Narodna in univerzitetna knjižnica
v Ljubljani

II 270128

Narodna in univerzitetna knjižnica
v Ljubljani

II 270128

VISOKA EKONOMSKO KOMERCIALNA ŠOLA
MARIBOR

Dr. MARIJAN BLEJEC

GOSPODARSKA STATISTIKA

MARIBOR, 1974

VISOKA EKONOMSKO KOMERCIALNA ŠOLA
M A R I B O R

dr. MARIJAN BLEJEC

GOSPODARSKA STATISTIKA

Maribor 1974

T 270128

II 270128

Izdajo tega učbenika je odobril Pedagoško znanstveni svet šole na svoji 8. redni seji dne 11. maja 1973.

Učbenik je prirejen vsebini predmeta na VEKŠ. Nova so poglavja od 1. do 6. in 14. poglavje. Poglavja od 6. do 13. so ponatisnjena nespremenjeno po učbeniku dr. M. Blejca: Statistične metode za ekonomiste, II. predelana in popravljena izdaja, Ljubljana 1961 z upoštevanjem popravkov v izdaji iz leta 1971.

Narodna in univerzitetna knjižnica LJUBLJANA

O 1763/1976

V S E B I N A

1. UVOD 1

Kaj je statistika? 1
Socialno-ekonomska statistika 1

2. PROUČEVANJE MNOŽIČNIH POJAVOV 3

Množični pojavi 3
Statistične enote 4
Statistični znaki 5
Statistične populacije 7
Statistični parametri 9

3. STATISTIČNO PROUČEVANJE 10

Vrste statističnih opazovanj 10

4. UREJEVANJE STATISTIČNEGA GRADIVA 13

Razvrščanje vrednosti znakov 13

5. PRIKAZOVANJE STATISTIČNIH PODATKOV 21

Statistične vrste 21
Statistične tabele 25

6. RELATIVNA ŠTEVILA 26

Strukture ali razčlenitvena števila 27
Grafično prikazovanje struktur 30
Statistični koeficienti in gostote 43
Izračunavanje statističnih koeficientov 43
Grafično prikazovanje statističnih koeficientov 52
Enostavni indeksi 57
Stvarni in krajevni indeksi 58
Časovni indeksi 60

7. FREKVENČNE DISTRIBUCIJE 66
Sestavljanje frekvenčne distribucije 66
Frekvenčne distribucije z neenakimi razredi 69
Grafično prikazovanje frekvenčnih distribucij 73
Kumulativna frekvenčna distribucija 79
Lorenzov grafikon 81

8. KVANTILI 84

Ranžirna vrsta. Rang 84
Kvantilni rang 85
Izračunavanje kvantilov iz negrupiranih podatkov 87
Izračunavanje kvantilnih rangov in kvantilov iz frekvenčnih distribucij 90

9. SREDNJE VREDNOSTI 95

Vrste srednjih vrednosti 95

Mediana 96
Lastnosti mediane 96
Modus 97
Izračunavanje modusa iz frekvenčnih distribucij 97
Lastnosti modusa 100
Aritmetična sredina 100
Lastnosti aritmetične sredine 101
Izračunavanje aritmetične sredine iz negrupiranih podatkov 103
Izračunavanje aritmetične sredine iz frekvenčnih distribucij 103
Direktna metoda 103
Pomožni znak u 105
Metoda kumulativ 107
Modificirana aritmetična sredina 108
Aritmetična sredina aritmetičnih sredin 108
Harmonična sredina 110
Povprečja iz relativnih števil 112
Odvisnost sumarnega relativnega števila od strukture 114
Standardizirani pokazatelji 115
Geometrijska sredina 117
Odnosi med različnimi vrstami srednjih vrednosti 121

10. MERE VARIACIJE ASIMETRIJE IN SPLOŠČENOSTI 123

Mere variacije 123
Vrste mer variacije 123
Variacijski razmak 124
Kvartilni odklon 124
Povprečen absolutni odklon 126
Varianca. Standardni odklon 128
Izračunavanje iz negrupiranih podatkov 128
Izračunavanje iz grupiranih podatkov 130
Skupna varianca 134
Zveza standardnega odklona z normalno distribucijo 136
Razmerje med Q, AD in SD za normalno distribucijo 137
Relativne mere variacije 137
Mere asimetrije in sploščenosti 140
Mere asimetrije 140
Mera sploščenosti 142

11. INDEKSI 143
Povprečni indeks cen 143
Agregatni indeksi cen 144
Izbira ponderov pri izračunavanju agregatnega indeksa cen 145
Verižni indeks cen 149
Nadomeščanje artiklov in spremembe v ponderaciji 152
Reprezentativni indeksi cen 154
Agregatni indeksi količin 158
Testi o agregatnih indeksih 160
Drugi agregatni indeksi 161

12. ČASOVNE VRSTE 164

Oblike časovnih vrst 164
Izvedene časovne vrste 165
Kumulativna časovna vrsta 165
Vrsta sredin 166
Vrsta drsečih vsot 168
Časovna vrsta drsečih sredin 170
Grafično prikazovanje časovnih vrst 174
Logaritemski grafikoni 177
Polarni grafikon 180
Z-diagram 182
Brunsmanov grafikon 184
Ganttov grafikon 186
Analiza časovnih vrst 189
Primerljivost podatkov v časovni vrsti 189
Elementarni pokazatelji dinamike 193
Komponente gibanja v časovnih vrstah 196
Osnovni modeli časovnih vrst 198
Vloga povprečij za analizo časovnih vrst 199
Trend 200
Prostoročna metoda 204
Metoda drsečih sredin 206
Metoda sredin med najnižjimi in najvišjimi točkami 206
Metoda izbranih točk 208
Metoda delnih sredin 210
Metoda delnih vsot 212
Metoda najmanjših kvadratov 214
Sezonske in periodične variacije 221
Metoda vsot 222
Metoda kvocientov na trend 225
Metoda kvocientov na vrsto drsečih sredin 225
Metoda verižnih kvocientov 228
Metoda grafičnega približka 232
Ciklična nihanja 234

13. KORELACIJA 238

Funkcijske odvisnosti 236
Korelacijske odvisnosti 238
Prikazovanje korelacijskih odvisnosti 238
Regresijska krivulja 243
Metode določanja regresijskih krivulj 245
Indeks korelacije 251
Standardna napaka ocene 253
Linearna korelacija 254
Izračunavanje pokazateljev linearne korelacije 256
Negrupirani podatki 256
Izračunavanje pokazateljev linearne korelacije iz grupiranih podatkov 262
Krivuljčna korelacija 267
Analitična metoda za določanje regresijskih krivulj 267
Transformacija znakov 268
Izračunavanje krivuljčne korelacije s polinomi 276

Korelacijsko razmerje 280
Korelacija ranga 285
Asociacija in kontingenca 288
Asociacija 288
Kontingenca 292
Parcialna korelacija 295
Multipla korelacija 296
Linearna multipla korelacija 297

14. VZORČENJE 300

Osnove vzorčenja 300
Prednosti in pomanjkljivosti metode vzorčenja 301
Uporaba vzorčenja pri statističnem opazovanju 302
Opis normalne distribucije 303
Standardizirana normalna distribucija 305
Osnovna populacija. Vzorec. Populacija vseh vzorcev 310
Zakonitosti v populaciji vseh vzorcev 313
Enostavno slučajnostno vzorčenje 314
Tehnika enostavnega slučajnega izbora 319

15. PREIZKUŠANJE DOMNEV 321
Preizkušanje domnev 321
Osnovna in ničelna domneva 321
Enostavne in sestavljene domneve 321
Preizkušanje enostavnih domnev 321
Napaka I. vrste in napaka II. vrste 323
Značilnost in neznačilnost razlik in vzorčnih izrazov 325
Stopnje značilnosti 325
Vzorčni izrazi 325

U V O D

KAJ JE STATISTIKA?

1.1 Statistiko opredelimo kot vedo, ki s številčnim proučevanjem množičnih pojavov, z metodami, ki so njej lastne, odkriva zakonitosti množičnega pojavljanja in podaja kakovostno analizo pojavov.

Iz te opredelitve spoznamo, da ima statistika svoje področje - množične pojave - in svojo metodo, da s kvantitativnim proučevanjem analizira kvalitativne odnose množičnih pojavov. Tako na primer količina proizvodnje, proizvedene v enoti časa, pokaže produktivnost dela, višina hektarskega pridelka uspeh agrotehnih 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 povezave z znanostjo, v katero spada proučevani pojav. Zato bi mogli imeti statistiko le za eno izmed metod za proučevanje pojavov. Vendar statistična metoda, razen v nekaterih posebnih primerih, ni bistveno vezana na določeno področje, marveč veljajo njene metode in zakonitosti neodvisno od predmeta za vse množične pojave.

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

SOCIALNO EKONOMSKA STATISTIKA

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

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

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

Splošne metode statističnega proučevanja 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, korelacija itd.), so druge take, da jih uporabljamo samo ali pretežno za proučevanje socialno-ekonomskih pojavov. Med temi so na primer analiza časovnih vrst, teorija indeksov in konjunkturna statistika.

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

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

Tako imamo posebej:

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

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

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

Analogno imamo nadalje gozdarsko statistiko, statistiko obrti, gradbeništva, prometa, trgovine, statistiko gostinstva, turizma, financ, statistiko cen, kulturno-prosvetno

statistiko, statistiko zdravstva, sodno statistiko itd.

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

DRUGO POGLAVJE

P R O U Č E V A N J E M N O Ž I Č N I H P O J A V O V

MNOŽIČNI POJAVI

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

Če analiziramo pojave, ki množično nastopajo, odkrijemo, da ti pojavi niso med seboj enaki, marveč se v svojih značilnostih razlikujejo.

Podjetja imajo različno število delavcev, različno mesečno proizvodnjo, različno porabo surovin, so iz različnih strok itd. Kupoprodaje se med seboj razlikujejo po času, prodajalcu, kupcu, ceni, količini, kakovosti prodanega blaga itd. Osebe so različnega spola, starosti, stanu, zaposlitve, imajo različno šolsko izobrazbo, mesečne prejemke, število otrok itd. Izdelki so različnih dimenzij, imajo različno kakovost, uporabnost itd., čeprav se zde na oko med seboj enaki.

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

Z 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 izdelkov, 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 samo nekatere predmet konkretnega proučevanja. Značilnosti, ki so v posameznem primeru predmet 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 delavcev, vsi izdelki niso enako uporabni itd.

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

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

Statistične enote

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

a) oseba (prebivalec, delavec, študent),
b) žival (konj, krava, prašič itd.),
c) stvar (avto, motor, izdelek),
d) pravna tvorba (zadruga, društvo),
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)
i) trenutek (v katerem opazujemo pojav).

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

Statistične enote so bodisi enostavne enote ali pa agregati - skupinice enot. Tako štejemo posameznega človeka za enostavno enoto. Družina, občina, društvo itd. pa so agregati, ker so sestavljeni iz enostavnih enot: članov družine, prebivalcev občine, članov društva itd. Združevanje enostavnih enot v enote višje stopnje - agregate, je za statistično proučevanje vsebinsko in tehnično izredno pomembno.

Statistični znaki

2.4 Statistični znaki dajo enotam vsebino. Znaki so tiste značilnosti statističnih enot, ki so predmet proučevanja. Statistične enote oziroma množični pojavi na splošno imajo namreč veliko najrazličnejših značilnosti. V konkretni raziskavi pa podobno kakor iz množice pojavov izločimo populacijo, iz množice vseh možnih značilnosti izločimo tiste, ki so važne za proučevanje in te imenujemo znake. Katere značilnosti so v posebnem primeru znaki, je odvisno od namena proučevanja.

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

Numerični so znaki, katerih vrednosti izražamo številčno. Po tem, katere vrednosti morejo zavzeti, delimo numerične znake v: a) nezvezne - diskontinuirne in b) zvezne - kontinuirne. Nezvezen 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 razmiku. Tako ne more imeti gospodinjstvo 3 1/2 članov, mati 1 1/4 otrok, podjetje 86 1/2 zaposlenih itd. Zvezni znaki so tisti numerični znaki, ki morejo teoretično imeti vsako izmed vrednosti v določenem razmiku. Zvezni znaki so na primer starost, teža izdelka, premer profila. (cele in dec. vrednosti)

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

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

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

2.9 Variiranje statističnih znakov. Značilno za znake je, 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 opredelitvijo variabilnosti. Vsaka oseba v dani populaciji namreč ne more biti različnega spola, če ima znak spol samo dve vrednosti: moški ali ženski.

Variabilnost statističnih znakov je ena izmed temeljnih lastnosti množičnih pojavov. Statistika se predvsem in v najrazličnejših oblikah ukvarja z variabilnostjo pojavov. Če množični pojavi ne bi bili variabilni, pojavov sploh ne bi proučevali s statističnimi metodami. Zato smo tudi navedli,

da ima statistični znak najmanj dve vrednosti, saj značilnost z eno samo vrednostjo ne more biti statistični znak, ker ne variira.

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

Statistične populacije

2.10 Opredelitev statistične populacije. Statistična populacija je množica vseh istovrstnih pojavov - statističnih enot, ki izpolnjujejo opredeljujoče pogoje, s katerimi je opredeljena. Statistično populacijo opredelimo tako, da navedemo, kakšne vrednosti znakov, ki so opredeljujoči pogoji, morajo imeti enote, da so enote populacije. V znakih, ki opredeljujejo populacijo, je torej variacija okrnjena.

Za populacijo ne velja samo množica vseh statističnih enot v gornjem smislu, marveč je populacija tudi množica podatkov za neki znak. Tako je populacija v tem smislu tudi množica vseh podatkov o starosti poročenih zaposlenih žena v SRS po stanju 31.12.1972.

2.11 Populacija mora biti nedvomno opredeljena: a) časovno, b) krajevno in c) stvarno.

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

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

Gornjo populacijo žensk smo opredelili s trenutkom. Populacije realnih enot imenujemo tudi momentne ali trenutne populacije. Momentnih populacij imamo veliko. Vse časovno enolično opredelimo z momentom, v katerem morajo posamezne enote populacije zadoščati vsem drugim pogojem, da so enote populacije.

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

2.12 Homogene in heterogene populacije. Po sorodnosti enot so populacije homogene - istorodne in heterogene - raznorodne. V homogenih populacijah enote med seboj niso preveč različne, medtem ko so heterogene populacije sestavljene iz raznorodnih enot. Homogenost populacije dosežemo, če jo opredelimo po čim več faktorialnih znakih. Delitev populacij na homogene in heterogene ni ostra in imamo bolj ali manj homogene in bolj ali manj heterogene populacije. Tako je populacija 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. Homogenost populacij je sicer pogojena z opredeljujočimi pogoji, ni pa istovetna. Opredelitev v faktorialnih znakih ima za posledico manjše variiranje oziroma večjo sorodnost ne samo v opredeljujočih temveč tudi v rezultativnih znakih, ki niso opredeljujoči. Krajevna opredelitev kmetijskih gospodarstev pogojuje homogeno populacijo glede na strukturo proizvodnje, način gospodarjenja ipd.

2.13 Delne populacije. Če v populaciji zaposlenih poročenih žena v Sloveniji konec leta 1972 dodamo pogoj, da je zaposlena kot delavka, iz prve populacije izločimo novo populacijo. To populacijo imenujemo delno populacijo prvotne, zato ker je vsaka enota druge populacije hkrati enota prve populacije.

Populacijo moremo po nekem znaku razdeliti v popoln sistem delnih populacij, če jo razdelimo v delne populacije po vseh vrednostih nekega znaka. Tako moremo populacijo vseh obrtnih obratov v Sloveniji konec leta 1972 razdeliti po strokah v popoln sistem delnih populacij. Delne populacije so bolj homogene kot osnovna populacija, ker se enote delnih populacij ujemajo še v dodatnem pogoju. Če populacijo razdelimo v popoln sistem delnih populacij po nadaljnjih faktorialnih znakih, heterogeno populacijo razdelimo v sistem homogenih populacij. To metodo večkrat uporabljamo pri analizi množičnih pojavov.

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

da ima statistični znak najmanj dve vrednosti, saj značilnost z eno samo vrednostjo ne more biti statistični znak, ker ne variira.

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

Statistične populacije

2.10 Opredelitev statistične populacije. Statistična populacija je množica vseh istovrstnih pojavov - statističnih enot, ki izpolnjujejo opredeljujoče pogoje, s katerimi je opredeljena. Statistično populacijo opredelimo tako, da navedemo, kakšne vrednosti znakov, ki so opredeljujoči pogoji, morajo imeti enote, da so enote populacije. V znakih, ki opredeljujejo populacijo, je torej variacija okrnjena.

Za populacijo ne velja samo množica vseh statističnih enot v gornjem smislu, marveč je populacija tudi množica podatkov za neki znak. Tako je populacija v tem smislu tudi množica vseh podatkov o starosti poročenih zaposlenih žena v SRS po stanju 31.12.1972.

2.11 Populacija mora biti nedvomno opredeljena: a) časovno, b) krajevno in c) stvarno.

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

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

Gornjo populacijo žensk smo opredelili s trenutkom. Populacije realnih enot imenujemo tudi momentne ali trenutne populacije. Momentnih populacij imamo veliko. Vse časovno enolično opredelimo z momentom, v katerem morajo posamezne enote populacije zadoščati vsem drugim pogojem, da so enote populacije.

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

2.12 Homogene in heterogene populacije. Po sorodnosti enot so populacije homogene - istorodne in heterogene - raznorodne. V homogenih populacijah enote med seboj niso preveč različne, medtem ko so heterogene populacije sestavljene iz raznorodnih enot. Homogenost populacije dosežemo, če jo opredelimo po čim več faktorialnih znakih. Delitev populacij na homogene in heterogene ni ostra in imamo bolj ali manj homogene in bolj ali manj heterogene populacije. Tako je populacija 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. Homogenost populacij je sicer pogojena z opredeljujočimi pogoji, ni pa istovetna. Opredelitev v faktorialnih znakih ima za posledico manjše variiranje oziroma večjo sorodnost ne samo v opredeljujočih temveč tudi v rezultativnih znakih, ki niso opredeljujoči. Krajevna opredelitev kmetijskih gospodarstev pogojuje homogeno populacijo glede na strukturo proizvodnje, način gospodarjenja ipd.

2.13 Delne populacije. Če v populaciji zaposlenih poročenih žena v Sloveniji konec leta 1972 dodamo pogoj, da je zaposlena kot delavka, iz prve populacije izločimo novo populacijo. To populacijo imenujemo delno populacijo prvotne, zato ker je vsaka enota druge populacije hkrati enota prve populacije.

Populacijo moremo po nekem znaku razdeliti v popoln sistem delnih populacij, če jo razdelimo v delne populacije po vseh vrednostih nekega znaka. Tako moremo populacijo vseh obrtnih obratov v Sloveniji konec leta 1972 razdeliti po strokah v popoln sistem delnih populacij. Delne populacije so bolj homogene kot osnovna populacija, ker se enote delnih populacij ujemajo še v dodatnem pogoju. Če populacijo razdelimo v popoln sistem delnih populacij po nadaljnjih faktorialnih znakih, heterogeno populacijo razdelimo v sistem homogenih populacij. To metodo večkrat uporabljamo pri analizi množičnih pojavov.

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

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

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

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

Statistični parametri

2.15 S statističnimi parametri opisujemo in proučujemo množične pojave. Z njimi številčno izražamo številčne in kakovostne značilnosti in odnose populacij. Nekatere parametre populacij dobimo s preprostim preštevanjem in seštevanjem podatkov za vse enote populacije. Tako je statistični parameter za populacijo kmetijskih gospodarstev število kmetijskih gospodarstev N. ki ga dobimo, če preštejemo vse enote populacije. Parameter "skupna vrednost proizvodnje vseh kmetijskih gospodarstev" dobimo, če seštejemo vrednost proizvodnje vseh kmetijskih gospodarstev v populaciji, $Y = \Sigma y_i$. Parameter: vsota podatkov za populacijo - imenujemo agregat. Parametre z analitično vrednostjo dobimo že s preštevanjem in seštevanjem v delnih populacijah. Tako dobimo npr. s preštevanjem po delnih populacijah število gospodarstev po vrsti proizvodnje, velikosti gospodarstev itd., s seštevanjem podatkov pa vrednost proizvodnje v vseh kmetijskih gospodarstvih po vrsti proizvodnje, velikosti gospodarstev itd.

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

TRETJE POGLAVJE

Statistično proučevanje

3.1 Statistično proučevanje socialnoekonomskih pojavov sestoji v celoti iz več etap. Te so po svojem značaju in vlogi bistveno različne, vse pa so med seboj tesno povezane in odvisne druga od druge. Statistično proučevanje obsega naslednje etape:

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

Medtem ko je določitev vsebine in namena opazovanja v najožji povezanosti s predmetom opazovanja in ustreznim področjem, je splošen načrt več ali manj delo onih, ki statistično proučevanje izvajajo.

Vrste statističnih opazovanj

3.2 Statistično opazovanje v ožjem smislu obsega zbiranje osnovnih podatkov od enot opazovanja. Glede na vrsto populacije, ki jo proučujemo, imamo več vrst statističnih opazovanj.

Podatke o populacijah realnih enot, kot so npr. prebivalstvo, kmetijska gospodarstva, industrijska podjetja ipd. zbiramo s popisom. S popisom zajamemo vse enote populacije, ki obstajajo ob določenem trenutku - kritičnem trenutku. Izvedba popisa je specifična glede na velikost populacije in glede na premikanje populacije. Velike populacije narekujejo posebno organizacijo dela, da v razmeroma kratkem času po kritičnem trenutku popišemo vse enote populacije. Veliko gibanje populacije (npr. prebivalstva) pa narekuje tako izbiro kritičnega trenutka, da je populacijo možno zajeti in da je stanje populacije čimbolj normalno. Iz obeh razlogov npr. popisi prebivalstva v poletnem času niso primerni.

Populacije dogodkov časovno opredelimo s časovnim razmikom. Temu primerno je tudi opazovanje. Medtem ko opazujemo dogodek neposredno, ko se zgodi ali v razmeroma kratkem času po njem, populacijo dogodkov sestavljajo vsi dogodki, ki so se zgodili v enem mesecu, enem letu ipd. Podatke o dogodkih zbiramo s statističnimi poročili, ki obsegajo podatke o dogodkih v določenem časovnem razdobju.

Značilno za nakazana opazovanja je, da zbiramo podatke za vse enote populacij. Dostikrat pa se zadovoljimo z ocenami

o populacijah, ki jih dobimo z delnim opazovanjem samo nekaterih enot. Posebno mesto med načini delnega opazovanja ima vzorčenje, ki je zasnovano na slučajnostnem izboru opazovanih enot. Ker je vzorčenje najpomembnejša metoda delnega opazovanja, jo bomo obravnavali v posebnem poglavju. Z opazovanjem zbrani podatki so podani v izpolnjenih statističnih obrazcih. Z urejevanjem in osnovno obdelavo prevedemo te podatke v osnovne statistične vrste, ki so dane v obdelovalnih tabelah. Te obdelovalne tabele pa so osnova za analitično obdelavo in analizo podatkov.

3.3 Ker je celotna izvedba statističnega proučevanja dolg in drag posel in zahteva sodelovanje velikega števila najrazličnejših strokovnjakov, je delo pri statističnem proučevanju običajno deljeno. Specializirane ustanove - statistični organi, ki so za to pooblaščeni, izvajajo statistična opazovanja in osnovno obdelavo podatkov za najrazličnejša področja in te podatke publicirajo v publikacijah. Te publikacije so osnova za vsebinsko analitično obdelavo in analizo.

Zaradi tega se s statističnim opazovanjem v ožjem smislu ne bomo podrobneje ukvarjali, pač pa bomo navedli nekaj osnovnih republiških in zveznih publikacij, ki jih izdajata Zavod SR Slovenije in Zvezni statistični zavod.

Mesečni statistični publikaciji Zavoda SRS za Slovenijo sta:

Mesečni statistični pregled SR Slovenije in

Družbeni razvoj SR Slovenije, ki daje kratko analizo podatkov.

Redna publikacija pa je

Statistični letopis, ki pa daje podatke o ozemlju in podnebju, družbeno politični ureditvi, prebivalstvu, zaposlenosti, družbenem produktu in narodnem dohodku, splošni gospodarski pregled in letne podatke o dejavnosti posameznih panog in podatke po občinah.

V samostojnih publikacijah izdaja zavod SR Slovenije podatke statističnih opazovanj, ki jih zavod izvaja. Razen tega v sistematičnih publikacijah izdaja: Statistične podatke po občinah, v Metodološkem gradivu pa šifrante, sezname, nomenklature in navodila za izvajanje posameznih statističnih opazovanj.

Podobni zvezni zavod za statistiko izdaja mesečne statistične podatke v reviji Indeks.

Od leta 1954 izhaja Statistični godišnjak SFRJ, ki podobno kot statistični letopis daje kompleksen pregled statističnih podatkov iz različnih področij.

Popularna izdaja zveznega statističnega zavoda je: Statistički kalendar SFRJ, ki izhaja v vseh nacionalnih jezikih Jugoslavije in v štirih tujih jezikih. V kalendarju najdemo osnovne statistične podatke za republike in SFRJ.

Redna letna publikacija je Demografska statistika, ki izhaja od leta 1950, Statistika spoljne trgovine, ki izhaja letno od 1946. Mesečni pregled medjunarodne statistike izhaja od leta 1968.

Občasne, a stalne publikacije so: Statistički bilteni, ki vsebujejo rezultate iz vseh dejavnosti statistike. Metodološki materiali vsebujejo metodološke osnove in navodila za izvajanje statističnih opazovanj.

Studije, analize i prikazi prinašajo analitične obdelave in prikaze posameznih gibanj in pojavov.

Razen tega izdaja Zvezni zavod za statistiko Statističko revijo, ki je trimesečni teoretični časopis. Statistička revija je organ JSD.

Razen teh izdajajo mednarodne organizacije niz statističnih publikacij, ki so vir podatkov. Te publikacije so podobne kot naše publikacije, bodisi mesečne, letne ali občasne in prinašajo podatke iz mednarodne statistike.

ČETRTO POGLAVJE

U R E J E V A N J E S T A T I S T I Č N E G A G R A D I V A

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

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

Razvrščanje vrednosti znakov

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

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

4.4 V bistvu iz istega razloga in po istih načelih kakor v vsakdanjem življenju grupiramo znake v skupine tudi v statistične namene. Natančne vrednosti dostikrat niso potrebne. Število vseh možnih vrednosti znaka je za nekatere znake veliko ali celo neomejeno. Zato je običajno prikladneje, da sorodne vrednosti združimo v skupinske 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.

4.5. Preden bomo obravnavali grupiranje posameznih vrst znakov v skupine, navedimo nekaj splošnih načel, ki veljajo za grupiranje vseh znakov.

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

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

Vrednosti enega in istega znaka moremo grupirati po različnih načelih, tako da za isti znak dobimo več različnih grupacij. Ker posamezne vrednosti združujemo v grupe po sorodnosti, daje načelo grupiranja merilo o sorodnosti. Načelo grupiranja je v tesni zvezi s predmetom opazovanja, namenom in potrebami proučevanja pojava.

Razen tega pri grupiranju vsaka grupa običajno dobi neko ime - novo vrednost, ki pojmovno združuje vse vrednosti znakov za ustrezno grupo.

Grupe dobe posebna imena: geografske grupe npr. imena občin, republik itd., časovne grupe datume, mesece, leta itd., stvarne grupe nazive grup poklicev itd.

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

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

Za proučevanje pojava je vsekakor bolje grupiranje po načelu, ki je v vsebinski zvezi s pojavom, ki ga proučujemo. Če upoštevamo to načelo, razdelimo teritorij v vsebinske geografske grupe - rajone. Geografske rajone izdelujemo po različnih merilih in za različne potrebe. Rajonizacijo pa spremljajo tehnične težave. Izvesti rajonizacijo po nekem vsebinskem načelu je običajno zelo težko, ker moramo teritorij in razmestitev pojava, ki ga proučujemo, zelo dobro poznati.

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

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

4.8 Grupiranje stvarnih znakov je bistveno različno za atributivne in za numerične znake.

4.9 Atributivne znake z velikim številom vrednosti običajno grupiramo tehnično po decimalni klasifikaciji. Osnova decimalne klasifikacije je v tem, da največ po deset osnovnih pojmov združimo v grupe prve stopnje, največ po deset grup prve stopnje v grupe druge stopnje itd. To grupiranje v grupe višjih stopenj ponavljamo toliko časa, da imamo največ deset grup zadnje stopnje. To načelo ni vsebinsko, temveč tehnično. Vsak pojem moremo namreč po decimalni klasifikaciji označiti s številom, ki ima toliko mest, kolikor stopenj ima grupiranje.

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

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

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

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

63 Izdelki iz lesa in plute (razen pohištva).
631 Furnir, vezane plošče, deske, umetno ali rekonstruiran in drugi les, obdelan neomejeno.

631-01	Furnir.	
631-01-11	Furnir bukov normalen	m^3/kg.
631-01-12	Furnir bukov za stole	m^3/kg.
631-01-20	Orehov furnir	m^3/kg.
631-01-30	Hrastov furnir	m^3/kg.
631-01-40	Furnir slepi	m^3/kg.
631-01-90	Furnir iz drugih listavcev	m^3/kg.

Iz zgleda vidimo smisel in pomen decimalne klasifikacije. Prva številka v oznaki 631-01-11, s katero je zaznamovan normalen bukov furnir (6) pove, da normalen bukov furnir spada po klasifikaciji v grupo izdelkov, klasificiranih pretežno po materialu. Prvi dve številki (63) pomenita, da spada ta izdelek

v podgrupo izdelkov iz lesa in plute (brez pohištva). Prve tri številke (631) oznake pomenijo, da spada navedeni izdelek v nadaljnjo podgrupo furnirjev, vezanih plošč, desk itd.

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

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

4.11 Grupiranje za vrednosti stvarno-numeričnih znakov mora ustrezati vsebinskim in tehničnim okoliščinam. Ker je analiza numeričnih znakov najbolj razvita, se pri grupiranju numeričnih znakov oziramo na formalno-tehnična načela grupiranja, če le niso nedopustna iz vsebinskih razlogov.

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

4.12 Grupe numeričnega znaka, ki jih imenujemo razrede, so določene s spodnjo mejo $x_{k,min}$ in z zgornjo mejo $x_{k,max}$. Nobena vrednost v razredu ni manjša od spodnje meje razreda in nobena ni večja od zgornje meje razreda. V posamezni razred spadajo vse vrednosti med spodnjo in zgornjo mejo razreda. Vsak razred ima svojo širino razreda i_k, ki je razlika med zgornjo in spodnjo mejo razreda.

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

in sredino razreda x_k, ki je polovica vsote mej razreda

$$x_k = \frac{x_{k,min} + x_{k,max}}{2} \qquad (4.2)$$

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

Iz gornjega zgleda vidimo, da niso vsi razredi omejeni s spodnjo in zgornjo mejo. Zadnji razred v našem zgledu ima le

sponjo mejo (8 ha). Take razrede, ki so omejeni samo navzdol ali samo navzgor, imenujemo odprte razrede. Odprte razrede tvorimo, če je število enot, ki imajo vrednost nad ali pod neko mejo, majhno, podatki pa so med seboj zelo različni.

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

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

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

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

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

4.13 Površine gospodarstev ne merimo natančno do mm^2, marveč jih običajno zaokrožujemo na are. Kot 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 zaokrožimo na 3 ha 28 a, ker je stvarna površina bliže 3 ha 28 a kakor površini 3 ha 27 a, po drugem načinu pa je zaokrožena vrednost 3 ha 27a, 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 gornjo grupacijo površin takole:

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

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

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

Tako na primer mesečne osebne dohodke grupiramo v grupe.

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

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

4.14 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 5 ni vrednosti. Tako navajanje razredov pa moti pri izračunavanju širine razredov. Razlike med tri in ena je dva, v razredu so pa tri vrednosti: 1, 2, 3. Širina razreda je tri, ne pa dva.

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

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

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

Podjetja so po številu delavcev grupirana v tele grupe:

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

Kvocienti med mejami so stalni (2,00), razen v razredu 61 do 125, v katerem je 2,08, zato da so meje naslednjih razredov lahko prešle na naravnejše meje.

PETO POGLAVJE

P R I K A Z O V A N J E S T A T I S T I Č N I H P O D A T K O V

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

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

Prednosti tabelaričnega prikaza statističnih podatkov so predvsem:

a) v tabeli moremo prikazati po potrebi razmeroma veliko podatkov;

b) v tabeli prikažemo podatke s poljubno natančnostjo;

c) tabelarični način je enostavnejši kakor grafični, ker so načela za sestavljanje dobrega tabelaričnega prikaza enotnejša kakor pri grafičnem prikazovanju, pri katerem za različne statistične vrste in v različne namene uporabljamo različne metode grafičnega prikazovanja.

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

Statistične vrste

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

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

5.3 V tabeli 5.1 je prikazana kot zgled krajevne ali geografske vrste porazdelitev narodnega dohodka v SFRJ v letu 1965 po socialističnih republikah.

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

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

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

Tabela 5.2 Proizvodnja eleketrične energije v SFRJ v razdobju 1952-1966 v mlj kWh (Vir: SGJ-72)

Leto	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962
Proiz-vodnja	2700	2982	3440	4340	5048	6252	7356	8106	8928	9924	11275

Leto	1963	1964	1965	1966	1967	1968	1969	1970	1971
Proiz-vodnja	13535	14189	15523	17174	18702	20641	23375	26023	29509

Statistična vrsta proizvodnje električne energije v tabeli 5.2 je intervalna časovna vrsta, ker prikazuje intervalne

podatke. Momentna časovna vrsta pa je npr. število zaposlenih v družbenem sektorju na dan 30. septembra v razdobju 1957-1971 v SFRJ v tabeli 5.3.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Prav poseben pomen imajo statistične vrste, kakršna je npr. število kmetijskih gospodarstev po številu članov v tabeli 5.6. Ta statistična vrsta prikazuje variabilnost števila članov za anketirana gospodarstva. Take statistične vrste imenujemo frekvenčne distribucije. Ker so za analizo populacij osnovne važnosti, jih bomo podrobneje poručevali v posebnem poglavju.

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

Tabela je s sistemom vrst in stolpcev razdeljena 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 s pomenom ustreznega stolpca, v katerem je polje oziroma podatek. Vsi ti pojmi so shematično nakazani v tabeli 5.7. Polje A v tabeli 5.7 je polje za podatek o vzdrževanih ženskah v starosti izpolnjenih treh let.

Tabela 5.7 Tabela 1 za obdelavo vzorca predhodnih rezultatov za popis prebivalstva 31.3.1953 v SFRJ

Delavnost	Delovni		Osebe z osebnimi dohodki		Vzdrževani		glava
Spol / Starost	Moški	Ženske	Moški	Ženske	Moški	Ženske	
0							
1							
2	V r s t a				polje		
3					stolpec	A	telo
4							
5							
6							
Čelo							

ŠESTO POGLAVJE

R E L A T I V N A Š T E V I L A

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 8,66 % od skupnega števila prebivalstva Jugoslavije, kakor pa če navedemo, da je bilo v Sloveniji po popisu 31.3.1953 1,466.425 prebivalcev od skupnega števila 16,936.573 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 popisom 15.3.1948 in 31.3.1953 povečalo v Sloveniji za 5,36 %. 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 1,466.425 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.

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 prikladnejše za prikazovanje in analizo (npr. 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 (npr. 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 (npr. proizvodnja jekla na 1 prebivalca je bila v letu 1957 v FLRJ 54,7 kg/prebivalca);

c) indekse, kadar primerjamo prirejene istovrstne podatke (npr. indeks spremembe števila prebivalstva v Sloveniji od popisa 1948 do popisa 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 le 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_1^o dobimo, da delni podatek delimo s celoto.

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

$$y_1^o = \frac{y_1}{y} \; ; \quad y_1\% = 100 \cdot \frac{y_1}{y} ; \quad y_1\%o = 1000 \cdot \frac{y_1}{y} \qquad (6.1)$$

Pri tem pomeni: y_1 = podatek za del populacije, y = podatek za populacijo, y_1^o = strukturni kvocient, $y_1\%$ = strukturni odstotek, $y_1\%o$ = strukturni delež, izražen v promilih.

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

Tabela 6.1 Vrednost izvoza iz FLRJ po stopnji obdelave izvoznih 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,o	12,3	19,9	31,9	16,6	16,o	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 (1.612 milijard dinarjev) in izračunati v odstotkih, 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 6.3.

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 6.2 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 (617 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	65,6	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 6.4 vidimo velike razlike v sestavu družbenega produkta po gospodarskih sektorjih. Medtem ko so na primer v splošnem sektorju plače samo 21,o % družbenega produkta, pomenijo plače v zadružnem sektorju 39,2 %, v privatnem pa celo 79,6 % od skupnega družbenega produkta. Analogno so velike razlike tudi v drugih elementih.

Tudi tabela 6.5 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.

G r a f i č n o p r i k a z o v a n j e s t r u k t u r

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

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.

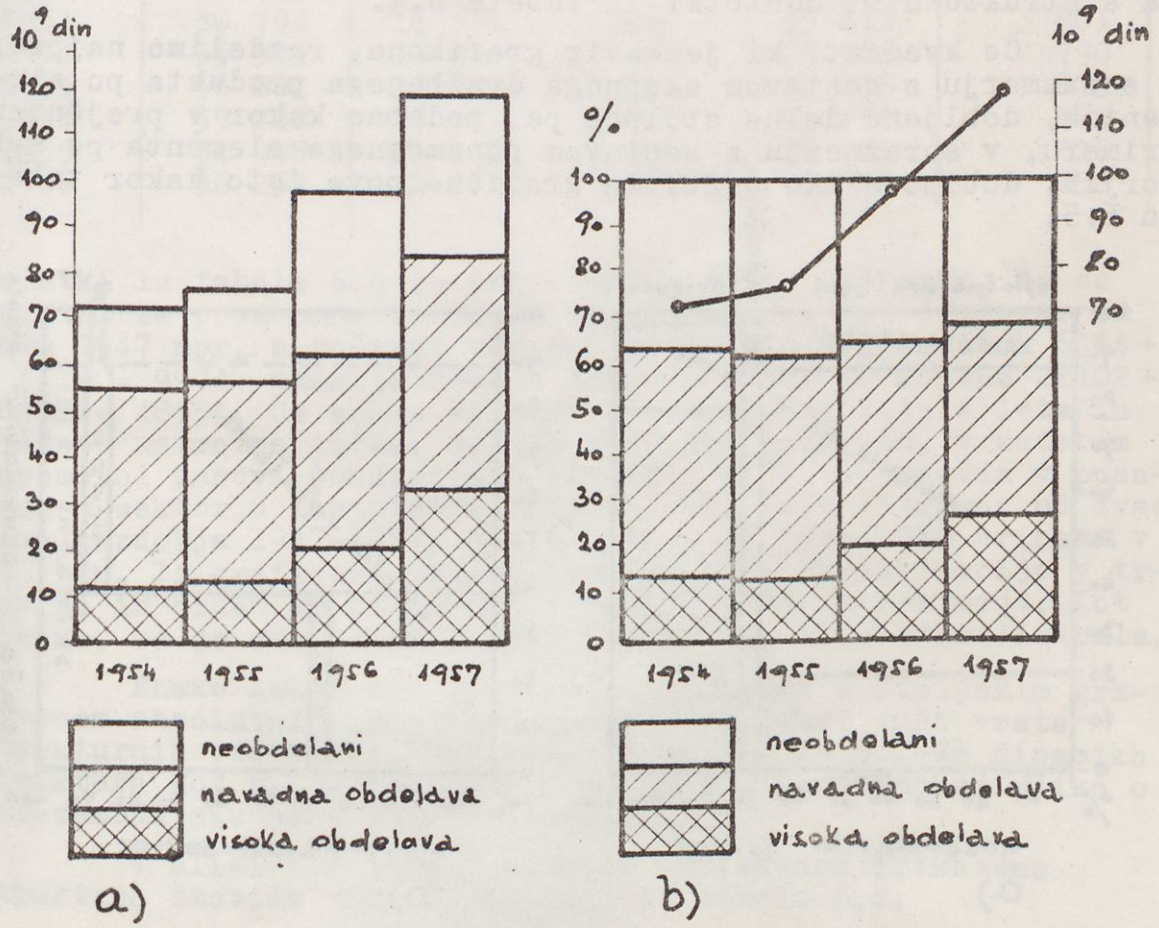
Vzemimo kot primer vrednost izvoza iz Jugoslavije po stopnji obdelave v razdobju 1954-1957 iz tabele 6.1. Strukturni stolpci iz absolutnih podatkov v tej tabeli so prikazani v sliki 6.1a.

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 6.1b, 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.

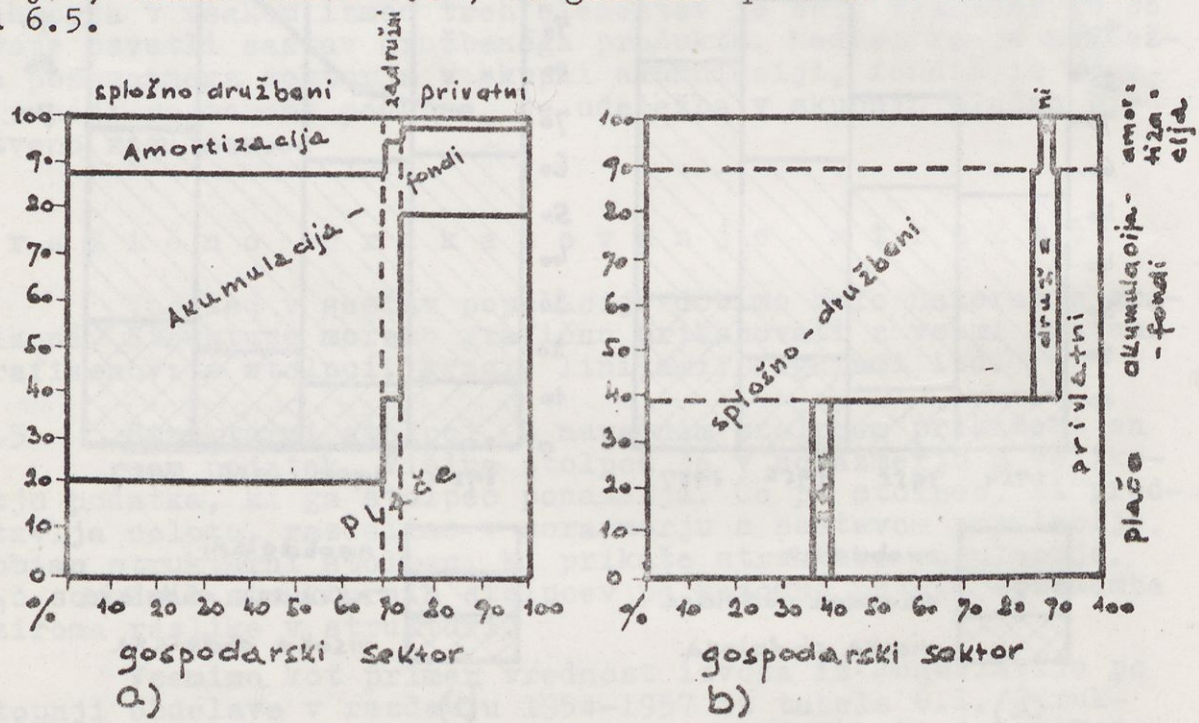


Slika 6.1 Struktura vrednosti izvoza iz SFRJ po stopnji obdelave izvoženih proizvodov (Vir: Tabela 6.1)

Vzemimo za primer strukturo družbenega produkta v SFRJ v letu 1956 po gospodarskih sektorjih in elementih iz tabele 6.2. Celoten družbeni produkt ponazorimo s kvadratom.

Strukturo družbenega produkta pa prikažemo tako, da kvadrat razdelimo v sorazmerju z deležem posameznega sektorja v celotnem družbenem produktu. Celoten družbeni produkt je tako razdeljen v 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 6.2a. Ta je fotografija tabele 6.4. 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 6.4.

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 6.2b; ta grafično pove isto kakor tabela 6.5.



Slika 6.2 Struktura družbenega produkta v SFRJ 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 1947-1956 v SFRJ.

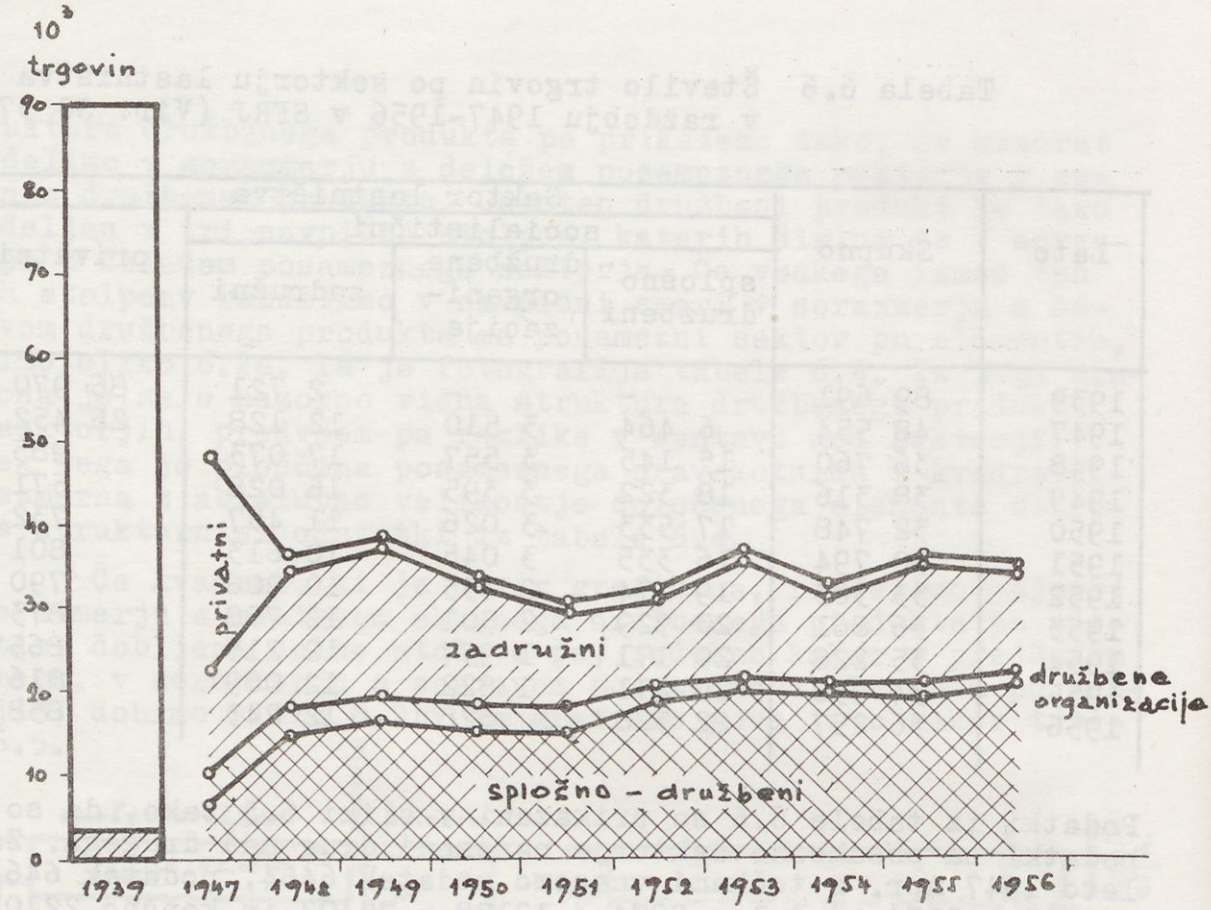
Tabela 6.6 Število trgovin po sektorju lastništva v razdobju 1947-1956 v SFRJ (Vir: SG 57)

Leto	Skupno	Sektor lastništva			
		socialistični			privatni
		splošno družbeni	družbene organizacije	zadružni	
1939	89 691	-	-	2 721	86 970
1947	48 554	6 464	3 510	12 128	26 452
1948	36 760	15 145	3 557	17 073	985
1949	38 316	18 322	3 395	16 028	571
1950	32 748	17 533	3 026	11 457	732
1951	30 794	16 335	3 045	10 613	801
1952	33 303	19 174	2 433	10 906	790
1953	36 861	20 929	2 301	12 758	873
1954	35 228	20 171	1 977	12 215	865
1955	34 958	20 231	1 822	12 089	816
1956	36 257	22 528	628	12 243	858

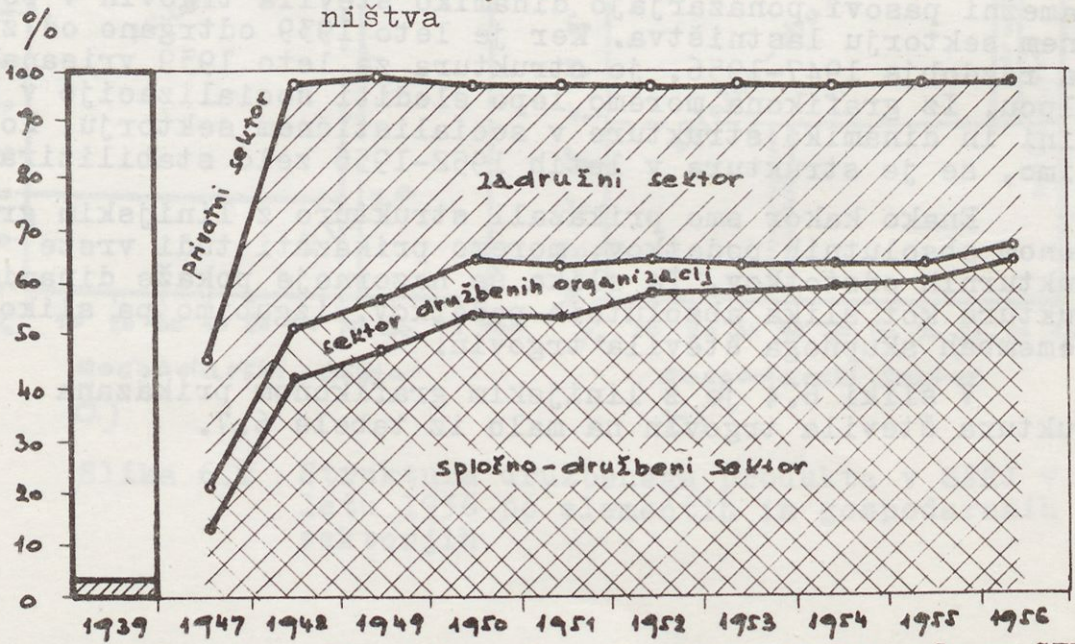
Podatki iz tabele 6.6 so prikazani v sliki 6.3 tako, da so podatki za posamezne sektorje naneseni drug nad drugega. Za leto 1947 npr. s točkami vnesemo podatek 6464, podatek 6464 + + 3510 = 9974, nadalje 9974 + 12128 = 22102 in končno 22102 + 26452 = 48554. Če enako vrišemo strukturo za ostala leta in 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 1947-1956, 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.

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 6.4 je z linijskim grafikonom prikazana struktura števila trgovin na malo iz tabele 6.7.



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



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

Tabela 6.7 Struktura števila trgovin na malo v razdobju 1947-1956 v SFRJ 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	46,4	2,7
1949	100,0	47,8	8,9	41,8	1,5
1950	100,0	53,6	9,2	35,0	2,2
1951	100,0	53,0	9,9	34,5	2,6
1952	100,0	57,6	7,3	32,7	2,4
1953	100,0	56,8	6,2	34,6	2,4
1954	100,0	57,2	5,6	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 strukturne deleže Y % pred 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 SFRJ v letu 1956.

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

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

Ločne stopinje smo izračunali: 3,6 . 52,0 % = 187°

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.

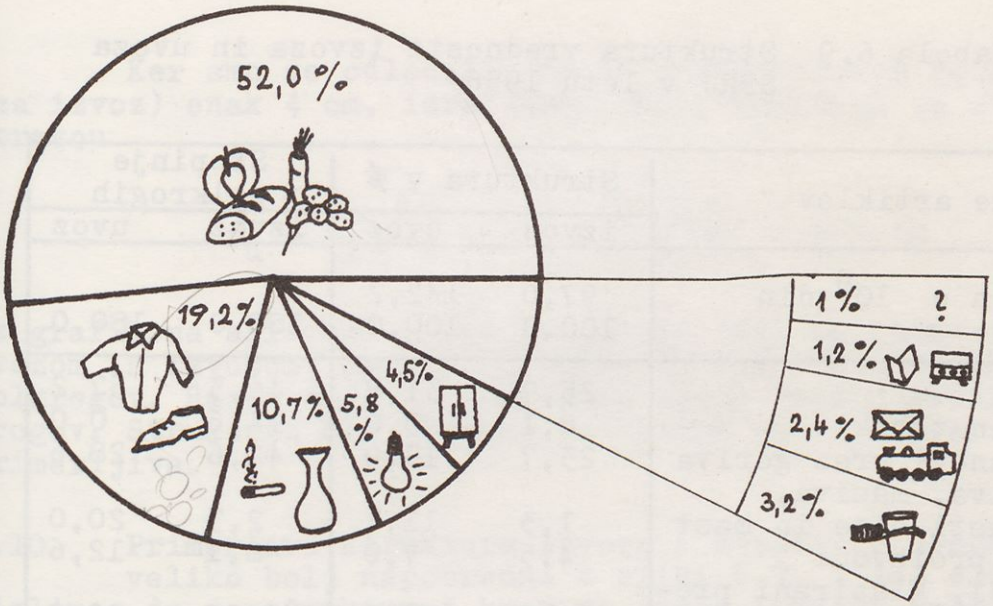
Če imamo več strukturnih vrst, ki jih med seboj primerjamo, strukturni krogi običajno niso prikladni. Primerljivost med strukturami, če so prikazane s krogi, je namreč veliko slabša, kakor če uporabimo stolpce ali linije. Kljub temu včasih kroge uporabljamo za prikazovanje struktur. Uporabni so npr. 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 npr. 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°.

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, proporcionalne velikosti pojava,



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

radije krogov, za katere želimo, da so njihove ploščine v sorazmerju z velikostjo pojava, izračunavamo po obrazcu

$$r_1 = r_o \sqrt{\frac{Y_1}{Y_o}} \qquad (6.4)$$

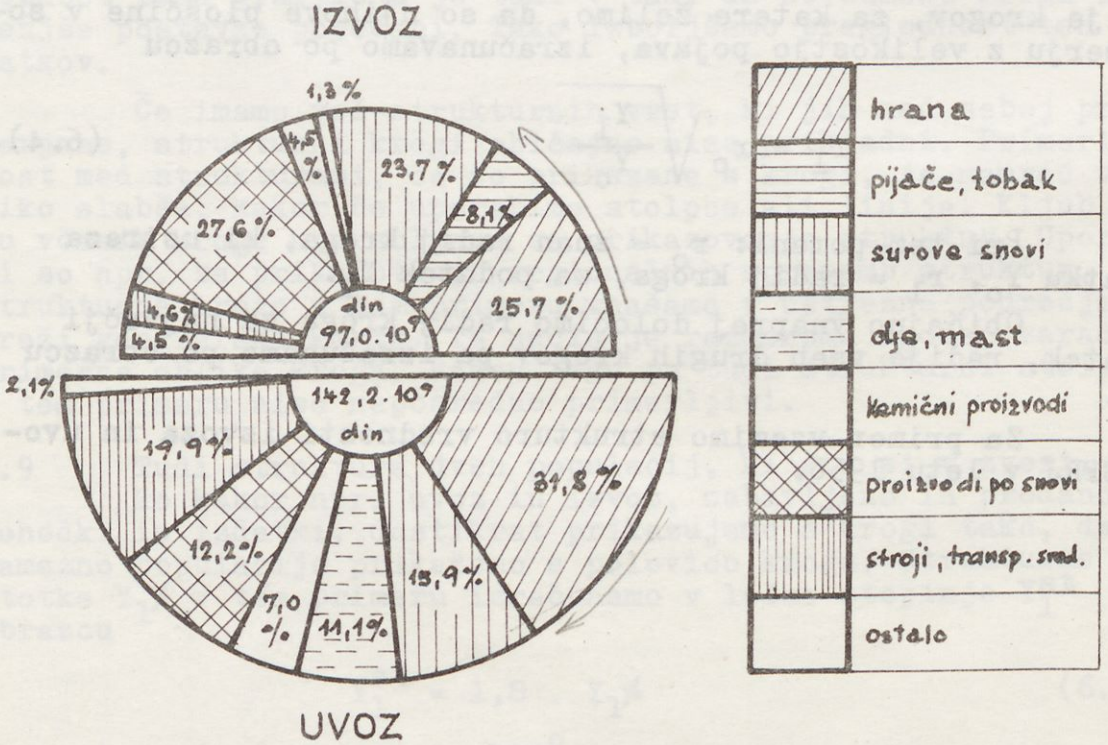
Pri tem pomeni: r_o = znan radij kroga, ki ustreza podatku Y_o, r_1 = radij kroga, za podatek Y_1.

Običajno vnaprej določimo radij kroga za največji podatek, radije vseh drugih krogov pa izračunamo po obrazcu 6.4.

Za primer vzemimo strukturo vrednosti izvoza in uvoza SFRJ v letu 1956.

Tabela 6.9 Struktura vrednosti izvoza in uvoza SFRJ v letu 1956

Skupine artiklov	Struktura v %		Stopinje v polkrogih	
	izvoz	uvoz	izvoz	uvoz
S k u p n o 10^9 din	97,0 100,0	142,2 100,0	180,0	180,0
Hrana	25,7	31,8	46,3	57,2
Pijače in tobak	8,1	0,0	14,6	0,0
Surove snovi brez goriva	23,7	15,9	42,6	28,6
Min.goriva, maziva, živ.rastl.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



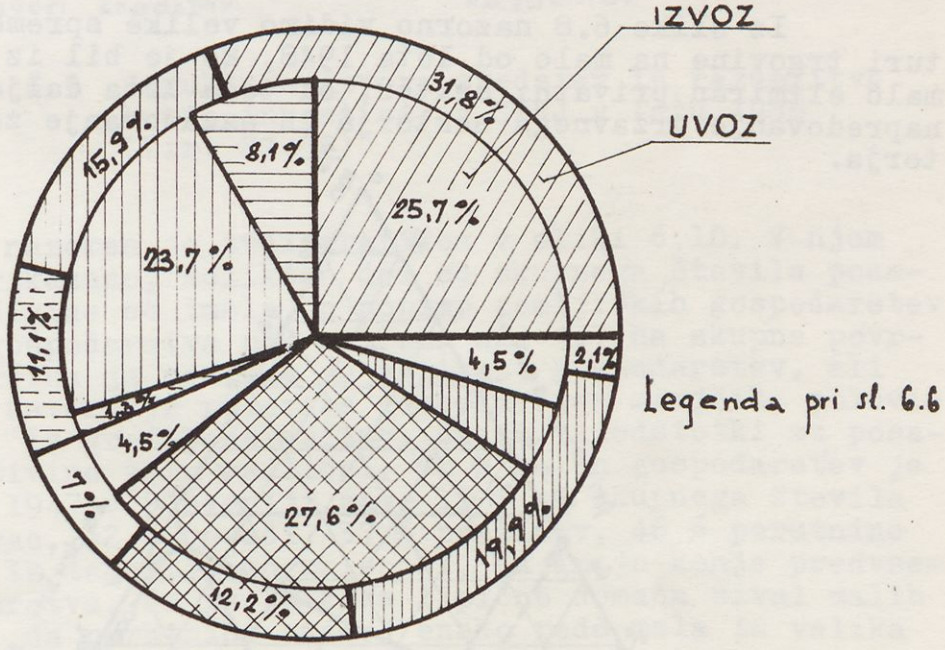
Slika 6.6 Struktura vrednosti izvoza in uvoza SFRJ v letu 1956 (Vir: tabela 6.9)

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 razvidno 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 6.7. 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.

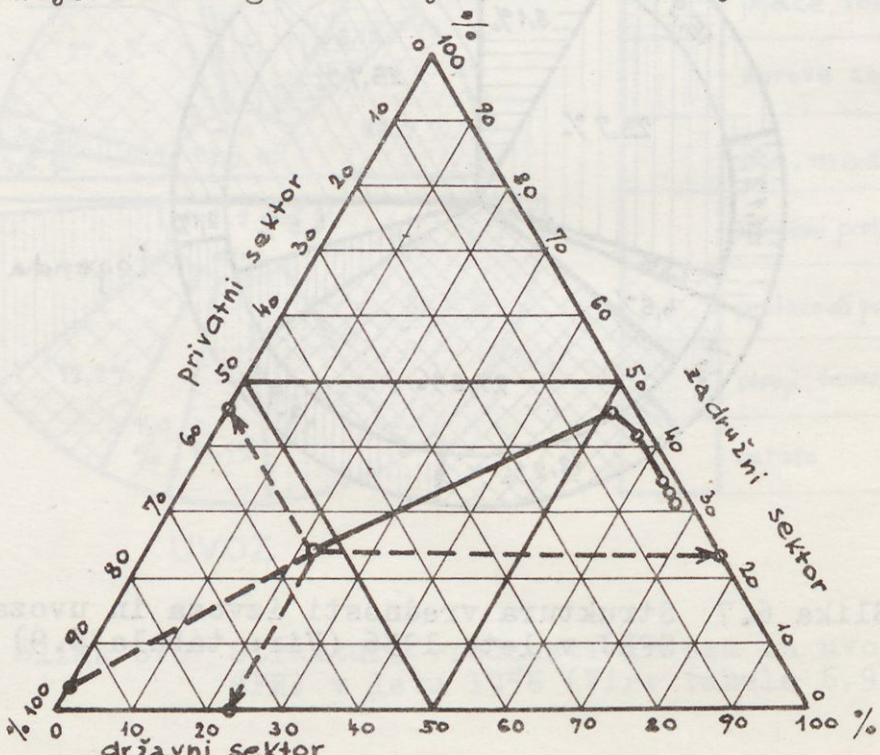


Slika 6.7 Struktura vrednosti izvoza in uvoza SFRJ v letu 1956 (Vir: tabela 6.9)

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 enostavno prikazati s toliko točkami, kolikor imamo strukturnih vrst.

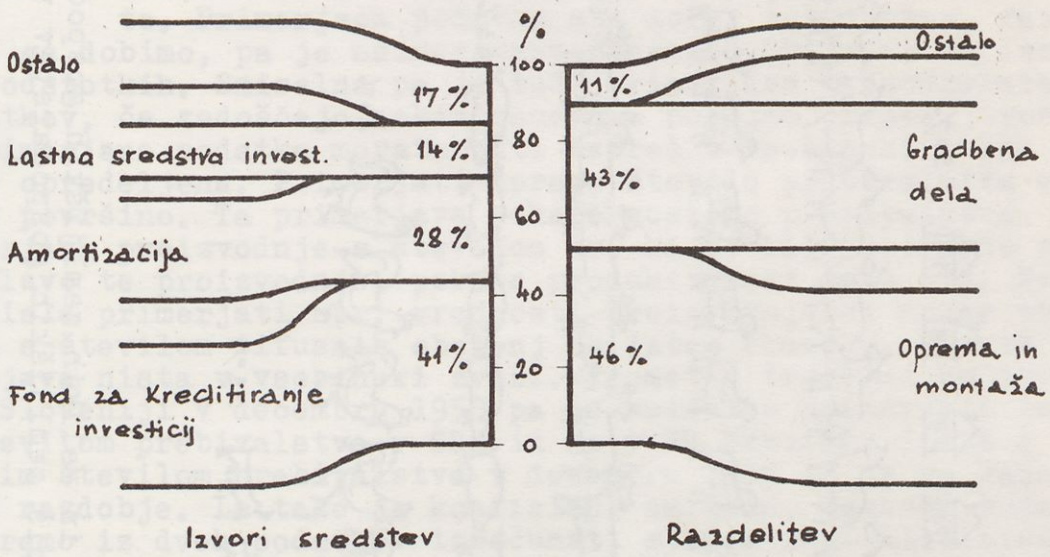
V sliki 6.8 je prikazana struktura števila trgovin po sektorju lastništva v razdobju 1947-1956 iz tabele 6.7. Iz grafikona vidimo, da je na vsaki stranici enakostraničnega trikotnika skala odstotkov za enega izmed sektorjev lastništva. Trikotnik je razdeljen z mrežo vzporednic k posameznim osem. V sliki 6.8 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 6.8 nazorno vidimo velike spremembe v strukturi trgovine na malo od leta 1948, ko je bil iz trgovine na malo elimiran 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 SFRJ v razdobju 1947-1956

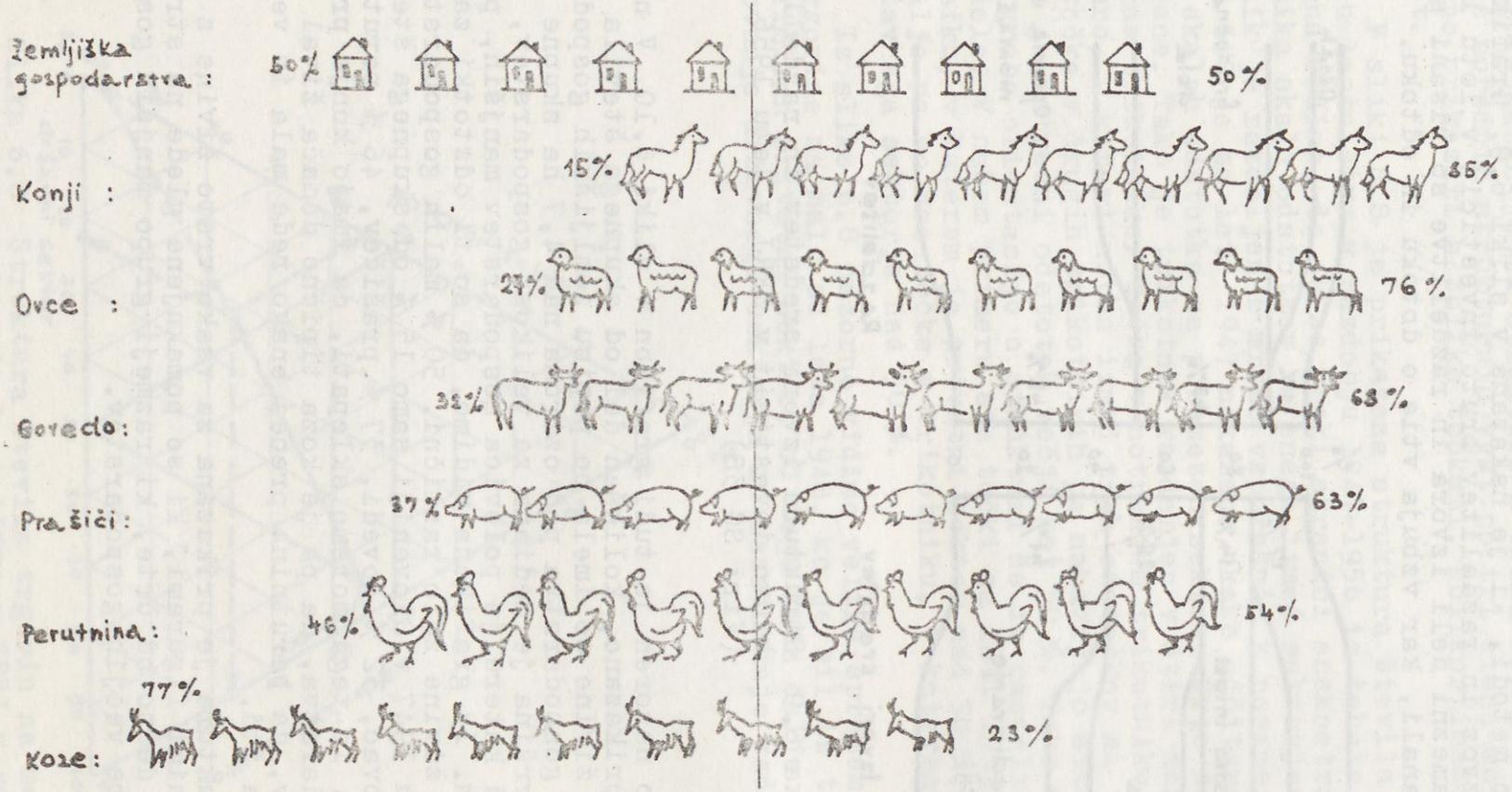
6.12 Druge metode prikazovanja struktur. Ker si moremo predstavljati bilančne podatke kot tok, ki na eni strani doteka, na drugi pa odteka, strukturo bilanc nazorno prikažemo po metodi, ki je nakazana v sliki 6.9. Grafikon ponazarja izvoz in razdelitev bruto investicij v letu 1956 v SFRJ. Posamezni deli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



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

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

Struktura je prikazana za vsako vrsto živine z desetimi shematičnimi figurami, ki so pomaknjene glede na strukturo na levo ali 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 npr. vrednosti proizvodnje na nekem območju s številom tifusnih obolenj na istem območju, ker ta dva pojava nista v vsebinski zvezi. Promet v trgovini na drobno v Sloveniji v decembru 1958 pa je smiselno primerjati le s številom prebivalstva v SRS in ne v SR 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 npr. koeficient nepismenih, ne vzamemo v primerjavo vsega prebivalstva, temveč se omejimo samo na prebivalstvo, staro nad deset let, da tako izključimo iz primerjave otroke. Enako ravnamo pri izračunavanju drugih pokazateljev, npr. 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 npr. primerjamo število prebivalstva v Sloveniji po popisu 31.3.1953 (1,466.425 prebivalcev) s površino Slovenije ob istem popisu (19992 km^2), dobimo gostoto prebivalstva v Sloveniji 31.3.1953.

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

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

6.15 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 (momemtna populacija), ker se število umrlih nanaša na določeno razdobje, npr. leto, število prebivalstva pa moremo ugotoviti samo za določen moment. V takih primerih si pomagamo s tem, da momenten podatek spremenimov 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 črtrtletja 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 razmaka (sredine mesecev, sredine dekad, sredine let itd.), izračunavamo povprečje po obrazcu

$$\bar{Y} = \frac{1}{r}\,(Y_1 + Y_2 + \ldots\ldots + Y_{r-1} + Y_r) \qquad (6.5)$$

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

Po tem obrazcu izračunavamo npr. 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 razmakov, izračunamo povprečje po obrazcu

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

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

Če imamo npr. 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 delavca 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 6.5, da je $\bar{Y} = Y_1$. Iz tega sledi, da moremo povprečje nadomestiti s stanjem v sredini intervala. Tako npr. nadomeščamo povprečno število prebivalstva s "srednjim številom prebivalstva", to je s številom prebivalstva sredi leta (30. junija).

Če je r = 1 obrazec 6.6 degenerira v

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

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 = dolčina časovnega razdobja, za katerega izračunavamo koeficient, E = 100, 1000, 10000, odvisno od tega, na koliko enot momentnega podatka izračunavamo koeficient.

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

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

$$K = \frac{2395816 \cdot 1000}{17067000 \cdot 5} = 28{,}1$$ 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 npr. 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 $\bar{Y}$ = 1,556.578, število umrlih v januarju 1958 X = 1464, E = 1000, i = 31/365, je koeficient mortalitete enak

$$K = \frac{1464 \cdot 1000}{1556578 \cdot 31/365} = 11{,}1$$ umrlih na leto na 1000 prebivalcev.

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

6.19 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 SRS 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 SRS)

Mesec	Zaloga konec meseca v t	Poraba v t
1957 D	664	
1958 J	774	494
F	888	489
M	789	562
A	546	468
M	568	465
J	725	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.

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

Povprečno zalogo $\bar{Y}$ pa izračunamo po obrazcu 6.6, ker imamo podatke za konec mesecev.

$$\bar{Y} = \frac{\frac{564}{2} + 774 + 888 + 789 + 546 + 568 + \frac{726}{2}}{6} = 710 \text{ t}$$

Povprečna zaloga je torej 710 t.

Ker je koeficient merjen s številom obratov v enem mesecu, naš interval pa obsega pol leta, je i = 6. Iz teh podatkov izračunamo, da je koeficient obračanja zalog K kremenove moke v prvem polletju 1958 enak

$$K = \frac{X\ E}{\bar{Y}\ i} = \frac{3022 \cdot 1}{710 \cdot 6} = 0{,}71$$

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

Z analognim računom smo dobili koeficiente obračanja zalog za isto podjetje za rjavi premog K = 12,0, pepeliko K = = 0,50, kalcinirano sodo K = 2,78. Kakor vidimo, so koeficienti obračanja zalog, ki so tudi operativno zelo pomemben 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

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

Po navedenem postopku izračunavamo še niz drugih koeficientov iz ekonomskih statistik. Tako je npr. proizvodnja na enega delavca pokazatelj produktivnosti dela, proizvodnja določenih

proizvodov na prebivalca, npr. 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 1967 po republikah. V tabeli 6.11 imamo osnovne podatke, iz katerih bomo izračunali koeficiente.

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

	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Število prebivalcev 30.6. 1957 v 10^3	18234	7448	4109	1582	3168	1454	473
Število trgovin na malo:							
31.12.1956	36257	14017	9296	4662	4634	2795	853
31.12.1957	37651	14601	9798	4722	4822	2810	898
Število zaposlenih v trgovini na malo:							
31.12.1956	112511	45954	27589	13689	14184	8523	2572
31.12.1957	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	483608	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 izračunamo povprečja, da bomo mogli število trgovin in število zaposlenih primerjati s prometom, ki je intervalnega značaja.

Za število trgovin v SFRJ je npr. iz zgornjih podatkov povprečno število trgovin enako polovici vsote števila trgovin po stanju 31.12.1956 in 31.12.1957.

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

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	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
trgovin	36954	14309	9547	4692	4728	2802	876
zaposlenih	114315	46071	28440	13855	14375	8832	2742

Iz podatkov v tabeli 6.11 in 6.12 izračunamo niz koeficientov, ki osvetljujejo trgovino na drobno v SFRJ. Tako moremo s številom prebivalstva primerjati vse druge podatke. Število prebivalstva na eno trgovino meri gostoto trgovinske mreže, število prebivalstva na enega zaposlenega preskrbljenost s trgovskim kadrom itd. Ne da bi se spuščali v globljo analizo prometa na glavo prebivalstva, ki je bistveno odvisen razen od kupne moči tudi od drugih faktorjev, npr. od socialne strukture prebivalstva, merimo s temi koeficienti kupno moč prebivalstva.

Primerjava števila trgovin s številom zaposlenih ali s skupnim prometom prikaže velikost trgovinskih obratov. Primerjava skupnega števila zaposlenih s skupnim prometom pokaže produktivnost dela v trgovini na 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 30.6.1957, podatke o številu trgovin in številu zaposlenih pa na dan 31.12.1956 in 31.12. 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 31.12.1956 ali 31.12.1957, ker imamo za te da-

tume 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 6.13 so izračunani vsi navedeni koeficienti.

Tabela 6.13 Koeficienti trgovine na malo v SFRJ v letu 1957 po republikah (Vir: SG 58)

	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Število prebivalcev na eno trgovino	493	520	431	337	670	519	539
Število prebivalcev na enega zaposlenega	159	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 industr. 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	6260	5020	5930
Promet z industr. proizvodi na sto din prometa s prehrambenimi proizvodi	207	267	194	184	158	176	120

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 (670). 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 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 BiH, 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 (267 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 SFRJ moremo torej izraziti tudi z 2.15 trgovin na 1000 prebivalcev. Kateri izmed obeh možnih koeficientov, ki jih moremo izračunati v vsakem primeru, je boljši, je odvisno od namena izračunavanja in predstave koeficienta. V mnogih 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 $p = a \cdot b$. Pri tem pomeni p ploščino, a in b stranici, med K, X in Y pa velja podobna zveza $K = X/Y$ ali $X = K.Y$. Stranici pravokotnika moremo uporabiti za prikaz koeficienta in enega absolutnega podatka, 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 6.14 podatke o številu prebivalstva in številu osebnih in tovornih avtomobilov in avtobusov. Iz teh podatkov moremo za vsako državo posebej izračunati koeficiente o številu motornih vozil na 1000 prebivalcev - koeficiente motorizacije. Ti podatki so po zgornjem načinu prikazani v sliki 6.11.

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

Država	Število prebivalstva v 000	Število motornih vozil v tisočih			Število motornih vozil na tisoč prebiv.		
		skupno	osebnih	kamion. avtobus.	skupno	osebnih	kamion. avtobus.
ZDA	165271	62020	51989	10031	375	315	60
Francija	43274	4820	3350	1470	111	77	34
Velika Britanija	51215	4732	3550	1182	92	69	23
Zap.Nemčija	49995	2909	1813	1096	58	36	22
Italija	48016	1210	879	341	25	18	7

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.

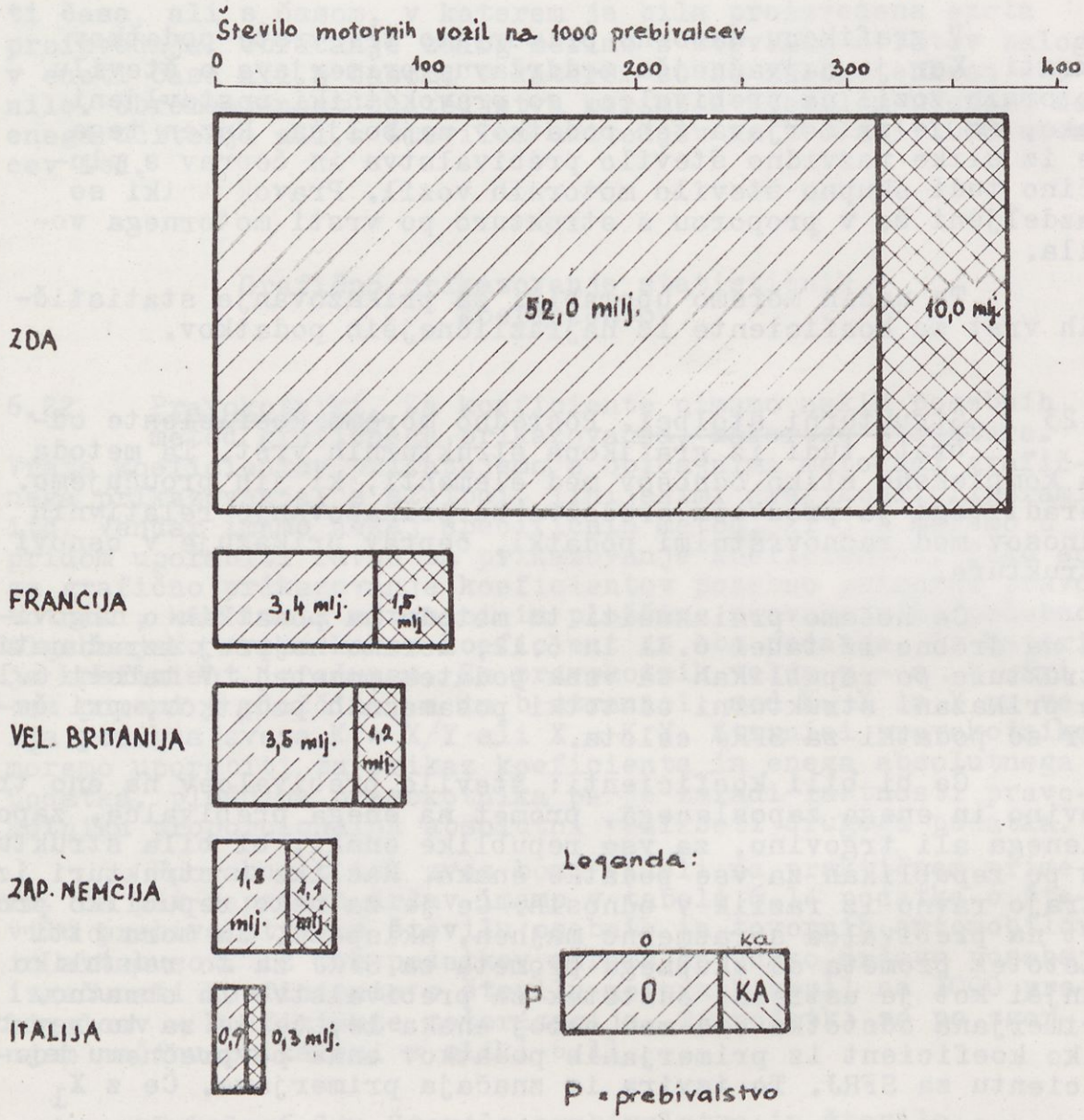
Ta način moremo uporabiti za prikazovanje statističnih vrst za koeficiente iz najrazličnejših podatkov.

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 SFRJ 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 odstotek prometa od skupnega prometa za SFRJ 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 SFRJ. To izvira iz značaja primerjave. Če z X_1 zaznamujemo prvi podatek za republiko, z X ustrezni skupni podatek za SFRJ, z Y_1 drugi podatek za republiko, z Y pa ustrezni skupni podatek za SFRJ, 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 SFRJ v letu 1947 po republikah

Pokazatelj	SFRJ	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	7,6	2,4
Zaposleni	100,0	40,3	24,9	12,1	12,6	7,7	2,4
Promet skupni	100,0	38,8	26,4	14,7	12,0	5,9	2,2
Prehrambeni	100,0	32,5	27,7	16,1	14,1	6,6	3,0
Industrijski	100,0	41,9	25,9	14,2	10,7	5,6	1,7

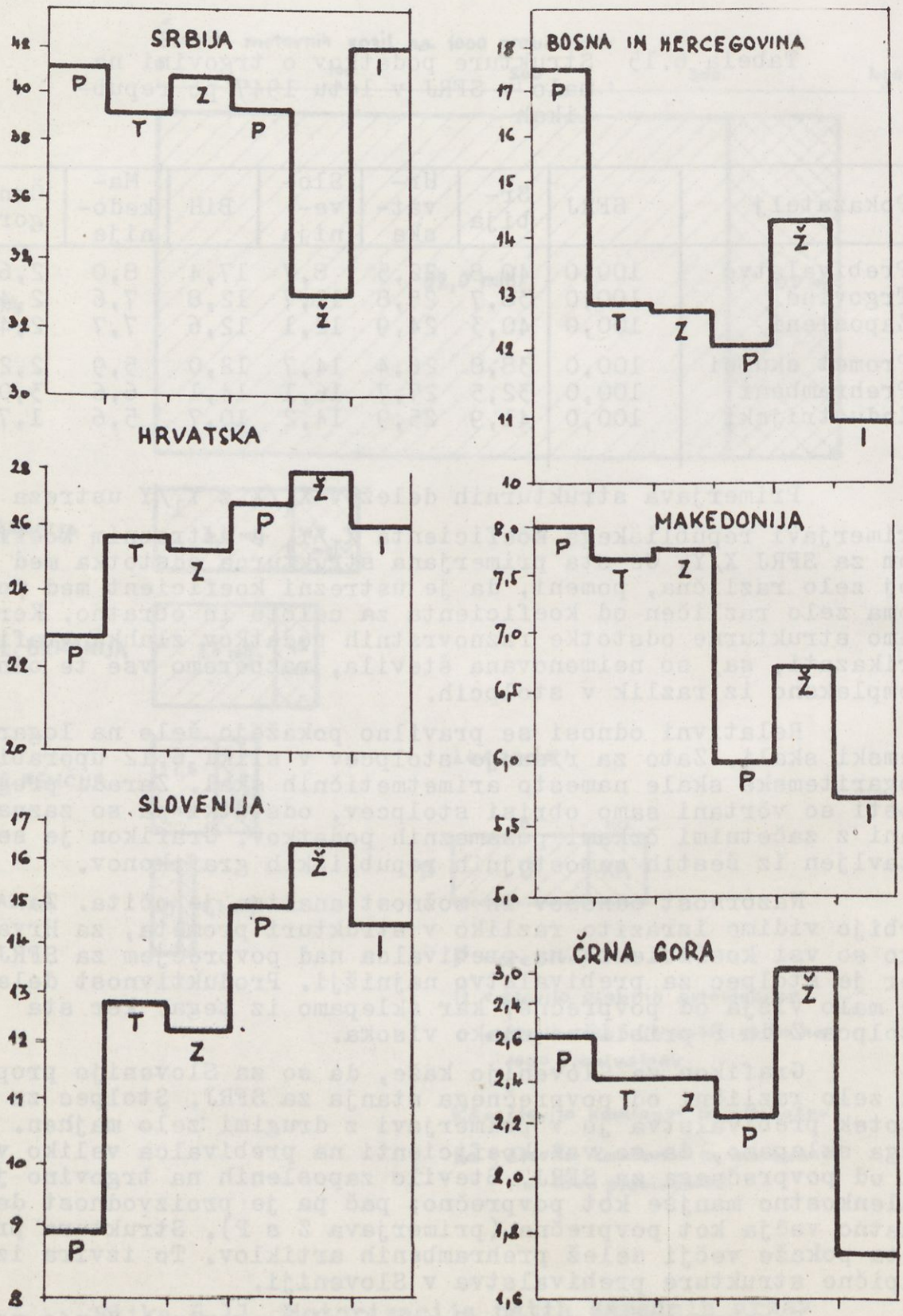
Primerjava strukturnih deležev X_1/X z Y_1/Y ustreza primerjavi republiškega koeficienta X_1/Y_1 z ustreznim koeficientom za SFRJ X/Y. Če sta primerjana strukturna odstotka med seboj zelo različna, pomeni, da je ustrezni koeficient med podatkoma zelo različen od koeficienta za celoto in obratno. Ker moremo strukturne odstotke raznovrstnih podatkov zlahka 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 6.12 uporabimo logaritemske skale namesto aritmetičnih skal. Zaradi preglednosti so včrtani 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 SFRJ, 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 SFRJ. 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 SFRJ. Število zaposlenih na trgovino je malenkostno manjše kot povprečno; pač pa je proizvodnost 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 podpovčrena (stolpec promet P manjši kot zaposleni Z).



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

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

Črna gora kaže precej podobno sliko o trgovinski mreži kot Srbija, ker so stolpci za prve štiri podatke (prebivalsto P, število trgovin T, število zaposlenih Z in skupni promet P) zelo podobni. Pač pa je očitna 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

$$I_{1/o} = 100 \cdot Y_1/Y_o \qquad (6.9)$$

Pri tem pomeni: Y_1 = podatek, ki ga primerjamo s podatkom Y_o. Y_o = podatek, na katerega primerjamo. Podatek, na katerega primerjamo, imenujemo bazo ali osnovo indeksa. $I_{1/o}$ = indeks. Z indeksom izražamo tekoči podatek, v stotinkah od baze Y_o. 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 11505 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 Jugo-

slavijo niso med seboj neposredno primerljivi, ker so eni dani v dinarjih, drugi pa v lirah. Če pa izračunamo indekse, dobimo, da je za Jugoslavijo indeks enak

$$I_{56/55} = 100 \cdot Y_{56}/Y_{55} = \frac{100 \cdot 1444,1}{1298,3} = 111,2$$

za Italijo pa

$$I_{56/55} = \frac{100 \cdot 11504}{10814} = 106,4$$

Indeksa pa sta primerljiva in sicer kažeta, da je bil porast v narodnem 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 1956), 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
Skopje	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.

npr.: za Beograd 100 . 287/170 = 169.

6.27 Za primer indeksov iz stvarnih vrst vzemimo skupne poprečne mesečne dohodke za tekstilno, prehrambeno in grafično industrijo v SFRJ v I. trimesečju 1958 po kvalifikacijah.

Tabela 6.17 Povprečni skupni mesečni prejemki delavcev po kvalifikacijah v I. trimesečju 1958 v SFRJ v dinarjih (Vir: SB 1923)

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. trimesečju 1958 v SFRJ. (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,
npr. za visoko kvalificirane delavce v tekst. stroki

$$I_{1/0} = 100 \cdot 20070/10910 = 184$$

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. Časovna vrsta 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 SFRJ in njenih sosedov v razdobju 1951-1957.

Tabela 6.19 Proizvodnja elektroenergije SFRJ in sosednih držav 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	1596
Italija	29223	30843	32619	35574	38124	40611
Jugoslavija	2550	2700	2982	3440	4340	5084
Madžarska	3506	4197	4615	4824	5428	5194
Romunija	2473	2886	3411	3697	4340	4930

V tabeli 6.19 težko proučimo in primerjamo dinamiko proizvodnje elektroenergije med državami. Posamezne države imajo zelo različno proizvodnjo, to pa zamegli 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 SFRJ in sosednih državah v razdobju 1951-1956 (baza 1951 = 100)

Država	Leto					
	1951	1952	1953	1954	1955	1956
Avstrija	100	109	119	133	146	159
Bolgarija	100	133	153	170	204	236
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:

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

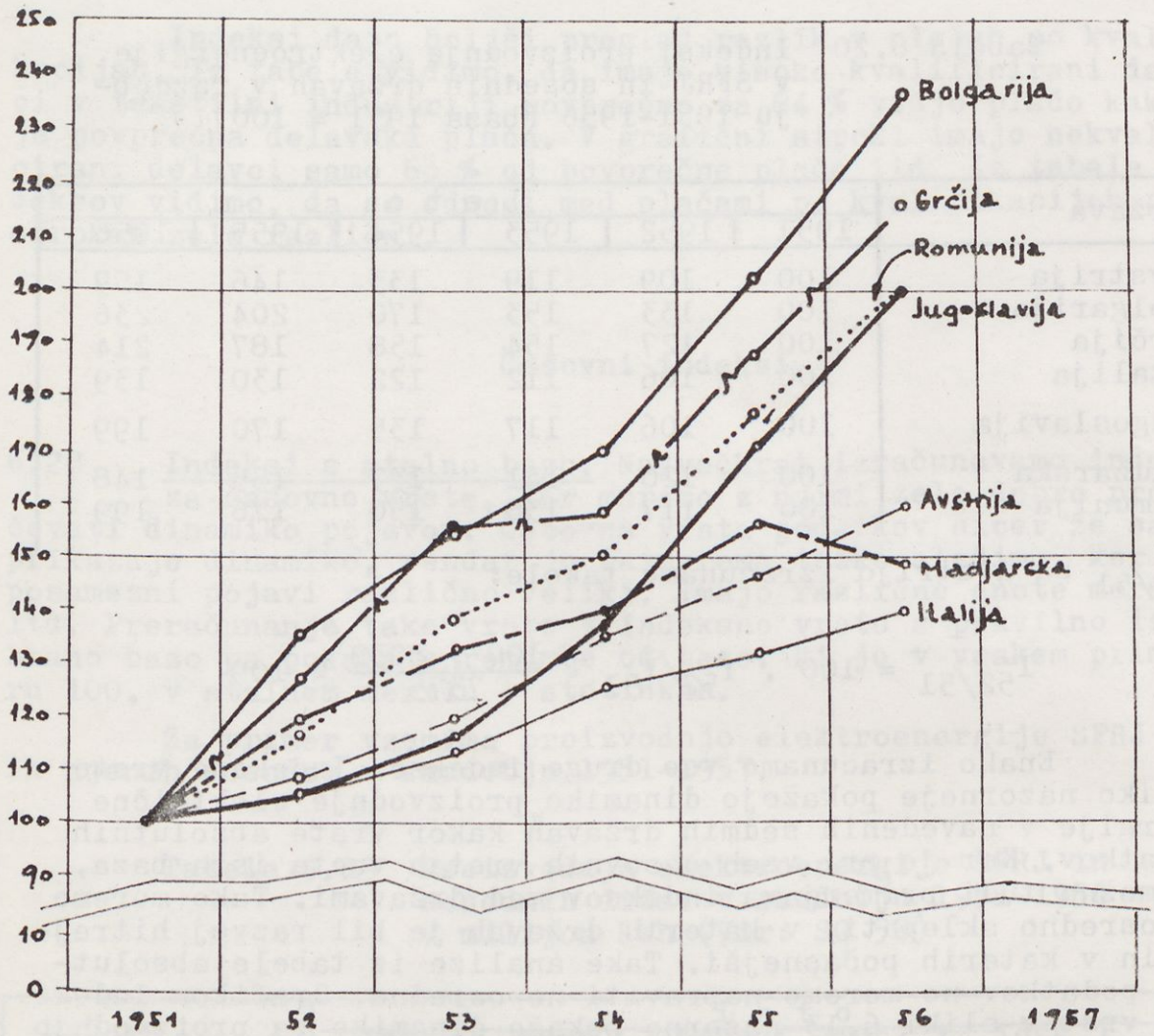
Enako izračunamo vse druge indekse. Indeksne vrste veliko nazorneje pokažejo dinamiko proizvodnje električne energije v navedenih sedmih državah kakor vrste absolutnih podatkov. Ker je pri vseh časovnih vrstah vzeta ista baza, je možna tudi primerjava indeksov med državami. Tako moremo neposredno sklepati, 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.

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

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

Pri tem pomeni: Y_k = tekoči podatek, Y_{k-1} = podatek predhodnega člena, I_k = verižni indeks.



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

Če vzamemo iz tabele 6.19 podatke o proizvodnji električne energije v Jugoslaviji v letih 1951-1956, je vrsta verižnih indeksov takale:

Tabela 6.21 Verižni indeksi za proizvodnjo električne energije v SFRJ 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 SFRJ od leta do leta večje, z izjemo v letu 1956.

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

$$I_{2/1} = 100 \frac{I_{2/o}}{I_{1/o}} \qquad (6.11)$$

Velja namreč:

$$100 \cdot \frac{I_{2/o}}{I_{1/o}} = 100 \cdot \frac{100.Y_2/Y_o}{100.Y_1/Y_o} = 100.Y_2/Y_1 = I_{2/1}$$

6.31 Vzemimo kot primer iz tabele 6.16 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 176)

Tabela 6.22 Indeksi koeficientov razvezanih zakonov na 1000 sklenjenih zakonov za glavni mesti republik

	Mesto	Beograd	Zagreb	Ljubljana	Sarajevo	Skopje	Titograd
Osnova	Ljubljana = 100	169	176	100	118	101	68
	Zagreb = 100	95	100	57	67	57	39

Za Beograd smo indeks z novo bazo izračunali:

$$100 \frac{169}{176} = 94$$

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 6.20 novo indeksno vrsto z bazo 1956 = 100, tako da vsak člen indeksne vrste delimo z indeksom za leto 1956 (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

L e t o	1951	1952	1953	1954	1955	1956
Baza 1951 = 100	100	106	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 katerikoli 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 enotno 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 npr. za bazo Slovenijo, v meddržavni primerjavi SFRJ 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 člene 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, npr. mesece pri indeksih proizvodnje, marveč vzamemo za bazo leto. Težko je tudi npr. reči v kmetijstvu, katero leto je normalno, ali je to leto ugodne ali neugodne vegetacije. Zato v kmetijstvu dostikrat vzamemo za 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 SFRJ vzamemo za primerjavo s predvojnim stanjem za bazo leto 1939, OZN pa leto 1937. Vendar moramo paziti, kdaj so primerjave pred in povojnega stanja vsebinsko utemeljene in smiselne. Za zelo dolga razdobja in v primeru velikih sprememb, primerjava na določeno staro stanje nima smisla.

Če je bila nekaj let po drugi svetovni vojni primerjava s predvojnim stanjem upravičena, ker smo tako približno dobili vtis o pojavih in spremembah, je primerjava s predvojnim stanjem sedaj nepomembna, ker nas zanima povojni razvoj, ne pa primerjava s predvojnim stanjem, ki je že razmeroma zelo odmaknjeno. Vendar ne vzamemo za bazo povojnega razvoja prvo povojno leto, ko 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 npr. 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

F R E K V E N Č N E D I S T R I B U C I J E

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	19,6	18,9	12,9	32,4	18,3	15,2	11,3	8,3
4,7	15,6	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 npr. 3 porabo od 3,0 m^3 do 3,9 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 pregledneje porabo lesa v gospodarstvih kakor individualni podatki v tabeli 7.1. Vendar frekvenčna distribucija ne da natančnih vrednosti. Iz nje npr. vidimo, da je eno gospodarstvo porabilo lesa od 2,0 do 2,9 m3, ne vemo pa, koliko so natančno porabili lesa v tem gospodarstvu. Enako vidimo, da je 14 gospodarstev porabilo 10,0 do 10,9 m3, 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 m3	x	0	1	2	3	4	5	6	7	8	9	10	11	12
Število gospodarstev	f	-	-	1	1	2	3	6	5	8	7	14	7	14

x	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
f	7	10	10	8	7	6	8	3	3	2	4	3	-	3	1	1	2	-

x	31	32	33
f	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 npr. porabili v kmetijskih gospodarstvih od 2 m3 do 34 m3 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 po-

večati 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 m3, 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

Seveda pa ne smejo biti razredi preširoki. S širjenjem razredov se natančnost bolj in bolj manjša. Tudi zakonisto 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 m3 razlik v porabi.

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 dado 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 od šestnajst razredov, dado običajno dosti dobro sliko gostitve. Število razredov se ravna po velikosti populacije. Za manjše populacije vzamemo manj, za večje pa več razredov. V našem primeru smo resnično dobili s štiridesetimi razredi nepregledno, s petimi razredi pregrobo sliko o variiranju 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 SFRJ 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 postopici; 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_o delimo z obsegom populacije N, imenujemo relativne frekvence f_k^o.

$$f_k^o = f_k/N \qquad (7.2)$$

Kakor smo iz frekvenc izračunali gostoto frekvence, tako moremo po obrazcu

$$\varphi_k = f_k^o/i_k = f_k/N \cdot i_k \qquad (7.3)$$

izračunati tudi gostoto relativne frekvence φ_k, če relativno frekvenco delimo s širino razreda.

Iz obrazca 7.3 zlahka dobimo, da je

$$f_k = N \cdot i_k \cdot \varphi_k \qquad (7.4)$$

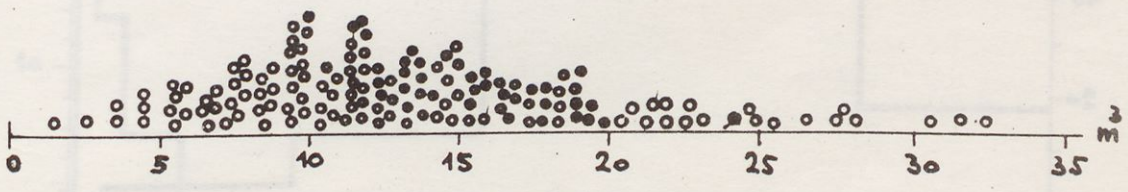
Frekvenca v danem razredu je torej premo sorazmerna z obsegom populacije N, s širino razreda i_k in gostoto relativne frekvence φ_k, ki je povezana z vsebino pojava.

7.4 Za frekvenčno distribucijo industrijskih podjetij v SFRJ konec leta 1957 so zgornje količine izračunane v tabeli 7.6.

Tabela 7.6 Industrijska podjetja v SFRJ konec leta 1957 po številu zaposlenih

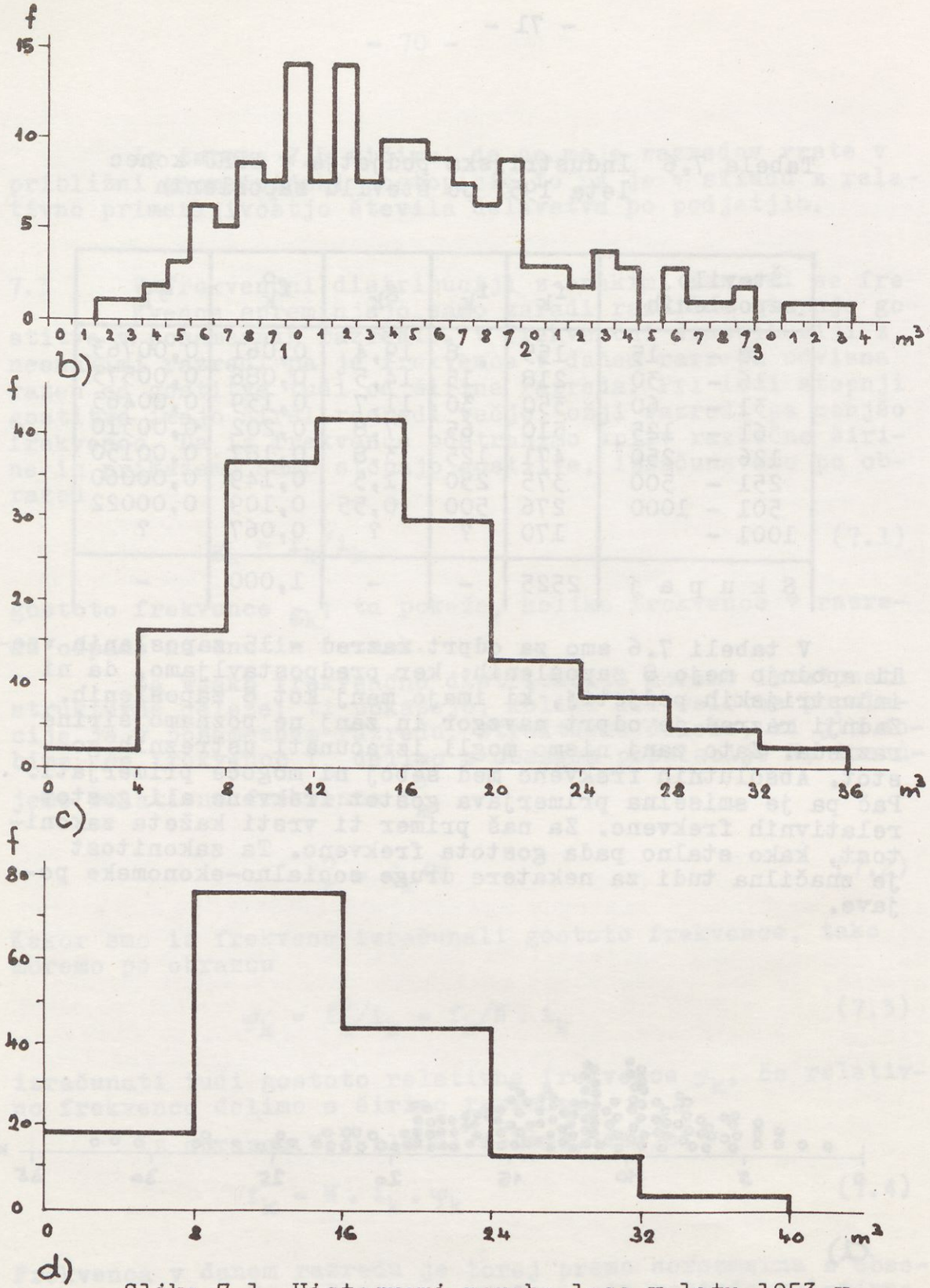
Število zaposlenih	f_k	i_k	g_k	f_k^o	φ_k
8 - 15	155	8	19,4	0,061	0,00763
16 - 30	218	15	14,5	0,086	0,00573
31 - 60	350	30	11,7	0,139	0,00463
61 - 125	510	65	7,8	0,202	0,00310
126 - 250	471	125	3,8	0,187	0,00150
251 - 500	375	250	1,5	0,149	0,00060
501 - 1000	276	500	0,55	0,109	0,00022
1001 -	170	?	?	0,067	?
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.



a)

Slika 7.1 (glej nadaljevanje na naslednji strani)



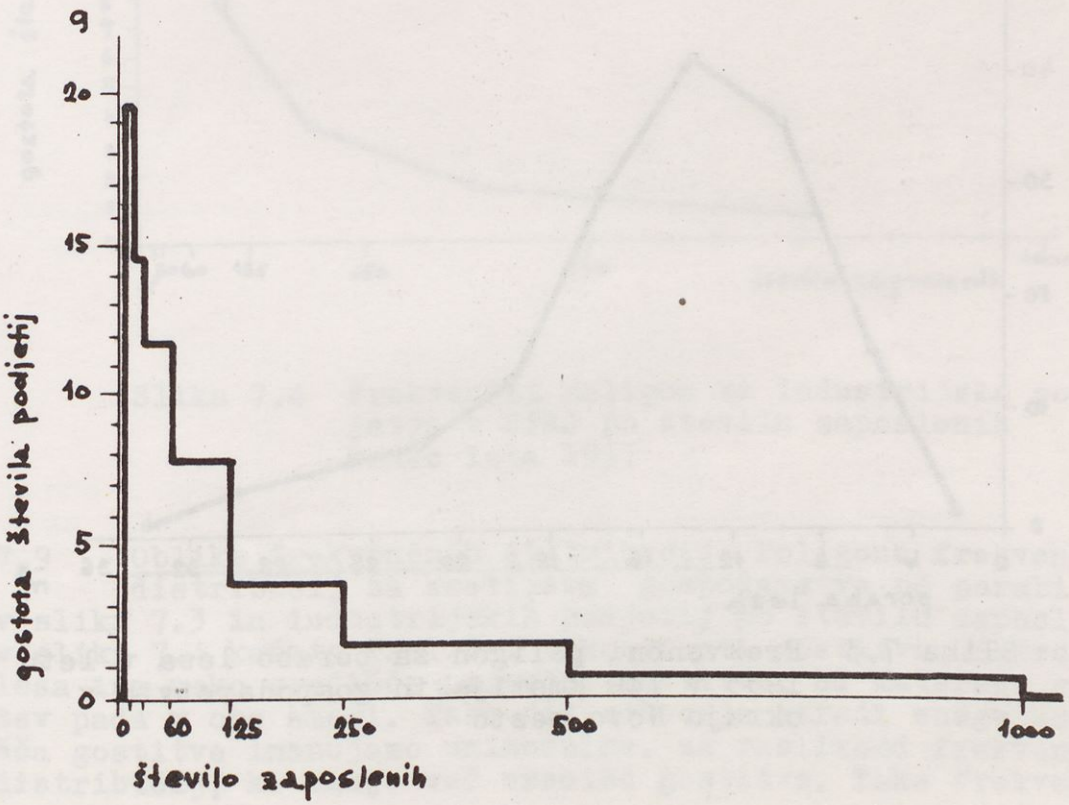
Slika 7.1 Histogrami porabe lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

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 m3, v drugi 4 m3, v tretji pa 8 m3. Iz histogramov še nazorneje kot iz tabel vidimo, da najbolj ustreza širina razreda 4 m3.

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



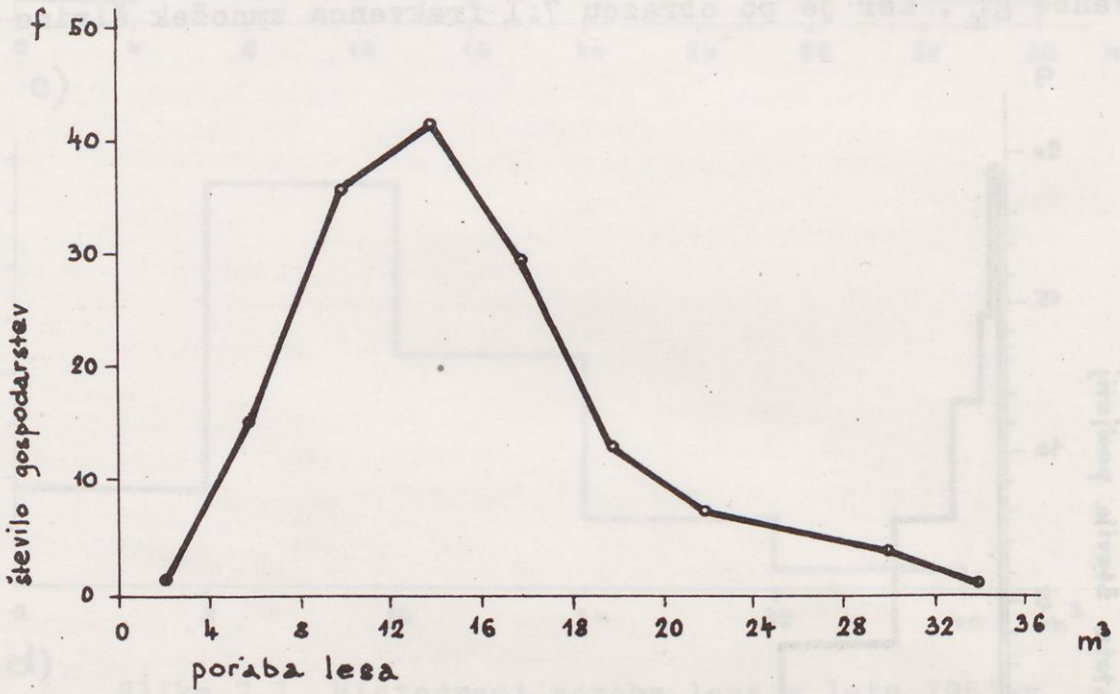
Slika 7.2 Histogram industrijskih podjetij po številu zaposlenih v SFRJ konec leta 1957

razreda 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 SFRJ konec leta 1957 po številu delavcev iz tabele 7.6. Histogram pravilno pokaže, da gostitev podjetij pada, če se število podjetij veča.

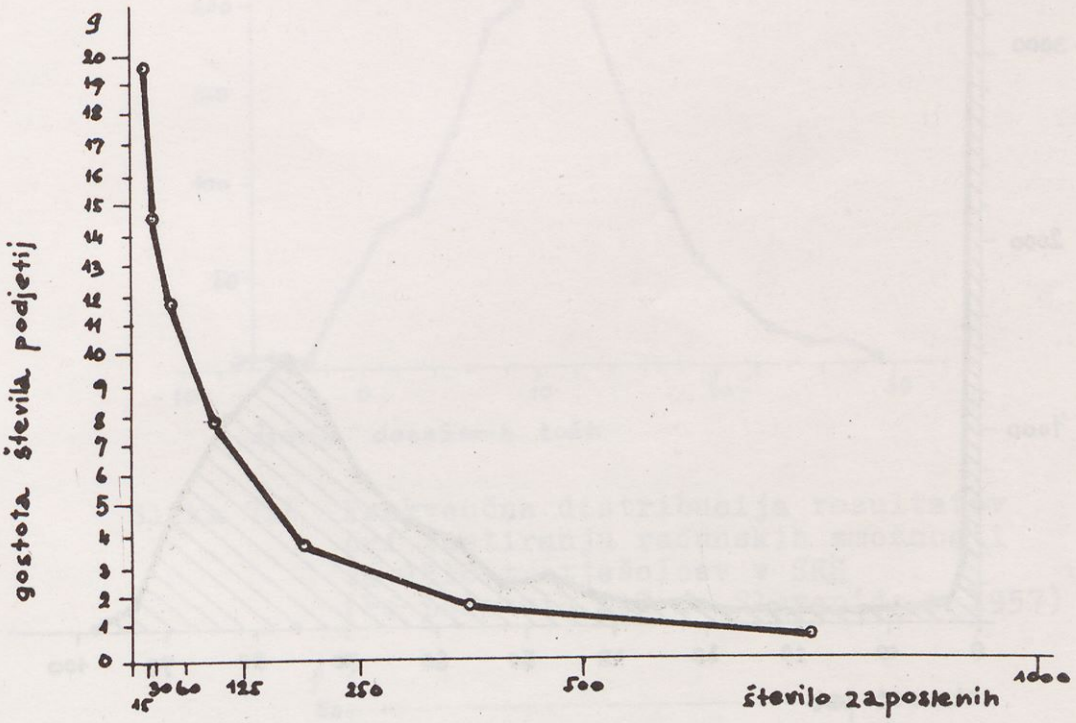
7.7 Poligon. Frekvenčni poligon za vrste z enakimi razredi narišemo tako, da nad sredine razmakov za posamezne razrede nanesemo točke, ki so od abscisne osi oddaljene v sorazmerju z velikostjo frekvence. Dobljene točke po vrsti zvežemo z daljicami. Če vzamemo, da je frekvenca v razredu pred prvim razredom 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

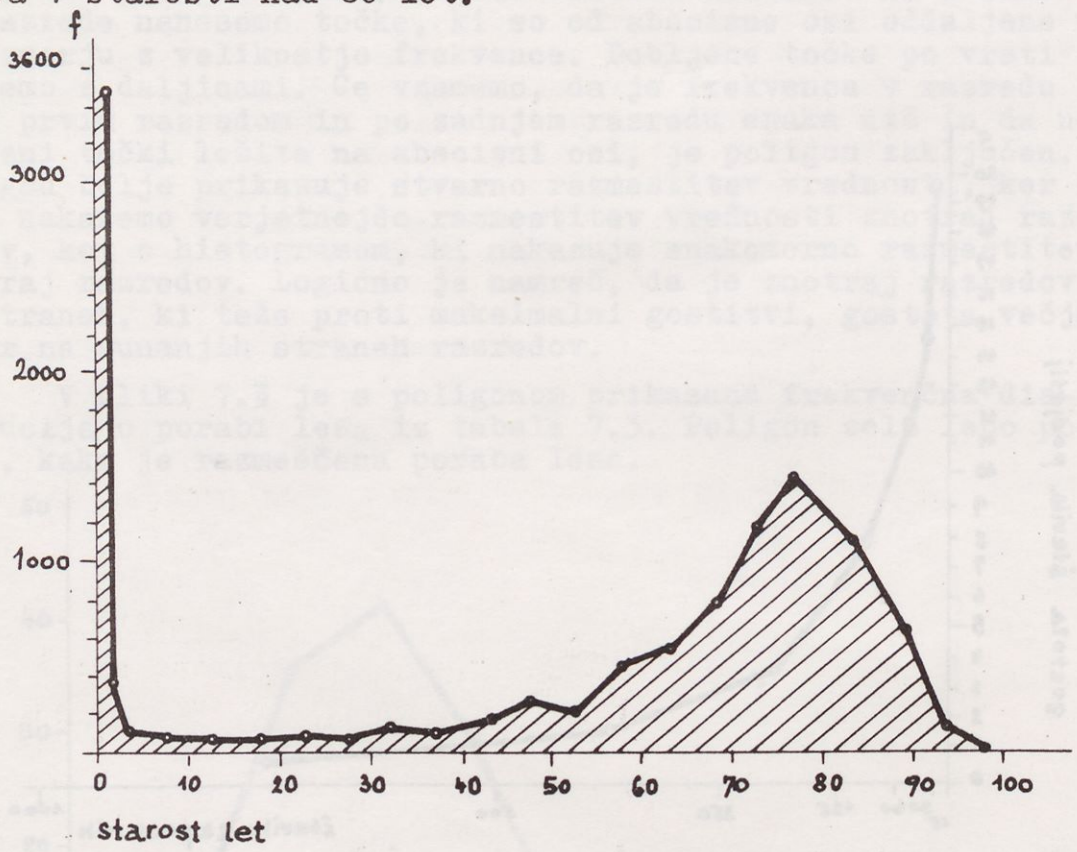
7.8 Frekvenčno distribucijo z različnimi širinami razredov pa prikažemo s poligonom tako, da nad sredine razredov, ki so risani v sorazmerju s širinami razredov, nanesemo točke, ki so od abscise oddaljene v sorazmerju z gostoto frekvence. Če točke med seboj povežemo, dobimo lomljeno črto, ki zelo dobro ponazarja razmestitev vrednosti. Frekvenčna distribucija industrijskih podjetij v SFRJ 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 SFRJ 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 dobimo, 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 1956 v Sloveniji. V grafikonu opazimo izredno veliko število umrlih otrok do enega leta. Število umrlih se zaradi majhne umrljivosti v nadaljnjih letih znatno zmanjša. V starejših letih se zaradi povečane umrljivosti 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.

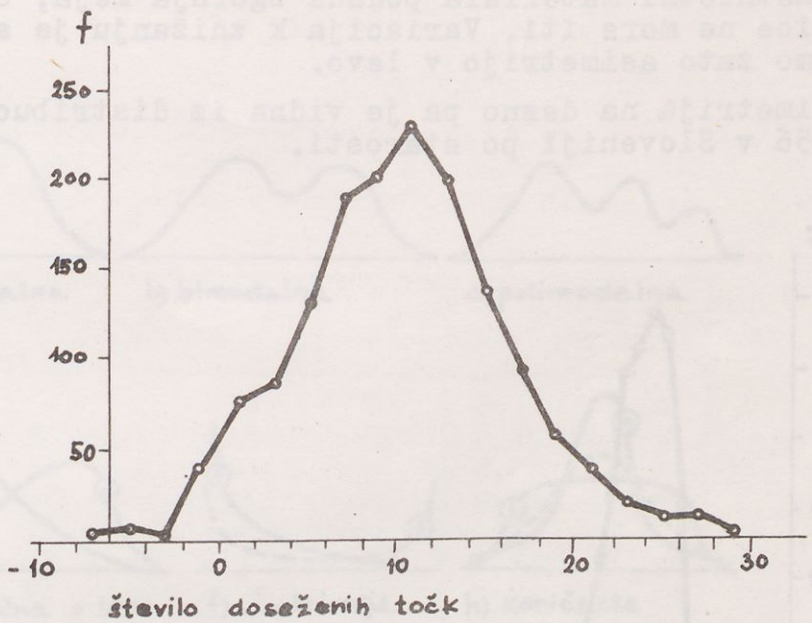


Slika 7.5 Število umrlih žensk v letu 1956 v Sloveniji po starosti ob smrti

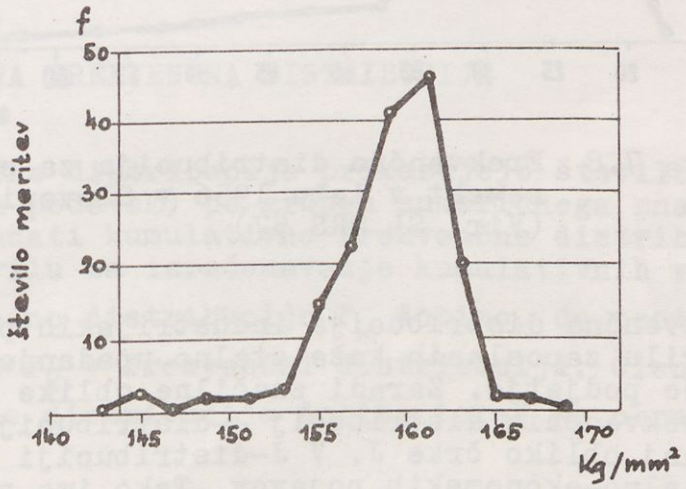
7.10 Poligon frekvenčne distribucije porabe lesa v kmetijskih gospodarstvih kaže bolj ali manj enako upadanje gostitve od središča gostitve. Frekvenčne distribucije, ki kažejo enako upadanje frekvenc na levo in desno od središča gostitve, imenujemo simetrične distribucije, za razliko od asimetričnih 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, govo-

rimo 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.



Slika 7.6 Frekvenčna distribucija rezultatov pri testiranju računskih zmožnosti za 1518 tretješolcev v SRS (Po podatkih DAT za Slovenijo v 1957)



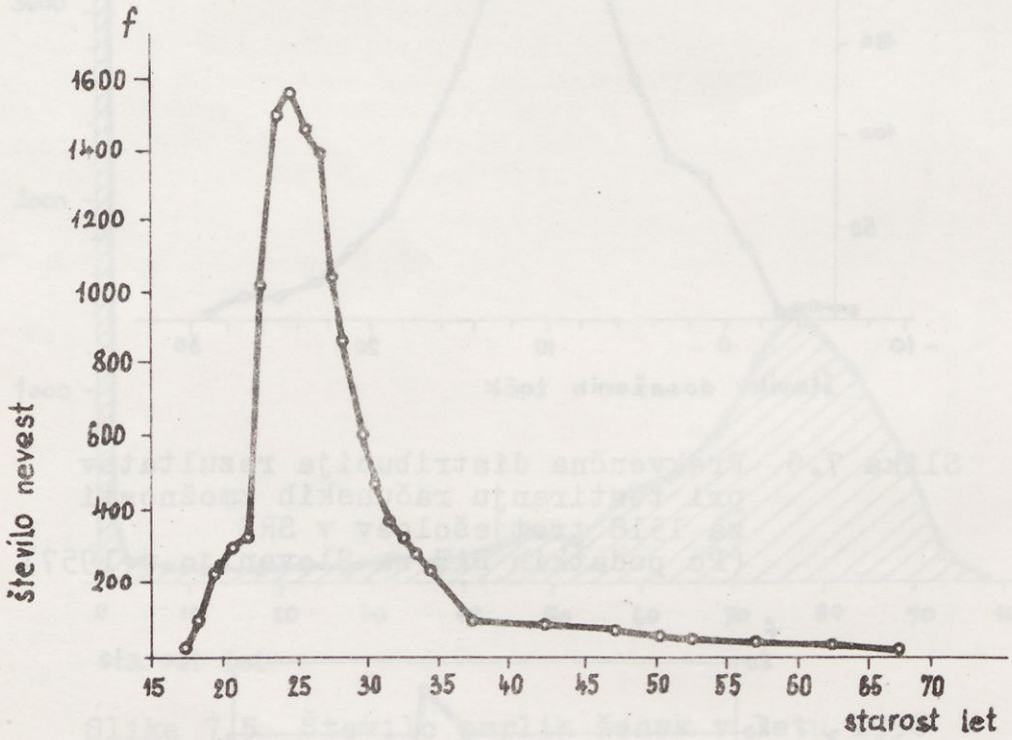
Slika 7.7 Frekvenčna distribucija o trdnosti za patentirano 5-mm jeseniško žico v kg/mm^2

(Po podatkih Zavoda za raziskavo materiala in konstrukcij SRS)

Normalni distribuciji je precej podobna frekvenčna distribucija rezultatov testiranja računskih zmožnosti za 1518 tretješolcev v SRS v letu 1957.

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 153 meritev trdnosti 5-mm jeseniške 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 znižanju je svobodnejša in imamo zato asimetrijo v levo.

Asimetrija na desno pa je vidna iz distribucije nevest v letu 1956 v Sloveniji po starosti.

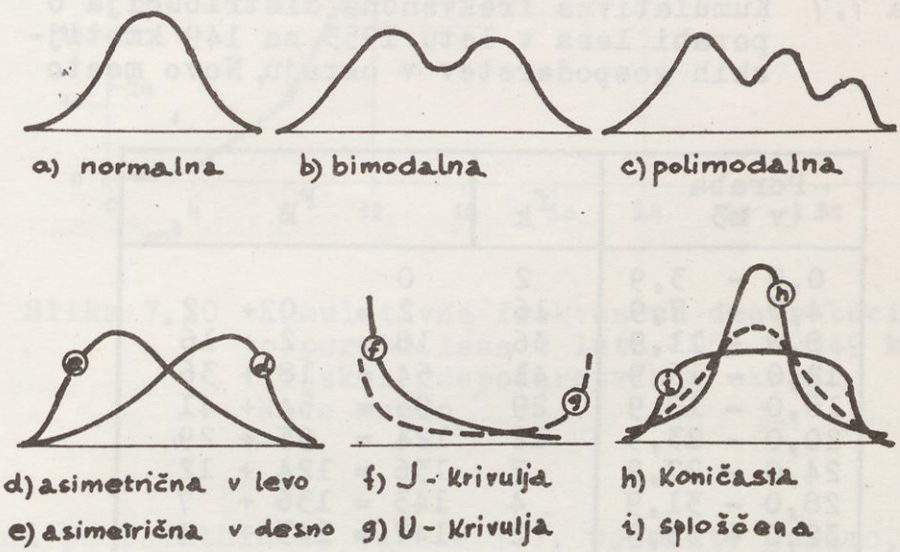


Slika 7.8 Frekvenčna distribucija za starost nevest v letu 1956 v Sloveniji (Vir: SL SRS 57)

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 bolj ali manj obliko črke J. V J-distribuciji se distribuira veliko socialno-ekonomskih pojavov. Tako ima podobno obliko frekvenčna distribucija za kmetijska gospodarstva v SRS.

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 primerjavo 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 podatek) 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 členi v kumulativni frekvenčni distribuciji F_k pomenijo, koliko enot ima vrednosti pod stopnjo mejo ustreznega razreda $x_{k,min}$. Tako pomeni npr. šesti člen kumulativne vrste, da je $F_6 = 124$ gospodarstev, ki so v letu 1953 porabila manj lesa kakor $x_{6,min} = 20$ m3 lesa.

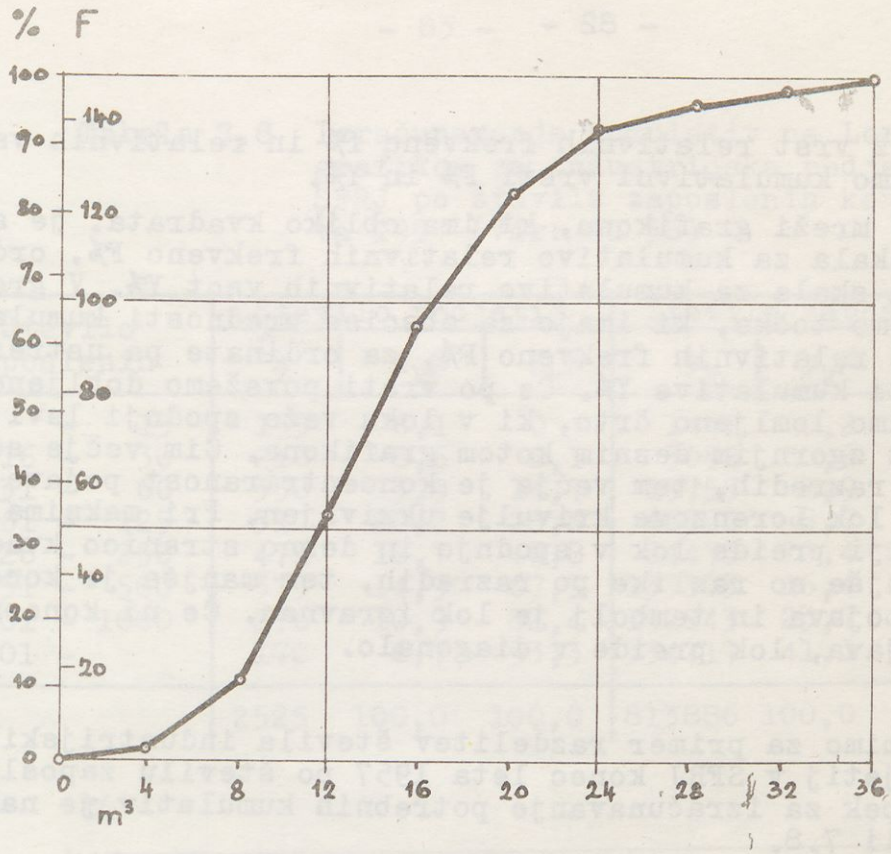
Tabela 7.7 Kumulativna frekvenčna distribucija o porabi lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m3	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 uporabljamo 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).

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



Slika 7.10 Kumulativna frekvenčna distribucija o porabi lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

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.

L o r e n z o v g r a f i k o n

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 SFRJ konec leta 1957 po številu zaposlenih. Postopek za izračunavanje potrebnih kumulativ je nakazan v tabeli 7.8.

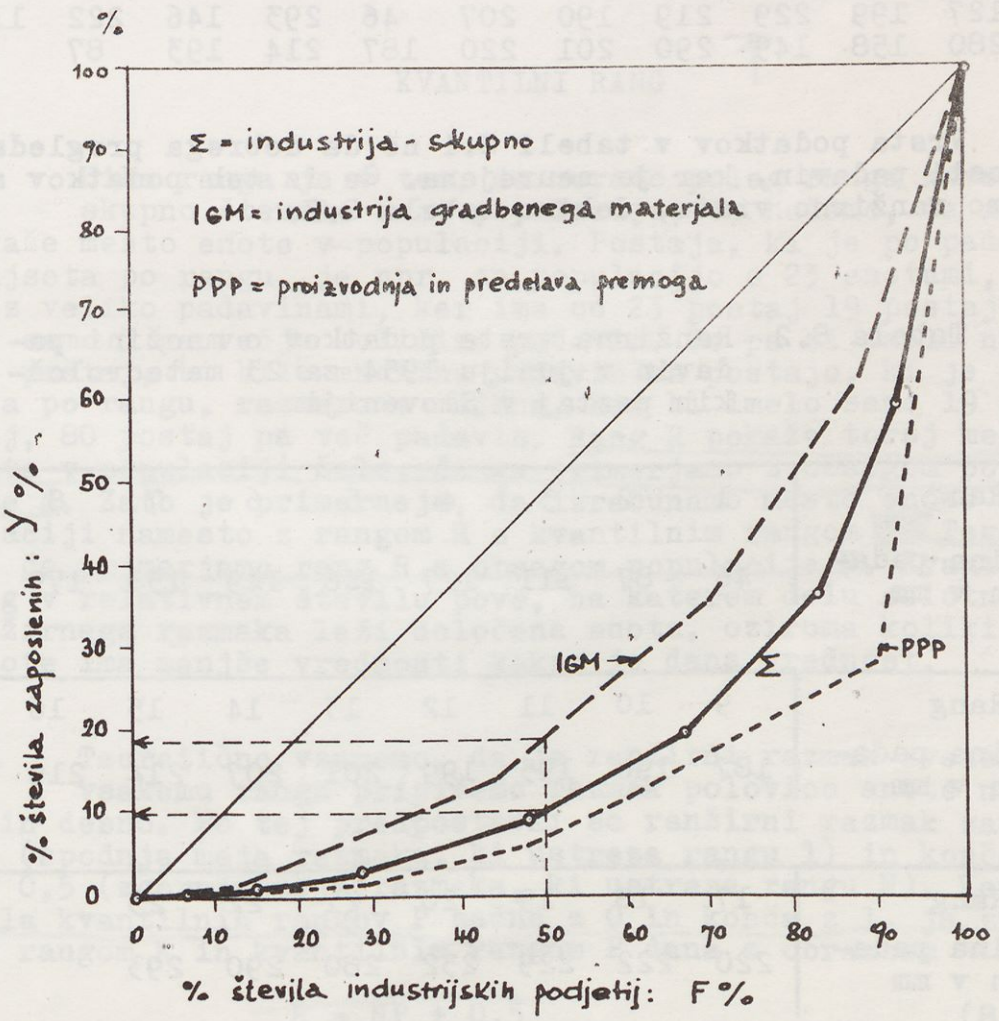
V tabeli 7.8 izračunani podatki (F% in Y%) so vneseni v Lorenzov grafikon v sliki 7.11.

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 podjetju 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 npr. polovica (manjša podjetja) podjetij v SFRJ 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 včrtati Lorenzove krivulje iz različno grupiranih podatkov. To posebno olajša mednarodne primerjave, ker ima vsaka država grupacije prilagojene svojim pogojem in potrebam.

Tabela 7.8 Izračunavanje kumulativ na Lorenzov grafikon za industrijska podjetja v SFRJ 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	28,6	46300	5,7	2,7
126 - 250	471	18,7	48,8	84475	10,4	8,4
251 - 500	375	14,9	67,5	133303	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 SFRJ za vsa industrijska podjetja (Σ), za industrijo gradbenega materiala (IGM) in proizvodnjo in predelavo premoga (PPP)

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: SG LRS 1956)

179	127	199	229	219	190	207	46	293	146	222	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
Množina padavin v mm	46	87	111	127	146	149	158	179

Rang	9	10	11	12	13	14	15	16
Množina padavin v mm	187	190	193	199	201	207	214	219

Rang	17	18	19	20	21	22	23
Množina padavin v mm	220	222	229	252	280	290	293

Iz ranžirne vrste pa moremo napraviti več zaključkov o prikazanem pojavu. Iz nje podobno kakor iz frekvenčne distribucije neposredno vidimo, v kakšnih mejah variira populacija, ker kaže podatek, ki je prvi po rangu, najmanjšo (46 mm), podatek, ki je poslednji po rangu, pa največjo množino padavin (293 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 določeno meteorološko postajo vemo, da je v juniju 1953 izkazala 252 mm padavin, 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 23 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 npr. 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 npr. 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 primera vidimo, da kvantilni rang P sam zase - ne da bi navajali obseg populacije - nazorno prikaže mesto določene enote v populaciji.

K V A N T I L I

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 npr. 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 npr. 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

$$Q_1 = x_{0,25};\ Q_2 = x_{0,50};\ Q_3 = x_{0,75} \qquad (8.3)$$

Analogno z decili

$$D_1 = x_{o,1o},\quad D_2 = x_{o,2o} \ldots\ldots D_9 = x_{o,9o} \qquad (8.4)$$

razdelimo celotno populacijo v deset po obsegu enakih delov s centili:

$$C_1 = x_{o,o1},\ C_2 = x_{o,o2} \ldots\ldots C_{98} = x_{o,98} \text{ in } C_{99} = x_{o,99} \qquad (8.5)$$

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 npr. 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_o in x_1 pade vrednost x; za katero iščemo P_x, tako da velja: $x_o < x < x_1$. Vrednosti x_o naj ustreza rang R_o.

c) Rang R_x, ki ustreza vrednosti x, dobimo z linearno interpolacijo po obrazcu

$$R_x = R_o + \frac{x - x_o}{x_1 - x_o} \qquad (8.6)$$

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 npr. padavinam x = 150 mm, dobimo po zgornjem postopku

x = 150 leži v ranžirni vrsti med vrednostima $x_o = 149$, za katero je rang $R_o = 6$, 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: $P_x = \frac{6{,}11 - 0{,}50}{23} = 0{,}244$

Kvantilni rang kraja, ki je imel v juniju 1953 $x = 150$ 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: $R_o < R_p < R_1$. Rangoma R_o in R_1 ustrezata vrednosti x_o in x_1.

d) Iz teh podatkov izračunamo z linearno interpolacijo x_p po obrazcu:

$$x_p = x_o + (x_1 - x) \cdot (R_p - R_o) \qquad (8.9)$$

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 juniju 1953 v Sloveniji

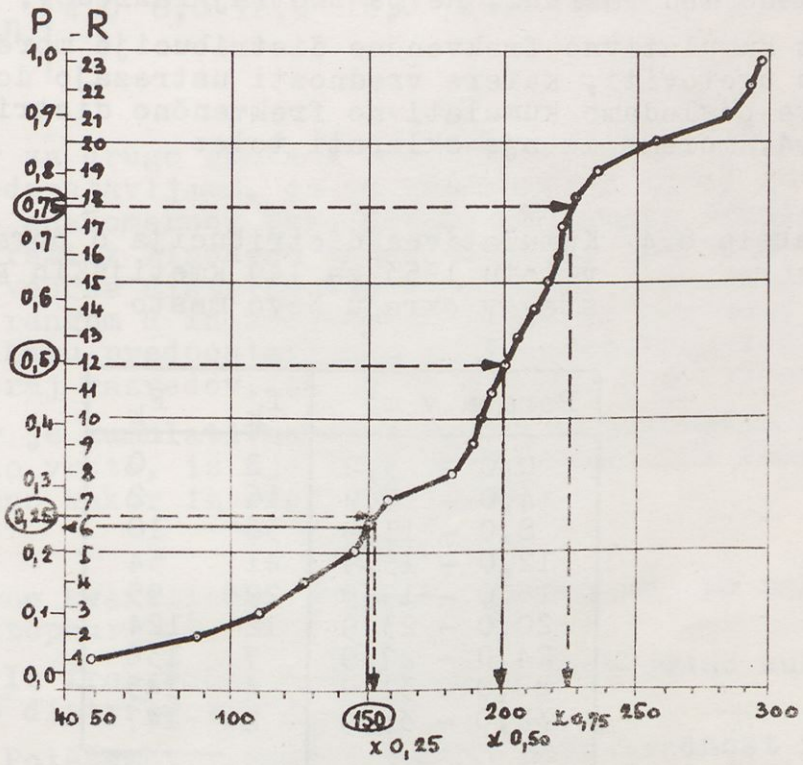
i	P	R_p	R_o	x_o	x_1	$x_1 - x_o$	$(x_1 - x_o)(R_p - R_o)$	$x_p = Q_i$
1	0,25	6,25	6	149	158	9	2,25	$151{,}25 = Q_1$
2	0,50	12,00	12	199	199	0	0	$199{,}00 = Q_2$
3	0,75	17,75	17	220	222	2	1,50	$221{,}50 = Q_3$

Za prvi kvartil izračunamo rang $R = 23 \cdot 0{,}25 + 0{,}5 = 6{,}25$ 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, polovica 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 vidno, 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 23 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 členi kumulativne vrste, spoznamo, da je kumulativna frekvenčna distribucija v zvezi z rangi.

Proučimo frekvenčno distribucijo o porabi lesa v 149 kmetijskih gospodarstvih v Novem mestu iz tabele 7.3. 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 razredih. 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 m3	f_k	F_k
0,0 - 3,9	2	0
4,0 - 7,9	16	2
8,0 - 11,9	36	18
12,0 - 15,9	41	54
16,0 - 19,9	29	95
20,0 - 23,9	12	124
24,0 - 27,9	7	136
28,0 - 31,9	4	143
32,0 - 35,9	2	147
		149 = N

Dvoje kmetijskih gospodarstev je porabilo manj kot 4,0 m3 lesa. Če vzamemo, da ima gospodarstvo z največjo porabo v tem razredu vpribližku porabo enako tej meji, je gospodarstvo, ki ima rang 2, porabilo 4,0 m3. Iz kumulative dalje izvemo, da je 18 gospodarstev porabilo pod 8,0 m3. Če

vzamemo, da je največja poraba v teh 18 gospodarstvih enaka tej meji, dobimo, da ima gospodarstvo, ki je porabilo 8,0 m3 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 tabeli 8.4)

Rang R F_k	2	18	54	95	124	136	143	147	149
$x_p = x_{k,min}$	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

$$R_x = F_o + f_o \frac{x - x_{o,min}}{i_o} \qquad (8.1o)$$

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 m3 lesa, izračunamo P_x po zgornjem postopku takole:

Poraba lesa x = 13,8 m3 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:

$$P_{x=13,8} = \frac{72,4 - 0,5}{149} = 0,482$$

Gospodarstvo s porabo x = 13,8 m3 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, tako da je: $F_o < R_p < F_1$. F_o ustrezen razred je kvantilni razred. Zanj poiščemo v frekvenčni distribuciji količine: $x_{o,min}$, i_o, f_o in F_o.

d) Iz teh količin izračunamo ustrezni kvantil x_p po obrazcu:

$$x_p = x_{o,min} + i_o \frac{R_p - F_o}{f_o} \qquad (8.13)$$

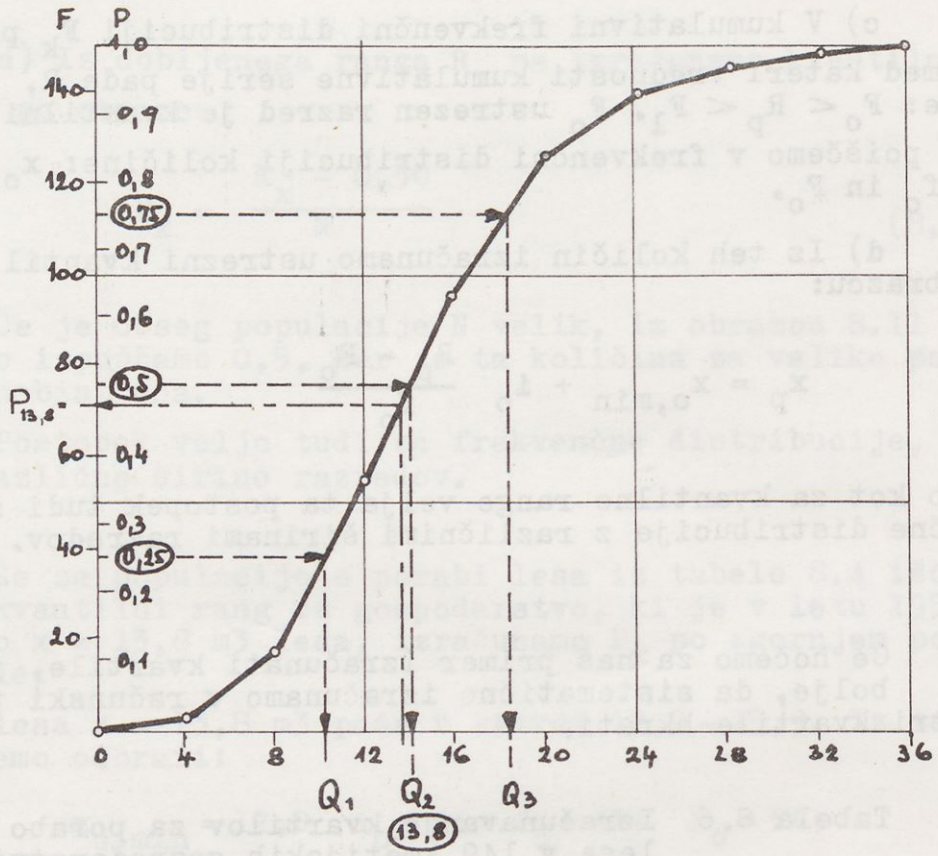
Enako kot za kvantilne range velja ta postopek tudi za frekvenčne distribucije z različnimi širinami razredov.

8.12 Če hočemo 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	$x_{o,min}$	i_o	f_o	F_o	$R_p - F_o$	$i_o \frac{R_p - F_o}{f_o}$	Q_i
1	0,25	37,75	8,0	4	36	18	19,75	2,19	$10,19 = Q_1$
2	0,50	75,00	12,0	4	41	54	21,00	2,05	$14,05 = Q_2$
3	0,75	112,25	16,0	4	29	95	17,25	2,38	$18,38 = Q_3$

8.13 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 m3 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 z decili oziroma, pri podrobnejših normnih tablicah, s centili. Centilne norme s pridom uporabljamo v psihologiji, moremo pa jih izdelati tudi za določene socialno-ekonomske probleme. Tako tablica decilnih ali centilnih norm o produktivnosti dela pomaga presoditi, kakšna je produktivnost dela za posameznega delavca v razmerju s celotnim kolektivom itd.

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.

Vrste srednjih vrednosti

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 dana z lego vrednosti, štejemo ostale tri: aritmetično, harmonično in geometrijsko sredino med izračunane srednje vrednosti.

M E D I A N A

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 1953 za 23 meteoroloških postaj v Sloveniji, enaka Me = 199 mm. Za porabo lesa v letu 1953 v kmetijskih gospodarstvih v Novem mestu pa smo iz frekvenčne distribucije o porabi lesa izračunali, da je mediana enaka Me = 14,05 m3 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 se njena 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.

M O D U S

9.5 Mediana pa ni vedno vrednost, ki dobro reprezentira vrednosti 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 1953 v Novem mestu iz tabele 7.3 sklepamo, da je najpogostejša vrednost v razredu od 12,0 do 16,0 m3 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 m3. 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 sosednih razredih. Iz histograma o porabi lesa v sliki 7.1 moremo sklepati, da je modus za porabo lesa verjetno pod sredino modalnega razreda s sredino 14 m3. 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 sosednemu raz-

redu 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_o, f_o) in (x_{+1}, f_{+1}). Pri tem pomenijo x_{-1}, x_o in x_{+1} sredine modalnega razreda in njegovih sosedov, f_{-1}, f_o in f_{+1} pa ustrezne frekvence. Pri tej predpostavki izračunamo modus po naslednjem postopku:

a) Podatke imamo grupirane v frekvenčni distribuciji z razredi enake širine i. Razredi morajo biti tako veliki, da frekvenčna distribucija izraža zakonitost gostitve frekvenc.

b) V frekvenčni distribuciji poiščemo razred 0, ki ima največjo frekvenco $f_{-1} < f_o > f_{+1}$. Razred z največjo frekvenco imenujemo modalni razred.

c) Iz modalne frekvence f_o in frekvenc obeh sosednih razredov f_{-1} in f_{+1} izračunamo $d_{-1} = f_o - f_{-1}$ in $d_{+1} = f_o - f_{+1}$.

d) Modus izračunamo po obrazcu

$$M_o = x_{o,min} + i \frac{d_{-1}}{d_{-1} + d_{+1}} \qquad (9.1)$$

Pri tem pomeni razen že navedenih izrazov: $x_{o,min}$ = 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 m3. 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 m3, ker ima največjo frekvenco ($f_o = 41$). Ker je za to distribucijo $f_{-1} = 36$, in $f_{+1} = 29$, velja dalje:

$$d_{-1} = f_o - f_{-1} = 41 - 36 = 5; \quad d_{+1} = f_o - f_{+1} = 41 - 29 = 12.$$

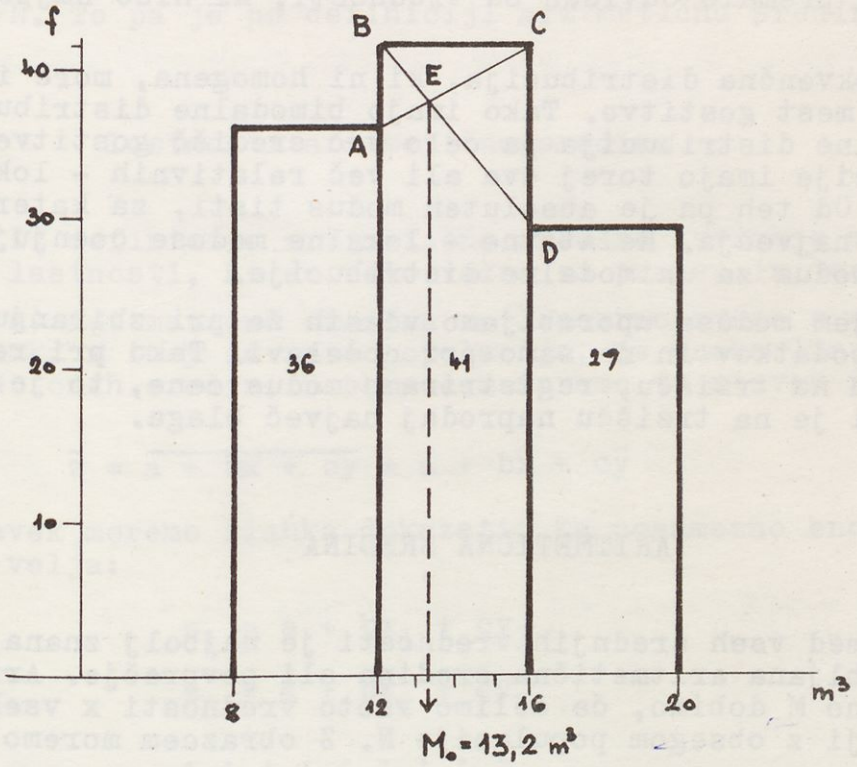
$$i = 4; \quad x_{o,min} = 12,0.$$

Iz teh podatkov je modus po obrazcu 9.1

$$M_o = 12,0 + 4 \cdot \frac{5}{5 + 12} = 13,17 \text{ m3}$$

Druga ocena modusa M_o = 13,17 m3 je stvarno pod prvo oceno 14,0 m3.

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 s C, oglišče B z D. Projekcija sečišča obeh veznic E je ocena modusa M_o. Slika pokaže, da grafičen način da v načem 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

L a s t n o s t i m o d u s a

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 modus za unimodalne distribucije.

Pojem modusa uporabljamo včasih že pri zbiranju statističnih podatkov in ne samo pri obdelavi. Tako pri registriranju cen na tržišču, registriramo modus cene, to je ceno, po kateri je na tržišču naprodaj največ blaga.

ARITMETIČNA SREDINA

9.10 Izmed vseh srednjih vrednosti je najbolj znana in uporabljana 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

$$M_x = \bar{x} = \frac{1}{N}(x_1 + x_2 + \ldots + x_N) = \frac{1}{N}\Sigma x = X/N \quad (9.2)$$

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

$$\bar{x} = a + b\bar{x} + c\bar{y}$$

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

$$\overline{bx} = b \cdot \bar{x} \qquad (9.5)$$

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 $\frac{dK}{dA} = 0$.

$$\frac{dK}{dA} = -2\Sigma(x - A) = 0\,; \quad \Sigma(x - A) = \Sigma x - NA = 0;$$

$$A = \frac{1}{N}\Sigma x$$

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.

d) 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 ocenjevati 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 m3	f_k	x_k	$f_k x_k$
0,0 - 3,9	2	2	4
4,0 - 7,9	16	6	96
8,0 - 11,9	36	10	360
12,0 - 15,9	41	14	574
16,0 - 19,9	29	18	522
20,0 - 23,9	12	22	264
24,0 - 27,9	7	26	182
28,0 - 31,9	4	30	120
32,0 - 35,9	2	34	68
	N = 149		$\Sigma f_k x_k$ = 2190

Po obrazcu 9.9 je $\bar{x}$ = 2190/149 = 14,70 m3.

Če primerjamo dobljeni rezultat s pravo aritmetično sredimo $\bar{x}$ = 2182,1/149 = 14,34 m3, 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:

$$x_k = x_o + i \cdot u_k \qquad (9.11)$$

Pri tem pomeni: x_o = sredina razreda, ki je približno v sredini frekvenčne distribucije ali blizu razreda z največjo frekvenco, i = širina razreda.

Če natančneje proučimo vrednosti znaka u, spoznamo, da sredinam razredov v posameznih razredih ustrezajo u ... -3, -2, -1,0, +1, +2, +3 ... Pri tem je vrednost u = 0 v razredu, ki mu ustreza x_o.

Zaradi obrazca 9.11 in stavka 9.3 pa velja

$$\bar{x} = x_o + i\bar{u} \qquad (9.12)$$

Ta postopek se običajno pokaže prikladnejši 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

$$\bar{x} = x_o + \frac{i}{N} \Sigma f_k u_k \qquad (9.13)$$

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 m3	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		Σfu = +26

Ker je x_o = 14,0, i = 4, N = 149 in Σfu = +26, 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 prikladnejši, kakor če vpeljemo pomožni 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

$$\bar{x} = x_o - i\, A/N \qquad (9.14)$$

Pri tem pomeni: x_o = 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 m3	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+...+143+147 = 719 = A.
	N = 149		

Ker je sredina zadnjega razreda $x_o = 34$, dobimo po obrazcu 9.14

$$\bar{x} = 34{,}0 - 4.719/149 = 14{,}70 \text{ m3}.$$

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 obeh 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

$$\bar{r} = \frac{\sum w_k r_k}{\sum w_k} \qquad (9.16)$$

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 1956.

Tabela 9.4 Izračunavanje skupnih povprečnih mesečnih prejemkov delavcev v poljedelstvu v SFRJ v septembru leta 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	353162870
Priučeni	9280	20515	190379200
Nekvalificirani	7700	78819	606906300
		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 hibam 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

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_N}} = \frac{N}{\frac{1}{x_i}} \qquad (9.17)$$

tehtana harmonična sredina pa analogno

$$H = \frac{w_1 + w_2 + \ldots\ldots + w_r}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots\ldots + \frac{w_r}{x_r}} = \frac{w_k}{\frac{w_k}{x_k}} \qquad (9.18)$$

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

$$\bar{x} = \frac{1 + 2 + 4 + 7 + 9}{5} = 4{,}6$$

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, 6, 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 posamezni 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:

$60/4 = 15;\ 60/5 = 12;\ 60/2 = 30;\ 60/6 = 10;\ 60/3 = 20.$

Skupno število proizvedenih artiklov pa dobimo, če te podatke seštejemo. Povprečna produktivnost dela je torej

$$\frac{5 \cdot 60}{60/4 + 60/5 + 60/2 + 60/6 + 60/3} = \frac{300}{87} = 3{,}45 \text{ minut za artikel}$$

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

$$\bar{x} = \frac{1}{5}(4 + 5 + 2 + 6 + 3) = 4.$$

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

9.28 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 X_k in Y_k

$$r_k = X_k/Y_k, \tag{9.19}$$

moremo ta obrazec pisati v več oblikah. Iz njega dobimo, da je

$$X_k = Y_k r_k \tag{9.20}$$

in

$$Y_k = X_k/r_k \tag{9.21}$$

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

$$r = \frac{X}{Y} = \frac{\Sigma X_k}{\Sigma Y_k} \tag{9.22}$$

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

$$r = \frac{\Sigma Y_k r_k}{\Sigma Y_k} \tag{9.23}$$

tehtana artimetična 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 pondera z relativnim številom $Y_k r_k$.

c) Če upoštevamo obrazec 9.21, pa je sumarno relativno število enako

$$r = \frac{\Sigma X_k}{\Sigma X_k / r_k} \qquad (9.24)$$

To pa je tehtana harmonična sredina grupnih relativnih števil r_k. Pri tem so ponderi količine, ki nastopajo v relativnem številu v številu. 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 številu, oziroma kadar je smiseln kvocient pondera 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 sumarnih 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, porabljen za proizvodnjo enega artikla", ki ga dobimo. če porabljen čas delimo s številom proizvedenih artiklov. Ker razpolagamo s časom, kolikor so posamezni delavci delali (60 minut), so ponderi količine, ki v relativnem številu nastopajo v številu. Po zgornjem pravilu je treba za sumarno relativno število uporabiti harmonično sredino.

Če poznamo površine in povprečne gostote prebivalstva po republikah, izračunamo gostoto prebivalstva za SFRJ 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

$$r = \frac{\Sigma Y_k r_k}{Y} = \Sigma \frac{Y_k}{Y} \cdot r_k = \Sigma Y_k^o rk \qquad (9.25)$$

Podobno dobimo iz obrazca 9.24

$$r = \frac{X}{\Sigma X_k / r_k} = \frac{1}{\Sigma \frac{X}{X} k / r_k} = \frac{1}{\Sigma X_k^o / r_k} \qquad (9.26)$$

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.

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 R.

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 SFRJ (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	9530	8950	0,157	0,218
Nekvalificirani	8030	6770	0,605	0,406
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 dvomljiva in meglena, 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 produktinosti 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 pondere. Tako je standardiziran koeficient urmljivosti 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 koeficienti 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 delavcevpo kvalifikaciji za kmetijstvo in gradbeništvo iz tabele 9.5 standardizirane povprečne prejemke delavcev za obe panogi. Za standardno strukturo vzemimo povprečno strukturo delavstva v SFRJ po kvalifikaciji v vseh panogah.

Ker za izračunavanje sumarnih povprečnih prejemkov vzamemo za standardno strukturo - strukturo delavstva po kvalifikaciji v SFRJ - dobimo v kmetijstvu povprečne mesečne prejemke večje (10576 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 povprečij, 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 SFRJ

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	9530	8950	0,263	2506,39	2353,85
Nekvalificirani	8030	6770	0,297	2384,91	2010,69
			1,000	10576,50	9855,34

GEOMETRIJSKA SREDINA

9.32 Geometrijska sredina G iz N vrednosti x_1, x_2, x_3, ... x_N je N koren iz produkta vseh vrednosti

$$G = \sqrt[N]{x_1 . x_2 . \ \dots \ x_N} \qquad (9.27)$$

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

$$G = \sqrt[w]{x_1^{w1} . \ x_2^{w2} . \ \dots \ x_N^{wN}} \qquad (9.28)$$

w_1, w_2 ... 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

$$\log G = \frac{1}{N}(\log x_1 + \log x_2 + \log x_3 + \dots + \log x_N) =$$

$$= \frac{1}{N} \Sigma \log x \qquad (9.29)$$

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 arimetični 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 indeks je logG = 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).

Zaznamujmo z Y_o, Y_1, Y_2, ... Y_N posamzne člene v časovni vrsti, s $K_1 = Y_1/Y_o$, $k_2 = Y_2/Y_1$, ... $k_N = Y_N/Y_{N-1}$ 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

$$Y_o.k_1.k_2 \ldots k_N = Y_N = Y_o k.k...k = Y_o k^N \qquad (9.31)$$

Če namreč Y_o 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_o 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

$$k = \sqrt[N]{Y_N/Y_o} \qquad (9.33)$$

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 katerikoli časovni razmak.

9.37 Pri obravnavanju verižnih indeksov v odstavku o enostavnih indeksih smo izračunali vrsto verižnih indeksov za proizvodnjo električne energije v SFRJ v letih 1951-1956. Verižni indeksi v teh letih so: 105,9; 110,4; 115,4; 126,1; 117,1.

Srednji verižni indeks v tem razdobju je po obrazcu 9.32 enak geometrijski sredini iz individualnih verižnih indeksov

$$I = \sqrt{105{,}9 \cdot 110{,}4 \cdot 115{,}4 \cdot 126{,}1 \cdot 117{,}1}$$

Z logaritmi izračunamo srednji verižni indeks po obrazcu 9.29

$$\log I = \frac{1}{5}(\log 105{,}9 + \log 110{,}4 + \log 115{,}4 + \log 126{,}1 + \log 117{,}1) =$$

$$= \frac{1}{5}(2{,}02490 + 2{,}04297 + 2{,}06221 + 2{,}10106 + 2{,}06856) =$$

$$= 2{,}05994$$

Z antilogaritmiranjem dobimo, da je srednji verižni indeks $I = 114{,}8$.

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 10.33 iz začetne in končne proizvodnje električne energije. Po podatkih iz tabele 6.19 je bila proizvodnja električne energije v letu 1951 v SFRJ $Y_{51} = 2550$, v letu 1956 pa $Y_{56} = 5084$ milijonov kWh.

Po obrazcu 10.33 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

$$H = \frac{2}{1/1 + 1/4} = 1,3; \quad G = \sqrt{1.4} = 2; \quad M = \frac{1 + 4}{2} = 2,5$$

Iz teh rezultatov vidimo, da je harmonična sredina, če jo izračunamo iz istih podatkov, najmanjša, aritmetična sredina največja, geometrijska sredina pa leži med obema. To ne velja samo za ta primer, ampak velja splošno pravilo: harmonična sredina je manjša kakor geometrijska, ta pa manjša kakor aritmetična sredina, če jih izračunamo iz istih podatkov.

$$H < G < M \qquad (9.34)$$

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 ($M < Me < Mo$).

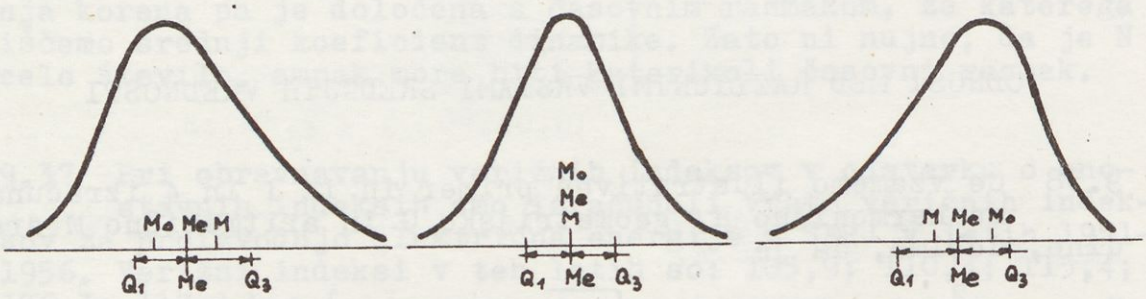
Za zvezne, ne preveč asimetrične distribucije velja, da je razlika med aritmetično sredino in modusom približno trikrat večja kot razlika med aritmetično sredino in mediano.

$$M - Mo = 3.(M - Me) \qquad (9.35)$$

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

$$Mo = M - 3.(M - Me) \qquad (9.36)$$

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 itd.; od podjetja do podjetja poraba surovin, število delavcev, vrednost proizvodnje, poraba električne energije, goriva itd.; od človeka do človeka starost, spol, stan, dohodek itd.

Medtem ko je srednja vrednost izraz splošnih pogojev, ki so za vso populacijo enaki - nespremenjeni, se zaradi 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: SB 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

$$Q_3 - Q_1$$

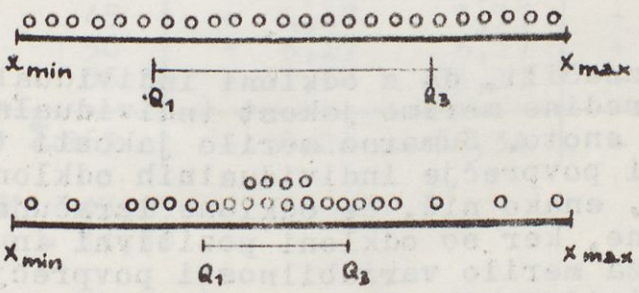
vzamemo za mero variacije kvartilni odklon

$$Q = \frac{Q_3 - Q_1}{2} \qquad (10.2)$$

ki je polovica kvartilnega razmaka.

Zgornji meri variacije pravilneje pokažeta jakost variacije kakor variacijskih 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 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 odstavku 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

$$Q = \frac{Q_3 - Q_1}{2} = \frac{50 - 38}{2} = 6$$

Za množino padavin v juniju 1954 za 23 meteoroloških postaj v Sloveniji smo kvartile izračunali že v odstavku o kvantilih

(Q_1 = 151,3 mm; Q_3 = 221,5 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 1953 pa je po podatkih iz odstavka o kvantilih (8.9) (Q_1 = 10,2 m3, Q_3 = 18,4 m3) 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 absolutni 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.

Če 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.

Resnično 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 absolutni 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 doklonov 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 ognemo 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 <u>Izračunavanje iz negrupiranih podatkov. Direktna metoda.</u> 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 = 32,83

Po obrazcu 10.6 dobimo

$$\sigma^2 = 362,8334/6 = 60,4722;$$

Standardni odklon SD pa je po obrazcu 10.7 enak

$$SD = \sqrt{60,4722} = 7,76 \text{ dinarjev.}$$

10.12 Metoda pomožnega znaka u. Ker so odkloni od povprečja navadno decimalna števila, to komplicira kvadriranje. Zato je izračunavanje variance in standardnega odklona po direktni metodi neprikladno. 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_o tako, da se stvarne vrednosti x od nje čim manj razlikujejo;

b) Izračunamo odklone stvarnih vrednosti od x_o; tako dobimo vrednosti $u = x - x_o$;

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, $x_o = 45$. Glede na to je potek računa nakazan v tabeli 10.3.

Tabela 10.3 Izračunavanje variance za ceno mleka za glavna mesta republik v SFRJ 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

$$K = \Sigma u^2 - U^2/N = 371 - (-7)^2/6 = 362{,}8333$$

$$\sigma^2 = K/N = 362{,}8333/6 = 60{,}4722; \quad \sigma = \sqrt{60{,}4722} = 7{,}76 \text{ din.}$$

Dobili smo torej iste rezultate kakor po direktni metodi.

Če se ne moremo odločiti za nobeno vrednost x_o, ali če izračunavamo variance z računskih strojem, vzamemo $x_o = 0$. 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 zamudneje 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, ne prikladen. 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:

1) 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:

... -3, -2, -1, 0, +1, +2, +3, ...

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 m3	f	u	fu	fu^2
0,0 - 3,9	2	-3	- 6	18
4,0 - 7,9	16	-2	-32	64
8,0 - 11,9	36	-1	-36	36
12,0 - 15,9	41	0	0	0
16,0 - 19,9	29	+1	+29	29
20,0 - 23,9	12	+2	+24	48
24,0 - 27,9	7	+3	+21	63
28,0 - 31,9	4	+4	+16	64
32,0 - 35,9	2	+5	+10	50
	149		+26	372
	N		U	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 prikladneje 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 m3	f	F	FF	
0,0 - 3,9	2	0		
4,0 - 7,9	16	2	0	
8,0 - 11,9	36	18	2	
12,0 - 15,9	41	54	20	
16,0 - 19,9	29	95	74	
20,0 - 23,9	12	124	169	
24,0 - 27,9	7	136	293	
28,0 - 31,9	4	143	429	
32,0 - 35,9	2	147	572	
		149	719	1559 = 2+20+...+429+572
		N	A	B

$K = 2.1559 + 719 - 719^2/149 = 367,4631.$

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 m3	f	F N 149	FF A 473	B 698 = 326+195+...+8+2
0,0 - 3,9	2	147	326	
4,0 - 7,9	16	131	195	
8,0 - 11,9	36	95	100	
12,0 - 15,9	41	54	46	
16,0 - 19,9	29	25	21	
20,0 - 23,9	12	13	8	
24,0 - 27,9	7	6	2	
28,0 - 31,9	4	2	0	
32,0 - 35,9	2	0		

$K = 2.698 + 473 - 473^2/149 = 367,4631$

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,

$$\sigma_M^2 = \frac{1}{N}\Sigma N_k(M_k - M)^2 \qquad (10.16)$$

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 kulturo 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	28358
20 - 24	10	670	35820
25 - 29	7	949	48212
30 - 34	4	1317	83469
35 -	3	1703	8089
	36	819	
	N	M	

Iz teh podatkov dobimo

$$\sigma_M^2 = \frac{1}{36}\left[12(479-819)^2 + \ldots + 3(1703-819)^2\right] = 140709$$

$$M_{\sigma^2} = \frac{1}{36}(12.28358 + \ldots + 3.8089) = 38726$$

$$\sigma^2 = \sigma_M^2 + M_{\sigma^2} = 38726 + 140709 = 179435$$

10.22 Pomembnejše kakor to, da moremo izračunati skupno varianco, če poznamo grupne podatke, pa je dejstvo, da moremo z zgornjim postopkom skupno varianco razstaviti po grupnem znaku v dva dela. Skupna varianca namreč meri vpliv vseh 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.

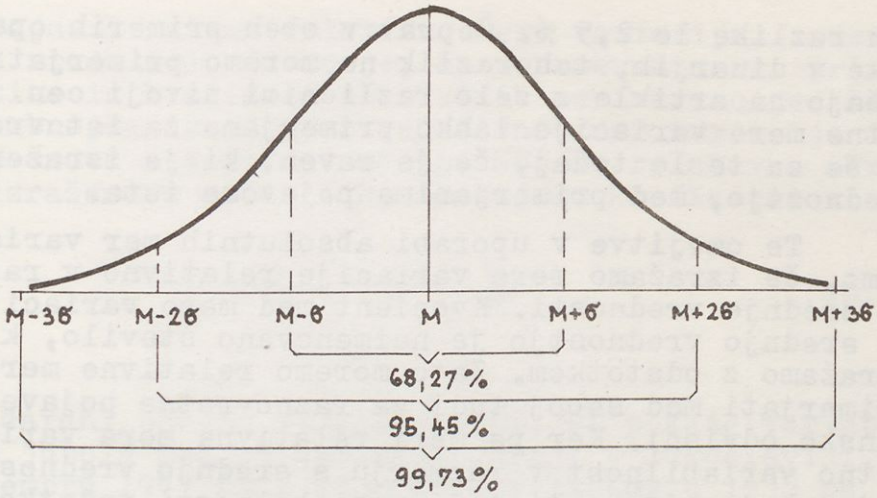
Variacijski razmak ali kvartilni razmak imata enostavno tolmačenje in pomen. Enako je tudi povprečen absoluten odklon razmeroma razumljiv. Teže pa je s standardnim odklonom, dokler ga ne obravnavamo v zvezi z nekaterimi lastnostmi distribucij. Standardni odklon namreč dobimo z razmeroma zapletenim računskim 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 $M + \sigma$ 68,27 % ali okroglo 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.

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 m3 do 20,87 m3 70 % vseh primerov, v razmaku $M - 2\sigma$ do $M + 2\sigma$, to je od 2,36 m3 do 27,04 m3 94 % vseh primerov, v razmaku $M - 3\sigma$ do $M + 3\sigma$, to je od 0 do 33.21 m3 pa 99.3 % vseh primerov. Razlike od odstotkov, ki jih pričakujemo pri normalni distribuciji, so torej neznatne.



Slika 10.2 Frekvence za normalno distribucijo v razmakih M $\pm\sigma$, M $\pm$ 2σ in M $\pm$ 3σ.

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:

$$Q = 0{,}6745\sigma \approx 2/3 \qquad (10.17)$$

$$AD = 0{,}7979\sigma \approx 4/5 \qquad (10.18)$$

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 npr. 20 din, je 5 din 25 % povečanje; če pa vzamemo, da je povprečna cena blaga za moške suknje npr. 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 variabilnosti 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 absolutni 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:

$$\frac{Q}{Me} \text{ ali } \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad (10.20)$$

Na osnovi povprečnega absolutnega odklona AD:

$$\frac{AD_{Me}}{Me} \text{ ali } \frac{AD_M}{M} \qquad (10.21)$$

Na osnovi standardnega odklona SD:

$$KV^o = \frac{\sigma}{M} \qquad (10.22)$$

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 = 113,6 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	
Enotni kruh kg	47,3	2,5	5,3
Svinjska mast uvožena kg	325,1	9,8	3,0
Jabolka kg	113,6	27,4	24,1
Jajca kos	17,8	3,8	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 variabilnosti 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.

$$M - Mo = 3(M - Me) \qquad (10.24)$$

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

$$KA_{Mo} = \frac{M - Mo}{\sigma} \qquad (10.25)$$

ali s koeficientom asimetrije, ki ima za osnovo razliko med aritmetično sredino in mediano

$$KA_{Me} = \frac{3(M - Me)}{\sigma} \qquad (10.26)$$

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.

$$KA_Q = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1} \qquad (10.27)$$

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 teh podatkov dobimo po zgornjih obrazcih koeficiente asimetrije

$$KA_{Mo} = \frac{14,70 - 13,17}{6,17} = +0,248$$

$$KA_{Me} = \frac{3(14,70 - 14,05)}{6,17} = +0,316$$

$$KA_Q = \frac{18,38 + 10,19 - 2 \cdot 14,05}{18,38 - 10,19} = +0,057$$

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 $Q_1 = 10{,}19$; $Q_3 = 18{,}38$; $D_1 = 7{,}33$; $D_9 = 23{,}53$. Iz teh podatkov dobimo, da je

$$KS = 1{,}9\,\frac{18{,}38 - 10{,}19}{23{,}53 - 7{,}33} = 0{,}96$$

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

$$\bar{I} = \frac{1}{N} \Sigma \frac{p_1}{p_o} \qquad (11.1)$$

Pri tem pomeni:

p_o = 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 važnosti 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 (npr. moka, kruh, mleko, meso itd.) za potrošnika veliko pomembnejše, kot spremembe za cene artiklov, za katere je poraba majhna (npr. televizij-

ski 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.

$$V = \Sigma p \cdot q \qquad (11.2)$$

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 itd.

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

$$V_0 = \Sigma p_0 q_0; \quad V_1 = \Sigma p_1 q_1; \quad V_2 = \Sigma p_2 q_2;$$
$$V_3 = \Sigma p_3 q_3; \ \ldots \qquad (11.3)$$

Iz te vrste nazorno vidimo, da so spremembe v skupni vrednosti rezultat sprememb cen in količin v posameznih razdobjih. Indeks vrednosti, ki ga iz teh 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. Stal-

nost 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

$$I_p = \frac{\Sigma p_0 q \cdot \frac{p_1}{p_0}}{\Sigma p_0 q} \qquad (11.7)$$

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 11.6 za izračunavanje agregatnega indeksa cen morejo biti teoretično katerekoli 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 čimbolj 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 11.6 konkreten obrazec za izračunavanje agregatnega indeksa cen

$$L_p = \frac{\Sigma p_1 p_o}{\Sigma p_o q_o} \qquad (11.9)$$

Metodo, po kateri pri izračunavanju agregatnih indeksov vzamemo za pondere 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 katerokoli 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 $p_1 > p_o$, ima ta podražitev za rezultat zmanjšanje porabe tega artikla: $q_1 < q_o$. Obratno znižana cena $p_1 < p_o$ navadno poveča porabo $q_1 > q_o$. Ker v Laspeyresovem obrazcu vzamemo za pondere količine iz bazičnega razdobja, so ponderi pri artiklih, ki so se podražili, preveliki, pri artiklih, ki so se pocenili, pa premajhni. Skupen učinek teh razlik je sistematična pristranost navzgor, ker so ponderi za artikle, ki se podražijo, preveliki, ponderi za artikle, ki se pocenijo, pa premajhni.

11.6 Vzemimo kot primer za izračunavanje agregatnega indeksa cen po Laspeyresovem obrazcu indeks cen mesa v razdobju 1954-1957 v SRS. 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 SR Slovenije.

Tabela 11.1 Izračunavanje agregatnega indeksa cen za surovo in prekajeno meso v splošnem indeksu cen na malo v SR Sloveniji po Laspeyresovem obrazcu (cene v dinarjih, količine v tisočih tonah)

Artikel	q_o	1954		1955		1956		1957	
		p_o	$p_o q_o$	p_1	$p_1 q_o$	p_2	$p_2 q_o$	p_3	$p_3 q_o$
Surovo meso s kostmi:									
Goveje	11,6	223	2586,8	250	2900,0	243	2818,8	251	2911,6
Telečje	2,9	271	785,9	304	881,6	296	858,4	305	884,5
Svinjsko	3,7	249	921,3	292	1080,4	291	1076,7	351	1298,7
Ovčje	0,2	166	33,2	192	38,4	180	36,0	202	40,4
Prekajeno svinjsko meso	2,1	320	672,0	364	764,4	375	787,5	402	844,2
Prekajena slanina	1,2	323	387,6	351	421,2	374	448,8	396	475,2
			5386,8		6086,0		6026,2		6454,6
			$\Sigma p_o q_o$		$\Sigma p_1 q_o$		$\Sigma p_2 q_o$		$\Sigma p_3 q_o$
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 pondere 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 od primera do 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 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 Fischer vzel v ta namen geometrijsko sredino med L_p in P_p

$$F_p = \sqrt{L_p \cdot P_p} = \quad L_p \cdot P_p = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \qquad (11.11)$$

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 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 Fischerjevem. Iz tabele vivimo, 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

$$I_{1/o} = I_1; \quad I_{2/o} = I_{1/o} \cdot I_2; \quad I_{3/o} = I_{2/o} \cdot I_3 \cdots \qquad (11.12)$$

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

$$I_1 = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}; \quad I_2 = \frac{\Sigma p_2 q_1}{\Sigma p_1 q_1}; \quad I_3 = \frac{\Sigma p_3 q_2}{\Sigma p_2 q_2} \cdots \qquad (11.13)$$

b) Iz vrste verižnih indeksov dobimo indeksno vrsto s stalno osnovo tako, da verižne indekse postopoma kumulativno množimo:

$$I_{1/o} = I_1; \quad I_{2/o} = I_{1/o} \cdot I_2; \quad I_{3/o} = I_{2/o} \cdot I_3 \text{ itd.} \qquad (11.14)$$

Pri tem pomenijo $I_{1/o}$, $I_{2/o}$, $I_{3/o}$ ... indekse s stalno osnovo 0.

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_o	p_1	q_1	p_oq_1	p_1q_1	p_2	q_2	p_oq_2	p_2q_2	p_3	q_3	p_oq_3	p_3q_3
Surovo meso s kostmi:													
Goveje	223	250	12,2	2720,6	3050,0	243	13,4	2988,2	3256,2	251	14,5	3233,5	3639,5
Telečje	271	304	3,8	1029,8	1155,2	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			7076,3	8386,8
				Σp_oq_1	Σp_1q_1			Σp_oq_2	Σp_2q_2			Σp_oq_3	Σp_3q_3
		$P_1 = 100 \frac{6884,5}{6096,1} = 112,9$				$P_2 = 100 \frac{7406,6}{6636,4} = 111,6$				$P_3 = 100 \frac{8383,8}{7076,3} = 118,5$			

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_o	p_1	q_o	p_oq_o	p_1q_o	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	2964,6	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	1539,2	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	36,0	40,4
Prekajeno svinjsko meso	320	323	2,1	672,0	764,4	375	2,6	946,4	975,0	402	2,5	937,5	1005,0
Prekajena slanina	323	351	1,2	387,6	421,2	374	1,5	526,5	561,0	396,	1,5	561,0	594,0
				5386,8	6086,0			6884,5	6793,0			7406,6	7887,5
				Σp_oq_o	Σp_1q_o			Σp_1q_1	Σp_2q_1			Σp_2q_2	Σp_3q_2
		$I_1 = 100 \frac{6086,0}{5383,8} = 113,0$				$I_2 = 100 \frac{6793,0}{6884,5} = 98,7$				$I_3 = 100 \frac{7887,5}{7406,6} = 106,5$			
		$I_{1/o} = 113,0$				$I_{2/o} = 113,0 \cdot 98,7 = 111,5$				$I_{3/o} = 111,5 \cdot 106,5 = 118,7$			

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 SR 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
Fischerjev 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

$$I_{2/o} = I_{1/o} \cdot I_{2/1} \qquad (11.15)$$

(Glej obrazec 6.11).

Če z O 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

$$I_{2/o} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \cdot \frac{\Sigma p_2 q_1}{\Sigma p_1 q_1} \qquad (11.16)$$

Z njim rešimo problem spremembe ponderacije in nadomestila artiklov z drugimi.

11.14 Za primer izračunavanja agregatnega indeksa cen pri spremembi ponderacijskega 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_o	q_o	p_1	q_1	p_2	$p_o q_o$	$p_1 q_o$	$p_1 q_1$	$p_2 q_1$
A	5	12	5	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	-	-
F	-	-	4	6	6	-	-	24	36
						150	166	201	264
						$\Sigma p_o q_o$	$\Sigma p_1 q_o$	$\Sigma p_1 q_1$	$\Sigma p_2 q_1$

$$I_{1/o} = 100 \frac{166}{150} = 110,7; \quad I_{2/1} = 100 \frac{264}{201} = 131,5;$$

$$I_{2/o} = I_{1/o} \cdot I_{2/1} = 110,7 \cdot 131,5 = 145,2$$

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 kak 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

11.15 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 indeksom, 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 združimo predmete v več grup sorodnih predmetov.

b) Iz vsakega homogenega dela izberemo nekaj predmetov. Gibanje cen za te izbrane reprezentatne 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 navadno 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 teorestka 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 po-

trošni koš stalen. Če izračunamo, kolikšni so stroški za preživljanje, ako hočemo zadostiti porabi po teoretski listi, dobimo absoluten iznos; ta je fiktiven in ga ne uporabljamo za analizo. Indeksi, ki jih izračunamo iz teh absolutnih - fiktivnih stroškov, pa zelo dobro kažejo, kako se giblje raven stvarnih življenjskih stroškov, če je potrošni koš po svoji 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 SR 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 izračunavanja, 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 teh

Tabela 11.7 Indeks življenjskih stroškov v Sloveniji. Izračunavanje skupinskega indeksa za hrano v mesecu novembru 1958
(Podatki ZS - SR Slovenije)

Artikel	Enota mere	Količina mesečne porabe	ø 1957		November 1958		Indeks
			cene	stroški	cene	stroški	
Kruh pšenični pol beli	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	262	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	83,6	84	811	81	96
Sol drobna	"	1	38,2	38	43,5	44	116
Kavin nadom. "Divka"	zav.	2	35	70	36,6	73	104
Kava prava	kg	0,1	2039	204	2125	213	104
				14055		13823	98

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 teh so v novembru vzeti 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	6,2
Zelje - presno	0,56	16,0	9,0
Zelje - kislo	-	-	-
Paradižnik - vkuhan	0,12	316,0	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 62,5 din. Ta povprečna cena je vnesena za ceno povrtnine v indeks življenjskih stroškov za prehrano.

AGREGATNI INDEKSI KOLIČIN

11.18 Dosedaj 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

$$L_q = \frac{\Sigma p_o q_1}{\Sigma p_o q_o} \qquad (11.17)$$

za pondere 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 pondere cen iz tekočega razdobja.

Enako je geometrijska sredina med Laspeyresovim in Paaschejevim obrazcem Fischerjev idealni obrazec za izračunavanje agregatnega indeksa količin

$$F_q = \sqrt{L_q \cdot P_q} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}} \qquad (11.19)$$

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 pondere, 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 $L_q = \frac{\Sigma p_0 q_3}{\Sigma p_0 q_0} = \frac{7076,3}{5386,8} = 131,3$

po Paaschejevem obrazcu $P_q = \frac{\Sigma p_3 q_3}{\Sigma p_3 q_0} = \frac{8386,8}{6454,6} = 129,9$

po Fischerjevem obrazcu $F_q = \sqrt{L_q \cdot P_q} = \sqrt{131,3 \cdot 129,9} =$

$= 131,1$

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 reprezenta-

tivni 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 $V_1 = \Sigma p_1 q_1$, moremo vrednost proizvodnje reducirati na cene v bazičnem razdobju $V_1' = \Sigma p_o q_1$, če jo delimo z ustreznim reprezentativnim indeksom cen

$$V_1' = \Sigma p_o q_1 \approx \frac{V_1}{P_p'}; \quad P_p' = \frac{\Sigma p_1' q_1'}{\Sigma p_o' q_o'} \tag{11.20}$$

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 $\Sigma p_1 q_1$ in $\Sigma p_o q_1$.

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_p', vrednost prometa v trgovini na drobno z indeksom cen v trgovini na drobno, promet v trgovini na debelo z indeksom cen v trgovini na debelo itd. Če ocenjujemo vrednost samo za določeno skupino, reduciramo vrednost z ustreznim grupnim indeksom cen itd.

Testi o agregatnih indeksih

11.21 <u>Test o zamenljivosti blaga.</u> Če v individualnem indeksu cen $i_{1/o} = p_1/p_o$ zamenjamo osnovo in tekoče razdobje $i_{o/1} = p_o/p_1$, je produkt obeh indeksov

$$i_{1/o} \cdot i_{o/1} = \frac{p_1}{p_o} \cdot \frac{p_o}{p_1} = 1 \tag{11.21}$$

enak ena.

Te lastnosti pa nimajo vsi agregatni indeksi in je posebna odlika le nekaterih vrst agregatnih indeksov. Če vzamemo npr. Laspeyresov obrazec za izračunavanje agregatnega indeksa cen, zanj ta zveza ne velja, ker je

$$L_{1/o} \cdot L_{o/1} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \cdot \frac{\Sigma p_o q_1}{\Sigma p_1 q_1} \neq 1 \tag{11.22}$$

Zgornja zveza ne velja niti za Paaschejev obrazec cen niti količin. Pač pa velja za Fischerjev obrazec. Tako dobimo, da je za indeks cen po Fischerjevem 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_1, 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

$$L_p \cdot L_q = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = I_v \qquad (11.25)$$

Te lastnosti tudi drugi agregatni indeksi nimajo, razen Fischerjevega 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 Fischerjev obrazev 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

$$V = \Sigma Tv \qquad (11.28)$$

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

$$I_v = \frac{\Sigma T_o v_1}{\Sigma T_o v_o} \qquad (11.29)$$

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

$$T = \Sigma Qt \qquad (11.31)$$

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

$$I_t = \frac{\Sigma Q_o t_o}{\Sigma Q_o t_1} \qquad (11.32)$$

V tem indeksu je agregat $\Sigma Q_o t_1$ v imenovalcu, $\Sigma Q_o t_o$ pa v števcu zato, ker je t recipročen pokazatelj produktiv-

nosti 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 $P = \Sigma N\bar{p}$, 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 npr. 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

Č A S O V N E V R S T E

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 slede na filmskem platnu, pa daje videz 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. Momemtne časovne vrste prikazujejo časovni razvoj pojavov, ki veljajo za posamezne momente, intervalne časovne vrste pa časovni razvoj pojavov, ki so intervalnega značaja. Momemtna časovna vrsta so npr. 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 SFRJ po letih sredi leta momentna, naravni prirastek po letih pa intervalna časovna vrsta.

Tabela 12.1 Število prebivalstva in naravni prirastek prebivalstva v SFRJ

Leto	Število prebivalstva sredi leta v tisočih	Naravni prirastek v tisočih
1947	15679	217
1948	15901	233
1949	16133	266
1950	16346	282
1951	16588	213
1952	16798	302
1953	17048	272
1954	17318	306
1955	17586	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 itd. Kumulativne vrste se v teh primerih nanašajo na začetek petletke, začetek leta, začetek meseca.

Kumulativno časovno vrsto dobimo tako, da vzamemo vrednost prvega člena kumulativne vrste enako $K_1 = 0$, 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 SFRJ.

Tabela 12.2 Izračunavanje kumulativne časovne vrste za proizvodnjo krogličnih ležajev v SFRJ 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	115 + 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 SFRJ 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 1946-1957, 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, neglede 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, npr. 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

$$\bar{Y} = \frac{1}{r} (y_1 + y_2 + \ldots + y_r) \qquad (12.2)$$

Za momente časovne vrste, za katere se vrednosti nanašajo na začetek ali konec osnovnih intervalov, pa izračunavamo sredine po obrazcu

$$\bar{Y} = \frac{1}{r} \left(\frac{1}{2} Y_o = Y_1 + Y_2 + \ldots + Y_{r-1} + \frac{1}{2} Y_r\right) \qquad (12.3)$$

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 katerokoli 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:

$$S_{k+1} = S_k + Y_k - Y_{k-r} = S_k + {}_r d_k \qquad (12.4)$$

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; ${}_r d_k = Y_k - Y_{k-r}$ = razlika ustreznih vrednosti za Y.

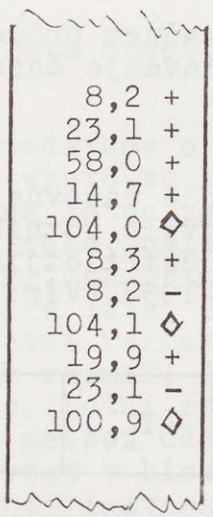
12.9 Za število potnikov v zračnem prometu v SFRJ v razdobju 1955-1958 je izračunavanje časovne vrste drsečih vsot nakazano v tabeli 12.3.

Tabela 12.3 Izračunavanje časovne vrste drsečih vsot za število potnikov v zračnem prometu po četrtletjih v SFRJ v razdobju 1955-1958 (Vir: Indeks)

Leto	Četrtletje	Y_k	Y_{k-4}	4^d_k	S_k
1955	1	8,2			
	2	23,1			
	3	58,0			
	4	14,7			
1956	1	8,3	8,2	+ 0,1	104,0
	2	19,9	23,1	- 3,2	104,1
	3	50,6	58,0	- 7,4	100,9
	4	15,2	14,7	+ 0,5	93,5
1957	1	9,8	8,3	+ 1,5	94,0
	2	23,6	19,9	+ 3,7	95,5
	3	60,4	50,6	+ 9,8	99,2
	4	17,9	15,2	+ 2,7	109,0
1958	1	10,9	9,8	+ 1,1	111,7
	2	26,0	23,6	+ 2,4	112,8
	3	72,3	60,4	+11,9	115,2
	4	21,0	17,9	+ 3,1	127,1
					130,2

Iz primera je jasno razviden postopek 8,2 + 23,1 + + 38,0 + 14,7 = 104,0; 104,0 + (+0,1) = 104,1; 104,1 + + (-3,2) = 100,9 ...

Če računamo časovno vrsto drsečih vsot z računskih strojem, ki ima registrirni trak, ni treba izračunavati vmesnih diferenc. Časovno vrsto drsečih vsot dobimo neposredno, če uporabljamo 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 (npr. 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

$$\bar{Y}_k = \frac{1}{r} \left(Y_{k-i} + Y_{k-i+1} + \ldots + Y_k + \ldots Y_{k+i-1} + Y_{k+i}\right) =$$

$$= \frac{1}{r} S_{k+i+1} \qquad (12.5)$$

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 1921-1939. (Vir: SG 57)

Leto	Y_k	Y_{k-5}	$5^d k$	S_k	$\bar{Y}_k$
1921	20,9				
1922	20,8				
1923	20,3				20,2
1924	20,2				19,8
1925	18,7				19,8
1926	18,8	20,9	- 2,1	100,9	19,8
1927	20,9	20,8	+ 0,1	98,8	20,0
1928	20,4	20,3	+ 0,1	98,9	20,1
1929	21,2	20,2	+ 1,0	99,0	20,3
1930	19,0	18,7	+ 0,3	100,0	19,9
1931	19,8	18,8	+ 1,0	100,3	19,2
1932	19,2	20,9	- 1,7	101,3	18,4
1933	16,9	20,4	- 3,5	99,6	18,0
1934	17,1	21,2	- 4,1	96,1	17,2
1935	16,8	19,0	- 2,2	92,0	16,6
1936	16,1	19,8	- 3,7	89,8	16,3
1937	15,9	19,2	- 3,3	86,1	15,9
1938	15,6	16,9	- 1,3	82,8	
1939	14,9	17,1	- 2,2	81,5	
				79,3	

12.12 b) Če pa sredine veljajo za sodo število osnovnih razmakov (r = 2i), drseče sredine izračunavamo po obrazcu:

$$\bar{Y}_k = \frac{1}{r}\left(\frac{1}{2}Y_{k-i} + Y_{k-i+1} + \ldots + Y_k + \ldots + Y_{k+i-1} + \frac{1}{2}Y_{k+i}\right) \quad (12.6)$$

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 12.6 preurediti v obliko

$$\bar{Y}_k = \frac{1}{2r}\,(Y_{k-i} + 2Y_{k-i+1} + \ldots + 2Y_k + \ldots\ 2Y_{k+i-1} + Y_{k+i}) =$$

$$= \frac{1}{2r}\,S'_k \qquad (12.7)$$

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 odsekih 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 vrsto $S'_{i+1} = Y_1 + 2Y_2 + \ldots\ 2Y_{i+1} + \ldots + 2Y_r + Y_{r+1}$.

d) Druge vsote dobimo postopno po obrazcu

$$S'_{k+1} = S'_k + {}_r d_{k+i} + {}_r d_{k+i+1} \qquad (12.8)$$

e) Iz vsot S'_k dobimo vsoto 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 SFRJ.

Tabela 12.5 Izračunavanje časovne vrste drsečih sredin za indeks industrije gradbenega materiala v SFRJ (Ø = 100) (Vir: SB 108)

		J	F	M	A	M	J	J	A	S	O	N	D
1951	Y	38	37	49	56	78	95	97	104	90	84	71	58
	d												
S′= 24	Ȳ							1707	1690	1667	1648	1635	1634
	Ȳ							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	Ȳ	1667	1712	1752	1788	1802	1800	1801	1810	1828	1858	1897	1936
	Ȳ	69,5	71,3	73,0	74,5	75,1	75,0	75,0	75,4	76,2	77,4	79,0	80,7
1953	Y	34	33	48	68	92	119	125	131	127	118	86	74
	d	+ 3	+ 6	+12	+18	+21	+18	+ 1	+ 9	+15	+20	+15	+18
S′= 24	Ȳ	1955	1965	1989	2024	2059	2092	2108	2095	2084	2081	2071	2062
	Ȳ	81,5	81,9	82,9	84,3	85,8	87,2	87,8	87,3	86,8	86,7	86,3	85,9
1954	Y	32	22	48	65	85	117	134	143	145	122	90	83
	d	- 2	-11	0	- 3	- 7	- 2	+ 9	=12	=18	+ 4	+ 4	+ 9
S′= 24	Ȳ	2069	2090	2120	2142	2150	2163	2195	2247	2282	2284	2292	2332
	Ȳ	86,2	87,1	88,3	89,3	89,6	90,1	91,5	93,6	95,1	95,2	95,5	97,2
1955	Y	55	51	54	61	97	145	138	144	135	121	89	81
	d	+23	+29	+ 6	- 4	+12	+28	+ 4	+ 1	-10	- 1	- 1	- 2
S′= 24	Ȳ	2364	2369	2360	2349	2347	2344	2339	2311	2276	2266	2256	2232
	Ȳ	98,5	98,7	98,3	97,9	97,8	97,7	97,5	96,3	94,8	94,4	94,0	93,0
1955	Y	52	26	44	61	87	131	139	157	155	154	106	89
	d	- 3	-25	-10	0	-10	-14	+ 1	+13	+20	+33	+17	+ 8
S′= 24	Ȳ	2219	2233	2266	2319	2369	2394	2394	2398	2420	2445	2492	2547
	Ȳ	92,5	93,0	94,4	96,6	98,7	99,8	99,8	99,9	100,8	101,9	103,8	106,1
1955	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	Ȳ	2603	2658	2691	2700	2713	2736						
	Ȳ	108,5	110,8	112,1	112,5	113,0	114,0						

Razlike med ustreznimi meseci zaporednih let smo izračunali takole: Prva možna razlika je v mesecu januarju 1952;

$$_{12}d_{13} = Y_{J,52} - Y_{J,51} = 31 - 38 = -7;$$

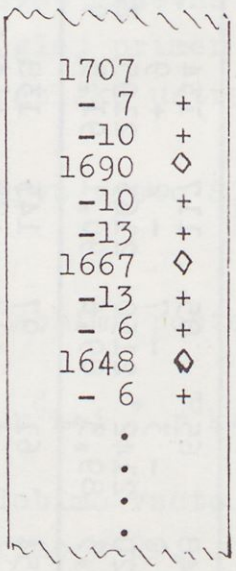
$$_{12}d_{14} = Y_{F,52} - Y_{F,51} = 27 - 37 = -10 \text{ itd.}$$

Prvo vsoto (za julij 1951) smo izračunali v celoti:

$$S'_{J,51} = 38 + 2.37 + 2.49 + \ldots + 2.84 + 2.71 + 2.58 + 31 = 1707.$$

Nadaljnje vsota pa so izračunane, kot kaže trak na seštevalnem stroju na sliki 12.2.

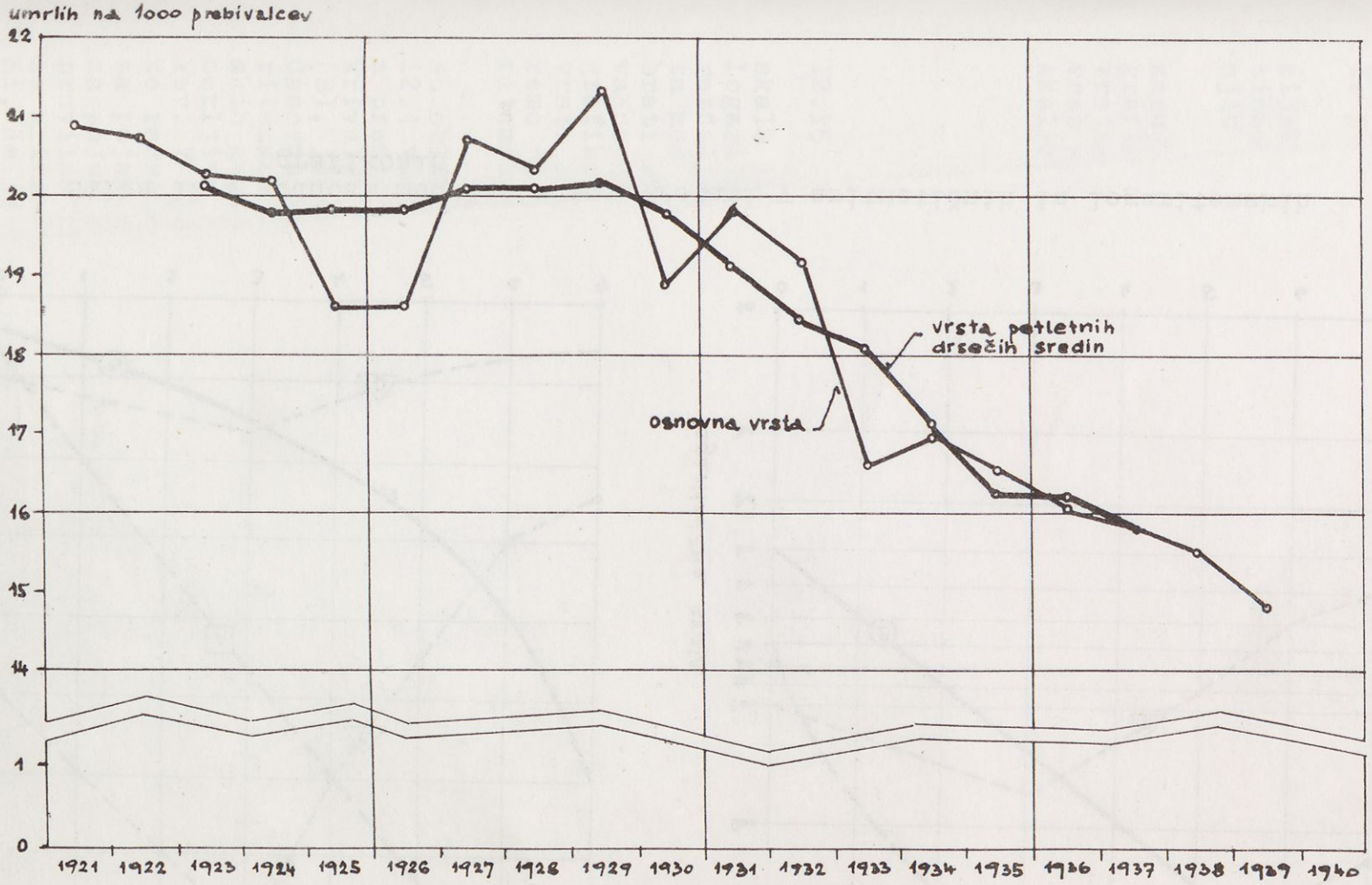
Iz obeh primerov vidimo, da je časovna vrsta drsečih sredin krajša od osnovne časovne vrste in da začetnim in končnim i členom v osnovni časovni vrsti ne moremo prirediti drseče sredine. To je hiba časovne vrste drsečih sredin, katere ne moremo odpraviti.



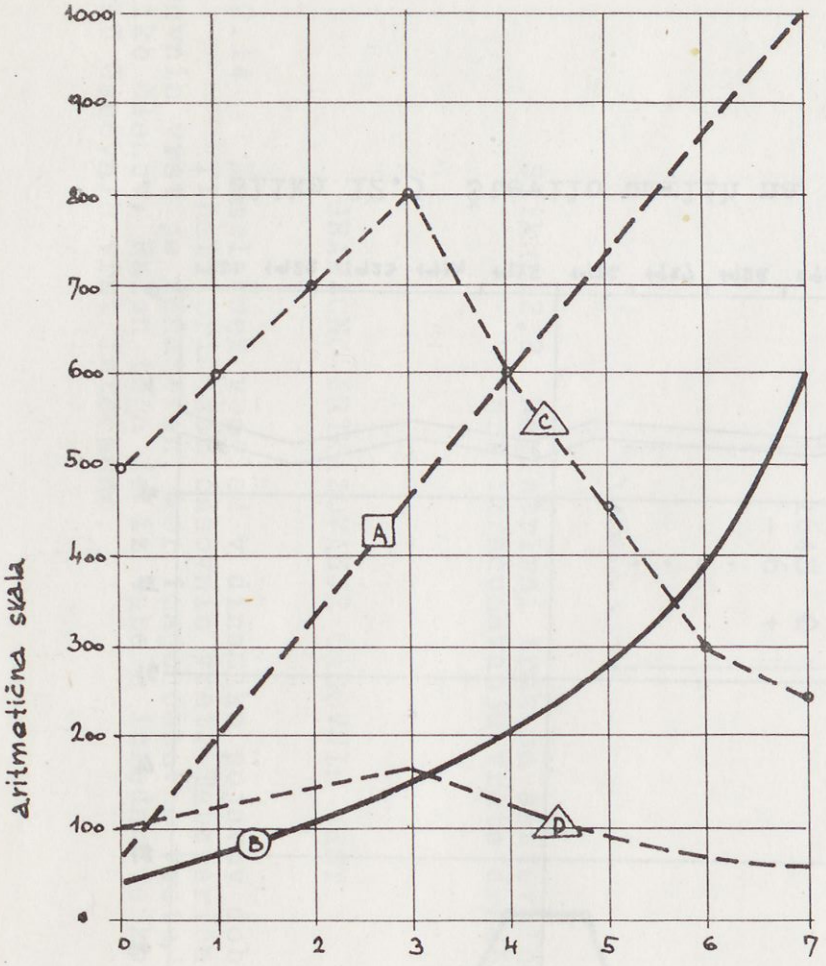
Slika 12.2 Registrirni trak na seštevalnem stroju pri izračunavanju vrste drsečih vsot S'_i

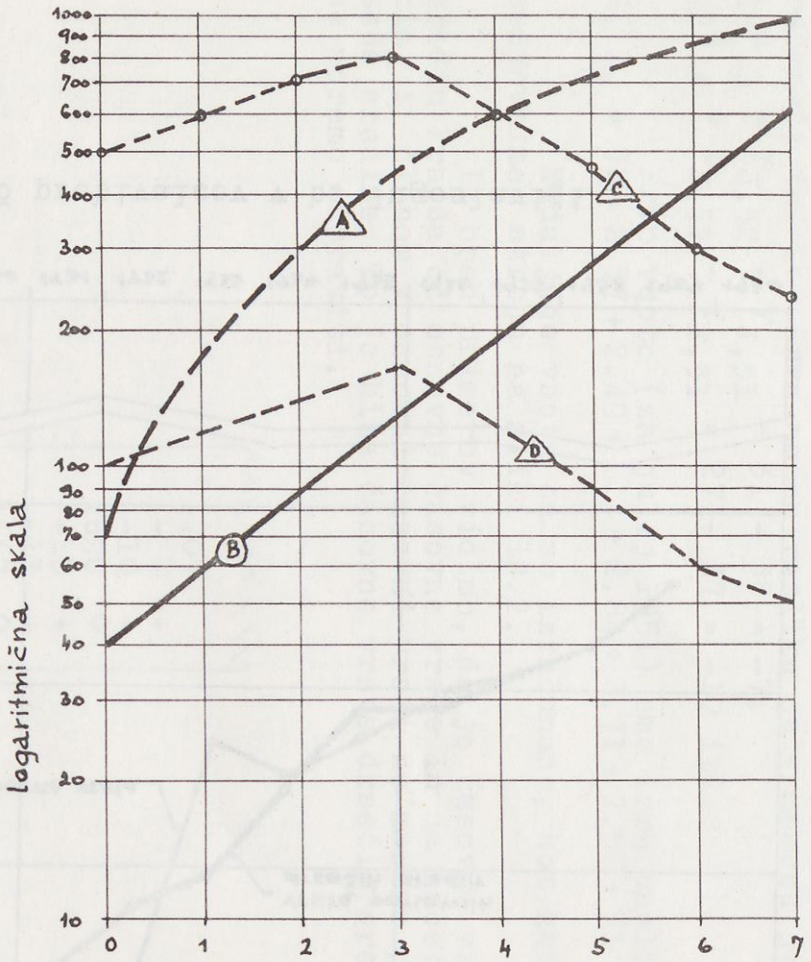
GRAFIČNO PRIKAZOVANJE ČASOVNIH VRST

12.14 Kompleksen vpogled v dinamiko pojavov dobimo z grafičnim prikazom časovnih vrst. Tabelarični prikaz časovnih vrst je nenazoren, ker imajo časovne vrste običajno veliko členov. Razen tega je iz tabele izredno težko primerjati več časovnih vrst istočasno.



Slika 12.3 Število umrlih na tisoč prebivalcev v b. Jugoslaviji





Slika 12.4 Odnosi med časovnimi vrstami v aritmetičnih in logaritemskih grafikonih

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. Eden 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 arimetično skalo pa moremo prikazovati samo istovrstne časovne vrste. Pri prikazovanju raznovrstnih časovnih vrst na istem grafikonu je namreč problematično razmerje med skalami.

Logaritemski grafikoni

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 (B), je na logaritemskem grafikonu prikazana s premico (B). 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 prikaza... razvoj proizvodnje električne energije v SFRJ in SRS.

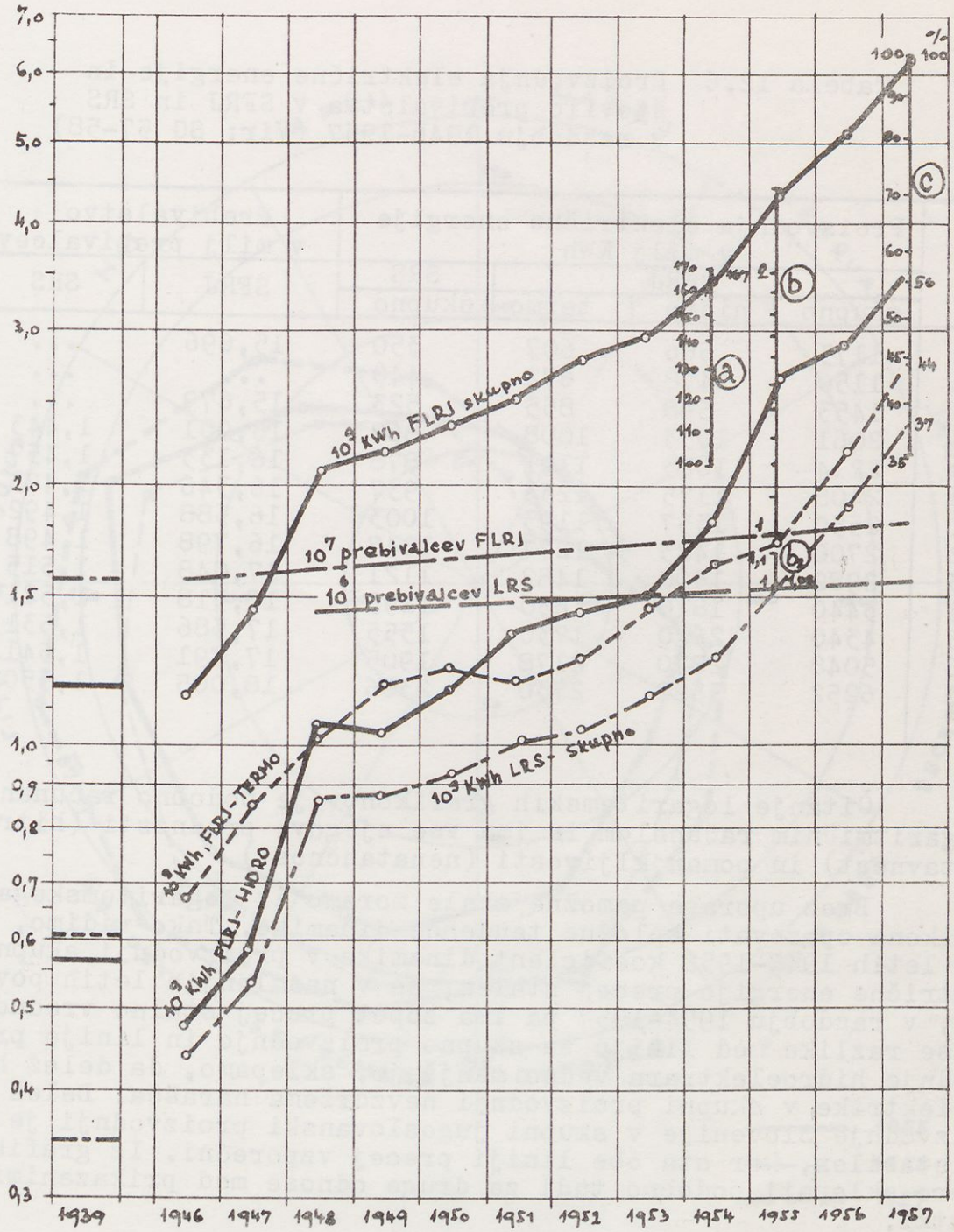
Logaritemska skala na grafikonu v sliki 12.5 je neimenovana in velja za vse prikazane časovne vrste ne glede na velikost in vrsto podatka. Za vsako vrsto posebej pa v grafikonu vpišemo decimalni faktor; z njim je treba vrednost, ki jo odberemo na skali, pomnožiti, da dobimo vrednost količine v enoti mere, ki je naznačena za decimalnim faktorjem. Iz grafikona 12.5 tako odberemo, da je bila skupna proizvodnja električne energije leta 1948 $2{,}06.10^9$ KWh ali 2,06 milijard KWh. Ker je logaritemsko skalo teže brati kakor aritmetično, logaritemske grafikone običajno proučujemo s pomožno premično skalo, ki 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 SFRJ za leto 1954 na bazo 1948 tako, da izhodišče pomožne skale 100 naravnamo na osnovo 1948. Na pomožni skali odberemo tam, kjer linija preseka skalo, indeks ca. 165 (nantančna vrednost je 167).

V primeru, ki je označen z b_1, določimo koeficient proizvodnje električne energije na enega prebivalca v SFRJ v letu 1955. Ta koeficient odberemo tako, da izhodišče 1 na pomožni skali naravnamo na linijo števila prebivalstva, na skali pa pri liniji proizvodnje odčitamo ustrezno vrednost 2,47. Vrednost koeficienta pa dobimo, če to vrednost pomnožimo s kvocientom 10^9 KWh/10^7 prebivalcev; tega dobimo iz decimalnih faktorjev, ki so vpisani ob ustreznih linijah za proizvodnjo in število prebivalstva. Koeficient je torej 247 KWh/prebiv. Analogen koeficient za isto leto za SR 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/10^6 prebivalcev. Iz grafikona ocenjeni koeficient je torej 1020 KWh/preb. v letu 1955 v SR Sloveniji.

Kako s pomožno logaritemsko skalo odčitamo strukturne odstotke, je nakazano v primeru C. Ker iščemo strukturne deleže skupne proizvodnje v SFRJ, izhodišče 100 na premični skali naravnamo v letu 1957 na skupno proizvodnjo SFRJ. Ob linijah proizvodnje pa za dele skupne proizvodnje odberemo, da je bilo od skupne proizvodnje v SFRJ v letu 1955 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.



Slika 12.5 Proizvodnja električne energije in število prebivalstva v SFRJ in SRS v razdobju 1946 - 1957

Tabela 12.6 Proizvodnja električne energije in število prebivalstva v SFRJ in SRS v razdobju 1946-1957 (Vir: SG 57-58)

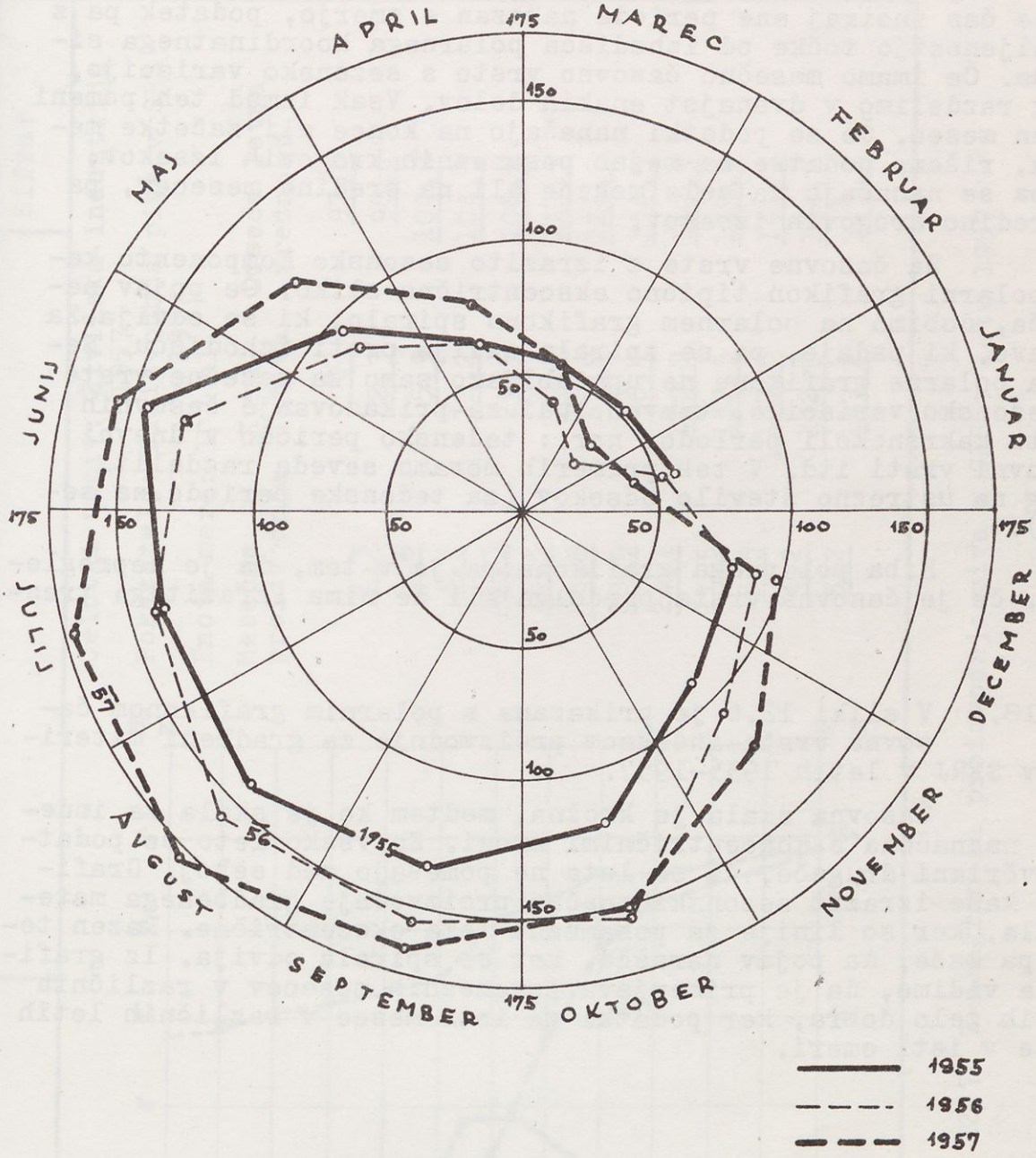
Leto	Proizvodnja električne energije v milj KWh				Prebivalstvo v milj prebivalcev	
	SFRJ			SRS	SFRJ	SRS
	skupno	hidro	termo	skupno		
1939	1173	566	607	350	15,696	...
1946	1150	478	672	440	...	...
1947	1453	598	855	523	15,679	...
1948	2061	1053	1008	867	15,901	1,443
1949	2214	1023	1191	878	16,133	1,457
1950	2408	1175	1233	937	16,346	1,473
1951	2550	1357	1193	1003	16,588	1,492
1952	2700	1423	1277	1047	16,798	1,498
1953	2982	1500	1482	1121	17,048	1,515
1954	3440	1810	1630	1293	17,318	1,521
1955	4340	2610	1730	1555	17,586	1,531
1956	5048	2870	2178	1905	17,791	1,541
1957	6252	3522	2730	2306	18,005	1,550

Čitanje logaritemskih grafikonov je podobno računanju z logaritmičnim računalom in ima vse njegove prednosti (hitrost, enostavnost) in pomanjkljivosti (nenatančnost).

Brez uporabe pomožne skale moremo iz logaritemskega grafikona opazovati splošne tendence dinamike. Tako vidimo, da je v letih 1948-1952 koeficient dinamike v proizvodnji skupne električne energije precej stalen, se v nadaljnjih letih povečuje, v razdobju 1954-1957 pa ima zopet precej stalno vrednost. Ker se razlike med linijo za skupno proizvodnjo in linijo proizvodnje hidroelektrarn vedno manjšajo, sklepamo, da delež hidroelektrike v skupni proizvodnji 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;



Slika 12.6 Indeksi proizvodnje gradbenega materiala v SFRJ v razdobju 1955-1957 (Vir: tabela 12.5)

ta zamenja pravokotni koordinatni sistem. V polarnem grafikonu je čas znotraj ene periode nakazan s smerjo, podatek pa z oddaljenostjo točke od izhodišča polarnega koordinatnega sistema. Če imamo mesečno časovno vrsto s sezonsko variacijo, krog razdelimo v dvanajst enakih delov. Vsak izmed teh pomeni po en mesec. Če se podatki nanašajo na konce ali začetke meseca, rišemo podatke na mejah 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 vrsts kakršnokoli 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.

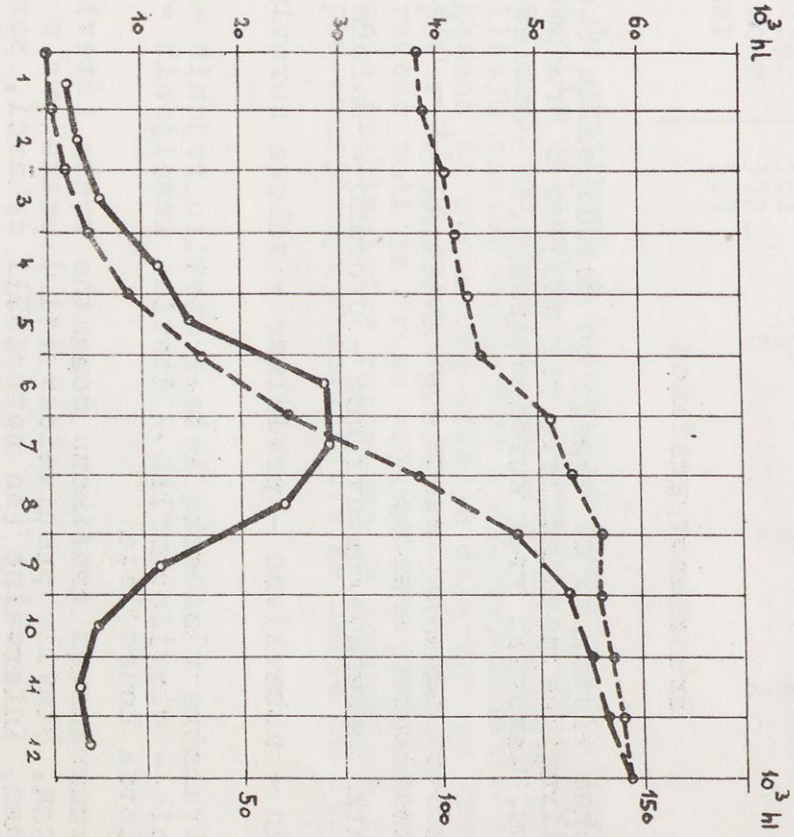
12.18. V sliki 12.6 je prikazana s polarnim grafikonom časovna vrsta indeksov proizvodnje za gradbeni material v SFRJ v letih 1955-1957.

Časovna skala je krožna, medtem ko je skala za indekse naznačena s koncentričnimi krogi. Za vsako leto so podatki včrtani drugače, da se leta ne pomešajo med seboj. Grafikon kaže izrazit sezonski značaj proizvodnje gradbenega materiala, ker so linije za posamezna leta ekscentrične. Razen tega pa kaže, da pojav narašča, ker se spirala odvija. 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 - d i a g r a m

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,

Proizvodnja piva v SRS 1957		Šifra:



Celotni naziv: Živilska industrija 1957

Pokazatelj: Proizvodnja piva v SRS 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 Z-diagram za proizvodnjo piva v SRS v letu 1957

na kateri je en del namenjen za tekoče vpisovanje 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 SRS 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 kot, moremo s polaganjem kartonov za več let drugega poleg drugega dobiti sliko dinamike za posamezni element za daljše razdobje. Če polagamo kartone za različna leta drugega nad drugega, moremo proučiti sezonsko variiranje pojava. Če pa polagamo kartone za različne elemente, ki so v vsebinski zvezi, drugega čez drugega, pa analiziramo povezave in odnose med posameznimi elementi.

Brunsmanov grafikon

12.20 Časovne vrste, ki se vežejo po metodi salda (korespondirajoče populacije), prikazujemo z Brunsmanovim grafikonom. Časovnih vrst korespondirajočih populacij je veliko:

število delavcev v začetku meseca + na novo došli - odšli = število delavcev konec meseca;

zaloge v začetku meseca + nabavljeno - prodano = zaloga konec meseca;

zaloga surovin + nabavljeno - predelano = zaloga surovin konec dekade;

število prebivalstva v začetku leta + število rojenih + število doseljenih - število umrlih - število izseljenih = število prebivalstva konec leta.

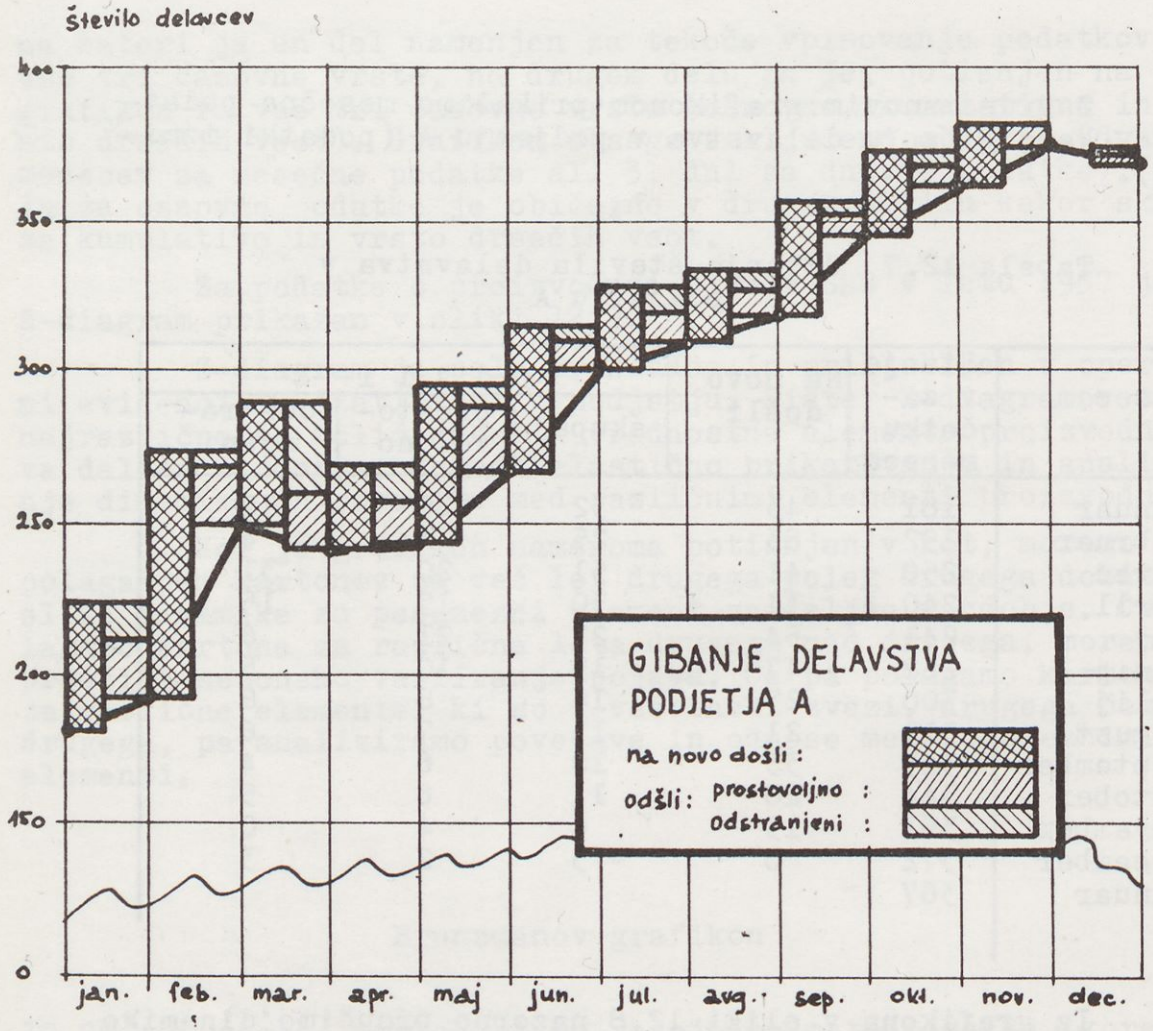
V Brunsmanovem grafikonu momentne vrste (število delavstva, zaloge, število prebivalcev itd.) prikažemo z linijskim grafikonom, intervalne (na novo došli delavci, porabljene zaloge, število izseljenih itd.) pa s stolpci. Stolpci in linije so kombinirani tako, da se vežejo med seboj.

12.21 Z Brunsmanovim 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	6
Junij	268	47	15	10	5
Julij	300	27	16	8	8
Avgust	311	21	13	7	6
September	319	35	10	6	4
Oktober	344	28	11	6	5
November	361	15	4	4	0
December	372	0	5	2	3
Januar	367				

Iz grafikona v sliki 12.8 nazorno proučimo dinamiko stanja števila delavstva in fluktuacijo števila delavstva. Brunsmanov grafikon kaže, da je bila fluktuacija delavstva v prvem polletju znatno večja kor v drugem. Dodatno je v grafikonu prikazana še struktura odišlih delavcev po vzroku odhoda. Stolpce v Brunsmanovem grafikonu uporabljamo za prikazovanje struktur tudi za druge pojave. Pri prebivalstvu stolpec za povečanje 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 Kumulativna časovna vrsta 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
		1030		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 npr. dejansko proizvedli 390 enot. Ta vrednost pade v planski kumulativi med 200 in 450. Iz tega sklepamo; če bi delali po planu, bi 390 enot delali vse prvo četrtletje, ostanek 190 enot pa bi izdelovali 190/250 del drugega četrtletja ali v odstotkih 76 % drugega četrtletja. Ganttov odstotek napišemo za ta primer v troštevilčnem številu 176. Pri tem pomeni prvo mesto število polnih razdobij, zadnji dve mesti pa odstotek naslednjega razdobja.

Iz tabele 12.8 vidimo dalje, da smo do začetka četrtega četrtletja, za katerega je Ganttov odstotek 307, izdelali količino, predvideno po planu v prvih treh četrtletjih in 7 % količine, predvidene za četrto četrtletje.

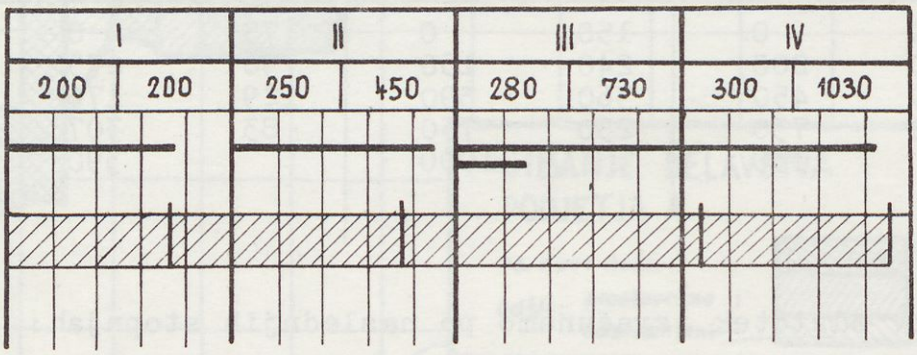
Ako s p_k zaznamujemo člene časovne vrste za plan, s P_k njegove kumulative, z d_k člene v časovni vrsti za dejansko opravljeno z D_k pa kumulativo te časovne vrste, v simbolih izrazimo izračunavanje Ganttovega odstotka takole:

a) poiščemo, med katera člena P_k in P_{k+1} v planski kumulativi pade vrednost dejanske kumulative D_r ,

$$P_k \leq D_r < P_{k+1} \qquad (12.9)$$

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 včrtujemo v grafikon z vodoravnim stolpcem. Pas se vleče skozi Ganttovemu odstotku ustrezno število celih razmakov in skozi ustrezni del načetega razmaka. Običajno razen Ganttovega odstotka rišemo v grafikone 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 postopma 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 več 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 kot 10 % zaradi vzroka, ki nima nobene zveze z rodnostjo. Te razlike so še večje pri pojavih, ki so, kakor na primer proizvodnja, odvisni ne od koledarskih, temveč od delovnih dni, ker so med meseci zaradi različnega števila nedelj in praznikov še večje variacije v številu delovnih dni. Koledar vpliva tudi na pojave, ki imajo tedensko periodičnost. Ker je na primer število prevoženih potnikov, prometnih nesreč ali nočnin v tujskem prometu največje okrog nedelje, je številčnost teh pojavov v mesecu bistveno odvisna 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 zvezni 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 SR Sloveniji v letu 1955 na enotine mesece.

Tabela 12.10 Reduciranje števila rojstev dečkov v SRS v letu 1955 na enotine mesece (Vir: MSP)

Mesec	Število rojstev Y	Število dni	Popravni faktor F	Popravljeno število rojstev $YF = Y_f$
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 meja, 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 po 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 itd. 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 itd.

Vpliv sprememb v spremenjenem številu prebivalstva odstranimo tako, da izračunavamo koeficiente "na prebivalca", "na delavca" itd.

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 natančno 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 izredno zamuden in težko izvedljiv, ker bi ga morali upoštevati že v osnovni evidenci. Vpliv cen popravimo tudi z indeksi cen tako, da podatke po tekočih cenah delimo z ustreznim indeksom cen. Ta način je operativno veliko lažji, je pa kljub temu, da je le ocena, zadovoljiv. 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.

$$I_{k/o} = 100\ Y_k/Y_o \qquad (12.11)$$

Absolutna razlika

$$D_k = Y_k - Y_{k-1} \qquad (12.12)$$

pokaže v absolutnih vrednostih spremembo pojava od člena do člena.

Tempo 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

$$I_k = 100 \cdot Y_k/Y_{k-1} \qquad (12.15)$$

le da relativno spremembo nakaže v indeksu.

Ker so vsi ti pokazatelji izpeljani iz osnovne časovne vrste, so med njimi enostavne zveze.

$$T_k = 100 \cdot D_k/Y_{k-1} = 100(k_k - 1) = I_k - 100$$
$$I_k = 100 \cdot k_k \qquad (12.16)$$

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	$D_k > 0$	$D_k = 0$	$D_k < 0$
Tempo 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. Bolj smiselno je, da primerjamo podatek tekočega meseca s podatkom iz predhodnega leta za isti mesec. Ta dva podatka sta si sorodnejša kakor zaporedna meseca. Ti indeksi imajo premično bazo, vendar niso verižni, ker osnova indeksov ni predhodni člen.

Tabela 12.13 Mesečni in kumulativni indeksi za proizvodnjo piva v letu 1957 v SRS (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	6,7	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	38,6	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 ta e primere izračunavamo indekse iz kumulativnih vrednosti dveh zaporednih let. Ta primerjava je logična, ker obe primerjani vrednosti obsegata isti del sezone v dveh zaporednih letih. Za proizvodnjo piva v SRS 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 1956 je 216). 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 spreminjata. 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 (npr. 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 nepričekovano, 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 ne 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.

$$\begin{aligned} &1)\ Y = T + P + C + E + S \\ &2)\ Y = T\,(1 + p + c + e + s) \\ &3)\ Y = T \cdot P \cdot C \cdot E \cdot S \\ &4)\ Y = T \cdot P \cdot C \cdot E + S \end{aligned} \qquad (12.17)$$

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 posamznih 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 zakonitosti dinamike pojava. Te zakonitosti izkoriščamo, kadar analiziramo dogajanja v preteklosti ali napovedujemo razvoj v prihodnosti. Z različnimi analitičnimi postopki namreč moremo določiti posamezno izmed teh 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 povprečje.

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, 6, 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 katero računamo povprečje, enaka 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.

T R E N D

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 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 predpisujejo, 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: $T = k + ab^x$
6. 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 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 krivilje najbolje ustreza stvarni smeri razvoja.

O obliki trenda se moremo poučiti tudi iz vrsti 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 prva diferenca 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

$$\log T = \log a + x \log b \qquad (12.19)$$

kar moremo pisati

$$T' = a' + b'x \qquad (12.20)$$

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

$$\frac{1}{T} = \frac{1}{T_\infty} + \frac{e^a}{T_\infty} e^{bx} \qquad (12.21)$$

ali dalje:

$$T' = k' + a'b'^x \qquad (12.22)$$

Pri tem pomeni: $T' = \frac{1}{T}$; $a' = \frac{e^a}{T_\infty}$; $b' = e^b$;

$$k' = \frac{1}{T_\infty}$$

Če logaritmiramo Gompertzovo funkcijo, pa dobimo:

$$\log T = \log K + b^x \log a \qquad (12.23)$$

ali

$$T' = k' + a'b'^x \qquad (12.24)$$

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 raz-

dobje, za katero 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 včrtati v grafikonu prostoročno. Ta metoda je subjektivna in zato nima posebne analitične vrednosti. Vendar je njena prednost v enostavnosti in hitrosti. Zato jo uporabljamo predvsem za približek in za osnovo pri izbiri analitične oblike trenda.

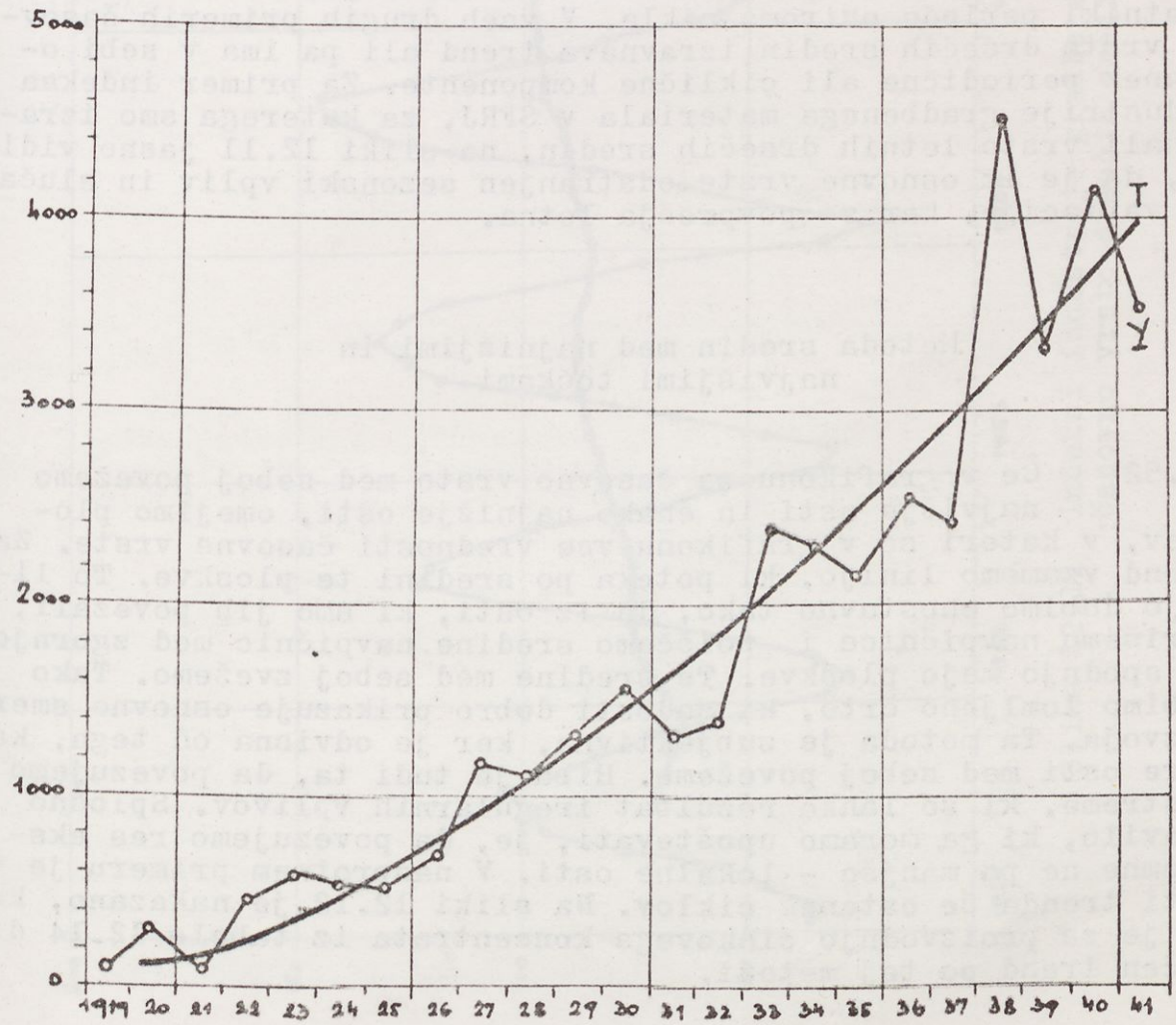
12.50 V sliki 12.10 je narisana časovna vrsta proizvodnje cinkovega koncentrata na območju SRS v razdobju med obema vojnama od 1919-1941, v njej pa prostoročno včrtan trend.

Tabela 12.14 Proizvodnja cinkovega koncentrata v SRS med obema vojnama v tonah (Vir: SB SRS 1952, št. 8)

Leto	1919	1920	1921	1922	1923	1924	1925	1926	1927
Proiz-vodnja	90	350	75	450	600	520	510	639	1143

Leto	1928	1929	1930	1931	1932	1933	1934	1935	1936
Proiz-vodnja	1076	1313	1547	1301	1358	2393	2282	2173	2567

Leto	1937	1938	1939	1940	1941
Proiz-vodnja	2401	4530	3252	4121	3483



Slika 12.10 Trend za proizvodnjo cinkovega koncentrata v SRS med obema vojnama, določen prostoročno

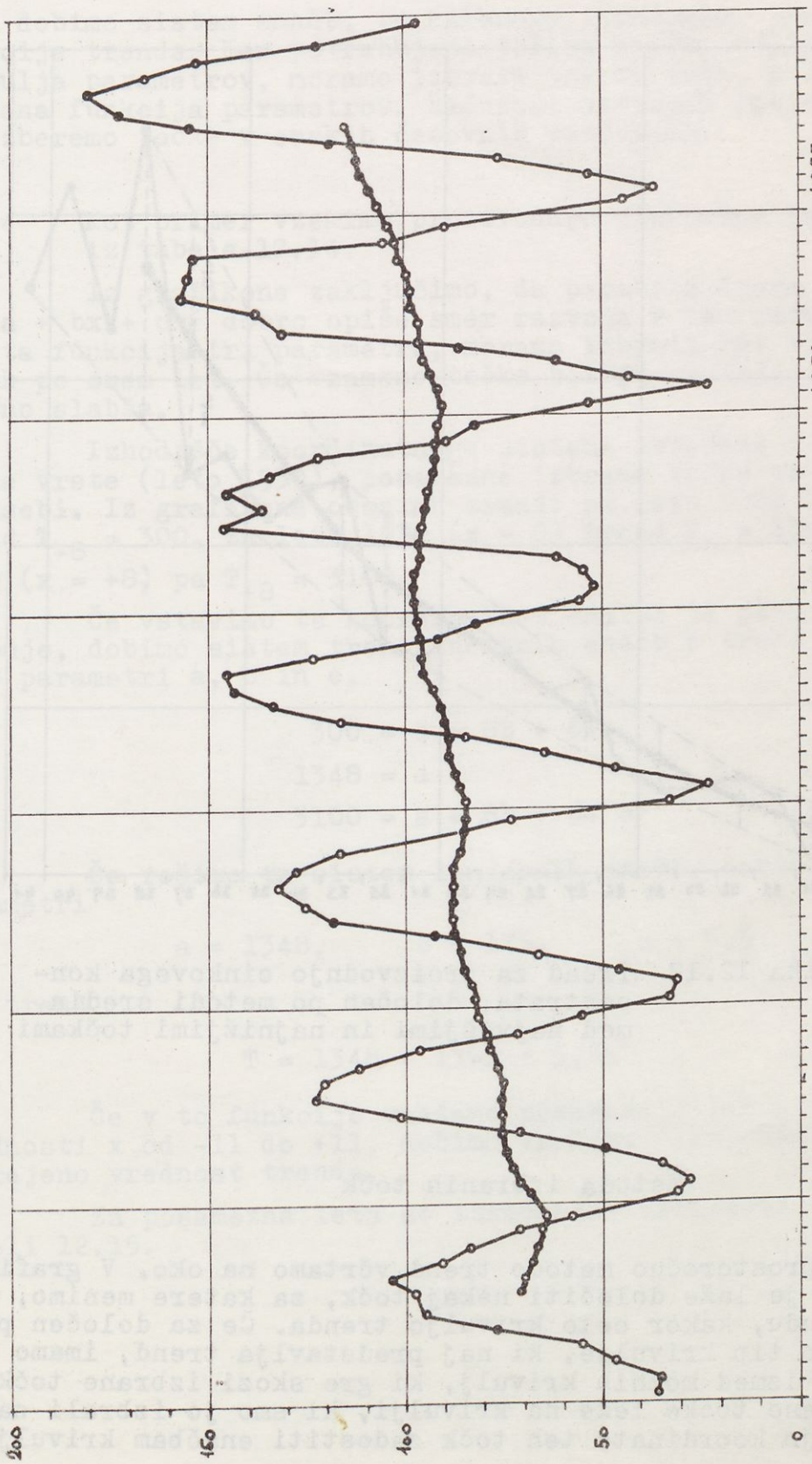
Metoda drsečih sredin

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.

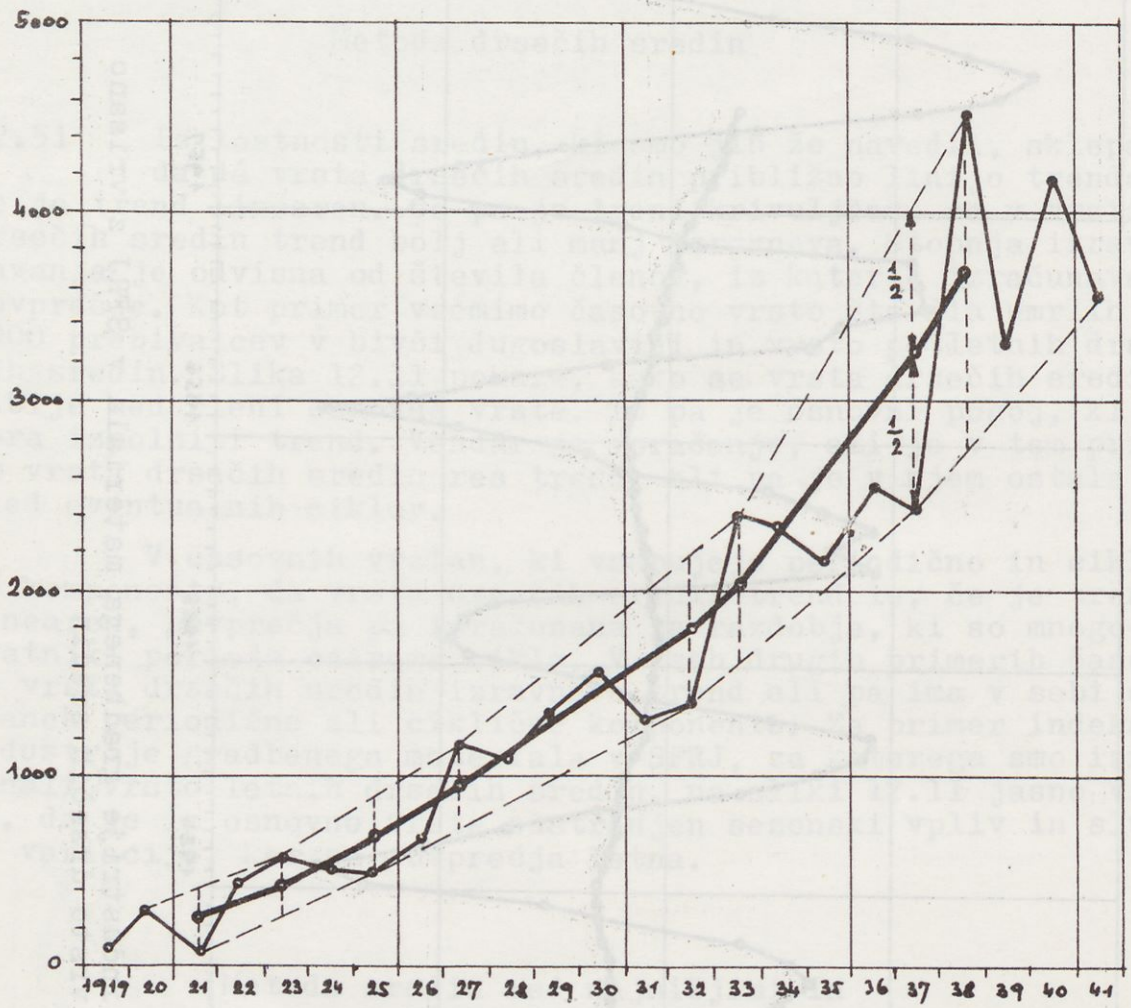
V časovnih vrstah, ki vsebujejo periodično in ciklično komponento, da vrsta drsečih sredin trend le, če je trend linearen, povprečja pa izračunana za razdobja, ki so mnogokratniki periode oziroma cikla. V vseh drugih primerih časovna vrsta drsečih sredin izravnava trend ali pa ima v sebi ostanek periodične ali ciklične komponente. Za primer indeksa industrije gradbenega materiala v SFRJ, 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 i poiščemo sredine navpičnic med zgornjo in spodnjo mejo 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.



Slika 12.11 Indeksi industrije gradbenega materiala v SFRJ z vrisano časovno vrsto drsečih sredin



Slika 12.12 Trend za proizvodnjo cinkovega koncentrata, določen po metodi sredin med najvišjimi in najnižjimi točkami

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.

$$300 = a - 8b + 64c$$
$$1348 = a$$
$$3100 = a + 8b + 64\,c$$

Če rešimo ta sistem linearnih enačb, dobimo, da so parametri

$$a = 1348, \qquad b = 175, \qquad c = 5{,}5$$

ali trend

$$T = 1348 + 175x + 5{,}5x^2$$

Če v to funkcijo vnašamo posameznim letom 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
x	-11	-10	- 9	- 8	- 7	- 6	- 5	- 4	- 3	- 2	- 1
T	89	148	219	300	392	496	611	738	872	1020	1178

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	1348	1528	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 objektivneje 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 povprečje. Za točke, ki imajo za absciso sredino časovnega razmaka, iz katerega izračunamo povprečje, za ordinato pa izračunamo 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 SRS, 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 $I_{-8}(-8,371)$, $T_0(0,1462)$, $T_{+8}(+8,3218)$ vnesemo v funkcijo

$$T = a + bx + cx^2.$$

Tako dobimo tri enačbe:

$$\begin{aligned} 371 &= a - 8b + 64c \\ 1462 &= a \\ 3218 &= a + 8b + 64c \end{aligned}$$

Iz tega sistema enačb zlahka dobimo, da je: a = 1462, 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
x	-11	-10	- 9	- 8	- 7	- 6	- 5	- 4	- 3	- 2	- 1
T	133	202	281	371	471	581	702	833	975	1127	1289

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
	1462	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 razultati 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 ... parametri, f_1, f_2, f_3 pa funkcije časa, dobimo sistem linearnih enačb za parametre a, b, c ... če vzamemo delne vsote členov iz časovne vrste.

Število vsot se ravna po številu parametrov. Če imamo na primer tri parametre, dobimo, da je

$$\begin{aligned} \underset{1}{\Sigma Y} &= a\underset{1}{\Sigma f_1} + b\underset{1}{\Sigma f_2} + c\underset{1}{\Sigma f_3} \\ \underset{2}{\Sigma Y} &= a\underset{2}{\Sigma f_1} + b\underset{2}{\Sigma f_2} + c\underset{2}{\Sigma f_3} \qquad (12.25) \\ \underset{3}{\Sigma Y} &= a\underset{3}{\Sigma f_1} + b\underset{3}{\Sigma f_2} + c\underset{3}{\Sigma f_3} \end{aligned}$$

Pri tem pomeni: $\underset{1}{\Sigma Y}$, $\underset{2}{\Sigma Y}$, $\underset{3}{\Sigma Y}$ = delne vsote členov v časovni vrsti

$\underset{1}{\Sigma f_1}$, $\underset{2}{\Sigma f_2}$ ... $\underset{3}{\Sigma f_3}$ = 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: $f_1 = 1$, $f_2 = x$, $f_3 = x^2$.

Izračun parametrov in vrste trenda je nakazan v tabeli 12.17.

Tabela 12.17 Izračunavanje paraboličnega trenda druge stopnje za proizvodnjo cinkovega koncentrata v SRS med obema vojnama, po metodi delnih vsot

Leto	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	123,723	-1237,23	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	447
1924	- 6	36	520	145,419	- 872,51	553
1925	- 5	25	510	150,843	- 754,21	671
Σ 7 1	-56	476	2595	-	-	2595
1926	- 4		639	156,268	- 625,07	801
1927	- 3	9	1143	161,692	- 485,08	941
1928	- 2	4	1076	167,116	- 334,23	1091
1929	- 1	1	1313	172,540	- 172,54	1253
1930	0	0	1547	177,964	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
Σ 7 2	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	3747
1941	+11	121	3483	237,629	+2613,92	4039
Σ 7 3	+56	476	22527	-	-	22526

Kakor pri prejšnji metodi smo tudi tu izpustili leti 1926 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:

$$a = 1425{,}59,\ b = 177{,}964,\ c = 5{,}4241$$

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 bolj 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

$$\Sigma(Y - T)^2 = \text{Min} \qquad (12.26)$$

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

$$T = af_1 + bf_2 + cf_3 \ \ldots \qquad (12.27)$$

f_1, f_2, f_3 = funkcije x,

pogoj iz obrazca 12.26 napišemo v konkretnejši obliki

$$\Sigma(Y - af_1 - bf_2 - cf_3 \ \ldots)^2 = \text{Min} \qquad (12.28)$$

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:

$$\Sigma Yf_1 = a\Sigma f_1^2 + b\Sigma f_2 f_1 + c\Sigma f_3 f_1$$
$$\Sigma Yf_2 = a\Sigma f_1 f_2 + b\Sigma f_2^2 + c\Sigma f_3 f_2 \qquad (12.29)$$
$$\Sigma Yf_3 = a\Sigma f_1 f_3 + b\Sigma f_2 f_3 + c\Sigma f_3^2$$

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

$$\Sigma Y = Na + b\Sigma x \qquad (12.30)$$
$$\Sigma Yx = a\Sigma x + b\Sigma x^2$$

Če je trend parabola druge stopnje $T = a + bx + cx^2$, je sistem normalnih enačb

$$\Sigma Y = aN + b\Sigma x + c\Sigma x^2$$
$$\Sigma Yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \qquad (12.31)$$
$$\Sigma Yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

Če pa je trend parabola tretje stopnje $T = a + bx + cx^2 + dx^3$, dobimo tele normalne enačbe

$$\Sigma Y = aN + b\Sigma x + c\Sigma x^2 + d\Sigma x^3$$
$$\Sigma Yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + d\Sigma x^4 \qquad (12.32)$$
$$\Sigma Yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + d\Sigma x^5$$
$$\Sigma Yx^3 = a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + d\Sigma x^6$$

Kakšne so normalne enačbe, če je trend parabola višje stopnje kot tretje, je razvidno iz teh 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 do tretje stopnje s konstantami iz tabele G. Postopek je tale:

a) iz osnovne vrste izračunamo izraza

$$\Sigma Y, \quad \Sigma Yx, \quad \Sigma Yx^2 \quad \text{in} \quad \Sigma Yx^3$$

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:

$$\begin{aligned} a_2 &= A_2 \Sigma Y - K_2 \Sigma Yx^2; \\ b_2 &= B_2 \Sigma Yx; \\ c_2 &= C_2 \Sigma Yx^2 - K_2 \Sigma Y \end{aligned} \tag{12.36}$$

Za paraboličen trend tretje stopnje

$$T = a_3 + b_3 x + c_3 x^2 + d_3 x^3 \tag{12.37}$$

so parametri:

$$\begin{aligned} a_3 &= A_3 \Sigma Y - K_2 \Sigma Yx^2; \\ b_3 &= B_3 \Sigma Yx - K_3 \Sigma Yx^3; \\ c_3 &= C_3 \Sigma Yx^2 - K_2 \Sigma Y; \\ d_3 &= D_3 \Sigma Yx^3 - K_3 \Sigma Yx \end{aligned} \tag{12.38}$$

12.62 Za proizvodnjo cinkovega koncentrata v obdobju med obema vojnama je treba izračunati po metodi najmanjših kvadratov paraboličen trend druge stopnje.

Trend izra unamo 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 $T = a_2 + b_2x + c_2x^2$. Iz te enačbe na znan način izračunamo vrednosti trenda, ki ustrezajo posameznim členom tako, da v $T = a_2 + b_2x + c_2x^2$ vnašamo ustrezne vrednosti za x.

V tabeli G najdemo, da je za N = 23:

$A_2 = 0{,}981366_{-1}$; $B_2 = 0{,}988142_{-3}$; $C_2 = 0{,}282326_{-4}$;

$K_2 = 0{,}124224_{-2}$

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	-2146	4292
1929	- 1	1313	-1313	1313
1930	0	1547	0	0
1931	+ 1	1301	+1301	1301
1932	+ 2	1358	+2713	5432
1933	+ 3	3293	+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	+180050	1882894
		ΣY	ΣYx	ΣYx^2

Po obrazcih za izračunavanje parametrov za parabolični trend dobimo:

$a_2 = A_2\Sigma Y - K_2\Sigma Yx^2 = 0,981366.10^{-1}.38071 - 0,124224.10^{-2}.$
$. 1882894 = 1397,15$

$b_2 = B_y\Sigma Yx = 0,988142.10^{-3}.180050 = 177,915$

$c_2 = C_2\Sigma Yx^2 - K_2\Sigma Y = 0.282326.10^{-4}.1882894 - 0.124224.10^{-2}.$
$. 38071 = 5,86567$

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 12.19 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	2676	2930	3196	3474	3763	4064

Pri kontroli o pravilnosti izračuna trenda upoštevamo pravilo, da je vsota členov v osnovni vrsti enaka vsoti trenda. Vsota trenda je $\Sigma T = 38072$, vsota v osnovni vrsti pa $\Sigma Y = 38071$. Razlika izvira iz zaokroževanja.

12.63 Za proizvodnjo svinčeno-cinkove rude v SR Sloveniji v razdobju po osvoboditvi od 1946-1957 je treba določiti smer razvoja. Osnovna vrsta podatkov je dana v tabeli 12.19.

Tabela 12.19 Proizvodnja svinčeno-cinkove rude v SR Sloveniji v 10^3 tonah (Vir:MSP SRS)

Leto	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957
Proizvodnja	133	145	211	226	232	231	238	312	347	419	455	461

Analiza osnovnih podatkov pokaže, da imajo podatki linearno smer, če so narisani na pollogaritmičnem grafikonu. Iz tega sklepamo, da je smer razvoja eksponencialna funkcija obli-

ke $T = ab^x$. Iz odstavka o primernih funkcijah trenda povezamemo, 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 logY. Ta časovna vrsta kaže linearno smer razvoja.

Vrsti logY prilagodimo po metodi najmanjših kvadratov premico. Iz enačbe vidimo, da sta parametra te premice logaritma parametrov eksponencialnega trenda. Z antilogaritmiranjem vrste logT dobimo časovno vrsto trenda za proizvodnjo svinčeno-cinkove rude.

Ker je število členov sodo, vzamemo za x vrsto:
-11, -9 ... -3, +1, +3 ... +9, +11.

Tabela 12.20 Izračunavanje eksponencialnega trenda za proizvodnjo svinčeno-cinkove rude v SR Sloveniji

Leto (1)	Y (2)	log Y (3)	x (4)	x.logY (5)=(3).(4)	logT = loga+xlogb (6)	T (7)
1946	133	2,1239	-11	-23,3629	2,1548	143
1947	145	2,1614	- 9	-19,4526	2,2031	160
1948	211	2,3243	- 7	-16,2701	2,2515	178
1949	226	2,3541	- 5	-11,7705	2,2998	199
1950	232	2,3655	- 3	- 7,0965	2,3481	223
1951	231	2,3636	- 1	- 2,3636	2,3965	249
1952	238	2,3766	+ 1	+ 2,3766	2,4448	278
1953	312	2,4942	+ 3	+ 7,4826	2,4932	311
1954	346	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	483
Vsota		29,0479		+13,8226	29,0478	
		ΣlogY		Σx.logY	ΣlogT	

Parametre izračunamo po tabeli G. Ker je logaritmični trend linearen, poiščemo pod N = 12 : $A_1 = 0{,}83333_{-1}$; $B_1 = 0{,}174825_{-2}$. Dalje sledi:

$$\text{Loga} = A_1 \Sigma \log Y = 0{,}833333.10^{-1} \cdot 29{,}0478 = 2{,}42065$$

$$\text{Logb} = B_1 \Sigma x.\log Y = 0{,}174825.10^{-2} \cdot 13{,}8226 = 0{,}02417.$$

Če te rezultate antilogaritmiramo, dobimo:

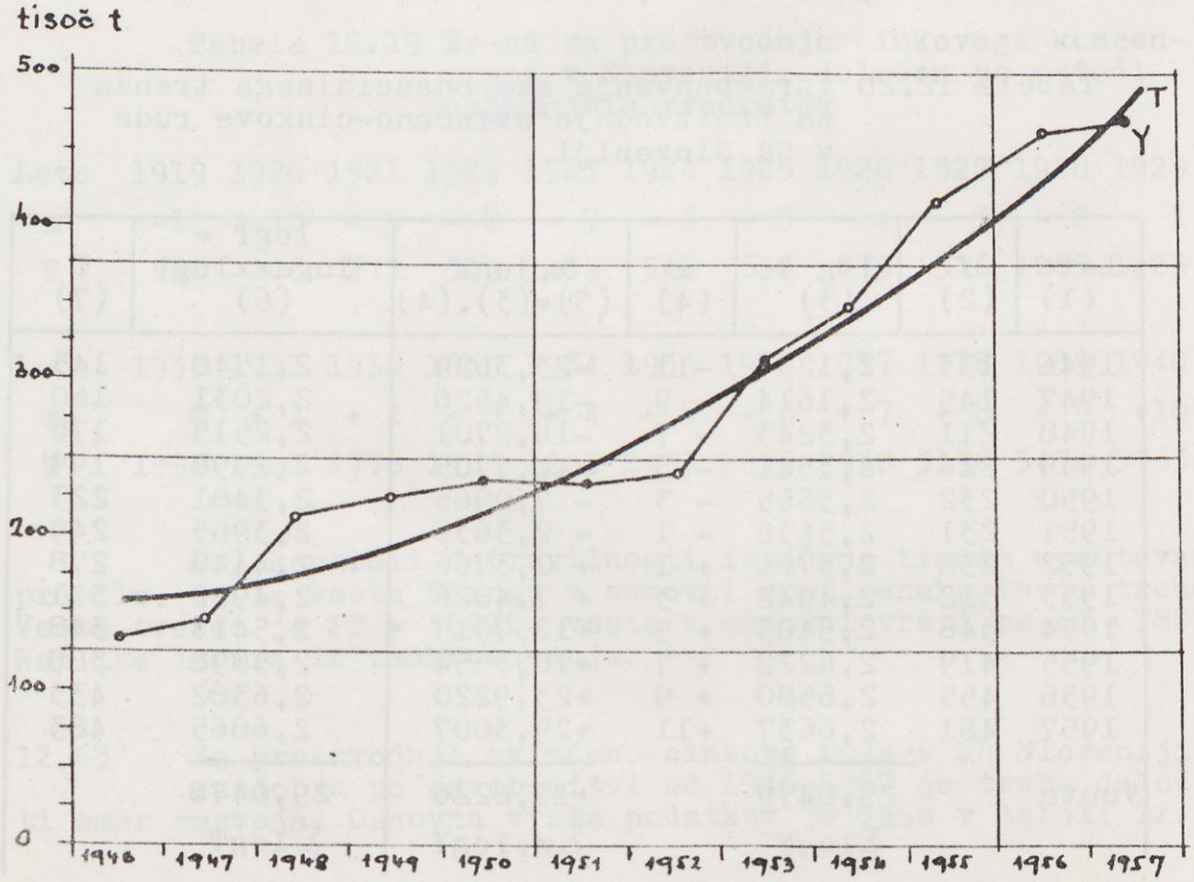
$$a = 263{,}42 \qquad b = 1{,}0572$$

$$T = a.b^{x} = 263{,}42 \; . \; 1{,}0572^{x}$$

a = 263,42 pomeni vrednost trenda sredi vrste, $b^2 = 1{,}1177$ pa je povprečni koeficient dinamike za proizvodnjo svinčeno-cinkove rude v razdobju 1946-1957.

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 SRS, 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 variacija 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 periodični vplivov je v tem, da se na določeno časovno razdovje (npr. leto, mesec, teden, dan) ponavljajo; to izzove na proučevanih pojavih periodično vedno enak učinek.

V splošnem periodični vplivi v gospodarstvu niso zaželeni. Periodičen značaj dovede ali do zastoja v času sezone ali do neizkoriščenja kapacitet izven sezone, če imamo zadostne kapacitete za čas maksima. Zato si prizadevamo, da periodične vplive omilimo, če jih že ne moremo odpraviti. Tipičen primer je nadomeščanje proizvodnje hidro in termo elektrike. Obe se medsebojno dopolnjujeta tako, da čimbolj 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 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,
d) metodo kvocientov na vrsto drsečih sredin,
e) metodo grafičnega približka.

M e t o d a v s o t

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 vsto 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

$$\bar{S} = KA \qquad (12.40)$$

Narodna in univerzitetna knjižnica LJUBLJANA

c) Če mesečne vsote S delimo s povprečno vsoto $\bar{S}$, dobimo sezonske indekse

$$S/\bar{S} = KA(1+-)/KA = 1+p \qquad (12.41)$$

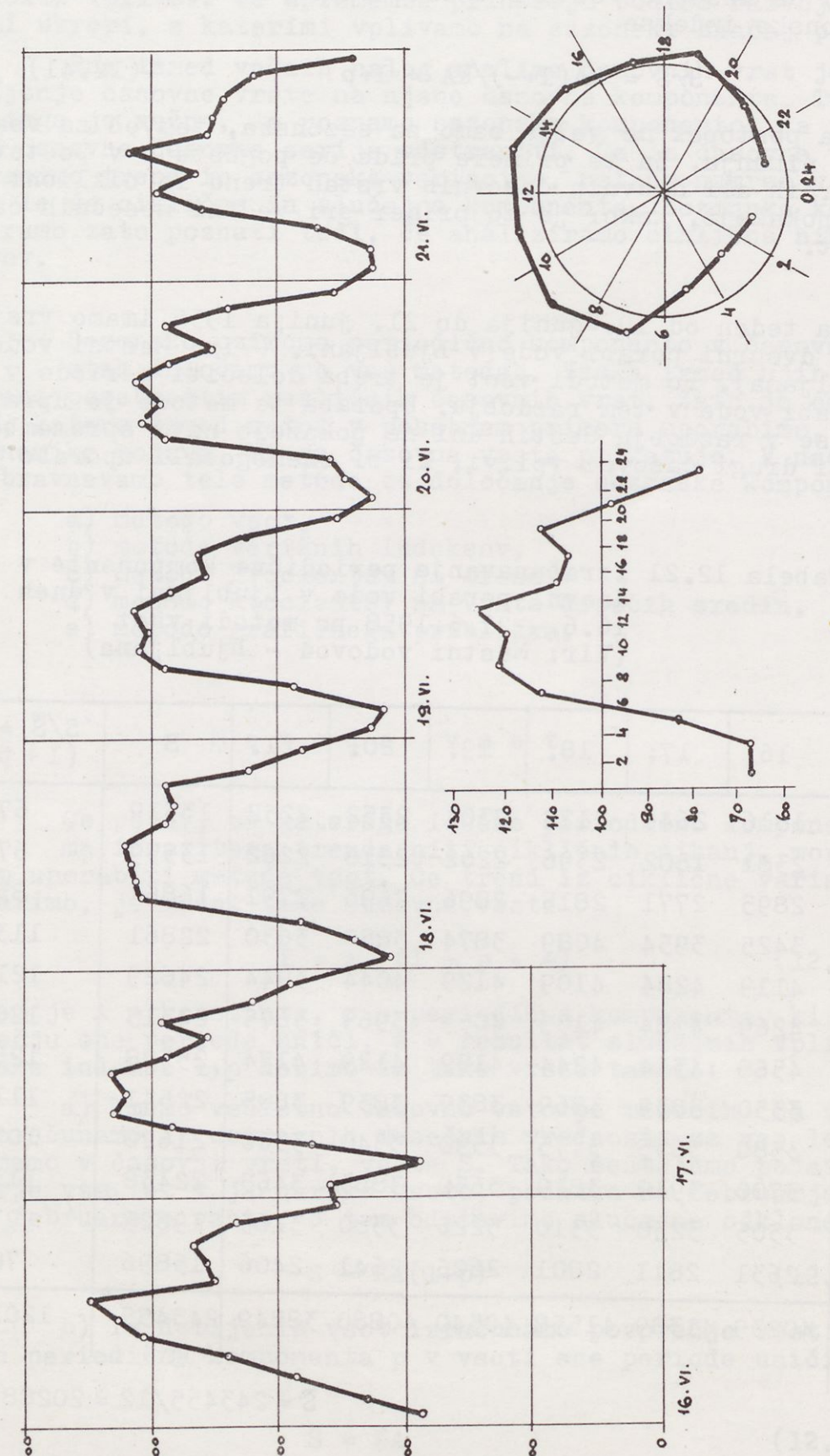
Ta postopek ne velja samo za sezonska, temveč za vsa periodična nihanja. Za te primere pride še pogosteje v poštev, ker se v urnih ali dnevnih časovnih vrstah trend in ciklični vplivi ne pokažejo, kakor je to primer pri večini mesečnih časovnih vrst.

12.67 Za teden od 19. 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.

Tabela 12.21 Izračunavanje periodične komponente v dnevni porabi vode v Ljubljani v dneh 16.6.- 21.6.1958 po metodi vsot (Vir: Mestni vodovod - Ljubljana)

Datum / Ura	16.	17.	18.	19.	20.	21.	S	$S/\bar{S}$ = (1 + p)
0^h- 2^h	1816	2646	2172	2301	2352	2232	13519	67
2^h- 4^h	2381	1802	2396	2202	2516	2262	13559	67
4^h- 6^h	2895	2771	2815	2896	2696	2791	16864	83
6^h- 8^h	3425	3954	4089	3874	3889	3630	22861	113
8^h-10h	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	3650	3885	3465	22438	111
20^h-22^h	3305	3230	3310	3225	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



Slika 12.14 Periodična komponenta v dnevni porabi vode v Ljubljani za razdobje 16.6.-21.6.1958

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

12.63 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 vplivov. 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 dá 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 drsečimi 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 $\bar{\bar{K}} = \Sigma K/12$. 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 SFRJ 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.

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 teh odsekih seveda ne moremo izračunati kvocientov. V tabeli 12.12 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 SFRJ 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 mastopiti 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 $\bar{K}_i$: npr. januar: $\bar{K}_1 = 1/4(44,6+41,7+55,8+40,6) = 1827/4 = 45,7$.

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	37,9	49,3	67,1	94,5	134,7	~~165~~,3	~~161~~,8	147,0	126,6	~~89~~,9	~~69~~,4
1953	41,7	40,3	~~57~~,9	~~80~~,7	~~107~~,2	136,5	142,4	150,0	146,3	136,2	99,7	86,1
1954	~~37~~,1	~~25~~,3	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~~,3	99,2	~~148~~,5	141,6	149,6	142,4	128,2	94,7	~~87~~,0
1956	~~56~~,2	28,0	~~46~~,6	63,1	~~88~~,1	131,3	139,3	157,2	~~153~~,8	151,1	102,1	83,9
1957	40,6	34,3	48,2	67,6	105,3	135,0						
ΣK	182,7	140,5	206,7	270,5	393,9	537,5	569,8	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

Vsota: 1193,3; povprečje $\bar{\bar{K}} = 99,4$

$1+p_i$	46	35	52	68	99	135	143	153	148	136	98	86

Vsota: 1199, povprečje: 100

5) Iz dobljenih mesečnih kvocientov $\bar{K}_i$ izračunamo povprečje $\bar{\bar{K}}$. Če bi bili povprečni mesečni kvocienti pravi sezonski indeksi, bi bilo povprečje teh kvocientov enako 100. Izračunano povprečje je v našem primeru $\bar{\bar{K}} = 99{,}68$. Razlika torej ni velika; zato tudi popravljena vrsta sezonskih indeksov ni bistveno različna od nje.

6) Povprečne mesečne kvociente $\bar{K}_i$ popravimo tako, da jih delimo s povprečjem povprečnih mesečnih kvocientov $\bar{\bar{K}}$ $1+p_i = \bar{K}_i/\bar{\bar{K}}$.

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

$$Y_x = a.b^x(1+p+e) \qquad (12.46)$$

Verižni kvocienti za ta tip časovne vrste so:

$$I_x = Y_x/Y_{x-1} = \frac{ab^x(1+p_i+e_1)}{ab^{x-1}(1+p_{i-1}+e_o)} = b\,\frac{1+p_i}{1+p_{i-1}} + e' \qquad (12.47)$$

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

$$\bar{I}_i = b\,\frac{1+p_i}{1+p_{i-1}} \qquad (12.48)$$

Zaradi lastnosti kvocientov sezonskih indeksov je geometrijska sredina povprečnih mesečnih indeksov enaka b.

$$G_I = \sqrt[12]{b\,\frac{1+p_1}{1+p_{12}} \cdot b\,\frac{1+p_2}{1+p_1} \cdots b\,\frac{1+p_{11}}{1+p_{10}}\; b\,\frac{1+p_{12}}{1+p_{11}}} = b \qquad (12.49)$$

S kvocientom povprečnega mesečnega indeksa $\bar{I}_i$ in geometrijske sredine dobimo čiste kvociente sezonskih indeksov

$$\bar{I}_i / G_I = b\,\frac{1+p_i}{1+p_{i-1}} \Big/ b = \frac{1+p_i}{1+p_{i-1}} = t_i \qquad (12.50)$$

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 $\bar{K} = 1/1+p_{12}$, prave sezonske indekse $1+p_i$ dobimo

$$K_i / \bar{K} = \frac{1+p_i}{1+p_{12}} \Big/ \frac{1}{1+p_{12}} = 1+p_i \qquad (12.52)$$

če K_i delimo s povprečnim $\bar{K}$.

12.72 Časovna vrsta indeksov industrije gradbenega materiala kaže smer razvoja, ki jo moremo v približku vzeti kot eksponencialno. Ker dá 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 SFRJ 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) itd.

Verižne indekse za januarje naslednjih let dobimo z deljenjem januarskih vrednosti z decembrskimi vrednostmi prejšnjega leta (npr. $I_{jan.1952}$ = 100.31/58 = 53,4 itd.)

b) Izračunamo povprečja verižnih indeksov $\bar{I}_i$ 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 $\bar{I}_i$

$$G_I = \sqrt[12]{56,9 \cdot 86,4 \cdot 144,5 \ldots 74,5 \cdot 84,6} = 101,06$$

G_I izračunamo z logaritmi.

d) Povprečne mesečne verižne indekse $\bar{I}_i$ delimo z $G_I = 101,06$, da odstranimo parameter b. Tako dobimo kvociente sezonskih indeksov $t_i = \bar{I}_i / G_I$. 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.

$K_1 = t_1 = 56,3$; $K_2 = K_1 t_2 = 56,3 \cdot 0.855 = 48,1$ ... itd.

Kontrola: $K_{12} = 100,0$.

f) Izračunamo povprečje iz K_i.

$$\bar{K} = \frac{1}{12}(56,3 + 48,1 + \ldots + 119,5 + 100,0) =$$

$$= \frac{1505,9}{12} = 125,5$$

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	56	78	95	97	104	90	84	71	58
1952	31	27	36	50	71	101	124	122	112	98	71	56
1953	34	36	48	68	92	119	125	131	127	118	86	74
1954	32	22	48	65	85	117	134	143	145	122	90	83
1955	55	51	54	61	97	145	138	144	135	121	89	81
1956	52	26	44	61	87	131	139	157	155	154	106	89
1957	44	38	54	76	119	154	172	179	166	152	121	97

Verižni indeksi

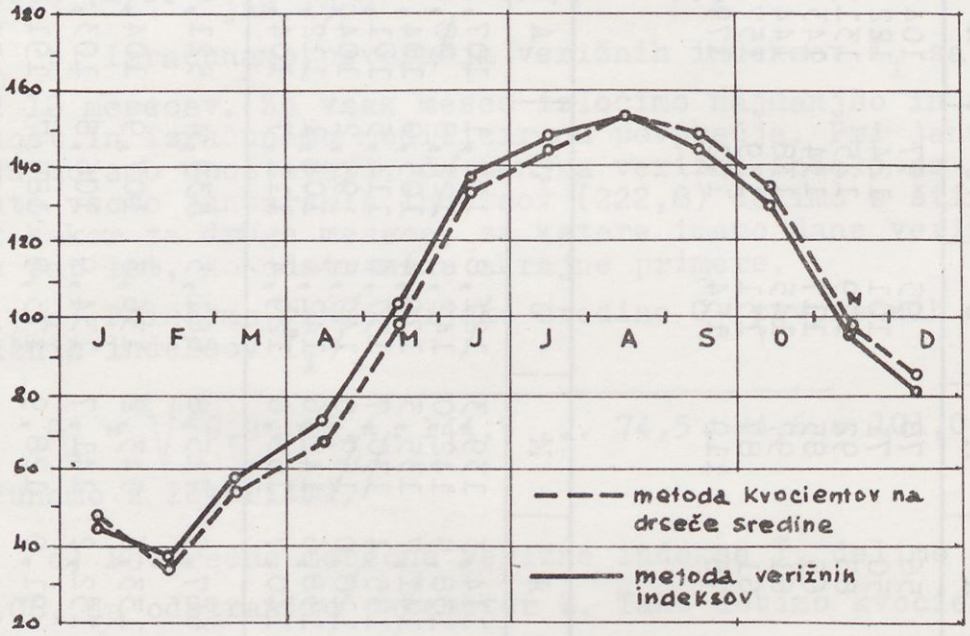
	J	F	M	A	M	J	J	A	S	O	N	D
1951	-	~~97~~,4	132,4	114,3	139,3	~~121~~,8	102,1	107,2	~~86~~,5	93,3	~~84~~,5	81,7
1952	53,4	87,1	133,3	138,9	142,0	142,3	~~122~~,8	~~98~~,4	91,8	87,5	72,4	~~78~~,9
1953	60,7	97,1	145,5	~~141~~,7	135,3	129,4	105,0	104,8	96,9	92,9	72,9	86,0
1954	~~45~~,2	68,8	~~218~~,6	135,4	~~130~~,7	137,6	114,5	101,4	~~101~~,4	~~84~~,1	73,8	~~92~~,2
1955	~~66~~,3	92,7	~~105~~,8	~~118~~,0	~~159~~,0	~~149~~,5	~~95~~,2	104,3	93,8	89,6	73,6	91,0
1956	64,2	~~50~~,0	169,2	138,6	142,6	142,6	106,1	~~112~~,9	98,7	~~99~~,4	~~68~~,8	84,0
1957	49,5	86,4	142,1	140,7	156,6	129,4	111,7	104,1	92,7	91,6	79,6	80,2
Vsota	227,8	432,1	722,5	667,9	715,8	681,3	539,4	521,8	473,9	454,9	372,3	422,9
I_i	56,9	86,4	144,5	133,6	143,2	136,3	107,9	104,4	94,8	91,0	74,5	84,6
t_i	56,3	85,5	143,0	132,2	141,7	134,9	106,8	103,3	93,8	90,0	73,7	83,7
K_i	56,3	48,1	68,8	91,0	128,9	173,9	185,7	191,8	179,9	162,0	119,5	100,0
$1+p_i$	45	38	55	73	103	139	148	153	143	129	95	80

g) Posamezne K_i delimo z dobljenim povprečjem $\bar{K}$ = = 125,5, kvociente pa pomnožimo s 100. Tako dobimo čiste sezonske indekse. Za januar je npr.

$$1+p_1 = 100 \frac{56,3}{125,5} = 45$$

Kontrola: Vsota sezonskih indeksov je enaka 1200. V našem primeru dobimo vsoto enako 1201. Razlika izvira iz razokroževanja.

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 SFRJ 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 nanašamo 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 naguban papir, pomenijo črtice, s katerimi smo označili težišča za posamezne mesece, logaritme nepopravljenih sezonskih indeksov. Z logaritemski 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 6, 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 $Y = T \cdot C \cdot P \cdot S$, 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

$$Y/TP = C \cdot S \qquad (12.49)$$

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. Časov-

ne vrste drsečih sredin zabrišejo oziroma izravnajo ciklična nihanja, če izračunamo sredine za predolga razdobja. Zato so primerna povprečja in neprevelikega števila členov. Zelo primerne so trimesečne ponderirane sredine, pri katerih vzamemo za pondere binomske koeficiente (1, 2, 1).

e) Ko smo tako izluščili 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

K O R E L A C I J A

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 socialnoekonomskih pojavov in množičnih pojavov na sploh, to je njihovo medsebojno odvisnost in povezanost. Pri proučevanju socialnoekonomskih 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 npr. 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:

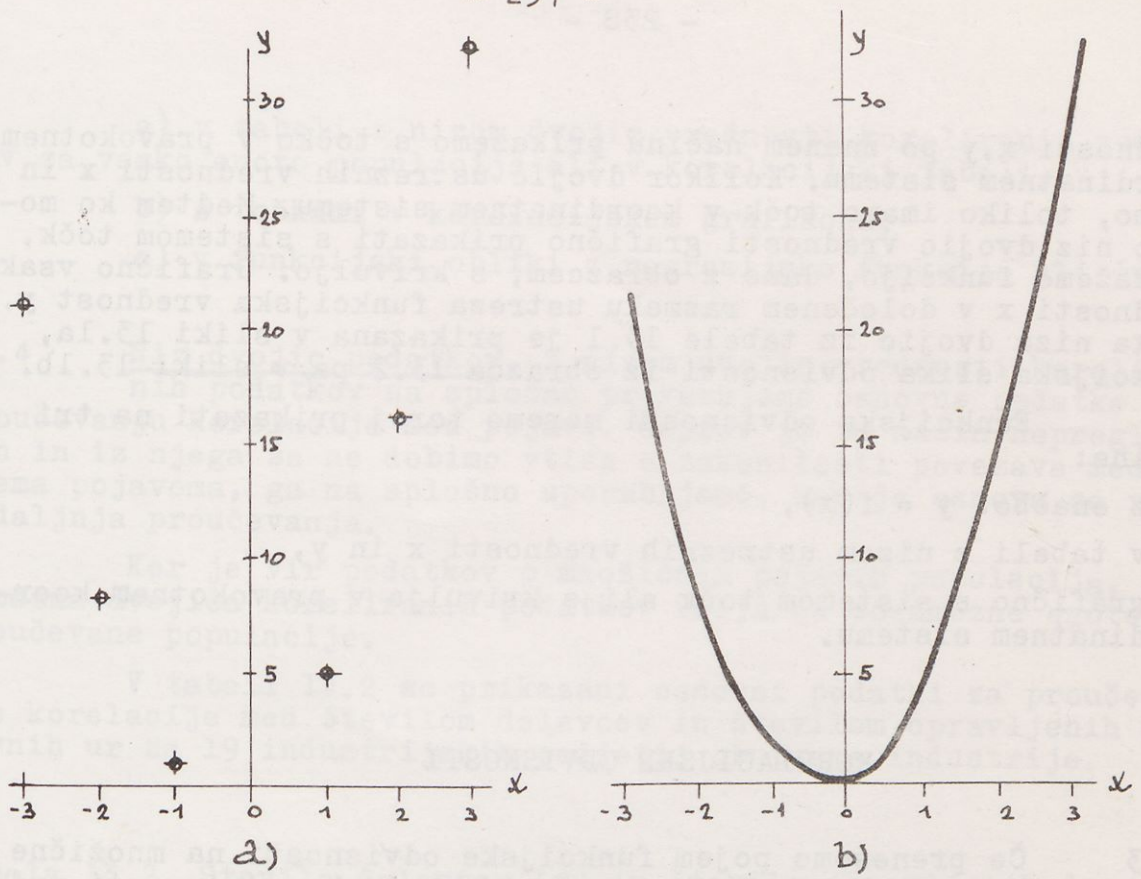
$$y = f(x) \qquad (13.1)$$

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 npr. za funkcijo

$$y = 2x + 3x^2 \qquad (13.2)$$

$x = 2$ ustrezna vrednost $y = 2.2 + 3.2^2 = 16$.



Slika 13.1 Slika funkcijske odvisnosti $y = 2x + 3x^2$

Razen z obrazcem moremo funkcijsko vrednost 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 dvojic ustreznih vrednosti x in y imamo, toliko imamo točk v koordinatnem sistemu. Medtem ko moremo niz dvojic vrednosti grafično prikazati s sistemom točk, prikažemo funkcijo, dano z obrazcem, s krivuljo. Grafično vsaki vrednosti x v določenem razmaku ustreza funkcijska vrednost y. Slika niza dvojic iz tabele 13.1 je prikazana v sliki 13.1a, funkcijska slika odvisnosti iz obrazca 13.2 pa v sliki 13.1b.

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 prenesemo pojem funkcijske odvisnosti na množične pojave, bi imela v primeru funkcijske odvisnosti med površino in proizvodnjo kmetijskega gospodarstva vsa gospodarstva z enako površino enako proizvodnjo. Vendar to ni tako. Čeprav sta površina gospodarstva in proizvodnja med seboj odvisni, imata gospodarstvi z enakima površinama le redko enako proizvodnjo. Še več, Čeprav sodimo, da ima večje gospodarstvo večjo proizvodnjo, velja ta odvisnost samo na splošno, v posameznih primerih pa more imeti tudi večje gospodarstvo manjšo proizvodnjo. Do tega pride zato, ker proizvodnja ni odvisna samo od površine, temveč še od mnogo drugih faktorjev. Poizkus, da bi odstranili vse dodatne faktorje in tako prišli do funkcijske povezave med dvema znakoma, se ne bi posrečil. Vedno ostanejo neki faktorji, katerih vpliv ne moremo odstraniti in 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 za 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 SRS (Vir: Mesečna statistična služba v industriji SRS)

x	154	51	57	15	92	112	23	150	58	81
y	35,3	10,1	11,0	2,9	17,9	25,4	4,1	31,3	10,0	15,1
x	28	100	37	111	12	50	7	80	49	
y	4,9	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 36 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 36 delavskih družin v Mariboru v novembru 1957 (Vir: Anketa o življenju delavcev in nameščenc ZS SRS)

x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7
y	260	330	490	310	580	560	330	480	590	330
x	16,5	10,9	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1
y	630	630	650	1100	760	460	700	540	750	370

x	24,5	23,4	28,4	25,8	3,5	29,8	27,5	26,5	26,7	26,7
y	630	740	1080	630	990	1210	1100	870	640	1110
x	32,9	37,5	30,8	31,8	38,4	35,9				
y	610	1590	1070	1600	1710	1808				

V tabeli 13.4 so podatki o odstotkih travniških in pašniških površin za okraje v SRS po stanju leta 1952.

Tabela 13.4 Odstotki travniških in pašniških površin po okrajih v SRS po stanju leta 1952. (Vir: Statistični bolten SRS)

O k r a j	Odstotek površin pod	
	travniki	pašniki
Celje mesto	22	8
Celje okolica	3	28
Črnomelj	9	20
Gorica	5	13
Kočevje	6	15
Kranj	7	7
Krško	16	9
Ljubljana mesto	28	3
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 npr. 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 npr. število šol v okraju, kulturnoprosvetna 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 SR 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)

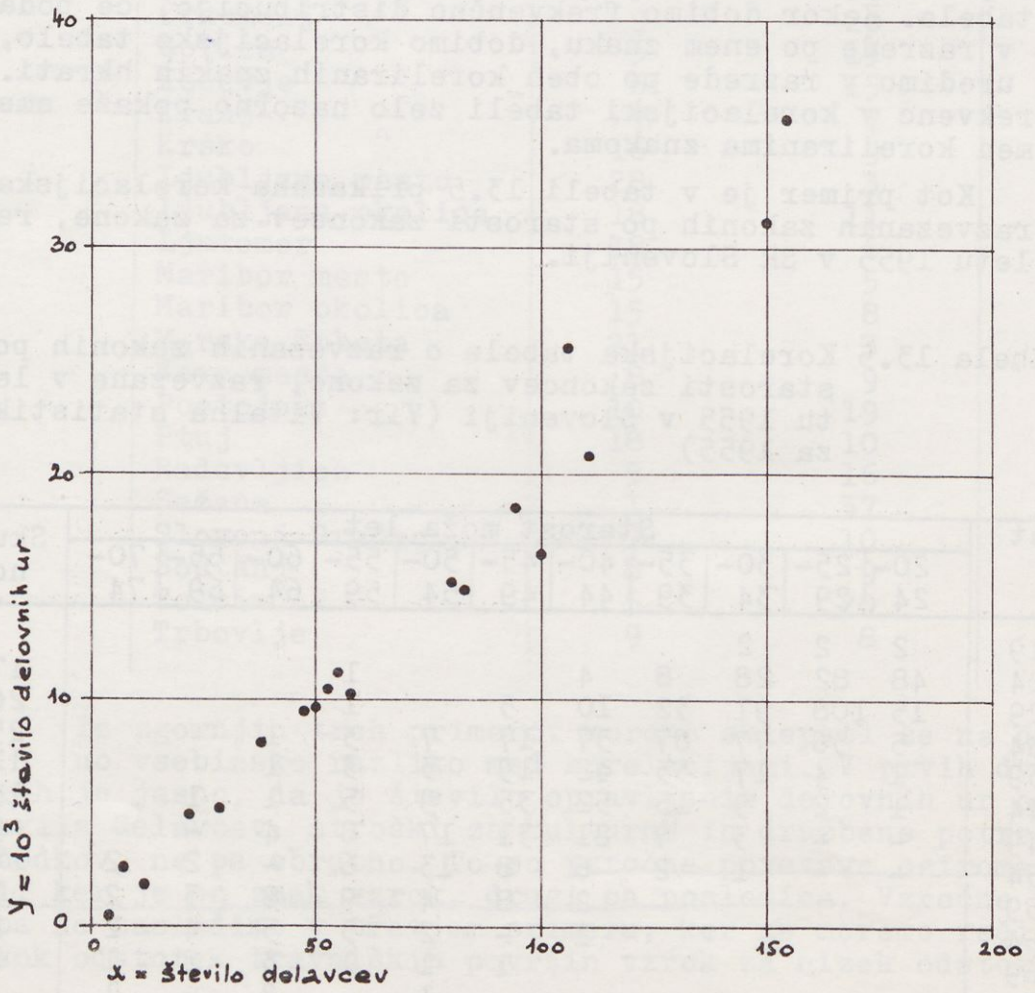
Starost žene let	Starost moža let											Skupno
	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	
15 - 19	2	2	2									6
20 - 24	48	82	28	8	4			1				171
25 - 29	15	108	91	32	10	5		1				262
30 - 34	5	36	77	67	37	17	7	3	1			250
35 - 39	1	4	17	33	42	19	3	3	1			123
40 - 44	1	2	9	15	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	164	175	124	57	40	25	12	8	1139

Podatki v tabeli pomenijo frekvence, tj. število razvezanih zakonov, v katerih je bila starost moža in žene v ustreznih razredih, npr. v 91 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 2236 kmetijskih gospodarstev, ki so bila anketirana v SFRJ, prikazuje odvisnost med skupno površino gospodarstev in denarnimi dohodki od gospodarstva v letu 1956.

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 zajemamo vse točke hkra-



Slika 13.2 Korelacijski grafikon med številom delavcev in številom opravljenih delovnih ur za 19 podjetij obutvene industrije v letu 1952

ti, je korelacija med dvema znakoma v korelacijskem grafikoni vidna neposredno.

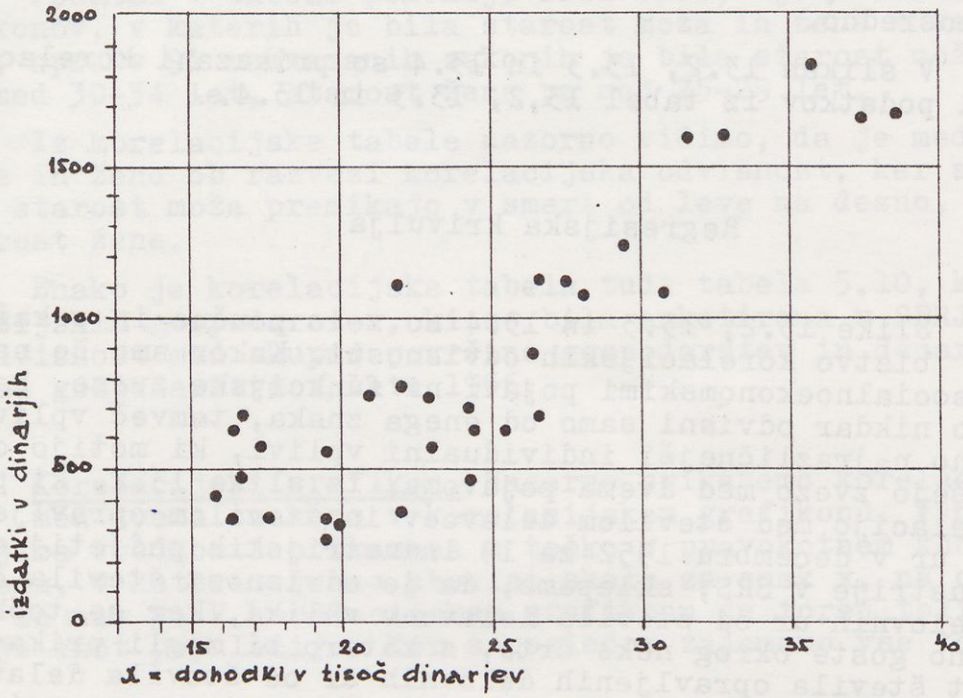
V slikah 13.2, 13.3 in 13.4 so prikazani korelacijski grafikoni podatkov iz tabel 13,2, 13.3 in 13.4.

Regresijska krivulja

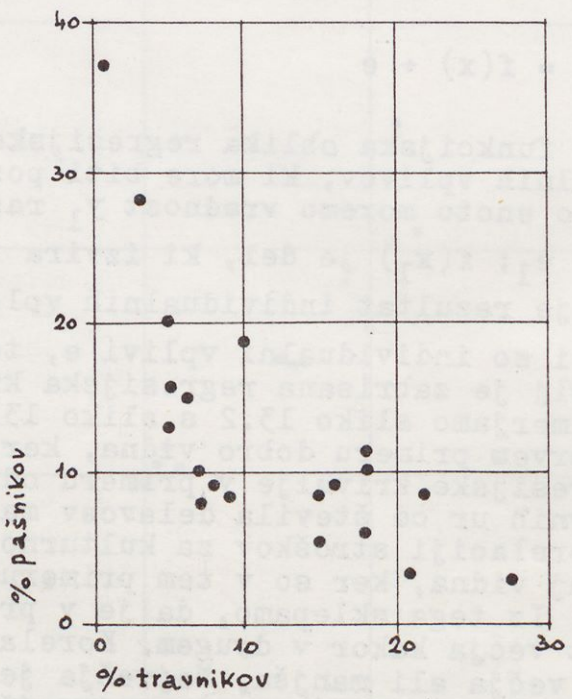
13.8 Slike 13.2, 13.3 in 13.4 so zelo poučne in pokažejo bistvo korelacijskih odvisnosti. Kakor smo že nakazali, med socialnoekonomskimi pojavi ni funkcijske zveze, ker pojavi niso nikdar odvisni samo od enega znaka, temveč vplivajo nanj vedno najrazličnejši individualni vplivi, ki motijo oziroma zabrišejo zvezo med dvema 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 SRS, sklepamo, da je odvisnost števila opravljenih delovnih ur od števila delavcev velika, ker se točke zelo določno goste okrog neke črte, s katero bi mogli prikazati odvisnost števila opravljenih delovnih ur od števila delavcev. To črto, ki gre med vrisanimi točkami, izmenujemo 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

$$y = f(x) + e \qquad (13.3)$$

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 v primeru neodvisnosti ($f(x) = C$). Če sta dva pojava neodvisna, na kore-



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 SRSv letu 1952

lacijskem 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 arimetični sredini koreliranih znakov.

V prvih dveh zgornjih primerih opazimo, da je smer regresijske krivulje taka, da se veča vrednost enega znaka, če se veča 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 13.4. 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 krivuljčni 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 krivilje 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. Seveda ne sme biti vrisana črta preveč kompliciranа, 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 odločitev 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

$$\bar{y} = \overline{f(x)} + \bar{e} \approx f(x) \qquad (13.5)$$

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 $\bar{x}_k$ in $\bar{y}_k$.

c) V korelacijskem grafikonu narišemo točke, ki imajo kot koordinati ustrezne grupne aritmetične sredine za x in y $(\bar{x}_k, \bar{y}_k)$. Tako imamo v grafikonu toliko točk, ki leže na regresijski krivulji ali v njeni bližini, kolikor imamo grup.

d) Vrisane točke grupnih sredin med seboj zvežemo z daljicami.

Lomljena črta, ki jo dobimo, je približno regresijska krivulja.

13.12 Po tej metodi določimo regresijsko krivuljo za odvisnost stroškov za kulturno in družbeno življenje za 36 delavskih gospodinjstev v Mariboru. Podatki so vzeti iz tabele 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)

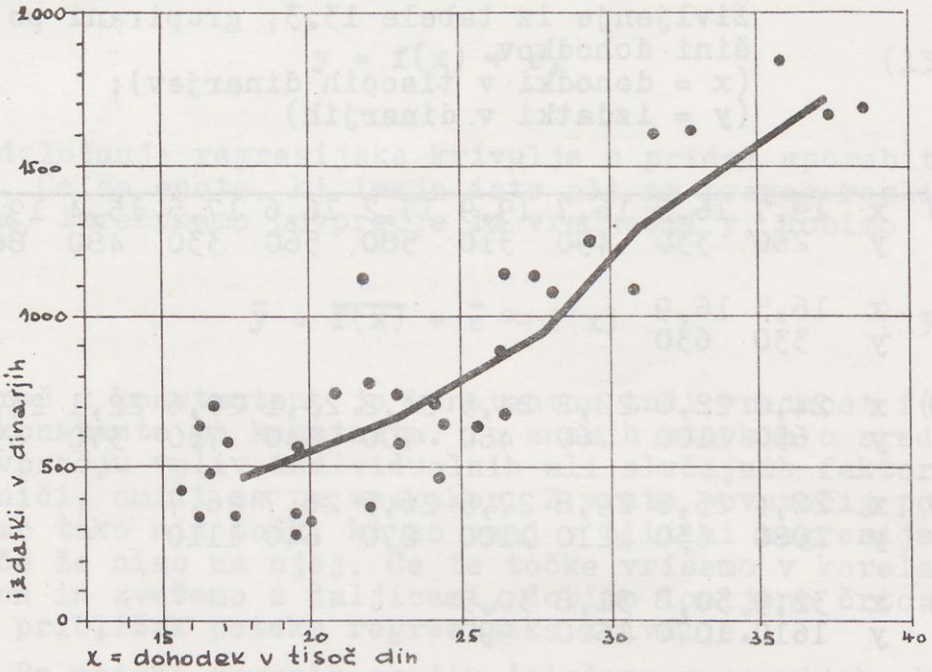
15,0-19,9	x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7
	y	260	330	490	310	580	560	330	480	860	590
	x	16,5	16,9								
	y	330	630								
20,0-24,9	x	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4
	y	650	1100	760	460	700	540	750	370	630	740
25,0-29,9	x	28,4	25,8	29,8	27,5	26,5	26,7	26,7			
	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 5000din in po njih razdelimo dvojice vrednosti x in y. Tako grupirani podatki so navedeni v tabeli 13.6.

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	X_k	Y_k	$\bar{x}_k$	$\bar{y}_k$
15,0 - 19,9	12	212,2	5750	17,7	479
20,0 - 24,9	10	230,6	6700	23,1	670
25,0 - 29,9	7	191,4	6640	27,3	949
30,0 - 34,5	4	126,0	5270	31,5	1318
35,0 -	3	111,8	5110	37,3	1703



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 katerikoli izmed metod za izračunavanje sredin iz frekvenčnih distribucij. Te grupne sredine včrtamo nad sredine razrednih razmakov.

Za korelacijo med starostjo moža in starostjo žene ob razvezi zakona iz tabele 13.5 je smiselno izračunati obe regresijski seriji 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 $\bar{y}_1 = 24,6$ 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 SR 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	26,7	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.

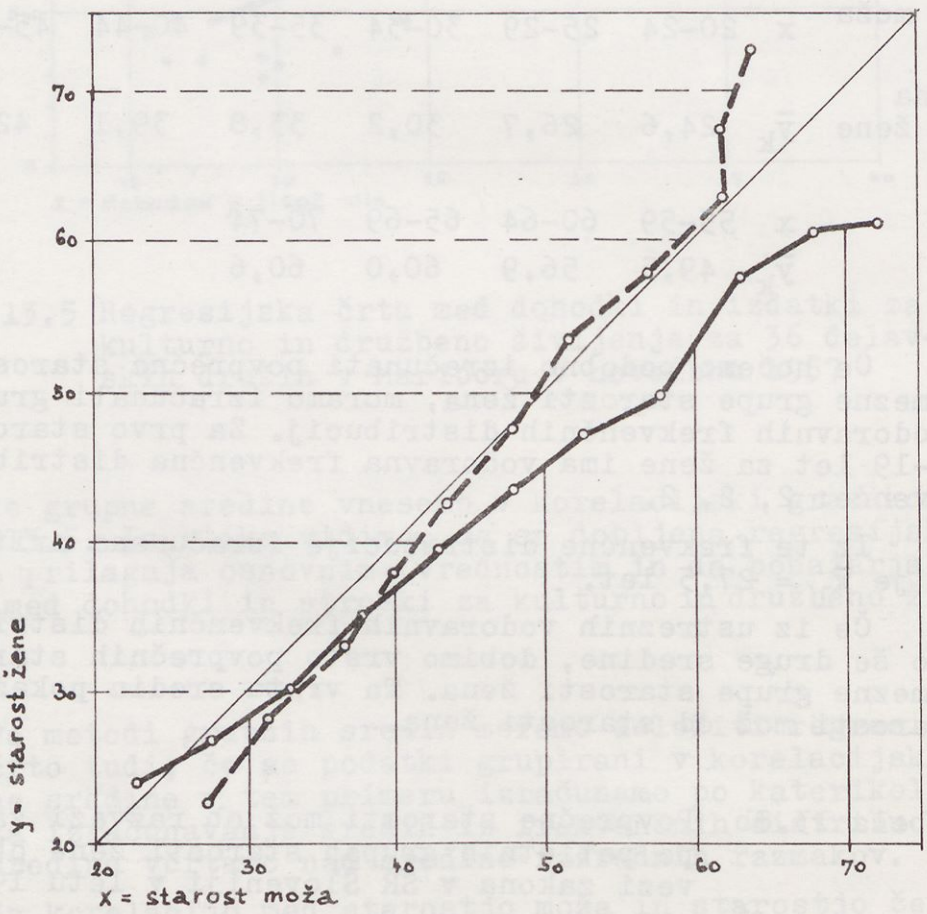
Iz te frekvenčne distribucije izračunana aritmetična sredina je $\bar{x}_1 = 27,5$ let.

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 SR 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	,61,6	61,4	62,5		

Zgornji vrsti, narisani v sliki 13.6, 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 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 ima 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.



Slika 13.6 Regresijski črti grupnih sredin za starost moža in žene ob razvezi zakona v letu 1955 v SR Sloveniji

13.14 Analitična metoda. Regresijsko krivuljo moremo določiti tudi analitično. Če se po analizi podatkov odločimo, da je določen tip funkcije

$$y' = f(x; a, b, c \ldots), \qquad (13.6)$$

ki ima več parametrov (a, b, c,,,), primeren za ragresijsko krivuljo, je treba ustrezno podatkom, ki jih proučujemo, poiskati vrednosti parametrov a, b, c... 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

$$\Sigma(y - y')^2 = F(a, b, c\ldots) \qquad (13.7)$$

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,,), najmanjša. Po tej metodi, ki jo imenujemo metodo najmanjših kvadratov, moremo določiti vrednosti parametrov za regresijsko kriviljo 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

$$y = \bar{y} + (y' - \bar{y}) + (y - y') \qquad (13.8)$$

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 $(y' - \bar{y})$

$$\sigma^2_{y.x} = \frac{1}{N} \Sigma (y' - \bar{y})^2 \qquad (13.9)$$

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 varianca odklonov $(y - y')$

$$\sigma^2_e = \frac{1}{N} \Sigma (y - y')^2 \qquad (13.10)$$

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:

$$\sigma^2_y = \sigma^2_{y.x} + \sigma^2_e \qquad (13.11)$$

Če to enačbo delimo s 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 vizualno oceno. Če poznamo regresijsko krivuljo o odvisnosti cene od količine blaga, ki je naprodaj, moremo pri znani količini blaga oceniti ceno itd.

Ta ocena je jasno tem boljša, čim bolj je y odvisen 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 68 % ocen ne bo napačnih za več kot σ_e, približno 95 % ne več kot $2\sigma_e$ in da bo le izjemoma ocena od prave vrednosti različna za več kot $3\sigma_e$. Tako moremo s standardno 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 odvisnosri 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

$$y' = \bar{y} + b_1(x-\bar{x}) \qquad (13.16)$$

podaja odvisnost y od x, druga

$$x' = \bar{x} + b_2(y-\bar{y}) \qquad (13.17)$$

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

$$b_1 = \frac{c_{xy}}{\sigma_x^2} \; ; \qquad b_2 = \frac{c_{xy}}{\sigma_y^2} \qquad (13.18)$$

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

$$c_{xy} = \frac{1}{N} \Sigma (x-\bar{x}).(y-\bar{y}) \qquad (13.19)$$

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

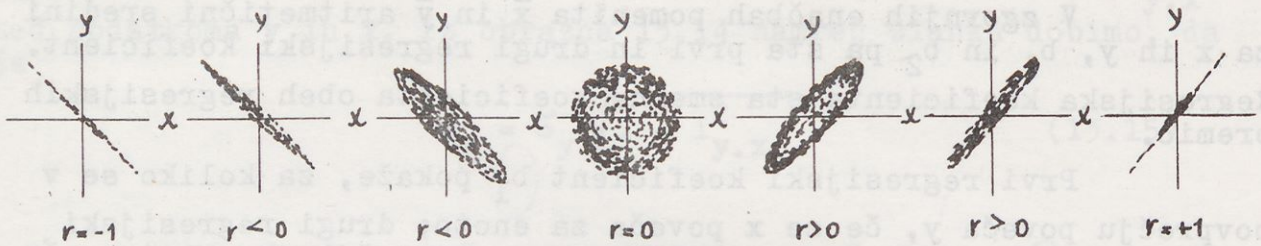
$$r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y} \qquad (13.20)$$

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 od 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 neprikladna, 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_o med stvarnimi vrednostmi za x, in poljubno vrednost y_o med stvarnimi vrednostmi za y,

c) izračunamo odklone stvarnih vrednosti x od x_o in stvarnih vrednosti y od y_o. Tako dobimo nove pomožne znake $u = (x-x_o)$ in $v = (y-y_o)$. Vrednosti u in v so v večini primerov prikladnejše 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.

13.21 Kot primer za izračunavanje pokazateljev korelacije po metodi pomožnih 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 SRS in pomenijo število doseženih točk pri posameznem testiranju.

Iz absolutnih podatkov za x in y smo presodili, da je primerno, da vzamemo za $x_o = 40$ in za $y_o = 50$. 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

Tabela 13.9 Shema za računanje pokazateljev linearne korelacije

i	x	y	$x-x_o = u$	$y-y_o = v$	u^2	uv	v^2
1	x_1	y_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	$-U^2/N$	$-UV/N$	$-V^2/N$
			$+x_o$	$+y_o$	$\Sigma u^2-U^2/N = K_u$	$\Sigma uv-UV/N = K_{uv}$	$\Sigma v^2-V^2/N = K_v$
			$U/N+x_o = \bar{x}$	$V/N+y_o = \bar{y}$	$K_u/N = \sigma_x^2$	$K_{uv}/N = c_{xy}$	$K_v/N = \sigma_y^2$
					$\sqrt{\sigma_x^2} = \sigma_x$		$\sqrt{\sigma_y^2} = \sigma_y$
					$b_1 = \frac{K_{uv}}{K_u}$	$r_{xy} = \frac{K_{uv}}{K_u \cdot K_v}$	$b_2 = \frac{K_{uv}}{K_v}$
					$\sigma_{ex} = \sigma_x\sqrt{1-r_{xy}^2}$	$\sqrt{1-r_{xy}^2}$	$\sigma_{ey} = \sigma_y\sqrt{1-r_{xy}^2}$

STANDARDNA NAPAKA OCENE

$y' = \bar{y} + b_1(x-\bar{x})$

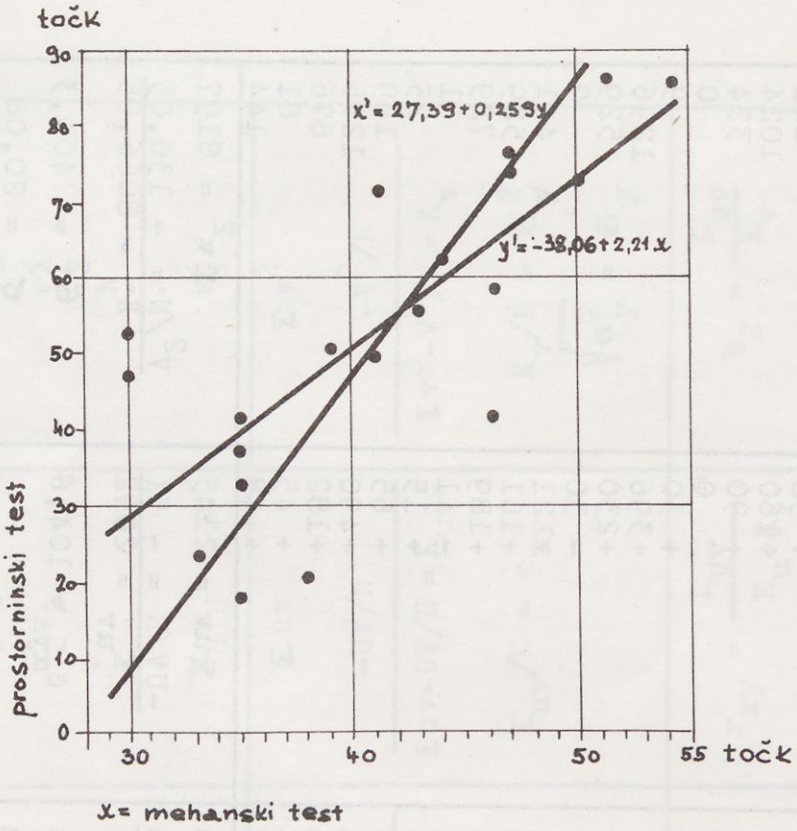
$x' = \bar{x} + b_2(x-\bar{y})$

Tabela 13.10 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	52	73	+10	+23	100	+230	529
10	30	53	-10	+ 3	100	- 30	9
11	41	71	+ 1	+21	1	+ 21	441
12	47	72	+ 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	Σu^2 = 968	Σuv = 2143	Σv^2 = 8197
			U/N = + 1	V/N = +2,55	$-U^2/N$ = -20	$-UV/N$ = - 51	V^2/N = - 130,05
			$+x_o$ = +40	$+y_o$ = + 50	K_u = 948	K_{uv} = 2092	K_v = 8066,95
			$\bar{x}$ = 41	$\bar{y}$ = 52,55	σ_x^2 = 47,4	c_{xy} = 104,6	σ_y^2 = 403,3
					σ_x = 6,88		σ_y = 20,08
					b_1 = +2,21	r_{xy} = 0,757	b_2 = +0,259
					σ_{ex} = 4,49	$\sqrt{1-r_{xy}^2}$ = 0,653	σ_{ey} = 13,08

$y' = 52,55 + 2,21(x-41) = -38,06 + 2,21x$

$x' = 41 + 0,259(y-52,55) = 27,39 + 0,259y$



Slika 13.8 Regresijski premici za korelacijo med dosežki pri mehanskem in prostorninskem testu za 20 dijakov klasične gimnazije v Mariboru

Korelacijski grafikon z včrtanima 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_o in y_o, da bi poenostavili osnovne vrednosti, vzamemo, da je $x = 0$ in $y_o = 0$. 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 prikladno, 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 SRS v decembru 1952. Izračun je izvršen v tabeli 13.12.

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 SRS 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	1646,8	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	6561	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
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
	$\underline{X} = 1267$	$\underline{Y} = 250,1$	$\Sigma x^2 = 119721$	$\Sigma xy = 24272,8$	$\Sigma y^2 = 4970,92$
	$\bar{x} = 66,68$	$\bar{y} = 13,16$	$-X^2/N = 84488,89$	$-XY/N = 16677,72$	$-Y^2/N = 3292,11$
			$K_x = 35232,11$	$K_{xy} = 7595,08$	$K_y = 1678,81$
			$\sigma_x^2 = 1854,32$	$c_{xy} = 399,74$	$\sigma_y^2 = 88,36$
					$\sigma_y = 9,40$
			$b_1 = 0,216$	$r = +0,988$	
				$\sqrt{1-r^2} = 0,154$	$\sigma_{ey} = 1,45$

$$y' = 13,16 + 0,216(x-66,68) = -1,24 + 0,216x$$

Ker je med številom delavstva x in številom opravljenih delovnih ur y vzročna zveza, izračunamo samo prvo regresijsko premico.

Ker je število ur dano v tisočih, regresijski koeficient 0,216 pove, da se v povprečju spremeni skupno število opravljenih ur za 216 ur, če se število delavcev spremeni za enega.

Iz r_{xy} in r_{xy}^2 vidimo, da je med številom delavcev in številom opravljenih delovnih ur močna povezanost in da je samo 2,39 % celotne variance nepojasnjene.

Standardna napaka ocene za število opravljenih delovnih ur je $\sigma_{ey} = 1,45$. To pomeni, da se 68 % ocen za število opravljenih delovnih ur, ki bi ga ocenili s številom delavcev z zgornjo regresijsko premico za katerokoli 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 $y_8' = -1,24 + 0,216 . x_8 = -1,24 + 0,216 . 150 = 31,16$. Razlika med pravo vrednostjo in oceno $y-y' = 31,3 - 31,16 = +0,14$ je zelo majhna.

Poizkusimo na enak način oceniti število opravljenih delovnih ur za trinajsto podjetje z $x_{13} = 37$ delavcev. Zanj je

$$y_{13}' = -1,24 + 0,216x_{13} = -1,24 + 0,216.37 = 6,75$$

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 izhodišče 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_y pomnožimo z ustreznimi vrednostmi znaka v, da dobimo vf_y, dobljene produkte pa ponovno pomnožimo z ustreznimi vrednostmi v, da dobimo v^2f_y. Oba dobljena stolpca seštejemo. Enako napravimo z drugo robno frekvenčno distribucijo. Tako dobimo naslednje količine:

$$\sum_u uf_u = U; \quad \sum_u u^2f_u = \Sigma u^2; \quad \sum_v vf_v = V; \quad \sum_v v^2f_v = \Sigma v^2$$

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	i_y	Σu^2	Σuv	Σv^2
N U	V	$-U^2/N$	$-UV/N$	$-V^2/N$
$i_x U/N$	$i_y V/N$	$\Sigma u^2 - U^2/N = K_u$	$\Sigma uv - UV/N = K_{uv}$	$\Sigma v^2 - V^2/N = K_v$
$+x_o$	$+y_o$	$i_x^2 K_u/N = \sigma_x^2$	$i_x i_y K_{uv}/N = c_{xy}$	$i_y^2 K_v/N = \sigma_y^2$
$i_x U/N + x_o = \bar{x}$	$i_y V/N + y_o = \bar{y}$	$\sqrt{\sigma_x^2} = \sigma_x$		$\sqrt{\sigma_y^2} = \sigma_y$
		$b_1 = \frac{c_{xy}}{\sigma_x^2}$	$r_{xy} = \frac{K_{uv}}{\sqrt{K_u K_v}}$	$b_2 = \frac{c_{xy}}{\sigma_y^2}$
		$\sigma_{ex} = \sigma_x \sqrt{1-r_{xy}^2}$	$\sqrt{1-r_{xy}^2}$	$\sigma_{ey} = \sigma_y \sqrt{1-r_{xy}^2}$

$$y' = \bar{y} + b_1(x - \bar{x})$$

$$x' = \bar{x} + b_2(y - \bar{y})$$

13.24 Kot primer za izračunavanje pokazateljev linearne korelacije in regresije vzemimo odvisnost med starostjo žen in mož ob razvezi zakona v SRS v letu 1955 iz tabele 13.5. V tabeli 13.14 je sistematično nakazan postopek izračunavanja pomožnih količin U, V, Σu^2, Σv^2, 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:

$U_1 = (-4).2 + (-3).2 + (-2).2 = -18$

$U_2 = (-4).48 + (-3).82 + (-2).28 + (-1).8 + (0).4 + (+3).1 = -499$

itd. do zadnje vrste, za katero imamo

$$U_{12} = (+2).1 + (+4).2 + (+6).1 = +16$$

Kontrola izračunavanja izrazov U_v, ki so najbolj problematični, je $\Sigma_v U_v = -1044 = U$

Nadaljnji račun po shemi 13.13 je nakazan v tabeli 13.15.

Tabela 13.14 Izračunavanje pokazateljev linearne korelacije in regresije za odvisnost med starostjo žene in moža ob razvezi zakona v SRS 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	f_v	vf_v	v^2f_v	$\Sigma_u uf_{uv}$	vU_v
	v \ u	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6				$U_u = V$	
15-19	-5	2	2	2	-	-	-	-	-	-	-	-	6	-30	150	-18	+90
20-24	-4	48	82	28	8	4	-	-	1	-	-	-	171	-684	2736	-499	+1996
25-29	-3	15	108	91	32	10	5	-	1	-	-	-	262	-786	2358	-590	+1770
30-34	-2	5	36	77	67	37	17	7	3	1	-	-	250	-500	1000	-305	+ 610
35-39	-1	1	4	17	33	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	2	1	9	+ 45	225	+ 34	+ 170
70-74	+6	-	-	-	-	-	-	1	-	2	-	1	4	+ 24	144	+ 16	+ 96
	f_u	72	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 Σ_v^2	U	Σuv
	u^2f_u	1152	2106	912	164	0	124	228	360	400	300	288	6034 = Σu^2				
	$\Sigma uf_{uv} = V_u$	-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 = Σuv				

Tabela 13.15 Izračun pokazateljev linearne korelacije za primer v tabeli 13.14 po shemi iz tabele 13.13

$i_x = 5$; $i_y = 5$;	$x_o = 42,5$	$y_o = 42,5$;	
$N = 1139$ $U = -1044$ $V = -1707$	$\Sigma_u^2 = 6034$	$\Sigma_{uv} = +5575$	$\Sigma_v^2 = 7583$
$i_x U/N = -4,58$ $i_y V/N = -7,49$	$-U^2/N = -956,92$	$-UV/N = -1564,63$	$-V^2/N = -2558,25$
$+x_o = 42,50$ $+y_o = 42,50$	$K_u = 5077,08$	$K_{uv} = +4010,37$	$K_u = 5024,75$
$\bar{x} = 37,92$ $\bar{y} = 35,01$	$\sigma_x^2 = 111,43$	$c_{xy} = +88,02$	$\sigma_y^2 = 110,29$
	$\sigma_x = 10,56$		$\sigma_y = 10,50$
	$b_1 = +0,790$	$r_{xy} = +0,794$	$b_2 = +0,798$
	$\sigma_{ex} = 6,42$	$\sqrt{1-r_{xy}^2} = 0,608$	$\sigma_{ey} = 6,38$

$$y' = 35,01 + 0,790(x-37,92) = 5,05 + 0,790x$$

$$x' = 37,92 + 0,798(y-35,01) = 9,98 + 0,798y$$

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 precenimo, 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.

A n a l i t i č n a m e t o d a z a d o l o č a n j e r e g r e s i j s k i h k r i v u l j

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,.. 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 ,..

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 ... v linearni 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

$$g(y') = af_1(x) + bf_2(x) + cf_3(x), \qquad (13.23)$$

je sistem normalnih enačb enak

$$\begin{aligned} \Sigma gf_1 &= a\Sigma f_1^2 + b\Sigma f_2 f_1 + c\Sigma f_3 f_1 \\ \Sigma gf_2 &= a\Sigma f_1 f_2 + b\Sigma f_2^2 + c\Sigma f_3 f_2 \\ \Sigma gf_3 &= a\Sigma f_1 f_3 + b\Sigma f_2 f_3 + c\Sigma f_3^2 \end{aligned} \qquad (13.24)$$

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.

T r a n s f o r m a c i j a z n a k o v

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 Transformacija 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 (logx; logy).

V vseh primerih smo v tabeli 13.16 uspeli s transformacijo razmeroma komplicirano 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, reciprok 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 reciprokov 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.

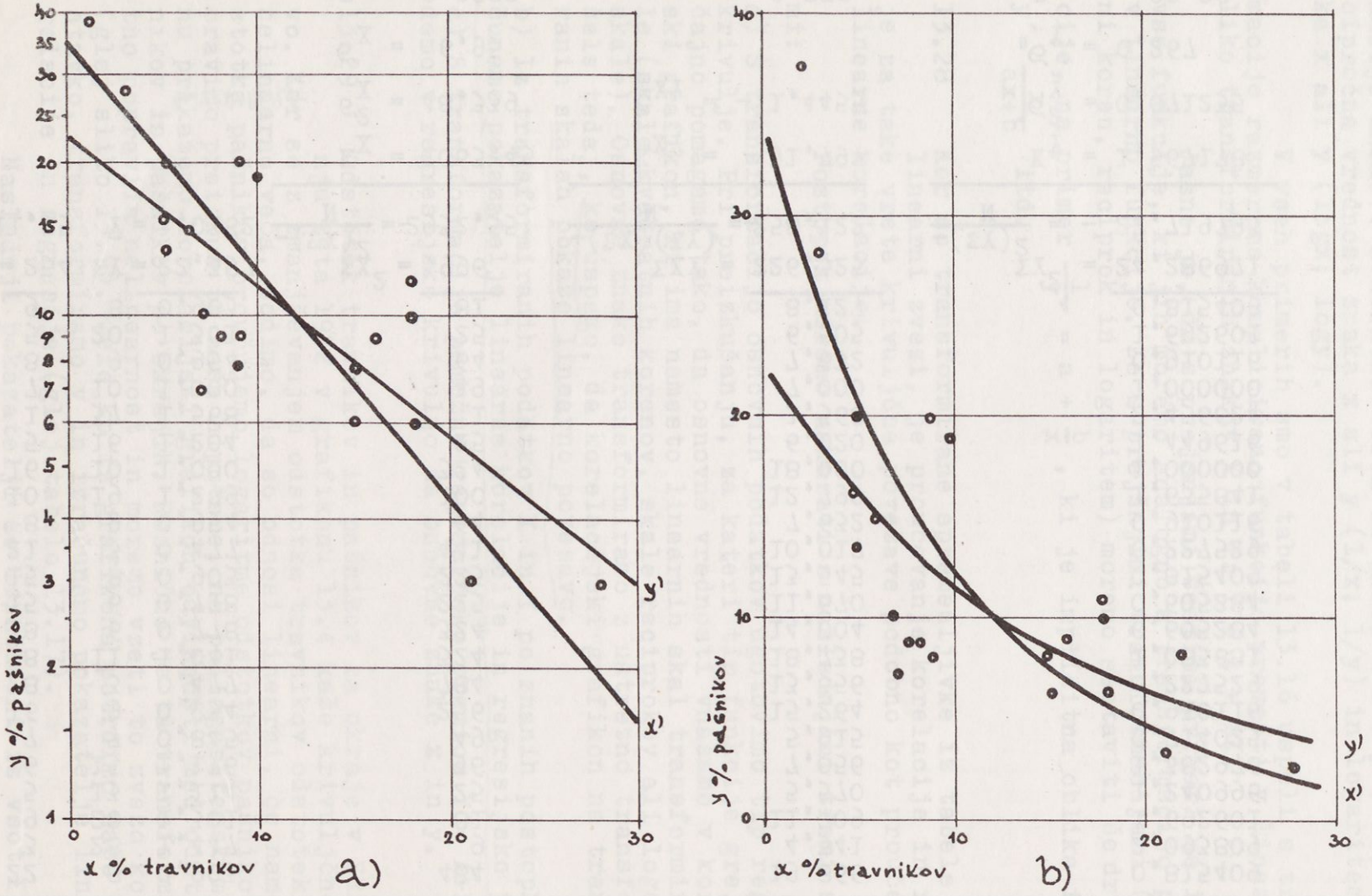
13.29 Odstotek travnikov in pašnikov za okraje v SRS po stanju leta 1952 v grafikonu 13.4 kaže krivuljčno povezavo. Ker se z zmanjševanjem odstotka travnikov odstotek pašnikov nelinearno veča, sodimo, da so odnosi linearni, če namesto odstotka pašnikov proučujemo logaritme odstotkov pašnikov. Da napravimo preizkus te naše hipoteze, na pollogaritemskem 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 pašniških površin za okraje v SRS 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	16	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	16	10	1,000	324	18,000	1,000000
17	5	16	1,204	25	6,020	1,449616
18	1	37	1,568	1	1,568	2,456624
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	$\Sigma X = 254$	$\Sigma Y = 22,357$		$\Sigma X^2 = 3996$	$\Sigma XY = 226,325$	$\Sigma Y^2 = 24,286971$
	$\bar{X} = 11,55$	$\bar{Y} = 1,0162$		$\frac{-(\Sigma X)^2}{N} = -2932,55$	$\frac{-(\Sigma X)(\Sigma Y)}{N} = -258,122$	$\frac{-(\Sigma Y)^2}{N} = -22,719793$
				$K_X = 1063,45$	$K_{XY} = -31,797$	$K_Y = 1,567178$
				$\sigma_X^2 = 48,339$	$C_{XY} = -1,4453$	$\sigma_Y^2 = 0,071235$
				$\sigma_X = 6,95$		$\sigma_Y = 0,267$
				$B_1 = 0,0299$	$r_{XY} = -0,778$	$B_2 = -20,29$

KOVARIANCA:



Slika 13.9 Korelacijska grafikona med odstotkom travnikov in odstotkom pašnikov za okraje v SRS v letu 1952 z včrtanima regresijskima krivuljama na: a) transformiranih skalah, b) osnovnih skalah

Regresijske krivulje pa so iz teh pokazateljev dalje:

$Y \dot{=} \overline{Y} + B_1(X-\overline{X}) = 1,0162 - 0,0299(X-11,55) = 1,3615 - 0,0299\,X = A + B_1 X$

$\log y' = 1,3615 - 0,0299x$

$A = \log a = 1,3615;\ a = 22,99 \qquad B_1 = \log b = -0,0299;\ b_1 = 0,9335$

Iz tega dalje sledi, da je y' v eksplicitni obliki

$$y' = 22,99 \cdot 0,9335^x \qquad y' = a \cdot b^x$$

Druga regresijska enačba pa ima obliko:

$X' = \overline{X} + B_2(Y-\overline{Y}) = 11,55 - 20,29(Y-1,0162) = 32,17 - 20,29Y$

Če upoštevamo transformacijo Y = logy, 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

$$Y' = \overline{Y} + B_1(X - \overline{X}) = 3,377 - 0,793(X-3,229) = 5,938 - 0,793X$$

Regresijska krivulja v osnovnih znakih x in y pa je

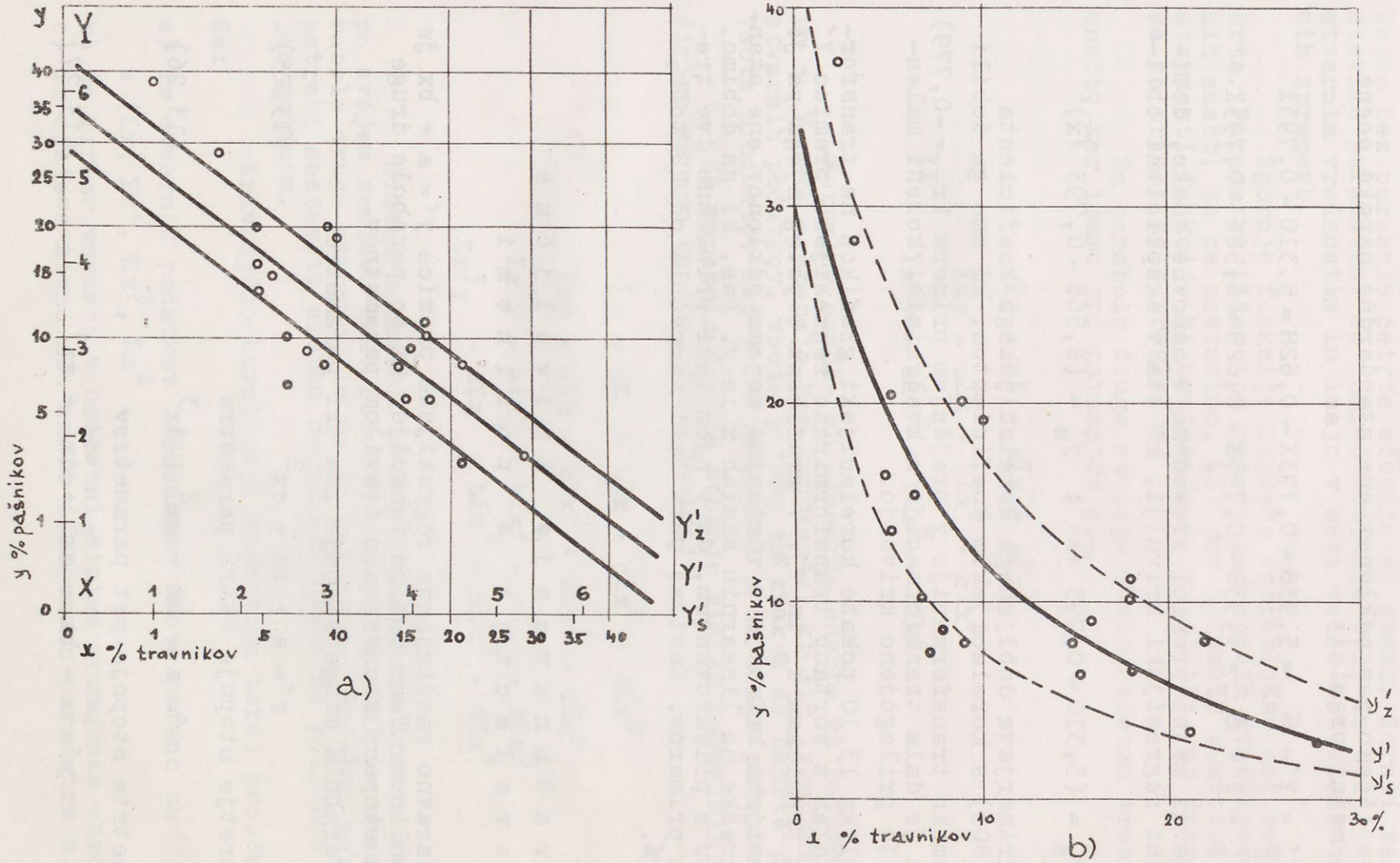
$$\sqrt{y'} = 5,938 - 0,793 \sqrt{x}$$

ali v eksplicitni obliki

$$y' = (5,938 - 0,793 \sqrt{x})^2$$

Tabela 13.18 Izračunavanje krivuljčne korelacije in regresije med odstotkom travnikov (x) in pašnikov (y) za okraje v SRS 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	10,0128	20
22	9	8	3,00	2,83	9	8,4900	8
			$\Sigma X = 71,04$	$\Sigma Y = 74,29$	$\Sigma X^2 = 254$	$\Sigma XY = 220,4711$	$\Sigma Y^2 = 275$
			$\bar{X} = 3,229$	$\bar{Y} = 3,377$	$-\frac{(\Sigma X)^2}{N} = 229,39$	$-\frac{(\Sigma X)(\Sigma Y)}{N} = -240,0000$	$-\frac{(\Sigma Y)^2}{N} = -250,86$
					$K_X = 24,61$	$K_{XY} = -19,5289$	$K_X = 24,14$
					$\sigma_x^2 = 1,1186$	$c_{xy} = -0,8877$	$\sigma_y^2 = 1,0973$
					$\sigma_x = 1,058$		$\sigma_y = 1,047$
					$B_1 = -0,793$	$r_{xy} = -0,801$	$B_2 = -0,809$
					$\sigma_{ex} = 0,634$	$\sqrt{1-r^2} = 0,599$	$\sigma_{ey} = 0,628$



Slika 13.10 Korelacijska grafikona med odstotkom travnikov in odstotkom pašnikov za okraje v SRS v letu 1952 z včrtano regresijsko krivuljo y´ in pasom ene standardne napake: a) na transformiranih skalah, b) na osnovnih skalah

Standardna napaka ocene $\sigma_{ey} = 0,628$

Če iščemo pas odklonov ene standardne napake ocene, dobimo dve mejni premici:

$$Y'_s = Y' - \sigma_{ey} = 5,938 - 0,793X - 0,628 = 5,310 - 0,793X$$

$$Y'_s = Y' + \sigma_{ey} = 5,938 - 0,793X + 0,628 = 6,566 - 0,793X$$

Če ti mejni premici prevedemo v osnovne znake, dobimo dve mejni regresijski krivulji, ki sta v eksplicitni obliki takile:

$$y'_s = (5,310 - 0,793\sqrt{x})^2; \quad y'_z = (6,566 - 0,793\sqrt{x})^2$$

Primerjava dobljenega korelacijskega koeficienta ($r_{XY} = -0,801$) s korelacijskim koeficientom, ki smo ga dobili z logaritemsko transformacijo v prejšnjem primeru ($r_{XY} = -0,779$) 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 %.

I z r a č u n a v a n j e k r i v u l j č n e k o r e l a c i j e s p o l i n o m i

13.31 Naravno razširjenje regresijske premice $y' = a + bx$ je polinom. Tako dobimo funkcije, ki so parabole druge ali višjih stopenj z ustreznim številom parametrov.

Parabola druge stopnje ima tri parametre

$$y' = a + bx + cx^2 \qquad (13.25)$$

parabola tretje stopnje štiri parametre

$$y' = a + bx + cx^2 + dx^3 \qquad (13.26)$$

parabola četrte stopnje pet parametrov

$$y' = a + bx + cx^2 + dx^3 + ex^4 \qquad (13.27)$$

Ta postopek moramo 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

$$\begin{aligned} \Sigma y &= aN + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned} \qquad (13.28)$$

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_{yx} izračunamo po znanem obrazcu

$$I_{yx} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}} \qquad (13.32)$$

13.32 Če pogledamo korelacijski grafikon med odstotkom travnikov in pašnikov za okraje v SRS v letu 1952 na sliki 12.4, ugotovimo, da moremo v tem primeru za regresijsko krivuljo vzeti tudi del parabole stopnje tipa $y' = a + bx + cx^2$.

Zanjo velja sistem normalnih enačb 13.28. Zato moramo iz osnovnih podatkov izračunati izrazre: Σy, Σxy, $\Sigma x^2 y$, Σx, Σx^2, Σx^4. V tabeli 13.19 je nakazan izračun teh izrazov.

Tabela 13.19 Izračunavanje krivuljčne 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	216	1296	225
6	7	7	49	343	49	343	2401	49
7	16	9	144	2304	256	4096	65536	81
8	28	3	84	2352	784	21952	614656	9
9	18	11	198	3564	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	9261	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	4096	81
21	5	20	100	500	25	125	625	400
22	9	8	72	648	81	729	6561	64
	254	275	2324	29956	3996	74990	1564476	4859

Če te izraze vnesemo v sistem normalnih enačb 13.28, dobimo

$$\begin{aligned} 22a + 254b + 3996c &= 275 \\ 254a + 3996b + 74990c &= 2324 \\ 3996a + 74990b + 1564476c &= 29956 \end{aligned}$$

Iz tega sistema normalnih enačb dobimo parametre:

$$a = +29{,}0468; \quad b = -2{,}3066; \quad c = +0{,}0555$$

Regresijska krivulja pa je

$$y' = +29{,}0468 - 2{,}3066x + 0{,}0555x^2$$

Standardna napaka ocene je po obrazcu 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_{xy} po obrazcu 13.32 enak

$$I_{yx} = \sqrt{1 - \frac{25,8686}{64,6136}} = \sqrt{0,5996} = 0,774$$

Če primerjamo dobljeni indeks korelacije $I_{yx} = 0,774$ s korelacijskima koeficientoma, katera smo dobili z logaritemsko transformacijo $r_{XY} = 0,779$, ali s transformacijo s kvadratnimi koreni $r_{XY} = 0,801$, 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 artimetičnimi sredinami σ_y^2 in aritmetične sredine grupnih varianc $M_\sigma 2$.

$$\sigma_y^2 = \sigma\frac{2}{y} + M_\sigma 2 \qquad (13.33)$$

Enako kot dosedaj vzamemo za merilo korelacije razmerja med pojasnjeno varianco (varianco med aritmetičnimi sredinami) $\sigma\frac{2}{y}$ 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 artimetične sredine najbolj prilagajo stvarnim podatkom.

Po tej definiciji je korelacijsko razmerje

$$\eta^2_{y.x} = \frac{\sigma^2_{\bar{y}}}{\sigma^2_y} = \frac{\frac{1}{N}\Sigma N(\bar{y}_k - \bar{y})^2}{\frac{1}{N}\Sigma(y-\bar{y})^2} \qquad (13.34)$$

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 dosedanje 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 neprikladno. 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 zrazov 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.

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 prve 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

$$\eta^2_{y.x} = \frac{\Sigma A_k^2/N_k - A^2/N}{2B + A - A^2/N} \qquad (13.36)$$

Če za analizo potrebujemo vrsto grupnih aritmetičnih sredin, jih izračunamo po obrazcu

$$\bar{y}_k = y_o + i_y \frac{A_k}{N_k} \qquad (13.37)$$

y_o = sredina najnižjega razreda.

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
$\Sigma y^2 = 260^2 + 330^2 + \ldots + 1810^2 = 30584100$						36	29470	868480900	29189980
						N	Y	Y^2	Y_k^2/N_k
								24124469 =	Y^2/N

Po obrazcu 13.35 dobimo $\eta^2_{y.x} = \frac{29189980 - 24124469}{30584100 - 24124469} = 0,784$

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	3	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	6	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		93		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	$1573,07 = \Sigma A_k^2/N_k$			

Po obrazcu 13.36 dobimo dalje, da je

$$\eta^2_{y.x} = \frac{1576,07 - 1059,27}{2.1204 + 1539 - 1059,27} = 0,179$$

Rezultat pokaže, da je $\eta^2_{yx} = 0,179$ od skupne variance dohodkov iz kmetijstva pojasnjena s skupno površino

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 prikladen, če kumulative izračunavamo s seštevalnim strojem, ki ima kontrolni trak.

Po obrazcu 13.37 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	134
nad 8 do 12 ha	195
nad 12 do 16 ha	189
nad 16 do 20 ha	288
nad 20 ha	313

KORELACIJA RANGA

13.38 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 pripišimo 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 eno-

stavnejše kot izračunavanje korelacijskega koeficienta r_{xy}. Koeficient korelacije ranga izračunavamo po Spearmanovem obrazcu

$$\varrho = 1 - \frac{b \Sigma d^2}{N(N^2-1)} \qquad (13.38)$$

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 napišemo ustrezne range R_x.

c) Enako razporedimo listke po velikosti vrednosti znaka y in tudi zanj vpišemo na listku 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, izračunani iz istih podatkov, nista enaki, ker sta izračunani različno. Vendar v večini primerov razlike niso velike in moremo 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 nasto sestavljena tabela 13.23.

Tabela 13.23 Izračunavanje korelacijskega koeficienta ranga med številom delavcev (x) in številom opravljenih delovnih ur (y) za 19 podjetij industrije obutve v SRS

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	6	0	0
49	9,9	7	8	-1	1
50	9,5	8	7	+1	1
51	10,1	9	10	-1	1
57	11,0	10	11	-1	1
58	10,0	11	9	+2	4
80	15,1	12	13	-1	1
81	15,0	13	12	+1	1
92	17,9	14	15	-1	1
100	16,7	15	14	+1	1
111	20,7	16	16	0	0
112	25,4	17	17	0	0
150	31,3	18	18	0	0
154	35,3	19	19	0	0
				0	12 = Σd^2

Ker je N = 19 in Σd^2 = 12, dobimo, da je po obrazcu 13.38

$$\varrho = 1 - \frac{6.12}{19(19^2-1)} = 0,9895$$

Če primerjamo ϱ = +0,9895 z ustreznim r_{xy} = 0,988, ki smo ga izračunali v tabeli 13.12, vidimo, da so razlike neznatne.

ASOCIACIJA IN KONTINGENCA

A s o c i a c i j a

13.40 Iz kombinacijske tabele med starostjo moža in žene ob razvezi zakona v SRS 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 SR 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 SR Sloveniji po popisu 1953 po spolu in delavnosti v tisočih prebivalcev (Vir: SG 1958)

Delavnost / Spol	Delovni	Vzdrževani	Skupno
Moški	474	219	693
Ženske	330	443	773
Skupno	804	662	1466

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 SR Sloveniji po popisu 1953

Delavnost / Spol	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 tema odstotkoma, in obratno: čim manjša je odvisnost delavnosti od spola, tem manjša je razlika med tema 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 delovnega prebivalstva v SR Sloveniji, če delavnost ne bi bila odvisna od spola. Ker bi bilo v tem primeru recimo 54,8 % od skupno 693 tisočev moških delovnih, bi bilo število delovnih 693 . 0,548 = 380 tisoč. Podobno izračunamo druge podatke.

Tako dobimo tabelo 13.26, 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 SRS, če delavnost ne bi bila odvisna od spola

Delavnost / Spol	Delovni	Vzdrževani	Skupno
Moški	380	313	693
Ženske	424	349	773
Skupno	804	662	1466

Če primerjamo tabelo 13.26 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 = 804/1466, moremo izračunati teoretično frekvenco tudi takole: 380 = 693.804/1466. 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

$$Q = \frac{ad-bc}{ad+bc} \qquad (13.42)$$

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 funkcijski 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

$$V = \frac{ad - bc}{\sqrt{(a+b).(a+c).(b+d).(c+d)}} \qquad (13.43)$$

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:

$$ad - bc = 474.443 - 219.330 = +137712$$

$$\delta = \frac{ad - bc}{N} = +137712/1466 = 93,937$$

$$Q = \frac{ad - bc}{ad + bc} = \frac{+137712}{474.443 + 219.330} = +0,488$$

$$V = \frac{ad - bc}{\sqrt{(a+b).(a+c).(b+d).(c+d)}} = \frac{+ 137712}{\sqrt{693.804.662.773}} = +0,258$$

Tabela 13.27 Število prebivalstva po spolu in delavnosti 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	163	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 dělovnostjo, opazimo izjemno nizek koeficient asociacije za Slovenijo (Q = +0,488) in izjemno visok za Makedonijo (Q = +0,646). Za druge republike pa so koeficienti asociacije med Q = +0,583 in Q = +0.587. To je v skladu z razmerami, ker so v Sloveniji zaradi različnega položaja žensk 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

K o n t i n g e n c a

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:

$$\chi^2 = \sum \frac{(f-f')^2}{f'} \qquad (13.45)$$

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 $(f-f')^2$

d) kvadrate diferenc delimo z ustreznimi teoretičnimi frekvencami f',

e) če seštejemo dobljene izraze $(f-f')^2/f'$ izračunane za vsa polja v kontingenčni tabeli, dobimo χ^2,

f) včasih je primerneje, če izračunavamo χ^2 po obrazcu

$$\chi^2 = \Sigma \frac{f^2}{f'} - N \qquad (13.46)$$

V tem primeru namesto diferenc $f-f'$ kvadriramo stvarne frekvence f; to je večinoma enostavnejše.

Iz osnovnega obrazca za χ^2 sklepamo, da je $\chi^2 = 0$, če ni razlik med stvarnimi in teoretičnimi frekvencami; če sta dva 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 moremo 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 SR Slovenijo v letu 1955.

Tabela 13.30 Sklenjeni zakoni po prejšnjem stanu ženina in neveste v SRS 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 SRS 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
razveden	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

$$\Phi^2 = \chi^2/N = 3495,997/13740 = 0.254$$

$$C = \sqrt{\frac{\chi^2}{\chi^2+N}} = \sqrt{\frac{3495,997}{3495,997 + 13740}} = 0,450$$

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 rezultat neposredne - vzročne odvisnosti, del povezanosti pa 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: $r_{13} = +0,34802;$

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 eleketrične energije (x_3), dobimo:

$$r_{12.3} = \frac{0,03383 - 0,34802 \cdot 0,72000}{\sqrt{1 - 0,34802^2}\ \sqrt{1 - 0,72000^2}} = -0,333$$

Dobljeni rezultat se zdi nerazumljiv, ker kaže, da produktivnost dela, merjena s proizvodnjo na enega zaposlenega, pri isti porabi eleketrične 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 Dosedaj smo obravnavali probleme odvisnosti nekega pojava od enega faktorja. Vemo pa, da posamezni socialno-ekonomski pojavi niso odvisni od enega, marveč 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

$$x_1' = f(x_2,x_3) \qquad (13.51)$$

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

$$I_{1.23} = \sqrt{1 - \frac{\sigma^2_{1.23}}{\sigma^2_1}} \qquad (13.52)$$

pri tem pomeni: σ^2_1 = skupna varianca za znak x_1; $\sigma^2_{1.23}$ = nepojasnjeni del skupne variance za znak x_1.

Podobno kot pri enostavni regresiji izračunamo tudi za multiplo regresijo standardno napako ocene po obrazcu

$$\sigma_{1.23} = \sigma_1\sqrt{1 - I^2_{1.23}} \qquad (13.53)$$

L i n e a r n a m u l t i p l a k o r e l a c i j a

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:

$$x_1' = \bar{x}_1 + b_{12.3}(x_2-\bar{x}_2) + b_{13.2}(x_3-\bar{x}_3) \qquad (13.54)$$

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

$$b_{12.3} = \frac{\sigma_1}{\sigma_2} \cdot \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}; \quad b_{13.2} = \frac{\sigma_1}{\sigma_3} \cdot \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \qquad (13.55)$$

Determinacijski koeficient multiple korelacije $R^2_{1.23}$, 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

$$R^2_{1.23} = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \qquad (13.56)$$

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 so 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 1.000 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 MW_h, 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; $\bar{x}_3 = 900{,}70$ MW_h

$\sigma_1 = 813{,}0$; $\sigma_2 = 244{,}9$; $\sigma_3 = 1231{,}5$

$r_{12} = +0{,}03383$; $r_{13} = +0{,}34802$; $r_{23} = 0{,}72000$.

Iz teh podatkov dobimo po obrazcih 13.55

$$b_{12.3} = \frac{813,0}{244,9} \cdot \frac{0,03383 - 0,34802 \cdot 0,72000}{1 - 0,72000^2} = -1,49404$$

$$b_{13.2} = \frac{813,0}{1231,5} \cdot \frac{0,34802 - 0,03383 \cdot 0,72000}{1 - 0,72000^2} = +0,44367$$

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

$$x_1' = 3027 - 1,49404x_2 + 0,44367x_3$$

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 13.56 determinacijski koeficient multiple korelacije $R^2_{1.23}$, dobimo

$$R^2_{1.23} = \frac{0,03383^2 + 0,34802^2 - 2.0,03383.0,34802.0,72000}{1 - 0,72000^2} = 0,21866$$

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

$$R_{1.23} = \sqrt{0,21866} = 0,468$$

Temu ustrezno je velika tudi standardna pogreška ocene 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 produktivnosti 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

V Z O R Č E N J E

Osnove vzorčenja

14.1 Vzporedno s hitrim ekonomskim in družbenim razvojem se je v zadnjih desetletjih vedno bolj večala potreba po statističnih podatkih v najrazličnejših področjih.

Stare metode zbiranja in obdelave podatkov niso mogle zadostiti novim zahtevam, ker so bile drage, počasne, zvezane z angažiranjem velikega števila ljudi itd. Razen tega so se te metode zaradi posebnih pojavov 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 npr. metoda izbora tipičnih enot, monografija, metoda opazovanja osnovne mase itd. pa so izdelane na poznavanju dela populacije. Vendar imajo vse te metode ocenjevanja večinoma iste pomanjkljivosti. Vse so subjektivne, za sestavljanje ocen pa moramo razmeroma zelo podrobno poznati osnovno populacijo; razen tega pri nobeni od njih ne vemo, kolikšna je razlika ocene od prave vrednosti, ki bi jo dobili, če bi opazovali populacijo, oziroma koliko so te ocene zanesljive.

Vseh teh pomanjkljivosti, ki jih imajo ocene gornjih tipov, pa nima metoda vzorčenja. Po tej metodi ocenjujemo parametre populacije iz dela enot populacije - vzorca. Enote vzorca izberemo iz populacije podobno kot številke pri loteriji po metodi slučajnega izbora tako, da ima vsaka enota populacije enako možnost, da je izbrana. Če je vzorec, kakor pravimo, slučajnosten, ima ta ocena kakovosti, ki jih nobena drugačna ocena nima. Za slučajnosten vzorec namreč veljajo zakonitosti, ki bistveno pripomorejo h kakovosti ocene. Te pa moremo odkriti z uporabo verjetnostnega računa. Zaradi tega je osnova metode vzorčenja verjetnostni račun. Zakonitosti vzorčenja pa veljajo le, če so enote izbrane naključno.

14.2 Metoda vzorčenja je v svojih posebnih oblikah v nekaterih znanostih že dolgo znana kot raziskovalna metoda. Tako je vzorčenje že razmeroma dolgo znano v agronomiji, biologiji, medicini, psihologiji itd. Vzrok, da se je metoda vzorčenja uveljavila najprej v teh znanostih, ni naključje. Raziskovalec v teh področjih je bil že od vsega začetka nujno

navezan na opazovanje samo dela populacije, ker celotne populacije ne obstajajo. Tako je za agronoma trideset preskusnih parcelic samo vzorec iz populacije neomejenega števila parcelic, ki bi jih 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 ob enakih pogojih. Enako vzamemo določeno število izdelkov, ki jih proizvedemo na določenem stroju, za vzorec iz neomejene populacije izdelkov, ki bi jih stroj proizvedel ob enakih pogojih.

V teh znanostih se je vzorčenje uveljavilo razmeroma zgodaj, ker s popolnim opazovanjem nismo mogli zajeti neomejenih - umišljenih populacij. Ravno to pa je vzrok, da se je vzorčenje razmeroma pozno začelo uveljavljati tudi v socialno-ekonomskih znanostih. Te znanosti namreč večinoma obravnavajo populacije, ki imajo končno število enot. Te pa je možno opazovati s klasičnimi metodami opazovanja, s popisi, s tekočimi službami itd. Vendar je potreba po statističnih podatkih prerasla možnosti, ki jih dajejo klasične metode. Zato so izdelali specifične metode vzorčenja, ki omogočajo objektivno in kakovostno ocenjevanje podatkov tudi za to področje. Te metode so sedaj splošno orodje statistike in v mnogih področjih izpodrivajo klasične metode opazovanja.

Prednosti in pomanjkljivosti metode vzorčenja

14.3 Prednosti vzorčenja pred drugimi metodami ocenjevanja, a v mnogih primerih tudi pred metodami popolnega opazovanja, se 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 boljši, bolje poučeni itd.

d) Zaradi izbire kadrov in možnosti kontrole so napake pri popisovanju kar najbolj skrčene.

e) Čas, v katerem dobimo rezultate, je znatno krajši kot pri popolnem opazovanju; to pa poveča uporabno vrednost podatkov.

f) Pogrešek, ki izvira iz tega, da nismo opazovali celotne populacije, temveč samo del, moremo pri vzorčenju objektivno

izmeriti in uravnavati. Te lastnosti nima nobega druga metoda ocenjevanja. Ta pogrešek, ki ga moremo oceniti, je dostikrat celo manjši kakor pa je rezultat popisnih napak pri celotnem opazovanju. Popisnih napak je predvsem kriv velik obseg populacije, sodelovanje velikega števila popisovalcev, ki jih ne moremo uspešno poučiti in kontrolirati,itd.

g) V raziskovalnem delu je metoda vzorčenja edino možna, ker umišljenih - hipotečnih populacij v nobenem primeru ne moremo zajeti.

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 območja.

c) Vzorčenje ne da zanesljivih podrobnih rezultatov. Zato z njim ne moremo obdelovati kombinacije več znakov.

d) Zanesljivost ocen za posamezne podatke je različna.

e) Načrt vzorčenja je bolj zapleten kakor načrt popolnega opazovanja. Enako je računanje ocen in pokazovalcev zanesljivosti ocen razmeroma zamotano.

Uporaba vzorčenja pri statističnem opazovanju

14.4 Vzorčenje moremo uporabiti v vseh stopnjah statističnega opazovanja.

Pri izvedbi statistične akcije more vzorec nadomestiti popolno opazovanje ali pa ga uporabimo kot dopolnilo k popolnemu opazovanju. Tako npr. z vzorcem v nekaterih letih v Jugoslaviji nadomestimo popis živine. Kot dopolnilo pa vzorčenje uporabimo, če podatke, ki jih potrebujemo zelo podrobno tudi za manjša območja, zberemo s popisom. Podatke, ki jih potrebujemo samo za večja območja pa istočasno opazujemo z vzorcem. Zato prve podatke zberemo od vseh enot populacije, druge pa samo od vnaprej naključno izbranih enot.

Pri obdelavi statističnega gradiva moremo uporabiti vzorčenje za dva namena. Obdelava velikih popisov traja razmeroma dolgo. V tem času podatki velikokrat že izgube aktualnost. Z vzorcem predhodnih rezultatov pa moremo dobiti ocene podatkov znatno pred koncem popolne obdelave. V tem primeru je popolna obdelava tudi potrebna, ker da vzorec samo globalne, ne pa podrobnih rezultatov.

Moremo pa vzorčenje uporabiti tudi za dopolnilno obdelavo. Čeprav s statističnim opazovanjem zberemo popolne

podatke za vsa vprašanja, včasih popolno obdelamo le nekatera od teh vprašanj. Druge pa obdelamo z vzorcem v dopolnilni obdelavi, ker sicer bodisi zaradi pomanjkanja sredstev ali drugih vzrokov ne bi teh podatkov sploh obdelali.

Z vzorcem moremo tudi kontrolirati izvedbo statistične akcije v vseh stopnjah. Z njim ocenimo delo popisovalcev ali obdelavo statističnih podatkov v celoti. Vzorčenje uporabljamo celo za popravo podatkov, ki smo jih zbrali s popisom. Pri popisu živine, pa tudi pri drugih statističnih akcijah, neposredno po popisu izvedemo kontrolni vzorec. Kakovostno dobri, zanesljivi in posebej za to poučeni kontrolorji z vzorcem ugotove, koliko so zbrani podatki različni od stvarnih. Podatke zberejo kontrolorji z drugačnimi sredstvi, bolj vestno in natančno kakor popisovalci. Tako ugotovimo učinek napak, ki jih z drugimi sredstvi kontrole ne moremo ugotoviti. Iz podatkov kontrolnega vzorca pa ocenimo za vsako vrsto živine oziroma podatek posebej, koliko je stvaren podatek oziroma za koliko je treba v skupnem ta podatek popraviti. Vzorčna kontrola ima namen ugotoviti skupni učinek napak, ne pa napake posameznika.

Čeprav je v problemih, ki smo jih nakazali v prejšnjem odstavku, število enot v vzorcu razmeroma majhen del celotne populacije, je število enot vseeno razmeroma veliko, saj je imel npr. 2 % vzorec popisa prebivalstva v Sloveniji v letu 1953 približno 30.000 enot.

Vzorčenje se je razvilo v eno izmed osnovnih metod pri proučevanju javnega mnenja, analize tržišča, kontrole proizvodnje in proizvodnega procesa itd. Pri teh problemih so vzorci razmeroma majhni in v dosti primerih ne presegajo nekaj deset enot populacije. Seveda rezultati takih vzorcev niso tako zanesljivi kakor rezultati vzorcev z več tisoči enot. Vendar je njihova zanesljivost glede na problem, ki ga s takimi vzorci rešujemo, večinoma ustrezna.

14.5 Normalna distribucija. Osnovna verjetnostna distribucija v vzorčenju je normalna porazdelitev. Mnogo slučajnostnih spremenljivk, ki jih sestavljamo v vzorčenju, se distribuira v normalni distribuciji, če ne neposredno, pa vsaj asimptotično. Normalna distribucija je tudi osnova za druge teoretične distribucije. V normalno distribucijo preide asimptotično večina teoretičnih distribucij.

Opis normalne distribucije

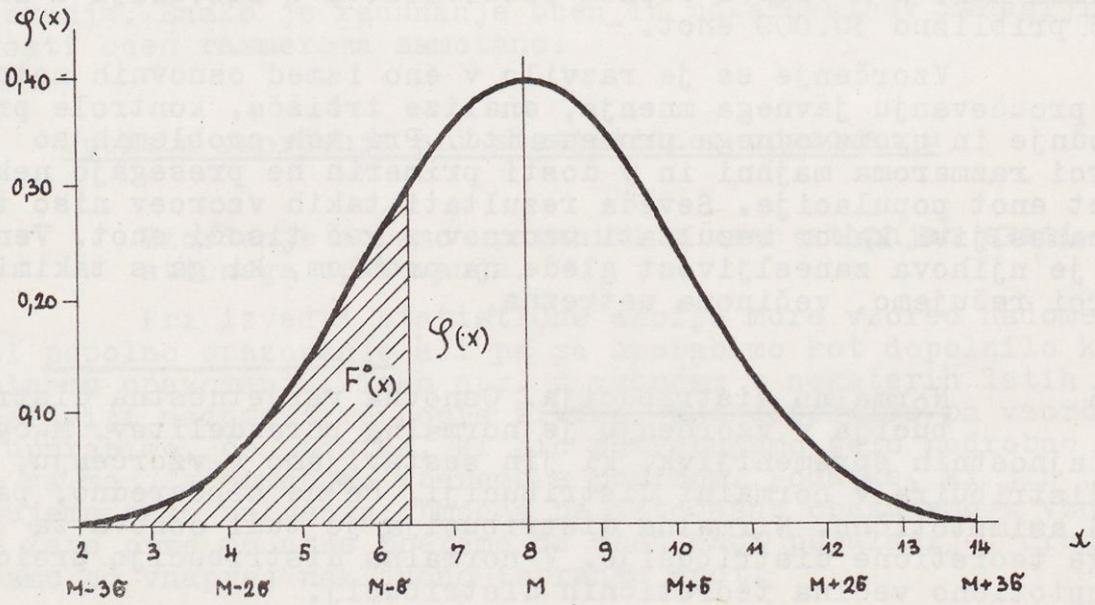
14.6 Normalna distribucija je določena z dvema parametroma: z aritmetično sredino M, ki je rezultat splošnih vplivov, in s standardnim odklonom σ, ki je rezultat posamičnih oziroma slučajnih vplivov.

Gostota relativne frekvence za normalno distribucijo je dana s funkcijo

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} \qquad (14.1)$$

Ta enačba pokaže, v kakšni odvisnosti je relativna gostota frekvence $\varphi(x)$ z vrednostjo x.

Če vzamemo vrednost znaka x za absciso, ustrezno vrednost relativne gostote frekvence $\varphi(x)$ pa za ordinato, dobimo sliko normalne distribucije (glej sliko 14.1). Iz enačbe (14.1) in slike spoznamo, da je normalna distribucija unimodalna, simetrična in zvonasta, zaradi česar je zanjo M = Me = Mo. Največja gostota relativne frekvence je za 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

14.7 Čeprav imajo normalno distribucije različne parametre M in SD, so si med seboj podobne. Slike za normalne distribucije z različnimi aritmetičnimi sredinami so samo premaknjeme v levo ali desno. Normalna distribucija pa je bolj stisnjena, če je standardni odklon manjši in bolj raztegnjena, če je standardni odklon velik.

Za vsak poseben primer normalne distribucije bi bilo zelo neprikladno neposredno računati gostoto relativne frekvence $\varphi(x)$, vrednosti za porazdelitveno funkcijo $F(\bar{x})$ itd. Ker so si vse normalne distribucije podobne, velja zanje nekaj splošnih lastnosti, ki to nalogo poenostavijo.

Iz slike 14.1 sklepamo, da je $F(x=m) = 0,50$ za vsako porazdelitev, ne glede na to, kolik je M. V vsaki normalni distribuciji je zaradi simetrije glede na aritmetično sredino polovica vrednosti pod, polovica pa nad aritmetično sredino M. Celo več! Za vsako normalno distribucijo, ne glede na to, kakšen je M in σ, je v razmiku od $M - \sigma$ do $M + \sigma$ relativna frekvenca enaka 0,6827. Podobno je v razmiku $M - 2\sigma$ do $M + 2\sigma$ relativna frekvenca enaka 0,9545, v razmiku $M - 3\sigma$ do $M + 3\sigma$ 0,9973 itd. V splošnem moremo reči, da je za normalno distribucijo relativna frekvenca v razmiku $M - z\sigma$ do $M + z\sigma$ odvisna samo od koeficienta z, nič pa od M in σ.

Enako je tudi s porazdelitveno funkcijo oziroma ploščino v razmiku $-\infty$ do $M + z\sigma$. Za porazdelitveno funkcijo za normalno distribucijo velja, da je

$$F^{o}(x = M + z) = F^{o}(z) \qquad (14.2)$$

za vse normalne distribucije odvisna samo od koeficienta z, nič pa od M in σ.

Standardizirana normalna distribucija

14.8 Vse normalne distribucije so si med seboj podobne. Površine v posameznih razmikih za vsako normalno distribucijo so 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.

Iz lastnosti standardiziranega odklona spoznamo, da se "z" za normalno populacijo distribuira normalno z $M_z = 0$ in $\sigma_z = 1$. 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

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad (14.3)$$

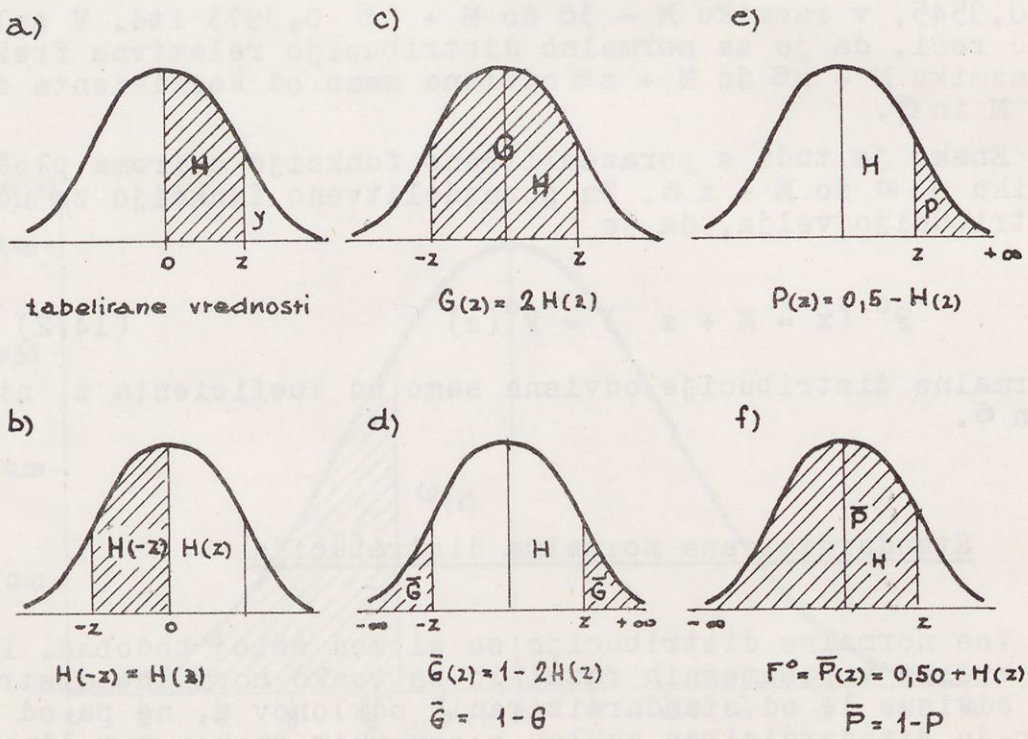
porazdelitvena funkcija $F^{o}(z)$ pa

$$F^{o}(z) = \int_{-\infty}^{z} \varphi(x)\, dx = \frac{1}{2\pi} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} \cdot dz \qquad (13.6)$$

Običajno imamo tabelirane za standardizirano normalno distribucijo površine za razmik od 0 do z

$$H(z) = \int_{0}^{z} \varphi(z) \cdot dz \qquad (14.4)$$

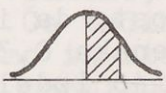
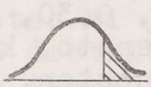
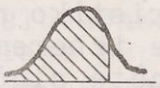
Za normalno distribucijo potrebujemo večkrat različno opredeljene površine, ta elirane pa imamo samo površine H v razmiku od 0 do z. Zato so na sliki 14.2 nakazani odnosi med posameznimi površinami, ki pridejo v praksi v poštev v različnih primerih.



Slika 14.2 Različne površine pod standardizirano normalno distribucijo

V tabeli 14.1 navajamo nekaj najznačilnejših odnosov med standardiziranim odklonom z in ustreznimi površinami H, G, $\bar{G}$, P in $\bar{P}$.

Tabela 14.1 Odnosi med standardiziranim odklonom z in ustreznimi relativnimi frekvencami H, G, $\bar{G}$, P in $\bar{P}$.

	v o d s t o t k i h				
					
z	H	G = 2H	$\bar{G}$ = 1-2H	P = ·5-H	$\bar{P}$ = ·5+H
0,6745	25,00	50,00	50,00	25,00	75,00
1,000	34,13	68,26	31,74	15,87	84,13
1,645	45,00	90,00	10,00	5,00	95,00
1,960	47,50	95,00	5,00	2,50	97,50
2,000	47,73	95,46	4,54	2,27	97,73
2,326	49,00	98,00	2,00	1,00	99,00
2,576	49,50	99,00	1,00	0,50	99,50
3,000	49,87	99,74	0,26	0,13	99,87
3,090	49,90	99,80	0,20	0,10	99,90
3,291	49,95	99,90	0,10	0,05	99,95

14.9 Normalna distribucija kot verjetnostna distribucija.

Najpomembnejša uporaba normalne distribucije je v vzorčenju. Normalno distribucijo vzamemo v vzorčenju bodisi kot predpostavko za določeno populacijo, ki jo proučujemo, ali kot zakonitost distribucije za količine, ki so pomembne v vzorčenju.

Normalna distribucija se prepleta prek vseh delov in problemov vzorčenja, tako pri velikih vzorcih kot pri malih, pri ocenjevanju kot pri preizkušanju domnev.

14.10 Teorija vzorčenja je zgrajena na verjetnostnem računu. Zato morajo biti izpolnjene osnovne postavke, ob katerih veljajo zakonitosti, ki smo jih odkrili za vzorčenje.

Če predpostavimo, da poznamo osnove verjetnostnega računa, se omejimo na osnovno črto, ki je važna v vzorčenju. Verjetnostni račun se ukvarja z računanjem in preračunavanjem verjetnosti za katerekoli 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 izbora za vsako enoto. Če dosežemo slučajnost izbora, se pojem apriorne verjetnosti sklada s pojmom relativne frekvence v statistiki. Če je v populaciji relativna frekvenca za neko karakteristiko npr. 0,30, pomeni, da je verjetnost, da iz te populacije izberemo enoto, ki ima to značilnost, enaka 0,30, če je izbor enote slučajnosten. Če imamo npr. populacijo izdelkov, v kateri je relativna frekvenca za neuporabne artikle 0,05, je verjetnost, da iz te populacije s slučajnostnim izborom izberemo izdelek, ki je neuporaben, enaka 0,05.

14.11 Zaradi zakona o velikih številih velja, da se za slučajnostni izbor enot aposteriorna in apriorna verjetnost skladata. To pomeni, da se kvocient med število slučajnostno izbranih enot, ki imajo dano značilnost, in skupnim številom vseh slučajnostno izbranih enot približuje relativni frekvenci za to značilnost, če število izbranih enot večamo.

Iz gornjega sledi da moremo vzeti porazdelitev relativnih frekvenc za verjetnostno distribucijo, če je izbor iz te populacije slučajnosten. Pri tem se relativna frekvenca za določeno vrednost ali razmik vrednosti sklada z apriorno verjetnostjo, da iz populacije slučajnostno izberemo enoto, ki ima ustrezno vrednost ali vrednosti iz ustreznega razmika. 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.

Če vzamemo populacijo, ki je normalno distribuirana, je npr. verjetnost, da iz nje slučajnostno izberemo enoto, ki je v mejah $M - \sigma$ do $M + \sigma$, enaka relativni frekvenci za ta razmik. Ta pa je po tabeli 14.1 enaka 0,6826. Če slučajnostni izbor iz populacije ponavljamo, pričakujemo, da v povprečju približno 2/3 vseh slučajnostno izbranih vrednosti leži v gornjem razmiku. Podobno dobimo ustrezne verjetnosti za poljuben razmik.

14.12 <u>Pojem tveganja.</u> V vsakdanjem življenju vzamemo dogodke, ki so zelo verjetni, praktično tako, kakor da se bodo za gotovo zgodili. Zato računamo, da nas pri sprehodu ne bo povozil avto, ker je velika verjetnost, da nas resnično ne bo. Ko kupujemo konzervo, računamo, da ni pokvarjena, ker je zelo verjetno, da res ni, itd. Pri vseh teh sklepih pa je določena stopnja tveganja. 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 lanske proizvodnje ni pokvarjena, kakor da ni pokvarjena konzerva iz letošnje proizvodnje. Tveganje je torej verjetnost, da se dogodek, ki ga pričakujemo, ne zgodi.

Pojmu tveganja v vsakdanjem življenju popolnoma ustreza pojem tveganja v verjetnostnem računu oziroma statistiki.

Vrednosti standardizirane normalne distribucije teoretično leže na vsem razmiku od $-\infty$ do $+\infty$. Vendar praktično vzamemo, da vse vrednosti leže v razmiku med -3,29 do +3,29. Če trdimo, da je vrednost enote, ki jo slučajnostno izberemo iz standardizirano normalno distribuirane populacije v razmiku od -3,29 do +3,29, je tveganje te izjave 0,001 (glej tabelo 14.1). To pomeni: le z verjetnostjo 0,001 moremo dobiti pri slučajnostnem izboru enoto, za katero je vrednost izven navedenega razmika. To tveganje pa moremo tolmačiti tudi drugače.

Glede na aposteriorno verjetnost moremo sklepati: če hoče določen raziskovalec preizkusiti trditev, da slučajnostno izbrana enota leži med -3,29 in +3,29 in ponavlja izbor enote večkrat, v velikem številu ponavljanj za 0,001 ali 1 ‰ enot ugotovi, da njihove vrednosti ne leže v imenovanem razmiku.

Stopnjo tveganja moremo poljubno spreminjati. Če vzamemo širši razmik, je tveganje manjše, če vzamemo ožji, je stopnja tveganja večja. Za vsak poseben primer moremo stopnjo tveganja ugotoviti tako, da izračunamo verjetnost, da se napovedani dogodek ne zgodi.

14.13 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 razmik, 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 dvostranka, drugi dve pa sta enostranski trditvi.

Stopnja tveganja, s katero dajemo izjave, ki smo jih pravkar navedli, more biti poljubna. Običajne stopnje tveganja so: 0,10, 0,05, 0,01 in 0,001. Stopnja tveganja 0,10 je zelo velika in pomeni, da moremo pričakovati, da se v poprečju v enem od desetih poskusov izkaže, da trditev, ki jo postavljamo, ne drži. Zato to stopnjo tveganja redko uporabljamo. Najpogosteje sklepamo s tveganjem 0,05. To pomeni, da moremo pri takem sklepanju pričakovati, da se v enem izmed dvajsetih primerov v poprečju izkaže, da naša trditev ne drži. Veliko bolj zanesljive so trditve s tveganjem 0,01 ali 0,001. Pri teh se izkaže, da trditev ne velja poprečno v enem od sto oziroma od tisoč preskusov.

Razmik, v katerem je z določeno stopnjo tveganja vrednost za slučajnostno izbrano enoto, imenujemo kritični razmik, meji pa kritični meji.

Glede na pomen P na splošno velja za slučajnostno izbrano enoto iz standardizirano normalne distribucije.

	Razmik	Tveganje
s tveganjem 2P	$-1,64 < z < +1,64$	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	$-1,28 < z$	0,10
	$-1,64 < z$	0,05
	$-2,32 < z$	0,01
	$-3,09 < z$	0,001
s tveganjem P	$z < +1,28$	0,10
	$z < +1,64$	0,05
	$z < +2,32$	0,01
	$z < +3,09$	0,001

Osnovna populacija. Vzorec, Populacija vseh vzorcev

14.14 Osnovna populacija. Populacijo, za katero skušamo oceniti parametre z vzorčenjem, imenujemo osnovno populacijo. Osnovna populacija je npr. populacija vseh kmetijsih gospodarstev v SR 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 npr. populacija industrijskih podjetij ali v ožjem smislu populacija vrednosti proizvodnje posameznih podjetij, populacija števila delavcev posameznih podjetij itd. Te populacije imenujemo končne osnovne populacije, ker imajo končno število enot.

Z vzorčenjem pa obravnavamo tudi osnovne populacije, ki imajo neomejeno število enot. Te populacije sicer stvarno ne obstajajo, vendar jih moremo miselno dojeti in razumeti. Te

vrste populacij obravnavamo predvsem pri raziskovalnem delu. Umišljena ali hipotetična populacija, kot imenujemo take vrste populacij, je npr. skupnost vseh ponovitev določenega poskusa ob enakih pogojih.

Enako imamo za hipotetično populacijo skupnost vseh izdelkov, ki bi jih proizvedel nek stroj ob enakih pogojih, če bi delal neomejeno. Tudi iz tega primera vidimo, da ima taka umišljena populacija neomejeno število enot, da pa obstaja samo miselno, ne pa stvarno. Praktično je namreč nemogoče ustvariti enake pogoje za razmeroma veliko - kaj šele za neomejeno proizvodnjo predmetov, ne glede na to, da ne mremo take populacije nikdar izčrpati.

Osnovne populacije z neomejenim številom enot moremo izpeljati tudi iz zveznih populacij. Parcelo z obsegom enega hektarja moremo razdeliti v populacijo s končnim številom površinic po 1 m^2. Če pa vzamemo drugo osnovo in imamo za enoto populacije vsako parcelico kvadratne oblike s površino 1 m^2 ne glede na to, kako je položena, in dopustimo, da sé med seboj prekrivajo, dobimo osnovno populacijo z neomejenim številom enot. Podobno populacijo dobimo iz osnovne zvezne populacije velikega kosa tkanine, zvezne populacije proizvodnje cementa v določenem dnevu itd.

Enote osnovne populacije so posamezno gospodarstvo, posamezno industrijsko podjetje, poskus ob enakih pogojih, parcelica po 1 m^2, 10 dkg cementa itd. Znaki teh enot so vse značilnosti, ki jih proučujemo. Parametri osnovnih populacij pa so vsota podatkov, srednje vrednosti, strukturni odstotki, mere variacije, mere korelacije itd.

14.15 Vzorec. Iz osnovne populacije izberemo delno populacijo, ki jo imenujemo vzorec. Enote vzorca morajo biti izbrane tako, da ima vsaka enota enako možnost, da je izbrana v vzorec. To dosežemo s slučajnostnim izborom. Vzorec je sam zase tudi populacija, ker ima svoje enote (enote osnovne populacije, ki smo jih izbrali), te pa znake, ki so isti kot znaki v osnovni populaciji. Kakor za vsako drugo populacijo moremo iz podatkov v vzorcu izračunati sredine, proporce, mere variacije in korelacije itd.

Ker je glede na število enot teorija vzorčenja različna, ločimo male vzorce od velikih. Stroge meje med malimi in velikimi vzorci ni. Običajno imamo vzorce, ki imajo nekaj deset enot, za male, če obseg vzorca ne prekorači sto enot. Število enot v velikih vzorcih je običajno znatno večje. Včasih znaša tudi nekaj tisoč ali celo deset tisoč enot. Medtem ko uporabljamo male vzorce predvsem v raziskovalnem delu pri preskušanju domnev, z velikimi vzorci običajno nadomeščamo popolna opazovanja.

Glede na način izbora ločimo vzorce s ponavljanjem od vzorcev brez ponavljanja. Ta razlika izvira iz načina, kako izbiramo enote. Pri vzorčenju s 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 kot obravnavanje vzorcev s ponavljanjem. Ta razlika pa ne pride v poštev pri hipotetičnih in zveznih populacijah in se njen pomen tudi pri končnih populacijah tem bolj izgublja, čim večja je osnovna populacija. Običajno v praksi za končne populacije jemljemo vzorce brez ponavljanja, ker so bolj logični, razen tega pa so ocene vzorcev brez ponavljanja zanesljivejše kakor ocene iz vzorcev s ponavljanjem.

14.16 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 vseh istih enot kot 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 poprečja, proporci, mere variacije in korelacije itd., ki jih izračunamo iz podatkov za posamezni vzorec. Ne da bi navajali določene primere, sklepamo, da ti znaki variirajo, ker so izračunani za vsak vzorec iz podatkov za posamezni vzorec. Skupnost vseh vzorcev je torej populacija, v kateri so posamezni vzorci enote, poprečja, proporci, mere variacije in korelacije itd. pa znaki teh enot.

14.17 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

$$\binom{N}{n} = \frac{N(N-1) \ldots (N-n+1)}{1.2. \ldots n} \qquad (14.5)$$

Število vseh možnih vzorcev s ponavljanjem pa je enako številu kombinacij s ponavljanjem po n enot iz N enot. To število pa je

$$\binom{N+n-1}{n} = \frac{N(N+1) \ldots (N+n-1)}{1.2. \ldots n} \qquad (14.6)$$

Če vzamemo za zgled razmeroma majhno osnovno populacijo z N = 75 enotami in vzorec z n = 25 enotami, pokaže račun, da je število vseh možnih vzorcev brez ponavljanja 52,59 trilijonov, število vseh možnih vzorcev s ponavljanjem pa 181890 trilijonov. Zgled pokaže, da je že za tako majhne populacije in vzorce obseg populacije vseh možnih vzorcev velikanski in veliko večji za vzorce s ponavljanjem kakor za vzorce brez ponavljanja.

Podajamo pregled osnovnih značilnosti za osnovno populacijo (OP), vzorec (V) in populacijo vseh vzorcev (PVMV):

Populacija	Število enot		Enota	Znak	Populacija	Parameter
	N končen	N neskončen				
OP	N	∞	gospodarstvo	x,A	vsa gospodarstva	M_x; P_A
V	n	n	gospodarstvo	x,A	vzorec	$\bar{x}$; p_A
PVMV	$\binom{N}{n}\Big/\binom{N+n-1}{n}$	∞	vzorec	$\bar{x}$, pA	skupnost vseh možnih vzorcev	$M_{\bar{x}}$; M_{pA} $Var\bar{x}$

Zakonitosti v populaciji vseh vzorcev

14.18 Ker je število vseh možnih vzorcev v praktičnih primerih izredno veliko, ne moremo neposredno proučevati populacije vseh možnih vzorcev, razen če ima populacija malo enot. Take populacije pa so brez praktičnega pomena. Ne moremo si misliti, da bi npr. frekvenčno porazdelitev vzorčnih poprečij sestavili tako, da bi iz populacije izbrali vse možne vzorce, za vsak vzorec izračunali aritmetično sredino $\bar{x}$, iz vseh dobljenih poprečij pa sestavili frekvenčno porazdelitev. Prav tako bi bilo težko neposredno računati aritmetično sredino poprečij vzorcev, varianco vseh poprečij itd. Po tej poti je torej težko ali pa celo nemogoče proučevati populacije vseh vzorcev.

Vendar s teoretičnim razglabljanjem pridemo do zvez, ki veljajo med osnovno populacijo in populacijo vseh možnih

vzorcev. Izkaže se, da moremo za dosti primerov iz podatkov za osnovno populacijo sklepati, kakšne so porazdelitve in parametri v populaciji vseh vzorcev. K uspešnemu odkrivanju zakonitosti zvez med osnovno populacijo in populacijo vseh vzorcev bistveno pripomore verjetnostni račun; ta je, kot smo že omenili, osnova vzorčenja. Te zakonitosti, ki so za nekatere probleme znane natančno, za druge pa samo približno, izkoristimo za praktično uporabo pri ocenjevanju in sklepanju iz vzorca na osnovno populacijo.

Za posamezne statistične količine so zakonitosti več ali manj različne in odvisne od podatka, ki ga proučujemo, od osnovne populacije in od tipa vzorca. Zato bomo za aritmetično sredino, ki je eden izmed najvažnejših in najpogostejših statističnih parametrov, razložili splošna načela vzorčenja.

Enostavno slučajnostno vzorčenje

Ocenjevanje aritmetične sredine

14.19 Vzorčenje, pri katerem ima vsaka enota populacije enako verjetnost, da je izbrana v vzorec, imenujemo enostavno slučajnostno vzorčenje.

Za enostavne slučajnostne vzorce, izbrane iz neomejene populacije, in za enostavne slučajnostne vzorce, ki so izbrani iz končnih populacij z izborom s ponavljanjem, veljajo za aritmetične sredine vzorcev $\bar{x}$ naslednje zakonitosti:

a) Aritmetična sredina iz poprečij $\bar{x}$ za vse možne vzorce je enaka aritmetični sredini $\bar{x}$ v osnovni populaciji. V skladu s pojmi iz verjetnostnega računa se ta zakonitost glasi tudi takole: Matematično upanje za slučajnostno spremenljivko $\bar{x}$ slučajnostnih vzorcev je enako aritmetični sredini v populaciji M_x

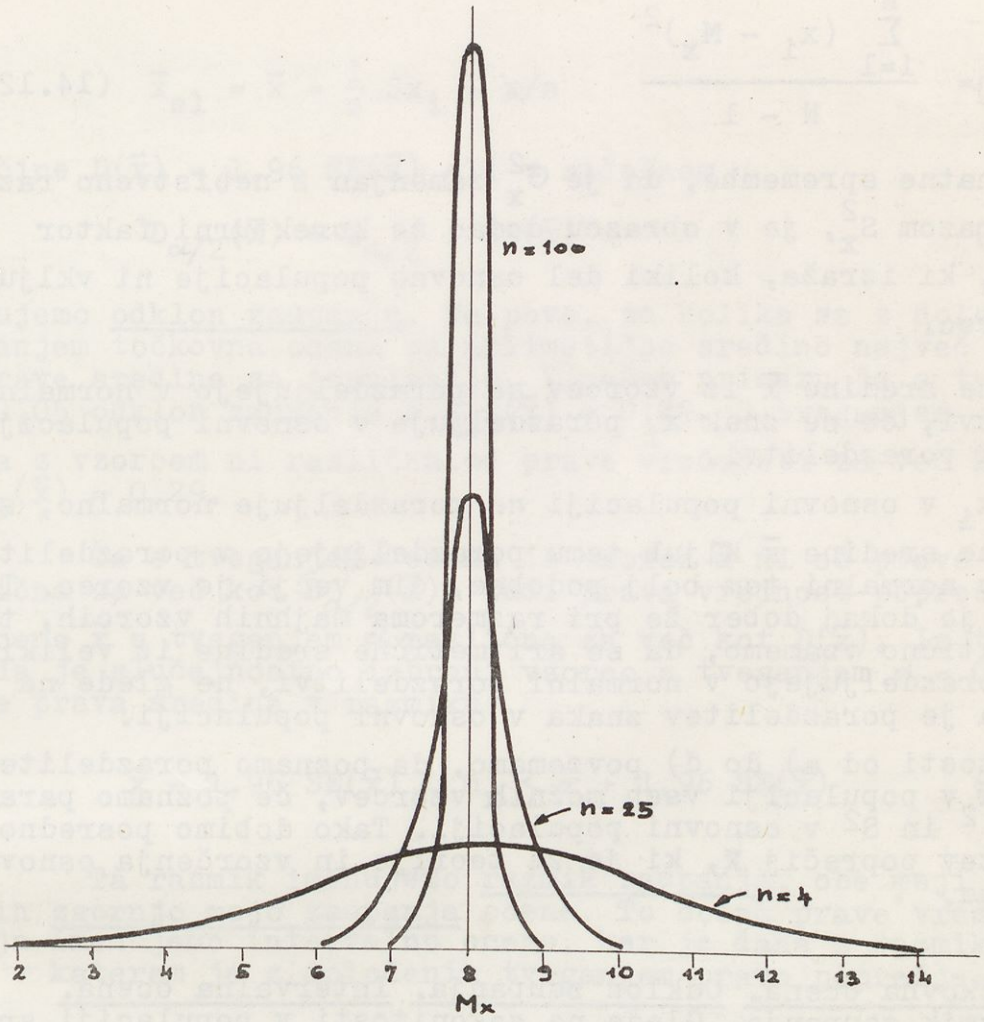
$$M_{(\bar{x})} = E_{(\bar{x})} = M_x \qquad (14.7)$$

b) Varianca aritmetičnih sredin slučajnostnih vzorcev s ponavljanjem $Var(\bar{x})$ je n krat manjša kot varianca znaka x v osnovni populaciji

$$Var_{\bar{x}} = \frac{\sigma_x^2}{n} \qquad (14.8)$$

Če iz variance aritmetičnih sredin vzorcev izračunamo kvadratni koren, dobimo standardni odklon za aritmetične sredine vzorcev. Na splošno ga imenujemo <u>standardni pogrešek ocene,</u> zaznamujemo pa ga s SE.

$$SE(\bar{x}) = \sqrt{Var\bar{x}} = \frac{\sigma x}{\sqrt{n}} \qquad (14.9)$$



14.3 Porazdelitev aritmetičnih sredin vseh vzorcev za n = 25 in n = 100 enot

c) V praksi običajno proučujemo končne populacije z vzorci brez ponavljanja. Zanje pa veljajo za aritmetične sredine vseh vzorcev nekoliko drugačne zakonitosti.

Matematično upanje za aritmetične sredine vzorcev brez ponavljanja je še vedno enako aritmetični sredini v osnovni populaciji:

$$M(\bar{x}) = E(\bar{x}) = M_x \qquad (14.10)$$

varianca za artimetične sredine vzorcev brez ponavljanja pa je enaka

$$Var(\bar{x}) = \frac{S_{\bar{x}}^2}{n} \cdot \frac{N-n}{N} = \frac{S_x^2}{n} \cdot (1-f) \qquad (14.11)$$

Pri tem je S_x^2 srednji kvadratični odklon za znak x, ki je izračunan po obrazcu

$$S_x^2 = \frac{\sum_{i=1}^{N} (x_i - M_x)^2}{N - 1} \qquad (14.12)$$

Razen neznatne spremembe, da je σ_x^2 zamenjan z nebistveno različnim izrazom S_x^2, je v obrazcu dodan še korekturni faktor $\frac{N-n}{N} = 1-f$, ki izraža, koliki del osnovne populacije ni vključen v vzorec.

d) Aritmetične sredine $\bar{x}$ iz vzorcev se porazdeljujejo v normalni porazdelitvi, če se znak x_i porazdeljuje v osnovni populaciji v normalni porazdelitvi.

Če se pa x_i v osnovni populaciji ne porazdeljuje normalno, se aritmetične sredine $\bar{x}$ kljub temu porazdeljujejo v porazdelitvah, ki so normalni tem bolj podobne, čim večji je vzorec. Ta približek je dokaj dober že pri razmeroma majhnih vzorcih, tako da praktično vzamemo, da se aritmetične sredine iz velikih vzorcev porazdeljujejo v normalni porazdelitvi, ne glede na to, kakšna je porazdelitev znaka v osnovni populaciji.

Iz zakonitosti od a) do d) povzemamo, da poznamo porazdelitev poprečij $\bar{x}$ v populaciji vseh možnih vzorcev, če poznamo parametre M, σ^2 in S^2 v osnovni populaciji. Tako dobimo posredno porazdelitev poprečij $\bar{x}$, ki je za teorijo in vzorčenja osnovnega pomena.

14.20 Točkovna ocena. Odklon zaupanja. Intervalna ocena. Razmik zaupanja. Glede na zakonitosti v populaciji sredin vzorcev in glede na lastnosti normalne porazdelitve v zvezi s stopnjo tveganja na splošno postavimo trditev, da je aritmetična sredina za enostaven slučajnostni vzorec $\bar{x}$ s tveganjem $\alpha = 0,05$ v razmiku

$$M_x - 1,96\ SE(\bar{x}) < \bar{x} < M_x + 1,96\ SE(\bar{x}) \qquad (14.13)$$

Aritmetične sredine vzorcev se tem bolj goste okrog prave aritmetične sredine, čim manjši je standardni pogrešek sredine. Ta pa je tem manjši, čim večji je vzorec.

Če iz populacije, za katero je $M_x = 30$ in $\sigma = 3$ izberemo enostaven slučajnostni vzorec z $n = 400$ enotami, leži poprečje, izračunano iz vzorca, s tveganjem $\alpha = 0,05$, v razmiku

$$30 - 1,96 \frac{3}{\sqrt{400}} < \bar{x} < 30 + 1,96 \frac{3}{\sqrt{400}}$$

$$2o,71 < \bar{x} < 30,29$$

Poprečje iz vzorca se torej s tveganjem $\alpha = 0,05$ odklanja od pravega poprečja za največ 0,29. Zato moremo vzeti poprečje iz vzorca $\bar{x}$ za oceno pravega poprečja M_x. To oceno $\bar{x}$ imenujemo točkovno oceno, ker je dana z eno samo vrednostjo.

$$\bar{x}_{sl} = \bar{x} = \frac{1}{n} Sx_i = x/n \qquad (14.14)$$

Količine $D(\bar{x}) = 1,96\ SE(\bar{x})$ ali v splošnem

$$D_{\alpha/2}(\bar{x}) = Z_{\alpha/2} \cdot SE(\bar{x}) \qquad (14.15)$$

imenujemo odklon zaupanja. Ta pove, za koliko se z določenim tveganjem točkovna ocena za aritmetično sredino največ odklanja od prave sredine za populacijo. V našem primeru je s tveganjem $\alpha = 0,05$ odklon zaupanja $D_{.025}(\bar{x}) = 0,29$. S tveganjem $\alpha = 0,05$ ocena z vzorcem ni različna od prave vrednosti za več kot $D_{.025}(\bar{x}) = 0,29$.

Če s tveganjem α ocena iz vzorca $\bar{x}$ ni od prave vrednosti različna za več kot $D_{\alpha/2}(\bar{x})$, tudi prava vrednost poprečja M_x ni od ocene $\bar{x}$ s tveganjem α različna za več kot $D(\bar{x})$. Iz tega sledi, da je slučajnostno izbrani vzorec s tveganjem $\alpha = 0,05$ tak, da je prava sredina v razmiku

$$\bar{x} - 1,96\ SE(\bar{x}) < M_x < \bar{x} + 1,96\ SE(\bar{x}) \qquad (14.16)$$

Ta razmik imenujemo razmik zaupanja, obe meji pa spodnjo in zgornjo mejo zaupanja ocene. To oceno prave vrednosti poprečja imenujemo intervalno oceno, ker je dana z razmikom zaupanja, v katerem je z določenim tveganjem pravo poprečje.

14.21 Ocenjevanje strukturnega deleža z velikimi vzorci. Razen aritmetične sredine je strukturni delež parameter, ki ga zelo pogosto ocenjujemo z vzorčenjem. Zgled za uporabo vzorčenja pri ocenjevanju strukturnih deležev so najrazličnejše ankete s področja javnega mnenja, analize trga ipd. S temi anketami skušamo oceniti, kolik odstotek prebivalstva ima določeno mnenje o kakem problemu, kolik delež prebivalstva je naklonjen ali nenaklonjen določenemu ukrepu, kolik delež prebivalstva je pripravljen, da nabavi določen izdelek, ki ga reklamiramo in podobno. Slično je ocena strukturnega deleža dobrodošel pokazovalec o kakovosti skupine izdelkov. Z vzorčno kontrolo moremo namreč oceniti delež izmeta, delež prvovrstnih izdelkov ipd.

Strukturni delež ocenjujemo podobno kot aritmetično sredino. Strukturni delež $p\,\%$, izračunan iz vzorca z n enotami, je nepristranska ocena za strukturni delež v populaciji

$$p\% = 100 \cdot \frac{n_a}{n} \qquad (14.17)$$

Pri tem je $p\%$ ocena za strukturni delež, n število enot v vzorcu, n_a pa število enot v vzorcu, ki ima proučevano značilnost, npr. število anketiranih oseb, ki so pripravljene kupiti določen izdelek. Podobno kot za aritmetično sredino je $p\%$ točkovna ocena za strukturni delež.

Če želimo dobiti razmično oceno, moramo najprej poiskati standardni pogrešek za oceno strukturnega deleža. Pravo vrednost za standardni pogrešek za strukturni delež dobimo po obrazcu

$$SE(p\%) = \sqrt{\frac{P\%\ (100-P\%)}{n}} \qquad (14.18)$$

Pri tem je P% prava vrednost za ocenjevani strukturni delež. Ker pa tega ne poznamo, ga moremo za velike vzorce nadomestiti z ocenjenim p%. Tako dobimo obrazec za oceno za standardni pogrešek za strukturni delež se(p%)

$$se(p\%) = \sqrt{\frac{p\%\ (100-p\%)}{n}} \qquad (14.19)$$

tako, da zamenjamo pravo vrednost P% z oceno p%. Razmik zaupanja za oceno strukturnega deleža dobimo po neenačbi

$$P_s\% = p\% - 1{,}96.se(p\%) < P\% < P_z\% = p\% + 1{,}96.se(\%) \qquad (14.20)$$

Pri tem je $P_s\%$ spodnja meja zaupanja, $P_z\%$ pa zgornja meja zaupanja za oceno strukturnega deleža s tveganjem $\alpha = 0{,}05$.

14.22 Vzemimo za zgled, da smo v vzorčno anketo pri analizi trga vključili n = 2500 gospodinjstev. Dobili smo naslednje rezultate: n_a = 500 gospodinjstev je pripravljenih, da brezpogojno nabavi mali gospodinjski stroj, če bi imel reklamirane priključke, n_b = 1000 gospodinjstev se ne more odločiti, n_c = 800 gospodinjstev je izjavilo, da niso pripravljena nabaviti reklamirane kombinacije, ker že imajo ustrezen strojček, n_d = = 200 gospodinjstev pa nima namena kupiti strojčka, čeprav ga še nima.

Izračunajmo najprej oceno za delež gospodinjstev, ki so pripravljena nabaviti reklamiran stroj.

Točkovna ocena za strukturni delež je po obrazcu

$$p_a\% = 100 \cdot \frac{500}{2500} = 20\ \%$$

standardni pogrešek se(p%) je po obrazcu 14.17

$$se(p\%) = \sqrt{\frac{20.(100-20)}{2500}} = 0{,}80\ \%$$

verjeten odklon je d% = 1,96 . 0,80 = 1,57 %, razmična ocena za strukturni delež pa po obrazcu 14.20

$$P_s\% = 20\ \% - 1{,}6\ \% < P\% < P_z\% = 20\ \% + 1{,}6\ \%.$$

Prava vrednost za ocenjevani strukturni delež leži s tveganjem 0,05 v razmiku

$$P_s\% = 18{,}4\ \% < P\% < P_z\% = 21{,}6\ \%$$

Ocene za strukturne deleže za druge odgovore so naslednje:

	p%	se(%)	d(p%)	$P_s\%$	$P_z\%$
a	20 %	·80 %	1,57 %	18,4 %	21,6 %
b	40 %	·98 %	1,92 %	38,1 %	41,9 %
c	32 %	·93 %	1,83 %	30,2 %	33,8 %
d	8 %	·54 %	1,06 %	6,9 %	9,1 %

Kot moremo presoditi, so rezultati dokaj zanesljivi.

Tehnika enostavnega slučajnega izbora

14.23 Pri praktični izvedbi enostavnega slučajnega vzorca, kakršne smo obravnavali dosedaj, 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.

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.

14.24 Loterijski način izbora enot. Pri loterijskem načinu izbora enot v enostavni vzorec izbiramo enote vzorca z žaro. Kot smo 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 zapo-

redno 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.

Prikladneje je, da v žaro vključimo samo deset enakih elementov, npr. kock, kroglic in podobno; nanje napišemo številke od 0 do 9. Če ima populacija npr. 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 npr. 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.

14.25 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 npr. takole: V žari imamo deset kroglic, ki so oštevilčene s številkami od 0 do 9. Iz nje 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. 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 slučajnih številk

5354 9142 0847 5393 5413 3505 7153 5334 9703 3221

enostavno sestavimo tele štirimestne številke, ki so zaporedne številke izbranih enot v okviru vzorčenja: 5354, 9142, 0847, 5393, 5413, 3505, 7153, 5334, 3221; če jih beremo od leve na desno.

Če beremo številke v tablici slučajnih številk od desne proti levi, pa dobimo tele zaporedne številke: 1223, 3079, 4335, 6517, 5053, 3145, 3935, 7480, 2419, 4535.

Če vzamemo, da štirimestna števila drse prek vrste slučajnih številk 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 številk zelo povečamo.

PETNAJSTO POGLAVJE

P R E I Z K U Š A N J E D O M N E V

15.1 Preizkušanje domnev. Razen s popolnim opazovanjem ali z ocenjevanjem z vzorci dobimo informacije o populacijah s preizkušanjem domnev o parametrih. Domnevo, da je v skupini izdelkov delež nekakovostnih izdelkov manjši kot P = 5 % preizkusimo s popolnim pregledom vseh izdelkov, ki domnevo potrdi ali zavrne. Izkaže pa se, da moremo do sklepa o določeni domnevi priti z vzorčenjem in to po pravilu z manjšimi vzorci kot jih uporabljamo pri ocenjevanju parametrov. Kot pri ocenjevanju tudi pri preizkušanju domnev sklepi, ki jih napravimo, veljajo z večjim ali manjšim tveganjem. Uporabnost preizkušanja domnev je tolikšna, da velik del statistične analize bazira na preizkušanju domnev v najrazličnejših inačicah.

15.2 Osnovna in ničelna domneva. Domneve ima smisel preizkušati le, če je možnih več, najmanj dvoje različnih stanj. Če je možno eno samo stanje pojava, preizkušanje te domneve nima smisla, ker vnaprej vemo, kakšno je stanje.

Zato domnevi, ki jo preizkušamo - imenujemo jo osnovno domnevo (H_1) - praviloma priredimo alternativno - ničelno domnevo (H_0). Zavrnitev ene naj bi pomenila sprejem druge.

Osnovni domnevi $H_1 : P = P_1 = 5\ \%$, da je delež nekakovostnih izdelkov v pošiljki enak 5 %, postavimo alternativno ničelno domnevo, $H_0 : P = P_0 = 3\ \%$, da je delež neustreznih izdelkov enak 3 %. Osnovni domnevi, $H_1 : \sigma > \sigma_0 = 2$, da je standardni odklon večji kot $\sigma = 2$, postavimo nasproti ničelno domnevo, $H_0 : \sigma < \sigma_0 = 2$, da je standardni odklon manjši kot 2.

15.3 Enostavne in sestavljene domneve. V zgledu o izmetu vsebuje ničelna in osnovna domneva po eno samo - točno določeno vrednost. Take domneve imenujemo enostavne za razliko od domnev, ki nastopajo v zgledu o standardnem odklonu, v katerem so domneve sestavljene, ker vsebujejo več vrednosti.

15.4 Preizkušnje enostavnih domnev. Prvine preizkušanja domnev ponazorimo na zgledu, v katerem sta H_0 in H_1 enostavni domnevi o aritmetičnih sredinah.

Preizkušamo domnevo H_1, da je pošiljka izdelkov izdelana iz surovine S_1, pri kateri je poprečna trdnost $H_1 : M = M_1 = 21{,}5$ proti ničelni domnevi, da so izdelki iz surovine S_0, pri

kateri je povprečna trdnost $H_0: M = M_0 = 20$. Predpostavljajmo, da se x porazdeljuje normalno s $\sigma = 2$.

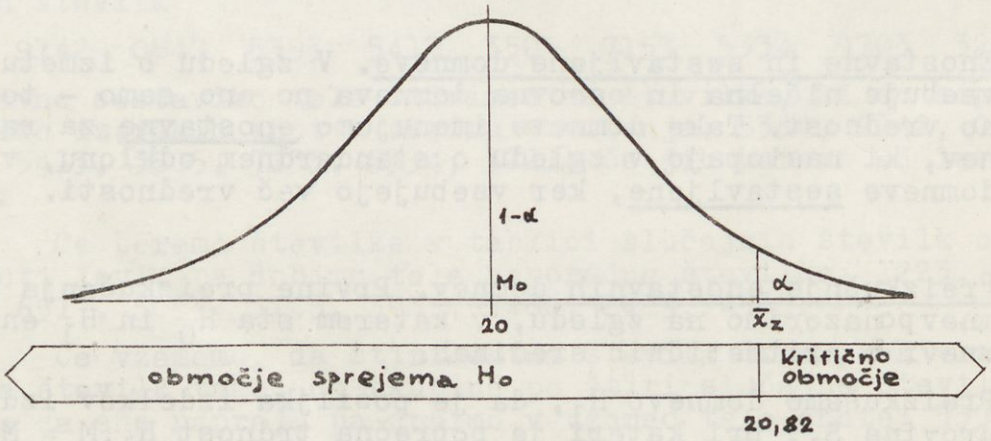
Dani domnevi preizkušajmo s preizkusnim vzorcem z n = 16 enotami. Iz teorije vzorčenja vemo, da velja za sredine iz vzorca

$$Pr\ (\bar{x} < \bar{x}_z = M_0 + z_p \frac{\sigma}{\sqrt{n}}) = 1 - P = 1 - \alpha \qquad (15.1)$$

Če vzamemo, da velja ničelna domneva $M = M_0 = 20$, in $\alpha = P = 0{,}05$, kateremu ustreza $z_{0,05} = 1{,}645$, velja po zgornji verjetnostni neenačbi

$$Pr(\bar{x} < \bar{x}_z = 20 + 1{,}645 \cdot \frac{2}{\sqrt{16}} = 20{,}82) = 0{,}95$$

V primeru, da velja ničelna domneva $H_0 : M = M_0 = 20$, je po zgornjem sklepu velika verjetnost (0,95), da je poprečje iz preizkusnega vzorca n = 16 enot pod $\bar{x}_z$. Zato sklepamo, da smo upravičeni, da ničelno domnevo sprejmemo, če je $\bar{x} < \bar{x}_z$ in zavrnemo, če je $\bar{x} > \bar{x}_z$. Razmik, v katerem sprejmemo ničelno domnevo, imenujemo območje sprejema ničelne domneve, razmik, v katerem ničelno domnevo zavrnemo, pa kritično območje. Vrednost $\bar{x}_z$ ali vrednosti, ki razmejujejo območje sprejema od kritičnega območja, imenujemo kritične meje. V našem zgledu je kritična meja za poprečje iz vzorca $\bar{x}_z = 20{,}82$.



Situacija v primeru, da velja $H_0 = M = M_0 = 20$

15.5 Napaka I. vrste in napaka II. vrste. Iz zgleda spoznamo, da je možno, da je poprečje $\bar{x}$ iz preizkusnega vzorca v kritičnem območju, čeprav velja H_0. Po nakazanem pravilu bi v tem primeru bili upravičeni, da ničelno domnevo zavrnemo, čeprav velja. Sklep je napačen. Napako, ki jo napravimo, da ničelno domnevo zavrnemo, čeprav velja, imenujemo napako I. vrste. Verjetnost, da napravimo napako I. vrste, je α. Verjetnost za napako I. vrste praviloma vnaprej predpišemo: običajne stopnje za napako I. vrste so: 0,05, 0,01 in 0,001.

Proučimo še situacijo, da velja osnovna domneva $H_1 = M = M_1 = 21{,}5$. Če ohranimo kritično mejo $\bar{x}_z = 20{,}82$, je verjetnost, da je poprečje iz preizkusnega vzorca $\bar{x}$ v kritičnem območju, v tem primeru enaka

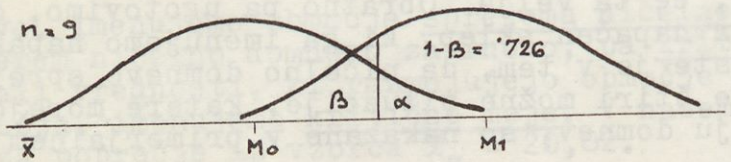
$$Pr(\bar{x} > \bar{x}_z = M_1 - z_{1-\beta} \frac{\sigma}{\sqrt{n}}) = 1 - \beta \qquad (15.2)$$

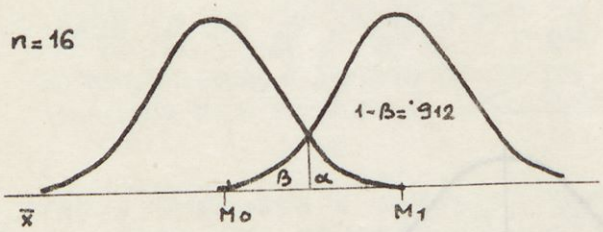
Če vstavimo znane izraze v to verjetnostno neenačbo pri pogoju normalnosti, ugotovimo, da je $1-\beta = 0{,}912$. Verjetnost, da bo v primeru, da velja osnovna domneva, poprečje v kritičnem območju, je zelo velika. Verjetnost $1 - \beta$, ki pokaže, s kakšno verjetnostjo zavrnemo ničelno domnevo (in s tem sprejmemo osnovo), če velja osnovna domneva, imenujemo moč preizkusa. Čim večja je moč preizkusa, tembolj verjetno je, da bomo zavrnili ničelno in sprejeli osnovno domnevo, če ta velja. Obratno pa ugotovimo, da je z verjetnostjo β možen napačen sklep, ki ga imenujemo napako II. vrste. Napaka II. vrste je v tem, da ničelno domnevo sprejmemo, če ta ne velja. Vse štiri možne situacije, katere morejo nastopiti pri preizkušanju domnev, so nakazane v primerjalnem pregledu.

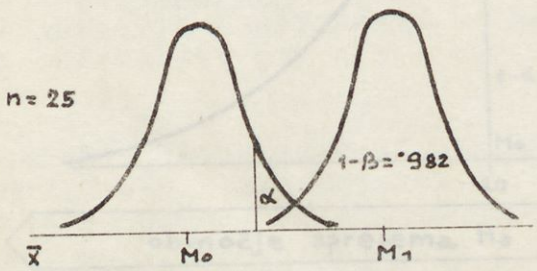
	Območje sprejema H_0	Kritično območje
Rezultat preizkusnega vzorca	$\bar{x} < \bar{x}_z$	$\bar{x} > \bar{x}_z$
Stanje \ sklep	sprejmemo H_0	sprejmemo H_1
$M = M_0$	sklep pravilen $1 - \alpha$	sklep napačen α
$M = M_1$	sklep napačen β	sklep pravilen $1 - \beta$

Primerjalni pregled možnih situacij pri preizkušanju enostavnih domnev.

15.6 Odvisnost moči preizkusa od velikosti vzorca. Pri danih domnevah je preizkus tem močnejši, čim večji je vzorec. Kako je v našem zgledu moč preizkusa odvisna od velikosti vzorca, je ponazorjeno v sliki in tabeli za vzorce $n = 9$, $n = 16$ in $n = 25$.







n	$1-\beta$	β
9	·726	·274
16	·912	·098
25	·982	·018

15.7 Značilnost in neznačilnost razlik in vzorčnih izrazov. V prikazanem zgledu preizkusa domnev o izvoru surovin smo za $\bar{x}$ določili:

a) kritično območje,

b) območje sprejema ničelne domneve.

Če pada $\bar{x}$ v kritično območje, sklepamo, da z razmeroma majhnim tveganjem zavrnemo ničelno domnevo H_0 in sprejmemo osnovno domnevo. Ta sklep strnemo v ugotovitev, da je preizkusni $\bar{x}$ značilno različen od M_0, ki ga postavljamo z ničelno domnevo. Značilne razlike pomenijo, da je vzorec s tveganjem, ki ni večji kakor α, odkril, da ne velja ničelna temveč osnovna domneva. Obratno pa sklepamo, da so razlike neznačilne, če $\bar{x}$ pade v območje sprejema ničelne domneve. Če so razlike neznačilne, po zgornjem pravilu sprejmemo ničelno domnevo. Ker pa ne poznamo verjetnosti β za napako II. vrste, vemo pa, da more biti tudi razmeroma velika, običajno sprejmemo ničelno domnevo, če so razlike neznačilne z razlago, da jo sprejmemo, ker ni razloga, da bi na osnovi vzorčnega rezultata dvomili v njo. Če pri preverjanju dela stroja v proizvodnem procesu s preizkusnim vzorcem ugotovimo, da je preizkusni $\bar{x}$ neznačilno različen od predpisanega M_0 (ki se sklada z ničelno domnevo), vzamemo, da teče proizvodni proces v redu, ker na osnovi preizkusnega vzorca nimamo osnove, da dvomimo v to.

Gornji sklep je ena izmed možnih odločitev pri neznačilnosti razlik pri preizkušanju domnev. Ker je kritična meja in s tem tudi značilnost vzorčnih izrazov bistveno odvisna od velikosti preizkusnega vzorca, sklepamo lahko tudi tako, da more neznačilnost izvirati iz tega, da je vzorec premajhen in je neznačilnost razlik razlog, da preizkusni vzorec povečamo in skušamo odkriti značilnost razlik z večjim vzorcem.

15.8 Stopnje značilnosti. Ker je kritično območje odvisno od tveganja α, se more zgoditi, da je $\bar{x}$ v kritičnem območju, če je α večji, da pa je v območju sprejema ničelne domneve za manjše tveganje α. Zato govorimo, da je razlika značilna na ravni $\alpha = 0{,}05$, če so razlike značilne stopnji tveganja $\alpha = 0{,}05$, niso pa značilne na stopnji $\alpha = 0{,}01$. Razlike smatramo za značilne na ravni $\alpha = 0{,}01$, če so značilne s tveganjem $\alpha = 0{,}01$, niso pa značilne na ravni $\alpha = 0{,}001$. In končno: razlike smatramo za značilne na ravni $\alpha = 0{,}001$, če se izkažejo značilne na ravni $\alpha = 0{,}001$. Te tri stopnje so seveda vzete konvencionalno. Ker je stopnja tveganja na ravni $\alpha = 0{,}10$ razmeroma veliko, sklepamo, da obstaja sum na značilnost razlik, če je vzorčna vrednost v kritičnem območju z $\alpha = 0{,}10$ in v območju sprejema za $\alpha = 0{,}05$.

15.9 Vzorčni izrazi. Poprečje in vse druge parametre pa enostavneje kot direktno preizkušamo prek vzorčnih izrazov, ki vsebujejo rezultate preizkusnega vzorca in domevne vrednosti.

Če je populacija v skladu z domnevnimi vrednostmi, vzorčni izrazi slede določenim zakonitostim.

Tako se izraz

$$\frac{\bar{x} - M_0}{\sigma} \sqrt{n} = z \qquad (15.3)$$

porazdeljuje standardizirano normalno, če velja ničelna domneva H_0: $M_0 = M$. Preizkušanje domnev o aritmetični sredini je z nakazanim vzorčnim izrazom enostavnejše, ker dobimo kritične vrednosti za z direktno iz tabel o standardizirani normalni porazdelitvi. Tako štejemo, da so razlike na določeni stopnji značilne ali neznačilne, če je absolutna vrednost za z v mejah

neznačilno	sum	značilno		
	10 %	5 %	1 %	1°/oo
	1,645	1,96	2,56	3,29

Če ničelna domneva ne velja, moremo gornji vzorčni izraz razdeliti v dva dela

$$\frac{\bar{x} - M_0}{\sigma} \sqrt{n} = \frac{\bar{x} - M}{\sigma} \sqrt{n} + \frac{M - M_0}{\sigma} \sqrt{n}$$

Prvi del $\frac{x-M}{\sigma} \sqrt{n}$ sledi zakonitost za standardizirano normalno porazdelitev, ker je v njej prava aritmetična sredina. Drugi del $\frac{M-M_0}{\sigma} \sqrt{n}$ pa je konstanta, ki zavisi od: razlike med pravo in domnevno aritmetično sredino $M - M_0$, merjeno v standardnih odklonih in od velikosti vzorca n. Razlika z večjo verjetnostjo odstopa od zakonitosti z, čim večja je razlika $M - M_0$, čim večji je vzorec n in čim manjši je standardni odklon σ.

Narodna in univerzitetna knjižnica
LJUBLJANA
*

24 -11- 1976

NARODNA IN UNIVERZITETNA
KNJIŽNICA
00000059352