NARODNA IN UNIVERZITETNA KNJIŽNICA
DS
II 130 281
095407046
COBISS ©

ZAVOD ZA STATISTIKO IN EVIDENCO LRS

BLEJEC MARIJAN

TEORIJA STATISTIKE

GRADIVO ZA STROKOVNE IZPITE ZA NAZIV
„NIŽJI STATISTIK"

LJUBLJANA, 1951

I 130281

II 130231

Narodna in univerzitetna knjižnica v Ljubljani

U 7046/1954

V S E B I N A

		Stran
1.	Uvod in razvoj statistike	1
11.	Nastanek in razvoj	1
12.	Vloga in pomen statistike v socializmu in planskem gospodarstvu	4
13.	Statistika in evidenca	5
14.	Razvoj statistike v Jugoslaviji	6
15.	Različni pomeni besede statistika	8
16.	Pomen matematike v statistiki	8
17.	Glavna področja uporabe statistike	9
18.	Glavne vrste socialno ekonomskih statistik	9
2.	Masovni pojavi, njih opredelitev in proučevanje v statistiki	10
20.	Masovni pojavi	10
21.	Statistična enota	11
22.	Znaki statistične enote	11
23.	Statistična masa	14
24.	Proučevanje masovnih pojavov	15
3.	Statistično opazovanje	18
31.	Program opazovanja	18
32.	Organizacijsko - tehnični del opazovanja	21
33.	Organizacijski plan opazovanja	30
34.	Kontrola statističnega opazovanja	31
4.	Grupiranje in statistične tabele	36
41.	Grupiranje	36
42.	Statistične tabele	42
5.	Urejevanje in osnovna obdelava statističnega gradiva	47
50.	Splošno	47
51.	Predpriprave izvedbe urejevanja	47
52.	Vrste urejevanj	50
53.	Urejevanje glede na vrsto uporabljenih obrazcev	52
54.	Vrste urejevanj v organizacijskem pogledu	54

		Stran
6.	Statistične serije	55
61.	Kaj je statistična serija	55
62.	Delitev statističnih serij po značaju osnovnega znaka serije	55
63.	Pripombe o členih statistične serije	58
64.	Delitev statističnih serij po značaju vrednosti členov	58
65.	Kumulativne serije	59
66.	Frekvenčna distribucija	60
67.	Povzetek	65
7.	Relativna števila	67
70.	Splošno	67
71.	Vrste relativnih števil	69
72.	Strukture ali razčlenitvena števila	69
73.	Statistični koeficienti ali gostote	73
74.	Indeksi - merska števila	79
75.	Povzetek	93
8.	Srednje vrednosti	95
80	Splošno	95
81	Aritmetična sredina	98
82.	Harmonična sredina	125
83.	Geometrična srednja vrednost	128
84.	Medijana	131
85.	Modus	136
86.	Povzetek	139
9.	Časovne serije	141
90.	Splošno	141
91.	Vrste časovnih serij	142
92.	Analiza časovnih serij	156
93.	Vrste gibanj masovnih pojavov	160
94.	Razstavljanje časovnih serij na sestavne dele	165
95.	Povzetek	177

		Stran
10.	Grafično prikazovanje	180
100.	Splošno	180
101.	Stolpci	183
102.	Pravokotniki	188
103.	Krogi	191
104.	Točke	193
105.	Črte - linije	194
106.	Kartogrami	208
107.	Figure	212
108.	Grafična kontrola plana	214
109.	Povzetek	224
11.	Mere variacije	226
110.	Splošno	226
111.	Vrste mer variacije	227
112.	Širina variacije	227
113.	Interquartilna vrednost	228
114.	Povprečen absoluten odklon	235
115.	Povprečen kvadratičen odklon ali varianca in standardna deviacija	237
116.	Relativne mere variacije. Koeficient variacije	252
117.	Povzetek	256
12.	Osnove korelacije	258
120.	Splošno	258
121.	Funkcionalne odvisnosti	258
122.	Korelacijske odvisnosti	260
123.	Korelacijska karta	267
124.	Črta grupnih aritmetičnih sredin	270
125.	Regresijski premici	273
126.	Determinacijski koeficient	283
127.	Korelacijski koeficient	291
128.	Krivuljčna multipla ali večkratna in parcielna ali delna korelacija	293
129.	Povzetek	294

		Stran
13.	Metoda vzorčenja	296
130.	Splošno o vzorčenju	296
131.	Osnovna statistična masa, vzorec in masa vseh vzorcev	300
132.	Slučajni izbor kot osnova vzorčenja	353
133.	Ocenjevanje s pomočjo vzorčenja	312
134.	Vrste vzorčenj	334
135.	Občutljivost vzorčenja za razne napake	339
136.	Povzetek	341

Opomba. Poglavja in odstavki so označeni s številkami po decimalni klasifikaciji. Tabele, slike in obrazci so oštevilčeni znotraj poglavij s tekočimi številkami. Primeri so označeni s črko P, dokazi pa s črko D na levem robu besedila.

1. UVOD IN RAZVOJ STATISTIKE

11. NASTANEK IN RAZVOJ STATISTIKE

11.1 V naravi, družbenem in ekonomskem življenju ne nastopajo posamezni pojavi posamič, temveč v velikem številu, množično. Tako ne obstoji samo en človek, temveč množica ljudi, človeštvo, ne eno samo kmetijsko gospodarstvo, temveč veliko število gospodarstev itd. Pojave, ki nastopajo množično, imenujemo množične, masovne pojave. Človeštvo, katerega metoda dela za napredek je raziskovanje in poznavanje dejstev, je v zgodovini že zgodaj uvidelo važnost opazovanja in raziskave ne samo posameznih, temveč masovnih pojavov. Način opazovanja in raziskovanja masovnih pojavov pa je dokaj različen od opazovanja in raziskovanja individualnih pojavov. Zaradi tega se je v teku zgodovine razvila posebna metoda in znanost o opazovanju masovnih pojavov, ki jo imenujemo statistiko.

11,2 Potreba po najenostavnejšem statističnem podatku, številu istovrstnih stvari, se je v zgodovini pojavila že zelo zgodaj. Iz vojaških in finančno - gospodarskih ozirov je vladar hotel vedeti število prebivalstva, posebno število moškega prebivalstva, ki je sposobno za bojevanje, višino premoženja svojih podložnikov itd. Tako zasledimo prvo primitivno štetje prebivalstva že več ko dva tisoč let pred našim štetjem na Kitajskem. Enako vsebujejo indijski viri iz 12. stoletja pred našim štetjem mnogo podatkov o razmerah v državi. Vse ostale visoko razvite države starega veka, kot Perzija, Grčija in Rimska država, so zaradi upravljanja države zbirali različne podatke (znani cenzusi v Rimu). Način zbiranja teh podatkov pa je bil naravno zelo enostaven.

Za perzijskega kralja Darija (6 st. p.n. štetjem) je n.pr. pri določanju števila vojaštva vsak vojak položil na odrejeno mesto kamen.

11.3 V fevdalnem družbenem redu, kjer je bila zemlja glavno sredstvo gospodarstva, je bila registracija zemljišč glavni interes masovnega opazovanja. Iz te dobe je znana poleg drugih " Knjiga strašne sodbe" angleškega vladarja Viljema Osvajalca. Tu je dan številčni opis posestnih, denarnih in službenih razmer prebivalstva. Ta registracija je bila izvršena z veliko strogostjo. V dobi reformacije je nastal interes za nov podatek: vero. Katoliška cerkev pa je predpisala obvezno registracijo podatkov o smrtih, rojstvih in porokah.

11.4 V kapitalizmu, ki je v ogromni meri razvil proizvodne sile, je masovna proizvodnja nujno zahtevala njeno masovno registriranje. V prvi fazi razvoja kapitalizma je kapitalist uvedel knjigovodstvo, ki je

kot sistematično registriranje vseh poslovnih dejstev podjetniku v začetku zadoščalo. Gospodarstvo pa se je razvijalo, trgovinski promet je porastel, konkurenčna borba se je zaostrila in sicer ne le v merilu podjetnika proti podjetniku, temveč tudi države proti državi. Pojavila se je potreba za poznavanjem poslovanja ne samo lastnega podjetja in ne samo strogo poslovnih podatkov, temveč o vseh gospodarskih in družbenih pojavih. Izvajanje davčnih ukrepov, carinskih predpisov in podobno je bilo nemogoče brez podatkov o prebivalstvu, proizvodnji, blagovnem prometu. Kakor je kapitalistu po utrjenem sistemu vodil knjigovostvo poseben za to usposobljen uslužbenec, knjigovodja, je zaradi vedno večje potrebe po statističnih podatkih tudi država začela ustanavljati statistične urade, katerih naloga je bila organizacija zbiranja in proučevanja statističnih podatkov. Poleg države so tudi posamezna večja podjetja ali družbe morale začeti na statističen način registrirati svoje poslovno dogajanje, ker samo knjigovodstvo ni več zadoščalo.

Z razvojem kapitalizma se je vedno očitneje izražal razredni značaj statistike. Statistika je bila v službi vladajočega razreda in mu je posredno služila pri tlačenju delovnega ljudstva. Zaradi tega je prišlo do protislovij, ki so dovedla statistiko v kapitalizmu v krizo. Ta protislovja so izvirala predvsem iz naslednjih treh dejstev:

a) Podatki o poslovnih dogodkih posameznih podjetij so poslovna tajnost, ki bi mogla koristiti drugemu podjetniku ali državi in s tem škodovati temu podjetju. Zaradi tega te podatke skriva; če pa jih mora dati (državi), jih potvarja. Zato je statistična slika gospodarstva nepopolna in potvorjena.

b) Zaradi fiskalnega namena ima delovno ljudstvo nezaupanje do statistike, se skuša izogniti dajanju podatkov ali pa daje netočne in napačne podatke.

c) Zbrani statistični podatki pa so, čeprav netočni in nepopolni, pokazali protislovja v kapitalizmu, ki se očituje v bedi tlačenega razreda in kopičenju dobrin pri vladajočem razredu. Zato se je vladajoči razred kljub temu, da je statistiko nujno potreboval, njenih rezultatov bal, ker so odkrivali dejansko stanje in bi mogli tlačenemu razredu služiti kot orožje v borbi za svoje pravice. Prav zato je bil kapitalist prisiljen statistične podatke skrivati ali potvarjati.

Protislovje v statistiki se izraža tudi v tem, da so metode, ki jih je [illegible] teorija, pogosto zašle v goli formalizem, posebno še glede na dostikrat slabo kvaliteto osnovnih podatkov.

11.5 Vzporedno s statistiko kot prakso, se je razvijala statistika kot znanost.

O statistiki kot znanosti moremo začeti govoriti v 17. stoletju. Skoro istočasno sta se začeli razvijati dve različni smeri znanstvenega proučevanja

masovnih pojavov na osnovi številčnih podatkov "politična aritmetika" v Angliji in "nauk o državi" v Nemčiji.

Politični aritmetiki, ki so dobili ime po delu Pettya "Political arithmetic" so prvi pričeli na osnovi številčnih podatkov ne samo opisovati masovne pojave, temveč tudi proučevati in analizirati zakonitost teh pojavov. Njihov namen je bil predvsem znanstveno proučevanje in šele v drugi vrsti praktična svrha njihovih izsledkov.

Prvi politični aritmetik je Graunt, ki je l. 1660 proučeval splošne zakone rojstev, smrti itd. na osnovi podatkov o rojstvih in smrtih v Londonu. Po tem uspehu si je statistika vedno hitreje utirala pot kot metoda znanstvenega proučevanja masovnih pojavov in sicer ne samo socialnih, temveč tudi ekonomskih. Zaradi tega se je politična aritmetika razvijala v dveh smereh: v socialni, t.j. v ožjem smislu demografski in ekonomski smeri.

Utemeljitelj prve (socialne) smeri, ki je prešla v tako imenovano biometrično šolo, je bil Graunt. Halley je sestavil na osnovi izsledkov te šole prve tablice umrljivosti.

Utemeljitelj druge (ekonomske) smeri pa je bil Petty, ki je skušal s statističnimi podatki analizirati in primerjati ekonomsko moč Anglije in Francije.

V nadaljnjem razvoju je posebno uspevala prva smer, medtem ko so na ekonomskih področjih pričeli uporabljati statistične podatke opreznejе. Politična aritmetika se je razvila predvsem v Angliji, Franciji in Nizozemski, manj v drugih državah.

Skoraj istočasno s pričetki politične aritmetike v Angliji, se je pojavil v Nemčiji " nauk o državi ".

L. 1660 je pričel Conring profesor na univerzi v Helmstadtu, s sistematičnim opisovanjem življenja države. Z njim je statistika dobila značaj univerzitetne znanosti.

Achenwal je nadaljeval s tradicijo Conringa in je prvi pričel uporabljati ime statistika.

Sredi 18. stoletja je nastala ideja primerjalnega podajanja podatkov v pregledni obliki tabel. Ustanovitelj tako imenovane "tabelarne statistike" je bil Achenwallov sodobnik Dar Anchersen. Tabelarni statistiki so poudarjali veliko važnost primerjanja statističnih podatkov, kar jim je posebno omogočala metoda prikazovanja podatkov v tabelah.

V statistiki do nastopa Queteleta sredi 19. stoletja ni bilo enotnosti in jasnih pojmov. Obvladovala so različna mišljenja v pogledu statistike. Vendar se je že videlo, da je z naraščajočim razvojem kapitalizma nujno povezan tudi širok razvoj statistike. V tej dobi ima za razvoj

statistiki velike zasluge Belgijec Quetelet. S svojim teoretičnim in praktičnim delom je dal statistiki [illegible] in praksi zaokrožen sistem. Poglobil je metode proučevanja zakonitosti masovnih pojavov. Njegova ideja je "povprečen človek" v fizičnem, ekonomskem in moralnem pogledu. Statistiko je nazval z značilnim imenom "socialna fizika". Njegovo delo pa ni bilo omejeno samo na Belgijo, temveč je bil vnet delavec za delo statistike v mednarodnem merilu, organizator prvih mednarodnih statističnih kongresov itd. Njegova ideja povprečnega človeka je po izdaji "socialne fizike"zaradi napačnega tolmačenja njegovih predpostavk izzvala ostro kritiko s strani antiquetelistov.

Po Quetelotu se je praktična in znanstvena statistika izredno hitro razvijala. Posebno se je razvila matematična statistika. Med znanstveniki, ki so delali v tem času na teoriji statistike so bili n.pr. . Laplace, Fourier, Poisson (dal zakon o velikih številih v matematični obliki), Pearson (uvedel v statistiko nove načine raziskave), Czuber, Yule, Fisher in drugi.

Važnost statistike ket sredstvo proučevanja socialno ekonomskih pojavov, so pravilno spoznali klasiki marksizma in leninizma. Dela Marksa so polna statističnih podatkov, na osnovi katerih analizira socialno ekonomski položaj proletariata.

Enako je Lenin svoje ugotovitve dosledno osnoval na statističnih podatkih. Podatke buržoazne statistike je preanaliziral, da je mogel podati objektivno sliko stanja in razvoja. Spoznal je, da sta statistika in evidenca nujno orodje in orožje socializma, da sta važen sestavni del upravljanja socialistične države.

12. VLOGA IN POMEN STATISTIKE V SOCIALIZMU IN PLANSKEM GOSPODARSTVU

Značaj in vloga statistike se v socializmu v osnovi menja. S prehodom oblasti v roke ljudstva in s socializacijo gospodarskih podjetij avtomatično izginejo protislovja, ki so v kapitalizmu dušila razvoj statistike. Ni več razloga za prikrivanje poslovnih podatkov, saj služijo vsa podjetja istemu cilju. Ker je država istovetna z ljudstvom, ni razloga, da ljudske množice ne bi imele zaupanja v državo, niti, da bi se država bala rezultatov statistike. Saj gre za to, da država pozna dejansko stanje, kar ji služi za pravilno upravljanje in vodstvo. Druga osnovna črta statistike in evidence v socializmu je, da postaneta osnovni sredstvi usmerjenega planskega dogajanja v družbi in gospodarstvu, sredstvi obvladanja stihije. Zato si socialistične družbe ni mogoče zamisliti brez splošne in kompleksne evidence in statistike.

Naloga evidence in statistike v planskem gospodarstvu je:

a) Številčni podatki o družbi in gospodarstvu so neobhodno potrebni pri sestavljanju planov.

b) Evidenca in statistika morata tekoče spremljati izpopolnjevanje planov in sta torej sredstvi kontrole plana in dela podjetij.

c) Evidenca in statistika pomagata analizirati iz-polnjenje plana in iskati faktorje, ki so pozitivno in negativno vplivali na njegovo realizacijo, kar moramo upoštevati pri sestavi novih planov.

d) Statistika mora podati popolno sliko socialno-ekonomskega dogajanja, tudi onega, ki ni zajet s planom.

13. STATISTIKA IN EVIDENCA:

V splošnem razvoju opazovanja masovnih pojavov, ki smo ga doslej podali, je podan razvoj obeh načinov opazovanja masovnih pojavov: evidence in statistike. Povezava evidence in statistike je namreč tako tesna in se prepletata, da je težko povleči ostro mejo med eno in drugo. Obe imata za predmet opazovanj masovnih pojavov v družbi in gospodarstva. Vendar se po načinu in namenu opazovanja bistveno razlikujeta.

13.1 Evidenca

je sistematično registriranje vseh pojavov, ki so v zvezi z določeno družbeno ali gospodarsko delavnostjo oz. s podjetjem ter je njen sestavni del delovanja te delavnosti ali podjetja. Brez evidence si je nemogoče zamisliti poslovanje katerega koli podjetja. Osnovne značilnosti evidence so naslednje: Vsaka evidenca je vezana s svojim zbiranjem in urejevanjem na točno določen okvir in na točno določen način registriranja, ki je večkrat določen z uredbo. Druga važna lastnost evidence je, da registrira vse posamezne pojave z vsemi značilnostmi in na točno določenem mestu.

Ogromno področje dela evidence je dovedlo v naši stvarnosti do izoblikovanja treh vrst evidence:

a) Operativna evidenca je namenjena operativnemu vodstvu podjetij, direkcij. itd. Njen namen je tekoča kontrola dogajanja in plana, zaradi takojšnjega operativnega posega v proces v primeru slabega dela. V njej se uporabljajo elementarne metode obdelave in je njena osnovna značilnost ažurnost.

b) Planska evidenca registrira izpolnjenje plana za daljše roke, toda ne operativnega plana, temveč izvršenje tako imenovanega osnovnega plana, ki obsega širše - osnovne grupe. Z njo globlje analiziramo izvršenje plana, raziskujemo proporce in tendence razvoja. Planska evidenca se ne poslužuje samo primitivnih metod masovnega opazovanja, temveč uporablja že metode analize masovnih pojavov.

c) Knjigovodstvena evidenca evidentira izvore, sestav in gibanje sredstev podjetja in sicer v denarnem iznosu. Knjigovodstvena evidenca je torej vezana na podjetje in na ona dogajanja, ki jih moremo denarno oceniti.

Njena osnovna značilnost je bilančna metoda.

S knjigovodstvom kontroliramo izpolnjenje plana z denarnimi pokazatelji.

13.2 Za statistiko pa ni osnovne važnosti vsak posamezni element s svojimi značilnostmi, temveč skupnost vseh elementov in značilnosti te skupnosti. Za statistiko je značilno opisovanje skupnosti masovnih pojavov in iskanje splošnih odnosov in zakonitosti med njimi.

Ako hočemo to doseči, pa je potrebno opazovati individualne elemente. V tej fazi je delo evidentičarja in statističarja enako. Vendar je razlika ta, da se delo za evidentičarja tu konča, medtem ko je za statističarja to šele osnova dela. Od tu dalje zanj posamezni pojav ni važen in proučuje le masovnost pojavljanja.

Medtem, ko je za evidenco značilna stalna oblika registriranja in urejevanja podatkov, si statistika za vsako opazovanje ustvari poseben sistem dela,ki je sicer v skladu s principi statističnega opazovanja, vendar prilagojen akciji,ki jo vršimo.

Enako ni v statistiki neobhodno potrebno registriranje vseh pojavov, ker je možno zaradi zakonitosti in lastnosti masovnega opazovanja odkriti zveze in zakonitosti že s tem,da opazujemo samo del elementov masovnega pojava.

Kljub tem razlikam je povezava med evidenco in statistiko tesna. Evidenca v sektorjih, ki jih obsega, nadomešča statistiki prvo fazo statističnega dela, zbiranje materiala. Ker je naloga statistike,dati številčno dokumentacijo vsega družbeno-ekonomskega dogajanja, je za statistiko delo tem lažje, čim širši in kompleksnejši obseg ima evidenca, ker ima s tem podano že gradivo za svoje delo. Za sektorje, kjer taka evidenca ne obstoja, pa je navezana na samostojno zbiranje. Na drugi strani pa more evidenca s tem, da se poslužuje statističnih metod, res znanstveno analizirati svoje podatke.

14. RAZVOJ STATISTIKE V JUGOSLAVIJI

Statistična služba v bivši Jugoslaviji ni imela značaja dobro organizirane statistične službe. Organizirane statistike v pravem smislu besede ni bilo. Vršile so se sicer nekatere akcije (popisi prebivalstva, kmetijska poročevalska služba itd.) vendar so vsa ta dela vršili posamezni organi sami zase, brez težnje, dati sliko celotnega ekonomskega stanja in koordinirati statistično službo.Oblika in značaj kapitalizma v bivši Jugoslaviji je bil tak, da dobr organizirane statistike niti ni potreboval.

Statistična služba v FLRJ ni zrasla na dediščini stare statistike bivše Jugoslavije. Njene korenine moramo iskati v narodno-osvobodilni borbi,

kjer je bila statistika nujno praktično orodje organiziranja zaledja in pomoči borcem. To se je v začetku vršilo po potrebiin posamič. Z višjo stopnjo organiziranosti in večjega obsega osvobojenega ozemlja, pa je dobila statistična služba organizacijsko obliko v ustanovitvi "Državnega statističnega urada FFJ 14.XII. 1944. Naloga tega urada je bila organizacija statistične službe v republikah in na terenu, zbiranje, koordinacija statističnih podatkov in služb. Takoj po osvoboditvi so bili ustanovljeni republiški uradi. Podatki statističnih uradov v tej fazi so bili prepotrebna osnova obnove države.

Z zakonom o " Splošnem državnem gospodarskem planu" je postal 4.6. 1946 državni statistični urad organ Zvezne planske komisije. Poleg statistične linije je obstojala tudi resorna statistična linija, ki je opravljala statistične posle resorov, vendar v soglasju s statistično linijo.

29.5.1948. leta je bila dana državnemu statističnemu uradu nova organizacijska oblika, Postal je Zvezni statistični urad z lastnimi republiškimi in okrajnimi organi. V sklopu Planske komisije je ostala le še planska evidenca, ki je imela svojo posebno organizacijsko obliko v zveznom in republiških "Biroih za evidenco". Družbeno- ekonomska evidenca in dokumentacija se je zbirala, oziroma vodila po dveh organih: Državnem statističnem uradu in Biroih za evidenco. Jedro dela statističnega urada je bila predvsem družbena plat, neplanirani pojavi ali organizacijski popisi. Jedro dela Biroev za evidenco pa so bili ekonomski, oziroma planirani pojavi, ki so bili organizacijsko vključeni v sistem poročevalskih služb.

Čeprav je bila naloga Državnega statističnega urada, da zbira in poda sliko popolnega socialno-ekonomskega dogajanja, ter naloge v polni meri izvršil.

Šele z ustanovitvijo "Zavoda za statistiko in evidenco", ki je bil ustanovljen 64951 kot del Gospodarskega sveta, je bila statistični službi dana organizacijska oblika, ki ustreza njenim nalogam. Z ustanovitvijo enotnega organa, ki po ukinitvi biroev za evidenco dejansko združuje dokumentacijo vse socialno ekonomske dejavnosti v državi, je dana v polni meri možnost razmaha statistične službe v FLRJ.

Statistična služba v FLRJ s popisi ali tekočimi poročevalskimi službami zajema vsa področja socialno- ekonomskega življenja. Ti podatki služijo oblasti za plansko vodenje gospodarstva, za kontrolo uspehov na vseh socialno- ekonomskih področjih in znanstveno proučavanje naše stvarnosti. Z borbo za kakovost statistične službe in podatkov daje znanstveno objektivno in pravilno sliko naše stvarnosti.

15. RAZLIČNI POMENI BESEDE STATISTIKA

Beseda statistika izvira iz latinske besede "status" in pomeni stanje, tudi državo. Iz pomena država je izšla beseda statistika, ki je v 18.st. pomenila sprva beseden, kasneje pa številčen opis države. Kasneje je beseda izgubila svoj ožji pomen opisa države in se je prenesla na splošno za označitev metode opazovanja masovnih pojavov.

Danes pomeni beseda statistika naslednje:

1) statistično delo, prakso zbiranja in proučevanje masovnih pojavov. Pri ugotovitvi, da je "Statistika v neki državi slaba", razumemo, da je statistična delavnost v tej državi majhna, nesistematska in da so njene osnove neznanstvene.

2) Rezultate masovnega opazovanja v obliki številčnih pregledov (preglede publikacije),

3) Organe državne uprave in vse, ki se bavijo z zbiranjem, obdelavo in proučevanjem statističnih podatkov;

4) Statistiko kot znanost in teorijo o proučevanju masovnih pojavov. Ta teorija je zrasla tekom vsega razvoja opazovanja masovnih pojavov in se še vedno razvija.

Osnovne črte, ki označujejo statistično proučevanje so:

a) Predmet statističnega proučevanja so masovni pojavi. Statistika dela zaključke o masi, ne pa o posameznih pojavih. Ako hočemo proučevati strukturo in odnose v zemljiški posesti, moramo opazovati veliko število gospodarstev, ne pa samo eno.

b) Rezultati statističnega proučevanja so številčni zaključki o statistični masi. Način izražanja v statistiki je kvantitativen, torej količinski.

c) Namen in cilj statističnega proučevanja je, s pomočjo številčnega izražanja proučevati kvalitativne- kakovostne odnose in zakonitosti v masovnih pojavih. Ako proučujemo n.pr. vpliv kakega agrotehničnoga ukrepa, je ta vpliv razviden iz primerjave hektarskega donosa s hektarskim donosom, ako tega ukrepa nismo izvedli.

16. POMEN MATEMATIKE V STATISTIKI

Ker podaja statistika svoje rezultate številčno, sledi, da v svojem delu nujno naleti na stik z matematiko vsaj s štirimi računskimi operacijami in preštevanjem. V začetku razvoja statistike je dejansko tudi obstojala samo ta zveza. Z razvojem statistike kot znanosti pa je uvedba novih metod jo zahtevala kot pomožno orodje matematiko. Ta povezava je v zgodovini razvidna že iz imen politična aritmetika in socialna fizika. Moderna teorija statistike je s svojimi metodami tako tesno

povezana z matematiko, da je razumevanje teh metod možno le z dobrim poznavanjem matematike in se je od splošne teorije statistike odcepila celo posebna veja "matematična statistika" katera uporablja v svojih metodah višjo analizo, predvsem verjetnostni račun, na katerem temelji vsa zgradba delnega opazovanja po metodi vzorčenja. Drug stik statistike z matematiko oz. geometrijo pa se pojavi pri grafičnem prikazovanju, kjer uporabljamo lastnosti geometričnih elementov za prikazovanje statističnih podatkov.

17. GLAVNA PODROČJA UPORABE STATISTIKE

Statistična znanost je nastala iz prakse, iz potrebe po opazovanju družbeno gospodarskih pojavov. Zato se je statistika tudi razvojno naprej uporabljala in razvijala pri proučevanju socialno-ekonomskih pojavov. Kasneje se je ugotovilo, da je možno metode, ki služijo za opazovanje masovnih pojavov v družbi in gospodarstvu, uporabiti tudi na drugih področjih, kjer nastopajo masovni pojavi. S tem je postala statistika metoda raziskovanja v biologiji, meteorologiji, fiziki, medicini, tehniki in drugje. S tem se je pospešil razvoj teh znanosti, odprle nove metode raziskovanja. Obenem pa se je potrdila pravilnost statističnih metod in razvile so se nove, kakršne je zahtevala posebnost posameznih vej znanosti.

18. GLAVNE VRSTE SOCIALNO-EKONOMSKIH STATISTIK

Kljub temu, da si je statistika utrla pot kot raziskovalna metoda v najrazličnejša področja, je vendar ostala predvsem metoda proučevanja socialno- ekonomskih pojavov. Zaradi obširnosti in posebnosti posameznih večjih področij socialno- ekonomskih pojavov so se poleg splošnih metod masovnega opazovanja, ki veljajo za proučevanje vseh masovnih pojavov, in ki so obsežene v splošni teoriji in metodi ter matematični statistiki, za posamezne veje socialno- ekonomskega življenja razvile posebne metode opazovanja.

Najvažnejše in najbolj izdelane so iz ekonomskega področja naslednje statistike:

Industrijska statistika, ki obsega metode proučevanja masovnih pojavov v industriji.

Poljedelska statistika, ki zahteva posebne metode proučevanja, v kmetijskih raziskavanjih.

Statistika blagovnega prometa (trgovine in gospodarstva).

Prometna statistika.

Iz socialnega področja pa so predvsem važne:

Demografska statistika, ki spremlja vse pojave v zvezi s prebivalstvom. Demografska statistika je ena izmed najstarejših vej statistike in tvori osnovno demografijo.

Zdravstvena statistike, ki obravnava vse masovne pojave s področja ljudskega zdravstva, tako zdravstveno službo kot zdravstveno stanje ljudstva.

Kulturno- prosvetna statistika obvlada obširno področje kulture in prosvete.

2. MASOVNI POJAVI, NJIH OPREDELITEV in PROUČEVANJE V STATISTIKI

20. MASOVNI POJAVI

Predno preidemo na samo statistično opazovanje, moramo podrobneje spoznati značaj masovnih pojavov.

Socialno-ekonomsko življenje in priroda sestoji iz nepregledne množice predmetov, oseb, dogodkov, stvari itd., ki so med seboj enaki, podobni, sorodni, istovrstni ali raznovrstni. Ti se med seboj povezujejo in po določenih zakonitostih vplivajo drug na drugega. Vso to množico predmetov, oseb, dogodkov itd. bomo do nadaljnjega označili z enotnim nazivom pojav.

Posamezen pojav je označen, oziroma opisan s nebroj značilnostmi, ki so lastne temu pojavu. Neka oseba je n.pr. moškega spola, stanuje v Ljubljani, stara je 28 let, po zaposlitvi ključavničar itd. Poleg teh značilnosti ima ta oseba še nebroj drugih, ki jo karakterizirajo. Z obravnavanjem posameznih pojavov bi ne mogli dobiti pregleda nad vsemi, ker jih ogromna množica. Ker določeni sorodni pojavi, ki imajo osnovne značilnosti skupne v času in prostoru množično nastopajo, imenujemo take pojave množični ali masovne pojave. Tako je prebivalstvo FLRJ 15. marca 1948 množičen pojav, ker moremo najti veliko število pojavov - oseb, ki so bile 15. marca 1948 prebivalci FLRJ. Enako so množični pojavi smrt človeka, ker moremo v času in prostoru najti mnogo dogodkov, ki so označeni kot smrt človeka itd. Statistika združuje istovrstne oziroma podobne pojave v kolektive in kolektivno opazuje skupnost istovrstnih pojavov in napravlja zaključke o celotni skupnosti, ne pa o poedinih pojavih.

Skupnost sorodnih enot, ki predstavljajo masovni pojav statistika določi s tem, da postavi osnovne značilnosti, ki jih mora posamezen pojav izpolnjevati, da spada v kolektiv. To skupnost imenujemo v statistiki statistično maso, posamezen pojav statistične mase statistično enoto, vsako značilnost statistične enote pa statistični znak. Vsak znak zavzame pri dani enoti neko določeno vrednost. Te vrednosti imenujemo vrednosti znaka.

Statistične znake, v katerih je masa opredeljena, imenujemo opredeljujoče znake.

Osnovne značilnosti, ki jih mora enota izpolnjevati, da je enota dane statistične mase so opredeljujoči pogoji statistične mase.

Ako vzamemo kot statistično maso "delavstvo zaposleno v industriji na dan 1.jan. 1950 v LRS" so značilnosti: delavec, zaposlen v industriji, dne 1. jan. 1950, LRS opredeljujoči pogoji, posamezen delavec, ki tem pogojem zadošča je statistična enota, starost, stopnja šolske izobrazbe, število otrok itd. so statistični znaki enote, delavec, starost 28 let, dokončana strokovna šola, 1 otrok itd. pa so vrednosti znakov, ki jih posamezna enota - delavec ima.

Vse enote statistične mase se vjemajo v vrednostih opredeljujočih pogojev in so s temi označene vse enote mase.

Opredeljujoči pogoji razdele vse pojave na dve skupini: Skupino, ki tem pogojem zadošča, in drugo, ki tem pogojem ne zadošča. Iz vseh pojavov torej z opredeljujočimi pogoji izdvojimo statistično maso, ki je v nadaljnjem predmet našega proučevanja.

21. STATISTIČNA ENOTA

Statistična enota more biti vsak pojav, ki v prostoru in času masovno nastopa in je predmet socialno - ekonomskega ali drugega znanstvenega proučevanja.

Statistična enota more torej biti:

a) oseba (delavec v Ljubljani na dan 15. marca 1948)
b) žival (govedo v OLO na dan 15. januarja 1950)
c) stvar (poslovna zgradba na dan 31. maja 1949)
d) pravna tvorba (zadruga v LRS na dan 5. jan. 1948)
e) administrativno - politična enota (KLO v LRS)
f) gospodarska tvorba (kmetijsko gospodarstvo v LRS na dan 1.julija 1950)
g) dogodek (smrtni primer v l. 1949 v LRS)
h) poizkus (poizkusna parcela, poizkusno cepljenje miši itd.), itd.

S formalne strani je važna delitev enot na :

a) realne enote, t.j. take, ki v času obstojajo (oseba, gospodarstvo)
b) dogodke, t.j. take, ki se v času dogode trenutno (smrt, nesreča)

22. ZNAKI STATISTIČNE ENOTE

Znaki dajo statistični enoti in masi vsebino, zato je poznavanje problematike znakov zelo važno. Kot znak more nastopiti vsaka značilnost statistične enote. Iz ogromnega števila možnih znakov, ki označujejo enote, pa izberemo samo one znake, ki so važni pri proučevanju določenega

vprašanja. Zato je izbira vsebinsko pravilnih znakov zelo važna. Pri popisu prebivalstva nas n.pr. zanima spol, starost, zaposlitev, ne pa teža, ki bi bila pri drugem problemu morda važna.

22.1 Delitev znakov

.11 Po vsebini delimo statistične znake na tri dele:

a) krajevne ali geografske znake, ki povedo, bodisi kje se enota nahaja, ali se je nahajala v zanjo značilnem času (n.pr. kraj rojstva, kraj stalnega bivališča, itd.)

b) časovne znake, ki označujejo čas, kdaj je bila enota opazovana ali kdaj se je dogodil zanjo značilen dogodek (n.pr. čas rojstva, smrti).

c) stvarni znaki so vsi ostali (n.pr. spol, starost, sektor lastništva, velikost posestva, zaposlitev itd.). Ta skupina znakov je najbolj obširna in po vsebini najbolj pestra.

.12 Po tem, ali izražamo vrednost znaka z besedami oz. opisom ali številčno, delimo znake na:

a) atributivne (spol, sektor lastništva, zaposlitev itd.)

b) numerične (starost, število otrok, višina plače itd.)

Krajevni in del stvarnih znakov je atributivnega značaja, časovni in del stvarnih znakov pa numeričnega značaja.

.13 Vrednost numeričnih znakov moremo ponazoriti na številčni premici. Pri nekaterih znakih more znak zavzeti samo določene-običajno cele - vrednosti na intervalu številčne premice (n.pr. število otrok more biti 0, 1, 2, 3 itd.). Take znake imenujemo nezvezne ali s tujim izrazom diskontinuirne znake. Numerične znake, ki morejo na intervalu številčne premice zavzeti vse vrednosti, pa imenujemo zvezne, kontinuirne. Zvezni znaki so časovni znaki, kjer more zavzeti znak vsak časovni moment in nekateri stvarno numerični znaki (n.pr. starost, višina, teža itd.).

Gornja vsebinska delitev znakov je najvažnejša in ima odraz v vseh nadaljnjih fazah statističnega opazovanja in proučevanja. Po istem sistemu se dele opredeljujoči pogoji mase, grupe, statistične serije itd.

.14 Za statistično obdelavo, ki sestoji iz preštevanja in seštevanja vrednosti znakov, enot statistične mase, je važen sortirni in adirni značaj znakov. Sortirni značaj imajo vsi statistični znaki, kajti po vseh znakih moremo prešteti, koliko enot ima posamezno vrednost znaka, n. pr. spol(350 moških in 370 žensk)medtem ko imajo adirni značaj, t.j. možnost seštevanja vrednosti znakov samo numerični, in še od teh samo nekateri. Brez smisla je namreč vsota starosti vseh ljudi v krajevnih ljudskih odborih ali vsota vseh cen na trgu v Ljubljani, medtem ko ima velik

vsebinski smisel vsota njivskih površin vseh gospodarstev v LR Sloveniji in podobno. Znak "velikost njiv" ima torej adirni značaj, ker moremo njih vrednosti seštevati. Seveda pa ima ta znak tudi sortirni značaj, ker moremo prešteti, koliko enot - gospodarstev ima določeno velikost njiv po velikostnih skupinah.

22.15 Ako opazujemo n.pr. donos pšenice, so znaki tega masovnega pojava: velikost njive, način sejanja, lega njive, gnojenje, višina donosa itd. Ako proučimo medsebojni odnos teh znakov, vidimo, da je znak "višina donosa" odvisen od velikosti njive, načina sejanja, lege njive, gnojenja itd. "Višina donosa" je torej rezultat teh znakov. Zaradi tega imenujemo "višino donosa" rezultativen znak. Ker so velikost njive, lega njive, gnojenje itd. faktorji, ki vplivajo na donos, pa imenujemo te znake faktorielne znake. Pri večini masovnih pojavov moremo znake razdeliti na faktorielne in rezultativne. Ta razdelitev je vsebinsko izredno važna.

22.2 Variacija vrednosti znakov. Znaki statističnih enot ne zavzemajo pri pri vsaki enoti statistične mase iste vrednosti, temveč morejo zavzeti različne vrednosti. Pravimo, da vrednosti znaka variraje, statistični znaki pa so torej variabilni. Ako proučujemo n.pr. pri statistični masi "prebivalstvo Ljubljane", znak starost, vidimo, da ima vsaka oseba različno starost, različno zaposlitev itd. Seveda morejo imeti posamezne enote med seboj enake vrednosti, vendar te ni v nasprotju z značajem variabilnosti znakov.

Variabilnost je temeljna lastnost statističnih znakov in jo opazujemo pri vseh znakih. Vendar je variabilnost numeričnih znakov posebno važna in jo moremo, kot bomo videli kasneje, številčno izraziti. Proučevanje variabilnosti numeričnih znakov predstavlja zaradi svojo posebnosti važen del statistične analize.

Teoretično more statistični znak zavzeti majhno število vrednosti, veliko, vendar omejeno, ali neomejeno število vrednosti. Ako vzamemo kot primer znak "spol", more zavzeti le dve vrednosti, in sicer moški in ženski. Znak "število otrok" more zavzeti le omejeno število vredhosti 0, 1, 2, 3 itd., enako znak "zaposlitev" n.pr. ključavničar, delavec, uradnik, profesor itd. Zvezne numerične znake moremo zavzeti na danem [illegible] neomejeno število vrednosti. V razdobju enega leta imamo neomejeno število možnih trenutkov. Enako je z znakom "starost" itd.

Katere izmed možnih vrednosti znakov in koliko enot zavzema posamezne vrednosti, pa je odvisno od opredeljujočih pogojev dane statistične mase in mase same. Ako vzamemo kot statistično maso enosobna kletna stanovanja, so najemnine v splošnem drugačne, kot, če opazujemo najemnine trosobnih stanovanj v prvem nadstropju

23. STATISTIČNA MASA

Ko poznamo statistične enote in znake, moremo označiti statistično maso kot skupnost statističnih enot, ki izpolnjujejo opredeljujoče pogoje, s katerimi je statistična masa določena. Enote statistične mase so torej v smislu opredeljujočih pogojev istovrstno. Statistična masa je torej določena z značilnostmi, ki so skupne vsem enotam dane mase t.j. z opredeljujočimi pogoji.

Statistično maso opredelimo s tem, da navedemo, katere vrednosti opredeljujočih znakov morajo enote imeti, da so enote statistične mase.

23.1 Statistično maso je treba opredeliti : krajevno, časovno in stvarno. Krajevno opredelimo statistično maso s tem, da določimo področje , na katerem mora enota obstajati.

Časovno pa opredelimo statistične mase različno. Če gre za maso realnih enot (n. pr. industrijsko podjetje, prebivalstvo itd.)jo opredelimo časovno z momentom (n.pr. industrijska podjetja na dan 1.jan.1949, prebivalstvo v noči med 15. in 16. marcem 1948). Ker statistične mase realnih enot opredelimo z momentom, jih imenujemo momentne mase. Če pa gre za maso dogodkov, pa z momentno opredelitvijo dejansko ne bi zajeli nobene ali malo enot. Zato mase dogodkov opredeljujemo z intervalom, mase pa imenujemo intervalne mase. (rojstva v letu 1949)

Stvarna opredelitev statistične mase je najtežja in zahteva točno poznavanje proučevanega področja.

23.2 Statistične mase delimo po obsegu na končne in neskončne.

Končna masa je n.pr. skupnost vseh kmetijskih gospodarstev v LRS po stanju na dan 15. maja 1947,končna zato, ker ima končno število enot. Neskončna masa pa je n.pr. skupnost vseh možnih poizkusov pod danimi pogoji, neskončna zato, ker je število enot neomejeno. Neskončno maso moremo proučevati samo z opazovanjem končnega dela enot celotne mase, ker ne moremo nikdar izčrpati celotne mase.

23.3 S stališča analitičnega proučevanja masovnih pojavov, so zelo važne homogene statistične mase. Homogene mase so tiste, ki so opredeljene po znakih (faktorielnih), ki vplivajo na vrednost znakov, ki jih proučujemo. Variacija vrednosti proučevanih znakov (rezultativnih) je pri homogenih masah zaradi tega razmeroma majhna, ker izvira samo iz slučajnih vzrokov. Znanstveno proučevanje masovnih pojavov stremi vedno za tem, da proučuje homogene mase. Ako pa imamo opravka z nehomogenimi - heterogenimi masami, jih skušamo razdeliti v delne homogene mase.

23.4 Ako opredelitvi statistične mase dodamo še en opredeljujoč pogoj dobimo med enotami mase take, ki temu novemu pogoju zadoščajo, druge pa ne. Enote, ki temu novemu pogoju zadoščajo, tvorijo samostojno statistično maso, ki jo imenujemo delno maso prvotne, ker so vse enote delne mase obenem tudi enote glavne mase. Ako proučujemo prebivalstvo v LRS na dan 15. marca 1948, je moško prebivalstvo delna masa celotnega prebivalstva LRS.

24. PROUČEVANJE MASOVNIH POJAVOV.

24.1 Značilnosti posameznih vrst proučevanja in opazovanja.

Vsaka realna znanost temelji na opazovanju in poizkusu. Vendar je značaj teh opazovanj pri posameznih znanostih poseben.

Ako vzamemo kot primer poizkuse v fiziki ali kemiji, moremo pri njih točno določiti vse opredeljujoče pogoje in odstraniti vse vplive, ki bi mogli motiti poizkus. Značilno za tak eksperiment je, da je rezultat poizkusa izvršenega pod enakimi okolnostmi, isti tudi pri večkratni ponovitvi poizkusa. S tem, da izpreminjamo en opredeljujoč pogoj, in opazujemo vpliv te spremembe na rezultat, moremo dognati odvisnost med vzrokom in posledico. Ako vzamemo kot primer prosti pad, je čas padanja vedno isti, ako izpustimo na istem kraju isti predmet iz določene višine na enak način. Zakonitost odvisnosti med dolžino poti in časom padanja pa ugotovimo, ako menjamo višino in pri različnih višinah merimo čas padanja.

Omenili smo, da pri fizikalnem poizkusu, izvedenemu pod enakimi pogoji, pridemo vedno do istega rezultata. To trditev moramo popraviti v toliko, da je to res le v primeru, ako poizkuse ponovimo pod popolnoma enakimi pogoji. Vendar je to izredno težko in odvisi od natančnosti priprav, ki jih pri poizkusu uporabljamo, oziroma od natančnosti izvedbe poizkusa sploh. Rezultati ponovljenih poizkusov so zaradi tega med seboj različni, vendar tem manj, čim bolj moremo ohraniti enake pogoje. Te razlike nastopijo zaradi dodatnih vplivov, ki ne izvirajo iz opredeljujočih pogojev, temveč iz tega, ker ni možno izvršiti poizkusov povsem natančno. Ako prosti pad izvršimo malomarno in ne pazimo na to, da vržemo predmet vedno točno iz iste višine in na enak način, ako ne razpolagamo z natančno uro za določanje časa, so časi padanja ponovljenih poizkusov seveda med seboj različni.

Pri poizkusih v kmetijstvu je obseg pogojev, ki jih ne moremo določiti in motijo izvajanje poizkusa pod določenimi pogoji, znatno večji. Nemožno je določiti celo vrsto pogojev, kot so mikrosestav zemlje, vpliv klime, razlike v legi njiv, razlike v kakovosti semena. Dodatnim vplivom, ki jih pri obstoječih možnostih opredelitve ne moremo odpraviti, pravimo v obeh primerih slučajni vplivi, napakam, ki izvirajo iz teh slučajnih vplivov, pa

slučajne [illegible]. Vpliv teh pogojev se javlja kot rezultat slučajnih vplivov. Vendar moremo pri pojavljanju slučajnih [illegible] opazovati neko zakonitost. Rezultati slučajnih vplivov delujejo namreč tako v smeri povečanja, kot zmanjšanja rezultata, vsota učinkov slučajnih vplivov pa je pri velikem številu poizkusov nič. Zaradi tega je z izračunavanjem tipične količine iz rezultatov obstoječih merjenj možno izločiti slučajne vplive. Lastnost, da izbije tipičnost oz. zakonitost vplivov šele pri opazovanju velikega števila enot, se v statistiki imenuje zakon velikih števil. Zakon velikih števil nas uči, da se pri opazovanju velikega števila pojavov, ono kar je individualno, slučajno, kompenzira, ostane pa ono, kar je tipično, masovno.

Na splošno se proučevanje v fiziki, kemiji, agronomiji in podobnih področjih izvrši tako, da se neposredno izvrše poizkusi pod določenimi opredeljujočimi pogoji. V socialno-ekonomskih in nekaterih drugih znanostih (n.pr. meteorologija, medicina), pa to ni mogoče. Predmet opazovanja ne moremo dobiti s poizkusom, ampak ga je ustvarilo življenje samo in ne za potrebe eksperimentiranja. V tem primeru ne preostane drugega, kot da med množico že obstoječih pojavov poiščemo take, ki zadoščajo opredeljujočim pogojem. Na prvi pogled je razvidno, da je v teh primerih obseg slučajnih vplivov še večji in je tipičnost - zakonitost še težje izluščiti. Vendar zakon velikih števil zagotovi, da se tudi v teh primerih pokaže tipičnost, ako opazujemo zadostno število pojavov. Kot primer vzemimo razmerje med rojstvi dečkov in deklic. Ako opazujemo to razmerje pri posameznem porodu ali za posamezne družine, zakonitost ni vidna, ker prevladujejo individualni vplivi. Ako pa pogledamo vsa rojstva za en matični okoliš, oblast ali LRS za daljše razdobje, vidimo, da se številke vedno bolj bližajo tipičnemu razmerju 106 : 100.

Obseg individualnih vplivov je pri nekaterih socialno-ekonomskih pojavih prevladujoč nad opredeljujočimi določilnimi pogoji. Za te pojave je naloga statistike samo opisovanje, ne pa iskanje zakonitosti.

Statistika torej opisuje in išče zakonitosti masovnih pojavov. To dosežo s tem, da ugotavlja število enot statistične mase (n.pr. število prebivalstva), tvori vsote vrednosti numeričnih znakov (n.pr. skupna površina njiv kmetijskih gospodarstev), proučuje sestav mase po različnih znakih (n.pr. spolno - starostna struktura prebivalstva), išče tipične količine, na splošno rečeno, kvantitativno podaja vplive, povezanosti in zakonitosti masovnih pojavov.

24.2 Etape statističnega proučevanja.

Statistično proučevanje se po značaju dela in vloga razdeli na tri velike dele ali etape:

1. Statistično opazovanje v ožjem smislu
2. Urejevanje in osnovna obdelava statističnega gradiva
3. Analitična obdelava, tolmačenje rezultatov in analiza.

.21 Naloga statističnega opazovanja v ožjem smislu je, da zbere točne in polnoštevilne podatke o vseh enotah statistične mase. To je možno zaradi tega, ker je običajno število enot zelo veliko, samo s planskim, sistematičnim registriranjem vseh enot statistične mase. Ta etapa smiselno sestoji iz dveh delov. Prvi del je točna določitev predmeta in znakov opazovanja in sestava plana zbiranja. Drugi del pa sestoji iz dejanske izvedbe opazovanja oz. zbiranja podatkov.

Značilnost te etape je veliko število sodelavcev in neposreden stik z enotami oziroma onimi, ki morejo dati podatke o enotah. Težišče dela prve etape je organizacijsko- tehničnega značaja. Osnovna zahteva prve etape je popolnost in pravilnost zbranih podatkov, Po izvršeni prvi etapi, so statistični podatki zbrani na obrazcih. Na osnovi izpolnjenih obrazcev je možna obdelava statističnih podatkov.

.22 Osnovna obdelava statističnega gradiva obstoji iz urejevanja statističnih podatkov po danih shemah grupiranja, preštevanja in seštevanja podatkov. Iz nepregledne množice podatkov dobimo z urejevanjem, preštevanjem in seštevanjem podatkov, sliko statistične mase v absolutnih številih. Posebnost te etape je obširen obseg dela, ki pa gamore vršiti le specializiran kader. Ta etapa zaradi svojih posebnih težav traja najdalj časa in je največja ovira, da podatki statističnega opazovanja niso obdelani v določenem času po izvršenem opazovanju. Čas skrajšamo bodisi s tem, da zaposlimo večje število obdelovalcev ali pa z mehaniziranjem obdelave. To etapo, ki je karakterizirana s sistematiziranjem statističnega gradiva, je namreč možno, ako razpolagamo s specialnimi statističnimi stroji, izredno skrajšati z mehaniziranjem.

K hitremu statističnemu in pravilnemu delu te faze pripomore v bistveni meri dobra shema in plan dela.

.23 Absolutni podatki, kot rezultati druge etape, že dajo sliko masovnega pojava in so že sami na sebi velikega praktičnega pomena. Vendar absolutna števila ne dajo še popolnega vpogleda v značilnosti in zakonitosti masovnih pojavov. To dosežemo šele z nadaljnjo obdelavo absolutnih števil, s posebnimi statističnimi metodami (srednje vrednosti, relativne vrednosti, korelacija itd.), ki omogočajo skupno z vsebinsko analizo znanstveno tolmačenje statističnih podatkov.

.24 Vse tri etape, čeprav so po značaju dela med seboj zelo različne pa so vsebinsko med seboj ozko povezane. Uporaba posebnih statističnih metod predpostavlja vsebinsko in tehnično pravilno izvedeno drugo etapo, to pa je možno le, ako je bilo s programskim delom prve etape zagotovljeno, da so bila vprašanja postavljena tako, da je bilo možno pravilno grupiranje itd. Da ne bi bilo neskladnosti tako v vsebini kot po teku posameznih etap dela, je potrebno predhodno sestaviti splošni plan celotnega statističnega proučevanja. Ta plan mora biti vseskozi sestavljen po vidikih končnega cilja statistične akcije. Splošni plan obsega poleg programov in planov vseh treh etap tudi finančni plan celotnega dela.

3. S T A T I S T I Č N O O P A Z O V A N J E

30. Statistično opazovanje je osnova statistišnega proučevanja. Od tega, ali je ta osnova dobro ali slabo izvedena, je odvisna kakovost in vrednost celotnega dela. Izvedba te prve etape je navezana na veliko število ljudi, ki jim ni statistika stalni posel. Poleg tega se običajno opazovana statistična masa hitro spreminja. Zaradi tega je po izvršenem opazovanju izredno težko, včasih celo nemogoče popraviti napake, ki so bile izvršene pri opazovanju. Jasno je razvidno, da moramo zaradi zgornjih razlogov izvedbo opazovanja predhodno dobro proučiti in njegov potek utrditi s premišljenim planom.

Plan statističnega opazovanja je sestavljen iz dveh delov:

Program opazovanja vsebuje elemente, ki povedo, kaj opazujemo. Ti elementi so namen, cilj in predmet opazovanja, ter opredelitev enote, mase in znakov opazovanja.

Organizacijsko - tehnični plan pa določa način in sredstva statističnega opazovanja. K njemu spadajo poleg določitve načina opazovanja in sredstev tudi določitev kraja in časa opazovanja, obseg, pripravljalna dela in kontrola podatkov.

31. PROGRAM OPAZOVANJA

31.1 Namen in cilj opazovanja

Predno postavimo plan statističnega opazovanja, je nujno potrebno, da je znan in jasen namen in cilj samega opazovanja. Jasno in točno postavljen cilj in namen omogoča nedvoumno opredelitev predmeta opazovanja in dobro izvedbo statističnega opazovanja. Ta je pogojena s tem, da s čim manjšim trudom oziroma stroški dosežemo najboljše in najhitrejše rešitev naloge. Ako je cilj opazovanja točno postavljen, more odpasti mnogo nepotrebnega dela. Zaradi nepoznavanja cilja se dostikrat zbira podatke, ki niso v neposredni zvezi s predmetom oziroma problemom, ki ga proučujemo.

Namen in cilj opazovanja običajno podajo oni, ki rezultate opazovanja potrebujejo. V večini primerov statističnih akcij v ožjem smislu so to organi državne oblasti. Zaradi tega je namen in cilj važnejših statističnih akcij dan v posebni uredbi vlade. Ta uredba, ki predstavlja pravno osnovo statistične akcije, vsebuje namen popisa, organe, zagotovi finančna sredstva,

dolžnost dajanja podatkov in višino kazni za one, ki podatkov ne bi hoteli dati.

Ko je cilj opazovanja jasen, preidemo k socialno - ekonomski in stvarni analizi pojavov, ki jih nameravamo opazovati. Ta analiza da jasno sliko vsebine pojavov, ki jih nameravamo opazovati, kar je važno za uspešno planiranje in izvajanje statističnega opazovanja.

31.2 Opredelitev predmeta opazovanja- statistične mase in enote.

Ena izmed osnovnih nalog programa statističnega opazovanja je točna določitev statistične enote, oziroma opredelitev statistične mase. Statistično maso opredelimo s tem, da določimo opredeljujoče pogoje, katerim morajo enote dane statistične mase zadoščati.

.21 Vsebinska opredelitev. Najvažnejša a najtežja je vsebinska opredelitev in opredelitev predmeta opazovanja. Običajno postavimo definicijo opazovanega pojava. Kot enoto opazovanja smatramo pojav, ki zadošča navedenim pogojem. Večina pojavov, ki so predmet statističnega opazovanja je zaradi velike prepletenosti družbenih in gospodarskih pojavov že teoretično težko definirati. Poleg tega pa je treba definicijo in stvarno opredelitev predmeta oz. enot podati tako, da je razumljiva za širši krog popisovalcev. Njih naloga namreč je, da na terenu na osnovi dane definicije odločajo, ali je določen pojav predmet popisa ali ne. Poleg tega pa je točno definiranje predmeta opazovanja potrebno tudi za one, ki se bodo zbranih podatkov posluževali,oziroma jih analizirali.

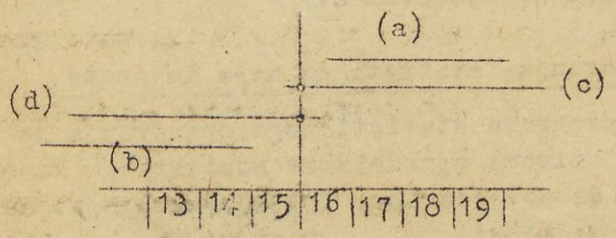
Ker je pri vsebinski opredelitvi včasih težko postaviti nedvoumno definicijo, se včasih poslužujemo enostavnih tehničnih pripomočkov, ki pa so izhod za silo. Med temi pripomočki je najobičajnejša tako imenovana cenzus norma. O cenzus normi govorimo, kadar z vrednostjo enega znaka nadomestimo sicer zapleteno definicijo pojava. Tako so kot kmetijsko gospodarstvo pri popisu zemljiških gospodarstev leta 1947. smatrali posest nad 500 m^2. Leta 1946. pa so bila pri popisu industrijskih podjetij popisana kot industrijska podjetja z več kot 5 delavci. Cenzus norma je v prvem primeru 500 m^2, v drugem pa pet delavcev.

.22 Časovna opredelitev. Časovno opredelimo statistično maso različno, odvisno od tega, kakšno maso obravnavamo: momentno ali intervalno. Že iz samih imen moramo povzeti, da opredelimo momentno maso z momentom, intervalno pa s časovnim intervalom.

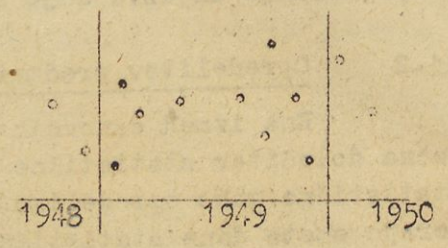
Nedvoumna časovna opredelitev momentne mase je možna le z momentom. (slika 3. 1). Strogo vzeto, so enote mase le pojavi, ki obstojajo v momentu, s

katerim je masa opredeljena. Ta moment imenujemo kritični moment. Pri popisu prebivalstva 1.1948. je bil kritični moment polnoč med 15 in 16 marcem. Otroci, rojeni po tem trenutku (a) in osebe, umrle pred tem trenutkom (b) niso bile enote opazovanja, medtem ko so otroci, rojeni pred kritičnim trenutkom (c) in osebe, umrle po kritičnem trenutku (d) bile predmet popisa.

Sl. 3.1



Sl. 3.2



3.

Pri intervalnih masah (slika 2), kjer je masa opredeljena s časovnim intervalom, so enote vsi dogodki, ki so se izvršili v tem intervalu (.). Dogodki, ki so se izvršili izven tega časa, niso enote mase (o). Kot primer moremo vzeti rojstva v letu 1949.

Pri časovni opredelitvi momentnih mas je važno, kateri trenutek izberemo kot kritičen trenutek. Trenutna slika stanja je zelo odvisna od časa, poleg tega pa je opazovanje v določenih časih težko izvesti. Tu velja pravilo, da gledano z vsebinske plati izberemo kot kritični tisti datum, ob katerem je stanje najbolj normalno oz. odgovarja cilju opazovanja, s tehnične plati pa tak trenutek, ob katerem je najlažje izvesti opazovanje. To je čas, ko opazovane enote geografsko niso preveč razstresene in so v najmanjšem gibanju (n.pr. živina v januarju).

.23 Krajevna opredelitev. Ako hočemo točno opredeliti statistično maso, jo moramo opredeliti tudi prostorno in sicer s teritorijem. Socialno ekonomska statistika ima za osnovo geografske opredelitve navadno administrativno razdelitev, (mrežo republik, oblasti, okrajev, krajev).

31.3 Znaki opazovanja.

.31 Vsebino statističnim enotam dajo znaki, oziroma vrednosti znakov. Kot vemo, more biti vsaka enota označena z ogromnim številom značilnosti. Iz te množice pa izberemo samo one, ki so važni za dano statistično opazovanje. Ta izbira znakov je zelo važna. Upoštevanje nepotrebnih znakov namreč po nepotrebnem obremenjuje opazovanje, izpuščanje važnih znakov pa onemogoča popolno analizo danega problema.

Pri izbiri znakov opazovanja moramo upoštevati naslednja pravila:

a) pri določanju znakov moramo biti v stalnem stiku z onimi, ki podatke potrebujejo

b) potrebna je tudi zveza z onimi, ki bodo dajali podatke (teren, poizkusno opazovanje)

c) izbrati smemo samo znake, ki so v neposredni zvezi z opazovanim pojavom

d) znakov, ki zahtevajo subjektivne, sumljive ali diskretne odgovore , se moramo izogibati, ker nanje ne bomo dobili odgovor - ali pa slabe.

e) ogibati se moramo znakov, ki zahtevajo na terenu zamotana proračunavanja.

f) opazovati smemo le one znake, ki jih nameravamo obdelati (veza opazovanja z obdelavo in analizo).

Znake, ki jih opazujemo pri statističnem opazovanju moremo razdeliti po namenu na tri dele: jedro opazovanja tvorijo znaki, ki dajo vsebino opazovanja. Poleg teh pa upoštevamo še dve vrsti znakov.

.32 Identifikacijski znaki so znaki, ki služijo identifikaciji ali razpoznavanju enot. V primeru popravkov, moramo namreč vedeti, na katero enoto se podatki nanašajo. Identifikacijski znaki so n.pr. ime in naslov enote opazovanja ali tudi enote poročanja itd. Ti podatki so individualnega značaja in nimajo z vsebino opazovanja nikake zveze.

.33 Ker obstoji med znaki večkrat neposredna zveza, postavljamo v teh primerih včasih tudi znake, katerih vrednost bi mogli dobiti že iz vrednosti drugih znakov in bi jih mogli izpustiti. Ker ti znaki služijo za kontrolo drugih znakov, jih imenujemo kontrolne znake.

32. ORGANIZACIJSKO - TEHNIČNI DEL OPAZOVANJA

Plan in potek organizacijsko - tehničnega dela opazovanja mora biti tak, da bo glede na prilike terena, sposobnost kadrov, finančna sredstva in vse druge pogoje, ki vplivajo na izvedbo statističnega opazovanja, cilj opazovanja čim boljše in čim hitreje izpolnjen. Medtem, ko je postavljanje programa statističnega opazovanja vsebinsko delo in možno ob ozkem sodelovanju statistikov s strokovnjaki iz področja, v katerega predmet opazovanja spada, je organizacijsko tehnični del predvsem delo statističnih organov.

32.1 Viri podatkov: Pri organizaciji statističnega opazovanja je važno, da predhodno analiziramo vire, iz katerih bomo črpali podatke, ker šele ta analiza odkrije pravilen način opazovanja. Pri tem moremo ločiti dvoje vrst virov podatkov:

a) Vire, ki statistično opazovanje nadomeste
b) Vire, ki so osnova izvedbe statističnega opazovanja.

Kar se tiče virov prve vrste, je važno, da predhodno pregledamo, ali

je bil pojav že proučevan oziroma ali obstoje viri, na podlagi katerih bi mogli dati odgovor na cilj opazovanja brez izvršitve opazovanja. Za [illegible] se moramo, da je neposredno statistično opazovanje najtežji in najdražji način, da pridemo do podatkov. Enako je zelo važno, ako obstoja že zbrano gradivo, ki ga je treba samo na novo obdelati v skladu s postavljenim ciljem.

Ako tega ni, je treba pregldati vire, ki bodo osnova izvedbe opazovanja. Ti viri morejo biti trojnega značaja:

a) neposredno registriranje obstoji v tem, da ugotovimo podatke o enotah neposredno n.pr. z merjenjem, tehtanjem (n.pr. tehtanje zrn pri metraži).

b) izjave oseb, katerim so znane enote, katere opazujemo. Izjave oseb so n.pr. vir podatkov pri popisih prebivalstva.

c) obstoječi zapiski so podlaga opazovanja skoro vseh opazovanj iz gospodarstva. V tem primeru se zbirajo podatki iz obstoječih evidenc. Na obstoječe zapiske se oslanja tudi proučevanje naravnega gibanja prebivalstva, za katero se črpa podatke iz tekoče matične službe, katere glavni namen ni registriranje za potrebe statistike, temveč je njen glavni namen drug.

32.2 Vrste opazovanj. Vrsta opazovanja je odvisna od tega, ali mora statistično opazovanje dati trenutno ali dinamično sliko opazovanega pojava in od tega, ali opazujemo realne enote ali dogodke.

.21 Kadar hočemo dobiti enkratno, trenutno sliko - stanje - masovnega pojava, ki sestoji iz realnih enot, se običajno poslužujemo najobičajnejšega načina-statističnega popisa. Glavne značilnosti popisov so naslednje:

a) Časovno je statistična masa pri popisih opredeljena s trenutkom-kritičnim trenutkom. Kritični trenutek mora biti izbran tako, da je statistična masa v najmanjšem gibanju (n.pr. pri popisih prebivalstva polnoč), da je njeno stanje normalno in pridejo do izraza tipične črte pojava (n.pr. pri popisih prebivalstva ne poletje) in da so enote čim lažje dosegljive (pri popisih živine ne poleti).

b) Statistični popis zbira podatke o vseh enotah statistične mase in je zaradi tega običajno akcija velikega obsega.

c) Statistični popis da podrobno sliko sestava statistične mase in podrobno opazuje večje število znakov.

d) Statistični popis je izmed vseh statističnih opazovanj akcija največjega obsega, ki pa jo je treba izvršiti izredno hitro. Zato zahteva posebno organizacijo in postopanje.

.22 Ako potrebujemo manj podrobno sliko o stanju statistične mase in obstoja možnost, da se more dati sumarna slika za manjše teritorialne enote, uporabljamo tako imenovano tabelarno metodo. Metodo imenujemo tabelarno statistiko zaradi tega, ker so dani podatki za manjšo skupnost enot že v osnovi sumarno, tabelarno in ne iščemo podatkov o posameznih enotah.

Čeprav je tabelarna statistika znak zastarele statistike, jo vendar še pogosto uporabljamo, ker more dati, čeprav subjektivno ocenjene, podatke o statističnih masah v razmeroma kratkem času. Tabelarno metodo uporabljamo pri nas v osnovni kmetijski [illegible], ker je osnovna enota KLC ali OLC.

.23 Dinamično sliko sestava mase realnih enot dobimo z nizom periodičnih popisov. Za periodične popise veljajo pravila:

a) periodični popisi naj se vrše vedno ob istem letnem času, da ne moti sezonski značaj pojavov,

b) razd[illegible] med posameznimi popisi naj bo vedno eno in isto,

c) čim bolj gibljiva in spremenljiva je statistična masa, tem krajše naj bo razdobje med dvema popisoma. Popise prebivalstva vršimo na vsakih 5 ali 10 let, popise živine vsako leto itd.

.24 Opazovanje mase dogodkov je možno izvršiti točno le s tekočim registriranjem. To pomeni, da vsak pojav registriramo z vsemi važnimi značilnostmi takoj ob nastopu. Na osnovi tekoče registracije so pri nas organizirane vse oblike socialistične evidence. Enako je na tej osnovi organizirana tudi matična služba.

Tekoče registriranje ima naslednje značilnosti:

a) Osnovna evidenca tekočega registriranja služi predvsem za redno operativno delo v področju, v katerem je uvedena in služi statistiki le posredno,

b) Sistem pokazateljev raznih vrst tekočih evidenc mora biti medsebojno usklajen. Vsebovati mora vse podatke, ki so potrebni državni administraciji ali za operativno ali plansko vodstvo. Vendar je škodljivo zbiranje odvečnih in nepotrebnih podatkov, ki obremenjujejo tako administrativni kot gospodarski aparat.

c) Težiti je treba k poenostavljenju evidenc, seveda v možnih mejah in ne k nepotrebnemu razširjanju, ker vodi to v birokratizem, ki ovira delo administrativnih in gospodarskih organov.

.25 Podatke tekoče registracije koristimo s tekočimi poročevalskimi službami, ki v sumarni, kratki obliki dajo možnost spremljanja dinamike pojavov na kratke roke. Čas poročanja in periodičnost odvisi od značaja opazovanega pojava in operativne potrebe. Imamo tedenske, dekadne in mesečne poročevalske službe. Poročevalske službe so orodje za operativno

kontrolo in vir podatkov za statistično opazovanje.

32.3 Vrste opazovanj po obsegu .

.31 Statistika običajno opazuje masovne pojave na podlagi kompletnega ali popolnega opazovanja. Pri popolnem opazovanju opazujemo vse enote statistične mase. S tem dobimo najpopolnejše podatke.

.32 Vendar pa je možno napraviti določene zaključke in odkriti zakonitosti masovnih pojavov že na osnovi opazovanja samo dela enot statistične mase. Za nekatere vrste mas, kot neomejene ali hipotetične je to tudi edin način, ker maso v celoti niti ne moremo registrirati. Opazovanje, pri katerem opazujemo samo del enot statistične mase in pri katerem napravimo iz tega dela zaključke na celoto, imenujemo delno opazovanje. Važen činitelj, da moremo iz delnega opazovanja sklepati na zakonitosti in celoto, je zakon velikih števil.

Delna opazovanja ločimo po svojem značaju na več vrst:

a) Monografija obsega opazovanje ene same tipične enote, ki pa jo obravnava podrobno in kompleksno (n.pr. Piro Monografija slovenske vasi).

b) Anketa sloni na opazovanju sorazmerno majhnega števila enot, ki so izbrane po predhodni analizi predmeta in predstavljajo tipične enote.

c) Kot opazovanje osnovne mase imenujemo delno opazovanje, pri katerem opazujemo samo enote, ki imajo glede na obravnavane pojave pomen, medtem ko izpuščamo enote, ki so glede na proučevan pojav malo pomembn. V SSSR so n.pr. l. 1939. pri popisu poljedelskih del zajeli s popisom 242000 kolhozov in sovjozov 99 °/o vseh poljedelskih del. S tem, da niso registrirali 19 milj. privatnih gospodarstev, niso registrirali samo 1 °/o poljedelskih del.

d) Vzorčenje pa ima za osnovo slučajni izbor enot, ki pridejo v delno opazovanje. Slučajni izbor enot zagotovi reprezentativnost delne mase in dopušča uporabo teorije vzorčenja, katere osnova je verjetnostni račun. Objektivnost ocene, možnost določanja pogreške in možnost vplivanja na velikost pogreške so tako važni činitelji, da se je tako teorija kot uporaba vzorčenja zelo izpopolnila in razširila. Zaradi važnosti in posebnosti metode jo bomo obravnavali v posebnem poglavju.

32.4 Načini opazovanja. Tehnika opazovanja mora biti od primera do primera različna in prilagojena danim prilikam in pogojem. V glavnem ločimo naslednje načine opazovanja.

.41 Zelo običajen način opazovanja je, da posebni organi, popisovalci, obiščejo popisne enote in zberejo podatke. Pri tem ločimo dva različna postopka.:

a) O samoregistraciji govorimo, kadar popisane enote same izpolnijo obrazce. Popisovalec v tem primeru obrazce raznese in jih po preteku nekoliko dni izpolnjene zopet zbere.

b) O ekspedicijskem načinu pa govorimo, ako popisovalec obišče popisne enote in sam, na podlagi izpraševanja izpolni obrazce.

Vsak izmed teh dveh načinov ima svoje prednosti in pomanjkljivosti. Samoregistracijo moremo uporabiti v primeru, da je splošna raven onih, ki dajejo podatke, dovolj velika, z drugimi besedami, da so zmožni sami izpolniti obrazce. Edino možen način je to tudi v primeru, ako morajo enote šele zbirati potrebne podatke, n.pr. dokumente. Možnost napak je na vsak način večja kot pri ekspedicijski metodi, kjer so popisovalci ljudje, ki so na posebnih tečajih izvežbani za svoj posel in zainteresirani nad pravilnostjo rezultatov. Vendar nastanejo pri ekspedicijskem načinu problemi, ako enot ni možno ob določenem času dobiti itd.

.42 O prijavnem načinu govorimo, kadar osebe, ki morajo dati podatke, pozovemo na določeno mesto, kjer jih statistični organi popišejo. Pri tem načinu odpade obhod terena.

.43 Poštno - telegrafski način predvideva pošiljanje in vračanje izpolnjenih obrazcev po pošti, oziroma telegrafu. S tem opravi tehnično delo popisovalcev takorekoč pošta. Telegrafska služba se uvede v primeru, ako je potrebno podatke o opazovanih pojavih neposredno po izvedenem opazovanju zbrati na določenem mestu (n.pr. telegrafska služba v kmetijski statistiki).

.44 Pri korespondentnem načinu opazovanja vrše opazovanje takozvani korespondenti, osebe, katerih glavni posel ni statistika, pač pa so strokovnjaki področja, iz katerega vršimo opazovanje in morejo dati zaradi tega dobre podatke. Mreža korespondentov je stalna. Korespondenti običajno tekoče poročajo po tabelarni metodi. Posebno se korespondentni način uporablja v kmetijski statistiki.

32.5 Enota porčanja. Od enote opazovanja, ki je vsebinska enota statističnega opazovanja in tvori njena skupnost statistično maso, moramo ločiti enoto poročanja. Enoto poročanja imenujemo organizacijsko - tehnično enoto, od katere moremo dobiti podatke o enoti opazovanja. Ako popisujemo n.pr. industrijske stroje, je posamezen stroj enota opazovanja, posamezno podjetje, ki ima stroj in daje podatke o tem stroju, pa je enota poročanja itd.

32.6 Sredstva opazovanja.

.61 Kot sredstvo statističnega opazovanja nastopajo statistični obrazci ali formularji. Statistični obrazec je list papirja, na katerem so nanizana vprašanja - znaki in na katerem se za posamezne enote vpiše odgovore - vrednosti znakov. Izpolnjen obrazec nadomešča enoto in omogoča, da vršimo na enotah tehnične postopke, ki jih predvideva obdelava.

Najvažnejš pri sestavljanju obrazcev je predvsem vsebina obrazca - vprašanja oz. znaki. Seveda pri tem ne smemo zanemariti tehnične plati obrazca. Obrazec mora biti sestavljen tako, da bo zadostil vsebini, izvedbi opazovanja in obdelavi.

Vsebine in izvedbe opazovanja se tiče predvsem to, kako so vprašanja postavljena in šele drugovrstno tehnična oblika. Tehnična oblika je predvsem prilagojena predvideni obdelavi.

.62 Vprašanja morajo biti postavljena tako, da se na čim lažji način dobi točen in nedvoumen odgovor na vprašanje. Pri tem moramo imeti pred očmi pravila za vsebinske znake, ki so navedeni v odstavku o znakih opazovanja.

Tehnično moremo podajati v obrazcu vprašanja na več načinov.

a) Najpogostejši način je, da je vpisano vprašanje, na katerega se v zato določenem mestu vpiše odgovor.

N.pr.

stan	samski ←	— prostor za odgovor

b) Zelo priporočljivo je, ako število vrednosti znakov ni preveliko, da vse vrednosti znaka naštejemo. Odgovarjajočo vrednost se kot odgovor po navodilih ali podčrta, obkroži redno številko itd. n.pr.

Stan: 1 samski, 2 poročen, 3 vdovec-a, 4 razveden.

c) Ako ima znak preveliko vrednosti znakov, moremo vrednosti znaka zgrupirati v grupe in našteti grupe vrednosti znakov. V nadaljnjem postopamo kot pod točko b) n.pr. šolska izobrazba 1. brez šolske izobrazbe, 2. osnovna šola, 3. nižja srednja šola, 4. višja srednja šola, 5. fakultetna izobrazba.

d) Moremo pa v primeru pod c) postaviti kombinacijo metode a in b s tem, da pustimo pri grupni vrednosti prostor, v katerega se vpiše v odgovarjajoči grupi podroben odgovor n.pr. šolska izobrazba.

1 brez šolske izobrazbe	
2 osnovna šola	
3 nižja srednja šola	
4 višja srednja šola	6 razredov gimnazije
5 fakultetna izobrazba	

.63 Organizacijske sestavine obrazca: Poleg vsebinske strani morajo statistični obrazci vsebovati tudi nekatere organizacijske sestavine. Med temi so najvažnejše:

a) Kratek naslov obrazca s šifro in številko v sistemu obrazcev, n.pr. Ind_3, PO_2 itd.

b) Naziv organa ali ustanove, ki je obrazec izdala.

c) Datum, številka in označitev pristojnega statističnega urada, ki je obrazec odobril.

d) Naziv ali oznaka poročevalne enote.

e) Oznaka časa opazovanja (kritični moment, oziroma razdobje poročanja.)

f) Označitev roka, v katerem je treba obrazce izpolniti in vrniti.

g) Označba, v koliko primerih je treba obrazec izpolniti.

h) Naslovi, komu je treba poslati izpolnjene obrazce.

i) Podpis osebe, ki je obrazec izpolnila in podpis osebe, ki odgovarja za resničnost in točnost podatkov.

Ti podatki so neobhodno potrebni za redno in pravilno delo statistične službe.

.64 Vrste obrazcev glede na število enot, ki jih vsebujejo.

Glede na število enot, ki jih moremo popisati z enim statističnim obrazcem, ločimo:

a) individualne obfazce,

b) kolektivne obrazce.

a) Individualni obrazec je namenjen za popisovanje ene same enote. Ima velike prednosti pri urejevanju gradiva (sortiranj), pač pa je potrošnja papirja pri njem znatna.

b) Kolektivni obrazec se uporablja za registriranje več enot na enem in istem obrazcu. Pri tem moramo ločiti dvoje različnih primerov. Enote, popisane na enem obrazcu morejo tvoriti skupno novo statistično enoto (n.pr. gospodinjski list pri popisu prebivalstva, kjer pridejo popisane na en kolektiven obrazec samo osebe enega gospodinjstva, ki tvori novo enoto). S kolektivnim obrazcem pa morejo biti podpisane tudi enote, med katerimi ni vsebinske zveze. Te obrazce moremo izpolnjevati samo na ekspedicijski ali prijavni način. Obdelava teh podatkov je posebna. Potrošnja papirja pa je znatno manjša kot pri individualnih obrazcih.

.65 Tehnična redakcija obrazcev. Kakovost zbranih podatkov je v veliki meri odvisna tudi od tehnične izdelave obrazcev. Smiselno pregledno sestavljen obrazec, ki vsebuje dovolj prostora za odgovore, ki je skratka sestavljen tako, da čim bolj olajša izpolnjevanje, ima velike prednosti tako pri zbiranju kot pri urejevanju gradiva.

.66 Delitev obrazcev po vlogi, ki jo imajo v statističnem opazovanju. Pri statističnem opazovanju ne nastopa en sam statistični obrazec, temveč običajno cela zbirka obrazcev, katerih vloga v statističnem opazovanju je različna. Najvažnejši so glavni obrazci, ki vsebujejo vsebinske podatke o statističnih enotah. Naloga vseh ostalih, pomožnih obrazcev je, da pomagajo izpolnjevati glavne obrazce, ali, ako nastopajo kot kontrolni obrazci, da omogočajo kontrolo polnoštevilnosti enot. Kontrolniki včasih vsebujejo tudi osnovne vsebinske podatke, da je na ta način možna hitra, predhodna obdelava najosnovnejših podatkov.

.67 Navodila za izpolnjevanje obrazcev. Obrazce je potrebno opremiti z vsebinskimi in tehničnimi navodili za izpolnjevanje. Ni potrebno poudarjati, da morajo biti navodila jasna, kratka, a kljub temu izčrpna. Navodila, ki se tičejo izpolnjevanja posameznih vprašanj, ja najboljše, da so tiskana neposredno pri vprašanjih.

32.7 Čas opazovanja.

Čas opazovanja je razdobje, v katerem so opazovanje statističnih enot dejansko izvede. Ta čas ni identičen s kritičnim datumom, niti s trenutkom izvršitve dogodka, temveč je običajno krajše razdobje (par dni) po kritičnem trenutku, oz. izvršitvi dogodka. Ta rok popisa ali registracije ne sme biti preveč odmaknjen od kritičnega trenutka popisa oz. izvršitve dogodka, ker je drugače težko ugotoviti stanje ob kritičnem trenutku.

32.8 Kraj opazovanja in geografska razdelitev terena.

Kraj opazovanja je določen z definicijo enote opazovanja in ni potrebno, da se v vsakem primeru sklada s krajem, kjer se enota v kritičnem trenutku nahaja (n.pr. začasno odsotne osebe pri popisu prebivalstva).

Ker celotno opazovanje ne izvrši na celotnem teritoriju ena sama oseba, temveč veliko število oseb - popisovalcev, od katerih vsak takorekoč samostojno opazuje, je jasno, da je potrebno opredeliti področje posameznega popisovalca, to je, določiti teritorij, na katerem bo izvršil opazovanje. Pri tem delu se vedno naslanjamo na obstoječo upravno razdelitev terena, ker je s tem opravljeno že veliko dela pri razdelitvi terena. Pri statističnih opazovanjih, ki vsebujejo manjše število enot, more biti najmanjša upravna enota že obenem popisno področje enega popisovalca. Kadar pa gre za opazovanje, pri katerih je število enot v najmanjših upravnih enotah večje, kot bi jih mogel en popisovalec popisati v času popisa, te najmanjše upravne enote razdelimo v še manjša področja, popisne okoliše, ki so tako veliki, da more popisovalec v času popisa popisati vse enote tega področja. Ti popisni okoliši morajo biti pri določenih opazovanjih zelo nenavadni n.pr. pri popisu prebivalstva: planinska koča ali tretje nadstropje stanovanjskega bloka.

ija
Razdelitev teritor na popisne okoliše mora biti taka:

a) da jo v njih toliko enot, kolikor jih more popisovalec v času popisa popisati;

b) da je možno iz popisnih okolišev sestaviti področja najmanjše administrativne enote;

c) da vsaka geografska točka pade v en in en sam popisni okoliš;

d) da so meje med okoliši jasne in po možnosti naravne.

Zaradi jasnosti običajno izdelamo sk e popisnih okolišev.

Ker je teritorialna razdelitev terena zelo važna za dobro in pravilno izvedbo statističnega opazovanja, se za posamezna večja opazovanja, kot je n.pr. popis prebivalstva, izdaja zapora nad izpremembami v upravni razdelitvi za čas priprav in izvajanja opazovanja. Včasih izvršimo tudi predhodni popis naselij itd.

32.9 Organi statističnega opazovanja:

Osebe, ki sodeljujejo pri statističnem opazovanju, moremo po njih vlogi v statističnem opazovanju deliti na organizatorje, kontrolorje, popisovalce in popisane osebe, po značaju njihovega sodelovanja pa na:

a) osebe, ki jim je statistika poklic,

b) osebe, ki imajo drugo glavno zaposlitev, vendar trajno sodelujejo pri določenih statističnih opazovanjih,

c) osebe, ki priložnostno sodelujejo v statističnem opazovanju kot popisovalci,

d) osebe, ki priložnostno sodelujejo kot osebe, ki dajejo podatke o popisnih enotah.

Osebo, ki jim je statistika poklic, tvorijo statistično linijo. Statistična linija organizira statistična opazovanja in pri manjšem številu enot opazovanje tudi sama izvrši. Statistična linija ima zvezne, republiške in okrajne organe. Njih izobraževanje je možno trajno, enako moremo vršiti tudi izbiro.

Statistični liniji je v privi fazi zbiranja podatkov zelo sorodna mreža evidentičarjev, katerih naloga je vodenje gospodarske evidence.

Osebe, ki imajo drugo glavno zaposlitev, a trajno sodelujejo pri določenih opazovanjih, so korespondenti. Korespondenti nastopajo obenem v vlogi popisanih oseb in popisovalcev.

Pri opazovanjih velikega obsega, statistična linija sama ne more izvesti opazovanja, ki je v vseh primerih masoven posel. Zaradi tega si pomaga z ljudmi, ki sodelujejo pri opazovanju priložnostno. Ti ljudje sodelujejo pri statističnem opazovanju kot popisovalci, včasih tudi kot kontrolorji. Ia ustrezajo pogojem, ki so potrebni in da morejo v redu vršiti

naloge popisovalcev, si pomagamo s tem, da izberemo ljudi, ki jim je predmet opazovanja znan po njih stroki, z vežbanjem na posebnih tečajih in tiskanimi navodili.

Osebe, ki priložnostno sodelujejo s tem, da dajejo podatke o popisnih enotah, se z nalogami, ki jih imajo pri statističnem opazovanju, seznanjajo s propagando v časopisih, radiu, lepaki in tiskanimi navodili.

32.10 Poizkusno (pilotno)opazovanje. Plan opazovanja, obrazci, navodila itd. so sestavljeni na osnovi izkušenj prejšnjih opazovanj, poznavanja teorije in metode statistike, poznavanja terena itd. Vendar kljub temu ne vemo, ali smo izbrali danim pogojem in nalogam najugodnejšo rešitev. Zaradi tega je treba celoten sistem opazovanja predhodno preizkusiti s poizkusnim, pilotnim opazovanjem. Pilotno opazovanjo izvedemo tako, da izvršimo v malem celotno opazovanje, vključno obdelavo. Opazovanje izvršimo v nekaj krajih, ki so tipični, raje slabi glede na možnost izvedbe, kot predobri. Opazovanje je treba izvesti s povprečnimi kadri, strokovnjake pa dodeliti le kot opazovalce, ki si nabirajo izkušnje in odkrivajo hibe sistema in plana. Šele, ko smo na osnovi pilotnega opazovanja preizkusili in popravili morebitne napake, napravimo dokončen plan, damo tiskati obrazce, navodila itd.

32.11 Propaganda. Pri velikih popisih, (popis prebivalstva, popis kmetijskih gospodarstev itd.) je uspeh popisa v bistveni meri odvisen od tega, kakšno je razumevanje popisanih oseb do popisa. Zaradi tega še pred popisom in med samim popisom z vsemi razpoložljivimi sredstvi (časopis, radio, film, plakati, ustno) tolmačimo važnost popisa in način izvedbe. Na ta način mnogo pridobimo, tako na lažji izvedbi kot tudi na kakovosti podatkov.

33. ORGANIZACIJSKI PLAN OPAZOVANJA

Organizacijski plan opazovanja obstoja iz navodil za izvršitev opazovanja. Ta plan predvideva vsebino, potek in roke posameznih faz, zadolžitev organov pri posameznih delih itd. Ker je statistično opazovanje zelo obsežno delo, sestavljeno iz velikega števila faz, ki so povezane med seboj, tako vsebinsko kot časovno, navodila v običajni obliki ne nudijo zadostnega pregleda nad planom. Zato se poslužujemo pri planiranju statističnega opazovanja dveh pripomočkov: terminske tabele in operograma.

33.1 Terminska tabela v grafični obliki pokaže roke, oziroma čase izvajanja posameznih faz. Roki in časi izvajanja posameznih faz so v terminski tabeli črtani s pasom, ki se vleče čez odgovarjajoč del časovne skale. Časovna skala vsebuje koledar, poleg tega pa ima vnešeno tudi skalo, ki označuje, koliko je posamezen dan oddaljen od najvažnejšega

trenutka statističnega opazovanja - kritičnega datuma. Iz terminske tabele je jasno razviden časoven redosled posameznih faz statističnega opazovanja. Ni pa razvidna zveza z ostalimi elementi statističnega opazovanja - organi in sredstvi. Zaradi tega uporabljamo poleg terminske tabele še drug grafični pripomoček in sicer operogram.

33.2 Operogram povezuje med seboj organe, sredstva in faze. Vidimo, da operogram ne vsebuje časa izvajanja faz in je zaradi tega potrebno, da izdelamo za posamezno statistično opazovanjo terminsko tabelo in operogram. Ker vsebuje operogram tri elemente, so v operogramu v eno smer tabele nanešeni n.pr. organi, v drugo faze, sredstva pa so vrisana z dogovornimi znaki v sama polja tabele. Konstrukcija je najlepše razvidna iz same slike. Iz operograma moremo že pri planiranju opazovanja razbrati smiselnost ali nepotrebnost posameznih del.

34. KONTROLA STATISTIČNEGA OPAZOVANJA

Gradivo, zbrano s statističnim opazovanjem, bo imelo vrednost za proučevanje le tedaj, če bo polnoštevilno in pravilno. Ker sodeluje pri statističnem opazovanju veliko ljudi, ki bodisi ne znajo ali nočejo delati točno po danih navodilih, moremo bistveno dvigniti kvaliteto zbranih podatkov z dobro organizirano kontrolo.

Kontrolo opazovanja moremo tako po svoji vlogi kot času razdeliti v kontrolo izvajanja statističnega opazovanja med samim opazovanjem in kontrolo zbranega gradiva po izvršenem opazovanju.

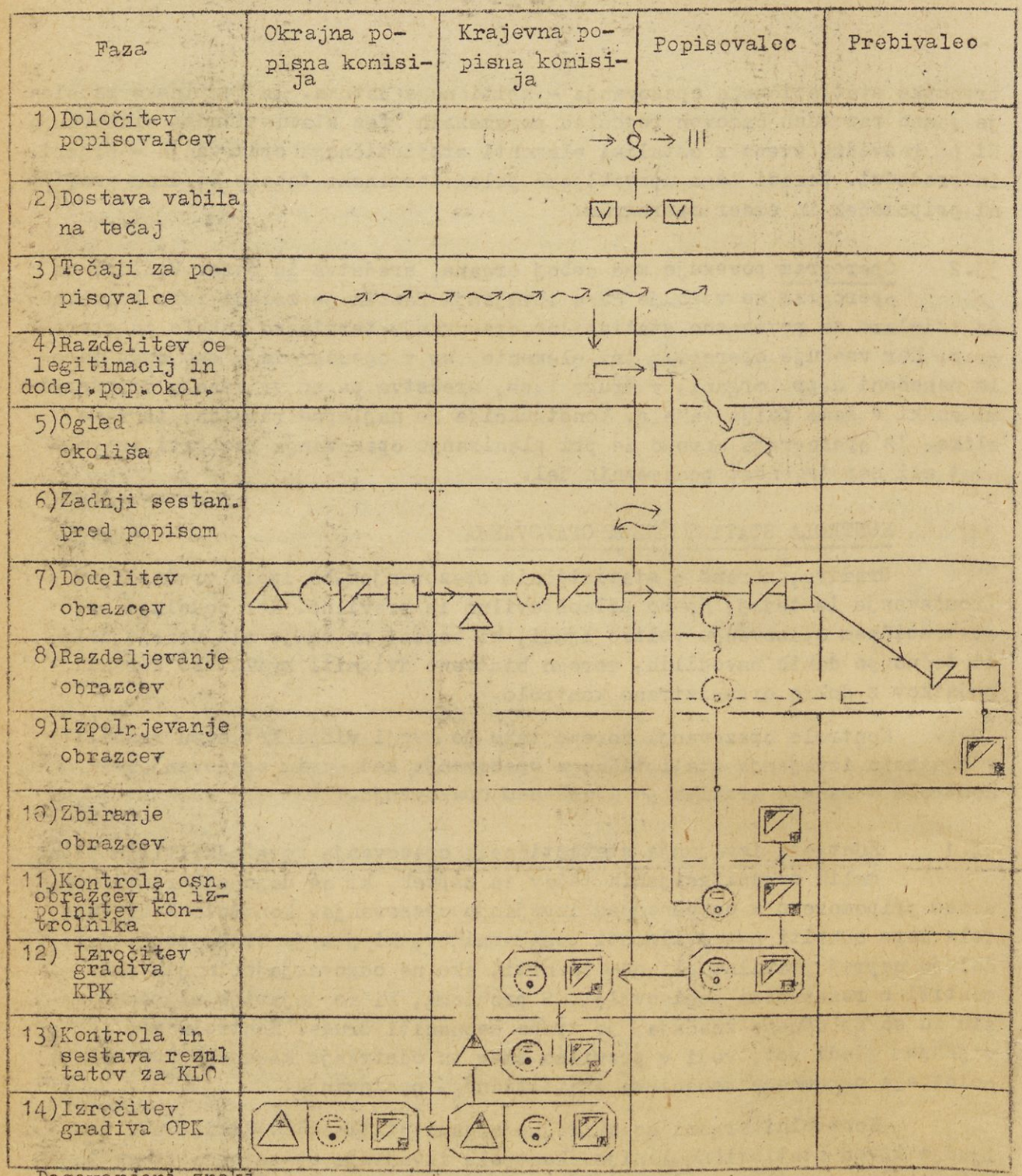
34.1 Kontrola izvajanja statističnega opazovanja more odkriti še polno malih organizacijskih težav in zaprek, ki se dajo odstraniti in s tem pripomoremo k uspešnejšemu izvajanju opazovanja. Kontrola kvalitete dela more odkriti nepravilnosti posameznikov, ki jih je treba poučiti, da delajo naprej pravilno ali odstraniti, ako ne odgovarjajo, in jih nadomestiti z rezervnimi popisovalci. S problemi, ki se pojavijo na kakem mestu in so splošnega značaja, je treba seznaniti druge. Kontrola med izvajanjem gladi pot, vodi v pravilno smer in odstranja kamenčke med kolesi celotnega ogromnega mehanizma statističnega opazovanja.

Kontrolni organi so običajno organi statistične linije ali pa ljudje izven statistične linije. Kontrola izvajanja opazovanja je sestaven del opazovanja in brez nje ne moremo priti do zadovoljivih rezultatov.

34.2 Kontrola zbranega gradiva po izvršenem opazovanju se od gornje kontrole v celoti razlikuje. Tu ne gre za kontrolo opazovanja, temveč za individualno kontrolo izpolnjenega gradiva t.j. posameznih enot in posameznih vprašanj.

Kontrola zbranega gradiva se nanaša na polnoštevilnost zbranih

Sl. 3.2 Operogram popisa prebivalstva l. 1948



Dogovorjeni znaki

§ - določitev odreditev
V - vabilo na tečaj
⬡ - popisni okoliš
⊏ - legitimacija

A →B = izroči B

▽ - popisnica
□ - list za gosp.
○ - kontrolnik
△ - rezultat KLO
⬭ - omot

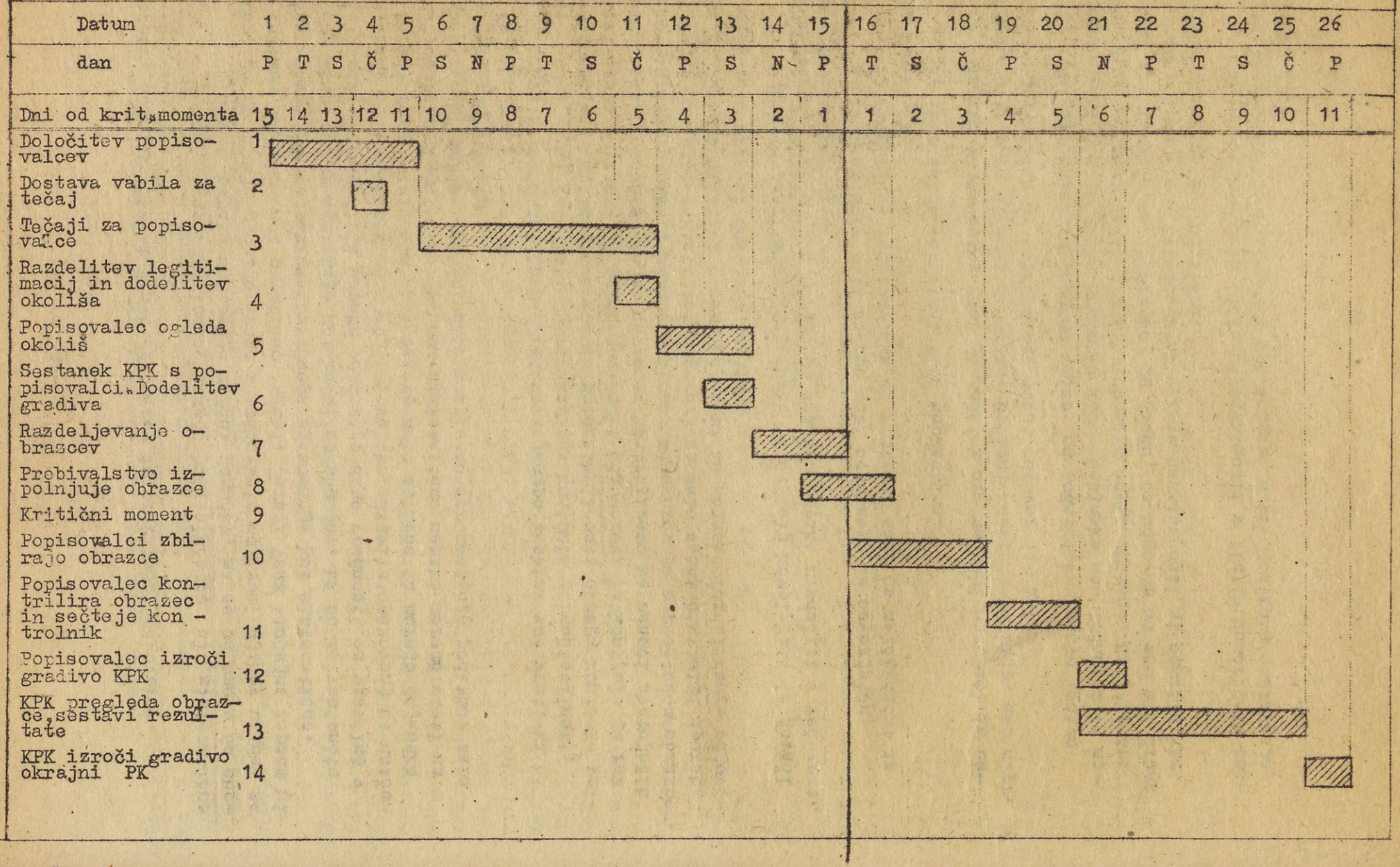
A - - →B = pokaže B

| - izpolni v celoti
| - izpolni delno
! - kontrolira

A ~→B = A pošlje B

Sl. 3.1 Terminska tabela

Popis prebivalstva 15. marca 1948



obrazcov, polnoštevilnost danih odgovorov in pravilnost odgovorov.

.21 Napake, ki se pojavljajo, morejo biti slučajne ali sistematične. Slučajne napake nastanejo zaradi pogojev, ki se od enote do enote menjajo in nimajo globlje zveze z opazovanim pojavom. Njihov učinek se pri velikem številu enot praktično izgubi. Zaradi tega slučajne napake niso tako nevarne za uspeh statističnega opazovanja kot sistematične. Sistematične napake nastajajo kot rezultat pogojev, ki so za vse enote isti. Zaradi sistematične napake odstopajo dane vrednosti od pravilne v isto smer. Njihov učinek se torej pri velikem številu opazovanj ne uničuje. Zaradi tega so sistematične napake zelo nevarne in morejo dejansko sliko pojava zmaličiti. Kot primer sistematične napake moremo vzeti namerno utajo njivskih površin, oziroma splošno spreminjanje podatkov zaradi gospodarskih koristi.

.22 Kontrol statističnega gradiva imamo po vsebini več vrst, ki bi jih mogli razdeliti v tri skupine: kontrola polnoštevilnost, čitljivosti in pravilnosti. Kontrola polnoštevilnosti obsega kontrolo polnoštevilnosti obrazcev in polnoštevilnost odgovorov. Ta kontrola je razmeroma lahka. Polnoštevilnost obrazcev kontroliramo na osnovi obstoječih spiskov in kontrolnikov, polnoštevilnost odgovorov pa s pregledom obrazcov. Čitljivost in jasno postavljeni odgovori so zelo važni činitelji uporabnosti gradiva. Nečitljivo pisani odgovori in nejasno formuliranje odgovorov more zelo ovirati obdelavo gradiva.

Kontrola pravilnosti odgovorov je najbolj kompleksna in zaradi tega najtežja. To kontrolo moremo po značaju napake razdeliti v več vrst. Značaj napake more biti:

a) Podatek more biti sam po sebi nesmiseln in neverjeten;

b) Primerjava podatkov, ki so v zvezi, kaže na nelogičnost in nepravilnost;

c) Podatek je verjeten, toda nepravilen.

Napake prve vrste, kjer je podatok sam na sebi nesmisel in neverjeten, obravnava stvarna kontrola. Stvarna kontrola odkrije, ako n.pr. na vprašanje, koliko je nekdo star, stoji odgovor 115 let.

Pri obravnavanju logične zveze med podatki ene enote ločimo dve vrsti kontrole. Logična kontrola išče vsebinske neskladnosti in nelogičnosti v odgovorih. Ako je n.pr. z neko osebo vpisano, da je stara 7 let in ima fakultetno izobrazbo, je jasno, da odgovor na eno ali drugo vprašanje ni pravilen. Stvarna kontrola te neskladnosti ni odkrila, ker ni iskala povezave med odgovori.

Računska kontrola išče nepravilnosti v tistih numeričnih znakih, ki so povezani z računskimi operacijami n.pr. kontrola seštevanja, množenja itd.

Stvarna, logična in računska kontrola, ki se vrši s pregledom posameznih obrazcev more nekatere napake, ki jih odkrije, tudi takoj popraviti, če je prava vrednost očitna, dostikrat pa napako samo odkrijo, medtem ko je treba popravek iskati pri viru -- popisovalcu ali popisani osebi. Za izvrševanje stvarne in logične kontrole je potrebno dobro poznavanje proučevanega predmeta.

Problem zase so podatki, ki so sicer verjetni, a kljub temu nepravilni. Nje namreč ne moremo odkriti niti s stvarno, niti z logično ali računsko kontrolo. Te napake more odkriti individualno le oseba, ki kontrolirano enoto dejansko pozna, to pomeni, je čim bližje terenu. Vendar moremo te napake odkriti tudi z drugo metodo, ki sicer ne odkriva teh napak individualno, za vsako enoto, temveč ugotovi njih skupni učinek. Ta takoimenovana statistična kontrola je osnovana na metodi vzorčenja. Določen odstotek slučajno izbranih enot pregledajo izredno vestni in usposobljeni kadri. S tem se število napak vseh vrst za te kontrolirane enote skrči na najmanjšo mero. S primerjavo kontroliranih podatkov s podatki, ki smo jih dobili za iste enote z osnovnim opazovanjem, posplošenjem teh razlik na celot , moremo oceniti skupni učinek napak in s tem skupnim učinkom podatke popraviti.

.23 Ostane še vprašanje, kdo naj vrši posamezne vrste kontrol. Kontrolo more vršiti popisovalec, posebej za to postavljene kontrolne skupine na terenu, oziroma statistični organi na vseh stopnjah, tako okrajni kot republiški. Kontrola se vrši običajno na več stopnjah. Za kontrolo, ki jo na samem terenu popisovalci vrše, govore naslednja dejstva: Ta kontrola se more izvesti neposredno po izvršenem opazovanju, ko dobimo od opazovanih enot še pravilne popravke. To velja posebno za samoregistracijo, pri kateri popisovalec pri prevzemu prekontrolira obrazce. Organi na terenu morejo zaradi poznavanja prilik odkriti dosti napak, ki jih višji organi ne bi mogli. Ekspeditivnost popravljanja napak je največja, ker je možen hiter, neposreden stik z opazovanimi enotami. Pomanjkljivost te kontrole pa je v tem, da je ta kadar razmeroma nekritičen, premalo vesten in ne pozna v podrobnostih predmeta opazovanja. Vendar kljub temu odkrijemo in popravimo s to prvo kontrolo na terenu večino napak, posebno napake glede polnoštevilnosti.

Kontrola na višjih stopnjah ima to prednost, da jo vrše strokovnjaki in je zato kvalitetnejša. Pomanjkljivost pa je v tem, da kljub temu ne poznajo individualnih prilik na terenu in da je popravljanje zaradi oddaljenosti od terena počasno.

Kontrola je običajno organizirana v več stopnjah, ki se med seboj dopolnjujejo.

4. GRUPIRANJE IN STATISTIČNE TABELE

41. GRUPIRANJE

41.0 Uvod. Prva naloga po izvršenem statističnem opazovanju je ureditev in sistemizacija zbranega statističnega gradiva. Šele z ureditvijo in sistemizacijo napravimo statistično maso pregledno in uporabno za analizo. Statistično gradivo, zbrano pri statističnem opazovanju tvori namreč za-enkrat nepregledno maso podatkov, iz katere je nemogoče napraviti katerikoli zaključek.

Osnovna metoda v težnji za sistematizacijo statističnih podatkov je grupiranje. Statistična masa je preglednejša, ako istovrstne enote združimo v grupe. Istovrstne glede na nek znak so one enote, ki imajo eno in isto vrednost znaka. Tako moremo razdeliti prebivalstvo po spolu v dve grupi tako, da tvori moško prebivalstvo eno grupo, žensko prebivalstvo pa drugo grupo. Enako moremo prebivalstvo grupirati po stanu itd. Statistično maso moremo torej po enem ali več znakih razdeliti v sistem delnih mas, grup enot, v katerih se vse enote ene grupe ujemajo v vrednosti znakov, po katerih smo maso razdelili. Na ta način kolektivno za vse enote grupe vemo, kakšne vrednosti grupnih znakov imajo enote grupe. Grupni znak imenujemo znake, po katerih je izvedeno grupiranje enot. S preštevanjem enot ali seštevanjem vrednosti nekega znaka v posameznih grupah moremo napraviti koristne številčne zaključke o statistični masi. S tem dobimo statistične serije, ki so osnovni načini podajanja statističnih podatkov. Grupiranje enot je torej osnovna metoda in problem urejevanja statističnega gradiva.

Vendar se grupiranje enot zamota, ako hočemo izvesti grupiranje enot po znakih, ki imajo veliko ali celo neomejeno število možnih vrednosti. V tem primeru zavzame posamezno vrednost zelo omejeno število enot, ali pa nobena. Preglednost, predvsem pa masovnost pojavljanja enot sta v tem primeru zelo majhna. Zaradi tega vse možne vrednosti znaka zreduciramo na manjše število vrednosti s tem, da podobne vrednosti znaka zgrupiramo v grupno vrednost znaka in te vrednosti med seboj izenačimo. Tako moremo n.pr. pri znaku starost tvoriti grupe vrednosti znaka n.pr. 0 - 1 leto, nad 1 leto - 2 leti, nad 2 leti - 3 let itd. Pri tem nas zanima samo, v katerem letu starosti je oseba, ne pa točna starost do dneva ali ure, kar je tudi možno določiti. Na ta način izreduciramo število možnih vrednosti iz neomejenega na 100, če smo vzeli starost po letih, ali na 20 vrednosti, ako smo vzeli grupe po 5 let.

Kot je razvidno iz gornje razlage, je treba strogo ločiti grupiranje enot od grupiranja vrednosti znakov. Tako smisel kot problematika enega in drugega sta popolnoma različna.

41.1 Grupiranje vrednosti znakov

Grupiranje vrednosti znakov prido v poštev pri vseh vrstah znakov, kor v vseh vrstah nastopijo znaki, ki imajo zelo veliko, ali pa celo neomejeno število vrednosti znakov.

Pri grupiranju vrednosti znakov moramo paziti, da so grupe vrednosti tvorjene tako, da vsaka možna vrednost spada samo v eno izmed grup, ker je v obratnem primeru grupiranje enot nemogoče.

Grupne vrednosti znakov imajo vse lastnosti prvotnih vrednosti znakov. Vsaka enota ima namreč eno izmed grupnih vrednosti. Pravtako moremo iz grupnih vrednosti tvoriti višje grupe. Ako vzamemo kot primer starost, ima vsaka oseba v nekem trenutku neko točno določeno starost, je pa tudi vsaka oseba v tem trenutku ali v prvem, drugem, tretjem itd. letu starosti. Letne grupe moremo dalje grupirati v petletke itd.

Čeprav moremo formalno grupirati vrednost znakov ne glede na masoven pojav in enote, katerim ta znak pripada, je nemogoče napraviti vsebinsko odgovarjajoče grupe vrednosti znakov brez poznavanja in analize predmeta in znakov. Zaradi tega je tvorjenje grup vrednosti znakov predvsem vsebinsko in šele drugovrstno formalno. Ako vrednosti znaka starost, ki je znak velikega števila raznovrstnih enot, zgrupiramo neglede na proučevan pojav, more grupacija po petletkah za enoto človek biti smiselna, za enoto n.pr. prašič pa brez smisla, ker bi vse vrednosti znaka padle v prvo grupe in je za ta primer treba stvoriti druge grupe, ki odgovarjajo naravi tega pojava.

.11 Grupiranje vrednosti časovnih znakov. Ker more znak čas zavzeti neomejeno število vrednosti znakov, mo ramo v vseh primerih izvesti grupiranje vrednosti znaka, da dobimo omejeno število vrednosti oz. grup.

Sorodnost oz. podobnost vrednosti znaka je izražena večinoma z oddaljenostjo dveh časovnih momentov. Grupe vrednosti časovnega znaka so n.pr. dan, teden, mesec, leto, poslovno leto, šolsko leto itd. More jo pa tvoriti časovne grupe tudi n.pr. vsi ponedeljki, vsi torki itd. določenega razdobja, če kaže dan v tednu kakšno značilnost za proučevan pojav in je v tem smislu ponedeljek enega tedna bolj podoben ponedeljku drugega tedna, kot pa ponedeljek torku istega tedna itd. Enako more jo tvoriti grupo tudi vsi isti meseci nekega daljšega razdobja (sezonski vpliv).

Težava nastopi pri tvorjenju časovnih grup tudi zaradi tega, ker je dolžina posamičnih časovnih grup med seboj različna, kar moti analizo.

.12 Grupiranje vrednosti krajevnih znakov. Vrednosti krajevnega znaka moremo grupirati s tvorjenjem geografskih področij.

.121 Najobičajnejše je tvorjenje geografsko- administrativnih grup. Taka grupa obsega vse krajevne točke administrativne enote n.pr. LR-ja. Pri geografskih grupah je lepo vidno razporedje grup osnovne

vrednosti grupiramo v kraje, kraje v okraje, okraje v republike, republike v celotno državo. Geografsko - administrativne grupe se običajno uporabljajo zaradi tega, ker jih je lahko tvoriti, oziroma so že dane z upravno razdelitvijo. Poleg tega potrebujejo državni organi podatke urejene po administrativnih enotah. Hiba tega grupiranja je, da je treba ob spremembah upravne razdelitve podatke preračunavati na novo upravno razdelitev, kar je precej zamuden posel. Hiba tega grupiranja je tudi, da grupiranje po upravni razdelitvi, ki je bila tvorjena po upravno- političnih vidikih, ne združuje sorodne kraje po ekonomskih vidikih.

.122 Zato za potrebe znanstvenega proučevanja tvorimo geografsko-ekonomske grupe ali rajone . Te dobimo, ako združujemo kraje z istimi ali podobnimi pogoji glede na proučevan pojav (živinorejski rajoni, sadjarski, žitni itd.). Enote, grupirane po rajonih, tvorijo šele grupe istovrstnih enot v vsebinskem smislu.

.123 Ker je tvorjenje rajonov zelo težaven posel, si običajno pomagamo s tem, da sestavimo rajon iz najnižjih geografsko- administrativnih grup. S tem se oslonimo tudi pri rajonizaciji na upravno razdelitev, kar olajša delo.

.13 Grupiranje vrednosti stvarnih znakov.

.131 Grupiranje vrednosti stvarno- atributivnih znakov je najtežje in posebno skoraj za vsak znak, Pri grupiranju vrednosti stvarno - atributivnih znakov je dano običajno načele grupiranja, po katerem se izvede sistematično grupiranje vrednosti znakov v obliki klasifikacij ali nomenklatur. Te enotne klasifikacije imajo zelo važno vlogo v našem sistemu evidenco in statistike. V naši praksi se uporablja cela vrsta enotnih klasifikacij in nomenklatur. Najvažnejše med njimi so: klasifikacija in nomenklatura blaga, klasifikacija proizvodov po osnovnih planskih grupah, začasna nomenklatura delavnosti, začasna nomenklatura poklicev, klasifikacija vzrokov smrti. Vse te klasifikacije so sistematizirane po principu grup in podgrup in tvorijo osnove sistematiziranja v svojem področju.

Pri grupiranju vrednosti stvarno atributivnih znakov nas večkrat zanimajo samo nekatere vrednosti, oziroma grupe vrednosti. V tem primeru vse vrednosti, ki ne spadajo v eno izmed teh grup, združimo zaradi popolnosti v grupo "ostalo".

.132 Pri grupiranju vrednosti stvarno-numeričnih znakov se moramo ozirati tako na vsebinsko plat proučevanja pojava kot na to, da grupacija zadošča tehničnim načelom grupiranja numeričnih znakov. Statistična analiza stvarno-numeričnih serij je namreč najbolj razvita in važna. Da pa moremo uporabiti te metode analizo, mora statistična serija izpolniti določene tehnične pogoje, ki jih upoštevamo, če le vsebinsko ni brez smisla.

Vrednosti stvarno-numeričnih znakov moremo zaradi njihove narave urediti po velikosti in odrediti njih mesto na številčni premici. V

zvezi s tem je treba vpeljati nekaj novih pojmov.

Grupa vrednosti stvarno - numeričnih znakov imenujemo razred. Najmanjšo vrednost znaka v razredu imenujemo spodnjo mejo razreda ($X_{k,min}$) največjo vrednost znaka v razredu pa zgornjo mejo razreda ($X_{k,max}$). Razliko pod zgornjo in spodnjo mejo imenujemo širino razreda.

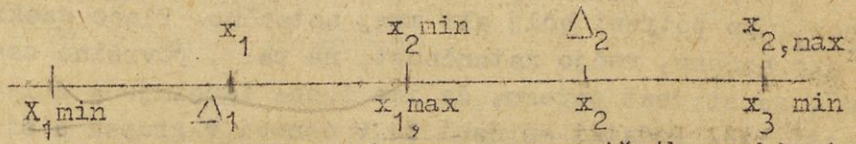
$$(\Delta_k = X_{k,max} - X_{k,min}),$$

polovično razliko vsote spodnje in zgornje meje pa sredino razreda

$$(\bar{X}_k = \frac{1}{2}(X_{k,min} + X_{k,max}).$$

V oklepajih so vpisani simboli, ki jih bomo še pogosto srečali. x je simbol za znak in sicer katerikoli, k je simbol za k-to grupo "min" označuje minimalno vrednost, max pa maksimalno vrednost $x_{k,min}$ pomeni torej minimalno vrednost znaka x v k-ti grupi.

Slika 4.1

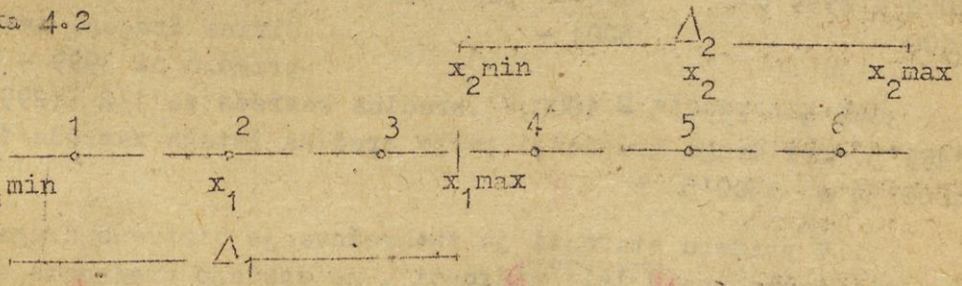


Osnovni pogoj,ki ga postavlja analiza stvarno - numeričnih serij, je enaka širini razreda pri vseh razredih $_k$ =

Pri nezveznih znakih je grupiranje razmeroma enostavno. Individualni vrednosti pripišemo interval polovico enote navzgor in navzdol. Tudi kot spodnjo mejo vzamemo formalno za polovico zmanjšano vrednost in kot zgornjo mejo za polovico zvečano vrednost.

1 - 3 otrok in ne	1 - 3 otrok
4 - 6 otrok	3 - 6 otrok
7 - 9 otrok	6 - 9 otrok

Slika 4.2



$3\ 1/2 - 1/2 = 3 = \Delta_1$ $\qquad$ $6\ 1/2 - 3\ 1/2 = 3 = \Delta_2$

$1/2\ [3\ 1/2 + 1/2] = 2 = X_1$ $\qquad$ $1/2\ [6\ 1/2 + 3\ 1/2] = 5 = X_2$ itd.

Pri grupiranju vrednosti zveznih znakov nastopi vprašanje, kam spada mejna vrednost, ki je pri zveznem znaku istočasno spodnja meja enega in zgornja meja naslednjega razreda. Ta vrednost more pripadati prvemu ali drugemu. Zaradi tega je treba to nedvomno naznačiti.

P Plače:

1.	2.	3.
pod 3.000 din	do 3.000 din	- 3.000 din
3.000 do pod 4.000 din	nad 3.000 " 4.000 din	3.000 din - 4.000 din
4.000 do pod 5.000 din	nad 4.000 " 5.000 din	4.000 din - 5.000 din
5.000 din dalje	nad 5.000 din	5.000 din -

V prvi grupaciji spada mejna vrednost v naslednji razred, v drugi grupaciji spada mejna vrednost v spodnji razred, v tretji pa to ni razvidno in je zaradi tega ta grupacija nepravilna.

V praksi pa za zvezne znake ni dana točna, ampak vedno zaokrožena vrednost, po potrebi bolj ali manj natančno. Plačo zaokrožujemo na dinarje, ali, ako hočemo, večjo natančnost, na pare, površino zaokrožujemo na are ali m^2, za starost povemo, da ima oseba izpolnjena 4 leta, ali 7 mesecev starosti itd. Podatki so dani že v osnovi v grupah s širino razreda, ki je enaka zaokroženi enoti.

P Ako zaokrožujemo n.pr. plačo v dinarjih, pomeni vrednost 2786 vse vrednosti razreda 2785'5 - 2786'5. Širina tega osnovnega razreda je torej 1, meje razreda pa dobimo, ako od dane vrednosti odštejemo in prištejemo 1/2 zaokrožene enote. Podobno kot za gornje grupacije dobimo za zaokroževane vrednosti:

1.	2.
- 2999 din	- 3000 din
3000 - 3999 din	3001 - 4000 din
4000 - 4999 din	4001 - 5000 din
5000 -	5001 -

Vendar moramo paziti, da je zgornja meja prvega razreda v prvi grupaciji 2999'5 din, v drugi pa 3000'50 din. Širina drugega razreda v prvem primeru ni 3999 - 3000 = 999 ampak 3999'5 - 2999'5 = 1000,- , sredina razreda pa 1/2 (2999'5 + 3999'5) = = 3499'5. Pri drugi grupaciji pa je sredina istega razreda 1/2 (3000'5 + + 4000'5) = 3500'5

P V primeru starosti je zaokroževanje bistveno drugačno. V grupo starosti izpolnjena 4 leta starosti , ne spadajo vrednosti od 3'5 do 4'5 temveč od 4 do pod 5 let. Starost pet let spada že v naslednji razred. Grupa od 0 do pod 1 leto označujemo z 0, pod 1 pod 2 1 itd. Grupa 0 - 3 torej pomeni starost od 0 do pod 4 leta itd. Tak način zaokroževanja se uporablja pri redkih znakih, je pa značilen za znak starost in starost običajno navajamo na ta način:

* razpredelje

P Primer: Grupacije starosti po petletkah:

0 - 4, 5 - 9, 10 - 14, 15 - 19, ...

V nekaterih primerih vsebina proučevanega pojava zahteva, da širina razreda ni stalna za vse razrede. Kadar proučujemo maso, pri kateri zavzemajo enote zelo majhne in zelo velike vrednosti, sorodnost vrednosti bolje izraža relativna razlika kot absolutna. V tem primeru uporabljamo razrede, katerih meje tvorijo geometrično postopico n.pr.: število delavstva do 5, 6 - 10, 11 - 20, 21 - 40, 41 - 80, 81 - 160, 161 - 320 itd.

Vsebinsko, tipološko grupiranje enot narekuje včasih, da tvorimo razrede vrednosti numeričnih znakov, ki ne zadoščajo nikakemu formalnemu kriteriju, ampak je grupiranje vrednosti znaka vsebinsko: n.pr. starostni kontingenti:

P 0 dojenčki, 1 - 6 predšolski otroci, 7 - 13 šoloobvezni otroci, 14 - 20 mladina, 21 - 64 odrasli, 65 starejši.

Statistične enote dane statistične mase zavzamejo za numerične znake vrednosti samo v nekem omejenem intervalu. Ker je pogostnost vrednosti na mejah zelo majhna, običajno vse te vrednosti združimo v odprte razrede in sicer "spodnji odprt razred" ali razred "pod" in "zgornji odprt razred" ali razred "nad". Primer znaka plača ima odprt razred "pod 3000 din" in "5000 dalje".

41.2 Grupiranje enot

.21 Enostavno in kombinirano grupiranje. Enote moremo grupirati po enem ali več znakih istočasno. V primeru grupiranja po enem znaku govorimo o enostavnem grupiranju enot, v primeru grupiranja po več znakih hkrati pa o kombiniranem grupiranju enot. Pri kombiniranem grupiranju enot se enote ene grupe vjemajo v kombinaciji vrednosti znakov, po katerih je kombinirano grupiranje izvedeno. Kombinirano grupiranje pomaga iskati odvisnosti v masovnih pojavih. Ako n.pr. gradivo popisa prebivalstva grupiramo samo po spolu, govorimo o enostavnem grupiranju, ako pa ga grupiramo po spolu in stanu hkrati, pa je to kombinirano grupiranje.

.22 Tipološko grupiranje enot. Celotna statistična masa predstavlja skupnost raznovrstnih enot. Grupiranje enot celotne mase v istovrstne grupe imenujemo tipološko grupiranje enot. Ime izhaja iz tega, ker združujemo enote, ki imajo iste tipične črte. Ker je razstavljanje mas na istovrstne grupe enot osnovne važnosti za možnost uporabe metod statistične analize, je jasno, da je tipološko grupiranje osrednjega pomena za vse statistično proučevanje. Istovrstnost enot je navezana na vrednosti, ki jih imajo znaki. Zaradi tega je potrebno, kadar hočemo izvršiti tipološko

grupiranje enot, izvršiti predhodno grupiranje vrednosti znaka ne formalistično, temveč vsebinsko. Grupe vrednosti znaka starost na 0, 1 - 6, 7 - 13, 14 - 20, 21 - 64, 65 - so osnova za tipološko grupiranje prebivalstva. Posestne skupine - 0'51 - 2'00 ha, 2'01 - 5'00 ha, 5'01 - 10'0 ha, 10'01 - 20'00 ha, 20'01 - 30'00 ha, 30'01 - 45'00 ha 45'01 - predstavljajo ogrodje za tipološko grupiranje gospodarstev. Večkrat tip ni določen z enim samim znakom, temveč z več znaki. Zaradi izključnega povdarka na vsebini je tipološko grupiranje zelo važno, na drugi strani pa zahteva popolno poznavanje predmeta in je možno le s predhodno podrobno vsebinsko analizo.

.23 Važnejše grupacije v naši statistični in ekonomski praksi.

Zahteve, ki se stavljajo na statistiko v naši družbeni stvarnosti v fazi izgradnje socializma, se očitujejo tudi v splošnih grupacijah masovnih pojavov iz socialno - ekonomskega življenja. Izražati morajo osnovne spremembe, ki so nastale ali se še dogajajo v naši stvarnosti, pomagati odkrivati zakonitosti družbenega in gospodarskega življenja.

Pri tem je najvažnejša razdelitev prebivalstva po socialnih skupinah. Enako je važna razdelitev najrazličnejših ekonomskih pojavov po sektorju lastništva. V proizvodnem področju je važna delitev na sredstva proizvodne (grupa A) in sredstva potrošnje (grupa B). Za potrebe planiranja grupiramo proizvodnjo po proizvodnih vejah in panogah in ministrstvih in administrativno - operativnih vodstvih. Enako ima velik pomen razdelitev po republikah in kompetencah (zvezni,republiški,lokalni značaj).

42. STATISTIČNE TABELE

42.0 Osnovno sredstvo prikazovanja statističnih podatkov je statistična tabela ali razpredelnica. Statistične podatke moramo prikazovati sistematično, pregledno, da je možno čim boljše koriščenje rezultatov, ki jih podatki predstavljajo. Zaradi razmeroma velikega števila podatkov, ki jih skušamo istočasno primerjati in analizirati, je važna naloga, da so podatki dani v taki obliki, da je to omogočeno.

42.1 Tehnično je tabela sestavljena iz sistema vrst in kolon-stolpcev, ki se križajo in ustvarjajo sistem okenc - polj, ki so namenjeni za vpisovanje podatkov. Tabela sestoji iz besednega - pojmovnega in številčnega dela. Besedni del posreduje vsebino podatkov, ki so podani v določenem okencu. Podatek v okencu se nanaša na pojme, ki odgovarjajo pripadajoči vrsti oz. stolpcu. Besedni oz. pojmovni del ima dva dela: glavo in čelo.

Sl. 4.3

shema tabele :

	GLAVA			
	vrsta			
		okence polje	kolona - stolpec	zbirni stolpec
telo				
	Zbirna vrsta			

shema tabele :

A / B	A_1		A_2		Sk. B
	A_{11}	A_{12}	A_{21}	A_{22}	
B_1					
B_2					
B_3			A_{21} B_3		B_3
B_4					
B_5					
Sk B			A_{21}		+

V okencu tabele 2 je podatek za pojm, ki je kombinacija pojmov B_3 in A_{21}.

V zbirnih vrstah oz. kolonah so podatki, ki se nanašajo samo glede na pojm A ali B, v kotnem koncu + pa podatek veljaven za celo opazovano maso.

42.2 Vrsta tabel.

Glede na to, kaj tabela prikazuje, delimo tabele na:

1) enostavne
2) sestavljene
3) kombinirane

.21 Enostavne tabele so tabele, ki prikazujejo eno samo statistično serijo. Enostavne tabele so tabele 3, 4, 6 in 7 v nadaljevanju odstavka.

.22 Dostikrat prikazujemo več serij, ki imajo vse isti osnovni znak. V tem primeru ni treba prikazati vsako serijo zase, temveč jih moremo prikazati skupno z enkratno navedbo osnovnega znaka. Take tabele, ki so sestavljene iz več enostavnih tabel, imenujemo sestavljene -primerjalne tabele. To služijo vsebinski primerjavi različnih podatkov:

Podatki popisa prebivalstva 15.3.1948

Tab. 4.1

L R	Število gospodinjstev
FLRJ	3,609.568
Srbija	1,485.744
Hrvatska	954.482
Slovenija	368.754
B i H	498.133
Makedonija	218.816
Črna gora	83.639

+

Tab. 4.2

L R	Število prebivalcev
FLRJ	15,772.107
Srbija	6,527.969
Hrvatska	3,756.807
Slovenija	1,391.873
B i H	2,565.283
Makedonija	1,152.986
Črna gora	377.189

=

Tab. 4.3

L R	Število	
	gospodinjstev	prebivalcev
F L R J	3,609.568	15,772.107
Srbija	1,485.744	6,527.969
Hrvatska	954.482	3,756.807
Slovenija	368.754	1,391.873
B i H	498.133	2,565.283
Makedonija	218.816	1,152.986
Črna gora	83.639	377.189

=

Drug način sestavljenih tabel dobimo, kadar prikazujemo dve seriji, ki imata različne osnovne znake in enake pomene vrednosti članov.

P Podatki popisa prebivalstva 15.3.1948 v F L R J

Tab. 4.4

S P O L	Število prebivalstva
Skupno FLRJ	15,772.107
Moški	7,582.476
Ženski	8,189,631

+

Tab. 4.5

VZDRŽEVANOST	Število prebivalstva
Skupno FLRJ	15,772.107
Aktivno	9,784.602
Pasivno	5,987.505

=

Tab. 4.6

6	Število prebivalstva
Skupno F L R J	15,772.107
Po spolu:	
Moški	7,582.476
Ženski	8,189.631
Po vzdrževanosti:	
Aktivni	9,784.602
Pasivni	5,987.505

Navedena primera sta najenostavnejši sestavljeni tabeli. Sestavljena tabela more biti večkrat sestavljena in je še kljub temu bolj pregledna, kot da bi podatke navedli v enostavnejših sestavljenih tabelah.

.23 Kombinirane-razčlenstvene tabele uporabljamo za prikazovanje rezultatov urejevanja kombiniranega grupiranja po več znakih. Tako dobimo kombinirane tabele z dvema vhodoma, ako kombiniramo dva znaka, s tremi vhodi, ako, kombiniramo tri znake itd.

Kot primer kombinirane tabele podajamo število prebivalstva v FLRJ, razdeljeno istočasno po republikah, spolu in vzdrževanosti.

P Tab. 4.7

Število prebivalstva po popisu 15.3.1948
v tisoč prebivalcev

Vzdrževal. Spol / LR	Aktivni			Pasivni			Skupno		
	M	Ž	SK	M	Ž	SK	M	Ž	SK
Srbija	2146	2028	4174	1025	1329	2354	3171	3357	6528
Hrvatska	1231	1244	2475	528	754	1282	1759	1998	3757
Slovenija	449	413	862	204	326	530	653	739	1392
B i H	718	722	1440	519	606	1125	1237	1328	2565
Makedonija	352	274	626	232	295	527	584	569	1153
Črna gora	100	109	209	78	90	168	178	199	377
FLRJ Skupno	4996	4789	9785	2586	3400	5981	7582	8190	15772

42.3 Vrsto tabel po namenu. Tabele se po namenu uporabe dele v obdelovalne in analitične - končne tabele.

Obdelovalne tabele služijo kot sredstva za vpisovanje rezultatov urejevanja statistilnih podatkov. Tehnika njihovega sestavljanja oziroma uporabe je zato prilagojena sistemu obdelave (glej odstavek o obdelovalnih tabelah).

Cilj analitične tabele pa je, da v čim preglednejši obliki poda

vsebino prikazovanega masovnega pojava. Sestavljena mora biti tako, da je mogoča čim boljša primerjava podatkov, ki so v smiselni zvezi. Kljub temu, da mora prikazati pojav kompleksno, mora biti jasna in pregledna. Da dosežemo ta cilj, morajo biti vse tabele sestavljene po nekih okvirnih splošnih načelih. To je na eni strani potrebno zaradi pravilne sestave, na drugi strani pa se je v tabelah, ki so sestavljene po istih načelih, lažje spoznati.

Osnovna načela analitičnih tabel so:

Tabela mora biti:

a) razumljiva. To dosežemo a.) s pravilnim in smiselnim naslovom, b.) navedbo vira, c.) kratkim, jasnim in smiselnim navajanjem pojmov v tekstualnem delu tabelo, d.) z navedbo enote mere, v vseh primerih, kjer je to potrebno, e.) s kratkimi pojasnili k posameznim podatkom. Pojasnila postavimo pod tabelo. Z znamenji navedemo, h kateremu podatku pojasnilo spada.

b) pregledna. Ta pogoj, ki je zelo važen, dosežemo s tehničnimi postopki sestavljanja tabel.
   a. Preobširne tabele bomo raje razdelili v več samostojnih tabel.
   b. Uporabljali bomo različen tisk za različne vrste podatkov.
   c. Z oštevilčenjem vrst in kolon se bomo v tabeli lažje razpoznali. Oštevilčene morajo biti vrste in kolone v začetku in na koncu.
   d. S smiselno uporabo tanjših in debelejših črt, ki se ravna bodisi po vsebini (več pojmov, ki spadajo skupaj, damo med debelejše črte) ali formalno (vsaka tretja ali peta črta debelejša) zvečamo preglednost.

c) enotna. Ta pogoj se tiče odnosov : med tabelami.
   a. Tabele, ki se tičejo istih pojavov, naj bodo sestavljene po enotnih načelih oznamenovanja in podajanja podatkov.
   b. Ker ne sme biti v tabeli nobeno polje prazno, moramo uporabljati vedno iste dogovorne - konvencionalne znake. Najvažnejši so:

.	podatek ne pride v poštev
. .	s podatkom ne razpolagamo
( )	začasen predhoden podatek
-	vrednost enaka nič
0 0'0	vrednost manjša od polovice uporabljene enote.

5. UREJEVANJE IN OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

50. SPLOŠNO. Prva faza obdelave zbranega in kontroliranega statističnega gradiva je urejevanje. Pod urejevanjem razumemo grupiranje in preštevanje enot in seštevanje vrednosti znakov po grupah. Cilj urejevanja so absolutni podatki o statistični masi oziroma proučevanem masovnem pojavu. S tem postopkom dobimo pregled nad notranjim sestavom statistične mase in odvisnosti v masovnih pojavih. Rezultati urejevanja so dani v tabelah v obliki statističnih serij. Čeprav so že sami absolutni podatki za proučevanje masovnega pojava zelo važni in za nekatere potrebe neobhodno potrebni in zadostni, se more izluščiti vse značilnosti masovnega pojava šele z nadaljnjimi metodami analize masovnih pojavov. Kot je rezultat statističnega opazovanja - zbrani obrazci, surov material glede na urejevanje, so serije z absolutnimi podatki dobljenimi z urejevanjem, surovina glede na analizo masovnega pojava.

Cilj urejevanja so torej statistične serije, ki že same na sebi nudijo vpogled v sestav, odvisnosti in zakonitosti ali pa nudijo gradivo, da z nadaljnjo analizo te značilnosti odkrijemo. Ta cilj je podan že s samim programom statističnega proučevanja in je vsebinska plat urejevanja z njim dana. Program določi cilj in splošne smernice.

51. **predpriprave izvedbe urejevanja**

51.1 Seznami vrednosti in grupnih vrednosti znakov. Rezultat statističnega urejevanja so statistične serije. Ker je, kot bomo kasneje slišali, statistična serija niz istovrstnih statističnih količin, katerih vsaka velja za eno vrednost ali grupo vrednosti določenega znaka, so potrebne stat. serije kot neobhodna predpriprava urejevanja nedvoumne nomenklature vseh vrednosti znakov, ki pridejo v poštev pri urejevanju. Ako gre za grupne vrednosti znakov, ki so določene z samo vsebino pojava, je potrebno navesti tudi, katere osnovne vrednosti spadajo v dano grupo vrednosti znaka.

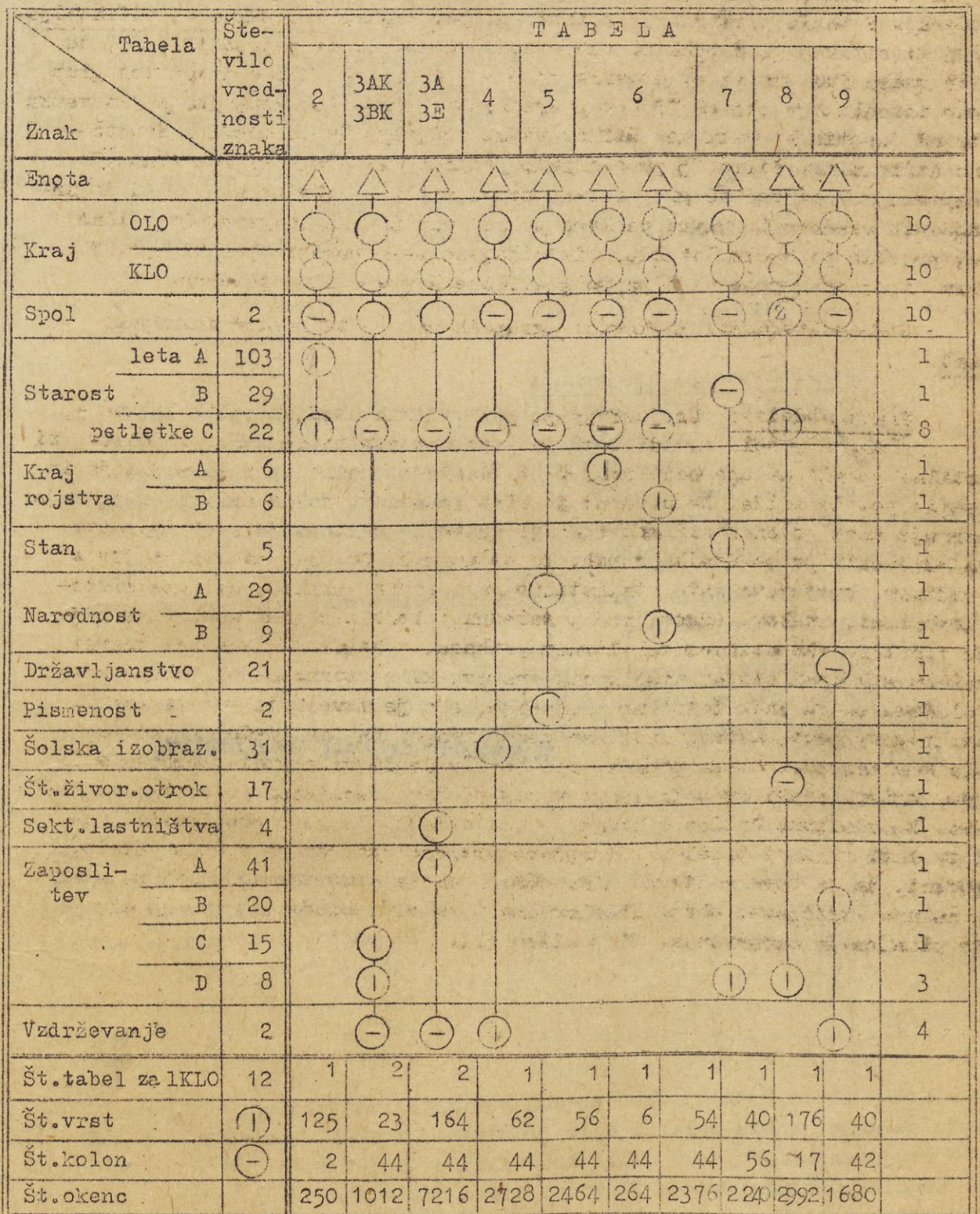
52.2 Šifre vrednosti, oziroma grup vrednosti znakov. Tehnika urejevanja priporoča, da se vrednostim znakov in grupnim vrednostim znakov pripišo kratko označbo, ki jo imenujemo šifro. Vsaka vrednost znaka ali grupna vrednost znaka ima svojo šifro. To znatno olajša delo, ker šifra nadomesti dolgo vpisovanje vrednosti. Kot šifre moremo uporabljati katere koli označbe (n.pr. +, 0 itd.) Običajno uporabljamo ali črkovne ali številčne šifre. Črkovne šifre se večkrat uporabljajo pri ročnem urejevanju. Ako so šifre

začetne črke grupnih vrednosti, si jih zelo lahko zapomnimo in so radi tega zelo uporabne. Številčne šifre so zelo uporabne, posebno za numerične znake in za znake, katerih vrednosti so urejene v kompliciran sistem grup in podgrup. V takih primerih je posebno priporočljiva decimalna klasifikacija. Pri tej klasifikaciji razdelimo vse vrednosti znakov na največ 10 grup, od katerih vsaka ima eno izmed številk od 0 - 9 za šifro. Vsako izmed teh grup ponovno razdelimo v največ 10 grup vrednosti. Izmed teh grup ima zopet vsaka eno izmed desetih številk kot šifro. Vsaka podgrupa je označena z dvoštevilčno šifro n.pr. šifra 25 pomeni peto podgrupo iz druge glavne grupe. To razdeljevanje v največ 10 grup moremo nadaljevati. S tem moremo priti do individualnih vrednosti. Šifra pa lepo pokaže sistemizacijo grup. Številčne šifre, posebno pa decimalna klasifikacija, se zelo pogosto uporabljajo. V primeru strojnega urejevanja pa so številčne šifre odlično uporabne.

Seznam vrednosti znakov oz. grupnih vrednosti znakov imenujemo šifrant.

51.3 Načrt obdelave. Urejevanje je pri velikih akcijah velik in kompleksen posel. Zaradi tega je potreben predhodni načrt obdelave, ki shematično pokaže naloge obdelave. Na ta način je možno izbrati najlažjo in najkrajšo pot do cilja. Že uvodoma je bilo omenjeno, da je naloga urejevanja grupiranje enot po enem ali kombinaciji znakov, preštevanje enot in seštevanje vrednosti po poedinih grupah. Da se osvetli določen masovni pojav z več vidikov, urejujemo isto statistično maso po več znakih, oziroma sistemih kombinacij znakov. Odnosi med posameznimi kombinacijami znakov so najlepše vidni na shematičnem tabelarnem prikazu, v katerem v poedini koloni prikažemo s krogci sistem kombinacij znakov, ki se povezujejo. Vsaka vrsta je določena za en znak določene grupacije, ako je izvedenih več grupacij. Mesto krogcev pove, kateri znaki se kombinirajo. En tak sistem znakov se obdela kombinirano. Shema pokaže, ali je mogoče že en sistem vsebovan v drugem, kateri znaki se pojavljajo v največ kombinacijah, od česar zavisi obdelava. Z vodoravno črtico v krogu ⊖, naznačimo, da je v obdelovalni tabeli ta znak v glavi tabele, z navpično črtico ⊕ pa, da je v čelu tabele. △ pomeni, da se znak sešteva. Z navedbo števila grupnih vrednosti za vsak znak moremo izračunati obseg obdelovalne tabele. Ta shema nudi važne elemente planiranja urejevanja. Glej sliko 5.1.

Sl. 51 Shema obdelovalnih tabel popisa prebivalstva leta 1948



51.4 Obdelovalne tabele: Rezultate urejevanje vpisujemo v obliki serij obdelovalne tabele. Obdelovalne tabele so na osnovi sheme obdelave sestavljene kombinirane tabele. En sistem kombiniranih znakov se kombinira v eni tabeli. Tehnika in načela sestavljanja obdelovalnih tabel je drugačna kot sestavljanje končnih, analitičnih. Obdelovalna tabela služi le enemu cilju - obdelavi in je zaradi tega njih sestavljanje značilno. Sistem obdelovalnih tabel mora biti tak, da je obdelava čim lažja in čim hitrejša. Zaradi tega je sestavljanje obdelovalnih tabel odvisno od načina obdelave. Obdelovalne tabele so včasih zelo obsežne, imajo poleg grupnih vrednosti vpisane šifre ali celo samo šifre, kar pri končnih tabelah ni dopustno itd.

52. VRSTE UREJEVANJ:

52.1 Tehnični načini urejevanja

Od razpoložljivih tehničnih sredstev je zelo odvisen način, hitrost in ekonomičnost urejevanja. Pri vseh načinih urejevanja moramo izvesti navedene predpriprave, vključno šifriranje gradiva opazovanja. Urejevanje oz. prva obdelava sestoji iz dveh različnih delov: preštevanja in seštevanja po grupah. Vsak izmed teh delov se oslanja na isto predhodno fazo, grupiranje enot v grupe.

Glede na tehnična sredstva delimo urejevanje na : ročno urejevanje in strojno urejevanje.

.11 Pri ročnem urejevanju uporabljamo več metod: črtkanje obstoji v da za določen znak pregledamo vse enote in za grupno vrednost znaka vsake enote napravimo v pripadajočem polju obdelovalne tabele črtico ali piko.

Tab. 5.1

Stan \ Spol	moški		ženske	
samski	𝍸 𝍸 ||	12	𝍸 𝍸 |||| |	16
poročen	𝍸 ||||	9	𝍸 𝍸 |	11
vdovec	|||	3	||||	4
razveden	|	1	||	2

S preštetjem črtic ali pik v posameznem polju dobimo število enot z isto grupno vrednostjo. (glej primer 1.) Glede na lažje preštevanje črtic imamo različne sisteme črtkanja:

𝍸 = 5 ☒ = 10

Metoda črtkanja ima precej pomanjkljivosti. Po izvršenem urejevanju po nekem znaku obrazci niso grupirani, temveč v istem neredu kot pred urejevanjem. To zelo ovira kombinirano urejevanje po več znakih hkrati. Kombinirano urejevanje je možno edino na ta način, da črtice vpisujemo neposredno v kombinacijsko tabelo. Pri velikem številu grupnih vrednosti in več znakih je tak postopek izredno težaven. Pri črtkanju so zelo pogoste

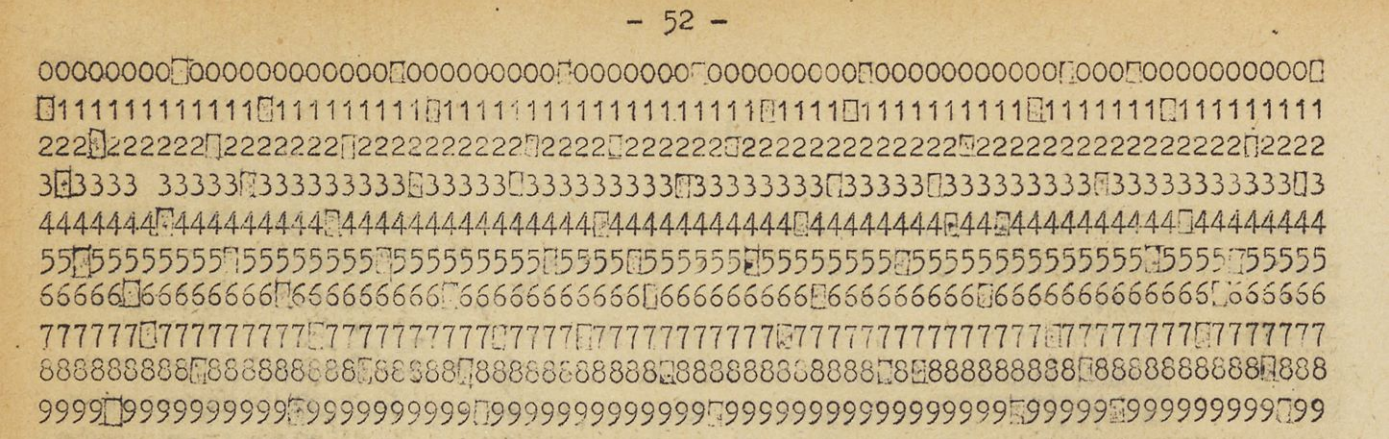
napake, ki pa jih kasneje ni mogoče odkriti. Z aradi tega se metoda črtkanja običajno uporablja za urejanje po onem znaku in v primeru, da število enot ni veliko.

.12 Metoda kuponov obstoji v tem, da imamo za vsako grupno vrednost znaka svoj blok z numeriranimi kuponi. Kuponi za različne grupne vrednosti se med seboj razlikujejo, bodisi po barvi ali oznakah. Kuponi so gumirani. Preštevanje po grupah se vrši na ta način, da dani grupni vrednosti ene enote odgovarjajoč kupon nalepimo na obrazec. Izredne številke zadnjega odtrganega kupona posameznega bloka odberemo število enot v posamezni grupi. Ta način je razmeroma drag (bloki) in se ne da uporabiti za kombinirano urejevanja in gradivo ostane po izvršenem urejevanju negrupirano. Zaradi tega se redko uporablja.

.13 Odlaganje listkov je izmed ročnih urejevanj najprimernejše in se najpogosteje uporablja. Delo sestoji iz treh faz: šifriranja, sortiranja in preštevanja. Sortiranje izvedemo tako, da razdelimo popisne liste v grupe po grupnih vrednostih znaka. Ker moremo vsako grupo razdeliti - razsortirati po novem grupnem znaku, moremo izvršiti sortiranje po kombinaciji poljubnega števila znakov. Možnost napak pri tem načinu je majhna in še te napake je možno odkriti. Po tem načinu je možno tehnično izvesti vsako urejevanje. Rezultate urejevanja dobimo s tem, da preštejemo število popisnih listov v vsaki grupi.

.14 Seštevanje vrednosti znakov po grupah izvedemo, bodisi s seštevanjem na pamet ali z računskimi stroji. Posebno kadar gre za seštevanje podatkov iz individualnih listov je priporočljiva uporaba računskih strojev, ker s tem odpade predhodno prepisovanje podatkov. Ako strojev nimamo, se poslužujemo različnih tehničnih olajšav kot plisiranja obrazcev, polaganja obrazcev drugega preko drugega, ako so podatki ekstenzivnega značaja nanešeni ob robu itd.

.15 Strojna obdelava: Urejevanje statističnega gradiva je "grlo" opazovanja masovnih pojavov. Ker je število enot običajno pri večjih akcijah zelo veliko in gre v stotisoče in milijone, urejevanje pa po svoji naravi in potrebi strokovnega osebja navezano na razmeroma majhno število ljudi, se vrši obdelava zelo dolgo. Ker rezultati obdelave služijo svojemu namenu včasih šele čez par let, je vprašanje, ali so podatki sploh še uporabni in važni. Zaradi tega je težnja po skrčenju časa obdelave statističnega gradiva dovedla do tega, da so izumili stroje, ki hitro in zanesljivo izvedejo urejevanje in seštevanje statističnih podatkov. Strojna obdelava je zasnovana na sistemu posebnih kartic, na katere so podatki prenešeni v številčnih šifrah v obliki luknjic.



Posebni stroji registrirajo na to, na katerem mestu so luknjice in izredno hitro izvrše sortiranje in preštevanje kartic - enot (do 20000 kartic na uro), in tudi seštevanje numeričnih znakov. Obdelava na strojih se vrši po naslednjih fazah: a šifriranje osnovnega gradiva s številčnimi šiframi, b) perforiranje - luknanje kartic na posebnih strojih perforirkah, kar pomeni, prenašanje šifer s sistema luknjic na kartice. Posamezna šifra ima na kartici točno določeno mesto - kodeks. Pri strojih sistema Pauers imamo na kartici prostora za 45 enomestnih šifer, pri strojih sistema Pauers imamo na kartici prostora za 45 enomestnih šifer, pri strojih sistema Hollerith pa za 90 enomestnih šifer. c) Verificiranje - kontrola perforiranja se izvrši na spcialnih strojih - verificirkah, ki so podobni perforirkam in je delo na njih zelo podobno perforiranju. d) Sortiranje se izvrši s pomočjo strojev sortirk, ki po danem znaku kartice razsortirajo in obenem na posebnih števcih pokažejo število enot v posameznih grupah. Ti stroji izvrše obenem sortiranje in preštevanje. e) Tabeliranje se izvede na strojih tabelirkah. Ti stroji izvajajo hkrati seštevanje več numeričnih podatkov.

Pri sistemu strojne obdelave traja najdalj perforiranje kartic, ki je v stvari prepisovanja podatkov na kartico in sicer v šifrah. Da bi skrčili tudi to fazo, so izdelali nov stroj, ki sam perforira kartice. Osnovne podatke ne vpisujemo v statistične obrazce, temveč direktno na kartice v obliki črtic na odgovarjajočih mestih. Stroj, ki je občutljiv za te črtice, napravljene s svinčnikom, sam perforira kartice. S tem odpadejo individualni obrazci, ki jih nadomeste kartice. Pri tem načinu se čas perforiranja zelo skrči, ker ga izvrši stroj mehanično.

53. UREJEVANJE GLEDE NA VRSTO UPORABLJENIH OBRAZCEV.

53.0 Način urejevanja statističnega gradiva je odvisen od vrste obrazca, ki se uporablja kot sredstvo opazovanja. Nalogo moremo postaviti tudi obratno, da je treba predvideni obdelavi izdelati primerne obrazce.

Pri obravnavanju statističnih obrazcev smo videli, da moremo izvršiti opazovanje z individualnimi ali kolektivnimi obrazci. O splošni prednosti in pomanjkljivosti enega ali drugega sistema smo že govorili. Pogledali bomo še prednosti in pomanjkljivosti glede na posamezne faze urejevanja. Te prednosti in pomanjkljivosti pridejo v poštev predvsem pri ročnem urejevanju. Pri strojnem načinu razlika v sistemih obrazcev ne pride dosti v poštev, ker obrazci služijo samo kot osnova za perforiranje in se vse ostale faze vrše s karticami.

53.1 Individualni obrazci imajo glede na to, da so za vsako enoto podatki vpisani na posebnem obrazcu, velike prednosti pri sortiranju in preštevanju. Seštevanje vrednosti znakov je pri sistemu individualnih obrazcev razmeroma nerodno, posebno, ako seštevamo brez računskih strojev, na pamet. Sestavljanje obrazcev tako, da so vrednosti ekstenzivnih znakov nanešene ob robu obrazca, plisiranje obrazcev ali prepisovanje podatkov na pomožne obrazce, so tehnični postopki, ki omogočajo ali pomagajo pri sistemu individualnih obrazcev izvrševati seštevanje ekstenzivnih znakov.

Iz gornjega sledi, da individualne obrazce uporabljamo predvsem v primerih, kadar opazovanje vsebuje predvsem sortirne znake (n.pr. popis prebivalstva).

53.2 Kolektivni obrazci so z vidika sortiranja enot neprimerni. Enote, nanizane v kolektivnem obrazcu, se ne dajo direktno urejevati. Vendar zaradi drugih prednosti, ki jih imajo, dostikrat uporabljamo kolektivne obrazce. Da pa kljub temu moremo statistične enote urediti, uporabljamo različne postopke. Sortiranje enot po nekaj osnovnih znakih se more izvršiti že pri izvedbi samega opazovanja tako, da ima popisovalec za posamezno grupo enot samostojen obrazec, v katerega vpisuje samo enote te grupe. S tem je izvršeno sortiranje že na samem terenu. Seveda je to možno, če je potrebno izvesti sortiranje le po par osnovnih znakih. Zaradi samega sistema popisovanja (popisni okoliši), je s kolektivnimi obrazci vedno možno dobiti regionalno razdelitev masovnega pojava. Ta način se uporablja pri nas posebno v kmetijski statistiki, kjer gospodarstva razdelimo po sektorju, privatna pa še po velikostnih razredih.

Drugi način, ki ga moremo uporabljati, je, da podatke za posamezne enote prepišemo iz kolektivnih obrazcev na individualne obdelovalne listke. Ta prepis se običajno izvrši v šifrah, zaradi tega morejo imeti ti listki zelo primerno obliko za sortiranje. Kljub temu, da izgleda način na prvi pogled nekoliko neprikladen, ima določene prednosti in ga moremo s pridom uporabljati.

Manjša urejevanja pa se dajo iz kolektivnih obrazcev zelo lahko izvesti z uporabo črtkanja.

Z vidika seštevanja vrednosti znakov pa imajo kolektivni obrazci veliko prednosti pred individualnimi. Podatki so za posamezne enote v kolektivnem obrazcu nanizani drug pod drugim, kar zelo olajša seštevanje (n.pr. popisi živine, orodja).

Iz gornjega napravimo sklep, da je glede na ročno obdelavo gradiva primeren kolektiven obrazec takrat, kadar vsebuje opazovanje predvsem adirne znake, katerih vrednosti moramo seštevati, ne pa preštevati.

54. VRSTE UREJEVANJ V ORGANIZACIJSKEM POGLEDU.

Urejevanjemmoremo organizacijsko vršiti na enem mestu - centralizirano, ali na večih mestih - decentralizirano. Včasih uporabljamo mešan sistem, pri katerem se urejevanje vrši delno centralizirano, delno decentralizirano.

54.1 Sistem centraliziranega urejevanja ima velike tehnične prednosti kot so: možnost izvedbe dobre organizacije dela, večje specializacije in usposabljanja osebja, enotnejše obdelave in kontrole in dobra izraba računskih in statističnih strojev. Zaradi tega izvajamo centralizirano urejevanje pri velikih akcijah.

54.2 Vendar sistem opazovanja ali druge prilike včasih narekujejo, da izvedemo urejevanje decentralizirano. Eden izmed primerov so ravno popisi v kmetijstvu s pomočjo kolektivnih obrazcev. Ta sistem je vezan na to, da popisovalci izvedejo sortiranje enot in seštevanje ekstenzivnih znakov že pri samem opazovanju, torej decentralizirano. Z decentraliziranim urejevanjem pa prevzamemo vse tveganje zaradi težkega enotnega metodološkega vodstva, težke organizacije dela, kontrole in podobno. Vendar je prednost tega sistema v tem, da se delo razdeli na širši krog ljudi in je zaradi tega hitrost obdelave večja.

54.3 Mešani sistem se običajno uporablja tako, da se decentralizirano izvede enostavna obdelava, ki se zaradi hitre izvedbe, vendar kot nezanesljivi uporabljajo kot predhodni podatki. Celotno obdelavo pa izvedemo centralizirano. Kot osnova kratke decentralizirane obdelave služijo večkrat kontrolniki, ki vsebujejo osnovne podatke o vseh enotah.

6. STATISTIČNE SERIJE

61. KAJ JE STATISTIČNA SERIJA ?

Z obdelavo osnovnega statističnega gradiva, grupiranjem in urejevanjem dobimo rezultate obdelave dane v absolutnih številih. En sam sumaren podatek, čeprav sam zase zelo važen, pa za proučevanje pojavov kljub temu ni dovolj. Zaradi tega podajamo in proučujemo navadno celo vrsto istovrstnih podatkov v takozvanih statističnih serijah naenkrat.

Statistična serija je vrsta istovrstnih količin, od katerih se vsaka nanaša na določeno vrednost oziroma grupo vrednosti enega in istega znaka. Statistična serija podaja kakovostno stran masovnega pojava v odvisnosti od znaka, na katerega se podatki nanašajo. V tabeli(1) navajamo primer enostavne serije

Tabela 6.1

Število prebivalstva v FLRJ po sektorju lastništva (po popisu 15. marca

Sektor lastništva	Število prebivalstva
Državni	3,431.541
Zadružni	408.097
Družb.org.	26.576
Privatni	11,905.893
Skupno	15,772.107

Vsaka statistična serija ima naslednje sestavne dele: Osnovni znak serije (sektor lastništva) je znak, na katerega vrednosti se nanašajo posamezni podatki. Na isti osnovni znak se morejo nanašati različne količine, zato je za serijo značilen pomen vrednosti prikazanih količin (število prebivalstva). Člen statistične serije je del serije, ki se nanaša na eno vrednost ali grupo vrednosti osnovnega znaka (n.pr. zadružni sektor, število prebivalstva 408.097) vrednost člena pa je številčen iznos posameznega člena serije (n.pr. 408.097 prebivalcev).

63. DELITEV STATISTIČNIH SERIJ PO ZNAČAJU OSNOVNEGA ZNAKA SERIJE

Po značaju časovnega znaka delimo statistične serije enako kot znake na:

a) časovne serije, ki se dele na:
   - aa) momentne, pri katerih se členi serije nanašajo na časovne momente. Z njimi prikazujemo podatke o momentnih masah (glej tabelo 6.2).
   - ab) intervalne, pri katerih se členi serije nanašajo na časovne intervale. Z njimi prikazujemo podatke o intervalnih masah (tabela 6.3).

b) Krajevne - teritorialne serije, kjer so podatki razdeljeni po po teritorialnih enotah - grupah (Tabela 6.7).

c) Stvarne serije delimo enako kot znake na:

ca) stvarno - atributivne serije: Nepismeno prebivalstvo v procentih od skupnega števila nad 10 let starega prebivalstva po narodnosti (Tabela 6.4).

cb) stvarno -numerične serije: Te pa na:

cba) zvezne n.pr. število zemljiških gospodarstev po velikostnih skupinah v LRS (Tabela 6.5).

cbb) nezvezne (število gospodinjstev po številu članov, (tabela 6.6).

Posamezni členi nezveznih stvarno - numeričnih serij imajo momenten značaj, ako vrednosti osnovnega znaka niso grupirane (do 15v tab. 6.6 primeru pa, da so grupirane, imajo členi intervalen značaj, ker se podatki nanašajo na interval (16-20, 21-25, 26 in več v primeru Tabela 6.6).

Tabela 6.2

Število prebivalstva v Sloveniji (od 1890 do 1931 področje b.dravske banovine od 1948 do 1950 področje LRS)

Datum	Število prebivalcev v 1000
1890	1.002
1900	1.026
1910	1.067
1921	1.060
31.3. 1931	1.144
15.3 1948	1.392
30.6. 1949	1.409
30.6. 1950	1.422

Tabela 6.3

Število natovorjenih vagonov v bivši Jugoslaviji v letu 1933

Mesec	Število natovorjenih vagonov
J	89,0
F	86,6
M	109,7
A	97,1
M	110,8
J	106,7
J	109,5
A	127,6
S	152,6
O	152,2
N	130,6
D	106,0

Tabela 6.4

Nepismeno prebivalstvo v FLRJ po narodnosti po popisu 15.3.1948. (v °/o od skupnega števila nad 10 let starega prebivalca

Narodnost	Procent nepismenih
Skupno	25,43 °/o
Srbi	27,71 "
Hrvati	18,11 "
Slovenci	2,32 "
Makedonci	30,24 "
Črnogorci	24,13 "
Muslimani	54,26 "

Tab. 6.5

Število zemljiških gospodarstev v LRS po popisu zemljišč 1947

Velikostna skupina po skupni površini	Število gospo - darstev
S k u p n o	211.299
Brez lastne zemlje	8.334
0,01 - 0,50 ha	37.371
0,51 - 1,00 "	17.269
1,01 - 2,00 "	24.182
2,01 - 3,00 "	18,340
3,01 - 5,00 "	26.184
5,01 - 8,00 "	24.200
8,01 - 10,00 "	11.234
10,01 - 15,00 "	19.036
15,01 - 20,00 "	10.540
20,01 - 30,00 "	8.442
30,01 - 45,00 "	3.595
45,01 -	2.572

Tab. 6.6

Število gospodinjstev po številu članov v LRS po popisu 15. marca 1948

Število članov	Število gospodinjstev
S k u p n o	368.754
1	68.027
2	55.376
3	63.309
4	60.773
5	46.183
6	31.247
7	19.290
8	11.396
9	6.309
10	3.230
11	1.651
12	898
13	284
14	126
15	73
16-20	97
21-25	71
26-	408
Neznano	6

Tab. 6.7

Število prebivalstva v FLRJ po LR po popisu 15. marca 1948

L R	Število prebiv. v 000
S k u p n o	15.752
Srbija	6.583
Hrvatska	3.749
Slovenija	1.389
B i H	2.562
Makedonija	1.152
Črna gora	377

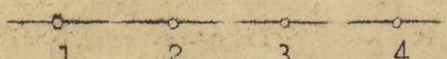
63. PRIPOMBE O ČLENIH STATISTIČNE SERIJE

Za nadaljnje proučevanje statističnih serij, posebno časovnih, je važno naslednje: vrednost členov intervalnih serij pripišemo sredini intervala. V seriji tabele 6.5 pripišemo 37.371 gospodarstev vrednosti (0,50 + 0,01) : 2 = 0,255 ha. 17.269 vrednosti (0,51 + 1,00) : 2 = 0,755 ha itd. Vrednost 97 v tabeli 6.6 pripišemo vrednosti (16 + 20) : 2 = 18 itd.

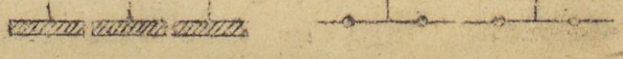
Obratno pa vrednosti momentnih serij, oziroma členov pripišemo interval, običajno v velikosti 1/2 širine razreda levo in desno od vrednosti. Tako se vrednosti členov v seriji tabela 6.6 nanašajo na interval 0,5 - 1,5, 1,5 - 2,5 itd. To olajša analizo in odrejanje intervala, pri morebitnem grupiranju momentnih členov (n.pr. v tabeli 6.6 se 73 gospodinjstev nanaša na interval 14,5 - 15,5, 97 pa na interval 15,5 - 20,5).

Pri zveznih numeričnih znakih, kjer je bilo izvedeno zaokroževanje vrednosti znakov že pred grupiranjem, je treba paziti, da so intervali dejansko na levo in desno za 1/2 zaokrožene vrednosti večji kot je navedeno v grupaciji (glej tabela 6.5). 0,01 - 0,50 se dejansko nanaša na 0,005 - 0,505, 0,51 - 1,00 na interval 0,505 - 1,005 itd. To sledi iz tega, ker so na 0,01 zaokrožene vse vrednosti v intervalu 0,005 - 0,[illegible] itd. Tudi širina razreda ni 0,50 - 0,01 = 0,49 ha, temveč 0,505 - 0,005 = 0,50.

Sl. 6.1 Pripisovanje intervala momentnim serijam



Sl. 6.2 Pripisovanje sredine intervala intervalnih serij



64. DELITEV STATISTIČNIH SERIJ PO ZNAČAJU VREDNOSTI ČLENOV

Značaj vrednosti členov statističnih serij more biti različen. Vrednosti členov morejo biti rezultat preštevanja enot po vrednostih ali grupah vrednosti (n.pr. serija tabela 6.5, tabela 6.6, tabela 6.7). Taka serija pokaže, koliko enot statistične mase ima posamezne vrednosti ali grup vrednosti osnovnega znaka. Take serije imenujemo zaradi tega variacijske serije ali frekvenčne distribucije. Njih proučevanje je zaradi važnosti variiranja znakov izredno važno. Podrobneje o frekvenčnih distribucijah je govora v odstavku 6.6.

Vrednosti členov pa morejo biti tudi rezultat seštevanja vrednosti znakov vseh enot, ki imajo grupno vrednost. Znak, ki ga seštevamo more biti osnovni znak serije (n.pr. površina po velikostnih skupinah) ali pa tudi kak drug znak (n.pr. število živine po velikostnih skupinah). Pri teh vrstah serij je važno, da moremo vrednosti posameznih členov med seboj seštevati, imajo kot pravimo, ekstenziven značaj. Tudi členi variacijske

serije imajo ekstenziven značaj.

Vrednosti členov pa morejo biti tudi količine, ki so rezultat nadaljne statistične obdelave in analizirajo pojav globlje kot same absolutne vrednosti. To so n.pr. relativna števila, srednje vrednosti itd. Te vrednosti niso med seboj seštevljive, imajo intenziven značaj (n.pr. tabela 6.4, kjer so navedeni strukturni odstotki).

65. KUMULATIVNE SERIJE

Iz numerične serije, katere členi imajo ekstenziven značaj, moremo izpeljati serijo, ki jo imenujemo kumulativno serijo. Pri frekvenčni distribuciji člen kumulativne serije pove, koliko enot ima vrednost nad ali pod določeno vrednostjo. Tvoriti moremo torej dve kumulativni seriji in sicer: kumulativno serijo imenovano "pod" in kumulativno serijo imenovano "nad". Kumulativno serijo dobimo s postopnim prištevanjem posameznih členov prvotne serije.

Gornjim pogojem, ki jih mora izpolnjevati statistična serija, da moremo tvoriti kumulativno serijo, zadošča serija v Tabeli 6.5

Tab. 6.8

Velikostna skupina	Število gospodarstev
Brez lastne zemlje	8.334
0,01 - 0,50	37.371
0,51 - 1,00	17.269
1,01 - 2,00	24.182
2,01 - 3,00	18.340
3,01 - 5,00	26.184
5,01 - 8,00	24.200
8,01 -10,00	11.234
10,01 -15,00	19.036
15,01 -20,00	10.540
20,01 -30,00	8.442
30,01 -45,00	3.595
45,01 -	2.572

Tab. 6.9

Kumulativna serija " pod "

Velikostna skupina	Število gospodarstev
Pod 0,0	8.334
" 0,50	45.705
" 1,00	62.974
" 2,00	87.156
" 3,00	105.496
" 5,00	131.680
" 8,00	155.880
" 10,00	167.114
" 15,00	186.150
" 20,00	196.690
" 30,00	205.132
" 45,00	208.727
	211.299

Kumulativna serija " nad "

Velikostna skupina	Število gospodarstev
Nad 0,0	211.299
" 0,01	202.965
" 0,51	165.594
" 1,01	148.325
" 2,01	124.143
" 3,01	105.803
" 5,01	79.619
" 8,01	55.419
" 10,01	44.185
" 15,01	25.149
" 20,01	14.609
" 30,01	6.167
" 45,01	2.572

Kumulativno serijo "pod" računamo: 8334 + 37.371 = 45.705; 45.705 + 17.269 = 62.974; 62.974 + 24.182 = 87.156 , na podoben način, le da pričnemo od spodaj, pa izračunavamo kumulativno serijo "nad".

Kumulativne serije uporabljamo pri analizi statističnih podatkov zelo pogosto.

66. FREKVENČNE DISTRIBUCIJE

66.1 Važnost proučevanja.

Proučevanje variabilnosti je v statistiki velikega pomena. Ker dajo frekvenčne distribucije sliko variiranja vrednosti znakov v statistični masi, se moramo zaradi tega z njimi podrobneje seznaniti. V primerih statističnih serij v prejšnjih poglavjih imamo več serij,ki jih štejemo k frekvenčnim distribucijam, tako n.pr. tabele 6,1, 6.5, 6.6, 6.7 ....

Zaradi važnosti in značilnosti variiranja numeričnih znakov je posebno važno proučevanje frekvenčnih distribucij numeričnih znakov. Kadar govorimo o frekvenčnih distribucijah, mislimo predvsem nanje.

66.2 Sestavine frekvenčne distribucije.

Pri proučevanju frekvenčnih distribucij moramo uvesti nekaj novih pojmov, ki so potrebni za njihovo analizo.

Vrednosti členov frekvenčne distribucije imenujemo frekvenco, kar pomeni številnost. Frekvenca izraža število enot, ki imajo vrednosti znaka določenega razreda ali pa določeno vrednost znaka. Odtod tudi ime frekvenčna distribucija ali distribucija frekvenc. Tako je v tabeli 6.5 frekvenca 8.334, 37.371 itd.

Vrednosti znaka variirajo med spodnjo mejo najnižjega razreda in zgornjo mejo najvišjega razreda. Ta interval, v katerem vrednosti variirajo, imenujemo variacijski interval ali širino variacije.

Vsota vseh frekvenc pomeni število vseh enot statistične mase, za katero velja frekvenčna distribucija. To vsoto imenujemo obseg mase. V tabeli 6.5 je vsota vseh frekvenc enaka 211.299, kar pomeni, da predstavlja distribucija razdelitev 211.299 gospodarstev.

Zaradi lažjega sporazumevanja pri analizi frekvenčnih distribucij vpeljimo splošne znake, katere bomo v kasnejših izvajanjih še večkrat potrebovali. Splošen člen serije zaznamujemo z k.

x_k = vrednost sredine k-tega razreda oz. k-ta vrednost

Δ_k = širina k-tega razreda

n_k = frekvenca, kateri pripada x_k

$n = \sum n_k$ = vsota vseh frekvenc = število enot v masi.

Frekvenca v nekem razredu je odvisna od lege razreda in širine razreda. Od lege razreda je namreč odvisna splošna gostitev vrednosti, na drugi strani pa velja, da je v širšem razredu tudi frekvenca večja. Med temi količinami stoji naslednja zveza:

$$n_k = \Delta_k \frac{n_k}{\Delta_k} = \Delta_k \cdot g_k \qquad g_k = \frac{n_k}{\Delta_k} \qquad (6.1)$$

g_k je količina, ki pove, kolik del frekvence odpade na enoto intervala v k-tem razredu. Ta kvocient imenujemo gostoto frekvence.

Ako vzamemo kot primer distribucijo iz tabele 6.5, dobimo:

Tab. 6.10

Število zemljiških gospodarstev v LRS
(po republiškem popisu ~~popisu~~ 1947)

Velikost gospodarstva	n_k	x_k	Δ_k	$g_k = \frac{n_k}{\Delta_k}$
Brez lastne zemlje	8.334	0	-	
0,01 - 0,50 ha	37.371	0,225	0,5	74.742
0,51 - 1,00 ha	17.269	0,755	0,5	34.538
1,01 - 2,00 ha	24.182	1,505	1,0	24.182
2,01 - 3,00 ha	18.340	2.505	1,0	18.340
3,01 - 5,00 ha	26.184	4.005	2,0	13.092
5,01 - 8,00 ha	24.200	6.505	3,0	9.067
8,01 - 10,00 ha	11.234	9.005	2,0	5.617
10,01 - 15,00 ha	19.036	12.005	5,0	3.807
15,01 - 20,00 ha	10.540	17.005	5,0	2.108
20,01 - 30,00 ha	8.422	25.005	10,0	842
30,01 - 45,00 ha	3.595	37.005	15,0	240
45,01 -	2.572			
skupno	n = 211.299			

Vidimo, da je gostota gospodarstev največja v najnižjem razredu potem pa stalno pada. To ni razvidno iz samih frekvenc n_k ker moti sliko različna širina razredov, ki vpliva na velikost frekvenc.

Ako imajo razredi enake širine, $\Delta_k = \Delta$, se n_k neposredno

sorazmerni g_k. Zakonitost gostitve je razvidna v tem primeru že iz frekvenc. Zato tvorimo običajno frekvenčne distribucije, ki imajo enake širine razreda.

66.3 Grafični prikaz.

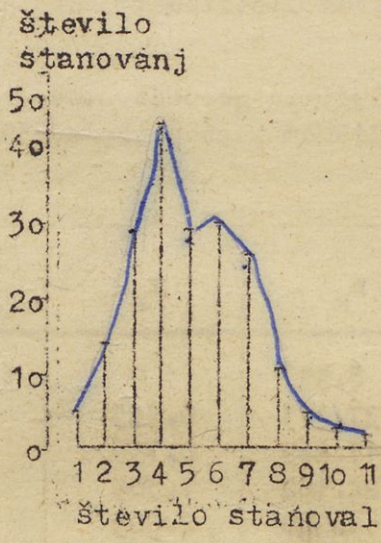
Frekvenčno distribucijo moremo prikazati na različne načine grafično. Ako gre za distribucijo nezveznega znaka jo moremo prikazati na tri načine: s črtami, histogramom in poligonom frekvenc.

Tab. 6.11

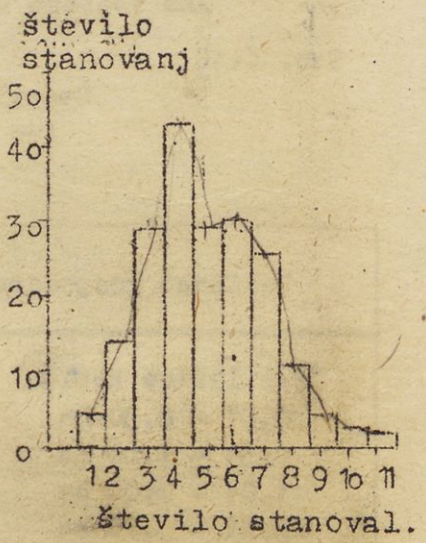
Stanovanja popisnega okoliša A po številu stanovalcev (po popisu stanovanj)

Število stanovalcev	Število stanovanj
1	5
2	14
3	29
4	43
5	29
6	30
7	26
8	11
9	5
10	3
11	2
Σ	197

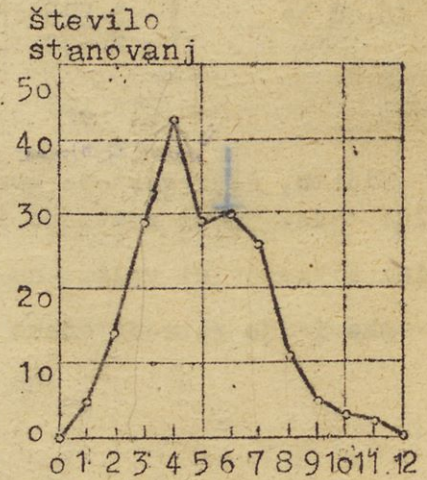
Sl. 6,1
črte



Sl.6.2
histogram



Sl. 6.3
poligon



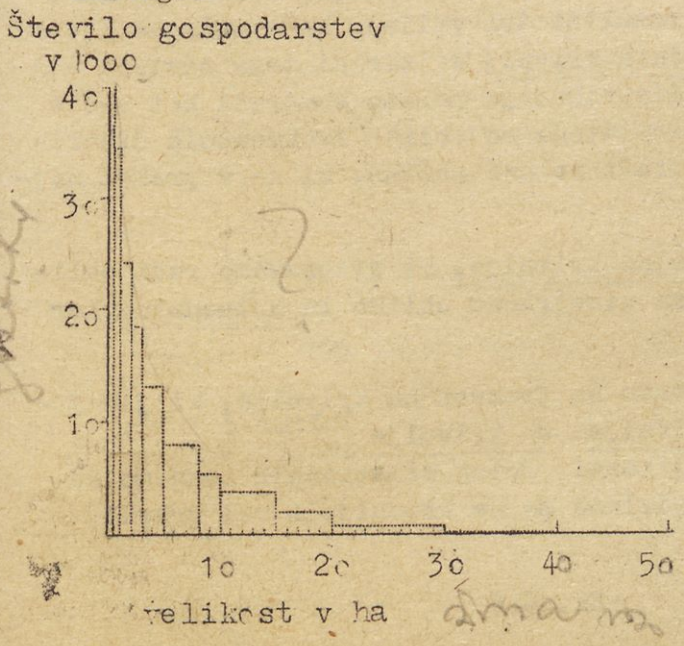
Ako gre za zvezen znak moremo prikazati frekvenčne distribucije s histo - gramom ali poligonom.

Pri histogramu je treba paziti, da nanašamo v ordinato gostote frekvenc, širino stolpcev pa v širini razrednega intervala. Površina stolpca je v tem primeru v sorazmerju s frekvenco. Glej sl. 6.4.

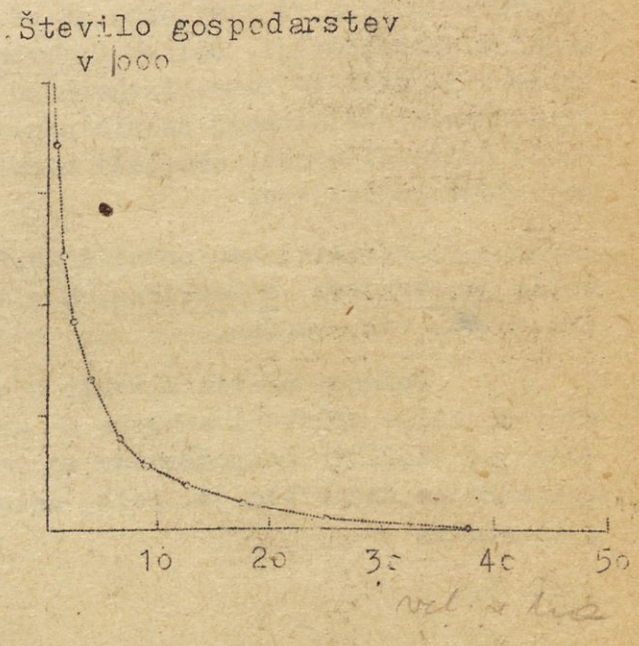
Pri risanju poligonov enako nanašamo na ordinato gostoto in sicer nad sredino razreda. Glej sliko 6.5.

Kot primer vzemimo serije iz tabele 6.10 število gospodarstev v LRS

Sl. 6,4
histogram



Sl. 6.5
poligon



66.4 Frekvenčne krivulje.

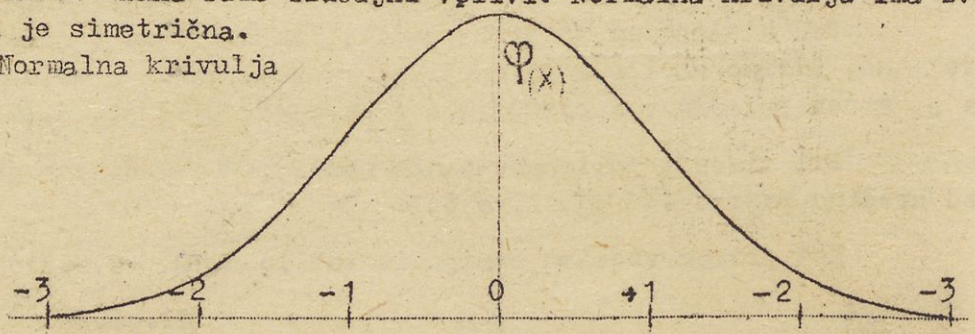
Če vzamemo razrede zvezne frekvenčne distribucije dovolj majhne, bo histogram, posebno pa poligon pri velikih masah vedno bolj podoben izglajeni krivulji. V primeru, da širina intervalov limitira proti 0, pa bo histogram oz. poligon prešel v krivuljo. Frekvenčna distribucija bo v tem primeru dana s frekvenčno krivuljo. Histogram oz. poligon sta samo približka za frekvenčno krivuljo. Za nekatere distribucije, predvsem teoretične, more biti gostota frekvenc frekvenčne krivulje dana v analitični obliki, v obliki obrazca. Ena izmed najbolj važnih in znanih je normalna distribucija. Enačba za gostoto frekvenc normalne distribucije se glasi:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \qquad (6.2)$$

(x je vrednost znaka, $\varphi(x)$ pa gostota frekvence pri vrednosti znaka x, e = 2,718.. je stalna vrednost). Vrednosti nekega znaka se razdelijo po normalni krivulji, ako

vplivajo na vrednost znaka samo slučajni vplivi. Normalna krivulja ima zvonasto obliko in je simetrična.

Sl. 6.6 Normalna krivulja

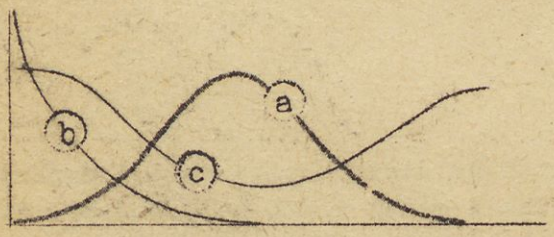


Vrednosti znaka pri enotah so rezultat delovanja najrazličnejših vplivov. Iz tega je razvidno, da se rezultat teh vplivov kaže tudi na frekvenčni distribuciji. Oblike frekvenčnih krivulj so zaradi tega odvisne od vplivov in sestave mase. Frekvenčno distribucijo moramo smatrati kot vsoto frekvenčnih distribucij delnih mas. Zaraditega so oblike frekvenčnih distribucij najrazličnejše. Obdelali bomo grafično one oblike, ki se v praksi najpogosteje pojavljajo.

Že zgoraj smo omenili normalno krivuljo, ki ji pravimo radi oblike da je zvonasta, simetrična, ker ima simetrično obliko in unimodalna, ker ima en sam vrh - modus.

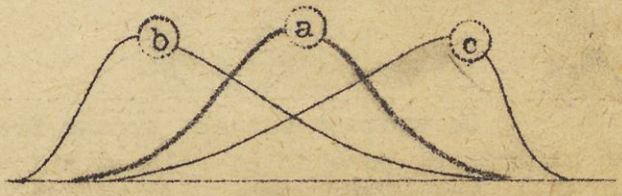
Poleg zvonastih krivulj poznamo še frekvenčno krivulje, ki jih zaradi značilne oblike imenujemo U- krivulje, J- krivulje. Obliko J krivulje pa ima primer [illegible] bucije zemljiških gospodarstev po velikosti. Poleg simetričnih poznamo še asimetrične distribucije, poleg unimodalnih pa še krivulje z več vrhovi, polimodalne

Sl. 6.7



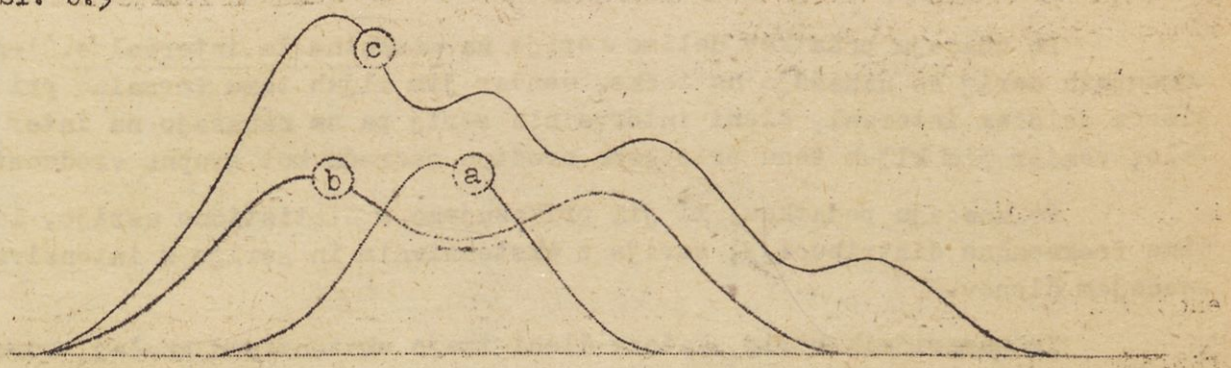
Krivulja po obliki:
a) zvonasta
b) J- krivulja
c) U- krivulja

Sl. 6.8



Krivulja po simetriji:
a) simetrična
b) asimetrična v levo
c) asimetrična v desno

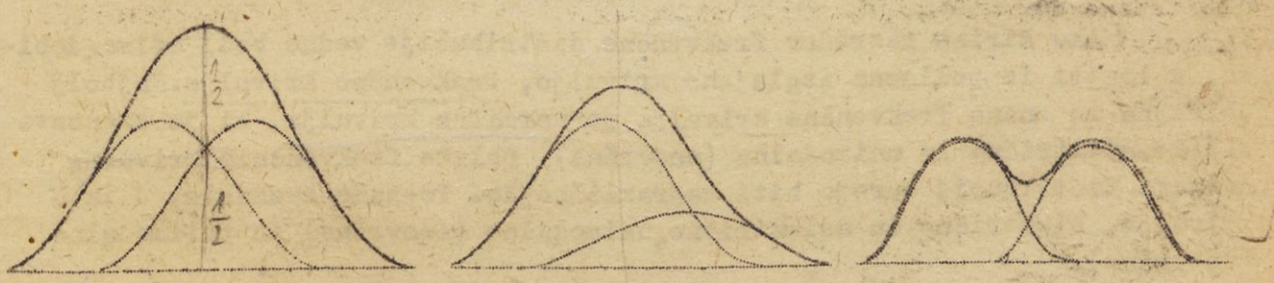
Sl. 6.9



Krivulja po številu vrhov:
a) unimodalna (z enim vrhom)
b) bimodalna (z dvema vrhovoma)
c) polimodalna (z več vrhovi)

Frekvenčna distribucija celotne mase je vsota frekvenčnih distribucij delnih mas. Četudi bi bile vse frekvenčne distribucije delnih mas, ki sestavljajo celotno maso, normalne, morejo kot vsota nastati najrazličnejše oblike.

Sl. 6.10



zvonasta asimetrična bimodalna

Že na teh enostavnih primerih dveh delnih distribucij je razvidno, kako morejo nastati posamezni tipi krivulj pri različnih sestavih mas.

67. POVZETEK

Istovrstne statistične podatke prikazujemo v statističnih serijah. Statistična serija je torej vrsta istovrstnih podatkov, od katerih se vsak nanaša na eno izmed vrednosti ali grupo osnovnega znaka, serije.

Statistične serije delimo po značaju osnovnega znaka, tako kot znake, na časovne, krajevne in stvarne. Stvarne serije naprej delimo na atributivne in numerične, numerične pa na zvezne in nezvezne.

Kumulativno serijo dobimo s postopnim seštevanjem posameznih členov prvotne serije.

Po značaju podatkov delimo serije na momentne in intervalne. Členi momentnih serij se nanašajo na točke, vendar jim kljub temu formalno pripišemo določen interval, členi intervalnih serij pa se nanašajo na intervale, vendar jim kljub temu pripišemo sredino razreda kot grupno vrednost.

Po značaju podatkov, ki jih prikazujemo s statistično serijo, ločimo frekvenčne distribucije, serije z ekstenzivnim in serije z intenzivnim značajem členov.

Iz numeričnih serij, katere členi imajo ekstenziven značaj, moremo izračunati kumulativne serije.

Frekvenčne distribucije so posebno važne zaradi tega, ker moremo iz njih proučiti variabilnost pojava.

Pri frekvenčni distribuciji moramo poznati naslednje pojme. Število enot v nekem razredu imenujemo frekvenco in jo zaznamujemo z n_k. Širina razreda (Δ_k) je razlika med zgornjo in spodnjo mejo razreda. Gostitev frekvenc v nekem razredu podaja gostota frekvenc, ki jo izračunamo tako, da frekvenco delimo s širino razreda $g_k = \frac{n_k}{\Delta_k}$. Gostota frekvenc je posebno važen pojav pri serijah z neenako širokimi razredi.

Frekvenčne distribucije prikazujemo grafično s črtami, histogrami (stolpci) ali poligoni.

Ako širine razredov frekvenčne distribucije vedno bolj ožimo, dobimo v limiti iz poligona izglajeno krivuljo, frekvenčno krivuljo. Najbolj običajna in znana frekvenčna krivulja je normalna krivulja, ki je zvonaste oblike, simetrična in unimodalna (enovršna). Oblike frekvenčnih krivulj, oziroma distribucij morejo biti najrazličnejše. Poznamo zvonaste, J in U krivulje, simetrične in asimetrične, unimodalne (enovršne) in polimodalne (večvršne).

7. RELATIVNA ŠTEVILA

70. SPLOŠNO

Pri proučevanju masovnih pojavov en sam statistični podatek ne pomeni mnogo. Ako vemo, da je bilo število delovnega prebivalstva v LRS po popisu l. 1948 862 tisoč, se takoj vprašamo, koliko je to veliko ali malo. Takoj iščemo primerjavo z drugimi statističnimi podatki. Vprašamo se, koliko je bilo delovnega prebivalstva v Sloveniji pred vojno, skušamo primerjati število delovnega prebivalstva v Sloveniji s podatki o delovnemu prebivalstvu za druge republike. Zanima nas tudi, koliki del celotnega števila prebivalstva LRS predstavlja delovno prebivalstvo itd. Zaradi tega statistične podatke združujemo v tabelah v statistične serije. S tem je dana osnova za primerjavo, ker imamo zbranih v pregledni obliki več med seboj primerljivih podatkov. Na vprašanje časovnega razvoja števila delovnega prebivalstva v LRS bomo dobili odgovor s časovno serijo števila delovnega prebivalstva v LRS po posameznih letih. Med republiško primerjavo števila delovnega prebivalstva bomo dobili v geografski seriji števila delovnega prebivalstva po republikah itd.

Vendar je navajanje absolutnih števil v obliki statističnih serij pomankljivo sredstvo medsebojne primerjave statističnih podatkov. Poglejmo primer delovnega prebivalstva po republikah po popisu 15. marca 1948.

Tab. 7.1

Republika	Število delovnega prebivalstva
F L R J	9,784.602
LR Srbija	4,173.549
LR Hrvatska	2,474.673
LR Slovenija	862.144
LR B i H	1,439.845
LR Makedonija	652.671
LR Črna gora	208.720

Iz te tabele moremo brez nadaljnega sklepati samo, da je število delovnega prebivalstva v LR Srbiji, LR Hrvatski in LR B i H večje, v LR Makedoniji in LR Črni gori pa manjše kot v Sloveniji. To pa ni mnogo. Korak naprej bi storili, ako bi poiskali, za koliko je število delovnega prebivalstva v posameznih republikah večje od števila v LR Sloveniji. V LR Srbiji je 3,311.405 več delovnega prebivalstva kot v LR Sloveniji, v LR Hrvatski 1,612.529 itd. Vendar tudi absolutne diference niso najboljše sredstvo primerjave.

Najboljše podajo odnose med posameznimi podatki relativna števila. Relativno število dobimo tako, da primerjana podatka med seboj delimo. Ako delimo število delovnega prebivalstva v LR Srbiji (4,173.549) z številom delovnega prebivalstva v LR Sloveniji (862.147) dobimo: 4,173.549 : 862.144 = = 4,84. To število pove, da je v Srbiji število delovnega prebivalstva 4,84 krat večje kot v Sloveniji. Na ta način dobimo v enem podatku najlepše in najnazorneje podan odnos med številom delovnega prebivalstva v LR Srbiji in LR Sloveniji. Ako izračunamo na enak način razmerja za ostale republike,

razpredelnica teh razmerij pokaže primerjavo vseh republik z LR Slovenijo.

Tab. 7.2

Republika	Delovno prebivačstvo LRS = 1
F L R J	11,34
Srbija	4,84
Hrvatska	2,87
Slovenija	1,oo
B i H	1,67
Makedonija	o,76
Črna gora	o,24

1138

Vidimo, da te številke kažejo število delovnega prebivalstva, merjeno s številom delovnega prebivalstva v LR Sloveniji. Število prebivalstva posameznih republik smo primerjali, merili s številom prebivalstva v LR Sloveniji, tako, da smo podatke za posamezne republike delili s podatkom za Slovenijo, ki predstavlja enoto. Število, s katerim primerjamo, oziroma merimo neki drugi podatek, imenujemo v statistiki osnovo ali bazo primerjave. V našem primeru smo vzeli kot bazo podatke za LR Slovenijo. Jasno je, da bi mogli vzeti kot bazo tudi katerikoli drugo republiko ali skuper podatek FLRJ. Ako vzamemo kot bazo skupno število prebivalstva v celi državi, dobimo vpogled v sestav delovnega prebivalstva po republikah. Posamezni podatki pomenijo, kakšen del celotnega delovnega prebivalstva odpade na posamezno republiko

Tab. 7.3

L R	
FLRJ	1,oo
Srbija	o,426
Hrvatska	o,253
Slovenija	o,o88
B i H	o,147
Makedonija	o,o67
Črna gora	o,o22

$$\frac{\text{LR Srbija}}{\text{FLRJ}} = \frac{4,173.549}{9,784.6o2} = o,426$$

$$\frac{\text{LR Hrvatska}}{\text{FLRJ}} = \frac{2,474.673}{9,784.6o2} = o,253$$

itd.

Relativna števila so zelo važen pripomoček analiziranja statističnih podatkov. Ako jih pravilno uporabljamo, moremo z enostavnimi načini napraviti iz statističnih podatkov zaključke, ki bi jih sicer ne mogli.

Osnovnega pomena pri uporabi relatimnih števil je, kdaj sta dva podatka med seboj primerljiva z relativnim številom. Primerjanje statističnih podatkov z relativnimi števili ni omejeno samo na istovrstne podatke, temveč moremo primerjati med seboj tudi raznovrstne podatke.

V našem zgornjem primeru smo primerjali med seboj istovrstne podatke in sicer: število delovnega prebivalstva ene republike s številom delovnega prebivalstva druge republike.

Moremo pa na primer primerjati število prebivalstva s površino. Skupno število prebivalstva v FLRJ (15.823 tisoč) in skupna površina (256.88o km2) sta raznovrstna podatka. Vendar moremo iz števila prebivalstva in površine izračunati relativno število, kiga pozhamo pod imenom gostota prebivalstva.

Gostoto prebivalstva izračunamo tako, da število prebivalstva delimo s površino.

$$\frac{15{,}825.000 \text{ preb.}}{256.880 \text{ km}^2} = 61{,}5 \text{ preb/km}^2$$

Kvocient 61,5 preb/km^2 (izgovori: 61,5 prebivalcev na km^2), pokaže, koliko prebivalcev pride povprečno na en kvadratni meter.

Iz tega vidimo, da moremo z relativnimi števili primerjati med seboj tako istovrstne kot raznovrstne podatke. Ako primerjamo istovrstne podatke, je pogoj smiselne primerjave le, da se primerjana podatka razlikujeta v enem samem opredeljujočem znaku. Ako pa primerjamo raznovrstne podatke, pa mora med njima obstojati smiselna vsebinska zveza, skladati pa se morata oba v vseh opredeljujočih pogojih.

71. Vrste relativnih števil:

Po tem, kakšne vrste podatkov med seboj primerjamo, razlikujemo več vrst relativnih števil:

Obravnavali bomo naslednje tri vrste relativnih števil:

a) Strukture ali razčlenitvena števila. O strukturah govorimo, kadar primerjamo delno maso s celotno maso.

b) Statistične koeficiente ali gostote. Pri teh relativnih številih primerjamo med seboj raznovrstne podatke, ki pa so med seboj v vsebinski povezavi.

c) Indeksi. Z njimi primerjamo istovrstne podatke. Indekse uporabljamo predvsem pri časovnih serijah.

72. Strukture ali razčlenitvena števila.

72.1 Enostavne strukture.

Vsak se spominja, da pri proučevanju volilnih rezultatov ne prebiramo toliko absolutnih števil, ampak merimo uspeh ali neuspeh na volitvah s procentom glasovanja za OF itd. Zakaj pokažejo v tem primeru procenti več kot absolutna števila in kako te procente izračunamo?

Ker je število volilnih upravičencev v posameznih volilnih okoliših, krajih ali okrajih različno, je jasno, da je v krajih, kjer je število volilnih upravičencev večje, tudi udeležba na volitvah večja. Absoluten uspeh na volitvah je torej odvisen od skupnega števila volilcev. Primerjavo torej moti različno število volilnih upravičencev. Ta vpliv pa moremo izločiti, ako se vprašamo, koliko povprečno volilcev od 100 volilnih upravičencev je volilo v enem ali drugem kraju. Ti podatki so med seboj bolj primerljivi, ker niso več odvisni od absolutnega števila volilnih upravičencev, temveč

samo od večje ali manjše državljanske zavesti. Ta procent izračunamo tako, da število oseb, ki so volile, delimo s številom vseh volilnih upravičencev in kvocient pomnožimo s loo.

Ker je število oseb, ki so volile, del celotne mase vseh volilnih upravičencev, moremo reči, da dobimo strukturni procent ali slovensko odstotni delež tako, da delno maso delimo s celotno maso in kvocient pomnožimo s loo.

Uporabnost struktur ali razčlenitvenih števil za medsebojne primerjave sestava različnih mas moremo še bolje videti na naslednjem primeru števila delovnega prebivalstva za LR Srbijo, Hrvatsko in Slovenijo. Primerjava absolutnih podatkov ne nudi vpogleda v zaposlitveno strukturo med republikami, ker je absolutno število delovnega prebivalstva za posamezne republike različno. Vpogled v sestav nudijo šele strukturne serije, ki jih izračunamo tako, da število prebivalstva posameznih zaposlitvenih skupin za neko republiko delimo s skupnim številom aktivnega prebivalstva za to republiko in kvocient pomnožimo s loo.

Primerjava strukture aktivnega prebivalstva po zaposlitvi za LR Slovenijo, LR Hrvatsko in LR Srbijo. (Podatki po popisu 15. marca 1948)
Glej tab. 7.4.

Republika / grupa zaposlitve	Absolutno v 1000 prebivalcih			Struktura v odstotkih		
	Slovenija	Hrvatska	Srbija	Slovenija	Hrvatska	Srbija
Skupno	862,1	2474,1	4173,5	100,00	100,00	100,00
Delavci, učenci v gospodarstvu	274,9	386,8	476,8	31,9	15,65	11,43
Nameščenci	92,8	195,6	291,8	1o,77	7,90	7,00
Kmetje: skupno	414,9	1747,5	3196,3	48,11	7o,6	76,6o
zadružni	4,1	13,8	92,7	1,o	o,8	2,9
privatni	41o,8	1733,6	31o3,7	99,o	99,2	97,1
Ribiči	o,o	3,1	1,3	o,oo	o,12	o,o3
Obrtniki	16,3	37,7	6 ,4	1,89	1,53	,47
Trgovci	4,2	9,6	17,5	o,49	o,39	o,42
Svobod. poklici	o,6	2,1	2,8	o,o7	o,o9	o,o7
Privatniki	3,7	18,8	33,4	o,43	o,76	o,8o
Upokojenci	23,9	35,2	35,8	2,77	1,42	o,86
Vzdrž. in drž.	3o,8	38,1	56,2	3,58	1,54	1,34
Neznano	o,o	o,1	o,2	o,oo	o,oo	o,oo

Strukturne procente smo izračunali tako, da smo število delovnega prebivalstva za posamezno grupo zaposlitve delili s skupnim številom delovnega prebivalstva v odgovarjajoči republiki. Na primer:

$$\frac{274,9}{862.1} \cdot 100 = 31,9\ \%\ \text{itd.}$$

Kot vidimo ja za kmete izračunana tudi delna struktura po sektorju.

Struktura odkrije možnosti analize, ki jih v absolutnih številih ni. Iz gornje strukture vidimo, da je v LRS skoro tretjina vsega aktivnega prebivalstva delavcev in približno polovica kmetov. V primerjavi med republikami pa na primer vidimo, da je relativno največ delavcev 31,9 % v LR Sloveniji, največ kmetov 76.6 % pa v LR Srbiji, da je zadružni sektor v kmetijstvu najmočnejši v Srbiji (2,9 % od zaposlenih v kmetijstvu) itd. Struktura po zaposlitvi je torej odkrila splošno ekonomsko strukturo republik.

72. Večkratka struktura.

V prejšnjem primeru je bilo aktivno prebivalstvo za neko republiko razčlenjeno samo po enem znaku: zaposlitvi. Statistično maso pa moremo razčleniti istočasno po dveh ali več znakih. Vzemimo kot primer skupno število prebivalstva v LR Sloveniji, razčlenjeno po spolu in delavnosti.

Tab. P 7.5 Število prebivalstva v LRS po popisu prebivalstva 15.III. 1948 (v tisočih)

Spol \ Delavec	Aktiven	Pasiven	Skupno
Moški	449,3	203,6	652,9
Ženske	412,8	326,2	739,0
Skupno	862,1	529,7	1391,9

Razčlenitvena števila dobimo, ako absolutne podatke v delni masi delimo s podatki v celotni masi. Ako pogledamo tabelo o številu prebivalstva, pa vidimo, da je n.pr. število aktivnega moškega prebivalstva 449,3 delna masa treh različnih statističnih mas: Skupnega števila prebivalstva 1391,9 tisoč, skupnega števila moškega prebivalstva 652,9 in skupnega števila aktivnega prebivalstva 862,1.

Od skupnega števila prebivalcev v LR Sloveniji je $\frac{449,3}{1391,9} \cdot 100 = 32,2\ \%$ aktivnega moškega prebivalstva.

Od skupnega števila moških prebivalcev v LRS je $\frac{449,3}{652,9} \cdot 100 = 68,8\ \%$ aktivnih.

Od skupnega števila aktivnih prebivalcev v LRS je $\frac{449,3}{862,1} \cdot 100 = 52,2$ % moških.

Iz tega sledi, da moremo pri masah, ki so razčlenjene po več znakih, izračunati več vrst struktur. Pri masah, razčlenjenih po dveh znakih, moremo izračunati tri vrste struktur, in sicer:

Tab. 7.5 a

Spol \ Delav.	Aktiven	Pasiven	Skupno
Moški	32,2	14,6	46,8
Ženske	29,8	23,4	53,2
Skupno	62,0	38,0	100,0

V tej vrsti strukture, ki jo imenujemo "kotni sto", ker stoji 100 v kotu, je razdeljeno celotno število prebivalstva. Dobili smo jo tako, da smo posamezne podatke delili s skupnim številom prebivalstva in kvociente pomnožili s 100.

Tab. 7.6

Spol \ Delav.	Aktiven	Pasiven	Skupno
Moški	68,8	31,2	100,0
Ženske	55,8	44,2	100,0
Skupno	62,0	38,0	100,0

Pri teh strukturah vidimo, da 100 ne stoji samo pri "kotni vrednosti", temveč v vsem stolpcu pri skupno. Prvo vrsto odstotkov smo dobili, da smo posamezen podatek v prvi vrsti delili s skupnim številom moškega prebivalstva (652,9), drugo pa, da smo posamezne podatke v drugi vrsti delili s skupnim številom ženskega prebivalstva (739,0). Ta strukturna tabela, imenovana tudi "stolpični sto", daje vpogled v odvisnot strukture po delavnosti od spola. Vidimo namreč, da je odstotek aktivnega ali pasivnega prebivalstva odvisen od spola, (pri moških 68,8 % pri ženskah pa samo 55.8 % aktivnega prebivalstva).

Tab. 7.7

Spol \ Delav.	Aktiven	Pasiven	Skupno
Moški	52,2	38,4	46,8
Ženske	47,8	61,6	53,2
Skupno	100,0	100,0	100,0

V zadnji izmed možnih strukturnih tabel pa vidimo, da je 100 postavljen v vseh okencih zadnje vrste "vrstični sto". Tabela prikazuje strukturo po spolu, za aktivno, pasivno in skupno prebivalstvo. Izračunali smo jo na ta način, da smo vsak stolpec delili s pripadajočim skupnim številom. Ta tabela pokaže odvisnost strukture po spolu od delavnosti.

Na podoben način moremo proučiti vse večkratno razčlenjene statistične mase. Vsaka izmed gornjih vrst struktur osvetli razčlenitev mase na svoj način.

73. STATISTIČNI KOEFICIENTI L' GOSTOTE

73.0 Kot smo omenili že uvodoma, so v statistiki smiselni in važni tudi odnosi raznovrstnih podatkov. Ako primerjamo število prebivalstva s površino, dobimo vpogled v gostoto prebivalstva. Primerjava proizvodnje električne, jekla, premoga v neki državi s skupnim številom prebivalstva pokaže stopnjo industrializacije in ekonomske moči te države. Podobnih primerov bi mogli našteti še mnogo.

Primerjavo izvedemo na ta način, da oba podatka med seboj delimo. Že v uvodnem poglavju o relativnih številih imamo primer gostote prebivalstva, ki katerem gre za primerjavo dveh raznovrstnih podatkov števila prebivalstva in površine.

Kadar izračunavamo relativna števila iz raznovrstnih podatkov, govorimo o statističnih koeficientih in gostotah. kazalci nivoja

Dve raznovrstni statistični količini, ki ju primerjamo, moreta po svojem nastanku izvirati iz iste statistične mase (n.pr.število goved in površina istih gospodarstev) ali pa iz dveh popolnoma različnih mas (n.pr.število prebivalstva in površina). Za prve moremo koeficiente izračunavati tudi za posamezne enote, za druge pa samo za celoto! Primerjani masi sta si v vsakem primeru v prirejenem položaju. Ker gre že za vsebinsko raznovrstnost primerjanih količin, se morata podatka, iz katerih izračunavamo statistični koeficient, skladati v vseh opredeljujočih pogojih.

Ako primerjamo n.pr. število motornih vozil s številom prebivalstva, se morata oba podatka nanašati na isti datum in isto področje, da je primerjava smiselna. Ker moremo primerjati med seboj raznovrstne mase, mora [illegible] pod kakšnimi pogoji je ta primerjava smiselna.

Nastopiti morejo trije različni primeri:

a) Oba podatka izhajata iz momentnih mas

b) Oba podatka izhajata iz intervalnih mas

c) Prvi podatek izhaja iz intervalne, drugi pa iz momentne mase.

73.1 Pri primerjavi podatkov dveh momentnih mas moremo brez težav doseči enako opredelitev obeh mas. Oba podatka morata imeti isti kritični datum

Koeficient izračunamo po obrazcu $K = \frac{R}{S}$ (7.1)

R je podatek prve mase, S podatke druge mase, K pa statistični koeficient.

P Kot primer za koeficient iz dveh momentnih mas vzemimo število prebivalstva in površino. Koeficient, ki ga iz njih izračunamo pomeni število prebivalcev na 1 km2, in ga imenujemo gostota prebivalstva. Primer vsebuje podatke za nekaj evropskih držav (Vir statistični bilten št. 1).

Tab. 7.8

Država	Število prebivalstva sredi leta 1948 v 000	Površina v km^2	Gostota prebivalstva
.	(1)	(2)	(3)=(1):(2)
Avstrija	6972	83851	83
Belgija	8557	30507	281
Češkoslovaška	12339	127827	98
Holandija	9793	33734	290
Italija	45706	301026	152
Jugoslavija	15825	256880	61,5

Gostota prebivalstva je zelo uporabljan koeficient, ki pokaže naseljenost prebivalstva. Za Avstrijo n.pr. je račun naslednji:

$$83 = \frac{6972000}{83851}$$

73.2 Enako je brez vseh težav mogoče izvesti primerjavo podatkov dveh intervalnih mas. Edini pogoj je smiselnost primerjave in isto razdobje za obe masi. Širina časovnega intervala v tem primeru ne moti, ker se izloči sama od sebe. Zaradi tega ni potrebno podatke predhodno reducirati na enoto.

D Dokaz: Ako je R absoluten podatek prve intervalne mase, S absoluten podatek druge mase, $\triangle$ pa širina intervala, so podatki praračunani na enoto časa enaki: $R' = \frac{R}{\triangle}$, $S' = \frac{S}{\triangle}$. Koeficient iz teh podatkov je enak: $K' = \frac{R'}{S'} = \frac{R}{\triangle} : \frac{S}{\triangle} = \frac{R\triangle}{\triangle S} = \frac{R}{S} = K$ Koeficient izračunan iz reduciranih podatkov je torej enak koeficientu, izračunanem iz absolutnih podatkov.

P Razmerje med številom rojenih in številom umrlih v nekem razdobju pokaže, če prebivalstvo zaradi naravnega gibanja narašča ali pada? Ta koeficient iz demografske statistike imenujemo "vitalni indeks". Izračunati je treba vitalni indeks za Češkoslovaško za dana razdobja (Po podatkih Statistička priručka 1948).

Tab. 7.9

Razdobje	Absolutno število		Vitalni indeks
	Rojstev	Smrti	
	(1)	(2)	(3)=(1):(2).100
1936 - 1940	1296,0	971,7	133,3
1945	276,0	251,3	110,0
1946	287,3	178,8	161,0

Za razdobje 1936-1940 je vitalni indeks enak:

$$\frac{1296,0}{971,7} \cdot 100 = 133,3$$

73.3 Težje pa je s primerjavo intervalne mase z momentno.

Pogoj za primerljivost je enaka opredelitev obeh mas. Ta pogoj pa v tem primeru ni mogoče brez nadaljnega izpolniti, ker je intervalna masa opredeljena z intervalom, momentna pa z momentom. Primerjava pa je kljub temu možna, ker moremo iz več podatkov za momentno maso izračunati povprečno vrednost. Povprečna vrednost momentnih podatkov pa ima intervalen značaj. Iz tega sledi, da moremo primerjati podatke intervalne mase s povprečji podatkov momentne mase. Da moremo izračunati povprečno vrednost za momentne podatke, je potrebno, da imamo v danem intervalu podatke za več momentov. Zaznamujemo jih z $M_1, M_2 \ldots M_k \ldots M_n$.

Ker je podatek o intervalni masi (R) odvisen od širine intervala, ga moramo v primeru, da širina intervala ($\triangle$) ni enota, preračunati na enoto tako, da ga delimo s širino razreda ($\triangle$).

Koeficient je smiseln, ako primerjamo podatek intervalne mase preračunan na enoto (R'), s povprečno vrednostjo momentne mase ($\bar{M}$)

$$K = \frac{R'}{\bar{M}} = \frac{R}{\triangle . \bar{M}} \qquad (7.2)$$

Ker izračunavamo te koeficiente na 100, 1000 ali 10000 enot podatka momentne mase, moramo K pomnožiti z 100, 1000 ali 10000, splošno z E, ako pomeni E pripadajoče okroglo število.

$$K/_E = \frac{R \cdot E}{\triangle . \bar{M}} \qquad (7.3)$$

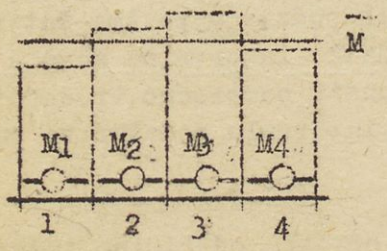
Ako so R rojstva, E = 1000, M število prebivalstva, enota časa leto, je $K/_E$ število rojstev na leto na 1000 prebivalcev. (natalitet)

Povprečno vrednost momentne mase ($\bar{M}$) pa glede na trenutke, za katere veljajo posamezne vrednosti M_k, izračunamo različno. V glavnem obstojata dve možnosti.

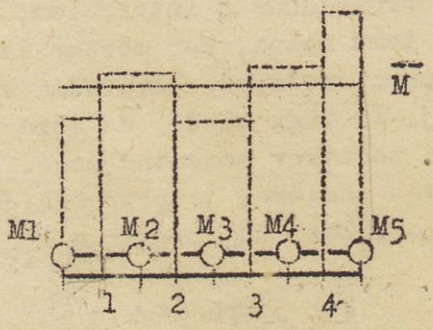
Ako je interval, na katerega se nanaša intervalna masa R, razdeljen na n delnih enakih intervalnih, so podatki momemntne mase (M_k) običajno dani za sredine ali meje intervalov.

Sl. 7.1

a)



b)



sl.a Ako padejo v celotni interval za vse vrednosti celi pripadajoči intervali okrog posameznih vrednosti M_k izračunamo povprečno vrednost $\bar{M}$ po obrazcu:

$$\bar{M} = \frac{1}{4} (M_1 + M_2 + M_3 + M_4)$$

ali splošno

$$\bar{M} = \frac{1}{n} (M_1 + M_2 + M_3 + \ldots + M_n) \qquad (7.4)$$

V primeru sl. b pa pade od vrednosti, ki so na meji celotnega intervala samo polovica mejnega intorvala v celotni interval. Zaradi tega moramo ta podatka pri izračunavanju povprečne vrednosti šteti samo polovično. Povprečno vrednost $\bar{M}$ izračunamo v tem primeru po obrazcu:

$$\bar{M} = \frac{1}{4} (\frac{1}{2} M_1 + M_2 + M_3 + M_4 + \frac{1}{2} M_5)$$

ali splošno

$$\bar{M} = \frac{1}{n} (\frac{1}{2} M_1 + M_2 + M_3 + \ldots + M_n + \frac{1}{2} M_{n+1}) \qquad (7.5)$$

Iz gornjega vidimo, da moremo smatrati kot povprečno vrednost eno samo vrednost M_1 samo, ako ta leži sredi intervala

$$\overset{M_1}{\circ}\text{------} \qquad \bar{M} = \frac{1}{1} M_1 = M_1$$

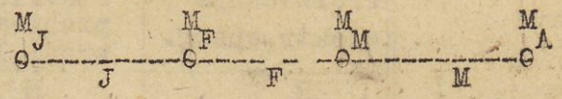
V primeru, da leže vrednosti na mejah celotnega intervala

$$\overset{M_1}{\circ}\text{----------}\overset{M_2}{\circ}$$

pa vzame mo kot povprečno vrednost: $\bar{M} = \frac{1}{1} (\frac{1}{2} M_1 + \frac{1}{2} M_2) = \frac{1}{2} (M_1 + M_2)$

P Število nesreč pri delu je bilo v neki panogi industrije v prvem četrtletju 183. Število delavstva 1. januarja je bilo 11.236, 1. februarja 12.287, 1. marca 12.627, 1. aprila pa 13.120 (Podatki izmišljeni). Izračunati je treba koeficient gostote nesreč na 1000 delavcev na leto.

$$K/_E = \frac{R \cdot E}{\triangle \cdot \bar{M}}$$



$$\bar{M} = \frac{1}{3}\left(\frac{1}{2} M_J + M_F + M_M + \frac{1}{2} M_A\right)$$

$$\bar{M} = \frac{1}{3}\left(\frac{1}{2} 11.236 + 12.287 + 12.627 + \frac{1}{2} 13.120\right) = \frac{1}{3} \times 37.092 = 12.364$$

$R = 183$

$E = 1000$

$\triangle = \frac{1}{4}$

$\bar{M} = 12364$

$K/_E = 59,2$

$$K/_E = \frac{183 \cdot 1000}{\frac{1}{4} \cdot 12364} = \frac{732000}{12364} = 59,2$$

P V letu 1949 je bilo rojenih v neki državi 38766 otrok, število prebivalstva v začetku leta je bilo 1456000, koncem leta pa 1477000. Kolika je nataliteta (nataliteta je število rojstev na leto na 1000 prebivalcev (Podatki izmišljeni).

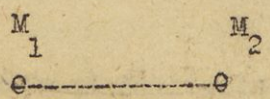
$R = 38766$

$E = 1000$

$\triangle = 1$

$\bar{M} = 1466500$

$K_E = N = 26,4$



$$\bar{M} = \frac{1}{2}(M_1 + M_2) = \frac{1}{2}(1456000 + 1477000) = \frac{2933000}{2} = 1466500$$

$$K_E = \frac{38.766 \cdot 1000}{1 \cdot 1466500} = \frac{3,8766 \cdot 10^7}{1,4665 \cdot 10^6} = 2,64 \cdot 10^1 = 26,4$$

P Najrazličnejši ekonomski podatki, izračunani na prebivalca ("na glavo") pokažejo raven ekonomskega stanja posameznih držav. Izračunati je treba proizvodnjo električne energije na prebivalca za nekaj držav. (Podatki mesečnega biltena OZN)

Tab. 7.1o

Država	Proizvodnja elektr.energ.	Število prebivalstva sredi leta	Proizvodnja el. energ. / preb.
	R_k	M_k	K_k
Belgija	6o1	8421	856
Bolgarija	4o	7o48	68
Češkoslovaška	555	12170	548
Italija	1495	45833	391
ZDA	21310	144o34	178o
Anglija	3548	49539	860

R_k = proizvodnja električne energije v $1o^6$ KWh na mesec v letu 1947 po državah.

M_k = število prebivalstva v $1o^3$ sredi leta po državah.

K_k = proizvodnja električne energije v letu na prebivalca

$$\triangle = \frac{1}{12} \qquad E = 1$$

$$K_k = \frac{R_k \cdot 10^6 \cdot 1}{\frac{1}{12} \cdot M_k \cdot 1o^3} = 1{,}2 \cdot 1o^4 \cdot \frac{R_k}{M_k}$$

$$K_{Bel} = \frac{1{,}2 \cdot 6o1 \cdot 1o^4}{8421} = 856$$

Nekateri statistični koeficienti in gostote so take narave, da so smiselne tudi njihove obratne vrednosti.

Smiselen je koeficient "število prebivalcev na eno motorno vozilo" in obraten koeficient "število motornih vozil na looo prebivalcev itd.

Enote mere statističnih koeficientov so izpeljane iz enot mere primerjanih mas. Gostota prebivalstva se meri s "številom prebivalstva na 1 km^2", nataliteta s "številom rojstev na leto na looo prebivalcev". Pravilno navajanje enot mere je za razumevanje vsebine statističnih koeficientov zelo važno.

73.4 Uporaba statističnih koeficientov in gostot je vsestranska in v vseh vejah statistike zelo razširjena. Zelo pogosto izračunavamo koeficiente "na prebivalca". Ti koeficienti pokažejo splošno raven določenega socialno ekonomskega pojava glede na najvažnejši činitelj, človeka. Poleg teh izračunavamo še druge.

V industrijski statistiki merimo proizvodnost dela " s proizvodnjo na enega delavca" ali na "izvršeno delovno uro". Skrb za naraščaj pokaže "število učencev v gospodarstvu na 100 kvalificiranih delavcev" itd.

V kmetijski statistiki je hektarski donos razmerje med pridelkom in površino. Število živine primerjamo s površino njivsko, obdelovalne ali krmske površine. Kot koeficient mehanizacije uporabljamo število traktorjev na 1000 ha orne površine ali obratno orno površino, ki odpade na en traktor.

V demografski statistiki so skoro vsi pokazatelji statistični koeficienti, oziroma gostote. Poleg natalitete, ki pomeni število rojstev na 1000 prebivalcev na leto, pomeni mortaliteta število smrti na 1000 prebivalcev na leto. Specifična morbiditeta, število obolenj za določeno bolezen na 1000 ali 10000 prebivalcev v enem letu itd.

V kulturno-prosvetni statistiki merimo s "številom šolskobveznih otrok na eno šolo" obremenjenost šol s "številom učencev na enega učitelja" obremenjenost učiteljstva s " število obiskov kulturno - prosvetnih prireditev na prebivalca na leto" kulturno raven prebivalstva itd.

V trgovinski statistiki je cena odnos med vrednostjo in količino. Blagovni promet na enega prebivalca je pokazatelj življenjskega standarda. Pod obrtljivostjo zalog razumemo odnos med prometom in zalogami itd.

74. INDEKSI – MERSKA ŠTEVILA

74.0 O indeksih govorimo takrat, kadar z relativnimi števili primerjamo med seboj istovrstne, prirejene količine. Medtem ko smo pri strukturah in statističnih koeficientih primerjali med seboj samo absolutna števila, moremo z indeksi primerjati poleg absolutnih tudi vse druge statistične količine, strukturne pokazatelje, koeficiente, srednje vrednosti itd.

Ako z indeksi primerjamo celo serijo podatkov, ostane običajno baza (podatek na katerega primerjamo vse člene statistične serije) za celo serijo ista. Vrednost baze postavimo vedno enako 100,0. Indeksi so neimenovana števila, ker delimo istovrstne podatke. Splošen obrazec za indeks je:

$$I_{k/o} = \frac{Y_k}{Y_o} \cdot 100 \qquad (7.6)$$

74.1 Enostavni indeksi.

Indekse, izračunane iz podatkov, ne gledena njihov notranji sestav imenujemo enostavne. Po tem, v kakšnem znaku se primerjana podatka razlikujeta, delimo indekse na:

a) stvarne
b) krajevne - teritorialne
c) časovne

Izmed teh treh vrst so najvažnejši časovni indeksi, ker pomagajo analizirati eno izmed najvažnejših lastnosti masovnih pojavov, časovni razvoj - dinamiko.

.11 Stvarni indeksi.

Kot primer serije stvarnih indeksov moremo vzeti odstotek nepismenih od lo let dalje v FLRJ po glavnih narodnostih (Vir: Statistični bilten FLRJ št.1).

Tab. 7.11

Narodnost	°/° nepismenih	Indeks FLRJ = loo.o
Skupno FLRJ	25.43	loo.o
Srbi	27.71	lo9.o
Hrvati	18.11	71.2
Slovenci	2.32	9.1
Makedonci	3o.24	119.o
Črnogorci	24.13	94.8
Muslimani - neopredeljeni	54.26	213.5

Indeks nad loo.o pomeni večjo, pod loo.o pa manjšo nepismenost, kot je povprečna v FLRJ. Kadar serija vsebuje tudi povprečno količino, običajno vzamemo povprečje kot bazo loo.o. Indeksi pokažejo v tem primeru odstopanja od povprečka. Indeksno serijo izračunamo, da podatek za posamezno narodnost delimo z odstotkom nepismenih v FLRJ in pomnožimo s loo.

Za Srbija je n.pr. $\frac{27.71}{25.43} \cdot loo = lo9.o$

.12 Krajevni - geografski indeksi.

Z geografskimi indeksi primerjamo statistične podatke, ki veljajo za različne kraje ali geografska področja.

P . Kot primer poglejmo serijo proizvodnje električne energije v KWh na prebivalca na leto iz odstavka 73.3 in vzemimo kot bazo ZDA.

Tab. 7.12

Država	Proizvodnja na preb. v KWh	Indeks ZDA = loo
Anglija	860	48.3
Belgija	856	48.1
Bolgarija	68	3.8
Češkoslovaška	548	30.8
Italija	391	22.o
ZDA	1780	loo.o

Indeksna serija pokaže, da nobena izmed navedenih držav nima niti polovico tolikšno proizvodnjo električne energije na enegg prebivalca kot ZDA. Indeksno serijo izračunamo tako, da podatek za vsako državo delimo s podatkom za ZDA in kvocient pomnožimo s loo. Za Anglijo je n.pr.

račun naslednji:

$$\frac{860}{1760} \cdot loo = 48.3$$

.13 Časovni indeksi.

Najvažnejši in najbolj izdelani so časovni indeksi. Ti se tudi najpogosteje uporabljajo. Kadar govorimo o indeksih in indeksni analizi, mislimo zato predvsem na časovne indekse in analizo dinamike masovnih pojavov. Pri proučevanju časovnega razvoja in odnosov večjega števila pojavov je indeksna analiza odlično sredstvo primerjave.

P Kot primer vzemimo časoven razvoj proizvodnje električne energije v ZDA, Angliji in Franciji. Za vsako državo izračunamo serijo indeksov in vzemimo pri vseh državah leto 1937 kot bazo.

Tab. 7.13 Proizvodnja električne energije v 10^6 KWh mesečno povprečje (Vir. Mesečni bilten OZN)

Leto	Absolutno v 10^6 kwh ZDA	Anglija	Francija	Indeksi 1937 = loo ZDA	Anglija	Francija
1937	9910	1909	1514	loo	loo	loo
1938	9484	2031	1548	96	lo6	lo2
1939	lo637	2201	1643	lo7	115	lo9
1940	11819	2398	1448	118	126	96
1941	13733	2697	1587	139	141	lo5
1942	15499	2971	1577	156	156	lo4
1943	18414	3079	1663	186	162	11o
1944	19016	3196	1282	192	168	85
1945	18540	3lo6	1464	187	163	97
1946	18594	3437	1847	188	180	122
1947	2131o	3548	2111	215	186	14o

Iz indeksnih serij moremo veliko lažje primerjati dinamiko proizvodnje električne energije, ker izhajamo pri vseh iz iste ravni loo, medtem ko so absolutne vrednosti med seboj zelo različne.

Iz indeksnih vrst vidimo, da je proizvodnja prva leta najbolj naraščala v Angliji, l. 1942 pa jo v dinamiki dohiti in prehitijo ZDA. Razvoj proizvodnje v Franciji zelo zaostaja za ZDA in Anglijo.

Indeks za ZDA za leto 1938 izračunamo tako:

$$\frac{9484}{9910} \cdot 100 = 96$$

Ker pri časovnih serijah za analizo dinamike pojavov ni važna samo primerjava na neki stalni člen (indeksi s stalno bazo) izračunavamo v eni in isti seriji večkrat tudi indekse, ki bazo od člena do člena menjajo. Take vrste indeksov imenujemo indekse s premično bazo. Kot bazo indeksov s premično bazo jemljemo običajno predhodni člen. Ker se ti členi vežejo kot členi verige, imenujemo te vrste indeksov verižne indekse. Včasih vzamemo kot bazo indeksov s premično bazo tudi odgovarjajoč mesec preteklega leta n.pr. podatke za januar 1950 primerjamo s podatki za januar leta 1949, podatke za februar 1950 s podatki za februar leta 1949 itd.

Splošen obrazec za izračunavanje verižnih indeksov se glasi:

$$I_{k/k-1} = \frac{Y_k}{Y_{k-1}} \cdot 100 \qquad (7.7)$$

Izračunati je treba serijo verižnih indeksov za časovno serijo proizvodnje električne energije v ZDA.

Tab. 7.14

Leto	Absolutno v 10^6 kwh	$I_{k/k-1}$
1937	9910	100
1938	9484	95.7
1939	10637	112.4
1940	11819	111.1
1941	13733	116.2
1942	15499	112.9
1943	18414	118.8
1944	19016	103.3
1945	18540	97.5
1946	18594	100.3
1947	21310	114.6

$\frac{9484}{9910} \cdot 100 = 95.7$

$\frac{10637}{9484} \cdot 100 = 112.4$

.14 Vprašanja v zvezi z izbiro baze indeksov.

S tehničnega vidika moremo vzeti kot bazo vrednost vsakega člena statistične serije ali kako drugo istovrstno število (n.pr. povprečno vrednost členov, normo, plan itd.). Vsebinsko pa je izbira baze težek posel. Nepravilno izbrana baza da indeksno serijo, iz katere he moremo napraviti dobrih ali pa napačne zaključke. Ker je izbira baze za vsak primer posebna, moremo dati samo nekaj splošnih pripomb.

a) Pri stvarnih in geografskih indeksih vzamemo kot bazo vrednost, ki velja za področje, o katerem imamo najboljšo predstavo (n.pr. v meddržavni primerjavi z geografskimi indeksi n.pr. FLRJ. Kadar izračunavamo indekse iz števil kot so relativna, sredine, itd. vzamemo kot bazo povprečno vrednost (glej primer v odstavku 74.11).

b) Pri časovnih indeksih vzamemo kot bazo čas normalnega stanja pojava. Ako proučujemo gospodarske pojave, ne vzamemo kot bazično medvojno leto, ali leto tik pred ali po vojni, temveč leto normalnega stanja. Pri nas izvajamo primerjavo na leto 1939, OZN pa na leto 1937. Za določene pojave, za katere je težko določiti normalno leto, vzamemo kot bazo večletno povprečje (n.pr. desetletni povpreček hektarskega donosa).

c) Pri istočasnem proučevanju več statističnih serij, moramo vzeti pri vseh serijah isti člen serije kot bazo. Ako nimajo vse serije iste baze, je treba podatke preračunati na enotno bazo. To je možno brez poznavanja osnovnih podatkov, kar sledi iz:

$$I_{k/o} = \frac{Y_k}{Y_o} \cdot 100 = \frac{\frac{Y_k}{Y_r} \cdot 100}{\frac{Y_o}{Y_r} \cdot 100} \cdot 100 = \frac{I_{k/r}}{I_{o/r}} \cdot 100$$

$$I_{k/o} = \frac{I_{k/r}}{I_{o/r}} \cdot 100 \qquad (8.8)$$

Serijo indeksov, preračunamo na novo bazo tako, da vsak člen indeksne serije delimo z indeksom, ki pripada novi bazi in kvocient pomnožimo s 100. To pravilo velja za stvarne, geografske in časovne indekse.

P Preračunati je treba indeksno serijo proizvodnje električne energije v ZDA z bazo 1937 = 100, na novo bazo 1939 = 100

Tab. 7.15

Leto	Indeks 1937 = 100	Indeks 1939 = 100
1937	100	93
1938	96	90
1939	107	100
1940	118	110
1941	139	130
1942	156	146
1943	186	174
1944	192	184
1945	187	175
1946	188	176
1947	215	201

$$I_{k/39} = \frac{I_{k/37}}{I_{39/37}} \cdot 100$$

$$I_{37/39} = \frac{100}{107} \cdot 100 = 93$$

$$I_{38/39} = \frac{96}{107} \cdot 100 = 90$$

$$I_{39/39} = \frac{107}{107} \cdot 100 = 100$$

itd.

74.2 Agregatni ali mnogočlenski indeksi.

Indekse , ki smo jih obravnavali dosedaj, imenujemo enostavne indekse. Ti indeksi se izračunavajo iz sumarnih količin, ne glede na njihov notranji sestav.

.21 Agregatni indeksi.

Pri podrobnejši proučitvi nekaterih količin pa moremo ugotoviti, da so sestavljene iz cele množice elementov. Ako vzamemo kot nazoren primer vrednost proizvodnje, je skupna vrednost proizvodnje vsota produktov količine in cene posameznih artiklov, ki to proizvodnjo sestavljajo. Ako zaznamujemo s q_i količino nekega artikla, s p_i pa ceno istega artikla, je vrednost proizvodnje Y izražena v matematični obliki $Y = \sum_i q_i p_i$. V nadaljnjem bomo radi enostavnejše oblike znak i izpuščali in bomo namesto $\sum_i q_i p_i$ pisali kratko $\sum qp$. Ako z o zaznamujemo bazično vrednost, z 1 pa tekočo vrednost, pomeni q_o količino v bazičnem razdobju, p_o ceno v bazičnem razdobju, q_1 količino v tekočem razdobju p_1 ceno v tekočem razdobju, ker se s časom spreminjajo količine in cene.

Pisano v teh oznakah je:

$$Y_o = \sum p_o q_o \qquad Y_1 = \sum p_1 q_1 \qquad I_{1/o} = \frac{Y_1}{Y_o} = \frac{\sum p_1 q_1}{\sum p_o q_o} \qquad (7.9)$$

$I_{1/o}$ pa nima analitične vrednosti, ker smo ga dobili s primerjavo dveh količin, ki so se menjale zaradi dveh vzrokov hkrati; zaradi spremembe cen in spremembe količin. V statistiki smatramo kot dobro samo primerjavo, ki pokaže vpliv spremembe enega samega sestavnega dela. Indeks, ki kaže spremembo vrednosti proizvodnje samo zaradi spremembe količin, izračunamo tako, da izračunano vrednost proizvodnje za obe razdobji po enakih v stalnih cenah. V simbolih moremo pisati:

$$I_q = \frac{\sum p_o q_1}{\sum p_o q_o} \cdot 100 \qquad \text{Formula Laspeyra} \qquad (7.10)$$

Na enak način moremo izračunati, kakšen je vpliv x spremembe cen na vrednost proizvodnje, ako vzamemo za primerjani razdobji iste količine

$$I_p = \frac{\sum p_1 q_1}{\sum p_o q_1} \cdot 100 \qquad \text{Formula Pasche-a} \qquad (7.11)$$

I_q pokaže povečanje obsega proizvodnje, zaradi tega ga imenujemo indeks obsega proizvodnje, I_p pa spremembo nivoja cen in ga zaradi tega imenujemo indeks cen. Bistvo Laspeyr-ove formule je v tem, da so ponderi (p_o) v indeksni seriji stalni, bistvo Pasche-jeve pa, da so ponderi (q_1) tekoče vrednosti, ki se torej menjajo.

$I_{1/o}$ je rezultat spremembe cen in obsega proizvodnje. Z I_p in I_q smo oba vpliva razdružili, kar sledi iz obrazca:

$$I_{1/o} = \frac{\sum p_1 q_1}{\sum p_o q_o} = \frac{\sum p_1 q_1}{\sum p_o q_1} \cdot \frac{\sum p_o q_1}{\sum p_o q_o} = I_p \cdot I_q \qquad (7.12)$$

Indeks skupnega vpliva spremembe cen in obsega proizvodnje je enak produktu indeksov cen in indeksa obsega proizvodnje.

Te vrste indeksov imenujemo zaradi tega, ker so izračunani iz veččlenskih količin agregatne - mnogočlenske indekse.

F Shematičen primer agregatnega indeksa:

Tab. 7.16

Art.	Razdobje 0		Razdobje 1				
	p_o	q_o	p_1	q_1	$p_o q_o$	$p_1 q_1$	$p_o q_1$
A	3	300	4	350	900	1400	1050
B	5	200	4	250	1000	1000	1250
C	4	700	5	650	2800	3250	2600
D	7	400	6	500	2800	3000	3500
E	2	300	5	600	600	3000	1200
	21	1900	24	2350	8100	11650	9600

$$I_q = \frac{\sum p_o q_1}{\sum p_o q_o} = \frac{9600}{8100} \cdot 100 = 118.5$$

$$I_p = \frac{\sum p_1 q_1}{\sum p_o q_1} = \frac{11650}{9600} \cdot 100 = 121,3$$

$$I_{1/o} = I_p \cdot I_q = 1,213 \cdot 1,185 \cdot 100 = 144,0$$

$$I_{1/o} = \frac{\sum p_1 q_1}{\sum p_o q_o} = \frac{11650}{8100} \cdot 100 = 144,0$$

Podobna vprašanja kot z vrednostjo proizvodnje nastopajo tudi pri drugih pojavih n.pr. proizvodnosti dela itd.

.22 Agregatni indeksi kot tehtana sredina individualnih indeksov.

$p_o q_1 = p_o q_o \left(\frac{q_1}{q_o}\right)$, ali z besedami: vrednost posameznega artikla računana po bazičnih cenah je enaka produktu vrednosti v bazičnem razdobju in enostavnemu indeksu količine tega artikla. Zaradi tega moremo pisati:

$$I_q = \frac{\sum p_o q_1}{\sum p_o q_o} = \frac{\sum p_o q_o \left(\frac{q_1}{q_o}\right)}{\sum p_o q_o} = \sum \frac{p_o q_o}{\sum p_o q_o} \cdot \frac{q_1}{q_o} = \sum \frac{p_o q_o}{\sum p_o q_o} \cdot i_q$$

$$i_q = \frac{q_1}{q_o}$$

$$I_q = \sum \frac{p_o q_o}{p_o q_o} \cdot i_q \qquad (7.13)$$

Agregatni indeks obsega proizvodnje je enak tehtani aritmetični sredini individualnih indeksov količin i_q. Ponderi so strukturni deleži vrednosti v bazičnem razdobju $\left(\frac{p_o q_o}{p_o q_o}\right)$

Na podoben način dobimo tudi:

$$I_p = \frac{\sum p_1 q_1}{\sum p_o q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{\frac{p_1}{p_o}}} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} \qquad i_p = \frac{p_1}{p_o}$$

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} \qquad (7.14)$$

Agregatni indeks cen je enak tehtani harmonični sredini individualnih indeksov cen i_p. Ponderi so vrednosti tekočega razdobja ($p_1 q_1$).

.22 Reprezentativni indeksi.

Z enostavnimi in agregatnimi indeksi primerjamo pojave s področij, ki so obsežni s splošnim popolnim sistemom evidence. Poleg teh pa obstojajo tudi pojavi, ki jih zaradi ogromnega števila individualnih pojavov in težkoč pri evidentiranju ne moremo opazovati v celoti. Nemogoče je opazovati potrošnjo vseh družin, cene vseh artiklov, razdelitev blaga po socialnih skupinah in podobno. Dinamiko teh pojavov proučujemo zaradi tega z reprezentativnimi indeksi. Pri reprezentativnem indeksu vzamemo v indeks iz vsake grupe samo po nekaj elementov, ki jih smatramo kot reprezentativne za to grupo. Podatke reprezentativnega indeksa smatramo kot sliko gibanja celote..

Pri sestavljanju reprezentativnih indeksov moramo izpolniti naslednje točke:

1./ Če gre za raznovrstno maso moramo razdeliti celotno maso v homogene delne mase.

2./ Iz vsakega homogenega dela moramo izbrati en ali več elementov, ki so za homogen del reprezentativni.

3./ Utež ali ponder je treba določiti v sorazmerju z obsegom vrednosti posameznega dela, glede na celotno maso.

4./ Skupinski, ali generalni reprezentativni indeks je vsota produktov uteži in indeksov za posamezne homogene grupe.

.23 Primeri uporabe indeksov v socialno-ekonomskih statistikah.

Enostavni indeksi se splošno uporabljajo kot sredstvo primerjanja in analize, posebno dinamike.

Kot primer agregatnega indeksa pa moremo vzeti vrednost proizvodnje in z njim v zvezi indeks obsega proizvodnje in agregatni indeks cen.

.231 Agregatnè indekse uporabljamo tudi pri drugem problemu ekonomske statistike in sicer pri merjenju proizvodnosti dela.

a) V primeru, da gre za en sam artikel, izračunavamo proizvodnost dela v statističnem koeficientu med proizvedeno količino Q in časom porabljenim na proizvodnjo T. Izračunati moremo dve vrsti koeficientov. Proizvodnost dela moremo meriti s proizvodnjo izdelano v enoti časa, ali s časom, v katerem je bila izdelana enota proizvodnje (komad, tona in podobno). Prvi koeficient izračunamo tako, da celotno proizvodnjo delimo s celotnim uporabljenim časom $q = \frac{Q}{T}$, drugo merilo proizvodnosti dela pa tako, da celoten uporabljen čas delimo s celotno proizvodnjo $t = \frac{T}{Q}$. Ako se proizvodnost dela zveča, se q zveča, t pa zmanjša. Iz teh dveh enostavnih obrazcev vidimo, da ni pri izračunavanju koeficientov proizvodnosti dela nobenih težav, ako gre za en sam artikel, oziroma vrsto proizvodnje. Stvarno stanje pa seveda ni tako enostavno. Posamezno podjetje in niti posamezen delavec ne izdeluje samo ene vrste proizvednje, temveč veliko število najrazličnejših artiklov. V tem primeru si seveda z zgornjima enostavnima obrazcema ne moremo pomagati. Težava nastopi v tem, ker raznovrstne artikle med seboj ne moremo sešteti in proizvodnjo nekega podjetja ali delavca ne moremo izraziti v enem številu, da bi jo mogli deliti s časom.

b) Kot izhod za silo moremo uporabiti pogojne enote. Vsak artikel vrednotimo z določenim številom pogojnih enot. Čimveč dela zahteva izdelava nekega artikla, s tem večjim številom pogojnih enot ga vrednotimo. Na ta način dosežemo, da moremo skupno proizvodnjo raznovrstnih artiklov izraziti z enim samim številom pogojnih enot. Ako Q_i pomeni proizvodnjo posameznega artikla, e_i v pogojnih enotah izraženo važnost istega artikla, T_i pa za izdelavo Q_i potreben čas, moremo sumarno proizvodnost dela izraziti s kvocientom:

$$q = \frac{\sum Q_i e_i}{\sum T_i} \qquad (7.15)$$

Vendar ta način ni najboljši, ker je vrednotenje s pogojnimi enotami zelo težko izvedljivo in zelo podvrženo osebnim vplivom.

Kot drug način vrednotenja artiklov moremo uporabiti ceno. Ta način je stvarnejši od pogojnih enot, ker je vrednotenje bolj življensko. Ako z C_i zaznamujemo ceno enote je $Q_i C_i = P_i$ vrednost proizvodnje tega artikla $\sum Q_i C_i = \sum P_i$ pa vrednost celotne proizvodnje. Proizvodnost dela moremo v tem primeru meriti z vrednostjo proizvodnje na enoto časa

$$p = \frac{\sum Q_i C_i}{\sum T_i} = \frac{\sum P_i}{\sum T_i} \qquad (7.16)$$

Vendar naletimo pri tem načinu na težave, ki zelo zmanjšajo vrednost tega koeficienta. Prvo na kar moramo opozoriti je to, da moramo vrednosti proizvodnje izračunavati po stalnih, enotnih cenah, ker je vrednost proizvodnje odvisna od količine in cene.

Ker je $P_i = \frac{P_i}{T_i} T_i = p_i T_i$ moremo pisati

$$p = \frac{\sum p_i T_i}{\sum T_i} \qquad (7.17)$$

Sumatna proizvodnost dela p je torej tehtana aritmetična sredina iz proizvodnosti dela za posamezne artikle p_i.

Odvisna je torej od proizvodnosti dela posameznih artiklov p_i in strukture časa $\frac{T_i}{\sum T_i}$. Sumarna proizvodnost dela se more torej spremeniti tudi v primeru, ako pokazatelj proizvodnosti dela vseh posameznih artiklov ostanejo isti, ako se spremeni struktura časa. Še več: Proizvodnost dela vseh posameznih artiklov se more zvečati, zaradi istočasne spremembe strukture časa pa more sumarni pokazatelj proizvodnosti dela pokazati znižanje proizvodnosti dela, kar je razvidno iz shematičnega primera:

Tab.7.17

Artikel	Razdobje o			Razdobje 1			$I = (p_i : p_o)100$
	Po	To	po	Pi	Ti	pi	
A	600	12	50	972	18	54	108,0
B	700	5	140	640	4	160	114,3
Skupno	1300	17	76,5	1612	22	73,3	95,3

Iz primera je razvidno, da so se posamezne proizvodnosti dela pri obeh artiklih izvedle (indeks 108,0 in 114,3), sumarna proizvodnost pa kaže padec (indeks 95,3), kar je popolnoma nerealno. Zaradi tega ta način v primerih, kadar se struktura delovnega časa spreminja ni uporaben.

Ker zaradi spremembe strukture časa zgornji način ni v polni meri uporaben, dobimo pravilnejše rezultate, ako izračunavamo sumarno proizvodnost dela pri stalni strukturi časa.

Ako v obrazcih izpustimo znak i, z o pa zanzamujemo bazične vrednosti, z 1 tekoče vrednosti, moremo sumarno proizvodnost dela izračunati, ako pri obeh vzamemo strukturo časa tekočega razdobja, po obrazcih:

$$p_o = \frac{\sum p_o T_1}{\sum T_1} \qquad p_1 = \frac{\sum p_1 T_1}{\sum T_1}$$

Indeks proizvodnostidela, ki kaže pravilno stanje, pa moremo v tem primeru izračunati po obrazcu:

$$I_p = \frac{p_1}{p_o} = \frac{\sum p_1 T_1}{\sum T_1} : \frac{\sum p_o T_1}{\sum T_1} = \frac{\sum p_1 T_1}{\sum p_o T_1} \qquad (7.18)$$

Iz zgornjega obrazca vidimo, da je to oblika agregatnega indeksa za količino p.

Ako uporabimo ta obrazec na našem shematičnem primeru vidimo, da pokaže rezultat dejansko sliko.

$$I_p = \frac{54.18 + 160.4}{50.18 + 140.4} \cdot 100 = \frac{1612}{1460} \cdot 100 = 110,4$$

e) Zgornji način izračunavanja indeksa proizvodnosti dela predpostavlja, da poznamo cene, oziroma vrednosti poedinih artiklov. Vendar moremo sumarni indeks proizvodnesti dela izračunati tudi v primeru, da ne poznamo cen, ampak samo proizvedene količine. To moremo doseči, ako merimo proizvodnost dela s časom, uporabljenim za izdelavo enote.

$$t_o = \frac{T_o}{Q_o} \qquad t_1 = \frac{T_1}{Q_1}$$

Indeks proizvodnosti dela izračunamo v tem primeru kot agregatni indeks, s količinami tekočega razdobja kot ponderi po naslednjem obrazcu:

$$I_q = \frac{\sum Q_1 t_o}{\sum Q_1 t_1} = \frac{\sum Q_1 t_o}{\sum T_1} \qquad (7.19)$$

Način izračunavanja je razviden iz shematičnega primera:

Artikel	Razdobje 0			Razdobje 1		
	Q_o	T_o	$t_o = \frac{T_o}{Q_o}$	Q_1	T_1	$t_o Q_1$
A	150	180	1,2	160	176	192
B	280	224	0,8	270	202	216
C	40	60	1,5	42	64	63
Skupno					442	471

$$I_q = \frac{\sum t_o Q_1}{\sum T_1} \cdot 100 = \frac{471}{442} \cdot 100 = 106,1$$

.232 Izmed reprezentativnih indeksov sta najbolj znana reprezentativen indeks cen, pri katerem zasledujemo cene grup reprezentativnih artiklov, in indeks življenjskih stroškov. Indeks življenjskih stroškov je sestavljen na naslednji način. Na katerikoli način (fiziološki minimum, potom ankete iti) določimo artikle in količine artiklov, ki so potrebni za preživljanje posameznika ali družine v določenem razdobju. Za te artikle iščemo od meseca do meseca cene. Vsota produktov cen in določenih količin predstavlja strošek za preživljanje v določenem razdobju. Te stroške primerjamo med seboj v časovnem indeksu, ki predstavlja indeks življenjskih stroškov. Čeprav strošek v absolutnem iznosu nima posebnega pomena, ker se ljudje ne preživljajo po tej shemi, pokaže indeks življenjskih stroškov gibanje cen artiklov široke potrošnje. Zaradi pregleda nadgibanjem strukture stroškov izračunavamo poleg skupnega tudi delne indekse za skupine: 1. Prehrana, 2. Obleka in obutev, 3. Kurjava in razsvetljava, 4. Stanovanje, 5. Ostalo. V zvezi z indeksom nominalne plače izračunavamo iz indeksa življenjskih stroškov indeks realne plače $I_r = \frac{I_n}{I_{žs}}$ (I_r = indeks realne plače, I_n = indeks nominalne plače, $I_{žs}$ = indeks življenjskih stroškov).

Kot primer izračuna indeksa življenskih stroškov podajamo ponatis indeksa za mesec nov. 1919 iz revije "Indeks", ki jo je pred vojno v Zagrebu izdajal Benko Drado.

Mesečni življenjski stroški na osnovi teoretične najmanjše potrošnje za odraslega neoženjenega delavca (Benko Grado)

Oznaka potrošnega blaga	Enota mere	Mesečna količina	Povprečna cena v din		Skupni meseč. strošek v din		Indeks	
			julij 1914	1.nov. 1940	julij 1914	1.nov. 1940	julij 1914	1.nov. 1940
1. Krušna moka št.6	kg	8	0,34	4,75	2,72	38,0		
2. Koruzna moka	kg	4	0,28	5,50	1,12	22,0		
3. Krompir	kg	5	0,09	1,98	0,45	9,90		
4. Fižol	kg	3	0,28	8,90	0,84	26,70		
5. Riž (splendov)	kg	1	0,42	15,75	0,42	15,75		
6. Sladkor	kg	1,5	0,92	17,50	1,38	26,25		
7. Goveje meso	kg	3	1,40	18,50	4,20	55,50		
8. Ribe(belica)	kg	1	0,78	10,66	0,78	10,66		
9. Svinjska mast	kg	1,5	1,80	26,00	2,70	39,00		
10. Mleko	l	10	0,20	3,00	2,00	30,00		
11. Sir	kg	1,20	1,65	18,66	1,98	22,39		
12. Maslo	kg	0,25	3,00	38,33	0,75	9,58		
13. Jajca	kom	15	0,05	1,72	0,75	25,80		
14. Sadje	kg	4	0,60	6,90	2,40	27,60		
15. Kava	kg	0,20	3,00	89,33	0,60	17,87		
16. Kavin dodatek	kg	0,50	0,96	21,33	0,48	10,67		
17? Sol	kg	0,30	0,26	1,50	0,07	0,45		
18. Razno(zelenjava itd)	-	-	-	-	1,00	17,00		
I. Skupno hrana					24,64	405,12	100	1644
1. Moška suknja	kom	1/72	60,0	1729,59	0,83	24,02		
2. Moška obleka	gar.	1/12	50,0	1364,02	4,17	113,66		
3. Moška srajca	kom	1/3	5,20	96,71	1,73	32,33		
4. Moške sp. hlače	kom	1/3	3,50	66,84	1,16	22,21		
5. Kratke nogavice	par	1/2	1,00	21,66	0,50	10,83		
6. Visoki čevlji	par	1/6	12,0	255,18	2,00	42,53		
7. Moški klobuk	kom	1/12	6,0	112,63	0,50	9,39		
II. Skupno obleka in obutev					10,89	254,97	100	2341
1. Stanovanje	soba s kuhinjo	1/2	18,0	283,97	9,0	141,99	100	1577
III. Skupno stanovanje								
1. Bukova drva	kg	210	0,02	0,40	4,20	84,0		
2. Petrolej	l	1	0,44	7,00	0,44	7,00		
IV. Skupno kurjava razsvetljava					4,64	91,00	100	1961
V. Razno 10 °/° od skupno I-IV. (alkohol!)					4,91	89,31		
Skupno					54,08	982,39	100	1817

24,64 : 405,12 = 40(12 : 2464 = 16

1597,2

188

75. POVZETEK.

Osnova statističnega proučevanja je primerjava. Izmed vseh vrst primerjav pa najboljše podajajo odnose med posameznimi podatki relativna števila. Relativno število dobimo, da podatek, katerega hočemo primerjati, delimo s podatkom, s katerim ga hočemo primerjati. Na ta način ugotovimo kolikokrat je dan podatek večji od podatka s katerim ga primerjamo, ki ga imenujemo baza ali osnova.

V socialno ekonomski statistiki uporabljamo v glavnem tri vrste relativnih števil.

S strukturnimi ali razčlenitvenimi števili primerjamo podatke delne mase s podatki celotne mase. Odstotni delež pokaže kakšen del celote, ki je predstavljena s loo, odpade na dano delno maso. Na ta način dobimo podroben vpogled v sestav mase. Ako je statistična masa razčlenjena po enem samem znaku, govorimo o enostavni strukturi, ako pa je razčlenjena istočasno po več znakih, pa izračunavamo večkratne strukture. Pri statistični masi, ki je razčlenjena po dveh znakih, moremo izračunati troje vrst struktur, kix so znane pod imeni "kotni loo", "vrstični loo" in "stolpični loo". Vsaka izmed teh ima posebne možnosti analize.

S statističnimi koeficienti in gostotami primerjamo raznovrstne prirejene podatke, ki pa so v vsebinski zvezi. Izračunavanje in pomen koeficientov je odvisen od tega, ali primerjamo med seboj intervalni masi, intervalno maso z momentno, ali dve momentni masi. Statistične koeficiente uporabljamo v vseh vejah socialno - ekonomske statistike.

Z indeksi primerjamo prirejene istovrstne podatke. Ako primerjamo dva istovrstna podatka, ne glede na njun notranji sestav, izračunavamo enostavne indekse. Po tem, v katerem znaku se primerjani podatka razlikujejo, ločimo stvarne, krajevne in časovne. Najvažnejši in najobičajnejši so časovni indeksi. Splošen obrazec za izračunavanje enostavnih indeksov je:

$$I_{1/o} = \frac{Y_1}{Y_o} \cdot 100, \quad Y_o \text{ je baza.}$$

Ako izračunavamo indekse iz časovne serije, je baza običajno stalna. Moremo pa baze od člena do člena spreminjati. V tem primeru govorimo o indeksih s premično bazo. Najpogosteje izračunavamo verižne indekse, pri katerih posamezne člene časovne serije primerjamo s predhodnim členom.

Pri izbiri baze moramo upoštevati predvsem vsebinsko plat proučevane serije.

O mnogočlenskih ali agregatnih indeksih govorimo kadar upoštegamo, da so posamezne sumarne količine sestavljene iz več sestavin. Vrednost proizvodnje je n.pr. vsota produktov cen in količin. Z agregatnimi indeksi moremo ugotoviti vpliv spremembe enega samega elementa, bodisi cen ali količin.

Z Laspeyrovo formulo $I_q = \frac{\sum p_o q_1}{\sum p_o q_o}$ ugotavljamo spremembo obsega proizvodnje pri stalnih cenah, z Paschejevo formulo

$$I_p = \frac{\sum p_1 q_1}{\sum p_o q_1}$$

pa spremembo cen pri stalnih količinah.

V ekonomski statistiki se dostikrat poslužujemo reprezentativnih indeksov, ki so izračunani na osnovi reprezentativnih elementov. V socialno ekonomski praksi so najvažnejši agregatni indeksi proizvodnosti dela, reprezentativni indeksi cen, indeksi obsega proizvodnje in indeks življenjskih stroškov.

8 SREDNJE VREDNOSTI

80. SPLOŠNO

80.0 Pojm

Na posamezno enoto statistične mase deluje veliko število najrazličnejših vplivov. Skupni rezultat delovanja vseh teh vplivov na posamezno enoto statistične mase se izraža v vrednosti rezultativnega znaka te enote. Hektarski donos parcel, zasejanih s pšenico, je odvisen od najrazličnejših vplivov. Ker se pogoji od parcele do parcele menjajo, se menjo tudi hektarski donos, ki je rezultat teh menjajočih se pogojev. Vendar pri podrobnejšem proučevanju, vidimo, da se opredeljujoči pogoji od enote do enote ne menjajo, temveč da ostanejo stalni, so torej splošni. Zaradi tega imenujemo opredeljujoče pogoje splošne pogoje. Ostali pogoji, ki niso v sklopu opredeljujočih pogojev, pa se od enote do enote menjajo. Zaradi tega jih imenujemo individualne pogoje. Ker je hektarski donos posamezne njive rezultat vseh pogojev, moremo ta rezultat pri posamezni njivi razdeliti na dva dela: a) rezultat vplivov splošnih, nespremenljivih pogojev in b) rezultat individualnih, spremenljivih pogojev. Ker je prvi del rezultat stalnih, od enote do enote nespremenljivih pogojev, je po svojem iznosu za vse enote enak in predstavlja tipično vrednost. Rezultat individualnih pogojev se od enote do enote izpreminja, tako po svoji velikosti, kot po predznaku, ker morajo delovati individualni pogoji v smeri povečanja ali manjšanja tipične vrednosti. Metoda srednjih vrednosti omogoča, da moremo rezultat splošnih pogojev izračunati.

Ako je statistična masa, katero proučujemo, homogena, se v sklopu individualnih vplivov nahajajo samo slučajni vplivi. Slučajni vplivi se zaradi raznosmernega delovanja pri zadostno veliki statistični masi v vsoti uničijo. Zaradi te lastnosti moremo z metodo srednje vrednosti iz statistične mase izluščiti tipično - srednjo vrednost, ki je rezultat vpliva splošnih pogojev. Zakonitost, da se slučajni vplivi v velikem številu uničujejo, je ozko povezana z zakonom o velikih številih, oziroma je pravzaprav isti zakon, podan v drugi obliki. Ta zakonitost nam bo še na marsikaterem mestu omogočala pri analiziranju masovnih pojavov. Na kakšen način nam omogoča izluščiti tipično količino bomo videli v poglavju o aritmetični sredini.

Tipične srednje vrednosti zavzemajo po pravici temeljno mesto v statistiki in tvorijo osnove tako enostavne, kot višje analize masovnih pojavov. Ker srednja vrednost predstavlja cel niz vrednosti v enem samem številu, je z njimi zelo olajšana primerljivost med različnimi masovnimi pojavi.

Še važnejše pa je naslednje. Ako primerjamo v nekem letu povprečen hektarski donos jare pšenice na nekem področju s povprečnim hektarskim donosom ozimne pšenice na istem področju, dobimo dve različni povprečji, ako je hektarski donos odvisen od tega, ali je pšenica posejana jeseni ali spomladi. S tem doženemo, v koliko vpliva izpremememba posameznega pogoja na hektarski donos. Možnost analiziranja enega samega pogoja dosežemo s tem, da ohranimo pri obeh masah vse opredeljujoče pogoje (leto, rajen itd.)iste, razen onega, za katerega skušamo ugotoviti rezultat spremembe na hektarski donos. Za bivšo Jugoslavijo je n.pr. desetletno povprečje (1930 - 1939) hektarskega donosa jare pšenice 1.5 q/ha, ozimine pa 11.4 q/ha. Iz teh dveh povprečij vidimo, da je vpliv jarovizacije na hektarski donos pšenice zelo velik.

80.1 Metoda grupnih srednjih vrednosti

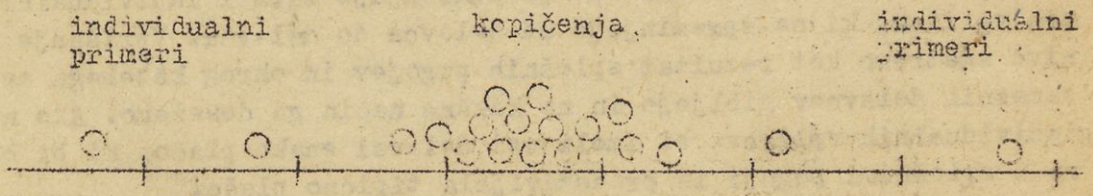
V prejšnjem odstavku smo omenili, da se rezultat individualnih vplivov uniči v povprečju le, ako imamo opravka s homogeno statistično maso, kjer nastopajo kot individualni vplivi samo slučajni vplivi. Zaradi tega nima z vidika iskanja tipičnih količin smisla izračunavati srednje vrednosti statističnih mas, ki niso homogene. Take srednje vrednosti, čeprav jih moremo formalno izračunati, so neznanstvene in imajo zelo omejeno analitično vrednost. Nesmiselno je za 11 ljudi, od katerih ima eden 1,000.000 din, ostalih 10 pa vsak po 10.000 din premoženja, trditi, da imajo povprečno po 100.000 din premoženja, kar pokaže formalen račun. Ta vrednost ni tipična niti za milionarja, niti za deset revežev in je torej neuporabna kot tipična vrednost. Da pa moremo kljub temu analizirati tudi nehomogene mase, si pomagamo s tem, da dano homogeno statistično maso razdelimo z grupiranjem na skupine homogenih statističnih mas. Sredine izračunamo za vsako homogeno-delno maso posebej. V tem primeru govorimo o grupnih sredinah. Skupne sredine zaradi primerjave z drugimi masami sicer tudi izračunavamo, vendar moramo biti pri uporabi teh sredin vedno oprezni in upoštevati, da je to skupna sredina (Več o grupnih sredinah glej v odstavku o aritmetični sredini aritmetičnih sredin 81.7).

80.2 Metoda individualnega primera

Proučujmo delovni učinek kvalificiranih delavcev določenega oddelka nekega podjetja. Statistično pogledano, je ta masa delavcev na prvi pogled homogena. Odkloni delovnega učinka posameznih delavcev izgledajo na prvi pogled podvrženi samo slučajnim vplivom. Vendar se izkaže, da se ta odklon, ako je slučajen, nahaja v določenih mejah. Ako je ta odklon le previelik in izpada iz sklopa ostalih vrednosti, ne moremo smatrati, da delavec, ki ima od drugih delavcev tako različen delovni učinek, spada v sklop ostalih delavcev. Statistično rečeno, da je ta individualen primer izpadel iz homogene mase, da predstavlja torej posebno kvaliteto. Ako je storilnost nekega delavca znantno večja od povprečne, je zelo verjetno, da delo vrši po postopku, ki je boljši od postopka drugih. Nikakor ne bi bilo pravilno, da

bi se taki individualni primeri vtopili v splošnem povprečju, temveč jih je treba proučiti in skušati ta način dela prenesti tudi na ostale delavce. Na isti način odkrije znaten odklon od povprečne storilnosti navzdol lenuha ali celo saboterja. Take izredne primere proučujemo zato vedno individualno in jih ne vključujemo v maso od katere računamo povprečje. Grafično je zgornji problem ponazorjen na spodnji sliki.

Sl. 8.1



Zgornji primer je samo eden, kjer je možno uporabiti metodo individualnega primera. Postopek je možen in smiseln pri masovnih pojavih iz najrazličnejših področij, reči moremo, da v vseh primerih, kjer uporabljamo sredine.

Iz vsega dosedanjega sledi, da ima metoda srednjih vrednosti polno analitično vrednost le, ako jo povežemo z metodo grupnih sredin in metodo individualnega primera. Od tod sledi tesna povezanost srednjih vrednosti z grupiranjem.

80.3 Vrste srednjih vrednosti, ki jih uporabljamo v socialno-ekonomski statistiki

Srednje vrednosti, ki jih uporabljamo v statistiki delimo na izračunane in srednje vrednosti po legi. Izračunane srednje vrednosti iz vrednosti znakov izračunavamo, srednje vrednosti po legi pa so določene z lego vrednosti.

Od izračunanih srednjih vrednosti bomo obravnavali:

a) aritmetično sredino, kot srednjo vrednost, ki se v praksi najpogosteje uporablja,
b) harmonično sredino, kot srednjo vrednost, ki se uporablja v nekaterih primerih namesto aritmetične sredine,
c) geometrično sredino, ki jo uporabljamo pri pokazateljih dinamike.

Med srednje vrednosti po legi pa štejemo: a) mediano b) modus.

81. ARITMETIČNA SREDINA

81.1 Navadna aritmetična sredina

.11 Proučujemo mesečne plače devetih kvalificiranih delavcev v neki industrijski panogi v letu 1951. Plače za posamezne delavce so naslednje: 4527 din, 4336 din, 4628 din, 4446 din, 4580 din, 4425 din, 4552 din, 4502 din. Posamezne plače se med seboj razlikujejo, vendar se drže na določenem nivoju. Ta nivo je pogojen s splošnimi pogoji, kot so n.pr. kvalifikacija, industrijska panoga, leto, za katero veljajo plače. Plače za posameznega delavca se od tega nivoja odklanjajo zaradi individualnih, slučajnih vplivov, ki se spreminjajo od delavca do delavca. Vprašanje je, kateri nivo smatramo kot rezultat splošnih pogojev in okrog katerega se plače posameznih delavcev gibljejo in na kakšne način ga dosežemo. Ako ne bi bilo individualnih vplivov, bi imeli vsi delavci enako plačo, ki bi bila pogojena z splošnimi pogoji in predstavljala tipično plačo.

Tipično plačo izračunamo v našem primeru na ta način, da seštejemo plače vseh devetih delavcev, dobimo skupen fond plač, ta fond pa delimo z 9. Ako zaznamujemo to vrednost, ki jo imenujemo povprečno plačo, z A, je:

$$A = \frac{1}{9}(4527 + 4336 + 4628 + 4446 + 4396 + 4580 + 4425 + 4525 + 4502) =$$

$$= 1/9 \cdot (40392) = 4488$$

Povprečna plača delavca znaša 4488 din.

Individualna plača posameznega delavca je torej sestavljena iz dveh delov: povprečne plače A = 4488 in rezultata vpliva individualnih pogojev. Ta del more biti, kot bomo videli, pozitiven ali negativen. Za posamezne delavce dobimo:

Tab. 8.1

Mesečna plača	Rezultat			
	splošnih pogojev		individualnih pogojev	
4527	= 4488	+	(+ 39)
4336	= 4488	+	(- 152)
4628	= 4488	+	(+ 140)
4446	= 4488	+	(- 42)
4396	= 4488	+	(- 92)
4580	= 4488	+	(+ 92)
4425	= 4488	+	(- 63)
4552	= 4488	+	(+ 64)
4502	= 4488	+	(+ 14)
40392	= 9.4488	+	0

Ako vseh devet enačb seštejemo dobimo na levi strani skupni plačilni fond 40392 din. Na desni da vsota prvega stolpca 9.4488 din, vsota drugega stolpca pa je enaka 0. Iz tega moremo zaključiti lastnosti:

a) Vsota rezultatov individualnih pogojev je enaka nič.

b) Vsota individualnih plač je enaka produktu povprečne plače s številom delavcev. Iz tega sledi, da je povprečna plača enaka kvocientu vsote individualnih plač in števila delavcev.

A = 40392/9 = 4488. Te lastnosti pa ne veljajo samo za ta primer, temveč jih moremo posplošiti.

.12 Posplošenje postopka. Postopek, ki smo ga izvedli na primeru plač, moremo uporabiti za najrazličnejše pojave. Zaradi tega ga bomo posplošili in pisali v simbolih, da bo preglednejši in splošno veljaven. Posamezne plače nadomestimo s simboli x_1, x_2 - x_n, ki pomenijo vrednosti kateregakoli numeričnega znaka neke statistične mase. Število delavcev pa zaznamujemo z splošnim številom enot neke statistične mase, ki ga zaznamujemo z n. Individualne odklone od tipične - povprečne vrednosti A, zaznamujemo z d_1, d_2 ..... d_n. V tem primeru moremo pisati:

$$x_1 = A + d_1$$
$$x_2 = A + d_2$$
$$x_3 = A + d_3$$
$$\cdot \quad \cdot$$
$$\cdot \quad \cdot$$
$$x_n = A + d_n$$

$$x_1 + x_2 + x_3 + \ldots + x_n = nA + (d_1 + d_2 + d_3 + \ldots d_n)$$

ali krajše:

$$\sum_{i=1}^{n} x_i = nA + \sum_{i=1}^{n} d_i$$

Znak $\sum$ pomeni vsoto vrednosti vseh enot statistične mase.

Ker je vsota rezultatov individualnih vplivov enaka nič.

$$\sum_{i=1}^{n} d_i = 0 \text{ velja: } \sum_{i=1}^{n} x_i = nA$$

Iz tega moremo izračunati splošen obrazec za količino A, ki jo imenujemo povprečje ali navadno aritmetično sredino:

$$A = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (8.1)$$

ali ako izpustimo znak i

$$A = \frac{1}{n} \sum x \qquad (8.2)$$

Iz gornjega obrazca sledi postopek izračunavanja.
Navadno aritmetično sredino izračunamo tako, da:
1./ Izračunamo vsote vrednosti znaka za vse enote.
2./ Vsoto vseh vrednosti delimo s številom enot.

Aritmetično sredino zaznamujemo za razliko od drugih sredin običajno z A. Ako gre za aritmetične sredine več znakov, jih zaradi tega, da jih med seboj razlikujemo, zaznamujemo z A_x, A_y itd.

Ker pa aritmetično sredino izmed vseh sredin najpogosteje izračunavamo, jo zaznamujemo običajno z krajšim simbolom: prečno črtico nad simbolom znaka, za katerega izračunamo aritmetično sredino.

$$\frac{1}{n}\sum x = A_x = \bar{x} \qquad \frac{1}{n}\sum y = A_y = \bar{y} \quad \text{itd.}$$

Iz zveze $\bar{x} = \frac{1}{n}\sum x$ navadno izračunavamo aritmetično sredino iz znane vsote vrednosti in števila primerov oziroma enot. Vendar je važno tudi dejstvo, da je:

$$\sum x = n\bar{x} \qquad (8.3)$$

Iz tega obrazca vidimo, da je vsoto vrednosti vseh enot neke statistične mase možno izračunati, ako poznamo aritmetično sredino in število enot. Ako imamo znano število gospodinjstev in povprečno število članov v enem gospodinjstvu, moremo izračunati skupno število članov v teh gospodinjstvih tako, da število gospodinjstev pomnožimo s povprečnim številom članov na eno gospodinjstvo itd.

P . . . Imamo statistično maso 45 gospodarstev. Od teh jih je 23 iz skupine do 2 ha skupne površine. Število krav v posameznih gospodarstvih je naslednje: 1,0,0,1,2,0,2,1,1,1,1,3,2,1,1,2,4,1,1,0,2,1,1.

Iz skupine od 2,01 - 5,00 ha skupne površine je 15 gospodarstev. Število krav po gospodarstvih je: 3,1,2,2,4,3,2,2,2,1,3,3,2,5,5. Ostalih 7 gospodarstev pa je iz skupine nad 5,01 ha skupne površine s sledečim številom krav po posameznih gospodarstvih: 4,5,8,2,5,6,3.

Vpliv velikosti gospodarstev na število krav je možno proučiti, ako tvorimo grupne aritmetične sredine in primerjamo med seboj te grupne sredine.

$$A_1 = \frac{1}{23}(1+0+0+1+2+0+2+1+1+1+1+3+2+1+1+2+4+1+1+0+2+1+1) = \frac{29}{23} = 1,10$$

$$A_2 = \frac{1}{15}(3+1+2+2+2+4+3+2+2+1+3+3+2+5+5) = \frac{40}{15} = 2,67$$

$$A_3 = \frac{1}{7}(4+5+8+2+5+6+3) = \frac{33}{7} = 4,71$$

Iz teh treh aritmetičnih sredin je razvidno, da z velikostjo gospodarstev raste tudi povprečno število krav na eno gospodarstvo, kar je za analizo zelo važna ugotovitev.

81.2 Važni stavki o aritmetičnih sredinah.

.21 Pri nadaljnih proučevanjih bomo potrebovali naslednje stavke o aritmetičnih sredinah:

a) Aritmetična sredina konstante je konstanta sama

$$\bar{a} = a \tag{8.4}$$

b) Aritmetična sredina vsote dveh znakov je enaka vsoti aritmetičnih sredin obeh znakov

$$\overline{(x + y)} = \bar{x} + \bar{y} \tag{8.5}$$

c) Iz gornjih dveh stavkov moremo izpeljati:

$$\overline{(a + x)} = \bar{a} + \bar{x}, \qquad \overline{bx} = b\bar{x} \tag{8.6}$$

D Dokaz: Veljavnost zgornjih stavkov se da zelo hitro razvideti na primeru.

a) Ako bi imelo vseh devet delavcev iz našega začetnega primera enako plačo a, bi imeli skupaj $(a+a+a+a+a+a+a+a+a) = 9\,a$. Ako skupen plačilni fond 9 a delimo s številom delavcev $9\,a : 9 = a$, dobimo kot aritmetično sredino vrednost a.

b) Ako vzamemo, da se delavčeva plača sestoji iz dveh delov, osnovne plače x in dodatka y je povprečna plača:

$$\overline{(x + y)} = \frac{1}{9}\left[(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) + \ldots + (x_9 + x_9)\right] =$$

$$= \frac{1}{9}\left[(x_1 + x_2 + x_3 + \ldots + x_9) + (y_1 + y_2 + y_3 + \ldots + y_9)\right] =$$

$$= \frac{1}{9}\left[(x_1 + x_2 + x_3 + \ldots + x_9\right] + \frac{1}{9}\left[y_1 + y_2 + y_3 + \ldots + y_9\right] =$$

$$= \bar{x} + \bar{y}$$

Stavek se more raztegniti na več znakov in je:

$$\overline{(x + y + z)} = \bar{x} + \bar{y} + \bar{z}$$

c) Na enak način je:

$$\overline{(a + x)} = \frac{1}{n}\left[(a + x_1) + (a + x_2) + (a + x_3) + \ldots + (a + x_n)\right] =$$

$$= \frac{1}{n}\left[n\,a + (x_1 + x_2 + x_3 + \ldots + x_n)\right] = a + \bar{x} \text{ in}$$

$$\overline{bx} = \frac{1}{n}\left[bx_1 + bx_2 + \ldots + bx_n\right] = \frac{1}{n}\left[b(x_1 + x_2 + x_3 + \ldots + x_n)\right] =$$

$$= b\,\frac{1}{n}\left[x_1 + x_2 + x_3 + \ldots + x_n\right] = b\bar{x}$$

.22 Olajšan način izračunavanja navadne aritmetične sredine. Iz gornjih stavkov moremo izpeljati postopek, ki v primerih, kadar so vrednosti precej velike in se gibljejo okoli neke vrednosti, skrajša izračunavanje aritmetične sredine.

Vzemimo neko poljubno okroglo vrednost x_o, ki je dejanskim vrednostim blizu. To vrednost imenujemo pogojno sredino.
Vrednost znaka x moremo pisati:

$x = x_o + (x - x_o)$, iz česar sledi: $\bar{x} = \overline{x_o + (x - x_o)} = x_o + \overline{(x - x_o)}$ (8.7)

Stavek se z besedami glasi: Aritmetična sredina nekega znaka je enaka vsoti pogojne sredine x_o in aritmetične sredine odklonov od pogojne sredine.

P Vrednost proizvodnje nekega podjetka po mesecih v ccc din je naslednja: kot pogojno sredino vzemimo $x_o = 1280$ tisoč

Tab. 8.2

	x	x - x_o
I	1287,6	7,6
F	1286,5	6,5
M	1282,4	2,4
A	1283,6	3,6
M	1289,2	9,2
I	1288,9	8,9
J	1286,4	6,4
A	1288,5	8,5
S	1284,6	4,6
O	1290,5	10,5
N	1291,2	11,2
D	1292,4	12,4
	15451,8	91,8

Povprečna mesečna proizvodnja je enaka:

$\bar{x} = 1280 + \frac{1}{12} 91{,}8 = 1280 + 7{,}65 = 1287{,}65$

81.3 Tehtana aritmetična sredina.

31 Navadno aritmetično sredino smo izračunali tako, da smo vrednost vseh enot dejansko sešteli. Poglejmo pa, ali se v primeru, da so vrednosti nekaterih enot med seboj enake, izračunanje aritmetične sredine da skrajšati.

Vzemimo 12 mater in opazujmo število otrok. Individualni podatki so naslednji: 1,2,1,2,2,3,3,1,1,1,2,1 otrok.

Aritmetično sredino moremo v tem primeru izračunati na dva načina:

1./ Posamezne vrednosti seštejemo in delimo z 12.

$$\bar{x} = \frac{1}{12}(1 + 2 + 1 + 2 + 2 + 3 + 3 + 1 + 1 + 1 + 2 + 1) = \frac{20}{12} = 1,67$$

2./ Vsoto členov, ki imajo enako vrednost, moremo dobiti tudi na ta način, da vrednost pomnožimo s številom enot, ki imajo to vrednost. Skupna vsota pa je enaka vsoti teh delnih produktov.

$$\bar{x} = \frac{6.1 + 4.2 + 2.3}{12} = \frac{6 + 8 + 6}{12} = \frac{20}{12} = 1,67$$

Vidimo, da smo dobili isti rezultat.

Ta način izračunavanja aritmetične sredine imenujemo tehtan in pravimo, da vsako vrednost, ki nastopa (1,2,3) tehtamo s številom enot, ki ima odgovarjajočo vrednost, (6,4,2). Števila, s katerimi tehtamo posamezne vrednosti, imenujemo teže. Tuj izraz za težo je ponder, samo izračunavanje pa imenujemo s tujim izrazom ponderirano aritmetično sredino.

.32 Posplošenje. Pri podrobnejši proučitvi ponderov vidimo, da so ponderi frekvence frekvenčne distribucije, ako se vrednosti grupirane v frekvenčni distribuciji. Za zgornji primer dobimo:

x_k	n_k
1	6
2	4
3	2
12	12

Število enot je enako vsoti vseh frekvenc, oziroma ponderov. Ako zaznamujemo grupne vrednosti z x_k, frekvence pa z n_k, se glasi splošen obrazec za izračunavanje tehtane aritmetične sredine:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \ldots + n_k x_k}{n_1 + n_2 + n_3 + \ldots + n_k} \qquad (8.8)$$

sli pisano krajše:

$$\bar{x} = \frac{1}{n} \sum_k n_k x_k \qquad n = \sum n_k \qquad (8,9)$$

Tehtano aritmetično sredino izračunavamo tako, da:

a) Grupne vrednosti pomnožimo s pripadajočimi ponderi.

b) Seštejemo produkte in ponderе.

c) Vsoto produktov iz a) delimo z vsoto ponderov.

Obrazec omogoča torej izračunavanje aritmetične sredine, ako so vrednosti dane v frekvenčni distribuciji.

P 197 stanovanj je v frekvenčni distribuciji razdeljenih po številu stanovalcev. Izračunati je treba aritmetično sredino.

Tab. 8.3

Število stanovalcev x_k	Število stanovanj n_k	$n_k x_k$
(1)	(2)	(3)=(1).(2)
1	5	5
2	14	28
3	29	87
4	43	172
5	29	145
6	30	180
7	26	182
8	11	88
9	5	45
10	3	30
11	2	22
$\sum$	197	984

$$\sum n_k x_k = 984$$

$$\sum n_k = 197$$

$$\bar{x} = \frac{984}{197} = 5$$

Povprečno število stanovalcev na eno stanovanje, je 5.

$\sum n_k x_k = 984$ pomeni skupno število vseh stanovalcev v 197 stanovanjih.

8.4 Izračunavanje aritmetične sredine iz relativnih frekvenc.

Ako v prejšnjem primeru posamezne frekvence delimo s skupnim številom vseh enot, dobimo relativne frekvence. Relativne frekvence so torej strukturni deleži za posamezne vrednosti znaka.

$$f_1 = \frac{n_1}{n} = \frac{5}{197} \quad f_2 = \frac{n_2}{n} = \frac{14}{197} \quad f_3 = \frac{n_3}{n} = \frac{29}{197} \ldots \quad f_{11} = \frac{n_{11}}{n} = \frac{2}{197}$$

Vsota vseh relativnih frekvenc je enaka 1, ker je:

$$f_1 + f_2 + f_3 + \ldots + f_r = \frac{n_1}{n} + \frac{n_2}{n} + \frac{n_3}{n} + \ldots + \frac{n_r}{n} =$$

$$= \frac{1}{n}(n_1 + n_2 + n_3 + \ldots n_R) = \frac{n}{n} = 1$$

Vprašanje je, ali je možno iz relativnih frekvenc f_k izračunati tehtano aritmetično sredino in kako. Izkaže se, da je to možno, in da je aritmetična sredina enaka vsoti produktov relativnih frekvenc z grupnimi vrednostmi.

$$\bar{x} = f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots + f_r x_r = \sum f_k x_k \qquad (8.10)$$

D Dokaz: Dokaz je razmeroma enostaven:

$$\bar{x} = \frac{1}{n} \sum n_k x_k = \sum \frac{n_k}{n} x_k = \sum f_k x_k$$

Ker je n konstanten, ga moremo nesti pod znak seštevanja, $\frac{n_k}{n}$ pa je relativna frekvenca f_k.

Ako so relativne frekvence dane v odstotkih, je treba vsoto produktov procentov in vrednosti znakov deliti na koncu z 100.

P Kot primer vzemimo izračunavanje povprečne tržne cene nekega artikla. Z oceno je bilo določeno, da se je 50 % nekega artikla prodajalo po ceni 8 din za kg, 30 % po ceni 7 din za kg, 20 % pa po ceni 11 din kg. Kolika je povprečna cena?

$$\bar{x} = \frac{30.7 + 50.8 + 20.11}{100} = \frac{210 + 400 + 220}{100} = \frac{830}{100} = 8{,}3 \text{ din}$$

81.5 Izračunavanje tehtane aritmetične sredine intervalnih frekvenčnih serij.

Kadar razredi členov frekvenčne serije niso posamezne vrednosti, temveč vsebujejo več vrednosti nezveznega znaka, ali pa, kadar gre za frekvenčno distribucijo zveznega znaka, govorimo o intervalnih serijah.

P Primer nezvezne intervalne serije imamo, ako v frekvenčni distribuciji stanovanj po stanovalcih iz 81.3 vzamemo v grupo po dve vrednosti.

število stanovalcev	število stanovanj
1 - 2	19
3 - 4	72
5 - 6	59
7 - 8	37
9 - 10	8
11 - 12	2
	197

V tej frekvenčni seriji za posamezne člene ne vemo točnih vrednosti, ki jih imamo v posameznih razredih. Vemo samo to, da je 19 stanovanj zasedenih z enim ali dvema stanovalcema, 72 s tremi ali štirimi itd. V tem primeru nimamo dane točne grupne vrednosti, temveč samo interval, v katerem se nahaja določeno število enot. Na prvi pogled izgleda, da v takih distribucijah ne bo možno izračunati aritmetične sredine, ker ne vemo, s katero vrednostjo naj množimo razredu odgovarjajočo frekvenco. Vendar moremo izračunati pravi aritmetični sredino, ki jo dobimo, ako izhajamo iz individulanih vrednosti, zelo podobno vrednost, ako vzamemo v posdinih razredih kot grupno vrednost - sredino razreda. Aritmetično sredino izračunamo v tem primeru kot tehtano aritmetično sredino sredin razredov s pripadajočimi frekvencami kot ponderi.

Tab. 8.4

število stanovalcev (1)	število stanovanj (2)	sredina razreda (3)	(4)=(2).(3)
1 - 2	19	1,5	28,5
3 - 4	72	3,5	252,o
5 - 6	59	5,5	324,5
7 - 8	37	7,5	277,5
9 - 1o	8	9,5	76,o
11 - 12	2	11,5	23,o
$\sum$	197	-	981,5

$$\bar{x} = \frac{981.5}{197} = 4,982$$

Vidimo, da je rezultat 4,982 zelo blizu pravemu rezultatu 5,•, čeprav je izračunan iz grupiranih podatkov.

P Rezultat je prilično dober tudi v primeru, ako grupiramo po tri vrednosti:

Tab.8.5

število stanovalcev (1)	število stanovanj (2)	sredina razreda (3)	(2).(3)
1 - 3	48	2	96
4 - 6	1o2	5	51o
7 - 9	42	8	336
1o - 12	5	11	55
$\sum$	197	-	997

$$\bar{x} = \frac{997}{197} = 5,o6$$

Jasno je, da moramo pričakovati tem večjo netočnost aritmetične sredine, čim večja je širina razreda.

Sredine razreda se poslužujemo kot grupne vrednosti tudi v primerih zvezne frekvenčne serije.

P Kor primer navajamo frekvenčno serijo teže po tisoč zrn iz 378 parcel po m^2. Podatki, ki so vzetiiz gradiva poizkusne metraže v Sloveniji l. 1948, se nanašajo na parcelice cele Slovenije. Izračunati je treba povprečno težo po looo zrn.

Tab. 8.6

teža v gramih	število zrn n_k	sredina razreda x_k	$n_k \cdot x_k$
(1)	(2)	(3)	(4)=(2).(3)
nad 20 - 24	1	22	22
nad 24 - 28	12	26	312
nad 28 - 32	28	30	840
nad 32 - 36	96	34	3264
nad 36 - 40	135	38	5130
nad 40 - 44	65	42	2730
nad 44 - 48	23	46	1058
nad 48 - 52	7	50	350
nad 52 - 56	4	54	216
nad 56 - 60	2	58	116
$\sum$	373	-	14038
$\frac{1}{n}\sum$	1		37.64 g

$$\bar{x} = \frac{14038}{373} = 37{,}64 \text{ g}$$

81.6 Postopki okrajšanega izračunavanja tehtane aritmetične sredine.

.61 Uporaba pogojne sredine. Frekvenčne serije imajo običajno enake širine razredov, ako ni kakega posebnega razloga, da vzamemo razrede z neenako širino (tipološko grupiranje). V primerih, kadar gre za serije z enakimi razredi, moremo z vpeljavo tako imenovane pogojne sredine, ki je poljubna grupna vrednost sredi serije, in jo poznamo že iz 81.22 izračunavanje znatno skrajšati. Ta metoda se posebno obnese, ako gre za grupne vrednosti, ki so večmestna števila.

Stopnje izračunavanja so v tem primeru naslednje:

1./ Sredi serije izberemo poljubno grupno vrednost kot pogojno sredino x_o.

2./ Tej vrednosti pripišemo novo grupno vrednost 0, ostalim členom pa od tega člena zaporedoma navzgor - 1, -2, -3,..., zaporedoma navzdol pa + 1, + 2, + 3 .... Te vrednosti vzamemo kot grupne vrednosti novega znaka u.

3./ Izračunamo tehtano aritmetične sredino znaka u s frekvencami kot ponderi.

4./ Aritmetična sredina prvotnega znaka x je enaka:

$$\bar{x} = x_o + \triangle \bar{u}$$

kjer je x_o pogojna sredina, $\triangle$ širina razreda, $\bar{u}$ pa tehtana aritmetična sredina znaka u.

P Kot primer vzemimo frekvenčno serijo iz tab. 8.4

Tab. 8.7

število stanovalcev	število stanovanj n_k	sredina razreda x_k	u_k	$n_k u_k$
(1)	(2)	(3)	(4)	(2).(4)=(5)
1 - 2	19	1,5	- 2	- 38
3 - 4	72	3,5	- 1	- 72
5 - 6	59	5,5	0	0
7 - 8	37	7,5	+ 1	+ 37
8 - 10	8	9,5	+ 2	+ 16
11 - 12	2	11,5	+ 3	+ 6
$\sum$	197	-	-	- 51

$x_o = 5,5$

$\triangle = 2$

$\bar{u} = \frac{-51}{197} = -0,259$

$\bar{x} = x_o + \triangle \bar{u}$

$\bar{x} = 5,5 - 2 \cdot 0,259 =$

$= 5,5 - 0,518 =$

$= 4,982$

Rezultat se sklada z rezultatom, dobljenim na drug način.

D Dokaz: Izpeljava postopka je naslednja:

Med x, x_o $\triangle$ in u obstoja za vse grupne vrednosti enaka zveza.

$x = x_o + \triangle u$ V našem primeru je: $x = x_o + \triangle \cdot u$

$1,5 = 5,5 + 2\ (-2)$
$3,5 = 5,5 + 2\ (-1)$
$5,5 = 5,5 + 2\ (\ 0)$
$7,5 = 5,5 + 2\ (+1)$
$9,5 = 5,5 + 2\ (+2)$
$11,5 = 5,5 + 2\ (+3)$

Ako pa obstoja ta zveza med posameznimi vrednostmi velja tudi za aritmetično sredino.

$$\bar{x} = \overline{(x_o + \triangle u)} = x_o + \triangle \bar{u}$$

$\bar{x}$ moremo torej izračunati po tem obrazcu iz količin x_o, $\triangle$ in u

.62 Uporaba kumulativnih serij. Zelo praktična je za izračunavanje aritmetičnih sredin tudi metoda kumulativnih serij.

Ta metoda ima naslednje stopnje:

1./ Iz členov frekvenčne serije izračunamo kumulativno serijo na način, kot je prikazan v primeru. Končni člen, ki izpade iz serije, je enak številu vseh enot in ga zaznamujemo s N_o.

2./ Izračunamo vsoto vseh členov kumulativne serije, razen N_o. To vsoto zaznamujemo s N'_o.

3./ Aritmetično sredino izračunamo po obrazcu:

$$\bar{x} = x_o + \triangle \frac{N'_o}{N_o} \qquad (8.12)$$

kjer je x_o vrednost sredine razreda prvega člena, $\triangle$ širina razreda, N_o in N'_o pa količini, izračunani iz kumulativne serije.

P Kot prvi primer ohranimo naš dosedanji primer razdelitve stanovanj po številu stanovalcev:

število stanovalcev (1)	število stanovanj (2)	Kumulativna serija 19 + 178 = 197 = N_o
1 - 2	19	72 + 1o6 = 178
3 - 4	72	59 + 47 = 1o6
5 - 6	59	37 + 1o = 47
7 - 8	37	8 + 2 = 1o
9 - 1o	8	2 + 0 = 2
11 - 12	2	0

178 + 1o6 + 47 + 1o + 2 = 343 = N'_o

$$\bar{x} = x_o + \triangle \frac{N'_o}{N_o} = 1,5 + 2\ \frac{343}{197} = 1,5 + \frac{686}{197} = 1,5 + 3,482 = 4,982$$

Primer je izdelan v razviti obliki, da je viden potek postopka. V praksi se račun skrajša, ker ne pišemo, na kakšen način smo izračunali posamezne količine. Da bo jasno razvidno, kako postopamo, da najhitreje izračunamo aritmetično sredino, postopke ponavljamo še na enem primeru. Postopek se posebno obnese, ako računamo kumulative na električnem stroju s kontrolnim trakom.

P ... Iz gradiva kontrole popisa živine leta 195o za tedanjo ljubljansko oblast smo kontrolirana gospodarstva do 1o ha skupne površine razdelili najprej po velikosti skupne površine v tri skupine. Prva skupina obsega gospodarstva do dva ha, druga nad dva do pet ha, tretja pa nad 5 do 1o ha skupne površine.

Za vsako izmed teh grup gospodarstev je bila sestavljena frekvenčna serija gospodarstev po številu krav.

Tab. 8.8

Vel. skupina / Štev. krav	0-2,oo ha	2,o1 - 5,oo ha	5,o1 - 1o,oo ha
0	25o	43	42
1	185	127	69
2	128	182	165
3	33	127	2o8
4	7	54	139
5	3	15	71
6	1	3	21
7	-	-	8
8	-	-	2
9	-	-	1
	6o7	551	726

Izračunati je treba za vse tri serije z metodo kumulativ aritmetično sredine.

Tab. 8.9

Vel. skupina / štev. krav	- 2,oo ha		2,o1 - 5,oo ha		5,o1 - 1o,oo ha	
	n_k	N_o=6o7	n_k	N_o=551	n_k	726=N_o
0	25o	357	43	5o8	42	684
1	185	172	127	381	69	615
2	128	44	182	199	165	45o
3	33	11	127	72	2o8	242
4	7	4	54	18	139	1o3
5	3	1	15	3	71	32
6	1	-	3	-	21	11
7	-	-	-	-	8	3
8	-	-	-	-	2	1
9	-	-	-	-	1	-
	N'_o	= 589	N'_o	= 1181	N'_o	= 2141

$$\bar{x} = x_o + \Delta \frac{N'_o}{N_o}$$

$$\bar{x}_1 = 0 + 1 \cdot \frac{589}{607} = 0{,}97 \text{ krav}$$

$$\bar{x}_2 = 0 + 1 \cdot \frac{1181}{551} = 2{,}14 \text{ krav}$$

$$\bar{x}_3 = 0 + 1 \cdot \frac{2141}{726} = 2{,}95 \text{ krav}$$

Iz aritmetičnih sredin je lepo razvidno, da povprečno število krav raste s površino gospodarstva, skupna površina je torej eden izmed faktorjev, ki vplivajo na stanje živine. Iz primera je razvidna velika enostavnost izračunavanja aritmetičnih sredin, ako se poslužimo metode kumulativnih serij.

D Dokaz: Ako proučimo postopek s kumulativno serijo na primeru, vidimo, da ga moremo privesti na metodo pogojne sredine:
Vzemimo primer razdelitve stanovanj po številu stanovalcev:

Tab.8.1o

Število razred	stanovalcev sred.razr.	Število stanovanj nk	197 = 2 + 8 + 37 + 59 + 72 + 19 = N_o
1 - 2	1,5	19	178 = 2 + 8 + 37 + 59 + 72
3 - 4	3,5	72	1o6 = 2 + 8 + 37 + 59
5 - 6	5,5	59	47 = 2 + 8 + 37
7 - 8	7,5	37	1o = 2 + 8
9 - 1o	9,5	8	2 = 2
11 - 12	11,5	2	0
		N'_o =	343 = 5,2+4,8+3,37+2,59+1,72+o,19

Ako računamo aritmetično sredino s pogojno sredino $x_o = 1{,}5$ (v splošnem sredina prvega razreda) je izračun aritmetične sredine naslednji:

Tab. 8.11

Število razred	stanovalcev sred.razr.	število stanovanj	u_k	$n_k \cdot u_k$
(1)	(2)	(3)	(4)	(3).(4)=(5)
1 - 2	1,5	19	0	o,19
3 - 4	3,5	72	+ 1	+ 1,72
5 - 6	5,5	59	+ 2	+ 2,59
7 - 8	7,5	37	+ 3	+ 3,37
9 - 1o	9,5	8	+ 4	+ 4,8
11 - 12	11,5	2	+ 5	+ 5,2
		197		343

Ako uporabljamo pogojno sredino je :

$$\bar{x} = x_o + \Delta\, \bar{u} \qquad \bar{x} = 1{,}5 + 2\,\frac{343}{197}$$

Iz primerjave sestava posameznih sestavin vidimo, da je vsota stolpca (3) v drugem načinu enaka N_o, vsota stolpca (5) pa N'_o

$\bar{u}$ je torej enak kvocientu količin N'_o in N_o, aritmetična sredina $\bar{x}$ pa $\bar{x} = x_o + \Delta\,\frac{N'_o}{N_o} \cdot x_o$ je sredina razreda prvega člena, Δ širina razreda, N_o in N'_o pa količino izračunane iz kumulativne serije.

81.7 Aritmetična sredina aritmetičnih sredin.

Iz grupnih aritmetičnih sredin moremo izračunati aritmetično sredino celotne mase. Iz povprečnih plač delavcev za posamezne stroke moremo izračunati povprečno plačo delavcev za celotno industrijo, iz povprečnih cen za nek artikel po podjetjih moremo izračunati povprečno ceno tega artikla za vso državo itd.

Skupno aritmetično sredino iz grupnih sredin izračunamo kot tehtano aritmetično sredino grupnih sredin. Kot pondere vzamemo posameznim grupnim sredinam odgovarjajoče število enot grupe. V simbolih moremo pisati:

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \ldots + n_r\bar{x}_r}{n_1 + n_2 + n_3 + \ldots + n_r} \qquad \bar{x} = \frac{1}{n}\sum_k n_k\,\bar{x}_k \qquad (8.13)$$

P Upravičenost tega postopka je razvidna iz primera:

Zaradi boljšega pregleda vzemimo manjšo statistično maso in sicer število delavcev za 2o podjetij. Podjetja, ki so iz treh različnih strok A,B,C, razdelimo po strokah. Individualni podatki so dani v spodnji razpredelnici.

Tab. 8.12

Stroka	Štev. podj.	Število delavcev v podjetjih	Skup.št. delavcev	Aritmetična sredina
(1)	(2)	(3)	(4)=Σ(3)	(5)=(4):(2)
A	8	35, 47, 52, 39, 58, 62, 44, 53	390	48,7
B	7	89,11€, 36, 92, 94,114, 85	626	89,4
C	5	32€,287,536,332,421	1902	36o,4
Skupno	2o		2918	145,9

Skupno aritmetično sredino za vseh 20 podjetij dobimo, ako skupno število vseh delavcev 2918, delimo z 20, to je s skupnim številom vseh podjetij. Skupno število vseh delavcev je enako vsoti skupnega števila delavcev po posameznih strokah 2918 = 390 + 626 + 1902, skupno število vseh podjetij pa vsoti števila podjetij po posameznih strokah 20 = 8 + 7 + 5. Vsoto števila delavcev moremo zaradi lastnosti aritmetične sredine pisati kot produkt števila enot in aritmetične sredine. 390 = 8.48,7, 626 = 7.89,4, 1902=5.360,4.

Skupno aritmetično sredino moremo torej pisati:

$$\bar{x} = 145,9 = \frac{2918}{20} = \frac{390 + 626 + 1902}{8 + 7 + 5} = \frac{8.48,7 + 7.89,4 + 5.360,4}{8 + 7 + 5}$$

kar je tehtana aritmetična sredina grupnih sredin s številom delavcev kot ponderi. V simbolih moremo pisati:

$$\bar{x} = \frac{n_A \bar{x}_A + n_B \bar{x}_B + n_C \bar{x}_C}{n_A + n_B + n_C} \qquad (8.14)$$

kar je poseben primer splošnega obrazca 8.13, ako je število grup enako tri.

D Dokaz: Splošen dokaz moremo razviti iz lastnosti aritmetične sredine, da je vsota vrednosti znakov enaka produktu števila enot in aritmetične sredine.

$$\sum X = n\bar{x}$$

Ta lastnost velja tako za skupno statistično maso, kot za grupe. Za posamezne grupe imamo:

$$\sum' x = n_1 \bar{x}_1 \quad \sum'' x = n_2 \bar{x}_2 \quad \sum''' x = n_3 \bar{x}_3 \quad \sum^{(r)} x = n_r \bar{x}_r \text{ skupno pa } \sum x = n\bar{x}$$

Ker je vsota delnih vsot enaka skupni vsoti vseh enot, je :

$$\sum x = \sum' x + \sum'' x + \sum''' x + \ldots + \sum^{(r)} x \quad \text{iz česar sledi :}$$

$$n\bar{x} = n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3 + \ldots + n_{(r)} \bar{x}_{(r)} \quad \text{ali, ker je :}$$

$$n = n_1 + n_2 + n_3 + \ldots n_r$$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3 + \ldots + n_r \bar{x}_r}{n_1 + n_2 + n_3 + \ldots + n_r} \qquad \bar{x} = \frac{\sum n_k \bar{x}_k}{\sum n_k} \qquad (8.15)$$

P Primer: Izračunaj iz grupnih sredin primera iz 81.62 povprečno število krav na gospodarstvo za vsa gospodarstva do 10 ha skupne površine.

$n_1 = 607$	$\bar{x}_1 = 0,970$	$n_1\bar{x}_1 = 588,79$
$n_2 = 551$	$\bar{x}_2 = 2,143$	$n_2\bar{x}_2 = 1180,79$
$n_3 = 726$	$\bar{x}_3 = 2,949$	$n_3\bar{x}_3 = 2140,97$
$n = 1884$		$x = 3910,55$

Zadnji stolpec pomeni število krav v posameznih skupinah. Morala bi biti torej cela števila. Razlika nastopi zaradi tega, ker se povprečja izračunana le na tri decimalke. Vendar ta razlika ne vpliva na skupno sredino, ker v skupnem razliki n_i niti ena glava živine.

Skupna aritmetična sredina se torej glasi:

$$\bar{x} = \frac{3910,55}{1884} = 2,074 \text{ krav na gospodarstvo}$$

V gospodarstvih do 10 ha skupne površine je povprečno 2,074 krav na eno gospodarstvo.

81.8 Odvisnost skupne ali sumarne aritmetične sredine od strukture po grupah.

.81 Hiba sumarnih aritmetičnih sredin je poleg tega, da ne predstavljajo tipične sredine, tudi v tem, da ni odvisna samo od vrednosti grupnih sredin, temveč tudi od strukture ponderov, oziroma grup. To sledi iz samega obrazca:

$$\bar{x} = \frac{1}{n}\sum n_k \bar{x}_k = \sum \frac{n_k}{n} \bar{x}_k \qquad (8.16)$$

kjer so $\frac{n_k}{n}$ relativne frekvence, oziroma strukturni deleži za posamezne grupe. Iz obrazca je razvidno, da je sumarna sredina sestavljena iz produktov strukturnih deležev $\frac{n_k}{n}$ in grupnih aritmetičnih sredin. Zaradi tega, ker zavisi sumarna sredina od strukture grup in grupnih sredin, nima polne analitične vrednosti. Ako primerjamo med seboj sumarni aritmetični sredini dveh statističnih mas, ne vemo, ali je razlika sredin rezultat spremembe grupnih sredin ali razlike v strukturi mase. Razlika sumarnih sredin je skupen rezultat obeh učinkov, ki sta raznosmerna. Kot bomo videli na naslednjem primeru, more sumarna sredina izpasti večja za drugo maso, čeprav do grupne sredine druge mase vse manjše od odgovarjajočih grupnih sredin prve mase. Ta rezultat izgleda na prvi pogled protisloven, vendar je možen, ako sta strukturi obeh primerjanih mas med seboj zelo različni.

P Proučujmo povprečno število članov na eno gospodinjstvo po socialnih skupinah za okraja Trbovlje in Tolmin. Podatki so vzeti iz publikacije o gospodinjstvih popisa prebivalstva v LR Sloveniji z dne 15. marca 1948. Ako imamo za vsak okraj dano povprečno število članov na eno gospodinjstvo in procentni delež gospodinjstev po socialnih skupinah, moremo izračunati sumarno povprečno število članov na eno gospodinjstvo kot tehtano aritmetično sredino po naslednjem obrazcu:

$$\bar{x} = \frac{\sum n_k \cdot 100\, \bar{x}_k}{100} \qquad (8.17)$$

Tab. 8.13

Socialna skupina	Tolmin			Trbovlje		
	%	$\bar{x}_k$	(2).(3)	%	$\bar{x}_k$	(5).(6)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Delavci in učenci	19,8	3,0	59,40	53,9	3,4	183,26
Akt. uslužbenci	12,0	2,0	24,00	11,5	3,0	34,50
Kmetje	56,6	4,3	243,38	12,8	5,2	66,56
Obrtniki-trgovci	3,5	3,4	11,90	3,3	3,9	12,87
Upokojenci in vzdrž.	7,2	2,1	15,12	18,1	2,8	50,68
Ostali	0,9	1,9	1,71	0,4	2,7	1,08
Skupaj	100,0	-	355,51	100,0	-	348,95
Povprečje	-	3,56	-	-	3,49	-

V koloni (2) in (5) je vnešena struktura števila gospodinjstev po socialni skupinah, v kolonah (3) in (6) pa povprečna števila članov na eno gospodinjstvo.

Čeprav je v okraju Trbovlje povprečno število članov v vseh socialnih skupinah znatno višje kot v oktaju Tolmin, (primerjaj stolpec (3) in (6)), je zaradi razlik v sovialni strukturi prebivalstva sumarno povprečno število članov na gosppdinjstvo za okraj Tolmin večje, kot pa v okraju Trbovlje. Do tega protislovnega rezultata pridemo zaradi tega, ker se v okraju Tolmin najmočnejša socialna skupina kmetje z 56,6 % in relativno visokim povprečjem 4,3, v okraju Trbovlje pa delavci z 53,9 % z relativno nizkim povprečjem 3,4 članov na eno gospodinjstvo. Ta primer je pokazal, do kakšnih pogrešnih rezultatov more dovesti uporaba sumarnih srednjih vrednosti.

.82 Enotni -- standardni ponderi. Opisane težave nastopajo v vseh primerih uporabe sumarnih sredin. V demografiji splošni pokazatelji rodnosti, smrtnosti itd., odvise od smrtnosti, oziroma rodnosti po letih in strukturi prebivalstva po starosti ali socialnih skupinah. Sumarna povprečna plača je odvisna od povprečnih plač po kvalifikaciji in stroki in strukturi delavstva po kvalifikaciji in strokah.

Primerjava sumarnih povprečij nima polne analitične vrednosti zaradi istočasne spremembe strukture in grupnih povprečij. To nas dovede do misli, da bi bila primerljivost sumarnih povprečij večja, ako bi pri vseh primerjalnih sumarnih povprečijh uporabljali enotno strukturo. To sumarno povprečje sicer ni dejansko, vendar dobro služi svojemu namenu -- primerjavi sumarnih povprečij. V tem primeru je sprememba med dvema sumarnima sredinama samo rezultat spremembe grupnih sredin, ker je struktura ostala enaka. Enotna struktura pomeni enotne pondere. Seveda obstoji pri vsakem danem primeru vprašanje, katero strukturo naj vzamemo kot enotno strukturo. Včasih vzamemo povprečno strukturo vseh proučevanih mas, včasih pa strukture, izdelane na osnovi teoretičnih vidikov. Kadar skušamo doseči medkarodno primerljivost sumarnih demografskih koeficientov, uporabljamo enotno starostno strukturo, ki jo imenujemo standardno strukturo, na osnovi nje izračunane koeficiente pa standardne koeficiente.

Zelo pogosto pa uporabljamo kot enotne pondere povprečne strukture.

Vzemimo v našem primeru kot enotno strukturo za izračunavanje sumarne sredine povprečno strukturo gospodinjstev v LRS.

Tab. 8.14

Socialna skupina	Struktura gosp. v LRS	Povprečno članov		(2).(3)	(2).(4)
		Tolmin	Trbovlje		
(1)	(2)	(3)	(4)	(5)	(6)
Delavci in nameščenci	32,8	3,0	3,4	98,40	101,52
Aktivni uslužbenci	15,0	2,0	3,0	30,00	45,00
Kmetje	37,1	4,3	5,2	159,53	192,92
Obrtniki-trgovci	4,5	3,4	3,9	15,30	17,55
Upokojenci-vzdrž.	9,8	2,1	2,8	20,58	27,44
Ostali	0,8	1,9	2,7	1,52	2,16
S k u p a j	100,0	-	-	325,33	386,59
Povprečno	-	3,25	3,87	-	-

Sumarno povprečje je vsota stolpca (5) deljena s 100, oziroma stolpca (6), deljena s 100.

Sumarni sredini, izračunani na osnovi enotne, povprečne strukture, pokažeta pravilno večje število članov na eno gospodinjstvo v okraju Trbovlje kot rezultat samo spremembe grupnih sredin. Vendar moramo opozoriti, da imata gornji sredini in z njo vse na enak način izračunane sredine samo analitičen značaj in ne predstavljata dejanskega povprečka. Zaradi tega zanje tudi ne velja, da je produkt povprečnega števila članov na eno gospodinjstvo in števila gospodinjstev enak skupnemu številu vseh oseb v vseh gospodinjstvih okraja. Ta stavek velja le v primeru, ako vzamemo kot pondet pravo strukturo okraja.

81.9 Aritmetična sredina iz relativnih števil.

.91 Sumarno relativno število, izračunano iz absolutnih števil. Vzemimo kot enostaven primer strukturni delež moškega prebivalstva po popisu prebivalstva 15. marca 1948 po republikah. Za vsako republiko posebej moremo izračunati delež moškega prebivalstva na ta način, da delimo število moškega prebivalstva v republiki s številom celotnega moškega prebivalstva v republiki. Enakoizračunamo tudi strukturni delež moškega prebivalstva za vso državo. Število moškega prebivalstva za vso državo delimo z številom prebivalstva za vso državo. Ako skupnih podatkov za vso državo nimamo, jih dobimo kot vsoto podatkov za posamezne republike. Odstotek moškega prebivalstva za vso državo imenujemo povprečen strukturni delež moškega prebivalstva v FLRJ.

Tab.8.15

Ljudska republika	Število prebivalstva po popisu 15. marca 1948 v 000		
	Skupno S	Moških M	% moških D
(1)	(2)	(3)	(4)
1. Srbija	6528	3171	0,485
2. Hrvatska	3757	1759	0,468
3. Slovenija	1392	653	0,469
4. B i H	2565	1237	0,482
5. Makedonija	1153	584	0,506
6. Črna gora	377	178	0,472
Skupno v FLRJ	15772	7582	0,481

$$\frac{3171}{6528} = 0{,}485 \qquad \frac{1759}{3757} = 0{,}468 \quad \text{itd.}$$

Po potrebi moremo strukturne deleže pomnožiti z 100, da dobimo strukturne deleže izražene v odstotkih ($0{,}485 \cdot 100 = 48{,}5\%$). V simbolih moremo pisati:

$$\frac{M_1}{S_1} = D_1 \qquad \frac{M_2}{S_2} = D_2 \quad \text{itd.}$$

M pomeni število moškega prebivalstva, S skupno število prebivalstva, D pa delež moških od skupnega števila prebivalcev. Številke 1, 2, 3 itd. pomenijo za katero republiko gre. Simboli brez številke naj pomenijo vrednosti za FLRJ. V tem primeru velja:

$$D = \frac{M}{S} = \frac{M_1 + M_2 + M_3 + M_4 + M_5 + M_6}{S_1 + S_2 + S_3 + S_4 + S_5 + S_6} = \frac{7582}{15772} = 0,481$$

Sumarni procent moških v FLRJ dobimo, ako vsoto moških za vse republike delimo z vsoto skupnega prebivalstva za vse republike.

Iz gornjega primera moremo zaključiti splošno pravilo za izračunavanje sumarne, povprečne relativne vrednosti R. Ako sta znani absolutni količini Š in I iz katerih izračunamo relativno število

$$R = \frac{\check{S}}{I}$$

za posamezne grupe. Zaradi boljše preglednosti smo zaznamovali količino, ki pri izračunavanju relativnega števila stoji v števcu z Š, količino, ki stoji v imenovalcu, pa z I.

Sumarno, povprečno relativno število moremo izračunati v tem primeru brez poznavanja grupnih, relativnih števil na ta način, da tvorimo iz grupnih vrednosti vsoto za prvo in drugo količino in tvorimo kvocient obeh vsot na enak način, kot bi izračunavali grupna relativna števila.

$$R = \frac{\check{S}}{I} = \frac{\check{S}_1 + \check{S}_2 + \check{S}_3 + \ldots \check{S}_r}{I_1 + I_2 + I_3 + \ldots I_r} = \frac{\Sigma \check{S}_k}{\Sigma I_k} \qquad (8.18)$$

Ta primer nastopa v praksi zelo pogosto. Iz površine njiv in skupne površine po krajevnih ljudskih odborih moremo izračunati odstotek njiv v vsej Sloveniji tako, da seštejemo vse njive in vse skupne površine in iz teh vsot pričnemo odstotek njiv. Iz števila učencev in števila učiteljev po okrajih moremo izračunati povprečno število učencev na enega učitelja, po okrajih, iz vsote števila učiteljev za celo republiko in iz vsote števila učencev za celo republiko pa povprečje, ne da bi poznali grupna relativna števila.

.92 Sumarno relativno število, izračunamo iz relativnih števil.

Vprašanje je, ali moremo izračunati povprečen strukturni delež moških v primeru, ako imamo dano za vsako republiko skupno število prebivalcev in strukturni delež moških.

Jasno je, da se povprečni strukturni delež izračunava vedno iz obrazca:

$$D = \frac{M}{S} = \frac{M_1 + M_2 + M_3 + \ldots + M_6}{S_1 + S_2 + S_3 + \ldots + S_6} \qquad (8.19)$$

Vprašanje, ki ga je treba rešiti, je torej samo, ali je možno iz števila prebivalstva in strukturnega deleža za neko republiko izračunati število moških oseb v tej republiki. Odgovor je pozitiven. Ako pomnožimo število prebivalstva neke republike z deležem moškega prebivalstva za to republiko, dobimo število moškega prebivalstva za to republiko. To sledi iz zveze:

$$D_1 = \frac{M_1}{S_1} \quad . \tag{8.20}$$

Ako pomnožimo to enačbo na levi in desni z S_1, dobimo

$$S_1 D_1 = M_1 \quad . \tag{8.21}$$

Ta zveza velja za vse republike. Veljavnost tega pravila moremo preiskusiti tudi na primeru. Alo na primer pogledamo LR Slovenijo, za katero je bilo število prebivalstva po popisu 15. marca 1948 1392 tisoč, strukturni delež pa o,469, vidimo, da je

$$1392 \cdot o{,}469 = 653$$

kar se strinja s podatkom v razpredelnici 8.15 Števec v obrazcu nadomestimo zaradi tega z vsoto produktov števila prebivalcev s strukturnimi deleži. Obrazec za izračunavanje povprečnega strukturnega deleža moških v FLRJ se zaradi tega glasi:

$$D = \frac{M}{S} = \frac{S_1 D_1 + S_2 D_2 + S_3 D_3 + S_4 D_4 + S_5 D_5 + S_6 D_6}{S_1 + S_2 + S_3 + S_4 + S_5 + S_6} \tag{8.22}$$

V tem obrazcu vidimo, da nastopajo samo količine, ki jih poznamo, število prebivalstva po republikah in strukturni deleži po republikah. Na drugi strani pa vidimo, da je ta obrazec enak obrazcu izračunavanja tehtane aritmetične sredine. Sumarni povprečni strukturni delež števila prebivalstva za vso FLRJ je torej enak tehtani aritmetični sredini republiških strukturnih deležev. Kot pondere vzamemmo število prebivalstva po republikah.

Dejansko izračunavanje izvršimo v naslednjih fazah:

1./ Množimo število prebivalstva (kolona 2) z strukturnimi deleži (kolona 3), dobimo kolono (4)

2./ Seštejemo število prebivalstva po republikah (kolono 2 in produkte v koloni 4).

3./ Vsoto kolone 4 delimo z vsoto kolone 2 in dobimo povprečni strukturni delež.

Tab. 8.16

Ljudska republika	Število prebivalstva S_k	Strukturni delež moških D_k	Izračunane količine $M_k = S_k D_k$
(1)	(2)	(3)	(4)
1. Srbija	6528	o,485	3171
2. Hrvatska	3757	o,468	1758
3. Slovenije	1392	o,469	653
4. B i H	2565	o,482	1237
5. Makedonija	1153	o,5o6	584
6. Črna gora	377	o,472	178
Skupno FLRJ	15772	..	7582

$$D = \frac{\sum S_k D_k}{\sum S_k} = \frac{7582}{15772} = o,481$$

Zgornji postopek moremo posplošiti in izdelati splošno pravilo. Ako so dana grupna relativna števila in grupne vrednosti količin, ki nastopajo pri relativnem številu v imenovalcu, izračunamo povprečno relativno število kot tehtano aritmetično sredino grupnih relativnih števil. Kot pondere pa vzamemo dane grupne vrednosti količin, ki nastopajo pri izračunavanju relativnega števila v imenovalcu. Ako ohranimo dosedanje simbole iz 81.91 R, Š in I, je R v tem primeru enak:

$$R = \frac{I_1 R_1 + I_2 R_2 + I_3 R_3 + \dots + I_r R_r}{I_1 + I_2 + I_3 + \dots I_r} = \frac{\sum_k I_k R_k}{\sum_k I_k} = A_R \qquad (8.23)$$

81.10 Lastnosti aritmetične sredine.

a) Aritmetična sredina količin, ki med seboj niso enake, ki torej variirajo, je večja od najmanjše in manjša od največje vrednosti količin, iz katerih smo izračunali aritmetično sredino. V simbolih se ta lastnost izraža v obliki neenačbe:

$$x_{min} < \bar{x} < x_{max} \qquad (8.24)$$

b) Sumarna aritmetična sredina, izračunana iz grupnih sredin ali grupnih relativnih števil je večja od najmanjše grupne sredine in manjša od največje grupne sredine. V simbolih moremo napisati:

$$\bar{x}_{min} < \bar{x} < \bar{x}_{max} \qquad (8.25$$

$\bar{x}_{min}$ pomeni najmanjšo grupno sredino, $\bar{x}_{max}$ pa največjo grupno sredino.

c) Vsota individualnih odklonov od aritmetične sredine je enaka 0. V simbolih moremo to lastnost napisati:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x}) = 0 \qquad (8.26)$$

ali krajše

$$\sum_i (x_i - \bar{x}) = 0 \qquad (8.27)$$

To lastnost smo opazili že na primeru v 81.11. Jasno je, da so nekateri odkloni pri tem pozitivni, nekateri pa negativni.

d) Iz individualnih vrednosti moremo izračunati vsoto kvadratov odklonov od neke poljubne vrednosti.

Vsota kvadratov odklonov od neke vrednosti je najmanjša, ako kot M vzamemo aritmetično sredino $\bar{X}$. V simbolih se ta lastnost glasi:

$$(x_1 - M)^2 + (x_2 - M)^2 + (x_3 - M)^2 + \ldots + (x_n - M)^2 = \text{Min ako } M = \bar{X} \qquad (8.28)$$

ali krajše:

$$\sum_i (x_i - M)^2 = \text{Min ako je } M = \bar{x} \qquad (8.29)$$

D Dokazi: Najprej dokažimo, da je vsota odklonov od aritmetične sredine enaka 0.

Ako levo stran enačbe, ki izraža to lastnost

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x})$$

preuredimo tako, da nesemo vse individualne vrednosti skupaj in vse aritmetične sredine skupaj, dobimo:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x}) =$$

$$= (x_1 + x_2 + x_3 + \ldots + x_n) - (\bar{x} + \bar{x} + \bar{x} + \ldots + \bar{x}) =$$

$$= (x_1 + x_2 + x_3 + \ldots + x_n) - n\bar{x}$$

Ker je po definiciji vsota vseh vrednosti enaka produktu števila enot in aritmetične sredine, je:

$$(x_1 + x_2 + x_3 + \ldots + x_n) - n\bar{x} = n\bar{x} - n\bar{x} = 0.$$

S tem je stavek 4. c $\sum_i (x_i - \bar{x}) = 0$ dokazan.

Ako velja zgornja lastnost, mora biti v izrazu

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots + (x_n - \bar{x})$$

nekaj diferenc pozitivnih, nekaj pa negativnih, ali pa vse enake 0. Ako bi bile vse diference $(x_1 - \bar{x})$ enake 0, pomeni, da so vse vrednosti med seboj enake, torej enake $\bar{x}$. Ta primer izključimo. Ako vrednosti med seboj niso enake, morajo biti nekatere diference pozitivne nekatere negativne. Z drugimi besedami to pomeni, da so nekatere vrednosti x_i večje, druge pa manjše od aritmetične sredine.

Aritmetična sredina torej leži med najmanjšo in največjo vrednostjo znaka.

Enako lastnost za sumarno sredino iz grupnih sredin, moremo dokazati na sledeč način:

$$\text{Iz} \quad \frac{1}{n} \sum n_k \bar{x}_k = \bar{x} \quad \text{moremo zaključiti}$$

$$n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \ldots + n_r\bar{x}_r - n\bar{x} = 0$$

$$n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \ldots + n_r\bar{x}_r - (n_1+n_2+n_3 + \ldots + n_r)\bar{x} = 0$$

$$n_1(\bar{x}_1 - \bar{x}) + n_2(x_2 - \bar{x}) + n_3(\bar{x}_3 - x) + \ldots + n_r(x_r - \bar{x}) = 0$$

To je k tretji lastnosti podobna lastnost za grupne sredine. V tem primeru moramo odklone grupnih sredin od sumarne sredine tehtati s številom enot v posameznih grupah. Leva stran enačbe je 0, v primeru, da so vse diference $(\bar{x}_k - \bar{x})$ enake 0, ali pa, da so nekatere pozitivne, nekatere pa negativne. Prvi primer enakih grupnih sredin izpustimo, ker ni zanimiv. Iz pogoja, da morajo biti nekatere diference pozitivne, nekatere negativne, pa sledi, da morajo biti nekatere grupne sredine večje, nekatere pa manjše od sumarne sredine.

Z drugimi besedami moremo izraziti to lastnost s tem, da leži sumarna sredina mad najmanjšo in največjo grupno sredino.

Ti stavki omogočajo kontrolo izračunavanja tako navadnih kot tehtanih in sumarnih sredin.

Ako pogledamo ali veljajo te lastnosti za naše primere vidimo, da je povsod izpolnjena.

Tab. 8.17

Številka odstavka	Primer	Enota mere	Individualna ali grupna vrednost		Povprečje
			najmanjša	največja	
81.11	Plača delavcev	din	4336	4628	4488
81.22	Vrednost proizv.	ooo din	1282,4	1292,4	1287,65
81.31	Število otrok na eno mater	otroci	1	3	1,67
81.3	Stanoval	stanovalci	1	11	5,o
81.4	Cena	din	7	11	8,3
81,5	Teža zrn	gr	2o	6o	37,64
81.7	Število delavcev v podjetjih	delavci	35	62	48,7
		delavci	36	116	89,4
		delavci	287	536	360,4
		povpr.štev. delavcev	48,7	36o,4	145,9
81.62	Število krav na gospodarstvo po vel.skupinah	število krav	0	6	o,97
		"	0	6	2,14
		"	0	9	2,95
		Povpreč. št.krav	o,97	2,95	2,o7
81.81	Člani na gosp. v okraju Tolmin	Povpreč.št. članov na gospodinj.	1,9	4,3	3,56
	Trbovlje		2,7	5,2	3,49
81.91	Strukturni delež moških po republikah	Strukturni delež moš.	46,8	5o,6	48,1

Četrta lastnost, ki ima neposredno zvezo z eno izmed najvažnejših mer variacije, je tudi zelo važna. Dokaz moremo izpeljati iz poznavanja tretje lastnosti.

Ker velja za vsak x $\quad x - M = x - \bar{x} + \bar{x} - M = (x - \bar{x}) + (\bar{x} - M)$

velja za vsak x tudi: $(x - M)^2 = \left[(x - \bar{x}) + (\bar{x} - M)\right]^2 =$

$$(x - \bar{x})^2 + (x - \bar{x})\,(\bar{x} - M) + (\bar{x} - M)^2$$

ali za posamezne:

$$(x_1 - M)^2 = (x_1 - \bar{x})^2 + 2(x_1 - \bar{x})\,(\bar{x} - M) + (\bar{x} - M)^2$$

$$(x_2 - M)^2 = (x_2 - \bar{x})^2 + 2(x_2 - \bar{x})\,(\bar{x} - M) + (\bar{x} - M)^2$$

$$(x_3 - M)^2 = (x_3 - \bar{x})^2 + 2(x_3 - \bar{x})\,(\bar{x} - M) + (\bar{x} - M)^2$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$(x_n - M)^2 = (x_n - \bar{x})^2 + 2(x_n - \bar{x})\,(\bar{x} - M) + (\bar{x} - M)^2$$

Ako vse te enačbe seštejemo dobimo:

$$\Sigma_i (x_i - M)^2 = \Sigma (x_i - \bar{x})^2 + 2(\bar{x} - M) \cdot \Sigma_i (x_i - \bar{x}) + n(\bar{x} - M)^2 \qquad (8.30)$$

ker je $\Sigma_i (x_i - \bar{x}) = 0$, se obrazec še poenostavi:

$$\Sigma (x_i - M)^2 = \Sigma (x_i - \bar{x})^2 + n(\bar{x} - M)^2 \qquad (9.31)$$

Iz te enačbe vidimo, da je vsota kvadratov odklonov od M sestavljena iz dveh delov. Prvi del je od M neodvisen, drugi del, ki je produkt dveh pozitivnih količin, števila enot in kvadrata razlike med M in aritmetično sredino, je del, ki ga moremo spreminjati. Vsota kvadratov odklonov od M bo tem manjša, čim manjši bo ta del. Ker pa je ta del pozitivna količina, je njena najmanjša možna vrednost nič. Nič pa je ta del takrat, ako je:

$x - M = 0$ ali $M = \bar{x}$. S tem je stavek dokazan.

82. HARMONIČNA SREDINA

82.1. Izpeljava iz primera.

Sumarni strukturni delež moških v FLRJ smo v poglavju 81.9 izračunali na dva različna načina. Ako imamo dane podatke o številu moških in številu prebivalstva po republikah, je sumarni strukturni delež enak kvocientu vsot moških in vsega prebivalstva za vse republike. V primeru, da imamo dano število prebivalstva in strukturne deleže po republikah, se je izkazalo, da je potrebno izračunati sumarni povprečni strukturni delež za vso državo kot tehtano aritmetično sredino strukturnih deležev po republikah, pri čemer vzamemmo število prebivalstva kot pondere. Dosedaj smo izračunali sumarni strukturni delež, ako smo izmed treh količin, ki nastopajo v našem primeru (S skupno število prebivalstva, M število moških, D strukturni delež) poznali po republikah po dva, in sicer v prvem primeru število prebivalstva in število moških po republikah (S in M) v drugem pa število prebivalstva in strukturni delež (Sin D) po republikah.

Vprašamo se, ali je možno izračunati sumarni povprečni strukturni delež, ako gre za tretjo možno kombinacijo, to je, ako imamo dano po republikah število moških in strukturne deleže (M in D).

Strukturni delež moških za celotno državo je najenostavnejše izračunati na prvi način, ako imamo dane podatke o številu moških M in številu prebivalstva S po republikah. Število moških po republikah imamo v našem primeru dano, nimamo pa danega skupnega števila prebivalstva. Ker obstoja med $S_k M_k$ in D_k zveza

$$D_k = \frac{M_k}{S_k} \tag{8.32}$$

imamo pa dane M in D, moremo izračunati za vsako republiko število prebivalstva, ako delimo število moškega prebivalstva z strukturnim deležem

$$S_k = \frac{M_k}{D_k} \tag{8.33}$$

Na ta način moremo iz danih podatkov o številu moških in strukturnem deležu tudi izračunati sumarni povprečni strukturni delež.

$$D = \frac{M}{S} = \frac{M_1 + M_2 + M_3 + M_4 + M_5 + M_6}{S_1 + S_2 + S_3 + S_4 + S_5 + S_6} = \frac{M_1 + M_2 + M_3 + M_4 + M_5 + M_6}{\frac{M_1}{D_1} + \frac{M_2}{D_2} + \frac{M_3}{D_3} + \frac{M_4}{D_4} + \frac{M_5}{D_5} + \frac{M_6}{D_6}} \tag{8.34}$$

Sum. povpr. struk. delež

Vidimo, da izračunavamo v primeru, da imamo dano število moških in strukturne deleže po republikah, sumarni delež na dosedaj povsem nov način. Tak način izračunavanja povprečnega sumarnega strukturnega deleža imenujemo harmonična sredina.

Upravičenost tega postopka bomo videli tudi na primeru:

Tab. 8.18

Ljudska republika	Dane količine		Izračunana količina
	Število moških $M_k = S_k \cdot D_k$	Strukturni delež D_k	$S_k = \frac{M_k}{D_k}$
(1)	(2)	(3)	(4)=(2):(3)
1. Srbija	3171	o,485	6528
2. Hrvatska	1759	o,468	3757
3. Slovenija	653	o,469	1392
4. B i H	1237	o,482	2565
5. Makedonija	584	o,506	1153
6. Črna gora	178	o,472	377
	7582	-	15772

$$D = \frac{7582}{15772} = o,481$$

Harmonično sredino smo izračunali v naslednjih fazah:

1./ Delili smo število moških (kolona 2) s strukturnim deležem (kolona 3) za vsako republiko. Dobili smo na ta način v koloni 4 število prebivalstva za posamezne republike.

2./ Sešteli smo število moških po republikah (kolona 2) in število prebivalstva po republikah (kolona 4).

3./ Vsoto števila moških smo delili z vsoto števila prebivalstva in dobili povprečen strukturni delež o,481.

82.2 Zgornji postopek moremo posplošiti in izdelati splošno pravilo. Ako so dana grupna relativna števila in grupne vrednosti količin, ki nastopajo pri izračunavanju relativnega števila v števcu, izračunamo sumarno, povprečno relativno število kot tehtano harmonično sredino grupnih relativnih števil. Kot pondere vzamemo dane grupne vrednosti količin, ki nastopajo v števcu relativnih števil.

Ako ohranimo oznake iz poglavja 81.9, kjer je R pomenil relativno število, Š količine, ki nastopajo v števcu, I pa količine, ki nastopajo v imenovalcu, je:

$$R = \frac{\check{S}_1 + \check{S}_2 + \check{S}_3 + \dots + \check{S}_r}{\frac{S_1}{R_1} + \frac{S_2}{R_2} + \frac{S_3}{R_3} + \dots + \frac{S_r}{R_r}} = \frac{\sum_k \check{S}_k}{\sum_k \frac{S_k}{R_k}} = H_R \qquad (8.35)$$

Tehtano harmonično sredino torej izračunamo tako, da vsoto ponderov delimo z vsoto kvocientov ponderov in odgovarjajočih grupnih vrednosti, za katere iščemo povprečje.

Ako so vsi ponderi med seboj enaki, se izračunavanje poenostavi:

$$H_R = \frac{\check{S} + \check{S} + \check{S} + \ldots + \check{S}}{\frac{\check{S}}{R_1} + \frac{\check{S}}{R_2} + \frac{\check{S}}{R_3} + \ldots + \frac{\check{S}}{R_n}} = \frac{n}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots + \frac{1}{R_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{R_i}} \qquad (8.36)$$

Ta način izračunavanja se imenuje navadna harmonična sredina.

P Avtomobil je na 640 km dolgi progi prevozil 100 km s hitrostjo 50 km na uro, 240 km s hitrostjo 80 km na uro in 300 km s hitrostjo 60 km na uro. Kolika je povprečna hitrost?

Povprečna hitrost je enaka kvocientu celotne poti in časa, v katerem je bila ta pot prevožena.

Neposredno je dana pot, ni pa dan čas, katerega moramo za posamezne odseke poti izračunati kot kvociente poti in hitrosti. Povprečna hitrost je torej enaka:

$$H = \frac{P_1 + P_2 + P_3}{\frac{P_1}{H_1} + \frac{P_2}{H_2} + \frac{P_3}{H_3}} = \frac{100 + 240 + 300}{\frac{100}{50} + \frac{240}{80} + \frac{300}{60}} = \frac{640}{2 + 3 + 5} = \frac{640}{10} = 64 \text{ km/uro}$$

82.3 V vsakem danem primeru je treba po vrsti podatkov določiti ali je pravilno za izračunavanje povprečje vzeti aritmetično ali harmonično sredino. Zaradi preglednosti podajamo obravnavane tri možnosti v oznakah v razpredelnici.

Ako je $R_k = \frac{\check{S}_k}{I_k}$, potem izračunavamo povprečno relativno število R.

Ako so dani grupni podatki izračunavamo povprečno relativno število R	$\check{S}_k, I_k$	I_k, R_k	$\check{S}_k, R_k$
Kot	kvocient vsot	tehtano sredino	
		aritmetično	harmonično
Po obrazcu	$R = \frac{\sum \check{S}_k}{\sum I_k}$	$R = \frac{\sum I_k R_k}{\sum I_k}$	$R = \frac{\sum \check{S}_k}{\sum \frac{\check{S}_k}{R_k}}$

(8.37)

p Število učencev na enega učitelja pokaže obremenjenost učiteljev. Ako razpolagamo s temi koeficienti po okrajih in imamo poleg tega dano še število učencev po okrajih, izračunamo povprečni koeficient s harmonično sredino.

Kvocient med blagovnim prometom in zalogami v trgovski statistiki pokaže obrtljivost žalog. Ako razpolagamo s podatki o blagovnem prometu in zalogah za posdine trgovine, izračunamo povprečno obrtljivost kot kvocient vsote blagovnega prometa in vstote zalog vseh trgovin.

Gostoto prebivalstva dobimo, ako število prebivalstva delimo s površino v km2. Ako imamo dano povprečno gostoto po oblasteh in število prebivalstva po oblasteh, izračunamo gostoto za celo republiko s pomočjo tehtane aritmetične sredine.

83. GEOMETRIČNA SREDNJA VREDNOST

Pri dinamičnih serijah izračunavamo povprečno spremembo pojava, ki ga serija prikazuje, s pomočjo geometrijske sredine.

Geometrijsko sredino izračunamo na ta način, da naj prej vse člene, iz katerih računamo geometrijsko sredino, med seboj pomnožimo, iz tega produkta pa izračunamo koren stopnje, ki je enaka številu členov. V simbolih je postopek naslednji:

$$G = \sqrt[n]{x_1 x_2 x_3 \dots x_n} \qquad (8.38)$$

Smiselnost in pomen te količine bo viden na sami dinamični seriji. Vzemimo, da je bila proizvodnja nekega artikla v nekem podjetju po letih in v prvi petletki naslednja:

Tab. 8.19

Leto	Proizvodnja	
	v tonah	$K_k = Y_k / Y_{k-1}$
1947	$Y_1 = 141$	-
1848	$Y_2 = 160$	1,135
1949	$Y_3 = 192$	1,200
1950	$Y_4 = 220$	1,146
1851	$Y_5 = 276$	1,255

Da bomo lažje zasledovali potek izračunavanja, bomo izražali posamezne količine tudi v simbolih. Proizvodnjo zaznamujemo z Y. Številka ob znaku Y pove, za katero leto gre. Dvig proizvodnje od leta do leta pokaže koeficient dinamike, ki ga izračunamo tako, da delimo proizvodnjo v nekem letu s proizvodnjo v preteklem letu.

$$K_1 = \frac{Y_2}{Y_1} = \frac{160}{141} = 1{,}135 \qquad K_2 = \frac{Y_3}{Y_2} = \frac{192}{160} = 1{,}200 \text{ itd.}$$

Koeficient dinamike pokaže, kolikokrat se je proizvodnja od leta do leta povečala. Koeficient dinamike 1,135 pomeni z drugimi besedami, da se je proizvodnja od leta 1947 do leta 1948 povečala za 13,5 °/°.

Iz narave koeficientov dinamike je razvidno, da dobimo vrednost proizvodnje za določeno leto, ako začetno proizvodnjo (leto 1946) pomnožimo z odgovarjajočim številom koeficientov dinamike. Proizvodnje v letu 1947 dobimo po tem pravilu, ako proizvodnjo v letu 1946 pomnožimo s K_1, K_2, K_3 in K_4.

$$Y_1 \cdot K_1 \cdot K_2 \cdot K_3 \cdot K_4 = Y_1 \cdot \frac{Y_2}{Y_1} \cdot \frac{Y_3}{Y_2} \cdot \frac{Y_4}{Y_3} \cdot \frac{Y_5}{Y_4} = Y_5$$

$$141 \cdot 1{,}135 \cdot 1{,}200 \cdot 1{,}145 \cdot 1{,}225 = 276$$

Vprašanje, ki ga moremo rešiti, je naslednje: S kakšnim enotnim koeficientom dinamike bi se morala spreminjati proizvodnja tekom petih let, da bi v letu 1951 dosegli isto vrednost proizvodnje 276 ton, kot smo jo dosegli dejansko. Pri tem izhajamo iz iste osnovne proizvodnje za leto 1947.

Zaradi lastnosti, ki je omenjena že zgoraj, bi v tem primeru dobili iz osnovne vrednosti $Y_1 = 141$ t, vrednost proizvodnje $Y_5 = 276$, ako bi vrednost Y_1 pomnožili s pripadajočim številom koeficientov dinamike. Ker je ta stalen, dobimo:

$$Y_5 = Y_1 \; K \cdot K \cdot K \cdot K = Y_1 \, K^4$$

Ker smo vrednost proizvodnje Y_5 mogli izračunati tudi iz danih koeficientov dinamike, moremo postaviti:

$$Y_5 = Y_1 \cdot K_1 \, K_2 \, K_3 \, K_4 = Y_1 \cdot K^4$$

$$276 = 141 \cdot 1{,}135 \cdot 1{,}200 \cdot 1{,}145 \cdot 1{,}255 = 141 \cdot K^4$$

Iz teh enačb moremo izračunati povprečen koeficient dinamike na dva načina:

a) Ako imamo dane koeficiente dinamike, je:

$$Y_1 \cdot K^4 = Y_1 K_1 K_2 K_3 K_4 , \qquad K^4 = K_1 \cdot K_2 \cdot K_3 \cdot K_4$$

Povprečen koeficient dinamike pa je četrti koren iz produkta dejanskih koeficientov dinamike.

$$K = \sqrt[4]{K_1 K_2 K_3 K_4} = \sqrt[4]{1{,}135 \cdot 1{,}200 \cdot 1{,}145 \cdot 1{,}255} = 1{,}183$$

Vidimo, da je povprečen koeficient dinamike enak geometrijski sredini dejanskih koeficientov dinamike.

b) Povprečen koeficient dinamike moremo izračunati tudi, ako imamo dano začetno in končno vrednost proizvodnje, ker je:

$$Y_5 = Y_1 K^4 \qquad K^4 = \frac{Y_5}{Y_1} \qquad K = \sqrt[4]{\frac{Y_5}{Y_1}} = \sqrt[4]{\frac{276}{141}} = 1{,}183$$

Rezultat je v obeh primerih enak.

Ako pomen povprečne spremembe tolmačimo s povprečnim tempom, moremo reči: Da bi dosegli enak končni nivo, kot smo ga dejansko, bi se morala proizvodnja vsako leto dvigniti za 18,3 %.

Izračunavanje povprečne spremembe je tehnično zvezano z razmeroma zamotanim računananjem, množenjem in korenjenjem. Z uporabo logaritmov moremo računanje znatno olajšati:

Ker je : $K = \sqrt[4]{K_1 K_2 K_3 K_4} = \sqrt[4]{1{,}135 \cdot 1{,}200 \cdot 1{,}146 \cdot 1{,}255}$:

je $\log K = \frac{1}{4} (\log K_1 + \log K_2 + \log K_3 + \log K_4) =$

log 1,135	0,05 500
log 1,200	0,07 918
log 1,146	0,05 918
log 1,254	0,09 830
$\sum \log K_k$	0,29 166
$\log K = \frac{1}{4}\sum \log K_k$	0,07 292
	$K = 1{,}183$

Vidimo, da smo logaritem geometrijske sredine iz $K_1 K_2 K_3 K_4$ izračunali kot aritmetično sredino logaritmov posameznih koeficientov dinamike.

Ako sta dani začetna in končna vrednost proizvodnje y_5 in y_1 moremo povprečen koeficient dinamike izračunati z logaritmi takole:

log 276	2,44 091
- log 141	2,14 922
$\log \frac{y_5}{y_1}$	0,29 169
$\text{Log } K = \frac{1}{4} \cdot \log \frac{y_5}{y_1}$	0.07 292
	K = 1,183

Posplošenje zgornjih postopkov dovede do naslednjih splošnih obrazcev:

1./ Ako imamo dane za neko časovno serijo individualne pokazatelje dinamike, je povprečen koeficient dinamike geometrijska sredina iz individualnih koeficientov dinamike.

$$K = \sqrt[n]{K_1 K_2 K_3 \dots K_n} \qquad (8.39)$$

2./ Ako je dan začetni (Y_1) in končni člen (Y_n) časovne serije, moremo povprečen koeficient dinamike izračunati po splošnem obrazcu:

$$K = \sqrt[n-1]{\frac{Y_n}{Y_1}} \qquad (8.40)$$

3./ Logaritem geometrijske sredine je enak aritmetični sredini logaritmov individualnih vrednosti:

$$\text{Log } K = \frac{1}{n} (\log K_1 + \log K_2 + \log K_3 + \dots + \log K_n = \frac{1}{n} \sum_i \log K_i \qquad (8.41)$$

84. MEDIJANA

84.1 Definicija in določevanje v primeru individualnih vrednosti.

Doseđanje vrste srednjih vrednosti smo iz individualnih vrednosti izračunavali. Poleg teh v statistiki uporabljamo kot srednje vrednosti tudi vrednosti, ki so dane z lego členov. Ena izmed takih količin je mediana.

Mediana nekega znaka je tista vrednost, za katero ima polovica enot manjše, polovica pa večje vrednosti od mediane. Mediana torej razdeli statistično maso v dva dela. Polovica enot ima manjše, polovica pa večje vrednosti od mediane.

P Poizkusimo najti mediano za starost sedmih oseb. Število enot smo vzeli tako majhno zaradi nazornosti. Individualne starosti oseb so naslednje:

21, 28, 24, 19, 23, 25, 27 .

Iz te serije podatkov je vrednost, od katere ima polovica enot manjše, polovica pa večje vrednosti, težko najti. Lažje bomo to dosegli, ako vrednosti uredimo po velikosti.

19, 21, 23, 24, 25, 27, 28 .

Iz te serije vidimo, da je vrednost četrtega člena tista, ki zadošča pogojem mediane, ker leži v sredini po velikosti urejenih enot. Od te vrednosti so tri vrednosti manjše, tri pa večje. Mediana je torej 24 let.

Srednji člen po velikosti urejene serije moremo lahko najti, ako je število enot liho.

P Vzemimo pa, da je število enot sodo. Imamo na primer podatke o starosti osmih oseb. Po velikosti urejena serija je naslednja:

28, 27, 29, 29, 31, 32, 35, 40 .

V tem primeru imamo več vrednosti, ki imajo lastnost, da je polovica dejanskih vrednosti manjših, polovica pa večjih. Ena taka starost je n. pr. 29 let in pol, ali 30 let, ali 30 let in 5 mesecev. Da se izognemo nedvoumnosti vzamemo kot mediano v tem primeru aritmetično sredino dveh členov, ki sta najbližja sredini serije. V našem primeru vzamemo kot mediano sredino med četrtim in petim členom:

$$\frac{29 + 31}{2} = 30 \text{ let}$$

Iz teh dveh enostavnih primerov moremo podati splošno pravilo, po katerem določamo vrednost mediane, ako so dane individualne vrednosti enot.

1./ Člene uredimo po velikosti vrednosti znaka.

2./ a) Ako je število enot n liho število, je mediana vrednost člena z zaporedno številko $\frac{n + 1}{2}$.

b) Ako je število enot n sodo število, je mediana aritmetična sredina vrednosti členov, ki imata zaporedni številki

$$\frac{n}{2} \text{ in } \frac{n}{2} + 1.$$

P Ako imamo skupino 97 vojakov, dobimo mediano višin teh vojakov, ako iz vrste, v kateri smo vojake uredili po velikosti, zmerimo višino devetinštiridesetega vojaka.

$$\frac{97 + 1}{2} = 49$$

84.2 Določanje mediane, ako so vrednosti znaka dane v frekvenčni distribuciji.

Običajno nimamo dane vrednosti znaka neke statistične mase v nizu vrednosti posameznih enot, temveč pregledno grupirane v frekvenčni distribuciji. V tem primeru moremo vrednost mediane določiti, če ne popolnoma točno pa vsaj približno iz frekvenčne distribucije. Način določanja odvisi od značaja osnovnega znaka.

.21 Vzemimo najprej, da je znak serije nezvezen in negrupiran.

P Vzemimo kot primer serijo iz števila stanovanj po številu stanovalcev iz 81.33.

Tab. 8.2c

Število stanovalcev	Število stanovanj	Kumulativa
x_k	n_k	N_k
1	5	5
2	14	19
3	29	48
4	43	91
5	29	120
6	30	150
7	26	176
8	11	187
9	5	192
10	3	195
11	2	197
	197	197

Ker ima serija 197 enot, je mediana vrednost $\frac{197 + 1}{2}$ - tega to je 99 - tega člena. Iz navedene frekvenčne distribucije bi mogli napisati niz individualnih vrednosti in sicer: 1,1,1,1,1,2,2,2. To pa bi bilo prezamudno.

Mediano moremo določiti na lažji način s pomočjo kumulativne serije. Iz kumulativne serije moremo zaključiti, da bi v urejeni vrsti individualnih vrednosti imeli člani z redno številko od 1 do 5 vrednost 1, od 6 do 19 vrednost 2, od 20 : 48 vrednost 3, od 49 do 91 vrednost 4, od 92 do 120 vrednost 5 itd. Stanovanje z zaporedno številko 99 ima torej pet stanovalcev. Me = 5 oseb. V polovici stanovanj stanuje torej 5 in manj, v polovici stanovanj pa 5 in več stanovalcev. Ako ta postopek posplošimo, določimo mediano frekvenčne distribucije nezveznega, negrupiranega znaka na naslednji način:

a) Poiščemo vrednost $\frac{n + 1}{2}$ b) Izračunamo kumulativno serijo N_K

c) V kumulativni seriji poiščemo ona člena za katera je $N_d < \frac{n+1}{2} < N_{d+1}$

d) $x_d = Me$

.22 Za zvezno serijo niso znane vrednosti posameznih enot, temveč vemo samo to, v katerem intervalu se posamezna vrednost nahaja. V tem primeru moremo iz frekvenčne distribucije zaključiti, v katerem razredu se nahaja vrednost mediane, točno lego vrednosti pa moremo samo oceniti. Iz tega je razvidno, da je možno dati temboljšo oceno mediane, čim ožji so razredi.

Postopek bomo izpeljali iz primera in ga posplošili.

P Določiti je treba vrednost mediane velikosti zemljiških gospodarstev v LRS.

Število zemljiških gospodarstev v LRS
po popisu zemljišč 1947

Tab.8.21

	Razred velikosti po skupni površini	Število gospodarstev	
		n_k	N_k
1	brez lastne zemlje	8334	0
2	0,01 - 0,50 ha	37371	8434
3	0,51 - 1,00 ha	17269	45705
4	1,01 - 2,00 ha	24182	62974
5	2,01 - 3,00 ha	18340	87156
6	3,01 - 5,00 ha	26184	105496
7	5,01 - 8,00 ha	24200	131680
8	0,01 - 10,00 ha	11234	155880
9	10,01 - 15,00 ha	19036	167114
10	15,01 - 20,00 ha	10540	186150
11	20,01 - 30,00 ha	8442	196690
12	30,01 - 45,00 ha	3595	208132
13	45,01 -	2572	207727
	Skupno	211299	211299

Vrednost mediane je velikost gospodarstva z redno številko $1/2(211.299 + 1) = 105.650$, ako so gospodarstva urejena po velikosti. Enako kot v prejšnjem primeru si pomagamo s kumulativno serijo. Iz kumulativne serije vidimo, da so v urejeni vrsti gospodarstva od redne številke 1 do 8334 brez lastne zemlje, da imajo gospodarstva od redne številke 8335 - 45705 od 0,005 - 0,505 ha skupne površine, gospodarstva od številke 45706-62974 od 0,505 do 1,005 ha površine, gospodarstva od 62975 - 87156 od 1,005 -2,005 ha, gospodarstva od 87157 - 105496 od 2,005 do 3,005 ha, gospodarstva z redno številko od 105497 - 131680 velikost od 3,005 - 5,005 ha itd.

Zaenkrat torej vemo, da ima gospodarstvo z redno številko 105650 skupne površine od 3,005 ha - 5,005 ha, z drugimi besedami, da je mediana v razredu 3,005 - 5,005 ha. Poleg tega vemo, da ima gospodarstvo z redno številko 105496 skupno površino 3,005 ha, gospodarstvo z redno številko 131680 pa 5,005 ha skupne površine.

Ker ne vemo notranje razdelitve vrednosti v tem razredu, predpostavljamo, da so vrednosti razporejene enakomerno, t.j. da je razlika v velikosti dveh zaporednih gospodarstev stalna. Ker je v tem razredu frekvenca 26.184 gospodarstev, širina razreda pa 2 ha, je razlika med dvema zaporednima gospodarstvima 2,00 : 26184. Ker leži med gospodarstvom z redno številko 105496, ki ima velikost 3,005 in gospodarstvom z redno številko 105650, katerega velikost je vrednost mediane, 105650 - 105496 = 154 enot, moramo vrednosti 3,005 ha dodati še

$$(154) \cdot \frac{2.00}{26184} = 154 \cdot \frac{2,00}{26184} = \frac{308}{26184} = 0,0118 \text{ ha}$$

Vrednost mediane je torej:

$$3,005 + 0,0118 = 3,0168 \text{ ha.}$$

Ako postopek posplošimo, izračunamo mediano za zvezen znak takole:

a) Poiščemo vrednost $\frac{n+1}{2}$

b) Izračunamo kumulativno serijo N_k

c) Poiščemo v kumulativni seriji $N_d < \frac{n+1}{2} < N_{d+1}$

b) Mediano izračunamo po obrazcu:

$$M_e = x_{d,min} + \left(\frac{n+1}{2} - N_d\right) \cdot \frac{\Delta_d}{n_d} \qquad (8.42)$$

Pri tem je n število enot celotne mase, $x_{d\,min}$ - spodnja meja razreda, v katerem leži mediana, n_d frekvenca v razredu, v katerem leži mediana, Δ_d pa širina tega razreda.

84.3 Lastnosti in pomen mediane.

a) Vrednost izraza $\sum_{i=1}^{n} (x_i - M) = \text{Min}$ ako je $M = M_e$ (8.43)

ali z besedami: Vsota absolutnih vrednosti odklonov je najmanjša, ako odklone računamo od mediane. Ta lastnost je zelo zanimiva in važna, ker vidimo, da mediana najboljše predstavlja vrednosti vseh enot, ako vzamemo kot merilo podrobnosti individualnih vrednosti od neke vrednosti absoluten odklon. Dokaz gornje lastnosti ni zamotan, vendar presega okvir našega programa.

b) Vrednost mediane je zelo neobčutljiva napram spremembam vrednosti posameznih členov. Izpremeni se le, ako se spremeni vrednost člena, katerega vrednost je mediana, ali ako izpremene nekatere enote svojo vrednost tako, da preidejo iz ene strani mediane na drugo.

D Dokaz: Ako ostane število enot spodaj ali zgoraj mediane isti, ostane, ne glede na vrednost, $x_{\frac{n+1}{2}}$ ista vrednost.

Mediana ima velik praktičen smisel in jo posebno uporabljamo v primerih, kadar zavzema znak zelo različne vrednosti. V tem primeru izgubi artitmetična sredina prevladujočo važnost, čeprav jo tudi za take primere izračunavamo.

85. MODUS.

85.1 Definicija.

Kot srednjo vrednost, ki je dana z lego vrednosti, uporabljamo tudi vrednost, ki se najpogosteje pojavlja. Imenujemo jo normala, modus, ali najpogostejša vrednost. Modus najlažje ugotovimo iz frekvenčne distribucije. Prav tako kot za mediano je določanje modusa za nezvezen znak različno od določanja za zvezen znak.

85.2 Vzemimo najprej, da je osnovni znak serije nezvezen in negrupiran:.

Tri skupine gospodarstev, ki so določene z velikostjo po površini, so razdeljena po številu krav.

Tab. 8.22
Štev. gospodarstev OBLO Ljubljana

št. krav \ velik.	do 2,o ha	2,o1-5,oo	5,oo-1o,o
0	25o	43	42
1	185	127	69
2	128	182	165
3	33	127	2o8
4	7	54	139
5	3	15	71
6	1	3	21
7			8
8			2
9			1
	6o7	551	726

V skupini do 2,o ha so najpogostejša (25o) gospodarstva, ki so brez krav $M_o = 0$

V skupini od 2,o1 do 5,oo ha so najpogostejša (182) gospodarstva, ki imajo po dve kravi $M_o = 2$

V skupini od 5,o1-1o,o ha so najpogostejša (2o8) gospodarstva, ki imajo po tri krave $M_o = 3$.

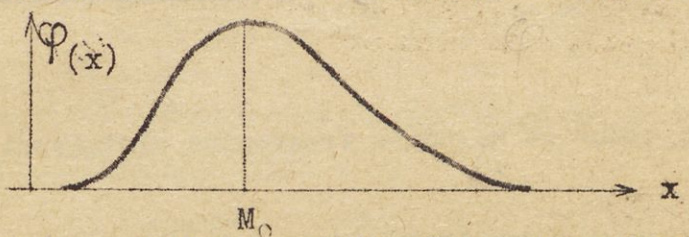
Vrednost modusa se veča torej z velikostjo posestva.

85.3 V primerih, da je osnovni znak serije zvezen je določanje modusa težje.

Ako imamo razdelitev frekvenc za zvezni znak dano s frekvenčno krivuljo, (glej odstavek 6 o serijah), je modus ona vrednost, za katero je:

gostota največja $$\varphi_{(x)} = \max_{M_o} \qquad x = M_o \qquad (8.44)$$

V tem primeru se modus določa analitično sl. 8.2



Običajno pa imamo opravka s frekvenčnimi distribucijami s končnimi razredi. V tem primeru gostota frekvence v nekem razredu $$\varphi_k = \frac{N_k}{\Delta_k}$$

pomeni povprečno gostoto frekvenc v danem razredu. Iz velikosti grupnih gostot frekvenc moremo torej ugotoviti samo nivo gostote frekvenc v posameznih razredih. Iz opazovanja teh gostot moremo sklepati, da je modus v onem razredu, v katerem je nivo gostot, torej φ_k največji. Ako je širina vseh razredov enaka, moremo napraviti isti sklep iz frekvenc n_k, ker so v tem primeru frekvence sorazmerne gostoti.

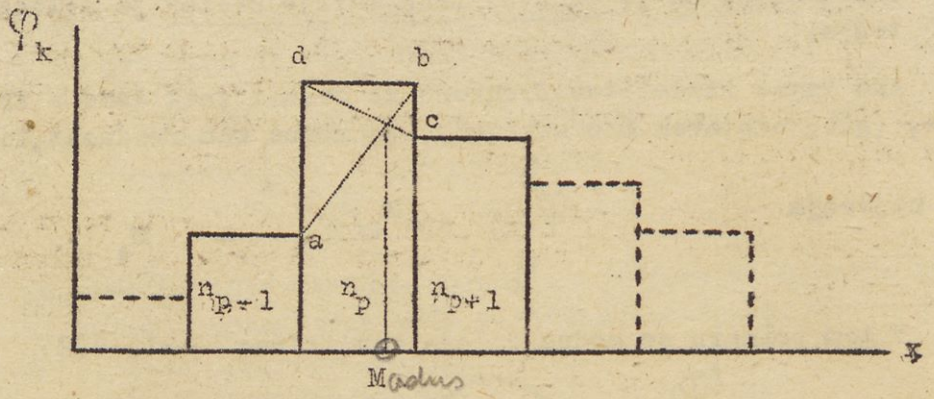
a) Približno mesto modusa v tem razredu se da v tem primeru izračunati iz največje frekvence in frekvenc razredov, ki so sosednji razredu z največjo frekvenco, po obrazcu:

$$M_o = x_{p,min} + \Delta \frac{n_p - n_{p-1}}{2n_p - n_{p-1} - n_{p+1}} \qquad (8.45)$$

Oznake pomenijo:
p je razred z največjo frekvenco.
$x_{p,min}$ je spodaja meja tega razreda.
n_{p-1}, n_p, n_{p+1} so zaporedne frekvence.
Δ je širina razreda.

b) Ako imamo frekvenčno distribucijo dano grafično v histogramu, moremo M_o določiti grafično, kot kaže slika. 8.3 Zvežemo a z b, c z d. Kjer se ti daljici sečeta, potegnemo navpičnico do abcisne osi. Kjer se navpičnica dotika abcise, je modus.

Sl. 8.3 Grafično določanje modusa



P Določiti modus iz navedene frekvenčne distribucije 9o oseb po starosti.

Tab. 8.23

K	Starostni razred	Število oseb n_k
1	nad 13-14 let	1
2	" 14-15 "	9
3	" 15-16 "	16
4	" 16-17 "	35
5	" 17-18 "	19
8	" 18-19 "	8
7	" 19-2o "	2
		9o

$p=4, x_{4,min}=16, n_3=16, n_4=35, n_5=19$

$$\Delta = 1$$

$$M_o = x_{4,min} + \Delta \cdot \frac{n_4-n_3}{2n_4-n_3-n_5}$$

$$M_o = 16 + 1 \cdot \frac{35-16}{2.35-16-19} = 16+\frac{19}{35}=$$

$$= 16{,}54 \text{ let}$$

85.4 Pomen in lastnosti modusa.

a) Modus je sredina, ki ima realno osnovo in je lahko določljiv. Zaradi tega se večkrat uporablja že pri statističnem opazovanju (m.pr. najpogostejša cena v censki statistiki itd).

b) Modus je zelo neobčutljiv napram spremembam vrednosti znakov in svoje vrednosti ne izpremeni vse dotlej, dokler ostane maksimum gostote isti.

c) Medtem ko ima ena distribucija samo po eno izmed vseh ostalih srednjih vrednosti, more imeti več modusov (lokalnih).

d) V primerih, da ima več vrednosti ali rzaredov eno in isto frekvenco, modus ni določljiv.

86. Povzetek.

Da je možna preglednost in primerljivost med statističnimi masami se v statistiki poslužujemo srednjih vrednosti. Srednja vrednost je število, ki izraža značilnosti vseh enot cele statistične mase. Pravilna uporaba srednje vrednosti je zasnovana na pravilnem grupiranju, grupnih sredinah in metodi individualnega primera. Srednjih vrednosti imamo več vrst.

Najvažnejša je aritmetična sredina. Ako imamo vrednosti dane za vsako enoto, izračunavamo aritmetično sredino po obrazcu za izračunavanje navadne aritmetične sredine.

$$\bar{x} = \frac{1}{n}\sum x_j$$

ako pa imamo dane vrednosti v frekvenčni distribuciji pa jo izračunavamo po obrazcu za tehtano aritmetično sredino.

$\bar{x} = \frac{1}{n}\sum n_k x_k$. Frekvence n_k imenujemo teže, ali s tujim izrazom pondere.

Za okrajšanje izračunavanja aritmetične sredine se poslužujemo več metod. Z vpeljavo pogojne sredine x_o se postopek skrajša, ker so vrednosti novega znaka $u = x_o$ prikladnejše za izračunavanje. $\bar{x}$ in $\bar{u}$ sta v enostavni medsebojni zvezi $\bar{x} = x_o + \Delta \bar{u}$.

Z uporabo metode kumulativnih serij se izračunavanje še bolj poenostavi. Iz frekvenčne distribucije izračunamo vrednosti kumulativ N_o in N'_o, iz teh pa

$C_1 = N'_o : N_o$. V tem primeru je $\bar{x} = x_{min} + \Delta C_1$. ($x_{min}$ je najnižja grupna vrednost v distribuciji).

Iz grupnih aritmetičnih sredin moremo izračunati sumarno aritmetično sredino po obrazcu $\bar{x} = \frac{1}{n}\sum n_k \bar{x}_k$ kot tehtano aritmetično sredino grupnih aritmetičnih sredin. Iz obrazca je razvidno, da je sumarna aritmetična sredina odvisna od grupnih sredin ($\bar{x}_k$) in strukture grup (n_k). Zaradi tega je njih analitična vrednost, kadar se ne poslužujemo stalnih struktur.

Povprečno relativno število moremo izračunati na več načinov. Način je odvisen od tega, iz kakšnih podatkov ga izračunavamo. Relativno število naj bo: $R = \frac{Š}{I}$ (R relativno število), Š količina v števcu, I količina v imenovalcu.

Ako imamo dano $Š_k$ in I_k je povprečno relativno število enako $R = \frac{\sum Š_k}{\sum I_k}$. Ako imamo pa dana R_k in I_k, pa povprečno relativno število izračunavamo po obrazcu: $R = \frac{\sum R_k I_k}{\sum I_k}$.

Lastnosti aritmetične sredine so naslednje:

1./ Aritmetična sredina leži med najmanjšo in največjo vrednostjo

$$x_{min} < \bar{x} < x_{max}.$$

2./ Vsota individualnih odklonov od $\bar{x}$ je enaka 0. $\sum (x_i - \bar{x}) = 0$

3./ Izraz $\sum (x - M)^2$ je najmanjši, ako je $M = \bar{x}$.

Ako imamo pri relativnem številu dane R_k in $š_k$, izračunavamo povprečno relativno število ne kot aritmetično, temveč kot harmonično sredino po obrazcu $H_R = \frac{\sum š_k}{\sum \frac{š_k}{R_k}}$. Notehtana oblika harmonične sredine je $H = \frac{n}{\sum \frac{1}{x_i}}$

Pri dinamičnih serijah izračunavamo povprečen porast kot geometrično sredino. Obrazec za izračunavanje geometrične sredine je

$K = \sqrt[n]{K_1 \; K_2 \dots K_n}$. Povprečen porast moremo izračunati tudi iz prvega in zadnjega člena časovne serije po obrazcu $K = \sqrt[n-1]{\frac{Y_n}{Y_1}}$

Iz gornje definicije sledi, da je $\log K = \frac{1}{n} \cdot (\log K_1 + \log K_2 + \log K_n)$. Po tem obrazcu običajno v praksi izračunavamo geometrično sredino.

Srednja vrednost po legi je mediana. Mediana je vrednost, od katere ima polovico enot statistične mase manjše, polovico pa večje vrednosti od mediane. Mediano moremo iz individualnih vrednosti dobiti, ako vrednosti uredimo po velikosti, iz frekvenčne distribucije pa moremo približno vrednost določiti iz izračunavanjem.

Druga srednja vrednost po legi je modus, normalna ali najpogostejša vrednost. Modus je vrednost, ki se v statistični masi najpogosteje pojavlja. Modus je lahko razumljiv in določljiv.

9. ČASOVNE SERIJE

90. SPLOŠNO

Socialno- ekonomski pojavi niso nespremenljivi, temveč se v svojem časovnem razvoju stalno spreminjajo. Te spremembe so rezultat najrazličnejših pogojev, ki vplivajo na pojav. Naloga statistike je, zasledovati časovni razvoj in iskati zakonitosti časovnega razvoja masovnih pojavov.

Sliko dinamike masovnih pojavov dobimo s časovnimi serijami. Časovna serija je niz istovrstnih statističnih podatkov, v katerem se poedine vrednosti nanašajo na zaporedne momente ali intervale. Pri filmski kameri je posamezna slika negibna, v časovnem zaporedju pa dajo slik gibanja. Enako da skupnost poedinih členov časovne serije, ki posamezno prikazujejo stanje, sliko gibanja.

Časovne serije so edino sredstvo, s katerimi moremo proučevati časovni razvoj masovnih pojavov. V njih se izraža dinamika in kakovostne spremembe masovnih pojavov. Časovna serija prikazuje spremembe pojavov v odvisnosti od časa. Po namenu, obliki in načinu proučevanja se povsem razlikujejo od ostalih vrst statističnih serij. Metode tvorjenja, prikazovanja in proučevanja so zaradi tega drugačne od metod, ki jih uporabljamo za proučevanje ostalih vrst statističnih serij.

Časoven razvoj gospodarskega dogajanja običajno ne proučujemo z eno samo časovno serijo, temveč s celim sistemom časovnih serij pojavov, ki tvorijo skupaj področje proučevanja.

90.1 Proučevanje časovnih serij v kapitalizmu

Značilno za kapitalizem je stihijski razvoj gospodarskih pojavov. Zanimanje za zakonitosti ekonomskega razvoja je jasno veliko, ker se more s spoznavanjem teh zakonitosti oni, ki jih pozna, okoristiti. Zaradi tega se je proučevanje časovnih serij v kapitalizmu pojavilo zgodaj (v 19. stol. "nauk o krizah")in doseglo kasneje kot nauk o konjunkturah neverjeten razvoj. Nauk o konjunkturi pa v svoji osnovi predpostavlja kapitalistično ureditev individualnega gospodarjenja. Njegova naloga je, na podlagi izsledkov o zakonitostih ekonomskega dogajanja sklepati na gospodarsko dogajanje v bodočnosti. Vendar pomeni leto 1929 velik poraz formalističnega analiziranja ekonomskih pojavov, ko je Harvardski konjunkturni institut napovedal prosperiteto, izbruhnila pa je svetovna gospodarska kriza.

90.2 Pomen analize časovnih serij v socializmu

N-amesto stihije v kapitalističnem gospodarstvu stopi v socializmu plansko gospodarstvo.

Po čl. 15 Ustave FLRJ usmerja gospodarsko življenje in razvoj

država s pomočjo splošnega gospodarskega plana. Pri tem se opira na državni in zadružni gospodarski sektor, ter izvaja splošno kontrolo nad zasebnim sektorjem v gospodarstvu.

Nauk o kunjunkturi zaradi tega za proučevanjenašega gospodarstva dejansko ne pride v poštev. Pač pa moremo koristiti mnoge statistične metode časovnih serij, ker je v našem gospodarstvu vseeno mnogo pojavov, ki jih moremo formalno analizirati s temi metodami. Ti pojavi nastopajo zaradi vpliva od zunaj, vpliva naravnih pojavov, privatnega sektorja i.t.d.

91. VRSTE ČASOVNIH SERIJ

91.1 Momentne in intervalne časovne serije

Po značaju podatkov, ki jih časovna serija predstavlja, delimo serije na momentne in intervalne.

.11 Pri momentnih ali trenutnih serijah se vrednosti posameznih členov nanašajo na časovne momente. Odtod tudi ime.

Kot primer momentne časovne serije moremo vzeti indeks življenjskih stroškov po podatkih serije "Indeks", ki jo je pred vojno izdajal Benko Grado v Zagrebu.

Tab. 9.1

Datum	Indeks
Ø 1914	100
1.dec. 1930	1646
1.dec 1931	1544
1.dec. 1932	1349
1.dec. 1933	1183
1.dec. 1934	1141
1.dec. 1935	1162
1.dec. 1936	1196
1.dec. 1937	1285

Enako je momentna časovna serija tudi stanje zalog koncem meseca po mesecih ali število živine ne določen datum po letih itd. Vsak posamezen člen momentne časovne serije daje sliko stanja ob določenem trenutku, torej popolnoma statično - negibno sliko. Daljši niz teh členov pa poda gibanje, oziroma spreminjanje tega stanja, daje torej dinamično sliko pojava.

P Ako pogledamo še en primer momentne časovne serije, bomo mogli odkriti važen pogoj, ki ga morajo izpolnjevati momentne časovne serije, ako hočemo, da bomo časovno serijo mogli pravilno analizirati. V časovni seriji števila prebivalstva v LRS moti analizo to, da razmaki med posameznimi stanji niso enaki. Med letom 1890 in 1900 je razdobje deset let, med letom 1949do 1950

Tab. 9.2

Število prebivalstva v Sloveniji

Datum	Število prebivalstva v 000 preb.
1890	1002
1900	1026
1910	1067
1921	1060
31.3. 1931	1144
15.3. 1948	1392
30.6. 1949	1409
30.6 1950	1422

ra samo eno leto. Zaradi tega ta časovna serija ne kaže prave slike dinamike. Medtem, ko je sprememba 24000 prebivalcev nastala 1890 do 1900 v razdobju deset let, je nastala razlika 13000 prebivalcev od 1949 do 1950 v enem letu. Čeprav je 13000 manj kot 24000, je porast na koncu serije večji kot na začetku, kjer se je število prebivalstva izpremenilo za 24000 v desetih letih. Zaradi tega stremimo za tem, da so intervali med členi momentne časovne serije enaki. Le na takih serijah moremo vršiti analiziranje dinamike.

Poleg tega časovno serijo, prikazano v tabeli 9.2 ne moremo uporabiti v celoti kot prikaz dinamike števila prebivalstva v Sloveniji, ker se od leta 1890 do 1931 podatki nanašajo na področje dravske banovine, od 1948 - 1950 pa na področje LR Slovenije.

.12 Člani intervalne časovne serije so količine, ki se nanašajo na intervale. Intervalno serijo dobimo n.pr. ,ako prikazujemo po mesecih industrijsko proizvodnjo, število rojstev, porok, smrti, blagovni promet itd.

P Kot primer navajamo število živorojenih otrok v FLRJ v letu 1949 po mesecih. Ker gre v tej seriji za absolutne vrednosti, je jasno, da vpliva dolžina intervala na število rojstev. V mesecih z manjšim številom dni je bilo naravno tudi manj rojstev, kot bi jih bilo pod istimi pogoji, ako bi bil mesec daljši. En del spremenljivosti podatkov gre na račun različne dolžine mesecev. Ker pa različna dolžina mesecev nima nič opraviti z vsebino pojava, analizo samo moti. Zaradi tega je boljše, da ta vpliv odstranimo. To moremo doseči s tem, da absolutne podatke po mesecih, kot pravimo, reduciramo na enako dolgo časovna razdobja. V našem primeru more biti to dan ali pa mesec z enakim številom dni (30 dni).

Tab. 9.3

Mesec	Število živorojenih otrok
J	39902
F	39845
M	43649
A	40359
M	39858
J	37557
J	41772
A	42269
S	40285
O	41936
N	37678
D	36165

Ako preračunamo podatke na dan moramo število rojstev za vsak mesec deliti s pripadajočim številom dni v mesecu.

Tab.9.4

Mesec	Število živorojenih otrok	Število dni v mesecu	Povprečno število rojstev na dan	Faktor reduciranja na 30 dni	Število živorojenih otrok reduc. na 30 dni
(1)	(2)	(3)	(4)=(2):(3)	(5) = $\frac{30}{(3)}$	(6)=(2).(5)
J	39902	31	1290	0,97	38600
F	39845	28	1420	1,07	42500
M	43649	31	1410	0,97	42300
A	40359	30	1340	1,00	40359
M	39858	31	1290	0,97	38600
J	37557	30	1250	1,00	37557
J	41772	31	1350	0,97	40400
A	42269	31	1370	0,97	41000
S	40285	30	1340	1,00	40285
O	41936	31	1350	0,97	40600
N	37678	30	1260	1,00	37678
D	36165	31	1170	0,97	35000

Ako podatke reduciramo na mesec z enotnim številom dni 30 dni, moramo za mesec z 31 dni absolutne podatke pomnožiti s faktorjem $\frac{30}{31} = 0,9677$, za februar z 28 dni s faktorjem $\frac{30}{28} = 1,0744$ za prestopna leta, ko ima 29 dni s $\frac{30}{29} = 1,0345$. Za mesec s 30 dni podatki ostanejo, ker je faktor $\frac{30}{30} = 1$.

Enako moramo paziti, ako imamo opravka z drugimi koledarskimi intervali. Posamezna tromesečja imajo n.pr, različno dolžino dni I = 90, II = 91, III = 92, IV = 92. Tudi dekade, računane po mesecih imajo 8, 10 ali 11 dni. Poleg tega za nekatere pojave, kot n.pr. proizvodnja ako ni bilo neprekinjeno, ni merodajno število koledarskih dni, temveč število delovnih dni. V teh primerih morejo motnje izvirajoče iz različnega števila delovnih dni v intervalu biti še večje.

91.2 Kumulativna časovna serija.

Ako pogledamo primer števila rojstev po mesecih v FLRJ, imamo v tabeli 9.4 tabelo intervalne časovne serije: serijo dejanskega absolutnega

števila rojstov (kolona (2)) serijo povprečnega števila rojstev na en dan (kolona (4)) in serijo števila rojstev, reduciranih na mesec s 30 dni (kolona (6)). Vse te tri serije so intervalne časovne serije. Vendar obstoja med njimi bistvena razlika. Člene serije dejanskega števila rojstev moremo seštevati (ekstenziven značaj), medtem ko je vsota členov povprečnega števila rojstev brez smisla (intenziven značaj).

Iz intervalne časovne serije, za katero ima seštevanje členov smisel, moremo tvoriti novo časovno serijo, ki jo imenujemo kumulativno časovno serijo. Kumulativno časovno serijo dobimo s postopnim seštevanjem členov osnovne intervalne časovne serije.

P Ako vzamemo kot primer časovno serijo števila rojstev in prejšnjega odstavka, izračunamo kumulativno časovno serijo takole:

Tab. 9.5

Mesec	Število rojstev v mesecu	Kumulativna serija
J	39902	0
F	39845	39902
M	43649	79747
A	40359	123396
M	39858	163755
J	37557	202613
J	41772	241170
A	42269	282942
S	40285	325211
O	41936	365496
N	37678	407432
D	36165	445110
	Skupaj	481275

Prvi člen kumulativne serije je 0. Drugi člen smo izračunali tako, da vrednosti 0 prištejemo prvi člen osnovne serije + 0 + + 39902 = 39902. Tretji člen kumulativne serije dobimo, da drugemu členu kumulativne serije prištejemo drugi člen osnovne serije 39902 + 39845 = 79747 . Četrti člen kumulativne serije (123396), dobimo, da tretjemu členu kumulativne serije prištejemo tretji člen osnovne serije 79747 + + 43649 = 123396 in tako dalje.

Iz tega moremo sestaviti splošno pravilo za izračunavanje ostalih členov kumulativne serije. Vsak člen kumulativne časovne serije moremo izračunati tako, da prištejemo predhodnemu členu kumulativne serije člen osnovne serije za ta mesec. Ako zaznamujemo člene kumulativne serije z K_k, člene osnovne serije pa z Y_k, moremo to pravilo pisati v shemi v tabeli 9,6.

Tab. 9.6

Osnovna serija	Kumulativna serija
1. Y_1	$K_1 = 0$
2. Y_2	$K_2 = K_1 + Y_1$
3. Y_3	$K_3 = K_2 + Y_2$
K-1 Y_{k-1}	$K_k = K_{k-1} + Y_{k-1}$
K Y_k	$K_{k+1} = K_k + Y_k$

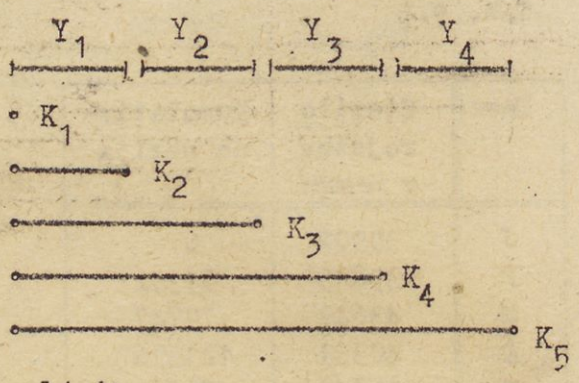
Ta shema izračunavanja kumulativne časovne serije je splošna in ne velja samo za naš primer

Ako pogledamo drugi člen kumulativne serije (za februar 39902) vidimo, da je to število rojstev v januarju ali z drugimi besedami: Število rojstev od začetka leta do začetka februarja. Člen kumulativne serije za marec je seštevek rojstev v januarju in februarju. Zaradi tega pomeni število rojstev od začetka leta do začetka marca.

Pomen nadaljnjih členov kumulativne serije je enak. Člen kumulativne serije za avgust 282942 pomeni n.pr., da je bilo do začetka avgusta v letu 1949 282942 rojstev. V sliki moremo pomen členov kumulativne serije ponazoriti tako, da z daljicami naznačimo za kateri interval posamezen člen velja.

Iz slike vidimo, da se pri intervalih členov kumulativne serije izpreminja samo eno krajišče, medtem ko ostane začetno pri vseh členih stalno. Zaradi tega moremo vsaj tehnično smatrati, da je kumulativna serija momentna časovna serija, ker je vsak člen kumulativne serije določen z enim samim krajiščem, ker je drugo krajišče stalno. Zato običajno vrednosti členov kumulativne serije pripišemo spremenljivim krajiščem intervalov, za katere členi veljajo.

Sl. 9.1



Kumulativne časovne serije se zelo pogosto uporabljajo v ekonomski statistiki, posebno pa v planski in v operativni evidenci, kjer izračunavamo kumulative za posamezna zaokrožena razdobja (mesečna kumulativa za leto, dnevna kumulativa za mesec).

91.3 Časovna serija drsečih vrednosti.

P Vzemimo časovno serijo števila natovorjenih vagonov v bivši Jugoslaviji v 000 vagonih po četrtletjih za razdobje 1933 - 1935. (Tab. 9.7).

Tab. 9.7

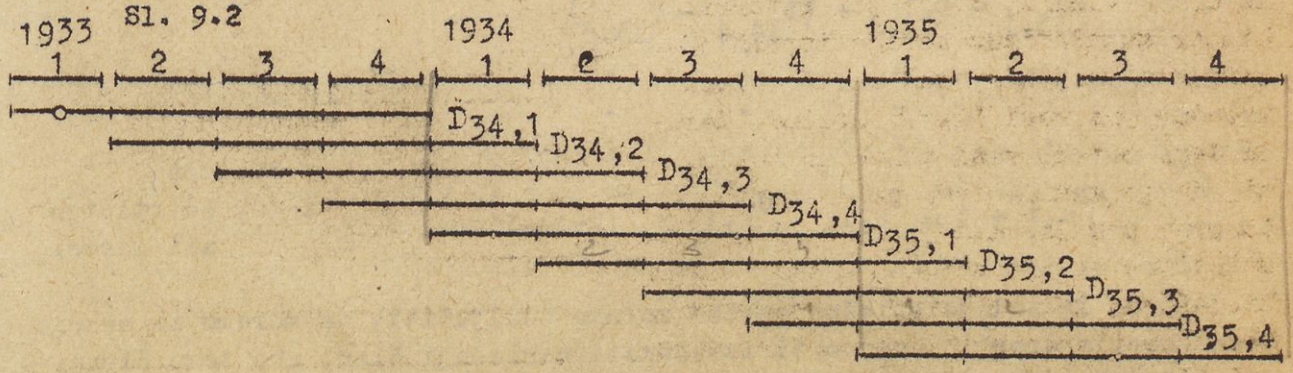
Četrtletje		Število natovorjenih vagonov v 000
1933.	I	285,3
	II.	314,6
	III.	389,7
	IV.	388,8
1934.	I.	303,4
	II.	324,5
	III.	373,3
	IV.	396,7
1935.	I.	284,8
	II.	331,2
	III.	415,1
	IV.	427,7

Število v letu 1933 natovorjenih vagonov dobimo, ako seštejemo podatke prvih štirih četrtletij. 285,3 + + 314,6 + 389,7 + 388,8 = 1378,4. Na enak način moremo dobiti število natovorjenih vagonov v letu 1934 in 1935. Poleg teh treh letnih vrednosti pa moremo dobiti še druge letne intervale in podatke, ki se sicer ne skladajo z zaključenim kolendarskim letom, vendar obsegajo razdobja enega leta. Razdobje enega leta dobimo tudi,

ako vzamemo drugo, tretje in četrto četrtletje leta 1933 in prvo četrtletje leta 1934. Število natovorjenih vagonov v tem razdobju dobimo, ako seštejemo odgovarjajoče tromesečne podatke. 314,6 + 389,7 + 388,8 + 303,4 = 1396,5. Na enak način dobimo leto, ako seštejemo podatke za tretje in četrto četrtletje leta 1933 in prvo in drugo četrtletje leta 1934: 389,5 + 388,8 + 303,4 + 324,5 = 1407,4. Postopek moremo nadaljevati. Na ta način moremo dobiti celo serijo letnih podatkov, ki veljajo za intervale, ki niso premaknjeni za celo leto, temveč samo za četrtletje. V sliki 9.2 je ponazorjeno gornje izvajanje z daljicami, ki predstavljajo tvorjene intervale:

drsece vrednosti

Sl. 9.2



Na spodnjih daljicah se lepo vidi, iz katerih intervalov so sestavljeni posamezni letni intervali. Ako prvemu letnemu intervalu odvzamemo prvo četrtletje leta 1933 in dodamo prvo četrtletje leta 1934, dobimo drugi interval. Iz drugega intervala moremo dobiti tretjega, ako od drugega odvzamemo drugo četrtletje leta 1933 in dodamo drugo četrtletje leta 1934 itd. Ker na ta način tvorjene letne vrednosti nekako drse preko osnovne časovne serije, imenujemo to serijo letnih vrednosti časovno serijo drsečih vrednosti. Vrednosti posameznih členov serije drsečih vrednosti pripišemo koncu odgovarjajočega intervala. Prvi člen pripišemo koncu četrtega četrtletja leta 1933, drugi člen pripišemo koncu prvega četrtletja leta 1934 itd.

Ako zaznamujemo z $Y_{33,1}$ člen osnovne serije za prvo četrtletje leta 1933, z $Y_{33,2}$ člen osnovne serije za drugo četrtletje istega leta, z $Y_{34,1}$ člen za prvo četrtletje leta 1934 z $D_{34,1}$ prvi člen $D_{34,2}$ drugi člen serije drsečih vrednosti, moremo v teh oznakah napisati, kako izračunamo iz osnovne intervalne serije serijo drsečih vrednosti.

$$D_{34,1} = Y_{33,1} + Y_{33,2} + Y_{33,3} + Y_{33,4}$$

$$D_{34,2} = Y_{33,2} + Y_{33,3} + Y_{33,4} + Y_{34,1}$$

$$D_{34,3} = Y_{33,3} + Y_{33,4} + Y_{34,1} + Y_{34,2}$$

Iz teh treh členov vidimo, da se v dveh zaporednih členih drseče

serije nekaj členov osnovne serije ponovi. Iz $D_{34,1}$ dobimo torej $D_{34,2}$, ako od $D_{34,1}$ odštejemo $Y_{33,1}$ in prištejemo $Y_{34,1}$. Pisano v obliki enačbe imamo:

$$D_{34,2} = D_{34,1} + (Y_{34,1} - Y_{33,1})$$

$$D_{34,3} = D_{34,2} + (Y_{34,2} - Y_{33,2})$$

$$D_{34,4} = D_{34,3} + (Y_{34,3} - Y_{33,3})$$

$$D_{35,1} = D_{34,4} + (Y_{34,4} - Y_{33,4})$$

. . . .

. . . .

. . . .

$$D_{k,i+1} = D_{k,i} + (Y_{k,i} - Y_{(k-1),i})$$

k-to leto
i-to, četrtletje ali mesec.

Iz zadnje splošne enačbo moremo zaključiti, da moremo iz nekega člena serije drsečih vrednosti izračunati naslednji člen, ako temu členu prištejemo odgovarjajočo razliko členov dveh zaporednih let. Ako zaznamujemo:

$$Y_{k,i} - Y_{k-1,i} = d_{ki} \text{ je } D_{ki+1} = D_{ki} + d_{ki}. \qquad (9.1)$$

Račun izvedemo na našem primeru v naslednjih stopnjah:

Tab. 9.8

		Y_{ki}	$Y_{(k-1),i}$	d_{ki}	D_{ki}
1933	1	285,3			
	2	314,6			
	3	389,7			
	4	388,8			
1934	1	303,4	285,3	+ 18,1	1378,4
	2	327,6	314,6	+ 13,0	1396,5
	3	373,9	389,7	- 15,8	1409,5
	4	396,7	388,8	+ 7,9	1393,7
1935	1	284,8	303,4	- 18,6	1401,6
	2	330,3	327,6	+ 2,7	1383,0
	3	415,8	373,9	+ 41,9	1385,7
	4	427,7	396,7	+ 31,0	1427,6
					1458,6

1. Osnovno časovno serijo napišemo kot kaže primer, dvakrat
Druga serija je za eno leto pomaknjena naprej.
2. Izračunamo diference iz istoležnih členov napisanih serij (n.pr. 303,4 - 285,3 = + 18,1.)
3. Izračunamo vsoto za prvo leto in jo vpišemo kot prvi člen 285,3 + + 314,6 + 389,7 + 388,8 = = 1378,4.
4. Naslednje člene serije drsečih vrednosti izračunamo, da znanemu členu serije prištejemo istoležno razliko d_{ki}. To je novi člen serije drsečih vrednosti, ki ga vpišemo v novo

vrsto. 1378,4 + 18,1 = 1396,5 , 1396,5 + 13,0 = 1409,5 itd.

DSerijo drsečih vrednosti ne izračunavamo samo v primeru četrtletnih podatkov. Dostikrat izračunavamo serije drsečih vrednosti tudi iz mesečnih, dnevnih ali dekadnih serij. Važno si je zapomniti le to, da vzamemo v vseh primerih vsoto takega števila členov, da tvorijo členi serije drsečih sredin zaključen interval. Pri mesečnih serijah vzamemo vsoto po dvanajst členov, da dobimo letne vsote, enako iz istega razloga pri četrtletnih serijah štiri člene. Pri dekadnih časovnih serijah vzamemo pri izračunavanju serije drsečih vrednosti po tri člene, da dobimo mesečne vrednosti, pri dnevnih serijah pa po sedem členov, da dobimo tedenske vsote. Le tako tvorjenje serije drsečih vrednosti morejo služiti za proučevanje dinamike pojavov (glej Z diagram po - glavje 10.)

91.4 Časovne serije sredin

Časovne serije imajo dostikrat zelo veliko število členov. Mislimo si samo dnevno časovno serijo števila delavstva za eno leto, ali mesečno časovno serijo proizvodnje za več let. Vprašanje je, na kakšen način bi mogli število členov časovne serije skrčiti, da bi p.. serija kljub temu pokazala glavne črte razvoja pojava, Ker moremo proizvodnjo seštevati, moremo v tem primeru skrčiti število členov časovne serije s tem, da mesečne proizvodnje seštejemo v letne. Na ta način smo število členov skrčili od 120 mesečnih vrednosti, ako smo imeli desetletno serijo, na deset letnih vrednosti. Pripomniti pa je treba, da letna časovna serija z deset členi ne more nadomestiti mesečne serije in je z njo odkazana samo osnovna smer razvoja proizvodnje, ne pa podrobno gibanje proizvodnje, kot moremo to razvideti iz mesečne serije. Vendar ta serija ne pokaže razvoja mesečne proizvodnje, temveč razvoj letne proizvodnje. Ako želimo, da v skrčeni obliki podamo razvoj mesečne proizvodnje, ne zadostuje, da izračunamo samo letno proizvodnjo, temveč moramo izračunati povprečno mesečno proizvodnjo za posamezna leta.

Ako vzamemo kot številčni primer število natovorjenih vagonov v b. Jugoslaviji v letih 1933 do 1938, dobimo naslednjo sliko:

Tab.9.9

	1933	1934	1935	1936	1937	1938
J	89,0	98,9	90,5	96,6	107,1	114,6
F	86,6	94,4	90,0	92,5	108,1	120,1
M	109,7	110,1	104,3	111,6	122,2	148,0
A	97,1	101,1	103,8	102,4	128,4	134,8
M	110,8	113,3	116,0	114,0	129,9	143,2
J	106,7	113,2	110,5	116,1	136,2	139,9
J	109,5	111,3	125,7	129,5	132,5	148,5
A	127,6	125,9	137,4	145,0	163,4	168,8
S	152,6	136,7	152,7	155,0	169,2	171,9
O	152,2	153,7	169,9	171,0	182,3	178,1
N	130,6	132,5	147,4	150,8	151,1	156,1
D	106,0	110,8	110,4	129,4	133,8	132,1
Σ	1378,4	1401,6	1458,6	1513,9	1684,6	1756,5
$\bar{Y} = \frac{1}{n}\Sigma$	114,9	116,8	121,6	126,2	140,4	146,4

Letne vsote podajajo pregledno v skrčeni obliki porast števila natovorjenih vagonov od 1933 do 1938. Vendar se ti podatki gibljejo na znatno višji ravni kot mesečni podatki. Razvoj mesečnega števila natovorjenih vagonov poda šele serija letnih povprečij mesečnih podatkov. Iz letne vsote dobimo mesečno povprečje, ako letne vsote delimo z 12.

$$\bar{Y}_{1933} = \frac{1}{12} \cdot 1378{,}4 = 114{,}87 \qquad \bar{Y}_{1936} = \frac{1}{12} \cdot 1513{,}9 = 126{,}16$$

$$\bar{Y}_{1934} = \frac{1}{12} \cdot 1401{,}6 = 116{,}80 \qquad \bar{Y}_{1937} = \frac{1}{12} \cdot 1684{,}6 = 140{,}38$$

$$\bar{Y}_{1935} = \frac{1}{12} \cdot 1458{,}6 = 121{,}55 \qquad \bar{Y}_{1938} = \frac{1}{12} \cdot 1756{,}5 = 146{,}38$$

Važno je, da moremo serijo sredin izračunati tudi v primerih momentnih časovnih serij ali intervalnih serij, za katere vsota členov nima smisla. Čeprav vsota členov vsebinsko nima smisla, so povprečne vrednosti tudi za te vrste serij smiselne in moremo brez nadaljnjega tvoriti serije povprečnih vrednosti tako momentnih kot intervalnih serij.

Serija sredin je v vsakem primeru intervalna serija. Posamezen člen serije sredin se nanaša na interval, ki je vsota intervalov, na katere se nanašajo členi, iz katerih smo tvorili sredino. V našem primeru se posamezna mesečna povprečja nanašajo na celo leto, od 1. januarja do 31. decembra posameznih let. Ker moremo tudi pri momentnih serijah okrog posameznih členov določiti pripadajoč interval (polovico intervala levo in desno), je tudi v teh primerih možno določiti interval, na katerega se nanaša sredina.

Pri vseh vrstah časovnih serij pripišemo vrednost vsakega člena neki točno določeni vrednosti. Pri osnovnih časovnih serijah je to sredina, pri kumulativnih časovnih serijah in serijah drsečih vrednosti pa konec odgovarjajočega intervala. Prav tako pripišemo tudi pri serijah sredin, vrednost sredine neki vrednosti, in sicer sredini odgovarjajočega intervala.

V našem primeru mesečnih povprečij pripišemo povprečno vrednost vsakega leta sredini leta, to je točno na konec meseca junija.

Serijo sredin pa ne moremo tvoriti brez predhodne proučitve, kaj moramo s serijo sredin doseči in kako bomo to dosegli. Problematika je bila nakazana že pri relativnih številih. V naslednjem odstavku jo hočemo samo poglobiti in izpopolniti. Že pri relativnih vrednostih smo videli, da se izračunavanje časovne sredine po dveh obrazcih:

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + Y_3 + \ldots + Y_n) \qquad (9.2)$$

ali

$$\bar{Y} = \frac{1}{n}\left(\frac{1}{2} Y_1 + Y_2 + Y_3 + \ldots + Y_n + \frac{1}{2} Y_{n+1}\right) \qquad (9.3)$$

Enkrat vzamemo končna člena v vsoto s celo, drugič pa samo s polovično vrednostjo. Katerega izmed obrazcev vzamemo v poštev ni poljubno, temveč odvisno od namena.

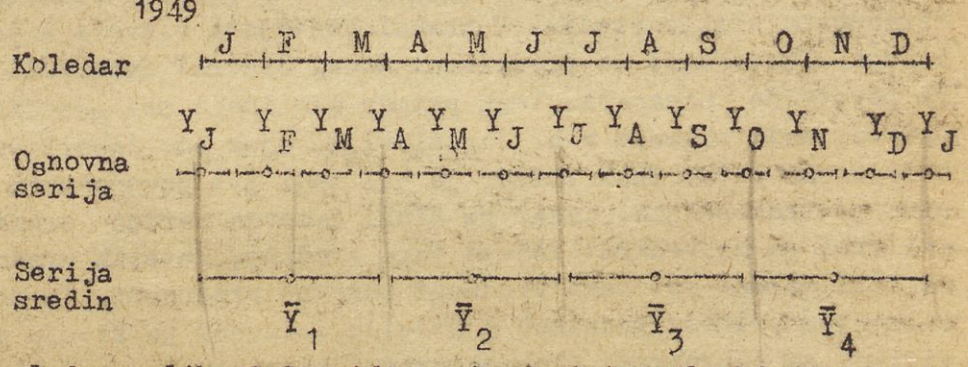
Vzemimo kot primer serijo števila delavstva v neki tovarni po stanju v začetku meseca (podatki izmišljeni).

Tab. 9.10

Prvega v mesecu		Število delavcev
1949	J	283
	F	327
	M	314
	A	343
	M	329
	J	362
	J	385
	A	410
	S	428
	O	442
	N	453
	D	480
1950	J	490

Izračunati je potrebno serijo povprečnega števila delavstva po koledarskih tromesečjih. Najenostavnejše bomo nalogo rešili, ako si pomagamo s tem, da narišemo intervale, ki odgovarjajo členom osnovne in povprečne serije. Čeprav je število delavcev na določen dan moment en podatek, mu pripišemo interval, ki sega od datuma, za katerega velja, za polovico meseca levo in desno.

Sl. 9.3



Ako pogledamo sliko 9.3, vidimo, da je interval, ki pripada povprečju za prvo četrtletje ($\bar{Y}_1$) sestavljen iz ene polovice intervala, ki pripada podatku za prvi januar 1949 (Y_J) celega intervala, ki pripada podatku za februar (Y_F) celega intervala, ki pripada podatku za prvi marec in polovice intervala, ki pripada podatku za prvi april. Povprečje za prvo tromesečje izračunamo torej kot tehtano sredino.

$$\bar{Y}_1 = \frac{1}{3}\left(\frac{1}{2}Y_J + Y_F + Y_M + \frac{1}{2}Y_A\right) = \frac{1}{3}\left(\frac{1}{2}283 + 327 + 314 + \frac{1}{2}343\right) = \frac{1}{3}954 = 318$$

Iz istega razloga je:

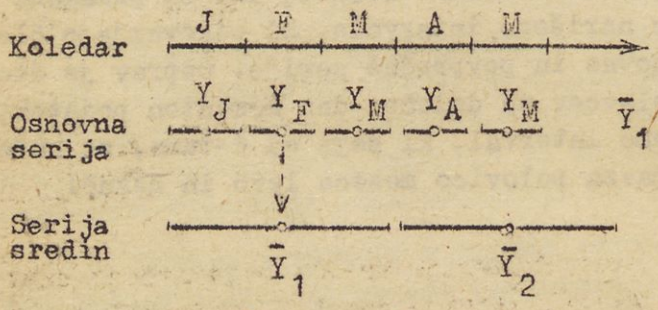
$$\bar{Y}_2 = \frac{1}{3}\left(\frac{1}{2}Y_A+Y_M+Y_J+\frac{1}{2}Y_J\right) = \frac{1}{3}\left(\frac{1}{2}\ 343+329+36[illegible]-\frac{1}{2}\ 385\right) = \frac{1}{3}\ 1055 = 351,7$$

$$\bar{Y}_3 = \frac{1}{3}\left(\frac{1}{2}Y_J+Y_A+Y_S+\frac{1}{2}Y_O\right) = \frac{1}{3}\left(\frac{1}{2}\ 385+410+428+\frac{1}{2}\ 442\right) = \frac{1}{3}\ 1251,5 = 417,2$$

$$\bar{Y}_4 = \frac{1}{3}\left(\frac{1}{2}Y_O+Y_N+Y_D+\frac{1}{2}Y_J\right) = \frac{1}{3}\left(\frac{1}{2}\ 441+453+480+\frac{1}{2}\ 490\right) = \frac{1}{3}\ 1399 = 466,3$$

Ako bi imeli podatke o številu delavstva po stanju sredi meseca, bi izračunavali tromesečna povprečja drugače. Zopet si pomagamo s sliko:

Sl. 9.4



Obstojajo sicer splošna pravila, po katerem obrazcu izračunavamo v posameznih primerih, vendar je najzanesljivejš[illegible], da v vsakem posebnem primeru, ki ni še razčiščem, ugotovimo s pomočjo slike kot zgoraj, kako izračunamo serijo sredin. V praksi obstojata v zvezi s tem običajno dve nalogi. Največkrat so predpisani točni časovni intervali, na katere se mora povprečje nanašati. T-ak primer smo imeli pri mesečni seriji števila natovorjenih vagonov v b. Jugoslaviji, kjer smo iskali povprečje za točna koledarska leta. Enak primer je bil tudi pri povprečnemu številu delav - stva po tromesečjih. Kadar pa služi časovna serija sredin za nadaljno analizo, pa največkrat iščemo take sredino, katerih vrednosti se nanašajo na isti moment kot kaka vrednost osnovne vrednosti. Ta pogoj je izpolnjen v gornjem primeru podatkov o številu delavcev po stanju sredi meseca. Kot vidimo, se februarski člen osnovne serije nanaša na isti časovni moment kot prvi člen serije tromesečnih sredin (glej sliko 9.4)

91.5 Serija drsečih sredin.

Ko smo tvorili serijo drsečih vrednosti, smo imeli pred očmi željo, da pripišemo letno vrednost ne samo enemu členu v letu, temveč vsakemu členu. Isto pomanjkljivost občutimo tudi pri seriji sredin. Sredina se nanaša samo na en trenutek intervala . To vidimo iz primerov serij sredin, ki smo jih tvorili. Za analizo časovnih serij pa je dostikrat potrebna serija sredin, v kateri vsaki osnovni vrednosti odgovarja po ena sredina. Kako to dosežemo, bomo najlažje razvideli iz kratkega primera.

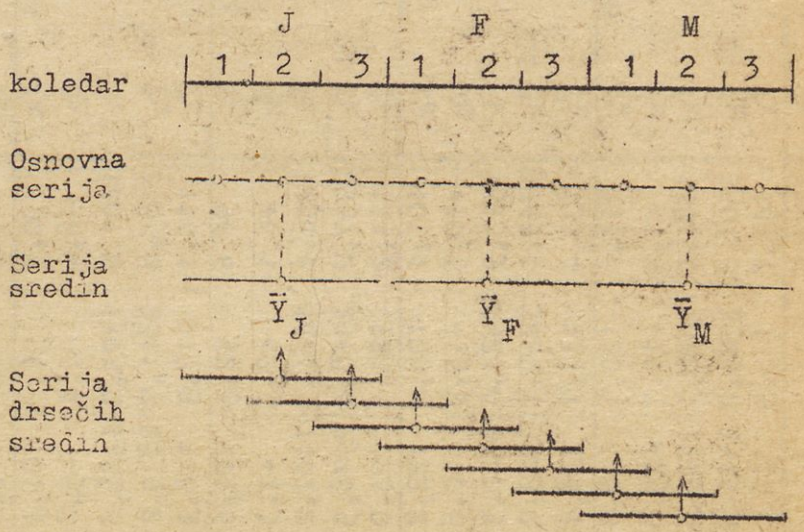
P Vzemimo vrednost proizvodnje nekega podjetja po dokadah za tri

mesece.

Tab. 9.11

Mesec	dekada	Vrednost proizv. v 000 din
J	1	123
	2	184
	3	211
F	1	143
	2	241
	3	203
M	1	161
	2	263
	3	284

Sl. 9.5



9.5

Iz slike vidimo, da členi navadne serije sredin odgovarjajo samo členom druge dekade. Enako kot smo tvorili serijo drsečih vrednosti pa moremo tvoriti tudi serijo drsečih sredin, da izračunamo serijo sredin iz treh zaporednih dekad. Povprečje iz prve, druge in tretje dekade se nanaša na drugo dekado januarja, povprečje iz druge in tretje dekade januarja in prve dekade februarja se nanaša na tretjo dekado januarja itd. Serijo drsečih sredin moremo v tem primeru izračunati tako, da člene serije drsečih vrednosti delimo s tri.

Tab. 9.12

Mesec	Dekada	Vrednost proizv. v 000 din		d_{ki}	D_{ki}	Y_{ki}
		$Y_{k,i}$	$Y_{k-1,i}$			
J	1	123				
	2	184				172,7
	3	211				179,3
F	1	143	123	+ 20	518	198,7
	2	241	184	+ 57	538	196,0
	3	203	211	- 8	595	168,7
M	1	161	143	+ 18	587	176,0
	2	263	241	+ 22	605	203,0
	3	284	203	+ 81	627	
					708	

$\frac{1}{3}$ 518 = 172,7 $\frac{1}{3}$ 538 = 179,3 itd.

Tab.9.13 Število natovorjenih vagonov v b.Jugoslaviji Serija izsečih sredin

		J	F	M	A	M	J	J	A	S	O	N	D
1933	Y_k	89,0	86,6	109,7	97,1	110,8	106,7	109,5	127,6	152,6	152,2	130,6	106,0
	d_k												
	D_k												
	24 $\bar{Y}_k$							2766,7	2784,4	2792,6	2797,0	2803,5	2812,5
	$\bar{Y}_k$							115,3	116,0	116,4	116,5	116,8	117,2
1934	Y_k	98,9	94,4	110,1	101,1	113,3	113,2	111,3	125,9	136,7	153,4	132,5	110,8
	d_k	+9,9	+7,8	+0,4	+4,0	+2,5	+6,5	+1,8	-1,7	-15,9	+1,2	+1,9	+4,8
	D_k	1378,4	1388,3	1396,1	1396,5	1400,5	1403,0	1409,5	1411,3	1409,6	1393,7	1391,9	1396,8
	24 $\bar{Y}_k$	2820,8	2820,9	2803,3	2788,6	2791,7	2798,4	2794,8	2782,0	2771,7	2768,7	2774,1	2774,8
	$\bar{Y}_k$	117,5	117,6	116,8	116,2	116,4	116,6	116,5	116,0	115,5	115,3	116,6	115,6
1935	Y_k	90,5	90,0	104,3	103,8	116,0	110,5	125,7	137,4	152,7	169,9	147,4	110,4
	d_k	-8,4	-4,4	-5,8	+2,7	+2,7	-2,7	+14,4	+11,5	+16,0	+16,5	+14,9	-0,4
	D_k	1401,6	1393,2	1388,8	1383,0	1385,7	1388,4	1385,7	1400,1	1411,6	1427,6	1444,1	1459,0
	24 $\bar{Y}_k$	2785,8	2811,7	2839,2	2871,7	2903,1	2917,6	2923,3	2931,9	2941,7	2947,6	2944,2	2947,8
	$\bar{Y}_k$	116,1	117,3	118,3	119,7	121,1	121,5	121,8	122,2	122,6	122,8	122,7	122,8
1936	Y_k	96,6	92,5	111,6	102,4	114,0	116,1	129,5	145,0	155,0	171,0	150,8	129,4
	d_k	+6,1	+2,5	+7,3	-1,4	-2,0	+5,6	+3,8	+7,6	+2,3	+1,1	+3,4	+19,0
	D_k	1458,6	1464,7	1467,2	1474,5	1473,1	1471,1	1476,7	1480,5	1488,1	1490,4	1491,5	1494,9
	24 $\bar{Y}_k$	2957,2	2968,6	2978,5	2981,9	2986,4	3008,8	3038,3	3064,7	3091,2	3127,8	3169,7	3205,8
	$\bar{Y}_k$	123,2	123,6	124,0	124,3	124,4	125,2	126,5	127,6	128,7	130,2	132,0	133,4
1937	Y_k	107,1	108,4	122,2	128,4	129,9	136,3	152,5	163,4	169,2	182,3	151,1	133,8
	d_k	+10,5	+15,9	+10,6	+26,0	+15,9	+20,2	+23,0	+18,4	+14,2	+11,3	+0,3	+4,4
	D_k	1513,9	1524,4	1540,3	1550,9	1576,9	1592,8	1613,0	1636,0	1654,4	1668,6	1679,9	1680,2
	24 $\bar{Y}_k$	3249,0	3290,4	3323,0	3348,5	3360,1	3364,8	3376,7	3395,9	3433,4	3465,6	3485,3	3502,2
	$\bar{Y}_k$	135,4	137,1	138,4	139,4	139,9	140,1	140,6	141,4	142,9	144,2	145,2	145,7
1938	Y_k	114,6	120,1	148,0	134,8	143,2	139,9	148,5	168,8	171,9	178,1	156,1	132,5
	d_k	+7,5	+11,7	+25,8	+6,4	+13,3	+3,6	-4,0	+5,4	+2,7	-4,2	+5,0	-1,3
	D_k	1684,6	1692,1	1703,8	1729,6	1736,0	1749,3	1752,9	1748,9	1754,3	1757,0	1752,8	1757,8
	24 $\bar{Y}_k$	3501,8	3503,2	3511,3	3509,8	3510,6	3514,3						
	$\bar{Y}_k$	145,7	145,9	146,3	145,9	146,0	146,3						
1939	D_k	1756,5											

Kot vidimo na prednjem primeru, se serija drsečih sredin ne začne takoj pri prvem in konča pri zadnjem členu osnovne serije, temveč je za nekaj členov krajša. Začne se namreč z drugo dekado januarja in konča že z drugo dekado marca. V splošnem se skrajša za polovico osnovnega razdobja.

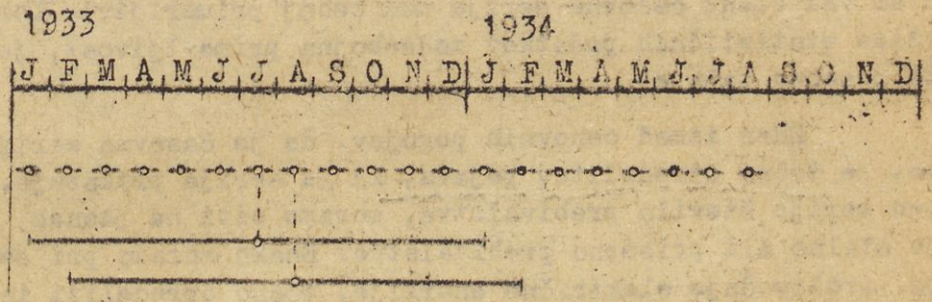
V prednjem primeru smo izračunali povprečje iz lihega števila členov. Ako izračunamo povprečje drseče serije sredin iz sodega števila členov, pa moramo uporabiti za računanje sredine obrazec

$$\bar{Y} = \frac{1}{n}\left(\frac{1}{2}Y_1 + Y_2 + Y_3 + \ldots + \frac{1}{2}Y_{n+1}\right)$$

Ker najpogosteje analiziramo mesečne časovne serije in ima leto sodo število mesecev, bomo izračunali kot primer serijo drsečih sredin za časovno serijo števila natovorjenih vagonov, ki smo jo že navedli v odstavku 91,4 v tabeli 9.9.

Kako izračunamo v teh primerih posamezne sredine, bomo videli iz slike 9,6

Sl. 9.6



$$\bar{Y}_{jul} = \frac{1}{12}\left(\frac{1}{2}Y_J + Y_F + Y_M + Y_A + Y_M + Y_J + Y_J + Y_A + Y_S + Y_O + Y_N + Y_D + \frac{1}{2}Y_J\right)$$

Ako pišemo ta obrazec v drugi obliki, dobimo:

$$\bar{Y}_{33,jul} = \frac{1}{12}\left(\frac{1}{2}Y_J + \frac{1}{2}Y_F + \frac{1}{2}Y_M + \ldots\ldots\ldots + \frac{1}{2}Y_N + \frac{1}{2}Y_D + \frac{1}{2}Y_J + \frac{1}{2}Y_F + \frac{1}{2}Y_M + \ldots\ldots\ldots + \frac{1}{2}Y_N + \frac{1}{2}Y_D + \frac{1}{2}Y_J\right)$$

Ako $\frac{1}{2}$ izpostavimo, vsoto letnih podatkov pa pišemo kot drsečo vrednost, je zgornji izraz enak :

$$\bar{Y}_{33,jul} = \frac{1}{24}\left(D_{34,jan} + D_{34,Feb}\right)$$

Povprečje, ki ga pripišemo mesecu juliju, je enako štirindvajsetemu delu vsote drsečih vsot za mesec januar in februar naslednjega leta.

Na isti način nadaljujemo in imamo:

$$\bar{Y}_{33.5} = \frac{1}{24}(D_{34.F} + D_{34.M}) \text{ itd.}$$

splošno $$\bar{Y}_k = \frac{1}{24}(D_{k+6} + D_{k+7}) \qquad (9.4)$$

P Po teh obrazcih je račun za naš primer prikazan v tabeli 9.13

92. ANALIZA ČASOVNIH SERIJ.

92.0 Splošno

Časovne serije dajo sliko dinamike pojava v razdobju, na katerega se časovna serija nanaša. V časovni seriji so v številčni obliki izraženi rezultati vseh vplivov, ki so med tem časom vplivali na masovni pojav. V nekaterih primerih so ti vplivi tako močni in take narave, da je vprašanje, ali so vsi členi časovne serije med seboj primerljivi. Ker je osnovno načelo analize statističnih podatkov medsebojna primerljivost, je temu treba posvetiti posebno pažnje.

Eden izmed osnovnih pogojev, da je časovna serija primerna za analizo, je točna opredelitev pojava, ki ga serija prikazuje. Ako imamo časovno serijo števila prebivalstva, moramo biti na jasnem, ali serija prikazuje stalno ali prisotno prebivalstvo. Enako moramo pri seriji, ki prikazuje n.pr. proizvodnjo električne energije, točno opredeliti in navesti, ali je to električna energija proizvedena v termo - ali hidrocentralah ali skupna. Važno je tudi, ali gre samo za proizvodnjo elektrike, ki je namenjena za uporabo izven podjetja, ki jo proizvaja, ali je všteta tudi energija, ki se uporabi znotraj samega podjetja.

Drug važen pogoj primerljivosti podatkov v časovni seriji je enakost časovnih intervalov . O važnosti enakosti intervalov med posameznimi členi časovne serije smo govorili že v enem izmed prejšnjih odstavkov. Dodamo še to, da se marsikatera metoda analiziranja časovnih serij da uporabiti le, ako je pojav prikazan s časovno serijo, katere členi si sledijo v enakih časovnih razdobjih.

Vendar sama enakost časovnih intervalov med posameznimi členi časovne serije še ni zadosten pogoj za primerljivost med členi. Ozirati se moramo pri tem tudi na morebitno periodičnost pojava, ki ga proučujemo. O periodičnosti pojavov bomo govorili še kasneje. Zaenkrat naj bo dovolj samo to, da se nekateri pojavi na določeno razdobje ponavljajo (n.pr. gradbena, statistična sezona itd.). Kadar imamo opravka s periodičnim pojavom, moramo vzeti člene vedno na enem in istem delu periode. Interval med dvema zaporednima členoma mora biti enak dolžini periode. Iz tega razloga se n.pr. vrše popisi živine v FLRJ vsako leto na isti datum (15. jan.). Razdobje med dvema popisoma je torej točno eno leto. Ako bi vršili popis živine vsakih 7 mesecev, bi na ta način sicer imeli med posameznimi členi časovne serije

enak razmak sedem mesecev, vendar bi dobili popolnoma pogrešno sliko, ker bi se en člen serije nanašal na januar, drugi na avgust, tretji na marec itd.

Dostikrat prekinejo primerljivost členov časovne serije dogodki, ki s samim pojavom, ki ga serija prikazuje, nimajo vsebinske zveze. Pri tem mislimo na teritorialne in administrativne spremembe. Število prebivalcev v nekem okraju se more v trenutku povečati ali zmanjšati, ako se z novo upravno spremembo spremeni teritorij okraja. Enako je tudi z drugimi podatki, s številom gospodarstev, številom živine itd. Ako je upravna sprememba nastala v času med dvema popisoma, podatki popisov pred to spremembo niso primerljivi s podatki popisov, ki so se vršili po tem popisu. Spremembe te vrste velikega obsega se vrše pri spremembah državnih mej v času vojn. V teh primerih je primerljivost predvojnih podatkov s povojnimi zelo omejena. Možnost analiziranja takih serij pa je omejena razen zaradi spremembe mej tudi zaradi sprememb, ki so nastale z vojno v socialnem, ekonomskem in političnem pogledu. Te vrste sprememb nastajajo tudi v primeru administrativnih ukrepov. Znani je primer, da je težka časovna primerljivost podatkov n.pr. republiške industrije, kjer so podjetja stalno prehajala iz ene kompetence v drugo. Spremembe v podatkih, ki s tem nastajajo, so velike, nastajajo pa zaradi administrativnih ukrepov. Da pa je kljub temu možno analizirati razvoj pojava s časovno serijo, se poslužujemo različnih metod:

a) Najboljša, čeprav najtežja, je rešitev, da preračunamo podatke za vse prejšnjo člene na novo stanje n.pr. število prebivalstva v LRS smo preračunali za predvojno dobo na ozemlje, ki ga danes obsega LRS. To delo ni samo zamudno, temveč včasih tudi nemogoče, ako pomislimo, da je bilo področje, ki ga danes obsega LRS v več državah, ki niso vršile popisov prebivalstva na isti dan in po istem načinu.

b) Drug način, ki je lažji, zaradi tega pa ne tako popoln, je naslednji: Za razdobje, v katerem je sprememba nastala, izračunamo podatke po starem in novem stanju. Na ta način dobimo za en člen dva podatka.

Podatek po starem stanju moremo primerjati s prejšnjimi členi, podatek po novem stanju pa z naslednjimi. Primerjava med samima podatkoma pa pokaže velikost spremembe zaradi ukrepa samega.

P Vzemimo kot primer število delavstva prvega v mesecu za neko stroko republiške industrije (podatki izmišljeni):

Tab. 9.14

Datum	Število delavcev
1.1	3648
1.2	3852
1.3	3760
1.4	3820, 8760
1.5	7752
1.6	6274

15. marca je prešlo v republiško industrijo večje število podjetij. Zaradi tega je za 1. april dano število delavstva za stari in novi obseg. 3820 moremo primerjati
1. marcem (3760) 8760 pa s 1. majem (7752). Primerjava med 8760 in 3820 pokaže, kako velika je sprememba,

ki je nastala zaradi prehoda podjetij.

Spremembe, ki nastanejo zaradi vzrokov, ki smo jih navedli v zgornjem odstavku, so običajno tako velike, da moramo serijo razdeliti v dva samostojna dela in analizirati vsak del časovne serije posebej.

92.1 Osnovni pokazatelji analize dinamike

Najosnovnejša analiza časovnih serij obstoji v tem, da opazujemo spreminjanje členov časovne serije. Pri tem se poslužujemo nekaj pokazateljev, ki opazovanje te osnovne dinamike olajšajo in poglobijo.

Vzemimo naš znan primer, časovno serijo števila natovorjenih vagonov v b. Jugoslaviji in sicer letno serijo mesečnih povprečij. Na njej bomo izračunali vse vrste osnovnih pokazateljev, ki podajajo spremembe pojava.

Dinamiko pojava moremo proučiti že iz serije absolutnih števil (Tab. 9.15, kolona 2).

Tab. 9.15

K	Leto	Štev. natovorjenih vagonov Y_k	$A_k =$	$I_{k/1} =$	P_k	V_k	$T_k = \frac{A_k}{Y_k}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1932	119,0	- 5,0	100,0	0,958	95,8	- 4,2 %
2	1933	114,0	+ 2,0	95,8	1,018	101,8	+ 1,8 "
3	1934	116,0	+ 5,5	97,4	1,047	104,7	+ 4,7 "
4	1935	121,5	+ 4,5	102,2	1,037	103,7	+ 3,7 "
5	1936	126,0	+14,4	105,9	1,114	111,4	+11,5 "
6	1937	140,4	+ 6,0	118,0	1,043	104,3	+ 4,3 "
7	1938	146,4		123.0			

Enostaven iz osnovne serije izračunan pokazatelj, ki podaja spremembo členov časovne serije, je absolutna razlika dveh zaporednih členov. Ako se pojav ne izpreminja, je absolutna razlika nič. Ako pojav narašča, je pozitivna, v primeru padanja pa je absolutna razlika negativna. Ako zaznamujemo splošen člen časovne serije z Y_k, je absolutna razlika, ki jo zaznamujemo z A_k enaka

$$A_k = Y_{k+1} - Y_k \qquad (9.5)$$

Za našo serijo je serija absolutnih razlik izračunana v stolpcu (3)! Padec je opazen samo od prvega leta na drugo ($A_1 = -5,0$) medtem ko opažamo v vseh naslednjih letih porast (absolutna razlika je večja od nič.)

Največji absoluten porast zaznamujemo od leta 1936 na leto 1937 ($A_5 = + 14,4$).

Spremembo glede na določen člen najlepše poda indeksna serija s stalno bazo. Za naš primer imamo izračunano indeksno serijo s stalno bazo 1932 v stolpcu (4).

Relativno spremembo od člena pokaže serija pokazateljev dinamike. Pokazatelj dinamike dobimo z deljenjem členov

$$P_1 = \frac{Y_2}{Y_1} = \frac{114,0}{119,0} = 0,958$$

ali splošno:

$$P_k = \frac{Y_{k+1}}{Y_k} \qquad (9.6)$$

Pokazatelji dinamike za naš primer so podani v stolpcu (5). Ako med členi ni izpremembe, je pokazatelj dinamike enak ena, padanje pomeni, ako je vrednost pokazatelja dinamike pod ena, porast pa, ako je njegova vrednost nad ena.

Prav tako pokaže relativno spremembo od člena do člena tudi serija verižnih indeksov, ki so s pokazatelji dinamike v tesni zvezi. Iz pokazatelja dinamike namreč dobimo verižni indeks, ako pokazatelj dinamike množimo s 100.

$$V_1 = \frac{Y_2}{Y_1} \cdot 100 = P_1 100 = 95,8$$

Splošno moremo pisati, da je:

$$V_k = \frac{Y_{k+1}}{Y_k} \cdot 100 \qquad (9.7)$$

Važen pokazatelj, ki podaja dinamiko časovnega razvoja je tudi temp razvoja. Temp je v procentih izražena relativna sprememba pojava od enega do drugega člena.

Ker je prva absolutna razlika $A_1 = - 5,0$, absolutna vrednost prvega člena Y_1 pa 119,0, je relativna razlika, izražena v procentih ali temp enak:

$$T_1 = \frac{A_1}{Y_1} \cdot 100 = \frac{-5,0}{119,0} \cdot 100 = -4,2\ \%$$

Na enak način moremo izračunati temp razvoja tudi za ostale člene. Splošno izračunamo temp razvoja po obrazcu:

$$T_k = \frac{Y_{k+1} - Y_k}{Y_k} \cdot 100 \qquad (9.8)$$

Enako je

$$T_k = \frac{A_k}{Y_k} \cdot 100 \qquad (9.9)$$

ali

$$T_k = V_k - 100 \qquad (9.10)$$

Temp razvoja je pozitiven, ako pojav narašča, in negativen, ako pada.

V naslednji razpredelnici bomo podali, kakšne vrednosti zavzema posamezen izmed pokazateljev, ako pojav pada, ako ostane na stalni ravni in ako narašča.

Tab. 9.16

Gibanje / Pokazatelj	padanje	stacionarnost	naraščanje
Absolutna diferenca	$A_k < 0$	$A_k = 0$	$A_k > 0$
Pokazatelj dinamike	$P_k < 1$	$P_k = 1$	$P_k > 1$
Verižni indeks (Koef. dinamike)	$V_k < 100$	$V_k = 100$	$V_k > 100$
Temp	$T_k < 0$	$T_k = 0$	$T_k > 0$

Navedeni pokazatelji se izračunavajo zelo pogosto pri enostavni analizi časovnih serij in jih s pridom uporabljamo tudi v operativni in planski evidenci.

93. VRSTE GIBANJ MASOVNIH POJAVOV

93.1 Proučitev

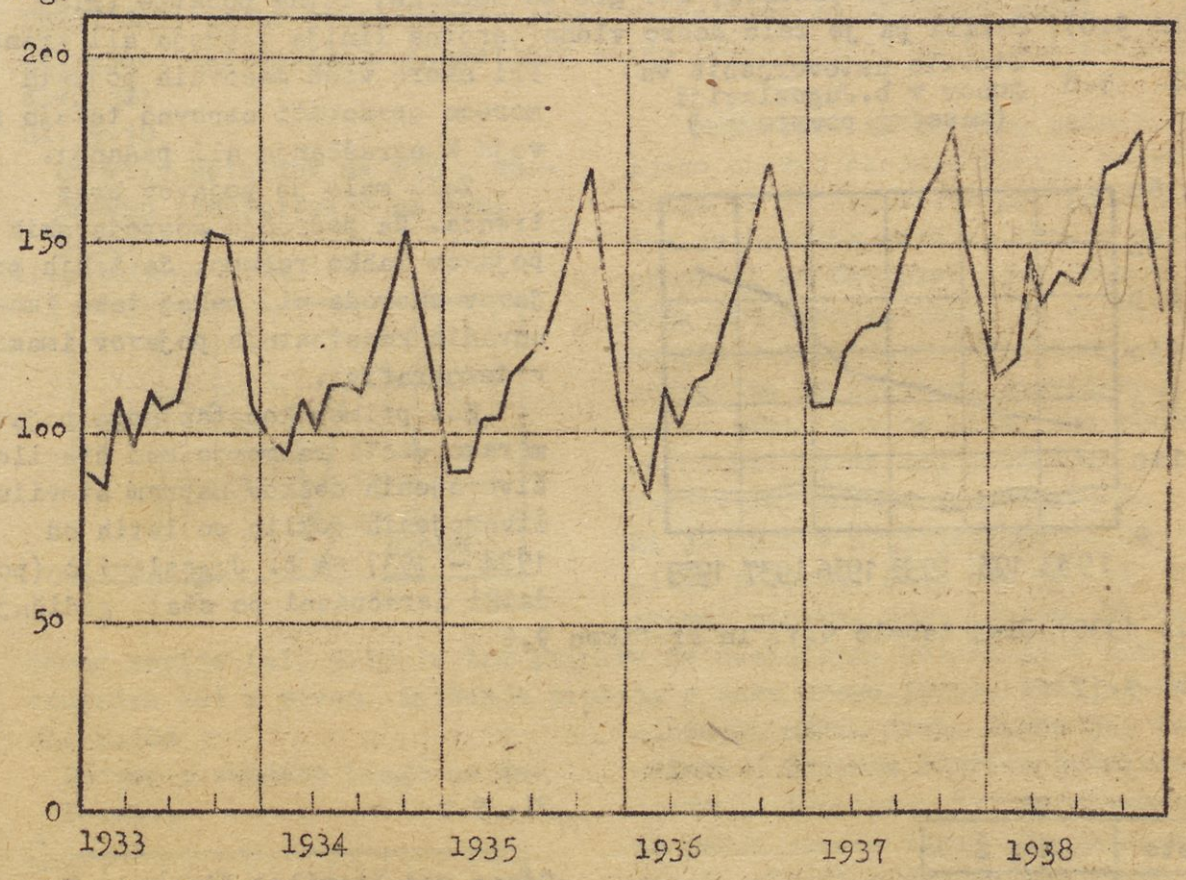
Proučevanje časovnih serij ni omejeno samo na izračunavanje gornjih pokazateljev. S posebnimi metodami, z osnovami katerih se bomo seznanili v naslednjih odstavkih, moremo prodreti znatno globlje, kot pa z njimi. Predno pa preidemo nanje, se moramo seznaniti z nekaj pojmi, ki bodo omogočili razumevanje teh načinov proučevanja. Poglejmo najprej nekaj primerov časovnih serij. Na osnovi njih bomo lažje razumeli, za kaj gre. Najboljše moremo časovno serijo predočiti z grafičnim prikazom. Podrobneje o grafičnem prikazovanju časovnih serij je govora v poglavju o časovnih serijah, vendar je grafikon časovne serije tudi brez tega razumljiv. Časovni grafikon je

prikazan v pravokotnem koordinantnem sistemu. Ena os služi kot časovna os, druga pa kot količinska os.

P Kot prvi primer vzemimo časovno serijo števila natovorjenih vagonov v b. Jugoslaviji v letih 1932 do 1938. (glej sliko 9.7).

Sl. 9.7

Število natovorjenih vagonov v b. Jugoslaviji v letih 1932 do 1938



Ako pozorneje premotrimo zgornjo sliko časovne serije, moremo napraviti nekatere koristne zaključke.

1. Število natovorjenih vagonov se iz leta v leto dviga. Osnovna tendenca razvoja je torej naraščanje. To naraščanje je rezultat splošnega ekonomskega razvoja v bivši Jugoslaviji. To osnovno tendenco razvoja imenujemo s tujim izrazom trend.

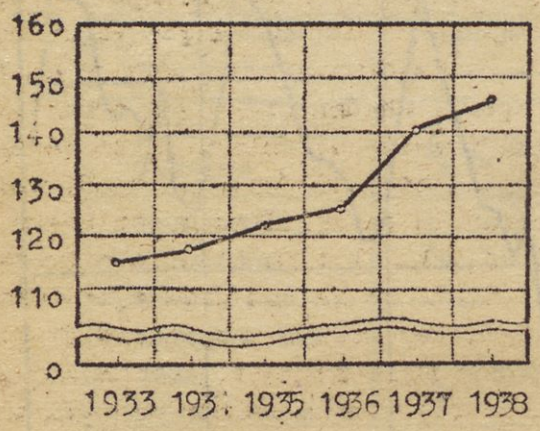
2. Poleg tega opazimo, da je število natovorjenih vagonov vsako leto največje v jesenskih mesecih (september, oktober), najmanjše pa v zimskih mesecih (januar, februar). Ta zakonitost se periodično ponavlja vsako leto. Vsako leto je najmanjši promet v mesecu januarju in februarju in največji v oktobru in novembru. Ta pojav nastane zaradi sezonskih razlik v gospodarstvu.

3. Vendar pa dejanska slika časovne serije vsako leto ni točno enaka, čeprav vsako leto periodični vplivi delujejo enako. Gibanja so si vsako leto le precej podobna. Ti odkloni nastanejo zaradi slučajnih činiteljev, ki se od meseca do meseca spreminjajo.

Že na tem samem primeru smo spoznali tri vrste vplivov, ki imajo za rezultat različna gibanja masovnih pojavov.

P Ako namesto mesečne serije število natovorjenih vagonov vzamemo letno serijo mesečnih povprečij, vidimo, da na tej seriji ni več razvidna periodičnost tovornega prometa, ker gre za sumarne, letne podatke (glej sliko 9.8). Ostala pa je zelo dobro vidna osnovna linija razvoja ali trend.

Sl. 9.8 Število natovorjenih vagonov v b. Jugoslaviji (mesečna povprečja)



Pri skoro vseh masovnih pojavih moremo opazovati osnovno težnjo razvoja k naraščanju ali padanju.

Zelo malo je pojavov brez trenda. Na področju gospodarskih pojavov lahko rečemo, da takih pojavov skoroda ni, nekaj tako imenovanih konstantnih pojavov imamo v demografiji.

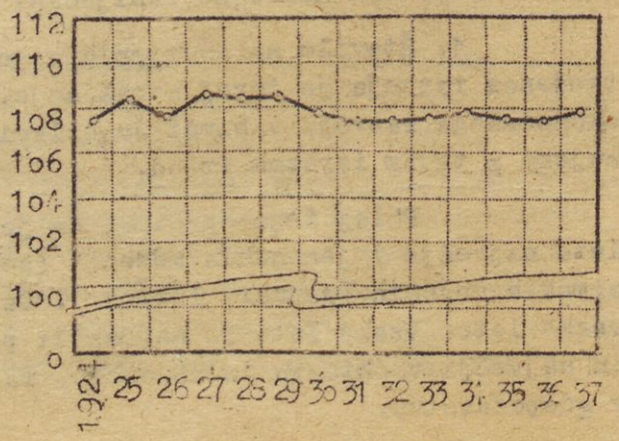
Kot primer konstantnega pojava moremo vzeti razmerje med številom živorojenih dečkov napram številu živorojenih deklic po letih od 1924 - 1937 za b. Jugoslavijo (podatki izračunani po stat. godišnjaku l. 1939) Glej tabelo 9.17 in grafikon 9.9

Tab. 9.17

Leto	M : Ž
1924	107.5
1925	108.3
1926	107.6
1927	108.7
1928	108.4
1929	108.4
1930	107.9
1931	107.2
1932	107.3
1933	107.6
1934	107.9
1935	107.4
1936	107.3
1937	107.9

Sl. 9.9

Odnos med številom živorojenih dečkov in deklic v b. Jugoslaviji



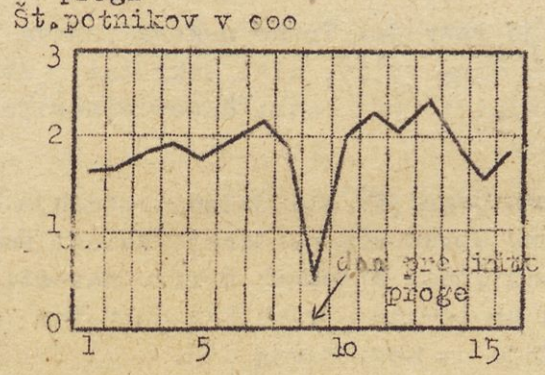
Vrednosti se gibljejo na istem nivoju in ni opazna niti težnja padanja, niti porast tega razmerja. Razliko med posameznimi leti izvirajo iz slučajnih vplivov.

Še eno vrsto vplivov, ki delujejo na časoven razvoj, poznamo.

P Mislimo si dnevno časovno serijo prevoženih potnikov ljubljanske ECŽ. Predpostavimo, da je na neki progi zaradi popravil ustavljen promet. Vprašanje je, na kakšen način se to odraža na časovni seriji. Jasno je, da se bo število skupno na vseh progah prevoženih potnikov za ta dan zmanjšalo. Drugi dan, ko bo proga popravljena, pa bo število potnikov zopet toliko, kot običajno. Ako narišemo shematično sliko (sl.9.10) bo za te dni slika naslednja:

P Sl. 9.10

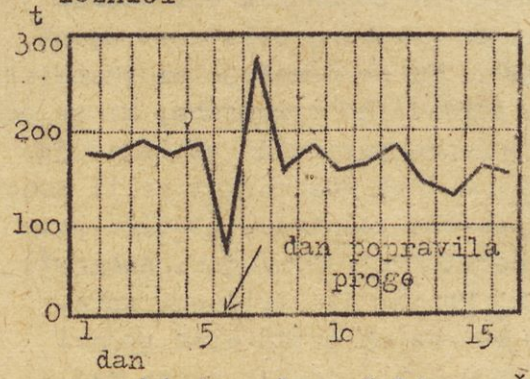
Število potnikov na ECŽ z enodnevno prekinitvijo na neki progi



Mislimo si pa drug primer. Sestavljamo dnevno časovno serijo prevoženega blaga po železnici. Zopet poglejmo, kakšna je časovna serija, ako je na eni progi za en dan prekinjen promet. Količina prevoženega blaga je za ta dan manjša kot običajno. Naslednji dan po popravilu pa količina prevoženega blaga ne bo taka, kot je bila, ako promet ne ne bi bil za en dan prekinjen, temveč večji, ker morajo prepeljati še blago, ki je ostalo neprepeljano na dan prekinitve prometa. Šele čez dan ali dva se promet ponovno normalizira. Slika časovne serije (sl. 9.10a) v tem primeru je naslednja. Učinek je v tem primeru drugačen kot v prvem. Znižanje prometa v enem dnevu izzove zvišanje v naslednjih nekaj dneh. Jasno je, da ta učinek v primeru števila potnikov na tramvaju izostane, ker se noben potnik ne pelje zaradi tega, ako se en dan ni mogel peljati s tramvajem, drugi dan dvakrat, medtem ko mora biti blago prepeljano drugi dan, ako ni moglo biti prepeljano na dan popravila.

Sl. 9.10a

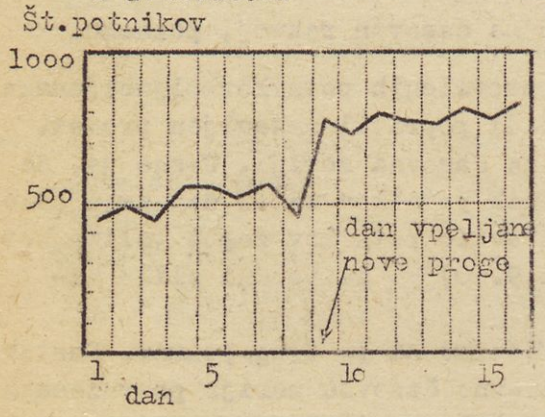
Dnevno prevoženo blago na železnici



P Kot tretji primer iz prometne statistike vzemimo dnevno število prevoženih potnikov z avtobusi v LRS. Vzemimo, da nekega dne vpeljejo novo avtobusno progo, ki je zelo zaželjena. Število prevoženih potnikov se iz enega dne na drugi dan nenadoma poveča in ostane na novem nivoju. Glej sliko 9.11.

Sl. 9.11

Število prevoženih potnikov z avtobusom.



Na teh treh primerih smo videli, da je enkraten vpliv (prekinitev proge, uvedba nove proge) prekinil normalni potek časovne serije. Učinek je v vsakem primeru drugačen. V prvem primeru imamo trenotno spremembo z reakcijo, v tretjem primeru pa je sprememba stalnega značaja.

Na ta način smo spoznali vse vrste časovnega gibanja masovnih pojavov.

93.2 Glavne vrste gibanj so torej:

1. Trend, ki podaja osnovno linijo razvoja. Trend nekega pojava bi dobili, ako na pojav ne bi deloval nikak drug vpliv, niti enkratni, niti periodični ali slučajni. Trend teče med realnimi vrednostmi časovne serije in ima enakomeren tok naraščanja ali padanja.

2. Enkratne spremembe, ki za moment, krajše ali daljše razdobje prekinejo redni tok časovne serije. Enkratne spremembe nastanejo zaradi enkratnih vplivov. Poleg zgornjih primerov enkratnih sprememb moremo navesti še dosti drugih. Stavka delavcev vpliva na proizvodnjo v obliki enkratne spremembe, nov izum more dvigniti proizvodnjo na novo raven itd.

3. Periodično gibanje povzročajo vplivi, ki se na določeno stalno razdobje ponavljajo. Periodični vplivi so zelo pogosti in se pojavljajo v veliki večini masovnih pojavov. Periodičnosti, ki se ponavljajo vsako leto, so pri ekonomskih proučevanjih najpogostejše in najvažnejše. V primeru letnih periodičnosti govorimo o sezonskih spremembah. Sezonskim spremembam je podvrženo število gradbenega delavstva, oziroma gradbena delavnost sploh, vsi pojavi s področja tujskega prometa, pojavi, ki se nanašajo na transport itd. Poleg sezonskih sprememb pa imamo tudi periodične spremembe, ki se ponavljajo na druga razdobja, ne pa na leto. Mesečno periodičnost izkazuje n.pr. blagovni promet, ki je živahnejši v prvih dneh meseca kot proti koncu. Tedensko periodičnost moremo opazovati pri prevozu potnikov. Število potnikov je največje v soboto, nedeljo in v ponedeljek. Dnevne periodičnosti moremo zaslediti pri najrazličnejših pojavih. Poraba električnega toka, plina in vode ima dnevno periodičnost. Vsak dan se ob istih urah pojavi največja in najmanjša poraba. Pri tem so za preskrbo posebno važne največje obremenitve, tako imenovane "konice". Enako je periodična dnevna obremenitev lokalnega tramvajskega ali avtobusnega prometa. Največje število potnikov je v urah, ko ljudje prihajajo in odhajajo v službo.

4. Slučajne spremembe. Slučajne spremembe nastanejo zaradi velikega števila individualnih vzrokov, ki se ne dajo predvideti vnaprej. Za posledico imajo odklone od serije, ki bi jo dobili, ako bi delovali na pojav samo trend, enkratni in periodični vplivi. Vendar slučajni vplivi niso tako izraziti, da iz serije ne bi bila razvidna trend in periodična nihanja.

94. RAZSTAVLJANJE ČASOVNIH SERIJ NA SESTAVNE DELE

94.1 Splošno.

Analiza časovne serije obstoji v tem, da časovno serijo razstavimo na zgornje sestavne dele: trend, rezultat enkratnih vplivov, periodičnih nihanj in slučajnih vzrokov. To razstavljanje moremo s posebnimi metodami dejansko izvršiti. Misliti si moremo, da je vrednost vsakega posameznega člena časovne serije sestavljena iz vrednosti trenda, enkratne spremembe, periodične spremembe in slučajnih sprememb.

$$Y_k = T_k + E_k + P_k + S_k \qquad (9.11)$$

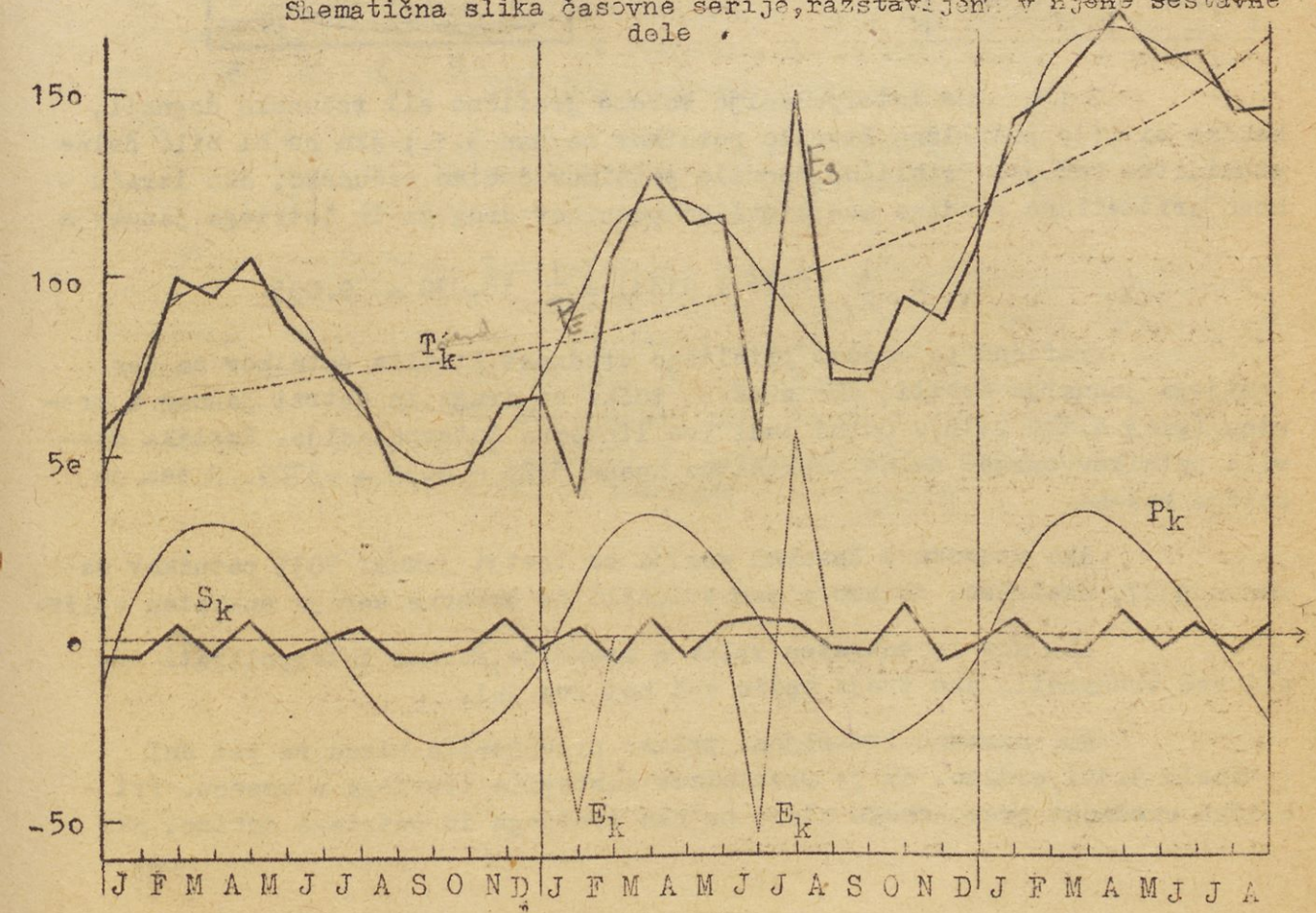
T_k = trend E_k = enkratna sprememba

P_k = periodična sprememba S_k = slučajna sprememba

Naloga analize je, določiti, oziroma razdružiti posamezne sestavne dele.

Sl. 9.12

Shematična slika časovne serije, razstavljene v njene sestavne dele.



V sliki je z debelo črto narisana dejanska časovna serija, s tenko pa njeni sestavni deli. S pikčastimi črtami so naznačeni posamezni sestavni deli dejanske linije.

94.2 Določanje in izločanje enkratnih sprememb iz časovne serije.

Ako imamo časovno serijo, ki je spremenila svoj redni tok zaradi enkratne spremembe, moremo na zelo enostaven način določiti, kolik je učinek te enkratne spremembe na časovno serijo.

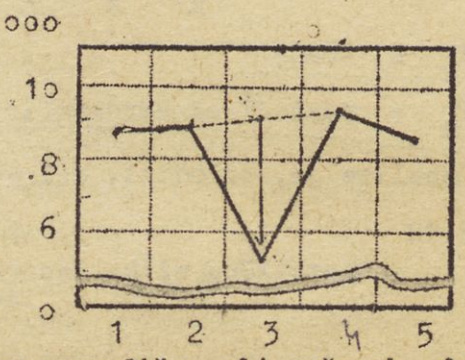
P Vzemimo n.pr., da ima dnevna serija prevoženih tramvajskih potnikov za nekaj dni vrednosti, kot kaže tabela 9.18.

Iz vrednosti serije, še bolj pa iz slike 9.13, je razvidno, da je enkraten vpliv nastopil 3.1

Tab. 9.18

Datum	Število potnikov
1.1.	8762
2.1.	8926
3.1.	5326
4.1.	9143
5.1.	8.627

Sl. 9.13 Slika serije 9.18



Z linearno interpolacijo moremo grafično ali računsko dognati, kakšno bi bilo približno število potnikov na dan 3.1., ako ne bi bilo delne prekinitve vožnje. Približno število potnikov dobimo računsko, ako izračunamo aritmetično sredino med številom potnikov drugega in četrtega januarja

$$Y'_3 = \frac{1}{2}\ (8926 + 9144) = \frac{1}{2}\ 18.070 = 9.035.$$

Grafično pa moremo približno vrednost števila potnikov na dan tretjega januarja dobiti, ako zvežemo točki za drugi in četrti januar s premico (glej sliko 9.13). Odtod tudi ime linearna interpolacija. Razlika števila potnikov zaradi delne prekinitve znaša 5326 - 9035 = -3709. S tem je naloga rešena.

Ako vzamemo v časovni seriji za tretji januar 9035 potnikov namesto 5327, smatramo, da smo s tem izločili iz časovne serije enkraten vpliv.

Ako gre za enkraten vpliv z reakcijo, moramo interpolirati dve ali več vrednosti, ako traja vpliv več kot dva dni.

P Ako vzamemo izmišljeni primer prevoženega blaga za pet dni (tabela 9.19), vidimo, da je prekinitev nastopila tretjega v mesecu. Približna vrednost prevoženega blaga na dan tretjega in četrtega dobimo, ako

vzamemo aritmetično sredino in vrednosti za tretji in četrti februar.

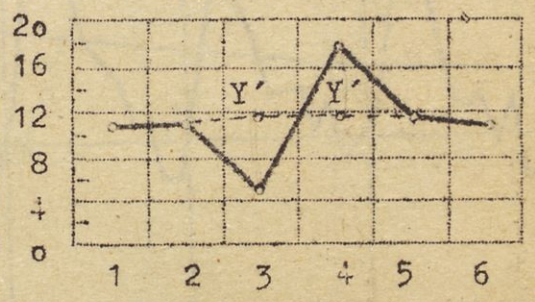
$\frac{1}{2}$ ⅀ (4762 + 17540) = $\frac{1}{2}$ 22302 = 11151. To količino smatramo za približek prevoženega blaga brez prekinitve na dan tretjega in četrtega.

Tab. 9.19

Datum	Prevoženo v t dejansko	Interpolirano
1.1	10.500	10.500
2.2	10.863	10.863
3.2	4.762	11.151
4.2	17.540	11.151
5.2	11.523	11.523
6.2	10.630	10.630

Sl. 9.14

Slika serije 9.19



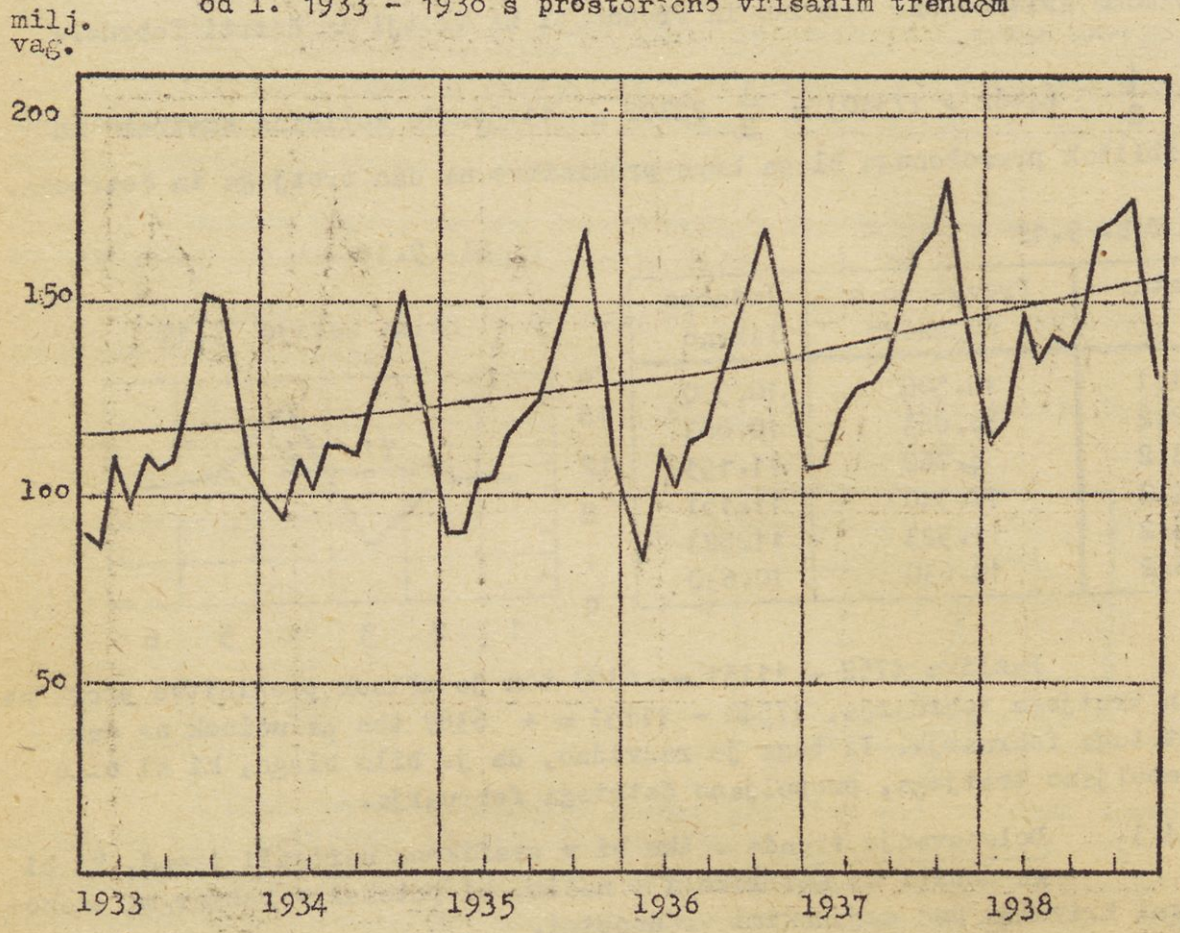
Razlika 4762 - 11151 = - 6389 ton je učinek prekinitve proge na dan tretjega februarja, 17540 - 11151 = + 6389 ton pa učinek na dan četrtega februarja. Iz tega je razvidno, da je bilo blago, ki ni bilo prepeljano tretjega, prepeljano četrtega februarja.

94.3 Določevanje trenda . Ako bi v grafikonu narisali trend, ki bi ga dobili na katerikoli način, bi potekal v razmerjema enakomerni krivulji med dejanskimi vrednostmi.

.31 Prostoročno. Bolj ali manj točno moramo včrtavati v grafikon trend tako, da narišemo med dejansko črto časovne serije prostoročno linijo, ki po našem mnenju najbolj odgovarja osnovni liniji razvoja. Čeprav ta način ni najboljši, ker je natančen in subjektiven, ga zaradi njegove enostavnosti pogosto uporabljamo, če ne drugače, vsaj predhodno kot osnovo za natančnejše načine. Po nekaj primerih moremo dobiti v prostoročnem risanju trenda že precejšnjo izvežbanost.

Sl. 9.15

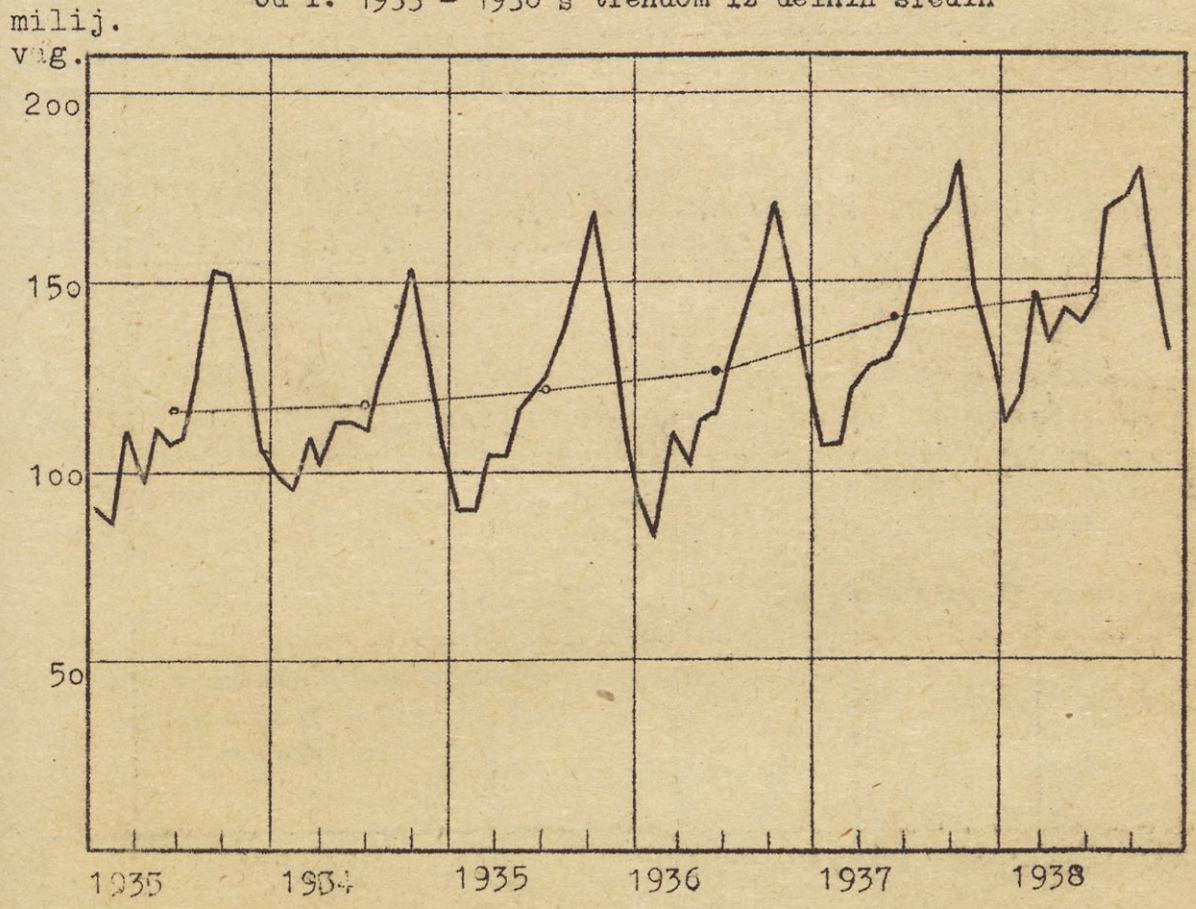
Serija števila natovorjenih vagonov v b. Jugoslaviji od l. 1933 - 1938 s prostoročno vrisanim trendom



.32 Določanje trenda s pomočjo dolnih sredin. Kot smo videli na primeru števila natovorjenih vagonov v b. Jugoslaviji, je mesečna časovna serija vsebovala naslednje sestavne dele: trend, sezonske spremembe in slučajne spremembe (glej sliko 9.7). Ako iz te serije izračunamo letno serijo mesečnih povprečij, vidimo, da ta serija ne vsebuje periodičnega dela, niti slučajnih odklonov, ker se obeji v povprečju ene periode uničijo (slika 9.8). Serijo sredin moremo zaradi tega smatrati kot trend, ker smo osnovno serijo s tem, da smo tvorili sredine, "očistili" sezonskih in slučajnih sprememb. Ako obe serij narišemo na eno samo sliko, vidimo, da serija sredin ustreza pogojem, katerim mora zadoščati trend, da teče umerjeno med osnovnimi vrednostmi.

Sl. 9.16

Število natovorjenih vagonov v b. Jugoslaviji od l. 1933 - 1938 s trendom iz delnih sredin



Opozoriti moramo, da serija sredin predstavlja trend le, ako tvorimo sredino iz členov enega leta, v splošnem iz členov ene periode.

Trend moremo s pomočjo sredin določiti tudi za časovne serije, ki nimajo periodičnega dela.

P Vzemimo kot primer število prebivalstva v Združenih državah Amerike v letih 1937 do 1948. Podatki se nanašajo na sredino leta. Trend izračunamo v tem primeru po naslednjem postopku:

1. Serijo razdelimo na dva dela. V vsakem delu naj bo enako število členov. V našem primeru so v enem delu podatki od leta 1937 - 1942, v drugem pa podatki od leta 1943 do 1948.

2. Izračunamo aritmetično sredino iz členov prve in druge polovice serije. Aritmetična sredina prve polovice je $Y_1 = \frac{1}{6}\ 789{,}4 = 131{,}5$

druge pa $Y_2 = \frac{1}{6}\ 846{,}0 = 141{,}0$.

Tab. 9.21

K	Leto	Število prebivalstva Y_k	Y_k	Trend Y'
1	1937	128,8		127.7
2	1938	129,8		129,2
3	1939	130,9		130,8
4	1940	132,0		132,4
5	1941	133,2		133,9
6	1942	134,7	789,4	135,5
7	1943	136,5	846,0	137,1
8	1944	138,1		138,7
9	1945	139,6		140,2
10	1946	141,2		141,8
11	1947	144,0		143,4
12	1948	146,6		144,9

3. Grafično narišemo trend tako, da $\bar{Y}_1$ nanesemo nad koncem leta 1939. $\bar{Y}_2$ pa nad koncem leta 1945. Premico, ki veže ti dve točki, vzamemo kot trend.

Enačbo te premice, ki predstavlja trend pa izračunamo po obrazcu:

$$Y = \frac{\bar{Y}_2 - \bar{Y}_1}{\frac{n}{2}} \left(t - \frac{n+1}{2} \right) + \frac{\bar{Y}_1 + \bar{Y}_2}{2} \qquad (9.12)$$

$\bar{Y}_1 = 131,6$

$\bar{Y}_2 = 141,0$

V našem primeru je enačba trenda enaka:

$$Y = \frac{141,0 - 131,6}{6} \left(t - \frac{13}{2} \right) + \frac{131,6 + 141,0}{2}$$

$$Y = \frac{9,4}{6} (t - 6,5) + 136,3$$

$$Y = 1,57 (t - 6,5) + 136,3$$

$$Y = 1,57\ t + 126,1$$

Trend za prvi člen ($t = 1$) je:

$$Y' = 126,1 + 1,57 \cdot 1 = 127,67$$

Za naslednje člene dobimo trend, ako postopoma prištevamo srednji prirast 1,57.

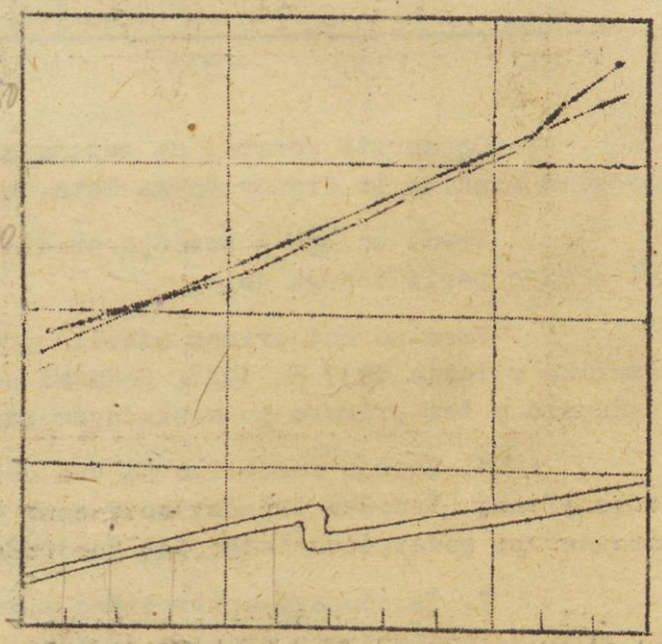
$$Y'_2 = 127,67 + 1,57 = 129,24$$

$$Y'_3 = 129,24 + 1,57 = 130,81$$

$$Y'_4 = 130,81 + 1,57 = 132,38$$

Sl. 9.16

Število prebivalstva v ZDA od l. 1937 - 1948 z vrisanim trendom

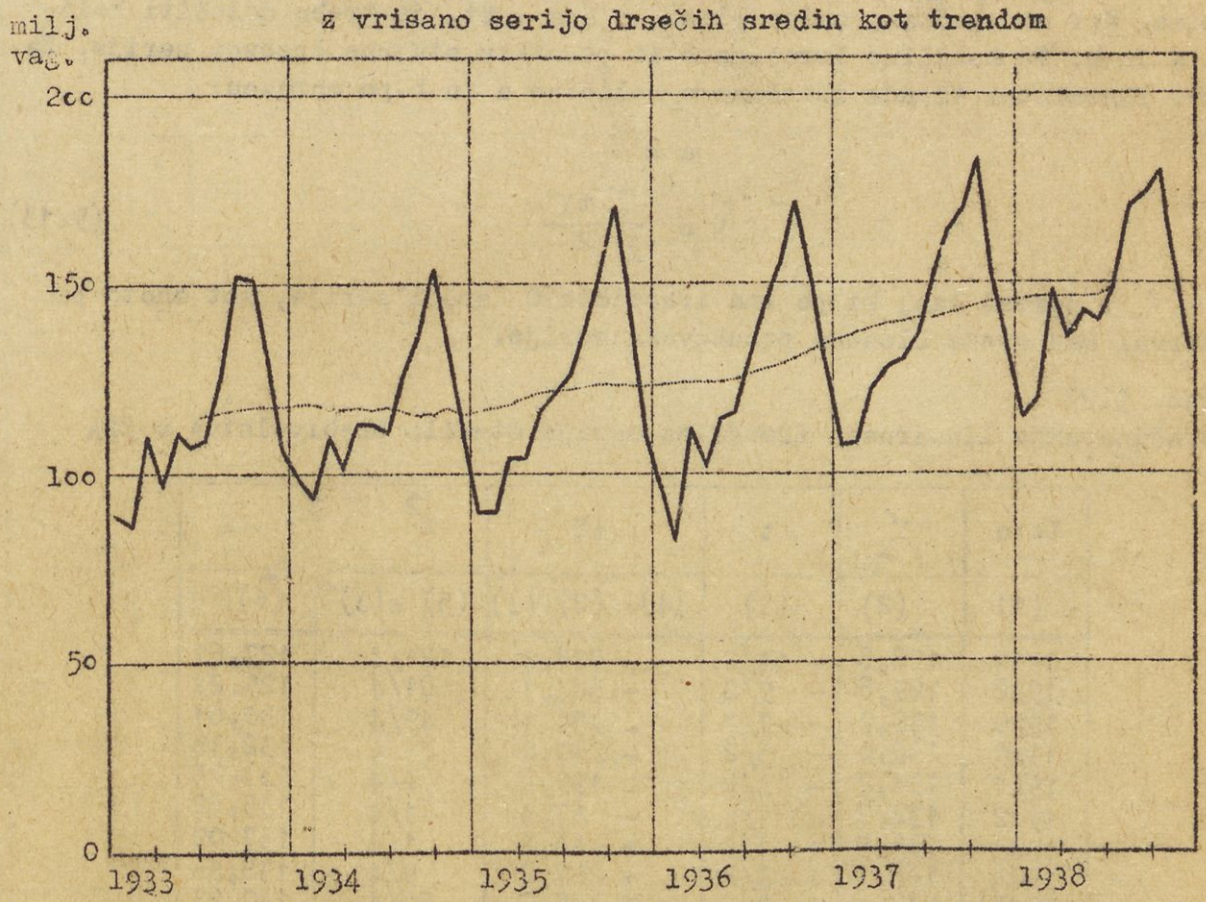


.33 Določanje trenda s pomočjo drsečih sredin

V primeru določanja trenda s pomočjo delnih sredin smo pri časovni seriji števila natovorjenih vagonov v b. Jugoslaviji za vsako leto dobili le eno samo vrednost trenda (letno povprečje). Črto trenda smo dobili grafično na ta način, da smo te vrednosti med seboj zvezali z daljicami. Boljše pa ponazarja trend serija drsečih sredin, ker ima serija drsečih sredin vrednost za vsak člen časovne serije, razen prvih in zadnjih. Vrednost trenda je torej določena za vsak mesec. Ker smo časovno serijo drsečih sredin že izračunali, jo bomo podali samo grafično, obenem z osnovno časovno serijo.

Sl. 9.17

Število natovorjenih vagonov v b. Jugoslaviji z vrisano serijo drsečih sredin kot trendom



Seveda velja isto kot za delne sredine, da moramo pri seriji drsečih sredin vzeti za izračunavanje sredine razdobje ene periode.

.34 Analitično določanje trenda

Po metodi najmanjših kvadratov moremo analitično določiti funkcijo, ki predstavlja krivuljo, ki se dejanski časovni seriji najbolše prilega.

Način se imenuje metoda najmanjših kvadratov zaradi tega, ker smatramo za najugodnejšo krivuljo tisto, za katero je vsota kvadratov odklonov dejanskih vrednosti od krivulje najmanjša. Pri analitičnem določevanju trenda moramo deliti delo v dve stopnji.

Najprej je treba ugotoviti, kateri tip funkcije najbolje odgovarja danemu primeru. Običajno vzamemo enostavne funkcije n.pr. premico $y = a+bt$, ali parabolo druge, tretje ali četrte stopnje: $y = a + bt + ct^2$ itd. Včasih iščemo funkcijo tudi v obliki eksponencialnne funkcije $y = ab^x$.
Pri izbiri tipa funkcije je najugodneje, ako si pomagamo s tem, da trend najprej narišemo prostoročno, po tej obliki pa se še odločimo, katera vrsta funkcije je najprimernejša.

Ko imamo izbrano vrsto funkcije, pa je treba določiti še vrednosti količin a, b, c itd. tako, da se bo dana funkcija najbolje prilegala dejanskim vrednostim.

Najenostavnejša funkcija, ki jo uporabljamo kot trend je vsekakor premica. Ker se enačba premice glasi: $y = a + bt$, je treba določiti količine a in b. Te količine izračunamo iz podatkov stvarne časovne serije. Za primer linearnega trenda izračunamo količine a in b po obrazcu:

$$a = \bar{Y}$$

$$b = \frac{\sum tY}{\sum t^2} \qquad (9.13)$$

t pomeni čas, ki pa ima izhodišče 0 sredi serije, kot enoto pa interval med dvema členoma proučevane serije.

Tab. 9.22
Izračunavanje linearnega trenda za serijo števila prebivalstva v ZDA

Leto	Y_k	t	tY	t^2	Trend y'_k
(1)	(2)	(3)	(4)= (2).(3)	(5) =$(3)^2$	(6)
1937	128,8	- 11/2	- 708,4	121/4	127,67
1938	129,8	- 9/2	- 584,1	81/4	129.24
1939	130,9	- 7/2	- 458 1	49/4	130,81
1940	132,0	- 5/2	- 330,0	25/4	132,38
1941	133,2	- 3/2	- 199,8	9/4	133,95
1942	134,7	- 1/2	- 67,4	1/4	135,52
1943	136,5	+ 1/2	+ 68,2	1/4	137.09
1944	138,1	+ 3/2	+ 207,2	9/4	138,66
1945	139,6	+ 5/2	+ 349,0	25/4	140.23
1946	141.2	+ 7/2	+ 494,2	49/4	141.80
1947	144,0	+ 9/2	+ 648,0	81/4	143.37
1948	146,6	+ 11/2	+ 806,3	121/4	144.94
$\sum$	1635,4	0	+ 225,1	572/4 = =143	
$\frac{1}{12}\sum$	136,3	0	X	X	

$y = ax + b$ $\quad a = \frac{(2)}{12} = \frac{1635,4}{12} = 136,3$

Tab.9.23 Izločitev trenda (serije drsečih sredin 12. serije števila natovorjenih vagonov v milij.v b.Jugoslaviji v letih 1933-1938

		J	F	M	A	M	J	J	A	S	O	N	D
1933	Y_{1i}	89,0	86,6	109,7	97,1	110,8	106,7	109,5	127,6	152,6	152,2	130,6	106,0
	$\bar{Y}_{1i}$							115,3	116,0	116,4	116,5	116,8	117,2
	Y''_{1i}							-5,8	+11,6	+36,2	+35,7	+13,8	-11,2
1934	Y_{2i}	98,9	94,4	110,1	101,1	113,3	113,2	111,3	125,9	136,7	153,4	132,5	110,8
	$\bar{Y}_{2i}$	117,5	117,6	116,8	116,2	116,4	116,6	116,5	116,0	115,5	115,3	116,6	115,6
	Y''_{2i}	-18,6	-23,2	-6,7	-15,1	- 3,1	- 3,4	- 5,2	+ 9,9	+21,2	+38,1	+15,9	-4,8
1935	Y_{3i}	90,5	90,0	104,3	103,8	116,0	110,5	125,7	137,4	152,7	169,9	[illegible]	110,4
	$\bar{Y}_{3i}$	116,1	117,3	118,3	119,7	121,1	121,5	121,8	122,2	122,6	122,8	122,7	122,8
	Y''_{3i}	-25,6	-27,3	-14,0	-15,9	-5,1	-11,0	+3,9	+15,2	+30,1	+47,1	+24,7	-12,4
1936	Y_{4i}	96,6	92,5	111,6	102,4	114,0	116,1	129,5	145,0	155,0	171,0	150,8	129,4
	$\bar{Y}_{4i}$	123,2	123,6	124,0	124,3	124,4	125,2	126,5	127,6	128,7	130,2	132,0	133,4
	Y''_{4i}	-26,6	-31,1	-12,4	-21,9	-20,4	- 9,1	+ 3,0	+17,4	+26,3	+40,8	+18,8	- 4,0
1937	Y_{5i}	107,1	108,4	122,2	128,4	129,9	136,3	152,5	163,4	169,2	182,3	151,1	135,8
	$\bar{Y}_{5i}$	135,4	137,1	138,4	139,4	139,9	140,1	140,6	141,4	142,9	144,2	145,2	145,7
	Y''_{5i}	-28,3	-28,7	-16,2	-11,0	-10,0	-3,8	+11,9	+22,0	+26,3	+38,1	+5,9	-11,9
1938	Y_{6i}	114,6	120,1	148,0	134,8	143,2	139,9	148,5	168,8	171,9	178,1	156,1	132,5
	$\bar{Y}_{6i}$	145,7	145,9	146,3	145,9	146,0	146,3						
	Y''_{6i}	-31,1	-25,8	+ 1,7	-11,1	- 2,8	-16,4						

P Ako vzamemo kot primer serijo števila prebivalstva v ZDA v letih od 1937 do 1948 je postopek izračuna naslednji (Tab.9.22)

1. Napišemo osnovno časovno serijo Y_k (kolona 2)

2. Točno sredi serije izberemo izhodišče koordinatnega sistema za t. K vsakemu členu pripišemo vrednost t. V našem primeru je izhodišče koncem leta 1942.

3. Izračunamo serijo produktov iz druge in tretje kolone in vpišemo rezultate v kolono (4).

4. Serijo kvadratov t-ja (3) vpišemo v kolono (5).

5. Seštejemo kolone (2)(4) in (5.)

6. Ako vsoto kolone (2) delimo z 12 ali splošno s številom členov časovne serije, dobimo vrednost količine a. $a = \frac{1}{12}\ 1635,4 = 136,3$.

7. Ako vsoto kolone (4) 225,1 delimo z vsoto kolone (5) 143, dobimo vrednost količine b.

$$b = \frac{225,1}{143} = 1,57$$

$y = 136,3 + 1,57t$

8. K posameznim členom odgovarjajočo vrednost trenda y izračunamo tako, da v enačbo premice vnesemo odgovarjajoče vrednosti t-ja. V našem primeru dobimo vrednost trenda za prvi člen serije, ako v enačbo $y = 136,3 + 1,57\ t$ vstavimo za t vrednost - 11/2.

$$Y_1 = 136,3 - 1,57 \cdot 11/2 = 136,3 - 8,63 = 127,67$$

Vrednost trenda za drugi člen dobimo, ako vrednost trenda za prvi člen prištejemo srednji prirast 1,57. Na enak način dobimo tudi ostale člene trenda.

$$Y_2 = Y_1 + 1,57 = 127,67 + 1,57 = 129,24$$
$$Y_3 = Y_2 + 1,57 = 129,24 + 1,57 = 130,81$$
$$Y_4 = Y_3 + 1,57 = 130,81 + 1,57 = 132,38$$

Serija trenda je vpisana v koloni (6). Ako primerjamo trend, ki smo ga izračunali s pomočjo metode najmanjših kvadratov s trendom, ki smo ga izračunali po metodi delnih sredin, vidimo, da se v našem primeru slučajno oba trenda popolnoma skladata. To sicer ni pravilo, vendar razlike nikdar niso znatne.

.35 Izločitev trenda iz časovne serije. Iz časovne serije moremo izločiti trend na enak način, kot smo izločili enkratne vplive - z odštetjem od osnovne serije.

P Kot primer izločimo na ta način trend iz časovne serije prevoženih

tovornih vagonov v b. Jugoslaviji. Kot trend uporabimo serijo drsečih sredin, ki jo imamo že izračunano. Račun je izveden v tabeli 9.23.

Po izločitvi trenda ostane časovna serija, ki vsebuje še sezonske in slučajne vplive.

94.3 Določitev jakosti periodičnih vplivov v časovni seriji

Ko smo iz časovne serije izločili učinek enkratnih vplivov in trend, dobimo serijo, ki vsebuje še rezultate periodičnih in slučajnih vplivov. Naša naloga je, da razdružimo še ta dva, oziroma, da izluščimo iz nje periodični sestavni del.

Ako napišemo po izločitvi trenda preostalo časovno serijo prevoženih natovorjenih vagonov v b. Jugoslaviji, more služiti ta serija kot osnova za izračunavanje sezonskega vpliva.

Tab. 9.[illegible]

Izračunavanje sezonskega vpliva za serijo natovorjenih vagonov v b. Jugoslaviji

Leto / mesec	1933	1934	1935	1936	1937	1938	(2)+(3)+(4)+(5)+(6)+(7)	$\frac{1}{5}$ (8)	P_i = (9)-p
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
J		-18,6	-25,6	-26,6	-28,3	-31,1	-130,2	-26,0	-25,8
F		-23,2	-27,3	-31,1	-28,7	-25,8	- 136,1	-27,2	-26,9
M		- 6,7	-14,0	-12,4	-16,2	+ 1,7	- 47,6	- 9,5	- 9,3
A		-15,1	-15,9	-21,9	-11,0	-11,1	- 75,0	-15,0	-14,7
M		- 3,1	- 5,1	-20,4	-10,0	- 2,8	- 41,4	- 8,3	- 8,1
J		- 3,4	-11,0	- 9,1	- 3,8	-16,4	- 43,7	- 8,7	- 8,4
J	-5,8	- 5,2	+ 3,9	+ 3,0	+11,9		+ 7,8	+ 1,6	+ 1,8
A	+11,6	+ 9,9	+15,2	+17,4	+22,0		+ 76,1	+15,2	+15,5
S	+36,2	+21,2	+30,1	+26,3	+26,3		+ 140,1	+28,0	+28,2
O	+35,7	+38,1	+47,1	+40,8	+38,0		+ 199,8	+40,0	+40,3
N	+13,8	+15,9	+24,7	+18,8	+ 5,9		+ 79,1	+15,8	+16,0
D	-11,2	- 4,8	-12,4	- 4,0	-11,9		- 44,3	- 8,9	- 8,6
							$\sum$	- 3,0	0
							$p = \frac{3,0}{12} =$	0,25	

Iz razpredelnice vidimo, da se serija, ki smo jo dobili z izločitvijo trenda, začenja šele z julijem 1933 in konča že z junijem 1938. Serija je torej v celoti za eno leto krajša kot osnovna serija. To skrčenje dolžine serije gre na račun tega, ker je tudi serija drsečih sredin, ki smo jo uporabili kot trend, za eno leto krajša.

Ker se sezonski vplivi vsako leto ponavljajo, je učinek sezonskega vpliva za isti mesec vsako leto enak. To se odkriva do neke mere tudi na naši seriji. Ako pogledamo člene zgornje serije in primerjamo med seboj podatke za mesec januar za vseh pet let, vidimo, da se ti podatki res gibljejo okrog neke stalne vrednosti. Ta nivo pa se od meseca do meseca menja. Iz razpredelnice je razvidno, da se gibljejo vrednosti v prvi polovici leta okrog negativnih vrednosti, v drugi polovici leta pa okrog pozitivnih vrednosti.

Vrednosti členov serije je za posamezne mesece med leti pa med seboj niso čisto enake, ker so ti podatki podvrženi tudi slučajnim vplivom. Ker pa se slučajni vplivi v povprečju uničijo, moremo smatrati, da predstavlja povprečje iz januarskih podatkov za vsa leta rezultat sezonskih vplivov za januar. Enako predstavlja povprečje za mesec februar rezultat sezonskih vplivov za februar itd. Na ta način dobimo serijo sestavljeno iz dvanajstih členov, ki predstavlja rezultat vplivov sezonskega značaja.

Ker pa se mora učinek periodičnih vplivov v povprečju ene periode (leta) uničiti, moramo posamezne člene sezonske serije popraviti še s tem, da od vsakega člena odštejemo povprečje iz kolone (9). Popravek, ki ga zaznamujemo s p, je enak

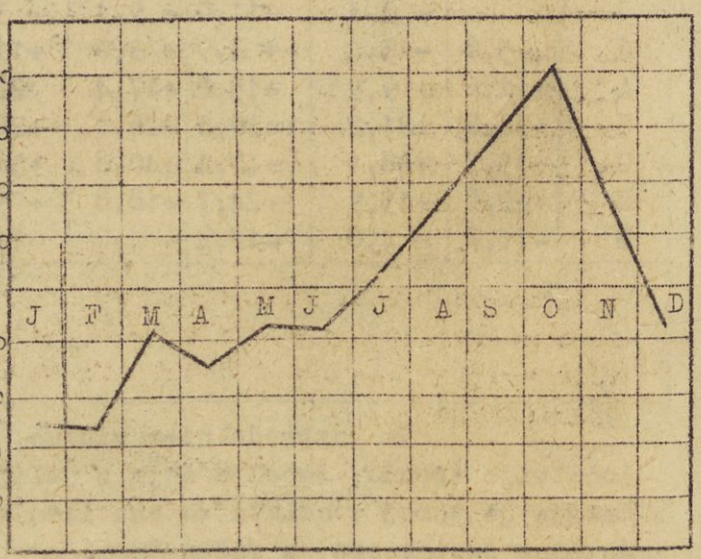
$$p = \frac{-0,3}{12} = -0,25.$$

Popravljeni podatki sezonskega sestavnega dela so vpisani v koloni (10).

Ako sezonske vplive narišemo v grafikonu, dobimo sliko:

Sl. 9.18

Sezonski sestavni del časovne serije
Število natovorjenih vagonov v b. Jugoslaviji



Iz grafikona 9.18 je jasno razvidno, da je učinek sezonskih vplivov pozitiven, od julija do novembra, najmočnejši pa je v oktobru. V ostalih mesecih je učinek negativen, najmočnejši pa v februarju. Največji promet je torej v oktobru, najmanjši pa v februarju.

S tem je naloga razdružitve časovne serije v njene sestavne dele izvršena.

94.4 Razstavitev časovne serije na njene sestavne dele

Ako pogledamo kot primer število natovorjenih vagonov v letu 1936, moremo sestaviti naslednjo analizo:

	Osnovna serija	Trend	Sezonski vpliv	Slučajni vplivi
1936 J	96,6	123,2	- 25,8	- 0,8
F	92,5	123,6	- 26,9	- 4,2
M	111,6	124,0	- 9,3	- 3,1
A	102,4	124,3	- 14,7	- 7,2
M	114,0	124,4	- 8,1	- 2,3
J	116,1	125,2	- 8,4	- 0,7
J	129,5	126,5	+ 1,8	+ 1,2
A	145,0	127,6	+15,5	+ 1,9
S	155,0	128,7	+28,2	+ 1,9
O	171,0	130,2	+40,3	+ [illegible]
N	150,8	132,0	+16,0	+ [illegible]
D	129,4	133,4	- 8,6	+ 4,6

Vrednosti osnovne serije so razstavljene na svoje sestavne dele: trend, sezonske vplive in slučajne vplive. Trend in sezonske vplive smo izračunavali v prejšnjih odstavkih, jakost slučajnih vplivov pa smo izračunali iz zveze $Y = T + P + S$ iz katere sledi: $S = Y - T - P$.

95. POVZETEK

Zelo važno je proučevanje dinamike masovnih pojavov. Sliko dinamike socialno - ekonomskih pojavov podajajo časovne serije. Z analizo časovnih serij skušamo odkriti zakonitosti masovnih pojavov, ki so vezane na časoven razvoj. V kapitalizmu je glavni cilj proučevanje časovnih serij predvsem izluščiti zakonitosti zato, da se iz njih predvideva ekonomski razvoj za bodočnost (konjunktúrna statistika). Ker je tudi pri nas veliko pojavov, ki so podvrženi podobnim vplivom, moremo te metode uporabiti tudi mi.

Časovne serije delimo na momentne (n.pr. serija števila prebivalstva po letih) in intervalne (n.pr. število rojstev po mesecih). Da je časovno serijo lažje analizirati, stremimo za tem, da so intervali, za katere veljajo posamezni členi, enako dolgi. V nasprotnem primeru podatke preračunamo na enoto.

Iz osnovnih časovnih serij moremo izračunati več vrst serij. Kumulativno serijo dobimo s postopnim prištevanjem členov osnovne serije.

Serijo drsečih vrednosti dobimo , ako tvorimo vsote členov

časovne serije v dolžini ene periode tako, da je vsak člen te nove serije premaknjen za en delni interval naprej.

S serijo sredin dosežemo, da se število členov časovne serije skrči, vendar je iz nje vidna osnovna črta razvoja.

Serija drsečih sredin je tvorjena na podoben način kot serija drsečih vrednosti, le da namesto y_k izračunamo povprečje.

Da moremo časovno serijo proučiti z metodami, ki so izdelane za njih analizo, morajo biti predvsem členi serije med seboj istovrstne narave, intervali posameznih členov pa enako dolgi.

Enostavno proučitev časovnih serij izvedemo s pomočjo izračunanja naslednjih pokazateljev:

a) absolutne diference $A_k = Y_{k+1} - Y_k$

b) indeksnih serij s stalno bazo $I_{k.o} = Y_k : Y_o$

c) pokazateljev dinamike $P_k = Y_{k+1} : Y_k$

d) serijo verižnih indeksov $V_k = P_k \cdot 100$

e) tempom razvoja $T_k = \frac{A_k}{Y_k} \cdot 100.$

Važna naloga analize časovnih serij obstoji iz razstavljanja časovnih serij na sestavne dele. Časovno serijo si namreč mislimo sestavljeno iz naslednjih vrst gibanj: trenda, ki podaja osnovno linijo razvoja.

enkratnih sprememb, ki so rezultat enkratnih vplivov,

periodičnih sprememb, ki so rezultat periodičnih ali v posebnem sezonskim vplivom in

slučajnih sprememb, ki so rezultat slučajnih vplivov.

V obliki obrazca moremo to dejstvo podati iz:

$$Y = T + E + P + S$$

Enkratne spremembe ugotavljamo s tem, da iščemo razliko med dejansko vrednostjo in vrednostjo, ki bi jo člen časovne serije imel, ako ne bi bilo enkratnega vpliva.

Trend določamo največ načinov. Prostoročna metoda je najenostavnejša, zato pa tudi najmanj natančna. Trend moremo izločiti iz časovne serije s serijo delnih sredin, natančnejšo pa s serijo drsečih sredin. Analitično

moremo določiti trend z metodo najmanjših kvadratov. Iz časovne serije izločimo trend tako, da ga od vrednosti osnovne serije odštejemo.

Ako iz časovne serije izločimo trend in enkratne spremembe, je ostanek rezultat periodičnih in slučajnih sprememb, ker se rezultati slučajnih vplivov v povprečju uničijo, moremo s tvorjenjem grupnih sredin izračunati, kakšne so periodične spremembe.

Na ta način moremo dano časovno serijo razstaviti na njen sestavne dele.

10. GRAFIČNO PRIKAZOVANJE

100. SPLOŠNO

100.0 Uvod

Osnova statističnega proučevanja je primerjava. S statističnimi serijami, srednjimi vrednostmi, relativnimi števili itd. skušamo napraviti statistične podatke pregledne in primerljive. Vsi dosedanji načini so težili za tem, da se poveča možnost primerjanja. To smo imeli v vidu tudi pri sestavljanju tabel, kjer smo podatke, ki jih je treba primerjati, pisali druge poleg drugega. Vendar moremo v tabeli primerjati naenkrat med seboj le dva do tri podatke, in si je težko ustvariti sliko vseh značilnosti proučevanega pojava. V tem primeru je idealno sredstvo grafičen prikaz, v katerem moremo, ako je dobro sestavljen, proučiti vse značilnosti pojava naenkrat.

Grafičen prikaz zviša primerljivost statističnih podatkov. Pač pa ni možno na grafikonu odbrati točno vrednost podatka, kar pa za analizo ni neobhodno potrebno. Zaradi tega grafikon dopolnjujo vedno tabela, ako ni mogoče že v sam grafikon vpisati številčne podatke.

100.1 Vrste grafikonov

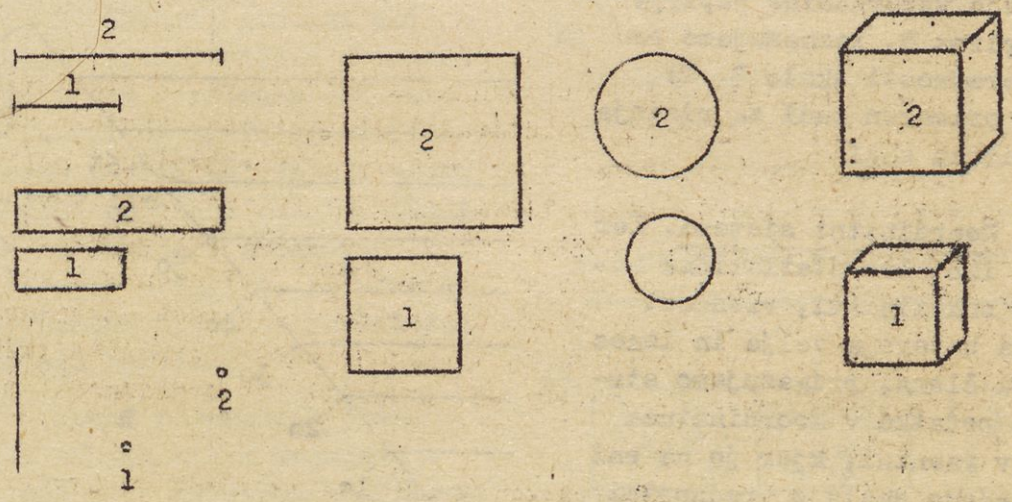
Grafikoni so najrazličnejše vrste. Med seboj se ločijo po tem, kakršnih osnov se poslužujemo za prikazovanje in po tem, kaj prikazujejo. Osnovni dve vrsti grafikonov so diagrami in kartogrami.

100.2 Elementi grafičnega prikaza

Elementi podajanja. Kot elementov grafičnega prikaza se poslužujemo geometrijskih elementov in sicer izmenljivih. Velikost elementa ponazoruje velikost statističnega podatka. Taki elementi so: daljice, kvadrati, pravokotniki, krogi, površine, prostornine, oddaljenost točke od točke, premice od premice. Izmed vseh elementov so najprikladnejši oni elementi, katerih velikost ima enorazsežnost. Ti so torej izpeljani iz pojma daljice. V čim višjo razsežnost gremo, tem slabša je primerljivost. Iz daljice izpeljani elementi so poleg daljice stolpci in oddaljenost točke od točke ali premice.

Poleg geometrijskih likov uporabljamo tudi figure, vendar le bolj zaradi poljudnosti podatkov.

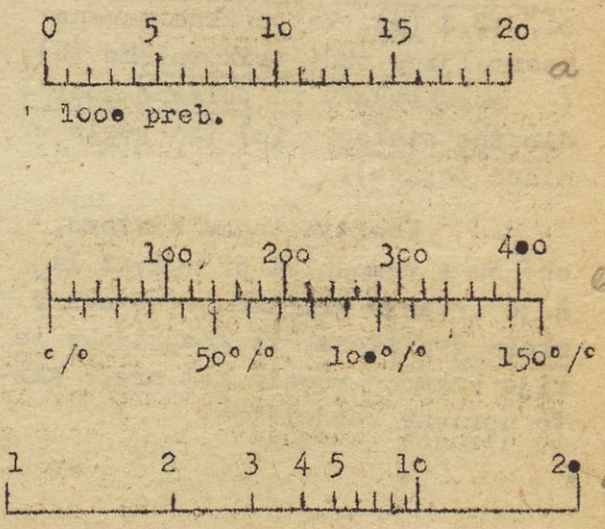
Sl. 10.1 Primerljivost elementov grafičnega prikazovanja



100.3 Skale - lestvice

Skala ali lestvica je merilo, ki pokaže odnos med velikostjo statističnih podatkov in dolžino. 1 cm dolžine more pomeniti 1000 prebivalcev, 1 ha površine itd. V grafičnem prikazovanju uporabljamo različne vrste skal. Običajno uporabljamo navadno aritmetično skalo, kjer so dolžine sorazmerne velikosti podatka. (a)

Sl. 10.2 Skale



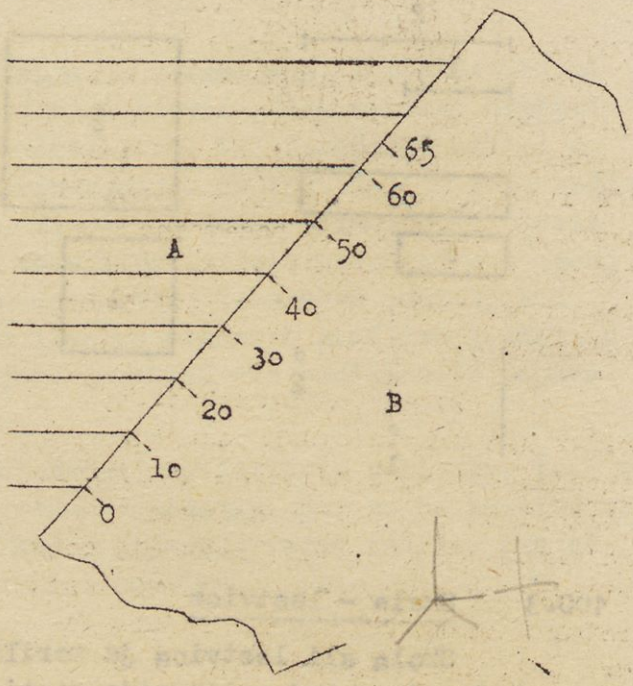
Dostikrat je koristno, da nanesemo na eni in isti premici dve ali več skal. Tako je n.pr. na primeru (b) možno odčitati absoluten podatek in indeks.

Poleg aritmetičnih skal v statistiki zelo pogosto uporabljamo logaritmične skale (c). Dolžine na tej skali so v sorazmerju z logaritmom podatka. Logaritmična skala je posebno priporočljiva, kadar gre za proučevanje relativnih odnosov.

Skalo v poljubnem merilu moremo dobiti kot kaže slika 10.3. Na črtan papir(A) položimo rob drugega lista (B) tako, da razdalja na robu lista (B), katero je treba razdeliti na 6 1/2 enot (= 35)

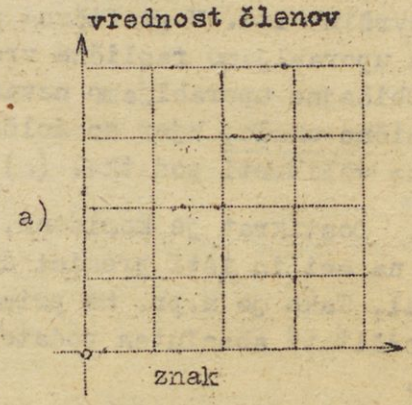
prečka 6 in pol črte lista (A). Kjer sekajo vzporednice papirja A rob papirja B, zaznamujemo posamezne vrednosti skalo B. Ta način je primeren tudi za risanje logaritmičnih skal.

Sl. 10.3 Sestavljanje skale

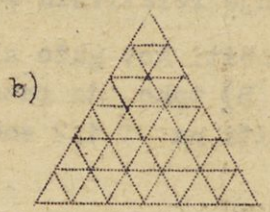


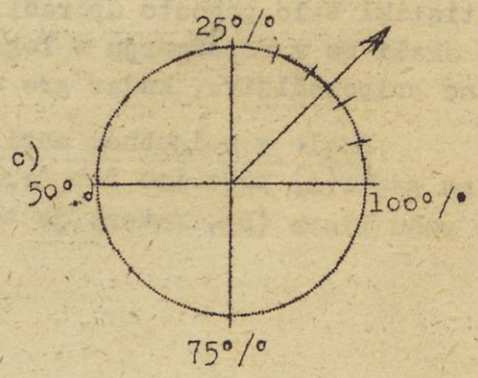
110.4 Koordinatni sistemi. Ker ima člen statistične serije dve značilnosti, vrednost znaka, za katerega velja in iznos vrednosti člena, prikazujemo statistične podatke v koordinatnem sistemu v ravnini, kjer je na eni osi nanešena skala z vrednostmi osnovnega znaka serije, na drugi strani pa skala za vrednosti členov. Najpogosteje uporabljamo pravokotni koordinatni sistem sl. 10.4 (a). Za posebne namene uporabljamo tudi trikotniške (b) (strukture) in polarne koordinatne sisteme (c) (za periodične pojave).

Sl. 10.4 Koordinatni sistemi



Koordinatnemu sistemu, obenem s pomožnimi črtami, ki lajšajo črtanje grafikonov, pravimo mreža grafikona. Za geografske serije vzamemo včasih kot mrežo karto upravne razdelitve.

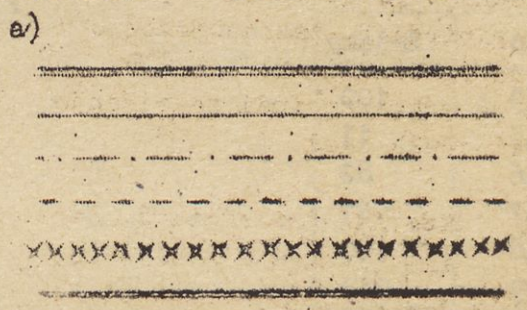




100.5 Črtkanje in šrafiranje

S točkami, stolpci ali površinami in črtami rišemo običajno na enem in istem grafikonu več različnih pojavov. Da razlikujemo črte za posamezne pojave med seboj, uporabimo različna črtkanja (sl. 10.5b) ali barve. Da razlikujemo posamezne stolpce oziroma površine, pa posamezne površine šrafiramo z različnimi šrafurami ali barvami.

Sl. 10.5 Črte in šrafure



Pomen posameznih črtk, šrafur ali drugih oznak podajamo v legendi, običajno na samem grafikonu. V težnji za čim boljšo nazornost grafikona, šrafure ali barve čim bolj prilagodimo prikazanemu pojavu. Mesto šrafur vnašamo v površine majhne figure (b), ki omogočajo neposredno zvezo s proučevanim pojavom. Tudi barve morejo večkrat smiselno predstavljati nek pojav. (Njive rjavo, travnik zelen, neplodno belo). Jakost nekega pojava prikažemo s svetlejšo ali temnejšo šrafuro (sl. 10.5c).

101. STOLPCI

Sl. 10.6 y_i

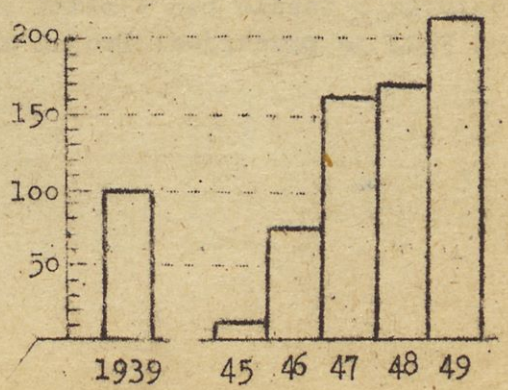


Stolpec kot element grafičnega prikazovanja v statistiki imenujemo vsak pravokotnik, katerega dolžina pomeni velikost podatka Y_k, širina Δ pa je za vse člene statistične serije enaka. (Sl. 10.6).

101.1 Enostavni stolpci.

S stolpci moremo prikazovati vse vrste statističnih serij: časovne, geografske, stvarno-atributivne in stvarno numerične.

Sl. 10.7 Indeks proizvodnje surovega železa v FLRJ



.11 Prikazovanja časovnih in stvarno numeričnih serij, s stolpci je brez posebnih težav, ker je vrstni red stolpcev s serijo točno

določen. Stolpce moremo risati strnjeno skupaj, razen pri časovnih serijah za one člene, za katere je časovno razdobje več kot leto.

Tab. 10.1

Leto	Indeks
1939	100
1945	11
1946	82
1947	161
1948	170
1949	213

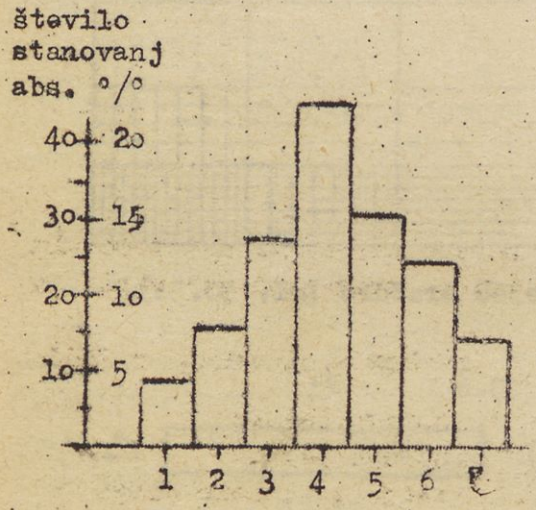
P Indeks proizvodnje surovega železa v FLRJ (po proračunski debati 1948) sl. 10.7.

Kadar prikazujemo stvarno- numerične frekvenčne distribucije s stolpci, imenujemo ta grafikon histogram. Zaradi popolnosti podajamo v sl. 10.8 histogram za serijo števila stanovanj po številu stanovalcev iz odstavka 6 z dvojno skalo, absolutno in procentualno.

Sl. 10.8

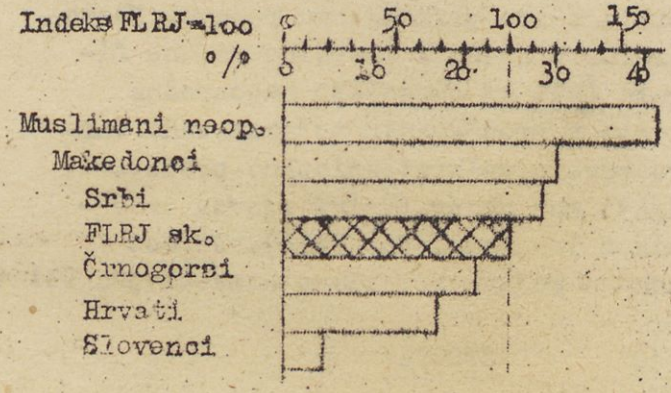
Število stanovalcev

Število stanovanj po številu stanovalcev



Sl. 10.9

Odstotek nepismenih v FLRJ po narodnosti



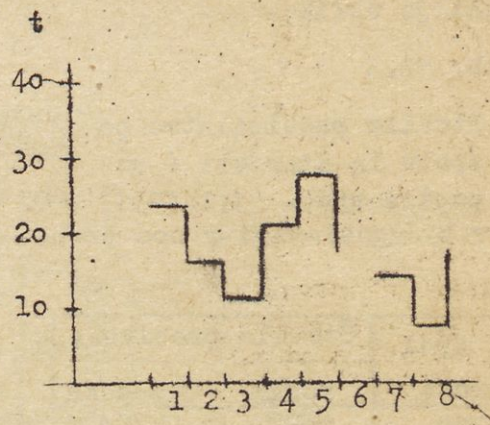
.12 Pri geografskih in stvarno atributivnih serijah obstoji težava zaradi tega, ker vrstni red členov ni vnaprej določen. Običajno stolpce razvrstimo po velikosti.

V našem primeru stvarno- atributivne serije iz odstotka nepismenih iz odstavka 6 tabele 6.4 v sliki 10.9 smo razvrstili po velikosti. Stolpec za skupno FLRJ je povdarjen. Dvojna skala kaže % nepismenosti in indeks FLRJ = 100.

.13 Včasih pri prikazovanju časovnih ali stvarno-numeričnih serij uporabljamo samo obrise stolpcev. S tem preidemo že skoro v prikaz z linijami. Ta grafikon imenujemo stopničasti (sl. 10.10).

V dopisih moremo stolpičaste diagrame risati s črko x na pisalnem stroju x pomeni določeno količino, n.pr. 10 živali, 1000 ton itd. sl. 10.11

Sl. 10.10 Dnevna proizvodnja



Sl. 10.11 Proizvodnja

	1 2 3 4 5 6 7 8 9 10
A	XXXXXXXX
B	XXXXX
C	XXXX
D	XXXX

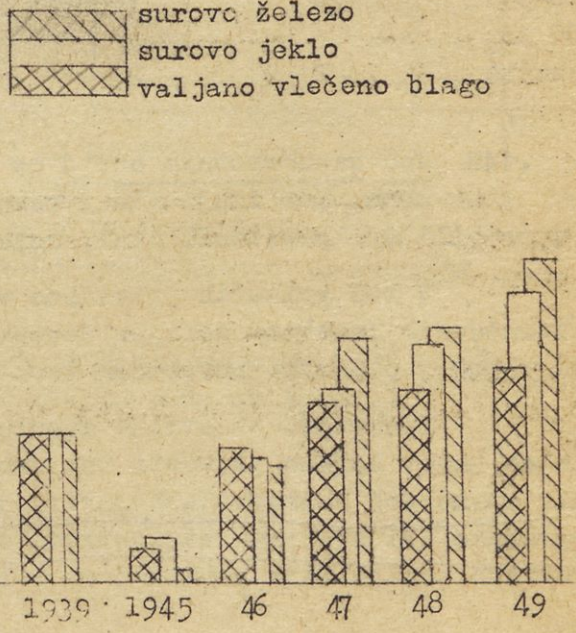
.14 Grafični prikaz več serij naenkrat. Kompleksno analizo moremo napraviti z istočasnim opazovanjem več serij. Zaradi tega skušamo grafično prikazati več serij na istem grafikonu.

.141 Prikaz več istovrstnih serij, je razmeroma lahko izvesti, ker velja ena in

P Tab. 10.2 Indeksi proizvodnje v FLRJ (po proračunski debati 1948)

Artikel / Leto	Surovo železo	Sur. jeklo	Valjano vlečeno blago
1939	100	100	100
1945	11	28	23
1946	82	86	91
1947	161	132	121
1948	170	157	129
1949	213	194	141

Sl. 10.12 Indeksi proizvodnje v FLRJ



istabskala za vse serije. Stolpce je le treba označiti s šrafurami, za katere serije veljajo. Stolpce razporedimo tako, da so oni, katerih primerjava je važnejša, blizu skupaj.

Klasičen primer prikazovanja več istovrstnih serij s stolpci je življenjsko drevo, s katerim podamo strukturo prebivalstva po starosti in spolu. Prednost damo primerjavi po starosti. Ime ima grafikon zaradi svoje značilje oblike.

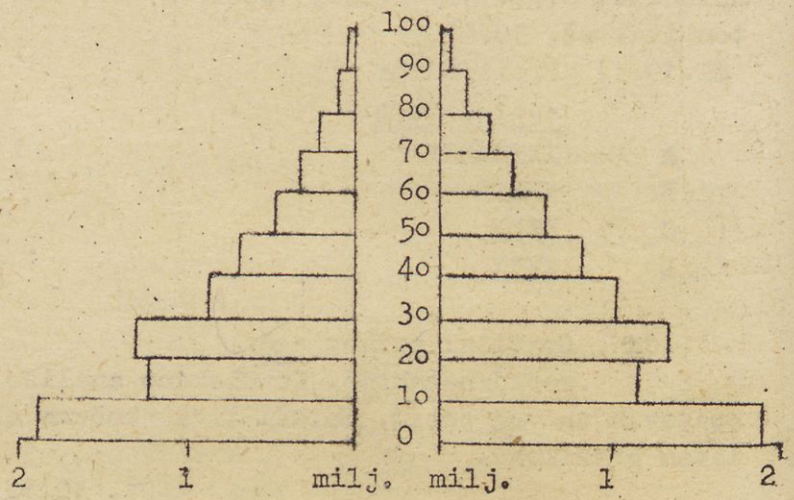
Tab. 10.3

Število prebivalstva po spolu in starosti (po popisu preb. 31.3.1931) v b.Jugoslaviji v ooo preb.

Spol / Starost	Število prebival. moški	ženske
Skupna	6892	7042
0 - 9	1898	1829
10-19	1206	1177
20-29	1289	1258
30-39	845	914
40-49	627	738
50-59	470	535
60-69	343	366
70-79	164	171
80-89	42	45
90-99	6	7
100 in neznano	2	2

Sl. 10.13

Življenjsko drevo števila prebivalstva po starosti in spolu v Jugoslaviji leta 1931



142 Več raznovrstnih serij ne moremo direktno prikazati na istem grafikonu, ker imajo členi vsake serije drugačno enoto mere. Tudi ako bi uporabili več različnih skal, ostane vprašanje razmerja med poedinimi skalami.

V teh primerih privedemo vse serije na isto enoto mere na ta način, da izračunamo indeksne serije. Seveda moramo vzeti kot bazo pri vseh serijah isti [illegible], običajno povprečja vseh členov posameznih serij.

Kadar gre za serije absolutnih vrednosti, ki podajajo sestav neke celotne mase, moremo privesti vse enote na isto mero, ako izračunamo strukture. Kadar imamo več raznovrstnih serij, jih s stolpci prikazujemo le, ako gre za geografske ali stvarno- atributivne serije. Raznovrstne časovne ali stvarno - numerične serije raje prikazujemo z linijskimi grafikoni.

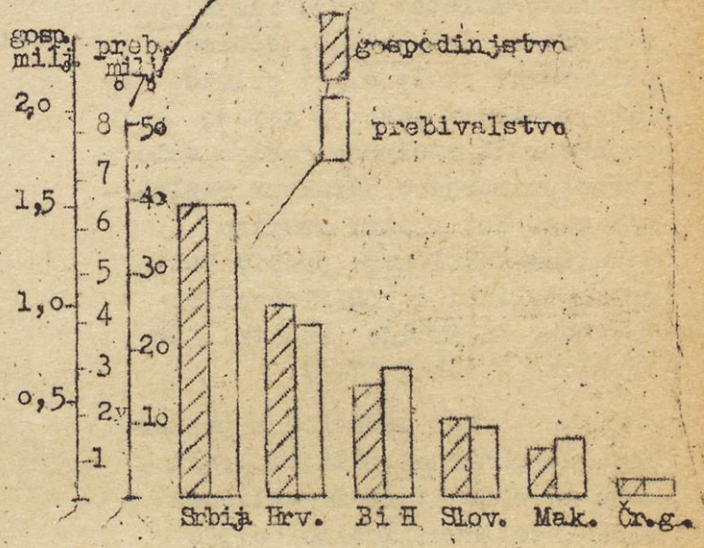
Tab. 10.4

Število gospodinjstev in prebiv. v FLRJ po repa - blikah (po popisu 15.3. 1948).

LR	Število v 000		Struktura	
	gosp.	preb.	gosp.	preb.
	G_k	P_k	g_k	p_k
FLRJ	3610	15772	1000	100
Srbija	1486	6528	41	41
Hrvatska	954	3757	27	24
Slovenija	369	1392	10	9
Makedonija	219	1153	6	7
B. i H.	498	2565	14	16
Črna Gora	84	377	2	2

Sl. 10.14

Število gospodinjstev in prebivalstva po republikah



Iz slike moremo poleg primerjave števila gospodinjstev in števila prebivalstva po republikah videti tudi, kakšno je notranje razmerje med številom gospodinjstev in prebivalstvom. Ako je stolpec za prebivalstvo večji od stolpca za gospodinjstvo za isto republiko, pomeni, da je povprečno število članov na eno gospodinjstvo večje od povprečja za FLRJ in obratno, ako je stolpec gospodinjstev večji, pomeni to, da je število članov na eno gospodinjstvo manjše od povprečja za FLRJ. V LR Hrvatski in Sloveniji so gospodinjstva manjša kot povprečno za FLRJ; v LR Srbiji in Črni gori enaka povprečju v LR B. i H. in Makedoniji pa večja od povprečnega za FLRJ. Trojna skala omogoča odčitanje strukture, števila prebivalstva in števila gospodinjstev.

Ta način je zelo uporaben v vseh primerih raznovrstnih geografskih in stvarno - atributivnih serij in nudi veliko možnost analize.

.12 Strukturni stolpci

Ako enostaven stolpec, katerega dolžina pomeni absolutno vrednost celotne mase, razdelimo v razmerju delnih mas, dobimo strukturni stolpec, ki prikaže notranji sestav mase. S serijo strukturnih stolpcev dobimo zelo dobro primerjavo struktur različnih mas. Kot primer vzemimo strukturo aktivnega prebivalstva po zaposlitvi za LR Slovenijo, Hrvatsko in Srbijo (Podatki iz odstavka 7 tabele 7.4).

Ker medsebojna primerjava struktur na gornjem grafikonu ni dobra zaradi različnih višin stolpcev, dobimo boljšo primerjavo struktur, ako višine stolpcev rišemo v sorazmerju s strukturnimi procenti. Višine vseh stolpcev so v tem primeru enake 100. (Sl 10.16). Ta slika pa ima to pomanjkljivost, da ne pokaže absolutnih obsegov celotnih mas, kar je podano v gornjem primeru z različnimi višinami stolpcev.

Združitev prikaza strukture v odstotkih in absolutnega obsega na istem grafikonu dosežemo z vpeljavo pravokotnika kot element grafičnega prikaza.

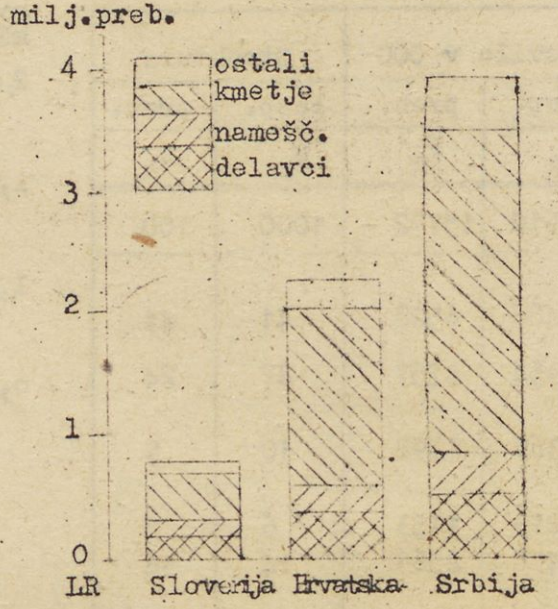
102.

11.2 Pravokotniki

V statistiki nastopa dosti podatkov, ki so v zvezi A.B = C. Skupno število prebivalstva je enako produktu gostote in površine, površina njiv produktu odstotnega deleža in skupne površine

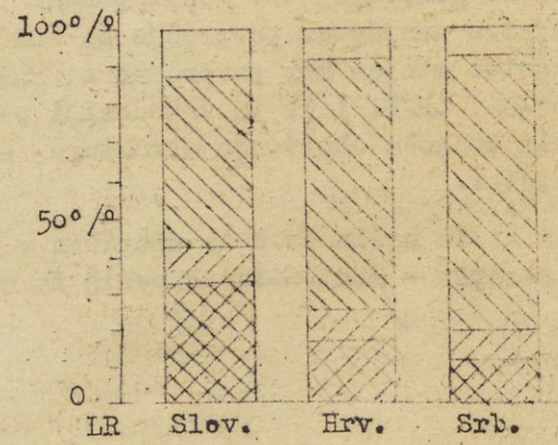
Sl. 10.15

Struktura aktivnega prebivalstva po zaposlitvi

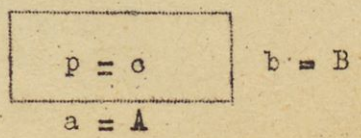


Sl. 10.16

Struktura aktivnega prebivalstva po zaposlitvi
Legenda ista kot pri 10.15



Sl. 10.17

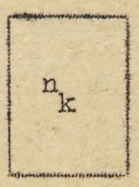


frekvenca produktu gostote in širine razreda itd. Vse te tri podatke moremo prikazati naenkrat s pravokotnikom, ker je produkt stranic enak ploščini ab = p (sl. 10.17). Ena stranica predstavlja en podatek, druga drugi, s ploščino pa je označen podatek, ki je izpeljan iz A in B kot produkt.

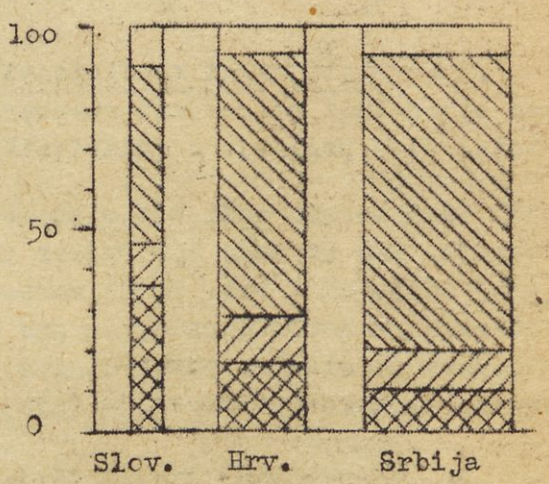
.21 Prikrojeni strukturni stolpci

Gornjo lastnost moremo uporabiti pri risanju struktur, ker je obseg delne mase enak produktu obsega celotne mase in odstotnega deleža. Ako vzamemo n (absolutni obseg) kot širino stolpca, s_k (strukturni procent) kot višino stolpca, je ploščina tega pravokotnika v sorazmerju z n_k (obsegom delne mase) (sl.10.18). Strukturne pravokotnike one in iste mase moremo zaradi tega, ker je širina vseh ista (skupen obseg za isto maso stalen), risati drugega nad drugim. Struktura zaposlitve je prikazana upoštevaje ta dejstva - v sliki 10.19.

Sl. 10.18

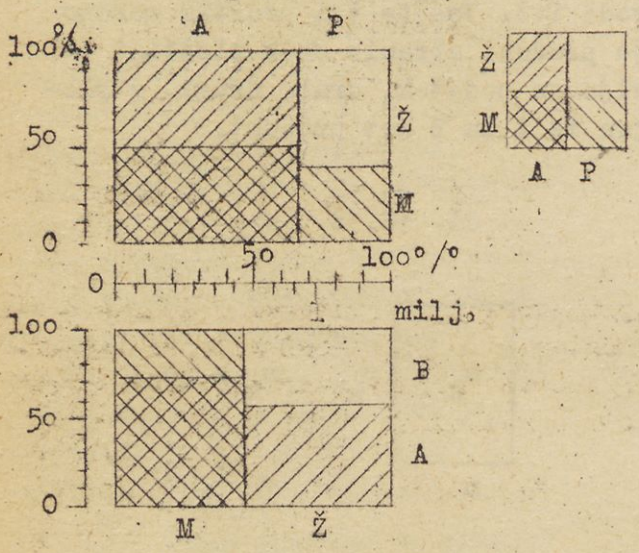


Sl. 10.19
Struktura aktivnega prebivalstva po zaposlitvi
Legenda ista kot pri 10.15

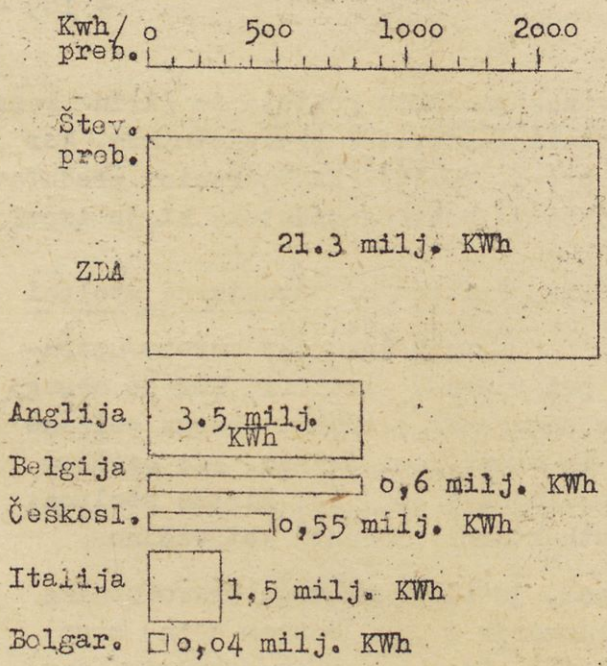


.22 Dvojna, oziroma mnogoterna struktura . Lastnost pravokotnikov moremo uporabiti tudi za prikaz dvojnih struktur. Stolpec, ki predstavlja obseg celotne mase, razdelimo najprej v sorazmerju z razdelitvijo po prvem znaku. Celotno maso dobimo sestavljeno iz serije pravokotnikov. Vsak delni pravokotnik pa moremo razdeliti v sorazmerju z razdelitvijo po drugem znaku. Ako vzamemo kot primer razdelitev prebivalstva po spolu in vzdrževanosti iz odstavka 7,6 in 7,7 dobimo sliko 10.20 Enako kot izračunavamo mnogoterno strukturo na več načinov, je tudi več vrst slik mnogoterne strukture.

Sl. 10.20
Struktura števila prebivalstva po spolu in aktivnosti



Sl. 10.21
Proizvodnja električne energije na prebivalca l. 1947 po državah

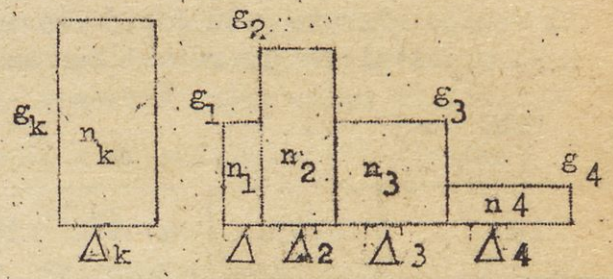


.23 Statistični koeficienti in gostote. Na način A.B = C sta povezani med seboj tudi primerjani raznovrstni masi in statistični koeficient ali gostota. Glej poglavje o relativnih številih. Ker je $K = \frac{R}{S}$ je $R = K.S.$ Zaradi tega moremo tudi v tem primeru uporabiti pravokotnik kot element prikazovanja. Ako vzamemo kot primer proizvodnjo električne energije na prebivalca (glej odstavek 7 tab. 7.10) moremo vzeti kot dolžino stolpcev podatek za katerega primerjavo predvsem gre (proizvodnja na prebivalca v KWh), kot širino stolpcev pa število prebivalstva. Ploščina pravokotnika v tem primeru pomeni skupno proizvodnjo energije (glej sliko).21.

.24 Frekvenčna distribucija z neenako širokimi razredi. Pravokotnikov se poslužujemo tudi za risanje frekvenčnih distribucij, kadar so širine razredov različno velike. Ker je : $n_k = g_k \Delta_k$ (glej odstavek 7) predstavlja širina pravokotnika širino razreda Δ_k, višina pravokotnika gostoto k'

ploščina pravokotnika pa absolutno vrednost frekvence n_k. Vsak drug način risanja frekvenčnih distribucij te vrste je napačen. Histogram za frekvenčno distribucijo zemljiških gospodarstev po velikosti je narisan pri poglavju o statističnih serijah (odstavek 6 slika 6.2).

Sl. 10.22

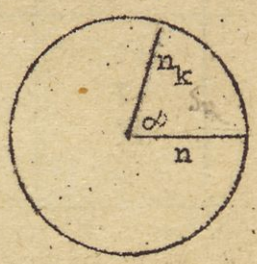


103. KROGI

Kroge uporabljamo včasih za prikazovanje strukture. Cel krog pomeni celotno maso, medtem ko delne mase predstavljajo posamezni krogovi izseki. Velikost krogovega izseka (kot α), je v sorazmerju z odgovarjajočim strukturnim odstotkom delne mase in sicer sta v naslednji zvezi

$$\alpha_k^o = 3{,}6 . s_k \qquad s_k = \frac{n_k}{n} . 100 \qquad (10.1)$$

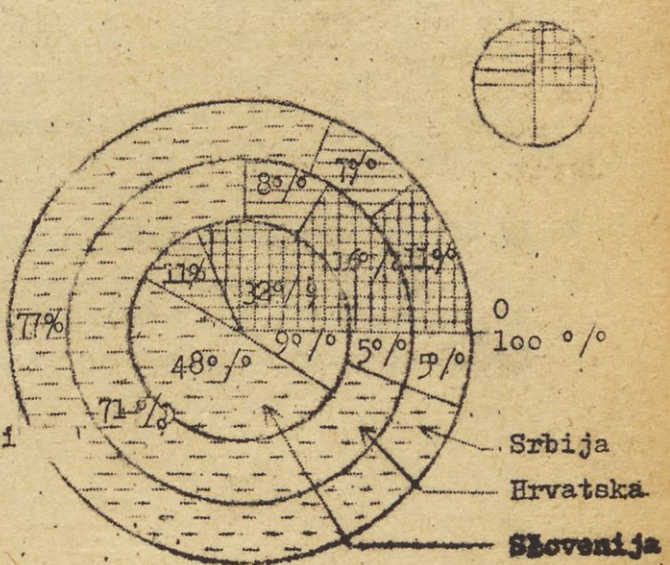
Sl. 10.23



Krog je pripraven za prikazovanje strukture predvsem zaradi tega, ker že po svoji obliki predstavlja celoto, izseki pa dele. Poleg tega so vrednosti kot 25 °/o, 50 °/o, 75 °/o zelo lahko ocenljivi.

Ker je primerjava večjih struktur prikazanih s krogi, težka, uporabljamo kroge za prikaz več struktur običajno le, če kroge vnesemo v geografsko mrežo (regionalna razdelitev pojava), ali pa kroge vrišemo drugega v drugega, kot kaže slika strukture aktivnega prebivalstva za LR Slovenijo, Hrvatsko in Srbijo. Važno je, da morajo biti v primeru, da vrišemo več struktur s krogi, ploščine krogov v sorazmerju z obsegi celotnih mas. Radij kroga za eno maso izberemo poljubno, vsi ostali radiji pa so določeni z razmerji med obsegi posameznih mas.

Sl. 10.24
Struktura po zaposlitvi za LR Slovenijo, Hrvatsko in Srbijo



Ker $\pi r^2 = CM$, $\pi r_k^2 = CM_k$ je zveza med radiji naslednja:

$$r_k = r_1 \sqrt{\frac{M_k}{M}}$$

P Narisati je treba s krogi strukturo aktivnega prebivalstva za LR Slovenijo, Hrvatsko in Srbijo (primer iz 7). Dano je skupno število aktivnega prebivalstva po republikah in struktura po zaposlitvi v %.

Tab. 10.5

LR	Srbija	Hrvatska	Slovenija
Skupno štev. akt. prebival. v 000	4174	2475	862
Skupno	struktura v %		
	100 %	100 %	100 %
Delavci	11	16	32
Nameščenci	7	8	11
Kmetje	77	71	48
Ostali	5	5	9

Srbija	Hrvatska	Slovenija
30 mm	23,1 mm	13,6 mm
kot ° krogovega izseka		
360°	360°	360°
40	58	115
25	29	40
277	255	173
18	18	32

Radij največjega kroga (Srbija) izberimo $r_1 = 3,0$ mm

$r_2 = r_1 \sqrt{\frac{M_2}{M_1}}$ radij kroga za Hrvatsko je: $r_2 = 30 \sqrt{\frac{2475}{4174}} = 23,1$ mm

Radij kroga za Slovenijo : $r_3 = 30 \sqrt{\frac{862}{4174}} = 13,6$ mm

Za LR Srbijo izračunamo n.pr. kot krogovega izseka za delavce takole:

$\alpha_k = 3,6\, s_k$ 3,6 10 Enako izračunaj tudi ostale kote.

Kadar primerjamo strukturi dveh mas, ki sta si v razmerju kot uvoz - izvoz, nabavljeno - prodano, rojeni - umrli itd. uporabljamo za posamezno maso kot celoto polkrog. Oba kroga sta postavljena tako, da tvorita skupno celoto.

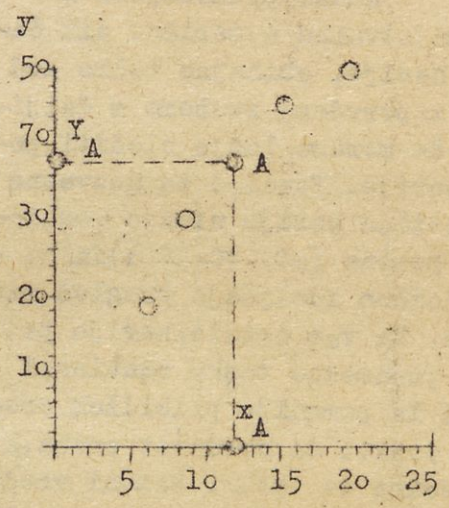
104. TOČKE

Ako imamo križani v koordinatnem sistemu dve skali, skali x in y, vsaka točka A v koordinatnem sistemu pomeni dvojico vrednosti, ki ju dobimo kot pravokotno projekcijo na osi koordinatnega sistema. (Glej sliko 10.24).

Ako vzamemo kot absciso skalo osnovni znak serije, kot ordinatno skalo pa skalo za vrednosti členov serije, moremo v koordinatnem sistemu s točkami prikazati statistično serijo. Točka A je slika onega člena serije, ki ima vrednost osnovnega znaka x_A, vrednost člena pa Y_A. Celotno statistično serijo moremo prikazati z nizom točk, ki smo jih narisali na podoben način.

Sl. 10.24

Pravokoten koordinatni sistem



Ako pogledamo naravo osnovnih znakov serij, vidimo, da veljajo členi serije edino pri nezveznih stvarno- numeričnih in momentnih časovnih serijah za točno določeno vrednost znaka. Pri intervalnih serijah se namreč podatek nanaša na ves grupni interval. Ker vemo, da za intervalne serije posameznemu razredu pripišemo sredino razreda kot grupno vrednost (- glej statistične serije -) bomo v primerih intervalnih serij risali točko, ki bo pomenila podatek za določen interval, nad sredino razreda.

Ako sta na obeh oseh koordinatnega sistema nanešeni skali dveh numeričnih znakov enot statistične mase, pomeni vsaka točka enoto, ki ima vrednosti znaka, ki odgovarjajo projekcijam na koordinatnih skalah. Tak točkovni grafikon, katerega primere imamo pri poglavju o osnovah korelacije (odstavek 12), pokaže odvisnost oz. povezanost dveh znakov statistične mase.

Ako hočemo prikazati povezanost dveh statističnih serij, moremo vzeti v koordinatni osi skali za vrednost členov prve in druge serije. Na ta način moremo prikazati s točkami tudi podatke geografskih ali stvarno atributivnih serij, kar v prvem primeru ni mogoče. Kot primer vzemimo podatke o površinah

in številu prebivalstva za nekaj držav. Podat so iz odstavka o relativnih številih tab. 7.8.

Da je možno lažje odčitanje podatkov, včrtamo v koordinatni sistem pomožne črte - mrežo.

105. ČRTE - LINIJE

Kadar prikazujemo s točkami stvarno numerične ali časovne serije, običajno točke med seboj zaporedoma zvežemo z daljicami, da moremo lažje slediti poteku serije. Veznice za nezvezne statistične serije nimajo vsebinskega pomena (sl. 10.26, a) in samo tehnično zboljšajo pregled nad serijo. Za vse ostale serije pa imajo posamezne točke vsebinski smisel in pomenijo približek vrednosti pojava za vrednost znaka, ki je abscisa te točke. Pomeni vrednosti vmesnih točk pri posameznih vrstah statističnih serij so razvidni iz slike 10.26.

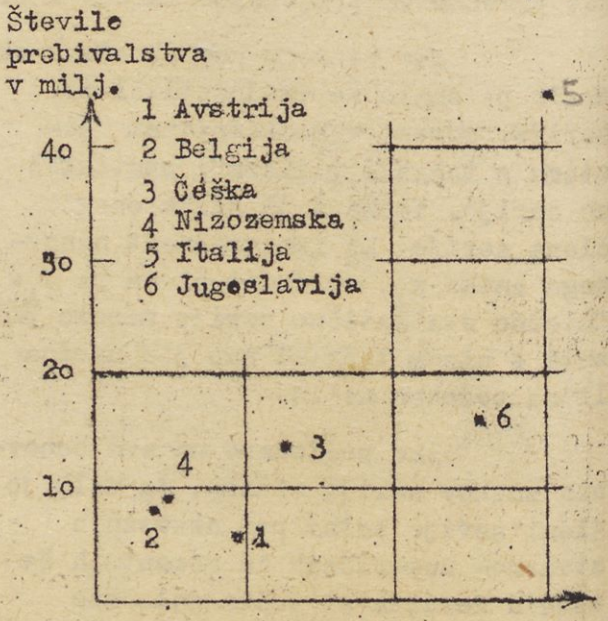
Tej vrsti garfikonov pravimo linijski grafikoni. Linijske grafikone uporabljamo pri stvarno numeričnih, predvsem pa časovnih serijah.

105.1 Stvarno numerične serije

.11 Primer linijskega grafikona na stvarno - numerične nezvezne serije imamo prikazan že pri statističnih serijah (gl. odst. 6). Ta način je brez posebne problematike. Nad nezvezne serije nanesemo vrednosti s točkami, te pa povežemo med seboj. V primeru, da gre za frekvenčno distribucijo, dobimo poligon frekvenc.

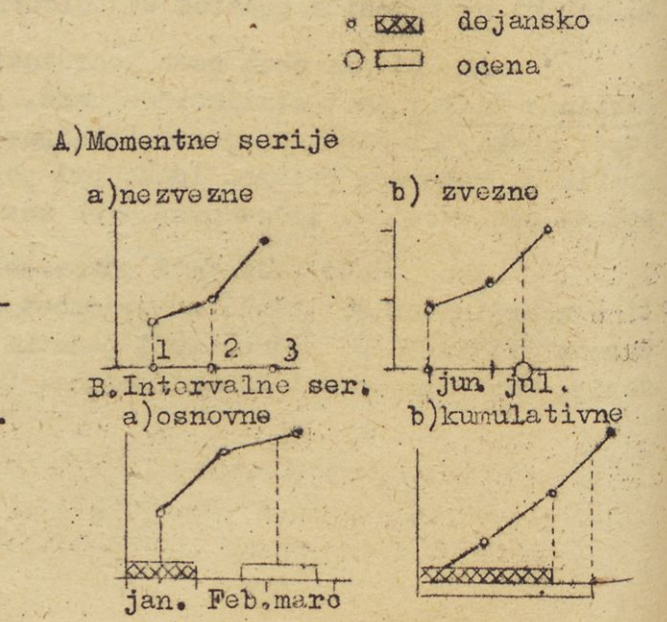
Sl. 10.25

Odnos med številom prebivalstva in površino za 6 evropskih držav



Sl. 10.26

Različni pomeni vrednosti točk linijskega diagrama



P Kot primer moremo vzeti tudi število gospodinjstev po številu članov za FLRJ in Slovenijo. Grafikon spremljajo tri skale in sicer: absolutna za podatke FLRJ, absolutna za LRS in odstotna skala. Razmerje med skalami moremo dobiti po načinu opisanem v odstavku 100.3.

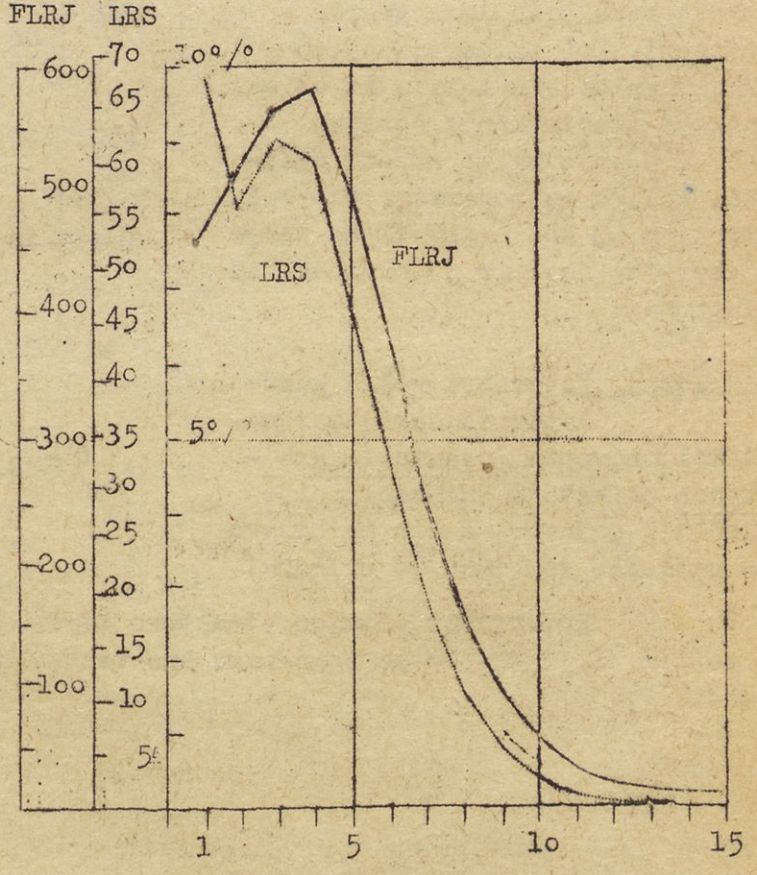
Tab. 10.6

Število gospodinjstev v FLRJ in LRS po številu članov v 000 gospodinjstev po popisu 1. marca 1948

Število članov	FLRJ	Slovenije
Σ	3609,6	368,8
1	451,1	68,0
2	509,4	55,4
3	561,2	63,3
4	574,9	60,8
5	492,6	46,2
6	372,3	31,2
7	250,9	19,3
8	156,9	11,4
9	93,4	6,3
10	53,7	3,2
11	31,0	1,7
12	19,9	0,9
13	11,6	0,3
14	7,7	0,1
15	5,3	0,1
16-20	10,8	0,1
21-25	2,8	0,1
26-	3,9	0,4

Sl. 10.27

Število gospodinjstev po številu članov v gospodinjstvih

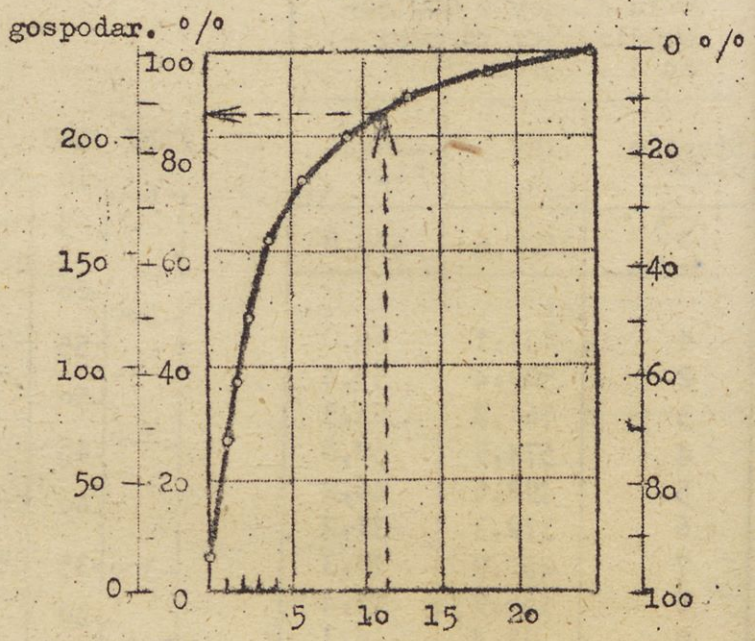


.12 Pri zveznih stvarno - numeričnih serijah nanašamo vrednosti členov serije nad sredino razreda. Čeprav je to približek, v večini primerov odgovarja. Primer linijskega grafikona zvezne stvarno - numerične serije je poligon frekvenc števila gospodarstev v LRS po velikostnih razredih. Glej odstavek o frekvenčnih distribucijah sl. 6.4.

.13 Linijski grafikon kumulativne serije iz stvarno numeričnih serij je zelo dober pripomoček analize in moremo iz njega napraviti zelo koristne zaključke. Za vsako vrednost osnovnega znaka in ne samo za meje razredov, moremo iz linijskega grafikona kumulativne serije razbrati, koliko vrednosti je pod in koliko nad izbrano vrednostjo. Pri merah variacije (11.5) je podan način, kako moremo iz slike kumulativne serije razbrati quartile in interquartilno vrednost. Poleg tega je linijski grafikon kumulativne serije boljši kot grafikon osnovne serije, ker so abscise vrednosti členov kumulativne serije točno določene (meje razredov), medtem ko je za zvezne stvarno-numerično serije sredina razreda samo približek grupne vrednosti. Kot primer navajamo sliko kumulativne serije iz odstavka o statističnih serijah (tab. 6.9). Iz slike moremo odčitati, koliko gospodarstev ima več, oziroma manj kot n.pr. 12 ha skupne površine in to v absolutnem in relativnem.

Sl. 10.28

Kumulativna serija števila zemljiških gospodarstev v LRS po velikosti skupne površine



.14 Na osnovi večih kumulativnih serij moremo s takoimenovanim Lorenzovim grafikonom prikazati koncentracijo določenih pojavov. Način in bistvo sta najlepše razvidna iz primera:

Kot primer vzemimo strukturo živine v LRS po velikostnih skupinah po popisu 1. 1947. Vzeli bomo samo one vrste, ki kažejo izrazito tipičnost. (Tab. 10.7).

A3 5

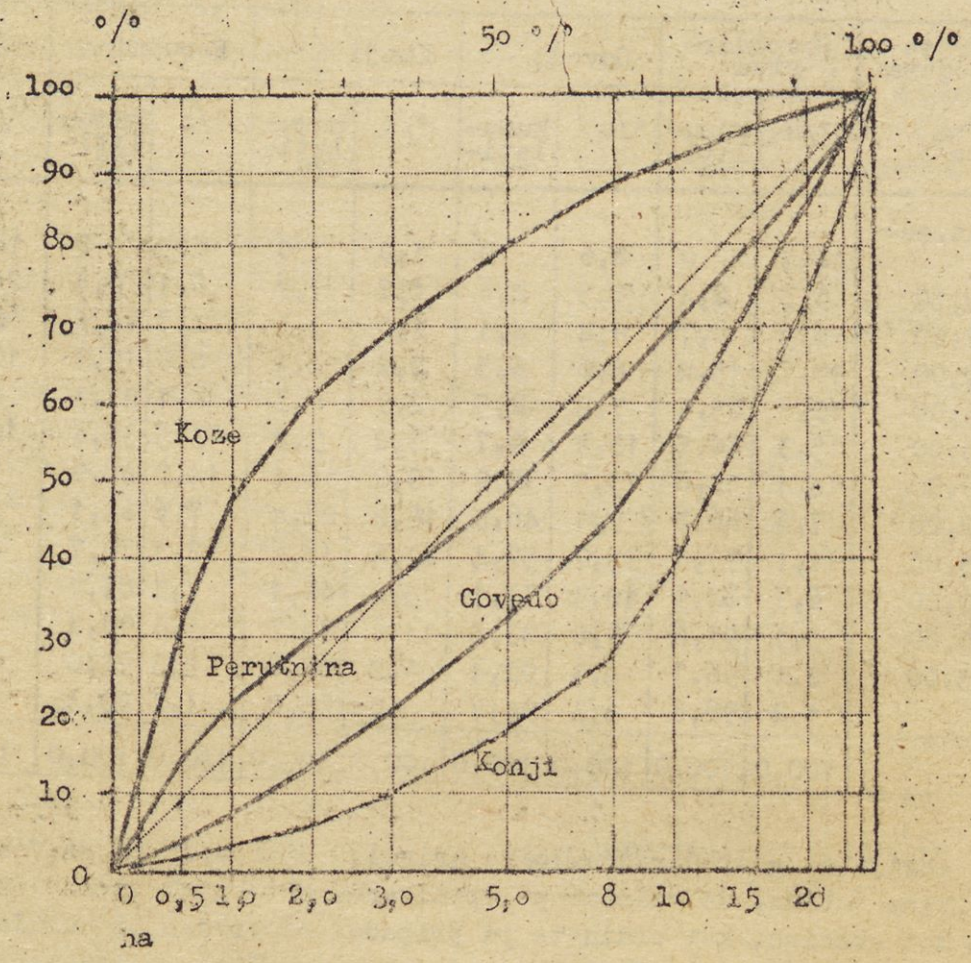
Tab. 10.7 Število gospodarstev in živine v LRS po velikostnih skupinah v LRS po popisu živine l. 1949

Vrsta pojava / Vel. skup v ha	Gospodarstva		Govedo		Konji		Perutnina		Koze	
	%	kumulative	%	kumulative	%	kumulative	%	kumulative	%	kumulative
brez last. zemljo	2,4	0	2,6	0	1,9	0	8,5	0	18,5	0
0,01 - 0,50	6,3	2,4	1,5	2,6	1,2	1,9	6,1	8,5	21,4	18,5
0,51 - 1,00	6,5	8,7	2,4	4,1	1,0	3,1	4,7	14,6	0,6	39,3
1,01 - 2,00	10,7	15,2	6,0	6,5	2,2	4,1	8,0	19,3	10,0	50,5
2,01 - 3,00	10,6	25,9	7,2	12,5	2,9	6,3	8,1	27,3	8,1	60,5
3,01 - 5,00	15,3	36,5	13,2	19,7	7,2	9,2	13,5	35,4	10,6	68,6
5,01 - 8,00	14,4	51,8	15,5	32,9	12,0	16,4	14,2	48,9	7,2	79,2
8,01 -10,00	7,6	66,2	9,1	48,3	10,2	28,4	7,6	63,1	3,7	86,4
10,01-15,00	11,1	73,8	15,7	57,4	21,5	38,6	12,4	70,7	4,3	90,1
15,01-20,00	6,7	84,9	10,3	73,1	15,6	60,1	7,3	83,1	2,3	94,4
20,01-30,00	5,1	91,6	9,0	83,4	13,1	75,7	5,6	90,4	1,7	96,7
30,01-45,00	2,1	96,7	4,3	92,4	6,2	88,8	2,3	96,0	1,0	98,4
45,01-	1,2	98,8	3,3	96,7	5,0	95,0	1,7	98,3	0,6	99,4
	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Osnovna serija so gospodarstva. Točke krivulje Lorenzovega grafikona za določeno vrsto živine dobimo, da vzamemo po vrsti vrednost kumulativne serije za gospodarstva kot absciso, kot ordinate pa pripadajoče vrednosti kumulativne serije za določeno vrsto živine. Te točke med seboj povežemo. Na ta način dobimo v kvadratu za vsako vrsto živine linijo, ki teče iz enega ogla kvadrata do drugega. Iz poteka krivulje za posamezno vrsto moremo sklepati na razporeditev tiste vrste živine po velikostnih skupinah. Ako krivulja v začetku polagoma, proti koncu pa strmo narašča, vemo, da je število te vrste živine pri velikih gospodarstvih znatno večje kot pri malih (konji). Obratno kaže pojav koncentracijo pri malih gospodarstvih, ako takoj v začetku strmo narašča proti koncu pa raste bolj počasi (koze). Ako bi se krivulja skladala z diagonalo, bi pomenila, da je te vrste živina enako pri malih kot pri velikih, - koncentracije ne bi bilo. Naša slika kaže to približno za perutnino. Poleg teh zaključkov moremo iz Lorenzovega grafikona odbrati še druge stvari, koliko in do katere velikosti imajo gospodarstva določen odstotek širine itd.

Sl. 10.29

Lorenzov grafikon za število živine v L R S



105.2 Časovne serije

Najpogosteje uporabljamo linijske diagrame za prikazovanje časovnih serij, zaradi tega so tudi načini in tehnika linijskih časovnih diagramov najbolj razviti. Osnovne značilnosti, ki jih mora linijski grafikon imeti, se ravnajo po zahtevi primerjave.

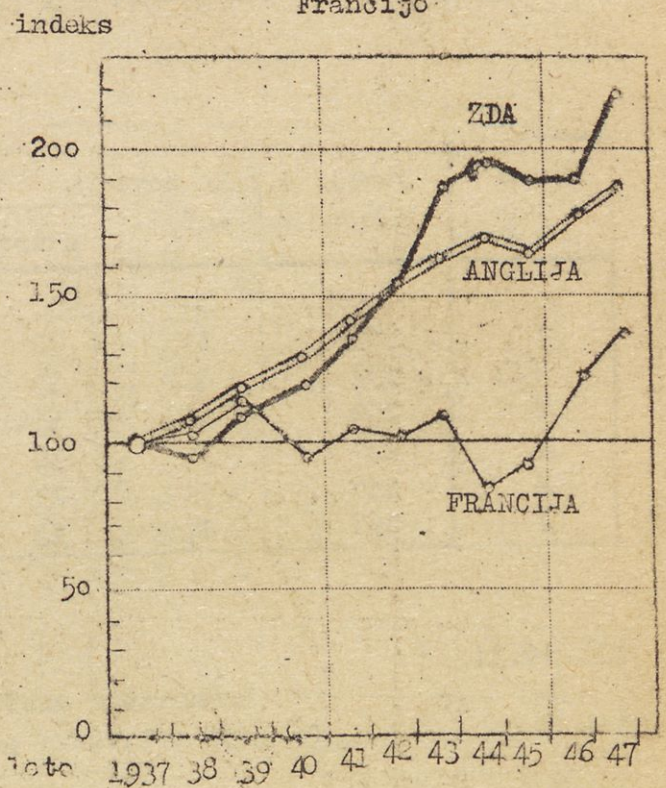
.21 Istovrstne časovne serije nimajo nobenih posebnih vprašanj in jih moremo brez nadaljnjega risati več na en in isti grafikon, kadar hočemo izvesti primerjavo med serijami, ker imajo eno in isto enoto mere.

Med istovrstne časovne serije štejemo tudi indeksne serije, ker so vsi indeksi neimenovana števila. Kot primer podajamo serije časovnih indeksov

proizvodnje električne energije v ZDA, Angliji in Franciji, ki smo jih izračunali pri relativnih številih (tab. 7.13).

Sl. 10.30

Indeks proizvodnje električne energije za ZDA, Anglijo in Francijo



Na časovnem grafikonu je dinamika pojava zelo dobro vidna, ker moremo naenkrat pregledati vse tri časovne serije. Točke rišemo sredi serije.

.22 Raznovrstne časovne serije, ki vsebinsko spadajo skupaj, morejo, prikazane na istem grafikonu zvičati medsebojno primerjavo. Vendar zaradi različnih enot mere, risanje na eni in isti sliki ni brez vsega mogoče. Da privedemo raznovrstne serije na istovrstne, izračunavamo in rišemo indeksne serije. Ako hočemo risati raznovrstne serije v absolutnih vrednostih, nastane tehnično vprašanje razmerja med merili za posamezne skale. Ako serija predstavlja neko zaključeno časovno enoto (n.pr. leto), običajno vzamemo med posameznimi skalami razmerje, ki se pojavi v povprečju. S tem dobimo tako razmerje med serijami, da kažejo časovne linije poleg dinamike tudi nesorazmerja med količinami glede na povprečne odnose.

.23 Brunsmanov grafikon. Časovno gibanje pojavov, ki se povezujejo kot stanje delavstva, odšli iz podjetja na novo došli, ali število prebivalstva, rojstva in smrti itd. moremo prikazati z grafikonom, ki je kombinacija linijskega grafikona in stolpcev. Te mase so med seboj povezane kot:

$$M_1 + R_1 - S_1 = M_2 ; \qquad (10.3)$$

novo stanje momentne mase (M_2) je enako staremu stanju (M_1) + intervalna masa, ki deluje na povečanje (R) -intervalna masa, ki deluje na zmanjšanje (S).

Kot primer podajamo gibanje delavstva nekega podjetja v polovici leta.

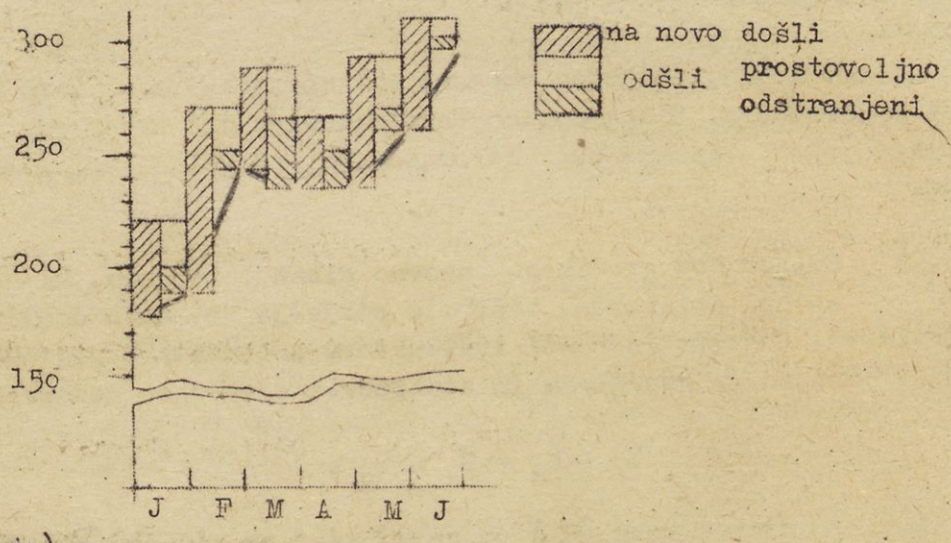
Tab. 10.8

Gibanje delavstva podjetja A (podatki izmišljeni)

Mesec	Stanje 1. v mes.	Na novo došli	odšli			Stanje zad. v mesecu
			skupno	prost.	odstr.	
J	181	43	32	20	12	192
F	192	83	25	21	4	250
M	250	41	51	20	31	240
A	240	31	30	15	15	241
M	241	54	27	21	6	268
J	268	47	15	10	5	300

Sl. 10.31

Brunsmanov grafikon
Gibanje delavstva podjetja A



Momentna masa je prikazana s točkami oz. linijami, intervalni masi (došli in odšli) pa s stolpci. V stolpcih moremo prikazati še strukturo , kot je vidno

za stolpce našlih.

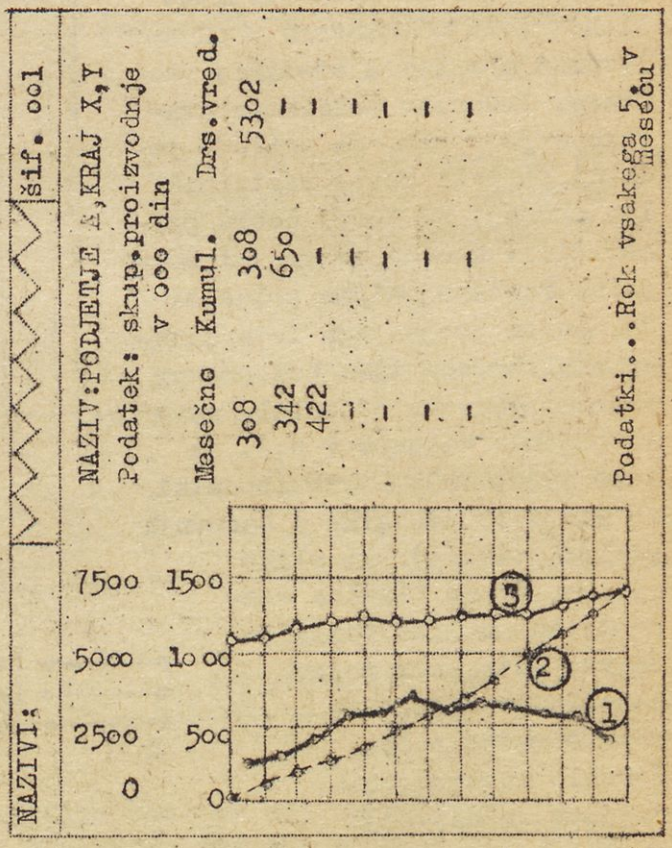
.24 Z - diagram. Poenotenje časovnih linijskih grafikonov je Z- diagram. Z-diagram je kartoteka, na kateri je poleg absolutnih podatkov časovne serije včrtan tudi grafikon časovne serije. Z diagrami vsebuje tabelo in sliko za naslednje serije:

a) absoluvnih vrednosti
b) kumulativnih vrednosti
c) drsečih vsot

(Glej poglavje 9 o časovnih serijah).

Sl. 10.32

Z - diagram



Te tri serije predstavljajo osnovne značilnosti dinamike časovnih serij. Zaradi značilne oblike, ki jo tvorijo gornje tri serije (črka Z, glej sliko 10.32) imenujemo te vrste diagrame Z- diagram.

Format (din A4) in oblika Z-diagram (grafikon v oglu) sta enotna. To in pa dejstvo, da je za vsako leto Omogoča in pojav poseben karton, da moremo izvršiti vsestransko primerjavo podatkov:

1) časovno dinamiko dobimo s polaganjem kartonov drugega poleg drugega (sl. 10.32 A);

2) sezonske primerjave dobimo s polaganjem kartonov drugega nad drugega (sl. 10.32 B);

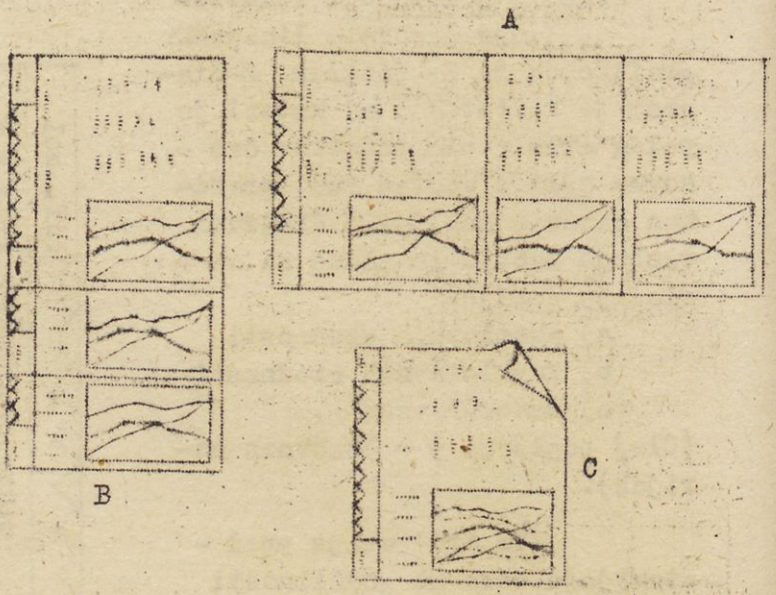
3) analitično primerjavo dobimo s polaganjem dveh kartonov drugega čez drugega (sl. 10.32 C)

.25 Polarni grafikon. Kadar hočemo proučevati periodičnost pojavov, večkrat uporabimo polarni koordinatni sistem. Cel krog pomeni razdobje ene periode. Krožnico razdelimo v sorazmerju delnih intervalov

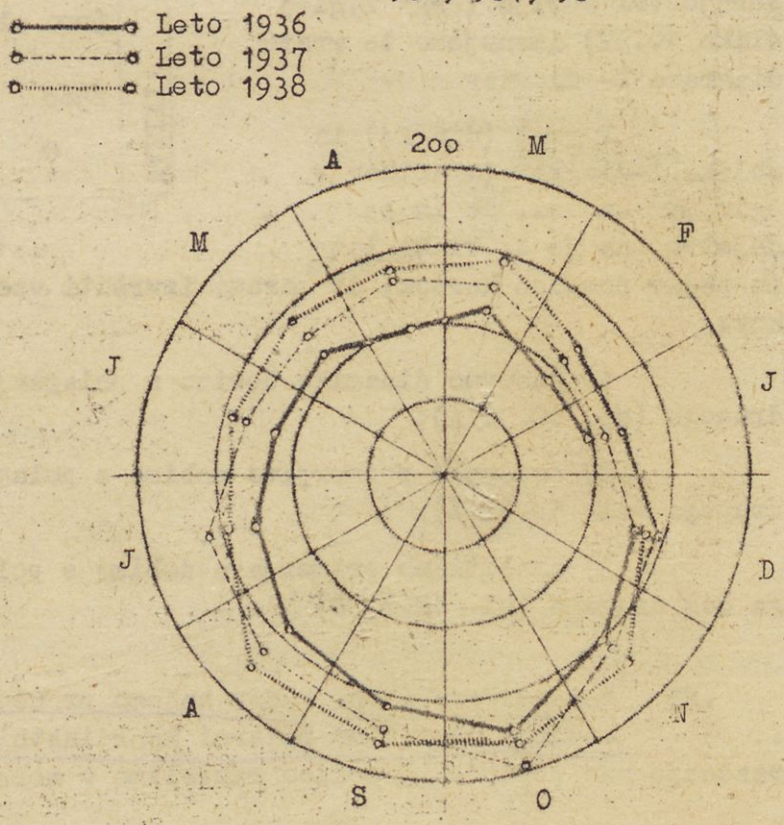
periode, običajno na dvanajst delov, kadar proučujemo (letne) sezonske spremembe. Časovno os tvori nadomestni krožnica, skala pa smer oziroma kot. Velikost členov serije pokažemo z oddaljenostjo točk od središča kroga. Prednost polarnega grafikona je v tem, da pridejo podatki za isti mesec različnih let skupaj, so torej dobro primerljivi. Obenem so meseci smiselno razporejeni in pride za decembrom prejšnjega leta januar naslednjega itd. Kot primer vzemimo podatke o številu natovorjenih vagonov v b.J. za leto 1936, 1937, 1938 (podatki so vzeti iz odstavka o časovnih serijah tab. 9). Iz slike vidimo, da se tipična ekscentrična oblika ponavlja za vsa leta. Ker pojav narašča, postajajo te oblike vedno večje. Število natovorjenih vagonov je največje v mesecih september, oktober, november, najmanjše pa v mesecih januar - februar.

Sl. 10.33

Vrste primerjav z Z-diagramom



Sl. 10.34 Število natovorjenih vagonov v b. Jugoslaviji v letih 1936, 1937, 1938



.26 Logaritmični diagrami

.261 Splošno Osnova statističnega proučevanja je, kot vemo, primerjava. V statistiki dajemo iz vzrokov, ki smo jih spoznali pri relativnih številih, prednost primerjavi z

z razmerji - relativnimi števili pred primerjavo z absolutnimi diferencami. Razlika ordinat dveh točk na linijskem grafikonu pomeni absolutno razliko dveh členov statistične serije. Iz tega načina prikazovanja moremo odbrati relativne odnose samo posredno. (Ena točka je dvakrat bolj oddaljena od absisne osi kot druga itd.). Ako pa vzamemo kot ordinatno skalo (skalo vrednosti členov serije) namesto navadne aritmetične skale logaritmično skalo, dolžina razlike dveh ordinat predstavlja logaritem kvocienta primerjanih pojavov, torej relativni odnos ($\text{Log}\frac{A}{B} = \text{Log } A - \log B$). Zaradi te lastnosti se v statistiki zlasti za proučevanje dinamike pojavov z grafikoni pogosto uporablja namesto aritmetične skale logaritmična skala.

Logaritmično skalo sestavimo na ta način, da od izhodišča na osi nanesemo namesto absolutnih vrednosti logaritme vrednosti. Skalo ne označimo z vrednostjo logaritmov, temveč z odgovarjajočimi absolutnimi vredhostmi. Vzamimo kot primer vrednosti od 1 - 10.

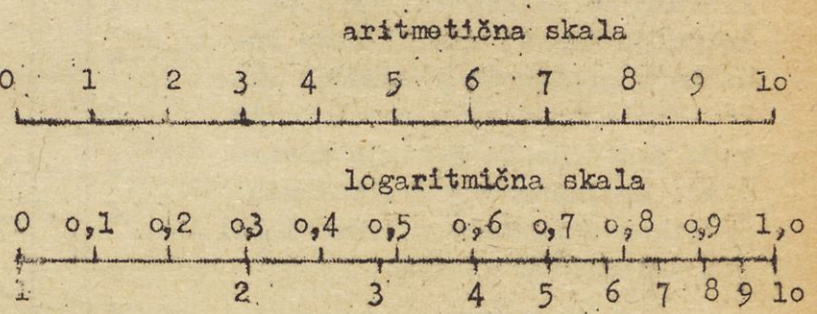
Tab. 10.9

$y = \log x$

x	Y
1	0,000
2	0,301
3	0,477
4	0,602
5	0,699
6	0,778
7	0,845
8	0,903
9	0,954
10	1,000

Sl. 10.35

Razlika med aritmetično in logaritmično skalo

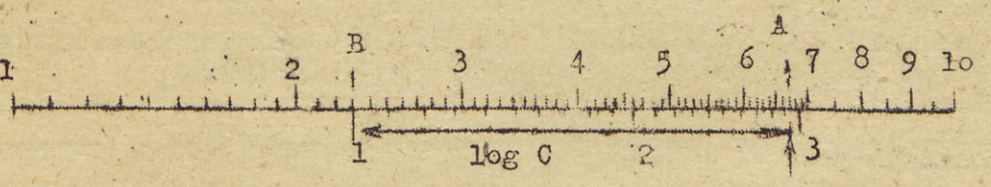


Na sliki 10.35 vidimo razliko med aritmetično in logaritmično skalo. Razdalje med posameznimi vrednostmi aritmetične skale so enake, ker je absolutna diferenca enaka, pri logaritmični skali pa so vedno manjše, ker se relativne razlike med dvema vrednostima manjšajo

n.pr. $\frac{2}{1} > \frac{12}{9}$

Da je logaritmična skala podrobnejša, moramo vnesti vanjo še vmesne vrednosti:

Sl. 10.36 Logaritmična skala



C = 2.94

ne

Jasno je, da bomo merili samo absolutne podatke v logaritmičnem merilu, temveč bomo morali z njim meriti tudi diference. Običajno imamo premakljivo pomožno logaritmično skalo, s katero moremo iz razdalje dveh primerjanih ordinat neposredno odčitati relativen odnos, ne pa logaritem odnosa. Na tej osnovi je zasnovano tudi logaritmično računalo.

P Odbrati je treba razmerje C med A = 67 in B = 2,3 s pomočjo pomožne logaritmične skale

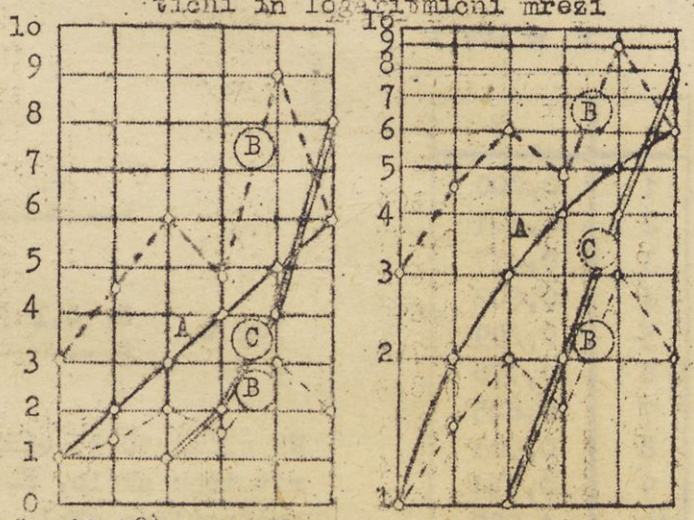
Sl. 10.37



Sl. 10.38

Iz logaritmičnih diagramov moremo odbrati vse vrste relativnih števil, strukture, koeficiente in indekse; oblike krivulj so seveda različne od krivulj, ako bi risali iste podatke v grafikonu z navadno skalo. Vendar par primerov in malo vaje zadošča, da senanje navadimo. Uporabnost logaritmičnih grafikonov je zelo velika zlasti pri dinamičnih serijah.

Slika serij podatkov na aritmetični in logaritmični mreži



Kot primer razlike oblike krivulj pri uporabi navadne ali logaritmične skale, podajamo nekaj shematičnih primerov (sl. 10.38)

Tab. 10.10

Pojav	Aritmetična mreža	Logaritmična mreža
A aritmetična postopnost	premica	Logaritmična kriv.
C geometrična postopnost	eksponencialna krivulja	premica
B Dva pojava z istimi razmerji	različne krivulje	enake, a premaknjene krivulje

.26. Logaritmični časovni diagram. Z logaritmični[m] diagrami moremo prikazati več istovrstnih ali raznovrstnih statističnih serij. Ker more ista skala služiti za raznovrstne serije, jo pustimo neimenovano, enoto mere pa naznačimo pri vsaki krivulji posebej.

Kot primer naj služi število prebivalstva in proizvodnja električne energije v ZDA v letih 1937-1947 (vir Mesečni bilten O Z N).

Tab. 10. 11

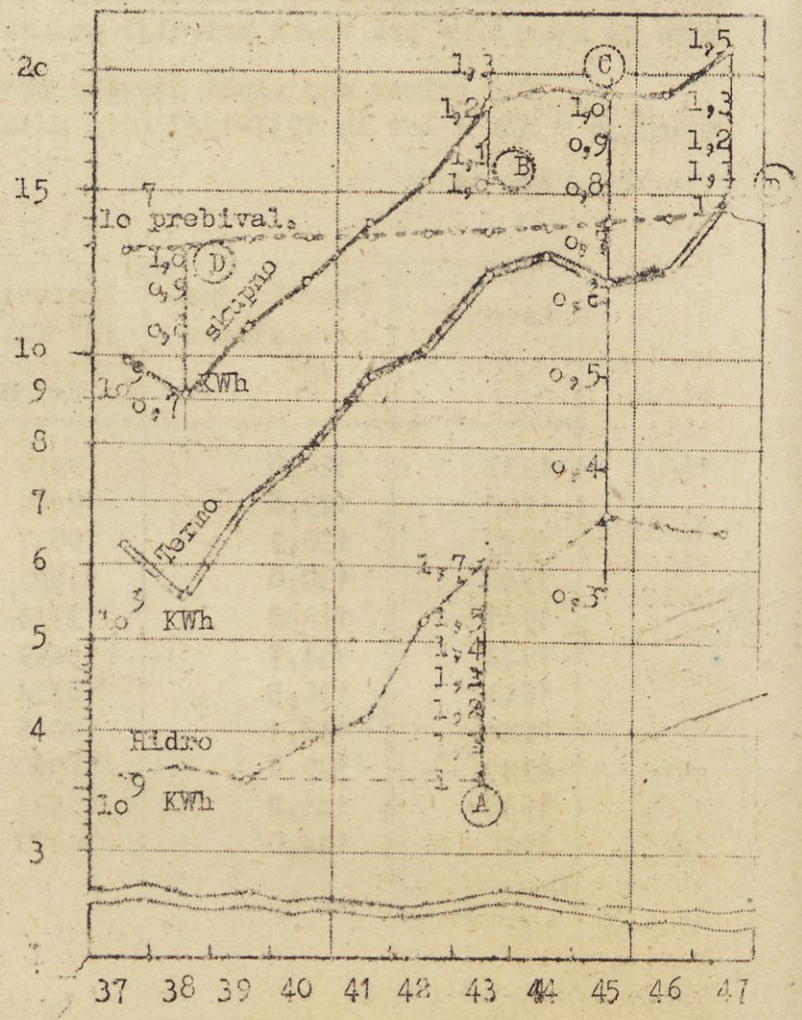
Leto	Število prebivalstva v 10^6	Proizvodnja elektr. energije v 10^6 KWh povpr. mesečno		
		Skupno	Hidro	Termo
1937	128,8	9910	3668	6242
1938	129,8	9484	3690	5794
1939	130,9	10637	3630	7007
1940	132,0	11829	3943	7876
1941	133,2	13733	4239	9494
1942	134,7	15499	5323	10176
1943	136,5	18414	6162	12252
1944	138,1	19016	6162	12854
1945	139,6	18540	6664	11876
1946	141,2	18594	6533	12061
1947	144,0	21310	6535	14775

Iz tega grafikona moremo odčitati s pomočjo premične skale za vsako serijo posebej indekse s poljubno stalno bazo (n.pr. $i = I_{43/39} = 170$), verižne indekse, (n.pr. $B = I_{43/42} = 123$) odstotne deleže (n.pr. 0,1945 H = 36 %, T = 64%), koeficient proizvodnje na 1 prebivalca (n.pr. $D^1 = 0,7 \cdot \frac{10^9}{10^7} =$ = 70 KWh/preb.; $D^2 = 1,48 \cdot \frac{10^9}{10^7} =$ = 148 KWh/preb.

Iz slike moremo tudi zaključiti, da raste prebivalstvo približno v geometrijski postopnosti, enako proizvodnja v letih (1948-1963) /glej sliko 10.38 in tab. 10.10/.

Sl. 10.39

Časoven razvoj proizvodnje električne energije v ZDA

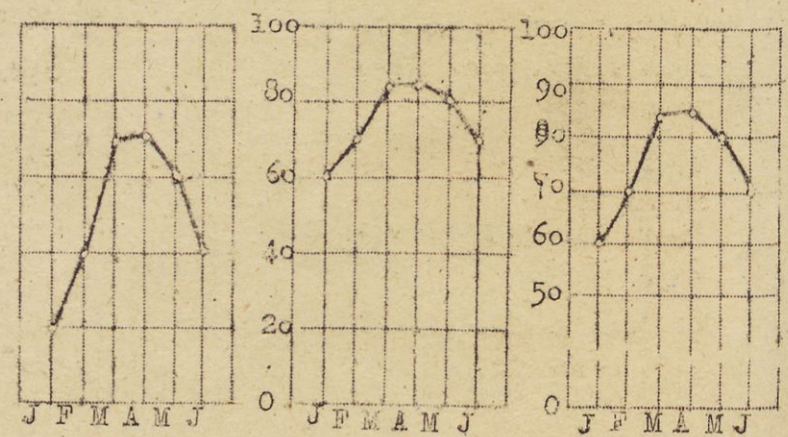


105.3 Pravila za risanje linijskih diagramov

Poleg splošnih pravil, ki smo jih navedli v začetku, moramo upoštevati, ako hočemo da bo linijski grafikon pregleden, še posebna pravila:

1) Navpična skala mora biti prirejena tako, da leži na skali tudi izhodišče. (Pri aritmetičnih ^{0} in pri

Sl. 10.40 Pravilo 1



logaritmičnih[1]). Če to ni mogoče, moramo narisati, kot da je mreža pretrgana, in pod njo dodati izhodišče

2) Izhodiščno črto, črto 100 pri indeksih in druge važnejše črte je treba risati debelejše od ostalih.

3) Če rišemo krivulje na logaritmičnem papirju, skušamo zaključiti skalo z 10, 100 itd.

4) Koordinatnih črt rišemo samo toliko, kolikor je nujno potrebno za preglednost.

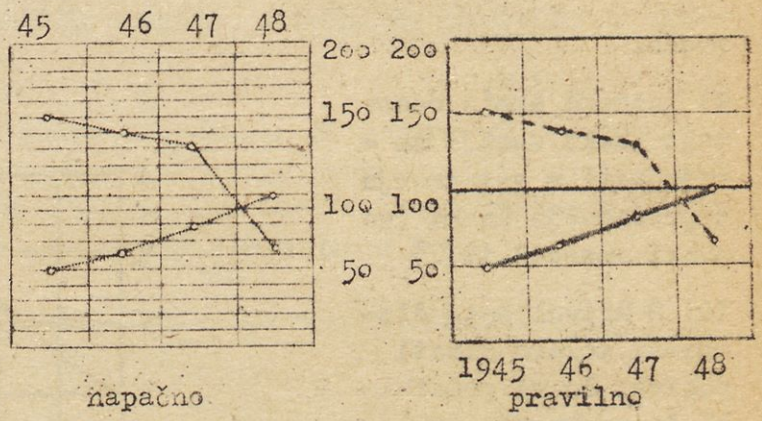
5) Če se časovni diagram ne nanaša na zaključene časovno razdobje, končni ordinati ne odebelimo. Ako se slika nanaša na različne dele časovne skale, moramo časovno skalo prekiniti na podoben način kot pri pravilu 1 za ordinato.

6) Skale običajno rišemo v grafikonu levo in spodaj.

7) Tolmačenje vodoravne skale damo na levi, navpične pa zgoraj.

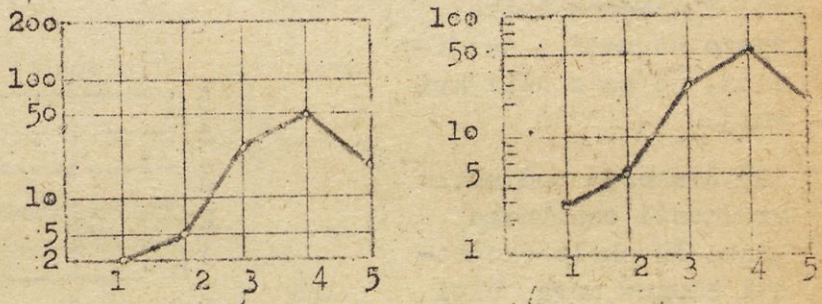
8) Črte pojava se morajo ostro ločiti od

Sl.1o.41 Pravilo (2),(4),(5),(6),(7) in (8)



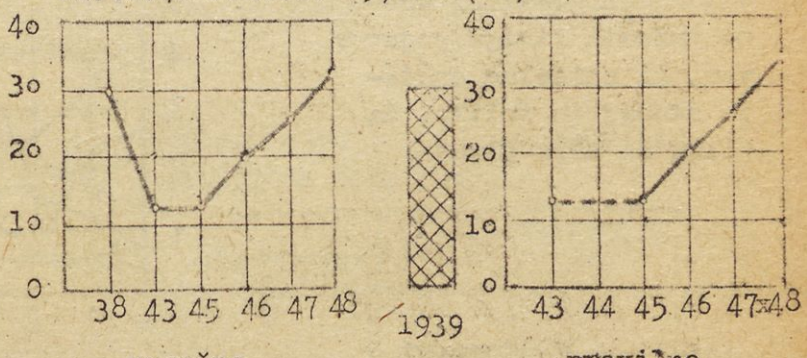
napačno pravilno

Sl. 1o.41a: Pravilo (3)



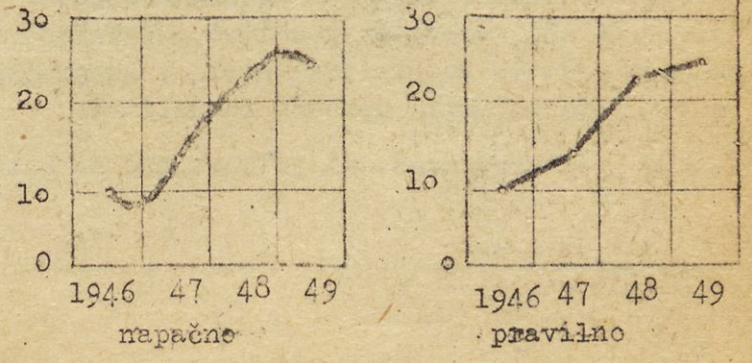
napačno pravilno

Sl.1o.42 Pravilo (9) in (1o)



napačno pravilno

Sl.1o.43 Pravilo (11)



napačno pravilno

mreže.

9) Razmaki skale osnovnega znaka serije morajo biti v sorazmerju z vrednostjo znaka (neenaki razredi itd.)

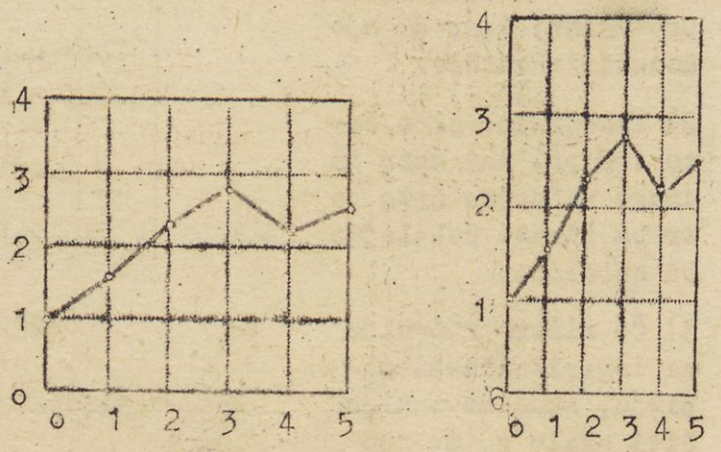
10) S krivuljčnim diagramom moremo risati pojave, če si podatki sledе na krajše razmake.

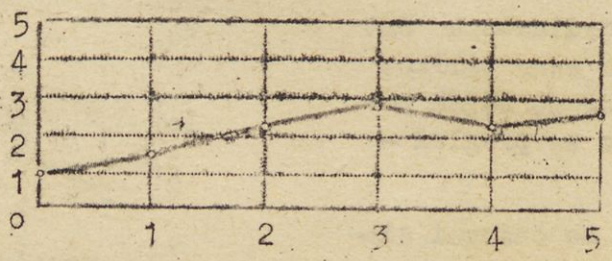
11) Posamezne točke smemo zvezati z daljicami, ne pa z deli krivulj.

12) Razmerje med skalo vrednosti osnovnega znaka in skalo vrednosti členov mora biti tako, da izgled ni pristranski.

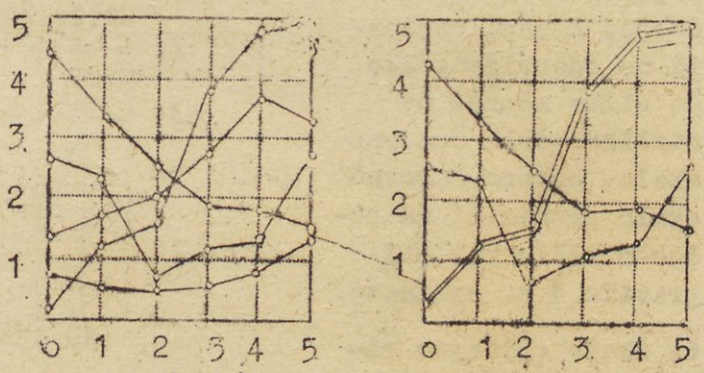
13) Na en grafikon smemo vrisati linije pojavov, ki smiselno spadajo skupaj in toliko, da slika ni nepregledna.

Sl.16.44 Pravilo (12)





Sl.16.45 Pravilo (13)



nepregledno pregledno

106. KARTOGRAMI

Krajevne geografske serije smo doslej prikazovali s serijami stolpcev. Vendar ta način ne nudi možnosti prikaza geografske razdelitve pojava. Zaradi možnosti proučitve geografsko razdelitve pojavov običajno prikazujemo podatke geografskih serij v geografskih kartah. "Mrežo" predstavlja v tem primeru karta upravne razdelitve.

Grafikone, ki prikazujejo na katerikoli način statistične podatke

v geografski karti imenujemo kartogram v širšem smislu. Te pa delimo v diagramske karte in prave kartograme.

166.1 Diagramska karta je zemljevid, v katerem so prikazani posameznim področjem pripadajoči podatki s stolpci, pravokotniki, krogi ali krivuljami. Ti diagrami so vnešeni v karto samo zaradi lažje krajevne primerjave. Pogosto prikazujemo v diagramski karti strukturo s krogi

Sl. 10.45 Struktura po zaposlitvi po republikah



Kot primer vzemimo aktivno prebivalstvo v FLRJ po zaposlitvi po republikah. (Vir Stat.bilten FLRJ št.1)

Tab. 10.12

Struktura aktivnega prebivalstva v FLRJ po zaposlitvi po republikah (Vir Stat.bilten FLRJ št.1)

LR	Srbija	Hrvatska	Slovenija	B i H	Makedonija	Črna gora
Število aktivnega prebivalstva v 000	4174	2475	862	1440	626	209
struktura sk.	100	100	100	100	100	100
Delavci	11	16	32	10	12	6
Nameščenci	7	8	11	6	8	8
Kmetje	77	71	48	81	75	78
Ostali	5	5	9	3	5	8

106.2 Pravi kartogrami

S Pravimi kartogrami prikazujemo krajevno gostitev pojava. Za risanje pravih kartogramov uporabljamo glede na to ali prikazujemo relativna ali absolutna števila kot element prikazovanja šrafure ali točke.

.21 Šrafure uporabljamo, kadar je pojav izražen v relativnih številih. Jakost pojava, oziroma velikost relativnega pojava v nekem področju je ponazorjena s svetlejšo ali temnejšo šrafuro. Običajno uporabimo za prikaz v kartogramu skalo šrafur z 7 do 11 odtenkov od belega polja, ki pokaže najmanjšo jakost do črne, ki pokaže največjo jakost pojava. Posamezna šrafura velja za razred vrednosti. Ti razredi so izbrani po predhodni proučitvi pojava, ki ga prikazujemo.

P Kot primer moremo prikazati gostoto prebivalstva v LRS po okrajih.

Tab. 10.13

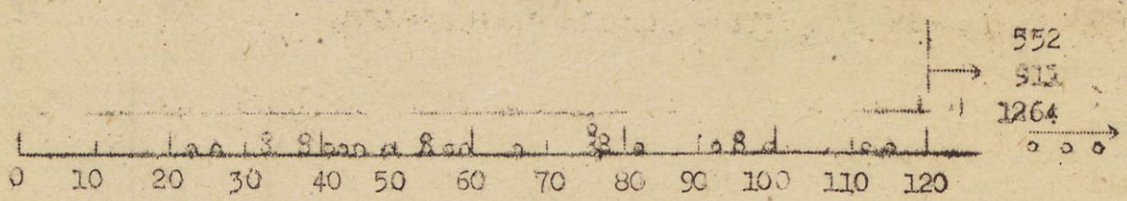
Gostota prebivalstva v LRS

Vir popis prebivalstva 15.3. 1948

Maribor mesto	1264	Gorica	66
Ljubljana mesto	911	Trebnje	59
Celje mesto	582	Novo mesto	57
Trbovlje	115	Ljubljana okol.	54
Lendava	112	Dravograd	54
Ljutomer	99	Grosuplje	49
Radgona	95	Mozirje	45
Murska Sobota	95	Jesenice	43
Ptuj	92	Črnomelj	41
Poljčane	82	Sežana	38
Krško	77	Idrija	38
Šalje okol.	77	Ilir. Bistrica	33
Kamnik	76	Postojna	33
Maribor okol.	76	Tolmin	26
Kranj	76	Kočevje	23

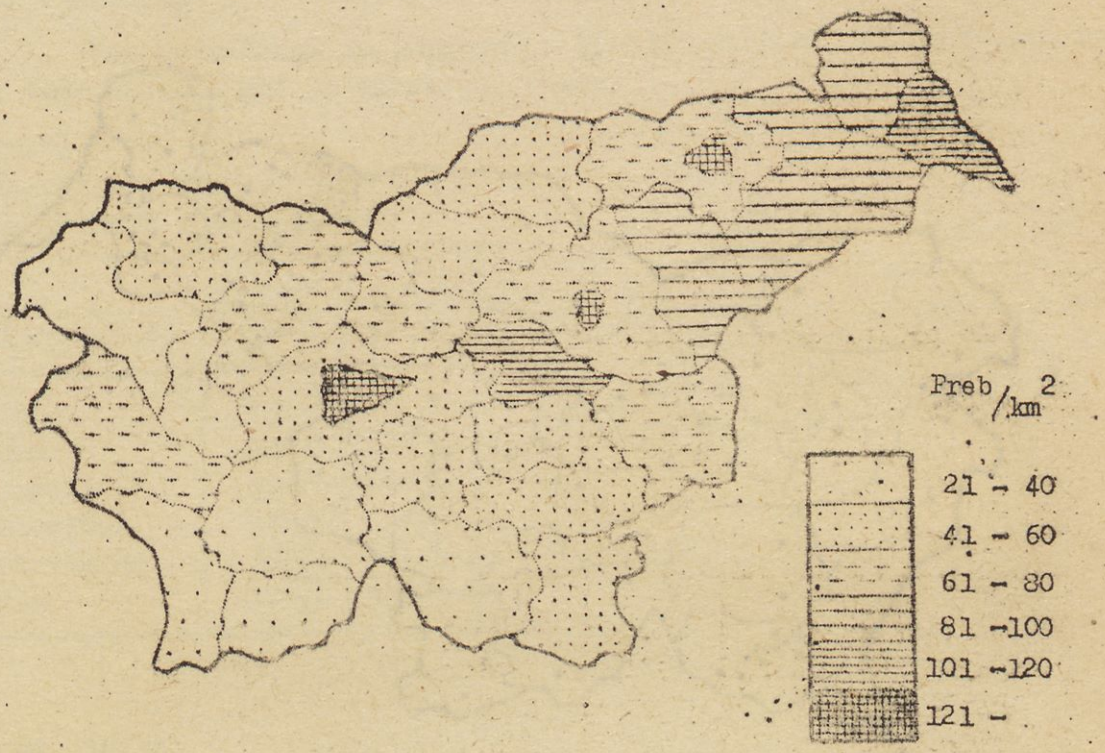
Sl. 10.46

Proučitev vrednosti in določitev razredov za šrafure



Sl. 10.47

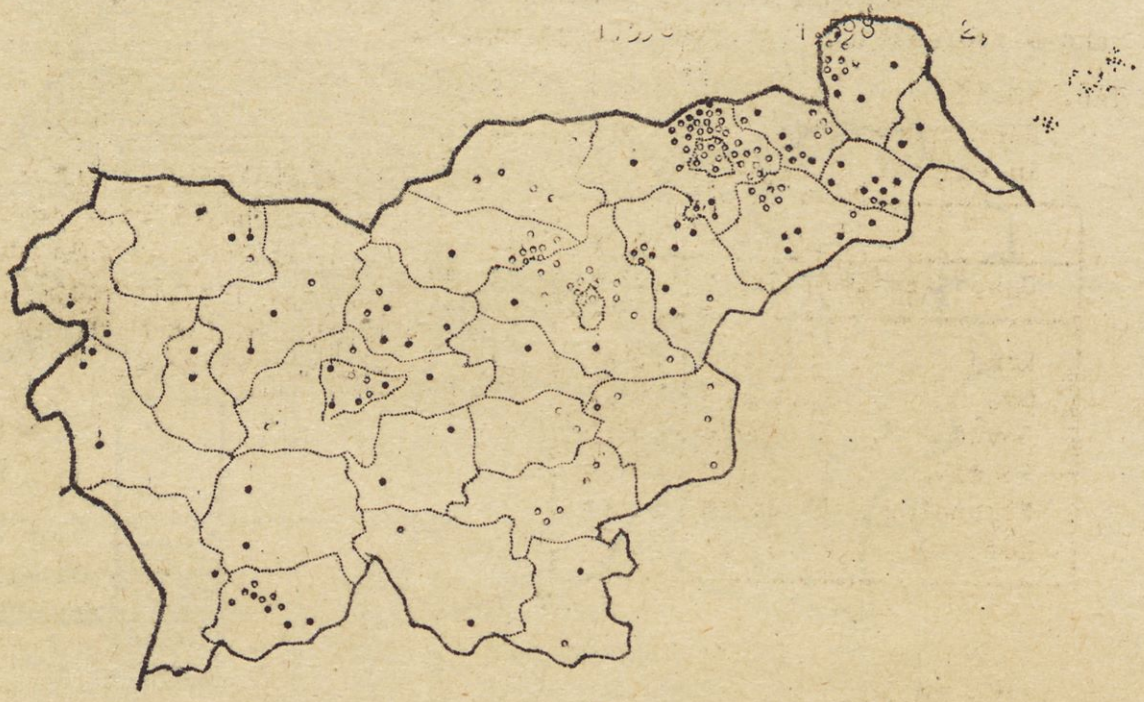
Gostota prebivalstva v LR Sloveniji po popisu 15.3. 1948



.22 Krajevno razdelitev absolutnih podatkov določenega pojava moremo prikazati s točkami. Točka pomeni v tem primeru določeno okroglo količino ali število pojava. Točke vrišemo v geografsko karto številčno v sorazmerju z velikostjo pojava, točno na mesta, kjer pojav obstoja. Ako rišemo kartogram sadnega drevja n.pr. jablan in pomeni vsaka točka 1000 jablan, bomo v kraj, kjer je število jablan 7.000, vnesli 8 pik v tisti del kraja, v katerem dejansko rastejo jablane.

Običajno rišemo kartograme s točkami na osnovi podatkov za najmanjše upravne enote (pri nas kraje) in v sorazmerju z absolutnim številom odgovarjajoče število točk pravilno razmestimo znotraj kroja. Kot primer podajam regionalno razdelitev števila jablan v LRS.

Sl. 10.48 Krajevna porazdelitev števila jablan v LRS



107. FIGURE

Nekateri načini grafičnega prikazovanja so nestrokovnjaku brez posebnih pojasnjevanj težko razumljiva. Zaradi tega uporabljamo, kadar je grafikon namenjen ljudem brez znanja statistike, kot osnovno sestavino grafičnega prikazovanja priproste figure. Tak grafikon je tem boljši, čim lažje je razumljiv.

Slike uporabljamo za prikazovanje statističnih podatkov na tri načine:

a) poleg običajnega grafikona narišemo sliko, ki nazorno kaže proučevan pojav.

b) velikost figur rišemo v sorazmerju z velikostjo pojava, ki ga predstavljajo

c) enotna figura pomeni določeno število enot. Velikost podatka je izražena na ta način, da narišemo podatku odgovarjajoče število figur.

Drug način je slab zaradi tega, ker bi morali velikost figur po pravilu risati v sorazmerju s prostornino, ker imajo figure tri razsežnosti.

Zaradi tega je primerljivost teh figur zelo slaba. Ako vzamemo sorazmerno podatku višine ali ploščine figur, pa je prikaz napačen.

Tem neprilikam se izognemo s tretjim načinom, ki je veliko boljši, tako s teoretičnega kot praktičnega vidika.

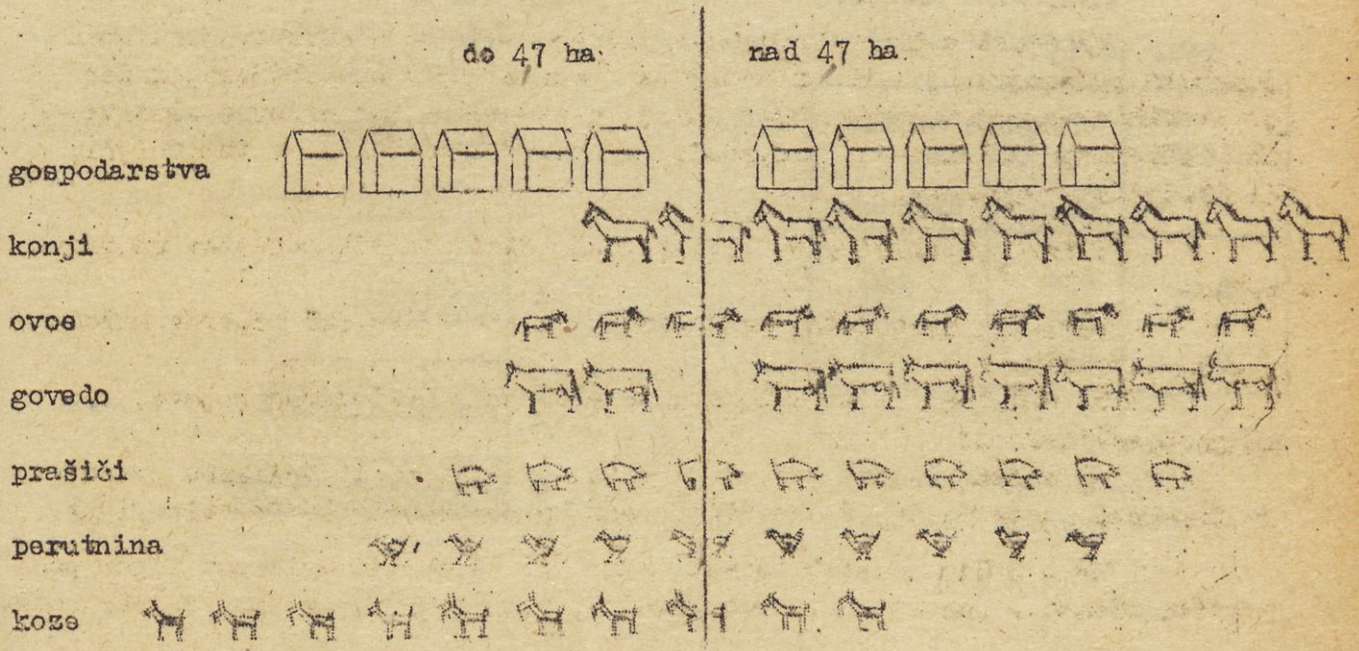
Tab. 10.14
Razmerje v živini v LRS pri malih in velikih gospodarstvih

	-47 ha	47 ha
Gospodarstev	50°/o	50°/o
Konj	15°/o	85°/o
Ovc	24°/o	76°/o
Goved	32°/o	68°/o
Prašičev	37°/o	63°/o
Perutnine	46°/o	54°/o
Koz	77°/o	23°/o

Razmerje v živini v LRS med malimi in velikimi gospodarstvi je različno. Po podatkih popisa 15.dec. 1947 (odbrano iz Lorenzovega grafikona sl.10.29) je slika naslednja.

Sl. 10.49

Razmerje v živini med petdeset odstotkov malih in petd. odstotkov velikih gospodarstev
1 figura = 10 °/o



Iz slike je razvidno, koliko odstotkov celotnega števila posameznih vrst surovine ima petdeset odstotkov malih in koliko petdeset odstotkov velikih kosov.

108. GRAFIČNA KONTROLA PLANA

108.0 Splošno

Učinkovita kontrola plana zahteva do podrobnosti izdelano in obširno evidenco, ki takoj opozori na vso odklone od planskega poteka in sicer na vseh stopnjah dela. Ker more sumarna evidenca zabrisati določene individualne odklone, ki se med seboj izenačijo, je potrebno, da gre evidenca plana prav do posameznega delovnega mesta. Ta velika množica posameznih podatkov pa je uporabna šele, ako jih uredimo in izračunamo posebne količine, (odstotek izpolnjenja za časovni interval, kumulativni odstotek, odstotek izpolnjenja po asortimentu itd.) Ker uporabljamo grafične metode, kadar skušamo primerjati večje število podatkov, moremo z grafičnimi metodami izvajati tudi kontrolo plana. Načini pa bodo glede na svojevrstnost podatkov posebni. Naloga grafične kontrole plana je, da spremlja izvajanje plana

1) količinsko
2) časovno roki)
3) po razmerju med posameznimi sestavinami (delovna sila-surovine- proizvodnja itd.) Pogoj, ki ga stavljamo na grafikon kontrole plana, je jasno registriranje odklonov od plana.

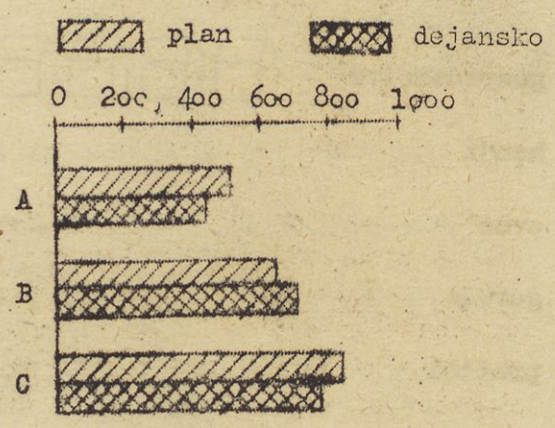
108.1 Enostavno grafično kontrolo po količini za istovrstne elemente dobimo, ako glede na količinsko skalo nizamo dvojice stolpcev za količino po planu in za dejansko:

P Pregleden primer:

Tab. 10.15

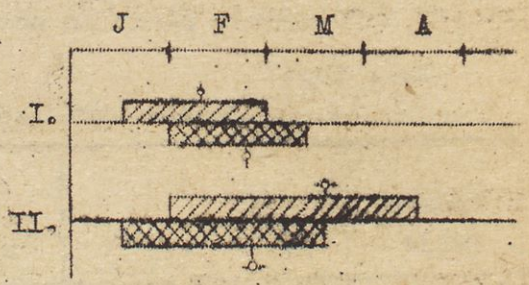
Podjetje	Vrednost proizv. v 000 din	
	plan	dejansko
A	500	430
B	650	700
C	830	780

Sl. 10.50
Primer kontrole plana po količini

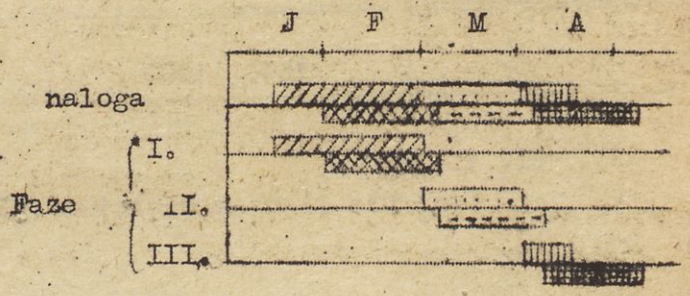


108.2 Kontrola po rokih izvedemo grafično, ako vpeljemo v grafikon namesto količinske skale časovno skalo. Ta grafikon pa omogoča kontrolo šele ob izvršitvi naloge. Delno moremo zvečati njegovo uporabnost z dogovornimi znamenji (sl.10.51) ali razbitjem na roke po fazah. (Sl. 10.52). Čas, v katerem je delo planirano ali izvedeno je v grafikonu naznačen s pasom.

Sl. 10.51
Primer kontrole plana po rokih



Sl. 10.52
Primer kontrole plana po rokih po fazah



108.3 Krivuljčni diagrami kontrole plana

Enostaven diagram kontrole plana dobimo, ako risemo časovni diagram, v katerega včrtavamo krivuljo plana in krivuljo dejansko izvršenega. Ako je črta dejansko izvršenega pod črto plana, plan za tisto razdobje ni izpolnjen in obratno. (Glej primer tabele 10.16 in slike 10).

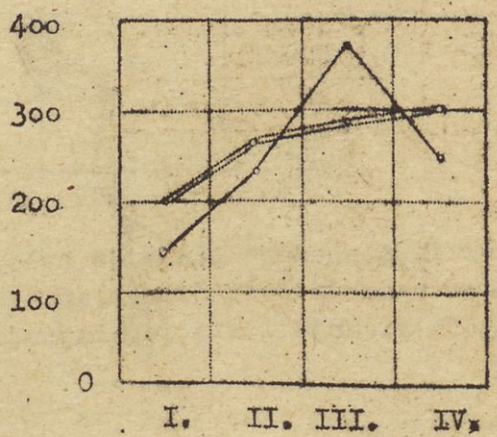
P primer:

Tab. 10.16

Tromesečje	Plan	Dejansko
I	200	150
II	250	240
III	280	360
IV	300	250

Z eno samo črto moremo izpolnitev plana prikazati s krivuljo procenta izpolnitve plana, bodisi za razdobje kumulativno, ali glede na odlotni plan. Ti procenti so izračunani v tabeli 10.17. V tej tabeli pomeni p_k četrtletni,

Sl. 10.53
Slika primera tabele 10.16



P_k pa kumulativni plan, d_k četrtletno, D_k pa kumulativno dejansko izvršenega. V koloni (5), (6) in (7) so zaporedoma podani četrtletni in kumulativni odstotek izpolnjenja plana in odstotek izpolnjenja plana glede na letni plan.

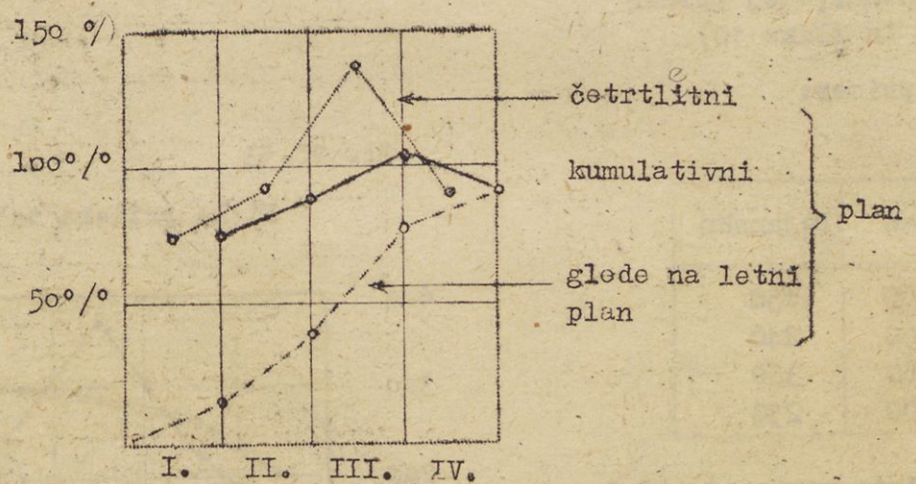
Tab. 10.17

Odstotki izpolnjenja plana

Tromesečje	p_k	d_k	P_k	D_k	$\frac{d_k}{P_k} \cdot 100$	$\frac{D_k}{P_k} \cdot 100$	$\frac{D_k}{\sum P_k} \cdot 100$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	200	150	0	0	75 %	-	0
II	250	240	200	150	96 %	75 %	15 %
III	280	360	450	390	129 %	87 %	38 %
IV	300	250	730	750	83 %	102 %	73 %
			1030	1000		97 %	97 %

Sl. 10.54

Kontrola plana z odstotki (slika primera iz tabele 10.17)



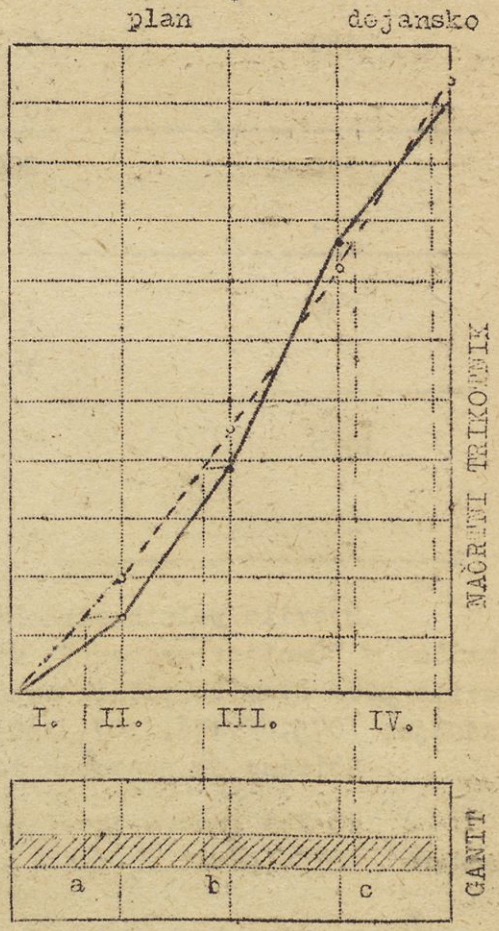
Odstotek izpolnjenega plana je zelo primeren za grafičen prikaz zaradi tega, ker imamo namesto dveh črt (plan - dejansko) samo eno (procent izpolnjenja plana). To omogoča risanje linij izpolnjenja plana za več podatkov na en in isti grafikon.

Načrtni trikotnik je krivuljčni diagram kumulativne serije plana in dejanskega. (V sliki 10.55 je načrtni trikotnik za primer v tabeli 10.16). I_me ima po svoji obliki. Iz prepletanja linije plana in izvršenega moremo sklepati na izpolnjenje plana. Načrtni trikotnik je dobro sredstvo kontrole. Edina hiba je obsežnost, ki onemogoča kontrolo več pojavov hkrati.

108.4 Ganttov grafikon

Ameriškemu racionalizatorju Ganttu je uspelo, združiti vse značilnosti načrtnega trikotnika v sliki, ki ima eno razsežnost -pasu. (Glej sliko 10.55). Dejansko izvršeno v nekem razdobju označi s časom, v katerem bi to količino izdelali, ako bi delali po planu. Ta čas je manjši od dejansko porabljenega časa, ako izvršeno zaostaja za planom,(glej sliko 10.55 a in b) in večji od dejanskega časa, ako je kumulativni plan prekoračen (glej sliko 10.55c). Grafična rešitev te naloge in enostavna slika Ganttovega grafikona je dana pri načrtnem trikotniku (v sliki 10.55 N amesto obsežnega načrtnega trikotnika imamo iste značilnosti podane v ozkem pasu.

Sl. 10.55
Zveza načrtnega trikotnika z Ganttovim grafikonom



To omogoča risanje velikega števila pojavov drugega za drugim in kompleksno analizo izvršenega plana.

.41 Izračunavanje Ganttovega procenta. Grafično določanje dolžin pasu Ganttovega grafikona bi bilo zelo zamudno. Zaradi tega bomo izdelali računski postopek za določanje vrednosti "Ganttovega procenta".

Vzomimo osnovni seriji p_k d_k in kumulativni seriji P_k D_k plana

Tab. 10.8

Četrtletje	p_k	P_k	d_k	D_k
I.	200	0	150	0
II.	250	200	240	150
III.	280	450	360	390
IV.	300	730	250	750
	1030	1030		1000

Za vsak člen kumulativne serije za dejansko izvršeno moramo poiskati koliko polnih in kakšen del naslednjega razdobja bi potrebovali, ako bi proizvodnja tekla točno po planu. 150 enot, kolikor so jih izdelali v prvem četrletju, bi morali izvršiti v treh četrtinah prvega četrtletja ($\frac{150}{200} \cdot 100 =$ = 75 %). 390 enot, kolikor so jih izvršili v prvih dveh četrtletjih bi morali izvršiti v prvem četrtletju. (200) in v 76 % drugega četrtletja ($\frac{390 - 200}{250} \cdot 100 = 76$ %) itd.

Račun je pregledno izvedon v naslednjem postopku

Dejansko kumulativno	Polna razdobja		deli naslednjega razdobja
150	0	+	$\frac{150}{200} \cdot 100$
	0		75 %
390	200	+	$\frac{390 - 200}{250} \cdot 100$
	1		76 %
750	730	+	$\frac{750 - 730}{300} \cdot 100$
	3		07 %
1000	730	+	$\frac{1000 - 730}{300} \cdot 100$
	3	+	90 %

Število polnih razdobij in procent naslednjega razdobja združimo tehnično v "Ganttov procent" ki je tromestno število, od katerega prvo število pove število polnih razdobij, ostali dve pa procent naslednjega razdobja, (075, 176, 307, 390)

Splošno je postopek izračunavanja Ganttovega odstotka naslednji:

a) Kot osnovo izdelamo tabelo osnovne serije plana p_k in kumulativne serije plana P_k.

Tab. 10.19

razdalje K	p_k	P_k	d_k	D_k
1	p_1	$0=P_1$	d_1	$0=D_1$
2	p_2	P_2	d_2	D_2
3	p_3	P_3	d_3	D_3
:	:	:	:	:
K	p_k	P_k	d_k	D_k
				D_{k+1}
r-1	p_{r-1}			
r	p_r	P_r		
r+1	p_{r+1}	P_{r+1}		
		P_{r+2}		

b) od razdobja do razdobja vpisujemo v tabelo podatke o dejanskem (d_k) in kumulativu dejanskega (D_k).

c) v kumulativni seriji plana poiščemo, med katere vrednosti pada D_{k+1}

$$P_r < D_{k+1} < P_{r+1}$$

d) r-1 je število celih intervalov Ganttovega procenta (prvo mesto Ganttovega procenta)

e) $\frac{D_{k+1} - P_r}{P_r} \cdot 100$ je procent naslednjega intorvala kinje še v Ganttovem procentu drugo in tretje mesto v Ganttovem procentu.

navaden procent izpolnjenja plana za periodo $\frac{c_k}{p_k} \cdot 100$.

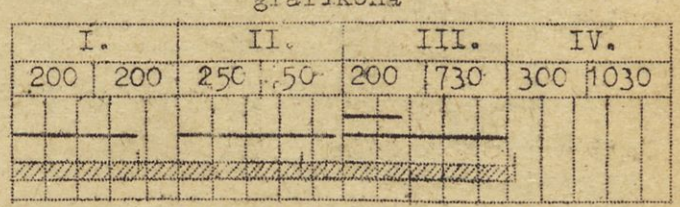
Oblika Ganttovega grafikona je enotna in so izdelane enotne mreže za Ganttov grafikon. Ganttov grafikon vsebuje naslednje sestavine

a) Vpisane vrednosti serije plana in kumulative plana,

b) navaden procent izpolnjenja plana,

c) Ganttov procent izpolnjenja plana

Sl. 10.56
Končna oblika Ganttovega grafikona



Za naš primer je slika Ganttovega grafikona dana v sliki 10.56, ki je narisan, kakšen izgleda ob koncu tretjega četrtletja. Posamezna razdobja so razdeljena na 5 delov po 20 %, da je lažje risanje in čitanje.

Tab.10.20

Element \ Dekada			I	II	III	I
Dobava surovin	Plan	dekadno	160	200	200	
		kumulativno	0	160	360	360
	De-jansko	dekadno	64	336	120	
		kumulativno	0	64	400	520
	Odstotek dekade		40 %	168 %	60 %	
	Ganttov odstotek		0	040 %	220 %	280 %
Poraba surovin	Plan	dekadno	160	200	200	
		kumulativno	0	160	360	560
	De-jansko	dekadno	64	156	60	
		kumulativno	0	64	220	280
	Odstotek dekade		40 %	78 %	30 %	
	Ganttov odstotek		0	040 %	130 %	160 %
Delovne ure	Plan	dekadno	800	1000	1000	
		kumulativno	0	800	1800	2800
	De-jansko	dekadno	440	960	400	
		kumulativno	0	440	1400	2000
	Odstotek dekade		55 %	96 %	40 %	
	Ganttov odstotek		0	055 %	160 %	220 %
Mezdni fond	Plan	dekadno	9600	12000	12000	
		kumulativno	0	9600	21600	33600
	De-jansko	dekadno	7200	9600	14400	
		kumulativno	0	7200	16800	31200
	Odstotek dekade		75 %	80 %	120 %	
	Ganttov odstotek		0	075 %	160 %	280 %
Proizvodnja	Plan	dekadno	80	100	100	
		kumulativno	0	80	180	280
	De-jansko	dekadno	24	56	30	
		kumulativno	0	24	80	110
	Odstotek dekade		30 %	56 %	30 %	
	Ganttov odstotek		0	030 %	100 %	130 %
Izvršitev naročil	Plan	dekadno	80	100	100	
		kumulativno	0	80	180	280
	De-jansko	dekadno	24	32	54	
		kumulativno	0	24	56	110
	Odstotek dekade		30 %	32 %	54 %	
	Ganttov odstotek		0	030 %	70 %	130 %

Z navpičnimi črticami zaznamujemo, do kam je šel stolpec Ganttovega grafikona za druga razdobja.

.42 Analiza s pomočjo Ganttovega grafikona. Najzanimivejša in obenem najvažnejša je analiza celotnega delovnega procesa s pomočjo Ganttovega grafikona. Oglejmo si primer, ki bo približno pokazal način primerjave. V tem primeru je za kontrolo plana seveda potrebno, da so znane norme za potrošnjo materiala, za delovno silo, kakor tudi za druge elemente delovnega procesa. Vzemimo, da za komad proizvodnje porabimo 2 kg osnovne surovine, 3 delavne ure kvalificiranega delavca (s povprečno mezdo 16 din), 4 delovne ure priučenega delavca (s povprečno mezdo po 12 din) in 3 delovne ure nekvalificiranega delavca (s povprečno mezdo 8 din na uro). Iz tega izvira, da porabimo za komad 10 delovnih ur, ki stanejo po planu 3 x 16+4 x 12+ +3 x 8 = 120 din.

Če je plan proizvodnje razdeljen na manjša razdobja na primer dekade, lahko iz norm preračunamo planirano porabo surovin, delovne sile za ta razdobja itd. Če vključimo še nabavo surovin in izpolnjevanje naročil, smo zajeli glavne dele proizvodnje od dobave surovin do izvršitve naročil.

Proizvodnja po planu naj doseže v prvi dekadi 80 enot, v drugi 100 enot in v tretji tudi 100. Ker poznamo norme, lahko iz teh podatkov izračunamo plan za fond plač, delovnih ur in porabe surovin. Ti podatki so vpisani v tabeli 10.20.

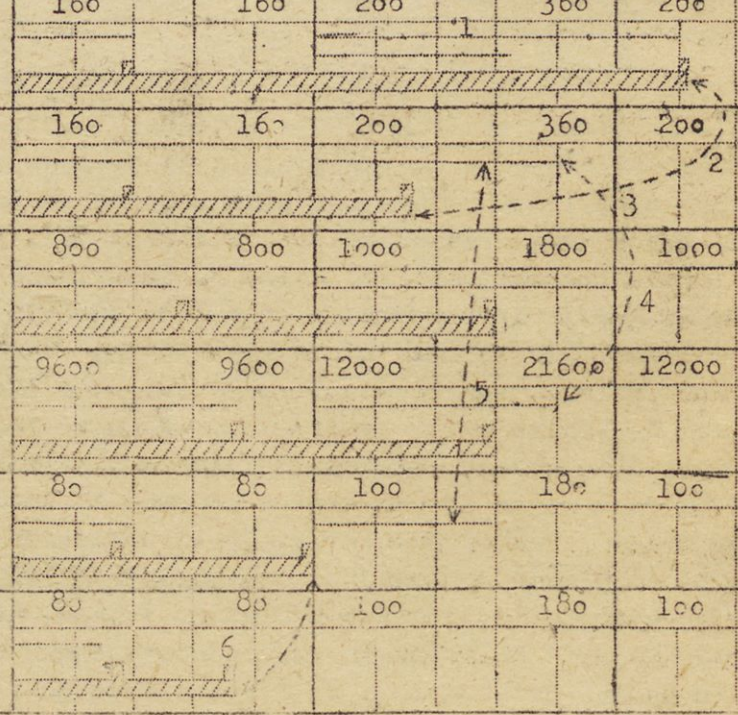
Grafikon ob koncu prve dekade je takšen, kakor je naznačeno v sliki 10.57. Če podamo analizo stanja prve dekade, pridemo do naslednjih zaključkov. Proizvodnja je znatno zaostala za plansko nalogo (1). Vzrok za to je verjetno pomanjkljiva dobava surovin, o čemer se da sklepati iz prve in druge črte, ker so bile vse dobavljene surovine tudi porabljene (2). Delovnih ur je bilo uporabljenih več kakor je predpisano po normi, kar je razvidno iz tega, ker je črta delovnih ur daljša od črte proizvodnje (3). Struktura delovne sile je bila nepravilna to je, za delo je bilo uporabljenih sorazmerno preveč kvalificiranih delavcev, kar lahko sklepamo iz razmerja črt med delavnimi urami in plačnim fondom. Primer sklepa iz uporabljenih skupnih delovnih ur, iz skupnega fonda plač na notranjo strukturo je zelo važen in je le eden izmed načinov uporabe te metode. Ta način se da uporabiti tudi drugod, tako n.pr. pri kontroli proizvodnje po skupinah. V tem primeru se po eni strani poslužujemo kontrole po količini, obenem pa po vrednosti. Če je razmerje sortimentov v skupini takšno, kakršno je planirano, sta obe črti, t.j. za količino in za vrednost, enako dolgi, pri izpremembah v sortimentu pa ena črta zastaja za drugo ali pa jo prehiteva. Za proizvodnjo je bilo porabljenih več surovin kakor je predpisano po normi, ker je črta proizvodnje zastala za črto porabe surovin (4). Naročila niso

Slika 1o.57

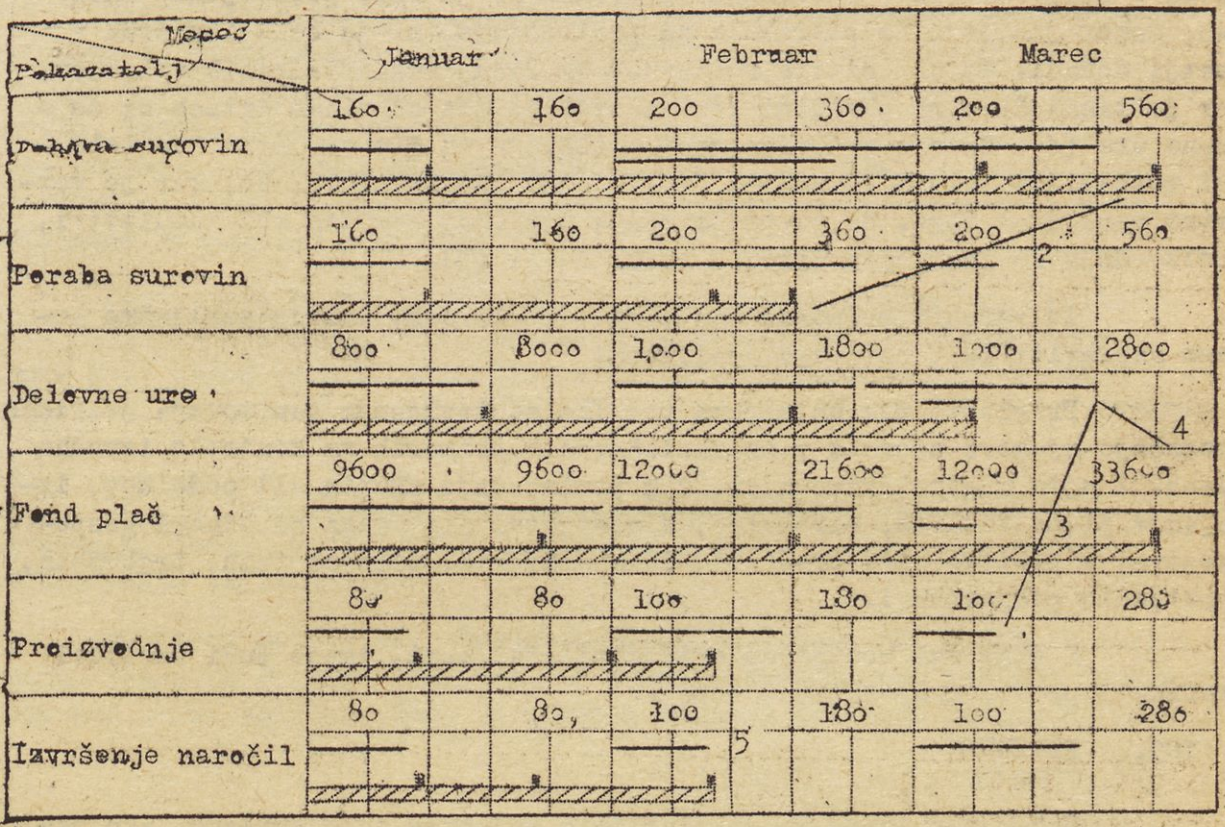
Mesec / Pokazatelj	Januar		Februar		Marec	
Dobava surovin	16o	16o	2oo	36o	2oo	56o
Poraba surovin	16o	16o	2oo	36o	2oo	56o
Delavne ure	8oo	8oo	1ooo	18oo	1ooo	28oo
Fond plač	96oo	96oo	12ooo	216oo	12ooo	336oo
Proizvodnja	8o	8o	1oo	18o	1oo	28o
Izvršenje naročil	8o	8o	1oo	18o	1oo	28o

Slika 1o.58

Mesec / Pokazatelj	Januar		Februar			
Dobava surovin	16o	16o	2oo	36o	2oo	56o
Poraba surovin	16o	16o	2oo	36o	2oo	56o
Delovne ure	8oo	8oo	1ooo	18oo	1ooo	28oo
Fondi plač	96oo	96oo	12ooo	216oo	12ooo	336oo
Proizvodnja	8o	8o	1oo	18o	1oo	28o
Izvršitev naročil	8o	8o	1oo	18o	1oo	28o



Slika 59



bila izvršena (5), vzrok za to je treba iskati v zastajanju proizvodnje, ker sta obe črti enako dolgi (6). Proizvodnost dela je znatno pod normo, kar lahko posnemamo iz črtic delovnih ur, fonda plač in proizvodnje, ker je črtica proizvodnje znatno krajša od črt delovnih ur fonda (7). S tem je analiza stanja za prvo dekado izčrpana.

Analiza stanja druge dekade (slika lo.58). Glavni vzrok neizpolnjevanja v prvi dekadi je bilo pomankanje surovin . Ta vzrok je bil v drugi dekadi odpravljen. Surovine so bile nabavljene v zadostni količini,kar kaže tanke črte nabavljenih surovin (1). Ob koncu druge dekade imamo še precejšnjo zalogo surovin, kar je razvidno iz črt nabave in porabe surovin (2). Delovnih ur je bilo porabljenih preko normo (3), struktura porabljena delovne sile je bila nepravilna (preveč nekvalificiranih delavcev) (4) verjetno zaradi tega, da bi se popravilo nepravilno stanje v prvi dekadi. Ta nepravilna struktura je imela za posledico, da se je pojavila velika množina odpadkov in izvržka, kar sklepamo iz tankih črtic porabe surovin in proizvodnje (5). Blaga je bilo prevzetega v obeh dekadah okrog 7o odst. količine, določene po planu za prvo dekado. Vzrok, da blago ni bilo prevzeto, je v tem primeru brzkone pri prevzemniku, ker ni bila

prevzeta niti količina, ki je bila razpoložljiva (6).

Analiza stanja tretje dekade (slika 10.59). Proizvodnja v tej dekadi je zelo majhna (30 odst. plana) (1). Vzrok za to ni pomanjkanje surovin (kumulativna črta nabavljenih surovin je daleč naprej od porabljenih surovin (2) niti pomanjkanje delovne sile (ur je bilo porabljenih veliko več, kakor jih je bilo določenih za proizvodnjo, ki je bila proizvedena v tretji dekadi). (5). Vzrok iz grafikona ni razviden. Ta vzrok bi bil n.pr. nepričakovano potrebno popravilo strojev, kar lahko označimo na sliki z znakom + pri proizvodnji. Delovnih ur je bilo porabljenih mnogo več, kakor jih je bilo potrebnih za proizvodnjo, ki je bila izvršena v tretji dekadi. Te ure so šle v glavnem na račun kvalificiranih delavcev, ker je tanka črta za plačilni fond znatno daljša od tanke črtice za delovne ure (4). Te ure kvalificiranih delavcev so šle verjetno za dela pri popravilu strojev. Naročila je podjetje izvrševalo slabo, kar je verjetno posledica tega, da je proizvodnja zastala, ker sta obe kumulativi, proizvodnje in izvršenih naročil enako dolgi (5).

Iz gornjih primerov smo videli v glavnem, kakšne zaključke moremo napraviti iz Ganttovega grafikona.

Prepričsta ter ustaljena pravila sestavkjenja Ganttovega grafikona omogočajo njegovo uporabo v najrazličnejših primerih za kontrolo izvedbe plana. Z njim kontroliramo n.pr. tok proizvodnje skupin ali oddelkov, izvajanje plana tovarne, direkcije, ministrstva ali vsega narodnega gospodarstva, v proizvodnji, preskrbi, industrijskih, transportnih, trgovskih, kmetijskih podjetjih i.p.

Kakor vsi kontrolno evidenčni grafikoni, imajo tudi Ganttovi grafikoni predvsem operativni pomen.

109. POVZETEK

Možnost primerjave statističnih podatkov zelo zvišamo, ako jih prikazujemo v grafični obliki. Grafikone delimo v dve veliki skupini :diagrame in kartograme.

Za prikazovanje statističnih podatkov se poslužujemo geometrijskih elementov: daljic, točk, kvadratov, pravokotnikov, krogov in tudi teles v treh razsežnostih. Velikost podatka posredujejo skale, podlaga grafikona pa je mreža grafikona, ki jo predstavljajo različni koordinatni sistemi. Katere podatke predstavljajo posamezni geometrijski elementi, naznačimo z različnimi črtkanji in šrafurami.

Najenostavnejši način prikazovanja statističnih podatkov so stolpci. S stolpci moremo prikazovati vse vrste statističnih serij. S stolpci moremo prikazovati tudi več istovrstnih ali raznovrstnih serij, serije struktur itd.

S pravokotniki prikazujemo podatke, ki so v zvezi A.B = C. Ena stranica ponazarja v tem primeru en podatek, druga stranica drugega, ploščina pravokotnika pa tretjega. To lastnost moremo uporabiti predvsem pri

večkratnih strukturalnih statističnih koeficientih.

Kroge uporabljamo za prikazovanje struktur. Vendar je njegova uporabnost omejena, ako prikazujemo serije struktur.

Točko moremo s pridom uporabiti posebno, kadar prikazujemo povezanosti med pojavi.

Ako točke v grafikonu zvežemo, dobimo linijski grafikon. Linijski grafikoni so izmed vseh vrst skoro najbolj uporabljani, posebno kadar prikazujemo časovne serije. Za njih izdelavo so sestavljena posebna ustaljena pravila, katera moramo upoštevati, ako hočemo, da bo grafikon brezhiben. Poleg običajnih časovnih diagramov, ki so risani bodisi na aritmetični ali logaritmični skali so linijski diagrami Lorenzov in Brunsmanov grafikon. Z diagrami in polarni grafikoni

Kartogrami so grafikoni, ki imajo za osnovo geografsko karto. Neprave kartograme dobimo, ako v mrežo administrativne razdelitve vnesemo katere koli izmed gornjih vrst diagramov. Pravih kartogramov pa imamo dvoje vrst. Relativne vrednosti podajamo s šrafurami, absolutne pa s točkami.

Figure uporabljamo, kadar so grafični prikazi namenjeni statistični nepoučenim ljudem. Najboljši je način, da podatek predstavlja določeno število enako velikih figur.

Grafične metode se z uspehom uporabljajo tudi pri kontroli plana. Posebno važen je Ganttov grafikon, ki omogoča vsestransko analizo izvajanja plana med samim potekom. Zato je njegova važnost predvsem operativnega značaja.

11. MERE VARIACIJE

110. SPLOŠNO:

Že pri splošnem obravnavanju srednjih vrednosti smo videli, da se vrednosti znakov od enote do enote spreminjajo, kot rečemo v statistiki, variirajo. Plača se spreminja od delavca do delavca, hektarski donos od parcele do parcele, število učencev od šole do šole itd. Ugotovili smo, da je vrednost znaka neke enote rezultat vseh vplivov, ki vplivajo na to enoto. Te vplive pa moremo razdeliti na dva dela. Prvi so vplivi, ki izvirajo iz splošnih, enotnih pogojev, ki se od enote do enote ne izpreminjajo; drugi pa so tako imenovani individualni vplivi, ki se od enote do enote menjajo. Splošne, enotne pogoje imenujemo opredeljujoče pogoje, ker je z njimi opredeljena statistična masa in so zato nujno stalni, ker so enaki za vse enote. Ugotovili smo, da rezultat splošnih, enotnih pogojev predstavlja srednja vrednost. Rezultat individualnih vplivov na eno samo enoto meri odklon od srednje vrednosti. Kajti, če individualnih vplivov ne bi bilo, bi enota imela vrednost, ki je enaka srednji vrednosti.

P Ako zaradi osvetlitve gornjega ponovimo primer plač devetih delavcev, ki smo ga navedli pri obravnavanju aritmetične sredine, vidimo, da je možno delitev plač v dva dela dejansko izvesti. Za navedene podatke je slika naslednja:

Tab. 11.1

Mesečna plača		Rezultat splošnih pogojev		individualnih pogojev
4527	=	4488	+	(+ 39)
4336	=	4488	+	(-152)
4628	=	4488	+	(+140)
4446	=	4488	+	(- 42)
4396	=	4488	+	(- 92)
4580	=	4488	+	(+ 92)
4425	=	4488	+	(- 63)
4552	=	4488	+	(+ 64)
4502	=	4488	+	(+ 14)
40392	=	9. 4488	+	0

Ako pogledamo na tem primeru rezultate individualnih vplivov, vidimo, da so od enote do enote različni. To je popolnoma jasno. Ako se menjajo od enote do enote individualni vplivi, se menja tudi rezultat teh vplivov.

Videli smo, da je rezultat individualnih vplivov na eno samo

40392 : 9 = 4488,
43
79

enoto možno količinsko izraziti. Ta rezultat izraža odklon plače posameznega delavca od srednje vrednosti. Ta odklon dobro podaja variacijo posamezne enote. Vprašanje je, ali je možno z enim številom izraziti variacijo vseh enot, oziroma celotne mase hkrati. Odgovor je pozitiven. Kot bomo videli, je možnih celo več mer variacije, od katerih pa so nekatere bolj, druge pa zopet manj uporabne.

111. VRSTE MER VARIACIJE

Izmed mer variacije bomo obravnavali samo najvažnejše in najbolj značilne. Med temi so:

1. Širina variacije.
2. Interquartilna vrednost.
3. Povprečen aritmetičen odklon.
4. Povprečen kvadratičen odklon ali varianca in standardna deviacija.

Najvažnejše merilo variacije, ki se v teoriji in praksi najpogosteje uporablja, je povprečen kvadratičen odklon ali varianca in iz nje izpeljana standardna deviacija. Zaradi tega jo bomo tudi najpodrobneje obravnavali.

112. ŠIRINA VARIACIJE

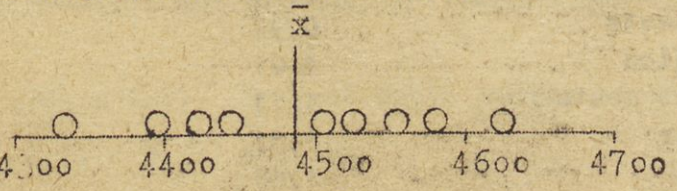
Kot najenostavnejše merilo variacije uporabljamo širino intervala, v katerem se gibljejo vrednosti določenega znaka neke mase. To merilo variacije imenujemo širino variacije. Njen iznos dobimo kot diferenco med največjo in najmanjšo vrednostjo.

$$W = x_{max} - x_{min} \qquad (11.1)$$

V našem primeru plač za devet delavcev iz 11.1 je najnižja plača, namreč 4336 din, najvišja plača pa 4628 din. Razpon, v katerem se dane plače gibljejo je 4628 − 4336 = 292 din. Širina variacije je torej 292 din.

Grafično moremo predočiti širino variacije lepo, ako narišemo distribucijo vrednosti

Sl. 11.1 Distribucija plač devetih delavcev



Širina variacije je kot merilo variacije zelo lahko določljiva in tudi razumljiva. Ima pa pomanjkljivosti, zaradi katerih je njegova vrednost zelo omejena. Širina variacije zavisi, kot je razvidno iz gornjega primera, samo od dveh vrednosti in še to od ekstremnih, katerih vrednost je

40397:9-44
43

rezultat izjemnih pogojev. Poleg tega je odvisna tudi od števila enot opazovane statistične mase. To je razvidno iz primera. Ako zgornjim devetim delavcem dodamo plače še drugih delavcev, je možno, da je plača nekega delavca še manjša od 4336 din in plača katerega delavca še večja kot 4628 din. Vidimo, da se širina variacije veča, ako večamo število enot v statistični masi.

113. INTERQUARTILNA VREDNOST

Da se izognemo pomanjkljivostim, ki izvirajo iz tega, da so pri širini variacije odločujoče skrajne vrednosti, ki za statistično maso niso tipične, se poslužujemo drugih mer variacije, ki skrajne vrednosti izključujejo. Ena izmed teh je interquartilna vrednost. Interquartilna vrednost je širina onega intervala, v katerem se nahaja polovica vseh vrednosti. Te vrednosti predstavljajo glavnino celotne mase.

Interquartilno vrednost izračunamo s pomočjo quartilov. Ako statistične enote uredimo po velikosti, mediana razdeli vse enote statistične mase v dva dela. Polovica enot ima manjše, polovica pa večje vrednosti od mediane (gl. odstavek 84).

Vsako tako nastalo polovico pa moremo naprej na enak način razdeliti v dva dela. Dobimo dve novi vrednosti, ki ju zaznamujemo z Q_1 in Q_3 in jima pravimo prvi in trotji quartil. Statistična masa je na ta način razdeljena na štiri dele. V vsakem delu je ena četrtina enot celotne mase. V intervalu $Q_3 - Q_1$ leži torej polovica vrednosti.

113.1 Izračunavanje iz individualnih podatkov

P Vzemimo kot primer strukturni delež nepismenih nad deset let starih po okrajih po popisu 15. marca 1948.

	Okraj	%
1.	Ljubljana mesto	0,60
2.	Kranj mesto	0,78
3.	Maribor mesto	0,84
4.	Celje mesto	0,97
5.	Radovljica	1.01
6.	Jesenice mesto	1.11
7.	Ljutomer	1,51
8	Radgona	1,68

$$Q_1 = \frac{1}{2}(1{,}68 + 1{,}74) = 1{,}71$$

9	Ptuj	1.74
10	Novo mesto	1.83

11.	Postojna	1.83
12.	Ljubljana okolica	1.90
13.	Gorica	1.95
14.	Trbovlje	2,13
15.	Kočevje	2.25
16.	Grosuplje	2.29

$$Q_2 = M_e = \frac{1}{2}(2{,}29 + 2{,}39) = 2{,}34$$

17.	Kamnik	2.39
18.	Krško	2.40
19.	Poljčane	2,44
20.	Tolmin	2,61
21.	Kranj okolica	2,64
22.	Celje okolica	2,66
23.	Črnomelj	2,67
24.	Trebnje	2,77

$$Q_3 = \frac{1}{2}(2{,}77 + 2{,}79) = 2{,}78$$

25.	Maribor okolica	2,79
26.	Idrija	3,41
27.	Lendava	3,61
28.	Murska Sobota	4,24
29.	Šoštanj	4,58
30.	Sežana	5,78
31.	Slovenj Gradec	6,63
32.	Ilir. Bistrica	6,67

Ker je skupno 32 okrajev, je mediana med vrednostjo za šestnajsti in sedemnajsti okraj. Po pravilu, ki smo ga obravnavali pri mediani, vzamemo kot mediano aritmetično sredino vrednosti za Grosuplje in Kamnik $Me = \frac{1}{2}(2{,}29 + 2{,}39) = 2{,}34$. Mediano moremo zaznamovati tudi s Q_2 in imenovati drugi quartil.

V prvi polovici imamo okraje od Ljubljana-mesta do Grosuplja, v drugi pa okraje od Kamnika do Ilir. Bistrice. Prvih šestnajst okrajev moremo prav tako razdeliti na dva dela po osem okrajev. Mediano spodnje polovice, ki jo imenujemo prvi quartil (Q_1), izračunamo v tem primeru na podoben način kot mediano, kot aritmetično sredino med vrednostjo za okraj Radgona in Ptuj.

$Q_1 = \frac{1}{2}(1{,}68 + 1{,}74) = 1{,}71$. Na enak način moremo izračunati tudi treji quartil: $Q_3 = \frac{1}{2}(2{,}77 + 2{,}79) = 2{,}78$.

Interquartilna vrednost je enaka:

$$Q_3 - Q_1 = 2{,}78 - 1{,}71 = 1{,}07\ \%.$$

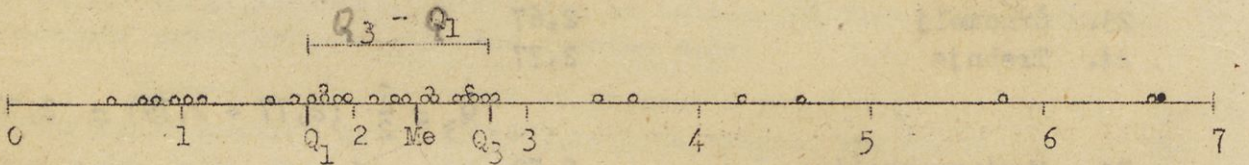
Poleg interquartilne vrednosti izračunamo tudi seminterquartilno vrednost, ki je enaka polovici interquartilne vrednosti:

$$\frac{1}{2}(Q_3 - Q_1) = \frac{1}{2} \cdot 1{,}07\ \% = 0{,}535\ \%.$$

Interquartilna vrednost ima lahko razumljiv vsebinski smisel in jo je lahko najti. Vrednosti, ki so manjše od prvega quartila in večje od tretjega quartila, analiziramo individualno, vrednosti v interquartilnem intervalu pa kot kolektiv, s tem, da povemo, v katerem intervalu se te vrednosti gibljejo. Da bo jasneje vidna distribucija in smisel quartilov in interquartilne vrednosti, podajamo primer v sliki.

Sl. 11.2

Distribucija okrajev po nepismenosti



Iz slike je razvidno, da skrajno velike vrednosti nastopajo zelo osamljeno in bi v tem primeru širina variacije dala zelo nestalno vrednost $w = 6{,}67 - 0{,}60 = 6{,}07\ \%$. Vendar kljub temu širine variacije ni zametovati, ker ima tudi določen pomen in poda celotno področje variacije.

V posplošeni obliki moremo podati določitev quartilov in izračunanje interquartilne vrednosti na naslednji način:

Q_1 je vrednost $\frac{n+2}{4}$ -ega člena po velikosti urejenih vrednosti

Q_2 je vrednost $\frac{2n+2}{4}$ -ega člena po velikosti urejenih vrednosti

Q_3 je vrednost $\frac{3n+2}{4}$ -ega člena po velikosti urejenih vrednosti.

n pomeni število enot celotne mase.

V našem primeru, kjer je bilo število okrajev $n = 32$ je bil Q_1 vrednost $\frac{32+2}{4} = 8\frac{1}{2}$ -ega člena, Q_2 vrednost $\frac{2 \cdot 32 + 2}{4} = 16\frac{1}{2}$ -ega člena, Q_3 pa vrednost $\frac{3 \cdot 32 + 2}{4} = 24\frac{1}{2}$ -ega člena. Ker $8\frac{1}{2}$, $16\frac{1}{2}$ in $24\frac{1}{2}$ niso cela števila, pomeni $8\frac{1}{2}$ aritmetično sredino vrednosti osmega in devetega člena, $16\frac{1}{2}$ aritmetično sredino med šestnajstim in sedemnajstim členom itd.

P Vzemimo, da je obseg mase 121. Po zgornjem pravilu je

Q_1 vrednost $\frac{121+2}{4} = 40\frac{3}{4}$ -tega člena po velikosti urejenih členov,

Q_2 vrednost $\frac{2 \cdot 121 + 2}{4} = 61$ -tega člena po velikosti urejenih členov

Q_3 vrednost $\frac{3 \cdot 121 + 2}{4} = 121\frac{1}{4}$ -tega člena po velikosti urejenih členov.

Na prvi pogled izgleda nesmisel $40\frac{3}{4}$ -ti člen ali $121\frac{1}{4}$ -ti člen, ker imamo dejansko štirideseti ali enainštirideseti člen itd. Vendar označimo kot vrednost $40\frac{3}{4}$ -tega člena vrednost, ki leži na treh četrtinah intervala med vrednostjo štiridesetega in enainštiridesetega člena.



To vrednost dobimo, ako vrednosti štiridesetega člena prištejemo $\frac{3}{4}(x_{41} - x_{40})$.

$$x_{40\frac{3}{4}} = x_{40} + \frac{3}{4}(x_{41} - x_{40})$$

Enako je vrednost $121\frac{1}{4}$ -tega člena vrednost

$$x_{121\frac{1}{4}} = x_{121} + \frac{1}{4}(x_{122} - x_{121})$$

113.2 Izračunavanje interkvartilne vrednosti iz frekvenčnih serij

Izračunavanje interquartilne vrednosti je v primeru, da so vrednosti grupirane v frekvenčni seriji, nekoliko drugačno. Enako kot v zgornjem primeru iščemo vrednosti:

$$\frac{n+2}{4}\text{-tega}, \quad \frac{2n+2}{4}\text{-tega}, \quad \frac{3n+2}{4}\text{-tega člena.}$$

Vendar pa se postopek skrajša z uporabo kumulativnih serij, na enak način, kot pri določanju mediane.

Izračunavanje interquartilne vrednosti za nezvezne in zvezne serije se izvrši v naslednjih stopnjah:

1. Izračunamo količine $\frac{n+2}{4}$, $\frac{2n+2}{4}$, $\frac{3n+2}{4}$

2. Izračunamo kumulativno serijo N_k $(N_{k+1} = N_k + n_k)$

3. V kumulativni seriji poiščemo člene, za katere je:

$$N_p < \frac{n+2}{4} < N_{p+1}, \quad N_d < \frac{2n+2}{4} < N_{d+1}, \quad N_t < \frac{3n+2}{4} < N_{t+1} \qquad (11.2)$$

(p pomeni prvi, d pomeni drugi, t pomeni pa tretji)

4. V tej točki se postopek za nezvezne serije različen od postopka za zvezne serije.

a) za nezvezne serije je:

$$Q_1 = x_p \qquad Q_2 = x_d \qquad Q_3 = x_t \tag{11.3}$$

b) za zvezne serije pa je:

$$Q_1 = x_{p,min} + \left(\frac{n+2}{4} - N_p\right) \cdot \frac{\Delta p}{n_p}$$

$$Q_2 = x_{d,min} + \left(\frac{2n+2}{4} - N_d\right) \cdot \frac{\Delta d}{n_d} \tag{11.4}$$

$$Q_3 = x_{t,min} + \left(\frac{3n+2}{4} - N_t\right) \cdot \frac{\Delta t}{n_t}$$

($x_{p,min}$, $x_{d,min}$, $x_{t,min}$ so spodnje meje odgovarjajočih razredov, Δ so širine razredov, n pa frekvence odgovarjajočih razredov.

5. Interquartilna vrednost je v obeh primerih enaka $Q_3 - Q_1$ semi-interquartilna vrednost pa $\frac{Q_3 - Q_1}{2}$.

Kot primer izračunavanja interquartilne frekvenčne distribucije vzemimo serije iz primera o aritmetičnih sredinah odstavek 81.62

Tab. 11.2

vel.skupina	A		B		C	
Število krav	- 2,00 ha		2,01 - 5,00 ha		5,01 - 10,00 ha	
x_k	n_k	N_k	n_k	N_k	n_k	N_k
0	250	0	43	0	42	0
1	185	250	127	43	69	42
2	128	435	182	170	165	111
3	33	563	127	352	208	276
4	7	596	54	479	139	484
5	3	603	15	533	71	623
6	1	606	3	548	21	694
7		607	-	551	8	715
8	-	607	-	607 551	2	723
9	-	607	-	551	1	725
$\sum$	-	607	-	551	-	726

A

1) $\frac{607+2}{4} = \frac{609}{4} = 152\frac{1}{4}$ $\frac{2607+2}{4} = \frac{1\,216}{4} = 304$, $\frac{3.607+2}{4} = \frac{1\,823}{4} = 455\frac{3}{4}$

2) Kumulativa izračunana zgoraj

3) $0 < 152\frac{1}{4} < 250$ $250 < 304 < 435$ $435 < 455\frac{3}{4} < 563$

$N_0 < 152\frac{1}{4} < N_1$ $N_1 < 304 < N_2$ $N_2 < 455\frac{3}{4} < N_3$

4) $Q_1 = 0$ $Q_2 = 1$ $Q_3 = 2$

5) $M_e = 1$ $Q_3 - Q_1 = 2$

B

1) $\frac{553}{4} = 138\frac{1}{4}$ $\frac{1\,104}{4} = 276$ $\frac{1\,655}{4} = 413\frac{3}{4}$

2) Kumulativa izračunana zgoraj.

3) $43 < 138\frac{1}{4} < 170$, $170 < 276 < 352$, $352 < 413\frac{3}{4} < 479$

$N_1 < 138\frac{1}{4} < N_2$ $N_2 < 276 < N_3$ $N_3 < 413\frac{3}{4} < N_4$

4) $Q_1 = 1$ $Q_2 = 2$ $Q_3 = 3$

5) $M_e = 2$ $Q_3 - Q_1 = 2$

C

1) $\frac{728}{4} = 182$, $\frac{1\,454}{4} = 363\frac{1}{2}$, $\frac{2\,180}{4} = 545$

2) Kumulativa izračunana zgoraj.

3) $111 < 182 < 276$, $276 < 363\frac{1}{2} < 484$, $484 < 545 < 623$

$N_2 < 182 < N_3$ $N_3 < 363\frac{1}{2} < N_4$ $N_4 < 545 < N_5$

4) $Q_1 = 2$ $Q_2 = 3$ $Q_3 = 4$

5) $M_e = 3$ $Q_3 - Q_1 = 2$

Skupni pregled da naslednjo sliko:

Tab. 11.3

Skupina	M_e	$Q_3 - Q_1$
A	1	2
B	2	2
C	3	2

P Kot primer za zvezno distribucijo dajemo teže po tisoč zrn za 373 parcel po m^2. Distribucija je že podana pri aritmetični sredini.

Tab. 11.4

	Teža v gr.	Frekvenca N_k	Kumulativa N_k
1.	nad 20 - 24	1	0
2.	" 24 - 28	12	1
3.	" 28 - 32	28	13
4.	" 32 - 36	96	41
5.	" 36 - 40	135	137
6.	" 40 - 44	65	272
7.	" 44 - 48	23	337
8.	" 48 - 52	7	360
9.	" 52 - 56	4	367
10.	" 56 - 60	2	371
			373

1) $\frac{n+2}{4} = \frac{375}{4} = 93\frac{3}{4}$, $\frac{2n+2}{4} = \frac{748}{4} = 187$, $\frac{3n+2}{4} = \frac{1\,021}{4} = 280\frac{1}{4}$

2) Kumulativa izračunana zgoraj

3) $41 < 93\frac{3}{4} < 137$, $137 < 187 < 272$, $272 < 280\frac{1}{4} < 337$

4) $Q_1 = 32 + (93\frac{3}{4} - 41)\frac{4}{96} = 32 + 52\frac{3}{4} \cdot \frac{4}{96} = 34{,}20$

$Q_2 = 36 + (187 - 137)\frac{4}{135} = 36 + 50\,\frac{4}{135} = 37{,}48$

$Q_3 = 40 + (280\frac{1}{4} - 272)\frac{4}{65} = 40 + 8\frac{1}{4}\,\frac{4}{65} = 40{,}49$

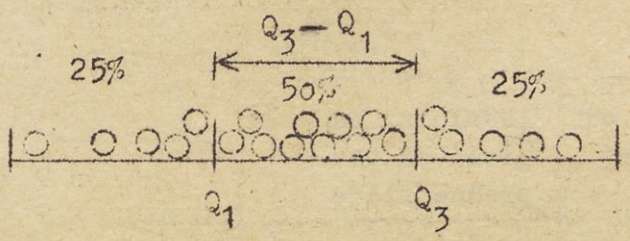
5) $Q_3 - Q_1 = 40{,}49 - 34{,}20 = 6{,}29$

113.3 Grafična ponazoritev interquartilnih vrednosti

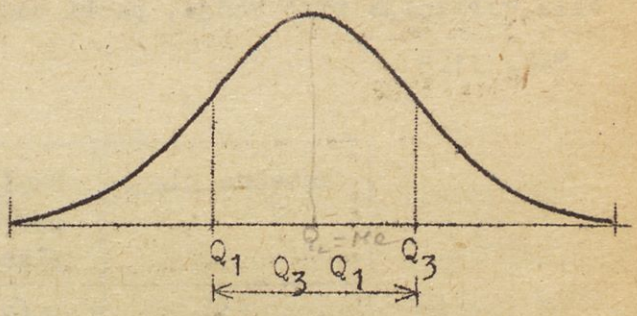
Grafično moremo ponazoriti smisel in pomen quartilov na treh slikah. Na prvi so narisani negrupirani podatki, na drugi frekvenčna distribucija, na tretji pa so dane kumulativne serije.

Sl.11.3 Določitev quartilov iz kumulativne serije

Pomen quartilov za individualne vrednosti

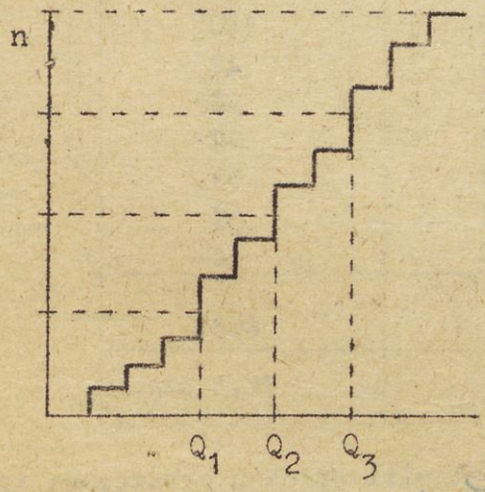


Sl.11. Pomen quartilov za frekvenčno distribucijo

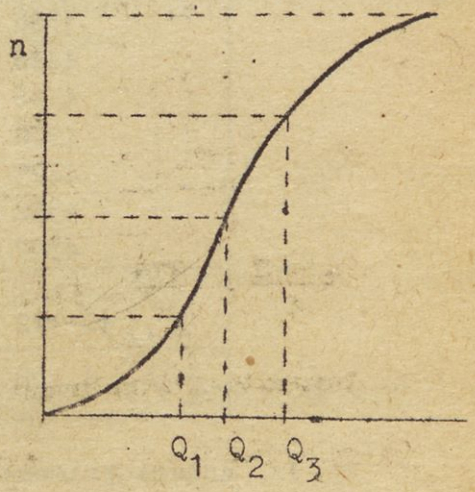


Sl.11.5 Določitev quartilov iz kumulativne serije odpade

a) nezvezni znak (iz tabele 11,2c)



b) zvezni znak (iz tabele 11,4)



aritmetični

114. POVPREČEN ABSOLUTEN ODKLON

Uvodoma smo navedli, da je individualen odklon od srednje vrednosti rezultat individualnih vplivov na dano enoto. Rekli smo, da moremo ta odklon uporabiti kot merilo jakosti delovanja teh vplivov. Vendar je individualno opazovanje teh odklonov zelo nepregledno, posebno v primeru,

ako je število enot veliko. Zaradi tega bomo izračunali povprečen odklon, ki bo predstavljal vse odklone v eni sami vrednosti. Jakost variiranja vseh enot bo torej količinsko izražena v enem številu s povprečnim odklonom. Vendar, ako podrobneje pregledamo naš primer plač devetih delavcev v tabeli 11.1, naletimo na težavo. V stolpcu (2) namreč vidimo, da povpreček odklonov nima smisla, ker je vsota odklonov enaka 0. Povprečen odklon moremo izračunati šele, ako se ne brigamo za smer učinka delovanja individualnih vplivov, temveč samo za njegovo absolutno jakost.

Kot merilo variacije bomo torej izračunavali povprečen absolutni odklon. V našem primeru je povprečen absoluten odklon 77.5 din. Ako bi bila variacija plač manjša, bi bil povprečen absoluten odklon manjši, ako pa bi bila variacija plač večja, pa bi bil povprečen absoluten odklon večji.

Tab. 11.5

Mesečna plača	Odkloni od povprečne plače	Absolutni odkloni od povprečne plače
(1)	(2)	(3)
x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
4527	+ 39	39
4336	- 152	152
4628	+ 140	140
4446	- 42	42
4396	- 92	92
4580	+ 92	92
4425	- 63	63
4552	+ 64	64
4502	+ 14	14
$\sum$ 40.392	0	698
$\frac{1}{n}\sum$ 4.488	-	77,5 din

Iz našega primera moremo sestaviti splošen obrazec za izračunavanje povprečnega absolutnega odklona.

$$\delta_{\bar{x}} = \frac{1}{n}\left[\lvert x_1 - \bar{x}\rvert + \lvert x_2 - \bar{x}\rvert + \lvert x_3 - \bar{x}\rvert + \ldots + \lvert x_n - \bar{x}\rvert\right] \qquad (11.5)$$

$$\delta_{\bar{x}} = \frac{1}{n}\sum \lvert x_i - \bar{x}\rvert$$

Navpične črtice pomenijo, da gre za absolutne vrednosti in da se na predznak ne oziramo.

Pri δ ,ki je znak za povprečen absoluten odklon, smo dali znak $\bar{x}$, kar pomeni, da so absolutni odkloni računani od aritmetične sredine. Odklone moremo namreč računati tudi od kake druge srednje vrednosti. Posebno važon je povprečen absoluten odklon, računan glede na mediano (δ_{Me}), ker je v tem primeru njegova vrednost izmed vseh najmanjša. Učinek kompleksa individualnih pogojev je v tem primeru najmanjši, v simbolih moremo to lastnost izraziti:

$$\delta_{Me} = \frac{1}{n}\sum_1 [x_i - Me] = \text{Min.} \qquad (11.6)$$

Ta lastnost se izkaže tudi v našem primeru:

Tab. 11.6

Mesečna plača	Absolutni odkloni od mediane
x_i	$[x_i - Me]$
4527	25
4337	166
4628	126
4446	56
4396	106
4580	78
4425	77
4552	50
4502	0
M_0 4502	$\sum$ 684
$\delta_{Me} = \frac{1}{9}\,684 = 76$	

δ_{Me} je v tem primeru za malenkost manjša od $\delta_{\bar{x}}$, kar pa ni slučaj, temveč izraz lastnosti, ki smo jo omenili zgoraj.

Iz frekvenčnih distribucij izračunavamo povprečen absoluten odklon tehtano,na ta način,da odklon od grupne vrednosti pomnožimo s pripadajočo frekvenco.

$$\delta_{\bar{x}} = \frac{1}{n}\sum_k n_k [x_k - \bar{x}] \qquad (11.7)$$

$$\delta_{Me} = \frac{1}{n}\sum_k n_k [x_k - Me] \qquad (11.8)$$

115. POVPREČEN KVADRATIČEN ODKLON ALI VARIANCA IN STANDARDNA DEVIACIJA

115.1 Varianca

Kot mera variacije se v statistiki najpogosteje uporablja povprečen kvadratičen odklon od aritmetične sredine. Z mednarodnim izrazom imenujemo to mero variacije varianca. Kot ime povprečen kvadratičen odklon pove, to mero variacije izračunamo na ta način, da tvorimo kvadrate odklonov individualnih vrednosti od aritmetične sredine in izračunamo povprečje iz teh kvadratov.

Po vzamemo naš primer plač, izračunamo varianco na naslednji način:

Tab. 11.7

	Mesečna plača	Odkloni od povprečne plače	Kvadratni odkloni
	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	4527	+ 39	1521
2	4336	- 152	23104
3	4628	+ 140	19600
4	4446	- 42	1764
5	4396	- 92	8464
6	4580	+ 92	8464
7	4425	- 63	3969
8	4552	+ 64	4096
9	4502	+ 14	196
$\sum$ 9	40392	0	71178
$\frac{1}{9}\sum$ 1	4488 $\bar{x}$	-	7909,7

$$V = \frac{1}{9} \cdot 71178 = 7909{,}7$$

V obliki obrazca moremo postopek napisati:

$$V = \frac{1}{9}\ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \ldots + (x_9 - \bar{x})^2$$

Splošen obrazec izračunavanja variance moremo po posplošitvi zgornjega pisati:

$$V = \frac{1}{n}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right] \qquad (11.9)$$

ali krajše:

$$= \frac{1}{n}\sum (x_i - \bar{x})^2 \qquad (11.10)$$

Operativna shema izračunavanja po tem obrazcu pa se glasi:

	1	x_i	$x_2 - \bar{x}$	$(x_i - \bar{x})^2$
	-	(1)	(2)=(1)- $\bar{x}$	(3) = (2)^2
	1	x_1	$x_1 - \bar{x}$	$(x_1 - \bar{x})^2$
	2	x_2	$x_2 - \bar{x}$	$(x_2 - \bar{x})^2$
$\sum$	n	$\sum x$	0	$\sum (x_i - \bar{x})^2$
$\frac{1}{n}\sum$	1	$\bar{x}$	-	$\frac{1}{n}\sum (x_i - \bar{x})^2 = V$

(11.11)

Stopnje izračunavanja so naslednje:

1. Sešteti vse vrednosti iz kolone (1)
2. Izračunati $\bar{x}$
3. Izračunati odklone $(x_i - \bar{x})$ kolona 2)
4. Kvadrirati individualne odklone (kolona 3)
5. Sešteti kvadrate odklonov
6. Vsoto kvadratov odklonov deliti s številom enot.

Čeprav je varianco razmeroma težko izračunati, je vendar zaradi svojih teoretičnih in praktičnih prednosti mera variacije, ki se, če že ne izključno, pa vsaj pretežno uporablja v statistiki. Skoro izključna pa je njena uporaba v višji statistični analizi, kjer je proučevanje variacije osnova vse analize. Zaradi tega stremimo edino za tem, da kolikor mogoče poenostavimo postopke izračunavanja variance.

115.2 Standardna deviacija

Do sedaj obravnavane mere variacije so imele iste enote mere, kot vrednosti, za katere smo mero variacije izračunavali. V našem primeru je bila širina variacije 292 din, interquartilna vrednost 141.25 din, povprečen absoluten doklon 77,5 din. Varianca pa ima enoto mere, ki je kvadrat osnovne enote mere podatkov, iz katerih varianco izračunavamo. To je razvidno iz tega, ker računamo kvadrate odklonov, ki imajo osnovno enoto mere. Odkloni od povprečne plače so dinarji, kvadrati teh odklonov pa kvadratni dinarji. Ta količina pa je logično nesmiselna in nepredstavljiva. Zaradi tega v podajanju končnih podatkov običajno izračunavamo kvadratni koren iz variance in ta kvadratni koren uporabljamo kot mero variacije, ker je kvadratni koren iz kvadratnih dinarjev zopet količina v dinarjih. Kvadratni koren iz variance imenujemo s tujim izrazom , ki pa se skoro izključno uporablja, standardna deviacija. Standardno deviacijo zaznamujemo z grško črko σ (sigma). Med varianco in standardno deviacijo torej obstoja zveza:

$$\sigma = \sqrt{V} \quad \text{ali} \quad V = \sigma^2 \tag{11.12}$$

Standardno deviacijo uporabljamo kot mero variacije tako pogosto, da običajno zaznamujemo varianco kar z znakom σ^2. S tem se izognemo večjemu številu simbolov, obenem pa je neposredno razvidna zveza med standardno deviacijo in varianco.

Obrazec izračunavanja standardne deviacije se po zgornjem glasi:

$$\sigma = \sqrt{\frac{\Sigma_i (x_i - \bar{x})^2}{n}} \tag{11.13}$$

P V našem primeru plač devetih delavcev je varianca enaka $\sigma^2 = 7909,7$, standardna deviacija pa je kvadratni koren iz variance:

$$\sigma = \sqrt{7909,7} = 89 \text{ din}$$

115.3 Postopek enostavnejšega izračunavanja variance in standardne deviacije iz negrupiranih podatkov

.31 Prva poenostavitev Izračunavanje odklonov individualnih vrednosti od aritmetične sredine in kvadriranje teh odklonov, ki so običajno večmestna števila je razmeroma zamudno.

Po obrazcu $$\sigma^2 = \frac{1}{n}\sum \overline{x_i^2} - \overline{x}^2 \qquad (11.14)$$

pa moremo varianco izračunati brez izračunavanja odklonov od aritmetične sredine. Od povprečnega kvadrata osnovnih vrednosti

$\frac{1}{n}\sum \overline{x_i^2}$, ki ga dobimo na ta način, da posamezne osnovne vrednosti kvadriramo, seštejemo in delimo s številom enot n, odštejemo kvadrat aritmetične sredine. Ta način v nekaterih primerih zelo olajša izračunavanje variance.

Operativna shema in postopek izračunavanja sta v tem primeru naslednja:

i	x_i	x_i^2
	(1)	(2) = (1)^2
1	x_1	x_1^2
2	x_2	x_2^2
⋮	⋮	⋮
n	x_n	x_n^2
$\sum$ n	$\sum x$	$\sum x^2$
$\frac{1}{n}\sum$ 1	$\bar{x}$	$\overline{x^2}$
$\sigma^2 = \overline{x^2} - (\bar{x})^2$		
$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2}$		

Stopnje izračunavanja so naslednje: (11.15)

1. Napišemo vsoto osnovnih vrednosti x_1 (kolona 1)
2. Posamezne vrednosti iz kolone 1 kvadriramo in vpišemo v kolono 2.
3. Seštejemo kolono 1 in kolono 2.
4. Vsoto v koloni 1 in koloni 2 delimo s številom enot. Na ta način dobimo aritmetično sredino in povprečen kvadrat.
5. Varianco dobimo, da od povprečnega kvadrata odštejemo kvadrat aritmetične sredine.
6. Standardna deviacija je kvadratni koren iz variance.

P Pri obravnavanju aritmetične sredine smo imeli v odstavku 81.31 primer 12 mater. Opazovali smo število otrok (x_i). Izračunati je treba standardno deviacijo števila otrok.

Tab. 11.8

i	x_i	x_i^2
	(1)	(2) = $(1)^2$
1	1	1
2	2	4
3	1	1
4	2	4
5	2	4
6	3	9
7	3	9
8	1	1
9	1	1
10	1	1
11	2	4
12	1	1
$\sum$ 12	20	40
$\frac{1}{n}\sum$ 1	1,67	3,33

$\sigma^2 = 3,33 - 1,67^2 = 3,33 - 2,79 =$

$= 0,54$

$\sigma = \sqrt{0,54} = 0,73$ otroka

.32 Uporaba pogojne sredine

Ako se vrednosti posameznih enot gibljejo okoli kake okrogle vrednosti, se izračunavanje variance zelo skrajša, ako vpeljemo pogojno sredino x_o. Pogojno sredino poznamo že iz poglavja o izračunavanju aritmetične sredine (odstavek 81.22). Koristnost uporabe pogojne sredine pa je pri varianci še znatno večja kot pri aritmetični sredini.

Enako kot pri aritmetični sredini izračunamo tudi sedaj najprej odklone posameznih vrednosti x_i pogojne sredine x_o. Te odklone zaznamujemo z u_i. Torej je $(x_i - x_o) = u_i$.

Velja splošen stavek, da je varianca prvotnega znaka x enaka varianci odklonov. CV simbolih moremo pisati, ako uporabimo način, ki smo ga uporabili zgoraj:

$$\sigma_x^2 = \sigma_u^2 = \overline{u^2} - (\bar{u})^2 \qquad u_i = x_i - x_o \qquad (11.16)$$

Ponavljamo, da je pogojna sredina poljubna, in jo izberemo tako, da so odkloni čim prikladnejše količine za izračunavanje:

Shema in postopek izračunavanja sta v tem primeru naslednja:

	i	x_i	$x_i - x_o = u_i$	u_i^2
		(1)	(2)=(1)-x_o	3= (2)^2
		x_1	$x_1 - x_o$	u_1^2
		⋮	⋮	⋮
$\sum$	n		$\sum u$	$\sum u^2$
$\frac{1}{n}\sum$	1		$\bar{u}$	$\overline{u^2}$

$$\sigma_x^2 = \overline{u^2} - \bar{u}^2$$

$$\sigma_x = \sqrt{\overline{u^2} - (\bar{u})^2}$$

(11.17)

Stopnje izračunavanja so naslednje:

a) Napišemo vrsto osnovnih vrednosti x_i (kolona 1).

b) Izberemo pogojno sredino x_o

c) Izračunamo odklone posameznih vrednosti x_i, dane v koloni 1 od pogojne sredine in jih vpišemo v kolono 2

d) Odklone iz kolone 2 kvadriramo in vpišemo v kolono 3.

e) Seštejemo vrednosti v koloni 2 in vrednosti v koloni 4 in vpišemo v vrsto zaznamovano z $\sum$.

f) Ti dve vrednosti delimo s številom enot n in vrednosti vpišemo v naslednjo vrsto. V koloni 2 smo dobili na ta način povprečen odklon od pogojne sredine $\bar{u}$, v koloni 3 pa povprečno kvadratno odklonov od pogojne sredine.

g) Varianco dobimo, ako od povprečja kvadratov odklonov odštejemo kvadrat povprečnega odklona.

Opozoriti moramo, da je povprečje kvadratov odklonov $\overline{u^2}$ nekaj povsem drugega kot pa kvadrat povprečnega odklona $\bar{u}^2$. Prvo količino dobimo, ako posamezen odklon kvadriramo in poiščemo povprečje iz teh kvadratov, drugo količino pa izračunamo na ta način, da izračunamo najprej povprečen odklon in ta povprečen odklon kvadriramo.

P Vzemimo kot primer vrednost proizvodnje nekega podjetja po mesecih v 000 din iz poglavja o aritmetičnih sredinah odstavek 81.22 in izračunajmo zanj standardno deviacijo:

Tab. 11.9

i	x_i	$x_i - 1285 = u_i$	u_i^2
	(1)	(2) = (1)- x_o	$(2)^2$
1	1287,6	+ 2,6	6,75
2	1286,5	+ 1,5	2,25
3	1282,4	- 2,6	6,76
4	1283,6	- 1,4	1,96
5	1283,2	+ 4,2	17,64
6	1288,9	+ 3,9	15,21
7	1286,4	+ 1,4	1,96
8	1288,5	+ 3,5	12,25
9	1284,6	- 0,4	0,16
10	1290,5	- 5,5	30,25
11	1291,2	+ 6,2	38,44
12	1292,4	+ 7,4	54,76
$\sum$ 12	-	+ 31,8	188,39
$\frac{1}{12}\sum$ 1	-	+ 2,65	15,6992

$$\bar{x} = 1285 + 2,65 = 1287,65$$

$$\sigma^2 = 15,6992 - (2,65)^2 = 15,6992 - 7,0225 = 8,6767$$

$$\sigma = \sqrt{8,6767} = 2,945$$

D Dokaz zgornjih dveh okrajšav izračunavanja moremo hitro izpeljati, ako poznamo stavke, ki smo jih navedli že pri aritmetičnih sredinah v odstavku 81,21

$$\bar{a} = a; \quad \overline{x + y} = \bar{x} + \bar{y} \text{ in } \overline{bx} = b\bar{x}$$

ker je varianca povprečen kvadratičen odklon od aritmetične sredine, moremo pisati:

$$\sigma_x^2 = \overline{(x - \bar{x})^2} = \overline{(x^2 - 2x\bar{x} + \bar{x}^2)} = \overline{x^2} - 2\bar{x}\bar{x} + \bar{x}^2 = \overline{x^2} - \bar{x}^2 \qquad (11.18)$$

S tem je dokazan prvi način

Ostane še dokaz načina, v katerem uporabljamo pogojno sredino.

Ker je $x_i - x_o = u_i$, sledi:

$$x_i = x_o + u_i$$

$$\bar{x} = x_o + \bar{u}$$

$$x - \bar{x} = u - \bar{u}$$

$$(x - \bar{x})^2 = (u - \bar{u})^2$$

$$\sigma_x^2 = \overline{(x - \bar{x})^2} = \overline{(u - \bar{u})^2} = \overline{u^2} - \bar{u}^2$$

$$\sigma_x^2 = \overline{u^2} - \bar{u}^2 \qquad u_i = x_i - x_o \qquad (11.19)$$

115.4 Izračunavanje variance in standardne deviacije iz podatkov, ki so grupirani v frekvenčni distribuciji

.41 Izračunavanje

Ako imamo vrednosti urejene v frekvenčni distribuciji, aritmetično sredino izračunavamo kot tehtano aritmetično sredino. Pri tem grupne vrednosti množimo s pripadajočimi frekvencami, ki jih vzamemo kot teže.

Ker je varianca aritmetična sredina kvadratov odklonov od aritmetične sredine, je jasno, da jo bomo morali v tem primeru izračunati tudi kot tehtano aritmetično sredino kvadratov odklonov.

torej:
$$\sigma_x^2 = \overline{(x - \bar{x})^2} = \frac{1}{n}\sum n_k (x_k - \bar{x})^2 \qquad (11.20)$$

Enako sta tudi v obrazcu za izračunavanje variance po obrazcu:

$\sigma_x^2 = \overline{x^2} - \bar{x}^2$ povprečen kvadrat in aritmetična sredina izračunana tehtani kot/ povprečji, ako gre za frekvenčno distribucijo.

$$\overline{x^2} = \frac{1}{n}\sum_k n_k x_k^2 \qquad \bar{x} = \frac{1}{n}\sum_k n_k x_k \qquad (11.21)$$

.42 Izračunavanje s pogojno sredino

Posebno ugodno izvemo izračunavanje variance in frekvenčne distribucije izvesti, ako je širina razredov pri vseh razredih enaka. V tem primeru iz razloga, ki smo ga navedli že pri tehtani aritmetični sredini, uvedemo tudi za izračunavanje variance pomožni znak

$$u = \frac{x - x_o}{\Delta}$$

Izkaže se, da so vrednosti u posebno enostavne, -3, -2, -1, -0, +1, +2, +3, ako je x_o grupna vrednost sredi serije, Δ pa širina razreda. V varianco znaka je v tem primeru možno izračunati razmeroma enostavno. Varianca znaka x je z varianco znaka u v naslednji zvezi:

$$\sigma_x^2 = \Delta^2 \sigma_u^2$$

Zaradi tega je možno varianco znaka x iz variance znaka u zelo enostavno izračunati.

D Dokaz moremo izpeljati v naslednji zvezi:

Ker je $u = \frac{x - x_o}{\Delta}$ je $x = x_o + \Delta u$

$$\bar{x} = x_o + \Delta \bar{u}$$

$$x - \bar{x} = \Delta$$

$$(x - \bar{x})^2 = \Delta (u - \bar{u})^2$$

$$(x - \bar{x})^2 = \Delta^2 (u - \bar{u})^2$$

$$\sigma_x^2 = \Delta^2 (u - \bar{u})^2 = \Delta^2 \sigma_u^2$$

Shema in postopek izračunavanja sta v tem primeru naslednja

k	x_k	n_k	$\frac{x-x_o}{\Delta} = u_k$	$n_k u_k$	$n_k u_k^2$
	(1)	(2)	(3)	(2).(3)=(4)	(3).(4)=(5)
1	x_1	n_1	u_1	$n_1 u_1$	$n_1 u_1^2$
2	x_2	n_2	u_2	$n_2 u_2$	$n_2 u_2^2$
			-2	.	.
			-1	$n_k \cdot 1$	$n_k \cdot 1^2$
			0	0	0
			+1	.	.
			+2	.	.
r	x_r	n_r	u_r	$n_r u_r$	$n_r u_r^2$
$\sum$		n		$n_k u_k$	$n_k u_k^2$
$\frac{1}{n} \sum$		1		$\bar{u}$	$\overline{u^2}$

(11.22)

$$\sigma_x^2 = \Delta^2 (\overline{u^2} - \bar{u}^2) \qquad \sigma_x = \Delta \sqrt{\overline{u^2} - \bar{u}^2}$$

Stopnje izračunavanja so naslednje:

a) Izpišemo v stolpec 1 grupne vrednosti v stolpec 2 pa odgovarjajoče frekvence.
b) Izberemo primerno grupno vrednost kot pogojno sredino in ugotovimo širino razreda.
c) V stolpcu 3 v vrsto, ki pripada pogojni sredini, postavimo vrednost 0, v vrste navzgor od 0 zaporedoma- 1, -2, -3, -4 od 0 navzdol pa vrednosti +1, +2, +3, +4 itd. To so vrednosti u.
d) pomnožimo odgovarjajoče vrednosti iz stolpca 2 in 3 in produkte vpišemo v stolpec 4.
e) Pomnožimo odgovarjajoče vrednosti iz stolpca 3 in 4 in produkte vpišemo v stolpec 5.
f) Seštejemo stolpec 2, stolpec 4 in stolpec 5.
g) Dobljene vsote delimo z vsoto vseh frekvenc (vsota stolpca 2). Na ta način dobimo količini $\bar{u}^2$ in $\bar{u}$.
h) Varianco izračunamo po obrazcu $\sigma_x^2 = \Delta^2\left(\overline{u^2} - \bar{u}^2\right)$

standardno deviacijo pa po obrazcu $\sigma_x = \Delta\sqrt{\overline{u^2} - \bar{u}^2}$

P Izračunati je treba standardno deviacijo teže po 1000 zrn iz 373 parcel z uporabo pogojne sredine.

Tab. 11.10

Teža v gramih	Grupna sredina	Frekvenca	$\frac{x_k - 38}{4}$	$n_k u_k$	$n_k u_k^2$
	x_k	n_k	u_k		
	(1)	(2)	(3)	(2).(3)=(4)	(3).(4)=(5)
nad 20 - 24	22	1	-4	-4	16
nad 24 - 28	26	12	-3	-36	+108
nad 28 - 32	30	28	-2	-56	+112
nad 32 - 36	34	96	-1	-96	+ 96
nad 36 - 40	38	135	0	0	0
nad 40 - 44	42	65	+1	+65	+ 65
nad 44 - 48	46	23	+2	+46	+ 46
nad 48 - 52	50	7	+3	+21	+ 21
nad 52 - 56	54	4	+4	+16	+ 16
nad 56 - 60	58	2	+5	+10	+ 10
		373		-34	666
$\frac{1}{373}$		1		-0,0912	1,785 = $\overline{u^2}$

$\sigma_x^2 = \Delta^2\left[\overline{u^2} - \bar{u}^2\right] = 4^2\left[1{,}785 - (0{,}0912)^2\right] = 16\left[1{,}785 - 0{,}008\right] = 16 \cdot 1{,}777 = 28{,}432$

$\sigma_x = \sqrt{28{,}432} = 5{,}33$ gr

Kot grupne vrednosti smo vzeli sredine razredov, kot pogojno sredino grupno vrednost 38 gr, širina razreda pa je $\Delta = 4$.

.43 Izračunavanje variance s pomočjo kumulativnih serij

Že pri izračunavanju tehtane aritmetične sredine smo videli, da moremo s kumulativnimi serijami zelo enostavno izračunati tehtano aritmetično sredino. Metodo kumulativnih serij pa moremo uporabiti tudi za izračunavanje variance.

Shema in postopek izračunavanja sta naslednja:

k	X_k	n_k	N_k	N'_k
	(1)	(2)	(3)	(4)
0	-	-	$N_o = N_1 + n_1$	$N_o = N'_1 + N_1$
1	x_1	n_1	$N_1 = N_2 + n_2$	$N'_1 = N'_2 + N_2$
2	x_2	n_2	$N_2 = N_3 + n_3$	$N'_2 = N'_3 + N_3$
3	x_3	n_3	$N_3 = N_4 + n_4$	$N'_3 = N'_4 + N_4$
r - 2	x_{r-2}	n_{r-2}	$N_{r-2} = N_{r-1} + n_{r-1}$	$N'_{r-2} = N'_{r-1} + N_{r-1}$
r - 1	x_{r-1}	n_{r-1}	$N_{r-1} = N_r + n_r$	$N'_{r-1} = N'_r + N_r$
r	x_r	n_r	$N_r = 0$	$N'_r = 0$

(11.23)

$$\sum_{k=1}^{r} N'_k = N''_o$$

$$C_1 = \frac{N'_o}{N_o} \qquad C_2 = \frac{N''_o}{N_o} \qquad \sigma_x^2 = \Delta^2 \left[2 C_2 - C_1 (C_1 - 1) \right]$$

Stopnje izračunavanja varianco so naslednje:

1. Najprej napišemo grupne vrednosti in frekvence distribucije, za katero hočemo izračunati varianco. [Stolpec (1) in (2)]
2. Iz frekvenc (stolpec 2) izračunamo od spodaj nevzgor prvo kumulativno serijo, na isti način kot pri izračunavanju sredine, oz. kot je prikazano na primeru. Prvo kumulativno serijo pišemo

v stolpec 3. Zadnji člen te serije je število vseh enot n in ga zaznamujemo z N_o.

3, Iz prve kumulativne serije (stolpec 3) izračunamo drugo kumulativno serijo na enak način, kot smo izračunali prvo kumulativo iz frekvenc in jo vpišemo v stolpec 4. Zadnji člen druge kumulative zaznamujemo z No'.

4. Seštejemo vrednosti druge kumulativne serije stolpec (4) brez vrednosti N_o! To vsoto zaznamujemo z No''.

5. Iz vrednosti N_o, N_o' in N_o'' izračunamo količine:

$$c_1 = \frac{N_o'}{N_o} \qquad c_2 = \frac{N''}{N_o} \qquad (11.24)$$

6. Iz količine c_1, c_2 in Δ izračunamo varianco po obrazcu:

$$\sigma_x^2 = \Delta^2 \left[2c_2 - c_1 (c_1 - 1) \right]. \qquad (11.25)$$

D Dokaz ni zamotan, vendar presega okvir programa. Upravičenost postopka bomo zaradi tega preizkusili na primeru, za katerega imamo izračunano varianco že po kateri izmed prejšnjih metod.

Vzemimo kar frekvenčno distribucijo teže po 1000 zrn pšenice, za katero smo varianco izračunali v prejšnjem odstavku.

Tab. 11.

K	x_k	n_k	N_k	N'_k	
0			No = 373	No' = 1458	
1	22	1	372	1086	
2	26	12	360	726	
3	30	28	332	394	
4	34	96	236	158	
5	38	135	101	57	2452 = N''
6	42	65	36	21	
7	46	23	13	8	
8	50	7	6	2	
9	54	4	2	0	
10	58	2	0	0	

$$c_1 = \frac{1458}{373} = 3,9088 \qquad c_2 = \frac{2452}{373} = 6,57373$$

$$\sigma^2 = \Delta^2 \left[2\, c_2 - c_1 \ (c_1 - 1) \right]$$

$$\sigma^2 = 4^2 \left[2.6,57373 - 3,9088 \cdot 2,9088 \right]$$

$$\sigma^2 = 16 \cdot 1,777 = 28,432$$

Ta postopek je zelo prikladon, ker pridemo do količin, ki jih potrebujemo za izračunavanje variance, s pomočjo samega seštevanja. Posebno pa se postopek obnese, ako imamo na razpolago računski stroj z registrirnim trakom. V tem primeru kumulativne serijo izračunamo z uporabo takoimenovanih "subtotalov".

11.5 Interval $\bar{x} \pm t\sigma$

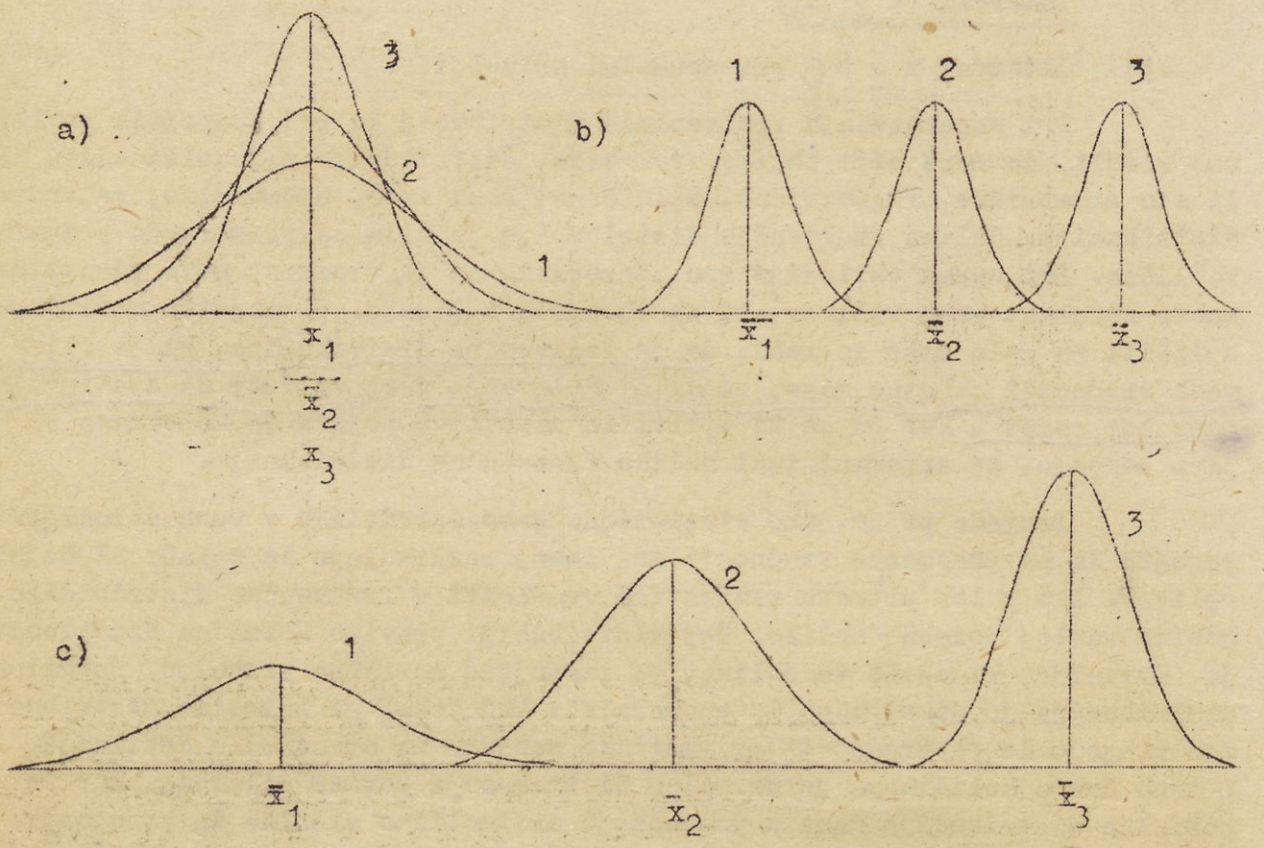
.51 Interval $\bar{x} \pm t\sigma$ pri normalni krivulji.

Pri obravnavanju frekvenčnih distribucij in pri kasnejših poglavjih smo videli, da more biti oblika frekvenčne distribucije najrazličnejša. Videli smo simetrične, asimetrične, sploščene, koničaste, enomodalne, večmodalne distribucije. Oblika frekvenčne distribucije je bila od primera do primera različna. Pri merah variacije smo opozorili, da je vrednost neke posamezne enote rezultat vseh vplivov, ki na to enoto vplivajo. To dejstvo moremo posplošiti na celo maso in reči, da je frekvenčna distribucije, ki je slika vseh vrednosti celotne mase, rezultat delovanja vseh vplivov na statistično maso kot celot. Ker pa se ti vplivi od masovnega pojava do masovnega pojava menjajo, se spremeni tudi oblika frekvenčne distribucije.

Drugače pa je, ako statistično maso opredelimo v vseh določljivih pogojih in se posamezne vrednosti med seboj razlikujejo le zaradi slučajnih vplivov. Ako v tem primeru razvrstimo vrednosti v frekvenčno distribucijo, dobimo vselej podobno obliko. Tej distribuciji pravimo normalna distribucija. Krivuljo, ki opiše to obliko, pa imenujemo normalno krivuljo. Normalna distribucija je simetrična in enomodalna. Med seboj se normalne distribucije razlikujejo le po vrednosti aritmetične sredine in standardne deviacije. S tema dvema količinama je normalna distribucija povsem določena. Ker je normalna distribucija dana z vrednostjo aritmetične sredine in standardne deviacije, se posamezne normalne krivulje med seboj razlikujejo po tem, da leže na različnih delih skale (odvisnost od aritmetične sredine) in da so bolj ali manj raztegnjene (odvisnost od standardne deviacije). To moremo lepo razvideti iz slike 11.6. Na vsaki sliki so narisane po tri normalne krivulje. Na sliki a) je pri vseh treh normalnih krivuljah aritmetična sredina ista, različna pa je standardna deviacija, na sliki b) so pri vseh treh enake standardne deviacije, različne pa so aritmetične sredine, na sliki c) pa so različne tako standardne deviacije kot tudi aritmetične sredine.

Sl. 11.6

Različne normalne krivulje



Iz slike a) vidimo, da vpliva standardna deviacija na samo obliko normalne krivulje, ki je tem bolj raztegnjena, čim večja je standardna deviacija. Iz slike b) pa vidimo, da se zaradi spremenjene vrednosti aritmetične sredine pri isti standardni deviaciji oblika krivulje ne izpremeni, ampak samo prestavi na drug del.

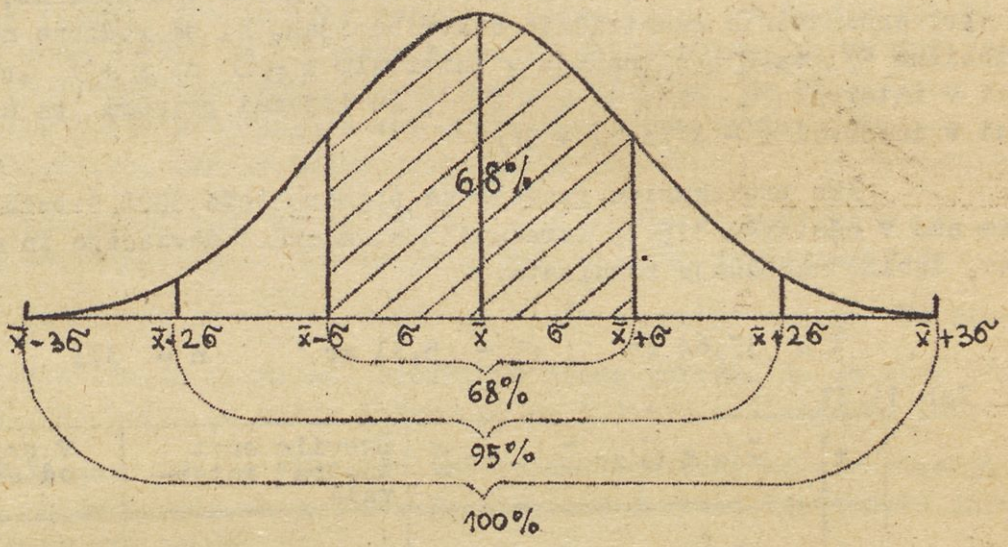
Vkljub medsebojni različnosti normalnih krivulj, ki je razvidna iz slike c), pa moremo ugotoviti naslednje važno dejstvo, ki velja za vse normalne krivulje, ne glede na to, za katero konkretno normalno krivuljo gre.

Okrog aritmetične sredine $\bar{x}$ odredimo interval na ta način, da enkrat od aritmetične sredine standardno doviacijo σ_x odštejemo, drugič prištejemo. Dobimo interval od $\bar{x} - \sigma_x$ do $\bar{x} + \sigma_x$. V tem intervalu se nahaja pri vseh normalnih krivuljah cca. 68 °/c (točno 68,27 °/o) enot celotne statistične mase.

Ako tvorimo interval $\bar{x} - 2\sigma$ do $\bar{x} + 2\sigma$ je jasno, da se v tem intervalu nahaja večji del in sicer cca. 95 °/o (točno 95,45 °/o) vseh vrednosti. Ako pa tvorimo interval $\bar{x} - 3\sigma$ do $\bar{x} + 3\sigma$ pa so zunaj intervala skoro vsi primeri (točno 99,73 °/o).

Grafično je gornje dejstvo ponazorjeno v sliki 11.7

Sl. 11.7



V splošnem moremo najti za vsako normalno krivuljo, koliki del vseh enot leži v intervalu od $\bar{x} - t\sigma$ do $\bar{x} + t\sigma$. Procent, koliko enot leži v tem intervalu, je odvisen samo od količine t. Ta t pove, koliko standardnih deviacij smo aritmetični sredini odšteli ali prišteli. V našem zgornjem primeru je pri intervalu $\bar{x} - \sigma$ do $\bar{x} + \sigma$ t enak 1, pri intervalu $\bar{x} - 3\sigma$ do $\bar{x} + 3\sigma$ pa je t enak 3. V splošnem moremo za vsako vrednost t-ja določiti, kakšen del vseh enot leži pri normalni krivulji v intervalu $\bar{x} - t\sigma$ do $\bar{x} + t\sigma$. Izdelane so tabele, ki podajajo te odnose. Iz njih smo vzeli samo najvažnejše vrednosti, ki so prikazane v tabeli :

Tab. 11.12

t	P
0,6745	50 °/o
1.00	68,27 °/o
1,96	95 °/o
2,00	95,45 °/o
2,5758	99 °/o
3,00	99,73 °/o

Iz tabele vidimo, da moramo vzeti interval $\bar{x} - 1{,}96\sigma$ do $\bar{x} + 1{,}96\sigma$ ako hočemo, da bo znotraj tega intervala točno 95 °/o vseh enot. Kot bomo videli, se ta interval zelo pogosto uporablja.

.52 Interval $\bar{x} \pm t\,\sigma$ pri splošnih distribucijah

V odstavku 115.51 navedena zakonitost velja strogo samo za normalne krivulje. Izkaže pa se, da moremo za distribucije, ki se od normalne ne razlikujejo preveč (enomodalne, simetrične) reči, da zanje zgornje zakonitosti veljajo približno in sicer tem bolj, čim bolj je distribucija podobna normalni krivulji. Ker je v praksi zelo veliko distribucij, ki so normalni distribuciji zelo slične, moremo to ugotovitev zelo koristno uporabiti. "Pri tem pri procentih seveda ne gremo v decimalke, ampak rečemo: pri enomodalnih simetričnih distribucijah, ki so podobne normalni, se približno 2/3 vseh enot nahaja v intervalu $\bar{x} - \sigma$ do $\bar{x} + \sigma$,približno 95°/o enot v intervalu $\bar{x} - 2\sigma$ do $\bar{x} + 2\sigma$, le izjemni primeri pa niso vključeni v interval $\bar{x} - 3\sigma$ do $\bar{x} + 3\sigma$."

P Ako preizkusimo pravilo na primeru teže 1000 žitnih zrn, za katere smo v odstavku 115.42 izračunali standardno deviacijo in aritmetično sredino, dobimo naslednje rezultate.

$\bar{X} = 37{,}64$ gr $\quad \sigma = 5{,}33$ gr $\quad n = 373$

Tab. 11.23

t	$\bar{x} - t\sigma$ do $\bar{x} - +\sigma$	Število enot znotraj intervala	V procentih od skupnega
1	32,31 - 42,97	272	73 °/o
2	26,98 - 48,30	351	94 °/o
3	21,65 - 53,63	369	99 °/o

Iz navedenega primera vidimo, da se v našem primeru vrednosti odklanjajo od pričakovanih vrednosti v prvem intervalu za 5 °/o (73 - 68 = 5 °/o) v drugem za 1 °/o (94 - 95 = - 1 °/o)in enako za 1 °/o v tretjem (99 - 100 = - 1 °/o). Rezultati so torej povsem zadovoljivi.

116. RELATIVNE MERE VARIACIJE. KOEFICIENT VARIACIJE

116.0 Pojm

Vse mere variacije, ki smo jih do sedaj obravnavali, so absolutne količine. Vendar bo že kratek primer pokazal, da absolutne mere variacije ne zadoščajo, posebno ako hočemo primerjati med seboj mere variacije različnih mas.

Vzemimo, da smo v vseh krajevnih ljudskih odborih registrirali cene. Za vsak artikel smo izračunali povprečno ceno za vso LRS in standardno deviacijo cen. Ta pokaže variabilnost cen v odvisnosti od kraja. Vzemimo dva artikla: čebnje in blago za moške obleke. Primerjava izgleda na prvi pogled nesmiselna, vendar bomo takoj videli, da je primerjava

variabilnosti smiselma, če le vzamemo ... vilno merilo. Vzemimo, da smo ugotovili za češnje $\bar{x} = 35$ din $\sigma_x = 15$ din, za blago za moške obleke določene kvalitete pa $\bar{x} = 3800$ din $\sigma_x = 100$ din. Ako skušamo primerjati med seboj standardno deviacijo češenj in standardno deviacijo blaga za moške obleke, vidimo, da je standardna deviacija pri blagu zelo velika napram standardni deviaciji pri češnjah, ne-glede na to, da ta primerjava nima posebnega smisla, kot nima smisla primerjava povprečnih cen za oba artikla. Sklep, da cene blaga bolj variirajo kot cene češenj pa bi se pokazal kot sporen tudi iz drugega vidika. Ako je cena češenj v nekem kraju za 20 din različna od povprečne cene 35 din za kg, cena blaga za moške obleke pa za 200 din različna od povprečne cene 3800 din za meter, pravimo, da je razlika v ceni od povprečne večja pri češnjah kot pri blagu. V tem primeru gledamo nehote relativno razliko, kar je edino pravilno in smiselno.

Relativno razliko pa dobimo, ako absolutno razliko v ceni primerjamo s povprečno ceno. Ako to relativno razliko izračunamo, je $\frac{20}{35} = 0{,}57$ za češnje in $\frac{200}{3800} = 0{,}053$ za blago za moške obleke. Ta primerjava pa je smisolna in pravilna. Relativen odnos pokaže, da je cena češenj za 57 °/o, cena blaga za moške obleke pa za 5,3 °/o različna od povprečne.

Iz gornjega se vidi, da smo upravičeni tudi mero variacije primerjati s srednjo vrednostjo, ker je standardna deviacija skupen izraz odklonov cen za vse kraje.

Na ta način pridemo do pojma relativne mere variacije. Relativne mere variacije izračunamo na ta način, da določeno mero variacije, ki je izraz individualnih vplivov, postavimo v odnos do pripadajoče srednje vrednosti, ki je izraz splošnih vplivov.

116.1 Ako semi - quartilno vrednost $\frac{Q_3 - Q_1}{2}$ delimo z mediano (Me) ali povprečjem med Q_1 in Q_3 $\left(\frac{Q_1 + Q_3}{2}\right)$, dobimo relativni merili variacije, ki izvirata iz quartilov.

Prvo se izračunava kot $\frac{Q_3 - Q_1}{2\ Me}$; (11.26)

drugo pa je enako $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ (11.27)

V našem primeru teže po 1000 zrn iz 373 parcel, kjer je $Q_1 = 34{,}20$ gr, $Q_3 = 40{,}49$ gr, $Me = 37{,}48$ je

$$\frac{Q_3 - Q_1}{2\ Me} = \frac{40{,}49 - 34{,}20}{2 \cdot 37{,}48} = \frac{6{,}29}{74{,}96} = 0{,}084 \text{ ali v odstotkih } 8{,}4\ °/o$$

Drugo merilo:

$$\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40,49 - 34,20}{40,49 + 34,20} = \frac{6.29}{74,96} \quad 0,084 \text{ ali v odstotkih } 8,4\ \%.$$

Razlika med obema merama je minimalna in se v našem primeru pri rezultatu sploh ne občuti. Druga mera je v toliko enostavnejša, ker se da izračunati že, ako poznamo samo prvi in tretji del oz. quartil. Ker rezultat ni dosti različen od prvega, se običajno izračunava po drugem načinu.

116.2 Povprečen absoluten doklon

primerjamo z mediano Me, ako smo odklone izračunavali od mediane. Ako pa smo odklone računali od aritmetične sredine, pa ga primerjamo z aritmetično sredino. Relativna mera variacije iz povprečnega absolutnega odklona se torej računa ali po obrazcu

$$\frac{\delta_{Me}}{Me} \quad \text{ali} \quad \frac{\delta_{\bar{x}}}{\bar{x}} \qquad (11.28)$$

P V našem primeru plač devetih delavcev za katerega smo dobili v odstavku 114. da je $\delta_{Me} = 76$ din, Me = 684, je $\frac{\delta_{Me}}{Me} = \frac{76}{684} = 0,111$ ali v odstotkih 11.1 %

116.3 Koeficient variacije

Najvažnejši izmed vseh relativnih mer variacije pa je koeficient variacije. Koeficient variacije izračunamo na ta način, da delimo standardno deviacijo σ_x z aritmetično sredino $\bar{x}$. Ako zaznamujemo koeficient variacije z KV se obrazec za izračunavanje koeficienta variacije glasi:

$$KV = \frac{\sigma_x}{\bar{x}} \quad \text{ali v odstotkih} \quad KV = \frac{\sigma_x}{\bar{x}} \cdot 100 \qquad (11.29)$$

Iz obrazca je razvidno, da je koeficient variacije neimenovano število, ker ima σ_x in $\bar{x}$ isto enoto mere.

Koeficient variacije je zelo dobro sredstvo primerjave variabilnosti različnih statističnih mas. Ker je KV neimenovano število je možna celo primerjava med variabilnostjo raznovrstnih mas.

Za primera, za katera smo pri obravnavanju variacije izračunavali standardno deviacijo, je koeficient variacije naslednji:

P V primeru odstavka 115.32 smo izračunali povprečno mesečno vrednost proizvodnje $\bar{x} = 1,287.650$ din in standardno deviacijo $\sigma_x = 2945$.

Koeficient variacije KV je enak :

$$KV = \frac{\sigma_x}{\bar{x}} 100 = \frac{2945}{1287650} 100 = 0{,}227\ \%$$

Iz koeficienta variacije vidimo, da je variabilnost mesečne proizvodnje v tem primeru zelo majhna. Koeficient variacije za proizvodnjo tega podjetja bi mogli brez nadaljnjega primerjati s koeficientom variacije proizvodnje za druga podjetja. Ne samo to. Smiselna bi bila tudi primerjava koeficienta variacije za vrednost proizvodnje s koeficientom variacije števila delavstva, izvršenih delovnih ur itd., medtem ko medsebojna primerjava odgovarjajočih standardnih deviacij ne bi imela nikakega smisla.

P Iz primera v odstavku 115.42 dobimo, da je povprečna teža 1000 zrn 37,64 gramov, standardna deviacija pa 5,33 gramov. Ti količini sta bili iz distribucije izračunani v poglavju o aritmetični sredini in standardni deviaciji.

Koeficient variacije je naslednji:

$$KV = \frac{\sigma_x}{\bar{x}} \times 100 = \frac{5{,}33}{37{,}64} 100 = 14{,}2\ \%$$

P Zaradi važnosti in uporabnosti koeficientov variacije za analizo poglejmo še primer iz kmetijske statistike:

Po podatkih PO 8 imamo izračunane povprečne pridelke v kg na rodno drevo jablan, hrušk, sliv in povprečen pridelek vina v hl na ha rodnih vinogradov. Dane imamo tudi standardne deviacije pridelkov po okrajih. Izračunati je treba koeficiente variacije in analizirati variabilnost pridelka sadja.

Tab. 11.14

Vrsta pridelka	Enota mere	$\bar{x}$	σ_x	KV
Jabolka	kg/dr	14,0	12,0	85,7 %
Hruške	"	16,1	9,6	59,6 "
Slive	"	5,1	4,9	96,1 "
Vino	hl/ha	25,0	12,7	50,8 "

$KV_{jab} = \frac{12}{14} \cdot 100 = 85{,}7\ \%$

Izmed vseh vrst pridelkov je pridelek vina še najbolj neodvisen od teritorija (KV je 50,8 %), najbolj pa krajevni faktorji vplivajo na pridelek sliv (KV = 96,1 %).

117. POVZETEK

Mere variacije merijo variabilnost pojava. Z njimi izražamo jakost individualnih vplivov, na enak način kot podajajo srednje vrednosti rezultat splošnih vplivov. Mer variacije je več vrst:

1. Širina variacije je diferenca med največjo in najmanjšo vrednostjo, ki v maši nastopa.

2. Interquartilna vrednost $(Q_3 - Q_1)$ podaja interval, v katerem se nahaja 50 %/o vrednosti, 25 °/o vrednosti je manjših od Q_1, 25 °/o pa večjih od Q_3.

Poleg interquartilne vrednosti izračunavamo tudi semi-interquartilno vrednost, ki je $\frac{1}{2}(Q_3 - Q_1)$.

3. Povprečen absoluten odklon $\delta_M = \frac{1}{n}\sum[x_1 - M]$ je mera variacije ki je odvisna od vrednosti vseh členov. Kot vrednost, od katere izračunavamo odklone, vzamemo kako srednjo vrednost. Običajno vzamemo ali aritmetično sredino $\delta_{\bar{x}} = \frac{1}{n}\sum[x_1 - \bar{x}]$, najpogosteje pa mediano $\delta_{Mo} = \frac{1}{n}\sum[x_I - Me]$, to pa zaradi tega, ker je v tem primeru vrednost povprečnega absolutnega odklona najmanjša.

4. Varianca in standardna deviacija sta meri variacije, ki se v statistiki najpogosteje uporabljata.

Varianca je povprečen kvadratičen odklon od aritmetične sredine $V = \sigma^2 = \frac{1}{n}\sum(x_I - \bar{x})^2$ oziroma $\sigma^2 = \frac{1}{n}\sum n_k (x_k - \bar{x})^2$, ako izračunavamo varianco iz frekvenčne distribucije. Krajše moremo pisati $\sigma^2 = \overline{(x-\bar{x})^2}$, standardna deviacija, pa je kvadratni koren iz variance $\sigma = \sqrt{V} =$

$$\sigma = \frac{1}{n}\sqrt{(x_i - \bar{x})^2}$$

Za izračunavanje variance imamo postopke, ki skrajšajo sicer zamudno izračunavanje. Z vpeljavo pogojne sredine x_c in novega znaka $u = x - x_o$ izračunamo varianco po obrazcu

$\sigma_x^2 = \overline{u^2} - \bar{u}^2$, kjer je $\overline{u^2} = \frac{1}{n}\sum u_i^2$, $\bar{u} = \frac{1}{n}\sum u_i$.

V primeru, da so vrednosti urejene v frekvenčni distribuciji, vpeljemo nov znak

$$u = \frac{x - x_o}{\Delta}.$$

x_o pogojna sredina, Δ širina razreda).

Varianca σ^2 se v tem primeru izračuna po obrazcu:

$\sigma_x^2 = \Delta^2[\overline{u^2} - \bar{u}^2]$. $\overline{U^2}$ in $\bar{u}$ sta v tem primeru izračunana kot tehtani sredini.

S kumulativnimi serijami izračunamo varianco po obrazcu

$\sigma_x^2 = \Delta^2[2C_2 - C_1(C_1 - 1)]$, kjer je $C_1 = \frac{No'}{No}$, $C_2 = \frac{No''}{No}$, No, No', No'' pa zadnje vrednosti prve, druge in tretje kumulative iz frekvenc.

.5 Ako delujejo na enote statistične mase samo slučajni vplivi, se vrednosti razdele v normalni distribuciji. V tem primeru se v intervalu $\bar{x} - t\sigma$ do $\bar{x} + t\sigma$ za vse normalne distribucije nahaja pri istem t-ju enak del vsega števila vrednosti. Ako je $t = 1$, se nahaja približno 68 %, ako je $t = 2$, se nahaja približno 95 %, ako je $t = 3$ pa samo izjemne vrednosti izven tega intervala. Ti odnosi veljajo približno tudi za frekvenčne distribucije, ki niso normalne, temveč so normalnim samo podobne.

.6 Relativne mere variacije so količine, ki merijo variabilnost pojava v odnosu s srednjimi vrednostmi. Med njimi so važne:

$\frac{Q_3 - Q_1}{2\,Me}$, $\frac{Q_3 - Q_1}{Q_3 + Q_1}$, $\frac{\delta_x}{\bar{x}}$, $\frac{\delta_{Me}}{Me}$ in pa koeficient variacije

$KV = \frac{\sigma_x}{\bar{x}}$ ali v odstotkih $KV = \frac{\sigma_x}{\bar{x}} \cdot 100$, ki je med vsemi najvažnejši in se najpogosteje uporablja. Koeficient variacije je koeficient, ki zelo razširi možnost primerjave variabilnosti posameznih pojavov.

12. OSNOVE KORELACIJE

120. SPLOŠNO

Dosedaj smo proučevali naenkrat vedno vsak znak statistične mase zase. Pri tem nas ni zanimalo, kakšne so vrednosti drugih znakov. Izkaže pa se, da tako proučevanje ne zajame ene izmed najvažnejših in najznačilnejših lastnosti masovnih pojavov, to je njih medsebojne povezanost in odvisnost. Na povezanost in odvisnost pojavov pa naletimo pri praktičnem in teoretičnem proučevanju na vsakem koraku.

Kmetijska proizvodnja je odvisna od velikosti obrata, delovne sile, površine, števila živine itd, industrijska proizvodnja podjetja je odvisna od števila delavstva, stopnje elektrifikacije, vrste proizvodnje itd. Nepismenost je odvisna od kulturno prosvetne dejavnosti, ki se izraža s številom šol ali učiteljev na 1000 prebivalcev. Število obolenj in smrti za nalezljivimi boleznimi je odvisno od zdravstvene službe. V naštetih primerih moremo ugotoviti kateri pojav je odvisen od drugega, govorimo o vzročni odvisnosti. Kmetijska proizvodnja je odvisna od velikosti obrata in ne narobe itd.

Pri nekaterih pojavih pa moremo opaziti odvisnosti in povezanosti, ki niso vzročne. Ako n.pr. proučujemo po okrajih nepismenost moških in žensk, sta ta dva pojava med seboj povezana. Čim večja je nepismenost moških, tem večja je tudi nepismenost žensk in narobe. Vendar ne moremo trditi, da je ta povezanost vzročna, ker ni nepismenost žensk posledica nepismenosti moških, niti narobe. Povezava nastopi v tem primeru zaradi tega, ker sta nepismenost moških in žensk v istem okraju odvisna od istih skupnih faktorjev, ki vplivajo tako na nepismenost pri ženskah, kot na nepismenost pri moških. Ti faktorji so splošni kulturni nivo itd.

Jasno je na drugi strani, da vsi pojavi med seboj niso povezani in odvisni, čeprav bi bila povezanost med njimi možna in smiselna. Dosti pa je tudi pojavov, med katerimi bi bilo nesmiselno pričakovati povezavo. Gotovo je nesmisel iskati povezavo med donosom pšenice in številom rojstev v okraju, med številom porok in površino pod žiti in podobno.

Statistični postopki proučevanja zvez in odvisnosti so zelo razviti in uspešni. Vendar bomo med njimi obdelali le najbolj enostavne.

121. FUNKCIONALNE ODVISNOSTI

Odvisnost in povezanost obravnavamo tudi v matematiki. V matematiki je odvisnost med dvema spremenljivkama x in y dana z nekim pravilom zveze med x in y. Po tem pravilu vsaki vrednosti x odgovarja ena ali več točno določenih vrednosti spremenljivke y. V tem primeru pravimo, da je x neodvisna spremeljivka, y pa odvisna spremeljivka od x. S simbolom pišemo $y = f(x)$ in pravimo: y je funkcija od x. Zveze take narave, kot so omenjene, imenujemo funkcionalne zveze. Ako velja med x in y funkcionalna zveza, pripada vsakemu x ena ali nekaj točno določenih vrednosti. Vrednost spremeljivke y je odvisna

samo od vrednosti x, je torej z vrednostjo x določena.

Zveza med x in y je običajno dana z enačbo funkcije, iz katere moramo za vsak x izračunati odgovarjajoči y. Ako vzamemo kot primer funkcijo $y = 2x + 3x^2$. Odgovarja za to funkcijsko zvezo n.pr.

$x = 2$ vrednost $y = 2 \cdot 2 + 3 \cdot 2^2 = 16$.

Poleg načina navajanja funkcijske zveze v obliki enačbe, pa moramo podati funkcijsko zvezo tudi v obliki sistema parov vrednosti, kjer navedemo k vsaki vrednosti neodvisne spremenljivke x vrednost odvisne spremenljivke y. Za našo funkcijo $y = 2x + 3x^2$ moramo na tà način pisati nekaj dvojic vrednosti v obliki kot kaže tabela 12.1

Tab. 12.1

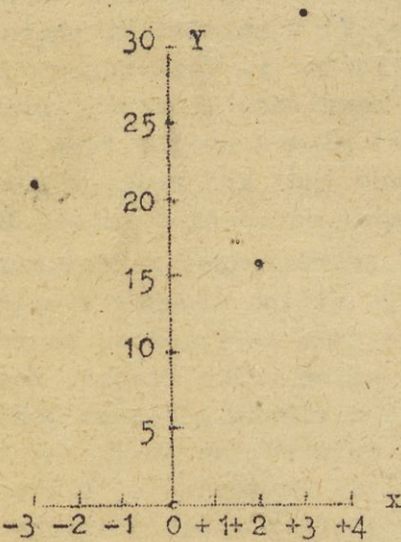
x	y
- 3	+ 21
- 2	+ 8
- 1	+ 1
0	0
+ 1	+ 5
+ 2	+ 16
+ 3	+ 33
+ 4	+ 56

Navedli smo samo k osmim vrednostim x odgovarjajoče vrednosti y. Na prvi pogled je vidno, da je v taki obliki nemogoče izčrpati in napisati dvojice za vse vrednosti x-a. To more dati edino pravilo, pisano v obliki enačbe

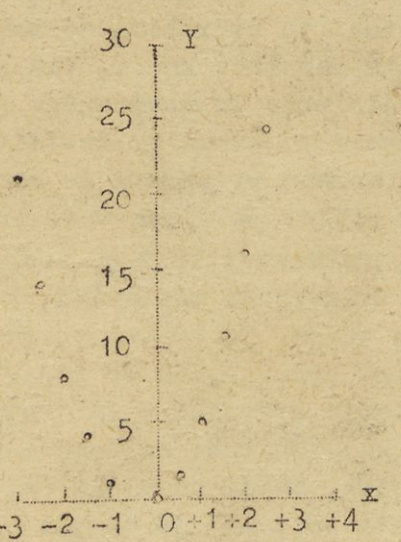
$$y = 2x + 3x^2 .$$

Dvojico vrednosti moramo grafično prikazati v pravokotnem koordinantnem sistemu kot točko. Ta način prikazovanja poznamo iz analitične geometrije. Vrednost x vzamemo kot abscisо (vodoravna os), vrednost y pa kot ordinato (navpična os). Ako narišemo samo nekaj točk funkcije, ne bo še vidna slika funkcije (sl.12.1). Več vrednosti že bolje poda potek funkcije (sl.12.2), medtem ko da zvezna krivulja/sliko funkcije (sl.12.3)

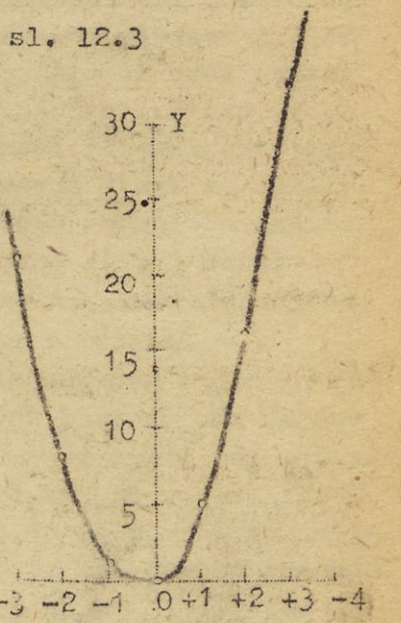
sl. 12.1



sl. 12.2



sl. 12.3



Fukkcionalne odvisnosti moramo torej prikazati na tri načine:

1./ V obliki enačbe
2./ V obliki sistema dvojic vrednosti x in y
3./ Grafično v pravokotnem koordinatnem sistemu kot sistem točk ali krivuljo.

122. KORELACIJSKE ODVISNOSTI

Pri proučevanju odvinosti pri masovnih pojavih bomo videli, da zveze med masovnimi pojavi niso funkcionalne. Posameznim vrednostim enega znaka ne odgovarja točno določena vrednost drugega znaka, ki je z njim povezan in te vrednosti ni možno iz poznavanja vrednosti prvega znaka točno določiti. Zaradi tega imenujemo zveze, ki nastopajo pri masovnih pojavih, za razliko od funkcionalnih, statistične zveze. S tujim izrazom, ki se je mednarodno udomačil imenujemo statistične zveze koleracijske zveze ali odvisnosti (iz angleščine correlation = vzajemnost).

Predno preidemo na podrobno proučevanje, poglejmo na primeru, katere izmed zgornjih načinov bomo mogli uporabiti tudi za prikazovanje statističnih odvisnosti. Zaenkrat podajanje statističnih odvisnosti v obliki enačb ne pride v poštev, pač pa jih moramo prikazati kot sistem dvojic ali grafično kot sistem točk.

P Vzemimo kot primer težo in višino 5o dijakov ljubljanskih gimnazij v starosti od 13 do 18 let (merjenja izvršena v letu 1951). Iz poznavanja vsakdanjega življenja moramo sklepati, da obstoji povezava med tema dvema pojavema. Teža se veča z višino. Podatki, ki so nam na razpolago so teža in višina vsakega posameznega dijaka. Imamo torej za vsako enoto (dijaka) dva znaka: težo in višino. Ker smo že navajeni zaznamovati statistične znake z x, y, u itd., moramo tudi tu uporabiti te znake, ki v matematiki pomenijo spremenljivke in zaznamujemo težo kot x, višino pa kot y. Teža odgovarja eni, višina pa drugi spremenljivki. Iz navedenega je razvidno, da moramo statistično odvisnost med dvema pojavema prikazati kot sistem dvojic teže (x) in višine (y). Enako moremo prikazati te dvojice tudi kot točke v koordinatnem sistemu. Za naš primer 5o dijakov so podatki navedeni v tabeli 12.2

Podatki v razpredelnici 12.2 so podani nepregledno, vendar moramo iz njih pri medsebojni primerjavi ugotoviti, da je višina dijakov v splošnem večja, ako je tudi teža večja. Ta stavek pa velja samo na splošno, za celotno maso, ker imamo posamezne primere, da so dijaki, ki so enako visoki, različno težki (n.pr. dijaka z redno številko 5 in 7 sta oba visoka 158,5 cm tehta pa eden 46, drugi pa 48,5 kg). Še več: najdemo celo primere, da je dijak, ki je manjši od svojega součenca, težji od njega (n.pr. dijak z redno številko 28, ki meri 166,5 cm in tehta 64,5 kg, je manjši od dijaka z redno številko 23, ki meri 176 cm, in težji od njega, ker tehta drugi dijak samo 56 kg, kar je v nasprotju s splošno težnjo.)

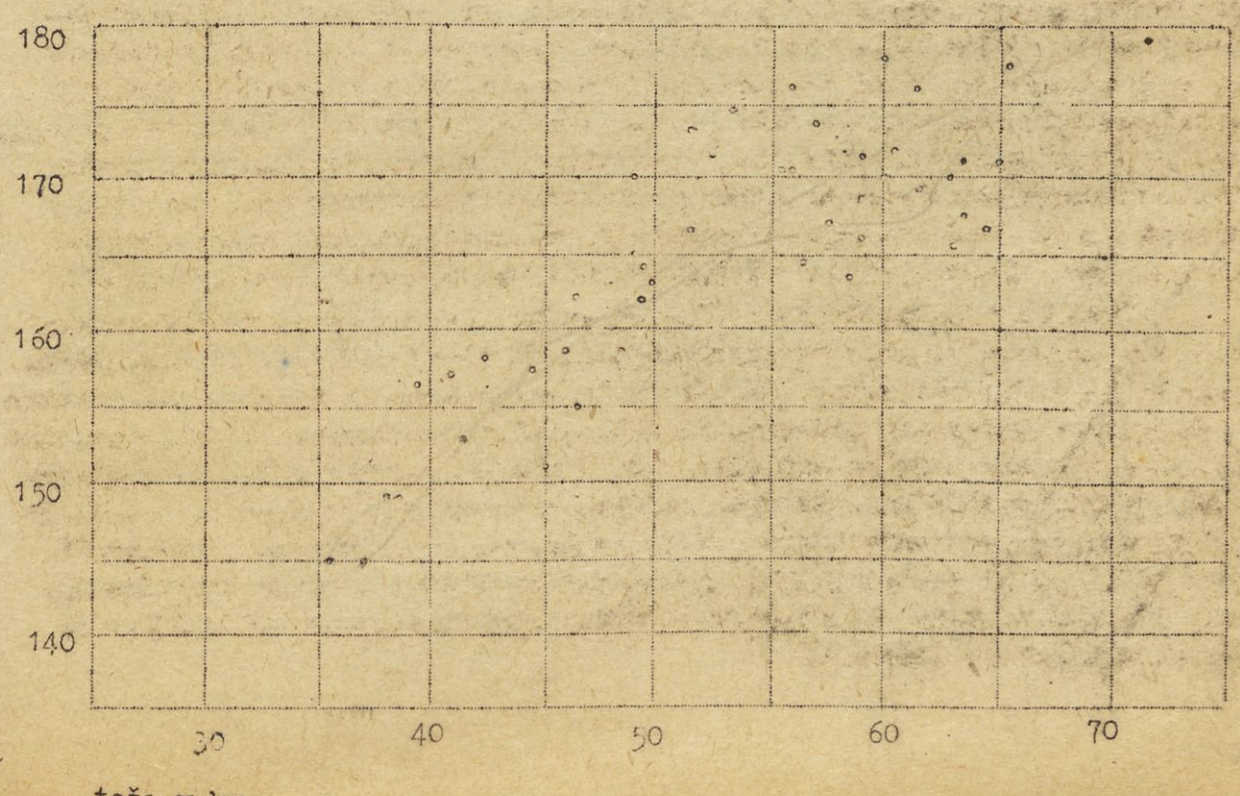
Teže in višine 5o dijakov ljubljanskih gimnazij
x = teža v kg y = višina v cm

Tab. 12.2

R.št.	x	y	R.št.	x	y	R.št.	x	y	R.št.	x	y
1	39,5	156,5	11	58.5	163.5	21	38,5	149,9	46	55.5	17o.5
2	41	157	12	56.5	164.5	22	37	145	47	6o	17o
3	44,5	157,5	13	53	164.5	23	56	176	48	6o	178
4	42,5	158	14	49.5	162	24	61.5	169.5	49	71.5	179
5	46	158.5	15	49.5	164	25	57	166	5o	61.5	176
6	46,5	155	16	38	149	26	57.5	167			
7	48,5	158.5	17	45	151	27	57.5	165.5			
8	46,5	155	18	37,5	151,5	28	64,5	166,5			
9	5o	163	19	41,5	153	29	51,5	166,5			
1o	46,5	162	2o	35,5	145	3o	59	168,5			
31	49	165	36	6o,5	172	41	65	171			
32	59	166	37	56	17o,5	42	57	173,5			
33	65,5	177,5	38	53,5	171	43	59	171,5			
34	63	165,5	39	49	17o	44	53,5	174,5			
35	63,5	167,5	4o	63	17o	45	51 5	173			

Zgornje podatke moremo veliko nazorneje podati grafično v pravokotnem koordinatnem sistemu, kjer je na abscisi x nanešena teža, v ordinati y pa višina.

Sl. 12.4 Korelacijski grafikon med težo in višino 50 dijakov



Vsaka posamezna točka predstavlja podatke o teži in višini enega dijaka. Iz slike je razvidno, da so točke v splošnem, ne pa v posameznih primerih vedno višje, ako se pomikamo od leve na desno. Vsebinsko to pomeni, da se višina veča, ako se veča teža. Slika je podobna meglici, ki se vleče poševno od spodaj navzgor, ako gremo od leve na desno. Grafikon lepo pokaže povezanost med obema podatkoma. Razvidno je tudi, da odvisnost med težo in višino ni funkcionalna, ker se z večanjem teže veča le raven višin ne pa posamezne višine. Ker grafikon te vrste pokaže korelacijsko povezanost med dvema pojavoma, ga imenujemo korelacijski grafikon.

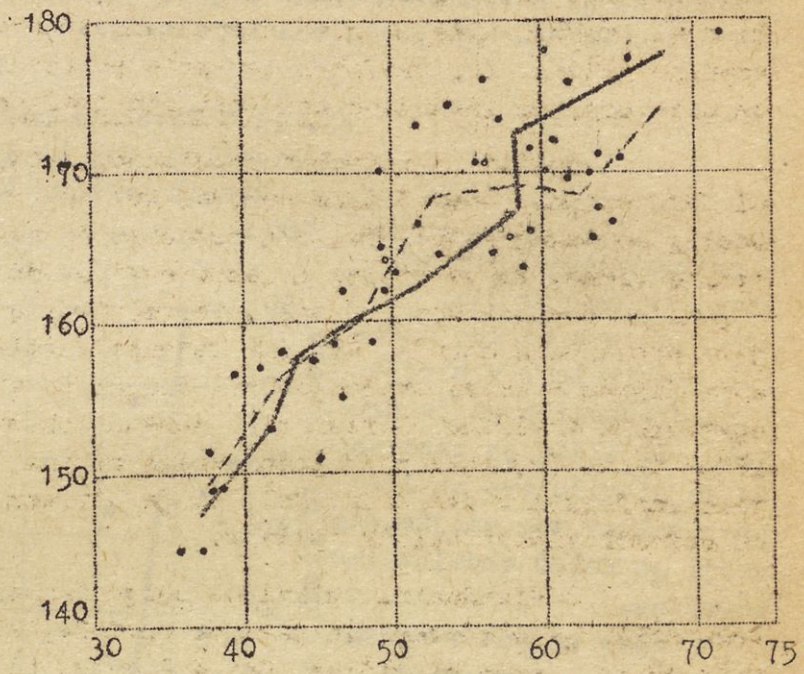
Rešiti je potrebno vprašanje, zakaj odvisnost med masovnimi pojavi ni fukcionalna, temveč take narave, kot je v primeru teže in višine dijakov. Razlog bo takoj razumljiv, ako proučimo že samo naš primer. Iz slike je nesporno vidno, da je višina dijakov odvisna od teže. Vendar višina posameznega dijaka ni odvisna samo od teže, temveč tudi od velikega števila drugih pogojev: socialnega stanja, splošne telesne razvitosti itd. Ti pogoji vplivajo pri enem dijaku v smeri zvečanja, pri drugem pa v smeri zmanjšanja teže, ki je opredeljena z višino. Zaradi tega moramo dobiti pri isti višini dijake z različno težo. Dejanska teža posameznega dijaka, ki je rezultat vseh vplivov, moremo razdeliti v dva dela: del, ki je odvisen od višine, in del, ki je odvisen od ostalih individualnih vplivov.

Individualni vplivi delujejo raznosmerno. To pomeni, da se dejanske teže odklanjajo navzdol in navzgor od teže, ki bi jo imeli dijaki, ako bi bila odvisna samo od višine. Oni del teže, ki zavisi od višine, je torej na grafikonu dan s črto, ki teče nekje med dejanskimi vrednostmi. Čeprav te črte ne poznamo, je možno iz grafikona točk, ki prikazuje povezavo, na oko ugotoviti približno smer in obliko zveze. Smer in oblika sta večkrat tako izrazita, da je možno črto povezanosti, ki teče med dejanskimi vrednostmi, narisati precej dobro že kar prostoročno. Iz zgodovinskih razlogov imenujemo črte odvisnosti regresijske črte ali krivulje.

Zaradi važnosti spoznanja različnih vrst povezav in zaradi tega, da bomo videli, kaj moremo razbrati iz korelacijskega grafikona, podajamo korelacijske grafikone za nekaj korelacij iz različnih področij.

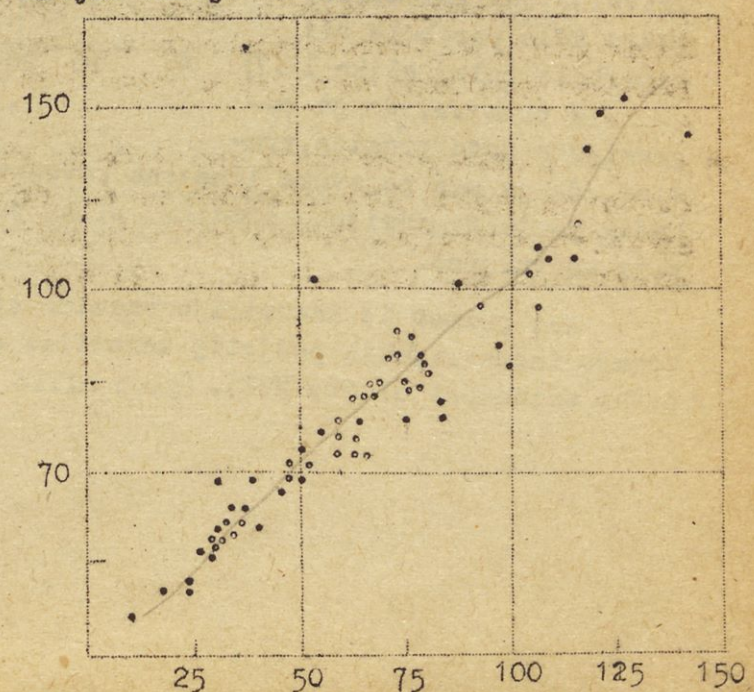
P Ako ponovimo primer teže in višine, je kljub temu, da je 50 enot razmeroma majhno število, možno precej dobro ugotoviti smer in tudi obliko regresijske krivulje. Precej dobro je smer in oblika prikazana, ako potegnemo med točkami kar premico. (glej sliko 12.5)

sl. 12.5 Korelacijska povezava z regresijsko črto med težo in višino 50 dijakov



Drugi primer vzemimo iz kmetijstva. Za vsak krajevni ljudski odbor okraja Maribor-okolica vzemimo iz popisa kmetijskih strojev in orodij število plugov in število bran. Že iz vsebine pojavov moramo rakljućiti, da med tema dvema podatkoma obstoja korelacija. Oba sta tesno povezana s tretjim podatkom, obdelovalno, oz. njivsko površino. Ker sta odvisna od istih vzrokov, je jasno, da mora obstojati tudi med njima samima povezava, ki pa ni vzročna. (glej sliko 12.6)

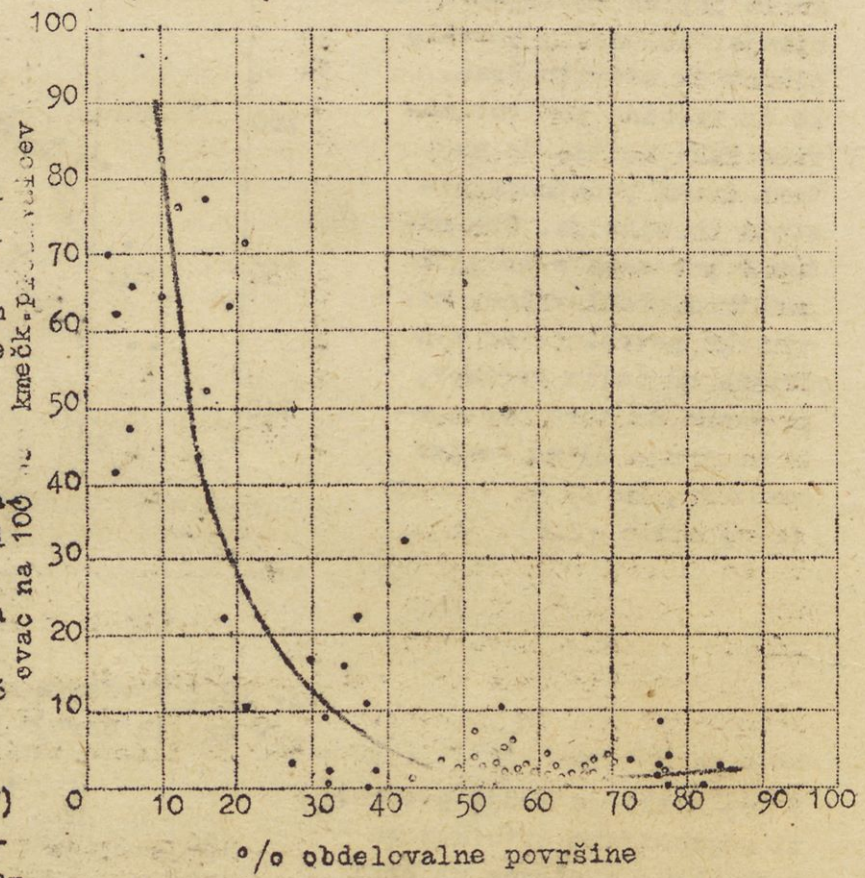
sl. 12.6 Povezava med številom plugov in bran v krajih okraja Maribor - okolica



V tej sliki je lažje potegniti regresijsko črto, ker so točke bolj nanizane na neki krivulji kot v prvem primeru. Individualni odkloni so v tem primeru razmeroma manjši. Regresijska črta je izrazita premica, ki gre približno pod kotom 45°,° skozi izhodišče. To pomeni, da je število plugov po krajevnih ljudskih odborih približno enako številu bran. Imamo sicer posamezne primere, v katerih so odkloni znatni, vendar so to izjeme.

Še en primer iz kmetijstva iz istega okraja (Maribor - okolica) nam bo pokazal, da regresijska krivulja ni v vseh primerih premica. Proučujmo zvezo med deležem obdelovalne površine in številom ovac na 100 kmečkih prebivalcev. Te vrste povezave, kot jo imamo v tem primeru, dosedaj še nismo imeli. Dosedaj se je z večanjem enega pojava večala tudi raven drugega. Pravimo, da med pojavoma obstoja pozitivna korelacija.

sl. 12.7 Povezava med deležem obdelovalne površine in številom ovc na 100 kmečkih gospodarstev v krajih okraja Maribor - okolica



V povezavo med deležem obdelovalne površine in številom ovc na 100 kmečkih prebivalcev v okraju Maribor-okolica pa je drugače. Iz slike je razvidno da je število ovc na 100 prebivalcev tem manjše, čim večji je delež obdelovalne površine.

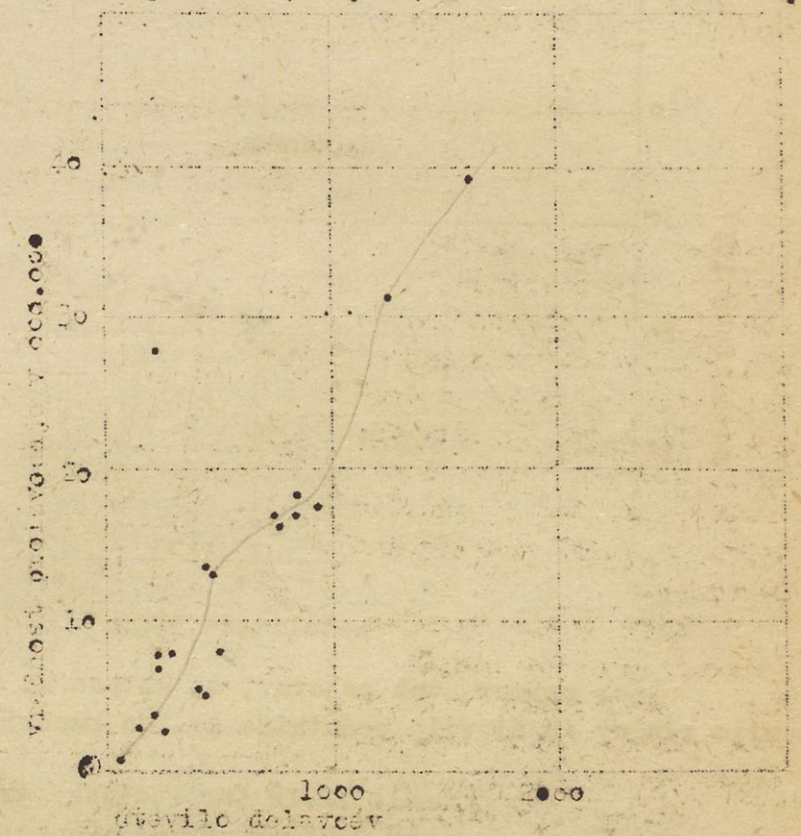
Kjer se kot v našem primeru en pojav manjša ako se drug pojav veča, pravimo, da med njima obstoji negativna korelacija.

Poleg tega vidimo, da v tem primeru regresijska krivulja ni premica, temveč krivulja, ki ima obliko enega dela parabole ali hiperbole (glej sliko 12.7)

V primeru, da je regresijska črta premica, pravimo, da je med pojavoma linearna povezava in govorimo o linearni regresiji, kadar pa je regresijska črta krivulja, pa pravimo, da je povezava kot regresija krivuljčna.

Kot primer iz industrije vzemimo število delavstva in vrednost proizvodnje industrijskih podjetij tekstilne stroke v LRS za mesec maj 1951. (Podatki evidenčne službe ZSE). (glej sliko 12.8)

Sl. 12.8 Povezanost med številom delavcev in vrednostjo proizvodnje delovnih v tekstilni industriji ZRS (... 1951)

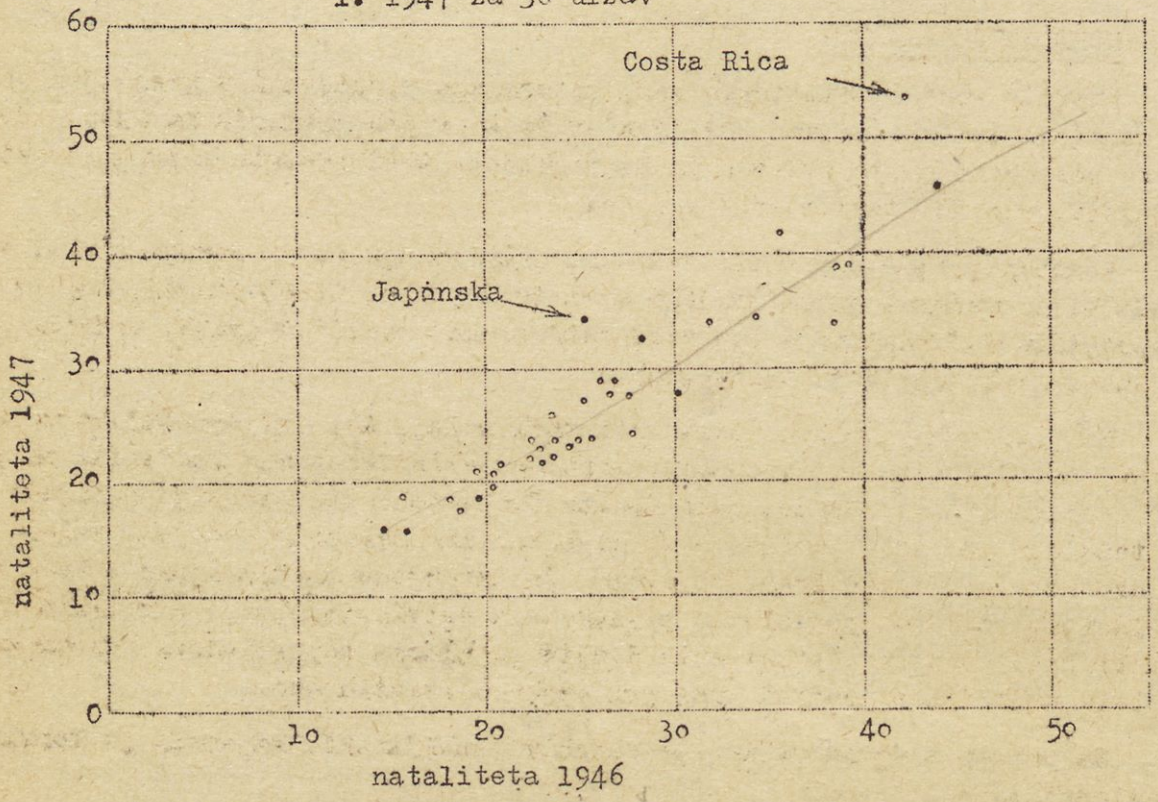


Iz slike je razvidna izrazita korelacijska povezanost in sicer pozitivna. Čim večje je število delavcev, tem večja je tudi vrednost proizvodnja.

Iz demografske statistike vzemimo kot primer nataliteto v letu 1946 in 1947 za 36 držav (podatki mesečnega biltena Urada za statistiko pri Združenih narodih). Nataliteta je število rojstev v enem letu na 1000 prebivalcev. Pogoji, ki vplivajo na višino natalitete so od države do države zelo različni. Ti pogoji pa so take narave, da se v razdobju enega leta bistveno ne spremene. Zaradi tega obstoja med obema podatkoma korelacija in sicer pozitivna. Če je nataliteta v letu 1946 visoka, je visoka tudi v letu 1947 in obratno, za državo za katero je bila nizka v letu 1946, je nizka tudi v letu 1947.

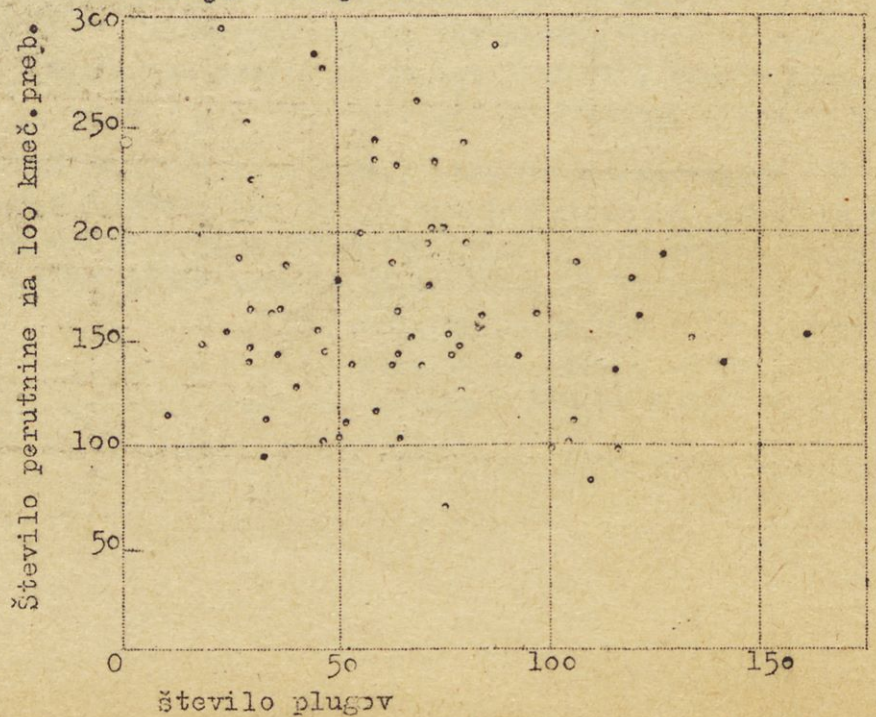
Veljavnost te analize potrjuje slika 12.9. Na sliki lepo vidimo, kako se točke, ki predstavljajo višino natalitete v letu 1946 in 1947, za posamezne države nizajo okrog regresijske črte, ki je premica. Le za dve državi (Japonsko in Costa Rico) se je nataliteta toliko spremenila, da imamo izredno velik odklon od premice.

sl. 12.9 Povezanost med natalitet. 1. 1946 in 1. 1947 za 36 držav



Kor primer dveh pojavov, za katera ni opazna korelacija, vzemimo število plugov in število perutnine na 100 kmečkih prebivalcev.

sl. 12.10 Korelacijski grafikon med številom plugov in številom perutnine na 100 kmečkih prebivalcev v krajih okraja Maribor - okolica



Iz slike je razvidno, da ne moremo med temi točkami potegniti premice ali krivulje, ki bi podajala smer povezave. Ta dva pojava med seboj nista povezana ali pa ta povezava ni opazna.

123. KORELACIJSKA KARTA

Število enot statističnih mas, katere smo proučevali v prejšnjem odstavku je bilo razmeroma majhno (največ 72). Vendar je že v teh primerih navajanje dvojic podatkov za posamezne enote obširno in nepregledno. Obširnost in nepreglednost pa se še stopnjujeta, ako je število enot veliko.

Podoben primer smo imeli tudi pri proučevanju enega samega znaka. Individualni podatki so večkrat predstavljali maso tudi po več tisoč števil, kar je bilo povsem nepregledno. Zaradi tega smo vrednosti znaka uredili v grupe, podatke pa podali urejene v frekvenčnih distribucijah.

Na podoben način bomo postopali tudi sedaj. Ker pa proučujemo povezanost dveh znakov, torej dva znaka istočasno, moramo statistično maso razdeliti ne po enem, temveč po dveh znakih istočasno. Tako dobimo frekvenčno distribucijo v obliki kombinacijske tabele, razdeljeno istočasno po dveh znakih. Ker kombinacijska tabela take vrste pokaže korelacijo med dvema pojavoma, jo imenujemo korelacijsko tabelo ali karto. Korelacijsko tabelo uporabljamo predvsem za statistične mase z velikim številom enot, ker bi bilo navajanje posameznih dvojic popolnoma nepregledno. Vendar korelacijska tabela nudi dober pregled tudi pri majhnem številu enot.

Za primer teže in višine 50 dijakov ljubljanskih gimnazij je korelacijska tabela naslednja:

Tab. 12.3 Korelacijska karta med težo in višino 50 dijakov

Višina v cm \ Teža v kg	30-	35-	40-	45-	50-	55-	60-	65-	70-	Skupno
175-						1	2	2	1	6
170-				1	2	4	4			11
165-				1	1	5	4			11
160-				3	2	2				7
155		1	3	5						9
150		1	1							2
145		4								4
Skupno		6	4	10	5	12	10	2	1	50

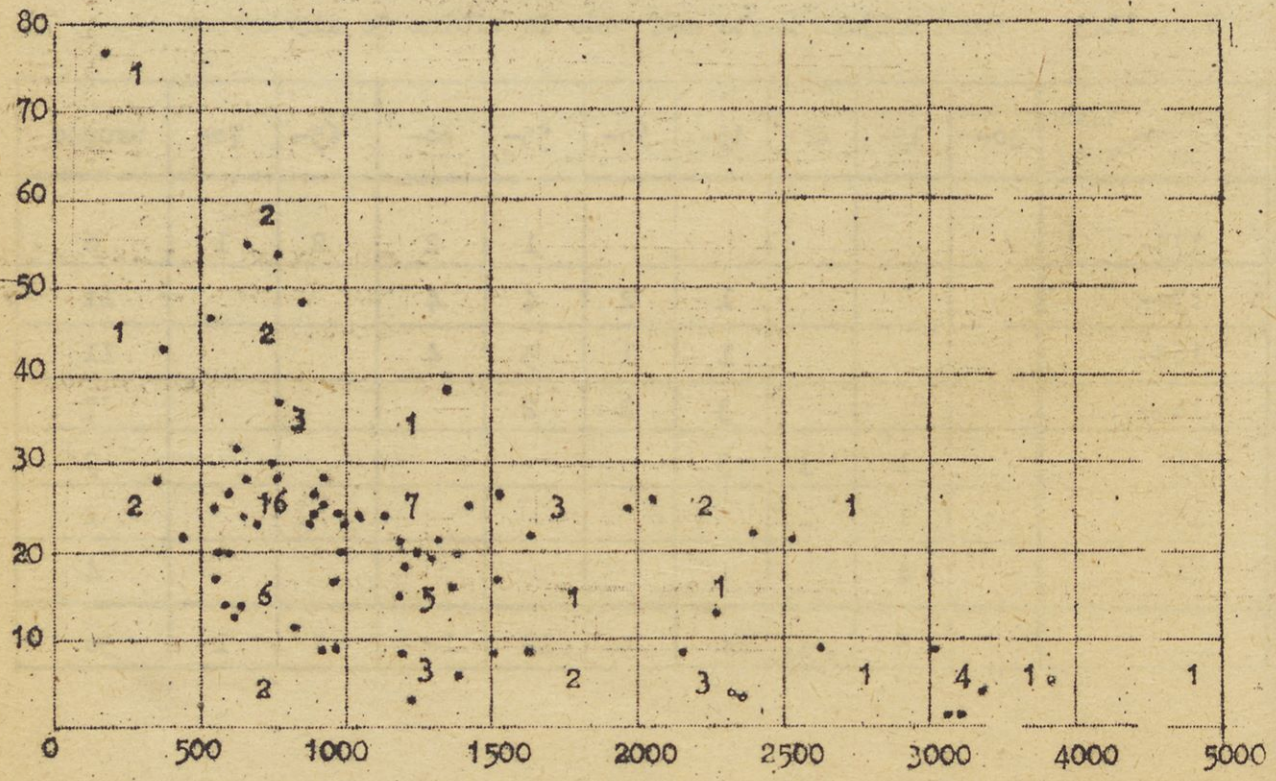
Števila v okencih so frekvence in povedo, koliko enot ima težo in višino v določenem razredu. Iz korelacijske tabele je n.pr. razvidno, da ima 5 dijakov težo od 45 do 5o kg in so visoki od 155 do 16o cm.

Korelacijsko tabelo moremo dobiti tudi na drug način. V korelacijskem grafikonu razdelimo celotno ravnino na pravokotnike na ta način, da k mejam razredov 35, 4o, 45, 5o .... kg napravimo vzporednice vodoravno k navpični osi, k mejam razredov višine 14o, 145, 15o, 155 .... kg pa vzporednice k vodoravni osi. Celotna ravnina je razdeljena na pravokotnike, ki so široki "5 kg", visoki pa "5 cm". Frekvenca v korelacijski tabeli pomeni število točk, ki leže v odgovarjajočem pravokotniku korelacijske karte.

Korelacijska tabela je v toliko manj natančna od korelacijske karte, da pove samo, koliko enot je v določenem delnem pravokotniku ravnine, ne pokaže pa točno mesto, kje znotraj pravokotnika leže posamezne točke.

Zvezo med korelacijsko karto in korelacijsko tabelo ponazorimo še na primeru korelacije med skupno površino in odstotnim deležem njiv. (72 KLO-jev okraja Maribor okolica)

sl. 12.11 Zveza med korelacijskim grafikonom in korelacijsko karto. Povezanost med skupno površino in deležem njiv v krajih okraja Maribor okolica



Tako iz korelacijskega grafikona kot iz same tabele je razvidno, da je korelacija v tem primeru negativna, ker se z večjo skupno površino manjša odstotni delež njiv.

Ker je korelacijska tabela enakovredna korelacijskemu grafikonu, moremo korelacijsko tabelo prav tako uporabiti za proučevanje korelacije med dvema pojavoma kot korelacijski grafikon. Iz nje moremo razbrati smer in obliko regresijske črte. Za statistične mase z velikim številom enot, podatke, ki naj pokažejo povezavo med dvema pojavoma, nikdar ne podajamo v obliki sistema individualnih dvojic vrednosti, temveč vedno v obliki korelacijske tabele.

Dva primera naj pokažeta uporabnost korelacijskih tabel.

Med številom stanovalcev in velikostjo stanovanj, ki je označeno s številom sob, je pozitivna korelacija. Poglejmo na kakšen način, je to razvidno na primeru 197 stanovanj nekega popisnega okoliša v Ljubljani. (Podatki vzeti iz popisa stanovanj). (glej tabelo 12.4)

Tab. 12.4 Korelacijska karta med številom sob in številom stanovalcev
x = število sob, y = število stanovalcev

x/y	1	2	3	4	5	6	7	8	9	1o	11	Sk.
5						4	2	2	1	2	1	12
4				1	3	6	9	5	3	1	1	29
3		2	2	12	14	16	11	4	1			62
2	1	3	11	24	12	4	4					59
1	4	9	16	6								35
Sk.	5	14	29	43	29	3o	26	11	5	3	2	197

Točno je razvidno, da je med številom stanovalcev in številom sob pozitivna korelacija, ker se frekvence pomikajo od leve na desno in od spodaj navzgor. V večjih stanovanjih je v splošnem večje število stanovalcev.

Kot drug primer vzemimo starost ženinov in nevest novoporočencev v LRS v letu 1949. (Tabela 12.5)

Tab. 12.5 Korelacijska karta med starostjo ženina in nevesto LRS leta 1949

Nev. st. / St. žen.	-2o	2o-24	25-29	3o-34	35-39	4o-44	45-49	5o-54	55-59	6o-	Skupno
6o-						1	1	2	6	11	21
5o-54						4	13	12	3	22	54
45-49		1		4	3	13	21	24	11	26	1o5
45-49		1	5	5	23	34	5o	39	32	36	225
4o-44		8	18	27	1o2	124	97	63	29	28	496
35-39	4	27	81	111	278	212	122	47	28	18	928
3o-34	3	68	227	245	283	163	78	31	16	11	1125
25-29	32	151	1715	8o8	64o	256	115	26	7	1	4172
2o-24	196	21oo	2649	642	389	1oo	26	12	4	4	6192
-2o	12o	868	531	96	43	15	6	1			168o
Skup.	335	3644	5226	1938	1761	923	529	257	136	139	14927

Tudi v tem primeru je razvidna odvisnost, ki obstoja med starostjo ženina in nevesto. Iz korelacijske karte je razvidno, da starost neveste zavisi od starosti ženina, vendar je ta zveza vse proj kot funkcionalna in so individualni vplivi zelo močni.

124. ČRTA GRUPNIH ARITMETIČNIH SREDIN

Že v odstavku 122 smo nakazali, da se zakonitost povezave ne pojavlja nujno v individualnih primerih, temveč, da se odvisnost in povezava kaže v spremembi splošne ravni vrednosti.

Proučimo to vprašanje na primeru korelacije med številom sob in številom stanovalcev. Tabela je zaradi boljšega razumevanja preširitve postavljena po koncih:

Tab. 12.6 Število sob in število stanovalcev
x = število stanovalcev, y = štev. sob

x / y	1	2	3	4	5	Σ
11				1	1	2
1o				1	2	3
9			1	3	1	5
8			4	5	2	11
7		4	11	9	2	26
6		4	16	6	4	3o
5		12	14	3		29
4	6	24	12	1		43
3	15	11	2			29
2	9	3	2			14
1	4	-				4
Σ	35	59	62	29	12	197

V stanovanjih z eno sobo stanujejo po ena do štiri osebe. Ako bi bilo število stanovalcev odvisno samo od števila sob, bi bilo vseh 35 enosobnih stanovanj zasedeno z enakim številom stanovalcev. Prav tako bi bilo v vseh 59 dvosobnih stanovanjih enako število stanovalcev, ker bi bil edini faktor, ki bi vplival na število stanovalcev, to je število sob, za vseh 59 stanovanj

isti in sicer dve sobi. Isto bi veljalo za dvo, tro, štiri in petsobna stanovanja. Na število stanovalcev pa ne vpliva samo število sob, temveč še drugi individualni vplivi, ki se od primera do primera menjajo. Zaradi tega je število stanovalcev tudi v stanovanjih, ki imajo enako število sob, med seboj različno. Vendar je vpliv števila sob na število prebivalcev iz primera očiten.

V povprečju se učinek individualnih vplivov uniči (glej poglavje o srednjih vrednostih). Aritmetična sredina podaja splošno raven vrednosti. Zaradi tega moremo regresijsko črto dobiti tudi s pomočjo aritmetičnih sredin. Za vsako vrsto stanovanj (enosobno, dvosobno, trisobno itd) izračunajmo aritmetično sredino. Ako je med številom sob in številom stanovalcev pozitivna korelacija, je nujno, da je za večja stanovanja tudi povprečno število stanovalcev na eno stanovanje večje. Odkloni od tega pravila so možni le v primeru, ako število enot v posameznih razredih ni tako veliko, da bi se učinek individualnih vplivov povsem uničil.

Ako izračunamo za posamezno velikost stanovanj povprečno število stanovalcev na eno stanovanje dobimo naslednje rezultate v tabeli 12.7.

Tab. 12.7

Število sob	Povprečno število stanovalcev na eno stanovanje
1	2,7
2	4,2
3	5,5
4	7,1
5	7,8

Iz te serije sredin (tabela 12.7) je jasno razvidna pozitivna korelacija med številom sob in povprečnim številom stanovalcev za eno stanovanje, ker se aritmetična sredine večajo z večanjem stanovanj. Ako te vrednosti včrtamo v grafikon podajajo regresijsko črto, ki ni včrtana na oko, temveč na osnovi izračunanih količin. Zaradi tega je tudi objektivnejša. Proučevanje pa moremo razviti tudi v drugo smer. Zgoraj smo iskali, kako se spreminja število stanovalcev v odvisnosti od števila sob. Moremo pa se vprašati tudi obratno: v kakšnih stanovanjih stanujejo eno, dvo, tro ---- enajst članske družine. Ako hočemo dognati to, moramo proučevati posamezne vrste tabele, ne pa kolone. Iz tabele je jasno razvidno, da večje družine stanujejo v večjih stanovanjih. Na enak način kot smo najprej izračunali povprečno število stanovalcev enosobnih, dvosobnih, trisobnih itd. stanovanj, moremo sedaj izračunati povprečno število sob stanovanj, ki jih zasedajo eno, dvo, tro in več članske družine. Ti podatki so dani v razpredelnici:

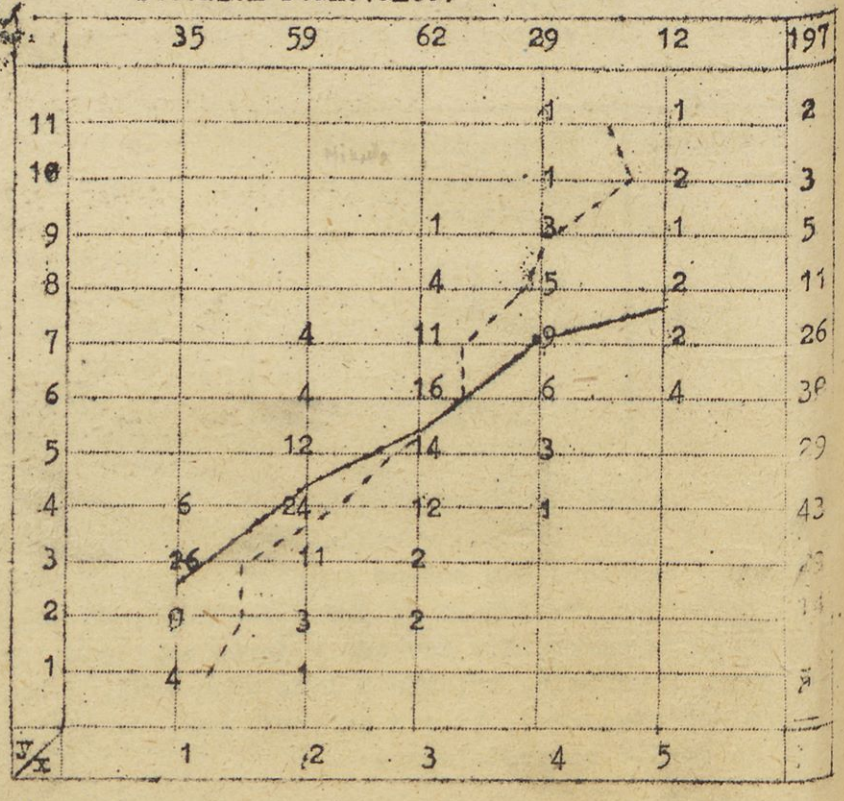
Tab. 12.8

Število stanovalcev na eno stanovanje	1	2	3	4	5	6	7	8	9	1o	11
Povprečno število sob	1,2	1,5	1,5	2,2	2,7	3,3	3,4	3,8	4	4,7	4,5

Ako vrišemo te podatke v korelacijski grafikon, dobimo drugo regresijsko črto, narisano na osnovi serije grupnih povprečij. Vendar ta črta tako vsebinsko kot po obliki ni ista kot prva. Medtem ko prva predstavlja regresijsko črto, ki pokaže odvisnost števila stanovalcev od števila sob, poda druga regresijska črta število sob v odvisnosti od števila družinskih članov. Tudi grafično sta to dve različni krivulji.

Tako dobljeni regresijski črti moremo vnesti tudi v korelacijsko karto (slika 12.12)

sl. 12.12
Regresijski črti povezanosti med številom sob in številom stanovalcev



Ako bi risali regresijsko črto na oko, bi tudi mogli vrisati dve krivulji in sicer tako, da bi iskali krivuljo, ki bi šla med realnimi vrednostmi gledano enkrat v navpični, drugič pa v vodoravni smeri.

Ako pogledamo sliko 12.12 moremo opaziti še neko važno značilnost. Lomljena črta, ki prikazuje odvisnost števila stanovalcev od števila sob je bolj izglajena, kot pa črta, ki prikazuje odvisnost števila sob od števila stanovalcev. Ta razlika nastopi zaradi tega, ker učinek individualnih vplivov pri drugi črti ni povsem odstranjen, ker so aritmetične sredine izračunane na osnovi razmeroma majhnega števila primerov (2,3,5 itd).

Splošen zaključek, ki ga moremo napraviti iz tega primera je:

1./ Regresijski črti, ki kažeta smer in obliko povezave dveh pojavov sta dve. Prva pokaže odvisnost prvega pojava od drugega, druga pa pokaže odvisnost drugega pojava od prvega.

2./ Regresijski črti moremo določiti s pomočjo izračunavanja grupnih sredin enega znaka pri različnih vrednostih drugega znaka.

125. REGRESIJSKI PREMICI

125.1 Enačbi

Ako pogledamo regresijski črti, ki smo jih dobili s pomočjo grupnih sredin, vidimo, da sta v našem primeru ti dve črti zelo podobni premicam, v kolikor ne odklanjajo posamezne sredine od te premice individualni vplivi, zaradi premajhnega števila primerov v posameznih grupah. Ker so tudi v sredinah odkloni individualnih vplivov raznosmerni, moremo včrtati premico že na oko. Imamo pa postopek, s katerim moremo v primerih, v katerih sta regresijski krivulji premici, enačbi regresijskih premic dobiti z izračunavanjem.

Iz analitične geometrije vemo, da je enačba premice, ki gre skozi točko (x_o , y_o) in je tangens naklonskega kota enak " k " enaka:

$$y - y_o = K (x - x_o) \qquad (12.1)$$

To je splošna enačba premice, ki jo moremo za vsak poseben primer tudi narisati.

Kot bomo videli kasneje na dokazu, je enačba regresijske premice, ki kaže smer odvisnosti znaka y od znaka x, dana z obrazcem:

$$y' - \bar{y} = b_1 (x - \bar{x}) \qquad (12.2)$$

V tej enačbi je $\bar{x}$ aritmetična sredina znaka x, $\bar{y}$ aritmetična sredina znaka y, koeficient b_1 pa imenujemo prvi regresijski koeficient. Regresijski koeficient b_1 izračunamo iz dvojic podatkov x in y po obrazcu:

$$b_1 = \frac{C_{xy}}{\sigma_x^2} \qquad (12.3)$$

σ_x^2 je varianca znaka x, C_{xy} pa je nov izraz. Imenujemo ga kovarianca.

Kovarianca C_{xy} je aritmetična sredina produktov odklonov znaka x od $\bar{x}$ in odklonov y od $\bar{y}$ za posamezne vrednosti. V obliki obrazca moremo kovarianco po zgornji opredelitvi pisati:

$$C_{xy} = \frac{1}{n}\sum_i (x_i - \bar{x}) (y_i - \bar{y}) \qquad (12.4)$$

Ker je izračunavanje odklonov od aritmetične sredine $x_i - \bar{x}$ in $y_i - \bar{y}$ razmeroma zamudno, običajno izračunavamo kovarianco po obrazcu, ki je zgornjemu istoveten:

$$c_{xy} = \frac{1}{n}(x_1 y_1 + x_2 y_2 + x_3 y_3 + \dots x_n y_n) - \bar{x}\bar{y} = \frac{1}{n}\Sigma x_i y_i - \overline{\bar{x}\bar{y}} = \overline{xy} - \bar{x} \cdot \bar{y} \quad (12.5)$$

$\overline{xy}$ pomeni aritmetično sredino vseh produktov individualnih vrednosti x z y. K vrednosti x pripadajočo vrednost znaka y na premici zaznamujemo z y, za razliko od dejanske vrednosti y za enoto i.

Enačba druge regresijske premice, ki kaže odvisnost znaka x od y pa je dana z obrazcem:

$$x' - \bar{x} = b_2 (y - \bar{y}) \qquad (12.6)$$

b_2 je drugi regresijski koeficient in se izračuna po obrazcu:

$$b_2 = \frac{c_{xy}}{\sigma_y^2} \qquad (12.7)$$

c_{xy} je kovarianca, σ_y^2 pa je varianca znaka y.

Iz istega razloga, kot smo pisali y', pišemo v drugi regresijski premici z x' vrednost znaka x na regresijski premici.

V preglednejši obliki moremo pisati enačbi regresijskih premic in izračunavanje potrebnih koeficientov na naslednji način:

$$y' - \bar{y} = b_1 (x - \bar{x}) \quad b_1 = \frac{c_{xy}}{\sigma_x^2} \quad \sigma_x^2 = \overline{x^2} - \bar{x}^2$$

$$c_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} \quad (12.8)$$

$$x' - \bar{x} = b_2 (y - \bar{y}) \quad b_2 = \frac{c_{xy}}{\sigma_y^2} \quad \sigma_y^2 = \overline{y^2} - \bar{y}^2$$

Iz enačb regresijskih premic je razvidno, da se obe regresijski premici sečeta v točki s koordinatama $(\bar{x}, \bar{y})$

125.2 Shema in postopek izračunavanja regresijskih premic sta naslednja:

	i	x_i	y_i	x_i^2	y_i^2	$x_i\ y_i$
		(1)	(2)	(3) = $(1)^2$	(4)=$(2)^2$	(5)=(1).(2)
	1	x_1	y_1	x_1^2	y_1^2	$x_1\ y_1$
	2	x_2	y_2	x_2^2	y_2^2	$x_2\ y_2$
	⋮	⋮	⋮	⋮	⋮	⋮
	n	x_n	y_n	x_n^2	y_n^2	$x_n\ y_n$
$\sum$	n	x	y	x^2	y^2	xy
$A=\frac{1}{n}\sum$	1	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{y^2}$	$\overline{xy}$
			B	$\bar{x}^2$	$\bar{y}^2$	$\bar{x}\cdot\bar{y}$
			C = A-B	σ_x^2	σ_y^2	c_{xy}

(12.9)

$$b_1 = \frac{c_{xy}}{\sigma_x^2} \qquad b_2 = \frac{c_{xy}}{\sigma_y^2}$$

$$y' - \bar{y} = b_1(x - \bar{x}), \quad x' - \bar{x} = b_2(y - \bar{y})$$

Postopek izračunavanja je naslednji:

1./ Za vse enote vpišemo dvojice osnovnih vrednosti znakov x in y v stolpca (1) in (2).

2./ Za vsako dvojico vrednosti x in y izračunamo kvadrat znaka x, kvadrat znaka y in produkt x . y. Te vrednosti vnesemo v pripadajoče stolpce (3), (4) in (5).

3./ Seštejemo vrednosti stolpcev od (1) do (5).

4./ Dobljene vsote delimo s številom enot in rezultate vpišemo v naslednjo vrsto. Na ta način dobimo v vrsti A količine,

$$\bar{x},\ \bar{y},\ \overline{x^2},\ \overline{y^2},\ \overline{xy}$$

5./ Pod te vrednosti podpišemo v vrsti B v stolpcu (3) $\bar{x}^2$, v stolpcu (4) $\bar{y}^2$, v stolpcu (5) pa $\bar{x} \cdot \bar{y}$.

6./ Ako vrednosti v vrsti B, odštejemo od vrednosti v vrsti A, dobimo v vrsti C končne količine

$$\sigma_x^2,\quad \sigma_y^2,\quad c_{xy}.$$

7./ Iz količin $\bar{x}$, $\bar{y}$, σ_x^2 σ_y^2 in c_{xy} izračunamo po navedenih obrazcih regresijska koeficienta in regresijski premici.

P Po zgornji shemi izračunajmo regresijski premici za težo in višino za 5o dijakov.

Zaradi dolžine računa bomo računanje z individualnimi vrednostmi v primeru samo nakazali za par enot. Ako vršimo računanje z računskim strojem vpisovanje računov za posamezne enote itak odpade, ker stroj kvadrate in produkte avtomatično sešteva.

Tab. 12.9

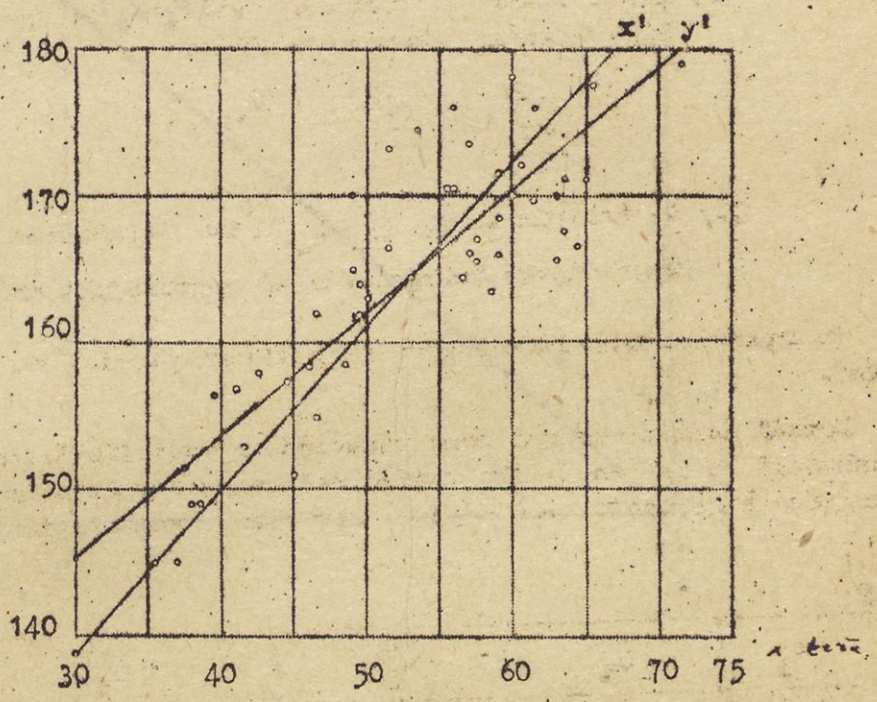
i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	(1)	(2)	(3) = $(1)^2$	(4) = $(2)^2$	(5)=(1).(2)
1	39,5	156,5	156o,25	24492,25	6181,75
2	41	157	1681	24649	6437
3	44,5	157,5	198o,25	248o6,25	7oo8,75
49	71,5	179	5112,25	32o41	12798,5o
5o	61,5	176	3782,25	3o976	1o824
Σ 5o	2641,5	822o,5	143480,25	1355239,25	437575
$A=\frac{1}{n}\Sigma$ 1	52,83	164,41	2869,6o5o	271o4,785o	8751,5ooo
		B =	2791,oo89	27o3o,6481	8685,78o3
		A -B = C	78,5961	74,1369	65,7197

$$b_1 = \frac{65,7197}{78,5961} = o,8363 \qquad b_2 = \frac{65,7197}{74,1369} = o,8865$$

$y' - 164,41 = o,8362\,(x - 52,83)$ $x' - 52,83 = o,8865\,(y - 164,41)$

Regresijski premici moremo vrisati v korelacijski grafikon

sl.12.13 Regresijski premici med težo in višino 50 dijakov ljubljanskih gimnazij



125.3 Dokaz enačb regresijskih premic zahteva nekaj znanja matematike, vendar ne presega znanja srednješolske matematike. Predvsem je potrebno poznavanje stavkov, ki so jih obravnavali pri srednjih vrednostih.

Izhajali bomo iz grafičnega prikaza korelacije med dvema pojavoma s točkami (glej sliko 12.14). En pojav zaznamujemo z x, drugega z y.

Proučujmo najprej odvisnost y-a od x. Regresijska premica, ki podaja smer regresije leži med točkami (glej sl.12.14). Naredimo predpostavko, da regresijska premica predstavlja rezultat vpliva x-a na y. K vsaki vrednosti x_i moremo na premici najti točko $(x_i\ y_i)$. Vrednost ordinate y_i' predstavlja torej rezultat vpliva znaka x na y.

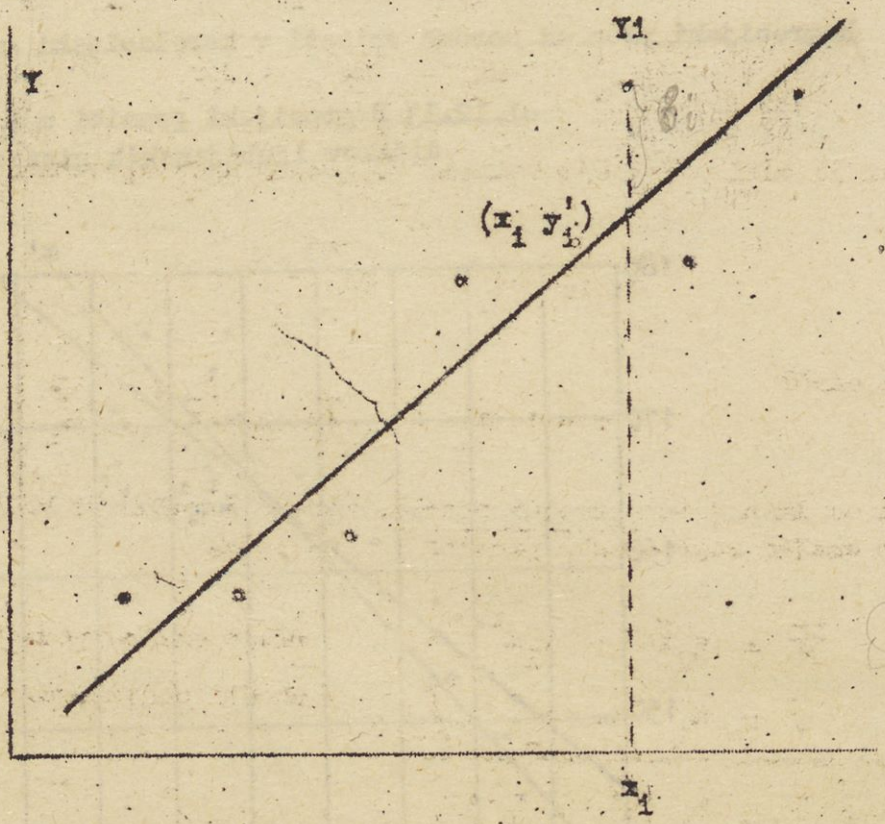
Odklon dejanske vrednosti y_i od y_i', $y_i - y_i' = \delta_i$, je torej rezultat ostalih, individualnih vplivov. Vrednost znaka y neke poljubne enote i moremo razdeliti na dva dela:

$$y_i = y_i' + (y_i - y_i') \qquad (12.10)$$

ali ker je

$$y_i - y_i' = \delta_i \qquad y_i = y_i' + \delta_i \qquad (12.11)$$

sl. 12.14



y_i' je odvisen samo od x_i, δ_i pa od individualnih pogojev, ki vplivajo na isto enoto. Ker je zveza med y_i' in x_i' linearna, je $y_i' = a + bx_i$. Iz tega sledi, da je $y_i = a + bx_i + \delta_i$

Individualni odkloni dejanskih vrednosti y_i vod premice, ki smo jih zaznamovali z δ_i, so pri nekaterih enotah pozitivni, pri drugih pa negativni. To zavisi od tega, ali točka leži nad ali pod premico.

Lastnost, da se učinek individualnih vplivov v povprečju uniči, katero smo navedli pri obravnavanju aritmetične sredine, moremo uporabiti tudi v našem primeru.

Ker je $y_i = a + bx_i + \delta_i$ in se δ_i v povprečju uniči, je:

$$\bar{y} = a + b\bar{x} + 0$$

Ako pomnožimo individualno enačbo $y_i = a + bx_i + \delta_i$ z x_i, dobimo:

$$y_i x_i = ax_i + bx_i^2 + \delta_i x_i$$

Ker je tudi povprečje količin $\delta_i \cdot x_i$ enako 0, dobimo:

$$\overline{xy} = a\bar{x} + b\overline{x^2} + 0$$

Iz enačb

$$\bar{y} = a + b\bar{x} \quad \text{in} \quad \overline{xy} = a\bar{x} + b\overline{x^2} \qquad (12.12)$$

ki se imenujeta normalni enačbi, moremo izračunati količini a in b, oziroma enačbo regresijske premice $y' = a + bx$.

$$\overline{xy} = a_1\bar{x} + b_1\overline{x^2}$$
$$\bar{y} = a_1 + b_1\bar{x} \quad / \cdot \bar{x}$$

Drugo enačbo pomnožimo z $\bar{x}$ in obe enačbi seštejemo.

$$\overline{xy} - \bar{x}\,\bar{y} = b_1(\overline{x^2} - \bar{x}^2)$$

Ker je $\overline{xy} - \bar{x}\cdot\bar{y} = C_{xy}$, $\overline{x^2} - \bar{x}^2 = \sigma_x^2$ je: $b_1 = \dfrac{C_{xy}}{\sigma_x^2}$

Iz zveze $y' = a + b_1 x$ in $\bar{y} = a + b_1\bar{x}$, pa moremo odstraniti a, ako drugo enačbo od prve odštejemo:

$$y' = a_1 + b_1 x$$
$$\bar{y} = a_1 + b_1\bar{x} \quad / \cdot (-1)$$

$$y' - \bar{y} = b_1(x - \bar{x})$$

Ker je $b_1 = \dfrac{C_{xy}}{\sigma_x^2}$ je prva regresijska enačba v končni obliki dana z enačbo:

$$y' - \bar{y} = \frac{C_{xy}}{\sigma_x^2}(x - \bar{x})$$

Na ta način smo prišli do enačbe prve regresijske premice.

Ako iščemo drugo regresijsko premico odvisnosti x-a od y-a se celoten postopek izpremeni v toliko, da zamenjamo x z y. Enačba druge regresijske premice se torej glasi:

$$x' - \bar{x} = \frac{c_{xy}}{\sigma_y^2} (y - \bar{y}).$$

Splošnejši in strožji dokaz bi mogli izvesti po takoimenovani metodi najmanjših kvadratov. Ta način ima za osnovo predpostavko, da je vsota kvadratov odklonov vrednosti y_i od premice najmanjša. Čeprav postopek izhaja iz drugih predpostavk je seveda končen rezultat, to je regresijski premici, isti kot smo jih dobili z našim dokazom.

Da je $c_{xy} = \overline{(x - \bar{x})(y - \bar{y})} = \overline{xy} - \bar{x} \cdot \bar{y}$ moremo dokazati na enostaven način:

$$\overline{(x - \bar{x})(y - \bar{y})} = \overline{xy - \bar{x}y - x\bar{y} + \bar{x}\bar{x}} = \overline{xy} - \overline{\bar{x} \cdot y} - \overline{x \cdot \bar{y}} + \overline{\bar{x} \cdot \bar{y}}$$

$$= \overline{xy} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} = \overline{xy} - \bar{x} \cdot \bar{y}.$$

125.4 Skrajšani postopek izračunavanja regresijskih premic

Pri izračunavanju varianc smo skrajšali postopek na ta način, da smo uvedli pogojno sredino. Enako v marsikaterih primerih vpeljava pogojne sredine olajša tudi izračunavanje enačb regresijskih premic.

Ker računamo pri odvisnosti z dvema znakoma x in y, moramo vpeljati tudi dve pogojni sredini, x_o za x in y_o za y. Kot vemo vzamemo kot pogojno sredino tako vrednost, da so odkloni individualnih vrednosti $x_i - x_o = u_i$ in $y_i - y_o = v_i$ količine, ki so čim prikladnejše za kvadriranje in množenje. Količine u_i in v_i izračunamo torej iz osnovnih vrednosti, tako, da vsaki vrednosti x-a odštejemo pogojno sredino x_o, vsaki vrednosti y-a pa pogojno sredino y_o.

Kot vemo iz prejšnjega poglavja, so količine, ki jih potrebujemo za izračunavanje regresijske premice naslednje:

$$\bar{x}, \bar{y}, \sigma_x^2, \sigma_y^2 \text{ in } c_{xy}.$$

Iz poglavja o aritmetičnih sredinah in varianci vemo, da je:

$$\bar{x} = x_o + \bar{u} \qquad \bar{y} = y_o + \bar{v} \qquad \sigma_x^2 = \sigma_u^2 \qquad \sigma_y^2 = \sigma_v^2.$$

$\bar{x}$, $\bar{y}$, σ_x^2 in σ_y^2 moremo izračunati iz samih količin x_o, y_o u in v. Zelo enostavno je dokazati, da moremo tudi C_{xy} izračunati iz količin u in v in sicer je $C_{xy} = C_{uv}$.

Dokaz: Ker je $x = x_o + u$, $\bar{x} = x_o + \bar{u}$ je $x - \bar{x} = u - \bar{u}$ Enako zaradi $y = y_o + v$, $\bar{y} = y_o + \bar{v}$ velja: $y - \bar{y} = v - \bar{v}$.

Individualni produkt $(x - \bar{x})(y - \bar{y})$ je enak $(u - \bar{u})(v - \bar{v})$. Zaradi tega je tudi aritmetična sredina produktov $(x - \bar{x})(y - \bar{y})$ enaka aritmetični sredini produktov $(u - \bar{u})(v - \bar{v})$, kar je istovetno stavku $C_{xy} = C_{uv}$.

Enačbi regresijskih premic sta torej enaki:

$$y' - (y_o + \bar{v}) = \frac{C_{uv}}{\sigma_u^2}\left[(x - (x_o + \bar{u})\right] \quad \text{in} \qquad (12.13)$$

$$x' - (x_o + \bar{u}) = \frac{C_{uv}}{\sigma_v^2}\left[(y - (y_o + \bar{v})\right] \qquad (12.14)$$

Shema in stopnje izračunavanja regresivnih premic so naslednje:

	i	x_i	y_i	$x_i - x_o = u_i$	$y_i - y_o = v_i$	u_i^2	v_i^2	$u_i v_i$
		(1)	(2)	(3)=(1)−x_o	(4)=(2)−y_o	(5)=(3)^2	(6)=(4)^2	(7)−(3)^2
	1	x_1	x_1	$x_1-x_o=u_1$	$y_1-y_o=v_1$	u_1^2	v_1^2	$u_1 v_1$
	2	x_2	y_2	$x_2-x_o=u_2$	$x_2-x_o=v_2$	u_2^2	v_2^2	$u_2 v_2$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	n	x_n	y_n	$x_n-x_o=u_n$	$y_n-y_o=v_n$	u_n^2	v_n^2	$u_n v_n$
$\sum$	n			u	v	u^2	v^2	uv
$A = \frac{1}{n}\sum$	1			$\bar{u}$	$\bar{v}$	$\bar{u}^2$	$\bar{v}^2$	$\bar{u}\bar{v}$
					B =	$\bar{u}^2$	$\bar{v}^2$	$\bar{u} \cdot \bar{v}$
					C = A − B	σ_u^2	σ_v^2	C_{uv}

$$y' - (y_o + \bar{v}) = b_1\left[x - (x_o + \bar{u})\right]$$
$$x' - (x_o + \bar{u}) = b_2\left[y - (y_o + \bar{v})\right]$$

$$b_1 = \frac{C_{uv}}{\sigma_u^2} \qquad b_2 = \frac{C_{uv}}{\sigma_v^2}$$

Stopnje izračunavanja so naslednje:

1./ Izpišemo seriji dvojic osnovnih vrednosti x in y v stolpca (1) in (2).

2./ Izberemo primerni vrednosti x_o in y_o kot pogojni sredini.

3./ Od vrednosti x v stolpcu (1) odštejemo pogojno sredino x_o. Dobimo vrednosti novega znaka u. Te vrednosti vpišemo v stolpec (3). Enako od vrednosti y odštejemo y_o in diferenco vpišemo v stolpec (4).

4./ V stolpec (5) vpišemo kvadrate količin iz stolpca (3), v stolpec (6) kvadrate količin iz stolpca (4), v stolpec (7) pa produkte stolpca (3) in (4).

5./ Seštejemo vrednosti v stolpcih (3) do (7).

6./ Dobljene vsote delimo s številom enot. Na ta način dobimo v vrsti A količine:

$$\bar{u},\ \bar{v},\ \overline{u^2},\ \overline{v^2},\ \overline{uv}.$$

7./ V stolpcu (5) podpišemo pod vrednost $\overline{u^2}$ kvadrat od $\bar{u}^2$, v stolpcu (6) podpišemo vrednost $\bar{v}^2$, v stolpcu (7) pa $\bar{u} \cdot \bar{v}$.

8./ Ako te količine odštejemo od količin v predhodni vrsti, dobimo vrednosti σ_u^2, σ_v^2 in c_{uv}.

9./ Iz danih količin izračunamo regresijska koeficienta in enačbi regresijskih premic po danih obrazcih.

P Kot primer vzemimo isti problem, kot v predhodnem odstavku, in izračunajmo regresijski premici z uporabo pogojnih sredin.

Tab. 12.1o

i	x_i	y_i	$x_i-5o = u_i$	$y_i-16o=v_i$	u_i^2	v_i^2	u_iv_i
	1	2	3=1-5o	4 = 2-16o	5=3^2	6=4^2	7=3.4
1	39,5	156,5	- 1o,5	- 3,5	11o,25	12,25	+ 36,75
2	41	157	- 9	- 3	81,oo	9,oo	+ 27,oo
3	44,5	157,5	- 5,5	- 2,5	3o,25	6,25	+ 13,75
49	71,5	179	+ 21,5	+ 19	462,25	361,oo	+ 4o,85
5o	61,5	176	+ 11,5	+ 16	132,25	256,oo	+ 17,4o
5o			+ 141,5	+ 22o,5	433o,25	4679,25	391o
		A	+ 2,83	+ 4,41	86,6o5	93,585	78,2ooo
				B =	8,oo89	19,4481	12,4803
				A - B = C	78,5961	74,1369	65,7197

$\bar{x} = 50 + 2.83 = 52.83$ $\bar{y} = 160 + 4.41 = 164.41$

$b_1 = \frac{65.7197}{78.5961} = 0,8363$ $b_2 = \frac{65.7197}{74.1369} = 0,8865$

$y' - 164.41 = 0,8362\ (x-52,83)$ $x' - 52.83 = 0,8865\ (y - 164.41)$

126. DETERMINACIJSKI KOEFICIENT KOT MERILO JAKOSTI ODVISNOSTI MED POJAVI.

Ako pogledamo korelacijske grafikone, ki smo jih navedli v prejšnjih odstavkih, moremo na prvi pogled ugotoviti, da so nekateri pojavi med seboj bolj, nekateri pa manj povezani oz. odvisni. Jakost povezave je odvisna od tega, kako se posamezne točke grupirajo okrog regresijske črte v primeru linearne regresije okrog regresijske premice. Povezava med dvema pojavoma je na vsak način večja pri pojavih, pri katerih se posamezne vrednosti manj odkkanjajo od regresijske premice, kot pri pojavih, pri katerih so ti odkloni veliki. V skrajnem primeru, da bi vse točke ležale na sami regresijski premici bi bila povezava najbolj tesna. Zveza bi bila v tem primeru funkcionalna.

Ako pogledamo korelacijske grafikone naših primerov povezave, moremo že na oko ugotoviti, da je povezava med številom plugov in številom perutnine na loo kmečkih prebivalcev od vseh primerov najbolj šibka. Tesnejša povezava je že med težo in višino ljubljanskih dijakov. Vendar izgleda, da je odvisnost med številom plugov in številom bran v krajevnih odborih okraja Maribor - okolica in med natalitetо za leto 1946 in 1947 za 37 držav sveta še večja. Vidimo, da je možna medsebojna primerjava jakosti povezanosti pri zelo različnih pojavih. Pri pojavih, ki so med seboj bolj povezani, je regresijska krivulja, oziroma regresijska premica bolj jasno razvidna, ker se posamezne točke veliko bolj grupirajo okoli nje.

Ker so odkloni od regresijske premice rezultat individualnih vplivov je povezava med dvema pojavoma tem tesnejša, čim manjša je jakost individualnih vplivov.

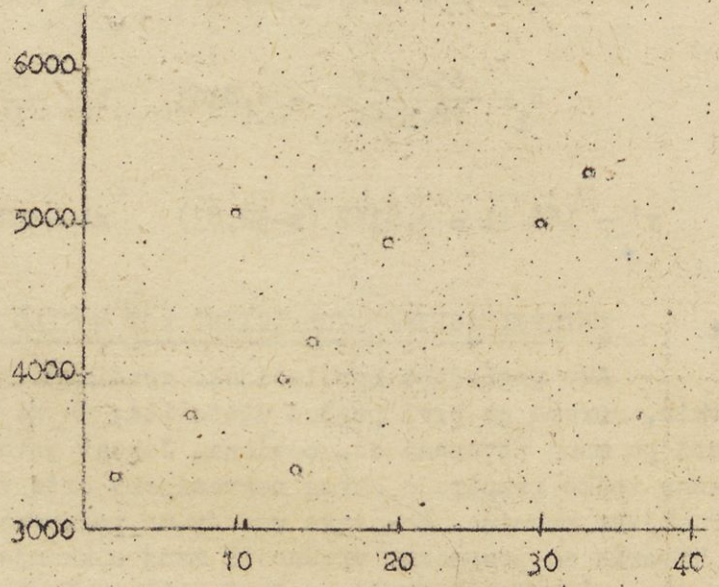
Podrobnejše obravnavajmo zgornje vprašanje na kratkem primeru.

Vzemimo devet nameščencev, in proučujmo njih plače in odvisnost plač od službenih let.

Tab. 12.11

i	Službena leta x_i	Mesečna plača y_i
1	2	3312
2	7	3726
3	13	3961
4	14	3890
5	19	4210
6	20	4870
7	30	5000
8	33	5308
9	10	5062

Korelacijski grafikon za ta primer je prikazan v sliki 12.15



Vplive, ki vplivajo na višino plače, moremo razdeliti v tri dele:

a) Splošni vplivi, ki so za vseh devet uslužbencev enaki n.pr. stroka, splošni nivo plač itd.

Kot vemo je rezultat splošnih vplivov dan s povprečno plačo ($\bar{y}$).

b) Službena leta (x), ki jih ne moremo smatrati kot splošen pogoj, ker se od primera do primera menjajo.
Ker regresijska premica $y' = \bar{y} + b_1 (x - \bar{x})$ podaja skupen rezultat vplivanja splošnih pogojev in službenih let, moremo vrednost $(y' - \bar{y})$ ali $b_1 (x - \bar{x})$ smatrati kot rezultat vplivanja službenih let na mesečno plačo.

c) Individualni vplivi, ki povzročajo, da se plače v posameznem primeru odklanjajo od plače, ki je dana s številom let in splošnimi pogoji. Individualni vplivi so n.pr. izredna funkcija, ki jo morda ima nek uslužbenec, ki odkloni plačo navzgor (n.pr. uslužbenec z redno številko 10) ali pridnost, ali malomarnost v službi itd. Ker je dejanska plača y rezultat vseh vplivov, y' pa rezultat splošnih pogojev in službenih let, moremo smatrati $y - y'$ kot rezultat individualnih vplivov.

V teh oznakah moremo razstaviti plačo na te tri dele:

$$y_i = \bar{y} + (y_i' - \bar{y}) + (y_i - y_i') \qquad (12.16)$$

Ako razrešimo v tej enačbi oklepaje, vidimo, da res dobimo na desni in na levi strani y_i.

Dejanska plača y_i je v tej enačbi razdeljena na tri dele:

$\bar{y}$ je rezultat splošnih pogojev (s) $(y' - \bar{y})$ je rezultat vpliva službenih let na plačo (x), $(y_i - y_i')$ pa rezultat ostalih individualnih vplivov (I).

Ako poiščemo regresijsko premico, moremo razdelitev posameznih plač na na tri dele dejansko izvršiti:

Najprej izračunajmo regresijsko premico:

Tab. 12.12

i	x	y	x = u	y-4000 = v	u^2	v^2	u.v
1	2	3312	2	- 688	4	473.344	- 1376
2	7	3726	7	- 274	49	75.076	- 1918
3	13	3961	13	- 39	169	1.521	- 507
4	14	3890	14	- 110	196	12.100	- 1540
5	15	4210	15	+ 210	225	44.100	+ 3150
6	20	4870	20	+ 870	400	756.900	+ 17400
7	30	5000	30	+ 1000	900	1000.000	+ 30000
8	33	5308	33	+ 1308	1089	1710.864	+ 43164
9	10	5062	10	+ 1062	100	1127.844	+ 10620
		$\sum$	144	3339	3132	5201.749	+ 98993
	$A = \frac{1}{n}\sum$		+ 16	+ 371	348	577972,1	10999,222
				B	256	137641	5936
				C = A-B	92	440331,1	5063,22
					σ_x^2	σ_y^2	c_{xy}

$$\bar{x} = 16 \qquad \bar{y} = 4000 + \bar{v} = 4371$$

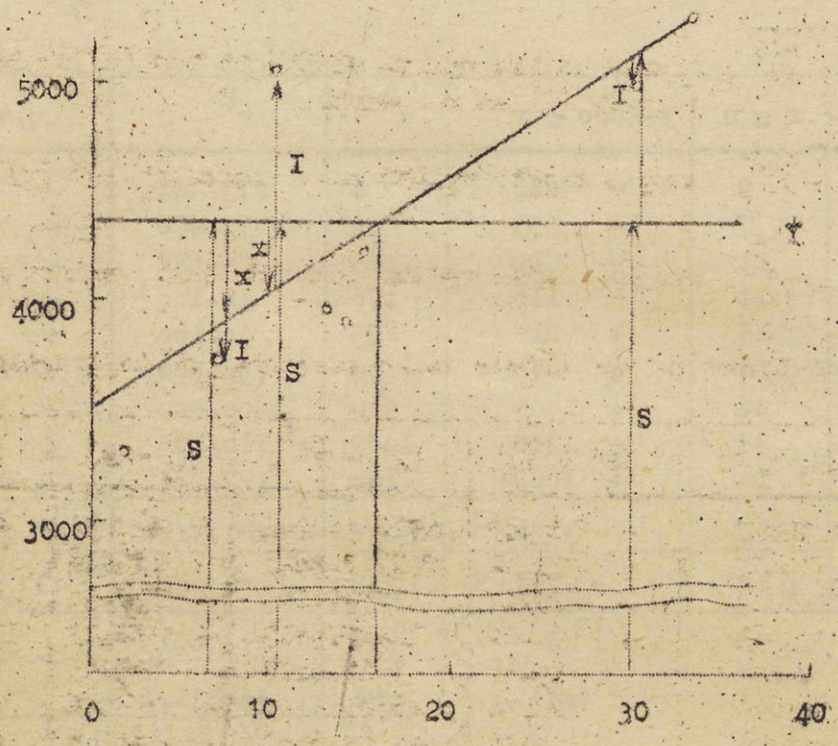
$$b_1 = \frac{5063,22}{92} = 55,035$$

$$y' = \bar{y} + b_1 (x - \bar{x})$$

$$y' = 4371 + 55,035\ (x - 16)$$

Ker imamo dano regresijsko premico, moremo izvesti razdelitev individualnih plač na rezultate posameznih vplivov tako grafično, kot računsko. Zaradi okrajšave zamenjujemo rezultat splošnih vplivov s S, rezultat službenih let z x, rezultat individualnih pogojev pa z I.

sl. 12.16 Razdelitev plač na sestavne dele



Na sliki so za tri uslužbence narisana delovanja posameznih vplivov. Računsko moremo podati razdelitev plač na dele takole:

Pab. 12.13

$S\,x$	=	S	+	XI	=	S	+	X	+	I
y_i	=	$\bar{y}$	+	$y_i - \bar{y}$	=	$\bar{y}$	+	$(y_i - \bar{y})$	+	$(y_i - \bar{y})$
3312	=	4371	+	(- 1059)	=	4371	+	(- 770,49)	+	(- 288,51)
3726	=	4371	+	(- 645)	=	4371	+	(- 495,32)	+	(- 149,68)
3961	=	4371	+	(- 410)	=	4371	+	(- 165,10)	+	(- 244,90)
3890	=	4371	+	(- 481)	=	4371	+	(- 110,07)	+	(- 370,93)
4210	=	4371	+	(- 161)	=	4371	+	(- 55,04)	+	(- 105,96)
4870	=	4371	+	(+ 499)	=	4371	+	(+ 220,14)	+	(+ 278,86)
5000	=	4371	+	(+ 629)	=	4371	+	(+ 770,49)	+	(- 141,49)
5308	=	4371	+	(+ 937)	=	4371	+	(+ 935,60)	+	(+ 1,40)
5062	=	4371	+	(+ 691)	=	4371	+	(- 330,21)	+	(+ 1021,21)
				0				0		0

Rezultat vpliva službenih let na plačo $y' - \bar{y}$ je enak $b_1 (x - 16)$ in ga moremo iz te zveze tudi izračunati $b_1 (x - 16) = 55{,}035 (x - 16)$.

Iz razdelitve se vidi, da individualne plače variirajo okrog aritmetične sredine $\bar{y} = 4371$. Celotna variacija $y_i - \bar{y}$ je rezultat vpliva let in individualnih vplivov. V zadnjih dveh stolpcih je celotna variacija razdeljena v dva dela: na rezultat vpliva službenih let $y' - \bar{y}$, ki je enak tudi $b_1 (x - \bar{x})$ in rezultat vpliva individualnih pogojev $(y_i - \bar{y})$.

Kot vidimo je: Vsota učinkov vsakega izmed treh, enaka 0. Skupna mera variacije vseh enot je varianca ali povprečen kvadratičen odklon. Zaradi tega je:

(12.17) $\sigma_y^2 = \overline{(y_i - \bar{y})^2}$ skupno merilo vpliva službenih let (x) in individualnih vplivov (I) za vse enote.

(12.18) $\sigma_o^2 = \overline{(y' - \bar{y})^2}$ skupno merilo vpliva samo službenih let (x) na vse enote.

(12.19) $\sigma_p^2 = \overline{(y_i - y')^2}$ skupno merilo vpliva individualnih vplivov na vse enote.

Ako izračunamo za naš primer vsako varianco posebej, dobimo:

Tab.12.14

$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y' - \bar{y})$	$(y' - \bar{y})^2$	$y_i - y'$	$(y_i - y')^2$
- 1059	1121481	- 770.49	593654.8401	- 288.51	83238.0201
- 645	416025	- 495.32	245341.9024	- 149.68	22404.1024
- 410	168100	- 165.10	27258.0100	- 244.90	59976.0100
- 481	231361	- 110.07	12115.4049	- 370.93	137589.0649
- 161	25921	- 55.04	3029.4016	- 105.96	11227.5216
+ 499	249001	+ 220.14	48461.6196	+ 278.86	77762.8996
+ 629	395641	+ 770.49	593654.8401	- 141.49	20019.4201
+ 937	877969	+ 935.60	875347.3600	+ 1.40	1.9600
+ 691	477481	- 330.21	109038.6441	+ 1021.21	1042869.8641
$\sum$ 0	3962980	0	2507902.0228	0	1455098.8628

$A = \frac{1}{9}\sum \sigma_y^2 =$ 440331 $\quad \sigma_o^2 =$ 278655.7803 $\quad \sigma_p^2 =$ 161676.5403

$A : \sigma_y^2$ 1 $\quad \frac{\sigma_o^2}{\sigma_y^2} =$ 0.6328 $\quad \frac{\sigma_p^2}{\sigma_y^2} =$ 0.3672

Vidimo, da je varianca σ_y^2, ki izraža skupno variacijo zaradi vpliva službenih let in individualnih vplivov na plačo, enaka vsoti variano σ_o in σ_p od katerih prva izraža variacijo samo zaradi službenih let, druga pa variacijo samo razadi individualnih vplivov.

$$44o331 = 278655.8 + 161676.5$$

V obliki enačbe moremo pisati:

$$\sigma_y^2 = \sigma_o^2 + \sigma_p^2 \qquad (12.2o)$$

Ako to enačbo delimo z σ_y^2 dobimo:

$$1 = \frac{\sigma_o^2}{\sigma_y^2} + \frac{\sigma_p^2}{\sigma_y^2} \qquad (12.21)$$

V našem primeru je:

$$\frac{44o331}{44o331} = \frac{278655.8}{44o331} + \frac{161676.5}{44o331} = o,6328 + o,3672$$

Kvocient $\frac{\sigma_o^2}{\sigma_y^2} = o,6328$ pove, kakšen del celotne variance plač izvira iz tega, da je mesečna plača odvisna od službenih let, kvocient $\frac{\sigma_p^2}{\sigma_y^2} = o,3672$ pa kakšen del celotne variance izvira iz drugih vzrokov.

Posebno važen je prvi del $\frac{\sigma_o^2}{\sigma_y^2}$, ki pove kakšen del variance izvira iz odvisnosti plače od službenih let. Običajno ga zaznamujemo z r^2 in ga imenujemo determinacijski koeficient. Iz količin σ_{xy}, σ_x in σ_y izračunamo determinacijski koeficient na enostaven način s obrazcem:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} \qquad (12.22)$$

V našem primeru plač moremo za kontrolo izračunati r^2 na oba načina:

iz zveze $r^2 = \frac{\sigma_o^2}{\sigma_y^2}$ in iz zveze $r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2}$

$$r^2 = \frac{278655.8}{44031} = 0,6323 \qquad r^2 = \frac{5063.222^2}{92 \cdot 44031} = \frac{25636217.0213}{40510452} = 0,6328$$

Vidimo, da dobimo isti rezultat.

Iz označbe $1 = \frac{\sigma_o^2}{\sigma_y^2} + \frac{\sigma_p^2}{\sigma_y^2}$ moremo zaključiti več stvari. Ker je vsak del zase

kvadrat, more $\frac{\sigma_o^2}{\sigma_p^2}$ zavzeti samo pozitivne vrednosti.

er je vsota dveh kvadratov enaka 1, $\frac{\sigma_o^2}{\sigma_y^2}$ ne more biti večji kot ena.

Možna vrednost determinacijskega koeficienta r^2 leži torej med 0 in 1. V simbolih mofemo pidati:

$$0 < r^2 < 1 \qquad (12.23)$$

Ako je $r^2 = \frac{\sigma_o^2}{\sigma_y^2} = 0$, sledi, da je $\frac{\sigma_p^2}{\sigma_y^2} = 1$.

V tem priemru vsa varianca izvira iz individualnih vplivov in nič iz vpliva x. Pojav y je v tem primeru od znaka, oz. pojava neodvisen.

Ako je $r^2 = \frac{\sigma_o^2}{\sigma_y^2} = 1$, sledi, da je $\frac{\sigma_p^2}{\sigma_y^2} = 0$. V tem primeru vsa varianca

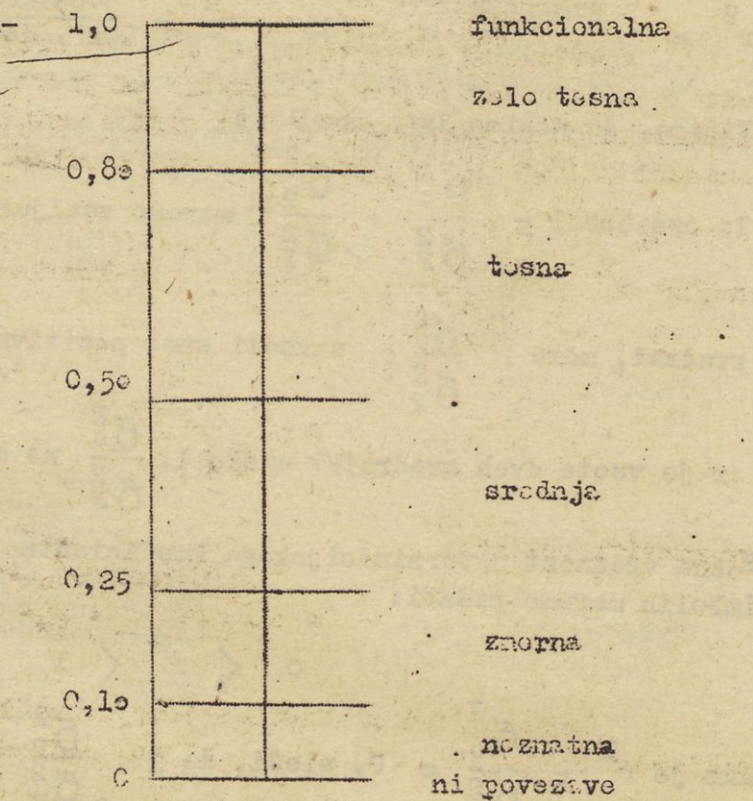
izvira iz odvisnoti y in x in je učinek individualnih vplivov enak 0. Grafičmo leže v tem primeru vse točke na regresijski premici. Ako je vrednost determinacijskega koeficienta majhna, je $\frac{\sigma_p^2}{\sigma_y^2}$ velik, rezultat individualnih vplivov torej

velik, povezanost majhna. Obratno, ako je vrednost r^2 blizu 1, pomeni, da velik del variacije izvira iz odvisnosti y-a od x, povezanost je torej velika.

r^2 more torej dobro služiti kot koeficient, ki pokaže jakost, oz. stopnjo povezave med dvema pojavoma. Na grafičnem prikazu je dana skala vrednosti r^2 in opisana jakost povezave, ki jo izraža pripadajoča vrednost r^2 (po Chaddocku).

Ako smo v nekem primeru izračunali, da je $r^2 = 0,6$ moremo zaključiti, da 6o°/° celotne variacije izvira iz znaka od katerega je pojav odvisen. Ako pogledamo grafično preglednico, vidimo, da gre v tem primeru za tesno povezavo.

Sl.12.17 Stopnja povezave pri različnih vrednostih determinacijskega koeficienta



D Dokaz, da je $r^2 = \frac{\sigma_o^2}{\sigma_y^2} = \frac{c_{xy}^2}{\sigma_x^2 \sigma_y^2}$ je razmeroma enostaven, ako poznamo lastnosti regresijske premice.

ker je individualen učinek znaka x na y $(y' - \bar{y})$ enak $b_1 \cdot (x - \bar{x})$, je varianca tega učinka

$$\sigma_o^2 = \overline{(y - \bar{y})^2} = \overline{b_1^2 (x - \bar{x})^2} = b_1^2 \cdot \sigma_x^2$$

Ker je $b_1 = \frac{c_{xy}}{\sigma_x^2}$ je $\sigma_o^2 = b_1^2 \cdot \sigma_x^2 = \frac{c_{xy}^2}{\sigma_x^2}$

Ker je $\sigma_o^2 = \frac{c_{xy}^2}{\sigma_x^2}$ je $r^2 = \frac{\sigma_o^2}{\sigma_y^2} = \frac{c_{xy}^2}{\sigma_x^2 \sigma_y^2}$ S tem je obrazec dokazan.

127. KORELACIJSKI KOEFICIENT

Koeficient determinacije r^2 ima svoj logični smisel in pomeni oni del variance proučevanega pojava, ki izvira iz znaka, s katerim je pojav v zvezi. Vendar se kljub temu običajno uporablja kot merilo povezave kvadratni koren iz koeficienta determinacije $\sqrt{r^2} = r$, ki se imenuje korelacijski koeficient.

Iz enačbe $r^2 = \frac{c_{xy}^2}{\sigma_x^2 \sigma_y^2}$ sledi, da je korelacijski koeficient

$$r = \frac{c_{xy}}{\sigma_x \sigma_y} \qquad (12.24)$$

ali v razviti obliki:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} \qquad (12.15)$$

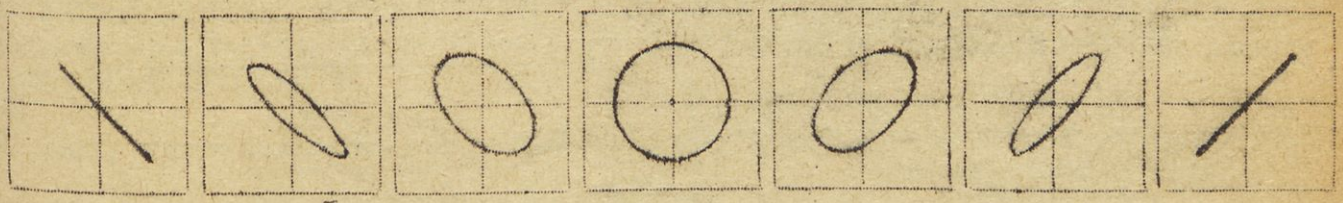
Korelacijski koeficient torej izračunamo na ta način, da kovarianco delimo s standardnima deviacijama znakov, ki sta med seboj povezana.

Korelacijski koeficient nima takega določenega pomena kot koeficient determinacije, vendar ima to dobro lastnost, da poda poleg jakosti povezave tudi smer odvisnosti med dvema pojavoma. Ker je območje r^2 v intervalu od 0 do 1, je območje korelacijskega koeficienta r zaradi tega, ker je kvadratni koren in r^2, od -1 do + 1.

Ker so standardne deviacije pozitivne količine, se predznak korelacijskega koeficienta ravna po predznaku kovariance c_{xy}.

Iz zveze $c_{xy} = \frac{1}{n}\sum (x - \bar{x})(y - \bar{y})$ moremo iz korelacisjkega grafikona hitro zaključiti,, ali je r pozitiven ali negativen in, ali je majhen ali velik. V grafičnem prikazu, v katerem so z elipsami naznačena območja, v katerih so nahajajo točke korelacijskega grafikona, je jasno razvidno, kakšni so korelacijski koeficienti pri posameznih oblikah.

sl.12.18 Vrednosti korelacijskega koeficienta, pri različnih vrstah korelacijskega grafikona



$r = -1$ $r \ll 0$ $r < 0$ $r = 0$ $r > 0$ $r \gg 0$ $r = +1$

Iz naših primerov, za katere imamo že izračunane c_{xy}, σ_x in σ_y, moremo korelacijski koeficient izračunati iz zveze

$$r = \frac{c_{xy}}{\sigma_x \sigma_y}, \text{ ali iz determinacijskega koeficienta } r^2.$$

127.1 Območje $\bar{y} + b_1 (x - \bar{x}) + \sigma_y \sqrt{1 - r^2}$. Ker je σ_p standardna deviacija odklonov od regresijske premice, moremo okrog regresijske premice določiti območje, v katerem se nahaja 68% vseh primerov.

To območje predstavlja pas med premico

$$y'_s = \bar{y} + b_1 (x - \bar{x}) - \sigma_p \text{ in} \qquad (12.26)$$

premico

$$y'_z = \bar{y} + b_1 (x - \bar{x}) + \sigma_p \qquad (12.27)$$

Standardno deviacijo odklonov od regresijske premice moremo določiti z izračunavanjem iz odklonov, ali pa iz enačbe:

$$1 = r^2 + \frac{\sigma_p^2}{\sigma_y^2} \qquad (12.27)$$

Iz te enačbe sledi dalje:

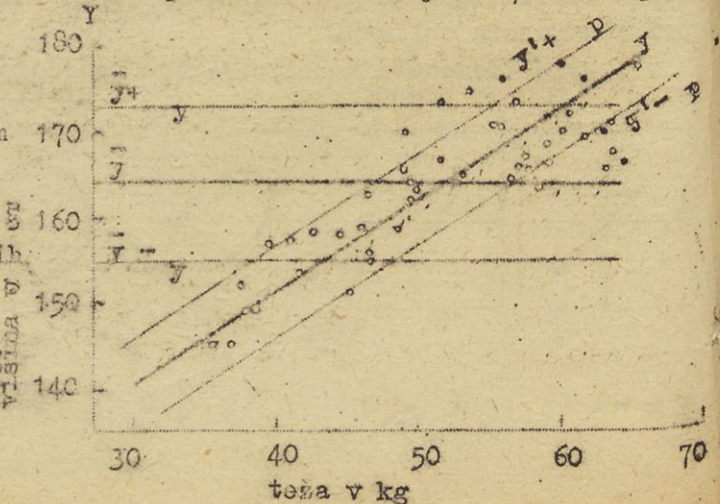
$$\sigma_p = \sigma_y \sqrt{1 - r^2} \qquad (12.28)$$

Standardno deviacijo odklonov od regresijske premice torej dobimo, ako standardno deviacijo odklonov od aritmetične sredine pomnožimo s koefeicientom $\sqrt{1 - r^2}$.

Ako je korelacijski koeficient blizu +1 ali -1, je $\sqrt{1 - r^2}$ in zaradi njega tudi σ_p majhna količina. Območje v katerem se nahaja v tem primeru 68% vseh enot je torej ozko.

Za primer korelacije med težo in višino 50 študentov ljubljanskih gimnazij je slika naslednja:

sl.12.19 Korelacijski grafikon med težo in višino 50 dijakov ljubljanskih gimnazij z vrisanim pasom v katerem je 68% vseh primerov



V sliki je označeno območje 68 °/° primerov okrog aritmetične sredine $\bar{y}$ in okrog regresijske premice y'. Vidimo, da je pas okrog regresijske premice ožji iz česar sledi, da regresijska premica bolje podaja vrednosti y kot aritmetična sredina.

128.X KRIVULJČNA, MULTIPLA ALI VEČKRATNA IN PARCIELNA ALI DELNA KORELACIJA

128.1 Krivuljčna korelacija

Dosedaj smo obravnavali primere, v katerih je regresijska črta najenostavnejša krivulja - premica. Kot pa smo videli na primeru korelacije med deležem obdelovalne površine in številom ovc na loo kmečkih gospodarstev za Kra-je okraja Maribor - okolica, more biti regresijska črta tudi krivulja in ne premica. V tem primeru govorimo o krivuljah regresije. Definicijo, da je determinacijski koeficient pri linearni korelaciji enak $r^2 = 1 - \frac{\sigma_p^2}{\sigma_y^2}$ moremo uporabiti tudi za krivuljčno korelacijo.

Determinacijski koeficient krivuljčne korelacije, ki ga zaznamujemo z R^2, je enak podobnemu izrazu

$$R^2 = 1 - \frac{\sigma_k^2}{\sigma_y^2} \qquad (12.29)$$

kjer pomeni σ_k^2 varianco odklonov od regresijske krivulje. R^2 podobno kot korelacijski koeficient r^2 pove, kakšen del celotne variance y-a je pojasnjen z regresijsko krivuljo.

128.2 Multipla ali večkratna korelacija

Masovni pojavi so običajno odvisni od večjih vplivov istočasno. Težadijaka je odvisna od višine, starosti, socialnega stanja itd., donos od temperature in padavin v času vegetacije. Mesečna proizvodnja podjetja je odvisna od števila delavstva, števila delovnih dni v mesecu, stopnje mehanizacije itd. Kot smo v primeru odvisnosti od enega pojava iskali regresijsko črto : v posebnem regresijsko premico, v primerih kadar gre za odvisnost od več pojavov istočasno, regresijske plokve oz. regresijske ravnine. Jakost povezave merimo zopet tako, da iščemo, kakšen del celotne variance izvira iz vplivov, s katerimi je pojav, ki ga proučujemo, v korelaciji.

128.3 Parcielna ali delna korelacija

Ako proučujemo n.pr. korelacijo med težo in višino, vidimo, da je korelacija med težo in višino zelo tesna. Pri podrobnejšem proučevanju pa vidimo, da

v veliki meri ta korelacija izvira iz tega, ker sta teža in višina odvisna od starosti. Majši dijaki so manjši in lažji, starejši dijaki so večji in težji. Ako hočemo analizirati samo odvisnost med težo in višino, moramo torej vpliv starosti odstraniti. Da to dosežemo, izračunavamo parcielne oz. delne korelacijske koeficiente, ki predstavljajo korelacijo, iz katere so odstranjeni vplivi drugih pojavov.

128.4 Uporaba korelacije v vzročni analizi

Kot smo že uvodoma omenili, je povezanost pojavov v nekaterih primerih vzročna. Pri povezavah take vrste moremo iz vsebine pojavov razbrati, kateri pojav je vzrok, kateri posledica. Padavine in temperatura v času vegetacije sta vzroka, ki vplivata na donos. Tudi proizvodnja in število izvršenih delovnih ur sta v vzročni povezavi in sicer je število izvršenih delovnih ur vzrok, vrednost proizvodnje pa posledica. Korelacija v teh primerih zelo dobro služi za analizo vzročne povezave med pojavi in se pri vzročni analizi vedno uporablja.

129. POVZETEK

Masovni pojavi so med seboj povezani. Vendar ta povezava ni funkcionalna temveč korelacijska. Pri korelacijski povezavi se povezanost ne očituje individualno, temveč kot težnja v primeru velikega števila primerov. Podatke pojavov, ki so med seboj v odvisnosti, prikazujemo ali kot sistem dvojic vrednosti v obliki korelacijskega grafikona ali v obliki korelacijske karte. Korelacijski grafikoni predstavljajo v pravokotnem koordinatnem sistemu narisane točke, od katerih vsaka predstavlja vrednost enem enote. Korelacijska karta pa je frekvenčna distribucija kombinacije dveh znakov, dana v kombinacijski tabeli.

Vrednost znaka individualne enote so rezultat treh vrst vplivov:

a) splošnih, b) znaka s katerim je v korelaciji, c) individualnih vplivov. Rezultati splošnih vplivov in znaka, s katerim je pojav povezan so dani z regresijsko črto. To moremo določiti na več načinov.

1./ Prostoročno kot krivuljo med dejanskimi točkami v korelacijskem grafikonu.
2./ Kot črto aritmetičnih sredin.
3./ Kot premico ali krivuljo, izračunano kot črto, ki se vsem točkam najbolj prilega.

Regresijski premici sta dve. Ena kaže odvisnost y-ó od x-a, druga pa odvisnost x-a znaka y. Enačbi regresijskih premic sta naslednji:

$$y' - \bar{y} = b_1 (x - \bar{x}) \quad \text{in} \quad (x' - \bar{x}) = b_2 (y - \bar{y}) \quad \text{kjer sta } b_1 \text{ in } b_2$$

smerna koeficienta premice in ju imenujemo regresijska koeficienta.

Izračunana sta po obrazcu:

$$b_1 = \frac{C_{xy}}{\sigma_x^2} \qquad b_2 = \frac{C_{xy}}{\sigma_y^2}$$

Količino C_{xy}, ki jo imenujemo kovarianco, izračunamo po obrazcu

$$C_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

Kot merilo povezanosti služi koeficient determinacije

$$r^2 = \frac{C_{xy}}{\sigma_x^2 \sigma_y^2}$$

ki pove kakšen del celotne variacije izvira iz odvisnosti y-a od x-a. Vrednosti koeficienta determinacije gredo od o do 1. Nič je v primeru, da povezave ni, ena pa v primeru funkcionalne povezave.

Običajno se uporablja kot merilo povezanosti korelacijski koeficient r, ki je kvadratni koren iz koeficienta determinacije. Izračunamo ga po obrazcu

$$r = \frac{C_{xy}}{\sigma_x \sigma_y} .$$

Korelacijski koeficient pokaže poleg jakosti povezave tudi smer povezave. Njegove vrednosti gredo od -1 do +1.

Standardno deviacijo okrog regresijske premice σ_p moremo izračunati po obrazcu $\sigma_p = \sigma_y \sqrt{1 - r^2}$. S pomočjo σ_p moremo okrog regresijske premice določiti območje, v katerem se nahajajo točke 68% vseh enot.

O krivuljčni korelaciji govorimo takrat kadar regresijska črta ni premica temveč krivulja.

Kadar iščemo odvisnot nekega pojava od več drugih pojavov govorimo o multipli, kadar pa pri proučevanju korelacijske povezave med dvema pojavoma odstranim vpliv tretjega pojava, pa o parcialni korelaciji.

13 METODA VZORČENJA

130. SPLOŠNO O VZORČENJU

130.0 Metode delnega opazovanja

Že pri oblikah statističnega opazovanja smo omenili, da uporabljamo poleg popolnega opazovanja tudi delna opazovanja. Popolno opazovanje izvajamo takrat, kadar popišemo v celoti vse enote statistične mase. Podatki, ki jih dobimo na ta način, so točna slika celote, v kolikor ni bilo v samem zbranem gradivu napak.

V nekaterih primerih pa je popolno opazovanje iz različnih vzrokov nemogoče izvršiti. Vzemimo kot primer, da nimamo dovolj finančnih sredstev, ali da ne razpolagamo z dovoljnim številom ljudi, ki bi izvedli popis. Dostikrat je tudi rok, v katerem moramo priti do podatkov, prekratek, da bi izvedli popis. Imamo tudi primere, da je preiskava nekega masovnega pojava važna, ne pa toliko, da bi zato trošili velike vsote denarja, kolikor stane popis. V področjih znanstvenega raziskavanja pa je kompletno opazovanje iz razlogov, ki jih bomo navedli kasneje, povsem nemogoče izvajati.

Zaradi tega se je v statistiki že zgodaj pojavilo vprašanje, ali je možno že na osnovi pregleda samo enega dela statistične mase napraviti kake zaključke o celotni masi. Začeli so izvajati ankete, monografije, znana je metoda izbora tipičnih enot. Na osnovi teh metod pridemo do podatkov, ki do neke mere morejo dati koristne zaključke o celotni statistični masi. Vendar imajo vse te metode dve bistveni slabosti, zaradi katerih je njih važnost za statistiko omejena. Prva izmed slabosti je ta, da moramo statistično maso precej dobro poznati, ako hočemo izbrati res tipične enote. To pa v večini primerov ni, ker običajno preiskujemo pojave, ki jih ne poznamo. Drugič pa je odločitev, kaj je tipično, zelo subjektivna. Kar nekdo smatra za tipično, se mogoče drugemu ne zdi, in obratno.

Zaradi tega si je osvojila mesto edino znanstvene in subjektivne metode delnega opazovanja, metoda slučajnega izbora. Že naziv "slučajni izbor" pove, da enote, ki jih opazujemo, niso izbrane "po prevdarku" kot tipične, temveč na slučajen način, "na slepo".

V preprosti, vendar v osnovi odgovarjajoči obliki, način slučajnega izbora vršimo vsakodnevno v življenju. Profesor si ustvari sliko o dijakovem znanju na ta način, da mu postavi par vprašanj. Ta vprašanja vleče dijak z listki, da se ohrani objektivnost in nepristranost in da se ne da namerno nekemu dijaku pretežka, drugemu prelahka vprašanja. Zdravnik, ki hoče preiskati kri bolnika, mu ne odvzame vse krvi, ampak si napravi sodbo o krvi na osnovi preiskave samo ene kapljice krvi. Da gospodinja preiskusi

kakovost jedi, ne poje vse jedi, temveč jo samo pokusi. Trgovec, ki kupuje v velikih količinah n.pr. jajca, se o kakovosti jajc ne prepriča na ta način, da pobije vse kupljeno blago, temveč preizkusi iz vsakega zaboja samo nekaj jajc. Ta jajca pa ne izbira, temveč jih vzame "na slepo". V tovarni električnih žarnic ne ugotavljajo trajanje žarnic tako, da puste izgoreti celotno proizvodnjo žarnic, temveč preizkusijo samo nekaj žarnic. Ako bi pustili izgoreti celotno proizvodnjo, bi sicer dobili popolnoma pravilno povprečje, pri tem pa uničili celotno proizvodnjo.

Podobnih primerov bi mogli našteti še več. Jasno je, da se znanstvena uporaba metode vzorčenja po svoji izdelanosti razlikuje od teh preprostih primerov. Vendar morajo biti v obeh primerih enote izbrane na slepo, ako hočemo, da bodo storjeni zaključki pravilni.

130.1 Področja uporabe metode vzorčenja v praksi.

Metodo vzorčenja moremo uporabljati na vseh področjih, kjer nastopajo masovni pojavi. Njegova uporaba ni omejena, temveč ga moremo uporabljati tudi na področjih, kjer popolno opazovanje po naravi predmeta ni možno. Tak primer imamo v znanostih kot je biologija, antropometrija, medicina, agronomija, meteorologija itd. Raziskave na teh področjih so možne edino z uporabo metod, ki so osnovane na vzorčenju. Tu predstavlja določeno število opazovanih enot ali izvršenih poizkusov vedno majhen izsek iz neomejene množice možnih poizkusov, izvršenimi pod enakimi pogoji. Seveda je ta neomejena množica samo umišljena in v resnici ne obstoja. Zanimivo je, da se je metoda vzorčenja začela uporabljati najprej ravno pri raziskovalnem delu v teh znanostih in ne v statistiki socialno-ekonomskih pojavov. To je do neke mere razumljivo. Potreba po teh metodah je bila veliko bolj pereča v raziskovalnem delu, ker na drug način opazovanja niso mogli izvesti, kot v statistiki v pravem smislu, kjer poznamo kompletno, opazovanje. Metoda vzorčenja si je šele v novejšem času priborila mesto tudi na področjih izven strogo raziskovalnega dela. Danes se v vseh naprednih državah uporablja kot splošno priznana metoda statističnega opazovanja na vseh področjih. Prav posebno je koristna na področjih, kjer je bodisi zaradi velikega štev-ila enot, zapletenega ali zamudnega registriranja popolno opazovanje zelo drago in zamudno.

Eno izmed takih področij je kmetijska statistika. Tu uporabljamo vzorčenje pri registraciji površin, živine itd. Posebno področje je ugotavljanje donosov. Za ugotavljanje donosov je po metodi vzorčenja izdelana posebna metoda metraže. Z metražo na osnovi meritev na parcelah z razmeroma majhne površine ugotavljamo povprečen donos na celi njivi, v celem okraju ali republiki.

Važna veja, kjer uporabljamo metodo vzorčenja je kontrola kvalitete proizvodnje. Enakost artiklov, ki jih izdeljuje nek stroj, ni absolutna. Izmere posameznih artiklov se zaradi najraznovrstnejših vzrokov malenkostno

razlikujejo tudi če stroj dela brezhibno. V primeru pa, da je delovanje stroja napačno, so razlike večje, kot nastopajo zaradi slučajnih razlik. Na osnovi vzorčenja izdelana metoda omogoča, da to okvaro pravočasno ugotovimo.

Pri statističnih popisih uporabljamo metodo vzorčenja po namenu in odnosu na več načinov:

1. V izvedbi popisa more vzorčenje služiti:

a) kot nadomestilo za popolen popis. Za to se odločimo iz različnih vzrokov. Večkrat je vzrok pomanjkanje finančnih sredstev, popisnih organov ali kratek rok, v katerem moramo priti do podatkov. Za vzorec se odločimo tudi v primeru, kadar more dati že vzorec zadostne podatke.

b) kot dopolnilo k popisu. Dostikrat pri popisu preveliko število vprašanj zavira potek popisa. Kljub temu pa so vsi ti podatki za proučevanje važni. V tem primeru iz vseh vprašanj izberemo one, ki so najbolj važni in za katere moramo dobiti podrobne podatke. Odgovore na ta vprašanja iščemo od vseh enot. Enotam, ki smo jih zbrali na slučajen način, pa dodamo še ostala vprašanja. Medtem ko prvi del vprašanj obdelamo v celoti, dodatna vprašanja obdelamo po vzorčni metodi. Tako moremo pri popisu prebivalstva spol, starost ,zaposlitev itd. popisati s popisom, z vzorcem pa ostale manj važne znake, kot znanje tujih jezikov, defektnost itd.

2. Pri obdelavi podatkov moremo uporabiti vzorčenje na dva načina:

a) kot predhodne rezultate. Popolna obdelava vsakega popisa traja običajno več mesecev ali tudi več let. Posledica tega je, da so podatki dani na razpolago zelo kasno po izvedbi,, ko je stanje v velikih primerih že zelo izpremenjeno. Zaradi tega je zelo koristno, da v takih primerih glavne podatke obdelamo z vzorcem. Ti podatki, ki niso podrobni, pač pa so na razpolago kratek čas po popisu, morejo zelo dobro služiti kot predhodni rezultati. Celotna obdelava, ki se vrši dalj časa, da potrebne končne rezultate. Na tak način smo v LRS dobili v letu 1948 z dvoodstotnim vzorcem predhodne rezultate popisa prebivalstva. Celotna obdelava se je končala šele v letu 1950, medtem ko smo z vzorcem prišli do glavnih podatkov že par mesecev po popisu.

b) kot dopolnilno obdelavo. Dostikrat zaradi katerega koli vzroka ne moremo popolno obdelati vse podatke in možne kombinacije statističnega gradiva. Običajno je vzrok pomanjkanje denarja, ljudi ali časa. V tem primeru važnejše podatke in kombinacije obdelamo popolno, ostale, katere ne potrebujemo preveč podrobno, pa z vzorcem.

3. Vzorčenje moremo uporabiti tudi kot kontrolo izvedbe popisa na terenu ali obdelave podatkov.

a) Kontrola popisa na terenu je sestavni del izvedbe vsakega popisa. Vendar moremo z običajnimi metodami kontrole prekontrolirati le polnoštevilnost obrazcev in odgovorov in grobe napake, ki jih dobimo iz nelogičnosti odgovorov. Z vzorčenjem izvršimo kontrolo tako, da po izvršenem popolnem popisu skrbno ponovno popišemo določeno število slučajno izbranih

enot in podatke za te enote primerjamo s podatki, ki smo jih dobili pri popolnem popisu. Iz tega vzorca moremo sklepati na kakovost celotnega popisa in ugotoviti, za koliko se dejansko stanje razlikuje od rezultatov, ki smo jih dobili s popisom. Na ta način izvajamo v FLRJ vsako leto kontrolo kvalitete popisa živine. Takoj po izvršenem celotnem popisu kontrolorji ponovno zelo skrbno popišejo izbrana gospodarstva. Iz tega vzorca izračunamo za vsako vrsto živine posebej razliko od dejanskega stanja.

b) Kontrolo kvalitete obdelave moremo z vzorcem izvesti v vseh fazah obdelave. Tako moremo z vzorcem izvesti kontrolo šifriranja, sortiranja, tabeliranja pri ročni obdelavi in kontrolo šifriranja in luknjanja kartic pri obdelavi na statističnih strojih.

130.2 Prednosti in pomanjkljivosti vzorčenja napram popolnemu popisu in drugim delnim opazovanjem.

Vzorčenje ima na vsak način prednost pred vsemi drugimi metodami delnega opazovanja. Ima pa tudi določene prednosti pred celotnim popisom, zaradi katerih se dostikrat odločimo za vzorec in ne za popis.

Najvažnejše prednosti vzorčenja, katere moramo upoštevati, kadar odločamo o tem, na kakšen način bomo izvedli statistično opazovanje, so naslednje:

1. Število enot, ki jih opazujemo pri vzorčenju je znatno manjše kot število enot celotnega popisa. V večini primerov vsebuje vzorec le nekaj odstotkov enot celotne statistične mase (n.pr. 2 o/o vzorec popisa prebivalstva leta 1948).

Že iz samega tega dejstva slede naslednje prednosti:

2. Število potrebnih ljudi za opazovanje je zelo omejeno. S tem je možna večja izbira, boljša izvežbanost in kontrola izvedbe opazovanja. Zaradi tega se dvigne kakovost podatkov.

3. Čas, v katerem pridemo do rezultatov z vzorčenjem je znatno krajši, kot pri popolnem popisu. Vrednost rezultatov je zaradi tega znatno večja.

4. Stroški vzorčenja so iz zgornjih razlogov veliko manjši kot stroški za popolen popis.

5. Zaradi manjšega števila enot, izbere dobrih terenskih delavcev in boljše kontrole, moremo pri vzorčenju število napak, ki izvirajo iz površnosti, nepoučenosti in malomarnosti, skrčiti na najmanjšo mero. Pogreška, ki izvira iz tega, da smo z vzorcem opazovali samo del celotne mase, pa se da znanstveno objektivno določiti. Velikost te pogreške je možno z velikostjo vzorca regulirati.

6. Kot smo že omenili, je vzorčenje v raziskovalnem delu edino možen način opazovanja.

Poleg teh dobrih lastnosti pa ima vzorčenje tudi pomanjkljivosti, ki omejujejo njegovo pomembnost in v marsikaterem primeru onemogočajo njegovo uporabo.

Ena izmed glavnih pomanjkljivosti je, da je možno vzorčenje s pridom uporabiti le za one množične pojave, ki imajo veliko število enot. Zaradi tega z vzorčenjem ne moremo dobiti uporabnih podatkov za manjše administrativne dele, kot, n.pr. [illegible] ali okraje, razen, ako vzamemo v vzorec velik del vseh enot. Vzorčenje moremo uporabiti torej največkrat v primeru, ko hočemo dobiti rezultate za vso republiko ali državo. Seveda imajo taki podatki omejen pomen.

Druga pomanjkljivost, preko katere ne moremo iti, je, da ne moremo z istim vzorcem zajeti velikega števila statističnih znakov, ker je zanesljivost posameznih podatkov v takem primeru zelo različna.

Iz tega moremo napraviti naslednji zaključek: Vzorčenju bomo dali prednost pred popisom v primerih, kadar bomo proučevali majhno število podatkov in potrebovali podatke samo za večje administrativne enote. Na vsak način pa je potrebno izvesti celoten popis, kadar opazujemo veliko število znakov in potrebujemo podatke za najmanjše administrativne enote. Metražo bomo n.pr. uporabili za ugotavljanje donosa, kadar bomo skušali dobiti povprečje za celoten rajon ali republiko. V primeru, da potrebujemo podatke o donosu za posamezne KLO-je ali celo gospodarstva, pa ta metoda ni uporabna.

131. OSNOVNA STATISTIČNA MASA, VZOREC, MASA VSEH VZORCEV

131.1 Osnovna statistična masa.

Statistično maso, o kateri hočemo dobiti rezultate in zanjo napraviti na osnovi vzorčenja določene zaključke, imenujemo osnovno ali celotno statistično maso. Vsa kmetijska gospodarstva v LRS, vse šole v FLRJ so celotne ali osnovne statistične mase. Osnovne statistične mase vsebujejo vse enote na določenem področju. Število enot osnovne statistične mase je običajno veliko. Za statistične mase z majhnim številom enot se namreč ne izplača izvajati vzorca, temveč izvedemo rajši popolen popis.

Vendar ločimo osnovne statistične mase po številu enot na take, ki imajo končno število enot in take, ki imajo neskončno število enot. Statistične mase s končnim številom enot, imenovane tudi končne osnovne mase so n.pr. kmečko gospodarstvo v LRS, industrijski delavci v FLRJ, obrtna podjetja v FLRJ itd.

Statistične mase z neskončnim številom enot ali neskončne osnovne mase nastopajo n.pr. pri zveznih statističnih masah. Zvezna statistična masa je ona, ki ni sestavljena iz ločenih enot, temveč je snovna in so enote

določene šele s predpisom (parcele zemljiška posest).Ako vzamemo kot primer njivo v izmeri 100 m x 100 m, moremo iz nje napraviti končno ali pa tudi neskončno statistično maso parcelic po 1 m2. Ako vzamemo okvir, ki je en meter dolg in en meter širok,vidimo, da moremo ta okvir položiti v njivi, ki meri 1 ha na več kot pa 10000 različnih načinov. Pri natančnejši preiskavi vidimo, da moremo ta okvir položiti na neskončno število različnih načinov. Na ta način smo dobili neskončno statistično maso.

Neskončna osnovna masa je tudi hipotetična masa.Ako zdravnik preizkusi neko novo zdravilo na 100 pacientih, teh 100 pacientov ne tvori osnovne mase, čeprav so to vsi primeri dejanskega zdravljenja s tem zdravilom. Teh 100 primerov smatramo namreč kot vzorec iz namišljene hipotetične osnovne mase, ki sestoji iz neskončnega števila pacientov, zdravljenih z istim zdravilom, pod enakimi pogoji. S hipotetičnimi statističnimi masami imamo opravka v vseh področjih, kjer s statističnimi metodami vršimo znanstveno raziskovalno-delo, pa naj bo to agronomija, medicina,biologija in podobno.

131.2 Vzorec.

Enote, ki smo jih na slučajen način izbrali iz osnovne statistične mase, so delna statistična masa osnovne mase, torej zase tvorijo tudi statistično maso. Opredeljujočim pogojem osnovne mase se pridruži še nov pogoj, da so bile izbrane pri slučajnem izboru. Delno statistično maso, ki smo jo iz osnovne mase dobili na ta način, da smo zbrali določeno število enot na slučajen način zaradi tega, da bi iz nje sklepali na razmere v celotni masi, imenujemo vzorčna masa ali na kratko vzorec. Ako iz 240000 kmetijskih gospodarstev na slučajen način izberemo 24000 gospodarstev, da bi na osnovi teh 24000 gospodarstev sklepali na celoto, je to 10 o/o vzorec.

.21 Po številu enot, ki jih izberemo v vzorec, delimo vzorce na male in velike vzorce. Razmejitve, do katerega števila enot smatramo vzorec kot majhnega in nad katerim kot velikega, ni. Vendar se izkaže, da imamo pri znanstvenih poizkusih običajno opravka z vzorci, pri katerih je število enot pod sto. Te vzorce štejemo kot male. Število poizkusnih parcelic pri poljskih poizkusih ali mišk pri poizkusih v medicini znaša vsega par deset parcelic ali mišk.

Pri uporabi vzorčenja pri statističnih popisih pa je število enot v vzorcu več sto ali tudi več tisoč enot. Take vzorce imenujemo velike vzorce. Pri vzorcu popisa prebivalstva leta 1948 je bilo vzeto v vzorec 7351 gospodinjstev in 27847 oseb. Iz tega moremo napraviti zaključek, da imamo opravka z malimi vzorci predvsem v poizkusnem delu, medtem ko v statistiki v ožjem smislu uporabljamo velike vzorce. Delitev na male in velike vzorce je potrebna zaradi tega, ker so zakonitosti vzorčenja pri velikih vzorcih znatno enostavnejše kot pri malih vzorcih.

.22 Iz končne osnovne mase moremo izbrati vzorec na dva načina: brez ponavljanja in s ponavljanjem.

Za neko statistično maso imamo izpolnjene popisnice. Iz tega gradiva moremo izvleči 100 enot, oziroma popisnic na dva načina. Ko izvlečemo popisnico in jo registriramo, jo moremo vložiti v popisno gradivo nazaj, predno izvlečemo naslednjo popisnico. Pri takem postopku je možno, da ena in ista popisnica nastopa v vzorcu večkrat. Ker se mora ena in ista enota pri takem načinu izvlačenja večkrat ponoviti, imenujemo tak vzorec "vzorec s ponavljanjem."

Ako izvlečene popisnice ne vlagamo sproti nazaj v popisno gradivo, temveč jih polagamo na stran, popisnica, ki je bila enkrat izvlečena, pri naslednjih izvlačenjih ne pride v poštev in ne more biti ponovno izvlečena. Pri takem načinu izvlačenja more posamezna enota v vzorcu nastopiti samo enkrat. " Zaradi tega imenujemo take vrste vzorec "vzorec brez ponavljanja".

Zakonitosti za vzorec s ponavljanjem so sicer enostavnejše, vendar pri popisih običajno uporabljamo vzorec brez ponavljanja, ker pridemo z njim do bolj natančnih rezultatov.

Ako je število enot osnovne statistične mase zelo veliko, so razlike v zakonitostih med vzorcem s ponavljanjem in vzorcem brez ponavljanja čezdalje manjše.

131.3 Masa vseh vzorcev.

Ako iz neke osnovne mase izvlečemo na slučajen način vzorec, je jasno, da to ni edini možen vzorec, ki ga moremo dobiti iz osnovne mase. Ako iz osnovne statistične mase izvlečemo ponovno vzorec z istim številom enot, dobimo prav gotovo v drugem vzorcu druge enote. Število vseh možnih različnih vzorcev iz ene in iste osnovne mase je odvisno od števila enot v osnovni masi in števila enot v vzorcu. Število možnih vzorcev je običajno zelo veliko število in se hitro veča, ako se veča število enot v osnovni statistični masi. Po pravilih, ki veljajo za tvorjenje statističnih mas, moremo skupnost vseh vzorcev smatrati kot statistično maso, vsak izmed možnih vzorcev pa kot statistično enoto mase vseh vzorcev. Opredeljujoči pogoji dane mase vseh vzorcev so naslednji: vsi vzorci imajo isto število enot in so vzeti na slučajen način iz ene in iste osnovne mase.

Pri vzorčenju nastopajo torej tri vrste statističnih mas: Osnovna masa, vzorec in masa vseh vzorcev. Ako vzamemo kot primer kmetijska gospodarstva, so enote osnovne mase in vzorca kmetijska gospodarstva, enote mase vseh vzorcev pa so skupnosti gospodarstev v posameznih vzorcih.

Ker so enote v masi vseh vzorcev drugačne, kot v osnovni masi in vzorcu, so tudi znaki drugačni. Statistični znak posameznega gospodarstva je n.pr. skupna površina, iz tega znaka izpeljan znak pri masi vseh vzorcev pa

povprečna površina. Statistični znak posameznega gospodarstva je, ali ima gospodarstvo vinograd ali ne, iz tega izpeljan znak pri masi vseh vzorcev pa je odstotek gospodarstev, ki imajo vinograd.

132. SLUČAJNI IZBOR KOT OSNOVNA VZORČENJA

Že uvodoma smo omenili, da je vzorčenje metoda delnega opazovanja, ki je zasnovana in velja le v tem primeru, da so enote, ki jih vključimo v vzorec izbrane na slučajen način, alim kot rečeno "na slepo". Ker je slučajnost vzorca osnova, na kateri je izgrajena vsa teorija vzorčenja in moremo le pod to predpostavko napraviti iz vzorca uporabne zaključke, je jasno, da je treba izbiri enot posvetiti vso pozornost. Zato je treba pri izbiri enot predvsem paziti, da je vzorec, ki smo ga izbrali, res slučajen, pri zbiranju podatkov na terenu pa predvsem, da se res popišejo enote, ki so določene z izborom.

132.1 Ogrodje ali okvir vzorčenja

Ako hočemo iz neke osnovne statistične mase izbor, je pred pogoj, da ta izbor lahko izvršimo, da imamo na kakršenkoli način podane vse enote osnovne statostične mase. Najenostavnejši našin je ta, da imamo vse enote osnovne mase podane v spisku, sezhamu, imeniku n.pr. naslovov gospodarstev, kmetijskih gospodarstev ali podjetij, šol itd.

Ta spisek enot mora biti popoln, to se pravi, da mora vsebovati vse enote osnovne mase. Enote osnovne mase moremo imeti v nekaterih primerih podane tudi na druge načine. Ako raziskujemo probleme iz kmetijstva, moremo imeti parcele nazorno podane z geografsko karto. Dostikrat so enote osnovne mase geografska področja v obliki kvadratov, ki jih dobimo, da geografsko karto razdelimo na kvadrate.

Spiske, sezname in imenike enot, geografske karte in druge pripomočke, ki tvorijo osnovo, da moremo iz osnovne mase izbrati vzorec, imenujemo ogrodje ali okvir vzorčenja.

Ogrodje mora biti prilagojeno vrsti izbora, kar pomeni, da morajo biti enote podane grupirane v geografske ali vsebinske grupe, ako je to glede na vrsto izbora potrebno.

Iz tehničnih razlogov je potrebno, da so enote v spiskih ali geografskih kartah oštevilčene z zaporednimi številkami. Zaporedne številke olajšajo tehnični postopek izbiranja enot.

132.2 Slučajni izbor brez omejitve

Najenostavnejši način slučajnega izbora je slučajni izbor brez omejitve. Pri izboru brez omejitve obstoja za vsako enoto osnovne mase ista močnost, da je izbrana v vzorec.

Izbor brez omejitve poznamo iz vsakdanjega življenja pri tomboli.

Pri tomboli vlečemo ~~številke iz žare na slepo in je odvisno samo od slučaja, katero številko~~ potegnemo. Nobena številka nima prednosti pred drugo. Pravimo, da je verjetnost, da bomo izvlekli iz žare neko določeno številko, za vse številke od 1 do 90, enaka. Verjetnost je nov pojm, katerega smo do sedaj omenili samo mimogrede. Moramo pa se z njim nekoliko podrobneje spoznati. Verjetnost in verjetnostni račun je namreč osnova teorije vzorčenja. Z njim pa se bomo spoznali samo v toliko, v kolikor je to nujno potrebno za razumevanje nadaljnjih izvajanj.

132.3 Verjetnost.

.31 Pojm. O verjetnosti govorimo vsak dan. Verjetno je, da ima vlak, ki pripelje iz Beograda, zamudo. Bolj verjetno je, da bo na nogometni tekmi zmagalo moštvo, ki ga štejemo kot boljšega, kot slabšo moštvo. Malo verjetno je, da nas povozi avto, ako upoštevamo cestne predpise. Podobnih verjetnih dogodkov bi mogli našteti nebroj. Čeprav ta "vsakdanja" verjetnost ni izračunana, temveč si jo za nek dogodek podzavestno ustvarimo, ima vendar isto osnovo kot strogo določen pojm verjetnosti v verjetnostnem računu.

Iz zgornjih primerov vidimo, da moremo verjetnost tudi v vsakdanjem smislu količinsko določiti. Govorili smo namreč o majhni, večji, veliki verjetnosti. Pojem velikosti verjetnosti si ustvarimo na naslednji način. Ako je imel brzovlak, ki pripelje dnevno iz Beograda, do sedaj dostikrat zamudo, je zelo verjetno, da jo bo imel tudi danes. To pa seveda ni gotovo, obstoja tudi možnost, da vlak zamude ne bo imel. Da je zelo mala verjetnost, da nas pri prehodu ceste ne bo povozil avto, ako bomo upoštevali cestne predpise, sklepamo iz tega, ker se je dejansko dosedaj zgodilo zelo malo nesreč v primeri z vsemi prehodi ceste. Pojem o velikosti verjetnosti smo si ustvarili na osnovi izkušenj že izvršenih dogodkov. Primerjali smo, čeprav ne računsko, število izvršenih dogodkov s številom primerov, v katerih bi se ta dogodek mogel zgoditi. Ako je imel brzovlak iz Beograda povprečno na 10 voženj enkrat zamudo, pravimo, da je verjetnost, da bo imel vlak tudi danes zamudo, majhna.

Tudi v verjetnostnem računu izražamo verjetnost z ulomkom med številom primerov, v katerih se je neki dogodek izvršil in številom primerov, v katerih bi se ta dogodek mogel izvršiti. Pri tem moramo število primerov večati v neskončno. Vzemimo, da imamo za nalogo določiti verjetnost, da s pravilno kocko vržemo šestico. Do rezultata pridemo na naslednji način. Kocko mečemo v zrak in pri vsakem metu zabeležimo, ali je padla šestica. Ako izvršimo z zelo veliko metov, moremo ugotoviti verjetnost, da vržemo s pravilno kocko šestico, tako, da število metov, pri katerih smo vrgli šestico, delimo s številom vseh izvršenih metov, ker je pri vsakem metu dana možnost, da vržemo šestico. Pri tem načinu smo ugotovili verjetnost šele z zelo velikim številom metov. Strogo vzeto, bi morali vreči kocko celo neskončno mnogokrat, ako bi hoteli priti do prave vrednosti verjetnosti. To pa

je tehnično zelo zamudno, oziroma sploh nemogoče. Na srečo poznamo iz verjetnostnega računa način, s katerim moremo v primerih, ki bodo prišli za nas v poštev, verjetnost določiti na veliko krajši način.

Pri metanju pravilne kocke moremo vreči naslednje številke: 1, 2, 3, 4, 5, 6. Ker je kocka pravilna, je za vsako izmed teh številk enaka možnost, da jo vržemo pri dejanskem metu. Vseh šest številk je med seboj enakomožnih. Ta enaka možnost se očituje v tem: Ako bi izvršili zelo veliko metov, na primer šest miljonov, bi bila potemtakem vsaka izmed šestih številk zastopana miljonkrat. Verjetnost, da vržemo n.pr. šestico je torej

$$P(6) = \frac{1{,}000.000}{6{,}000.000} = \frac{1}{6}$$

Ako bi izvršili dvanajst milijonov metov, bi zaradi enakomerne razvrstitve enako možnih dogodkov dobili vsako številko dva milijonkrat

$$P(6) = \frac{2{,}000.000}{12{,}000.000} = \frac{1}{6}$$

Na tem primeru vidimo, da verjetnost ni odvisna od števila izvršenih primerov. V vsakem primeru se v ulomku z določenim krajšanjem pride do istega rezultata 1/6. Pri proučitvi tega ulomka vidimo, da je 6 število vseh enakomožnih metov.

Verjetnost, da se zgodi eden izmed enakomožnih dogodkov, moremo torej izračunati v naprej, brez dejanskega izvršenja. Ako posplošimo zgornji primer, je verjetnost, da se zgodi neki enakomožen dogodek, enaka ulomku 1/n, kjer je n število vseh enakomožnih dogodkov.

$$P(e) = \frac{1}{n} \qquad (13.1)$$

P (e) pomeni verjetnost za enakomožen dogodek, n pa je število vseh enakomožnih dogodkov.

P Ako imamo 52 igralnih kart in izvlečemo na slepo eno karto, je za vsako izmed 52 igralnih kart enaka možnost, da jo izvlečemo. Verjetnost, da potegnemo pikovega asa je torej: P (as) = 1/52.

P Pri tomboli je v žari 90 številk. Za vsako izmed številk od 1 do 90 je enaka možnost, da jo pri slepem vlečenju izvlečemo. Verjetnost, da izvlečemo n.pr. številko 49, je torej P (49) = 1/90.

P Imamo popisno gradivo popisa kmetijskih gospodarstev, ki vsebuje 240.000 popisnic. Iz tega gradiva izvlečemo na slepo eno popisnico. Kakšna je verjetnost, da je to popisnica posestnika Franceta Porente iz Zgornje Hrušice ? Ker je za vsako izmed popisnic enaka možnost, da jo izvlečemo, je verjetnost, da izvlečemo ravno popisnico Franceta Porente iz Zgor. Hrušice 1/240.000. To je zelo majhna verjetnost. Pomeni pa naslednje: Ako bi izvlečenje popisnice ponovili zelo velikokrat, bi izvlekli popisnico Franceta Porente iz Zgor. Hrušice enkrat pri 240.000 izvlačenjih.

•32 Seštevanje verjetnosti. Do sedaj smo obravnavali verjetnost, da se dogodi eden izmed enakomožnih dogodkov. Pri primeru s kocko smo videli, da je verjetnost, da vržemo številko 1 1/6 in ista verjetnost 1/6 da vržemo številko 2. Vprašanje je, kolika je verjetnost, da vržemo s pravilno kocko številko 1 ali 2. Najprej poizkusimo z dejanskim metanjem kocke. Ako vržemo kocko 600.000 krat, dobimo med temi meti številko ena 100.000 krat in številko dve 100.000 krat. Skupno smo vrgli številko ena ali dve 200.000 krat. Po definiciji je verjetnost, da vržemo s kocko številko ena ali dve enaka $\frac{200.000}{600.000.}$ Ako ulomek skrajšamo, dobimo, da je verjetnost, da vržemo s kocko številko ena ali dve enaka 2/6. Do istega rezultata pa pridemo, ako seštejemo verjetnost, da s kocko vržemo številko ena in verjetnost, da vržemo številko 2.

$$\frac{1}{6} + \frac{1}{6} = \frac{2}{6} \qquad P\,(1 \text{ ali } 2) = P\,(1) + P\,(2) \qquad (13,2)$$

Ako se vprašamo, kakšna je verjetnost, da s pravilno kocko vržemo številko od ena do pet, ni treba storiti nič drugega, kot sešteti verjetnosti, da vržemo številko ena, dve, tri, štiri, pet.

$$P\,(1 \text{ ali } 2 \text{ ali } 3 \text{ ali } 4 \text{ ali } 5) = P(1) + P(2) + P(3) + P(4) + P(5) =$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{5}{6} .$$

To je zelo velika verjetnost in pomeni, da pri velikem številu metov v petih od šestih metov vržemo številko, ki je enaka ali manjša od pet.

Verjetnost, da vržemo katero koli izmed šestih številk je po zgornjem enaka $\frac{6}{6} = 1$. Ker je gotovo, da s pravilno kocko v vsakem metu vržemo eno izmed šestih številk, je verjetnost 1 istovetna z gotovostjo. Zelo verjetni, čeprav ne povsem gotovi so dogodki, katerih verjetnost je blizu 1 n.pr. 0,99. Ta verjetnost pomeni, da se pri velikem številu dogodkov dan dogodek izvrši v 99 primerih od 100 primerov.

132.4 Reprezentativnost slučajno izbranega vzorca.

P Ako pogledamo primer kmetijskih gospodarstev in na njih uporabimo gornje stavke, moramo napraviti zelo koristne zaključke. Vzemimo, da je od 240.000 gospodarstev 40.000 gospodarstev v višinskih legah, 200.000 pa v nižinskih legah. Ako iz statističnoga gradiva izvlečemo na slepo eno gospodarstvo, je bolj verjetno, da izvlečemo popisnico za gospodarstvo ki leži v nižini, kot popisnico, da leži v višinskih predelah. Točna verjetnost za nižinska gospodarstva je $\frac{200.000}{240.000} = \frac{5}{6}$ za višinsko pa $\frac{40.000}{240.900} = \frac{1}{6}$.

Ako izvlečemo na slepo več popisnic, ki jih smatramo kot vzorec, moremo pričakovati, da v povprečju izvlečemo ~~izvlečemo~~ na šest popisnic pet nižinskih in eno višinsko gospodarstvo. Ker to velja točno le v primeru, da izvršimo neskončno izvlačenj, velja za vzorec, ki ima končno število enot, le približno. To razmerje velja tem bolj, čim večji je vzorec. Podobno analizo bi mogli napraviti tudi za druge značilnosti, kot velikost gospodarstev, lego glede na okraje itd. Iz gornjega moremo sklepati, dabo vzorec vseboval več gospodarstev onih skupin ali okrajev, v katerih je število gospodarstev večje. Vzorec da torej reprezentativno sliko celote, samo, ako je izbor enot res slučajen. Reprezentativnost vzorca je zajamčena s tem, da so enote izbrane slučajno.

Iz tega je razvidno, kako važno je, da je vzorec izbran res na slučajni način. Zato mora biti način izbiranja tak, da zagotovi slučajnost izbora. To je ena izmed najvažnejših del v prvi stopnji vzorčenja. Za izbiranje enot na slepo obstoja več metod.

132.5 Načini slučajnega izbiranja enot vzorca.

.51 Vlečenje iz žare. Že na primeru smo omenili, da odgovarja vlečenje števil iz žare slučajnemu izboru brez omejitve. To lastnost vlečenja števil iz žare moremo uporabiti pri vzorčenju. Do slučajnega vzorca brez omejitve moremo ~~[illegible]~~ tako imamo vse popisnice celotnega gradiva zmešane v žari, iz katere nato na slepo vlečemo popisnico za popisnico. Čeprav je postopek enostaven, naletimo pri tem načinu na tehnične ovire. Obseg žare bi moral biti prevelik, zaradi česar je težko izvesti izbor na slepo na ta način.

Prikladnejše je, da žaro uporabimo v naslednji obliki. V žari imamo namesto popisnic listke z zaporednimi številkami. Popisno gradivo je tudi oštevilčeno z zaporednimi številkami. Iz žare potegnemo na slepo odgovarjajoče število listkov. V vzorec vključimo enote, ki imajo zaporedne številke, ki smo jih izvlekli iz žare.

Še enostavnejši način bi pa bil naslednji: Popisno gradivo imamo oštevilčeno s tekočimi številkami. V žari pa nimamo listke z vsemi zaporednimi številkami, temveč deset listkov ali krogljic, na katerih so označene številke od 0 do 9.

P Ako je skupno število enot celokupne statistične mase n.pr. 8.900, dobimo zaporedno številko enote, ki pride v vzorčenje, na naslednji način: deset krogljic, ki jih imamo v žari, dobro premešamo in potegnemo na slepo eno krogljico. Številka, ki je označena na krogljici pomeni tisočice zaporedne številke (potegnili smo n.pr. 4). Nato vržemo krogljico nazaj v žaro in znova potegnemo krogljico. Številka na tej krogljici pomeni stotice zaporedne številke (potegnili smo n.pr. 8). Enako ponovimo postopek za desetice (potegnili smo n.pr. 6) in ednice (potegnili smo n.pr. 1). Iz teh štirih

številk moremo sestaviti zaporedno številko enote, ki jo vključimo v vzorec. V našem primeru bi bila to enota z zaporedno številko 4861. Enako moremo določiti zaporedne številke za vse ostale enote, ki jih vključimo v vzorec.

.52 Tablice slučajnih številk. Kljub zgornjim poenostavitvam je izbiranje slučajnih enot s pomočjo žare vseeno neprikladno. Ker so številke, ki so slučajne za en primer, slučajne tudi za druge statistične akcije, so sestavili tako imenovane tablice slučajnih številk, ki zamenjujejo žaro. Določanje slučajnih tekočih številk s pomočjo tablic slučajnih števil se zelo poenostavi in je poleg tega znatno hitrejše. Zaradi tega jih uporabljamo skoro v vseh primerih vzorčenja. Tablice slučajnih števil, ki so jih izdelali različni avtorji (Romanovski, Tippett) so nastale na ta način, da so slučajne številke, ki so jih izvlekli iz žare ali dobili na kateri koli drug način, zapisali in jih uporabljamo pri različnih primerih vzorčenja.

Za primer uporabe tablic slučajnih številk, navajamo začetek tablic slučajnih številk, ki jih je sestavil Romanovski.

8574	5490	4069	0163	1241	1270	3768	4287	8486	9370
4575	6276	2709	4732	0301	8730	1672	5474	1585	1237
4999	8828	0291	0258	9430	1281	8148	7695	6015	3112
7627	6090	9572	0426	1218	4703	9764	3171	6567	1210
4315	5778	1508	9466	7012	1845	6474	4083	9658	9171
6987	8055	0026	8093	7121	8061	0452	2984	6916	6010
0387	9994	0103	3705	4252	5806	1301	4848	9942	1027
5581	2184	9763	8160	5917	1851	3464	6626	8904	1004
6531	8780	1572	1400	6529	1274	4844	2649	0976	4698
5735	5350	9828	5652	3698	5365	1580	7026	2630	9280
6092	0979	6190	2410	0650	3211	2402	4720	9314	3013
1791	3893	7019	3530	3463	6165	9063	5292	4224	4343
9746	5248	3866	3797	8070	5821	2595	2072	1339	5398
0118	1348	6571	0497	4376	2543	3898	0534	4308	2171
0986	3888	4252	5736	7093	8166	1869	4680	0564	8615

P Primer uporabe tablice slučajnih števil: I_z osnovne mase, ki ima 736 enot, moramo izbrati na slučajen način 30 enot.

Najprej oštevilčimo enote za zaporednimi številkami od 1 do 736. Pri tem ni važen vrstni red enot, temveč samo to, da ima vsaka enota svojo številko. Slučajna števila v zgornji tablici so štirimestna. Ker pridejo v našem primeru v poštev največ tromestna števila, vzamemo v poštev samo prva tri mesta štirištevilčnih števil. Z iste upravičenosti bi mogli vzeti tudi zadnja tri mesta. Od tromestnih števil, ki jih na ta način dobimo, pridejo v našem primeru v poštev samo ona, ki so manjša od 736. Ako pričnemo v prvi koloni izpadá 574, pride pa v poštev 457, 499, v nadaljnjem izpade 762 itd. Ako izvajamo postopek naprej, dobimo naslednjih trideset slučajnih števil.

457 011 524 499 098 134 431 549 388 698
627 406 038 609 270 558 577 029 653 218
150 573 535 002 609 097 010 179 389 157

•53 Sistematični izbor enot vzorca. Kljub olajšavam pa je slučajni izbor s pomočjo slučajnih števil vseeno razmeroma zapleten, posebno še v primeru, ako izbirajo enote organi na terenu, ki v tem poslu niso izvežbani, niti dobro poučeni o važnosti tega, da so enote izbrane slučajno.

Zaradi tega uporabljamo dostikrat izbor, ki ga imenujemo mehaničen ali sistematičen izbor. Vprašanje je, ali moremo smatrati kot slučajen vzorec enote, ako na primer pri desetodstotnem vzorcu ne izbiramo enote s tablicami slučajnih števil, temveč vzemamo v vzorec kar vsako deseto enoto po vrstnem redu spiska enot. Ta postopek je seveda najenostavnejši od vseh, ki smo jih omenili do sedaj, vendar ima slabo lastnost, da ni v vseh primerih reprezentativen in slučajen, kar zelo zniža njegovo ceno. Statistično maso, oziroma spisek enot, ki služi kot ogrodje izbora, je treba podrobno proučiti, predno na njem izvedemo mehaničen, oziroma sistematičen izbor.

Zaradi česa morejo nastopiti primeri, da skupnost enot, ki smo jih izbrali z mehaničnim izborom, ne more služiti kot vzorec, ker ni reprezentativna in slučajna, bomo videli najlepše na primeru.

P Vzemimo da smo pri popisu prebivalstva uporabljali kolektivne obrazce in vanj zapisovali zapovrstjo člane družin, družino za družino. Za vsako družino smo vpisovali najprej družinskega poglavarja, nato ženo, potem pa otroke. Vzemimo zaradi tega, da bo nepravilnost postopka še bolj razvidna, da imajo vse družine, ki smo jih popisali po pet članov. Družinskega poglavarja zaznamujemo z G, ženo z Ž, sina z S, hčerko pa z H. Vrstni red v popisnem spisku je potem naslednji:

G Ž S S H, G Ž S S H, G Ž S S S, G Ž H H S, G S S H H ...

Ako vzamemo v tem popisnem seznamu vsako deseto osobo in začnemo s prvo popisnico, je na prvi pogled razvidno, da skupnost teh oseb ni reprezentativna za celoto, ker dobimo v zorec same družinske poglavarje, kar ni reprezentativno niti za spol (sami moški), niti za starost (sami starejši), delavnost ali vzdrževanost (pretežno delavni) itd. Do teh neprilik je prišlo zaradi tega, ker se določene lastnosti popisnih enot periodično ponavljajo (na vsakih pet enot).

Iz tega primera moremo napraviti splošen sklep, da mehaničen izbor ne moremo uporabiti, kadar je v vrstnem redu kak sistem, oziroma periodičnost. Samo v primeru, da je vrstni red enot v spisku slučajen, moremo skupnost ki smo jih izbrali s sistematičnim izborom, smatrati kot reprezentativen vzorec. O tem, ali je v vrstnem redu periodičnost ali ne, se moremo prepričati samo s proučitvijo spiska in s poizkusom.

Enote, ki jih vključujemo v izbor morajo biti tako izbrane, da pride v poštev pri izboru celotna statistična masa in ne samo en del. Ako moramo iz osnovne statistične mase, ki ima 6000 enot, izbrati v vzorec 300 enot, ne bomo vzeli vsako deseto enoto po spisku, ker pridemo na ta način samo do polovice spiska. Taka skupnost pa seveda ni reprezentativna. Vzeti moramo tak količnik, da bomo z mehaničnim izborom prešli celotno maso. V našem primeru moramo zaradi tega izbrati vsako dvajseto enoto. V izbor pridejo na ta način enote z zaporednimi številkami: 20, 40, 60, 80.... 3000, 3020, 3040 ..... 5900, 5920, 5940, 5960, 5980, 6000. Vidimo, da so v tem primeru izbrane enote raztresene po vsem spisku. Količnik 20 smo dobili na ta način, da smo število enot celokupne mase delili s številom enot, kolikor jih vsebuje vzorčna masa.

Ako zaznamujemo število enot celotne statistične mase z N in število enot vzorčne mase z n, izračunamo količnik K na naslednji način:

$$K = \frac{N}{n} \qquad (13,3)$$

Ako pa je dano z odstotkom, kakšen del enot celokupne mase naj obsega vzorec, izračunamo količnik po naslednjem obrazcu:

$$K = \frac{100}{P} \qquad (13,4)$$

Pri tem pomeni K količnik, P pa procent, kakšen del celote predstavlja vzorec.

Ta obrazec sledi iz naslednjega:

$$P = \frac{n}{N} \cdot 100 \text{ (strukturni procent)},$$

Ker iz zgornjega sledi:

$$\frac{N}{n} = \frac{100}{P}, \text{ moremo pisati: } K = \frac{100}{P}$$

Ako je število enot v celotni statistični masi n.pr. 28345, izbrati pa moramo 912 enot, je količnik enak:

$$K = \frac{N}{n} = \frac{28345}{930} = 30,8$$

Ker moremo uporabiti samo cel količnik, ga zaokrožimo na najbližjo celo vrednost. V vzorec bomo vzeli torej vsako enaintrideseto enoto. Prvo enoto izberemo na slučajen način, ostale pa mehanične.

Ako vršimo n.pr. 2,5 °/o izbor, je količnik mehaničnega izbora enak:

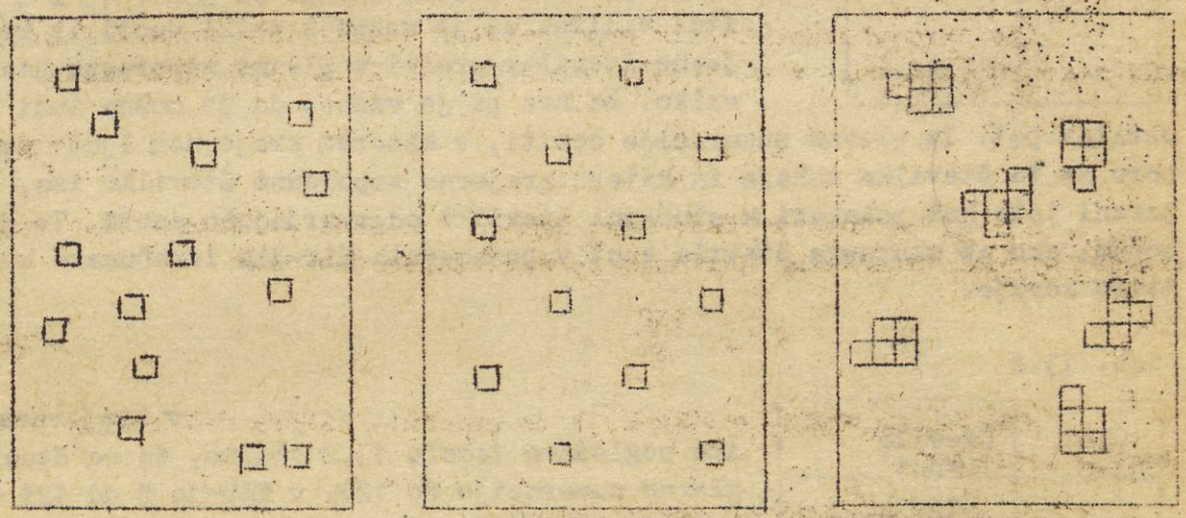
$$K = \frac{100}{P} = \frac{100}{2,5} = 40$$

V vzorec pride torej vsaka štirideseta enota.

Kadar gre za vzorec na površinah, govorimo o **sistematičnem** vzorcu, ako so poizkusne parcelice druga od druge enako oddaljene

Na naslednjih treh slikah imamo narisano, kako izgledajo posamezne vrsto vzorcev na površinah.

Sl. 13.1 Izbor na površinah



1. slučajen izbor 2. Mehaničen izbor 3. izbor v blokih

.26 Postopek izbora v primeru delnih spiskov. Oštevilčenje celotnega popisnega gradiva z enotnim sistemom zaporednih številk je zelo zamudno. V primerih, da gradivo ni zbrano centralno, pa je celo nemogoče. Običajno ima popisno gradivo zaporedne številke, ki pa se začnejo z eno za vsak krajevni ljudski odbor ali popisni okoliš. Te zaporedne številke niso namenjene za potrebe vzorčenja, vendar jih moremo za vzorčenje izkoristiti. Na zelo enostaven način moremo dobiti zvezo med delnimi numeracijami in glavno numeracijo s pomočjo kumulativ.

P Vzemimo, da je spisek napravljen po krajevnih ljudskih odborih, in v vsakem izmed njih samostojno numeriran. Vzemimo, da je v začetnih krajevnih odborih število enot, kot kaže tabela 13,1.

Tab. 13.1

KLO	Število enot
A	120
B	216
C	83
D	163
E	516
F	216
.	.
.	.
.	.

Krajevna numeracija v krajevnem ljudskem odboru A se sklada z glavno numeracijo. Prvo gospodarstvo v krajevnem ljudskem odboru B pa ima po glavni numeraciji zaporedno številko 120 + 1 = = 121, drugo 120 + 2 = 122, zadnje ali 216 -to pa 120 + 216 = 366. Prvo gospodarstvo v krajevnem ljudskem odboru C ima za eno višjo glavno številko kot 366. 366 + 1 = 337, drugo za dve večjo 336 + 2 = 338, zadnje pa 336 + 83 = 419 itd. Vidimo, da je možno k vsaki enoti iz krajevne številke preiti v glavno zaporedno številko. Za nas pa je važno, da je možna tudi obratna pot: Iz glavne numeracije dobiti, v katerem krajevnem ljudskem odboru se ta številka nahaja in katero krajevno zaporedno številko ima, kar pomeni isto kot poiskati k slučajni številki odgovarjajočo enoto. To je možno, ako iz skupnega števila enot v posameznih KLO-jih izračunamo kumulativno serijo.

Tab. 13.2

KLO	Število enot	0
A	120	120
B	216	336
C	83	419
D	163	582
E	516	1098
F	216	1314

Ako pogledamo tabelo 13,2 vidimo, da so v krajevnem odboru A številke glavne numeracije do 120, v KLO-ju B od 121 - 336, v C od 337 - 419 itd.

Zaporedna številka 596 predstavlja na primer potemtakem (596 - 582 = 14) štirinajsto enoto v v KLO-ju E, zaporedna številka 1216 (1216 - 1098 = 118), stoosemnajsto enoto v KLO-ju F itd. Postopek je zelo enostaven in znatno olajša posel.

133. OCENJEVANJE S POMOČJO VZORČENJA

133.1 Ocenjevanje aritmetične sredine celotne mase z vzorcem

.11 Ocena. Kot oceno aritmetične sredine celote moremo smatrati aritmetično sredino vzorca. Zelo verjetno je namreč, da se aritmetična sredina celote ne razlikuje od vrednosti ocene za več kot neko določeno vrednost, ki jo je možno določiti. Ocena aritmetične sredine, ki jo zaznamujemo z $\bar{x}'$, kjer pomeni črtica, da je to ocena, je torej

$$\bar{x}' = \frac{1}{n} Sx \qquad (13.5)$$

Sx pomeni, da gre za vsoto vrednosti x v vzorcu, za razliko od $\sum x$, ki pomeni vsoto vrednosti x za celotno statistično maso.

Kot primer vzemimo 50 kmetijskih gospodarstev in opazujmo število goved v vsakem gospodarstvu. Teh 50 gospodarstev naj predstavlja osnovno maso. Podatki so dani v spisku, tabela 13.3, ki služi kot ogrodje vzorčenja. Iz te osnovne mase izberimo 5 slučajnih vzorcev po 25 enot in izračunajmo za vsakega posebej aritmetično sredino. Vsaka izmed teh aritmetičnih sredin vzorca predstavlja samostojno oceno aritmetične sredine celotne mase. Iz primera bomo videli, v koliko se te ocene razlikujejo od prave aritmetične sredine celotne mase.

Tab. 13.3 Število krav za 50 kmetijskih gospodarstev

Zap. št.	Štev. krav	Zap. št.	Štev. krav	Zap. št.	Štev. krav	Zap. št.	Štev. krav	Zap. št.	Štev. krav
1	4	11	4	21	1	31	2	41	1
2	2	12	1	22	6	32	1	42	5
3	1	13	3	23	1	33	1	43	4
4	2	14	3	24	3	34	3	44	2
5	3	15	2	25	1	35	3	45	3
6	1	16	1	26	2	36	3	46	1
7	1	17	5	27	2	37	1	47	1
8	5	18	1	28	1	38	2	48	2
9	1	19	1	29	2	39	2	49	1
10	4	20	2	30	4	40	4	50	1

Iz tablice slučajnih števil v odstavku 132.52 moremo dobiti 25 slučajnih števil na ta način, da izpišemo 25 dvomestnih števil, ki so manjše od 50. Ako pričnemo n.pr. v tretjem stolpcu od zgoraj navzdol, moremo iz posameznega štirimestnega števila sestaviti šest dvomestnih, na ta način, da beremo zaporedno po dve mesti in sicer naprej in nazaj. Iz 0163 moremo sestaviti n.pr. 01, 16, 63, 36, 61, 10. Od teh števil so štiri manjša od 50 in jih moremo uporabiti kot slučajne številke v našem vzorcu. Na enak način postopamo naprej. Tako dobimo naslednje serije po 25 števil oz. gospodarstev. Zadnja serija je izbrana kot mehaničen vzorec in smo vzeli vsako drugo gospodarstvo (glej tabelo 13.4)

Tab. 13.4

Prvi vzorec		Drugi vzorec		Tretji vzorec		Četrti vzorec		Peti vzorec	
Sluč. štev.	Štev. krav	Sluč. štev.	Štev. krav	Sluč. štev.	Štev. krav	Sluč. štev.	Štev. krav	Sluč. štev.	Štev. krav
01	4	18	1	29	2	3	1	1	4
16	1	14	3	19	1	30	4	3	1
36	3	40	4	31	2	12	1	5	3
10	4	4	2	11	4	24	3	7	1
47	1	41	1	10	4	41	1	9	1
32	1	25	1	13	3	14	3	11	4
23	1	24	3	21	1	42	5	13	3
37	1	23	1	17	5	21	1	15	2
02	2	10	4	7	1	1	4	17	5
25	1	01	4	12	1	10	4	19	1
20	2	42	5	15	2	43	4	21	1
04	2	35	3	16	1	28	1	23	1
41	1	30	4	34	4	34	3	25	1
21	1	03	1	6	1	49	1	27	2
14	4	37	1	3	1	18	1	29	2
40	4	49	1	43	4	7	1	31	1
46	1	36	3	27	2	17	5	33	1
49	1	13	3	49	1	25	1	35	3
09	1	45	3	47	1	19	1	37	1
39	2	31	2	38	2	29	2	39	2
08	5	43	4	39	2	36	3	41	1
50	1	47	1	48	2	6	1	43	4
7	1	26	2	32	1	50	1	45	3
6	1	39	2	2	2	5	3	47	1
05	3	34	3	39	2	46	1	49	1
$\sum$	49		62		52		56		50
$\bar{x}' =$	1,96		2,48		2,08		2,24		2,00

Prava aritmetična sredina je enaka:

$$\bar{x} = \frac{1}{N} \sum x = \frac{1}{50} \cdot 113 = 2,26$$

Ocene pa so po vrsti: $\bar{x}'_1 = 1,96$; $\bar{x}'_2 = 2,48$; $\bar{x}'_3 = 2,08$; $\bar{x}'_4 = 2,24$; $\bar{x}'_5 = 2,00$

Iz navedenih rezultatov vidimo, da je ena ocena (2,24) zelo blizu pravi aritmetični sredini, druge pa se bolj razlikujejo, vendar tudi

največji odklon znaša samo 17 °/o.

Odkloni, ki so se pojavili v našem primeru, pa so razmeroma vseeno zelo veliki. To izvira iz tega, ker je število enot v vzorcu razmeroma majhno. Od števila enot v vzorcu namreč bistveno zavisi kakovost ocene. Ker vsebujejo v praksi vzorci več sto ali tisoč enot, je jasno, da je kakovost teh ocen znatno boljša. O tem se bomo prepričali na primeru iz 2 °/o vzorca popisa prebivalstva.

P Pri popisu prebivalstva v LRS 15.marca 1948 je bil iz celotnega števila vseh gospodinjstev (N = 368.754) izbran vzorec, sestavljen iz 7351 "na slepo" izbranih gospodinjstev. Med drugimi znaki smo opazovali tudi število članov gospodinjstev. Iz tega vzorca je treba oceniti povprečno število članov na eno gospodinjstvo. Podatki vzorca so dani v spodnji frekvenčni distribuciji. Ker predstavlja oceno povprečja aritmetična sredina vzorca, moramo iz dane distribucije izračunati aritmetično sredino.

Tab 13.3

Štev. čl.	Število gospodinjstev	7351
1	1313	6038
2	1120	4918
3	1255	3663
4	1219	2444
5	941	1503
6	632	871
7	375	496
8	266	230
9	130	110
10	53	57
11	27	28
12	19	9
13	6	3
14	2	1
15	1	0
	7351	20371

$N_o = 7351 \qquad N'_o = 20371$

$$\bar{x}' = x_o + \frac{N'_o}{N_o} = 1 + \frac{20370}{7351} = 3,7712$$

Aritmetično sredino smo izračunali z metodo kumulativ. Ocena je v tem primeru zelo dobra, ker je napravljena na osnovi velikega vzorca (7351 enot). Ocena povprečnega števila članov na eno gospodinjstvo v LRS se od pravega povprečnega števila članov na eno gospodinjstvo, ki je 3,7748, razlikuje šele v tretji decimalki. Relativno znaša ta odklon komaj

$$\frac{0,0036}{3,7748}$$, kar ni niti 1 °/o.

.12 Distribucija aritmetičnih sredin vzorca. Standardna pogreška.

V obeh primerih prejšnjega odstavka smo imeli dano poleg ocene sredine tudi pravo sredino, ki smo jo izračunali iz celotne osnovne mase. S primerjavo ocene s pravo sredino smo se mogli prepričati o kakovosti ocene. Vendar ta način v praksi ni vedno mogoč. Oceno izračunavamo ravno zaradi tega, da ni treba izračunati popolnih podatkov. Primerjava ocene s pravim

rezultato torej ni možna. Teorija vzorčenja pa nas uči, da moremo za - nesljivost ocene določiti iz samega vzorca, brez poznavanja popolnih podatkov.

V primeru števila goved 50 kmetijskih gospodarstev smo izbrali 5 vzorcev po 25 gospodarstev. Vsak izmed teh vzorcev je vseboval druga gospodarstva. Na prvi pogled je jasno, da bi dobili še druge vzorce, ako bi še naprej izbirali na slučajen način po 25 gospodarstev. Število vseh možnih različnih vzorcev je zelo veliko. Kot primer navedimo, da je v našem primeru, kjer je število enot osnovne mase in vzorca razmeroma majhno, število vseh možnih različnih vzorcev po 25 enot cca. 12,8 bilijonov. Prav toliko pa je tudi ocen aritmetične sredine. Pri večjih masah je število vseh možnih vzorcev in število ocen še ogromno večje. Kljub temu, da je število vzorcev in število ocen tako veliko, poznamo nekatere zakonitosti, ki veljajo za vse mase vzorcev,

Skupnost vseh možnih vzorcev tvori maso vzorcev. Posamezni vzorec je enota te mase, aritmetična sredina vzorca pa znak.

Splošne zakonitosti mase vzorcev pa so naslednje:

1. Povprečje iz aritmetične sredine vseh vzorcev je enako aritmetični sredini osnovne statistične mase. V simbolih moremo to zakonitost pisati:

$$\bar{\bar{x}}' = \bar{x} \qquad (13.6)$$

2. Ako ocene aritmetične sredine, ki smo jih dobili iz vseh možnih vzorcev, zgrupiramo v frekvenčno distribucijo, dobimo normalno distribucijo. Ocene aritmetične sredine se torej distributirajo okrog prave aritmetične sredine v obliki normalne krivulje. Ta zakonitost velja za vse vzorce, ki imajo več kot trideset enot.

3. Standardna deviacija ocen aritmetične sredine vseh vzorcev, ki jo imenujemo na kratko standardna napaka, je enaka:

$$\sigma_{\bar{x}'} = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad (13.7)$$

V tem obrazcu pomeni $\sigma_{\bar{x}'}$ standardno deviacijo vseh ocen aritmetičnih sredin σ_x standardno deviacijo znaka x v osnovni masi, n število enot v vzorcu, N pa število enot v osnovni masi.

Ker pa σ_x ne moremo izračunati, ker ne poznamo celotne mase, moremo $\sigma_{\bar{x}'}$ s pomočjo oceniti. Oceno standardne pogreške $\tilde{\sigma}_{\bar{x}'}$ izračunamo po obrazcu:

$$\sigma'_{\bar{x}'} = \frac{s_x}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}} \qquad (13.8)$$

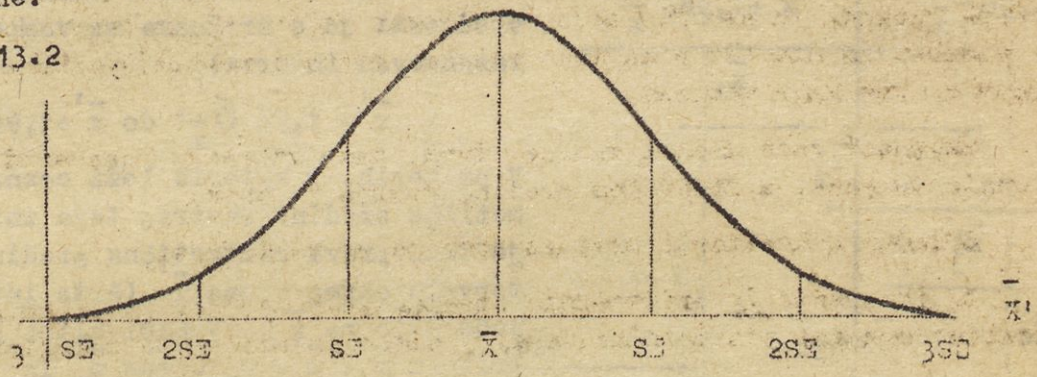
$$s_x = \frac{1}{n} S (x - \bar{x}')^2 \qquad (13.9)$$

V tem obrazcu pomeni $\sigma_{\bar{x}'}$ oceno standardne pogreške aritmetične sredine, s_x standardno deviacijo znaka x v vzorcu, n število enot v vzorcu, N pa število enot v osnovni masi. Standardno pogreško moremo torej izračunati iz podatkov, ki jih imamo na razpolago iz samega vzorca.

.13 Verjeten odklon. Meje zaupanja.

.131 Pravi verjeten odklon in meje zaupanja. S temi tremi zakonitostmi imamo točno določeno frekvenčno distribucijo vseh ocen aritmetične sredine.

Sl. 13.2



Iz slike 13.2 vidimo, da pade največ ocen v neposredno okolico prave aritmetične sredine (normalna krivulja je tu najvišja). Ocen, ki so zelo različne od prave aritmetične sredine, pa je malo. Iz tega sklepamo, da je zelo verjetno, da je odklon ocene, ki smo jo dobili z danim vzorcem, od aritmetične sredine preje majhen kot pa velik. Še več. Iz zakonitosti, ki veljajo za normalno distribucijo (glej odstavek), moremo odrediti interval, v katerem se nahaja določeno število vseh vzorcev.

V intervalu $\bar{x} - \sigma_{\bar{x}'}$ do $\bar{x} + \sigma_{\bar{x}'}$ se nahaja 68,4 °/o vseh ocen.

V intervalu $\bar{x} - 1{,}96\,\sigma_{\bar{x}'}$ do $\bar{x} + 1{,}96\,\sigma_{\bar{x}'}$ se nahaja 95 °/o vseh ocen,

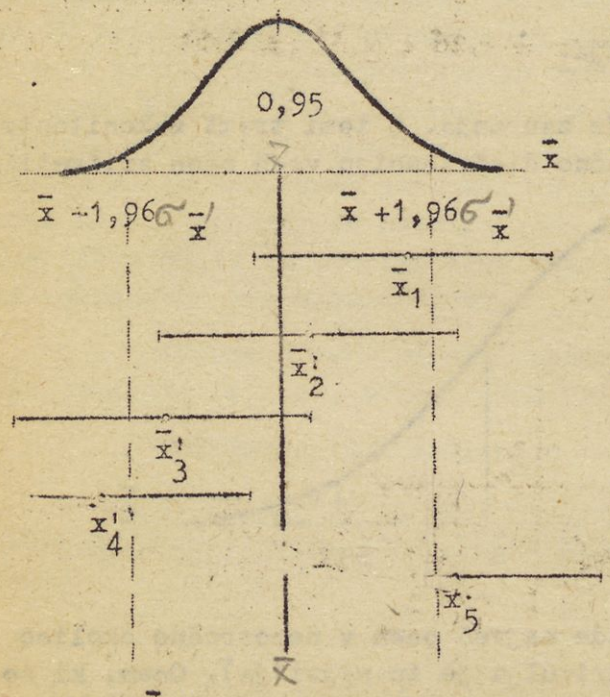
v intervalu $\bar{x} - 2{,}58\,\sigma_{\bar{x}'}$ do $\bar{x} + 2{,}58\,\sigma_{\bar{x}'}$ se nahaja 99 °/o vseh ocen.

To pomeni: Ako izbiramo na slučajen način veliko vzorcev in za vsak vzorec izračunamo aritmetično sredino, bo v povprečju v 95 primerih od 100, ali v 19 od 20 primerov aritmetična sredina vzorca ležala v intervalu $\bar{x} - 1{,}96\,\sigma_{\bar{x}'}$ do $\bar{x} + 1{,}96\,\sigma_{\bar{x}'}$. Obratno moremo reči, da bo v povprečju v petih primerih od 100 ali v enem od dvajsetih ležala aritmetična sredina vzorca

izven intervala $\bar{x} - 1{,}96\,\sigma_{\bar{x}'}$ do $\bar{x} + 1{,}96\,\sigma_{\bar{x}'}$

Iz tega je razvidno, da je samo 5 °/o ocen od prave aritmetične sredine različnih za več kot $1{,}96\,\sigma_{\bar{x}'}$. Obratno pa moremo trditi, da je samo 5 °/o vzorcev takih, da se prava aritmetična sredina razlikuje od ocene za več kot $1{,}96\,\sigma_{\bar{x}'}$. To je zelo lepo razvidno iz slike 13,3

Sl. 13,3



Naslikana je normalna krivulja z intervalom $\bar{x} - 1{,}96\,\sigma_{\bar{x}'}$ do $\bar{x} + 1{,}96\,\sigma_{\bar{x}'}$. Pod njo je narisano območje tega intervala. S pikami je označeno nekaj aritmetičnih sredin vzorcev. Okrog vsake vrednosti je s črticama na vsako stran zaznamovan interval

$$\bar{x}' - 1{,}96\,\sigma_{\bar{x}'} \text{ do } \bar{x}' + 1{,}96\,\sigma_{\bar{x}'}.$$

V primerih, v katerih leži ocena aritmetične sredine znotraj tega intervala, je tudi prava aritmetična sredina v intervalu okrog ocene $\bar{x}$ (črta intervala seče črto za $\bar{x}$). V naši sliki je to v primeru $\bar{x}'_1$, $\bar{x}'_2$, $\bar{x}'_3$. Ocene $\bar{x}'_4$ in $\bar{x}'_5$ pa ... interval $\bar{x} - 1{,}96\,\sigma_{\bar{x}'}$.

... je prava aritmetična sredina $\bar{x}$ izven intervala okrog ocene. To vidimo iz ... črta, ki ponazarja interval okrog ocene $\bar{x}'_4$ ... $\bar{x}'_5$ ne seže do prave aritmetične sredine $\bar{x}$.

Iz tega moremo sklepati, da je v vseh primerih, v katerih je ocena $\bar{x}'$ v intervalu $\bar{x} - 1{,}96\,\sigma_{\bar{x}'}$ do $\bar{x} + 1{,}96\,\sigma_{\bar{x}'}$ tudi prava aritmetična sredina $\bar{x}$ v intervalu $\bar{x}' - 1{,}96\,\sigma_{\bar{x}'}$ do $\bar{x}' + 1{,}96\,\sigma_{\bar{x}'}$. Obratno pa velja, ako je ocena $\bar{x}$ izven intervala $\bar{x} \pm 1{,}96\,\sigma_{\bar{x}'}$.

Zaradi tega moremo reči, da leži v 95 primerih od 100 vzorcev prava aritmetična sredina v intervalu, ki ga dobimo, ako oceni $\bar{x}'$ odštejemo in prištejemo $1{,}96\,\sigma_{\bar{x}'}$. Meji intervala $\bar{x}' - 1{,}96\,\sigma_{\bar{x}'}$ in $\bar{x}' + 1{,}96\,\sigma_{\bar{x}'}$ imenujemo meji zaupanja, vrednost $1{,}96\,\sigma_{\bar{x}'}$, ki ga zaznamujemo z $\varepsilon_{\bar{x}}$ pa verjeten odklon aritmetične sredine.

Ako hočemo preizkusiti gornje sklepe na našem primeru 50 gospodarstev, moramo izračunati še σ_x oziroma $\sigma_{\bar{x}'}$.

Da ne ponavljamo zanih postopkov izračunavanja standardne deviacije, podajamo $\bar{x}$ in za $\sigma_{\bar{x}}$ kar rezultate:

verjetni odklon

$\bar{x} = 2{,}26 \qquad \sigma_x = 1{,}357$

$$\sigma_{\bar{x}'} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1{,}357}{\sqrt{25}}\sqrt{\frac{50-25}{50-1}} = \frac{1{,}357}{\sqrt{25}}\sqrt{\frac{25}{49}} = \frac{1{,}357}{7} = 0{,}194$$

$e_{\bar{x}} = 1{,}96\,\sigma_{\bar{x}'} = 1{,}96 \cdot 0{,}194 = 0{,}380$

$\bar{x} - e_{\bar{x}} = 2{,}26 - 0{,}38 = 1{,}88 \qquad \bar{x} + e_{\bar{x}} = 2{,}26 + 0{,}38 = 2{,}64$

$\bar{x} = 2{,}26$

$\bar{x}'_1 = 1{,}96$

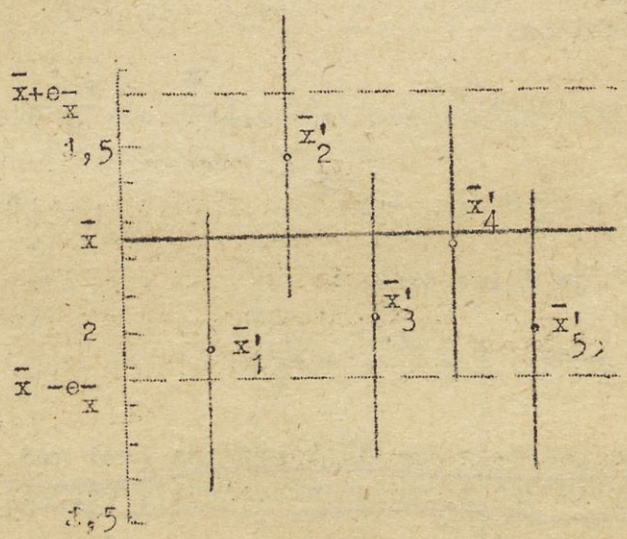
$\bar{x}'_2 = 2{,}48$

$\bar{x}'_3 = 2{,}04$

$\bar{x}'_4 = 2{,}24$

$\bar{x}'_5 = 2{,}00$

Sl. 13,4



Iz slike 13.4, kjer je vrisanih vsdh pet ocen z odgovarjajočimi verjetnostnimi odkloni, je razvidno, da leži aritmetična sredina $\bar{x}$ v intervalih okrog vseh petih ocen, kar je v skladu z zgornjimi izvajanji.

.132 Ocena verjetnega odklona in mej zaupanja. Z določitvijo verjetnega odklona in mej zaupanja smo storili že velik korak naprej. Poleg ocene aritmetične sredine $\bar{x}'$ imamo še verjeten odklon $e_{\bar{x}}$. Ta pove, da 95 od 100 sredin vzorcev ni od prave aritmetične sredine različnih za več, kot je velik verjeten odklon. Iz ocene sredine $\bar{x}'$ in verjetnega odklona $e_{\bar{x}}$

moremo izračunati meji zaupanja $\bar{x}' - e_{\bar{x}}$ in $\bar{x}' + e_{\bar{x}}$. Verjetnost, da prava aritmetična sredina leži v intervalu med mejama zaupanja je 0,95. V povprečju se torej v devetnajstih primerih od dvajset vzorcev prava aritmetična sredina nahaja v intervalu med mejama zaupanja. S tem je torej dana natančnost ocene. Vendar nastopi pri izračunavanju verjetnega odklona težava. Verjeten odklon $e_{\bar{x}} = 1{,}96\ \sigma_{\bar{x}'}$ izračunavamo s pomočjo standardne pogreške, ki pa vsebuje standardno deviacijo osnovne mase. To pa je nemogoče izračunati, ker nimamo dane podatke za osnovno maso, temveč samo za vzorec. Kot pa smo videli, je možno standardno pogreško oceniti iz podatkov vzorca. Ako v obrazec $e_{\bar{x}} = 1{,}96\ \sigma_{\bar{x}'}$ vstavimo namesto prave standardne pogreške oceno $\sigma'_{\bar{x}'}$, dobimo oceno verjetnega odklona.

$$e'_{\bar{x}} = 1{,}96\ \sigma'_{\bar{x}'} = 1{,}96 \cdot \frac{s_x}{\sqrt{n-1}} \cdot \sqrt{\frac{N-n}{N}} \qquad (13.10)$$

$$s_x = \frac{1}{n} S\ (x_i - \bar{x})^2 \qquad (13.11)$$

Ocena mej zaupanja pa je dalje enaka

$$\bar{x}' - e'_{\bar{x}} < \bar{x} < \bar{x}' + e'_{\bar{x}} \qquad (13.12)$$

Ta neenačba pomeni, da prava aritmetična sredina leži med mejama zaupanja. Na ta način nam je uspelo dobiti oceno sredine $\bar{x}'$, verjetnega odklona $e'_{\bar{x}}$ in mej kaupanja $\bar{x}' - e'_{\bar{x}}$ in $\bar{x}' + e'_{\bar{x}}$ iz podatkov samega vzorca.

P Preizkusimo ocenjevanje na naših petih vzorcih in 50 gospodarstev. Ocene $\bar{x}'$ imamo že izračunane. Po znanem postopku smo za vsak vzorec izračunali s_x

verjeten odklon $e'_{\bar{x}} = \frac{s_x \cdot 1{,}96}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}}$

ker je za vse vzorce faktor $\frac{1{,}96}{n-1} \sqrt{\frac{N-n}{N}} = \frac{1{,}96}{24} \sqrt{\frac{25}{50}} = 0{,}283$

stalna količina,

je $e'_{\bar{x}} = 0{,}283 \cdot s_x$.

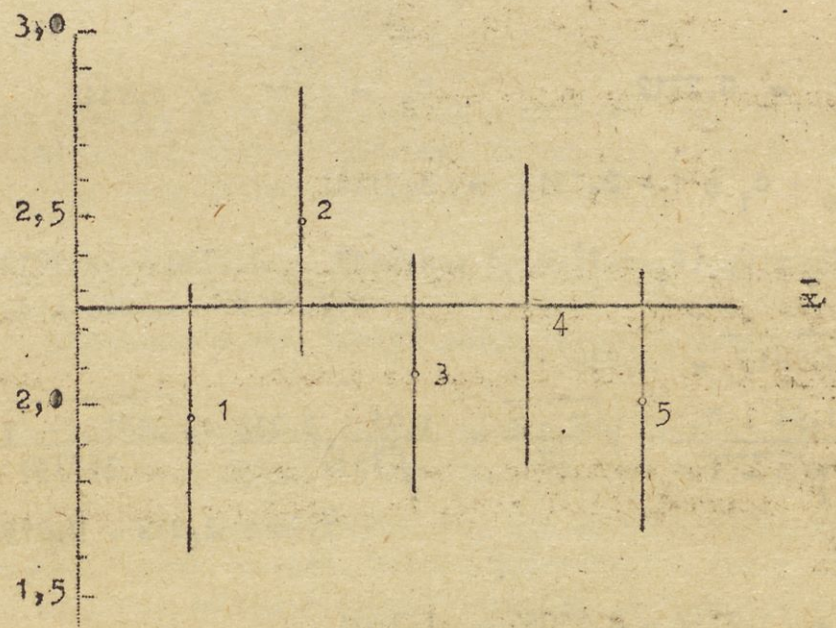
Podatki za vseh pet vzorcev so dani v naslednji razpredelnici:

Tab. 13.6

Vzorec	$\bar{x}'$	s_x	$e'_{\bar{x}} = 0{,}283 \cdot s_x$	$\bar{x}' - e'_{\bar{x}}$	$\bar{x}' + e'_{\bar{x}}$
1	1,96	1,28	0,362	1,598	2,322
2	2,48	1,236	0,35	2,130	2,830
3	2,08	1,197	0,338	1,742	2,418
4	2,24	1,420	0,402	1,838	2,642
5	2,00	1,232	0,239	1,651	2,349

Grafično moremo prikazati rezultate v sliki 13.5 na podoben način kot v prejšnjem poglavju v sliki 13,4

Sl. 13.5



Vidimo, da je v vseh petih primerih prava aritmetična sredina znotraj ocenjenega intervala. Intervali okrog ocene aritmetične sredine so različno visoki, ker so izračunani s pomočjo posameznih ocen.

P V odstavku 133.11 smo izračunali oceno aritmetične sredine povprečnega števila članov na 1 gospodinjstvo v LRS po popisu l. 1948. Vendar nismo še znali izračunati mej zaupanja, ki podajo kakovost oziroma natančnost ocene. Ponovimo primer in zanj izračunajmo meje zaupanja.

Tab. 13.7

Število članov	Število gospodinjstev	N_o 7351	N'_o 20371
1	1313	6038	14333
2	1180	4918	9415
3	1255	3663	5753
4	1219	2444	3308
5	941	1503	1805
6	632	871	934
7	375	496	438
8	266	230	208
9	130	110	98
10	53	57	41
11	29	28	13
12	19	9	4
13	6	3	1
14	2	1	-
15	1	-	-

Število vseh gospodinjstev N = 368754, število gospodinjstev v vzorcu n pa 7351.

$\bar{x}'$ in s_x sc izračunani iz frekvenčne distribucije vzorca z metodo kumulativ.

$$N'_o = 36350$$

$$C_1 = \frac{N'_o}{N_o} = 2,7712 \qquad C_2 = \frac{N''_o}{N_o} = 4,9449$$

$$\bar{x}' = x_o + C_1 = 1 + 2,7712 = 3,7712$$

$$s_x^2 = 2C_2 - C_1 (C_1 - 1) = 2 \cdot 4,9449 - 2,7712 \cdot (2,7712 - 1) =$$
$$= 9,8898 - 4,9083 = 4,9815$$

$$s_x = \sqrt{4,9815} = 2,232$$

$$e'_{\bar{x}} = \frac{1,96 \cdot s_x}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}} = \frac{1,96 \cdot 2,232}{\sqrt{7350}} \sqrt{\frac{368754 - 7351}{368754}} =$$
$$= 1,96 \cdot 2,232 \cdot 0,015473 = 0,0505$$

$$\bar{x}' - e'_{\bar{x}} = 3,7712 - 0,0505 = 3,7207$$

$$\bar{x}' + e'_{\bar{x}} = 3,7712 + 0,0505 = 3,8217$$

Končni rezultat moremo podati v naslednji obliki: Prava aritmetična sredina leži z verjetnostjo 0,95 med 3,7207 in 3,8217. V simbolih moremo izraziti: $P\ (3{,}7207 < \bar{x} < 3{,}8217) = 0{,}95$

Dejanska aritmetična sredina 3,7745 članov na eno gospodinjstvo, ki smo jo izračunali kasneje iz popolnega gradiva leži res v intervalu med mejama zaupanja, kar je v skladu s teorijo.

V dosedanjih primerih smo standardno pogreško $\sigma_{\bar{x}}$ množili z 1,96, da smo dobili verjeten odklon. Pri mejah zaupanja, ki smo jih izračunali na tej osnovi, pričakujemo z verjetnostjo 0,95, da leži prava aritmetična sredina v mejah zaupanja. Ako ta faktor spremenimo, se spremeni tudi verjetnost, s katero leži aritmetična sredina v mejah zaupanja. Ako vzamemo n.pr. namesto 1,96, 1,65, je verjetnost, da prava aritmetična sredina leži znotraj tega intervala 0,90, ako pa vzamemo 2,59 pa je verjetnost 0,99. Najobičajneje v praksi iščemo rezultate z verjetnostjo 0,95. Zaradi tega pomnožimo standardno pogreško z 1,96. Ker pa moremo ta faktor spreminjati, ga splošno zaznamujemo z t tako, da je ocena verjetnega odklona

$$e'_{\bar{x}} = t\,\sigma'_{\bar{x}} \qquad (13.3)$$

V tem obrazcu je t količina, ki zavisi od tega, kakšno zanesljivost podatkov iščemo. Ako iščemo podatke, ki veljajo z večjo verjetnostjo, je t in zaradi tega tudi interval zanesljivosti večji.

133.2 Ocenjevanje skupne vsote nekega znaka iz vzorca. Poleg aritmetične sredine iščemo v statistiki zelo pogosto skupno vsoto nekega znaka za celo statistično maso. Iščemo skupno število prebivalstva, živine, površine itd.

Iz aritmetične sredine moremo izračunati skupno vsoto vseh vrednosti, ako jo pomnožimo s skupnim številom vseh enot $\sum x = N\bar{x}$.

Oceno za vsoto pa moremo dobiti, ako oceno aritmetične sredine pomnožimo s skupnim številom enot. Enako moremo dobiti oceno verjetnega odklona in mej zaupanja za vsoto na enostaven način tako, da vse te količine za aritmetično sredino pomnožimo s številom vseh enot v osnovni masi. Tako je:

$$\sum' x = N\bar{x}' \qquad (13.14),$$

$$e'_{\sum x} = N\,e'_{\bar{x}} \qquad (13.15)$$

$$N\bar{x}' - Ne'_{\bar{x}} < \sum x < N\bar{x} + Ne'_{\bar{x}} \qquad (13.16)$$

$\sum'_x$ pomeni oceno (') skupne vsote $[\sum]$

P Ako vzamemo kot primer število krav 50 gospodarstev, moremo skupno število krav v vseh 50 gospodarstvih iz prvega vzorca oceniti tako, da pomnožimo oceno sredine verjetnega odklona in mej zaupanja s 50.

$$\bar{x}' = 1{,}96 \qquad \sum_x' = 50 \cdot 1{,}96 = 98$$

$$e'_{\bar{x}} = 0{,}362 \qquad e'_{\sum x} = 50 \cdot 0{,}362 = 18$$

$$80 < \sum x < 116$$

Število govedi v vseh 50 gospodarstvih leži z verjetnostjo 0,95 med 80 in 116 glav. Dejansko število 113 realnično pade v meje intervala.

P Na ta način moremo iz števila članov 7351 gospodinjstev oceniti skupno število prebivalstva v LRS.

V prejšnjem odstavku smo ocenili aritmetično sredino števila članov na eno gospodinjstvo. Rezultate te ocene moremo uporabiti za ocenitev skupnega števila prebivalstva. Skupno število vseh gospodinjstev je po popisu 368754. N je torej 368.754.

$$\bar{x}' = 3{,}7712 \qquad e'_{\bar{x}} = 0{,}0505$$

$$\sum x = N\bar{x}' = 368\,754 \cdot 3{,}7712 = 1{,}390.645$$
$$e'_{\sum x} = Ne'_{\bar{x}} = 368\,754 \cdot 0{,}0505 = 18.627$$
$$\sum x - e_{\sum x} = 1{,}390.645 - 18627 = 1{,}372.018$$

$$\sum \bar{x} + e'_{\sum x} = 1{,}390.645 + 18627 = 1{,}409.272$$

Rezultate moremo koristiti na naslednji način: Ocena skupnega števila prebivalstva je 1,380.645. 95 % verjetnosti pa je, da pravi rezultat ni od ocene različen za več kot 18.627 oseb. 95 % verjetnosti je, da pravo število prebivalstva leži med 1,372.018 in 1,409.272. Dejansko se ocena števila prebivalstva razlikuje od pravega števila prebivalstva (1,391.873) samo za 1228 oseb.

D Dokaz zgornjega stavka je zelo enostaven. Napišemo v obliki neenačbe, da leži aritmetična sredina med obema mejama zaupanja

$$\bar{x}' - e'_{\bar{x}'} < \bar{x} < \bar{x}' + e'_{\bar{x}'}$$

To neenačbo pomnožimo s skupnim številom vseh enot osnovne mase N

Tako dobimo:

$$N\bar{x}' - N e'_{\bar{x}} < N\bar{x} < N\bar{x}' + N e'_{\bar{x}}$$

Ker je

$$\bar{x} = \frac{1}{N}\sum' x \quad \text{je} \quad N\bar{x} = \sum x.$$

Zaradi tega moremo pisati:

$$\sum\nolimits' x - e'_{\sum x} < \sum x < \sum\nolimits' x + e'_{\sum x} \qquad (13.17)$$

kjer je

$$\sum\nolimits' x = N\bar{x}' \quad \text{ocena vsote,}$$

$$e'_{\sum x} = N e'_{\bar{x}} \quad \text{pa ocena verjetnega odklona vsote.}$$

133.3 Ocenjevanje strukture iz vzorca. Ako hočemo izračunati, kakšen del od 50 gospodarstev ima eno kravo, moramo število gospodarstev, ki imajo eno kravo deliti s številom vseh gospodarstev (strukturni delež). Ako preštejemo gospodarstva z eno kravo, jih imamo v celoti 20. Strukturno število je torej:

$\frac{20}{50} = 0{,}40$. Ako to število pomnožimo s sto, dobimo delež izražen v odstotkih. Zaenkrat pa tega ne bomo storili, ampak bomo kot strukturni delež uporabljali decimalne ulomke. Ta ulomek zaznamujemo s splošnim znakom p.

Kot ocena strukturnega deleža p more služiti strukturni delež izračunan iz vzorca. Ta strukturni delež zaznamujemo zaradi tega, ker je ocena, s črtico,

$$p' = \frac{n_1}{n} \qquad (13.18)$$

Ako poiščemo strukturne deleže iz petih vzorcev po 25 enot, dobimo pet samostojnih ocen strukturnega deleža števila gospodarstev za eno kravo. Iz podatkov števila gospodarstev z eno kravo v vzorcu izračunamo strukturni delež tako, da število gospodarstev z eno kravo delimo s skupnim številom gospodarstev v vzorcu $p' = \frac{n_1}{n}$. Za prvi vzorec je $\frac{14}{25} = 0{,}5$

Tab. 13.8

Vzorec	Število ospo-rgtev z eno kravo v vzorcu	Strukturni delež
1	14	0,56
2	8	0,32
3	10	0,40
4	13	0,52
5	13	0,52

Ena ocena se točno sklada s pravim strukturnim deležem, druge pa se več ali manj odklanjajo od 0,40.

Enako kot ocene aritmetične sredine od prave aritmetične sredine, se tudi ocene strukturnih deležev distribuirajo okrog p v obliki normalne krivulje. Oceno standardne pogreške strukturnega deleža pa dobimo, ako v oceni standardne pogreške za aritmetično sredino s_x nadomestimo $s\sqrt{p'(1-p')}$. Ocena standardne pogreške strukturnega deleža je torej enaka:

$$\sigma'_{p'} = \sqrt{\frac{p'(1-p')}{n-1}} \sqrt{\frac{N-n}{N}} \qquad (13.19)$$

Ocena verjetnega odklona

$$e'_{p'} = t\,\sigma'_{p'} , \qquad (13.20)$$

Ocena mej zaupanja oziroma pravega strukturnega deleža pa je:

$$p' - e'_{p'} < p < p' + e'_{p'} . \qquad (13.21)$$

P Pri popisu prebivalstva v LRS 15. marca 1948 je bil iz celotnega števila prebivalcev (N = 1,391.873) vzet slučajen vzorec 27.847 oseb. Med drugim smo z vzorcem proučevali tudi stan. V vzorcu je bilo najdeno razmerje, kot ga kaže tabela 13.9. Kakšna je ocena za strukturni delež samskih in strukturni delež poročenih prebivalcev ?

Tab. 13.9

Stan	Štev. oseb v vzorcu
Samski	15827
Poročen	9978
Vdovec, razveden	2041
Skupno	27847

$$p'_1 = \frac{15827}{27847} = 0,568356$$

$$p'_2 = \frac{9978}{27847} = 0,358315$$

Ocena procenta samskih je 56,8 °/o, ocena procenta poročenih pa 35,8 °/o.

(4,7 ostalo)

Da ugotovimo natančnost podatkov, moramo izračunati še standardno pogreško, verjeten odklon in meje zaupanja. Standardno pogreško izračunamo po obrazcu:

$$\sigma'_{p_1} = \sqrt{p_1'(1 - p_1')} \cdot \frac{1}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}}$$

$$\sigma'_{p_2} = \sqrt{p_2'(1 - p_2')} \cdot \frac{1}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}}$$

$$\sqrt{p_1'(1 - p_1')} = \sqrt{0,568356 \cdot 0,431644} = 0,495305$$

$$\sqrt{p_2'(1 - p_2')} = \sqrt{0,358315 \cdot 0,641685} = 0,479505$$

$$\frac{1}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}} = \sqrt{\frac{1391873 - 27847}{27846 \cdot 1391873}} = \sqrt{\frac{1364026}{38758095558}} = 0,0059324$$

$$\sigma_{p_1}' = 0,495305 \cdot 0,0059324 = 0,0029383$$

$$\sigma_{p_2}' = 0,479505 \cdot 0,0059324 = 0,0028446$$

$$e'_{p_1} = 1,96 \cdot \sigma'_{p_1} = 1,96 \cdot 0,0029383 = 0,00575916$$

$$e'_{p_2} = 1,96 \cdot \sigma'_{p_2} = 1,96 \cdot 0,0028446 = 0,00557545$$

$$p_1' - e'_{p_1} < p_1 < p_1' + e'_{p_1}$$

$$0,568356 - 0,005759 < p_1 < 0,568356 + 0,005759$$

$$0,562597 < p_1 < 0,574125$$

$$p_2' - e'_{p_2} < p_2 < p_2' + e'_{p_2}$$

$$0,358315 - 0,005576 < p_2 < 0,358315 + 0,005576$$

$$0,352739 < p_2 < 0,363891$$

Z verjetnostjo 0,95 moremo pričakovati, da se odstotek samskih v celotnem prebivalstvu nahaja med 56,26 °/o do 57,41 °/o, odstotek poročenih pa med 35,27 °/o do 36,39 °/o.

Dejanski odstotek samskih po popisu znaša 57,1 °/o, poročenih pa 35,6 °/o. Oceni se zelo malo razlikujeta od teh podatkov in oba strukturna deleža padeta v interval med mejama zaupanja.

133,4 Ocenjevanje absolutnega števila enot dela statistične mase. V zgornjem primeru smo ocenjevali strukturne deleže samskih in poročenih. Iz strukturnega deleža moremo dobiti absolutno število samskih ali poročenih tako, da strukturni delež p pomnožimo s številom vseh enot N.

$$p_1 = \frac{N_1}{N}\,;\; N_1 = N\,p_1 \qquad (13.22)$$

Prav tako moremo dobiti oceno števila samskih ali poročenih, ako oceno strukturnega deleža pomnožimo s skupnim številom vseh oseb N.

V obliki obrazca moremo zgornji stavek napisati:

$$N_1' = Np_1' \,. \qquad (13.23)$$

Na isti način postopamo tudi pri oceni verjetnega odklona e_{N_1}'

$$e_{N_1}' = N \cdot e_{p_1}' \qquad (13.24)$$

Pravo število leži torej v intervalu

$$N_1' - e_{N_1}' < N_1 < N_1' + e_{N_1}' \qquad (13.25)$$

P Iz primera v prejšnjem odstavku moremo iz ocene strukturnih deležev hitro izračunati ocene skupnega števila samskih in poročenih oseb v LRS po popisu 1948. Število vsega prebivalstva je N = 1,391.873. Izvedimo izračun za samske osebe.

$$p_1' = 0{,}568356 \qquad e_{p_1}' = 0{,}005575$$

$$N_1' = Np_1' = 1{,}391873 \cdot 0{,}568356 = 791079$$

$$e_{N_1}' = Ne_{p_1}' = 1{,}391873 \cdot 0{,}005575 = 8016$$

$$N_1' = e_{N_1}' = 791079 - 8016 = 783063$$

$$N_1' + e_{N_1}' = 791079 + 8016 = 799095$$

$$783063 < N_1 < 799095$$

Z verjetnostjo 0,95 pričakujemo, da je pravo število samskih oseb po popisu med 783063 in 799095. Dejanski rezultat 796159 se nahaja res v tem intervalu.

133.5 Odvisnost verjetnega odklona $e_{\bar{x}}$ od elementov vzorčenja.

Ako pogledamo obrazec za verjeten odklon ocene aritmetične sredine:

$$e_{\bar{x}} = \frac{t\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad (13.26)$$

vidimo, da verjeten odklon, ki podaja natančnost ocene, zavisi od naslednjih elementov:

1. Verjeten odklon je premosorazmeren s standardno deviacijo σ_x znaka v osnovni masi. Ker standardna deviacija meri variabilnost pojava, vidimo, da je ocena aritmetične sredine tem bolj natančna, čim bolj je variabilna statistična masa, katero proučujemo. Z drugimi besedami moremo reči, da bomo dobili najbolj natančne ocene za homogene statistične mase.

2. Verjeten odklon je odvisen tudi od števila enot v vzorcu n. Odvisnost je obratno sorazmerna in sicer ne neposredno z n, temveč z $\sqrt{n}$. Iz tega vidimo, da je verjeten odklon tem manjši, čim večja je velikost vzorca. Verjeten odklon pa se zmanjša samo dvakrat, ako je vzorec štirikrat večji, trikrat, ako je vzorec devetkrat večji itd.

3. Ako pogledamo še $\sqrt{\frac{N-n}{N-1}}$, moremo ta izraz pisati v obliki $\sqrt{1 - \frac{n-1}{N-1}}$, kar je pri razmeroma velikem n in N približno enako $\sqrt{1 - \frac{n}{N}}$.

Ker je $\frac{n}{N}$ delež enot vzorca od celotne mase, vidimo, da je verjeten odklon in z njim natančnost ocene odvisna od tega, kakšen del celotne mase smo vključili v vzorec. Ker vzorec predstavlja običajno samo par odstotkov od celote, je $\frac{n}{N}$ majhna količina, $\sqrt{1 - \frac{n}{N}}$ pa vrednost, ki je zelo blizu 1. Zaradi tega ne napravimo velike napake, ako pri izračunavanju verjetnega odklona za vzorce, ki predstavljajo majhen del celote, ta faktor izpustimo in postavimo verjeten odklon enak:

$$e_{\bar{x}} = \frac{t\sigma_x}{\sqrt{n}} \qquad (13.27)$$

4. Verjeten odklon je odvisen nadalje tudi od faktorja t. Ta faktor je določen z zanesljivostjo, ki jo zahtevamo za podatke. Kot vemo, običajno postavimo t enak 1,96.

N in σ_x sta dana z osnovno maso, t pa je določen z zanesljivostjo, ki jo zahtevamo. Edini element, s katerim moremo po želji vplivati na verjeten odklon, je velikost vzorca n. Z večanjem vzorca manjšamo verjeten odklon in večamo natančnost ocen, z manjšim vzorcem pa se zmanjša tudi natančnost.

133.6 Vprašanja vzorčenja v zvezi z velikostjo vzorca in natančnostjo ocen. V praksi vzorčenja se postavljata v glavnem dve nalogi. Ena izmed njih je, da iz danega vzorca izračunamo oceno in verjeten odklon, oziroma natančnost. Druga naloga pa je, da izračunamo, pri kako velikem vzorcu bomo dosegli željeno natančnost.

.61 Natančnost pri dani velikosti vzorca. To nalogo smo v prejšnjih odstavkih že rešili za ocene štirih vrst statističnih količin: aritmetično sredino, vsoto vrednosti znaka, strukturni delež in obseg delne mase. Vendar smo ocenjevali samo absolutno natančnost, oziroma verjeten odklon. Ako absoluten verjeten odklon aritmetične sredine delimo z aritmetično sredino, dobimo natančnost merjeno relativno. Relativna mera natančnosti je prav tako važna kot absolutna. Verjetne odklone ostalih količin moramo seveda deliti s pripadajočimi vrednostmi n.pr. verjeten odklon vsote z vsoto, verjeten odklon strukturnega deleža s strukturnim deležem itd. V razpredelnici 13.28 ponavljamo izračunavanje ocen in absolutnih in relativnih verjetnih odklonov za vse štiri količine, ki smo jih ocenjevali v prejšnjih odstavkih.

.62 Velikost vzorca pri dani natančnosti.

Velikokrat se pojavlja v praksi vzorčenja vprašanje, kako velik moramo vzeti vzorec, da dosežemo vnaprej dano natančnost ?Naloga je v tem primeru ravno obratna od naloge v prejšnjem odstavku, v katerem smo pri dani velikosti izračunali natančnost. V sedanjem primeru iščemo število enot v vzorcu (n), ako imamo dane ali ocenjene

σ_x, t, N in e ali E. e je absolutna, E pa relativna natančnost

Za oceno aritmetične sredine moremo iz obrazca:

$$e_{\bar{x}} = \frac{t\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$ dobiti n, ako ta obrazec razrešimo na

n. Najprej enačbo obrnimo in pomnožimo z $t\sigma_x$

$$\frac{t\sigma_x}{e_{\bar{x}}} = \sqrt{n \frac{N-1}{N-n}}$$, količino $\frac{t\sigma_x}{e_{\bar{x}}}$ zaznamujemo s C in enačbo kvadrirajmo

Tab. 13.28

Proučevana količina R		Verjeten odklon			
		Absolutno		Relativno	
	Prava R vrednost / Ocena R'	Prava e vrednost	Ocena e'	Prava E vrednost	Ocena E'
Aritmetična sredina	$\bar{x}$ / $\bar{x}' = \frac{1}{n} S x$	$\sigma_x A$	$s_x A'$	$\frac{\sigma_x}{\bar{x}} A = KV_x \cdot A$	$\frac{s'_x}{\bar{x}'} A = KV'_x \cdot A'$
Vsota vrednosti znaka	$\sum x$ / $\sum x = \frac{N}{n} S x$	$N \sigma_x A$	$N s_x A'$		
Strukturni delež	$p = \frac{N_1}{N}$ / $p' = \frac{n_1}{n}$	$\sqrt{p(1-p)}A$	$\sqrt{p'(1-p')}A'$	$\sqrt{\frac{1-p}{p}} A$	$\sqrt{\frac{1-p'}{p'}} A'$
Obseg delne mase	N_1 / $N'_1 = \frac{N}{n} n_1$	$N\sqrt{p(1-p)}A$	$N\sqrt{p'(1-p')}A'$		
	Prava vrednost leži v intervalu $R' - e'_R < R < R' + e'_R$	$A = \frac{1}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ Ako je N velik je $A = \frac{1}{\sqrt{n}}$		$A' = \frac{1}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}}$ Ako je N velik je $A' = \frac{1}{\sqrt{n-1}}$	
		$\sigma_x^2 = \frac{1}{N} \Sigma (x-\bar{x})^2$		$s_x^2 = \frac{1}{n} S (x-\bar{x}')^2$	

Dobimo

$$C^2 \cdot (N - n) = n(N - 1)$$

$$C^2N - C^2n = nN - n$$

Vse člene, ki vsebujejo n, denimo na desno stran:

$C^2N = n(N - 1 + C^2)$. Iz te enačbe je n enak:

$$n = \frac{NC^2}{(N - 1) + C^2} \qquad (13.29)$$

pri čemer je $$C = \frac{t\,\sigma_x}{e_{\bar{x}}} \qquad (13.30)$$

Kadar je statistična masa zelo velika (N je zelo velik), v imenovalcu C^2 napram N ne pride v poštev, v tem primeru je približna vrednost

$$n = C^2 \qquad (13.31)$$

ker se N v številcu in imenovalcu krajša.

P Kako velik vzorec bi morali vzeti, ako bi v primeru ocene aritmetične sredine števila članov na eno gospodinjstvo v popisu prebivalstva v letu 1948 iskali rezultat, ki se od pravega povprečja z verjetnostjo 0,95 ne bi razlikoval za več kot 0,1 člana na gospodinjstvo. Druge elemente vzemimo iz primera odstavka 133, 132

$$N = 368754, \quad \sigma_x = 2{,}232, \quad t = 1{,}96.$$

Verjeten odklon $e_{\bar{x}}$, ki je predpisan, je enak $e_{\bar{x}} = 0{,}1$ člana

$$C = \frac{t \cdot \sigma_x}{e_{\bar{x}}} = \frac{1{,}96 \cdot 2{,}232}{0{,}1} = 43{,}747$$

$$C^2 = 1913{,}80$$

Ker je N zelo velik, moremo že kar C^2 smatrati kot približno število enot ($n = 1914$), točna vrednost pa je enaka:

$$n = \frac{N \cdot C^2}{N - 1 + C^2} = \frac{368\,754 \cdot 1913{,}80}{368\,753 + 1913{,}80} = \frac{705{,}721405{,}20}{369666{,}80} = 1909$$

Vidimo, da je v našem primeru razlika 5 enot malenkostna.

Na enak način moremo iz obrazca

$$n = \frac{N\,C^2}{N-1+C^2}$$

izračunati število enot v vzorcu tudi za ostale količine, ki smo se jih naučili ocenjevati. C se za vsako vrsto ocenjenih količin izračuna po spodnji razpredelnici:

Razpredelnica vrednosti C

Ocenjena količina	Natančnost dana	
	absolutno e	relativno E
$\bar{x}$	$C = \frac{t\,\sigma_x}{e_{\bar{x}}}$	$C = \frac{t \cdot KV}{E}$
$\sum x$	$C = \frac{t\,\sigma_x N}{e_{\sum x}}$	
p	$C = \frac{t\,\sqrt{p(1-p)}}{e_p}$	$C = \frac{t}{E}\sqrt{\frac{1-p}{p}}$
N_1	$C = \frac{t\,\sqrt{N_1(N-N_1)}}{e_{N_1}}$	

$$n = \frac{C^2 N \quad \sigma}{(N-1)+C^2} \qquad (13.31)$$

Ako je N velik, je $n = C^2$

P Oceniti je treba skupno število krav v LRS. Približna vrednost koeficienta variacije za število krav na eno gospodarstvo je (KV = 0,75). Dopusten je 5 °/o odklon ocene od prave vrednosti (E = 0,05). Zanesljivost podatkov mora biti 95 °/o. (t = 1,96). Skupno število gospodarstev je N = 211.300. Kako velik vzorec moramo vzeti, da bomo zadostili gornjim pogojem ?

Narodna in univerzitetna knjižnica Ljubljana

Količina C za oceno vsote dane z relativno natančnostjo je enaka:

$$C = \frac{t \cdot KV}{E} = \frac{2,075}{0,05} = \frac{1.5}{0,05} = 30$$

$$C^2 = 900$$

$$n = \frac{N \cdot C^2}{N-1+C^2} = \frac{211300 \cdot 900}{211299 + 900} = \frac{180170000}{212199} = 849$$

V vzorec bomo vzeli 849 gospodarstev.

134. VRSTE VZORČENJ

134.1 Enostavno vzorčenje. Najbolj preprosto izmed vseh vzorčenj je enostavno vzorčenje. Enostavno vzorčenje obstoji v tem, da iz celotne mase, ne glede na njen sestav, izberemo na slučajen način vzorec. Za izvedbo enostavnega vzorca zaradi tega ni potrebno predhodno poznavanje sestava mase. Vsi dosedanji načini in primeri, ki smo jih navedli, so se nanašali na enostaven vzorec. Zaradi njegove enostavnosti ga v praksi zelo pogosto uporabljamo.

134.2 Stratificirano vzorčenje ali vzorčenje v legah. Natančnost ocene v mnogih primerih zelo zvečamo, ako se poslužimo tako imenovanega stratificiranega vzorca. Beseda stratificirano vzorčenje je angleška tujka in pomeni vzorčenje v plasteh ali legah (stratum = plast). Smisel stratificiranega vzorčenja obstoji v naslednjem: Verjeten odklon, ki podaja natančnost ocene, je odvisen v vseh primerih od standardne deviacije pojava, ki ga proučujemo z vzorcem. Ker je standardna deviacija v nehomogenih statističnih masah zelo velika, je natančnost ocene majhna, ako proučujemo nehomogene pojave z enostavnim vzorčenjem. V takih primerih dostikrat uspemo zboljšati ocene, ako pred vzorčenjem razdelimo nehomogeno statistično maso v homogene dele. Te homogene dele imenujemo lege ali s tujim izrazom stratume. V vsakem stratumu izvedemo samostojen enostaven vzorec, iz delnih ocen za posamezne stratume pa sestavimo skupno oceno. Ako je stratifikacija uspešna, je natančnost ocene, izračunana na osnovi stratificiranega vzorca večja kot natančnost ocene, izračunana na osnovi enostavnega vzorčenja. Seveda je uvedba stratificiranega vzorca bolj komplicirana, ker moramo predhodno izvesti analizo pojava in poznati sestav osnovne mase. Poleg tega moramo celotno maso pred vzorčenjem razdeliti v homogene dele.

Ako ocenjujemo s stratificiranim vzorcem aritmetično sredino nekega pojava, je ocena celotne aritmetične sredine tehtana aritmetična sredina delnih ocen. Kot ponder služi število enot celotne mase v pripadajočih legah. Ako vzamemo, da je masa razdeljena v tri stratume, ki jih označimo

z 1, 2, 3, je ocena skupne sredine enaka:

$$\bar{x}' = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2 + N_3 \bar{x}_3}{N_1 + N_2 + N_3} \qquad (13.32)$$

Iz ocen verjetnih odklonov aritmetičnih sredin po legah, pa moremo izračunati oceno verjetnega odklona po naslednjem obrazcu:

$$e_{\bar{x}}' = \sqrt{\frac{N_1^2 e_{\bar{x}_1}'^2 + N_2^2 e_{\bar{x}_2}'^2 + N_3^2 e_{\bar{x}_3}'^2}{(N_1 + N_2 + N_3)^2}} \qquad (13.33)$$

P Da bomo uvideli učinkovitost stratifikacije, vzemimo shematičen primer. V neki republiki je 100000 kmetijskih gospodarstev. Od teh je 50000 malih, 30000 srednjih in 20000 velikih gospodarstev. Ocenjujemo povprečno površino njiv na eno gospodarstvo.

Ko smo izvedli enostavno vzorčenje iz 2000 enot, smo dobili naslednje rezultate:

$$\bar{x}' = 1{,}92 \qquad s_x^2 = 3{,}61 \qquad s_x = 1{,}9$$

Verjeten odklon $e_{\bar{x}}' = \frac{t \cdot s_x}{\sqrt{n-1}} \sqrt{1 - \frac{n}{N}} = \frac{1{,}96 \cdot 1{,}9}{\sqrt{2000-1}} \sqrt{1 - \frac{2}{100}} =$

$$= \frac{1{,}96 \cdot 1{,}9 \sqrt{0{,}98}}{\sqrt{1999}} = 0{,}0824$$

Pri stratificiranem vzorčenju smo vzeli v vzorec isto skupno število enot 2000. Enote so v stratumu izbrane proporcionalno, sorazmerno. To pomeni, da je v vsakem stratumu vzet isti odstotek gospodarstev. Ker je 2000 dva odstotka od 100000, vzamemo po posameznih stratumih vzorec po 2 % gospodarstev. Ta iznese 1000 gospodarstev pri malih, 600 pri srednjih in 400 pri velikih.

Za stratificirano vzorčenje potrebni podatki in računi so naslednji:

Stratum	N_k	n_k	$\bar{x}'_k$	s_x^2	$e_{\bar{x}}^2 = \frac{4s_x^2}{n_k}$	$\frac{N_k}{N}$	$\left(\frac{N_k}{N}\right)^2$	$\frac{N_k}{N}\bar{x}'_k$	$\left(\frac{N_k}{N}\right)^2 e'^2_{\bar{x}_k}$
	(1)	(2)	(3)	(4)	(5) $\frac{4\cdot(4)}{(2)}$	(6)	(7)	(8)	(9)=(7)·(5)
Mali kmetje	50000	1000	0,5	0,25	0,001	0,5	0,25	0,25	0,00025
Srednji kmetje	30000	600	2,0	1,00	0,0067	0,3	0,09	0,60	0,00060
Veliki kmetje	20000	400	5,0	2,00	0,0200	0,2	0,04	1,0	0,00080
Skupno	100000	2000	-	-	-	1,0	=	1,85	0,00165

$\bar{x}' = 1.85$; $e^{2'}_{\bar{x}} = 0,00165$ $\quad e'_{\bar{x}} = 0,041$

Podatki vnešeni v koloni (1) in (2) tabele slede iz poznavanja zgrodja vzorčenja. Kot vidimo, je v vsakem stratumu vzeto v vzorec 2 °/o enot $\frac{1000}{50000} = \frac{600}{30000} = \frac{400}{20000} = 0,02$. Ker je število vzorčnih enot posameznem stratumu proporcionalno številu enot osnovne mase govorimo v tem primeru o proporcionalnem - sorazmernem stratificiranju.

$\bar{x}_k$ in s_k^2 so količine, ocenjene in izračunane na osnovi vzorčnih enot v posameznih stratumih. Iz teh podatkov moremo izračunati kvadrate verjetnih odklonov v posameznih stratumih. Da poenostavimo račun, je vzeta namesto njene približna vrednost, t ki je točno 1,96 postavimo približno enak 2, faktor $\sqrt{1 - \frac{n}{N}}$ postavimo enak 1, ker je v našem primeru $\sqrt{0,98} = 0,99$. Količino $n_k - 1$ pa postavimo enako n_k ker so te količine razmeroma velika števila. N je enak vsoti vseh N_k. Iz števila enot so v koloni 6 izračunani ponderi za izračun skupne ocene aritmetične sredine, v koloni 7 pa količine potrebne za izračunavanje verjetnega odklona za skupno aritmetično sredino. Kot kaže zgornja shema, izračunamo količine v koloni 8. Vsota je enaka oceni skupne aritmetične sredine. Vsota količin v koloni 9 pa je enaka oceni kvadrata verjetnega odklona skupne aritmetične sredine. Kvadratni koren iz vsote te

kolone pa je enak oceni verjetnega odklona skupne aritmetične sredine. Ako primerjamo verjeten odklon ocene, ki smo jo dobili s stratifikacijo (0,085) vidimo, da je natančnost stratificirane ocene znatno večja kot enostavna.

Eno izmed važnih vprašanj stratificiranega vzorcenje, ali je proporcionalna razmestitev vzorčnih enot po stratumih najboljša, ali moremo morda pri drugačni razmestitvi vzorčnih enot doseči večjo natančnost. Pri analizi proporcionalnega vzorčenja se vrine namreč misel, da smo vzeli, ako vršimo proporcionalno stratifikacijo, pri stratumih z majhno standardno deviacijo (bolj homogeni del) preveč, pri stratumih z veliko standardno deviacijo (bolj heterogeni del) pa premalo enot. Zaradi tega slutimo, da bi morali tudi v mejah iste velikosti skupnega vzorca (2000 enot), doseči z boljšo razmestitvijo vzorčnih enot večjo natančnost rezultatov.

Res se izkaže, da najboljša razmestitev enot ni proporcionalna razmestitev, temveč razmestitev, pri kateri je število enot v posameznih stratumih sorazmerno produktu števila enot v tem stratumu in standardni deviaciji

$$(n_k = N_k \sigma_k a) \qquad (13.34)$$

Ker σ_x običajno nimamo izračunane, si pomagamo z ocenami. V našem primeru vzemimo kot oceno σ_x kar količine s_x. Potrebno število enot v posameznih stratumih izračunamo takole:

$$n_1 = N_1 s_{x_2} a = 50000 \cdot 0,5\ C = 25000\ C$$

$$n_2 = N_2 s_{x_2} a = 30000 \cdot 1,0\ C = 30000\ C$$

$$n_3 = N_3 s_{x_3} a = 20000 \cdot 1,41\ C = 28200\ C$$

Ne poznamo še množitelja a. Ker pa vemo, da je $n_1 + n_2 + n_3 = 2000$, moremo postaviti $2000 = (25000 + 30000 + 28200)\ a$. Iz te enačbe moremo izračunati a,

$$a = \frac{2000}{83200} = 0,024$$

Iz tega dobimo, da moramo vzeti v prvem stratumu $n_1 = 25000 \cdot 0,024 = 600$ enot, v drugem $n_2 = 30000 \cdot 0,024 = 720$ enot, v tretjem pa $n_3 = 28200 \cdot 0,024 = 680$ enot, skupno torej 2000. Ostali izračun je popolnoma enak kot zgornji. Vidimo, da vzamemo v prvem stratumu znatno manj enot, kot pri proporcionalnem vzorcu, ker je standardna deviacija v majhna (0,5) v tretjem stratumu pa več, ker je standardna deviacija velika (1,41).

Stratum	N_k	n_k	$\bar{x}'_k$	$s^2_{x_k}$	$e'^2_k = \frac{4s^2_x}{n_k}$	$\frac{N_k}{N}$	$\left(\frac{N_k}{N}\right)^2$	$\frac{N_k}{N}\bar{x}'_k$	$\left(\frac{N_k}{N}\right)^2 e'^2_{\bar{x}}$
	(1)	(2)	(3)	(4)	(5) $\frac{4\,(4)}{(2)}$	(6)	(7)	(8) (3) (6)	(9) (5) (7)
Mali kmetje	50000	600	0,5	0,25	0,00167	0,5	0,25	0,25	0,00042
Srednji kmetje	30000	720	2,0	1,00	0,00555	0,3	0,09	0,60	0,00050
Veliki kmetje	20000	680	5,0	2,00	0,01176	0,2	0,04	1,00	0,00047
Skupno	100000	2000	-	-	-	1,0		1,85	0,00139

$\bar{x}' = 1,85; \quad e'^2_{\bar{x}} = 0,00139 \quad e'_{\bar{x}} = 0,037$

Ako pregledamo vse tri vrste vzorčenja, vidimo, da je najmanj natančna ocena po metodi enostavnega vzorčenja ($e'_{\bar{x}} = 0,085$), znatno natančnejša ocena s proporcionalnim vzorčenjem ($e'_{\bar{x}} = 0,041$), najnatančnejši rezultat pa da stratifikacija z najboljšo razmestitvijo ($e'_{\bar{x}} = 0,037$)

Na podoben način ocenjujemo s stratificiranim vzorcem tudi ostale količine. Ocene in kvadrate verjetnih odklonov teh ocen izračunavamo kot kaže razpredelnica:

Ocenj. količina	Stratificirani vzorec	
	Ocena	e^2
$\bar{x}'$	$\sum_k \frac{N_k}{N} \bar{x}'_k$	$\sum_k \left(\frac{N_k}{N}\right)^2 e^2_{\bar{x}_k}$
x	$\sum_k N_k \bar{x}'_k$	$\sum_k N_k^2 e^2_{\bar{x}_k}$
p'	$\sum_k \frac{N_k}{N} p'_k$	$\sum_k \left(\frac{N_k}{N}\right)^2 e^2_{p'_k}$
N'_1	$\sum_k N'_{1k}$	$\sum_k N^2_{1k} e^2_{p'_k}$

(13.35)

$\sum$ pomeni vsoto količin, ki stoje za tem znakom za vse stratume. Ostale količine so znane. Črke k pomenijo, da gre za vrednost v stratumih.

134.3 Vzorčenje v več stopnjah. Vzemimo, da moramo izvesti v LRS vzorec o številu živine. Enota pri tem vzorčenju je kmetijsko gospodarstvo. Ako na slučajen način izberemo potrebno število gospodarstev, bodo ta gospodarstva raztresena po vsej Sloveniji. Izbrana gospodarstva se bodo nahajala skoro v vseh krajevnih ljudskih odborih. Tak vzorec pride razmeroma zelo drag in dolgotrajen. Upoštevati moramo namreč to, da je pri popisovanju vzorčnih enot najzamudnejša in najdražja pot, da pridemo do enote, ki jo moramo popisati. To pot moremo običajno opraviti za eno samo enoto, ker je druga enota od te zopet zelo oddaljena. To težkočo odpravi vzorčenje v več stopnjah, ki bi ga v našem primeru izvedli na sledeč način: Ker bi radi dosegli, da bi izbrana gospodarstva ne bila tako raztresena, izvedemo dvojno vzorčenje. Najprej vzamemo kot statistično maso skupnost vseh krajevnih odborov in jih iz njih izberemo na slučajen način določeno število n.pr. 100. Posamezna gospodarstva v nadaljnjem ne izbiramo iz vseh gospodarstev oz. krajevnih ljudskih odborov, temveč samo iz sto na slučajen način izbranih.

Na ta način se vsa izbrana gospodarstva nahajajo samo v sto krajevnih odborih, ne pa v vseh. Čeprav moramo pri vzorčenju te vrste vzeti v vzorec večje število enot, kot pri enostavnem vzorčenju, da dosežemo isto natančnost, vendar ga dostikrat uporabljamo zaradi prednosti, ker enote pri vzorčenju v več stopnjah niso raztresene po teritoriju, temveč grupirane na manjšem številu točk. S tem se stroški in čas znatno skrčita.

Ker smo v našem primeru izbrali najprej KLO-je, v teh pa gospodarstva, imenujemo tak vzorec, vzorec v dveh stopnjah. Dostikrat pa vršimo tudi vzorce v več stopnjah. Ako bi ugotavljali n.pr. hektarski donos pšenice, bi mogli izbrati v prvi stopnji KLO-je. V njih bi izbrali v drugo stopnjo gospodarstva. V tretji stopnji bi izbirali v izbranih gospodarstvih parcele, posejane s pšenico, v celoti pa na izbranih parcelah površine po m2.

135. OBČUTLJIVOST VZORČENJA ZA RAZNE NAPAKE

Ker z vzorcem sklepamo na razmere v celotni masi, je jasno, da se vsaka napaka, ki jo napravimo v vzorcu, v pomnoženi obliki očituje na rezultatu. Ako vršimo enodstotni vzorec, ima napaka, ki jo napravimo pri vzorčni masi na eni sami enoti isti učinek da sklep na celoto, kot bi pri popisu napravili isto napako pri sto enotah.

Pri obravnavanju napak, ki se morejo pojaviti pri vzorčenju, pa moramo

takoj povdariti, da med napake ne štejemo standardne pogreške oziroma verjetnega odklona. Ta slučajen odklon ne izvira iz napačnega izvršenja vzorca ali popisa vzorčnih enot, temveč iz same narave vzorčenja, ker je osnovan na opazovanju dela mase. Slučajen odklon ne moremo šteti kot napako v pravem smislu besede tudi zaradi tega, ker ga moremo od primera do primera oceniti in ohraniti v predpisanih mejah.

Napake pri vzorčenju moremo razdeliti v napake v teh fazah:

1) Pri sestavi ogrodja in izbiri enot,

2) Pri zbiranju podatkov za zbrane enote,

3) Pri obdelavi podatkov.

Pri sestavi ogrodja vzorčenja največkrat grešimo zaradi tega, ker imamo na razpolago pomanjkljivo gradivo. Ogrodje v teh primerih ne moremo sestaviti tako popolno, da bi izbor enot bil popolnoma slučajen. Vsaka pristranost pri izboru enot pa ima za posledico napačne rezultate. Sestava ogrodja in izbor enot morata težiti za tem, da je čim bolj izpolnjena osnova vzorčenja - slučajnost izbora.

Nejasna opredolitev enote opazovanja in posameznih vprašanj more pri vzorčenju imeti še hujše posledice kot pri popolnem opazovanju.

Problemi v zvezi z zbiranjem podatkov o vzorčnih enotah na terenu so povsem svojstveni in različni od zbiranja podatkov pri popolnem popisu. Zato je potrebna prav posebno skrbna izbira in poučitev oseb, ki bodo vršile popis vzorčnih enot. Prednost je seveda v tem, da more biti izbira popisovalcev bolj skrbna, kot pri popolnem popisu, ker je število oseb znatno manjše. Osebe, ki sodelujejo pri zbiranju podatkov vzorčenja, morajo biti v osnovah poučene o metodi vzorčenja, da jim moremo opravičiti posebnosti zbiranja podatkov za vzorec. Pri marsikaterem navodilu bi mogli v primeru, da popisovalec ne ve za kaj gre, imeti občutek nepotrebnosti in s tem že nagnjenje, da navodila ne izpolnjujejo. Osebi, ki vrši vzorčno popisovanje, je treba raztolmačiti, zakaj je potrebno, da popišemo ravno določene enote in ne drugih, katere bi mogel popisati z manjšim trudom. Zato je treba popisovalce pri vzorčenju dobesedno vzgojiti. V dosti primerih teži za tem, da popiše enote, ki so mu lažje dostopne. Ako osebe, ki jo mora popisati, ne dobi pri prvem obisku, dostikrat popiše drugo osebo, ki je slučajno doma in ne ponovi obiska. S tem vnese v vzorec pristranost, ker so doma večinoma gospodarsko nedelavne osebe (stari, otroci, gospodinje).

Pri vzorčni kontroli popisa živine v LRS leta 1950 je rezultate v nekaterih primerih oziroma predelih povsem izkrivila naslednja sistematična napaka. Kontrolorji so niso držali točno danih navodil in se jim je zdelo nesmiselno kontrolirati gospodarstva, za katera so vedeli, da so lastniki poštenjaki in so kontrolirali gospodarstva onih, za katere so mislili, da so

živine pujili. S tem je bil ocenjeni odstotek utaje seveda prevelik. Drugo leto, ko se je pazilo, da so bila kontrolirana res predpisana gospodarstva, so bili tudi podatki realni.

Napake v obdelavi morejo biti računske narave ali pa obstojajo v tem, da napačno uporabljamo metode in obrazce ocenjevanja. Zaradi tega morajo sodelovati v tej fazi enako kot pri načrtovanju vzorca statistiki - strokovnjaki za vzorčenje.

Kot smo videli, je samo vzorčenje razmeroma zapletena metoda opazovanja. Kdor jo hoče uporabljati, mora dobro poznati osnove teorije in biti obenem tudi statistik - praktik. Skoro vsako vzorčenje namreč zahteva posebne prijeme in predstavlja probleme, ki jih moramo od primera do primera reševati drugače.

Lažja je uporaba in razumevanje statističnih podatkov, ki jih dobimo z vzorčenjem. Vendar morajo tudi oni, ki te rezultate uporablja, poznati osnove vzorčenja, da ve, kaj predstavljajo vzorčne ocene, kakšen pomen imajo meje zaupanja in standardne pogreške. Rezultati vzorčenja morajo biti namreč vedno "zavarovani" z verjetnim odklonom, oziroma mejami zaupanja. Nepoznavanje osnov vzorčenja ima večkrat za posledico nezaupanje v njegove rezultate.

136. POVZETEK

Od vseh delnih opazovanj smatramo kot najboljšo metodo vzorčenja. Metoda vzorčenja je [illegible] na objektivnih znanstvenih temeljih verjetnostnega računa. Področja uporabe metode vzorčenja so različne znanosti kot: biologija, antropometrija, medicina, agronomija itd., kjer so raziskave možne edino s pomočjo metod, zasnovanih na vzorčenju. V socialno - ekonomski statistiki uporabljamo vzorčenje kot nadomestilo ali dopolnilo pri popisih, pri obdelavi pa za sestavo predhodnih rezultatov ali kot dopolnilno obdelavo. Z vzorčenjem pa moremo tudi kontrolirati posamezne faze statističnega opazovanja, n.pr. kvaliteto dela na terenu, ali obdelave. Prednosti vzorčenja pred popolnim opazovanjem so številne. Število enot opazovanja pri vzorčenju je znatno manjše kot pri popolnem opazovanju. Zaradi tega se znatno skrajša čas opazovanja, dvigne kakovost in zmanjšajo stroški. Vendar vzorčenja ne moremo uporabiti, ako potrebujemo podatke na majhne teritorialne enote ali pa veliko število znakov.

V teoriji vzorčenja poznamo tri vrste statističnih mas. Statistično maso, o kateri hočemo sestaviti pregled, imenujemo osnovno statistično maso. S slučajnim izborom izbrane enote tvorijo vzorčno maso, ali kratko vzorec. Ker moremo iz celokupne statistične mase, ki je v splošnem zelo veliko

število različnih vzorcev, moremo smatrati vsak posamezen možen vzorec kot enoto statistične mase, ki jo sestavljajo vsi možni vzorci. Po številu enot ločimo končne in neskončne osnovne mase. Vzorce pa delimo po številu enot na male in velike. Osnova vzorčenja je v tem, da se enote, ki sestavljajo vzorec iz osnovne statistične mase, izbrane na slučajen način. Slučajnost vzorca je predpogoj za reprezentativnost vzorca. Gradivo oziroma okvir vzorčenja sestavljajo spiski in seznami vseh enot in je temelj za izvršitev slučajnega izbora. Načinov izbiranja enot slučajnega vzorca imamo več. Poslužovati se moremo običajne žare, ali tablice slučajnih števil. Pod določenimi pogoji moremo uporabiti tudi sistematičen izbor, ki obstoji v tem, da izberemo v vzorec tudi sistematičen izbor, ki obstoji v tem, da izberemo v vzorec enoto, ki pride na vrsto na vsako določeno število.

Z vzorcem moremo oceniti najrazličnejše podatke osnovne statistične mase. Naš program pa vsebuje ocenjevanje naslednjih količin: aritmetične sredine, vsote enega znaka, strukturnega deleža in števila enot dolne mase. Ocena, dobljena s pomočjo vzorčenja, vsebuje naslednje sestavne dele:

a) oceno mere R';

b) standardno pogreško $\sigma_{R'}$, ki je dejansko standardna deviacija mase vseh možnih vzorcev;

c) verjeten odklon $r_{R'} = t\,\sigma_{R'}$, ki pove, da se prava vrednost z določeno verjetnostjo odklanja od ocene za manj kot je verjetni odklon.

d) meje zaupanja $R' - e_{R'}$ in $R' + e_{R'}$, ki določajo interval, v katerem se z določeno verjetnostjo nahaja prava vrednost R.

$$R' - e_{R'} < R < R' + e_{R'}$$

Z vzorcem moremo količine ocenjevati s standardne pogreške, verjeten odklon in meje zaupanja samo oceniti. Zaradi tega se poslužujemo samo ocen teh količin.

Verjeten odklon aritmetične sredine

$$e_{\bar{x}} = \frac{t\,\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

je premosorazmeren σ, ki izraža variabilnost znaka, in t, ki je odvisen od tega, kako zanesljive podatke hočemo imeti. Edini sestavni del, na katerega moremo vplivati in od katerega v bistveni meri je odvisen verjeten odklon je velikost vzorca, ki je izražena s številom enot n. Čim večji je vzorec, tem manjši je verjeten odklon. Naloga, ki se pri vzorčenju postavlja more biti dvojna. Prvič moremo pri predpisani velikosti vzorca iskati verjeten odklon, drugič pa moremo pri predpisanem odklonu iskati velikost vzorca, pri katerem bomo to predpisano natančnost dosegli.

Poznamo več vrst vzorčenj. Najenostavnejše je enostavno vzorčenje, pri katerem izberemo enote neglede na njen sestav. Pri stratificiranem vzorčenju razdelimo osnovno maso v homogene dele in izvršimo v vsakem homogenem delu samostojno enostavno vzorčenje. Ocena celotne mase je aritmetična sredina delnih ocen. O vzorčenju v več stopnjah govorimo, kadar iz celotnega števila KLOjev izberemo na slučajen način nekaj KLOjev, v teh pa zopet druge enote n.pr. gospodarstva, osebe itd.

Vzorčenje je izredno občutljivo za vse vrste napak, ki se morejo zgoditi bodisi pri zbiranju podatkov ali obdelavi. Zaradi tega je potrebna posebna pazljivost.

Popravki:

stran		namesto	pravilno
63	peta vrsta od spodaj	navačna distribucija	normalna distribucija
133	zadnja vrsta	$N_d \quad \frac{N+1}{2} \quad N_{d+1}$	$N_d < \frac{n+1}{2} < N_{d+1}$
137	dvanajsta vrsta	torej $_k$	torej φ_k
170	sedma vrsta	$t - \frac{u+1}{2}$	$(t - \frac{u+1}{2})$
170	sl. 9.16	vnesti skale vodoravno od 37 - 48, navpično 120, 130, 140	
188	sedma vrsta od spodaj	11.2	102.
213	v tabeli 10.14 in sliki 10.49	47 ha	4,7 ha
230	v sliki 11.2	$_3 - Q_1$	$Q_3 - Q_1$
235	pri sliki 11.3	Določitev ...	odpade
242	osemnajsta vrsta	$_r$	$r\sum$
253	šestindvajseta vrsta	semi -quartilna	semiinterquartilna
253	zadnja vrsta	$\frac{3 - Q_1}{}$	$\frac{Q_3 - Q_1}{2Mo}$
327	četrta vrsta	v obrazcu za $\sigma'_{p'}$	drugi označaj odpade
331	v tabeli 13.28 v zadnji vrsti	$\sigma_x \quad sx$	$\sigma_x^2 \quad s_x^2$

NARODNA IN UNIVERZITETNA
KNJIŽNICA
COBISS ©
00000497279