UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MARIJAN BLEJEC

STATISTIČNE METODE ZA EKONOMISTE

LJUBLJANA 1972

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MARIJAN BLEJEC

STATISTIČNE METODE ZA EKONOMISTE

LJUBLJANA 1972

67083/1.

D1986/13151

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

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

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

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

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

I = Indeks, Zvezni zavod za statistiko

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

SB-LRS = Statistični Bilten LRS

MSP = Mesečni statistični pregled: Zavod za statistiko LRS.

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

Ljubljana, februar, 1959

M. Blejec

Prvo poglavje

U V O D

KAJ JE STATISTIKA?

1.1 Z imenom statistika razumemo več stvari. Statistike so številčni *podatki*, s katerimi opisujemo pojave iz najrazličnejših področij socialno-ekonomskega življenja. Tako so statistike zbrane v najrazličnejših publikacijah, statističnih letopisih, specialnih statističnih publikacijah, n. pr. o industrijski statistiki, kmetijski statistiki, statistiki prebivalstva itd.

Kot statistiko razumemo tudi delo pri *zbiranju statističnih podatkov*. Tako imamo statistiko zunanje trgovine; tu zbiramo podatke iz področja zunanje trgovine; industrijsko statistiko, ki zbira podatke iz industrijske dejavnosti; kmetijsko statistiko, ki se ukvarja z zbiranjem podatkov iz kmetijstva, itd.

Z imenom statistika obeležujemo tudi *organe*, ki zbirajo statistične podatke. Tako pomeni državna statistika mrežo vseh organov, ki zbirajo podatke iz najrazličnejših področij socialno-ekonomskega življenja, da prikažejo številčno sliko dogajanja, ki rabi državnim in drugim organom za njihovo delo in analizo.

Statistika kot *znanost* pa pomeni teorijo in metode statističnega proučevanja. Statistiko kot znanost *opredelimo* takole: statistika je veda, ki s kvantitativnim proučevanjem množičnih pojavov, z metodami, ki so njej lastne, odkriva zakonitosti množičnega pojavljanja in podaja kvalitativno analizo pojavov.

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

Ker je v statističnem proučevanju pojavov vedno poudarek na metodi proučevanja, je v uporabi teh metod statistična metoda v ozki povezavi z neko drugo znanostjo. Zato bi mogli imeti statistiko le za eno izmed metod za proučevanje pojavov. Vendar statistična metoda razen v nekaterih posebnih primerih, ni bistveno vezana na določeno področje, marveč veljajo njene metode in zakonitosti neodvisno od predmeta za vse množične pojave. Tako matematska statistika, ki razvija metode proučevanja množičnih pojavov in katere osnova so postavke verjetnostnega računa, razvija svoje metode neodvisno od področja uporabe. Zato je statistična znanost v ozki

zvezi z matematiko, saj se tudi ta ukvarja s številčnimi odnosi, vendar pa ne s kvantitativnim proučevanjem množičnih pojavov.

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

PODROČJA, V KATERIH UPORABLJAMO STATISTIKO

1.2 Beseda statistika je izšla iz latinske besede „status" - država. Ta izraz je za proučevanje množičnih pojavov zgodovinsko upravičen. Razvojno je statistika res rabila najprej za opisovanje ekonomskih in socialnih razmer v razvitih državah starega in srednjega veka. Vendar je ta pojem že dolgo le ime za nekaj drugega kakor za metodo opisovanja socialnih in ekonomskih razmer v državi. Statistika je razvila svoje metode na splošno proučevanje množičnih pojavov, od katerih je mnogo takih, ki jih ne uporabljamo za opis in analizo ekonomskih in socialnih razmer v državi. Razen tega pa je statistika tako razširila področje uporabe, da so socialno-ekonomski pojavi samo eden, čeprav ne najmanjši sektor v uporabi statističnih metod.

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

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

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

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

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

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

Statistiko z velikim pridom uporabljamo tudi v meteorologiji.

Tudi agronomija je tipično področje za uporabo statističnih metod pri planiranju eksperimentov.

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

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

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

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

SOCIALNO-EKONOMSKA STATISTIKA

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

prostoru v končnem številu.

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

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

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

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

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

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

Analogno imamo nadalje *gozdarsko statistiko, statistiko obrti, gradbeništva, prometa, trgovine, statistiko gostinstva, turizma, financ, statistiko cen, kulturno-prosvetno statistiko, statistiko zdravstva, sodno statistiko* itd.

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

Drugo poglavje

PROUČEVANJE MNOŽIČNIH POJAVOV

MNOŽIČNI POJAVI

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

Če analiziramo zgornje pojave, ki množično nastopajo, odkrijemo, da ti pojavi niso med seboj enaki, marveč se razlikujejo v najrazličnejših značilnostih. Če primerjamo podjetja, vidimo, da imajo med seboj različno število delavcev, različno mesečno proizvodnjo, različno porabo surovin, so iz različnih strok itd. Kupoprodaje se med seboj razlikujejo po času, prodajalcu, kupcu, ceni, količini, kvaliteti prodanega blaga itd. Osebe so različnega spola, starosti, stanu, zaposlitve, se med seboj razlikujejo po šolski izobrazbi, mesečnih prejemkih, številu otrok itd. Artikli imajo med seboj različne dimenzije, kvaliteto, uporabnost itd., čeprav se zde na oko med seboj enaki.

2.2 Če natančneje proučimo možnosti glede analiziranja množičnih pojavov, spoznamo, da je treba področje proučevanja opredeliti. Če namreč opazujemo prebivalstvo, ne moremo nikrati opazovati vsega človeštva v vseh časih, ker je to neizvedljivo, niti nas v konkretni raziskavi to ne zanima. Omejimo se samo na del prebivalstva tako, da opredelimo, katerim pogojem mora ustrezati pojav, da je predmet naše konkretne raziskave. Z *opredeljujočimi pogoji* razmejimo pojave, ki jih proučujemo, od pojavov, ki v konkretnem primeru niso predmet proučevanja. Skupnost pojavov, ki jih opredelimo zato, da jih proučimo, imenujemo *statistično množico* ali *populacijo*. Čeprav je beseda populacija privzeta iz pojma prebivalstva, uporabljamo ta izraz za vse vrste statističnih množic. Populacija v statističnem smislu ni samo prebivalstvo, temveč tudi skupnost industrijskih podjetij v Jugoslaviji na določen datum, skupnost vseh artiklov, proizvedenih na določenem stroju v določenem času itd.

Vsak posamezen pojav populacije imenujemo statistična *enota*. Tako je v zgornjih populacijah enota posamezen prebivalec, posamezno industrijsko podjetje, posamezen artikel.

Vsaka enota populacije ima veliko značilnosti. Izmed teh pa so sa-

mo nekatere predmet našega proučevanja. Te značilnosti, ki so v posameznemu primeru predmet našega proučevanja, imenujemo *statistične znake*, ali kratko, *znake*. Vsak znak ima za posamezne enote različne *vrednosti*. Tako niso vsi ljudje enako stari, vsa podjetja nimajo enako število delavstva, vsi artikli niso enako uporabni itd.

Vedtem ko je spol znak posamezne enote, je število oseb v populaciji, ki so moškega spola, značilnost celotne populacije, čeprav je ta podatek odvisen od tega, kakšnega spola so posamezni prebivalci v populaciji. Enako je odstotek neuporabnih artiklov v proizvodnji za določen dan značilnost celotne proizvodnje, ne pa značilnost za posamezen artikel, čeprav smo ta podatek izvedli iz podatkov o uporabnosti posameznih artiklov. Značilnosti populacije, kot so na primer število moškega prebivalstva, odstotek neuporabnih artiklov v dnevni proizvodnji, povprečna plača delavcev, povprečen hektarski donos pšenice v socialističnih kmetijskih obratih, itd. imenujemo *parametre* populacije.

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

Statistične enote

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

a/ oseba (prebivalec, delavec, študent),
b/ žival (konj, krava, prašič itd.),
c/ stvar (avto, motor, artikel),
d/ pravna tvorba (zadruga, društvo),
e/ administrativna enota (občina, okraj),
f/ gospodarska tvorba (industrijsko podjetje, trgovina, kmetijsko gospodarstvo),
g/ dogodek (kupoprodaja, smrt, nesreča itd.),
h/ poskus (poskusna obdelava parcele, cepljenje živali).

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

Statistične enote so bodisi *enostavne enote* ali pa *agregati* - skupinice enostavnih enot. Tako štejemo posameznega človeka za enostavno enoto, družina, občina, društvo itd., pa so agregati, ker so te enote sestavljene iz enostavnih enot: članov družine, prebivalcev občine, članov

društva itd.

Statistični znaki

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

2.5 **Vrste statističnih znakov.** Vsebinsko delimo znake na a) krajevne ali geografske, b) časovne in c) stvarne.

Krajevni ali geografski znak je lahko kraj, ki je v zvezi z enoto proučevanja ali z dogodkom, ki je značilen za proučevano enoto. Tako je krajevni znak kraj rojstva, kraj stalnega bivališča, kraj zaposlitve, sedež podjetja, kraj nesreče itd.

Časovni so vsi znaki, ki so v zvezi s časom, v katerem se je zgodil kak dogodek, ki je v zvezi s proučevano enoto. Časovni znak je torej lahko čas rojstva, čas izdelave artikla, čas ustanovitve podjetja itd.

Vsi drugi znaki so *stvarni*. Stvarnih znakov je največ in z njimi označujemo vse značilnosti pojavov, ki niso krajevne ali časovne.

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

Atributivni so vsi znaki, za katere vrednosti izražamo opisno, z besedami. Tako je na primer atributiven znak spol, ker so njegove vrednosti: moški, *ženski, izražene z besedami*. Iz istega vzroka je atributiven znak tudi zaposlitev, panoga dejavnosti, vrsta zgradbe, vzrok smrti itd.

Numerični so vsi znaki, katerih vrednosti izražamo številčno. Po tem, katere vrednosti morejo zavzeti, delimo numerične znake v a) nezvezne - diskontinuirne in b) zvezne - kontinuirne. *Nezvezni znak* je na primer število članov v gospodinjstvu, število otrok, ki jih ima mati, število delavcev itd. Nezvezni znaki morejo imeti samo neke, običajno cele vrednosti na določenem razmaku. Tako ne more imeti gospodinjstvo 3 1/2 članov, mati 1 1/4 otrok, podjetje 83 1/2 zaposlenih itd. *Zvezni znaki* so tisti numerični znaki, ki morejo teoretično imeti vsako izmed vrednosti v določenem razmaku. Tako je zvezni znak starost, teža artikla, dohodek itd.

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

Vendar moremo tudi nekatere atributivne znake urediti po velikosti. To so znaki, ki imajo v osnovi numeričen značaj. Stopnjo šolske izobrazbe izražamo atributivno v naslednjih stopnjah: brez šolske izobrazbe, nedokončana osnovna šola, dokončana osnovna šola, nedokončana nižja srednja šola, dokončana nižja srednja šola, nedokončana višja srednja šola, dokončana višja srednja šola, nedokončana visoka šola in dokončana visoka šola. Te vrednosti so urejene po velikosti. To je možno, ker je v osnovi stopnje šolske izobrazbe numeričen znak - število razredov, ki dopušča, da razporedimo vrednosti atributivnega znaka po velikosti. Enak primer je kvalifikacija delavcev: visoko kvalificiran, kvalificiran, priučen, nekvalificiran, ali obdelanost predmetov v zunanji trgovini: visoka obdelava, srednja obdelava, polizdelki, surovine, oziroma neobdelani proizvodi.

2.7 Znake delimo po drugem kriteriju na: a) ekstenzivne in b) intenzivne. *Ekstenzivni* znaki nakazujejo količino, množino, *intenzivni* pa kakovost. Tako je ekstenziven znak vrednost proizvodnje podjetja v določenem razdobju, število delavstva po podjetjih, količina določenega blaga na trgu itd. Intenzivni znaki pa so na primer cena artikla, produktivnost dela, merjena z vrednostjo proizvoda na enoto časa, hektarski donos itd.

Vsebinska razdelitev znakov na ekstenzivne in intenzivne se tehnično sklada z razdelitvijo znakov na: a) *adirne* in b) *sortirne*. Ekstenzivni znaki imajo adirni in sortirni značaj. Vrednosti ekstenzivnih znakov imajo adirni značaj, ker jih moremo seštevati (proizvodnjo več podjetij v proizvodnjo stroke: vrednosti posameznih vrst blaga na trgu v skupno vrednost blaga na trgu). Ker moremo podjetja razvrstiti tudi v grupe po velikosti proizvodnje, imajo ekstenzivni znaki tudi sortirni značaj. Drugače je z intenzivnimi znaki. Vrednosti intenzivnih znakov nima smisla seštevati: nimajo torej adirnega značaja. Pač pa moremo enote po vrednostih intenzivnega znaka razporejati v grupe. Intenzivni znaki imajo torej samo sortirni značaj.

2.8 Za analizo je zelo važna delitev znakov po funkciji, ki jo imajo. V tej zvezi razdeljujemo znake na: a) *faktorialne* in b) *rezultativne*. Faktorialni znaki so izraz faktorjev, ki vplivajo na pojav. Rezultativni znaki pa so rezultat vpliva vseh faktorjev. Tako je na primer količina uporabljenega umetnega gnojila faktorialen, hektarski donos pa rezultativen znak. Enako je število delavstva v podjetju faktorialen, vrednost proizvodnje pa rezultativen znak. Količina blaga na trgu je faktorialen,

cena blaga pa je rezultativen znak. Zveze med faktorialnimi in rezultativnimi znaki kažejo, kako se izpreminja vrednost rezultativnega znaka, če se spreminjajo faktorialni znaki. Proučevanje teh zvez je ena izmed važnih metod v statistični analizi.

2.9 Variiranje statističnih znakov. Vsak znak ima določeno število vrednosti znakov. Število vseh teoretično možnih vrednosti je najmanj dve, največ neomejeno. Tako ima na primer spol dve vrednosti: moški, ženski, stan štiri vrednosti: samski, poročen, razvezan in vdovec, poklic veliko, vendar končno število različnih možnih vrednosti, človeška starost pa teoretično neomejeno število vrednosti. Značilno za znake je to, da imajo posamezne enote različne vrednosti znakov. Temu pojavu pravimo *variiranje* statističnih znakov. Statistični znaki so torej variabilni. Da imajo nekatere enote tudi iste vrednosti enega in istega znaka, ni v nasprotju z definicijo variiranja. Vsaka oseba v dani populaciji namreč ne more biti različnega spola, če ima znak spol samo dve varianti: moški ali ženski.

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

Posebno važna je variabilnost numeričnih znakov. Kakor bomo videli kasneje, moremo variabilnost numeričnih znakov najpogrobneje proučevati.

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

Statistične populacije

2.10 Opredelitev statistične populacije. Statistična populacija je skupnost vseh istovrstnih pojavov - statističnih enot, ki izpolnjujejo *opredeljujoče pogoje*, s katerimi je opredeljena. Statistično populacijo opredelimo tako, da navedemo, kakšne vrednosti znakov, ki so opredeljujoči pogoji, morajo imeti enote, da so enote populacije. V znakih, ki opredeljujejo populacijo, je torej variacija okrnjena. Če proučujemo razmere zaposlenih poročenih žensk na področju Slovenije po stanju 31.12. 1958, je s tem že opredeljena populacija, ki jo proučujemo. Proučujemo samo poročene, zaposlene ženske. V stanu, zaposlenosti in spolu populacija ne variira. Enako tudi ni variacije v časovnem znaku, ker je populacija opredeljena s teoretično momentanim stanjem konec leta 1958. V krajevnem zna-

ku pa je variabilnost populacije okrnjena, ker proučujemo žene v Sloveniji.

Za populacijo ne velja samo skupnost vseh statističnih enot v zgornjem smislu, marveč je populacija tudi skupnost podatkov za neki znak. Tako je populacija v tem smislu tudi skupnost vseh podatkov o starosti poročenih zaposlenih žena v LRS po stanju 31.12. 1958.

2.11 Kakor vidimo iz zgornjega primera, mora biti populacija nedvoumno opredeljena: a) časovno, b) krajevno in c) stvarno.

Zgornjo populacijo smo časovno opredelili tako, da smo navedli: enote populacije so samo tiste ženske, ki zadoščajo vsem opredeljujočim pogojem konec leta 1958. Krajevno je zgornja populacija opredeljena s tem, da smo navedli: enote populacije so samo ženske, ki zadoščajo ostalim opredeljujočim pogojem in so v Sloveniji. Stvarno je populacija opredeljena s tremi znaki: stanom, spolom in zaposlenostjo.

V splošnem *krajevni opredelitev* ne dela posebnih težav. Običajno se pri krajevni opredelitvi naslonimo na administrativno razdelitev teritorija. To olajša krajevno opredelitev. Razen tega pa imajo večjo operativno vrednost podatki, ki so dani za teritorije, skladne z upravno razdelitvijo.

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

Zgornjo populacijo žensk smo opredelili z momentom, ker je populacija realnih enot. Zato populacije realnih enot včasih tudi imenujemo *momentne populacije*. Momentnih populacij imamo veliko. Vse časovno enolično opredelimo z momentom, v katerem morajo posamezne enote populacije zadoščati vsem ostalim pogojem, da so enote populacije. Če se je katera izmed zaposlenih žena poročila v januarju, ne pride v populacijo. Enako tudi ni enota populacije poročena žena, ki je bila še zaposlena v decembru 1958, je pa v decembru izstopila iz službe. Ta momentna opredelitev populacij realnih enot je potrebna zaradi nedvoumne razmejitve pojavov v pojave, ki so enote populacije, od pojavov, ki niso enote populacije.

Če bi populacijo stvarnih enot skušali opredeliti z intervalom, te nedvoumnosti ne bi dosegli. Če bi zgornjo populacijo zaposlenih poročenih žena opredelili časovno z mesecem decembrom 1958, bi bilo dvomljivo, katere žene pridejo v populacijo in katere ne. Problematični bi bili vsi primeri žena, ki so se v decembru bodisi zaposlile ali izstopile iz službe.

Časovna opredelitev populacij dogodkov z momentom ni smiselna. V določenem momentu se more zgoditi in se zgodi samo omejeno in slučajno število dogodkov. Tako nima smisla iskati populacijo rojstev, nesreč, smrti, proizvodnje itd. ob določenem momentu, ker ta populacija poleg svoje omejenosti niti ni smiselna. Populacijo dogodkov opredelimo z razdobjem in pridejo v populacijo vsi dogodki, ki so se zgodili v določenem razdobju in zadoščajo vsem ostalim pogojem. Te populacije imenujemo analogno z momentnimi *intervalne populacije*, ker so opredeljene s časovnimi inter-

vali. Tako so intervalne populacije proizvodnja, rojstva, nesreče, prodaja itd.

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

Stvarno je zgornja populacija žensk opredeljena s spolom: ženske, zaposlenostjo: zaposlene in stanom: poročene. Za to populacijo je stvarna opredelitev enostavna. V splošnem pa je stvarna opredelitev najtežja in zahteva dobro poznavanje predmeta, ki ga proučujemo.

2.12 Homogene in heterogene populacije. Po sorodnosti enot so populacije homogene - istorodne in heterogene - raznorodne. Za homogene populacije je bistveno, da enote med seboj niso preveč različne, medtem ko so heterogene populacije sestavljene iz raznorodnih enot. Homogeno populacijo dobimo, če jo opredelimo po faktorialnih znakih. Delitev populacij na homogene in heterogene ni ostra in imamo bolj ali manj homogene in bolj ali manj heterogene populacije. Tako je populacija obrtnih podjetij za določeno stroko, v katerih obrtnik razen sebe zaposluje enega samega delavca, bolj homogena kot populacija obrtnih podjetij za isto stroko ne glede na število zaposlenih. Čeprav druga populacija ni heterogena, je vendar manj homogena kot prva.

2.13 Delne populacije. Če v zgornji populaciji zaposlenih poročenih žena v Sloveniji konec leta 1958 dodamo dodatni pogoj, da je zaposlena kot delavka, smo iz prve populacije izdvojili novo populacijo. To populacijo imenujemo delno populacijo zato, ker je vsaka enota druge populacije hkrati enota prve populacije.

Populacijo moremo po nekem znaku razdeliti v popoln sistem delnih populacij, če jo razdelimo v delne populacije po vseh vrednostih nekega znaka. Tako moremo populacijo vseh obrtnih obratov v Sloveniji konec leta 1958 razdeliti po strokah v popoln sistem delnih populacij. Delne populacije so bolj homogene kot osnovna populacija, ker se enote delnih populacij ujemajo še v dodatnem pogoju. Če populacijo razdelimo v sistem delnih populacij po faktorialnem znaku, moremo heterogeno populacijo razdeliti v sistem homogenih populacij. To metodo večkrat uporabljamo pri analizi množičnih pojavov.

2.14 Korespondirujoče populacije. Med nekaterimi momentnimi in intervalnimi populacijami je zveza, ki je nakazana v naslednjem primeru: Če vzamemo momentno populacijo prebivalstvo v začetku leta, smrt vsakega izmed teh prebivalcev vpliva na momentno populacijo - prebivalstvo konec leta tako, da se populacija za eno enoto zmanjša. Vsako rojstvo pa ima obraten učinek in se populacija prebivalstva za eno enoto zveča. Vsa rojstva v tem letu so intervalna populacija vseh rojstev, vse smrti v tem letu pa intervalna populacija vseh smrti. Iz primera vidimo, da se dve momentni populaciji (prebivalstvo v začetku in prebivalstvo na koncu leta) vežeta z dvema intervalnima populacijama (populacija vseh smrti in populacija vseh

rojstev v letu). Število enot v posameznih populacijah pa je v tejle zvezi

$$P_1 = P_0 + R - S \qquad (2.1)$$

Pri tem pomeni: P_1 = število prebivalstva konec leta; P_0 = število prebivalstva v začetku leta; R = število rojstev v letu; S = število smrti v letu.

Podobno so korespondirujoče populacije tudi druge populacije. Tako se veže na primer: začetna zaloga-nabavljeno-porabljeno-končna zaloga; delavstvo na začetku meseca - na novo prišli-odšli-delavstvo na koncu meseca ita.

Statistični parametri

2.15 S statističnimi parametri opisujemo in analiziramo množične pojave. Z njimi številčno izražamo kvantitativne in kvalitativne značilnosti in odnose populacij. Statistične parametre izračunavamo iz podatkov vseh ali samo nekaj enot populacije. Nekatere parametre populacij dobimo s preprostim preštevanjem in seštevanjem podatkov za vse enote populacije. Tako je statistični parameter za populacijo kmetijskih gospodarstev število kmetijskih gospodarstev, ki ga dobimo, če preštejemo vse enote populacije. Parameter skupno vrednost proizvodnje vseh kmetijskih gospodarstev v populaciji dobimo, če seštejemo vrednost proizvodnje vseh kmetijskih gospodarstev v populaciji.

Parametre, ki imajo že analitično vrednost, dobimo s preštevanjem in seštevanjem v delnih populacijah. Tako dobimo na primer s preštevanjem po delnih populacijah število gospodarstev po vrsti proizvodnje, velikosti gospodarstev ita., s seštevanjem podatkov pa vrednost proizvodnje v vseh kmetijskih gospodarstvih po vrsti proizvodnje, velikosti gospodarstev ita.

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

ZNAČILNOSTI PRI PROUČEVANJU MNOŽIČNIH POJAVOV

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

V fiziki, kemiji in v agronomiji proučujemo pojave s poskusi. Po-

skuse v *fiziki* in *kemiji* delamo laboratorijsko v natančno določenih pogojih, da odstranimo vse vplive, ki bi mogli motiti poskus. Če bi mogli odstraniti vse dodatne vplive, ki motijo poskus, in rezultate meriti z idealno točnimi merilnimi instrumenti, bi zadoščal en sam poskus, ker bi dobili pri vsaki ponovitvi v izenačenih pogojih isti rezultat. Vendar v praktičnih primerih rezultati pri ponovitvah niso povsem enaki. Pokažejo se majhne razlike, ker je nemogoče poskus ponoviti natančno enako. Zaradi dodatnih vplivov, ki jih z našimi eksperimentalnimi sredstvi ne moremo odpraviti, so rezultati med seboj različni. Te razlike imenujemo slučajne razlike, vplive, ki jih povzroče, pa slučajne vplive. Vendar se te razlike v velikem številu poskusov pojavljajo v nekih zakonitostih. Zaradi slučajnih vplivov se rezultati posameznih eksperimentov odklanjajo od prave vrednosti navzgor in navzdol. Po *zakonu o velikih številih* pa se v vsoti rezultatov velikega števila poskusov kompenzirajo. Zato imamo povprečen rezultat, ne pa rezultat posameznega poskusa za pravi rezultat, ki bi ga dobili, če ne bi bilo subjektivnih momentov eksperimentatorja, nemožnosti popolne določitve eksperimentalnih pogojev in napak v merilnih instrumentih. Da dobimo objektivnejše povprečje, je dobro, če vsak poskus izvedejo različni eksperimentatorji z različnimi merilnimi instrumenti. Tako merilni instrumenti kot eksperimentator lahko povzročijo sistematične napake. Napaka v instrumentu je lahko take narave, da sistematično daje vedno premajhne rezultate. Isto more biti tudi z eksperimentatorjem. Če ponavljamo poskuse z istimi merilnimi instrumenti ali eksperimentatorji, se sistematične napake ne kompenzirajo in povprečen rezultat se ne približa dejanski vrednosti.

Podobno je pri poiskusih v drugih znanostih. Za razliko od laboratorijskih fizikalnih poskusov je na primer pri poskusih v *agronomiji* kompleks slučajnih faktorjev močnejši kot pri fizikalnih poiskusih. Mikrosestav zemlje, razlike v semenu, lega parcele itd. so faktorji, ki jih pri posameznih poskusih ne moremo natančno izenačiti. Pri poskusih v agronomiji moramo na primer določen gnojilni poskus izvesti v več ponovitvah, da izločimo vpliv slučajnih faktorjev, katerih iz objektivnih razlogov ne moremo odstraniti pri individualnih poskusih.

Kakor pri fizikalnih poskusih, tako smo tudi pri poskusih v agronomiji ugotovili isto važno dejstvo. Zakonitost, ki jo preizkušamo s poskusom, v individualnem primeru motijo slučajni vplivi. V veliki množici istovrstnih poskusov pa se učinek slučajnih vplivov v povprečju kompenzira in zaradi delovanja zakona o velikih številih izbije tipičnost pojava. Ta postopek je kvalitetno povsem specifičen. Točnejših rezultatov pri eksperimentih ne dosegamo z izboljšavo laboratorijske tehnike, temveč s ponavljanjem poskusov pod danimi pogoji: iz rezultatov skupine poiskusov pa dobimo bolj natančen rezultat. Taka pot je v dosti primerih boljša, ker je dosti ceneje ponavljati poskus z danimi sredstvi, kakor pa z dragimi aparaturami neposredno odstranjevati slučajne vplive. Če pa dosežemo meje možnosti pri opredelitvi pogojev, pa so kljub temu slučajni vplivi močni, kakor je to na primer pri poskusih v agronomiji, pa edino s ponovnimi poskusi dosežemo natančnejše rezultate.

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

Iz zgornjih primerov vidimo, da z množičnim opazovanjem pojavov presežemo meje natančnosti, ki jih moremo doseči z individualnim opazovanjem.

2.17 Če skušamo postopke, ki smo jih navedli kot primerne za proučevanje v fiziki in agronomiji, prenesti v *socialno-ekonomske znanosti*, takoj spoznamo neke bistvene razlike v možnosti opazovanja. Tako v fiziki in v agronomiji smo za proučevanje določene zakonitosti izvedli poskuse bodisi v laboratoriju ali na polju. Predmet opazovanja v socialno-ekonomskih znanostih pa običajno ne moremo ustvariti s poskusom. Če proučujemo, kako je struktura porabe odvisna od skupnih dohodkov po gospodinjstvih, ne moremo zaradi te raziskave sestaviti eksperimentalnih gospodinjstev z natančno določeno strukturo članov, jim odrediti določene skupne dohodke in opazovati razlike v strukturi, če imajo različne skupne dohodke. Enako ne moremo zaradi raziskave o višini proizvodnje v podjetjih kovinske stroke z 51 do 100 zaposlenih delavcev zaradi raziskave ustanoviti niz podjetij kovinske stroke, ki imajo 51 do 100 zaposlenih delavcev in opazovati njihovo proizvodnjo. V socialno-ekonomskih znanostih v splošnem enot opazovanja ne moremo (razen v nekaterih posebnih primerih) stvoriti s poskusi kakor v fiziki in agronomiji. Tu uporabljamo druge metode. Življenje samo ustvarja pojave v najrazličnejših variantah. Ko proučujemo, kako je odvisna struktura porabe od višine dohodkov, ne sestavimo eksperimentalnih družin, temveč iz množice dejanskih gospodinjstev poiščemo taka, ki ustrezajo pogojem, ki jih ima naše raziskovanje. Če proučujemo višino proizvodnje podjetij v kovinski stroki, ki imajo zaposlenih od 51 do 100 delavcev, ne ustanavljamo podjetja zaradi proučevanja, temveč iz obstoječih industrijskih podjetij poiščemo podjetja iz kovinske stroke, ki imajo zaposlenih od 51 do 100 delavcev, in zanje proučujemo proizvodnjo. To je ena izmed temeljnih razlik med proučevanjem s poskusi in med proučevanjem, ki je običajno v socialno-ekonomskih raziskavah.

Vendar razmejitev ni tako ostra in moremo v določenih pogojih in za določene potrebe tudi v socialno-ekonomskih proučevanjih uporabiti eksperimentalno metodo. Če na primer proučujemo odvisnost prodaje nekega artikla od embalaže in mesta v trgovini, kjer prodajamo ta artikel, moremo [illegible] problem raziskati eksperimentalno v pravem pomenu besede. V določeni veleblagovnici organiziramo prodajo tega artikla po natančno določenem planu: koliko časa, kdaj in kje prodajamo določeni artikel na ta ali drug način. Ta razpored more biti sestavljen kot čisti eksperiment, kakor planiramo eksperimente na primer v agronomiji ali drugod. Takih primerov moremo našteti še več. Vsi ti eksperimenti pa so večinoma manjšega obsega in

internega značaja. Z eksperimenti more na primer poljetje zase proučevati učinke različnih pogojev na produktivnost dela, probleme notranjega transporta, učinek odmora na delovno storilnost itd.

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

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

Tabela 2.1. Spolni indeks novorojenčkov za Ljubljano, Slovenijo in Jugoslavijo v letih 1947–1955

	Leto								
	1947	1948	1949	1950	1951	1952	1953	1954	1955
Ljubljana	111,3	114,4	107,1	103,3	97,9	109,0	111,6	105,3	111,2
Slovenija	107,5	106,0	105,9	107,2	105,2	106,6	108,8	103,8	108,8
Jugoslavija	108,2	107,3	106,7	107,5	107,4	106,8	107,0	106,8	107,0

Iz podatkov je razvidno, kako raste stabilnost podatkov, čim večje so populacije, ki jih proučujemo. Medtem ko je število rojstev v Ljubljani po letih ca 2500 letno, je število rojstev v Sloveniji povprečno 33 tisoč, v Jugoslaviji pa 470 tisoč na leto. Skladno z velikostjo populacij je za Ljubljano razlika med največjim in najmanjšim letnim vitalnim indeksom $R = 114{,}4 - 97{,}9 = 16{,}5$, za Slovenijo $R = 108{,}8 - 103{,}8 = 5{,}0$, za Jugoslavijo pa $R = 108{,}2 - 106{,}7 = 1{,}5$.

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

Iz primera lepo vidimo veljavnost zakona velikih števil. Zakonitosti se pokažejo v množičnem opazovanju, medtem ko v individualnih primerih oziroma v malem številu primerov prevladujejo individualni vplivi, ki zatrišejo splošno zakonitost.

Primer pa je pokazal tudi to, da v nekaterih primerih le za velike populacije izbije iz pojava zakonitost. To je vselej, kadar so individualni vplivi zelo močni.

Vendar dostikrat nimamo možnosti opazovati tako velikega števila primerov kakor v zgornjem primeru analize spolnega indeksa novorojenčkov. Zato se za male populacije navadno omejimo le na statistično opisovanje množičnega pojava, ne iščemo pa zakonitosti. Statistično proučevanje socialno-ekonomskih pojavov zato včasih omejimo le na opisovanje populacij.

ETAPE STATISTIČNEGA PROUČEVANJA

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

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

a/ *določitev vsebine in namena statističnega proučevanja z analizo pojava, ki ga proučujemo,*
b/ *izdelava splošnega plana statističnega proučevanja,*
c/ *opazovanje v ožjem smislu,*
d/ *urejevanje in osnovna obdelava,*
e/ *analitična obdelava in analiza podatkov.*

2.20 Za vsako statistično proučevanje je potrebno takoj v začetku še pred katerim koli drugim delom *določiti namen in vsebino* proučevanja. Natančno moramo opredeliti pojav, ki ga proučujemo, in določiti znake, ki jih bomo proučevali. Če hočemo na primer proučevati delovno silo v kmetijstvu, je treba natančno določiti obseg proučevanja. Lahko se omejimo samo na delovno silo v privatnem sektorju, na delovno silo, ki je zaposlena izključno v kmetijstvu, itd. Enako je od namena proučevanja odvisno, ali se bomo omejili pri proučevanju samo na številčno stanje delovne sile v kmetijstvu, ali nas zanima stanje po velikostnih skupinah za različne vrste kmetijskih gospodarstev, ali so za proučevanje važni tudi živ-

ljenjski pogoji itd. Če je jasno postavljen namen statističnega proučevanja, moremo predmet opredeliti in določiti posamezne pojme.

Glede na namen in potrebe analize je treba v prvi stopnji tudi določiti, ali je potrebno podrobnejše proučevanje, ali je zadostno, da dobimo za posamezne proučevane pojave le ocene.

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

Ko postavljamo namen in vsebino statističnega proučevanja, sodelujejo pri tem predvsem konzumenti statističnih podatkov in strokovnjaki iz področja, iz katerega je problem, ki ga skušamo analizirati.

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

Zahteve, ki jih stavimo na splošni plan, ni lahko izpolniti. Zato je za sestavljanje splošnega plana potrebno poznavanje teoretičnih in tehničnih prijemov v vseh stopnjah statističnega proučevanja, tako v opazovanju kakor pri obdelavi in analizi.

Splošni plan mora glede na namen in razpoložljiva materialna in finančna sredstva določiti metodo opazovanja, način osnovne in analitične obdelave in način prikazovanja oziroma publiciranja podatka. Na te splošne postavke splošnega plana se mora nasloniti operativni plan v posameznih stopnjah.

2.22 Statistično opazovanje v ožjem smislu. Statistično opazovanje v ožjem pomenu besede je zbiranje statističnih podatkov o statističnih enotah. S statističnim opazovanjem zberemo podatke o statistični populaciji v obliki, ki je primerna za osnovno obdelavo. Imamo več vrst statističnih opazovanj. Vsako izmed njih ima svoje tipičnosti glede na meto-

do dela in glede na to, kakšna je vrednost in zanesljivost rezultatov, ki jih z njim dobimo. Izvajanje statističnega opazovanja je bolj tehničnega kot pa vsebinskega značaja. V zvezi z njim moramo rešiti niz tehničnih problemov, katerih teža je odvisna od populacije, ki jo proučujemo, in vrste opazovanja, ki ga uporabljamo.

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

2.24 Analitična obdelava. V osnovni obdelavi dobimo podatke urejene, preštete in seštete. Vendar ti podatki še ne zadostujejo za analizo pojava. S preračunavanjem osnovnih absolutnih podatkov s posebnimi metodami, ki jih bomo v nadaljevanju podrobneje obravnavali, analiziramo pojave, ki jih proučujemo. Iz osnovnih podatkov namreč dalje izračunavamo relativna števila, srednje vrednosti, mere variacije, mere korelacije, analiziramo dinamiko pojavov itd.

Iz navedenega je jasneje vidno, kako potrebna je povezava vseh stopenj proučevanja: samo tak postopek zagotavlja uspešno delo in zaključek statističnega proučevanja.

Tretje poglavje

STATISTIČNO OPAZOVANJE

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

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

OPREDELITEV PREDMETA OPAZOVANJA STATISTIČNE POPULACIJE

3.2 Že v programu statističnega proučevanja v načelu opredelimo enoto proučevanja oziroma populacijo, ki jo opazujemo. Vendar moramo pred izvedbo statističnega opazovanja populacijo in enote opazovanja opredeliti tako, da je nedvomno jasno in preprosto razumljivo, kateri pojavi so predmet opazovanja in kateri ne. Populacijo moramo, kakor smo že nakazali, opredeliti: stvarno, krajevno in časovno.

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

Včasih se zaradi enostavnosti in nedvoumnosti, kaj zbrani podatki

pomenijo, omejimo na en sam znak in z njim opredelimo enoto opazovanja. Tako smo v letu 1948 vključili v popis industrijskih podjetij kot industrijska podjetja vsa proizvodna podjetja z več ko petimi delavci. Leta 1947 pa smo popisali kot kmetijsko gospodarstvo vsako posest z nad 500 m^2 zemlje. Čeprav zgornji definiciji nista natančni, smo z njima dosegli enoličnost in nedvoumnost tako pri opazovanju kakor pri tolmačenju podatkov. Meji pet delavcev za popis industrija in 500 m^2 za popis kmetijskih gospodarstev imenujemo *census normo*. S census normo enostavno z enim samim znakom vsebinsko opredelimo pojav. Čeprav ta način teoretično ni najboljši, ima take tehnične prednosti, da ga često uporabljamo.

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

Časovna opredelitev je tudi brez posebnih težav in pri popisih opredelimo populacijo časovno s teoretičnim momentom - kritičnim momentom, pri tekočih registracijah pa navadno s koledarskim razdobjem.

VRSTE OPAZOVANJ

3.3 Glede na proučevano populacijo uporabljamo različne vrste statističnega opazovanja. Medtem ko populacije realnih enot, kot so prebivalstvo, stavbe, gospodarstva, živina itd., opazujemo s *popisi*, populacije dogodkov, kot so smrti, rojstva, nesreče, itd., opazujemo s *tekočo registracijo*. Glede na potrebe, materialne in finančne možnosti pa opazujemo *vse enote populacije*, če potrebujemo popolno in podrobno sliko o populaciji, ali samo *del populacije*, če se zadovoljimo z ocenami.

Popis

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

Osnovna značilnost popisov je, da popišemo vse enote populacije po stanju na določen moment - kritični trenutek. Popisati vse enote populacije, je idealno stanje, ki se mu bolj ali manj približamo. Odstotek zajetja, ki pove kolik del populacije smo uspeli popisati, meri popolnost zajetja. Odstotek zajetja je odvisen od proučevane populacije in uspešnosti organizacije opazovanja. Populacije, ki se gibijejo, je teže popolno popisati kakor populacije, ki se malo gibijejo ali pa mirujejo. Popolnost zajetja pri popisu prebivalstva je zaradi gibanja prebivalstva večji problem kakor pri popisih zgradb, popisih kmetijskih gospodarstev itd.

3.5 S popisom dobimo podobno kakor s fotografskim posnetkom trenutno sliko populacije. Nemogoče je n.pr. s statističnim popisom obseči prebi-

valstvo v mesecu ali letu, ker je stanje prebivalstva vsak hip drugačno.

Zato pri vsakem popisu določimo trenutek, na katerega se nanašajo zbrani podatki. Ta trenutek imenujemo *kritični trenutek*. Čeprav populacije ne moremo opazovati natančno ob kritičnem trenutku in popisujemo enote po kritičnem trenutku, jo vendar skušamo popisati po stanju, kakršno je bilo ob kritičnem trenutku. Kako dolg naj bo čas popisovanja, ki se ne ujema s kritičnim trenutkom, je odvisno od populacije in tehničnih možnosti popisovanja. Posebno za populacije, ki se gibljejo, čas popisovanja ne sme biti preveč oddaljen od kritičnega datuma, niti predolg, ker drugače težko ugotovimo stanje ob kritičnem trenutku. Čas popisovanja pri popisu prebivalstva mora biti razmeroma kratek in morajo popisovalci popisati prebivalstvo v nekaj dneh po kritičnem trenutku. Pri popisu zgradb pa n.pr. ni nujno, da je tako kratek, ker populacija zgradb ni tako gibljiva kakor prebivalstvo.

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

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

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

3.6 S popisom dobimo trenutno sliko pojava. Vendar je za proučevanje populacij dostikrat zelo važna *dinamika*. Te pa enkratni popis ne da. Dobimo jo pa z več zaporednimi popisi v določenih časovnih razmakih, podobno kakor v kinu iz slik v časovnem zaporedju dobimo vtis o gibanju. Če hočemo, da dajo rezultati zaporednih popisov osnovo za proučevanje dinamike, moramo upoštevati tale tri načela:

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

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

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

Tekoča registracija
Statistična poročila

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

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

Redkeje ko pri popisih je namen tekoče registracije zbiranje podatkov samo za statistično proučevanje pojava. Matična služba, register prebivalstva, operativne evidence v podjetjih, knjigovodski podatki itd., so samo posredno vir statističnih podatkov. Te podatke zbiramo s *statističnimi poročili*, ki sumarno zajamejo dogodke v določenem razdobju(dnevu, tednu, mesecu, letu). Statistična poročila so ena izmed običajnih oblik registracije socialno-ekonomskih pojavov. Tako imamo n.pr. statistična poročila iz statistike cen, gostinstva, turizma, gradbeništva, industrije, matične službe itd.

Metode delnega opazovanja

3.8 Značilno za popis in tekočo registracijo je, da z njimi opazujemo vse enote populacije. Tako dobimo zelo podrobno sliko proučevanega pojava. Slaba stran popolnega opazovanja pa je v tem, da zahteva veliko materialnih sredstev in časa.

Kadar za proučitev pojava ni potrebna podrobna slika opazovane

populacije ali če nimamo na razpolago zadostno sredstev, podatke o populaciji ocenjujemo. Najpopularnejše so metode ocenjevanja, pri katerih skušamo sklepati in ocenjevati podatke o populaciji, če poznamo razmeroma majhen del enot populacije.

Najobjektivnejša metoda pri ocenjevanju statističnih podatkov je *vzorčenje*. Pri vzorčenju so enote, ki jih opazujemo, da iz njih ocenimo celoto, izbrane s postopkom slučajnega izbora enot. Metoda slučajnega izbora ima niz prednosti pred drugimi metodami ocenjevanja. Ocene, ki jih dobimo z vzorčenjem, so objektivne, na kvaliteto ocen moremo enostavno vplivati, zanesljivost ocen pa moremo objektivno določiti. Ker je ta metoda ocenjevanja v socialno-ekonomskih proučevanjih zelo pomembna, za njeno razumevanje pa je potrebno poznavanje osnov statistične analize, na tem mestu vzorčenje samo omenjamo, podrobneje pa ga bomo obravnavali v posebnem poglavju kasneje.

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

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

3.10 *Monografijo* uporabljamo, kadar hočemo zelo podrobno osvetliti določen problem na enem ali več tipičnih enotah populacije. Namen monografije ni posploševanje na celoto, temveč podrobna osvetlitev posameznih primerov. Znana so podrobna monografska proučevanja slovenske vasi, monografska proučevanja o produktivnosti dela, o delovnih razmerah delavcev ita. Monografska proučevanja so zelo uspešna tudi pri odkrivanju novih kvalitet pojavov.

VIRI PODATKOV

3.11 Pri organiziranju statističnega opazovanja je važno proučiti vire, iz katerih črpamo podatke. Predhodna analiza virov mnogokrat poenostavi opazovanje, včasih pa celo uspemo, da pridemo do podatkov o proučevanem pojavu brez statističnega opazovanja. Zavedati se namreč moramo, da je statistično opazovanje najdražji in razmeroma zamotan način, da pridemo do statističnih podatkov.

Zato predhodno pregledamo vse *publicirane in nepublicirane podatke* o problemu, ki ga proučujemo. Včasih dobimo odgovor na postavljeni problem že neposredno iz podatkov, ki so jih zbrali in obdelali drugi. Tako uspemo, da analiziramo pojav s podatki iz enega ali več različnih virov brez posebnega statističnega opazovanja in osnovne obdelave. Večina podatkov, ki jih publicira uradna statistika, so osnovni podatki. Njih namen je, da jih dalje analizira konzument za svoje potrebe in po svojih vidikih.

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

3.13 Dostikrat vzamemo za osnovo proučevanja *dosedanje zapiske - evidence*, ki jih zbira ena ali druga organizacija ali enota v nestatistične namene. Ti zapiski - imenujemo jih tudi sekundaren vir statističnih podatkov - so osnova podatkov iz vseh gospodarskih statistik, statistike industrije, trgovine, gostinstva itd. Prav tako je sekundaren vir statističnih podatkov matična služba, ki zbira podatke iz naravnega gibanja prebivalstva, in registracija prebivalstva, ki registrira mehansko gibanje prebivalstva.

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

3.14 Dostikrat pa ne razpolagamo za proučevan pojav niti s takimi zapiski. Takrat smo primorani, da podatke o enotah populacije zberemo drugače. Eden izmed takih načinov je *posredno opazovanje* po osebah, ki enoto opazovanja poznajo in morejo dati o njej zahtevane podatke. Posredno opazovanje je zelo razširjeno. Tako dajejo gospodarji kmetijskih gospodarstev podatke o svojih gospodarstvih, gospodinje o porabi v svojih gospo-

ainjstvih, starešine gospodinjstev podatke o članih, obrtniki o svojih podjetjih itd.

3.15 V nekaterih primerih, posebno če gre za eksperimente, pa je vir osnovnih statističnih podatkov *neposredno opazovanje*. Neposredno opazujemo pojave, če sami izdelamo eksperimente oziroma merimo enote opazovanja.

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

En tak primer posrednega in neposrednega opazovanja je določanje donosa po kmetijskih gospodarstvih. S posrednim opazovanjem se zanesemo na izjave gospodarja kmetijskega gospodarstva, pri neposrednem opazovanju pa donos v posameznih gospodarstvih izmerimo. Iz tega primera vidimo tudi prednosti in pomanjkljivosti posrednega in neposrednega opazovanja. Gospodarjevo napoved dobimo zlahka, podatek pa more biti nezanesljiv; merjenje donosa pa da zanesljivejše podatke, je pa bolj zamotano.

NAČINI POSREDNEGA OPAZOVANJA

3.17 Med osebami, ki podatke dajejo, in organi statističnega opazovanja je več vrst stikov. Običajno popisovalec ali anketar *obišče* osebo, ki more dati o proučevani enoti podatke. Tako pri popisih prebivalstva popisovalci obiščejo posamezna gospodinjstva in osebe v svojem popisnem okolišu. Enako pri vzorčenju obiščejo anketarji izbrane enote, da jih popišejo.

3.18 Pri popisovalčevem obisku pa ločimo dve tehniki popisovanja. Pri *samoregistraciji* popisovalec pri prvem obisku razdeli obrazce, razloži tehniko in namen popisa in prosi popisano osebo, da do določenega roka izpolni obrazce: te potem pri ponovnem obisku izpolnjene pobere. Samoregistracija ima svoje prednosti in hibe. Prednost je v tem, da oseba, ki podatke daje, lahko v miru in premišljeno izpolni obrazce, kar je v prid kvalitete, če popisana oseba vpisuje podatke iz dokumentov. Prednost samoregistracije je tudi ta, da popisovalec ali anketar pri prvem obisku obrazce in navodila lahko pusti pri sorodnikih popisane osebe, če popisane osebe ni doma, in naroči, da prizadeta oseba izpolni obrazce do pred-

pisanega roka. V tem primeru sploh ni potreben neposreden stik med popisovalcem in osebo, ki podatke daje. Pomanjkljivost samoregistracije pa je v tem, da je uporabna le tedaj, če je kulturna raven oseb, ki podatke dajejo, tako visoka, da te osebe pravilno in vestno same izpolnijo obrazce.

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

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

3.21 Pri *poštnem načinu* statistično mrežo na terenu zamenja poštna mreža. Poštni način je znatno cenejši kot opazovanje z obiski ali prijavni način, ker odpadejo vsi vmesni organi med osebo ali organizacijo, ki podatke daje, in centralnim organom, ki zbira podatke. Pri poštnem načinu so stroški znatno manjši, organizacija pa enostavnejša kot pri drugih metodah. Enako je tudi čas izvedbe opazovanja pri disciplinirani enotah pri poštni metodi zelo kratek. Hiba poštnega načina pa je v tem, da moremo pri poštnem načinu uporabiti le samoregistracijo, in so osebe, ki dajejo podatke, navezane na pisana navodila brez osebnega stika s statističnimi organi. Zaradi tega je pri poštnem načinu kvaliteta podatkov slaba, odstotek neodgovorov pa v splošnem velik. Poštni način uporabljamo le, če vemo, da so osebe, ki podatke dajejo, disciplinirane in zmožne, da same,

brez pomoči popisovalcev izpolnijo in pošljejo obrazce organizaciji, ki podatke zbira. Primeri uspelih opazovanj po poštnem načinu so tekoče statistične službe v industriji, trgovini, gradbeništvu itd. Podjetja so po službeni liniji zavezana, da v roku pošljejo podatke o svoji dejavnosti, zagotovljena pa je tudi strokovnost izpolnjevanja.

Če opazujemo s poštnim načinom privatne osebe, obrtne obrate, kmetijska gospodarstva itd. običajno kombiniramo poštni način z metodo obiskov tako, da skušamo dobiti po pošti odgovore za čim večji del populacije. Iz enot, za katere nismo prejeli odgovorov, pa izberemo vzorec. Izbrane enote obiščejo anketarji, da dobimo od njih podatke z osebnim kontaktom.

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

3.23 *O korespondentnem načinu* govorimo, kadar imamo za proučevanje določenega pojava na terenu stalno mrežo korespondentov, ki občasno pošiljajo podatke o dejavnosti, za katero so določeni. Korespondenti po strokimi so statistiki in imajo drugo glavno zaposlitev. Običajno pa so korespondenti zaposleni v stroki, iz katere je pojav, ki ga proučujemo. Tako imamo v kmetijski statistiki stalno mrežo korespondentov, ki so izbrani tako, da so razmeščeni po celem teritoriju, ki ga proučujemo. Kmetijski korespondenti so običajno agronomi ali boljši kmetje, ki so zmožni, da ocenjujejo pojave iz kmetijstva in občasno pošiljajo podatke v centralni urad, tam pa podatke obdelujejo. Korespondenti pošiljajo podatke o stanju posevkov, o vegetaciji, o tem, kako kaže donos itd.

ZNAKI OPAZOVANJA

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

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

mo obdelati.

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

3.25 Razen vsebinskih znakov včasih postavljamo še *kontrolne znake*. Kontrolni znaki so v zvezi z vsebinskimi znaki tako, da moremo iz primerjave podatkov s kontrolnimi znaki sklepati na pravilnost odgovorov na osnovna vsebinska vprašanja.

Identifikacijski znaki niso v zvezi z vsebino pojava, ampak jenjin funkcija v tem, da z njimi po potrebi ugotovimo za katero enoto veljajo dani podatki. Identifikacijski znaki so n.pr. ime in priimek, naslov poročevalske enote itd. Ti podatki so potrebni, če iščemo za enoto opazovanja dodatne podatke, popravke itd.

SREDSTVA OPAZOVANJA

3.26 Rezultate opazovanja vnašamo v statistične *obrazce, pobisnice* ali *vprašalne pole*. Na statističnih obrazcih zbiramo podatke o posameznih enotah opazovanja. Po končanem opazovanju izpolnjeni statistični obrazci zamenjajo opazovane enote, ker vsebuje posamezen obrazec vse informacije, ki jih iščemo o enoti.

Statistični obrazec je kos papirja, na katerem so nanizani vsi znaki statističnega opazovanja. Prirejen je tako, da more popisana oseba, popisovalec ali anketar vanj vpisovati značilnosti za popisano enoto.

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

3.27 Vrste obrazcev. Glede na to, ali je obrazec namenjen za popis ene same ali več enot opazovanja, imamo *individualne* in *kolektivne obrazce*. Z individualnim obrazcem popisujemo eno samo enoto, s kolektivnim pa več enot. Enote na kolektivnem obrazcu morejo biti med seboj vsebinsko povezane, ni pa to nujno. Pri popisu prebivalstva kolektiven obrazec za gospodinjstvo združuje podatke vseh članov enega gospodinjstva. V tem pri-

meru je med enotami, ki so popisane na kolektivnem obrazcu, vsebinska zveza. Večkrat pa je kolektiven obrazec namenjen za popis vseh enot enega popisnega okoliša in med enotami v kolektivnem obrazcu ni globlje vsebinske zveze.

Izbira med individualnim ali kolektivnim obrazcem je odvisna od več vzrokov. Kolektiven obrazec moremo uporabiti le pri ekspedicijskem načinu, za samoregistracijo pa ni uporaben. Medtem ko so individualni obrazci primerni za vse vrste obdelave, moremo s kolektivnimi obrazci izvesti obdelavo samo po nekaterih metodah. Če so enote grupirane v kolektivnem obrazcu po kakšnem vsebinskem kriteriju, je kolektiven obrazec primeren za ročno obdelavo, ker moremo na njem neposredno izvesti osnovne operacije obdelave, to je seštevanje podatkov. S kontrolnimi obrazci pri popisu prebivalstva 31.3. 1953, ki so bili obenem kolektivni obrazci za popisni okoliš, smo v kratkem času po popisu dobili z elementarno osnovo obdelavo podatke o številu prebivalstva po spolu in prisotnosti.

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

3.28 Sistem obrazcev. Pri statističnem opazovanju navadno ne uporabljamo samo enega obrazca, temveč *sistem obrazcev*. Vsi obrazci pa imajo en sam namen: ztrati čim popolnejše in pravilnejše odgovore na stavljena vprašanja. Po funkciji delimo obrazce v glavne, pomožne in kontrolne.

Glavni obrazec je običajno v enem opazovanju en sam. Ta obrazec vsebuje listo znakov oziroma vprašanj, na katera mora dati popisana enota odgovor. Glavni obrazec je namenjen za nadaljnjo obdelavo.

Pomožni obrazci pomagajo k čim boljšim in popolnejšim odgovorom na vprašanja v glavnem obrazcu. Pri popisu prebivalstva je n.pr. posebno vprašanje dobiti zanesljive podatke na ekonomska vprašanja, kot so: poklic, položaj v poklicu, panoga dejavnosti ita. Zato smo pri popisu 31.3. 1953 uvedli pomožni obrazec: tega je za vsako zaposleno osebo pred popisom izpolnilo podjetje oziroma ustanova, kjer je bila popisana oseba zaposlena. Popisovalec ali popisana oseba je vnesla odgovore, ki so bili zanesljivejši kakor odgovori, ki bi jih dala posamezna popisana oseba sama, iz pomožnega obrazca v glavni obrazec.

Naloga *kontrolnih obrazcev* je predvsem kontrola polnoštevilnosti izpolnjenih obrazcev. Včasih, kot je bil to n.pr. popis prebivalstva 31.3. 1953, pa rabi kontrolni obrazec tudi kot kolektiven obrazec za osnovne podatke o popisani populaciji.

3.29 Elementi obrazca. Statistični obrazci so izdelani po nekih splošnih načelih, vendar imamo različne variante obrazcev glede na populacijo, način in organizacijo opazovanja in obdelave. Kakor vidimo iz primerov obrazca za popis prebivalstva 31.3. 1953, za mesečno industrijsko službo v

letu 1956 in vprašalnega lista za kmetijsko anketo 15.I. 1958, imajo vsi trije neke skupne črte.

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

Vsak obrazec ima svoj *naslov:* ta v kratkem poda vsebino opazovanja in vlogo obrazca, če ima opazovanje več obrazcev.

Razen z naslovom obrazec označimo s kratko oznako-*kratico.* Tako ima prikazani obrazec popisa prebivalstva oznako „Obrazec PS-1", obrazec kmetijske ankete „Obrazec PA-4", obrazec za mesečno industrijsko poročilo pa „IND-1-1958".

V glavi obrazca je tudi *naziv organa,* ki je obrazec izdal (n.pr. Zvezni zavod za statistiko), *kdo je obrazec odobril,* če ga izdaja organizacija, ki ni statistična, in *naziv in naslov enote,* ki daje podatke.

V obrazcu je naveden tudi *čas opazovanja* (popis prebivalstva 31. 3. 1953, kmetijska anketa 15.I. 1958, mesečno industrijsko poročilo: za mesec ..... 1958), *rok izpolnitve* (mesečno industrijsko poročilo': Rok 3. v mesecu), *število izvodov* (mesečno industrijsko poročilo: v 4 izvodih) in *komu pošlje* poročevalska enota obrazce pri samoregistraciji (mesečno industrijsko poročilo: oddа Zavodu za statistiko tistega okraja, na katerega teritoriju je sedež uprave podjetja).

Obrazec ima običajno tudi mesto za *podpis odgovorne osebe* ali osebe, ki je podatke dala.

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

3.31 Najmanj oseben je obrazec, v katerem *nanizamo znake,* popisane enote pa vnašajo odgovore v ustrezen prazen prostor poleg označitve znaka. Ta tehnika je bila n.pr. uporabljena pri popisu prebivalstva 31.3. 1953 v popisnici v sliki 3.1. Znaka tehnika je uporabljena tudi v vprašalnem li-

stu za kmetijsko anketo v sliki 3.2. Ta metoda je zelo uporabna, ker je enostavna in pregledna. Posebno je priporočljiva pri ekspedicijskem načinu opazovanja.

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

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

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

a) ~~samski~~, poročen, ~~razvezan~~, ~~vdovec~~

b) samski, poročen, razvezan, vdovec

c) samski ☐
poročen ☑
razvezan ☐
vdovec ☐

d) samski 1
poročen (2)
razveden 3
vdovec 4

Včasih kljub temu, da je število možnih odgovorov razmeroma veliko, v obrazcu naznačimo nekaj večjih grup odgovorov; oseba, ki podatke vpisuje, pa vpiše podroben odgovor v okence za ustrezno grupo n.pr.

Šolska izobrazba	Število razredov
brez	
osnovna šola	
srednja šola	8 razredi
visoka šola	

Federativna ljudska republika Jugoslavija
ZVEZNI ZAVOD ZA STATISTIKO IN EVIDENCO

Obrazec PS - 1

Popis prebivalstva 31. marca 1953

Popisnica

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

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

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

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

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

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

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

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

3. ROJSTNI DATUM dan _____ mesec _____ leto _____

4. ROJSTNI KRAJ a) naselje (mesto, vas) _____ b) okraj _____
   c) ljudska republika (osebe, rojene v inozemstvu, navedejo državo) . . . . . _____

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

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

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

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

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

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

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

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

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

Slika 3.1 Prva stran popisnice pri popisu prebivalstva 31.3.1953 v FLR Jugoslaviji

1	2	3	4	5	6	7	8	9	10	11	12
Prvo polletje	Drugo polletje	15 januarja 1959					15.I.1958	15 januarja 1959			
Koliko je bilo l. 1958 porabljeno umetnih gnojil (kg.)		Žrebeta do 2 let	Mladi konji nad 2 leti	Kobile in breje žrebice	Žrebci in skopljenci	Skupno število konj		Delovni konji in kobile	Skupno število mezgov in mul	Skupno število oslov	

ZVEZNI ZAVOD ZA STATISTIKO

VPRAŠALNI LIST

za

KMETIJSKO ANKETO

15-I-1959

Obrazec PA-4

Okraj _____

Občina _____

Naselje _____

Hišna št _____

Številka popisnega okoliša v občini _____

Zap. štev. gospodarstva iz PA-3 _____

Priimek, očetovo ime in ime starešine gospodarstva

Vrsta gospodarstva (gospodinjstva) | Kateg. gosp. | Poraba umetn. gnojil

Skupna lastna površina ha _____ arov _____

Obdelovalna lastna površina ha _____ arov _____

Telcta do 1 leta	Bikci, junčki in junice nad 1—2 let	Mlado govedo nad 2 leti	Krave in breje junice	Plemenski biki	Voli	Skupno število govedi		Krave ki se uporabljajo za delo	V teku l. 1958					
									Oteljeno telet leta 1958	Skupno število kupljenih govedi	zaklano			Skupno število poginule govedi
											Telet (do 1 leta)	Bikcev, juncev in junic nad 1 leto	Odraslih živali	
15 januarja 1959 leta							15.I.1958							
13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

Slika 3.2 Vprašalni list za kmetijsko anketo 15. I. 1959

Zvezni zavod za statistiko, Beograd

Zavod LRS za statistiko, Ljubljana
poštni predal 282 — telefon št. 22-379

IND-1-1958

Mesečno poročilo industrijskega podjetja

za mesec 1958

Polni naziv podjetja ..... telef. št. .....

LR Slovenija — okraj: ..... Kraj: ..... Ulica in št. .....

Ne izpolnjuje podjetje

Leto in mesec 8
Okraj
Občina
Panoga, stroka in vrsta dejavnosti
Sektor lastništva
Grupni znak

Izpolnjuje Zavod za statistiko OLO

Zaporedna številka iz kartoteke organiz. enot

Vsako podjetje, ki je obvezno pošiljati to poročilo, izpolnjuje obrazec v 4 izvodih, katere 6. v mesecu odda Zavodu za statistiko tistega okraja, na katerega teritoriju se nahaja sedež uprave podjetja

Rok 6. v mesecu

V. p.: 324

Tabela I — Izgotovljena proizvodnja, zaloge in realizacija blaga v količinah

Proizvodi po nomenklaturi IND-1-1956		Merska enota	Dosežena izgotovljena proizvodnja		Skupna zaloga konec meseca poročanja	Realizacija blaga od začetka leta do konca meseca poročanja	Pričakovana proizvodnja v sledečem mesecu
Šifra	Naziv proizvoda		v mesecu poročanja	od začetka leta do konca meseca poročanja			
1	2	3	4	5	6	7	8

Slika 3.3 Glava mesečnega poročila industrijskega podjetja v letu 1958

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

3.33 Navodila. Obrazec bi bilo običajno nemogoče pravilno izpolniti, če bi vseboval le listo znakov ali spisek vprašanj. Mnoga vprašanja niso takoj jasna in moramo zanje dati pojasnila v *navodilih za izpolnjevanje obrazca*.

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

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

Navodila za izpolnjevanje so tiskana na glavnem obrazcu pri popisu prebivalstva 31.3. 1953. V sliki vidimo, da so splošna navodila na začetku popisnice, navodila k ustreznim vprašanjem pa pri posameznih vprašanjih.

3.34 Posebnost prikazanega obrazca za popis prebivalstva je v tem, da ima ob desni strani prostor, določen za vnašanje številčnih šifer - oznak za posamezne vrednosti znakov. Šifre so potrebne za strojno obdelavo podatkov.

Pri obrazcu za kmetijsko anketo v sliki 3.2 pa vidimo, da je prostor za vnašanje podatkov o številu živine sistematično nanesen ob robu obrazca in da ima obrazec nenavadno-podolgovato obliko. Ta posebna oblika in prostor za podatke je pogojen z načrtom ročne obdelave podatkov. Seštevanje podatkov je tako znatno poenostavljeno. Obrazce, za katere moramo podatke seštevati, naložimo drugega čez drugega. Tako zlahka seštevamo podatke brez prepisovanja v obdelovalne tabele.

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

KRAJ OPAZOVANJA

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

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

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

Razdelitev teritorija popisa v popisne okoliše mora biti izvedena tako, da moremo iz teh najmanjših popisnih teritorialnih enot sestaviti najmanjše politično-upravne teritorialne enote, t.j. občine. To načelo je zelo koristno, če potrebujemo podatke po upravno-teritorialnih enotah. Če upoštevamo zgornje načelo, moremo iz okolišev brez težav sestaviti podatke za katere koli upravno-teritorialno enoto.

3.37 Kraj opazovane enote more biti v načelu različen od *kraja popisa*. Pri posrednem opazovanju daje podatke oseba, ki ni nujno v istem kraju kot enota opazovanja. Tako daje centrala podatke za svoje podružnice, ki niso v istem kraju, kmetijski posestnik podatke o svojem imetju (živini, zemljiščin itd.), ki ni v istem kraju, kakor sedež gospodarstva.

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

ORGANI STATISTIČNEGA OPAZOVANJA

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

Program statističnega opazovanja sestavljajo *statistični strokovnjaki* v tesnem sodelovanju s konzumenti statističnih podatkov in *strokovnjaki iz področja*, kamor spada konkretno statistično opazovanje. Pri sestavljanju programa sodeluje razmeroma malo visoko kvalificiranih kadrov.

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

Važni sodelavci pri statističnem opazovanju so *popisovalci*. Popisovalci neposredno statistično opazujejo na terenu. Njihov glavni poseb običajno ni statistika, in so izbrani za to delo pri vsakem popisu posebej. Čeprav to niso statistiki, si pri vsakem popisu prizadevamo, da izberemo za popisovalce osebe, ki imajo tako zaposlitev, da jim predmet opazovanja ni tuj. Tako n.pr. pri anketah iz kmetijstva izberemo za popisovalce agronome, pri anketah iz gozdarstva gozdarje ali logarje, pri popisih iz šolstva in prosvete učitelje ali profesorje ita. To znatno pripomore k kvalitetni izvedbi opazovanja. Popisovalce usposobimo za njihov posel na posebnih tečajih ali s podrobnimi navodili za izvajanje statističnega opazovanja, pri katerem sodelujejo.

Vmesni organi med središčem, ki statistično opazovanje vodi in popisovalci, so običajno *kontrolorji*. Ti kontrolirajo delo popisovalcev na terenu in jim pomagajo reševati zamotanejše probleme, ki nastanejo med opazovanjem.

Razen kontrolorjev, ki so bolje poučeni in izvežbani za opazovanje kakor popisovalci, in so običajno, če le mogoče, statistični organi, sestavimo pri velikih popisih *popisne komisije*. Popisne komisije in kontrolorji imajo nalogo pomagati drugim organom pri opazovanju. Obenem popisne komisije zbirajo gradivo od popisovalcev in skrbe za pravilno izvajanje opazovanja na svojem področju.

Razen popisovalcev so organi opazovanja še *osebe, ki dajejo podatke* o opazovanih enotah. To so v vsakem primeru osebe, ki predmet opazovanja najbolje poznajo. Tako dajejo pri popisih prebivalstva starši razen zase še podatke za svoje otroke, gospodarji kmetijskih gospodarstev za svoje gospodarstvo, statistični v industrijskih podjetjih za

svoje podjetje, itd.

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

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

NAPAKE IN KONTROLA STATISTIČNEGA OPAZOVANJA

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

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

3.40 Vrste napak. Napake v statističnem opazovanju morejo biti slučajne ali sistematične. *Slučajne napake* izvirajo iz malomarnega opazo-

vanja in izpolnjevanja obrazcev. Učinek slučajnih napak na rezultate opazovanja ni prevelik, če se napake ne pojavljajo masovno, ker se po zakonu o velikih številih slučajne napake izravnavajo.

Hujši je učinek sistematičnih napak. *Sistematične napake* so napake, ki se pojavljajo iz nekega določenega vzroka s sistematično enakim učinkom na posameznih enotah opazovanja. Učinek sistematične napake se v velikem številu enot ne izravna, temveč kopiči. Zato so sistematične napake nevarnejše kakor slučajne. Vir sistematičnih napak more biti vsak izmed vzrokov, ki smo jih naštelí zgoraj. Tako more biti vzrok sistematične napake nejasno opredeljen pojem ali slabo zastavljeno vprašanje. Če n.pr. iščemo po podjetjih vrednost proizvodnje v določenem razdobju, a ne določimo natančno, kaj imamo za vrednost industrijske proizvodnje, more biti to vir sistematične napake, ker podjetja v dobri veri vključujejo ali izključujejo neke od svojih dejavnosti v vrednost proizvodnje.

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

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

3.41 Kontrola napak. Zato je potrebna dobra in učinkovita *kontrola* samega opazovanja in kontrola zbranega gradiva, da dobimo čim boljše podatke o opazovanem pojavu.

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

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

Kontrola *polnoštevilnosti zajetja* se omeji na pregled gradiva in številčno primerjavo zbranih obrazcev s kontrolniki, spiski oziroma registri enot. Ta kontrola odkrije ali kake enote niso popisane in ali so neke enote popisane večkrat.

Kontrola o *polnoštevilnosti odgovorov* je preprosta. V tej stop-

nji kontrole obrazce pregledamo, ali so na vsa stavljena vprašanja vpisani odgovori.

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

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

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

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

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

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

Podatke, ki so verjetni, vendar nepravilni, more odkriti le dober poznavalec krajevnih razmer ali oseba, ki popisano enoto pozna. Te napake odkrije tudi ponoven popis, ki ga izvedemo bolj natančno z dokumenti, pregledom itd. Jasno je, da taka kontrola ne pride v poštev, ker je pravzaprav ponovna izvedba opazovanja s strožjimi merili.

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

3.42 Podatke kontrolirajo organi opazovanja v vseh stopnjah. Kontrola podatkov na vsaki stopnji ima svoje pomanjkljivosti in prednosti.

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

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

Četrto poglavje

UREJEVANJE STATISTIČNEGA GRADIVA

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

Osnovna obdelava statističnih podatkov je urejevanje statističnih podatkov. Urejevanje sestoji iz grupiranja, preštevanja in seštevanja statističnih enot in podatkov. Namen urejevanja so absolutni podatki o statistični populaciji tako, da jih je z analitičnimi metodami možno obdelati in analizirati. Z urejevanjem dobimo osnovne statistične vrste, ki nudijo uvid v sestav populacij in zveze med posameznimi pojavi. Tako pri popisu prebivalstva z urejevanjem ugotovimo sestav prebivalstva po spolu, stanu, starosti, šolski izobrazbi, poklicu itd., skratka po vseh znakih in kombinacijah znakov, ki smo jih zbrali s popisom. Rezultate urejevanja prikazujemo s statističnimi vrstami v statističnih tabelah. Statistična vrsta je osnovni način v podajanju statističnih podatkov, tabela pa eden izmed načinov za prikazovanje statističnih vrst.

Urejevanje statističnega gradiva sestoji iz več različnih stopenj, ki so med seboj vezane časovno in vsebinsko. Urejevanje statističnega gradiva je za velika opazovanja znaten tehnični posel, ki traja razmeroma dolgo. Ves napor za izboljšanje tehnike urejevanja je usmerjen v to, da skrajšamo čas urejevanja. Vrednost podatkov določenega opazovanja je tem večja, čim prej so podatki po izvedenem opazovanju na razpolago. Obdelavo pospešimo, če zaposlimo več obdelovalcev. Vendar je število obdelovalcev omejeno, ker so potrebne za urejevanje razmeroma kvalificirane osebe. Sedaj na splošno uvajajo uporabo modernih sredstev strojne obdelave podatkov. Ta je pri opazovanjih večjega obsega že skoraj izpodrinila ročno obdelavo. Vendar so v urejevanju določena dela, ki se ne dajo mehanizirati.

4.2 Po kontroli statističnega gradiva sestoji urejevanje statističnega gradiva iz različnih stopenj dela. Plan urejevanja mora določiti in rešiti vprašanje za izvedbo naslednjih stopenj urejevanja:

a) Grupiranje vrednosti znakov.
b) Šifriranje oziroma signiranje zbranih statističnih podatkov.
c) Načrt za osnovno obdelavo.
d) Način obdelave podatkov.

Kakor smo že nakazali pri obravnavanju znakov, imajo nekateri znaki

razmeroma veliko,, včasih tudi neomejeno število vrednosti. Naloga *grupiranja* je združiti sorodne vrednosti znakov, n.pr. sorodne poklice, sorodne artikle ita., v grupe tako, da število grup ni preveliko. Ti grupni znaki dajo sicer bolj grobo sliko, vendar so zaradi razmeroma manjšega števila grup bolj prikladni za obdelavo in dajo preglednejšo sliko o pojavu. Grupiranje znaka je ozko povezano z vsebino in nameni proučevanja.

Čisto tehničnega značaja pa je *šifriranje* oziroma signiranje podatkov. Da je nadaljnja obdelava podatkov čimbolj avtomatična, posamezne podatke v gradivu zaznamujemo s kratkimi oznakami - šiframi, pri nadaljnji obdelavi pa se oziramo le na šifre in ne več na osnovne podatke.

Načrt obdelave določa, po katerih kombinacijah znakov ali grupnih znakov bomo obdelali gradivo, da bo dalo odgovor na dano proučevanje. Pri popisu prebivalstva ne iščemo strukture prebivalstva samo ločeno po spolu in starosti, temveč je pomembna struktura, ki razdeli prebivalstvo po spolu in starosti istočasno. Tako dobimo uvid ne le v spolni in starostni sestav, ampak v spolno-starostni sestav prebivalstva. Kombinirana obdelava dveh ali več znakov istočasno je ena izmed osnovnih metod obdelave in analize podatkov.

Načrt obdelave mora glede na namen proučevanja določiti: katere kombinacije znakov so za proučevanje pomembne in ali nameravamo po teh kombinacijah znake preštevati ali seštevati.

V načrtu osnovne obdelave je treba tudi določiti, kako bomo obdelovali statistične podatke. Že pri sestavljanju obrazcev (glej primere obrazcev v slikah 3.1, 3.2, 3.3), posebno pa pri šifriranju podatkov, je treba upoštevati, ali nameravamo obdelovati podatke ročno ali strojno.

Rezultati urejevanja so prikazani v obdelovalnih tabelah, ki so prikrojene načinu obdelave. Iz obdelovalnih tabel šele sestavimo končne publikacijske ali analitične tabele, katerih namen je prikazati populacijo in dati osnovo za analizo pojava.

GRUPIRANJE VREDNOSTI ZNAKOV

4.3 Grupne karakteristike uporabljamo že *v vsakdanjem govoru* in izražanju. Tako izražamo starost z zaokroženo starostjo v letih. Če zaokrožujemo starost na izpolnjena leta, vzamemo za 26 let stare vse tiste, ki so stari več kot 26 in manj kot 27 let. Če pa zaokrožujemo starost na najbližje leto, štejemo za 26 let stare vse, ki so stari od 25 in pol in 26 in pol let. Vse, katerih starost je v navedenem razmaku, štejemo da so stari enako - 26 let. Prav tako tudi plače ne navajamo v parah, marveč v tisočih.

Znako v pojmu delavec ali uradnik združujemo vse delavske ali uradniške poklice. Če za nekoga pravimo, da je delavec, zanj sicer ne vemo natančno, kakšen je njegov poklic; vemo pa, v kateri grupi poklicev je.

Vsi delavski poklici so združeni pod skupnim pojmom delavec. Tudi časovno v vsakdanjem govoru uporabljamo grupe in rečemo: ta in ta nesreča se je zgodila v nedeljo 25. I. 1959. Tako je čas dogodka za dane potrebe zadosti določen, čeprav se je mogla po teh podatkih nesreča dogoditi v razmaku 24 ur. Včasih zadoščajo še širše grupe. Na vprašanje, kdaj se je neka oseba poročila, ta odgovori: v letu 1936. Ta podatek, čeprav je dan z razmakom enega leta, povsem zadošča za opredelitev časa poroke. Enako ravnamo tudi s krajevnimi opredelitvami. Včasih na vprašanje odkod smo, odgovorimo: iz Ljubljane, ne navedemo pa ulice, hišne številke, nadstropja in stopnišča. Ko rečemo, da smo iz Ljubljane, povemo, da je naše stanovanje kjer koli na območju Ljubljane. Da so iz Ljubljane, odgovorijo na vprašanje, odkod so, vsi, ki so doma na območju Ljubljane.

Iz zgornjih primerov vidimo, da v vsakdanjem življenju in govoru ne zaokrožujemo samo numeričnih znakov, ampak na splošno vse. V kako velikih grupah se izražamo, je odvisno od potrebe in situacije, v kateri smo. Včasih je potrebno, da plačo navedemo natančneje kakor v tisočih dinarjev. Včasih ni dovolj, če kdo reče, da je uradnik, in je treba poklic natančneje določiti. Enako je s krajem stalnega bivališča. Če vas v Ameriki nekdo vpraša, odkod ste, je dovolj, da rečete, da ste iz Jugoslavije, čeprav je to razmeroma širok pojem. Če vas pa nekdo, ki ve, da ste Ljubljančan, v Ljubljani vpraša, kod ste doma, morate navesti ulico; torej niti opredelitev, da ste Ljubljančan, ni zadostna.

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

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

Grupiranje pri vseh znakih opazovanja mora biti izvedeno *enolično*. Vse vrednosti znaka grupiramo v grupni znak tako, da vsaka osnovna vrednost znaka spada v eno in samo eno grupo. Napačno bi bilo grupiranje, če bi določena vrednost bila v dveh ali več grupah istočasno, druga vrednost pa ne bi bila v nobeni grupi. To načelo zahteva, da smo posebno pazljivi pri mejnih vrednostih. Tako je pri časovnih grupah za dneve problem, kam spada moment ob koncu enega dne in v začetku drugega dne. Pri geografskih grupah je večkrat sporno, kam spadajo točke, ki so teoretično na meji med dvema geografskima grupama itd.

Grupe vseh znakov imajo tudi to lastnost: grupe znakov moremo

grupirati v *grupe višje vrste* enako, kakor grupiramo osnovne vrednosti znakov. Tako moremo občine, ki so grupe prve stopnje, grupirati v okraje, ki so grupe druge stopnje, te v republike kot grupe tretje stopnje itd. Enako moremo ure združevati v dneve, te v mesece, mesece v leta, leta v dekade. Prav tako grupiramo tudi stvarne znake. Posamezne poklice grupiramo v najožje grupe poklicev, te grupe v širše grupe itd. Ta postopek ponavljamo, dokler grupiranje ne doseže do najširših grup oziroma do grupe, ki obseže vse vrednosti. Vsaka podrobnejša grupacija v tem stopnjevanju grup podrobneje določa poklice. Enako grupiramo predmete porabe, vzroke smrti itd.

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

Razen tega pri vsakem grupiranju vsaka grupa običajno dobi neko *novo ime - vrednost*, ki pojmovno združuje vse vrednosti znakov za ustrezno grupo. Tako imajo svoja posebna imena: geografske grupe n.pr. imena občin, okrajev itd., časovne grupe datume, mesece, leta itd., stvarne grupe nazive grup poklicev, itd.

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

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

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

4.7 Za *časovne znake* je sorodnost momentov običajno dana s časovnim razmakom med momenti. Po tem načelu grupiramo momente v naravne enote: minute, ure, dneve, tedne, mesece, leta. Izmed teh grup so najbolj problematični meseci, ker so različno dolgi (od 28 do 31 dni). Razlika med najkrajšim in najdaljšim mesecem je 3 dni ali približno 10 %. Še večje so razlike, če namesto koledarskih dni upoštevamo v grupo samo delovne dni. S posebnimi prijemi reduciramo podatke tako, da lahko mesečne podatke primerjamo med seboj.

Za proučevanje nekaterih pojavov uporabljamo drugačna načela grupiranja časa. Za proučevanje periodičnosti v časovnem dogajanju so si sorodnejši vsi januarji, vsi februarji, vsi marci vseh let v določenem razdobju kakor pa januar, februar, marec itd. istega leta, čeprav so prvi časovno bolj oddaljeni. Če proučujemo na primer turizem, gradbeništvo, poljedelstvo, nam je takoj jasno, da je res tako. Pogoji za vsako od teh dejavnosti so v istih mesecih v zaporednih letih bolj sorodni kot pogoji v različnih mesecih istega leta. Enak primer je grupiranje dni v tednu, če proučujemo promet, prometne nesreče, porabo pijač itd. V tem primeru tvorimo grupe vseh ponedeljkov, vseh torkov itd. Enako tvorimo grupe iz določenih ur posameznih zaporednih dni, če proučujemo nihanje v dnevni porabi električne energije ali vode, v lokalnem prometu itd.

4.8 Grupiranje *stvarnih znakov* je bistveno različno za atributivne in numerične znake. Grupiranje atributivnih znakov je najtežje. Čeprav postavimo za grupiranje atributivnih znakov n.pr. poklice, vzroke smrti, dejavnost, artikle v zunanji trgovini, surovine v industriji itd., neko načelo grupiranja, s tem grupiranje še ni izvedeno, kakor je n.pr. izvedeno krajevno grupiranje, če se odločimo za upravno-teritorialni princip ali če določimo načelo grupiranja za časovne znake. Za stvarno-atributivne znake so z načelom postavljene samo smernice za grupiranje. Grupiranje posameznih poklicev, vzrokov smrti, dejavnosti itd. pa je zelo obširno delo, ki ga morejo opraviti le strokovnjaki. Ti po načelih grupiranja

sistematično klasificirajo vse mogoče posamezne poklice, vzroke smrti, dejavnosti itd., in izdelajo podrobne klasifikacije - nomenklature, v katerih je za vsak poklic, vzrok smrti, artikel itd. navedeno, v katero grupo spada. Klasifikacije in nomenklature so torej sistematično po grupah in podgrupah urejeni seznami poklicev, vzrokov smrti, artiklov itd.

4.9 Za atributivne znake z velikim številom vrednosti običajno grupacijo tehnično izvedemo po *decimalni klasifikaciji*. Osnova decimalne klasifikacije je v tem, da največ po deset osnovnih pojmov združimo v grupe prve stopnje, največ po deset grup prve stopnje v grupe druge stopnje itd. To grupiranje v grupe višjih stopenj ponavljamo toliko časa, da imamo največ deset grup zadnje stopnje. Ta princip ni vsebinskega, temveč tehničnega značaja. Vsak pojem moremo namreč po decimalni klasifikaciji označiti s številom, ki ima toliko mest, kolikor stopenj ima grupiranje.

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

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

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

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

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

Iz primera vidimo smisel in pomen decimalne klasifikacije. Medtem ko prva številka v oznaki 631-01-11, s katero je zaznamovan normalen bukov furnir (6), pove, da normalen bukov furnir spada po klasifikaciji v

grupo izdelkov, klasificiranih pretežno po materialu, prvi dve številki (33) pomenita, da spada ta artikel v podgrupo izdelkov iz lesa in plute (brez pohištva). Prve tri (331) številke oznake pomenijo, da spada navedeni artikel v nadaljnjo podgrupo furnirjev, vezanih plošč, desk itd.

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

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

4.11 Grupiranje *stvarno-numeričnih znakov* mora ustrezati vsebinskim in tehničnim pogojem. Ker je analiza numeričnih znakov najbolj razvita, se pri grupiranju numeričnih znakov oziramo na formalno-tehnična načela grupiranja, če je taka grupacija ni nedopustna iz vsebinskih razlogov.

Če vzamemo za primer grupiranje zveznega numeričnega znaka skupno površino privatnih kmetijskih gospodarstev, smo imeli pri popisu zemljiških gospodarstev v letu 1947 tele grupe:

nad 0,00 ha do 0,50 ha
nad 0,50 do 1,00 ha
nad 1,00 do 2,00 ha
nad 2,00 do 3,00 ha
nad 3,00 do 5,00 ha
nad 5,00 do 8,00 ha
nad 8,00 do 10,00 ha
nad 10,00 do 15,00 ha
nad 15,00 do 20,00 ha
nad 20,00 do 30,00 ha
nad 30,00 do 45,00 ha
nad 45,00 ha

Ta grupacija je bila pogojena s stanjem v skupni zemljiški posesti posameznih kmetijskih gospodarstev v letu 1947. Jasno je, da je bila grupacija njivske površine drugačna in ta grupacija ne ustreza za njivsko površino. Enako ta grupacija danes glede na velikost privatnih kmetijskih gospodarstev ne ustreza več. Iz tega vidimo, da je grupiranje določenega znaka pogojeno z vsebino pojava.

4.12 Če pogledamo posamezne grupe, ki jih za numerične znake imenujemo *razrede*, vidimo, da je vsak razred določen s svojo *spodnjo mejo* $x_{k,min}$ in svojo *zgornjo mejo* $x_{k,max}$. Nobena vrednost v razredu ni manjša od spodnje meje razreda in nobena večja od zgornje meje razreda. V posamezen razred spadajo vse vrednosti med spodnjo in zgornjo mejo razreda. Vsak ra-

zred ima svojo *širino razreda* i_k, ki je razlika med zgornjo in spodnjo mejo razreda

$$i_k = x_{k,max} - x_{k,min} \tag{4.1}$$

in *sredino razreda* x_k, ki je polovica vsote mej razreda

$$x_k = \frac{x_{k,min} + x_{k,max}}{2} \tag{4.2}$$

Kakor smo naveali, ima na splošno vsaka grupa neko svojo karakteristiko. Tako vzamemo, da je pri numeričnih znakih sredina razreda reprezentant vseh vrednosti razreda.

Iz zgornjega primera vidimo, da niso vsi razredi omejeni s svojo spodnjo in zgornjo mejo. Tako ima zadnji razred samo spodnjo mejo(45,00ha). Take razrede, ki so omejeni samo navzdol ali samo navzgor, imenujemo *odprte razrede*. Odprte razrede tvorimo takrat, če je število enot, ki ima vrednosti nad ali pod neko mejo, majhno, podatki pa med seboj zelo različni.

Za zgornjo grupacijo so širine razredov po vrsti: 0,50 ha, 0,50 ha, 1,00 ha, 1,00 ha, 2,00 ha, 3,00 ha, 2,00 ha, 5,00 ha, 5,00 ha, 10,00 ha, 15,00 ha. Za zadnji razred ne moremo izračunati širine, ker je odprt. Sredine razredov pa so tele: 0,25 ha, 0,75 ha, 1,50 ha, 2,50 ha, 4,00 ha, 6,50 ha, 9,00 ha, 12,50 ha, 17,50 ha, 25,00 ha, 37,50 ha. Za zadnji odprt razred tudi sredine razreda ne moremo izračunati.

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

Če hočemo meje razredov vključiti v zgornji razred, pa moramo pisati grupacijo takole:

0,00 ha do pod 0,50 ha
0,50 ha do pod 1,00 ha
1,00 ha do pod 2,00 ha
ita.

Obe zgornji grupaciji sta enolični in kompletni. Grupacija

0,00 ha do 0,50 ha
0,50 ha do 1,00 ha
1,00 ha do 2,00 ha
ita.

pa ni enolična, ker mejne vrednosti spadajo v dva razreda.

Površine gospodarstev ne merimo natančno do mm^2, marveč jih običajno zaokrožujemo na are. Kakor smo že omenili, pa je zaokroževanje numeričnih znakov dvojno: zaokrožujemo jih na *najbližjo* celo vrednost ali pa na

največjo celo vrednost v enoti, v kateri navajamo podatek. Tako gospodarstvo, ki ima 3 ha 27 a 58 m^2, po prvem načinu zaokrožujemo na 3 ha 28a, ker je stvarna površina bliže 3 ha 28 a kakor površini 3 ha 27 a, po drugem načinu pa je zaokrožena vrednost 3 ha 27 a, ker je 3 ha 27 a največja površina, navedena v arih, ki je manjša kakor stvarna površina gospodarstva.

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

0,00 ha do 0,49 ha
0,50 ha do 0,99 ha
1,00 ha do 1,99 ha
itd.

Na prvi pogled se zdi, da zgornja grupacija ni kompletna, ker je zgornja meja naznačena n.pr. z 0,49 ha, spodnja meja naslednjega razreda pa z 0,50. Če pa upoštevamo, da so to zaokroženi podatki, ni nobenega dvoma o kompletnosti te grupacije. Vendar je od tega, kako zaokrožujemo, odvisno, katere vrednosti so meje razredov. Če zaokrožujemo na najbližjo celo vrednost v najmanjši enoti, zaokrožena vrednost 0,49 pomeni vse površine od 0,485 do 0,495 ha, zaokrožena vrednost 0,50 ha pa vse površine od 0,495 do 0,505 ha. Meja med prvim in drugim razredom je torej 0,495 ha. Če pa zaokrožujemo na največjo celo vrednost v najmanjši enoti, pa vrednost 0,49 pomeni vse površine od vključno 0,49 do pod 0,50, vrednost 0,50 ha vse površine od vključno 0,50 do pod 0,51 itd. Meja razreda pri tem zaokroževanju je 0,50 ha. Iz primera vidimo, da način zaokroževanja vpliva na meje razredov in po njih na sredine razredov. V splošnem da naravnejše razrede zaokroževanje na največjo celo vrednost, kar vidimo že iz našega primera. Pri prvem načinu zaokroževanja dobimo meje razredov 0,495 ha, 0,995 ha, 1,995 ha itd., pri drugem pa 0,50 ha, 1,00 ha 2,00 ha itd.

V zgornjem primeru smo iz vsebinskih razlogov vzeli različne širine razredov. To je prav posebno vidno iz primera meje 8,00 ha; ta je vzeta v grupacijo zato, da smo mogli pregrupirati podatke v grupe, ki so bile osnova pri agrarni reformi.

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

Tako na primer plače grupiramo v grupe

v tisočih dinarjev
nad 5 do 10
nad 10 do 15
nad 15 do 20
nad 20 do 25
nad 25 do 30
nad 30 do 35
nad 35 do 40
nad 40 do 45

Širina razreda je v tej grupaciji enaka (pet tisoč dinarjev). Zato se tudi sredine razredov 7,5; 12,5; 17,5; 22,5; 27,5; 32,5; 37,5; 42,5 večajo v aritmetični postopici.

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

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

4.14 Sorodnost vrednosti dveh numeričnih vrednosti merimo z razliko med vrednostima. Po tem merilu sta plači 7000 in 8000 enako različni kot plači 37000 in 38000 dinarjev. Na prvi pogled pa spoznamo, da je razlika med plačama 7000 in 8000 dinarjev večja kot med plačama 37000 in 38000 dinarjev, čeprav je v absolutnem razlika v obeh primerih tisoč dinarjev. Za merilo razlik in sorodnosti ni vedno ugodna absolutna razlika, temveč je boljša relativna razlika. Kadar so možne velike razlike med vrednostmi znaka, so primernejše merilo o sorodnosti vrednosti relativne kakor pa absolutne razlike. Za take primere sestavljamo razrede, za katere niso razlike, ampak kvocienti mej, konstante. Meje razredov za tako grupacijo niso aritmetična, ampak *geometrična postopica*.

Primer take grupacije v naši statistiki je grupiranje industrijskih podjetij po številu delavcev.

Podjetja so po številu delavcev grupirana v tele grupe:

število delavcev	kvocient mej	število delavcev	kvocient mej
do 15		126 do 250	
	2,00		2,00
16 do 30		251 do 500	
	2,00		2,00
31 do 60		501 do 1000	
	2,08		?
61 do 125		1001 in več	

Kvocienti med mejami so stalni (2,00) razen v razredu 31 do 125, v katerem je 2,08, ker so meje prešle na naravnejše meje.

ŠIFRIRANJE OSNOVNEGA GRADIVA

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

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

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

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

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

Tudi *črkovne šifre* so prikladne, posebno če je število grup manjše. V tem primeru je dobro vzeti za šifre prve črke oznak za vrednosti ali grupe. Tako vzamemo za stan šifre: samski -S, poročen -P, razvezan -R in vdovec -V; ali za spol: moški -M, ženski -Ž. Za take primere so črkovne šifre primernejše od številčnih, ker si jih laže zapomnimo.

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

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

OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

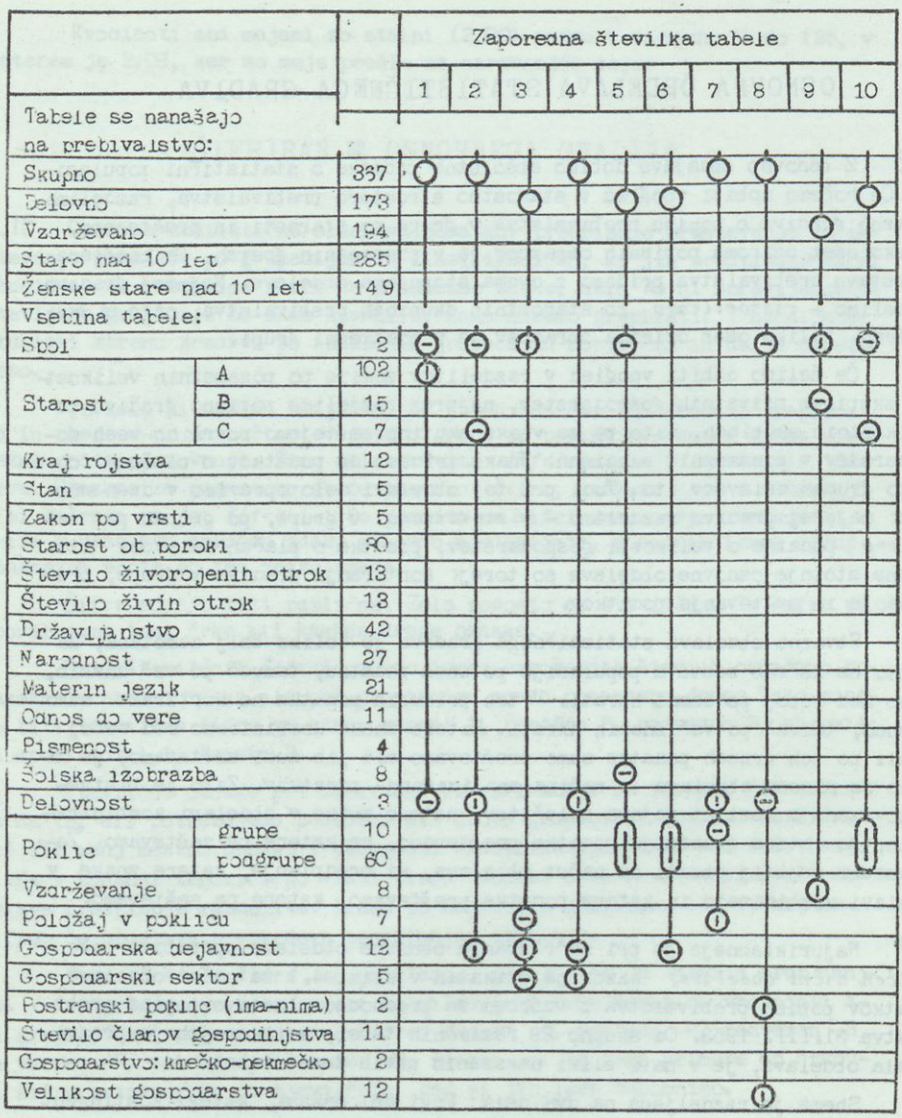
4.16 Z osnovno obdelavo dobimo absolutne podatke o statistični populaciji. Če hočemo dobiti vpogled v starostno strukturo prebivalstva, razdelimo popisno gradivo o popisu prebivalstva v grupe po starosti in preštejemo, koliko oseb oziroma popisnih obrazcev je v posameznih grupah. Do pregleda o sestavu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

Če želimo dobiti vpogled v razdelitev zemlje po posameznih velikostnih skupinah privatnih gospodarstev, najprej razdelimo popisno gradivo po velikostnih skupinah, nato pa za vsako skupino seštejemo površino vseh gospodarstev v posameznih skupinah. Enako pridemo do podatkov o plačnem fondu po grupah delavcev itd. Tudi pri tej obdelavi delo opravimo v dveh stopnjah: najprej gradivo razdelimo - *razsortiramo* v grupe, po grupah pa *seštejemo* podatke o velikosti gospodarstev, podatke o plačah delavcev itd. Glavne stopnje osnovne obdelave so torej: sortiranje gradiva v grupe, preštevanje in seštevanje podatkov.

Stvarna obdelava statističnega gradiva je toliko bolj zamotana, da navadno ne iščemo sestava populacije po enem podatku, temveč po več znakih, ki so med seboj povezani, hkrati. V teh primerih podatke ne sortiramo samo po enem, temveč po več znakih hkrati. Katere znake kombiniramo med seboj in ali po teh grupah podatke samo preštevamo ali jih tudi seštevamo, je odvisno od namena obdelave in načina analiziranja podatkov. Zato je bistveno povezana z vsebino pojava odločitev, katere znake v obdelavi kombiniramo in, za katere grupacije podatke preštevamo, za katere pa seštevamo. Zato moramo najbolj paziti na načrt obdelave, ki predpisuje, katere znake v obdelavi kombiniramo in katere podatke preštevamo, katere pa seštevamo.

4.17 Najpriklacneje je pri načrtovanju osnovne obdelave uporabiti *shematičen načrt obdelave*, kakor je prikazan v sliki 4.1 za načrt obdelave podatkov popisa prebivalstva z vzorcem za predhodno obdelavo popisa prebivalstva 31.III. 1953. Od skupno 29 različnih tabel, kolikor jih je predvidevala obdelava, je v naši sliki nakazanih prvih deset tabel.

Shema je razdeljena na dva dela: Prvi del nakaže, kateri kontingent prebivalstva se nanaša na posamezno tabelo, drugi del pa se nanaša na vsebino tabele. V tem delu so naznačeni znaki popisa, v posebni koloni pa število grup za vsak posamezen znak. V shemi je vsaka tabela ponazorjena z več krogci, ki so med seboj povezani z navpično črto. Krogci so postavljeni pri tistih znakih, ki so kombinirani v posamezni tabeli. Tako vidimo, da je v prvi tabeli predvideno, da skupno prebivalstvo razdelimo po spolu, po starosti po najpodrobnejši grupaciji z 102 grupami, in aktivnosti, ki ima tri grupe. Z znamenji ⊖ in ⓘ v krogcih je naznačeno, ali je posamezen znak v glavi ⊖ ali čelu ⓘ tabele. Analogno moremo iz sheme razbrati vsebino tabele za druge obdelovalne tabele. Ker pri popisu prebivalstva osno-



○ Kontingent prebivalstva

Ⓘ Znak v čelu tabele

⊖ Znak v glavi tabele

Slika 4.1. Shema načrta obdelave vzorca predhodnih rezultatov popisa prebivalstva 31.III.1953. (Vir: Popis stanovništva 1953 knjiga XI.)

vna obdelava sestoji samo iz preštevanja gradiva po posameznih grupacijah, ni potrebno, da pri vsaki izmed prikazanih tabel to posebej navedemo. Če pa je obdelava tolj raznolika, pa razen krogcev uporabljamo še druge dogovorne znake in simbole: trikotnike, kvadrate itd. Z njimi naznačimo posamezne operacije: preštevanje, seštevanje ali elementarna preračunavanja.

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

Shema načrta obdelave pa v končni obliki ne nakaže samo obdelave gradiva, marveč tudi, kakšne so obdelovalne tabele, v katere vnašamo podatke obdelave. Za tabelo 1 v shemi načrta obdelave vzorca predhodnih rezultatov popisa prebivalstva 31.III. 1953 je obdelovalna tabela narisana v sliki 5.8 v poglavju o prikazovanju statističnih podatkov. Skladno s shemo, v kateri je pri tabeli 1 vodoravna črtica postavljena v krogcih pri spolu z dvema grupama in aktivnosti s tremi grupami, navpična črtica pa v krogcu pri starosti z 102 grupami, je v obdelovalni tabeli v sliki 5.8 spol in aktivnost v glavi, starost pa v čelu tabele.

4.18 *Seštevanje statističnih podatkov* ni brez zapletljajev. Neposredno moremo seštevati le istovrstne podatke. Če seštevamo dnevno proizvodnjo za določeno vrsto cigaret za določeno podjetje, dobimo mesečno proizvodnjo cigaret te vrste. Če seštevamo mesečne plače za posamezne uslužbence, dobimo skupaj fond plač za vse uslužbence v določeni ustanovi itd. V navedenih primerih je seštevanje podatkov smiselno in dobimo s seštevanjem podatek za neko novo enoto. Iz dnevne proizvodnje dobimo mesečno proizvodnjo, iz posameznih plač fond plač za ustanovo.

Problem pa se zamota že pri seštevanju istovrstnih količin. Vzemimo število kilometrov, ki jih prepotuje v določenem razdobju posamezen potnik. Če seštejemo število kilometrov, ki jih je prepotoval posamezen potnik, za vse potnike, dobimo vsoto, katere pomen ni več tako jasen kakor v zgornjih primerih. S seštevanjem poti posameznih potnikov dobimo skupno število kilometrov, ki so jih prepotovali vsi potniki. Ta vsota pa se zdi, da je brez pomena. Razumljivo je, da je potnik, ki je prepotoval prvi dan 100 kilometrov, drugi dan 130 km, tretji dan 150 km, skupno prepotoval 380 kilometrov, manj jasna pa je vsota, če seštevamo poti različnih potnikov. Vendar imajo te vrste podatkov v transportni statistiki velik pomen. Da se pa izognemo napačnemu pojmovanju tega podatka, damo vsoti poti vseh potnikov novo enoto mere - potniški kilometri. Če je prvi potnik prepotoval 30 kilometrov, drugi 35 kilometrov, tretji pa 34 kilometrov, je podjetje, kjijh je prevozilo, prevozilo 129 potniških kilometrov. Po enakem sklepu dobimo v transportni statistiki vlakovne kilometre, če seštevamo dolžino poti posameznih vlakov: če seštevamo prevožene kilometre, ki so jih prevozili posamezni železniški vozovi, dobimo vozovne kilometre itd. Tudi v industrijski statistiki naletimo na podobne probleme. Število ur, ki jih je v določenem mesecu delal posamezen delavec, seštevamo v skupno število delavec - ur ali skupno število delovnih ur v podjetju. Isti problem imamo, če

seštevamo čas dela pri strojih; tu dobimo s seštevanjem ur dela za posamezen stroj strojne ure za podjetje ita.

Medtem ko imajo zgoraj navedene vsote smisel, čeprav ni neposreden, seštevanje nekaterih istovrstnih podatkov nima smisla. Tako je brez smisla vsota starosti posameznikov za določeno populacijo, vsota cen za določeno vrsto blaga za različne trge ita.

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

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

4.20 Raznovrstne količine, ki imajo različne enote mere, na splob ne moremo seštevati. Podjetje, ki proizvaja različne artikle, svoje mesečne proizvodnje ne more sešteti in izraziti v enem podatku. Enako trgovina ne more izraziti svojega skupnega prometa v naturalnih enotah mere, ker prodaja popolnoma različne artikle.

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

$$V = \Sigma\, p.q \qquad (4.3)$$

Pri tem pomeni: V = skupna vrednost, p = cena za posamezen artikel, q = = ustrezna količina, Σ = znak za seštevanje.

4.21 Načelo, da posamezne *količine tehtamo* glede na njihov pomen in vrednost, kakor ga uporabljamo pri sestavljanju skupne vrednosti, pri kateri količine tehtamo s ceno, v ekonomski statistiki razširimo.

Za premog moremo uporabiti za težo ali ponder ceno posameznih vrst

premoga, ker je cena skladna s kvaliteto premoga. Vendar je objektivnejše merilo kakovosti kalorična vrednost premoga. Po tem načelu vzamemo za enoto tono premoga z določeno kalorično vrednostjo. Ta premog, ki ga označimo z 1, imenujemo pogojno enoto. Premog, za katerega je kalorična vrednost samo tri četrtine kakovosti pogojne enote, označimo s koeficientom 0,75, premog, ki ima za 20 % višjo kalorično vrednost, pa s koeficientom 1,20 itd. Podobno kakor smo dobili skupno vrednost, če smo produkte količin s cenami seštelí, izrazimo proizvodnjo premoga v pogojnih enotah premoga določene kakovosti tako, da proizvodnjo za posamezno vrsto premoga pomnožimo s ustreznim koeficientom, te produkte pa seštejemo. Dobljena vsota pomeni, koliko tonam pogojne kakovosti ustreza stvarna proizvodnja premoga. Ta podatek ima večji ekonomski pomen kakor pa navadna vsota proizvodnje za vse vrste premoga ne glede na kakovost.

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

Podobno tudi v kmetijski statistiki seštevamo raznovrstne živali. Kot pogojno žival vzamemo 400 kg težko govedo. Vse druge živali izražamo s koeficienti v razmerju s pogojno-normalno živaljo. Tako ponderiramo posamezne vrste živali glede na velikost in proizvodno moč. Ta način omogoča, da seštevamo govejo živino, konje, prašiče, ovce itd. v skupno število normalnih živali. Če vzamemo za primer konje, imamo za posamezne skupine tele koeficiente: Žrebeta do enega leta 0,6; žrebeta od enega leta do treh let starosti 1,0; kobile nad tri leta 1,3; žrebički 1,3; konji nad tri leta 1,5. Če je na nekem področju 100 žrebet do enega leta, 200 žrebet od enega leta do treh let, 250 kobil nad tri leta, 50 žrebcev in 400 konj nad tri leta, je skupno: 100.0,3 + 200.1,0 + 250.1,3 + 50.1,3 + + 400.1,5 = 1250 normalnih živali. Čeprav je ta podatek fiktiven, ga moremo s pridom uporabljati za primerjave med območji, ker v enem podatku izraža vse živali na določenem območju, v določenem velikostnem razredu itd.

Ročna obdelava

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

4.22 *Pri odlaganju* obrazcev popisne obrazce razdelimo v grupe tako,

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

Številčen pregled o razdelitvi populacije v grupe po predpisani kombinaciji znakov dobimo, če *preštejemo* po grupah število enot oziroma obrazcev in rezultate vpišemo v obdelovalno tabelo. Če je v obdelavi predpisano seštevanje določenega podatka po grupah, *seštejemo* ustrezen podatek za vse enote v posameznih grupah. Tako moremo seštevati površino, donos, število posameznih vrst živine v obrazcih, ki smo jih prej razsortirali po velikostnih skupinah po skupni površini. Pri tem ne štejemo, da je obdelava strojna, če podatke seštevamo z običajnimi računskimi stroji.

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

Vendar moremo odlaganje uporabiti samo pri opazovanjih z individualnimi obrazci, ne moremo pa tega storiti, če imamo kolektivne obrazce. Če so obrazci kolektivni, navadno prepišemo šifre podatkov na individualne obdelovalne listke, te pa potem po znanem postopku odlaganja samostojno obdelujemo.

Lep primer o povezavi posameznih etap statističnega proučevanja je obrazec za kmetijsko anketo v sliki 3.2, v katerem so podatki, za katere je določeno, da jih v osnovni obdelavi seštevamo, nanizani na robeh obrazca. Obdelava podatkov ankete je predvidevala sortiranje po grupah. Po sortiranju obrazce posamezne grupe nanizamo drugega nad drugim tako, da so podatki, ki jih gledе na načrt obdelave seštevamo, sami od sebe nanizani v stolpcih. Zato je seštevanje podatkov znatno olajšano.

4.23 *Črtkanje* uporabljamo pri manjših obdelavah. Če hočemo dobiti za dano populacijo prebivalstva sestav po spolu, delamo pri črtkanju enostavno takole: Sestavimo obdelovalno tabelo, ki ima za posamezno grupo razmeroma velika polja. V primeru razdelitve populacije po spolu ima obdelo-

valna tabela samo dve polji: eno za moške, drugo za ženske. Nato po vrsti pregledamo, v katero grupo spadajo posamezne enote; to naznačimo za vsako enoto posebej v ustreznem polju v obdelovalni tabeli s črtico. Če je prva popisnica za moškega, napravimo črtico v polju za moške. Če je druga popisnica tudi za moškega, napravimo zraven prve črtice drugo črtico. Če je tretja popisnica za žensko, napravimo črtico v polju za ženske. Ta postopek ponavljamo za vsako popisnico, dokler populacijo ne izčrpamo. Če preštejemo črtice v polju za moške, dobimo, koliko je število moških v populaciji. Analogno dobimo število žensk, če preštejemo črtice v polju za ženske. Skupno število vseh črtic pa je enako številu enot v populaciji. Da si olajšamo štetje črtic po končani obdelavi, včasih že med obdelavo sestavljamo skupine po pet črtic, tako da s peto črtico v vsaki skupini po pet črtic prve štiri prečrtamo: 𝍸 Vzemimo populacijo 25 oseb. Zanje je spol naznačen z M za moške in z Ž za ženske. Osnovni podatki o enotah so: M Ž M M Ž M M Ž Ž Ž M Ž M Ž M M Ž Ž Ž Ž M M Ž Ž Ž. Po metodi črtkanja dobimo takle rezultat obdelave:

Moški	𝍸 𝍸 /	11
Ženske	𝍸 𝍸 ////	14
		25 = N

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

Vzemimo, da je populacija sestavljena iz deset enot, od katerih so neke popisnice za delavce -D, druge pa za uslužbence -U. Imamo pa še podatke o plačah teh desetih enot (v tisočih). Osnovni podatki za posamezne enote so: D-19, U-13, D-21, D-11, U-15, U-17, D-23, U-13, U-17, U-14. Če predpisuje obdelava, da je treba izračunati skupni fond plač po grupah za delavce in uslužbence, je obdelava po zgornjem postopku takale:

		vsota	število
Delavci -D	19, 21, 11, 23	74	4
Uslužbenci -U	13, 15, 17, 16, 17, 14	92	6
		$\Sigma x = X =$ 166	10 = N

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

Črtkanje moremo uporabiti pri individualnih in kolektivnih obrazcih ali če so podatki navedeni za vse enote na enem listu papirja, kakor je to v naših ilustrativnih primerih. To je njegova prednost. Ima pa črtkanje svoje hibe, ki omejujejo uporabnost in uporabo. Črtkanje je neprikladno za obdelave, v katerih kombiniramo več znakov. Za take obdelave je iskanje ustreznega polja v obdelovalni tabeli zamudno. Razen tega je

za populacije velikega obsega črtkanje neuporabno, ker nimamo kontrole. Napake moremo naknadno odkriti le s ponovno obdelavo. Pri črtkanju si sicer večkrat pomagamo tako, da osnovno populacijo razstavimo v več obdelovalnih delovnih populacij. Vendar tudi ta rešitev večkrat ne pomaga. Zato črtkanje uporabljamo le za majhne populacije in ne preveč zamotane obdelave.

Strojna obdelava

4.25 Kakor smo že omenili, je osnovna obdelava statističnega gradiva obsežen in časovno dolgotrajen posel. Večletna obdelava popisov velikega obsega, kot so popisi prebivalstva, popisi industrije, kmetijstva itd., je v škodo vrednosti podatkov, ker so, ko jih objavimo, že zastareli. Pri velikih statističnih akcijah čas osnovne obdelave znatno skrčimo z mehanizacijo obdelave.

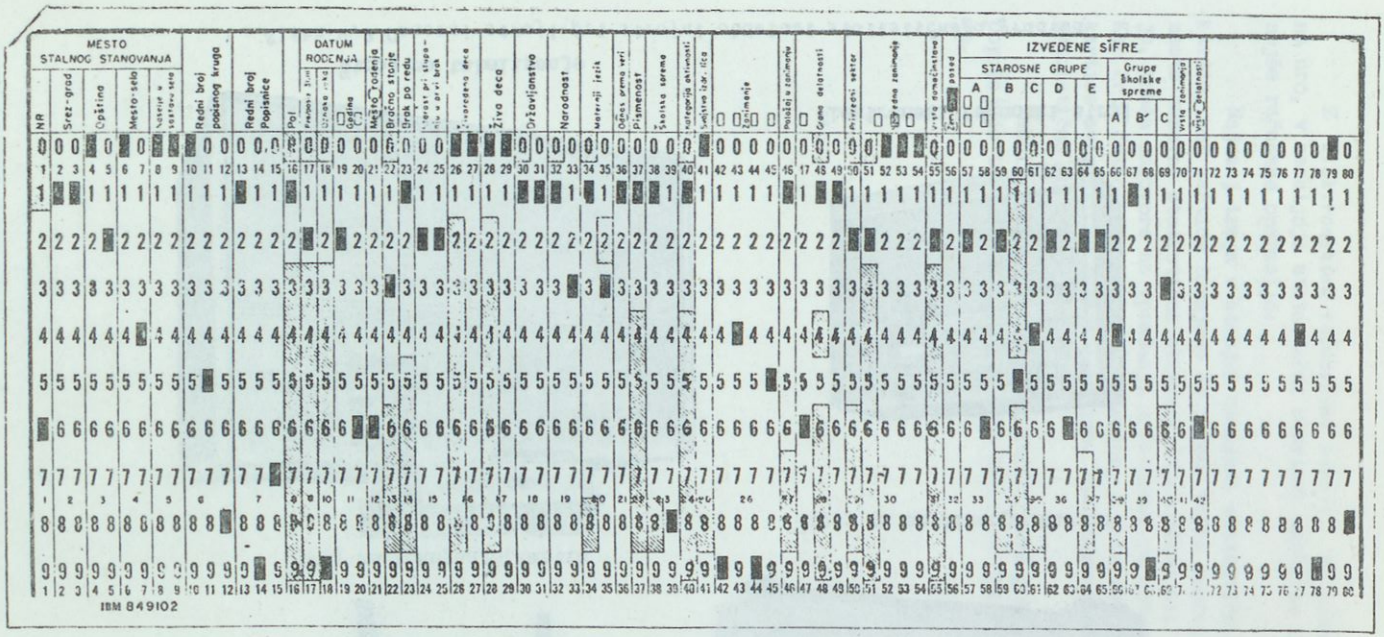
Osnova strojne obdelave je posebna kartica, na katero s sistemom luknjic prenesemo podatke iz popisnega obrazca. Da je možno z luknjicami prenesti podatke iz popisnega obrazca na kartico, mora biti osnovno gradivo šifrirano s številčnimi šiframi. Najuspešnejše so šifre po decimalni klasifikaciji. Standardna kartica ima 80 kolon-kodeksov. Vsaka kolona ima v navpični smeri deset mest od 0 do 9 oziroma z X in Y dvanajst mest. Za znake z enomestno šifro uporabimo en kodeks, za znake z večmestnimi šiframi pa večmestne kodekse.

V prvi stopnji strojne obdelave s posebnim *strojem za luknjanje kartic* prenesemo šifre iz osnovnih obrazcev na kartice. S strojem za luknjanje kartic za vsak znak z luknjicami pravokotne ali okrogle oblike vnesemo šifre na kartico. V sliki 4.2 imamo kartico za obdelavo popisa prebivalstva 31. III. 1953 v FLRJ. Prvi kodeks na kartici je za enomestno šifro za ljudsko republiko, drugi, dvomestni, velja za okraj, tretji, dvomestni kodeks za občino, itd.

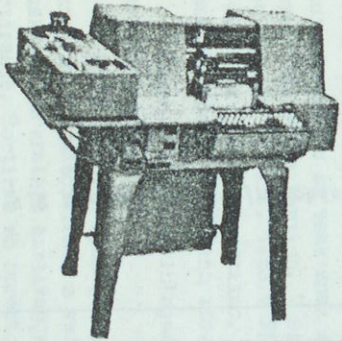
Razen strojev za luknjanje kartic s posebnimi stroji kontroliramo pravilnost luknjanja.

Ko je gradivo zluknjano, ga na *strojih za sortiranje* sortiramo. S sortirkami razsortiramo z veliko hitrostjo kartice po šifrah za določen kodeks. Z večkratnim sortiranjem po različnih kodeksih dobimo kombinirano sortiranje. Sortirke so opremljene s števci, ki sproti avtomatično štejejo število kartic - enot v posameznih grupah.

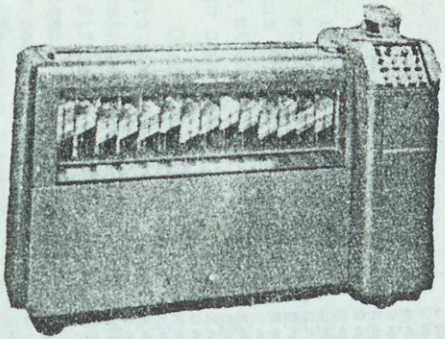
Na *tabelirki* avtomatično z veliko hitrostjo seštevamo po grupah več desetin različnih podatkov istočasno. Ta stroj rezultate tabeliranja avtomatično tiska na poseben papir: tega moremo pripraviti tako, da dobimo iz stroja neposredno stiskane končne tabele za osnovne podatke. Ker je možno s primerno kombinacijo luknjic doseči, da tabelirka ne piše samo številke, marveč tudi črke, iz tabelirke dobimo končno tabelo z izpisanim čelom tabele.



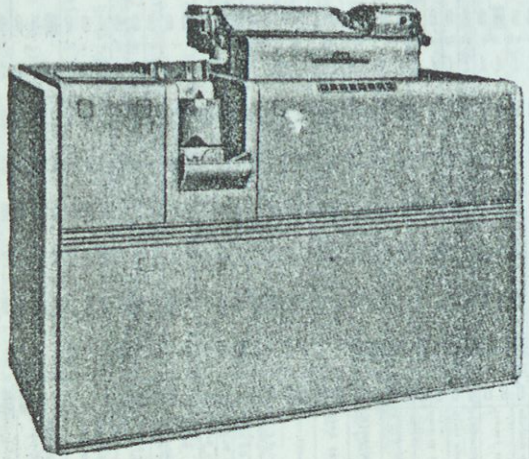
Slika 4.2 Kartica za strojno obdelavo popisa prebivalstva 31.III.1953 v FLRJ



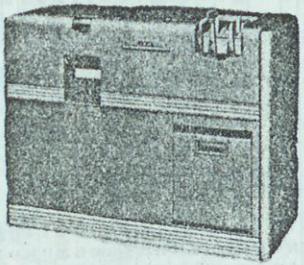
Stroj za luknjanje kartic



Stroj za sortiranje kartic



Stroj za tabeliranje



Elektronski računski stroj

Slika 4.3. Osnovni stroji pri strojni obdelavi statističnega gradiva

Z *elektronskimi računskimi stroji*, ki izvrše več tisoč operacij na uro, v nadaljnji statistični obdelavi množično izračunavamo najrazličnejše računske operacije.

Najdalj traja v strojni obdelavi statističnih podatkov luknjanje kartic, ker kartice luknjamo še ročno. Korak naprej so kartice, ki nadomeste osnovni statistični obrazec. Popisovalec ali tisti, ki podatke daje, vnaša odgovore na vprašanja s črticami na ustreznih mestih v kartici. Ker črta črtice s posebnim svinčnikom, v obdelavi stroj na drugem koncu kartice avtomatično luknja luknjice. Tako odpade časovno dolgotrajno poiavtomatično luknjanje kartic.

Tehnika strojne obdelave statističnega gradiva pa se še stalno izpopolnjuje tako glede skrčenja časa obdelave kakor glede možnosti obdelave.

Peto poglavje

PRIKAZOVANJE STATISTIČNIH PODATKOV

5.1 Statistične podatke redkokaj prikazujemo posamič, ker je bistvo statističnega proučevanja pojavov v primerjavi podatkov. Zato sorodne statistične podatke zarežujemo v *statistične vrste*, te pa prikazujemo v *tabelah* ali v *grafikonih*. Tabelarični in grafični prikaz sta specifična načina prikazovanja statističnih podatkov. Tabelarično in grafično prikazovanje statističnih podatkov se po svojih sredstvih razlikujeta med seboj. Imata pa isti namen: čim nazorneje in preglednejе prikazati proučevano populacijo, tako da je analiza pojava čim lažja.

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

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

Prednost *grafičnega prikazovanja* v primerjavi s tabelaričnim prikazom pa je predvsem v tem: a) v grafikonu moremo s pogledom odkriti zveze in odnose več podatkov hkrati; to v tabelaričnem prikazu ni možno; b) grafikon, s katerim hočemo popularizirati določeno dejstvo, je bolj privlačen in neposreden kakor tabelaričen prikaz; c) z analitičnimi grafikoni statističnih podatkov moremo včasih analizirati pojav brez dodatne analitične obdelave. Dostikrat pa da grafikon smernice za podrobno statistično obdelavo, kot je to n.pr. pri proučevanju odvisnosti med pojavi, proučevanju dinamike ekonomskih pojavov itd.

STATISTIČNE VRSTE

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

Glede na to, po kakšnem znaku se razlikujejo med seboj posamezni členi statistične vrste, imamo: krajevne, časovne in stvarne statistične vrste. Stvarne statistične vrste pa dalje delimo v atributivne in numerične vrste, glede na to, ali statistična vrsta prikazuje podatke razporejene po atributivnem ali po numeričnem znaku. Kakor vidimo iz same de-

finicije statistične vrste, se posamezni členi nanašajo na določeno vrednost osnovnega znaka.

5.3 V tabeli 5.1 je prikazana kot primer *krajevne ali geografske vrste* razdelitev narodnega dohodka v FLRJ v letu 1953 po ljudskih republikah.

Tabela 5.1 Narodni dohodek v FLRJ po ljudskih republikah v letu 1956 (v milijardah dinarjev). (Vir: SG 1958).

Republika	Narodni dohodek v 10^9 din
Jugoslavija	1 444
Srbija	519
Hrvatska	395
Slovenija	238
BiH	188
Makedonija	82
Črna gora	22

Drug primer krajevne vrste je povprečna cena mleka v glavnih mestih republik v FLRJ v letu 1953, v tabeli 5.2

Tabela 5.2 Povprečna cena mleka v glavnih mestih republik v FLRJ v letu 1956 (Vir: SG 1956)

Mesto	Cena v dinarjih
Ljubljana	31,0
Zagreb	38,0
Beograd	44,0
Sarajevo	43,5
Skoplje	50,0
Titograd	60,0

5.4 *Časovna vrsta* je n.pr. proizvodnja električne energije v FLRJ v razdobju 1950-1957 v tabeli 5.3, ki prikazuje dinamiko v proizvodnji električne energije v tem razdobju.

Tabela 5.3 Proizvodnja električne energije v FLRJ v razdobju 1950-1957. (Vir: SG 1958)

	1950	1951	1952	1953	1954	1955	1956	1957
Proizvodnja v milijonih KWh	2 408	2 550	2 700	2 982	3 440	4 340	5 048	3 252

Statistično vrsto proizvodnje električne energije v tabeli 5.3 imenujemo *intervalno časovno vrsto*, ker prikazuje intervalne podatke. *Momentna časovna vrsta* pa je število zaposlenih v družbenem sektorju na dan 30. septembra v razdobju 1952-1957 v FLRJ v tabeli 5.4.

Tabela 5.4 Število zaposlenih v družbenem sektorju dne 30. septembra v letu v razdobju 1952-1957 v FLRJ (Vir: SB 110).

	1952	1953	1954	1955	1956	1957
Število zaposlenih v tisočih	1 703	1 855	2 038	2 224	2 245	2 413

Statistična vrsta v tabeli 5.4 je momentna časovna vrsta, ker prikazuje število zaposlenih v določenem momentu v vsakem letu.

5.5 *Stvarno atributivna statistična vrsta* je n.pr. število zaposlenih v družbenem sektorju dne 30. septembra 1957 po gospodarskih dejavnostih, ki je prikazana v tabeli 5.5.

Tabela 5.5 Število zaposlenih v družbenem sektorju na dan 30. septembra 1957 v FLRJ po panogah dejavnosti (Vir: SB 110)

Panoga dejavnosti	Število zaposlenih v tisočih
Skupno	2 550
Industrija in rudarstvo	953
Kmetijstvo	213
Gozdarstvo	84
Gradbeništvo	243
Promet	192
Trgovina in gostinstvo	237
Obrt	137
Komunala	42
Kultura-prosveta	133
Socialno skrbstvo-zdravstvo	115
Organi državne oblasti in uprave	120
Pravosodje	12
Finance in kredit	30
Ostalo	19

5.6 Primer *zvezne numerične vrste* so povprečni denarni izdatki za eno gospodinjstvo za 317 gospodinjstev v Sloveniji iz ankete o kmetijskih gospodarstvih v letu 1956 po višini skupnih denarnih dohodkov v tabeli 5.6.

Tabela 5.6 Skupni povprečni denarni izdatki za eno gospodinjstvo za 317 gospodinjstev v anketi o kmetijskih gospodarstvih v Sloveniji v letu 1956, po višini skupnih denarnih dohodkov (Vir: SB 113).

Skupni denarni dohodki v tisočih din	Povprečni skupni denarni izdatki za gospodinjstvo v tisočih din
do 50	27,5
50 - 100	52,0
100 - 150	80,8
150 - 200	100,2
200 - 300	135,3
300 - 400	184,0
400 - 500	197,8
nad 500	285,0

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

Tabela 5.7 Število gospodarstev v kmetijski anketi v Sloveniji v letu 1956 po številu članov v gospodinjstvu. (Vir: SB 113).

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

Prikazane statistične vrste so po svojem značaju med seboj različne tudi po drugih lastnostih, ne samo po tem, da imajo za osnovo različne vrste znakov. V tabeli 5.1 imamo *sestav* skupnega narodnega dohodka v FLRJ po republikah. Istega značaja je tudi statistična vrsta v tabeli 5.5,

ki prikazuje sestav števila zaposlenih po panogah dejavnosti.

Prav poseben pomen imajo statistične vrste, kakršna je n.pr. število kmetijskih gospodarstev po številu članov v tabeli 5.7. Ta statistična vrsta prikazuje variabilnost števila članov za anketirana gospodarstva. Take vrste statističnih vrst imenujemo *frekvenčne distribucije*. Ker so za proučevanje populacij osnovne važnosti, jih bomo podrobneje proučevali v posebnem poglavju. Podobne narave je tudi statistična vrsta o številu zaposlenih po panogah dejavnosti v tabeli 5.5. Tudi ta statistična vrsta pokaže variabilnost znaka: panoga dejavnosti. Vendar so metode proučevanja statističnih vrst, ki kažejo variabilnost po določenem numeričnem znaku, mnogo bolj razvite kakor za nenumerične.

Tudi analiza časovnih vrst je specifična. S časovnimi vrstami proučujemo *dinamiko* socialno-ekonomskih pojavov. Ker je tudi analiza časovnih vrst specifična, tudi časovne vrste proučujemo v posebnem, za ekonomista važnem poglavju o analizi časovnih vrst. Dinamika proizvodnje električne energije v FLRJ je prikazana v tabeli 5.3, dinamika števila zaposlenih v družbenem sektorju pa v tabeli 5.4.

Statistična vrsta o povprečni ceni mleka v posameznih glavnih mestih republik v tabeli 5.2 in statistična vrsta o povprečnih skupnih denarnih izdatkih za gospodinjstvo po grupah skupnih denarnih dohodkov v tabeli 5.3, prikazujeta, kako je cena mleka odvisna od kraja, povprečni mesečni izdatki na gospodinjstvo pa od skupnih denarnih dohodkov v kmetijskem gospodarstvu. Ti vrsti prikazujeta važen in splošen problem socialno-ekonomskih pojavov - *odvisnost* med pojavi. Tudi o odvisnosti med socialno-ekonomskimi pojavi je govora v posebnem poglavju.

Iz navedenega vidimo, da imamo najrazličnejše statistične vrste. Vsaka od njih po svoje prikaže določeno značilnost ali lastnost proučevanih pojavov. Razen tega vidimo, da so v dosti primerih statistične vrste le osnova za nadaljnjo posebno obdelavo in analizo.

STATISTIČNE TABELE

5.8 Iz primerov v prejšnjem odstavku vidimo, da statistične vrste prikazujemo v tabelah. Statistične podatke najpogosteje prikazujemo v statističnih tabelah, ker je tabela najpreglednejši in najracionalnejši prikaz podatkov.

V tabeli s sistemom *vrst* in *stolpcev* razdelimo kos papirja v *polja*, v katera vpisujemo po določenem sistemu statistične podatke. Dogovorno velja, da imajo vsa polja v isti vrsti isto vsebinsko oznako, ki je naznačena za vsako vrsto posebej v posebnem tekstualnem delu - *čelu* tabele. Enako imajo vsa polja v istem stolpcu enako vsebinsko oznako, ki je ustrezno naznačena v tekstualnem delu - *glavi* tabele. Pomen podatka v določenem polju dobimo, če križamo pomen ustrezne vrste oziroma stolpca, v katerem je polje oziroma podatek. Vsi ti pojmi so nakazani v tabeli 5.8.

Polje A v tabeli 5.8 je polje za podatek o vzgrževanih ženskah v starosti izpolnjenih treh let.

Tabela 5.8 Obdelovalna tabela 1 za obdelavo vzorca predhodnih rezultatov za popis prebivalstva 31. III. 1959 v FLRJ. (Glej sliko 4.1).

Delovnost	Delovni		Osebe z osebnimi dohodki		Vzgrževani		glava
Spol / Starost	moški	ženske	moški	ženske	moški	ženske	
0							telo
1							
2	vrsta				polje		
3						A	
4			polje				
5							
6							
čelo							

5.9 Tabela je *enostavna*, če v nji prikažemo eno samo statistično vrsto. Enostavne tabele so torej vse tabele od 5.1 do 5.7 v prejšnjem odstavku, ker je v njih prikazana ena sama statistična vrsta. Vse te tabele imajo bodisi en sam stolpec (tabele 5.1, 5.2, 5.5, 5.6 in 5.7), ali eno samo vrsto (tabeli 5.3, 5.4).

5.10 Dostikrat imamo dve ali več različnih statističnih vrst, ki imajo za osnovo isti znak in so med seboj v vsebinski zvezi. Take statistične vrste moremo prikazati v eni sami statistični tabeli. Tako tabelo imenujemo *sestavljeno tabelo*, ker je sestavljena iz več enostavnih tabel. Sestavljena tabela je tabela 5.9, ki prikazuje število trgovinskih podjetij, število zaposlenega osebja v teh podjetjih in promet v trgovini na debelo v FLRJ v letu 1957 po republikah.

Tabela 5.9 Število podjetij in število zaposlenih konec leta 1957 in promet v letu 1957 v trgovini na debelo v FLRJ po republikah. (Vir: SG 58).

Republika	Število podjetij	Število zaposlenih v tisočih	Promet v milijardah dinarjev
FLRJ	1 607	61,8	834,0
Srbija	529	24,9	329,0
Hrvatska	372	15,8	211,7
Slovenija	272	8,7	145,8
BiH	269	7,7	93,2
Makedonija	124	3,6	42,2
Črna gora	41	1,0	12,1

V tabeli 5.9 prikazani podatki so zaruženi iz treh enostavnih tabel, od katerih prva prikazuje število podjetij, druga število zaposlenih, tretja pa promet v trgovini na debelo po republikah.

5.11 Če obdelujemo populacijo samo po enem znaku, dobimo za rezultat enostavne statistične vrste, ki jih prikazujemo v enostavnih statističnih tabelah. Dostikrat pa populacijo razdeljujemo po dveh, treh ali celo več znakih hkrati. Ta obdelava da podroben vpogled v odvisnosti pri množičnih pojavih in je posebno pomembna, če proučujemo zvezo med faktorialnimi in rezultativnimi znaki. Rezultate take večkratne istočasne obdelave prikazujemo v *kombinacijskih tabelah.* V kombinacijski tabeli križamo po dva ali več znakov. Kombinacijska tabela je n.pr. razdelitev kmetijskih gospodarstev v anketi o kmetijskih gospodarstvih v letu 1953 po velikostnih skupinah in dohodkih od kmetijskega gospodarstva za 2233 anketiranih gospodarstev v FLRJ v tabeli 5.10.

Tabela 5.10 Število privatnih kmetijskih gospodarstev v kmetijski anketi v FLRJ po velikosti gospodarstev in denarnih dohodkih od kmetijstva v letu 1956 (Vir: SB 113).

Skupna površina v ha	Denarni dohodki od kmetijstva v tisočih dinarjev							
	-100	100-200	200-300	300-400	400-500	500-800	800-	skupno
do 4 ha	942	208	43	19	7	5	2	1 224
nad 4 - 8 ha	337	172	91	35	14	2	4	655
nad 8 - 12 ha	72	72	40	30	9	4	7	234
nad 12 - 16 ha	21	20	14	7	2	2	1	67
nad 16 - 20 ha	3	5	7	3	1	5	2	29
nad 20 ha	2	7	8	2	2	2	4	27
Skupno	1 380	479	203	93	35	20	20	2 233

V kombinacijski tabeli 5.10 imamo razen običajnih stolpcev in vrst še seštevke stolpcev in vrst. Tako dobimo *zbirno vrsto* in *zbirni stolpec*, katerega stavljamo ali na konec tabele, kakor je to v našem primeru, ali pa na začetek tabele.

5.12 Za primer kombinacijske tabele s kombinacijo več znakov hkrati vzemimo tabelo 4.9 iz Statističnega godišnjaka FNRJ 1957. Tabela prikazuje prebivalstvo v Jugoslaviji, staro nad 10 let po delovnosti, pismenosti, starosti in spolu po republikah. V tabeli je torej kombiniranih pet znakov hkrati.

Tabela 5.11 Prebivalstvo v FLRJ staro nad 10 let po delovnosti, pismenosti, starosti in spolu po republikah (v tisočih; Vir: SG 1957).

Starost - spol		FLRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
SKUPNO PREBIVALSTVO								
Skupno	skupno	13 381	5 586	3 204	1 191	2 123	959	319
	ženske	6 992	2 887	1 715	637	1 107	477	169
čita in piše								
Skupno	skupno	9 977	4 027	2 682	1 159	1 270	616	223
	ženske	4 485	1 719	1 330	620	475	246	94
10 - 14 let	skupno	1 420	545	327	128	257	118	45
	ženske	667	258	161	63	108	56	21
15 - 19	skupno	1 533	604	361	130	284	111	42
	ženske	700	273	174	64	119	50	20
20 - 34	skupno	3 536	1 502	900	343	478	236	76
	ženske	1 627	669	460	184	182	97	34
35 - 49	skupno	1 878	752	572	263	162	97	33
	ženske	795	287	279	141	45	30	12
50 in več	skupno	1 609	624	521	295	89	55	26
nepoznano	ženske	696	232	256	167	21	13	6
nepismeno								
Skupno	skupno	3 404	1 559	522	32	853	343	96
	ženske	2 506	1 168	384	18	632	230	75
10 - 14 let	skupno	185	61	18	1	84	18	3
	ženske	119	38	10	1	59	10	2
15 - 19	skupno	224	92	21	1	81	25	4
	ženske	166	69	14	1	32	17	3
20 - 34	skupno	688	312	83	4	217	74	16
	ženske	577	263	54	2	186	58	14
35 - 49	skupno	940	441	147	7	228	93	24
	ženske	680	324	109	3	163	61	19
50 in več	skupno	1 368	654	270	18	243	133	49
nepoznano	ženske	964	475	197	11	161	85	36

Starost - spol		FIRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
DELOVNO PREBIVALSTVO								
Skupno	skupno	8 383	3 570	2 027	803	1 266	549	170
	ženske	2 975	1 235	747	330	423	156	55
čita in piše								
Skupno	skupno	9 718	2 783	1 765	782	865	391	131
	ženske	2 087	820	603	320	217	94	34
10 - 14 let	skupno	281	117	64	12	60	19	9
	ženske	127	54	30	6	25	8	5
15 - 19	skupno	1 142	440	272	103	225	77	24
	ženske	461	177	115	47	83	30	10
20 - 34	skupno	2 713	1 176	671	263	376	178	51
	ženske	871	375	245	110	85	44	12
35 - 49	skupno	1 411	586	407	187	130	75	25
	ženske	332	123	115	65	15	9	4
50 in več	skupno	1 168	464	350	217	74	42	21
nepoznano	ženske	295	90	98	93	9	3	2
nepismeno								
Skupno	skupno	1 668	787	262	21	401	156	39
	ženske	888	445	134	10	207	62	21
10 - 14 let	skupno	54	14	4	-	32	3	1
	ženske	37	9	3	-	24	1	-
15 - 19	skupno	169	65	17	1	66	17	2
	ženske	115	44	12	-	48	10	2
20 - 34	skupno	370	183	40	2	102	37	5
	ženske	266	137	29	1	72	22	4
35 - 49	skupno	485	246	78	5	98	48	9
	ženske	230	131	42	2	34	17	5
50 in več	skupno	589	279	122	13	103	52	21
nepoznano	ženske	240	124	58	7	29	12	10

Vrste tabel po vlogi v statističnem proučevanju

5.13 Glede na vlogo, ki jo imajo v statističnem proučevanju, delimo ta-

bele na: *obdelovalne, publikacijske* in *analitične.*

Osnovni namen obdelovalnih tabel je sistematično zapisovanje rezultatov osnovne obdelave statističnega gradiva, medtem ko je osnovni namen publikacijskih tabel sistematično prikazovanje, analitičnih pa analiza statističnih podatkov.

5.14 Oblika *obdelovalnih tabel* je prilagojena obdelavi podatkov. Zato se obdelovalne tabele pri ročni obdelavi razlikujejo od obdelovalnih tabel pri strojni obdelavi. Oblika in velikost obdelovalnih tabel je drugotnega pomena in pri sestavljanju dajemo prednost momentom obdelave. Zato obdelovalne tabele pri nekaterih statističnih obdelavah zaružujejo kombinacije razmeroma velikega števila znakov. Tako so pri popisu prebivalstva 31. marca 1953 obdelovalne tabele vsebovale kombinacije tudi po šest in več znakov hkrati in so bile sestavljene iz več delnih tabel. V obdelovalni tabeli 3, so bili n.pr. po okrajih kombinirani tile znaki: spol (2), aktivnost (3), poklic (74), položaj v poklicu (7) in sektor lastništva(5). Podatki v oklepajih pri vsakem znaku pomenijo število vrednosti znaka. Imenovana tabela je torej imela za vsak okraj 2×3×74×7×5 = 15540 polj brez delnih in končnih zbirnih stolpcev in vrst. Iz tega primera jasno vidimo, da ta tabela ni primerna niti za publiciranje, še manj pa za neposredno analizo podatkov.

5.15 V *publikacijskih tabelah* je še poudarek na tem, da damo strnjeno čim več statističnih podatkov. Vendar je pri njih določilen vsebinski moment. V publikacijskih tabelah so podatki prikazani tako, da so podatki, ki so med seboj v zvezi, čim bolj primerljivi. Publikacijske tabele so v glavnem vse tabele, ki so objavljene v statističnih publikacijah statističnih organov, kot so: letopisi, bilteni, rezultati popisov itd. Čeprav v publikacijskih tabelah že močno čutimo, da so sestavljene tako, da moremo čim laže prikazane podatke analizirati, je njihov osnovni namen še vedno ta, da dajo osnovne - običajno absolutne podatke o populaciji, ki jo prikazujejo, pojava pa ne analizirajo. Zato osnovne statistične publikacije redko vsebujejo več kakor absolutne podatke o statistični populaciji. Tipična publikacijska tabela je tabela 5.11.

5.16 Namen *analitične tabele* pa je prikazati statistične podatke tako, da je iz njih možno neposredno analizirati določeno značilnost pojava. Zato mora biti analitična tabela sestavljena tako, da je primerljivost podatkov čim boljša, osnovna značilnost pojava pa čim bolje vidna. Ker je razmeroma težko, včasih pa celo nemogoče, istočasno primerjati veliko podatkov hkrati, so analitične tabele običajno manjšega obsega in vsebujejo samo podatke, ki so v neposredni zvezi s problemom, ki ga hočemo z analitično tabelo prikazati in analizirati. Zato v analitičnih tabelah običajno kombiniramo po dva, največ tri znake hkrati. Kombinacija več znakov analizo podatkov znatno zamota. V analitičnih tabelah ne prikazujemo samo absolutnih števil, kakor v obdelovalnih tabelah in v veliki večini publikacijskih tabel, temveč vključujemo vanje vse vrste statističnih parametrov: relativna števila, srednje vrednosti, mere variacije itd. Analitič-

ne tabele so navadno sestavni del pismene analize proučevanega pojava.

Tehnična načela za sestavljanje tabel

5.17 Vsaka tabela mora biti sestavljena po nekih enotnih tehničnih načelih, če hočemo, da je razumljiva in pregledna.

Tabela mora imeti samostojen *naslov*, iz katerega je v kratkem razvidno, kaj prikazuje, področje, ki ga obsega, čas, na katerega se podatki v tabeli nanašajo, in kako so podatki v tabeli razdeljeni. Tako je n.pr. iz naslova tabele 5.1 takoj vidno: kaj prikazuje: narodni dohodek; področje: FLRJ; kdaj: v letu 1953; razdeljen: po republikah.

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

V nadaljevanju naslova navedemo *vir*, iz katerega je bila tabela črpana. To je potrebno, ker iz vira sklepamo na kakovost podatkov, po potrebi pa moremo iz osnovnega vira dobiti podrobne - dodatne podatke.

V tabeli je treba naznačiti *enoto mere*, ki jo stavljamo nad tabelo ali v podnaslov, če je enota mere enotna za vse podatke v tabeli (n. pr. tabela 5.III), ali v glavo oziroma čelo tabele, če je enota mere za posamezne podatke različna (n. pr. tabela 5.9).

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

V analitičnih tabelah se izogibamo podatkov z velikim številom mest in podatke ustrezno *zaokrožujemo*. V glavnem so pri večmestnih podatkih pomembna tri mesta. Vendar zaokrožujemo v teh primerih na naravna ali konvencionalna merila n.pr. vrednosti na tisoče, milijone ali milijarde dinarjev, ne pa na desettisoče, deset milijone itd.).

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

5.18 Grupe za posamezne *znake razvrstimo* po različnih načelih. Časovne grupe in numerične znake razporedimo po *naravnem vrstnem redu* po velikosti v čelu tabele od zgoraj navzdol, v glavi tabele pa od leve na desno (glej tabele 5.3, 5.4, 5.10 in 5.11).

Za druge vrste znakov vrstni red ni vnaprej določen. Zanje uporabljamo različne načine. Če je število grup znaka veliko in ni nobenega merila, po katerem bi združevali grupe nižjega reda v grupe višjega reda, uredimo vrednosti po *abecednem redu*. V publikaciji osnovnih podatkov o prebivalstvu - knjiga XIV, v kateri so publicirani podatki za naselja po upravni razdelitvi v letu 1953, so v republikah okraji, v okrajih občine, v občinah pa naselja, razvrščena po abecedi. To veliko pripomore k hitremu iskanju potrebnih krajev.

Včasih je vrstni red *ustaljen* po kakem drugem znaku. Tako imamo za vrstni red republik ustaljen vrstni red: Srbija, Hrvatska, Slovenija, BIH, Makedonija in Črna gora; ta je pogojen z velikostjo posameznih republik na splošno. (Glej tabeli 5.9 in 5.11). Enako je tudi ustaljen vrstni red za panoge dejavnosti v tabeli 5.5.

Če prikazujemo eno samo statistično vrsto, je včasih ugodno, da navedemo grupe po *velikosti podatkov* v statistični vrsti. (Glej tabelo 5.2.) Če imamo sestavljeno tabelo, pa včasih uredimo člene po velikosti podatka, ki je v tabeli vodilen.

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

Tak znak je n.pr. šolska izobrazba, za katero razvrstimo grupe po stopnji od grupe: brez šolske izobrazbe do končne, najvišje stopnje: dokončana visoka šola. Enako moremo posamezne stopnje kvalifikacije urediti po vrstnem redu: visoko kvalificiran, kvalificiran, priučen in nekvalificiran.

5.19 Polja v tabeli so razmeroma majhna. Zato *pojasnila* k posameznim podatkom ali poljem ne moremo dati v sami tabeli. Da skrajšamo in poenostavimo pojasnila k posameznim okencem, uporabljamo *dogovorjene kratice* oziroma oznake, ki enostavno opisujejo določene značilnosti v tabeli. V uradnih statističnih publikacijah statističnih uradov v FLRJ uporabljamo tele dogovorjene oznake. V ustrezno polje postavljamo:

\- če pojava ni,
... če ne razpolagamo s podatkom,
0 če je podatek manjši kakor je 0,5 od enote mere,
0,0 če je podatek manjši kakor je 0,05 od enote mere.

Nepopoln, nezanesljivo kontroliran podatek damo v okrogel oklepaj (37), da je podatek popravljen pa označimo z zvezdico 37*.

Ø pomeni povprečje, z majhnimi številkami 1,2,3 ob podatku pa naznačimo ustrezno opombo k podatku pod črto.

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

GRAFIČNO PRIKAZOVANJE

5.21 V uvodu v prikazovanje statističnih podatkov smo dejali, da statistične podatke prikazujemo razen s tabelami tudi z *grafikoni*. Prikazali smo tudi prednosti in pomanjkljivosti enega in drugega načina.

Grafično prikazani podatki dajo enostavno in nazorno sliko statističnih vrst. Čeprav more biti tabela teoretično poljubno velika, moremo v nji istočasno primerjati le zelo omejeno število podatkov. Število istočasno primerjanih podatkov je v grafičnem prikazu znatno večje. Grafikon ne prikaže natančnih podatkov. Vendar to ni bistveno, ker je njegov namen, da da grobo sliko in vtis o pojavu. Zato običajno grafikon dopolnjuje tabela, če niso podatki številčno vpisani že v sam grafikon.

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

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

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

Elementi grafičnega prikazovanja

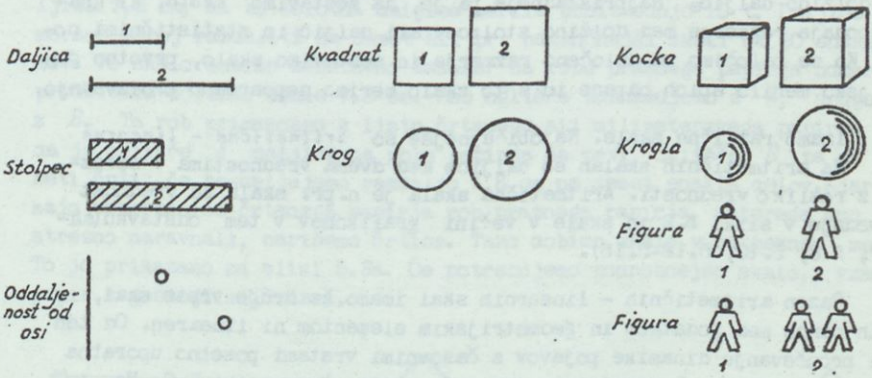
5.22 Vsaka statistična vrsta je sestavljena iz niza statističnih številčnih podatkov, od katerih se vsak nanaša na posamezno vrednost ali grupo vrednosti za osnovni znak v statistični vrsti. Ker je vsak statistični podatek dan številčno, moremo uporabiti za njegov prikaz najrazličnejše geometrijske elemente, ki se dajo meriti. Ustrezna povezava teh elementov večinoma da nazorno sliko pojava. Za prikazovanje statističnih podatkov med drugimi uporabljamo tele geometrijske elemente:

a) *daljico*, b) pravokotnike s stalno širino, ki jih imenujemo *stolpce*, c) *oddaljenost točke* od dane osi, d) *kvadrate*, e) *kroge*,

f) *kocke*, g) *krogle* itd.

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

Če primerjamo daljico in stolpec s kvadratom in krogom ali kocko in kroglo, vidimo, da je najlažja primerjava različno velikih količin pri daljici in stolpcih: elementi pa so tem slabše primerljivi, čim večja je razsežnost geometrijskega elementa. To je lepo razvidno iz slike 5.1, v kateri so narisani dvakrat večji posamezni zgoraj našteti elementi.



Slika 5.1 Elementi grafičnega prikazovanja

Razen zgornjih elementov, od katerih je vsak določen samo z enim podatkom, uporabljamo za prikazovanje tudi druge elemente, ki imajo več različnih razsežnosti. Tako *pravokotnik*, ki ima svojo višino in dolžino, uporabljamo za prikazovanje dveh statističnih podatkov hkrati, zaradi zveze med širino, dolžino in ploščino pa celo s tremi. Enako je *krogov izsek* določen z dvema podatkoma: radijem in kotom, ki ga izsek oklepa itd. Te elemente uporabljamo v kompleksnejših primerih grafičnega prikazovanja.

V grafikonih, ki so namenjeni popularizaciji podatkov, dostikrat uporabljamo za element grafičnega prikaza *idealizirane figure*, ki so vezane z vsebino pojava. Tako podatke o prebivalstvu prikažemo z idealizirano figuro moškega ali ženske (glej sliko 5.1), železniški promet z idealizirano lokomotivo, morski promet z idealiziranim prikazom parnika itd. Tudi

pri figurah uporabljamo dve metodi: Po prvi je volumen prikazane figure sorazmeren s podatkom, ki ga prikazujemo, po drugem pa velikost podatka ponazorimo z ustreznim številom enako velikih figur.

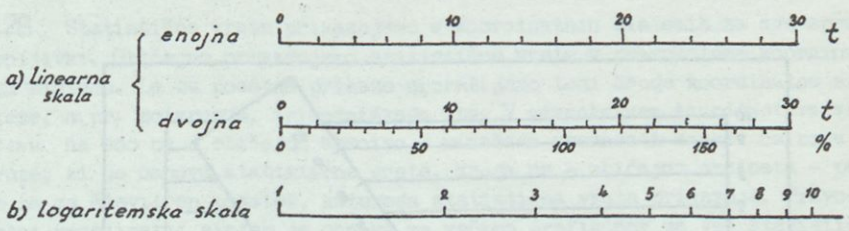
Skale - lestvice

5.23 Statistični podatki imajo različne enote mere. Tako merimo vrednost proizvodnje v kosih, metrih, kg ipd., površine v ha, narodni dohodek v milijonih dinarjev, število prebivalcev v tisočih prebivalcev itd. Enote mere geometrijskih elementov, kot je daljica, stolpec ali oddaljenost točke od osi, merimo v dolžinskih merah mm, cm itd. ploščine kvadratov in krogov v ploščinskih merah mm^2, cm^2 itd., volumen kocke in krogle pa v prostorninskih merah: mm^3, cm^3 itd.

Če hočemo geometrijske elemente, kot so daljica, kvadrat , kocka itd. uporabiti za prikazovanje statističnih podatkov, moramo določati razmerje med enoto mere za statistični podatek in enoto mere za geometrijski element, s katerim prikazujemo podatek. Če vzamemo, da 1 cm pomeni 10 t proizvodnje, je s tem razmerjem dana zveza katere koli vrednosti proizvodnje z dolžino daljice Najprikladneje pa je, da sestavimo *skalo*, ki grafično podaja razmerje med dolžino stolpcev ali daljic in statističnimi podatki. Ko se odločimo za določeno razmerje in sestavimo skalo, prvotno geometrijsko merilo sploh odpade in s to skalo merimo neposredno proizvodnjo.

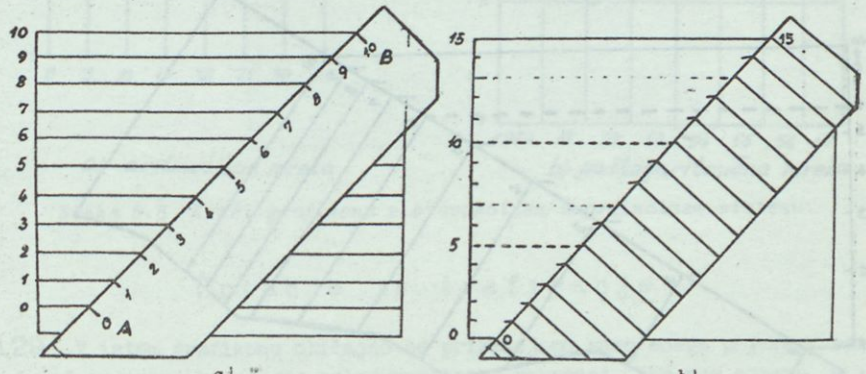
5.24 Imamo različne skale. Najobičajnejše so *aritmetične - linearne skale*. Na aritmetičnih skalah so daljice med dvema vrednostima v sorazmerju z razliko vrednosti. Aritmetična skala je n.pr. skala za vrednost proizvodnje v sliki 5.2 in skale v večini grafikonov v tem odstavku(slike 5.7; 5.8; 5.10; 5.12-5.15).

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



Slika 5.2 Skale

5.26 Risanje skal v ustreznem merilu. Pri risanju grafikonov je vsakdanji problem, kako narisati skalo v merilu, ki je v skladu z velikostjo grafikona. Pri tem uporabljamo enostavno metodo, ki pomaga risati skale v poljubnem merilu. Vzemimo, da moramo narisati tako linearno skalo na daljici 7,8 cm, da bo celotna daljica merila proizvodnjo 10 t. Daljico 7,8cm moramo torej razdeliti na deset ali pri natančnejši skali na 20 enakih delov. To skalo dobimo enostavno takole: na robu praznega papirja odmerimo predpisano dolžino skale 7.8 cm. Eno oglišče zaznamujemo z *A*, drugo pa z *B*. Ta rob primaknemo k listu črtanega ali milimeterskega papirja tako, da je oglišče *A* daljice na robu papirja na prvi, oglišče *B* pa na deseti črti, če razdeljujemo razdaljo 7,8 cm na deset enakih delov. Kjer sekajo vzporednice črtanega papirja rob praznega papirja, katerega smo ustrezno naravnali, narišemo črtice. Tako dobimo skalo v ustreznem merilu. To je prikazano na sliki 5.3a. Če potrebujemo podrobnejšo skalo, vzamemo papir z gostejšimi črtami.

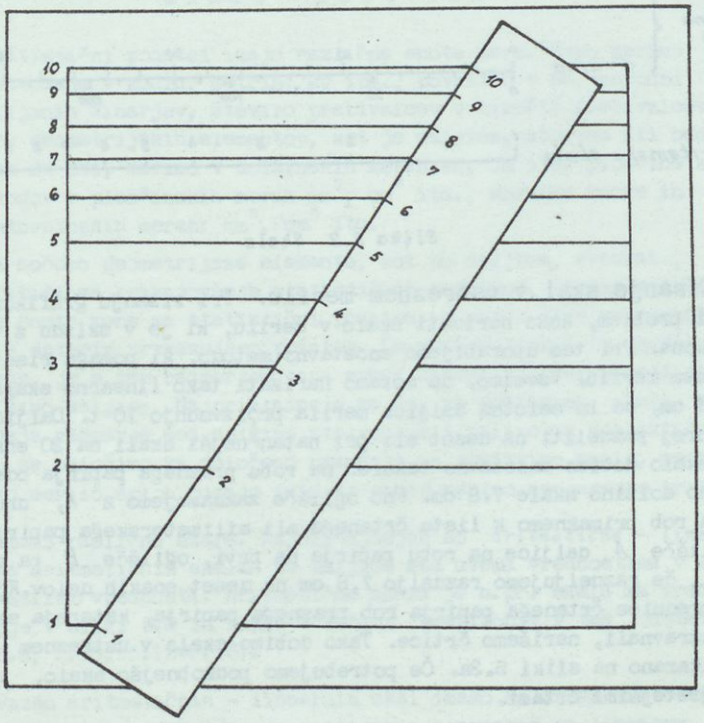


Slika 5.3 Risanje linearnih skal v poljubnem merilu

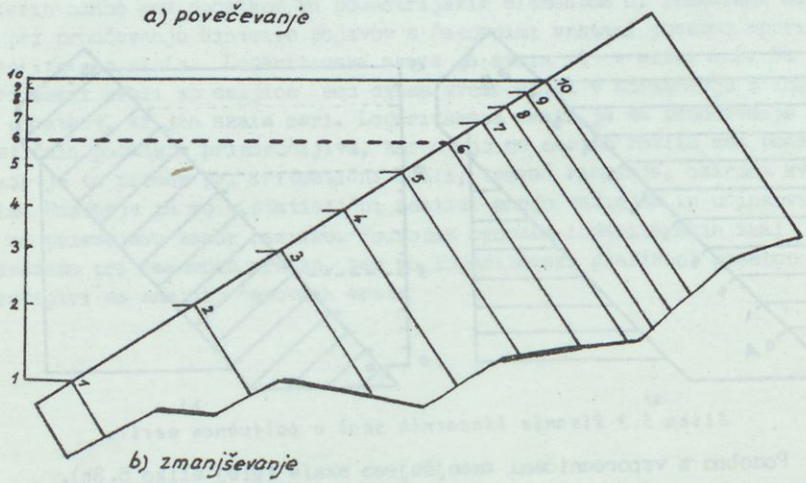
Podobno z vzporednicami zmanjšujemo skale (glej sliko 5.3b).

5.27 Zgornji postopek pa ne velja samo za linearne skale, temveč ga upo-

rabljamo tudi za druge vrste skal. V sliki 5.4 imamo nakazano, kako poljubno *povečujemo ali zmanjšujemo logaritemske skale.*



a) povečevanje

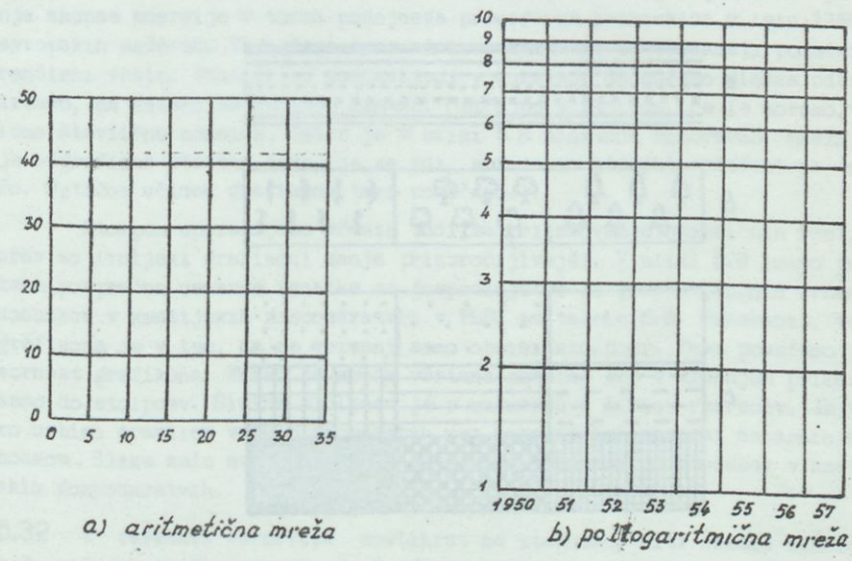


b) zmanjševanje

Slika 5.4 Povečevanje in zmanjševanje logaritemskih skal

Mreže

5.28 Statistične vrste prikazujemo v koordinatnih sistemih za dve spremenljivki. Običajno prikazujemo statistične vrste v pravokotnem koordinatnem sistemu. Le za posebne prikaze uporabljamo tudi druge koordinatne sisteme, n.pr. polarnega, trikotniškega ita. V *pravokotnem koordinatnem sistemu* na eno os - običajno absciso - nanašamo vrednosti znaka oziroma grupe, ki je osnova statistične vrste, druga os - običajno ordinata - pa je os za številčen podatek, katerega statistična vrsta prikazuje. Pravokotni koordinatni sistem je osnova za večino grafikonov za vse statistične vrste. Vendar je najbolj upravičen za prikazovanje časovnih oziroma numeričnih statističnih vrst. Običajno v pravokotnem koordinatnem sistemu vrišemo osnovne linije za nekaj najvažnejših vrednosti. Tako dobimo mrežo grafikona. Če v grafikon vrišemo mrežo, ga moremo veliko laže brati. V sliki 5.5 sta vrisani navadna aritmetična in poliogaritmična mreža grafikona v pravokotnem koordinatnem sistemu.



Slika 5.5 Mreži grafikona v pravokotnem koordinatnem sistemu

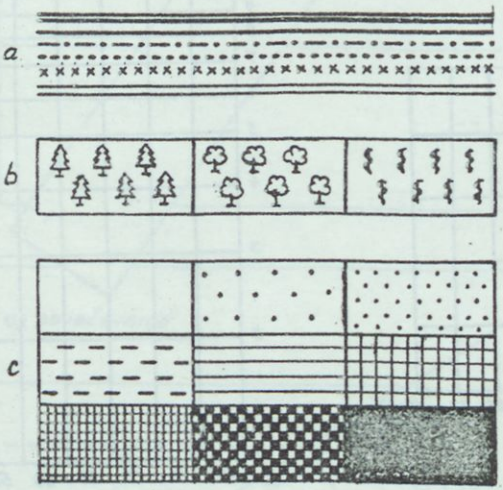
Črtanje in šrafiranje

5.29 V istem grafikonu običajno ne prikazujemo samo enega pojava, marveč več pojavov, ki so med seboj vsebinsko povezani. Ker vse pojave v istem grafikonu prikazujemo z linijami, stolpci, površinami ita., bi bil grafikon nepregleden, če bi uporabljali enake linije za različne pojave, enako zaznamovali površine, ki se nanašajo na različne pojave ita. Preglednost povečamo tako, da linije za različne pojave *različno črtamo*,

površine za različne pojave pa *različno šrafiramo*. V *legendi* - pojasnilu pa označimo, kaj posamezni znaki, črte, šrafure itd. pomenijo.

V sliki 5.3 a je načrtano nekaj vzorcev *črtanj* za linijske grafikone. Če so tehnične možnosti, je priključno, da namesto različnega črtanja linij uporabljamo različne barve. Barve moremo izbirati skladno s pojavom, ki ga prikazujemo. Tako v kmetijski statistiki rišemo linije, ki se nanašajo na travnike, zeleno, linije, ki se nanašajo na njive, rjavo, na gozd temno-zeleno itd.

Pri šrafurah je izbira še bolj pisana. Tu imamo možnost spreminjati barve in vzorce v šrafuri. Za šrafure v glavnem veljajo trije načini: a) šrafure brez zveze s pojavom; b) šrafure v skladu z intenziteto pojava, ki ga prikazuje; c) šrafure asociativno povezane s pojavom. V sliki 5.3 b imamo šrafure sestavljene tako, da podajajo predmet, ki ga prikazuje površina. Prva ponazarja gozd, druga sadovnjak, tretja pa vinograd. V sliki 5.3 c pa so šrafure razporejene po vrstnem redu od bele mimo različnih nians do črne. Take šrafure koristno uporabljamo za prikazovanje različne intenzitete pojava in pomeni temnejša šrafura večjo intenziteto.



Slika 5.6 Črte in šrafure

Vrste grafikonov

5.30 Običajni grafikoni so: a) *stolpci*, b) *linijski grafikoni*, c) *grafikoni s figurami*, d) *kartogrami*. Vsak izmed teh načinov ima svoje prednosti in pomanjkljivosti, ker ima vsak značilnosti, ki jih drug način nima, in obratno. Stolpce moremo uporabljati za prikazovanje vseh statističnih vrst, vendar niso v vseh primerih najprikladnejši. Linijski diagrami ima-

jo niz prednosti, ki jih drugi načini grafičnega prikazovanja nimajo. Zato so analitični grafikoni običajno linijski. Figure uporabljamo za popularizacijo pojavov, ki jih z njimi prikazujejo, nimajo pa posebnega študijskega značaja. Regionalno razmestitev pojava pa lepo prikažemo s kartogrami. Pri teh navadno mrežo v koordinatnem sistemu zamenja geografska karta.

Stolpci

5.31 Z *enostavnimi stolpci* prikazujemo grafično vse statistične vrste. To je prednost stolpcev. Te lastnosti linijski grafikoni n. pr. nimajo, čeprav so zaradi drugih lastnosti bolj čisłani. Zato s stolpci prikazujemo predvsem tiste statistične vrste, ki jih ne moremo prikazovati z linijskimi grafikoni. To so: geografske vrste in stvarno atributivne vrste.

Stolpce običajno v grafikonu razporedimo po velikosti členov, da podatke laže primerjamo med seboj. V sliki 5.8 je s stolpci prikazana potrošnja skupne energije v tonah pogojnega premoga na prebivalca v letu 1955 v evropskih državah. V grafikonu so stolpci nanizani po velikosti podatkov v ranžirni vrsti. Stolpci so postavljeni vodoravno, da moremo zlahka odbrati državo, na katero se stolpec nanaša. Običajno v stolpce, če le moremo, vnesemo številčne podatke. Skala je v sliki 5.8 nanešena vodoravno, mreža pa je v grafikon včrtana tako, da se zdi, kot da so stolpci položeni na mrežo. Optični učinek grafikona tako povečamo.

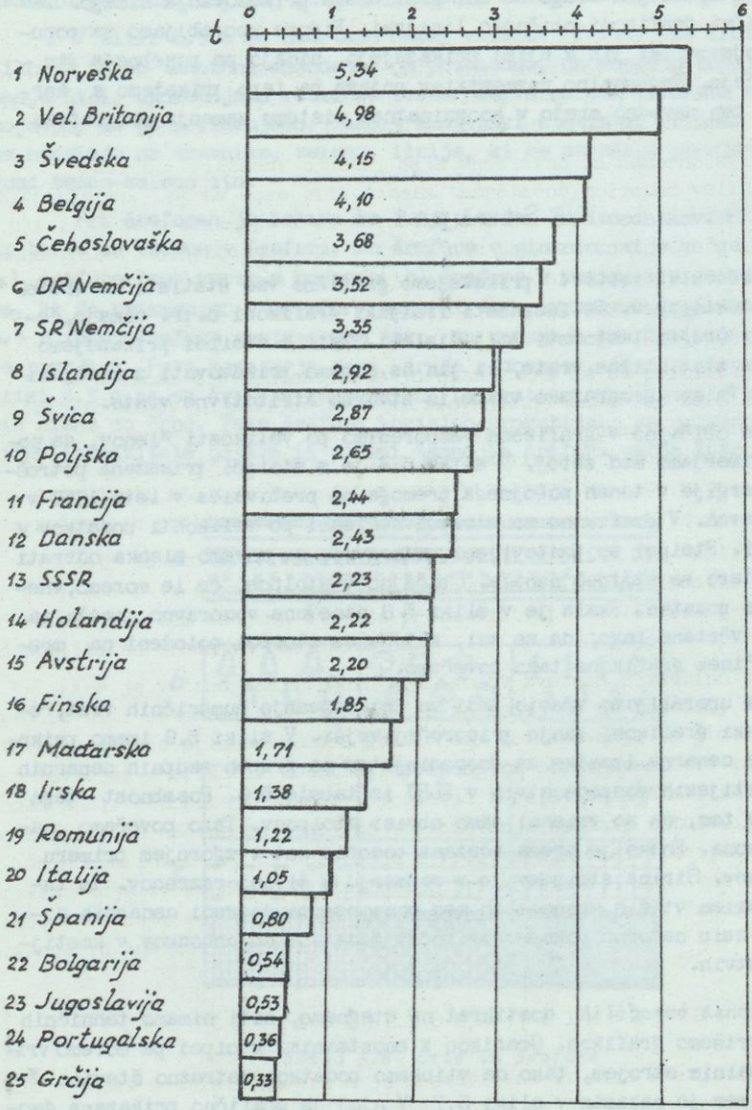
Stolpce uporabljamo včasih tudi za prikazovanje numeričnih vrst, čeprav so linijski grafikoni zanje priporočljivejši. V sliki 5.9 imamo prikazane povprečne denarne izdatke za gospodinjstvo po grupah skupnih denarnih dohodkov v kmetijskih gospodarstvih v FLRJ iz tabele 5.6. Posebnost tega grafikona je v tem, da so vrisani samo obrisi stolpcev. Tako povečamo nazornost grafikona. Enako je mreža včrtana podobno kot v zgornjem primeru samo do stolpcev. Širina stolpcev je v razmerju s širino razredov. Le tako dobimo pravilen vtis o spremembah med posameznimi razredi denarnih dohodkov. Slika zelo nazorno pokaže odvisnost izdatkov od dohodkov v kmetijskih gospodarstvih.

5.32 V *tipkanih poročilih* dostikrat ne utegnemo, niti nimamo tehničnih možnosti, da vrišemo grafikon. Grafikon z enostavnimi stolpci pa moremo vrisati kar s pisalnim strojem, tako da vtipkamo podatkom ustrezno število X ali I. Tak primer je nakazan v sliki 5.7 V njej je grafično prikazana geografska vrsta za narodni dohodek po republikah v letu 1953 iz tabele 5.1.

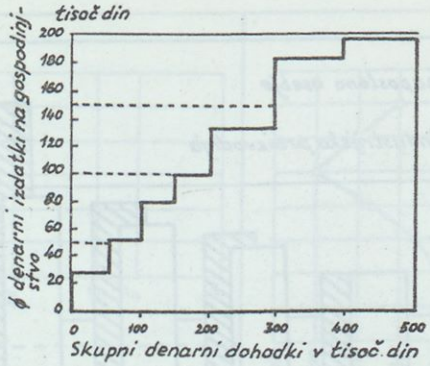
	v 10^2 mlj	1 2 3 4 5
Srbija	519	X X X X X X X X X X I
Hrvatska	395	X X X X X X X X
Slovenija	238	X X X X X
BiH	188	X X X X
Makedonija	82	X I
Črna gora	22	I

X = 50 mlj din
I = 25 mlj din

Slika 5.7 Narodni dohodek v FLRJ v letu 1956 po republikah



Slika 5.8 Poraba skupne energije na prebivalca v tonah pogojnega premoga v letu 1955 (Vir: SG 58)



Slika 5.9 Povprečni denarni izdatki po grupah skupnih denarnih dohodkov v kmetijskih gospodarstvih v letu 1956 v Sloveniji (Vir: Tabela 5.6)

5.33 *Več statističnih vrst* s stolpci prikazujemo poredko, ne glede na to, da moremo na istem grafikonu prikazati le istovrstne količine. Razen tega je tudi tehnično težko prikazati veliko število raznovrstnih stolpcev. V sliki 5.10 sta s stolpci prikazani indeksni vrsti za zaposleno osebje in industrijsko proizvodnjo v FLRJ v razdobju 1953–57. Indeksi, ki jih podrobneje obravnavamo pri relativnih številih, kažejo razmerja podatkov v statistični vrsti v primerjavi na določen člen – osnovo primerjave. V naši sliki je osnova primerjave leto 1953. Iz stolpcev nazorno vidimo, da proizvodnost dela po letih raste, ker se stolpci za indekse proizvodnje hitreje večajo kot stolpci za indekse zaposlenih.

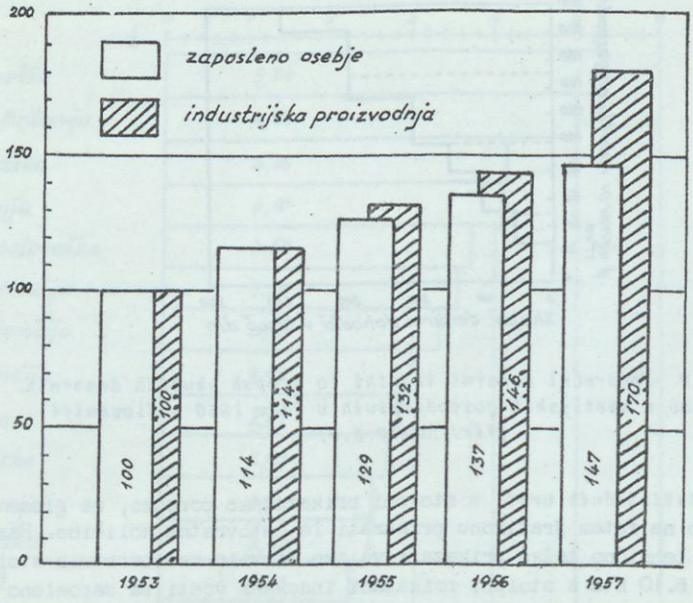
Stolpci so v sliki 5.10 risani tako, kakor da bi bili prekriti. To poveča učinek grafikona. Z različnima šrafurama poudarimo, da grafikon prikazuje dve različni časovni vrsti.

Linijski grafikoni

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

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

Za risanje grafikonov z linijskimi grafikoni veljajo ustaljena pravila, s katerimi dosežemo, da je grafikon pregleden in nazoren in da nedvoumno prikaže proučevani pojav.



Slika 5.10 Indeksi zaposlenega osebja in indeks industrijske proizvodnje v razdobju 1953-1957 v FLRJ (Vir: SG 58)

Ta *pravila* moremo v kratkem združiti v naslednje točke:

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

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

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

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

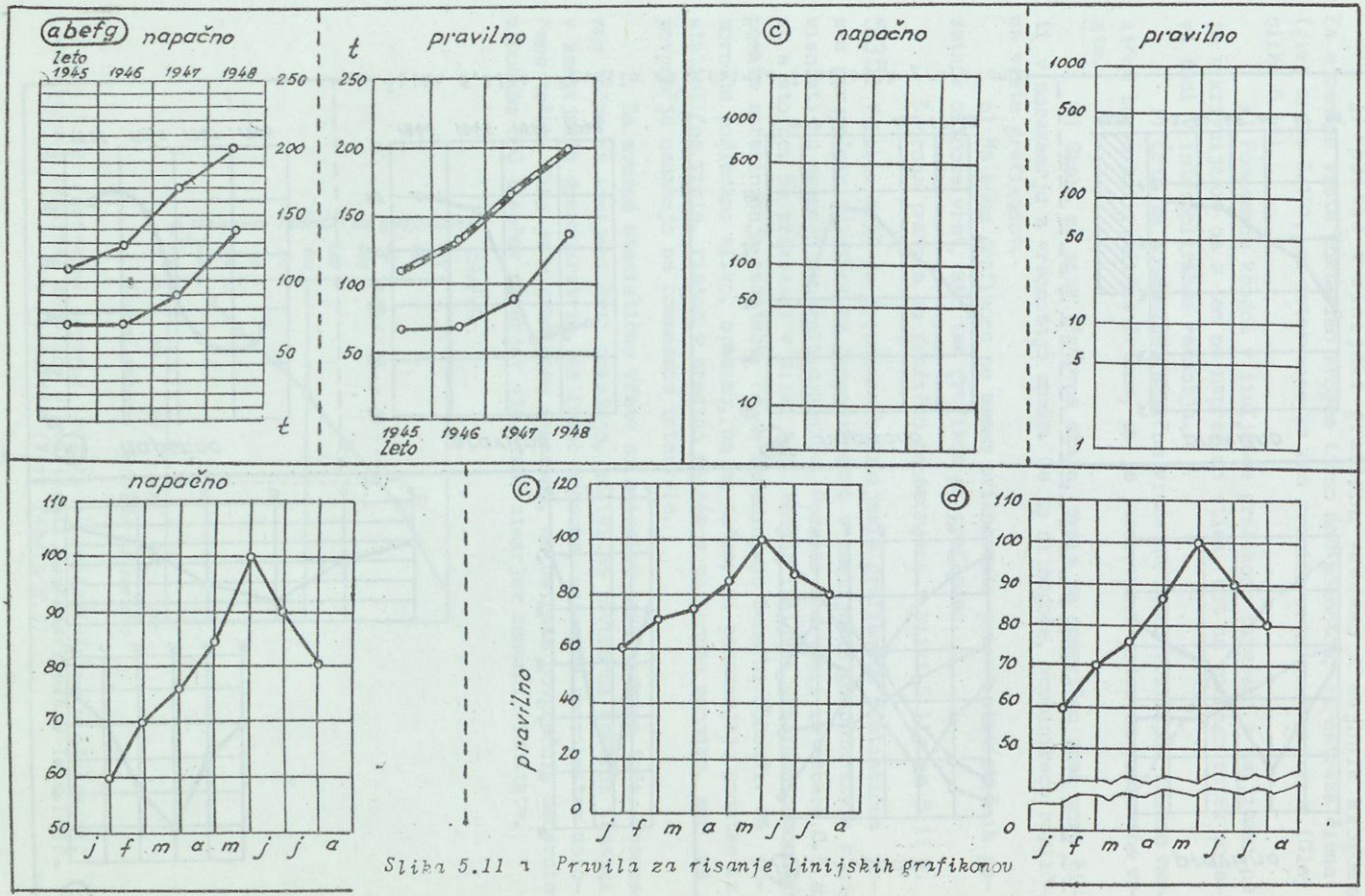
e) Da laže čitamo podatke, ima grafikon po pravilu včrtano mrežo.

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

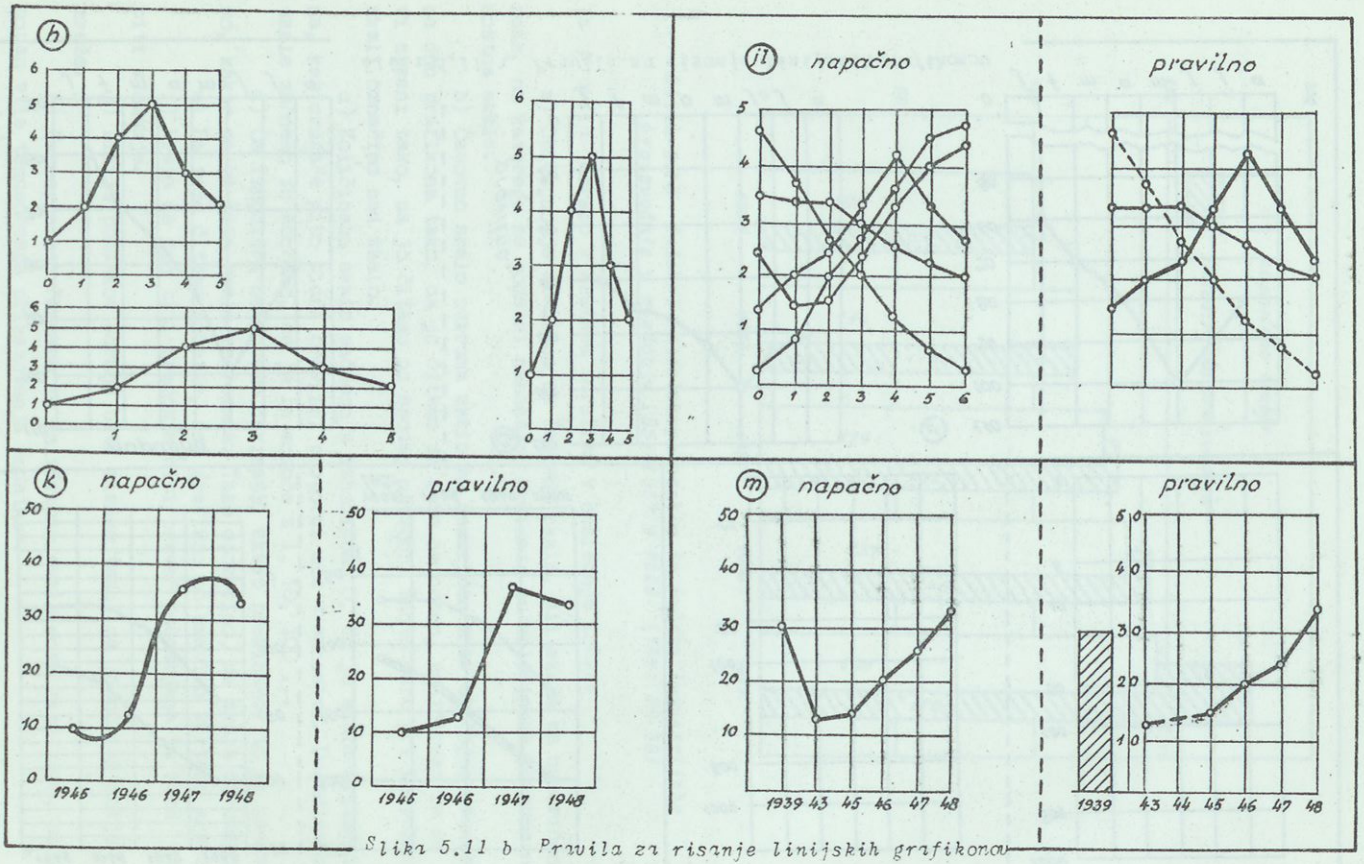
g) Linije podatkov morajo biti risane tako, da jasno izstopijo iz mreže.

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

i) Če v istem delu prikazujemo več sorodnih grafikonov, mora biti



Slika 5.11 : Pravila za risanje linijskih grafikonov



Slika 5.11 b Pravila za risanje linijskih grafikonov

razmerje med skalama v vseh grafikonih isto.

j) Podatke v linijski grafikon vnesemo z majhnim krogcem. Za momentne vrste postavimo krogce točno nad momentom, ki ustreza členu (glej sliko 5.13), za intervalne pa nad sredino ustreznih razmakov. (Glej sliko 5.14).

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

l) Črte za posamezne različne vrste na istem grafikonu rišemo različne barve ali različno črtane, da je jasno viden potek za posamezne vrste.

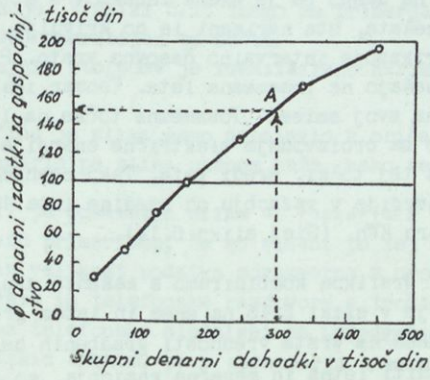
m) Skala za znak (časovna skala, skala za numeričen znak mora biti v sorazmerju z vrednostjo znaka. Če to ni mogoče, kombiniramo različne vrste grafikonov.

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

Zgornja pravila so grafično ponazorjena v sliki 5.11 a in 5.11 b.

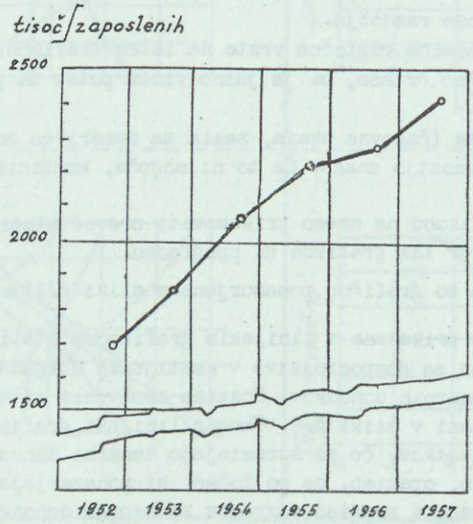
5.35 V sliki 5.12 je prikazana z linijskim grafikonom statistična vrsta povprečnih izdatkov za gospodinjstvo v kmetijskih gospodarstvih v odvisnosti od skupnih denarnih dohodkov. Podatke smo vzeli iz tabele 5.3 in so s stolpci že prikazani v sliki 5.8. Vendar linijski grafikon nazorneje pokaže spreminjanje izdatkov, če se spreminjajo denarni dohodki. Če pozorneje pogledamo sliko, opazimo, da so točke, ki ponazarjajo podatke iz statistične vrste, risane v sredino razredov za skupne dohodke, ker se povprečja nanašajo na posamezne intervale.

Za zvezne statistične vrste so smiselne tudi veznice točk. Posamezne točke na lomljeni črti pokažejo, koliki so povprečni denarni izdatki v kmetijskih gospodarstvih, ki imajo dohodke, ustrezne koordinatam določene točke. Tako je v grafikonu nakazano, da ustreza povprečnim denarnim dohodkom 300 tisočev dinarjev 150 tisoč dinarjev denarnih izdatkov.



Slika 5.12 Povprečni denarni izdatki za gospodinjstvo v odvisnosti od skupnih denarnih dohodkov v kmetijskih gospodarstvih v letu 1956 v Sloveniji (Vir: Tabela 5.6)

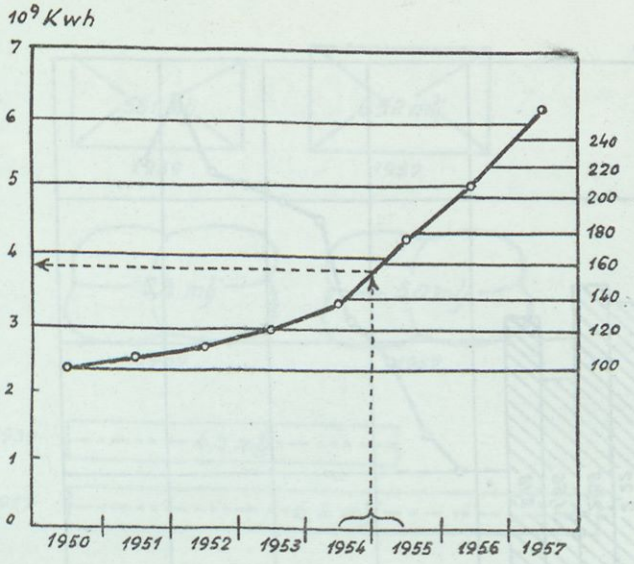
5.36 V sliki 5.13 je prikazana statistična vrsta števila zaposlenih v družbenem sektorju v FLRJ po letih v razdobju 1952-1957, iz tabele 5.4. Ker se podatki nanašajo vsako leto na konec septembra, je točka, ki prikazuje število zaposlenih na ta dan, postavljena nad ustrezno vrednostjo v letnih razmakih. Tudi za ta primer so točke na veznici linearna interpolacija za število zaposlenih za vmesne časovne momente.



Slika 5.13 Število zaposlenih v družbenem sektorju v FLRJ v razdobju 1952-1957 (Vir: Tabela 5.4)
Točke vnešene nad konec septembra v letnih razmakih

5.37 Slika 5.14, ki prikazuje proizvodnjo električne energije, je zanimiva, ker ima dvojno skalo. Na levo je skala za količino električne energije, merjeno naturalno, na desno pa je skala indeksov s stalno osnovo leto 1950. Da se skali ne mešata, sta narisani le do krivulje, vsaka s svoje strani. Ta grafikon prikazuje intervalno časovno vrsto. Posamezni vrisani podatki se torej nanašajo na posamezna leta. Vendar imajo tudi točke na lomljeni črti grafikona svoj smisel. Posamezne točke na liniji pomenijo linearno interpolacijo za proizvodnjo električne energije v letu, za katerega je čas, ki ustreza tej točki, sredi leta. Tako ocenimo, da je bila proizvodnja električne energije v razdobju od sredine leta 1954 do sredine leta 1955 3,85 milijard KWh. (Glej sliko 5.14).

5.38 Postikrat linijski grafikon kombiniramo s kakšnim drugim načinom grafičnega prikaza. Tako je v sliki 5.15 na enem in istem grafikonu prikazana povprečna letna in mesečna vrsta vrednosti gradbenih del v LRS. Ker je nemogoče drugače vskladiti letna in mesečna razdobja, so letna povprečja narisana s stolpci, mesečni podatki v letu 1958 pa z linijskim grafikonom.



Slika 5.14 Proizvodnja električne energije v FLRJ v razdobju 1950-1957 (Vir: Tabela 5.3)

Figure

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

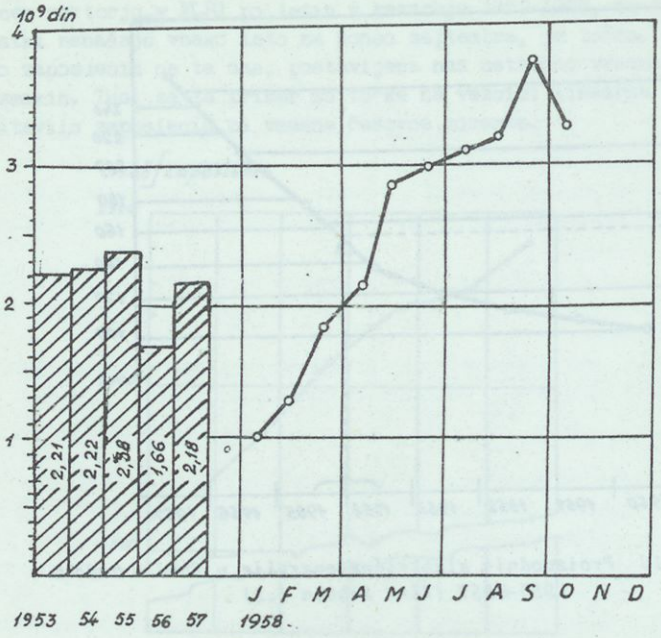
Figure vključujemo v grafikone na tri načine. Včasih spremlja *običajen grafikon* s stolpci ali linijami slika, ki laiku približa pojav, ki ga grafikon prikazuje. V sliki 5.13 imamo za primerjavo poštnih uslug v letu 1957 z letom 1939 uporabljen ta način.

Poleg običajnih stolpcev je idealizirano narisan pojav, ki ga stolpci prikazujejo.

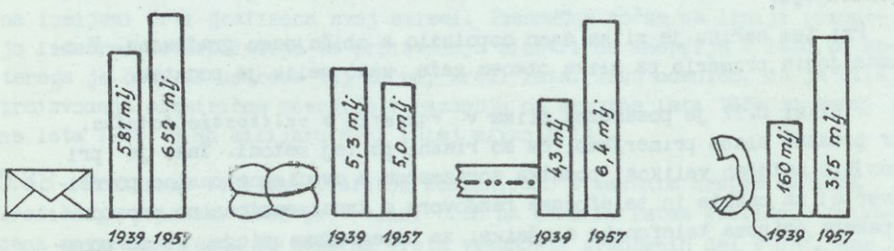
Pri tem načinu je slika samo dopolnilo k običajnemu grafikonu. V obeh nadaljnjih primerih pa slika obenem kaže, kako velik je podatek.

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

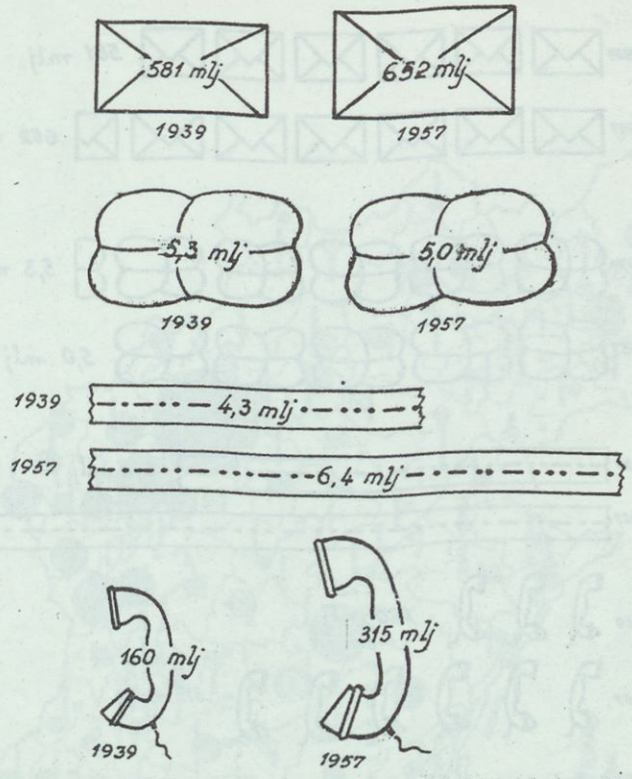
5.41 Najbolje je, da prikažemo podatke s figurami tako, da velikost po-



Slika 5.15 Vrednost gradbenih del v LRS v milijonih dinarjev (Vir: Indeks)



Slika 5.16 PTT usluge v FLRJ v letu 1957 v primerjavi z letom 1939 (Vir: Indeks)

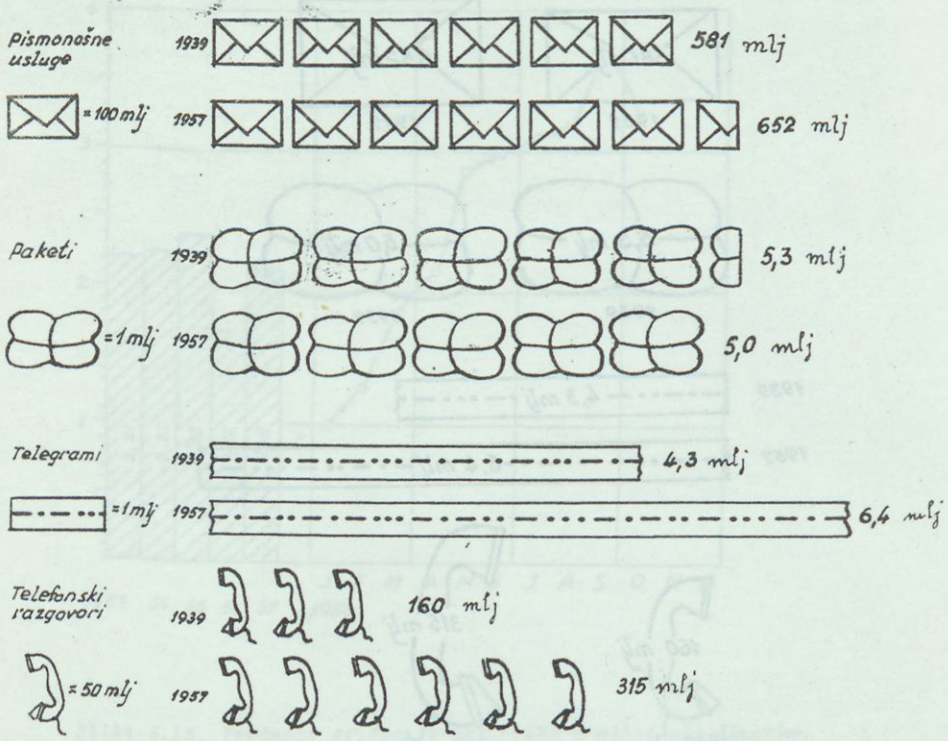


Slika 5.17 PTT usluge v FLRJ v letu 1957 v primerjavi z letom 1939 (Vir: Indeks)

datka ponazorimo z *ustreznim številom enakih, idealiziranih figur.* V tem primeru odpadejo vsi dvomi in nevšečnosti zaradi različne razsežnosti pojavov, slabe primerljivosti prostornin itd. Po tem načinu so prikazani podatki o PTT uslugah v sliki 5.18. Vsaka slikica danega elementa predstavlja določeno velikost pojava. Tako pri pismonošnih uslugah ena kuverta ponazarja 100 milijonov pismonošnih usług, en paket 1 milijon paketov, 1 cm telegrafskega traku 1 milijon telegrafskih usług in ena telefonska slušalka 50 milijonov telefonskih razgovorov. Kakor vidimo iz slike 5.18, moremo z delom osnovne slikice nakazati podatek, ki je manjši od podatka, ki ga prikazuje cela figura.

Kartogrami

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



Slika 5.18 PTT usluge v FLRJ v letu 1957 v primerjavi z letom 1939 (Vir: Indeks.)

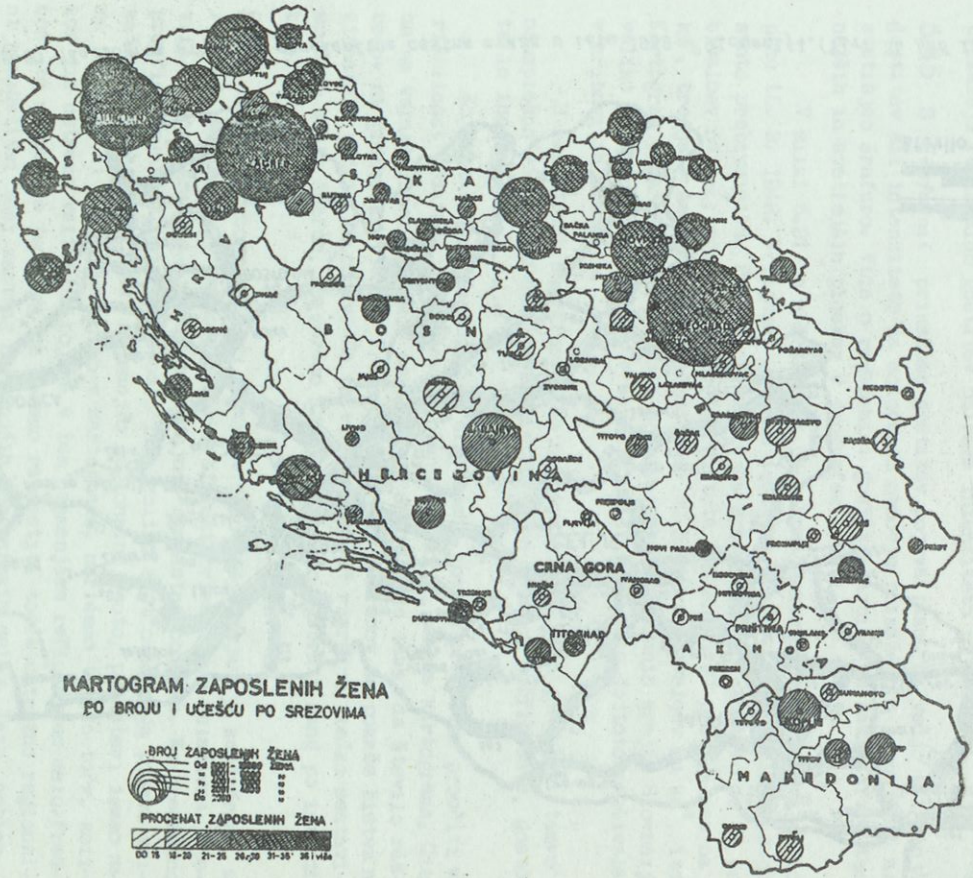
grafska karta. Vse grafikone, ki nam kakor koli prikažejo razmestitev določenega pojava v geografski karti, imenujemo kartograme v širšem smislu.

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

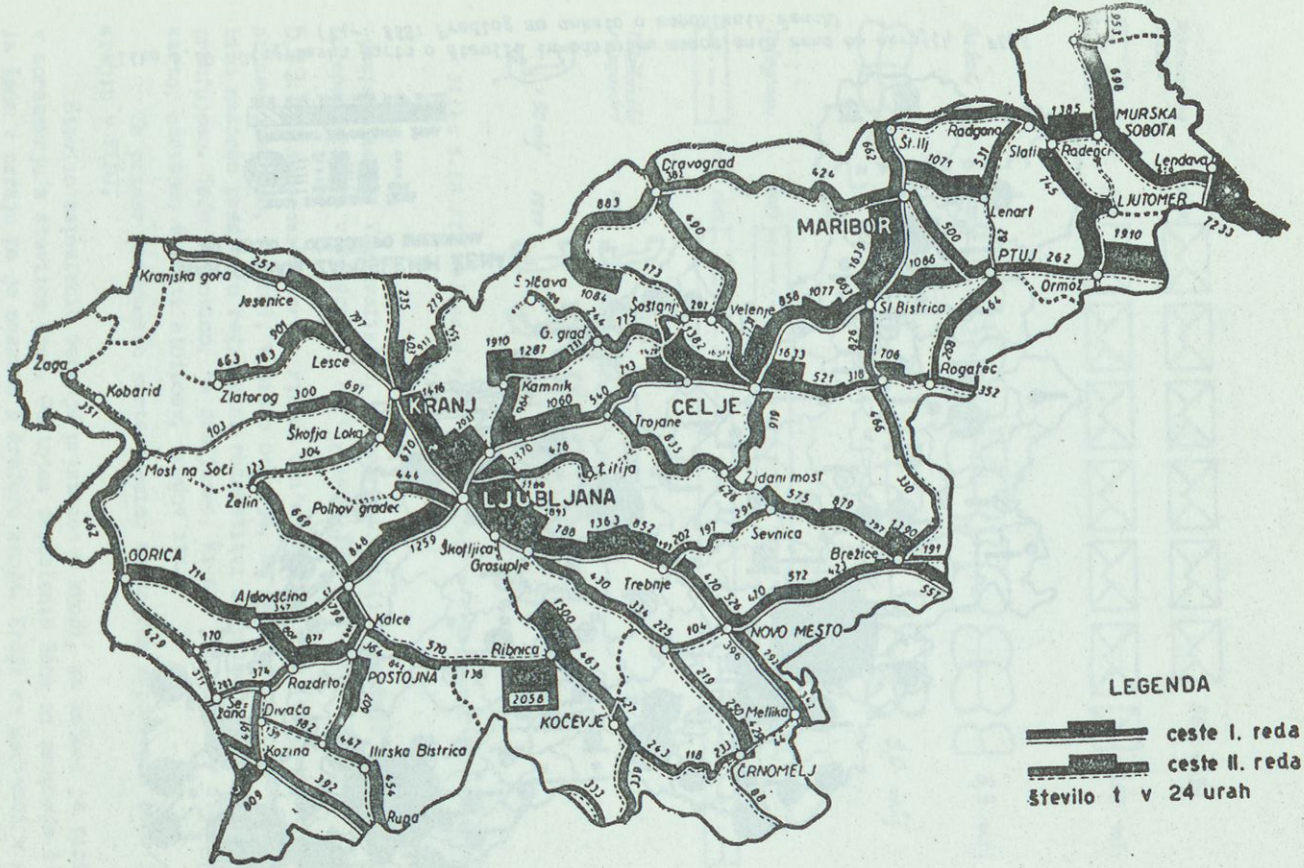
Za primer diagramske karte vzemimo kartogram zaposlenih žena po okrajih v FLRJ.

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

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



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



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

V kartogramu (sl. 5.20) je s pasovi, ki so postavljeni vzdolž posameznih cest, prikazana povprečna obremenitev cest v 24 urah v tonah. Debelina pasov je v sorazmerju z obremenitvijo. Tako je nazorno prikazana obremenitev posameznih prometnih žil.

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

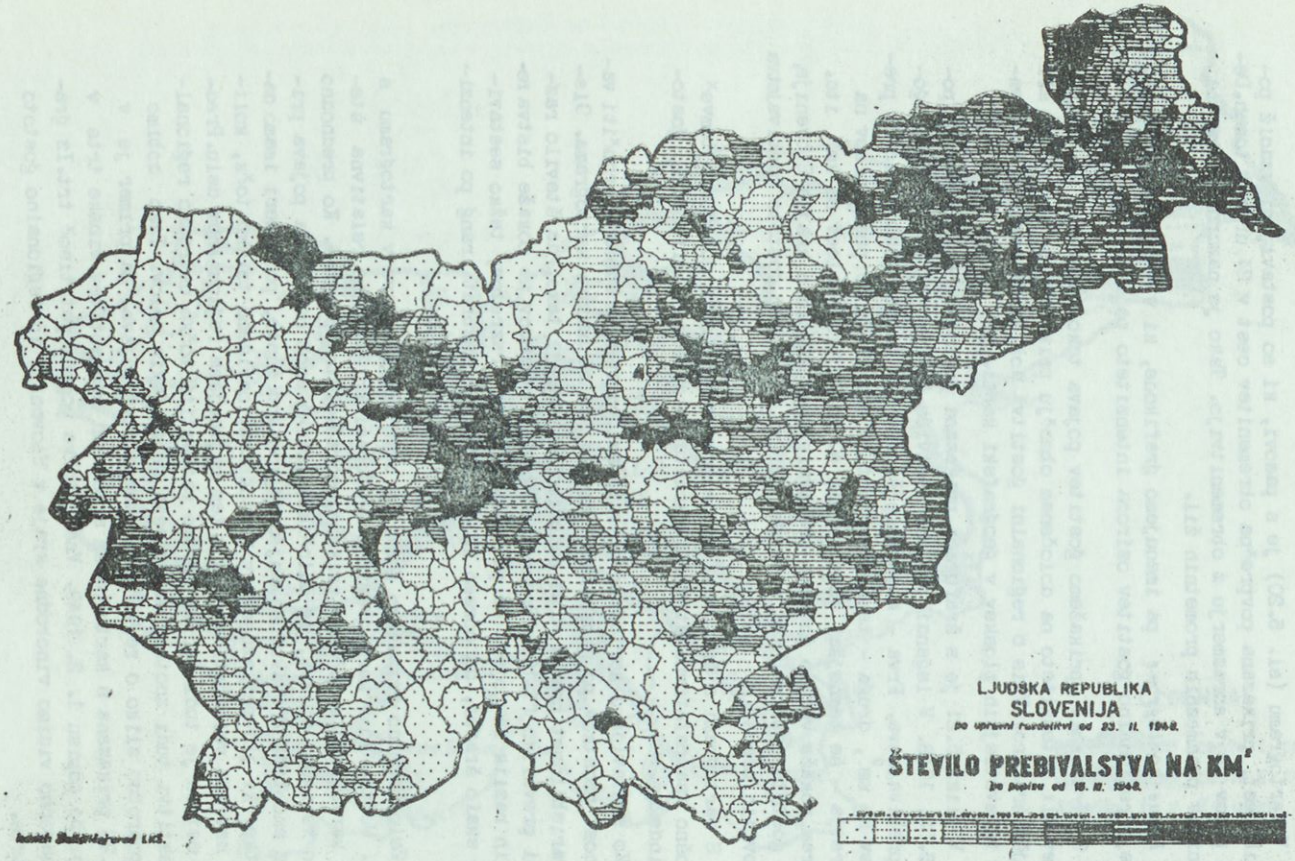
5.45 S *šrafurami* prikažemo gostitev pojava tako, da večjo ali manjšo gostitev ali intenziteto na določenem območju prikažemo s temnejšo ali svetlejšo šrafuro. Vtis o regionalni gostitvi dobimo iz razmestitve temnejših in svetlejših ploskev v geografski karti.

V sliki 5.21 je s šrafurami prikazana gostota prebivalstva po popisu 15. 3. 1948. V legendi je dana lestvica šrafur, ki je skladna z gostoto prebivalstva. Prva - najsvetlejša šrafura pomeni gostoto do 30 prebivalcev na km^2, druga - temnejša, gostota nad 30 do 40 prebivalcev na km^2, tretja - še temnejša, gostoto nad 40 do 50 prebivalcev na km^2 itd. Kartogram pokaže veliko gostoto prebivalstva v vzhodnih krajih Slovenije, veliko gostoto prebivalstva v savski dolini in majhno gostoto prebivalstva v hribovitih predelih.

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

Ko sestavljamo kartogram s šrafurami, moramo predhodno proučiti variabilnost podatkov po geografskih enotah, ki so osnova kartograma. Glede na variabilnost podatkov se odločimo za razrede tako, da število razredov ni preveliko (3-10). Premajhno število razredov ne pokaže bistva regionalnih razlik za pojav. Če pa vzamemo preveč razredov, težko sestavimo tako skalo šrafur, da moremo iz nje nedvoumno razbrati rang po intenziteti.

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



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



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

Šesto poglavje

RELATIVNA ŠTEVILA

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

Vendar za proučevanje statističnih podatkov ni zadostna le površna primerjava, kakršno dobimo iz statistične tabele osnovnih podatkov. Primerjavo podatkov zelo poglobimo in olajšamo z relativnimi števili. Veliko boljšo predstavo o številu prebivalstva v Sloveniji v primerjavi s številom prebivalstva Jugoslavije, dobimo, če rečemo, da je v Sloveniji 8,33 % od skupnega števila prebivalstva Jugoslavije, kakor pa če navedemo, da je bilo v Sloveniji po popisu 31. 3. 1953 1,433.425 prebivalcev od skupnega števila 13,983.573 prebivalcev v Jugoslaviji. Enako dobimo najboljši pregled o prirastku števila prebivalstva v Sloveniji, če ga izrazimo relativno in rečemo, da se je število prebivalstva v razdobju med obema popisoma 15. 3. 1948 in 31. 3. 1953 povečalo v Sloveniji za 5,33 %. Tako določenega in hitrega vtisa o povečanju niti ne dobimo, če povemo, da je bilo število prebivalstva 15. 3. 1948. 1,391.873, 31. 3. 1953 pa 1,433.425 prebivalcev, niti če povemo, da se je število prebivalstva v petih letih povečalo za 74.552 prebivalcev. Tudi cena pri proizvajalcu je relativno število, ker je razmerje med proizvodnimi stroški in proizvedeno količino itd.

Relativno število iz dveh podatkov, ki ju primerjamo, je v splošnem kvocient med primerjanima podatkoma. Jasno je, da enako kakor med seboj primerjamo le podatke, za katere je primerjava smiselna, izračunavamo tudi relativna števila le iz podatkov, ki so v medsebojni vsebinski povezavi. Te kvociente izražamo v oblikah, ki so priklad nejše za prikazovanje in analizo (n. pr. v procentih, indeksih itd.).

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

a) *strukture* ali *razčlenitvena števila*, kadar primerjamo del s celoto (n. pr. od skupnega števila prebivalstva v Jugoslaviji je bilo v letu 1953 8,7 % v Sloveniji.)

b) *koeficiente* in *gostote*, kadar primerjamo raznovrstne podatke, ki so v medsebojni zvezi (n. pr. proizvodnja jekla na 1 prebivalca je bila v letu 1957 v FLRJ 54,7 kg/prebivalca);

c) *indekse*, kadar primerjamo prirejene istovrstne podatke (n. pr. indeks spremembe števila prebivalstva v Sloveniji od popisa 1948 do po-

pisa 1953 je 105,4, če vzamemo število prebivalstva v letu 1948 kot 100,0).

STRUKTURE ALI RAZČLENITVENA ŠTEVILA

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

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

Ta princip na splošno uporabljamo vselej, kadar hočemo pregledno in primerljivo prikazati notranji sestav statističnih populacij. Največkrat ne izračunavamo samo enega strukturnega pokazatelja, temveč cele vrste struktur, ki nazorno prikažejo sestav populacij. Strukture običajno izražamo v delih od celote ali v odstotkih. Če pa so deli izjemno majhni, izražamo strukture tudi v promilih.

Strukturni koeficient Y_i^0 dobimo, da delni podatek delimo s celoto.

Strukturni odstotek dobimo tako, da delni podatek delimo s skupnim, kvocient pa pomnožimo s 100. Če izražamo *strukturni delež v promilih*, pa kvocient pomnožimo s 1000. Računski postopek moremo nakazati tudi z obrazci:

$$Y_i^0 = \frac{Y_i}{Y}; \qquad Y_i\% = 100 \cdot \frac{Y_i}{Y}; \qquad Y_i‰ = 1000 \cdot \frac{Y_i}{Y} \qquad (3.1)$$

Pri tem pomeni: Y_i = podatek za del populacije, Y = podatek za populacijo, Y_i^0 = strukturni koeficient, Y_i % = strukturni odstotek, Y_i ‰ = = strukturni delež, izražen v promilih.

6.3 Uporabnost enostavnih strukturnih serij za statistično analizo moremo prikazati na primeru izvoza iz FLRJ v razdobju 1954–1957 po stopnji obdelave izvoženih artiklov.

Tabela 6.1 Vrednost izvoza iz FLRJ po stopnji obdelave izvoženih proizvodov v razdobju 1954-1957. (Vir SG 58).

Stopnja obdelave	Vrednost v 10^9 dinarjev				Struktura			
	1954	1955	1956	1957	1954	1955	1956	1957
Neobdelani proizvodi	27,1	29,2	35,5	37,5	37,6	37,9	36,6	31,7
Proizvodi srednje obdelave	33,0	35,5	41,6	49,1	45,8	46,1	42,9	41,4
Proizvodi visoke obdelave	12,0	12,3	19,9	31,9	16,6	16,0	20,5	26,9
S k u p n o	72,1	77,0	97,0	118,5	100,0	100,0	100,0	100,0

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

6.4 Večkratne strukture. Če je populacija razdeljena po več znakih hkrati, moremo zanjo izračunati več vrst struktur. Vsaka od njih po svoje osvetli prikazano populacijo. Vzemimo kot primer razdelitev družbenega produkta v FLRJ v letu 1956 po elementih in gospodarskih sektorjih.

Tabela 6.2 Družbeni produkt v FLRJ v letu 1956 po elementih in gospodarskih sektorjih v 10^9 dinarjev (Vir: SG 58).

Gospodarski sektor	Plače	Akumulacija in fondi	Amortizacija	Skupno
Splošno družbeni	231	721	149	1 101
Zadružni	20	28	3	51
Privatni	366	78	16	460
S k u p n o	617	827	168	1 612

V tej tabeli moremo vzeti kot celoto najprej skupen družbeni produkt (1.312 milijard dinarjev) in izračunati v ostotkih, kakšen delež te celote je vsak izmed podatkov. Plače v splošno družbenem sektorju predstavljajo po tem principu

$$100 \frac{231}{1612} = 14,4\ \%$$

od skupnega družbenega produkta

Vsi odstotki, izračunani na tej osnovi, so dani v tabeli 3.8.

Tabela 6.3 Struktura družbenega produkta v FLRJ v letu 1956 po elementih in gospodarskih sektorjih

Gospodarski sektor	Plače	Akumulacija in fondi	Amortizacija	Skupno
Splošno družbeni	14,4	44,7	9,2	68,3
Zadružni	1,2	1,8	0,2	3,2
Privatni	22,7	4,8	1,0	28,5
S k u p n o	38,3	51,3	10,4	100,0

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

Iz tabele 3.2 o sestavu družbenega produkta pa moremo izračunati tudi druge vrste struktur. Plače v splošno družbenem sektorju so del skupnega družbenega produkta, so pa istočasno del družbenega produkta v splošno družbenem sektorju (1101 milijard din) in del skupnih plač (317 milijard din). Zaradi tega moremo razen zgornje tabele o sestavu izračunati še dve tabeli struktur.

Tabela 6.4 Struktura družbenega produkta posameznih gospodarskih sektorjev po elementih

Gospodarski sektor	Plače	Akumulacija in fondi	Amortizacija	Skupno
Splošno družbeni	21,0	65,5	13,5	100,0
Zadružni	39,2	54,9	5,9	100,0
Privatni	79,6	16,9	3,5	100,0
S k u p n o	38,3	51,3	10,4	100,0

Tabela 6.5 Struktura elementov družbenega produkta po gospodarskih sektorjih

Gospodarski sektor	Plače	Akumulacija in fondi	Amortizacija	Skupno
Splošno družbeni	37,4	87,2	88,7	68,3
Zadružni	3,3	3,4	1,8	3,2
Privatni	59,3	9,4	9,5	28,5
S k u p n o	100,0	100,0	100,0	100,0

V tabeli 3.4 vidimo velike razlike v sestavu družbenega produkta po gospodarskih sektorjih. Medtem ko so na primer v splošnem sektorju plače samo 21,0 % družbenega produkta, pomenijo plače v zadružnem sektorju 38,2 %, v privatnem pa celo 79,3 % od skupnega družbenega produkta. Analogno so velike razlike tudi v drugih elementih.

Tudi tabela 3.5 je zelo poučna. Udeležba posameznega sektorja v vsakem izmed treh elementov je zelo različna in po svoje osvetli sestav družbenega produkta. Medtem ko je udeležba posameznega sektorja v skupni akumulaciji, fondih in amortizaciji razmeroma podobna, je udeležba v skupnih plačah bistveno različna.

Grafično prikazovanje struktur

Vpogled v sestav populacije dobimo zelo nazorno z grafikoni. Strukture moremo grafično prikazovati z vsemi vrstami grafikonov: s stolpci, krogi, linijami, figurami ita.

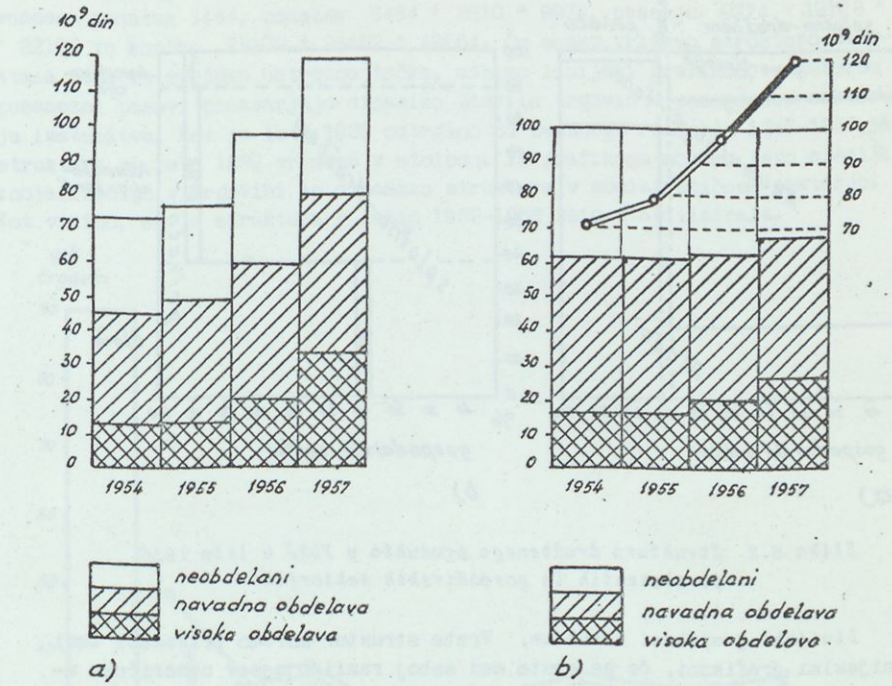
6.5 Strukturni stolpci. Z navadnim stolpcem prikažemo en sam podatek. Višina stolpca je v sorazmerju z velikostjo podatka, ki ga stolpec ponazarja. Če pa stolpec, ki predstavlja celoto, razdelimo v sorazmerju s sestavom populacije, dobimo strukturni stolpec, ki prikaže strukturo populacije. Več sorodnih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

Vzemimo kot primer vrednost izvoza iz Jugoslavije po stopnji obdelave v razdobju 1954–1957 iz tabele 3.1. Strukturni stolpci iz absolutnih podatkov v tej tabeli so prikazani v sliki 3.1a.

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

Veliko bolj neposredno je sprememba strukture izvoza vidna iz slike 3.1b, v kateri je struktura izvoza v posameznih letih prikazana v odstotkih z enako visokimi stolpci. Pri tem sicer izgubimo pregled o spremembah v skupni vrednosti izvoza, vendar boljši pregled o spremembah v sestavu izvoza odtehta to pomanjkljivost. To pomanjkljivost včasih odpravimo tako, da v isti sliki z linijskim grafikonom ponazorimo gibanje skupne vrednosti.

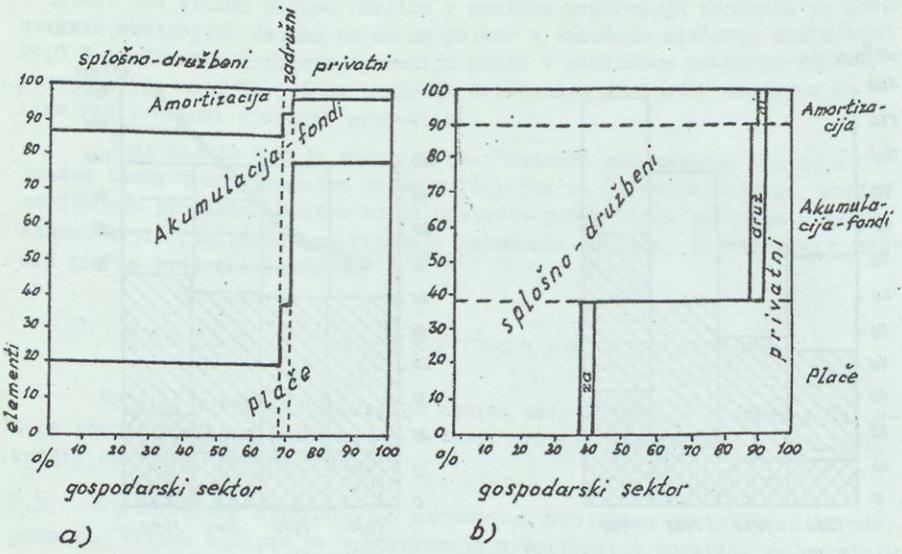
6.6 Strukturni stolpci večkratnih struktur. Širina strukturnih stolpcev iz prejšnjega primera je enaka za vsa leta. Moremo pa stolpce risati široke v sorazmerju z obsegom populacije. Vrsta takih strukturnih stolpcev pokaže spremembe v skupni velikosti pojava. Kljub temu pa so vsi strukturni stolpci enako visoki, da je iz njih neposredno vidna sprememba v strukturi. Ta postopek s pridom izkoriščamo za prikazovanje večkratnih struktur.



Slika 6.1 Struktura vrednosti izvoza iz FLRJ po stopnji obdelave izvoženih proizvodov (Vir: Tabela 6.1)

Vzemimo za primer strukturo družbenega produkta v FLRJ v letu 1953 po gospodarskih sektorjih in elementih iz tabele 3.2. Celoten družbeni produkt ponazorimo s kvadratom. Strukturo družbenega produkta pa prikažemo tako, da kvadrat razdelimo v sorazmerju z deležem posameznega sektorja v celotnem družbenem produktu. Celoten družbeni produkt je tako razdeljen v tri navpične stolpce, katerih širina je v sorazmerju z deležem posameznega sektorja. Če vsakega izmed teh treh stolpcev razdelimo v navpični smeri v sorazmerju s sestavom družbenega produkta za posamezni sektor po elementih, dobimo sliko 3.2a. Ta je fotografija tabele 3.4. Iz tega grafikona je zelo nazorno vidna struktura družbenega produkta po sektorjih, predvsem pa razlike v sestavu med sektorji. Poleg tega je ploščina posameznega pravokotnika v kvadratu sorazmerna z absolutno velikostjo določenega elementa oziroma s strukturnimi odstotki iz tabele 3.4.

Če kvadrat, ki je okvir grafikona, razdelimo najprej v sorazmerju s sestavom skupnega družbenega produkta po elementih, dobljene delne stolpce pa, podobno kakor v prejšnjem primeru, v sorazmerju s sestavom posameznega elementa po sektorjih, dobimo sliko 3.2b; ta grafično pove isto kakor tabela 3.5.



Slika 6.2 Struktura družbenega produkta v FLRJ v letu 1956 po elementih in gospodarskih sektorjih

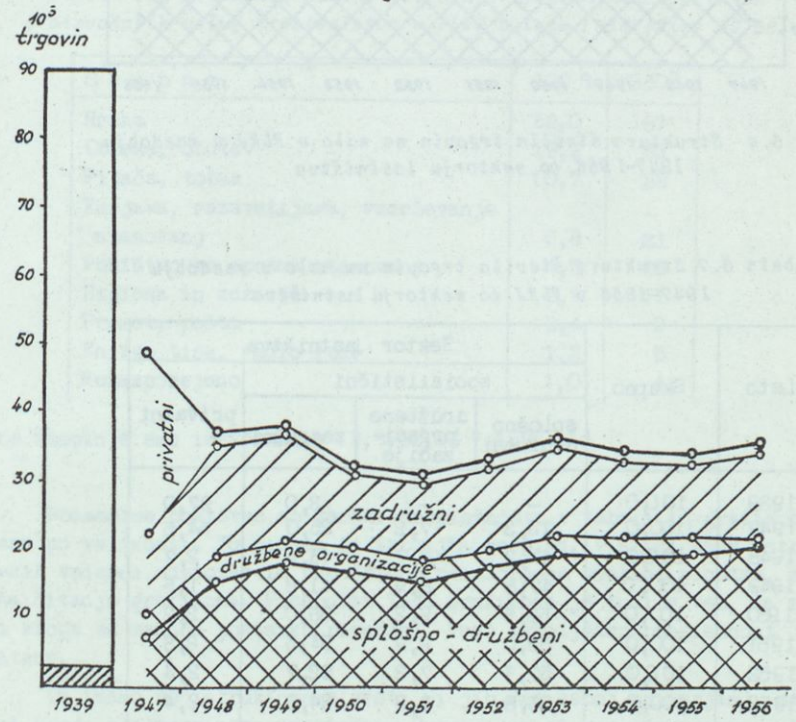
6.7 *Linijski grafikoni struktur.* Vrste struktur moremo prikazati tudi z linijskimi grafikoni, če se vrste med seboj razlikujejo v numeričnem ali časovnem znaku.

Kot primer vzemimo dinamiko strukture socializacije trgovine na malo v razdobju 1947-1953 v FLRJ.

Tabela 6.6 Število trgovin po sektorju lastništva v razdobju 1947-1956 v FLRJ (Vir: SG 57)

Leto	Skupno	Sektor lastništva			
		socialistični			privatni
		splošno družbeni	družbene organizacije	zadružni	
1939	89 391	-	-	2 721	86 970
1947	48 554	3 434	3 510	12 128	29 452
1948	36 730	15 145	3 557	17 073	985
1949	38 313	18 322	3 395	13 028	571
1950	32 748	17 533	3 023	11 457	732
1951	30 794	13 335	3 045	10 313	801
1952	33 303	19 174	2 433	10 903	790
1953	36 831	20 929	2 301	12 758	873
1954	35 228	20 171	1 977	12 215	835
1955	34 958	20 231	1 822	12 089	816
1956	36 257	22 528	328	12 243	858

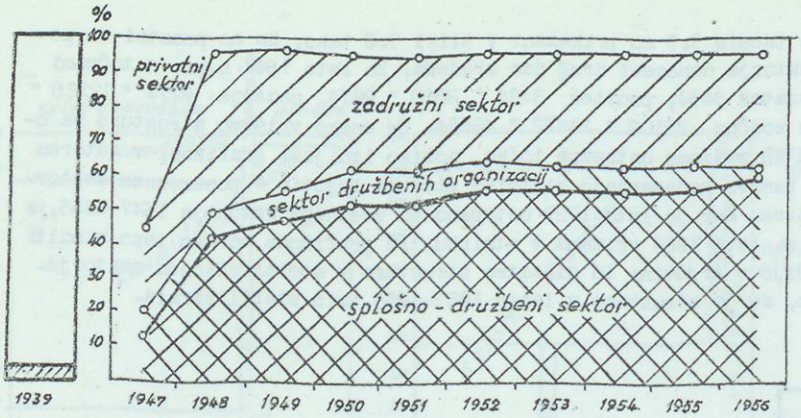
Podatki iz tabele 3.3 so prikazani v sliki 3.3 tako, da so podatki za posamezne sektorje naneseni drug nad drugega. Za leto 1947 n.pr. s točkami vnesemo podatek 3434, podatek 3434 + 3510 = 9974, nadalje 9974 + 12128 = = 22102 in končno 22102 + 23452 = 48554. Če enako vrišemo strukturo za ostala leta in zvežemo ustrezne točke, dobimo linijski grafikon, v katerem posamezni pasovi ponazarjajo dinamiko števila trgovin v posameznem sektorju lastništva. Ker je leto 1939 odtrgano od zveznega razdobja 1947-1953, je struktura za leto 1939 vrisana v stolpcu. Iz grafikona moremo lepo slediti socializacijo v trgovini in dinamiko strukture v socialističnem sektorju. Kot vidimo, se je struktura v letih 1952-1956 zelo stabilizirala.



Slika 6.3 Struktura števila trgovin na malo v FLRJ v razdobju 1947-1956 po sektorjih lastništva

Enako kakor smo prikazali strukturo z linijskim grafikonom absolutnih podatkov, moremo prikazati tudi vrste strukturnih odstotkov. Ta slika še nazorneje pokaže dinamiko strukture kot slika absolutnih podatkov, izgubimo pa sliko o spremembah skupnega števila trgovin.

V sliki 3.4 je z linijskim grafikonom prikazana struktura števila trgovin na malo iz tabele 3.7.



Slika 6.4 Struktura števila trgovin na malo v FLRJ v razdobju 1947-1956 po sektorju lastništva

Tabela 6.7 Struktura števila trgovin na malo v razdobju 1947-1956 v FLRJ po sektorju lastništva

Leto	Skupno	Sektor lastništva			
		socialistični			privatni
		splošno družbeni	družbene organizacije	zadružni	
1939	100,0	-	-	3,0	97,0
1947	100,0	13,3	7,2	25,0	54,5
1948	100,0	41,2	9,7	43,4	2,7
1949	100,0	47,8	8,9	41,8	1,5
1950	100,0	53,3	9,2	35,0	2,2
1951	100,0	53,0	9,9	34,5	2,3
1952	100,0	57,3	7,8	32,7	2,4
1953	100,0	56,8	8,2	34,3	2,4
1954	100,0	57,2	5,3	34,7	2,5
1955	100,0	58,0	5,2	34,6	2,3
1956	100,0	62,1	1,7	33,8	2,4

6.8 Strukturni krogi. Za prikazovanje struktur se izkaže primeren tudi krog, ker krog nazorno prikazuje celoto. Za strukturne dele celote vzamemo *krogove izseke*, kot posameznega izseka pa je v sorazmerju z velikostjo strukturnega deleža. Ker četrtina kroga ponazarja 25 %, polovica kroga pa 50 %, moremo s pogledom enostavno oceniti strukturne deleže.

Ker imamo običajno kotomere z ločnimi stopinjami, moramo struktur-

ne deleže $Y\%$ prea risanjem preračunati po obrazcu

$$Y_1^{st} = 3{,}6 . Y_1\% \qquad (6.2)$$

v ločne stopinje Y_1^{st}.

Kot primer bomo prikazali strukturo osebne potrošnje materialnih dobrin in proizvodnih uslug za prebivalstvo v FLRJ v letu 1953.

Tabela 6.8 Struktura osebne potrošnje materialnih dobrin in proizvodnjih uslug prebivalstva v FLRJ v letu 1956 (Vir: SG 58)

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

Ločne stopinje smo izračunali: $3{,}6 . 52{,}0\% = 187^{\circ}$ itd.

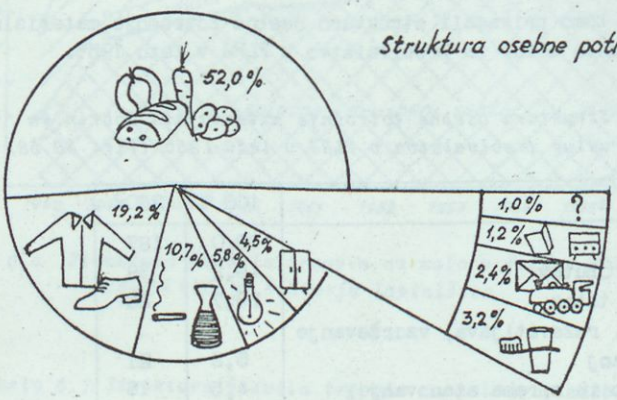
Posamezne postavke so zaradi preglednosti v tabeli in grafikonu nanizane po velikosti. To olajša in izboljša pregled. V sliki so v ustreznih izsekih vpisani odstotki in shematično ponazorjena vsebina podatka, kar olajša čitanje grafikona. Posebnost tega grafikona je tudi v tem, da so izseki kroga za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

Če imamo več strukturnih vrst, ki jih med seboj primerjamo, strukturni krogi običajno niso prikladni. Primerljivost med strukturami, če so prikazane s krogi, je namreč veliko slabša, kakor če uporabimo stolpce ali linije. Kljub temu včasih kroge uporabljamo za prikazovanje struktur. Uporabni so n.pr. za prikazovanje regionalnih sprememb struktur. Strukturne kroge v tem primeru vnašamo v ustrezna področja v mreži geografske karte. Ta način je razmeroma ugoden zaradi primerne oblike kroga. Razen tega pa tudi strukturni stolpci v tem primeru niso neposredno primerljivi.

6.9 Tudi strukture dveh populacij, ki sta si v zvezi, tako kakor n.pr. uvoz in izvoz, nabavljeno in prodano, dohodki in izdatki, dostikrat prikazujemo s krogi tako, da posamezno populacijo prikažemo s polovico kroga. Strukturne odstotke $Y_1\%$ v tem primeru izračunamo v ločne stopinje Y_1^{st} po obrazcu

$$Y_1^{st} = 1{,}8 . Y_1\% \qquad (6.3)$$

ker celoto predstavlja le 180°.



Slika 6.5 Struktura osebne potrošnje materialnih dobrin in materialnih uslug za prebivalstvo FLRJ v letu 1956 (Vir: Tabela 6.8)

Če ne upoštevamo absolutno velikost pojava, imata obe polovici kroga isti radij. Zaželeno pa je, da z velikostjo kroga nakažemo velikost pojava. Ker mora biti ploščina, ne pa radij kroga ali polkroga, proporcionalna velikosti pojava, radije krogov, za katere želimo, da so njihove ploščine v sorazmerju z velikostjo pojava, izračunavamo po obrazcu

$$r_1 = r_0 \sqrt{\frac{Y_1}{Y_0}} \qquad (6.4)$$

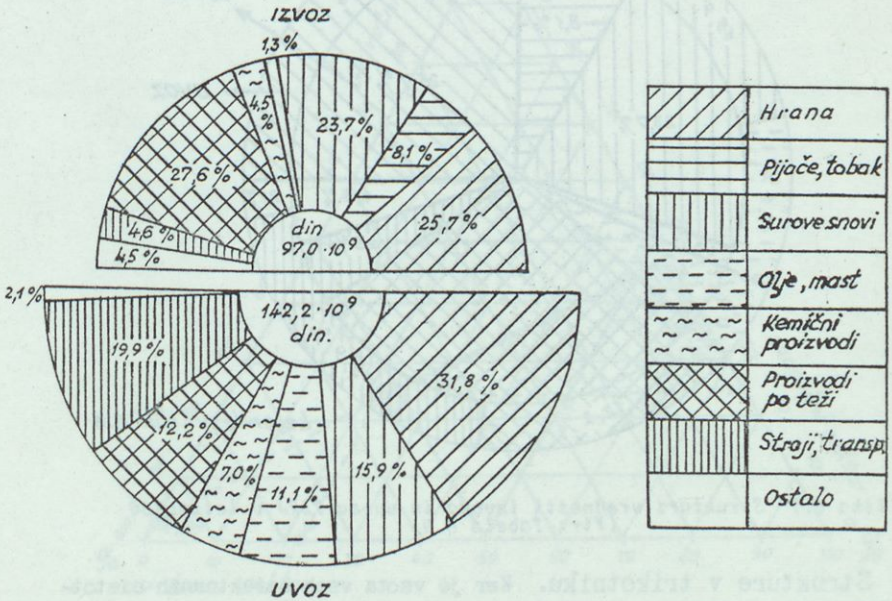
Pri tem pomeni: r_0 = znan radij kroga, ki ustreza podatku Y_0, r_1 = radij kroga za podatek Y_1.

Običajno vnaprej določimo radij kroga za največji podatek, radije vseh drugih krogov pa izračunamo po obrazcu 3.4.

Za primer vzemimo strukturo vrednosti izvoza in uvoza FIRJ v letu 1953.

Tabela 6.9 Struktura vrednosti izvoza in uvoza FLRJ v letu 1956

Skupine artiklov	Struktura v %		Stopinje v polkrogih	
	izvoz	uvoz	izvoz	uvoz
S k u p n o 10^9 din	97,0	142,2		
	100,0	100,0	180,0	180,0
Hrana	25,7	31,8	43,3	57,2
Pijače in tobak	8,1	0,0	14,6	0,0
Surove snovi brez goriva	23,7	15,9	42,6	28,3
Min.goriva, maziva živ.rast.olje in mast	1,3	11,1	2,3	20,0
Kemični proizvodi	4,5	7,0	8,1	12,6
Proizvodi klasirani pretežno po snovi	27,6	12,2	49,7	22,0
Stroji in transportna sredstva	4,6	19,9	8,3	35,8
Ostalo	4,5	2,1	8,1	3,8



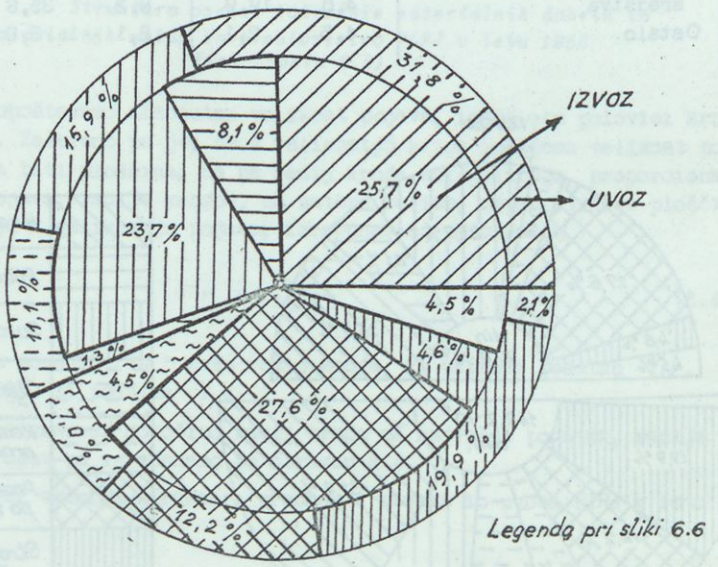
Slika 6.6 Struktura vrednosti izvoza in uvoza FLRJ v letu 1956 (Vir: Tabela 6.9)

Ker smo se odločili, da bo radij večjega polkroga (za izvoz) enak 4 cm, izračunamo radij polkroga za uvoz po obrazcu

$$r_u = r_I\sqrt{\frac{U}{I}} = 4.\sqrt{\frac{97,0}{142,2}} = 3^{\cdot}3 \text{ cm}$$

Iz grafikona sklepamo poleg strukture tudi na razmerje med uvozom in izvozom, kar je razviano iz različnih velikosti polkrogov. Razen tega je absolutni iznos vpisan sredi polkrogov. Struktura med izvozom in uvozom pa ni neposredno primerljiva.

6.10 Primerjava strukture izvoza s strukturo uvoza pa je veliko bolj neposredna v sliki 3.7. V tej sliki si mislimo,da je strukturni krog za izvoz (manjši krog) položen koncentrično na strukturni krog za uvoz (večji krog). Ker so izseki kroga za ustrezne skupine proizvodov blizu skupaj, je primerjava struktur veliko boljša. Da še izboljšamo primerjavo ustreznih krogovih izsekov, je krožnica manjšega kroga na mestih, kjer se stikata ustrezni skupini, nepoudarjena.



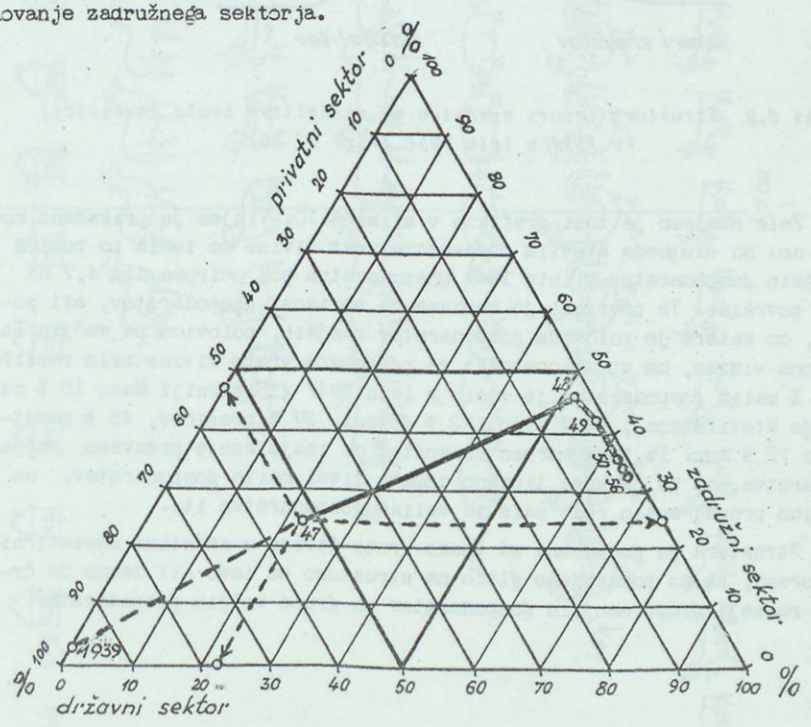
Slika 6.7 Struktura vrednosti izvoza in uvoza FLRJ v letu 1956 (Vir: Tabela 6.9)

6.11 Strukture v trikotniku. Ker je vsota vseh strukturnih odstotkov za vsako populacijo enaka 100 %, moremo strukture populacij, ki so razdeljene v tri dele, prikazati s točko v tako imenovani trikotniški mreži oziroma koordinatnem sistemu. Ta način je za strukture s tremi členi zelo prikladen. Več sorodnih strukturnih vrst po tri člene moremo v njem eno-

stavno prikazati s tolikotočkami, kolikor imamo strukturnih vrst.

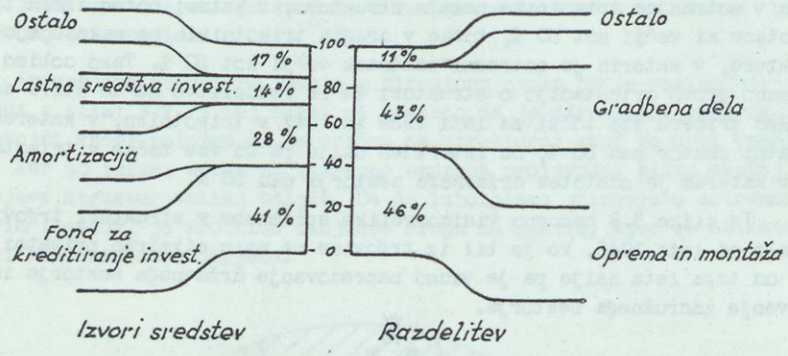
V sliki 3.8 je prikazana struktura števila trgovin po sektorju lastništva v razdobju 1947-1953 iz tabele 3.7. Iz grafikona vidimo, da je na vsaki stranici enakostraničnega trikotnika skala odstotkov za enega izmed sektorjev lastništva. Trikotnik je razdeljen z mrežo vzporednic k posameznim osem. V sliki 3.8 je za leto 1947 nakazano, kako odčitamo na posameznih skalah odstotke za posamezen sektor. Črte za 50 % so odebeljene. Tako je trikotnik razdeljen v štiri manjše enakostranične trikotnike. Točka v notranjem trikotniku nakaže strukturo, v kateri noben izmed treh odstotkov ni večji kot 50 %, točke v drugih trikotnikih pa nakazujejo strukture, v katerih je ustrezni odstotek večji kot 50 %. Tako dobimo neposredno grobo orientacijo o strukturi že iz razdelitve na te štiri dele. V našem primeru sta točki za leti 1939 in 1947 v trikotniku, v katerem je privatni sektor nad 50 %, od leta 1948 dalje pa so vse točke v trikotniku, v katerem je odstotek državnega sektorja nad 50 %.

Iz slike 3.8 nazorno vidimo velike spremembe v strukturi trgovine na malo od leta 1948, ko je bil iz trgovine na malo eliminiran privatni sektor, od tega leta dalje pa je vidno napredovanje državnega sektorja in nazadovanje zadružnega sektorja.



Slika 6.8 Struktura števila trgovin na malo po sektorju lastništva v FLRJ v razdobju 1947-1956

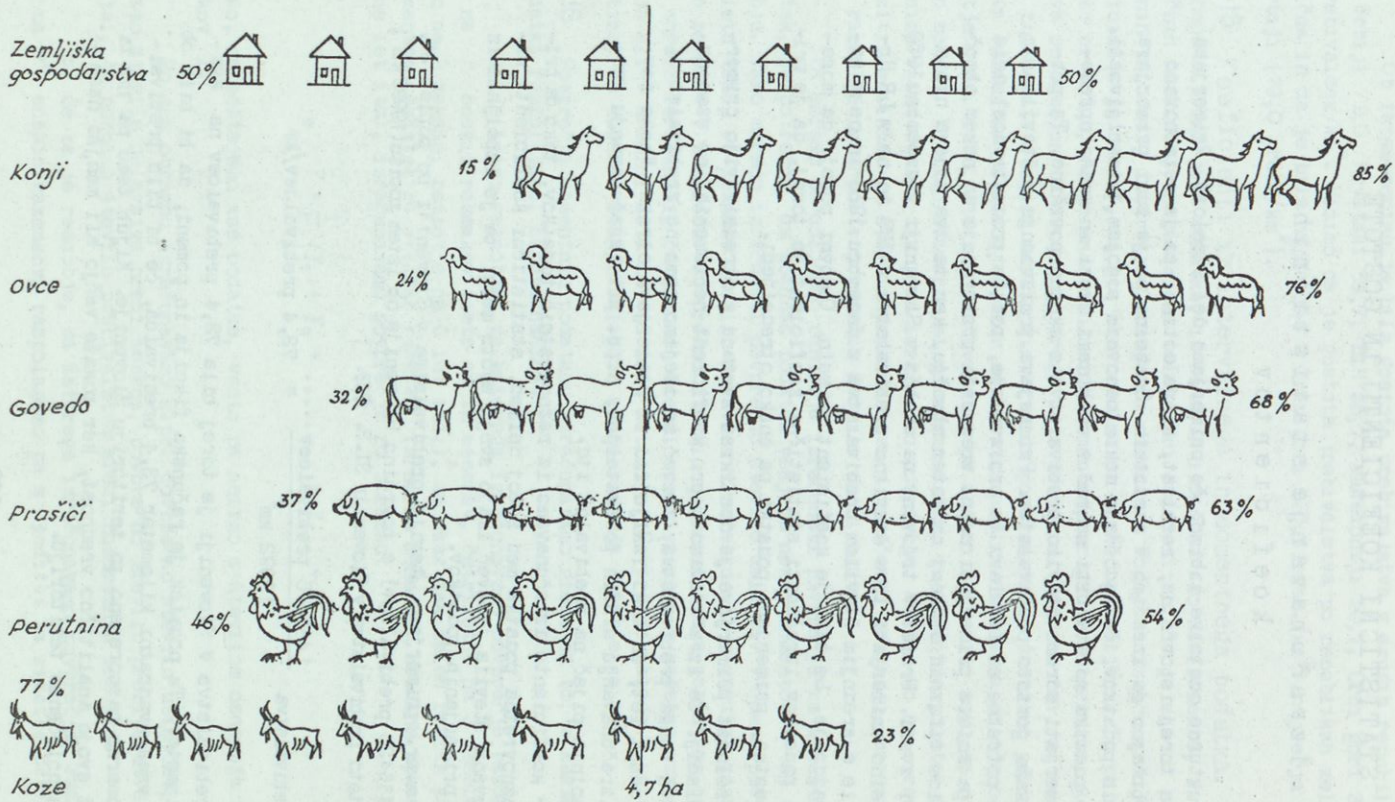
6.12 Druge metode prikazovanja struktur. Ker si moremo predstavljati bilančne podatke kot tok, ki na eni strani dohaja, na drugi pa odteka, strukturo bilanc nazorno prikažemo po metodi, ki je nakazana v sliki 3. 8. Grafikon ponazarja izvor in razdelitev bruto investicij v letu 1953 v FLRJ. Posamezni deli izvora in razdelitve so risani shematično kot kanali, kar vzbuja vtis o dohodku in odtoku.



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

6.13 Zelo nazoren je tudi grafikon v sliki 3.10. V njem je prikazano, kolikšen del od skupnega števila posameznih vrst živine so imela po popisu zemljiških gospodarstev v letu 1947 gospodarstva pod oziroma nad 4,7 ha skupne površine. Ta površina je mediana za velikost gospodarstev, ali površina, od katere je polovica gospodarstev manjših, polovica pa večjih. Iz grafikona vidimo, da so ti odstotki za posamezne vrste živine zelo različni. 50 % malih gospodarstev je imelo v letu 1947 v Sloveniji samo 15 % od skupnega števila konj, 24 % ovac, 32 % govedi, 37 % prašičev, 43 % perutnine in 77 % koz. Iz tega moremo sklepati, da imajo konje predvsem večja gospodarstva, da pa je koza tipično domača žival malih gospodarstev, da perutnino precej enako rede mala in velika gospodarstva itd.

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



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

STATISTIČNI KOEFICIENTI IN GOSTOTE

Izračunavanje statističnih koeficientov

6.14 Strukturne odstotke dobimo, če primerjamo del s celoto. Primerjana podatka sta torej istovrstna, rezultat, ki ga dobimo, pa je neimenovano število: Običajno ga izražamo v odstotkih. Smiselna pa je tudi primerjava raznovrstnih podatkov, če zadoščajo nekim osnovnim pogojem primerljivosti. Primerjana podatka morata biti namreč v vsebinski zvezi in enako opredeljena. Primerjati moremo število prebivalstva s skupno površino. Ta primerjava pokaže gostoto prebivalstva. Primerjava proizvodnje s številom ur, ki so bile potrebne za izdelavo te proizvodnje, pokaže produktivnost dela itd. Nima pa smisla primerjati n.pr. vrednosti proizvodnje na nekem območju s številom tifoznih obolenj na istem območju, ker ta dva pojava nista v vsebinski zvezi. Promet v trgovini na drobno v Sloveniji v decembru 1958 pa je smiselno primerjati le s številom prebivalstva v LRS in ne v LR Hrvatski in le s srednjim številom prebivalstva v decembru 1958 in ne za kako drugo razdobje. Le tako je koeficient smiseln. Osnovni pogoj, da moremo iz dveh podatkov izračunati statistični koeficient je torej, da je primerjava umestna, primerjani podatki pa enako opredeljeni.

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

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

Če n.pr. primerjamo število prebivalstva v Sloveniji po popisu 31. 3.1953 (1433425 prebivalcev) s površino Slovenije ob istem popisu (19992 km^2), dobimo gostoto prebivalstva v Sloveniji 31.3. 1953.

$$\text{gostota prebivalstva } \frac{1433425 \text{ prebivalcev}}{19992 \text{ km}^2} = 73{,}4 \text{ prebivalcev/km}^2$$

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

Ob istem popisu je bila gostota prebivalstva v drugih republikah: v Srbiji 79,0, Hrvatski 39,7, BiH 55,7, Makedoniji 50,7, Črni gori 30,4 prebivalcev/km^2. Vidimo da je gostota prebivalstva po republikah zelo različna in da je bila najmanjša v Črni gori (30,4 preb./km^2) največja pa v Srbiji (79,0 preb./km^2).

6.15 Koeficienti iz intervalnega in momentnega podatka. Ker izhajata primerjana podatka, iz katerih izračunavamo koeficiente, iz različnih populacij, naletimo na težavo pri opredelitvi, če je ena od primerjanih populacij intervalna, druga pa momentna. Pogoj za izračunavanje koeficienta je namreč, da sta obe populaciji enako opredeljeni. To pa na pogled ni možno, če je ena populacija momentna, druga pa intervalna, ker je prva opredeljena z momentom, druga pa z razmakom. Zato se zdi, da ne moremo primerjati število umrlih (intervalna populacija) s številom prebivalstva (momentna populacija), ker se število umrlih nanaša na določeno razdobje n.pr. leto, število prebivalstva pa moremo ugotoviti samo za določen moment. V takih primerih si pomagamo s tem, da momenten podatek spremenimo v intervalen tako, da izračunamo povprečje. Medtem ko ne moremo izraziti, kolikšno je število prebivalstva v določenem razdobju, moremo za to razdobje izračunati povprečno število prebivalstva.

Da moremo za določeno razdobje izračunati povprečje za momentni podatek, je potrebno, da poznamo podatek vsaj za nekaj momentov v tem razdobju. Tako moremo izračunati povprečno število delavcev v nekem podjetju v letu 1958, če poznamo število delavcev v začetku ali v sredini posameznih mesecev v tem letu. Povprečno število prebivalstva v petletnem razdobju moremo izračunati, če poznamo število prebivalstva v začetku vsakega leta ali v sredini leta za vsa leta petletja. Povprečno zalogo v letu pa dobimo, če poznamo stanje zalog ob koncu vsakega četrtletja itd.

6.16 Povprečja momentnih podatkov izračunavamo glede na to, s kakšnimi podatki razpolagamo, na dva načina.

a) Če imamo razdobje, za katerega izračunavamo povprečje, razdeljeno na r enakih razmakov (leto na 12 mesecev, četrtletje na 3 mesece, mesec na 3 dekade, petletje na 5 let itd.), za vsak tak delen razmak pa dan momenten *podatek za sredino razmika* (sredine mesecev, sredine dekad, sredine let itd), izračunavamo povprečje po obrazcu

$$\bar{Y} = \frac{1}{r}\,(Y_1 + Y_2 + \ldots\ldots + Y_{r-1} + Y_r) \qquad (3.5)$$

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

Po tem obrazcu izračunavamo n.pr. povprečno število delavcev, če poznamo število delavstva sredi meseca, povprečno število prebivalstva v petletki, če poznamo število prebivalstva sredi posameznih let itd.

b) Če pa je razdobje, za katerega izračunavamo povprečja, razdeljeno na r enakih razmakov, razpolagamo pa s *podatki za začetke oziroma*

konce teh razmikov, izračunamo povprečje po obrazcu

$$\bar{Y} = \frac{1}{r}\left(\frac{1}{2}Y_0 + Y_1 + Y_2 + \ldots + Y_{r-1} + \frac{1}{2}Y_r\right) \qquad (6.6)$$

tako, da vsoto polovičnih vrednosti podatkov na začetku in koncu celotnega intervala (Y_0 in Y_r) in celih vrednosti ostalih podatkov na začetku ali koncih delnih razmakov delimo s številom osnovnih razmakov r.

Če imamo n.pr. število delavstva na začetku oziroma na koncu posameznih mesecev, izračunati pa je treba povprečno število delavstva v letu, vzamemo polovico števila delavstva 1. januarja, število delavstva 1. februarja, 1. marca, 1. aprila itd. do 1. decembra in polovično število delavstva 31. decembra. Vsoto vseh teh vrednosti pa delimo z 12.

Če vzamemo, da je razdobje, za katerega hočemo izračunati povprečje, sestavljeno iz enega samega osnovnega intervala ($r = 1$), dobimo po obrazcu 3.5, da je $\bar{Y} = Y_1$. Iz tega sledi, da moremo povprečje nadomestiti s stanjem v sredini intervala. Tako n.pr. nadomeščamo povprečno število prebivalstva s „srednjim številom prebivalstva", to je s številom prebivalstva sredi leta (30. junija).

Če je $r = 1$ obrazec 3.3 degenerira v

$$\bar{Y} = \left(\frac{1}{2}y_0 + \frac{1}{2}y_1\right) = \frac{1}{2}(y_0 + y_1);$$

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

6.17 Ko imamo izračunano povprečje iz momentnih podatkov, izračunamo koeficient po obrazcu

$$K = \frac{X \cdot E}{\bar{Y} \cdot i} \qquad (6.7)$$

pri čemer pomeni: K = koeficient, X = intervalni podatek, $\bar{Y}$ = povprečje momentnega podatka, i = dolžina časovnega razdobja, za katerega izračunavamo koeficient, E = 100, 1000, 10000, odvisno od tega, na koliko enot momentnega podatka izračunavamo koeficient.

6.18 Ker je n.pr. koeficient natalitete število rojenih v enem letu na 1000 prebivalcev, izračunamo koeficiente natalitete za FLRJ v razdobju 1951-1955 takole:

Povprečno število prebivalstva v tem razdobju je $\bar{Y} = 17037000$, število rojenih $X = 2395813$, razdobje $i = 5$ let, koeficient pa je računan na $E = 1000$ prebivalcev. Iz teh podatkov dobimo:

$$K = \frac{2395813 \cdot 1000}{17067000 \cdot 5} = 28{,}1$$ rojstev na leto na 1000 prebivalcev.

Podobno izračunavamo vse druge koeficiente, ki imajo za osnovo primerjavo intervalnega podatka z momentnim.

Po tem obrazcu moremo izračunati koeficiente tudi za razdobja, ki so krajša od enotinega razdobja. Tako n.pr. koeficient mortalitete, ki je po definiciji število umrlih v letu na 1000 prebivalcev, za Slovenijo v januarju 1958 izračunamo takole: Ker je bilo srednje število prebivalstva v januarju 1958 $\bar{Y} = 1556578$, število umrlih v januarju 1958 $X = 1464$, $E = 1000$, $i = 31/335$, je koeficient mortalitete enak

$$K = \frac{1434 \cdot 1000}{1553578 \cdot 31/335} = 11{,}1$$ umrlih na leto na 1000 prebivalcev.

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

6.19 Koeficient obračanja zalog pove, kolikokrat se zaloga blaga ali surovin obrne v enoti časa. Izračunavamo ga tako, da primerjamo promet določenega blaga ali porabo surovin z ustreznimi povprečnimi zalogami.

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

Tabela 6.10 Mesečna zaloga in poraba kremenovega peska v podjetju A iz nekovinske stroke v prvem polletju 1958. (Vir: ZS LRS)

Mesec	Zaloga konec meseca v t	Poraba v t
1957 D	334	
1958 J	774	494
F	888	489
M	789	562
A	546	468
M	568	465
J	726	544

Izračunati je treba koeficient obračanja zalog kremenovega peska v prvem polletju 1958, izražen s številom obratov zalog v enem mesecu.

Po postopku za izračunavanje koeficientov moramo iz podatkov v tabeli 6.10 izračunati skupno porabo kremenovega peska v prvem polletju 1958 in povprečno zalogo v tem razdobju. Skupna poraba v polletju je vsota mesečne porabe za ustreznih šest mesecev.

$$X = 494 + 489 + 562 + 468 + 465 + 544 = 3\ 022\ t$$

Povprečno zalogo $\bar{Y}$ pa izračunamo po obrazcu 3.3, ker imamo podatke za konec mesecev.

$$\bar{Y} = \frac{\frac{684}{2} + 774 + 888 + 789 + 546 + 568 + \frac{726}{2}}{6} = 710\ t$$

Povprečna zaloga je torej 710 t.

Ker je koeficient merjen s številom obratov v enem mesecu, naš interval pa obsega pol leta, je $i = 3$. Iz teh podatkov izračunamo, da je koeficient obračanja zalog K kremenove moke v prvem polletju 1958 enak

$$K = \frac{X\ E}{\bar{Y}\ i} = \frac{3022.1}{710.6} = 0,71$$

Koeficient obračanja zalog $K = 0,71$. To pomeni, da zaloge kremenovega peska v podjetju A napravijo mesečno 0,71 obrata.

Z analognim računom smo dobili koeficiente obračanja zalog za isto podjetje za rjavi premog $K = 12,0$, pepelko $K = 0,50$, kalcinirano sodo $K = 2,78$. Kakor vidimo, so koeficienti obračanja zalog, ki so tudi operativno zelo pomemben pokazatelj, za posamezne surovine in goriva zelo različni.

Obračanje zalog pa moremo izraziti tudi z dolžino enega obrata. Ta koeficient, ki je recipročni pokazatelj zgornjega pokazatelja, izračunamo po obrazcu

$$K = \frac{\bar{Y} \cdot i}{X} \qquad (6.8)$$

Če dolžino enega obrata izrazimo v delovnih dneh, je za zgornji primer $i = 151$, koeficient obračanja zalog za kremenov pesek v podjetju A, izražen z dolžino enega obrata pa je enak

$$K = \frac{710 \cdot 151}{3022} = 35,4\ dni$$

Po navedenem postopku izračunavamo še niz drugih koeficientov iz ekonomskih statistik. Tako je n.pr. proizvodnja na enega delavca pokazatelj produktivnosti dela, proizvodnja določenih proizvodov na prebivalca, n.pr. jekla, pokazatelj industrializacije, poraba določenega proizvoda na prebivalca pokazatelj standarda, razmerje med prometom in zalogami pokazatelj obračanja zalog, število prodanih knjig na leto na 1000 prebivalcev pokazatelj kulturne ravni itd.

6.20 Kot primer kompleksne analize s koeficienti vzemimo trgovino na malo v letu 1957 po republikah. V tabeli 7.11 imamo osnovne podatke, iz katerih bomo izračunali koeficiente.

Tabela 6.11 Trgovina na malo v FLRJ v letu 1957 po republikah
(Vir: SG 58)

	FLRJ	Srbija	Hrvatska	Slovenija	BIH	Makedonija	Črna gora
Število prebivalstva 30.VI.1957 v 10^3	18284	7448	4109	1582	3168	1454	473
Število trgovin na malo:							
31.XII.1956	36257	14017	9296	4662	4634	2795	853
31.XII.1957	37651	14601	9798	4722	4822	2810	898
Število zaposlenih v trgovini na malo:							
31.XII.1956	112511	45954	27589	13689	14184	8523	2572
31.XII.1957	116119	46187	29292	14021	14566	9141	2912
Promet v letu 1957 v 10^6							
S k u p n o	750776	291478	198339	110259	90123	44331	16246
od tega:							
Prehrambeni proizvodi	232637	75604	64414	37519	32744	15348	7008
Industrijski proizvodi	483808	202410	125065	68917	51662	27133	8421

Iz tabele vidimo, da so prvi trije podatki momentni, drugi pa intervalni. Ker imamo srednje število prebivalstva, ta podatek vzamemo za povprečno število prebivalstva. Za število trgovin in število zaposlenih pa imamo stanje v začetku in na koncu leta 1957. Zaradi tega najprej izračunajmo povprečja, da bomo mogli število trgovin in število zaposlenih primerjati s prometom, ki je intervalnega značaja.

Za število trgovin v FLRJ je n.pr. iz zgornjih podatkov povprečno število trgovin enako polovici vsote števila trgovin po stanju 31.XII.1956 in 31.XII.1957.

$$\bar{Y} = \frac{1}{2}(36257 + 37651) = 36954$$

Podobno izračunamo vsa ostala povprečja. Tako dobimo vrsti za povprečno število trgovin in zaposleno osebje v trgovini na malo po republikah.

Tabela 6.12 Povprečno število trgovin in zaposlenega osebja v trgovini na malo v letu 1957 po republikah

Povprečno število	FLRJ	Srbija	Hrvatska	Slovenija	BIH	Makedonija	Črna gora
trgovin	36954	14309	9547	4692	4728	2802	876
zaposlenih	114315	43071	28440	13855	14375	8832	2742

Iz podatkov v tabeli 3.11 in 3.12 izračunamo niz koeficientov, ki osvetljujejo trgovino na drobno v FLRJ. Tako moremo s številom prebivalstva primerjati vse druge podatke. Število prebivalstva na eno trgovino meri gostoto trgovinske mreže, število prebivalstva na enega zaposlenega preskrbljenost s trgovskim kadrom itd. Ne da bi se spuščali v globljo analizo prometa na glavo prebivalstva, ki je bistveno odvisen razen od kupne moči tudi od drugih faktorjev, n.pr. od socialne strukture prebivalstva, merimo s temi koeficienti kupno moč prebivalstva.

Primerjava števila trgovin s številom zaposlenih ali s skupnim prometom prikaže velikost trgovinskih obratov. Primerjava skupnega števila zaposlenih s skupnim prometom pokaže produktivnost dela v trgovini na malo, razmerje med prometom z industrijskimi in prehrambenimi proizvodi pa da vpogled v strukturo prometa, ki je odvisna od socialnega sestava prebivalstva, življenjske ravni itd.

Vprašanje je, ali je možno vse te koeficiente izračunati iz zgornjih podatkov.

Primerjava števila trgovin in zaposlenega osebja s številom prebivalstva iz osnovnih podatkov ni možna, ker imamo število prebivalstva na dan 30. VI. 1957, podatke o številu trgovin in številu zaposlenih pa na dan 31. XII. 1953 in 31. XII. 1957. Primerjani momentni populaciji nista časovno enako opredeljeni. Ker pa moremo imeti srednje število prebivalstva za povprečje v letu 1957, smo upravičeni, da primerjamo srednje število prebivalstva s povprečnim številom trgovin in zaposlenih v letu 1957, ker se vsi ti podatki nanašajo na leto 1957. Število zaposlenih moremo primerjati s številom trgovin po stanju 31. XII. 1953 ali 31.XII.1957, ker imamo za te datume podatke o številu trgovin in številu zaposlenih. Moremo pa primerjati tudi povprečno število zaposlenih s povprečnim številom trgovin, ker veljata oba podatka za isto razdobje. V tabeli 3.13 so izračunani vsi navedeni koeficienti.

Tabela koeficientov v trgovini na malo ilustrativno pokaže razlike v trgovini na drobno po republikah. Najmanj prebivalcev odpade na eno trgovino v Sloveniji (337), največ pa v BIH (370). Enako velja za število prebivalstva, ki odpade na enega zaposlenega (Slovenija 114, BIH 221). Ti podatki kažejo gostoto trgovinske mreže. Tudi skupni promet na enega prebivalca kaže velike razlike med republikami (od 28,4 tisoč dinarjev za BIH

Tabela 6.13 Koeficienti trgovine na malo v FLRJ v letu 1957 po republikah (Vir; SG 58).

	FLRJ	Srbija	Hrvatska	Slovenija	FIH	Makedonija	Črna gora
Število prebivalcev na eno trgovino	493	520	431	387	370	519	539
Število prebivalcev na enega zaposlenega	159	162	144	114	221	165	172
Skupni promet na enega prebivalca v tisočih din	41,2	39,1	48,3	69,7	28,4	30,5	34,3
Promet s prehrambenimi proizvodi na prebivalca v tisočih din	12,8	10,2	15,7	23,7	10,3	10,5	14,8
Promet z industrijskimi proizvodi na prebivalca v tisočih din	26,5	27,1	30,4	43,6	16,3	18,6	17,8
Število zaposlenih na eno trgovino (povprečno)	3,09	3,22	2,98	2,95	3,04	3,15	3,11
Skupni promet na eno trgovino v milijon din	20,3	20,4	20,8	23,5	19,0	15,8	18,5
Skupni promet na enega zaposlenega v tisočih din	6570	6320	6970	7970	6230	5020	5930
Promet z industrijskimi proizvodi na sto din prometa s prehrambenimi proizvodi	207	267	194	184	158	176	120

do 69,7 za Slovenijo). Vendar ta koeficient nima polne analitične veljave, ker je njegova vrednost odvisna od velikega števila faktorjev (socialne strukture prebivalstva, kupne moči itd.). Promet s prehrambenimi artikli na prebivalca je precej odvisen od odstotka nekmečkega prebivalstva, promet z industrijskimi proizvodi na prebivalca pa moremo imeti bolj za pokazatelj kupne moči in standarda prebivalstva. Tudi ta koeficient kaže velike razlike med republikami. Medtem ko je za FIH, Črno goro in Makedonijo izredno nizek, je za druge republike, posebno za Slovenijo, znatno večji.

Velikost trgovinskih obratov, ki je razvidna iz povprečnega števila zaposlenih na eno trgovino, med republikami malo variira. Iz tega sklepamo, da je tip trgovin po republikah precej enoten. Enako težnjo kaže tudi skupen promet na eno trgovino in so znatnejše razlike edino le za Slovenijo in Makedonijo. Promet na enega zaposlenega, ki ga štejemo za pokazatelja produktivnosti dela v trgovini, kaže znatna odstopanja navzgor

za Slovenijo in znatno odstopanje navzdol za Makedonijo. Razmerje med prometom z industrijskimi proizvodi in prometom s prehrambenimi artikli kaže med republikami velike razlike. Največji koeficient dobimo za Srbijo (237 din za 100 dinarjev prometa s prehrambenimi artikli), najmanjši pa za Črno goro (120 din na 100 din prometa s prehrambenimi artikli). Razlike izvirajo iz različnega sestava kmečkega in nekmečkega prebivalstva in različne kupne moči prebivalstva. Zato koeficient nima polne analitične vrednosti.

Iz primera iz trgovinske statistike vidimo, kako so statistični koeficienti uporabni za kompleksno analizo statističnih podatkov. Podobno moremo analizirati podatke za najrazličnejša druga socialno-ekonomska področja.

6.21 Recipročni koeficienti. Medtem ko izračunavamo strukturne odstotke vedno tako, da del primerjamo s celoto, ne pa obratno, so statistični koeficienti smiselni tudi, če izračunamo recipročne vrednosti. Tako moremo pokazatelj gostote trgovinske mreže, ki je dan s številom prebivalstva na eno trgovino, izraziti tudi z recipročnim koeficientom številom trgovin na 1000 prebivalcev. Medtem ko pri prvotnem koeficientu primerjamo število prebivalstva s številom trgovin, z drugim primerjamo število trgovin s številom prebivalstva. Koeficient 493 prebivalca na eno trgovino v FLRJ moremo torej izraziti tudi z 2,15 trgovin na 1000 prebivalcev. Kateri izmed obeh možnih koeficientov, ki jih moremo izračunati v vsakem primeru, je boljši, je odvisno od namena izračunavanja in predstave koeficienta. V mnogo primerih sta oba koeficienta enako uporabljiva in vsak po svoje pokaže isto značilnost. Tako merimo produktivnost dela s količino, proizvedeno v enoti časa, ali s časom, v katerem je bila proizvedena enota proizvodnje. Obračanje zalog merimo s številom obratov zalog v enoti časa ali s časom, v katerem so se zaloge enkrat obrnile. Obremenjenost učiteljstva merimo s številom učencev na enega učitelja ali s številom učiteljev na 100 ali 1000 učencev itd.

Grafično prikazovanje statističnih koeficientov

6.22 Pravokotniki. Za koeficiente nimamo nekih posebnih metod grafičnega prikazovanja, kakor za strukture. Vrste koeficientov prikazujemo z običajnimi metodami grafičnega prikazovanja: s stolpci, linijskimi grafikoni, figurami itd. Vendar imamo nekaj specifičnih metod, ki jih moremo s pridom uporabiti ravno za prikazovanje koeficientov. Tako so za grafično prikazovanje koeficientov posebno primerni pravokotniki, ker imata stranici in ploščina pravokotnika podobno algebrajsko zvezo kakor koeficient in oba podatka, iz katerih je koeficient izračunan. Za pravokotnik velja $p = a.b$. Pri tem pomeni p ploščino a in b pa stranici, med K, X in Y pa velja podobna zveza $K = X/Y$ ali $Y = K.Y$. Stranici pravokotnika moremo uporabiti za prikaz koeficienta in enega absolutnega po-

datka, ploščina pravokotnika pa je zaradi lastnosti pravokotnika proporcionalna absolutni velikosti drugega podatka.

Uporabnost teh zvez bomo videli na praktičnem primeru. Za pet zapadnih držav imamo v tabeli 3.14 podatke o številu prebivalstva in številu osebnih in tovornih avtomobilov in avtobusov. Iz teh podatkov moremo za vsako državo posebej izračunati koeficiente o številu motornih vozil na 1000 prebivalcev - koeficiente motorizacije. Ti podatki so po zgornjem načinu prikazani v sliki 3.11.

Tabela 6.14 Število prebivalstva in število motornih vozil za pet zapadnih držav v letu 1955 (Vir: SG 58)

Država	Število prebivalstva v tisočih	Število motornih vozil v tisočih			Število motornih vozil na tisoč preb.		
		skupno	osebnih	kamionov avtobusov	skupno	osebnih	kamionov avtobusov
ZDA	165271	62020	51989	10031	375	315	60
Francija	43274	4820	3350	1470	111	77	34
Vel. Britanija	51215	4732	3550	1182	92	69	23
Zap. Nemčija	49995	2909	1813	1096	58	33	22
Italija	48016	1210	879	341	25	18	7

V grafikonu nazorno primerjamo tri vrste podatkov hkrati. Ker je najvažnejša meddržavna primerjava o številu motornih vozil na prebivalca, so pravokotniki postavljeni tako, da je primerjava teh podatkov najboljša. Razen tega je iz slike razvidno število prebivalstva in čeprav s ploščino tudi skupno število motornih vozil. Pravokotniki so razdeljeni še v proporcu s strukturo po vrsti motornega vozila.

Ta način moremo uporabiti za prikazovanje statističnih vrst za koeficiente iz najrazličnejših podatkov.

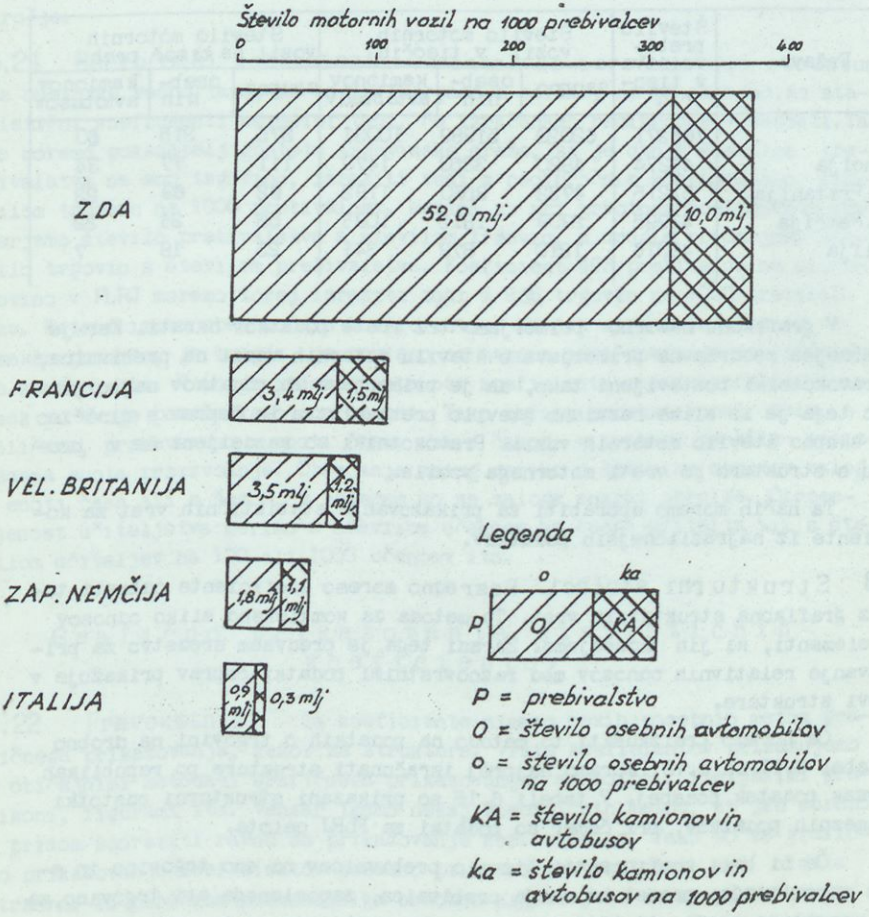
6.23 Strukturni stolpci. Posredno moremo koeficiente odbrati tudi iz grafikona strukturnih vrst. Ta metoda da kompleksno sliko odnosov med elementi, ki jih proučujemo. Zaradi tega je predvsem sredstvo za prikazovanje relativnih odnosov med raznovrstnimi podatki čeprav prikazuje v osnovi strukture.

Če hočemo preizkusiti to metodo na podatkih o trgovini na drobno iz tabel 6.11 in 6.12, moramo najprej izračunati strukture po republikah za vsak podatek posebej. V tabeli 6.15 so prikazani strukturni odstotki posameznih podatkov, pri čemer so podatki za FLRJ celota.

Če bi bili koeficienti: število prebivalcev na eno trgovino in enega zaposlenega, promet na enega prebivalca, zaposlenega ali trgovino, za vse republike enaki, bi bila struktura po republikah za vse podatke enaka. Razlike v strukturi izvirajo ravno iz razlik v odnosih. Če je za neko republiko promet na prebivalca sorazmerno majhen, sklepamo, da mora biti procent prometa od skupnega prometa za FLRJ za to republiko manjši kot

je ustrezni odstotek za prebivalstvo in obratno. Primerjana odstotka sta med seboj enaka le, če je za to republiko koeficient iz primerjanih podatkov enak povprečnemu koeficientu za FLRJ. To izvira iz značaja primerjave. Če z X_1 zaznamujemo prvi podatek za republiko, z X ustrezni skupni podatek za FLRJ, z Y_1 drugi podatek za republiko, z Y pa ustrezni skupni podatek za FLRJ, pomeni X_1/X strukturni delež za prvi podatek, Y_1/Y pa strukturni delež za drugi podatek. Če ju primerjamo, dobimo, da je

$$X_1/X : Y_1/Y = X_1/Y_1 : X/Y$$



Slika 6.11 Motorizacija petih zapadnih držav v letu 1955

Tabela 6.15 Strukture podatkov o trgovini na malo v FLRJ v letu 1947 po republikah

Pokazatelj	FLRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Prebivalstvo	100,0	40,8	22,5	8,7	17,4	8,0	2,6
Trgovine	100,0	38,7	25,8	12,7	12,8	7,3	2,4
Zaposleni	100,0	40,3	24,9	12,1	12,3	7,7	2,4
Promet skupni	100,0	38,8	23,4	14,7	12,0	5,9	2,2
Prehrambeni	100,0	32,5	27,7	13,1	14 1	3,3	3,0
Industrijski	100,0	41,9	25,9	14,2	10,7	5,3	1,7

Primerjava strukturnih deležev X_i/X z Y_i/Y ustreza primerjavi republiškega koeficienta X_i/Y_i z ustreznim koeficientom za FLRJ X/Y. Če sta primerjana strukturna odstotka med seboj zelo različna, pomeni, da je ustrezni koeficient med podatkoma zelo različen od koeficienta za celoto in obratno. Ker moremo strukturne odstotke raznovrstnih podatkov zlanka grafično prikazati, saj so neimenovana števila, razberemo vse te odnose kompleksno iz razlik v stolpcih.

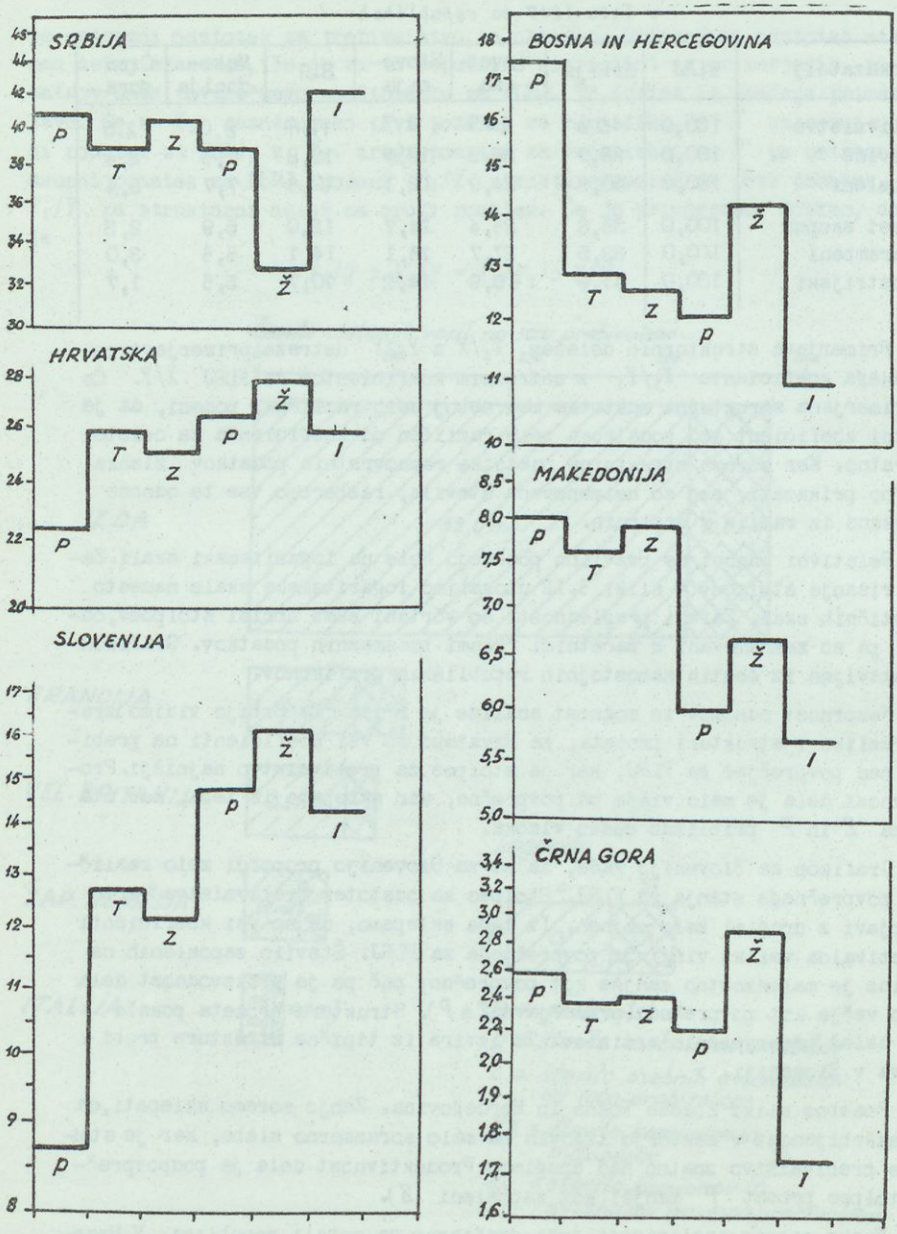
Relativni odnosi se pravilno pokažejo šele na logaritemski skali. Zato za risanje stolpcev v sliki 3.12 uporabimo logaritemske skale namesto aritmetičnih skal. Zaradi preglednosti so vrtani samo obrisi stolpcev, odstotki pa so zaznamovani z začetnimi črkami posameznih podatkov. Grafikon je sestavljen iz šestih samostojnih republiških grafikonov.

Nazornost odnosov in možnost analize je očita. Za Srbijo vidimo izrazito razliko v strukturi prometa, za Hrvatsko so vsi koeficienti na prebivalca nad povprečjem za FLRJ, ker je stolpec za prebivalstvo najnižji. Produktivnost dela je malo višja od povprečne, kar sklepamo iz tega, ker sta stolpca Z in P približno enako visoka.

Grafikon za Slovenijo kaže, da so za Slovenijo proporci zelo različni od povprečnega stanja za FLRJ. Stolpec za odstotek prebivalstva je v primerjavi z drugimi zelo majhen. Iz tega sklepamo, da so vsi koeficienti na prebivalca veliko višji od povprečnega za FLRJ. Število zaposlenih na trgovino je maloštevilno manjše kot povprečno, pač pa je produktivnost dela znatno večja kot povprečna (primerjava Z s P). Struktura prometa pokaže večji delež prehrambenih artiklov. To izvira iz tipične strukture prebivalstva v Sloveniji.

Posebno sliko pokaže Bosna in Hercegovina. Zanjo moremo sklepati, da je preskrbljenost v sektorju trgovin na malo sorazmerno slaba, ker je stolpec za prebivalstvo znatno nad drugimi. Produktivnost dela je podpovprečna (stolpec promet P manjši kot zaposleni Z).

Enako moremo analizirati tudi grafikone za ostali republiki. V Makedoniji število trgovin in zaposlenega osebja niti ni tako slabo v primerjavi s FLRJ, pač pa je promet sorazmerno majhen; to kaže, da je kupna moč majhna, produktivnost dela v trgovini pa nizka.



Slika 6.12 Struktura trgovine na malo v FLRJ v letu 1957 po republikah

Črna gora kaže precej podobno sliko o trgovinski mreži kot Srbija, ker so stolpci za prve štiri podatke (prebivalstvo P, število trgovin T, število zaposlenih Z in skupni promet P), zelo podobni. Pač pa je očita velika sprememba v strukturi prometa (promet s prehrambenimi artikli-$Ž$ visoko nad industrijskimi - I).

ENOSTAVNI INDEKSI

6.24 O indeksih govorimo, kadar z relativnimi števili primerjamo istovrstne, prirejene podatke. Strukture in koeficiente izračunavamo iz absolutnih podatkov. Indekse pa moremo izračunavati iz vseh vrst statističnih podatkov, ne pa samo iz absolutnih. Tako moremo z indeksi primerjati med seboj različne koeficiente, strukturne odstotke in druge izvedene pokazatelje.

Indeks vedno izračunavamo po osnovnem obrazcu

$$I_{1/0} = 100.Y_1/Y_0 \qquad (6.9)$$

Pri tem pomeni: Y_1 = podatek, ki ga primerjamo s podatkom Y_0. Y_0 = podatek, na katerega primerjamo. Podatek, na katerega primerjamo, imenujemo *bazo* ali *osnovo* indeksa. $I_{1/0}$ = indeks. Z indeksom izražamo tekoči podatek, v stotinkah od baze Y_0. Indeks pod 100 pomeni, da je pojav manjši od baze, indeks 100 pomeni, da sta primerjani podatek in baza, enaka. Indeks pa je nad 100, če je podatek večji od baze. Ker so indeksi neimenovana števila in imajo enoten merski sistem, iz njih zelo nazorno dobimo vtis o velikosti sprememb oziroma razlik. Razen tega moremo med seboj primerjati indekse raznovrstnih podatkov, ki drugače ne bi bili neposredno primerljivi.

6.25 Po podatkih mednarodnega pregleda v SG 1958 je imela Jugoslavija v letu 1955 narodnega dohodka 1298,3 milijard dinarjev, v letu 1956 pa 1444,1 milijard dinarjev. Italija je imela v letu 1955 10814 milijard lir, v letu 1956 pa 11504 milijard lir narodnega dohodka. Iz absolutnih podatkov si ne moremo ustvariti slike, v kateri državi je bilo povečanje narodnega dohodka večje. Podatki za Italijo in Jugoslavijo niso med seboj neposredno primerljivi, ker so eni dani v dinarjih, drugi pa v lirah. Če pa izračunamo indekse, dobimo, da je za Jugoslavijo indeks enak:

$$I_{56/55} = 100.Y_{56}/Y_{55} = \frac{100.1444,1}{1298,3} = 111,2$$

za Italijo pa

$$I_{56/55} = \frac{100.11504}{10814} = 106,4$$

Indeksa pa sta primerljiva in sicer kažeta, da je bil porast v na-

rodnem dohodku relativno v Jugoslaviji večji kakor v Italiji.

Če so spremembe, ki jih izražamo z indeksi, velike, indekse izračunavamo na cele, na eno decimalko pa jih računamo le, če prikazujemo majhne spremembe. Nikdar pa ne izračunavamo indeksov na več decimalk, ker tako indeks izgubi svojo osnovno kvaliteto - nazornost.

Stvarni in krajevni indeksi

6.26 . Ker se podatka, ki smo ju primerjali med seboj v zgornjem primeru, razlikujeta v času (leto 1955 in leto 1953), imenujemo ta indeks časovni indeks. Ker pa se moreta primerjana podatka razlikovati tudi v krajevnem ali stvarnem znaku, imamo tudi krajevne in stvarne indekse. Običajno izračunavamo indekse za cele vrste podatkov. V teh primerih vzamemo običajno za bazo vseh indeksov en in isti člen.

Kot primer *krajevnih indeksov* vzemimo koeficient razvezanih zakonov na 1000 sklenjenih zakonov v letu 1955 v glavnih mestih ljudskih republik.

Tabela 6.16 Število razvezanih zakonov na 1000 sklenjenih zakonov v letu 1955 v glavnih mestih ljudskih republik (Vir: Vitalna statistika 1955)

Mesto	Število razvezanih zakonov na 1000 sklenjenih zakonov	
	koeficient	indeks
Beograd	287	169
Zagreb	300	176
Ljubljana	170	100
Sarajevo	201	118
Skoplje	172	101
Titograd	115	68

Za bazo vzamemo člen, za katerega najbolje poznamo razmere.

Ker kot Slovenci najbolje poznamo razmere v Ljubljani, zato za bazo primerjave vzamemo Ljubljano in nanjo izračunamo vse indekse. Iz tabele 6.16 vidimo, da je problem razvez zakonov posebno pereč v Zagrebu in Beogradu. Tu je koeficient za 76 % oziroma za 65 % večji kot za Ljubljano. V primerjavi z Ljubljano ima majhen koeficient Titograd, (indeks 68).

Indekse smo izračunali po osnovnem obrazcu 6.9.

n.pr.: za Beograd 100.287/170 = 169.

6.27 Za primer indeksov iz *stvarnih vrst* vzemimo skupne povprečne mesečne dohodke za tekstilno, prehrambeno in grafično industrijo v FLRJ v

I.tromesečju 1958 po kvalifikacijah.

Tabela 6.17 Povprečni skupni mesečni prejemki delavcev po kvalifikacijah v I.tromesečju 1958 v FLRJ v dinarjih (Vir: SB 123)

Stroka industrije	Skupaj	Kvalifikacija			
		visoko kvalificirani	kvalificirani	priučeni	nekvalificirani
Tekstilna	10910	20070	11440	9480	9050
Prehrambena	9750	15040	11570	8620	7840
Grafična	12710	20620	13580	9160	8000

Primerljivejšo sliko teh podatkov dobimo, če vzamemo povprečne delavske prejemke za bazo in izračunamo indekse za vsako stroko. Ti indeksi so prikazani v tabeli 6.18.

Tabela 6.18 Indeksi skupnih mesečnih prejemkov delavcev po kvalifikaciji v I.tromesečju 1958 v FLRJ.(Baza: povprečna delavska plača = 100)

Stroka industrije	Skupaj	Kvalifikacija			
		visoko kvalificirani	kvalificirani	priučeni	nekvalificirani
Tekstilna	100	184	105	87	83
Prehrambena	100	154	118	88	80
Grafična	100	162	107	72	63

Indekse smo izračunali po osnovnem obrazcu za izračunavanje indeksov,

n.pr.: za visoko kvalificirane delavce v tekstilni stroki

$$I_{i/o} = 100.20070/10910 = 184$$

Indeksi dajo boljši pregled razlik v plačah po kvalifikacijah. Iz tabele vidimo, da imajo visoko kvalificirani delavci v tekstilni industriji povprečno za 84 % višjo plačo, kakor je povprečna delavska plača. V grafični stroki imajo nekvalificirani delavci samo 63 % od povprečne plače itd. Iz tabele indeksov vidimo, da so odnosi med plačami po kvalifikacijah po strokah zelo različni.

Časovni indeksi

6.28 Indeksi s stalno bazo. Največkrat izračunavamo indekse za časovne vrste, ker moremo z njimi zelo dobro proučevati dinamiko pojavov. Ča-

sovna vrsta osnovnih podatkov sicer že sama prikazuje dinamiko, vendar jo razmeroma težko sledimo, ker so posamezni pojavi različno veliki, imajo različne enote mere itd. Preračunanje take vrste v indeksno vrsto s pravilno izbrano bazo pa pokaže spremembe od baze, ki je v vsakem primeru 100, v stalnem merilu - stotinkah.

Za primer vzemimo proizvodnjo elektroenergije FLRJ in njenih sosedov v razdobju 1951-1957.

Tabela 6.19 Proizvodnja elektroenergije FLRJ in sosednih državah v razdobju 1951-1956 v milijon KWh (Vir: SG 58)

Država	Leto					
	1951	1952	1953	1954	1955	1956
Avstrija	7375	8032	8764	9847	10751	11718
Bolgarija	1015	1352	1557	1730	2073	2393
Grčija	745	943	1144	1173	1391	1593
Italija	29223	30843	32619	35574	38124	40611
Jugoslavija	2550	2700	2982	3440	4340	5084
Madžarska	3503	4197	4615	4824	5428	5194
Romunija	2478	2883	3411	3397	4340	4930

V tabeli 3.19 težko proučimo in primerjamo dinamiko proizvodnje elektroenergije med državami. Posamezne države imajo zelo različno proizvodnjo; to pa zameglí primerljivost dinamike. To hibo absolutnih podatkov odpravimo z indeksnimi vrstami. Če se odločimo, da vzamemo za bazo leto 1951, ker je začetno leto, dobimo naslednjo tabelo indeksov:

Tabela 6.20 Indeksi proizvodnje elektroenergije v FLRJ in sosednih državah v razdobju 1951-1956 (baza 1951 = 100)

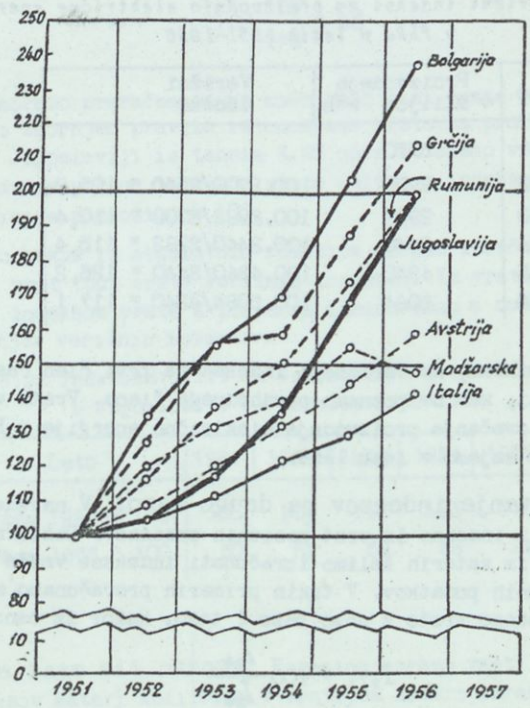
Država	Leto					
	1951	1952	1953	1954	1955	1956
Avstrija	100	109	119	133	143	159
Bolgarija	100	133	153	170	204	233
Grčija	100	127	154	158	187	214
Italija	100	106	112	122	130	139
Jugoslavija	100	106	117	135	170	199
Madžarska	100	120	131	137	155	148
Romunija	100	117	138	150	176	199

$I_{52/51}$ za Avstrijo izračunamo takole:

$$I_{52/51} = 100 \cdot Y_{52}/Y_{51} = \frac{100 \cdot 8032}{7375} = 109$$

Enako izračunamo vse druge indekse. Indeksne vrste veliko nazorneje pokažejo dinamiko proizvodnje električne energije v navedenih sedmih državah kakor vrste absolutnih podatkov. Ker je pri vseh časovnih vrstah vzeta ista baza, je možna tudi primerjava indeksov med državami. Tako moremo neposredno sklepati, v katerih državah je bil razvoj hitrejši in v katerih počasnejši. Take analize iz tabele absolutnih podatkov ne moremo napraviti neposredno. Grafikon indeksnih vrst v sliki 6.13 nazorno pokaže dinamiko za proizvodnjo električne energije za zgornjih sedem držav.

6.29 Indeksi s premično bazo. V zgornjem primeru smo za posamezno državo primerjali proizvodnjo vsakega leta s proizvodnjo istega leta 1951. Tako indeksno vrsto imenujemo indeksno vrsto s *stalno bazo*.



Slika 6.13 Indeksi proizvodnje električne energije za FLRJ in sosedne države v razdobju 1951-1956

Večkrat pa izračunavamo indekse tudi tako, da v isti časovni vrsti menjamo osnovo ali bazo primerjave. Take indeksne vrste imenujemo indeksne vrste s *premično bazo*. Izmed indeksnih vrst s premično bazo najpogosteje uporabljamo *verižne indekse*. Za dano časovno vrsto izračunavamo

vrsto verižnih indeksov tako, da za vsak podatek, za katerega izračunamo verižni indeks, vzamemo za bazo predhodni člen. Tako dobimo vrsto verižnih indeksov, ki pokažejo relativne spremembe od člena do člena. Verižne indekse torej izračunavamo po splošnem obrazcu

$$I_k = 100.Y_k/Y_{k-1} \qquad (6.10)$$

Pri tem pomeni: Y_k = tekoči podatek; Y_{k-1} = podatek predhodnega člena; I_k = verižni indeks.

Če vzamemo iz tabele 7.19 podatke o proizvodnji električne energije v Jugoslaviji v letih 1951-1953, je vrsta verižnih indeksov takale:

Tabela 6.21 Verižni indeksi za proizvodnjo električne energije v FLRJ v letih 1951-1956

Leto	Proizvodnja v milijon kWh	Verižni indeks
1951	2550	
1952	2700	100.2700/2550 = 105,9
1953	2982	100.2982/2700 = 110,4
1954	3440	100.3440/2982 = 115,4
1955	4340	100.4340/3440 = 126,2
1956	5084	100.5084/4340 = 117,1

Iz primera vidimo, da verižnega indeksa za prvi člen časovne vrste ne moremo izračunati, ker ne poznamo predhodnega člena. Vrsta verižnih indeksov kaže, da je zvečanje proizvodnje električne energije v FLRJ od leta do leta večje, z izjemo v letu 1953.

6.30 Preračunavanje indeksov na drugo bazo. V navedenih primerih smo izračunavali indekse iz vrst osnovnih podatkov. Dostikrat pa imamo indeksne vrste, iz katerih želimo izračunati indeksne vrste z novo bazo, nimamo pa osnovnih podatkov. V takih primerih preračunamo staro indeksno vrsto v indeksno vrsto z novo bazo 1 tako, kakor iz osnovne vrste

$$I_{2/1} = 100\frac{I_{2/0}}{I_{1/0}} \qquad (6.11)$$

Velja namreč:

$$100 \cdot \frac{I_{2/0}}{I_{1/0}} = 100 \cdot \frac{100.Y_2/Y_0}{100.Y_1/Y_0} = 100.Y_2/Y_1 = I_{2/1}$$

6.31 Vzemimo kot primer iz tabele 3.13 indeksno vrsto koeficientov razvezanih zakonov na 1000 sklenjenih zakonov. Baza te indeksne vrste je Ljubljana. Želimo pa izračunati novo indeksno vrsto, v kateri je baza mesto z najvišjim indeksom (Zagreb). Po obrazcu 6.11 indeks za vsako mesto

primerjamo z bazičnim indeksom (Zagreb 173)

Tabela 6.22 Indeksi koeficientov razveznih zakonov na 1000 sklenjenih zakonov za glavna mesta republik

Osnova	Mesto	Beograd	Zagreb	Ljubljana	Sarajevo	Skopje	Titograd
	Ljubljana=100	169	173	100	118	101	38
	Zagreb=100	96	100	57	67	57	39

za Beograd smo indeks z novo bazo izračunali:

$$100 \frac{169}{176} = 96$$

Enako moremo preračunati na novo bazo tudi vrste časovnih indeksov. Tako dobimo po zgornjem pravilu iz indeksne vrste za proizvodnjo električne energije v Jugoslaviji iz tabele 3.20 novo indeksno vrsto z bazo 1953 je 100, tako da vsak člen indeksne vrste delimo z indeksom za leto 1953 (199) in kvociente pomnožimo s 100.

Podobno kakor iz absolutnih podatkov moremo izračunavati iz časovnih indeksnih vrst tudi vrste verižnih indeksov. Iz vrste verižnih indeksov pa dobimo indeksne vrste s poljubno stalno bazo s postopnim množenjem oziroma deljenjem verižnih indeksov.

Tabela 6.23 Indeksna vrsta za proizvodnjo električne energije v Jugoslaviji v razdobju 1951-1956

Leto	1951	1952	1953	1954	1955	1956
Baza 1951 = 100	100	103	117	135	170	199
Baza 1956 = 100	50	53	59	68	85	100

6.32 Izbira baze ali osnove. Formalno moremo vzeti za bazo izračunavanja indeksov kateri koli člen v vrsti, za katero izračunavamo indekse, ali tudi vrednost izven nje. Vsebinsko pa je odločitev o bazi primerjave odvisna od namena, primernosti in smiselnosti primerjave. Zaradi tega je za izbiro baze nemogoče dati splošno pravilo, marveč samo nekaj splošnih načel, ki pomagajo pri pravilni izbiri baze.

Pri *stvarnih in geografskih indeksih* vzamemo za bazo po pravilu pojav oziroma območje, ki si ga najbolje predstavljamo oziroma ga najbolje poznamo. Tako v medrepubliški primerjavi vzamemo n.pr. za bazo Slovenijo, v medržavni primerjavi FLRJ itd. Če primerjamo z indeksi relativna števila ali druge izvedene pokazatelje za posamezne dele populacije, je

najprimerneje in najenostavneje, da vzamemo za bazo sumarni pokazatelj za celoto. Taki indeksi pokažejo odklone od nekega povprečnega stanja. Zaradi tega smo za vrsto indeksov mesečnih prejemkov po vrstah zaposlenih vzeli za bazo povprečne prejemke v stroki itd. Take vrste indeksov najobjektivneje pokažejo razlike po grupah.

Poseben problem je izbira baze pri *časovnih indeksih*. Za časovne indekse velja splošno pravilo, da vzamemo za bazo čas, ko je pojav normalen in ustaljen. Kdaj pa je pojav normalen, je težko določiti. Vendar moremo iz tega pravila vsaj zaključiti, katere čiene ne smemo vzeti za bazo, ker je laže ugotoviti, kdaj pojav ni normalen. Zaradi tega za primerjavo večine pojavov ne vzamemo vojna leta, leta gospodarskih kriz in tako dalje. Enako v zdravstveni statistiki ne vzamemo za bazo leta epidemij. Večja verjetnost je, da je pojav normalen v daljšem razdobju kot v krajšem. Zaradi tega običajno ne jemljemo za bazo kratka razdobja, n.pr. mesece pri indeksih proizvodnje, marveč vzamemo za bazo leto. Težko je tudi n.pr. reči v kmetijstvu, katero leto je normalno, ali je to leto ugodne ali neugodne vegetacije. Zato v kmetijstvu dostikrat vzamemo za bazo večletno (petletno ali desetletno) povprečje.

Povojni razvoj običajno primerjamo s stanjem pred vojno. Za bazo primerjave vzamemo čim kasnejše leto pred vojno, vendar tako, da nanj še ne vplivajo priprave na vojno. V FLRJ vzamemo za primerjavo s predvojnim stanjem za bazo leto 1939, OZN pa leto 1937. Vendar moramo paziti, kdaj so primerjave pred in povojnega stanja vsebinsko utemeljene in smiselne. Za zelo dolga razdobja in v primeru velikih sprememb, primerjava na določeno staro stanje nima smisla.

Če je bila nekaj let po drugi svetovni vojni primerjava s predvojnim stanjem upravičena, ker smo tako približno dobili vtis o pojavih in spremembah, je primerjava s predvojnim stanjem sedaj nepomembna, ker nas zanima povojni razvoj, ne pa primerjava s predvojnim stanjem, ki je že razmeroma zelo odmaknjeno. Vendar ne vzamemo za bazo povojnega razvoja prvo povojno leto, ki so bile razmere še neustaljene, temveč kasnejše razdobje normaliziranega stanja.

Včasih vzamemo za bazo indeksov tudi zadnje - tekoče leto. Taka indeksna vrsta primerja preteklo stanje s podatkom, ki je časovno najbližji in tudi najzanimivejši, ker predstavlja trenutno stanje.

Velikih sprememb tudi ne kaže prikazovati z indeksi. Tako nima smisla n.pr. izračunati indeksa proizvodnje valjanih aluminijevih proizvodov na bazo 1939, ko je bila proizvodnja 15 ton, s proizvodnjo v letu 1957, ko smo proizvedli 4527 ton valjanih aluminijevih proizvodov. Indeks, ki ga izračunamo iz teh podatkov, je 30180 in nenazoren. Kljub temu, da indeksi v splošnem izboljšajo primerjavo, ta indeks nima smisla, ker je nepredstavljiv.

Sedmo poglavje

FREKVENČNE DISTRIBUCIJE

SESTAVLJANJE FREKVENČNE DISTRIBUCIJE

7.1 Med statističnimi vrstami imajo posebno mesto vrste, ki prikazujejo razporeditev vrednosti za numerične znake. Take vrste imenujemo *frekvenčne distribucije* ali *pogostnostne razdelitve*. Za dano populacijo dobimo frekvenčno distribucijo, če za posamezne razrede znaka poiščemo, koliko enot ima vrednost znaka v ustreznem razredu. Tehnično dobimo frekvenčno distribucijo preprosto s črtkanjem ali odlaganjem listov.

Vzemimo za primer porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v takratnem okraju Novo mesto. Ta gospodarstva so bila anketirana v anketi o porabi lesa leta 1953. Osnovni podatki o porabi lesa v posameznih anketiranih gospodarstvih so navedeni v tabeli 7.1.

Tabela 7.1 Poraba lesa v m^3 v 149 kmetijskih gospodarstvih v letu 1953 v okraju Novo mesto
(Vir: Anketa o porabi lesa v letu 1953)

11,6	8,0	8,5	12,2	11,6	11,8	10,2	15,1	2,1	18,5	7,2	10,1	16,9	14,0
12,7	18,0	14,2	16,2	20,5	20,1	14,6	15,7	23,5	12,2	9,0	12,7	12,2	12,1
19,4	6,5	6,0	9,8	8,5	5,0	3,6	10,1	10,3	8,6	8,4	5,5	10,5	6,2
6,0	13,9	12,8	12,4	10,9	15,7	13,6	13,4	12,8	19,7	14,3	7,5	19,8	22,7
16,0	9,0	9,4	14,2	11,0	10,5	5,9	29,4	14,0	17,5	10,0	19,5	23,4	14,3
24,4	12,4	17,3	11,3	11,8	26,7	16,5	16,9	12,1	17,0	26,7	24,4	28,6	33,5
20,1	10,9	13,9	10,9	10,3	7,6	10,3	21,7	17,7	17,4	10,8	15,1	15,8	15,7
15,2	23,2	21,8	16,2	19,2	22,7	19,0	13,7	14,3	24,9	23,8	19,9	14,6	12,0
21,1	31,6	29,9	26,5	6,7	8,2	6,2	4,9	7,9	9,4	9,3	16,2	19,3	18,9
12,9	32,4	18,3	15,2	11,3	8,3	4,7	15,3	13,3	17,3	27,5	15,1	16,1	13,2
18,4	14,8	17,1	10,5	8,4	7,5	12,2	9,7	18,1					

Uredimo zgornjo nepregledno množico podatkov v frekvenčno distribucijo, v kateri vzemimo po 1 m^3 široke razrede! V tabeli 7.2 pomeni n.pr. 8 porabo od 8,0 m^3 do 8,9 m^3.

Število enot v posameznem razredu imenujemo *frekvenca*. Po pravilu frekvence vedno zaznamujemo s črko *f*.

Frekvenčna distribucija v tabeli 7.2 kaže veliko preglednejše porabo lesa v gospodarstvih kakor individualni podatki v tabeli 7.1. Vendar

frekvenčna distribucija ne da natančnih vrednosti. Iz nje n.pr. vidimo, da je eno gospodarstvo porabilo lesa od 2,0 do 2,9 m^3, ne vemo pa, koliko so natančno porabili lesa v tem gospodarstvu. Enako vidimo, da je 14 gospodarstev porabilo 10,0 do 10,9 m^3, ne vemo pa natančne porabe lesa. S frekvenčno distribucijo se je preglednost povečala, natančnost informacij pa zmanjšala.

Tabela 7.2 Frekvenčna distribucija 149 kmetijskih gospodarstev v Novem mestu po porabi lesa v letu 1953

Poraba lesa m^3	x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Število gospodarstev	f	-	-	1	1	2	3	6	5	8	7	14	7	14	7	10

x	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
f	10	8	7	6	8	3	3	2	4	3	-	3	1	1	2	-	1	1	1

Iz frekvenčne distribucije dobimo nekatere zelo koristne ugotovitve. Predvsem je frekvenčna distribucija slika o variiranju znaka. Iz tabele 7.2 vidimo, da so n.pr. porabili v kmetijskih gospodarstvih od 2 m^3 do 34 m^3 lesa. Razen tega si iz frekvenčne distribucije ustvarimo sliko o porabi lesa znotraj teh meja. Bežen pogled pokaže, da je gospodarstev z majhno porabo malo, medtem ko se število gospodarstev v posameznih razredih do neke porabe v glavnem veča, od tu naprej pa so frekvence vedno manjše. Frekvenčna distribucija nakazuje zakonitost, ki jo opazimo pri večini populacij. V dani populaciji se vrednosti najbolj goste okrog nekega središča, od katerega so odkloni tem redkejši, čim večji so. Vendar se ta zakonitost v frekvenčni distribuciji v tabeli 7.2 ne kaže popolnoma. Imamo veliko odklonov od tega pravila. Frekvence v sosednih razredih so včasih večje, včasih manjše. Število enot v posameznih razredih je namreč razmeroma majhno, zato prevladujejo še slučajni vplivi. Če bi bilo število enot v posameznih razredih večje, slučajni vplivi ne bi bili tako močni in bi se zgornja zakonitost izražala močneje. Frekvence v razredih moremo povečati na dva načina. Če bi namesto 149 gospodarstev imeli desetkrat večjo populacijo s 1490 gospodarstvi, bi bile tudi frekvence večje. Zakonitost bi se izražala močneje. Vendar tako večanje frekvenc običajno ne moremo uporabiti. Frekvence pa povečamo tudi, če vzamemo širše razrede. S tem se sicer zmanjša natančnost prikaza, zakonitost, po kateri variira vrednost, pa je prikazana bolje, ker so slučajni vplivi manjši.

Če v tabeli 7.2 združimo po štiri razrede, dobimo novo frekvenčno distribucijo; ta je prikazana v tabeli 7.3.

Frekvenčna distribucija v tabeli 7.3 je preglednejša, ker je število razredov manjše. Razen tega pa je zaradi razmeroma velikih frekvenc dobro vidna zakonitost gostitve. Frekvence se sprva večajo, dokler ne dosežejo

največjo gostitev v razredu 12,0 - 15,9 m^3, od tu pa stalno padajo.

Tabela 7.3 Frekvenčna distribucija za porabo lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m^3	f
0,0 - 3,9	2
4,0 - 7,9	16
8,0 - 11,9	36
12,0 - 15,9	41
16,0 - 19,9	29
20,0 - 23,9	12
24,0 - 27,9	7
28,0 - 31,7	4
32,0 - 35,9	2
	149 = N

Razreda pa ne smejo biti razredi preširoki. S širjenjem razredov se natančnost bolj in bolj manjša. Tudi zakonitost gostitve je vedno manj vidna. Če v zgornji frekvenčni distribuciji združimo po dva razreda, dobimo skrčeno distribucijo, ki je prikazana v tabeli 7.4.

Tabela 7.4 Frekvenčna distribucija porabe lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

Poraba v m^3	f
0,0 - 7,9	18
8,0 - 15,9	77
16,0 - 23,9	41
24,0 - 31,9	11
32,0 - 39,9	2
	149 = N

Ta frekvenčna distribucija sicer še vedno nakazuje osnovno zakonitost, vendar daje pregrobo sliko, ker so v istem razredu gospodarstva, ki imajo do 8 m^3 razlik v porabi lesa.

Zato je pri sestavljanju frekvenčnih distribucij važno število in velikost razredov. Če vzamemo premajhne razrede, na frekvence preveč vplivajo slučajni vplivi in zabrišejo osnovno zakonitost gostitve, preveliki razredi pa dajo pregrobo sliko. Nimamo splošnega pravila, kakšne in koliko razredov naj ima frekvenčna distribucija; pač pa je praksa pokazala tole: razredi, ki razdele razmak med najmanjšo in največjo vrednostjo v populaciji na osem do šestnajst razredov, dajo običajno dosti dobro sli-

ko gostitve. Število razredov se ravna po velikosti populacije. Za manjše populacije vzamemo manj, za večje pa več razredov. V našem primeru smo resnično dobili s štiridesetimi razredi nepregledno, s petimi razredi pregrobo sliko o variranju znaka, medtem ko je frekvenčna distribucija z desetimi razredi pokazala vse značilnosti gostitve.

Frekvenčne distribucije z neenakimi razredi

7.2 Čeprav je iz tehničnih razlogov najbolje, da so vsi razredi v frekvenčni distribuciji enako široki, včasih iz vsebinskih razlogov sestavljamo frekvenčne distribucije, v katerih so razredi različno široki. Vzemimo za primer industrijska podjetja! Če ima eno 10, drugo pa 50 zaposlenih, je med njima večja vsebinska razlika kakor pa med podjetjema s 1000 in 1040 delavci, čeprav je v obeh primerih razlika 40 delavcev. Zato industrijska podjetja po številu delavstva prikazujemo v frekvenčni distribuciji z različnimi širinami razredov.

Tabela 7.5 Industrijska podjetja v FLRJ ob koncu leta 1957 po številu zaposlenih (Vir: SG 58)

Število zaposlenih	f
- 15	155
16 - 30	218
31 - 60	350
61 - 125	510
126 - 250	471
251 - 500	375
501 - 1000	276
1000 -	170
	2525 = N

Iz tabele 7.5 vidimo, da se meje razredov vrste v približni geometrijski postopnici; to pa je v skladu z relativno primerljivostjo števila delavstva po podjetjih.

7.3 V frekvenčni distribuciji z enakimi razredi se frekvence spreminjajo samo zaradi različne stopnje gostitve v posameznih razredih. V frekvenčni distribuciji z neenakimi razredi pa je frekvenca v danem razredu odvisna razen od gostitve tudi od širine razreda. Pri isti stopnji gostitve imajo širši razredi večjo, ožji razredi pa manjšo frekvenco. Da iz frekvence odstranimo vpliv različne širine in prikažemo samo stopnjo gostitve, izračunavamo po obrazcu

$$g_k = f_k / i_k \qquad (7.1)$$

gostoto frekvence g_k: ta pokaže, koliko frekvence v razredu odpade na enotin razmak.

Za vsako frekvenčno distribucijo moremo izračunati strukturne deleže; ti pokažejo, koliki del celotne populacije je v posameznem razredu. Strukturne deleže, ki jih dobimo, če frekvenco f_k delimo z obsegom populacije N, imenujemo *relativne frekvence* f_k^o.

$$f_k^o = f_k/N \tag{7.2}$$

Kakor smo iz frekvenc izračunali gostoto frekvence, tako moremo po obrazcu

$$\varphi_k = f_k^o/i_k = f_k/N \cdot i_k \tag{7.3}$$

izračunati tudi *gostoto relativne frekvence* φ_k, če relativno frekvenco delimo s širino razreda.

Iz obrazca 7.3 zlahka dobimo, da je

$$f_k = N \cdot i_k \cdot \varphi_k \tag{7.4}$$

Frekvenca v danem razredu je torej premo sorazmerna z obsegom populacije N, s širino razreda i_k in gostoto relativne frekvence φ_k, ki je povezana z vsebino pojava.

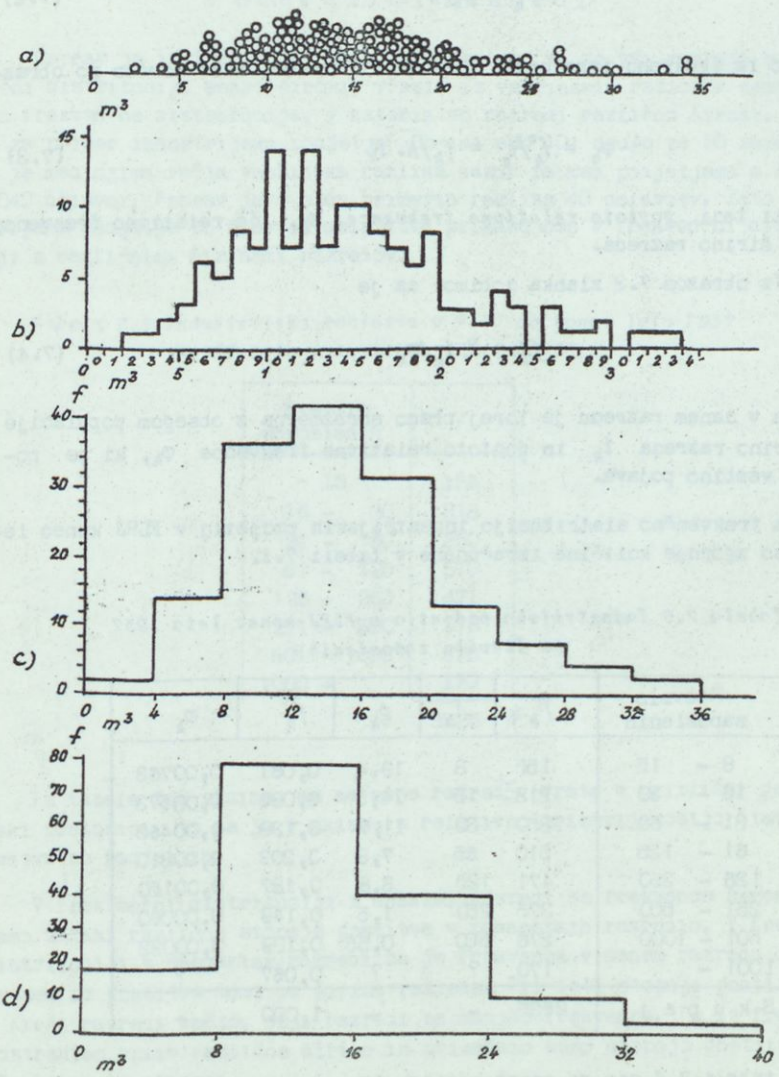
7.4 Za frekvenčno distribucijo industrijskih podjetij v FLRJ konec leta 1957 so zgornje količine izračunane v tabeli 7.3.

Tabela 7.6 Industrijska podjetja v FLRJ konec leta 1957 po številu zaposlenih

Število zaposlenih	f_k	i_k	g_k	f_k^o	φ_k
8 – 15	155	8	19,4	0,061	0,00763
16 – 30	218	15	14,5	0,086	0,00573
31 – 60	350	30	11,7	0,139	0,00463
61 – 125	510	65	7,8	0,202	0,00310
126 – 250	471	125	3,8	0,187	0,00150
251 – 500	375	250	1,5	0,149	0,00060
501 – 1000	276	500	0,55	0,109	0,00022
1001 –	170	?	?	0,067	?
S k u p a j	2525	–	–	1,000	–

V tabeli 7.6 smo za odprt razred -15 zaposlenih vzeli spodnjo mejo 8 zaposlenih, ker predpostavljamo, da ni industrijskih podjetij, ki imajo

manj kot 8 zaposlenih. Zadnji razred je odprt navzgor in zanj ne poznamo širine razreda. Zato zanj nismo mogli izračunati ustreznih gostot. Absolutnih frekvenc med seboj ni mogoče primerjati. Pač pa je smiselna primerjava gostot frekvenc ali gostot relativnih frekvenc. Za naš primer ti vrsti kažeta zakonitost, kako stalno pada gostota frekvenc. Ta zakonitost je značilna tudi za nekatere druge socialno-ekonomske pojave.



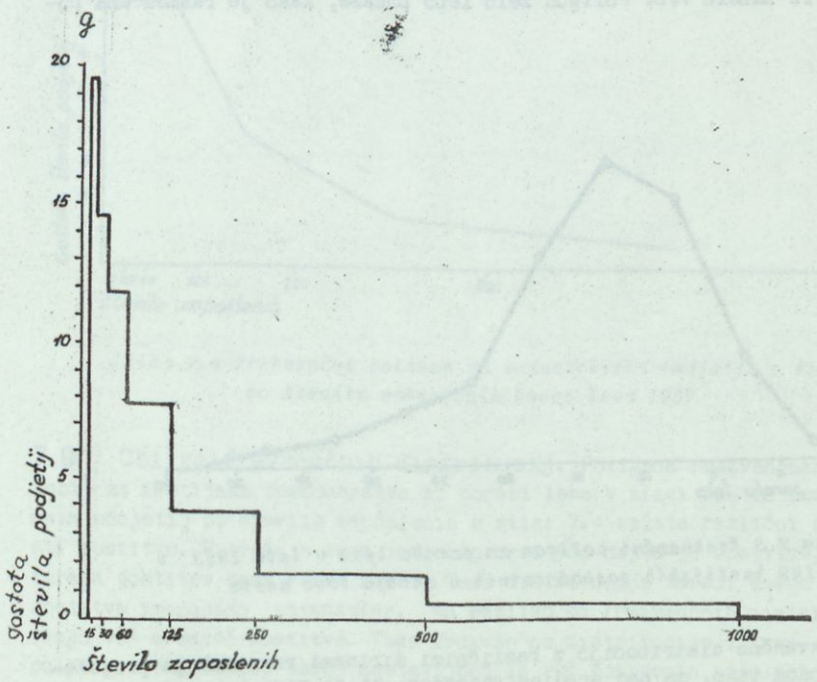
Slika 7.1 Histogrami porabe lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

Grafično prikazovanje frekvenčnih distribucij

7.5 Histogram. Frekvenčne distribucije grafično prikazujemo na dva načina: s *histogrami* in *poligoni*.

Frekvenčne distribucije z enakimi razredi v *histogramu* prikažemo z enako širokimi stolpci, katerih višina je v sorazmerju z ustreznimi frekvencami. Površina vseh stolpcev je zato sorazmerna z obsegom populacije. Običajno ne črtamo celih stolpcev, temveč rišemo samo obrise. V sliki 7.1 je na skali za porabo lesa prikazana najprej razmestitev individualnih vrednosti porabe lesa za 149 kmetijskih gospodarstev v Novem mestu iz tabele 7.1, nato pa vse tri frekvenčne distribucije iz tabel 7.2, 7.3, 7.4. V prvi je širina razreda 1 m^3, v drugi 4 m^3, v tretji pa 8 m^3. Iz histogramov še nazorneje kot iz tabel vidimo, da najbolj ustreza širina razreda 4 m^3.

7.6 Če prikazujemo s histogramom frekvenčne distribucije z različnimi širinami razredov, rišemo frekvence v razredih s pravokotniki, v katerih so širine sorazmerne s širinami razredov i_k, višine pa sorazmerne z gostoto frekvence g_k. Ker je po obrazcu 7.1 frekvenca zmnožek širine raz-



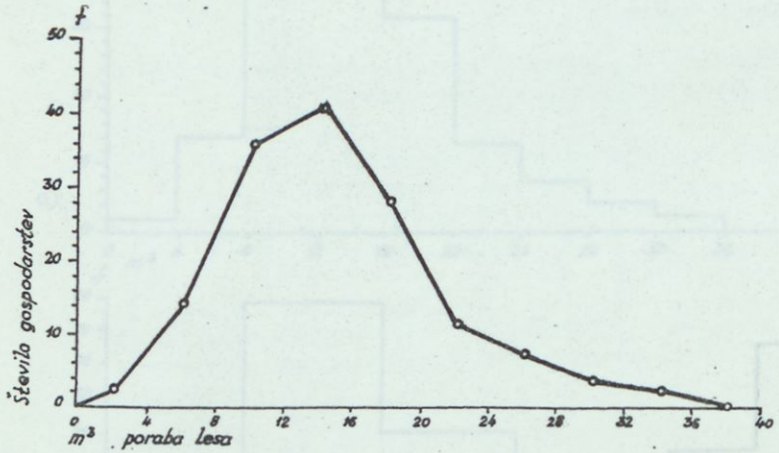
Slika 7.2 Histogram industrijskih podjetij po številu zaposlenih v FLRJ konec leta 1957

reda in gostote frekvence, so ploščine stolpcev v sorazmerju z ustrezno frekvenco f_k, ploščina vseh stolpcev pa v sorazmerju z obsegom populacije N.

Edino tako dobimo pravilno sliko o razmestitvi pojava. V sliki 7.2 je s histogramom prikazana frekvenčna distribucija industrijskih podjetij v FLRJ konec leta 1957 po številu delavcev iz tabele 7.3. Histogram pravilno pokaže, da gostitev podjetij pada, če se število podjetij veča.

7.7 Poligon. Frekvenčni poligon za vrste z enakimi razredi narišemo tako, da nad sredine razmakov za posamezne razrede nanesemo točke, ki so od abscisne osi oddaljene v sorazmerju z velikostjo frekvence. Dobljene točke po vrsti zvežemo z daljicami. Če vzamemo, da je frekvenca v razredu pred prvim razredom in po zadnjem razredu enaka nič in da ustrezni točki ležita na abscisni osi, je poligon zaključen. Poligon bolje prikazuje stvarno razmestitev vrednosti, ker z njim nakažemo verjetnejšo razmestitev vrednosti znotraj razredov, kot s histogramom, ki nakazuje enakomerno razmestitev znotraj razredov. Logično je namreč, da je znotraj razredov na straneh, ki teže proti maksimalni gostitvi, gostota večja kakor na zunanjih straneh razredov.

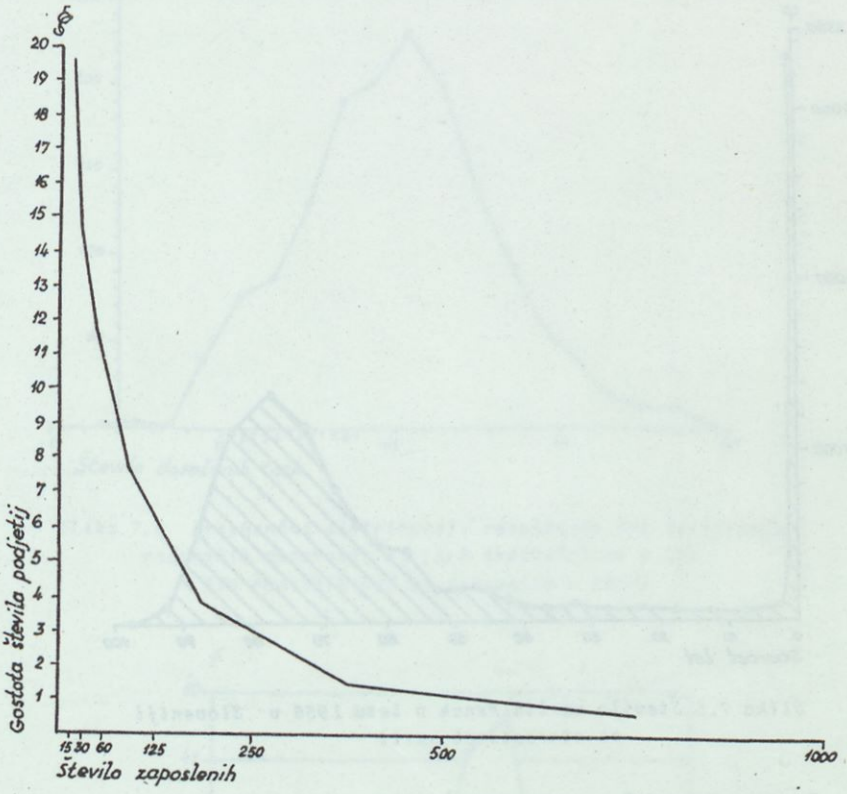
V sliki 7.3 je s poligonom prikazana frekvenčna distribucija o porabi lesa iz tabele 7.3. Poligon zelo lepo pokaže, kako je razmeščena poraba lesa.



Slika 7.3 Frekvenčni poligon za porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

7.8 Frekvenčno distribucijo z različnimi širinami razredov pa prikažemo s poligonom tako, da nad sredine razredov, ki so risani v sorazmerju s širinami razredov, nanesemo točke, ki so od abscise oddaljene v sorazmerju z gostoto frekvence. Če točke med seboj povežemo, dobimo lomljeno črto,

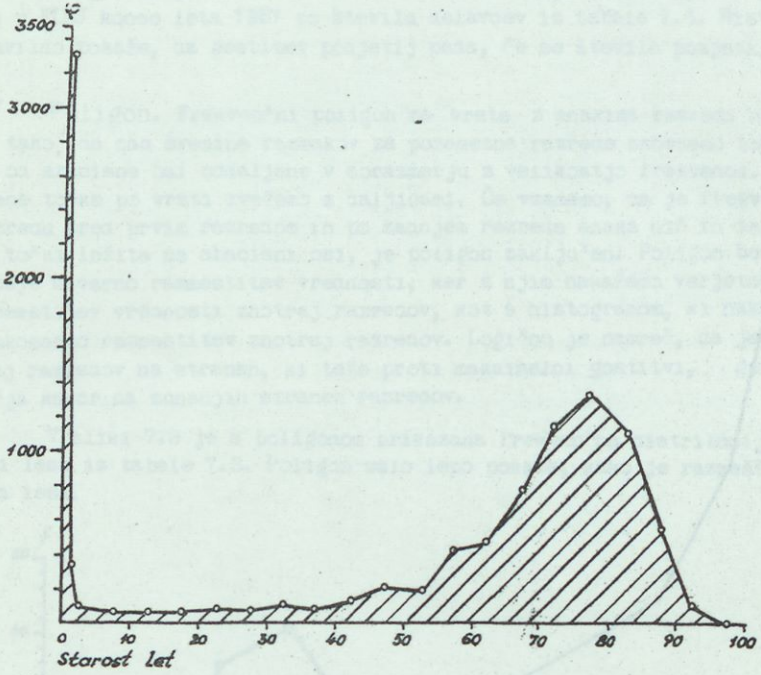
ki zelo dobro ponazarja razmestitev vrednosti. Frekvenčna distribucija industrijskih podjetij v FLRJ po številu zaposlenih je s poligonom narisana v sliki 7.4. Poligon še nazorneje kakor histogram kaže, kako upada gostitev podjetij, če se število zaposlenih veča.



Slika 7.4 Frekvenčni poligon za industrijska podjetja v FLRJ po številu zaposlenih konec leta 1957

7.9 Oblike frekvenčnih distribucij. Poligona frekvenčnih distribucij za kmetijska gospodarstva po porabi lesa v sliki 7.3 in industrijskih podjetij po številu zaposlenih v sliki 7.4 kažeta različni zakonitosti gostitve. Potrošnja lesa ima neko središče največje gostitve, od katerega gostitev pada v obe smeri. Take distribucije zaradi enega središča gostitve imenujemo *unimodalne*, za razliko od frekvenčnih distribucij, ki imajo več središč gostitve. Take frekvenčne distribucije sodimo, če je populacija nehomogena, torej sestavljena iz več homogenih populacij. Te vrste distribucij imenujemo *bimodalne*, če imajo dva vrha, in *polimodalne*, če imajo več vrhov. Z dvema vrhovoma je v sliki 7.5 prikazana frekvenčna distribucija umrlih žensk po starosti v letu 1953 v Sloveniji. V grafiko-

nu opazimo izredno veliko število umrlih otrok do enega leta. Število umrlih se zaradi majhne umrljivosti v naslednjih letih znatno zmanjša. V starejših letih se zaradi povečane umrljivosti zelo poveča in doseže največje število umrlih v starosti 75–79 let. Od te starosti dalje zopet pada zaradi vedno manjšega števila prebivalstva v starosti nad 80 let.

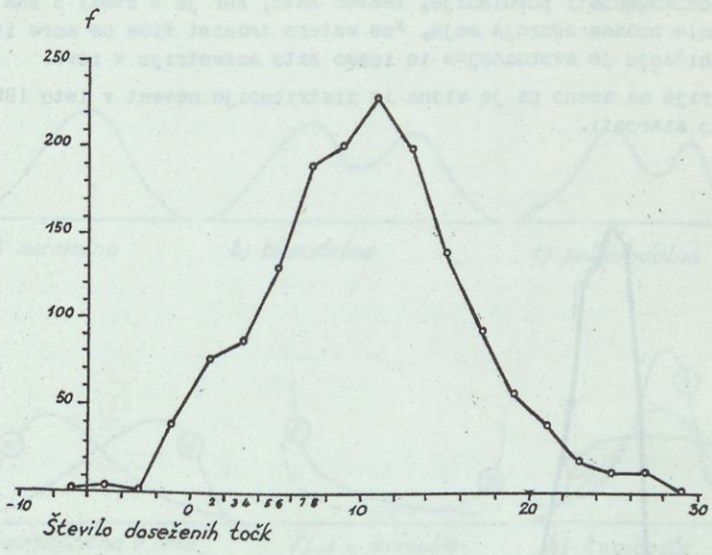


Slika 7.5 Število umrlih žensk v letu 1956 v Sloveniji po starosti ob smrti

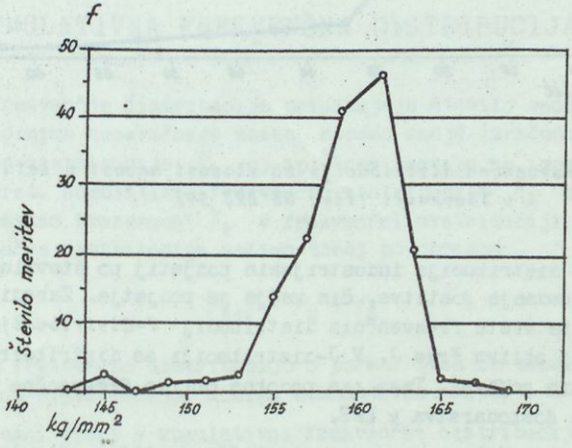
7.10 Poligon frekvenčne distribucije porabe lesa v kmetijskih gospodarstvih kaže bolj ali manj enako upadanje gostitve od središča gostitve. Frekvenčne distribucije, ki kažejo enako upadanje frekvenc na levo in desno od središča gostitve, imenujemo *simetrične distribucije*, za razliko od *asimetričnih distribucij*, pri katerih je upadanje gostitve na eno ali drugo stran hitrejše. Če je upadanje gostitve počasnejše na levo od središča gostitve, pravimo, da je frekvenčna distribucija *asimetrična v levo*, če pa je upadanje gostitve počasnejše desno od središča, govorimo o *asimetriji v desno*. Simetrično distribucijo dobimo navadno za homogene populacije.

Če je populacija popolnoma homogena in so razlike rezultat samo slučajnih vplivov, variiranje pa ni omejeno v nobeno stran, se vrednosti distribuirajo v frekvenčni distribuciji, ki je simetrična in *zvonaste* oblike. To distribucijo imenujemo *normalno distribucijo*. Normalna distribucija je narisana v sliki 7.9a.

Normalni distribucija je precej podobna frekvenčna distribucija rezultatov testiranja računskih zmožnosti za 1518 tretješolcev v LRS v letu 1957.



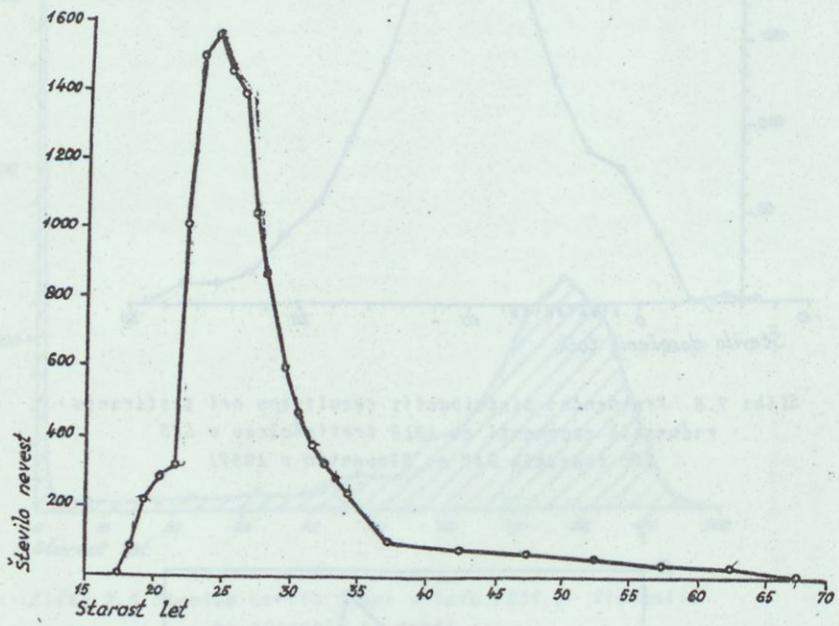
Slika 7.6 Frekvenčna distribucija rezultatov pri testiranju računskih zmoznosti za 1518 tretješolcev v LRS (Po podatkih DAF za Slovenijo v 1957)



Slika 7.7 Frekvenčna distribucija o trdnosti za patentirano 5-mm jeseniško žico v kg/mm²
(Po podatkih Zavoda za raziskavo materiala in konstrukcij LRS)

Asimetrijo povzročajo različni vzroki. Vir asimetrije more biti heterogenost populacije ali pa tudi naravna omejitev variiranja. Tako imamo v sliki 7.7 narisano frekvenčno distribucijo za 158 meritev trdnosti 5-mm jeklene patentirane žice. Ta distribucija ne kaže tendence asimetrije v levo zaradi nehomogenosti populacije, temveč zato, ker je v zvezi z lastnostmi materiala podana zgornja meja, čez katero trdnost žice ne more iti. Variacija k zniževanju je svobodnejša in imamo zato asimetrijo v levo.

Asimetrija na desno pa je vidna iz distribucije nevest v letu 1956 v Sloveniji po starosti.

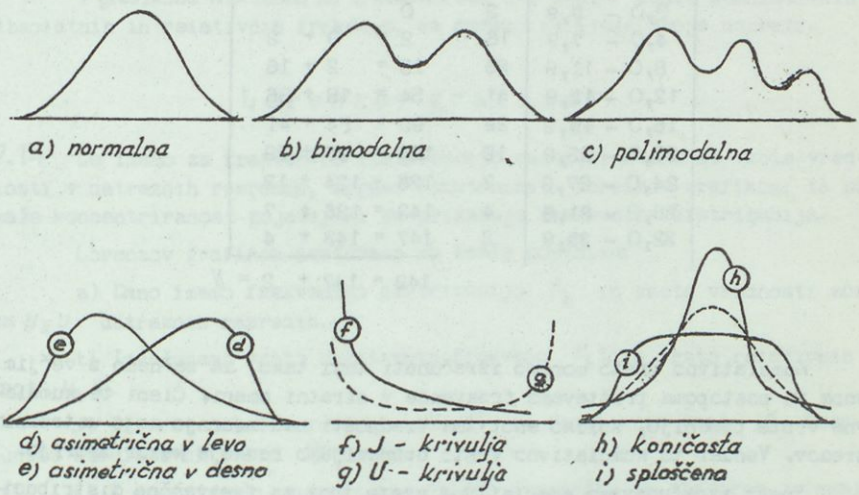


Slika 7.8 Frekvenčna distribucija za starost nevest v letu 1956 v Sloveniji (Vir: SL LRS 57)

7.11 Frekvenčna distribucija industrijskih podjetij po številu zaposlenih kaže stalno upadanje gostitve, čim večje je podjetje. Zaradi značilne oblike imenujemo te vrste frekvenčnih distribucij *J-distribucije*, ker imajo tolj ali manj obliko črke J. V J-distribuciji se distribuira veliko socialno-ekonomskih pojavov. Tako ima podobno obliko frekvenčna distribucija za kmetijska gospodarstva v LRS.

J-distribucijo dobimo, kadar je središče gostitve zelo blizu naravne omejitve. Za trdnost žice iz primera v sliki 7.8 bi dobili tipično sliko J-distribucije, če bi se splošna kvaliteta žice približala maksimalni trdnosti.

Če imamo normalno distribucijo za idealno, so nekatere distribucije v primerjavi z njo *koničaste*, druge pa *sploščene* (glej sliko 7.9 h, i).



Slika 7.9 Oblike frekvenčnih distribucij

KUMULATIVNA FREKVENČNA DISTRIBUCIJA

7.12 Ker frekvenčne distribucije prikazujejo število enot (ekstenziven pojatek) po grupah numeričnega znaka, moremo zanje izračunati kumulativne frekvenčne distribucije F_k po splošnem pravilu za izračunavanje kumulativnih vrst. Kumulativno frekvenčno distribucijo F_k dobimo, če postopoma seštevamo frekvence f_k v frekvenčni distribuciji. Člene kumulativne frekvenčne distribucije dobimo torej po obrazcu

$$F_{k+1} = F_k + f_k \qquad (7.5)$$

Če za frekvenčno distribucijo o porabi lesa iz tabele 7.3 izračunamo kumulativno distribucijo, dobimo kumulativno vrsto v tabeli 7.7.

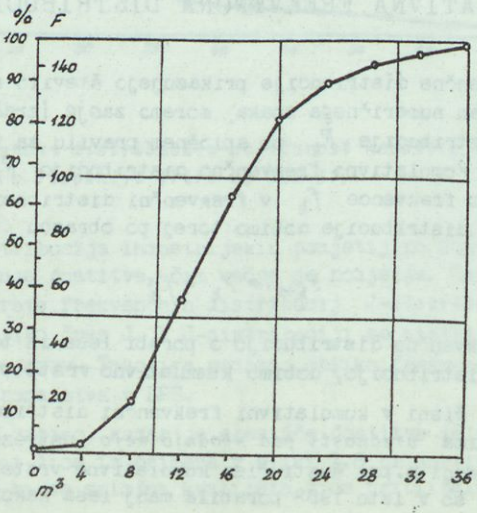
Posamezni člani v kumulativni frekvenčni distribuciji F_k pomenijo, koliko enot ima *vrednosti pod spodnjo mejo* ustreznega razreda $x_{k,min}$. Tako pomeni n.pr. šesti člen kumulativne vrste, da je $F_6 = 124$ gospodarstev, ki so v letu 1958 porabila manj lesa kakor $x_{6,min} = 20$ m^3 lesa.

Tabela 7.7 Kumulativna frekvenčna distribucija o porabi lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m^3	f_k	F_k
0,0 – 3,9	2	0
4,0 – 7,9	16	2 = 0 + 2
8,0 – 11,9	36	18 = 2 + 16
12,0 – 15,9	41	54 = 18 + 36
16,0 – 19,9	29	95 = 54 + 41
20,0 – 23,9	12	124 = 95 + 29
24,0 – 27,9	7	136 = 124 + 12
28,0 – 31,9	4	143 = 136 + 7
32,0 – 35,9	2	147 = 143 + 4
		149 = 147 + 2 = N

Kumulativno vrsto moremo izračunati tudi tako, da začnemo z večjim členom in postopoma prištevamo frekvence v obratni smeri. Členi te kumulativne vrste pomenijo, koliko enot ima vrednosti nad zgornjo mejo ustreznih razredov. Vendar to kumulativno vrsto upravljamo redkeje kakor zgornjo.

Enako izračunavamo kumulativne vrste tudi za frekvenčne distribucije z različnimi širinami razredov. Tudi pomen členov je isti kakor zgoraj. Kumulativne frekvenčne distribucije moremo izračunati tudi iz vrst relativnih frekvenc (glej tabelo 7.8).



Slika 7.10 Kumulativna frekvenčna distribucija o porabi lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

7.13 Če je frekvenčna distribucija unimodalna, ima grafični prikaz kumulativne frekvenčne distribucije značilno obliko črke S. O tem se moremo prepričati tudi v grafikonu za kumulativno frekvenčno distribucijo o porabi lesa v kmetijskih gospodarstvih v sliki 7.10, ki je načrtana po podatkih iz tabele 7.7. Opomniti moramo, da vrednosti kumulativnih frekvenc v grafikonu nanašamo nad spodnjo mejo ustreznih razredov na abscisi.

V grafikonu 7.10 imamo na ordinatni osi dve skali: skalo kumulativnih absolutnih in relativnih frekvenc, da moremo iz njega oboje odbrati.

Lorenzov grafikon

7.14 Če imamo za frekvenčno distribucijo razen frekvenc še vsote vrednosti v ustreznih razredih, moremo konstruirati Lorenzov grafikon; ta pokaže koncentriranost pojava, ki ga prikazuje frekvenčna distribucija.

Lorenzov grafikon sestavimo po temle postopku:

a) Dano imamo frekvenčno distribucijo f_k in vsoto vrednosti znaka y_k v ustreznih razredih.

b) Izračunamo vrsto relativnih frekvenc $f_k\%$ in vrsto relativnih vsot $y_k\%$.

c) Iz vrst relativnih frekvenc $f\%$ in relativnih vsot $y\%$ izračunamo kumulativni vrsti $F\%$ in $Y\%$.

d) V mreži grafikona, ki ima obliko kvadrata, je abscisna os skala za kumulativo relativnih frekvenc $F\%$, ordinatna os pa skala za kumulativo relativnih vsot $Y\%$. V grafikonu narišemo točke, ki imajo za abscise vrednosti kumulativne vrste relativnih frekvenc $F\%$, za ordinate pa ustrezne vrednosti kumulative $Y\%$. Če po vrsti povežemo dobljene točke, dobimo lomljeno črto, ki v loku veže spodnji levi kot grafikona z zgornjim desnim kotom grafikona. Čim večje so razlike po razredih, tem večja je koncentriranost pojava in tembolj je lok Lorenzove krivulje ukrivljen. Pri maksimalni koncentraciji preide lok v spodnjo in desno stranico kvadrata. Čim manjše so razlike po razredih, tem manjša je koncentriranost pojava in tembolj je lok izravnan. Če ni koncentracije pojava, lok preide v diagonalo.

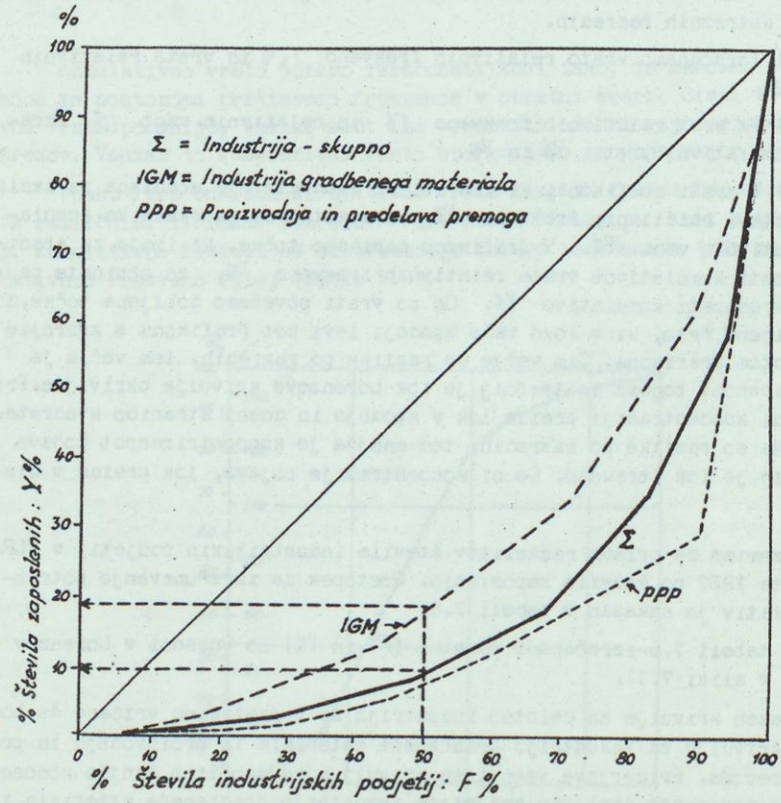
7.15 Vzemimo za primer razdelitev števila industrijskih podjetij v FLRJ konec leta 1957 po številu zaposlenih. Postopek za izračunavanje potrebnih kumulativ je nakazan v tabeli 7.8.

V tabeli 7.8 izračunani podatki ($F\%$ in $Y\%$) so vneseni v Lorenzov grafikon v sliki 7.11.

Razen krivulje za celotno industrijo so v grafikonu vrisane še Lorenzove krivulje za industrijo gradbenega materiala in proizvodnjo in predelavo premoga. Primerjava vseh treh krivulj pokaže očitno manjšo koncentracijo zaposlenega osebja v podjetjih industrije gradbenega materiala in večjo koncentracijo v podjetjih proizvodnje in predelave premoga. Iz grafikona lahko odberemo tudi druge važne podatke. Iz njega vidimo, da ima

Tabela 7.8 Izračunavanje kumulativ za Lorenzov grafikon za industrijska podjetja v FLRJ po številu zaposlenih konec leta 1957
(Vir: SG 58)

Število zaposlenih	Število podjetij			Število zaposlenih		
	f_k	f_k%	F_k%	y_k	y_k%	Y_k%
– 15	155	6,1	0	1441	0,2	0
16 – 30	218	8,6	6,1	5052	0,6	0,2
31 – 60	350	13,9	14,7	15755	1,9	0,8
61 – 125	510	20,2	28,3	46800	5,7	2,7
126 – 250	471	18,7	48,8	84475	10,4	8,4
251 – 500	375	14,9	67,5	133803	16,4	18,8
501 – 1000	276	10,9	82,4	192843	23,7	35,2
1001 –	170	6,7	93,3	334717	41,1	58,9
	2525	100,0	100,0	813886	100,0	100,0



Slika 7.11 Lorenzov grafikon za industrijska podjetja po številu zaposlenih v FLRJ za vsa industrijska podjetja (Σ), za industrijo gradbenega materiala (IGM) in proizvodnjo in predelavo premoga (PPP)

n.pr. polovica (manjša podjetja) podjetij v FLRJ nekaj manj kot 10 % od vseh zaposlenih, medtem ko ima druga polovica (večja podjetja) drugih 90 % zaposlenih. V industriji gradbenega materiala je zaradi manjše koncentracije v polovici manjših podjetij približno 19 % vseh zaposlenih v tej stroki. Enako moremo odbirati druge podobne odnose.

Lorenzov grafikon ima tudi druge tehnične prednosti. Sestavljanje Lorenzovih krivulj ni vezano na ustaljeno grupacijo in moremo v istem grafikonu vrtati Lorenzove krivulje iz različno grupiranih podatkov. To posebno olajša mednarodne primerjave, ker ima vsaka država grupacije prilagojene svojim pogojem in potrebam.

Osmo poglavje

K V A N T I L I

RANŽIRNA VRSTA. RANG

8.1 Osnovni statistični podatki, ki smo jih zbrali s statističnim opazovanjem, so dani v neurejeni množici podatkov. Te podatke običajno uredimo v frekvenčno distribucijo. Moremo pa jih pregledno prikazati tudi v *ranžirni vrsti*, urejene po velikosti od najmanjšega do največjega. V Tabeli 8.1 so prikazani osnovni podatki za množino padavin v mm v juniju 1954 za 23 meteoroloških postaj v Sloveniji.

Tabela 8.1 Množina padavin v mm v juniju 1954 za 23 meteoroloških postaj v Sloveniji (Vir: SL LRS 1956)

179 127 189 229 219 190 207 46 298 222 143 111 252 280
158 149 290 201 220 187 214 193 87 .

Vrsta podatkov v tabeli 8.1 ne da dobrega pregleda o velikosti padavin, ker je neurejena. Če iz teh podatkov sestavimo ranžirno vrsto, dobimo tabelo 8.2.

Tabela 8.2 Ranžirna vrsta podatkov o množini padavin v juniju 1954 za 23 meteoroloških postaj v Sloveniji

Rang	1	2	3	4	5	6	7	8	9	10	11	12
Množina padavin v mm	46	87	111	127	143	149	158	179	187	190	193	199

Rang	13	14	15	16	17	18	19	20	21	22	23
Množina padavin v mm	201	207	214	219	220	222	229	252	280	290	298

Iz ranžirne vrste pa moremo napraviti več zaključkov o prikazanem pojavu. Iz nje podobno kakor iz frekvenčne distribucije neposredno vidimo, v kakšnih mejah varlira populacija, ker kaže podatek, ki je prvi po rangu, najmanjšo (43 mm), podatek, ki je poslednji po rangu, pa največjo množino padavin (298 mm).

Razen tega vsaki enoti dodamo nov znak – *rang* enote v ranžirni vrsti. Rang daje o posamezni enoti dodatno informacijo, ki je osnovna vrednost znaka, – v našem primeru množina padavin – ne daje. Če namreč za dotično meteorološko postajo vemo, da je v juniju 1958 izkazala 252 mm pa-

davin; iz tega podatka ne moremo sklepati, ali je to za Slovenijo za to razdobje malo ali veliko. Če pa dodatno povemo, da je v ranžirni vrsti 23 postaj, ta postaja pa dvajseta po rangu, sklepamo, da je ta množina padavin za Slovenijo razmeroma velika, ker imajo od 28 postaj samo tri postaje večjo množino padavin kot 252 mm, devetnajst postaj pa manjšo.

KVANTILNI RANG

8.2 Hiba ranga je v tem, da moramo poleg ranga navesti še skupno število enot populacije, če hočemo, da rang pokaže mesto enote v populaciji. Postaja, ki je po padavinah dvajseta po rangu, je n.pr. za populacijo s 23 enotami, postaja z veliko padavinami, ker ima od 23 postaj 19 postaj manjšo, samo 3 pa večjo količino padavin. Če pa bi imeli n.pr. sto postaj, bi bila množina padavin za postajo, ki je dvajseta po rangu, razmeroma majhna, ker bi imelo samo 19 postaj manj, 80 postaj pa več padavin. Rang R pokaže torej mesto enote v populaciji šele, če ga primerjamo z obsegom populacije N. Zato je primerneje, da izračunamo mesto enote v populaciji namesto z rangom R s *kvantilnim rangom* P. Tega dobimo, če primerjamo rang R z obsegom populacije N. Kvantilni rang v relativnem številu pove, na katerem delu celotnega ranžirnega razmaka leži določena enota, oziroma koliki del celote ima manjše vrednosti, kakor je dana vrednost.

8.3 Teoretično vzamemo, da je ranžirni razmak zvezen in vsakemu rangu pripišemo razmak polovico enote na levo in desno. Po tej predpostavki se ranžirni razmak začne z 0,5 (spodnja meja razmaka, ki ustreza rangu 1) in konča z $N + 0,5$ (zgornja meja razmaka, ki ustreza rangu N). Ker se skala kvantilnih rangov P začne z 0 in konča z 1, je zveza med rangom R in kvantilnim rangom P dana z obrazcem

$$R = NP + 0,5 \qquad (8,1)$$

Če je $P = 0$, je $R = 0,5$, če pa je $P = 1$, je $R = N + 0,5$.

Iz zgornje zveze dobimo, da velja tudi

$$P = \frac{R - 0,5}{N} \qquad (8.2)$$

Če vzamemo zgornji primer meteorološke postaje z rangom $R = 20$ in $N = 23$, dobimo po obrazcu

$$P = \frac{20-0,5}{23} = 0,85$$

Kvantilni rang $P = 0,85$ pove, da ima 0,85 del celotne populacije manj padavin kakor je $R = 20$ ustrezna količina padavin (252 mm). Iz pri-

mera vidimo, da kvantilni rang P sam zase - ne da bi navajali obseg populacije - nazorno prikaže mesto določene enote v populaciji.

KVANTILI

8.4 Z obrazcema 8.1 in 8.2 moremo reševati dva, po svojem bistvu različna problema.

Za posamezno enoto populacije moremo določiti mesto enote v populaciji, če izračunamo vrednosti x ustrezni kvantilni rang P_x. Kvantilni rangi so torej karakteristike posameznih enot.

Moremo pa analogno reševati tudi drug problem. Če se vprašamo, kakšna vrednost x_p ustreza n.pr. kvantilnemu rangu $P = 0,50$, vrednost, ki jo dobimo, ne označuje posamezne enote, marveč populacijo. x_p ki ustreza danemu kvantilnemu rangu P, imenujemo *kvantil*. Kvantil $x_{0,50}$ je n.pr. vrednost, od katere ima polovica enot populacije manjše, polovica pa večje vrednosti. Ta vrednost je vsekakor važen parameter populacije in ga imenujemo *mediana*.

Kvantili $x_{0,25}$, $x_{0,50}$ in $x_{0,75}$ so vrednosti, ki razdele populacijo v štiri dele tako, da je pod $x_{0,25}$, med $x_{0,25}$ in $x_{0,50}$, med $x_{0,50}$ in $x_{0,75}$, in nad $x_{0,75}$ po četrtina po velikosti urejenih vrednosti populacije. Te vrednosti, ki razdele populacijo v štiri po obsegu enake dele, imenujemo *kvartile* in zaznamujemo z

$$Q_1 = x_{0,25} \; ; \quad Q_2 = x_{0,50} \; ; \quad Q_3 = x_{0,75} \tag{8.3}$$

Analogno z *decili*

$$D_1 = x_{0,10}, \quad D_2 = x_{0,20} \ldots\ldots\ldots D_9 = x_{0,90} \tag{8.4}$$

razdelimo celotno populacijo v deset po obsegu enakih delov, s *centili* :

$$C_1 = x_{0,01}, \quad C_2 = x_{0,02} \ldots\ldots C_{98} = x_{0,98} \text{ in } C_{99} = x_{0,99} \tag{8.5}$$

pa v sto po obsegu enakih delov.

Izračunavanje kvantilov iz negrupiranih podatkov

8.5 Ker predpostavljamo, da je rang zvezna količina, moremo določiti range za vsako vrednost med najmanjšo in največjo vrednostjo v populaciji in ne samo za podatke, navedene v ranžirni vrsti. Za vmesne vrednosti uporabimo linearno interpolacijo. Tako n.pr. menimo, da ustreza v zgornjem

primeru rangu 12,5 sredina med 199 mm in 201 mm padavin, torej 200 mm padavin.

Če predpostavljamo zveznost rangov, izračunavamo iz ranžirne vrste za poljubno vrednost x med najmanjšo in največjo vrednostjo populacije ustrezni kvantilni rang P_x po naslednjem postopku:

a) Imamo ranžirno vrsto vrednosti enot populacije.

b) V ranžirni vrsti poiščemo, med kateri vrednosti x_0 in x_1 pade vrednost x; za katero iščemo P_x, tako da velja: $x_0 < x < x_1$. Vrednosti x_0 naj ustreza rang R_0.

c) Rang R_x, ki ustreza vrednosti x, dobimo z linearno interpolacijo po obrazcu

$$R_x = R_0 + \frac{x-x_0}{x_1-x_0} \qquad (8,6)$$

d) Kvantilni rang P_x izračunamo iz ranga R_x in obsega populacije po obrazcu

$$P_x = \frac{R_x-0,5}{N} \qquad (8,7)$$

Če iščemo, kakšen kvantilni rang ustreza n.pr. padavinam $x = 150$ mm, dobimo po zgornjem postopku:

$x = 150$ leži v ranžirni vrsti med vrednostima $x_0 = 149$, za katero je rang $R_0 = 6$, in vrednostjo $x_1 = 158$

Iz teh podatkov izračunamo rang R_x, ki ustreza $x = 150$ mm, po obrazcu 8.6.

$$R_x = 6 + \frac{150-149}{158-149} = 6,11$$

Po obrazcu 8.7 pa dobimo dalje: $P_x = \frac{6,11 - 0,50}{28} = 0,244$

Kvantilni rang kraja, ki je imel v juniju 1958 $x = 150$ mm padavin, je 0,244 ali izraženo v centilih: Kraj z $x = 150$ mm padavin je v 24.centilu.

8.6 Obratni problem, da k danemu kvantilnemu rangu P poiščemo ustrezni kvantil x_P, pa rešimo takole:

a) Za populacijo imamo ranžirno vrsto.

b) Iz danega P po obrazcu

$$R_P = NP + 0,5 \qquad (8.8)$$

izračunamo ustrezni rang R.

c) V ranžirni vrsti poiščemo, med katera cela ranga pade izračunani R_p tako, da velja: $R_0 < R_p < R_1$. Rangoma R_0 in R_1 ustrezata vrednosti x_0 in x_1.

d) Iz teh podatkov izračunamo z linearno interpolacijo kvantil x_p po obrazcu:

$$x_p = x_0 + (x_1 - x_0).(R_p - R_0) \qquad (8.9)$$

Kako iz podatkov o množini padavin v juniju 1953 izračunamo kvantile, če upoštevamo tabelo 8.2, je nakazano v tabeli 8.3.

Tabela 8.3 Izračunavanje kvartilov za množino padavin v v juniju 1953 v Sloveniji

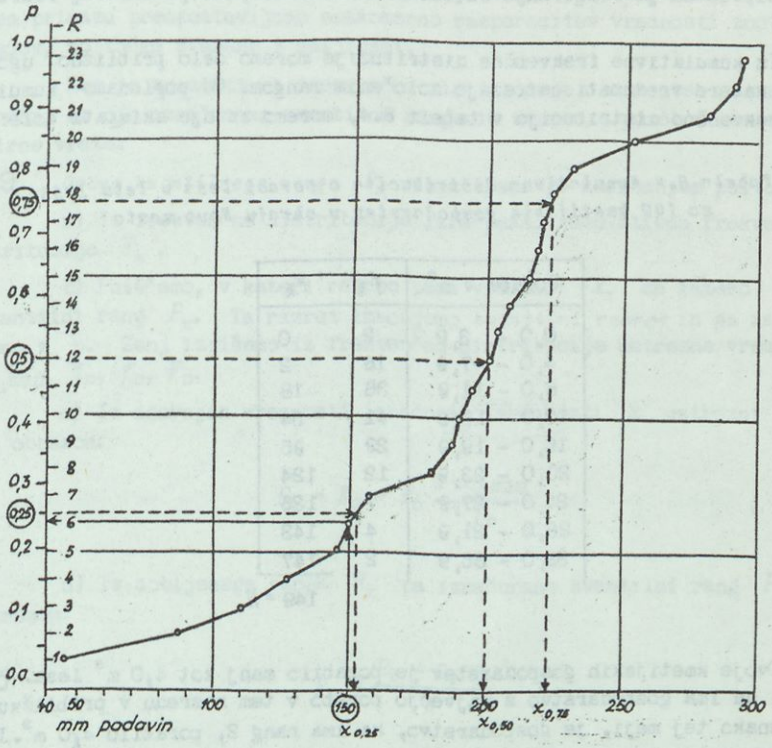
i	P	R_p	R_0	x_0	x_1	x_1-x_0	$(x_1-x_0)(R_p-R_0)$	$x_p = Q_i$
1	0,25	6,25	6	149	158	9	2,25	151,25 = Q_1
2	0,50	12,00	12	199	199	0	0	199,00 = Q_2
3	0,75	17,75	17	220	222	2	1,50	221,50 = Q_3

Za prvi kvartil izračunamo rang $R = 23.0,25 + 0,5 = 6,25$ in analogno ostala dva.

Poseben primer je drugi kvartil. Zanj je kvantilnemu rangu ustrezni rang celo število. Zato je nadaljnji postopek nepotreben, ker je kvantil kar vrednost, ki ustreza temu rangu. V našem primeru je to 199 mm.

Kvartili razmeroma dobro označujejo populacijo. Četrtina postaj je imela pod 151,25 mm padavin, polovico pod 199 mm, in četrtina nad 221,5 mm.

8.7 V sliki 8.1 imamo v grafikonu za primer padavin iz tabele 8.2 vrisano ranžirno vrsto tako, da je abscisa vrednost podatka, ordinata pa ustrezni rang. Točke predstavljajo osnovno ranžirno vrsto, lomljena črta, ki veže točke, pa linearno interpolacijo za vmesne vrednosti. V grafikonu je razen skale rangov vrisana tudi skala kvantilnih rangov P, za katero je nazorno viano, da se na skali rangov začne z 0,5 in konča z $N + 0,5$. Na grafikonu je tudi nakazano, kako moremo grafično najti dani vrednosti x ustrezen kvantilni rang P_x in danim kvantilnim rangom P ustrezne kvantile x_p. Grafično so rešeni problemi, ki smo jih zgoraj rešili računsko.



Slika 8.1 Zveza med kvantili, rangi in kvantilnimi rangi za množino padavin za 28 meteoroloških postaj v Sloveniji v juniju 1953.

Izračunavanje kvantilnih rangov in kvantilov iz frekvenčnih distribucij

8.8 Če je populacija, ki jo proučujemo, obsežna, je razvrščanje vrednosti v ranžirno vrsto neprikladno. Za obsežne populacije pa moremo izračunati približne vrednosti kvantilnih rangov in kvantilov iz kumulativnih frekvenčnih distribucij. Če vemo, kaj pomenijo člani kumulativne vrste, spoznamo, da je kumulativna frekvenčna distribucija v zvezi z rangi.

Proučimo frekvenčno distribucijo o porabi lesa v 142 kmetijskih gospodarstvih v Novem mestu iz tabele 7.8. Iz nje vidimo, da ima frekvenčna distribucija nekatere lastnosti ranžirne vrste. V ranžirni vrsti so vrednosti, ki leže levo od dane vrednosti, manjše, desno pa večje. V frekvenčni distribuciji pa so vse vrednosti enot v danem razredu večje od vrednosti v vseh spodnjih razredih in manjše od vrednosti v vseh zgornjih raz-

redih. Tudi v frekvenčni distribuciji so torej vrednosti razporejene po velikosti, le da je rangiranje izvedeno med razredi, ne pa znotraj razredov.

Iz kumulativne frekvenčne distribucije moremo celo približno ugotoviti, katere vrednosti ustrezajo določenim rangom. Če pogledamo kumulativno frekvenčno distribucijo v tabeli 8.4, moremo iz nje sklepati tole:

Tabela 8.4 Kumulativna distribucija o porabi lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m^3	f_k	F_k
0,0 - 3,9	2	0
4,0 - 7,9	16	2
8,0 - 11,9	36	18
12,0 - 15,9	41	54
16,0 - 19,9	22	95
20,0 - 23,9	12	124
24,0 - 27,9	7	136
28,0 - 31,9	4	143
32,0 - 35,9	2	147
		149 = N

Dvoje kmetijskih gospodarstev je porabilo manj kot 4,0 m^3 lesa. Če vzamemo, da ima gospodarstvo z največjo porabo v tem razredu v približku porabo enako tej meji, je gospodarstvo, ki ima rang 2, porabilo 4,0 m^3. Iz kumulative dalje izvemo, da je 18 gospodarstev porabilo pod 8,0 m^3. Če vzamemo, da je največja poraba v teh 18 gospodarstvih enaka tej meji, dobimo, da ima gospodarstvo, ki je porabilo 8,0 m^3 lesa, rang 18 itd. Iz frekvenčne distribucije sklepamo, da imajo gospodarstva, ki so porabila toliko, kot so spodnje meje razredov, range enake ustreznim vrednostim kumulativne serije. Iz tega moremo sestaviti tabelo, ki je zelo podobna ranžirni vrsti, le da so rangi dani samo za meje razredov.

Tabela 8.5 Okrnjena ranžirna vrsta za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto (dobljena iz kumulativne vrste v tab.8.4)

Rang R F_k	2	18	54	95	124	136	143	147	149
$x_R = x_{k,\min}$	4,0	8,0	12,0	16,0	20,0	24,0	28,0	32,0	36,0

Ker za druge enote v ranžirni vrsti ne poznamo vrednosti, predpostavljamo, da so vrednosti znotraj razredov razmeščene enakomerno. Tako dobimo vrednosti drugih enot v ranžirni vrsti z linearno interpolacijo.

Grafikon kumulativne frekvenčne distribucije v sliki 8.2 ponazarja približen odnos med rangom R in vrednostjo znaka x. Ker tudi pri grafičnem prikazu predpostavljamo enakomerno razporeditev vrednosti znotraj razredov, so točke zvezane z daljicami.

Ker je kumulativna frekvenčna distribucija nadomestek za ranžirno vrsto, iz nje izračunamo kvantilne range in kvantile podobno kakor iz ranžirne vrste.

8.9 *Oceno kvantilnega ranga* P_x izračunamo po naslednjem postopku:

a) Iz frekvenčne distribucije izračunamo kumulativno frekvenčno distribucijo F_k.

b) Poiščemo, v kateri razred pade vrednost x, za katero iščemo kvantilni rang P_x. Ta razred imenujemo *kvantilni razred* in ga zaznamujemo z o. Zanj izpišemo iz frekvenčne distribucije ustrezne vrednosti: $x_{o,min}$, i_o, f_o, F_o.

c) Iz zgornjih vrednosti izračunamo vrednosti x ustrezni rang R_x po obrazcu:

$$R_x = F_o + f_o \frac{x - x_{o,min}}{i_o} \qquad (8.10)$$

d) Iz dobljenega ranga R_x pa izračunamo kvantilni rang P_x po obrazcu:

$$P_x = \frac{R_x - 0,50}{N} \qquad (8.11)$$

Če je obseg populacije N velik, iz obrazca 8.11 običajno izpuščamo 0,5, ker je ta količina za velike populacije nebistvena.

Postopek velja tudi za frekvenčne distribucije, ki imajo različne širine razredov.

8.10 Če za populacijo o porabi lesa iz tabele 8.4 iščemo kvantilni rang za gospodarstvo, ki je v letu 1953 porabilo $x = 13,8\ m^3$ lesa, izračunamo P_x po zgornjem postopku takole:

Poraba lesa $x = 13,8\ m^3$ pade v razred 12,0–15,9. Iz tabele 8.4 moremo odbrati:

$$x_{o,min} = 12,0; \quad i_o = 4; \quad f_o = 41; \quad F_o = 54;$$

Če vstavimo te podatke v obrazec 8.10, dobimo:

$$R_x = 54 + 41 \cdot \frac{13,8-12,0}{4} = 72,4$$

in dalje po obrazcu 8.11: $P_{x=13,8} = \frac{72,4 - 0,5}{149} = 0,483$

Gospodarstvo s porabo $x = 13,8\ m^3$ lesa je v 48. centilu.

8.11 Analogno moremo iz frekvenčne distribucije izračunati tudi *kvantile*. Postopek je naslednji:

a) Iz frekvenčne distribucije izračunamo kumulativno frekvenčno distribucijo F_k.

b) Iz danega P izračunamo ustrezni rang R_P po obrazcu

$$R_P = NP + 0,5 \qquad (8.12)$$

Če je populacija velika, v obrazcu 8.12 izpustimo 0,5.

c) V kumulativni frekvenčni distribuciji F_k poiščemo, med kateri vrednosti kumulativne serije pade R_P tako da je: $F_0 < R_P < F_1$. F_0 ustrezen razred je kvantilni razred. Zanj poiščemo v frekvenčni distribuciji količine: $x_{0,min}$, i_0, f_0 in F_0.

d) Iz teh količin izračunamo ustrezni kvantil x_p po obrazcu:

$$x_p = x_{0,min} + i_0 \frac{R_P - F_0}{f_0} \qquad (8.13)$$

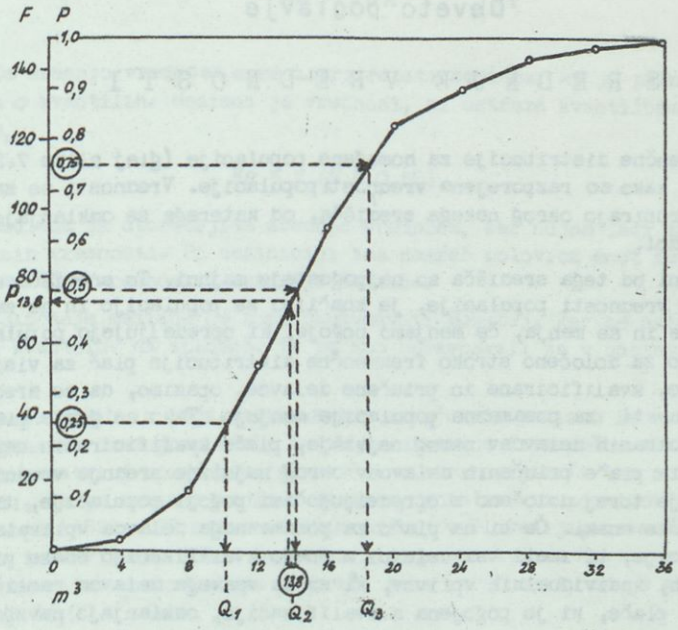
Enako kot za kvantilne range velja ta postopek tudi za frekvenčne distribucije z različnimi širinami razredov.

8.12 Če nočemo za naš primer izračunati kvartile, je najbolje, da sistematično izračunamo v računski tabeli vse tri kvartile hkrati.

Tabela 8.6 Izračunavanje kvartilov za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto

i	P	R_P	$x_{0,min}$	i_0	f_0	F_0	$R_P - F_0$	$i_0 \frac{R_P - F_0}{f_0}$	Q_i
1	0,25	37,75	8,0	4	33	18	19,75	2,19	10,19 = Q_1
2	0,50	75,00	12,0	4	41	54	21,00	2,05	14,05 = Q_2
3	0,75	112,25	16,0	4	29	85	17,25	2,38	18,38 = Q_3

8.13 Podobno kakor v sliki 8.1 za negrupirane podatke moremo grafično oceniti vse zgornje količine iz grafikona kumulativne serije. V sliki 8.2 je nakazano, kako dani vrednosti $x = 13,8\ m^3$ poiščemo ustrezni kvantilni rang, in obratno: kako moremo grafično oceniti kvartile.



Slika 8.2 Grafično ocenjevanje kvantilnih rangov in kvantilov iz slike kumulativnih frekvenčnih distribucij za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu v letu 1953

8.14 V praksi večkrat izdelujemo decilne ali centilne norme, s katerimi dobimo neposredno zvezo vrednosti znaka z decili oziroma, pri podrobnejših normnih tablicah, s centili. Centilne norme s pridom uporabljamo v psihologiji, moremo pa jih izdelati tudi za določene socialno-ekonomske probleme. Tako tablica decilnih ali centilnih norm o produktivnosti dela pomaga presoditi, kakšna je produktivnost dela za posameznega delavca v razmerju s celotnim kolektivom ita.

Deveto poglavje

S R E D N J E V R E D N O S T I

9.1 Frekvenčne distribucije za homogene populacije (glej slike 7.3, 7.6, 7.7) kažejo, kako so razporejene vrednosti populacije. Vrednosti se za te populacije grupirajo okrog nekega središča, od katerega se odklanjajo navzgor in navzdol.

Odkloni od tega središča so najpogosteje majhni. To središče, ki reprezentira vrednosti populacije, je značilno za populacijo in ne za posamezne enote in se menja, če menjamo pogoje, ki opredeljujejo populacijo. Če primerjamo za določeno stroko frekvenčne distribucije plač za visoko kvalificirane, kvalificirane in priučene delavce, opazimo, da se središča *srednje vrednosti* za posamezne populacije menjajo. Tako se goste plače visoko kvalificiranih delavcev okrog najvišje, plače kvalificiranih delavcev okrog nižje in plače priučenih delavcev okrog najnižje srednje vrednosti. To središče je torej določeno z opredeljujočimi pogoji populacije, ti pa so za vse enote enaki. Če bi na plačo za posameznega delavca vplivala samo kvalifikacija, bi imeli vsi delavci z enako kvalifikacijo enake plače. Zaradi drugih, individualnih vplivov, ki so za vsakega delavca različni, pa se plače, od plače, ki je pogojena s kvalifikacijo, odklanjajo navzgor in navzdol. Čim manjši so individualni vplivi, tem bolje srednja vrednost reprezentira vrednosti populacije in obratno: čim večji so individualni vplivi, tem slabše srednja vrednost reprezentira vrednosti populacije. Zato je srednja vrednost reprezentant vrednosti v populaciji le, če je populacija homogena.

V r s t e s r e d n j i h v r e d n o s t i

9.2 Parametrov, ki pokažejo centralno tendenco vrednosti populacije, imamo več. Od teh so v socialno-ekonomski statistiki pomembne naslednje srednje vrednosti:

a/ *mediana*,
b/ *modus*,
c/ *aritmetična sredina*,
d/ *harmonična sredina*,
e/ *geometrijska sredina*,

Medtem ko sta mediana in modus dani z lego vrednosti, štejemo ostale tri: aritmetično, harmonično in geometrijsko sredino med izračunane srednje vrednosti.

MEDIANA

9.3 Za srednjo vrednost more dobro rabiti mediana, ki jo poznamo že iz poglavja o kvantilih. Mediana je vrednost, ki ustreza kvantilnemu rangu $P = 0,50$,

$$Me = x\ (P = 0,50)$$

Mediana je uporabljiva srednja vrednost, ker nujno leži sredi individualnih vrednosti. Po definiciji ima namreč polovica enot manjše, polovica pa večje vrednosti, kakor je mediana.

Določanje mediane ne bomo obravnavali posebej, ker se sklada z določanjem kvantilov na splošno. Te pa smo obravnavali že v odstavku o kvantilih.

Na primerih iz tega odstavka smo izračunali mediano iz ranžirne vrste, če podatki niso grupirani, in iz kumulativne frekvenčne distribucije, če so podatki grupirani v frekvenčni distribuciji. Iz teh primerov smo dobili, da je mediana množine padavin, določena iz podatkov o padavinah v juniju 1958 za 28 meteoroloških postaj v Sloveniji, enaka $Me = 199$mm. Za porabo lesa v letu 1958 v kmetijskih gospodarstvih v Novem mestu pa smo iz frekvenčne distribucije o porabi lesa izračunali, da je mediana enaka $Me = 14,05$ m^3 lesa.

9.4 Lastnosti mediane. Prednost mediane je predvsem v tem, da je lahko razumljiva in zato kot opisni parameter priporočljiva. Velika prednost mediane pred izračunanimi sredinami je tudi ta, da za določanje mediane ni nujno, da poznamo vrednosti za vse enote populacije. Zadosti je, da poznamo vrednosti za enote, ki leže okoli sredine v ranžirni vrsti. Ta lastnost mediane pride posebno prav, če izračunavamo srednjo vrednost iz frekvenčne distribucije, ki ima odprte razrede. Za odprte razrede namreč dostikrat niti približno ne vemo, kakšne vrednosti vsebujejo, ker so omejeni samo navzgor ali navzdol. Mediana je vselej primerna srednja vrednost tudi takrat, če so ekstremne vrednosti take, da sumimo, ali sploh sodijo v homogeno populacijo, ali niso mogoče izraz nekih drugih kvalitet.

Pomanjkljivost mediane pa je v tem, da je le preveč neobčutljiva za spremembe vrednosti in sejnena vrednost ne spremeni vse dotlej, dokler so spremembe take, da vrednosti ne preidejo iz ene polovice v drugo. Vsota absolutnih odklonov posameznih vrednosti od neke vrednosti je najmanjša, če odklone računamo od mediane. Mediana je torej sorazmerno dober reprezentant posameznih vrednosti.

MODUS

9.5 Mediana pa ni vedno vrednost, ki dobro reprezentira vrednosti po-

populacije. Za asimetrične ali polimodalne distribucije je mediana vrednost, ki je različna od večine vrednosti v populaciji. Če naj bo srednja vrednost reprezentant populacije, je v takih primerih veliko primerneje, da vzamemo za sredino vrednost, okrog katere se vrednosti populacije najbolj goste. Če pogledamo poligone frekvenčnih distribucij, moremo za vsako frekvenčno distribucijo zlahka ugotoviti približno mesto največje gostitve. Vrednost, okrog katere se najbolj goste vrednosti populacije, imenujemo *modus* ali *najpogostejšo vrednost.*

Modus moremo ugotoviti samo za razmeroma obsežne populacije, ki so grupirane v frekvenčni distribuciji. Iz individualnih podatkov je modus nemogoče ugotoviti.

Če imamo namesto frekvenčne distribucije frekvenčno krivuljo, modus poiščemo enostavno. V tem primeru je modus abscisa tiste točke na frekvenčni krivulji, za katero je ordinata (gostota frekvence) največja. To je lepo razvidno iz slike 9.2.

Izračunavanje modusa iz frekvenčnih distribucij

9.6 V praktičnih primerih za populacije nimamo frekvenčnih krivulj, temveč frekvenčne distribucije, ki jih moremo narisati v histogramu. Vendar moremo tudi iz frekvenčne distribucije oziroma histograma sklepati, kje je gostota frekvence največja, ker histogram nakaže približno obliko frekvenčne krivulje. Iz histograma sklepamo, da je modus v razredu, v katerem je frekvenca največja. V primeru za porabo lesa v 149 kmetijskih gospodarstvih v letu 1958 v Novem mestu iz tabele 7.3 sklepamo, da je najpogostejša vrednost v razredu od 12,0 do 13,0 m^3 lesa. Kot prvi približek modusa vzamemo kar sredino razreda z največjo frekvenco-modalnega razreda. V našem primeru je prva ocena modusa za porabo lesa 14 m^3. Sredina razreda pa je prava vrednost modusa le, če je distribucija simetrična. Drugače leži modus nad sredino razreda ali pod njo, kar je odvisno od tega, kakšna je frekvenčna distribucija.

Natančnejšo oceno modusa dobimo, če upoštevamo tudi frekvence v sosednjih razredih. Iz histograma o porabi lesa v sliki 7.12 moremo sklepati, da je modus za porabo lesa verjetno pod sredino modalnega razreda s sredino 14 m^3. Ker je frekvenca pred modalnim razredom večja kot frekvenca za modalnim razredom, bi bila frekvenčna krivulja, ki bi jo načrtali med stolpce histograma, nagnjena proti sosednjemu razredu z višjo frekvenco. Predpostavimo, da je v okolici modusa frekvenčna krivulja v približku parabola druge stopnje, ki gre skozi točke s koordinatama (x_{-1}, f_{-1}), (x_0, f_0) in (x_{+1}, f_{+1}). Pri tem pomenijo x_{-1}, x_0 in x_{+1} sredine modalnega razreda in njegovih sosedov, f_{-1}, f_0 in f_{+1} pa ustrezne frekvence. Pri tej predpostavki izračunamo modus po naslednjem postopku:

a/ Podatke imamo grupirane v frekvenčni distribuciji z razredi enake širine i. Razredi morajo biti tako veliki, da frekvenčna distribucija izra-

za zakonitost gostitve frekvenc.

b/ V frekvenčni distribuciji poiščemo razred 0, ki ima največjo frekvenco $f_{-1} < f_0 > f_{+1}$. Razred z največjo frekvenco imenujemo *modalni razred*.

c/ Iz modalne frekvence f_0 in frekvenc obeh sosednih razredov f_{-1} in f_{+1} izračunamo $d_{-1} = f_0 - f_{-1}$ in $d_{+1} = f_0 - f_{+1}$

d/ Modus izračunamo po obrazcu

$$M_0 = x_{0,min} + i \frac{d_{-1}}{d_{-1} + d_{+1}} \qquad (9.1)$$

Pri tem pomeni razen že navedenih izrazov: $x_{0,min}$ = spodnja meja modalnega razreda.

9.7 Če pogledamo primer o porabi lesa v 149 kmetijskih gospodarstvih v letu 1953 v Novem mestu, vidimo, da je od treh distribucij, ki smo jih sestavili za izračunavanje modusa, prikladna distribucija, v kateri je razredna širina $i = 4$ m^3. V distribuciji v tabeli 7.2 se zakonitost gostitve še ne kaže, v distribuciji v tabeli 7.4 pa je zaradi prevelike širine razredov zakonitost gostitve že zabrisana.

Iz frekvenčne distribucije v tabeli 7.3 sklepamo, da je modalni razred razred 12,0 – 15,9 m^3, ker ima največjo frekvenco ($f_0 = 41$). Ker je za to distribucijo $f_{-1} = 36$, in $f_{+1} = 29$, velja dalje:

$$d_{-1} = f_0 - f_{-1} = 41 - 36 = 5; \quad d_{+1} = f_0 - f_{+1} = 41 - 29 = 12.$$

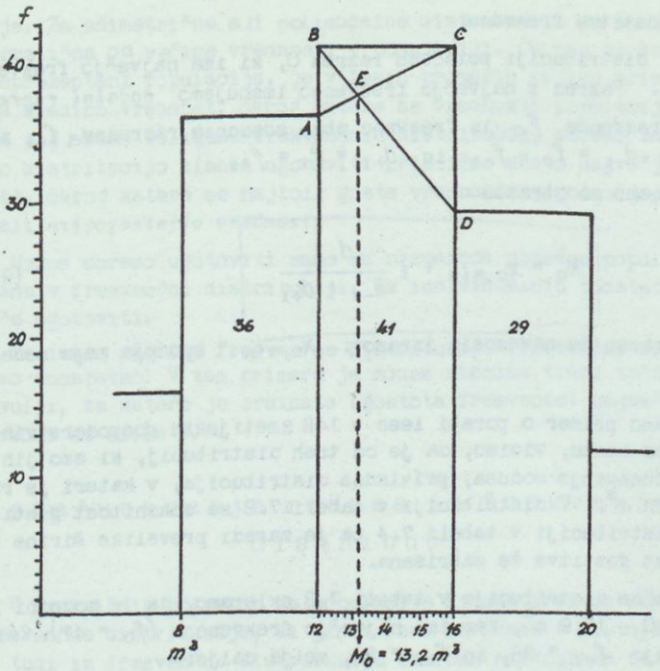
$$i = 4; \quad x_{0,min} = 12,0.$$

Iz teh podatkov je modus po obrazcu 9.1

$$M_0 = 12,0 + 4 \cdot \frac{5}{5 + 12} = 13,17 \text{ m}^3$$

Druga ocena modusa $M_0 = 13,17$ m^3 je stvarno pod prvo oceno 14,0 m^3.

9.8 Modus pa moremo z istimi predpostavkami zelo enostavno oceniti tudi iz histograma frekvenčne distribucije. Kakor kaže slika 9.1, zvežemo v histogramu oglišče A z C, oglišče B z D. Projekcija sečišča obeh veznic E je ocena modusa M_0. Slika pokaže, da grafičen način da v našem primeru praktično isti rezultat kakor račun.



Slika 9.1 Grafičen način določanja modusa iz frekvenčne distribucije za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu

Lastnosti modusa

9.9 Modus je srednja vrednost, ki dobro reprezentira vrednosti populacije, ker je že po definiciji vrednost, ki se v populaciji najpogosteje pojavlja. Enako kakor mediana je tudi modus srednja vrednost, ki je dana z lego vrednosti populacije. Zato je tudi modus neobčutljiv za spremembe vrednosti posameznih enot vse dotlej, dokler gostitev na nekem drugem mestu ne prekorači stopnje gostitve v modusu. To more biti dobra lastnost modusa, ker ni odvisen od vrednosti, ki za populacijo niso tipične, more pa biti tudi hiba, ker je le premalo odvisen od vrednosti, ki niso najpogostejše.

Frekvenčna distribucija, ki ni homogena, more imeti tudi več mest gostitve. Tako imajo *bimodalne distribucije* dva, *polimodalne distribucije* pa celo več središč gostitve. Take distribucije imajo torej dva ali več relativnih - lokalnih modusov. Od teh pa je absoluten modus tisti, za katerega je gostitev največja. Relativne - lokalne moduse ocenjujemo enako kot moduse za unimodalne distribucije.

Pojem modusa uporabljamo včasih že pri zbiranju statističnih po-

datkov in ne samo pri obdelavi. Tako pri registriranju cen na tržišču registriramo modus cene, to je ceno, po kateri je na tržišču največ prodaj blaga.

ARITMETIČNA SREDINA

9.10 Izmed vseh srednjih vrednosti je najbolj znana in uporabljena aritmetična sredina ali povprečje. Aritmetično sredino M dobimo, če delimo vsoto vrednosti x vseh enot v populaciji z obsegom populacije N. Z obrazcem moremo to izraziti

$$M_x = \bar{x} = \frac{1}{N}(x_1 + x_2 + \ldots + x_N) = \frac{1}{N}\Sigma x = X/N \qquad (9.2)$$

Pri tem pomeni: $M_x = \bar{x}$ = aritmetična sredina. Znak M_x včasih zamenjamo z znakom $\bar{x}$ (x prečna); $\Sigma x = X$ = vsota vseh vrednosti v populaciji; Σ = splošen znak za seštevanje izraza, ki stoji za njim; x = individualne vrednosti. Te oznake so splošne in jih bomo uporabljali tudi v drugih primerih.

9.11 Aritmetična sredina je izpeljana iz predpostavke, da je vrednost x za posamezno enoto vsota rezultatov splošnih (M) in individualnih vplivov (e)

$$x = M + e$$

Nadaljnja predpostavka pa je, da se rezultati individualnih vplivov v vsoti uničijo. $\Sigma e = 0$.

Če upoštevamo zgornje predpostavke in seštejemo vrednosti x za celo populacijo, dobimo

$$\Sigma x = NM + \Sigma e = NM.$$

Iz te enačbe zaključimo, da je rezultat splošnih vplivov $M = X/N$. To pa je po definiciji aritmetična sredina.

Lastnosti aritmetične sredine

9.12 Iz definicije aritmetične sredine izvirajo različne lastnosti, ki so zelo važne za prakso in teorijo.

a) Aritmetična sredina za linearno zvezo $z = a + bx + cy$ med znaki x in y za isto populacijo, je enaka linearni zvezi aritmetičnih sredin. Z obrazcem moremo ta stavek napisati:

$$\bar{z} = \overline{a + bx + cy} = a + b\bar{x} + c\bar{y} \qquad (9.3)$$

Ta stavek moremo zlahka dokazati. Za posamezno enoto in za vsoto velja:

$$z_1 = a + bx_1 + cy_1$$

$$z_2 = a + bx_2 + cy_2$$

$$\cdots\cdots\cdots$$

$$z_N = a + bx_N + cy_N$$

$$\Sigma z = Na + b\Sigma x + c\Sigma y$$

Ker je po definiciji $\Sigma z = N\bar{z}$, $\Sigma x = N\bar{x}$ in $\Sigma y = N\bar{y}$, velja dalje za vsoto:

$$N\bar{z} = Na + bN\bar{x} + cN\bar{y}$$

Če to enačbo delimo z N, dobimo

$$\bar{z} = a + b\bar{x} + c\bar{y}$$

Ta sestavek ni omejen samo na linearno zvezo dveh znakov in ga moremo po potrebi razširiti.

Iz gornjega stavka sledita neposredno nova stavka: Aritmetična sredina konstante je konstanta

$$\bar{a} = a \qquad (9.4)$$

Aritmetična sredina produkta znaka s konstanto je enaka produktu konstante in aritmetične sredine znaka

$$\overline{bx} = b.\bar{x} \qquad (9.5)$$

b) Vsota odklonov individualnih vrednosti od aritmetične sredine je enaka 0.

$$\Sigma(x-\bar{x}) = 0 \qquad (9.6)$$

Dokaz je preprost: $\Sigma(x - \bar{x}) = \Sigma x - N\bar{x} = X - X = 0$

c) Vsota kvadratov odklonov individualnih vrednosti od neke konstante A je najmanjša, če je A enak aritmetični sredini $\bar{x}$.

$$K = \Sigma(x - A)^2 = \text{Min, če je } A = \bar{x} \qquad (9.7)$$

Dokaz: Če naj bo K minimalen, mora biti $\frac{dK}{dA} = 0$

$$\frac{dK}{dA} = -2\Sigma(x-A) = 0; \quad \Sigma(x-A) = \Sigma x - NA = 0; \quad A = \frac{1}{N}\Sigma x$$

Ta lastnost aritmetičnih sredin je zelo važna za nadaljnje proučevanje populacij. Iz vsote kvadratov odklonov od aritmetične sredine je namreč izpeljan poseben parameter, *varianca*, ki je najvažnejša mera jakosti individualnih vplivov.

a) Sumarno aritmetično sredino populacije izračunamo iz aritmetičnih sredin delnih populacij po obrazcu

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 \ldots + N_r\bar{x}_r}{N_1 + N_2 + \ldots + N_r} = \frac{\Sigma N_k\bar{x}_k}{\Sigma N_k} \qquad (9.8)$$

Ta način za izračunavanje aritmetične sredine imenujemo *tehtano izračunavanje* za razliko od enostavnega načina izračunavanja po obrazcu 9.2.

Dokaz: Skupna vsota X je enaka vsoti delnih vsot X_k za x

$$X = X_1 + X_2 + \ldots + X_k + \ldots + X_r$$

Ker velja tako za skupno kot za dele $X = N\bar{x}$, je dalje:

$$N\bar{x} = N_1\bar{x}_1 + N_2\bar{x}_2 + \ldots + N_k\bar{x}_k + \ldots + N_r\bar{x}_r$$

Če delimo to enačbo z N, dobimo obrazec 9.8.

Izračunavanje aritmetične sredine iz negrupiranih podatkov

9.13 Iz individualnih podatkov izračunamo aritmetično sredino po osnovnem obrazcu 9.2 tako, da vsoto podatkov delimo z obsegom populacije N.

Če po obrazcu 9.2 izračunamo aritmetično sredino za množino padavin v juniju 1953 za 23 meteoroloških postaj iz tabele 9.1, dobimo

$$M_x = \frac{179+127+199+\ldots+214+193+87}{23} = \frac{4399}{23} = 191{,}3 \text{ mm}$$

Povprečno je odpadlo na eno postajo 191,3 mm padavin. Aritmetična sredina pokaže, na kakšni ravni so bile padavine v Sloveniji v juniju 1953 in je izraz splošnega stanja padavin v tem razdobju. Dobljeno povprečje pomeni: če bi bila celotna množina padavin v vseh 23 postajah enakomerno porazdeljena po posameznih postajah, bi na eno postajo odpadlo 191,3 mm padavin.

Izračunavanje aritmetične sredine iz frekvenčnih distribucij

9.14 Direktna metoda. Izračunavanje aritmetične sredine po obrazcu 9.2 je za velike populacije zaradi velikega števila sumandov zamudno. Ker daje frekvenčna distribucija približno sliko vseh vrednosti populacije, moremo aritmetično sredino oceniti iz frekvenčne distribucije.

Če predpostavljamo, da sredina razreda x_k v frekvenčni distribuciji reprezentira vrednosti v posameznem razredu, dobimo oceno za vsoto vrednosti v tem razredu, če sredino razreda pomnožimo s frekvenco f_k. Vsoto vrednosti za celo populacijo pa dobimo, če produkte $f_k x_k$ za vse razrede seštejemo. Če izkoristimo to lastnost frekvenčnih distribucij, izračunamo oceno aritmetične sredine iz frekvenčne distribucije po obrazcu

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_r x_r}{f_1 + f_2 + \ldots + f_r} = \frac{\Sigma f_k x_k}{\Sigma f_k} = \frac{1}{N} \Sigma f_k x_k \qquad (9.9)$$

Ta obrazec je samo posebna oblika splošnega obrazca 9.8. Ta aritmetična sredina je *tehtana aritmetična sredina*. Pri tem pa so frekvence *teže* ali *ponderi*.

Oceno aritmetične sredine izračunavamo po tem obrazcu v naslednjih točkah:

a) Za frekvenčno distribucijo poiščemo frekvencam f_k ustrezne sredine razredov x_k.

b) Izračunamo produkte frekvenc f_k z ustreznimi sredinami razredov x_k.

c) Če vsoto produktov $\Sigma f_k x_k$ delimo z obsegom populacije N, dobimo oceno aritmetične sredine $\bar{x}$.

Ta ocena je tem boljša, čim manjši so razredi, vendar tudi pri razmeroma velikih razredih dobimo še vedno zadovoljive rezultate, če je frekvenčna distribucija unimodalna in ne preveč asimetrična. Za zelo asimetrične distribucije, posebno za distribucije tipa J, pa daje tako izračunavanje sistematično napačne ocene. Če je distribucija asimetrična v desno, dobimo sistematično prevelike ocene, pri asimetriji v levo pa sistematično premajhne ocene.

Izračunavanje aritmetične sredine po obrazcu 9.9 velja za distribucije z enakimi in z neenakimi razredi. Ne moremo pa po tem obrazcu oceniti aritmetične sredine za frekvenčne distribucije z odprtimi razredi, ker zanje ne moremo izračunati sredine razredov. Edino če za odprt razred poznamo razen frekvence f_r še vsoto vrednosti v odprtem razredu X_r, moremo izračunati oceno aritmetične sredine, če upoštevamo obrazec 9.8, po obrazcu

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_{r-1} x_{r-1} + X_r}{N} \qquad (9.10)$$

9.15 Izračunavanje ocene aritmetične sredine po direktni metodi je v tabeli 9.1 nakazano za frekvenčno distribucijo za porabo lesa v letu 1953 za kmetijska gospodarstva v Novem mestu.

Tabela 9.1 Izračunavanje ocene aritmetične sredine potrošnje lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto v letu 1953 po direktni metodi

Poraba v m^3	f_k	x_k	$f_k x_k$
0,0 - 3,9	2	2	4
4,0 - 7,9	13	6	93
8,0 - 11,9	36	10	360
12,0 - 15,9	41	14	574
16,0 - 19,9	29	18	522
20,0 - 23,9	12	22	264
24,0 - 27,9	7	26	182
28,0 - 31,9	4	30	120
32,0 - 35,9	2	34	68

$N = 149 \quad \Sigma f_k x_k = 2190$

Po obrazcu 9.9 je $\bar{x} = 2190/149 = 14,70\ m^3$.

Če primerjamo dobljeni rezultat s pravo aritmetično sredino $\bar{x} = 2182,1/149 = 14,64\ m^3$, ki smo jo izračunali iz osnovnih podatkov v tabeli 8.1, vidimo, da je razlika minimalna.

9.16 Pomožni znak u. Za frekvenčne distribucije, ki imajo enake razrede, moremo izračunavanje aritmetične sredine poenostaviti, če vpeljemo namesto sredin razredov pomožni znak u_k, ki je z osnovnim znakom x_k v naslednji linearni zvezi:

$$x_k = x_0 + i.u_k \qquad (9.11)$$

Pri tem pomeni: x_0 = sredina razreda, ki je približno v sredini frekvenčne distribucije ali bližu razreda z največjo frekvenco, i = širina razreda.

Če natančneje proučimo vrednosti znaka u, spoznamo, da sredinam razredov v posameznih razredih ustrezajo $u \ldots -3, -2, -1,0, +1, +2, +3 \ldots$ Pri tem je vrednost $u = 0$ v razredu, ki mu ustreza x_0.

Zaradi obrazca 9.11 in stavka 8.8 pa velja

$$\bar{x} = x_0 + i\bar{u} \qquad (9.12)$$

Ta postopek se običajno pokaže priklad nejši kakor direktna metoda, ker poenostavi množenje. Po tem postopku izračunavamo aritmetično sredino

po tehle točkah:

a) V dani frekvenčni distribuciji izberemo nekje v sredini ali v razredu, okoli katerega so frekvence največje, izhodišče pomožnega znaka u. Glede na to poljubno izhodišče postavimo v posamezne razrede ustrezne vrednosti pomožnega znaka u

$$\ldots -3, -2, -1, 0, +1, +2, +3 \ldots$$

b) Pomnožimo frekvence f_k z ustreznimi vrednostmi u_k. Tako dobimo produkte $f_k u_k$.

c) Seštejemo dobljene produkte in vsoto $\Sigma f_k u_k$ vnesemo v obrazec

$$\bar{x} = x_0 + \frac{i}{N} \Sigma f_k u_k \qquad (9.13)$$

9.17 Za porabo lesa v Novem mestu je izračunavanje aritmetične sredine po metodi pomožnega znaka u nakazano v tabeli 9.2.

Tabela 9.2 Izračunavanje ocene aritmetične sredine za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi pomožnega znaka u

Poraba v m^3	f	u	fu
0,0 - 3,9	2	-3	-6
4,0 - 7,9	16	-2	-32
8,0 - 11,9	36	-1	-36
12,0 - 15,9	41	0	0
16,0 - 19,9	29	+1	+29
20,0 - 23,9	12	+2	+24
24,0 - 27,0	7	+3	+21
28,0 - 31,9	4	+4	+16
32,0 - 35,9	2	+5	+10
	$N = 149$		$\Sigma fu = +23$

Ker je $x_0 = 14{,}0$, $i = 4$, $N = 149$ in $\Sigma fu = +23$ dobimo po obrazcu 9.13

$$\bar{x} = 14{,}0 + 4 \frac{+26}{149} = 14{,}70 \text{ m}^3$$

Rezultat se sklada z rezultatom, ki smo ga dobili po direktni metodi.

9.18 Metoda kumulativ. Če so podatki grupirani v frekvenčni distribuciji z enakimi razredi, moremo aritmetično sredino izračunati s kumulativami. Ta postopek je običajno še priklanejši, kakor že vpeljemo po-

možni znak u. Pri tej metodi odpade vmesno množenje. Razen tega pa imamo kot postranski rezultat izračunano kumulativno frekvenčno distribucijo; ta je važna sama zase ali pa za izračunavanje kvantilov.

Po metodi kumulativ izračunamo aritmetično sredino po naslednjem postopku:

a/ Iz frekvenčne distribucije f izračunamo kumulativno frekvenčno distribucijo F.

b/ Seštejemo člene v kumulativni vrsti, razen zadnjega, ki leži pod črto, ki pomeni obseg populacije N. Vsoto členov kumulativne vrste zaznamujemo z A.

c/ Aritmetično sredino $\bar{x}$ izračunamo iz dobljenih izrazov po obrazcu

$$\bar{x} = x_0 - i\,A/N \qquad (9.14)$$

Pri tem pomeni: x_0 = sredina zadnjega razreda v frekvenčni distribuciji, i = širina razreda; A = vsota členov v kumulativni vrsti; N = obseg populacije.

9.19 Aritmetična sredina za porabo lesa v letu 1953 v Novem mestu je izračunana po metodi kumulativ v tabeli 9.3.

Tabela 9.3 Izračunavanje aritmetične sredine za porabo lesa v 149 kmetijskih gospodarstvih v letu 1953 po metodi kumulativ

Poraba v m^3	f	F	
0,0– 3,9	2	0	
4,0– 7,9	16	2	
8,0–11,9	36	18	
12,0–15,9	41	54	
16,0–19,9	29	95	
20,0–24,9	12	124	
24,0–27,9	7	136	
28,0–31,9	4	143	
32,0–35,9	2	147	$2+18+54+\ldots+143+147 = 719 = A$
	$N = 149$		

Ker je sredina zadnjega razreda $x_0 = 34$, dobimo po obrazcu 9.14

$$\bar{x} = 34{,}0 - 4.719/149 = 14{,}70\ m^3$$

Dobljeni rezultat se sklada z rezultatoma, ki smo ju dobili po prejšnjih metodah.

Modificirana aritmetična sredina

9.20 Kakor smo videli, je aritmetična sredina odvisna od vseh vrednosti populacije. Zato nanjo vplivajo tudi ekstremne vrednosti, ki so včasih rezultat izjemnih pogojev in jih ne moremo šteti, da sodijo v prikazano populacijo. Te vrednosti pačijo sliko populacije in tudi aritmetično sredino populacije. Zato za take primere izračunavamo sredino, ki je nekak kompromis med aritmetično sredino in mediano.

Iz ranžirne vrste dobimo mediano, če sukcesivno izpuščamo po en spodnji in en zgornji člen v ranžirni vrsti. Na koncu tega postopka ostane samo en člen - mediana, ali dva člena, katerih sredina je mediana. Kompromisna rešitev pa je v tem, da izpuščamo po en zgornji in spodnji člen le toliko časa, dokler ne izključimo vse netipične vrednosti, iz ostanka pa izračunamo aritmetično sredino. Tako odstranimo vpliv ekstremnih - izjemnih vrednosti, ki včasih občutno vplivajo na aritmetično sredino. Kljub vsemu pa je ta sredina izračunana iz večine vrednosti populacije. Modificirano aritmetično sredino dobimo torej tako, da iz serije individualnih vrednosti odstranimo po en, dva ali več parov skrajnih - izjemnih vrednosti, iz ostalih pa izračunavamo povprečje. Ta postopek uporabljamo na primer pri analizi časovnih serij. Zaradi izjemnih razmer v določenih razdobjih nekateri podatki pačijo tipičnost aritmetične sredine. Če pa te ekstreme ne upoštevamo in uporabimo modificirano povprečje, dobimo realnejšo sliko.

Aritmetična sredina aritmetičnih sredin

9.21 Če poznamo aritmetične sredine $\bar{x}_k$ in obsege N_k za delne populacije, ki sestavljajo populacijo, moremo iz teh podatkov izračunati aritmetično sredino $\bar{x}$ za populacijo po obrazcu

$$\bar{x} = \frac{\Sigma N_k \bar{x}_k}{\Sigma N_k} \qquad (9.15)$$

Ta obrazec smo dokazali v odstavku o lastnostih aritmetičnih sredin. Lastnosti, da moremo izračunati skupno srednjo vrednost, če poznamo ustrezne srednje vrednosti za delne populacije, nimata niti mediana niti modus. Pri tem je treba iz delnih populacij sestaviti skupno populacijo in iz nje poiskati mediano ali modus.

Zgornji način izračunavanja aritmetične sredine imenujemo *tehtan* ali *ponderiran način*, ker je iz delnih sredin izračunana skupna sredina tako, da upoštevamo velikost - *težo* ali *ponder* - za posamezno delno aritmetično sredino. Obrazec 9.15 je poseben primer splošnega obrazca za izračunavanje tehtane aritmetične sredine

$$\bar{r} = \frac{\Sigma w_k r_k}{\Sigma w_k} \qquad (9.13)$$

Pri tem pomeni: $\bar{r}$ = tehtana aritmetična sredina količin r_k, w_k = ponder za posamezne vrednosti r_k.

9.22 V tabeli 9.4 je nakazano izračunavanje skupnih povprečnih mesečnih prejemkov delavcev v poljedelstvu, če poznamo povprečne mesečne prejemke in skupno število delavstva po kvalifikaciji v letu 1953.

Tabela 9.4 Izračunavanje skupnih povprečnih mesečnih prejemkov delavcev v poljedelstvu v FLRJ v septembru letu 1957 (Vir: SB 114)

Kvalifikacija	Povprečni mesečni prejemki v dinarjih $\bar{x}_k$	Število delavcev N_k	$N_k\bar{x}_k$
Visoko kvalificirani	14140	2555	36127700
Kvalificirani	12470	28321	353132870
Priučeni	9280	20515	190379200
Nekvalificirani	7700	78819	606903300
		130210 ΣN_k	1186576070 $\Sigma N_k\bar{x}_k$

$$\bar{x} = \frac{1186576070}{130210} = 9113 \text{ din}$$

Popolnoma nepravilno bi bilo, če bi izračunali navadno aritmetično sredino iz podatkov o mesečnih prejemkih po kvalifikaciji. Rezultat, ki bi ga dobili, $\bar{x}_{nav} = 1/4\ (14140 + 12470 + 9280 + 7700) = 10900$, je seveda različen od zgornjega in brez logičnega smisla, medtem ko pravo sumarno povprečje pove, kakšni bi bili prejemki delavcev v kmetijstvu, če bi imeli vsi enake prejemke pri istem fondu plač.

Sumarna povprečja pa moramo uporabljati zelo previdno. V splošnem sumarne aritmetične sredine ne reprezentirajo vrednosti v populaciji. Sumarno povprečje more biti celo vrednost, ki jo ima malo število enot. Vendar kljub tem hitam dostikrat izračunavamo sumarna povprečja, ki so, če jih pravilno tolmačimo, dobro sredstvo analize.

HARMONIČNA SREDINA

9.23 Med izračunane srednje vrednosti štejemo tudi harmonično sredino, ki je po definiciji recipročna vrednost aritmetične sredine reciprokov iz osnovnih podatkov. Po tej definiciji je harmonična sredina H enaka

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_N}} = \frac{N}{\sum \frac{1}{x_i}} \qquad (9.17)$$

tehtana harmonična sredina pa analogno

$$H' = \frac{w_1 + w_2 + \ldots\ldots + w_r}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots\ldots + \frac{w_r}{x_r}} = \frac{\sum w_k}{\sum \frac{w_k}{x_k}} \qquad (9.18)$$

9.24 Če iz istih podatkov izračunamo aritmetično sredino in harmonično sredino, dobimo različne rezultate. Za podatke 1,2,4,7,9 je aritmetična sredina enaka

$$x = \frac{1 + 2 + 4 + 7 + 9}{5} = 4{,}6$$

harmonična sredina pa

$$H' = \frac{5}{\frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{7} + \frac{1}{9}} = 2{,}5$$

Iz shematičnega primera vidimo, da so razlike znatne.

Harmonično sredino izračunavamo redkeje kakor aritmetično. Kadar se vrednosti populacije distribuirajo v asimetrični distribuciji, a tako, da je distribucija recipročnih vrednosti simetrična, pa harmonična sredina bolje kaže centralno tendenco kakor aritmetična sredina. Zato v takih primerih dajemo prednost harmonični sredini pred aritmetično.

9.25 S harmonično sredino izračunavamo včasih tudi povprečje iz relativnih števil.

Vzemimo, da je pet delavcev delalo po eno uro. Pri tem so dosegli naslednjo proizvodnost dela: 4, 5, 2, 3, 3 minute za en artikel. Proizvodnost dela merimo s časom, ki so ga potrebovali za izdelavo enega artikla.

Sumarni pokazatelj o produktivnosti dela za vseh pet delavcev dobimo, če skupno porabljen čas delimo s številom proizvedenih artiklov. Vseh pet delavcev je delalo skupno 5.60 = 300 minut. Število artiklov, ki jih je proizvedel posamezen delavec, dobimo, če za vsakega delavca čas (60 minut) delimo s časom, ki ga je potreboval za izdelavo enega artikla (pokazatelj produktivnosti dela). Za posamezne delavce dobimo po vrsti:

$60/4 = 15 \; ; \quad 60/5 = 12 \; ; \quad 60/2 = 30 \; ; \quad 60/6 = 10 \; ; \quad 60/3 = 20.$

Skupno število proizvedenih artiklov pa dobimo če te podatke seštejemo. Povprečna produktivnost dela je torej

$$\frac{5.60}{60/4 + 60/5 + 60/2 + 60/6 + 60/3} = \frac{300}{87} = 3{,}45 \text{ minut za artikel}$$

Iz računa vidimo, da smo izračunali povprečno proizvodnost dela s harmonično sredino. Če iz zgornjega računa izpustimo ponder 60, dobimo

$$\frac{5}{1/4 + 1/5 + 1/2 + 1/6 + 1/3} = 3{,}45 \text{ minut za artikel}$$

Ker so ponderi med seboj enaki, dobimo isti rezultat, če vzamemo tehtano ali netehtano obliko.

Če iz zgornjih podatkov izračunamo aritmetično sredino, pa dobimo

$$\bar{x} = \frac{1}{5}(4+5+2+3+3) = 4.$$

Preizkus pokaže da harmonična sredina da pravi rezultat. Vseh pet delavcev, ki so delali skupno 300 minut, bi proizvedlo 87 artiklov. Če bi vsi delavci delali s povprečno proizvodnostjo dela 3,45 minut za artikel bi enako napravili $300/3{,}45 = 87$ artiklov. To pa je v skladu z definicijo povprečja. Če bi vseh pet delavcev delalo z enotno produktivnostjo dela, 4 minute za kos, ki smo jo dobili z aritmetično sredino, pa bi v skupno 300 minutah izdelali le 75 artiklov in ne 87, kolikor so jih v resnici.

Podrobneje o uporabi harmonične sredine pri izračunavanju povprečnih relativnih števil bomo zvedeli v naslednjem odstavku.

POVPREČJA IZ RELATIVNIH ŠTEVIL

9.26 Problem izračunavanja povprečij iz relativnih števil je za socialno-ekonomsko statistiko tako osnoven in važen, da mu bomo posvetili poseben odstavek.

Relativna števila so v splošnem kvocienti primerjanih količin. Če vzamemo, da je grupno relativno število r_k razmerje dveh absolutnih podatkov Y_k in Y_k

$$r_k = X_k / Y_k \ , \qquad (9.19)$$

moremo ta obrazec pisati v več oblikah. Iz njega dobimo, da je

$$X_k = Y_k r_k \qquad (9.20)$$

in

$$Y_k = X_k / r_k \qquad (9.21)$$

Če upoštevamo te zveze, moremo izračunavati sumarno relativno število r na tri načine:

a) Sumarno relativno število je razmerje med vsotami absolutnih grupnih vrednosti za X_k in Y_k.
Po tej definiciji je sumarno relativno število r enako

$$r = \frac{X}{Y} = \frac{\Sigma X_k}{\Sigma Y_k} \qquad (9.22)$$

S tem obrazcem izračunavamo sumarno relativno število, kadar imamo po grupah *absolutne podatke* za X_k in Y_k.

b) Če upoštevamo obrazec 9.20, je sumarno relativno število

$$r = \frac{\Sigma Y_k r_k}{\Sigma Y_k} \qquad (9.23)$$

tehtana aritmetična sredina grupnih relativnih števil r_k. Pri tem so vrednosti, ki so v imenovalcu relativnega števila, ponderi. Tako izračunavamo povprečno relativno število, kadar razpolagamo z grupnimi relativnimi števili in vrednostmi, ki nastopajo v *imenovalcu* relativnih števil, oziroma kadar je *smiseln produkt* ponderа z relativnim številom $Y_k r_k$.

c) Če upoštevamo obrazec 9.21, pa je sumarno relativno število enako

$$r = \frac{\Sigma Y_k}{\Sigma Y_k / r_k} \qquad (9.24)$$

To pa je *tehtana harmonična sredina* grupnih relativnih števil r_k. Pri tem so ponderi količine, ki nastopajo v relativnem številu v števcu. Po tem obrazcu izračunavamo sumarno relativno število, kadar razpolagamo z grupnimi relativnimi vrednostmi r_k in grupnimi količinami X_k, ki nastopajo v relativnem številu v *števcu* oziroma kadar je *smiseln kvocient* ponderа in relativnega števila X_k / r_k.

Obrazci 9.22, 9.23 in 9.24 oziroma zgornja pravila ne veljajo samo za relativna števila v ožjem smislu, temveč tudi za izračunavanje sumar-

nih aritmetičnih sredin, saj so aritmetične sredine tudi relativna števila, ker so razmerja med vsoto vrednosti in številom enot.

9.27 Po zgornjih pravilih ugotovimo, kdaj uporabimo za izračunavanje sumarnih relativnih števil tehtano aritmetično ali harmonično sredino.

Po teh pravilih ugotovimo, da je treba v primeru produktivnosti dela, ki smo ga navedli v odstavku 9.25, izračunati harmonično sredino. Produktivnost dela je merjena z relativnim številom „čas, potrebljen za proizvajanje enega artikla", ki ga dobimo, če potrebljen čas delimo s številom proizvedenih artiklov. Ker razpolagamo s časom, kolikor so posamezni delavci delali (30 minut), so ponderi količine, ki v relativnem številu nastopajo v števcu. Po zgornjih pravilih je treba za sumarno relativno število uporabiti harmonično sredino.

Če poznamo površine in povprečne gostote prebivalstva po republikah, izračunamo gostoto prebivalstva za FLRJ po obrazcu za tehtano aritmetično sredino, ker so ponderi „površine" v koeficientu „gostota prebivalstva" v imenovalcu.

Če imamo po podjetjih dano število zaposlenih žensk in odstotek zaposlenih žensk od skupnega števila zaposlenih, izračunamo povprečen odstotek zaposlenih žensk za celo stroko po obrazcu za izračunavanje harmonične sredine. Relativno število „odstotek zaposlenih žensk" je kvocient med številom zaposlenih žensk in številom skupno zaposlenih v podjetju. Ponder – število zaposlenih žensk je v tem primeru v števcu relativnega števila.

Za podjetje imamo koeficiente o obračanju zalog posameznih vrst surovin (merjeno s časom enega obrata za posamezno vrsto surovine) in ustrezne vrednosti porabljenih surovin. Sumarni koeficient obračanja zalog izračunamo po obrazcu za izračunavanje tehtane aritmetične sredine. Pokazatelj o obračanju zalog „čas enega obrata" je namreč kvocient med povprečnimi zalogami in vrednostjo porabljenih surovin. Ponder-vrednost porabljenih surovin-pa je v relativnem številu v imenovalcu.

Če imamo po okrajih podatke o površini in skupnem pridelku pšenice v določenem letu, izračunamo povprečen hektarski donos pšenice po obrazcu 9.22 tako, da vsoto pridelkov po okrajih delimo z vsoto površin.

Odvisnost sumarnega relativnega števila od strukture

9.28 Če preuredimo obrazec 9.23, dobimo

$$r = \frac{\sum Y_k r_k}{Y} = \sum \frac{Y_k}{Y} \cdot r_k = \sum Y_k^0 r_k \qquad (9.25)$$

Podobno dobimo iz obrazca 9.24

$$r = \frac{X}{\Sigma X_k / r_k} = \frac{1}{\Sigma \frac{X_k}{X} / r_k} = \frac{1}{\Sigma X_k^o / r_k} \qquad (9.26)$$

Pri tem pomeni: $X_k^o = X_k / X$ strukturni delež za grupni podatek X_k; $Y_k^o = Y_k / Y$ strukturni delež za podatek Y_k.

Iz teh dveh obrazcev zaključimo, da sumarno relativno število r ni odvisno samo od grupnih relativnih števil r_k, temveč tudi od strukture ustreznih ponderov X_k^o ali Y_k^o.

To je ena izmed osnovnih hib sumarnih relativnih števil, zaradi katere imamo včasih zelo resne pomisleke o uporabnosti sumarnih pokazateljev. Za dve populaciji, ki imata ista grupna relativna števila, sta sumarni relativni števili med seboj različni, če sta strukturi ponderov različni. Še več. Razlike v strukturi ponderov morejo povzročiti, da je sumarno relativno število za populacijo A večje kot za populacijo B, čeprav so vsa grupna relativna števila za populacijo A manjša kot za populacijo B.

9.29 Vzemimo za primer povprečne mesečne prejemke delavcev v kmetijstvu in gradbeništvu po kvalifikaciji

Tabela 9.5 Povprečni prejemki delavcev v kmetijstvu in gradbeništvu po kvalifikaciji v letu 1957 v FLRJ (Vir SB 114)

Kvalifikacija	Povprečni mesečni prejemki v dinarjih		Struktura števila delavcev	
	kmetijstvo	gradbeništvo	kmetijstvo	gradbeništvo
Visoko kvalificirani	15310	15200	0,020	0,078
Kvalificirani	12430	11920	0,218	0,298
Priučeni	9580	8950	0,157	0,218
Nekvalificirani	8080	6770	0,605	0,403
S k u p n o	9370	9437	1,000	1,000

Iz tabele 9.5 vidimo, da so povprečni prejemki delavcev v kmetijstvu za vse kvalifikacije višji kot v gradbeništvu. Kljub temu pa so sumarni povprečni prejemki v kmetijstvu nižji kot v gradbeništvu. Vzrok je različna struktura števila delavstva v kmetijstvu in gradbeništvu. Iz zadnjih dveh stolpcev v tabeli 9.5 vidimo, da je sestav delavstva po kvalifikaciji v poljedelstvu znatno nižji kot v gradbeništvu. Ta razlika v sestavu delavstva po kvalifikaciji ima tako močan vpliv na sumarno povprečje, da izpadejo povprečni prejemki delavstva v celoti v kmetijstvu nižji kot v gradbeništvu, čeprav je za posamezne kvalifikacije situacija prav obratna.

Standardizirani pokazatelji

9.30 Omenjena lastnost sumarnih pokazateljev je tako značilna, da se vprašamo, ali imajo sumarni pokazatelji sploh analitičen pomen in smisel, ker je primerljivost sumarnih pokazateljev zelo avomiljiva in megljena, če ne vemo, ali izvira razlika iz razlik v grupnih pokazateljih ali iz razlik v strukturi ponderov.

Sumarni koeficient mortalitete je po tem odvisen od umrljivosti po posameznih starostnih skupinah za ženske in moške in spolne in starostne strukture prebivalstva. Prebivalstvo z razmeroma nizko stopnjo umrljivosti po posameznih starostnih skupinah more imeti visok sumarni koeficient umrljivosti, če so v populaciji predvsem stari ljudje. Prebivalstvo z razmeroma visoko stopnjo umrljivosti po posameznih starostnih skupinah pa more imeti nizek sumarni koeficient umrljivosti, če sestoji predvsem iz mladih ljudi.

Podjetje more izkazati visok sumarni koeficient produktivnosti dela, če proizvaja predvsem artikle, za katere je splošna produktivnost dela visoka, čeprav je produktivnost dela za posamezen artikel v podjetju razmeroma nizka in obratno.

Primerljivost tako izračunanih sumarnih pokazateljev motijo razlike v strukturi. Da izločimo vpliv razlik v strukturi, izračunavamo *standardizirane pokazatelje* tako, da vzamemo za vse sumarne pokazatelje, ki jih primerjamo, *stalno - standardno strukturo za ponderje*. Tako je standardiziran koeficient umrljivosti ponderirana sredina specifičnih koeficientov umrljivosti po starosti, pri tem pa vzamemo enotno standardno starostno strukturo za vse države, za katere primerjamo podatke. Ti koeficienti imajo večjo analitično vrednost kot navadni koeficienti umrljivosti, ker razlike med standardiziranimi kkeficienti kažejo samo razliko v umrljivosti, ne pa v starostni in spolni strukturi. Standardna struktura, ki jo vzamemo za osnovo za izračunavanje sumarnih koeficientov, mora biti seveda taka, da čim bolje ustreza stvarnim strukturam za posamezne populacije, katere med seboj primerjamo. Standardna struktura je zato včasih povprečna struktura vseh populacij, ki jih primerjamo, včasih idealna struktura, ki je pogojena z analizo pojava in podobno.

9.31 - Izračunajmo za prejemke delavcev po kvalifikaciji za kmetijstvo in gradbeništvo iz tabele 9.5 standardizirane povprečne prejemke delavcev za obe panogi. Za standardno strukturo vzemimo povprečno strukturo delavstva v FLRJ po kvalifikaciji v vseh panogah.

Ker za izračunavanje sumarnih povprečnih prejemkov vzamemo za standardno strukturo - strukturo delavstva po kvalifikaciji v FLRJ - dobimo v kmetijstvu povprečne mesečne prejemke večje (10573 din) kakor v gradbeništvu (9855 din). Standardizirana povprečja resnično pokažejo, da je raven plač v kmetijstvu višja kot v gradbeništvu, ker razlike v sumarnih prejemkih izvirajo samo iz razlik v nivoju plač po kvalifikacijah, ne pa tudi iz razlik v strukturi. Seveda so standardizirana povprečja različna od povpre-

čij, ki smo jih dobili v tabeli 9.5. Te razlike so rezultat tega, da je stvarna struktura v posamezni panogi različna od povprečne strukture za vse panoge.

Tabela 9.6 Izračunavanje standardiziranih mesečnih prejemkov v kmetijstvu in gradbeništvu v letu 1957 v FLRJ

Kvalifikacija	Povprečni mesečni prejemki v din		Standardna struktura	(1).(3)	(2).(3)
	kmetijstvo	gradbeništvo			
	(1)	(2)	(3)	(4)	(5)
Visoko kvalificirani	15310	15200	0,075	1148,25	1140,00
Kvalificirani	12430	11920	0,365	4536,95	4350,80
Priučeni	9580	8950	0,263	2506,39	2353,85
Nekvalificirani	8030	6770	0,287	2384,91	2010,39
			1,000	10576,50	9855,34

GEOMETRIJSKA SREDINA

9.32 Geometrijska sredina G iz N vrednosti $x_1, x_2, x_3, \ldots x_N$ je N koren iz produkta vseh vrednosti

$$G = \sqrt[N]{x_1 . x_2 . \ldots x_N} \qquad (9.27)$$

Iz te definicije sklepamo, da ima smisel izračunavati geometrijsko sredino le tedaj, če noben izmed členov ni negativen ali nič.

Tehtana geometrijska sredina je za razliko od navadne geometrijske sredine, ki je dana v obrazcu 9.27 enaka

$$G = \sqrt[w]{x_1^{w_1} . x_2^{w_2} . \ldots x_N^{w_N}} \qquad (9.28)$$

$w_1; w_2 \ldots w_N$ = ponderi: $w = \Sigma w_k$

9.33 Direktno izračunavanje geometrijske sredine po zgornjih osnovnih obrazcih je razen za najenostavnejše primere neizvedljivo. Pač pa moremo geometrijsko sredino razmeroma enostavno izračunati z logaritmi. Če namreč obrazca 9.27 in 9.28 logaritmiramo, dobimo

$$\log G = \frac{1}{N}(\log x_1 + \log x_2 + \log x_3 + \ldots + \log x_N) = \frac{1}{N}\Sigma \log x \qquad (9.29)$$

in

$$\log G = \frac{1}{w}(w_1 . \log x_1 + w_2 . \log x_2 + w_3 \log x_3 + \ldots + w_N \log x_N) = \frac{\Sigma w_k \log x_k}{\Sigma w_k} \qquad (9.30)$$

Logaritem navadne geometrijske sredine je enak navadni aritmetični sredini logaritmov osnovnih podatkov, logaritem tehtane geometrijske sredine pa je enak tehtani aritmetični sredini logaritmov osnovnih podatkov. S tema obrazcema posredno izračunavamo geometrijske sredine z logaritmiranjem.

9.34 Zaradi razmeroma zamotanega izračunavanja in težje interpretacije geometrijsko sredino izračunavamo redkeje kot druge vrste sredin. Vendar so problemi, pri katerih je upravičeno edino izračunavanje geometrijske sredine. Tako izračunavamo geometrijsko sredino v nekaterih posebnih problemih individualnih indeksov oziroma relativnih števil na splošno.

Povečanju cene za 100 % (indeks 200) smiselno ne ustreza znižanje cene za 100 %, ker bi v tem primeru bila cena nič, temveč znižanje za 50 % (indeks 50). Kompenzacija raznosmernih učinkov se pokaže pri geometrijski sredini, ker je geometrijska sredina iz indeksa 200 in 50

$$\sqrt{200.50} = 100.$$

Ne pokaže pa se na primer pri aritmetični sredini, ker je aritmetična sredina med 200 in 50 enaka

$$\frac{200 + 50}{2} = 125$$

Zato je za izračunavanje sredin iz individualnih indeksov bolj upravičena geometrijska kot aritmetična sredina.

9.35 Za pet artiklov, za katere so indeksi cen: 80, 100, 135, 67, 150, je srednji indeks enak

$$G = \sqrt[5]{80.100.135.67.150}$$

To količino izračunamo z logaritmi takole.

Tabela 9.7 Izračunavanje geometrijske sredine iz individualnih indeksov

Indeks	log
80	1,90309
100	2,00000
135	2,13033
67	1,82607
150	2,17609
	10,03558:5 = 2,00712

Logaritem srednjega indeksa je $\log G = 2{,}00712$, srednji indeks pa je $G = 100{,}4$.

Aritmetična sredina individualnih indeksov $M = \frac{1}{5}(80+100+135+67+150) = 106{,}4$; to je znatno višji rezultat kakor geometrijska sredina.

9.36 Geometrijsko sredino uporabljamo tudi za izračunavanje srednjega verižnega indeksa oziroma koeficienta dinamike iz časovnih vrst (koeficient dinamike je kvocient dveh zaporednih členov v časovni seriji).

Zaznamuj-mo z $Y_0, Y_1, Y_2, \ldots Y_N$ posamezne člene v časovni vrsti, s $k_1 = Y_1/Y_0$, $k_2 = Y_2/Y_1$, ... $k_N = Y_N/Y_{N-1}$ pa posamezne stvarne koeficiente dinamike. k naj bo srednji koeficient dinamike, s katerim bi se moral pojav v času od 0 do N spreminjati, da bi dosegli isti končni člen Y_N.

Med temi količinami so naslednje zveze

$$Y_0 . k_1 . k_2 \ldots k_N = Y_N = Y_0 k . k \ldots k = Y_0 k^N \qquad (9.31)$$

Če namreč Y_0 postopoma množimo s k_1, k_2 ... k_N, dobimo zaradi pomena koeficientov dinamike postopoma vrednosti členov časovne serije in končno zadnji člen Y_N. Glede na definicijo srednjega koeficienta dinamike k pa dobimo Y_N tudi, če začetno vrednost Y_0 N-krat pomnožimo s k, ali če jo pomnožimo s k^N.

Iz zveze v obrazcu 9.31 moremo izpeljati dva obrazca.

$$k = \sqrt[N]{k_1 . k_2 \ldots k_N} \qquad (9.32)$$

in

$$k = \sqrt[N]{Y_N / Y_0} \qquad (9.33)$$

Srednji koeficient dinamike je po obrazcu 9.32 geometrijska sredina iz posameznih koeficientov dinamike.

Po drugem obrazcu pa izračunamo srednji koeficient dinamike, če poznamo za razmak, za katerega izračunavamo srednji koeficient dinamike, začetno in končno vrednost pojava. Srednji koeficient dinamike je v tem primeru koren iz kvocienta končnega in začetnega člena časovne serije. Stopnja korena pa je določena s časovnim razmakom, za katerega iščemo srednji koeficient dinamike. Zato ni nujno, da je N celo število, ampak more biti kateri koli časovni razmak.

9.37 Pri obravnavanju verižnih indeksov v odstavku o enostavnih indeksih smo izračunali vrsto verižnih indeksov za proizvodnjo električne energije v FLRJ v letih 1951-1963. Verižni indeksi v teh letih so: 105,9; 110,4; 115,4; 123,1; 117,1.

Srednji verižni indeks v tem razdobju je po obrazcu 9.32 enak geometrijski sredini iz individualnih verižnih indeksov

$$I_k = \sqrt{105,9 . 110,4 . 115,4 . 126,1 . 117,1}$$

Z logaritmi izračunamo srednji verižni indeks po obrazcu 9.29

$$\log I = \frac{1}{5}(\log 105{,}9 + \log 110{,}4 + \log 115{,}4 + \log 126{,}1 + \log 117{,}1) =$$

$$= \frac{1}{5}(2{,}02490 + 2{,}04297 + 2{,}03221 + 2{,}10106 + 2{,}03856) = 2{,}05994$$

Z antilogaritmiranjem dobimo, da je srednji verižni indeks $\bar{I} = 114{,}8$

Da bi bilo zvečanje proizvodnje električne energije v razdobju 1951-1956 enako stvarnemu, bi se morala proizvodnja električne energije večati s srednjim verižnim indeksom 114.8. Isti rezultat dobimo, če računamo srednji verižni indeks po obrazcu 10.38 iz začetne in končne proizvodnje električne energije. Po podatkih iz tabele 3.19 je bila proizvodnja električne energije v letu 1951 v FLRJ $Y_{51} = 2550$, v letu 1956 pa $Y_{56} = 5084$ milijonov KWh.

Po obrazcu 10.38 pa je

$$k = \sqrt[5]{\frac{5084}{2550}}$$

Z logaritmiranjem dobimo

$$\log k = \frac{\log 5084 - \log 2550}{5} = \frac{3{,}70621 - 3{,}40654}{5} = 0{,}05993,$$

z antilogaritmiranjem pa

$$k = 1{,}148$$

ODNOSI MED RAZLIČNIMI VRSTAMI SREDNJIH VREDNOSTI

9.38 Če vzamemo ilustrativen primer in iz 1 in 4 izračunamo harmonično H, geometrijsko G in aritmetično M sredino, dobimo, da je

$$H = \frac{2}{1/1 + 1/4} = 1{,}6\,; \quad G = \sqrt{1 \cdot 4} = 2\,; \quad M = \frac{1+4}{2} = 2{,}5$$

Iz teh rezultatov vidimo, da je harmonična sredina, če jo izračunamo iz istih podatkov, najmanjša, aritmetična sredina največja, geometrijska sredina pa leži med obema. To ne velja samo za ta primer, ampak velja splošno pravilo: harmonična sredina je manjša kakor geometrijska, ta pa manjša kakor aritmetična sredina, če jih izračunamo iz istih podatkov.

$$H < G < M \qquad (9.34)$$

9.39 Tudi med vrednostmi modusa Mo, mediane Me in aritmetične sredine M opazimo neke stalne odnose. Za unimodalne simetrične distribucije so vrednosti modusa, mediane in aritmetične sredine enake $(Mo = Me = M)$. Za unimodalne distribucije, ki so asimetrične v desno, je modus manjši, aritmetična sredina pa večja kot mediana $(Mo < Me < M)$. Za unimodalne distribucije, ki so asimetrične v levo, pa se vrstni red zamenja in je aritmetična sredina manjša, modus pa večji kot mediana $(M < Me < Mo)$.

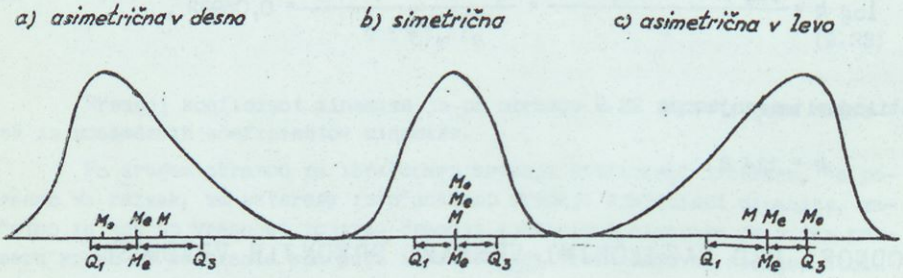
Za zvezne, ne preveč asimetrične distribucije velja, da je razlika med aritmetično sredino in modusom približno trikrat večja kot razlika med aritmetično sredino in mediano.

$$M - Mo = 3.(M - Me) \tag{9.35}$$

Ta obrazec moremo uporabiti za ne preveč asimetrične unimodalne distribucije, da z njim ocenimo modus, čeprav nimamo podatke grupirane v frekvenčni distribuciji. Ker moremo M in Me izračunati tudi iz negrupiranih podatkov, ocenimo vrednost modusa z obrazcem

$$Mo = M - 3.(M - Me) \tag{9.36}$$

Odnosi med modusom, mediano in aritmetično sredino za posamezne vrste distribucij so nakazani v sliki 9.2.



Slika 9.2: Odnosi med Mo, M in Me za različne tipe unimodalnih distribucij

Deseto poglavje

MERE VARIACIJE, ASIMETRIJE IN SPLOŠČENOSTI

MERE VARIACIJE

10.1 Že večkrat smo poudarili, da je za množične pojave značilna *variabilnost*. Tako se od kmetijskega gospodarstva do gospodarstva spreminja velikost, pridelek, število članov, število živine ita; od podjetja do podjetja poraba surovin, število delavcev, vrednost proizvodnje, poraba električne energije, goriva ita; od človeka do človeka starost, spol, stan, dohodek ita.

Medtem ko je srednja vrednost izraz splošnih pogojev, ki so za vso populacijo enaki - nespremenjeni, se zaradi individualnih vplivov posamezne enote od nje odklanjajo navzgor in navzdol. Čim manjši so individualni vplivi, tem manjši so odkloni in tem manjša je variabilnost pojava in obratno: čim močnejši so individualni vplivi, tem večji so odkloni oziroma variabilnost pojava. Variabilnost je izraz delovanja individualnih vplivov, ki morejo biti manjši ali večji. Variabilnost moremo torej meriti.

Vrste mer variacije

10.2 Variabilnost merimo z več parametri. Vsak izmed njih ima drugo osnovo, vendar vsi - vsak po svoje kaže isto lastnost - variabilnost pojava.

Obravnavali bomo naslednje mere variacije:

a/ variacijski razmak R,
b/ kvartilni odklon Q,
c/ povprečen absolutni odklon AD,
d/ standardni odklon SD-σ, oziroma varianco σ^2, iz katere je SD izpeljan.

Od teh mer variacije sta variacijski razmak R in kvartilni odklon Q dana z lego nekaterih členov populacije, povprečni absolutni odklon AD in standardni odklon SD pa sta izračunana iz vseh vrednosti populacije in sta zaradi tega solidnejši meri variabilnosti. Zaradi svojih lastnosti je izmed vseh mer variacije najvažnejši standardni odklon SD.

Variacijski razmak

10.3 Najenostavnejša, a tudi najbolj problematična mera variacije je variacijski razmak

$$R = x_{max} - x_{min} \qquad (10.1)$$

ki je po definiciji razlika med najmanjšo in največjo vrednostjo v populaciji.

Najmanjšo in največjo vrednost zlahka najdemo, če so podatki urejeni v ranžirni vrsti.

Če vzamemo, da je bila cena mleka v glavnih mestih republik v marcu 1958 (Vir: SP 126)

31, 38, 44, 45, 50, 55,

je za te podatke variacijski razmak

$$R = 55 - 31 = 24 \text{ dinarjev.}$$

Včasih variacijski razmak posredno izražamo tudi tako, da navedemo najnižjo in najvišjo vrednost. V našem primeru bi mogli torej reči: Cena mleku v glavnih mestih republik je bila v marcu 1958 med 31 in 55 dinarjev.

Variacijski razmak je nestabilna mera variacije, ker je izračunan iz obeh skrajnih vrednosti populacije, ki moreta biti rezultat nekih izjemnih - netipičnih vplivov. R ni odvisen od razmestitve vrednosti znotraj variacijskega razmaka. Ta more biti najrazličnejša. Vse vrednosti morejo biti v variacijskem intervalu razmeščene enakomerno ali pa se goste okrog nekega središča. V prvem primeru je variabilnost znatno večja kot v drugem, variacijski razmak pa tega ne pokaže, ker sta ekstremni vrednosti isti.

Kljub tem hibam pa variacijski razmak v praksi uporabljamo vselej, kadar hočemo jakost variacije na hitro orientacijsko oceniti in ni časa, niti potrebe za izračunavanje drugih, natančnejših mer variacije. Eden izmed takih primerov je tekoča kontrola kvalitete dela strojev. Variacijski razmak, ki ga določimo iz meritev razmeroma majhnega števila artiklov, rabi za orientacijo jakosti variacije, ki je eden izmed osnovnih pokazateljev kvalitete proizvodnje.

Kvartilni odklon

10.4 Variacijski razmak je slabo merilo variabilnosti zato, ker je odvisen samo od skrajnih vrednosti. To hibo odpravimo tako, da merimo variacijo z razmakom, v katerem je samo del populacije. Tako z razmakom med prvim in devetim decilom omejimo razmak, v katerem je 80 % vseh vrednosti, ker smo izločili po 10 % najmanjših in največjih vrednosti. Ta razmak

$$D_9 - D_1$$

imenujemo *decilni razmak.*

Podobno s prvim in tretjim kvartilom omejimo razmak, v katerem je 50 % vrednosti in iz njega izključimo 25 % najmanjših in 25 % največjih vrednosti.

Običajno pa namesto *kvartilnega razmaka*

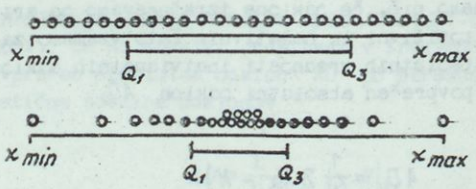
$$Q_3 - Q_1$$

vzamemo za mero variacije *kvartilni odklon*

$$Q = \frac{Q_3 - Q_1}{2} \qquad (10.2)$$

ki je polovica kvartilnega razmaka.
Zgornji meri variacije pravilneje pokažeta jakost variacije kakor variacijski razmak, ker nanju ne vplivajo skrajne - netipične vrednosti.

Od zgornjih mer variacije pogosteje uporabljamo kvartilni kot decilni razmak. V sliki 10.1 je shematično prikazano, kako je variacijski razmak neobčutljiv za razlike v razmestitvi vrednosti znotraj variacijskega razmaka, in kako se različna razmestitev pokaže na kvartilnem razmaku.



Slika 10.1 Variacijski in kvartilni razmak pri različni razmestitvi vrednosti

Določanje zgornjih mer variacije ne dela težav, ker smo v postavku o kvantilih nakazali, kako izračunavamo decile in kvartile.

10.5 Za ceno mleka v glavnih mestih republik v marcu 1958, za katero smo izračunali že variacijski razmak, je kvartilni odklon

$$Q = \frac{Q_3 - Q_1}{2} = \frac{50 - 38}{2} = 3$$

Za množino padavin v juniju 1954 za 28 meteoroloških postaj v Sloveniji smo kvartile izračunali že v postavku o kvantilih

$$(Q_1 = 151{,}3 \text{ mm} ; \quad Q_3 = 221{,}5 \text{ mm}).$$

Iz teh podatkov je kvartilni odklon

$$Q = \frac{221,5 - 151,3}{2} = 35,1 \text{ mm}$$

Za porabo lesa v 149 kmetijskih gospodarstvih v letu 1958 pa je po podatkih iz dodatka o kvantilih (6.9) ($Q_1 = 10,2$ m^3, $Q_3 = 18,4$ m^3) kvartilni odklon enak

$$Q = \frac{18,4 - 10,2}{2} = 4,1 \text{ m}^3.$$

Povprečen absolutni odklon

10.6 Uvodoma smo naznačili, da z odkloni individualnih vrednosti od sredine merimo jakost individualnih vplivov na posamezno enoto. Sumarno merilo jakosti teh vplivov bi moglo biti povprečje individualnih odklonov. To pa je, kakor vemo, enako nič, če odklone izračunavamo od aritmetične sredine, ker so odkloni pozitivni in negativni. Zato vzamemo za merilo variabilnosti povprečje absolutnih vrednosti individualnih odklonov od sredine. Po tej definiciji je povprečen absolutni odklon AD_M

$$AD_M = \frac{1}{N} \Sigma |x - M| \qquad (10.3)$$

Ker je povprečen absoluten odklon najmanjši, če računamo odklone od mediane, izračunavamo AD tudi iz absolutnih odklonov od mediane.

$$AD_{Me} = \frac{1}{N} \Sigma |x - Me| \qquad (10.4)$$

AD_{Me} je vsebinsko zaradi te lastnosti mediane bolj upravičen kot AD_M.

AD izračunamo torej tako, da vsoto absolutnih vrednosti individualnih odklonov od aritmetične sredine ali mediane delimo z obsegom populacije N.

10.7 Za ceno mleka v glavnih mestih republik v marcu 1958 je izračunavanje AD_M in AD_{Me} nakazano v tabeli 10.1.

Ob vsoto absolutnih odklonov $\Sigma |x - M| = 37,34$ in $\Sigma |x - Me| = 37,0$

delimo z $N = 6$, dobimo, da je $AD_M = \frac{37,34}{6} = 6,23$ din in $AD_{Me} = \frac{37,0}{6} =$

$= 6,17$ din.

Podobno se je pokazalo, da je povprečen absolutni odklon, računan od mediane, manjši kot računan od aritmetične sredine.

Tabela 10.1 Izračunavanje povprečnega absolutnega odklona AD_M in AD_{Me} za ceno mleka v glavnih mestih republik v marcu 1958

x	$x-M$	$\|x-M\|$	$x-M_e$	$\|x-M_e\|$
31	- 12,83	12,83	- 13,5	13,5
38	- 5,83	5,83	- 6,5	6,5
44	+ 0,17	0,17	- 0,5	0,5
45	+ 1,17	1,17	+ 0,5	0,5
50	+ 6,17	6,17	+ 5,5	5,5
55	+ 11,17	11,17	+ 10,5	10,5
Σ 263	+ 0,02	37,34	- 4,0	37,0

M = 43,83

Me = 44,50

10.8 Izračunavanje povprečnega absolutnega odklona iz grupiranih podatkov. Če so podatki grupirani v frekvenčni distribuciji, izračunavamo povprečen absoluten odklon AD po obrazcu za izračunavanje tehtane aritmetične sredine odklonov

$$AD = \frac{1}{N} \Sigma f \, |x-M| \qquad (10.5)$$

Podobno kakor pri aritmetični sredini moremo tudi za izračunavanje AD uporabiti pomožni znak u, da si olajšamo izračunavanje, vendar zaradi razmeroma majhne uporabe AD tega postopka ne bomo dalje razvijali.

Varianca. Standardni odklon

10.9 Podobno osnovo kakor povprečen absolutni odklon ima tudi *varianca* σ^2, ki je po definiciji povprečje kvadratov odklonov od aritmetične sredine

$$\sigma^2 = \frac{1}{N} \Sigma (x - M)^2 \qquad (10.6)$$

Pri AD odpravimo predznake odklonov tako, da vzamemo absolutne odklone, pri σ^2 pa tako, da odklone kvadriramo.

Podobno kot AD je tudi σ^2 izračunana mera variacije in je odvisna od vsake izmed vrednosti populacije. Vendar so druge lastnosti variance osnovne važnosti za statistično analizo. Zaradi tega je, enako kakor je aritmetična sredina osnovna srednja vrednost, varianca osnovna mera variacije, čeprav je njeno izračunavanje zaradi kvadriranja razmeroma zamudno.

Ker je varianca povprečje kvadratov odklonov, je izražena v neprikladni enoti mere – kvadratu osnovne enote. Temu se ogнemo tako, da izračunamo kvadratni koren iz variance. To mero variacije imenujemo *standardni odklon* SD ali σ.

Varianca in standardni odklon sta torej v enostavni zvezi

$$SD = \sigma = \sqrt{\sigma^2} \qquad (10.7)$$

10.10 Izračunavanje iz negrupiranih podatkov. Direktna metoda Varianco in standardni odklon moremo po zgornji definiciji oziroma obrazcih izračunavati tako, da:

a) Iz osnovnih podatkov izračunamo aritmetično sredino M.
b) Izračunamo odklone posameznih vrednosti od aritmetične sredine $x - M$.
c) Posamezne odklone kvadriramo $(x - M)^2$.
d) Varianco dobimo, če kvadrate odklonov seštejemo, dobljeno vsoto pa delimo z obsegom populacije N.
e) Kvadratni koren iz variance σ^2 je standardni odklon SD.

10.11 Za ceno mleka v marcu 1958 za glavna mesta republik izračunamo varianco, kakor je nakazano v tabeli 10.2.

Tabela 10.2 Izračunavanje standardnega odklona za ceno mleka v glavnih mestih republik v mesecu marcu 1958 po direktni metodi

x	$x-M$	$(x-M)^2$
31	– 12,83	164,6089
38	– 5,83	33,9889
44	+ 0,17	0,0289
45	+ 1,17	1,3689
50	+ 6,17	38,0689
55	+ 11,17	124,7689
263	+ 0,02	362,8334
X	$\Sigma(x-M)$	$\Sigma(x-M)^2$

$M = 263/6 = 43,83$

Po obrazcu 10.6 dobimo

$$\sigma^2 = 362,8334/6 = 60,4722 ;$$

Standardni odklon SD pa je po obrazcu 10.7 enak

$$SD = \sqrt{60,4722} = 7,73 \text{ dinarjev.}$$

10.12 Metoda pomožnega znaka u. Ker so odkloni od povprečja navadno decimalna števila, to komplicira kvadriranje. Zato je izračunava-

nje variance in standardnega odklona po direktni metodi nepriklicno. S pomožnim znakom u pa moremo v večini primerov s postopkom, ki smo ga spoznali že pri izračunavanju aritmetične sredine, izračunati varianco znatno krajše.

Po tej metodi:

a) Izberemo poljubno okroglo vrednost x_0, tako, da se stvarne vrednosti x od nje čim manj razlikujejo;

b) Izračunamo odklone stvarnih vrednosti od x_0; tako dobimo vrednosti $u = x - x_0$;

c) Kvadriramo posamezne vrednosti u.

d) Seštejemo vrednosti za u in kvadrate u^2. Tako dobimo $\Sigma u = U$ in Σu^2;

e) Iz teh količin izračunamo varianco in standardni odklon po obrazcih

$$K = \Sigma u^2 - U^2/N; \quad \sigma^2 = K/N; \quad SD = \sqrt{\sigma^2} \qquad (10.8)$$

10.13 Če izračunamo po tej metodi varianco za ceno mleka iz zgornjega primera, vidimo, da je skrčenje računa znatno. Po pregledu podatkov sodimo, da je primerna vrednost za računanje vrednosti u, $x_0 = 45$. Glede na to je potek računa nakazan v tabeli 10.3.

Tabela 10.3 Izračunavanje variance za ceno mleka za glavna mesta republik v FLRJ v marcu 1958 po metodi pomožnega znaka u

x	$u = x-45$	u^2
31	-14	196
38	- 7	49
44	- 1	1
45	0	0
50	+ 5	25
55	+10	100
	-7	371
	U	Σu^2

$$K = \Sigma u^2 - U^2/N = 371 - (-7)^2/3 = 362,8333$$

$$\sigma^2 = K/N = 362,8333/3 = 60,4722; \quad \sigma = \sqrt{60,4722} = 7,76 \text{ din.}$$

Dobili smo torej iste rezultate kakor po direktni metodi.

Če se ne moremo odločiti za nobeno vrednost x_0, ali če izračunavamo variance z računskim strojem, vzamemo $x_0 = 0$. Tako odpade izračunavanje u. Varianco v tem primeru izračunamo po obrazcih

$$K = \Sigma x^2 - X^2/N; \quad \sigma^2 = K/N; \quad SD = \sqrt{\sigma^2} \qquad (10.9)$$

10.14 Izračunavanje iz grupiranih podatkov. Direktna metoda. Za velike populacije je izračunavanje variance iz negrupiranih podatkov še zamudnejše kakor izračunavanje aritmetične sredine. Zaradi tega za velike populacije tudi variance ocenjujemo iz podatkov, ki so grupirani v frekvenčne distribucije. Ocena variance iz frekvenčne distribucije je direktno

$$\sigma_x^2 = \frac{1}{N} \Sigma f_k (x_k - \bar{x})^2 \qquad (10.10)$$

tehtana aritmetična sredina kvadratov odklonov sredin razredov x_k od aritmetične sredine $\bar{x}$. Vendar se ta postopek izkaže v primerjavi z drugima postopkoma, s katerima moremo tudi izračunavati varianco iz frekvenčnih distribucij, neprikladen. Zato zanj ne navajamo primera, ker v splošnem te metode ne uporabljamo.

10.15 Metoda pomožnega znaka u. Varianco moremo ocenjevati iz frekvenčne distribucije z metodo pomožnega znaka u, ki je samo razširitev metode pomožnega znaka u za izračunavanje aritmetične sredine iz grupiranih podatkov.

Po tej metodi izračunamo varianco tako, da:

a) Enako kakor za aritmetično sredino izberemo razred, ki je nekje sredi frekvenčne distribucije. Vanj postavimo novo izhodišče znaka $u = 0$, v druge razrede pa navzdol in navzgor od izhodišča ustrezne vrednosti pomožnega znaka u:

$$\ldots -3, -2, -1, 0, +1, +2, +3, \ldots$$

b) Enako kakor za aritmetično sredino, izračunamo produkte frekvenc f z ustreznimi vrednostmi znaka u, da dobimo fu.

c) Produkte fu ponovno pomnožimo z ustreznimi vrednostmi znaka u; tako dobimo vrednosti fu^2.

d) Seštejemo frekvence f, produkte fu in produkte fu^2; Tako dobimo $\Sigma f = N$, $\Sigma fu = U$ in Σfu^2.

e) Iz teh količin izračunamo varianco po obrazcih

$$K = \Sigma fu^2 - U^2/N; \quad \sigma^2 = i^2 K/N; \quad SD = \sqrt{\sigma^2} \qquad (10.11)$$

Pri tem je: i = širina razreda.

10.16 Za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto je prikazan izračun variance po tej metodi v tabeli 10.4.

Tabela 10.4 Izračunavanje variance σ^2 in standardnega odklona σ za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi pomožnega znaka u

Poraba v m^3	f	u	fu	fu^2
0,00 - 3,9	2	-3	- 3	18
4,0 - 7,9	16	-2	-32	64
8,0 - 11,9	36	-1	-36	36
12,0 - 15,9	41	0	0	0
16,0 - 19,9	29	+1	+29	29
20,0 - 23,9	12	+2	+24	48
24,0 - 27,9	7	+3	+21	63
28,0 - 31,9	4	+4	+16	64
32,0 - 35,9	2	+5	+10	50
	149		+23	372
	N		U	Σfu^2

Po obrazcih 10.11 dobimo dalje

$$K = 372 - (+26)^2/149 = 367,4631$$

$$\sigma^2 = 4^2 \cdot 367,4631/149 = 39,4591$$

$$\sigma = \sqrt{39,4591} = 6,28 \text{ m}^3$$

10.17 Metoda kumulativ. Tudi za izračunavanje variance moremo koristno uporabiti lastnosti kumulativnih vrst. Ta metoda je enako kakor metoda pomožnih znakov u samo razširitev iste metode pri izračunavanju aritmetične sredine.

Po metodi kumulativ izračunavamo varianco tako, da:

a) Enako kakor pri izračunavanju aritmetične sredine iz frekvenčne distribucije f izračunamo kumulativno vrsto F. Zadnji člen kumulativne vrste (pod ali nad črto) je obseg populacije N.

b) Po enakem postopku kumuliranja izračunamo iz prve kumulativne vrste F drugo kumulativno vrsto FF. Zadnji člen druge kumulativne vrste FF (pod ali nad črto) je A.

c) Seštejemo vrednosti členov druge kumulativne vrste (brez člena pod ali nad črto). To vsoto zaznamujemo z B.

d) Iz teh količin izračunamo varianco po obrazcih

$$K = 2B + A - A^2/N; \quad \sigma^2 = i^2K/N; \quad SD = \sqrt{\sigma^2} \qquad (10.12)$$

e) Kumulativne vrste moremo izračunavati od zgoraj navzdol ali obratno od

spodaj navzgor. Za asimetrične distribucije je priporočeno začeti izračunavati kumulative na strani asimetrije.

10.18 V tabeli 10.5 je postopek nakazan za porabo lesa v Novem mestu.

Tabela 10.5 Izračunavanje variance za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi kumulativ

Poraba v m^3	f	F	FF	
0,0 – 3,9	2	0		
4,0 – 7,9	13	2	0	
8,0 – 11,9	33	18	2	
12,0 – 15,9	41	54	20	
16,0 – 19,9	29	95	74	
20,0 – 23,9	12	124	139	
24,0 – 27,9	7	133	298	
28,0 – 31,9	4	143	429	
32,0 – 35,9	2	147	572	
		149	719	1559=2+20+...+429+572
		N	A	B

$K = 2.1559 + 719 - 719^2/149 = 337,4331$

Kumuliranje frekvenčne distribucije v obratni smeri je nakazano v tabeli 10.6.

Tabela 10.6 Izračunavanje variance za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi kumulativ

Poraba v m^3	f	F / N 149	FF / A 473	B 398 = 323+195+...+8+2
0,0 – 3,9	2	147	323	
4,0 – 7,9	13	131	195	
8,0 – 11,9	33	95	100	
12,0 – 15,9	41	54	43	
16,0 – 19,9	29	25	21	
20,0 – 23,9	12	13	8	
24,0 – 27,9	7	3	2	
28,0 – 31,9	4	2	0	
32,0 – 35,9	2	0		

$K = 2.398 + 473 - 473^2/149 = 337,4331$

Ne glede na smer kumuliranja smo po metodi kumulativ dobili enake rezultate za K kot pri metodi pomožnega znaka u. Zgornja postopka sta

pokazala, da je v našem primeru kumuliranje od spodaj navzgor prikladnejše, ker da manjše vmesne rezultate. Kot smo omenili zgoraj, je to zaradi tega, ker je distribucija asimetrična v desno.

10.19 Sheppardov popravek. Varianca, ki jo izračunamo po zgornjih postopkih, je samo ocena prave vrednosti variance, ki bi jo dobili, če bi jo izračunali iz individualnih podatkov. Analiza te ocene pa pokaže, da dobimo za unimodalne, ne preveč asimetrične distribucije za zvezne znake po tej metodi sistematično prevelike vrednosti za varianco. Ta napaka je tem večja, čim večji so razredi. Varianco, izračunano po prejšnjih metodah, moremo popraviti po obrazcu

$$\sigma^2_{cor} = \sigma^2 - i^2/12 \qquad (10.13)$$

tako, da od prvotne ocene variance, ki jo dobimo po metodi znaka u ali po metodi kumulativ, odštejemo dvanajstino kvadrata širine razreda.

Ta popravek imenujemo *Sheppardov popravek.*

Za frekvenčno distribucijo o potrošnji lesa v Novem mestu v letu 1953 je popravljena varianca enaka

$$\sigma^2_{cor} = 39,4591 - 4^2/12 = 38,1258$$

$$\sigma_{cor} = \sqrt{38,1258} = 6,17 \text{ m}^3$$

Skupna varianca

10.20 Enako kakor sumarno aritmetično sredino, moremo tudi skupno varianco izračunati iz grupnih podatkov.

Če poznamo po grupah število enot N_k, aritmetične sredine M_k in grupne variance σ^2_k, izračunamo skupno varianco za celo populacijo po obrazcu

$$\sigma^2 = M_{\sigma^2} + \sigma^2_M \qquad (10.14)$$

Pri tem je

$$M_{\sigma^2} = \frac{1}{N} \Sigma N_k \sigma^2_k \qquad (10.15)$$

tehtana aritmetična sredina grupnih varianc,

$$\sigma^2_M = \frac{1}{N} \Sigma N_k (M_k - M)^2 \qquad (10.16)$$

pa tehtana varianca med grupnimi aritmetičnimi sredinami M_k.

10.21 Vzemimo kot primer stroške za kulturno in družbeno življenje v novembru 1957 za 36 delavskih družin v Mariboru. V tabeli 10.7 so navedeni N_k, M_k in σ_k^2 po višini dohodkov.

Tabela 10.7 Podatki za N_k, M_k in σ_k^2 za stroške za kulturno in družbeno življenje za 36 delavskih družin v novembru 1957 v Mariboru po grupah dohodkov

Višina dohodkov v tisoč din	N_k	M_k	σ_k^2
15 - 19	12	479	28858
20 - 24	10	670	35820
25 - 29	7	949	48212
30 - 34	4	1317	83462
35 -	3	1703	8089
	36	819	
	N	M	

Iz teh podatkov dobimo

$$\sigma_M^2 = \frac{1}{36}\,[12(479-819)^2 + \ldots + 3(1703-819)^2] = 140709$$

$$M_{\sigma^2} = \frac{1}{36}\,(12.28858 + \ldots + 3.8089) = 38726$$

$$\sigma^2 = \sigma_M^2 + M_{\sigma^2} = 38723 + 140709 = 179435$$

10.22 Pomembnejše kakor to, da moremo izračunati skupno varianco, če poznamo grupne podatke, pa je dejstvo, da moremo z zgornjim postopkom skupno varianco razstaviti po grupnem znaku v dva dela. Skupna varianca namreč meri vpliv vseh individualnih faktorjev, medtem ko meri σ_M^2 vpliv grupnega znaka, M_{σ^2} pa vpliv drugih individualnih faktorjev na variabilnost pojava.

Po zgornjem postopku moremo torej varianco analizirati v tem smislu, da ugotovimo, koliko od skupne variance je rezultat faktorja, ki smo ga vzeli za grupni znak.

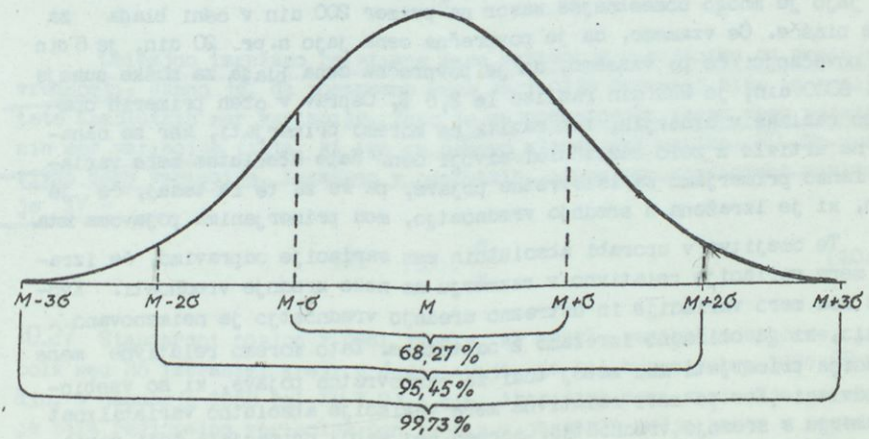
V primeru stroškov za kulturno in družbeno življenje za 36 družin v novembru 1957 v Mariboru vidimo, da je velik del (78 %) skupne variance stroškov za kulturno in družbeno življenje pojasnjene z višino dohodkov. Višina dohodkov je torej eden izmed bistvenih faktorjev, ki vplivajo na stroške za kulturno in družbeno življenje.

10.23 Zveza standardnega odklona z normalno distribucijo.

Variacijski razmak ali kvartilni razmak imata enostavno tolmačenje in pomen. Enako je tudi povprečen absoluten odklon razmeroma razumljiv. Teže pa je s standardnim odklonom, dokler ga ne obravnavamo v zvezi z nekaterimi lastnostmi distribucij. Standardni odklon namreč dobimo z razmeroma zapletenim računskim postopkom, ki zamegli jasno predstavo o smislu tega parametra.

Čeprav normalno distribucijo podrobneje obravnavamo kasneje, je prav, da navedemo nekaj njenih lastnosti v zvezi s standardnim odklonom. Kakor smo navedli že pri obravnavanju frekvenčnih distribucij, je normalna distribucija ena izmed osnovnih distribucij. Pomembna je po svojih *teoretičnih* lastnostih in praktičnem pomenu. Razen tega se v stvarnosti pod določenimi pogoji zelo pogosto pojavljajo distribucije, ki so po svojih značilnostih normalni več ali manj podobne, ker so simetrične, unimodalne in zvonaste oblike.

Za vse normalne distribucije na splošno velja, da leži
v razmaku $M - \sigma$ do $M + \sigma$ 38,27 % ali okrogio 2/3 vseh vrednosti,
v razmaku $M - 2\sigma$ do $M + 2\sigma$ 95,45 % ali približno 95 % vseh vrednosti populacije,
v razmaku $M - 3\sigma$ do $M + 3\sigma$ pa 99,73 % ali praktično vse vrednosti.
Odkloni od aritmetične sredine, ki so večji od σ, se v splošnem pojavijo v tretjini vseh primerov, odkloni, ki so večji kot 2σ, samo v 5 % vseh primerov, odkloni, ki so večji kot 3σ, pa so zelo zelo redki. Ti odnosi veljajo strogo za normalno distribucijo. Približno pa veljajo tudi za druge unimodalne, simetrične in zvonaste distribucije.



Slika 10.2 Frekvence za normalno distribucijo v razmakih $M \pm \sigma$, $M \pm 2\sigma$ in $M \pm 3\sigma$.

Koliko veljajo zgornje zakonitosti za nenormalne distribucije, ki pa so njej podobne, si moremo ustvariti sliko iz naše distribucije o porabi lesa. Zanjo dobimo, da leži v razmaku $M - \sigma$ do $M + \sigma$ to je od 8,53 m^3 do 20,87 m^3 70 % vseh primerov, v razmaku $M - 2\sigma$ do $M + 2\sigma$, to je od 2,33 m^3 do 27,04 m^3 94 % vseh primerov, v razmaku $M - 3\sigma$ do $M + 3\sigma$, to je od 0 do 33.21 m^3 pa 99.3 % vseh primerov. Razlike od odstotkov, ki jih pričakujemo pri normalni distribuciji, so torej neznatne.

Razmerje Q, AD in SD za normalno distribucijo

10.24 Za normalne distribucije so med Q, AD in SD stalni odnosi, ki jih navajamo zaradi tega, ker iz njih lahko sklepamo na približne odnose za druge simetrične, zvonaste in unimodalne distribucije.

Za normalno distribucijo velja:

$$Q = 0{,}6745\ \sigma \approx 2/3\ \sigma \qquad (10.17)$$

$$AD = 0{,}7979\ \sigma \approx 4/5\ \sigma \qquad (10.18)$$

Relativne mere variacije

10.25 Absolutne mere variacije, ki smo jih spoznali v prejšnjih odstavkih, moremo med seboj primerjati le redko. Nemogoče je primerjati absolutne mere variabilnosti za raznovrstne pojave, čeprav so v vsebinski zvezi. Ne moremo, recimo, primerjati variabilnost porabe alkoholnih pijač na prebivalca z variabilnostjo za porabo cigaret na prebivalca ali z variabilnostjo za porabo kruha. Mere variacije vsakega izmed teh pojavov imajo različno enoto mere in zato med seboj niso primerljive. Absolutnih mer variacij pa ne moremo primerjati tudi pri istovrstnih pojavih. 5 din razlike v ceni jajc je mnogo pomembnejše kakor na primer 200 din v ceni blaga za moške plašče. Če vzamemo, da je povprečna cena jajc n.pr. 20 din, je 5 din 25 % povečanje: če pa vzamemo, da je povprečna cena blaga za moške suknje n.pr. 8000 din, je 200 din razlike le 2,5 %. Čeprav v obeh primerih opazujemo razlike v dinarjih, teh razlik ne moremo primerjati, ker se nanašajo na artikle z zelo različnimi nivoji cen. Zato absolutne mere variacije lahko primerjamo za istovrstne pojave, pa še za te le tedaj, če je raven, ki je izražena s srednjo vrednostjo, med primerjanima pojavoma ista.

Te omejitve v uporabi absolutnih mer variacije odpravimo, če izražamo mere variacije relativno v razmerju do neke srednje vrednosti. Kvocient med mero variacije in ustrezno srednjo vrednostjo je neimenovano število, ki ga običajno izražamo z odstotkom. Zato moremo relativne mere variacije primerjati med seboj tudi za raznovrstne pojave, ki so vsebinsko odvisni. Ker pa meri relativna mera variacije absolutno variabilnost v razmerju s srednjo vrednostjo, moremo med seboj primerjati tudi varia-

bilnosti podatkov na različnih nivojih.

10.26 Relativne mere variacije moremo izračunati iz vseh vrst absolutnih mer variacije: R, Q, AD in SD. Vsako izmed teh primerjamo z ustrezno srednjo vrednostjo. Tako variacijski razmak R primerjamo običajno z aritmetično sredino med skrajnima vrednostima $\frac{1}{2}(x_{min} + x_{max})$, kvartilni odklon Q z mediano ali aritmetično sredino med Q_1 in Q_3, povprečen absoluten odklon AD z mediano ali aritmetično sredino, standardni odklon SD pa z aritmetično sredino M.

Tako dobimo naslednje relativne mere variacije:
Na osnovi variacijskega razmaka R:

$$\frac{2(x_{max} - x_{min})}{x_{max} + x_{min}} \qquad (10.19)$$

Na osnovi kvartilnega odklona Q:

$$\frac{Q}{Me} \quad \text{ali} \quad \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad (10.20)$$

Na osnovi povprečnega absolutnega odklona AD:

$$\frac{AD_{Me}}{Me} \quad \text{ali} \quad \frac{AD_M}{M} \qquad (10.21)$$

Na osnovi standardnega odklona SD:

$$KV^0 = \frac{\sigma}{M} \qquad (10.22)$$

Običajno izražamo relativne mere variacije v odstotku od srednje vrednosti. Jasno je, da posamezne mere variacije ohranijo hibe in kvalitete absolutnih mer variacije. Zato je najpomembnejša izmed vseh relativnih mer variacije tista, ki ima za osnovo standardni odklon SD. To relativno mero variacije, izraženo v odstotkih, imenujemo koeficient variacije $KV\,\%$

$$KV\% = 100 \cdot \frac{\sigma}{M} \qquad (10.23)$$

10.27 Standardni odklon v ceni jabolk, ki pokaže variabilnost cene jabolk med 35 izbranimi kraji v Jugoslaviji, je bil v septembru 1957 SD=13,2 din, v januarju 1958 pa SD = 27,4 din. Absolutna mera variacije pokaže, da je bila regionalna variacija cen jabolk v januarju večja kakor v septembru. Če pa računamo, da je bila povprečna cena jabolk v septembru 1957 M=40,4 din, v januarju 1958 pa M = 113,3 din, sklepamo, da je bila variabilnost cene

jabolk v septembru pomembnejša kakor v januarju, ker je koeficient variacije cen $KV\%$ v septembru 1957

$$KV\% = 100 \cdot \frac{13,2}{40,4} = 32,7\%$$

v januarju 1958 pa

$$KV\% = 100 \frac{27,4}{113,6} = 24,1\ \%$$

V septembru je standardni odklon približno tretjino od povprečja, v januarju pa samo približno četrtino.

Še bolj drastično vidimo pomen relativne mere variacije v regionalnem variiranju cen za več različnih artiklov.

Tabela 10.8 Koeficienti variacije za pet artiklov široke porabe med 35 izbranimi kraji v Jugoslaviji v januarju 1958 (Vir: Po podatkih v SB 126)

Naziv artikla	v dinarjih		$KV\%$
	M	SD	
Enoten kruh kg	47,8	2,5	5,3
Svinjska mast uvožena kg	325,1	9,8	8,0
Jabolka kg	113,6	27,4	24,1
Jajca kos	17,8	8,8	21,3
Prava kava kg	2086,0	54,2	2,6

V tabeli 10.8 vidimo, da je standardni odklon za ceno prave kave SD = = 54,2 din: ta je od vseh navedenih artiklov največji, stvarno pa najmanj pomemben, ker znaša komaj $KV\%$ = 2,6 % od povprečne cene. Iz tabele vidimo, da sta koeficienta variacije za jabolka in jajca sorazmerno izredno velika. Vzrok temu je tipično lokalni značaj teh artiklov, če jih primerjamo z drugimi tremi.

MERE ASIMETRIJE IN SPLOŠČENOSTI

10.28 Za frekvenčne distribucije je razen centralne tendence in variabilnosti značilna tudi asimetrija in sploščenost. Distribucije so *simetrične, asimetrične v levo* ali *asimetrične v desno*. Na nekaterih opazimo manjšo, na drugih večjo stopnjo asimetrije.

Razen tega distribucije z isto srednjo vrednostjo, isto mero varia-

bilnosti in stopnjo asimetrije, niso nujno med seboj enake, marveč so ene bolj, druge manj *sploščene*. Sploščenost normalne distribucije imamo za idealno. Z njo primerjamo sploščenost drugih distribucij.

Mere asimetrije

10.29 Že pri srednjih vrednostih smo navedli, da so za simetrične, unimodalne distribucije vrednosti modusa, mediane in aritmetične sredine enake. ($Mo=Me=M$). Za unimodalne distribucije, ki so asimetrične v desno, je modus manjši, aritmetična sredina pa večja kakor mediana ($Mo<Me<M$). Za unimodalne distribucije, ki so asimetrične v levo, pa se vrstni red zamenja in je aritmetična sredina manjša, modus pa večji kot mediana ($M<Me<Mo$).

10.30 Za zvezne, ne preveč asimetrične distribucije, velja, da je razlika med aritmetično sredino in modusom približno trikrat večja kakor razlika med aritmetično sredino in mediano.

$$M - Mo = 3(M - Me) \tag{10.24}$$

Pri isti variabilnosti so razlike med M, Me in Mo tem večje, čim večja je stopnja asimetrije. To lastnost frekvenčnih distribucij izkoriščamo in merimo asimetrijo s koeficientom asimetrije, ki ima za osnovo razliko med aritmetično sredino in modusom

$$KA_{Mo} = \frac{M - Mo}{\sigma} \tag{10.25}$$

ali s koeficientom asimetrije, ki ima za osnovo razliko med aritmetično sredino in mediano

$$KA_{Me} = \frac{3(M - Me)}{\sigma} \tag{10.26}$$

10.31 Za simetrične distribucije sta tretji in prvi kvartil enako oddaljena od mediane. Pri asimetričnih distribucijah pa je tretji kvartil bolj oddaljen od mediane, če je distribucija asimetrična v desno, in prvi kvartil bolj oddaljen od mediane, če je distribucija asimetrična v levo. Pri isti variabilnosti so razlike tem večje, čim večja je stopnja asimetrije. To lastnost distribucij izkorišča koeficient asimetrije, ki ima za osnovo razliko v oddaljenosti tretjega in prvega kvartila od mediane.

$$KA_Q = \frac{(Q_3-Me)-(Me-Q_1)}{Q_3-Q_1} = \frac{Q_3+Q_1-2Me}{Q_3-Q_1} \tag{10.27}$$

Vsi trije koeficienti, KA_{Mo}, KA_{Me} in KA_Q so pozitivni, če je distribucija asimetrična v desno, enaki nič za simetrične distribucije, in negativni, če so distribucije asimetrične v levo. Absolutno pa so koeficienti tem večji, čim večja je stopnja asimetrije.

10.32 Za frekvenčno distribucijo o porabi lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto v letu 1953 smo v prejšnjih odstavkih dobili: $Q_1 = 10,19$; $Q_3 = 18,38$; $Mo = 13,17$; $Me = 14,05$; $M = 14,70$ in $SD = 6,17$.

Iz ten podatkov dobimo po zgornjih obrazcih koeficiente asimetrije

$$KA_{Mo} = \frac{14,70-13,17}{3,17} = +0,248$$

$$KA_{Me} = \frac{3(14,70-14,05)}{6,17} = +0,313$$

$$KA_Q = \frac{18,38+10,19-2.14,05}{18,38-10,19} = +0,057$$

Vsi trije koeficienti asimetrije pokažejo, da je distribucija o porabi lesa asimetrična v desno. Ker prva dva koeficienta običajno ležita med -1 in $+1$, moreta pa ležati v mejah med -3 do $+3$, menimo, da stopnja asimetrije ni zelo velika.

MERA SPLOŠČENOSTI

10.33 Za mero sploščenosti vzamemo razmerje med kvartilnim in decilnim razmakom

$$KS = 1,9\,\frac{Q_3 - Q_1}{D_9 - D_1} \qquad (10.28)$$

Koeficient sploščenosti KS je enak ena, če je distribucija normalna, za koničaste distribucije je $KS < 1$, za sploščene pa je $KS > 1$.

Za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu smo dobili, da je $Q_1 = 10,19$; $Q_3 = 18,38$; $D_1 = 7,33$; $D_9 = 23,58$. Iz teh podatkov dobimo, da je

$$KS = 1,9\,\frac{18,38 - 10,19}{23,58 - 7,33} = 0,96$$

Distribucija je po sploščenosti zelo podobna normalni, ker je koeficient KS zelo blizu 1.

Enajsto poglavje

INDEKSI

11.1 V poglavju o relativnih številih smo kot eno izmed treh vrst relativnih števil obravnavali tudi enostavne indekse. Z enostavnimi indeksi primerjamo spremembo določene količine v času, kraju in podobno. Tako z enostavnimi indeksi primerjamo za določen artikel ceni v dveh različnih momentih, porabo v dveh različnih razdobjih, gostoto prebivalstva za dve republiki itd. Zelo važni pa so v ekonomski statistiki indeksi cen.

Povprečni indeks cen

11.2 Problem o primerjanju cen je precej enostaven, če proučujemo spremembe v cenah za posamezen artikel. Vendar enostavni indeksi v praksi ne zadovoljijo vseh potreb po proučevanju sprememb pri ekonomskih pojavih. Merjenje in proučevanje sprememb cen za posamezne artikle se izkaže za zelo neprikladno, ker je njih število veliko in je proučevanje sprememb individualnih cen zaradi velikega števila razdrobljenih podatkov nepregledno in neučinkovito. V praktičnih primerih je zadosti, če poznamo, kako se spreminja raven cen, ne pa, kakšne so spremembe cen za posamezne artikle. To nas navede na misel, da prikažemo spremembe v ravni cen za skupino artiklov s povprečnim indeksom, ki ga izračunamo iz individualnih indeksov cen za vsak artikel. Povprečni indeks iz individualnih indeksov moremo izraziti z obrazcem

$$\bar{I} = \frac{1}{N} \sum \frac{p_1}{p_0} \qquad (11.1)$$

Pri tem pomeni:

p_0 = cena v bazičnem trenutku, p_1 = cena v tekočem trenutku, N = število artiklov, $\bar{I}$ = povprečni indeks.

Povprečni indeks, izračunan po obrazcu 11.1, pa je razmeroma slabo merilo sprememb. Njegova hiba je v tem, da v njem ne upoštevamo važnost posameznega artikla. Važnost za posamezen artikel je dana s proizvodnjo, če gre za indeks cen pri proizvajalcu, ali s porabo, če gre za cene v trgovini na malo. Iz vsakdanje prakse vemo, da so spremembe v ceni za artikle, ki jih množično porabljamo (n.pr. moka, kruh, mleko, meso itd.), za potrošnika veliko pomembnejše, kot spremembe za cene artiklov, za katere je poraba majhna (n.pr. televizijski aparati, avtomobili, različni specialni električni aparati itd.). Zato povprečne indekse po obrazcu 11.1 razmeroma redko izračunavamo, pač pa namesto njih uporabljamo agregatne indekse, ki upoštevajo različno težo za posamezne artikle.

AGREGATNI INDEKSI CEN

11.3 Če proučimo skupno vrednost določene skupnosti različnih artiklov, ugotovimo, da je skupna vrednost vsota produktov količin in cen za posamezne artikle.

$$V = \Sigma p . q \qquad (11.2)$$

Pri tem pomeni: V = skupna vrednost, p = cena za posamezen artikel, q = količina za posamezen artikel.

Če vzamemo, da pomenijo q količine, ki so jih prodali v trgovini na drobno, p pa ustrezajoče cene v trgovini na drobno, pomeni V skupen promet v trgovini na drobno. Če pa so na primer q količine, proizvedene v določenem podjetju, p pa ustrezajoče cene pri proizvajalcu, pomeni V skupno vrednost gotove proizvodnje za to podjetje ita.

Če z 0, 1, 2, 3, ... zaznamujemo, da podatki veljajo za posamezna različna razdobja 0, 1, 2, 3, ..., moremo skupno vrednost za posamezno razdobje pisati

$$V_0 = \Sigma p_0 q_0; \quad V_1 = \Sigma p_1 q_1; \quad V_2 = \Sigma p_2 q_2; \quad V_3 = \Sigma p_3 q_3; \quad \ldots \qquad (11.3)$$

Iz te vrste nazorno vidimo, da so spremembe v skupni vrednosti rezultat sprememb cen in količin v posameznih razdobjih. Indeks vrednosti, ki ga iz ten podatkov izračunamo,

$$I_V = \frac{V_1}{V_0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \qquad (11.4)$$

je torej izraz spremembe cen in spremembe količin.

11.4 Indeks vrednosti izraža samo razlike v spremembi cen le, če so količine v vseh razdobjih enake. To pa stvarno niso. Tako poraba kakor proizvodnja se časovno spreminjata in ne moremo pričakovati, da so količine časovno stalne. Stalnost količin pa moremo predpostaviti in izračunati vrednosti, v katerih vzamemo količine q za vsa razdobja enaka. Tako dobimo vrednosti

$$V_0' = \Sigma p_0 q; \quad V_1' = \Sigma p_1 q; \quad V_2' = \Sigma p_2 q; \quad V_3' = \Sigma p_3 q \qquad (11.5)$$

Indeks, izračunan iz teh vrednosti

$$I_p = \frac{\Sigma p_1 q}{\Sigma p_0 q} \qquad (11.6)$$

dejansko ni več indeks vrednosti, temveč indeks cen, ker kaže le spremembe cen.

Če ta obrazec delno preuredimo

$$I_L = \frac{\Sigma p_0 q \cdot \frac{p_1}{p_0}}{\Sigma p_0 q} \qquad (11.7)$$

spoznamo, da je zgornji indeks stvarno povprečje iz individualnih indeksov cen p_1/p_0; pri tem pa so vrednosti za posamezne artikle po cenah v bazičnem razdobju $p_0 q$ ponderi.

Indeks cen, izračunan po obrazcu 11.7 kaže spremembe v ravni cen posredno prek sprememb v skupni vrednosti, pri predpostavki, da se količine ne menjajo. Tako izračunavanje indeksov imenujemo agregatno izračunavanje indeksov, ker ta indeks pokaže spremembe v cenah prek sprememb v agregatu-skupni vrednosti.

Izbira ponderov pri izračunavanju agregatnega indeksa cen

Stalne količine q v obrazcu 11.3 za izračunavanje agregatnega indeksa cen morejo biti teoretično katere koli količine. Vsebinsko pa je jasno, da indeks cen tem bolje prikaže resnične odnose, čim bolj količine, ki jih vzamemo za ponderе, ustrezajo stvarnosti. Zato po možnosti vzamemo za izračunavanje agregatnega indeksa cen količine, ki so čim bolj podobne stvarni porabi ali proizvodnji v proučevanem razdobju. Pri tem pa imamo različne možnosti.

11.5 Laspeyresov obrazec. Pogosto jemljemo pri izračunavanju agregatnega indeksa cen za ponderе količine iz bazičnega razdobja

$$q = q_0 \qquad (11.8)$$

Tako dobimo iz obrazca 11.3 konkreten obrazec za izračunavanje agregatnega indeksa cen

$$L_p = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \qquad (11.9)$$

Metodo, po kateri pri izračunavanju agregatnih indeksov vzamemo za ponderе stvarne podatke iz bazičnega razdobja, imenujemo Laspeyresovo metodo za izračunavanje agregatnih indeksov, obrazec pa Laspeyresov obrazec za izračunavanje agregatnega indeksa.

Ta obrazec ima prednost, da so ponderi za kateri koli razdobje isti. Ta prednost, ki je nekatere druge metode nimajo, daje Laspeyresovi metodi pri praktičnem izračunavanju agregatnih indeksov posebno ceno.

Če uporabimo za izračunavanje indeksa cen Laspeyresov obrazec, da indeks pristrane rezultate. Ta pristranost izvira iz tega, ker sta cena in prodana ali porabljena količina med seboj odvisni. Če vzamemo, da je cena nekega artikla v tekočem razdobju večja kakor v bazičnem $p_1 > p_0$, ima ta podražitev za rezultat zmanjšanje porabe tega artikla: $q_1 < q_0$. Obratno znižana cena $p_1 < p_0$ navadno poveča porabo $q_1 > q_0$. Ker v Laspeyresovem obrazcu vzamemo za ponderе količine iz bazičnega razdobja, so ponderi pri artiklih, ki so se podražili, preveliki, pri artiklih, ki so se pocenili, pa premajhni. Skupen učinek teh razlik je sistematična pristranost navzgor, ker so ponderi za artikle, ki se podražijo, preveliki, ponderi za artikle, ki se pocenijo, pa premajhni.

11.6 Vzemimo kot primer za izračunavanje agregatnega indeksa cen po Laspeyresovem obrazcu indeks cen mesa v razdobju 1954–1957 v LRS. V obračun vzemimo artikle, ki jih v skupini „surovo in prekajeno meso" vsebuje splošni indeks cen na malo za Slovenijo. Izključili smo le kokoš, ker je zanjo ponder minimalen. Podatki so vzeti iz censke in trgovinske statistike ZS LR Slovenije.

Tabela 11.1 Izračunavanje agregatnega indeksa cen za surovo in prekajeno meso v splošnem indeksu cen na malo v LR Sloveniji po Laspeyresovem obrazcu (cene v dinarjih, količine v tisočih tonah)

Artikel	q_0	1954		1955		1956		1957	
		p_0	p_0q_0	p_1	p_1q_0	p_2	p_2q_0	p_3	p_3q_0
Surovo meso s kostmi:									
Goveje	11,6	223	2586,8	250	2900,0	243	2818,8	251	2911,6
Telečje	2,9	271	785,9	304	881,6	296	858,4	305	884,5
Svinjsko	3,7	249	921,3	292	1080,4	291	1076,7	351	1298,7
Ovčje	0,2	166	33,2	192	38,4	180	36,0	202	40,4
Prekajeno svinjsko meso	2,1	320	672,0	334	764,4	375	787,5	402	844,2
Prekajena slanina	1,2	323	387,6	351	421,2	374	448,8	396	475,2
			5386,8		6086,0		6026,2		6454,6
			Σp_0q_0		Σp_1q_0		Σp_2q_0		Σp_3q_0
Indeks:			100,0	$L_1=100\frac{6086,0}{5386,8}$		$L_2=100\frac{6026,2}{5386,8}$		$L_3=100\frac{6454,6}{5386,8}$	
				= 113,0		= 111,9		= 119,8	

11.7 Paaschejev obrazec. Po Paaschejevi metodi pa vzamemo za ponderе pri indeksu cen tekoče količine q_1. Po tem osnutku izračunavamo indeks cen po obrazcu

$$P_p = \frac{\Sigma\, p_1 q_1}{\Sigma\, p_0 q_1} \qquad (11.10)$$

Če pogledamo obrazec 11.10 , vidimo, da je treba po njem za primera za primera izračunavati tako števce kot imenovalce indeksa, medtem ko pri Laspeyresovi metodi imenovalec $\Sigma\, p_0 q_0$ ostane pri vseh indeksih isti. Razen tega moramo pri Paaschejevi metodi poznati pondere - količine za vsa tekoča razdobja, za katera izračunavamo indekse. S praktičnega vidika je torej izračunavanje po Paaschejevi metodi bolj zamotano kakor po Laspeyresovi. Prednost Paaschejeve metode pa je v tem, da je indeks bližji stvarnim razmeram kot indeks, izračunan po Laspeyresovi metodi. Vendar imamo pri indeksih cen po Paaschejevem obrazcu iz istega razloga kot pri Laspeyresovem obrazcu pristranost z obratnim učinkom, ker stoji umetno sestavljena količina $(\Sigma\, p_0 q_1)$ v imenovalcu indeksa. Zato pričakujemo, da da Paaschejeva metoda premajhne rezultate.

11.8 Če izračunamo za skupino surovo in prekajeno meso indeks cen še po Paaschejevi metodi, dobimo rezultate, ki so dani v tabeli 11.2

Primerjava indeksov cen, ki smo jih dobili po Laspeyresu, z ustreznimi indeksi cen, ki smo jih dobili po Paascheju, kaže, da so resnično indeksi cen, izračunani z Laspeyresovim obrazcem višji kot indeksi cen, izračunani s Paaschejevim obrazcem. To pa je skladno z učinkom, ki izvira iz odvisnosti potrošene količine od cene.

11.9 Fischerjev idealni obrazec. Da omilimo pri indeksih cen učinke, ki izvirajo iz tega, da pri Laspeyresovi formuli vzamemo za pondere $q = q_0$ bazične, pri Paaschejevi pa $q = q_1$ tekoče količine, vzamemo včasih za pondere povprečje iz tekočih in bazičnih količin $q = 1/2(q_0 + q_1)$, včasih pa povprečne količine v celotnem razdobju, za katerega proučujemo gibanje cen.

Ker pa daje Laspeyresova formula sistematično prevelike, Paaschejeva pa sistematično premajhne indekse cen, menimo, da je pravi indeks nekje med obema. Zato vzamemo za pravilnejšo vrednost srednjo vrednost med L_p in P_p. Zaradi drugih kvalitet, ki jih ima tako izračunavanje indeksov cen, je Fisher vzel v ta namen geometrijsko sredino med L_p in P_p

$$F_p = \sqrt{L_p \cdot P_p} = \sqrt{\frac{\Sigma\, p_1 q_0}{\Sigma\, p_0 q_0} \cdot \frac{\Sigma\, p_1 q_1}{\Sigma\, p_0 q_1}} \qquad (11.11)$$

kot pravo vrednost indeksa cen, in jo imenoval idealno. Zato imenujemo ta obrazec *Fischerjev idealni obrazec.*

Indeks, izračunan s Fischerjevim idealnim obrazcem, nima istega logičnega smisla kot Laspeyresov ali Paaschejev, je pa nekak kompromis med

Tabela 11.2 Izračunavanje agregatnega indeksa cen za surovo in prekajeno meso v letih 1954-1957 po Paaschejevi metodi (cene v dinarjih, količine v tisoč tonah)

Artikel	1954	1955				1956				1957			
	p_0	p_1	q_1	p_0q_1	p_1q_1	p_2	q_2	p_0q_2	p_2q_2	p_3	q_3	p_0q_3	p_3q_3
Surovo meso s kostmi:													
Goveje	223	250	12,2	2720,6	3050,0	243	13,4	2988,2	3256,2	251	14,5	3233,5	3639,5
Teleče	271	304	3,8	1029,8	1155,2	296	5,2	1409,2	1539,2	305	5,8	1571,8	1769,0
Svinjsko	249	292	4,0	996,0	1168,0	291	3,7	921,3	1076,7	351	3,7	921,3	1298,7
Ovčje	166	192	0,2	33,2	38,4	180	0,2	33,2	36,0	202	0,2	33,2	40,4
Prekajeno svinjsko meso	320	364	2,6	832,0	946,4	375	2,5	800,0	937,5	402	2,6	832,0	1045,2
Prekajena slanina	323	351	1,5	484,5	526,5	374	1,5	484,5	561,0	396	1,5	484,5	594,0
				6096,1	6884,5			6636,4	7406,6			7073,3	8383,8
				Σp_0q_1	Σp_1q_1			Σp_0q_2	Σp_2q_2			Σp_0q_3	Σp_3q_3
		$P_1 = 100 \frac{6884,5}{3096,1} = 112,9$				$P_2 = 100 \frac{7406,3}{6636,4} = 111,3$				$P_3 = 100 \frac{8386,8}{7076,3} = 118,5$			

Tabela 11.3 Izračunavanje skupinskih indeksov cen za surovo in prekajeno meso v razdobju 1954-1957 v Sloveniji po metodi verižnih indeksov

Artikel	1954	1955				1956				1957			
	p_0	p_1	q_0	p_0q_0	p_1q_0	p_2	q_1	p_1q_1	p_2q_1	p_3	q_2	p_2q_2	p_3q_2
Surovo meso s kostmi:													
Goveje	223	250	11,6	2586,8	2900,0	243	12,2	3050,0	2964,3	251	13,4	3256,2	3363,4
Telečje	271	304	2,9	785,9	881,6	296	3,8	1155,2	1124,8	305	5,2	1589,2	1586,0
Svinjsko	249	292	3,7	921,3	1080,4	291	4,0	1168,0	1164,0	351	3,7	1076,7	1298,7
Ovčje	166	192	0,2	33,2	38,4	180	0,2	38,4	36,0	202	0,2	33,0	40,4
Prekajeno svinjsko meso	320	323	2,1	672,0	764,4	375	2,6	943,4	975,0	402	2,5	937,5	1005,0
Prekajena slanina	323	351	1,2	387,3	421,2	374	1,5	523,5	561,0	393	1,5	561,0	594,0
				5386,8	6086,0			6884,5	3793,0			7406,6	7887,5
				Σp_0q_0	Σp_1q_0			Σp_1q_1	Σp_2q_1			Σp_2q_2	Σp_3q_2
	$I_1 = 100 \frac{6086,0}{5383,8} = 113,0$					$I_2 = 100 \frac{6793,0}{6884,5} = 98,7$				$I_3 = 100 \frac{7887,5}{7403,3} = 106,5$			
	$I_{1/0} = 113,0$					$I_{2/0} = 113,0 \cdot 98,7 = 111,5$				$I_{3/0} = 111,5 \cdot 106,5 = 118,7$			

obema. V tabeli 11.4 so prikazane indeksne vrste za primer skupinskega indeksa za surovo in prekajeno meso, izračunane po vseh treh obrazcih: Laspeyresovem, Paaschejevem in Fisherjevem. Iz tabele vidimo, da je učinek ponderacije tem večji, čim večje je razdobje med bazičnim in tekočim razdobjem.

Verižni indeks cen

11.10 Vpliv ponderacije na indeks je tem večji, čim daljše je razdobje med bazo ali osnovo indeksa in tekočim časom. Zato je pristranost verižnih indeksov razmeroma majhna, ker se osnova razlikuje od tekočega časa le za enoto časa. Iz verižnih indeksov pa moremo izračunati indekse s stalno osnovo, s postopnim množenjem verižnih indeksov

$$I_{1/0} = I_1;\quad I_{2/0} = I_{1/0}.I_2;\quad I_{3/0} = I_{2/0}.I_3 \ldots \qquad (11.12)$$

Čeprav nakazani postopek velja sicer le za enostavne indekse, ga moremo uporabiti kot dober približek tudi za agregatne indekse.

Prek verižnih indeksov izračunamo indeksno vrsto s stalno osnovo v naslednjih stopnjah:

a) Najprej izračunamo po Laspeyresovem obrazcu vrsto verižnih indeksov

$$I_1 = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0};\quad I_2 = \frac{\Sigma p_2 q_1}{\Sigma p_1 q_1};\quad I_3 = \frac{\Sigma p_3 q_2}{\Sigma p_2 q_2} \ldots \qquad (11.13)$$

b) Iz vrste verižnih indeksov dobimo indeksno vrsto s stalno osnovo tako, da verižne indekse postopoma kumulativno množimo:

$$I_{1/0} = I_1;\quad I_{2/0} = I_{1/0}.I_2;\quad I_{3/0} = I_{2/0}.I_3 \text{ itd.} \qquad (11.14)$$

Pri tem pomenijo $I_{1/0}$, $I_{2/0}$, $I_{3/0}$... indekse s stalno osnovo 0.

11.11 Za primer skupinskega indeksa cen za surovo in prekajeno meso, za katerega smo že izračunali indeksno vrsto po drugih metodah, je v tabeli 11.3 nakazano, kako izračunavamo indeksne vrste z verižnimi indeksi.

11.12 V tabeli 11.4 so prikazane indeksne vrste za skupinski indeks cen za surovo in prekajeno meso v razdobju 1954–1957 v LR Sloveniji.

Iz tabele 11.4 vidimo, da je Laspeyersov indeks sistematično nad Paaschejevim, kar je v skladu s pričakovanjem. Indeksna vrsta, ki smo jo dobili z metodo verižnih indeksov, pa se najbolj prilagaja Paaschejevim indeksom. V splošnem razlike niso velike, in sicer zato, ker je štiriletno razdobje razmeroma kratko in strukturne razlike v porabi niso velike.

Tabela 11.4 Primerjalna tabela indeksnih vrst za skupinski indeks cen za surovo in prekajeno meso, izračunane po različnih postopkih

Vrsta indeksa	1954	1955	1956	1957
Laspeyresov indeks	100,0	113,0	111,9	119,8
Paaschejev indeks	100,0	112,9	111,6	118,5
Fisherjev idealni indeks	100,0	112,9	111,8	119,1
Verižni indeks	100,0	113,0	111,5	118,7

Nadomeščanje artiklov in spremembe v ponderaciji

11.13 Pri izračunavanju agregatnih indeksov za daljša razdobja dostikrat naletimo na težave zaradi sprememb, ki nastanejo v porabi s spremembo standarda, navad v porabi itd. Posledica tega je sprememba količin enega ali drugega artikla ali celo opustitev enega artikla in uvajanje drugega artikla v široko porabo. Ta proces je posebno hiter za tehnične artikle. Zato stalna ponderacija, ki jo imamo pri Laspeyresovem agregatnem indeksu cen, postaja v časovnem indeksu bolj in bolj nestvarna. Če pa nek artikel nadomesti drugega ali če se kvaliteta bistveno menja, indeksa niti ne moremo izračunati, ker nimamo cen za tekoče razdobje za vse artikle, ki so v osnovi vzeti kot ponder. Za artikle, ki se pojavijo v tekočem razdobju na novo, pa nimamo v osnovnem razdobju pondera, ker takrat še ni bil predmet široke porabe.

Vprašanje spremembe v ponderacijskem sistemu in nadomestilo določenih artiklov z novimi artikli rešimo s postopkom, ki je zelo podoben izračunavanju verižnega indeksa. Za enostavne indekse velja obrazec

$$I_{2/0} = I_{1/0} \cdot I_{2/1} \qquad (11.15)$$

Glej obrazec 6.11).

Če z 0 zaznamujemo bazično razdobje, z 1 razdobje, za katero spremenimo ponderacijski sistem, z 2 pa tekoče razdobje in prenesemo veljavnost obrazca (11.15) na agregatne indekse, dobimo obrazec

$$I_{2/0} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_2 q_1}{\Sigma p_1 q_1} \qquad (11.16)$$

Z njim rešimo problem spremembe ponderacije in nadomestila artiklov z drugimi.

11.14 Za primer izračunavanja agregatnega indeksa cen pri spremembi pon-

deracijskega sistema vzemimo shematični primer v tabeli 11.5.

Tabela 11.5 Shematičen primer, kako izračunamo agregatni indeks cen, če spremenimo ponderacijski sistem in nadomestimo artikle z novimi artikli

Artikel	p_0	q_0	p_1	q_1	p_2	p_0q_0	p_1q_0	p_1q_1	p_2q_1
A	5	12	6	18	8	60	72	108	144
B	6	5	5	4	7	30	25	20	28
C	6	7	7	7	8	42	49	49	56
D	9	2	10	-	-	18	20	-	-
E	-	-	4	3	3	-	-	24	36
						150	166	201	264
						Σp_0q_0	Σp_1q_0	Σp_1q_1	Σp_2q_1

$$I_{1/0} = 100\frac{166}{150} = 110,7; \quad I_{2/1} = 100\frac{264}{201} = 131,5$$

$$I_{2/0} = I_{1/0} \cdot I_{2/1} = 110,7 \cdot 131,5 = 145,2$$

0 = bazično razdobje, 1 = razdobje, v katerem smo spremenili ponderacijski sistem, 2 = tekoče razdobje.

Iz primera vidimo, da smo v razdobju 1 nadomestili artikel D z artiklom E. Razen tega se je spremenila v tem razdobju celotna ponderacija: to vidimo, če primerjamo q_0 in q_1.

Sprememba ponderacijskega sistema po metodi, ki smo jo nakazali, je eden izmed načinov popravljanja indeksnih vrst cen, če je razdobje med osnovo in tekočim razdobjem dolgo. Vendar včasih tudi tak postopek ne odpravi vseh hib, ki jih ima dolgoročna vrsta indeksov. Indeksi so vedno bolj problematični po računski plati. Razen tega indeksi, računani za zelo dolga razdobja, bolj in bolj izgubljajo tudi vsebinski smisel.

Reprezentativni indeksi cen

11.5 Agregatni indeks cen, v katerega vključimo vse elemente določenega pojava, je na primer agregatni indeks cen pri proizvajalcu za določeno podjetje, ki ne izdeluje veliko različnih artiklov.

V splošnem pa je takih primerov malo. Največ pojavov, ki jih obravnavamo z agregatnimi indeksi, sestoji iz toliko individualnih elementov, da jih je tehnično nemogoče v celoti zajeti. V splošni indeks cen na drobno ne moremo vključiti vseh artiklov, ki so v maloprodaji, ker je število teh artiklov veliko. Dinamiko cen za take primere ne opazujemo z agregatnim in-

deksom, ki je sestavljen iz vseh artiklov v maloprodaji, ampak z *reprezentativnim indeksom cen* v trgovini na malo. Ta pa je sestavljen iz podatkov o cenah za nekatere artikle. Izkaže se namreč, da moremo s takim reprezentativnim indeksom zadovoljivo spremljati dinamiko cen za cele skupine predmetov. Gibanje cen za skupine sorodnih artiklov kaže namreč isti razvoj. Sprememba cene artiklov v eni skupini ima običajno neki skupni vzrok. Sprememba v ceni surovine ali v plačah delavcev v določeni stroki povzroči spremembo v ceni za vse artikle, ki jih proizvajajo v tej stroki itd.

Pri reprezentativnem indeksu cen vzamemo iz vsake grupe samo določeno število predmetov, ki so reprezentanti te skupine. Gibanje cen za te predmete imamo značilno za gibanje cen vseh artiklov v grupi.

Da je agregatni indeks, v katerega ne vključimo vseh elementov določenega pojava, zares reprezentativen indeks za celo grupo, moramo paziti na naslednje:

a) Če hočemo sestaviti reprezentativni agregatni indeks cen za raznovrstno populacijo predmetov, predhodno zaružimo predmete v več grup sorodnih predmetov.

b) Iz vsakega homogenega dela izberemo nekaj predmetov. Gibanje cen za te izbrane reprezentante skupine imamo za gibanje ravni cen za vse artikle skupine. Za reprezentante skupine običajno izberemo take predmete, katerih proizvodnja ali poraba pomeni velik del proizvodnje ali porabe za skupino. Včasih izberemo tudi najznačilnejše predmete, ki so po svoji naravi taki, da njihove cene izražajo gibanje cen v skupini. Če izberemo iz posamezne skupine veliko artiklov, jih običajno izberemo po metodi slučajnega izbora, da indeks zadosti predpisom vzorčenja.

c) Iz podatkov o cenah izbranih artiklov, ki so reprezentanti skupine, izračunamo skupinski indeks: ta more biti bodisi agregatni indeks ali navadni povprečni indeks. Če vzamemo za osnovo pri izračunavanju indeksa grupe agregatni indeks, vzamemo za pondere po možnosti količine prometa ali proizvodnje za vse artikle grupe, ki jo izbrani artikel reprezentira in ne samo promet ali proizvodnjo za ta artikel.

d) Iz skupinskih indeksov sestavimo skupni indeks kot tehtani agregatni indeks. Pri tem pa vzamemo za ponder za posamezno grupo vrednost cele grupe in ne samo vrednost artiklov, ki so reprezentanti grupe.

Po tem principu so na splošno sestavljeni vsi agregatni indeksi cen.

11.16 Indeks življenjskih stroškov. Posebna oblika indeksa cen je indeks življenjskih stroškov. Ta indeks ni splošni indeks cen v trgovini na malo, marveč je po svojem sestavu in namenu posebna oblika indeksa cen na malo. Indeks življenjskih stroškov ima namen spremljati gibanje življenjskih stroškov v posameznih skupinah prebivalstva. Nemogoče je spremljati življenjske stroške vseh družin ali gospodinjstev za določeno socialno skupino. Statistično proučujemo probleme iz tega področja tako, da

anketiramo določene družine in zanje proučujemo spremembe v potrošnji in strukturi porabe. Take ankete dajejo stvarno sliko o porabi in stvarne življenjske stroške. Drug instrument za proučevanje standarda pa je indeks življenjskih stroškov. Pri indeksu življenjskih stroškov sestavimo skupnost predmetov široke porabe, v količini, za katero mislimo, da je za določen tip gospodinjstev pri določenem standardu potrebna. To skupnost predmetov imenujemo *potrošni koš*. Ta tako imenovana teoretska lista je sestavljena skladno s stvarnimi potrebami. Čeprav je teoretična, vendar vsebuje vse osnovne artikle, ki so potrebni in običajni za življenje na določeni stopnji standarda.

Teoretska lista je osnova za izračunavanje indeksa življenjskih stroškov. Od meseca do meseca ali od leta do leta izračunavamo na osnovi stalne teoretske liste, kako se menjajo življenjski stroški, če ostane standard oziroma potrošni koš stalen. Če izračunamo, kolikšni so stroški za preživljanje, ako hočemo zadostiti porabi po teoretski listi, dobimo absoluten iznos: ta je fiktiven in ga ne uporabljamo za analizo. Indeksi, ki jih izračunamo iz teh absolutnih - fiktivnih stroškov, pa zelo dobro kažejo, kako se gibije raven stvarnih življenjskih stroškov, če je potrošni koš po svoji strukturi in izbiri artiklov primeren.

11.17 V tabeli 11.6 je nakazan primer, kako je izračunan indeks življenjskih stroškov za štiričlansko delavsko družino za Slovenijo za november 1958. Za osem osnovnih skupin stroškov so vpisani stroški, skupinski indeksi in skupni indeks življenjskih stroškov. Za osnovo smo vzeli povprečne cene v letu 1957.

Tabela 11.6 Indeks življenjskih stroškov za november 1958 za štiričlansko delavsko družino v Sloveniji (podatki ZS LR Slovenije)

	Mesečno Ø 1957	November 1958	XI.1958 Ø 1957
	stroški v dinarjih		indeks
S k u p n i stroški	25987	26630	102
Hrana	14055	13823	98
Pijače in tobak	1922	2276	118
Obleka in obutev	4200	4327	103
Najemnina	1136	1136	100
Kurjava in razsvetljava	1237	1299	105
Pohištvo	947	966	102
Storitve	1511	1776	118
Kulturne potrebe	979	1027	105

V tabeli 11.7 je za ilustracijo, kako izračunavamo in kakšna je struktura indeksa življenjskih stroškov, nakazan podroben izračun indeksa življenjskih stroškov za prehrano. Količine, ki so vzete za osnovo izraču-

navanja, so po izbiri in količini enotne za vso Jugoslavijo.

Zanimivo je v indeksu stroškov za hrano rešen problem mesa, tolšč, povrtnine in sadja, ki so sezonskega značaja. Stalne artikle za celo leto je za te skupine težko, če ne celo nemogoče določiti. Zato je v indeksu za te grupe predpisana le skupna količina mesa, tolšč, povrtnine in sadja na splošno. Glede na posebnosti sezone pa za vsak mesec po dodatni listi izračunavamo povprečno ceno za vsako izmed ten skupin iz artiklov, ki so v tem mesecu naprodaj. Tako se veliko bolj približamo stvarnosti, kakor pa da iščemo za te skupine neke reprezentante, ki se prodajajo vse leto.

Za povrtnino je bila izračunana povprečna cena v novembru po shemi v tabeli 11.8. V tabeli 11.8 je prikazana lista vseh artiklov, ki pridejo v poštev v katerem koli mesecu v letu. Od ten so v novembru vzeti

Tabela 11.7 Indeks življenjskih stroškov v Sloveniji Izračunavanje skupinskega indeksa za hrano v mesecu novembru 1958
(Podatki ZS - LR Slovenije)

v din

Artikel	Enota mere	Količina mesečne porabe	Ø 1957		November 1958		Indeks
			cene	stroški	cene	stroški	
Kruh pšenični polbel	kg	50	49,7	2483	53	2650	107
Moka	"	4	56	224	58	232	104
Riž	"	0,5	160	80	211	106	132
Meso	"	10	286,1	2861	310,3	3103	108
Tolšče	"	5,5	357,4	1966	347,3	1910	97
Maslo surovo	"	0,25	608	152	630	158	104
Sir mehak (sirček)	"	2	116	232	127	254	109
Mleko	l	30	31,9	956	32,1	963	101
Jajca	kos	24	18	433	22,9	550	127
Fižol v zrnju	kg	3	71	212	73	219	103
Krompir	"	20	16,5	331	13,1	232	79
Povrtnina	"	25	78,8	1970	62,5	1563	79
Sadje presno	"	12	79,5	954	46,8	562	59
Slive suhe	"	0,5	304	152	241	121	80
Sladkor	"	4	145	580	170	680	117
Marmelada	"	0,5	145	73	157	79	108
Čokolada za kuhanje	"	0,1	88,6	84	811	81	96
Sol drobna	"	1	38,2	38	43,5	44	116
Kavin nad. „Divka"	zav.	2	35	70	36,3	73	104
Kava prava	kg	0,1	2039	204	2125	213	104
Σ				14055		13823	98

v poštev le tisti, ki so v tem mesecu na trgu. Povprečna cena je ponderirana sredina; pri tem pa so ponderi vzeti skladno z dejansko porabo.

Tabela 11.8 Izračunavanje povprečne skupinske cene za povrtnino za indeks življenjskih stroškov za mesec november 1957

Artikel	Ponder w	Cena p	Vrednost wp
Čebula	0,16	59	9,4
Korenje	0,16	38,6	3,2
Zelje - presno	0,56	16,0	9,0
Zelje - kislo	-	-	-
Paradižnik - vkuhan	0,12	313,0	37,9
Paradižnik - presen	-	-	-
Paprika zelena	-	-	-
Špinača	-	-	-
Grah zelen v stročju	-	-	-
Fižol v stročju	-	-	-
	1,00		62,5

Povprečna cena povrtnine v novembru je za indeks življenjskih stroškov 32,5 din. Ta povprečna cena je vnešena za ceno povrtnine v indeks življenjskih stroškov za prehrano.

AGREGATNI INDEKSI KOLIČIN

11.18 Doslej smo obravnavali le agregatne indekse cen. Vendar enako - le da pomen cen in količin zamenjamo - izračunavamo tudi agregatne indekse količin, tako da spreminjamo količine in vzamemo cene za stalne. Tako dobimo, analogno indeksom cen, indeks količin po *Laspeyresovem obrazcu*, če vzamemo v agregatnem indeksu

$$L_q = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \qquad (11.17)$$

za ponderе cene v osnovnem razdobju in *Paaschejev obrazec* za izračunavanje indeksa količin, če v obrazcu vzamemo

$$P_q = \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0} \qquad (11.18)$$

za ponderе cen iz tekočega razdobja.

Enako je geometrijska sredina med Laspeyresovim in Paaschejevim obrazcem *Fisherjev idealni obrazec* za izračunavanje agregatnega indeksa

količin

$$F_q = \sqrt{L_q \cdot P_q} = \sqrt{\frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_1 q_0}} \qquad (11.19)$$

Z agregatnimi indeksi za količine v praksi izračunavamo fizični obseg proizvodnje v industrijskih podjetjih in računamo vrednost proizvodnje po stalnih, bazičnih cenah. Enako kot so pri izračunavanju indeksa cen sčasoma stalne količine, ki jih vzamemo za ponderе, problematične, so pri indeksih obsega proizvodnje za daljša razdobja problematični ponderi tudi stare, stalne cene. V bistvu se pri izračunavanju indeksov količin borimo z enakimi težavami kot pri agregatnih indeksih cen.

11.19 Če izračunamo za primer indeks obsega potrošnje za porabo mesa v Sloveniji v tabelah 11.1 in 11.2, je indeks obsega porabe v letu 1957, če vzamemo za osnovo leto 1954

po Laspeyresovem obrazcu $$L_q = \frac{\sum p_0 q_3}{\sum p_0 q_0} = \frac{7076.3}{5386.8} = 131.3$$

po Paascheјevem obrazcu $$P_q = \frac{\sum p_3 q_3}{\sum p_3 q_0} = \frac{8386.8}{3454.6} = 129.9$$

po Fisherjevem obrazcu $$F_q = \sqrt{L_q \cdot P_q} = \sqrt{131.3 \cdot 129.9} = 131.1$$

11.20 Ko izračunavamo indeks obsega proizvodnje ali obseg prometa v trgovini na drobno ali na debelo, naletimo dostikrat na hude težave. Izračunavanje reprezentativnih indeksov obsega nima istega logičnega ozadja kakor reprezentativni indeksi cen. Zato smo navezani le na kompletno izračunavanje in preračunavanje vrednosti po stalnih cenah. To pa je sila težavno, če agregat sestavlja veliko število različnih predmetov. Dostikrat si pomagamo pri ocenjevanju vrednosti proizvodnje, prometa ali porabe po enotnih stalnih cenah z enostavnim sklepom. Če je stvarna vrednost proizvodnje, porabe ali prometa v tekočem razdobju $V_1 = \sum p_1 q_1$, moremo vrednost proizvodnje *reducirati na cene v bazičnem razdobju* $V_1' = \sum p_0 q_1$, če jo delimo z ustreznim reprezentativnim indeksom cen

$$V_1' = \sum p_0 q_1 \approx \frac{V_1}{P_r'} \quad ; \quad P_r' = \frac{\sum p_1' q_1'}{\sum p_0' q_0'} \qquad (11.20)$$

Pri cenah in količinah so napravljene črtice, ker so to cene in količine za reprezentativne artikle te grupe, ne pa za vse artikle grupe, kakor je to v izrazih $\sum p_1 q_1$ in $\sum p_0 q_1$.

Če je dana vrednost po tekočih cenah, moremo torej oceniti vrednost po bazičnih cenah, če stvarno vrednost $V_1 = \Sigma p_1 q_1$ delimo z ustreznim indeksom cen $P'_{1/0}$, vrednost prometa v trgovini na drobno z indeksom cen v trgovini na drobno, promet v trgovini na debelo z indeksom cen v trgovini na debelo itd. Če ocenjujemo vrednost samo za določeno skupino, reduciramo vrednost z ustreznim grupnim indeksom cen itd.

Testi o agregatnih indeksih

11.21 Test o zamenljivosti časa. Če v individualnem indeksu cen $i_{1/0} = p_1/p_0$ zamenjamo osnovo in tekoče razdobje $i_{0/1} = p_0/p_1$, je produkt obeh indeksov

$$i_{1/0} \cdot i_{0/1} = \frac{p_1}{p_0} \cdot \frac{p_0}{p_1} = 1 \qquad (11.21)$$

enak ena.

Te lastnosti pa nimajo vsi agregatni indeksi in je posebna odlika le nekaterih vrst agregatnih indeksov. Če vzamemo n.pr. Laspeyresov obrazec za izračunavanje agregatnega indeksa cen, zanj ta zveza ne velja, ker je

$$L_{1/0} \cdot L_{0/1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \neq 1 \qquad (11.22)$$

Zgornja zveza ne velja niti za Paaschejev obrazec cen niti količin. Pač pa velja za Fisherjev obrazec. Tako dobimo, da je za indeks cen po Fisherjevem obrazcu

$$F_{1/0} \cdot F_{0/1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \cdot \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = 1 \qquad (11.23)$$

11.22 Test o zamenljivosti količin. Če v individualnem indeksu cen $i_p = p_1/p_0$ zamenjamo simbol za ceno p s simbolom za količino q, dobimo tak indeks količin $i_q = q_1/q_0$, da je produkt obeh enak indeksu vrednosti $I_V = V_1/V_0$

$$i_p \cdot i_q = \frac{p_1}{p_0} \cdot \frac{q_1}{q_0} = \frac{V_1}{V_0} = i_V \qquad (11.24)$$

Te lastnosti Laspeyresov indeks cen nima. Če zamenjamo faktorje in pomnožimo oba indeksa, dobimo

$$L_p \cdot L_q = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \neq \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = I_V \qquad (11.25)$$

Te lastnosti tudi drugi agregatni indeksi nimajo, razen Fisherjevega idealnega obrazca. Zanj velja

$$F_p \cdot F_q = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = I_V \qquad (11.26)$$

Zaradi teh dveh lastnosti, ki enačijo operacije agregatnih indeksov z operacijami, ki jih moremo izvesti z individualnimi indeksi, imajo Fisherjev obrazec za idealen.

Drugi agregatni indeksi

11.23 Z agregatnimi indeksi ne obravnavamo samo gibanja cen in obsega proizvodnje. Z njimi proučujemo dinamiko najrazličnejših drugih ekonomskih pojavov. Praktično z agregatnimi indeksi proučujemo dinamiko vseh pojavov, ki se dajo numerično prikazati v obliki agregata Σpq kot vsota produktov dveh količin.

11.24 Eden izmed važnih problemov, ki ga rešujemo z agregatnimi indeksi, je indeks produktivnosti dela. Produktivnost dela za posamezen artikel v merimo z vrednostjo proizvodnje na enoto časa

$$v = \frac{V}{T} \qquad (11.27)$$

Skupna vrednost proizvodnje za celo podjetje je

$$V = \Sigma T v \qquad (11.28)$$

vsota produktov porabljenega časa in individualnih produktivnosti dela za posamezne artikle, torej agregat. Agregatni indeks produktivnosti dela dobimo po istem načelu kot indeks cen. Laspeyresov obrazec za izračunavanje indeksa o produktivnosti dela je na primer

$$I_v = \frac{\Sigma T_0 v_1}{\Sigma T_0 v_0} \qquad (11.29)$$

Ta obrazec za izračunavanje indeksov o produktivnosti dela poznamo pod imenom indeks produktivnosti dela pri *stalni strukturi časa*.

Če pa vzamemo za posamezen artikel za merilo produktivnosti dela čas t, ki ga porabimo za proizvodnjo enote produkta,

$$t = \frac{T}{Q} \qquad (11.30)$$

je skupno porabljen čas za celotno proizvodnjo v podjetju agregat

$$T = \Sigma Q t \qquad (11.31)$$

Pri tem pomenijo: T = čas, porabljen za proizvodnjo posameznega artikla, Q = proizvedena količina za posamezen artikel, t = proizvodnost dela za posamezen artikel.

Podobno kakor pri drugih problemih, moremo iz tega agregata izračunati agregatni indeks o produktivnosti dela po obrazcu

$$I_t = \frac{\Sigma Q_0 t_0}{\Sigma Q_0 t_1} \qquad (11.32)$$

V tem indeksu je agregat $\Sigma Q_0 t_1$ v imenovalcu, $\Sigma Q_0 t_0$ pa v števcu zato, ker je t recipročen pokazatelj produktivnosti dela. Čim večja je produktivnost dela, tem manjši je čas, porabljen za izdelavo enote proizvoda. Ker so v obrazcu 11.32 ponderi stalne količine, imenujemo ta indeks indeks produktivnosti dela pri *stalni strukturi količin.*

Podobno izračunamo agregatne indekse plač po kvalifikaciji, ker je skupni plačni fond vsota produktov števila delavcev in povprečnih plač po kvalifikaciji $P = \Sigma N \bar{v}$, in najrazličnejše druge agregatne indekse iz vseh področij ekonomskih statistik.

11.25 V dosedanjih obrazcih smo obravnavali indekse s stališča proučevanja dinamike pojavov. Pri tem smo vzeli oznako 0 za bazično razdobje, 1 pa za tekoče razdobje. Vendar moremo brez težav, vse, kar smo povedali o agregatnih indeksih, uporabiti tudi za primerjave, ki niso časovne. Tako moremo z agregatnim indeksom o produktivnosti dela primerjati produktivnost dela med podjetji iste stroke. Pri tem z 0 označimo podatke za prvo, z 1 pa podatke za drugo podjetje, ki ju primerjamo. Enako moremo agregatne indekse uporabiti pri izračunavanju indeksov o izpopolnjenju plana. Pri tem z 1 zaznamujemo stvarne, z 0 pa planske podatke. Prav tako izračunavamo indekse prihranka v materialu. Za take primere v indeksu pomenijo q_0 normirane količine porabljenih surovin, q_1 pa stvarne količine porabljenih surovin.

Seveda v analizah te vrste odpadejo določene metode, ki so umestne le pri časovnih agregatnih indeksih. Ena izmed njih je n.pr. metoda verižnih indeksov itd. V glavnem pa celotno teorijo indeksov, ki smo jo nakazali v tem poglavju, uporabljamo v različnih primerih in ni omejena le na proučevanje dinamike pojavov.

Dvanajsto poglavje

ČASOVNE VRSTE

12.1 Socialno-ekonomski pojavi niso nespremenljivi. Spremembe so rezultat delovanja najrazličnejših faktorjev, ki v tej ali oni obliki vplivajo na pojav. Sliko dinamike pojavov dobimo s časovnimi vrstami. Ena slika daje statični prikaz. Niz slik, ki si sledе na filmskem platnu, pa daje vtis gibanja. Enako da en sam podatek statično sliko pojava, niz istovrstnih podatkov v enakih časovnih razmakih pa daje dinamiko pojava. Časovna vrsta je niz istovrstnih podatkov, ki se nanašajo na sukcesivne časovne razmake ali momente. Z njimi proučujemo časovni razvoj pojavov, ker prikazujejo spremembe pojavov v odvisnosti od časa. Po namenu in načinu proučevanja se povsem razlikujejo od ostalih vrst statističnih vrst. Osnovni namen proučevanja časovnih vrst je, opazovati časovni razvoj socialno-ekonomskih pojavov, iskati njegove zakonitosti in napovedati nadaljnji razvoj. V pogojih planskega gospodarstva je proučevanje časovnih vrst važno tako pri sestavljanju plana kot pri kontroli izvajanja plana. Predvidevanje in napovedovanje razvoja ekonomskih pojavov, če poznamo zakonitosti in sedanje stanje, se je razvilo v posebno disciplino vede o konjunkturah, ki s proučevanjem časovnih vrst najvažnejših ekonomskih pokazateljev skuša sklepati na nadaljnji ekonomski razvoj. Seveda to napovedovanje ne more biti popolnoma zanesljivo, ker je nemogoče vnaprej napovedati in upoštevati vse faktorje, ki vplivajo na ekonomski razvoj. Napoved bi veljala strogo le v primeru, če bi bile izpolnjene predpostavke, pod katerimi je napoved izdelana.

OBLIKE ČASOVNIH VRST

12.2 Momentne in intervalne časovne vrste. Časovne vrste po značaju podatkov, ki jih prikazujejo, delimo na *momentne* in *intervalne*. Momentne časovne vrste prikazujejo časovni razvoj pojavov, ki veljajo za posamezne momente, intervalne časovne vrste pa časovni razvoj pojavov, ki so intervalnega značaja. Momentna časovna vrsta so n. pr. podatki o številu delavstva konec posameznega meseca, primer intervalne časovne vrste pa je proizvodnja določenega podjetja po mesecih. Enako je v tabeli 12.1 število prebivalstva v FLRJ po letih sredi leta momentna, naravni prirastek po letih pa intervalna časovna vrsta.

Tabela 12.1 Število prebivalstva in naravni prirastek prebivalstva v FLRJ

Leto	Število prebivalstva sredi leta v tisočih	Naravni prirastek v tisočih
1947	15679	217
1948	15901	233
1949	16133	266
1950	16343	232
1951	16588	213
1952	16798	302
1953	17048	272
1954	17318	306
1955	17583	271
1956	17791	262
1957	18005	236

Izvedene časovne vrste

12.3 Iz osnovnih časovnih vrst moremo izračunati več izvedenih časovnih vrst, ki pripomorejo k analizi dinamike pojavov. Izmed teh bomo navedli: kumulativno časovno vrsto, časovno vrsto sredin, časovno vrsto drsečih vrednosti in časovno vrsto drsečih sredin.

Kumulativna časovna vrsta

12.4 Kumulativno časovno vrsto izračunavamo samo za intervalne vrste ekstenzivnih podatkov, ker ima samo zanje kumuliranje oziroma seštevanje členov smisel. Kumulativno časovno vrsto dobimo po znanem principu postopnega seštevanja členov osnovne vrste. Členi kumulativne vrste pomenijo vrednosti v razmaku od začetka proučevanega razdobja do določenega trenutka, ki se od člena do člena menja. Kumulativno časovno vrsto izračunavamo običajno za zaključene časovne razmake, kot so, mesečne vrednosti v razdobju enega leta, dnevne vrednosti v razdobju enega meseca, letne vrednosti v razdobju ene petletke ita. Kumulativne vrste se v teh primerih nanašajo na začetek petletke, začetek leta, začetek meseca.

Kumulativno časovno vrsto dobimo tako, da vzamemo vrednost prvega člena kumulativne vrste enako $K_1=0$ nadaljnje člene kumulativne vrste pa izračunamo po znanem obrazcu

$$K_{k+1} = K_k + Y_k \qquad (12.1)$$

K_k = kumulativa; Y_k = vrednost v tekočem razdobju.

12.5 Za primer izračunavanja kumulativne časovne vrste vzemimo mesečno vsoto proizvodnje krogličnih ležajev v FLRJ.

Tabela 12.2 Izračunavanje kumulativne časovne vrste za proizvodnjo krogličnih ležajev v FLRJ v letu 1957 (Vir: Indeks)

Mesec	Mesečna proizvodnja v t	Kumulativa
J	17	0 = 0
F	16	0 + 17 = 17
M	22	17 + 16 = 33
A	21	33 + 22 = 55
M	19	55 + 21 = 76
J	19	76 + 19 = 95
J	20	95 + 19 = 114
A	5	114 + 20 = 134
S	18	134 + 5 = 139
O	14	139 + 18 = 157
N	18	157 + 14 = 171
D	18	171 + 18 = 189
Skupaj		189 + 18 = 207

Šesti člen kumulativne serije 95 ton pomeni, da je bilo od začetka leta 1957 do začetka meseca junija v FLRJ proizvedeno 95 ton krogličnih ležajev. Podoben pomen imajo drugi členi kumulativnih vrst.

Vrsta sredin

12.6 Časovne vrste z velikim številom členov so dostikrat nepregledne in je osnovna dinamika zaradi prevelike dolžine zabrisana. Za take vrste zmanjšamo število členov z združevanjem členov v vsote ali povprečja. Iz časovne vrste o številu tujih gostov po mesecih v razdobju 1948-1957, ki ima 144 členov, dobimo časovno vrsto z 12 členi, če seštejemo število tujih gostov po mesecih v letne vsote. Seveda ta časovna vrsta ne podaja tako natančno gibanje proučevanega pojava kot osnovna vrsta. S tem zabrišemo marsikatero značilnost pojava.

Medtem ko ima vrsta vsot smisel le, če je vrsta, ki jo prikazujemo, ekstenzivna, moremo z grupnimi povprečji skrčiti število členov za vsako časovno vrsto, ne glede na značaj členov.

Člene osnovne časovne vrste združujemo po naravnih časovnih enotah. Tako združujemo mesečne podatke v četrtletne ali letne, dnevne podatke v tedenske ali mesečne, letne podatke pa običajno v petletja tako, da iz njih dobimo zaokrožena petletja ali desetletja, n.pr. 1901-1904, 1905-

1909, 1910-1914 ... 1945-1949, 1950-1954.

12.7 Če izračunamo časovno vrsto povprečij, je treba paziti, na katere momente se nanašajo osnovne vrednosti časovne vrste. Od tega je namreč odvisno, kako izračunavamo sredine.

Če se nanašajo vrednosti na sredino elementarnega razmaka, kot je primer pri intervalnih časovnih vrstah in pri nekaterih momentih, izračunavamo povprečja po obrazcu

$$\bar{Y} = \frac{1}{r}(Y_1 + Y_2 + \cdots\cdots + Y_r) \qquad (12.2)$$

Za momente časovne vrste, za katere se vrednosti nanašajo na začetek ali konec osnovnih intervalov, pa izračunavamo sredine po obrazcu

$$\bar{Y} = \frac{1}{r}(\tfrac{1}{2}Y_0 + Y_1 + Y_2 + \ldots + Y_{r-1} + \tfrac{1}{2}Y_r) \qquad (12.3)$$

Po obrazcu 12.2 izračunavamo na primer povprečno mesečno proizvodnjo po letih, če imamo mesečne podatke o proizvodnji ali povprečno število prebivalstva po petletjih, če razpolagamo s številom prebivalstva po letih sredi leta. Obrazec 12.3 pa upoštevamo, če izračunavamo na primer povprečno število delavstva po letih, če imamo podatke o številu delavstva v začetku vsakega meseca itd. Na to moramo paziti zato, da se povprečja nanašajo na zaokrožena koledarska razdobja: tedne, mesece, leta, petletja.

Vrsta drsečih vsot

12.8 Iz mesečnih podatkov o proizvodnji s seštevanjem dobimo časovno vrsto za letno proizvodnjo. Ti letni podatki se nanašajo na letne razmake koledarskih let od januarja do decembra. Za neke analize pa so poleg teh važni tudi drugi letni razmaki, na primer od februarja danega leta do januarja naslednjega leta, od marca do februarja naslednjega leta itd. Ti letni razmaki se med seboj delno prekrivajo tako, da je vsak razmak, če ga primerjamo s predhodnim, pomaknjen za en mesec. Letni razmaki tako rekoč drsijo od člena do člena za en mesec. Odtod tudi ime drseče vsote. Časovna vrsta drsečih vsot v bistvu ni krajša od osnovne časovne vrste, ima pa niz lastnosti, ki ji dajo posebno analitično vrednost. Ena izmed teh je ta, da se v časovni vrsti drsečih vsot izravnavajo slučajni in periodični vplivi, če obsega vsota časovno periodo. O tem bomo podrobneje slišali v naslednjih poglavjih.

Zaradi zveze med zaporednimi členi časovne vrste drsečih vsot izračunavamo časovno vrsto drsečih vrednosti razmeroma enostavno. Če imamo izračunan prvi člen časovne vrste drsečih vsot (dobimo ga tako, da seštejemo vrednosti od januarja do decembra za prvo leto v časovni vrsti), dobimo naslednji člen v časovni vrsti drsečih sredin tako, da januarsko vrednost od prve letne vsote odštejemo, prištejemo pa vrednost za januar

naslednjega leta. Iz tako dobljenega drugega člena dobimo tretji člen časovne vrste drsečih vsot, če prištejemo februarski člen prihodnjega in odštejemo februarski člen prejšnjega leta iz osnovne časovne vrste. Posamezni členi vrste drsečih vsot pomenijo letno proizvodnjo v preteklem letu od začetka januarja, od začetka februarja itd.

Nakazano izračunavanje vrste drsečih vsot ne velja samo za mesečne podatke, marveč na splošno za katero koli vrsto ekstenzivnih intervalnih podatkov. Enako iz dnevnih podatkov izračunavamo časovno vrsto tedenskih drsečih vrednosti o številu prevoženih potnikov. Iz letnih podatkov o številu živorojenih pa moremo sestaviti petletno vrsto drsečih vsot za število živorojenih itd.

Drseče vrednosti v vseh primerih izračunavamo po splošnem obrazcu:

$$S_{k+1} = S_k + Y_k - Y_{k-r} = S_k + {}_rd_k \qquad (12.4)$$

Pri tem pomeni: S_k in S_{k+1} = ustrezne vrednosti v vrsti drsečih vsot; Y_k = člen k v osnovni vrsti; Y_{k-r} = člen osnovne vrste, ki je za r členov pred Y_k; r = število členov, iz kolikor so računane vsote: ${}_rd_k$ = $= Y_k - Y_{k-r}$ = razlika ustreznih vrednosti za Y.

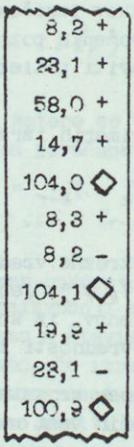
12.9 Za število potnikov v zračnem prometu v FLRJ v razdobju 1955–1958 je izračunavanje časovne vrste drsečih vsot nakazano v tabeli 12.3.

Tabela 12.3 Izračunavanje časovne vrste drsečih vsot za število potnikov v zračnem prometu po četrtletjih v FLRJ v razdobju 1958-1958 (Vir: Indeks)

Leto	Četrtletje	Y_k	Y_{k-4}	${}_4d_k$	S_k
1955	1	8,2			
	2	23,1			
	3	58,0			
	4	14,7			
1956	1	8,3	8,2	+ 0,1	104,0
	2	19,9	23,1	- 3,2	104,1
	3	50,6	58,0	- 7,4	100,9
	4	15,2	14,7	+ 0,5	93,5
1957	1	9,8	8,3	+ 1,5	94,0
	2	23,3	19,9	+ 3,7	95,5
	3	60,4	50,3	+ 9,8	99,2
	4	17,9	15,2	+ 2,7	109,0
1958	1	10,9	9,8	+ 1,1	111,7
	2	26,0	23,3	+ 2,4	112,8
	3	72,3	60,4	+11,9	115,2
	4	21,0	17,9	+ 3,1	127,1
					130,2

Iz primera je jasno razviden postopek 8,2 + 23,1 + 38,0 + 14,7 = = 104,0: 104,0 + (+ 0,1) = 104,1: 104,1 + (- 8,2) = 100,9 ...

Če računamo časovno vrsto drsečih vsot z računskim strojem, ki ima registrirni trak, ni treba izračunavati vmesnih diferenc. Časovno vrsto drsečih vsot dobimo neposredno, če upoštevamo subtotale. Ta postopek je nakazan v sliki 12.1



Slika 12.1 Registrirni trak iz seštevalnega stroja: Izračunavanje časovne vrste drsečih vsot

Časovna vrsta drsečih sredin

12.10 Vrednosti členov v časovni vrsti sredin priredimo sredini razmakov, na katere se povprečja nanašajo (n.pr. sredini let pri letnih povprečjih, tretjemu letu petletke pri petletnih povprečjih itd.). Členi v časovni vrsti sredin so torej prirejeni samo nekaterim členom osnovne časovne vrste. Za analizo pa je dostikrat potrebno, da imamo prirejena povprečja vsem, ne pa samo nekaterim členom v osnovni časovni vrsti. To dosežemo s časovno vrsto drsečih sredin, ki jo izračunavamo podobno kot vrsto drsečih vsot.

Ker se morajo posamezni členi vrste drsečih sredin nanašati na iste momente kot členi osnovne vrste, izračunavamo člene vrste drsečih vsot glede na osnovno časovno vrsto na dva načina:

a) Če imajo razmaki, na katere se nanašajo sredine, liho število osnovnih razmakov: $r = 2i + 1$, izračunavamo časovno vrsto drsečih sredin po obrazcu

$$\overline{Y}_k = \frac{1}{r}(Y_{k-i} + Y_{k-i+1} + \dots + Y_k + \dots Y_{k+i-1} + Y_{k+i}) = \frac{1}{r} S_{k+i+1} \quad (12.5)$$

Tak primer je izračunavanje tedenskih ali petletnih povprečij.

12.11 Vzemimo časovno vrsto o številu umrlih na 1000 prebivalcev v b. Jugoslaviji v letih 1921-1939 in zanjo izračunajmo časovno vrsto petletnih sredin? Ker izračunavamo povprečje iz lihega števila osnovnih razmakov, uporabimo obrazec 12.5.

Iz obrazca 12.5 zaključimo, da je najbolje, če po postopku, ki smo ga že navedli, izračunamo vrsto drsečih petletnih vsot S_k, petletne vsote pa delimo s pet.

Tabela 12.4 Izračunavanje časovne vrste petletnih drsečih sredin za število umrlih na tisoč prebivalcev v b. Jugoslaviji v letih 1921-1936. (Vir: SG 57)

Leto	Y_k	Y_{k-5}	$_5d_k$	S_k	$\overline{Y}_k$
1921	20,9				
1922	20,8				
1923	20,3				20,2
1924	20,2				19,8
1925	18,7				19,8
1926	18,8	20,9	- 2,1	100,9	19,8
1927	20,9	20,8	+ 0,1	98,8	20,0
1928	20,4	20,3	+ 0,1	98,9	20,1
1929	21,2	20,2	+ 1,0	99,0	20,3
1930	19,0	18,7	+ 0,3	100,0	19,9
1931	19,8	18,8	+ 1,0	100,3	19,2
1932	18,2	20,9	- 1,7	101,3	18,4
1933	16,9	20,4	- 3,5	99,3	18,0
1934	17,1	21,2	- 4,1	96,1	17,2
1935	16,8	19,0	- 2,2	92,0	16,6
1936	16,1	19,8	- 3,7	89,8	16,3
1937	15,9	19,2	- 3,3	83,1	15,9
1938	15,6	16,9	- 1,3	82,8	
1939	14,9	17,1	- 2,2	81,5	
				79,3	

12.12 b) Če pa sredine veljajo za sodo število osnovnih razmakov $(r = 2i)$, drseče sredine izračunavamo po obrazcu:

$$\overline{Y}_k = \frac{1}{r}\left(\frac{1}{2}Y_{k-i} + Y_{k-i+1} + \ldots + Y_k + \ldots + Y_{k+i-1} + \frac{1}{2}Y_{k+i}\right) \qquad (12.6)$$

Tak primer imamo pri letnih povprečjih iz mesečnih $(r = 12)$ ali četrtletnih $(r = 4)$ podatkih.

Če je število osnovnih razmakov sodo $(r = 2i)$, je najprikladneje obrazec 12.3 preurediti v obliko

$$\bar{Y}_k = \frac{1}{2r}(Y_{k-i} + 2Y_{k-i+1} + \ldots + 2Y_k + \ldots 2Y_{k+i-1} + Y_{k+i}) = \frac{1}{2r}S'_k \quad (12.7)$$

Po tem obrazcu izračunavamo vrsto drsečih sredin, če je $r = 2i$ po naslednjem postopku:

a) Osnovno časovno vrsto Y_k pišemo dvojno tako, da je drugič premaknjena za r členov. Tako dobimo časovni vrsti Y_k in Y_{k-r}. Moremo pa osnovno časovno vrsto pisati tudi enkrat, v osebkih po r členov (glej primer v tabeli 12.5).

b) Izračunamo razlike ustreznih zaporednih členov $_r d_k = Y_k - Y_{k-r}$

c) Izračunamo prvo vsoto $S'_{i+1} = Y_1 + 2Y_2 + \ldots 2Y_{i+1} + \ldots + 2Y_r + Y_{r+1}$;

d) Druge vsote dobimo postopno po obrazcu:

$$S'_{k+1} = S'_k + {}_r d_{k+i} + {}_r d_{k+i+1} \quad (12.8)$$

e) Iz vsot S'_k dobimo vsote drsečih sredin, če po obrazcu 12.7 vsote S'_k delimo z $2r$.

12.13 Če imamo vrsto mesečnih podatkov, po pravilu izračunavamo vrsto drsečih sredin za letna razdobja. Za ta primer je po zgornjih obrazcih: $r = 2i = 12$; $i = 6$. Ker začnemo z $S'_{i+1} = S'_{6+1} = S'_7$, se prvi člen vrste drsečih sredin nanaša na 7. mesec - julij prvega leta. Zanj izračunamo S'_7 po obrazcu 12.7 $S'_7 = (Y_1 + 2Y_2 + \ldots + 2Y_7 + \ldots + 2Y_{12} + Y_{13})$; Y_{13} = podatek za januar drugega leta. Po obrazcu 12.8 pa dobimo dalje $S_8 = S'_7 + {}_{12}d_{13} + {}_{12}d_{14}$; $S'_9 = S'_8 + {}_{12}d_{14} + {}_{12}d_{15}$ itd. ... ${}_{12}d_{13} = Y_{13} - Y_1$; ${}_{12}d_{14} = Y_{14} - Y_2$ itd.

Kot primer vzemimo indeks industrije gradbenega materiala v FLRJ.

Tabela 12.5 Izračunavanje časovne vrste drsečih sredin za indeks industrije gradbenega materiala v FLRJ (Ø 1956 = 100) (Vir: SB 108)

		J	F	M	A	M	J	J	A	S	O	N	D
1951	Y	38	37	49	56	78	95	97	104	90	84	71	58
$S'=24$	d												
	$\bar{Y}$							1707	1690	1667	1648	1635	1634
	$\bar{Y}$							71,1	70,4	69,5	68,7	68,1	68,1
1952	Y	31	27	36	50	71	101	124	122	112	98	71	56
	d	- 7	- 10	- 13	- 6	- 7	+ 6	+ 27	+ 18	+ 22	+ 14	0	- 2
$S'=24$	$\bar{Y}$	1667	1712	1752	1788	1802	1800	1801	1810	1828	1858	1897	1936
	$\bar{Y}$	69,5	71,3	73,0	74,5	75,1	75,0	75,0	75,4	76,2	77,4	79,0	80,7

		J	F	M	A	M	J	J	A	S	O	N	D
1953	Y	34	33	48	68	92	119	135	131	127	118	86	74
	d	+ 3	+ 6	+ 12	+ 18	+ 21	+ 18	+ 1	+ 9	+ 15	+ 20	+ 15	+ 18
$S'=24$	$\bar{Y}$	1955	1965	1989	2024	2059	2092	2108	2095	2084	2061	2071	2062
	$\bar{\bar{Y}}$	81,5	81,9	82,9	84,3	85,8	87,2	87,8	87,3	86,8	86,7	86,3	85,9
1954	Y	32	22	48	65	85	117	134	143	145	122	90	88
	d	- 2	- 11	0	- 3	- 7	- 2	+ 9	+ 12	+ 18	+ 4	+ 4	+ 9
$S'=24$	$\bar{Y}$	2062	2080	2120	2142	2150	2163	2195	2247	2282	2284	2292	2302
	$\bar{\bar{Y}}$	86,2	87,1	88,3	89,3	89,6	90,1	91,5	93,6	95,1	95,2	95,5	97,2
1955	Y	55	51	54	61	97	145	138	144	135	131	89	81
	d	+ 23	+ 29	+ 6	- 4	+ 12	+ 28	+ 4	+ 1	- 10	- 1	- 1	- 2
$S'=24$	$\bar{Y}$	2364	2369	2360	2349	2347	2344	2339	2311	2276	2266	2256	2232
	$\bar{\bar{Y}}$	98,5	98,7	98,3	97,9	97,8	97,7	97,5	93,3	94,8	94,4	94,0	93,0
1956	Y	52	26	44	31	87	131	139	157	155	154	106	89
	d	- 3	- 25	- 10	0	- 10	- 14	+ 1	+ 13	+ 20	+ 33	+ 17	+ 8
$S'=24$	$\bar{Y}$	2219	2223	2266	2319	2369	2394	2394	2398	2420	2445	2492	2547
	$\bar{\bar{Y}}$	92,5	93,0	94,4	96,6	98,7	99,8	99,8	99,9	100,8	101,9	103,8	106,1
1957	Y	44	38	54	76	119	154	172	179	166	152	121	97
	d	- 8	+ 12	+ 10	+ 15	+ 32	+ 23	+ 33	+ 22	+ 11	- 2	+ 15	+ 8
$S'=24$	$\bar{Y}$	2603	2658	2691	2700	2713	2736						
	$\bar{\bar{Y}}$	1085	1108	1121	1125	1130	1140						

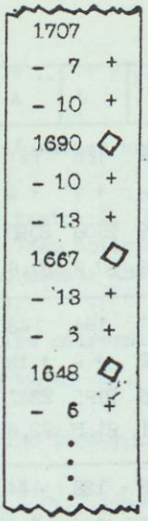
Razlike med ustreznimi meseci zaporednih let smo izračunali takole: Prva možna razlika je v mesecu januarju 1952; ${}_{12}d_{13} = Y_{J,\,52} - Y_{J,\,51} = 31 - 38 = -7$; ${}_{12}d_{14} = Y_{F,\,52} - Y_{F,\,51} = 27 - 37 = -10$ itn.

Prvo vsoto (za julij 1951) smo izračunali v celoti:

$$S'_{d,\,51} = 38 + 2.37 + 2.49 + \ldots + 2.84 + 2.71 + 2.58 + 31 = 1707$$

Nadaljnje vsote pa so izračunane, kot kaže trak na seštevalnem stroju na sliki 12.2.

Iz obeh primerov vidimo, da je časovna vrsta drsečih sredin krajša od osnovne časovne vrste in da začetnim in končnim i členom v osnovni časovni vrsti ne moremo priredit drseče sredine. To je hiba časovne vrste drsečih sredin, katere ne moremo odpraviti.



Slika 12.2 Registrirni trak na seštevalnem stroju pri izračunavanju vrste drsečih vsot $S_i^!$.

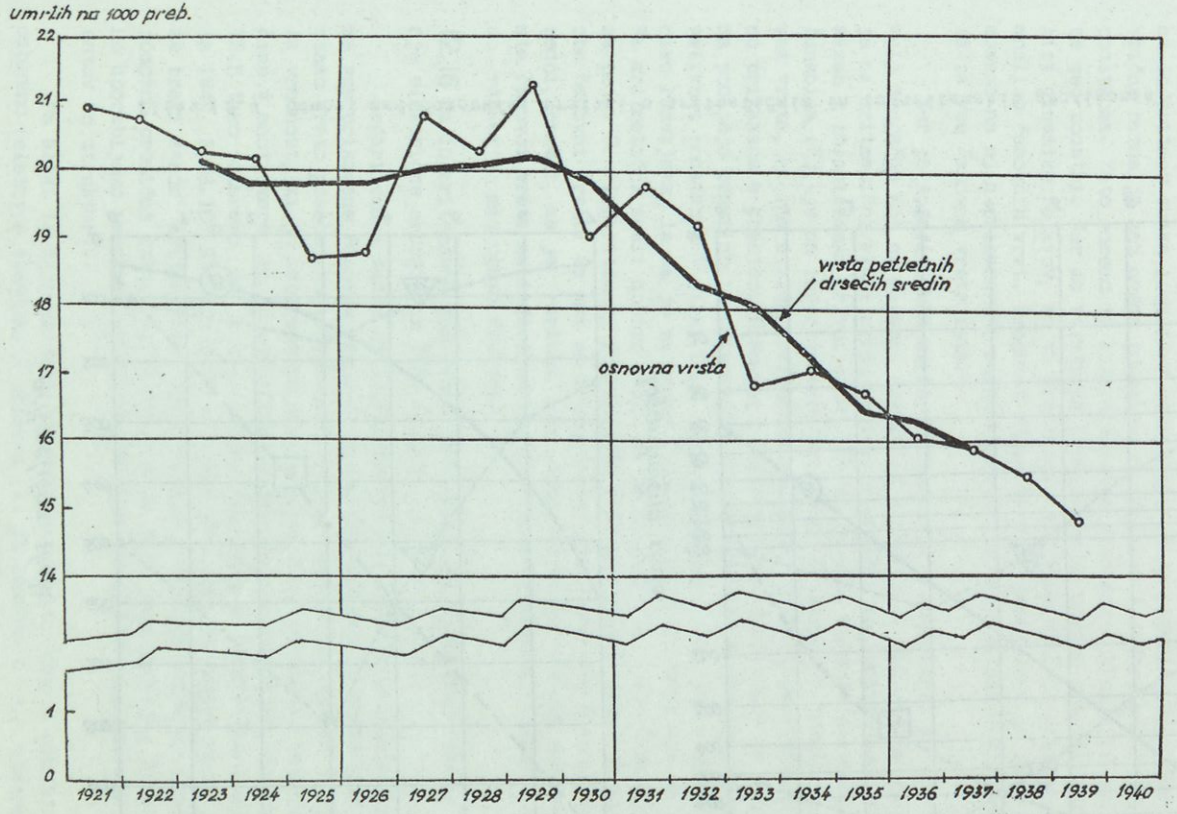
GRAFIČNO PRIKAZOVANJE ČASOVNIH VRST

12.14 Kompleksen upogled v dinamiko pojavov dobimo z grafičnim prikazom časovnih vrst. Tabelarični prikaz časovnih vrst je nenazoren, ker imajo časovne vrste običajno veliko členov. Razen tega je iz tabele izredno težko primerjati več časovnih vrst istočasno.

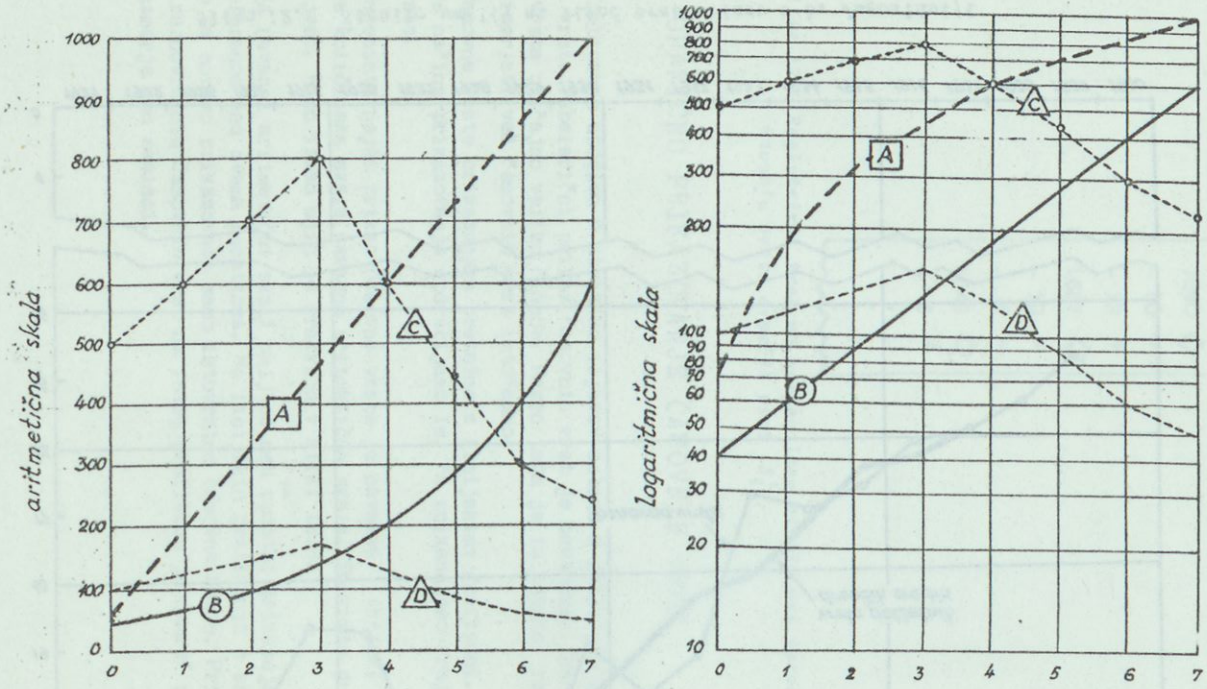
Časovne vrste prikazujemo pretežno z linijskimi grafikoni. Stolpce ali druge načine prikazovanja uporabljamo le, če prikazujemo krajše časovne vrste.

Najenostavnejši prikaz časovne vrste je navaden *linijski grafikon*. V njem je količinska skala navadna aritmetična skala. Te vrste grafikonov imamo največ. Enen izmed njih je grafikon v sliki 12.3.

Na navadni aritmetični skali daljice med podatki prikazujejo absolutne diference med dvema podatkoma. Na linijskih grafikonih z aritmetično skalo ne moremo prikazovati samo istovrstne časovne vrste. Pri prikazovanju raznovrstnih časovnih vrst na istem grafikonu je namreč problematično razmerje med skalami.



Slika 12.3 Število umrlih na tisoč prebivalcev v b. Jugoslaviji



Slika 12.4 Odnosi med časovnimi vrstami v aritmetičnih in logaritemskih grafikonih

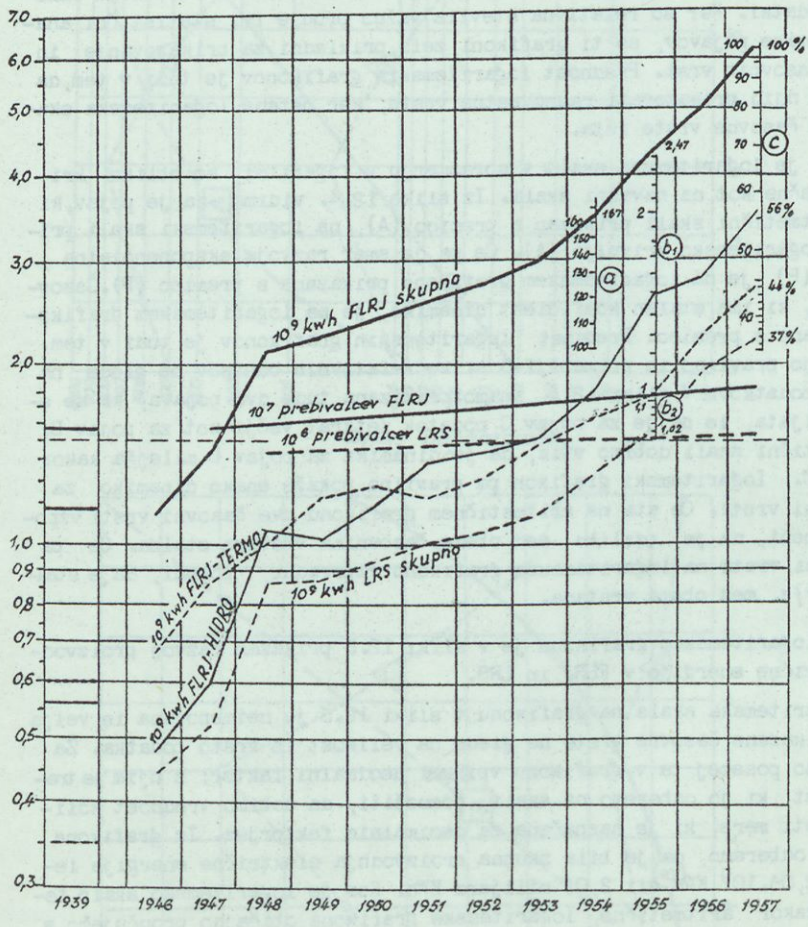
12.15 Za prikazovanje dinamike pojavov je posebno prikladen grafikon, ki ima namesto aritmetične količinske skale logaritemsko skalo. Ker je diferenca logaritmov enaka logaritmu kvocienta, je na logaritmičnem grafikonu navpična razdalja med dvema podatkoma v razmerju z logaritmom kvocienta med podatkoma. Zato moremo iz logaritemskega grafikona odbrati relativne odnose med podatki. Ker so relativna števila važno orodje pri statistični analizi dinamike pojavov, so ti grafikoni zelo prikladni za prikazovanje in analizo časovnih vrst. Prednost logaritemskih grafikonov je tudi v tem, da moremo na njih prikazovati raznovrstne vrste, ker ostane logaritemska skala za vse časovne vrste ista.

Ker je logaritemska skala v sorazmerju z logaritmi, so oblike krivulj drugačne kot na navadni skali. Iz slike 12.4. vidimo, da je pojav, ki je na aritmetični skali prikazan s premico (A), na logaritemski skali prikazan z logaritemsko krivuljo (A). Če pa je smer razvoja eksponencialna funkcija (E), je na logaritemskem grafikonu prikazana s premico (E). Časovna vrsta, ki ima stalen koeficient dinamike, je na logaritemskem grafikonu prikazana s premico. Prednost logaritemskih grafikonov je tudi v tem, da podajajo pravilno in primerljivo sliko relativnih odnosov ne glede na velikost podatkov. V sliki 12.4 imamo prikazana tudi dva pojava, ki se enako razvijata, le da je za pojav C podatek petkrat večji kot za pojav D. Na aritmetični skali dobimo vtis, da je dinamika za pojav D milejša kakor za pojav C. Logaritemski grafikon pa pravilno pokaže enako dinamiko za obe časovni vrsti. Če sta na aritmetičnem grafikonu dve časovni vrsti vzporedni, pomeni, da je *razlika* med obema časovnima vrstama stalna. Če pa sta časovni vrsti na logaritemskem grafikonu vzporedni, pomeni, da je stalno *razmerje* med obema vrstama.

12.16 Na logaritemskem grafikonu je v sliki 12.5 prikazan razvoj proizvodnje električne energije v FLRJ in LRS.

Logaritemska skala na grafikonu v sliki 12.5 je neimenovana in velja za vse prikazane časovne vrste ne glede na velikost in vrsto podatka. Za vsako vrsto posebej pa v grafikonu vpišemo decimalni faktor; z njim je treba vrednost, ki jo odberemo na skali, pomnožiti, da dobimo vrednost količine v enoti mere, ki je naznačena za decimalnim faktorjem. Iz grafikona 12.5 tako odberemo, da je bila skupna proizvodnja električne energije leta 1948 $2{,}03 \cdot 10^9$ KWh ali 2,03 milijard KWh. Ker je logaritemsko skalo teže brati kakor aritmetično, logaritemske grafikone običajno proučujemo s pomožno premično skalo, ki je vrisana na robu pasu papirja. To pomožno skalo uporabljamo predvsem za določanje relativnih števil (indeksov, koeficientov in struktur).

Na sliki 12.5 je na treh primerih naznačeno, kako s pomožno skalo odberemo relativna števila. V primeru, ki je označen z a, odberemo *indeks* skupne proizvodnje električne energije za FLRJ za leto 1954 na bazo 1948



Slika 12.5 Proizvodnja električne energije in število prebivalstva FLRJ in LRS v razdobju 1946 - 1957

Tabela 12.6 Proizvodnja električne energije in število prebivalstva v FLRJ in LRS v razdobju 1946 - 1957 (Vir: SG 57-58)

Leto	Proizvodnja električne energije v mlj KWh				Prebivalstvo v mlj prebivalcev	
	FLRJ			LRS	FLRJ	LRS
	skupno	hidro	termo	skupno		
1939	1173	566	607	350	15,696	...
1946	1150	478	672	440	...	...
1947	1453	598	855	523	15,879	...
1948	2061	1053	1008	867	15,901	1,443
1949	2214	1023	1191	878	16,133	1,457
1950	2408	1175	1233	937	16,343	1,473
1951	2550	1357	1193	1003	16,588	1,492
1952	2700	1423	1277	1047	16,798	1,498
1953	2982	1500	1482	1121	17,048	1,515
1954	3440	1810	1630	1293	17,318	1,521
1955	4340	2610	1730	1655	17,586	1,531
1956	5048	2870	2178	1905	17,791	1,541
1957	6252	3522	2730	2306	18,005	1,550

tako, da izhodišče pomožne skale 100 naravnamo na osnovo 1948. Na pomožni skali odberemo tam, kjer linija preseka skalo, indeks ca 135 (natančna vrednost je 137).

V primeru, ki je označen z b_1, določimo *koeficient* proizvodnje električne energije na enega prebivalca v FLRJ v letu 1955. Ta koeficient odberemo tako, da izhodišče 1. na pomožni skali naravnamo na linijo števila prebivalstva, na skali pa pri liniji proizvodnje odčitamo ustrezno vrednost 2,47. Vrednost koeficienta pa dobimo, če to vrednost pomnožimo s kvocientom 10^9 KWh/10^7 prebivalcev; tega dobimo iz decimalnih faktorjev, ki so vpisani ob ustreznih linijah za proizvodnjo in število prebivalstva. Koeficient je torej 247 KWh/preb. Analogen koeficient za isto leto za LR Slovenijo dobimo, če premaknemo izhodišče na pomožni skali na linijo prebivalstva za Slovenijo, ob liniji proizvodnje pa na pomožni skali odberemo vrednost 1,02. Po istem načelu kot prej to vrednost pomnožimo s kvocientom 10^9 KWh/10^3 prebivalcev. Iz grafikona ocenjeni koeficient je torej 1020 KWh/preb. v letu 1955 v LR Sloveniji.

Kako s pomožno logaritemsko skalo odčitamo *strukturne odstotke*, je nakazano v primeru *C*. Ker iščemo strukturne deleže skupne proizvodnje v FLRJ, izhodišče 100 na premični skali naravnamo v letu 1957 na skupno proizvodnjo FLRJ. Ob linijah proizvodnje pa za dele skupne proizvodnje odberemo, da je bilo od skupne proizvodnje v FLRJ v letu 1957 približno 56 % skupne proizvodnje hidroelektrike in približno 44 % proizvodnje termoelektrike. Proizvodnja Slovenije pa je približno 37 % od skupne jugoslovanske proizvodnje.

Iz teh treh primerov vidimo, kako iz logaritmičnih grafikonov odberemo vse vrste relativnih števil.

Čitanje logaritemskih grafikonov je podobno računanju z logaritmičnim računalom in ima vse njegove prednosti (hitrost, enostavnost) in pomanjkljivosti (nenatančnost).

Prez uporabe pomožne skale moremo iz logaritemskega grafikona opazovati splošne tendence dinamike. Tako vidimo, da je v letih 1948 - 1952 koeficient dinamike v proizvodnji skupne električne energije precej stalen, se v nadaljnjih letih povečuje, v razdobju 1954 - 1957 pa ima zopet precej stalno vrednost. Ker se razlike med linijo za skupno proizvodnjo in linijo proizvodnje hidroelektrarn vedno manjšajo, sklepamo, da delež hidroelektrike v skupni proizvodnji nevzdržema narašča. Delež proizvodnje Slovenije v skupni jugoslovanski proizvodnji je dokaj stabilen, ker sta obe liniji precej vzporedni. Iz grafikona moremo sklepati podobno tudi za druge odnose med prikazanimi podatki.

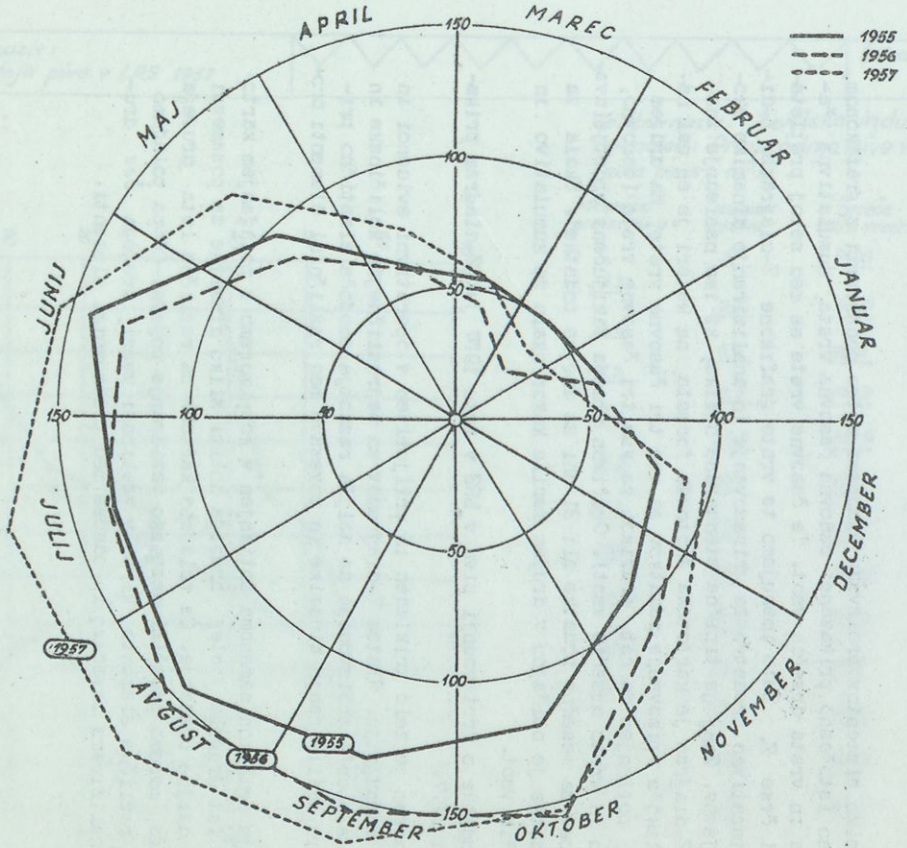
Polarni grafikon

12.17 Časovne vrste, ki imajo periodično komponento, včasih prikazujemo v polarnem koordinatnem sistemu. Za periodične časovne vrste je namreč pomembnejša in bolj smiselna primerjava med ustreznimi deli za različne periode, kakor pa primerjava med zaporednimi členi v časovni vrsti. S polarnim grafikonom najbolje primerjamo ustrezne člene v različnih periodah. Osnova polarnega grafikona je polarni koordinatni sistem; ta zamenja pravokotni koordinarni sistem. V polarnem grafikonu je čas znotraj ene periode nakazan s smerjo, podatek pa z oddaljenostjo točke od izhodišča polarnega koordinatnega sistema. Če imamo mesečno časovno vrsto s sezonsko variacijo, krog razdelimo v dvanajst enakih delov. Vsak izmed teh pomeni po en mesec. Če se podatki nanašajo na konce ali začetke meseca, rišemo podatke na meje posameznih krogovih izsekov, če pa se nanašajo na cele mesece ali na sredine mesecev, pa v sredino krogovih izsekov.

Za časovne vrste z izrazito sezonsko komponento kaže polarni grafikon tipično ekscentrično sliko. Če pojav narašča, dobimo na polarnem grafikonu spiralo, ki se odvija. Za pojave, ki padajo, pa se spirala zavija proti izhodišču. Seveda polarne grafikone ne uporabljamo samo za mesečne vrste s sezonsko variacijo, temveč tudi za prikazovanje časovnih vrst s kakršno koli periodo; npr. tedensko periodo v dnevni časovni vrsti itd. V teh primerih moramo seveda razdeliti krog na ustrezno število odsekov (za tedenske periode na sedem).

Hiba polarnega grafikona pa je v tem, da je nepregleden, če je časovna vrsta predolga ali če nima izrazitega trenda.

12.18 V sliki 12.3 je prikazana s polarnim grafikonom časovna vrsta indeksov proizvodnje za gradbeni material v FLRJ v letih 1955 - 1957.



Slika 12.6 Indeksi proizvodnje gradbenega materiala v FLRJ v razdobju 1955 - 1957 (Vir: Tabela 12.5)

Časovna skala je krožna, medtem ko je skala za indekse naznačena s koncentričnimi krogi. Za vsako leto so podatki vrtani drugače, da se leta ne pomešajo med seboj. Grafikon kaže izrazit sezonski značaj proizvodnje gradbenega materiala, ker so linije za posamezna leta ekscentrične. Razen tega pa kaže, da pojav narašča, ker se spirala davija. Iz grafikona vidimo, da je primerjava posameznih mesecev v različnih letih zelo dobra, ker podatki za isti mesec v različnih letih leže v isti smeri.

Z - Diagram

12.19 Dinamiko elementov proizvodnje moremo dobro ponazoriti z grafikonom, v katerem so istočasno prikazane: osnovna časovna vrsta, kumulativna časovna vrsta in vrsta drsečih vsot. Te časovne vrste se med seboj prepletajo v obliki črke Z. Zato imenujemo te vrste grafikone Z-diagrame. Z-diagram je operativno sredstvo za prikazovanje in analiziranje dinamike ekonomskih pojavov. Zanj je tipična standardna oblika, ki ima naslednje značilnosti: Z-diagram je kartoteka enotnega formata, na kateri je en del namenjen za tekoče vnašanje podatkov za vse tri časovne vrste, na drugem delu pa je, potisnjen na rob, grafikon za vse tri časovne vrste (osnovno, kumulativno in vrsto drsečih vsot). Grafikon obsega zaključeno obdobje (dvanajst mesecev za mesečne podatke ali 31 dni za dnevne podatke). Skala za osnovne podatke je običajno v drugem merilu kakor skala za kumulativo in vrsto drsečih vsot.

Za podatke o proizvodnji piva v LRS v letu 1957 je Z-diagram prikazan v sliki 12.7.

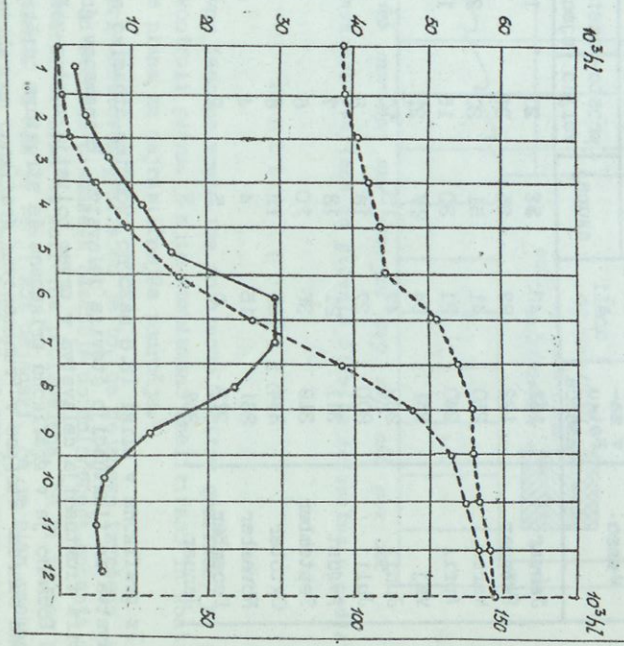
Z-diagram je zelo prikladen in priljubljen v operativni evidenci in statistiki v podjetju. Sistem Z-diagramov za najrazličnejše količinske in vrednostne elemente proizvodnje za daljša razdobja omogoča elastično prikazovanje in analiziranje dinamike in povezave med različnimi elementi proizvodnje.

Ker je grafikon namenoma potisnjen v rob, moremo s polaganjem kartonov za več let drugega *poleg* drugega dobiti sliko dinamike za posamezni element za daljše razdobje. Če polagamo kartone za različna leta drugega *nad* drugega, moremo proučiti sezonsko variiranje pojava. Če pa polagamo kartone za različne elemente, ki so v vsebinski zvezi, drugega *čez* drugega, pa analiziramo povezave in odnose med posameznimi elementi.

Kratek naziv:
Proizvodnja pive v LRS 1957

Šifra:

Celotni naziv: Živilska industrija 1957
Pokazatelj: Proizvodnja pive v LRS v 1957
Enota mere: 10^3 hl



Mesečna proizvodnja	Kumulativa	Drseča letna vrednost
	0	97,6
2,3	2,3	98,0
3,6	5,9	100,8
6,3	12,2	104,3
11,3	23,5	108,9
15,5	39,5	111,3
28,8	67,8	126,8
28,9	96,7	134,5
24,1	120,8	140,5
11,5	132,3	139,9
5,5	137,8	141,7
4,2	142,0	144,3
5,2	147,2	147,2

Slika 12.7 X-diagram za proizvodnjo piva v LRS v letu 1957

Prunsmanov grafikon

12.20 Časovne vrste, ki se vežejo po metodi salda (korespondirajoče populacije), prikazujemo z Prunsmanovim grafikonom. Časovnih vrst korespondirajočih populacij je veliko:

število delavcev v začetku meseca + na novo došli - odšli = število delavcev konec meseca;

zaloge v začetku meseca + nabavljeno - prodano = zaloga konec meseca;

zaloga surovin + nabavljeno - predelano = zaloga surovin konec dekade;

število prebivalstva v začetku leta + število rojenih + število doseljenih - število umrlih - število izseljenih = število prebivalstva konec leta.

V Prunsmanovem grafikonu momentne vrste (število delavstva, zaloge, število prebivalcev itd.) prikažemo z linijskim grafikonom, intervalne (na novo došli delavci, porabljene zaloge, število izseljenih itd.), pa s stolpci. Stolpci in linije so kombinirani tako, da se vežejo med seboj.

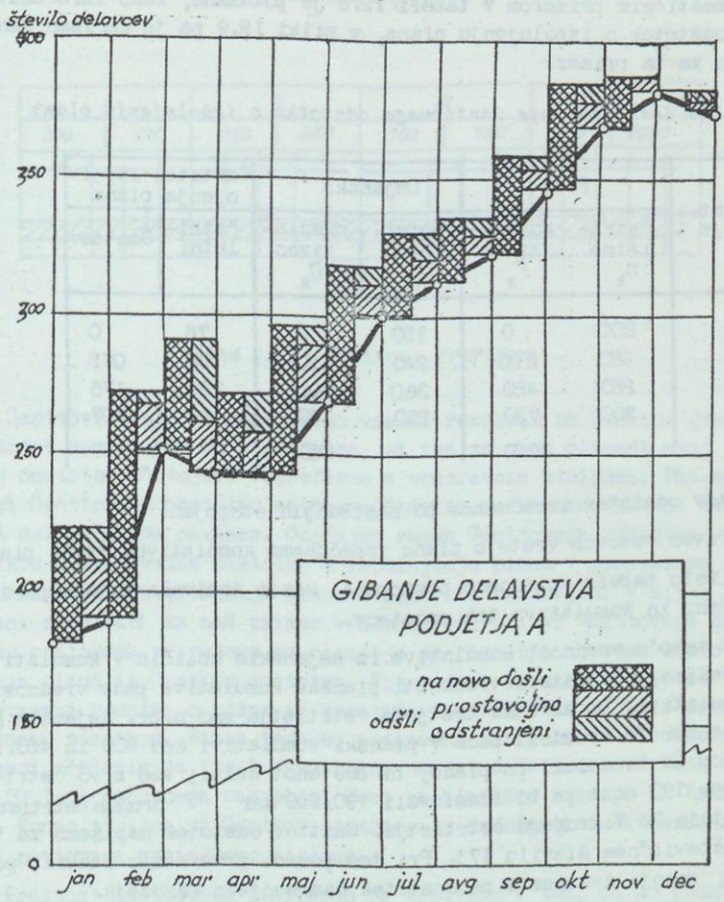
12.21 Z Prunsmanovim grafikonom prikažimo mesečne podatke o gibanju delavstva v podjetju A (podatki izmišljeni).

Tabela 12.7 Gibanje števila delavstva v podjetju A

Mesec	Stanje v začetku meseca	Na novo došli	Odšli		
			skupaj	prostovoljno	odstranjeni
Januar	181	43	32	20	12
Februar	192	83	25	21	4
Marec	250	41	51	20	31
April	240	31	30	15	15
Maj	241	54	27	21	3
Junij	268	47	15	10	5
Julij	300	27	16	8	8
Avgust	311	21	13	7	6
September	319	35	10	6	4
Oktober	344	28	11	6	5
November	361	15	4	4	0
December	372	0	5	2	3
Januar	367				

Iz grafikona v sliki 12.8 nazorno proučimo dinamiko stanja števila delavstva in fluktuacijo števila delavstva. Prunsmanov grafikon kaže, da je bila fluktuacija delavstva v prvem polletju znatno večja kakor v drugem. Dodatno je v grafikonu prikazana še struktura odšlih delavcev po vzroku odhoda. Stolpce v Prunsmanovem grafikonu uporabljamo za prikazovanje struktur tudi za druge pojave. Pri prebivalstvu stolpec za pove-

čanje prebivalstva razdelimo na novorojene in priseljene, stolpec zmanjšanja prebivalstva na umrle in izseljene itd.



Slika 12.8 Brunsmanov grafikon za gibanje števila delavstva v podjetju A

Ganttov grafikon

12.22 Kumulativne časovne vrste ima za osnovo Ganttov grafikon, ki ga uporabljamo pri kontroli plana. Z njim kompleksno analiziramo in kontroliramo izpolnjenje plana za najrazličnejša področja.

Osnovno pri Ganttovem grafikonu je to, da dejansko izvršeno količino v razdobju od začetka planskega razdobja do določenega momenta merimo s časom, v katerem bi bila ta količina proizvedena, če bi proces potekal ves čas po planu. Ta čas je večji kot dejanski, če je plan prekoračen; se ujema z dejanskim, če je plan izpolnjen in je manjši kakor dejanski, če plan ni izpolnjen.

S shematičnim primerom v tabeli 12.8 je pokazano, kako izračunavamo Ganttov odstotek o izpolnjenju plana, v sliki 12.9 pa je narisan Ganttov grafikon za ta primer.

Tabela 12.8 Izračunavanje Ganttovega odstotka o izpolnjenju plana

Četrtletje	Plan		Dejansko		Odstotek izpolnjenja plana	
	četrtletno p_k	kumulativno P_k	četrtletno d_k	kumulativno D_k	četrtletni	Ganttov
I	200	0	150	0	75	0
II	250	200	240	150	96	075
III	280	450	360	390	129	176
IV	300	730	250	750	83	307
		1080		1000		390

Ganttov odstotek izračunamo po naslednjih stopnjah:

a) Zraven časovne vrste o planu izračunamo kumulativno vrsto plana.

b) V isto tabelo vpisujemo postopoma, kakor dobivamo podatke, dejansko doseženo in kumulativo teh podatkov.

c) Za tekočo vrednost kumulative iz dejanskih količin v kumulativi za plan odredimo, med kateri vrednosti planske kumulative pade vrednost dejanske kumulative. Do začetka tretjega četrtletja smo n.pr. dejansko proizvedli 390 enot. Ta vrednost pade v planski kumulativi med 200 in 450. Iz tega sklepamo: Če bi delali po planu, bi 390 enot delali vse prvo četrtletje, ostanek 190 enot pa bi izdelovali 190/250 del drugega četrtletja ali v odstotkih 76 % drugega četrtletja. Ganttov odstotek napišemo za ta primer v troštevilčnem številu 173. Pri tem pomeni prvo mesto število polnih razdobij, zadnji dve mesti pa odstotek naslednjega razdobja.

Iz tabele 12.8 vidimo dalje, da smo do začetka četrtega četrtletja, za katerega je Ganttov procent 307 izdelali količino predvideno po planu v prvih treh četrtletjih in 7 % količine, predvidene za četrto četrtletje.

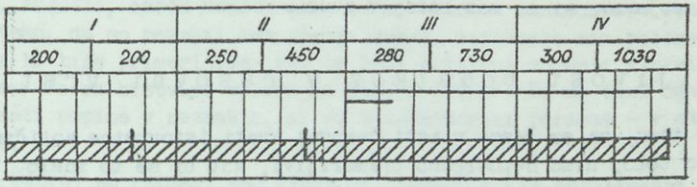
Ako s p_k zaznamujemo člene časovne vrste za plan, s P_k njegove kumulative, z d_k člene v časovni vrsti za dejansko opravljeno z D_k pa kumulativo te časovne vrste, v simbolih izrazimo izračunavanje Ganttovega odstotka takole:

a) poiščemo, med katera člena P_k in P_{k+1} v planski kumulativi pade vrednost dejanske kumulative D_r

$$P_k \leq D_r < P_{k+1} \tag{12.9}$$

b) Ganttov odstotek izračunamo po obrazcu: Prvo mesto Ganttovega odstotka je $k-1$. Naslednji dve mesti pa izračunamo po obrazcu

$$G\% = \frac{D_r - P_k}{P_k} 100 \qquad (12.10)$$



Slika 12.9 Ganttov grafikon

Ganttovi odstotki pa so samo vmesni rezultat za Ganttov grafikon. Te odstotke namreč rišemo v grafikon, ki ima za vsak element svoj pas. Ganttov odstotek vrtujemo v grafikon z vodoravnim stolpcem. Pas se viže skozi Ganttovemu odstotku ustrezno število celih razmakov in skozi ustrezni del načetega razmaka. Običajno razen Ganttovega odstotka rišemo v grafikon tudi navadne odstotke o izpolnjenju plana v posameznem razdobju. Ti nakazujejo izpolnjenje plana v osnovnih razdobjih. V sliki 12.9 so prikazani rezultati za naš primer v standardni obliki Ganttovega grafikona. Vanj postopoma vrisujemo od razdobja do razdobja navaden odstotek izpolnjenja plana in Ganttov odstotek. Kot vidimo iz slike 12.9, so v grafikon vpisani podatki o planu in kumulativi plana. Tako dobimo predstavo o velikosti podatkov. Razen tega so posamezni elementarni časovni razmaki (meseci, četrtletja itd.) razdeljeni na pet enakih delov. Vsak od njih pomeni 20 % od osnovnega razdobja. Tako si olajšamo risanje in branje grafikona. Konec stolpca za Ganttov odstotek za določen razmak zaznamujemo s črto in postopoma dorisujemo stolpec.

Prednost Ganttovega grafikona je predvsem v tem, da so podatki za vsak element naneseni v pasu. Zato moremo na isto sliko nanesti niz elementov: en element za več podjetij, ali različne elemente za isto podjetje. Tako moremo kompleksno analizirati izpolnjenje plana tudi v odnosih med elementi proizvodnje, ne pa samo izolirano za posamezne elemente proizvodnje.

ANALIZA ČASOVNIH VRST

12.23 Dinamiko pojavov analiziramo tako, da opazujemo spreminjanje vrednosti členov v časovnih vrstah in iščemo zakonitosti tega spreminjanja. Naloga enostavne analize časovnih vrst je enostavna primerjava med členi v isti časovni vrsti. Člene v časovni vrsti primerjamo lahko na več načinov.

Vsak izmed njih po svoje osvetli dinamiko pojava. Z metodami, ki so specifične za analizo časovnih vrst, pa analiziramo zakonitosti dinamike ene same vrste ali, s korelacijsko analizo, zakonitosti odvisnosti v dinamiki večih pojavov, ki so med seboj v zvezi.

Primerljivost podatkov v časovni vrsti

12.24 Kljub temu, da so členi v isti časovni vrsti istovrstne količine, dostikrat med seboj niso neposredno primerljivi, kot bi se to zdelo na prvi pogled. Da jih moremo primerjati, morajo izpolniti določene pogoje. Če jih ne izpolnjujejo, je treba časovno vrsto pred analizo preurediti tako, da tem pogojem zadošča.

Osnovni pogoj za primerljivost členov v isti časovni vrsti je pravilna in *nedvoumna opredelitev* pojava, katerega časovna vrsta prikazuje. Ta opredelitev mora biti vso dobo opazovanja enaka in se ne sme spreminjati.

12.25 Ker so spremembe pojava, ki ga časovna vrsta prikazuje, bistveno odvisne od časa, je zelo koristno, če so časovni *razmaki med posameznimi členi enaki*. Tako dosežemo, da je pri enakem delovanju faktorjev, sprememba pojava od člena do člena enaka; tega v nasprotnem primeru ne bi bilo. Zato imamo popise prebivalstva v razdobjih pet ali deset let, popise živine vsako leto na isti datum itd.

Kljub temu, da ta pogoj na prvi pogled ni težko izpolniti, naletimo v določenih primerih na težave. Problem so namreč osnovne vrste, za katere se členi nanašajo na krajša razdobja, mesece, dekade, tedne itd. Število rojstev po mesecih ne moremo neposredno primerjati, ker se število koledarskih dni v mesecu spreminja od 28 do 31 dni. Razlika med najkrajšim in najdaljšim mesecem je torej tri dni ali približno 10 %. Če bi bili vsi ostali faktorji, ki vplivajo na število rojstev, stalni, bi bilo število rojstev od meseca do meseca različno tudi za več ko 10 % zaradi vzroka, ki nima nobene zveze z rodnostjo. Te razlike so še večje pri pojavih, ki so, kakor na primer proizvodnja, odvisni ne od koledarskih, temveč od delovnih dni, ker so med meseci zaradi različnega števila nedelj in praznikov še večje variacije v številu delovnih dni. Koledar vpliva tudi na pojave, ki imajo tedensko periodičnost. Ker je na primer število prevoženih potnikov, prometnih nesreč ali nočnin v tujskem prometu največje okrog nedelje, je številčnost teh pojavov v mesecu bistveno odvisno od števila nedelj in praznikov v mesecu. Razlike med posameznimi meseci zaradi tega vzroka so tudi do 50 %.

12.26 Za *momentne časovne vrste* si pomagamo tako, da si prizadevamo, da so časovni intervali med posameznimi členi čim bolj enaki. Kakor smo že navedli, so popisi prebivalstva na enaka razdobja – vsakih pet ali deset let na isti datum, popisi živine vsako leto na isti datum. Le tako iz časovne vrste dobimo hiter pregled o dinamiki enega in drugega pojava.

Vendar enakost razmakov med posameznimi členi še ni zadosten pogoj za primerljivost med členi. Vzemimo, da popisi živine ne bi bili na letne, temveč na primer sedem-mesečne razmake. Popis živine bi bil na primer januarja, avgusta, marca naslednjega leta, oktobra naslednjega leta itd. Kljub temu, da so razmaki med popisi enaki, spremembe med posameznimi popisi ne bi bile primerljive, ker bi bili kritični momenti v različnih delih leta. To pa bistveno vpliva na stalež živine. Da se temu ognemo, moramo vzeti popise v razmakih, ki so mnogokratniki periode - v našem primeru leta.

Čeprav so razmaki med členi pri mesečnih ali četrtletnih momentih časovnih vrstah različni, je učinek razlik na časovno vrsto tako malenkosten, da podatkov ne popravljamo.

12.27 Pri *intervalnih časovnih vrstah* pa je ta učinek znaten in moramo podatke reducirati na enake razmake. To dosežemo tako, da za osnovne podatke, ki se nanašajo na dejanska mesečna razdobja, izračunamo namesto mesečnih vrednosti povprečja na en dan, enotin mesec s 30 koledarskimi ali 25 delovnimi dnevi. Tako odstranimo učinek različne dolžine meseca. Ali reduciramo podatke na 25 ali 30 dni, je odvisno od pojava, ki ga analiziramo. Rojstva reduciramo na koledarski mesec 30 dni, enako proizvodnjo v panogah, ki imajo zvezen delovni proces, na mesec 25 dni pa reduciramo proizvodnjo v panogah in podjetjih, v katerih je delovni proces prekinjen. Tako se najbolj prilagodimo dejanskemu številu dni v posameznih mesecih.

Popravni faktorji za posamezne vrste poprav so nakazani v tabeli 12.9

Tabela 12.9 Popravni faktorji za mesečne časovne vrste

Število koledarskih dni v mesecu	Popravni faktor	Število delovnih dni v mesecu	Popravni faktor
28	30/28 = 1,0714	22	25/22 = 1,1364
29	30/29 = 1,0345	23	25/23 = 1,0870
30	30/30 = 1,0000	24	25/24 = 1,0417
31	30/31 = 0,9677	25	25/25 = 1.0000
		26	25/26 = 0,9615
		27	25/27 = 0,9259

Podatke mesečnih časovnih vrst reduciramo na enotino dolžino meseca tako, da jih pomnožimo z ustreznimi popravnimi faktorji. Vsote popravljenih podatkov so brez smisla, ker so popravljeni podatki povprečja, čeprav so osnovni podatki intervalnega značaja.

12.28 Reducirajmo število rojstev dečkov v LR Sloveniji v letu 1955 na enotine mesece.

Tabela 12.10 Reduciranje števila rojstev dečkov v LRS v letu 1955 na enotne mesece (Vir: MSP)

Mesec	Število rojstev Y	Število dni	Popravni faktor F	Popravljeno število rojstev $Y^{*}=Y_f$
J	1571	31	0,9677	1520
F	1421	28	1,0714	1522
M	1577	31	0,9677	1526
A	1457	30	1,0000	1457
M	1488	31	0,9677	1440
J	1391	30	1,0000	1391
J	1384	31	0,9677	1339
A	1379	31	0,9677	1334
S	1310	30	1,0000	1310
O	1366	31	0,9677	1322
N	1407	30	1,0000	1407
D	1330	31	0,9677	1287

12.29 Na velikost pojavov dostikrat vplivajo *administrativni ukrepi*, ki z vsebino proučevanega pojava nimajo neposredne zveze. Eden izmed običajnih vzrokov so *upravno-teritorialne spremembe*, s katerimi se spremeni geografska opredelitev pojava. Taka sprememba je lahko sprememba državnih mej, ki je posledica vojne, ali sprememba upravno-teritorialnih mej znotraj države. Različna geografska opredelitev onemogoča primerljivost podatkov v časovni vrsti. V tem primeru je potrebno podatke časovne vrste za nazaj preračunati na novo območje, kar je običajno težaven in zamuden posel. Včasih si pomagamo tako, da za čas, ko je nastopila sprememba, registriramo podatek o starem in novem stanju. Tako vidimo, kolika je ob prehodu sprememba zaradi spremembe v teritoriju.

Če časovno vrsto za nazaj ne korigiramo, moramo analizirati vsak del posebej, nikakor pa ne smemo smatrati vrste kot enotno.

Isti problem so tudi spremembe zaradi drugih administrativnih ukrepov, na primer prehajanje podjetij iz ene kompetence v drugo ipd. Tudi v teh primerih ravnamo s spremembami enako kakor pri teritorialnih spremembah.

12.30 Ker se število prebivalstva stalno spreminja, povečana proizvodnja še ne pomeni nujno povečane produktivnosti dela, povečana potrošnja še ne povečanja življenjskega standarda ipd.

Vpliv sprememb v spremenjenem številu prebivalstva odstranimo tako, da izračunavamo *koeficiente "na prebivalca", "na delavca"* ipd.

Časovno primerljivost podatkov včasih motijo tudi *spremembe v strukturi*. Sprememba starostne strukture prebivalstva, ki more nastopiti v

daljšem časovnem razdobju, močno vpliva na umrljivost. Primerjava koeficientov mortalitete za daljše razdobje zato lahko da napačno sliko v spremembah v umrljivosti prebivalstva. Spremenjena struktura proizvodnje tako v podjetju kot v večjih gospodarskih enotah ali v celotnem narodnem gospodarstvu odločilno vpliva na sumarne koeficiente o produktivnosti dela, o povprečni plači itd. Vpliv spremenjene strukture na določene pojave v časovni vrsti odpravimo s *standardizacijo* podatkov, to je z izračunavanjem koeficientov v stalni strukturi. O standardizaciji koeficientov je bilo več govora pri koeficientih.

12.31 Proizvodnjo, promet tako v zunanji ali notranji trgovini in druge pojave izražamo v vrednosti. Časovno primerjavo vrednostnih podatkov motijo spremembe v cenah, ker se vrednost proizvodnje, prometa itd. spreminja zaradi sprememb v obsegu in sprememb cen. Vpliv, ki ga povzročajo spremembe cen, odstranimo tako, da izražamo proizvodnjo, promet itd. v *stalnih cenah*. Vendar je ta način izračuna zamuden in težko izvedljiv, ker bi ga morali upoštevati že v osnovni evidenci. Vpliv cen popravimo tudi z indeksi cen tako, da podatke po tekočih cenah delimo z ustreznim indeksom cen. Ta način je operativno veliko lažji, je pa kljub temu, da je le dokaj zadovoljiv. Paziti je treba le, da je indeks, s katerim popravljamo podatke, ustrezen po vsebini in značaju.

Elementarni pokazatelji dinamike

12.32 Enostavno analiziramo dinamiko pojava s primerjavo členov v časovni vrsti. Najlaže primerjamo člene časovne vrste z indeksi. Zato z njimi izvedemo več elementarnih pokazateljev dinamike.

Raven Y_k je osnovna časovna vrsta. Že iz sprememb v ravni sklepamo na smer in jakost dinamike pojava.

Če raven časovne vrste izrazimo v primerjavi z nekim stalnim karakterističnim členom v časovni vrsti, dobimo vrsto *indeksov s stalno osnovo*.

$$I_{k/0} = 100\, Y_k / Y_0 \qquad (12.11)$$

Absolutna razlika

$$D_k = Y_k - Y_{k-1} \qquad (12.12)$$

pokaže v absolutnih vrednostih spremembo pojava od člena do člena.

Temo rasti

$$T_k = 100 \frac{Y_k - Y_{k-1}}{Y_{k-1}} = 100 \frac{D_k}{Y_{k-1}} \qquad (12.13)$$

pokaže relativno razliko od člena do člena

Koeficient dinamike

$$k_k = Y_k/Y_{k-1} \qquad (12.14)$$

pokaže v obliki koeficienta relativne spremembe od člena do člena.

Enak značaj ima *verižni indeks*

$$I_k = 100.Y_k/Y_{k-1} \qquad (12.15)$$

le da relativno spremembo nakaže v indeksu.

Ker so vsi ti pokazatelji izpeljani iz osnovne časovne vrste, so med njimi enostavne zveze.

$$T_k = 100.D_k/Y_{k-1} = 100(k_k - 1) = I_k - 100$$
$$I_k = 100.k_k \qquad (12.13)$$

V tabeli 12.11 je prikazano, kakšni so posamezni pokazatelji dinamike, če pojav raste, zastane ali pada.

Tabela 12.11 Vrednosti posameznih pokazateljev dinamike pri različnem gibanju pojava

	Pojav		
	raste	zastane	pada
Nivo Y_k	$Y_{k+1} > Y_k$	$Y_{k+1} = Y_k$	$Y_{k+1} < Y_k$
Indeks s stalno osnovo	$I_{k+1/o} > I_{k/o}$	$I_{k+1/o} = I_{k/o}$	$I_{k+1/o} < I_{k/o}$
Absolutna razlika	$D_k > 0$	$D_k = 0$	$D_k < 0$
Temp rasti	$T_k > 0$	$T_k = 0$	$T_k < 0$
Koeficient dinamike	$k_k > 1$	$k_k = 1$	$k_k < 1$
Verižni indeks	$I_k > 100$	$I_k = 100$	$I_k < 100$

12.33 Za svetovno proizvodnjo boksita imamo v tabeli 12.12 izračunane vse vrste elementarnih pokazateljev dinamike.

Tabela 12.12 Pokazatelji dinamike za svetovno proizvodnjo boksita. (v tisočih tonah; (Vir: SG 58)

Leto	Y_k	Y_k/o	D_k	T_k	k_k	I_k
1950	7.700	100				
1951	10.300	134	+2600	+33,8	1,34	134
1952	11.900	155	+1600	+15,5	1,16	116
1953	13.000	169	+1100	+ 9,2	1,09	109
1954	14.850	193	+1850	+14,2	1,14	114
1955	15.350	199	+ 500	+ 3,4	1,03	103
1956	16.900	219	+1550	+10,1	1,10	110

12.34 Nakazani pokazatelji dinamike pa niso edini pokazatelji, s katerimi prikazujemo dinamiko pojava. Za nekatere časovne vrste se izkažejo za primernejše druge količine. Izmed teh naj omenimo samo mesečne časovne vrste za pojave, ki so sezonskega značaja. Če je pojav sezonski, primerjava tekočega meseca s predhodnim mesecem nima logične povezave in smisla. Polj smiselno je, da primerjamo podatek tekočega meseca s podatkom iz predhodnega leta za isti mesec. Ta dva podatka sta si sorodnejša kakor zaporedna meseca. Ti indeksi imajo premično bazo, vendar niso verižni, ker osnova indeksov ni predhodni člen.

Tabela 12.13 Mesečni in kumulativni indeksi za proizvodnjo piva v letu 1957 v LRS (v tisoč hl; Vir: MSP)

Mesec	1956		1957		Indeksi	
	Y	kum.	Y	kum.	mesečni	kumulativni
J	1,9	0	2,3	0	121	0
F	0,8	1,9	3,6	2,3	450	121
M	2,8	2,7	6,3	5,9	225	218
A	3,7	5,5	11,3	12,2	169	222
M	13,1	12,2	15,5	23,5	118	193
J	13,3	25,3	28,8	39,0	216	154
J	21,2	38,3	28,9	67,8	136	176
A	18,6	59,8	24,1	96,7	130	161
S	11,6	78,4	11,5	120,8	99	154
O	3,7	90,0	5,5	132,3	149	147
N	1,6	93,7	4,2	137,8	263	147
D	2,3	95,3	5,2	142,0	226	149
Skupaj		97,6		147,2		151

Enako za take primere izračunavamo indekse iz kumulativnih vrednosti dveh zaporednih let. Ta primerjava je logična, ker obe primerjani vredno-

sti obsegata isti del sezone v dveh zaporednih letih. Za proizvodnjo piva v LRS v letu 1957 so te vrste pokazatelji izračunani v tabeli 12.13.

Časovna vrsta mesečnih indeksov med dvema letoma pokaže, da je – če izvzamemo mesece z nizko proizvodnjo piva – največje zvečanje v mesecu juniju (indeks: junij 1957/junij 1953 je 213). Indeks kumulativ pa kaže v drugi polovici leta precej stalno vrednost okrog 150.

Komponente gibanja v časovnih vrstah

12.35 Časovna vrsta je kvantitativen izraz časovnega delovanja vseh faktorjev, ki vplivajo na pojav, ki ga prikazuje. Teh faktorjev je veliko in se njihova jakost in učinek časovno spreminja. Zato se členi časovne vrste spreminjajo; pravimo, da se pojav giblje.

Nemogoče je iz časovne vrste izluščiti, kolikšna je sprememba zaradi posameznega faktorja. Moremo pa iz nje razbrati skupen učinek faktorjev, ki imajo soroden vpliv na pojav, ki ga proučujemo. Na časovni vrsti po tem vidiku opazujemo naslednje vrste sprememb:

a) *trend* – T, ki podaja osnovno linijo razvoja,
b) *ciklične spremembe* – C, ki izvirajo iz dolgoročnih vzrokov,
c) *periodične spremembe* – P, med katerimi so posebno važne sezonske – S, ki izvirajo iz vzrokov, ki se ponavljajo na stalno razdobje,
d) *iregularne spremembe* – I, ki so rezultat enkratnih epizodičnih dogodkov ali rezultat stalnih slučajnih vzrokov.

V časovni vrsti ne opazimo vedno vseh komponent. Ta ali ona komponenta more izpasti. Časovna vrsta letnih podatkov ne kaže sezonskega gibanja, marsikateri pojav nima cikličnih nihanj ali epizodičnih sprememb. Vsak pojav pa ima neko osnovno smer razvoja. Le redki so primeri konstantnih pojavov in še ti niso z ekonomskih področij. Na vsaki časovni vrsti opazimo tudi slučajna nihanja, ki so izraz manjših vzrokov, ki nastopajo v eni ali drugi obliki.

12.36 Trend. Vsak ekonomski pojav ima neko osnovno linijo razvoja. Ta je opazna le v daljših časovnih razdobjih. To osnovno smer razvoja imenujemo trend. Če trend nakazuje razvoj v daljšem razdobju, ga včasih imenujemo *sekularni trend*. Trend je rezultat faktorjev, ki teže k stalnemu razvoju. Tako je trend povečanja proizvodnje električne energije izraz tehničnega napredka, trend zniževanja umrljivosti rezultat razvoja zdravstvene službe in medicine nasploh itd. Trend proizvodnje električne energije kaže stalno naraščanje, trend gibanja umrljivosti pa stalno padanje. Trend za določen pojav more tudi spremeniti smer. Tako more proizvodnja nekega artikla določeno razdobje naraščati, v nadaljnjem razvoju pa zaradi nadomestitve z drugim, kvalitetnejšim artiklom, padati. Enako more proizvodnja rudnika za neko razdobje stalno rasti, v nadaljnjem razdobju pa zaradi osiromašenja rudnika padati. Kakor smo že navedli, je konstantnih pojavov

malo. Eden izmed njih je na primer razmerje med številom rojstev dečkov in deklic. To razmerje ima časovno isto vrednost. Odkloni od konstante so le slučajni in ne kažejo težnje k časovni spremembi tega razmerja. Ta primer je prikazan v drugem poglavju v tabeli 2.1.

Zaradi drugih vplivov se dejanske vrednosti od trenda odklanjajo navzgor in navzdol, vendar v večini primerov že iz slike za časovno vrsto razberemo osnovno smer razvoja.

12.37 Periodična nihanja. Za veliko pojavov je značilen periodičen, zlasti sezonski značaj. Sezonski značaj marsikaterih pojavov povzroče predvsem klimatski vplivi. Tako je s klimatskimi vplivi povezana gradbena dejavnost, turizem in gostinstvo, kmetijstvo in vse druge dejavnosti, ki so odvisne od klimatskih vplivov ali posredno povezane z eno izmed navedenih. Tako je prehrambena industrija vezana na poljedelstvo, promet na turizem itd. Sezonski značaj pa ne izvira nujno iz klimatskih vplivov, marveč more biti vzrok tudi drugje. Tako morejo vplivati sezonsko določene ustaljene navade ali prireditve na sezonski značaj. Tako je potniški promet, poštni promet, trgovinski promet itd. odvisen od praznikov, ki so v vsakem letu ob istem času (n.pr. Novo leto). Enako izzove sezonski značaj določenega pojava prireditev, ki se letno ponavlja na isti datum; tak primer so gospodarski velesejmi.

12.38 Ciklična gibanja. V časovnih vrstah za daljša razdobja opazimo nihanja okrog trenda. Ta nihanja so več ali manj regularna, ni pa niti njihova dolžina niti oblika stalna kakor pri periodičnih nihanjih. Ta nihanja pa kljub temu, da niso popolnoma regularna, niso slučajna in so odvisna od dogajanj v preteklosti. Imenujemo jih ciklična gibanja. Ciklična gibanja so tipična za ekonomske pojave in jih opazujemo na nizu pojavov. Jakost cikličnih nihanj je za različne pojave različna. Odvisna je od tega, ali gre za pojav, ki je bolj ali manj občutljiv za spremembe v gospodarskem življenju. Zaradi povezanosti gospodarstva je med cikli za posamezne gospodarske panoge ali pojave zveza, ker se tako prosperiteta ali depresija prenaša od panoge na panogo.

12.39 Iregularne variacije. Razen navedenih treh vrst gibanj, trenda, periodičnih nihanj in cikličnih sprememb na časovni vrsti opazujemo še iregularne variacije, ki so rezultat enkratnih ali pa slučajnih vzrokov. Zaradi prekinitve toka nastane zastoj v prometu električne cestne železnice. Potres vpliva na niz dejavnosti. Zaradi prometne nesreče je zastoj v železniškem prometu, zaradi poplave je uničena letina, epidemija vpliva na morbiditeto in mortaliteto itd. Navedeni dogodki so *enkratni vplivi*, ki imajo za rezultat, da se za krajše razdobje pojav odkloni od regularnega razvoja.

Za razliko od enkratnih - epizodičnih vplivov, ki nastopijo nepričakovano, vendar za krajši čas, *slučajne variacije*, ki so rezultat slučajnih vplivov, nastopajo stalno in v vseh pojavih. Slučajne variacije so manjše fluktuacije, rezultat manjših dogodkov, ki jih ne obravnavamo, ker ni za to potrebe ali pa jih niti ne moremo obravnavati individualno.

Osnovni modeli časovnih vrst

12.40 V vsaki časovni vrsti ne opazimo vseh komponent, bodisi da te komponente pojav sploh ne vsebuje ali pa je v časovni vrsti zabrisana. Marsikateri pojav nima sezonske oziroma periodične komponente, v nekaterih časovnih vrstah pa je zabrisana. Tako na dnevni časovni vrsti novorojenih opazimo tedenske periodičnosti, medtem ko je v tedenski vrsti števila prevoženih potnikov zabrisana tedenska periodičnost. Enako je v vrsti letnih podatkov zabrisano sezonsko gibanje pojava. Marsikateri pojav nima cikličnih nihanj ali epizodičnih sprememb. Skoraj vsak pojav pa ima svoj trend. Le malo je pojavov, ki bi bili konstante v tem smislu, da nimajo neke smeri razvoja. Na vsaki časovni vrsti moremo opaziti tudi slučajne variacije, ki so izraz manjših vzrokov, ki vedno nastopajo.

Vrednosti členov v časovni vrsti Y so torej skupnost *trenda* T, *cikličnih gibanj* C, *sezonskih* oziroma *periodičnih vplivov* P in *iregularnih variacij*, ki so *epizodične* E ali *slučajne* S.

Zvezo teh komponent moremo nakazati v nekaj osnovnih modelih, ki jih uporabljamo pri analizi časovnih vrst.

$$
\begin{aligned}
&1.)\quad Y = T + P + C + E + S \\
&2.)\quad Y = T(1 + p + c + e + s) \\
&3.)\quad Y = T \,.\, P \,.\, C \,.\, E \,.\, S \\
&4.)\quad Y = T \,.\, P \,.\, C \,.\, E + S
\end{aligned}
\qquad (12.17)
$$

V prvem modelu so posamezne komponente vezane aditivno. Enako je to primer v drugem modelu, vendar je za ta primer učinek vsake komponente sorazmeren s trendom. Tretji model ima za osnovo faktorialno povezavo vseh komponent, medtem ko četrti predpostavlja aditivno slučajno komponento, druge komponente pa so vezane faktorialno.

Če v stvarnem primeru, ki ga proučujemo, kakšne komponente ni, ustrezni simbol v modelu izpade. V prvih dveh modelih so periodične in ciklične komponente in iregularne variacije (epizodične in slučajne) pozitivne ali negativne količine. Če učinka kake komponente ni, je njihova vrednost enaka 0. V prvem modelu so posamezne količine (P, C, E, S) izražene z absolutnimi, v drugem (p, c, e, s) pa z relativnimi odkloni od trenda. V tretjem in četrtem modelu so komponente P, C, E, S izražene s koeficienti, ki so ena, če učinka ni, in večji ali manjši od ene, če deluje v smeri povečanja oziroma zmanjšanja. V stvarnosti prvi model srečamo poredko, čeprav je najenostavnejši. Težko si namreč predstavljamo, da bi bila velikost učinka posameznih komponent neodvisna od velikosti pojava, ki je dan s trendom. Vendar je analiza prvega modela najenostavnejša. Zato skušamo druge modele privesti na prvega, če se le da. Drugi model preuredimo v prvega, če osnovno časovno vrsto Y delimo s trendom, tretji model pa prevedemo v prvega, če ga logaritmiramo.

Osnovna analiza časovnih vrst gre za tem, da razdružimo osnovno časovno vrsto na njene sestavne dele. S tem prodremo že precej v zakoni-

tosti dinamike pojava. Te zakonitosti izkoriščamo, kadar analiziramo dogajanja v preteklosti ali napovedujemo razvoj v prihodnosti. Z različnimi analitičnimi postopki namreč moremo določiti posamezno izmed teh komponent v časovni vrsti. Vendar moremo časovno vrsto analizirati le, dokler v kompleksu vplivov ni kakovostnih sprememb. To je jasno. Ako na primer neko novo odkritje da nove proizvodne možnosti, se smer razvoja bistveno spremeni. Ker imamo za take primere dve smeri razvoja, moramo časovno vrsto razdeliti v dva dela in ju obravnavati ločeno. Enako analiziramo sezonsko komponento kot statično le, dokler moremo predpostavljati, da je ta nespremenjena. Če se v proučevanem razdobju sezonska komponenta spremeni, moramo časovno vrsto razdeliti in analizirati sezonsko komponento za vsak odsek posebej.

Vloga povprečij za analizo časovnih vrst

12.41 Enostavno orodje, ki pomaga analizirati časovne vrste, so povprečja. Pod določenimi pogoji moremo s povprečji odstraniti neke komponente v časovni vrsti, druge pa ohraniti. Povprečja iz časovnih vrst vplivajo v eni ali drugi obliki na vse komponente.

Če je *trend* v razmaku, iz katerega izračunavamo povprečje, linearen, povprečje, ki ga centriramo na sredino razmaka, leži na trendu. Če je trend na odseku, iz katerega tvorimo povprečje, krivuljčen, povprečje, ki ga centriramo na sredino razmaka, stoji nad trendom, če je trend konkaven, in pod trendom, če je trend konveksen. Razlike so tem večje, čimbolj je trend ukrivljen. Povprečje, centrirano na obratišče, pa leži na trendu. Iz tega dalje sklepamo, da se časovna vrsta drsečih sredin, če jo izračunamo iz trenda, na odseku, na katerem je trend linearen, sklada s trendom, če pa je trend krivuljčen, dobimo črto, ki je bolj izravnana kot trend. Ker je krivina trenda na daljših odsekih večja, je izravnavanje trenda tem večje, čim širši so razmaki, za katere izračunavamo povprečja.

Izravnavanje ni tako izdatno, če člene, iz katerih izračunavamo povprečja, tehtamo tako, da imajo stranski členi manjšo, *srednji* pa večjo težo. Običajno tehtamo člene z binomskimi koeficienti (1, 2, 1) za povprečja iz treh členov, (1, 3, 3, 1) za povprečja iz štirih členov, (1, 4, 3, 4, 1) za povprečja iz petih členov. Izdelani so še različni drugi sistemi tehtanja. To pa presega naš okvir.

Če je zveza med komponentami aditivna, se učinek *periodičnih* oziroma *sezonskih vplivov* v povprečju uniči, če je dolžina razdobja, za katerega računamo povprečje, enako periodi ali mnogokratniku periode. Pri faktorialnih zvezah, kakor je nakazana v modelu 3 v 12.17, pa se v povprečju uničijo logaritmi sezonske komponente.

V vrsti drsečih sredin se uniči *ciklična komponenta*, če so povprečja izračunana iz razdobja celotnega cikla.

Učinki *slučajnih vplivov* se v povprečju vedno bolj manjšajo, čim daljše je razdobje, iz katerega izračunavamo povprečja.

Če ta pravila izkoriščamo kombinirano, povprečje veliko pripomore k analizi časovnih vrst.

TREND

12.43 Trend proučujemo iz dveh razlogov. Proučujemo ga samega zase, da spoznamo smer razvoja pojava, da ga primerjamo s trendi pojavov, ki so med seboj odvisni ali pa da proučujemo vpliv trenda na druge komponente (periodične in ciklične). Trend pa dostikrat iščemo zato, da iz odklonov stvarne časovne vrste od trenda proučujemo sezonska, ciklična ali iregularna nihanja.

Metod za določevanje trenda je več. Vsaka izmed njih ima za osnovo različne predpostavke in postopke. Zato je važna in odgovorna naloga izbrati v vsakem posebnem primeru primerno obliko in metodo za določanje trenda. Trend določamo z mehaničnimi sredstvi prostoročno ali z izravnavanjem s sredinami, ali analitično, s prilagajanjem analitičnih funkcij danim podatkom.

Katero metodo uporabimo za določanje trenda, je odvisno od proučevane časovne vrste in namena analize. Če potrebujemo trend zaradi proučevanja cikličnih ali sezonskih odklonov, je logično, da predpostavljamo, da je trend črta, ki poteka med realnimi vrednostmi tako, da se izenačujejo periodični ali ciklični vplivi. To črto določamo običajno mehanično. Če pa potrebujemo trend za napovedovanje, ga iščemo v obliki analitične krivulje, za katero moremo z ekstrapolacijo izračunati potek trenda tudi v prihodnost.

12.44 Najvažnejši problem je izbira pravilnega tipa krivulje, ki naj predstavlja trend. Enoličnega pravila za določevanje trenda ni. Splošne smernice predrisujejo, da mora biti krivulja, ki jo uporabimo za trend, smiselna, ne preveč zamotana in objektivna.

Po sliki časovne vrste, poznavanju pojava, ki ga vrsta prikazuje in poznavanju lastnosti funkcij, ki pridejo v poštev, izberemo v vsakem določenem primeru, kateri tip krivulje je vsebinsko in tehnično najprimernejši.

Kot trend uporabljamo ali iregularno prilagojene krivulje ali analitične krivulje. Od splošnih analitičnih krivulj pridejo najpogosteje v poštev tele funkcije:

1. Premica: $T = a + bx$
2. Parabola druge, tretje ... stopnje: $T = a + bx + cx^2 + dx^3 \ldots$
3. Parabola: $T = a + b\sqrt{x}$
4. Eksponencialna funkcija: $T = ab^x$
5. Modificirana eksponencialna funkcija: $T = k + ab^x$

(12.18)

3. Gompertzova krivulja: $T = Ka^{b^x}$

7. Pearl-Reedova logistična krivulja

$$T = \frac{T_\infty}{1 + e^{a+bx}} \qquad (12.18)$$

V teh funkcijah pomeni T trend, x čas, a, b, c, d, k, T_∞ pa so parametri trenda.

12.45 Določitev trenda ima dva dela. Najprej je treba *izbrati ustrezen tip krivulje*, ko pa tega imamo, je treba za konkretno časovno vrsto *določiti parametre* trenda. Prvi del je vsebinske, drugi pa tehnične narave.

Da se moremo odločiti v vsakem posebnem primeru za ustrezen tip krivulje, moramo poznati potek posameznih tipov krivulj in njihove lastnosti. Zato je najbolje, da časovno vrsto, za katero določamo trend, narišemo v linijski grafikon z aritmetičnimi ali logaritemskimi skalami in presodimo, kateri tip krivulje najbolje ustreza stvarni smeri razvoja.

O obliki trenda se moremo poučiti tudi iz vrst diferenc zaporednih členov. Če je trend linearen, so prve diference $DT_k = T_{k+1} - T_k$ stalne količine. Če je trend parabola druge stopnje (tipa 2), so druge diference, ki jih dobimo iz vrste prvih diferenc $D^2T_k = DT_{k+1} - DT_k$, konstantne. Enako ugotovimo, da je trend parabola tretje stopnje, če so tretje diference $D^3T_k = D^2T_{k+1} - D^2T_k$, dobljene iz vrste drugih diferenc, konstantne.

Če je trend eksponencialna funkcija, so prve diference logaritmov trenda konstante. Enako preizkusimo, ali je trend eksponencialen, z verižnimi indeksi ali z vrsto koeficientov dinamike. Če je trend eksponencialen, so verižni indeksi ali koeficienti dinamike konstantne količine. Relativen prirastek je namreč pri eksponencialni funkciji konstanten.

Če se pojav giblje v obliki Pearl-Reedove krivulje, tvorijo prve diference simetrično unimodalno krivuljo. Pri Gompertzovi krivulji pa dobimo unimodalno krivuljo, ki je asimetrična v desno. Pri logistični krivulji relativen prirastek linearno pada.

Jasno je, da zgornja pravila veljajo za trend, ne pa za stvarno časovno vrsto. Ta ne vsebuje samo trenda, ampak tudi druge komponente. Zato so časovne vrste povprečij, če so izračunana tako, da odstranijo periodične in ciklične vplive in zmanjšajo slučajne variacije, veliko primernejše za zgornje preizkuse.

12.46 Za tehnično izračunavanje parametrov trenda je pri vseh metodah najprimernejše, da so parametri v linearni zvezi. Za prve tri tipe krivulj to velja, ne velja pa za druge. Vendar moremo eksponencialno krivuljo privesti na obliko, za katero to ne velja, če jo logaritmiramo

$$\log T = \log a + x \log b \qquad (12.19)$$

ker moremo pisati

$$T' = a' + b'x \tag{12.20}$$

pri tem je $T' = \log T$; $a' = \log a$; $b' = \log b$

Modificirane eksponencialne krivulje ne moremo privesti v želeno obliko, vendar moremo bolj zamotani funkciji, logistično in Gompertzovo funkcijo, prevesti v modificirano eksponencialno funkcijo.

Če vzamemo reciprok logistične funkcije, dobimo

$$\frac{1}{T} = \frac{1}{T_\infty} + \frac{e^a}{T_\infty} e^{bx} \tag{12.21}$$

ali dalje:

$$T' = k' + a'b'^x \tag{12.22}$$

Pri tem pomeni: $T' = \frac{1}{T}$; $k' = \frac{1}{T_\infty}$; $a' = \frac{e^a}{T_\infty}$; $b' = e^b$

Če logaritmiramo Gompertzovo funkcijo, pa dobimo:

$$\log T = \log K + b^x \log a \tag{12.23}$$

ali

$$T' = k' + a'b'^x \tag{12.24}$$

Pri tem pomeni: $T' = \log T$; $k' = \log K$; $a' = \log a$; $b' = b$

Medtem ko je eksponencialna funkcija zelo primerna funkcija za smer razvoja za razdobja, ki niso predolga, se izkaže, da sta logistična in Gompertzova krivulja primerni za prikazovanje naravnega razvoja za zelo dolga razdobja. Za zelo dolga razdobja opazimo, da se v logistični krivulji spreminja število prebivalstva. Enake zakonitosti opazujemo tudi pri drugih živih bitjih. Ugotovili so, da moremo logistično in Gompertzovo krivuljo s pridom uporabiti tudi za trend proizvodnje. Logistična ali Gompertzova krivulja, s katero ponazorimo trend proizvodnje, nakazuje štiri obdobja: obdobje nizke proizvodnje pomeni stopnjo eksperimentiranja in uvajanja, drugo obdobje, ki pomeni vzpon v družbeno proizvodnjo, tretje obdobje po infleksiji, ko proizvodnja raste, vendar v vedno manjšem obsegu, in zadnje obdobje stabilizacije pri nasičenosti.

12.47 Splošno velja pravilo, da moremo trend izračunavati le za daljša razdobja. V nasprotnem primeru sicer dobimo prilagojeno krivuljo, vendar pa ta ni trend, marveč precej verjetno vsebuje predvsem ostanek ciklov, pa tudi drugih vplivov. Tudi pri daljših razdobjih moramo paziti, da je razdobje, za katerega določamo trend, tako, da vključuje cele cikle, ne pa samo dele. Če ta pogoj ni izpolnjen, more nezaključeni cikel znatno

vplivati, da dobljena krivulja ni verna slika osnovne smeri razvoja.

Kljub tej rezervi dostikrat prilagajamo krivulje tudi krajšim časovnim vrstam. Vendar moramo biti v teh primerih oprezni pri tolmačenju take smeri razvoja, saj je kratkoročna in more vsebovati razen trenda tudi ciklična nihanja in druge vplive.

12.48 Metode določanja trenda. Omenili smo že, da imamo več metod za določanje trenda. Od njih bomo obravnavali naslednje:

a) prostoročno metodo,

b) metodo drsečih sredin,

c) metodo sredin med najvišjimi in najnižjimi točkami.

Te metode posredujejo trend v mehanični ali grafični obliki. Od metod, ki dajejo trend analitično v obliki krivulj, pa bomo obravnavali:

a) metodo izbranih točk,

b) metodo delnih sredin,

c) metodo delnih vsot,

d) metodo najmanjših kvadratov.

Prostoročna metoda

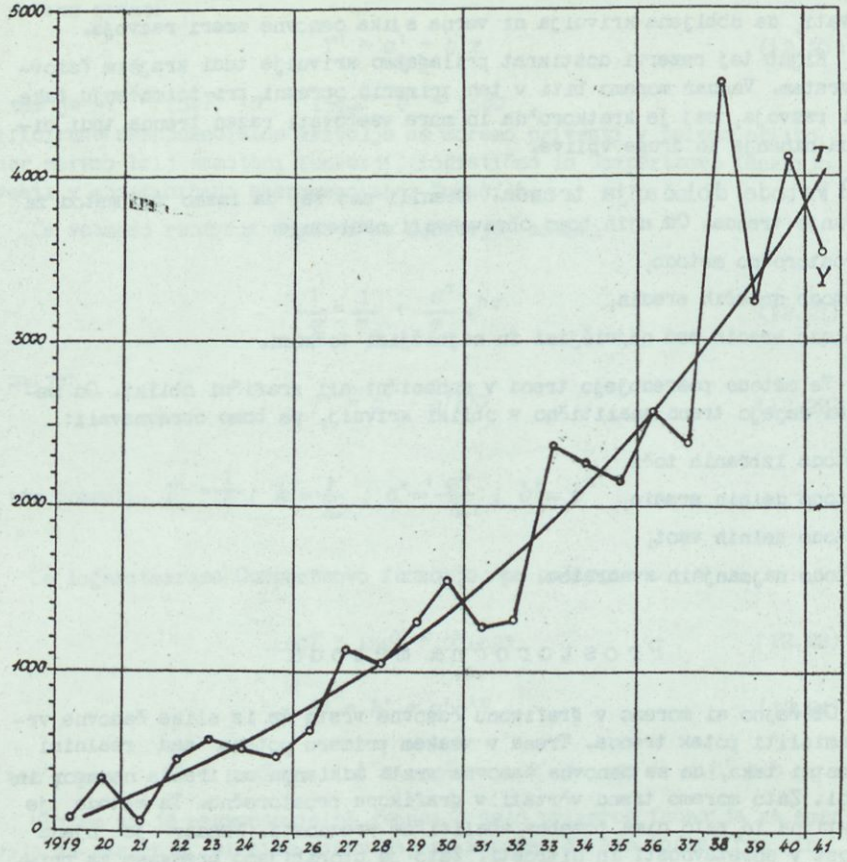
12.49 Običajno si moremo v grafikonu časovne vrste že iz slike časovne vrste zamisliti potek trenda. Trend v vsakem primeru poteka med realnimi vrednostmi tako, da se osnovna časovna vrsta odklanja od trenda navzgor in navzdol. Zato moremo trend vrtati v grafikonu prostoročno. Ta metoda je subjektivna in zato nima posebne analitične vrednosti. Vendar je njena prednost v enostavnosti in hitrosti. Zato jo uporabljamo predvsem za približek in za osnovo pri izbiri analitične oblike trenda.

12.50 V sliki 12.10 je narisana časovna vrsta proizvodnje cinkovega koncentrata na območju LRS v razdobju med obema vojnama od 1919-1941, v njej pa prostoročno vrtan trend.

Tabela 12.14 Proizvodnja cinkovega koncentrata v LRS med obema vojnama v tonah. (Vir: SB LRS 1952 št.8)

Leto	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931
Proizvodnja	90	350	75	450	600	520	510	639	1148	1076	1818	1547	1301

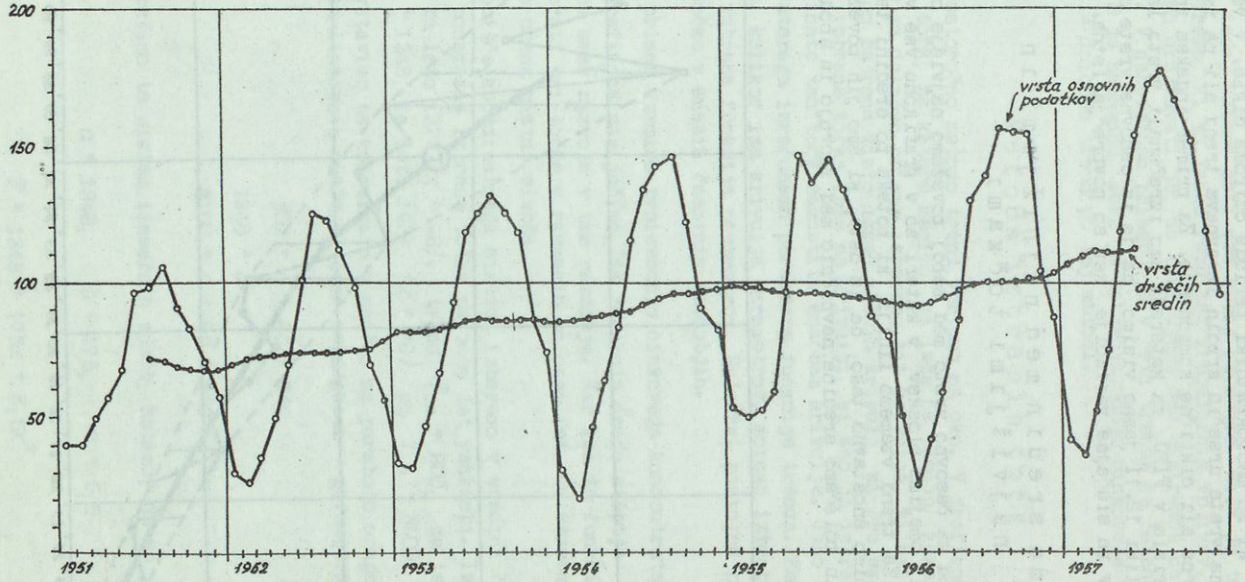
Leto	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
Proizvodnja	1358	2898	2282	2178	2537	2401	4530	3252	4121	3483



Slika 12.10 Trend za proizvodnjo cinkovega koncentrata v LRS med obema vojnama, določen prostoročno

Metoda drsečih sredin

12.51 Iz lastnosti sredin, ki smo jih že navedli, sklepamo, da dá vrsta drsečih sredin približno linijo trenda, če je trend linearen. Če pa je trend krivuljčen, se v vrsti drsečih sredin trend bolj ali manj izravnava. Stopnja izravnavanja je odvisna od števila členov, iz katerih izračunavamo povprečje. Kot primer vzemimo časovno vrsto števila umrlih na 1000 prebivalcev v bivši Jugoslaviji in vrsto petletnih drsečih sredin. Slika 12.11 pokaže, kako se vrsta drsečih sredin giblje med členi osnovne vrste. To pa je osnovni pogoj, ki ga mora izpolniti trend. Vendar je vprašanje, ali je v tem primeru vrsta drsečih sredin res trend, ali pa je v njem ostala sled eventualnih ciklov.

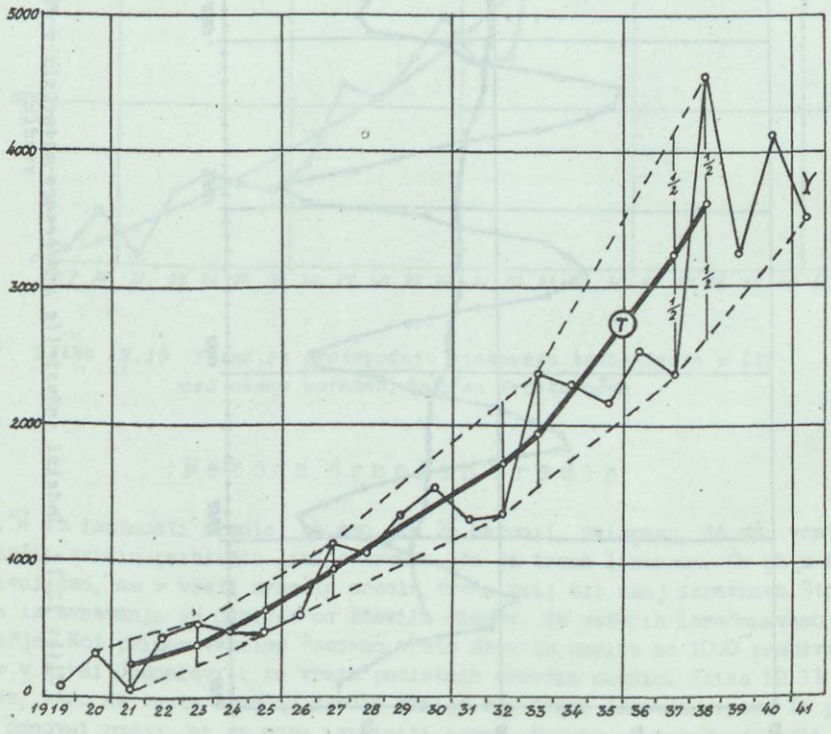


Slika 12.11 Indeksi industrije gradbenega materiala v FLRJ z vrisano časovno vrsto drsečih sredin

V časovnih vrstah, ki vsebujejo periodično in ciklično komponento, da vrsta drsečih sredin trend le, če je trend linearen, povprečja pa izračunana za razdobja, ki so mnogokratniki periode oziroma cikla. V vseh drugih primerih časovna vrsta drsečih sredin izravnava trend ali pa ima v sebi ostanek periodične ali ciklične komponente. Za primer indeksa industrije gradbenega materiala v FLRJ, za katerega smo izračunali vrsto letnih drsečih sredin, na sliki 12.11 jasno vidimo, da je iz osnovne vrste odstranjen sezonski vpliv in slučajne variacije, ker so povprečja letna.

Metoda sredin med najnižjimi in najvišjimi točkami

12.52 Če v grafikonu za časovno vrsto med seboj povežemo najvišje osti in enako najnižje osti, omejimo ploskev, v kateri so v grafikonu vse vrednosti časovne vrste. Za trend vzamemo linijo, ki poteka po sredini te ploskve. To linijo dobimo enostavno tako, da iz osti, ki smo jih povezali, narišemo navpičnice in poiščemo sredine navpičnic med zgornjo in spodnjo mejo



Slika 12.12 Trend za proizvodnjo cinkovega koncentrata, določen po metodi sredin med najvišjimi in najnižjimi točkami

ploskve. Te sredine med seboj zvežemo. Tako dobimo lomljeno črto, ki zadosti dobro prikazuje osnovno smer razvoja. Ta metoda je subjektivna, ker je odvisna od tega, katere osti med seboj povežemo. Hiba je tudi ta, da povezujemo ekstreme, ki so lahko rezultat iregularnih vplivov. Splošno pravilo, ki ga moramo upoštevati, je, da povezujemo res ekstremne ne pa manjše - lokalne osti. V nasprotnem primeru je v črti trenda še ostanek ciklov. Na sliki 12.12 je nakazano, kako je za proizvodnjo cinkovega koncentrata iz tabele 12.14 določen trend po tej metodi.

Metoda izbranih točk

12.53 S prostoročno metodo trend včrtamo na oko. V grafikonu je laže določiti nekaj točk, za katere menimo, da leže na trendu, kakor celo krivuljo trenda. Če za določen primer izberemo tip krivulje, ki naj predstavlja trend, imamo za trend tisto izmed možnih krivulj, ki gre skozi izbrane točke. Če naj izbrane točke leže na krivulji, ki smo jo izbrali za trend, morajo koordinate teh točk zadostiti enačbam krivulje. Tako dobimo sistem enačb, iz katerega izračunamo parametre funkcije trenda. Ker potrebujemo toliko enačb, kolikor ima krivulja parametrov, moramo izbrati toliko točk, kolikor ima izbrana funkcija parametrov. Računski postopek poenostavimo, če izberemo točke v enakih časovnih razdobjih.

12.54 Kot primer vzemimo proizvodnjo cinkovega koncentrata iz tabele 12.14.

Iz grafikona zaključimo, da parabola druge stopnje $T = a + bx + cx^2$ dobro opiše smer razvoja v tem razdobju. Ker ima ta funkcija tri parametre, moramo izbrati tri točke v razmakih po osem let. Če vzamemo točke bližje skupaj, je ocena znatno slabša.

Izhodišče koordinatnega sistema izberemo v sredini časovne vrste (leto 1930), posamezne izbrane točke pa osem let vsaksebi. Iz grafikona ocenimo trend: za leto 1922 $(x = -8)$, trend $T_{-8} = 300$, za leto 1930 $(x = 0)$ trend $T_0 = 1348$, za leto 1939 $(x = +8)$ pa $T_{+8} = 3100$.

Če vstavimo te koordinate v enačbo za parabolo druge stopnje, dobimo sistem treh linearnih enačb s tremi neznankami - parametri a, b, in c.

$$300 = a - 8b + 64c$$
$$1348 = a$$
$$3100 = a + 8b + 64c$$

Če rešimo ta sistem linearnih enačb, dobimo, da so parametri

$$a = 1348, \quad b = 175, \quad c = 5{,}5$$

ali trend

$$T = 1348 + 175x + 5{,}5x^2$$

Če v to funkcijo vnašamo posamezna leta ustrezne vrednosti x od -11 do $+11$, dobimo vsakemu členu osnovne vrste prirejeno vrednost trenda.

Za posamezna leta so izračunane vrednosti trenda dane v tabeli 12.15.

Tabela 12.15 Trend za proizvodnjo cinkovega koncentrata, določen po metodi izbranih točk

Leto	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931
x	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	−0	+1
T	89	148	219	300	392	496	611	738	872	1020	1178	1348	1528

Leto	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
x	+2	+3	+4	+5	+3	+7	+8	+9	+10	+11
T	1720	1922	2138	2361	2596	2842	3100	3369	3648	3939

Metoda delnih sredin

12.55 Tudi če je trend krivuljčen, moremo na krajšem odseku časovnega razvoja meniti, da je trend linearen. Navedli smo že, da v primeru linearnega trenda povprečje vrednosti členov leži na trendu. Zato moremo priti do objektivnejše izbranih točk na trendu, če namesto na oko izbranih točk vzamemo kot točke na trendu povprečja iz določenega števila členov. Po tej metodi razdelimo časovno vrsto na toliko enakih delov, kolikor ima funkcija trenda parametrov. Za vsak dobljeni del časovne vrste izračunamo povprečja. Za točke, ki imajo za abscisо sredino časovnega razmaka, iz katerega izračunamo povprečje, za ordinato pa izračunano povprečje, menimo, da leže na trendu. Od tu dalje je postopek enak postopku po metodi izbranih točk. Koordinate točk vstavimo v funkcijo trenda, da dobimo sistem enačb za parametre. Iz njih izračunamo parametre, ki jih vstavimo v funkcijo trenda. Z vstavljanjem vrednosti x za posamezne momente pa dobimo časovno vrsto trenda.

Čeprav je navedena metoda objektivnejša kakor metoda izbranih točk, je njena slaba stran v tem, da točke leže na trendu le, če je trend na ustreznem delnem odseku linearen. Če pa ni, da postopek delnih sredin preveč izravnan trend. Zato razdobja, iz katerih računamo sredine, ne smejo biti predolga. Če pa vzamemo prekratka razdobja, se preveč pokažejo slučajni vplivi. Treba je torej paziti, da spravimo v sklad oba učinka in izberemo pravilno dolžino.

12.56 Vzemimo za primer časovno vrsto proizvodnje cinkovega koncentrata v LRS, za katero smo trend določili že po metodi izbranih točk. Po analizi

podatkov menimo, da je trend parabola druge stopnje. Potrebujemo torej tri sredine, ker ima parabola druge stopnje tri parametre.

Ker ima vrsta 23 členov, kar ni deljivo s tri, povprečja izračunamo iz sedmih členov tako, da med njimi izpustimo po en člen. Tako dobimo simetrično sliko členov 7 + 1 + 7 + 1 + 7 = 23. Če vstavimo izhodišče za koordinatni sistem v sredino vrste, so abscise za povprečja: −8, 0, +8, ordinate pa povprečja ustreznih členov osnovne vrste. Tako dobimo:

$$T_{-8} = \frac{1}{7}(90+350+75+450+600+520+510) = 371$$

$$T_0 = \frac{1}{7}(1143+1076+1313+1547+1301+1358+2393) = 1462$$

$$T_{+8} = \frac{1}{7}(2173+2567+2401+4530+3252+4121+3483) = 3218$$

Koordinate točk $I_{-8}(-8, 371)$, $T_0(0, 1432)$, $T_{+8}(+8, 3218)$ vnesemo v funkcijo

$$T = a + bx + cx^2.$$

Tako dobimo tri enačbe:

$$371 = a - 8b + 64c$$
$$1432 = a$$
$$3218 = a + 8b + 64c$$

Iz tega sistema enačb zlahka dobimo, da je: $a = 1432$, $b = 177,94$, $c = 5,1953$. Analitična oblika trenda je torej

$$T = 1462 + 177,94x + 5,1953x^2$$

Če po prejšnjem načinu v to funkcijo vnašamo posamezne vrednosti za x, dobimo vrsto trenda v tabeli 12.16.

Tabela 12.16 Trend za proizvodnjo cinkovega koncentrata v Sloveniji, določen po metodi delnih sredin

Leto	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930
x	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0
T	133	202	281	371	471	581	702	833	975	1127	1289	1462

Leto	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
x	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11
T	1645	1839	2043	2257	2482	2717	2962	3218	3484	3761	4048

Metoda delnih vsot

12.57 Slaba stran metode sredin je izravnavanje trenda. Temu se izognemo z metodo delnih vsot. Ta metoda predpostavlja: a) v vsoti se rezultati slučajnih vplivov uničujejo; b) vsota rezultatov periodičnih vplivov v eni periodi je enaka nič; c) vsota rezultatov cikličnih vplivov v enem ciklu je enaka nič.

Če veljajo te predpostavke, je vsota členov časovne vrste v razmaku, ki je mnogokratnik periode in cikla, samo rezultat trenda. Če je na primer med komponentami aditivna zveza

$$Y = T + P + C + S$$

velja torej

$$\Sigma Y = \Sigma T + 0 + 0 + 0$$

Če je $T = af_1 + bf_2 + cf_3 + \ldots$ funkcija trenda, $a, b, c \ldots$ parametri, f_1, f_2, f_3 pa funkcije časa, dobimo sistem linearnih enačb za parametre $a, b, c \ldots$ če vzamemo delne vsote členov iz časovne vrste.

Število vsot se ravna po številu parametrov. Če imamo na primer tri parametre, dobimo, da je

$$\begin{aligned} {}_1\Sigma Y &= a{}_1\Sigma f_1 + b{}_1\Sigma f_2 + c{}_1\Sigma f_3 \\ {}_2\Sigma Y &= a{}_2\Sigma f_1 + b{}_2\Sigma f_2 + c{}_2\Sigma f_3 \\ {}_3\Sigma Y &= a{}_3\Sigma f_1 + b{}_3\Sigma f_2 + c{}_3\Sigma f_3 \end{aligned} \qquad (12.25)$$

Pri tem pomeni: ${}_1\Sigma Y, {}_2\Sigma Y, {}_3\Sigma Y$ = delne vsote členov v časovni vrsti ${}_1\Sigma f_1$, ${}_2\Sigma f_2 \ldots {}_3\Sigma f_3$ = delne vsote funkcij časa. Nadaljnji potek določanja funkcije trenda je enak kakor pri prejšnjih metodah. Iz enačb izračunamo parametre in jih vnesemo v funkcijsko obliko trenda. Iz te izračunamo časovno vrsto trenda, če vanjo vnašamo posamezne vrednosti za x.

12.58 Za proizvodnjo cinkovega koncentrata v Sloveniji je trend parabola druge stopnje: $T = a + bx + cx^2$.

Torej je: $f_1 = 1, \quad f_2 = x, \quad f_3 = x^2$.

Izračun parametrov in vrste trenda je nakazan v tabeli 12.17.

Tabela 12.17 Izračunavanje parabolicnega trenda druge stopnje za proizvodnjo cinkovega koncentrata v LRS med obema vojnama, po metodi delnih vsot

Leto (1)	x (2)	x^2 (3)	y (4)	$b + cx$ (5)	$x(b+cx)$ (6)=(5)x	$T=a+x(b+cx)$ (7)=a+(6)
1919	-11	121	90	118,299	-1301,29	124
1920	-10	100	350	128,728	-1287,28	188
1921	- 9	81	75	129,147	-1162,32	263
1922	- 8	64	450	134,571	-1076,57	349
1923	- 7	49	600	139,995	- 979,97	446
1924	- 6	36	520	145,419	- 872,51	553
1925	- 5	25	510	150,843	- 754,21	671
$\sum_1$ 7	-56	476	2595	-	-	2594
1926	- 4		339	156,238	- 325,07	801
1927	- 3	9	1143	161,692	- 485,08	941
1928	- 2	4	1076	167,116	- 334,23	1091
1929	- 1	1	1318	172,540	- 172,54	1253
1930	0	0	1547	177,934	- 0	1426
1931	+ 1	1	1301	183,388	+ 183,39	1609
1932	+ 2	4	1358	188,812	+ 377,62	1803
1933	+ 3	9	2393	194,236	+ 582,71	2008
$\sum_2$ 7	0	28	10131	-	-	10131
1934	+ 4		2282	199,660	+ 798,64	2224
1935	+ 5	25	2173	205,084	+1025,42	2451
1936	+ 6	36	2567	210,509	+1263,05	2689
1937	+ 7	49	2401	215,933	+1511,53	2937
1938	+ 8	64	4530	221,357	+1770,86	3196
1939	+ 9	81	3252	226,781	+2041,03	3467
1940	+10	100	4121	232,205	+2322,05	3748
1941	+11	121	3483	237,629	+2613,92	4040
$\sum_3$ 7	+56	476	22527	-	-	22528

Kakor pri prejšnji metodi smo tudi tu izpustili leti 1923 in 1934, da dobimo trikrat po sedem členov. Sistem enačb, iz katerega izračunamo parametre, sestavimo iz vsot, ki so izračunane v stolpcih 2, 3, 4.

$$7a - 56b + 476c = 2595$$
$$7a + 28c = 10131$$
$$7a + 56b + 476c = 22527$$

Iz tega sistema linearnih enačb dobimo, da je:

$$a = 1425,59, \quad b = 177,964, \quad c = 5,4241$$

V stolpcih 5, 6 in 7 je nakazano, kako izračunamo časovno vrsto trenda.

Kontrola računa je pravilo, da so pri metodi vsot delne vsote izračunanega trenda enake ustreznim delnim vsotam v osnovni časovni vrsti. V našem primeru se vsote natančno skladajo, razen zadnje, pri kateri je razlika zaradi zaokroževanja 1.

Metoda najmanjših kvadratov

12.59 V splošnem vzamemo za trend tisto krivuljo določenega tipa, ki se dani osnovni vrsti podatkov najbolje prilega. Če po analizi podatkov ugotovimo, da je trend parabola druge stopnje, kot trend izmed vseh parabol vzemimo tisto, ki se osnovni vrsti najbolj prilega.

Običajno vzamemo za merilo stopnje prilagojenosti vsoto kvadratov odklonov osnovne vrste od krivulje. Čim bolje je krivulja prilagojena osnovnim podatkom, tem manjša je ta vsota in narobe. Kot trend v tem primeru vzamemo tisto krivuljo, za katero je vsota kvadratov odklonov najmanjša.

Ta pogoj moremo pisati v obliki:

$$\Sigma(Y - T)^2 = \text{Min} \tag{12.26}$$

Ta vsota je funkcija parametrov krivulje, ki smo jo izbrali kot trend. Naloga je najti, pri katerih vrednostih parametrov je ta izraz minimalen.

Če je trend funkcija, v kateri so parametri v aditivni zvezi

$$T = af_1 + bf_2 + cf_3 \quad \ldots \tag{12.27}$$

f_1, f_2, f_3 = funkcije x,

pogoj iz obrazca 12.23 napišemo v konkretnejši obliki

$$\Sigma(Y - af_1 - bf_2 - cf_3 \ldots)^2 = \text{Min} \tag{12.28}$$

Po znanih stavkih o ekstremih dobimo s parcialnim odvajanjem po a, b, c, ... sistem linearnih enačb. Tem enačbam, ki jih imenujemo *normalne enačbe*, morajo zadostiti parametri, da je izpolnjen zgornji pogoj. Za trend s tremi parametri je sistem normalnih enačb naslednji:

$$\begin{aligned}
\Sigma Yf_1 &= a\Sigma f_1^2 + b\Sigma f_2 f_1 + c\Sigma f_3 f_1 \\
\Sigma Yf_2 &= a\Sigma f_1 f_2 + b\Sigma f_2^2 + c\Sigma f_3 f_2 \\
\Sigma Yf_3 &= a\Sigma f_1 f_3 + b\Sigma f_2 f_3 + c\Sigma f_3^2
\end{aligned} \tag{12.29}$$

12.60 Največkrat vzamemo kot funkcije časa f_1, f_2, f_3 ... potence od x. ($f_1 = 1$, $f_2 = x$; $f_3 = x^2$ ...).

Če je trend premica: $T = a + bx$, je sistem normalnih enačb po zgornjih pravilih

$$\Sigma Y = Na + b\Sigma x \qquad (12.80)$$
$$\Sigma Yx = a\Sigma x + b\Sigma x^2$$

Če je trend parabola druge stopnje: $T = a + bx + cx^2$, je sistem normalnih enačb

$$\Sigma Y = aN + b\Sigma x + c\Sigma x^2$$
$$\Sigma Yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \qquad (12.81)$$
$$\Sigma Yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

Če pa je trend parabola tretje stopnje: $T = a + bx + cx^2 + dx^3$ dobimo tele normalne enačbe

$$\Sigma Y = aN + b\Sigma x + c\Sigma x^2 + d\Sigma x^3$$
$$\Sigma Yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + d\Sigma x^4 \qquad (12.82)$$
$$\Sigma Yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + d\Sigma x^5$$
$$\Sigma Yx^3 = a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + d\Sigma x^6$$

Kakšne so normalne enačbe, če je trend parabola višje stopnje kot tretje, je razvidno iz ten primerov. Vendar krivulje, ki imajo preveliko število parametrov, iz vsebinskih razlogov niso primerne za trend.

12.61 Zgornje normalne enačbe znatno poenostavimo, če vzamemo, da je izhodišče koordinatnega sistema za x natančno sredi vrste. V tem primeru so namreč vsote vseh lihih potenc od x enake 0. Če je število členov v časovni vrsti liho $N = 2i + 1$, vzamemo izhodišče $x = 0$ v $(i+1)$ členu časovne vrste. Vrednost x so torej: $-i$, $-i+1$ ..... -2, -1, 0, $+1$, $+2$, ... $i - 1$, i.

Težavneje je, če je število členov sodo število: $N = 2i$. V tem primeru vzamemo prvi člen časovne vrste $x = -2i + 1$, za nadaljnje člene pa x večamo za 2. Tako dobimo vrsto za x. Za prvo polovico časovne vrste so x negativna, za drugo polovico pa pozitivna cela liha števila. Vrsta za x je v tem primeru:

$$-2i + 1,\ -2i + 3,\ \ldots\ -5,\ -3,\ -1,\ +1,\ +3,\ +5,\ \ldots\ 2i - 3,\ 2i - 1$$

Če je izpolnjen ta pogoj za vrste do 30 členov razmeroma enostavno izračunamo parametre parabol do tretje stopnje s konstantami iz tabele G. Postopek je tale:

a) iz osnovne vrste izračunamo izraze

$$\Sigma Y, \quad \Sigma Yx, \quad \Sigma Yx^2 \quad \text{in} \quad \Sigma Yx^3$$

b) parametre izračunamo z naslednjimi enačbami:

Za *linearen trend:*

$$T = a_1 + b_1 x \tag{12.33}$$

so parametri:

$$a_1 = A_1 \Sigma Y \qquad b_1 = B_1 \Sigma Yx \tag{12.34}$$

Za *paraboličen trend druge stopnje:*

$$T = a_2 + b_2 x + c_2 x^2 \tag{12.35}$$

so parametri:

$$\begin{aligned} a_2 &= A_2 \Sigma Y - K_2 \Sigma Yx^2 \; ; \quad b_2 = B_2 \Sigma Yx \; ; \\ c_2 &= C_2 \Sigma Yx^2 - K_2 \Sigma Y \end{aligned} \tag{12.36}$$

Za *paraboličen trend tretje stopnje:*

$$T = a_3 + b_3 x + c_3 x^2 + d_3 x^3 \tag{12.37}$$

so parametri:

$$\begin{aligned} a_3 &= A_3 \Sigma Y - K_3 \Sigma Yx^2 \; ; \quad b_3 = B_3 \Sigma Yx - K_3 \Sigma Yx^3 \; ; \\ c_3 &= C_3 \Sigma Yx^2 - K_3 \Sigma Y \; ; \quad d_3 = D_3 \Sigma Yx^3 - K_3 \Sigma Yx \end{aligned} \tag{12.38}$$

12.62 Za proizvodnjo cinkovega koncentrata v obdobju med obema vojnama je treba izračunati po metodi najmanjših kvadratov paraboličen trend druge stopnje.

Trend izračunamo po naslednjih stopnjah:

Iz osnovne vrste, ki ima 23 členov, izračunamo količine: ΣY, ΣYx, ΣYx^2. Iz njih in s konstantami za izračunavanje parametrov paraboličnega trenda v tabeli G izračunamo parametre a_2, b_2, c_2 za trend $T = a_2 + b_2 x + c_2 x^2$. Iz te enačbe na znan način izračunamo vrednosti trenda, ki ustrezajo posameznim členom tako, da v $T = a_2 + b_2 x + c_2 x^2$ vnašamo ustrezne vrednosti za x.

V tabeli G najdemo, da je za $N = 23$:

$$A_2 = 0{,}981366_{-1} \; ; \quad B_2 = 0{,}988142_{-3} \; ; \quad C_2 = 0{,}282326_{-4} \; ;$$

$$K_2 = 0{,}124224_{-2}$$

Tabela 12.18 Izračunavanje paraboličnega trenda druge stopnje za proizvodnjo cinkovega koncentrata v Sloveniji v razdobju med obema vojnama

Leto (1)	x (2)	Y (3)	Yx (4)=(3).(2)	Yx^2 (5)=(4).(2)
1919	-11	90	- 990	10890
1920	-10	350	-3500	35000
1921	- 9	75	- 675	6075
1922	- 8	450	-3600	28800
1923	- 7	600	-4200	29400
1924	- 6	520	-3120	18720
1925	- 5	510	-2550	12750
1926	- 4	639	-2556	10224
1927	- 3	1143	-3429	10287
1928	- 2	1073	-2143	4292
1929	- 1	1313	-1313	1313
1930	0	1547	0	0
1931	+ 1	1301	+1301	1301
1932	+ 2	1358	+2713	5432
1933	+ 3	2293	+6879	20637
1934	+ 4	2282	+9128	36512
1935	+ 5	2173	+10865	54325
1936	+ 6	2567	+15402	92412
1937	+ 7	2401	+16807	117649
1938	+ 8	4530	+36240	289920
1939	+ 9	3252	+29268	263412
1940	+10	4121	+41210	412100
1941	+11	3483	+38313	421443
		38071 ΣY	+180050 ΣYx	1882894 ΣYx^2

Po obrazcih za izračunavanje parametrov za parabolični trend dobimo:

$a_2 = A_2\Sigma Y - K_2\Sigma Yx^2 = 0,981366.10^{-1}.38071 - 0,124224.10^{-2}.1882894 = 1397,15$

$b_2 = B_2\Sigma Yx = 0,988142.10^{-3}.180050 = 177,915$

$c_2 = C_2\Sigma Yx^2 - K_2\Sigma Y = 0.282826.10^{-4} . 1882894 - 0.124224.10^{-2} . 38071 = 5,86567$

Enačba trenda je: $T = 1397,15 + 177,915x + 5,86567x^2$

Po znanem postopku dobimo iz te funkcije vrsto trenda v tabeli 12.19.

Tabela 12.18 Trend za proizvodnjo cinkovega koncentrata v Sloveniji določen po metodi najmanjših kvadratov

Leto	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
x	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
T	150	206	271	349	439	541	654	779	916	1065	1225

Leto	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
x	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11
T	1397	1581	1776	1984	2203	2433	2673	2930	3193	3474	3733	4034

Pri kontroli o pravilnosti izračuna trenda upoštevamo pravilo, da je vsota členov v osnovni vrsti enaka vsoti trenda. Vsota trenda je ΣT = = 38072, vsota v osnovni vrsti pa ΣY = 38071. Razlika izvira iz zaokroževanja.

12.63 Za proizvodnjo svinčeno-cinkove rude v LR Sloveniji v razdobju po osvobojitvi od 1943-1957 je treba določiti smer razvoja. Osnovna vrsta podatkov je dana v tabeli 12.19.

Tabela 12.19 Proizvodnja svinčeno-cinkove rude v LR Sloveniji v 10^3 tonah (Vir: MSP LRS)

Leto	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957
Proizvodnja	138	145	211	226	232	231	238	312	347	419	455	461

Analiza osnovnih podatkov pokaže, da imajo podatki linearno smer, če so narisani na pologaritmičnem grafikonu. Iz tega sklepamo, da je smer razvoja eksponencialna funkcija oblike $T = ab^x$. Iz podstavka o primernih funkcijah trenda povzamemo, da moremo to eksponencialno funkcijo prevesti v enostavnejšo obliko $\log T = \log a + x.\log b$, če jo logaritmiramo. To pa je linearna funkcija za x.

Postopek določitve trenda je torej naslednji: Najprej za časovno vrsto osnovnih podatkov poiščemo ustrezno časovno vrsto logaritmov $\log Y$. Ta časovna vrsta kaže linearno smer razvoja.

Vrsti $\log Y$ prilagodimo po metodi najmanjših kvadratov premico. Iz enačbe vidimo, da sta parametra te premice logaritma parametrov eksponencialnega trenda. Z antilogaritmiranjem vrste $\log T$ dobimo časovno vrsto trenda za proizvodnjo svinčeno-cinkove rude.

Ker je število členov sodo, vzamemo za x vrsto:
$-11, -9, \ldots -3, -1, +1, +3 \ldots +9, +11$

Tabela 12.20 Izračunavanje eksponencionalnega trenda za proizvodnjo svinčeno-cinkove rude v LR Sloveniji

Leto (1)	Y (2)	log Y (3)	x (4)	$x.\log Y$ (5)=(3).(4)	$\log T = \log a + x \log b$ (6)	T (7)
1946	133	2,1239	-11	-23,3629	2,1548	143
1947	145	2,1614	- 9	-19,4523	2,2031	160
1948	211	2,3243	- 7	-16,2701	2,2515	178
1949	223	2,3541	- 5	-11,7705	2,2998	199
1950	232	2,3355	- 3	- 7,0965	2,3481	223
1951	231	2,3636	- 1	- 2,3636	2,3965	249
1952	238	2,3763	+ 1	+ 2,3766	2,4448	278
1953	312	2,4942	+ 3	+ 7,4826	2,4932	311
1954	347	2,5403	+ 5	+12,7015	2,5415	348
1955	419	2,6222	+ 7	+18,3554	2,5898	389
1956	455	2,6580	+ 9	+23,9220	2,6382	435
1957	461	2,6637	+11	+29,3007	2,6865	486
Vsota		29,0478		+13,8223	29,0478	
		$\Sigma \log Y$		$\Sigma x.\log Y$	$\Sigma \log T$	

Parametre izračunamo po tabeli G. Ker je logaritmični trend linearen, poiščemo pod $N = 12$: $A_1 = 0,83333_{-1}$, $P_1 = 0,174825_{-2}$. Dalje sledi:

$$\text{Log}a = A_1 \Sigma \log Y = 0,833333.10^{-1} \cdot 29,0478 = 2,42065$$

$$\text{Log}b = P_1 \Sigma x.\log Y = 0,174825.10^{-2} \cdot 13,8223 = 0,02417$$

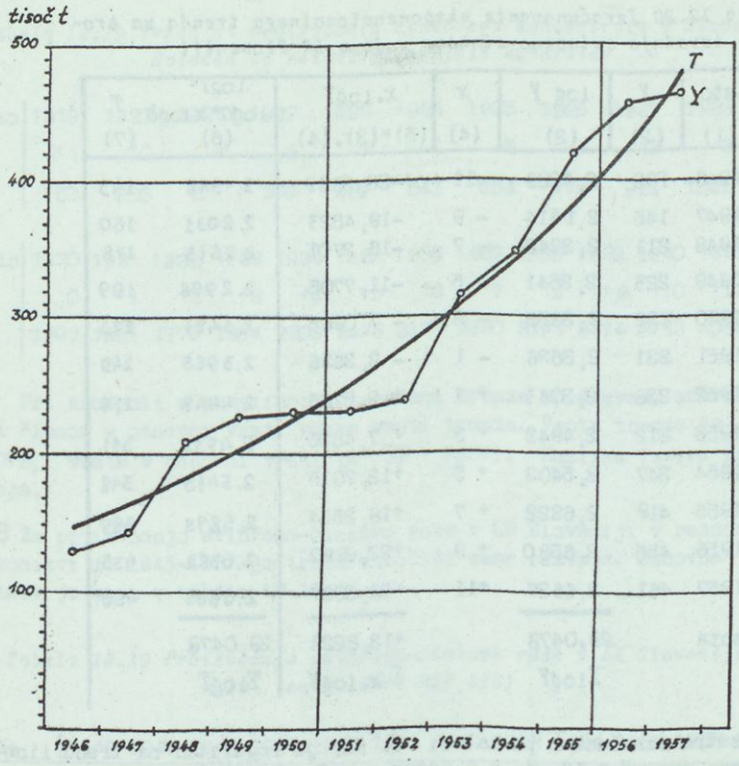
Če te rezultate antilogaritmiramo, dobimo:

$$a = 263,42 \qquad b = 1,0572$$

$$T = a.b^x = 263,42 \cdot 1,0572^x$$

$a = 263,42$ pomeni vrednost trenda sredi vrste, $b^2 = 1,1177$ pa je povprečni koeficient aktivnosti za proizvodnjo svinčeno-cinkove rude v razdobju 1943-1957.

Na sliki 12.13 je narisana osnovna vrsta podatkov in eksponencialni trend. Iz slike je dobro viden tipičen cikel v letih 1948-1953.



Slika 12.13 Trend za proizvodnjo svinčeno-cinkove rude v LRS določen po metodi najmanjših kvadratov

SEZONSKE IN PERIODIČNE VARIACIJE

12.64 Sezonske in periodične variacije proučujemo iz več namenov. Pomembno je, da poznamo sezonske variacije v preteklosti zaradi medsebojne primerjave in analize. Sezonsko komponento proučujemo tudi zaradi planiranja in napovedovanja za krajša razdobja. Sezonska variacije in periodične variacije nasploh so važne prvine v ekonomskih pojavih. Zato jih moramo poznati, da moremo proučevati običajno negativen vpliv teh variacij. Če časovno primerjamo sezonske variacije za različna razdobja, osvetlimo in analiziramo spremembe, ki so nastale v sezonskem variiranju kot rezultat različnih ukrepov ali iz drugih vzrokov. Tako na primer različna tehnika gradenj spremeni sezonsko komponento v gradbeni dejavnosti, različna politika subvencioniranja v gostinstvu spremeni sezonsko komponento v turizmu itd.

Poznavanje sezonskih in periodičnih variacij je važno tudi za kratkoročno planiranje najrazličnejših pojavov. Za vrsto dejavnosti je nepravilno linearno razdeljevanje letnega plana na mesečne plane in moramo zanje upoštevati sezonski značaj določenih pojavov. To pa moremo le, če poznamo sezonsko komponento za te pojave.

Sezonski značaj pojavov pa je samo eden izmed periodičnih vplivov. Tako ima na primer promet v trgovini na drobno mesečno periodičnost in opažamo v začetku vsakega meseca večji, proti koncu meseca pa manjši promet. Tudi tedenska periodičnost je tipična za veliko število pojavov. Tako pri potniškem prometu in z njim zvezanim številom prometnih nesreč, v prometu v gostinstvu itd. opazujemo tedensko periodičnost. Tudi dnevna periodičnost je zelo pogosta. Potrošnja električne energije ali vode, gostota prometa itd. imajo dnevno periodičnost. Znane so dnevne konice v potrošnji električne energije ali vode, konice v lokalnem prometu itd.

Bistvo periodičnih vplivov je v tem, da se na določeno časovno razdobje (n.pr. leto, mesec, teden, dan) ponavljajo: to izzove na proučevanih pojavih periodično vedno enak učinek.

V splošnem periodični vplivi v gospodarstvu niso zaželeni. Periodičen značaj dovede ali do zastoja v času sezone ali do neizkoriščenja kapacitet izven sezone, če imamo zadostne kapacitete za čas maksima. Zato si prizadevamo, da periodične vplive omilimo, če jih že ne moremo odpraviti. Tipičen primer je nadomeščanje proizvodnje hidro in termo elektrike. Obe se medsebojno dopolnujeta tako, da čimbolj omilimo sezonski značaj proizvodnje hidroelektrarn. Periodične vplive moramo nujno poznati in proučevati, če skušamo njih učinek ublažiti in odpraviti.

Nakazali smo, da je bistvo periodičnih vplivov v tem, da se na določene časovne razmake ponavljajo. Vendar tudi ta učinek ni statičen. V daljšem razdobju opažamo spremenljivost sezonskih vplivov. Te spremembe prinesejo bodisi razvoj ali zavestni ukrepi, s katerimi vplivamo na sezonski značaj pojavov.

Ena izmed važnih nalog analize časovnih vrst je razstavljanje časovne vrste na njene osnovne komponente. Tudi zaradi tega je važno, da poznamo sezonsko komponento, da jo moremo iz osnovne časovne serije odstraniti. Če iz časovne vrste odstranimo trend in sezonske variacije, ostane namreč v časovni vrsti le še ciklična in slučajna komponenta. Sezonsko komponento moramo zato poznati tudi, če analiziramo ciklična nihanja pojavov.

12.65 Sezonsko oziroma periodično komponento v časovnih vrstah določamo po več metodah. Vsaka izmed njih je prilagojena posebnostim različnih časovnih vrst. Zato je odločitev o tem, katero izmed teh metod v posebnem primeru uporabimo, odvisna od dinamike pojava, ki ga časovna vrsta prikazuje. V nadaljevanju obravnavamo tele metode za določanje sezonske komponente:

a) metodo vsot,
b) metodo verižnih indeksov,
c) metodo kvocientov na trend,

a) metodo kvocientov na vrsto drsečih sredin
e) metodo grafičnega približka.

Metoda vsot

12.66 Če pojav, za katerega iščemo periodično komponento, nima izrazitega trenda niti cikličnih nihanj, moremo s pridom uporabiti metodo vsot. Če trend iz ciklične variacije zanemarimo, je model take časovne vrste

$$Y = A.(1+p+e) \qquad (12.38)$$

Pri tem je A = konstanta, p = periodična komponenta, ki se v povprečju ene periode uniči, e = rezultat slučajnih vplivov. Sezonske indekse $1+p$ dobimo iz take vrste takole:

a) Imamo večletno časovno vrsto po mesecih. Za vsak mesec izračunamo iz ustreznih mesečnih vrednosti za vsa leta, ki jih imamo v časovni vrsti, vsote S. Tako seštejemo podatke za januarje vseh let v januarsko vsoto, podatke za februarje vseh let v februarsko vsoto. S tem odpravimo slučajne odklone, in je

$$S = KA(1+p) \qquad (12.39)$$

b) Iz dobljenih vsot izračunamo povprečje vsot $\bar{S}$. Ker se periodična komponenta p v vsoti ene periode uniči, dobimo

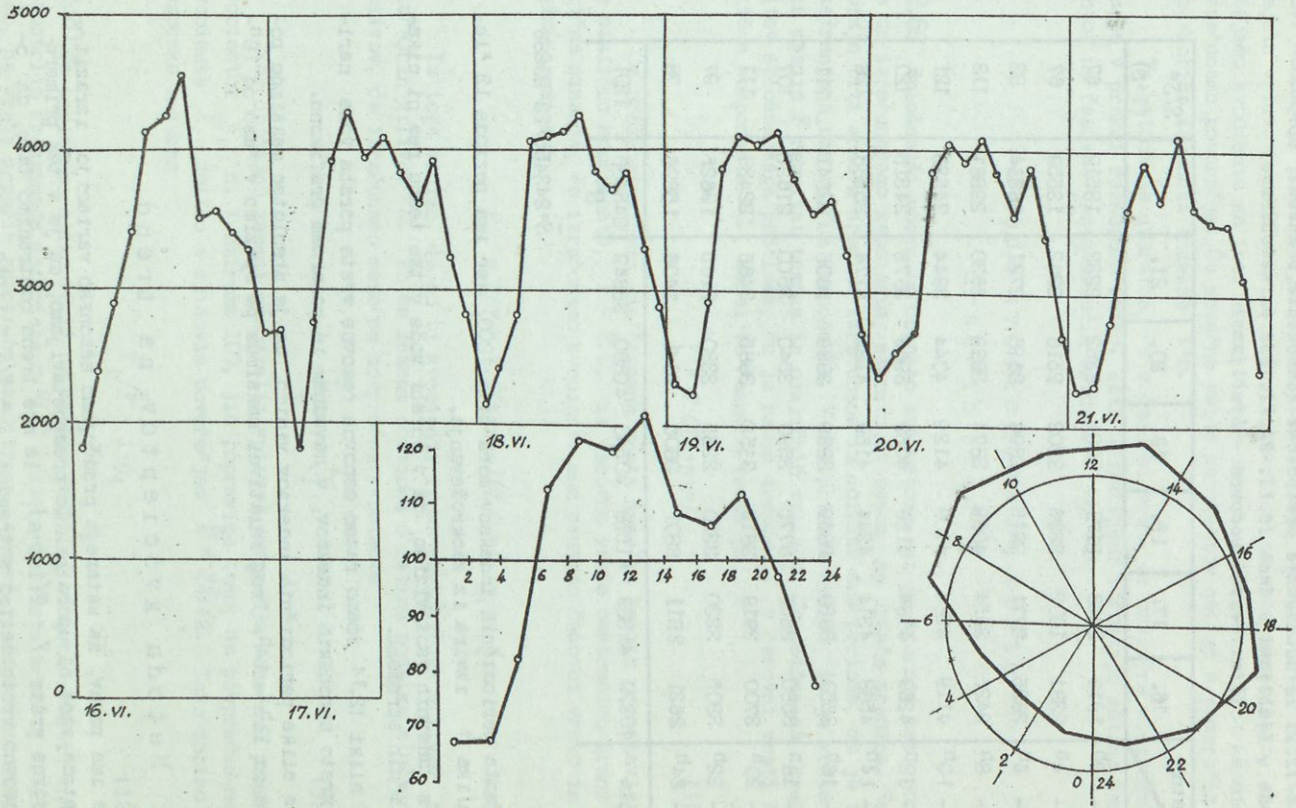
$$\bar{S} = KA \qquad (12.40)$$

c) Če mesečne vsote S delimo s povprečno vsoto $\bar{S}$, dobimo sezonske indekse

$$S/\bar{S} = KA(1+p)/KA = 1+p \qquad (12.41)$$

Ta postopek ne velja samo za sezonska, temveč za vsa periodična nihanja. Za te primere pride še pogosteje v poštev, ker se v urnih ali dnevnih časovnih vrstah trend in ciklični vplivi ne pokažejo, kakor je to primer pri večini mesečnih časovnih vrstah.

12.67 Za teden od 16. junija do 21. junija 1958 imamo vrsto o dveurni porabi vode v Ljubljani. (Vir: Mestni vodovod - Ljubljana.) Po metodi vsot je treba določiti periodo v dnevni porabi vode v tem razdobju. Uporaba te metode je upravičena, ker se v razdobju šestih dni ne pokažejo niti spremembe trenda niti drugi časovni vplivi, ki bi onemogočali uporabo te metode.



Slika 12.14. Periodična komponenta v dnevni porabi vode v Ljubljani za razdobje 16.VI.-21.VI.1958

Tabela 12.21 Izračunavanje periodične komponente v dnevni porabi vode v Ljubljani v dneh 16.VI.-21.VI.1958 po metodi vsot (Vir: Mestni vodovod - Ljubljana)

Datum / Ura	16.	17.	18.	19.	20.	21.	S	'S/S= (1+p)
$0^h - 2^h$	1816	2646	2172	2301	2352	2232	13519	67
$2^h - 4^h$	2381	1602	2396	2202	2516	2262	13559	67
$4^h - 6^h$	2895	2771	2815	2896	2693	2791	16834	83
$6^h - 8^h$	3425	3954	4089	3874	3889	3630	22861	113
$8^h - 10^h$	4119	4284	4109	4129	4044	3944	24629	121
$10^h - 12^h$	4269	4184	4159	4059	3969	3675	24315	120
$12^h - 14^h$	4568	4314	4244	4159	4129	4174	25588	126
$14^h - 16^h$	3550	3839	3869	3839	3839	3605	22541	111
$16^h - 18^h$	3580	3635	3775	3590	3540	3500	21620	107
$18^h - 20^h$	3700	3919	3819	3350	3885	3465	22438	111
$20^h - 22^h$	3305	3230	3310	3235	3380	3165	19625	97
$22^h - 24^h$	2631	2811	2801	2606	2641	2406	15896	78
vsota	40239	41389	41558	40540	40880	38849	243455 S	1201

$\bar{S}$=243455/12=20288

Vsota periodičnih indeksov mora biti 1200, ker ima perioda 12 členov. Razlika 1 izvira iz zaokroževanja.

Iz dnevnih vsot vidimo, da v porabi vode v tem tednu res ni nikakih cikličnih sprememb.

V sliki 12.14 imamo razen osnovne časovne vrste porabe vode narisano še vrsto sezonskih indeksov, v navadnem in polarnem grafikonu.

Iz slike periodičnih indeksov vidimo, da je absoluten maksimum porabe vode od $12^h - 14^h$. Drug relativni maksimum pa opazimo v času od $18^h - 20^h$.

Metoda kvocientov na trend

12.68 Če ima pojav, za katerega proučujemo sezonsko variacijo, izrazit trend, metoda vsot ni uporabna. Če predpostavljamo, da je v tem primeru model časovne vrste $Y = T(1+p+e)$, iz nje trend odstranimo tako, da osnovno časovno vrsto delimo s trendom $K = Y/T = 1+p+e$. Vrsta kvocientov, ki jo dobimo, je, kakor kaže shema, izraz periodičnih in slučajnih vpli-

vov. Nadaljnji postopek je podoben prejšnjemu. Da odstranimo rezultate slučajnih vplivov, poiščemo povprečja kvocientov za vsak mesec tako, da seštejemo kvociente za isti mesec v vseh letih. Iz mesečnih vsot kvocientov izračunamo povprečje. Če mesečne vsote kvocientov delimo s povprečno vsoto, dobimo sezonske indekse $1+p$.

Ta metoda je uporabna, če v časovni vrsti ni ciklične komponente. To pa je v praksi razmeroma redko. Zato uporabljamo pogosteje metodo kvocientov na časovno vrsto drsečih sredin.

Metoda kvocientov na vrsto drsečih sredin

12.69 Splošnejši pomen in uporabo ima metoda, ki je zgornji zelo podobna, le da izračunavamo kvociente, namesto na trend, na drseče sredine. Seveda morajo biti drseče sredine izračunane iz dolžine ene periode, da je iz njih odstranjena periodična komponenta. V tem primeru menimo, da je vrsta drsečih sredin $\bar{Y}$ rezultat trenda in cikličnih vplivov. Ta metoda daá torej rešitev splošnejšega problema, če je razen trenda T v časovni vrsti še ciklična komponenta C. Model te časovne vrste je

$$Y = T.C(1+p+e) \qquad (12.42)$$

Če predpostavljamo, da časovna vrsta drsečih sredin predstavlja trend in ciklično komponento, $\bar{Y} = T.C$, iz časovne vrste odstranimo trend in ciklična nihanja, če izračunamo kvociente med osnovno časovno vrsto in drseč-imi sredinami

$$K_{ri} = Y/\bar{Y} = T.C(1+p+e)/T.C = 1+p_i+e_{ri} \qquad (12.43)$$

Iz tega vidimo, da so ti kvocienti izraz periodičnih vplivov p in slučajnih vplivov e. Po že znanem postopku očistimo podatke slučajnih vplivov, če izračunamo mesečne povprečne kvociente

$$\bar{K}_i = 1+p_i \qquad (12.44)$$

Ker običajno zaradi preostalih vplivov povprečje mesečnih povprečnih kvocientov $\bar{K}_i$ ni 1 oziroma 100, jih popravimo tako, da povprečne mesečne kvociente $\bar{K}_i$ delimo z njihovim povprečjem $\bar{\bar{K}} = \Sigma\bar{K}/12$. Tako dobimo prave sezonske indekse

$$\bar{K}_i/\bar{\bar{K}} = 1+p_i \qquad (12.45)$$

12.70 Če pogledamo sliko 12.11, v kateri je poleg osnovne vrste indeksov industrije gradbenega materiala v FLRJ vrisana še vrsta drsečih sredin, vidimo, da vrsta drsečih sredin ni samo trend, temveč vsebuje tudi ciklična nihanja. Zato pride pri izračunavanju sezonske komponente za ta primer v

poštev metoda kvocientov na vrsto drsečih sredin.

Tabela 12.22 Izračunavanje sezonskih indeksov za vrsto indeksov gradbene industrije po metodi kvocientov na drseče sredine

	J	F	M	A	M	J	J	A	S	O	N	D
1951							136,4	147,7	128,5	122,3	104,2	85,2
1952	44,6	37,2	48,3	37,1	94,5	134,7	165,3	161,8	147,0	126,3	89,9	69,4
1953	41,7	40,3	57,9	80,7	102,2	136,5	142,4	150,0	143,3	136,2	99,7	86,1
1954	37,1	25,8	54,3	72,7	94,9	129,8	146,5	152,8	152,8	152,5	94,2	85,4
1955	55,8	51,7	54,9	62,8	99,2	148,5	141,3	149,6	142,4	128,2	94,7	87,0
1956	56,2	28,0	46,6	68,1	89,1	131,3	139,3	157,2	158,8	151,1	102,1	88,9
1957	40,6	34,3	48,2	67,6	105,3	135,0						
$\sum K$	182,7	140,5	206,7	270,5	393,9	537,5	539,8	609,6	588,5	542,1	390,7	340,6
$\bar{K}_i$	45,7	35,1	51,7	67,6	98,5	134,4	142,4	152,4	147,1	135,5	97,7	85,2
$I+p_i$	46	35	52	68	99	135	143	153	148	136	98	86

Vsota: 1138,3; povprečje $\bar{\bar{K}} = 99,44$

Vsota: 1199; povprečje: 100

Iz tabele 12.5 in iz slike 12.11 vidimo, da se vrsta dvanajstmesečnih drsečih sredin začne pol leta kasneje in konča pol leta prej kakor osnovna časovna vrsta. Na ten odsekih seveda ne moremo izračunati kvocientov. V tabeli 12.22 je prikazan postopek, kako izračunavamo sezonsko komponento po tej metodi. Ta je izračunana po naslednjih stopnjah:

1) Ker smo izračunali vrsto letnih drsečih sredin za indeks gradbenega materiala v FLRJ že v tabeli 12.5, jo uporabimo za izračunavanje sezonske komponente.

2) Iz časovne vrste osnovnih podatkov in vrste drsečih sredin izračunamo vrsto kvocientov $K = Y/\bar{Y}$. Kvociente zaradi večje nazornosti pomnožimo s 100. Prvi kvocient izračunamo za mesec julij 1951. Indeks gradbenega materiala iz osnovnega gradiva za ta mesec je 97, ustrezna vrednost dvanajstmesečne sredine pa 71,1. Kvocient je torej 100.97/71,1 = 136,4. Enako izračunamo vse nadaljnje kvociente.

3) Ker morejo zaradi enkratnih vplivov nastopiti izredni vplivi, ki motijo regularni potek, za vsak mesec izločimo minimalni in maksimalni kvocient, na primer: iz januarja 37,1 in 56,2.

4) Iz preostalih kvocientov izračunamo modificirane povprečne mesečne kvociente $\bar{K}_i$: n. pr. januar: $\bar{K}_1 = 1/4(44,3+41,7+55,8+40,3) = 182,7/4 = 45,7$.

5) Iz dobljenih mesečnih kvocientov $\overline{K_i}$ izračunamo povprečje $\overline{\overline{K}}$. Če bi bili povprečni mesečni kvocienti pravi sezonski indeksi, bi bilo povprečje teh kvocientov enako 100. Izračunano povprečje je v našem primeru $\overline{\overline{K}} = 99{,}38$. Razlika torej ni velika: zato tudi popravljena vrsta sezonskih indeksov ni bistveno različna od nje.

3) Povprečne mesečne kvociente $\overline{K_i}$ popravimo tako, da jih delimo s povprečjem povprečnih mesečnih kvocientov $\overline{\overline{K}}$. $1+p_i = \overline{K_i}/\overline{\overline{K}}$.

7) Vsota sezonskih indeksov, ki jih tako dobimo, je enaka 1199, in torej ni bistveno različna od pričakovane vrednosti 1200.

Izračunana vrsta sezonskih indeksov in njihov grafikon pokažeta tipično sliko sezonskega nihanja v industriji gradbenega materiala. Amplituda tega nihanja je velika, kar je vsekakor negativen pojav.

Metoda verižnih indeksov

12.71 Razen omenjenih metod v praksi večkrat uporabljamo še metodo verižnih kvocientov. Ta metoda ima to prednost, da predhodno ni treba izračunavati niti trenda niti vrste drsečih sredin.

Metoda verižnih indeksov teoretično predpostavlja, da je trend eksponencialna funkcija tipa $T = a.b^x$. Če to predpostavljamo, je model časovne vrste

$$Y_x = a.b^x(1+p+e) \tag{12.43}$$

Verižni kvocienti Y_x/Y_{x-1} za ta tip časovne vrste so:

$$I_x = Y_x/Y_{x-1} = \frac{ab^x(1+p_i+e_1)}{ab^{x-1}(1+p_{i-1}+e_0)} = b\,\frac{1+p_i}{1+p_{i-1}} + e' \tag{12.47}$$

Kakor vidimo iz tega obrazca, je od trenda v kvocientu ostal le še parameter b, od periodične komponente pa kvocient dveh zaporednih sezonskih indeksov. Kakor pri prejšnjih metodah, slučajne variacije e' odstranimo, če poiščemo povprečne mesečne verižne indekse

$$\overline{I_i} = b\,\frac{1+p_i}{1+p_{i-1}} \tag{12.48}$$

Zaradi lastnosti kvocientov sezonskih indeksov je geometrijska sredina povprečnih mesečnih indeksov enaka b.

$$G_I = \sqrt[12]{b\,\frac{1+p_1}{1+p_{12}} \cdot b\,\frac{1+p_2}{1+p_1} \cdots\cdots b\,\frac{1+p_{11}}{1+p_{10}}\, b\,\frac{1+p_{12}}{1+p_{11}}} = b \tag{12.49}$$

S kvocientom povprečnega mesečnega indeksa $\overline{I_i}$ in geometrijske sredine dobimo čiste kvociente sezonskih indeksov

$$\overline{I_i}/G_I = b\frac{1+p_i}{1+p_{i-1}} \Big/ b = \frac{1+p_i}{1+p_{i-1}} = t_i \qquad (12.50)$$

Če dobljene čiste kvociente sezonskih indeksov t_i postopoma množimo med seboj, dobimo naslednje izraze

$$\begin{aligned}
K_1 &= t_1 = \frac{1+p_1}{1+p_{12}} \\
K_2 &= K_1 t_2 = \frac{1+p_1}{1+p_{12}} \cdot \frac{1+p_2}{1+p_1} = \frac{1+p_2}{1+p_{12}} \\
K_3 &= K_2 t_3 = \frac{1+p_2}{1+p_{12}} \cdot \frac{1+p_3}{1+p_2} = \frac{1+p_3}{1+p_{12}} \\
K_4 &= K_3 t_4 = \frac{1+p_3}{1+p_{12}} \cdot \frac{1+p_4}{1+p_3} = \frac{1+p_4}{1+p_{12}} \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
K_{12} &= K_{11} t_{12} = \frac{1+p_{11}}{1+p_{12}} \frac{1+p_{12}}{1+p_{11}} = \frac{1+p_{12}}{1+p_{12}}
\end{aligned} \qquad (12.51)$$

Členi vrste kumulativnih produktov K_i so, kakor vidimo iz 12.51, sezonski indeksi, deljeni z $(1+p_{12})$. Ker je zaradi lastnosti sezonskih indeksov povprečje iz K_i enako $\overline{K} = 1/1+p_{12}$, prave sezonske indekse $1+p_i$ dobimo

$$K_i/\overline{K} = \frac{1+p_i}{1+p_{12}} \Big/ \frac{1}{1+p_{12}} = 1+p_i \qquad (12.52)$$

če K_i delimo s povprečnim $\overline{K}$.

12.72 Časovna vrsta indeksov industrije gradbenega materiala kaže smer razvoja, ki jo moremo v približku vzeti kot eksponencialno. Ker da metoda verižnih indeksov zadovoljive rezultate, tudi če trend ni povsem eksponencialen, jo moremo upravičeno uporabiti v našem primeru.

V tabeli 12.23 je za primer indeksov industrije gradbenega materiala FLRJ nakazano, kako izračunavamo sezonske indekse po metodi verižnih indeksov. Sezonsko komponento izračunamo po teh stopnjah:

a) Iz vrste osnovnih podatkov izračunamo vrsto verižnih indeksov I_x. Verižnega indeksa za januar 1951 ne moremo izračunati, ker ne poznamo indeksa za december 1950. Prvi verižni indeks iz osnovne časovne vrste je izračunan za februar 1951 (100.37/38 = 97,4) drugi za marec 1951 (100.49/37 = = 132,4) ita.

Verižne indekse za januarje naslednjih let dobimo z deljenjem januarskih vrednosti z decembrskimi vrednostmi prejšnjega leta (n.pr. $I_{jan.1952}$ = 100.31/58 = 53,4 ita).

b) Izračunamo povprečja verižnih indeksov $\overline{I}_i$ za vsakega izmed 12 mesecev. Za vsak mesec izločimo najmanjšo in največjo vrednost in izračunamo modificirana povprečja. Pri januarskem povprečju moramo upoštevati, da manjka verižni indeks za januar 1951. Zato vsoto januarskih indeksov (222,8) delimo s štiri, ne pa s pet kakor za druge mesece, za katere imamo dane verižne indekse za pet let, ko odstranimo skrajne primere.

c) Poiščemo geometrijsko sredino G_I povprečnih mesečnih verižnih indeksov $\overline{I}_i$

$$G_I = \sqrt[12]{56,9.86,4.144,5 \ldots\ldots 74,5.84,6} = 101,06$$

G_I izračunamo z logaritmi.

d) Povprečne mesečne verižne indekse $\overline{I}_i$ delimo z G_I = 101,06, da odstranimo parameter b. Tako dobimo kvociente sezonskih indeksov t_i = = $\overline{I}_i/G_I$. Geometrijska sredina G_I = 101,06 je vmesni rezultat, ki pove, da je povprečni mesečni koeficient dinamike b = 101,06.

e) S postopnim množenjem kvocientov t_i dobimo K_i.

$$K_1 = t_1 = 53,8 \; ; \quad K_2 = K_1 t_2 = 56,3.0.855 = 48,1 \ldots\ldots \text{ ita.}$$

Kontrola: K_{12} = 100,0.

f) Izračunamo povprečje iz K_i.

$$\overline{K} = \frac{1}{12}(56,3 + 48,1 + \ldots + 119,5 + 100,0) = \frac{1505,9}{12} = 125,5$$

g) Posamezne K_i delimo z dobljenim povprečjem $\overline{K}$ = 125,5, kvociente pa pomnožimo s 100. Tako dobimo čiste sezonske indekse. Za januar je n.pr.

$$1+n_1 = 100\frac{56,3}{125,5} = 45$$

Kontrola: Vsota sezonskih indeksov je enaka 1200. V našem primeru dobimo vsoto enako 1201. Razlika izvira iz zaokroževanja.

Tabela 12.23 Izračunavanje sezonskih indeksov za časovno vrsto indeksov industrije gradbenega materiala po metodi verižnih indeksov

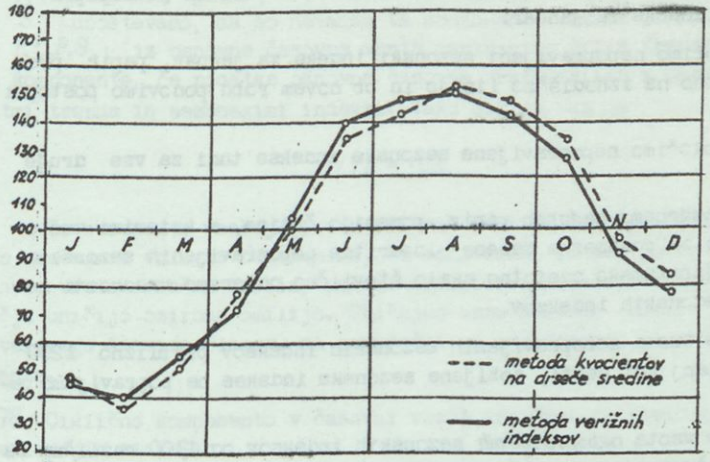
a) Osnovni podatki

	J	F	M	A	M	J	J	A	S	O	N	D
1951	38	37	49	53	78	95	97	104	90	84	71	58
1952	31	27	33	50	71	101	124	122	112	98	71	53
1953	34	33	48	38	92	119	125	131	127	118	83	74
1954	32	22	48	35	85	117	134	143	145	122	90	83
1955	55	51	54	31	97	145	138	144	135	121	89	81
1956	52	23	44	31	87	131	139	157	155	154	103	89
1957	44	36	54	73	119	154	172	179	133	152	121	97

b) Verižni indeksi

	J	F	M	A	M	J	J	A	S	O	N	D
1951		97,4	132,4	114,3	139,3	121,8	102,1	107,2	83,5	93,3	84,5	81,7
1952	53,4	87,1	133,3	138,9	142,0	142,3	122,8	98,4	91,8	87,5	72,4	73,9
1953	30,7	97,1	145,5	141,7	135,3	129,4	105,0	104,8	93,9	92,9	72,9	83,0
1954	43,2	68,8	218,3	135,4	130,7	137,3	114,5	101,4	101,4	84,1	73,8	92,2
1955	33,3	92,7	103,8	118,0	152,0	149,5	95,2	104,3	93,8	89,3	73,3	91,0
1956	34,2	50,0	139,2	138,3	142,3	142,3	103,1	112,9	98,7	99,4	38,8	84,0
1957	49,5	83,4	142,1	140,7	153,3	129,4	111,7	104,1	92,7	91,3	79,3	80,2
Vsota	227,8	432,1	722,5	337,9	715,8	381,3	539,4	521,8	473,9	454,9	372,3	422,9
$\bar{I}_i$	53,9	83,4	144,5	133,3	143,2	133,3	107,9	104,4	94,8	91,0	74,5	84,3
t_i	53,3	85,5	143,0	132,2	141,7	134,9	103,8	108,3	93,8	90,0	73,7	83,7
K_i	53,3	48,1	38,8	91,0	128,9	173,9	185,7	191,8	179,9	132,0	119,5	100,0
$1+p_i$	45	38	55	73	103	139	148	153	143	129	95	80

12.73 Primerjava sezonskih indeksov za industrijo gradbenega materiala, ki smo jih izračunali po obeh metodah, ne kaže velikih razlik in sta obe metodi pokazali tipično obliko sezonskega značaja te dejavnosti. Primerjava je lepo vidna iz slike 12.15.



Slika 12.15 Sezonski indeksi industrije gradbenega materiala v FLRJ v razdobju 1951-1957

Metoda grafičnega približka

12.74 Dosti dober približek sezonskih indeksov dobimo z razmeroma majhnim trudom tudi grafično. Metoda po svojem bistvu ni različna od metode kvocientov na vrsto drsečih sredin, le da ves posel opravimo grafično. Sezonske indekse določimo po naslednjih točkah:

a) Na pollogaritmičen papir narišemo v razmeroma velikem merilu vrsto osnovnih mesečnih podatkov.

b) Izračunamo letna povprečja in jih vrišemo v zgornji grafikon v sredine ustreznih let.

c) Skozi točke povprečij prostoročno narišemo krivuljo, tako da najbolje ponazarja skupno črto trenda in ciklov. Povprečja uporabimo pri tem kot oporne točke.

d) Na list praznega papirja narišemo pravokotno na rob črto izhodišča z vrednostjo 100. Ta papir uporabimo za črtanje odklonov Y od črte trenda in cikla. Rob papirja naravnamo tako, da izhodišče (100) leži na včrtani črti trenda in cikla za januar prvega leta. S črtico naznačimo odklon stvarne vrednosti od trenda. Na istem robu ponovimo ta postopek za januarje za vsa druga leta.

f) Ko s črticami naznačimo odklone za januarje za vsa leta prečrtamo črtici za najmanjši in največji odklon. Za druge črtice pa na oko ocenimo težišče. Težišče zaznamujemo na pasu tako, da se loči od individualnih odklonov. Težišče moremo kontrolirati in popraviti. Praviloma je vsota

negativnih in pozitivnih odklonov od težišča enaka 0. Dobljeno težišče je modificirana geometrijska sredina odklonov od trenda in je nepopravljeni sezonski indeks za januar.

g) Ko dobimo nepopravljeni sezonski indeks za januar, papir preganemo pravokotno na izhodiščno linijo in ob novem robu ponovimo postopek za februar.

Enako določimo nepopravljene sezonske indekse tudi za vse druge mesece.

h) Če razgrnemo nagubani papir, pomenijo črtice, s katerimi smo označili težišča za posamezne mesece, logaritme nepopravljenih sezonskih indeksov. Z logaritemsko odstotno skalo številčno odberemo vrednosti nepopravljenih sezonskih indeksov.

i) Če je vsota nepopravljenih sezonskih indeksov približno 1200 (različna za manj kot šest), dobljene sezonske indekse ne popravljamo dalje.

Če pa je vsota *nepopravljenih* sezonskih indeksov od 1200 različna za več kot 3, izračunamo kvocient med vsoto nepopravljenih sezonskih indeksov in 1200. To razmerje zaznamujemo na nagubanem papirju, na katerem smo včrtali individualne odklone. Novo izhodiščno črto 100 načrtamo iz tega novega nivoja.

j) Z logaritemsko odstotno skalo odberemo od nove izhodiščne črte popravljene sezonske indekse.

CIKLIČNA NIHANJA

12.75 Razen trenda in periodičnih komponent opazimo na nekaterih časovnih vrstah še ciklična nihanja. Ciklični vplivi imajo za rezultat, da pojav niha okrog trenda. Ciklična nihanja so povsem različna od sezonskega ali na splošno periodičnega nihanja. Za razliko od njih se ciklična nihanja ne pojavljajo niti po dolžini niti po obliki in amplitudah tako pravilno kakor sezonska oziroma v splošnem periodična nihanja.

Proučevanje ciklov je v ekonomiji zelo važno iz več razlogov. Poznavanje ciklov za določene gospodarske panoge oziroma elemente pokaže, koliko je določen ekonomski pojav občutljiv za splošne ciklične vplive. Ciklična gibanja je potrebno upoštevati tudi pri napovedovanju in planiranju za posamezne gospodarske aktivnosti. Če poznamo ciklična nihanja v preteklosti, jih moremo prenesti po analogiji v prihodnost.

Naloga statistike je, da dá ekonomistu metode, s katerimi more iz osnovne časovne vrste izluščiti ciklična nihanja in ugotoviti njihove značilnosti. Od teh metod bomo nakazali le eno - *rezidualno metodo*, po kateri z enostavnimi sredstvi izluščimo iz osnovne časovne vrste ciklično komponento, ki je osnova za proučevanje ciklov.

Če vzamemo, da je časovna vrsta rezultat vseh vrst vplivov: trenda T, sezonskih vplivov P, cikličnih vplivov C in slučajnih vplivov S, upoštevamo, da so navadno te komponente vezane po modelu $Y = T.C.P.S$, iz osnovne časovne vrste odstranimo vpliv trenda in sezonske komponente, če podatke osnovne časovne vrste delimo z ustreznimi vrednostmi trenda in sezonskimi indeksi. Tako dobimo, da je

$$Y/TP = C.S \qquad (12.49)$$

Ti kvocienti vsebujejo samo še ciklične in slučajne vplive. Rezultate slučajnih vplivov odstranimo, če izračunamo iz dobljenih kvocientov časovno vrsto drsečih sredin, ker se rezultati slučajnih vplivov v povprečju uničijo oziroma omilijo. Običajno izračunavamo povprečja iz ne prevelikega števila členov, da povprečja ne izravnavajo še ciklične komponente.

12.76 Ciklično komponento v časovni vrsti določimo po rezidualni metodi v naslednjih stopnjah:

a) Za časovno vrsto, ki jo proučujemo, poiščemo časovno vrsto trenda.

b) Razen tega moramo zanjo poznati sezonske indekse.

c) Posamezne člene osnovne časovne vrste Y delimo z ustreznimi vrednostmi trenda T in sezonskim indeksom P.

d) Časovno vrsto kvocientov očistimo slučajnih vplivov tako, da iz njih izračunamo časovno vrsto drsečih sredin. Časovne vrste drsečih sredin zabrišejo oziroma izravnajo ciklična nihanja, če izračunamo sredine za predolga razdobja. Zato so primerna povprečja iz nepreveликega števila členov. Zelo primerne so trimesečne ponderirane sredine, pri katerih vzamemo za pondere binomske koeficiente (1, 2, 1).

e) Ko smo tako izločili iz osnovne časovne vrste ciklično komponento, opisujemo cikle z različnimi parametri: izračunavamo povprečno dolžino cikla, povprečno amplitudo in povprečja za posamezne karakteristične dele ciklov. Razlike v dolžinah, amplitudah in drugih značilnostih ciklov opisujemo s povprečnimi absolutnimi odkloni AD za posamezno značilnost cikla. Povprečni absolutni odkloni so za proučevanje ciklov primernejše merilo variacije kakor standardni odklon SD.

12.77 Dostikrat moramo pri analizi ciklov primerjati značilnosti cikličnih nihanj, za različne raznovrstne pojave. Ciklična nihanja za raznovrstne pojave zelo dobro primerjamo, če ciklično komponento izražamo v standardiziranih z-odklonih. Te dobimo tako, da individualne odklone zaradi cikličnih nihanj reduciramo s standardnim odklonom SD_C. Tako dobimo za vsak posamezen pojav ciklično komponento izraženo v standardiziranih z-odklonih

$$z_C = C/SD_C \qquad (12.50)$$

Statistične vrste standardiziranih cikličnih nihanj so dobro primerljive med seboj, ker so neimenovana števila in jih moremo zato primerjati za različne pojave med seboj. Razen tega pa z njimi reduciramo na enotno variabilnost ciklična nihanja vseh pojavov, tako da so ciklična nihanja med seboj primerljiva.

V podrobnejšo in konkretnejšo problematiko merjenja ciklov se ne bomo spuščali.

Trinajsto poglavje

KORELACIJA

13.1 V prejšnjih poglavjih smo proučevali posamezne znake samostojno, brez zveze z drugimi znaki populacije. Te metode analize statističnih podatkov zelo podrobno obravnavajo značilnosti za posamezne znake. Vendar obsegajo samo del analize statističnih podatkov. Ne zajamejo namreč enega izmed najvažnejših problemov v analizi socialno-ekonomskih pojavov in množičnih pojavov na sploh, to je njihovo medsebojno odvisnost in povezanost. Pri proučevanju socialno-ekonomskih pojavov zelo često naletimo na problem povezanosti in odvisnosti med pojavi. Obseg proizvodnje kmetijskega obrata je odvisen od velikosti obrata, strukture osnovnih sredstev, delovne sile, klimatskih faktorjev itd. Vrednost proizvodnje industrijskega obrata je odvisna od števila delavstva, mehanizacije, produktivnosti dela, vrste proizvodnje itd. Cena je odvisna od količine blaga, ki je na trgu, plača od kvalifikacije, službene dobe itd. V medsebojni povezavi sta tudi nepismenost moških in nepismenost žensk po občinah, starost ženina in neveste, itd. Podobnih primerov moremo našteti veliko. Vendar niso vsi pojavi med seboj povezani, čeprav bi po vsebini mogli odvisnost pričakovati. Imamo tudi primere nesmiselne povezave. Nesmiselno je n.pr. proučevati odvisnost med številom porok in množino padavin po letih, cene kmetijskih pridelkov od obolelosti za tuberkulozo itd.

FUNKCIJSKE ODVISNOSTI

13.2 O funkcijski odvisnosti med x in y v matematičnem smislu govorimo, če je dano neko pravilo zveze med neodvisno spremenljivko x in odvisno spremenljivko y, po katerem določeni vrednosti neodvisne spremenljivke x, ustreza ena ali več natančno določenih vrednosti odvisne spremenljivke y. S simbolom pišemo:

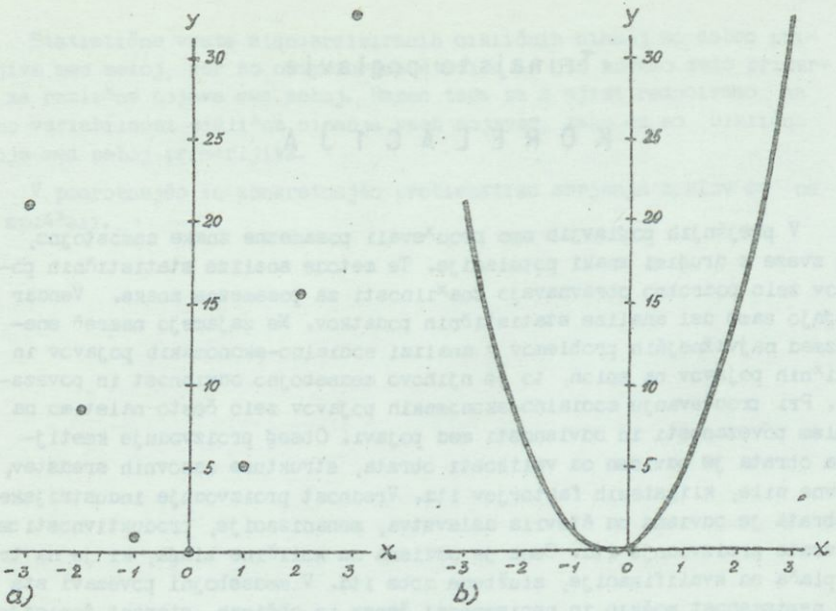
$$y = f(x) \tag{13.1}$$

in pravimo: y je funkcija od x.

Funkcijska zveza med y in x je običajno nakazana z *enačbo funkcije*, iz katere moremo za vsak x izračunati ustrezno vrednost y. Tako je n.pr. za funkcijo

$$y = 2x + 3x^2 \tag{13.2}$$

$x = 2$ ustrezna vrednost $y = 2.2 + 3.2^2 = 16$.



Slika 13.1 Slika funkcijske odvisnosti $y = 2x + 3x^2$

Razen z obrazcem moremo funkcijsko odvisnost med x in y nakazati tudi z *nizom dvojic* ustreznih vrednosti x in y. Tako je v tabeli 13.1 dana funkcijska zveza iz zgornje enačbe za osem parov vrednosti x in y

Tabela 13.1

x	y
-3	+21
-2	+ 8
-1	+ 1
0	+ 0
+1	+ 5
+2	+16
+3	+33
+4	+56

S sistemom dvojic pa ne moremo podati funkcijske zveze med x in y za vse vrednosti x. To moremo prikazati edino z enačbo med x in y.

Niz dvojic vrednosti x in y, moremo prikazati tudi *grafično* v pravokotnem koordinatnem sistemu. Posamezno dvojico vrednosti x, y po znanem načinu prikažemo s točko v pravokotnem koordinatnem sistemu. Kolikor dvo-

jic ustreznih vrednosti x in y imamo, toliko imamo točk v koordinatnem sistemu. Medtem ko moremo niz dvojic vrednosti grafično prikazati s sistemom točk, prikažemo funkcijo, dano z obrazcem, s krivuljo. Grafično vsaki vrednosti x v določenem razmaku ustreza funkcijska vrednost y. Slika niza dvojic iz tabele 18.1 je prikazana v sliki 18.1a, funkcijska slika odvisnosti iz obrazca 18.2 pa v sliki 18.1b.

Funkcijske odvisnosti moremo torej prikazati na tri načine:

a) z enačbo: $y = f(x)$,
b) v tabeli z nizom ustreznih vrednosti x in y.
c) grafično s sistemom točk ali s krivuljo v pravokotnem koordinatnem sistemu.

KORELACIJSKE ODVISNOSTI

13.3 Če preneseмo pojem funkcijske odvisnosti na množične pojave, bi imela v primeru funkcijske odvisnosti med površino in proizvodnjo kmetijskega gospodarstva vsa gospodarstva z enako površino enako proizvodnjo. Vendar to ni tako. Čeprav sta površina gospodarstva in proizvodnja med seboj odvisni, imata gospodarstvi z enakima površinama le redko enako proizvodnjo. Še več. Čeprav sodimo, da ima večje gospodarstvo večjo proizvodnjo, velja ta odvisnost samo na splošno, v posameznih primerih pa more imeti tudi večje gospodarstvo manjšo proizvodnjo. Do tega pride zato, ker proizvodnja ni odvisna samo od površine, temveč še od mnogo drugih faktorjev. Poizkus, da bi odstranili vse dodatne faktorje in tako prišli do funkcijske povezave med obema znakoma, se ne bi posrečil. Vedno ostanejo neki faktorji, katerih vpliv ne moremo odstraniti in jih štejemo med slučajne faktorje. Pri množičnih pojavih moremo torej opazovati le splošno tendenco odvisnosti, v posameznih primerih pa zakonitost zaradi delovanja dodatnih individualnih vplivov ni nujna. Zato imenujemo za razliko od *funkcijskih* te vrste odvisnosti *korelacijske odvisnosti*.

Proučevanje korelacijskih odvisnosti je različno od proučevanja funkcijskih odvisnosti, čeprav imata obe vrsti proučevanja svoje stične točke.

Prikazovanje korelacijskih odvisnosti

Korelacijske odvisnosti prikazujemo na enake tri načine kakor funkcijske odvisnosti:

a) v tabeli z nizom dvojic vrednosti koreliranih podatkov za vsako enoto populacije, ali v korelacijski tabeli

b) s točkami v korelacijskem grafikonu,

c) v funkcijski obliki z regresijsko funkcijo ali črto.

13.4 Niz dvojic podatkov. Z nizom dvojic vrednosti koreliranih podatkov na splošno prikazujemo osnovne podatke pri proučevanju korelacije med pojavi. Čeprav je ta način nepregleden in iz njega še ne dobimo vtisa o zakonitosti povezave med dvema pojavoma, ga na splošno uporabljamo, ker je osnova za vsa nadaljnja proučevanja.

Ker je vir podatkov o množičnih pojavih populacija, posamezna dvojica koreliranih podatkov velja za posamezne enote proučevane populacije.

V tabeli 13.2 so prikazani osnovni podatki za proučevanje korelacije med številom delavcev in številom opravljenih delovnih ur 19 industrijskih podjetij obutvene industrije.

Tabela 13.2 Število delavcev (x) in število opravljenih delovnih ur (v tisočih) (y) v decembru 1952 za 19 industrijskih podjetij z manj kot 200 delavcev za obutveno industrijo v LRS. (Vir: Mesečna statistična služba v industriji LRS)

x	154	51	57	15	92	112	23	150	58	81	28
y	35,3	10,1	11,0	2,9	17,9	25,4	4,1	31,3	10,0	15,1	4,9
x	100	37	111	12	50	7	80	49			
y	16,7	7,3	20,7	2,2	9,7	0,8	15,1	9,9			

V tabeli 13.3 so prikazani dohodki in stroški za kulturno in družbeno življenje za 33 delavskih družin v Mariboru v novembru 1957.

Tabela 13.3 Dohodki v tisočih dinarjev (x) in stroški za kulturno in družbeno življenje v dinarjih (y) za 33 delavskih družin v Mariboru v novembru 1957. (Vir: Anketa o življenju delavcev in nameščencev ZS LRS)

x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7	16,5
y	260	330	490	310	580	560	330	480	590	330	630
x	10,9	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4
y	630	350	1100	760	460	700	540	750	370	630	740
x	28,4	25,8	3,5	29,8	27,5	26,5	26,7	26,7	32,9	37,5	30,8
y	1080	630	990	1210	1100	870	640	1110	610	1590	1070
x	31,8	38,4	35,9								
y	1600	1710	1808								

V tabeli 13.4 so podatki o odstotkih travniških in pašniških površin za okraje v LRS po stanju leta 1952.

Tabela 13.4 Odstotki travniških in pašniških površin po okrajih v LRS po stanju leta 1952. (Vir: Statistični bilten LRS)

Okraj	Odstotek površin pod	
	travniki	pašniki
Celje mesto	22	8
Celje okolica	3	28
Črnomelj	9	20
Gorica	5	13
Kočevje	6	15
Kranj	7	7
Krško	13	9
Ljubljana-mesto	28	3
Ljubljana-okolica	18	11
Ljutomer	18	6
Maribor-mesto	15	5
Maribor-okolica	15	8
Murska Sobota	21	3
Novo mesto	8	9
Postojna	10	19
Ptuj	18	10
Radovljica	5	16
Sežana	1	37
Slovenj Gradec	7	10
Šoštanj	8	9
Tolmin	5	20
Trbovlje	9	8

13.5 Iz zgornjih treh primerov moremo sklepati še na določeno vsebinsko razliko med korelacijami. V prvih dveh primerih je jasno, da je število opravljenih delovnih ur odvisno od števila delavcev, stroški za kulturne in družbene potrebe pa od dohodkov, ne pa obratno. To so *vzročne* povezave oziroma odvisnosti, ker je en znak vzrok, drugi pa posledica. Vzročne povezave pa ne zasledimo v tretjem primeru, ker ne moremo reči, da je visok odstotek travniških površin vzrok za nizek odstotek pašniških površin ali obratno. Kljub temu, da med njima ni vzročne odvisnosti, pa sta ta dva podatka med seboj povezana, ker na oba vplivajo isti pojavi, ki imajo v našem primeru za posledico, da je travnikov mnogo, pašnikov pa malo in obratno. Dva taka odločilna skupna faktorja sta n. pr. nadmorska višina in vrsta tal.

Na podobno nevzročno čisto korelacijsko povezavo tudi naletimo, če proučujemo korelacijo med odstotkom nepismenih žensk in moških po okrajih. Čeprav med njima ni vzročne povezanosti, sta med seboj v korelaciji, ker na oba deluje isti kompleks faktorjev, od katerih je odvisna nepismenost. To so n.pr. število šol v okraju, kulturno prosvetna dejavnost v preteklosti itd.

Tudi korelacija med starostjo moža in žene ob razvezi zakona iz tabele 13.5 je primer čiste korelacijske povezave.

Medtem ko pri vzročnih odvisnostih po pravilu iščemo samo odvisnost posledice od vzroka, moremo pri čistih korelacijskih odvisnostih iskati odvisnost znaka x od y in obratno odvisnost znaka y od x.

13.6 **Korelacijska tabela.** Že pri majhnem številu enot je prikazovanje koreliranih podatkov v zgornji obliki zelo obširno in nepregledno. Zato za večje populacije prikazujemo korelirane podatke v kombinacijski tabeli, ki jo imenujemo korelacijska tabela. Kakor dobimo frekvenčno distribucijo, če podatke uredimo v razrede po enem znaku, dobimo korelacijsko tabelo, če podatke uredimo v razrede po obeh koreliranih znakih hkrati. Položaj frekvenc v korelacijski tabeli zelo nazorno pokaže smer povezave med koreliranima znakoma.

Kot primer je v tabeli 13.5 prikazana korelacijska tabela o razvezanih zakonih po starosti zakoncev za zakone, razvezane v letu 1955 v LR Sloveniji.

Tabela 13.5 Korelacijska tabela o razvezanih zakonih po starosti zakoncev za zakone, razvezane v letu 1955 v Sloveniji
(Vir: Vitalna statistika za 1955)

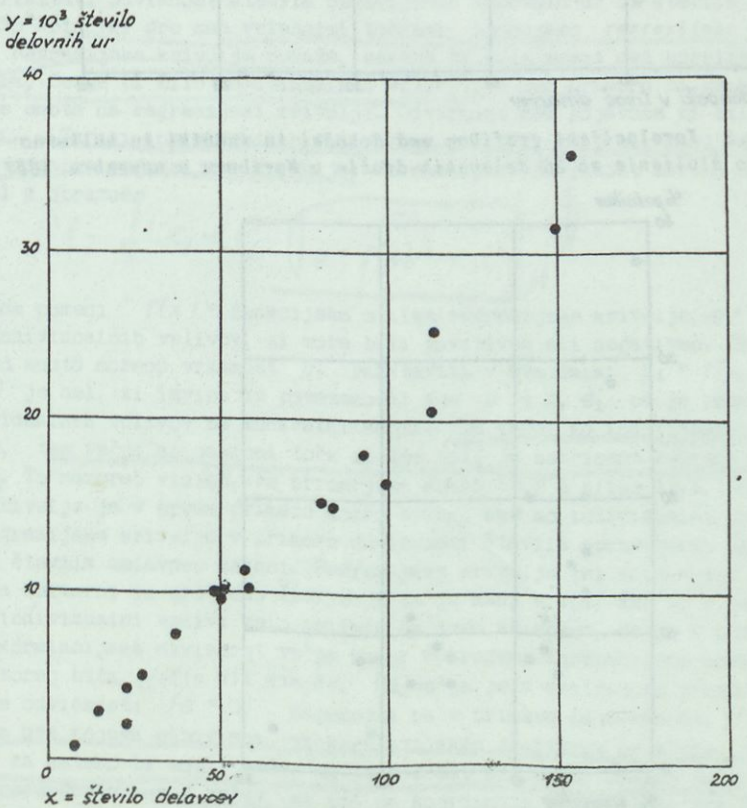
Starost žene let	Starost moža let											Skupno
	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	
15–19	2	2	2									6
20–24	48	82	28	8	4			1				171
25–29	15	108	91	32	10	5		1				262
30–34	5	36	77	67	37	17	7	3	1			250
35–39	1	4	17	33	42	19	3	3	1			123
40–44	1	2	9	15	52	35	9	5	1	1		130
45–49	–	–	3	7	21	31	17	8	4	–		91
50–54	–	–	1	2	6	8	13	6	–	2	2	40
55–59					2	8	4	9	8	3	2	36
60–64					1	–	2	3	5	4	2	17
65–69						1	1	1	3	2	1	13
70–74							1	–	2	–	1	4
Skupno	72	234	228	134	175	124	57	40	25	12	8	1139

Podatki v tabeli pomenijo frekvence, t. j. število razvezanih zakonov, v katerih je bila starost moža in žene v ustreznih razredih, n. pr.: v 21 razvezanih zakonih je bila starost moža ob razvezi med 30-34 let, starost žene pa med 25-29 let.

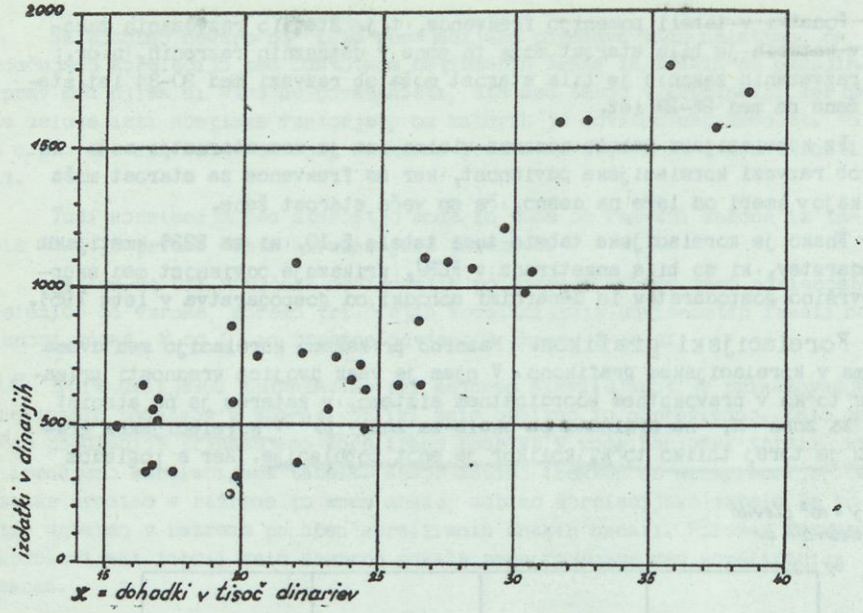
Iz korelacijske tabele nazorno vidimo, da je med starostjo moža in žene ob razvezi korelacijska odvisnost, ker se frekvence za starost moža premikajo v smeri od leve na desno, če se veča starost žene.

Enako je korelacijska tabela tudi tabela 5.10, ki za 2233 kmetijskih gospodarstev, ki so bila anketirana v FLRJ, prikazuje odvisnost med skupno površino gospodarstev in denarnimi dohodki od gospodarstva v letu 1953.

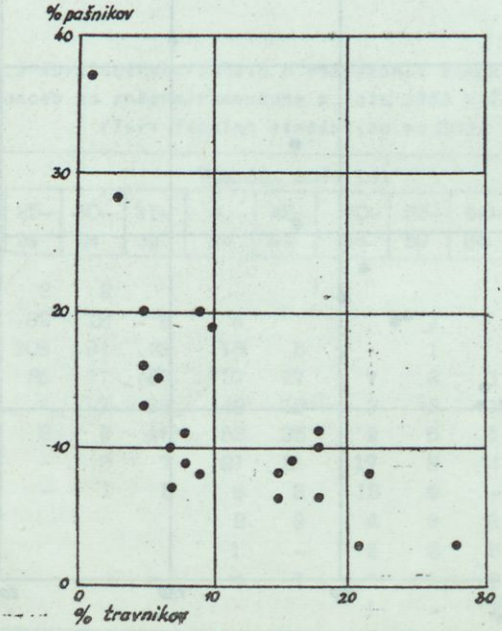
13.7 Korelacijski grafikon. Nazorno prikažemo korelacijo med dvema znakoma v korelacijskem grafikonu. V njem je vsak dvojica vrednosti prikazana s točko v pravokotnem koordinatnem sistemu, v katerem je na abscisi skala za znak x, na ordinati pa skala za znak y. V korelacijskem grafikonu je torej toliko točk, kolikor je enot populacije. Ker s pogledom



Slika 13.2 Korelacijski grafikon med številom delavcev in številom opravljenih delovnih ur za 19 podjetij obutvene industrije v letu 1952



Slika 13.3 Korelacijski grafikon med dohodki in izdatki za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957



Slika 13.4 Korelacijski grafikon med odstotkom travnikov in odstotkom pašnikov za okraje v LRS v letu 1952

zajemamo vse točke hkrati, je korelacija med obema znakoma v korelacijskem grafikonu vidna neposredno.

V slikah 13.2, 13.3 in 13.4 so prikazani korelacijski grafikoni podatkov iz tabel 13.2, 13.3 in 13.4.

Regresijska krivulja

13.8 Slike 13.2, 13.3, 13.4, so zelo poučne in pokažejo bistvo korelacijskih odvisnosti. Kakor smo že nakazali, med socialno-ekonomskimi pojavi ni funkcijske zveze, ker pojavi niso nikdar odvisni samo od enega znaka, temveč vplivajo nanj vedno najrazličnejši individualni vplivi, ki motijo oziroma zabrišejo zvezo med obema pojavoma. Iz slike 13.2, ki prikazuje korelacijo med številom delavcev in številom opravljenih delovnih ur v decembru 1952 za 19 industrijskih podjetij obutvene industrije v LRS, sklepamo, da je odvisnost števila opravljenih delovnih ur od števila delavcev velika, ker se točke zelo določno gosto okrog neke črte, s katero bi mogli prikazati odvisnost števila opravljenih delovnih ur od števila delavcev. To črto, ki gre med vrisanimi točkami, imenujemo *regresijsko krivuljo*. Regresijska krivulja pokaže, kakšna bi bila zveza med koreliranima znakoma, če ne bi bilo individualnih vplivov. V tem primeru bi bile točke za vse enote na regresijski krivulji. Odvisnost med pojavoma bi bila funkcijska. Zaradi individualnih vplivov pa se točke bolj ali manj odklanjajo od idealne regresijske krivulje. Korelacijsko odvisnost moremo torej pisati z obrazcem

$$y = f(x) + e \qquad (13.8)$$

Pri tem pomeni: $f(x)$ = funkcijska oblika regresijske krivulje, e = rezultat individualnih vplivov, ki more biti pozitiven ali negativen. Za konkretno enoto moremo vrednost y_1 razstaviti v dva dela: $y_1 = f(x_1) + e_1$; $f(x_1)$ je del, ki izvira iz povezanosti med y in x, e_1 pa je rezultat individualnih vplivov na konkretni enoti. Čim večji so individualni vplivi e, tem večji so odkloni točk in tem bolj je zabrisana regresijska krivulja. To nazorno vidimo, če primerjamo sliko 13.2 s sliko 13.3. Regresijska krivulja je v prvem primeru dobro vidna, ker so individualni odkloni od regresijske krivulje v primeru odvisnosti števila opravljenih delovnih ur od števila delavcev majhni. Regresijska krivulja pri korelaciji stroškov za kulturno in družbeno življenje pa je manj vidna, ker so v tem primeru individualni vplivi zelo veliki. Iz tega sklepamo, da je v prvem primeru korelacijska odvisnost večja kakor v drugem. Korelacijska odvisnost more torej biti *večja ali manjša*. Največja je v ekstremnem primeru funkcijske odvisnosti $(e = 0)$, najmanjša pa v primeru neodvisnosti $(f(x) = 0)$. Če sta dva pojava neodvisna, na korelacijskem grafikonu ne moremo začrtati črte, za katero bi mogli reči, da je regresijska črta. Takrat se točke brez reda goste okrog točke, ki ima za koordinati aritmetični sredini koreliranih znakov.

V prvih dveh zgornjih primerih opazimo, da je smer regresijske krivulje taka, da se večja vrednost enega znaka, če se večja vrednost drugega. Takim vrstam povezav pravimo *pozitivne povezave*. Če pa se z večanjem enega znaka vrednost drugega manjša, pa pravimo, da je *povezava negativna*. Negativna povezanost je med odstotkoma travniške in pašniške površine po okrajih, prikazana v sliki 13.1. Iz gostitve točk moremo sklepati, da je za okraje z večjim odstotkom travnikov odstotek pašnikov v splošnem manjši in obratno.

Iz slike o odvisnosti števila opravljenih delovnih ur od števila delavcev vidimo, da je črta, ki ponazarja regresijsko krivuljo, premica, v obeh drugih primerih pa krivulja. V prvem primeru govorimo o *linearni povezavi*, v obeh drugih pa o *krivuljični povezavi* med dvema pojavoma.

Metode določanja regresijskih krivulj

13.9 Eden izmed osnovnih problemov regresijske in korelacijske analize je določitev regresijske krivulje. Določamo jo v splošnem po treh metodah:

a) prostoročno,

b) z grupnimi sredinami,

c) analitično.

13.10 Prostoročno določanje regresijske krivulje. Iz slik 13.2, 13.3 in 13.4, vidimo, da je smer regresijske krivulje bolj ali manj vidna že iz korelacijskega grafikona. Zato moremo pri nekaterih slabše, pri drugih pa bolje včrtati regresijsko krivuljo na oko. Za regresijsko krivuljo, imamo krivuljo, ki jo včrtamo med točke tako, da se meglici točk najbolje prilega. Sevsda ne sme biti vrisana črta preveč komplicirana, marveč čimbolj izglajena. V naših primerih regresijsko krivuljo najlaže včrtamo za pojava, prikazana v sliki 13.2, za katera je povezava največja. Zato je regresijska krivulja najbolj vidna. Težja je določitev v drugih dveh primerih.

Metoda prostoročnega včrtovanja regresijskih črt je prikladna, ker je izredno preprosta. Ima pa to napako, da je subjektivna. Uporabljamo jo običajno v prvi stopnji študija regresije kot osnovo za druge metode. V dosti primerih pa je kljub subjektivnosti ravno zaradi svoje hitrosti v primerjavi z drugimi metodami zelo koristna.

13.11 Metoda grupnih sredin. Če izhajamo iz obrazca

$$y = f(x) + e, \qquad (13.4)$$

moremo za določanje regresijske krivulje s pridom uporabiti grupne sredine. Če za enote, ki imajo iste ali ne preveč različne vrednosti x, izračunamo povprečje iz vrednosti y, dobimo:

$$\overline{y} = \overline{f(x)} + \overline{e} \approx f(x) \qquad (13.5)$$

Ker je namreč x konstanten, je konstantna tudi vrednost $f(x)$, povprečje konstante pa konstanta. Po znanih stavkih o sredinah pa se v povprečju vpliv individualnih ali slučajnih faktorjev sicer ne uniči, omili se pa vsekakor. Z vrsto povprečij po grupah x dobimo tako niz točk, ki so vsaj v bližini regresijske krivulje, če že niso na njej. Če te točke vrišemo v korelacijski grafikon in zvežemo z daljicami, dobimo lomljeno črto; ta je zelo dober približek poteka regresijske krivulje.

Po metodi grupnih sredin določamo regresijsko krivuljo torej takole:

a) Dvojice vrednosti x, y grupiramo v razrede po vrednostih x

Grupe morajo biti take, da število enot v njih ni niti preveliko niti premajhno.

b) Za vsako grupo izračunamo aritmetični sredini za vrednosti x in y.

Tako dobimo vrsto dvojic grupnih aritmetičnih sredin $\overline{x}_k$ in $\overline{y}_k$.

c) V korelacijskem grafikonu narišemo točke, ki imajo kot koordinati ustrezne grupne aritmetične sredine za x in y $(\overline{x}_k , \overline{y}_k)$. Tako imamo v grafikonu toliko točk, ki leže na regresijski krivulji ali v njeni bližini, kolikor imamo grup.

d) Vrisane točke grupnih sredin med seboj zvežemo z daljicami.

Lomljena črta, ki jo dobimo, je približno regresijska krivulja.

13.12 Po tej metodi določimo regresijsko krivuljo za odvisnost stroškov za kulturno in družbeno življenje za 33 delavskih gospodinjstev v Mariboru. Podatki so vzeti iz tabele 13.3.

Tabela 13.6 Dohodki in izdatki za kulturno in družbeno življenje iz tabele 13.3, grupirani po višini dohodkov.
(x = Dohodki v tisočih dinarjev; y = izdatki v dinarjih)

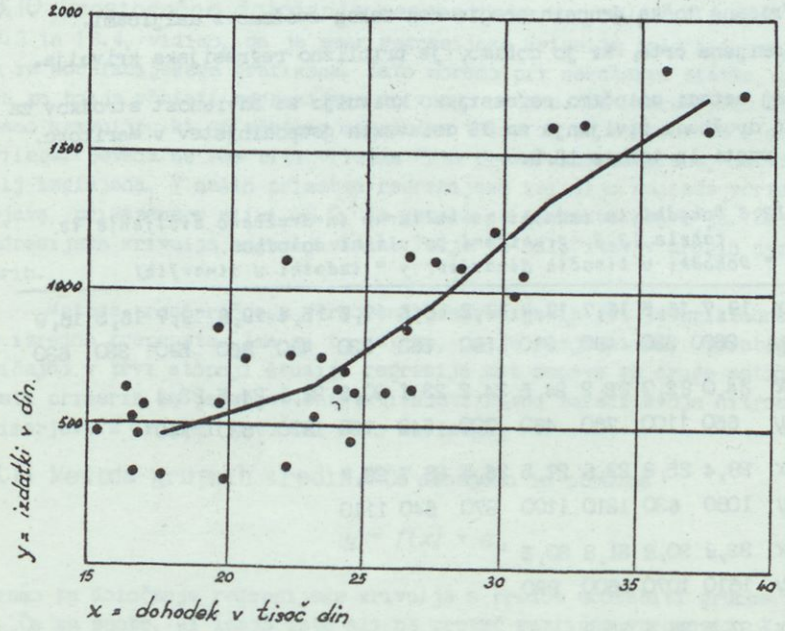
15,0–19,9	x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7	16,5	16,9
	y	260	330	490	310	580	560	330	480	860	590	330	630
20,0–24,9	x	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4		
	y	650	1100	760	460	700	540	750	370	630	740		
25,0–29,9	x	28,4	25,8	29,8	27,5	26,5	26,7	26,7					
	y	1080	630	1210	1100	870	640	1110					
30,0–34,9	x	32,9	30,8	31,8	30,5								
	y	1610	1070	1600	990								
35,0–	x	37,5	38,4	35,9									
	y	1590	1710	1810									

Iz grupiranih podatkov izračunamo grupne sredine za x in y.

Ker je med dohodki in stroški vzročna zveza, vzemimo dohodke za neodvisno spremenljivko (x), stroške pa za odvisno spremenljivko (y). Po analizi podatkov glede na velikost dohodkov vzamemo razrede dohodkov po 5000 din in po njih razdelimo dvojice vrednosti x in y. Tako grupirani podatki so navedeni v tabeli 13.3.

Tabela 13.7 Grupne sredine za dohodke in izdatke za kulturno in družbeno življenje za 36 delavskih družin v Mariboru iz tabele 13.6

Dohodki v tisočih din	N_k	Y_k	Y_k	$\overline{x_k}$	$\overline{y_k}$
15,0 - 19,9	12	212,2	5750	17,7	479
20,0 - 24,9	10	230,3	3700	23,1	370
25,0 - 29,9	7	191,4	3340	27,3	949
30,0 - 34,5	4	123,0	5270	31,5	1318
35,0 -	3	111,8	5110	37,3	1703



Slika 13.5 Regresijska črta med dohodki in izdatki za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957

Če grupne sredine vnesemo v korelacijski grafikon, dobimo sliko 13.5. Iz slike vidimo, da se dobljena regresijska črta resnično prilagaja osnovnim vrednostim in da ponazarja smer odvisnosti med dohodki in stroški za kulturno in družbeno življenje.

13.13 Po metodi grupnih sredin moremo določiti regresijsko črto tudi, če so podatki grupirani v korelacijski tabeli. Grupne sredine v tem primeru izračunamo po kateri koli izmed metod za izračunavanje sredin iz frekvenčnih distribucij. Te grupne sredine včrtamo nad sredine razredniih razmakov.

Za korelacijo med starostjo moža in starostjo žene ob razvezi zakona iz tabele 13.5 je smiselno izračunati obe regresijski serij sredin, ker zveza med starostima ni vzročna. Ker so v korelacijski tabeli podatki že grupirani, grupiranje odpade.

Če najprej izračunamo povprečja za starost žene ob razvezi zakona za posamezne starostne grupe moža, dobimo vrsto sredin za regresijsko črto, ki naj pokaže odvisnost starosti žene od starosti moža. Za prvo starostno grupo moža (20–24 let) izračunamo povprečno starost žene iz prve navpične frekvenčne distribucije, (frekvence te distribucije so po vrsti: 2, 48, 15, 5, 1, 1). Za njo dobimo, da je povprečna starost žen, ki se razvežejo in so njihovi možje stari od 20–24 let ali povprečno 22,5 let, enaka $\bar{y}_1 = 24,3$ let.

Če izračunamo povprečje iz analognih frekvenčnih distribucij še za druge grupe starosti moža, dobimo vrsto povprečnih starosti žena, ki je prikazana v tabeli 13.8a.

Tabela 13.8a Povprečne starosti žena ob razvezi zakona po petletnih grupah starosti moža ob razvezi v LR Sloveniji v letu 1955

Starost moža let	x	20–24	25–29	30–34	35–39	40–44	45–49	50–54
Povprečna starost žene let	$\bar{y}_k$	24,6	23,7	30,2	33,8	39,1	42,8	47,5
	x	55–59	60–64	65–69	70–74			
	$\bar{y}_k$	49,5	56,9	60,0	60,6			

Če hočemo podobno izračunati povprečne starosti moža za posamezne grupe starosti žena, moramo izračunati grupne sredine iz vodoravnih frekvenčnih distribucij. Za prvo starostno skupino 15–19 let za žene ima vodoravna frekvenčna distribucija tele frekvence: 2, 2, 2.

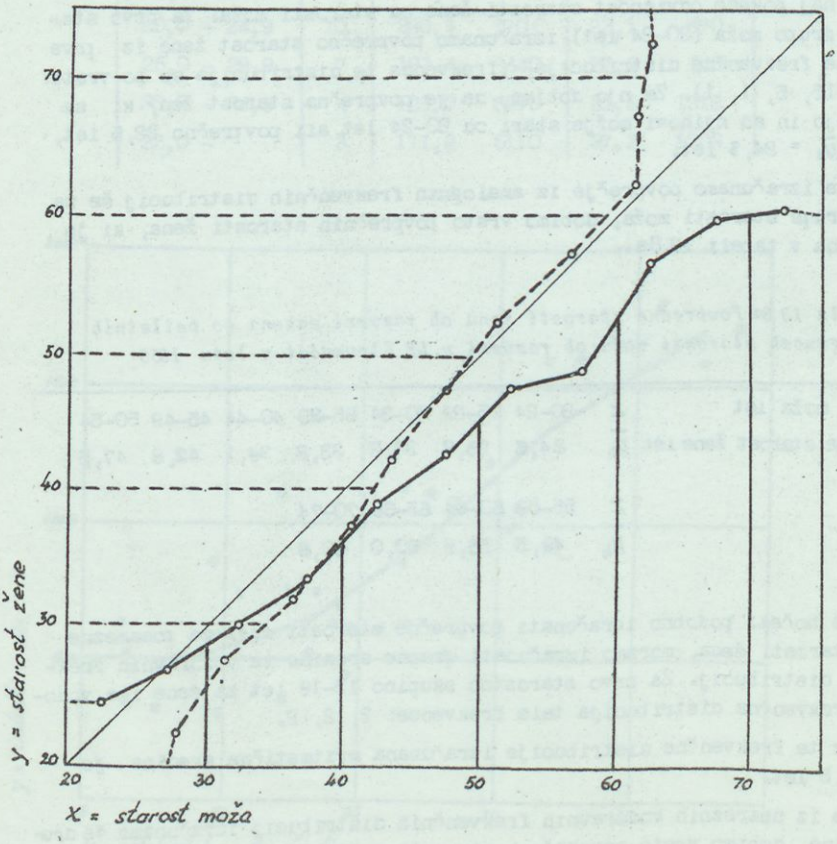
Iz te frekvenčne distribucije izračunana aritmetična sredina je $x_1 = 27,5$ let.

Če iz ustreznih vodoravnih frekvenčnih distribucij izračunamo še druge sredine, dobimo vrsto povprečnih starosti moža za posamezne grupe starosti žena. Ta vrsta sredin pokaže odvisnost starosti mož od starosti žena.

Tabela 13.8b Povprečne starosti mož ob razvezi zakona po petletnih grupah starosti žene ob razvezi zakona v LR Sloveniji v letu 1955

Starost žene let	y	15–19	20–24	25–29	30–34	35–39	40–44	45–49
Povprečna starost moža let	$\bar{x}_k$	27,5	27,9	31,2	36,4	40,7	43,8	47,6
	y	50–54	55–59	60–64	65–69	70+74		
	$\bar{x}_k$	51,2	56,7	61,3	61,4	62,5		

Zgornji vrsti, narisani v sliki 13.3, sta zelo ilustrativni. V grafikonu je vrisana tudi premica pod kotom 45°, ki kaže enako starost žene in moža. Regresijska črta odvisnosti starosti žene od starosti moža pokaže, da je povprečna starost žen ob razvezi zakona, razen v razredu 20+24 let, manjša



Slika 13.6 Regresijski črti grupnih sredin za starost moža in žene ob razvezi zakona v letu 1955 v LR Sloveniji

od povprečne starosti moža in da se ta razlika s starostjo moža veča. Povprečna starost mož ob razvezi pa je do starostne grupe 40–44 let za žene večja, po tej starosti pa je za malenkost manjša od starosti žene. Iz slike vidimo, da imamo dve regresijski črti, ki se v grafikonu po pravilu križata. Ker je populacija velika, je potek približka regresijskih črt zadosti umirjen in dobro ponazarja regresijsko krivuljo.

13.14 Analitična metoda. Regresijsko krivuljo moremo določiti tudi analitično. Če se po analizi podatkov odločimo, da je določen tip funkcije

$$y' = f(x; a, b, c \ldots), \tag{13.3}$$

ki ima več parametrov $(a, b, c \ldots)$, primeren za regresijsko krivuljo, je treba ustrezno podatkom, ki jih proučujemo, poiskati vrednosti parametrov $a, b, c \ldots$ tako, da se regresijska krivulja danim vrednostim najbolj prilega. Kot merilo boljše ali slabše prilagojenosti krivulje vzamemo vsoto kvadratov odklonov stvarnih vrednosti y od y' na regresijski krivulji

$$\Sigma(y - y')^2 = F(a, b, c \ldots) \tag{13.7}$$

Za regresijsko krivuljo imamo izmed vseh krivulj istega tipa tisto, za katero je vsota kvadratov odklonov stvarnih vrednosti y od vrednosti y' na regresijski krivulji, $F(a, b, c \ldots)$, najmanjša. Po tej metodi, ki jo imenujemo *metodo najmanjših kvadratov*, moremo določiti vrednosti parametrov za regresijsko krivuljo in tako regresijsko krivuljo samo. Kadar ugotavljamo regresijsko krivuljo analitično, uporabljamo najpogosteje metodo najmanjših kvadratov.

Indeks korelacije

13.15 Če vzamemo, da je y kvantitativen izraz delovanja vseh faktorjev, ki vplivajo na določen pojav, moremo vse faktorje razdeliti v tri grupe: a) na splošne vplive, ki so za vse enote isti: b) na faktor x, s katerim je y v korelaciji: c) na ostale individualne vplive skupno s slučajnimi.

Tako je na primer plača uslužbencev odvisna od : a) splošnih vplivov: kategorije uslužbencev, časa, za katerega plača velja itd., b) let službe, s katerimi je plača v korelacijski zvezi, c) drugih individualnih vplivov, kot so: vestnost pri delu, šolska izobrazba, delovno mesto itd.

Prav tako je na primer proizvodnja kmetijskega obrata odvisna od: a) splošnih faktorjev: klimatskih pogojev, specializiranosti itd. b) velikosti gospodarstva, s katero je proizvodnja v korelacijski povezanosti in c) od drugih individualnih faktorjev, ki vplivajo na proizvodnjo, kot so: razlike tal, različna produktivnost dela, različno seme itd.

Če poznamo aritmetično sredino $\bar{y}$ in regresijsko krivuljo y' med x in y, moremo celotno vrednost y za posamezno enoto razdeliti na te tri komponente, od katerih vsaka prikazuje rezultat vplivov po zgornjih treh skupinah.

y je rezultat vseh faktorjev, ki vplivajo na pojav; splošnih, koreliranega in drugih individualnih; y' je rezultat splošnih faktorjev in koreliranega faktorja x; $\bar{y}$ pa je rezultat samo splošnih faktorjev, zato moremo pisati

$$y = \bar{y} + (y'-\bar{y}) + (y-y') \tag{13.8}$$

Pri tem pomeni: $\bar{y}$ = rezultat splošnih vplivov; $(y'-\bar{y})$ = rezultat vpliva koreliranega znaka x; $(y-y')$ = rezultat vpliva drugih individualnih faktorjev.

Tako smo uspeli razdeliti celotno vrednost y po značaju faktorjev v tri ločene komponente.

Če je vpliv faktorja x velik, je komponenta $(y'-\bar{y})$ velika. Če je vpliv faktorja x majhen, so vrednosti $(y'-\bar{y})$ majhne, če pa y od x ni odvisen, je $(y'-\bar{y})$ za vse enote enak nič. Za skupno merilo velikosti vpliva x na y uporabimo varianco odklonov $(y'-\bar{y})$

$$\sigma^2_{y.x} = \frac{1}{N}\Sigma\,(y'-\bar{y})^2 \tag{13.9}$$

Enako so odkloni $(y-y')$ veliki, če so drugi individualni faktorji močni. Če je vpliv teh faktorjev majhen, so ti odkloni majhni, če jih ni, je $(y-y')$ enak nič. Povezava je v tem primeru funkcijska. Skupno merilo jakosti individualnih vplivov je analogno zgornjemu merilu variance odklonov $(y-y')$

$$\sigma^2_e = \frac{1}{N}\Sigma\,(y-y')^2 \tag{13.10}$$

Po stavku o razstavljanju variance na vsoto varianc po faktorjih (glej obrazec 10.14) velja, da je skupna varianca σ^2_y vsota dveh varianc: variance zaradi vpliva faktorja x na y $\sigma^2_{y.x}$ in variance zaradi vpliva drugih individualnih faktorjev σ^2_e.

Ker tako pojasnimo, da en del celotne variance izvira iz povezave y z x, imenujemo $\sigma^2_{y.x}$ *pojasnjeno varianco*, za razliko od σ^2_e, ki izvira iz drugih, neznanih faktorjev in jo imenujemo *nepojasnjeno varianco*. Z obrazcem moremo zgornje napisati:

$$\sigma^2_y = \sigma^2_{y.x} + \sigma^2_e \tag{13.11}$$

Če to enačbo delimo z σ^2_y, dobimo obrazec

$$\frac{\sigma^2_{y.x}}{\sigma^2_y} + \frac{\sigma^2_e}{\sigma^2_y} = 1 \qquad (13.12)$$

V tem obrazcu kvocient $\sigma^2_{y.x}/\sigma^2_y$ pove, koliki del skupne variance je pojasnjene s povezanostjo y z x, σ^2_e/σ^2_y pa, koliki del skupne variance odpade na druge, nepojasnjene vplive.

Velikost izraza $\sigma^2_{y.x}/\sigma^2_y$ se ravna po tem, kako velik je vpliv x na y. Če y od x ni odvisen, je vrednost tega izraza nič, če pa je odvisnost funkcijska, je ta izraz največji, in sicer enak ena. Vrednost kvocienta $\sigma^2_{y.x}/\sigma^2_y$ je torej med 0 in 1 in pomeni večjo ali manjšo stopnjo vpliva x na y. Zato kvocient $\sigma^2_{y.x}/\sigma^2_y$ uporabljamo za merilo stopnje odvisnosti y od x in ga imenujemo *determinacijski koeficient*. Čim večji je vpliv x na y, tem večja je namreč vrednost $\sigma^2_{y.x}/\sigma^2_y$.

Čeprav je determinacijski koeficient $\sigma^2_{y.x}/\sigma^2_y$ vsebinsko zelo nazoren, saj je delež pojasnjene variance od skupne variance, dostikrat namesto njega uporabljamo kot merilo povezanosti kvadratni koren tega izraza. To merilo korelacije imenujemo *indeks korelacije*. Indeks korelacije je torej razmerje med standardnim odklonom za pojasnjeno variabilnost in skupnim standardnim odklonom

$$I_{y.x} = \sigma_{y.x}/\sigma_y \qquad (13.13)$$

Če upoštevamo enačbi 13.12 in 13.13, moremo indeks korelacije pisati v obliki, ki jo splošno uporabljamo

$$I_{y.x} = \sqrt{1 - \frac{\sigma^2_e}{\sigma^2_y}} \qquad (13.14)$$

Standardna napaka ocene

13.16 Če je povezava med x in y tesna, se točke za dejanske vrednosti x in y razmeroma malo odklanjajo od regresijske krivulje. Ako za dano enoto poznamo samo vrednost x, moremo vzeti ustrezno vrednost y' na regresijski črti za oceno stvarne vrednosti y. To ocenjevanje ima veliko praktično vrednost. Če poznamo regresijsko krivuljo med količino padavin in hektarskim donosom, moremo iz količine padavin v določenem letu oceniti hektarski donos. Če imamo regresijsko krivuljo med oceno na oko in stvarnim hektarskim donosom, moremo z njo popraviti vizuelno oceno. Če poznamo regresijsko krivuljo o odvisnosti cene od količine blaga, ki je naprodaj, moremo pri znani količini blaga oceniti ceno ita.

Ta ocena je jasno tem boljša, čim bolj je y odvisen od x. Najboljša bi bila seveda, če bi bila odvisnost funkcijska. Takrat bi se iz

regresijske krivulje izračunana vrednost y' skladala s pravo vrednostjo y. Vendar, kakor vemo, takih primerov v stvarnosti ni. Za vsako posamezno oceno ne vemo, koliko smo pri ocenjevanju pogrešili, ali z drugimi besedami: kolika je razlika ocene y' od stvarne vrednosti y. Vendar moremo izračunati standardni odklon teh napak. Napaka pri ocenjevanju izvira iz individualnih vplivov, ki jih ne poznamo. Standardni odklon napake je torej enak standardnemu odklonu iz nepojasnjene variance. To pa moremo izračunati, če poznamo indeks korelacije $I_{y.x}$ med podatkoma y in x. Iz obrazca 13.14 namreč zlahka dobimo, da je

$$\sigma_e = \sigma_y \sqrt{1 - I^2_{y.x}} \qquad (13.15)$$

Standardni odklon nepojasnjene variance zato imenujemo tudi *standardno napako ocene*. Če predpostavljamo, da se odkloni ocen od pravih vrednosti distribuirajo normalno, moremo sklepati, da 38 % ocen ne bo napačnih za več kot σ_e, približno 95 % ne več kot $2\sigma_e$ in da bo le izjemoma ocena od prave vrednosti različna za več kot $3\sigma_e$. Tako moremo s standardno pogreško oceniti napako ocene, ki jo dobimo z regresijsko enačbo.

Koeficient $\sqrt{1 - I^2_{y.x}}$, s katerim reduciramo standardni odklon σ_y, da dobimo standardno napako ocene σ_e, imenujemo na splošno *koeficient alienacije* ali *koeficient nepovezanosti*.

LINEARNA KORELACIJA

13.17 Dosedanja razglabljanja o regresiji in korelaciji veljajo na splošno, neodvisno od tipa regresijske krivulje. V vsakem primeru izračunavamo regresijsko krivuljo, ki pokaže obliko odvisnosti med x in y, indeks korelacije, ki pokaže jakost povezanosti in standardno napako ocene, ki pokaže kvaliteto ocenjevanja z regresijsko krivuljo.

Vse te količine najenostavneje izračunamo, če je odvisnost med x in y linearna ali drugače povedano, če je regresijska krivulja premica. Linearne odvisnosti so v praksi zelo pogoste. Tudi če je povezava krivuljčna, je na krajših odsekih premica zelo dober približek regresijske krivulje. Zato je linearna regresija za prakso zelo važna in jo najpogosteje uporabljamo.

Če upoštevamo načela metode najmanjših kvadratov, izračunavamo pokazatelje linearne korelacije po naslednjih osnovnih obrazcih:

Ker imamo v splošnem dve regresijski krivulji, imamo tudi dve regresijski premici. Prva

$$y' = \bar{y} + b_1 (x - \bar{x}) \qquad (13.13)$$

podaja odvisnost y od x, druga

$$x' = \bar{x} + b_2(y-\bar{y}) \tag{13.17}$$

pa pokaže odvisnost x od y.

V zgornjih enačbah pomenita $\bar{x}$ in $\bar{y}$ aritmetični sredini za x in y, b_1 in b_2 pa sta *prvi* in *drugi regresijski koeficient*. Regresijska koeficienta sta smerna koeficienta obeh regresijskih premic.

Prvi regresijski koeficient b_1 pokaže, za koliko se v povprečju poveča y, če se x poveča za enoto; drugi regresijski koeficient b_2 pa pokaže, za koliko se v povprečju spremeni x, če se y spremeni za enoto. Regresijska koeficienta b_1 in b_2 izračunamo po obrazcih

$$b_1 = \frac{c_{xy}}{\sigma_x^2} \; ; \quad b_2 = \frac{c_{xy}}{\sigma_y^2} \tag{13.18}$$

Pri tem pomenita σ_x^2 in σ_y^2 varianci za znaka x in y. c_{xy} pa je nov izraz, ki ga imenujemo *kovarianco*. Kovarianca je po definiciji povprečen produkt odklonov x in y od ustreznih aritmetičnih sredin $\bar{x}$ in $\bar{y}$ in jo moremo izraziti z obrazcem

$$c_{xy} = \frac{1}{N} \Sigma (x-\bar{x}).(y-\bar{y}) \tag{13.19}$$

Za razliko od variance more biti vrednost kovariance pozitivna ali negativna.

Indeks korelacije pri linearni korelaciji imenujemo *koeficient korelacije*. Zaznamujemo pa ga z r_{xy}.

Koeficient korelacije izračunamo po osnovnem obrazcu

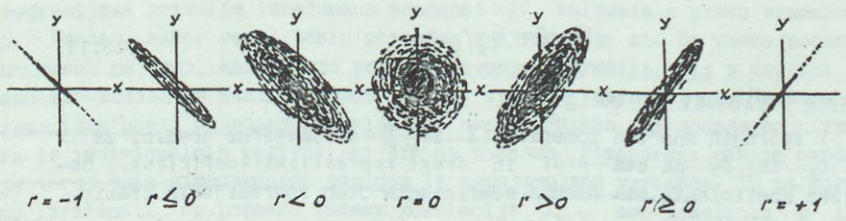
$$r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y} \tag{13.20}$$

Vrednosti korelacijskega koeficienta leže med -1 in $+1$. Korelacijski koeficient je pozitiven, če je povezava med x in y pozitivna; to pomeni, da se veča tudi vrednost druge spremenljivke, če se veča vrednost prve.

Korelacijski koeficient pa je negativen, če je povezava negativna. Taka pa je povezava, če se z večanjem enega znaka vrednost drugega manjša.

Koeficient korelacije ima vrednosti -1 oziroma $+1$, če je povezava med x in y funkcijska in linearna, 0 pa, kadar med x in y ni linearne zveze.

Slika 13.7 nazorno kaže velikost korelacijskega koeficienta za nekaj najznačilnejših oblik linearnih odvisnosti med x in y. Te povezave so ponazorjene s korelacijskim grafikonom.



Slika 13.7 Vrednosti korelacijskega koeficienta pri različnih razmestitvah točk v korelacijskem grafikonu

Kvadrat korelacijskega koeficienta r_{xy}^2 imenujemo *determinacijski koeficient*. Determinacijski koeficient sicer nima predznaka, ki bi nakazoval smer povezave, pač pa pove, kakšen del celotne variance je pojasnjen z linearno zvezo med x in y. Zato je analitično pomembnejši kakor korelacijski koeficient. Koeficient determinacije ima vrednosti med 0 in 1.

Izračunavanje pokazateljev linearne korelacije

13.18 Negrupirani podatki. Če je število enot v populaciji, za katero proučujemo korelacijo, razmeroma majhno, so osnovni podatki dani z nizom dvojic vrednosti za x in y. Če pogledamo obrazce v prejšnjem odstavku, vidimo, da so osnovnih količin, s katerimi izračunavamo pokazatelje linearne korelacije, razen kovariance C_{xy}, znamo izračunati vse količine.

13.19 Direktna metoda. Po direktni metodi izračunavamo osnovne količine: $\bar{x}, \bar{y}$, σ_x^2 σ_y^2 in C_{xy} natančno po obrazcih, ki so dani z definicijami.

Izkaže se, da je direktna metoda tehnično nepriklicna, ker so povprečja običajno decimalna števila, odkloni od povprečij zaradi tega večmestna števila, kvadriranje in množenje teh števil pa zamudno. Zato je ne bomo natančneje obravnavali.

13.20 Pomožna znaka u in v. Pri izračunavanju pokazateljev korelacije iz individualnih podatkov si največkrat pomagamo s *pomožnima znakoma* u *in* v. Z analognim postopkom smo si že olajšali izračunavanje variance. Po tej metodi izračunamo pokazatelje linearne korelacije po tehle točkah:

a) Imamo osnovne podatke x in y.

b) Izberemo neko poljubno vrednost x_0 med stvarnimi vrednostmi za x, in poljubno vrednost y_0 med stvarnimi vrednostmi za y.

c) Izračunamo odklone stvarnih vrednosti x od x_0 in stvarnosti y od y_0. Tako dobimo nove pomožne znake $u = (x-x_0)$ in $v = (y-y_0)$.

Vrednosti u in v so v večini primerov priklanejše manjše vrednosti kakor x in y.

d) Poiščemo kvadrate posameznih vrednosti u^2 in v^2 in izračunamo produkte $u.v$.

e) Seštejemo kolone za u, v, u^2, uv, v^2. Tako dobimo $\Sigma u = U$, $\Sigma v = V$, Σu^2, Σuv, Σv^2.

f) Iz količin N, U, V, Σu^2, Σuv in Σv^2 izračunamo pokazatelje korelacije po shemi v tabeli 13.9.

Tabela 13.9 Shema za računanje pokazateljev linearne korelacije

i	x	y	$x-x_0 = u$	$y-y_0 = v$	u^2	uv	v^2
1	x_1	x_1	u_1	v_1	u_1^2	u_1v_1	v_1^2
2	x_2	y_2	u_2	v_2	u_2^2	u_2v_2	v_2^2
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
N	x_N	y_N	u_N	v_N	u_N^2	u_Nv_N	v_N^2
			$\Sigma u = U$	$\Sigma v = V$	Σu^2	Σuv	Σv^2
			U/N	V/N	$-U^2/N$	$-UV/N$	$-V^2/N$
			$+ x_0$	$+ y_0$	$\Sigma u^2 - U^2/N = K_u$	$\Sigma uv - UV/N = K_{uv}$	$\Sigma v^2 - V^2/N = K_v$
			$U/N + x_0 = \bar{x}$	$V/N + y_0 = \bar{y}$	$K_u/N = \sigma_x^2$	$K_{uv}/N = c_{xy}$	$K_v/N = \sigma_y^2$
					$\sqrt{\sigma_x^2} = \sigma_x$		$\sqrt{\sigma_y^2} = \sigma_y$
					$b_1 = \frac{K_{uv}}{K_u}$	$r_{xy} = \frac{K_{uv}}{\sqrt{K_u \cdot K_v}}$	$b_2 = \frac{K_{uv}}{K_v}$
					$\sigma_{ex} = \sigma_x\sqrt{1-r_{xy}^2}$	$\sqrt{1-r_{xy}^2}$	$\sigma_{ey} = \sigma_y\sqrt{1-r_{xy}^2}$

$$y' = \bar{y} + b_1(x - \bar{x})$$

$$x' = \bar{x} + b_2(x - \bar{y})$$

13.21 Kot primer za izračunavanje pokazateljev korelacije po metodi pozitivnih znakov u in v vzemimo linearno korelacijo in regresijo med rezultati mehanskega (x), in prostorninskega (y) testa za 20 tretješolcev mariborske klasične gimnazije. Podatke smo vzeli iz kompleksnega testiranja zmožnosti srednješolske mladine v LRS in pomenijo število doseženih točk pri posameznem testiranju.

Tabela 13.19 Izračunavanje pokazateljev linearne korelacije med rezultati mehanskega (x) in prostorninskega (y) testa za 20 tretješolcev mariborske klasične gimnazije po metodi pomožnega znaka u in v

i	x	y	u	v	u^2	uv	v^2
1	46	58	+ 6	+ 8	36	+ 48	64
2	46	41	+ 6	− 9	36	− 54	81
3	38	21	− 2	−29	4	+ 58	841
4	35	18	− 5	−32	25	+160	1024
5	35	32	− 5	−18	25	+ 90	324
6	39	50	− 1	0	1	0	0
7	30	47	−10	− 3	100	+ 30	9
8	51	86	+11	+36	121	+396	1296
9	50	73	+10	+23	100	+230	529
10	30	53	−10	+ 3	100	− 30	9
11	41	71	+ 1	+21	1	+ 21	441
12	47	73	+ 7	+23	49	+161	529
13	33	23	− 7	−27	49	+189	729
14	41	49	+ 1	− 1	1	− 1	1
15	43	55	+ 3	+ 5	9	+ 15	25
16	35	37	− 5	−13	25	+ 65	169
17	54	85	+14	+35	196	+490	1225
18	47	76	+ 7	+26	49	+182	676
19	35	41	− 5	− 9	25	+ 45	81
20	44	62	+ 4	+12	16	+ 48	144
			$U = +20$	$V = +51$	$\Sigma u^2 = 968$	$\Sigma uv = 2143$	$\Sigma v^2 = 8197$
			$U/N = +1$	$V/N = +2,55$	$-U^2/N = -20$	$-UV/N = -51$	$V^2/N = -130,05$
			$+x_0 = +40$	$+y_0 = +50$	$K_u = 948$	$K_{uv} = 2092$	$K_v = 8033,95$
			$\overline{x} = 41$	$\overline{y} = 52,55$	$\sigma_x^2 = 47,4$	$c_{xy} = 104,6$	$\sigma_y^2 = 403,8$
					$\sigma_x = 6,88$		$\sigma_y = 20,08$
					$b_1 = +2,21$	$r_{xy} = 0,757$	$b_2 = +0,259$
					$\sigma_{ex} = 4,49$	$\sqrt{1-r_{xy}^2} = 0,653$	$\sigma_{ey} = 13,08$

$$y' = 52,55 + 2,21(x - 41) = -38,06 + 2,21x$$

$$x' = 41 + 0,259(y - 52,55) = 27,39 + 0,25y$$

Iz absolutnih podatkov za x in y smo presodili, da je primerno, da vzamemo za $x_0 = 40$ in za $y_0 = 50$. Znaki u in v so pokazali, da so te vrednosti upravičene. Vrednosti u in v so znatno manjše kakor osnovne vrednosti x in y. Kakor vidimo iz tabele 13.10, poteka račun natančno po shemi iz tabele 13.9. Za obe regresijski premici so v tabeli 13.11 dane zveze med x in y po pet točk

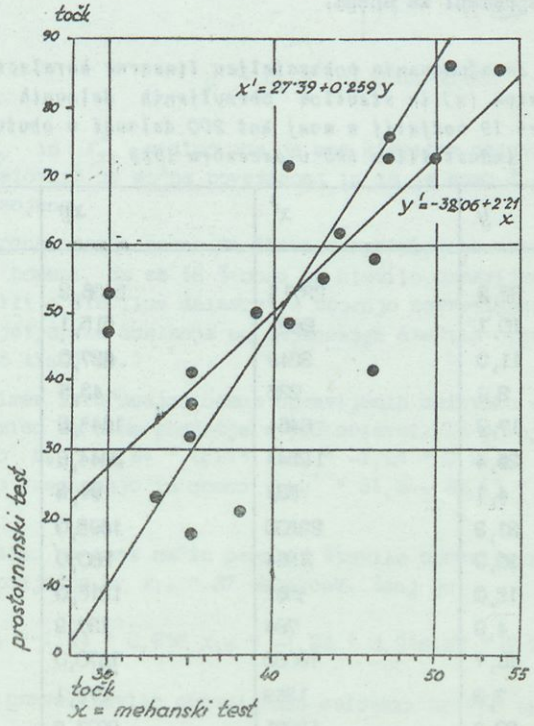
Tabela 13.11 Regresijski premici med številom točk pri mehanskem (x) in prostorninskem (y) testu iz primera v tabeli 13.10

Prva regresijska premica:

x	30	35	40	45	50	55
y'	28	39	50	61	72	83

Druga regresijska premica:

y	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
x'	33	34	35	36	38	39	40	42	43	44	46	47	48	49	51



Slika 13.8 Regresijski premici za korelacijo med dosežki pri mehanskem in prostorninskem testu za 20 dijakov klasične gimnazije v Mariboru

Korelacijski grafikon z vrtanima regresijskima premicama je za primer iz tabele 13.10 prikazan v sliki 13.8.

13.22 Če podatki zelo variirajo in ne moremo najti nekih ustreznih vrednosti za x_0 in y_0, da bi poenostavili osnovne vrednosti, vzamemo, da je $x = 0$ in $y_0 = 0$. Če to storimo, odpade iz zgornjega postopka točka C, ker ni potrebno izračunavati u in v. Ves nadaljnji postopek pa je isti, le da upoštevamo, da je x identičen z u, y pa z v.

Tako izračunavanje pokazateljev linearne korelacije je posebno priklado, če računamo z računskim strojem. Tedaj so pomožni znaki u in v nepotrebni in postopek samo zapletejo.

Po tej metodi izračunajmo pokazatelje linearne korelacije za primer korelacije med številom delavstva in številom opravljenih delovnih ur za 19 podjetij z manj kot 200 delavcev iz obutvene industrije v LRS v decembru 1952. Izračun je izvršen v tabeli 13.12.

Ker je med številom delavstva x in številom opravljenih delovnih ur y vzročna zveza, izračunamo samo prvo regresijsko premico.

Ker je število ur samo v tisočih, regresijski koeficient 0,213 pove, da se v povprečju spremeni skupno število opravljenih ur za 213 ur, če se število delavcev spremeni za enega.

Tabela 13.12 Izračunavanje pokazateljev linearne korelacije med številom delavstva (x) in številom opravljenih delovnih ur (y) v tisočih ur za 19 podjetij z manj kot 200 delavci v obutveni industriji v LRS v decembru 1952

Zap. št.	x	y	x^2	xy	y^2
1	154	35,3	23716	5436,2	1246,09
2	51	10,1	2601	515,1	102,01
3	57	11,0	3249	627,0	121,00
4	15	2,9	225	43,5	8,41
5	92	17,9	8464	1643,8	320,41
6	112	25,4	12544	2844,8	645,16
7	23	4,1	529	94,3	16,81
8	150	31,3	22500	4695,0	979,69
9	58	10,0	3364	580,0	100,00
10	81	15,0	3561	1215,0	225,00
11	28	4,9	784	137,2	24,01
12	100	16,7	10000	1670,0	278,89
13	37	7,3	1369	270,1	53,29
14	111	20,7	12321	2297,7	428,49

Zap. št.	x	y	x^2	xy	y^2
15	12	2,2	144	26,4	4,84
16	50	9,5	2500	475,0	90,25
17	7	0,8	49	5,6	0,64
18	80	15,1	6400	1208,0	228,01
19	49	9,9	2401	485,1	98,01
	$X = 1267$	$Y=250,1$	$\Sigma x^2 = 119721$	$\Sigma xy = 24272,8$	$\Sigma y^2 = 4970,92$
	$\bar{x} = 66,68$	$\bar{y} = 13,16$	$-X^2/N = -84488,89$	$-XY/N = 16677,72$	$-Y^2/N = 3292,11$
			$K_x = 35232,11$	$K_{xy} = 7595,08$	$K_y = 1678,81$
			$\sigma_x^2 = 1854,32$	$C_{xy} = 399,74$	$\sigma_y^2 = 88,33$
			-		$\sigma_y = 9,40$
			$b_1 = 0,216$	$r = +0,988$	-
			-	$\sqrt{1-r^2} = 0,154$	$\sigma_{ey} = 1,45$

$$y' = 13,16 + 0,213(x-66,68) = -1,24 + 0,213x$$

Iz r_{xy} in r_{xy}^2 vidimo, da je med številom delavcev in številom opravljenih delovnih ur močna povezanost in da je samo 2,89 % celotne variance nepojasnjene.

Standardna napaka ocene za število opravljenih delovnih ur je $\sigma_{ey} = 1,45$. To pomeni, da se 38 % ocen za število opravljenih delovnih ur, ki bi ga ocenili s številom delavcev z zgornjo regresijsko premico za katero koli podjetje, ne odklanja od dejanskega števila opravljenih ur za več kakor 1,45 tisoč ur.

Kot primer izračunajmo oceno opravljenih delovnih ur z dobljeno regresijsko premico za osmo podjetje s 150 delavci. Iz gornje regresijske premice dobimo $y_8' = -1,24 + 0,213 \cdot x_8 = -1,24 + 0,213 \cdot 150 = 31,13$. Razlika med pravo vrednostjo in oceno $y-y' = 31,3 - 31,13 = +0,14$ je zelo majhna.

Poizkusimo na enak način oceniti število opravljenih delovnih ur za trinajsto podjetje z $x_{13} = 37$ delavcev. Zanj je

$$y_{13}' = -1,24 + 0,216\, x_{13} = -1,24 + 0,216 \cdot 37 = 6,75$$

Ker je pravo število opravljenih delovnih ur 7,3 tisoč, je napaka ocene

$$\Delta y_{13} = y_{13} - y_{13}' = 7,3 - 6,75 = +0,55 \text{ ali } 550 \text{ ur.}$$

13.23 Izračunavanje pokazateljev linearne korelacije iz grupiranih podatkov. Že uvodoma smo navedli, da prikazujemo za velike populacije korelirane podatke v korelacijski tabeli. Ker je korelacijska tabela slika zaokroženih vrednosti za dvojice vrednosti znakov, moremo iz nje izračunati le ocene pokazateljev korelacije in regresije.

Osnovne količine, iz katerih izračunavamo pokazatelje linearne korelacije in regresije, so: $\bar{x}$, $\bar{y}$, σ_x^2 σ_y^2 in c_{xy}. Vse te količine, razen kovariance, moremo izračunati po znanih postopkih iz obeh robnih frekvenčnih distribucij v korelacijski tabeli. Novo je edino izračunavanje kovariance. Vendar moremo tudi za njo iz njene definicije sestaviti postopek, kako jo izračunamo. Ta je sicer bolj zamotan, vendar izvedljiv. Postopkov za izračunavanje pokazateljev za linearno korelacijo iz grupiranih podatkov je več. Od teh bomo obravnavali le postopek pomožnih znakov u in v. Ta je po osnovi enak metodi pomožnega znaka u za izračunavanje aritmetične sredine ali variance iz frekvenčnih distribucij. Pogoj za uporabo te metode je, da so razredi po *posameznem* znaku enako široki.

Pokazatelje linearne korelacije in regresije po metodi pomožnih znakov u in v izračunamo po naslednjih točkah:

a) Dano imamo kombinacijsko tabelo za korelirana znaka z vpisanima robnima frekvenčnima distribucijama.

b) Za znak x uvedemo pomožen znak u, tako da v razred, ki leži nekje sredi vrste, vnesemo izhodiščno 0, za druge razrede pa na levo in desno vrednosti ... −3, −2, −1, 0, +1, +2, +3, ..., enako kot to storimo, ko izračunavamo varianco.

Analogno uvedemo za y pomožen znak v. Vrednosti znakov u in v vnesemo v kombinacijsko tabelo v ustrezne razrede.

c) Kakor pri izračunavanju variance robno frekvenčno distribucijo f_v pomnožimo z ustreznimi vrednostmi znaka v, da dobimo vf_v, dobljene produkte pa ponovno pomnožimo z ustreznimi vrednostmi v, da dobimo $v^2 f_v$. Oba dobljena stolpca seštejemo. Enako napravimo z drugo robno frekvenčno distribucijo. Tako dobimo naslednje količine:

$$\sum_u uf_u = U; \quad \sum_u u^2 f_u = \Sigma u^2; \quad \sum_v vf_v = V; \quad \sum_v v^2 f_v = \Sigma v^2$$

d) Manjka še izraz $\sum_{uv} uvf_{uv} = \Sigma uv$. Tega pa dobimo takole. Za vsako vrsto v korelacijski tabeli izračunamo količino U_v tako, da v dani vrsti vrstne frekvence f_{uv} pomnožimo z ustreznimi vrednostmi znaka u in te produkte seštejemo $U_v = \sum_u uf_{uv}$. Te vsote produktov vpišemo v stolpec U_v. Kontrola: Vsota vseh dobljenih U_v je enaka U; $\sum_v U_v = U$.

Nato podatke v stolpcu U_v pomnožimo z ustreznimi vrednostmi znaka v in produkte seštejemo. Tako dobimo količino $\sum_v vU_v = \sum_{uv} uvf_{uv} = \Sigma uv$. Podobno dobimo Σuv, če postopek izvedemo s frekvencami v stolpcih.

e) Tako imamo količine N, U, V, Σu^2, Σuv, Σv^2, iz katerih izračunamo pokazatelje linearne korelacije po shemi v tabeli 13.13, ki je podobna shemi v tabeli 13.9, le da moramo pri grupiranih podatkih upoštevati širini razredov i_x in i_y.

Tabela 13.13 Shema za izračunavanje pokazateljev za linearno korelacijo in regresijo, če so podatki grupirani v korelacijski tabeli

i_x N U	i_y V	Σu^2	Σuv	Σv^2
i_xU/N	i_yU/N	$-U^2/N$	$-UV/N$	$-V^2/N$
$+x_0$	$+y_0$	$\Sigma u^2-U^2/N = K_u$	$\Sigma uv-UV/N = K_{uv}$	$\Sigma v^2-V^2/N = K_v$
$i_xU/N+x_0 = \bar{x}$	$i_yV/N+y_0 = \bar{y}$	$i_x^2K_u/N = \sigma_x^2$	$i_xi_yK_{uv}/N = c_{xy}$	$i_y^2K_v/N = \sigma_y^2$
		$\sqrt{\sigma_x^2} = \sigma_x$		$\sqrt{\sigma_y^2} = \sigma_y$
		$b_1 = \frac{c_{xy}}{\sigma_x^2}$	$r_{xy} = \frac{K_{uv}}{\sqrt{K_uK_v}}$	$b_2 = \frac{c_{xy}}{\sigma_y^2}$
		$\sigma_{ex} = \sigma_x\sqrt{1-r_{xy}^2}$	$\sqrt{1-r_{xy}^2}$	$\sigma_{ey} = \sigma_y\sqrt{1-r_{xy}^2}$

$$y' = \bar{y} + b_1(x - \bar{x})$$

$$x' = \bar{x} + b_2(y - \bar{y})$$

13.24 Kot primer za izračunavanje pokazateljev linearne korelacije in regresije vzemimo odvisnost med starostjo žen in mož ob razvezi zakona v LRS v letu 1955 iz tabele 13.5. V tabeli 13.14 je sistematično nakazan postopek izračunavanja pomožnih količin U, V, Σu^2, Σv^2 in Σuv, v tabeli 13.15 pa izračunavanje pokazateljev linearne korelacije iz teh količin po shemi iz tabele 13.13.

Količine U_v smo izračunali posebej, in sicer je:

$$U_1 = (-4).2 + (-3).2 + (-2).2 = -18$$

$$U_2 = (-4).48 + (-3).82 + (-2).28 + (-1).8 + (0).1 + (+3).1 = -499$$

Tabela 13.14 Izračunavanje pokazateljev linearne korelacije in regresije za odvisnost med starostjo žene in moža ob razvezi zakona v LRS v letu 1955

y starost žene	x starost moža	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74					
	v \ u	−4	−3	−2	−1	0	+1	+2	+3	+4	+5	+6	f_v	vf_v	$v^2 f_v$	$\sum_u uf_{uv}$	vU_v
15–19	−5	2	2	2	–	–	–	–	–	–	–	–	6	−30	150	−18	+90
20–24	−4	48	82	28	8	4	–	–	1	–	–	–	171	−384	2736	−432	+1296
25–29	−3	15	108	91	32	10	5	–	1	–	–	–	262	−786	2358	−530	+1770
30–34	−2	5	36	77	67	37	17	7	3	1	–	–	250	−500	1000	−305	+610
35–39	−1	1	4	17	38	42	19	3	3	1	–	–	123	−123	123	−45	+45
40–44	0	1	2	9	15	52	35	9	5	1	1	–	130	0	0	+34	0
45–49	+1	–	–	3	7	21	31	17	8	4	–	–	91	+91	91	+92	+92
50–54	+2	–	–	1	2	6	8	13	6	–	2	2	40	+80	160	+70	+140
55–59	+3	–	–	–	–	2	8	4	9	8	3	2	36	+108	324	+102	+306
60–64	+4	–	–	–	–	1	–	2	3	5	4	2	17	+68	272	+65	+260
65–69	+5	–	–	–	–	–	1	1	1	3	3	1	9	+45	225	+34	+170
70–74	+6	–	–	–	–	–	–	1	–	2	–	1	4	+24	144	+16	+96
	f_u	7	234	228	164	175	124	57	40	25	12	8	1139	−1707	7583	−1044	+5575
	uf_u	−288	−702	−456	−164	0	+124	+114	+120	+100	+60	+48	−1044 = U	V	$\sum v^2$	U	$\sum uv$
	$u^2 f_u$	1152	2106	912	164	0	124	228	360	400	300	288	6024 = $\sum u^2$				
	$\sum vf_{uv} = V_u$	−258	−738	−561	−284	−119	+8	+57	+48	+72	+39	+29	−1707 = V				
	uVu	1032	2214	1122	284	0	8	114	144	288	195	174	5575 = $\sum uv$				

itd. do zadnje vrste, za katero imamo:

$$U_{12} = (+2).1 + (+4).2 + (+6).1 = +13$$

Kontrola izračunavanja izrazov U_v, ki so najbolj problematični je: $\sum_v U_v = -1044 = U$

Nadaljnji račun po shemi 13.13 je nakazan v tabeli 13.15.

Tabela 13.15 Izračun pokazateljev linearne korelacije za primer v tabeli 13.14 po shemi iz tabele 13.13.

$i_x = 5$; $i_y = 5$; $x_0 = 42,5$; $y_0 = 42,5$;

$N=1139$ $U = -1044$ $V = -1707$	$\sum u^2 = 6034$	$\sum uv = +5575$	$\sum v^2 = 7583$
$i_x U/N = -4,58$ $i_y V/N = -7,49$	$U^2/N = -956,92$	$UV/N = -1564,63$	$V^2/N = -2558,25$
$+x_0 = 42,50$ $+y_0 = 42,50$	$K_u = 5077,08$	$K_{uv} = +4010,37$	$K_v = 5024,75$
$\bar{x} = 37,92$ $\bar{y} = 35,01$	$\sigma_x^2 = 111,43$	$c_{xy} = +88,02$	$\sigma_y^2 = 110,29$
	$\sigma_x = 10,53$		$\sigma_y = 10,50$
	$b_1 = +0,790$	$r_{xy} = +0,794$	$b_2 = +0,798$
	$\sigma_{ex} = 6,42$	$\sqrt{1-r_{xy}^2} = 0,608$	$\sigma_{ey} = 6,38$

$$y' = 35,01 + 0,790(x-37,92) = 5,05 + 0,790x$$

$$x' = 37,92 + 0,798(y-35,01) = 9,98 + 0,798y$$

KRIVULJČNA KORELACIJA

13.25 Čeprav je povezanost med pojavi pogosto linearna, imamo mnogo primerov nelinearne - krivuljčne povezave. Te primere rešujemo tako, kakor smo nakazali v splošnem razglabljanju o korelacijskih odvisnostih.

Najtežji problem pri krivuljčni korelaciji je določitev tipa krivulje povezave. Tip krivulje običajno določimo tako, da narišemo korelacijski grafikon, prostoročno včrtamo vanj regresijsko krivuljo, nato pa preocenimo, kateri tip krivulje najbolj ustreza tej črti. Ta način zahteva, da praktično poznamo potek posameznih tipov krivulj. Včasih je tip regresijske krivulje nakazan vnaprej z določeno hipotezo, ali pa je dan z znano zakonitostjo odvisnosti. Dostikrat pa ugotovimo tip regresijske krivulje

s transformacijo podatkov. S pravilno transformacijo osnovnih podatkov včasih dosežemo, da je zveza med transformiranimi podatki linearna: to olajša nadaljnjo analizo odvisnosti.

Analitična metoda za določanje regresijskih krivulj

13.26 Najobjektivneje določimo regresijsko krivuljo in analiziramo korelacijo med dvema pojavoma analitično, z metodo najmanjših kvadratov.

Po tej metodi najprej z analizo korelacijskega grafikona ali iz vsebine koreliranih pojavov določimo ustrezno funkcijsko obliko $y' = f(x)$ za regresijsko krivuljo. Najpogosteje je regresijska krivulja ena izmed tehle funkcij:

1. Premica $y' = a + bx$
2. Parabola druge stopnje $y' = a + bx + cx^2$
3. Parabola tretje stopnje $y' = a + bx + cx^2 + dx^3$
4. Parabola $y' = a + b\sqrt{x}$
5. Parabola $y' = (a + bx)^2$ (13.21)
6. Parabola $y' = ax^b$
7. Eksponencialna funkcija $y' = ab^x$
8. Hiperbola $y' = a + \frac{b}{x}$
9. Hiperbola $y' = \frac{1}{a + bx}$

Vse zgornje funkcije so splošne oblike, ker vsebujejo parametre $a, b, c, d \ldots$ Zato ni dovolj, da z analizo podatkov v posebnem primeru ugotovimo najprimernejši tip funkcije, ki naj predstavlja regresijsko krivuljo, temveč je treba najti danim podatkom ustrezne vrednosti parametrov $a, b, c, d \ldots$

Vrednost parametrov določimo po metodi najmanjših kvadratov z naslednjo predpostavko: Za regresijsko krivuljo imamo izmed vseh funkcij določenega tipa tisto, za katero je vsota kvadratov odklonov dejanskih vrednosti za y od regresijske krivulje $y' = f(x)$ najmanjša. To načelo zagotavlja, da je regresijska krivulja res krivulja, ki se danim vrednostim najbolje prilega. Z obrazcem moremo zgornji pogoj napisati

$$\Sigma (y - y')^2 = \text{Min} \qquad (13.22)$$

Parametre za regresijsko krivuljo določimo razmeroma enostavno, če je regresijska krivulja funkcija, v kateri so parametri $a, b, c \ldots$ v line-

arni zvezi. Temu pogoju izmed gornjih funkcij neposredno zadoščajo tipi funkcij 1, 2, 3, 4 in 8. Druge pa moremo v to obliko privesti z določenimi transformacijami.

Po pravilih za določanje ekstremov dobimo, da morajo parametri a, b, c ..., če so med seboj v linearni zvezi, zadostiti sistemu linearnih enačb, ki jih imenujemo *normalne enačbe*.

Če je splošna oblika regresijske funkcije s tremi parametri

$$g(y') = af_1(x) + bf_2(x) + cf_3(x), \tag{13.23}$$

je sistem normalnih enačb enak

$$\begin{aligned} \Sigma g f_1 &= a\Sigma f_1^2 + b\Sigma f_2 f_1 + c\Sigma f_3 f_1 \\ \Sigma g f_2 &= a\Sigma f_1 f_2 + b\Sigma f_2^2 + c\Sigma f_3 f_2 \\ \Sigma g f_3 &= a\Sigma f_1 f_3 + b\Sigma f_2 f_3 + c\Sigma f_3^2 \end{aligned} \tag{13.24}$$

Pri tem je: $g = g(y_i)$ funkcijske vrednosti g za stvarne vrednosti y_i za posamezne enote: $f_1 = f_1(x_i)$, $f_2 = f_2(x_i)$, $f_3 = f_3(x_i)$ pa so funkcijske vrednosti funkcij f_1, f_2 in f_3 za stvarne vrednosti x_i za posamezne enote populacije. Znak Σ pomeni seštevanje ustreznih izrazov za vse enote populacije.

Transformacija znakov

13.27 Dostikrat si pomagamo pri proučevanju krivuljčne korelacije s transformacijo znakov x in y. Že v prejšnjem odstavku smo navedli, da moremo nekatere razmeroma zamotane funkcije s transformiranjem privesti v enostavnejšo obliko.

Za nekaj najobičajnejših tipov krivulj so transformacije in transformirani znaki dani v tabeli 13.16.

Tabela 13.16 Transformacije regresijskih funkcij

Osnovna funkcija		Transformacija		Transformirana funkcija
eksplicitna	implicitna	znakov	parametrov	
$y' = (a+bx)^2$	$\sqrt{y'} = a + bx$	$x=X;\ \sqrt{y}=Y$	$a=A;\ b=B$	$Y' = A + BX$
$y' = a + b\sqrt{x}$	$y' = a + b\sqrt{x}$	$\sqrt{x}=X;\ y=Y$	$a=A;\ b=B$	$Y' = A + BX$
$y' = ax^b$	$\log y' = \log a + b\log x$	$\log x=X;\ \log y=Y$	$\log a=A;\ b=B$	$Y' = A + BX$
$y' = ab^x$	$\log y' = \log a + x\log b$	$x=X;\ \log y=Y$	$\log a=A;\ \log b=B$	$Y' = A + BX$
$y' = a + \frac{b}{x}$	$y' = a + \frac{b}{x}$	$\frac{1}{x}=X;\ y=Y$	$a=A;\ b=B$	$Y' = A + BX$
$y' = \frac{1}{a+bx}$	$\frac{1}{y'} = a + bx$	$x=X;\ \frac{1}{y}=Y$	$a=A;\ b=B$	$Y' = A + BX$

Iz teh primerov vidimo, da so najobičajnejše transformacije znakov: kvadratni koren iz znaka x ali y $(\sqrt{x}; \sqrt{y})$; recipročna vrednost znaka x ali y $(1/x; 1/y)$ in logaritem iz znaka x ali y $(\log x; \log y)$.

V vseh primerih smo v tabeli 13.16 uspeli s transformacijo razmeroma kompliciranо funkcijo prevesti v linearno obliko transformiranih znakov X in Y: $Y' = A + BX$.

Jasno je, da s prejšnjimi šestimi primeri ne izčrpamo vse funkcije, ki jih moremo analogno s transformacijo prevesti v linearno funkcijo. Že s prejšnjimi transformacijami (kvadratni koren, recipрok in logaritem) moremo sestaviti še druge funkcije, na primer $\frac{1}{y'} = a + \frac{b}{x}$, ki je implicitna oblika funkcije: $y' = \frac{x}{ax+b}$ itd.

13.28 Ker so transformirane spremenljivke iz tabele 13.16 v linearni zvezi, je proučevanje korelacije in regresije za take vrste krivuljčne povezave podobno kot proučevanje linearne korelacije.

Postopek moremo nakazati z naslednjimi tremi stopnjami:

a) S transformacijo osnovnih podatkov ugotovimo tip regresijske krivulje. Pri preizkušanju, za kateri tip funkcije gre, si običajno pomagamo tako, da osnovne vrednosti vnašamo v korelacijski grafikon, ki ima namesto linearnih skal transformirane skale (skale kvadratnih korenov, skale recipрokov ali logaritemske skale). Osnovne znake transformiramo z ustrezno transformacijo šele tedaj, ko uspemo, da korelacijski grafikon na transformiranih skalah pokaže linearno povezavo.

b) Iz transformiranih podatkov X in Y po znanih postopkih izračunamo pokazatelje linearne korelacije in regresijsko premico.

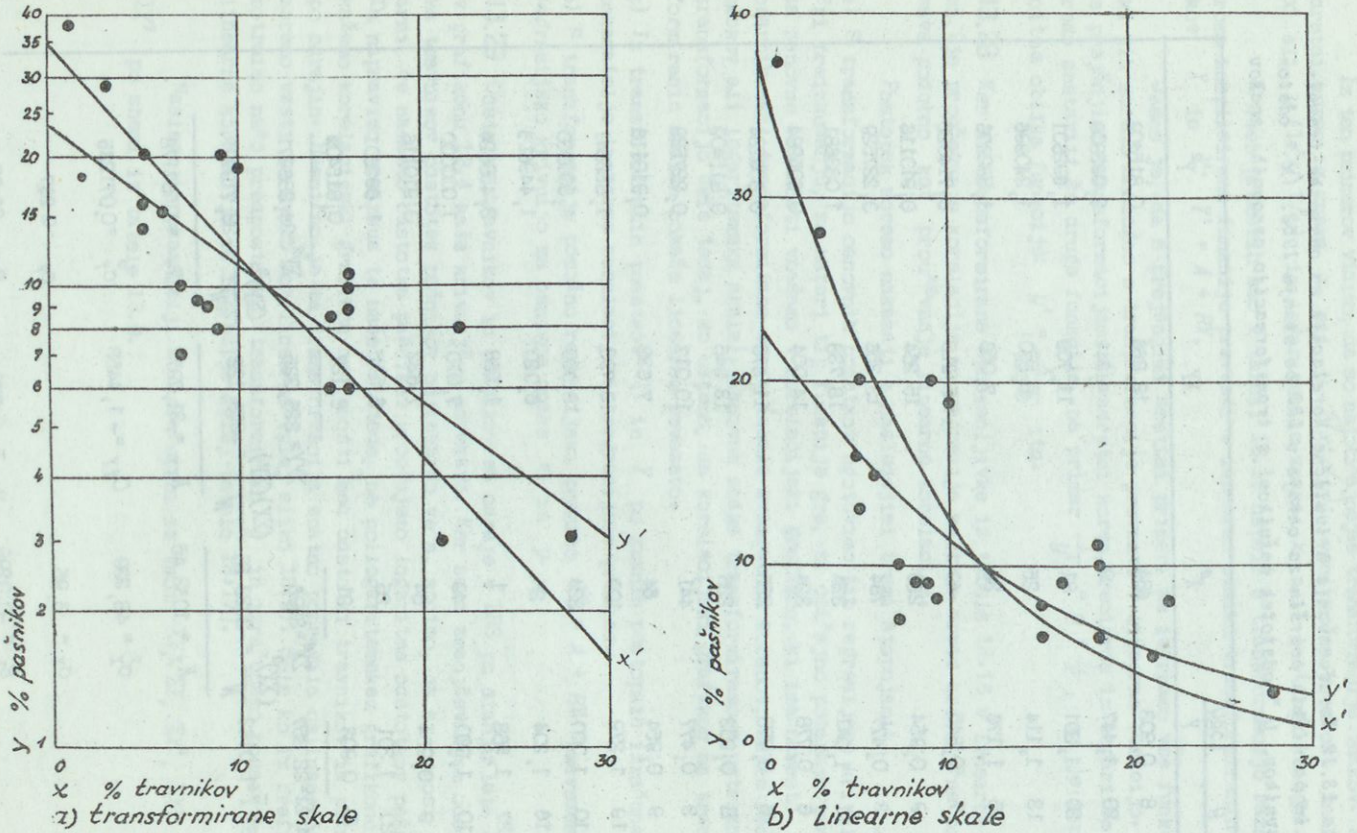
c) S transformacijo pomožno regresijsko premico $Y' = A + BX$ prevedemo v regresijsko krivuljo za osnovne znake x in y.

13.29 Odstotek travnikov in pašnikov za okraje v LRS po stanju leta 1952 v grafikonu 13.4 kaže krivuljčno povezavo. Ker se z zmanjševanjem odstotka travnikov odstotek pašnikov nelinearno veča, sodimo, da so odnosi linearni, če namesto odstotka pašnikov proučujemo logaritme odstotkov pašnikov. Da napravimo preizkus te naše hipoteze, na poliogaritemskem grafikonu prikažemo korelacijski grafikon odvisnosti med odstotki travnikov in pašnikov po okrajih. Resnično je ta transformacija znatno popravila nelinearnost in moremo vzeti to zvezo kot linearno (glej sliko 13.9). Šele ko iz grafikona potrdimo našo predpostavko, transformiramo y in izračunamo pokazatelje linearne korelacije in regresije (glej tabelo 13.17).

Naslednji pokazatelji so izračunani iz vsot ΣX, ΣY, ΣX^2, ΣXY in ΣY^2 po shemi iz tabele 13.9.

Tabela 13.17 Izračunavanje krivuljčne korelacije za odstotke travniških in pasniških površin za okraje v LRS po stanju 1952. (x = odstotek travnikov, y = odstotek pašnikov) s transformacijo osnovnih znakov

i	x / X	y	log y / Y	X^2	XY	Y^2
1	22	8	0,903	484	19,866	0,815409
2	3	28	1,447	9	4,341	2,093809
3	9	20	1,301	81	11,709	1,692601
4	5	13	1,114	25	5,570	1,240996
5	6	15	1,176	36	7,056	1,382976
6	7	7	0,845	49	5,915	0,714025
7	13	9	0,954	256	15,264	0,910116
8	28	3	0,477	784	13,356	0,227529
9	18	11	1,041	324	18,738	1,083681
10	18	6	0,778	324	14,004	0,605284
11	15	6	0,778	225	11,670	0,605284
12	15	8	0,903	225	13,545	0,815409
13	21	3	0,477	441	10,017	0,227529
14	8	9	0,954	64	7,632	0,910116
15	10	19	1,279	100	12,790	1,635841
16	18	10	1,000	324	18,000	1,000000
17	5	16	1,204	25	6,020	1,449613
18	1	37	1,568	1	1,568	2,458624
19	7	10	1,000	49	7,000	1,000000
20	8	9	0,954	64	7,632	0,910116
21	5	20	1,301	25	6,505	1,692601
22	9	8	0,903	81	8,127	0,815409
22	$\Sigma X=254$	$\Sigma Y=22,357$		$\Sigma X^2=3966$	$\Sigma XY=228,325$	$\Sigma Y^2=24,286971$
	$\bar{X}=11,55$	$\bar{Y}=1,0162$		$\frac{(\Sigma X)^2}{N}=2932,55$	$\frac{(\Sigma X)(\Sigma Y)}{N}=258,122$	$\frac{(\Sigma Y)^2}{N}=22,719783$
				$K_Y=1033,45$	$K_{XY}=-31,797$	$K_Y=1,567178$
				$\sigma_X^2=48,339$	$C_{XY}=-1,4453$	$\sigma_Y^2=0,071235$
				$\sigma_X=6,95$		$\sigma_Y=0,267$
				$B_1=-0,0299$	$r_{XY}=-0,778$	$B_2=-20,29$



Slika 13.9 Korelacijska grafikona med odstotkom travnikov in odstotko. pašnikov za okraje v LRS v letu 1952 z včrtanima regresijskima krivuljama na: a) transformiranih skalah, b) osnovnih skalah

Regresijske krivulje pa so iz teh pokazateljev dalje:

$$Y' = \overline{Y} + B_1(X-\overline{X}) = 1,0132 - 0,0299(X-11,55) = 1,3315 - 0,0299\,X = A + BX$$

$$\log y' = 1,3315 - 0,0299x$$

$$A = \log a = 1,3315 ; \quad a = 22,99 \qquad B = \log b = -0,0299 : \quad b = 0,9335$$

Iz tega dalje sledi, da je y' v eksplicitni obliki:

$$y' = 22,99 \cdot 0,9335^x$$

Druga regresijska enačba pa ima obliko:

$$X' = \overline{X} + B_2(Y-\overline{Y}) = 11,55 - 20,29(Y-1,0132) = 32,17 - 20,29Y$$

Če upoštevamo transformacijo $Y = \log y$, se druga regresijska enačba v osnovni obliki glasi:

$$x' = 32,17 - 20,29 \log y$$

V sliki 13.9 sta prikazani v korelacijskem grafikonu obe regresijski krivulji, na sliki a na transformiranih skalah, na sliki b pa na osnovnih. Na sliki a sta zaradi transformacije skal regresijski krivulji premici.

13.30 Korelacijski grafikon pokaže, da so odnosi odstotkov travnikov in pašnikov linearno povezani tudi, če znak x in y transformiramo s kvadratnim korenom iz osnovnih podatkov. To lepo vidimo na sliki 13.10 a, na kateri so podatki iz tabele 13.4 vrisani v korelacijski grafikon, v katerem sta obe skali skali za kvadratni koren. Regresijski krivulji sta v tem primeru funkciji tipa

$$\sqrt{y}' = a_1 + b_1\sqrt{x} \quad \text{in} \quad \sqrt{x}' = a_2 + b_2\sqrt{y}$$

Postopek je popolnoma analogen postopku v prejšnjem primeru in je nakazan v tabeli 13.18.

Regresijska premica za transformirane spremenljivke je

$$Y' = \overline{Y} + B_1(X - \overline{X}) = 3,877 - 0,793(X-3,229) = 5,938 - 0,793X$$

Regresijska krivulja v osnovnih znakih x in y pa je

$$\sqrt{y}' = 5,938 - 0,793\sqrt{x}$$

ali v eksplicitni obliki

$$y' = (5,938 - 0,793\sqrt{x})^2$$

Tabela 13.18 Izračunavanje krivuljične korelacije in regresije med odstotkom travnikov (x) in pašnikov (y) za okraje v LRS po stanju leta 1952 s transformacijo znakov

i	x	y	$\sqrt{x}$ = X	$\sqrt{y}$ = Y	X^2	XY	Y^2
1	22	8	4,69	2,83	22	13,2727	8
2	3	28	1,73	5,29	3	9,1517	28
3	9	20	3,00	4,47	9	13,4100	20
.....					.....	.....	.....
11	15	6	3,87	2,45	15	9,4815	6
.....					.....	.....	.....
21	5	20	2,24	4,47	5	10,0126	20
22	9	8	3,00	2,83	9	8,4900	8
			$\Sigma X = 71,04$	$\Sigma Y = 74,29$	$\Sigma X^2 = 254$	$\Sigma XY = 280,4711$	$\Sigma Y^2 = 275$
			$\bar{X} = 3,229$	$\bar{Y} = 3,377$	$-\frac{(\Sigma X)^2}{N} = -229,39$	$-\frac{(\Sigma X)(\Sigma Y)}{N} = -240,0000$	$-\frac{(\Sigma Y)^2}{N} = -250,86$
					$K_X = 24,61$	$K_{XY} = 19,53$	$K_Y = 24,14$
					$\sigma_x^2 = 1,1186$	$c_{xy} = -0,8877$	$\sigma_y^2 = 1,0973$
					$\sigma_x = 1,058$		$\sigma_y = 1,047$
					$B_1 = -0,793$	$r_{xy} = -0,801$	$B_2 = -0,809$
					$\sigma_{ex} = 0,634$	$\sqrt{1-r^2} = 0,599$	$\sigma_{ey} = 0,628$

Standarana napaka ocene $\sigma_{ey} = 0,628$

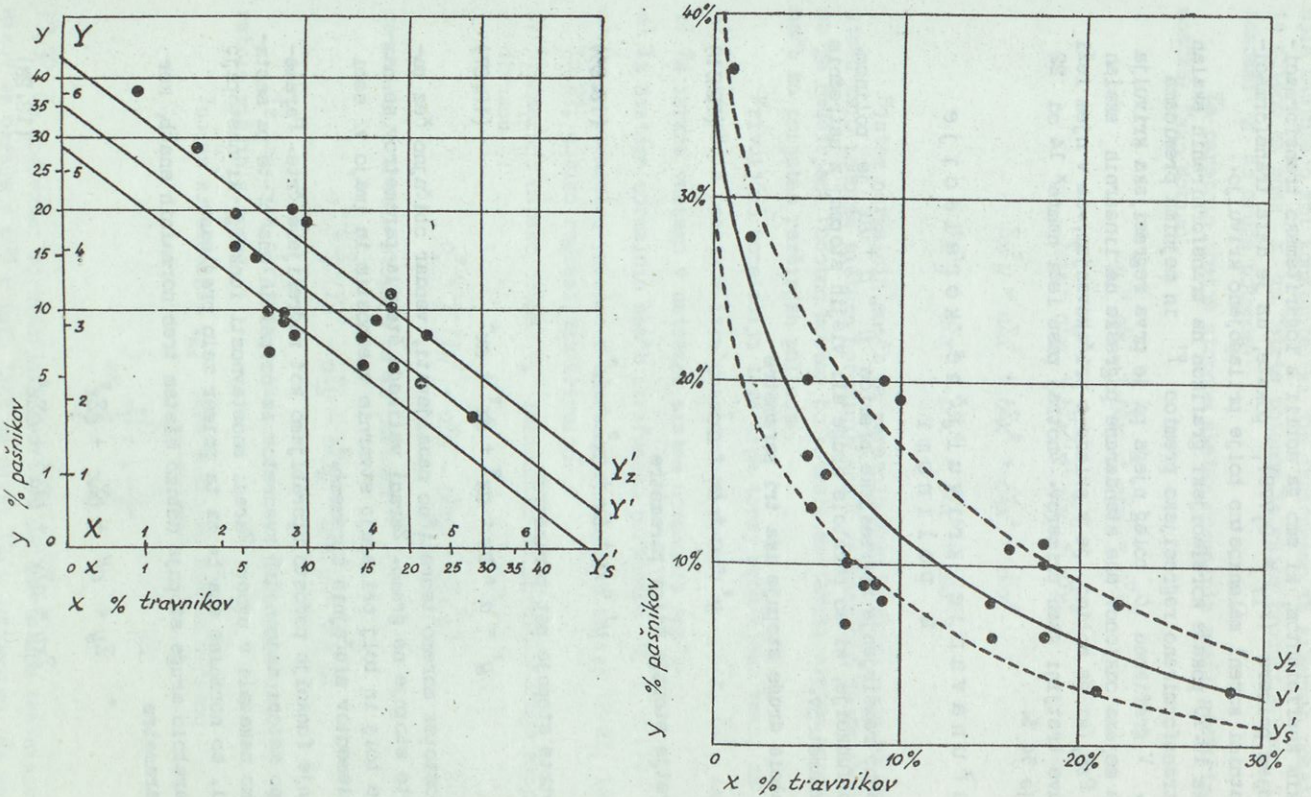
Če iščemo pas odklonov ene standarane napake ocene, dobimo dve mejni premici:

$$Y'_s = Y' - \sigma_e = 5,938 - 0,793X - 0,628 = 5,310 - 0,793X$$

$$Y'_z = Y' + \sigma_e = 5,938 - 0,793X + 0,628 = 6,566 - 0,793X$$

Če ti mejni premici prevedemo v osnovne znake, dobimo dve mejni regresijski krivulji, ki sta v eksplicitni obliki takile:

$$y'_s = (5,310 - 0,793\sqrt{x})^2 \quad ; \quad y'_z = (6,566 - 0,793\sqrt{x})^2$$



Slika 13.10 Korelacijska grafikona med odstotkom travnikov in odstotkom pašnikov za okrije v LRS v letu 1952 z včrtano regresijsko krivuljo y' in pasom ene standardne napake: a) na transformiranih skalah b) na osnovnih skalah

Primerjava dobljenega korelacijskega koeficienta ($r_{XY} = -0{,}810$) s korelacijskim koeficientom, ki smo ga dobili z logaritemsko transformacijo v prejšnjem primeru ($r_{XY} = -0{,}778$), pokaže, da je dala transformacija s kvadratnimi koreni malenkostno bolje prilagojeno krivuljo.

Slika 13.10 pokaže korelacijski grafikon na transformiranih skalah z včrtano transformirano regresijsko premico Y' in mejnima premicama Y'_s in Y'_z. V grafikonu b poleg njega pa je prva regresijska krivulja z včrtanima mejama odklonov ene standardne pogreške na linearnih skalah x in y. Pas, ki ga dobimo, je v skladu s pričakovanjem, da v njem leži približno dve tretjini vseh primerov. Znotraj pasu leži namreč 14 od 22 točk, kar je 64 %.

Izračunavanje krivuljčne korelacije s polinomi

13.31 Naravno razširjenje regresijske premice $y' = a + bx$ je polinom. Tako dobimo funkcije, ki so parabole druge ali višjih stopenj z ustrežnim številom parametrov.

Parabola druge stopnje ima tri parametre

$$y' = a + bx + cx^2 \tag{13.25}$$

parabola tretje stopnje štiri parametre

$$y' = a + bx + cx^2 + dx^3 \tag{13.26}$$

parabola četrte stopnje pet parametrov

$$y' = a + bx + cx^2 + dx^3 + ex^4 \tag{13.27}$$

Ta postopek moremo teoretično nadaljevati, vendar običajno čez parabolo četrte stopnje ne gremo. Zaradi velikega števila parametrov se namreč krivulje bolj in bolj prilegajo stvarnim vrednostim in imajo v sebi vedno več elementov slučajnih sprememb.

Zgornje funkcije pogosto uporabljamo kot regresijske črte. Parametre iščemo po metodi najmanjših kvadratov iz normalnih enačb: te pa sestavimo, kot smo nakazali v uvodu. Zaradi enostavnosti funkcij, ki nastopajo v tej obliki, so normalne enačbe za ta primer zelo pregledne.

Za parabolo druge stopnje dobimo sistem treh normalnih enačb, ker imamo tri parametre

$$\begin{aligned} \Sigma y &= aN + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned} \tag{13.28}$$

Normalne enačbe so linearne enačbe za parametre a, b in c. V njih nastopajo kot konstante, ki jih izračunamo iz stvarnih podatkov, vsote potenc za znak x do četrte potence in vsote produktov med y in potencami x do druge potence.

Za parabolo tretje stopnje sistem normalnih enačb razširimo za eno enačbo. Tako dobimo:

$$\begin{aligned} \Sigma y &= aN + b\Sigma x + c\Sigma x^2 + d\Sigma x^3 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + d\Sigma x^4 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + d\Sigma x^5 \\ \Sigma x^3 y &= a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + d\Sigma x^6 \end{aligned} \qquad (13.29)$$

Sistem normalnih enačb za parabolo tretje stopnje je po svojem sestavu zelo podoben sistemu normalnih enačb za parabolo druge stopnje, le da je razširjen. Podobno moremo po potrebi sestaviti sistem normalnih enačb za poljubno razširjen polinom.

Krivuljčno korelacijo tega tipa torej proučujemo tako, da:

a) Iz stvarnih podatkov izračunamo vsote potenc za x (Σx, Σx^2, Σx^3, Σx^4...)

b) Te izraze vnesemo v ustrezni sistem normalnih enačb.

c) Iz sistema normalnih enačb izračunamo parametre a, b, c ...

d) Če te parametre vnesemo v ustrezno funkcijsko obliko 13.25, 13.26 ali 13.27, dobimo regresijsko krivuljo.

f) Standardno napako ocene σ_e izračunamo za parabolo druge stopnje po obrazcu

$$\sigma_e^2 = \frac{1}{N}(\Sigma y^2 - a\Sigma y - b\Sigma xy - c\Sigma x^2 y) \qquad (13.30)$$

za parabolo tretje stopnje pa po obrazcu

$$\sigma_e^2 = \frac{1}{N}(\Sigma y^2 - a\Sigma y - b\Sigma xy - c\Sigma x^2 y - d\Sigma x^3 y) \qquad (13.31)$$

Iz teh dveh obrazcev spoznamo sistem, kako izračunavamo σ_e^2 za parabole višjih stopenj.

Indeks korelacije $I_{y.x}$ izračunamo po znanem obrazcu

$$I_{y.x} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}} \qquad (13.32)$$

13.32 Če pogledamo korelacijski grafikon med odstotkom travnikov in pašnikov za okraje v LRS v letu 1952 na sliki 12.4, ugotovimo, da moremo v

tem primeru za regresijsko krivuljo vzeti tudi del parabole stopnje tipa $y' = a + bx + xc^2$.

Zanjo velja sistem normalnih enačb 13.28. Zato moramo iz osnovnih podatkov izračunati izraze: Σy, Σxy, $\Sigma x^2 y$, Σx, Σx^2, Σx^3, Σx^4. V tabeli 13.19 je nakazan izračun teh izrazov.

Tabela 13.19 Izračunavanje krivuljične regresije med odstotki travnikov in pašnikov za okraje v letu 1952

i	x	y	xy	x^2y	x^2	x^3	x^4	y^2
1	22	8	176	3872	484	10648	234256	64
2	3	28	84	252	9	27	81	784
3	9	20	180	1620	81	729	6561	400
4	5	13	65	325	25	125	625	169
5	6	15	90	540	36	213	1296	225
6	7	7	49	343	49	343	2401	49
7	16	9	144	2304	256	4093	65536	81
8	28	3	84	2352	784	21952	614356	9
9	18	11	198	3534	324	5832	104976	121
10	18	6	108	1944	324	5832	104976	36
11	15	6	90	1350	225	3375	50625	36
12	15	8	120	1800	225	3375	50625	64
13	21	3	63	1323	441	9231	194481	1
14	8	9	72	576	64	512	4096	81
15	10	19	190	1900	100	1000	10000	361
16	18	10	180	3240	324	5832	104976	100
17	5	16	80	400	25	125	635	256
18	1	37	37	37	1	1	1	1396
19	7	10	70	490	49	343	2401	100
20	8	9	72	576	64	512	4093	81
21	5	20	100	500	25	125	625	400
22	9	8	72	648	81	729	6561	64
	254	275	2324	29956	3993	74990	1564476	4859

Če te izraze vnesemo v sistem normalnih enačb 13.28, dobimo:

$$22a + 254b + 3993c = 275$$

$$254a + 3996b + 74990c = 2324$$

$$3996a + 74990b + 1564476c = 29956$$

Iz tega sistema normalnih enačb dobimo parametre:

$$a = +29,0468 \; ; \quad b = -2,3066 \; ; \quad c = +0,0555$$

Regresijska krivulja pa je

$$y' = +29,0468 - 2,3066x + 0,0555x^2$$

Standardna napaka ocene je po obrazcu 13.30 enaka:

$$\sigma_e^2 = \frac{1}{22}\left[(4859 - (+29,0468).275 - (-2,3066).2324 - (+0,0555).29956\right] = 25,8686$$

$$\sigma_e = \sqrt{25,8686} = 5,086$$

Ker je

$$\sigma_y^2 = \frac{1}{N}\left[\Sigma y^2 - Y^2/N\right] = \frac{1}{22}\left(4859 - \frac{275^2}{22}\right) = 64,6136$$

dobimo, da je indeks korelacije $I_{y.x}$ po obrazcu 13.32 enak

$$I_{y.x} = \sqrt{1 - \frac{25,8686}{64,6136}} = \sqrt{0,5996} = 0,774$$

Če primerjamo dobljeni indeks korelacije $I_{y.x} = 0,774$ s korelacijskima koeficientoma, katera smo dobili z logaritemsko transformacijo $r_{XY} = 0,778$ ali s transformacijo s kvadratnimi koreni $r_{XY} = 0,804$, spoznamo, da smo s parabolo druge stopnje tipa $y' = a+bx+cx^2$ dobili manjši indeks korelacije, čeprav ima ta funkcija tri, transformirani funkciji pa samo po dva parametra. Iz tega moremo zaključiti, da transformacija, posebno pa transformacija s kvadratnim korenom, da boljše rezultate.

KORELACIJSKO RAZMERJE

13.33 Pri obravnavanju regresijskih krivulj smo videli, da moremo kot regresijsko črto vzeti tudi vrsto grupnih sredin. Ta regresijska črta je sorazmerno najbolje prilagojena stvarnim vrednostim, ker je aritmetična sredina za vsako grupo najboljši reprezentant vrednosti v svojem razredu. Če

za nek pojav dobimo, da so grupne sredine konstantne vrednosti, sklepamo, da znak, po katerem smo izvedli grupacijo, ne vpliva na znak, za katerega smo izračunali aritmetične sredine. Nasprotno pa pomeni, da je ta vpliv velik, če so razlike med grupnimi aritmetičnimi sredinami velike.

V primeru grupnih aritmetičnih sredin imamo varianco med grupnimi aritmetičnimi sredinami za pojasnjen del variance, varianco znotraj grup pa za nepojasnjen del variance. To načelo je istovetno z uvodnim tolmačenjem o pojasnjenem in nepojasnjenem delu variance.

Že pri izračunavanju skupne variance iz grupnih varianc (glej obrazce 10.14; 10.15 in 10.16) smo dobili, da je skupna varianca σ_y^2 vsota variance med aritmetičnimi sredinami $\sigma_{\bar{y}}^2$ in aritmetične sredine grupnih varianc M_{σ^2}.

$$\sigma_y^2 = \sigma_{\bar{y}}^2 + M_{\sigma^2} \tag{13.33}$$

Enako kot doslej vzamemo za merilo korelacije razmerje med pojasnjeno varianco (varianco med aritmetičnimi sredinami) $\sigma_{\bar{y}}^2$ in skupno varianco σ_y^2. Po tej definiciji izračunamo razmerje, ki ga imenujemo *korelacijsko razmerje*, zaznamujemo pa z η^2 (eta kvadrat). Korelacijsko razmerje pokaže maksimalno stopnjo povezanosti dveh pojavov, ker se aritmetične sredine najbolj prilagajo stvarnim podatkom.

Po tej definiciji je korelacijsko razmerje

$$\eta^2_{y.x} = \frac{\sigma_{\bar{y}}^2}{\sigma_y^2} = \frac{\frac{1}{N}\Sigma N_k(\bar{y}_k - \bar{y})^2}{\frac{1}{N}\Sigma(y-\bar{y})^2} \tag{13.34}$$

Razen dobrih lastnosti pa ima korelacijsko razmerje svoje hibe. Težko je namreč presoditi, kdaj so razredi pravilni, ker tako premajhni kot preveliki razredi popačijo rezultate. Sredine iz premajhnih razredov morejo biti rezultat slučajnih vplivov, grupne sredine iz prevelikih razredov pa ne podajajo regresijske smeri. Zato po pravilu razrede s premajhnim številom enot vključujemo s sorodnimi razredi.

Velika prednost korelacijskega razmerja pa je ta, da ga moremo izračunati tudi, če je en znak atributiven, drugi pa numeričen, medtem ko je bilo doseganje obravnavanje korelacije omejeno le na odvisnosti med numeričnimi znaki. Grupe moremo namreč tvoriti tako za numerične kot za atributivne znake. S korelacijskim razmerjem moremo izračunati mero odvisnosti plač od kvalifikacije, pridelka od različnih vrst agrotehnike, kvalitete proizvodnje od izmene, starosti ob sklenitvi zakona od socialne skupine itd.

13.34 Izračunavanje iz negrupiranih podatkov. Izračunavanje korelacijskega razmerja z osnovnim obrazcem je sicer izvedljivo, vendar nepri-

kladno. Zato tehnično izvedemo izračun korelacijskega razmerja takole:

a) Sestavimo za znak x grupe s primernim številom enot po grupah.

b) Izdelamo tabelo, v kateri so podatki koreliranega znaka y vpisani po grupah za x.

c) Izračunamo grupne vsote Y_k za znak y in vsoto vseh grupnih vsot $Y = \Sigma Y_k$

d) Izračunamo vsoto kvadratov vseh individualnih vrednosti za y. Tako dobimo Σy^2

e) Vsako grupno vsoto Y_k kvadriramo in delimo s številom enot v ustrezni grupi N_k. Te izraze za vse grupe seštejemo in dobimo izraz $\Sigma Y_k^2/N_k$

f) Iz skupne vsote Y in skupnega števila enot N izračunamo analogen izraz Y^2/N

g) Korelacijsko razmerje izračunamo iz dobljenih izrazov po obrazcu

$$\eta^2_{y.x} = \frac{\Sigma Y_k^2/N_k - Y^2/N}{\Sigma y^2 - Y^2/N} \qquad (13.35)$$

13.35 Izračunajmo korelacijsko razmerje za odvisnost stroškov za kulturno in družbeno življenje od skupnih dohodkov za 36 družin v Mariboru iz tabele 13.3. Grupe dohodkov vzemimo po 5.000 din od 15.000 din dalje. Izračunavanje korelacijskega razmerja je pregledno nakazano v tabeli 13.20.

Tabela 13.20 Izračunavanje korelacijskega razmerja η^2 za odvisnost stroškov za kulturno in družbeno življenje za 36 delavskih družin v Mariboru

Dohodek v tisoč din x	Stroški za družbeno in kulturno življenje v din y					N_k	Y_k	Y_k^2	Y_k^2/N_k
15 – 19	260	330	490	310	580				
	560	330	480	860	590				
	330	630				12	5750	33062500	2755208
20 – 24	650	1100	760	460	700				
	540	750	370	630	740	10	6700	44890000	4489000
25 – 29	1080	630	1210	1100	870				
	640	1110				7	6640	44089600	6298514
30 – 34	1610	1070	1600	990		4	5270	27772900	6943225
35 –	1590	1710	1810			3	5110	26112100	8704033
$\Sigma y^2 = 230^2+330^2+\ldots+1810^2 =$					30584100	36	29470	868480900	29189980
						N	Y	Y^2	$\Sigma Y_k^2/N_k$
									24124439 = Y^2/N

Po obrazcu 13.35 dobimo $\eta^2_{y.x} = \frac{29189980 - 24124469}{30584100 - 24124469} = 0,784$

Če primerjamo dobljeno korelacijsko razmerje $\eta^2_{y.x} = 0,784$ z determinacijskim koeficientom $r^2_{xy} = 0,518$, ki smo ga izračunali za iste podatke, vidimo, da je razlika znatna. To kaže na nelinearno odvisnost stroškov za kulturno življenje od plače, ker gre samo 0,518 od skupno pojasnjene variance 0,784 na račun linearne povezanosti.

13.36 Izračunavanje iz grupiranih podatkov. Če za podatke v grupah nimamo individualnih podatkov, temveč frekvenčne distribucije, je najprikladneje izračunati korelacijsko razmerje s kumulativnimi vrstami. Postopek je naslednji:

a) Imamo tabelo grupnih frekvenčnih distribucij f_k in sumarno frekvenčno distribucijo f.

b) Po znanih načelih (glej izračunavanje variance s kumulativami!) izračunamo za grupne frekvenčne distribucije prve kumulative F_k, iz njih pa količine A_k.

c) Za sumarno frekvenčno distribucijo izračunamo prvo kumulativno vrsto F, drugo kumulativno vrsto FF in iz nje količino B.

Kontrola: Zadnji člen druge sumarne kumulativne vrste A je enak vsoti grupnih A_k: $\Sigma A_k = A$.

d) Iz grupnih količin A_k in N_k izračunamo izraze A_k^2/N_k in jih seštejemo. Tako dobimo $\Sigma A_k^2/N_k$

e) Enako izračunamo iz sumarnega A in N količino A^2/N.

f) Korelacijsko razmerje izračunamo iz teh količin po obrazcu

$$\eta^2_{y.x} = \frac{\Sigma A_k^2/N_k - A^2/N}{2B + A - A^2/N} \qquad (13.33)$$

Če za analizo potrebujemo vrsto grupnih aritmetičnih sredin, jih izračunamo po obrazcu

$$\bar{y}_k = y_0 + i_y \frac{A_k}{N_k} \qquad (13.37)$$

y_0 = sredina najnižjega razreda

13.37 Izračunajmo kot primer korelacijsko razmerje za odvisnost denarnih dohodkov v kmetijskih gospodarstvih od velikosti gospodarstva iz tabele 5.10.

V tabeli 13.21 je pregledno nakazano, kako izračunamo osnovne količine, ki jih potrebujemo, če izračunamo korelacijsko razmerje po zgornjem postopku. Ta način je posebno priladen, če kumulative izračunavamo s seštevalnim strojem, ki ima kontrolni trak.

Tabela 13.21 Izračunavanje korelacijskega razmerja za odvisnosti dohodka od kmetijstva od velikosti gospodarstva v letu 1956 (Vir: Tabela 5.10)

Dohodek v tisoč din	Površina v ha													
	-4		4-8		8-12		12-16		16-20		20-			
	f_1	F_1	f_2	F_2	f_3	F_3	f_4	F_4	f_5	F_5	f_6	F_6	f	F	FF
600-700	2	0	4	0	7	0	1	0	2	0	4	0	20	0	
500-600	5	2	2	4	4	7	2	1	5	2	2	4	20	20	0
400-500	7	7	14	6	9	11	2	3	7	7	2	6	35	40	20
300-400	19	14	35	20	30	20	7	5	8	8	2	8	96	75	60
200-300	46	33	91	55	40	50	14	12	7	11	8	10	206	171	135
100-200	203	79	172	146	72	90	20	26	5	18	7	18	279	377	306
0-100	942	282	337	318	72	162	21	46	3	23	2	25	1380	856	683
N_k		1224		655		234		67		29		27		$2236 = N$	$1539 = A$ $1204 = B$
A_k		417		549		340		98		69		71	$1539 = \Sigma A_k$		
A_k^2		173889		301401		115600		8649		4761		5041	$A^2/N = \frac{1539^2}{2236} = 1059,27$		
A_k^2/N_k		142,07		460,02		494,02		129,09		164,17		186,70	$1576,07 = \Sigma A_k^2/N$		

Po obrazcu 13.33 dobimo dalje, da je

$$\eta^2_{y.x} = \frac{1576,07 - 1059,27}{2.1204 + 1539 - 1059,27} = 0,179$$

Rezultat pokaže, da je $\eta^2_{y,x} = 0,179$ od skupne variance dohodkov iz kmetijstva pojasnjene s skupno površino.

Po obrazcu 13.27 dobimo iz podatkov v tabeli 13.21 tole vrsto grupnih sredin.

Tabela 13.22 Grupne sredine dohodkov od kmetijstva po velikostnih skupinah

Površina v ha	Povprečni dohodek v tisoč dinarjih
do 4 ha	84
nad 4 do 8 ha	135
nad 8 do 12 ha	195
nad 12 do 16 ha	189
nad 16 do 20 ha	288
nad 20 ha	313

KORELACIJA RANGA

13.33 Rang je v ozki zvezi z vrednostjo znaka. Večji vrednosti ustreza večji rang in obratno. To lastnost rangov s pridom uporabljamo tudi pri proučevanju korelacije med pojavi. Razvrstimo enote po velikosti po obeh koreliranih znakih in enotam pripišemo rang, enkrat po prvem, drugič pa po drugem znaku! Če sta proučevana znaka v ozki pozitivni povezavi, se rangi obeh znakov med seboj precej dobro ujemajo. V skrajnem primeru – če je povezava funkcionalna, so celo med seboj identični. Če je med pojavoma majhna povezava ali je sploh ni, pa med rangi ni nobene zveze. Če osnovne podatke y nadomestimo z rangi, imamo namesto niza dvojic vrednosti x in y niz dvojic rangov za x in y. Če je med znakoma x in y povezava tesna, je tesna tudi med rangoma. Če pa je povezava rahla ali je sploh ni, moremo to opaziti tudi na rangih.

To lastnost rangov izkoriščamo in izračunavamo koeficient korelacije iz rangov namesto iz osnovnih podatkov. Če število enot ni preveliko, je izračunavanje korelacije ranga znatno enostavnejše kot izračunavanje korelacijskega koeficienta r_{xy}. Koeficient korelacije ranga izračunavamo po Spearmanovem obrazcu

$$\rho = 1 - \frac{6\Sigma d^2}{N(N^2-1)} \qquad (13.38)$$

Pri tem je: ρ = koeficient korelacije ranga; Σd^2 = vsota kvadratov razlik med rangoma za oba znaka. Ker je koeficient korelacije ranga navaden koeficient korelacije med rangi, je vrednost Spearmanovega koeficienta korelacije ranga med -1 do $+1$.

Izračunavanje Spearmanovega koeficienta ranga ρ je enostavno in ga izračunavamo po naslednjih stopnjah:

a) Najprikladneje je, če osnovne podatke vpišemo na obdelovalne listke tako, da na vsak listek napišemo ustrezna podatka x in y za posamezno enoto.

b) Ko prepišemo vse podatke na obdelovalne listke, razporedimo listke po velikosti vrednosti za znak x in na listke vpišemo ustrezne range R_x.

c) Enako razporedimo listke po velikosti vrednosti znaka y in tudi zanj vpišemo na listke range R_y.

d) Za vsako enoto izračunamo razliko rangov $d = R_x - R_y$ in razlike rangov kvadriramo.

e) Kvadrate razlik rangov d^2 za posamezne enote seštejemo in vsoto kvadratov diferenc Σd^2 vnesemo v obrazec (13.38)

Vrednosti r_{xy} in ρ, izračunana iz istih podatkov, nista enaki, ker sta izračunana različno. Vendar v večini primerov razlike niso velike in moremo ρ vsaj v prvi stopnji proučevanja korelacije s pridom uporabiti kot orientacijski podatek. Prednost korelacije ranga je tudi v tem, da zmanjša učinek ekstremnih vrednosti: to je ena izmed hib navadnega korelacijskega koeficienta r_{xy}.

S korelacijo ranga razširimo proučevanje korelacije tudi na nekatere nenumerične znake. Veliko atributivnih znakov je namreč takih, da moremo njihove vrednosti razporediti po velikosti in jim določiti rang. Za take znake sicer ne moremo izračunati korelacijskega koeficienta r_{xy}, ker so vrednosti znaka atributivne narave, moremo pa zanje izračunati korelacijski koeficient ranga ρ. Tako moremo kolektiv delavcev razporediti po delovni vnemi, ki jo sicer ne moremo ali zelo težko izrazimo numerično, učence po zanimanju za posamezne predmete, itd.

Korelacija ranga je posebno prikladna, kadar proučujemo odvisnost več znakov. V teh primerih dobimo razmeroma hitro in enostavno celoten pregled o korelaciji med pojavi.

13.39 Da bomo imeli primerjavo z že izračunanim r_{xy}, vzemimo kot primer izračunavanje korelacije ranga za korelacijo med številom delavcev in številom opravljenih delovnih ur za 19 podjetij industrije obutve iz tabele 13.12.

Podatki in način izračunavanja so dani v tabeli 13.23. Rangi so bili predhodno določeni z obdelovalnimi listki. Iz obdelovalnih listkov je bila nato sestavljena tabela 13.23.

Tabela 13.23 Izračunavanje korelacijskega koeficienta ranga ρ med številom delavcev (x) in številom opravljenih delovnih ur (y) za 19 podjetij industrije obutve v LRS

x	y	R_x	R_y	d	d^2
7	0,8	1	1	0	0
12	2,2	2	2	0	0
15	2,9	3	3	0	0
23	4,1	4	4	0	0
28	4,9	5	5	0	0
37	7,3	6	3	0	0
49	9,9	7	8	-1	1
50	9,5	8	7	+1	1
51	10,1	9	10	-1	1
57	11,0	10	11	-1	1
58	10,0	11	9	+2	4
80	15,1	12	13	-1	1
81	15,0	13	12	+1	1
92	17,9	14	15	-1	1
100	16,7	15	14	+1	1
111	20,7	16	16	0	0
112	25,4	17	17	0	0
150	31,3	18	18	0	0
154	35,3	19	19	0	0
				0	$12 = \Sigma d^2$

Kjer je $N = 19$ in $\Sigma d^2 = 12$, dobimo, da je po obrazcu 13.38

$$\rho = 1 - \frac{6 \cdot 12}{19(19^2-1)} = 0{,}9895$$

Če primerjamo $\rho = +0{,}9895$ z ustreznim $r_{xy} = 0{,}988$, ki smo ga izračunali v tabeli 13.12, vidimo, da so razlike neznatne.

ASOCIACIJA IN KONTINGENCA

Asociacija

13.40 Iz kombinacijske tabele med starostjo moža in žene ob razvezi zakona v LRS v letu 1955 v tabeli 13.5 iz grupiranja frekvenc sklepamo na korelacijo med obema podatkoma. Vprašanje je, ali je možno tudi iz kombinacijske tabele za atributivne znake sklepati na povezavo med atributivnimi znaki.

Vzemimo za primer najenostavnejšo kombinacijsko tabelo 2 x 2, ki ima po vsakem znaku samo po dve vrednosti. Taka tabela je na primer tabela o številu prebivalstva v LR Sloveniji po popisu leta 1953 po delavnosti in spolu. V primeru odvisnosti med dvema atributivnima znakoma, ki imata samo po dve vrednosti, govorimo o asociaciji med znaki.

Tabela 13.24 Število prebivalstva v LR Sloveniji po popisu 1953 po spolu in delavnosti v tisočih prebivalcev (Vir SG 1958)

Spol \ Delavnost	Delovni	Vzdrževani	Skupno
Moški	474	219	693
Ženske	330	443	773
Skupno	804	662	1 466

Že na prvi pogled moremo presoditi, da je delavnost odvisna od spola, ker je pri moških število delovnih večje kot število vzdrževanih, medtem ko je situacija pri ženskah obratna in je število vzdrževanih žena večje kot število delovnih. Še jasneje te odnose vidimo iz struktur v tabeli 13.25.

Tabela 13.25 Struktura števila prebivalstva v LR Sloveniji po popisu 1953

Spol \ Delavnost	Delovni	Vzdrževani	Skupno
Moški	68,4	31,6	100,0
Ženske	42,7	57,3	100,0
Skupno	54,8	45,2	100,0

Medtem ko je moških delovnih 68,4 %, je odstotek delovnih žena samo 42,7 %.

Čim večja je odvisnost delavnosti od spola, tem večja je tudi razlika med temi odstotkoma, in obratno: čim manjša je odvisnost delavnosti od spola, tem manjša je razlika med temi odstotkoma. Če bi bila delavnost neodvisna od spola, bi bila odstotka enaka, in sicer enaka povprečnemu odstotku 54,8 %. Ker vemo za skupno število moških in žensk, moremo izračunati, kolikšno bi bilo število prebivalstva v LR Sloveniji, če delavnost ne bi bila odvisna od spola. Ker bi bilo v tem primeru recimo 54,8 % od skupno 393 tisočev moških delovnih, bi bilo število delovnih 393 . 0,548 = = 380 tisoč. Podobno izračunamo druge podatke.

Tako dobimo tabelo 13.23, ki pokaže, kakšen bi bil sestav prebivalstva po spolu in delavnosti, če delavnost ne bi bila odvisna od spola.

Tabela 13.26 Teoretično število prebivalstva v LRS, če delavnost ne bi bila odvisna od spola

Spol \ Delavnost	Delovni	Vzdrževani	Skupno
Moški	380	313	693
Ženske	424	349	773
Skupno	804	662	1466

Če primerjamo tabelo 13.23 s tabelo 13.24, opazimo, da se obe v robnih frekvencah ujemata, ne ujemata pa se v frekvencah v notranjih poljih. Razlik ne bi bilo, če bi delavnost stvarno ne bila odvisna od spola. Razlike so tem večje, čim bolj je delavnost odvisna od spola.

Ker je 0,548 = 804/1433, moremo izračunati teoretično frekvenco tudi takole: 380 = 393.804/1433. Iz tega moremo izdelati splošno pravilo: teoretično frekvenco za dano polje izračunamo, če pomnožimo polju ustrezni robni frekvenci, produkt pa delimo s skupno frekvenco N.

Če iz simbolov a, b, c in d, sestavimo štiripoljno tabelo

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

(13.39)

je teoretična frekvenca za prvo polje enaka:

$$a' = \frac{(a+b) \cdot (a+c)}{N} \qquad (13.40)$$

Analogno dobimo teoretične frekvence za druga polja.

13.41 Ker so diference med stvarnimi in teoretičnimi frekvencami tem večje, čim večja je odvisnost - *asociacija* med obema znakoma, moremo vzeti razliko $a - a'$ za merilo odvisnosti.

Tako dobimo, da je

$$\delta = a - a' = a - \frac{(a+b) \cdot (a+c)}{N} = \frac{ad - bc}{N} \qquad (13.41)$$

pozitiven pri pozitivni in negativen pri negativni odvisnosti ali asociaciji. Njegova absolutna vrednost pa je tem večja, čim večja je asociacija med znakoma

13.42 Ker to merilo razen teh splošnih oznak po svoji vrednosti ni določeno, ker je odvisno tudi od velikosti populacije, je Yule vpeljal kot merilo asociacije *koeficient asociacije* Q

$$Q = \frac{ad-bc}{ad+bc} \qquad (13.43)$$

Koeficient asociacije Q ima vrednosti med -1 do +1. Q je negativen, če je asociacija negativna, in pozitiven, če je asociacija pozitivna. Q je nič, če pojava nista odvisna, ena pa pri funkcionalni povezavi. Koeficient asociacije Q ima torej vse lastnosti dobrega merila odvisnosti.

13.43 Če damo eni vrednosti posameznega atributa numerično vrednost 0, drugi pa 1, moremo iz podatkov štiripoljne tabele formalno izračunati navadni koeficient korelacije r_{xy}. Tako dobimo novo merilo asociacije V. Tega izračunamo po obrazcu

$$V = \frac{ad - bc}{\sqrt{(a+b).(a+c).(b+d).(c+d)}} \qquad (13.44)$$

Koeficient asociacije V ima iste lastnosti kakor korelacijski koeficient r_{xy}.

13.44 Kot primer analizirajmo asociacijo med spolom in delavnostjo za število prebivalstva po popisu 1953 po republikah.

Osnovni podatki so dani v tabeli 13.27.

Za Slovenijo smo koeficiente asociacije izračunali takole:

$$ad - bc = 474.443 - 219.330 = +137712$$

$$\delta = \frac{ad-bc}{N} = +137712/1466 = 93,937$$

$$Q = \frac{ad-bc}{ad+bc} = \frac{+137712}{474.443+219.330} = +0,488$$

$$V = \frac{ad - bc}{\sqrt{(a+b).(a+c).(b+d).(c+d)}} = \frac{+\ 137712}{\sqrt{398.804.332.773}} = +0{,}258$$

Tabela 13.27 Število prebivalstva po spolu in delovnosti po republikah v milj. (Vir: Popis prebivalstva 1953)

Republika	Spol	Delovnost		S
		D	V	
Srbija	M	2306	1106	3412
	Ž	1266	2302	3568
	S	3572	3408	6980
Hrvatska	M	1281	571	1852
	Ž	747	1319	2066
	S	2028	1890	3918
Slovenija	M	474	219	693
	Ž	330	443	773
	S	804	662	1466
BiH	M	843	543	1386
	Ž	423	1039	1462
	S	1266	1582	2848
Makedonija	M	394	266	660
	Ž	156	489	645
	S	550	755	1305
Črna gora	M	115	87	202
	Ž	55	133	218
	S	170	250	420

V tabeli 13.28 so vpisani koeficienti asociacije Q za vse republike. Izračunali smo jih enako kakor za Slovenijo.

Če pogledamo koeficiente asociacije Q, ki merijo asociacijo med spolom in delovnostjo, opazimo izjemno nizek koeficient asociacije za Slovenijo (Q =+0.488) in izjemno visok za Makedonijo (Q = +0.343). Za druge republike pa so koeficienti asociacije med Q = +0.583 in Q = +0.597. To je v skladu z razmerami, ker so v Sloveniji zaradi različnega položaja ženske razlike v delovnosti po spolu manjše kakor v drugih republikah, posebno v Makedoniji.

Tabela 13.28 Koeficienti asociacije Q med delovnostjo in spolom prebivalstva po republikah. (Po podatkih iz tabele 13.27.)

Republika	Q
Srbija	+0.583
Hrvatska	+0.597
Slovenija	+0.488
BiH	+0.584
Makedonija	+0.646
Črna gora	+0.593

Kontingenca

13.45 Vzemimo splošnejši primer, v katerem posamezen znak v kombinacijski kontingenčni tabeli nima dveh vrednosti, marveč več. V tem primeru govorimo o kontingenci.

Za posamezna polja v kontingenčni tabeli dobimo teoretične frekvence f'_{kg} za primer neodvisnosti med pojavoma podobno kot za asociacijo.

Stvarni frekvenci f_{kg} ustrezajočo teoretično frekvenco za primer neodvisnosti f'_{kg}, namreč dobimo, če pomnožimo ustrezni robni frekvenci f_k in f_g, produkt pa delimo s skupno frekvenco N. Z obrazcem to izrazimo takole:

$$f'_{kg} = \frac{f_k \cdot f_g}{N} \qquad (13.44)$$

Enako kakor pri asociaciji, so tudi tu majhne razlike med stvarnimi f_{kg} in teoretičnimi f'_{kg} frekvencami izraz majhne, velike razlike pa izraz velike odvisnosti med pojavoma. Iz teh razlik med teoretičnimi in stvarnimi frekvencami $f-f'$ je sestavljeno merilo kontingence χ^2:

$$\chi^2 = \Sigma \frac{(f-f')^2}{f'} \qquad (13.45)$$

Glede na zgornji obrazec izračunamo χ^2 (*hi* kvadrat) tako, da:

a) iz robnih frekvenc po obrazcu 13.44 izračunamo tabelo teoretičnih frekvenc f',

b) izračunamo diference med ustreznimi stvarnimi f in teoretičnimi f' frekvencami,

c) dobljene diference kvadriramo $(f-f')^2$

c) kvadrate diferenc delimo z ustreznimi teoretičnimi frekvencami f'

d) če seštejemo dobljene izraze $(f-f')^2/f'$ izračunane za vsa polja v kontingenčni tabeli, dobimo χ^2.

e) Včasih je primerneje, če izračunavamo χ^2 po obrazcu

$$\chi^2 = \Sigma \frac{f^2}{f'} - N \qquad (13.43)$$

V tem primeru namesto diferenc $f-f'$ kvadriramo stvarne frekvence f; to je večinoma enostavneje.

Iz osnovnega obrazca za χ^2 sklepamo, da je $\chi^2 = 0$, če ni razlik med stvarnimi in teoretičnimi frekvencami; če sta pojava neodvisna, je χ^2 večji, če so razlike večje, če je torej odvisnost med pojavi večja. Ta lastnost, ki jo ima χ^2, ustreza merilu za kontingenco.

Vendar χ^2 ne more neposredno primerjati za različne pojave in različne populacije. Zato večkrat iz njega izračunavamo *povprečno kvadratično kontingenco* Φ^2

$$\Phi^2 = \frac{\chi^2}{N} \qquad (13.47)$$

ki omili vpliv razlik v obsegu populacij, ali *Pearsonov koeficient kontingence* C

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \qquad (13.48)$$

Za C je sicer zgornja meja odvisna od števila razredov, vendar C nikdar ni večji kot 1.

13.46 Kot primer za izračunavanje χ^2 in koeficientov kontingence Φ^2 in C vzemimo kontingenčno tabelo med stanom ženinov in nevest pred sklenitvijo zakona za LR Slovenijo v letu 1955.

Tabela 13.30 Sklenjeni zakoni po prejšnjem stanu ženina in neveste v LRS v letu 1956

Prejšnji stan ženina	Prejšnji stan neveste			Skupaj
	samska	vdova	razvedena	
samec	12158	155	200	12513
vdovec	350	129	45	524
razveden	479	59	165	703
Skupaj	12987	343	410	13740

Teoretična frekvenca za prvo polje v prejšnji tabeli (samec - samska) je po obrazcu 13.44

$$f'_{11} = \frac{12987 \cdot 12513}{13740} = 11827$$

za naslednje polje (samec - vdova)

$$f'_{12} = \frac{343 \cdot 12513}{13740} = 312$$

Analogno izračunamo f' za vsa druga polja.

Tako dobimo tabelo teoretičnih frekvenc pod predpostavko, da prejšnji stan ženina in neveste nista odvisna.

Tabela 13.31 Sklenjeni zakoni po prejšnjem stanu ženina in neveste v LRS v letu 1956 (teoretične frekvence f')

Prejšnji stan ženina	Prejšnji stan neveste			Skupaj
	samska	vdova	razvezana	
samski	11827	312	374	12513
vdovec	495	13	16	524
raveden	665	18	20	703
Skupaj	12987	343	410	13740

Tabela 13.32 Razlike med stvarnimi in teoretičnimi frekvencami ($(f - f')$ iz podatkov v tabelah 13.30 in 13.31)

Prejšnji stan ženina	Prejšnji stan neveste			Skupaj
	samska	vdova	razvezana	
samski	+331	-157	-174	0
vdovec	-145	+116	+ 29	0
razveden	-186	+ 41	+145	0
Skupaj	0	0	0	0

Če pogledamo tabelo diference $f - f'$, vidimo, da so diference velike in pozitivne za ustrezne kombinacije samski - samska, vdovec - vdova, razvezan - razvezana. Iz te tabele je dobro vidna medsebojna odvisnost prejšnjega stanu zakoncev.

χ^2 izračunamo dalje po obrazcu 13.45

$$\chi^2 = \sum \frac{(f-f')^2}{f'} = \frac{(+331)^2}{11827} + \frac{(-157)^2}{312} + \frac{(-174)^2}{374} + \ldots + \frac{(+41)^2}{18} + \frac{(+145)^2}{20} = 3495,997$$

Iz tega je dalje

$$\Phi^2 = \chi^2/N = 3495,997/13740 = 0.2544$$

$$C = \sqrt{\frac{\chi^2}{\chi^2+N}} = \sqrt{\frac{3495,997}{3495,997 + 13740}} = 0,450$$

Dobljeni koeficient C moremo primerjati z analognimi koeficienti za druge republike, za Slovenijo za druga leta itd., medtem ko χ^2 ni neposredno primerljiv za različno velike populacije.

PARCIALNA KORELACIJA

13.47 Dva pojava sta med seboj povezana, če sta vzročno odvisna ali pa, če sta odvisna od istih faktorjev, ki nanju delujejo. Običajno je del povezanosti rezultati neposredne - vzročne odvisnosti, del povezanosti pa je rezultat posredne povezanosti zaradi vpliva drugih faktorjev. Če raziskujemo neposredno odvisnost med dvema pojavoma tako, da odstranimo vpliv enega ali več faktorjev, govorimo o parcialni ali delni korelaciji.

Izračunavanje parcialnih korelacijskih koeficientov, ki pokažejo odvisnost med dvema pojavoma, če odstranimo skupen vpliv drugih faktorjev, je razmeroma enostavno, če je odvisnost linearna. Posebno je enostavno izračunati koeficient parcialne korelacije, če odstranimo vpliv enega samega faktorja.

Parcialni korelacijski koeficient $r_{12.3}$ med pojavoma 1 in 2, če odstranimo vpliv faktorja 3, dobimo iz ustreznih enostavnih korelacijskih koeficientov za linearno korelacijo r_{12}, r_{13}, r_{23} po obrazcu

$$r_{12.3} = \frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \qquad (13.49)$$

13.48 Vzemimo za primer, kako je produktivnost dela, ki jo merimo z vrednostjo proizvodnje na enega delavca (x_1), odvisna od števila delavcev v podjetju (x_2). Pri tem proučimo, kaj je s to odvisnostjo, če odstranimo vpliv porabljene električne energije v podjetju (x_3) na povezanost med številom delavstva in produktivnostjo dela.

Proučevanje obsega 27 podjetij tekstilne industrije v Sloveniji v letu 1957, ki imajo zaposlenih pod 1.000 delavcev in je njih osnovna dejavnost proizvodnja tkanin. Enostavni korelacijski koeficienti, izračunani iz podatkov za zgornja podjetja, so:

med proizvodnostjo dela in številom zaposlenih: $r_{12} = +0,03383$;

med proizvodnostjo dela in porabo električne energije: $r_{13} = +0,84802$;

med številom zaposlenih in porabo električne energije: $r_{23} = +0,72000$.

Korelacijski koeficient r_{12} med proizvodnostjo dela in številom zaposlenih pokaže zelo majhno povezanost med tema dvema pojavoma. Če izračunamo po obrazcu 13.49 parcialni korelacijski koeficient med proizvodnostjo dela (x_1) in številom zaposlenih (x_2), če odstranimo vpliv porabe električne energije (x_3), dobimo:

$$r_{12.3} = \frac{0,03383 - 0,84802 \cdot 0,72000}{\sqrt{1 - 0,84802^2}\sqrt{1 - 0,72000^2}} = -0,333$$

Dobljeni rezultat se zdi nerazumljiv, ker pokaže, da produktivnost dela, merjena s proizvodnjo na enega zaposlenega, pri isti porabi električne energije pada, če se število zaposlenih veča. Rezultat pa je stvaren, če upoštevamo, da se razmerje med porabljeno električno energijo in številom zaposlenih slabša, če se pri isti porabi električne energije število zaposlenih veča.

MULTIPLA KORELACIJA

13.49 Doslej smo obravnavali probleme odvisnosti nekega pojava od enega faktorja. Vemo pa, da posamezni socialno - ekonomski pojavi niso odvisni od enega, temveč od več faktorjev. Proizvodnja v podjetju je odvisna od števila zaposlenih, porabljene električne energije, števila strojev itd. Cena je odvisna od količine na trgu in povpraševanja. Hektarski donos je odvisen od gnojenja, padavin, zemlje, lege parcele itd.

Analogno kot proučujemo odvisnost pojava od enega faktorja, moremo proučevati odvisnost pojava od več faktorjev hkrati. V tem primeru govorimo o *multipli regresiji* in *multipli korelaciji*.

Če proučujemo z multiplo regresijo, kako je x_1 odvisen od dveh faktorjev x_2, x_3, moremo pisati, da je

$$x_1 = f(x_2, x_3) + e \qquad (13.50)$$

Pri tem je $f(x_2, x_3)$ rezultat vpliva x_2 in x_3 na x_1, e pa je rezultat drugih, individualnih faktorjev na x_1.

Namesto regresijske krivulje pri enostavni regresiji, imamo pri multipli regresiji regresijsko ploskev, ki je podana z enačbo

$$x_1' = f(x_2, x_3) \tag{13.51}$$

Enako kot pri enostavni regresiji, vzamemo tudi pri multipli regresiji, da je po metodi najmanjših kvadratov regresijska funkcija tista, za katero velja, da *je vsota kvadratov odklonov stvarnih* vrednosti x_1 od ustreznih vrednosti x_1' najmanjša.

Enako merimo jakost povezave z *indeksom multiple korelacije*

$$I_{1.23} = \sqrt{1 - \frac{\sigma_{1.23}^2}{\sigma_1^2}} \tag{13.52}$$

pri tem pomeni: σ_1^2 = skupna varianca za znak x_1 ; $\sigma_{1.23}^2$ = nepojasnjeni del skupne variance za znak x_1 .

Podobno kot pri enostavni regresiji izračunamo tudi za multiplo regresijo *standardno napako ocene* po obrazcu

$$\sigma_{1.23} = \sigma_1 \sqrt{1 - I_{1.23}^2} \tag{13.53}$$

Linearna multipla korelacije

13.50 Kot pri enostavni korelaciji tudi pri multipli korelaciji najpogosteje proučujemo linearne zveze in odvisnosti.

Za linearno multiplo regresijo za odvisnost znaka x_1 od dveh znakov x_2 in x_3 ima regresijska funkcija, ki je enačba ravnine, obliko:

$$x_1' = \bar{x}_1 + b_{12.3}(x_2 - \bar{x}_2) + b_{13.2}(x_3 - \bar{x}_3) \tag{13.54}$$

Pri tem pomenijo: $\bar{x}_1$, $\bar{x}_2$ in $\bar{x}_3$ povprečja za ustrezne znake, $b_{12.3}$ in $b_{13.2}$ pa *multiple regresijske koeficiente*. Multiple regresijske koeficiente za ta primer izračunamo iz standardnih odklonov σ_1 , σ_2 , σ_3 in ustreznih enostavnih korelacijskih koeficientov r_{12} , r_{13} in r_{23} po obrazcih

$$b_{12.3} = \frac{\sigma_1}{\sigma_2} \cdot \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \; ; \quad b_{13.2} = \frac{\sigma_1}{\sigma_3} \cdot \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \tag{13.55}$$

Determinacijski koeficient multiple korelacije $R_{1.23}^2$, ki pokaže, kolik del od skupne variance za znak x_1 je pojasnjene z linearno odvisnostjo z znakom x_2 in x_3 , pa izračunamo iz r_{12} , r_{13} , r_{23} po obrazcu

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}} \qquad (13.53)$$

Kvadratni koren iz determinacijskega koeficienta je *multipli korelacijski koeficient* $R_{1.23}$.

Standardno napako ocene $\sigma_{1.23}$ pa izračunamo po obrazcu

$$\sigma_{1.23} = \sigma_1 \sqrt{1 - R^2_{1.23}} \qquad (13.57)$$

Če primerjamo vse zgornje pojme s pojmi za enostavno linearno regresijo in korelacijo, opazimo, da se posamezni pojmi ostali isti, le da so razširjeni na več znakov.

Zgornjo teorijo moremo, če je potrebno, razširiti na večje število znakov. Vendar to presega naš okvir.

13.51 Vzemimo za primer multiple korelacije odvisnost produktivnosti dela (x_1) v 27 podjetjih s pod 1000 zaposlenimi v tekstilni industriji v Sloveniji v letu 1957, od števila zaposlenih (x_2) in porabe električne energije (x_3).

Če merimo produktivnost dela z vrednostjo proizvodnje v letu 1957 v tisočih dinarjev na enega zaposlenega, porabo električne energije pa v MWh, iz osnovnih podatkov o proizvodnosti dela, številu zaposlenih in porabi električne energije za posamezna podjetja dobimo tele delne rezultate:

$\bar{x}_1 = 2910$ tisoč din ; $\bar{x}_2 = 345{,}78$ zaposlenih; $\bar{x}_3 = 900{,}70$ MWh

$\sigma_1 = 813{,}0$; $\sigma_2 = 244{,}9$; $\sigma_3 = 1231{,}5$

$r_{12} = +0{,}03383$; $r_{13} = +0{,}34802$; $r_{23} = 0{,}72000$

Iz teh podatkov dobimo po obrazcih 13.55

$$b_{123} = \frac{813{,}0}{244{,}9} \cdot \frac{0{,}03383 - 0{,}34802 \cdot 0{,}72000}{1 - 0{,}72000^2} = -1{,}49404$$

$$b_{132} = \frac{813{,}0}{1231{,}5} \cdot \frac{0{,}34802 - 0{,}03383 \cdot 0{,}72000}{1 - 0{,}72000^2} = +0{,}44367$$

Če vstavimo zgornje podatke v obrazec 13.54 dobimo naslednjo funkcijo

$$x'_1 = 2910 - 1{,}49404(x_2 - 345{,}78) + 0{,}44367(x_3 - 900{,}70)$$

ali, če jo dalje razvijemo

$$x_1' = 3027 - 1,49404x_2 + 0,44387x_3$$

Ta regresijska funkcija pokaže, kako je produktivnost dela x_1 linearno odvisna od števila zaposlenih x_2 in porabe električne energije x_3.

Vendar je ta povezava zelo rahla. Če namreč izračunamo po obrazcu 13.53 determinacijski koeficient multiple korelacije $R^2_{1.23}$, dobimo

$$R^2_{1.23} = \frac{0,08383^2 + 0,34802^2 - 2.0,08383.0,34802.0,72000}{1 - 0,72000^2} = 0,21866$$

Iz tega rezultata sklepamo, da je samo približno 21 % od skupne variance v produktivnosti dela pojasnjene z odvisnostjo produktivnosti dela od števila delavstva in porabljene električne energije. Iz determinacijskega koeficienta $R^2_{1.23}$ dobimo, da je koeficient multiple korelacije enak

$$R_{1.23} = \sqrt{0,21866} = 0,458$$

Temu ustrezno je velika tudi standardna pogreška ocene $\sigma_{1.23}$.

Po obrazcu 13.57 namreč dobimo, da je

$$\sigma_{1.23} = 813,0\sqrt{1-0,21866} = 719 \text{ tisoč dinarjev.}$$

Standardna pogreška ocene je tako velika, da iz dobljene regresijske funkcije ne moremo ocenjevati produktivnost dela, če poznamo za podjetje število zaposlenih in porabljeno električno energijo.

Praviloma $R_{1.23}$ ni manjši kakor r_{12} ali r_{13}. To je razumljivo, ker je pojav, ki je v odvisnosti z dvema znakoma, z obema vsaj toliko določen kakor z vsakim izmed njiju.

Štirinajsto poglavje

NORMALNA DISTRIBUCIJA

Pomen proučevanja teoretičnih distribucij

14.1 V dosedanjih poglavjih smo opisovali in analizirali predvsem stvarne distribucije, ki so slika opazovanih populacij.

Vendar smo na nekaj mestih nakazali, da imamo razen teh stvarnih frekvenčnih distribucij tudi frekvenčne distribucije, ki jih dobimo s teoretičnim razglabljanjem. Ena izmed njih je normalna distribucija. Zanjo smo že nakazali nekaj lastnosti v zvezi s standardnim odklonom. Razen normalne distribucije imamo tudi druge teoretične distribucije. Od teh bomo obravnavali: binomialno, Poissonovo, hipergeometrično, t-distribucijo, χ^2-distribucijo in F-distribucijo. Do teoretičnih distribucij pridemo matematično, če študiramo distribucije nekaterih količin, ki ustrezajo določenim pogojem.

Proučevanje teoretičnih distribucij je pomembno z več vidikov:

a) Če stvarna populacija zadošča pogojem, ki jim ustreza določena teoretična distribucija, je stvarna frekvenčna distribucija podobna, če že ne enaka teoretični distribuciji. Če primerjamo stvarno frekvenčno distribucijo s teoretično, ki jo dobimo v določenih pogojih, moremo sklepati, ali stvarna distribucija zadošča tem pogojem ali ne. Če na primer na določen pojav vplivajo samo slučajni vplivi, variacija pa v nobeno smer ni okrnjena, se vrednosti te populacije distribuirajo v normalni distribuciji. Če sta stvarna frekvenčna distribucija za določen pojav in ustrezna teoretična - normalna distribucija skladni, sklepamo, da so na populacijo vplivali samo slučajni vplivi brez omejitve v variiranju.

b) Včasih ne poznamo stvarne frekvenčne distribucije. Poznamo pa za populacijo določene parametre. Če predpostavljamo, da vrednosti populacije zadoščajo danim pogojem, moremo teoretično ugotoviti, kakšna je približno stvarna distribucija. Če n.pr. za proizvodnjo določenih profilov predpostavimo, da so odkloni od povprečne proizvodnje za posamezne artikle slučajni, poznamo pa aritmetično sredino in standardni odklon za to proizvodnjo, moremo določiti, kako se distribuirajo individualne vrednosti profilov, ne da bi poznali individualne vrednosti populacije.

c) Teoretične frekvenčne distribucije pa so posebno važne v vzorčenju, ki je, kakor smo že omenili, najpomembnejša metoda, s katero ocenjujemo in sklepamo v statistiki. Če vzorec zadošča osnovnim pogojem, se posamezne količine, ki nastopajo v vzorčenju, distribuirajo v določenih te-

oretičnih distribucijah. Razen za nepomembno majhne populacije teh ne moremo neposredno sestaviti. Z upoštevanjem postavk verjetnostnega računa pa moremo distribucije nekaterih od teh količin razviti teoretično. Tako n.pr. s teoretičnim študijem odkrijemo zakon o velikih številih, ki smo ga uvodoma nakazali in ilustrirali s praktičnim primerom.

V vzorčenju veljajo zakonitosti o distribuiranju določenih količin le, če vzorec zadošča osnovnim pogojem. Teorija vzorčenja je sestavljena pod predpostavko, da so enote, ki jih izbiramo v vzorec, izbrane slučajnostno. To pomeni, da mora imeti vsaka enota populacije enako možnost, da jo izberemo v vzorec. Pod to predpostavko je izdelana teoretična osnova vzorčenja. Jasno je, da teorija velja v praktičnem primeru le, če je ta osnovni pogoj izpolnjen. Zato se v vzorčenju na vso moč trudimo, da zadostimo temu osnovnemu pogoju.

Teoretične distribucije proučujemo torej s treh vidikov. Z njimi ugotavljamo, ali stvarne distribucije zadoščajo določenim predpostavkam. Če poznamo zakonitost variiranja v populaciji moremo določiti stvarno distribucijo. Teoretične distribucije pa so tudi osnova pomembne metode ocenjevanja - vzorčenja.

NORMALNA DISTRIBUCIJA

14.2 Normalna distribucija je ena izmed najpomembnejših teoretičnih distribucij. V proučevanjih množičnih pojavov imamo veliko stvarnih populacij, ki vsaj v grobem zadoščajo pogojem, v katerih se določena količina distribuira normalno. Razen tega pa je normalna distribucija osnovna distribucija v vzorčenju. Mnogo količin, ki jih sestavljamo v vzorčenju, se distribuira teoretično v normalni distribuciji. Razen tega je normalna populacija predpostavka za niz zakonitosti za vzorce. Normalna distribucija je tudi osnova za druge teoretične distribucije, kot so: t-distribucija, χ^2-distribucija, F-distribucija. V normalno distribucijo preide v limiti večina teoretičnih distribucij, med njimi tudi: binomialna, Poissonova, hipergeometrična, t-distribucija, χ^2-distribucija, F-distribucija.

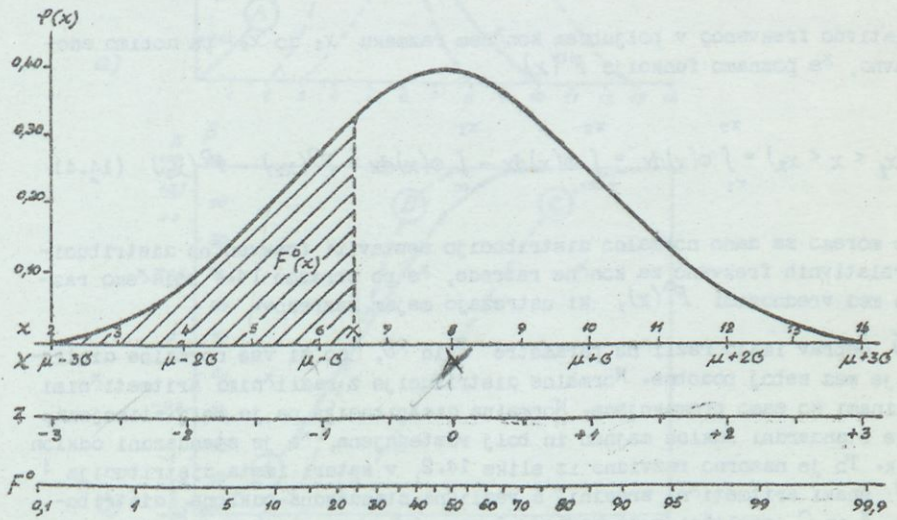
Opis normalne distribucije

14.3 Normalna distribucija je določena z dvema parametroma: z aritmetično sredino M, ki je rezultat splošnih vplivov in standardnim odklonom σ, ki je rezultat individualnih oziroma slučajnih vplivov. Normalna distribucija je dana v funkcijski obliki

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} \qquad (14.1)$$

Ta enačba pokaže, v kakšni odvisnosti je relativna gostota frekvence $\varphi(x)$ z vrednostjo x.

Če vzamemo vrednost x za absciso, ustrezno vrednost relativne gostote frekvence $\varphi(x)$ pa za ordinato, dobimo normalno krivuljo, ki je slika normalne distribucije (glej sliko 14.1). Iz te slike spoznamo, da je normalna distribucija unimodalna, simetrična in zvonasta distribucija, da je zato zanjo $M = Me = Mo$. Največja gostota frekvence je v točki $x = M$, gostota pa vztrajno pada, čim bolj je x oddaljen od aritmetične sredine. Normalna krivulja se abscisni osi asimptotično približuje, če x raste čez vse meje.



Slika 14.1 Normalna distribucija za $M = 8$ in $SD = 2$

14.4 Če za normalno distribucijo, ki je dana s funkcijo v obrazcu 14.1 izračunamo določen integral

$$\int_{-\infty}^{x} \varphi(x)dx = F^{\circ}(x) \qquad (14.2)$$

pomeni $F^{\circ}(x)$ kumulativno relativno frekvenco v normalni distribuciji v razmaku od $-\infty$ do X.

$F^{\circ}(x)$ je funkcija zgornje meje v razmaku od $-\infty$ do X in pomeni grafično ploščino, ki je v sliki 14.1 omejena z delom krivulje nad razmakom od $-\infty$ do X, z ustreznim delom razmaka na abscisni osi in vrednosti x ustrezno ordinato za normalno krivuljo. (Glej sliko 14.1). Če narišemo kumulativo relativne frekvence za normalno distribucijo v odvisno-

sti od x, dobimo značilno S krivuljo, ki smo jo srečali že pri stvarnih frekvenčnih distribucijah. S-krivuljo za normalno distribucijo vzamemo za idealno.

V sliki 14.2 so narisane tri normalne distribucije, pod njo pa ustrezne kumulativne distribucije relativnih frekvenc.

Iz pomena relativne frekvence, ki jo poznamo že iz poglavja o frekvenčnih distribucijah, sklepamo dalje, da je

$$F^0(\infty) = \int_{-\infty}^{+\infty} \varphi(x)dx = 1 \qquad (14.3)$$

Relativno frekvenco v poljubnem končnem razmaku x_1 do x_2 pa dobimo enostavno, če poznamo funkcijo $F^0(x)$

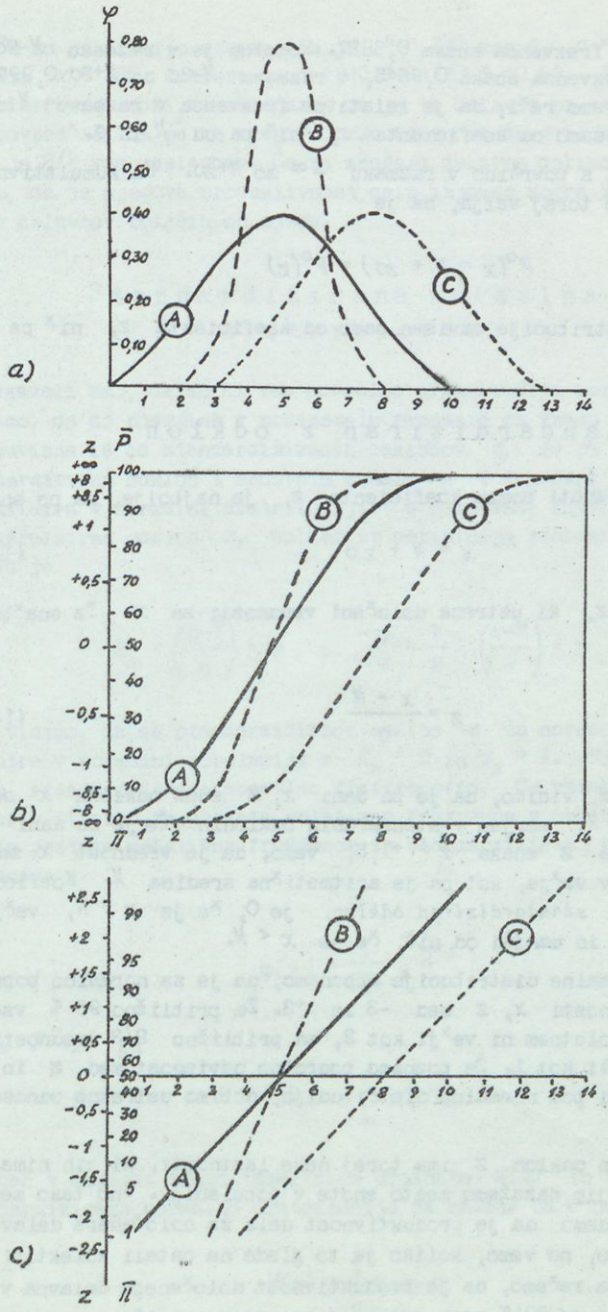
$$f^0(x_1 < x < x_2) = \int_{x_1}^{x_2} \varphi(x)dx = \int_{-\infty}^{x_2} \varphi(x)dx - \int_{-\infty}^{x_1} \varphi(x)dx = F^0(x_2) - F^0(x_1) \qquad (14.4)$$

Tako moremo za dano normalno distribucijo sestaviti frekvenčno distribucijo relativnih frekvenc za končne razrede, če po obrazcu 14.4 poiščemo razlike med vrednostmi $F^0(x)$, ki ustrezajo mejam razredov.

14.5 Čeprav imajo različne parametre M in SD, so si vse normalne distribucije med seboj podobne. Normalne distribucije z različnimi aritmetičnimi sredinami so samo premaknjene. Normalna distribucija pa je bolj stisnjena, če je standardni odklon majhen in bolj raztegnjena, če je standardni odklon velik. To je nazorno razvidno iz slike 14.2, v kateri imata distribucija A in B enaki aritmetični sredini, a različna standardna odklona, distribuciji A in C različni aritmetični sredini in enaka standardna odklona, distribuciji B in C pa različni aritmetični sredini in različna standardna odklona. Pod njima so prikazane ustrezne kumulative relativnih frekvenc za krivulje A, B in C. Iz slike 14.2 vidimo, da sta S-krivulji za distribuciji, ki imata enak standardni odklon, a različni aritmetični sredini, enake oblike, le da sta premaknjeni vzdolž abscisne osi. S-krivulja za normalno distribucijo, za katero je standardni odklon manjši, pa je strmejša.

Za vsak poseben primer normalne distribucije bi bilo zelo neprikladno neposredno iskati različne količine, kot so ordinate za normalno distribucijo $\varphi(x)$, kumulativne relativne frekvence $F^0(x)$ ita. Ker so si vse normalne distribucije podobne, velja zanje nekaj splošnih lastnosti, ki problem poenostavijo.

Če pogledamo normalno distribucijo v sliki 14.1, sklepamo, da je $F(x=M) = 0,50$ za vsako distribucijo, ne glede na to, kolik je M. V vsaki normalni distribuciji je zaradi simetrije polovica vrednosti pod, polovica pa nad aritmetično sredino M. Opazimo celo več. Za vsako normalno distribucijo, ne glede na to, kakšen je M in σ, je v razmaku od $M-\sigma$



Slika 14.2 *Različne normalne distribucije in njihove kumulative*

do $M+\sigma$ relativna frekvenca enaka 0,3827. Podobno je v razmaku $M-2\sigma$ do $M+2\sigma$ relativna frekvenca enaka 0,9545, v razmaku $M-3\sigma$ do $M+3\sigma$ 0,9973 ita. V splošnem moremo reči, da je relativna frekvenca v razmaku $M - z\sigma$ do $M + z\sigma$ odvisna samo od koeficienta z, nič pa od M in σ.

Enako je tudi s površino v razmaku $-\infty$ do $M+z\sigma$. Za kumulativno relativno frekvenco torej velja, da je

$$F^0(x = M + z\sigma) = F^0(z)$$

za vse normalne distribucije odvisen samo od koeficienta z, nič pa od M in σ.

Standardiziran odklon

14.6 Če hočemo spoznati pomen koeficienta z, je najbolje, da po enačbi

$$x = M + z\sigma \tag{14.5}$$

poiščemo vrednost z, ki ustreza določeni vrednosti za x. Iz enačbe 14.5 dobimo, da je

$$z = \frac{x - M}{\sigma} \tag{14.6}$$

Če proučimo ta izraz, vidimo, da je za dani x, z enak odklonu x od aritmetične sredine M, merjen v standardnih odklonih σ. Če je za dani x ustrezna vrednost za z enaka $z = +1,5$, vemo, da je vrednost x za 1,5 standardnih odklonov večja, kot pa je aritmetična sredina M. Koeficient z, ki ga imenujemo *standardiziran odklon*, je 0, če je $x = M$, večji od nič, če je $x > M$, in manjši od nič, če je $x < M$.

Iz poteka normalne distribucije spoznamo, da je za normalno populacijo za večino vrednosti x, z med -3 in $+3$. Za približno 95 % vseh vrednosti z v absolutnem ni večji kot 2, za približno 2/3 vrednosti pa je z absolutno manjši kot 1. Če poznamo podrobno odvisnost med z in ustreznimi površinami pod normalno distribucijo, dobimo ustrezne odnose za vsako vrednost z.

Standardiziran odklon z ima torej neke lastnosti, ki jih nima osnovni podatek. Z njim nakažemo mesto enote v populaciji. Eno tako merilo so centili. Če navedemo da je produktivnost dela za določenega delavca $x = 390$ kosov na uro, ne vemo, koliko je to glede na ostali kolektiv: veliko ali malo. Če pa rečemo, da je produktivnost določenega delavca v 80. centilu, pa vemo, da je 80 % vseh delavcev kolektiva slabših od njega, 20 % pa boljših. Podobno ta odnos izrazimo s standardiziranim odklonom z. Predpostavljajmo, da je produktivnost dela v kolektivu distribuirana nor-

malno. Za to populacijo vzemimo da je M = 340 kosov, σ = 30. Produktivnosti dela x = 390 ustreza z = (390-340)/30 = +0,88. Iz dobljenega standardiziranega odklona sklepamo, da je ta delavec po svoji produktivnosti nadpovprečen, da pa se njegova produktivnost še giblje v razmaku, v katerem je 2/3 vseh delavcev. Če za drugega delavca dobimo z_2 = +2,5, sklepamo, da je njegova produktivnost dela izjemno dobra in je samo nekaj promilov delavcev boljših od njega.

Standardizirana normalna distribucija

14.7 Nakazali smo, da so si vse normalne distribucije med seboj podobne. Videli smo, da so površine v posameznih razmakih za vsako normalno distribucijo odvisne le od standardiziranih odklonov z, ne pa od M in σ. Ker je standardiziran odklon z osnovnim znakom x v linearni zvezi, se tudi z distribuira v normalni distribuciji. Če poiščemo, kakšna sta M_z in SD_z za standardiziran odklon z, dobimo iz definicije standardiziranega odklona, da je

$$M_z = \overline{\left(\frac{x-M}{\sigma}\right)} = 0 \quad ; \quad \sigma_z^2 = \frac{1}{N}\Sigma\left(\frac{x-M}{\sigma}\right)^2 = 1 \qquad (14.7)$$

Iz tega vidimo, da se standardizirani odklon z za normalno populacijo distribuira v normalni populaciji z M_z = 0 in σ_z = 1. To distribucijo imenujemo *standardizirano normalno distribucijo*. Če vstavimo te posebne vrednosti v obrazec za gostoto relativne frekvence v obrazcu 14.1, dobimo, da je gostota relativne frekvence za standardizirano normalno distribucijo enaka

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad (14.8)$$

površine $F^0(z)$ pa

$$F^0(z) = \int_{-\infty}^{z} \varphi(z)dz \qquad (14.9)$$

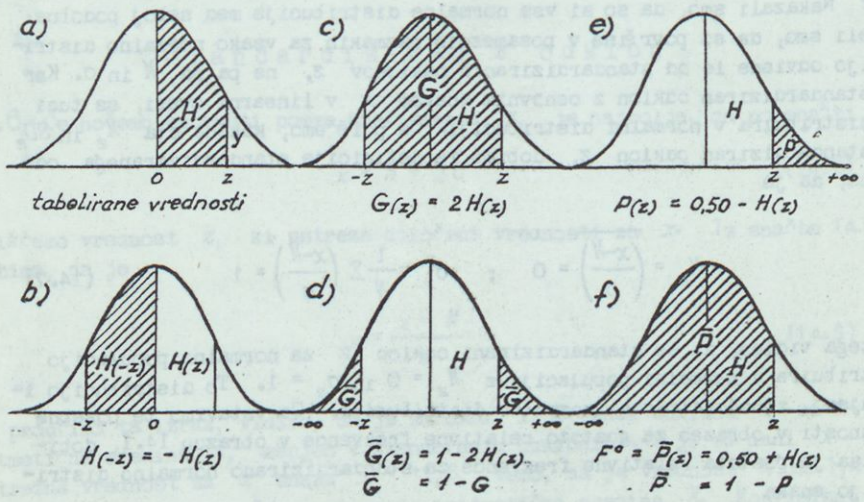
V tablicah v dodatku imamo tabelirane ordinate $\varphi(z)$ in površine $H(z)$ za standardizirano normalno distribucijo za razmak od 0 do z.

$$H(z) = \int_{0}^{z} \varphi(z)dz \qquad (14.10)$$

V tablicah so dane ordinate za pozitivne vrednosti za z. Ker je normalna distribucija simetrična, velja

$$H(z) = -H(-z) \qquad (14.11)$$

Ker za normalno distribucijo potrebujemo večkrat različno opredeljene površine, tabelirane pa imamo samo površine H v razmaku od 0 do z, so v sliki 14.3 nakazani odnosi med posameznimi površinami, ki pridejo v praksi v poštev v različnih primerih.



Slika 14.3 Različne površine pod standardizirano normalno distribucijo

V tablici ordinat $\varphi(z)$ in površin H za standardizirano normalno distribucijo so zaradi nazornosti standardizirani odkloni z pomnoženi s 100, ordinate $\varphi(z)$ in površine H pa z 10000.

14.8 Tablice za standardizirano normalno distribucijo uporabljamo za vse normalne distribucije. Gostoto absolutne frekvence $g(x)$ za stvarno normalno frekvenčno distribucijo, ki ima obseg N, aritmetično sredino M in standardni odklon σ izračunamo po obrazcu

$$g(x) = N\,\varphi(x) = N\,\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{(x-M)^2}{2\sigma^2}} = \frac{N}{\sigma}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{z^2}{2}} = \frac{N}{\sigma}\,\varphi(z) \qquad (14.12)$$

tako, da poiščemo najprej z, ki ustreza danemu x. Nato v tablicah za

standardizirano normalno distribucijo poiščemo ustrezno gostoto relativne frekvence $\varphi(z)$. To pa pomnožimo s kvocientom N/σ.

Vzemimo za primer normalno distribucijo o produktivnosti dela za kolektiv z $N = 400$ delavci. Povprečna produktivnost dela $M = 340$ kosov na uro, $SD = 30$ kosov. Ordinato za $x = 390$ dobimo po nakazanem postopku. Najprej izračunamo $x = 390$ ustrezajoči z.

$$z = \frac{390-340}{60} = 0{,}83$$

V tablicah za normalno distribucijo dobimo, da je ustrezna ordinata za standardizirano distribucijo $\varphi(z) = 0{,}2827$. Ordinata na stvarni distribuciji pa je po obrazcu 14.12 enaka

$$y(390) = \frac{400}{60} 0{,}2827 = 1{,}89$$

Absolutne frekvence za normalno populacijo z obsegom N pa dobimo, tako, da ustrezne relativne frekvence iz tablice za standardizirano distribucijo pomnožimo z obsegom populacije N.

Predpostavljamo, da je distribucija o produktivnosti dela normalna. Da ugotovimo, koliko delavcev je preseglo produktivnost dela $x = 390$ kosov na uro, poiščemo z ustrezno vrednost

$$P(z) = 0{,}50 - H(z) = 0{,}50 - 0{,}2937 = 0{,}2033$$

Če to relativno frekvenco pomnožimo z obsegom populacije $N = 400$, dobimo $400 \cdot 0{,}2033 = 81{,}32$ delavcev.

Če veljajo zgornje predpostavke, produktivnost dela x 390 kosov na uro, presega 81 delavcev.

14.9 Čeprav imamo v dodatku podrobno tabelo površin, navajamo v tabeli 14.1 nekaj najznačilnejših odnosov med standardiziranim odklonom z in ustreznimi površinami H, G, $\bar{G}$, P in $\bar{P}$.

14.10 Če predpostavljamo normalnost populacij, moremo s tablicami za normalno distribucijo reševati različne probleme. Za ilustracijo vzemimo tole nalogo. Stroj izdeluje določene profile. Karakteristika stroja je, da izdeluje profile s povprečnim premerom $M = 30$ mm in standardnim odklonom $\sigma = 0{,}2$ mm. Artikli pa so uporabni, če je premer profila med $x_s = 29{,}7$ in $x_z = 30{,}4$. Ugotoviti je treba, kolik je odstotek izmeta na tem stroju in koliko moremo proizvodnjo popraviti, če moremo aritmetično sredino M regulirati in naravnati na optimalno stanje.

Tabela 14.1 Odnosi med standardiziranim odklonom z, in ustreznimi relativnimi frekvencami H, G, $\bar{G}$, P in $\bar{P}$

z	v odstotkih				
	H	G	$\bar{G}$	P	$\bar{P} = F^0$
0,674	25,00	50,00	50,00	25,00	75,00
1,000	34,13	68,26	31,74	15,87	84,13
1,645	45,00	90,00	10,00	5,00	95,00
1,960	47,50	95,00	5,00	2,50	97,50
2,000	47,73	95,46	4,54	2,27	97,73
2,326	49,00	98,00	2,00	1,00	99,00
2,576	49,50	99,00	1,00	0,50	99,50
3,000	49,87	99,74	0,26	0,13	99,87
3,090	49,90	99,80	0,20	0,10	99,90
3,291	49,95	99,90	0,10	0,05	99,95

Standardizirana odklona za meji uporabnosti sta:

$$z_s = \frac{x_s - M}{\sigma} = \frac{29,7-30,0}{0,2} = -1,5 \; ; \quad z_z = \frac{x_z - M}{\sigma} = \frac{30,4-30,0}{0,2} = +2,0$$

Če poiščemo v tabeli za normalno distribucijo tem standardiziranim odklonom ustrezne relativne frekvence, dobimo:

$H(-1,5) = -H(1,5) = -0,4332$; $H(2,0) = 0,4773$. Pod $z = -1,5$ je $0,5 - 0,4332 = 0,0668$, nad $z = +2,0$ pa $0,5-0,4773 = 0,0227$ relativne frekvence. Skupen odstotek izmeta je torej $0,0668 + 0,0227 = 0,0895$ ali ca 9 odstotkov.

Kakor vidimo, povprečje za profil, ki ga proizvajamo, ni sredi razmaka, v katerem so artikli uporabni; to pa bi bilo glede na obliko normalne distribucije najboljše. Ker moramo spreminjati povprečje profila, ga naravnajmo na sredino razmaka. Tako dobimo novo povprečje $M_1 = \frac{29,7+30,4}{2} = 30,05$ mm.

Glede na predpis je torej artikel še uporaben, če se od novega povprečja ne odklanja za več kot 0,35 mm. Temu odklonu ustrezni standardiziran odklon je

$$z = \frac{0,35}{0,2} = 1,75$$

Zanj je po tablici $H(1,75) = 0,4599$. Ker je $\bar{G} = 1 - 2H =$ $= 1,0 - 2.0,4599 = 0,0802$, vidimo, da smo uspeli izmet skrčiti za en odstotek, ker znaša novi odstotek 8,0 %.

Prilagoditev normalne distribucije stvarni frekvenčni distribuciji

14.11 Kakor smo že omenili v uvodu, s teoretičnimi distribucijami rešujemo tudi tale problem. Imamo stvarno distribucijo, za katero predpostavljamo, da je nastala pod določenimi pogoji. Vzemimo, da vemo, kakšna bi morala biti ta distribucija teoretično, če bi ustrezala tem pogojem. Če primerjamo obe distribuciji med seboj, nadalje ugotovimo, ali je stvarna distribucija taka, da moremo sklepati, če zadošča tem pogojem ali ne.

Pri reševanju takih in podobnih problemov stvarni distribuciji večkrat prilagodimo normalno, ki ima iste parametre kakor stvarna distribucija: torej isto aritmetično sredino M in isti standardni odklon σ. Ker imamo tabelirane vrednosti za standardizirano normalno distribucijo, si pri takih nalogah pomagamo z njo. V splošnem moremo prilagoditi normalno distribucijo z dvema postopkoma: z metodo ordinat in metodo površin. Obravnavali pa bomo le metodo površin. Normalno distribucijo po metodi površin prilagodimo takole:

a) Za stvarno frekvenčno distribucijo, ki ji hočemo prilagoditi normalno distribucijo, izračunamo aritmetično sredino M in standardni odklon σ.

b) Za meje razredov izračunamo standardizirane odklone z_k / σ.

c) Za standardizirane odklone za meje razredov poiščemo ustrezne vrednosti kumulativnih relativnih frekvenc F^0 za normalno distribucijo. Te dobimo iz tablice za standardizirano normalno distribucijo v tabeli v dodatku. Ker imamo tabelirane H, ne pa F^0, izračunamo kumulativne relativne frekvence po znanem obrazcu $F^0 = 0,5 + H$. Pri tem moramo paziti na negativne predznake pri z in H.

d) Iz kumulativ relativnih frekvenc F^0 dobimo kumulativo absolutnih frekvenc F', če posamezne člene kumulativne vrste pomnožimo z obsegom populacije N.

e) Absolutne frekvence za prilagojeno normalno distribucijo f' dobimo, če poiščemo razlike med zaporednimi členi v kumulativni vrsti frekvenc za normalno distribucijo.

14.12 Za primer vzemimo frekvenčno distribucijo teže za $N = 531$ baterskih svitkov v predilnici ene izmed tekstilnih tovarn v Sloveniji v letu 1958. Delavci proizvedene baterske svitke sami kontrolirajo, teže posameznih baterskih svitkov pa vpisujejo v posebno obdelovalno tabelo. Za teh

N = 561 baterskih svitkov smo dobili, da je povprečna registrirana teža M = 18,8607 kg, standardni odklon σ = 0,07266. V tabeli 14.2 je prikazano, kako stvarni distribuciji za težo baterskih svitkov prilagodimo po metodi površin ustrezno normalno distribucijo, ki ima enako aritmetično sredino M in enak standardni odklon σ kot stvarna distribucija.

Tabela 14.2 Prilagoditev normalne distribucije po metodi površin za frekvenčno distribucijo registrirane teže baterskih svitkov (N = 561, M = 18,8607 kg; SD = 0,07266 kg)

Teža kg	f	z	F^o	F'	f'
18,525	3	-4,61	0,0000	0,0	–
18,575	–	-3,93	0,0000	0,0	0,3
18,625	14	-3,24	0,0006	0,3	2,6
18,675	11	-2,56	0,0052	2,9	14,3
18,725	23	-1,87	0,0307	17,2	49,6
18,775	71	-1,18	0,1190	66,8	108,3
18,825	229	-0,49	0,3121	175,1	149,9
18,875	129	+0,20	0,5793	325,0	129,7
18,925	50	+0,88	0,8106	454,7	73,6
18,975	11	+1,57	0,9418	528,3	26,0
19,025	18	+2,26	0,9881	554,3	5,8
19,075	2	+2,95	0,9984	560,1	0,8
19,125		+3,64	0,9999	560,9	0,1
	561				561,0

Standardiziran odklon za spodnjo mejo za prvi razred dobimo neposredno po obrazcu:

$$z = \frac{x-M}{\sigma} \quad \frac{18,5250-18,8607}{0,07266} = -4,620$$

Za druge meje razredov pa dobimo ustrezne z tako, da spodnji meji prvega razreda postopoma prištevamo širino razreda i, merjeno v standardnih odklonih σ. Torej: i/SD = 0,050/0,07266 = 0,688

Če je prva meja z_1 = -4,620; je druga: $z_2 = z_1 + i/SD$ = = -4,620 + 0,688 = -3,932; tretja $z_3 = z_2 + i/SD$ = -3,932 + 0,688 = = -3,244 itd.

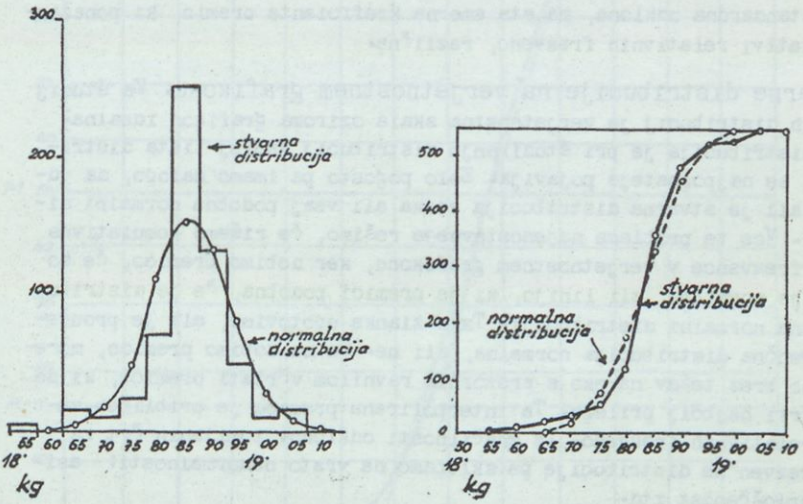
Relativno kumulativno vrsto izračunamo s tabelami površin za standardizirano normalno distribucijo. N.pr. za z = -1,87 dobimo:

$F^{0}(-1,87) = 0,50 + H(-1,87) = 0,50 - H(1,87) = 0,50 - 0,4393 = 0,0307$ ita.

Ko iz relativnih kumulativ dobimo absolutno kumulativo, če relativne kumulativne člene pomnožimo z $N = 561$, izračunamo razlike med zaporednimi členi kumulativ. Tako dobimo frekvence za prilagojeno distribucijo. N.pr.: $0,3 - 0,0 = 0,3$; $2,9 - 0,3 = 2,3$ ita.

Iz dobljenih rezultatov sklepamo, da se stvarna distribucija razmeroma slabo sklada s prilagojeno normalno distribucijo.

V sliki 14.4 a sta vrisani stvarna in prilagojena normalna distribucija, v sliki 14.4 b pa njeni kumulativi. Čeprav se obe distribuciji preveč ne skladata, dobro vidimo, da se kumulativa prilagojene distribucije vije med kumulativo za stvarno distribucijo.



a.) Frekvenčna distribucija b.) Kumulativna distribucija

Slika 14.4 Stvarna frekvenčna distribucija teze $N = 561$ baterskih svitkov in prilagojena normalna distribucija

Verjetnostni grafikoni

14.13 Verjetnostna skala. Kumulativa za normalno distribucijo ima značilno obliko črke S. Kumulative za večino unimodalnih distribucij imajo tudi obliko črke S, čeprav niso normalne. Oblika seveda ni tako pravilna kot pri normalni distribuciji, vendar je težko presoditi, ali ustreza normalni distribuciji ali ne. Če pa skalo na grafikonu kumulativ transformiramo, pri določeni transformaciji uspemo, da je na takem grafikonu kumulativa relativnih frekvenc za vsako normalno distribucijo premica. To dose-

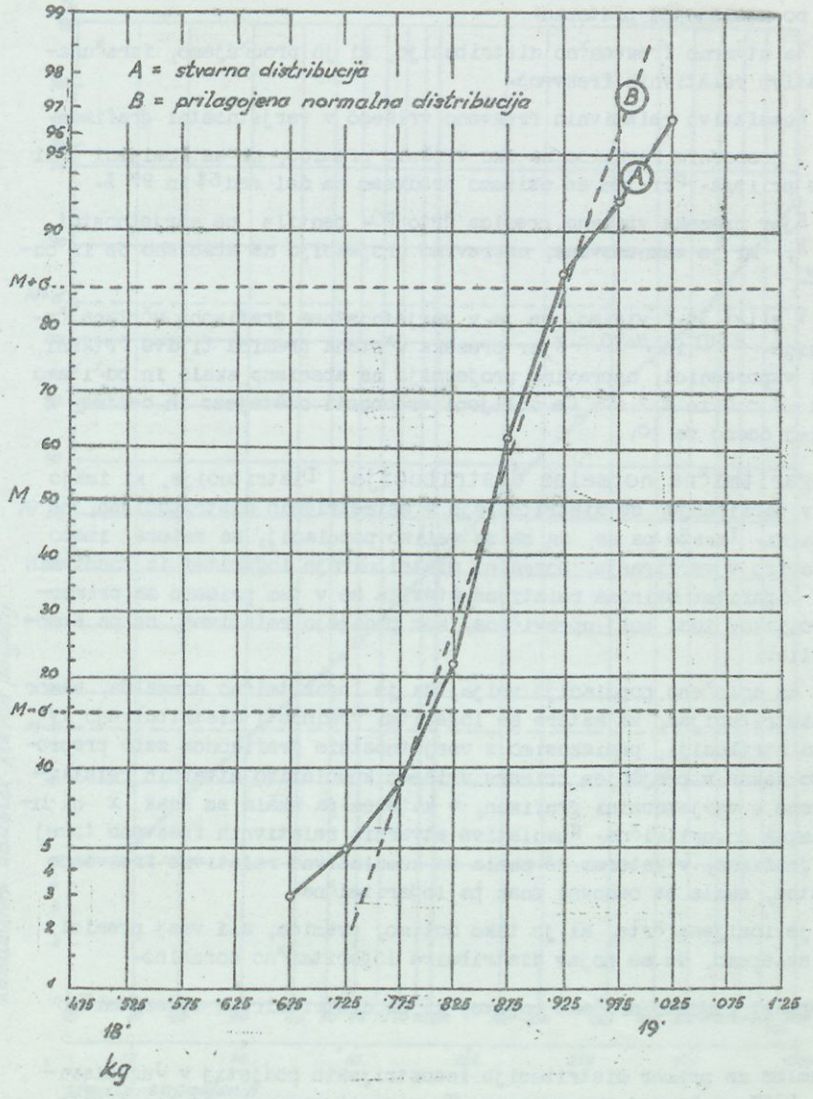
žemo, če namesto linearne skale relativnih frekvenc uvedemo tako imenovano *verjetnostno skalo relativnih frekvenc*. Če linearni z-skali priredimo ustrezno skalo kumulativnih relativnih frekvenc F°, dobimo verjetnostno skalo. Verjetnostna skala, ki jo dobimo po tem principu iz tablice za površine za standardizirano normalno distribucijo, je vrisana kot tretja skala v sliki 14.1, v kateri so prikazane skale za x, z in F°. Ta skala je v linearni zvezi s skalo za osnovni znak x in standardizirani znak z. Zato je v grafikonu, v katerem je abscisa skala za x, ordinata pa skala za kumulativno relativno frekvenco F°, za vsako normalno distribucijo kumulativa relativnih frekvenc premica. To nazorno vidimo v sliki 14.2. Tu so razen na linearni skali narisane kumulative relativnih frekvenc tudi na verjetnostni skali oziroma grafikonu. Iz primerjave distribucij sklepamo, da so premice za normalne distribucije, ki imajo različne aritmetične sredine, samo vzporedno premaknjene. Če pa imata distribuciji različna standardna odklona, pa sta smerna koeficienta premic, ki ponazarjata kumulativi relativnih frekvenc, različna.

14.14 Stvarne distribucije na verjetnostnem grafikonu. Za študij frekvenčnih distribucij je verjetnostna skala oziroma grafikon idealna. Normalna distribucija je pri študiranju distribucij vselej tista distribucija, ki se najpogosteje pojavlja. Zelo pogosto pa imamo nalogo, da ugotovimo, ali je stvarna distribucija enaka ali vsaj podobna normalni distribuciji. Vse te probleme najenostavneje rešimo, če rišemo kumulativne relativne frekvence v verjetnostnem grafikonu, ker dobimo premico, če so distribucije normalne, ali linijo, ki je premici podobna, če je distribucija podobna normalni distribuciji. Tako zlahka ugotovimo, ali je proučevana frekvenčna distribucija normalna, ali ne. Če ne dobimo premico, moremo grafično brez težav na oko s prozornim ravnilom včrtati premico, ki se lomljeni črti najbolj prilega. Ta interpolirana premica je približek kumulative relativnih frekvenc. Iz značilnosti odklonov lomljenih črt za stvarne frekvenčne distribucije pa sklepamo na vrsto nenormalnosti: asimetrijo, sploščenost ita.

V sliki 14.5 je na verjetnostnem grafikonu narisana kumulativa relativnih frekvenc za distribucijo baterskih svitkov po registrirani teži iz tabele 14.3. Lomljena črta A je slika za kumulativo relativnih frekvenc za stvarno distribucijo, premica B pa je slika kumulative za relativne frekvence za prilagojeno normalno distribucijo. Lomljena črta za ta primer ni premica, temveč ima še rahlo obliko črke S. Iz tega sklepamo, da je stvarna distribucija koničasta, če jo primerjamo z normalno.

Tabela 14.3 Vrsta kumulativnih relativnih frekvenc (KRF) za distribucijo registrirane teže za $N = 561$ baterskih svitkov iz tabele 14.2

Teža v kg	-18,575	-18,625	-18,675	-18,725	-18,775	-18,825
KRF	0,5	0,5	3,0	5,0	9,1	21,8
Teža v kg	-18,875	-18,925	-18,975	-19,025	-19,075	
KRF	62,6	85,7	94,6	99,7	100,0	



Slika 14.5 Slika KRF in prilagojene normalne distribucije za težo baterskih svitkov iz tabele 14.2 in 14.3

14.15 Ocenjevanje M in σ iz verjetnostnega grafikona. Ne da bi navedli primer, omenimo, da moremo z verjetnostnim grafikonom za stvarne distribucije grafično oceniti aritmetično sredino M in standardni odklon σ po naslednjem postopku:

a) Za stvarno frekvenčno distribucijo, ki jo proučujemo, izračunamo kumulativo relativnih frekvenc.

b) Kumulativo relativnih frekvenc vrišemo v verjetnostni grafikon.

c) S prozornim ravnilom na oko vrišemo premico, ki se lomljeni črti čim bolje prilega. Pri tem se oziramo predvsem na del med 5% in 95 %.

d) Kjer preseka vrisana premica črto 50. centila na verjetnostni skali z M, ki je zaznamovana, napravimo projekcijo na abscisno os in odberemo M.

e) V sliki 14.5 vidimo, da je v verjetnostnem grafikonu včrtana črtkana linija $M-SD$ in $M+SD$. Kjer preseka včrtana premica ti dve črtkani vodoravni vzporednici, napravimo projekcijo na abscisno skalo in odčitamo vrednosti $M - \sigma$ in $M + \sigma$. Če dobljeni vrednosti odštejemo in delimo z dva, dobimo oceno za σ.

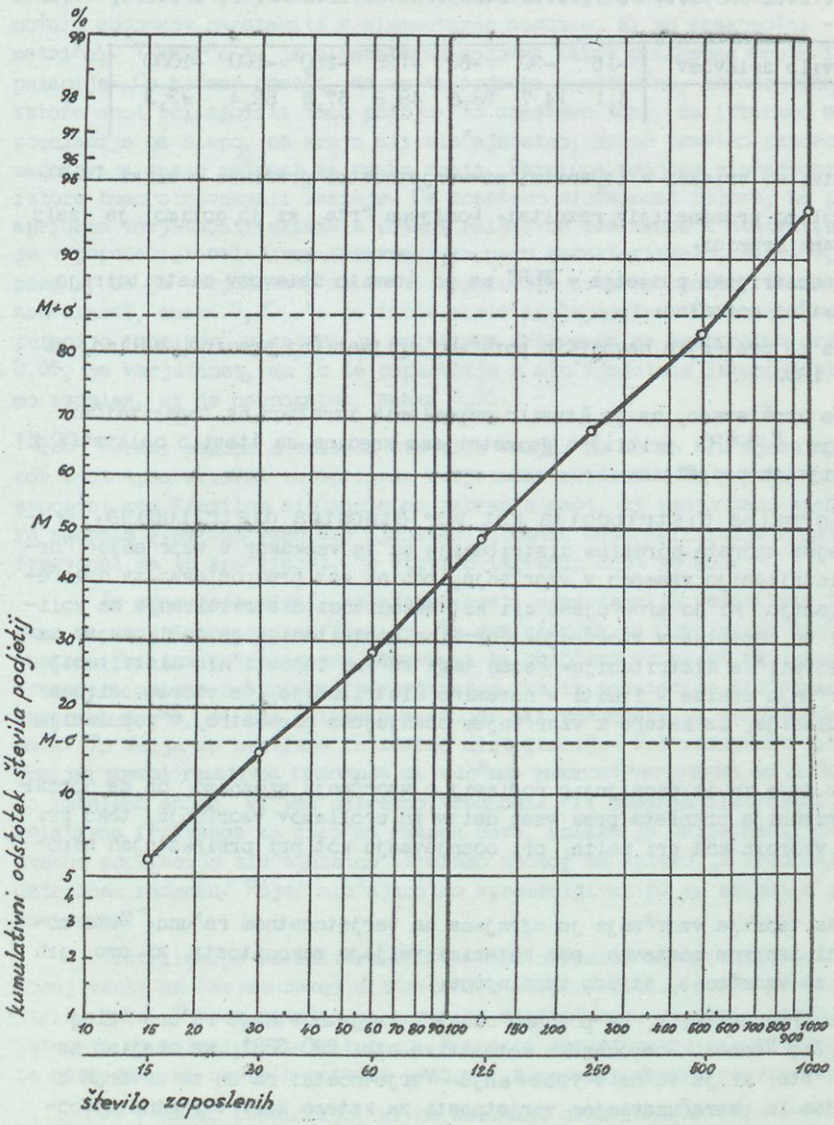
14.16 Logaritmična normalna distribucija. Distribucije, ki imajo omejitev v variiranju, se distribuirajo v asimetričnih distribucijah, ne pa v normalni. Izkaže pa se, da se za veliko populacij, za katere imamo določeno oviro v variiranju, normalno distribuirajo logaritmi iz osnovnih podatkov. Logaritmi oziroma relativna števila so v tem primeru za primerljivost podatkov tudi bolj upravičena, ker podajajo relativne, ne pa absolutne razlike.

Ali za določeno populacijo velja, da je logaritmično normalna, kakor pravimo distribucijam, za katere se logaritmi vrednosti distribuirajo v normalni distribuciji, preizkusimo z verjetnostnim grafikonom zelo preprosto. Enako kakor v prejšnjem primeru vrišemo kumulativo stvarnih relativnih frekvenc v verjetnostni grafikon, v katerem pa skala za znak x ni linearna, ampak logaritmična. Kumulative stvarnih relativnih frekvenc torej vrišemo v grafikon, v katerem je skala za kumulativne relativne frekvence verjetnostna, skala za osnovni znak pa logaritmična.

Če je lomljena črta, ki jo tako dobimo, premica, ali vsaj premici podobna, sklepamo, da se pojav distribuira logaritmično normalno.

14.17 V praksi pogosto srečamo pojave, ki se distribuirajo logaritmično normalno.

Vzemimo za primer distribucijo industrijskih podjetij v Jugoslaviji v letu 1957 po številu zaposlenih. To distribucijo smo že proučevali v poglavju o frekvenčnih distribucijah. Ugotovili smo, da je izrazita J-distribucija. Meje razredov v tej distribuciji so približno geometrijsko zaporedje, ker skušamo z njo pokazati relativne razlike med podjetji. V tabeli 14.4 imamo za to distribucijo kumulativo relativnih frekvenc, ki smo jo izračunali iz relativnih frekvenc v tabeli 7.3.



Slika 14.6 Frekvenčna distribucija industrijskih podjetij v FLRJ v letu 1957 po številu zaposlenih, narisana na logaritmično-verjetnostnem grafikonu

Tabela 14.4 Kumulativne relativne frekvence (KRF) za distribucijo industrijskih podjetij po številu zaposlenih v letu 1957. (Po tabeli 7.6)

Število delavcev	-15	-30	-60	-125	-250	-500	-1000
KRF	6,1	14,7	28,6	48,8	67,5	82,4	93,3

Ti podatki so vrisani v logaritmično-verjetnostni grafikon v sliki 14.3

Dobimo presenetljiv rezultat. Lomljena črta, ki jo dobimo, je zelo približana premici.

Industrijska podjetja v FLRJ se po številu delavcev distribuirajo logaritmično normalno.

Če po prejšnjih navodilih poiščemo aritmetično sredino, dobimo, da je $M = 180$.

Če upoštevamo, da je število zaposlenih izraženo na logaritmični skali, je $M = 180$ približno geometrijska sredina za število delavstva v industrijskih podjetjih.

14.18 Normalna distribucija kot verjetnostna distribucija. Najpomembnejša uporaba normalne distribucije pa je vsekakor v vzorčenju. Normalno distribucijo vzamemo v vzorčenju, bodisi kot predpostavko za določeno populacijo, ki jo proučujemo, ali kot zakonitost distribuiranja za količine, ki so pomembne v vzorčenju. Normalna distribucija je tudi osnova za druge teoretične distribucije. Razen tega večina teoretičnih distribucij pri vzorčenju preide v limiti v normalno distribucijo, če vzorec dovolj velik iz populacije, iz katere z vzorčenjem ocenjujemo parametre, v populacijah vsaj velik.

Iz tega in iz naslednjih poglavij o vzorčenju spoznamo, da se normalna distribucija prepleta prek vseh delov in problemov vzorčenja, tako pri velikih vzorcih kot pri malih, pri ocenjevanju kot pri preizkušnjah hipotez.

14.19 Vsa teorija vzorčenja je zgrajena na verjetnostnem računu. Zato morajo biti osnovne postavke, pod katerimi veljajo zakonitosti, ki smo jih odkrili za vzorčenje, strogo izpolnjene.

Če predpostavimo, da poznamo osnove verjetnostnega računa (glej dr. Alojzij Vadnal: Gospodarska matematika str. 280-288), se omejimo na osnovno črto, ki je važna v vzorčenju. Verjetnostni račun se ukvarja z računanjem in preračunavanjem verjetnosti za katere koli verjetne dogodke. Ti verjetni dogodki dobe v vzorčenju določeno obliko. V vzorčenju je dejanje, ki ga opredelimo kot poskus, izbiranje enot iz osnovne populacije. Dogodek v smislu verjetnostnega računa pa je vrednost določenega znaka ali značilnosti več znakov za izbrane enote. Pojem aposteriorne verjetnosti, ki je opredeljena kot limita kvocienta števila ugodnih dogodkov

in števila poskusov v enakih pogojih, če število poskusov večamo čez vse meje, predpostavlja, da so neki osnovni pogoji poskusov enaki. Pojem apriorne verjetnosti pa je zasnovan na predpostavki, da moremo skupnost vseh možnih dogodkov razstaviti v elementarne dogodke, ki so enakomožni - simetrični. V vzorčenju je elementaren dogodek izbor posamezne enote iz populacije. Če hočemo doseči, da so ti dogodki enakomožni, moramo tehniko izbora enot prilagoditi temu pogoju. To dosežemo tako, da izbiramo enote populacije na slepo, na srečo ali slučajnostno, kakor pravimo izboru, ki zagotavlja enako možnost za vsako enoto. Tehnične prijeme slučajnostnega izbora bomo obravnavali kasneje. Če dosežemo slučajnost izbora, se pojem apriorne verjetnosti sklada s pojmom relativne frekvence v statistiki. Če je v populaciji relativna frekvenca za neko karakteristiko, n.pr. 0,30, pomeni, da je verjetnost, da iz te populacije izberemo enoto, ki ima to značilnost, enaka 0,30, če je izbor enote slučajnosten. Če imamo n.pr. populacijo izdelkov, v kateri je relativna frekvenca za neuporabne artikle 0,05, je verjetnost, da iz te populacije s slučajnostnim izborom izberemo izdelek, ki je neuporaben, enaka 0,05.

14.20 Zaradi zakona o velikih številih velja, da se za slučajnostni izbor enot aposteriorna in apriorna verjetnost skladata. To pomeni, da se kvocient med številom slučajnostno izbranih enot, ki imajo dano značilnost, in skupnim številom vseh slučajnostno izbranih enot, približuje relativni frekvenci za to značilnost, če število izbranih enot večamo.

Iz zgornjega sledi, da moremo vzeti distribucijo relativnih frekvenc za verjetnostno distribucijo, če je izbor iz te populacije slučajnosten. Pri tem se relativna frekvenca za določeno vrednost ali razmak vrednosti sklada z apriorno verjetnostjo, da iz populacije slučajnostno izberemo enoto, ki ima ustrezno vrednost ali vrednosti iz ustreznega razmaka. To velja za nezvezne in zvezne distribucije. Pri nezveznih distribucijah pomeni relativna frekvenca za določeno vrednost verjetnost, da slučajnostno izberemo enoto, ki ima ustrezno vrednost. Pri zveznih distribucijah pa relativna frekvenca za določen razmak pove, kolika je verjetnost, da iz zvezne populacije slučajnostno izberemo enoto, za katero je vrednost v ustreznem razmaku. Pojem slučajnostne spremenljivke pa se sklada s pojmom znaka.

Distribucijo relativnih frekvenc za normalno distribucijo moremo torej vzeti za verjetnostno distribucijo za populacijo, ki je normalno distribuirana. Tabela površin za normalno distribucijo je obenem tabela verjetnosti za normalno distribucijo. Normalna verjetnostna distribucija je iz razlogov, ki smo jih navedli uvodoma, posebno pomembna za vzorčenje.

Če vzamemo populacijo, ki se normalno distribuira, je n.pr. verjetnost, da iz normalne populacije slučajnostno izberemo enoto, ki je v mejah $M - \sigma$ do $M + \sigma$, enaka relativni frekvenci za ta razmak. Ta pa je po tabeli 14.1 enaka 0,3823. Če slučajnostni izbor iz populacije ponavljamo, pričakujemo, da v povprečju približno 2/3 vseh slučajnostno izbranih vrednosti leži v zgornjem razmaku. Podobno dobimo ustrezne verjetnosti

za poljuben razmak.

14.21 Pojem tveganja. V vsakdanjem življenju vzamemo dogodke, ki so zelo verjetni, praktično tako, kakor da se bodo za gotovo zgodili. Zato računamo, da nas pri sprehodu ne bo povozil avto, ker je velika verjetnost, da nas resnično ne bo. Ko kupujemo konzervo, računamo, da ni pokvarjena, ker je zelo verjetno, da res ni ita. Pri vseh teh sklepih pa je določena stopnja rizika - tveganja. More se zgoditi, da nas povozi avto. More se zgoditi, da je konzerva pokvarjena, čeprav računamo, da ni. To tveganje je tem večje, čim večja je verjetnost, da se dogodek, ki ga pričakujemo, ne zgodi. Večje tveganje je, da konzerva iz lanskoletne proizvodnje ni pokvarjena, kakor da ni pokvarjena konzerva iz letošnje proizvodnje. Tveganje je torej verjetnost, da se dogodek, ki ga pričakujemo, ne zgodi.

Pojmu tveganja v vsakdanjem življenju popolnoma ustreza pojem tveganja v verjetnostnem računu oziroma statistiki.

Vrednosti standardiziranc normalne distribucije teoretično leže na vsem razmaku od $-\infty$ do $+\infty$. Vendar praktično vzamemo, da vse vrednosti leže v razmaku med $-3{,}29$ do $+3{,}29$. Če trdimo, da je vrednost enote, ki jo slučajnostno izberemo iz standardiziranc normalno distribuirane populacije v razmaku od $-3{,}29$ do $+3{,}29$, je tveganje te izjave 0,001 (glej tabelo 14.1).

To pomeni: da z verjetnostjo 0,001 moremo dobiti pri slučajnostnem izboru enoto, za katero je vrednost izven navedenega razmaka. To tveganje pa moremo tolmačiti tudi drugače. Glede na aposteriorno verjetnost moremo sklepati: Če hoče določen raziskovalec preiskusiti našo trditev, da slučajnostno izbrana enota leži med $-3{,}29$ in $+3{,}29$, in ponavlja izbor enote večkrat, v velikem številu ponavljanj za 0,001, ali 1 ‰ enot ugotovi, da vanje vrednosti ne leže v imenovanem razmaku.

Stopnjo tveganja moremo poljubno spreminjati. Če vzamemo širši razmak, je tveganje manjše, če vzamemo ožji, je stopnja tveganja večja. Za vsak poseben primer moremo stopnjo tveganja ugotoviti tako, da izračunamo verjetnost, da se napovedani dogodek ne zgodi.

14.22 Za normalno distribucijo moremo v splošnem z določenim tveganjem napovedati, kakšna je vrednost enot, ki jo slučajnostno izberemo iz normalne populacije, na tri različne načine:

a) nakazati moremo razmak, *v katerem* je z določeno stopnjo tveganja vrednost, ki jo ima slučajnostno izbrana enota;

b) nakazati moremo vrednost, *nad katero* z določeno stopnjo tveganja ni vrednost, ki jo ima slučajnostno izbrana enota;

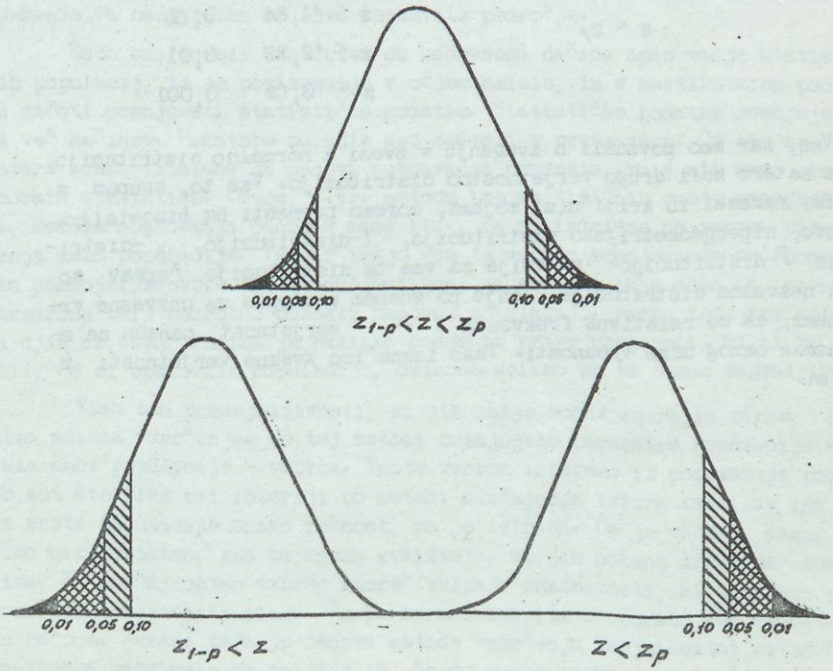
c) nakazati moremo mejo, *pod katero* z določeno stopnjo tveganja ni vrednost za enoto, ki jo slučajnostno izberemo iz normalne populacije.

Prva je dvostranska, drugi dve pa sta enostranski trditvi.

Stopnja tveganja, s katerim dajemo izjave, kot smo jih pravkar navedli, more biti poljubna. Običajne stopnje tveganja so: 0,10; 0,05, 0,01 in 0,001. Stopnja tveganja 0,10 je zelo velika in pomeni, da moremo pričakovati, da se *v povprečju v enem od desetih* poskusov izkaže, da traitev, ki jo postavljamo, ne drži. Zato to stopnjo tveganja redko uporabljamo. Najpogosteje sklepamo s tveganjem 0,05. To pomeni, da moramo pri takem sklepanju pričakovati, da se v enem izmed dvajsetih primerov v povprečju izkaže, da naša traitev ne drži. Veliko bolj zanesljive so traitve s tveganjem 0,01 ali 0,001. Pri teh se izkaže, da traitev ne velja povprečno v enem od sto oziroma od tisoč preizkusov.

Razmak, izven katerega je z določeno stopnjo tveganja vrednost za slučajnostno izbrano enoto, imenujemo kritični razmak, meje pa kritične meje.

V sliki 14.7 so prikazani kritični razmaki in kritične meje za vse tri vrste traitev.



Slika 14.7 Kritični razmaki in kritične meje za normalno distribucijo

Glede na pomen P (glej sliko 14.8), na splošno velja za slučajnostno izbrano enoto iz standardizirano normalne distribucije:

		Razmak	Tveganje
s tveganjem $2P$	$z_{1-P} < z < z_P$	$-1,64 < z < +1,64$	0,10
		$-1,96 < z < +1,96$	0,05
		$-2,58 < z < +2,58$	0,01
		$-3,29 < z < +3,29$	0,001
s tveganjem P	$z_{1-P} < z$	$-1,28 < z$	0,10
		$-1,64 < z$	0,05
		$-2,32 < z$	0,01
		$-3,09 < z$	0,001
s tveganjem P	$z < z_P$	$z < +1,28$	0,10
		$z < +1,64$	0,05
		$z < +2,32$	0,01
		$z < +3,09$	0,001

Vzorčevalni relativni vzorci

14.23 Vse, kar smo povedali o tveganju v zvezi z normalno distribucijo, velja za katero koli drugo verjetnostno distribucijo. Vse to, skupno s kritičnimi razmaki in kritičnimi mejami, moremo prenesti na binomialno, Poissonovo, hipergeometrijsko distribucijo, t-distribucijo, χ^2-distribucijo in F-distribucijo. Vse velja za vse te distribucije, čeprav so prve tri nezvezne distribucije. Zanje po znanem postopku za nezvezne vrste vzamemo, da se relativna frekvenca - verjetnost nanaša na enotin razmak okrog dane vrednosti. Tako izenačimo zvezne verjetnosti z nezveznimi.

Vzorčenje - veliki vzorci

Petnajsto poglavje

VZORČENJE - VELIKI VZORCI

OSNOVE VZORČENJA

15.1 Vzporedno s hitrim ekonomskim in družbenim razvojem se je v zadnjih desetletjih vedno bolj večala potreba po statističnih podatkih v najrazličnejših področjih.

Stare metode zbiranja in obdelave podatkov niso mogle zadostiti novim zahtevam, ker so bile drage, počasne, zvezane z angažiranjem velikega števila ljudi itd. Razen tega so se te metode zaradi posebnih pogojev izkazale za neuporabne za nova nerazvita področja.

Zato smo morali odstopiti od osnovnega načina opazovanja statističnih populacij, to je popisovanja v ožjem smislu, in v marsikaterem področju začeti ocenjevati statistične podatke. Statistične podatke ocenjujemo na več načinov. Nekatere od njih smo našteli v prvem delu. Od teh so nekatere ocene izdelane na osnovi poznavanja predmeta, bolj ali manj brez uporabe statistike. Druge, n.pr. metoda izbora tipičnih enot, monografija, metoda opazovanja osnovne mase itd., pa so izdelane na osnovi poznavanja dela populacije. Vendar imajo vse te metode ocenjevanja večinoma iste pomanjkljivosti. Vse so subjektivne, za sestavljanje ocen pa moramo razmeroma zelo podrobno poznati osnovno populacijo; razen tega pri nobeni od njih ne vemo, kolika je razlika ocene od prave vrednosti, ki bi jo dobili, če bi opazovali populacijo, oziroma koliko so te ocene zanesljive.

Vseh teh pomanjkljivosti, ki jih imajo ocene zgornjih tipov, pa nima metoda vzorčenja. Po tej metodi ocenjujemo parametre populacije iz dela enot populacije - vzorca. Enote vzorca izteremo iz populacije podobno kot številke pri loteriji po metodi slučajnega izbora tako, da ima vsaka enota populacije enako možnost, da je izbrana. Če je vzorec, kakor pravimo slučajnosten, ima ta ocena kvalitete, ki jih nobena drugačna ocena nima. Za slučajnosten vzorec namreč veljajo zakonitosti, ki bistveno pripomorejo k kakovosti ocene. Te pa moremo odkriti z uporabo verjetnostnega računa. Zaradi tega je osnova metode vzorčenja verjetnostni račun. Zakonitosti vzorčenja pa veljajo le, če so enote izbrane naključno.

15.2 Metoda vzorčenja je v svojih posebnih oblikah v nekaterih znanostih že dolgo znana kot raziskovalna metoda. Tako je vzorčenje že razmeroma dolgo znano v agronomiji, biologiji, medicini, psihologiji itd. Vzrok, da se je metoda vzorčenja uveljavila najprej v teh znanostih, ni slučaj. Raziskovalec v teh področjih je nujno navezan na opazovanje samo dela populacije, ker celotne populacije ne obstajajo. Tako je za agronoma trideset poizkusnih parcelic samo vzorec iz populacije neomejenega števila

parcelic, ki bi jih obravnavali enako kakor stvarnih trideset parcelic. Prav tako je v poskusih v biologiji ali medicini trideset ali štirideset poskusov na miših samo vzorec iz neomejenega števila poskusov pod enakimi pogoji. Enako vzamemo določeno število artiklov, ki jih proizvedemo na določenem stroju, za vzorec iz neomejene populacije artiklov, ki bi jih stroj proizvedel v enakih pogojih.

V teh znanostih se je vzorčenje uveljavilo razmeroma zgodaj, ker s popolnim opazovanjem nismo mogli zajeti neomejenih umišljenih populacij. Ravno to pa je vzrok, da se je vzorčenje razmeroma kasno začelo uveljavljati tudi v socialno-ekonomskih znanostih. Te znanosti namreč večinoma obravnavajo populacije, ki imajo končno število enot. Te pa je možno opazovati s klasičnimi metodami opazovanja, s popisi, s tekočimi službami ita. Vendar je potreba po statističnih podatkih prerasla možnosti, ki jih dajejo klasične metode. Zato so izdelali specifične metode vzorčenja, ki omogočajo objektivno in kvalitetno ocenjevanje podatkov tudi za to področje. Te metode so sedaj splošno orodje statistike in v mnogih področjih izpodrivajo klasične metode opazovanja.

Prednosti in pomanjkljivosti metode vzorčenja

15.3 *Prednosti* vzorčenja pred drugimi metodami ocenjevanja, a v nekaterih primerih tudi pred metodami popolnega opazovanja, so tele:

a) Število enot v vzorcu je znatno manjše kakor pri popolnem opazovanju.

Pri velikih populacijah je število enot v vzorcu komaj nekaj promilov ali odstotkov od skupnega obsega populacije.

Manjše število enot pa ima niz prednosti za vzorčenje.

b) Stroški vzorčenja so znatno manjši kakor stroški za popolno opazovanje.

c) Kakovost zbranih podatkov je v marsikaterem primeru celo boljša kot pri popisu, ker je za izvedbo potrebno manj ljudi. Ti pa so zaradi lažje izbire kvalitetnejši, bolje poučeni ita.

d) Zaradi izbire kadrov in možnosti kontrole so napake pri popisovanju kar najbolj skrčene.

e) Čas, v katerem dobimo rezultate, je znatno krajši kot pri popolnem opazovanju; to pa znatno poveča operativno vrednost podatkov.

f) Pogreška, ki izvira iz tega, da nismo opazovali celotne populacije, temveč samo del, moremo pri vzorčenju objektivno izmeriti in regulirati. Te lastnosti nima nobena druga metoda ocenjevanja. Ta pogreška, ki jo moremo oceniti, je dostikrat celo manjša kakor pa je rezultat popisnih napak pri celotnem opazovanju. Popisnih napak je predvsem kriv velik obseg

populacije, sodelovanje velikega števila popisovalcev, ki jih ne moremo kvalitetno poučiti in kontrolirati, ita.

g) V raziskovalnem delu je metoda vzorčenja edino možna, ker umišljenih - hipotečnih populacij v nobenem primeru ne moremo zajeti popolno.

Pomanjkljivosti vzorčenja pa so predvsem v tem:

a) Vzorčenje moremo uporabiti le za populacije z velikim številom enot.

b) Zato ne moremo dobiti z njim zadovoljivih ocen za manjša, temveč le za večja področja.

c) Vzorčenje ne da zanesljivih podrobnih rezultatov. Zato z njim ne moremo obdelovati kombinacije več znakov.

d) Zanesljivost ocen za posamezne podatke je različna.

e) Plan vzorčenja je bolj kompliciran kakor plan kompletnega opazovanja. Enako je izračunavanje ocen in pokazateljev zanesljivosti ocen razmeroma zamotano.

Uporaba vzorčenja pri statističnem opazovanju

15.4 Vzorčenje moremo uporabiti v vseh stopnjah statističnega opazovanja.

Pri izvedbi statistične akcije more vzorec *nadomestiti popolno opazovanje* ali pa ga uporabimo kot *dopolnilo* k popolnemu opazovanju. Tako n. pr. z vzorcem v nekaterih letih v Jugoslaviji nadomestimo popis živine. Kot dopolnilo pa vzorčenje uporabimo, če podatke, ki jih potrebujemo zelo podrobno tudi za manjša področja, zberemo s popisom. Podatke, ki jih potrebujemo samo za večja področja pa istočasno opazujemo z vzorcem. Zato prve podatke zberemo od vseh enot populacije, druge pa samo od vnaprej naključno izbranih enot.

Pri *obdelavi* statističnega gradiva moremo uporabiti vzorčenje za dva namena. Obdelava velikih popisov traja razmeroma dolgo. V tem času podatki velikokrat že izgube aktualnost. Z vzorcem *predhodnih rezultatov* pa moremo dobiti ocene podatkov znatno prej zaključkom popolne obdelave. V tem primeru je popolna obdelava tudi potrebna, ker da vzorec samo globalne, ne pa podrobne rezultate.

Moremo pa vzorčenje uporabiti tudi za *dopolnilno obdelavo*. Čeprav s statističnim opazovanjem zberemo popolne podatke na vsa vprašanja, včasih popolno obdelamo le nekatera od teh vprašanj. Druge pa obdelamo z vzorcem v dopolnilni obdelavi, ker sicer bodisi zaradi pomanjkanja sredstev ali drugih vzrokov ne bi te podatke sploh obdelali.

Z vzorcem moremo tudi *kontrolirati* izvedbo statistične akcije v vseh stopnjah. Z njim ocenimo delo popisovalcev ali obdelavo statističnih podatkov v celoti. Vzorčenje uporabljamo celo za popravo podatkov, ki smo jih

zbrali s popisom. Pri popisu živine, pa tudi pri drugih statističnih akcijah, neposredno po popisu izvedemo kontrolni vzorec. Kakovostno dobri, zanesljivi in posebej za to poučeni kontrolorji z vzorcem ugotove, koliko so zbrani podatki različni od stvarnih. Podatke zberejo kontrolorji z drugačnimi sredstvi, bolj vestno in natančno kakor popisovalci. Tako ugotovimo učinek napak, ki jih z drugimi sredstvi kontrole ne moremo ugotoviti. Iz podatkov kontrolnega vzorca pa ocenimo za vsako vrsto živine oziroma podatek posebej, koliko je stvaren oziroma koliko je treba v skupnem ta podatek popraviti. Vzorčna kontrola ima namen, ugotoviti skupen učinek napak, ne pa napake posameznika.

Čeprav je v problemih, ki smo jih nakazali v prejšnjem odstavku, število enot v vzorcu razmeroma majhen del celotne populacije, je število enot vseeno razmeroma veliko, saj je imel n.pr. 2 % vzorec popisa prebivalstva v Sloveniji ca 30000 enot ita.

Vzorčenje se je razvilo v eno izmed osnovnih metod pri proučevanju javnega mnenja, analize trišča, kontrole proizvodnje in proizvodnega procesa ita. Pri teh problemih so vzorci razmeroma majhni in v dosti primerih ne presegajo nekaj deset enot populacije. Seveda rezultati takih vzorcev niso tako zanesljivi kakor rezultati vzorcev z več tisoči enot. Vendar je njihova zanesljivost, glede na problem, ki ga s takimi vzorci rešujemo, večinoma ustrezna.

Osnovna populacija. Vzorec. Populacija vseh vzorcev

15.5 Osnovna populacija. Populacijo za katero skušamo oceniti parametre z vzorčenjem, imenujemo *osnovno populacijo*. Osnovna populacija je n.pr. populacija vseh kmetijskih gospodarstev v LR Sloveniji na določen dan, če z vzorčenjem ocenjujemo podatke za skupnost kmetijskih gospodarstev v Sloveniji na ta dan. V razširjenem pojmu populacije imamo za osnovno populacijo tudi skupnost podatkov o številu goveje živine za posamezna gospodarstva v Sloveniji na določen dan. Osnovna populacija je tudi n.pr. populacija industrijskih podjetij ali v ožjem smislu populacija vrednosti proizvodanje, posameznih podjetij, populacija števila delavcev posameznih podjetij ita. Te populacije imenujemo *končne osnovne populacije*, ker imajo končno število enot.

Z vzorčenjem pa obravnavamo tudi osnovne populacije, ki imajo *neomejeno število enot*. Te populacije sicer stvarno ne obstajajo, vendar jih moremo miselno dojeti in razumeti. Te vrste populacije obravnavamo predvsem pri raziskovalnem delu. *Umišljena* ali *hipotetična populacija*, kakor imenujemo take vrste populacij, je n.pr. skupnost vseh ponovitev določenega poizkusa v enakih pogojih.

Enako imamo za hipotetično populacijo skupnost vseh artiklov, ki bi jih proizvedel nek stroj v enakih pogojih, če bi delal neomejeno. Tudi iz

tega primera vidimo, da ima taka umišljena populacija neomejeno število enot, da pa obstaja samo miselno, ne pa stvarno. Praktično je namreč nemogoče ustvariti enake pogoje za razmeroma veliko - kaj šele za neomejeno proizvodnjo artiklov, ne glede na to, da ne moremo take populacije nikdar izčrpati.

Osnovne populacije z neomejenim številom enot moremo izpeljati tudi iz *zveznih populacij*. Parcelo z obsegom enega hektarja moremo razdeliti v populacijo s končnim številom površinic po 1 m^2. Dobimo jo, če razdelimo celotno površino z mrežo v kvadrate po 1 m^2. Če pa vzamemo drugo osnovo in imamo za enoto populacije vsako parcelico kvadratne oblike s površino 1 m^2, ne glede na to, kako je položena, in dopustimo, da se med seboj prekrivajo, dobimo osnovno populacijo z neomejenim številom enot. Podobno populacijo dobimo iz osnovne zvezne populacije velikega kosa tkanine, zvezne populacije proizvodnje cementa v določenem dnevu itd.

Enote osnovne populacije so posamezno gospodarstvo, posamezno industrijsko podjetje, poskus pod enakimi pogoji, parcelica po 1 m^2, 10 dkg cementa itd. Znaki teh enot so vse značilnosti, ki jih proučujemo. *Parametri osnovnih populacij* pa so vsota podatkov, srednje vrednosti, strukturni odstotki, mere variacije, mere korelacije itd.

15.6 Vzorec. Iz osnovne populacije izberemo delno populacijo, ki jo imenujemo *vzorec*. Enote vzorca morajo biti izbrane tako, da ima vsaka enota enako možnost, da je izbrana v vzorec. To dosežemo s slučajnim izborom. Vzorec je sam zase tudi populacija, ker ima svoje enote (enote osnovne populacije, ki smo jih izbrali), te pa znake, ki so isti kakor znaki v osnovni populaciji. Kakor za vsako drugo populacijo moremo iz podatkov v vzorcu izračunati sredine, proporce, mere variacije in korelacije itd.

Ker je glede na število enot teorija vzorčenja različna, ločimo *male vzorce od velikih*. Stroge meje med malimi in velikimi vzorci ni. Običajno imamo vzorce, ki imajo nekaj deset enot, za male, če obseg vzorca ne prekorači sto enot. Število enot v velikih vzorcih je običajno znatno večje. Včasih znaša tudi nekaj tisoč ali celo deset tisoč enot.

Medtem ko uporabljamo male vzorce predvsem v raziskovalnem delu, z velikimi vzorci običajno nadomeščamo popolna opazovanja.

Glede na način izbora ločimo *vzorce s ponavljanjem* od *vzorcev brez ponavljanja*. Ta razlika izvira iz tega, kako izbiramo enote. Pri vzorčenju s ponavljanjem moremo enoto, ki smo jo enkrat izbrali v vzorec, izbrati ponovno. Glede na sistem izbiranja enot ima izbrana enota to možnost. Pri vzorcu brez ponavljanja pa vsako izbrano enoto izključimo iz nadaljnjega izbora. Ta razlika je teoretično važna. Praktično pa se različen sistem izbora pokaže le, če je vzorec razmeroma velik del osnovne populacije. Teoretično obravnavanje vzorcev brez ponavljanja je nekoliko bolj zamotano kakor obravnavanje vzorcev s ponavljanjem. Ta razlika pa ne pride v poštev pri hipotetičnih in zveznih populacijah in se njen pomen tudi pri končnih populacijah tem bolj izgublja, čim večja je osnovna populacija. O-

bičajno v praksi za končne populacije jemljemo vzorce brez ponavljanja, ker so bolj logični, razen tega pa so ocene vzorcev brez ponavljanja zanesljivejše kakor ocene iz vzorcev s ponavljanjem.

15.7 Populacija vseh vzorcev. Vzorec, ki smo ga po postopku slučajnega izbora izbrali iz osnovne populacije, je samo eden izmed vseh možnih vzorcev, ki jih moremo dobiti iz osnovne populacije, če enote vzorca izbiramo po istem postopku. Če izbiranje slučajnega vzorca ponovimo, prav gotovo ne bomo izbrali v vzorec iste enote, kakor v prvem vzorcu. Če se vprašamo, kakšno je število vseh možnih različnih vzorcev, ki jih moremo iz osnovne populacije izbrati, spoznamo, da je to število že za razmeroma majhne populacije in vzorce izredno veliko. Za večje populacije in vzorce pa je število vseh možnih vzorcev praktično neomejeno.

Če proučimo skupnost vseh možnih vzorcev, se prepričamo, da sestavljajo vsi možni vzorci populacijo v statističnem smislu. Vsak vzorec moremo imeti za enoto populacije vseh možnih vzorcev. Ta populacija ima vse lastnosti statističnih populacij. Sestavljena je iz istovrstnih enot - vzorcev, ki so izbrani v enakih pogojih. Enote te populacije - vzorci imajo svoje znake, ki so v tem primeru povprečja, proporci, mere variacije in korelacije ita., ki jih izračunamo iz podatkov za posamezni vzorec. Ne da bi navajali določene primere, sklepamo, da ti znaki variirajo, ker so izračunani za vsak vzorec iz podatkov za posamezni vzorec. Skupnost vseh vzorcev je torej populacija, v kateri so posamezni vzorci enote, povprečja, proporci, mere variacije in korelacije ita. pa znaki teh enot.

15.8 Kakor smo že navedli, je število vseh možnih vzorcev za dano populacijo razmeroma veliko že za ne preveč velike osnovne populacije in vzorce. Če poznamo število enot v osnovni populaciji N in število enot v vzorcu n, pa moremo število vseh možnih vzorcev izračunati. Število vseh možnih vzorcev brez ponavljanja je enako številu vseh možnih kombinacij brez ponavljanja po n enot iz N enot. To moremo računsko izraziti s simbolom

$$\binom{N}{n} = \frac{N.(N-1).....(N-n+2)(N-n+1)}{1.\ 2\ \ .....\ (n-1)\ .\ n} \qquad (15.1)$$

Število vseh možnih vzorcev s ponavljanjem pa je enako številu kombinacij s ponavljanjem po n enot iz N enot. To število pa je

$$\binom{N+n-1}{n} = \frac{N.(N+1).....(N+n-2)(N+n-1)}{1.\ 2\ \ .....\ (n-1)\ .\ n} \qquad (15.2)$$

Če vzamemo za primer razmeroma majhno osnovno populacijo z $N = 75$ enotami in vzorec z $n = 25$ enotami, pokaže račun, da je število vseh možnih vzorcev brez ponavljanja 52,59 trilijonov, število vseh možnih vzorcev s ponavljanjem pa 181890 trilijonov. Primer pokaže, da je že za tako majhne populacije in vzorce obseg populacije vseh možnih vzorcev velikanski in

veliko večji za vzorce s ponavljanjem kakor za vzorce brez ponavljanja.

Zakonitosti v populaciji vseh vzorcev

15.9 Ker je število vseh možnih vzorcev v praktičnih primerih izredno veliko, ne moremo neposredno proučevati populacije vseh možnih vzorcev, razen če ima populacija malo enot. Take populacije pa so brez praktičnega pomena. Ne moremo si misliti, da bi n.pr. frekvenčno distribucijo povprečij v vzorcih sestavili tako, da bi iz populacije izbrali vse možne vzorce, za vsak vzorec izračunali aritmetično sredino $\bar{x}$, iz vseh dobljenih povprečij pa sestavili frekvenčno distribucijo. Prav tako bi bilo težko neposredno izračunati aritmetično sredino povprečij vzorcev, varianco vseh povprečij ita. Po tej poti je torej težko ali pa celo nemogoče proučevati populacije vseh vzorcev.

Vendar s teoretičnim razglabljanjem pridemo do zvez, ki veljajo med osnovno populacijo in populacijo vseh možnih vzorcev. Izkaže se, da moremo za dosti primerov iz podatkov za osnovno populacijo sklepati, kakšne so distribucije in parametri v populaciji vseh vzorcev. K uspešnemu odkrivanju zakonitosti zvez med osnovno populacijo in populacijo vseh vzorcev bistveno pripomore verjetnostni račun: ta je, kakor smo že omenili, osnova vzorčenja. Te zakonitosti, ki so za nekatere probleme znane natančno, za druge pa samo približno, izkoristimo za praktično uporabo pri ocenjevanju in sklepanju iz vzorca na osnovno populacijo.

Za posamezne statistične količine so zakonitosti več ali manj različne in odvisne od podatka, ki ga proučujemo, od osnovne populacije in od tipa vzorca. Zato bomo za aritmetično sredino, ki je eden izmed najvažnejših in najpogostejših statističnih parametrov, razložili splošna načela vzorčenja, v nadaljevanju pa posebej obdelali še druge parametre in probleme, ki so važni v socialno-ekonomskih proučevanjih.

ENOSTAVNO VZORČENJE

Ocenjevanje aritmetične sredine

15.10 Standardna pogreška ocene. Vzemimo, da je osnovna populacija znaka X neomejena populacija, ki je normalno razporejena, ima aritmetično sredino M_x in standardni odklon σ. Iz te populacije izberimo slučajne vzorce z obsegom n in proučimo v populaciji vseh možnih vzorcev, kaj je z aritmetičnimi sredinami vzorcev $\bar{x}$.

Za populacijo aritmetičnih sredin v vseh možnih vzorcih iz neomejene normalne populacije veljajo naslednje zakonitosti:

a) Aritmetične sredine

$$\bar{x} = \frac{1}{n} Sx \tag{15.3}$$

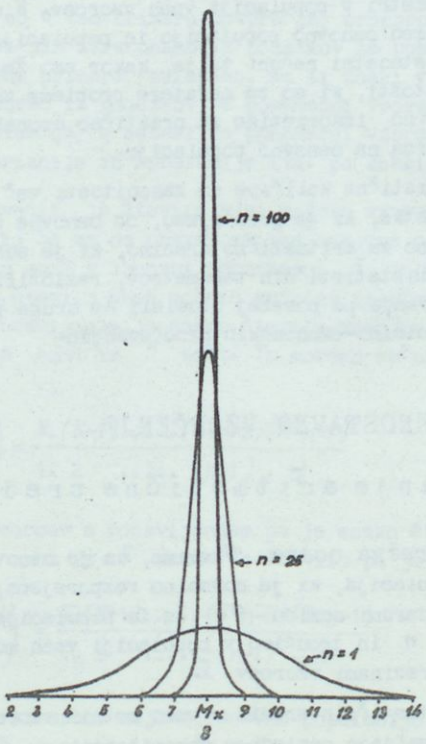
vseh možnih vzorcev se distribuirajo v normalni distribuciji, če se x distribuira v osnovni populaciji normalno.

b) Aritmetična sredina aritmetičnih sredin vseh vzorcev $M_{\bar{x}}$, ki jo imenujemo tudi pričakovano vrednost x, zaznamujemo pa z $E(\bar{x})$, je enaka aritmetični sredini M_x v osnovni populaciji.

$$E(\bar{x}) = M_{\bar{x}} = M_x \tag{15.4}$$

c) Standardni odklon aritmetičnih sredin vseh vzorcev $\sigma_{\bar{x}}$, ki ga imenujemo tudi standardna pogreška aritmetične sredine in zaznamujemo z $SE_{\bar{x}}$ je enak

$$SE_{\bar{x}} = \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \tag{15.5}$$



Slika 15.1 Distribucija aritmetičnih sredin vseh vzorcev za $n = 25$ in $n = 100$ enot

kvocientu standardnega odklona v osnovni populaciji σ_x in kvadratnega korena iz števila enot n v vzorcu.

V prejšnjih obrazcih smo uporabili za aritmetično sredino v osnovni populaciji ali populaciji vseh možnih vzorcev črke M z ustreznim indeksom, za aritmetično sredino v vzorcu pa znak $\bar{x}$. Znak Σ pomeni seštevanje v populaciji, za seštevanje v vzorcu pa uporabljamo znak S. To simboliko uporabljamo na splošno.

15.11 V sliki 15.1 sta narisani za normalno populacijo, za katero je $M_x = 8,0$ in standardni odklon $\sigma_x = 2,0$, distribucija osnovnega znaka x in distribucija aritmetičnih sredin $\bar{x}$, enkrat iz vzorcev po $n = 25$ enot, drugič iz vzorcev po $n = 100$ enot. Slika nazorno pokaže, kakšen je rezultat zgornjih zakonitosti. Aritmetične sredine v vzorcih se grupirajo okrog aritmetične sredine osnovne populacije M_x. V skladu z normalno distribucijo, v kateri se distribuirajo aritmetične sredine vzorcev $\bar{x}$, imamo veliko več vzorcev, za katere se $\bar{x}$ malo odklanjajo od M_x, kakor pa vzorcev, za katere se $\bar{x}$ bolj odklanja od aritmetične sredine osnovne populacije $M_x = 8,0$. Ti odkloni so veliko manjši, če smo vzeli vzorce po $n = 100$ enot, kot pa če smo vzeli vzorce z $n = 25$ enotami.

Ker se večina aritmetičnih sredin vzorcev razmeroma malo odklanja od aritmetične sredine osnovne populacije $M_x = 8,0$, imamo aritmetično sredino $\bar{x}$, ki jo izračunamo iz podatkov za vzorec, za oceno prave vrednosti aritmetične sredine M_x v osnovni populaciji. Če so enote vzorca izbrane slučajnostno in je zagotovljena enaka možnost za vsak vzorec, moremo z veliko večjo verjetnostjo pričakovati, da bo ocena z vzorcem od prave vrednosti manj kot pa bolj različna.

15.12 Če ima vsak vzorec enako možnost, da ga izberemo, je normalna distribucija aritmetičnih sredin vzorcev verjetnostna distribucija za aritmetično sredino v vzorcu. Za take vzorce je verjetnost, v skladu z normalno distribucijo, večja, da izberemo vzorec, za katerega so razlike med oceno $\bar{x}$ in pravo vrednostjo M_x manjše, kakor pa obratno. Enako možnost za vsak vzorec dosežemo, če izberemo enote vzorca tako, da imamo zagotovljeno slučajnost izbora. Zato je slučajnost izbora osnovnega pomena za dobro ocenjevanje in pravilno tolmačenje rezultatov vzorčenja.

Za normalno distribucijo vemo, da s tveganjem 0,05 vrednost enote, ki jo slučajnostno izvlečemo iz populacije, ni od aritmetične sredine populacije različna za več kot 1,96 SD.

Če to pravilo, ki velja na splošno za vse normalne distribucije priredimo za normalno distribucijo aritmetičnih sredin $\bar{x}$ vseh vzorcev, moremo reči:

S tveganjem 0,05 se ocena aritmetične sredine $\bar{x}$, ki jo po obrazcu

$$\bar{x} = \frac{1}{n} Sx \qquad (15.6)$$

izračunamo iz podatkov slučajnega vzorca, od prave vrednosti aritmetične sredine osnovne populacije M_x ne odklanja za več kot

$$e_{\bar{x}} = 1{,}93\ SE_{\bar{x}} = 1{,}93\ \sigma_x/\sqrt{n} \qquad (15.7)$$

To pa praktično pomeni, da moremo s tem izrazom oceniti kakovost ocene.

Tako za naš primer normalne populacije s standardnim odklonom $\sigma = 2{,}0$ pričakujemo, da se ocena aritmetične sredine $\bar{x}$, ki jo ocenimo s slučajnim vzorcem z $n = 25$ enotami, s tveganjem 0,05 ne razlikuje od prave vrednosti M_x za več kot $1{,}93 \cdot 2{,}0/\sqrt{25} = 0{,}784$. Če vzorec zvečamo in vzamemo $n = 100$ enot, pa odklon s tveganjem 0,05 ni večji kot 0,392. To oceno odklona napovemo, ne da bi poznali M_x.

Iz tega primera vidimo, da zanesljivost ocene aritmetične sredine merimo s standardno pogreško ocene $SE_{\bar{x}}$. Nanjo pa vplivamo, če spreminjamo velikost vzorca. Obrazec 15.7 pokaže, da je standardna pogreška ocene in z njo zanesljivost ocene obratno proporcionalna s kvadratnim korenom iz števila enot v vzorcu n. To pomeni, da moramo vzeti štirikrat večji vzorec, če želimo standardno pogreško ocene zmanjšati na polovico, devetkrat večji vzorec, če jo hočemo skrčiti na tretjino ita.

15.13 Meje zaupanja za oceno aritmetične sredine. V prejšnjem odstavku smo aritmetično sredino osnovne populacije M_x ocenili z aritmetično sredino vzorca $\bar{x}$. Aritmetično sredino osnovne populacije pa ocenjujemo tudi z razmakom.

Kakor vemo, se ocena aritmetične sredine $\bar{x}$ od prave vrednosti M_x z 0,05 tveganjem ne odklanja od prave vrednosti M_x za več kot $e_{\bar{x}} = 1{,}93\ SE_{\bar{x}}$ (Če stopnjo tveganja spremenimo, se ustrezno spremeni tudi koeficient z; za tveganje 0,05 je $z = 1{,}93$, za tveganje 0,01 je $z = 2{,}58$ za tveganje 0,001 pa $z = 3{,}29$).

Če vemo, v kakšni okolici okrog M_x je ocena $\bar{x}$, vemo tudi obratno, v kakšni okolici okrog ocene $\bar{x}$ leži prava vrednost M_x. Zato moremo reči: Če s slučajnim vzorcem ocenimo aritmetično sredino populacije z $\bar{x}$, sklepamo, da je ta vzorec iz populacije, za katero je aritmetična sredina M_x od ocene $\bar{x}$ različna za manj kot $e_{\bar{x}} = 1{,}93 \cdot SE_{\bar{x}}$. Tveganje, da je razlika večja kot $e_{\bar{x}} = 1{,}93\ SE_{\bar{x}}$ je 0,05. Glede na koeficient 1,93 moremo namreč z verjetnostjo 0,05 izbrati vzorec, za katerega je razlika med $\bar{x}$ in M_x večja kot $1{,}93\ SE_{\bar{x}}$.

V drugačni obliki moremo ta stavek povedati tudi takole: Če s slučajnim vzorcem dobimo aritmetično sredino $\bar{x}$, leži prava vrednost aritmetične sredine M_x za osnovno populacijo v mejah

$$\bar{x} - e_{\bar{x}} < M_x < \bar{x} + e_{\bar{x}} \qquad (15.8)$$

To pomeni: Le povprečno v petih od sto vzorcev dobimo tak $\bar{x}$, da leži M_x izven meja, ki so dane v 15.8.

Dobljeni razmak imenujemo *razmak zaupanja*, $e_{\bar{x}}$ *odklon zaupanja*, $\bar{x} - e_{\bar{x}}$ *spodnjo mejo zaupanja*, $\bar{x} + e_{\bar{x}}$ *pa zgornjo mejo zaupanja* za oceno aritmetične sredine.

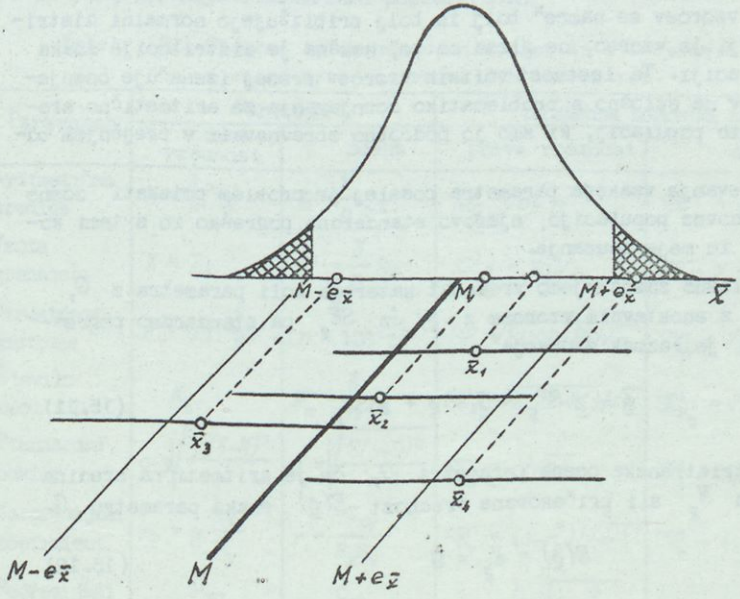
Aritmetično sredino osnovne populacije, razen z aritmetično sredino vzorca $\bar{x}$, ki jo imenujemo *točkovno oceno*, ocenjujemo tudi z *razmakom zaupanja*. Razmak zaupanja včasih pišemo razen z obrazcem 15.8 tudi v obliki

$$M_x = \bar{x} \pm e_{\bar{x}} \qquad (15.9)$$

ali pa navedemo njegovo spodnjo M_s in zgornjo mejo M_z takole

$$M_s = \bar{x} - e_{\bar{x}}; \quad M_z = \bar{x} + e_{\bar{x}} \qquad (15.10)$$

15.14 V sliki 15.2 je nakazan odnos med pravo aritmetično sredino M, oceno z vzorcem $\bar{x}$ in razmakom zaupanja za nekaj vzorcev iz normalne populacije. Iz slike nazorno vidimo, da 95 % ocen aritmetične sredine $\bar{x}$ leži v razmaku $M - 1.96\, SE_{\bar{x}}$ do $M + 1,96\, SE_{\bar{x}}$ in le 5 % izven tega razmaka. Obratno pa opazimo, da je za vse ocene za $\bar{x}$ (95 %) v tem razmaku (glej $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$) prava vrednost M v razmaku zaupanja $\bar{x} - e_{\bar{x}}$ do $\bar{x} + e_{\bar{x}}$, za one (5 %), ki leže izven razmaka $M - 1,96\, SE_{\bar{x}}$ do $M + 1,96\, SE_{\bar{x}}$ pa leži tudi prava vrednost M izven razmaka zaupanja (glej $\bar{x}_4$).



Slika 15.2 Odnos med pravo aritmetično sredino M, oceno z vzorcem $\bar{x}$ in razmakom zaupanja

Ocenjevanje drugih parametrov z enostavnimi vzorci

15.15 V prejšnjem odstavku smo tolmačili problematiko ocenjevanja aritmetične sredine z vzorčenjem iz neomejene normalne populacije. Vendar v praksi ne ocenjujemo samo aritmetične sredine. Tudi vse populacije niso neomejene in tudi niso distribuirane v normalni distribuciji. Glede na to, kakšna je osnovna populacija in kaj ocenjujemo, pa se zakonitosti vzorcev menjajo. Vendar so v vseh primerih ocenjevanja osnove iste. V vseh primerih, enako kot pri aritmetični sredini, ocenjujemo parameter s slučajnim vzorcem točkovno ali z razmakom zaupanja z določeno stopnjo tveganja.

Teorija in praksa ocenjevanja posameznih parametrov iz malih vzorcev je za vsak parameter več ali manj različna in vezana na določene predpostavke o osnovni populaciji. Zato je ocenjevanje z malimi vzorci neenotno in razmeroma zapleteno.

Za velike vzorce pa je problematika znatno enostavnejša. Z določenimi približki, ki niso v škodo zanesljivosti rezultatov, moremo posamezne parametre ocenjevati z velikimi vzorci razmeroma enostavno in enotno.

Prva pomembna lastnost velikih vzorcev je ta, da so vzorčne distribucije ocen parametrov vedno manj odvisne od distribucije znaka v osnovni populaciji, čim večji je vzorec. Distribucije ocen parametrov v populaciji vseh možnih vzorcev se namreč bolj in bolj približujejo normalni distribuciji, čim večji je vzorec, ne glede na to, kakšna je distribucija znaka v osnovni populaciji. Ta lastnost velikih vzorcev precej izenačuje ocenjevanje parametrov na splošno s problematiko ocenjevanja za aritmetično sredino iz normalnih populacij, ki smo jo podrobno obravnavali v prejšnjem odstavku.

Za ocenjevanje vsakega parametra posebej je problem poiskati oceno parametra za osnovno populacijo, njegovo standardno pogreško in s temi količinami razmak in meje zaupanja.

Če na splošno zaznamujemo vrednost katerega koli parametra z G, oceno parametra z enostavnim vzorcem z $\hat{g}$, z SE_g pa standardno pogreško za oceno $\hat{g}$, je razmak zaupanja

$$\hat{g} - z\,SE_g < G < \hat{g} + z\,SE_g \qquad (15.11)$$

$\hat{g}$ je nepristranska ocena parametra G, če je aritmetična sredina vseh možnih ocen M_g ali pričakovana vrednost $E(\hat{g})$ enaka parametru G.

$$E(\hat{g}) = M_g = G \qquad (15.12)$$

Iz obrazca 15.4 vidimo, da je aritmetična sredina vzorca $\bar{x}$ nepristranska ocena. Vse ocene niso nepristranske. Pri pristranskih ocenah pa

želimo, da so vsaj *dosledne*. Za dosledne imamo tiste ocene, za katere M_g konvergira k G, če velikost vzorca večamo. Tako je n.pr. izraz

$$s'^2 = \frac{S(x - \bar{x})^2}{n} \tag{15.13}$$

ki je po svojem sestavu enak obrazcu za izračunavanje variance v populaciji, dosledna ocena za varianco v populaciji, izraz

$$s^2 = \frac{S(x-\bar{x})^2}{n-1} = \frac{Sx^2 - X^2/n}{n-1} \tag{15.14}$$

pa je nepristranska ocena variance σ^2, ker zanjo velja

$$E(s^2) = M_{s^2} = \sigma^2 \tag{15.15}$$

Ocene in standardne pogreške za nekaj najvažnejših parametrov

15.16 V tabeli 15.1 so obrazci za izračunavanje ocen iz enostavnih slučajnih vzorcev za nekaj najvažnejših parametrov. Navedeni so tudi obrazci za izračunavanje standardnih pogrešk ocen.

Tabela 15.1 Obrazci za izračunavanje ocen in standardnih pogrešk ocen za nekaj najvažnejših parametrov

Parameter	Parameter		Standardna pogreška ocene	
	prava vrednost	ocena	prava vrednost	ocena
Aritmetična sredina	$M = \frac{1}{N}\Sigma x$	$\bar{x} = \frac{1}{n}Sx$	$SE_{\bar{x}} = \sigma/\sqrt{n}$	$SE'_{\bar{x}} = s/\sqrt{n}$
Vsota vrednosti	$X = \Sigma x$	$X' = \frac{N}{n}Sx$	$SE_{X'} = N\sigma/\sqrt{n}$	$SE'_{X'} = Ns/\sqrt{n}$
Strukturni odstotek	$P = 100\frac{N_a}{N}$	$p = 100\frac{n_a}{n}$	$SE_p = \sqrt{P(100-P)}/\sqrt{n}$	$SE'_p = \sqrt{p(100-p)}/\sqrt{n-1}$
Število enot z a	N_a	$N'_a = \frac{N}{n}n_a$	$SE_{N'_a} = \sqrt{N_a(N-N_a)}/\sqrt{n}$	$SE'_{N'_a} = \sqrt{N'_a(N-N'_a)}/\sqrt{n-1}$
Standardni odklon	$\sigma = \sqrt{\frac{\Sigma(X-M)^2}{N}}$	$s = \sqrt{\frac{S(x-\bar{x})^2}{n-1}}$	$SE_s = \sigma/\sqrt{2n}$	$SE'_s = s/\sqrt{2n}$
Korelacijski koeficient	$r_0 = \frac{C_{xy}}{\sigma_x\sigma_y}$	$r = \frac{C_{xy}}{s_x s_y}$	$SE_r = (1-r_0^2)/\sqrt{n-1}$	$SE'_r = (1-r^2)/\sqrt{n-1}$
Regresijski koeficient	$b_0 = \frac{C_{xy}}{\sigma_x^2}$	$b = \frac{C_{xy}}{s_x^2}$	$SE_b = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r_0^2}{n-2}}$	$SE'_b = \frac{s_y}{s_x}\sqrt{\frac{1-r^2}{n-2}}$

Obrazci za izračunavanje pravih vrednosti standardnih pogrešk pa so v praktičnih primerih za ocenjevanje običajno neuporabni, ker ne poznamo pravih vrednosti parametrov v osnovni populaciji. Zato so razen obrazcev za izračunavanje pravih vrednosti standardnih pogrešk dani še obrazci, s katerimi dobimo zadosti dobro oceno standardnih pogrešk iz podatkov vzorca. Z njimi izračunavamo standardne pogreške in meje zaupanja iz podatkov v vzorcu. V glavnem so posamezni simboli pojasnjeni v sami tabeli ali pa jih poznamo iz prejšnjih poglavij. Kjer nimamo samostojnih znakov, smo uporabili ' za znak za oceno (n. pr. SE', X', N_a' itd.) Tako zaznamovanje ocen je zelo priključno, vendar nepotrebno, če imamo samostojen izraz za oceno (n. pr.: S, $\bar{x}$ itd.). c_{xy}, ki je nepristranska ocena kovariance, je analogno kot za nepristransko oceno variance izračunana po obrazcu

$$c_{xy} = \frac{S(x-\bar{x})(y-\bar{y})}{n-1} = \frac{S_{xy}-XY/n}{n-1} \qquad (15.16)$$

Če pogledamo standardne pogreške katere koli izmed zgornjih količin, vidimo, da imajo vse v imenovalcu izraz $\sqrt{n}$. V *nekaterih* primerih imamo odstopanje od tega pravila, vendar to odstopanje ni bistveno, ker je samo n zmanjšan za eno oziroma dve. To pa ne vpliva bistveno na rezultat. Standardne pogreške vseh zgornjih ocen so torej obratno proporcionalne s kvadratnim korenom števila enot v vzorcu. To je važna lastnost vseh standardnih pogrešk: ocene parametrov so zanesljivejše, čim večji je vzorec. Zanesljivost pa se veča s kvadratnim korenom iz števila enot v vzorcu n.

Popravek za standardne pogreške za vzorce brez ponavljanja

15.17 Zgornji obrazci za standardne pogreške veljajo za vzorce s ponavljanjem oziroma za vzorce iz neomejene osnovne populacije. Vendar predvsem prve štiri parametre (M, X, P, N_a) navadno ocenjujemo za končne populacije z vzorcem brez ponavljanja. Za take vrste vzorcev prave vrednosti za standardne pogreške v tabeli 15.1 popravimo tako, da jih pomnožimo s faktorjem

$$\sqrt{\frac{N-n}{N-1}} \qquad (15.17)$$

ocene standardnih pogrešk za te štiri parametre pa s faktorjem

$$\sqrt{\frac{N-n}{N}} \qquad (15.17a)$$

Kvocient $n/N = f$ imenujemo vzorčni delež. Vzorčni delež pove, kakšen del celote je zajet v vzorcu. Če to upoštevamo, moremo popravni faktor

pisati v obliki

$$\sqrt{\frac{N-n}{N-1}} \approx \sqrt{\frac{N-n}{N}} = \sqrt{1-f} \approx 1 - \frac{f}{2} \qquad (15.18)$$

Prvi približek je upravičen, ker je N običajno velik in je enica v imenovalcu nebistvena. Drugi približek pa je upravičen le, če je vzorčni delež manjši kot $f = 0{,}10$. V praksi vzorčenja je vzorec običajno drobec odstotka ali nekaj odstotkov od osnovne populacije. Za take vzorce da zadnji približek uporabne rezultate. Če pa je vzorec velik del populacije, je treba faktor izračunati iz izraza $\sqrt{1-f}$.

Popravni faktor zniža standardno pogreško, ker je v vsakem primeru manjši kot 1. Iz tega vidimo, da je zanesljivost ocen z vzorčenjem brez ponavljanja večja kakor z vzorčenjem s ponavljanjem. Če je vzorec n.pr. 1 % od celotne populacije, je popravni faktor enak 0,995. Popravek za ta vzorec ni bistven, ker je popravni faktor zelo blizu 1. Za vzorec, ki ima 10 % enot osnovne populacije, pa je popravni faktor enak 0,95. Za tako velike vzorce je popravek že znatnejši, vendar še vedno ni velik.

Iz obrazcev v tabeli 15.1 in pomena popravnega faktorja vidimo, da je zanesljivost ocen odvisna predvsem od velikosti vzorca, ne pa od velikosti osnovne populacije. Le za vzorce brez ponavljanja ima velikost populacije malenkosten vpliv na zanesljivost ocene. Vendar tudi zanje standardna pogreška ni neposredno odvisna od velikosti osnovne populacije N, temveč le od vzorčnega deleža f. Ta lastnost vzorcev je zelo važna, ker se na prvi pogled zdi, da so ocene za velike populacije nezanesljivejše kakor ocene za manjše populacije: to pa ni res.

Tehnika enostavnega slučajnega izbora

15.18 Pri praktični izvedbi enostavnega slučajnega vzorca, kakršne smo obravnavali doslej, se pojavi vprašanje, kako je treba izbrati enote vzorca iz populacije, da vzorec zadosti pogoju, da je za vse enote populacije enaka možnost izbora. Take vrste vzorcev imenujemo *enostavne slučajne vzorce*, izbor pa *izbor brez omejitve*. V praksi vzorčenja uporabljamo tudi vzorce, ki zadoščajo drugačnim osnovam. Nekatere od njih obravnavamo v kasnejših odstavkih, drugi pa presegajo naš okvir.

Pri vsakem enostavnem vzorčenju iz končne populacije je osnova izbora *okvir vzorca*. Okvir vzorca je register vseh enot v osnovni populaciji. Običajno je okvir vzorca spisek enot za osnovno populacijo, v katerem so enote oštevilčene s tekočimi številkami. Okvir vzorčenja ni vedno običajen spisek. Kot okvir vzorčenja se zelo izkaže kartoteka osnovnih enot, ki so oštevilčene s tekočimi številkami. Če gre za vzorec površin, more biti okvir vzorčenja tudi zemljevid, v katerem so vrisane osnovne enote. Tudi v zemljevidu zaznamujemo enote populacije - površinice z zaporednimi številkami. Vendar ni nujno, da imajo tekoče številke zvezo z regionalno razmestitvijo površinic.

15.19 Loterijski način izbora enot. Pri loterijskem načinu izbora enot v enostavni vzorec izbiramo enote vzorca z žaro. Kot so že omenili, ima vsak okvir vzorčenja enote populacije oštevilčene z zaporednimi številkami. Pri loterijskem načinu slučajnega izbora imamo v žari listke z zaporednimi številkami za vse enote populacije. Listkov je toliko, kolikor je enot v populaciji. Iz žare na slepo izvlečemo listek. V vzorec vključimo enoto, ki ima zaporedno številko, napisano na listku. Če izbiramo vzorec s ponavljanjem, listek, potem ko zabeležimo zaporedno številko za izbrano enoto, vrnemo v žaro in listke premešamo. Pri tem postopku more biti ista enota ponovno vključena v izbor. Pri vzorčenju brez ponavljanja pa vsak izbran listek odložimo. Po tem postopku moremo vsako enoto izbrati v vzorec samo enkrat.

Loterijski način izbora enot vzorca pa je praktično zamuden in okoren.

Priključeneje je, da v žaro vključimo samo deset enakih elementov n.pr. kock, kroglic in podobno; nanje napišemo številke od 0 do 9. Če ima populacija n.pr. 10000 enot, dobimo zaporedno številko za vsako enoto vzorca, če iz žare štirikrat potegnemo po eno kroglico (s ponavljanjem). Iz štirih številk, ki smo jih potegnili, sestavimo štirimestno število, ki je zaporedna številka enote, ki jo vključimo v izbor. Tako vključimo v izbor n.pr. enoto, ki ima v okviru vzorčenja zaporedno številko 7503, če smo zaporedoma potegnili številke: 7, 5, 0 in 3. Postopek ponavljamo, dokler ne izberemo ustreznega števila enot vzorca.

Enostaven slučajen vzorec

15.20 Tablice slučajnih števil. Vendar tudi postopek z desetimi elementi v praksi malokdaj uporabljamo, razen v nekaterih posebnih primerih za majhne populacije. Običajno izbiramo enote v enostaven slučajnostni izbor s tablicami slučajnih številk. Te sestavimo n.pr. takole: V žari imamo deset kroglic, ki so oštevilčene s številkami od 0 do 9. Iz nje na slepo izbiramo s ponavljanjem po eno izmed kroglic in zapisujemo izide teh izborov. Vrsta številk, ki jo dobimo, če izbiranje ponavljamo, je osnova slučajnega izbora. Iz te vrste številk moremo v zaporedju sestavljati števila s poljubnim številom mest glede na velikost populacije. Tablice slučajnih številk, ki so jih sestavili različni avtorji, uporabljamo pri vsakem enostavnem slučajnem izboru. Tako odpade pisanje tekočih številk na listke ali izvlačenje iz žare. Med tabelami v dodatku je tudi kratka tablica slučajnih številk. Čeprav je ta tablica razmeroma kratka, jo moremo vseeno izkoristiti za razmeroma velike vzorce. Tablico slučajnih številk moremo namreč brati naprej in nazaj, od zgoraj navzdol ali od spodaj navzgor ali v diagonalah. Razen tega se morejo posamezne skupine slučajnih številk, ki sestavljajo zaporedno številko za izbrane enote, med seboj prekrivati. Tako za populacijo z $N = 10000$ enotami iz prve vrste slučajnih številk v tabeli

5354 9142 0847 5393 5413 3505 7153 5334 9703 6221

enostavno sestavimo tele štirimestna števila, ki so zaporedne številke

izbranih enot v okviru vzorčenja: 5354, 9142, 0847, 5393, 5416, 6505, 7156, 5334, 9708, 3221; če jih beremo od leve na desno.

Če beremo številke v tablici slučajnih števil od desne proti levi, pa dobimo tele zaporedne številke: 1226, 3079, 4365, 6517, 5056, 6145, 3935, 7480 2419, 4535.

Če vzamemo, da štirimestna števila drse prek vrste slučajnih števil tako, da skupine po štiri slučajne številke premikamo za eno naprej, pa dobimo v vzorec tele enote: 5354, 3549, 5491, 9142, 1420 itd. Tako obseg tablic slučajnih števil zelo povečamo. Stvaren izbor s tablicami slučajnih števil je prikazan v primeru.

Primeri enostavnega slučajnega vzorčenja

15.21 Vzemimo za primer enostavno vzorčenje gozdne parcele kvadratne oblike s 40 ha površine. Gozdna parcela je razdeljena v 20 × 20 = 400 parcelic po 10 a površine. Osnovna populacija ima torej $N = 400$ enot. V tabeli 15.2 je podana shematično cela gozdna površina. V posameznih poljih je vpisan volumen lesne mase v m^3 na ustreznih parcelicah. Parcelice po 10 a so oštevilčene z zaporednimi številkami ob robeh kvadrata. Celo parcelo provizorično razdelimo z navpično mejo po sredini v dva dela. V levem delu parcele imamo vrste od 0 do 19, v desnem delu parcele pa vrste od 20 do 39. Mesto ustrezne parcele v določeni vrsti pa je določeno s številkami od 0 do 9. Te številke so označene v shemi v tabeli 5.2 zgoraj in spodaj. Iz oštevilčenja za vrste in stolpce moremo sestaviti tromestno število od 000 do 399 za vsako parcelico. Prvi dve številki označujeta vrsto, tretja pa mesto v vrsti. Tako je n. pr. s trimestnim številom 154, označena parcelica, ki leži v 15. vrsti v 4. stolpcu. Če pogledamo v tabelo 15.2, ugotovimo, da je volumen lesne mase na tej parcelici 10 m^3. Parcelica z zaporedno številko 338 leži v 33. vrsti, to je v desni polovici gozda v osmem stolpcu. Volumen lesne mase te parcele je 21 m^3. Podobno identificiramo vsako parcelico.

Vzemimo da z enostavnim vzorčenjem brez ponavljanja ocenjujemo povprečen volumen na eno parcelo: $\bar{X}$; skupen volumen lesa za ves gozd: Y; standardni odklon variabilnosti volumna lesa po parcelicah: S_x, odstotek parcelic, na katerih je volumen lesa 20 m^3 in več: p in število parcelic z volumnom lesa 20 m^3 in več: N_a.

Vse te količine bomo ocenili z istim slučajnim vzorcem. V praksi slučajnostno izbiramo parcelice vzorca in na terenu izmerimo volumen lesne mase za izbrane parcelice. V našem primeru bomo potrebne podatke o volumnu črpali iz sheme v tabeli 15.2.

Po načrtu izbora moramo oceniti podatke za populacijo z navadnostratnim slučajnim vzorcem osnovnih parcelic. V vzorec torej izberemo $n = 80$ parcelic.

Tabela 15.2 Shematičen prikaz gozda z N = 400 parcelicami po 10 a. Vpisani podatki v posameznih poljih so volumen lesne mase v m^3, uokvirjene parcele pa so parcele, ki smo jih izbrali z dvajsetodstotnim enostavnim slučajnim izborom

	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	
0	17	15	14	13	12	14	15	21	20	15	19	15	19	22	16	12	9	17	11	10	20
1	22	16	24	15	28	15	21	17	19	18	13	16	18	25	15	14	10	12	9	12	21
2	23	17	25	28	21	13	27	25	20	26	12	17	23	20	24	16	9	11	14	16	22
3	28	27	16	26	28	14	24	20	13	22	18	21	16	28	22	17	12	13	15	17	23
4	16	21	28	27	21	28	23	18	21	15	14	23	17	18	19	25	28	17	23	23	24
5	21	27	22	17	28	25	26	19	14	16	23	18	20	25	22	24	29	16	17	25	25
6	15	16	14	15	17	14	17	15	13	17	19	24	15	30	28	17	20	23	21	18	26
7	14	13	14	12	15	18	16	22	20	20	15	23	17	29	33	32	24	29	16	19	27
8	12	10	11	10	8	11	12	15	17	25	19	20	28	32	34	31	33	22	19	18	28
9	11	8	9	6	11	9	10	12	21	21	20	22	16	31	30	32	30	31	21	19	29
10	10	6	7	8	9	6	11	14	27	23	29	18	30	29	31	24	16	28	18	21	30
11	7	5	5	11	7	8	12	16	22	21	19	24	29	25	29	19	18	24	22	20	31
12	4	3	0	7	9	10	14	25	18	20	16	22	26	30	25	21	20	27	24	19	32
13	5	4	2	8	11	13	27	23	19	18	21	18	25	18	23	18	24	23	19	18	33
14	6	7	3	10	14	17	20	24	22	27	26	19	22	18	20	19	21	18	17	18	34
15	12	8	4	7	10	5	13	15	24	19	19	21	18	27	19	24	20	22	23	18	35
16	8	0	1	6	11	8	15	14	22	20	17	23	18	20	18	16	23	23	21	18	36
17	8	4	5	7	9	12	16	17	26	21	21	16	24	18	22	20	25	14	17	20	37
18	10	9	8	10	11	13	14	15	19	19	22	24	20	16	19	17	16	15	13	16	38
19	14	13	12	11	9	11	17	23	18	23	17	22	25	20	22	15	13	12	16	14	39
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	5	7	8	9	

Tabela 15.3 Slučajne številke (sš), zaporedne številke (zš) parcelic in ustrezni volumni (v) za dvajsetodstotni vzorec iz sheme v tabeli 15.2.

(sš = slučajne številke; zš = zaporedne številke; v = volumen)
dv = slučajna številka ponovljena)

sš	zš	v	sš	zš	v	sš	zš	v	sš	zš	v	sš	zš	v
535	135	13	839	–	–	152	152	4	189	189	19	786	386	16
090	90	11	781	381	24	257	257	16	095	95	av	397	397	12
142	142	6	271	271	23	832	–	–	381	381	av	364	364	18
321	321	22	714	314	29	634	234	22	283	283	32	988	–	–
939	–	–	241	241	23	826	–	–	084	84	8	031	31	av
091	91	8	523	123	7	352	352	av	939	–	–	650	250	23
463	63	15	351	351	21	281	281	20	867	–	–	404	4	12
251	251	13	031	31	27	043	43	27	660	260	19	052	52	22
596	196	17	752	352	18	842	–	–	438	38	13	285	285	31
784	384	19	517	117	16	121	121	3	050	50	21	574	174	9
519	119	21	202	202	19	284	284	34	486	86	12	071	71	13
843	–	–	363	333	20	010	10	22	987	–	–	890	–	–
816	–	–	076	76	16	172	172	5	501	101	dv	564	164	11
915	–	–	914	–	–	890	–	–	790	390	17	603	203	22
606	206	9	698	298	21	420	20	26	141	141	7	189	189	av
540	140	6	047	47	18	251	251	av	302	302	30	457	57	13
046	43	23	908	–	–	787	387	15	215	215	14	616	216	10
780	380	22	843	–	–	966	–	–	529	129	20	539	139	18
729	329	19	496	96	10	620	220	19	096	96	av	397	397	av
548	148	22	707	307	23	137	137	23	287	287	22	232	232	16
101	101	6	473	73	12	095	95	9	043	43	av			
085	85	11	041	41	21	743	346	21	274	274	33			

Zaporedne številke izbranih parcelic poiščemo s tablicami slučajnih števil. Ker ima populacija $N = 400$ enot, vzamemo trimestna slučajna števila. Najenostavneje je, da vzamemo kar prva tri mesta štirimestnih skupin slučajnih števil po posameznih stolpcih. Od teh tromestnih števil pa moremo izkoristiti samo del tromestnih števil; ker ima populacija samo $N = 400$ enot, tromestnih slučajnih števil pa je 1000. Zdi se, da so vsa slučajna števila nad 400 za naš primer neuporabna. Taka izraba tablic slučajnih števil ni učinkovita. Polje jih izkoristimo, če napravimo pravilo, da slučajna števila 000–399 izkoristimo neposredno iz slučajnih števil, od 400 do 799 pa dobimo zaporedne številke izbranih parcel tako, da od njih odštejemo $N = 400$. Če n.pr. naletimo na slučajno število 535, izberemo v vzorec parcelico z zaporedno številko 535 – 400 = 135. Slučajna števila od 800 do 999 ne moremo uporabiti in jih izključimo. Navedeni način splošno uporabljamo v vseh podobnih primerih. Če ima populacija n.pr. $N = 3000$ enot, slučajne številke od 0000 do 2999 izkoriščamo neposredno, slučajnim številkam od 3000 do 5999 odštevamo 3000, slučajnim številkam od 6000 do 8999 pa 6000. Slučajne številke od 9000 do 9999 pa ne moremo vključiti v izbor.

V tabeli 15.3 je nakazan izbor enot vzorca. V prvem stolpcu so vpisane po vrsti vsa slučajna števila, ki smo jih dobili iz tablic slučajnih števil po nakazanem načinu. V drugem stolpcu so iz slučajnih števil dobljene po zgornjem načinu zaporedne številke enot, ki so izbrane v vzorec. Iz tabele vidimo, da slučajna števila nad 800 izpadejo. V tretji stopnji v shemi 15.2 poiščemo parcelice z ustreznimi zaporednimi številkami. Te parcelice so v shemi dokvirjene, ustrezni volumni pa so vpisani v tretjem stolpcu v tabeli 15.3.

Iz vzorca $n = 80$ parcelic, za katerega so individualni podatki vzorca nakazani v shemi v tabeli 15.2 in v tabeli 15.3, dobimo tele ocene:

$\bar{x} = Sx/n = 1320/80 = 17{,}4$; $X' = NSx/n = 400.1320/80 = 6950$;

$s^2 = 50{,}7$; $s = 7{,}12$; $p = 100.n_1/n = 100.32/80 = 40\ \%$

$M_1' = N.n_1/n = 400.32/80 = 160$

Če po vrsti za navedene ocene parametrov izračunamo standardne pogreške in meje zaupanja, dobimo tele rezultate.

V vseh obrazcih za izračunavanje standardnih pogrešk za zgornje ocene je $\sqrt{(1-f)/n}$. Zato ga izračunajmo vnaprej:

$$\sqrt{(1-f)/n} = \sqrt{(1-0{,}20)/80} = 0{,}10$$

Standardna pogreška ocene aritmetične sredine je

$$\sigma_{\bar{x}} = s\sqrt{(1-f)/n} = 7{,}12.0{,}10 = 0{,}712$$

Maksimalni odklon je $e_{\bar{x}} = z\, SE_{\bar{x}} = 1,96 \cdot 0,712 = 1,40$

meje zaupanja pa:

$$17,4 - 1,40 < M_x < 17,4 + 1,40$$

$$16,00 < M_x < 18,80$$

Za oceno skupnega volumna dobimo po tabeli 15.1.

$$SE'_{X'} = N s \sqrt{(1-f)/n} = 400 \cdot 7,12 \cdot 0,10 = 284,8$$

$$SE'_{X'}\ \% = 100\, SE'_{X'}/X' = 100 \cdot 284,8/6950 = 4,1\ \%$$

$$e'_{X'} = z \cdot SE'_{X'} = 1,96 \cdot 284,8 = 558$$

$$X' - e'_{X'} = 6950 - 558 = 6392$$

$$X' + e'_{X'} = 6950 + 558 = 7508$$

$$6392 < X < 7508$$

Za oceno strukturnega odstotka parcel z nad 200 m^3 lesne mase dobimo

$$SE'_p = \sqrt{p(100-p)}\ \sqrt{(1-f)/(n-1)} \approx \sqrt{40(100-40)}\ 0,10 = 4,9\ \%$$

$$e'_p = z\, SE'_p = 1,96 \cdot 4,9 = 9,6\ \%$$

$$p - e'_p = 40 - 9,6 = 30,4\ \%$$

$$p + e'_p = 40 + 9,6 = 49,60\ \%$$

$$30,4\ \% < P < 49,6\ \%$$

Za oceno števila parcel z več kot 20 m^3 lesne mase pa je

$$SE'_{N'_a} = \sqrt{N'_a(N-N'_a)}\ \sqrt{(1-f)/(n-1)} \approx \sqrt{130(400-130)}\ 0,10 = 19,6$$

$$e_{N'_a} = z\, SE'_{N'_a} = 1,96 \cdot 19,6 = 38,4$$

$$N'_a - e'_{N'_a} = 160 - 38,4 = 121,6$$

$$N'_a + e'_{N'_a} = 160 + 38,4 = 198,4$$

$$122 < N_a < 198$$

Za oceno vsakega izmed zgornjih parametrov smo ocenili: točkovno oceno $(x, X', p$ in $N'_a)$, oceno standarne pogreške SE, maksimalni odklon e in meje zaupanja.

Če primerjamo dobljene rezultate s pravimi rezultati za populacijo, vidimo, da $M = 17,17$, $Y_0 = 7110$, $P = 40\ \%$, $N = 160$, sodejo v razmak zaupanja, kakor smo pričakovali. Za odstotek in število parcel nad 20 m^3

lesne mase pa se ocena in prava vrednost slučajno celo skladata.

15.22 V občini A z $N = 2185$ kmetijskimi gospodarstvi skušamo z vzorčenjem oceniti povprečno obdelovalno površino na eno gospodarstvo in skupno obdelovalno površino. Za ta namen iz osnovne populacije izberemo enostaven slučajen vzorec z $n = 353$ enotami.

Za izbrana gospodarstva smo z anketo ugotovili obdelovalno površino. Iz teh podatkov smo dobili naslednje rezultate: Vsota površin v $n = 353$ izbranih gospodarstvih je $S_x = 1443,1$ ha.

Iz teh podatkov izračunamo po obrazcih iz tabele 15.1:

Ocena povprečne obdelovalne površine:

$$\bar{x} = \frac{1}{n} S_x = 1443,1/353 = 2,2145 \text{ ha}$$

Ocena skupne obdelovalne površine:

$$X' = \frac{N}{n} S_x = \frac{2185}{653} \cdot 1443,1 = 4839 \text{ ha}$$

Da ocenimo standardni pogreški in intervala zaupanja za zgornji oceni $\bar{x}$ in X' iz vzorca, ocenimo standardni odklon za obdelovalno površino v osnovni populaciji.

Ker poznamo vsoto kvadratov obdelovalnih površin $S_x^2 = 4435,11$ v izbranih gospodarstvih, izračunamo nepristrano oceno variance po obrazcu

$$s^2 = \frac{S_x^2 - \frac{(S_x)^2}{n}}{n-1} = \frac{4435,11 - \frac{1443,1^2}{353}}{353-1} = 1,8903$$

Ocena standardnega odklona pa je

$$s = \sqrt{1,8903} = 1,3751 \text{ ha}$$

Iz teh podatkov je dalje standardna pogreška za oceno aritmetične sredine:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N}} = \frac{1,3751}{\sqrt{353}} \sqrt{\frac{2185-353}{2185}} = 0,0451 \text{ ha}$$

Maksimalen odklon je enak

$$e_{\bar{x}} = 1,96 \cdot SE_{\bar{x}} = 1,96 \cdot 0,0451 = 0,0884 \text{ ha}$$

meje zaupanja pa

$$M_s = \bar{x} - e_{\bar{x}} = 2,2145 - 0,0884 = 2,1261 \text{ na}$$

$$M_z = \bar{x} + e_{\bar{x}} = 2,2145 + 0,0884 = 2,3029 \text{ ha}$$

Če primerjamo obrazce za ocenjevanje za aritmetično sredino z obrazci za ocenjevanje vsote podatkov v tabeli 15.1, vidimo, da dobimo standardno pogrešku, maksimalen odklon in meje zaupanja za vsoto podatkov, če ustrezne količine za aritmetično sredino pomnožimo s številom enot v osnovni populaciji N.

Tako je

$$SE_{Y'} = N.SE_{\bar{x}} = 2185.0,0451 = 98,54$$

$$e_{Y'} = N.e_{\bar{x}} = 2185.0,0884 = 193 \text{ ha}$$

$$Y'_s = N.M_s = 2185.2,1261 = 4646 \text{ ha}$$

$$Y'_z = N.M_z = 2185.2,3029 = 5032 \text{ ha}$$

Prava vrednost skupne obdelovalne površine je torej v mejah 4343 ha do 5082 ha. Primerjava tega podatka s pravim rezultatom, ki smo ga dobili s popisom, pokaže, da leži prava vrednost $Y = 4750$ ha znotraj mej razmaka zaupanja.

15.23 Pri popisu prebivalstva 15. marca 1948 smo za Slovenijo, ki je imela po tem popisu $N = 1,391.878$ prebivalcev, z dvoodstotnim vzorcem ocenili prednodne rezultate.

Iz osnovnega gradiva smo po metodi slučajnega izbora izbrali $n = 27847$ popisnic. Med drugim je dal vzorec naslednjo strukturo prebivalstva po stanu:

Tabela 15.4 Ocena strukture prebivalstva po stanu v Sloveniji po popisu 15.3.1948 z 2 % vzorcem. (Vir: Neobjavljeni podatki vzorca, SZ LRS)

Stan	Podatki vzorca						Podatki popisa
	n_i	$p=100\frac{n_i}{n}$	SE_p	$e_p = 1,96SE_p$	$p_s = p-e_p$	$p_z = p+e_p$	P
Skupno	27847	100,00					100,0
Samski	15828	56,84	0,294	0,577	56,26	57,42	57,1
Poročeni	9978	35,83	0,284	0,557	35,27	36,39	35,6
Vdovci in razvezani	2041	7,33	0,155	0,304	7,03	7,63	7,3

V tabeli 15.4 so po vrsti navedeni podatki vzorca, ocene strukturnih odstotkov, standardne pogreške ocen odstotkov, maksimalni odkloni, meje zaupanja in prave vrednosti strukturnih odstotkov, ki smo jih dobili kasneje s popolno obdelavo.

Za vse podatke, razen za standardno pogreško ocen, je v glavi tabele 15.4 nakazano, kako so podatki v posameznih stolpcih izračunani. Standardna pogreška ocene za posamezne strukturne odstotke pa je bila izračunana po obrazcu

$$SE'_p = \sqrt{\frac{p(100-p)}{n-1}} \sqrt{\frac{N-n}{N}} \qquad (15.18)$$

ki je naveden v tabeli 15.1.

Rezultati popisa v zadnjem stolpcu so v skladu z našim pričakovanjem, ker vsi ležé v razmaku zaupanja.

Iz ocen strukturnih odstotkov dobimo neposredno ocene skupnega števila prebivalstva po stanu, kar po obrazcu

$$N'_a = \frac{N}{100} p_a \qquad (15.19)$$

ocene strukturnih deležev pomnožimo s skupnim številom prebivalstva. Analogno dobimo iz podatkov za strukturne deleže tudi druge količine za oceno števila prebivalstva.

Tabela 15.5 Ocene skupnega števila prebivalstva po stanu v Sloveniji po popisu 15.3.1948

Stan	V tisočih		
	N'_a	$SE_{N'_a}$	$e_{N'_a}$
Skupno	1392		
Samski	791	4,1	8,0
Poročeni	499	4,0	7,8
Razvezani in vdovci	102	2,2	4,2

Ocenjevanje razlik dveh parametrov

15.24 Ker je primerjava osnova statistične analize, dostikrat izračunavamo razlike parametrov za različni populaciji. Če sta vzorca v populacijah, ki ju primerjamo, med seboj neodvisna, je diferenca ocen parametrov ocena diference parametrov. Če zaznamujemo oceno parametra G_1 v prvi populaci-

ji z $\hat{g}_1$, oceno parametra G_2 v drugi populaciji z $\hat{g}_2$, je ocena razlike parametrov $(\Delta G)'$ enaka

$$(\Delta G)' = \Delta\hat{g} = \hat{g}_2 - \hat{g}_1 \qquad (15.20)$$

Če sta $\hat{g}_2$ in $\hat{g}_1$ nepristranski ali vsaj dosledni oceni parametrov G_2 in G_1, se razlike $(\hat{g}_2 - \hat{g}_1)$ za velike vzorce distribuirajo približno normalno s sredino ΔG in standardno pogreško ocene $SE_{\Delta\hat{g}}$.

$$M_{\hat{g}_2-\hat{g}_1} = G_2 - G_1 \qquad (15.21)$$

$$SE^2_{\hat{g}_2-\hat{g}_1} = SE^2_{\hat{g}_2} + SE^2_{\hat{g}_1} \qquad (15.22)$$

Kvadrat standardne pogreške diference ocen je za velike in neodvisne vzorce enak vsoti kvadratov standardnih pogrešk za oceni $\hat{g}_1$ in $\hat{g}_2$. Dalje ocenjujemo diference parametrov enako kot ocene parametrov. Zanje izračunavamo maksimalne odklone, meje zaupanja ita.

15.25 Treba je oceniti kolika je razlika v odstotku prvorazrednih artiklov za proizvodnjo v dveh oddelkih v istem podjetju. Za ta namen smo v oddelku A vzeli slučajni vzorec $n_1 = 500$ artiklov in ugotovili, da je $p_1 = 50$ % artiklov prvorazrednih. V oddelku B smo vzeli vzorec $n_2 = 300$ artiklov in ugotovili, da je $p_2 = 70$ %.

Ocena razlike v odstotku prvorazrednih artiklov je

$$\Delta p = p_2 - p_1 = 70 - 50 = 20\ \%$$

Po obrazcu 15.22 in tabeli 15.1 je kvadrat standardne pogreške ocene razlike

$$SE^2_{\Delta p} = \frac{n_1(100-p_1)}{n_1-1} + \frac{n_2(100-p_2)}{n_2-1} = \frac{50 \cdot 50}{500-1} + \frac{70 \cdot 30}{300-1} \approx 12{,}03$$

Standardna pogreška ocene diference je

$$SE_{\Delta p} = \sqrt{12{,}03} = 3{,}47\ \%$$

maksimalni odklon s tveganjem 0,05

$$e_{\Delta p} = 1{,}96 \cdot 3{,}47 = 6{,}8\ \%$$

meje zaupanja pa $\Delta p = 20\% \pm 6{,}8\%$

Prava vrednost razlike v odstotku prvorazrednih artiklov je s 5 % tveganjem med 13,2 % in 23,8 %.

15.26 V anketi o porabi prehrambenih artiklov smo anketirali $n_1 = 1000$ delavskih družin in $n_2 = 500$ uslužbenskih družin. Za mesečno porabo nekega artikla dobimo za delavsko družino oceno $\bar{x}_1 = 4,8$ kg in $s_1 = 1,5$ kg, za uslužbensko družino pa $\bar{x}_2 = 5,4$ kg in $s_2 = 1,2$ kg (podatki izmišljeni).

Izračunati je treba meje zaupanja za razlike v povprečni porabi med delavskimi in uslužbenskimi družinami.

Ocena razlike v povprečni porabi je

$$\Delta\bar{x} = 5,4 - 4,8 = 1,1 \text{ kg}$$

Standardna pogreška ocene razlike je po obrazcu 15.22 in obrazcu v tabeli 15.1 enaka

$$SE^2_{\Delta\bar{x}} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{1,5^2}{1000} + \frac{1,2^2}{500} = 0,00513$$

Standardna pogreška razlike pa je

$$SE_{\Delta\bar{x}} = \sqrt{0,00513} = 0,0717 \text{ kg}$$

Če vzamemo meje zaupanja s tveganjem 0,05, je maksimalen odklon

$$e_{\Delta\bar{x}} = 1,96 \cdot 0,0717 = 0,14,$$

meje zaupanja pa

$$\Delta\bar{x} = 1,1 \text{ kg} \pm 0,14 \text{ kg}$$

Prava vrednost razlik v povprečni porabi je torej s 5 % tveganjem v mejah od 0,93 do 1,24 kg.

Določanje števila enot v enostavnem slučajnem vzorcu

15.27 V prejšnjem odstavku smo velikost vzorca določili vnaprej in zanj razen ocene podatkov izračunali standardne pogreške in meje zaupanja. Velikokrat pa je naloga obratna. Predpisano imamo zanesljivost, ki naj jo imajo ocene, dobljene iz vzorca, in je treba poiskati tako velik vzorec, da zadosti tem pogojem.

Zanesljivost podatkov je dana s standardno pogreško, še večkrat pa konzument podatkov določi maksimalen odklon, ter navede, za koliko se ocena lahko največ odklanja od prave vrednosti.

Po obrazcu za izračunavanje standardne pogreške za aritmetične sredine v tabeli 15.1 je za vzorčenje s ponavljanjem maksimalni odklon

$$e_{\bar{x}} = z\sigma/\sqrt{n} \qquad (15.23)$$

Iz tega obrazca dobimo, da je n enak

$$n = \left(\frac{z\sigma}{e_{\bar{x}}}\right)^2 \qquad (15.24)$$

Če poznamo standardni odklon v osnovni populaciji in imamo predpisan maksimalen odklon ocene $e_{\bar{x}}$ z določenim tveganjem, od česar je odvisen koeficient z, moremo po obrazcu 15.24 določiti, kako velik vzorec moramo vzeti, da bomo dosegli predpisano zanesljivost za oceno aritmetične sredine.

Podobno določimo potrebno velikost vzorca, da dosežemo predpisano zanesljivost podatkov, tudi za vse druge parametre. Za ocene parametrov M, X, P in N_a so obrazci za izračunavanje števila enot za vzorce s ponavljanjem dani v tabeli 15.3. Prvi obrazec velja za primer, da je maksimalni odklon predpisan absolutno – e, drugi pa v primeru, če je dan relativno – $e\,\%$ v odnosu na velikost parametra.

Tabela 15.6 Obrazci za izračunavanje velikosti vzorca s ponavljanjem pri predpisani absolutni in relativni zanesljivosti podatkov

Parameter	Število enot v vzorcu s ponavljanjem, če je maksimalni odklon predpisan	
	absolutno – e	relativno – $e\,\%$
Aritmetična sredina M	$n = \left(\frac{z\sigma}{e_{\bar{x}}}\right)^2$	$n = \left(\frac{z \cdot KV\%}{e\%}\right)^2$
Vsota podatkov X	$n = \left(\frac{Nz\sigma}{e_x}\right)^2$	
Strukturni odstotek P	$n = \frac{z^2 \cdot P(100-P)}{e_p^2}$	$n = \left(\frac{100z}{e\%}\right)^2 \frac{100-P}{P}$
Število enot z a N_a	$n = \frac{z^2 N_a (N-N_a)}{e_{N_a}^2}$	

Če imamo število enot v vzorcu s ponavljanjem n, izračunamo potrebno število enot v vzorcu brez ponavljanja n' za ocene katerega koli izmed zgornjih parametrov, po enotnem obrazcu

$$n' = \frac{N.n}{N+n-1} \doteq \frac{n}{1+f} \qquad (15.25)$$

Pri tem pomeni: N = število enot v osnovni populaciji; n = število enot v vzorcu s ponavljanjem; n' = število enot v vzorcu brez ponavljanja; $f = n/N$ = vzorčni delež v vzorcu s ponavljanjem.

Hiba zgornjih obrazcev je v tem, da moramo poznati parametre za osnovno populacijo, če hočemo izračunati potrebno velikost vzorca. Teh pa običajno nimamo. Zato si pri planiranju vzorcev pomagamo tako, da parametre, ki jih potrebujemo, ko določamo velikost vzorcev, ocenimo s predhodno analizo pojava, razpoložljivimi podatki prejšnjih popisov ali vzorcev in vzamemo za osnovo približne vrednosti teh količin. Včasih te količine ocenimo s predhodnim manjšim vzorcem. Čeprav ne dobimo natančne potrebne velikosti vzorcev, rezultati vseeno zadoščajo.

15.28 Drug problem, ki nastane v zvezi z določanjem potrebne velikosti vzorcev, pa je v tem, da običajno z vzorčenjem ne ocenjujemo enega samega, marveč razmeroma veliko podatkov. Vzemimo, da bi za vse ocene zahtevali isto relativno zanesljivost in predpišemo relativni maksimalni odklon, n.pr. 5%. Če za vsako oceno izračunamo potrebno velikost vzorca po obrazcih v tabeli 15.3, dobimo za vsak podatek drugo potrebno število enot v vzorcu. V tem primeru je težka odločitev, katero izmed teh števil naj vzamemo za osnovo enotnega vzorca. Če vzamemo najmanjše število enot, za vse druge podatke dobimo ocene, ki so premalo zanesljive. Če pa vzamemo največje število, je za večino podatkov zanesljivost prevelika. V tem primeru vzamemo nek srednje velik vzorec. Pri tem se opiramo predvsem na najvažnejše podatke in skušamo vzeti tako velik vzorec, da zadosti predpisani zanesljivosti predvsem zanje.

Zveza med ocenami za M, X, P in N_a

15.29 V praksi z velikimi vzorci največkrat ocenjujemo prve štiri parametre, ki so navedeni v tabeli 15.1: To so: aritmetična sredina M, agregat X, število enot z dano značilnostjo N_a in strukturi delež P.

Število enot z dano značilnostjo N_a je samo poseben primer agregata X. Če pripišemo namreč enotam, ki nimajo značilnosti a, vrednost $x = 0$, enotam, ki imajo to značilnost, pa vrednost $x = 1$, je N_a agregat tega znaka. Agregat je vsota vrednosti znaka za populacijo. Če seštevamo vrednosti 1 za tiste enote, ki imajo značilnost a, dejansko preštevamo enote, ki imajo to značilnost.

Med agregatom X in aritmetično sredino M velja enostavna in znana zveza

$$M = X/N \qquad (15.26)$$

Za standardno pogreško ocene aritmetične sredine $SE_{\bar{x}}$ in standardno pogreško ocene agregata $SE_{X'}$ velja podobna zveza

$$SE_{\bar{x}} = SE_{X'}/N \qquad (15.27)$$

Podobno velja, da je za strukturni odstotek P in število enot z dano značilnostjo N_a

$$P = 100\, N_a/N \qquad (15.28)$$

$$SE_p = 100\, SE_{N'_a}/N \qquad (15.29)$$

Agregat X je torej splošen parameter, iz katerega moremo izvesti parametre M, N_a in P.

Zato v nadaljevanju obravnavamo samo ocenjevanje in problematiko agregatov, ker moremo iz agregata X priti do vseh drugih treh količin.

Zgornji obrazci veljajo za vse vrste ocen in ne samo za ocenjevanje z enostavnim vzorčenjem.

Ocenjevanje po metodi razmerij

15.30 Vzemimo, da ocenjujemo skupno vrednost proizvodnje v obrtnih podjetjih za določeno stroko. Iz populacije vseh obrtnih podjetij te stroke izteremo na slučajen način vzorec podjetij. Po osnovnem obrazcu za enostavno ocenjevanje agregata v tabeli 15.1 izračunamo iz podatkov vzorca oceno povprečne proizvodnje na eno podjetje, to povprečje pa pomnožimo s skupnim številom podjetij. Zanesljivost te ocene je verjetno majhna, ker je vrednost proizvodnje po podjetjih zaradi različne velikosti obrtnih obratov zelo variabilna. Predpostavljajmo, da proizvodnost dela, merjena z vrednostjo proizvodnje na enega delavca po podjetjih, ni preveč različna. Če ta predpostavka velja, dobimo zanesljivejšo oceno skupne proizvodnje, če ocenimo povprečno proizvodnjo na delavca, to pa pomnožimo s skupnim številom delavcev v tej stroki.

Oceno skupne vrednosti proizvodnje X po tej metodi dobimo z obrazcem

$$Y'_r = Y\frac{Sx}{Sy} = Yr' \qquad (15.30)$$

tako, da vsoto vrednosti proizvodnje v izbranih podjetjih Sx delimo z vsoto števila zaposlenih Sy v teh podjetjih, dobljeno oceno razmerja med

vrednostjo proizvodnje in številom delavstva r' pa pomnožimo s skupnim številom delavstva. Ta metoda da tem boljše ocene agregatov, čim večja je korelacija med znakoma x in y. Če bi bila proizvodnja na enega delavca po podjetjih enaka, bi bila odvisnost med številom zaposlenih in proizvodnjo funkcionalna. r' bi bil v tem skrajnem primeru prava vrednost razmerja med vrednostjo proizvodnje in številom delavstva, pa četudi bi ga izračunali iz podatkov za eno samo podjetje.

Čeprav je ocena po metodi razmerja samo doslejna, ne pa nepristranska ocena agregata, jo z velikim pridom uporabljamo vselej, kadar imamo za podatek, ki je v korelaciji s podatkom, za katerega ocenjujemo agregat, skupno vrednost za populacijo. To pa se v praksi pogosto primeri. V že izvršenih popisih in evidencah je veliko podatkov, ki jih moremo izkoriščati za ocenjevanje po metodi razmerij. Če smo n.pr. v preteklem letu popisali živino po kmetijskih obratih, moremo podatke lanskoletnega popisa uporabiti za osnovo za ocenjevanje po metodi razmerja za vzorec v letošnjem letu. Podatki o številu živine v lanskem in letošnjem letu po posameznih kmetijskih gospodarstvih so v razmeroma visoki korelacijski odvisnosti. Po metodi razmerij pomnožimo lanskoletno število živine z razmerjem skupnega števila živine za izbrana gospodarstva v letošnjem in lanskem letu.

Prav také metodo razmerij s pridom uporabljamo pri vzorčni popravi popisnih podatkov. Neposredno po popisu ponovno skrbneje popišemo enote, ki jih s slučajnim vzorčenjem izberemo v kontrolni vzorec. Popisne podatke popravimo tako, da po metodi razmerij pomnožimo popisni podatek z razmerjem med vsoto podatkov, ki smo jih dobili v kontrolnem vzorcu in vsoto popisnih podatkov za enote v vzorcu. Upravičeno pričakujemo, da bo ocena po metodi razmerij dobra, ker je zelo verjetno, da so popisni podatki in podatki kontrole med seboj zelo odvisni.

Metoda razmerij se obnese tudi pri različnih drugih problemih. Če imamo za večje število pošiljk deklarirane teže (n.pr. carina) popravimo skupno deklarirano težo vseh pošiljk po metodi razmerij tako, da izmed vseh pošiljk izberemo slučajnostni vzorec in izbrane pošiljke stehtamo. Če vsoto pravih tež za izbrane pošiljke delimo s skupno vsoto deklariranih tež za iste pošiljke, dobljeno razmerje pa pomnožimo s skupno deklarirano težo vseh pošiljk, dobimo popravljen podatk o teži. Podobno popravljamo ocene na oko ita.

Standardna pogreška ocene po metodi razmerij je enaka

$$SE_{\bar{x}'_r} = \frac{N \sigma_z}{\sqrt{n}} \sqrt{1-f} \qquad (15.31)$$

Pri tem je σ_z^2 varianca izraza $z_i = y_i(r_i - \bar{r})$

Indeks i pomeni, da so to izrazi za posamezno enoto, $r_i = x_i/y_i$ = razmerje za posamezno enoto, $\bar{r} = X/Y$ pa je razmerje za osnovno populacijo. Iz tega obrazca vidimo, da je standardna pogreška ocene po metodi

razmerij tem boljša, čim manjša je variabilnost razmerij v posameznih enotah oziroma čim večja je korelacija med x in y.

15.31 Za občino z $N = 2185$ gospodarstvi z vzorcem $n = 355$ gospodarstev ocenjujemo letno proizvodnjo mleka v tej občini. Ker razpolagamo s podatkom o skupnem številu krav $Y = 4376$, uporabimo metodo razmerij.

V $n = 355$ gospodarstvih, ki smo jih izbrali s slučajnim vzorcem, smo dobili naslednje podatke: skupno število krav v anketiranih gospodarstvih je bilo $Sy = 1291$ krav, skupna količina proizvedenega mleka v teh gospodarstvih pa je bilo $Sx = 2053$ tisoč litrov. Po metodi razmerij je ocena skupne proizvodnje mleka

$$Y' = Y\frac{Sx}{Sy} = 4376 \cdot \frac{2053}{1291} = 6959 \text{ tisoč litrov}$$

Če primerjamo dobljeno oceno s pravim rezultatom $Y = 3945$, vidimo, da se rezultata presenetljivo skladata. Vendar menimo, da je ta skladnost bolj ali manj slučajna. Iz istega vzorca ocenjena količina mleka po enostavnem otrazcu pa je 3850. Pogreška je znatno večja.

Najboljši pokazatelj uspešnosti metode razmerij v primerjavi z enostavno oceno je razmerje kvadratov standardnih pogrešk po eni in drugi metodi. To razmerje namreč pokaže, kolikokrat večji vzorec moramo vzeti, če hočemo z enostavnim načinom ocenjevanja dobiti tako zanesljivo oceno, kakor smo jo dobili z metodo razmerij. Ker je v našem primeru za enostavno oceno $SE^2_{Y's} = 11300$, za metodo razmerij pa $SE^2_{Y'r} = 2230$, dobimo da je

$$\frac{SE^2_{Y's}}{SE^2_{Y'r}} = \frac{11300}{2230} = 5,07$$

Da bi z enostavno oceno dobili tako zanesljivo oceno, kakor smo jo dobili po metodi razmerij, bi morali vzeti v vzorec s ponavljanjem petkrat večje število enot. Iz primera vidimo, da je dobiček lahko znaten.

VZORČENJA Z OMEJITVAMI

15.32 Osnova dosedanjih ocen so bili enostavni vzorci. Vzorec je enostaven, če ima vsaka enota osnovne populacije enako možnost, da jo izberemo v vzorec. Enostaven vzorec imenujemo zaradi tega tudi vzorec brez omejitev.

Vendar ta način v praksi dostikrat ni najboljši niti po tehnični možnosti izvedbe niti po zanesljivosti ocen. Zato večkrat uporabljamo tudi druge načine vzorčenja. Od teh bomo obravnavali stratificirano vzorčenje, vzorčenje v skupinicah, vzorčenje v več stopnjah, vzorčenje v več fazah, sistematično vzorčenje in kvota vzorčenja. Vsaka izmed teh metod odpravi to ali ono pomanjkljivost enostavnega vzorčenja.

Stratificirano vzorčenje

15.33 Zanesljivost ocene za aritmetično sredino, pa tudi drugih parametrov je bistveno odvisna od variabilnosti v osnovni populaciji in velikosti vzorca. To vidimo iz obrazcev o standardnih pogreških v tabeli 15.1. Na zanesljivost ocen moremo pri enostavnem vzorčenju vplivati edino tako, da zvečamo velikost vzorca. To pa je v vsakem primeru zvezano z občutnimi stroški, ker se zanesljivost vzorca veča le s kvadratnim korenom iz velikosti vzorca. Na variabilnost pojava pri enostavnem vzorčenju ne moremo vplivati. Vemo le, da je zanesljivost ocene tem boljša, čim bolj je populacija homogena, oziroma čim manjša je variabilnost.

Iz tega povzamemo, da bi bila ocena parametra za nehomogeno populacijo verjetno bolj zanesljiva, če bi populacijo razstavili v homogene dele - stratume in za vsak homogen del - stratum s samostojnim vzorcem ocenili parameter za stratum. Iz ocenin ocen za posamezne stratume pa bi sestavili skupno oceno za osnovno populacijo. Res se izkaže, da je zanesljivost ocene s stratificiranjem zelo povečana, če osnovno populacijo razdelimo v stratume po znaku, ki bistveno vpliva na proučevani pojav. Zelo uspešna je običajno regionalna stratifikacija, pri kateri razdelimo populacijo v stratume glede na geografske rejone. Ena od uspešnih regionalnih stratifikacij je razdelitev na mesto in odželo. Vendar so uspešne tudi stratifikacije po drugih znakih, odvisno od tega, kaj ocenjujemo.

Hiba stratificiranega vzorčenja je v tem, da moramo osnovno populacijo razmeroma dobro poznati, če hočemo izvesti razdelitev osnovne populacije v homogene dele. Ker običajno ne proučujemo popolnoma neznanih pojavov, vzamemo večkrat za osnovo stratifikacije prejšnje popise, vzorce ali druge vire.

Za ocene agregatov in standardnih pogrešk agregatov za stratificirano vzorčenje veljajo precej enostavni obrazci.

Ocena agregata s stratificiranim vzorčenjem X'_{str} je vsota ocen agregata v posameznih stratumih

$$X'_{str} = X'_1 + X'_2 + \dots + X'_p = \Sigma X'_k = \Sigma N_k \bar{x}_k \qquad (15.32)$$

Kvadrat standardne pogreške ocene agregata pa je vsota kvadratov standardnih pogrešk agregatov v posameznih stratumih

$$S\bar{x}^2_{X'_{str}} = S\bar{x}^2_{X'_1} + S\bar{x}^2_{X'_2} + \dots + S\bar{x}^2_{X'_p} = \Sigma S\bar{x}^2_{X'_k} \qquad (15.33)$$

Standardno pogreško v posameznih stratumih pa dobimo v tabeli 15.1. Stavek o seštevanju kvadratov standardnih pogrešk v posameznih stratumih velja, če so ocene v posameznih stratumih med seboj neodvisne.

Razmestitev enot vzorca po stratumih

V stratificiranem vzorcu ocenjujemo agregat v vsakem stratumu s samostojnim slučajnim vzorcem. Če poznamo skupno število enot v stratificiranem vzorcu n, je vprašanje, koliko enot od skupnega števila enot v vzorcu n vzamemo v posameznih stratumih.

Imamo več načinov glede razmestitve skupnega števila enot po stratumih. Po prvem je število enot v posameznem stratumu n_k proporcionalno skupnemu številu enot v posameznih stratumih N_k.

15.34 Za *proporcionalno razmestitev* določimo velikost vzorcev v posameznih stratumih n_k, če poznamo število enot v vzorcu n in skupno število enot po stratumih N_k po obrazcu

$$n_k = \alpha N_k \; ; \qquad \alpha = \frac{n}{N} \tag{15.34}$$

15.35 Vzemimo za primer ocenjevanje skupnega števila krav v občini z $N = 2185$ kmetijskimi gospodarstvi, ki smo jo razdelili v tri kmetijske predele *(A-B-C)*.

V tabeli 15.7 je nakazan izračun o številu anketiranih gospodarstev v posameznih stratumih, če je razmestitev enot proporcionalna in je enotni vzorčni delež $f_k = f = 0{,}30$. V tabeli 15.7 je izračunana tudi standardna pogreška stratificiranega vzorčenja.

Tabela 15.7 Proporcionalna razmestitev enot vzorca po stratumih za občino z $N = 2185$ gospodarstev, za vzorec s skupno $n = 655$ gospodarstev in izračun standardne pogreške ocene za skupno število krav

Stratum	N_k (1)	n_k (2)	f_k (3)	$\frac{1-f_k}{f_k}$ (4)	σ_k^2 (5)	$N_k\sigma_k^2$ (6)=(1).(5)	$SE^2_{\bar{X}_k}$ (7)=(4).(6)
A	855	257	0,30	2,333	4,66	3984	9297
B	705	211	0,30	2,333	1,84	1297	3027
C	625	187	0,30	2,333	0,93	581	1356
	2185 N	655 n					13680 $SE^2_{\bar{X}\,str.}$

Število enot v posameznih stratumih za proporcionalno razmestitev dobimo enostavno tako, da skupno število enot v stratumu pomnožimo s skupnim

vzorčnim deležem $f = n/N = 355/2185 = 0,30$.

Da ugotovimo uspešnost stratifikacije, izračunajmo še standardno pogreško za enostavno oceno. Ker poznamo skupno število gospodarstev $N = 2185$, vzorčni delež $f = 0,30$ in skupno varianco za število krav po gospodarstvih $\sigma^2 = 2,90$, je najprikladneje, da izračunamo kvadrat standardne pogreške agregata po obrazcu

$$SE_{X'}^2 = N\sigma^2(1-f)/f \qquad (15.35)$$

Obrazec je neposredno izpeljan iz obrazca za izračunavanje standardne pogreške agregata iz tabele 15.1. (Napravljen je nebistven popravek, ker smo namesto $N-1$ vzeli N). Zaradi enostavnosti ta obrazec uporabimo tudi pri izračunavanju pogrešk v stratumih pri vseh razmestitvah. Po tem obrazcu je

$$SE_{X'}^2 = 2185 \cdot 2,90 \cdot \frac{1-0,30}{0,30} = 14784$$

Iz rezultata vidimo, da je standardna pogreška ocene s stratificiranim vzorcem res manjša kakor pri enostavnem vzorčenju.

Uspeh stratifikacije pokaže kvocient med kvadratoma standardnih pogrešk ocen za enostavno in stratificirano vzorčenje. Ta kvocient

$$SE_{X'}^2 / SE_{X',str,p}^2 = 14784/13380 = 1,08$$

pove, da bi morali vzeti za osem odstotkov večji slučajni vzorec, če bi hoteli z njim dobiti tako zanesljive podatke, kakor smo jih dobili s stratificiranim vzorčenjem.

15.36 Optimalna razmestitev. Vendar proporcionalna razmestitev enot vzorca po posameznih stratumih ni najboljša. Že z vsebinsko analizo moremo sklepati, da je verjetno boljša razmestitev ona, ki upošteva tudi variabilnost podatkov v posameznih stratumih. Verjetno zadostuje v stratumih z majhno variabilnostjo manjši vzorec kakor v stratumih, kjer je variabilnost podatkov velika. Res se izkaže, da je standardna pogreška manjša, če pri razmestitvi enot po stratumih razen skupnega števila enot v stratumih upoštevamo tudi variabilnost. Najboljšo razmestitev enot po stratumih dobimo, če izračunamo velikost vzorca v stratumih po obrazcu

$$n_h = \alpha N_h \sigma_h \; ; \quad \alpha = \frac{n}{\Sigma N_h \sigma_h} \qquad (15.33)$$

15.37 Za skupno število krav iz prejšnjega primera je v tabeli 15.8 prikazano, kako izračunamo velikost vzorcev po stratumih za optimalno razme-

stitev. Razen tega je izračunana tudi standardna pogreška te ocene.

Tabela 15.8 Izračunavanje velikosti vzorcev in standardne pogreške pri optimalni razmestitvi enot po stratumih za število krav v občini z 2185 gospodarstev

Stratum	N_h (1)	σ_h (2)	$N_h \sigma_h$ (3)=(1).(2)	n_h (4)	f_h (5)	$1-f_h/f_h$ (3)	$N_h \sigma_h^2$ (7)	$SE_{\bar{X}}^2$ (8)=(3).(7)
A	855	2,16	1846,80	355	0,415	1,4096	3984	5616
B	705	1,36	958,80	185	0,262	2,8168	1227	3653
C	625	0,96	600,00	115	0,184	4,4348	581	2577
	2185		3405,60	355				11843
	N		$\Sigma N_h \sigma_h$	n				$SE_{str,o}^2$

Za prvi stratum dobimo število enot $n_1 = n \frac{N_1 \sigma_1}{\Sigma N_h \sigma_h} = 355 \frac{1843,80}{3405,30} = 355$ ita.

Primerjava dobljenega kvadrata standardne pogreške s kvadratom standardne pogreške za slučajen vzorec

$$SE_s^2 / SE_{str,o}^2 = 14784/11843 = 1,25$$

pokaže, da je uspešnost ocene pri optimalni razmestitvi veliko večja kot pri proporcionalni in moramo vzeti za 25 % večji enostaven vzorec, če hočemo, da bo ocena agregata tako zanesljiva, kot jo dobimo s stratificiranim vzorčenjem z optimalno razmestitvijo.

15.38 Optimalna razmestitev enot po stratumih pri različnih stroških. So pa še drugi elementi, ki vplivajo na razmestitev enot po stratumih. Dostikrat so stroški popisovanj za posamezno enoto zaradi terenskih ali drugih razlik med stratumi različni. Navadno ni toliko važno, koliko enot popišemo, kakor to, kolikšni so stroški za popisovanje. Zato ni nebistveno, če upoštevamo pri razmestitvi enot po stratumih tudi razlike v stroških popisovanja. Enostaven premislek pokaže, da je za skupno oceno koristneje, če v stratumih, v katerih je popisovanje dražje, anketiramo manj enot, ker za prihranjena sredstva anketiramo v stratumih, v katerih je popisovanje cenejše, večje število enot. Če pri določanju števila enot po stratumih upoštevamo še stroške za popisovanje ene enote v posameznih stratumih (c_h) dobimo, da je pri stalnih skupnih stroških za anketiranje C najboljša razdelitev enot v stratumih, če število enot po stratumih izračunavamo po obrazcu

$$n_k = \alpha \frac{N_k \sigma_k}{\sqrt{c_k}} \quad ; \quad \alpha = \frac{C}{\Sigma N_k \sigma_k \sqrt{c_k}} \qquad (15.37)$$

15.39 Za število goveje živine iz prejšnjega primera je izračun nakazan v tabeli 15.9

Tabela 15.9 Izračunavanje velikosti vzorcev in standardne pogreške ocene pri optimalni razmestitvi pri različnih stroških za število krav v občini z $N = 2185$ gospodarstvi

Stratum	N_k (1)	σ_k (2)	c_k (3)	$\frac{N_k \sigma_k}{\sqrt{c_k}}$ (4)	$N_k \sigma_k \sqrt{c_k}$ (5)=(3)(4)	n_k (3)	f_k (7)	$\frac{1-f_k}{f_k}$ (8)	$N_k \sigma_k^2$ (9)	SE_x^2 (10)=(8)(9)
A	855	2,13	200	130,98	26196	328	0,384	1,604	3984	6390
B	705	1,36	150	78,59	11788	197	0,279	2,584	1297	3351
C	625	0,96	100	60,00	6000	150	0,240	3,167	581	1840
	2185				43984	675				11581
	N				$\Sigma N_k \sigma_k \sqrt{c_k}$	n				$SE^2_{str,c}$

Da moremo rezultate nove razmestitve primerjati z razmestitvijo pri optimalni razmestitvi, smo vzeli, da so stroški za anketiranje pri prejšnji razmestitvi enaki stroškom pri tej razmestitvi. $C = \Sigma c_k n_k = 355.200 + 185.150 + 115.100 = 110250$ dinarjev, konstanta α pa je: $C/\Sigma N_k \sigma_k \sqrt{c_k} = 110250/43984 = 2,5033$. S to konstanto pomnožimo podatke v stolpcu (4) in dobimo n_k. Iz tabele 15.9 dobimo, da so stroški po novi razmestitvi enot po stratumih

$$C = \Sigma c_k n_k = 328.200 + 197.150 + 150.100 = 110150 \text{ dinarjev.}$$

Razlika sto dinarjev izvira iz tega, ker zaokrožimo število enot na celo. Če primerjamo optimalno razmestitev z optimalno razmestitvijo pri različnih stroških, vidimo, da za isti denar kot pri optimalni razmestitvi anketiramo v skupnem dvajset enot več. To dosežemo tako, da v najdražjem stratumu vzamemo 27 enot manj, za ta denar pa anketiramo v srednjem stratumu 12, v najcenejšem pa 35 enot več. Skupen učinek te premestitve se pokaže na zmanjšani skupni varianci ocene. Koeficient uspešnosti optimalne razmestitve pri različnih stroških dobimo, če primerjamo ustrezna kvadrata standardnih pogrešk

$$SE^2_{str,o} / SE^2_{str,c} = 11846/11581 = 1,023$$

Vzorčenje v skupinicah

15.40 Vendar stratifikacija ne odstrani eno izmed osnovnih tehničnih slabosti pri enostavnem vzorčenju. Če iz osnovne populacije izberemo na slučajen način vzorec, so te enote posamič razmeščene po vsej populaciji. To posebno nazorno opazimo, če so enote populacije razmeščene po terenu in je anketiranje teh enot razmeroma obsežno terensko delo. V posameznih primerih moramo iti v oddaljen kraj, da anketiramo eno samo enoto. Kljub temu, da je taka razmestitev za kakovost ocene ugodna, je iz tehničnih ozirov težko izvedljiva in draga. Na podobne težave pa naletimo tudi, če enote niso raztresene po terenu, ampak zbrane na enem in istem mestu. Iskanje popisnih listov, ki ustrezajo slučajnim številkam vzorca, v skladišču za popis prebivalstva, kjer je vskladiščenih več milijonov popisnih obrazcev, je tudi razmeroma težavno. Enako je pri kontroli kvalitete izdelkov drago, zaradi kontrole kvalitete enega artikla, odpirati in kvariti embalažo zavitka, v katerem je več artiklov.

Zato v dosti primerih uporabljamo postopek vzorčenja v skupinicah. Vzorčenje v skupinicah izvedemo tako, da pred izbiranjem vzorca združimo po nekaj enot v skupinice. V skupinice združujemo enote, ki so vezane po nekem merilu položaja. Tako združujemo na terenu v skupinice enote, ki so si krajevno blizu, n. pr. kmetijska gospodarstva v skupinice v okviru katastrskih občin, delavce v obrtnih podjetjih v skupinice po podjetjih, prebivalstvo v družine ali popisne okoliše, artikle v skupinice po paketih itd. Če namesto vzorca osnovnih enot – gospodarstev, delavcev, prebivalcev, artiklov itd. izberemo v vzorec skupinice, katastrske občine, podjetja, popisne okoliše, pakete itd., so v vzorcu skupinic osnovne enote združene tako, da jih tehnično laže in ceneje dosežemo. V določen kraj potujemo, da popišemo vsa gospodarstva v eni katastrski občini ali vse zaposlene v enem podjetju. V skladišču, kjer imamo vskladiščeno gradivo popisa prebivalstva, moramo pri enostavnem vzorcu poiskati določen popisni okoliš samo zato, da iz njega izvlečemo eno samo popisnico, pri vzorčenju v skupinicah pa pridejo v poštev vse popisnice tega okoliša. Enako pri kontroli kvalitete odpremo pri metodi skupinic paket, da pregledamo vse artikle, ne pa samo enega itd.

Če z x_{ki} zaznamujemo vrednost znaka, za katerega ocenjujemo agregat X, za i enoto v k skupinici, z $X_k = \sum_i x_{ki}$ vsoto podatkov v k skupinici, z M = skupno število skupinic, z m = število skupinic v vzorcu, ocenimo vrednost agregata X po obrazcu

$$X' = \frac{M}{m} \sum_k X_k \qquad (15.38)$$

Vzorčenje v skupinicah je tem bolj učinkovito, čim večje so razlike med enotami v skupinicah in čim manjše so razlike med skupinicami. To je ravno obratno kot pri stratificiranem vzorčenju, ker je stratifikacija

tem bolj uspešna, čim večje so razlike med stratumi in čim manjše so razlike v stratumih.

Vzorčenje v dveh ali več stopnjah

15.41 Če v vzorcu kmetijskih gospodarstev v FLRJ v skupinicah popišemo vsa kmetijska gospodarstva v izbranih občinah, je število enot v posameznih občinah toliko, da imamo po metodi skupinic razmeroma veliko dela s popolnim popisom v izbranih občinah. Zato ocenimo agregat za posamezne izbrane občine s samostojnimi enostavnimi vzorci gospodarstev. Ta način vzorčenja imenujemo vzorčenje v dveh stopnjah. Ime je metoda dobila po tem, ker izbiramo dva vzorca. V prvi stopnji izberemo vzorec občin, v drugi stopnji pa v izbranih občinah vzorce kmetijskih gospodarstev. Obrazec 15.38 za ocenjevanje agregata z vzorcem v skupinicah se za vzorčenje v dveh stopnjah toliko spremeni, da prave vrednosti agregatov X_k za skupinice zamenjamo z ocenami X'_k, ki jih dobimo z vzorci v drugi stopnji. Obrazci za ocenjevanje z vzorčenjem v dveh stopnjah so naslednji:

$$X'_k = \frac{N_k}{n_k} \sum_i x_{ki} \; ; \quad X'' = \frac{M}{m} \sum_k X'_k \qquad (15.39)$$

Iz vzorcev v drugi stopnji najprej ocenimo agregate za enote prve stopnje X'_k, iz teh pa po drugem obrazcu izračunamo oceno agregata X''.

15.42 Vzorčenje v dveh ali več stopnjah zelo pogosto uporabljamo. Čeprav je pri vzorčenju v več stopnjah v skupnem številu potrebno večje število osnovnih enot kot pri enostavnem vzorčenju za isto zanesljivost rezultatov, običajno dobiček v stroških popisa odtehta vse teoretične hibe vzorčenja v več stopnjah.

Prednost vzorčenja v več stopnjah pred enostavnim vzorčenjem se pokaže že pri izbiranju enot. Pri enostavnem vzorčenju potrebujemo za zgornji primer spisek vseh gospodarstev v FLRJ, ki rabi za okvir vzorčenja. Pri vzorčenju v dveh stopnjah pa potrebujemo spisek vseh občin, spiske kmetijskih gospodarstev pa samo za občine, ki so bile izbrane v prvi stopnji.

Podatke ocenjujemo tudi z vzorčenjem v več stopnjah. Vzorčenje v štirih stopnjah moremo uporabiti za ocenjevanje donosa za pšenico. V prvi stopnji izberemo enote prve stopnje – občine, v teh pa v drugi stopnji gospodarstva, ki goje pšenico. V izbranih gospodarstvih izberemo enote tretje stopnje – parcele, zasejane s pšenico, v teh pa enote četrte stopnje – kvadrate površin po 1 m^2. Oceno sestavimo po stopnjah podobno kakor za vzorčenje v dveh stopnjah.

Tudi odstotek pepela v premogu moremo za določen rudnik oceniti z vzorčenjem v več stopnjah. Iz tekoče proizvodnje premoga izberemo na slučajen način vagonček – enote prve stopnje. Iz vsakega izbranega vagončka

v drugi stopnji izberemo določeno število kosov premoga. Skupnost vseh teh kosov razdrobimo in dobro premešamo, da dobimo homogeno populacijo. Iz te mešanice izbiramo v tretji stopnji nove enote, n. pr. po 10 dkg premogovega prahu. Ta vzorec predstavlja populacijo celotne proizvodnje premoga. Odstotek pepela, ki ga dobimo iz tega vzorca, je ocena za odstotek pepela v celotni proizvodnji.

Pri vzorčenju v več stopnjah variabilnost podatkov v vsaki izmed stopenj prispeva k standardni pogreški ocene agregata. Čim večja je variabilnost podatkov v posameznih stopnjah, tem večja je tudi standardna pogreška skupne ocene.

Vzorčenje v dveh ali več fazah

15.43 Vzemimo, da za določeno področje po metodi razmerij ocenjujemo skupno letno proizvodnjo mleka v kmetijskih gospodarstvih.

Pri metodi razmerij moramo poznati vrednost agregata za neki podatek, ki je z ocenjevanim v čim večji korelaciji. Vprašanje je, ali moremo idejo metode razmerij ohraniti, če v našem primeru skupnega števila krav ne poznamo. Če analiziramo podatka število krav in letna proizvodnja mleka, moremo presoditi, da je za posamezno gospodarstvo veliko laže dobiti podatke o številu krav kakor o proizvodnji mleka. Zato se izkaže, da je pametno, da proizvodnjo mleka po metodi razmerij ocenimo z vzorčenjem v dveh fazah po obrazcu

$$X_f'' = Y' r'' = \frac{N}{n_1} {}_1y \frac{{}_2s_x}{{}_2s_y} \qquad (15.40)$$

Skupno število krav Y', ki ga ne poznamo, ocenimo z razmeroma velikim vzorcem (prva faza), ker je zanesljivost te ocene zaradi velike variabilnosti števila krav po gospodarstvih pri majhnem vzorcu majhna, zbiranje podatkov o številu krav pa ni zvezano s posebnimi težavami. Razmerje med proizvodnjo mleka in številom krav r'' pa cenimo po metodi razmerij z manjšim vzorcem gospodarstev, ker razmerje zanesljivo ocenimo tudi z majhnim vzorcem. Ta gospodarstva iz vzorca v prvi fazi izberemo v drugi fazi.

15.44 Idejo o vzorčenju v dveh fazah uporabimo tudi *pri stratificiranem vzorčenju.* Stratifikacija zahteva, da popolnoma poznamo populacijo po znaku, po katerem stratificiramo, da moremo opredeliti vsako enoto osnovne populacije v enega izmed stratumov. To pa je včasih zelo obsežno in težavno delo. Zato ocenimo stratificirano oceno z vzorčenjem v dveh fazah. Da ni treba celotno osnovno populacijo razdeljevati v stratume, vzamemo v prvi fazi razmeroma velik vzorec enot, za katerega pa zberemo samo podatke o znaku, po katerem populacijo stratificiramo. Iz podatkov vzorca v prvi fazi ocenimo po obrazcu za enostavno ocenjevanje število enot z značilnostjo posameznega stratuma N_k'. Iz vzorca v prvi fazi, ki je dal strukturo po stratumih, izberemo v posameznih stratumih samostojne vzorce v dru-

gi fazi. Za enote druge faze pa šele zberemo podatke, za katere ocenjujemo agregat. Iz vzorcev v drugi fazi ocenimo za vsak stratum oceno aritmetične sredine $\bar{x}''$, skupno oceno pa sestavimo po obrazcu

$$X''_{f,\,str} = \sum_k N'_k \bar{x}''_k = \sum_k N \frac{n'_k}{n} \frac{S_{x_2}}{n''_k} \qquad (15.41)$$

15.45 To metodo dostikrat uporabljamo pri anketiranju, da korigiramo podatke zaradi *neodgovorov*. Pri večini anket določeno število enot, ki so izbrane v vzorcu, iz katerega koli vzroka ne odgovori na anketo. Odstotek neodgovorov je bistveno odvisen od izvedbe ankete. Če n.pr. anketirane osebe prosimo, da po pošti pošljejo odgovore na stavljena vprašanja, je odstotek oseb, ki na anketo ne odgovore, znatno večji, kakor če anketiramo z anketarji, ki z osebnim stikom pridejo do podatkov. Enako je odstotek neodgovorov odvisen od vsebine enkete. Na vprašanja o intimnih ali premoženjskih razmerah imamo v splošnem velik odstotek neodgovorov.

Popolnoma napačno ravnamo, če sklepamo takole: Ker računamo, da n.pr. v anketi, ki naj zajame 1000 oseb, 20 % oseb ne bo poslalo odgovora, vzamemo v vzorec 1250 oseb in dobimo odgovore za želeno število oseb. Populacija oseb, ki ne odgovori na anketo, je po svojih lastnostih verjetno bistveno različna od populacije, ki odgovori. Če delamo na zgornji način, dobimo oceno, ki ne velja za celo populacijo, marveč za tisti del prebivalstva, ki je voljan odgovoriti na anketna vprašanja. Ker pa hočemo dobiti oceno za celo populacijo, to je tudi za tisti del, ki n.pr. ni voljan odgovarjati na poštne ankete, je pravilen postopek takle. Izvedemo vzorec z n_1 enotami. Te enote anketiramo in od njih za n_1^+ enot dobimo odgovore, za n_1^- pa ne. Tako razen podatkov za n_1^+ enot, za katere je vsota podatkov $S_{x_1}^+$, dobimo še stratifikacijo tako, da imamo v prvi grupi osebe, ki odgovarjajo na anketna vprašanja, v drugi grupi pa osebe, ki ne odgovarja na poštne ankete. Iz vzorca tistih n_1^- ki niso odgovorili izberemo v drugi fazi vzorec z n_2^- enotami. Zanje pa z drugimi sredstvi kakor pri prvi anketi skušamo zbrati podatke. Tako se na te osebe ne obračamo po pošti, marveč osebno poskušamo priti do podatkov s pozivedtami ita. Vsoto podatkov za teh n_2^- enot zaznamujemo z $S_{x_2}^-$. Skupno oceno za agregate iz tega dvofaznega vzorca dobimo

$$X''_{f,\,str} = \frac{N}{n_1}\left(S_{x_1}^+ + S_{x_2}^-/f_2^-\right) \qquad (15.42)$$

Pri tem je razen znanih izrazov, $f_2^- = n_2^-/n_1^-$ vzorčni delež enot v drugi fazi v populaciji neodgovorov.

Sistematično vzorčenje

15.46 Tehnika izbora enot vzorca s tablicami slučajnih števil je posebno za velike populacije precej neprikladna. Zato dostikrat uporabljamo sistematičen izbor enot v vzorcu. Ta da v določenih okoliščinah zadovoljive, včasih pa celo boljše rezultate kakor čisto slučajni vzorec. Pri sistematičnem izboru vzamemo v vzorec n.pr. vsako deseto enoto po spisku, če je vzorec desetodstotni. Ta izbor je sila preprost. Zanj je potreben samo utrjen spisek enot, ki je okvir vzorčenja. Prvo enoto izberemo slučajnostno izmed slučajnih števil, ki so manjša od stopinje izbora. Stopinja izbora je število enot, ki jih v zaporedni vrsti preskočimo od ene izbrane enote do naslednje enote, ki jo izberemo v vzorec.

Stopinjo izbora s določimo po obrazcu

$$s = 1/f \qquad (15.43)$$

če poznamo vzorčni delež f ali po obrazcu

$$s = N/n \qquad (15.44)$$

če imamo dano število enot v vzorcu.

Običajno stopinja s ni celo število. V teh primerih vzamemo za stopinjo najbližje celo število. Na koncu sistematičnega izbora pa s ponovnim sistematičnim ali slučajnim izborom izključimo odvečne ali dopolnimo manjkajoče število enot. Če hočemo n.pr. v našem primeru občine z $N = 2185$ gospodarstvi izbrati v vzorec $n = 355$ gospodarstev, je stopinja ustreznega sistematičnega vzorca

$$s = 2185/355 = 3,3$$

Če vključimo v vzorec vsako tretjo enoto po spisku, dobimo v vzorec $n = 728$ enotami namesto 355 enot. Ker smo dobili v vzorec 73 enot preveč, moremo iz teh 728 izbranih enot s sistematičnim izborom s stopinjo

$$s = 728/73 = 10$$

izločiti odvečne enote.

15.47 Če je populacija sestavljena iz več delnih populacij oziroma spiskov enot, izberemo s slučajnim izborom samo prvo enoto v prvi delni populaciji oziroma spisku; prve enote v vseh naslednjih populacijah pa so določene z ostankom iz predhodne - delne populacije oziroma spiska. Če izbiramo n.pr. vzorec s stopinjo 10 in v prvem spisku določimo zadnjo enoto vzorca, ostanejo pa v tem spisku še štiri enote, v naslednjem spisku začnemo s peto enoto, tako da je stopinja med zadnjo izbrano enoto v enem spisku in prvo izbrano enoto v naslednjem spisku zopet 10.

Za površine je sistematičen izbor določen s parcelami, ki so enako oddaljene druga od druge. Ta postopek dostikrat uporabljamo v kmetijski statistiki, v gozdarstvu itd.

15.48 Shematičen prikaz regionalnega sistematičnega vzorca je dan v sliki 15.3.

Sistematičen izbor je posebno prikladen pri decentraliziranem izboru vzorca, ker je lahko razumljiv in izvedljiv.

Vendar ima sistematičen izbor svoje hibe, zaradi katerih moramo v vsakem posebnem primeru podrobno analizirati zaporedje enot, preden se odločimo za sistematičen izbor.

Zavedati se namreč moramo, da sistematičen izbor ni slučajen, zato moremo z njim nadomestiti slučajen izbor samo v določenih pogojih.

Zaradi ilustracije bomo nakazali shematičen primer, v katerem pretiramo učinek, ki more nastati pri sistematičnem izboru.

Vzemimo, da smo v popisu prebivalstva v kolektiven obrazec vpisovali podatke o posameznih prebivalcih po vrsti po družinah. Vse popisane družine so petčlanske v temle sestavu: oče, mati in trije otroci. Mislimo si, da so družinski člani navedeni v spisku po vrsti: oče, mati in otroci po starosti. V spisku imamo v daljši vrsti prebivalce razporejene v naslednjem vrstnem redu:

MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○MŽ○○○

Če iz tega spiska izberemo sistematičen vzorec s stopnjo $s = 10$ in začnemo n.pr. s sedmo enoto po spisku, dobimo izbrane enote, ki so v zgornji vrsti podčrtane.

Izbrali smo samo ženske oziroma matere. Seveda tak izbor ni reprezentativen in ga ne moremo vzeti za osnovo za ocenjevanje podatkov o prebivalstvu. Sistematičen izbor je dal neuporabne podatke zato, ker je v zaporedju enot v spisku periodično ponavljanje. Perioda ponavljanja pa je ubrana s stopinjo izbora. V praktičnih primerih učinek sistematičnega izbora ni tako drastičen, ker nimamo take periodičnosti v stvarnih spiskih, vendar more biti ta pojav, čeprav v milejši obliki, vir velike pristranosti v rezultatih ocenjevanja. Zato je treba pred uporabo sistematičnega izbora temeljito proučiti okvir vzorčenja in v primeru, da se pokaže zgornji učinek, opustiti sistematičen izbor in ohraniti slučajni izbor enot.

Omenjeni učinek more nastati tudi pri regionalnem vzorčenju, ker morejo tudi na terenu nastopiti periodični pasovi boljših ali slabših pogojev v vegetaciji in tako vplivati na kvaliteto ocene.

Kvotno vzorčenje

15.49 Kljub najrazličnejšim izboljšanim postopkom mora biti vzorec razmeroma velik, da dosežemo zadovoljive rezultate. Kakor smo videli iz

prejšnjih primerov, smo pri popisu prebivalstva za vzorec predhodnih. podatkov v Sloveniji vzeli ca 30000 enot, da smo dosegli sorazmerno zanesljivost podatkov ca 1 %.

Dostikrat pa nimamo možnosti, da bi ocenjevali populacijo z velikimi vzorci. Veliki vzorci zahtevajo, čeprav so samo majhen del populacije, dosti časa in stroškov. S slučajnimi vzorci nekaj sto enot pa so ocene tako nezanesljive, da so v večini primerov neuporabne za namene, za katere jih sestavljamo. Tipičen primer so ankete javnega mnenja analize tržišča ita. Za reševanje teh problemov vzorčenje ne sme biti preobširno, rezultati pa morajo biti zadosti zanesljivi, da jih moremo analizirati.

V takih primerih uporabljamo za ocenjevanje metodo kvotnega vzorca. Kvotni vzorec nastopa na predpostavk slučajnega vzorčenja. Zato kvotni vzorec nima kvalitet, ki jih ima slučajen vzorec. Ima pa druge kvalitete. Za rezultate, ki jih dobimo s kvotnim vzorcem, ne moremo sicer določiti standardne pogreške ocene. Glede na to, kako je ocena sestavljena, pa moremo sklepati, da je dobljena ocena boljša, kakor bi jo dobili s slučajnim vzorcem, ki bi bil enako velik. Princip kvotnega vzorca je v tem, da enote vzorca zavestno izberemo tako, da je struktura vzorca, ki rabi za ocenjevanje, za določene ključne znake enaka kot v celotni populaciji. Kvotni vzorec mora biti po strukturi enak populaciji v tistih znakih, ki so bistveni faktorji za podatke, ki jih ocenjujemo. Tako vzamemo pri anketah o javnem mnenju za osnovo kvotnega vzorca strukturo po starosti, poklicu, kraju ita. Iz strukture za populacijo moremo glede na skupno število enot vzorca določiti, koliko oseb, ki jih anketiramo (kvota) mora imeti določeno starost, poklic in biti iz določenega kraja. Po tej opredelitvi običajno anketar ne izbira posamezne osebe slučajnostno, marveč pustimo, da svobodno izbira osebe, ki danim pogojem zadoščajo. Ta vzorec ima v osnovi precej elementov stratifikacije s proporcionalno razmestitvijo.

15.50 Kot poskus, kako je kvotni vzorec uporaben, so v Zavodu za statistiko LRS po principu kvotnega vzorca iz gradiva popisa živine na dan 15. I. 1953 izbrali $n' = 700$ kmetijskih gospodarstev. Gospodarstva so bila izbrana tako, da je struktura vzorca po okrajih in velikostnih skupinah ustrezala strukturi v celotni populaciji. Ta dva znaka smo šteli za dva izmed bistvenih faktorjev, ki vplivajo na število živine. V Tabeli 15.9 je prikazana razmestitev izbranih gospodarstev po tedanjih okrajih in po velikostnih skupinah. Ta razmestitev je bila glede na strukturo v celotni populaciji postavljena vnaprej. Iz gradiva o popisu živine pa so bila izbrana gospodarstva, ki so ustrezala tej strukturi.

Tabela 15.10 Razmestitev n = 700 gospodarstev v poizkusni anketi po okrajih in velikostnih skupinah

Okraj	Skupno	Skupna površina v ha										
		a	0,5	-1	1-2	2-3	3-5	5-8	8-10	10-15	15-20	20-
LR Slovenija	700	16	10	51	76	73	120	112	53	88	45	56
Celje mesto	2	-	-	-	-	-	1	1	-	-	-	-
Celje okolica	72	1	1	5	9	8	12	12	6	9	4	5
Črnomelj	18	-	-	1	1	1	1	3	2	5	2	2
Gorica	44	1	1	4	7	5	9	7	3	4	1	2
Kočevje	19	1	-	1	1	1	2	3	2	4	2	2
Kranj	24	-	-	1	1	2	3	3	2	4	3	5
Krško	48	1	-	4	5	5	9	9	4	7	2	2
Ljubljana mesto	8	-	-	1	1	1	1	1	1	1	1	-
Ljubljana okolica	61	1	1	4	4	4	8	9	5	10	6	9
Ljutomer	35	5	-	4	5	5	7	4	1	2	1	1
Maribor mesto	3	-	-	-	1	1	1	-	-	-	-	-
Maribor okolica	55	4	1	4	7	6	10	8	3	5	3	4
Murska Sobota	79	1	1	5	12	14	23	15	3	4	1	-
Novo mesto	47	-	-	4	5	4	7	7	5	8	4	3
Postojna	26	-	-	1	1	1	3	4	3	5	4	4
Ptuj	61	1	3	9	10	7	10	9	4	5	2	1
Radovljica	10	-	-	-	-	1	1	2	1	2	1	2
Sežana	27	-	-	2	2	2	5	6	3	4	2	1
Slovenj Gradec	14	-	1	-	1	1	1	1	1	2	1	5
Šoštanj	12	-	-	-	1	1	1	2	1	2	1	3
Tolmin	25	-	1	1	2	2	4	5	2	3	2	2
Trbovlje	10	-	-	-	-	1	1	1	1	2	2	2

a = gospodarstva kmetijskih delavcev

V tabeli 15.11 so prikazani rezultati kvotnega vzorca za strukturo gospodarstev po številu konj, govedi in prašičev v primerjavi s podaki popolne obsežne popisa. Glede na to, da je vzorec majhen, so rezultati zelo natni.

Tabela 15.11 Primerjalna slika med rezultati ankete po kvotnem vzorcu (A) in rezultati popisa (P) za strukturo o številu gospodarstev po številu konj, govedi in prašičev (v tisočih)

Število konj	Število gospodarstev	
	P	A
Skupaj	146,6	146,6
0	110,8	108,9
1	17,8	16,8
2	14,2	16,1
3	3,2	4,2
4-	0,7	0,6

Število govedi	Število gospodarstev	
	P	A
Skupaj	146,6	146,7
0	15,0	16,6
1	23,7	23,5
2	33,7	33,7
3	28,1	25,1
4	18,9	22,6
5	11,3	10,3
6-	15,9	14,7

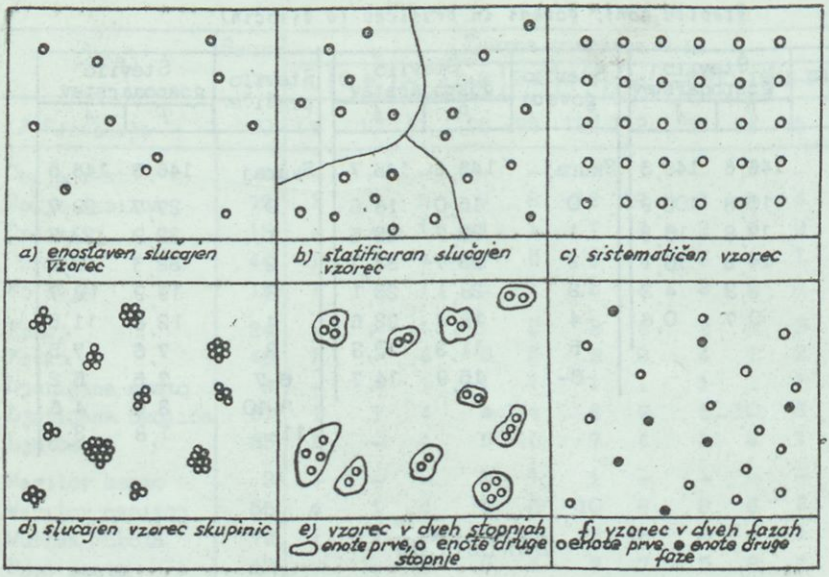
Število prašičev	Število gospodarstev	
	P	A
Skupaj	146,6	146,5
0	27,7	29,7
1	32,9	32,7
2	32,3	32,3
3	19,9	19,7
4	12,8	11,5
8	7,6	7,5
6-7	7,5	5,2
8-10	8,9	4,6
11-	1,8	3,4

ZAKLJUČEK

15.51 Kakor vidimo, imamo dvoje vrst ocen: enostavne ocene in ocene po metodi razmerij: dvoje vrst izborov: čisti slučajni izbor in sistematičen izbor in več vrst vzorčenj: enostavno vzorčenje, stratificirano vzorčenje, vzorčenje v več stopnjah in vzorčenje v več fazah. V praktičnih primerih glede na pojav, ki ga proučujemo, vzamemo tako kombinacijo vseh zgornjih elementov, ki da najugodnejšo izvedbo in najučinkovitejši rezultat. Tako moremo oceniti po metodi razmerij s stratificiranim vzorcem v dveh stopnjah, v katerem je izbor enot v prvi stopnji slučajen, v drugi stopnji pa sistematičen, potrošnjo za kulturno in družbeno življenje delavcev.

Po tem načrtu moremo oceniti, koliko trošijo n.pr. delavci za kulturno in družbeno življenje. Stratuma sta n.pr.: podjetja v mestu in podjetja na deželi. V prvi stopnji v posameznem stratumu s slučajnim vzorcem izberemo podjetja, v teh pa s sistematičnim vzorcem vsakega desetega delavca. Za osnovo ocene po metodi razmerij uporabimo plačo, ker predpostavljamo, da so stroški za kulturno in družbeno življenje odvisni od plače, razpolagamo pa po plačnih spiskih s skupnim plačnim fondom za podjetja. Obrazec za to oceno je

$$\bar{Y}' = \sum \frac{M}{m} {}_k SY_k \frac{{}_i Sx_{ki}}{{}_i Sy_{ki}} \qquad (15.45)$$



Slika 15.3 Regionalna razmestitev enot vzorca pri različnih vrstah vzorčenja

Pri tem pa pomeni: x_{ki} = strošek za kulturno življenje izbranega delavca i v podjetju k; y_{ki} = plača delavca i v podjetju k; $\bar{S}_i$ = seštevanje podatkov vzorca v podjetju Y_k = plačni fond izbranega podjetja k; S_k = seštevanje podatkov za vzorec podjetij, M = število podjetij v stratumih, m = število podjetij v vzorcu v stratumu, Σ = seštevanje podatkov po stratumih.

Razen navedenih vrst vzorčenja imamo še druge vrste vzorčenj, ki izkoriščajo še različne druge lastnosti populacij in vzorcev. Vendar te metode presegajo naš okvir in jih zato ne navajamo. Za obravnavane vrste vzorčenja imamo v sliki 15.3 nazorno prikazano, kakšna je geografska razmestitev pri posameznih vrstah vzorčenja. Te slike nakažejo bistvene značilnosti za posamezne vrste vzorčenj in razlike med njimi.

Šestnajsto poglavje

VZORČENJE - MALI VZORCI

16.1 Pri velikih vzorcih je situacija ugodna. Ugotovili smo namreč, da se za večino parametrov ocene za velike vzorce distribuirajo približno v normalnih distribucijah. Zato so načela za ocenjevanje za večino parametrov enotna. Problem pa se zaplete pri majhnih vzorcih.

Za male vzorce so odkrite zakonitosti le za vzorce iz populacij, ki zadoščajo določenim pogojem, ne pa za populacije na splošno. Razen tega se različni izrazi, izračunani iz podatkov vzorcev za male vzorce ne distribuirajo vsaj približno normalno, temveč v različnih, včasih zamotanih distribucijah. Ne glede na vse to pa so ocene iz malih vzorcev še nezanesljive in imajo zato omejen praktičen pomen. Zato za male vzorce včasih ne ocenjujemo parametre s točkovnimi ocenami in razmaki zaupanja, temveč z njimi le preizkušamo določene hipoteze o proučevanih populacijah. O preizkušanju hipotez je govora v posebnem poglavju.

TEORETIČNE DISTRIBUCIJE ZA MALE VZORCE

16.2 Razen v normalni distribuciji, ki je osnova tako za velike kot za male vzorce, se posamezni izrazi za male vzorce distribuirajo v različnih verjetnostnih distribucijah. Za nekatere od njih, ki so v praksi vzorčenja pogoste in jih v nadaljevanju uporabljamo, navajamo njihove osnovne značilnosti.

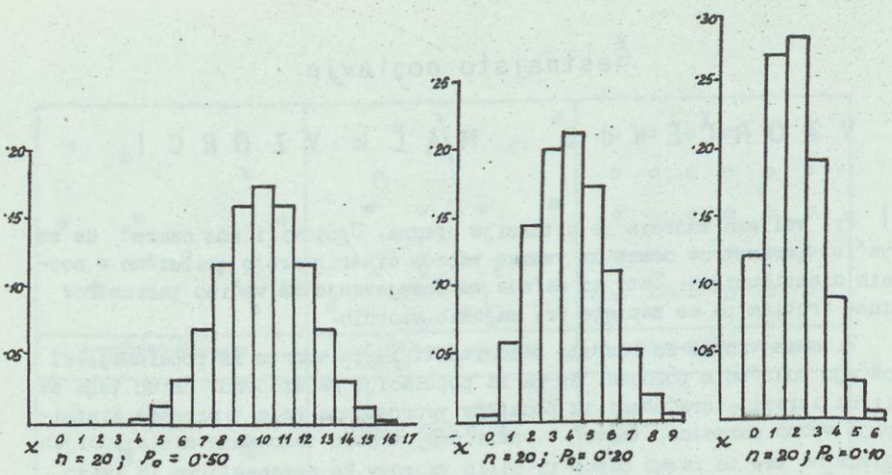
Od teoretičnih distribucij za male vzorce bomo omenili tele nezvezne distribucije: binomialno, hipergeometrično, Poissonovo distribucijo. Od zveznih teoretičnih distribucij pa so najpomembnejše: standardizirana normalna distribucija, t-distribucija, χ^2-distribucija in F-distribucija.

Binomalna distribucija

16.3 Če iz populacije, v kateri je delež enot z dano značilnostjo enak P_0, izvlečemo slučajnostni vzorec z n enotami, je verjetnost, da je v vzorcu x enot s to značilnostjo, enaka

$$B(P_0;n;x) = \binom{n}{x} P_0^x (1-P_0)^{n-x} \qquad (16.1)$$

Po tem obrazcu moremo izračunati verjetnosti za vse vrednosti x od 0 do n. Tako dobimo distribucijo verjetnosti.



Slika 16.1 Binomialna distribucija

Binomialna distribucija je simetrična, če je $P_0 = 0{,}50$. Če pa P_0 ni $0{,}50$, je asimetrična tem bolj, čim bolj je P_0 od $0{,}5$ različen. To lepo vidimo iz slike 16.1

Za binomialno distribucijo je matematično upanje x za proporc $p = x/n$ enako

$$M_p = E\left(\frac{x}{n}\right) = \sum_0^n \frac{x}{n} B(P_0; n; x) = P_0 \qquad (16.2)$$

varianca pa

$$\sigma_p^2 = \frac{P_0(1-P_0)}{n} \qquad (16.3)$$

Ocena proporca z malim vzorcem s ponavljanjem je torej nepristrana ocena za delež v populaciji P_0. Obrazec za standardno pogreško ocene pa je enak kot pri velikih vzorcih.

Hipergeometrična distribucija

16.4 Če iz končne populacije z N enotami, v kateri ima N_1 enot določeno značilnost, izberemo slučajnostni vzorec brez ponavljanja z n enotami, je verjetnost, da je v vzorcu x enot, ki imajo to značilnost, enaka

$$H(N; N_1; n; x) = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} \qquad (16.4)$$

Po tem obrazcu moremo izračunati verjetnosti za vse vrednosti x od 0

do n. Tako dobimo distribucijo verjetnosti, ki jo imenujemo hipergeometrično distribucijo. Ta je binomialni tem bolj podobna, čim večji je obseg populacije N. Za hipergeometrično distribucijo je matematično upanje za slučajnostno spremenljivko $p = x/n$ enako

$$M_p = E\left(\frac{x}{n}\right) = \sum_0^n \frac{x}{n} H(N;N_1;n;x) = \frac{N_0}{N} = P_0 \qquad (16.5)$$

varianca za $p = x/n$ pa je enaka

$$\sigma_p^2 = \frac{P_0(1-P_0)}{n} \cdot \frac{N-n}{N-1} \qquad (16.3)$$

Ocena proporca z malim vzorcem brez ponavljanja je torej tudi nepristrana ocena proporca v populaciji P_0. Če primerjamo σ_p^2 obrazcu 13.3 z ustreznim obrazcem v tabeli 15 1, spoznamo, da se skiada z varianco strukturnega deleža z velikim vzorcem brez ponavljanja.

Poissonova distribucija

16.5 Če je v populaciji delež P_0 za določeno značilnost zelo majhen, n pa velik, je verjetnost, da z vzorcem n enot dobimo x enot s to značilnostjo, enaka

$$P(\alpha;x) = e^{-\alpha}\frac{\alpha^x}{x!} \qquad (16.7)$$

Po tem obrazcu moremo izračunati verjetnosti za vse x od 0 navzgor. Verjetnostno distribucijo, ki jo tako dobimo, imenujemo Poissonovo distribucijo. Poissonova distribucija velja za pojave, ki se sorazmerno redko pojavljajo.

Za Poissonovo distribucijo velja za število enot z dano značilnostjo x

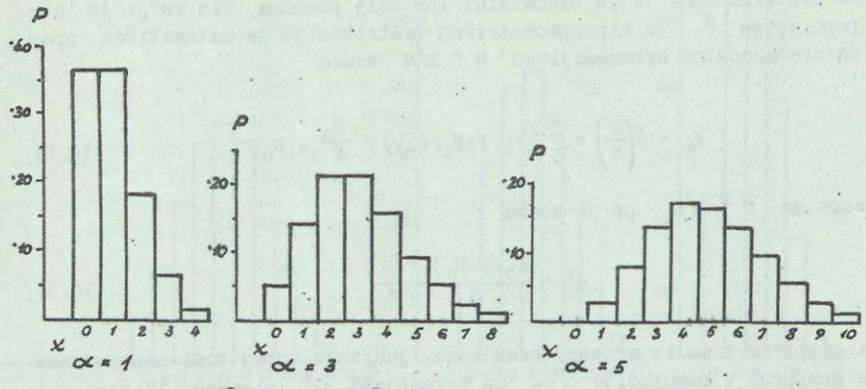
$$M_x = E(x) = \sum_0^\infty xP(\alpha;x) = \alpha\ ; \quad \sigma_x^2 = \alpha \qquad (16.8)$$

Za redke dogodke je x nepristrana ocena za α. Poissonova distribucija je teoretično izpeljana pod predpostavko, da je limita produkta števila enot n in deleža P_0

$$\lim nP_0 = \alpha \qquad (16.9)$$

če gre n čez vse meje, P_0 pa proti nič. Vendar je Poissonova distribucija zelo dober približek tudi za populacije, v katerih je proporc majhen, vendar končen (do 0,05).

V sliki 13.2 je narisana Poissonova distribucija za $\alpha = 1$, $\alpha = 3$ in $\alpha = 5$.



Slika 16.2 Poissonova distribucija

Stopinje prostosti

16.6 Vse zvezne distribucije za male vzorce, ki jih bomo prikazali, so odvisne od števila enot v vzorcu, oziroma natančneje, od števila stopinj prostosti v vzorčnem sistemu. Zato najprej obravložimo pojem stopinj prostosti.

Če je $n = 5$ vrednosti: x_1, x_2, x_3, x_4, x_5 vezanih na pogoj, da je aritmetična sredina

$$\bar{x} = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) \qquad (16.10)$$

konstantna, je variabilnost teh petih vrednosti omejena. Od petih vrednosti samo štiri svobodno variirajo, medtem ko je peta določena z zvezo v obrazcu 16.10. Ta sistem vrednosti ima torej štiri stopinje prostosti, ker je peta vrednost določena, če poznamo štiri vrednosti. Če zgornjim petim vrednostim dodamo še pogoj, da je zanje ocena variance

$$s_x^2 = \frac{S(x-\bar{x})^2}{n-1} \qquad (16.11)$$

stalna, se število stopinj prostosti zniža še za eno in imamo samo tri stopinje prostosti.

16.7 Vzemimo, da so v vzorcu štiri skupine po pet vrednosti. Vsaka skupina po pet vrednosti pa je vezana s stalno grupno aritmetično sredino. Po zgornjem pravilu imamo v vsaki grupi po pet vrednosti, a le štiri stopinje prostosti, ker je od petih podatkov eden določen z grupno aritmetično sredino. Ker imamo štiri grupe, v vsaki pa po štiri stopinje prostosti, je

skupno število stopinj prostosti 4.(5-1) = 16.

16.8 Vzemimo za nadaljnji primer kontingenčno tabelo 4 × 3 za število pretrgov na prstančevih strojih v eni izmed tekstilnih tovarn v Sloveniji za štiri številke preje po vrstah napak.

Tabela 16.1 Število pretrgov na prstančevih strojih za številke Nm 14, 34, 40, 50 po vrstah napak (M = mehanske; P = posluževalne; O = ostale. (Vir: Operativna evidenca ene izmed tekstilnih tovarn v LRS)

Nm preje	Vrsta napak			Skupaj
	M	P	O	
14	12	48	5	65
34	45	258	58	331
40	100	88	26	214
50	55	41	12	108
Skupaj	212	435	101	748

Če vzamemo, da morejo v tabeli 16.1 frekvence v posameznih poljih svobodno variirati, se prosto spreminja $4 \times 3 = 12$ podatkov. Če pa ta sistem podatkov omejimo in postavimo pogoj, da smejo frekvence variirati le tako, da ostanejo robne frekvence iste, se število podatkov, ki svobodno variirajo, zmanjša. V prvem stolpcu imamo samo tri stopinje prostosti, ker so frekvence vezane s pogojem, da je skupna vsota frekvenc enaka 212. Enako je z drugim stolpcem, v katerem so zaradi istega vzroka tudi samo tri stopinje prostosti. V zadnjem stolpcu pa ne more svobodno variirati nobena frekvenca, ker je v vsaki vrsti tretja frekvenca določena s pogojem, da so tudi vsote frekvenc po vrstah stalne. Tako imamo v tem sistemu le $3 + 3 = 6$ stopinj prostosti, ker je od skupnega števila $4 \times 3 = 12$ podatkov $2 + 4 = 6$ vezanih na pogoj, da morajo ostati robne frekvence stalne. V splošnem je za kontingenčno tabelo $k \times g$ število stopinj prostosti: $m = (k-1).(g-1)$.

Iz navedenih primerov vidimo, da za določen sistem na splošno ugotovimo število stopinj prostosti preprosto tako, da število vseh podatkov, ki variirajo, zmanjšamo za število vezi med njimi, ki so postavljene kot pogoj in omejujejo svobodno variiranje vzorčnih podatkov.

Standardizirana normalna distribucija

16.9 Zvezne verjetnostne distribucije opisujemo z gostoto relativne frekvence ali z gostoto verjetnosti.

Tako je za standardizirano normalno distribucijo $N(0, 1)$ gostota verjetnosti

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad (13.12)$$

To je stvarno funkcijska zveza med normiranim odklonom in ordinato verjetnostne krivulje. Standardizirano normalno distribucijo smo podrobno obravnavali v štirinajstem poglavju.

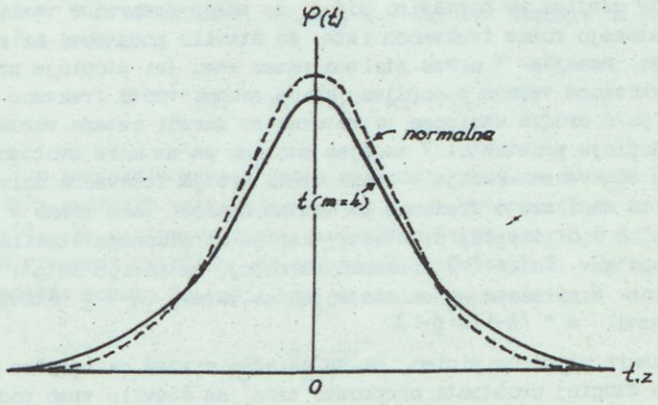
t-distribucija

16.10 t-distribucija je simetrična, unimodalna, zvonasta distribucija in je normalni zelo podobna.

Gostota verjetnosti za t-distribucijo $\varphi(t)$ je dana s funkcijo

$$\varphi(t) = C_t\left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \qquad (13.13)$$

pri tem je C_t konstanta, ki je odvisna od števila stopinj prostosti m. Iz obrazca 13.13 vidimo, da je t-distribucija odvisna od stopinj prostosti m.



Slika 16.3 t-distribucija

V sliki 13.3 je narisana t-distribucija za $m = 4$ in standardizirana normalna distribucija, proti kateri se bliža t-distribucija, če se število stopinj prostosti veča. Vse t-distribucije, za katere je m večji kot štiri, leže med narisanima krivuljama.

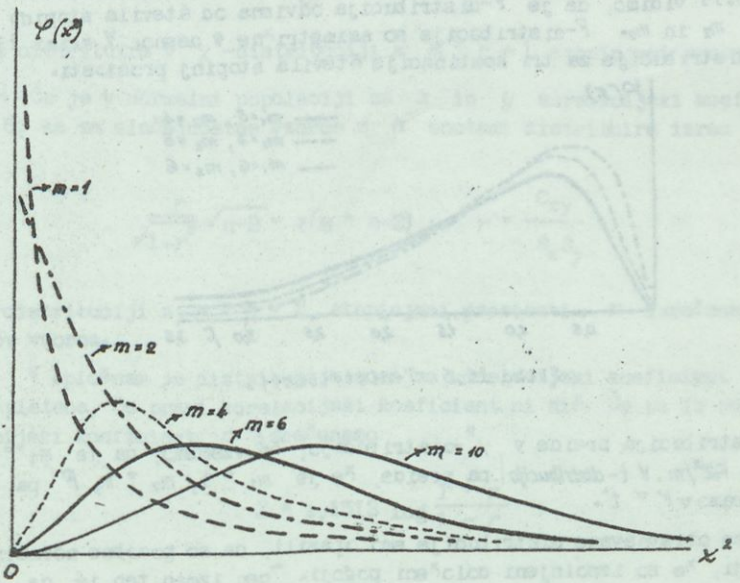
χ^2 - distribucija

16.11 Od števila stopinj prostosti je odvisna tudi χ^2-distribucija. Gostota verjetnosti za χ^2-distribucijo je dana s funkcijo

$$\varphi(\chi^2) = C_{\chi^2}(\chi^2)^{\frac{m-2}{2}} e^{-\chi^2/2} \qquad (16.14)$$

Pri tem je C_{χ^2} konstanta, ki je odvisna od števila stopinj prostosti.

χ^2-distribucija je unimodalna distribucija, ki je asimetrična v desno. V sliki 13.4 so prikazane χ^2-distribucije za več stopinj prostosti ($m=1$, $m=2$, $m=4$, $m=3$ in $m=10$).



Slika 16.4 χ^2-distribucija

Iz slike 13.4 vidimo, da je χ^2-distribucija vedno manj asimetrična, čim večje je število stopinj prostosti in je za $m = 10$ že precej podobna normalni distribuciji. Res se χ^2-distribucija, če večamo število stopinj prostosti, vedno bolj približuje normalni distribuciji. Za χ^2-distribucije, ki imajo nad trideset stopinj prostosti moremo vzeti, da se izraz

$$\sqrt{2\chi^2} - \sqrt{2m-1} = z \qquad (16.15)$$

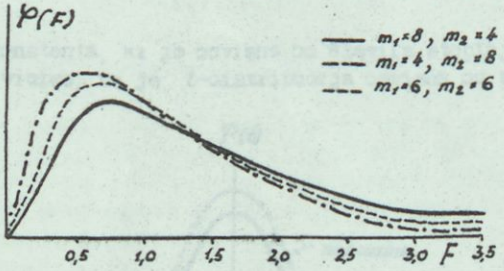
distribuira standardizirano normalno.

F-distribucija

16.12 Če avojnega števila stopinj prostosti je odvisna F-distribucija. Zanjo je gostota verjetnosti

$$\varphi(F) = C_F F^{\frac{m_1-2}{2}} \left(1 + \frac{m_1}{m_2} F\right)^{-\frac{m_1+m_2}{2}} \qquad (16.16)$$

Pri tem je C_F konstanta, ki je odvisna od števila stopinj prostosti. Iz obrazca 16.16 vidimo, da je F-distribucija odvisna od števila stopinj prostosti m_1 in m_2. F-distribucije so asimetrične v desno. V sliki 16.5 imamo F-distribucije za tri kombinacije števila stopinj prostosti.



Slika 16.5 F-distribucija

F-distribucija preide v χ^2-distribucijo, če vzamemo, da je $m_1 = m$ in $m_2 = \infty$ *in* $F\chi^2/m$. *V t-distribucijo* pa preide, če je $m_1 = 1$, $m_2 = m$, F pa transformiramo v $F = t^2$.

Za vse obravnavane distribucije smo opazili, da so podobne normalni distribuciji, če so izpolnjeni določeni pogoji. Razen izmed teh je, da je vzorec velik.

DISTRIBUCIJE VZORČNIH IZRAZOV

V nadaljevanju je za male vzorce nakazanih nekaj izrazov, ki se distribuirajo v navedenih verjetnostnih distribucijah.

16.13 Iz normalne populacije, v kateri je aritmetična sredina M in standardni odklon σ, izbiramo slučajnostne vzorce z n enotami. Za take vzorce velja:

Izraz

$$\frac{\bar{x}-M}{\sigma}\sqrt{n} = z\,; \quad \bar{x} = \frac{1}{n}Sx \tag{16.17}$$

se distribuira v standardizirani normalni distribuciji $N(0,1)$.

Izraz

$$\frac{\bar{x}-M}{s_x}\sqrt{n} = t(m = n-1); \quad s_x^2 = \frac{S(x-\bar{x})^2}{n-1} \tag{16.18}$$

se distribuira v t-distribuciji z $m = n-1$ stopinjami prostosti.

Izraz

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2(m = n-1) \tag{16.19}$$

pa se distribuira v χ^2-distribuciji z $m = n-1$ stopinjami prostosti.

16.14 Če je v normalni populaciji za x in y korelacijski koeficient $r_0 = 0$, se za slučajnostne vzorce z n enotami distribuira izraz

$$\frac{r}{\sqrt{1-r^2}}\sqrt{n-2} = t(m = n-2)\,; \quad r = \frac{c_{xy}}{s_x s_y}. \tag{16.20}$$

v t-distribuciji z $m = n-2$ stopinjami prostosti. r izračunamo iz podatkov vzorca.

V splošnem je distribucija ocen za korelacijski koeficient razmeroma zapletena, če pravi korelacijski koeficient ni nič. Če pa iz ocen za korelacijski koeficient r izračunamo

$$Z = 1{,}1513 \log\frac{1+r}{1-r} \tag{16.21}$$

se izraz

$$(Z - Z_0)\sqrt{n-3} = z \tag{16.22}$$

distribuira približno v standardizirani normalni distribuciji. Pri tem je Z_0 isti izraz kot v obrazcu 16.21, samo da ga izračunamo iz podatkov za r_0 v populaciji.

16.15 Če iz dveh normalnih populacij, ki imata različni aritmetični sredini $(M_1 \neq M_2)$, a enako varianco σ^2, izberemo iz prve slučajnostni vzo-

rec z n_1 enotami, iz druge pa vzorec z n_2 enotami, se izraz

$$\frac{(\bar{x}_2-\bar{x}_1)-(M_2-M_1)}{s_d}\sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}} = t(m=n_1+n_2-2) \; ; \; s_d^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \qquad (16.23)$$

distribuira v t-distribuciji z $m=n_1+n_2-2$ stopinjami prostosti. Pri tem sta $\bar{x}_1$ in s_1^2 količini, izračunani iz prvega, $\bar{x}_2$ in s_2^2 pa količini, izračunani iz drugega vzorca. Opozoriti moramo, da predpostavljamo, da je varianca σ^2 v obeh populacijah ista.

16.16 Če iz dveh normalnih populacij, ki imata različni varianci $(\sigma_1^2 \neq \sigma_2^2)$, izberemo iz prve vzorec z n_1 enotami, iz druge pa vzorec z n_2 enotami, se izraz

$$\frac{s_1^2}{\sigma_1^2} \Big/ \frac{s_2^2}{\sigma_2^2} = F(m_1 = n_1-1 \; ; \; m_2 = n_2-1) \qquad (16.24)$$

distribuira v F-distribuciji z $m_1 = n_1-1$ in $m_2 = n_2-1$ stopinjami prostosti. To velja, če vzamemo v kvocientu na levi strani izraza v obrazcu 16.24 v števec večji izraz $(s_1^2/\sigma_1^2 > s_2^2/\sigma_2^2)$.

16.17 Če iz populacije, za katero je korelacijski koeficient r_{01}, izberemo vzorec z n_1 enotami, iz populacije, za katero je korelacijski koeficient r_{02} pa vzorec z n_2 enotami, se izraz

$$\frac{(Z_2-Z_1)-(Z_{02}-Z_{01})}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}} = z \qquad (16.25)$$

distribuira v približno standardizirani normalni distribuciji. Z_1, Z_2 ; Z_{01}, Z_{02} = ustrezni izrazi po obrazcu 16.21.

16.18 Če iz populacije, za katero so teoretične frekvence za dani sestav v populaciji enake f_0, izbiramo slučajnostne vzorce z n enotami, za katere frekvence za sestav populacije zaznamujemo z f, se izraz

$$\sum \frac{(f-f_0)^2}{f_0} = \chi^2(m) \qquad (16.26)$$

distribuira približno v χ^2-distribuciji z ustreznim številom stopinj prostosti m. Če imamo razdelitev po enem samem znaku s k grupami, je število prostosti $m = k-1$, ker je za frekvence pogoj, da je skupna vsota frekvenc enaka n. Če imamo populacijo razdeljeno v tabeli $k \times g$ po dveh

znakih hkrati, pa je število stopinj prostosti $m = (k-1)(g-1)$; to smo nakazali že v odstavku o stopinjah prostosti.

To so osnovne zakonitosti za ocene parametrov, ki jih najpogosteje ocenjujemo.

MEJE ZAUPANJA ZA OCENE Z MALIMI VZORCI

16.19 Če poznamo zakonitosti vzorcev, ki smo jih navedli v prejšnjem odstavku, moremo zlahka določiti meje zaupanja za parametre, ki jih ocenjujemo z malimi vzorci.

Ker je standardizirana normalna distribucija simetrična distribucija, zanjo velja $z_{1-P} = -z_P$. Zato moremo trditi: Če iz podatkov slučajnostnega vzorca izračunamo izraz, ki se distribuira standardizirano normalno, ta izraz s tveganjem $2P$ leži v mejah med $-z_P$ in z_P. S tveganjem P pa ni večji kot $+z_P$. Z enakim tveganjem ni manjši kot $-z_P$ (glej odstavek 14.22).

Podobno velja za izraze, ki se distribuirajo v t-distribuciji. Ker je t-distribucija simetrična distribucija, tudi zanjo velja: $t_{1-P} = -t_P$. Zato moremo enako trditi: Če iz slučajnostnega vzorca izračunamo izraz, ki se distribuira v t-distribuciji, ta izraz s tveganjem $2P$ leži v mejah med $-t_P$ in $+t_P$. S tveganjem P pa ni večji kot $+t_P$ in z enakim tveganjem P ni manjši kot $-t_P$.

Za izraze, ki se distribuirajo v χ^2-distribuciji, pa moremo trditi: Če iz slučajnostnega vzorca izračunamo izraz, ki se distribuira v χ^2-distribuciji, ta izraz s tveganjem $2P$ leži v mejah med χ^2_{1-P} in χ^2_P. S tveganjem P pa ni večji kot χ^2_P.

Za izraze, ki se distribuirajo v F-distribuciji trdimo analogno: Če iz podatkov za slučajnostni vzorec izračunamo izraz, ki se distribuira v F-distribuciji, je vrednost tega izraza s tveganjem P manjša kot F_P.

Če to upoštevamo, moremo izračunati meje zaupanja za vse parametre, ki nastopajo v obrazcih 16.17 do 16.23.

Medtem ko imamo v tabelah v dodatku za standardizirano normalno distribucijo ordinate in verjetnosti za podrobne vrednosti standardiziranega odklona, zadostujejo za druge distribucije le kritične vrednosti t_P, χ^2_P in F_P, katere s tveganjem P ne presegajo izrazi, ki se distribuirajo v teh distribucijah. Za P so v tablicah vzete običajne stopnje tveganja (0,10; 0,05; 0,01 in 0,001).

Meje zaupanja za posamezen parameter iz obrazcev 16.17 do 16.23 dobimo, če najprej določimo, v katerem razmaku je z določenim tveganjem ustrezna količina z, t, χ^2 ali F. Dobljene neenačbe pa preuredimo tako, da so znane količine ločene od ocenjevanega parametra.

16.20 Vzemimo za prvi primer *oceno aritmetične sredine*, če poznamo varianco populacije, dobimo iz obrazca 13.17

$$-z_p < \frac{\bar{x} - M}{\sigma}\sqrt{n} < +z_p \qquad (16.27)$$

Če ta obrazec preuredimo, dobimo

$$\bar{x} - z_p \frac{\sigma}{\sqrt{n}} < M < \bar{x} + z_p \frac{\sigma}{\sqrt{n}} \qquad (16.28)$$

16.21 Po tem postopku moremo n.pr. rešiti tale problem. Za stroj, za katerega vemo, da je standardni odklon njegovih proizvodov $\sigma = 2$ mm, ne poznamo pa aritmetične sredine njegovih proizvodov, z vzorcem $n = 50$ proizvodov dobimo, da je $\bar{x} = 28{,}0$ mm. Kolike so meje zaupanja za to oceno, če vzamemo stopnjo tveganja $2P = 0{,}05$. Ker je $z_{0{,}025} = 1{,}96$, dobimo po obrazcu 13.28

$$28 - 1{,}96 \frac{2}{\sqrt{50}} < M < 28 + 1{,}96 \frac{2}{\sqrt{50}}$$

$$27{,}45 \text{ mm} < M < 28{,}55 \text{ mm}$$

16.22 Če pa standardnega odklona ne poznamo in ga ocenimo iz vzorca, dobimo meje zaupanja za *aritmetično sredino* populacije podobno kot zgoraj iz obrazca 13.18.

$$\bar{x} - t_p \frac{s}{\sqrt{n}} < M < \bar{x} + t_p \frac{s}{\sqrt{n}} \qquad (16.29)$$

16.23 Vzemimo zgornji primer, da standardnega odklona za proizvodnjo ne poznamo. Zato ga ocenimo iz istega vzorca $n = 50$ proizvodov, iz katerega smo ocenili aritmetično sredino. Iz njega dobimo, da je $s = 2{,}1$. Meje zaupanja določimo takole: Ker je število stopinj prostosti za t enako $m = n-1 = 50-1 = 49$, najdemo v tabeli za t-distribucijo, da je $t(m=49) = 2{,}01$. Ker za $m = 49$ nimamo v tabeli ustreznih kritičnih vrednosti, vzamemo interpolirano vrednost, oziroma vrednost za $m = 50$, ker so razlike neznatne. Če to upoštevamo, dobimo, da so meje zaupanja za aritmetično sredino proizvodnje po obrazcu 13.29 enake

$$28 - 2{,}01 \frac{2{,}1}{\sqrt{50}} < M < 28 + 2{,}01 \frac{2{,}1}{\sqrt{50}}$$

$$27{,}40\text{mm} < M < 28{,}60\text{mm}$$

16.24 Za oceno *variance* iz normalne populacije dobimo po obrazcu 16.18, da velja s tveganjem $2P$

$$\chi^2_{1-P} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{P} \qquad (16.30)$$

Če ta izraz preuredimo, dobimo meje zaupanja za varianco

$$\frac{(n-1)s^2}{\chi^2_P} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-P}} \qquad (16.31)$$

16.25 Vzemimo, da za pošiljko jabolk ugotavljamo kakšna je stopnja sortiranosti jabolk glede na težo. Pošiljka je tem kvalitetnejša, čim manjše so razlike v teži za posamezna jabolka. Kot pokazatelj za stopnjo enovitosti vzemimo varianco v teži. Za ta namen slučajnostno izberimo iz pošiljke $n = 25$ jabolk. Zanje ugotovimo, da je ocena variance $s^2 = 6,2$. Kakšne so meje zaupanja za to oceno?

Ker je $n = 25$, je število stopinj $m = n-1 = 25-1 = 24$. Vzemimo, da ocenjujemo meje zaupanja s tveganjem $0,10$. Iz teh podatkov najdemo v tabeli za χ^2-distribucijo, da je $\chi^2_{0,95}(m = 24) = 13,85$ in $\chi^2_{0,05}(m=24) = 36,42$. Če te podatke vnesemo v obrazec 16.31, dobimo:

$$\frac{(25-1)\cdot 6,2}{36,42} < \sigma^2 < \frac{(25-1)\cdot 6,2}{13,85}$$

$$4,09 < \sigma^2 < 10,73$$

Meje zaupanja za standardni odklon dobimo, če poiščemo kvadratna korena iz mej zaupanja za varianco.

$$2,02 \text{ dkg} < \sigma < 3,30 \text{ dkg}$$

Če omejimo razmak zaupanja enostransko in za isti vzorec iščemo, katero vrednost prava vrednost standardnega odklona ne prekorači s tveganjem $P = 0,01$, dobimo iz obrazca 16.19:

$$\chi^2_{1-P} < \frac{(n-1)s^2}{\sigma^2} \quad : \quad \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-P}} \qquad (16.32)$$

Tako dobimo

$$\sigma^2 < \frac{(25-1)6,2}{10,86} = 13,7$$

ker je po tablici za χ^2-distribucijo $\chi_{0,99}(m = 24) = 10,86$.

16.26 Ker menimo, da je natančnost dobljenih podatkov premajhna, vzemimo večji vzorec in izberimo n.pr. $n = 75$ jabolk. Za ta vzorec dobimo, da je ocena variance $s^2 = 7,3$. Meje zaupanja za to oceno izračunamo enako kot v zgornjem primeru po obrazcu 13.32. Glede na to, da je stopinj prostosti $m = n-1 = 75 - 1 = 74$, zahtevamo pa rezultate z isto stopnjo tveganja $2P = 0,10$, moramo poiskati $\chi^2_{0,95}(m = 74)$ in $\chi^2_{0,05}(m = 74)$. Teh količin pa v tabeli za χ^2-distribucijo ni. Pač pa vemo, da velja za χ^2-distribucije, če je število stopinj prostosti veliko, obrazec 13.15. Zato moremo za te primere ustrezni χ^2 izračunati prek standardiziranega odklona z po obrazcu

$$\chi_P^2 = \tfrac{1}{2}(\sqrt{2m-1} + z_P)^2 \qquad (16.33)$$

Tako dobimo, da je v našem primeru

$$\chi^2_{0,95}(m=74) = \tfrac{1}{2}(\sqrt{2,74 - 1} - 1,645)^2 = 54,9 ;$$

$$\chi^2_{0,05}(m=74) = \tfrac{1}{2}(\sqrt{2,74 - 1} + 1,645)^2 = 94,8$$

ker je $z_{0,95} = -1.645$, $z_{0,05} = +1,645$

Meje zaupanja pa so dalje po obrazcu 13.32

$$\frac{(75-1)7,6}{94,8} < \sigma^2 < \frac{(75-1)7,6}{54,9}$$

$$5,93 < \sigma^2 < 10,24$$

ali standardni odklon

$$2,44\,\mathrm{dkg} < \sigma < 3,20\,\mathrm{dkg}$$

Iz rezultata vidimo, da se je razmak zaupanja resnično skrčil. Ocena je boljša.

16.27 Meje zaupanja za *korelacijski koeficient* določimo prek transformacije podatkov v Z po obrazcu 13.21. Po obrazcu 13.22 s tveganjem $2P$ velja

$$-z_P < (Z - Z_0)\sqrt{n-3} < +z_P \qquad (16.34)$$

Če to neenačbo preuredimo, dobimo

$$Z - \frac{z_P}{\sqrt{n-3}} < Z_0 < Z + \frac{z_P}{\sqrt{n-3}} \qquad (16.35)$$

meje zaupanja za Z_0 s tveganjem $2P$. Meje zaupanja za korelacijski koeficient pa dobimo, če v tablici, v kateri imamo zvezo med korelacijskim koeficientom r in Z, poiščemo mejnim Z ustrezne vrednosti za korelacijski koeficient.

16.28 Vzemimo za primer povezanost med številom doseženih točk pri mehanskem in prostorninskem testu za $n = 20$ tretješolcev mariborske klasične gimnazije. Teh 20 dijakov vzemimo kot vzorec iz hipotetične populacije tretješolcev v enakih pogojih. V tabeli 13.11 smo za ta primer dobili, da je $r = 0{,}73$. Iz tabele r-Z dobimo, da $r = 0{,}73$ ustreza $Z = 1{,}00$.

Če vzamemo stopnjo tveganja v razmaku zaupanja enako $2P = 0{,}05$, je ustrezni $z_{0,025} = 1{,}96$.

Če te količine vnesemo v obrazec 15.35 dobimo

$$1{,}00 - \frac{1{,}96}{\sqrt{20-3}} < Z_0 < 1{,}00 + \frac{1{,}96}{\sqrt{20-3}}$$

$$0{,}52 < Z_0 < 1{,}48$$

V tabeli r-Z najdemo, da $Z_s = 0{,}52$ ustreza $r_s = 0{,}48$, $Z_z = 1{,}48$ pa $r_z = 0{,}90$. Meje zaupanja za korelacijski koeficient med številom točk pri mehanskem in prostorninskem testu so torej

$$0{,}48 < r_0 < 0{,}90$$

Iz primera vidimo, da je zanesljivost rezultata razmeroma majhna.

16.29 Analogno kot za druge parametre izračunavamo meje zaupanja tudi za *regresijski koeficient*. Meje zaupanja za regresijski koeficient so

$$b - t_P \frac{\sqrt{s_y^2/s_x^2 - b^2}}{\sqrt{n-2}} < b_0 < b + t_P \frac{\sqrt{s_y^2/s_x^2 - b^2}}{\sqrt{n-2}} \qquad (16.36)$$

s koeficientom tveganja $2P$. t_P določimo z $m = n-2$ stopinjami prostosti.

16.30 Če vzamemo za primer regresijski koeficient med mehanskim in pro-

storninskim testom, za $n=20$ tretješolcev mariborske klasične gimnazije iz primera v tabeli 13.11, dobimo po podatkih te tabele, da je $b = 2,21$, razmerje med variancama $s_y^2/s_x^2 = 8,5$. Če določimo koeficient tveganja $2P = 0,05$, dobimo po tabeli za t-distribucijo za $m = n-2 = 20-2 = 18$: $t_{0,025}(18) = 2,10$. Iz teh podatkov dobimo meje zaupanja

$$2,21 - 2,1\frac{\sqrt{8,5-2,21^2}}{\sqrt{20-2}} < b_0 < 2,21 + 2,1\frac{\sqrt{8,5-2,21^2}}{\sqrt{20-2}}$$

$$1,27 < b_0 < 3,15$$

Razmak zaupanja, ki smo ga dobili, je razmeroma zelo širok. Ocena je razmeroma slaba.

16.31 Meje zaupanja za *strukturni delež*. Če ga ocenjujemo z malim vzorcem, bi morali določati z binomialnimi distribucijami. Vendar je direktsn način zamuden. Izkaže pa se, da moremo poiskati meje zaupanja za strukturni delež iz majhnih vzorcev prek F-distribucije po temle postopku: Če iz slučajnostnega vzorca z n enotami dobimo x enot z določeno značilnostjo, določimo spodnjo mejo zaupanja $P_s\,\%$ za strukturni delež $P_0\,\%$ po obrazcu:

$$P_s\,\% = \frac{100\,x}{x+(n-x+1)F_s} \qquad (16.37)$$

Pri tem je: F_s = kritična vrednost za F-distribucijo z $m_1=2\cdot(n-x+1)$ in $m_2 = 2x$ stopinjami prostosti s stopnjo tveganja P. Zgornje meje zaupanja $P_z\,\%$ pa izračunamo po obrazcu

$$P_z\,\% = \frac{100(x+1)F_z}{(x+1)F_z + n-x} \qquad (16.38)$$

Pri tem je: F_z = kritična vrednost za F-distribucijo z $m_1 = 2(x+1)$ $m_2 = 2(n-x)$ stopinjami prostosti in stopnjo tveganja P. Prava vrednost strukturnega deleža $P_0\,\%$ je v razmaku zaupanja

$$P_s\,\% < P_0\,\% < P_z\,\% \qquad (16.39)$$

s tveganjem $2P$.

16.32 Vzemimo, da smo iz populacije delavcev v nekem podjetju na slepo izbrali $n = 30$ delavcev. Od teh je $x = 19$ izpolnilo določeno nalogo v ustreznem času. Kakšne so meje zaupanja za odstotek delavcev, ki so sposobni, da rešijo proučevani problem v ustreznem času? Stopnja tveganja

$2P = 0{,}10$.

Po prejšnjem postopku moramo najprej v tabeli za F-distribucijo poiskati F_s z $m_1 = 2(n-x+1) = 2(30-19+1) = 24$ in $m_2 = 2x = 2.19 = 38$ stopinjami prostosti; in F_z z $m_1 = 2(x+1) = 2(19+1) = 40$ in $m_2 = 2(n-x) = 2(30-19) = 22$ stopinjami prostosti. Glede na predpisano tveganje dobimo iz tabel za F-distribucijo v tabeli za $P = 0{,}05$

$$F_s = 1{,}81 \quad \text{in} \quad F_z = 1{,}94.$$

Te vrednosti dobimo z linearno interpolacijo, ki v prvem približku zadošča.

Če te vrednosti vnesemo v obrazec 15.37 in 16.38, dobimo

$$P_s\% = \frac{100.19}{19+(30-19+1)1{,}81} = 47\ \%\ ; \quad P_z\% = \frac{100.(19+1).1{,}94}{(19+1).1{,}94+(30-19)} = 78\ \%$$

$$47\ \% < P_0\% < 78\ \%$$

Pravi odstotek je s tveganjem 0,10 med 47 % in 78 %.

16.33 Po obrazcu 15.23 pa moremo sestaviti meje zaupanja za *razliko med aritmetičnima sredinama* dveh populacij, ki imata enaki varianci. Iz tega obrazca namreč dobimo, da je s tveganjem $2P$ prava razlika med aritmetičnima sredinama dveh populacij v razmaku zaupanja

$$\left(\bar{x}_2 - \bar{x}_1\right) - t_p s_d \sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}} < M_2 - M_1 < \left(\bar{x}_2 - \bar{x}_1\right) + t_p s_d \sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$$

$$t = t_p(m = n_1+n_2-2)\ ; \quad s_d^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \qquad (16.40)$$

Ta obrazec je zelo koristen. Z njim moremo z vzorci oceniti učinek spremenjenih pogojev na določen parameter. Po tej metodi iz dveh populacij, v katerih so pogoji, ki jih proučujemo, različni, izberemo iz prve vzorec z n_1, iz druge pa vzorec z n_2 enotami. Tako moremo po tem postopku oceniti razlike v produktivnosti dela po stroju, razlike v navadah porabe določenih artiklov med dvema porabnikoma, razlike med donosoma pri spremenjenih agrotehničnih ukrepih, razlike v trpežnosti tkanin pri spremenjenem tehnološkem procesu, razlike v prodaji pri različnem načinu embalaže itd. Običajno vzamemo eno populacijo za tako imenovano kontrolno grupo, na katero primerjamo učinek različnih oziroma spremenjenih pogojev.

16.34 Vzemimo za primer podatke iz ankete delavskih družin v novembru 1957 Zavoda LRS za statistiko. Že pri korelaciji smo proučevali izdatke za kulturno in družbeno življenje v odvisnosti od skupnih dohodkov. Postavimo si nalogo, da iz majhnega vzorca $n_1 = 12$ družin, ki imajo od 15 do 20 tisoč dinarjev dohodkov, in iz vzorca $n_2 = 10$ družin, ki imajo od 20 do 25 tisoč

dinarjev dohodkov, ocenimo, koliko najmanj znaša razlika v povprečnih izdatkih za kulturno in družbeno življenje za ti dve skupini (s tveganjem 0,05).

Vzorca sta dala tele osnovne podatke:

Grupa dohodkov 15–20 tisoč din: $n_1 = 12$

Izdatki v din: 260; 330; 490; 310; 580; 560; 330; 480; 860; 590; 330; 630.

Grupa dohodkov: 20–25 tisoč din: $n_2 = 10$

Izdatki v din: 650; 1100; 760; 460; 700; 540; 750; 370; 630; 740.

Da izračunamo varianco s_d^2 moramo poznati ocene variance v prvem s_1^2 in drugem vzorcu s_2^2. Te pa izračunamo po znanem obrazcu

$$s^2 = \frac{Sx^2 - X^2/n}{n-1}$$

Ker je $\sum_1 x^2 = 3095500$; $X_1 = 5750$; je

$$s_1^2 = \frac{3095500 - 5750^2/12}{12 - 1} = 30936$$

Ker je $\sum_2 x^2 = 4847200$; $X_2 = 6700$; je

$$s_2^2 = \frac{4847200 - 6700^2/10}{10 - 1} = 39800$$

Po obrazcu 16.40 je

$$s_d^2 = \frac{30936(12-1) + 39800(10-1)}{12 + 10 - 2} = 34925$$

$$s_d = \sqrt{34925} = 187$$

Ker se vprašamo, koliko najmanj znaša stvarna razlika v povprečnih izdatkih za kulturno in družbeno življenje med obema grupama, ocenimo razmak zaupanja enostransko in poiščemo tisto vrednost, ki jo prava razlika med sredinama prekorači s tveganjem $P = 0{,}05$. Ta pa je

$$\bar{x}_2 - \bar{x}_1 - t_p s_d \sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}} < M_2 - M_1 \qquad (16.41)$$

ker je $\bar{x}_1 = X_1/n_1 = 5750/12 = 479$ din

$\bar{x}_2 = X_2/n_2 = 6700/10 = 670$ din

je ocena razlike: $\bar{x}_2 - \bar{x}_1 = 670 - 479 = 191$ din

Koeficient t je $t_{0,05}\,(m = n_1 + n_2 - 2 = 12+10-2 = 20) = 1{,}72$

Tako dobimo po obrazcu 16.41

$$191 - 1{,}72 \cdot 187 \sqrt{\frac{12+10}{12 \cdot 10}} < M_2 - M_1$$

$$53 \text{ din} < M_2 - M_1$$

Razlika v povprečnih izdatkih za kulturno in družbeno življenje med grupama je s tveganjem 0,05 najmanj 53 dinarjev.

Sedemnajsto poglavje

VZORČENJE-PREIZKUŠANJE HIPOTEZ

TEORETIČNE OSNOVE PREIZKUŠANJA HIPOTEZ

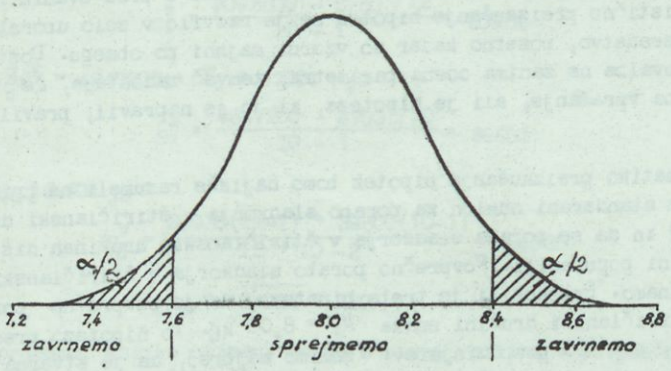
17.1 Z vzorčenjem ocenjujemo parametre na dva načina: s točkovno oceno in razmakom zaupanja. Te metode pa dajo zadovoljive rezultate le, če je vzorec zadosti velik. V nasprotnem primeru so ocene običajno tako nezanesljive, da je vprašanje, če so uporabne. Vendar moremo z vzorci reševati še druge probleme. Z njimi preizkušamo določene hipoteze o proučevanih populacijah. Statistično preizkušanje hipotez se je razvilo v zelo uporabno raziskovalno sredstvo, posebno kadar so vzorci majhni po obsegu. Dostikrat namreč raziskovalca ne zanima ocena parametra, temveč zadostuje, če vzorec odgovori na vprašanje, ali je hipoteza, ki jo je napravil, pravilna ali ne.

17.2 Problematiko preizkušanja hipotez bomo najlaže razumeli na primeru. Vzemimo, da je standardni odklon za porabo sladkorja v štiričlanski družini $\sigma = 2{,}0$ kg in da se poraba sladkorja v štiričlanskih družinah distribuira v normalni populaciji. Povprečno porabo sladkorja v štiričlanski družini pa ne poznamo. Preskusiti je treba hipotezo, da je povprečna poraba sladkorja v štiričlanski družini enaka $M_H = 8{,}00$ kg. To hipotezo preskušamo z vzorcem $n = 100$ gospodinjstev. Vzemimo najprej, da je stvarna povprečna potrošnja sladkorja v štiričlanskih gospodinjstvih resnično enaka hipotetični potrošnji $M = M_H = 8{,}00$ kg. V tem primeru se povprečja, ki jih dobimo iz vzorcev po $n = 100$ slučajnostno izbranih gospodinjstev, po znanih zakonitostih distribuirajo v normalni distribuciji z $M = 8{,}00$ kg in standardno pogreško $SE_{\bar{x}} = \sigma/\sqrt{n} = 2{,}0/\sqrt{100} = 0{,}2$ kg. Iz lastnosti normalnih distribucij pa sklepamo dalje, da je za 95 % od vseh vzorcev povprečje $\bar{x}$ v razmaku od $M_H - 1{,}96SE_{\bar{x}} = 8{,}00 - 1{,}96 \cdot 0{,}2 = 7{,}61$ kg in $M_H + 1{,}96SE_{\bar{x}} = 8{,}00 + 1{,}96 \cdot 0{,}2 = 8{,}39$ kg.

Ker v primeru, da se stvarna potrošnja ujema s hipotetično, z veliko verjetnostjo (0,95) izberemo vzorec, za katerega povprečje pade v razmak 7,31 kg do 8,39 kg, napravimo naslednje pravilo:

Pravilo A: Hipotezo, da je povprečna poraba sladkorja za štiričlansko gospodinjstvo enaka 8 kg mesečno, ($M_H = 8{,}00$ kg) sprejmemo, če povprečje v preskusnem vzorcu pade v razmak 7,61 do 8,39 kg. Hipotezo pa zavrnemo, če je povprečje, ki ga dobimo s preskusnim vzorcem, zunaj tega razmaka. Meje, s katerimi razmejimo razmak, v katerem hipotezo sprejmemo, od razmaka, v katerem jo zavrnemo, imenujemo kritične meje.

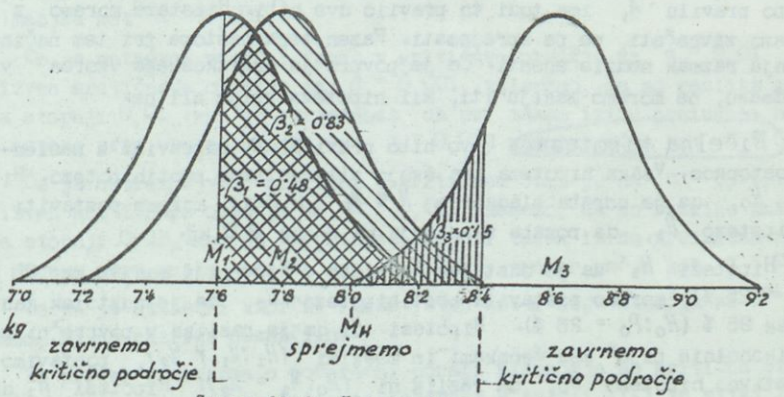
17.3 Napake prve vrste. Če pravilo A pozorneje proučimo, ugotovimo, da je v njem določeno tveganje. Če je namreč resnična poraba sladkorja enaka hipotetični $M = M_H = 8{,}00$ kg, moremo z zgornjim pravilom dobiti tudi vzorce, iz katerih sklepamo, da hipoteza, da je stvarna poraba enaka 8,00 kg, ne velja. Z verjetnostjo $\alpha = 2P = 0{,}05$ namreč pričakujemo, da po zgornjem pravilu sredina v vzorcu pade izven zgornjega razmaka. Iz tega sledi, da po zgornjem pravilu z verjetnostjo $\alpha = 0{,}05$ izberemo tak vzorec, da hipotezo zavrnemo, kljub temu da je pravo povprečje v skladu s hipotezo. Tveganju, ki ga pri tem načinu sklepanja napravimo s tem, da moremo z določeno verjetnostjo hipotezo zavrniti, čeprav velja, pravimo *napaka prve vrste*. Pravilo A in napaka prve vrste sta za naš primer grafično ponazorjeni v sliki 17.1.



Slika 17.1 Pravilo A in napaka prve vrste

17.4 Napaka druge vrste. V zgornjem pravilu pa tiči še druga nevarnost, ki je hujša kot napaka prve vrste. Vzemimo, da poznamo dejansko povprečno porabo sladkorja, ki pa ni enaka hipotetični, temveč na primer $M = 7{,}8$ kg. Povprečja vzorcev se distribuirajo okrog pravega povprečja $M = 7{,}8$ kg v normalni distribuciji s standardno pogreško $S_{\bar{x}}^{E} = 0{,}2$ kg. V sliki 17.2 so razen distribucije povprečij vzorcev, ki velja, če je pravo povprečje enako $M = 7{,}8$ kg, vrisane še kritične meje, ki smo jih dobili po pravilu A, če bi veljala hipoteza $M = M_H = 8{,}0$. Iz slike vidimo, da je v tem primeru verjetnost, da izberemo vzorec, za katerega povprečje leži izven razmaka, ki ga omejujeta kritični vrednosti 7,31 in 8,39 enaka 0,17, verjetnost, da je povprečje slučajnostnega vzorca v razmaku med kritičnima mejama, pa 0,83.

Iz tega vidimo, da je, čeprav hipoteza ne velja (stvarno povprečje $M = 7{,}8$ kg različno od hipotetičnega povprečja $M_H = 8{,}0$ kg), velika verjetnost ($\beta = 0{,}83$), da izberemo vzorec, za katerega je povprečje znotraj razmaka, v katerem hipotezo $M_H = 8{,}0$ kg sprejmemo.



Slika 17.2 Napaka druge vrste

Po pravilu A v tem primeru sklepamo, da je stvarno povprečje M enako hipotetičnemu, ker povprečje vzorca pade znotraj kritičnih meja 7,31 in 8,39. Ta sklep pa je napačen, ker hipotezo sprejmemo, stvarno povprečje pa je različno od hipotetičnega. Kot vidimo iz primera, je verjetnost, da hipotezo sprejmemo, čeprav ne velja, zelo velika ($\beta = 0{,}83$). V povprečju bi torej od 100 vzorcev, ki bi jih izbrali, s 83 vzorci hipotezo sprejeli, čeprav ne velja. Napako, ki je v tem, da napačno hipotezo sprejmemo, čeprav ni pravilna, imenujemo *napako druge vrste*, zaznamujejo jo pa z β.

V sliki 17.2 sta razen za vzorčno distribucijo z $M = 7{,}8$ vrisani tudi vzorčni distribuciji povprečij vzorcev, če je $M = 7{,}3$ ali $M = 8{,}3$. V sliki so vrisane tudi ustrezne napake druge vrste. Za distribucijo z $M = 7{,}3$ je napaka druge vrste 0,48, za $M = 8{,}3$ pa 0,15. Iz teh primerov sklepamo, da je napaka druge vrste odvisna od prave vrednosti povprečja. Napaka druge vrste je tem večja, čim manjše so razlike med stvarno in hipotetično vrednostjo in obratno, tem manjša, čim večje so razlike med hipotetično in stvarno vrednostjo.

Ker more biti napaka druge vrste zelo velika in jo razen tega niti ne poznamo, ker ne poznamo prave vrednosti parametra, je zelo nevarno hipotezo sprejeti, če je povprečje za preizkusni vzorec v razmaku 7,61 kg - 8,39 kg. Zato pravilo A popravimo tako, da namesto njega postavimo pravilo B.

Pravilo B. Hipotezo, da je poraba sladkorja v štiričlanski družini enaka $M_H = 8{,}0$ kg zavrnemo, če pade povprečje $\bar{x}$, ki ga izračunamo iz preizkusnega vzorca $n = 100$ gospodinjstev v kritično območje pod 7,31 kg ali nad 8,39 kg. Če pa je iz preizkusnega vzorca izračunano pov-

preč je $\bar{x}$ znotraj razmaka 7,31–8,39 kg, hipoteze ne sprejmemo, temveč se vzdržimo izjave, ker je prevelik tveganje, da bi hipotezo sprejeli.

Kljub temu, da je sklepanje po pravilu B pravilneje kakor sklepanje po pravilu A, ima tudi to pravilo dve hibi. Hipoteze moremo z njim samo zavračati, ne pa sprejemati. Razen tega nastopa pri tem načinu sklepanja razmak nedoločenosti. Če je povprečje preizkusnega vzorca v tem razmaku, ne moremo zaključiti, ali hipoteza velja ali ne.

17.5 Ničelna hipoteza. Prvo hibo pravila B odpravimo z naslednjim postopkom. Vsaka hipoteza ima svojo alternativno protihipotezo. Hipotezi H_0, da je poraba sladkorja $M = M_H = 8{,}0$ kg, moremo postaviti protihipotezo H_1, da poraba sladkorja ni enaka 8,0 kg.

Hipotezi H_1 da je odstotek kadilcev v Sloveniji manjši kot 25 % ($H_1 : P_0 < 25$ %), moremo postaviti protihipotezo H_0, da je odstotek kadilcev enak 25 % ($H_0 : P_0 = 25$ %). Hipotezi H_1 da je razlika v povprečni porabi alkoholnih pijač med ženskami in moškimi ($H_1 : M_1 \neq M_2$), postavimo alternativno hipotezo H_0, da razlik ni ($H_0 : M_1 = M_2$). Hipotezi H_1 da stroj ne izdeluje artikle predpisane kvalitete, postavimo protihipotezo H_0, da stroj izdeluje artikle predpisane kvalitete ita.

Hipoteza in alternativna protihipoteza se med seboj izključujeta. Če uspemo da zavrnemo protihipotezo, sprejmemo osnovno hipotezo. Zato sklepanje po pravilu B preoblikujemo tako, da hipotezi, ki jo skušamo preizkusiti, postavimo ustrezno alternativno protihipotezo, ki jo imenujemo *ničelna hipoteza*. V nadaljnjem preizkušamo ničelno hipotezo. Ker moremo po pravilu B hipoteze samo zavrniti, ne pa sprejeti, to pomeni: Če uspemo, da zavrnemo ničelno hipotezo, sprejmemo osnovno hipotezo. Tako z ničelno hipotezo sprejmemo osnovno hipotezo. Z ničelno hipotezo preizkušamo hipoteze po naslednjem pravilu C.

Pravilo C. Osnovni hipotezi H_1 postavimo ustrezno alternativno protihipotezo H_0, ki jo po pravilu skušamo zavrniti. To protihipotezo imenujemo ničelno hipotezo. Osnovni hipotezi H_1, da poraba sladkorja ni enaka 8,0 kg, postavimo nasproti ničelno hipotezo H_0, da je poraba sladkorja v štiričlanskih družinah enaka ($H_0 : M = M_H = 8{,}0$ kg).

Če iz vzorca $n = 100$ gospodinjstev izračunana povprečna potrošnja pade v kritično območje izven razmaka 7,31 kg do 8,39 kg, sklepamo, da je stvarna povprečna potrošnja sladkorja *značilno* različna od $M_H = 8{,}0$ kg, ki smo jo postavili za ničelno hipotezo. Če pa je vzorčna aritmetična sredina $\bar{x}$ znotraj razmaka 7,31 kg do 8,39 kg, vzamemo, da razlike *niso značilne*. Da razlike niso značilne, še ne pomeni, da razlik ni, temveč le, da jih z izvršenim preizkusom nismo odkrili.

Zaradi napake prve vrste more preizkus z verjetnostjo α pokazati značilne razlike, čeprav ničelna hipoteza drži. Ker pa moremo napako prve vrste spreminjati, običajno navajamo značilnost razlik na eni izmed treh stopenj značilnosti: 0,05, 0,01 in 0,001.

Če je ustrezni vzorčni izraz (v našem primeru je to $\bar{x}$) v kritičnem območju za $\alpha = 0{,}001$, ničelno hipotezo zavrnemo na stopnji 0,001. To pomeni, da je verjetnost, da pri takem izidu preizkusa ničelna hipoteza velja, manjša kot $\alpha = 0{,}001$.

Če je ustrezni vzorčni izraz v kritičnem območju za $\alpha = 0{,}01$, vendar izven kritičnega območja za $\alpha = 0{,}001$, sklepamo, da so razlike značilne na stopnji 0,01, ker je verjetnost, da pri takem izidu preizkusa napravimo napako prve vrste, manjša kot $\alpha = 0{,}01$, vendar večja kot $\alpha = 0{,}001$.

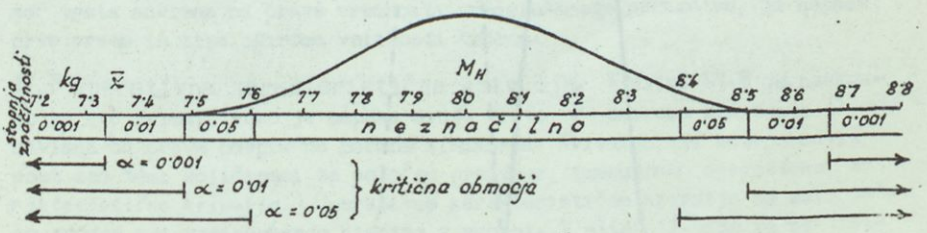
Če je ustrezni vzorčni izraz v kritičnem območju za $\alpha = 0{,}05$, vendar izven kritičnega območja za $\alpha = 0{,}01$, vzamemo, da so razlike značilne na stopnji 0,05, ker je verjetnost, da pri takem izidu preizkusa napravimo napako prve vrste, manjša kot $\alpha = 0{,}05$, vendar večja kot $\alpha = 0{,}01$.

Če pa je ustrezni vzorčni izraz izven kritičnega območja za $\alpha = 0{,}05$, vzamemo, da so razlike neznačilne.

Za primer preizkusa o povprečni porabi sladkorja so kritična območja za posamezne stopnje značilnosti nakazane v tabeli 17.1 in sliki 17.3

Tabela 17.1 Kritična območja za posamezne stopnje značilnosti za povprečno porabo sladkorja

Stopnja značilnosti	Kritično območje	
$\alpha = 0{,}05$	$\bar{x} < 7{,}61$ kg	$8{,}39$ kg $< \bar{x}$
$\alpha = 0{,}01$	$\bar{x} < 7{,}49$ kg	$8{,}51$ kg $< \bar{x}$
$\alpha = 0{,}001$	$\bar{x} < 7{,}34$ kg	$8{,}66$ kg $< \bar{x}$



Slika 17.3 Stopnje značilnosti in kritična območja za povprečno porabo sladkorja

Če dobimo z vzorcem 100 gospodinjstev povprečno porabo $\bar{x} = 8{,}62$, vzamemo, da je poraba sladkorja značilno različna od 8,0 kg na ravni

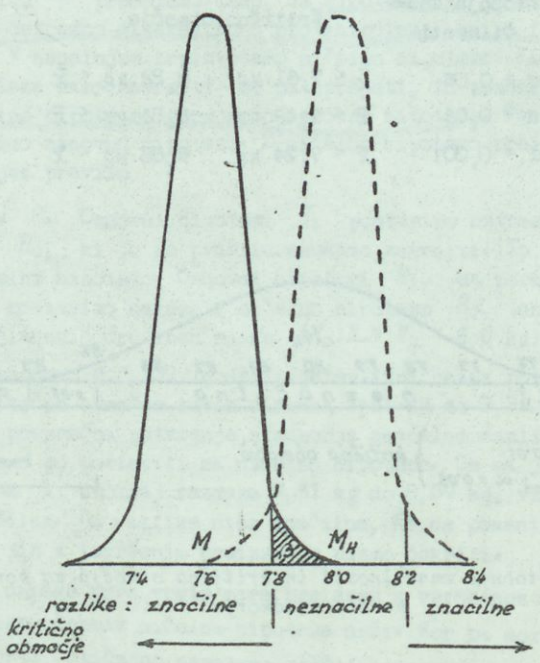
$\alpha = 0{,}01$, ker za raven $\alpha = 0{,}01$ pade v kritično območje, za raven $\alpha = 0{,}001$ pa izven njega. Da je v tem primeru razlika značilna na ravni $\alpha = 0{,}01$, pomeni, da je največ 1 % verjetnosti, da bi dobili z vzorcem tako aritmetično sredino, če bi bila povprečna poraba enaka 8,0 kg, kakršna je ničelna hipoteza o porabi. Zanesljivost sklepa je torej zelo velika.

Neznačilne razlike pa še ne pomenijo, da razlik med stvarnimi in hipotetičnimi vrednostmi ni, temveč nakažejo samo to, da s preizkusnim vzorcem razlik nismo mogli odkriti, da pa lahko obstajajo. Sklepa, da razlik ni, ne moremo narediti, ker ne poznamo napake druge vrste.

Vzemimo, da smo s preizkusnim vzorcem ugotovili, da je povprečna poraba sladkorja v vzorcu $\bar{x} = 7{,}7$ kg. Razlike od hipotetične porabe $M = 8$ kg se pokažejo neznačilne, ker pade $\bar{x}$ v razmak 7,31 kg do 8,39 kg, za katerega je $\alpha = 0{,}05$. To pomeni, da je možno, da med stvarno in hipotetično povprečno porabo ni razlik, ali pa so razlike premajhne, da bi jih mogli odkriti z vzorcem $n = 100$ enot.

Če pri istem preizkusnem vzorcu večamo kritično območje, se napaka prve vrste zveča. To pa je v škodo zanesljivosti sklepov.

Kritično območje, v katerem so razlike značilne, pa moremo povečati tudi tako, da povečamo preizkusni vzorec. Velike razlike med stvarnimi in hipotetičnimi vrednostmi odkrijemo z majhnimi vzorci, majhne razlike pa odkrijejo šele večji vzorci.



Slika 17.4 Napaka druge vrste za preizkusni vzorec n=400 enot

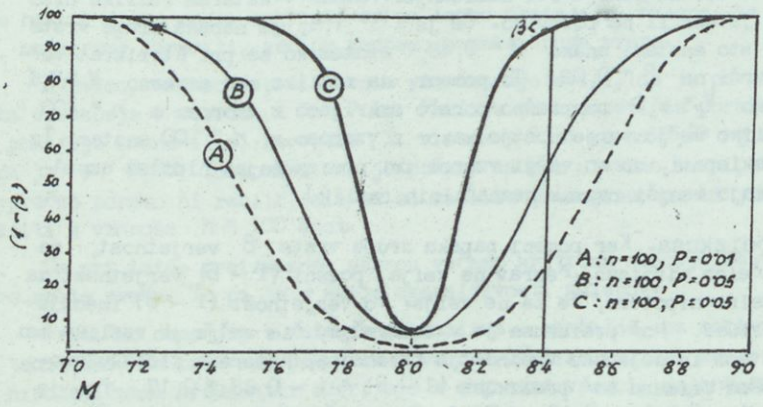
Nadaljujmo s primerom o porabi sladkorja in vzemimo, da je preizkusni vzorec $n = 400$ štirikrat večji kot v prejšnjem primeru. V sliki 7.4 je nakazan za isto ničelno hipotezo $M = M_H = 8,0$ kot zgoraj, ustrezen kritični razmak z $\alpha = 0,05$ napako prve vrste (7,80 do 8,20). V sliki 7.3 je vrisana tudi distribucija povprečij vzorcev, če je stvarno povprečje enako $M = 7,3$ kg.

Iz slike moremo zaključiti naslednje. Razmak v katerem razlike niso značilne, se je skrčil na polovico. Če je $M = 7,3$, je napaka druge vrste pri vzorcu s sto enotami enaka $\beta = 0,48$, medtem ko se pri štirikrat večjem vzorcu skrči na $\beta = 0,02$. To pomeni, da razliko med stvarno $M = 7,3$ in hipotetično $M_H = 8,0$ povprečno porabo odkrijemo z vzorcem z $n = 400$ enotami z veliko večjo verjetnostjo kakor z vzorcem z $n = 100$ enotami. Iz tega moremo sklepati, da so večji vzorci pri preizkušnjah hipotez uspešnejši, ker imajo manjši razmak neznačilnih razlik!

17.6 Moč preizkusa. Ker pomeni napaka druge vrste β verjetnost, da sprejmemo ničelno hipotezo, čeprav ne velja, pomeni $(1 - \beta)$ verjetnost, da zavrnemo ničelno hipotezo, če ta ne velja. To verjetnost $(1 - \beta)$ imenujemo *moč preizkusa*. Moč preizkusa je v tem večja, čim večja je razlika med ničelno hipotezo in dejansko vrednostjo parametra, katerega preizkušamo. Tako je v našem primeru moč preizkusa $(1 - \beta) = 1 - 0,83 = 0,17$, če je $M = 7,8$ kg; $1 - \beta = 1 - 0,48 = 0,52$, če je $M = 7,6$ kg in $(1 - \beta) = 1 - 0,15 = 0,85$, če je $M = 8,6$ kg. Za isto populacijo se moč preizkusa poveča, če je vzorec večji. Tako iz prejšnjih rezultatov sklepamo, da je moč preizkusa, če je $M = 7,3$ kg enaka $(1 - \beta) = (1 - 0,48) = 0,52$, če preizkušamo našo ničelno hipotezo z vzorcem $n = 100$ enotami, in $(1 - \beta) = (1 - 0,02) = 0,98$, če isto hipotezo preizkušamo z vzorcem $n = 400$ enotami. To pomeni: če preizkušamo ničelno hipotezo, da je povprečna poraba sladkorja v štiričlanskih družinah enaka $M_H = 8,0$ kg, stvarna povprečna poraba pa $M = 7,3$ kg, odkrijemo značilne razlike z vzorcem s 100 enotami samo z verjetnostjo $(1 - \beta) = 0,52$, z vzorcem z 400 enotami pa z verjetnostjo $(1 - \beta) = 0,98$. Iz tega sklepamo, da sta napaka druge vrste in moč testa odvisna od prave vrednosti preizkušanega parametra, od napake prve vrste in tipa oziroma velikosti vzorca.

17.7 Operativna karakteristična krivulja. V sliki 17.5 je nakazano za naš primer, kako je napaka druge vrste β oziroma moč testa $(1 - \beta)$ odvisna od prave povprečne porabe sladkorja. Krivulji, ki kažejo odvisnost med temi količinami za določen preizkus, imenujemo *operativno karakteristično krivuljo*. Operativne karakteristične krivulje so zelo važno orodje pri preizkušanju hipotez z vzorci. V sliki 17.5 so za naš primer preizkusa o povprečni potrošnji sladkorja narisane tri karakteristične krivulje. Operativna karakteristična krivulja A velja za preizkus z vzorcem z $n = 100$ enotami, če je napaka prve vrste $\alpha = 0,01$. Operativna karakteristična krivulja B velja za preizkus z vzorcem $n = 100$ enotami in napako prve vrste $\alpha = 0,05$. Če primerjamo operativni karakteristični krivulji za primer A in B, vidimo, da je moč preizkusa $(1 - \beta)$

pri krivulji B večja, vendar gre ta dobiček na račun tega, ker smo vzeli napako prve stopnje večjo. Iz tega sklepamo, da se moč testa zveča, če povečamo napako prve vrste. Seveda tako povečanje moči preizkusa ni vselej umestno. Operativna karakteristična krivulja pa se znatno popravi v primeru C, ki jo dobimo za preizkus z vzorcem $n = 400$ enotami in napako prve vrste $\alpha = 0{,}05$.



Slika 17.5 Operativna karakteristična krivulja

Ocenimo iz slike 17.5 napako druge vrste oziroma moč testa za vse tri plane preizkusnih vzorcev, če je prava povprečna poraba $M = 8{,}4$ kg. Za ta primer dobimo:

Plan preizkusa A: $n = 100$; $\alpha = 0{,}01$; $\beta = 0{,}71$; $(1 - \beta) = 0{,}29$

Plan preizkusa B: $n = 100$; $\alpha = 0{,}05$; $\beta = 0{,}48$; $(1 - \beta) = 0{,}52$

Plan preizkusa C: $n = 400$; $\alpha = 0{,}05$; $\beta = 0{,}02$; $(1 - \beta) = 0{,}98$

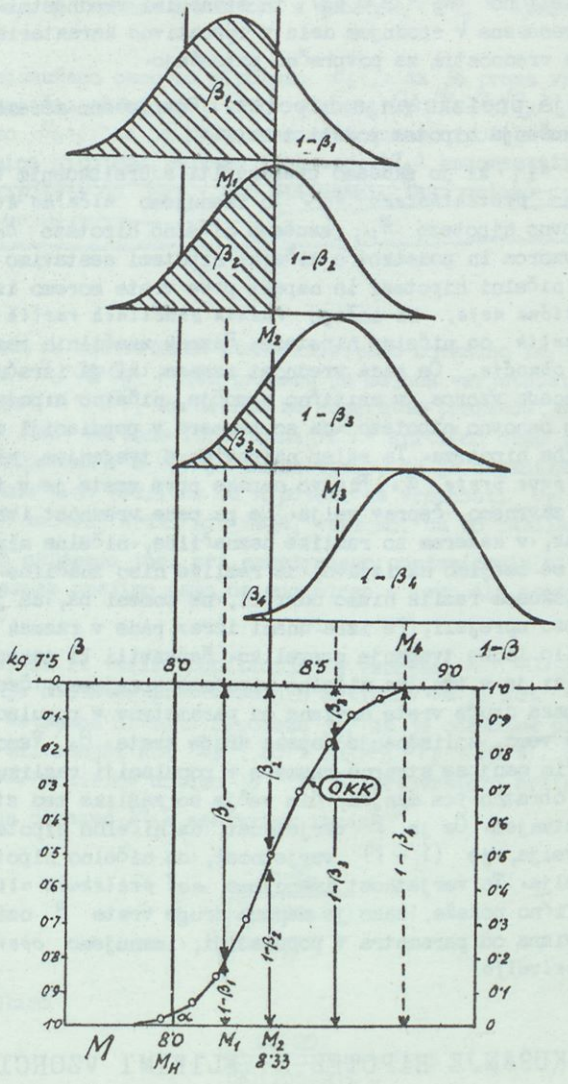
Pri prvem planu bi odkrili značilnost razlik med stvarno ($M = 8{,}4$ kg) in hipotetično ($M_H = 8{,}0$ kg) povprečno porabo sladkorja v štiričlanskih družinah le z verjetnostjo $0{,}28$, pri drugem z verjetnostjo $0{,}51$, pri tretjem pa z verjetnostjo $0{,}94$.

17.8 Operativne karakteristične krivulje pa moremo dobiti tudi za *enostranske preizkuse*. Če preizkušamo z vzorcem z $n = 100$ enotami z napako prve vrste $\alpha = P = 0{,}05$ hipotezo, da je povprečna poraba sladkorja v štiričlanskih delavskih družinah večja kot $M_H = 8{,}00$ kg, postavimo ničelno hipotezo, da je povprečna poraba manjša kot $M_H = 8{,}0$ kg ali enaka $M_H = 8{,}00$ kg.

Iz podatkov, ki jih imamo ($\sigma = 2{,}0$ kg) moremo izračunati kritično mejo za ta preskus. Ta je: $M_H + z_P \sigma/\sqrt{n} = 8{,}0 + 1{,}645 \cdot 2{,}0/\sqrt{100} = 8{,}33$kg

Iz tega napravimo pravilo:

Poraba sladkorja je z napako $\alpha = 0{,}05$ značilno večja kot $8{,}00$ kg, če dobimo s preizkusnim vzorcem $n = 100$ enot povprečje večje kot $\bar{x} = 8{,}33$ kg. Če je povprečje manjše kot $8{,}33$ kg, stvarna povprečna potrošnja ni značilno večja od $8{,}00$ kg.



Slika 17.6 Sestavljanje operativne karakteristične krivulje (OKK) za ničelno hipotezo H: potrošnja sladkorja je enaka ali manjša kot $M_H = 8{,}0$ kg

V sliki 17.3 nazorno vidimo, kako je sestavljena operativna karakteristična krivulja za ta primer. V sliki so narisane vzorčne distribucije za štiri različne povprečne porabe sladkorja M. V navpičnici je narisana kritična meja 8,38 kg. Iz slike spoznamo, da je napaka druge vrste β (ničelno hipotezo sprejmemo, čeprav ne velja) tem večja, čim manjša je razlika med hipotetično (M_H = 8,0 kg) in stvarnimi vrednostmi. Napake druge vrste so prenesene v spodnjem delu v operativno karakteristično krivuljo k ustreznim vrednostim za povprečno potrošnjo.

17.9 Posplošenje preizkušanja hipotez. Posplošeno moremo dosedanji primer preizkušanja hipotez podati takole:

Osnovni hipotezi H_1, ki jo skušamo preizkusiti s preizkusnim vzorcem, priredimo ustrezno *protihipotezo* H_0. To imenujemo *ničelno hipotezo*. Da sprejmemo osnovno hipotezo H_1, skušamo ničelno hipotezo H_0 zavreči. Iz podatkov vzorca in podatkov o ničelni hipotezi sestavimo določen *izraz*. Ustrezno ničelni hipotezi in napaki prve vrste moremo izračunati za ta izraz *kritične meje*, ki ločijo *razmak značilnih razlik* od *razmaka neznačilnih razlik* od ničelne hipoteze. Razmak značilnih razlik imenujemo *kritično območje*. Če pada vrednost izraza, ki ga izračunamo iz podatkov preizkusnega vzorca, v kritično območje, ničelno hipotezo zavrnemo in sprejmemo osnovno hipotezo, da so razmere v populaciji drugačne, kakor pa je ničelna hipoteza. Ta sklep napravimo s tveganjem, ki je v skladu z *napako prve vrste* α. Bistvo napake prve vrste je v tem, da ničelno hipotezo zavrnemo, čeprav velja. Če pa pade vrednost izračunanega izraza v razmak, v katerem so razlike neznačilne, ničelne hipoteze ne sprejmemo, ampak se omejimo na izjavo, da razlike niso značilne. To pomeni, da s tem preizkusom razlik nismo odkrili, ne pomeni pa, da jih ni. Če bi ničelno hipotezo sprejeli, če izračunani izraz pade v razmak neznačilnih razlik, bi bilo lahko tveganje preveliko. Napravili bi lahko *napako druge vrste* β, ki je v tem, da ničelno hipotezo sprejmemo, čeprav ne velja. Ker je napaka druge vrste odvisna od parametrov v populaciji, teh pa ne poznamo, ne vemo, kolikšna je napaka druge vrste β. Vemo le to, da je tem večja čim manj se stvarne razmere v populaciji razlikujejo od hipotetičnih, in obratno tem manjša, čim večje so razlike med stvarnim in hipotetičnim stanjem. Če je β verjetnost, da ničelno hipotezo sprejmemo, čeprav ne velja, je $(1 - \beta)$ verjetnost, da ničelno hipotezo zavrnemo, če ta ne velja. To verjetnost imenujemo *moč preizkusa ali testa*. Krivuljo, ki grafično pokaže, kako je napaka druge vrste β oziroma moč testa $(1 - \beta)$ odvisna od parametra v populaciji, imenujemo *operativno karakteristično krivuljo*.

PREIZKUŠANJE HIPOTEZ Z VELIKIMI VZORCI

17.10 Preizkušanje hipotez z velikimi vzorci je precej enotno za vse parametre, ker vemo iz odstavka 15.15, da se za velike vzorce izraz

$$\frac{\bar{g} - G_0}{S\bar{E}_{\bar{g}}} = z \qquad (17:1)$$

distribuira standardizirano normalno. Pri tem pomeni: G_0 = prava vrednost parametra, $\bar{g}$ = ocena parametra z vzorcem, $S\bar{E}_{\bar{g}}$ = standardna pogreška ocene.

Če preizkušamo osnovno hipotezo H_1:, da je prava vrednost parametra G_0 različna od hipotetične vrednosti G_H ($H_1: G_0 \neq G_H$) postavimo ničelno hipotezo H_0:, da je prava vrednost enaka hipotetični ($H_0: G_0 = G_H$). Če velja ničelna hipoteza, moremo v obrazcu 17.1 nadomestiti pravo vrednost G_0 s hipotetično G_H:, ker sta enaki. Tako dobimo

$$\frac{\bar{g} - G_H}{S\bar{E}_{\bar{g}}} = z \qquad (17.2)$$

Ta izraz se distribuira standardizirano normalno le, če velja ničelna hipoteza $H_0: G_0 = G_H$. V tem primeru je majhna verjetnost (enaka napaki prve stopnje $\alpha = 2P$), da dobimo za konkreten preizkus z izven razmaka $-z_p$ do $+z_p$. Če pade izračunani z v kritično območje, zato ovržemo ničelno hipotezo $H_0: G_0 = G_H$ in sklepamo, da je dejanska vrednost parametra G_0 značilno različna od hipotetične vrednosti G_H. Če pa pade izračunani z znotraj kritičnih mej, vzamemo, da so razlike neznačilne.

Podobno sklepamo tudi pri enostransko postavljenih hipotezah, le da kritični razmak omejimo samo na eno stran – levo ali desno. (Glej odstavek 14.22).

17.11 Vzemimo, da iz velike populacije, z vzorcem $n = 400$ artiklov preizkušamo, ali je odstotek prvorazrednih artiklov večji kot $P_H = 20\,\%$ ($H_1: P_0 > P_H$). Ustrezna ničelna hipoteza je, da je odstotek prvovrstnih artiklov enak ali manjši kot $P_H = 20\,\%$ ($H_0: P_0 < P_H = 20\,\%$). S preizkusnim vzorcem $n = 400$ artiklov dobimo $p = 25\,\%$ prvovrstnih artiklov.

Glede na obrazec 17.2 sestavimo izraz

$$z = \frac{p - P_H}{\sqrt{\frac{P_H(100 - P_H)}{n}}} \qquad (17.3)$$

Iz njega pa dobimo

$$z = \frac{25 - 20}{\sqrt{\frac{20(100 - 20)}{400}}} = +2,5$$

Glede na zgornje pogoje sklepamo, da je stvarni odstotek prvorazrednih artiklov v proizvodnji na stopnji 0,05 značilno večji kot $P_H = 20\,\%$, če

dobimo, da je po obrazcu 17.3 izračunani z večji kot $z_{0,05} = +1,64$, na stopnji 0,01, če je večji kot $z_{0,01} = +2,32$ in značilno večji na stopnji 0,001, če je večji kot $z_{0,001} = +3,09$.

Ker je izračunani $z = +2,5$, vzamemo, da ima stvarna proizvodnja značilno večji odstotek prvovrstnih artiklov na stopnji 0,01, ker je $z = +2,5$ večji kot $z_{0,01} = +2,32$ in manjši kot $z_{0,001} = 3,09$.

PREIZKUŠANJE HIPOTEZ Z MALIMI VZORCI

17.12 Z malimi vzorci preizkušamo hipoteze podobno kot z velikimi. Razlika je le v tem, da za preizkušanje hipotez o posameznih parametrih uporabljamo izraze, ki se distribuirajo v distribucijah za male vzorce.

Če to upoštevamo, preizkušamo v splošnem hipoteze z malimi vzorci po temle postopku:

a) Osnovni hipotezi H_1 o določenem parametru ali lastnosti populacije poiščemo ustrezno ničelno hipotezo H_0.

b) V zbirki obrazcev od 18.17 do 18.23 za male vzorce poiščemo ustrezni izraz, ki vsebuje preizkušani parameter ali ima lastnost, ki ustreza ničelni hipotezi.

c) V ta obrazec vstavimo ustrezne količine, izračunane iz preizkusnega vzorca, namesto parametrov populacije pa parametre, kot jih predpisuje ničelna hipoteza. Če ničelna hipoteza velja, je izračunana vrednost izraza v skladu z zakonitostmi o malih vzorcih in s tveganjem α ne prekorači ustreznih kritičnih vrednosti. Zato z dano napako prve vrste α sklepamo, da ničelna hipoteza ne velja, če je izračunana vrednost v kritičnem območju.

d) V tablicah o teoretičnih distribucijah poiščemo teoretični distribuciji in stopinjam prostosti ustrezno kritično vrednost na različnih stopnjah.

e) Če je stvarno izračunana vrednost manjša kot kritična vrednost na stopnji 0,05, vzamemo da razlike *niso značilne*. Če je stvarno izračunana vrednost večja od kritične vrednosti za stopnjo 0,05, vendar manjša kot kritična vrednost za stopnjo 0,01, vzamemo da so razlike stvarnega parametra od hipotetičnega *značilne na stopnji 0,05*. Če je izračunana vrednost izraza večja kot kritična vrednost za stopnjo 0,01, vendar manjša kot kritična vrednost za stopnjo 0,001, vzamemo da so stvarne razmere v populaciji značilno različne od hipotetične *na stopnji 0,01*. Če pa je izračunani izraz večji kakor ustrezna kritična vrednost na stopnji 0,001, vzamemo da so razlike *značilne na stopnji 0,001*.

17.13 Preizkušanje hipotez o aritmetični sredini. Vzemimo kot prvo nalogo, ki jo moremo rešiti s preizkušanjem hipoteze z malimi vzorci

tale problem: Cena določenega artikla dnevne porabe se je znižala. Preiskati je treba, ali je znižanje vplivalo na povečanje prodaje za ta artikel. Osnovna hipoteza je, da se je povprečna prodaja na trgovino M povečala $(H_1: M_2 > M_1;\ \Delta M = M_2 - M_1 > 0)$ M_1 = povprečna prodaja pred pocenitvijo, M_2 = povprečna prodaja po pocenitvi. Tej hipotezi ustreza ničelna hipoteza, da se prodaja zaradi spremembe cene ni povečala $(H_0: M_2 = M_1,\ \Delta M = 0)$. To ničelno hipotezo moremo najlepše preizkusiti takole: Izberemo slučajnostni vzorec n trgovin in zanje zberemo podatke o mesečni prodaji za obravnavani artikel v mesecu pred (x_1) in mesecu po (x_2) pocenitvi. Za vsako trgovino posebej moremo iz teh podatkov izračunati spremembo v prodaji $d = x_2 - x_1$. Ker preizkušamo ničelno hipotezo, da je povprečje povečanja prodaje $M_d = 0$, je ustrezni vzorčni izraz obrazec 13.18. Ker proučujemo podatek d, v obrazcu 13.18 zamenjamo $\bar{x}$ z $\bar{d}$, s z s_d, pravo povprečje pa s hipotetičnim $M_d = 0$. Tako dobimo

$$t = \frac{\bar{d}}{s_d}\sqrt{n} \qquad (17.4)$$

Če velja ničelna hipoteza $M_d = 0$, je ta izraz v skladu z zakonitostmi s t-distribucijo z $m = n-1$ stopinjami prostosti. Zato vzamemo, da je pravo povprečje razlik v prodaji značilno različno od nič, če po obrazcu 17.4 izračunani izraz večji kakor ustrezna kritična vrednost za t.

17.14 Vzemimo za ilustracijo postopka shematičen primer z vzorcem $n = 9$ trgovin

Tabela 17.2 Preizkušanje hipoteze o zvečanju prometa z določenim artiklom po pocenitvi (shematičen primer)

Zap. št. trgovine	x_1	x_2	d	d^2
1	23	25	+2	4
2	20	19	−1	1
3	27	30	+3	9
4	23	24	+1	1
5	21	22	+1	1
6	18	17	−1	1
7	29	32	+3	9
8	21	21	0	0
9	29	30	+1	1
			+9	27
			Sd	Sd^2
			D	

$$\bar{d} = Sd/n = +9/9 = +1 \; ; \quad s_d^2 = \frac{Sd^2 - D^2/n}{n-1} = \frac{27 - 9^2/9}{9-1} = 2,25$$

$$sd = \sqrt{2,25} = 1,5$$

Če te količine vstavimo v obrazec 17.4, dobimo

$$t = \frac{+1}{1,5}\sqrt{9} = +2,0$$

Izračunani t je $t = +2,0$. Ker je število stopinj prostosti $m = n-1 = 9-1 = 8$, preizkus pa enostranski, dobimo v tabeli za t-distribucijo ustrezni kritični vrednosti $t_{0,05}(m=8) = 1,86$ in $t_{0,01}(m=8) = 2,31$. Iz teh podatkov sklepamo, da je povečanje prodaje značilno različno od nič s tveganjem 0,05, ker pade izračunana vrednost $+2,00$ med 1,86 in 2,31. Navedeni postopek uporabljamo v praksi vselej, kadar ugotavljamo spremembe v povprečjih za dva podatka, za katera moremo sestaviti pare za posamezne enote.

17.15 Preizkušanje hipotez o varianci. Z obrazcem 13.19 preizkušamo hipoteze o varianci v proučevani populaciji

$$\chi^2 = \frac{(n-1)s^2}{\sigma_H^2} \tag{17.5}$$

z $m = n-1$ stopinjami prostosti.

17.16 Podjetje, ki je dobavilo drugemu podjetju stroj za proizvodnjo vijakov, je navedlo kot karakteristiko kvalitete dela stroja standardni odklon dolžine vijakov $\sigma = 1$ mm. Podjetje, ki je stroj kupilo, skuša preizkusiti to karakteristiko in postavlja hipotezo, da je stvarni standardni odklon večji kot 1 mm ($H_1: \sigma_0 > \sigma_H = 1$). Tej hipotezi postavijo ustrezno ničelno hipotezo, da je stvarni standardni odklon enak 1 mm ($H_0: \sigma_0 = \sigma_H = 1$). To hipotezo preizkušamo z majhnim vzorcem $n = 30$ vijakov, ki jih proizvedemo v preizkusni proizvodnji. Iz teh artiklov ocenimo standardni odklon $s = 1,1$ mm. Preizkusiti je treba značilnost razlik med stvarno in hipotetično variabilnostjo. Ker potrebujemo v obrazcu variance, ne pa standardne odklone, izračunamo: $\sigma_H^2 = 1$; $s^2 = 1,21$.

Če te podatke vstavimo v obrazec 17.5, dobimo

$$\chi^2 = \frac{(30-1)\cdot 1,21}{1} = 35,09$$

Ustrezna kritična vrednost $\chi^2_{0,05}(m=30-1=29) = 42,53$

Ker je izračunana vrednost χ^2 manjša kot kritična vrednost za tveganje $0,05$, vzamemo, da razlike niso značilne. To pomeni, da preizkusni vzorec ni odkril, da stvarna variabilnost ni v skladu z naznačeno kvaliteto.

17.17 Preizkušanje hipotez o korelacijskem koeficientu. Če preizkušamo hipotezo, da sta dva pojava v korelaciji $(H_1: r_0 \neq 0)$, postavimo ničelno hipotezo, da korelacijske povezave ni $(H_0: r_0 = 0)$. To ničelno hipotezo preizkušamo z obrazcem 13.20, ki velja za primer, če dva pojava nista povezana.

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \qquad (17.6)$$

z $m = n-2$ stopinjami prostosti.

17.18 Vzemimo, da preizkušamo, ali je kvaliteta dela, ki jo merimo z odstotkom prvovrstnih izdelkov, odvisna od starosti delavca. Ničelno hipotezo, da kvaliteta izdelkov ni odvisna od starosti delavca, preizkušamo s preizkusnim vzorcem $n = 50$ delavcev. Za ta vzorec dobimo $r = +0,45$. Iz teh podatkov izračunamo po obrazcu 17.3, da je

$$t = \frac{+0,45}{\sqrt{1-0,45^2}}\sqrt{50-2} = +3,49$$

Ker je preizkus obojestranski, najdemo v tabeli za t-distribucijo za $m = n-2 = 50-2 = 48$ za $\alpha = 2P = 0,01$, $t_{0,005}(m=48) = 2,38$ za $t_{0,0005}(m=48) = 3,51$. Zato vzamemo, da je korelacija značilna, in sicer s tveganjem $0,01$. To pomeni, da je verjetnost, da bi dobili tak korelacijski koeficient, če starost in kvaliteta dela ne bi bili povezani, manjša kot $0,01$.

17.19 Če preizkušamo hipotezo, da korelacijski koeficient za določena pojava v populaciji *različen od* r_H, ki pa ni nič, moramo uporabiti obrazec 13.22.

Ker je osnovna hipoteza, da je korelacijski koeficient različen od $r_H (H_1: r_0 \neq r_H)$, je ničelna hipoteza, da je korelacijski koeficient v populaciji enak hipotetičnemu $(H_0: r_0 = r_H)$.

V tem primeru se izraz

$$z = (Z-Z_H)\sqrt{n-3} \qquad (17.7)$$

distribuira v standardizirani normalni distribuciji. Pri tem so Z in Z_0 količine, ki jih izračunamo po obrazcu 13.21.

17.20 Če za korelacijo med mehanskim in prostorninskim testom za tretješolce postavljamo hipotezo, da je korelacijski koeficient različen od

$r_H = 0,50$, je ustrezna ničelna hipoteza, da je korelacijski koeficient med dosežki obeh testov enak $r_H = 0,50$. To ničelno hipotezo preizkušamo z vzorcem $n = 30$ tretješolcev. Iz podatkov vzorca dobimo $r = 0,76$.

Če transformiramo korelacijske koeficiente, ustreza po tabeli r–Z korelacijskemu koeficientu $r_H = 0,50$ koeficient $Z_H = 0,55$; $r = 0,76$ pa $Z = 1,00$. Iz teh podatkov dobimo

$$z = (1,00-0,55)\sqrt{20-3} = +1,85$$

Ker bi moral z ležati izven razmaka $-1,96$ do $+1,96$, da bi bile razlike s tveganjem $\alpha = 0,05$ značilne, vzamemo, da korelacijski koeficient ni značilno različen od $r_H = 0,50$.

17.21 Preizkušanje hipotez o razlikah med aritmetičnimi sredinami. V enem izmed prejšnjih primerov smo preizkušali značilnost razlik med aritmetičnimi sredinami za dve odvisni populaciji. Za ta primer smo preizkušali značilnost razlik v povprečjih prek razlik za ustrezne enote. Vendar niso v vseh primerih primerjani populaciji med seboj odvisni kakor pri proučevanju vpliva spremembe cene na prodajo. Za preizkušanje razlik v povprečjih za dve neodvisni populaciji pa uporabljamo obrazec 16.23. Kakor vidimo iz podstavka 16.15, moremo uporabiti ta obrazec le, če predpostavljamo, da sta varianci v obeh populacijah enaki.

Običajno po tem obrazcu preizkušamo hipotezo, da aritmetični sredini v obeh populacijah nista enaki $(H_1: M_2 \neq M_1)$. Tej hipotezi ustreza ničelna hipoteza, da sta aritmetični sredini v obeh populacijah enaki $(H_0: M_2 = M_1)$. Če hočemo po tem obrazcu preizkusiti značilnost razlik med aritmetičnima sredinama v dveh populacijah, *moramo*, kakor kaže obrazec 16.23, iz vsake izmed proučevanih populacij izbrati vzorec. Iz vzorca iz prve populacije z n_1 enotami izračunamo $\bar{x}_1$ in s_1^2. Enako izračunamo tudi iz drugega vzorca z n_2 enotami $\bar{x}_2$ in s_2^2. Te količine skupno s hipotetičnimi $(M_2 - M_1 = 0)$ vstavimo v obrazec 16.23. Tako dobimo

$$t = \frac{\bar{x}_2 - \bar{x}_1}{s_d}\sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}} \quad ; \quad s_d^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} \qquad (17.8)$$

Če je izračunati t z ustreznim tveganjem v kritičnem razmaku $-t_p$ do $+t_p$, pri čemer moramo upoštevati stopinje prostosti $m = n_1+n_2-2$, vzamemo, da sta M_1 in M_2 med seboj značilno različna na stopnji $\alpha = 2P$.

17.22 Vzemimo za primer proučevanje izdatkov za obleko in obutev na enega člana v letu 1957 v kmetijskih gospodarstvih s površino 3–8 ha v Sloveniji. Hipoteza, ki jo moramo preizkusiti, je, da je povprečje v izdatkih za obleko in obutev na enega člana v kmetijskih gospodarstvih v primorskih okrajih (Gorica, Koper) različna od povprečnih izdatkov za obleko in obutev na enega člana v kmetijskih gospodarstvih v štajerskih okra-

jih (Maribor, Ptuj, Murska Sobota). Hipoteza je dvostranska, ker iščemo samo značilnost razlik, ne predpostavljamo pa, v katerem predelu je povprečje večje. Iz podatkov ankete o življenju kmetov ZS LRS dobimo za slučajnostno izbrana gospodarstva tele osnovne podatke (v tisočih dinarjev na člana):

Primorska: $n_1 = 7$; $x_{1i} = 8,3$; $5,2$; $14,2$; $8,8$; $6,2$; $16,4$; $16,4$;

Štajerska: $n_2 = 14$; $x_{2i} = 3,6$; $4,5$; $2,7$; $2,3$; $1,0$; $3,2$; $3,1$; $8,6$; $9,1$; $1,3$; $11,0$; $2,1$; $5,6$; $4,4$;

Iz teh podatkov dobimo dalje:

$$X_1 = 74,8 \quad ; \quad \bar{x}_1 = 10,69 \quad ; \quad Sx_1^2 = 937,96 \quad ;$$

$$s_1^2 = \frac{Sx_1^2 - X_1^2/n_1}{n_1 - 1} = \frac{937,96 - 74,8^2/7}{7 - 1} = 23,11$$

$$X_2 = 60,5 \quad ; \quad \bar{x}_2 = 4,32 \quad ; \quad Sx_2^2 = 382,83 \quad ;$$

$$s_2^2 = \frac{Sx_2^2 - X_2^2/n_2}{n_2 - 1} = \frac{382,83 - 60,5^2/14}{14 - 1} = 9,84$$

Iz teh podatkov je po obrazcu 17.8

$$s_d^2 = \frac{(7-1)\cdot 23,11 + (14-1)\,9,84}{7 + 14 - 2} = 13,39$$

$$s_d = \sqrt{13,39} = 3,70$$

Če te podatke vstavimo v obrazec 17.8, dobimo

$$t = \frac{4,32 - 10,39}{3,70}\sqrt{\frac{7\cdot 14}{7+14}} = -3,73$$

Ker je število stopinj prostosti $m = n_1 + n_2 - 2 = 7 + 14 - 2 = 19$, sta za tveganje $\alpha = 2P = 0,01$ kritični meji za t $-2,86$ do $+2,86$, za tveganje $\alpha = 2P = 0,001$ pa $-3,88$ do $+3,88$. Iz tega sklepamo, da so razlike med obema predeloma značilne na stopnji $0,01$; to je razmeroma visoka značilnost.

17.23 Preizkušnje hipoteze o razlikah med variancami. Z obrazcem 13.24 preizkušamo hipoteze o razlikah v variancah v dveh populacijah. Če je osnovna hipoteza, da sta v dveh populacijah varianci različni

$(H_1: \sigma_1^2 \neq \sigma_2^2)$ je ustrezna ničelna hipoteza, da sta varianci v obeh populacijah enaki $(H_0: \sigma_1^2 = \sigma_2^2)$. Če to upoštevamo, izraz

$$F = s_1^2 / s_2^2 \qquad (17.9)$$

zanešča zakonitostim v F-distribuciji z $m_1 = n_1 - 1$ in $m_2 = n_2 - 1$ stopinjami prostosti. Po ustaljenem pravilu sklepamo, da ničelna hipoteza ne velja in da so razlike v variancah značilne, če je ta izraz, izračunan iz stvarnih vrednosti, večji kakor kritična meja za določeno napako prve vrste. Pripomniti moramo, da mora biti kvocient varianc vedno večji od ena in zato primerjamo večjo varianco z manjšo.

17.24 Če preizkušamo značilnost razlik med variancama za prejšnji primer izdatkov za obleko in obutev na enega člana v kmetijskih gospodarstvih na Primorskem in Štajerskem, dobimo

$$F = s_1^2 / s_2^2 = 23,11 / 9,34 = 2,47$$

Če primerjamo izračunani $F = 2,47$ s kritično vrednostjo, ki jo dobimo v tablici za F-distribucijo za $m_1 = n_1 - 1 = 7 - 1 = 3$ in $m_2 = 14 - 1 = 13$ stopinjami prostosti $F_{0,05}(m_1 = 3; m_2 = 13) = 2,90$, spoznamo, da razlike niso značilne, ker je izračunani F manjši kot kritična vrednost na stopnji 0,05.

17.25 Preizkušanje hipotez o razlikah med korelacijskimi koeficienti. Obrazec 13.25 uporabljamo za preizkušanje značilnosti razlike med dvema korelacijskima koeficientoma. Po osnovni hipotezi sta korelacijska koeficienta r_{01} in r_{02} za ave populaciji med seboj različna $(H_1: r_{01} \neq r_{02})$. Tej hipotezi ustreza ničelna hipoteza, da sta korelacijska koeficienta v obeh populacijah enaka. $(H_0: r_{01} = r_{02})$. Če velja ničelna hipoteza, se izraz

$$z = \frac{Z_2 - Z_1}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \qquad (17.10)$$

distribuira standardizirano normalno. Ta izraz dobimo iz obrazca 13.25, ker je zaradi ničelne hipoteze $r_{01} = r_{02}$ tudi $Z_{01} = Z_{02}$. Kakor vemo, so Z_1 in Z_2 transformirane količine iz tabele r-Z, ki ustrezajo oceni r_1 iz prve populacije z n_1 enotami, in oceni r_2 iz druge populacije z n_2 enotami.

Po znanem pravilu vzamemo, da je razlika med korelacijskima koeficientoma značilna, če je po obrazcu 17.10 izračunani izraz v kritičnem območju, to je izven razmaka $-z_p$ do $+ z_p$.

17.26 Vzemimo, da preizkušamo hipotezo o različni korelaciji med dohodki

in izdatki za tobak in pijačo med obema skupinama prebivalstva. Vzemimo, da smo za prvo skupino z vzorcem $n_1 = 30$ gospodinjstev dobili ustrezni $r_1 = 0,48$, z vzorcem $n_2 = 50$ gospodinjstev iz druge skupine pa $r_2 = 0,39$. Ker hočemo preizkusiti značilnost razlik med korelacijskima koeficientoma za obe skupini, izračunamo izraz 17.10. Prej pa v tabeli $r - Z$ poiščemo r_1 in r_2 ustrezni Z_1 in Z_2. Iz tabele $r-Z$ dobimo $Z_1 = 0,52$; $Z_2 = 0,85$. Če te podatke vstavimo v obrazec 17.10, dobimo:

$$z = \frac{0,85 - 0,52}{\sqrt{\frac{1}{30-3} + \frac{1}{50-3}}} = +1,37$$

Ker je izračunani z manjši kot $+1,93$, ki je kritična zgornja meja za dvostransko preizkušnjo za standardizirano normalno distribucijo s tveganjem $0,05$, sklepamo, da razlika med korelacijskima koeficientoma ni značilna.

17.27 Preizkušanje hipotez o frekvenčnih distribucijah. Z izrazom v obrazcu 13.23 preizkušamo hipoteze o frekvenčnih distribucijah v populacijah. Z njim preizkušamo osnovno hipotezo, da frekvence v osnovni populaciji f_o niso enake teoretičnim frekvencam f'. ($H_1 : f_o \neq f'$). Tej hipotezi ustreza ničelna hipoteza, da so frekvence v osnovni populaciji f_o enake teoretičnim frekvencam f', ki jih dobimo z določeno predpostavko o populaciji ($H_o : f_o = f'$).

Če velja ničelna hipoteza, se izraz

$$\chi^2 = \Sigma \frac{(f-f')^2}{f'} \qquad (17.11)$$

distribuira v χ^2-distribuciji z m stopinjami prostosti. Obrazec 17.11 dobimo iz obrazca 13.27 tako, da f_o zamenjamo z f', ker se zaradi ničelne hipoteze skladata. f v obrazcu 17.11 pa so stvarne frekvence iz preizkusnega vzorca.

Zato vzamemo, da je stvarna distribucija f_o značilno različna od teoretične hipotetične f' s stopnjo $\alpha = P$, če izračunana vrednost za χ^2 prekorači kritično vrednost χ^2_p. χ^2_p so za različne stopnje α in stopinje prostosti m dane v tabeli za χ^2-distribucijo.

17.28 Vzemimo za primer distribucijo tež baterskih svitkov v eni izmed tekstilnih tovarn v Sloveniji iz odstavka 14.12.

Preizkušamo osnovno hipotezo, da registracija tež baterskih svitkov ni nepristranska in pravilna, ker imajo delavci, ki sami registrirajo težo baterskih svitkov, vzrok, da navajajo napačne teže. Tej osnovni hipotezi postavimo ustrezno ničelno hipotezo, da registracija tež ni pri-

stranska. Ker v primeru, da registracija ni pristranska, predpostavljamo, da so razlike v teži zgolj slučajne, je teoretična distribucija tež baterskih svitkov normalna distribucija.

V tabeli 14.2 imamo stvarno frekvenčno distribucijo $n = 561$ baterskih svitkov, ki jih vzamemo za vzorec iz hipotetične populacije vseh baterskih svitkov, ki bi bili tehtani v enakih pogojih. Razen tega je v tej tabeli izračunana tudi teoretična normalna distribucija, ki ima enako aritmetično sredino in standardni odklon kot stvarna distribucija. V tabeli 17.3 je prikazan izračun χ^2 za ta primer.

Tabela 17.3 Izračunavanje χ^2 za frekvenčno distribucijo $n = 561$ baterskih svitkov po registrirani teži

Teža kg	f		f'		$f - f'$	$(f - f')^2$	$(f - f')^2/f'$
18,55	3		0,0				
18,60	-		0,3				
18,65	14	28	2,6	17,2	+10,8	116,64	6,78
18,70	11		14,3				
18,75	23		49,6		-26,6	707,56	14,27
18,80	71		108,3		-37,3	1391,29	12,85
18,85	229		149,9		+79,1	6256,81	41,74
18,90	129		129,7		- 0,7	0,49	0,00
18,95	50		73,6		-23,6	556,96	7,57
19,00	11		23,0		-15,0	225,00	8,65
19,05	18		5,8				
19,10	2	20	0,8	3,7	+13,3	176,89	26,40
19,15	-		0,1				
	$n = 561$		561,0		0		98,26 = χ^2

Ker se izraz 17.1 „distribuira v χ^2-distribuciji le, če so teoretične frekvence večje kot 5, smo v našem primeru skrajne frekvence, ki so manjše kot 5, združili s sosednimi razredi. Seveda smo morali združiti tudi ustrezne razrede v stvarni frekvenčni distribuciji.

Vprašanje, ki ga moramo rešiti, je še število stopinj prostosti. Po združitvi razredov z majhnimi frekvencami imamo $k = 8$ grup. Variiranje v frekvencah pa je omejeno s tremi pogoji: s skupnim številom enot n, enako aritmetično sredino $\bar{x}$ in enako varianco s^2. Število stopinj prostosti se v tem sistemu zniža za 3. Skupno število stopinj prostosti

je torej $m = k-3 = 8-3 = 5$. $m = 5$ ustrezna kritična vrednost $\chi^2_{0,001}(m=5) = 20,52$. Izračunana vrednost za χ^2 je veliko večja kot kritična vrednost z $\alpha = 0,001$. Iz tega sklepamo, da so razlike med stvarno distribucijo f_o in hipotetično distribucijo f' zelo značilne in da je pravilna osnovna hipoteza, da je registracija teže baterskih svitkov pristranska. Če podrobneje proučimo razlike med vzorčno in teoretično distribucijo, opazimo izredno velike razlike v razredu 18,85 kg, v katerem dobimo kopičenje v vzorcu $f-f' = +79,1$. Ker je 18,85 kg predpisana teža za baterski svitek, je to kopičenje psihološko utemeljeno. Enako opazimo kopičenje, čeprav ni tako intenzivno, tudi za teži 18,35 kg in 19,05 kg. Tudi to kopičenje je utemeljeno, ker sta ti teži meji uporabnosti baterskih svitkov.

17.29 Preizkušanje hipoteze o odvisnosti med atributivnimi znaki. S χ^2-distribucijo moremo preizkušati tudi značilnost odvisnosti med atributivnimi znaki. Že pri obravnavanju kontingence smo uvedli kot merilo kontingence χ^2, ki ga izračunamo po osnovnem obrazcu 13.45. Pri tem teoretične frekvence za primer, da sta znaka nepovezana, izračunamo iz robnih frekvenc po obrazcu 13.44.

Z enako upravičenostjo kot v prejšnjem primeru za normalno distribucijo moremo sklepati, da se izraz

$$\chi^2 = \sum \frac{(f-f')^2}{f'} \quad ; \quad f' = f'_{kg} = \frac{f_k f_g}{n} \qquad (17.12)$$

distribuira v χ^2-distribuciji, če so frekvence f rezultati vzorca iz populacije, v kateri sta znaka, ki ju proučujemo, neodvisna. S tem izrazom moremo torej preizkušati hipotezo, da sta proučevana znaka med seboj odvisna. Ustrezna ničelna hipoteza pa je, da pojava nista odvisna.

Ker se pri tej ničelni hipotezi izraz 17.12 distribuira v χ^2-distribuciji, sklepamo, da sta proučevana znaka značilno odvisna, če je izračunani χ^2 večji kot ustrezna kritična vrednost za $\chi^2_{0,05}$.

17.30 V tabeli 13.1 smo navedli kontingenčno tabelo med številko preje N_m in vrsto napak za pretrge na prstančevih strojih. S χ^2-preizkusom moremo preizkusiti hipotezo, da so vrste napak odvisne od številke – N_m preje.

Prvo teoretično frekvenco izračunamo po obrazcu 17.2.

$f'_{11} = \frac{35 \cdot 212}{748} = 18,4$. Podobno izračunamo tudi vse druge teoretične frekvence.

Iz teh podatkov izračunamo χ^2 po obrazcu 17.12.

Tabela 17.4 Število pretrgov na prstančevih strojih za številke preje Nm 14, 34, 40 in 50 po vrstah napak (M=mehanske, P=posluževalne: O=ostale) v eni izmed slovenskih tekstilnih tovarn.

Stvarna frekvenčna distribucija

f

Nm preje	Vrsta napak			S_k
	M	P	O	
14	12	48	5	65
34	45	258	58	331
40	100	88	26	214
50	55	41	12	108
S_k	212	435	101	748

Teoretična frekvenčna distribucija

f'

Nm preje	Vrsta napak			S_k
	M	P	O	
14	18,4	37,8	8,8	65,0
34	102,3	209,9	48,7	360,9
40	60,7	124,5	28,9	214,1
50	30,6	62,8	14,6	108,0
S_k	212,0	435,0	101,0	748,0

$$\chi^2 = \frac{(12-18,4)^2}{18,4} + \frac{(48-37,8)^2}{37,8} + \ldots + \frac{(12-14,6)^2}{14,6} = 115,42$$

Že v odstavku 13.8 smo ugotovili, da je za kontingenčne tabele število stopinj prostosti $m = (k-1).(g-1)$. Za naš primer dobimo torej $m = (3-1).(4-1) = 6$. Ker je kritična vrednost za $\chi^2_{0,001}(6) = 22,46$, sklepamo, da je sestav napak, zaradi katerih se preja na prstančevih strojih trga, značilno različen za različne vrste preje.

ANALIZA VARIANCE

17.31 Preizkušanje hipotez o varianci se je razvilo v samostojno in zelo pomembno področje statistične analize. Vse planiranje eksperimentov ima za osnovo analizo variance v najrazličnejših oblikah. V tem odstavku bomo nakazali osnovo analize variance in še najenostavnejši primer uporabe.

Vzemimo da merimo produktivnost dela (čas, ki ga porabijo za izdelavo enega izdelka) štirih delavcev. Za vsakega delavca imamo podatke o času za pet izdelkov. Proučujmo variabilnost teh podatkov. Ti podatki so shematično nakazani v tabeli 17.5, v kateri so v vsaki vrsti nanizani podatki za posameznega delavca. Predpostavljamo, da je produktivnost dela vseh štirih delavcev enaka. Seveda zaradi slučajnih vplivov čas za posamezne izdelke ni enak, marveč variira.

Če ni razlik v produktivnosti dela med delavci, moremo varianco v produktivnosti dela oceniti na več načinov.

Tabela 17.5 Shema podatkov o produktivnosti dela za štiri delavce, ki so proizvedli po 5 izdelkov

Delavec 1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	Y_1	$\bar{x}_1$
Delavec 2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	Y_2	$\bar{x}_2$
Delavec 3	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	Y_3	$\bar{x}_3$
Delavec 4	x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	Y_4	$\bar{x}_4$
						Y	$\bar{x}$

Prvo oceno variance dobimo, če izračunamo po znanem obrazcu vsoto kvadratov odklonov od skupne aritmetične sredine $\bar{x}$, dobljeno vsoto pa delimo s številom stopinj prostosti m $4,5 - 1 = 19$. Z obrazcem moremo to pisati

$$s^2 = \frac{\sum_k \sum_i (x_{ki} - \bar{x})^2}{4.5-1} = \frac{\sum_k \sum_i x_{ki}^2 - Y^2/4.5}{4.5-1} = \frac{V}{4.5-1}$$

Varianco pa moremo oceniti tudi iz petih podatkov za posameznega delavca:

$$s_1^2 = \frac{\sum_i (x_{1i} - \bar{x}_1)^2}{5-1} = \frac{\sum_i x_{1i}^2 - Y_1^2/5}{5-1}$$

$$s_2^2 = \frac{\sum_i (x_{2i} - \bar{x}_2)^2}{5-1} = \frac{\sum_i x_{2i}^2 - Y_2^2/5}{5-1}$$

$$s_3^2 = \frac{\sum_i (x_{3i} - \bar{x}_3)^2}{5-1} = \frac{\sum_i x_{3i}^2 - Y_3^2/5}{5-1}$$

$$s_4^2 = \frac{\sum_i (x_{4i} - \bar{x}_4)^2}{5-1} = \frac{\sum_i x_{4i}^2 - Y_4^2/5}{5-1}$$

Tako dobimo štiri neodvisne ocene variance, vsako z $m=5-1$ stopinjami prostosti. Iz teh štirih grupnih varianc moremo izračunati povprečje iz posameznih grupnih varianc. Če bi bilo število podatkov za posamezne grupe različno, bi morali seveda izračunati tehtano povprečje in uporabiti za pondere število stopinj prostosti. Ker pa je v našem primeru število podatkov v grupah enako, dobimo isti rezultat, če izračunamo navadno povprečje. Tako dobimo, da je

$$s_e^2 = \bar{s}^2 = \frac{1}{4} \sum_k s_k^2 = \frac{1}{4} \sum_k \frac{\sum_i (x_{ki} - \bar{x}_k)^2}{5-1} = \frac{\sum_k \sum_i x_{ki}^2 - \sum_k Y_k^2/5}{4(5-1)} = \frac{V_e}{4(5-1)}$$

Ta ocena je izračunana z $m = 4(5-1)$ stopinjami prostosti, ker smo jo izračunali iz štirih grup po pet podatkov, od katerih je po eden vezan na grupno aritmetično sredino.

Razen teh dveh ocen pa imamo še tretjo oceno, ki jo dobimo iz grupnih aritmetičnih sredin. Če variirajo individualne vrednosti, variirajo tudi sredine, izračunane iz teh vrednosti. Vendar je variacija med sredinami manjša kot med individualnimi vrednostmi. Iz zakonitosti o variiranju sredin vemo, da je varianca sredin $\sigma^2_{\bar{x}} = \sigma^2/n$, ali v našem primeru $\sigma^2_{\bar{x}} = \sigma^2/5$.

Če z obrazcem

$$s^2_{\bar{x}} = \frac{S_k(\bar{x}_k - \bar{\bar{x}})^2}{4-1} = \frac{S_k \bar{x}_k^2 - (S_k \bar{x}_k)^2/4}{4-1}$$

izračunamo oceno variance grupnih aritmetičnih sredin, to torej ni ocena za variabilnost produktivnosti dela, marveč ocena za $\sigma^2/5$. Če hočemo, da bo to ocena za variabilnost o produktivnosti dela, moramo $s^2_{\bar{x}}$ pomnožiti s 5. Tretja ocena za varianco o produktivnosti dela je torej

$$s^2_K = 5\, s^2_{\bar{x}} = \frac{5[S_k(\bar{x}_k - \bar{\bar{x}})^2]}{4-1} = \frac{S_k X_k^2/5 - X^2/20}{4-1} = \frac{K_K}{4-1}$$

Ta ocena je izračunana z $m = 4-1 = 3$ stopinjami prostosti, ker smo jo izračunali iz štirih grupnih sredin, ki so vezane na skupno sredino.

Če pozorneje pogledamo obrazce za posamezno vrsto ocen za varianco o produktivnosti dela opazimo, da je med njimi dobljena zveza.

$$K = S_{ki} x^2_{ki} - X^2/20 = S_{ki} x^2_{ki} - S_k X_k^2/5 + S_k X_k^2/5 - X^2/20 = K_e + K_K$$

Enaka zveza je tudi med stopinjami prostosti

$$m = 4.5-1 = 4.5-4+4-1 = 4(5-1)+4-1 = m_e + m_K$$

Ker sta s^2_K in s^2_e oceni iste variance, samo ocenjeni enkrat iz variabilnosti med grupami, drugič pa iz variabilnosti v grupah, izraz

$$F = s^2_K / s^2_e$$

ustreza zakonitosti F-distribucije z $m_1 = 4-1 = 3$ in $m_2 = 4(5-1) = 16$ stopinjami prostosti. To pomeni, da ta izraz s tveganjem 0,05 ni večji kakor

kritična vrednost $F_{0,05}(3,13) = 3,24$.

Zgornji postopek je izdelan pod predpostavko, da povprečna produktivnost za posamezne delavce ni različna $(M_1=M_2=M_3=M_4)$. Vprašanje je, kaj se zgodi s posameznimi ocenami variance, če je produktivnost dela med delavci različna. Če pogledamo S_e^2, spoznamo, da različna raven produktivnosti dela za posamezne delavce nanjo ne vpliva, ker je odvisna samo od variabilnosti v produktivnosti dela za posameznega delavca. Pač pa ima različna produktivnost dela za posameznega delavca bistven vpliv na S_K^2. Ker je ta ocena izračunana iz ocen grupnih sredin, je S_K^2 tem večji, čim večje so stvarne razlike v produktivnosti dela med delavci. Ker je S_K^2 v izrazu S_K^2/S_e^2 v števcu, je izračunani F tem večji, čim večje so razlike v produktivnosti dela med delavci. To sklepanje moremo izkoristiti za preizkušanje hipoteze o razlikah med aritmetičnimi sredinami za več delavcev. Osnovna hipoteza je, da je produktivnost dela med delavci različna $(H_1: M_1 \neq M_2 \neq M_3 \neq M_4)$. Ustrezna ničelna hipoteza je, da ni razlik v produktivnosti dela med delavci $(H_0: M_1=M_2=M_3=M_4)$. Po nakazanem postopku izračunamo F. Če je izračunani F večji kot ustrezna kritična vrednost $F_{0,05}(3,13) = 3,30$, vzamemo, da so med aritmetičnimi sredinami v produktivnosti dela med delavci značilne razlike. Z napako prve vrste $\alpha = 0,05$ se v tem primeru izkaže, da je produktivnost med delavci različna.

17.32 Vrednost postopka analize variance, ki smo ga nakazali za primer o produktivnosti dela, je v tem, da moremo preizkusiti značilnost razlik med aritmetičnimi sredinami za vse grupne sredine hkrati, medtem ko moremo s preizkusom o značilnosti razlik med povprečjema dveh populacij po odstavku 17.21 primerjati le dvoje povprečij. Pogoj za uporabo analize variance je v tem, da je izdelana pod predpostavko, da je variabilnost znotraj grup za vse grupe enaka in da je populacija normalno distribuirana.

Z analizo variance moremo reševati najrazličnejše probleme. Preizkušati moremo razlike v porabi za več socialnih skupin hkrati, kontrolirati enakomernost dela za skupine strojev, proučevati razlike v donosih na parcelah, ki so bile obravnavane z različnimi agrotehničnimi postopki itd.

17.33 Postopek analize variance enega faktorja v splošnejšem primeru, da je število enot v grupah različno, moremo nakazati v tehle točkah:

a) Iz vsake izmed k grup, katerih razlike proučujemo, izberemo vzorec z n_k enotami in zanje poiščemo vrednosti znaka x_{ki}, ki ga proučujemo. Indeks k pri x pomeni grupo, i pa tekočo številko enote znotraj grupe.

b) Za vsako grupo izračunamo iz individualnih vrednosti grupne vsote $Y_k = \sum_i x_{ki}$, iz teh pa skupno vsoto $Y = \sum_k Y_k$

c) Izračunamo vsoto kvadratov individualnih vrednosti $SSx_{ki}^2 = Q_{KI}$, iz grupnih vsot izraz $SY_k^2/n_k = Q_K$, iz skupne vsote pa izraz $Y^2/n = Q_S$ $(n = \Sigma n_k)$

d) Iz teh izrazov obračunamo analizo variance po standardni shemi

Tabela 17.6 Shema obračuna analize variance

Vir variacije	Vsota kvadratov odklonov K	Stopinje prostosti m	$s^2 = K/m$	F
med grupami	$Q_K - Q = K_K$	$k - 1$	$K_K/(k-1) = s_K^2$	$s_K^2/s_e^2 = F$
znotraj grup	$Q_{KI} - Q_K = K_I$	$n - k$	$K_I/(n-k) = s_e^2$	1
Skupaj	$Q_{KI} - Q = K_{KI}$	$n - 1$		

Razlike med grupnimi sredinami so značilne na stopnji P, če je izračunani F večji kot kritična vrednost $F_p(m_1=k-1; m_2=n-k)$.

17.34 Za platišča za kamione, ki jih proizvaja ena izmed tovarn v Sloveniji, moramo preizkusiti ali je kvaliteta platišč odvisna od različnih žarilnih temperatur. Za ta namen smo vzeli za vsako žarilno temperaturo po šest platišč in vsako izmed teh obdelovali pri različnih žarilnih temperaturah: 450°, 550° in 650°. Karakteristika, ki je izraz trpežnosti platišč, je število pulzacij do uničenja. Osnovne podatke imamo v tabeli 17.7.

Tabela 17.7 Osnovni podatki preizkave, ce je trpežnost platišč odvisna od žarilne temperature (x = število pulzacij)

Žarilna temperatura	x_{ki}						Y_k
450°	335	394	314	282	211	253	1789
550°	906	334	401	560	390	513	3404
650°	253	336	264	311	133	247	1544
							6737 = Y

$n_k = \bar{n} = 3, \; k = 3, \; n = 18$

$$Q_{KI} = \underset{ki}{SS} x_{ki}^2 = 335^2 + 394^2 + \ldots\ldots + 247^2 = 3088593$$

$$Q_K = \underset{k}{S} Y_k^2/n_k = \frac{1789^2}{3} + \frac{3404^2}{3} + \frac{1544^2}{3} = 2861945$$

$$Q = Y^2/n = 6737^2/18 = 2521509$$

Tabela 17.7 Analiza variance kvaliteto platišč od žarilne temperature

Vir variacije	SK	m	s^2	F
Žarilna temperatura	2861945-2521509 = 340436 = K_K	3-1 = 2	s_k^2 = 170218	11,3
Slučajni vplivi	3088583-2861945 = 226648 = K_I	18-3 = 15	s_e^2 = 15110	1
Skupno	567084	17		

Razlike so zelo značilne, ker je izračunani $F = 11,3$ večji kakor $F_{0,01}(2,15) = 3,83$. To pomeni da žarilna temperatura vpliva na kakovost platišč.

STATISTIČNA KONTROLA KVALITETE

17.35 Preizkušanje hipotez po načelih, navedenih v prejšnjih odstavkih, uporabljamo v najrazličnejših problemih. Posebna uporaba teh metod v praksi pa je statistična kontrola kvalitete, ki se je razvila v samostojno disciplino in operativno sredstvo kontrole predvsem v masovni industrijski proizvodnji.

Vzorčna kontrola je izpodrinila stoodstotno kontrolo iz več vzrokov. Dolgotrajnost, stroški, monotonost, ki vpliva na kvaliteto kontrole, poškodbe na kontroliranih artiklih, vse to so slabosti stoodstotne kontrole. Vseh zgornjih slabosti vzorčna kontrola nima, v primerih ko je kontrola zvezana z uničenjem produkta pa je vzorčna kontrola edino možna. Vzorčna kontrola je torej hitra, ekonomična in kvalitetna.

Statistično kontrolo proizvodnje delimo v dve večji samostojni grupi: *kontrolo prevzema* in *kontrolo proizvodnega procesa*. Čeprav imata obe isto osnovo-preizkušanje hipotez, so tako tehnični prijemi kot uporaba pri obeh zelo različni.

Statistična kontrola prevzema

17.36 Statistična kontrola prevzema je direktna uporaba statističnih metod preskušanja hipotez, brez posebnih sprememb. Pri tem običajno preskušamo hipoteze o aritmetični sredini določenih karakteristik proizvodnje M, o variabilnosti teh karakteristik σ in hipoteze o strukturi proizvodnje D (po kvaliteti, uporabnosti itd.).

Te vrste kontrole vršimo v različnih fazah proizvodnje: pri prevzemu surovin, prehodu polizdelkov iz oddelka v oddelek v istem podjetju in pri kontroli gotovega blaga.

Ker običajno kontroliramo, ali proizvodnja ustreza pogojem, ki jih stavlja kupec oziroma potrošnik, imenujemo pri kontroli kvalitete napako prve vrste, ki je v tem, da ustrezno proizvodnjo zavrnemo, *tveganje proizvajalca*, napako druge vrste, ki je v tem, da sprejmemo proizvodnjo, ki ne ustreza določenim pogojem pa *tveganje kupca*.

Kontrolo prevzema uporabljamo pri kontroli množične proizvodnje v tejle obliki. Celotno proizvodnjo razdelimo glede na pogoje proizvodnje (dnevna proizvodnja; proizvodnja, proizvedena na posameznih strojih itd.) v skupine artiklov. Po vnaprej določenem planu kontrole iz posamezne skupine izberemo slučajnostno določeno število artiklov. Glede na kvaliteto teh artiklov posamezno skupino sprejmemo ali zavrnemo. Zavrnjene skupine prekontroliramo v celoti, in v njih defektne artikle nadomestimo z dobrimi. Tako stvaren odstotek izmeta zmanjšamo, ker so defektni artikli v zavrnjenih skupinah nadomeščeni z dobrimi. Posamezni plani statistične kontrole prevzema imajo za osnovo različne količine. Ena izmed najpogostejših je *povprečen odstotek izmeta po opravljeni kontroli*. Ta pa je odvisna od kvalitete proizvodov pred kontrolo. Če je kvaliteta pred kontrolo dobra, je odstotek izmeta po kontroli tudi nizek. Enako je odstotek izmeta nizek, če je kvaliteta proizvodnje pred kontrolo zelo slaba, ker v tem primeru zavrnemo razmeroma veliko število skupin, in v njih nadomestimo slabe artikle z dobrimi. Pri neki določeni kvaliteti proizvodnje pred kontrolo pa je odstotek izmeta po kontroli maksimalen. Pri vzorčnih planih je zato razen povprečnega odstotka izmeta po kontroli pomemben tudi *maksimalen povprečen odstotek izmeta*, ki ga dobimo pri kontroli proizvodnje v kateri odstotek izmeta srednje velik.

17.37 Enojni, dvojni, trojni, sekvencialni plan. Pri dani velikosti kontroliranih skupin in povprečnemu odstotku izmeta po opravljeni kontroli moremo najti, pri zadostni velikosti vzorca neko celo število c_1 tako, da skupino sprejmemo, če je število defektnih artiklov v vzorcu manjše ali enako c_1 in zavrnemo, če je število slabih artiklov v vzorcu večje kot c_1. Tak plan imenujemo *enojni vzorčni plan*.

Pri velikih razlikah med stvarno in hipotetično kvaliteto odkrijemo razlike tudi z manjšimi vzorci. Zato se včasih obnese *dvojni plan*. Po tem planu v prvi fazi izberemo vzorec n_1 enot. Če je število slabih artiklov v tem vzorcu manjše ali enako c_1, skupino sprejmemo. Če je število slabih artiklov večje ali enako c_2 skupino zavrnemo. Če pa je število slabih artiklov večje kot c_1 in manjše kot c_2, izberemo dopolnilni vzorec z n_2 enotami. Ta vzorec je osnova za zavrnitev ali sprejem podobno kot pri enostavnem planu. Če število slabih artiklov v skupnem vzorcu z n_1+n_2 enotami enako ali manjše kot c_1' skupino sprejmemo, v nasprotnem primeru pa jo zavrnemo. Če pričakujemo da so med skupinami velike razlike, je ta plan bolj ekonomičen kot enojni, ker slabe skupine navadno izločimo že pri prvem vzorcu.

Podobno moremo sestaviti *trojni plan*, pri katerem pridemo pri nekaterih vzorcih do dokončne odločitve šele pri tretjem vzorcu. Pri tem planu vzorec n_1 enot povečamo na n_1+n_2 enot, če ne dosežemo odločitve pri prvem vzorcu, na $n_1+n_2+n_3$ enot pa v primeru, če tudi z drugim vzorcem ne moremo skupine niti zavreči niti sprejeti.

Razširitev te ideje je *sekvencialni plan*, pri katerem večamo preskusni vzorec za en ali za majhno število artiklov. Pri sekvencialnem planu dodajamo nove preskusne artikle toliko časa da pridemo do odločitve, da skupino sprejmemo ali zavrnemo. Dokler pa je število slabih artiklov tolikšno, da je situacija nedoločena, nadaljujejo z večanjem vzorca. Po določenem številu postopkov pridemo do cilja, in sicer tem prej, čim večje so razlike med stvarnim in hipotetičnim odstotkom. Sekvencialni plan je bolj zamotan kot navaden ali dvojen, ima pa to dobro lastnost, da dovede do rezultata v povprečju z najmanjšim številom preskusov. Zato ga uporabljamo predvsem za kontrolo proizvodov, za katere je preskus drag.

17.38 Kot primer vzemimo tabelo iz knjige: Freeman, Friedman, Mostelier, Wallis: Sampling Inspection. Iz priročnika je vzeta tabela za kontrolo skupin s 500-800 enotami za primer da pričakujemo povprečni odstotek defektnih artiklov po kontroli od 1,2 % do 2,2 %, zgornjo limito povprečja defektnih artiklov po kontroli pa 2,5 % do 3,5 %.

Tabela 17.8 Plan kontrole skupin s 500 do 800 enotami s povprečnim odstotkom defektnih artiklov po kontroli 1,2% do 2,2 % in zgornjo limito povprečja defektnih artiklov 2,5 % do 3,5 %

Tip vzorca	Vzorec	Velikost vzorca		c_1	c_2
		posamezнega	kumulativnega		
Enojni	prvi	40	40	2	3
Dvojni	prvi	25	25	1	4
	drugi	50	75	3	4
Sekvencialni	prvi	10	10	+	2
	drugi	10	20	0	3
	tretji	10	30	1	3
	četrti	10	40	1	4
	peti	10	50	2	4
	šesti	10	60	2	4
	sedmi	10	70	4	5

Skupino sprejmemo če je število defektnih artiklov v vzorcu $x < c_1$ in zavrnemo če je $x > c_2$. Vzorec nadaljujemo, če je $c_1 < x < c_2$.
+ pomeni: skupine ne moremo sprejeti pri prvem vzorcu.

Če n.pr. iz skupine S 300 enotami po dvojnem planu izberemo $n_1 = 25$ artiklov in med njimi najdemo 3 defektne, moramo vzorec glede na prejšnjo tabelo povečati za $n_2 = 50$; ker je $1 < 3 < 4$! Če najdemo v skupnem vzorcu $n_1+n_2=75$ enot 7 defektnih artiklov, skupino glede na prejšnjo tabelo zavrnemo.

Podobno je pri sekvencionalnem planu. Če iz skupine S $N=300$ enotami izberemo najprej 10 artiklov in izmed teh ne najdemo nobenega defektnega, moramo vzeti naslednjih 10 enot. Če v skupnem vzorcu z 20 enotami dobimo dva defektna artikla, moramo postopek ponoviti in vzorec povečati še za 10 enot, ker je $0 < 2 < 3$. Če v skupnem vzorcu s 30 enotami dobimo 3 defektne artikle pa skupino po tretjem vzorcu zavrnemo, ker je $3 \geq 3$.

Podobne tabele kot je tabela 17.8 so sestavljene za različne kombinacije povprečnega odstotka defektnih artiklov po kontroli in za različne velikosti skupin, tako da moramo najti ustrezno tabelo za vsako kombinacijo, ki je praktično pomembna.

Statistična kontrola proizvodnega procesa

17.39 S kontrolo gotovega blaga, kakor smo jo nakazali v prejšnjem odstavku, ugotavljamo, kontroliramo in popravljamo kvaliteto proizvodov po končanem proizvodnem procesu. Veliko efektnejša pa je kontrola, ki jo izvajamo v teku samega proizvodnega procesa. Veliko bolje je, če vzroke slabe proizvodnje sproti odstranjujemo, ne pa da po končani proizvodnji ugotovimo, da smo delali slabo.

Artikli, ki jih proizvajamo na določenem stroju niso med seboj popolnoma enaki. Zaradi drobnih *vzrokov* se med seboj, čeprav nepomembno, razlikujejo. Nemogoče, pa tudi nepotrebno, je te male vzroke razlik med artikli odstranjevati, če so za uporabnost artiklov nepomembni. Te drobne vzroke zaružujemo v skupino *slučajnih vplivov*, ki so pogojeni s kvaliteto stroja, ki artikle proizvaja. Ti slučajni odkloni so večji pri nenatančnih in iztrošenih strojih in manjši pri natančnih in novih strojih.

Razen slučajnih vplivov, s katerimi v dani proizvodnji računamo in jih ne skušamo odpravljati, ker za to ni potrebe, pa morejo v toku proizvodnega procesa nastopiti tudi *bistveni vzroki*, zaradi katerih se proizvedeni artikli spremene. Ti bistveni vzroki imajo za rezultat, da se proizvodnja polagoma ali nenadoma spremeni tako, da ni uporabna.

Naloga statistične kontrole proizvodnega procesa je v tem, da v teku proizvodnega procesa odkrije, kdaj nastopijo bistvene spremembe v proizvodnji zaradi teh vzrokov. Statistična kontrola nas torej opozori kdaj se je na samem stroju izvršila sprememba, ki ima za rezultat bistvene razlike v proizvodnji.

17.40 Kontrolna karta za raven - $\bar{x}$ karta. Tehnična osnova statistične kontrole proizvodnega procesa so kontrolne karte. Kontrolna kar-

ta je grafičen pripomoček pri kontroli, ki jasno, preprosto in učinkovito pokaže, kdaj se je v proizvodnem procesu izvršila bistvena sprememba. Teoretična osnova kontrolne karte pa je preskušanje hipotez. Tehnika pa je tako poenostavljena, da more kontrolne karte uporabljati tudi oseba brez posebnega statističnega znanja.

Vzemimo da za določen stroj poznamo učinek slučajnih vplivov, ki je izražen s standardnim odklonom σ za določeno karakteristiko proizvoda. Idealen povprečen premer določenega profila, ki ga množično proizvajamo, naj bo $M_H = 12$ mm, standardni odklon, ki je izraz slučajnih vplivov, pa je za to karakteristiko na tem stroju $\sigma = 1$ mm. Če vzamemo, da iz tekoče proizvodnje izbiramo vzorec $n = 9$ enot, se po znanih stavkih povprečja teh vzorcev hipotetično distribuirajo normalno z $M_H = 12$ mm in $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1/\sqrt{9} = 1/3$ mm. Aritmetične sredine vzorcev po $n = 9$ enot iz take proizvodnje se torej z verjetnostjo 0,95 nahajajo v mejah $M_H \pm 2\sigma/\sqrt{n}$ (1,96 praktično zaokrožen na 2). Za naš primer so torej povprečja vzorcev z veliko verjetnostjo (0,95) v mejah $12 \pm 2.1/\sqrt{9} = 11,33$ mm do 12,37 mm.

Če poznamo te meje, moramo kontrolirati proizvodni proces tako, da od časa do časa izteremo iz tekoče proizvodnje vzorec po $n = 9$ artiklov, zanj izmerimo premere, iz njih izračunamo povprečje in pogledamo ali je izračunano povprečje znotraj meja 11,33 do 12,37 ali ne. Če pade vzorčno povprečje znotraj tega razmaka, so razlike neznačilne in nimamo osnove, da dvomimo da je stvarno povprečje proizvodnje različno od predpisanega. Zato pustimo, da proizvodni proces teče dalje, kljub eventualni napaki druge vrste. Če pa povprečje preskusnega vzorca pade v kritično območje, v razmak pod 11,33 mm ali nad 12,37, je to znak da so nastopili bistveni vzroki, zaradi katerih se stvarno povprečje značilno odklanja od predpisanega $M_H = 12$ mm. Proizvodnjo v tem primeru ustavimo in poiščemo vzrok spremembe, čeprav s približno verjetnostjo 0,05 tvegamo, da je proizvodnja v redu da je to slučajno. Vzorce v teku proizvodnega procesa od časa do časa (vsak dan: vsako izmeno: vsako uro itd.) ponavljamo. Časovni razmak med vzorci je odvisen od tega, kakšna je možnost, da nadstopijo nenadni bistveni vzroki in kakšna je gospodarska škoda, če so artikli slabi. Vzorčne rezultate vnašamo v kontrolno karto, ki je časovni grafikon, v katerem je abscisa čas, v ordinato pa nanašamo povprečja vzorcev. V kontrolni karti je vrisan pas oziroma meje značilnih razlik, tako da je takoj razvidno, če pade povprečje v področje, ko je treba proizvodnjo ustaviti.

Razen pasu odklonov dveh standardnih pogrešk običajno vrisujemo še pas treh standardnih pogrešk. Medtem ko je tveganje pri dveh standardnih pogreškah še precejšnje, moremo z zelo veliko verjetnostjo računati, da so nastopile v proizvodnji bistvene razlike, če povprečje vzorca pade izven meja treh standardnih pogrešk.

Če vzamemo, da so bili premeri devetih artiklov v prvem preskusnem vzorcu x: 11,0 12,5 13,0 12,8 11,0 11,5 12,2 12,4 13,1 je za ta vzorec povprečje $\bar{x} = 12,17$ mm. Ta rezultat je vrisan v kontrolni karti v sliki 17.7 s točko za prvi dan.

17.4 Kontrolna karta za variabilnost s - karta. V proizvodnem procesu se more zaradi bistvenih vzrokov spremeniti tudi variabilnost pro-

izvoanje n.pr. zaradi bistvene okvare na stroju.

Ne samo spremembe povprečij, temveč tudi spremembe v variabilnosti povzroče večji odstotek neuporabnih artiklov v proizvodnji. Čeprav iz objektivnih razlogov slučajnih variacij iz proizvodnje ne moremo odpraviti, je važno, da kontroliramo, da kdaj je variabilnost v mejah, ki so pogojene z normalnimi pogoji proizvodnje in kdaj je ta meja prekoračena. Variabilnost pogosto kontroliramo s standardnim odklonom v vzorcih S, zaradi česar kontrolno karto za variabilnost običajno imenujemo kar *s-karto*. Teoretična osnova S-karte je obrazec 13.19, s katerim glede na predvideno variabilnost σ, število enot v kontrolnem vzorcu n in stopnjo tveganja $\alpha = p$ določimo zgornjo mejo, katero standardni odklon S s tveganjem $\alpha = p$ ne prekorači, če je stvarna variabilnost enaka σ. Po tem obrazcu dobimo da je

$$s < \frac{\sqrt{\chi^2_p}\,\sigma}{\sqrt{n-1}}$$

V našem primeru, kjer je osnovna variabilnost $\sigma = 1$ in imamo v vzorcih po $n = 9$ enot dobimo zgornjo mejo za $\alpha = 0{,}05$ enako

$$s < \frac{\sqrt{15{,}51}\cdot 1}{\sqrt{9-1}} = 1{,}39 \text{ mm}$$

za $\alpha = 0{,}01$ pa

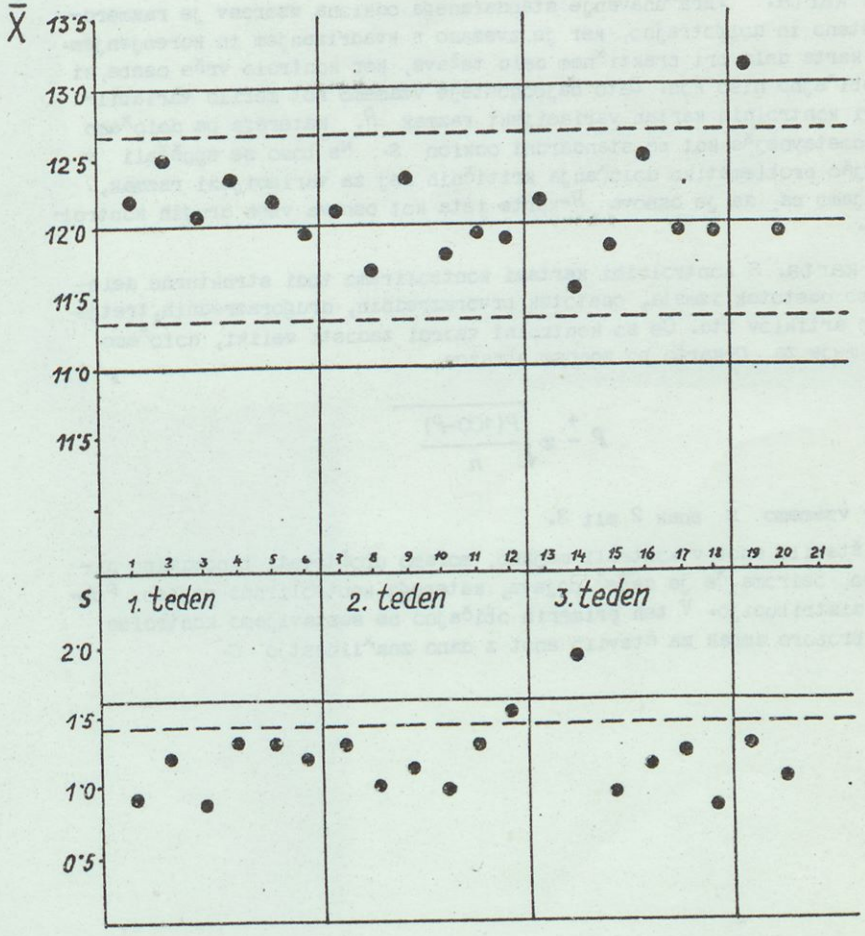
$$s < \frac{\sqrt{20{,}09}\cdot 1}{\sqrt{9-1}} = 1{,}58 \text{ mm}$$

Ker je $\chi^2_{0,05}(8) = 15{,}51$: $\chi^2_{0,01}(8) = 20{,}09$

Spodnje meje v S-karti so nepomembne, ker eventualna zmanjšana variabilnost ne pomeni poslabšanja proizvodnje. Zato je v teh primerih neumestno da ustavimo proizvodnjo.

V našem primeru smo za prvi dan iz individualnih podatkov izračunali, da je $s = 0{,}81$, za druge dneve pa je variiranje variabilnosti v vzorcih razvidno iz kontrolne S-karte v sliki.

17.42 Analiza **kombinirane** $\bar{x}$-karte in S-karte v sliki 17.7 je naslednja. Enajsti dni je proizvodnja tekla v redu tako glede na povprečje kot glede na variabilnost profilov. Dvanajsti dan pa so se pojavile motnje v variabilnosti in je bil S iz vzorca v kritičnem območju z $\alpha = 0{,}05$. Ker pa je bila variabilnost še vedno le v mejah značilnosti z $\alpha = 0{,}05$, proizvodnje nismo ustavili ampak smo računali, da je možno, da je ta S rezultat slučaja. Vendar se je šele dva dni pokazalo da je vzrok za povečano variabilnost bistven faktor, ker se je točka za S za štirinajsti dan



Slika 17.7 Kombinirana $\bar{x}$-s-karta

pojavila v območju značilnosti 0,01. Stroj je bil zato pregledan, ugotovljena je bila napaka in odstranjen vzrok, tako da v nadaljnjih dneh ni bilo več motenj v variabilnosti. Devetnajsti dan proizvodnje pa je poskočila

aritmetična sredina preskusnega vzorca v območje značilnosti $a = 0,01$. Zato smo morali delo stroja ponovno pregledati in ugotoviti napako. Dejstvo, ali kontrolna karta pokaže nepravilnost v ravni ali variabilnosti je tudi indikacija glede vrste napake.

17.48 **R**-karta. Izračunavanje standardnega odklona vzorcev je razmeroma zapleteno in dolgotrajno, ker je zvezano s kvadriranjem in korenjenjem. Zato ta karta dela pri praktičnem delu težave, ker kontrolo vrše osebe, ki nalogi običajno niso kos. Zato najpogosteje vzamemo kot merilo variabilnosti pri kontrolnih kartah variacijski razmak R, katerega pa določamo znatno enostavnejše kot pa standardni odklon S. Ne bomo se spuščali v podrobnejšo problematiko določanja kritičnih mej za variacijski razmak, pripominjamo pa, da je osnova R-karte ista kot osnova vseh drugih kontrolnih kart.

17.44 p-karta. S kontrolnimi kartami kontroliramo tudi strukturne deleže: kot so odstotek izmeta, odstotek prvorazrednih, drugorazrednih, tretjerazrednih artiklov ita. Če so kontrolni vzorci zadosti veliki, določamo kritične meje za p-karto po znanem obrazcu

$$P \pm z\sqrt{\frac{P(100-P)}{n}}$$

pri čemer vzamemo z enak 2 ali 3.

Če pa je število enot v kontroli majhno, moramo upoštevati binomialno distribucijo, oziroma, če je delež pojava, katerega kontroliramo majhen, Poissonovo distribucijo. V teh primerih običajno ne sestavljamo kontrolne karte za proporc ampak za število enot z dano značilnostjo C.

LITERATURA

Jugoslovanska:

Blejec M. Teorija statistike, (Ljubljana 1951

Blejec M. Statistične metode za psihologe, (Ljubljana 1959

Ećimović J. Osnovi statističke reprezentativne metode; (Beograd 1951

Obradović S. Sentić M. Osnovi statističke analize; (Beograd 1956

Serdar V. Udžbenik statistike 1957

Vogelnik D. Osnovi statistike; (Beograd 1950)

Vogelnik D. Blejec M. Statistični repertorij; (Ljubljana 1953

Zarković S. Statističke metode u industrijskim istraživanjima; (Beograd 1949

Zarković S. Kontrola kvalitete robe; (Beograd 1949

Zarković S. Kvalitativna kontrola proizvodnog procesa; (Beograd 1951.

Tuja:

Allen R.G.D. Statistics for Economists; (London 1949

Boldrini R. Principii di statistica metodologica; (Milano 1942

Croxton F.E. Cowden D.J. Applied General Statistics; (Englewood Cliffs N.J. 1955

Dixon W.J. Massey F.J. Introduction to Statistical analysis; (New York 1957

Nemčinov V.S. Selskohazjastvena statistika s osnovami obščei teorii; (Moskva 1945

Riggleman J.R. Frisbee I.N. Business Statistics; (New York 1951

Tintner G. Mathematics and Statistics for Economists; (New York 1954

Wallis W.A.R. Roberts H.W. Statistics- a New Approach; (New York 1953

Winkler W. Grundriss der Statistik; (Wien 1947

Žižek F. Grundriss der Statistik; (Berlin 1931

Yule G.U. Kendall M.G. An Introduction to the Theory of Statistics; (New York 1950; (v slovenščino prevedel Vogelnik D. 1955)

V S E B I N A

1. UVOD 5

Kaj je statistika ? 5
Področja, v katerih uporabljamo statistiko 6
Socialno-ekonomska statistika 7

2. PROUČEVANJE MNOŽIČNIH POJAVOV 9

Množični pojavi 9
Statistične enote 10
Statistični znaki 11
Statistične populacije 13
Statistični parametri 15
Značilnosti pri proučevanju množičnih pojavov 16
Etape statističnega proučevanja 20

3. STATISTIČNO OPAZOVANJE 23

Opredelitev predmeta opazovanja - statistične populacije 23
Vrste opazovanj 24
Popis 24
Tekoča registracija - Statistična poročila 25
Metode delnega opazovanja 26
Viri podatkov 28
Načini posrednega opazovanja 29
Znaki opazovanja 31
Sredstva opazovanja 32
Kraj opazovanja 40
Organi statističnega opazovanja 41
Napake in kontrola statističnega opazovanja 42

4. UREJEVANJE STATISTIČNEGA GRADIVA 46

Grupiranje vrednosti znakov 47
Šifriranje osnovnega gradiva 56
Osnovna obdelava statističnega gradiva 57
Ročna obdelava 61
Strojna obdelava 64

5. PRIKAZOVANJE STATISTIČNIH PODATKOV 68

Statistične vrste 68
Statistične tabele 72
Vrste tabel po vlogi v statističnem opazovanju 76
Tehnična načela za sestavljanje tabel 78

Grafično prikazovanje 80
Elementi grafičnega prikazovanja 80
Skale - lestvice 82
Mreže 85
Črtanje in šrafiranje 85
Vrste grafikonov 86
Stolpci 87
Linijski grafikoni 89
Figure 95
Kartogrami 97

6. RELATIVNA ŠTEVILA 104
Strukture ali razčlenitvena števila 105
Grafično prikazovanje struktur 108
Statistični koeficienti in gostote 120
Izračunavanje statističnih koeficientov 120
Grafično prikazovanje statističnih koeficientov 128
Enostavni indeksi 133
Stvarni in krajevni indeksi 134
Časovni indeksi 135

7. FREKVENČNE DISTRIBUCIJE 141
Sestavljanje frekvenčnih distribucij 141
Frekvenčne distribucije z neenakimi razredi 144
Grafično prikazovanje frekvenčnih distribucij 147
Oblike frekvenčnih distribucij 149
Kumulativna frekvenčna distribucija 153
Lorenzov grafikon 155

8. KVANTILI 158
Ranžirna vrsta. Rang 158
Kvantilni rang 159
Kvantili 160
Izračunavanje kvantilov iz negrupiranih podatkov 160
Izračunavanje kvantilov iz grupiranih podatkov 163

9. SREDNJE VREDNOSTI 168
Vrste srednjih vrednosti 168
Mediana 168
Lastnosti mediane 168
Modus 168
Izračunavanje modusa iz frekvenčnih distribucij 170
Lastnosti modusa 172

Aritmetična sredina 173
Lastnosti aritmetične sredine 173
Izračunavanje aritmetične sredine iz negrupiranih podatkov 175
Izračunavanje aritmetične sredine iz frekvenčnih distribucij 176
Direktna metoda 176
Pomožni znak u 177
Metoda kumulativ 178
Modificirana aritmetična sredina 180
Aritmetična sredina aritmetičnih sredin 180
Harmonična sredina 182
Povprečja iz relativnih števil 183
Standardnost sumarnega relativnega števila od strukture 185
Standardizirani pokazatelji 187
Geometrijska sredina 188
Odnosi med različnimi vrstami srednjih vrednosti 191

10. MERE VARIACIJE ASIMETRIJE IN SPLOŠČENOSTI 193
Mere variacije 193
Vrste mer variacije 193
Variacijski razmak 193
Kvartilni odklon 194
Povprečen absolutni odklon 196
Varianca. Standardni odklon 197
Izračunavanje iz negrupiranih podatkov 198
Izračunavanje iz grupiranih podatkov 200
Skupna varianca 203
Zveza standardnega odklona z normalno distribucijo 205
Razmerje med Q, AD in SD za normalno distribucijo 206
Relativne mere variacije. 206
Mere asimetrije in sploščenosti 208
Mere asimetrije 209
Mera sploščenosti 210

11. INDEKSI 211
Povprečni indeks cen 211
Agregatni indeksi cen 212
Izbira ponderov pri izračunavanju agregatnega indeksa cen 213
Verižni indeks cen 218
Nadomeščanje artiklov in spremembe ponderacije 219
Reprezentativni indeksi cen 220
Agregatni indeksi količin 224
Testi o zamenljivosti časa in faktorjev 226

Drugi agregatni indeksi 227

12. ČASOVNE VRSTE 229

Oblike časovnih vrst 229
Izvedene časovne vrste 230
Kumulativna časovna vrsta 230
Vrsta sredin 231
Vrsta drsečih vsot 232
Časovna vrsta drsečih sredin 234
Grafično prikazovanje časovnih vrst 238
Logaritemski grafikoni 241
Polarni grafikon 244
Z-diagram 243
Brunsmanov grafikon 248
Ganttov grafikon 249
Analiza časovnih vrst 251
Primerljivost podatkov v časovni vrsti 252
Elementarni pokazatelji dinamike 255
Komponente gibanja v časovnih vrstah 258
Osnovni modeli časovnih vrst 260
Vloga povprečij za analizo časovnih vrst 261
Trend 262
Prostoročna metoda 265
Metoda drsečih sredin 263
Metoda sredin med najnižjimi in najvišjimi točkami 268
Metoda izbranih točk 269
Metoda delnih sredin 270
Metoda delnih vsot 272
Metoda najmanjših kvadratov 274
Sezonske in periodične variacije 280
Metoda vsot 282
Metoda kvocientov na trend 284
Metoda kvocientov na vrsto drsečih sredin 285
Metoda verižnih kvocientov 287
Metoda grafičnega približka 291
Ciklična nihanja 292

* 13. KORELACIJA 295

Funkcionalne odvisnosti 295
Korelacijske odvisnosti 297
Prikazovanje korelacijskih odvisnosti 297
Regresijska krivulja 303
Metode določanja regresijskih krivulj 304
Indeks korelacije 309
Standardna napaka ocene 311

Linearna korelacija 312
Izračunavanje pokazateljev linearne korelacije 314
Negrupirani podatki 314
Izračunavanje pokazateljev linearne korelacije iz grupiranih podatkov 320
Krivuljčna korelacija 323
Analitična metoda za določanje regresijskih krivulj 324
Transformacija znakov 325
Izračunavanje krivuljčne korelacije s polinomi 332
Korelacijsko razmerje 335
Korelacija ranga 340
Asociacija in kontingenca 343
Asociacija 343
Kontingenca 347
Parcialna korelacija 350
Multipla korelacija 351
Linearna multipla korelacija 352

14. NORMALNA DISTRIBUCIJA 355

Pomen proučevanja teoretičnih distribucij 355
Normalna distribucija 353
Opis normalne distribucije 353
Standardiziran z odklon 330
Standardizirana normalna distribucija 331
Prilagoditev normalne distribucije stvarni frekvenčni distribuciji 335
Verjetnostni grafikoni 337
Normalna distribucija kot verjetnostna distribucija 372
Pojem tveganja 374

15. VZORČENJE - VELIKI VZORCI 377

Osnove vzorčenja 377
Prednosti in pomanjkljivosti vzorčenja 378
Uporaba vzorčenja pri statističnem opazovanju 379
Osnovna populacija. Vzorec. Populacija vseh vzorcev 380
Zakonitosti v populaciji vseh vzorcev 383
Enostavno vzorčenje 383
Ocenjevanje aritmetične sredine 383
Ocenjevanje drugih parametrov z enostavnimi vzorci 388
Ocene in standardne pogreške za nekaj najvažnejših parametrov 389
Popravek za standardne pogreške za vzorce brez ponavljanja 390
Tehnika enostavnega slučajnega izbora 391
Primeri enostavnega slučajnega vzorčenja 393
Ocenjevanje razlik dveh parametrov 400
Določanje števila enot v enostavnem slučajnem vzorcu 402

Zveza med ocenami za M, X, P in N_a 404
Ocenjevanje po metodi razmerij 405
Vzorčenja z omejitvami 407
Stratificirano vzorčenje 408
Vzorčenje v skupinicah 413
Vzorčenje v dveh ali več stopnjah 414
Vzorčenje v dveh ali več fazah 415
Sistematično vzorčenje 417
Kvotno vzorčenje 418
Zaključek 421

16. VZORČENJE - MALI VZORCI 423

Teoretične distribucije za male vzorce 423
Binomialna distribucija 423
Hipergeometrična distribucija 424
Poissonova distribucija 425
Stopinje prostosti 426
Standardizirana normalna distribucija 427
t-distribucija 428
χ^2-distribucija 429
F-distribucija 430
Distribucije vzorčnih izrazov 430
Meje zaupanja za ocene z malimi vzorci 433

17. VZORČENJE - PREIZKUŠANJE HIPOTEZE 441

Teoretične osnove preizkušanja hipotez 441
Preizkušanje hipotez z velikimi vzorci 450
Preiskušanje hipotez z malimi vzorci 452
Analiza variance 462
Statistična kontrola kvalitete 467
Statistična kontrola prevzema 467
Statistična kontrola proizvodnega procesa 470

TABELE

Tabela A. Kvadrati števil 1-1000 475
Tabela B. Normalna distribucija 477
Tabela C. t-distribucija 479
Tabela D. χ^2-distribucija 480
Tabela F. F-distribucija 481
Tabela F. Pretvarjanje korelacijskih koeficientov r v koeficiente Z 485
Tabela G. Konstante za izračunavanje parametrov paraboličnega trenda prve, druge in tretje stopnje 486
Tabela H. Slučajna števila 487

DODATEK, ZBIRKA OZNAK IN OBRAZCEV

Dr. Marijan Blejec

redni profesor na Ekonomski fakulteti Univerze
v Ljubljani

STATISTIČNE METODE ZA EKONOMISTE

Izdala Ekonomska fakulteta Univerze v Ljubljani
Založilo Tajništvo Univerze v Ljubljani
Tiskala Univerzitetna tiskarna v Ljubljani
Slike prerisal Zoran Didek
Matrice tipkali: Z. Vogrič in F. Celarc

UNI
Univerzitetna knjižnica Maribor
R
67083/1
098613151
COBISS