	-2	-11	0	-3	-7	-2	+9	+12	+18	+4	+4	+9
S'= 24	$\bar{Y}$	2069	2090	2120	2142	2150	2163	2195	2247	2282	2284	2292	2332
	$\bar{\bar{Y}}$	86,2	87,1	88,3	89,3	89,6	90,1	91,5	93,6	95,1	95,2	95,5	97,2
1955	Y	55	51	54	61	97	145	138	144	135	121	89	81
	d	+23	+29	+6	-4	+12	+28	+4	+1	-10	-1	-1	-2
S'= 24	$\bar{Y}$	2364	2369	2360	2349	2347	2344	2339	2311	2276	2266	2256	2232
	$\bar{\bar{Y}}$	98,5	98,7	98,3	97,9	97,8	97,7	97,5	96,3	94,8	94,4	94,0	93,0
1956	Y	52	26	44	61	87	131	139	157	155	154	106	89
	d	-3	-25	-10	0	-10	-14	+1	+13	+20	+33	+17	+8
S'= 24	$\bar{Y}$	2219	2233	2266	2319	2369	2394	2394	2398	2420	2445	2492	2547
	$\bar{\bar{Y}}$	92,5	93,0	94,4	96,6	98,7	99,8	99,8	99,9	100,8	101,9	103,8	106,1
1957	Y	44	38	54	76	119	154	172	179	166	152	121	97
	d	-8	+12	+10	+15	+32	+23	+33	+22	+11	-2	+15	+8
S'= 24	$\bar{Y}$	2603	2658	2691	2700	2713	2736						
	$\bar{\bar{Y}}$	108,5	110,8	112,1	112,5	113,0	114,0						

1707 ◇

- 7 +

-10 +

1690 ◇

-10 +

-13 +

1667 ◇

-13 +

- 6 +

1648 ◇

- 6 +

.

.

.

Slika 21.1 Izračunanje vrste drsečih vsot S'_k.

Iz obeh zgledov vidimo, da je časovna vrsta drsečih sredin krajša od osnovne časovne vrste in da začetnim in končnim i členom v osnovni časovni vrsti ne moremo prirediti drseče sredine. To je hiba časovne vrste drsečih sredin, s katero moramo računati.

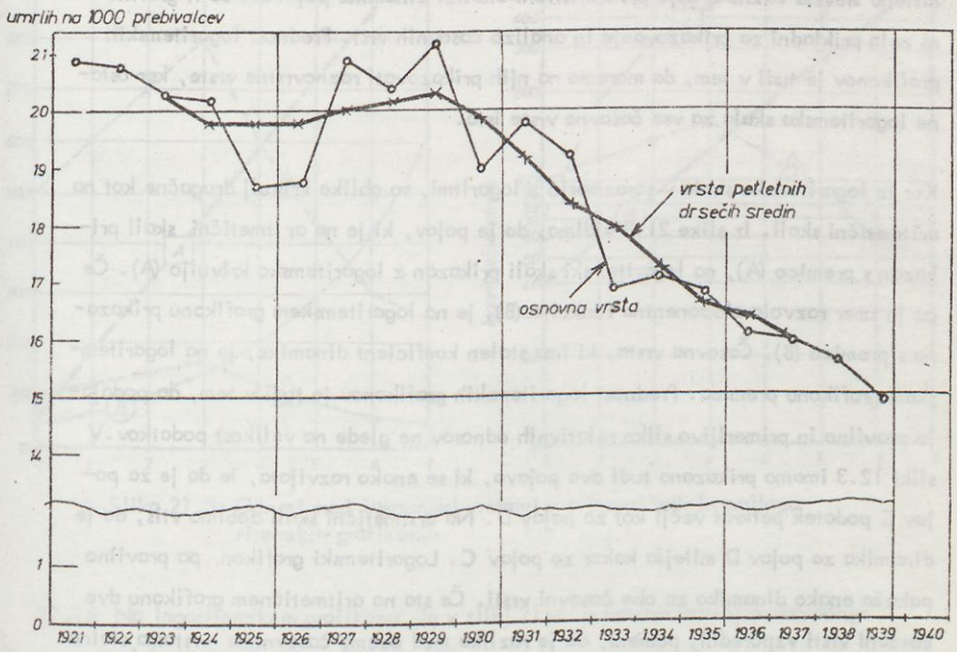
GRAFIČNO PRIKAZOVANJE ČASOVNIH VRST

21.14 Kompleksen vpogled v dinamiko pojavov dobimo z grafičnim prikazom časovnih vrst. Tabelarični prikaz časovnih vrst je nenazoren, ker imajo časovne vrste običajno veliko členov. Razen tega je v tabeli izredno težko primerjati več časovnih vrst istočasno.

Časovne vrste prikazujemo pretežno z linijskimi grafikoni. Stolpce ali druge načine prikazovanja uporabljamo le za krajše časovne vrste.

Najenostavnejši prikaz časovne vrste je navaden linijski grafikon. V njem je ko-

ličinska skala navadna aritmetična skala. Teh vrst grafikonov imamo največ. Eden izmed njih je grafikon v sliki 21.2.



Slika 21.2 Število umrlih na tisoč prebivalcev v predvojni Jugoslaviji

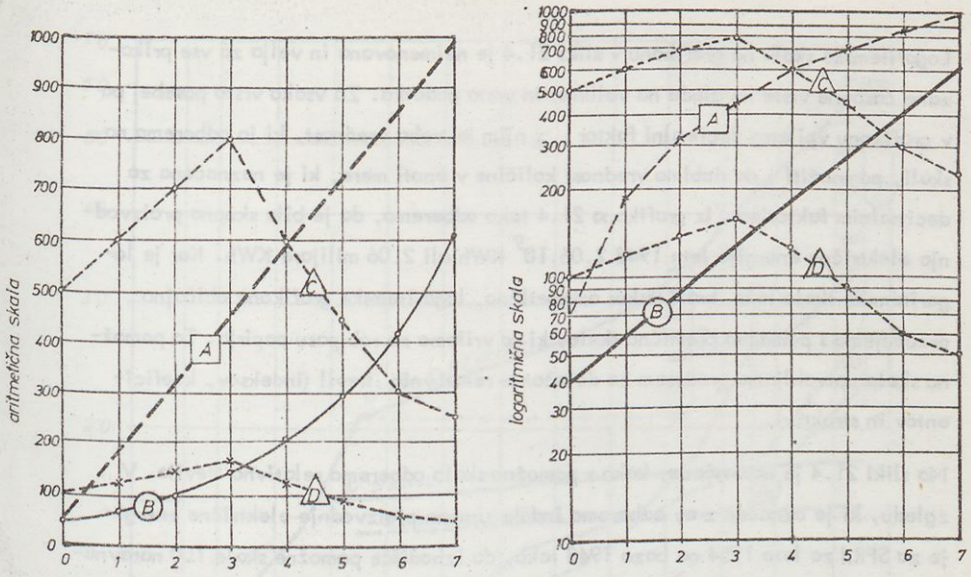
Na aritmetični skali daljice med podatki prikazujejo absolutne diference med dvema podatkoma. Na linijskih grafikonih z aritmetično skalo pa moremo praviloma prikazovati samo istovrstne časovne vrste. Pri prikazovanju raznovrstnih časovnih vrst na istem grafikonu je namreč problematično razmerje med skalami.

Pollogaritemski grafikoni

21.15 Za prikazovanje dinamike pojavov je posebno prikladen grafikon, ki ima namesto aritmetične količinske skale logaritemsko skalo. Ker je razlika logaritmov enaka logaritmu kvocienta, je na logaritemskem grafikonu navpična razdalja med dvema podatkoma v razmerju z logaritmom kvocienta med podatkoma. Zato mo-

remo iz logaritemskega grafikona odbrati relativne odnose med podatki. Ker so relativna števila važno orodje pri statistični analizi dinamike pojavov, so ti grafikoni zelo prikladni za prikazovanje in analizo časovnih vrst. Prednost logaritemskih grafikonov je tudi v tem, da moremo na njih prikazovati raznovrstne vrste, ker ostane logaritemska skala za vse časovne vrste ista.

Ker je logaritemska skala v sorazmerju z logaritmi, so oblike krivulj drugačne kot na aritmetični skali. Iz slike 21.3 vidimo, da je pojav, ki je na aritmetični skali prikazan s premico (A), na logaritemski skali prikazan z logaritemsko krivuljo (A). Če pa je smer razvoja eksponentna funkcija (B), je na logaritemskem grafikonu prikazana s premico (B). Časovna vrsta, ki ima stalen koeficient dinamike, je na logaritemskem grafikonu premica. Prednost logaritemskih grafikonov je tudi v tem, da podajajo pravilno in primerljivo sliko relativnih odnosov ne glede na velikost podatkov. V sliki 12.3 imamo prikazana tudi dva pojava, ki se enako razvijata, le da je za pojav C podatek petkrat večji kot za pojav D. Na aritmetični skali dobimo vtis, da je dinamika za pojav D milejša kakor za pojav C. Logaritemski grafikon pa pravilno pokaže enako dinamiko za obe časovni vrsti. Če sta na aritmetičnem grafikonu dve časovni vrsti vzporedni, pomeni, da je razlika med obema časovnima vrstama stalna. Če pa sta časovni vrsti na logaritemskem grafikonu vzporedni, pomeni, da je stalno razmerje med obema vrstama.



Slika 21.3 Odnosi med časovnimi vrstami v aritmetičnih in poliogaritemskih grafikonih

21.16 Na logaritemskem grafikonu je v sliki 21.4 prikazan razvoj proizvodnje električne energije v SFRJ in SRS.

Tabela 21.6. Proizvodnja električne energije in število prebivalstva v SFRJ in SRS v razdobju 1946-1957 (Vir: SG 57-58)

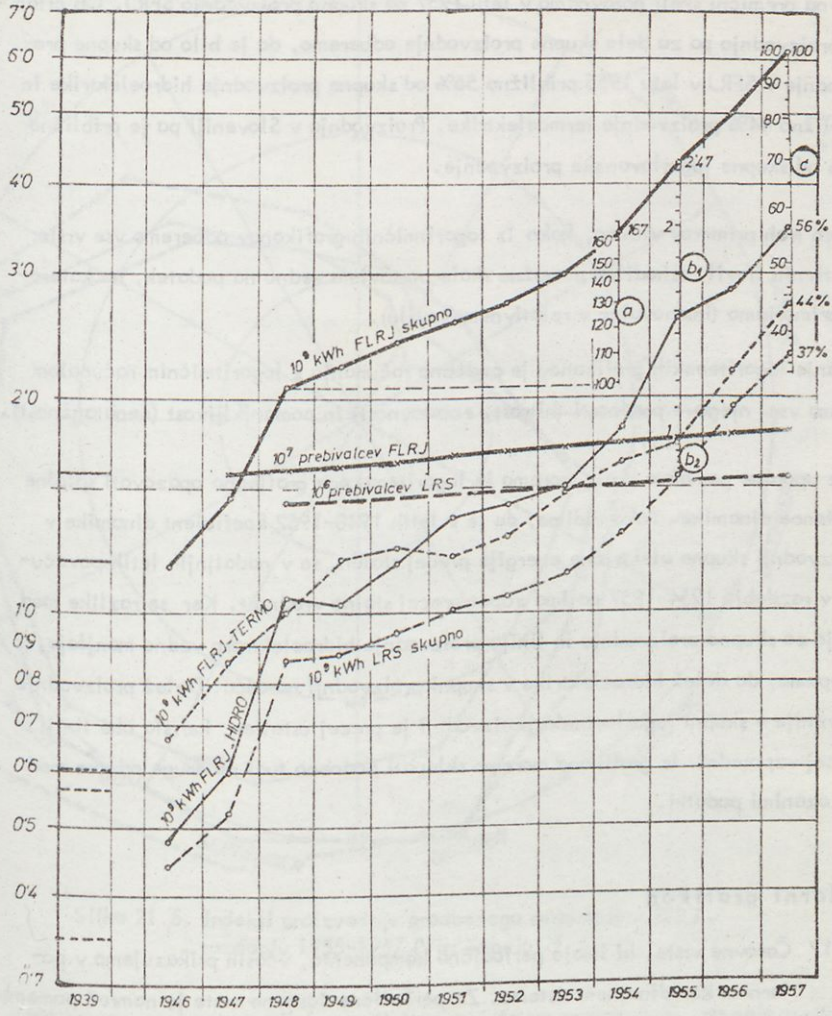
Leto	Proizvodnja električne energije v milj. KWh				Prebivalstvo v milj. prebiv.	
	SFRJ skupno	Hidro	Termo	SRS skupno	SFRJ	SRS
1939	1173	566	607	350	15,696	..
1946	1150	478	672	440	..	..
1947	1453	598	855	523	15,679	..
1948	2061	1053	1008	867	15,901	1,443
1949	2214	1023	1191	878	16,133	1,457
1950	2408	1175	1233	937	16,346	1,473
1951	2550	1357	1193	1003	16,588	1,492
1952	2700	1423	1277	1047	16,798	1,498
1953	2982	1500	1482	1121	17,048	1,515
1954	3440	1810	1630	1293	17,318	1,521
1955	4340	2610	1730	1555	17,586	1,531
1956	6252	3522	2730	2306	18,005	1,550
1957	6252	3522	2730	2306	18,005	1,550

Logaritemska skala na grafikonu v sliki 21.4 je neimenovana in velja za vse prikazane časovne vrste ne glede na velikost in vrsto podatka. Za vsako vrsto posebej pa v grafikonu vpišemo decimalni faktor; z njim je treba vrednost, ki jo odberemo na skali, pomnožiti, da dobimo vrednost količine v enoti mere, ki je naznačena za decimalnim faktorjem. Iz grafikona 21.4 tako odberemo, da je bila skupna proizvodnja električne energije leta 1948 $2,06.10^9$ KWh ali 2,06 milijard KWh. Ker je logaritemsko skalo teže brati kakor aritmetično, logaritemske grafikone običajno proučujemo s pomožno premično skalo, ki jo vrišemo na rob pasu papirja. To pomožno skalo uporabljamo predvsem za določanje relativnih števil (indeksov, koeficientov in struktur).

Na sliki 21.4 je naznačeno, kako s pomožno skalo odberemo relativna števila. V zgledu, ki je označen z a, odberemo indeks skupne proizvodnje električne energije za SFRJ za leto 1954 na bazo 1948 tako, da izhodišče pomožne skale 100 naravnamo na osnovo 1948. Na pomožni skali odberemo tam, kjer linija preseka skalo, indeks ca 165 (natančna vrednost je 167).

Z b_1 je naznačeno, kako določimo koeficient proizvodnje električne energije na enega prebivalca v SFRJ v letu 1955. Ta koeficient odberemo tako, da izhodišče 1 na pomožni skali naravnamo na linijo števila prebivalstva, na skali pa pri liniji proizvodnje odčitamo ustrezno vrednost 2,47. Vrednost koeficienta pa dobimo, če to vrednost pomnožimo s kvocientom 10^9 KWh/10^7 prebivalcev; tega dobimo iz decimalnih faktorjev, ki so vpisani ob ustreznih linijah za proizvodnjo in število prebivalstva. Koeficient je torej 247 KWh/preb. Analogen koeficient za isto leto za SR Slovenijo dobimo, če premaknemo izhodišče na pomožni skali na črto za prebivalstvo v Sloveniji, ob črti za proizvodnjo pa na pomožni skali odberemo vrednost 1,02. Po istem pravilu kot prej to vrednost pomnožimo s kvocientom 10^9 KWh/10^6 prebivalcev. Iz grafikona ocenjeni koeficient je torej 1020 KWh/preb. v letu 1955 v SR Sloveniji.

Kako s pomožno logaritemsko skalo odčitamo strukturne odstotke, je nakazano v zgledu c. Ker iščemo strukturne deleže za skupno proizvodnjo v SFRJ, izhodišče



Slika 21.4 Proizvodnja električne energije in število prebivalstva v SFRJ in SRS v razdobju 1946-1957

100 na premični skali naravnamo v letu 1957 na skupno proizvodnjo SFRJ. Ob črtah za proizvodnjo pa za dele skupne proizvodnje odberemo, da je bilo od skupne proizvodnje v SFRJ v letu 1955 približno 56% od skupne proizvodnje hidroelektrike in približno 44% proizvodnje termoelektrike. Proizvodnja v Sloveniji pa je približno 37% od skupne jugoslovanske proizvodnje.

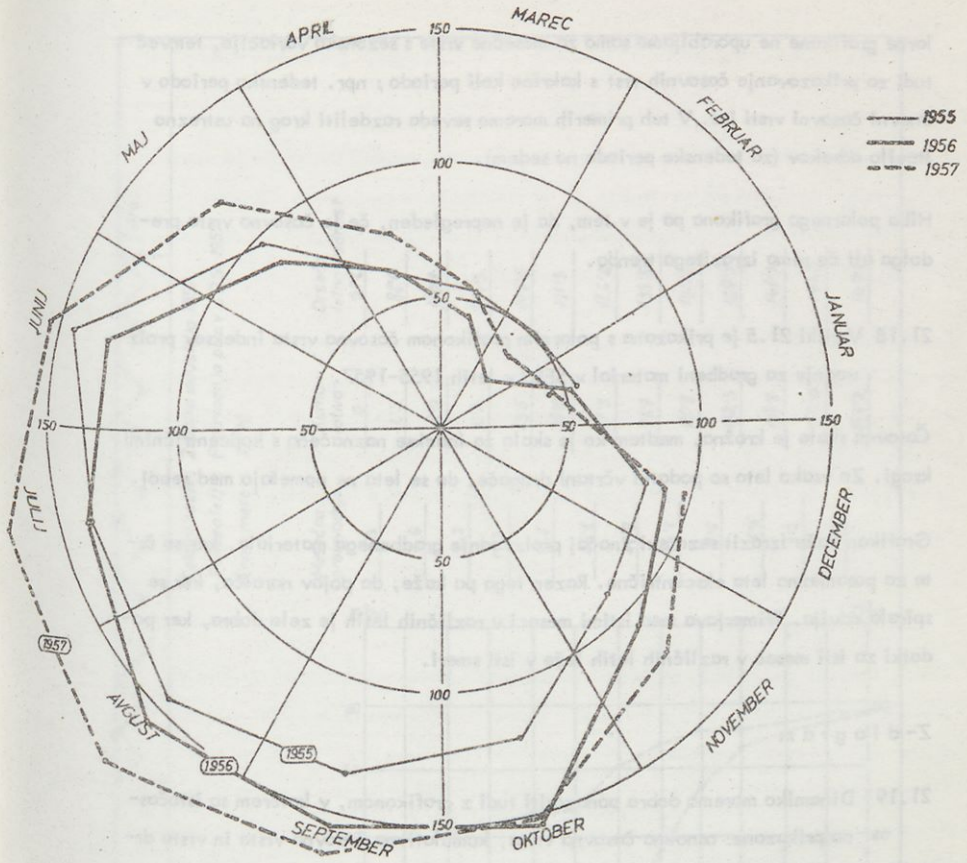
Iz teh treh primerov vidimo, kako iz logaritmičnih grafikonov odberemo vse vrste relativnih števil. Izhodišče pomožne skale postavimo vedno na podatek, na katerega primerjamo (imenovalec v relativnem številu).

Čitanje logaritemskih grafikonov je podobno računanju z logaritmičnim računalom in ima vse njegove prednosti (hitrost, enostavnost) in pomanjkljivost (nenatančnost).

Brez uporabe pomožne skale moremo iz logaritemskega grafikona opazovati splošne tendence dinamike. Tako vidimo, da je v letih 1948-1952 koeficient dinamike v proizvodnji skupne električne energije precej stalen, se v nadaljnjih letih povečuje, v razdobju 1954-1957 pa ima zopet precej stalno vrednost. Ker se razlike med linijo za skupno proizvodnjo in linijo proizvodnje hidroelektrarn vedno manjšajo, sklepamo, da delež hidroelektrike v skupni proizvodnji narašča. Delež proizvodnje Slovenije v skupni jugoslovanski proizvodnji je precej ustaljen, ker sta obe liniji precej vzporedni. Iz grafikona moremo sklepati podobno tudi za druge odnose med prikazanimi podatki.

Polarni grafikon

21.17 Časovne vrste, ki imajo periodično komponento, včasih prikazujemo v polarnem koordinatnem sistemu. Za periodične časovne vrste je namreč pomembnejša in bolj smiselna primerjava med istoležnimi členi za različne periode kakor pa primerjava med zaporednimi členi v časovni vrsti. S polarnim grafikonom najbolje primerjamo ustrezne člene v različnih periodah. Osnova polarnega grafikona je polarni koordinatni sistem; ta zamenja pravokotni koordinatni sistem. V polarnem grafikonu je čas znotraj ene periode nakazan s smerjo, podatek pa z oddaljenostjo točke od izhodišča polarnega koordinatnega sistema. Če imamo mesečno časovno vrsto s sezonsko variacijo, krog razdelimo v dvanajst enakih delov. Vsak izmed teh delov pomeni po en mesec. Če se podatki nanašajo na konce ali začetke mesecev, rišemo



Slika 21.5. Indeksi proizvodnje gradbenega materiala v SFRJ v razdobju 1955-1957 (Vir: Tabela 21.5)

podatke na meje posameznih krogovih izsekov, če pa se nanašajo na cele mesece ali na sredine mesecev, pa v sredino krogovih izsekov.

Za časovne vrste z izrazito sezonsko komponento kaže polarni grafikon tipično ekscentrično sliko. Če pojav narašča, dobimo na polarnem grafikonu spiralo, ki se odvija. Za pojave, ki upadajo, pa se spirala zavija proti izhodišču. Seveda po-

larne grafikone ne uporabljamo samo za mesečne vrste s sezonsko variacijo, temveč tudi za prikazovanje časovnih vrst s kakršno koli periodo; npr. tedensko periodo v dnevni časovni vrsti itd. V teh primerih moramo seveda razdeliti krog na ustrezno število odsekov (za tedenske periode na sedem).

Hiba polarnega grafikona pa je v tem, da je nepregleden, če je časovna vrsta predolga ali če nima izrazitega trenda.

21.18 V sliki 21.5 je prikazana s polarnim grafikonom časovna vrsta indeksov proizvodnje za gradbeni material v SFRJ v letih 1955-1957.

Časovna skala je krožna, medtem ko je skala za indekse naznačena s koncentričnimi krogi. Za vsako leto so podatki včrtani drugače, da se leta ne pomešajo med seboj.

Grafikon kaže izrazit sezonski značaj proizvodnje gradbenega materiala, ker so črte za posamezna leta ekscentrične. Razen tega pa kaže, da pojav narašča, ker se spirala odvija. Primerjava med istimi meseci v različnih letih je zelo dobra, ker podatki za isti mesec v različnih letih leže v isti smeri.

Z-d i a g r a m

21.19 Dinamiko moremo dobro ponazoriti tudi z grafikonom, v katerem so istočasno prikazane: osnovna časovna vrsta, kumulativna časovna vrsta in vrsta drsečih vsot. Te časovne vrste se med seboj prepletajo v obliki črke Z. Zato imenujemo te vrste grafikona Z-diagrame. Z-diagram je operativno sredstvo za prikazovanje in analiziranje dinamike ekonomskih pojavov. Zanj je tipična standardna oblika, ki ima naslednje značilnosti: Z-diagram je kartoteka enotnega formata, na kateri je en del namenjen za tekoče vpisovanje podatkov za vse tri časovne vrste Z-diagrama, na drugem delu pa je, potisnjen na rob, grafikon za vse tri časovne vrste (osnovno, kumulativno in vrsto drsečih vsot). Grafikon obsega zaključeno obdobje (dvanajst mesecev za mesečne podatke ali 31 dni za dnevne podatke). Skala za osnovne podatke je običajno v drugem merilu kakor skala za kumulativo in vrsto drsečih vsot.

Slika 21.6. Z-diagram za proizvodnjo piva v SRS v letu 1957 (1/2 stvarne velikosti).

Za podatke o proizvodnji piva v SRS v letu 1957 je Z-diagram prikazan v sliki 21.6.

Z-diagram je zelo prikladen in priljubljen v operativni evidenci in statistiki v podjetju. Sistem Z-diagramov za najrazličnejše količinske in vrednostne elemente proizvodnje za daljša razdobja omogoča elastično prikazovanje in analiziranje dinamike in povezave med različnimi elementi proizvodnje.

Ker je grafikon namenoma potisnjen v kot, moremo s polaganjem kartonov za več let drugega poleg drugega dobiti sliko dinamike za posamezni element za daljše razdobje. Če polagamo kartone za različna leta drugega nad drugega, moremo proučiti sezonsko variiranje pojava. Če pa polagamo kartone za različne podatke, ki so v vsebinski zvezi, drugega čez drugega, pa analiziramo povezave in odnose med posameznimi elementi.

Brunsmanov grafikon

21.20 Časovne vrste, ki se vežejo po metodi salda (korespondirajoče populacije) prikazujemo z Brunsmanovim grafikonom. Časovnih vrst korespondirajočih populacij je veliko: število delavcev v začetku meseca + na novo došli - odšli = število delavcev konec meseca; zaloge v začetku meseca + nabavljeno - prodano = zaloga konec meseca; zaloga surovin v začetku dekade + nabavljeno - predelano = zaloga surovin konec dekade; število prebivalstva v začetku leta + število rojenih + število doseljenih - število umrlih - število izseljenih = število prebivalstva konec leta.

V Brunsmanovem grafikonu trenutne vrste (število delavstva, zaloge, število prebivalcev itd.) prikažemo z linijskim grafikonom, razmične (na novo došli delavci, porabljene zaloge, število izseljenih itd.) pa s stolpci. Stolpci in linije so kombinirani tako, da se vežejo med seboj.

21.21 Z Brunsmanovim grafokonom prikažemo mesečne podatke o gibanju delavstva v podjetju A (podatki so izmišljeni):

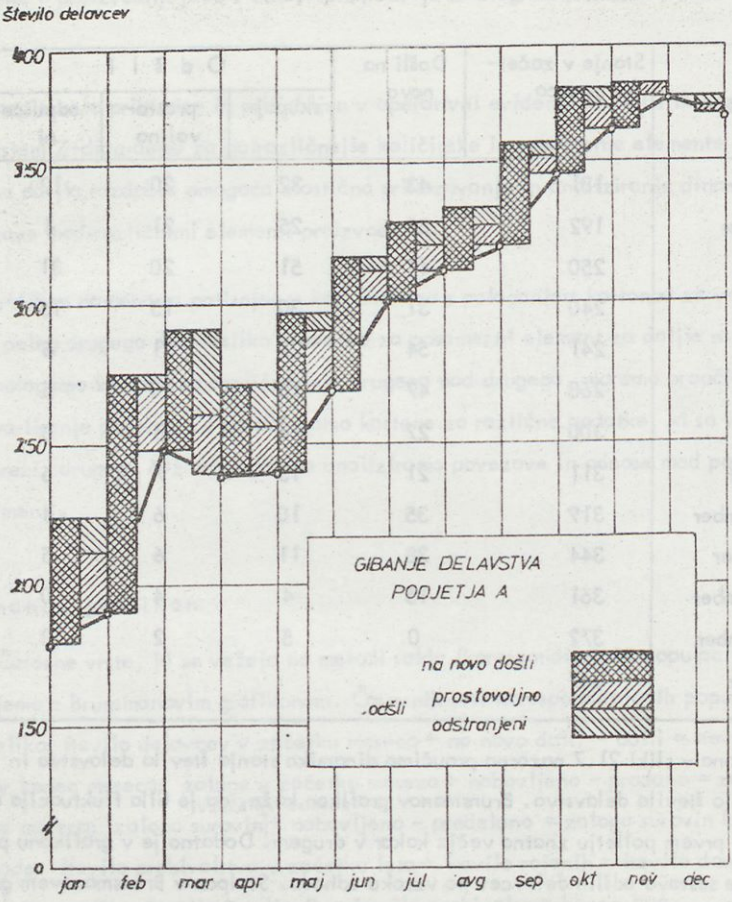
Tabela 21.7. Gibanje števila delavstva v podjetju A

Mesec	Stanje v začetku meseca	Došli na novo	Odšli		
			skupaj	prostovoljno	odpuščeni
Januar	181	43	32	20	12
Februar	192	83	25	21	4
Marec	250	41	51	20	31
April	240	31	30	15	15
Maj	241	54	27	21	6
Junij	268	47	15	10	5
Julij	300	27	16	8	8
Avgust	311	21	13	7	6
September	319	35	10	6	4
Oktober	344	28	11	6	5
November	361	15	4	4	0
December	372	0	5	2	3
Januar	367				

Iz grafikona v sliki 21.7 nazorno proučimo dinamiko stanja števila delavstva in fluktuacijo števila delavstva. Brunsmanov grafikon kaže, da je bila fluktuacija delavstva v prvem polletju znatno večja kakor v drugem. Dodatno je v grafikonu prikazana še sestava odšlih delavcev po vzroku odhoda. Stolpce v Brunsmanovem grafikonu uporabljamo za prikazovanje sestav tudi za druge pojave. Pri prebivalstvu stolpec za povečanje prebivalstva razdelimo na novorojene in priseljene, stolpec zmanjšanja prebivalstva na umrle in izseljene itd.

Ganttov grafikon

21.22 Kumulativne časovne vrste ima za osnovo Ganttov grafikon, ki ga uporabljamo pri kontroli plana. Z njim kompleksno analiziramo in kontroliramo izpolnjenje plana za najrazličnejša področja.



Slika 21.7. Brunsmanov grafikon za gibanje števila delavstva v podjetju A.

Osnovno pri Ganttovem grafikonu je to, da dejansko količino v razdobju od začetka planskega razdobja do določenega trenutka merimo s časom, v katerem bi bila ta količina realizirana, če bi proces potekal ves čas po planu. Ta čas je večji kot dejanski, če je plan prekoračen; se ujema z dejanskim, če je plan izpolnjen in je manjši kakor dejanski, če plan ni izpolnjen.

S shematičnim zgledom v tabeli 21.8 je pokazano, kako izračunamo Ganttov odstotek o izpolnjenju plana, v sliki 21.8 pa je narisan Ganttov grafikon za ta zgled.

Tabela 21.8. Izračunanje Ganttovega odstotka o izpolnjenju plana

Četrtletje	Plan		Dejansko		Odstotek izpolnjenje plana	
	četrtletno P_k	kumulativno P_k	četrtletno d_k	kumulativno D_k	četrtletni	Ganttov
I	200	0	150	0	75	0
II	250	200	240	150	96	075
III	280	450	360	390	129	176
IV	300	730	250	750	83	307
		1030		1000		390

Ganttov odstotek računamo po tehle stopnjah:

a) Zraven časovne vrste o planu izračunamo kumulativno vrsto plana.

b) V isto tabelo vpisujemo postopoma kakor dobivamo podatke, dejansko doseženo in kumulativo teh podatkov.

c) Za tekočo vrednost kumulative iz dejanskih količin v kumulativi za plan odredimo, med kateri vrednosti planske kumulative pade vrednost dejanske kumulative.

Do začetka tretjega četrtletja smo npr. dejansko proizvedli 390 enot. Ta vrednost pade v planski kumulativi med 200 in 450. Iz tega sklepamo: Če bi delali po planu, bi 390 enot delali vse prvo četrtletje, ostanek 190 enot pa bi izdelovali 190/250 del drugega četrtletja, ali v odstotkih 76% drugega četrtletja. Ganttov odstotek napišemo za ta primer v troštevilčnem številu 176. Pri tem pomeni prvo mesto število polnih osnovnih razdobij, zadnji dve mesti pa odstotek naslednjega razdobja.

Iz tabele 21.8 vidimo dalje, da smo do začetka četrtega četrtletja, za katerega je Ganttov procent 307, izdelali količino predvideno po planu za prva tri četrtletja in 7% količine, predvidene za četrto četrtletje.

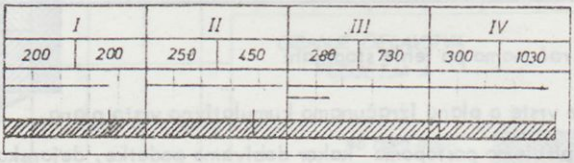
Ako s p_k zaznamujemo člene časovne vrste za plan, s P_k njegove kumulative, z d_k člene v časovni vrsti za dejansko opravljeno, z D_k pa kumulativo te časovne vrste, v simbolih izrazimo računanje Ganttovega odstotka takole:

a) Poiščemo, med katera člena P_k in P_{k+1} v planski kumulativi pade vrednost dejanske kumulative D_r.

$$P_k \leq D_r < P_{k+1} \qquad (21.9)$$

b) Ganttov odstotek izračunamo po obrazcu: Prvo mesto Ganttovega odstotka je k - 1. Naslednji dve mesti pa izračunamo po obrazcu:

$$G_k = \frac{D_r - P_k}{P_k} \cdot 100 \qquad (21.10)$$



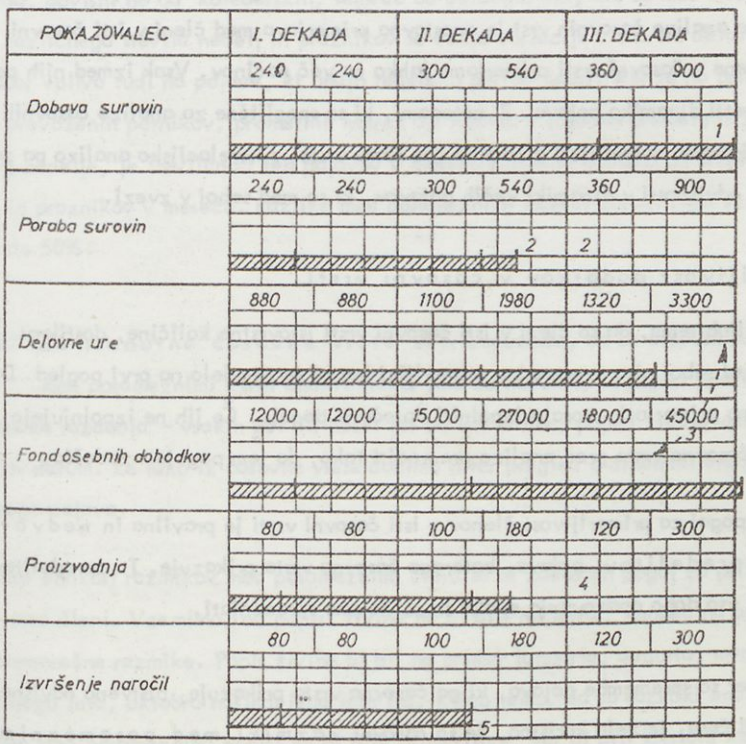
Slika 21.8. Ganttov grafikon

Ganttovi odstotki pa so samo vmesni rezultat za Ganttov grafikon. Te odstotke namreč rišemo v grafikon, ki ima za vsak element svoj pas. Ganttov odstotek včrtujemo v grafikon z vodoravnim stolpcem. Pas se vleče skozi Ganttovemu odstotku ustrezno število celih razmikov in skozi ustrezni del načetega razmika. Običajno razen Ganttovega odstotka rišemo v grafikon tudi navadne odstotke o izpolnjenju plana v posameznem razdobju. Ti nakazujejo izpolnjenje plana v osnovnih razdobjih. V sliki 12.8 so prikazani rezultati za naš zgled v standardni obliki Ganttovega grafikona. Vanj postopoma s črto vrisujemo od razdobja do razdobja navaden odstotek izpolnjenja plana, s pasom pa Ganttov odstotek. Kot vidimo iz slike 12.8 so v grafikon vpisani podatki o planu in kumulativi plana. Tako dobimo predstavo o velikosti podatkov. Razen tega so posamezni osnovni časovni razmiki (meseci, četrtletja itd.)

razdeljeni na pet enakih delov. Vsak od njih pomeni 20% od osnovnega razdobja. Tako si olajšamo risanje in branje grafikona. Konec stolpca za Ganttov odstotek za določen razmik zaznamujemo s črto in postopoma dorisujemo stolpec.

Prednost Ganttovega grafikona je predvsem v tem, da so podatki za vsak element naneseni v pasu. Zato moremo na isto sliko nanesti niz elementov: en element za več podjetij, ali različne elemente za isto podjetje. Tako moremo kompleksno analizirati izpolnjenje plana tudi v odnosih med elementi proizvodnje, ne pa samo izolirano za posamezne elemente proizvodnje.

21.23 V sliki 21.9 je prikazan zgled kompleksnega Ganttovega grafikona za šest sestavin za tri dekade. Proučitev Ganttovega grafikona nam pokaže niz po-



Slika 21.9. Kompleksen Ganttov grafikon

membnih odnosov med posameznimi podatki. Dobava surovin je razen v prvi dekadi dosti večja kot poraba. Delovnih ur in osebnih dohodkov je bilo posebno v tretji dekadi nesorazmerno s proizvodnjo, kar moremo tolmačiti s tem, da so v tretji dekadi delali na remontu, pri katerem je sodelovala predvsem visokokvalificirana delovna sila (fond osebnih dohodkov je nesorazmeren z opravljenimi delovnimi urami /3/). Izvršitev naročil kaže, da je bila proizvodnja v prvih dveh dekadah sproti tudi oddana (4 in 5), medtem ko je proizvodnja iz tretje dekade ostala.

ANALIZA ČASOVNIH VRST

21.24 Dinamiko pojavov analiziramo tako, da opazujemo spreminjanje vrednosti členov v časovnih vrstah in iščemo zakonitosti tega spreminjanja. Naloga enostavne analize časovnih vrst je enostavna primerjava med členi v isti časovni vrsti. Člene v časovni vrsti primerjamo lahko na več načinov. Vsak izmed njih po svoje osvetli dinamiko pojava. Z metodami, ki so specifične za analizo časovnih vrst analiziramo zakonitosti dinamike ene same vrste, s korelacijsko analizo pa zakonitosti odvisnosti v dinamiki večih pojavov, ki so med seboj v zvezi.

Primerljivost podatkov v časovni vrsti

21.25 Kljub temu, da so členi v isti časovni vrsti istovrstne količine, dostikrat med seboj niso neposredno primerljivi kot bi se to zdelo na prvi pogled. Da jih moremo primerjati, morajo izpolniti določene pogoje. Če jih ne izpolnjujejo, je treba časovno vrsto pred analizo preurediti tako, da tem pogojem zadošča.

Osnovni pogoj za primerljivost členov v isti časovni vrsti je pravilna in nedvoumna opredelitev pojava, katerega časovna vrsta prikazuje. Ta opredelitev mora biti vso dobo opazovanja enaka in se ne sme spreminjati.

21.26 Ker so spremembe pojava, ki ga časovna vrsta prikazuje, bistveno odvisne od časa, je zelo koristno, če so časovni razmiki med posameznimi členi enaki. Tako dosežemo, da je pri enakem delovanju dejavnikov, spremem-

ba pojava od člena do člena enaka; tega v nasprotnem primeru ne bi bilo. Zato imamo popise prebivalstva v razdobjih pet ali deset let, popise živine vsako leto na isti datum itd.

Kljub temu, da ta pogoj na prvi pogled ni težko izpolniti, so v določenih primerih težave. Problem so namreč osnovne vrste, za katere se člani nanašajo na krajša razdobja, mesece, dekade, tedne itd. Število rojstev po mesecih ne moremo neposredno primerjati, ker se število koledarskih dni v mesecu spreminja od 28 do 31 dni. Razlika med najkrajšim in najdaljšim mesecem je torej tri dni ali približno 10%. Če bi bili vsi ostali dejavniki, ki vplivajo na število rojstev, stalni, bi bilo število rojstev od meseca do meseca različno tudi za več kot 10% zaradi vzroka, ki nima nobene zveze z rodnostjo. Te razlike so še večje pri pojavih, ki so, kakor na primer proizvodnja, odvisni ne od koledarskih, temveč od delovnih dni, ker so med meseci zaradi različnega števila nedelj in praznikov še večje variacije v številu delovnih dni. Koledar vpliva tudi na pojave, ki imajo tedensko periodičnost. Ker je na primer število prevoženih potnikov, prometnih nesreč ali nočnin v tujskem prometu največje okrog nedelje, je številčnost teh pojavov v mesecu bistveno odvisno od števila nedelj in praznikov v mesecu. Razlike med posameznimi meseci zaradi tega vzroka so tudi do 50%.

21.27 Za trenutne časovne vrste si prizadevamo, da so časovni razmiki med posameznimi členi enaki. Kakor smo že navedli, so popisi prebivalstva na enaka razdobja - vsakih pet ali deset let na isti datum, popisi živine vsako leto na isti datum. Le tako iz časovne vrste dobimo hiter pregled o dinamiki enega in drugega pojava.

Vendar enakost razmikov med posameznimi členi še ni zadosten pogoj za primerljivost med členi. Vzemimo, da popisi živine ne bi bili na letne, temveč na primer sedemmesečne razmike. Popis živine bi bil na primer januarja, avgusta, marca naslednjega leta, oktobra naslednjega leta itd. Kljub temu, da so razmiki enaki, spremembe med posameznimi popisi ne bi bile primerljive, ker bi bili kritični momenti

v različnih delih leta. To pa bistveno vpliva na stalež živine. Da se temu ognemo, izvajamo popise v razmikih, ki so mnogokratniki periode - v našem primeru leta.

21.28 Čeprav so razmiki med členi pri mesečnih ali četrtletnih trenutnih časovnih vrstah različni, je učinek razlik na časovno vrsto tako malenkosten, da podatkov ne popravljamo.

Pri razmičnih časovnih vrstah pa je ta učinek znaten in moramo podatke pred analizo preračunati na enake razmike. To dosežemo tako, da za osnovne podatke, ki se nanašajo na dejanska mesečna razdobja, izračunamo namesto mesečnih vrednosti poprečja na en dan, enotin mesec s 30 koledarskimi ali 25 delovnimi dnevi. Tako odstranimo učinek različne dolžine meseca. Ali preračunamo podatke na 25 ali 30 dni, je odvisno od pojava, ki ga analiziramo. Rojstva preračunamo na koledarski mesec 30 dni, enako proizvodnjo v panogah, ki imajo zvezen delovni proces; na mesec 25 dni pa preračunavamo proizvodnjo v panogah in podjetjih, v katerih je delovni proces prekinjen. Tako se najbolj prilagodimo dejanskemu številu dni v posameznih mesecih.

Popravni faktorji za posamezne vrste poprav so nakazani v tabeli 21.9.

Tabela 21.9 Popravni faktorji za mesečne časovne vrste.

Število koledarskih dni v mesecu	Popravni faktor	Število delovnih dni v mesecu	Popravni faktor
28	30/28 = 1,0714	22	25/22 = 1,1364
29	30/29 = 1,0345	23	25/23 = 1,0870
30	30/30 = 1,0000	24	25/24 = 1,0417
31	30/31 = 0,9677	25	25/25 = 1,0000
		26	25/26 = 0,9615
		27	25/27 = 0,9259

Podatke mesečnih časovnih vrst preračunamo na enotino dolžino meseca tako, da jih pomnožimo z ustreznimi popravnimi faktorji. Vsote popravljenih podatkov so

brez smisla, ker so popravljeni podatki povprečje, čeprav so osnovni podatki razmičnega značaja.

21.29 Reduciranje števila rojstev v SRS v letu 1969 na enotine mesece (Vir: MSP)

Mesec	Število rojstev Y	Število dni	Popravni faktor F	Popravljeno število rojstev $VF = Y_f$
J	2143	31	0,9677	2074
F	2275	28	1,0714	2437
M	2546	31	0,9677	2464
A	2480	30	1,0000	2480
M	2454	31	0,9677	2375
J	2190	30	1,0000	2190
J	2399	31	0,9677	2322
A	2248	31	0,9677	2175
S	2283	30	1,0000	2283
O	2388	31	0,9677	2311
N	1997	30	1,0000	1997
D	2775	31	0,9677	2685

21.30 Na velikost pojavov dostikrat vplivajo administrativni ukrepi, ki z vsebino proučevanega pojava nimajo neposredne zveze. Eden izmed običajnih vzrokov so upravno-teritorialne spremembe, s katerimi se spremeni geografska opredelitev pojava. Taka sprememba je lahko sprememba državnih meja, ki je posledica vojne, ali sprememba upravno-teritorialnij mej znotraj države. Različna geografska opredelitev onemogoča primerljivost podatkov v časovni vrsti. V tem primeru je potrebno podatke časovne vrste za nazaj preračunati na novo območje, kar je običajno težaven in zamuden posel. Včasih si pomagamo tako, da za čas, ko je nastopila sprememba, navedemo podatek po starem in novem stanju. Tako vidimo, kolika je ob prehodu sprememba zaradi spremembe v teritoriju.

Če časovno vrsto za nazaj ne popravimo, moramo proučiti vsak del posebej, nikakor pa ne smemo proučevati vrsto kot enotno.

Isti problem so tudi spremembe zaradi drugih administrativnih ukrepov, na primer ko je šlo za prehajanje podjetij iz ene pristojnosti v drugo itd. Tudi v teh primerih ravnamo s spremembami enako kakor pri teritorialnih spremembah.

21.31 Ker se število prebivalstva stalno spreminja, povečana proizvodnja še ne pomeni nujno povečane produktivnosti dela, povečana poraba še ne povečanje življenjske ravni itd.

Vpliv sprememb v spremenjenem številu prebivalstva odstranimo tako, da izračunamo koeficiente "na prebivalca", "na delavca" itd.

Časovno primerljivost podatkov včasih motijo tudi spremembe v sestavi. Sprememba starostne sestave prebivalstva, ki more nastopiti v daljšem časovnem razdobju, močno vpliva na umrljivost. Primerjava koeficientov mortalitete za daljše razdobje zato lahko da napačno sliko v spremembah v umrljivosti prebivalstva. Spremenjena sestava proizvodnje tako v podjetju kot v večjih gospodarskih združbah ali v celotnem narodnem gospodarstvu odločilno vpliva na sumarne koeficiente o produktivnosti dela, o poprečnih osebnih dohodkih itd. Vpliv spremenjene sestave na določene pojave v časovni vrsti odpravimo s standardizacijo podatkov, to je s koeficienti pri stalni sestavi. O standardizaciji koeficientov je bilo več govora pri koeficientih.

21.32 Proizvodnjo, promet v zunanji ali notranji trgovini in druge pojave izražamo v vrednosti. Časovno primerjavo vrednostnih podatkov motijo spremembe v cenah, ker se vrednost proizvodnje, prometa itd. spreminja zaradi sprememb v obsegu in sprememb cen. Vpliv, ki ga povzročajo spremembe cen, odstranimo tako, da izražamo proizvodnjo, promet itd. v stalnih cenah. Vendar je ta način izredno zamuden in težko izvedljiv, ker bi ga morali upoštevati že v osnovni evidenci. Vpliv cen popravimo tudi z indeksi cen tako, da podatke po tekočih cenah delimo z ustreznim indeksom cen. Ta način je operativno veliko lažji, je pa kljub temu, da je le ocena, zadovoljiv. Le paziti je treba, da je indeks, s katerim popravlja-

mo podatke, ustrezen po vsebini in značaju.

Enostavni pokazovalci dinamike

21.33 Enostavno analiziramo dinamiko pojava s primerjavo členov v časovni vrsti.

Najlaže primerjamo člene časovne vrste z indeksi. Zato z njimi izvedemo več elementarnih pokazovalcev dinamike.

Raven Y_k je osnovna časovna vrsta. Že iz sprememb v ravni sklepamo na smer in jakost dinamike pojava.

Če raven časovne vrste izrazimo v primerjavi z nekim stalnim značilnim členom v časovni vrsti, dobimo vrsto indeksov s stalno osnovo

$$I_{k/o} = 100\, Y_k / Y_o \tag{21.11}$$

Absolutna razlika

$$D_k = Y_k - Y_{k-1} \tag{21.12}$$

pokaže v absolutnih vrednostih spremembo pojava od člena do člena.

Temp rasti ali stopnja rasti

$$T_k = 100\,\frac{Y_k - Y_{k-1}}{Y_{k-1}} = 100\,\frac{D_k}{Y_{k-1}} \tag{21.13}$$

pokaže relativno razliko od člena do člena.

Koeficient dinamike

$$K_k = Y_k / Y_{k-1} \tag{21.14}$$

pokaže v obliki koeficienta relativne spremembe od člena do člena.

Enak značaj ima verižni indeks

$$I_k = 100\,.\,Y_k / Y_{k-1} \tag{21.15}$$

le, da relativno spremembo nakaže v indeksu.

Ker so vsi ti pokazovalci izpeljani iz osnovne časovne vrste, so med njimi enostavne zveze:

$$T_k = 100 \cdot D_k/Y_{k-1} = 100(K_k - 1) = I_k - 100$$
$$I_k = 100 \cdot K_k \qquad (21.16)$$

V tabeli 21.11 je prikazano, kolikšni so posamezni pokazovalci dinamike, če pojav raste, zastane ali pada.

Tabela 21.11. Vrednosti posameznih pokazovalcev dinamike pri različnem gibanju pojava

Pokazovalec dinamike	P o j a v		
	raste	zastane	pada
Raven Y_k	$Y_{k+1} > Y_k$	$Y_{k+1} = Y_k$	$Y_{k+1} < Y_k$
Indeks s stalno osnovo	$I_{k+1/o} > I_{k/o}$	$I_{k+1/o} = I_{k/o}$	$I_{k+1/o} < I_{k/o}$
Absolutna razlika	$D_k > 0$	$D_k = 0$	$D_k < 0$
Temp rasti	$T_k > 0$	$T_k = 0$	$T_k < 0$
Koeficient dinamike	$K_k > 1$	$K_k = 1$	$K_k < 1$
Verižni indeks	$I_k > 100$	$I_k = 100$	$I_k < 100$

21.34 Za svetovno proizvodnjo boksita imamo v tabeli 21.12 izračunane vse vrste elementarnih pokazovalcev dinamike.

Tabela 21.12. Pokazovalci dinamike za svetovno proizvodnjo boksita (v tisoč tonah ; vir: SG 58)

Leto	Y_k	$I_{k/o}$	D_k	T_k	K_k	I_k
1950	7.700	100				
1951	10.300	134	+ 2600	+ 33,8	1,34	134
1952	11.900	155	+ 1600	+ 15,5	1,16	116
1953	13.000	169	+ 1100	+ 9,2	1,09	109
1954	14.850	193	+ 1850	+ 14,2	1,14	114
1955	15.350	199	+ 500	+ 3,4	1,03	103
1956	16.900	219	+ 1550	+ 10,1	1,10	110

21.35 Nakazani pokazovalci dinamike pa niso edini pokazovalci, s katerimi prikazujemo dinamiko pojava. Pogosto prikažemo dinamiko tudi z drugačnimi primerjavami. Izmed teh naj omenimo samo mesečne časovne vrste za pojave, ki so sezonskega značaja. Če je pojav sezonski, primerjava tekočega meseca s predhodnim mesecem nima logične povezave in smisla. Bolj smiselno je, da primerjamo podatek tekočega meseca s podatkom iz predhodnega leta za isti mesec. Ta dva podatka sta si sorodnejša kakor zaporedna meseca. Ti indeksi imajo premično osnovo, vendar niso verižni, ker osnova indeksov ni predhodni člen.

Enako računamo tudi indekse iz kumulativnih vrednosti dveh zaporednih let. Ta primerjava je logična, ker obe primerjani vrednosti obsegata isti del sezone v dveh zaporednih letih.

V tabeli 21.13 je prikazana standardna primerjava z indeksi iz tekočih podatkov Indeksa za julij 1971. Za proizvodnjo vseh vrst premoga so izračunani štiri vrste indeksi:

$\frac{\text{I-VII 71}}{\text{I-VII 70}}$ = indeks iz kumulative proizvodnje do julija med letoma 1971 in 1970

$\frac{\text{VII 71}}{\text{Ø 70}}$ = indeks proizvodnje v tekočem mesecu (julij 71) s poprečjem preteklega leta;

$\frac{VII\ 71}{VII\ 70}$ = indeks tekočega meseca z istim mesecem preteklega leta (VII 70) in

$\frac{VII\ 71}{VI\ 71}$ = verižni indeks.

Tabela 21.13 Proizvodnja premoga v SFRJ v tisoč tonah

(Vir: Indeks VIII, 1971)

Vrsta premoga	Ø 70	I-VII 70	I-VII 71	VII 70	VI 71	VII 71	I-VII 71 / I-VII 70	VII 71 / Ø 70	VII 71 / VII 70	VII 71 / VI 71
Premog skupno	2369	15329	17539	2211	2247	2498	114	105	113	111
Črni	54	356	409	51	55	57	115	106	112	104
Rjavi	749	5147	5348	666	783	720	104	96	108	92
Lignit	1566	9826	11836	1494	1609	1721	120	110	115	107
Pogojni 4000 kal	1769	11578	13061	1638	1831	1839	113	104	112	100
Koks	109	753	757	114	107	109	101	100	96	102

Sestavine dinamike v časovnih vrstah

21.36 Časovna vrsta je številčen izraz časovnega delovanja vseh dejavnikov, ki vplivajo na pojav, ki ga prikazuje. Teh dejavnikov je veliko in se njihova jakost in učinek časovno spreminjata. Zato se členi časovne vrste spreminjajo; pravimo, da se pojav giblje.

Nemogoče je iz časovne vrste izluščiti, kolikšna je sprememba zaradi vsakega posameznega dejavnika posebej. Iz časovne vrste pa je moč razbrati skupen učinek dejavnikov, ki imajo soroden vpliv na pojav, ki ga proučujemo. Na časovni vrsti po tem vidiku opazujemo naslednje vrste sprememb:

a) trend - T, ki podaja osnovno smer razvoja,

b) ciklične spremembe - C, ki izvirajo iz dolgoročnih vzrokov,

c) periodične spremembe - P, ki izvirajo iz vzrokov, ki se ponavljajo na stalno razdobje - periodo,

d) iregularne spremembe - I, ki so rezultat enkratnih epizodičnih dogodkov ali rezultat stalnih slučajnih vzrokov.

Časovna vrsta ni vedno rezultat delovanja vseh sestavin. Ta ali ona sestavina ne nastopi bodisi zaradi tipa vrste ali ker ni vsebinsko pogojena. Časovna vrsta letnih podatkov ne kaže sezonskega gibanja, čeprav je pojav mogoče sezonski, marsikateri pojav nima cikličnih nihanj ali epizodičnih sprememb zaradi narave pojava. Vsak pojav pa ima neko osnovno smer razvoja. Le redki so primeri nespremenljivih pojavov. Na vsaki časovni vrsti opazimo slučajna nihanja, ki so izraz manjših vzrokov, ki nastopajo v eni ali drugi obliki.

21.37 T r e n d . Vsak ekonomski pojav ima časovno neko osnovno smer razvoja. Ta pa je opazna le v daljših časovnih razdobjih. Osnovno smer razvoja imenujemo trend. Če trend nakazuje razvoj za dolgo razdobje, ga včasih imenujemo s e k u l a r n i t r e n d . Trend je rezultat dejavnikov, ki pogojujejo stalen razvoj pojava. Tako je trend povečanja proizvodnje električne energije izraz tehničnega napredka, trend zniževanja umrljivosti rezultat razvoja zdravstvene službe in medicine nasploh itd. Trend proizvodnje električne energije kaže stalno naraščanje, trend gibanja umrljivosti pa stalno padanje. Trend pa more v določenem razdobju tudi spremeniti smer. Tako more proizvodnja nekega izdelka določeno razdobje naraščati, v nadaljnjem razvoju pa zaradi nadomestitve z drugim izdelkom, ki je bolj kakovosten, pada. Enako more proizvodnja v rudniku za neko razdobje naraščati, v nadaljnjem razdobju pa se zaradi osiromašenja rudnika zmanjšuje. Navedli smo že, da je konstantnih pojavov malo. Eden izmed njih je na primer spolni indeks - razmerje med številom rojstev dečkov in deklic. To razmerje ima časovno precej stalno vrednost brez težnje naraščanja ali padanja. Odkloni od konstante so le slučajni in ne kažejo težnje k časovni spremembi tega razmerja. Ta primer je prikazan v drugem poglavju v tabeli 2.1.

Zaradi drugih vplivov se dejanske vrednosti od trenda odklanjajo navzgor in navzdol, vendar v večini primerov že iz slike za časovno vrsto razberemo osnovno smer razvoja.

21.38 Periodična nihanja. Za veliko pojavov je značilen periodičen, zlasti sezonski značaj. Sezonski značaj marsikaterih pojavov povzroče predvsem klimatski vplivi. Tako je s klimatskimi vplivi povezana gradbena dejavnost, turizem in gostinstvo, kmetijstvo in vse druge dejavnosti, ki so odvisne od klimatskih vplivov ali posredno povezane z eno izmed navedenih. Prehrambena industrija je vezana na poljedelstvo, promet na turizem itd. Sezonski značaj pa ne izvira nujno iz klimatskih vplivov, marveč more biti vzrok tudi drugje. Tako vplivajo ustaljene navade ali prireditve na sezonski značaj pojavov. Potniški promet, poštni promet, trgovinski promet itd. je odvisen od praznikov, ki so v vsakem letu ob istem času (npr. Novo leto). Enako izzove sezonski značaj določenega pojava prireditev, ki se letno ponavlja na isti datum; tak zgled so npr. gospodarski velesejmi. Pri najrazličnejših pojavih pa opazimo periodična nihanja, za katere je perioda krajša kot leto. Tako opazujemo npr. v blagovnem prometu mesečna nihanja, v prometu tedenska ali dnevna nihanja ipd.

21.39 Ciklična nihanja. V časovnih vrstah za daljša razdobja opazimo nihanja okrog trenda. Ta nihanja so več ali manj regularna, ni pa niti njihova dolžina niti oblika stalna kakor pri periodičnih nihanjih. Ta nihanja pa kljub temu, da niso popolnoma regularna, niso slučajna in so odvisna od dogajanj v preteklosti. Imenujemo jih ciklična gibanja. Ciklična gibanja so tipična za ekonomske pojave in jih zasledimo na nizu pojavov. Jakost cikličnih nihanj je za različne pojave različna. Odvisna je od tega, ali gre za pojav, ki je bolj ali manj občutljiv za spremembe v gospodarskem življenju. Zaradi povezanosti gospodarstva je med cikli za posamezne gospodarske panoge ali pojave zveza, ker se tako prosperiteta kot depresija prenaša od panoge na panogo.

21.40 Iregularne variacije. Razen navedenih treh vrst gibanj, trenda, periodičnih nihanj in cikličnih sprememb opazujemo na časovni vrsti še iregularne variacije, ki so rezultat enkratnih ali pa slučajnih vzrokov. Zaradi prekinitve toka nastane kratkoročen zastoj v prometu električne cestne železnice. Potres vpliva na niz dejavnosti. Zaradi prometne nesreče je zastoj v železniškem prometu,

zaradi poplave je uničena letina, epidemija vpliva na morbiditeto in mortaliteto itd. Navedeni dogodki so e n k r a t n i v p l i v i , ki imajo za rezultat, da se za krajše razzobje pojav odkloni od rednega poteka.

Za razliko od enkratnih - epizodičnih vplivov, ki nastopijo nepričakovano, vendar za krajši čas, nastopajo stalno in v vseh pojavih s l u č a j n e v a r i a c i j e , ki so rezultat slučajnih vplivov. Slučajne variacije so manjše spremembe, ki so rezultat nepomembnejših vplivov. Slučajne variacije le redko obravnavamo posamično, ker za to bodisi ni možnosti ali potrebe.

Osnovni modeli časovnih vrst

21.41 V vsaki časovni vrsti ne opazimo vseh sestavin, bodisi da nekatere sestavine pojav sploh ne vsebuje ali pa so v časovni vrsti zabrisane. Marsikateri pojav nima sezonske oziroma periodične sestavine, v nekaterih časovnih vrstah pa je zabrisana. Tako na dnevni časovni vrsti novorojenih ne opazimo tedenske periodičnosti, ker je ni, medtem ko je v tedenski vrsti števila prevoženih potnikov zabrisana tedenska periodičnost. Enako je v vrsti letnih podatkov zabrisano sezonsko gibanje pojava. Marsikateri pojav nima cikličnih nihanj ali epizodičnih sprememb. Skoraj vsak pojav pa ima svoj trend, le malo je pojavov, ki bi bili stalni v tem smislu, da nimajo neke smeri razvoja. Na vsaki časovni vrsti moremo opaziti tudi slučajne variacije, ki so izraz manjših vzrokov, ki vedno nastopajo.

Vrednosti členov v časovni vrsti Y so torej skupnost trenda T, cikličnih gibanj C, sezonskih oziroma periodičnih vplivov P in iregularnih variacij, ki so epizodične E ali slučajne S.

Zveze teh sestavin moremo nakazati v nekaj osnovnih modelih, ki jih uporabljamo pri analizi časovnih vrst:

1) $Y = T + P + C + E + S$

2) $Y = T\,(1 + p + c + e + s)$

3) $Y = T \cdot C\,(1 + p + e + s)$ (21.17)

4) $Y = T \cdot P \cdot C \cdot E \cdot S$

5) $Y = T \cdot P \cdot C \cdot E + S$

V prvem modelu so sestavine vezane aditivno. Enako je v drugem modelu, vendar je v njem učinek vsake sestavine sorazmeren s trendom. Tretji model ima za osnovo faktorialno povezavo med trendom in ciklom, medtem ko četrti predpostavlja faktorialno povezavo vseh sestavin. Pri petem so vse sestavine, razen slučajne, vezane faktorialno.

Če v stvarnem primeru, ki ga proučujemo, kake sestavine ni, ustrezni simbol v modelu izpade. V prvih dveh modelih so periodične in ciklične sestavine in iregularne variacije (epizodične in slučajne) pozitivne ali negativne količine. Če učinka kake sestavine ni, je njihova vrednost enaka 0. V prvem modelu so posamezne količine (P, C, E, S) izražene z absolutnimi, v drugem (p, c, e, s) pa z relativnimi odkloni od trenda. V tretjem modelu je trend dan absolutno, cikel s koeficientom, ki je glede na smer učinka nad ali pod 1, p, e in s pa v relativnem odklonu glede na skupen učinek trenda in cikla. V četrtem in petem modelu so sestavine P,C,E,S izražene s koeficienti, ki so ena, če učinka ni in večji ali manjši od ene, če deluje v smeri povečanja oziroma zmanjšanja. V stvarnosti prvi model srečamo poredko, čeprav je najenostavnejši. Težko si namreč predstavljamo, da bi bila velikost učinka posameznih sestavin neodvisna od velikosti pojava, ki je dan s trendom. Vendar je analiza prvega modela najenostavnejša. Zato skušamo druge modele prevesti na prvega, če se le da. Drugi model preuredimo v prvega, če osnovno časovno vrsto Y delimo s trendom, tretjega, če Y delimo s skupnim učinkom T.C. Četrti model prevedemo v prvi model z logaritmiranjem.

Osnovna analiza časovnih vrst gre za tem, da razdružimo osnovno časovno vrsto na njene sestavne dele. S tem prodremo že precej v zakonitosti dinamike pojava. Te

zakonitosti izkoriščamo, kadar analiziramo dogajanja v preteklosti ali napovedujemo razvoj v prihodnosti. Z različnimi analitičnimi postopki namreč moremo določiti posamezno izmed teh sestavin v časovni vrsti. Vendar moremo časovno vrsto analizirati le, dokler v kompleksu vplivov ni kakovostnih sprememb. To je jasno. Ako na primer neko novo odkritje da nove proizvodne možnosti, se smer razvoja bistveno spremeni. Ker imamo za take primere dve smeri razvoja, moramo časovno vrsto razdeliti v dva dela in ju obravnavati ločeno. Enako analiziramo sezonsko sestavino kot statično le, dokler moremo predpostavljati, da je nespremenjena. Če se v proučevanem razdobju sezonski vpliv spremeni, moramo časovno vrsto razdeliti na obdobja in analizirati sezonskost za vsako obdobje posebej.

Vloga poprečij pri proučitvi časovnih vrst

21.42 Enostavno orodje, ki pomaga proučiti časovne vrste, so poprečja. Ob določenih pogojih moremo s poprečji neke sestavine v časovni vrsti odstraniti, druge pa ohraniti. Poprečja iz časovnih vrst vplivajo v eni ali drugi obliki na vse sestavine.

Če je t r e n d v razmiku, iz katerega računamo poprečje, linearen, poprečje, ki ga centriramo na sredino razmika, leži na trendu. Če je trend na odseku, iz katerega računamo poprečje, krivuljčen, je poprečje, centriramo na sredino razmika, nad trendom, če je trend konkaven in pod trendom, če je trend konveksen. Razlike so tem večje, čimbolj je trend ukrivljen. Poprečje, centrirano na obratišče, pa leži na trendu. Iz tega sklepamo, da se časovna vrsta drsečih sredin, če jo izračunamo iz trenda, na odseku, na katerem je trend linearen, sklada s trendom, če pa je trend krivuljčen, dobimo črto, ki je bolj izravnana kot trend. Ker je krivina trenda na daljših odsekih večja, je izravnavanje trenda tem večje, čim širši so razmiki, za katere izračunavamo poprečja.

Izravnavanje ni tako izdatno, če člene, iz katerih računamo poprečja, tehtamo tako, da imajo robni členi manjšo, osrednji pa večjo težo. Običajno tehtamo člene z binomskimi koeficienti (1,2,1) za poprečja iz treh členov, (1,3,3,1) za poprečja iz štirih členov, (1,4,6,4,1) za poprečja iz petih členov. Izdelani so še različni drugi načini tehtanja.

Če je zveza med sestavinami aditivna, se učinek periodičnih oziroma sezonskih vplivov v poprečju uniči, če je dolžina razdobja, za katerega računamo poprečje, enako periodi ali mnogokratniku periode. Pri faktorialnih zvezah, kakor je nakazana v 12.17 v četrtem modelu, pa je nič poprečje logaritmov periodične sestavine.

V vrsti drsečih sredin se uniči ciklična sestavina, če so poprečja izračunana iz razdobja celotnega cikla.

Učinki slučajnih vplivov se v poprečju manjšajo, čim daljše je razdobje, iz katerega računamo poprečja. Če ta pravila izkoriščamo povezano, poprečje veliko pripomore k proučitvi časovnih vrst.

TREND

21.43 Trend proučujemo iz dveh razlogov. Samega zase proučujemo, da spoznamo smer razvoja pojava, da ga primerjamo s trendi za pojave, ki so med seboj odvisni ali pa da proučujemo vpliv trenda na druge sestavine (periodične in ciklične). Trend pa dostikrat iščemo zato, da iz odklonov stvarne časovne vrste od trenda proučujemo druge sestavine - sezonska, ciklična ali iregularna nihanja.

Metod za določanje trenda je več. Vsaka izmed njih ima za osnovo različne predpostavke in postopke. Zato je pomembno in odgovorno izbrati v vsakem posebnem primeru primerno obliko in metodo za določanje trenda. Trend določamo s prostoročnim vrisavanjem, z izravnavanjem s sredinami ali analitično s prilagajanjem analitičnih funkcij danim podatkom.

Katero metodo uporabimo za določanje trenda, je odvisno od proučevane časovne vrste in namena analize. Če potrebujemo trend zaradi proučevanja cikličnih ali sezonskih odklonov, je logično, da predpostavljamo, da je trend črta, ki poteka med realnimi vrednostmi tako, da se izenačujejo periodični in ciklični vplivi. To črto določamo običajno mehanično. Če pa potrebujemo trend za napovedovanje, iščemo

za trend funkcijsko obliko, iz katere moremo z ekstrapolacijo oceniti potek trenda tudi v prihodnost.

21.44 Najvažnejši problem je izbira pravilnega tipa funkcije, ki naj predstavlja trend. Enoličnega pravila za določevanje funkcije trenda ni. Splošne smernice določajo, da mora biti funkcija, ki jo uporabimo za trend, smiselna, ne preveč zamotana.

Po sliki časovne vrste, poznavanju pojava, ki ga vrsta prikazuje in poznavanju lastnosti funkcij, ki pridejo v poštev, izberemo v vsakem določenem primeru, kateri tip funkcije je vsebinsko in tehnično najprimernejši.

Kot trend uporabljamo ali iregularno prilagojene črte ali analitične krivulje. Za opis trenda pridejo najpogosteje v poštev tele funkcije:

1. Premica: $T = a + bx$
2. Parabola druge, tretje ... stopnje: $T = a + bx + cx^2 + dx^3 \ldots$
3. Parabola: $T = a + b\sqrt[n]{x}$
4. Eksponentna funkcija: $T = ab^x$
5. Kvadratična eksponentna funkcija: $T = ab^x \cdot c^{x2}$
6. Modificirana eksponentna funkcija: $T = k + ab^x$ (21.18)
7. Gompertzova krivulja: $T = Ka^{b^x}$
8. Pearl-Reedova logistična krivulja:

$$T = \frac{T_\infty}{1 + ab^x}$$

V teh funkcijah pomeni T trend, x čas, a,b,c,d,k, T_∞ pa so parametri trenda.

21.45 Določitev trenda ima dva dela. Najprej je treba izbrati ustrezen tip funkcije, ko pa tega imamo, je treba za konkretno časovno vrsto določiti parametre trenda. Prvi del je vsebinske, drugi pa tehnične narave.

Da se moremo odločiti v vsakem posebnem primeru za ustrezen tip krivulje, moramo poznati potek posameznih tipov krivulj in njihove lastnosti. Zato je najbolje, da časovno vrsto, za katero določamo trend, narišemo v linijski grafikon z ustreznimi skalami in presodimo, kateri tip krivulje najbolje ustreza stvarni smeri razvoja.

O obliki trenda se moremo poučiti tudi iz vrst diferenc zaporednih členov. Če je trend linearen, so prve diference $DT_k = T_{k+1} - T_k$ stalne količine. Če je trend parabola druge stopnje (tipa 2), so druge diference $D^2T_k = DT_{k+1} - DT_k$, ki jih dobimo iz vrste prvih diferenc konstantne. Enako ugotovimo, da je trend parabola tretje stopnje, če so tretje diference $D^3T_k = D^2T_{k+1} - D^2T_k$, dobljene iz vrste drugih diferenc, konstantne.

Če je trend eksponentna funkcija, so prve diference logaritmov trenda konstante. Enako preizkusimo, ali je trend eksponenten, z verižnimi indeksi ali z vrsto koeficientov dinamike. Če je trend eksponenten, so verižni indeksi ali koeficienti dinamike konstantni. Relativen prirastek je namreč pri eksponentni funkciji konstanten.

Če se pojav razvija v logistični funkciji, tvorijo prve diference simetrično unimodalno krivuljo. Pri Gompartzovi krivulji pa tvorijo prve diference krivuljo, ki je asimetrična v desno. Pri logistični krivulji relativen prirastek linearno pada.

Jasno je, da ta pravila veljajo za trend, ne pa za stvarno časovno vrsto. Ta ne vsebuje samo trend, ampak tudi druge sestavine. Zato so časovne vrste poprečij, ki so izračunana tako, da odstranijo periodične in ciklične vplive in zmanjšajo slučajne variacije, veliko primernejše za gornje preizkuse.

21.46 Za računanje parametrov trenda je pri vseh metodah najprimerneje, da so parametri v linearni zvezi. Za prve tri tipe krivulj to velja, ne velja pa za druge. Vendar moremo marsikatero funkcijo, med njimi npr. eksponentno, prevesti v to obliko, če jo logaritmiramo

$$\log T = \log a + x \log b \qquad (21.19)$$

ker moremo pisati

$$T' = a' + b'x \qquad (21.20)$$

pri tem je $T' = \log T$; $a' = \log a$; $b' = \log b$.

Modificirane eksponentne krivulje ne moremo prevesti v željeno obliko, vendar moremo bolj zamotani funkciji, logistično in Gompertzovo funkcijo, prevesti v modificirano eksponentno funkcijo.

Če vzamemo za logistično funkcijo reciprok, dobimo

$$\frac{1}{T} = \frac{1}{T_{\infty}} + \frac{a}{T_{\infty}} \cdot b^{x} \qquad (21.21)$$

ali dalje:

$$T' = k' + a' b'^{x} \qquad (21.22)$$

Pri tem pomeni:

$$T' = \frac{1}{T}; \quad k' = \frac{1}{T_{\infty}}; \quad a' = \frac{a}{T_{\infty}}, \quad b' = b$$

Če logaritmiramo Gompertzovo funkcijo, pa dobimo:

$$\text{Log } T = \log K + b^{x} \log a \qquad (21.23)$$

ali

$$T' = k' + a' b'^{x} \qquad (21.24)$$

Pri tem pomeni: $T' = \log T$; $k' = \log K$; $a' = \log a$; $b' = b$.

Medtem ko je premica primerna funkcija za smer razvoja za kratka razdobja, eksponentna funkcija pa za srednjeročna, se izkaže, da logistična in Gompertzova krivulja dobro opišeta razvoj za zelo dolga razdobja. Za zelo dolga razdobja opazimo, da se v logistični krivulji spreminja npr. število prebivalstva. Enake zakonitosti opazujemo tudi pri drugih živih bitjih. Ugotovili so, da moremo logistično in Gompertzovo krivuljo s pridom uporabiti tudi za sekularen trend za ekonomske pojave. Logistična ali Gompertzova krivulja, s katero ponazorimo npr. trend proizvodnje, nakazuje

štiri obdobja: obdobje nizke proizvodnje, ki pomeni stopnjo preizkušanja in uvajanja, drugo obdobje, ki pomeni vzpon v družbeno proizvodnjo, tretje obdobje - po prevoju, ko proizvodnja raste, vendar degresivno , in zadnje obdobje stabilizacije pri nasičenosti.

21.47 Splošno velja pravilo, da moremo trend izračunati le za daljša razdobja. V nasprotnem primeru sicer dobimo prilagojeno funkcijo ali krivuljo, vendar pa ta ni trend marveč precej verjetno vsebuje predvsem ostanek cikla, pa tudi drugih vplivov. Tudi pri daljših razdobjih moramo paziti, da je razdobje, za katerega določamo trend, tako, da vključuje cele cikle in periode, ne pa samo dele. Če ta pogoj ni izpolnjen, moreta nezaključeni cikel ali perioda znatno vplivati, da dobljena krivulja ni verna slika osnovne smeri razvoja.

Kljub vsemu pa dostikrat prilagajamo krivulje tudi krajšim časovnim vrstam. Vendar moramo biti v teh primerih oprezni pri tolmačenju take smeri razvoja, saj je kratkoročna in more vsebovati razen trenda tudi ciklična nihanja in druge vplive.

Preskus o značilnosti trenda

21.48 Čeprav predpostavljamo, da imajo vsi ekonomski pojavi v daljšem razdobju dano smer razvoja, je vendar včasih časovna vrsta taka, da iz nje ni možno neposredno razbrati smer razvoja. Včasih je časovno razdobje prekratko, včasih pa so variacije podatkov iz drugih vzrokov prevelike. V takih primerih ni umestno določati trend, čeprav bi računanje, ki bi ga izvedli mehanično, pokazalo neko smer razvoja. Zato s preskusi ugotavljamo značilnost trenda. Metod za določanje značilnosti smeri razvoja je več. Nakazali pa bomo le uporabo Spearmanovega koeficienta korelacije ranga.

Če vzamemo čas v časovni vrsti kot znak x, podatke v časovni vrsti pa kot znak y moremo med x in y izračunati Spearmanov koeficient korelacije ranga. Če je smer razvoja monotono naraščajoča, je Spearmanov koeficient korelacije r_s visok in pozitiven, če pa je smer razvoja monotono padajoča, pa je Spearmanov koeficient absolutno velik, a negativen. S t-preskusom

$$t = \frac{r_S}{\sqrt{1 - r_S^2}} \cdot \sqrt{n - 2} \; ; \quad m = n - 2$$

pri čemer je: n = število členov v časovni vrsti ; r_S = Spearmanov koeficient med časom in podatkom Y.

21.49 Za preskus značilnosti trenda vzemimo volumen (v tisočih tonah) izvoza (Y) v predvojni in povojni Jugoslaviji

Leto x	Y	R_x	R_y	d_R	Leto x	Y	R_x	R_y	d_R
1920	915	1	1	0	1946	486	1	1	0
1921	1584	2	2	0	1947	2491	2	2	0
1922	2214	3	3	0	1948	4097	3	11	+ 8
1923	3026	4	7	+ 3	1949	3757	4	9	+ 5
1924	3916	5	13	+ 8	1950	3082	5	4	- 1
1925	4398	6	15	+ 9	1951	3228	6	5	- 1
1926	4885	7	19	+ 12	1952	3479	7	8	+ 1
1927	4251	8	14	+ 6	1953	2880	8	3	- 5
1928	4527	9	16	+ 7	1954	3443	9	7	- 2
1929	5330	10	20	+ 10	1955	3388	10	6	- 4
1930	4733	11	18	+ 7	1956	4067	11	10	- 1
1931	3323	12	8	- 4	1957	4680	12	12	0
1932	2398	13	5	- 8	1958	5143	13	15	+ 2
1933	2930	14	6	- 8	1959	4766	14	13	- 1
1934	3584	15	11	- 4	1960	5205	15	16	+ 1
1935	3327	16	9	- 7	1961	5061	16	14	- 2
1936	2868	17	4	- 13	1962	5934	17	17	0
1937	4557	18	17	- 1	1963	6225	18	19	+ 1
1938	3702	19	12	- 7	1964	6266	19	20	+ 1
1939	3402	20	10	- 10	1965	6128	20	18	- 2
			$\sum d_R^2$	= 1060	1966	7518	21	21	0
					1967	8316	22	24	+ 2
					1968	7609	23	22	- 1
					1969	8154	24	23	- 1
					1970	8369	25	25	0
								$\sum d_R^2$	= 160

$$r_{S_1} = 1 - \frac{6\sum d_R^2}{N(N^2-1)} = 1 - \frac{6 \cdot 1060}{20(20^2-1)} = 0,2030$$

$$r_{S_2} = 1 - \frac{6 \cdot 160}{25(25^2-1)} = 0,9385$$

$$t_1 = \frac{0,2030.\sqrt{20-2}}{\sqrt{1-0,2030^2}} = 0,880; \quad t_2 = \frac{0,9385 .\sqrt{25-2}}{\sqrt{1-0,9385^2}} = 13,036$$

Ker je $t_{0,10}(m=18) = 1,734$ in $t_{0,001}(m=23) = 3,767$, sklepamo, da je bil trend izvoza v Jugoslaviji neznačilen, medtem ko je v SFRJ v razdobju 1946-1970 trend visoko značilen. Seveda pa rezultat o značilnosti trenda ne pove nič o obliki trenda.

21.50 Metode za določanje trenda. Omenili smo že, da imamo več metod za določanje trenda. Od njih bomo obravnavali tele:

a) prostoročno metodo,

b) metodo drsečih sredin,

c) metodo sredin med najvišjimi in najnižjimi točkami.

Te metode posredujejo trend v mehanični ali grafični obliki. Od metod, ki dajejo trend analitično v obliki krivulj, pa bomo obravnavali:

a) metodo izbranih točk,

b) metodo delnih sredin,

c) metodo delnih vsot,

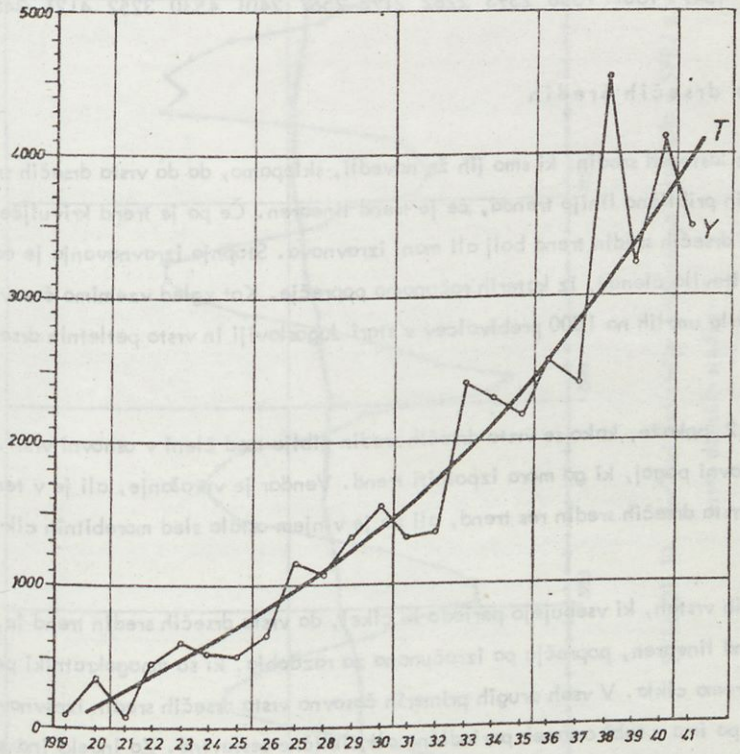
d) metodo najmanjših kvadratov.

Prostoročna metoda

21.51 Običajno si moremo v grafikonu že iz slike časovne vrste zamisliti potek trenda. Trend v vsakem primeru poteka med realnimi vrednostmi tako, da se osnovna časovna vrsta odklanja od trenda navzgor in navzdol. Zato moremo trend

včrtati v grafikonu prostoročno. Ta metoda je subjektivna in zato nima posebne analitične vrednosti. Vendar je njena prednost v enostavnosti in hitrosti. Zato jo uporabljamo predvsem za približek in za osnovo pri izbiri analitične oblike trenda.

21.52 V sliki 21.10 je narisana časovna vrsta proizvodnje cinkovega koncentrata na območju Slovenije v razdobju med obema vojnama od 1919-1941, v njej pa prostoročno včrtan trend.



Slika 21.10. Trend za proizvodnjo cinkovega koncentrata v Sloveniji med obema vojnama, določen prostoročno

Tabela 21.14. Proizvodnja cinkovega koncentrata v Sloveniji med obema vojnama v tonah (Vir: SB SRS 1952, št. 8)

Leto:	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
Proizvodnja	90	350	75	450	600	520	510	639	1143	1076	1313

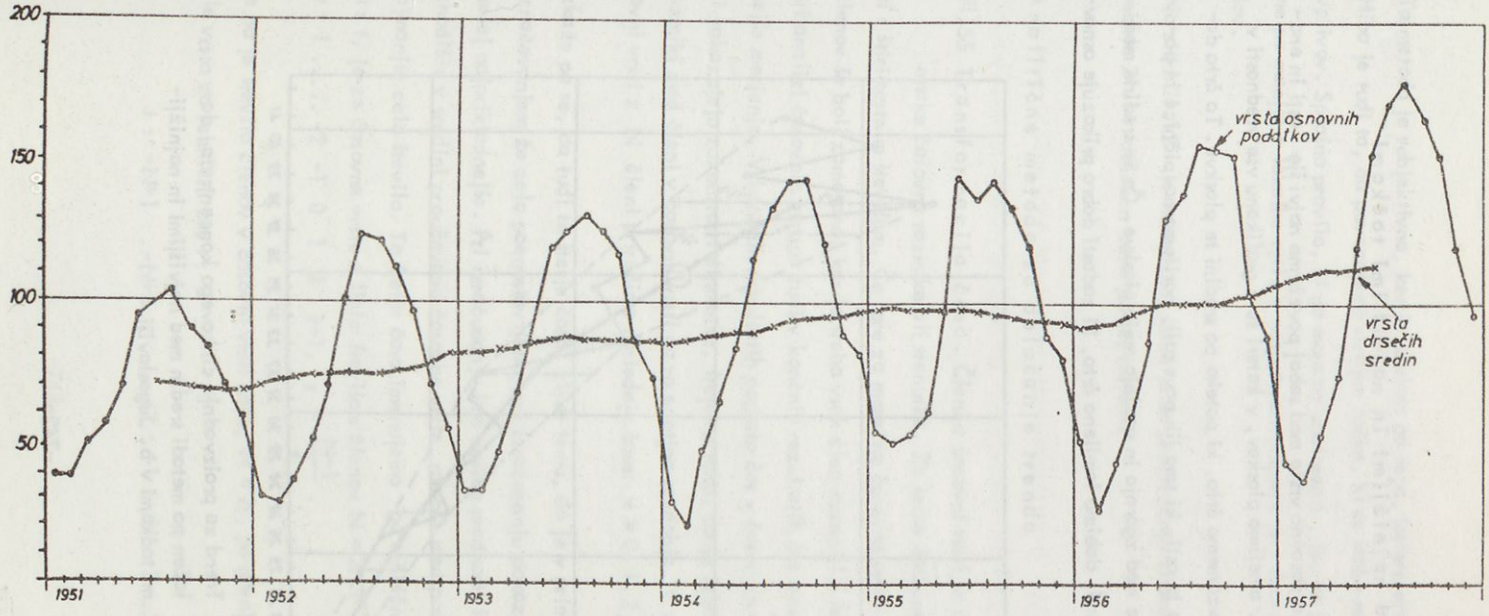
Leto:	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
Proizvodnja	1547	1301	1358	2393	2282	2173	2567	2401	4530	3252	4121	3483

Metoda drsečih sredin

21.53 Iz lastnosti sredin, ki smo jih že navedli, sklepamo, da da vrsta drsečih sredin približno linijo trenda, če je trend linearen. Če pa je trend krivuljčen, se v vrsti drsečih sredin trend bolj ali manj izravnava. Stopnja izravnavanja je odvisna od števila členov, iz katerih računamo poprečje. Kot zgled vzemimo časovno vrsto števila umrlih na 1000 prebivalcev v stari Jugoslaviji in vrsto petletnih drsečih sredin.

Slika 21.2 pokaže, kako se vrsta drsečih sredin giblje med členi v osnovni vrsti. To pa je osnovni pogoj, ki ga mora izpolniti trend. Vendar je vprašanje, ali je v tem primeru vrsta drsečih sredin res trend, ali pa je v njem ostala sled morebitnih ciklov.

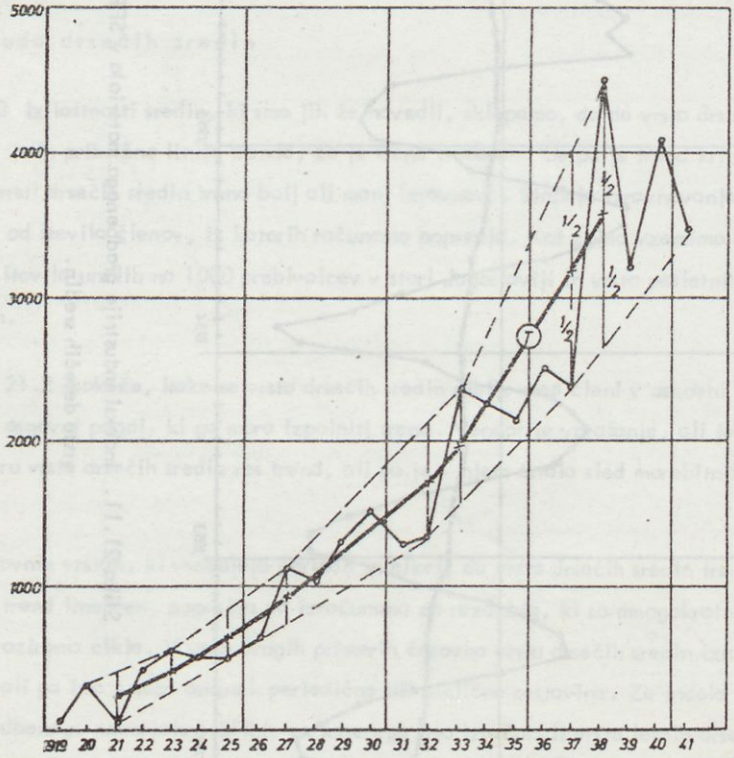
V časovnih vrstah, ki vsebujejo periodo in cikel, da vrsta drsečih sredin trend le, če je trend linearen, poprečja pa izračunana za razdobja, ki so mnogokratniki periode oziroma cikla. V vseh drugih primerih časovna vrsta drsečih sredin izravnava trend ali pa ima v sebi ostanek periodične ali ciklične sestavine. Za indeks industrije gradbenega materiala v SFRJ, za katerega smo izračunali vrsto letnih drsečih sredin, na sliki 21.11 jasno vidimo, da je iz osnovne vrste odstranjen sezonski vpliv in slučajne variacije, ker so poprečja letna.



Slika 21.11. Indeksi industrije gradbenega materiala v SFRJ z vrisano časovno vrsto desečih sredin.

Metoda sredin med najnižjimi in najvišjimi točkami

21.54 Če v grafikonu za časovno vrsto med seboj povežemo najvišje osti in enako najnižje osti, omejimo ploskev, v kateri so v grafikonu vse vrednosti v časovni vrsti. Za trend vzamemo črto, ki poteka po sredini te ploskve. To črto dobimo enostavno tako, da iz osti, ki smo jih povezali, narišemo navpičnice in poiščemo sredine navpičnic med zgornjo in spodnjo mejo ploskve. Če te sredine med seboj zvežemo z daljicami, dobimo lomljeno črto, ki zadosti dobro prikazuje osnovno smer razvoja.



Slika 21.12. Trend za proizvodnjo cinkovega koncentrata, določen po metodi sredin med najvišjimi in najnižjimi točkami v b. Jugoslaviji.

Ta metoda je subjektivna, ker je odvisna od tega, katere osti med seboj povežemo. Hiba je tudi ta, da povezujemo skrajne točke, ki so lahko rezultat iregularnih vplivov. Splošno pravilo, ki ga moramo upoštevati , je, da povezujemo res skrajne ne pa manjše - lokalne osti. V nasprotnem primeru je v črti trenda še ostanek ciklov.

Na sliki 21.12 je nakazano, kako je za proizvodnjo cinkovega koncentrata iz tabele 21.14 določen trend po tej metodi.

Analitične metode za določanje trenda

21.55 Transformacija časa. Členi v časovni vrsti se nanašajo na zaporedne enake časovne razmike ali trenutke. Za letne časovne vrste so členi označeni s štirimestnim številom. Če gre za mesečno časovno vrsto pa je označevanje členov še bolj zamotano, ker je treba vsak člen označiti z letom in mesecem. Zato pri analizi časovnih vrst pa tudi v končnih rezultatih čas transformiramo na enostavnejše merjenje. V končnih rezultatih pogosto čas v časovnih vrstah ali funkcijah, ki nakazujejo zakonitosti dinamike, štejemo tako, da je izhodišče v prvem členu, razmiki med členi v časovni vrsti pa so enotina razdobja. V praktičnih primerih časovni vrsti z N členi priredimo naslednje čase: $x = 0, 1, 2, 3, \ldots, (N-1)$.

Izkaže pa se, da tudi to štetje časa, kljub temu, da je v primerjavi s koledarskim označevanjem že zelo poenostavljeno, pri izračunanju pokazovalcev dinamike ni vselej najprikladnejše. Pri izračunanju ima veliko prednosti štetje, za katerega je izhodišče v sredini proučevane časovne vrste, razmik med posameznimi členi pa najmanjše celo število. To štetje časa, imenujemo ga tehnični čas, zaznamujemo pa s t, je za časovne vrste z lihim številom členov $N = 2r+1$ $t: -r = \frac{N-1}{2}$, $-r+1 \ldots\ldots -2\ -1\ 0\ 1\ 2\ \ r-1,\ r = \frac{N-1}{2}$.

Če pa je število členov v časovni vrsti sodo, $N = 2r$, pa gornjim pogojem ustreza tale vrsta za tehnični čas t:

$t: -N+1\quad -N+3\quad \ldots\ldots -3\quad -1\quad 1\quad 3\quad \ldots\ N-3\quad N-1$

Med tehničnim časom t in časom z izhodiščem v prvem členu x, je glede na to, ali je število členov N liho ali sodo tale zveza:

$$N = 2r+1 \; ; \quad t = x - \frac{N-1}{2}$$

$$N = 2r \; ; \quad t = 2x - (N-1) \tag{21.25}$$

Metoda izbranih točk

21.56 Metoda izbranih točk je manj subjektivna kot prostoročna grafična metoda.

Po tej metodi iz grafikona najprej proučimo časovno vrsto in ugotovimo tip funkcije, ki ustreza kot trend. Glede na število parametrov v funkciji trenda določimo v grafikonu osnovne časovne vrste toliko točk, za katere sodimo, da leže na trendu, kolikor parametrov ima funkcija trenda. Te točke so, jasno, določene subjektivno. Včasih si pomagamo tako, da najprej narišemo prostoročno trend, na tej krivulji trenda pa izberemo ustrezno število točk na trendu. Točke pa ne smejo biti izbrane na prekratkem razmiku, ker se sicer subjektivnim vplivom pridruži še slučajnostni. Če ima funkcija dva parametra, izberemo torej eno točko bolj na začetku, drugo pa pri koncu časovne vrste. Če ima funkcija tri parametre, izberemo eno točko v sredini časovne vrste, dve pa pri robnih členih, najbolje enako odmaknjenih od sredine, itd.

Metodo izbranih točk pogosto uporabljamo takrat, kadar je funkcija trenda taka, da parametri niso v linearni zvezi in je uporaba drugih metod otežkočena, če že ne onemogočena.

21.57 Za zgled vzemimo izvoz Anglije v razdobju 1820–1930 v milj. Lstg po stalnih cenah.

Tabela 21.15. Izvoz iz Anglije v razdobju 1820–1930 v milj. Lstg

Leto L	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930
Izvoz Y	27	37	58	99	176	236	320	407	416	573	459	461
DY		+10	+21	+41	+77	+60	+84	+87	+ 9	+157	-114	+ 2
t	-11	-9	-7	-5	-3	-1	+1	+ 3	+ 5	+ 7	+9	+11

Grafična izravnava prvih diferenc podatkov odkrije, da sestavljajo prve diference precej simetrično unimodalno krivuljo. Iz tega sklepamo, da je primerno, da uporabimo kot funkcijo trenda logistično funkcijo. (Glej sliko 21.14).

Ker je število členov v proučevani časovni vrsti sodo, izberemo tehnični čas po obrazcu

$$t = \frac{L - 1875}{5} \tag{21.26}$$

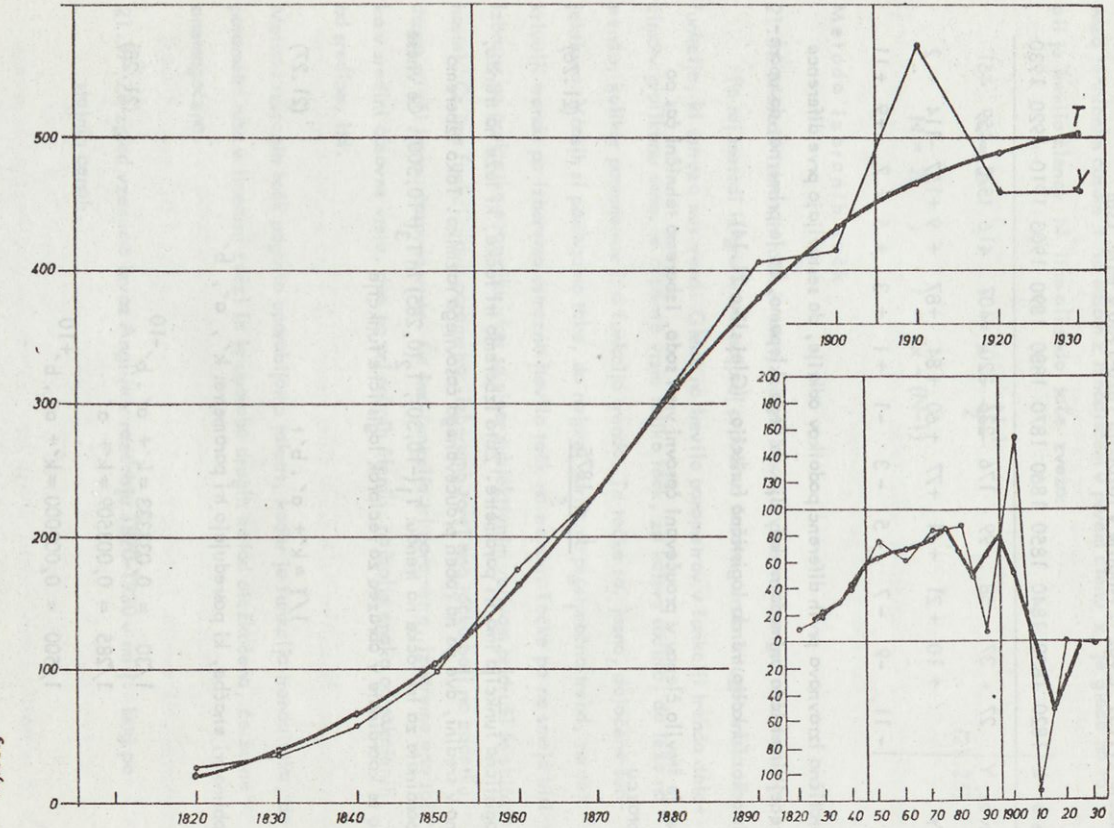
Logistična funkcija ima tri parametre. Zato izberemo tri točke, ki leže na trendu, eno v sredini, dve pa na robeh proučevanega časovnega razmika. Tako izberemo koordinate za tri točke na trendu: $T_1(-10,30)$, $T_2(0,285)$ in $T_3(+10,500)$. Če vnesemo te koordinate v obrazec za reciprok logistične funkcije

$$1/T = k' + a' \cdot b'^{t} \tag{21.27}$$

dobimo tri enačbe, ki posredujejo tri parametre: k', a', b'

$$\begin{aligned} 1/30 &= 0,033333 = k' + a' \cdot b'^{-10} \\ 1/285 &= 0,003509 = k' + a' \\ 1/500 &= 0,002000 = k' + a' \cdot b'^{+10} \end{aligned} \tag{21.28}$$

Glede na to, da smo točke izbrali simetrično na sredino vrste, je sistem enačb dokaj enostaven. Če odštejemo drugo enačbo od prve, tretjo pa od druge, izločimo parame-



Slika 21.13. Logistični trend za skupno vrednost izvoza Anglije v razdobju 1820–1930

Slika 21.14. Preskus za logistični trend

ter k'. Iz dobljenega sistema dveh enačb z dvema parametroma:

$$0,029824 = a' \cdot b'^{-10} \cdot (1-b'^{+10})$$

$$0,001509 = a' \cdot (1-b'^{+10})$$

izločimo parameter a', če delimo drugo enačbo s prvo enačbo.

$$b'^{10} = \frac{0,001509}{0,029824} = 0,050597$$

Z logaritmiranjem dobimo, da je $b' = 0,7420$.
Vrednost b'^{10} vstavimo v zadnjo enačbo v gornjem sistemu dveh enačb. Tako dobimo, da je

$$0,001509 = a' \cdot (1 - 0,050597)$$

oziroma

$$a' = \frac{0,001509}{1 - 0,050597} = 0,001589$$

Iz druge enačbe v sistemu enačb 21.28 sledi, da je

$$0,003509 = k' + 0,001589$$

oziroma $k' = 0,001920$.
Funkcija recipročnih vrednosti za trend je torej:

$$T' = k' + a'b'^{t} = 0,001920 + 0,001589 \cdot 0,742^{t}.$$

Če upoštevamo, da je po obrazcih $T_{\infty} = 1/k' = 1/0,001920 = 520,8$
$a = a'/k' = 0,001589/0,001920 = 0,827$ in $b' = b' = 0,742$, dobimo, da je funkcija logističnega trenda za izvoz Anglije

$$T = \frac{T_{\infty}}{1 + a \cdot b^{t}} = \frac{520,8}{1 + 0,827 \cdot 0,742^{t}} \qquad (21.29)$$

Pri tem moramo opozoriti, da je med časom t in stvarnim časom dana zveza z obrazcem 21.26.

Računanje trenda za leta, ki so dana v tabeli 21.15, je nakazano v tabeli 21.16, slika osnovne časovne vrste in ustreznega trenda pa v sliki 21.13.

Če transformiramo tehnični čas t na letni x z izhodiščem v letu 1900, dobimo zvezo

$$t = \frac{x + 25}{5}$$

Če vnesemo to izraz v obrazec 21.29, dobimo

$$T = \frac{520,8}{1 + 0,827 \cdot 0,742^{\frac{x+25}{5}}} = \frac{520,8}{1 + 0,827 \cdot 0,742^{5} \cdot (0,742^{1/5})^{x}} =$$

$$= \frac{520,8}{1 + 0,18607 \cdot 0,94207^{x}} \qquad (21.30)$$

Tabela 21.16. Izračunanje logističnega trenda za izvoz iz Anglije od leta 1820 do 1930

Leto	Y	t	$a' \cdot b'^{t}$	+	k'	=	T'	$T = 1/T'$
1820	27	- 11	0,042332	+	0,001920	=	0,044252	23
1830	37	- 9	23306	+	1920	=	25226	40
1840	58	- 7	12832	+	1920	=	14752	68
1850	99	- 5	7647	+	1920	=	9567	105
1860	176	- 3	4210	+	1920	=	6130	163
1870	236	- 1	2318	+	1920	=	4238	236
1880	320	+ 1	1276	+	1920	=	3196	313
1890	407	+ 3	703	+	1920	=	2623	381
1900	416	+ 5	387	+	1920	=	2307	434
1910	573	+ 7	213	+	1920	=	2133	469
1920	459	+ 9	117	+	1920	=	2037	491
1930	461	+ 11	65	+	1920	=	1985	504

Metoda delnih sredin

21.58 Metodo izbranih točk moremo nadomestiti z objektivnejšo metodo, metodo delnih sredin. Subjektivno izbrane vrednosti nadomesti objektivno določena

srednja vrednost.

Pri metodi delnih sredin izkoristimo predpostavko, da se v poprečju uničijo slučajnostni vplivi ter periodični in ciklični vplivi, če je razdobje, za katerega računamo sredine, enako dolžini ali mnogokratniku periode ali cikla. Dalje vemo, da poprečje iz trenda leži na trendu, če je v razdobju, za katerega računamo sredino, trend linearen. Vse te lastnosti dajo dosti omejitev za uporabnost metode delnih sredin. Če upoštevamo te omejitve, moremo po metodi delnih sredin neoporečno določati linearen trend. V tem so vključeni tudi primeri, da je trend funkcija, ki jo moremo s transformacijo prevesti v linearno funkcijo. Taka funkcija je npr. eksponentna funkcija $T = a.b^x$, ki jo z logaritmiranjem prevedemo v $\log T = \log a + x \log b$. Enako hiperbolo tipa $T = \frac{a}{x-b}$ s transformacijo $1/T = \frac{x-b}{a} = a' + b'x$ prevedemo v linearno funkcijo, itd. Za eksponentno funkcijo velja iz gornjega, da na trendu leže geometrijske sredine, za hiperbole gornjega tipa pa harmonične sredine iz osnovnih podatkov. Postopek računanja funkcije trenda po metodi delnih sredin je torej v mnogočem podoben metodi izbranih točk. Za vrste, za katere je trend krivuljčen, vzamemo približek, da je kljub temu, če je v celoti trend krivuljčen, na krajših odsekih trend linearen. Za take primere računamo sredine iz čim krajših razdobij. To pa je v nasprotju z drugo omejitvijo, po kateri mora biti zaradi slučajnostnih, cikličnih in periodičnih vplivov razmik, iz katerega računamo sredine, čim daljši oziroma mora imeti dano dolžino.

21.59 Za zgled izračunajmo po metodi delnih sredin trend oziroma smer razvoja za naravni prirastek prebivalstva za SFRJ v razdobju 1955-1964. Analiza grafikona nakaže, da je trend linearen. Zato potrebujemo dva pogoja, ker ima linearen trend dva parametra.

Tabela 21.17. Izračunanje trenda za naravni prirastek prebivalstva za SFRJ v razdobju 1955-1964 po metodi delnih sredin

Leto	t	Y	
1955	-9	15,5	
1956	-7	14,8	
1957	-5	13,2	$\bar{Y}_{-5} = 71,7 : 5 = 14,34$
1958	-3	14,7	
1959	-1	13,5	
1960	+1	13,6	
1961	+3	13,7	
1962	+5	12,0	$\bar{Y}_{+5} = 63,2 : 5 = 12,64$
1963	+7	12,5	
1964	+9	11,4	

Po gornjih predpostavkah ležita točki T(-5, 14,34) in T(+5, 12,64) na linearnem trendu $T = a+bt$. Če te vrednosti vnesemo v enačbo trenda, dobimo dve enačbi:

$$14,34 = a - 5b$$
$$12,64 = a + 5b$$

Zaradi prednosti tehničnega časa t ta sistem enačb rešimo enostavno. S seštevanjem in odštevanjem dobimo vrednosti parametrov a in b.

$$26,98 = 2a \; ; \; a = 13,49$$
$$-1,70 = 10b \; ; \; b = -0,170$$
$$T = 13,49 - 0,170\,t$$

Če t transformiramo v x z izhodiščem x = 0 v prvem členu, $t = 2x - 9$, dobimo:

$$T = 13,49 - 0,170\,(2x - 9) = 15,02 - 0,340\,x$$

Poprečna letna sprememba naravnega prirastka v razdobju 1955-1964 je b = -0,340.

Metoda delnih vsot

21.60 Izravnavanju trenda, do katerega pride pri delnih sredinah, se izognemo z metodo delnih vsot.

Enako kakor pri metodi delnih sredin tudi pri metodi delnih vsot časovno vrsto razdelimo v toliko - najprimerneje po obsegu enakih delov, kolikor ima funkcija trenda parametrov.

Če predpostavimo, da je proučevani pojav odvisen le od trenda in slučajnostnih vplivov, moremo vzeti, da se slučajnostni vplivi v vsoti v razdobjih, iz katerih računamo vsote, uničijo. Tako dobimo iz osnovnega modela

$$Y = T + S \tag{21.31}$$

toliko enačb tipa

$$\sum Y = \sum T \tag{21.32}$$

kolikor parametrov ima funkcija.

V primeru, če so v časovni vrsti še ciklični ali periodični vplivi, dobimo po znanih pravilih za ta primer enačbe tipa 21.32 le, če so razdobja, za katera računamo vsote mnogokratniki iz dolžine periode oziroma ciklov, če predpostavimo, da so sestavine vezane aditivno.

Metoda delnih vsot je kot objektivna metoda za določanje trenda posebno primerna takrat, če je trend krivuljčen, zveza med parametri v funkciji pa ni aditivna oziroma se niti s transformacijo ne da privesti v tako obliko. Od funkcij, ki pridejo v praksi najpogosteje v poštev, so take funkcije: modificirana eksponentna funkcija, Gompertzova funkcija in logistična funkcija.

Pri metodi delnih vsot predpostavljamo, da so sestavine med seboj v aditivni zvezi. V primerih, v katerih kaže osnovna časovna vrsta ali analiza pojava drugače, osnovno časovno vrsto s transformacijo prevedemo v tako obliko, da to velja (npr. če je med sestavinami faktorska zveza z logaritmiranjem).

21.61 Kot zgled za uporabo metode delnih vsot za določanje trenda vzemimo razvoj števila potniških avtomobilov v SFRJ v razdobju 1946-1960. Vsebinska analiza kakor tudi podatki kažejo razvoj, ki je tipičen za različne ekonomske pojave. Ob osvoboditvi smo razpolagali z določenim številom osebnih avtomobilov. K temu osnovnemu številu se je od leta do leta vključevalo število novih avtomobilov in sicer v geometrijskem zaporedju. Tej predpostavki ustreza kot funkcija trenda modificirana eksponentna funkcija

$$T = k + a.b^x \qquad (21.33)$$

oziroma model časovne vrste

$$Y = T + S = k + a \cdot b^x + S \qquad (21.34)$$

Tabela 21.18. Število potniških avtomobilov v SFRJ v razdobju 1946-1960 (v tisoč avtomobilov)

Leto	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956
Y	6,2	6,6	7,2	7,3	6,5	6,9	8,5	10,2	11,3	12,6	14,7

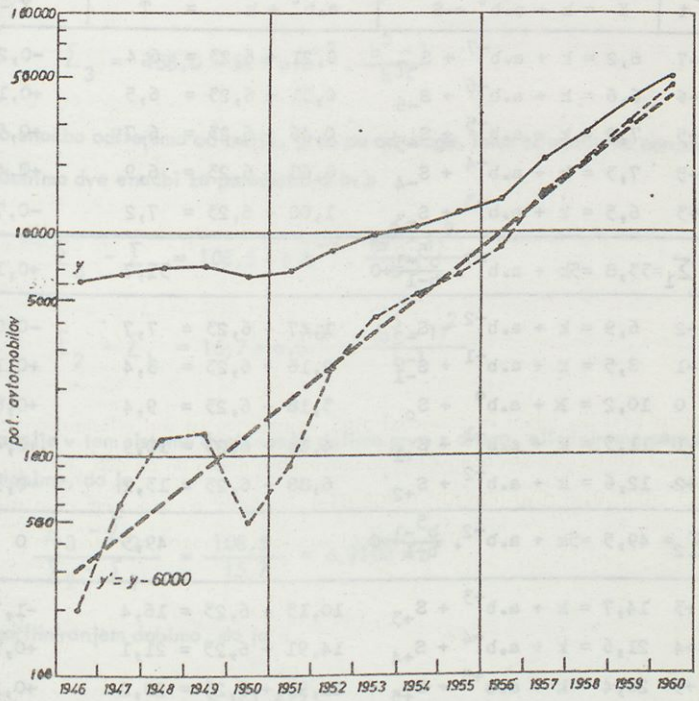
Leto	1957	1958	1959	1960
Y	21,6	28,4	39,0	54,3

Iz osnovne časovne vrste je razvidno, da je začetno število osebnih avtomobilov okrog 6000. Najprej napravimo grafični preizkus domneve, da je trend modificirana eksponentna funkcija. V ta namen smo v sliki 21.15 v logaritemskem grafikonu narisali osnovno časovno vrsto Y in časovno vrsto, od katere smo odšteli šest tisoč $Y' = Y - 6,0$.

Slika potrdi domnevo, ker je smer razvoja za časovno vrsto $Y' = Y - 6,0$ na logaritemskem grafikonu linearna, medtem ko je linija za osnovno časovno vrsto krivuljčna.

Na osnovi te analize prilagodimo osnovni časovni vrsti modificirano eksponentno funkcijo.

V tabeli 21.19 je sistematično prikazana sestava posameznih sestavin za posamezna leta in delne vsote. Ker celotno razdobje, ki ga proučujemo, obsega 15 let, je razdeljeno v tri delna razdobja po pet let. V drugem stolpcu je letom prirejen tehnični čas. V tabeli je nakazano, kako se posamezni elementi trenda seštevajo.



Slika 21.15 Grafični preizkus o funkciji trenda za število osebnih avtomobilov v SFRJ

Parameter k seštejemo enostavno v 5k. Člen $a.b^x$ v vsakem delnem razdobju tvori geometrijsko zaporedje in ga seštejemo po pravilih za vsoto členov iz geometrijskega zaporedja v izraze, ki so naznačeni pri vsaki delni vsoti. Za slučajnostne vplive S predpostavljamo, da se v vsoti v petletnem razmiku uničijo.

Tabela 21.19. Izračunanje trenda za razvoj števila potniških avtomobilov v SFRJ v razdobju 1946-1960 po metodi delnih vsot (v tisoč avtomobilov)

Leto	t	Izračunanje parametrov: Y = k + a.b^t + S	Izračunanje trenda: a.b^t + k = T	Y - T
1946	-7	6,2 = k + a.b^-7 + S_-7	0,21 + 6,23 = 6,4	-0,2
1947	-6	6,6 = k + a.b^-6 + S_-6	0,31 + 6,23 = 6,5	+0,1
1948	-5	7,2 = k + a.b^-5 + S_-5	0,46 + 6,23 = 6,7	+0,5
1949	-5	7,3 = k + a.b^-4 + S_-4	0,68 + 6,23 = 6,9	+0,4
1950	-3	6,5 = k + a.b^-3 + S_-3	1,00 + 6,23 = 7,2	-0,7
Vsota		$\sum_1$=33,8 =5k + a.b^-7. $\frac{b^5-1}{b-1}$+0	33,7	+0,1
1951	-2	6,9 = k + a.b^-2 + S_-2	1,47 + 6,23 = 7,7	-0,8
1952	-1	3,5 = k + a.b^-1 + S_-1	2,16 + 6,23 = 8,4	+0,1
1953	0	10,2 = k + a.b^0 + S_0	3,18 + 6,23 = 9,4	+0,8
1954	+1	11,3 = k + a.b^+1 + S_+1	4,68 + 6,23 = 10,9	+0,4
1955	+2	12,6 = k + a.b^+2 + S_+2	6,88 + 6,23 = 13,1	-0,5
Vsota		$\sum_2$= 49,5 =5k + a.b^-2. $\frac{b^5-1}{b-1}$+0	49,5	0
1956	+3	14,7 = k + a.b^+3 + S_+3	10,13 + 6,23 = 16,4	-1,7
1957	+4	21,6 = k + a.b^+4 + S_+4	14,91 + 6,23 = 21,1	+0,5
1958	+5	28,4 = k + a.b^+5 + S_+5	21,95 + 6,23 = 28,2	+0,2
1959	+6	39,0 = k + a.b^+6 + S_+6	32,31 + 6,23 = 38,5	+0,5
1960	+7	54,3 = k + a.b^+7 + S_+7	47,56 + 6,23 = 53,8	+0,5
Vsota		$\sum_3$=158,0 =5k + a.b^+3. $\frac{b^5-1}{b-1}$+0	158,0	0

Tako dobimo v prvem delu tabele tri enačbe za parametre k, a in b

$$\Sigma_1 = 33,8 = 5k + a.b^{-7} \cdot \frac{b^5 - 1}{b-1}$$

$$\Sigma_2 = 49,5 = 5k + a.b^{-2} \cdot \frac{b^5 - 1}{b-1}$$

$$\Sigma_3 = 158,0 = 5k + a.b^{+3} \cdot \frac{b^5 - 1}{b-1}$$

Drugo enačbo odštejemo od tretje, prvo pa od druge. Tako eliminiramo parameter k in dobimo dve enačbi za parametra a in b

$$\Sigma_3 - \Sigma_2 = 108,5 = a.b^{-2} \cdot \frac{(b^5 - 1)^2}{b-1}$$

$$\Sigma_2 - \Sigma_1 = 15,7 = a.b^{-7} \cdot \frac{(b^5 - 1)^2}{b-1}$$

Če nadalje v tem sistemu dveh enačb delimo prvo z drugo, eliminiramo parameter a in dobimo, da je

$$\frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1} = \frac{108,5}{15,7} = 6,9108 = b^5$$

Z logaritmiranjem dobimo, da je

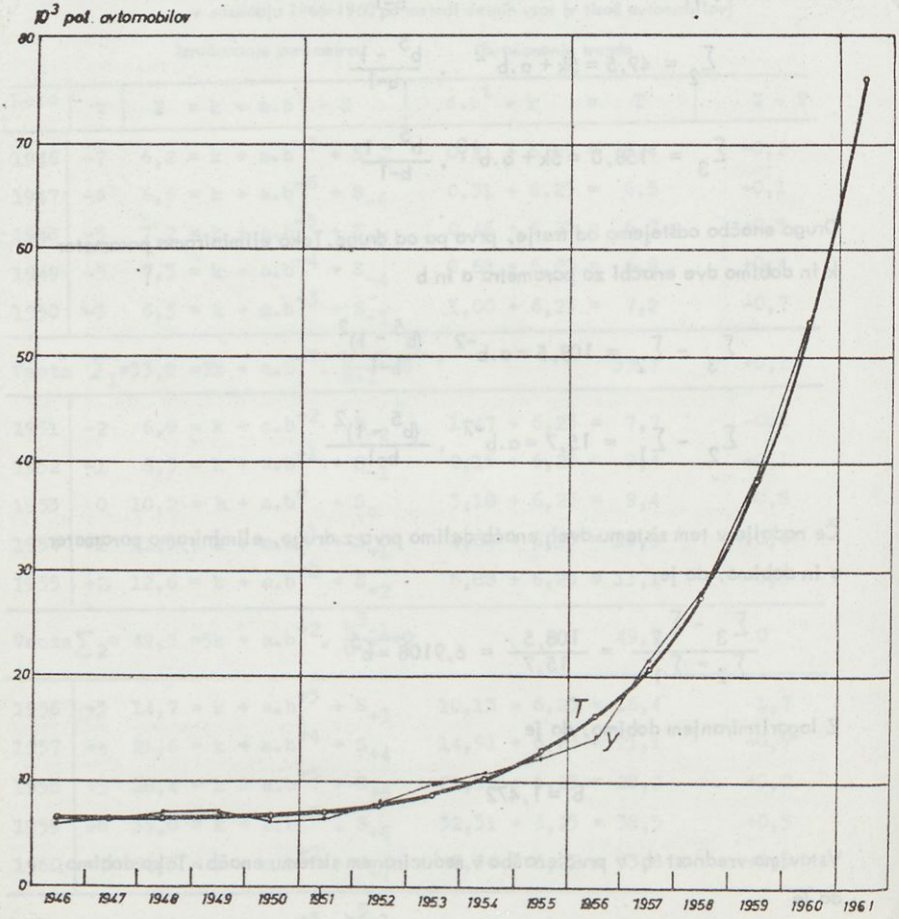
$$b = 1,472$$

Vstavimo vrednost b v prvo enačbo v reduciranem sistemu enačb. Tako dobimo, da je

$$a = (\Sigma_3 - \Sigma_2) \cdot b^2 \cdot \frac{b - 1}{(b^5 - 1)^2} = 108,5 \cdot 1,472^2 \cdot \frac{1,472-1}{(6,9108-1)^2} = 3,176.$$

Parameter k dobimo, če vnesemo vrednosti za parameter a in b v drugo enačbo osnovnega sistema treh enačb

$$k = \frac{1}{5}\left(\Sigma_2 - \frac{\Sigma_3 - \Sigma_2}{b^5 - 1}\right) = \frac{1}{5}\left(49{,}5 - \frac{108{,}5}{5{,}9108}\right) = 6{,}23.$$



Slika 21.16. Število potniških avtomobilov v SFRJ s prilagojeno modificirano eksponentno funkcijo kot trendom

Končno dobimo, da je trend za število potniških avtomobilov v SFRJ dan z enačbo

$$T = 6\,230 + 3\,176 \cdot 1{,}472^{t}$$

$$T = 6\,230 + 217{,}6 \cdot 1{,}472^{x}$$

Parameter b = 1,472 pomeni, da se je število osebnih avtomobilov v SFRJ, če ne upoštevamo začetno število avtomobilov, v petnajstletnem razdobju dvigalo s poprečnim koeficientom dinamike b = 1,472 ali s poprečno stopnjo rasti 47,2%.

V drugem delu tabele je nakazano izračunanje trenda za posamezna leta. Kot preizkus pravilnosti računa se morajo vsote osnovnih vrednosti in vsote trenda za posamezne dele ujemati. V našem zgledu se kontrolne vsote ujemajo razen v prvem delu, kjer razlika +0,1 izvira iz zaokroževanja.

Slika 21.16 in zadnji stolpec v tabeli 21.19, ki prikazuje razlike med osnovno časovno vrsto in trendom, potrjujeta, da je izbira modificirane eksponentne funkcije v tem primeru upravičena.

Da je upravičena predpostavka o linearni zvezi med trendom in slučajnimi vplivi, potrdi dejstvo, da so slučajnostni vplivi, ki so izraženi z odkloni, med stvarnimi vrednostmi in trendom (Y-T) v absolutni vsoti v vseh odsekih približno enaki, kljub temu, da se nivo znatno spreminja.

Če predpostavimo, da se koeficient dinamike povečanja števila potniških avtomobilov v letu 1961 in 1962 ne spremeni, moremo funkcijo trenda ekstrapolirati in dobimo vrednost trenda in s tem približno število potniških avtomobilov v SFRJ v letu 1961:

$$T_{61} = T_{+8} = 6\,230 + 3176 \cdot 1{,}472^{+8} = 76200$$

Pri tej predpostavki bi bila na račun razvoja avtomobilizma vrednost trenda v letu 1961-76200 potniških avtomobilov, v letu 1962 pa 109300. Vendar je treba biti pri ekstrapolaciji trenda previden. Ekstrapolacija za daljše časovno razdobje more

biti zelo problematična, ker se more zakonitost razvoja spremeniti, npr. z novim režimom pri uvozu potniških avtomobilov (dejansko število potniških avtomobilov v 1961 je bilo 75 600, v letu 1962 pa 97 900).

21.62 Postopek za računanje modificirane eksponentne funkcije trenda $T = k + ab^x$ po metodi delnih vsot posplošimo na poljubno število členov v časovni vrsti! Če je skupno število členov v časovni vrsti $N = 3r$, pri tem pa je $r = 2i + 1$ liho število, ima tehnični čas t za posamezne člene v časovni vrsti vrednosti ... -3, -2-1 0+1+2+3... Parametre k, a in b računamo v tem primeru po obrazcih

$$b = \sqrt[r]{\frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1}} \quad ;$$

$$a = (\Sigma_3 - \Sigma_2)\, b^i \frac{b - 1}{(b^r - 1)^2} \qquad (21.35)$$

$$k = \frac{1}{r}\left(\Sigma_2 - \frac{\Sigma_3 - \Sigma_2}{b^r - 1}\right)$$

Če je število členov v osnovni časovni vrsti $N = 3r$, pri čemer je $r = 2i$ sodo število, ima tehnični čas t za posamezne člene v časovni vrsti vrednosti ... -3, -1, +1, +3 ...

Parametre trenda k, a in b za modificirano eksponentno funkcijo izračunamo v tem primeru po obrazcih

$$b = \sqrt[2r]{\frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1}}$$

$$a = (\Sigma_3 - \Sigma_2)\, b^{r-1} \frac{b^2 - 1}{(b^{2r} - 1)^2} \qquad (21.36)$$

$$k = \frac{1}{r}\left(\Sigma_2 - \frac{\Sigma_3 - \Sigma_2}{b^{2r} - 1}\right)$$

Pri tem poprečen koeficient dinamike reduciranih vrednosti (Y - k) ni b, ampak b^2.

Metoda najmanjših kvadratov

21.63 Zelo pogosto uporabljamo za določanje parametrov trenda metodo najmanjših kvadratov. Čeprav je teoretično uporaba metode najmanjših kvadratov za ocenjevanje parametrov za prilagojeno funkcijo pogojena s slučajnostjo odklonov, jo kljub temu uporabljamo za prilagajanje časovnim vrstam, čeprav ta pogoj ni vedno izpolnjen. Pogosto odkloni od osnovne smeri razvoja niso le rezultat slučajnostnih vplivov, ampak so pogojeni s periodičnimi ali cikličnimi vplivi.

Splošna tehnika za uporabo metode najmanjših kvadratov pri določanju trenda je enaka kot pri določanju regresijskih krivulj, le da y' zamenjamo s T. Zato je ne bomo ponavljali, ker je že obdelana v odstavku 18.38.

Metodo najmanjših kvadratov uporabljamo brez računskih težav za vse primere, v katerih je funkcija trenda dana z zvezo

$$g(T) = \sum_{k=o}^{r} a_k f_k(x) \qquad (21.37)$$

kar omejuje tipe funkcije na to, da so parametri med seboj v aditivni zvezi. Ta pogoj je le tehnične narave, da moremo parametre določiti po enotnem načinu prek sistema normalnih enačb

$$\sum_{i=1}^{N} g(Y_i) f_h(x_i) = \sum_{k=o}^{r} a_k \sum_{i=1}^{N} f_k(x_1) f_h(x_i) \quad ; \; h = 0,1 \ldots r \qquad (21.38)$$

Prav posebno pa so kot funkcija trenda prikladne cele racionalne funkcije

$$g(T) = \sum_{k=o}^{r} a_k x^k \qquad (21.39)$$

ki jih moremo pisati v tehničnem času t

$$g(T) = \sum_{k=0}^{r} b_k t^k \qquad (21.40)$$

Zaradi pomena t je $\sum t^{k+h} = 0$, če je k+h liho število. Tako se sistem normalnih enačb znatno poenostavi.

21.64 Linearen trend. Ta prednost se pokaže že za linearen trend. Če vpeljemo tehnični čas t, sta normalni enačbi za parametra linearnega trenda

$$T = b_o + b_1 t \qquad (21.41)$$

enaki

$$\sum Y = b_o N \; ; \; \sum Yt = b_1 \sum t^2 \qquad (21.42)$$

Parametra b_o in b_1 sta torej:

$$b_o = \frac{1}{N} \sum Y \; ; \quad b_1 = \frac{\sum Y t}{\sum t^2} \qquad (21.43)$$

$\sum t^2$ izračunamo za poljuben N po obrazcih

$$\sum_{1}^{N} t^2 = \frac{1}{2} \binom{N+1}{3} \quad \text{če je } N = 2i+1$$

$$\sum_{1}^{N} t^2 = 2 \binom{N+1}{3} \quad \text{če je } N = 2i \qquad (21.44)$$

Za $2 \leq N \leq 30$ so po obrazcih 21.44 izračunane vrednosti $\sum t^2$ v tabeli 21.20.

Tabela 21.20. $\sum t^2$ za $2 \leqslant N \leqslant 30$

N		2	3	4	5	6	7	8	9	10
$\sum t^2$		2	2	20	10	70	28	168	60	330
N	11	12	13	14	15	16	17	18	19	20
$\sum t^2$	110	572	182	910	240	1360	408	1938	570	2660
N	21	22	23	24	25	26	27	28	29	30
$\sum t^2$	770	3542	1012	4600	1300	5850	1638	7308	2030	8990

Da bi dobili parametra b_o in b_1, je treba iz časovne vrste izračunati le $\sum Y$ in $\sum Yt$ ter uporabiti obrazca 21.43 in tabelo 21.20.

Linearen trend, transformiran na prvi člen kot izhodišče, izračunamo po obrazcih

$$T = \left[b_o - \frac{b_1(N-1)}{2}\right] + b_1 x \quad \text{za } N = 2i+1$$

$$T = \left[b_o - b_1(N-1)\right] + 2b_1 x \quad \text{za } N = 2i \qquad (21.45)$$

21.65 Če za naravni prirastek na tisoč prebivalcev za SFRJ po nakazani metodi izračunamo linearen trend, dobimo:

Leto	Y	t	Yt	T
1955	15,5	-9	-139,5	15,18
1956	14,8	-7	-103,6	14,80
1957	13,2	-5	- 66,0	14,43
1958	14,7	-3	- 44,1	14,05
1959	13,5	-1	- 13,5	13,68
1960	13,6	1	13,6	13,30
1961	13,7	3	41,1	12,93
1962	12,0	5	60,0	12,55
1963	12,5	7	87,5	12,18
1964	11,4	9	102,6	11,80
	$\sum Y = 134,9$		$\sum Yt = -61,9$	$134,90 = \sum T$

$\sum t^2 = 330$

$$b_o = \frac{1}{N}\sum Y = \frac{134,9}{10} = 13,49; \quad b_1 = \frac{\sum Yt}{\sum t^2} = \frac{-61,9}{330} = -0,1876$$

$T = b_o + b_1 t = 13,49 - 0,1876\, t$ ali transformiran

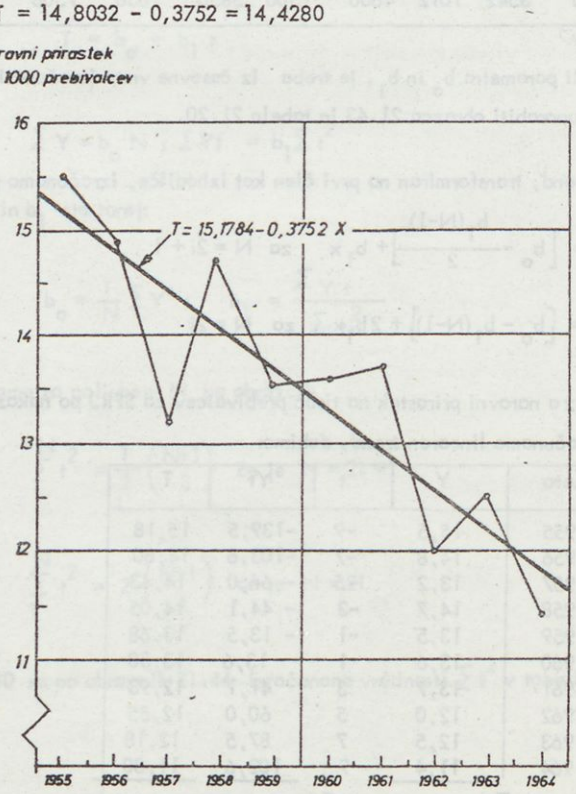
$$T = \left[b_o - b_1 (N-1)\right] + 2 b_1 x = 15,1784 - 0,3752\, x$$

Ker je $T = T_k - T_{k-1} = 2 b_1$, dobimo vrsto trenda, če začetnemu trendu

$T_o = 15,1784$ kumulativno prištevano $2 b_1 = -0,3752 = \Delta T$

$T_1 = T_o + \Delta T = 15,1784 - 0,3752 = 14,8032$

$T_2 = T_1 + \Delta T = 14,8032 - 0,3752 = 14,4280$



Slika 21.17. Trend za naravni prirastek prebivalstva za SFRJ v razdobju 1955-1964.

21.66 Ortogonalni polinomi. Funkcijo trenda, ki je dana s polinomom stopnje p

$$T = \sum_{k=o}^{p} a_k x^k \tag{21.46}$$

moremo transformirati v linearno zvezo ortogonalnih polinomov $X_k(x)$

$$T = \sum_{k=o}^{p} A_k \cdot X_k(x) \tag{21.47}$$

Ortogonalni polinom stopnje k

$$X_k(x) = \sum_{h=o}^{k} b_h x^h \tag{21.48}$$

je cela racionalna funkcija stopnje k, ki ima v odnosu do ortogonalnih polinomov drugih stopenj tole lastnost

$$\sum_{i=1}^{N} X_k(x_i) X_g(x_i) = 0 \; ; \quad k \neq g \tag{21.49}$$

Če vzamemo, da je polinom stopnje nič $X_o = 1$ konstanta (najenostavnejše je, da je ta konstanta 1) in polinom stopnje ena $X_1 = t$, spoznamo, da sta si X_o in X_1 ortogonalna, ker velja

$$\sum X_o X_1 = \sum 1 . t = 0$$

Če upoštevamo lastnosti ortogonalnih polinomov, sistem p+1 normalnih enačb degenerira v p+1 enačb

$$\sum Y X_k = A_k \sum X_k^2 \quad (k = 0,1 \ldots p) \tag{21.50}$$

vsaka z enim parametrom.

Iz sistema enačb 21.50 dobimo, da so parametri trenda

$$A_k = \frac{\sum Y X_k}{\sum X_k^2} \qquad (21.51)$$

Iz lastnosti ortogonalnih polinomov sledi dalje, da je

$$\sum_1^N (Y - T)^2 = \sum_1^N (Y - \sum_k A_k X_k)^2 = \sum Y^2 - \sum_k \frac{(\sum Y X_k)^2}{\sum X_k^2} \qquad (21.52)$$

pri čemer je

$$KS_k = \frac{(\sum Y X_k)^2}{\sum X_k^2} \qquad (21.53)$$

prispevek polinoma k variabilnosti Y.

21.67 Binomske funkcije. Celo racionalno funkcijo trenda stopnje p (21.46) moremo pisati tudi kot polinom binomskih funkcij $B_k(x) = \binom{x}{k}$

$$T = \sum_{k=0}^{p} c_k \binom{x}{k} \qquad (21.54)$$

Pri tem se izkaže, da so parametri

$$c_k = \Delta^k T_o^p \qquad (21.55)$$

zaradi lastnosti binomskih koeficientov k-te razlike trenda stopnje p v prvem členu. Zaradi te lastnosti moremo za funkcijo trenda, ki je cela racionalna funkcija stopnje p, dobiti vrsto trenda z enostavnim kumuliranjem iz izvornih $\Delta^k T_o^p$ v prvem členu.

Tako dobimo npr. vrsto vrednosti za polinom druge stopnje

$$T = 30 + 2\binom{x}{1} + 1\binom{x}{2}$$

za cele vrednosti x po tejle shemi:

x	$\Delta^2 T$	ΔT	T
0	1	2	30
1	1	2+1=3	30+2=32
2	1	3+1=4	32+3=35
3	1	4+1=5	35+4=39
4	1	5+1=6	39+5=44
5	1	6+1=7	44+6=50

Veljajo splošni obrazci:

$$\Delta^k T^P_{x+1} = \Delta^k T^P_x + \Delta^{k+1} T^P_x \qquad (21.56)$$

ali po vrsti

$$\begin{aligned} T^P_{x+1} &= T^P_x + \Delta T^P_x \\ \Delta T^P_{x+1} &= \Delta T^P_x + \Delta^2 T^P_x \\ \Delta^2 T^P_{x+1} &= \Delta^2 T^P_x + \Delta^3 T^P_x \end{aligned} \qquad (21.57)$$

Ker je ta način izredno prikladen za računanje vrste vrednosti za cele vrednosti x cele racionalne funkcije, nakažimo zvezo med $\Delta^k T_o$ in parametri a_k za polinome prve, druge in tretje stopnje iz 21.46.

Premica $T = a_o + a_1 x$

$$T_o = a_o ; \quad \Delta T_o = a_1$$

Parabola 2. stopnje $T = a_o + a_1 x + a_2 x^2$

$$T_o = a_o ; \quad \Delta T_o = a_1 + a_2 ; \quad \Delta^2 T_o = 2 a_2$$

Parabola tretje stopnje $T = a_o + a_1 x + a_2 x^2 + a_3 x^3$

$T_o = a_o$; $\Delta T_o = a_1 + a_2 + a_3$; $\Delta^2 T_o = 2a_2 + 6a_3$; $\Delta^3 T_o = 6a_3$

Tako dobimo npr. za polinom

$$T = 30 + 5x + 2x^2 + 3x^3$$

$T_o = 30$; $\Delta T_o = 5+2+3=10$; $\Delta^2 T_o = 2.2+6.3 = 22$ $\Delta^3 T_o = 6.3 = 18$

Iz tega pa

x	$\Delta^3 T$	$\Delta^2 T$	ΔT	T
0	18	22	10	30
1	18	40	32	40
2	18	58	72	72
3	18	76	130	144
4	18	94	206	274
5	18	112	300	480

Določanje polinomov trenda po metodi ortogonalnih polinomov binomskih funkcij

21.68 Metoda ortogonalnih polinomov binomskih funkcij ima veliko prednosti pred drugimi metodami, ker združuje analitične lastnosti ortogonalnih polinomov in tehnične prednosti polinomov binomskih funkcij. Množično množenje osnovnih podatkov vrednosti ortogonalnih polinomov zamenja kumulativno seštevanje. Z uporabo matrik normiranih koeficientov je poenostavljeno računanje parametrov trenda A_k, če ga pišemo kot polinom ortogonalnih polinomov X, $T = \sum A_k X_k$, enako pa je z uporabo iste matrike poenostavljeno tudi računanje parametrov $\Delta^k T_o^p$ za trend kot polinom binomskih funkcij $T = \sum_{k=o}^{p} \Delta^k T_o^p \binom{x}{k}$, centriran na prvi člen v osnovni časovni vrsti. Vrsto trenda računamo iz polinoma binomskih funkcij s kukumalivnim seštevanjem diferenc ustreznega reda. Z vmesnim izračunanjem parametrov polinoma ortogonalnih polinomov dobimo brez dodatnega računanja parametre trendov nižjih

stopenj in doprinos posamezne stopnje k variabilnosti časovne vrste.

21.69 Splošen postopek računanja funkcije trenda stopnje p

$$T_x^p = \sum_{h=o}^{p} \Delta^h T_o^p \binom{x}{h}$$

za časovno vrsto Y_x ima tele faze:

a) Iz osnovne časovne vrste Y_x izračunamo vsote kumulativ C_h ($h = 0,1,2 \ldots p$) do stopnje p.

b) Iz matrike konstant $d_{hk} | D_k$ in vektorja C_h izračunamo parametre trenda po obrazcih

$$\frac{d_{kh} | D_k}{C_h} \; ; \quad \sum_{h=k}^{o} d_{kh} C_h = E_k \; ; \quad A_k = (-1)^k \frac{E_k}{D_k} \; ; \quad s_k^2 = \frac{E_k^2}{D_k} \tag{21.58}$$

$$T_x^p = \sum_{k=o}^{p} A_k X_k \; ; \quad X_k = \sum_{h=k}^{o} d_{kh} \binom{x}{h} \tag{21.59}$$

Trend, podan kot polinom binomskih funkcij pa je

$$T_x^p = \sum_{h=o}^{p} \Delta^h T_o^p \binom{x}{h} \; ; \quad \Delta^h T_o^p = \sum_{k=h}^{p} A_k d_{kh} \tag{21.60}$$

Vsoto kvadratov odklonov stvarnih vrednosti Y_x od vrednosti trenda T_x^p računamo po obrazcu

$$\sum_{x=o}^{N-1} (Y_x - T_x^p)^2 = \sum_{x=o}^{N-1} Y_x^2 - \sum_{k=o}^{p} \frac{E_k^2}{D_k} = \sum_{x=o}^{N-1} Y_x^2 - \sum_{k=o}^{p} s_k^2 \tag{21.61}$$

Varianco iz odklonov pa izračunamo po obrazcu

$$\sigma_e^2 = \frac{1}{N} \left[\sum_x Y_x^2 - \sum_k s_k^2 \right] \tag{21.62}$$

Tabela 21.21. Izračunanje paraboličnega trenda tretje stopnje za indeks proizvodnje v elektroindustriji 1946-1965

Leto	Y	C_o	C_1	C_2	$\Delta^3 T_o^3$	$\Delta^2 T_x^3$	ΔT_x^3	T_x^3	Y_x/T_x^3
1	2	3	4	5	6	7	8	9	10
1946	18	0	0	0	1,40780	-2,20884	15,43334	T_o^3=18,4	0,98
1947	23	18	0	0		-0,80104	13,27450	33,9	0,82
1948	52	46	18	0		+0,60676	12,47346	47,2	1,10
1949	73	98	64	18		2,01456	13,08022	59,7	1,22
1950	72	171	162	82		3,42236	15,09478	72,7	0,99
1951	78	243	333	244		4,83016	18,51714	87,8	0,89
1952	100	321	576	577		6,23796	23,34730	106,3	0,94
1953	132	421	897	1153		7,64576	29,58526	129,7	1,02
1954	155	553	1318	2050		9,05356	37,23102	159,3	0,97
1955	198	708	1871	3368		10,46136	46,28458	196,5	1,01
1956	229	906	2579	5239		11,86916	56,74594	242,8	0,94
1957	310	1135	3485	7818		13,27696	68,61510	299,5	1,04
1958	378	1445	4620	11303		14,68476	81,89206	368,2	1,03
1959	464	1823	6065	15923		16,09256	96,57682	450,0	1,03
1960	576	2287	7888	21988		17,50036	112,66938	546,6	1,05
1961	645	2863	10175	29876		18,90816	130,10974	659,3	0,98
1962	725	3508	13038	40051		20,31596	149,07790	789,5	0,92
1963	913	4233	16546	53089		21,72376	169,39386	938,5	0,97
1964	1205	5146	20779	69635		23,13156	191,11762	1107,9	1,09
1965	1262	6351	25925	90414		24,53936	214,24918	1299,1	0,97
		7613	32276	116339	352828				
		C_o	C_1	C_2	C_3				

ali ocenimo po obrazcu

$$s_e^2 = \frac{1}{N-p-1}\left[\sum_x Y_x^2 - \sum_k s_k^2\right] \qquad (21.63)$$

če gre za slučajnostne odklone od trenda.

21.70 Za zgled vzemimo indeks proizvodnje v elektroindustriji SFRJ v razdobju 1946-1965.

Časovna vrsta ima $N = 20$ členov. Kot pokaže grafikon, še jasneje pa prispevek posameznih stopenj k variabilnosti vrste Y_x, je za opis osnovne smeri razvoja navedenega pojava primerna parabola tretje stopnje.

Prva faza - računanje kumulativnih vsot - C_o, C_1, C_2, C_3 za izračunanje parametrov je nakazana v stolpcih 2-5 v tabeli 21.21.

Drugo fazo - izračunanje parametrov $\overset{k}{\triangle} T_o^3$ izvršimo po obrazcih 21.58, 21.59 in 21.60. Iz zbirke tabel prepišemo matriko $d_{kh}|D_k$ za $N = 20$, podpišemo kumulativne vsote C_h, izračunamo po shemi $E_k = \sum_h d_{kh} C_h$, parametre $A_k = (-1)^k \frac{E_k}{D_k}$ in prispevke posameznih delov k varianci $s_k^2 = \frac{E_k^2}{D_k}$

Tabela 21.22. Izračun parametrov trenda za indeks proizvodnje v elektroindustriji

	3	2	1	0	D_k	E_k	A_k	s_k^2
0				1	20	7613	380,65000	2897888,4
1			2	-19	2660	-80095	30,11090	2411732,7
2		2	-18	57	17556	85651	4,87873	417868,2
3	20	-170	612	-969	4903140	-345155	0,07039	24297,1
	352828	116339	32276	7613				
	C_3	C_2	C_1	C_o				

Iz matrike koeficientov d_{kh} in parametrov ortogonalnih polinomov A_k dobimo

parametre $\Delta^k T_o^p$ za polinome binomskih funkcij.

$$\Delta^3 T_o^3 = d_{33} A_3 = 20 \cdot 0,07039 = 1,40780$$

$$\Delta^2 T_o^2 = d_{22} A_2 = 2 \cdot 4,87873 = 9,75746$$

$$\Delta^2 T_o^3 = \Delta^2 T_o^2 + d_{32} A_3 = 9,75746 + (-170) \cdot 0,07039 = -2,20884$$

$$\Delta T_o^1 = d_{11} A_1 = 2 \cdot 30,11090 = 60,22180$$

$$\Delta T_o^2 = \Delta T_o^1 + d_{21} A_2 = 60,22180 + (-18) \cdot 4,87873 = -27,59534$$

$$\Delta T_o^3 = \Delta T_o^2 + d_{31} A_3 = -27,59534 + 612 \cdot 0,07039 = 15,48334$$

$$T_o^o = d_{oo} A_o = 1 \cdot 380,6500 = 380,65000$$

$$T_o^1 = T_o^o + d_{10} A_1 = 380,65000 + (-19) \cdot 30,11090 = -191,45710$$

$$T_o^2 = T_o^1 + d_{20} A_2 = -191,45710 + 57 \cdot 4,87873 = 86,63051$$

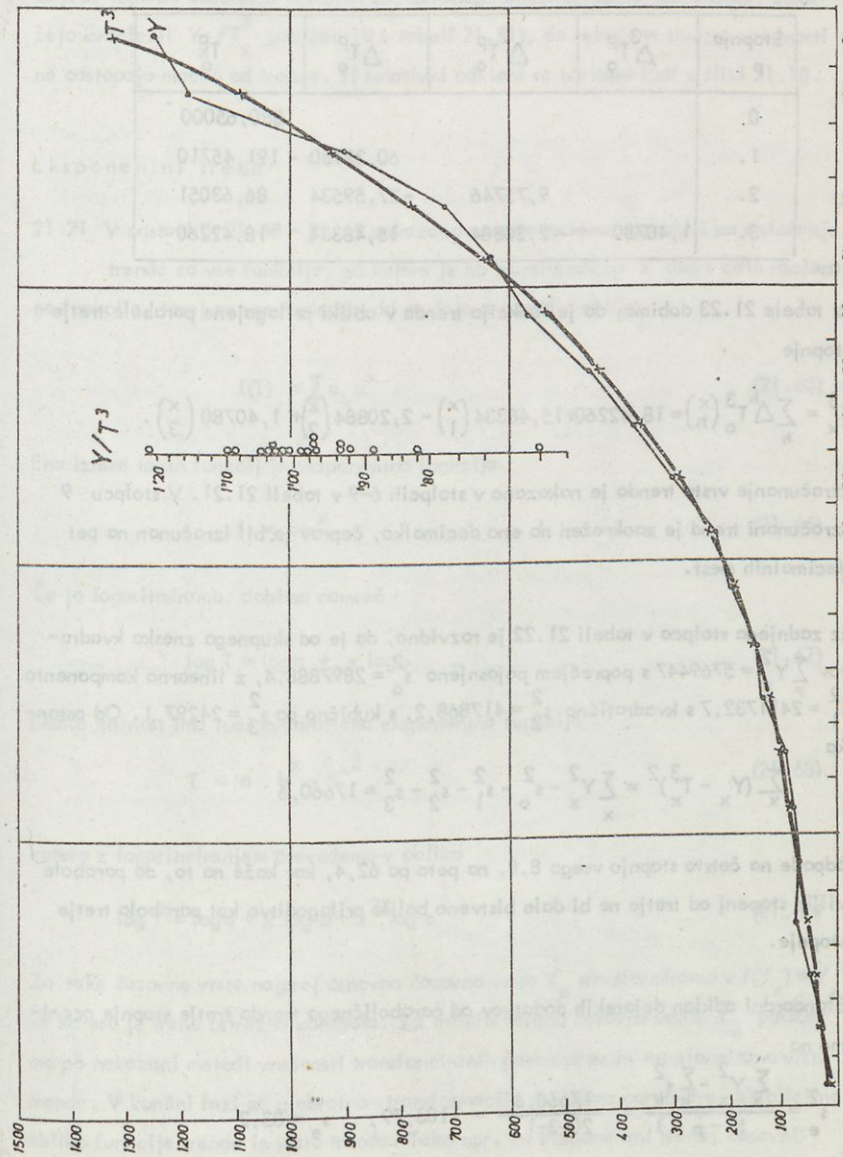
$$T_o^3 = T_o^2 + d_{3o} A_3 = 86,63051 + (-969) \cdot 0,07039 = 18,42260$$

Tako smo postopno izračunali parametre za vse prilagojene polinome do tretje stopnje.

Kot je razvidno iz sheme, še bolj neposredno pa iz izračuna za naš zgled, dobimo ustrezen parameter za stopnjo polinoma p $\Delta^h T_o^p$ po obrazcu

$$\Delta^h T_o^p = \Delta^h T_o^{p-1} + d_{ph} A_h \qquad (21.64)$$

V tabeli 21.23 so ponovljeni parametri prilagojenih polinomov do tretje stopnje.



Slika 21.18. Trend za indeks proizvodnje v elektroindustriji v SFRJ v razdobju 1946-1965

Tabela 21.23. Parametri za prilagojene polinome do 3. stopnje

Stopnja p	$\Delta^3 T_o^p$	$\Delta^2 T_o^p$	ΔT_o^p	T_o^p
0				380,65000
1.			60,22180	- 191,45710
2.		9,75746	- 27,59534	86,63051
3.	1,40780	- 2,20884	15,48334	18,42260

Iz tabele 21.23 dobimo, da je funkcija trenda v obliki prilagojene parabole tretje stopnje

$$T_x^3 = \sum_h \Delta^h T_o^3 \binom{x}{h} = 18,42260 + 15,48334 \binom{x}{1} - 2,20884 \binom{x}{2} + 1,40780 \binom{x}{3}.$$

Izračunanje vrste trenda je nakazano v stolpcih 6-9 v tabeli 21.21. V stolpcu 9 izračunani trend je zaokrožen na eno decimalko, čeprav je bil izračunan na pet decimalnih mest.

Iz zadnjega stolpca v tabeli 21.22 je razvidno, da je od skupnega zneska kvadratov $\sum_x Y_x^2 = 5769447$ s poprečjem pojasnjeno $s_o^2 = 2897888,4$, z linearno komponento $s_1^2 = 2411732,7$ s kvadratično $s_2^2 = 417868,2$, s kubično pa $s_3^2 = 24297,1$. Od ostanka

$$\sum_x (Y_x - T_x^3)^2 = \sum_x Y_x^2 - s_o^2 - s_1^2 - s_2^2 - s_3^2 = 17660,6$$

odpade na četrto stopnjo vsega 8,8, na peto pa 62,4, kar kaže na to, da parabole višjih stopenj od tretje ne bi dale bistveno boljše prilagoditve kot parabola tretje stopnje.

Standardni odklon dejanskih podatkov od paraboličnega trenda tretje stopnje ocenimo na

$$s_e^2 = \frac{\sum_x Y_x^2 - \sum_k s_k^2}{N - p - 1} = \frac{17660,6}{20-3-1} = 1103,79; \quad s_e = 33,2$$

Čeprav izgleda absolutna vrednost standardnega odklona razmeroma velika, pokažejo kvocienti Y_x/T_x^3 (stolpec 10 v tabeli 21.21), da relativne stvarne vrednosti ne odstopajo močno od trenda. Ti relativni odkloni so narisani tudi v sliki 21.18.

Eksponentni trend

21.71 V odstavkih 21.68 - 21.69 nakazano metodo moremo uporabiti za določanje trenda za vse funkcije, za katere je po transformaciji x dan s celo racionalno funkcijo, torej za vse funkcije, ki se dajo prevesti v obliko

$$f(T) = \sum_k a_k x^k \qquad (21.65)$$

Ena izmed takih funkcij je eksponentna funkcija

$$T = ab^x \qquad (21.66)$$

Če jo logaritmiramo, dobimo namreč

$$\log T = \log a + x \log b \qquad (21.67)$$

Enako lastnost ima tudi kvadratična eksponentna funkcija

$$T = a \cdot b^x \cdot c^{x^2} \qquad (21.68)$$

katero z logaritmiranjem prevedemo v obliko

$$\log T = \log a + x \cdot \log b + x^2 \cdot \log c \qquad (21.69)$$

Za take časovne vrste najprej osnovno časovno vrsto Y_x transformiramo v $f(Y_x) = \dot{Y}_x$, za katero je trend ustrezna parabola. Za transformirano časovno vrsto $\dot{Y}_x$ poiščemo po nakazani metodi vrednosti transformiranih parametrov in transformirano vrsto trenda. V končni fazi pa z obratno transformacijo poiščemo parametre, eksplicitno obliko funkcije trenda in vrsto trenda. Tako npr. za eksponentni trend, časovni vrsti Y_x poiščemo ustrezno vrsto $\dot{Y}_x = \log Y_x$, zanjo izračunamo parametre $\log a$

logb in vrsto linearnega trenda za log T_x. Z antilogaritmiranjem parametrov loga, logb dobimo parametre a in b, z antilogaritmiranjem vrste log T_x pa vrsto trenda T_x.

Opozoriti moramo, da lastnost, da je vsota kvadratov odklonov stvarnih vrednosti od trenda najmanjša, ne velja v tem primeru za osnovne podatke in trend, temveč za transformirane vrste

$$\sum_x \left[f(Y_x) - f(T_x) \right]^2 = \min \qquad (21.70)$$

21.72 Kot zgled vzemimo letne indekse industrijske proizvodnje potrošnega blaga v SFRJ v razdobju 1952-1964 (1952=100) Vir: Jugoslavija 1945-1964 - Statistički pregled.

S preskusom značilnosti trenda z verižnimi indeksi preskusimo, ali je trend osnovnih podatkov eksponenten. Ta preskus moremo izvršiti s preskusom značilnosti Spearmanovega koeficienta korelacije med rangi za vrsto verižnih indeksov in časom.

Tabela 21.24. Preskus o eksponentnosti trenda za indeks industrije pri proizvodnji potrošnega blaga v SFRJ.

x leto	Y_x	$Y_x / Y_{x-1} = K_x$	R_x	R_{k_x}	d_R
1952	100				
1953	113	1,13	1	6	+ 5
1954	130	1,15	2	7,5	+ 5,5
1955	146	1,12	3	5	+ 2
1956	162	1,11	4	3,5	- 0,5
1957	197	1,22	5	12	+ 7
1958	216	1,10	6	2	- 4
1959	248	1,15	7	7,5	+ 0,5
1960	289	1,17	8	10,5	+ 2,5
1961	312	1,08	9	1	- 8
1962	346	1,11	10	3,5	- 6,5
1963	405	1,17	11	10,5	- 0,5
1964	469	1,16	12	9	- 3
					0 $\sum d_R^2 = 246,50$

$$r_s = 1 - \frac{6 \cdot 246,5}{12(12^2 - 1)} = 0,1381$$

t preskus je pokazal, da je r_s neznačilno različen od nič, iz česar sklepamo, da ni vzroka, da ne štejemo vrsto verižnih indeksov kot vrsto, v kateri ni trenda. Iz tega sklepamo, da moremo vzeti za naš zgled za trend eksponentno funkcijo.

Postopek za računanje funkcije in vrste eksponentnega trenda za indeks industrijske proizvodnje potrošnega blaga v SFRJ je nakazan v tabeli 21.25.

Tabela 21.25. Izračunanje eksponentnega trenda za indeks industrijske proizvodnje potrošnega blaga v SFRJ

leto	x	Y_x	$y_x=\log Y_x-2$	Cy_x	$\log T_x$	T_x	$\log Y_x - \log T_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1952	0	100	0,0000	0	2,000926	100,2	-0,0009
1953	1	113	0,0531	0	2,0564	113,9	- 33
1954	2	130	0,1139	0,0531	2,1118	129,4	+ 21
1955	3	146	0,1644	0,1670	2,1673	147,0	- 27
1956	4	162	0,2095	0,3314	2,2227	167,0	- 132
1957	5	197	0,2945	0,5409	2,2782	189,8	+ 163
1958	6	216	0,3345	0,8354	2,3336	215,6	+ 9
1959	7	248	0,3945	1,1699	2,3891	245,0	+ 54
1960	8	289	0,4609	1,5644	2,4445	278,3	+ 164
1961	9	312	0,4942	2,0253	2,5000	316,2	- 58
1962	10	346	0,5391	2,5195	2,5554	359,3	- 163
1963	11	405	0,6075	3,0586	2,6109	408,2	- 34
1964	12	469	0,6712	3,6661	2,6663	463,8	+ 49
				4,3373=C_o			
				15,9316=C_1			

Za linearen trend za časovne vrste z N = 13 je matrika koeficientov

d_{kh}	D_k	$E_k = \sum_h d_{kh} C_h$	$A_k = (1)^k \frac{E_k}{D_k}$	$\sum_x y_x^2 = 2,00781813$
1 : 13		4,3373	0,333638	$-s_o^2 = -1,44709010$
1- 6 : 182		-10,0922	0,055452	$-s_1^2 = -0,55962912$

C_k	15,9316	4,3373
ΔT_o^k	0,055452	0,000926
$\Delta \log T_x$	$\log T_x$	- 2

$(\log Y - \log T)^2 = 0,00109891$

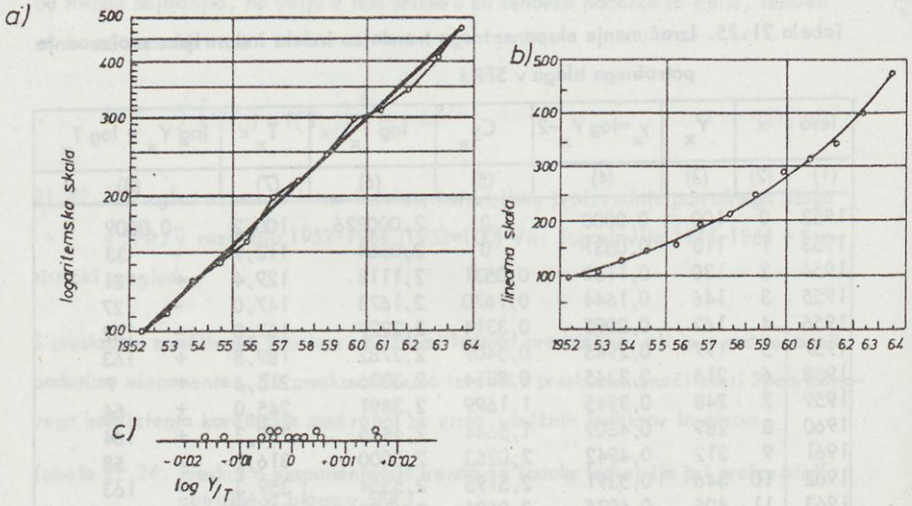
$s^2_{\log Y/T} = \frac{0,00109891}{13-1-1} =$

$= 0,00009901$

$s_{\log Y/T} = 0,0100$

$s'_{y/T} = 1,023$

Da bi računali z manjšimi števili, smo od logaritmov indeksov Y_x v stolpcu 4 odšteli konstanto 2. To vpliva le na parameter $\log T_x$, ki ga korigiramo tako, da mu prištejemo 2. C_o dobimo kot člen prve kumulative. $\log T_x$ so sproti zaokroževani na 4 decimalna mesta, čeprav so bile računane kumulative s parametri na šest decimalnih mest.



Slika 21.19. Eksponentni trend za indeks industrijske proizvodnje potrošnega blaga v SFRJ v razdobju 1952-1964 (1952 = 100)

Kot pričakujemo je v pasu

$$\log T_x \pm s_{\log Y/T}$$

ca 2/3 vseh vrednosti $\log Y_x$,

v pasu $T_x / s'_{y/T}$ do $T_x \cdot s'_{y/T}$

ca 2/3 vseh vrednosti Y_x, pri čemer je $s'_{y/T}$ antilogaritem od $s_{\log y/T}$. V našem zgledu je $s'_{y/T} = 1{,}023$ in meri relativen odklon stvarnih vrednosti od eksponentnega trenda.

PERIODIČNE VARIACIJE

21.73 Periodične variacije proučujemo za več namenov. Pomembno je, da poznamo periodične variacije v preteklosti zaradi medsebojne primerjave in analize. Periodične variacije proučujemo tudi za napovedovanja za krajša razdobja.

Za socialno-ekonomska proučevanja so še posebej pomembne sezonske variacije, to so periodične variacije z letno periodo. Poznavanje sezonskih in periodičnih variacij je važno tudi za kratkoročno planiranje najrazličnejših pojavov. Za vrsto dejavnosti je nepravilno linearno razdeljevanje letnega plana na mesečne plane in moramo zanje upoštevati sezonski značaj določenih pojavov. To pa moremo le, če poznamo sezonsko sestavino za te pojave.

Sezonski značaj pojavov pa je samo eden izmed periodičnih vplivov. Tako ima na primer promet v trgovini na drobno mesečno periodičnost in opažamo v začetku vsakega meseca večji, proti koncu meseca pa manjši promet. Tudi tedenska periodičnost je tipična za veliko število pojavov. Tako pri potniškem prometu in z njim zvezanim številom prometnih nesreč, v prometu v gostinstvu itd. opazujemo tedensko periodičnost. Tudi dnevna periodičnost je zelo pogosta. Potrošnja električne energije ali vode, gostota prometa itd. imajo dnevno periodičnost. Znane so dnevne konice v potrošnji električne energije ali vode, konice v krajevnem prometu itd.

Bistvo periodičnih vplivov je v tem, da se na določeno časovno razdobje (npr. leto, mesec, teden, dan) ponavljajo: to izzove na proučevanih pojavih periodično vedno enak učinek.

V splošnem periodični vplivi v gospodarstvu niso zaželeni. Periodičen značaj dovede ali do zastoja v času sezone ali do neizkoriščenja kapacitet izven sezone, če imamo zadostne kapacitete za čas maksima. Zato si prizadevamo, da periodične vplive omilimo, če jih že ne moremo odpraviti. Tipičen primer je nadomeščanje proizvodnje hidro in termo elektrike. Obe se medsebojno dopolnjujeta, tako da čimbolj omilita sezonski značaj proizvodnje hidroelektrarn. Periodične vplive moramo nujno poznati in proučevati, če skušamo njihov učinek ublažiti in odpraviti.

Nakazali smo, da je bistvo periodičnih vplivov v tem, da se na določene časovne razmike ponavljajo. Vendar tudi ta učinek ni statičen. V daljšem razdobju opažamo spremenljivost sezonskih vplivov. Te spremembe prinesejo bodisi razvoj ali zavestni ukrepi, s katerimi vplivamo na sezonski značaj pojavov.

Ena izmed važnih nalog analize časovnih vrst je razstavljanje časovne vrste na njene osnovne sestavine. Tudi zaradi tega je važno, da poznamo sezonsko sestavino, da jo moremo iz osnovne časovne vrste odstraniti. Če iz časovne vrste odstranimo trend in sezonske variacije, ostane namreč v časovni vrsti le še ciklična in slučajna sestavina. Sezonsko sestavino moramo zato poznati tudi, če analiziramo ciklična nihanja pojavov.

21.74 Sezonsko oziroma periodično sestavino v časovnih vrstah določamo po več metodah. Vsaka izmed njih je prilagojena posebnostim različnih časovnih vrst. Zato je odločitev o tem, katero izmed teh metod v posebnem primeru uporabimo, odvisna od dinamike pojava, ki ga časovna vrsta prikazuje oziroma od modela, ki ponazarja povezavo sestavin v časovni vrsti.

Čeprav smo uvodoma nakazali, da je značilno za sezonsko oziroma periodično sestavino, da se v razdobju ene periode ponavlja z enakim učinkom, spoznamo, da se more tudi sezonska sestavina v času spremeniti oziroma spreminjati.

Razvoj turizma prav gotovo ne vpliva samo na trend, temveč glede na različne spreminjajoče se sestavine tudi na sezonsko sestavino. Vpeljavanje nove tehnologije v gradbeništvo ima za posledico, da se sezonska sestavina z leti spreminja. Težnja, ki prevladuje, je v tem, da se sezonska sestavina časovno manjša.

Statičnost periodičnosti je v glavnem samo predpostavka, ki velja kot približek za krajša razdobja. Za daljša razdobja pa opažamo tudi pri sezonski oziroma periodični sestavini določeno dinamiko.

Od metod, s katerimi proučujemo sezonsko oziroma periodično sestavino bomo obravnavali:

a) metodo vsot,

b) metodo popravljenih vsot,

c) metodo koeficientov dinamike,

d) metodo kvocientov na drseče sredine.

Vsaka izmed teh metod je vezana na določene pogoje, katerim mora zadoščati osnovna časovna vrsta. Še najsplošnejše uporabna je metoda kvocientov na drseče sredine. To metodo uporabljamo tudi v primeru, če proučujemo dinamično sestavino. S prvimi tremi metodami pa po pravilu proučujemo statično periodično sestavino, ki v splošnem dajo neke poprečne sezonske indekse.

Sezonsko sestavino izražamo kot sezonske indekse

$$P_m = 100\,(1 + p_m) \qquad (21.71)$$

Če sezonske indekse delimo s 100, dobimo sezonsko sestavino, izraženo s koeficientom. Učinek sezonske sestavine na ustrezen člen dobimo, če upoštevamo (1+p) kot faktor, s katerim pomnožimo učinke drugih sestavin.

Metoda vsot

21.75 Če v časovni vrsti v proučevanem časovnem razmiku ni opaziti izrazite smeri razvoja, ki je rezultat trenda in ciklične sestavine, za določitev periodične sestavine s pridom uporabljamo metodo vsot, ki je od vseh najenostavnejša.

Za take časovne vrste je model časovne vrste takle:

$$Y_{lm} = A\,(1 + p_m + e_{lm}) \qquad (21.72)$$

Pri tem je Y_{lm} = osnovni podatek za člen m v periodi l; p_m = periodična sestavina za člen m; e_{lm} = rezultat posamičnih vplivov.

Za ta primer dobimo pri znanih lastnostih sredin vrsto periodičnih indeksov po temle postopku:

a) Da bi odstranili posamične vplive, izračunamo za vsako obdobje periode iz ustreznih mesečnih vrednosti za vse periode proučevanega razdobja vsote $S_m = \sum_l Y_{lm}$. Zaradi lastnosti, ki jih pripisujemo posamičnim vplivom, velja

$$S_m = NA\,(1 + p_m) \tag{21.73}$$

Dobljene vsote so sorazmerne sezonskim indeksom $1+p_m$.

b) Vsote S_m prevedemo v periodične s popravnim faktorjem

$$c = \frac{100 \cdot P}{\sum_{m=1}^{P} S_m} \tag{21.74}$$

v čiste periodične indekse. Pri tem je P število členov v periodi. Po tem je periodični indeks

$$P_m = 100\,(1 + p_m) = c \cdot S_m \tag{21.75}$$

Dnevna, tedenska in mesečna periodična nihanja imajo krajšo periodo kot sezonska nihanja. Zato metodo vsot uporabljamo najpogosteje za določanje le-teh. Za krajša razdobja moremo šteti, da sta trend ali ciklična sestavina konstantna. Tak primer je npr. z dnevno periodičnostjo porabe vode, porabo električne energije, številom potnikov v krajevnem prometu itd., za katere določamo periodično sestavino že iz podatkov za nekaj dni ali nekaj tednov.

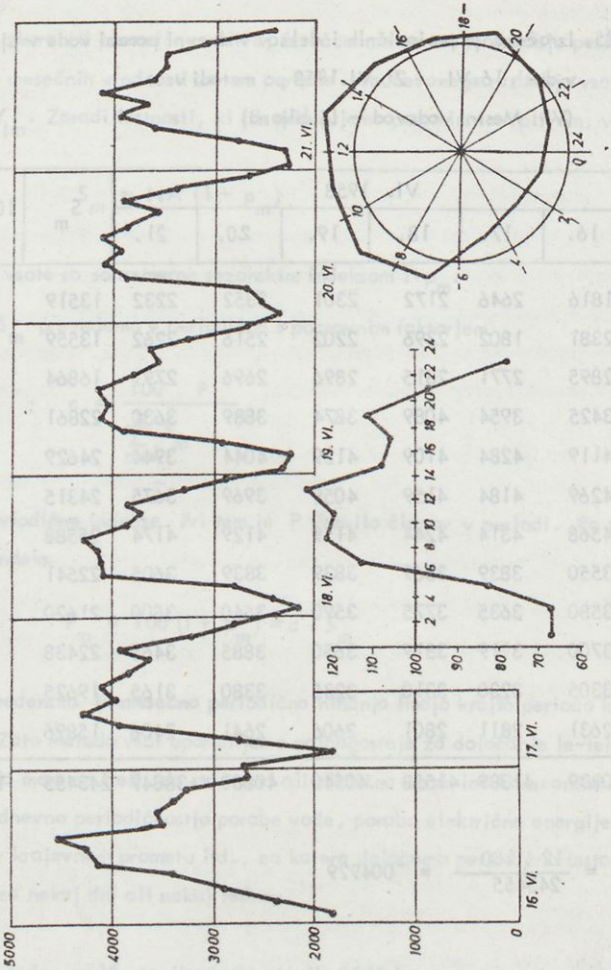
21.76 Za teden od 16. junija do 21. junija 1958 imamo vrsto o dveurni porabi vode v Ljubljani (Vir: Mestni vodovod Ljubljana). Po metodi vsot je treba določiti periodo v dnevni porabi vode v tem razdobju. Uporaba te metode je upravičena, ker se v razdobju šestih dni ne pokažejo niti spremembe trenda niti drugi časovni vplivi, ki bi onemogočili uporabo te metode.

Tabela 21.25. Izračunanje periodičnih indeksov v dnevni porabi vode v Ljubljani v dneh 16.VI. - 21.VI.1958 po metodi vsot.

(Vir: Mestni vodovod - Ljubljana)

Datum / Ura	VI. 1958						S_m	$100(1+p_m)$
	16.	17.	18.	19.	20.	21.		
0^h - 2^h	1816	2646	2172	2301	2352	2232	13519	67
2 - 4	2381	1802	2396	2202	2516	2262	13559	67
4 - 6	2895	2771	2815	2896	2696	2791	16864	83
6 - 8	3425	3954	4089	3874	3889	3630	22861	113
8 - 10	4119	4284	4109	4129	4044	3944	24629	121
10 - 12	4269	4184	4159	4059	3969	3675	24315	120
12 - 14	4568	4314	4244	4159	4129	4174	25588	126
14 - 16	3550	3839	3869	3839	3839	3605	22541	111
16 - 18	3580	3635	3775	3590	3540	3500	21620	107
18 - 20	3700	3919	3819	3650	3885	3465	22438	111
20 - 22	3305	3230	3310	3235	3380	3165	19625	97
22 - 24	2631	2811	2801	2606	2641	2406	15896	78
	40239	41389	41558	40540	40880	38849	243455	1201

$$c = \frac{12 \cdot 100}{243455} = \cdot 004929$$



Slika 21.20. Periodična sestavina v dnevni porabi vode v Ljubljani za razdobje 16.VI. - 21.VI. 1958

21.77 Metoda vsot s popravkom za smer razvoja. S primernim popravkom moremo metodo vsot uporabiti tudi za časovne vrste, za katere v proučevanem razdobju zasledimo smer razvoja. Ta smer razvoja je za proučevano razdobje, ki je običajno krajše, lahko rezultat trenda in cikličnih vplivov. Metoda

je teoretično vezana na predpostavko, da je v proučevanem razdobju smer razvoja eksponentna. Zato to metodo s pridom uporabljamo takrat, kadar je v proučevanem časovnem razdobju skupen učinek trenda in cikla monotono naraščajoč ali monotono padajoč.

Metoda je vezana na model časovne vrste

$$Y_{lm} = A \cdot b^x (1 + p_m) + e_{lm} \; ; \quad x = Pl + m \tag{21.76}$$

Ta model ustreza mnogim praktičnim primerom. Eksponentna funkcija je marsikdaj dober približek smeri razvoja v proučevanem razdobju. Naravno je tudi, da je periodična sestavina (1+p) s smerjo razvoja Ab^x vezana faktorialno. Če predpostavimo gornji model, iz osnovne časovne vrste Y_{lm} izluščimo sezonsko ali splošno periodično sestavino po temle postopku:

a) Iz osnovne časovne vrste, katere člani Y_{lm} se nanašajo na leto l in mesec m, izračunamo mesečne vsote S_m in letne vsote S_l.

b) Število let oziroma period v osnovni časovni vrsti je sodo $N = 2r$ ali liho $N = 2r+1$. Iz prvih r letnih vsot izračunamo vsoto S_s, iz zadnjih r letnih vsot pa vsoto S_z. (Če je N sodo število, so torej v vsoti S_s in S_z vključena vsa leta, če pa je N liho število, srednji člen ne upoštevamo.

c) Iz S_s in S_z ocenimo popravek za smer razvoja b po obrazcu

$$b = \sqrt[6N]{S_z / S_s} \tag{21.77}$$

če je število let sodo ($N = 2r$), in po obrazcu

$$b = \sqrt[6(N+1)]{S_z / S_s} \tag{21.78}$$

če je število let liho ($N = 2r+1$)

d) Iz popravnega faktorja b izračunamo geometrijsko zaporedje

$$1, \; b, \; b^2 \; \ldots\ldots \; b^{11} \tag{21.79}$$

e) Mesečne vsote S_m popravimo s potencami popravnega faktorja b

$$S'_m = S_m / b^{m-1} \tag{21.80}$$

f) Iz popravljenih vsot S'_m izračunamo sezonske indekse P_m po obrazcu

$$P_m = 100\,(1 + p_m) = kS'_m \tag{21.81}$$

pri čemer je

$$k = 1200 / \sum_m S'_m \tag{21.82}$$

popravni faktor.

21.78 Za zgled vzemimo časovno vrsto nočitev v SFRJ v razdobju 1960-1965

$$b = \sqrt[6N]{S_z / S_s} \qquad \sqrt[6.6]{\frac{89063}{70187}} = \sqrt[36]{1{,}268939} = 1{,}00663$$

Tabela 21.26. Izračunanje sezonske sestavine za število nočitev v SFRJ v razdobju 1960–1965 (po metodi popravljenih vsot).
(Vir: Jugoslavija 1955–1964 - Statistički pregled. Indeks).

(v tisočih)

	J	F	M	A	M	J	J	A	S	O	N	D	S_l
1960	549	541	574	697	1111	2214	5467	6526	2627	1176	840	700	23022
1961	680	621	657	746	1243	2299	5635	6779	2773	1225	825	699	24182
1962	674	582	594	719	1101	2344	5528	6683	2644	950	591	573	22983
1963	560	496	565	735	1227	2623	6612	7491	2939	1264	698	659	25869
1964	644	605	660	781	1685	3204	7625	8921	3544	1245	859	766	30539
1965	812	698	781	965	1659	3431	8194	9512	3671	1330	842	760	32655
S_m	3919	3543	3831	4643	8026	16115	39061	45912	18198	7190	4655	4157	159250
b^{m-1}	1,00000	1,00663	1,01330	1,02002	1,02678	1,03359	1,04044	1,04734	1,05428	1,06127	1,06831	1,07539	
S'_m	3919	3520	3781	4552	7817	15591	37543	43837	17261	6775	4357	3866	152819=$\sum S'_m$
$1+p_m$	30,8	27,6	29,7	35,7	61,4	122,4	294,8	344,2	135,5	53,2	34,2	30,4	1199,9
P_m	31	28	30	36	61	122	295	344	136	53	34	30	1200

$S_s = 23022 + 24182 + 22983 = 70187$

$S_z = 25869 + 30539 + 32655 = 89063$

$k = \frac{1200}{152819} = 0,0078524$

21.79 Potence b^{m-1} dobimo z računskim strojem s postopnim množenjem s konstantnim faktorjem k. Brez računskega stroja pa je najprikladneje računati popravljene vsote z logaritmiranjem. Najprej izračunamo s postopnim seštevanjem mnogokratnike od logb in jih odštejemo od ustreznih logaritmov $\log S_m$. Logaritme popravljenih vsot dobimo tako po obrazcu

$$\log S'_m = \log S_m - (m-1) \log b \qquad (21.83)$$

Če perioda ni sestavljena iz 12 členov, kot je to pri računanju sezonske sestavine iz mesečnih časovnih vrst, v obrazcu 21.77 in 21.78 nadomestimo 6 s P/2, pri čemer pomeni P = število členov v eni periodi, popravni faktor pa je $k = 100\ P/\sum_m S'_m$ Metoda vsot s popravkom za smer razvoja je najenostavnejša metoda za računanje sezonske sestavine, če je v časovni vrsti razen periodičnih in posamičnih vplivov še smer razvoja.

Hibe oziroma omejitve te metode pa so:

a) uporaba je omejena na predpostavko o eksponentni smeri razvoja;
b) predpostavljamo, da so posamični vplivi na druge sestavine vezani linearno in
c) pri računanju niso izločene skrajne vrednosti.

Po nakazani metodi pa vseeno dobimo razmeroma dobre rezultate tudi, če smer razvoja ni strogo eksponentna, ampak vsaj monotono naraščajoča ali padajoča. Linearna zveza posamičnih vplivov ni vselej v skladu z realnim stanjem, ker so ti vplivi dostikrat odvisni od ravni pojava. Skrajnih vplivov pri nakazani metodi ni mogoče izključiti, medtem ko jih pri metodah, ki jih obravnavamo v naslednjih odstavkih odstranimo z razmeroma enostavnim postopkom modificiranih sredin. Vpliv skrajnih odklonov pa je tem manjši, čim daljše je proučevano razdobje.

Metoda koeficientov dinamike

21.80 Če je smer razvoja v proučevanem razdobju eksponentna funkcija, razen metode popravljenih vsot za določanje sezonske sestavine uporabljamo tudi metodo koeficientov dinamike.

V tem primeru je model proučevane časovne vrste

$$Y_x = A \cdot b^x (1 + p_m + e_x) \qquad (21.84)$$

Vrsta koeficientov dinamike je

$$k_{|m} = k_x = Y_x / Y_{x-1} = Ab^x (1 + p_m + e_x) / Ab^{x-1} (1 + p_{m-1} + e_{x-1}) =$$

$$= b \cdot \frac{1 + p_m}{1 + p_{m-1}} + e'_x \qquad (21.85)$$

Od smeri razvoja ostane v vrsti $k_{|m}$ le parameter b, verižni kvocienti periodične sestavine in rezultat posamičnih vplivov e'_x.

Posamične vplive e'_x odstranimo z modificiranimi sredinami iz $k_{|m}$ tako, da iz vrste koeficientov dinamike za vsak mesec izločimo $k_{|m}$ za leto z najmanjšim in z največjim koeficientom dinamike, iz ostalih pa izračunamo poprečje. Tako dobimo

$$\bar{k}_m = \frac{1}{N-2} \sum_{|} k_{|m} = b \frac{1 + p_m}{1 + p_{m-1}} \qquad (21.86)$$

Z modificiranimi sredinami izločimo vpliv enkratnih izjemnih faktorjev, ki bi sicer pri navadnih sredinah pretirano vplivali na poprečje.

Iz lastnosti verižnih kvocientov sezonske sestavine sledi, da je geometrijska sredina iz poprečij $\bar{k}_m$ enaka parametru b, ker se v produktu zvrste isti sezonski indeksi v števcu in imenovalcu

$$G_{\bar{k}} = \sqrt[12]{\bar{k}_1, \bar{k}_2 \ldots\ldots \bar{k}_{12}} = b \qquad (21.87)$$

Če poprečne koeficiente dinamike $\bar{k}_m$ delimo z dobljenim parametrom b, dobimo

$$t_m = \bar{k}_m / b = \frac{1 + p_m}{1 + p_{m-1}} \qquad (21.88)$$

Iz vrste t_m dobimo s kumulativnim množenjem izraze

$$C_1 = 1 \cdot t_1 = 1 \cdot \frac{1+p_1}{1+p_{12}} = \frac{1+p_1}{1+p_{12}}$$

$$C_2 = C_1 \cdot t_2 = \frac{1+p_1}{1+p_{12}} \cdot \frac{1+p_2}{1+p_1} = \frac{1+p_2}{1+p_{12}} \qquad (21.89)$$

$$\vdots$$

$$C_m = C_{m-1} \cdot t_m = \frac{1+p_{m-1}}{1+p_{12}} \cdot \frac{1+p_m}{1+p_{m-1}} = \frac{1+p_m}{1+p_{12}}$$

$$\vdots$$

$$C_{12} = C_{11} \cdot t_{12} = \frac{1+p_{11}}{1+p_{12}} \cdot \frac{1+p_{12}}{1+p_{11}} = \frac{1+p_{12}}{1+p_{12}} = 1$$

Zaradi lastnosti sezonske sestavine ($\sum_m p_m = 0$) vrsto C_m popravimo s

$K = 1200 / \sum_m C_m$ in dobimo čiste sezonske indekse po obrazcu

$$P_m = 100\,(1+p_m) = K \cdot C_m \qquad (21.90)$$

21.81 Ker moremo nekatere stopnje pri računanju sezonske sestavine poenostaviti, če uporabljamo logaritme, po metodi koeficientov dinamike najenostavneje določimo sezonsko sestavino po tehle stopnjah:

a) Za proučevano časovno vrsto, ki sestoji iz mesečnih podatkov za več let, izračunamo vrsto koeficientov dinamike. (Za naš zgled o številu nočnin v SFRJ je ta vrsta prikazana v tabeli 21,27a).

b) Iz letnih vrst k_{lm} za posamezne mesece izločimo najmanjši in največji koeficient dinamike, da izločimo izjemne enkratne vplive. Iz ostalih členov izračunamo poprečne koeficiente dinamike $\bar{k}_m$ za posamezne mesece. (Za naš zgled so izračunani v zadnji vrsti tabele 21.27 a).

c) Za vrsto $\bar{k}_m$ poiščemo ustrezne vrednosti logaritmov $\log \bar{k}_m$ (stolpec 3 v tabeli 21,27b).

d) Iz vrste $\log\bar{k}_m$ izračunamo poprečje $\log b = \frac{1}{12}\sum_m \log \bar{k}_m$

e) Izračunamo $\log t_m = \log\bar{k}_m - \log b$ (21.91)

(v našem zgledu stolpec 4 v tabeli 21.27b)

f) Iz $\log t_m$ izračunamo kumulativno vrsto po obrazcu

$$\log C_m = \log C_{m-1} + \log t_m \qquad (21.92)$$

pri čemer je $\log C_{12} = 0$.

(Stolpec 5 v tabeli 21.27b)

g) $\log C_m$ antilogaritmiramo, da dobimo C_m (stolpec 6 v tabeli 21.27b)

h) Izračunamo popravni faktor $K = 1200/\sum_m C_m$. (V našem zgledu je $K = 1200/38,0382 = 31,5472$).

i) Po obrazcu $100\ (1 + p_m) = K \cdot C_m$ izračunamo sezonske indekse (stolpec 7 v tabeli 21.27b).

Tabela 21.27. Izračunanje sezonske komponente za število noči tev v SFRJ po metodi koeficientov dinamike

a) Koeficienti dinamike k_{Im}

	1960	1961	1962	1963	1964	1965	$\sum_I k_{Im}$	$\bar{k}_m$
J	-	0,971	0,964*	0,977	0,977	1,060*	2,925	0,975
F	0,985*	0,913	0,864	0,886	0,939	0,860*	3,602	0,901
M	1,060	1,058	1,021*	1,139*	1,090	1,119	4,327	1,082
A	1,214	1,135*	1,210	1,301	1,183	1,236*	4,908	1,227
M	1,594	1,667	1,531*	1,669	2,157*	1,720	6,650	1,663
J	1,993	1,850*	2,129	2,138*	1,901	2,068	8,091	2,023
J	2,469	2,451	2,358*	2,521*	2,380	2,388	9,688	2,422
A	1,194	1,203	1,209*	1,133*	1,170	1,161	4,728	1,182
S	0,403	0,409*	0,396	0,392	0,397	0,386*	1,588	0,397
O	0,448*	0,442	0,359	0,430	0,351*	0,362	0,593	0,398
N	0,714*	0,673	0,622	0,552*	0,690	0,633	2,618	0,655
D	0,833*	0,847	0,970*	0,944	0,892	0,903	3,586	0,897

b) Izračunanje sezonske komponente

	$\bar{k}_m$	$\log \bar{k}_m$	$\log t_m$	$\log C_m$	C_m	P_m
1	2	3	4	5	6	7
J	0,975	0,9890-1	0,9875-1	0,9875-1	0,9717	30,7
F	0,901	0,9547-1	0,9532-1	0,9407-1	0,8724	27,5
M	1,082	0,0342	0,0327	0,9734-1	0,9406	29,7
A	1,227	0,0888	0,0873	0,0607	1,1500	36,3
M	1,663	0,2209	0,2194	0,2801	1,9060	60,1
J	2,023	0,3060	0,3045	0,5846	3,8420	121,2
J	2,422	0,3842	0,3827	0,9643	9,2750	292,6
A	1,182	0,0726	0,0711	1,0384	10,9249	344,7
S	0,397	0,5988-1	0,5973-1	0,6357	4,3220	136,4
O	0,398	0,5999-1	0,5984-1	0,2341	1,7144	54,1
N	0,655	0,8162-1	0,8147-1	0,0488	1,1190	35,3
D	0,897	0,9528-1	0,9513-1	0,0001	1,0002	31,5
		$\sum \log \bar{k}_m = 0,0181$	0,0001		38,0382	1 200,1

Metoda kvocientov na vrsto drsečih sredin

21.82 Če smer razvoja v proučevanem razdobju ni eksponentna funkcija ali vsaj monotono naraščajoča ali monotono padajoča, metoda popravljenih vsot in metoda koeficientov dinamike nista uporabni. Če je trend T poljuben, v časovni vrsti pa ni cikličnih vplivov, iz osnovne časovne vrste z modelom:

$$Y_x = T_x (1 + p_m + e_x) \qquad (21.93)$$

odpravimo vpliv trenda, če posamezne člene v časovni vrsti delimo z ustreznimi vrednostmi trenda. Kvocienti

$$K_x = Y_x / T_x = A (1 + p_m + e_x) \qquad (21.94)$$

ustrezajo modelu 21.72 in moremo iz vrste kvocientov K_x določiti sezonsko sestavino po metodi vsot, ki smo jo opisali v odstavku 21.75.

Vendar pogosto zasledimo v časovni vrsti razen trenda tudi ciklično sestavino.

Z vrsto dvanajstmesečnih drsečih sredin Y_x iz osnovne časovne vrste izločimo periodične, enkratne in slučajne vplive. Zato moremo vrsto drsečih sredin Y_x vzeti za uporaben približek skupnega učinka trenda in ciklične sestavine $Y = TC$.

Vzemimo splošen model časovne vrste

$$Y_x = T_x \cdot C_x (1 + p_m + e_x) \qquad (21.95)$$

Če upoštevamo pomen vrste dvanajstmesečnih drsečih sredin Y_x, kvocienti

$$K_{lm} = Y_x / \tilde{Y}_x = A \cdot (1 + p_m + e_{lm}) \qquad (21.96)$$

na vrsto drsečih sredin vsebujejo samo periodično sestavino in slučajnostne vplive in ustrezajo modelu 21.72. Iz kvocientov na drseče sredine izračunamo dalje sezonske indekse po metodi vsot. Metoda kvocientov na vrsto drsečih sredin je seveda

uporabna tudi v primeru, če v časovni vrsti ni ciklične sestavine. Zato je metoda kvocientov na drseče sredine splošno uporabna metoda za določanje sezonskih indeksov. Hiba je le v tem, da more vrsta dvanajstmesečnih drsečih sredin podajati deformirano TC. Razen tega pa se proučevano razdobje skrči za eno periodo.

21.83 Postopek določanja sezonskih indeksov po metodi kvocientov na vrsto drsečih sredin izvedemo po gornjem v tehle stopnjah:

a) Za proučevano časovno vrsto Y_{lm} izračunamo vrsto dvanajstmesečnih sredin $\tilde{Y}_{lm}$.

b) Izračunamo vrsto kvocientov $K_{lm} = Y_{lm} / \tilde{Y}_{lm}$ osnovnih podatkov Y_{lm} z ustreznimi vrednostmi vrste drsečih sredin.

c) v vrstah kvocientov K_{lm} za posamezne mesece izračunamo modificirana poprečja. Najprej izločimo za vsak m oba skrajna kvocienta, iz drugih pa izračunamo poprečje $\bar{K}_m$.

d) Iz $\bar{K}_m$ izračunamo popravni faktor $c = 1200 / \sum_m \bar{K}_m$.

e) Sezonske indekse izračunamo po obrazcu

$$P_m = 100 \cdot (1 + p_m) = c \cdot \bar{K}_m$$

21.84 Za zgled vzemimo časovno vrsto nočitev v SFRJ, za katero smo že izračunali sezonske indekse po metodi popravljenih vsot in metodi koeficientov dinamike.

Od šestih let imamo za vsak mesec le pet kvocientov K_{lm}, iz katerih so izračunane modificirane sredine. Z * sta za vsak mesec zaznamovana najmanjši in največji koeficient, ki ju pri računanju sredine izpustimo in izračunamo aritmetično sredino iz treh koeficientov

Popravni faktor c je v našem zgledu

$$c = 1200/12,000 = 100,0$$

Točno enak 100,0. Popravni faktor c po svoji sestavi odstopa od 100,0 le zaradi slučajnostnih vplivov.

Tabela 21.28. Izračunanje sezonskih indeksov za nočitve v SFRJ v razdobju 1960–1965 po metodi koeficientov na vrsto drsečih sredin

		J	F	M	A	M	J	J	A	S	O	N	D
1960	Y_{lm}	549	541	574	697	1111	2214	5467	6526	2627	1176	840	700
	$\tilde{Y}_{lm}$							1924	1933	1940	1945	1953	1962
	K_{lm}							2,841	3,376	1,354	0,605	0,430*	0,357*
1961	Y_{lm}	680	621	657	746	1243	2299	5635	6779	2773	1225	825	699
	$\tilde{Y}_{lm}$	1972	1990	2007	2015	2016	2015	2015	2013	2009	2005	1998	1994
	K_{lm}	2,345*	0,312*	0,327*	0,370*	0,617	1,141*	2,796*	3,367*	1,380	0,611*	0,413	0,351
1962	Y_{lm}	674	582	594	719	1101	2344	5528	6683	2644	950	591	573
	$\tilde{Y}_{lm}$	1992	1983	1974	1957	1936	1921	1911	1902	1898	1897	1903	1920
	K_{lm}	0,338	0,293	0,301	0,367	0,569*	1,220	2,893	3,513*	1,393*	0,501	0,311*	0,298
1963	Y_{lm}	560	496	565	735	1227	2623	6612	7491	2939	1264	698	659
	$\tilde{Y}_{lm}$	1977	2055	2101	2127	2144	2152	2159	2168	2176	2182	2203	2246
	K_{lm}	0,283	0,241*	0,269	0,346	0,572	1,219	3,062*	3,456	1,351*	0,579	0,317	0,293*
1964	Y_{lm}	644	605	660	781	1685	3204	7625	8921	3544	1245	859	766
	$\tilde{Y}_{lm}$	2313	2414	2499	2524	2530	2541	2552	2563	2573	2585	2591	2600
	K_{lm}	0,278*	0,251	0,264*	0,309*	0,666*	1,261	2,988	3,481	1,378	0,482*	0,332	0,295
1965	Y_{lm}	812	698	781	965	1659	3431	8194	9512	3671	1330	842	760
	$\tilde{Y}_{lm}$	2633	2681	2711	2720	2723	2722						
	K_{lm}	0,308	0,260	0,288	0,355	0,609	1,261*						
	$\sum K_{lm}$	0,929	0,804	0,858	1,068	1,798	3,700	8,722	10,313	4,112	1,685	1,062	0,944
	$\bar{K}_m$	0,310	0,268	0,286	0,356	0,599	1,233	2,907	3,438	1,371	0,562	0,354	0,315
	P_m	31	27	29	36	60	123	291	344	137	56	35	31

21.85 Po treh metodah: metodi popravljenih vsot, metodi koeficientov dinamike in metodi kvocientov na vrsto drsečih sredin smo namenoma izračunali sezonske indekse za isti pojav: število nočitev v SFRJ v razdobju 1960-1965. Če primerjamo po treh nakazanih metodah izračunane vrste sezonskih indeksov (tabela 21.29), vidimo, da razlike niso znatne oziroma so nepomembne. Razlike so majhne predvsem zato, ker časovna vrsta vsaj v grobem zadošča pogoju, da je monotono naraščajoča in moremo osnovno smer razvoja vzeti kot eksponentno funkcijo.

Tabela 21.29. Sezonski indeksi za število nočitev v SFRJ, izračunani po različnih metodah:

Mesec	Metoda popravljenih vsot	metoda koeficientov dinamike	metoda kvocientov na drseče sredine
J	31	31	31
F	28	28	27
M	30	30	29
A	36	36	36
M	61	60	60
J	122	121	123
J	295	293	291
A	344	345	344
S	136	136	137
O	53	54	56
N	34	35	35
D	30	31	31
	1200	1200	1200

Katera izmed obravnavanih metod pride v poštev v konkretnih primerih, je odvisno od proučevanega pojava oziroma časovne vrste. Če analiza nakaže, da na časovno vrsto niso vplivali neki izraziti vplivi, smer razvoja v proučevanem razdobju pa je naraščajoča ali padajoča, je primerna metoda popravljenih vsot, ki je izmed njih računsko najenostavnejša. Če je smer razvoja eksponentna in zasledimo močne en-

kratne vplive, uporabimo metodo koeficientov dinamike, pri poljubni smeri razvoja pa je najprimernejša metoda kvocientov na drseče sredine s pridržki, ki smo jih že nakazali.

21.86 Na zgledu o nočninah v SRS, SFRJ in nočnin tujih gostov bomo nazorno pokazali razlike v sezonski sestavini, ki so pogojene z vsebinskimi razlikami v turizmu v SRS v primerjavi s SFRJ in turizmom tujih gostov v primerjavi z vsemi turisti v SFRJ.

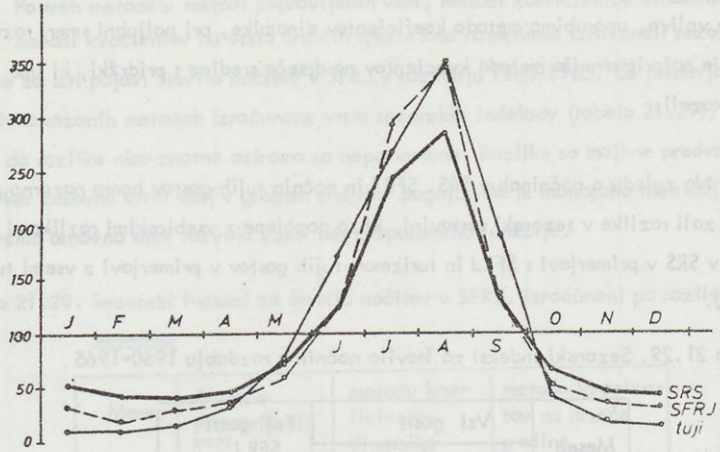
Tabela 21.29. Sezonski indeksi za število nočnin v razdobju 1960-1965

Mesec	Vsi gosti		Tuji gosti SFRJ
	SFRJ	SRS	
J	31	52	10
F	28	43	10
M	30	41	14
A	36	47	33
M	60	73	76
J	121	124	172
J	293	245	268
A	345	288	353
S	136	132	192
O	54	67	43
N	35	45	16
D	31	43	13

Iz tabele 21.29 in slike 21.21 so vidne izrazite razlike. Glede na pogoje letnega turizma je sezonska sestavina za SFRJ izrazitejša od SRS, ki ima tudi močno zimsko sezono. Še posebej je to očitno za sezonsko sestavino za tuje goste, ker je tuj turizem še bolj vezan na letno sezono kot domači.

Kot sintetični opisni pokazovalec jakosti periodičnih vplivov uporabljamo poprečni absolutni odklon od nivoja 100.

$$AD_p = \frac{1}{12} \sum |P_m - 100| = \frac{1}{6} \sum (P_m^+ - 100) \qquad (21.97)$$



Slika 21.21. Sezonski indeksi o številu nočnin v razdobju 1960-1965

Poprečni absolutni odklon od poprečnega sezonskega indeksa 100 izračunamo enostavno tako, da seštejemo sezonske indekse, ki so večji od 100, ki jih zmanjšamo za 100 in vsoto delimo s 6.

Za sezonsko sestavino vseh turistov v SFRJ je

$$AD_P = \frac{21 + 193 + 245 + 36}{6} = \frac{495}{6} = 82,5$$

Do enakega rezultata pridemo tudi po osnovnem obrazcu

$$AD_P = \frac{|31-100| + |28-100| + \ldots + |35-100| + |31-100|}{12} = 82,5$$

vendar po daljši poti.

Za vse turiste v SRS je $AD_P = 64,8$, za tuje turiste v SFRJ pa je $AD_P = 97,5$ kar je v skladu z zakonitostjo obravnavanih pojavov.

Dinamična sezonska komponenta

21.87 Pri dosedanjih metodah in zgledih smo predpostavljali, da je sezonska sestavina statična in da so za določen mesec razlike med kvocienti dejanskih podatkov na ustrezno drsečo sredino rezultat slučajnostnih in enkratnih vplivov. Iz

vzrokov, ki smo jih že navedli, pa pri dinamični sezonski sestavini kvocienti na vrsto drsečih sredin kažejo za posamezen mesec določeno smer spreminjanja, sezonska sestavina ni neodvisna od časa, temveč se z leti spreminja. Če predvidevamo, da je razvoj sezonske sestavine v proučevanem razdobju opisan s parabolo druge stopnje, moremo tak dinamičen sezonski indeks opisati s parabolo druge stopnje

$$1 + p_{lm} = 1 + p_{m.o} + p_{m.1} X_l^1 + p_{m.2} X_l^2 \qquad (21.98)$$

pri čemer so $p_{m,o}$ sezonski koeficienti, ki nakazujejo poprečno statično sestavino, $p_{m,1}$ sezonski koeficienti, ki nakazujejo linearne spremembe v sezonski sestavini in $p_{m,2}$ sezonski koeficienti, ki nakazujejo kvadratične spremembe v sezonski sestavini. Zaradi lastnosti sezonskih indeksov velja:

$$\sum_m p_{m.o} = 0; \quad \sum_m p_{m.1} = 0; \quad \sum_m p_{m.2} = 0 \qquad (21.99)$$

Model časovne vrste z dinamično sezonsko sestavino druge stopnje je torej takle

$$Y_{lm} = T_{lm} \cdot C_{lm} (1 + p_{m.o} + p_{m.1} X_l^1 + p_{m.2} X_l^2 + e_{lm}) \qquad (21.100)$$

Dinamično sezonsko sestavino proučimo v tehle fazah:

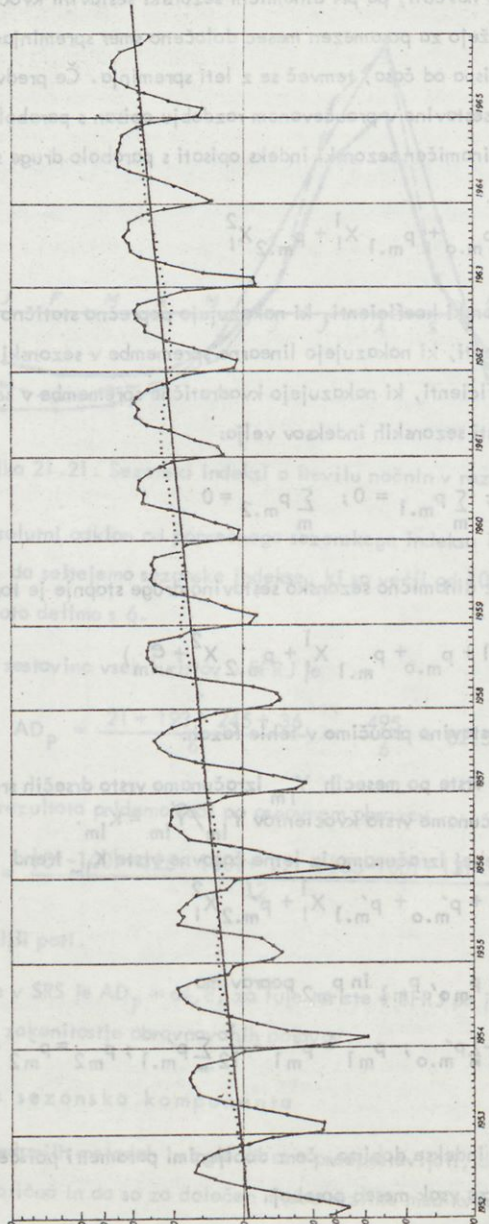
1. Iz osnovne časovne vrste po mesecih Y_{lm} izračunamo vrsto drsečih sredin $\tilde{Y}_{lm}$.
2. Iz Y_{lm} in $\tilde{Y}_{lm}$ izračunamo vrsto kvocientov $Y_{lm}/\tilde{Y}_{lm} = K_{lm}$.
3. Za vsak mesec posebej izračunamo iz letne časovne vrste K_{lm} trend

$$K_{lm} = 1 + p'_{m.o} + p'_{m.1} X_l^1 + p'_{m.2} X_l^2 \qquad (21.101)$$

4. Dobljene parametre p_{mo}, p_{m1} in p_{m2} popravimo

$$p_{m.o} = p'_{m.o} - \frac{1}{12}\sum_m p'_{m.o}; \quad p_{m1} = p'_{m1} - \frac{1}{12}\sum_m p'_{m.1}; \quad p_{m2} = p'_{m2} - \frac{1}{12}\sum_m p'_{m2} \qquad (21.102)$$

5. Izravnane sezonske indekse dobimo, če z dobljenimi parametri poiščemo vrste sezonskih indeksov za vsak mesec posebej.

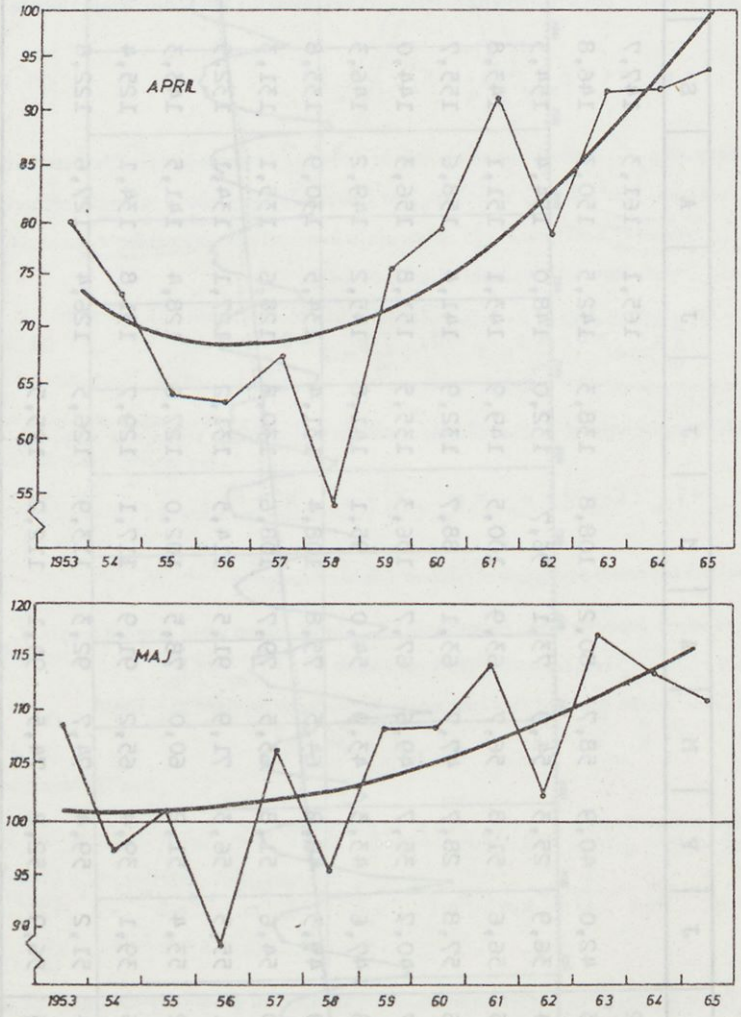


Slika 21.22. Indeks industrije gradbenega materiala v SFRJ v razdobju 1952–1965

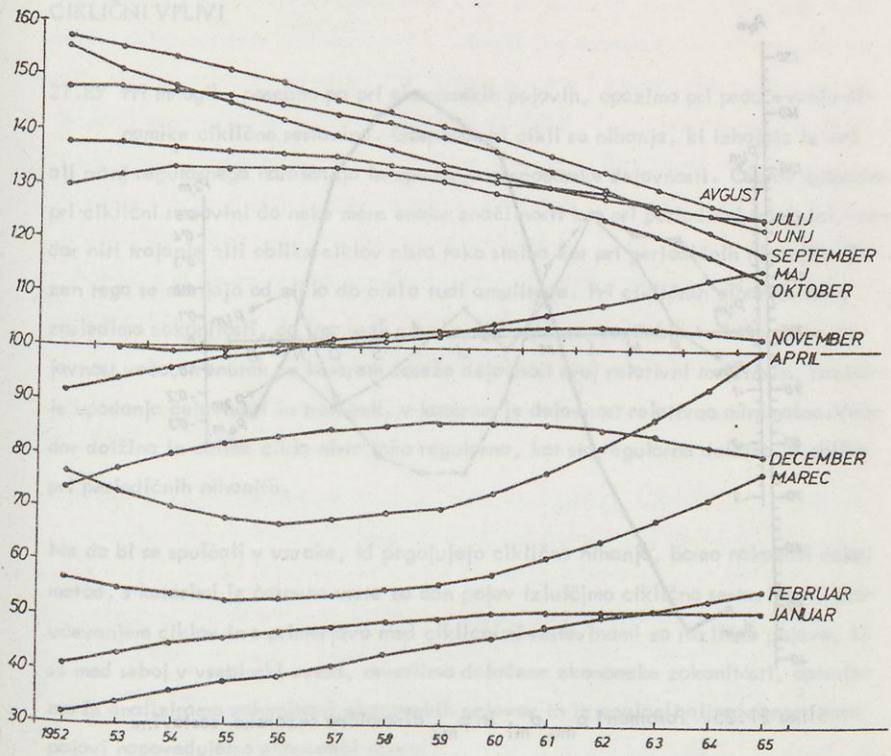
Tabela 21.30. Kvocienti na drseče sredine indeksov gradbene industrije SFRJ 1952-1965

	J	F	M	A	M	J	J	A	S	O	N	D
1952							165,1	161,3	147,7	126,3	90,6	69,8
1953	42,0	40,9	58,7	80,2	108,8	138,3	142,5	150,3	146,8	138,2	100,2	87,4
1954	36,9	25,3	54,9	73,1	96,7	132,0	148,0	154,4	154,3	129,2	94,3	87,3
1955	56,6	51,8	56,7	63,9	100,5	149,9	143,1	151,1	143,8	130,0	95,8	88,7
1956	57,8	28,7	47,7	63,1	88,7	132,9	141,4	158,6	155,7	153,2	103,4	84,6
1957	40,7	35,7	49,6	67,7	106,3	135,5	151,8	156,3	144,0	132,7	106,5	84,2
1958	47,6	43,3	43,9	54,0	95,1	141,8	145,2	149,2	146,3	140,7	101,8	89,4
1959	49,3	46,8	64,5	75,8	108,4	131,4	134,5	130,9	133,8	127,9	101,3	91,7
1960	54,6	51,8	65,5	79,7	108,6	130,8	128,6	135,1	131,3	119,1	105,0	91,6
1961	55,3	56,3	71,9	91,5	114,5	131,2	127,1	134,1	132,5	120,8	101,8	87,2
1962	53,4	51,2	60,0	78,5	102,0	127,6	128,4	141,5	143,3	133,5	106,3	81,9
1963	39,1	39,4	65,2	91,9	117,1	129,7	131,8	134,1	125,4	122,9	105,5	84,7
1964	51,2	59,4	74,7	92,3	113,9	126,5	126,4	127,6	122,8	120,5	103,5	91,0
1965	56,9	52,8	74,5	94,1	112,2	125,5						

21.88 Za zgled vzemimo časovno vrsto indeksov za industrijo gradbenega materiala v SFRJ v razdobju 1952-1965. Proučevano razdobje obsega štirinajst let. Glede na izrazito razvojno dinamiko proučevanega pojava in razmeroma dolgo časovno razdobje moremo predvidevati, da sezonska komponenta ni statična, temveč dinamična.



Slika 21.23. Vrste kvocientov na trend z vrisanimi trendi po mesecih za indeks industrije gradbenega materiala v letih 1952-1965 za meseca april in maj



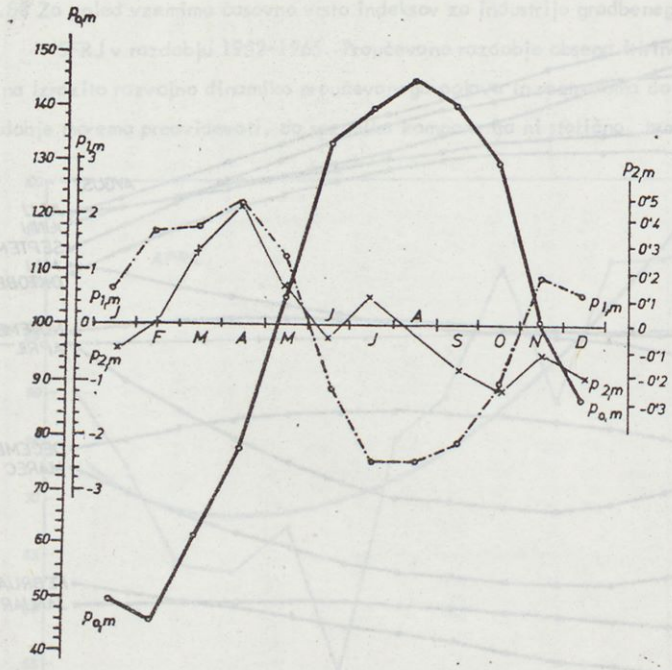
Slika 21.24. Trendi za dinamično sezonsko sestavino za indeks industrije gradbenega materiala v letih 1952-1965

V sliki 21.23 so za posamezne mesece vrisani po letih kvocienti na vrsto drsečih sredin in trend sezonske komponente. V sliki 21.23 so na enem grafikonu narisani trendi za sezonsko komponento za vse mesece v letu.

Slika 21.25 pa v sintetični obliki kaže spreminjanje parametrov sezonske sestavine po mesecih. Čeprav že slike za posamezne mesece kažejo določeno tendenco manjšanja sezonske sestavine, to še bolj nazorno pokaže slika trendov za dinamično sestavino.

Opazna je splošna tendenca zmanjševanja sezonske sestavine, ker imajo vsi bolj ali manj izrazito smer k ravni 100.

Na poseben način je to vidno iz grafikona 21.25.



Slika 21.25. Parametri p_{mo}, p_{m1} in p_{m2} dinamične sezonske sestavine

Če primerjamo statično sestavino p_{mo} z linearnim delom dinamične p_{m1}, vidimo med njima negativno povezanost. Negativnemu p_{mo} ustreza pozitiven p_{m1} in obratno, pozitivnemu p_{mo} negativen p_{m1}. To pomeni, da se v razvoju negativna sezonska sestavina linearno veča, pozitivna pa pada, kar je v skladu s tem, da trendi težijo k 100. Med linearno sestavino p_{m1} in kvadratično sestavino p_{m2} pa je opaziti, čeprav ne tako izrazito pozitivno korelacijo, kar kaže na to, da so naraščajoči trendi konkavni, padajoči pa konveksni. Iz tega sledi vsebinski sklep, da je dinamika spremembe sezonske sestavine v drugem delu proučevanega razdobja hitrejša kot v prvem.

21.89 Pri mnogih, posebno pa pri ekonomskih pojavih, opazimo pri proučevanju dinamike ciklično sestavino. Gospodarski cikli so nihanja, ki izhajajo iz več ali manj regularnega naraščanja in upadanja gospodarske dejavnosti. Čeprav opazimo pri ciklični sestavini do neke mere enake značilnosti kot pri periodični sestavini, vendar niti trajanje niti oblike ciklov nista tako stalna kot pri periodičnih nihanjih. Razen tega se menjajo od cikla do cikla tudi amplitude. Pri cikličnih nihanjih torej zasledimo zakonitosti, da ima vsak cikel svojo dolžino, razdobje, v katerem se dejavnost veča, trenutek, v katerem doseže dejavnost svoj relativni maksimum, razdobje upadanja dejavnosti in trenutek, v katerem je dejavnost relativno minimalna. Vendar dolžina in oblika cikla nista tako regularna, kot sta regularna dolžina in oblika pri periodičnih nihanjih.

Ne da bi se spuščali v vzroke, ki pogojujejo ciklična nihanja, bomo nakazali nekaj metod, s katerimi iz časovne vrste za dan pojav izluščimo ciklično sestavino. S proučevanjem ciklov in s primerjavo med cikličnimi sestavinami za različne pojave, ki so med seboj v vsebinski zvezi, osvetlimo določene ekonomske zakonitosti, opisujemo in analiziramo zakonitosti ekonomskih pojavov in iz analogij ali odvisnosti med pojavi napovedujemo ekonomski razvoj.

Preskus o značilnosti ciklične sestavine

21.90 Pri proučevanju cikličnih nihanj je včasih umestno, da pred uporabo določenih računskih metod preskusimo, ali časovna vrsta, ki prikazuje določen pojav, kaže značilnosti, ki so tipične za ciklična nihanja. Ena izmed metod za preskušanje značilnosti ciklične komponente, je neparametrična metoda sekvenc, s katero enostavno in hitro preskusimo, ali je v pojavu ciklična sestavina značilna ali ne.

Vzemimo, da ima trend za proučevano časovno vrsto na celotnem razmiku isto težnjo naraščanja ali upadanja in da na pojav vplivajo razen trenda samo še slučajnostni oziroma iregularni vplivi. V časovni vrsti, ki prikazuje tak pojav, razlike vrednosti

dveh zaporednih členov od člena do člena spreminjajo svoj predznak slučajnostno po zakonitostih slučajnostnega pojavljanja. V tem primeru ni verjetna daljša sekvenca stanovitega naraščanja ali padanja pojava, niti stanovito zakonito padanje ali naraščanje. Število sekvenc, to je število razdobij, v katerih pojav oziroma časovna vrsta konstantno narašča ali pada, torej v primeru slučajnostnega pojavljanja ni niti zelo majhno niti zelo veliko. V primeru, če so v dinamiki pojava vzroki, ki pogojujejo ciklična gibanja, pa, je število sekvenc bodisi izredno majhno ali izjemno veliko. Iz tega se ponuja rešitev, da s testom sekvenc preskusimo ali je v časovni dinamiki nekega pojava značilna ciklična sestavina ali ne. V primeru preskusa ciklične sestavine uporabimo test sekvenc po temle postopku:

S + oziroma - nakažemo v časovni vrsti od člena do člena ali je pojav narastel ali padel.

Ker proučujemo ciklična gibanja le v daljših časovnih vrstah, za katere je $N > 10$, uporabimo zakonitost oziroma približek, da se izraz

$$z = \frac{r - M_r}{SD_r} \qquad (21.103)$$

porazdeljuje v standardizirani normalni porazdelitvi. Ciklična sestavina je značilna na stopnji $\alpha = 0,05$, če je $1,96 < |z| < 2,58$, na stopnji $\alpha = 0,01$, če je $2,58 < |z| < 3,29$ in značilna na stopnji $\alpha = 0,001$, če je $|z| < 3,29$.

$$M_r = \frac{2\,n^+ \cdot n^-}{n^+ + n^-} + 1 \; ; \quad SD_r = \sqrt{\frac{(M_r - 1)(M_r - 2)}{n^+ + n^- - 1}} \qquad (21.104)$$

Pri tem pomeni n^+ = število dvigov, n^- = število padcev, r = število sekvenc.

21.91 Za zgled vzemimo angleško industrijsko proizvodnjo v razdobju 1818-1913. Za to razdobje ugotovimo $n^+ = 57$ letnih dvigov in $n^- = 38$ letnih padcev, medtem ko je število sekvenc $r = 30$. Iz teh podatkov dobimo:

$$M_r = \frac{2.57.38}{57+38} + 1 = 46,6$$

$$SD_r = \sqrt{\frac{(46,6-1) \cdot (46,6-2)}{57+38-1}} = 4,65$$

$$z = \frac{r - M_r}{SD_r} = \frac{30 - 46,6}{4,65} = -3,57$$

Ker je $z = -3,57$ absolutno večji kot je kritična vrednost $z_{0,001} = 3,29$, sklepamo, da je ciklična sestavina v proučeni časovni vrsti zelo značilna.

Merjenje ciklov

21.92 Če iz osnovne časovne vrste izločimo trend, periodične oziroma sezonske vplive in iregularne oziroma slučajnostne vplive, ostane čista ciklična sestavina. Ta zasnova je osnova rezidualne metode določanja ciklične sestavine. Jasno je, da ni v vseh časovnih vrstah periodične oziroma sezonske komponente, bodisi zaradi narave pojava, ki ga proučujemo, bodisi zaradi časovne vrste same. V vrsti letnih podatkov namreč ne zasledimo sezonskih variacij, kljub temu, da je pojav sezonskega značaja.

Razen rezidualne metode, po kateri dobimo časovno vrsto ciklov v določenem razdobju, proučujemo ciklična gibanja tudi sumarno v obliki poprečnega cikla, katerega analiziramo po nekaterih značilnih obdobjih in točkah in nekaterih značilnostih kot so poprečna dolžina cikla, poprečna oblika itd.

Rezidualna metoda za določanje ciklične sestavine

Izločitev trenda in sezonskih vplivov

21.93 Pri rezidualni metodi iz časovne vrste postopoma izločimo trend, sezonsko sestavino in iregularne oz. slučajnostne vplive.

Če predpostavimo, da je model časovne vrste

$$Y = T \cdot P \cdot C \cdot S. \qquad (21.105)$$

moremo izločiti trend in sezonske oziroma periodične vplive na tri načine, ki dajo končno enak rezultat:

Po prvem načinu iz osnovne časovne vrste najprej izločimo trend

$$Y/T = P \cdot C \cdot S \tag{21.106}$$

iz dobljene vrste pa izločimo sezonsko sestavino tako, da kvociente na trend delimo s sezonskimi indeksi

$$(Y/T)/P = C \cdot S \tag{21.107}$$

Drug možen način je, da iz osnovne časovne vrste najprej izločimo sezonske vplive. Osnovno časovno vrsto delimo s sezonskimi indeksi

$$Y/P = T \cdot C \cdot S \tag{21.108}$$

dobljeno časovno vrsto, katero smo očistili sezonskih vplivov, pa naknadno delimo s trendom, da odstranimo trend

$$(Y/P)/T = C \cdot S \tag{21.109}$$

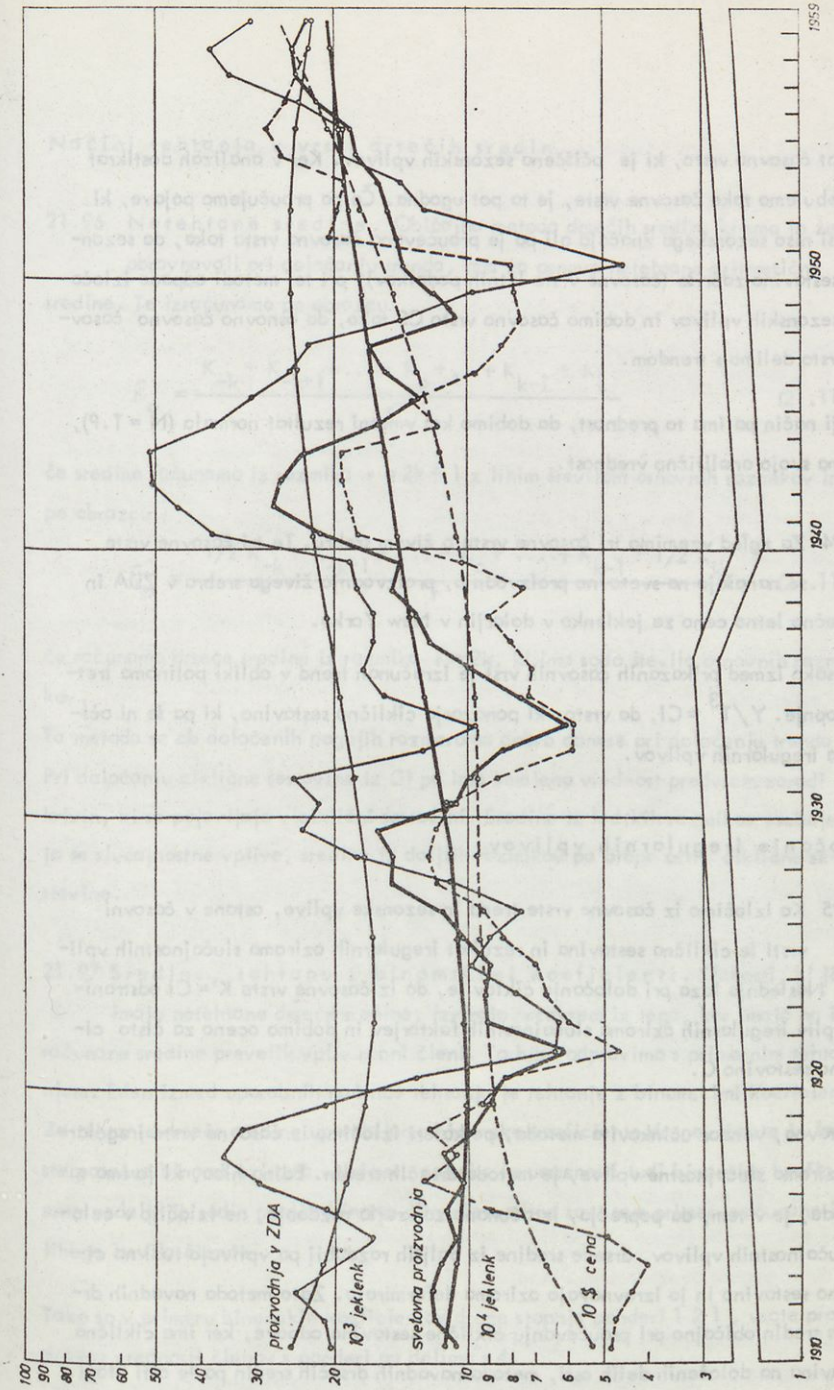
Po tretji metodi pa najprej izračunamo tako imenovano normalo, ki je skupni učinek trenda in sezonskih vplivov

$$N = T \cdot P \tag{21.110}$$

in jo dobimo, če vrednosti trenda pomnožimo s sezonskimi indeksi. Če osnovno časovno vrsto delimo z normalo, dobimo

$$Y/N = Y/T \cdot P = C \cdot S \tag{21.111}$$

Prvo metodo redko uporabljamo, ker vmesni rezultat $P \cdot C \cdot S$ nima posebnega vsebinskega pomena. Druga metoda je zelo uporabna, ker dobimo kot vmesni re-



Slika 21.26. Svetovna proizvodnja in proizvodnja živega srebra v ZDA in cena za jeklenko živega srebra v New Yorku v razdobju 1910–1959

Binomski koeficienti četrte stopnje so 1 4 6 4 1 , vsoto produktov pa delimo s 16.

V primeru binomskih koeficientov šeste stopnje pa so ponderi

1 6 15 20 15 6 1 ,

vsoto produktov pa delimo s 64.

Ustrezne sredine so centrirane na srednji člen.

V splošnem da vrsta drsečih sredin, ki so ponderirane z binomskimi koeficienti, boljše rezultate, ker se tudi pri daljših razmikih boljše prilegajo ciklu in ne pačijo cikla tako močno kakor enostavne netehtane sredine.

Časovno vrsto drsečih sredin, ki so tehtane z binomskimi koeficienti, dobimo enostavno iz vrst vsot sosednjih členov, če postopek seštevanja sosednjih vrednosti večkrat ponovimo.

21.98 Metoda drsečih lokov. Da čim bolje izločimo iregularne vplive brez izravnavanja amplitud v ciklični komponenti, razen binomskih koeficientov uporabljamo tudi druge načine tehtanja. Najobičajnejši je sistem drsečih lokov. Po tej metodi skozi več točk oziroma vrednosti po metodi najmanjših kvadratov prilagodimo parabolo določene stopnje, od vseh vrednosti prilagojene parabole pa uporabimo samo vrednost v sredini razmika, v katerem izvedemo prilagoditev. Tako za vsak člen v časovni vrsti (razen ustreznemu številu robnih členov) izračunamo vrednost, ki predstavlja po tej predpostavki vrednost cikla. Ker za vsak člen prilagodimo vrednost z lokom neke krivulje, imenujemo to metodo metodo drsečih lokov.

Tehnično dobimo srednjemu členu ustrezno vrednost na paraboli določene stopnje, če iz zaporednih členov z ustreznimi ponderi izračunamo tehtano aritmetično sredino, ki jo pripišemo srednjemu členu.

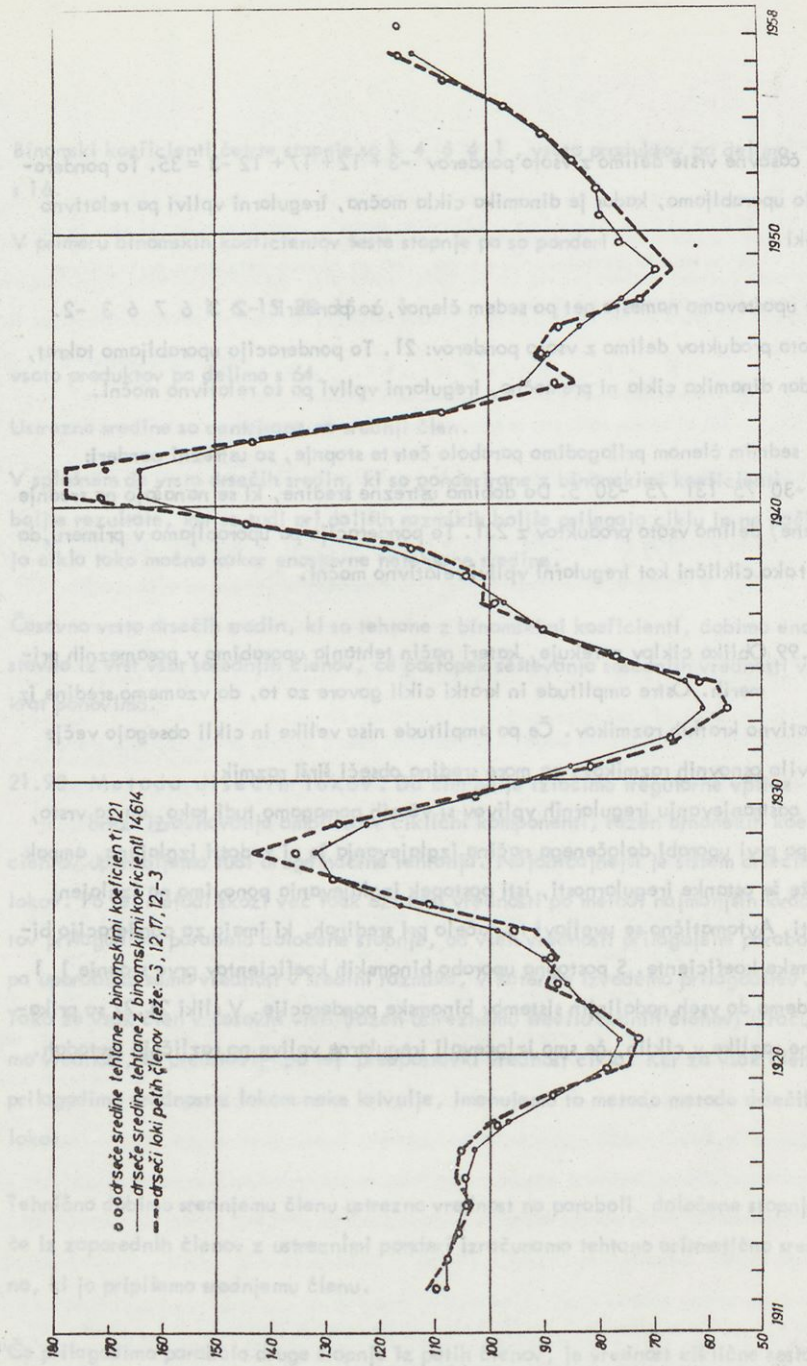
Če prilagodimo parabolo druge stopnje iz petih členov, je vrednost ciklične sestavine za srednji člen tehtana aritmetična sredina iz petih zaporednih členov s ponderi: -3 12 17 12 -3. Vsoto produktov ponderov z ustreznimi zaporednimi vrednost-

mi časovne vrste delimo z vsoto ponderov -3 + 12 + 17 + 12 -3 = 35. To ponderacijo uporabljamo, kadar je dinamika cikla močna, iregularni vplivi pa relativno šibki.

Če upoštevamo namesto pet po sedem členov, so ponderi: -2 3 6 7 6 3 -2. Vsoto produktov delimo z vsoto ponderov: 21. To ponderacijo uporabljamo takrat, kadar dinamika cikla ni premočna, iregularni vplivi pa so relativno močni.

Če sedmim členom prilagodimo parabolo četrte stopnje, so ustrezni ponderi: 5 -30 75 131 75 -30 5. Da dobimo ustrezne sredine, ki se nanašajo na srednje člene, delimo vsoto produktov z 231. To ponderacijo pa uporabljamo v primeru, da so tako ciklični kot iregularni vplivi relativno močni.

21.99 Oblika ciklov narekuje, kateri način tehtanja uporabimo v posameznih primerih. Ostre amplitude in kratki cikli govore za to, da vzamemo sredine iz relativno kratkih razmikov. Če pa amplitude niso velike in cikli obsegajo večje število osnovnih razmikov, pa more sredina obseči širši razmik.
Pri odstranjevanju iregularnih vplivov si včasih pomagamo tudi tako, da za vrsto, ki po prvi uporabi določenega načina izglajevanja še ni zadosti izglajena, ampak kaže še ostanke iregularnosti, isti postopek izglajevanja ponovimo na izglajeni vrsti. Avtomatično se uveljavi to načelo pri sredinah, ki imajo za ponderacijo binomske koeficiente. S postopno uporabo binomskih koeficientov prve stopnje 1 1 pridemo do vseh nadaljnjih sistemov binomske ponderacije. V sliki 21.27 so prikazane razlike v ciklih, če smo izločevali iregularne vplive po različnih metodah.



Slika 21.27. Indeksi ciklične sestavine za svetovno proizvodnjo živega srebra, določeni po različnih metodah.

21.100 Splet faktorjev, ki imajo za rezultat ciklična nihanja, ne vpliva samo na en pojav, ampak na niz ekonomskih pojavov, ki so med seboj povezani hkrati. Zato ima posebno vlogo pri analiziranju ekonomskih pojavov primerjava cikličnih sestavin za več pojavov, ki so v medsebojni povezanosti oziroma odvisnosti. Tako isti splet faktorjev vpliva istočasno na proizvodnjo najrazličnejših izdelkov, na cene teh izdelkov, na izvoz in uvoz surovin in končnih izdelkov itd.

Ciklična sestavina za posamezen pojav je dana v časovni vrsti indeksov cikla, ki pokažejo, kako se v določenem času normala (trend + sezonski vplivi) poveča zaradi cikličnih vplivov. Ker je ciklična sestavina za določen pojav prikazana z indeksi, moremo formalno primerjati cikle za različne pojave, ne glede na enote mere za osnovne podatke (količinske, vrednostne).

Primerljivost ciklov za različne pojave pa moti predvsem to, da je učinek faktorjev, ki povzročajo cikle, na nekatere pojave močnejši, na druge pa šibkejši. To se pokaže na različnih amplitudah ciklov za različne pojave. Različno občutljivost posameznih pojavov na ciklične faktorje odpravimo s standardizacijo ciklične sestavine.

Standardizacijo cikla izvedemo po tehle točkah:

a) iz indeksov ciklične sestavine C za posamezen pojav izračunamo poprečje $\bar{C}$ in standardni odklon SD_C,

b) iz $\bar{C}$, SD_C in indeksov ciklične sestavine izračunamo standardizirane z_C-odklone

$$z_C = \frac{C - \bar{C}}{SD_C} \qquad (21.114)$$

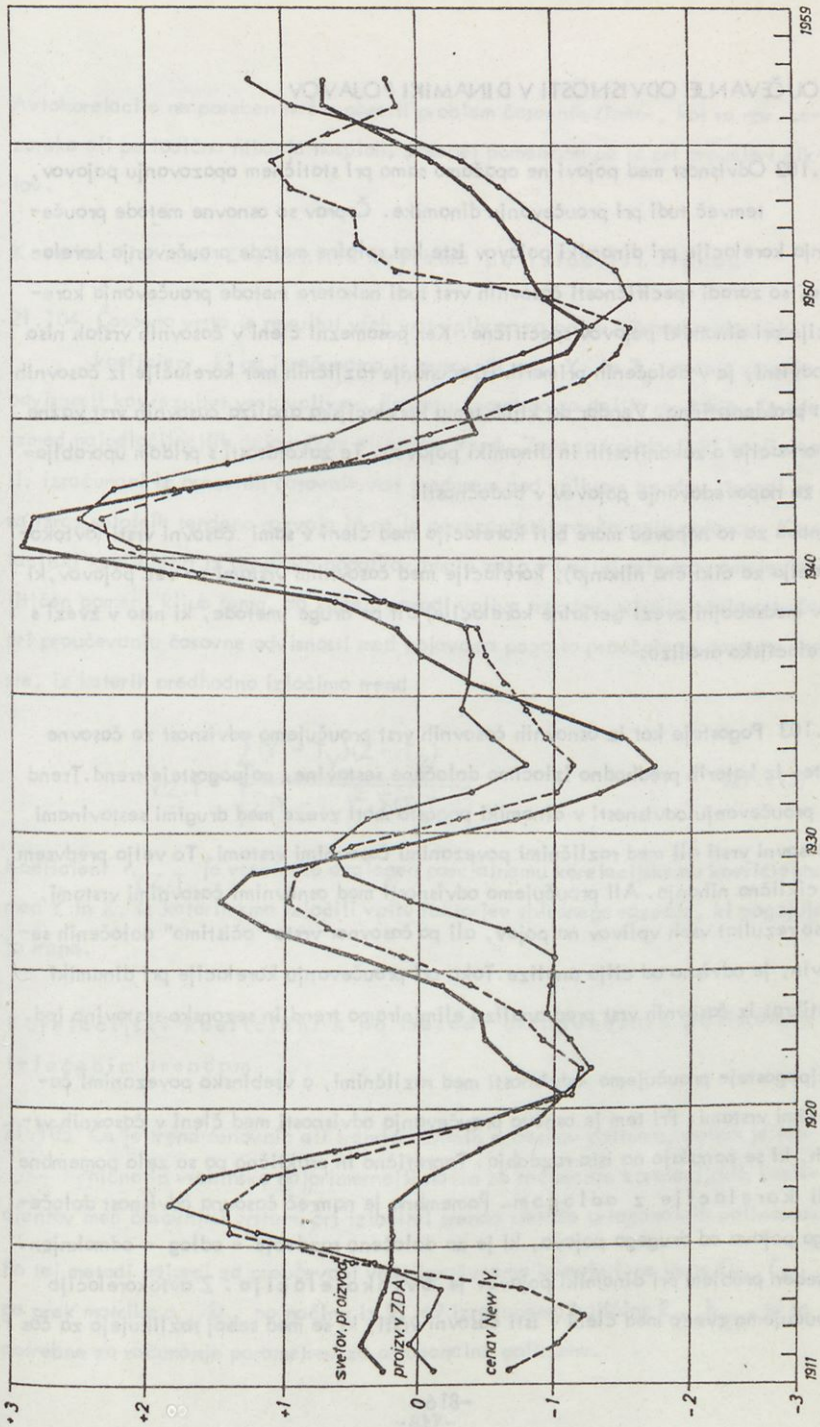
S standardizacijo ciklične sestavine izenačimo razlike v amplitudah ciklov za različne pojave in omogočimo boljšo primerjavo med cikli za pojave, ki so v medsebojni zvezi in odvisnosti.

21.101 Za zgled vzemimo časovne vrste svetovne proizvodnje, proizvodnje v ZDA in povprečne letne cene živega srebra. Neočiščene časovne vrste kvocientov na trend $Y/T^3 = K = CI$ so očiščene iregularnih vplivov z vrsto drsečih sredin $\tilde{K}$, ponderiranimi z binomskimi koeficienti 1 2 1. Časovno vrsto $\tilde{K} = C$ štejemo za čisto ciklične indekse. Za vsako časovno vrsto drsečih sredin so izračunani $\bar{C}$ in SD_C, ki so osnova za standardizacijo cikla po obrazcu 21.114. Standardizirani ciklični indeksi za vse tri časovne vrste so narisani v grafikonu v sliki 21.28. Iz slike lepo sledimo zakonitosti odvisnosti med proučevanimi tremi cikli.

Za proučevane tri časovne vrste za obdobje 1911 do 1959 je:

	$\bar{C}$	SD_C
Svetovna proizvodnja	99,7	24,8
Proizvodnja v ZDA	98,7	46,2
Cena za jeklenko	99,5	30,1

SD_C je razen tega, da z njim standardiziramo cikel, pomemben pokazovalec cikličnega vpliva.



Slika 21.28. Standardizirani ciklični indeksi za proizvodnjo živega srebra

PROUČEVANJE ODVISNOSTI V DINAMIKI POJAVOV

21.102 Odvisnost med pojavi ne opažamo samo pri statičnem opazovanju pojavov, temveč tudi pri proučevanju dinamike. Čeprav so osnovne metode proučevanja korelacije pri dinamiki pojavov iste kot splošne metode proučevanja korelacije, so zaradi specifičnosti časovnih vrst tudi nekatere metode proučevanja korelacije pri dinamiki pojavov specifične. Ker posamezni členi v časovnih vrstah niso neodvisni, je v določenih primerih izračunanje različnih mer korelacije iz časovnih vrst problematično. Vendar da kljub temu korelacijska analiza časovnih vrst važne informacije o zakonitostih in dinamiki pojavov. Te zakonitosti s pridom uporabljamo za napovedovanje pojavov v bodočnosti.
Osnova za to napoved more biti korelacija med členi v sami časovni vrsti (avtokorelacija za ciklična nihanja), korelacije med časovnimi vrstami za več pojavov, ki so v medsebojni zvezi (serialne korelacije) ali pa druge metode, ki niso v zvezi s korelacijsko analizo.

21.103 Pogosteje kot iz osnovnih časovnih vrst proučujemo odvisnost za časovne vrste, iz katerih predhodno izločimo določene sestavine, najpogosteje trend. Trend pri proučevanju odvisnosti v dinamiki pogosto moti zveze med drugimi sestavinami v časovni vrsti ali med različnimi povezanimi časovnimi vrstami. To velja predvsem za ciklična nihanja. Ali proučujemo odvisnosti med osnovnimi časovnimi vrstami, ki so rezultat vseh vplivov na pojav, ali pa časovno vrsto "očistimo" določenih sestavin, je odvisno od cilja analize. Tako pri proučevanju korelacije pri dinamiki dostikrat iz časovnih vrst pred analizo eliminiramo trend in sezonsko sestavino ipd.

Najpogosteje proučujemo odvisnosti med različnimi, a vsebinsko povezanimi časovnimi vrstami. Pri tem je osnova proučevanja odvisnosti med členi v časovnih vrstah, ki se nanašajo na ista razdobja. Teoretično in praktično pa so zelo pomembne tudi k o r e l a c i j e z o d l o g o m. Pomembna je namreč časovna odvisnost določenega pojava od drugega pojava, ki je za določeno razdobje - odlog - odmaknjen. Poseben problem pri dinamiki pojavov je a v t o k o r e l a c i j a. Z avtokorelacijo proučujemo zvezo med členi v isti časovni vrsti, ki se med seboj razlikujejo za čas t.

Avtokorelacija na poseben način osvetli problem časovnih ritmov, kot so npr. sezonska ali periodična nihanja nasploh, posebej pomembna pa je pri proučitvi ciklov.

Korelacija med časovnima vrstama po izločitvi trenda

21.104 Časovna vrsta je rezultat vseh vrst vplivov na pojav. Zato korelacijski koeficient, ki ga izračunamo iz osnovnih vrst Y_x in Z_x podaja stopnjo odvisnosti kot rezultat vseh vplivov. Pogosto, posebno za daljša razdobja, je eden izmed najodločilnejših dejavnikov dinamike trend. Zato so korelacijski koeficienti, izračunani iz osnovnih časovnih vrst predvsem pod vplivom trendov, trendi pa so izraz splošnih tendenc razvoja in ne le povezanosti proučevanih pojavov. Korelacijski koeficienti iz osnovnih podatkov imajo zato v večini primerov omejen analitičen pomen, kljub temu, da kažejo zaradi vpliva trendov, visoko odvisnost. Zato pri proučevanju časovne odvisnosti med pojavoma pogosto proučujemo časovne vrste, iz katerih predhodno izločimo trend

$$r_{YZ.T} = \frac{\sum (Y - T_Y)(Z - T_Z)}{\sqrt{\sum (Y - T_Y)^2 \cdot \sum (Z - T_Z)^2}} \qquad (21.115)$$

Koeficient $r_{YZ.T}$ je vsebinsko analogen parcialnemu korelacijskemu koeficientu med Y in Z, iz katerih smo izločili vpliv faktorjev splošnega razvoja, ki pogojujejo trend.

Korelacijski koeficient s po metodi ortogonalnih polinomov izločenim trendom

21.105 Če je trend osnovnih ali transformiranih podatkov polinom, potem je tehnično in vsebinsko najprimernejši način za računanje korelacijskih koeficientov med časovnimi vrstami pri izločitvi trenda metoda ortogonalnih polinomov.

Po tej metodi najprej za proučevani vrsti izračunamo kumulativne vsote C_{Yk}, C_{Zk}, iz teh pa prek matrike d_{kh}/D_k po načinu iz 21.69 izračunamo količine E_{Yk}, E_{Zk}, ki so potrebne za računanje parametrov za ortogonalne polinome.

Korelacijski koeficient med časovnima vrstama, iz katerih smo trend stopnje p izločili, izračunamo po obrazcu

$$r_{YZ.T^p} = \frac{\sum_x Y_x Z_x - \sum_k \frac{E_{Yk} \cdot E_{Zk}}{D_k}}{\sqrt{\left(\sum_x Y_x^2 - \sum_k \frac{E_{yk}^2}{D_k}\right)\left(\sum_x Z_x^2 - \sum_{k=0}^{p} \frac{E_{Zk}^2}{D_k}\right)}} \qquad (21.116)$$

Če je $p = 0$, je $r_{YZ.T^0}$ korelacijski koeficient med osnovnimi podatki Y_x, Z_x, $r_{YZ.T^1}$ je korelacijski koeficient med vrstama Y_x in Z_x, iz katerih je izločen trend 1. stopnje. $r_{YZ.T^2}$ je korelacijski koeficient med Y_x in Z_x, iz katerih je eliminiran trend v obliki parabole druge stopnje itd.

Korelacijske koeficiente iz vrst, iz katerih smo postopoma izločali trend 0, 1, 2, 3. stopnje računamo s tablicami ortogonalnih polinomov: $d_{kh}|D_k$ po shemi 21.31.

Tabela 21.31. Shema za računanje korelacijskih koeficientov z izločenim trendom.

				D_k	E_{Yk}	E_{Zk}	$\sum Y^2$	$\sum YZ$	$\sum Z^2$	
			d_{00}	D_0	E_{Y0}	E_{Z0}	$-E_{Y0}^2/D_0 = K_{Y0}$	$-E_{Y0}E_{Z0}/D_0 = K_{YZ0}$	$-E_{Z0}^2/D_0 = K_{Z0}$	$r_0 = \frac{K_{YZ0}}{\sqrt{K_{Y0}K_{Z0}}}$
d_{kh}		d_{11}	d_{10}	D_1	E_{Y1}	E_{Z1}	$-E_{Y1}^2/D_1 = K_{Y1}$	$-E_{Y1}E_{Z1}/D_1 = K_{YZ1}$	$-E_{Z1}^2/D_1 = K_{Z1}$	$r_1 = \frac{K_{YZ1}}{\sqrt{K_{Y1}K_{Z1}}}$
	d_{22}	d_{21}	d_{20}	D_2	E_{Y2}	E_{Z2}	$-E_{Y2}^2/D_2 = K_{Y2}$	$-E_{Y2}E_{Z2}/D_2 = K_{YZ2}$	$-E_{Z2}^2/D_2 = K_{Z2}$	$r_2 = \frac{K_{YZ2}}{\sqrt{K_{Y2}K_{Z2}}}$
d_{33}	d_{32}	d_{31}	d_{30}	D_3	E_{Y3}	E_{Z3}	$-E_{Y3}^2/D_3 = K_{Y3}$	$-E_{Y3}E_{Z3}/D_3 = K_{YZ3}$	$-E_{Z3}^2/D_3 = K_{Z3}$	$r_3 = \frac{K_{YZ3}}{\sqrt{K_{Y3}K_{Z3}}}$
C_{Y3}	C_{Y2}	C_{Y1}	C_{Y0}							
C_{Z3}	C_{Z2}	C_{Z1}	C_{Z0}							

Stopnje za računanje ustreznih korelacijskih koeficientov pa so:

a) Iz osnovnih časovnih vrst izračunamo kumulativne vsote:

$$C_{Yh},\ C_{Zh};\ h = 0 \ldots k$$

b) Izračunamo izraze:

$$E_{Yk} = \sum_{h} d_{hk} \cdot C_{Yh}$$

$$E_{Zk} = \sum_{h} d_{hk} \cdot C_{Zh}$$

c) Iz osnovnih časovnih vrst Y_x , Z_x izračunamo vsote:

$$\sum Y_x^2 \quad \sum Y_x Z_x, \quad \sum Z_x^2.$$

d) Kot kaže shema postopno izračunamo izraze E_{Yk}^2/D_k; $E_{Yk}E_{Zk}/D_k$; E_{Zk}^2/D_k in jih odštevamo od predhodnih količin K, da dobimo K stopnje h.

21.106 Za zgled vzemimo časovni vrsti indeksov v industriji usnja in obutve ter tekstilni industriji v SFRJ v razdobju 1946-1964. Osnovni indeksi proizvodnje so dani v tabeli 21.32.

Tabela 21.32. Indeksi proizvodnje v industriji usnja in obutve ter tekstilni industriji v SFRJ v razdobju 1946-1964 (1952=100)

Leto		1946	1947	1948	1949	1950	1951	1952	1953	1954
Indeks industrije	tekstila	67	103	130	133	127	110	100	97	120
	usnja in obutev	63	103	140	162	150	127	100	92	105

Leto		1955	1956	1957	1958	1959	1960	1961	1962	1963	1964
Indeks industrije	tekstila	138	145	167	181	196	224	230	256	295	334
	usnja in obutev	123	137	157	172	199	240	258	272	318	388

Za naš zgled je računanje prikazano v tabeli 21.33.

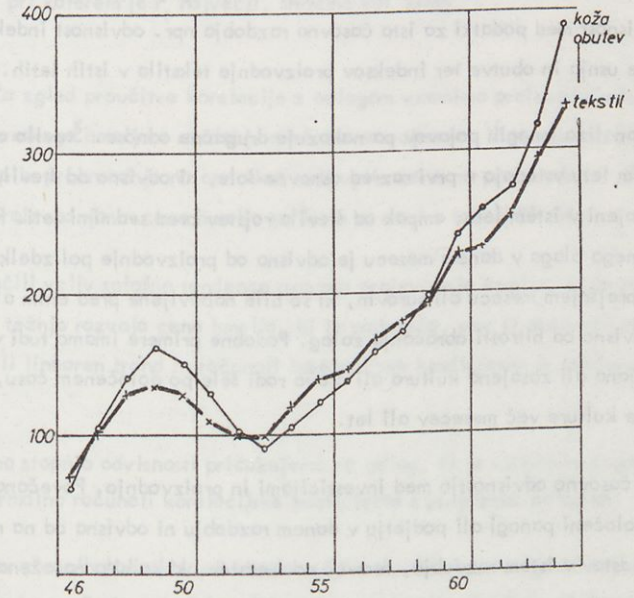
Tabela 21.33. Izračunanje korelacijskih koeficientov z izločitvijo trendov za odvisnost med proizvodnjo tekstilne in usnjarske industrije

N = 19

d_{kh}				D_k	$E_{.Yk}$	$E_{.Zk}$	$\sum_x Y_x^2$	$\sum_x Y_x Z_x$	$\sum_x Z_x^2$	$r_k = \frac{K_{kYZ}}{\sqrt{K_{kY} \cdot K_{kZ}}}$
							618053	658448	706088	
			1	19	3153	3306	-523232	-548622	-575244	$r_{YZ.T_0} = 0,986$
							$K_o =$ 94821	109826	130844	
		1	-9	570	-6640	-7261	-77350	-84584	-92495	$r_{YZ.T_1} = 0,975$
							$K_1 =$ 17471	25242	38349	
	2	-17	51	13566	10866	18255	-8703	-14622	-24565	$r_{YZ.T_2} = 0,965$
							$K_2 =$ 8768	10620	13784	
5	-40	136	-204	213180	-21905	-33941	-2251	-3488	-5404	$r_{YZ.T_3} = 0,965$
							$K_3 =$ 6517	7132	8380	

C_{Yh}	411383	109796	21737	3153
C_{Zh}	444407	116015	22493	3306

Korelacija med indeksom industrijske proizvodnje tekstila ter usnja in obutve je velika tudi med odkloni od trenda 3. stopnje. To kaže na izrazito pozitivno povezanost v dinamiki med obema strokama, ki je ilustrirana tudi z grafikonom osnovnih časovnih vrst. Ker gresta obe stroki vzporedno z razvojem standarda, je visoka zveza utemeljena in pomembna.



Slika 21.29. Indeks industrijske proizvodnje za tekstil in usnje ter obutev

21.107 Kot drug zgled nakažimo indeks proizvodnje gumi jastih cevi v SFRJ v primerjavi z indeksom skupne industrijske proizvodnje SFRJ v razdobju 1946-1964. Ker obe vrsti nakazujeta smer razvoja v eksponentni funkciji, so bili analizirani logaritmi iz osnovnih podatkov. Za ti vrsti je korelacijski koeficient iz logaritmov osnovnih podatkov $r_{logY.logZ.0} = 0,95$. Če pa izločimo linearni trend logaritmov, dobimo $r_{logY.logZ.1} = -0,255$. Visok pozitiven korelacijski koeficient gre na račun pozitivnega trenda v obeh časovnih vrstah. Če tega izločimo, se izka-

že, da je korelacijski koeficient med učinki drugih vplivov celo negativen.

Korelacija z odlogom

21.108 V prejšnjem odstavku smo proučevali korelacijo časovnih vrst med dvema pojavoma. Za tovrstno korelacijo je razen drugega značilno tudi to, da iščemo odvisnost med podatki za ista časovna razdobja npr. odvisnost indeksov proizvodnje usnja in obutve ter indeksov proizvodnje tekstila v istih letih.

Vsebinska analiza mnogih pojavov pa nakazuje drugačne odnose. Število otrok, ki v določenem letu vstopajo v prvi razred osnovne šole, ni odvisno od števila otrok, ki so bili rojeni v istem letu, ampak od števila rojstev pred sedmimi leti. Proizvodnja določenega blaga v danem mesecu je odvisna od proizvodnje polizdelkov proizvedenih v prejšnjem mesecu ali surovin, ki so bile nabavljene pred enim ali več meseci, odvisno od hitrosti obračanja zalog. Podobne primere imamo tudi v kmetijstvu. Zasejana ali zasajena kultura ali drevo rodi šele po določenem času, ki je za nekatere kulture več mesecev ali let.

Enako je s časovno odvisnostjo med investicijami in proizvodnjo. Povečana proizvodnja v določeni panogi ali podjetju v danem razdobju ni odvisna od na novo naloženih sredstev v istem razdobju, temveč od sredstev, ki so bila naložena v preteklosti.

V vseh teh in podobnih primerih vsebinsko nima smisla iskati zvezo oziroma odvisnost med podatki, ki se nanašajo na ista časovna razdobja ali trenutke, temveč je bolj smiselno in upravičeno proučevati odvisnost med podatki, ki so "odloženi" za določeno razdobje. Tako ima npr. smisel iskati zvezo med številom rojenih v letu 1946 in številom vpisanih v prvi razred osnovne šole v letu 1953, rojenih v letu 1947 s številom vpisanih v letu 1954 itd. tj., istih generacij.

Medtem ko smo v prejšnjem odstavku iskali korelacijo med vrstama Y_x in Z_x (oziroma iz odklonov od trenda, iščemo v navedenih zgledih korelacijo med vrstama Y_x in Z_{x+t}. V takih primerih govorimo o korelaciji s časovnim odlogom, pri čemer je t časovni odlog.

Pogosto dolžina odloga za dva pojava ni znana vnaprej, posebno, če sta primerjani vrsti rezultat različnih pogojev z različnimi dolžinami odloga. V tem primeru je naloga statistične analize, da razen stopnje korelacije z odlogom ugotovi tudi dolžino odloga. Da bi dobili dolžino odloga, izračunamo vrsto korelacijskih koeficientov z različnimi odlogi. Med korelacijskimi koeficienti z odlogom je eden absolutno največji. t, pri katerem je r_t največji, štejemo kot odlog.

21.109 Za zgled proučitve korelacije z odlogom vzemimo proizvodnjo in ceno slovenskega hmelja v razdobju med obema vojnama. V tem primeru pričakujemo tako vpliv proizvodnje na ceno kot tudi vpliv cene na proizvodnjo. Osnovni podatki za proizvodnjo in ceno hmelja po letih so dani v tabeli 21.34.

Da bi izločili vpliv splošne tendence razvoja proizvodnje hmelja, ki je naraščajoča in splošno težnjo razvoja cene hmelja, ki je padajoča, smo iz osnovnih časovnih vrst izločili linearen trend in računali korelacijske koeficiente iz odklonov od trenda.

Maksimalno stopnjo odvisnosti pričakujemo za odlog, ki je vsebinsko pogojen, čeprav moremo formalno računati korelacijske koeficiente s poljubnim odlogom.

Iz odklonov zbranih podatkov od trendov so izračunani korelacijski koeficienti za odlog cen od $t = -5$ do $t = +5$ let. Pri računanju korelacijskih koeficientov z odlogom se število dvojic podatkov zmanjša za toliko, za kolikor členov je en podatek odložen v primerjavi z drugim.

Tabela 21.34. Proizvodnja in cena slovenskega hmelja v razdobju med obema vojnama (Proizvodnja v 100 kg, cena v din)

Leto	Proizvodnja			Cena		
	Y_x	$T(Y_x)$	$Y_x - T(Y_x)$	Z_x	$T(Z_x)$	$Z_x - T(Z_x)$
1919	30	47	-17	15	52	-37
1920	35	55	-20	18	50	-32
1921	42	64	-22	50	49	+ 1
1922	45	72	-27	17	48	-31
1923	62	81	-19	100	47	+53
1924	60	89	-29	85	45	+40
1925	70	98	-28	100	44	+56
1926	120	106	14	95	43	+52
1927	200	114	86	40	41	- 1
1928	290	123	167	20	40	-20
1929	230	131	99	6	39	-33
1930	130	140	-10	10	37	-27
1931	110	148	-38	9	36	-27
1932	85	157	-72	16	35	-19
1933	110	165	-55	70	34	+36
1934	130	173	-43	24	32	- 8
1935	205	182	+23	23	31	- 8
1936	201	191	+10	22	30	- 8
1937	220	199	+21	18	28	-10
1938	157	207	-50	30	27	+ 3
1939	230	216	+14	46	26	+20
1940	220	224	- 4	25	25	0

Za odlog $t = +1$ je v tabeli 21.35 nakazano, katere člene vzamemo v pare pri izračunanju korelacijskega koeficienta r_{+1} iz Y_x in Z_{x+1}

Tabela 21.35. Odloženi vrsti prpizvodnje in cena hmelja za t =+1

X	Y_x	Z_{x+1}
1919	-17	-32
1920	-20	+ 1
1921	-22	- 31
.	.	.
.	.	.
.	.	.
1936	+10	-10
1937	+21	+ 3
1938	-50	+20
1939	+14	0

Korelacijski koeficienti z odlogom za proizvodnjo in ceno hmelja v razdobju med obema vojnama so

$r_{-5} = 0,269$ $r_{-4} = 0,657$ $r_{-3} = 0,792$ $r_{-2} = 0,629$ $r_{-1} = 0,152$

$r_o = -0,198$ $r_{+1} = -0,477$ $r_{+2} = -0,434$ $r_{+3} = -0,346$ $r_{+4} = -0,112$

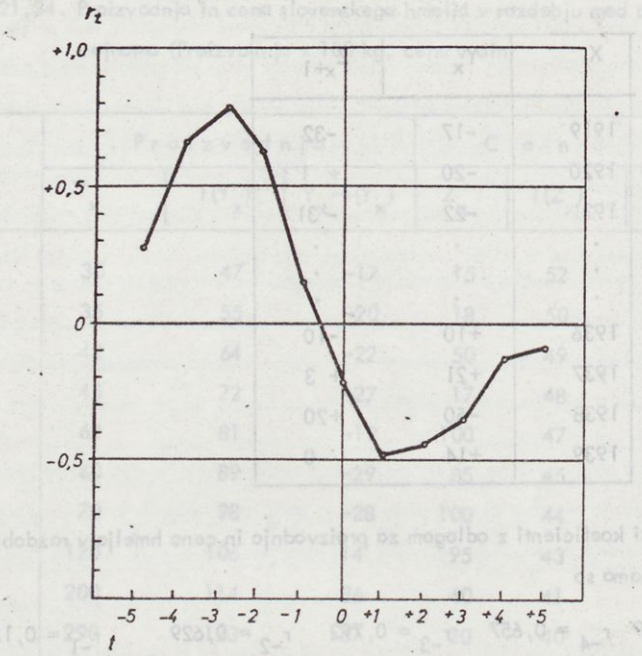
$r_{+5} = 0,088$

Ti podatki so prikazani v sliki 21.30.

Najvišjo pozitivno odvisnost med proizvodnjo in ceno hmelja opazimo za $t = -3$ ($r_{-3} = 0,792$), najvišjo negativno odvisnost pa za $t = +1$ ($r_{+1} = -0,477$).

Spremembe cen hmelja so pokazale najvišjo pozitivno odvisnost s proizvodnjo po treh letih, kar je v skladu z vegetacijsko dobo hmelja. (Visoka cena - visoka proizvodnja po treh letih ; nizka cena - nizka proizvodnja po treh letih).

Čeprav negativna zveza za $t = +1$ ni tako izrazita kot prejšnja ($r_{+1} = -0,477$) kaže pa na to, da je proizvodnja hmelja najmočneje negativno povezana s ceno čez eno leto, še vedno dosti močno čez dve leti, kar je v skladu s ponudbo in povpraševanjem po hmelju.



Slika 21.30. Korelacijski koeficienti z odlogom med proizvodnjo in ceno za slovenski hmelj v razdobju 1919-1940

21.110 Avtokorelacija. V prejšnjem odstavku smo nakazali problem korelacije z odlogom med dvema časovnima vrstama. Korelacijo z odlogom pa moremo proučevati tudi na eni sami časovni vrsti, če proučujemo odvisnost med členi iste vrste, ki so odmaknjeni za t členov. V tem primeru govorimo o avtokorelaciji. S proučevanjem avtokorelacije odkrivamo razen drugega tudi določene zakonitosti v ritmičnem pojavljanju v proučevani časovni vrsti. Zaradi tega je avtokorelacija pomemben pripomoček pri analizi časovnih vrst oziroma dinamike pojavov.

Kot pri korelaciji z odlogom proučujemo korelacijo med vrstama Y_x in Z_{x+t}, proučujemo pri avtokorelaciji korelacijo med vrstama Y_x in Y_{x+t}, tj. med členi iste časovne vrste odloženimi za t.

21.111 Za zgled za avtokorelacijo vzemimo časovno vrsto indeksov volumna za rastlinsko proizvodnjo v SFRJ, iz katere smo predhodno z vrsto drsečih sredin izločili trend. Osnova za izračunanje korelacijskih koeficientov za avtokorelacijo za indeks volumna rastlinske proizvodnje v SFRJ v razdobju 1948-1959 so indeksi na vrsto drsečih sredin:

Tabela 21.36. Osnova za računanje korelacijskih koeficientov avtokorelacije za indeks volumna rastlinske proizvodnje v SFRJ v razdobju 1948-1959

x	I_x	$\frac{I_x - 100}{u_x}$	u_{x+1}	u_{x+2}
1948	103	+ 3	+11	+24
1949	111	+11	-24	+27
1950	76	-24	+27	-28
1951	127	+27	-28	+20
1952	72	-28	+20	-14
1953	120	+20	-14	+14
1954	86	-14	+14	-18
1955	114	+14	-18	+15
1956	82	-18	+15	-13
1957	115	+15	-13	+12
1958	87	-13	+12	
1959	112	+12		

Iz vrst u_x in u_{x+1} izračunani koeficient avtokorelacije je $r_1 = -0,93$, iz u_x in u_{x+2} izračunani koeficient avtokorelacije pa $r_2 = +0,82$. Če bi izračunali še nadaljnje koeficiente avtokorelacije r_3, r_4, bi za našo vrsto opazili, da so korelacijski koeficienti z odlogom lihega števila let negativni, za sodo število let pa pozitivni, kar kaže na tipičen dvoleten cikel. Pri zaznamovanju korelacijskih koeficientov avtokorelacije pri indeksu odpade predznak, ker je $r_{+t} = r_{-t} = r_t$.

21.112 Multipla korelacija pri dinamiki pojavov. Idejo multiple korelacije moremo s pridom prenesti z vsemi svojstvenimi prvinami tudi na proučevanje dinamike.

Vzemimo za zgled zvezo

$$Y_x = a + bZ_{x.T} + c\,Y_{x-t_1} + d\,U_{x-t_2} + e_x \qquad (21.117)$$

V tem modelu zasledimo korelacijski element z izločitvijo trenda $Z_{x.T}$, element avtokorelacije Y_{x-t_1} in element korelacije z odlogom U_{x-t_2}, pri katerem je U odložen za t_2.

Tako sestavljene zveze in njihovo proučevanje je svojstveno sredstvo za analizo dinamike in služi predvsem za napovedovanje oziroma predvidevanje določenih ekonomskih pojavov. Proučevanje multiple regresije v dinamiki pojavov seveda ni vezano le na linearne odvisnosti, ampak moremo uporabljati tudi nelinearne odvisnosti.

21.113 Durbin-Watsonov preskus. Ko iz časovne vrste izločimo posamezne sestavine ali vplive posameznih dejavnikov, ki jih ugotovimo z regresijsko analizo, so ostanki $e_x = y_x - y'_x$ rezultat neproučenih in slučajnostnih dejavnikov. Če so ti ostanki časovno neodvisni, moremo sklepati, da smo iz časovne vrste izločili vse bistvene dejavnike in so ostali le slučajnostni vplivi. Če pa so ostanki med seboj odvisni, kaže, da nismo izločili vpliva vseh bistvenih dejavnikov. Značilna serialna korelacija med ostanki e_x kaže na odvisnost oziroma na to, da so v časovni vrsti ostali še bistveni dejavniki. Značilnost serialnega korelacijskega koeficienta r_1 z odlogom $t=1$ moremo proučiti neposredno. Izkaže pa se, da je možno posredno proučiti serialno korelacijo z odlogom $t=1$ z von Neumannovim preskusom za slučajnostni razpored vrednosti, ki je obravnavan v odstavku 16.56. Če v obrazcu 16.39 iz tega odstavka izraz

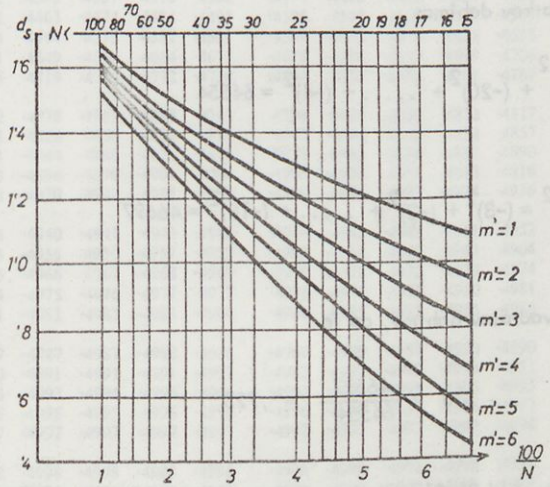
$$\eta = \frac{\sum_{k=1}^{N-1} (y_{k+1} - y_k)^2}{\sum_{k=1}^{N} (y_k - \bar{y})^2} \qquad (21.118)$$

priredimo za preskus ostankov e_x, dobimo Durbin-Watsonov vzorčni izraz d

$$d = \frac{\sum_{x=1}^{N-1} (e_{x+1} - e_x)^2}{\sum_{x=1}^{N} e_x^2} \qquad (21.119)$$

ker je po definiciji poprečje iz ostankov enako nič. Vzorčni izraz d je v ozki zvezi s korelacijskim koeficientom avtokorelacije r_1 in velja približna zveza $d = 2(1 - r_1)$. Za preskus značilnosti Durbin-Watsonovega izraza d je vsebinsko pomembna enostranska preskušnja, ali je d značilno manjši od 2, kar pomeni, da je koeficient avtokorelacije r značilno pozitiven.

V nomogramu v sliki 21.31 so dane spodnje kritične vrednosti d pri $\alpha = P = ,05$. m' pomeni število neodvisnih spremenljivk, ki smo jih vključili v časovno regresijo za y'.



Slika 21.31 Kritične meje d_s za Durbin-Watsonov preskus za $\alpha = P = .05$

21.114 Za zgled vzemimo odklone proizvodnje hmelja od linearnega trenda v Sloveniji v razdobju med obema vojnama iz tabele 21.34

Tabela 21.37 Preskus o značilnosti časovne odvisnosti (r_1) z Durbin-Watsonovim preskusom za odklone proizvodnje hmelja od linearnega trenda za Slovenijo med obema vojnama.

Leto	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
e_x	-17	-20	-22	-27	-19	-29	-28	14	86	167	99
$e_{x+1}-e_x$	- 3	-2	-5	8	-10	1	42	72	81	-68	-109

Leto	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940
e_x	-10	-38	-72	-55	-43	23	10	21	-50	14	-4
$e_{x+1}-e_x$	-28	-34	17	12	66	-13	11	-71	64	-18	

Iz gornjih podatkov dobimo:

$$\sum_{x=1}^{N} e_x^2 = (-17)^2 + (-20)^2 + \ldots\ldots + (-4)^2 = 64554$$

$$\sum_{x=1}^{N} (e_{x+1}-e_x)^2 = (-3)^2 + (-2)^2 + \ldots\ldots + (-18)^2 = 46697$$

Iz obeh vsot kvadratov dobimo, da je

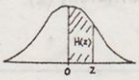
$$d = \frac{46697}{64554} = ,723$$

Iz nomograma v sliki 21.31 odberemo, da je pri N = 22 in m = 1 (ker smo izločili le linearen trend) za naš zgled kritična spodnja meja $d_s = 1,24$. Ker je izračunani d = ,723 znatno manjši od spodnje kritične meje $d_s = 1,24$, sklepamo na visoko časovno odvisnost med odkloni e_x, kar kaže na to, da je v časovni vrsti ostala značilna ciklična sestavina.

***

DODATEK: TABELE

Tabela A NORMALNA PORAZDELITEV



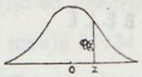
Ploščine H(z)

z	0	1	2	3	4	5	6	7	8	9
0·0	·0000	·0040	·0080	·0120	·0160	·0199	·0239	·0279	·0319	·0359
0·1	·0398	·0438	·0478	·0517	·0557	·0596	·0636	·0675	·0714	·0754
0·2	·0793	·0832	·0871	·0910	·0948	·0987	·1026	·1064	·1103	·1141
0·3	·1179	·1217	·1255	·1293	·1331	·1368	·1406	·1443	·1480	·1517
0·4	·1554	·1591	·1628	·1664	·1700	·1736	·1772	·1808	·1844	·1879
0·5	·1915	·1950	·1985	·2019	·2054	·2088	·2123	·2157	·2190	·2224
0·6	·2258	·2291	·2324	·2357	·2389	·2422	·2454	·2486	·2518	·2549
0·7	·2580	·2612	·2642	·2673	·2704	·2734	·2764	·2794	·2823	·2852
0·8	·2881	·2910	·2939	·2967	·2996	·3023	·3051	·3078	·3106	·3133
0·9	·3159	·3186	·3212	·3238	·3264	·3289	·3315	·3340	·3365	·3389
1·0	·3413	·3438	·3461	·3485	·3508	·3531	·3554	·3577	·3599	·3621
1.1	·3643	·3665	·3686	·3708	·3729	·3749	·3770	·3790	·3810	·3830
1·2	·3849	·3869	·3888	·3907	·3925	·3944	·3962	·3980	·3997	·4015
1·3	·4032	·4049	·4066	·4082	·4099	·4115	·4131	·4147	·4162	·4177
1·4	·4192	·4207	·4222	·4236	·4251	·4265	·4279	·4292	·4306	·4319
1·5	·4332	·4345	·4357	·4370	·4382	·4394	·4406	·4418	·4429	·4441
1·6	·4452	·4463	·4474	·4484	·4495	·4505	·4515	·4525	·4535	·4545
1·7	·4554	·4564	·4573	·4582	·4591	·4599	·4608	·4616	·4625	·4633
1·8	·4641	·4649	·4656	·4664	·4671	·4678	·4686	·4693	·4699	·4706
1·9	·4713	·4719	·4726	·4732	·4738	·4744	·4750	·4756	·4761	·4767
2·0	·4772	·4778	·4783	·4788	·4793	·4798	·4803	·4808	·4812	·4817
2·1	·4821	·4826	·4830	·4834	·4838	·4842	·4846	·4850	·4854	·4857
2·2	·4861	·4864	·4868	·4871	·4875	·4878	·4881	·4884	·4887	·4890
2·3	·4893	·4896	·4898	·4901	·4904	·4906	·4909	·4911	·4913	·4916
2·4	·4918	·4920	·4922	·4925	·4927	·4929	·4931	·4932	·4934	·4936
2·5	·4938	·4940	·4941	·4943	·4945	·4946	·4948	·4949	·4951	·4952
2·6	·4953	·4955	·4956	·4957	·4959	·4960	·4961	·4962	·4963	·4964
2·7	·4965	·4966	·4967	·4968	·4969	·4970	·4971	·4972	·4973	·4974
2·8	·4974	·4975	·4976	·4977	·4977	·4978	·4979	·4979	·4980	·4981
2·9	·4981	·4982	·4982	·4983	·4984	·4984	·4985	·4985	·4986	·4986
3·0	·4987	·4987	·4987	·4988	·4988	·4989	·4989	·4989	·4990	·4990
3·1	·4990	·4991	·4991	·4991	·4992	·4992	·4992	·4992	·4993	·4993
3·2	·4993	·4993	·4994	·4994	·4994	·4994	·4994	·4995	·4995	·4995
3·3	·4995	·4995	·4995	·4996	·4996	·4996	·4996	·4996	·4996	·4997
3·4	·4997	·4997	·4997	·4997	·4997	·4997	·4997	·4997	·4997	·4998
3·5	·4998	·4998	·4998	·4998	·4998	·4998	·4998	·4998	·4998	·4998
3·6	·4998	·4998	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999
3·7	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999
3·8	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999
3·9	·5000	·5000	·5000	·5000	·5000	·5000	·5000	·5000	·5000	·5000

Opomba: obrazložitev glej 13.6 - 13.9

Tabela B NORMALNA PORAZDELITEV

Ordinate $\varphi(z)$

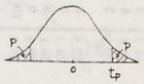


z	0	1	2	3	4	5	6	7	8	9
0·0	·3989	·3989	·3989	·3988	·3986	·3984	·3982	·3980	·3977	·3973
0·1	·3970	·3965	·3961	·3956	·3951	·3945	·3939	·3932	·3925	·3918
0·2	·3910	·3902	·3894	·3885	·3876	·3867	·3857	·3847	·3836	·3825
0·3	·3814	·3802	·3790	·3778	·3765	·3752	·3739	·3725	·3712	·3697
0·4	·3683	·3668	·3653	·3637	·3621	·3605	·3589	·3572	·3555	·3538
0·5	·3521	·3503	·3485	·3467	·3448	·3429	·3410	·3391	·3372	·3352
0·6	·3332	·3312	·3292	·3271	·3251	·3230	·3209	·3187	·3166	·3144
0·7	·3123	·3101	·3079	·3056	·3034	·3011	·2989	·2966	·2943	·2920
0·8	·2897	·2874	·2850	·2827	·2803	·2780	·2756	·2732	·2709	·2685
0·9	·2661	·2637	·2613	·2589	·2565	·2541	·2516	·2492	·2468	·2444
1·0	·2420	·2396	·2371	·2347	·2323	·2299	·2275	·2251	·2227	·2203
1·1	·2179	·2155	·2131	·2107	·2083	·2059	·2036	·2012	·1989	·1965
1·2	·1942	·1919	·1895	·1872	·1849	·1826	·1804	·1781	·1758	·1736
1·3	·1714	·1691	·1669	·1647	·1626	·1604	·1582	·1561	·1539	·1518
1·4	·1497	·1476	·1456	·1435	·1415	·1394	·1374	·1354	·1334	·1315
1·5	·1295	·1276	·1257	·1238	·1219	·1200	·1182	·1163	·1145	·1127
1·6	·1109	·1092	·1074	·1057	·1040	·1023	·1006	·0989	·0973	·0957
1·7	·0940	·0925	·0909	·0893	·0878	·0863	·0848	·0833	·0818	·0804
1·8	·0790	·0775	·0761	·0748	·0734	·0721	·0707	·0694	·0681	·0669
1·9	·0656	·0644	·0632	·0620	·0608	·0596	·0584	·0573	·0562	·0551
2·0	·0540	·0529	·0519	·0508	·0498	·0488	·0478	·0468	·0459	·0449
2·1	·0440	·0431	·0422	·0413	·0404	·0396	·0387	·0379	·0371	·0363
2·2	·0355	·0347	·0339	·0332	·0325	·0317	·0310	·0303	·0297	·0290
2·3	·0283	·0277	·0270	·0264	·0258	·0252	·0246	·0241	·0235	·0229
2·4	·0224	·0219	·0213	·0208	·0203	·0198	·0194	·0189	·0184	·0180
2·5	·0175	·0171	·0167	·0163	·0158	·0154	·0151	·0147	·0143	·0139
2·6	·0136	·0132	·0129	·0126	·0122	·0119	·0116	·0113	·0110	·0107
2·7	·0104	·0101	·0099	·0096	·0093	·0091	·0088	·0086	·0084	·0081
2·8	·0079	·0077	·0075	·0073	·0071	·0069	·0067	·0065	·0063	·0061
2·9	·0060	·0058	·0056	·0055	·0053	·0051	·0050	·0048	·0047	·0046
3·0	·0044	·0043	·0042	·0040	·0039	·0038	·0037	·0036	·0035	·0034
3·1	·0033	·0032	·0031	·0030	·0029	·0028	·0027	·0026	·0025	·0025
3·2	·0024	·0023	·0022	·0022	·0021	·0020	·0020	·0019	·0018	·0018
3·3	·0017	·0017	·0016	·0016	·0015	·0015	·0014	·0014	·0013	·0013
3·4	·0012	·0012	·0012	·0011	·0011	·0010	·0010	·0010	·0009	·0009
3·5	·0009	·0008	·0008	·0008	·0008	·0007	·0007	·0007	·0007	·0006
3·6	·0006	·0006	·0006	·0005	·0005	·0005	·0005	·0005	·0005	·0004
3·7	·0004	·0004	·0004	·0004	·0004	·0004	·0003	·0003	·0003	·0003
3·8	·0003	·0003	·0003	·0003	·0003	·0002	·0002	·0002	·0002	·0002
3·9	·0002	·0002	·0002	·0002	·0002	·0002	·0002	·0002	·0001	·0001

Opomba: obrazložitev glej od 13.6. - 13.9.

Tabela C t-PORAZDELITEV

Kritične vrednosti t_P



m	$\alpha = P$					
	.10	.05	.025	.01	.005	.0005
	$\alpha = 2P$					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

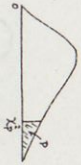
Opomba: obrazložitev glej 15.3.

TABELA Č χ^2 - PORAZDELITEV

Kritične vrednosti χ^2_P

$1 - P$

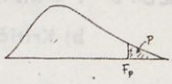
m	0.005	0.010	0.025	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.990	0.995	0.999	m
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	0.0642	0.148	0.275	0.455	0.708	1.07	1.64	2.71	3.84	5.02	6.63	7.88	10.8	1
2	0.0100	0.0201	0.0506	0.103	0.211	0.446	0.713	1.02	1.39	1.83	2.41	3.22	4.61	5.99	7.38	9.21	10.6	13.8	2
3	0.0717	0.115	0.216	0.352	0.584	1.00	1.42	1.87	2.37	2.95	3.67	4.64	6.25	7.81	9.35	11.3	12.8	16.3	3
4	0.207	0.297	0.484	0.711	1.06	1.65	2.19	2.75	3.36	4.04	4.88	5.99	7.78	9.49	11.1	13.3	14.9	18.5	4
5	0.412	0.554	0.831	1.15	1.61	2.34	3.00	3.66	4.35	5.13	6.06	7.29	9.24	11.1	12.8	15.1	16.7	20.5	5
6	0.676	0.872	1.24	1.64	2.20	3.07	3.83	4.57	5.35	6.21	7.23	8.56	10.6	12.6	14.4	16.8	18.5	22.5	6
7	0.989	1.24	1.69	2.17	2.83	3.82	4.67	5.49	6.35	7.28	8.38	9.80	12.0	14.1	16.0	18.5	20.3	24.3	7
8	1.34	1.65	2.18	2.73	3.49	4.59	5.53	6.42	7.34	8.35	9.52	11.0	13.4	15.5	17.5	20.1	22.0	26.1	8
9	1.73	2.09	2.70	3.33	4.17	5.38	6.39	7.36	8.34	9.41	10.7	12.2	14.7	16.9	19.0	21.7	23.6	27.9	9
10	2.16	2.56	3.25	3.94	4.87	6.18	7.27	8.30	9.34	10.5	11.8	13.4	16.0	18.3	20.5	23.2	25.2	29.6	10
11	2.60	3.05	3.82	4.57	5.58	6.99	8.15	9.24	10.3	11.5	12.9	14.6	17.3	19.7	21.9	24.7	26.8	31.3	11
12	3.07	3.57	4.40	5.23	6.30	7.81	9.03	10.2	11.3	12.6	14.0	15.8	18.5	21.0	23.3	26.2	28.3	32.9	12
13	3.57	4.11	5.01	5.89	7.04	8.63	9.93	11.1	12.3	13.6	15.1	17.0	19.8	22.4	24.7	27.7	29.8	34.5	13
14	4.07	4.66	5.63	6.57	7.79	9.47	10.8	12.1	13.3	14.7	16.2	18.2	21.1	23.7	26.1	29.1	31.3	36.1	14
15	4.60	5.23	6.26	7.26	8.55	10.3	11.7	13.0	14.3	15.7	17.3	19.3	22.3	25.0	27.5	30.6	32.8	37.7	15
16	5.14	5.81	6.91	7.96	9.31	11.2	12.6	14.0	15.3	16.8	18.4	20.5	23.5	26.3	28.8	32.0	34.3	39.3	16
17	5.70	6.41	7.56	8.67	10.1	12.0	13.5	14.9	16.3	17.8	19.5	21.6	24.8	27.6	30.2	33.4	35.7	40.8	17
18	6.26	7.01	8.23	9.39	10.9	12.9	14.4	15.9	17.3	18.9	20.6	22.8	26.0	28.9	31.5	34.8	37.2	42.3	18
19	6.84	7.63	8.91	10.1	11.7	13.7	15.4	16.9	18.3	19.9	21.7	23.9	27.2	30.1	32.9	36.2	38.6	43.8	19
20	7.43	8.26	9.59	10.9	12.4	14.6	16.3	17.8	19.3	21.0	22.8	25.0	28.4	31.4	34.2	37.6	40.0	45.3	20
21	8.03	8.90	10.3	11.6	13.2	15.4	17.2	18.8	20.3	22.0	23.9	26.9	29.6	32.7	35.5	38.9	41.4	46.8	21
22	8.64	9.54	11.0	12.3	14.0	16.3	18.1	19.7	21.3	23.0	24.9	27.3	30.8	33.9	36.8	40.3	42.8	48.3	22
23	9.26	10.2	11.7	13.1	14.8	17.2	19.0	20.7	22.3	24.1	26.0	28.4	32.0	35.2	38.1	41.6	44.2	49.7	23
24	9.89	10.9	12.4	13.8	15.7	18.1	19.9	21.7	23.3	25.1	27.1	29.6	33.2	36.4	39.4	43.0	45.6	51.2	24
25	10.5	11.5	13.1	14.6	16.5	18.9	20.9	22.6	24.3	26.1	28.2	30.7	34.4	37.7	40.6	44.3	46.9	52.6	25
26	11.2	12.2	13.8	15.4	17.3	19.8	21.8	23.6	25.3	27.2	29.2	31.8	35.6	38.9	41.9	45.6	48.3	54.1	26
27	11.8	12.9	14.6	16.2	18.1	20.7	22.7	24.5	26.3	28.2	30.3	32.9	36.7	40.1	43.2	47.0	49.6	55.5	27
28	12.5	13.6	15.3	16.9	18.9	21.6	23.6	25.5	27.3	29.2	31.4	34.0	37.9	41.3	44.5	48.3	51.0	56.9	28
29	13.1	14.3	16.0	17.7	19.8	22.5	24.6	26.5	28.3	30.3	32.5	35.1	39.1	42.6	45.7	49.6	52.3	58.3	29
30	13.8	15.0	16.8	18.5	20.6	23.4	25.5	27.4	29.3	31.3	33.5	36.3	40.3	43.8	47.0	50.9	53.7	59.7	30
35	17.2	18.5	20.6	22.5	24.8	27.8	30.2	32.3	34.3	36.5	38.9	41.8	46.1	49.8	53.2	57.3	60.3	66.6	35
40	20.7	22.2	24.4	26.5	29.1	32.3	34.9	37.1	39.3	41.6	44.2	47.3	51.8	55.8	59.3	63.7	66.8	73.4	40
45	24.3	25.9	28.4	30.6	33.4	36.9	39.6	42.0	44.3	46.8	49.5	52.7	57.5	61.7	65.4	70.0	73.2	80.1	45
50	28.0	29.7	32.4	34.8	37.7	41.4	44.3	46.9	49.3	51.9	54.7	58.2	63.2	67.5	71.4	76.2	79.5	86.7	50
75	47.2	49.5	52.9	56.1	59.8	64.5	68.1	71.3	74.3	77.5	80.9	85.1	91.1	96.2	100.8	106.4	110.3	118.6	75
100	67.3	70.1	74.2	77.9	82.4	87.9	92.1	95.8	99.3	102.9	106.9	111.7	118.5	124.3	129.6	135.6	140.2	149.4	100



Opomba: obrazložitev glej 15.2

TABELA D F-PORAZDELITEV

a) Kritične vrednosti $F_{P=.05}$



$p = .05$

m_2 \ m_1	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Opomba: obrazložitev glej 15.4

TABELA D F-PORAZDELITEV

b) Kritične vrednosti $F_{p=.01}$

$p = .01$

m_2 \ m_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

Opomba: obrazložitev glej 15.4

TABELA D F-PORAZDELITEV

c) Kritične vrednosti $F_{P=.001}$

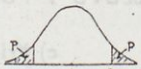
$p = .001$

m_2 \ m_1	1	2	3	4	5	6	8	12	24	∞
1	4052S4	500000	540379	562500	576405	585937	598144	610667	623497	636619
2	998.5	999.0	999.2	999.2	999 3	999.3	999.4	999.4	999.5	999.5
3	167.5	148.5	141.1	137.1	134.6	132.8	130.6	128.3	125.9	123.5
4	74 14	61.25	56.18	53.44	51.71	50.53	49.00	47.41	45.77	44.05
5	47.04	36.61	33.20	31.09	29.75	28.84	27.64	26.42	25.14	23.78
6	35.51	27.00	23.70	21.90	20.81	20.03	19.03	17.99	16.89	15.75
7	29.22	21.69	18.77	17.19	16.21	15.52	14.63	13.71	12.73	11.69
8	25.42	18.49	15.83	14.39	13.49	12.86	12.04	11.19	10.30	9.34
9	22.86	16.39	13.90	12.56	11.71	11.13	10.37	9.57	8.72	7.81
10	21.04	14.91	12.55	11.28	10.48	9.92	9.20	8.45	7.64	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.35	7.63	6.85	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	7.71	7.00	6.25	5.42
13	17.81	12.31	10.21	9.07	8.35	7.86	7.21	6.52	5.78	4.97
14	17.14	11.78	9 73	8.62	7.92	7.43	6.80	6.13	5.41	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.47	5.81	5.10	4.31
16	16.12	10.97	9.00	7.94	7.27	6.81	6.19	5.55	4.85	4.06
17	15.72	10.66	8.73	7.68	7.02	6.56	5.96	5.32	4.63	3.85
18	15.38	10.39	8.49	7.46	6.81	6.35	5.76	5.13	4.45	3.67
19	15.08	10.16	8.28	7.26	6.61	6.18	5.59	4.97	4.29	3.52
20	14.82	9.95	8.10	7.10	6.46	6.02	5.44	4.82	4.15	3.38
21	14.59	9.77	7.94	6.95	6.32	5.88	5.31	4.70	4.03	3.26
22	14.38	9.61	7.80	6.81	6.19	5.76	5.19	4.58	3.92	3.15
23	14.19	9.47	7.67	6.69	6.08	5.65	5.09	4.48	3.82	3.05
24	14.03	9.34	7.55	6.59	5 98	5.55	4.99	4.39	3.74	2.97
25	13.88	9.22	7.45	6.49	5.88	5.46	4.91	4.31	3.66	2.89
26	13.74	9.12	7.36	6.41	5.80	5.38	4.83	4.24	3.59	2.82
27	13.61	9.02	7.27	6.33	5.73	5.31	4.76	4.17	3.52	2.75
28	13.50	8.93	7.19	6.25	5.66	5.24	4.69	4.11	3.46	2.70
29	13.39	8.85	7.12	6.19	5.59	5.18	4.64	4.05	3.41	2.64
30	13.29	8.77	7.05	6.12	5.53	5.12	4.58	4.00	3.36	2.59
40	12.61	8.25	6.60	5.70	5.13	4.73	4.21	3.64	3.01	2.23
60	11.97	7.76	6.17	5.31	4.76	4.37	3.87	3.31	2.69	1.90
120	11.38	7.31	5.79	4.95	4.42	4.04	3.55	3.02	2.40	1.56
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.27	2.74	2.13	1.00

Opomba: obrazložitev glej 15.4

TABELA E STUDENTIZIRAN RAZMIK $q = (\bar{y}_{max} - \bar{y}_{min})/s_y$

Kritične vrednosti $q_{\alpha=2P}$



$\alpha = 2P = .05$

m \ k	2	3	4	5	6	7	8	9	10
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1
2	6.09	8.3	9.8	10.9	11.7	12.4	13.0	13.5	14.0
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.60	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.71	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.84	4.10	4.30	4.46	4.60	4.72	4.83
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.48	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

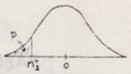
$\alpha = 2P = .01$

m \ k	2	3	4	5	6	7	8	9	10
1	90	135	164	186	202	216	227	237	246
2	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7
3	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7
4	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3
5	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.2
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37
8	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
11	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99
12	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44
16	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
40	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

Opomba: obrazložitev glej 20.6

TABELA F PRESKUS Z MEDIANO

Kritične vrednosti za n_s^+



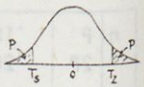
n	P 2P	·10 ·20	·05 ·10	·025 ·05	·01 ·02	·005 ·01
6		0	0	0	–	–
7		1	0	0	0	–
8		1	1	0	0	0
9		2	1	1	0	0
10		2	1	1	0	0
11		2	2	1	1	0
12		3	2	2	1	1
13		3	3	2	1	1
14		4	3	2	2	1
15		4	3	3	2	2
16		4	4	3	2	2
17		5	4	4	3	2
18		5	5	4	3	3
19		6	5	4	4	3
20		6	5	5	4	3
21		7	6	5	4	4
22		7	6	5	5	4
23		7	7	6	5	4
24		8	7	6	5	5
25		8	7	7	6	5
26		9	8	7	6	6
27		9	8	7	7	6
28		10	9	8	7	
29		10	9	8	7	7
30		10	10	9	8	7
31		11	10	9	8	7
32		11	10	9	8	8
33		12	11	10	9	8
34		12	11	10	9	8
35		13	12	11	10	9
36		13	12	11	10	9
37		14	13	12	10	10
38		14	13	12	11	10
39		15	13	12	11	11
40		15	14	13	12	11

n	P 2P	·10 ·20	·05 ·10	·025 ·05	·01 ·02	·005 ·01
41		15	14	13	12	11
42		16	15	14	13	12
43		16	15	14	13	12
44		17	16	15	13	13
45		17	16	15	14	13
46		18	16	15	14	13
47		18	17	16	15	14
48		19	17	16	15	14
49		19	18	17	15	15
50		19	18	17	16	15
52		20	19	18	17	16
54		21	20	19	18	17
56		22	21	20	18	17
58		23	22	21	19	18
60		24	23	21	20	19
62		25	24	22	21	20
64		26	24	23	22	21
66		27	25	24	23	22
68		28	26	25	23	22
70		29	27	26	24	23
72		30	28	27	25	24
74		30	29	28	26	25
76		31	30	28	27	26
78		32	31	29	28	27
80		33	32	30	29	28
82		34	33	31	30	28
84		35	33	32	30	29
86		36	34	33	31	30
88		37	35	34	32	31
90		38	36	35	33	32
92		39	37	36	34	33
94		40	38	37	35	34
96		41	39	37	36	34
98		42	40	38	37	35
100		43	41	39	37	36

Opomba: obrazložitev glej 16.41 in 16.42

TABELA G PRESKUS S PREDZNAČENIMI RANGI - WILCOXONOV PRESKUS

Spodnja T_s in zgornja T_z kritična vrednost za vzorce z n = 6(1) 20 enotami

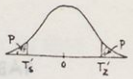


n	P 2P	·10 ·20		·05 ·10		·025 ·05		·01 ·02		·005 ·01	
		T_s	T_z	T_s	T_z	T_s	T_z	T_s	T_z	T_s	T_z
6		3	18	2	19	0	21		-		-
7		5	23	3	25	2	26	0	28		
8		8	28	3	31	3	33	1	35		-
9		10	35	8	37	5	40	3	42	1	44
10		14	41	10	45	8	47	5	50	3	52
11		17	49	13	53	10	56	7	59	5	61
12		21	57	17	61	13	65	10	68	7	71
13		26	65	21	70	17	74	12	79	10	81
14		31	74	25	80	21	84	16	89	13	92
15		36	84	30	90	25	95	19	101	16	104
16		42	94	35	101	30	106	23	113	20	116
17		48	105	41	112	35	118	28	125	24	129
18		55	116	47	124	40	131	33	138	28	143
19		62	128	54	136	46	144	38	152	33	157
20		69	141	60	150	52	158	43	167	38	172

Opomba: obrazložitev glej 16.44 in 16.45

TABELA H PRESKUS Z VSOTO RANGOV - MAN-WITHNEYEV PRESKUS

Spodnja T_s' in zgornja T_z' kritična vrednost za vsote rangov T'



n_1	P	n_2 = 3		4		5		6		7		8		9		10	
		T_s'	T_z'	T_s'	T_z'	T_s'	T_z'	T_s'	T_z'	T_s'	T_z'	T_s'	T_z'	T_s'	T_z'	T_s'	T_z'
1	.10													1	10	1	11
	.05													-	-	-	-
	.01													-	-	-	-
	.001													-	-	-	-
2	.10	3	9	3	11	4	12	4	14	4	16	5	17	5	19	6	20
	.05			-		3	13	3	15	3	17	4	18	4	20	4	22
	.01			-	-	-	-	-	-	-	-	-	-	-	-	-	-
	.001			-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	.10	7	14	7	17	8	19	9	21	10	23	11	25	11	28	12	30
	.05	6	15	6	18	7	20	8	22	8	25	9	27	10	29	10	32
	.01			-		-		-	-	6	27	6	30	7	32	7	35
	.001			-		-		-	-	-	-	-	-	-	-	-	-
4	.10			13	23	14	26	15	29	16	32	17	35	19	37	20	40
	.05			11	25	12	28	13	31	14	34	15	37	16	40	17	43
	.01			-		10	30	11	33	11	37	12	40	13	43	13	47
	.001			-		-	-	-	-	-	-	-	-	10	46	10	50
5	.10					20	35	22	38	23	42	25	45	27	48	28	52
	.05					19	36	20	40	21	44	23	47	24	51	25	55
	.01					-	-	17	43	18	47	19	51	20	55	21	59
	.001					-	-	-	-	15	50	15	55	16	59	16	64
6	.10							30	48	32	52	34	56	36	60	38	64
	.05							28	50	29	55	31	59	33	63	35	67
	.01							24	54	25	59	27	63	28	68	29	73
	.001							21	57	21	63	22	68	23	73	24	78

n_1	P	n_2 = 7		8		9		10	
		T_s'	T_z'	T_s'	T_z'	T_s'	T_z'	T_s'	T_z'
7	.10	41	64	44	68	46	73	49	77
	.05	39	66	41	71	43	76	45	81
	.01	34	71	36	76	37	82	39	87
	.001	29	76	30	82	31	88	33	93
8	.10			55	81	57	87	60	92
	.05			51	85	54	90	56	96
	.01			45	91	48	96	50	102
	.001			40	96	41	103	42	110

n_1	P	n_2 = 9		10	
		T_s'	T_z'	T_s'	T_z'
9	.10	70	101	73	107
	.05	66	105	69	111
	.01	59	112	61	119
	.001	52	119	54	126
10	.10			87	123
	.05			82	128
	.01			74	136
	.001			66	144

Opomba: obrazložitev glej: 16.46 in 16.47

TABELA I PRESKUS S SEKVENCAMI

Spodnja r_s in zgornja r_z kritična vrednost za število sekvenc r za $\alpha = P = .05$ in $\alpha = 2P = 0.10$

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4				9	9														
5			2	2	10	10	11	11											
6		2	2	3	3	11	12	12	13	13	13	13							
7		2	2	3	3	3	13	13	14	14	14	14	15	15	15				
8		2	3	3	3	4	4	14	14	15	15	16	16	16	16	17	17	17	17
9		2	3	3	4	4	5	5	15	16	16	16	17	17	18	18	18	18	18
10		2	3	3	4	5	5	5	6	16	17	17	18	18	18	19	19	19	20
11		2	3	4	4	5	5	6	6	7	17	18	19	19	19	20	20	20	21
12	2	2	3	4	4	5	6	6	7	7	7	19	19	20	20	21	21	21	22
13	2	2	3	4	5	5	6	6	7	7	8	8	20	20	21	21	22	22	23
14	2	2	3	4	5	5	6	7	7	8	8	9	9	21	22	22	23	23	23
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	22	23	23	24	24
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	23	24	25	25
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	25	25	26
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	26	26
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	27
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

r_s n_2

TABELA J FISHERJEVA TRANSFORMACIJA $Z = \frac{1}{2} \ln \frac{1+r}{1-r}$

a) $Z = Z(r)$

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	.100	.110	.121	.131	.141	.151	.161	.172	.182	.192
.2	.203	.213	.224	.234	.245	.255	.266	.277	.288	.299
.3	.310	.321	.332	.343	.354	.365	.377	.388	.400	.412
.4	.424	.436	.448	.460	.472	.485	.497	.510	.523	.536
.5	.549	.563	.576	.590	.604	.618	.633	.648	.662	.678
.6	.693	.709	.725	.741	.758	.775	.793	.811	.829	.848
.7	.867	.887	.908	.929	.950	.973	.996	1.020	1.045	1.071
.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422

r	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
.90	1.472	1.478	1.483	1.488	1.494	1.499	1.505	1.510	1.516	1.522
.91	1.528	1.533	1.539	1.545	1.551	1.557	1.564	1.570	1.576	1.583
.92	1.589	1.596	1.602	1.609	1.616	1.623	1.630	1.637	1.644	1.651
.93	1.658	1.666	1.673	1.681	1.689	1.697	1.705	1.713	1.721	1.730
.94	1.738	1.747	1.756	1.764	1.774	1.783	1.792	1.802	1.812	1.822
.95	1.832	1.842	1.853	1.863	1.874	1.886	1.897	1.909	1.921	1.933
.96	1.946	1.959	1.972	1.986	2.000	2.014	2.029	2.044	2.060	2.076
.97	2.092	2.109	2.127	2.146	2.165	2.185	2.205	2.227	2.249	2.273
.98	2.298	2.323	2.351	2.380	2.410	2.443	2.477	2.515	2.555	2.599
.99	2.646	2.700	2.759	2.826	2.903	2.994	3.106	3.250	3.453	3.800

b) $r = r(Z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	.100	.110	.119	.129	.139	.149	.159	.168	.178	.187
.2	.197	.207	.216	.226	.236	.245	.254	.264	.273	.282
.3	.291	.300	.310	.319	.327	.336	.345	.354	.363	.371
.4	.380	.389	.397	.405	.414	.422	.430	.438	.446	.454
.5	.462	.470	.478	.485	.493	.500	.508	.515	.523	.530
.6	.537	.544	.551	.558	.565	.572	.578	.585	.592	.598
.7	.604	.611	.617	.623	.629	.635	.641	.647	.653	.658
.8	.664	.670	.675	.680	.686	.691	.696	.701	.706	.711
.9	.716	.721	.726	.731	.735	.740	.744	.749	.753	.757
1.0	.762	.766	.770	.774	.778	.782	.786	.790	.793	.797
1.1	.800	.804	.808	.811	.814	.818	.821	.824	.828	.831
1.2	.834	.837	.840	.843	.846	.848	.851	.854	.856	.859
1.3	.862	.864	.867	.869	.872	.874	.876	.879	.881	.883
1.4	.885	.888	.890	.892	.894	.896	.898	.900	.902	.903
1.5	.905	.907	.909	.910	.912	.914	.915	.917	.919	.920
1.6	.922	.923	.925	.926	.928	.929	.930	.932	.933	.934
1.7	.935	.937	.938	.939	.940	.941	.942	.944	.945	.946
1.8	.947	.948	.949	.950	.951	.952	.953	.954	.954	.955
1.9	.956	.957	.958	.959	.960	.960	.961	.962	.963	.963
2.0	.964	.965	.965	.966	.967	.967	.968	.969	.969	.970
2.1	.970	.971	.972	.972	.973	.973	.974	.974	.975	.975
2.2	.976	.976	.977	.977	.978	.978	.978	.979	.979	.980
2.3	.980	.980	.981	.981	.982	.982	.982	.983	.983	.983
2.4	.984	.984	.984	.985	.985	.985	.986	.986	.986	.986
2.5	.987	.987	.987	.987	.988	.988	.988	.988	.989	.989
2.6	.989	.989	.989	.990	.990	.990	.990	.990	.991	.991
2.7	.991	.991	.991	.992	.992	.992	.992	.992	.992	.992
2.8	.993	.993	.993	.993	.993	.993	.993	.994	.994	.994
2.9	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995

Opomba: obrazložitev glej 18.26

TABELA K ORTOGONALNI POLINOMI BINOMSKIH FUNKCIJ

Matrike d_{kh} ID_k za N = 2(1) 30

N = 1

1	:	1

N = 2

1	:	2
2	:	2
−1	:	2

N = 3

1	:	3
−1	:	2
1	:	6

N = 4

1	:	4
2	:	4
−3	:	20
2	:	4
−2	:	4
1	:	4
4	:	20
−1	:	20

N = 5

1	:	5
−2	:	10
2	:	14
−1	:	10
1	:	70

N = 6

1	:	6
2	:	70
−5	:	70
3	:	84
−6	:	84
5	:	84
10	:	180
−15	:	180
12	:	180
−5	:	180
14	:	28
−14	:	28
9	:	28
−4	:	28
1	:	28
252	:	252
−126	:	252
56	:	252
−21	:	252
6	:	252
−1	:	252

N = 7

1	:	7
−3	:	28
2	:	84
−5	:	84
5	:	84
−1	:	6
3	:	154
−10	:	154
18	:	154
−21	:	154
14	:	154
−1	:	84
5	:	84
−14	:	84
28	:	84
−42	:	84
42	:	84

N = 8

1	:	8
2	:	168
−7	:	168
2	:	168
−6	:	168
7	:	168
4	:	264
−10	:	264
12	:	264
−7	:	264
7	:	616
−20	:	616
30	:	616
−28	:	616
14	:	616
−1	:	2184
7	:	2184
−28	:	2184
56	:	2184
−70	:	2184
84	:	2184
−126	:	2184
112	:	2184
−70	:	2184
30	:	2184
−7	:	2184

N = 9

1	:	9
−4	:	60
6	:	2772
−21	:	2772
28	:	2772
5	:	990
−14	:	990
21	:	990
−14	:	990
5	:	990
14	:	2002
−35	:	2002
45	:	2002
−35	:	2002
14	:	2002
18	:	468
−36	:	468
40	:	468
−30	:	468
15	:	468
−4	:	468

N = 10

1	:	10
−9	:	330
2	:	330
−4	:	132
1	:	132
6	:	132
−42	:	8580
10	:	8580
−35	:	8580
56	:	8580
−42	:	8580
18	:	2860
−30	:	2860
45	:	2860
−40	:	2860
10	:	2860
12	:	780
−30	:	780
40	:	780
−35	:	780
20	:	780
−6	:	780

N = 11

1	:	11
−5	:	110
15	:	858
−30	:	4 290
33	:	8 008
−33	:	15 912
33	:	15 912
−30	:	15 912
15	:	15 912
−5	:	15 912
1	:	15 912
5	:	15 912
−9	:	15 912
14	:	15 912
−7	:	15 912
2	:	15 912
−9	:	15 912
3	:	15 912
−11	:	15 912
11	:	15 912

N = 12

1	:	12
−11	:	572
55	:	12 012
−33	:	5 148
33	:	8 008
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912
−33	:	15 912
33	:	15 912

N = 13

1	:	13
−6	:	182
22	:	2 002
−11	:	572
11	:	572
−11	:	572
99	:	68 068
−165	:	68 068
135	:	68 068
−63	:	68 068
14	:	68 068
−22	:	6 188
55	:	6 188
−70	:	6 188
56	:	6 188
−28	:	6 188
7	:	6 188

TABELA K (nadaljevanje)

N = 14					1 :	14
				2	−13 :	910
			1	−6	13 :	728
		10	−55	132	−143 :	97 240
	14	−70	165	−220	143 :	136 136
28	−126	280	−385	330	−143 :	235 144
N = 15					1 :	15
				1	−7 :	280
			6	−39	91 :	37 128
		5	−30	78	−91 :	39 780
	70	−385	990	−1 430	1 001 :	6 466 460
126	−630	1 540	−2 310	2 145	−1 001 :	10 581 480
N = 16					1 :	16
				2	−15 :	1 360
			2	−14	35 :	5 712
		20	−130	364	−455 :	1 007 760
	14	−84	234	−364	273 :	470 288
12	−66	176	−286	−286	−143 :	201 522
N = 17					1 :	17
				1	−8 :	408
			2	−15	40 :	7 752
		1	−7	21	−28 :	3 876
	2	−13	39	−65	52 :	16 796
6	−36	104	−182	195	−104 :	100 776
N = 18					1 :	18
				2	−17 :	1 938
			3	−24	68 :	23 256
		2	−15	48	−68 :	23 256
	2	−14	45	−80	68 :	28 424
36	−234	728	−1 365	1 560	−884 :	6 953 544
N = 19					1 :	19
				1	−9 :	570
			2	−17	51 :	13 566
		5	−40	136	−204 :	213 180
	14	−105	360	−680	612 :	2 288 132
3	−21	70	−140	170	−102 :	89 148
N = 20					1 :	20
				2	−19 :	2 660
			2	−18	57 :	17 556
		20	−170	612	−969 :	4 903 140
	35	−280	1 020	−2 040	1 938 :	22 881 320
42	−315	1 120	−2 380	3 060	−1 938 :	31 201 800
N = 21					1 :	21
				1	−10 :	770
			6	−57	190 :	201 894
		5	−45	171	−285 :	432 630
	14	−119	459	−969	969 :	5 720 330
63	−504	1 904	−4 284	5 814	−3 876 :	121 687 020
N = 22					1 :	22
				2	−21 :	3 542
			1	−10	35 :	7 084
		2	−19	76	−133 :	96 140
	14	−126	513	−1 140	1 197 :	8 748 740
28	−238	952	−2 261	3 230	−2 261 :	40 562 340

TABELA K (nadaljevanje)

N = 23						1 :	23
					1	−11 :	1 012
				2	−21	77 :	35 420
			1	−10	41	−77 :	32 890
		14	−133	570	−1 330	1 463 :	13 123 110
	2	−18	76	−190	285	−209 :	340 860
N = 24						1 :	24
					2	−23 :	4 600
				6	−66	253 :	394 680
			20	−210	924	−1 771 :	17 760 600
		2	−20	90	−220	253 :	394 680
	36	−342	1 520	−3 990	6 270	−4 807 :	177 928 920
N = 25						1 :	25
					1	−12 :	1 300
				2	−23	92 :	53 820
			5	−55	253	−506 :	1 480 050
		10	−105	495	−1 265	1 518 :	14 307 150
	6	−60	280	−770	1 265	−1 012 :	7 803 900
N = 26						1 :	26
					2	−25 :	5 850
				1	−12	50 :	16 380
			10	−115	552	−1 150 :	7 803 900
		14	−154	759	−2 024	2 530 :	40 060 020
	12	−126	616	−1 771	3 036	−2 530 :	48 384 180
N = 27						1 :	27
					1	−13 :	1 638
				6	−75	325 :	712 530
			1	−12	60	−130 :	101 790
		14	−161	828	−2 300	2 990 :	56 448 210
	63	−693	3 542	−10 626	18 975	−16 445 :	2 032 135 560
N = 28						1 :	28
					2	−27 :	7 308
				2	−26	117 :	95 004
			4	−50	260	−585 :	2 103 660
		7	−84	450	−1 300	1 755 :	19 634 160
	42	−483	2 576	−8 050	14 950	−13 455 :	1 354 757 040
N = 29						1 :	29
					1	−14 :	2 030
				2	−27	126 :	113 274
			5	−65	351	−819 :	4 207 320
		14	−175	975	−2 925	4 095 :	107 987 880
	21	−252	1 400	−4 550	8 775	−8 190 :	500 671 080
N = 30						1 :	30
					2	−29 :	8 990
				3	−42	203 :	302 064
			10	−135	756	−1 827 :	21 360 240
		70	−910	5 265	−16 380	23 751 :	3 671 587 920
	36	−450	2 600	−8 775	17 550	−16 965 :	2 145 733 200

Opomba: obrazložitev glej 21.68 in 21.69

TABELA L SLUČAJNOSTNA ŠTEVILA

59 58 00 64 78	75 56 97 88 00	88 83 55 44 86	23 76 80 61 56	04 11 10 84 08
38 50 80 73 41	23 79 34 87 63	90 82 29 70 22	17 71 90 42 07	95 95 44 99 53
30 69 27 06 68	94 68 81 61 27	56 19 68 00 91	82 06 76 34 00	05 46 26 92 00
65 44 39 56 59	18 28 82 74 37	49 63 22 40 41	08 33 76 56 76	96 29 99 08 36
27 26 75 02 64	13 19 27 22 94	07 47 74 46 06	17 98 54 89 11	97 34 13 03 58
91 30 70 69 91	19 07 22 42 10	36 69 95 37 28	28 82 53 57 93	28 97 66 62 52
68 43 49 46 88	84 47 31 36 22	62 12 69 84 08	12 84 38 25 90	09 81 59 31 46
48 90 81 58 77	54 74 52 45 91	35 70 00 47 54	83 82 45 26 92	54 13 05 51 60
06 91 34 51 97	42 67 27 86 01	11 88 30 95 28	63 01 19 89 01	14 97 44 03 44
10 45 51 60 19	14 21 03 37 12	91 34 23 78 21	88 32 58 08 51	43 66 77 08 83
12 88 39 73 43	65 02 76 11 84	04 28 50 13 92	17 97 41 50 77	90 71 22 67 69
21 77 83 09 76	38 80 73 69 61	31 64 94 20 96	63 28 10 20 23	08 81 64 74 49
19 52 35 95 15	65 12 25 96 59	86 28 36 82 58	69 57 21 37 98	16 43 59 15 29
67 24 55 26 70	35 58 31 65 63	79 24 68 66 86	76 46 33 42 22	26 65 89 08 02
60 58 44 73 77	07 50 03 79 92	45 13 42 65 29	26 76 08 36 37	41 32 64 43 44
53 85 34 13 77	36 06 69 48 50	58 83 87 38 59	49 36 47 33 31	96 24 04 36 42
24 63 73 87 36	74 38 48 93 42	52 62 30 79 92	12 36 91 86 01	03 74 28 38 73
83 08 01 24 51	38 99 22 28 15	07 75 95 17 77	97 37 72 75 85	51 97 23 78 67
16 44 42 43 34	36 15 19 90 73	27 49 37 09 39	85 13 03 25 52	54 84 65 47 59
60 79 01 81 57	57 17 86 57 62	11 16 17 85 76	45 81 95 29 79	65 13 00 48 60
03 99 11 04 61	93 71 61 68 94	66 08 32 46 53	84 60 95 82 32	88 61 81 91 61
38 55 59 55 54	32 88 65 97 80	08 35 56 08 60	29 73 54 77 62	71 29 92 38 53
17 54 67 37 04	92 05 24 62 15	55 12 12 92 81	59 07 60 79 36	27 95 45 89 09
32 64 35 28 61	95 81 90 68 31	00 91 19 89 36	76 35 59 37 79	80 85 30 05 14
69 57 26 87 77	39 51 03 59 05	14 06 04 06 19	29 54 96 96 16	33 56 46 07 80
24 12 26 65 91	27 69 90 64 94	14 84 54 66 72	61 95 87 71 00	90 89 97 57 54
61 19 63 02 31	92 96 26 17 73	41 83 95 53 82	17 26 77 09 43	78 03 87 02 67
30 53 22 17 04	10 27 41 22 02	39 68 52 33 09	10 06 16 88 29	55 98 66 64 85
03 78 89 75 99	75 86 72 07 17	74 41 65 31 66	35 20 83 33 74	87 53 90 88 23
48 22 86 33 79	85 78 34 76 19	53 15 26 74 33	35 66 35 29 72	16 81 86 03 11
60 36 59 46 53	35 07 53 39 49	42 61 42 92 97	01 91 82 83 16	98 95 37 32 31
83 79 94 24 02	56 62 33 44 42	34 99 44 13 74	70 07 11 47 36	09 95 81 80 65
32 96 00 74 05	36 40 98 32 32	99 38 54 16 00	11 13 30 75 86	15 91 70 62 53
19 32 25 38 45	57 62 05 26 06	66 49 76 86 46	78 13 86 65 59	19 64 09 94 13
11 22 09 47 47	07 39 93 74 08	48 50 92 39 29	27 48 24 54 76	85 24 43 51 59
31 75 15 72 60	68 98 00 53 39	15 47 04 83 55	88 65 12 25 96	03 15 21 92 21
88 49 29 93 82	14 45 40 45 04	20 09 49 89 77	74 84 39 34 13	22 10 97 85 08
30 93 44 77 44	07 48 18 38 28	73 78 80 65 33	28 59 72 04 05	94 20 52 03 80
22 88 84 88 93	27 49 99 87 48	60 53 04 51 28	74 02 28 46 17	82 03 71 02 68
78 21 21 69 93	35 90 29 13 86	44 37 21 54 86	65 74 11 40 14	87 48 13 72 20

TABELA L (nadaljevanje)

98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07	35 44 13 18 80
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94	37 54 87 30 43
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98	94 62 46 11 71
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39	00 38 75 95 79
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27	77 93 89 19 36
74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21	80 81 45 17 48
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55	36 04 09 03 24
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40	88 46 12 33 56
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46	15 02 00 99 94
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15	01 84 87 69 38
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74	64 27 85 80 44
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50	68 47 66 46 59
73 18 95 02 07	47 67 72 52 69	62 29 06 44 64	27 12 46 70 18	41 36 18 27 60
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71	93 82 34 31 78
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95	07 70 61 78 13
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53	31 22 30 84 20
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75	94 11 90 18 40
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95	77 76 22 07 91
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14	83 48 34 70 55
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06	94 54 13 74 08
77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80	72 89 35 55 07
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51	65 34 46 74 15
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46	31 85 33 84 52
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91	08 00 74 54 49
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42	43 86 07 28 34
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22	93 17 49 39 72
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00	71 14 84 36 43
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61	62 32 71 84 23
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60	81 60 41 88 80
65 13 85 68 06	87 64 88 52 61	34 31 36 58 61	45 87 52 10 69	85 64 44 72 77
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49	65 58 44 96 98
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05	40 03 03 74 38
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47	15 50 12 95 78
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41	03 85 65 45 52
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56	64 69 11 92 02
07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38	04 71 36 69 94
60 52 68 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39	01 21 20 04 55
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83	92 30 15 04 98
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 57 18 39	06 41 01 93 62
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51	31 02 47 31 67

Opomba: obrazložitev glej 14.24

TABELA M LOGARITMI TROMESTNIH ŠTEVIL

N	0	1	2	3	4	5	6	7	8	9	Proportional parts 1	2	3	4	5
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16
14	1461	1492	1523	1553	1584	1611	1644	1673	1703	1732	3	6	9	12	15
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	4	5	7	9
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	5	7
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	7
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	2	4	5	6
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	4	5	6
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	5
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	3	3	4
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4
N	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5

TABELA M (nadaljevanje)

N	0	1	2	3	4	5	6	7	8	9	Proportional parts 1	2	3	4	5
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	1	2	3	4
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	3
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	3	3
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	3	3
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	3	3
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	3	3
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	1	1	2	2	3
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2
N	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5

STVARNO KAZALO

A

absolutna razlika 723

agregat 23

ocena z dvostopenjskim vzorčenjem 402

ocena z enostavnim vzorčenjem 360

ocena s sistematičnim vzorčenjem 403

agregatni indeks 295

cen 304

količin 310

verižni 299

α-verjetnost za napako I.vrste

alienacija 543

analitični grafikon 111

analitična obdelava 42

analitična tabela 107

analiza odločitev 505-520

anketa, panel 51

anketar 54, 68

analiza variance 485, 657

čisto slučajnostni poskus 663

dvofaktorski poskus 685

kompleksna 657

latinski kvadrat 678

rangi 497

slučajnostni bloki 673

aritmetična skala 113

aritmetična sredina 220, 224, 475

aritmetičnih sredin 233

modificirana 232

ocena z enostavnim slučajnostnim vzorčenjem 348

ocena razlik z neodvisnimi vzorci 430

ocena razlik z odvisnimi vzorci 432

ocena z malimi vzorci 426

aritmetična sredine

preskušanje domnev z malimi vzorci 476

subjektivna ocena 427

asimetrija 196

mere 276, 277

aposteriorna verjetnost 333, 334

apriorna verjetnost 333, 334

asociacija 587,601
pokazovalci 601
preskušanje domnev 605
preskus z mediano 606
Yatesov popravek 605
atributivni znak 25
avtokorelacija 826
avtomatizirana obdelava 97

B

baza, indeksa 182
β - verjetnost za napako II. vrste 462
Bayes
inverzna verjetnost 144,464
ocena za aritmetično sredino 428,429
sodilo 512,516
sodilo z vzorčenjem 517.
Bernoullijeva porazdelitev 438
bimodalna porazdelitev 200
binomska funkcija 764
binomska porazdelitev 315,438
asimptotična 439
parametri 439
rekurzijski obrazec 449
biologija 17
bloki
slučajnostni 662,670
delni 689
bivariatna porazdelitev 556,557
Brunsmanov grafikon 712

C

census norma 44
centil 44
centilne norme 217
centralna enota, računalnik 97
centralni limitni teorem 350
centralni momenti 277
cikel-standardizacija 813
ciklična nihanja 728,726,732,803
c-karta 643
Coob-Douglasova funkcija 624
C- Pearsonov koeficient kontingence 593

Č

časovna opredelitev 44
časovna vrsta 101
analiza 718
drseče sredine 698
grafično prikazovanje 702
izvedena 693
kumulativna 693
nezvezna numerična 102
sredin 694
trenutna 692,719
razmičma 692
vloga poprečij 731
časovno vzorčenje 409
ocena za časovni agregat 410
ocena za strukturni delež 411

časovni indeksi 177,183
čas popisovanja 48
časovni znak 24,78
čelo, tabela 104
čista korelacijska odvisnost 527
čisti slučajnostni poskus 662
čisti učinek, faktorski poskus 683
črkovne šifre 87
črtanje 117
črtkanje 95

D

decil 207
  decilne norme 217
decilni razmik 251
decimalna klasifikacija 79
delež, strukturni 139
delež, vzorčni 400
delna populacija 32
delni bloki 689
delno opazovanje 50
demografija 17
demografska statistika 19
demografski pokazovalci 314
determinacijski koeficient 541
  za linearno odvisnost 545
  za multiplo korelacijo 622
  za parcialno korelacijo 623
  za partkorelacijo 623
diagramska karta 131
dinamika pojavov 49,69,103,691
Dodge-Romigove tabele 649
dogajanje 23
dogodek 23
dogodek, verjetnostni račun
  elementarni 334
  enakomožen 334
  simetričen 334
dogovorna kratica (tabela) 110
domneva
  enostavna 455
  ničelna 454
  osnovna 454
  preskušanje 453,471
  sestavljena 455
drseči loki 810
Durbin-Watsonov preskus 828
2^P poskus 688
dvofaktorski poskus 662,684
dvofazno vzorčenje 404
dvojna struktura 142
dvojni kontrolni načrt 652,654
dvostopenjsko vzorčenje 403
dvostranski sklep 465
dvostranska trditev 336

E

ekspedicijski način 55
eksponentni trend 733,773
ekstenziven znak 27

elektronski računalnik 97
elementarni dogodek 334
elementi obrazca 60
enakomožen dogodek 334
enojni kontrolni načrt 652,654
enostavna domneva 455
enostavna funkcija stroškov 389
enostavna struktura 138
enostavna tabela 106
enostavni grafikon 111
enostavni indeksi 174-183
enostavni stolpci 118,119
enostavno slučajnostno vzorčenje 348,361
enota (statistična) 22,23
  agregat 23
  enostavna 23
enota (računalniška)
  centralna 97
  izhodna 97
  vhodna 97
  za računanje 97
enota mere 108
enota pogojna 93
enota, poskusna 660
enostranski sklep 336,466
evidenca 53

F

Faktorialni znak 28
figure (graf.prik) 128,157
ϕ^2 - pokazovalec asociacije 602
F-porazdelitev 315,317,422
frekvenca 186
  stvarna (pri kontingenci) 591
  teoretična (pri kontingenci) 590
frekvenčna porazdelitev 103,185
  preskus domneve z malimi vzorci 483
Fisherjeva transformacija Z 559
Fisherjev indeksni obrazec 299,306
faktorski poskusi 683-688
frakcionalni poskus 689
funkcija dohodka, linearna 513
funkcija stroškov, enostavna 389

G

Ganttov grafikon 713
  kompleksen 717
  Ganttov odstotek 715
  geografska vrsta 100
geografski znak 24
geometrijska sredina 220,242
Ginijev koeficient koncentracije 287
glava, tabela 104
glavni obrazec 60
Gompertzova funkcija 734,735
gostota 157,138
gostota frekvence 189
gostota prebivalstva 158
gostota relativne frekvence 190
gotovost- odločitev pri 508
gozdarska statistika 19
grafikon 99,110
  analitični 111

Brunsmanov 712
enostavni 111
figure 118,128-130
Ganttov 713
korelacijski 529
linijski 118,122
logaritemsko verjetnostni 332
polarni 708
pollogaritemski 703
struktur linijski 147
verjetnostni 328
ocenjevanje M in SD 223
grafično prikazovanje 99,110,136
časovne vrste 702
prvine 111
statističnih koeficientov 168-174
struktur 144
koeficientov s strukturnimi deleži 170-174
grupa "drugo" 81
grupa "neznano" 81
grupa višje vrste 77
grupiranje vrednosti znakov 74
časovni znaki 78
krajevni znaki 78
nezvezni znaki 84
numerični znaki 81
splošna načela 76
stvarni znaki 79

H

harmonična sredina 220,234
hipergeometrična porazdelitev 315,317,437
asimptotična 438
rekurzijski obrazec 448
χ^2- pokazovalec asociacije 602
pokazovalec kontingence 592
porazdelitev 315-317,418
adicijski teorem 419
hipotetična populacija 343
Hirschmanov koeficient koncentracije 290
Histogram 191
homogena populacija 31
homogenost strukturnih deležev 598
Hurviczevo pravilo 511

I

Identifikacijski znak 58
individualni obrazec 19
industrijska statistika 19
indeks 138,174,295
agregatni 295
cen 303
časovni 177
enostavni 174-183
izbira osnove 182
količin 310

s stalno bazo 177,723
premična baza 179
preračunavanje na drugo bazo 181
produktivnosti dela 312
recipročnih pokazovalcev 312
verižni 180,299
življenjskih stroškov 309
indeks korelacije 541
indeks koristnosti 507
indeks splošni 38
intenziven znak 27
interakcija 683
intervalna populacija 30
intervalnost znaka 26
intervju 55
intraklasna korelacija 398
inverzna verjetnost 144,464
iregularni vplivi 727,728,808
istorodna populacija 31
izbor enot slučajnostnega vzorčenja 365
izbor slučajnostnih trenutkov 414
izguba - priložnostna 506
izhodna kakovost
karakteristike načrtov 648
najslabša 649
poprečna 648
pričakovana 648,651
izvedena časovna vrsta 693

K

kakovost - izhodna 648,649
kakovost - preverjanje 629
skupin 645
procesov 629
kakovost vhodna 648,649
karta, kontrolna 632
kartogram 118,130
diagramska karta 131
s šrafurami 133
s točkami 135
klasifikacija, decimalna 79
kmetijska statistika 19
kocka, (graf.prikazovanje) 112
koeficient alienacije 543
koeficient dinamike 244
koeficient izrabljenosti 273
koeficient nepovezanosti 543
koeficient multiple korelacije 615
koeficient natalitete 161
koeficient obračanja zalog 161
koeficient parcialne korelacije 619
koeficient variacije 267,268
koeficient,statistični 138
koeficient interklasne korelacije 398
kolektiven obrazec 59
kompleksna analiza variance 657
koncentracija,mere 283-295

kontrola normalna, poostrena 637
kontrola pri prevzemu 647,648
kontrolna karta 636
kontrola napak pri opazovanju 69-71
kontrolni načrt
dvojni 652-654
enojni 652
sekvenčni 652-654
večkratni 652-654
kontrolni slučajnostni vzorec 72
kontrolno območje 638
kontrolni znaki 72
kontingenca 587
merila stopnje 591
kontingenčna tabela 588
korelacija v dinamiki 816
avtokorelacija 826
multipla 828
z odlogom 822
po izločitvi trenda 817
korelacija ranga 584
korelacijska analiza 657
korelacijski grafikon 529
korelacijske odvisnosti 524
korelacijska tabela 528
korelacijski koeficient 556
preskušanje domnev 559
korelacijski koeficient za kontingenco 594
korelacijsko razmerje 395,580
preskušanje značilnosti 583
preskušanje o linearnosti 584
korespondirujoča populacija 33
korespondentni način 57
krajevna opredelitev 44
krajevni znak 24,78
krajevna vrsta 100
kraj opazovanja 66
kraj popisa 66
kritična meja 456,471
za normalno porazdelitev 336,337
za F-porazdelitev 422
za χ^2- porazdelitev 419
za t-porazdelitev 421
kritični trenutek 44,47
krivuljčna odvisnost 533,566,567
krivuljčna multipla korelacija 624
krog, graf.prik. 112
krogla, graf.prik. 112
krogov izsek.graf.prik. 113,149
Kruskal-Wallisov preskus 497
kumulativna časovna vrsta 693
kumulativna frekvenčna porazdelitev 200,224
kvadrat, graf.prik. 112
kvadrat latinski 662,676
kvantil 203
kvantilni rang 205
kvartil 207
kvartilni odklon 251
kvartilni razmik, 251
kvotno vzorčenje 406

L

Laspeyersov indeks 297,298,304
Latinski kvadrat 662,676-683
  analiza variance 678
legenda 117
lestvica 113
limitni teorem, centralni 350
linearna funkcija dohodka 513
linearna korelacija 555
linearna multipla regresija 611,614
linearna odvisnost 543
determinacijski koeficient 545
  preskušanje s korelacijskim razmerjem 584
  pokazovalci 545
linearna parcialna korelacije 617
linearna regresija 547
  model 547
  najmanjši kvadrati 547
  razmik zaupanja 555
linearna skala 113
linijski grafikon 118,122
  pravila za risanje 122
logaritemska skala 114
logaritemsko normalna porazdelitev 320
logična kontrola 71
logistična funkcija 734,735
longitudialno opazovanje 45,46
Lorenzova krivulja 285
Lorenzov grafikon 283
loterijski način izbora 366

M

maksimaks pravilo 510
maksimin pravilo 510
mali vzorci 344
  razmične ocene 425
  preskušanje domnev 475
Man Withneyev preskus 495
matrika posledic 505
mediana 207,220
  preskus z- 493
meja, kritična 456,471
meja razreda 81
meja zaupanja za aritmetično sredino 352
mejne vrednosti 76
mere asimetrije 276
  po legi 276
  izračunane 277
mere sploščenosti 277
  po legi 277
  izračunane 277
mere variacije 249
metoda najmanjših kvadratov 539,547
metoda razmerij 380
Military standard 105 tabele 649
minimaks sodilo 512
množični pojavi 21
moč preskusa 458,468,472
modeli časovnih vrst 729
modificirana aritmetična sredina 232
modificirana eksponentna funkcija 733,752
modus 221
momenti 277

monografija 50,51
mreža, grafikona 116
multipla korelacija 610,622
determinacijski koeficient 622
koeficient 615
krivuljčna 624
multipla regresija 610
koeficient 615
linearna 611
krivuljčna 624

N

načela grupiranja 77
načrt obdelave, shema 89
načrtovanje poskusov 657
prvine 658
najslabša izhodna kakovost 649
največji verjeten odklon 665
napaka I.vrste 456.457,471
krivulja verjetnosti za 468
verjetnost a 461
napaka II.vrste 456,472
krivulja verjetnosti za 468
verjetnost B 468
odvisnost od ničelne domneve 465
napaka ocene, standardna 542
napaka, sistematična 69
napaka slučajna
napaka pri opazovanju 69
naslov obrazca 60
naslov tabele 108
navodila za izpolnjevanje obrazcev 65
negativna odvisnost 532
negotovost odločitev pri 508
neodgovori 405
neopredeljena vhodna kakovost 648
neparametrično preskušanje domnev 491
nepojasnjena varianca 541
neposredno opazovanje 54
nepovezanost, koeficient 543
nepristranska ocena za varianco 353
nesprejemljiva vhodna kakovost 648
Neumann von- preskus 502
nevtralna vhodna kakovost 648
neznačilna razlika 471
nezvezne porazdelitve 437
zveze med njimi 447
rekurzijski obrazci 448
nezvezni znak 25
grupiranje vrednosti 84
ničelna domneva 454,471
normalna kontrola 637
normalna porazdelitev 197,315
dveh spremenljivk 555
gostota 317
logaritemsko normalna 330
porazdelitvena funkcija 318
prilagoditev stvarni 325
standardizirana 319
normiran Pearsonov koeficient kontingence C 593
numerični znak 25
grupiranje vrednosti 81

O

obdelava statističnega gradiva
avtomatizirana 97
osnovna 42,88,94
ročna 94
obdelovalna tabela 107
območje
kritično 456,471
sprejema H_o 466
stabilnosti 638
obračanje zalog
koeficient 161
indeksi 313
obrazec 59
ocena, vzorčna 363
ocena M in SD iz verjetnostnega grafikona 320
odvisnost
funkcijska 521
korelacijska 104,522
linearna 543
krivuljčna 566
negativna 532
pozitivna 532
odklon zaupanja 351
odločitev pri
gotovosti 508
negotovosti 508
statistična analiza 505
tveganju 508
odlog 822
odstotek Ganttov 715
odstotek zajetja 47
odprt razred 82
okoliš popisni 66
okoliš statistični 78
okvir vzorčenja 366
operativna karakteristična krivulja 466-472
operator 412
opredelitev populacije 30
opredelitev predmeta opazovanja 43
opredeljujoči pogoji 22,29
opazovanje statistično 43
optimalna razmestitev enot v vzorcu 387
ordinalnost znaka 26
ortogonalni polinom 763
osebnostna komponenta 659
osnova indeksa 182
osnovna domneva 454,471

P

Paaschejev indeks 297,303
panel anketa 51
parameter 23,33
parameter trenda 733
parcialna korelacija 617 623
partkorelacija 620
Pearsonov koeficient kontingence 593
Pearl-Reedova krivulja 733
periodične variacije 726,728,777
metoda koeficientov dinamike 786
metoda kvocientov na vrsto drsečih sredin 791
metoda vsot 779

metoda vsot s popravkom za smer razvoja 782
Phillipsove tabele 649
pogojne enote 93
pojasnjena varianca 540
pokazovalci dinamike, enostavni 723
Polarni grafikon 708
poligon frekvenčni 191
polimodalna frekvenčna porazdelitev 195,200
polje, tabela 104
pollogaritemski grafikon 703
pomnilnik 97
pomožni momenti 279
ponavljanje poskusov 660
ponderiranje 93
standardna sestava 241
popis 45
popisna komisija 68
popisni okoliš 66
popisnica 58
popolna kontrola pri prevzemu 646
popoln sistem delnih populacij 33
popisovalec 54,68
poprečna kvadratična kontingenca ϕ^2 593
poprečna razlika 263,477
poprečen absoluten odklon 252
poprečen indeks cen 303
populacija 22,29
delna 32
heterogena, raznorodna 31
homogena, istorodna 31
popoln sistem delnih 33
vseh možnih vzorcev 345
Porazdelitev
binomska 438
bivariatna 556
F- 422
frekvenčna 103,185
χ^2
hipergeometrična 437
nezvezna 437
normalna 315
Poissonova 315,443
Studentova t-421
porazdelitvena funkcija
za normalno porazdelitev 319,321
za standardizirano normalno porazdelitev 418
za χ^2
poskus 233
2^p 688
dvofaktorski 684
faktorski 683
frakcionalni 689
načrtovanje 657
poskusna enota 660
poskusno gradivo 660
posledica, matrika 505
posteriorna analiza 517
poostrena kontrola 637

postopek 660
posredno opazovanje 53
potrošni koš 310
pozitivna odvisnost 310
premična baza 179
pravokotnik, graf, prik. 168
preposteriorna analiza 519
prevzem- preverjanje kakovosti 630
preskus
asiciacije z mediano 606
Durbin-Watsonov 828
Neumann von 828
mediana 493
moč 458, 472
predznačeni rangi 494
predznaki 492
sekvence 499
slučajnosti razporeda enot 498
z variancami 502
Tukeyev 669, 675
zamenljivosti količin indeksi 301
zamenljivosti osnove 300
značilnosti 803
preskusni vzorec 471
preskušanje domnev 453
asociacija 605
enostavna domneva 455
korelacijski koeficient 559
mali vzorci 475
posplošenje 471
razlike med parametri 474
odvisnost s χ^2 598
regresijski koeficient 551
strukturni delež 473
veliki vzorci 472
pričakovana izguba 470
pričakovana izhodna kakovost 648, 651
prikazovanje statističnih podatkov 99
grafično 110
prilagoditev normalne porazdelitve 325
priložnostna izguba 506
primerjava 666
vnaprej načrtovana 666
aposteriorne 666, 668
preverjanje kakovosti 629
pri prevzemu 630
procesov 629
skupin 629, 645
proporcionalna razmestitev (stratif. vzorčenje) 385
proučevanje odvisnosti 521
proučevanje, statistično 40
prožnostni koeficient 625
prvi regresijski koeficient 544, 556
prvine načrtovanja poskusov 658
shema 661
psevdo-Lorenzov grafikon 286
publikacijska tabela 107

Q

Q Yule-ov koeficient asociacije 603

R

računska kontrola 71

rajon 78

rang 203

analiza variance 497

kvantilni 205

vezani 585

ranžirna vrsta 203

računalnik, elektronski 97

raven Y_k 723

razčlenitvena števila 138

razmernost znaka 26

razmestitev enot(strat.vzorčenje) 385

razmična časovna vrsta 101, 692

razmična populacija 30

razmik studentiziran q 669

razmik zaupanja 350

raznorodna populacija 31

razpored enot, preskus o slučajnosti 498

razred 81

realna enota 23

regionalna razmestitev enot vzorca 415

regresija linearna 546

regresija multipla 610

regresijska analiza 657

regresijska krivulja 533, 538

regresijska premica 550

regresijski koeficient 544, 556

preskušanje domnev 551

regresijski model

celih potenc 577

linearni 547

transformiranega tipa 569

večparametrski 576

rekurzijski obrazec za nezvezne porazdelitve 448

binomska porazdelitev 449

hipergeometrijska 448

Poissonova 450

recipročni koeficienti 167, 312

relativna števila 137

relativne mere variacije 266

za relativna števila 271

relativni modus 224

relativni odklon 270

reprezentativni indeks cen 308

reziduálna metoda 805

rezultativen znak 28

risanje skal v merilu 114

ročna obdelava 94

S

samoregistracija 55

sekularen trend 727

sekvenca 499

preskus 499

sekvenčni kontrolni načrt 652, 654

sestava 103
sestava ponderov 239
seštevanje statističnih podatkov 91
sestavine dinamike 726
sestavljena domneva 455,464
sestavljena tabela 106
sezonska sestavina 49,728,732,778
sezonska sestavina, dinamična 796
shema za načrt obdelave 89
Sheppardov popravek
centralnih momentov 281
variance 259
signiranje 87
simetrična porazdelitev 196
simetričen dogodek 334
sistem obrazcev 59
sistematična napaka 69
sistematično vzorčenje 400
skala, lestvica 113
skala
aritmetična 113
linearna 113
logaritemska 114,706
risanje v merilu 114
verjetnostna 328
s-karta 634
sklep dvostranski, enostranski 465
skupina operatorjev 412
skupinica 23
skupinice, vzorčenje 395
skupinski indeks 301
skupna varianca 260
slučajne napake 69
slučajnostna spremenljivka 315
izvedena iz normalne porazdelitve 423
normalna porazdelitev dveh 555
slučajnost uporabe postopka 660
slučajnostna števila- tablice 367
generacija z računalnikom 369
slučajnostni bloki, 662,670
analiza variance 673
slučajnostni izbor pri poskusu 658
slučajnostni trenutki, izbor 414
slučajnostni značaj faktorja 681
slučajnost razporeda enot 498
socialno ekonomska statistika 18
sortiranje 94
Spearmanov koeficient korelacije ranga 584
sprejemljiva vhodna kakovost 648
splošni načrt statističnega proučevanja 41
spolni indeks 38
sredina razreda 82
standardizacija ciklov 813
standardna sestava ponderov 241
standardna napaka ocene 542
stalni značaj faktorja 681
stalne cene 722
stalna osnova, indeksi 177
sredine, tehtane z binomskimi koeficienti 809
sredstva opazovanja 58
srednji koeficient dinamike 244
srednje vrednosti 219
odnosi med njimi 246
standardiziran z-odklon 270

standardizirana normalna porazdelitev 319,417
standardizirani pokazovalci 240
standardni odklon 254
povezava z normalno porazdelitvijo 262
standardni pogrešek ocene 348
stanje- pri analizi odločitev 505
statistična gostota 157
statistična tabela 73,104
statistična vrsta 73,100
statistični koeficient 157
grafično prikazovanje 168
statistični okoliš 78
statistično opazovanje 41
statistično poročilo 49
statistično proučevanje 40
stolpec 112, 118
enostavni 118
strukturni 145
več vrst 121
v tabeli 104
v tipkanih poročilih 120
stopinja prostosti 419
stopnja rasti 723
stopnja značilnosti 383
stratificirano vzorčenje 383
stratum 383
strukture 138
strukture 138-
dvojne 142
enostavne 138
grafično prikazovanje 144
figure 157
krogi 149
stolpci 145
tokovi 156
trikotnika 154
večkratna 141
strukturni delež 139
ocenjevanje s časovnim vzorčenjem 411
ocenjevanje s slučajnostnim vzorčenjem 361
preskušanje domnev 473
preskus homogenosti skupin 598
Studentiziran razmik q 669
Studentova porazdelitev 421
stvarna kontrola 71
stvarna opredelitev 44
stvarne frekvence pri kontingenci 591
stvarni indeks 182
stvarni znak 24
grupiranje vrednosti znakov 79
subjektivna ocena za aritmetično sredino 427
subjektivna verjetnost 462
sumarno relativno število 239

Š

šifra 65,87
črkovna 87
številčna 87

šifrant 87
šifriranje 74,87
širina razreda 84
šrafiranje 117
šrafura, kartogram 133
število enot z dano značilnostjo 361,364

T

tabela 99-109
tabela, kontingenčna 588
tabela, korelacijska 528
tabelarna analiza 657
tablice slučajnostnih števil 367
tehtanje 93
tehtano računanje sredin 237-242
tekoča registracija 45
temp rasti 723
teoretična frekvenca (pri kontingenci) 590
teoretična porazdelitev 315,417
teoretska lista 310
terminal 97
točka, kartogram s 133
točka preloma linearne funkcije dohodka 513
točkovna ocena 350
tokovi, strukturni grafikon 156
tolerančni razmik 630
transformacija za r_{xy} 559
transverzalno opazovanje 45
trditev enostranska, dvostranska 336
trend 726-731
  eksponenten 773
  krivuljčen 756
  linearen 760
  metode za določanje 738
    delne sredine 748
    delne vsote 751
    drseče sredine 740
    izbrane točke 746
    najmanjši kvadrati 759
    ortogonalni polinomi 763
    ortogonalni polinomi binomskih funkcij 766
    prostoročna 738
    sredine med najvišjimi in najnižjimi točkami 742
  preskus o značilnosti 736
trenutna časovna vrsta 101,692,719
trikotnik, strukturni 154
trikotniška mreža 116
Tukeyev preskus 675,669
tveganje 335
tveganje, odločitev pri 508
t-porazdelitev 315,317,421

V

V- korelacijski koeficient za asociacijo
variacija, mere 249
varianca 227,254,481
  metode računanja 254

nepojasnjena 541
nepristranska ocena 353
ocena z malim vzorcem 434
pojasnjena 540
preskusi z 481
analiza variance 485
variiranje 28
variacijski razmik 250
varnostno območje 638
velika števila, zakon 37
večkratna struktura 140
večkratni kontrolni načrt 652,654
veliki vzorci 344
velikost vzorca, enostavnega 376
verižni indeks 180,723
verjeten
dogodek 333
odklon 665
verjetnost
aposteriorna 333
apriorna 333
inverzna 144,464
subjektivna 462
za napako I. vrste 461
za napako II. vrste 462
verjetnostna porazdelitev 334, 417
normalna 333
zvezne 417-437
verjetnostni račun 333
verjetnostna skala 328
vezana vzorca- neparametrično preskušanje 608
vezani rang 585
vhodna enota 97
vhodna kakovost 647
neopredeljena 648
nesprejemljiva 648
sprejemljiva 648
vir podatkov 52
sekundaren 53
vir, v tabeli 108
vnaprej načrtovana primerjava 666
vrednost znaka 22,28
vprašalna pola 58
vrsta, statistična 73
vsota rangov, preskus 495
vzorčenje 50,339
časovno 409
kvotno 406
mali vzorci 417
okvir 366
sistematično 400
stratificirano 383
v dveh fazah 404
pri razmerju 405
pri stratifikaciji 405
v skupinicah 395
v dveh stopnjah 403
z omejitvami 383
vzorčni delež 400
vzorec
brez ponavljanja 344
preskusni 471
s ponavljanjem 344
velik 339
veliki, preskušanje domnev 472
vzročna odvisnost 527

W

Wicoxonov preskus predznačenih rangov 494

Y

Yatesov popravek za asociacijo 605

$\bar{y}$-me-R karta 635

$\bar{y}$-s karta 632,638

Yuleov koeficient asociacije 603

Z

Z-diagram 710

z- odklon 270

zakon o velikih številih 37,334

zaokroževanje podatkov 75,83

zbirna vrsta, stolpec 107

značaj faktorjev, slučajnosten, stalen 681

znak 24-, 58-

znak, vrednost znaka, 25

značilnost

stopnja 460

razlik 471

vzorčnih izrazov 459

zvezna populacija 344

Ž

žrtvovanje informacij 689

NARODNA IN UNIVERZITETNA KNJIŽNICA
COBISS
000000095125

Narodna in univerzitetna knjižnica Ljubljana

