UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MARIJAN BLEJEC

UVOD V STATISTIKO

ponatis enajste izdaje

LJUBLJANA 1996

И

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MARIJAN BLEJEC

UVOD V STATISTIKO

ponatis enajste izdaje

LJUBLJANA 1996

prof. dr. Marijan Blejec
UVOD V STATISTIKO
Založila in izdala Ekonomska fakulteta v Ljubljani
Tiskala Univerzitetna tiskarna v Ljubljani
Naklada 1000 izvodov

PREDGOVOR K PONATISU DRUGE IZDAJE

Druga izdaja učbenika Uvod v statistiko je prirejena po učbeniku M. Blejec: Statistične metode za ekonomiste, druga predelana in razširjena izdaja. Ekonomska fakulteta, Ljubljana 1973.

Učbenik je razdeljen v 12 poglavij, od katerih vsako obravnava bodisi posamezno fazo statističnega proučevanja ali statistične metode analize.

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

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

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

I = Indeks, Zvezni zavod za statistiko

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

SB-LRS	=	Statistični Bilten LRS
MSP	=	Mesečni statistični pregled; Zavod za statistiko LRS.

Ljubljana, februar 1978

M. Blejec

VSEBINA

39

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

U V O D

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

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

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

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

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

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

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

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

PODROČJA, V KATERIH UPORABLJAMO STATISTIKO

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

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

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

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

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

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

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

Statistiko z velikim pridom uporabljamo tudi v meteorologiji.

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

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

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

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

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

SOCIALNO-EKONOMSKA STATISTIKA

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

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

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

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

Tako imamo posebej:

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

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

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

Analogno imamo nadalje gozdarsko statistiko, statistiko obrti, 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

PROUČEVANJE MNOŽIČNIH POJAVOV

MNOŽIČNI POJAVI

2.1 V naravi in družbi pojavi ne nastopajo posamič, marveč v velikem številu, ne izolirano, marveč povezani med seboj. Če opazujemo te pojave individualno in izolirano, se zdi njihovo pojavljanje brez reda. V resnici pa veljajo zanje določene zakonitosti. Te odkrijemo šele, če opazujemo ne samo posamezen pojav, temveč skupnosti pojavov. Omenili smo že, da statistika opazuje in analizira m n o ž i č n e p o j a v e . 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 delavstva, vsi izdelki niso enako uporabni itd.

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

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

S t a t i s t i č n e e n o t e

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

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

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

Statistične enote so bodisi e n o s t a v n e e n o t e ali pa a g r e g a t i - s k u p i n i c e enostavnih enot. Tako štejemo posameznega človeka za enostavno enoto.

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

Statistični znaki

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

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

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

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

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

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

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

2.6 Vrednosti časovnih in numeričnih znakov se dajo urediti po velikosti po nedvoumnem vrstnem redu. Te lastnosti krajevni in atributivni znaki v splošnem nimajo.

Možnost, da lahko uredimo vrednosti znakov po velikosti, je velika prednost časovnih in numeričnih znakov; omogoča namreč poglobljeno analizo, ki je za geografske in atributivne znake ne moremo izvesti. Možnost analize atributivnih znakov v primerjavi z numeričnimi je torej okrnjena. Medtem ko moremo vse metode za analizo atributivnih znakov uporabiti tudi za numerične znake, obratno niso vse metode za numerične znake uporabne za atributivne.

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

N o m i n a l n o s t je lastnost znakov, da moremo posamezne vrednosti znakov med seboj razlikovati. Nominalen značaj znakov je torej izražen z $y_1 \neq y_2$. Samo nominalen značaj imajo stvarno atributivni znaki, po katerih moremo vrednosti znakov samo razlikovati.

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

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

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

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

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

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

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

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

2.8 Numerične znake delimo po drugem merilu na: a) ekstenzivne in b) intenzivne.

E k s t e n z i v n i znaki nakazujejo količino, i n t e n z i v n i pa kakovost. Tako so ekstenzivni znaki npr. vrednost proizvodnje podjetja v določenem razdobju, število delavcev po podjetjih, količina določenega blaga na trgu itd. Intenzivni znaki pa so na primer cena izdelka, produktivnost dela, merjena z vrednostjo proizvoda na enoto časa, hektarski pridelek itd.

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

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

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

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

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

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

Statistične populacije

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

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

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

Gornjo populacijo smo časovno opredelili tako, da smo navedli: enote populacije so samo tiste ženske, ki zadoščajo vsem opredeljujočim pogojem konec leta 1972. Krajevno je zgornja populacija opredeljena s tem, da navedemo: enote populacije so samo ženske, ki zadoščajo drugim opredeljujočim pogojem in so v Sloveniji. Stvarno je populacija opredeljena s tremi znaki: s stanom, s spolom in z zaposlenostjo.

V splošnem k r a j e v n a o p r e d e l i t e v 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.

Č a s o v n a o p r e d e l i t e v je bolj problematična in je odvisna od narave populacije, ki jo proučujemo.

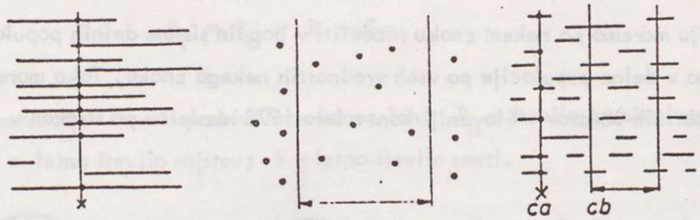
Zgornjo populacijo žensk smo opredelili s trenutkom, ker je populacija realnih enot. Zato populacije realnih enot včasih tudi imenujemo m o m e n t n e a l i t r e - n u t n e populacije. Momentnih populacij imamo veliko. Vse časovno enolično opredelimo z momentom, v katerem morajo posamezne enote populacije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

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

Časovna opredelitev populacij dogodkov z momentom ni smiselna. V določenem momentu se more zgoditi in se zgodi samo omejeno in slučajno število dogodkov. Tako nima smisla iskati populacijo rojstev, nesreč, smrti, proizvodnje itd., ob določenem momentu. Populacijo dogodkov opredelimo z razdobjem in vanjo štejemo vse dogodke, ki so se zgodili v določenem razdobju in zadoščajo vsem drugim pogojem. Te populacije imenujemo i n t e r v a l n e a l i r a z m i č n e p o p u l a c i j e , ker so opredeljene s časovnimi razmiki. Intervalne populacije so npr. proizvodnja, rojstva, nesreče, prodaja itd.

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

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



a) realne enote opredelitev: momentna — b) dogodki opredelitev: interval — c) dogajanje opredelitev ca: moment cb: interval

Slika 2.1 Momentna in intervalna opredelitev

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

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

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

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

Slika 2.2 Delne populacije

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

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

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

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

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

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

Statistični parametri

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

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

Te parametre moremo shematično nakazati s simboli:

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

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

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

ZNAČILNOSTI PRI PROUČEVANJU MNOŽIČNIH POJAVOV

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

V fiziki, kemiji in v agronomiji proučujemo pojave večinoma s poskusi. Poskuse v f i z i k i in k e m i j i delamo laboratorijsko v natančno določenih pogojih, da čimbolj odstranimo vse vplive, ki bi mogli poskus motiti. Če bi mogli odstraniti vse dodatne vplive in rezultate meriti z idealno točnimi merilnimi instrumenti, bi zadoščal en sam poskus, ker bi dobili pri vsakem poskusu ob enakih pogojih isti rezultat. Vendar v praktičnih primerih rezultati pri ponovitvah niso povsem enaki. Pokažejo se majhne razlike, ker je nemogoče poskus ponoviti natančno enako. Zaradi dodatnih vplivov, ki jih z obstoječimi poskusnimi sredstvi ne moremo odpraviti, so rezultati med seboj različni. Te razlike imenujemo slučajne razlike, faktorje, ki jih povzroče, pa slučajne faktorje. Vendar se te razlike v velikem številu poskusov pojavljajo v nekih zakonitostih. Zaradi slučajnih faktorjev se rezultati posameznih poskusov odklanjajo od prave vrednosti navzgor in navzdol. Po zakonu o velikih številih pa se slučajnostni odkloni v vsoti rezultatov iz velikega števila poskusov izravnavajo. Zato je poprečen rezultat bliže rezultatu, ki bi ga dobili, če ne bi bilo subjektivnih momentov eksperimentatorja, nemožnosti popolne določitve poskusnih pogojev in napak v meritvah. Da dobimo objektivnejše poprečje, je dobro, če vsak poskus izvedejo različni eksperimentatorji z različnimi merilnimi instrumenti. Tako merilne priprave kot eksperimentator lahko povzročijo sistematične napake. Napaka v instrumentu je lahko take narave, da sistematično daje vedno premajhne rezultate. Isto more biti tudi z eksperimentatorjem. Če ponavljamo poskuse z istimi merilnimi instrumenti ali eksperimentatorji, se sistematične napake ne izravnajo in poprečen rezultat se ne približa dejanski vrednosti - je pristranski.

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

Kakor pri fizikalnih poskusih tako smo tudi pri poskusih v agronomiji ugotovili isto važno dejstvo. Zakonitost, ki jo preizkušamo s poskusom, v posamičnem primeru motijo slučajni vplivi. V veliki množici istovrstnih poskusov pa se učinek slučajnih vplivov v poprečju izravna in se zaradi delovanja zakona o velikih številih pokaže tipičnost pojava. Ta postopek je povsem specifičen. Točnejših rezultatov pri poskusih ne dosegamo z izboljšavo laboratorijske tehnike, temveč s ponavljanjem poskusov ob danih pogojih iz rezultatov skupine poskusov dobimo natančnejši rezultat. Taka pot je v dosti primerih ustrezna, ker je običajno ceneje ponavljati poskus z danimi sredstvi, kakor pa z dragimi aparaturami neposredno odstranjevati slučajne vplive. V primeru, da dosežemo meje možnosti pri opredelitvi pogojev, so pa slučajni vplivi kljub temu močni, kakor je to npr. pri poskusih v agronomiji, edino s ponovnimi poskusi oziroma poprečji dosežemo natančnejše rezultate.

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

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

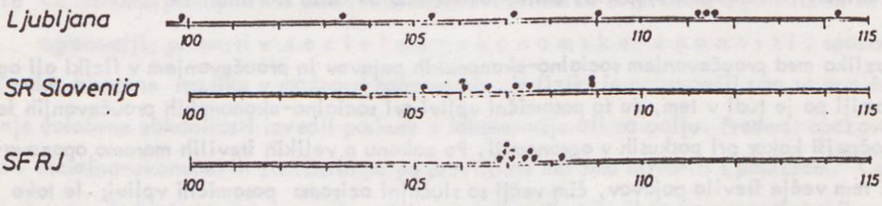
2.18 Če skušamo postopke, ki smo jih navedli kot primerne za proučevanje v fiziki in agronomiji, prenesti v s o c i a l n o - e k o n o m s k e z n a n o s t i, spoznamo neke bistvene razlike v možnosti opazovanja. V fiziki in agronomiji smo za proučevanje določene zakonitosti izvedli poskuse v laboratoriju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

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

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

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

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



Slika 2.3 Spolni indeks novorojenčkov

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

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

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

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

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

ETAPE STATISTIČNEGA PROUČEVANJA

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

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

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

b) izdelava splošnega načrta statističnega proučevanja,

c) opazovanje v ožjem smislu,

d) urejevanje in osnovna obdelava,

e) analitična obdelava in analiza podatkov.

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

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

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

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

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

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

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

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

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

2.25 Analitična obdelava. V osnovni obdelavi dobimo podatke urejene, preštete in seštete. Vendar ti podatki še ne zadostujejo za analizo pojava. Množične pojave analiziramo z obdelavo osnovnih absolutnih podatkov s posebnimi metodami, ki jih bomo v nadaljevanju podrobneje obravnavali. Iz osnovnih podatkov namreč dalje izračunavamo relativna števila, srednje vrednosti, mere variacije, mere korelacije, pokazovalce dinamike pojavov ipd. s katerimi prikažemo in analiziramo zakonitosti množičnih pojavov.

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

TRETJE POGLAVJE

STATISTIČNO OPAZOVANJE

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

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

OPREDELITEV PREDMETA OPAZOVANJA - STATISTIČNE POPULACIJE

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

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

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

K r a j e v n a   o p r e d e l i t e v je običajno brez težav, ker se opremo na obstoječe občinske, republiške ali državne meje.

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

VRSTE OPAZOVANJ

3.3 Glede na proučevano populacijo uporabljamo različne vrste statističnega opazovanja. Medtem ko populacije realnih enot, kot so prebivalstvo, stavbe, gospodarstva, živina itd., opazujemo s p o p i s i , populacije dogodkov, kot so smrti, rojstva, nesreče, itd., opazujemo s t e k o č o r e g i s t r a c i j o . Glede na potrebe, materialne in finančne možnosti pa opazujemo v s e e n o t e p o p u l a c i j e , če potrebujemo popolno in podrobno sliko o populaciji, ali samo d e l p o p u l a c i j e , če se zadovoljimo z ocenami.

3.4 Longitudinalno in transverzalno opazovanje

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

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

Če sestavimo matriko teh dogodkov, dobimo:

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

V numeričnem zgledu vidimo kakšno sliko moremo dobiti v praksi.

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

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

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

Opazovanje, v katerem proučujemo dinamiko populacij prek sprememb posameznih enot v določenem razdobju, imenujemo l o n g i t u d i n a l n a   o p a z o v a n j a. Prikazani zgledi so samo shematični primeri longitudinalnega opazovanja. V vmesnem razdobju more na posamezno enoto vplivati ne en sam dogodek temveč več dogodkov, ki spremene značilnosti enot. V takem primeru je seveda shema in opazovanje znatno bolj zamotano.

Enota	
1	AAAAAAAAABBBBBBBBBBB
2	AAAAAAAAAAAAAAAAAAAA
3	BBBBBBBBBAAAAAAAAAAA
4	AAAAAABBBBBBBBBBBBBB
5	BBBBAAAAAABBBBBBBBBB
6	AAAAAABBBBBBBBBBBBAA

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

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

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

P o p i s

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

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

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

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

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

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

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

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

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

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

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

Tekoča registracija – Statistična poročila

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

Tudi za tekočo registracijo je značilno popolno zajetje vseh enot populacije. Redkeje ko pri popisih je namen tekoče registracije zbiranje podatkov samo za statistično proučevanje pojava. Matična služba, register prebivalstva, operativne evidence v podjetjih, knjigovodski podatki itd. so samo posredno vir statističnih podatkov. Te podatke zbiramo s statističnimi poročili, ki podajajo splošno sliko za dogodke v določenem razdobju (dnevu, tednu, mesecu, letu). Statistična poročila so ena izmed običajnih oblik registracije socialno-ekonomskih pojavov. Tako imamo npr. statistična poročila iz statistike cen, gostinstva, turizma, gradbeništva, industrije, matične službe itd.

Delna opazovanja

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

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

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

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

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

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

3.12 Panel ankete. Proučevanje dinamike pojavov z vrsto neodvisnih vzorcev je problematično. Ker so v vsak vzorec izbrane druge enote: osebe, gospodinjstva, gospodarstva ipd., je vzorčni pogrešek razmeroma velik. Četudi se ne bi stanje v populaciji spremenilo, bi verjetno v dveh zaporednih vzorcih dobili zelo različne rezultate, kar bi bila posledica tega, da so v posameznih vzorcih druge enote. To hibo odpravimo tako, da vzorec istih enot uporabimo za niz opazovanj in opazujemo spremembe na posameznih enotah tega vzorca v sukcesivnih opazovanjih. Taka anketa, ki jo imenujemo panel anketa, ima za osnovo longitudinalno opazovanje in je odlično sredstvo za proučevanje dinamike pojavov. Zato panel ankete uporabljamo v primerih, kadar gre za proučevanje časovnih sprememb, ki so rezultat najrazličnejših vplivov. S panel anketami proučujemo spremembe v obnašanju gospodinjstev pod vplivom najrazličnejših dejavnikov, ki vplivajo na spremembe v standardu prebivalstva. S panel anketami moremo meriti reakcijo na oglaševanje, na cene, na konkurenco, na spreminjanje javnega mnenja ipd. Medtem ko je v primerih, ko gre za dolgoročnejše raziskave, en vzorec dalj časa osnova za panel anketo, more za določene raziskave, npr. vpliva na oglašanje rabiti panel anketa s samo dvema opazovanjema: pred in po reklamiranju. S panel anketami oziroma longitudinalnim opazovanjem izločimo predvsem razlike med enotami opazovanja.

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

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

VIRI PODATKOV

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

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

3.14 Podatki že izvedenih statističnih opazovanj niso vedno obdelani tako, da bi mogli rezultate neposredno uporabiti za analizo problema, ki ga proučujemo.

Preden se odločimo, da izvedemo samostojno opazovanje, proučimo še osnovno gradivo že izvedenih statističnih akcij. Dostikrat se izkaže, da sicer obdelani rezultati ne dajo odgovora na proučevan problem, moremo pa do njega priti s posebno obdelavo že zbranega gradiva po ustreznih vidikih in kombinacijah. Če nam to uspe, ni treba zbirati osnovnih podatkov. Tako izkoriščamo osnovno gradivo najrazličnejših popisov, kot so popisi prebivalstva, industrije, obrti itd. in prihranimo stroške zbiranja podatkov.

3.15 Za osnovo pri proučevanju uporabljamo tudi dosedanje zapise - evidence, ki jih zbira ena ali druga organizacija ali enota v nestatistične namene.

Ti zapisi - imenujemo jih tudi sekundaren vir statističnih podatkov - so osnova za podatke iz vseh gospodarskih statistik, statistike industrije, trgovine, gostinstva itd. Prav tako je sekundaren vir statističnih podatkov matična služba, ki zbira podatke iz naravnega gibanja prebivalstva in register prebivalstva, ki beleži mehansko gibanje prebivalstva.

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

3.16 Dostikrat pa ne razpolagamo za proučevan pojav niti s takimi zapisi. Takrat smo primorani, da podatke o enotah populacije zberemo drugače. Eden izmed takih načinov je posredno opazovanje po osebah, ki enoto opazovanja poznajo in morejo dati o njej zahtevane podatke. Posredno opazovanje je zelo razširjeno. Tako dajejo gospodarji kmetijskih gospodarstev podatke o svojih gospodarstvih, gospodinje o porabi v svojih gospodinjstvih, starešine gospodinjstev podatke o članih, obrtniki o svojih podjetjih itd.

V nekaterih primerih, posebno če gre za poskuse je vir osnovnih statističnih podatkov n e p o s r e d n o   o p a z o v a n j e. Neposredno opazujemo pojave, če sami izvedemo poskus oziroma merimo enote opazovanja.

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

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

NAČINI POSREDNEGA OPAZOVANJA

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

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

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

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

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

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

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

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

ZNAKI OPAZOVANJA

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

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

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

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

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

SREDSTVA OPAZOVANJA

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Obrazec označimo razen z naslovom še s kratko oznako - k r a t i c o . Tako ima prikazani obrazec popisa prebivalstva oznako "Obrazec PS-1", obrazec kmetijske ankete "Obrazec PA-4", obrazec za mesečno industrijsko poročilo pa "IND-1-1958". V glavi obrazca je tudi   n a z i v   o r g a n a , ki je obrazec izdal (npr. Zvezni zavod za statistiko), k d o j e o b r a z e c o d o b r i l, če ga izdaja organizacija, ki ni statistična, in n a z i v i n n a s l o v e n o t e, ki daje podatke.

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

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

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

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

Federativna ljudska republika Jugoslavija
ZVEZNI ZAVOD ZA STATISTIKO IN EVIDENCO

Obrazec PS-1

Popis prebivalstva 31. marca 1953

POPISNICA

Izpolni popisovalec	
Številka popisnega okoliša	
Zaporedna številka kontrolnika	
Okraj-mesto	
Občina	
Naselje (mesto, vas)	
Sestavni del naselja	
Ulica in hišna številka	

Izpolni zavod za statistiko in evidenco ↓

10

Ali je popisana oseba		
	trajno prisotna (*da, ne*)	
	začasno odsotna (*da, ne*)	
	začasno prisotna (*da, ne*)	

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

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

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

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

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

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

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

3. ROJSTNI DATUM dan ______ mesec ______ leto ______

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

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

20

5. DRUŽINSKI STAN [Pri osebah, rojenih po 31. marcu 1939, narediomo črtico (–); pri osebah, ki so se rodile pred tem datumom, vpišemo enega izmed odgovorov: *samski-samska, poročen-poročena, odovec-odova, razvezan-razvezana.*] . . . . . ______

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

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

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

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

30

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

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

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

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

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

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

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

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

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

b) samski, <u>poročen</u>, razvezan, vdovec

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

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

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

1	2	3	4	5	6	7	8	9	10	11	12
Prvo polletje	Drugo polletje	15 januarja 1959					15.1.1958	15 januarja 1959			
Koliko je bilo l. 1958 porabljena umetnih gnojil (kg.)		Žrebeta do 2 let	Mladi konji nad 2 leti	Kobile in breje žrebice	Žrebci in skopljenci	Skupno število konj		Delovni konji in kobile	Skupno število mezgov in mul	Skupno število oslov	

ZVEZNI ZAVOD ZA STATISTIKO

VPRAŠALNI LIST

za

KMETIJSKO ANKETO

15-I-1960

Obrazec PA-4

Okraj ______

Občina ______

Naselje ______

Hišna št ______

Številka popisnega okoliša v občini ______

Zap. štev. gospodarstva iz PA-3 ______

Priimek, očetovo ime in ime starešine gospodarstva

Vrsta gospodarstva (gospodinjstva) | Kateg. gosp. | Posebne ugotovitve

Skupna lastna površina ha ______ arov ______

Obdelovalna lastna površina ha ______ arov ______

Teleta do 1 leta	Bikci, junčki in junice nad 1—2 let	Mlado govedo nad 2 leti	Krave in breje junice	Plemenski biki	Voli	Skupno število govedi		Krave ki se uporabljajo za delo	Oteljeno telet leta 1958	Skupno število kupljenih govedi	V letu l. 1958 zaklano: Telet (do 1 leta)	Bikcev, juncev in junic nad 1 leto	Odraslih živali	Skupno število poginule govedi
15 januarja 1959 leta							15.1.1958							
13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

Slika 3.2 Prva stran vprašalnega lista za kmetijsko anketo 15.1.1958.

Zvezni zavod za statistiko, Beograd

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

IND-1-1958

Mesečno poročilo industrijskega podjetja

za mesec 1958

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

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

Ne izpolnjuje podjetje

Leto in mesec 8

Okraj

Občina

Panoga, stroka in vrsta dejavnosti

Sektor lastništva

Grupni znak

Izpolnjuje Zavod za statistiko OLO

Zaporedna številka iz kartoteke organiz. enot

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

Rok 6. v mesecu

V. p.: 324

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

Proizvodi po nomenklaturi IND-1-1958: Šifra	Naziv proizvoda	Merska enota	Dosežena izgotovljena proizvodnja: v mesecu poročanja	od začetka leta do konca meseca poročanja	Skupna zaloga konec meseca poročanja	Realizacija blaga od začetka leta do konca meseca poročanja	Pričakovana proizvodnja v sledečem mesecu
1	2	3	4	5	6	7	8

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

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

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

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

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

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

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

KRAJ OPAZOVANJA

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

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

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

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

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

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

ORGANI STATISTIČNEGA OPAZOVANJA

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

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

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

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

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

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

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

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

NAPAKE IN KONTROLA PRI STATISTIČNEM OPAZOVANJU

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

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

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

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

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

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

3.41 Kontrola napak. Da bi dobili čim boljše podatke o opazovanem pojavu, je potrebna dobra in učinkovita kontrola zbranega gradiva.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ČETRTO POGLAVJE

UREJEVANJE IN OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

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

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

Osnovni podatki, ki jih dobimo z urejevanjem, so osnova za preračunavanja in analizo.

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

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

a) grupiranje vrednosti znakov.

b) šifriranje oziroma signiranje zbranih statističnih podatkov.

c) načrt za osnovno obdelavo.

d) način obdelave podatkov.

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

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

Načrt obdelave določa, po katerih kombinacijah znakov ali grupah znakov bomo obdelali gradivo, da bomo dobili odgovor za dano proučevanje. Pri popisu prebivalstva ne iščemo sestave prebivalstva samo ločeno po spolu in starosti, temveč je pomembna sestava, ki razdeli prebivalstvo po spolu in starosti hkrati. Tako dobimo vpogled ne le v spolno in starostno sestavo ločeno, ampak v spolno-starostno sestavo prebivalstva. Kombinirana obdelava dveh ali več znakov hkrati je ena izmed osnovnih metod obdelave in osnova za analizo odvisnosti med pojavi. Z načrtom obdelave glede na namen proučevanja določimo,katere kombinacije znakov so za proučevanje pomembne in ali nameravamo po teh kombinacijah znake preštevati ali seštevati.

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

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

GRUPIRANJE VREDNOSTI ZNAKOV

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

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

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

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

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

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

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

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

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

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

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

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

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

Za proučevanje pojava je vsekakor boljše 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. Čeprav je za določeno proučevanje taka razdelitev najboljša, so težave v tem, da je treba tako rekoč za vsako proučevanje posebej podrobno razdeliti teritorij na vsebinske grupe. Te grupe v splošnem sekajo upravno-teritorialno razdelitev. Običajno vzamemo kompromisno rešitev, da manjše administrativne enote npr. občine ne grupiramo dalje v republike, marveč po vsebinskem načelu po pretežnosti v vsebinske grupe - rajone. Ta način, čeprav ni najbolj natančen, omogoča rajonizacijo brez posebnih dodatnih težav. Razen tega pa moremo po potrebi grupirati iste podatke po obeh načelih: upravno-teritorialnem in vsebinskem.

Najmanjše krajevne grupe, ki so sestavljene v statistične namene, so statistični okoliši. Ti so sestavljeni tako, da ne sekajo upravno-teritorialnih enot.

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 dne). Razlika med najkrajšim in najdaljšim mesecem je tri dni ali približno deset odstotkov. Še večje so razlike, če namesto koledarskih dni upoštevamo samo delovne dni. S posebnimi prijemi reduciramo podatke tako, da lahko mesečne podatke primerjamo med seboj.

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

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

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

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

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

0 Prehrambeni proizvodi.

1 Pijače in tobak.

2 Surovine, ki niso prehrambene (razen goriv).

3 Mineralna goriva, maziva in sorodni proizvodi.

4 Živalska in rastlinska olja in maščobe.

5 Kmetijski proizvodi.

6 Izdelki, klasificirani pretežno po materialu.

7 Stroji in transportne priprave.

8 Razni gotovi izdelki.

9 Razne transakcije in nikjer omenjeno blago.

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

63 Izdelki iz lesa in plute (razen pohištva)

631 Furnir, vezane plošče, deske, umetno ali rekonstruiran in drugi les, obdelan neomejeno

631-01 Furnir

631-01-11 Furnir bukov normalen m^3/kg

631-01-12 Furnir bukov za stole "

631-01-20 Orehov furnir "

631-01-30 Hrastov furnir "

631-01-40 Furnir slepi "

631-01-90 Furnir iz drugih listavcev "

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

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

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

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

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

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

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

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

$$y_k = \frac{y_{k,min} + y_{k,max}}{2} \tag{4.2}$$

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

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

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

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

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

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

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

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

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

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

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

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

Tako na primer mesečne osebne dohodke grupiramo v grupe

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

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

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

Vsem z medam 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). Takó grupiranje in obravnavanje nezveznih znakov izenačimo z zveznimi. Pri zveznih znakih razmikom pripišemo kot reprezentanta sredino razreda, pri nezveznih pa posamezni vrednosti priredimo enotin razmik.

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

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

Podjetja so po številu delavcev grupirana v tele grupe:

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

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

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

1 2 4 8 15 30 60 125 250 500 1000

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

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

10 25 60 150 400 1000

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

Cikel mej razredov v približnem geometrijskem zaporedju 3 je:

1 3 10

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

Cikel zaporedja s približnim razmerjem 4 pa je

1 4 16 60 250 1000

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

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

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

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

ŠIFRIRANJE OSNOVNEGA GRADIVA

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

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

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

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

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

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

D r u g e o z n a k e uporabljamo redkeje. V praksi pa za znake, ki imajo samo dve grupi, včasih namesto številčnih ali črkovnih šifer uporabljamo oznaki + in –.

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

OSNOVNA OBDELAVA STATISTIČNEGA GRADIVA

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

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

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

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

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

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

Tabele se nanašajo na prebivalstvo:			1	2	3	4	5	6	7	8	9	10
			Zaporedna številka tabele									
Skupno		367	○	○		○			○	○		
Delovno		173			○		○	○				
Vzdrževano		194									○	
Staro nad 10 let		285										
Ženske stare nad 10 let		149										
Vsebina tabela:												
Spol		2	⊜	⊜	(II)		⊜			⊜	(II)	⊜
Starost	A	102	⦶									
	B	15									⊖	
	C	7		⊖								⊖
Kraj rojstva		12										
Stan		5										
Zakon po vrsti		5										
Starost ob poroki		30										
Število živorojenih otrok		13										
Število živih otrok		13										
Državljanstvo		42										
Narodnost		27										
Materin jezik		21										
Odnos do vere		11										
Pismenost		4										
Šolska izobrazba		8					⊖					
Delovnost		3	⊖	(II)		⦶		⦶	(II)	⊖		⦶
Poklic	grupe	10					⦶	⦶	⊖			⦶
	podgrupe	7										
Vzdrževanje		8									⦶	
Položaj v poklicu		7			⊜				⊖			
Gospodarska dejavnost		12		⦶	⦶	⊜		⊖				
Gospodarski sektor		5			⊖	⦶						
Postranski poklic (ima-nima)		2								(II)		
Število članov gospodinjstva		11										
Gospodarstvo: kmečko-nekmečko		2										
Velikost gospodarstva		12								⦶		
število stolpcev		⊖	6	14	35	12	16	12	10	6	15	14
število vrstic		⦶	102	36	24	15	70	70	21	24	16	70
število polj			612	504	840	180	1120	840	210	144	240	980

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tudi v kmetijski statistiki upoštevamo enako načelo. V statistiki živinoreje vzamemo za pogojno žival 400 kg težko govedo. Vse druge živali izražamo s koeficienti v razmerju s to pogojno-normalno živaljo. Tako ponderiramo posamezne vrste živali glede na velikost in proizvodno moč, Ta način omogoča, da seštevamo govejo živino, konje, prašiče, ovce itd. v skupno število normalnih živali. Če vzamemo za zgled konje, imamo za posamezne skupine tele koeficiente: Žrebeta do enega leta 0,6; žrebeta od enega leta do treh let starosti 1,0; kobile nad tri leta 1,4; žrebci 1,3; konji nad tri leta 1,5. Če je na nekem področju 100 žrebet do enega leta, 200 žrebet od enega leta do treh let, 250 kobil nad tri leta, 50 žrebcev in 400 konj nad tri leta, je skupno: 100.0,6+ 200.1,0+ 250.1,4+ 50.1,3+ 400.1,5 = 1275 normalnih živali. Čeprav je ta podatek fiktiven, ga moremo s pridom uporabljati za primerjave med območji; ker z enim podatkom izrazimo skupnost vseh živali na določenem območju, v določenem velikostnem razredu itd.

Osnovna obdelava podatkov

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

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

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

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

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

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

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

Drug način, po katerem združujemo po deset enot je: |X| s štirimi oglišči, štirimi stranicami in dvema diagonalama kvadrata registriramo deset vrednosti.

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

	Prvi način		Drugi način	
Moški	~~////~~ ~~////~~ /	11	⊠ .	11
Ženske	~~////~~ ~~////~~ ////	14	⊠ : .	14
		25 = N		25 = N

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

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

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

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

Črtkanje je uporabno vselej: pri individualnih in kolektivnih obrazcih ali če so podatki navedeni za vse enote na enem listu papirja kakor je to v naših zgledih. To je njegova prednost. Črtkanje pa ima svoje hibe, ki omejujejo uporabnost. Črtkanje je neprikladno za obdelave, v katerih kombiniramo več znakov. Za take obdelave je iskanje ustreznega polja v obdelovalni tabeli zamudno. Razen tega je za populacije velikega obsega črtkanje neuporabno, ker nimamo kontrole. Napake moremo naknadno odkriti le s ponovno obdelavo. Pri črtkanju si pomagamo tako, da osnovno populacijo razstavimo v več obdelovalnih - delovnih populacij. Vendar tudi ta rešitev ni vselej ustrezna. Zato črtkanje uporabljamo le za manjše populacije, ne pa za zamotane obdelave.

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

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

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

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

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

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

PETO POGLAVJE

PRIKAZOVANJE STATISTIČNIH PODATKOV

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

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

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

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

STATISTIČNE VRSTE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tudi analiza časovnih vrst je specifična. S časovnimi vrstami proučujemo d i n a m i k o socialno-ekonomskih pojavov. Ker je analiza časovnih vrst specifična, tudi časovne vrste proučujemo v posebnem, za ekonomista važnem poglavju o analizi časovnih vrst. Dinamika proizvodnje električne energije v SFRJ je nakazana v tabeli 5.2., dinamika števila zaposlenih v družbenem sektorju pa v tabeli 5.3.

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

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

STATISTIČNE TABELE

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

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

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

Delavnost	Delovni		Osebe z osebnimi dohodki		Vzdrževani	
starost / spol	Moški	Ženske	Moški	Ženske	Moški	Ženske
O						
1						
2	V r s t a				S	
3					t	A
4			polje		o	
5					l	
6					p	
Čelo					e	
					c	

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

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

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

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

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

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

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

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

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

Vrste tabel po vlogi v statističnem proučevanju

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

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

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

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

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

T e h n i č n a n a č e l a z a s e s t a v l j a n j e t a b e l.

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

P o j a s n i l a k tabeli dajemo včasih v podnaslovu ali, če so daljša, pod tabelo; to naznačimo v naslovu z ustreznim znakom (zvezdico, križcem ali črko).

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

V tabeli je treba naznačiti e n o t o m e r e, ki jo vnesemo nad tabelo ali v podnaslov, če je enota mere enotna za vse podatke v tabeli (npr. tabela 5.9). Če so enote mere za posamezne podatke različne (npr. tabela 5.8), jih vnesemo v ustrezna polja v glavi oziroma čelu tabele. Opredelitve v o p i s n e m d e l u tabele - glava in čelo tabele - morajo biti kratko in jasno označene. Če zaradi pomanjkanja prostora to v sami tabeli ni mogoče, damo dodatna pojasnila o posameznih pojmih v pripombah pod tabelo.

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

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

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

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

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

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

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

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

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

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

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

GRAFIČNO PRIKAZOVANJE

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

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

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

Zato moremo grafikone deliti po svoji vlogi v dve vrsti: e n o s t a v n e g r a f i k o n e; z njimi ponazarjamo osnovne statistične podatke tako, da so razumljivi čim večjemu številu porabnikov; a n a l i t i č n e g r a f i k o n e, ki so analitično sredstvo raziskovalca.

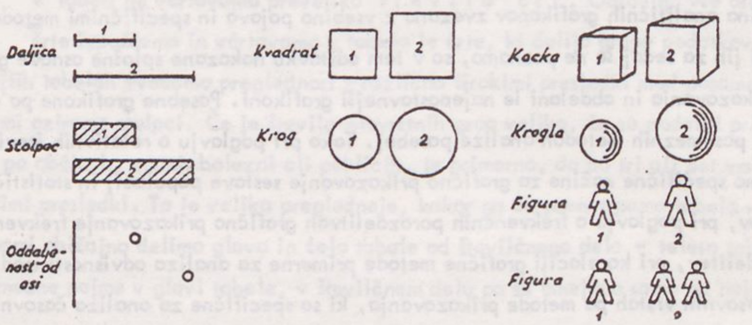
Ker je večina analitičnih grafikonov zvezana z vsebino pojava in specifičnimi metodami analize, ki jih za sedaj še ne poznamo, so v tem odstavku nakazane splošne osnove grafičnega prikazovanja in obdelani le najenostavnejši grafikoni. Posebne grafikone pa obravnavamo pri posameznih metodah analize posebej.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

Osnove grafičnega prikazovanja

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

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

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



Slika 5.1 Prvine grafičnega prikazovanja

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

Razen gornjih elementov, od katerih je vsak določen samo z enim podatkom, uporabljamo za prikazovanje tudi druge elemente, ki imajo več različnih razsežnosti. Pravokotnik, ki ima svojo višini in dolžino, uporabljamo za prikazovanje dveh statističnih podatkov hkrati; zaradi zveze med širino, dolžino in ploščino pa celo treh. Enako je k r o g o v   i z s e k določen z dvema podatkoma: radijem in kotom, ki ga izsek oklepa itd. Te elemente uporabljamo v kompleksnejših primerih grafičnega prikazovanja.

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

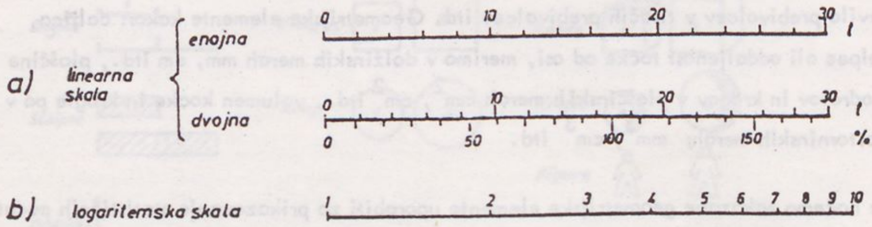
S k a l e - l e s t v i c e .

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

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

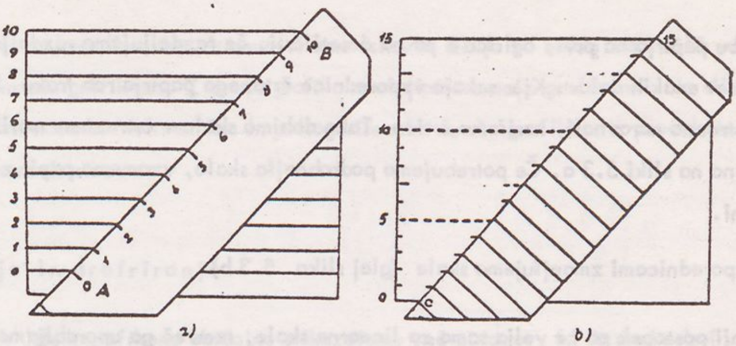
5.23 Uporabljamo različne skale ali lestvice. Najobičajnejše so a r i t m e t i č n e - l i n e a r n e   s k a l e . Na aritmetičnih skalah so daljice med dvema vrednostima v sorazmerju z razliko vrednosti. Aritmetična skala je npr. skala za vrednost proizvodnje v sliki 5.2 in skale v večini grafikonov v tem odstavku (slike 5.7, 5.8, 5.10, 5.12, 5.15).

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

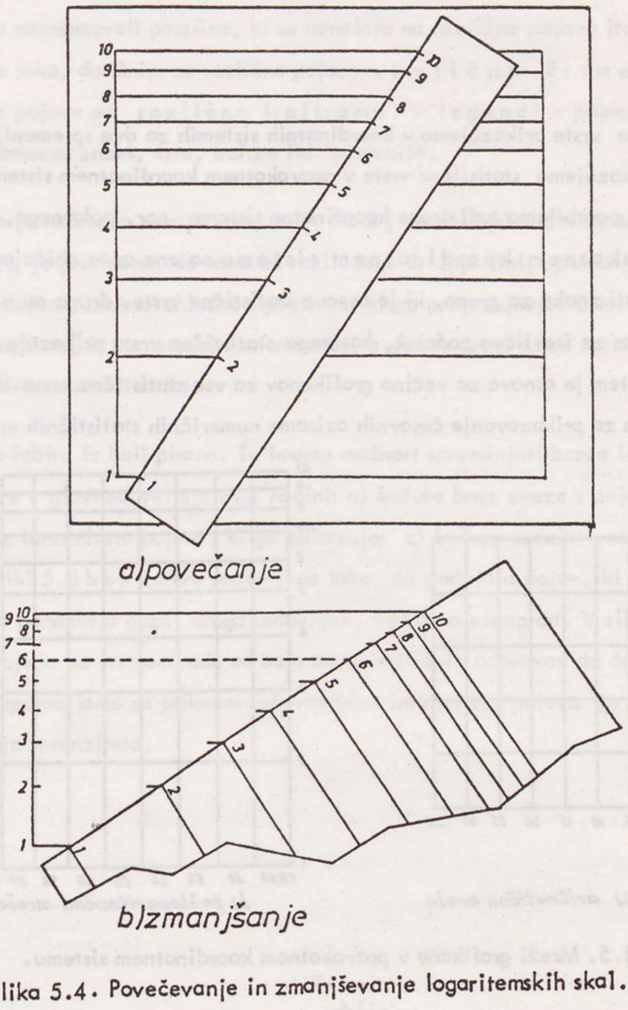


Slika 5.2 Skale ali lestvice

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



Slika 5.3 Risanje linearnih skal v poljubnem merilu



Slika 5.4. Povečevanje in zmanjševanje logaritemskih skal.

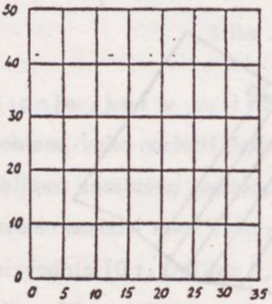
daljice na robu papirja na prvi, oglišče B pa na deseti črti, če razdeljujemo razdaljo 7,8 cm na deset enakih delov. Kjer sekajo vzporednice črtanega papirja rob traku, katerega smo ustrezno naravnali, narišemo črtice. Tako dobimo skalo v ustreznem merilu. To je prikazano na sliki 5.3 a. Če potrebujemo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

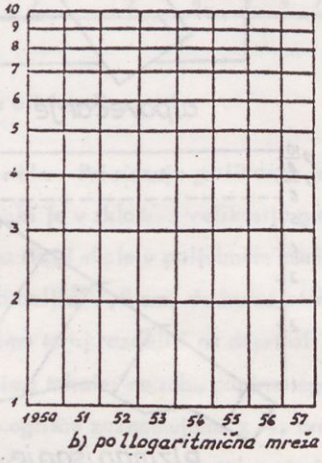
Podobno z vzporednicami zmanjšujemo skale (glej sliko 5.3 b).

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

Mreže

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





Slika 5.5. Mreži grafikona v pravokotnem koordinatnem sistemu.

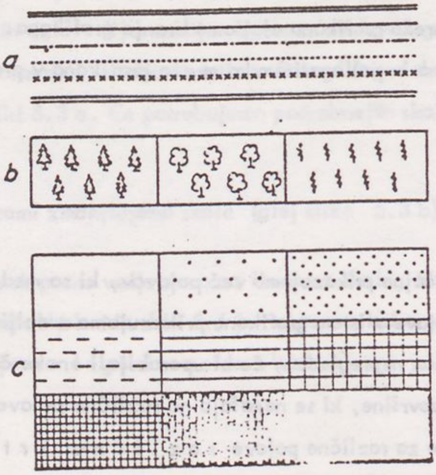
pravokotnem koordinatnem sistemu vrišemo osnovne linije za nekaj važnejših vrednosti, dobimo mrežo grafikona. Z mrežo grafikona olajšamo branje grafikona. Na sliki 5.5 sta vrisani navadna aritmetična in pollogaritemska mreža grafikona v pravokotnem koordinatnem sistemu.

Črtanje in šrafriranje

5.28 V istem grafikonu običajno prikazujemo več pojavov, ki so med seboj vsebinsko povezani. Ker vse pojave v istem grafikonu prikazujemo z daljicami, stolpci, površinami itd., bi bil grafikon nepregleden, če bi uporabljali enake črte za različne pojave, enako zaznamovali površine, ki se nanašajo na različne pojave itd. Preglednost povečamo tako, da linije za različne pojave različno črtamo, površine za različne pojave pa različno šrafiramo. V legendi - pojasnilu pa označimo, kaj posamezni znaki, črte, šrafure itd. pomenijo.

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

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



Slika 5.6. Črte in šrafure

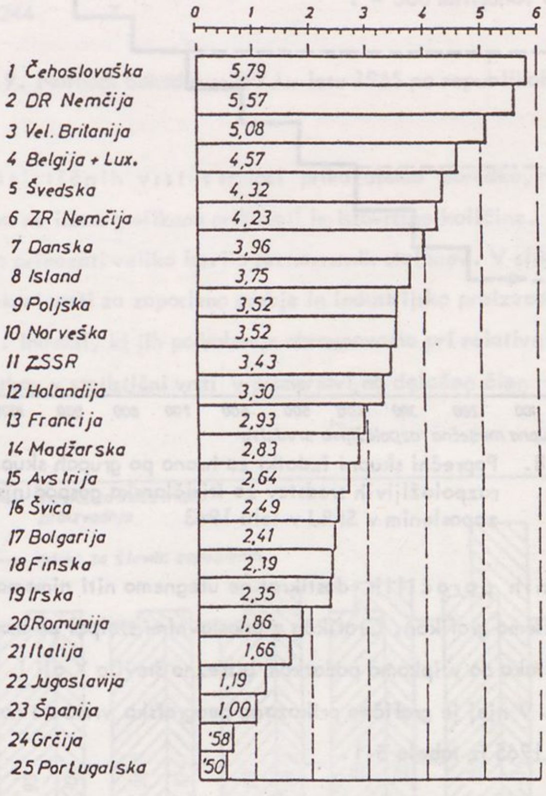
Vrste grafikonov

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

S t o l p c i

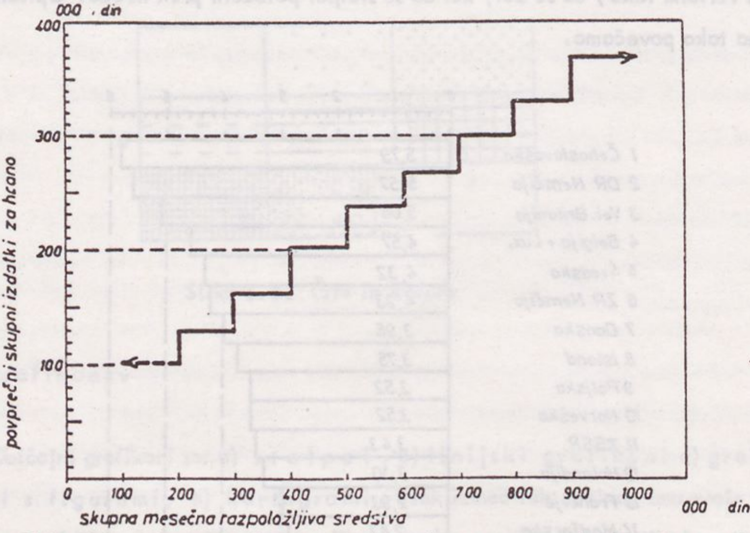
5.30 Z e n o s t a v n i m i s t o l p c i prikazujemo grafično vse statistične vrste. To je prednost stolpcev. Te lastnosti linijski grafikoni npr. nimajo, čeprav so zaradi drugih lastnosti bolj čislani. S stolpci prikazujemo predvsem tiste statistične vrste, ki jih ne moremo prikazovati z linijskimi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

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



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

Stolpce uporabljamo včasih tudi za prikazovanje numeričnih vrst, čeprav so linijski grafikoni zanje priporočljivejši. V sliki 5.8 imamo prikazane povprečne denarne izdatke za hrano za gospodinjstvo po grupah skupnih razpoložljivih sredstev za štiričlansko gospodinjstvo z enim zaposlenim v SFRJ v letu 1963 iz tabele 5.5. Posebnost tega grafikona je v tem, da so vrisani samo obrisi stolpcev. Tako povečamo nazornost grafikona. Enako je mreža včrtana podobno kot v gornjem primeru samo do stolpcev. Širina stolpcev je v razmerju s širino razredov. Le tako dobimo pravilen vtis o spremembah med posameznimi razredi denarnih dohodkov. Slika zelo nazorno pokaže odvisnost izdatkov od dohodkov v kmetijskih gospodarstvih.



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

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

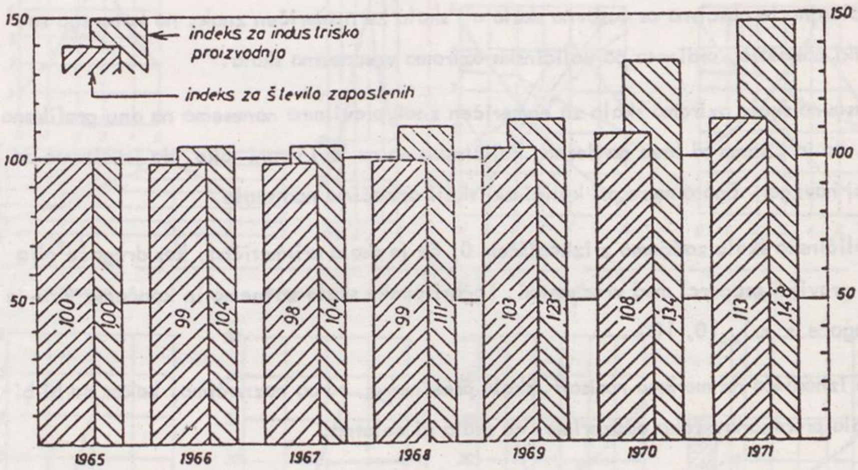
Narodni dohodek v mlj din		1 2 3 123456789012345678901234567890
Srbija	28.712	XXXXXXXXXXXXXXXXXXXXXXXXXXXXI
Hrvatska	19.422	XXXXXXXXXXXXXXXXXXXI
Slovenija	11.263	XXXXXXXXXXXI
BiH	9.022	XXXXXXXXX
Makedonija	3.907	XXXX
Črna gora	1.244	X

X = 1 milijarda din

I = 500 milijonov din

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

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



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

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

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

Linijski grafikoni

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

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

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

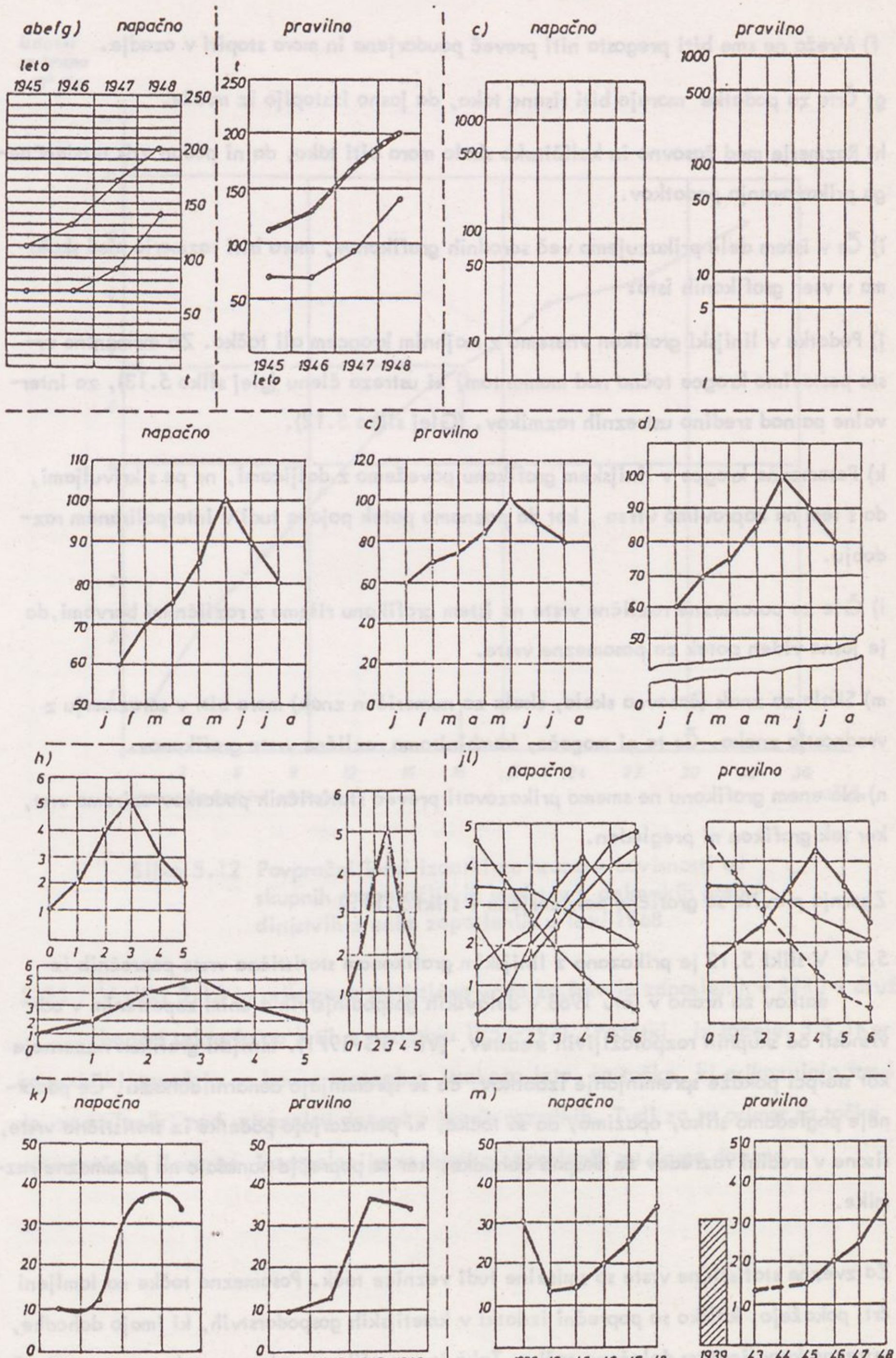
a) Običajno je abscisna os časovna skala ali skala za numeričen znak, na katerega se podatki nanašajo, ordinata pa količinska oziroma vrednostna skala.

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

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

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

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



Slika 5.11 Pravila za risanje linijskih grafikonov

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

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

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

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

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

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

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

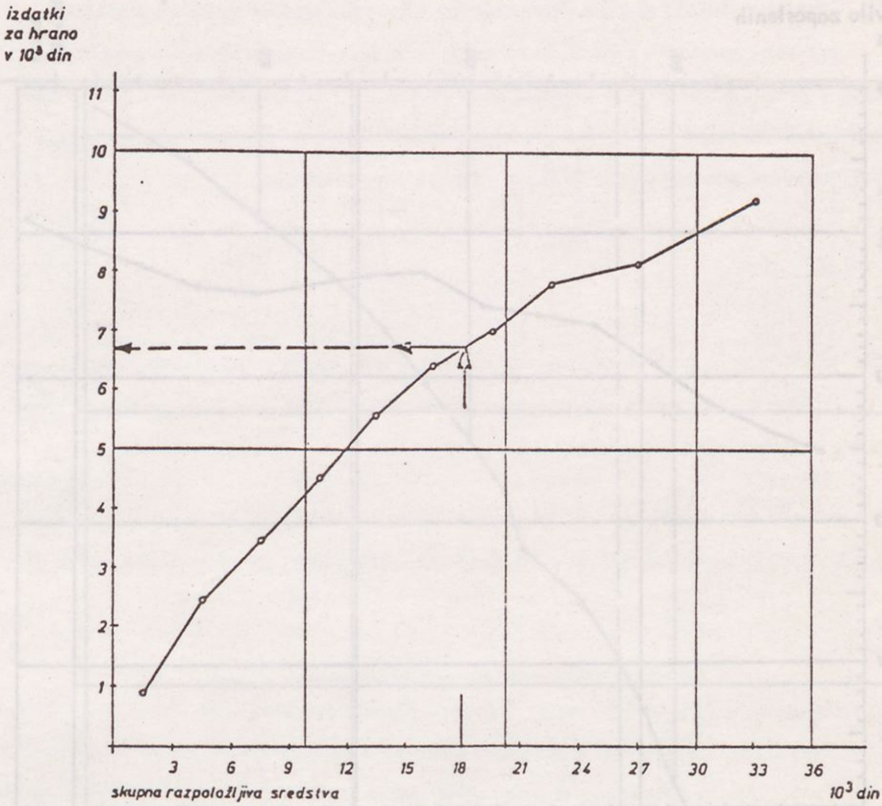
m) Skala za znak (časovna skala, skala za numeričen znak) mora biti v sorazmerju z vrednostjo znaka. Če to ni mogoče, kombiniramo različne vrste grafikonov.

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

Zgornja pravila so grafično ponazorjena v sliki 5.11.

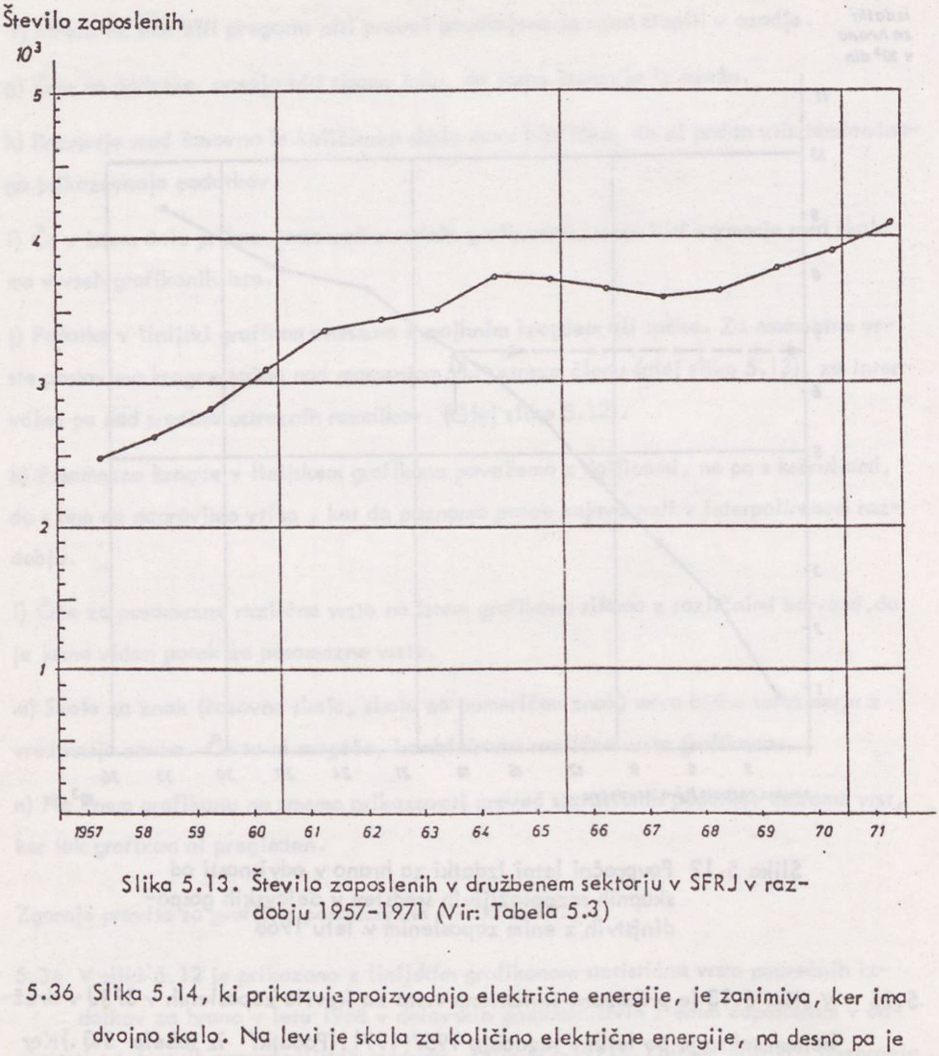
5.34 V sliki 5.12 je prikazana z linijskim grafikonom statistična vrsta poprečnih izdatkov za hrano v letu 1968 v delavskih gospodinjstvih z enim zaposlenim v odvisnosti od skupnih razpoložljivih sredstev. (Vir: SG-1971). Linijski grafikon nazorneje kot stolpci pokaže spreminjanje izdatkov, če se spreminjajo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

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



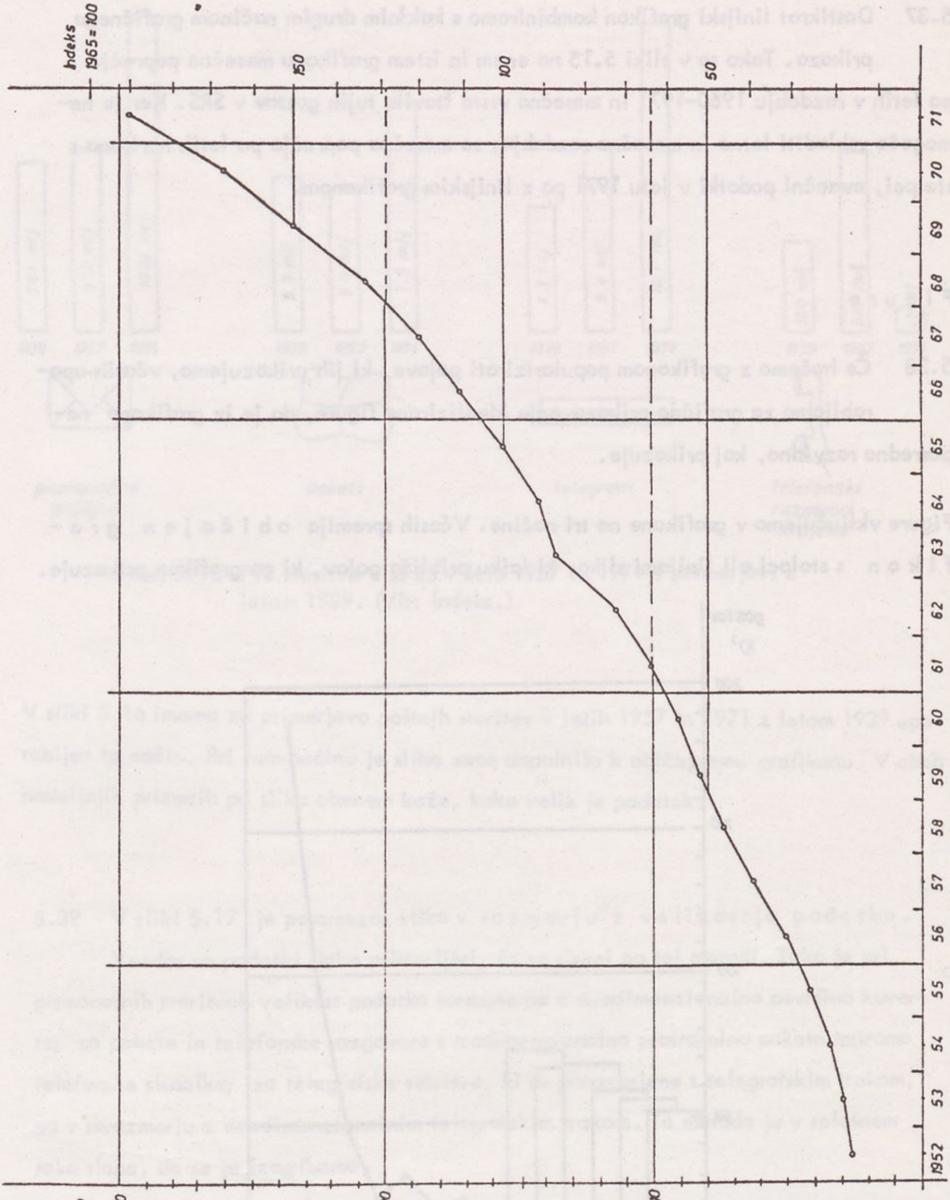
Slika 5.12 Povprečni letni izdatki za hrano v odvisnosti od skupnih razpoložljivih sredstev v delavskih gospodinjstvih z enim zaposlenim v letu 1968

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



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

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



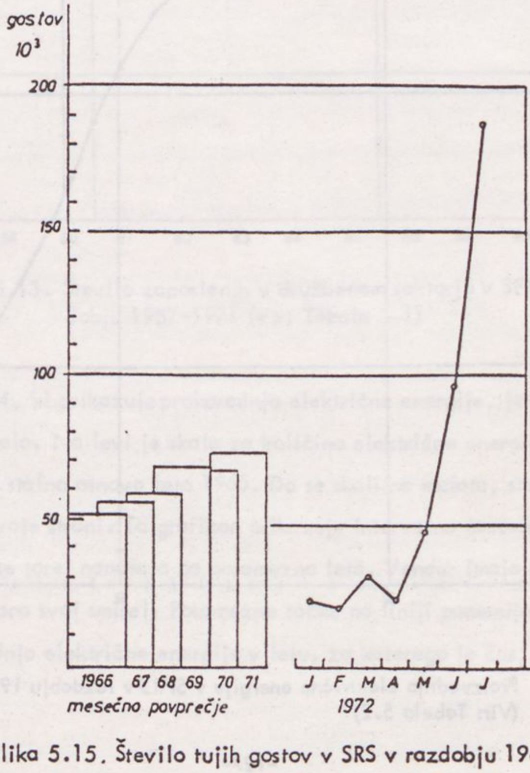
Slika 5.14. Proizvodnja električne energije v SFRJ v razdobju 1952-1971.
(Vir: Tabela 5.2)

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

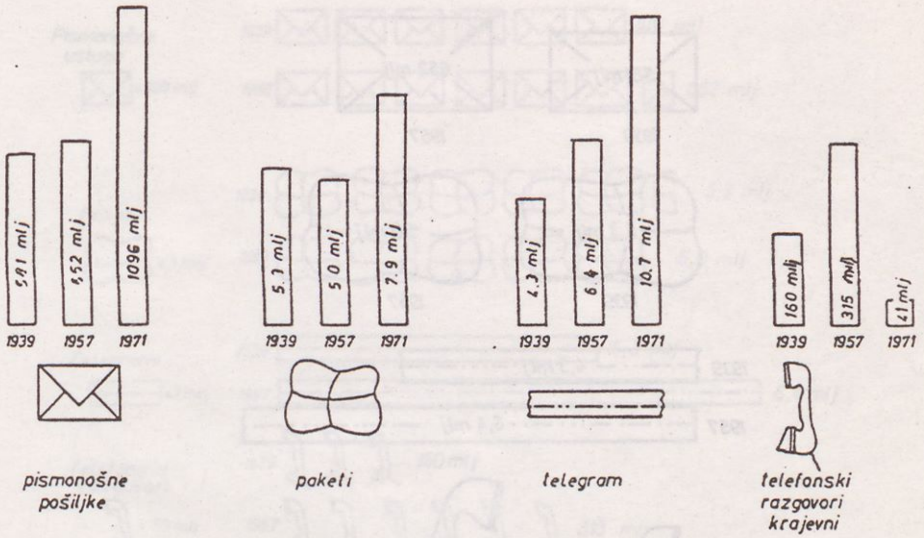
F i g u r e

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

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



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



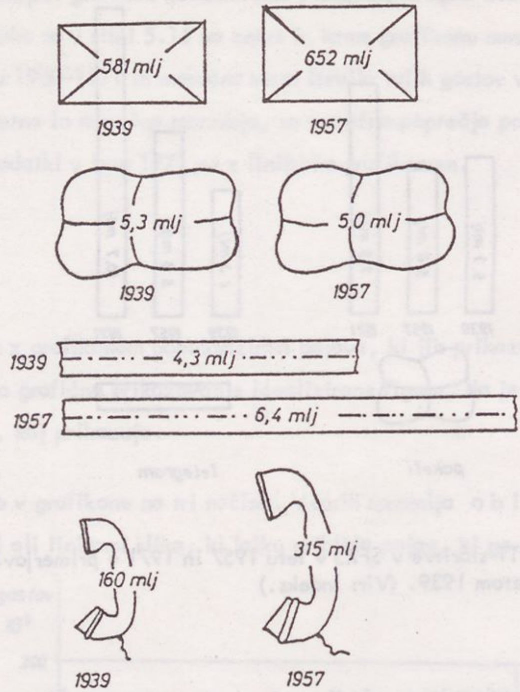
Slika 5.16 PTT storitve v SFRJ v letu 1957 in 1971 v primerjavi z letom 1939. (Vir: Indeks.)

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

5.39 V sliki 5.17 je posamezna slika v razmerju z velikostjo podatka.

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

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

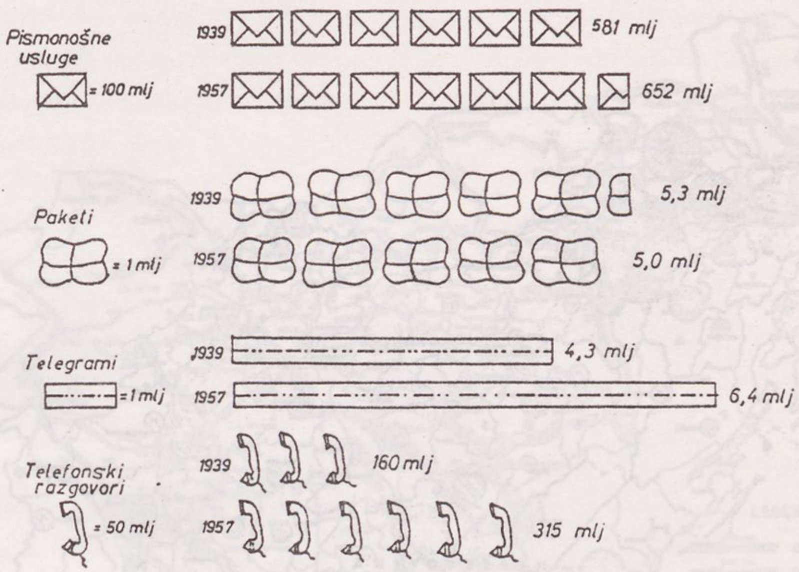


Slika 5.17. PTT storitve v SFRJ v letu 1957 v primerjavi z letom 1939.

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

Kartogrami

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



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

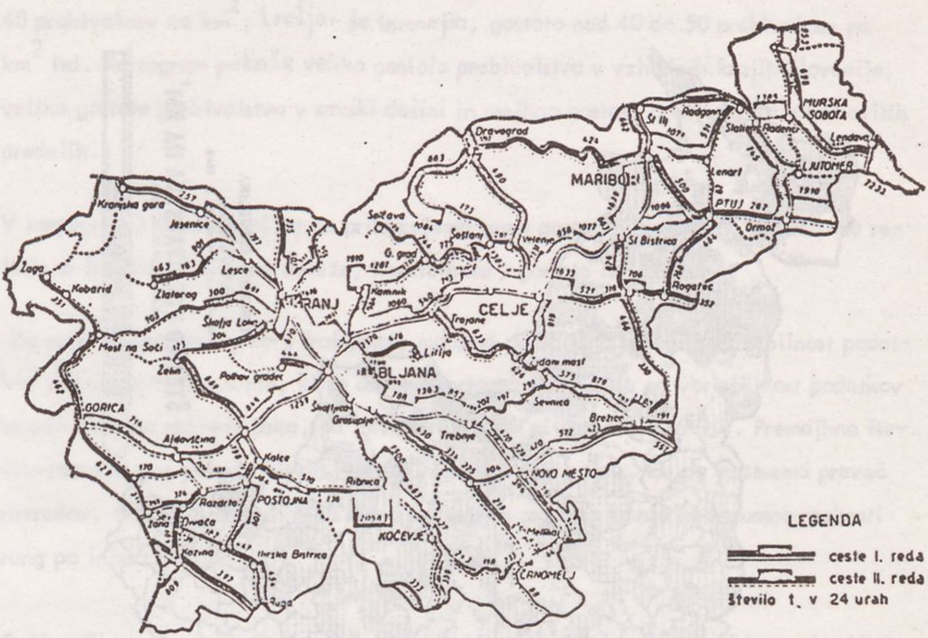
ne, ki kakor koli prikažejo razmestitev določenega pojava v geografski karti, imenujemo kartograme v širšem smislu.

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

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

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

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



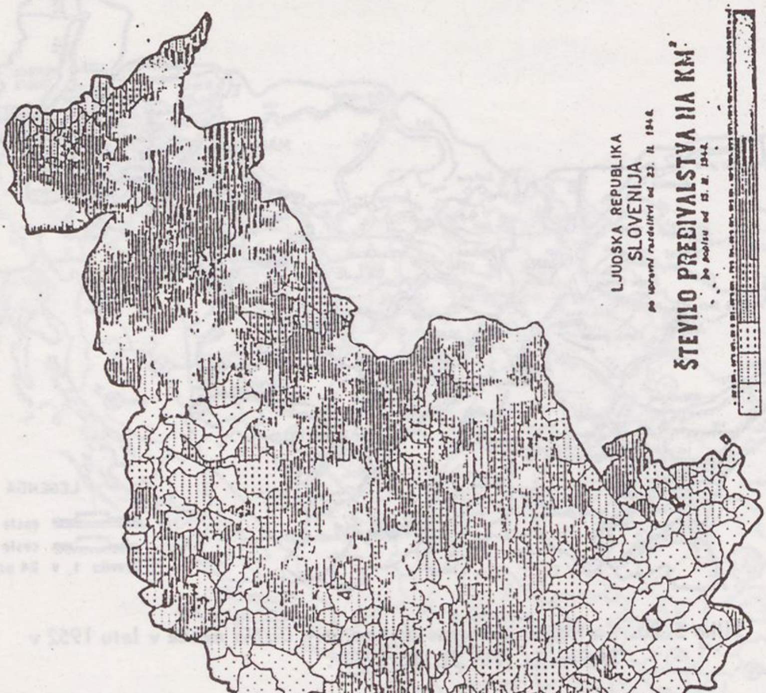
Slika 5.20. Kartogram prometne obremenitve cestne mreže v letu 1952 v Sloveniji. (Vir: SL LRS 1953).

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

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

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

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



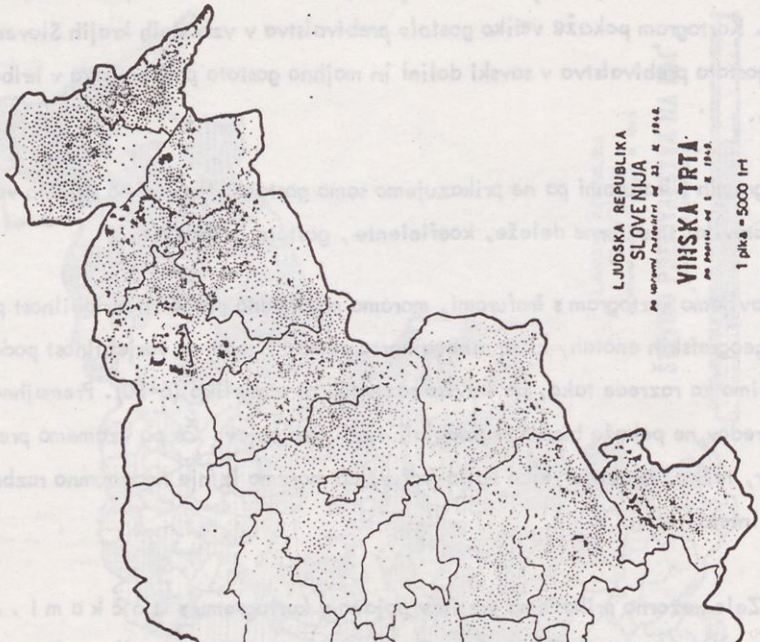
Slika 5.21. Kartogram gostote prebivalstva v Sloveniji po popisu 15.3.1948. (Vir: Statistični atlas Slovenije.)

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

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

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

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



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

ŠESTO POGLAVJE

RELATIVNA ŠTEVILA

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

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

Relativno število iz dveh podatkov, ki ju primerjamo, je v splošnem kvocient med primerjanima podatkoma. Razumljivo je, da enako kakor med seboj primerjamo le podatke, za katere je primerjava smiselna, izračunavamo tudi relativna števila le iz podatkov, ki so v medsebojni vsebinski povezavi. Relativna števila izračunamo v oblikah, ki so prikladnejše za prikazovanje in analizo (npr. v odstotkih, indeksih ipd.).

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

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

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

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

STRUKTURE ALI RAZČLENITVENA ŠTEVILA.

Enostavne strukture.

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

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

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

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

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

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

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

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

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

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

Večkratne strukture.

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

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

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

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

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

od skupnega družbenega produkta.

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

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

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

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

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

Tabela 6.4. Struktura družbenega produkta sektorjev po elementih.

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

Tabela 6.5 Struktura elementov družbenega produkta po gospodarskih sektorjih

Gospodarski sektor	Neto osebni dohodki	Akumulacija in fondi	Amortizacija	Skupno
Družbeni	60,4	91,4	91,1	78,5
Privatni	39,6	8,6	8,9	21,5
Skupaj	100,0	100,0	100,0	100,0

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

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

Splošne zveze za dvojne strukture

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

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

Ustrezni strukturni deleži na celotno populacijo so

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

Vrstične ali stolpične strukturne vrste zapišimo takole

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

$$
\begin{array}{c|c}
\dfrac{N_{AB}}{N_A} = P_{B/A} & \dfrac{N_A}{N_A} = 1 \\
\hline
\dfrac{N_B}{N} = P_B & \dfrac{N}{N} = 1
\end{array}
\qquad (6.5)
$$

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

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

Iz gornjih tabel dobimo naslednje zveze:

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

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

iz te zveze pa dobimo dalje obrazec

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

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

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

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

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

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

Tabela 6.7 Struktura kakovosti po oddelkih

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

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

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

Razen stavka o množenju verjetnosti za odvisne dogodke

$$Pr(A \cap B) = Pr(B) \cdot Pr(A/B) = Pr(A) \cdot Pr(B/A) \tag{6.8}$$

dobimo pomemben Bayesov obrazec za inverzno verjetnost

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

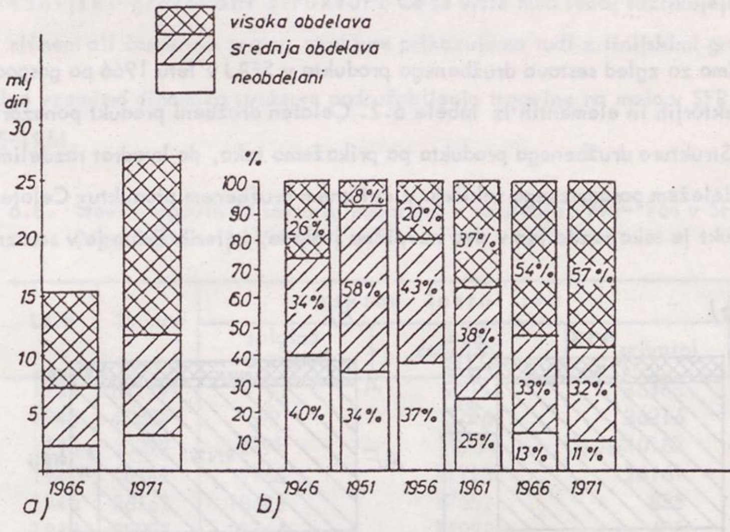
s katerim bomo v kasnejših poglavjih izvajali pomembne sklepe.

Grafično prikazovanje struktur

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

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

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



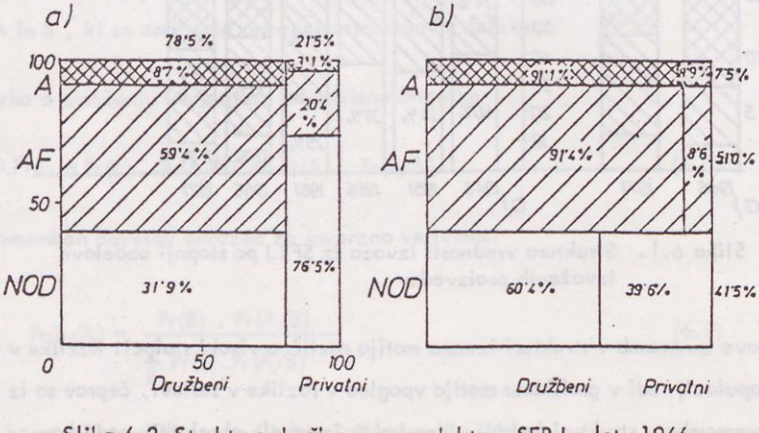
Slika 6.1. Struktura vrednosti izvoza iz SFRJ po stopnji obdelave izvoženih proizvodov

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

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

6.11 S t r u k t u r n i s t o l p c i z a v e č k r a t n e s t r u k t u r e. Širine strukturnih stolpcev iz prejšnjega zgleda so enake za vsa leta. Širine stolpcev pa moremo risati v sorazmerju z obsegom populacije. Vrsta takih strukturnih stolpcev pokaže tudi razlike v obsegu populacije. Ker pa so vsi strukturni stolpci enako visoki, je iz njih neposredno vidna sprememba v strukturi. Ta postopek s pridom izkoriščamo predvsem za prikazovanje večkratnih struktur.

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



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

ju z deležem posameznega sektorja. Če vsakega izmed teh stolpcev razdelimo v navpični smeri v sorazmerju s sestavo družbenega produkta za posamezni sektor po elementih, dobimo sliko 6.2a. Ta slika je fotografija tabele 6.4. Iz grafikona je zelo nazorno vidna struktura družbenega produkta po sektorjih, predvsem pa razlike v sestavi med sektorji. Poleg tega pa je ploščina posameznega pravokotnika v kvadratu sorazmerna z absolutno velikostjo določenega elementa oziroma s strukturnimi odstotki iz tabele 6.4.

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

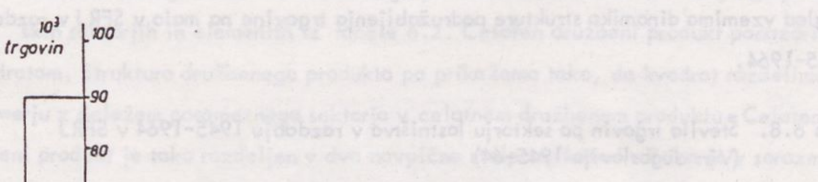
6.13 L i n i j s k i g r a f i k o n i s t r u k t u r . Če se vrste med seboj razlikujejo v numeričnem ali časovnem znaku, strukture prikazujemo tudi z linijskimi grafikoni.

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

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

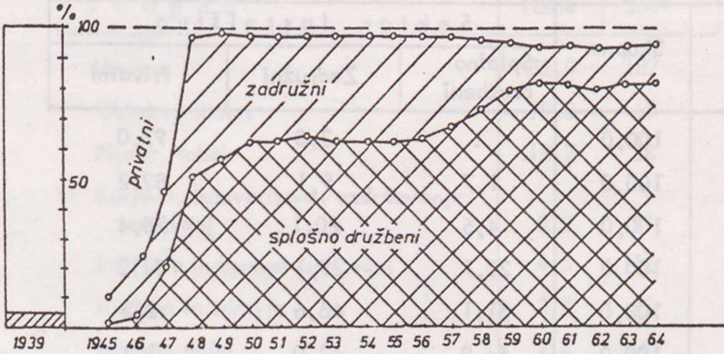
Leto	Skupno	sektor lastništva		
		splošno družbeni	zadružni	privatni
1939	89179	.	2711	86468
1945	40624	692	3716	36216
1946	53292	2391	10734	40167
1947	48242	9968	12113	26161
1948	36562	18675	17052	835
1949	38096	21664	15999	433
1950	32516	20487	11415	614
1951	30541	19281	10560	700
1952	32952	21477	10791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

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



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

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



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

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

6.14 Strukturni krogi. Za prikazovanje struktur se izkaže kot primeren tudi krog, ker nazorno prikazuje celoto. Strukturne deleže ponazorimo s krokovimi izseki, kot za posamezen izsek pa je v sorazmerju z velikostjo strukturnega deleža. Ker četrtina kroga ponazarja 25%, polovica kroga pa 50%, s pogledom enostavno ocenimo velikost strukturnih deležev.

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

$$Y_1^{st} = 3,6\ Y_1\ \% \tag{6.10}$$

v ločne stopinje Y_1^{st}.

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

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

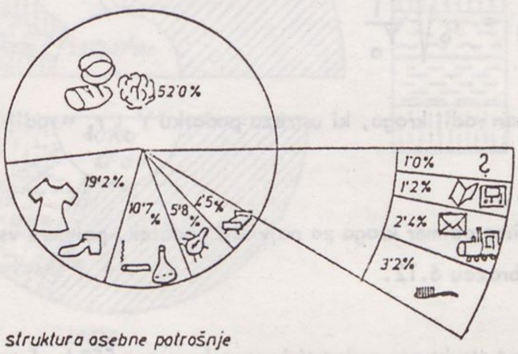
6.15 Kot zgled bomo prikazali strukturo za osebno porabo materialnih dobrin in proizvodnih stroškov za prebivalstvo v SFRJ v letu 1956.

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

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

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

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



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

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

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

$$Y_1^{st} = 1{,}8\, Y_1\% \tag{6.11}$$

ker celoto predstavlja le 180°.

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

$$r_1 = r_o \sqrt{\frac{Y_1}{Y_o}} \tag{6.12}$$

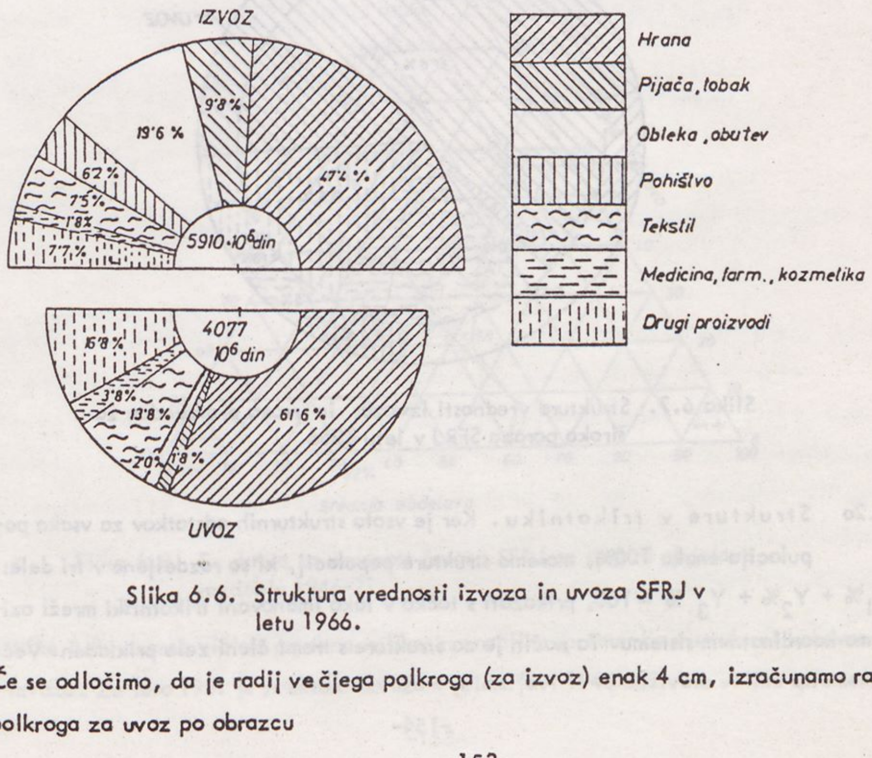
Pri tem pomeni: r_o = znan radij kroga, ki ustreza podatku Y_o; r_1 = radij kroga za podatek Y_1.

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

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

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

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



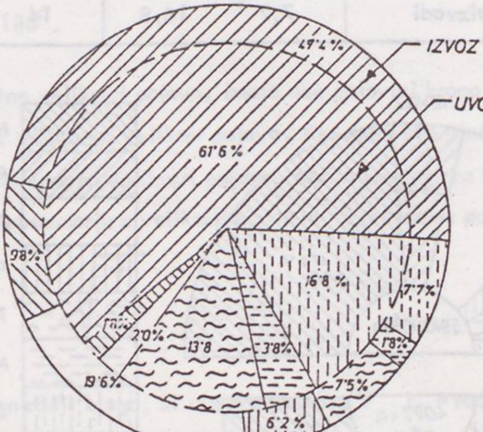
Slika 6.6. Struktura vrednosti izvoza in uvoza SFRJ v letu 1966.

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

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

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

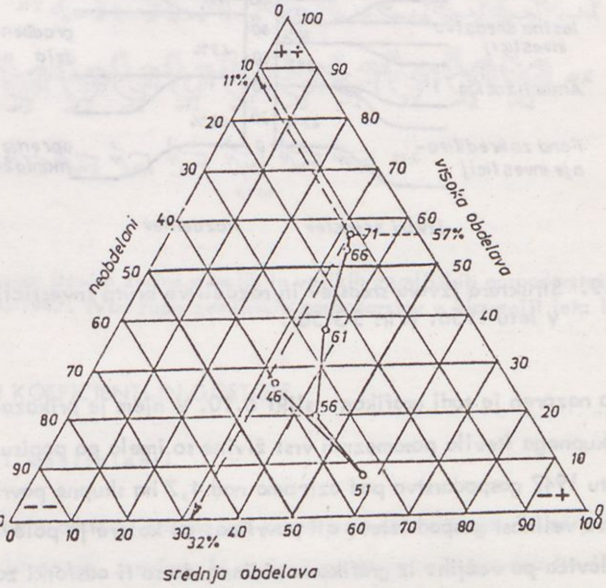
6.19 Primerjava strukture izvoza s strukturo uvoza pa je veliko bolj neposredna v sliki 6.7. V tej sliki si mislimo, da je strukturni krog za uvoz (manjši krog) položen koncentrično na strukturni krog za izvoz (večji krog). Ker so izseki kroga za ustrezne skupine proizvodov blizu skupaj, je primerjava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



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

6.2o Strukture v trikotniku. Ker je vsota strukturnih odstotkov za vsako populacijo enaka 100%, moremo strukture populacij, ki so razdeljene v tri dele: $Y_1\% + Y_2\% + Y_3\ \% = 100$, prikazati s točko v tako imenovani trikotniški mreži oziroma koordinatnem sistemu. Ta način je za strukture s tremi členi zelo prikladen. Več sorodnih strukturnih vrst s tremi členi v njej enostavno prikažemo s tolikimi točkami, kolikor imamo strukturnih vrst. Te točke glede na značaj vrst povezujemo v linijske grafikone, vektorje itd.

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

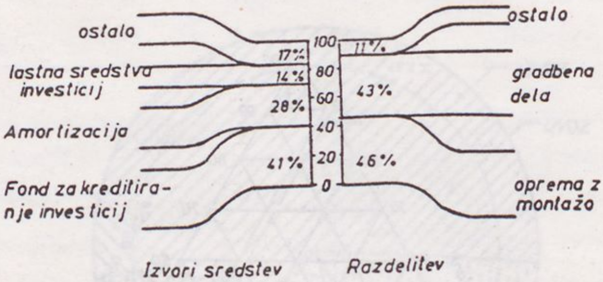


Slika 6.8. Struktura za vrednost izvoza SFRJ po stopnji obdelave v razdobju 1946-71.

Iz slike 6.8 nazorno vidimo posebno velike in značilne spremembe v strukturi kakovosti izvoza. Za leto 1951 je struktura izvoza v primerjavi 1946 doživela velike spremembe v zvezi z izdelki srednje in visoke obdelave. Od leta 1956 naprej pa je vidna trajna težnja v spremembi strukture v smeri zmanjšanja izvoza neobdelanih proizvodov v dobro obdelanih proizvodov.

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

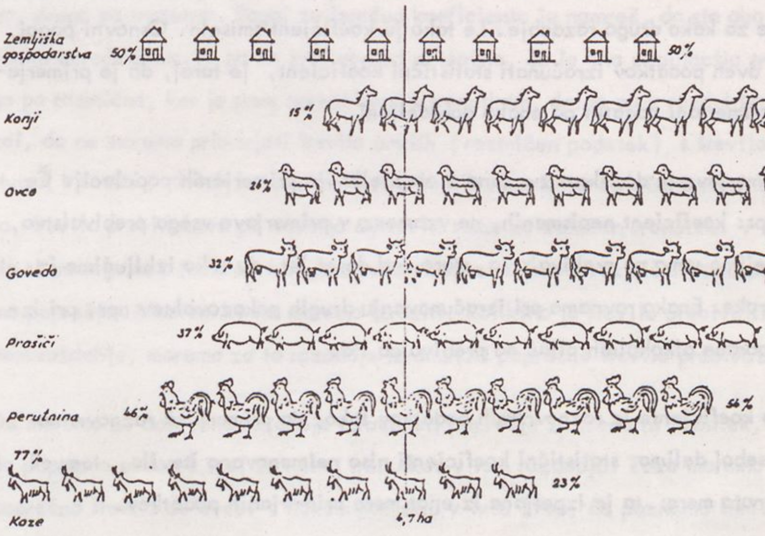
6.22 Druge metode za prikazovanje struktur. Ker si moremo predstavljati bilančne podatke kot tok, ki na eni strani doteka, na drugi pa odteka, strukturo bilanc nazorno prikažemo po metodi, ki je nakazana v sliki 6.9. Grafikon ponazarja izvor in razdelitev bruto investicij v letu 1956 v SFRJ. Posamezni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



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

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

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



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

STATISTIČNI KOEFICIENTI IN GOSTOTE

Statistični koeficienti

6.24 Strukturne odstotke dobimo, če primerjamo del s celoto. Primerjana podatka sta torej istovrstna, rezultat, ki ga dobimo, pa je neimenovano število; običajno ga izražamo v odstotkih. Smiselna pa je tudi primerjava raznovrstnih podatkov, če zadoščajo nekim osnovnim pogojem primerljivosti. Primerjana podatka morata biti namreč v vsebinski zvezi in enako opredeljena. Če primerjamo število prebivalstva s skupno površino, pokaže ta primerjava gostoto prebivalstva. Primerjava proizvodnje s številom ur, ki so bile potrebne za izdelavo te proizvodnje, pokaže produktivnost dela itd. Nima pa smisla primerjati npr. vrednosti proizvodnje na nekem območju s številom tifusnih obolenj na istem območju, ker ta dva pojava nista v vsebinski zvezi. Promet v trgovini na drobno v Sloveniji v decembru 1968 pa je smiselno primerjati le s številom prebivalstva v SR Sloveniji in ne v SR Hrvatski in le s srednjim številom prebivalstva v decembru 1968 in ne za kako drugo razdobje. Le tako je koeficient smiseln. Osnovni pogoj, da moremo iz dveh podatkov izračunati statistični koeficient, je torej, da je primerjava umestna, primerjani podatki pa enako opredeljeni.

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

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

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

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

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

Ob istem popisu je bila gostota prebivalstva v drugih republikah: v Srbiji 79,0, 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).

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

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

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

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

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

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

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

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

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

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

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

Če vzamemo, da je razdobje, za katerega hočemo izračunati poprečje, sestavljeno iz enega samega osnovnega razmika (r = 1), dobimo po obrazcu 6.13, da je $\bar{X} = X_1$

Iz tega sledi, da moremo poprečne nadomestiti s stanjem v sredini razdobja. Tako npr. nadomeščamo poprečno letno število prebivalstva s "srednjim številom prebivalstva", to je s številom prebivalstva sredi leta (30. junija).

Če je r = 1, obrazec 6.14 degenerira v

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

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

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

$$K = \frac{Y \, E}{\bar{X} \cdot i} \tag{6.15}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Po postopku za računanje koeficientov iz podatkov v tabeli 6.12 izračunamo skupno porabo kremenovega peska v prvem polletju 1958 in poprečno zalogo v tem razdobju. Skupna poraba v polletju je vsota mesečne porabe za ustreznih šest mesecev.

$$Y = 494 + 489 + 562 + 468 + 465 + 544 = 3022 \text{ t}$$

Poprečno zalogo $\bar{X}$ pa izračunamo po obrazcu 6.14, ker imamo podatke za konce mesecev.

$$\bar{X} = \frac{\frac{664}{2} + 774 + 888 + 789 + 546 + 568 + \frac{726}{2}}{6} = 710 \text{ t}$$

Poprečna zaloga je torej 710 t.

Ker je koeficient merjen s številom obratov v enem mesecu, naše razdobje pa obsega šest mesecev, je $i = 6$. Iz teh podatkov izračunamo, da je koeficient obračanja zalog za kremenov pesek v prvem polletju 1958 enak

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

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

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

Obračanje zalog pa moremo izraziti tudi z dolžino enega obrata. Ta koeficient, ki je recipročni pokazovalec za gornji pokazovalec, izračunamo po obrazcu

$$K = \frac{\bar{X} \cdot i}{\bar{Y}} \tag{6.8}$$

Če dolžino enega obrata izrazimo v delovnih dneh, je za gornji zgled i = 181, koeficient obračanja zalog za kremenov pesek v podjetju A, izražen z dolžino enega obrata pa je enak

$$K = \frac{710 \cdot 181}{3022} = 42{,}5 \text{ dni}$$

6.30 Po navedenem postopku računamo še niz drugih koeficientov iz ekonomskih statistik. Tako je npr. proizvodnja na enega delavca pokazovalec o produktivnosti dela, proizvodnja določenih proizvodov na prebivalca npr. jekla, pokazovalec industrializacije, poraba določenega proizvoda na prebivalca pokazovalec standarda, razmerje med prometom in zalogami pokazovalec obračanja zalog, število prodanih knjig na leto na 1000 prebivalcev pokazovalec za kulturno raven itd.

6.31 Kot zgled za kompleksno analizo s koeficienti vzemimo trgovino na drobno v letu 1957 po republikah. V tabeli 6.11 imamo osnovne podatke, iz katerih bomo izračunali koeficiente.

Tabela 6.11 Trgovina na drobno v SFRJ v letu 1957 po republikah (Vir: SG-59)

Pokazovalec	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Število prebivalstva 30.VI.1957 v 10^3	18234	7448	4109	1582	3168	1454	473
Število trgovin na drobno:							
31.XII.1956	36257	14017	9296	4662	4634	2795	853
31.XII.1957	37651	14601	9798	4722	4822	2810	898
Število zaposlenih v trgovini na drobno							
31.XII.1956	112511	45954	27589	13689	14184	8523	2572
31.XII.1957	116119	46187	29292	14021	14566	9141	2912
Promet v letu 1957 v 10^6							
Skupno	750776	291478	198339	110259	90123	44331	16246
od tega:							
Živila	232637	75604	64414	37519	32744	15348	7008
Industrijski proizvodi	483608	202410	125065	68917	51662	27133	8421

Iz tabele 6.11 vidimo, da so prvi trije podatki trenutni, drugi pa razmični. Ker imamo srednje število prebivalstva, ta podatek vzamemo za poprečno število prebivalstva. Za število trgovin in število zaposlenih pa imamo stanje na začetku in na koncu leta 1957. Zaradi tega najprej izračunamo poprečje, da bi mogli število trgovin in število zaposlenih primerjati s prometom, ki je razmičnega značaja.

Za število trgovin v SFRJ je npr. iz gornjih podatkov poprečno število trgovin enako polovici vsote števila trgovin po stanju 31.XII.1956 in 31.XII.1957

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

Podobno izračunamo vsa druga poprečja. Tako dobimo vrsti za poprečno število trgovin in za zaposleno osebje v trgovini na drobno po republikah.

Tabela 6.12. Poprečno število trgovin in zaposlenega osebja v trgovini na drobno v letu 1957 po republikah

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

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

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

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

Primerjava števila trgovin in zaposlenega osebja s številom prebivalstva iz osnovnih podatkov ni možna, ker imamo število prebivalstva na dan 30.VI.1957, podatke o številu trgovin in številu zaposlenih pa na dan 31.XII.1956 in 31.XII.1957. Primerjani trenutni populaciji nista časovno enako opredeljeni. Ker pa moremo imeti srednje število prebivalstva za poprečje v letu 1957, smo upravičeni, da primerjamo srednje število prebivalstva s poprečnim številom trgovin in zaposlenih v letu 1957, ker se vsi ti podatki nanašajo na leto 1957. Število zaposlenih moremo primerjati s številom trgovin po stanju 31.XII.1956 ali 31.XII.1957, ker imamo za te datume podatke o številu trgovin in številu zaposlenih. Moremo pa primerjati tudi poprečno število zaposlenih s poprečnim številom trgovin, ker veljata oba podatka za isto razdobje. V tabeli 6.13 so izračunani vsi nakazani koeficienti.

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

Pokazovalec	SFRJ	Srbija	Hrvatska	Slovenija	BiH	Makedonija	Črna gora
Število prebivalcev na eno trgovino	493	520	431	337	670	519	539
Število prebivalcev na enega zaposlenega	159	165	144	114	221	165	172
Skupni promet na enega prebivalca v tisoč din	41,2	39,1	48,3	69,7	28,4	30,5	34,2
Promet z živili na prebivalca v tisoč din	12,8	10,2	15,7	23,7	10,3	10,5	14,8
Promet z industrijskimi proizvodi na prebivalca v tisoč din	26,5	27,1	30,4	43,6	16,3	18,6	17,8
Število zaposlenih na eno trgovino (poprečno)	3,09	3,22	2,98	2,95	3,04	3,15	3,11
Skupni promet na eno trgovino v mlj. din	20,3	20,4	20,8	23,5	19,0	15,8	18,5
Skupni promet na enega zaposlenega v mlj. din	6570	6320	6970	7970	6260	5020	5930
Promet z industrijskimi proizvodi na sto din prometa z živili	207	267	194	184	158	176	120

Tabela koeficientov v trgovini na drobno ilustrativno pokaže razlike v trgovini na drobno po republikah. Najmanj prebivalcev pride na eno trgovino v Sloveniji (337), največ pa v BiH (670). Enako velja za število prebivalstva, ki pride na enega zaposlenega (Slovenija 114, BiH 221). Ti podatki kažejo gostoto trgovinske mreže. Tudi skupni promet na enega prebivalca kaže velike razlike med republikami (od 28,4 tisoč dinarjev za BiH do 69,7 za Slovenijo). Vendar ta koeficient nima polne analitične veljave, ker je njegova vrednost odvisna od velikega števila faktorjev (socialne strukture prebivalstva, kupne moči itd.). Promet z živili na prebivalca je precej odvisen od odstotka nekmečkega prebivalstva, promet z industrijskimi proizvodi na prebivalca pa moremo imeti bolj za pokazovalec o kupni moči in standardu prebivalstva. Tudi ta koeficient kaže velike razlike med republikami. Medtem ko je za BiH, Črno goro in Makedonijo izredno nizek, je za druge republike, posebno za Slovenijo, znatno večji.

Velikost trgovinskih obratov, ki je razvidna iz poprečnega števila zaposlenih na eno trgovino, med republikami malo variira. Iz tega sklepamo, da je tip trgovin po republikah precej enoten. Enako težnjo kaže tudi skupen promet na eno trgovino in so znatnejše razlike edino le za Slovenijo in Makedonijo. Promet na enega zaposlenega, ki ga štejemo za pokazovalec produktivnosti dela v trgovini, kaže znatno odklonitev navzgor za Slovenijo in znatno odklonitev navzdol za Makedonijo. Razmerje med prometom z industrijskimi proizvodi in prometom z živili kaže med republikami velike razlike. Največji koeficient dobimo za Srbijo (267 din na 100 dinarjev prometa z živili), najmanjši pa za Črno goro (120 din na 100 din prometa z živili). Razlike izvirajo iz različne sestave kmečkega in nekmečkega prebivalstva in različne kupne moči prebivalstva. Zato koeficient nima polne analitične vrednosti.

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

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

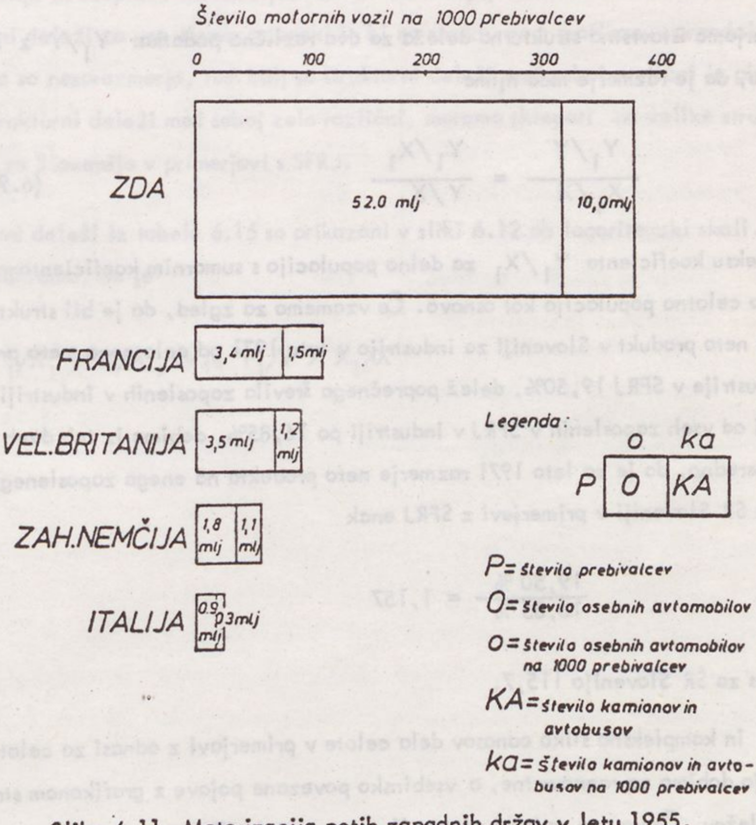
Grafično prikazovanje statističnih koeficientov

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

6.34 Uporabnost teh zvez bomo videli na praktičnem zgledu. Za pet zapadnih držav imamo v tabeli 6.14 podatke o številu prebivalstva in številu osebnih in tovornih avtomobilov in avtobusov. Iz teh podatkov moremo za vsako državo posebej izračunati koeficiente o številu motornih vozil na 1000 prebivalcev - koeficiente motorizacije. Ti podatki so po zgornjem načinu prikazani v sliki 6.11.

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

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



Slika 6.11. Motorizacija petih zapadnih držav v letu 1955.

V grafikonu nazorno primerjamo tri vrste podatkov hkrati. Ker je najvažnejša 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 sorazmerju s strukturo po vrsti motornega vozila.

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

6.35 Grafikoni strukturnih deležev. Posredno moremo koeficiente odbrati tudi iz grafikona strukturnih vrst. Ta metoda dá kompleksno sliko odnosov med pojavi, ki jih proučujemo. Zaradi tega je tak grafikon predvsem sredstvo za prikazovanje relativnih odnosov med raznovrstnimi podatki, čeprav prikazuje v osnovi strukture.

Če primerjamo istovrstna strukturna deleža za dva različna podatka: Y_1/Y z X_1/X, spoznamo, da je razmerje med njima

$$\frac{Y_1/Y}{X_1/X} = \frac{Y_1/X_1}{Y/X} \tag{6.9}$$

enak indeksu koeficienta Y_1/X_1 za delno populacijo s sumarnim koeficientom Y/X za celotno populacijo kot osnovo. Če vzamemo za zgled, da je bil strukturni delež za neto produkt v Sloveniji za industrijo v letu 1971 od celotnega neto produkta iz industrije v SFRJ 19,50%, delež poprečnega števila zaposlenih v industriji v Sloveniji od vseh zaposlenih v SFRJ v industriji pa 16,85%, dobimo iz teh dveh podatkov neposredno, da je za leto 1971 razmerje neto produkta na enega zaposlenega v industriji v SR Sloveniji v primerjavi z SFRJ enak

$$\frac{19{,}50\,\%}{16{,}85\,\%} = 1{,}157$$

ali indeks za SR Slovenijo 115,7.

Nazorno in kompleksno sliko odnosov dela celote v primerjavi z odnosi za celotno populacijo dobimo za raznovrstne, a vsebinsko povezane pojave z grafikonom strukturnih deležev. Čeprav so podatki po pravilu raznovrstni, jih v tem grafikonu narišemo brez težav v isti grafikon, ker so prikazani z neimenovanimi števili – strukturnimi deleži. Da so vidna relativna razmerja, uporabimo za njihov prikaz logaritemske skale. Vrsto strukturnih deležev za isto podpopulacijo moremo risati bodisi z vrstami stolpcev, ali še nazorneje, v grafikonu, v katerem ob logaritmični skali strukturnih deležev na ustreznih mestih vnesemo strukturne deleže s črtico, puščico itd.

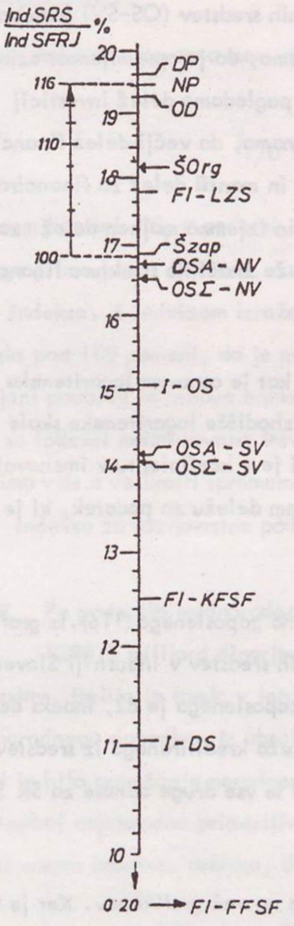
6.36 Za zgled vzemimo osnovne gospodarske pokazovalce za industrijo v Sloveniji v primerjavi s SFRJ v letu 1970. V tabeli 6.15 so po elementih podani za SR Slovenijo in SFRJ glavni absolutni podatki, v zadnjem stolpcu pa strukturni odstotki za SR Slovenijo. Če bi bila razmerja med posameznimi elementi npr. med neto produktom in številom zaposlenih, ki pokaže produktivnost, med sedanjo in nabavno vrednostjo osnovnih sredstev, ki pokaže izrabljenost osnovnih sredstev, med osebnimi dohodki in številom zaposlenih, ki pokaže raven poprečnih osebnih dohodkov; med viri finansiranja in skupnimi investicijami za SR Slovenijo, enaka kot za SFRJ, bi bili strukturni deleži za vse elemente enaki in bi na strukturnem grafikonu sovpadali. Čim večja pa so nesorazmerja, tem bolj so strukturni deleži med seboj različni in obratno, če so strukturni deleži med seboj zelo različni, moremo sklepati na velike strukturne razlike za Slovenijo v primerjavi s SFRJ.

Strukturni deleži iz tabele 6.15 so prikazani v sliki 6.12 ob logaritemski skali. Iz slike odberemo, da je

$$Y_1/X_1 > Y/X \text{ , če je } Y_1/Y > X_1/X$$

Tabela 6.15 Primerjava strukturnih deležev za industrijo SR Slovenije v industriji SFRJ za leto 1970. (Vir: SG-1972)

Pokazovalec	v milj. din SRS	v milj. din SFRJ	SRS/SFRJ %
Število industrijskih organizacij	426	2374	18,15
Število zaposlenih	244636	1451745	16,85
Osnovna sredstva 31.12.1970			
Nabavna vrednost	18150	110450	16,43
Sedanja vrednost	9437	67136	14,06
Aktivna osnovna sredstva 31.12.1970			
Nabavna vrednost	17091	102385	16,69
Sedanja vrednost	8521	60192	14,16
Družbeni proizvod	10848	55285	19,62
Neto produkt	9388	48147	19,50
Osebni dohodki (neto)	4076	21154	19,27
Investicije v osnovna sredstva	2254	14985	15,04
Investicije v družbeni standard	82	741	11,05
Vir financiranja investicij			
Lastna in združena sredstva	1314	7396	17,77
Kreditirano iz sredstev in fondov federacije	938	7496	12,51
Financirano iz sredstev in fondov federacije	2	93	2,15



DP	=	družbeni produkt
NP	=	neto produkt
OD	=	osebni dohodek (neto)
ŠOrg	=	število organizacij
Šzap	=	število zaposlenih
OS	=	osnovna sredstva - skupno
OSA	=	osnovna sredstva - aktivna
NV	=	nabavna vrednost
SV	=	sedanja vrednost
I-OS	=	investicije v osnovna sredstva
I-DS	=	investicije v družbeni standard
FI-LZS	=	financiranje investicij: iz lastnih in združenih sredstev
FI-KFSF	=	financiranje investicij s krediti sredstev in fondi federacije
FI-FFSF	=	financiranje investicij: financirano s sredstvi in fondi federacije

Slika 6.12. Strukturni grafikon za odnose v industriji SR Slovenije v letu 1970

Iz te zveze moremo iz grafikona neposredno odčitati: družbeni produkt (DP) in neto produkt (NP) na osebni dohodek (OD) je v SR Sloveniji malenkostno večji kot v poprečju za SFRJ. Obenem pa je osebni dohodek (OD) na zaposlenega (Z) v industriji precej večji v Sloveniji kot je jugoslovansko poprečje. Zaradi prvih dveh zvez je to izraz višje produktivnosti dela. Ker je strukturni delež za skupna (OS Σ -SV) in aktivna osnovna sredstva (OSA-SV) po sedanji vrednosti znatno nižji kot po nabavni vrednosti (OS Σ -NV) (OSA-NV), moremo računati, da je izrabljenost osnovnih sredstev v Sloveniji znatno večja kot v poprečju za vso državo. Ker so deleži osnovnih sredstev (OS-SV) (posebno sedanja vrednost) nižja od številnih zaposlenih (Z) sklepamo, da je opremljenost oziroma sestav v Sloveniji pod jugoslovanskim poprečjem. Če pogledamo delež investicij v osnovna sredstva (I-OS) in strukturo financiranja, spoznamo, da večji delež financiranja investicij iz lastnih in združenih sredstev (FI-LZS) in manjši delež za financiranje iz kreditov sredstev in fondov federacije (FI-KFSF) in izjemno majhen delež za financiranje iz federalnih sredstev in fondov (FI-FFSF) kaže značilno strukturo financiranja investicij v industriji v SR Sloveniji.

S premakljivo logaritemsko skalo, ki je v enakem merilu kot je osnovna logaritemska skala, bi mogli nakazane odnose izraziti z indeksi. Če izhodišče logaritemske skale (1 ali 100) naravnamo v grafikonu na strukturni delež, ki je v koeficientu v imenovalcu, odberemo indeks ustreznega koeficienta ob strukturnem deležu za podatek, ki je v koeficientu v števcu.

V grafikonu je za zgled nakazan indeks za neto produkt na zaposlenega (116). Iz grafikona odberemo podobno, da je indeks izrabljenosti osnovnih sredstev v industriji Slovenije 118 v primerjavi s SFRJ. Indeks osnovnih sredstev na zaposlenega je 82, indeks deleža financiranja investicij iz lastnih sredstev je 118, deleža kreditiranega iz sredstev in fondov federacije pa 83. Podobno bi mogli analizirati še vse druge odnose za SR Slovenijo v primerjavi s SFRJ.

Uporabnost grafikona moremo razširiti iz enega grafikona na več grafikonov. Ker je en grafikon za en del narisan ob eni premici, moremo nazorno prikazati odnose za vse republike hkrati za eno panogo ali za različne panoge po republikah. Časovna vrsta strukturnih grafikonov pokaže spreminjanje v indeksih odnosov ipd.

ENOSTAVNI INDEKSI

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

Indeks računamo po osnovnem obrazcu

$$I_{1/0} = 100 \cdot Y_1/Y \qquad (6.10)$$

Pri tem pomeni: Y_1 = podatek, ki ga primerjamo s podatkom Y_o. Y_o = podatek, na katerega primerjamo. Podatek, na katerega primerjamo, imenujemo b a z o ali o s n o v o indeksa. Z indeksom izražamo tekoči podatek v stotinkah ali poenih od osnove Y_o. Indeks pod 100 pomeni, da je pojav manjši od osnove, indeks 100 pomeni, da sta primerjani podatek in osnova enaka. Indeks pa je nad 100, če je podatek večji od osnove. Ker so indeksi neimenovana števila in imajo enoten merski sistem, iz njih zelo nazorno dobimo vtis o velikosti sprememb oziroma razlik. Razen tega moremo med seboj primerjati indekse za raznovrstne podatke, ki drugače ne bi bili neposredno primerljivi.

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

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

za Italijo pa

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

Indeksa pa sta primerljiva in sicer kažeta, da je bil porast narodnega dohodka relativno večji v Jugoslaviji kakor v Italiji.

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

Stvarni in krajevni indeksi

6.39 Ker se podatka, ki smo ju primerjali med seboj v zgornjem zgledu, razlikujeta v času (leto 1955 in leto 1956), imenujemo take indekse časovne indekse. Primerjana podatka pa se moreta razlikovati tudi v krajevnem ali stvarnem znaku. V tem primeru računamo krajevne in stvarne indekse. Običajno računamo indekse za cele vrste podatkov. V takih primerih vzamemo običajno za bazo vseh indeksov en in isti člen.

6.40 Kot zgled za vrsto krajevnih indeksov vzemimo poprečno pogodbeno prodajno ceno za m^2 stanovanjske površine za glavna mesta republik in pokrajin v letu 1971.

Tabela 6.16. Poprečna pogodbena prodajna cena za 1 m^2 (Vir: SG-72)

Mesto	Poprečna pogodbena prodajna cena za 1 m^2 v letu 1971	
	din	indeks
Beograd	2903	84
Ljubljana	3474	100
Sarajevo	3145	91
Skopje	2050	59
Titograd	2422	70
Zagreb	3877	112

Za osnovo vzamemo člen, za katerega najbolje poznamo razmere. Ker kot Slovenci najbolje poznamo razmere v Ljubljani, zato vzamemo za osnovo primerjave Ljubljano in nanjo izračunamo vse indekse. Iz tabele 6.16 vidimo, da je stanovanjska površina le v Zagrebu višja od Ljubljane, da pa je indeks za Skopje najnižji (59). Pri proučitvi teh podatkov pa moramo upoštevati ne le razlike v stroških temveč tudi v stanovanjskem standardu.

6.41 Za zgled indeksov iz vrste za stvarni znak vzemimo poprečne mesečne osebne dohodke v gospodarskih in negospodarskih dejavnostih v letu 1969 po stopnji strokovnosti.

Tabela 6.17 Poprečni neto osebni dohodki po stopnji strokovnosti v gospodarstvu in negospodarstvu (Ø 1969) (Vir: SG-72)

Stopnja strokovnosti	Gospodarstvo	Negospodarstvo	Indeks Negospodarstvo / gospodarstvo
Skupno	995	1211	121,7
Visoka strokovna izobrazba	1943	1963	101,0
Višja strokovna izobrazba	1506	1269	84,3
Srednja strokovna izobrazba	1122	1106	98,6
Nižja strokovna izobrazba	881	856	97,2
Visokokvalificiran delavec	1278	1246	97,5
Kvalificiran delavec	930	998	107,3
Polkvalificiran delavec	784	703	89,7
Nekvalificiran delavec	720	641	89,0

Indeksi nazorneje kakor pa absolutni podatki v enem številu pokažejo razmerje med gospodarstvom in negospodarstvom. Indeks za kvalificiranega delavca pokaže za 7.3 poena višjo raven v negospodarstvu v razmerju z gospodarstvom. Enako je indeks večji od 100 za zaposlene z visoko strokovno izobrazbo, kar je razumljivo. Za vse druge vrste delavcev pa so indeksi manjši kot 100 in sicer so največje razlike pri zaposlenih z višjo strokovno izobrazbo. Zaradi različne sestave skupno zaposlenih v gospodarstvu v primerjavi z negospodarstvom pa je indeks izredno velik 121,7.

Časovni indeksi

6.42 Indeksi s stalno osnovo. Največkrat računamo indekse za časovne vrste in z njimi proučujemo dinamiko pojavov. Časovna vrsta osnovnih podatkov sicer že sama prikazuje dinamiko, vendar ji razmeroma težko sledimo, ker so posamezni pojavi različno veliki, imajo različne enote mere itd. Preračunanje take vrste v indeksno vrsto s pravilno izbrano bazo pa pokaže spremembe od osnove, ki je v vsakem primeru 100, v stalnem merilu - stotinkah.

6.43 Za zgled časovnih indeksov vzemimo časovne vrste v razdobju 1965-71 za nekaj pomembnih podatkov za SFRJ.

Tabela 6.18 Razvoj za nekaj pojavov za SFRJ v razdobju 1965-71 (Vir: SG-72).

Leto	ŠP	ZDS	ND	EE	PR	NF	S
1965	19434	3583	84476	15523	29957	2063	1769
1966	19644	3491	91733	17174	29292	2222	1867
1967	19840	3466	94014	18702	26467	2374	1832
1968	20029	3487	97692	20641	26732	2494	1924
1969	20209	3622	107856	23375	26496	2699	2047
1970	20371	3765	114269	26023	28422	2854	2078
1971	20554	3944	123598	29509	30902	2961	2206

ŠP = število prebivalcev v 10^3

ZDS = zaposleni v družbenem sektorju v 10^3

ND = narodni dohodek v 10^6 po stalnih cenah v letu 1966

EE = skupna proizvodnja elektro-energije v 10^6 KWh

PR = skupna proizvodnja premoga v 10^3 ton

NF = proizvodnja surove nafte v 10^3 ton

S = surovo jeklo SM in EL v 10^3 ton

Tako enote mere kot ravni, na katerih so posamezni pojavi, otežujejo, če že ne onemogočajo primerjavo v dinamiki med posameznimi pojavi, zato izračunajmo vrste časovnih indeksov in vzemimo leto 1965 (prvo leto gospodarske reforme) za osnovo.

V tabeli 6.19 so iz absolutnih podatkov izračunane vrste indeksov. Ti indeksi so med seboj primerljivi, ker so vsi izračunani na isto osnovo in so neimenovana števila. Iz njih sklepamo, kako je npr. rastel narodni dohodek znatno hitreje kakor število prebivalstva, da je od indeksov proizvodnje energetskih sredstev proizvodnja premoga stag-

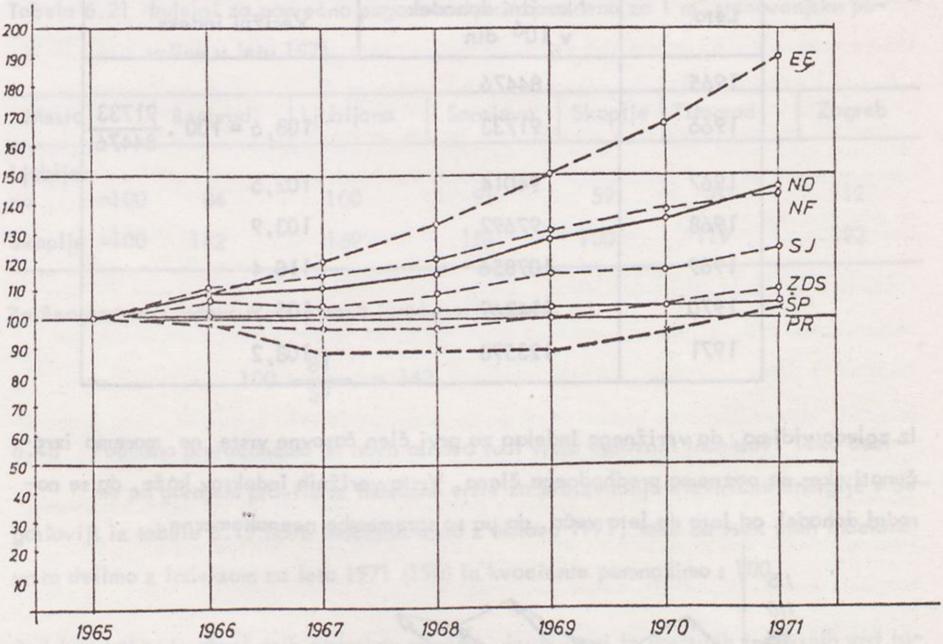
Tabela 6.19. Indeksi razvoja za nekaj pojavov za SFRJ v razdobju 1965-71

Leto	ŠP	ZDS	ND	EE	PR	NF	SJ
1965	100	100	100	100	100	100	100
1966	101	97	109	111	98	108	106
1967	102	97	111	120	88	115	104
1968	103	97	116	133	89	121	109
1969	104	101	128	151	88	131	116
1970	105	105	135	168	95	138	117
1971	106	110	146	190	103	144	125

nirala, medtem ko kažejo indeksi za elektroenergijo močan vzpon.

Indeksne vrste iz tabele 6.19 so še nazorneje prikazane v sliki 6.13.

6.44 Indeksi s premično osnovo. V gornjem primeru smo za posamezne pojave primerjali podatek za vsako leto s podatkom za leto leta 1965. Take indeksne vrste imenujemo indeksne vrste s stalno osnovo ali bazo.



Slika 6.13 Indeksi razvoja za nekaj pojavov iz gospodarstva za SFRJ v razdobju 1965-71

Večkrat pa računamo indekse tudi tako, da v isti časovni vrsti menjamo osnovo ali bazo primerjave. Take indeksne vrste imenujemo indeksne vrste s **premično bazo**. Izmed indeksnih vrst s premično bazo najpogosteje uporabljamo **verižne indekse**. Za dano časovno vrsto računamo vrsto verižnih indeksov tako, da za vsak podatek, za katerega računamo verižni indeks, vzamemo za osnovo predhodni člen. Tako dobimo vrsto verižnih indeksov, ki pokažejo relativne spremembe od člena do člena. Verižne indekse torej računamo po splošnem obrazcu

$$I_k = 100 \cdot Y_k / Y_{k-1} \qquad (6.11)$$

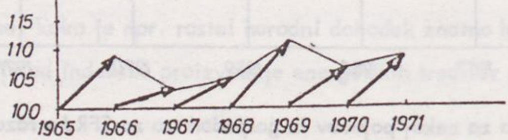
Pri tem pomeni: Y_k = tekoči podatek: Y_{k-1} = podatek za predhodni člen; I_k = verižni indeks.

6.45 Če vzamemo iz tabele 6.18 podatke o narodnem dohodku v SFRJ v letih 1965-1971, je vrsta verižnih indeksov takale:

Tabela 6.20 Verižni indeksi za narodni dohodek v SFRJ v razdobju 1965-1971

Leto	Narodni dohodek v 10^6 din	Verižni indeks
1965	84476	
1966	91733	$108,6 = 100 \cdot \frac{91733}{84476}$
1967	94014	102,5
1968	97692	103,9
1969	107856	110,4
1970	114269	105,9
1971	123598	108,2

Iz zgleda vidimo, da verižnega indeksa za prvi člen časovne vrste ne moremo izračunati, ker ne poznamo predhodnega člena. Vrsta verižnih indeksov kaže, da se narodni dohodek od leta do leta veča, da pa so spremembe neenakomerne.



Slika 6.14 Verižni indeksi za ND v SFRJ

6.46 **Preračunavanje indeksov na drugo osnovo.** V navedenih primerih smo izračunali indekse iz vrst osnovnih podatkov. Dostikrat pa imamo indeksne vrste, iz katerih želimo izračunati indeksne vrste z novo osnovo, nimamo pa osnovnih podatkov. V takih primerih preračunamo indeksno vrsto v indeksno vrsto z novo bazo 1 tako kakor iz osnovne vrste

$$I_{2/1} = 100 \frac{I_{2/0}}{I_{1/0}} \tag{6.12}$$

Velja namreč

$$100 \cdot \frac{I_{2/0}}{I_{1/0}} = 100 \cdot \frac{100 . Y_2/Y_0}{100 . Y_1/Y_0} = 100 . Y_2/Y_1 = I_{2/1}$$

6.47 Vzemimo za zgled iz tabele 6.16 indeksno vrsto za poprečno pogodbeno prodajno ceno za 1 m^2. Osnova te indeksne vrste je Ljubljana. Želimo pa izračunati novo indeksno vrsto, v kateri je osnova mesto z najnižjim indeksom (Skoplje). Po obrazcu 6.12 indeks za vsako mesto primerjamo z bazičnim indeksom (Skoplje 59).

Tabela 6.21 Indeksi za poprečno pogodbeno prodajno ceno za 1 m^2 stanovanjske površine v letu 1971

Mesto		Beograd	Ljubljana	Sarajevo	Skoplje	Titograd	Zagreb
Ljubljana	=100	84	100	91	59	70	112
Skoplje	=100	142	169	154	100	119	192

Za Beograd smo indeks z novo bazo izračunali takole:

$$100 \frac{84}{59} = 142$$

6.48 Podobno preračunamo na novo osnovo tudi vrste časovnih indeksov. Tako dobimo po gornjem pravilu iz indeksne vrste za proizvodnjo električne energije v Jugoslaviji iz tabele 6.19 novo indeksno vrsto z osnovo 1971, tako da vsak člen indeksne vrste delimo z indeksom za leto 1971 (190) in kvociente pomnožimo s 100.

Podobno kakor iz absolutnih podatkov moremo izračunati iz časovnih indeksnih vrst tudi vrste verižnih indeksov. Iz vrste verižnih indeksov pa dobimo indeksne vrste s poljubno stalno osnovo s postopnim množenjem oziroma deljenjem verižnih indeksov.

Tabela 6.22 Indeksna vrsta za proizvodnjo električne energije v Jugoslaviji v razdobju 1965-1971.

L e t o	1965	1966	1967	1968	1969	1970	1971
1965 = 100	100	111	120	133	151	168	190
1971 = 100	53	58	63	70	79	88	100
verižni indeks	.	111	108	111	114	111	113

6.49 Izbira baze ali osnove. Formalno moremo vzeti za osnovo pri računanju indeksov kateri koli člen v vrsti, za katero računamo indekse, ali tudi vrednost, ki je izven nje. Vsebinsko pa je odločitev o osnovi primerjave odvisna od namena, primernosti in smiselnosti primerjave. Zaradi tega je za izbiro baze nemogoče dati enotno pravilo, marveč samo nekaj splošnih načel, ki pomagajo pri pravilni izbiri baze.

Pri stvarnih in geografskih indeksih vzamemo za osnovo po pravilu pojav oziroma območje, ki si ga najbolje predstavljamo oziroma ga najbolje poznamo. Tako v medrepubliški primerjavi vzamemo npr. za osnovo SR Slovenijo, v meddržavni primerjavi SFRJ itd. Če primerjamo z indeksi relativna števila ali druge izvedene pokazovalce za posamezne dele populacije, je najprimerneje in najenostavneje, da vzamemo za bazo sumarni pokazovalec za celoto. Taki indeksi pokažejo odklone od nekega povprečnega stanja.

Poseben problem je izbira osnove pri časovnih indeksih. Za časovne indekse velja splošno pravilo, da vzamemo za osnovo čas, ko je pojav normalen in ustaljen. Kdaj pa je pojav normalen, je težko določiti. Vendar moremo iz tega pravila vsaj sklepati, katere člene ne smemo vzeti za osnovo, ker je laže ugotoviti, kdaj pojav ni normalen. Zaradi tega za primerjavo večine pojavov ne vzamemo vojna leta, leta gospodarskih kriz in tako dalje. Enako v zdravstveni statistiki ne vzamemo za osnovo leta epidemij. Večja verjetnost je, da je pojav normalen v daljšem razdobju kot v krajšem. Zaradi tega običajno ne jemljemo za osnovo kratka razdobja, npr. mesece pri indeksih proizvodnje, marveč vzamemo za osnovo leto. Težko je tudi npr. reči v kmetijstvu, katero leto je normalno, ali je to leto ugodne ali neugodne vegetacije.

Zato v kmetijstvu dostikrat vzamemo za osnovo večletno (petletno ali desetletno) poprečje.

Povojni razvoj običajno primerjamo s stanjem pred vojno. Za osnovo primerjave vzamemo čim kasnejše leto pred vojno, vendar tako, da nanj še ne vplivajo priprave na vojno. V SFRJ vzamemo za primerjavo s predvojnim stanjem za bazo leto 1939, OZN pa leto 1937. Vendar moramo paziti, kdaj so primerjave pred in povojnega stanja vsebinsko utemeljene in smiselne. Za zelo dolga razdobja in v primeru velikih sprememb primerjava na določeno staro stanje nima smisla.

Če je bila nekaj let po drugi svetovni vojni primerjava s predvojnim stanjem upravičena, ker smo tako približno dobili vtis o pojavih in spremembah, je primerjava s predvojnim stanjem sedaj nepomembna, ker nas zanima povojni razvoj, ne pa primerjava s predvojnim stanjem, ki je že razmeroma zelo odmaknjeno. Vendar ne vzamemo za osnovo povojnega razvoja prvo povojno leto, ko so bile razmere še neustaljene, temveč kasnejše razdobje normaliziranega stanja.

Včasih vzamemo za osnovo indeksov tudi zadnje - tekoče leto. Taka indeksna vrsta primerja preteklo stanje s podatkom, ki je časovno najbližji in tudi najzanimivejši, ker predstavlja trenutno stanje.

Velikih sprememb tudi ne kaže prikazovati z indeksi. Tako nima smisla npr. računati indeksa proizvodnje valjanih aluminijevih proizvodov na bazo 1939, ko je bila proizvodnja 15 ton, s proizvodnjo v letu 1957, ko smo proizvedli 4527 ton valjanih aluminijevih proizvodov. Indeks, ki ga izračunamo iz zgornjih podatkov, je 30180 in nenazoren. Kljub temu, da indeksi v splošnem izboljšajo primerjavo, ta indeks nima smisla, ker je nepredstavljiv.

Zato v kmetijstvu dostikrat vzamemo za osnovo večletno (petletno ali desetletno) povprečje.

Povojni razvoj običajno primerjamo s stanjem pred vojno. Za osnove primerjave vzamemo čim kasnejše leto pred vojno, vendar tako, da nanj še ne vplivajo priprave na vojno. V SFRJ vzamemo za primerjave s predvojnim stanjem za bazo leto 1939, OZN pa leto 1937. Vendar moramo paziti, kadar so primerjave pred in povojnega stanja vsebinsko utemeljene in smiselne. Za zelo dolgo razdobje in v primeru velikih sprememb primerjave na določeno staro stanje nima smisla.

Če je bilo nekaj let po drugi svetovni vojni primerjave s predvojnim stanjem upravičeno, ker smo tako približno dobili vtis o pojavih in spremembah, je primerjava s predvojnim stanjem sedaj nepomembna, ker nas zanima povojni razvoj, ne pa primerjava s predvojnim stanjem, ki je že razmeroma zelo odmaknjeno. Vendar ne vzamemo za osnovo povojnega razvoja prvo povojno leto, ko so bile razmere še neustaljene, temveč kasnejše razdobje normaliziranega stanja.

Včasih vzamemo za osnovo indeksov tudi zadnje – tekoče leto. Tako indeksne vrste primerjajo pretekla stanja s podatkom, ki je časovno najbližji in tudi najzanimivejši, ker predstavlja trenutno stanje.

Velikih sprememb tudi ne kaže prikazovati z indeksi? Tako nima smisla npr. računati indeksa proizvodnje valjanih aluminijevih proizvodov na bazo 1939, ko je bila proizvodnja 13 ton, s proizvodnjo v letu 1957, ko smo proizvedli 4327 ton valjanih aluminijevih proizvodov. Indeks, ki ga izračunamo iz zgornjih podatkov, je 33280 [illegible]. Kljub temu, da indeksi v splošnem izboljšajo primerjavo, ta indeks nima smisla, ker je nepredstavljiv.

SEDMO POGLAVJE

7. FREKVENČNE PORAZDELITVE

SESTAVLJANJE FREKVENČNE PORAZDELITVE

7.1 Med statističnimi vrstami imajo posebno mesto vrste, ki prikazujejo razporeditev vrednosti za numerične znake. Take vrste imenujemo **frekvenčne porazdelitve** ali **pogostnostne porazdelitve**. Za dano populacijo dobimo frekvenčno porazdelitev, če za posamezne razrede za proučevani znak poiščemo, koliko enot ima vrednost znaka v ustreznem razredu.

Vzemimo za zgled porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v takratnem okraju Novo mesto. Ta gospodarstva so bila anketirana v anketi o porabi lesa leta 1953. Osnovni podatki o porabi lesa v posameznih anketiranih gospodarstvih so navedeni v tabeli 7.1.

Tabela 7.1 Poraba lesa v m^3 v 149 kmetijskih gospodarstvih v letu 1953 v okraju Novo mesto. (Vir: Anketa o porabi lesa v letu 1953).

11,6	8,0	8,5	12,2	11,6	11,8	10,2	15,1	2,1	18,5	7,2	10,1
12,7	18,0	14,4	16,2	20,5	20,1	14,6	15,7	23,5	12,2	9,0	12,7
19,4	6,5	6,0	9,8	8,5	5,0	3,6	10,1	10,3	8,6	8,4	5,5
6,0	13,9	12,8	12,4	10,9	15,7	13,6	13,4	12,8	19,7	14,3	7,5
16,0	9,0	9,4	14,2	11,0	10,5	5,9	29,4	14,0	17,5	10,0	19,5
24,4	12,4	17,3	11,3	11,8	26,7	16,5	16,9	12,1	17,0	26,7	24,4
20,1	10,9	13,9	10,9	10,3	7,6	10,3	21,7	17,7	17,4	10,8	15,1
15,2	23,2	21,8	16,2	19,2	22,7	19,0	13,7	14,3	24,9	23,8	19,9
12,1	31,6	29,9	26,5	6,7	8,2	4,9	7,9	9,4	9,3	16,2	19,6
12,9	32,4	18,3	15,2	11,3	8,3	4,7	15,6	13,3	17,3	27,5	15,1
18,4	14,8	17,1	10,5	8,4	7,5	12,2	9,7	18,1	16,9	14,0	12,1
10,5	6,2	22,7	14,3	15,7	14,6	12,0	18,9	16,1	13,2	28,6	6,2
12,2	19,8	23,4	33,5	15,8							

Uredimo zgornjo nepregledno množico podatkov v frekvenčno porazdelitev, v kateri vzemimo po 1 m^3 široke razrede! V tabeli 7.2 pomeni npr. 3 porabo od 3,0 m^3 do 3,9 m^3.

Tabela 7.2 Frekvenčna porazdelitev 149 kmetijskih gospodarstev v Novem mestu po porabi lesa v letu 1953.

Poraba lesa m^3	y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Število gospodarstev	f	-	-	1	1	2	3	6	5	8	7	14	7	14	7	10

y	15	15	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
f	10	8	7	6	8	3	3	2	4	3	-	3	1	1	2	-	1	1	1

Število enot v posameznem razredu imenujemo frekvenca. Po pravilu zaznamujemo frekvence s črko f.

Frekvenčna porazdelitev v tabeli 7.2 kaže veliko preglednejše porabo lesa v gospodarstvih kakor posamični podatki v tabeli 7.1. Vendar frekvenčna porazdelitev ne da natančnih vrednosti. Iz nje npr. vidimo, da je eno gospodarstvo porabilo od 2,0 do 2,9 m^3 lesa, ne vemo pa, koliko so natančno porabili lesa v tem gospodarstvu. Enako je iz frekvenčne porazdelitve razvidno, da je 14 gospodarstev porabilo 10,0 do 10,9 m^3, ne vemo pa natančne porabe lesa. S frekvenčno porazdelitvijo se je preglednost povečala, natančnost informacij pa zmanjšala.

Iz frekvenčne porazdelitve dobimo nekatere zelo koristne ugotovitve. Predvsem je frekvenčna porazdelitev slika o variiranju znaka. Iz tabele 7.2 vidimo, da je bila npr. poraba lesa v kmetijskih gospodarstvih od 2 m^3 do 34 m^3. Razen tega si iz frekvenčne porazdelitve zlahka ustvarimo sliko o porabi lesa znotraj teh meja. Bežen pogled pokaže, da je gospodarstev z majhno porabo malo, medtem ko se število gospodarstev v posameznih razredih do neke porabe v glavnem veča, za večjo porabo pa so frekvence vedno manjše. Frekvenčna porazdelitev nakazuje zakonitost, ki jo opazimo pri večini populacij. V dani populaciji se vrednosti najbolj goste okrog nekega središča, od katerega so odkloni tem redkejši, čim večji so. Vendar se ta zakonitost v frekvenčni porazdelitvi v tabeli 7.2 ne kaže popolnoma. Imamo več odstopanj od tega pravila. Frekvence v sosednih razredih so včasih večje, včasih manjše. Število enot v posameznih razredih je namreč razmeroma majhno, zato prevladujejo še slučajni vplivi. Če bi bilo število enot v posameznih razredih večje, slučajni vplivi ne bi bili tako močni in bi se gornja zakonitost izražala močneje. Frekvence v razredih moremo povečati na dva načina. Če bi namesto 149 gospodarstev imeli desetkrat večjo populacijo s 1490 gospodarstvi, bi bile tako frekvence večje. Zakonitost bi se izražala močneje. Vendar tako večanje frekvenc običajno ne pride v poštev. Frekvence pa povečamo tudi, če vzamemo širše razrede. S tem se sicer zmanjša natančnost prikaza, zakonitost variiranja vrednosti pa je prikazana bolje, ker so slučajni vplivi manjši.

Če v tabeli 7.2 združimo po štiri razrede, dobimo novo frekvenčno porazdelitev; ta je prikazana v tabeli 7.3.

Tabela 7.3 Frekvenčna porazdelitev za porabo lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto.

Poraba v m^3	f
0,0 - 3,9	2
4,0 - 7,9	16
8,0 - 11,9	36
12,0 - 15,9	41
16,0 - 19,9	29
20,0 - 23,9	12
24,0 - 27,9	7
28,0 - 31,9	4
32,0 - 35,9	2
	149 = N

Frekvenčna porazdelitev v tabeli 7.3 je preglednejša, ker je število razredov manjše. Razen tega pa je zaradi razmeroma velikih frekvenc dobro vidna zakonitost gostitve. Frekvence se sprva večajo, dokler ne dosežejo največjo gostitev v razredu 12,0 - 13,9 m^3, od tu pa stalno padajo.

Seveda pa ne smejo biti razredi preširoki. S širjenjem razredov se natančnost bolj in bolj manjša. Tudi zakonitost gostitve je vedno manj vidna. Če v gornji frekvenčni porazdelitvi združimo po dva razreda, dobimo skrčeno porazdelitev, ki je prikazana v tabeli 7.4.

Tabela 7.4 Frekvenčna porazdelitev za porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

Poraba v m^3	f
0,0 - 7,9	18
8,0 - 15,9	77
16,0 - 23,9	41
24,0 - 31,9	11
32,0 - 39,9	2
	149 = N

Ta frekvenčna porazdelitev sicer še vedno nakazuje osnovno zakonitost, vendar daje pregrobo sliko, ker so v istem razredu gospodarstva, ki imajo do 8 m^3 razlik v porabi lesa.

Zato je pri sestavljanju frekvenčnih porazdelitev važno število in velikost razredov. Če vzamemo premajhne razrede, na frekvence preveč vplivajo slučajni vplivi in zabrišejo osnovno zakonitost gostitve, preveliki razredi pa dajo pregrobo sliko. Nimamo splošnega pravila, kakšne in koliko razredov naj ima frekvenčna porazdelitev, pač pa je praksa pokazala tole: razredi, ki razdele razmik med najmanjšo in največjo vrednostjo v populaciji na osem do šestnajst razredov, dajo običajno dosti dobro sliko gostitve. Število razredov se ravna po velikosti populacije. Za manjše populacije vzamemo manj, za večje pa več razredov. V našem primeru smo resnično dobili s štiridesetimi razredi nepregledno, s petimi razredi pregrobo sliko o porabi lesa, medtem ko je frekvenčna porazdelitev z desetimi razredi pokazala vse značilnosti gostitve.

7.2 Frekvenčne porazdelitve z neenakimi razredi. Čeprav je iz tehničnih razlogov najbolje, da so vsi razredi v frekvenčni porazdelitvi enako široki, včasih iz vsebinskih razlogov sestavljamo frekvenčne porazdelitve, v katerih so razredi različno široki. Vzemimo za zgled industrijska podjetja! Če ima eno podjetje 10, drugo pa 50 zaposlenih, je med njima večja vsebinska razlika kakor pa med podjetjema s 1000 in 1040 delavci, čeprav je v obeh primerih razlika 40 delavcev. Zato industrijska podjetja po številu delavstva prikazujemo v frekvenčni porazdelitvi z različnimi širinami razredov.

Tabela 7.5 Industrijska podjetja v SFRJ v letu 1970 po poprečnem številu zaposlenih (Vir: SG 72)

Število zaposlenih	f
- 15	79
16 - 29	61
30 - 60	150
61 - 125	378
126 - 250	522
251 - 500	475
501 - 1000	347
1001 - 2000	221
2001 -	141
	2374 = N

Iz tabele 7.5 vidimo, da se meje razredov vrste v približnem geometrijskem zaporedju z razmerjem $k = 2$; to pa je v skladu z relativno primerljivostjo števila delavstva po podjetjih.

7.3 V frekvenčnih porazdelitvah z enakimi razredi so razlike med frekvencami predvsem izraz različne stopnje gostitve v posameznih razredih. V frekvenčnih porazdelitvah z neenakimi razredi pa je frekvenca v danem razredu odvisna razen od gostitve tudi od širine razreda. Pri isti stopnji gostitve imajo širši razredi večjo, ožji razredi pa manjšo frekvenco. Da iz frekvence odstranimo vpliv različne širine razredov in prikažemo samo stopnjo gostitve, izračunavamo po obrazcu

$$g_k = f_k / i_k \tag{7.1}$$

gostoto frekvence g_k ; ta pokaže, koliko frekvence v razredu odpade na enotin razmik.

Za vsako frekvenčno porazdelitev moremo izračunati strukturne deleže; ti pokažejo, koliki del celotne populacije je v posameznem razredu. Strukturne deleže, ki jih dobimo, če frekvence f_k delimo z obsegom populacije N, imenujemo relativne frekvence f_k^o

$$f_k^o = f_k / N \tag{7.2}$$

Kakor iz frekvenc računamo gostoto frekvence, tako moremo po obrazcu

$$\varphi_k = g_k^o = f_k^o / i_k = f_k / N . i_k \tag{7.3}$$

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

Iz obrazca 7.3 zlahka dobimo, da je

$$f_k = N . i_k . \varphi_k \tag{7.4}$$

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

7.4 Za frekvenčno porazdelitev industrijskih podjetij v SFRJ v letu 1970 so nakazani pokazovalci dani v tabeli 7.6.

Tabela 7.6. Industrijska podjetja v SFRJ v letu 1970 po poprečnem številu zaposlenih

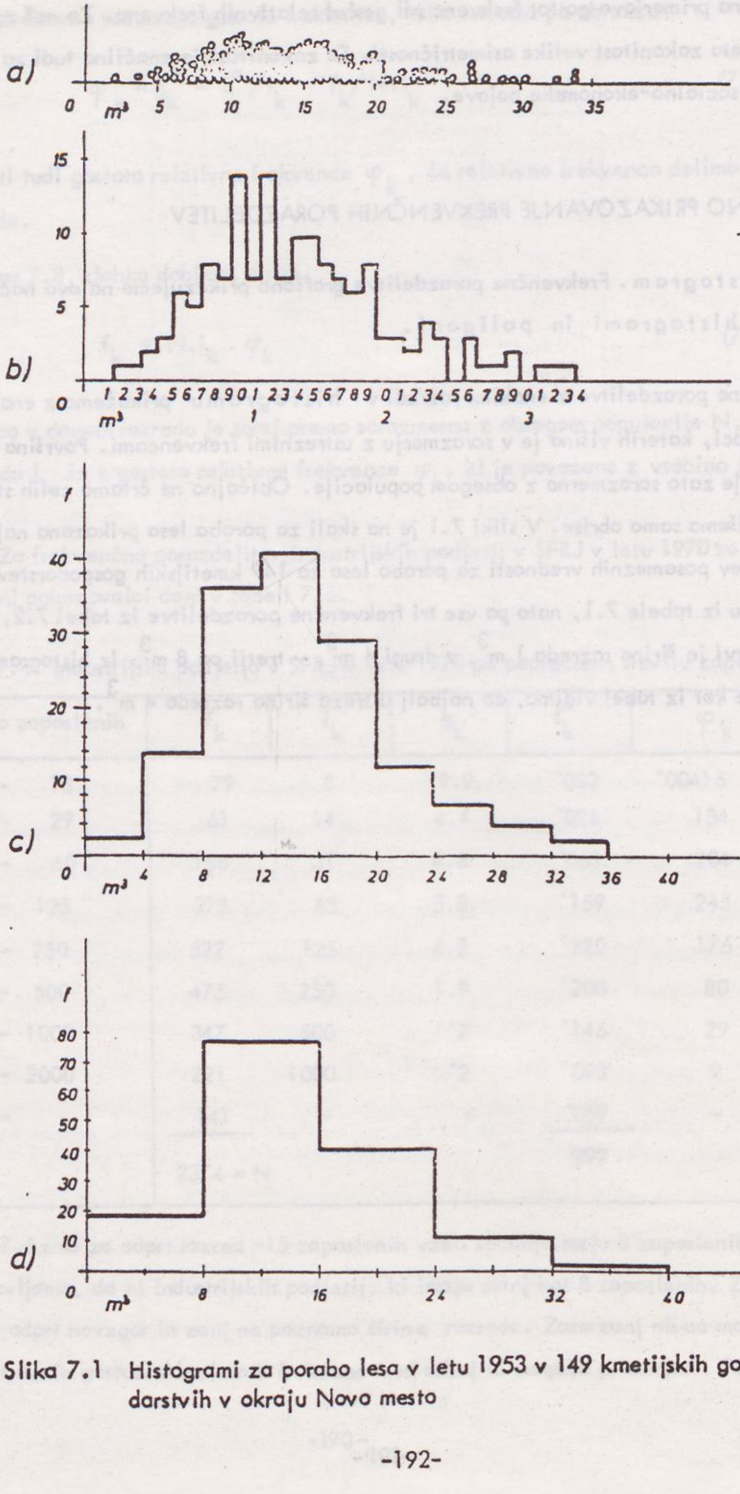
Število zaposlenih	f_k	i_k	g_k	f_k^o	φ_k
8 - 15	79	8	9.9	.033	.00416
16 - 29	61	14	4.4	.026	184
30 - 60	150	31	4.8	.063	204
61 - 125	378	65	5.8	.159	245
126 - 250	522	125	4.2	.220	176
251 - 500	475	250	1.9	.200	80
501 - 1000	347	500	.7	.146	29
1001 - 2000	221	1000	.2	.093	9
2001 -	141	-	-	.059	-
Skupaj	2374 = N			.999	

V tabeli 7.6 smo za odprt razred -15 zaposlenih vzeli spodnjo mejo 8 zaposlenih, ker predpostavljamo, da ni industrijskih podjetij, ki imajo manj kot 8 zaposlenih. Zadnji razred je odprt navzgor in zanj ne poznamo širine razreda. Zato zanj nismo mogli izračunati ustreznih gostot. Absolutnih frekvenc med seboj ni mogoče primerjati. Pač pa je smiselna primerjava gostot frekvenc ali gostot relativnih frekvenc. Za naš zgled ti vrsti kažeta zakonitost velike asimetričnosti. Ta zakonitost je značilna tudi za nekatere druge socialno-ekonomske pojave.

GRAFIČNO PRIKAZOVANJE FREKVENČNIH PORAZDELITEV

7.5 Histogram. Frekvenčne porazdelitve grafično prikazujemo na dva načina: s histogrami in poligoni.

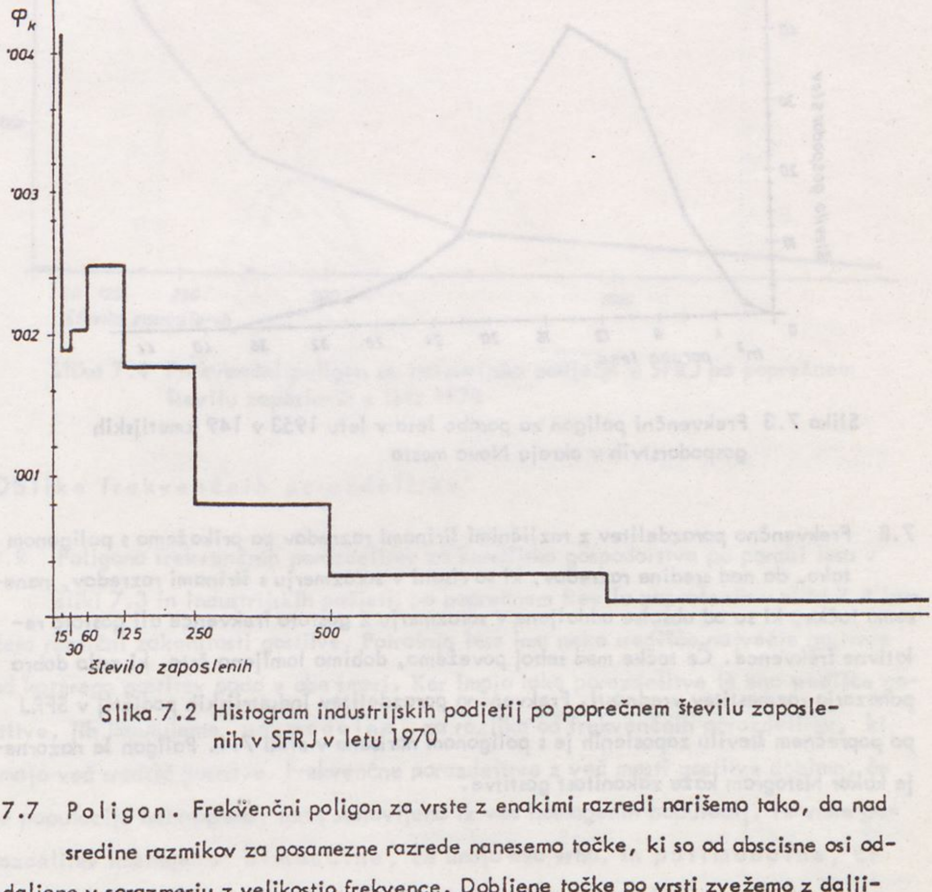
Frekvenčne porazdelitve z enakimi razredi v histogramu prikažemo z enako širokimi stolpci, katerih višina je v sorazmerju z ustreznimi frekvencami. Površina vseh stolpcev je zato sorazmerna z obsegom populacije. Običajno ne črtamo celih stolpcev, temveč rišemo samo obrise. V sliki 7.1 je na skali za porabo lesa prikazana najprej razmestitev posameznih vrednosti za porabo lesa za 149 kmetijskih gospodarstev v Novem mestu iz tabele 7.1, nato pa vse tri frekvenčne porazdelitve iz tabel 7.2, 7.3, 7.4. V prvi je širina razreda 1 m^3, v drugi 4 m^3, v tretji pa 8 m^3. Iz histogramov še nazorneje kot iz tabel vidimo, da najbolj ustreza širina razreda 4 m^3.



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

7.6 Če prikazujemo s histogramom frekvenčne porazdelitve z različnimi širinami razredov, rišemo gostoto frekvence g_k ali gostote relativnih frekvenc φ_k. Ker je po obrazcu 7.1 frekvenca zmnožek širine razreda in gostote frekvence, so ploščine stolpcev v sorazmerju z ustrezno frekvenco f_k, ploščina vseh stolpcev pa v sorazmerju z obsegom populacije N. Če pa rišemo v histogramu gostote relativnih frekvenc, so ploščine sorazmerne z relativnimi frekvencami.

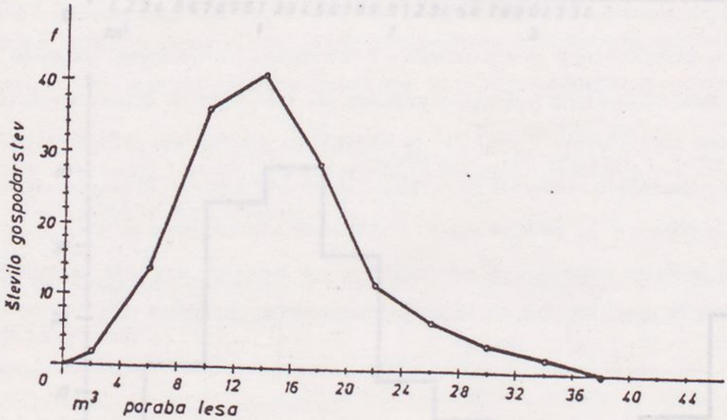
Edino tako dobimo pravilno sliko o razmestitvi pojava. V sliki 7.2 je s histogramom prikazana frekvenčna porazdelitev industrijskih podjetij v SFRJ v letu 1970 po poprečnem številu delavcev iz tabele 7.6.



Slika 7.2 Histogram industrijskih podjetij po poprečnem številu zaposlenih v SFRJ v letu 1970

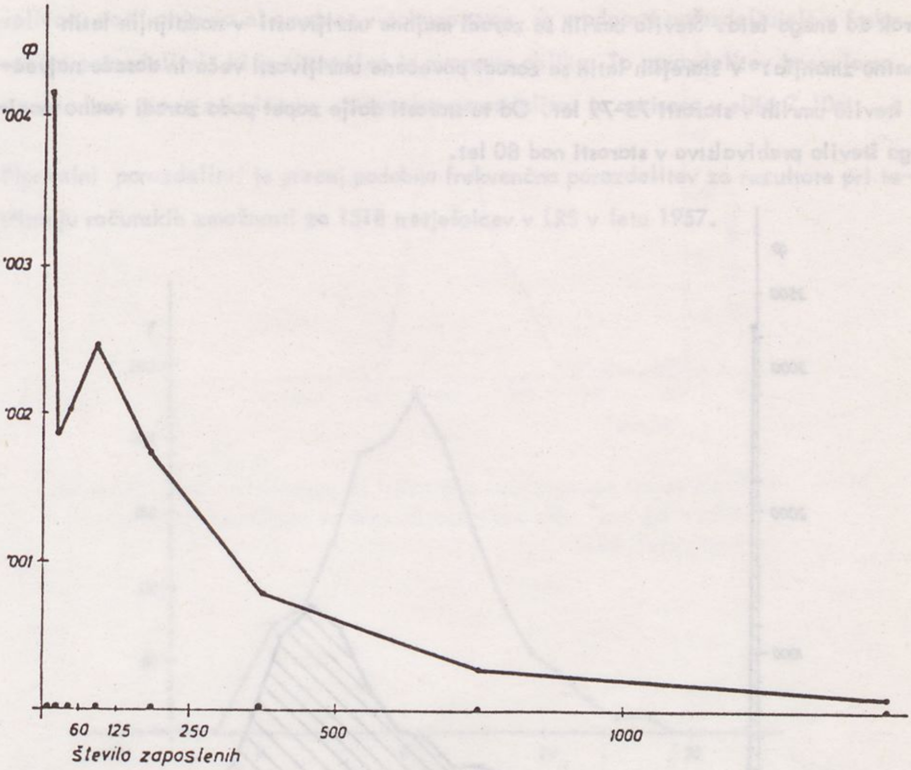
7.7 Poligon. Frekvenčni poligon za vrste z enakimi razredi narišemo tako, da nad sredine razmikov za posamezne razrede nanesemo točke, ki so od abscisne osi oddaljene v sorazmerju z velikostjo frekvence. Dobljene točke po vrsti zvežemo z daljicami. Če vzamemo, da je frekvenca v razredu pred prvim razredom in po zadnjem razredu enaka nič in da ustrezni točki ležita na abscisni osi, je poligon sklenjen. Poligon bolje prikazuje stvarno razmestitev vrednosti, ker z njim nakažemo verjetnejšo razmestitev vrednosti znotraj razredov kot s histogramom, ki nakazuje enakomerno razmestitev znotraj razredov. Logično je namreč, da je znotraj razredov na straneh, ki teže proti največji gostitvi, gostota večja kakor na zunanjih straneh razredov.

V sliki 7.3 je s poligonom prikazana frekvenčna porazdelitev o porabi lesa iz tabele 7.3. Poligon zelo lepo pokaže razmestitev gospodarstev po porabi lesa.



Slika 7.3 Frekvenčni poligon za porabo lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

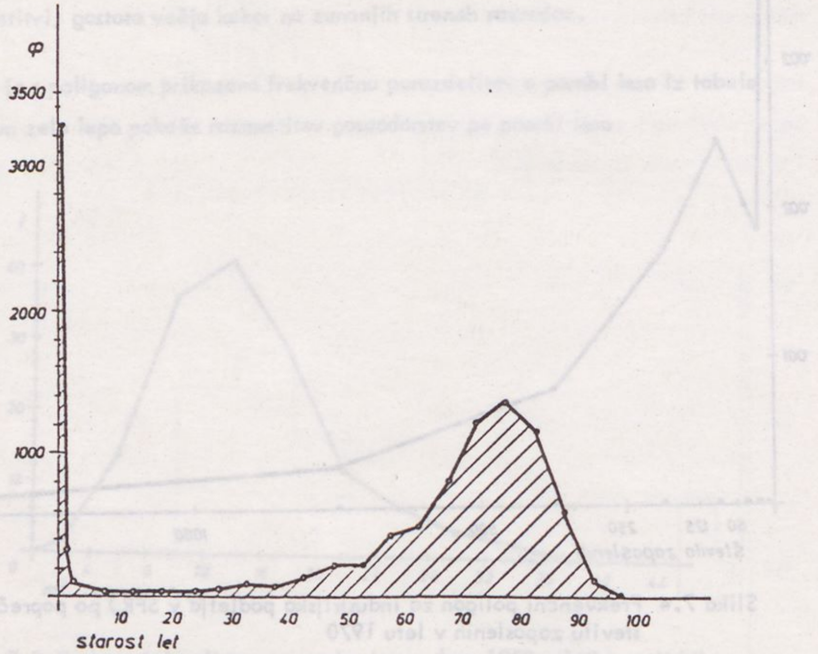
7.8 Frekvenčno porazdelitev z različnimi širinami razredov pa prikažemo s poligonom tako, da nad sredine razredov, ki so risani v sorazmerju s širinami razredov, nanesemo točke, ki so od abscise oddaljene v sorazmerju z gostoto frekvence ali gostoto relativne frekvence. Če točke med seboj povežemo, dobimo lomljeno črto, ki zelo dobro ponazarja razmestitev vrednosti. Frekvenčna porazdelitev industrijskih podjetij v SFRJ po poprečnem številu zaposlenih je s poligonom narisana v sliki 7.4. Poligon še nazorneje kakor histogram kaže zakonitost gostitve.



Slika 7.4 Frekvenčni poligon za industrijska podjetja v SFRJ po poprečnem številu zaposlenih v letu 1970

O b l i k e f r e k v e n č n i h p o r a z d e l i t e v

7.9 Poligona frekvenčnih porazdelitev za kmetijska gospodarstva po porabi lesa v sliki 7.3 in industrijskih podjetij po poprečnem številu zaposlenih v sliki 7.4 kažeta različni zakonitosti gostitve. Potrošnja lesa ima neko središče največje gostitve od katerega gostitev pada v obe smeri. Ker imajo take porazdelitve le eno središče gostitve, jih imenujemo u n i m o d a l n e , za razliko od frekvenčnih porazdelitev, ki imajo več središč gostitve. Frekvenčne porazdelitve z več mesti gostitve dobimo, če je populacija heterogena, torej sestavljena iz več homogenih populacij. Te vrste porazdelitev imenujemo b i m o d a l n e , če imajo dva vrha, in p o l i m o d a l n e , če imajo več vrhov. V sliki 7.5 je prikazana frekvenčna porazdelitev umrlih žensk po starosti v letu 1956 v Sloveniji. V grafikonu opazimo izredno veliko število umrlih otrok do enega leta. Število umrlih se zaradi majhne umrljivosti v nadaljnjih letih znatno zmanjša. V starejših letih se zaradi povečane umrljivosti veča in doseže največje število umrlih v starosti 75-79 let. Od te starosti dalje zopet pada zaradi vedno manjšega števila prebivalstva v starosti nad 80 let.

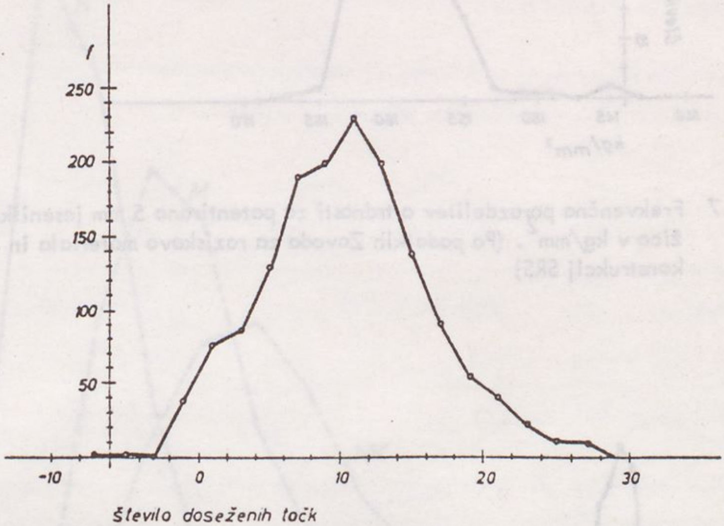


Slika 7.5 Število umrlih žensk v letu 1956 v Sloveniji po starosti ob smrti

7.10 Poligon za frekvenčno porazdelitev o porabi lesa v kmetijskih gospodarstvih kaže bolj ali manj enako upadanje gostitve od središča gostitve. Frekvenčne porazdelitve, ki kažejo enako upadanje frekvenc na levo in desno od središča gostitve, imenujemo simetrične porazdelitve, za razliko od asimetričnih porazdelitev, pri katerih je upadanje gostitve na eno ali drugo stran hitrejše. Če je upadanje gostitve počasnejše na levo od središča gostitve, pravimo, da je frekvenčna porazdelitev asimetrična v levo, če pa je upadanje gostitve počasnejše desno od središča, govorimo o asimetriji v desno. Simetrično porazdelitev dobimo navadno za homogene populacije.

Če je populacija popolnoma homogena in so razlike za enote rezultat samo slučajnih vplivov, variiranje pa ni omejeno v nobeno stran, se vrednosti porazdeljujejo v frekvenčni porazdelitvi, ki je simetrična in zvonaste oblike. To porazdelitev imenujemo normalno porazdelitev. Normalna porazdelitev je narisana v sliki 7.10a.

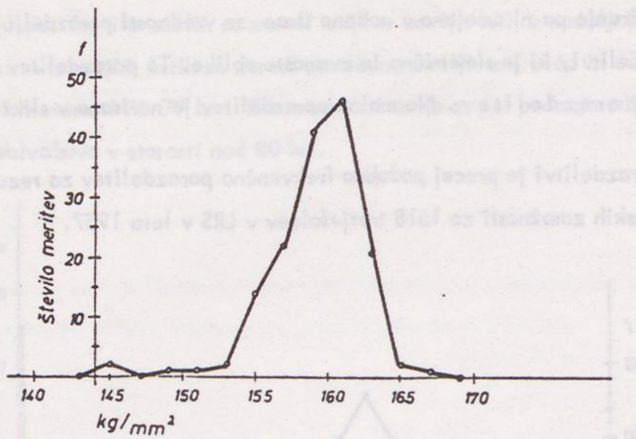
Normalni porazdelitvi je precej podobna frekvenčna porazdelitev za rezultate pri testiranju računskih zmožnosti za 1518 tretješolcev v LRS v letu 1957.



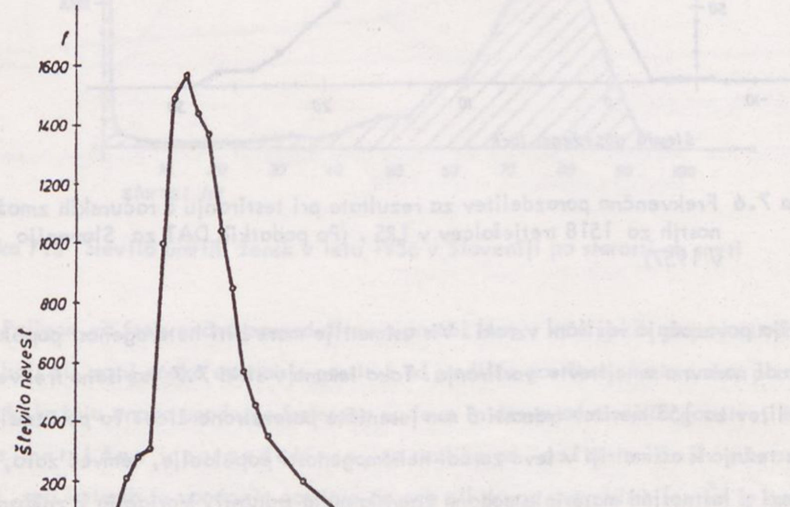
Slika 7.6 Frekvenčna porazdelitev za rezultate pri testiranju o računskih zmožnostih za 1518 tretješolcev v LRS. (Po podatkih DAT za Slovenijo v 1957)

Asimetrijo povzročajo različni vzroki. Vir asimetrije more biti heterogenost populacije ali pa tudi naravna omejitev v variiranju. Tako imamo v sliki 7.7 narisano frekvenčno porazdelitev za 153 meritev trdnosti 5 mm jeseniške patentirane žice. Ta porazdelitev ne kaže težnjo k asimetriji v levo zaradi nehomogenosti populacije, temveč zato, ker je v zvezi z lastnostmi materiala podana zgornja meja trdnosti. Variacija k zniževanju je svobodnejša in imamo zato asimetrijo v levo.

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

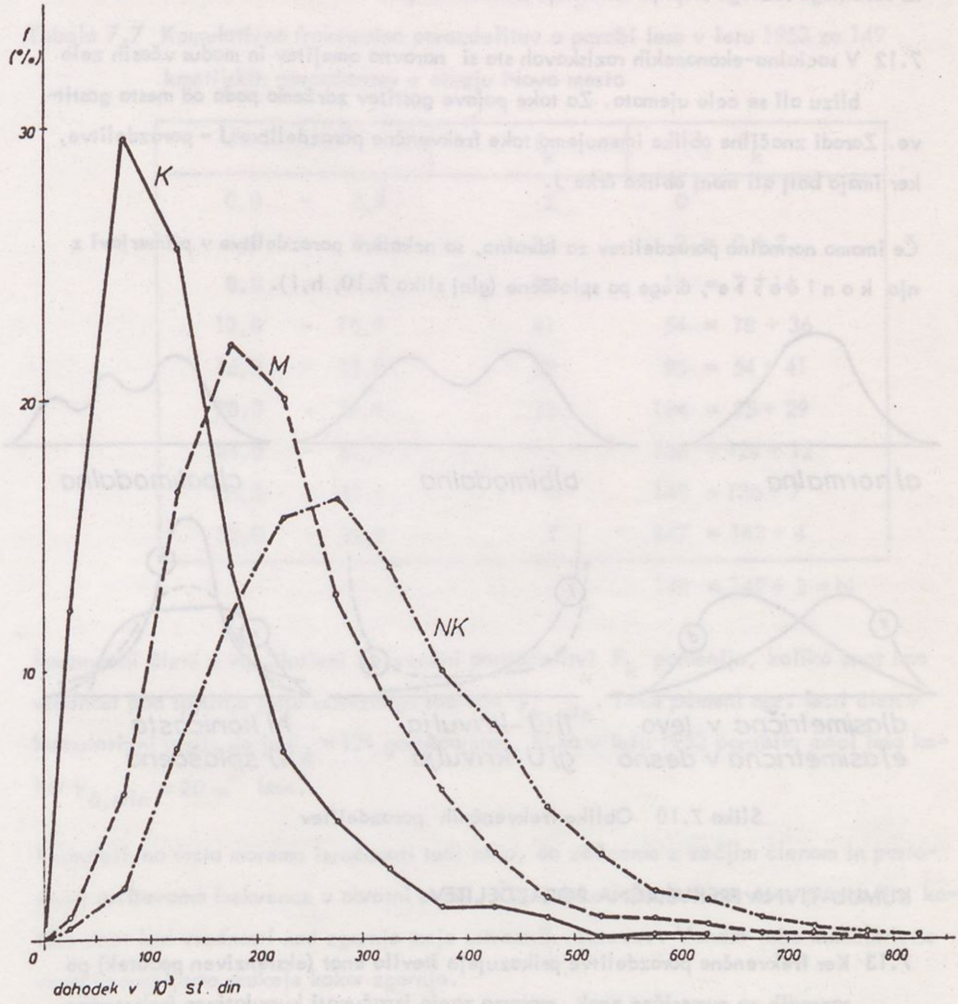


Slika 7.7 Frekvenčna porazdelitev o trdnosti za patentirano 5 mm jeseniško žico v kg/mm^2. (Po podatkih Zavoda za raziskavo materiala in konstrukcij SRS)



Slika 7.8 Frekvenčna porazdelitev za starost nevest v letu 1956 v Sloveniji. (Vir: SL LRS 57)

7.11 Kako je stopnja asimetrije odvisna od tega, kako blizu je jedro gostitve od naravne meje, je nazorno vidno na sliki 7.9. V njej so prikazane frekvenčne po-

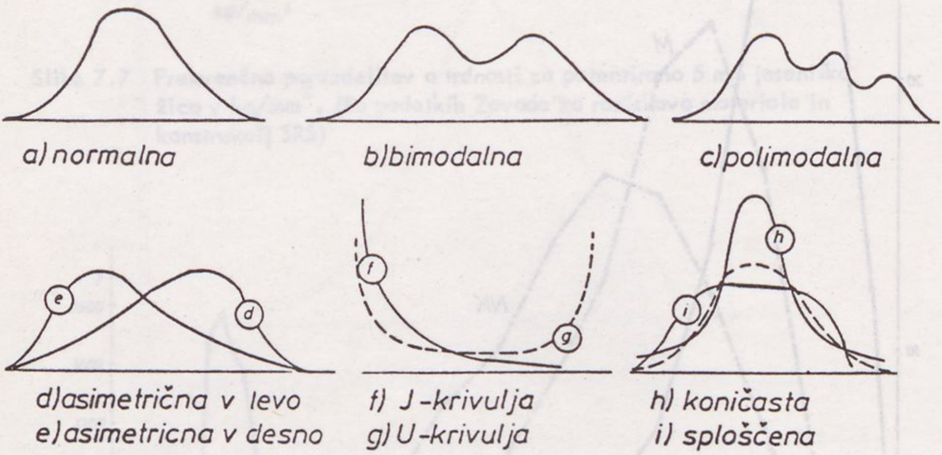


Slika 7.9 Frekvenčne porazdelitve K = kmečkih, M = mešanih in NK = nekmečkih gospodinjstev v SR Sloveniji v letu 1960 po letnem dohodku

razdelitve gospodinjstev v SR Sloveniji v letu 1960 po letnem dohodku. Gospodinjstva so razdeljena v tri skupine: K = kmečka, M = mešana in NK = nekmečka. Ker je za kmečka gospodinjstva jedro gostitve – modus najbližje naravni omejitvi dohodka (dohodek o), je asimetrija v desno za to skupino največja. Za nekmečka gospodinjstva je iz obratnega razloga stopnja asimetrije znatno manjša.

7.12 V socialno-ekonomskih raziskavah sta si naravna omejitev in modus včasih zelo blizu ali se celo ujemata. Za take pojave gostitev zdržema pada od mesta gostitve. Zaradi značilne oblike imenujemo take frekvenčne porazdelitve J – porazdelitve, ker imajo bolj ali manj obliko črke J.

Če imamo normalno porazdelitev za idealno, so nekatere porazdelitve v primerjavi z njo k o n i č a s t e , druge pa sploščene (glej sliko 7.10, h,i).



Slika 7.10 Oblike frekvenčnih porazdelitev

KUMULATIVNA FREKVENČNA PORAZDELITEV

7.13 Ker frekvenčne porazdelitve prikazujejo število enot (ekstenziven podatek) po razredih za numeričen znak, moremo zanje izračunati kumulativne frekvenčne porazdelitve F_k po splošnem pravilu za računanje kumulativnih vrst. Kumulativno frekvenčno porazdelitev F_k dobimo, če postopoma prištevamo frekvence f_k v frekvenčni porazdelitvi. Člene v kumulativni frekvenčni porazdelitvi dobimo torej po obrazcu

$$F_{k+1} = F_k + f_k \qquad (7.5)$$

7.14 Če za frekvenčno porazdelitev o porabi lesa iz tabele 7.3 izračunamo kumulativno porazdelitev, dobimo naslednjo kumulativno vrsto:

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

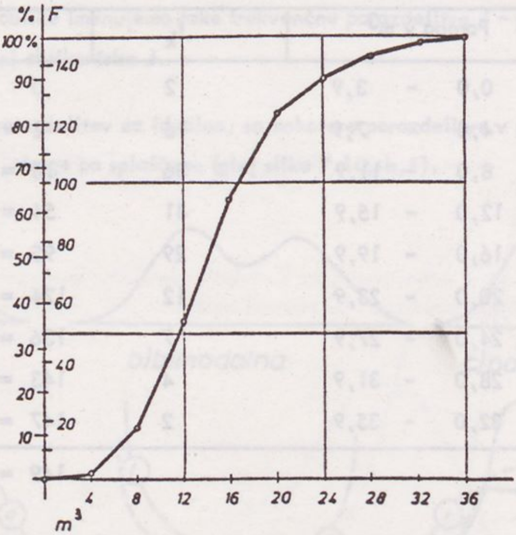
Poraba v m^3	f_k	F_k
0,0 - 3,9	2	0
4,0 - 7,9	16	2 = 0 + 2
8,0 - 11,9	36	18 = 2 + 16
12,0 - 15,9	41	54 = 18 + 36
16,0 - 19,9	29	95 = 54 + 41
20,0 - 23,9	12	124 = 95 + 29
24,0 - 27,9	7	136 = 124 + 12
28,0 - 31,9	4	143 = 136 + 7
32,0 - 35,9	2	147 = 143 + 4
		149 = 147 + 2 = N

Posamezni členi v kumulativni frekvenčni porazdelitvi F_k pomenijo, koliko enot ima vrednost pod spodnjo mejo ustreznega razreda $y_{k,min}$. Tako pomeni npr. šesti člen v kumulativni vrsti, da je $F_6 = 124$ gospodarstev, ki so v letu 1953 porabila manj lesa kakor $y_{6,min} = 20\ m^3$ lesa.

Kumulativno vrsto moremo izračunati tudi tako, da začnemo z večjim členom in postopoma prištevamo frekvence v obratni smeri. Členi take kumulativne vrste pomenijo, koliko enot ima vrednosti nad zgornjo mejo ustreznih razredov. Vendar tako kumulativno vrsto uporabljamo redkeje kakor zgornjo.

Enako računamo kumulativne vrste tudi za frekvenčne porazdelitve z različnimi širinami razredov. Pomen členov je isti. Kumulativne frekvenčne porazdelitve moremo izračunati tudi iz vrst relativnih frekvenc.

7.13 Če je frekvenčna porazdelitev unimodalna, ima grafični prikaz za kumulativno frekvenčno porazdelitev značilno obliko črke S. O tem se moremo prepričati tudi v grafikonu za kumulativno frekvenčno porazdelitev o porabi lesa v kmetijskih gospodarstvih v sliki 7.11, ki je načrtana po podatkih iz tabele 7.7. Opomniti moramo, da vrednosti kumulativnih frekvenc v grafikonu nanašamo nad spodnjo mejo ustreznih razredov na abscisi.



Slika 7.11 Kumulativna frekvenčna porazdelitev o porabi lesa v letu 1953 v 149 kmetijskih gospodarstvih v okraju Novo mesto

V grafikonu imamo na ordinatni osi dve skali: skalo kumulativnih absolutnih in relativnih frekvenc.

OSMO POGLAVJE

KVANTILI

RANŽIRNA VRSTA, RANG

8.1 Osnovni statistični podatki, ki jih zberemo s statističnim opazovanjem, so dani v neurejeni množici podatkov. Te podatke običajno uredimo v frekvenčno porazdelitev. Moremo pa jih pregledno prikazati tudi v ranžirni vrsti, v kateri so urejeni po velikosti od najmanjšega do največjega ali obratno. V tabeli 8.1 so prikazani osnovni podatki za poprečen promet v trgovini na drobno na prebivalca v SR Sloveniji v letu 1971 po občinah.

Tabela 8.1 Promet v trgovini na malo na prebivalca v SR Sloveniji v letu 1971 po občinah v din (Vir: SG-72)

Ajdovščina	AJD	6357	Maribor	MAR	8884
Brežice	BRE	6509	Metlika	MET	6453
Celje	CEL	12983	Mozirje	MOZ	5077
Cerknica	CER	4718	Murska Sobota	MSO	5367
Črnomelj	ČRN	4068	Nova gorica	NGO	12039
Domžale	DOM	6205	Novo mesto	NME	6445
Dravograd	DRA	4826	Ormož	ORM	2752
Gornja Radgona	GRA	4359	Piran	PIR	9678
Grosuplje	GRO	3457	Postojna	POS	8107
Hrastnik	HRA	5082	Ptuj	PTU	5406
Idrija	IDR	5900	Radlje ob Dravi	RJE	4051
Ilirska Bistrica	IBI	5504	Radovljica	RAD	9132

Izola	IZO	8666	Ravne na Koroškem	RAV	5163
Jesenice	JES	8541	Ribnica	RIB	5539
Kamnik	KAM	5728	Sevnica	SEV	3978
Kočevje	KOČ	7089	Sežana	SEŽ	15607
Koper	KOP	14398	Slov. Bistrica	SBI	4135
Kranj	KRA	8538	Slov. Konjice	SKO	4222
Krško	KRŠ	4400	Slov. Gradec	SLG	6332
Laško	LAŠ	4453	Šentjur pri Celju	ŠTJ	2244
Lenart	LRT	2377	Škofja Loka	ŠLO	6667
Lendava	LVA	3973	Šmarje pri Jelšah	ŠMA	3557
Litija	LIT	3743	Tolmin	TOL	8021
Logatec	LOG	5350	Trbovlje	TRB	7714
Ljubljana-Bežigrad	LJB	11293	Trebnje	TRE	4029
Ljubljana-Center	LJC	49433	Tržič	TRŽ	6966
Ljubljana-Moste-Polje	LJM	4499	Velenje	VEL	7161
Ljubljana-Šiška	LJŠ	8598	Vrhnika	VRH	5801
Ljubljana-Vič-Rudnik	LJV	3988	Zagorje ob Savi	ZAG	4907
Ljutomer	LJU	4974	Žalec	ŽAL	4492

Podatki v tabeli 8.1 so urejeni po abecednem redu občin. Za potrebe evidence je tak pregled primeren, statistično pa dobimo veliko boljši pregled, če podatke uredimo po velikosti in jih podamo v ranžirni vrsti. Podatki o prometu v trgovini na drobno na prebivalca po občinah v SRS v letu 1971 so podani v ranžirni vrsti v tabeli 8.2.

Iz ranžirne vrste pa moremo napraviti različne sklepe o prikazanem pojavu. Iz nje podobno kakor iz frekvenčne porazdelitve neposredno vidimo, v kakšnih mejah variira populacija, ker kaže podatek (2244 din), ki je prvi po rangu, najmanjši promet, oni, ki je zadnji po rangu, pa največji promet na malo (49433 din).

Razen tega je v ranžirni vrsti vsaki enoti dodana vrednost novega znaka-ranga. Rang daje o posameznih enotah dodatno informacijo, katere osnovna vrednost znaka - v našem zgledu poprečen promet v občini, ne daje. Če npr. za občino Nova Gorica vemo, da je bil poprečen promet na drobno v letu 1971 12039 din, iz tega podatka ne moremo sklepati, ali je to malo ali veliko. Če pa dodatno povemo, da je v ranžirni vrsti

Tabela 8.2 Ranžirna vrsta za občine v SR Sloveniji za promet na drobno na prebivalca v letu 1971

R	01	02	03	04	05	06	07	08	09	10
Vrednost	2244	2372	2752	3457	3557	3743	3973	3982	3988	4029
Oznaka	ŠTJ	LRT	ORM	GRO	ŠMA	LIT	LEN	SEV	LJ-V	TRE
R	11	12	13	14	15	16	17	18	19	20
Vrednost	4051	4068	4135	4222	4358	4400	4453	4492	4499	4718
Oznaka	RJE	ČRN	SBI	SKO	GRA	KRŠ	LAŠ	ŽAL	LJ-M	CER
R	21	22	23	24	25	26	27	28	29	30
Vrednost	4826	4904	4947	5077	5082	5163	5350	5367	5406	5504
Oznaka	DRA	ZAG	LJU	MOZ	HRA	RAV	LOG	MSO	PTU	IBI
R	31	32	33	34	35	36	37	38	39	40
Vrednost	5539	5728	5804	5900	6205	6332	6357	6445	6453	6509
Oznaka	RIB	KAM	VRH	IDR	DOM	SLG	AJD	NME	MET	BRE
R	41	42	43	44	45	46	47	48	49	50
Vrednost	6667	6966	7089	7161	7714	8021	8107	8538	8541	8598
Oznaka	ŠLO	TRŽ	KOČ	VEL	TRB	TOL	POS	KRA	JES	LJ-Š
R	50	52	53	54	55	56	57	58	59	60
Vrednost	8666	8884	9132	9687	11293	12039	12083	14398	15607	49433
Oznaka	IZO	MAR	RAD	PIR	LJB	NGO	CEL	KOP	SEŽ	LJ-C

60 občin, občina Nova Gorica pa 56- po rangu, sklepamo, da je ta poprečen promet za Slovenijo razmeroma velik, ker imajo od 60 občin samo 4. občine večji promet na prebivalca, 55 občin pa manjšega.

KVANTILNI RANG

8.2 Hiba ranga je v tem, da moramo poleg ranga navesti še skupno število enot populacije, če hočemo, da rang pokaže mesto enote v populaciji. Občina, ki je 56. po rangu, je npr. za populacijo s 60 enotami občina z razmeroma velikim prometom. Če pa je neka občina 56. po rangu v ranžirni vrsti za vseh 500 občin v SFRJ, pa bi bil ta podatek razmeroma majhen, ker bi imelo 444 občin večji promet. Rang R pokaže torej mesto enote v populaciji šele, če ga primerjamo z obsegom populacije N. Zato je primerneje, da izračunamo mesto enote v populaciji namesto z rangom R s kvantilnim rangom P. Tega dobimo, če primerjamo rang R z obsegom populacije N. Kvantilni rang v relativnem številu pove, na katerem delu celotnega ranžirnega razmika leži določena enota oziroma koliki del celote ima manjše vrednosti kakor je dana vrednost.

8.3 Teoretično vzamemo, da je ranžirni razmik zvezen in vsakemu rangu pripišemo razmik polovico enote na levo in desno. Po tej predpostavki se ranžirni razmik začne z 0,5 (spodnja meja razmika, ki ustreza rangu 1) in konča z $N + 0{,}5$ (zgornja meja razmika, ki ustreza rangu N). Ker se skala kvantilnih rangov P začne z 0 in konča z 1, je zveza med rangom R in kvantilnim rangom P dana z obrazcem

$$R = NP + 0{,}5 \qquad (8.1)$$

Če je $P = 0$, je $R = 0{,}5$, če pa je $P = 1$, je $R = N + 0{,}5$

Iz gornje zveze dobimo, da velja tudi

$$P = \frac{R - 0{,}5}{N} \qquad (8.2)$$

Če vzamemo naš zgled o prometu na drobno v Novi Gorici z rangom $R = 56$ in $N = 60$, dobimo po obrazcu

$$P = \frac{56 - 0{,}5}{60} = 0{,}93$$

Kvantilni rang $P = 0{,}93$ pove, da ima 0,93 del celotne populacije občin manjši povprečni promet na prebivalca kot občina Nova Gorica. Iz zgleda vidimo, da kvantilni rang P sam zase, ne da bi navajali obseg populacije, nazorno prikaže mesto določene enote v populaciji.

KVANTILI

8.4 Z obrazcem 8.1 in 8.2 moremo reševati dva različna problema.

Za posamezno enoto populacije moremo določiti mesto enote v populaciji, če izračunamo vrednosti y ustrezen kvantilni rang P. Kvantilni rangi so torej značilnosti posameznih enot.

Moremo pa analogno reševati tudi drug problem. Če se vprašamo, kakšna vrednost y_P ustreza npr. kvantilnemu rangu $P = 0,50$, vrednost, ki jo dobimo, ne označuje posamezne enote, marveč je značilnost za populacijo. y_P, ki ustreza danemu kvantilnemu rangu P, imenujemo **kvantil**. Kvantil $y_{0,50}$ je npr. vrednost, od katere ima polovica enot populacije manjše, polovica pa večje vrednosti. Ta vrednost je vsekakor važen parameter populacije in jo imenujemo **mediana**.

$$Me = y_{P=0.50} \tag{8.3}$$

Kvantili $y_{0,25}$, $y_{0,50}$ in $y_{0,75}$ so vrednosti, ki razdele populacijo v štiri dele tako, da je pod $y_{0,25}$, med $y_{0,25}$ in $y_{0,50}$, med $y_{0.50}$ in $y_{0,75}$ in nad $y_{0,75}$ po četrtina po velikosti urejenih vrednosti populacije. Te vrednosti, ki razdele populacijo v štiri po obsegu enake dele, imenujemo **kvartile** in zaznamujemo z

$$Q_1 = y_{P=0.25}\,; \quad Q_2 = y_{P=0.5o}\,; \quad Q_3 = y_{P=0.75} \tag{8.4}$$

Podobno z **decili**

$$D_1 = y_{0,10},\ D_2 = y_{0,20} \ldots\ldots\ldots\ldots D_9 = y_{0,90} \tag{8.5}$$

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

$$C_1 = y_{0,01} \quad C_2 = y_{0,02} \ldots\ldots C_{98} = y_{0,98} \text{ in } C_{99} = y_{0,99} \tag{8.6}$$

pa v sto po obsegu enakih delov.

Z vrsto kvantilov prikažemo gostitev vrednosti populacije vendar drugače kot s frekvenčno porazdelitvijo. V frekvenčni porazdelitvi z enakimi razredi se glede na gostitev pojava menja frekvenca v razredih, s kvantili pa razmejimo razrede, v katerih so frekvence enake: N/4 pri kvantilih, N/10 pri decilih in N/100 pri centilih, od stopnje gostitve pojava pa je odvisna širina razredov oziroma razmiki med zaporednimi kvantili.

Preračunanje kvantilov iz negrupiranih podatkov

8.5 Ker predpostavljamo, da je rang zvezna količina, moremo določiti range za vsako vrednost med najmanjšo in največjo vrednostjo v populaciji in ne samo za podatke, navedene v ranžirni vrsti. Za vmesne vrednosti uporabimo linearno interpolacijo. Tako npr. menimo, da ustreza v našem zgledu rangu 12,5 kot kvantil sredina med vrednostjo za rang 12 in 13: (4068 + 4135) : 2 = 4101,5 din

Če predpostavljamo zveznost rangov, izračunamo iz ranžirne vrste za poljubno vrednost y med najmanjšo y_{min} in največjo vrednostjo populacije y_{max} ustrezni kvantilni rang P_y po naslednjem postopku:

a) Imamo ranžirno vrsto vrednosti enot populacije.

b) V ranžirni vrsti poiščemo, med kateri vrednosti y_0 in y_1 pade vrednost y, za katero iščemo P_y, tako da velja: $y_0 \leq y < y_1$. Vrednosti y_0 naj ustreza rang R_0.

c) Rang R_y, ki ustreza vrednosti y, dobimo z linearno interpolacijo po obrazcu

$$R_y = R_0 + \frac{y - y_0}{y_1 - y_0} \tag{8.7}$$

d) Kvantilni rang P_y izračunamo iz ranga R_y in obsega populacije po obrazcu

$$P_y = \frac{R_y - 0,5}{N} \tag{8.8}$$

Če iščemo, kateri kvantilni rang ustreza npr. prometu na drobno na prebivalca $y = 5000$ din, dobimo po gornjem postopku: $y = 5000$ din leži v ranžirni vrsti med vrednostima $y_0 = 4947$, za katero je rang $R_o = 23$ in vrednostjo $y_1 = 5077$.

Iz teh podatkov izračunamo rang $R_{y=5000}$ po obrazcu 8.7

$$R = 23 + \frac{5000 - 4947}{5077 - 4947} = 23,41$$

Po obrazcu 8.8 pa dobimo dalje: $P_y = \dfrac{23,41 - 0,50}{60} = 0,382$

Kvantilni rang za občino, s prometom na drobno na prebivalca y = 5000 din, je $P_{y=5000} = 0.382$.

8.6 Obratni problem, da k danemu kvantilnemu rangu P_y poiščemo ustrezen kvantil y_p, pa rešimo takole:

a) Za populacijo imamo ranžirno vrsto.

b) Iz danega P po obrazcu

$$R_p = NP + 0,5 \tag{8.9}$$

izračunamo ustrezni rang R.

c) V ranžirni vrsti poiščemo, med katera cela ranga pade izračunani R_p, tako da velja: $R_o \leqslant R_p < R_1$. Rangoma R_o in R_1 ustrezata vrednosti y_o in y_1.

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

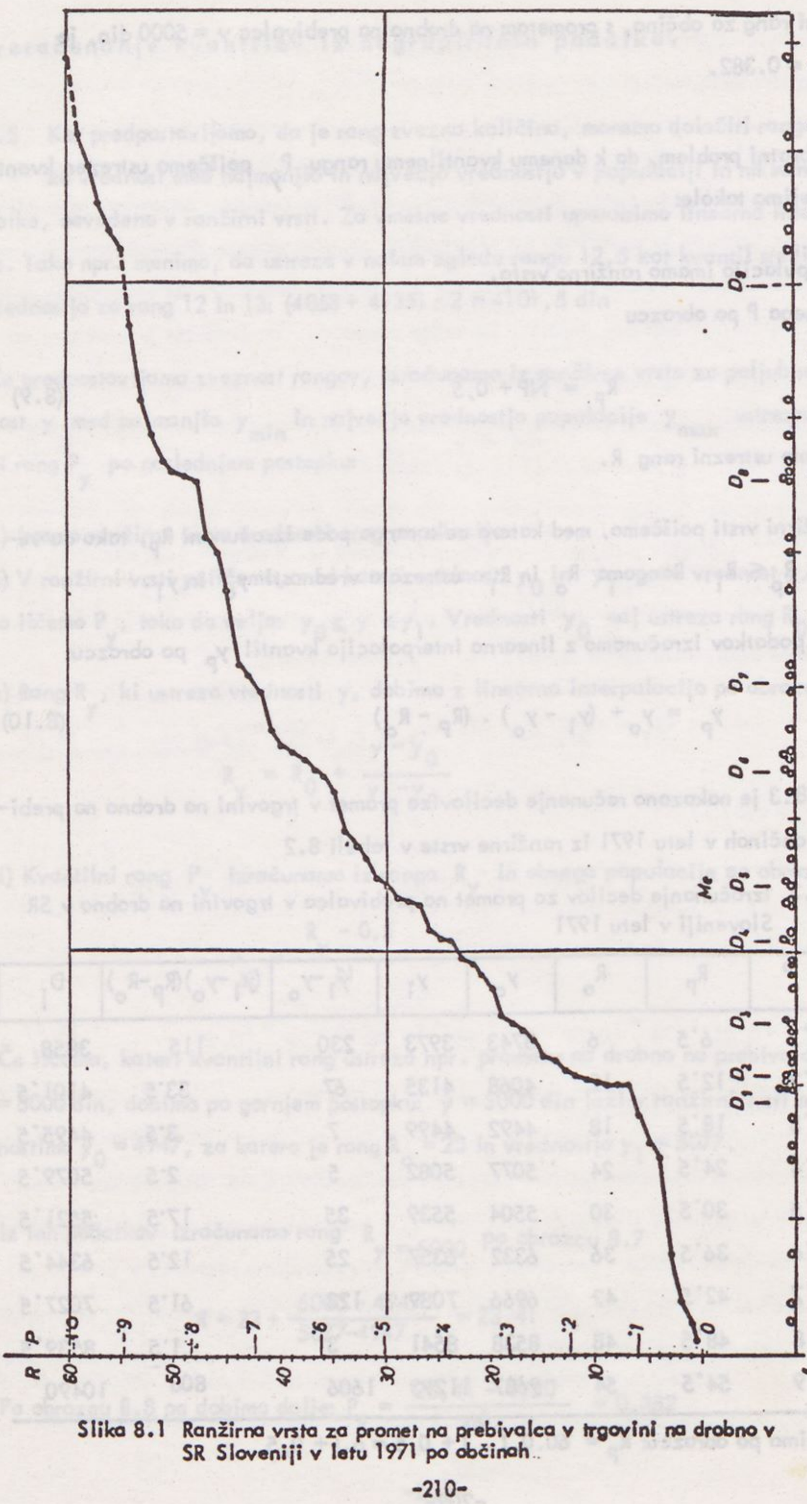
$$y_p = y_o + (y_1 - y_o) \cdot (R_p - R_o) \tag{8.10}$$

V tabeli 8.3 je nakazano računanje decilov za promet v trgovini na drobno na prebivalca po občinah v letu 1971 iz ranžirne vrste v tabeli 8.2

Tabela 8.3 Izračunanje decilov za promet na prebivalca v trgovini na drobno v SR Slovenija v letu 1971

i	P	R_p	R_o	y_o	y_1	y_1-y_o	$(y_1-y_o)(R_p-R_o)$	D_i
1	˙1	6˙5	6	3743	3973	230	115	3858
2	˙2	12˙5	12	4068	4135	67	33˙5	4101˙5
3	˙3	18˙5	18	4492	4499	7	3˙5	4495˙5
4	˙4	24˙5	24	5077	5082	5	2˙5	5079˙5
5	˙5	30˙5	30	5504	5539	35	17˙5	5521˙5
6	˙6	36˙5	36	6332	6357	25	12˙5	6344˙5
7	˙7	42˙5	42	6966	7089	123	61˙5	7027˙5
8	˙8	48˙5	48	8538	8541	3	1˙5	8539˙5
9	˙9	54˙5	54	9687	11293	1606	803	10490

R_p dobimo po obrazcu: $R_p = 60.0,1 \cdot i + 0,5 = 6.i + 0,5$



Slika 8.1 Ranžirna vrsta za promet na prebivalca v trgovini na drobno v SR Sloveniji v letu 1971 po občinah

V sliki 8.1 so z "o" narisane vrednosti o prometu na drobno na prebivalca, nakazana je frekvenčna porazdelitev, decili in kumulativna vrsta frekvenc, ki posreduje preslikave vrednosti y v kvantilne range P_y in obratno preslikave kvantilnih rangov P v kvantile y_P.

8.7 V sliki 8.1 imamo v grafikonu za promet na prebivalca v trgovini na drobno v SR Sloveniji po občinah vrisano ranžirno vrsto tako, da je abscisa vrednost podatka, ordinata pa ustrezni rang. Točke predstavljajo osnovno ranžirno vrsto, lomljena črta, ki veže točke, pa linearno interpolacijo za vmesne vrednosti. V grafikonu je razen skale rangov vrisana tudi skala kvantilnih rangov P, za katero je nazorno vidno, da se na skali rangov začne z 0.5 in konča z N + 0,5.

8.8 Rang grafikon. Rang grafikon je kombinacija ranžirne vrste in frekvenčne porazdelitve. V njem so enote razvrščene po velikosti od največje do najmanjše vrednosti v tako ozkih razredih, da moremo zadosti dobro določiti tudi posamično vrednost. V našem zgledu je vsaka občina označena s tremi črkami, v načelu s prvimi tremi črkami imena občine, če pa je ime občine sestavljeno iz več besed pa iz začetnic vseh besed imena vendar tako, da skupno ne presega treh črk. Vsaka občina je obeležena tudi z zaporedno številko po velikosti urejenih enot - z rangi. Tako moremo za vsako občino odbrati rang in približno velikost podatka. Tako je npr. občina Piran (oznaka PIR) po prometu v trgovini na drobno na sedmem mestu med vsemi šestdesetimi občinami v SR Sloveniji in ima v letu 1971 med 9500 in 9750 din prometa na prebivalca. Ker je N = 60 občin, je prvi kvartil med 15. in 16. občino (med poprečjem za Tolmin in Trbovlje), dalje je promet med Ribnico in Ilirsko Bistrico enak mediani za promet na drobno na prebivalca ipd. Levo od skale so v ustrezne razrede vnešeni tudi poprečni podatki za posamezne republike in SFRJ skupno. Tako moremo napraviti tudi primerjave občinskih podatkov z republiškimi poprečji. Tak poprečen promet v trgovini na drobno na prebivalca, kot je približno v poprečju za SFRJ, imajo občine: Ljubljana-Moste, Žalec, Laško, Krško in Gornja Radgona. Enako sklepamo, da ima tri četrtine občin poprečje večje kot je jugoslovansko poprečje in le ena četrtina večje kot je poprečje za SR Slovenijo. Skladnost tretjega kvartila s poprečjem za SR Slovenijo je izraz velike asimetrije proučevane porazdelitve. Ker je informacija za vsako občino na ustreznem mestu dana z istomestno oznako: dvomesten rang, črtica, tromestna oznaka občine, daje rang grafikon kot celota

din/preb
50000 01-LJC
40000
16000 02-SEŽ
15000 03-KOP
14000
13000 04-CEL 05-NGo
12000 06-LjB
11000
10000
9750 07-PIR
9500
9250 08-RAD
9000 09-MAR
8750 10-IZO 11-LjŠ 12-JES 13-KRA
8500
8250
SLO 8000 14-POS 15-TOL
7750
7500 16-TRB
7250
7000 17-VEL 18-KOČ
19-TRŽ
6750 20-ŠLo 21-BRE
6500 22-MET 23-NMe 24-AJD 25-SGr
6250 26-DOM
6000
27-IDR 28-VRH
5750
HRV 29-KAM 30-RIB 31-IBi
5500
5250 32-PTU 33-MSo 34-LOG
35-RAV 36-HRA 37-MOZ
Voj 5000 38-LJU 39-ZAG 40-DRA
4750 41-CER
4500
SFRJ 42-LjM 43-ŽAL 44-LAŠ 45-KRŠ 46-GRA
4250
OSr 47-SKo 48-SBi 49-ČRN 50-RJE 51-TRE
4000
SR 52-LJV 53-SEV 54-LEN
3750
ČGr 55-LIT 56-ŠMA
3500
MAK 57-GRO
3250
BIH
3000
58-ORM
2750
2500
2250 59-LRT
60-ŠTj
2000
KOS 1500

Slika 8.2 Rang grafikon prometa na prebivalca v trgovini na drobno v SR Sloveniji po občinah v letu 1971

vtis histograma. Razen nakazanih informacij o poprečnem prometu moremo iz rang grafikonov dobiti še druge informacije, ki so kombinacija naziva, ranga in vrednosti podatka za posamezno enoto.

Računanje kvantilnih rangov in kvantilov iz frekvenčnih porazdelitev

8.9 Če je populacija, ki jo proučujemo, obsežna, je razvrščanje vrednosti v ranžirno vrsto neprikladno. Za obsežne populacije pa moremo izračunati približne vrednosti kvantilnih rangov in kvantilov iz kumulativnih frekvenčnih porazdelitev. Če vemo, kaj pomenijo členi v kumulativni vrsti, spoznamo, v kakšni zvezi je kumulativna frekvenčna porazdelitev z rangi.

Proučimo frekvenčno porazdelitev o porabi lesa v 149 kmetijskih gospodarstvih v Novem mestu iz tabele 7.3. Iz nje vidimo, da ima frekvenčna porazdelitev nekatere lastnosti ranžirne vrste. V ranžirni vrsti so vrednosti, ki leže levo od dane vrednosti, manjše, desno pa večje. V frekvenčni porazdelitvi so vse vrednosti enot v danem razredu večje od vrednosti v vseh spodnjih razredih in manjše od vrednosti v vseh zgornjih razredih. Tudi v frekvenčni porazdelitvi so torej vrednosti razporejene po velikosti, le da je ranžiranje izvedeno med razredi, ne pa znotraj razredov.

Iz kumulativne frekvenčne porazdelitve moremo celo približno ugotoviti, katere vrednosti ustrezajo določenim rangom. Če preučimo kumulativno frekvenčno porazdelitev v tabeli 8.4, moremo iz nje sklepati tole:

Tabela 8.4 Kumulativna frekvenčna porazdelitev o porabi lesa v letu 1953 za 149 kmetijskih gospodarstev v okraju Novo mesto

Poraba v m^3	f_k	F_k
0,0 - 3,9	2	0
4,0 - 7,9	16	2
8,0 - 11,9	36	18
12,0 - 15,9	41	54
16,0 - 19,9	29	95
20,0 - 23,9	12	124
24,0 - 27,9	7	136
28,0 - 31,9	4	143
32,0 - 35,9	2	147
		149 = N

Dvoje kmetijskih gospodarstev je porabilo manj kot 4,0 m^3 lesa. Če vzamemo, da ima gospodarstvo z največjo porabo v tem razredu v približku porabo enako tej meji, je gospodarstvo, ki ima rang 2, porabilo 4,0 m^3. Iz kumulative dalje izvemo, da je 18 gospodarstev porabilo pod 8,0 m^3. Če vzamemo, da je največja poraba v teh 18 gospodarstvih enaka tej meji, dobimo, da ima gospodarstvo, ki je porabilo 8,0 m^3 lesa, rang 18 itd. Iz frekvenčne porazdelitve sklepamo, da imajo gospodarstva, ki so porabila toliko kot so spodnje meje razredov, range enake ustreznim vrednostim v kumulativni vrsti. Če to upoštevamo, moremo sestaviti tabelo, ki je zelo podobna ranžirni vrsti, le da so rangi dani samo za meje razredov.

Tabela 8.5 Okrnjena ranžirna vrsta za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto. (Dobljena iz kumulativne vrste v tabeli 8.4)

Rang R	F_k	2	18	54	95	124	136	143	147	149
y_R	y_{min}	4,0	8,0	12,0	16,0	20,0	24,0	28,0	32,0	36,0

Ker za druge enote v ranžirni vrsti ne poznamo vrednosti, predpostavljamo, da so vrednosti znotraj razredov razmeščene enakomerno. Tako dobimo vrednosti drugih enot v ranžirni vrsti z linearno interpolacijo. Grafikon kumulativne frekvenčne porazdelitve v sliki 8.3. ponazarja približen odnos med rangom R in vrednostjo znaka y. Ker tudi pri grafičnem prikazu predpostavljamo enakomerno razporeditev vrednosti znotraj razredov, so točke zvezane z daljicami.

Ker je kumulativna frekvenčna porazdelitev nadomestek za ranžirno vrsto, računamo iz nje kvantilne range in kvantile podobno kakor iz ranžirne vrste.

8.10 Oceno kvantilnega ranga P_y izračunamo po naslednjem postopku:

a) Iz frekvenčne porazdelitve izračunamo kumulativno frekvenčno porazdelitev F_k.

b) Poiščemo, v kateri razred pade vrednost y za katero iščemo kvantilni rang P_y. Ta razred imenujemo kvantilni razred in ga zaznamujmo z o. Zanj izpišemo iz frekvenčne porazdelitve ustrezne vrednosti:

$$y_{o,min}, i_o, f_o, F_o.$$

c) Iz gornjih vrednosti izračunamo vrednosti y ustrezni rang R_y po obrazcu:

$$R_y = F_o + f_o \frac{y - y_{o.min}}{i_o} \tag{8.11}$$

d) Iz dobljenega ranga R_y pa izračunamo kvantilni rang P_y po obrazcu:

$$P_y = \frac{R_y - 0{,}50}{N} \tag{8.12}$$

Če je obseg populacije N velik, iz obrazca 8.12 običajno izpuščamo 0,5, ker je ta količina za velike populacije nebistvena.

Postopek velja tudi za frekvenčne porazdelitve z različnimi širinami razredov.

8.11 Če za populacijo o porabi lesa iz tabele 8.4 iščemo kvantilni rang za gospodarstvo, ki je v letu 1953 porabilo $y = 13{,}8\ m^3$ lesa, izračunamo P_y po zgornjem postopku takole: Poraba lesa $y = 13{,}8\ m^3$ pade v razred 12,0–15,9. Iz tabele 8.4 odberemo:

$$y_{o,min} = 12{,}0\ ;\ i_o = 4\ ;\ f_o = 41\ ;\ F_o = 54\ ;$$

Če vstavimo te podatke v obrazec 8.11, dobimo:

$$R_y = 54 + 41 \frac{13{,}8 - 12{,}0}{4} = 72{,}4$$

in dalje po obrazcu 8.12:

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

Gospodarstvo s porabo $y = 13{,}8\ m^3$ lesa je v 48. centilu.

8.12 Podobno računamo iz frekvenčne porazdelitve tudi k v a n t i l e . Postopek je naslednji:

a) Iz frekvenčne porazdelitve izračunamo kumulativno frekvenčno porazdelitev F_k.

b) Iz danega P izračunamo ustrezni rang R_P po obrazcu

$$R_P = NP + 0{,}5 \tag{8.12}$$

Če je populacija velika, v obrazcu 8.13 izpustimo 0,5.

c) V kumulativni frekvenčni porazdelitvi F_k poiščemo, med kateri vrednosti kumulativne vrste pade R, tako da je: $F_o \leq R_p < F_1$. F_o ustrezen razred je kvantilni razred. Zanj poiščemo v frekvenčni porazdelitvi količine:

$$y_{o,min}, i_o, f_o \text{ in } F_o .$$

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

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

Enako kot za kvantilne range velja nakazani postopek tudi za frekvenčne porazdelitve z različnimi širinami razredov.

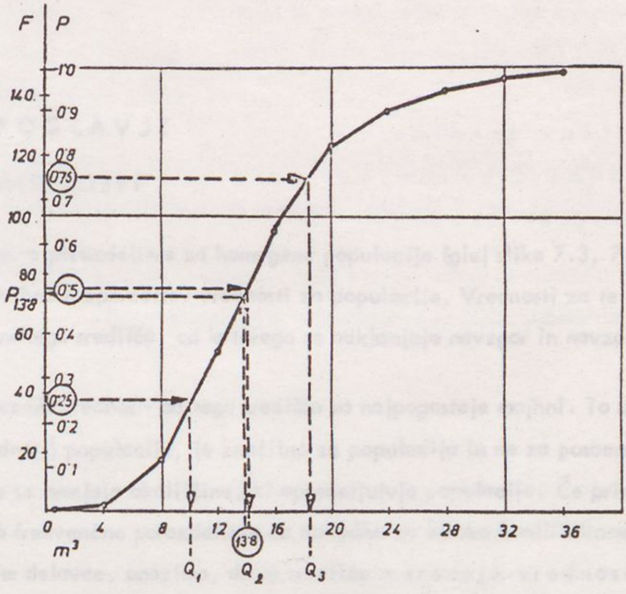
8.13 Da za naš zgled izračunamo kvartile, je najbolje, da sistematično izračunamo v računski tabeli vse tri kvartile hkrati.

Tabela 8.6 Izračunavanje kvartilov za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto.

i	P	R_p	$y_{0,min}$	i_o	f_o	F_o	$R_p - F_o$	$i_o \frac{R_p - F_o}{f_o}$	Q_i
1	0,25	37,75	8,0	4	36	18	19.75	2,19	$10,19 = Q_1$
2	0,50	75,00	12,0	4	41	54	21.00	2,05	$14,05 = Q_2$
3	0,75	112,25	16,0	4	29	95	17,25	2,38	$18,38 = Q_3$

8.14 Podobno kakor v sliki 8.1 za negrupirane podatke, moremo grafično oceniti vse zgornje količine iz grafikona kumulativne vrste. V sliki 8.3 je nakazano, kako dani vrednosti $y = 13,8\ m^3$ poiščemo ustrezni kvantilni rang in obratno: kako moremo grafično oceniti kvartile.

8.15 V praksi večkrat izdelujemo decilne ali centilne norme, s katerimi dobimo neposredno zvezo vrednosti znaka z decili oziroma pri podrobnejših normnih tablicah s centili. Centilne norme s pridom uporabljamo v psihologiji, moremo pa jih izdelati tudi za določene socialno-ekonomske probleme. Tako tablica decilnih ali centilnih norm o produktivnosti dela pomaga presoditi, kakšna je produktivnost dela za posameznega delavca v razmerju s celotnim kolektivom itd.



Slika 8.3 Grafično ocenjevanje kvantilnih rangov in kvantilov iz slike kumulativnih frekvenčnih porazdelitev za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu v letu 1953

Če je populacija velika, v obrazcu 8.13 izpustimo 0,5.

c) V kumulativni frekvenčni porazdelitvi [illegible]

[illegible]

Zanj poiščemo v frekvenčni porazdelitvi izdelane

$y_{o,min}$, i_o, f_o in F_o.

d) Iz teh količin izračunamo ustrezni kvantil y_p po obrazcu

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

Enako kot za kvantilne range velja nakazani postopek tudi za frekvenčne porazdelitve z različnimi širinami razredov.

8.13 Da za naš zgled izračunamo kvartile, je najbolje, da sistematično izračunamo v računski tabeli vse tri kvartile hkrati.

Tabela 8.6 Izračunavanje kvartilov za porabo lesa v 149 kmečkih gospodarstvih v okraju Novo mesto.

i	P_i	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Q_i
1	0,25	37,75	8,0	4	35	18	19,75	2,19	10,19 = Q_1
2	0,50	75,00	12,0	4	41	54	21,00	2,05	14,05 = Q_2
3	0,75	112,25	16,0	4	29	95	17,25	2,38	18,38 = Q_3

8.14 Podobno kakor v sliki 8.1 za negrupirane podatke, moremo grafično oceniti vse zgornje količine iz grafikona kumulativne vrste. V sliki 8.3 je nakazano, kako dani vrednosti $y = 13,9\ m^3$ poiščemo ustrezni kvantilni rang in obratno kako moremo grafično oceniti kvartile.

8.15 V praksi večkrat izdelujemo decilne ali centilne norme, s katerimi dobimo neposredno zvezo vrednosti znaka z decili oziroma pri podrobnejših normah tabelicah.

DEVETO POGLAVJE

SREDNJE VREDNOSTI

9.1 Frekvenčne porazdelitve za homogene populacije (glej slike 7.3, 7.6, 7.7) kažejo posebno razporeditev vrednosti za populacije. Vrednosti za te populacije se goste okrog nekega središča, od katerega se odklanjajo navzgor in navzdol.

Odkloni posameznih vrednosti od tega središča so najpogosteje majhni. To središče, ki predstavlja vrednosti populacije, je značilno za populacijo in ne za posamezne enote in se menja, če se menjajo okoliščine, ki opredeljujejo populacijo. Če primerjamo za določeno stroko frekvenčne porazdelitve za dohodke za visoko kvalificirane, kvalificirane in proučene delavce, opazimo, da so središča - srednje vrednosti za posamezne populacije različna. Dohodki za visoko kvalificirane delavce se goste najvišje, dohodki kvalificiranih delavcev nižje in dohodki priučenih delavcev najnižje. Ta središča so torej za posamezne populacije določene z opredeljujočimi pogoji populacije, ti pa so za vse enote enaki. Če bi na dohodek za posameznega delavca vplivala samo kvalifikacija, bi imeli vsi delavci z enako kvalifikacijo enake dohodke. Zaradi drugih - posamičnih vplivov, ki so za vsakega delavca različni, pa se dohodki odklanjajo navzgor in navzdol od dohodkov, ki so pogojeni s kvalifikacijo.

Čim manjši so posamični vplivi, tem bolje srednja vrednost predstavlja vrednosti populacije in obratno: čim večji so posamični vplivi, tem slabše srednja vrednost predstavlja vrednosti populacije. Zato je srednja vrednost predstavnik vrednosti v populaciji le, če je populacija enovita.

Vrste srednjih vrednosti

9.2 Parametrov, ki pokažejo osrednjo težnjo vrednosti populacije, poznamo več.

Od teh so v socialno-ekonomski statistiki pomembne naslednje srednje vrednosti:

a) mediana, - Me
b) modus, - Mo
c) aritmetična sredina, - M
d) harmonična sredina, - H
e) geometrijska sredina, - G

Medtem ko sta mediana in modus dana z lego vrednosti, štejemo aritmetično, harmonično in geometrijsko sredino med izračunane srednje vrednosti.

MEDIANA

9.3 Za srednjo vrednost more dobro rabiti mediana, ki jo poznamo že iz poglavja o kvantilih. Mediana je vrednost, ki ustreza kvantilnemu rangu P = 0,50,

$$Me = y_{P=.50}$$

Mediana je uporabna srednja vrednost, ker nujno leži sredi posamičnih vrednosti. Po opredelitvi ima namreč polovica enot manjše, polovica pa večje vrednosti kakor je mediana.

Določanje mediane ne bomo obravnavali posebej, ker se sklada z določanjem kvantilov na splošno. Te pa smo obravnavali že v odstavku o kvantilih.

9.4 Lastnosti mediane. Prednost mediane je predvsem v tem, da je lahko razumljiva srednja vrednost in zato kot opisni parameter priporočljiva. Velika prednost mediane pred izračunanimi sredinami je tudi ta, da za določanje mediane ni nujno, da poznamo vrednosti za vse enote populacije. Zadosti je, da poznamo vrednosti za enote, ki leže okoli sredine v ranžirni vrsti. Ta lastnost mediane pride posebno prav, če iščemo srednje vrednosti za frekvenčne porazdelitve, ki imajo odprte razrede.

Za odprte razrede namreč dostikrat niti približno ne vemo, kakšne vrednosti vsebujejo, ker so omejeni samo navzgor ali samo navzdol. Mediana je primerna srednja vrednost tudi takrat, če so skrajne vrednosti take, da sumimo, ali sploh sodijo v homogeno populacijo, ker so mogoče izraz nekih drugih kakovosti.

Pomanjkljivost mediane pa je v tem, da je le preveč neobčutljiva za spremembe vrednosti in se njena vrednost ne spremeni vse dotlej, dokler so spremembe take, da vrednosti ne preidejo iz ene polovice v drugo.

Pomembna lastnost mediane je tudi, da je vsota absolutnih odklonov posameznih vrednosti najmanjša, če odklone računamo od mediane. Mediana je torej sorazmeroma dober predstavnik posameznih vrednosti

$$\sum_{i=1}^{N} |y_i - A| = \min \; ; \quad \text{če je } A = Me \tag{9.1}$$

MODUS

9.5 Mediana pa ni vedno dober predstavnik vrednosti populacije. Za asimetrične ali polimodalne porazdelitve je mediana vrednost, ki je različna od večine vrednosti v populaciji. Če naj bo srednja vrednost predstavnik populacije, je v takih primerih veliko primerneje, da vzamemo za sredino vrednost, okrog katere se vrednosti populacije najbolj goste. Če pogledamo poligone frekvenčnih porazdelitev, moremo za vsako frekvenčno porazdelitev zlahka ugotoviti približno mesto največje gostitve. Vrednost, okrog katere se vrednosti populacije najbolj goste, imenujemo m o d u s ali n a j p o g o s t e j š o v r e d n o s t.

Modus moremo ugotoviti samo za razmeroma obsežne populacije, ki so grupirane v frekvenčni porazdelitvi. Iz posamičnih podatkov je modus neposredno nemogoče ugotoviti.

Če imamo frekvenčno porazdelitev dano s frekvenčno krivuljo, poiščemo modus enostavno. V tem primeru je modus abscisa tiste točke na frekvenčni krivulji, za katero je ordinata (gostota frekvence) največja.

9.6 R a č u n a n j e m o d u s a i z f r e k v e n č n e p o r a z d e l i t v e. Za populacije običajno nimamo frekvenčnih krivulj, temveč frekvenčne porazdelitve, ki jih moremo narisati v histogramu. Ker histogram nakaže približno obliko frekvenčne krivulje, moremo tudi iz frekvenčne porazdelitve oziroma histograma sklepati, kje je gostota frekvence največja. Iz histograma za frekvenčno porazdelitev z enakimi razdeli sklepamo, da je modus v razredu, v katerem je frekvenca največja. Za porabo lesa v 149 kmetijskih gospodarstvih v letu 1953 v Novem mestu iz tabele 7.3 sklepamo, da je najpogostejša vrednost v razredu od 12,0 do 15,9 m^3 lesa. Kot prvi približek modusa vzamemo kar sredino razreda z največjo frekvenco - sredino modalnega razreda. V našem zgledu je prva ocena modusa za porabo lesa $Mo = 14\ m^3$. Sredina razreda pa je prava vrednost modusa le, če je porazdelitev simetrična. Drugače leži modus nad sredino razreda ali pod njo, odvisno od oblike frekvenčne porazdelitve.

Natančnejšo oceno za modus dobimo, če upoštevamo tudi frekvence v sosednih razredih. Iz histograma o porabi lesa v sliki 7.1 sklepamo, da je modus za porabo lesa verjetno pod sredino modalnega razreda s sredino 14 m^3. Ker je frekvenca pred modalnim razredom večja kot frekvenca za modalnim razredom, bi bila frekvenčna krivulja, ki bi jo načrtali med stolpce histograma, nagnjena proti razredu z višjo frekvenco. Predpostavimo, da je v okolici modusa frekvenčna krivulja v približku parabola druge stopnje, ki gre skozi točke s koordinatama $(y_{-1} \cdot f_{-1})$, $(y_o \cdot f_o)$ in $(y_{+1} \cdot f_{+1})$. Pri tem pomenijo y_{-1}, y_o, y_{+1} sredine modalnega razreda in njegovih sosedov, f_{-1}, f_o in f_{+1} pa ustrežne frekvence. Pri tej predpostavki izračunamo modus po naslednjem postopku:

a) Podatke imamo grupirane v frekvenčni porazdelitvi z razredi z enako širino i. Razredi morajo biti tako veliki, da frekvenčna porazdelitev izraža zakonitost gostitve frekvenc.

b) V frekvenčni porazdelitvi poiščemo razred z največjo frekvenco $f_{-1} < f_o > f_{+1}$. Razred z največjo frekvenco (o) imenujmo m o d a l n i r a z r e d.

c) Modus izračunamo po obrazcu

$$M_o = y_{o.min} + i \frac{f_o - f_{-1}}{2f_o - f_{-1} - f_{+1}} \tag{9.2}$$

Pri tem pomeni razen že navedenih izrazov: $y_{0,min}$ = spodnja meja za modalni razred.

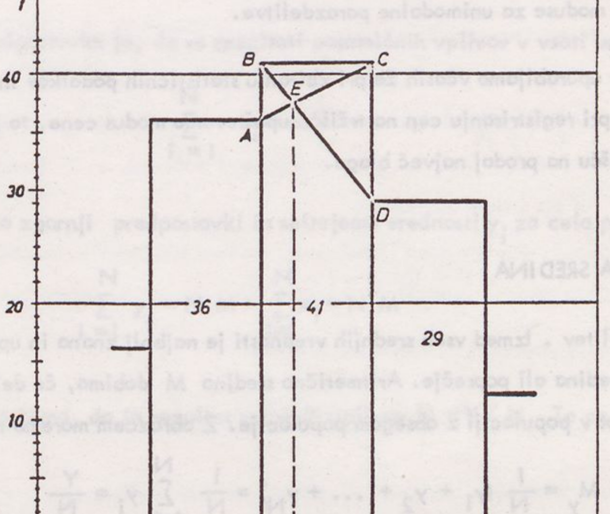
9.7 Za zgled o porabi lesa v 149 kmetijskih gospodarstvih v letu 1953 v Novem mestu je od treh porazdelitev, ki smo jih sestavili, za izračunavanje modusa prikladna porazdelitev, v kateri je razredna širina $i = 4\ m^3$. V porazdelitvi z $i = 1\ m^3$ se zakonitost gostitve še ne kaže, v porazdelitvi z $i = 8\ m^3$ pa je zaradi prevelike širine razredov zakonitost gostitve že zabrisana.

Iz frekvenčne porazdelitve v tabeli 7.3 sklepamo, da je modalni razred razred 12,0 - 15,9 m^3, ker ima največjo frekvenco ($f_o = 41$). Ker je za to porazdelitev $f_{-1} = 36$, $f_{+1} = 29$; $i = 4$; $y_{o.min} = 12,0$, je modus po obrazcu 9.2

$$Mo = 12,0 + 4 \cdot \frac{41 - 36}{2.41 - 36 - 29} = 13.18\ m^3$$

Kot smo pričakovali, je druga ocena za modus $Mo = 13,18\ m^3$ pod prvo oceno $14,0\ m^3$.

9.8 Modus pa moremo ob istih predpostavkah zelo enostavno oceniti iz histograma za frekvenčno porazdelitev tudi grafično. Kakor kaže slika 9.1, zvežemo v histogramu



Slika 9.1 Grafična določitev modusa iz frekvenčne porazdelitve za porabo lesa v 149 kmetijskih gospodarstvih v Novem mestu

oglišče A s C, oglišče B z D. Projekcija sečišča obeh veznic E na abscisi je ocena modusa Mo. Slika pokaže, da grafičen način dá isti rezultat kakor račun.

9.9 Lastnosti modusa. Modus je srednja vrednost, ki dobro predstavlja vrednosti za populacijo, ker je že po opredelitvi vrednost, ki se v populaciji najpogosteje pojavlja. Enako kakor mediana je modus srednja vrednost, ki je dana z lego vrednosti populacije. Zato je tudi modus neobčutljiv za spremembe vrednosti posameznih enot vse dotlej, dokler gostitev na nekem drugem mestu ne prekorači stopnje gostitve v modusu. To moremo šteti za dobro lastnost modusa, ker ni odvisen od vrednosti, ki za populacijo niso tipične, more pa biti tudi hiba, ker je modus le premalo odvisen od posamičnih vrednosti.

Nehomogena populacija more imeti tudi več mest gostitve. Bimodalne porazdelitve imajo dva, polimodalne porazdelitve pa celo več središč gostitve. Take porazdelitve imajo torej dva ali več relativnih - lokalnih modusov. Od teh pa je absolutni modus tisti, za katerega je gostitev največja. Relativne - lokalne moduse ocenjujemo enako kot moduse za unimodalne porazdelitve.

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

ARITMETIČNA SREDINA

9.10 Opredelitev. Izmed vseh srednjih vrednosti je najbolj znana in uporabljana aritmetična sredina ali poprečje. Aritmetično sredino M dobimo, če delimo vsoto vrednosti vseh enot v populaciji z obsegom populacije. Z obrazcem moremo to izraziti

$$M_y = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{Y}{N} \qquad (9.3)$$

Pri tem pomeni: M_y = aritmetična sredina. Znak M_y včasih zamenjamo z znakom $\bar{y}$ (y prečna) ; $\sum_{i=1}^{N} y_i = Y$ = vsota vseh vrednosti v populaciji; $\sum$ = splošen znak za seštevanje izraza, ki stoji za njim; y_i = posamične vrednosti. Nakazane oznake uporabljamo na splošno.

9.11 Za zgled vzemimo podatke o številu izdelkov, ki jih je skupina desetih delavcev izdelala v eni izmeni. Če so posamični podatki:

38, 35, 42, 47, 39, 48, 32, 42, 44, 50, je poprečno število izdelkov po obrazcu 9.3

$$M = \frac{38 + 35 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots + 44 + 50}{10} = 41.7 \text{ kosov}$$

Poprečno število $M =$ 41,7 pove, da bi moral vsak delavec izdelati po 41,7 kosov, če bi hoteli, da bi bila skupna proizvodnja enaka stvarni.

9.12 Lastnosti aritmetične sredine

a) Aritmetična sredina je izpeljana ob predpostavki, da je vrednost y za posamezno enoto vsota rezultatov splošnih (M) in posamičnih vplivov (e)

$$y_i = M + e_i \tag{9.4}$$

Nadaljnja predpostavka je, da se rezultati posamičnih vplivov v vsoti uničijo

$$\sum_{i=1}^{N} e_i = o.$$

Če upoštevamo zgornji predpostavki in seštejemo vrednosti y_i za celo populacijo, dobimo

$$\sum_{i=1}^{N} y_i = N.M + \sum_{i=1}^{N} e_i = N.M$$

Iz te enačbe dobimo, da je rezultat splošnih vplivov $M = Y / N$. To pa je aritmetična sredina.

b) Aritmetična sredina za linearno zvezo iz več znakov

$$u = a_o + \sum_{k=1}^{r} a_k y_k \tag{9.5}$$

je enaka linearni zvezi iz aritmetičnih sredin

$$u = a_o + \sum_{k=1}^{r} a_k y_k \; ; \; M_u = a_o + \sum_{k=1} a_k M_k \qquad (9.6)$$

Dokaz je dokaj enostaven. Če enačbo

$$u = \sum_{i=1}^{N} u_i = \sum_i (a_o + \sum_k a_k y_{ki}) = a_o N + \sum_k a_k \sum_i y_{ki} = a_o N + \sum_k a_k Y_k$$

ki jo dobimo, če seštejemo linearne enačbe 9.5 za vse enote populacije, delimo z obsegom populacije N, dobimo zvezo 9.6. Kot poseben primer zveze 9.6 je

$$M_a = a \qquad (9.7)$$

Poprečje konstante je enako konstanti in

$$M_{by} = b.M_y \qquad (9.8)$$

poprečje znaka, pomnoženega s konstanto, je enako produktu konstante s poprečjem znaka.

Vzemimo za zgled, da je skupni dohodek D delavca sestavljen iz osnovnega dela 0, dohodka iz rednega delovnega časa R, za katere je urni dohodek r, in izrednega dela, ki ga je opravljal P ur po urnem dohodku p

$$D = 0 + R \,.\, r + P \,.\, p$$

Po gornjem stavku je poprečni dohodek na delavca

$$M_D = M_o + R M_r + P M_p$$

c) Vsota odklonov posamičnih vrednosti y_i od aritmetične sredine M_y je nič

$$\sum_{i=1}^{N} (y_i - M_y) = 0 \qquad (9.9)$$

Dokaz je preprost. Ker je vsota razlik enaka razliki vsot, spoznamo, da je

$$\sum y_i - N \,.\, M_y = 0$$

v skladu z opredelitvijo aritmetične sredine

d) Vsota kvadratov odklonov posamičnih vrednosti y_i od neke konstante A je najmanjša, če je A enak aritmetični sredini M_y.

$$SK = \sum_{i=1}^{N} (y_i - A)^2 : \quad \text{če je: } A = M_y \tag{9.10}$$

Dokaz: Če naj bo izraz SK minimalen, mora biti $\frac{dSK}{dA} = 0$

$$\frac{dSK}{dA} = -2\sum_i (y_i - A) = 0 ; \quad \sum (y_i - A) = \sum y_i - NA = 0 ; \quad A = \frac{1}{N}\sum y_i = M_y$$

Ta lastnost za aritmetično sredino je zelo važna za nadaljnje proučevanje populacij. Iz vsote kvadratov odklonov od aritmetične sredine je namreč izpeljan poseben parameter, v a r i a n c a , ki je najvažnejša mera jakosti za posamične vplive.

e) Sumarno aritmetično sredino populacije M izračunamo iz aritmetičnih sredin M_k delnih populacij z obsegi N_k po obrazcu.

$$M = \frac{N_1M_1 + N_2M_2 + \dots + N_rM_r}{N_1 + N_2 + \dots + N_r} = \frac{\sum_{k=1}^{r} N_kM_k}{\sum_{k=1}^{r} N_k} = \frac{1}{N}\sum_k N_kM_k \tag{9.11}$$

Ta način za računanje aritmetične sredine imenujemo t e h t a n o r a č u n a n j e za razliko od enostavnega računanja po obrazcu 9.3. Dokaz: Skupna vsota Y je enaka vsoti delnih vsot Y_k za y

$$Y = Y_1 + Y_2 + \dots + Y_r = \sum_k Y_k \quad .$$

Ker velja za skupno populacijo in delne populacije $Y = NM_y$, je dalje:

$$NM = N_1M_1 + N_2M_2 + \dots + N_rM_r = \sum_k N_kM_k$$

Če delimo to enačbo z N dobimo obrazec 9.11. Splošnejše moremo tehtan način za izračunanje aritmetične sredine pisati

$$M = \frac{\sum_k w_k}{\sum_k w_k} \qquad (9.12)$$

Pri tem pomeni: w_k = teža - ponder

Računanje aritmetične sredine iz frekvenčnih porazdelitev

9.13 Neposredno računanje aritmetične sredine po osnovnem obrazcu 9.3 je za velike populacije zaradi velikega števila sumandov zamudno. Ker daje frekvenčna porazdelitev približno sliko vseh vrednosti populacije, moremo aritmetično sredino oceniti iz frekvenčne porazdelitve.

Če predpostavljamo, da sredina razreda y_k v frekvenčni porazdelitvi predstavlja vrednosti v posameznem razredu, dobimo oceno za vsoto vrednosti v tem razredu, če sredino razreda pomnožimo s frekvenco f_k. Oceno za vsoto vrednosti za celo populacijo pa dobimo, če produkte $f_k y_k$ za vse razrede seštejemo. Če to uporabimo, računamo oceno aritmetične sredine iz frekvenčne porazdelitve po obrazcu

$$M_y = \frac{f_1 y_1 + f_2 y_2 + \ldots + f_r y_r}{f_1 + f_2 + \ldots + f_r} = \frac{\sum f_k y_k}{\sum f_k} = \frac{1}{N} \sum_{k=1}^{r} f_k y_k \qquad (9.13)$$

Ta obrazec je samo posebna oblika splošnega obrazca 9.12 za tehtan način računanja aritmetične sredine. Pri tem so frekvence teže ali ponderi.

Po obrazcu 9.13 dobljena ocena je tem boljša, čim ožji so razredi, vendar dobimo tudi pri razmeroma širokih razredih še vedno zadovoljive rezultate, če je frekvenčna porazdelitev unimodalna in ne preveč asimetrična. Za zelo asimetrične porazdelitve, posebno za porazdelitve tipa J, pa daje tako računanje sistematično napačne ocene. Če je porazdelitev asimetrična v desno, dobimo sistematično prevelike ocene, pri asimetriji v levo pa sistematično premajhne ocene.

Računanje aritmetične sredine po obrazcu 9.13 velja za porazdelitve z enakimi in z neenakimi razredi. Ne moremo pa po tem obrazcu ocenjevati aritmetične sredine za frekvenčne porazdelitve z odprtimi razredi, ker zanje ne moremo izračunati sredine razredov. Če za odprt razred poznamo razen frekvence f_r še vsoto vrednosti v odprtem razredu Y_r, izračunamo oceno za aritmetično sredino po obrazcu

$$M_y = \frac{1}{N} (f_1 y_1 + f_2 y_2 + \ldots + f_{r-1} y_{r-1} + Y_r) \qquad (9.14)$$

9.14 Računanje ocene aritmetične sredine po neposredni metodi je v tabeli 9.1 nakazano za frekvenčno razdelitev za porabo lesa v letu 1953 za kmetijska gospodarstva v Novem mestu.

Tabela 9.1 Računanje ocene aritmetične sredine za potrošnjo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto v letu 1953 po neposredni metodi.

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

Po obrazcu 9.13 je $M_y = 2190/149 = 14,70\ m^3$.

Če primerjamo dobljeni rezultat s pravo aritmetično sredino $M_y = 2182,1/149 = 14,64\ m^3$, izračunano iz osnovnih podatkov v tabeli 7.1, vidimo, da je razlika malenkostna.

9.15 Pomožni znak u. Za frekvenčne porazdelitve, ki imajo enake razrede, računanje aritmetične sredine poenostavimo, če vpeljemo namesto sredin razredov y_k pomožni znak u_k, ki je z osnovnim znakom y_k v naslednji linearni zvezi:

$$y_k = y_o + i \cdot u_k \qquad (9.15)$$

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

Če natančneje proučimo vrednosti znaka u, spoznamo, da sredinam razredov v posameznih razredih ustrezajo u ... − 3, −2, −1, 0, +1, +2, +3... Pri tem je vrednost u = o v razredu, ki mu ustreza y_o.
Zaradi obrazca 9.13 in 9.6 pa velja

$$M_y = y_o + iM_u \qquad (9.16)$$

Ta postopek je prikladnejši kakor neposredna metoda, ker poenostavi množenje. Po tem postopku izračunamo aritmetično sredino po tehle točkah.

a) V dani frekvenčni porazdelitvi izberemo nekje v sredini ali v razredu, okoli katerega so frekvence največje, izhodišče za pomožni znak u. Glede na to poljubno izhodišče postavimo v posamezne razrede ustrezne vrednosti pomožnega znaka u

... −3, −2, −1, 0, + 1, +2, +3 .....

b) Pomnožimo frekvence f_k z ustreznimi vrednostmi u_k. Tako dobimo produkte $f_k u_k$.

c) Seštejemo dobljene produkte in vsoto $\sum f_k u_k$ vnesemo v obrazec

$$M_y = y_o + i \cdot \frac{1}{N} \sum f_k u_k \qquad (9.17)$$

9.16 Za porabo lesa v Novem mestu je izračunanje aritmetične sredine s pomožnim znakom u nakazano v tabeli 9.2.

Tabela 9.2 Računanje ocene za aritmetično sredino za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi pomožnega znaka u.

Poraba v m³ y	f_k	u_k	$f_k u_k$
0,0 - 3,9	2	− 3	− 6
4,0 - 7,9	16	− 2	− 32
8,0 - 11,9	36	− 1	− 36
12,0 - 15,9	41	0	0
16,0 - 19,9	29	+ 1	+ 29
20,0 - 23,9	12	+ 2	+ 24
24,0 - 27,9	7	+ 3	+ 21
28,0 - 31,9	4	+ 4	+ 16
32,0 - 35,9	2	+ 5	+ 10
	N = 149		$\sum fu$ = + 26

Ker je $y_o = 14,0$, $i = 4$, $N = 149$ in $\sum fu = +26$, dobimo po obrazcu 9.17

$$M_y = 14,0 + 4 \cdot \frac{1}{149} (+ 26) = 14,70 \; m^3$$

Rezultat se sklada z rezultatom, ki smo ga dobili po neposredni metodi.

9.17 Metoda kumulativ. Če so podatki grupirani v frekvenčni porazdelitvi z enakimi razredi, aritmetično sredino enostavno izračunamo s kumulativami. Pri tej metodi odpade vmesno množenje. Razen tega pa imamo kot postranski rezultat izračunano kumulativno frekvenčno porazdelitev; ta pa je važna sama zase ali za računanje kvantilov.

Po metodi kumulativ izračunamo aritmetično sredino po naslednjem postopku:

a) Iz frekvenčne porazdelitve izračunamo iz f_k kumulativno frekvenčno porazdelitev F_k.

b) Seštejemo člene v kumulativni vrsti, razen zadnjega, ki leži pod črto, ki pomeni obseg populacije N. Vsoto členov iz kumulativne vrste zaznamujemo s C.

c) Aritmetično sredino M_y izračunamo iz dobljenih podatkov po obrazcu

$$M_y = y_o - i \cdot \frac{C}{N} \qquad (9.17)$$

Pri tem pomeni: y_o = sredina zadnjega razreda v frekvenčni porazdelitvi; i = širina razreda; C = vsota členov v kumulativni vrsti; N = obseg populacije.

9.18 Aritmetična sredina za porabo lesa v letu 1953 v Novem mestu je izračunana po metodi kumulativ v tabeli 9.3.

Tabela 9.3 Izračunanje aritmetične sredine za porabo lesa v 149 kmetijskih gospodarstvih v letu 1953 po metodi kumulativ.

Poraba v m^3	f	F
0,0 - 3.9	2	0
4,0 - 7,9	16	2
8,0 - 11,9	36	18
12,0 - 15,9	41	54
16,0 - 19,9	29	95
20,0 - 23,9	12	124
24,0 - 27,9	7	136
28,0 - 31,9	4	143
32,0 - 35,9	2	147
	N = 149	

$2+18+54+\ldots+143+147 = 719 = C$

Ker je sredina zadnjega razreda $y_o = 34$, dobimo po obrazcu 9.17

$$M_y = 34{,}0 - 4.719/149 = 14{,}70\ m^3$$

Dobljeni rezultat se sklada z rezultatom, ki smo ga dobili po prejšnjih metodah.

Modificirana aritmetična sredina

9.19 Aritmetična sredina je odvisna od vseh vrednosti populacije. Zato nanjo vplivajo tudi skrajne vrednosti, ki so včasih rezultat izjemnih okoliščin in jih ne moremo šteti, da sodijo v prikazano populacijo. Te vrednosti pačijo sliko populacije in tudi aritmetično sredino populacije. Zato za take primere izračunamo sredino, ki je nekak kompromis med aritmetično sredino in mediano.

Iz ranžirne vrste dobimo mediano, če postopoma paroma izpuščamo po en spodnji in en zgornji člen v ranžirni vrsti. Na koncu tega postopka ostane samo en člen - mediana, ali dva člena, katerih sredino štejemo za mediano. Kompromisna rešitev pa je v tem, da izpuščamo po en zgornji in en spodnji člen le toliko časa, dokler ne izključimo vseh netipičnih vrednosti, iz ostanka pa izračunamo aritmetično sredino. Tako odstranimo vpliv skrajnih - izjemnih vrednosti, ki včasih občutno vplivajo na aritmetično sredino. Kljub vsemu pa je ta sredina izračunana iz večine vrednosti populacije. Modificirano aritmetično sredino dobimo torej tako, da iz ranžirne vrste posamičnih vrednosti odstranimo po en, dva ali več parov skrajnih - izjemnih vrednosti, iz ostalih pa izračunamo poprečje. Ta postopek uporabljamo pogosto pri analizi časovnih vrst. Zaradi izjemnih razmer v določenih razdobjih nekateri podatki pačijo tipičnost aritmetične sredine. Če pa teh skrajnih vrednosti ne upoštevamo in uporabimo modificirano poprečje, dobimo realnejšo sliko.

Aritmetična sredina aritmetičnih sredin

9.20 Če poznamo aritmetične sredine M_k in obsege N_k za delne populacije, ki sestavljajo populacijo, iz teh podatkov izračunamo aritmetično sredino M za populacijo po obrazcu

$$M = \frac{\sum N_k M_k}{\sum N_k} \qquad (9.18)$$

Ta obrazec smo dokazali v odstavku o lastnostih aritmetičnih sredin. Lastnosti, da moremo izračunati skupno srednjo vrednost, če poznamo ustrezne srednje vrednosti za delne populacije, nimata niti mediana niti modus. Pri obeh je treba iz delnih populacij sestaviti skupno populacijo in iz nje poiskati mediano ali modus.

Zgornje računanje za aritmetično sredino imenujemo 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 9,18 je poseben primer splošnega obrazca za računanje tehtane aritmetične sredine

$$M_R = \frac{\sum w_k R_k}{\sum w_k} \qquad (9.19)$$

Pri tem pomeni: M_R = tehtana aritmetična sredina količin R_k, w_k = ponder-teža za posamezne vrednosti R_k.

9.21 V tabeli 9.4 je nakazano računanje skupnih poprečnih mesečnih prejemkov delavcev v elektrogospodarstvu, če poznamo poprečne mesečne prejemke in skupno število delavstva po vrstah podjetij.

Tabela 9.4. Računanje skupnih poprečnih mesečnih prejemkov zaposlenega osebja v proizvodnji, prenosu in distribuciji električne energije.

Stroka	Število zaposlenih N_k	Ø dohodki IX - 1970 M_k	$N_k M_k$
Hidroelektrarne	895	2349	2102355
Termoelektrarne	1292	1736	2242912
Podjetje za prenos elektroenergije	170	2211	375870
Podjetje za razdelitev elektroenergije	3740	1738	6500120
Skupno	$\sum N_k = 6097$	1840	$\sum M_k N_k = 11221257$

Poprečni dohodek za stroko je

$$M = \frac{\sum N_k M_k}{\sum N_k} = \frac{11221257}{6097} = 1840 \text{ din}$$

Popolnoma nepravilno bi bilo, če bi izračunali navadno aritmetično sredino iz podatkov o mesečnih prejemkih po kvalifikaciji. Rezultat, ki bi ga dobili,

$M_{nav} = 1/4\ (2349 + 1736 + 2211 + 1738) = 2008.5$ je seveda različen od zgornjega in brez logičnega smisla, medtem, ko pravo sumarno poprečje pove, kakšni bi bili prejemki delavcev v elektrogospodarstvu, če bi imeli vsi enake prejemke pri istem fondu dohodkov.

Sumarna poprečja pa moramo uporabljati zelo previdno. V splošnem sumarne aritmetične sredine ne predstavljajo vrednosti v populaciji. Sumarno poprečje more biti celo vrednost, ki jo ima malo število enot. Vendar kljub hibam dostikrat izračunavamo sumarna poprečja, ki so, če jih pravilno tolmačimo, dobro sredstvo za analizo.

HARMONIČNA SREDINA

9.22 Med izračunane srednje vrednosti štejemo tudi harmonično sredino, ki je po definiciji recipročna vrednost aritmetične sredine reciprokov iz osnovnih podatkov. Po tej opredelitvi je harmonična sredina H enaka

$$H = \frac{N}{\frac{1}{y_1} + \frac{1}{y_2} \dots + \frac{1}{y_N}} = \frac{N}{\sum \frac{1}{y_i}} \tag{9.20}$$

Tehtana harmonična sredina pa podobno

$$H = \frac{w_1 + w_2 + \dots + w_r}{\frac{w1}{y_1} + \frac{w2}{y_2} + \dots + \frac{w_r}{y_r}} = \frac{\sum w_k}{\sum \frac{w_k}{y_k}} \tag{9.21}$$

9.23 Če iz istih podatkov izračunamo aritmetično sredino in harmonično sredino, dobimo različne rezultate. Za podatke 1,2,4,7,9, je aritmetična sredina enaka

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

harmonična sredina pa

$$H = \frac{5}{\frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{7} + \frac{1}{9}} = 2,5$$

Iz shematičnega zgleda vidimo, da so razlike znatne.

Harmonično sredino uporabljamo redkeje kakor aritmetično. Kadar se vrednosti populacije porazdeljujejo v asimetrični porazdelitvi, a tako, da je porazdelitev recipročnih vrednosti simetrična, pa harmonična sredina bolje kaže osrednjo težnjo kakor aritmetična sredina. Zato v takih primerih dajemo prednost harmonični sredini.

9.24 S harmonično sredino računamo včasih tudi poprečja iz relativnih števil.

Vzemimo, da je pet delavcev delalo po eno uro. Pri tem so dosegli naslednjo proizvodnost dela: 4,5,2,6,3 minute za en izdelek. Proizvodnost dela merimo s časom, ki so ga potrebovali za izdelavo enega izdelka.

Sumarni pokazovalec o produktivnosti dela za vseh pet delavcev dobimo, če skupno porabljen čas delimo s številom proizvedenih izdelkov. Vseh pet delavcev je delalo skupno $5.60 = 300$ minut. Število izdelkov, ki jih je izdelal posamezen delavec, dobimo, če za vsakega delavca čas (60 minut) delimo s časom, ki ga je potreboval za izdelavo enega izdelka (pokazovalec produktivnosti dela). Za posamezne delavce dobimo po vrsti:

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

Skupno število izdelkov pa dobimo, če te podatke seštejemo. Poprečna produktivnost dela je torej

$$\frac{5.60}{60/4 + 60/5 + 60/2 + 60/6 + 60/3} = \frac{300}{87} = 3,45 \text{ minut/izdelek}$$

Iz računa vidimo, da smo izračunali poprečno proizvodnost dela s harmonično sredino. Če v gornjem računu krajšamo s ponderom 60, dobimo

$$\frac{5}{1/4 + 1/5 + 1/2 + 1/6 + 1/3} = 3,45 \text{ minut/izdelek}$$

Ker so ponderi med seboj enaki, dobimo isti rezultat, če vzamemo tehtano ali netehtano obliko.

Če iz gornjih podatkov izračunamo aritmetično sredino, pa dobimo

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

Preizkus pokaže, da harmonična sredina da pravi prezultat. Vseh pet delavcev, ki so delali skupno 300 minut, je izdelalo 87 izdelkov. Če bi vsi delavci delali s poprečno proizvodnostjo dela 3,45 minut za izdelek, bi napravili enako število $300/3,45 = 87$ izdelkov. To pa je v skladu z opredelitvijo poprečja. Če bi vseh pet delavcev delalo z enotno produktivnostjo dela 4 minute za kos, ki smo jo dobili z aritmetično sredino, pa bi v skupno 300 minutah izdelali le 75 artiklov in ne 87, kolikor so jih v resnici.

Podrobneje o uporabi harmonične sredine pri računanju poprečnih relativnih števil bomo zvedeli v naslednjem odstavku.

POPREČJA IZ RELATIVNIH ŠTEVIL

9.25 Problem računanja poprečij iz relativnih števil je za socialno-ekonomsko statistiko tako osnoven in pomemben, da mu bomo posvetili poseben odstavek.

Relativna števila so v splošnem kvocienti primerjanih količin. Če vzamemo, da je grupno relativno število R_k razmerje dveh absolutnih podatkov Y_k in X_k

$$R_k = Y_k / X_k \tag{9.22}$$

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

$$Y_k = X_k R_k \tag{9.23}$$

in

$$X_k = Y_k / R_k \tag{9.24}$$

Če upoštevamo te zveze, računamo sumarno relativno število R na tri načine:

a) Sumarno relativno število je razmerje med vsotami absolutnih grupnih vrednosti za X_k in Y_k.

Po tej opredelitvi je sumarno relativno število R enako

$$R = \frac{Y}{X} = \frac{\sum Y_k}{\sum X_k} \tag{9.25}$$

S tem obrazcem računamo sumarno relativno število, če imamo po grupah a b s o l u t n e p o d a t k e za Y_k in X_k.

b) Če upoštevamo obrazec 9.23, je sumarno relativno število

$$R = \frac{\sum X_k R_k}{\sum X_k} \tag{9.26}$$

t e h t a n a a r i t m e t i č n a s r e d i n a grupnih relativnih števil R_k. Pri tem so vrednosti, ki so v imenovalcu relativnega števila, ponderi. Tako računamo poprečno relativno število, če razpolagamo z grupnimi relativnimi števili in vrednostmi, ki nastopajo v imenovalcu relativnih števil, oziroma če je smiseln produkt pondera z relativnim številom $X_k \cdot R_k$.

c) Če upoštevamo obrazec 9.24, pa sumarno relativno število izračunamo po obrazcu

$$R = \frac{\sum Y_k}{\sum Y_k / R_k} \qquad (9.27)$$

To pa je tehtana harmonična sredina grupnih relativnih števil R_k. Pri tem so ponderi količine, ki nastopajo v relativnem številu v števcu. Po tem obrazcu računamo sumarno relativno število, če razpolagamo z grupnimi relativnimi vrednostmi R_k in grupnimi količinami Y_k, ki nastopajo v relativnem številu v števcu, oziroma če je smiseln kvocient pondera in relativnega števila Y_k/R_k.

Obrazci 9.25, 9.26 in 9.27 oziroma gornja pravila ne veljajo samo za relativna števila v ožjem smislu, temveč tudi za računanje sumarnih aritmetičnih sredin, saj so aritmetične sredine tudi relativna števila, ker so razmerja med vsoto vrednosti in številom enot.

9.26 Po zgornjih pravilih ugotovimo, kdaj uporabimo za računanje sumarnih relativnih števil tehtano aritmetično ali tehtano harmonično sredino.

Po teh pravilih ugotovimo, da je treba v primeru produktivnosti dela, ki smo ga navedli v odstavku 9.24, računati harmonično sredino. Produktivnost dela je merjena z relativnim številom "čas, porabljen za izdelavo enega izdelka", ki ga dobimo, če porabljen čas delimo s številom izdelkov. Ker razpolagamo s časom, koliko so posamezni delavci delali (60 minut), so ponderi količine, ki v relativnem številu nastopajo v števcu. Po gornjih pravilih je treba za sumarno relativno število uporabiti harmonično sredino.

Če poznamo površine in poprečno gostoto prebivalstva po republikah, izračunamo gostoto prebivalstva za SFRJ kot tehtano aritmetično sredino, ker so ponderi površine v koeficientu "gostota prebivalstva" v imenovalcu.

Če imamo po podjetjih dano število zaposlenih žensk in odstotek zaposlenih žensk od skupnega števila zaposlenih, izračunamo poprečen odstotek zaposlenih žensk za celo stroko po obrazcu za računanje harmonične sredine. Relativno število "odstotek zaposlenih žensk" je kvocient med številom zaposlenih žensk in številom skupno zaposlenih v podjetju. Ponder - število zaposlenih žensk - je v tem primeru v števcu relativnega števila.

Za podjetje imamo koeficiente o obračanju zalog posameznih vrst surovin (merjeno s časom enega obrata za posamezno vrsto surovine) in ustrezne vrednosti porabljenih surovin. Sumarni koeficient obračanja zalog izračunamo po obrazcu za izračunanje tehtane aritmetične sredine. Pokazovalec o obračanju zalog "čas enega obrata" je namreč kvocient med poprečnimi zalogami in vrednostjo porabljenih surovin. Ponder - vrednost porabljenih surovin - pa je v relativnem številu v imenovalcu.

Če imamo po občinah podatke o površini in skupnem pridelku pšenice v določenem letu, izračunamo popreček hektarski pridelek pšenice po obrazcu 9.25 tako, da vsoto pridelkov po občinah delimo z vsoto površin.

Odvisnost sumarnega relativnega števila od sestave ponderov

9.27 Če preuredimo obrazec 9.26, dobimo

$$R = \frac{\sum X_k R_k}{X} = \sum \frac{X_k}{X} . R_k = \sum X_k^o R_k \qquad (9.28)$$

Podobno dobimo iz obrazca 9.27

$$R = \frac{Y}{\sum Y_k / R_k} = \frac{1}{\sum \frac{Y_k}{Y} / R_k} = \frac{1}{\sum Y_k^o / R_k} \qquad (9.29)$$

Pri tem pomeni: $X_k^o = X_k / X$ strukturni delež za grupni podatek X_k ; $Y_k^o = Y_k / Y$ strukturni delež za grupni podatek Y_k.

Iz teh dveh obrazcev sklepamo, da sumarno relativno število ni odvisno samo od grupnih relativnih števil, R_k, temveč tudi od sestave ustreznih ponderov X_k^o ali Y_k^o.

To je ena izmed osnovnih hib sumarnih relativnih števil, zaradi katere imamo včasih zelo resne pomisleke o uporabnosti sumarnih pokazovalcev. Za populaciji, ki imata enaka grupna relativna števila, sta sumarni relativni števili med seboj različni, če sta sestavi ponderov različni. Še več. Razlike v sestavi ponderov morejo povzročiti, da je sumarno relativno število za populacijo A večje kot za populacijo B, čeprav so vsa grupna relativna števila za populacijo A manjša kot za populacijo B.

9.28 Vzemimo za zgled poprečne mesečne prejemke delavcev v kmetijstvu in gradbeništvu po kvalifikaciji.

Tabela 9.5. Poprečni prejemki delavcev v kmetijstvu in gradbeništvu po kvalifikaciji v letu 1957 v SFRJ (Vir SB 114).

Kvalifikacija	Poprečni mesečni prejemki v dinarjih		Sestava o številu delavcev	
	Kmetijstvo	Gradbeništvo	Kmetijstvo	Gradbeništvo
Visoko kvalificirani	15310	15200	0,020	0,078
Kvalificirani	12430	11920	0,218	0,298
Priučeni	9530	8950	0,157	0,218
Nekvalificirani	8030	6770	0,605	0,406
Skupno	8910	9690	1,000	1,000

Iz tabele 9.5 vidimo, da so poprečni prejemki delavcev v kmetijstvu za vse kvalifikacije višji kot v gradbeništvu. Kljub temu pa so sumarni poprečni prejemki v kmetijstvu nižji kot v gradbeništvu. Vzrok je različna sestava za število delavstva v kmetijstvu in gradbeništvu. Iz zadnjih dveh stolpcev v tabeli 9.5. vidimo, da je sestava delavstva po kvalifikaciji v poljedelstvu znatno nižja kot v gradbeništvu. Ta razlika v sestavi delavstva po kvalifikaciji ima tako močan vpliv na sumarno poprečje, da so poprečni prejemki delavstva v celoti v kmetijstvu nižji kot v gradbeništvu, čeprav so za posamezne kvalifikacije prejemki obratni.

Standardizirani pokazovalci

9.29 Omenjena lastnost sumarnih pokazovalcev je tako značilna, da se vprašamo, ali imajo sumarni pokazovalci sploh analitičen pomen in smisel, ker je primerljivost sumarnih pokazovalcev zelo dvomljiva in meglena, če ne vemo, ali izvira razlika iz razlik v grupnih pokazovalcih ali iz razlik v sestavi ponderov.

Splošen koeficient mortalitete je po tem odvisen od umrljivosti po posameznih starostnih skupinah za ženske in moške in spolno-starostne sestave prebivalstva. Prebivalstvo z razmeroma nizko stopnjo umrljivosti po posameznih starostnih skupinah more imeti visok splošen koeficient umrljivosti, če so v populaciji predvsem stari ljudje. Prebivalstvo z razmeroma visoko stopnjo umrljivosti po posameznih starostnih skupinah pa more imeti nizek splošen koeficient umrljivosti, če sestoji predvsem iz mladih ljudi.

Podjetje izkaže visok sumarni koeficient za produktivnost dela, če izdeluje predvsem izdelke, za katere je splošna produktivnost dela visoka, čeprav je produktivnost dela za posamezen izdelek v podjetju razmeroma nizka in obratno.

Primerljivost tako izračunanih sumarnih pokazovalcev motijo razlike v sestavi. Da izločimo vpliv razlik v sestavi, izračunavamo standardizirane pokazovalce tako, da vzamemo za vse sumarne pokazovalce, ki jih primerjamo, stalno-standardno sestavo za ponderje. Tako je standardiziran koeficient umrljivosti tehtana sredina specifičnih koeficientov umrljivosti po starosti, pri tem pa vzamemo enotno standardno starostno sestavo za vse države, za katere primerjamo podatke. Ti koeficienti imajo večjo analitično vrednost kot navadni koeficienti umrljivosti, ker razlike med standardiziranimi koeficienti kažejo samo razlike v umrljivosti, ne pa v starostni in spolni strukturi. Standardna struktura, ki jo vzamemo za osnovo za izračunanje sumarnih koeficientov, mora biti seveda taka, da čim bolje ustreza stvarnim sestavam za posamezne populacije, ki jih med seboj primerjamo. Standardna sestava je zato včasih poprečna sestava iz vseh populacij, ki jih primerjamo, včasih pa idealna sestava, ki je pogojena z analizo pojava in podobno.

9.30 Izračunajmo za prejemke delavcev po kvalifikaciji za kmetijstvo in gradbeništvo iz tabele 9.5 standardizirane poprečne prejemke delavcev za obe panogi. Za standardno sestavo vzemimo povprečno sestavo delavstva v SFRJ po kvalifikaciji v vseh panogah.

Tabela 9.6 Izračunanje standardiziranih mesečnih prejemkov v kmetijstvu in gradbeništvu v letu 1957 v SFRJ

Kvalifikacija	Poprečni mesečni prejemki v din		Standardna sestava		
	kmetijstvo	gradbeništvo		(1).(3)	(2).(3)
Visoko kvalificirani	15310	15200	0,075	1148.25	1140.00
Kvalificirani	12430	11920	0,365	4536.95	4350,80
Priučeni	9530	8950	0,263	2506.39	2353.85
Nekvalificirani	8030	6770	0,297	2384,91	2010.69
			1.000	10576.50	9855.34

Za izračunanje sumarnih poprečnih prejemkov vzamemo za standardno sestavo sestavo delavstva po kvalifikaciji v SFRJ. Tako dobimo, da so v kmetijstvu poprečni mesečni prejemki večji (10576 din) kakor v gradbeništvu (9855 din). Standardizirana poprečja resnično pokažejo, da je raven dohodkov v kmetijstvu višja kot v gradbeništvu, ker razlike v sumarnih dohodkih izvirajo samo iz razlik v ravni dohodkov po kvalifikacijah, ne pa tudi iz razlik v sestavi zaposlenih. Seveda so standardizirana poprečja različna od poprečij, ki smo jih dobili v tabeli 9.5. Te razlike so rezultat tega, da je stvarna sestava v posamezni panogi različna od poprečne sestave za vse panoge.

GEOMETRIJSKA SREDINA

9.31 Geometrijska sredina G iz N vrednosti $y_1, y_2 \dots y_N$ je N-ti koren iz produkta vseh N vrednosti

$$G = \sqrt[N]{y_1 \cdot y_2 \dots y_N} \qquad (9.30)$$

Iz te opredelitve sklepamo, da ima smisel računati geometrijsko sredino le tedaj, če noben izmed členov ni negativen ali nič.

9.32 Tehtana geometrijska sredina je za razliko od navadne geometrijske sredine, ki je dana v obrazcu 9.30, enaka

$$G = \sqrt[\sum w_k]{y_1^{w_1} \cdot y_2^{w_2} \ldots y_r^{w_r}} \qquad (9.31)$$

9.33 Neposredno računanje geometrijske sredine po zgornjih osnovnih obrazcih je, razen za najenostavnejše primere, neizvedljivo. Pač pa moremo geometrijsko sredino razmeroma enostavno izračunati z logaritmi. Če namreč obrazca 9.30 in 9.31 logaritmiramo, dobimo

$$\log G = \frac{1}{N} \sum \log y_i \qquad (9.32)$$

in

$$\log G = \frac{1}{\sum w_k} \sum w_k \log y_k \qquad (9.33)$$

Logaritem navadne geometrijske sredine je enak navadni aritmetični sredini logaritmov iz osnovnih podatkov, logaritem za tehtano geometrijsko sredino pa je enak tehtani aritmetični sredini iz logaritmov osnovnih podatkov. S tema obrazcema posredno računamo geometrijske sredine z logaritmiranjem.

9.34 Zaradi razmeroma zamotanega računanja in težjega tolmačenja geometrijsko sredino računamo redkeje kot druge vrste sredin. Vendar so problemi, pri katerih je upravičeno edino računanje geometrijske sredine. Tako računamo geometrijsko sredino v nekaterih posebnih problemih z indeksi oziroma relativnimi števili na splošno.

Povečanju cene za 100% (indeks 200) smiselno ne ustreza znižanje cene za 100%, ker bi v tem primeru bila cena nič, temveč znižanje za 50% (indeks 50). Izravnava raznosmernih učinkov se pokaže pri geometrijski sredini, ker je geometrijska sredina iz indeksa 200 in 50

$$\sqrt{200.50} = 100$$

Ne pokaže pa se na primer pri aritmetični sredini, ker je aritmetična sredina med 200 in 50 enaka 125.
Zato je za izračunanje časovnih sredin iz posamičnih indeksov bolj upravičena geometrijska kot aritmetična sredina.

9.35 Geometrijsko sredino uporabljamo tudi za računanje srednjega koeficienta dinamike ali verižnega indeksa poprečne stopnje rasti.

Zaznamujmo z $Y_o, Y_1, Y_2, \ldots Y_N$ posamezne člene v časovni vrsti, s $k_1 = Y_1/Y_o$, $k_2 = Y_2/Y_1, \ldots\ldots k_N = Y_N/Y_{N-1}$ pa posamezne koeficiente dinamike. k naj bo srednji koeficient dinamike, s katerim bi se moral pojav v času od o do N spreminjati, da bi dosegli isti končni člen Y_N.

Med temi količinami so naslednje zveze

$$Y_o . k_1 . k_2 \ldots\ldots k_N = Y_N = Y_o k . k \ldots\ldots k = Y_o k^N \qquad (9.34)$$

Če namreč Y_o postopoma množimo s $k_1, k_2 \ldots k_N$, dobimo zaradi pomena koeficientov dinamike postopoma vrednosti členov v časovni vrsti in končno zadnji člen Y_N. Glede na opredelitev srednjega koeficienta dinamike k pa dobimo Y_N tudi, če začetno vrednost Y_o N-krat pomnožimo s k, ali če jo pomnožimo s k^N.

Iz zveze v obrazcu 9.34 moremo izpeljati dva obrazca

$$k = \sqrt[N]{k_1 . k_2 \ldots\ldots k_N} \qquad (9.35)$$

in

$$k = \sqrt[N]{Y_N / Y_o} \qquad (9.36)$$

ali splošneje

$$k = \sqrt[t_1 - t_o]{Y_{t_1} / Y_{t_o}} \qquad (9.37)$$

Srednji koeficient dinamike je po obrazcu 9.35 geometrijska sredina iz posameznih koeficientov dinamike.

Po drugem obrazcu pa računamo srednji koeficient dinamike, če poznamo za razmik, za katerega računamo srednji koeficient dinamike, začetno Y_{t_o} in končno vrednost pojava Y_{t_1}. Srednji koeficient dinamike je v tem primeru koren iz kvocienta iz Y_{t_1} in Y_{t_o}. Stopnja korena pa je določena s časovnim razmikom, za katerega iščemo srednji koeficient dinamike. Zato ni nujno, da je N celo število, ampak more biti $(t_1 - t_o)$ kateri koli časovni razmik, kar je določneje vidno iz obrazca 9.37.

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

Srednji verižni indeks v tem razdobju je po obrazcu 9.35 enak geometrijski sredini iz posameznih verižnih indeksov

$$\bar{I} = \sqrt[5]{105,9.110,4.115,4.126,1.117,1}$$

Z logaritmi izračunamo srednji verižni indeks po obrazcu 9.32

$$\log \bar{I} = \frac{1}{5}(\log 105,9 + \log 110,4 + \log 115,4 + \log 126,1 + \log 117,1) = \frac{1}{5}(2,02490 +$$

$$+2,04297 + 2,06221 + 2,101106 + 2,06856) = 2,0599.$$

Z antilogaritmiranjem dobimo, da je srednji verižni indeks $\bar{I} = 114,8$, poprečna stopnja rasti $T = \bar{I} - 100 = 14,8\%$

Da bi bilo zvečanje proizvodnje električne energije v razdobju 1951-1956 enako stvarnemu, bi se morala proizvodnja električne energije večati s srednjim verižnim indeksom 114,8.

Isti rezultat dobimo, če računamo srednji verižni indeks po obrazcu 9.36 iz začetne in končne proizvodnje električne energije.

9.37 Po republikah imamo podatke o družbenem produktu, po cenah iz leta 1966 za leti 1963 in 1971 po republikah. Ker je od 1963 do 1971 razdobje osmih let, je poprečen koeficient dinamike osmi koren iz kvocientov Y_{71} in Y_{63}

Tabela 9.8 Poprečni koeficienti dinamike za družbeni produkt v razdobju 1963-1971, po republikah in pokrajinah (Vir:SG-72)

Republika-pokrajina	1963	1971	poprečen koeficient dinamike	poprečen verižni indeks	poprečna stopnja rasti %
SFRJ	80577	134234	1,065	106,5	+ 6.5
BiH	9890	15749	1,060	106,0	+ 6.0
Črna gora	1433	2463	1,070	107,0	+ 7.0
Hrvatska	21478	35680	1,066	106,6	+ 6.6
Makedonija	4059	7403	1,078	107,8	+ 7.8
Slovenija	12220	20814	1,069	106,9	+ 6.9
Srbija - skupno	31497	52125	1,065	106,5	+ 6.5
Srbija - ožje področje	20565	33942	1,065	106,5	+ 6.5
Vojvodina	9471	15429	1,063	106,3	+ 6.3
Kosovo	1461	2754	1,082	108,2	+ 8.2

Za posamezne republike in pokrajine so osemletni poprečni koeficienti dinamike v razmiku od 1,060 za BiH do 1,082 za Kosovo.

ODNOSI MED RAZLIČNIMI VRSTAMI SREDNJIH VREDNOSTI

9.38 Če vzamemo ilustrativen zgled in iz 1 in 4 izračunamo harmonično H, geometrijsko G in aritmetično M sredino, dobimo, da je

$$H = \frac{2}{1/1 + 1/4} = 1,6 ; \quad G = \sqrt{1.4} = 2 ; \quad M = \frac{1+4}{2} = 2,5$$

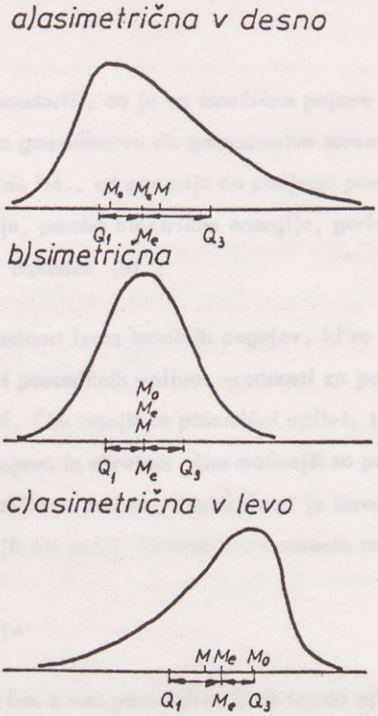
Iz teh rezultatov vidimo, da je harmonična sredina, če jo izračunamo iz istih podatkov, najmanjša, aritmetična sredina največja, geometrijska sredina pa leži med obema. To ne velja samo za ta primer, ampak velja splošno pravilo; harmonična sredina je manjša kakor geometrijska, ta pa manjša kakor aritmetična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

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

9.39 Tudi med vrednostmi modusa Mo, mediane Me in aritmetične sredine M opazimo neke stalne odnose. Za unimodalne simetrične porazdelitve so vrednosti modusa, mediane in aritmetične sredine enake ($M_o = M_e = M$). Za unimodalne porazdelitve, ki so asimetrične v desno, je modus manjši, aritmetična sredina pa večja kot mediana ($Mo < Me < M$). Za unimodalne porazdelitve, ki so asimetrične levo, pa se vrstni red zamenja in je aritmetična sredina manjša, modus pa večji kot mediana ($M < Me < Mo$).

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

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



Slika 9.2. Odnosi med Mo, M in Me za različne tipe unimodalnih porazdelitev

Ta obrazec moremo uporabiti za ne preveč asimetrične unimodalne porazdelitve, da z njim ocenimo modus, čeprav nimamo podatke grupirane v frekvenčni porazdelitvi. Ker moremo M in Me izračunati tudi iz ne grupiranih podatkov, ocenimo vrednost modusa z obrazcem

$$Mo = M - 3. (M - Me) \tag{9.36}$$

Odnosi med modusom, mediano in aritmetično sredino za posamezne vrste porazdelitev so nakazani v sliki 9.2.

DESETO POGLAVJE

MERE VARIACIJE, ASIMETRIJE IN SPLOŠČENOSTI

MERE VARIACIJE

10.1 Že večkrat smo poudarili, da je za množične pojave značilna variabilnost. Tako se od kmetijskega gospodarstva do gospodarstva spreminja velikost, pridelek, število članov, število živine itd., od podjetja do podjetja poraba surovin, število delavcev, vrednost proizvodnje, poraba električne energije, goriva itd., od človeka do človeka starost, spol, stan, dohodek itd.

Medtem ko je srednja vrednost izraz splošnih pogojev, ki so za vso populacijo enaki - nespremenjeni, se zaradi posamičnih vplivov vrednosti za posamezne enote od nje odklanjajo navzgor in navzdol. Čim manjši so posamični vplivi, tem manjši so odkloni in tem manjša je variabilnost pojava in obratno; čim močnejši so posamični vplivi, tem večji so odkloni oziroma variabilnost pojava. Variabilnost je izraz delovanja posamičnih vplivov, ki morejo biti manjši ali večji. Variabilnost moremo torej meriti.

Vrste mer variacije

10.2 Variabilnost merimo z več parametri. Vsak izmed njih ima drugo osnovo, vendar vsi - vsak po svoje, kaže isto lastnost - variabilnost pojava.

Obravnavali bomo naslednje mere variacij e:

a) variacijski razmik R,

b) kvartilni odklon Q,

c) poprečni absolutni odklon AD,

d) standardni odklon SD– σ, oziroma varianco σ^2, iz katere je SD izpeljan,

e) poprečna razlika Δ_1.

Od teh mer variacije sta variacijski razmik R in kvartilni odklon Q dana z lego nekaterih členov v populaciji, poprečni absolutni odklon AD, standardni odklon SD in poprečna razlika Δ_1 pa so izračunani iz vseh vrednosti populacije in so zaradi tega solidnejše mere variabilnosti. Zaradi svojih lastnosti je izmed vseh mer variacije najvažnejši standardni odklon SD.

Variacijski razmik– R

10.3 Najenostavnejša, a tudi najbolj problematična mera variacije je variacijski razmik

$$R = y_{max} - y_{min} \tag{10.1}$$

ki je po definiciji razlika med največjo in najmanjšo vrednostjo v populaciji.

Najmanjšo in največjo vrednost zlahka najdemo, če so podatki urejeni v ranžirni vrsti. Če vzamemo, da je bil v letu 1969 v SRS najnižji blagovni promet na prebivalca v občini Šentjur (1.497) in najvišji v občini Ljubljana-Center (30.011 din), je variacijski razmik za blagovni promet na prebivalca

$$R = 30.011 - 1.497 = 28.514 \text{ din}$$

Včasih variacijski razmik posredno izražamo tudi tako, da navedemo najnižjo in najvišjo vrednost.

Variacijski razmik je nestabilna mera variacije, ker je izračunan iz obeh skrajnih vrednosti populacije, ki moreta biti rezultat nekih izjemnih – netipičnih vplivov. V našem primeru je to gotovo s podatkom za Ljubljano-Center. R ni odvisen od razmestitve vrednosti znotraj variacijskega razmika. Ta more biti najrazličnejša. Vse vrednosti morejo biti v variacijskem razmiku razmeščene enakomerno ali pa se goste okrog nekega središča. V prvem primeru je variabilnost znatno večja kot v drugem, variacijski razmik pa tega ne pokaže, ker sta skrajni vrednosti v obeh primerih isti. Kljub tem hibam pa variacijski razmik uporabljamo vselej, kadar hočemo jakost variacije na hitro orientacijsko oceniti in ni časa, niti potrebe za izračun drugih, natančnejših mer variacije. Eden izmed takih primerov je tekoče preverjanje kakovosti dela strojev. Variacijski razmik, ki ga določimo iz meritev razmeroma majhnega števila izdelkov, rabi za orientacijo jakosti variacije, ki je eden izmed osnovnih pokazovalcev kakovosti proizvodnje.

Kvartilni odklon - Q

10.4 Variacijski razmik je slabo merilo variabilnosti zato, ker je odvisen samo od skrajnih vrednosti. To hibo odpravimo tako, da merimo variacijo z razmikom, v katerem je samo del populacije. Tako s prvim in devetim decilom omejimo razmik, v katerem je 80% vseh vrednosti, ker smo izločili po 10% najmanjših in največjih vrednosti. Ta razmik

$$D_9 - D_1$$

imenujemo decilni razmik.

Podobno s prvim in tretjim kvartilom omejimo razmik, v katerem je 50% vrednosti in iz njega izključimo 25% najmanjših in 25% največjih vrednosti.

Običajno pa namesto kvartilnega razmika

$$Q_3 - Q_1$$

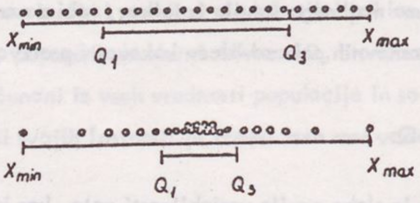
vzamemo za mero variacije kvartilni odklon

$$Q = \frac{1}{2}\ (Q_3 - Q_1) \qquad (10.2)$$

ki je polovica kvartilnega razmika.

Gornji meri variacije pravilneje pokažeta jakost variacije kakor variacijski razmik, ker nanju ne vplivajo skrajne - netipične vrednosti. Vendar pogosteje uporabljamo kvartilni kot decilni razmik. V sliki 10.1 je shematično prikazano, kako je variacijski razmik neobčutljiv za razlike v razmestitvi vrednosti znotraj variacijskega razmika, in kako se različna razmestitev pokaže na kvartilnem razmiku.

Določanje gornjih mer variacije ne dela težav, ker smo v odstavku o kvantilih nakazali, kako izračunamo decile in kvartile.



Slika 10.1 Variacijski in kvartilni razmik pri različni razmestitvi vrednosti

10.5 Za blagovni promet na prebivalca med občinami v SRS v letu 1969, za katerega smo izračunali že variacijski razmik, je kvartilni odklon

$$Q = \frac{1}{2}(Q_3 - Q_1) = \frac{5102,5 - 2632,5}{2} = 1235 \text{ din}$$

Poprečni absolutni odklon - AD

10.6 Uvodoma smo naznačili, da z odkloni posamičnih vrednosti od sredine merimo jakost posamičnih vplivov na posamezno enoto. Sumarno merilo jakosti teh vplivov bi moglo biti poprečje posamičnih odklonov. To pa je, kakor vemo, enako nič, če odklone računamo od aritmetične sredine, ker so odkloni pozitivni in negativni. Zato vzamemo za merilo variabilnosti poprečje absolutnih vrednosti posamičnih odklonov od sredine. Po tej definiciji je poprečni absolutni odklon od aritmetične sredine

$$AD_M = \frac{1}{N}\sum|y - M| \qquad (10.3)$$

Ker je poprečni absolutni odklon najmanjši, če računamo odklone od mediane, računamo AD tudi iz absolutnih odklonov od mediane

$$AD_{Me} = \frac{1}{N}\sum|y - Me| \qquad (10.4)$$

AD_{Me} je vsebinsko zaradi te lastnosti mediane bolj upravičen kot AD_M.

10.7 Če z $Y_s = \sum_{y < M_e} y_i$ zaznamujemo vsoto posamičnih podatkov, katerih vrednost je manjša kot mediana, z $Y_z = \sum_{y > M_e} y_i$ pa vsoto podatkov, ki so večji kakor mediana, izračunamo AD_{Me} po obrazcu

$$AD_{Me} = \frac{1}{N}\,(Y_z - Y_s) \tag{10.5}$$

AD_{Me} pa moremo računati tudi po obrazcu

$$AD_{Me} = \frac{1}{2}\,(M_z - M_s) \tag{10.6}$$

pri čemer je M_s poprečje iz členov pod, M_z pa poprečje iz členov nad mediano. Ta zveza je po sestavu podobna kvartilnemu odklonu, le da namesto kvartilov, ki sta mediani za prvo in drugo polovico populacije, vzamemo ustrezne aritmetične sredine.

Za blagovni promet na prebivalca v letu 1969 je:

$Y_s = 79347$, $Y_z = 183078$, število občin $N = 60$

Po obrazcu 10.5 je

$$AD_{Me} = \frac{1}{60}\,(183978 - 79347) = 1743{,}85 \text{ din}$$

AD_M pa dobimo po obrazcu

$$AD_M = 2 \cdot P_z \cdot P_s\,(M_z - M_s) \tag{10.7}$$

pri tem pomeni: $P_s = N_s/N$ in $P_z = N_z/N$ strukturna deleža števila enot pod in nad aritmetično sredino M, $M_s = Y_s/N_s$ in $M_z = Y_z/N_z$ pa poprečji členov, ki so pod oziroma nad skupnim poprečjem M.

Za blagovni promet na prebivalca v SRS po občinah je $N_s = 42$, $N_z = 18$, $Y_s = 126249$, $Y_z = 183978$. Iz teh podatkov dobimo, da je: $P_s = 42/60 = 0{,}70$; $P_z = 18/60 = 0{,}30$;

$M_s = 126249/42 = 3005,93$ in $M_z = 183978/18 = 7615,33$.

Po obrazcu 10,7 pa je

$$AD_M = 2.0,30 \cdot 0,70 \cdot (7615,33 - 3005,93) = 1935,95 \text{ din}$$

Varianca. Standardni odklon $- \sigma^2, \sigma$ - SD

10.8 Podobno osnovo kakor poprečen absolutni odklon ima tudi varianca σ^2, ki je po definiciji poprečje kvadratov odklonov od aritmetične sredine

$$\sigma^2 = \frac{1}{N}\sum(y_i - M)^2 \qquad (10.8)$$

Pri AD odpravimo predznake odklonov tako, da vzamemo absolutne odklone, pri σ^2 pa tako, da odklone kvadriramo.

Podobno kot AD je tudi σ^2 izračunana mera variacije in je odvisna od vsake izmed vrednosti populacije. Vendar so druge lastnosti variance osnovne važnosti za statistično analizo. Zaradi tega je, enako kakor je aritmetična sredina osnovna srednja vrednost, varianca osnovna mera variacije, čeprav je njeno izračunanje razmeroma zamudno.

10.9 Varianca je poprečje kvadratov odklonov od aritmetične sredine. Zato je izražena v neprikladnih enotah mere - kvadratu enote mere znaka. Temu se izognemo tako, da izračunamo kvadratni koren iz variance. Dobljena mera variacije, ki jo imenujemo standardni odklon, zaznamujemo pa z σ ali SD pa je merjena v isti enoti mere kot osnovni znak. Varianca in standardni odklon sta v naslednji enostavni zvezi

$$SD = \sigma = \sqrt{\sigma^2} \qquad (10.9)$$

Računanje iz negrupiranih podatkov

10.10 Računanje variance po osnovnem obrazcu 10.8 se izkaže za zelo neprikladno. Po njem bi morali najprej izračunati aritmetično sredino za proučevan znak M, izračunati odklone posameznih vredbosti od aritmetične sredine $(y_i - M)$, vsak posamezen odklon kvadrirati $(y_i - M)^2$, dobljene kvadrate sešteti in deliti z obsegom populacije. Zato varianco iz posamičnih podatkov raje računamo po enem izmed obrazcev

$$\sigma_y^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{N}}{N} = \frac{N\sum y_2^2 - (\sum y_i)^2}{N^2} \qquad (10.10)$$

$$\sigma_y^2 = \frac{Ky}{N} \; ; \; Ky = Q_i - Q \; ; \; Q_i = \sum y_i^2 \quad Q = \frac{(\sum y)^2}{N} \qquad (10.11)$$

Prvi obrazec dobimo iz osnovnega obrazca 10.8 z enostavnim računom

$$\sigma_y^2 = \frac{1}{N}\sum(y_i - M)^2 = \frac{1}{N}\left[\sum y^2 - 2M\sum y + NM^2\right] = \frac{1}{N}\left[\sum y_i^2 - \frac{(\sum y)^2}{N}\right] \qquad (10.12)$$

Nakazani način je še posebno prikladen, če računamo varianco z računskim strojem, ki istočasno zbira vsote podatkov in vsote kvadratov.

10.11 Če iz podatkov za 20 največjih mest v SFRJ izračunamo varianco poprečnih cen na drobno (v starih din) v letu 1970 za kokošja jajca, dobimo iz osnovnih podatkov

71	71	68	71	74	73	63	78	73	66	60	78	74
71	66	62	56	59	58	66						

$$\sum y_i = Y = 71 + 71 + 68 + \ldots + 58 + 66 = 1358$$

$$\sum y_i^2 = 71^2 + 71^2 + 68^2 + \ldots + 58^2 + 66^2 = 93028$$

$$\sigma_y^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{N}}{N} = \frac{93028 - \frac{1358^2}{20}}{20} = 40,99$$

standardni odklon pa

$$\sigma_y = \sqrt{40,99} = 6,4 \text{ din}$$

10.12 Ker ostane varianca enaka, če vsem osnovnim podatkom prištejemo ali odštejemo isto konstanto, je včasih prikladno, da od osnovnih podatkov pred računanjem odštejemo osnovnim podatkom ustrezno konstanto y_o, tako da so reducirane vrednosti u čim manjše

$$u_i = y_i - y_o$$

Varianco osnovnega znaka σ_y^2 dobimo, če računamo varianco iz reduciranih podatkov u po obrazcu

$$\sigma_y^2 = \sigma_u^2 = K_u/N \quad ; \quad K_u = \sum u^2 - \frac{(\sum u)^2}{N} \qquad (10.13)$$

10.13 Če v našem zgledu za ceno jajc odštejemo od osnovnih podatkov $y_o = 68$, dobimo reducirane vrednosti:

3	3	0	3	6	5	-5	10	5	-2	-8	10	6
3	-2	-6	-12	-9	-10	-2						

$$U = \sum u_i = 3 + 3 + 0 + \dots + (-9) + (-10) + (-2) = -2$$

$$\sum u_i^2 = 3^2 + 3^2 + 0^2 + \dots + (-9)^2 + (-10)^2 + (-2)^2 = 820$$

$$\sigma^2 = \frac{\sum u^2 - \frac{U^2}{N}}{N} = \frac{820 - (-2)^2/20}{20} = 40{,}99$$

Dobili smo isti rezultat kot po obrazcu 10.10

Računanje iz grupiranih podatkov

10.14 **Neposredna metoda.** Za velike populacije je računanje variance iz negrupiranih podatkov še zamudneje kakor računanje aritmetične sredine. Zaradi tega za velike populacije tudi variance ocenjujemo iz podatkov, ki so grupirani v frekvenčne porazdelitve. Ocena variance iz frekvenčne porazdelitve je

$$\sigma_y^2 = \frac{1}{N}\sum f_k (y_k - M_y)^2 \qquad (10.14)$$

tehtana aritmetična sredina kvadratov odklonov sredin razredov y_k od aritmetične sredine M_y. Vendar se ta postopek izkaže neprikladen v primerjavi z drugima postopkoma, s katerima moremo tudi izračunati varianco iz frekvenčnih porazdelitev. Zato zanj ne navajamo primera, ker v splošnem te metode ne uporabljamo.

10.15 Metoda pomožnega znaka u. Metoda pomožnega znaka u je samo razširitev metode pomožnega znaka u za računanje aritmetične sredine iz grupiranih podatkov.

Po tej metodi računamo varianco tako, da:

a) Enako kakor za aritmetično sredino izberemo razred, ki je nekje sredi frekvenčne porazdelitve. Vanj postavimo izhodišče znaka u = o, v druge razrede pa navzdol in navzgor od izhodišča ustrezne vrednosti pomožnega znaka u:
... -3, -2, -1, 0, +1, +2, +3, ...

b) Enako kakor za aritmetično sredino, izračunamo produkte frekvenc f_k z ustreznimi vrednostmi znaka u_k, da dobimo $f_k u_k$

c) Produkte $f_h u_k$ ponovno pomnožimo z ustreznimi vrednostmi znaka u_k: tako dobimo vrednosti $f_k u_k^2$.

d) Seštejemo frekvence f_k, produkte $f_k u_k$ in produkte $f_k u_k^2$.

e) Iz teh količin izračunamo varianco po obrazcih

$$\sigma_y^2 = \frac{i^2}{N} K_u \quad : \quad K_u = \sum f_k u_k^2 - \frac{(\sum f_k u_k)^2}{N} \qquad (10.15)$$

Pri tem je: i = širina razreda.

10.16 Za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto je prikazano računanje variance po tej metodi v tabeli 10.1.

Tabela 10.1 Izračun variance σ^2 in standardnega odklona σ za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi pomožnega znaka u

Poraba v m^3	f	u	fu	fu^2
0,0 - 3,9	2	-3	- 6	18
4,0 - 7,9	16	-2	- 32	64
8,0 - 11,9	36	-1	- 36	36
12,0 - 15,9	41	0	0	0
16,0 - 19,9	29	+1	+ 29	29
20,0 - 23,9	12	+ 2	+ 24	48
24,0 - 27,9	7	+ 3	+ 21	63
28,0 - 31,9	4	+ 4	+ 16	64
32,0 - 35,9	2	+ 5	+ 10	50
	149		+ 26	372
	N		U	$\sum fu^2$

Po obrazcih 10.11 dobimo dalje

$$K_u = 372 - (+26)^2/149 = 367,4631$$

$$\sigma_y^2 = 4^2 \cdot 367,4631/149 = 39,4591$$

$$\sigma_y = \sqrt{39,4591} = 6,28\ m^3$$

10.17 Metoda kumulativ. Tudi za računanje variance moremo koristno uporabiti lastnosti kumulativnih vrst. Ta metoda je enako kakor metoda pomožnih znakov u samo razširitev iste metode pri računanju aritmetične sredine.

Po metodi kumulativ računamo varianco tako, da:

a) Enako kakor pri računanju aritmetične sredine iz frekvenčne porazdelitve iz f izračunamo kumulativno vrsto F. Zadnji člen kumulativne vrste (pod črto) je obseg populacije N.

b) Po enakem postopku kumuliranja izračunamo iz prve kumulativne vrste F drugo kumulativno vrsto FF. Zadnji člen druge kumulativne vrste FF (pod črto) je C_1.

c) Seštejemo vrednosti členov druge kumulativne vrste (brez člena pod črto). To vsoto zaznamujmo s C_2.

d) Iz teh količin izračunamo varianco po obrazcih

$$K_u = 2\,C_2 + C_1 - C_1^2/N \; ; \quad \sigma_y^2 = \frac{i^2}{N}\, K_u \qquad (10.16)$$

e) Kumulativne vrste moremo računati od zgoraj navzdol ali obratno od spodaj navzgor. Za asimetrične porazdelitve je prikladneje začeti izračunavati kumulative na strani asimetrije.

10.18 V tabeli 10.2 je postopek nakazan za porabo lesa v Novem mestu.

Tabela 10.2. Računanje variance za porabo lesa v 149 kmetijskih gospodarstvih v okraju Novo mesto po metodi kumulativ.

Poraba v m^3	f	F	FF
0,0 - 3,9	2	0	
4,0 - 7,9	16	2	0
8,0 - 11,9	36	18	2
12,0 - 15,9	41	54	20
16,9 - 19,9	29	95	74
20,0 - 23,9	12	124	169
24,0 - 27,9	7	136	293
28,0 - 31,9	4	143	429
32,0 - 35,9	2	147	572
		149	719
		N	C_1

$1559 = 2 + 20 + \ldots + 429 + 572 = C_2$

$K_u = 2.1559 + 719 - 719^2/149 = 367{,}4631$, torej enako kot po metodi pomožnega znaka.

10.19 Sheppardov popravek. Varianca, ki jo izračunamo po zgornjih postopkih, je samo ocena za pravo vrednost variance, ki bi jo dobili, če bi jo izračunali iz posamičnih podatkov. Analiza te ocene pa pokaže, da dobimo za unimodalne, ne preveč asimetrične porazdelitve za zvezne znake po tej metodi sistematično prevelike vrednosti za varianco. Ta napaka je tem večja, čim širši so razredi. Varianco, izračunano po prejšnjih metodah, moremo popraviti po obrazcu

$$\sigma^2_{y,pop} = \sigma^2_y - i^2/12 \tag{10.17}$$

tako, da od prvotne ocene variance, ki jo dobimo po metodi znaka u ali po metodi kumulativ, odštejemo dvanajstino kvadrata širine razreda.

Ta popravek imenujemo Sheppardov popravek.

Za frekvenčno porazdelitev o porabi lesa v Novem mestu v letu 1953 je popravljena varianca enaka

$$\sigma^2_{y,pop} = 39{,}4591 - 4^2/12 = 38{,}1258$$

$$\sigma_{y,pop} = \sqrt{38{,}1258} = 6{,}17\ m^3$$

Skupna varianca

10.20 Enako kakor sumarno aritmetično sredino moremo tudi skupno varianco izračunati iz grupnih podatkov. Če poznamo po grupah število enot N_k, aritmetične sredine M_k in grupne variance σ^2_k, izračunamo skupno varianco za celo populacijo po obrazcu

$$\sigma^2 = M_{\sigma^2} + \sigma^2_M \tag{10.18}$$

Pri tem je

$$M_{\sigma^2} = \frac{1}{N}\sum N_k \sigma^2_k \tag{10.19}$$

tehtana aritmetična sredina grupnih varianc,

$$\sigma^2_M = \frac{1}{N}\sum N_k (M_k - M)^2 \tag{10.20}$$

pa tehtana varianca med grupnimi aritmetičnimi sredinami M_k.

To zelo pomembno zvezo moremo dokazati enostavno. Iz znanih zvez velja $N_k \sigma_k^2 = \sum_i (y_{ki} - A)^2 - N_k (M_k - A)^2$. Ta zveza velja za poljuben A. Če postavimo A = M in seštejemo po grupah, dobimo $\sum_k N_k \sigma_k^2 = \sum_k \sum_i (y_{ki} - M)^2 - \sum_k N_k (M_k - M)^2$. Če to zvezo delimo z N, dobimo $M_{\sigma^2} = \sigma^2 - \sigma_M^2$.

10.21 Vzemimo za zgled stroške za kulturno in družbeno življenje v novembru 1957 za 36 delavskih družin v Mariboru. V tabeli 10.3 so navedeni N_k, M_k in σ_k^2 po višini dohodkov.

Tabela 10.3 Podatki za N_k, M_k in σ_k^2 za stroške za kulturno in družbeno življenje za 36 delavskih družin v novembru 1957 v Mariboru po skupinah dohodkov.

Višina dohodkov v tisoč din	N_k	M_k	σ_k^2
15 - 19	12	479	28358
20 - 24	10	670	35820
25 - 29	7	949	48212
30 - 34	4	1317	83469
35 -	3	1703	8089
	36=N	819=M	

Iz teh podatkov dobimo

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

$$M_{\sigma^2} = \frac{1}{36} (12 \cdot 28358 + \ldots + 3.8089) = 38726$$

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

10.22 Pomembnejše kakor to, da moremo izračunati skupno varianco, če poznamo grupne podatke, pa je dejstvo, da moremo z zgornjim postopkom skupno varianco razstaviti po grupnem znaku v dva dela. Skupna varianca namreč meri vpliv vseh posamičnih faktorjev, medtem ko meri σ_M^2 vpliv grupnega znaka, M_{σ^2} pa vpliv drugih, posamičnih faktorjev na variabilnost pojava.

Po zgornjem postopku moremo torej varianco analizirati v tem smislu, da ugotovimo, koliko od skupne variance je rezultat dejavnika, ki ga vzamemo za grupni znak.

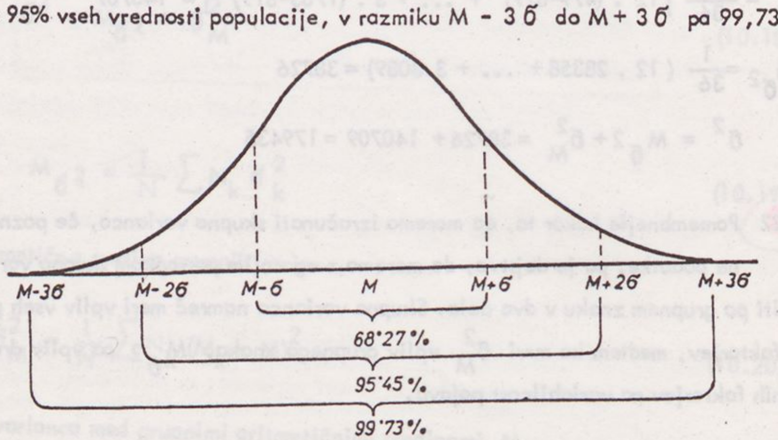
Za stroške za kulturno in družbeno življenje za 36 družin v novembru 1967 v Mariboru vidimo, da je velik del (78%) od skupne variance stroškov za kulturno in družbeno življenje pojasnjen z višino dohodkov. Višina dohodkov je torej eden izmed bistvenih faktorjev, ki vplivajo na stroške za kulturno in družbeno življenje.

10.23 Zveza standardnega odklona z normalno porazdelitvijo

Variacijski razmik ali kvartilni razmik imata enostavno tolmačenje in pomen. Enako je tudi poprečni absolutni odklon razmeroma razumljiv. Teže pa je s standardnim odklonom, dokler ga ne obravnavamo v zvezi z nekaterimi lastnostmi porazdelitev. Standardni odklon namreč dobimo z razmeroma zapletenim računskim postopkom, ki zamegli predstavo o smislu tega parametra.

Čeprav normalno porazdelitev podrobneje obravnavamo kasneje, je prav, da navedemo nekaj njenih lastnosti v zvezi s standardnim odklonom. Kakor že vemo, je normalna porazdelitev ena izmed osnovnih teoretičnih porazdelitev. Pomembna je po svojih teoretičnih lastnostih in praktičnem pomenu. V stvarnosti se ob določenih pogojih pogosto pojavljajo porazdelitve, ki so po svojih značilnostih normalni več ali manj podobne, ker so simetrične, unimodalne in zvonaste oblike.

Za vse normalne porazdelitve na splošno velja, da leži v razmiku $M - \sigma$ do $M + \sigma$ 68,27% ali okroglo 2/3 vseh vrednosti, v razmiku $M - 2\sigma$ do $M + 2\sigma$ 95,45% ali približno 95% vseh vrednosti populacije, v razmiku $M - 3\sigma$ do $M + 3\sigma$ pa 99,73%



Slika 10.2. Deleži za normalno porazdelitev v razmikih $M \pm \sigma$, $M \pm 2\sigma$ in $M \pm 3\sigma$.

ali praktično vse vrednosti. Odkloni od aritmetične sredine, ki so absolutno večji od σ, se v splošnem pojavijo v tretjini vseh primerov, odkloni, ki so večji kot 2σ, samo v 5% vseh primerov, odkloni, ki so večji kot 3σ, pa so zelo zelo redki. Ti odnosi veljajo strogo za normalno porazdelitev. Približno pa veljajo tudi za druge unimodalne, simetrične in zvonaste porazdelitve.

Koliko veljajo gornje zakonitosti za unimodalne zvonaste porazdelitve na splošno, si moremo ustvariti sliko iz porazdelitve o porabi lesa v Novem mestu. Zanjo dobimo, da leži v razmiku $M - \sigma$ do $M + \sigma$, to je od 8,53 m^3 do 20,87 m^3 70% vseh primerov, v razmiku $M - 2\sigma$ do $M + 2\sigma$, to je od 2,36 m^3 do 27,04 m^3 94% vseh primerov, v razmiku $M - 3\sigma$ do $M + 3\sigma$, to je od 0 do 33,21 m^3 pa 99,3% vseh primerov. Razlike od deležev pri normalni porazdelitvi so torej neznatne.

Poprečna razlika - Δ_1

10.24 Poprečni absolutni odklon in varianca imata za osnovo odklone od neke srednje vrednosti (Me ali M). Mera variacije, ki ni zasnovana na odklonih od neke srednje vrednosti, je poprečna razlika Δ_1. Poprečna razlika je poprečje iz vseh možnih pozitivnih razlik vseh vrednosti v populaciji. Če je število enot enako N, je vseh možnih pozitivnih razlik $\binom{N}{2}$. Za ranžirno vrsto posamičnih vrednosti je poprečna razlika Δ_1 dana z obrazcem

$$\Delta_1 = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N} \sum_{j<i} (y_i - y_j); \qquad y_1 \leq y_2 < \cdots \leq y_i \leq \cdots < y_N \qquad (10.21)$$

Iz podatkov, ki so urejeni v ranžirni vrsti, izračunamo Δ_1 s kumulativami

ker je

$$\sum_{i=1}^{N} \sum_{j<i} (y_i - y_j) = \sum_{i=1}^{N} \left[(i-1)\, y_i - (N-i) y_i \right] = \sum_{i=1}^{N} \left[2\, i y_i - (N+1) y_i \right] = (N-1)C_o - 2C_1$$

Iz kumulative ranžirne vrste dobimo pomožne vrednosti C_o in C_1. Za kumulativne vrste vemo, da je $C_o = \sum_{i=1}^{N} y_i$ in $C_1 = \sum_{i=1}^{N} (N-i) y_i$ iz česar je $\sum_{i=1}^{N} i\, y_i = N C_o - C_1$, Δ_1 pa

$$\Delta_1 = \frac{2\left[(N-1)\,C_o - 2C_1\right]}{N(N-1)} \tag{10.22}$$

10.25 Za zgled izračunajmo Δ_1 za naslednje izmišljene podatke: 7,9,13,3,15,8,5.

V ranžirno vrsto urejeni podatki so naslednji:

y_i	3	5	7	8	9	13	15	
Fy_i	0	3	8	15	23	32	45	$60 = C_o$

$$0 + 3 + 8 + \ldots + 45 = 126 = C_1$$

Po obrazcu 10.22 je

$$\Delta_1 = \frac{2\left[(7-1)\ 60 - 2.126\right]}{7.(7-1)} = 5,14$$

Neposredno računanje je takole: Iz ranžirne vrste izračunane vse možne pozitivne razlike so

y_i	3	5	7	8	9	13	15
$y_i - y_1$		2	4	5	6	10	12
$y_i - y_2$			2	3	4	8	10
$y_i - y_3$				1	2	6	8
$y_i - y_4$					1	5	7
$y_i - y_5$						4	6
$y_i - y_6$							2

$$\Delta_1 = \frac{\sum di}{\binom{N}{2}} = \frac{108}{21} = 5,14$$

10.26 Če so podatki grupirani v frekvenčni porazdelitvi, dobimo približek poprečne razlike po obrazcu

$$\Delta_1 = \frac{2\left[\sum Y_k\ (F_k + F_{k+1}) - NY\right]}{N(N-1)} \tag{10.23}$$

pri čemer so: Y_k = grupna vsota razredu k

F_k = kumulativna frekvenca v razredu k

$Y = \sum Y_k$

N = število enot

10.27 Za zgled vzemimo industrijska podjetja v SFRJ po številu zaposlenih v letu 1963. Postopek računa za Δ_1 je nakazan v tabeli 10.4.

Tabela 10.4 Izračunanje poprečne razlike Δ_1 za število zaposlenih v industrijskih podjetjih v SFRJ v letu 1963 (Vir: SG-65, str.76)

Število zaposlenih	Število podjetij f_k	F_k	$F_k + F_{k+1}$	Število zaposlenih Y_k	$(F_k + F_{k+1})Y_k$
- 15	51	0	51	507	25857
16 - 29	82	51	184	1831	336904
30 - 60	202	133	468	9236	4322448
61 - 125	464	335	1134	42038	47671092
126 - 250	581	799	2179	104848	228463792
251 - 500	520	1380	3280	184033	603628240
501 - 1000	310	1900	4110	219033	900225630
1001 - 2000	195	2210	4615	268391	1238624465
2001 -	102	2405	4912	386879	1900349648
	N = 2507			Y = 1216796	4923648076

$$\Delta_1 = \frac{2\left[4923648076 - 2507 \cdot 1216796\right]}{2507\,(2507-1)} = 596$$

Razmerje med Q, AD in SD za normalno porazdelitev

10.27 Za normalne porazdelitve so med Q, AD in SD stalni odnosi, ki jih navajamo zaradi tega, ker iz njih lahko sklepamo na približne odnose za druge simetrične, zvonaste in unimodalne porazdelitve.

Za normalno porazdelitev velja:

$$Q = 0.6745\,\sigma \doteq 2/3\,\sigma \qquad (10.24)$$

$$AD = 0.7979\,\sigma \doteq 4/5\,\sigma \qquad (10.25)$$

Relativne mere variacije

10.28 Absolutne mere variacije, ki smo jih spoznali v prejšnjih odstavkih, moremo med seboj le redko primerjati. Nemogoče je primerjati absolutne mere variabilnosti za raznovrstne pojave, čeprav so v vsebinski zvezi. Ne moremo, recimo, primerjati variabilnost porabe alkoholnih pijač na prebivalca z variabilnostjo za porabo cigaret na prebivalca ali z variabilnostjo za porabo kruha. Mere variacije vsakega izmed teh pojavov imajo različno enoto mere in zato med seboj niso primerljive. Absolutnih mer variacij pa ne moremo primerjati tudi pri istovrstnih pojavih. 15 para razlike v ceni jajc je mnogo pomembnejše kakor na primer 8 din v ceni blaga za moške plašče. Če vzamemo, da je poprečna cena jajc npr. '75 din, je '15 din 20% povečanje; če pa vzamemo, da je poprečna cena blaga za moške suknje npr. 400 din, je 8 din razlike le 2 %. Čeprav v obeh primerih opazujemo razlike v dinarjih, teh razlik ne moremo primerjati, ker se nanašajo na artikle z zelo različnimi višinami cen. Zato absolutne mere variacije lahko primerjamo za istovrstne pojave, pa še za te le tedaj, če je raven, ki je izražena s srednjo vrednostjo, med primerjanima pojavoma ista. Te omejitve v uporabi absolutnih mer variacije odpravimo, če izražamo mere variacije relativno v razmerju do neke srednje vrednosti. Kvocient med mero variacije in ustrezno srednjo vrednostjo je neimenovano število, ki ga običajno izražamo z odstotkom. Zato moremo relativne mere variacije primerjati med seboj tudi za raznovrstne pojave, ki so vsebinsko odvisni. Ker pa meri relativna mera variacije absolutno variabilnost v razmerju s srednjo vrednostjo, moremo med seboj primerjati tudi variabilnosti podatkov na različnih ravneh.

10.29 Relativne mere variacije moremo izračunati iz vseh vrst absolutnih mer variacije: R, Q, AD, SD in Δ_1. Vsako izmed teh primerjamo z ustrezno srednjo vrednostjo. Variacijski razmik R primerjamo običajno z aritmetično sredino med skrajnima vrednostima $1/2\,(y_{min} + y_{max})$, kvartilni odklon Q z mediano ali aritmetično sredino med Q_1 in Q_3, poprečni absolutni odklon AD z mediano ali aritmetično sredino, standardni odklon SD z aritmetično sredino M, poprečno razliko Δ_1 z aritmetično sredino M.

Tako dobimo naslednje relativne mere variacije:

Na osnovi variacijskega razmika

$$\frac{2(y_{max} - y_{min})}{y_{min} + y_{max}} \qquad (10.26)$$

Na osnovi kvartilnega razmika Q_3-Q_1

$$\frac{Q_3 - Q_1}{M_e} \quad \text{ali} \quad \frac{2(Q_3 - Q_1)}{Q_1 + Q_3} \qquad (10.27)$$

Na osnovi poprečnega absolutnega odklona AD

$$\frac{AD_{Me}}{M}, \quad \frac{AD_M}{M} \qquad (10.28)$$

Na osnovi standardnega odklona σ

$$KV = \frac{\sigma}{M} \qquad (10.29)$$

Na osnovi poprečne razlike Δ_1

$$\frac{\Delta_1}{M} \qquad (10.30)$$

Koeficiente variacije izražamo kakor je nakazano v gornjih izrazih ali v odstotkih tako, da zgornja razmerja pomnožimo s 100. Pobliže se bomo seznanili z relativnima merama variacije, ki sta zasnovani na poprečnem absolutnem odklonu in na standardnem odklonu.

10.30 Če proučimo relativno mero variacije zasnovano na AD_{Me} dobimo iz obrazca 10.6

$$\frac{AD_{Me}}{M} = Y^o_z - Y^o_s = 2\,Y^o_z - 1 = 1 - 2\,Y^o_s \qquad (10.31)$$

Pri tem sta Y^o_s in Y^o_z strukturna deleža za agregat Y pod oziroma nad mediano.

Za relativno mero variacije, zasnovano na AD_M pa je

$$\frac{AD_M}{M} = (Y^o_z - Y^o_s) - (N^o_z - N^o_s) = 2\,(Y^o_z - N^o_z) = 2\,(N^o_s - Y^o_s) \qquad (10.32)$$

Pri AD_M/M pomeni Y^o_s in Y^o_z strukturni delež za agregat Y pod oziroma nad poprečjem M, N^o_s oziroma N^o_z pa delež enot pod oziroma nad poprečjem M.

10.31 Za zgled izračunajmo AD_{Me}/M za blagovni promet na prebivalca iz 8.1. Ker je

$$Y^o_z = Y_z/Y = 183978/263325 = 0,69867,$$

je ustrezen koeficient variacije

$$2\,Y^o_z - 1 \quad = \quad 2.0,69867 - 1 = 0,39734$$

$$AD_M/M \text{ ustreza } N^o_z = 18/60 = 0,30 \text{ in } Y^o_z = Y_z/Y =$$

$$= 137076/263325 = 0,52056\text{ . Iz obrazca 10.32 sledi}$$

$$AD_M/M \quad = 2\,(.52056 - .30) = .44112$$

10.32 Koeficient variacije, izračunan na osnovi standardnega odklona ima operativno obliko

$$KV = \sqrt{\frac{Q_i}{Q} - 1}\ ; \quad Q_i = \sum y_i^2\ ; \quad Q = \frac{Y^2}{N} \qquad (10.33)$$

Standardni odklon za ceno jabolk, ki pokaže variabilnost cene jabolk med 35 izbranimi kraji v Jugoslaviji, je bil v septembru 1957 SD = 13,2 din, v januarju 1958 pa SD = 27,4 din. Absolutna mera variacije pokaže, da je bila regionalna variacija cen jabolk v januarju večja kakor v septembru. Če pa računamo, da je bila poprečna cena jabolk v septembru 1957 40,4 din, v januarju 1958 pa 113,6 din, sklepamo, da je bila variabilnost v ceni jabolk v septembru pomembnejša kakor v januarju, ker je koeficient variacije cen KV v septembru 1957

$$KV\% = 100\ .\ \frac{13,2}{40,4} = 32,7\ \%$$

v januarju 1958 pa

$$KV\% = 100 \cdot \frac{27,4}{113,6} = 24,0\% .$$

V septembru je standardni odklon približno tretjino od poprečja, v januarju pa samo približno četrtino.

Še bolj izrazito vidimo pomen relativne mere variacije v regionalnem variiranju cen za več različnih predmetov.

Tabela 10.5. Koeficienti variacije za pet predmetov široke porabe med 35 izbranimi kraji v Jugoslaviji v januarju 1958 in za 20 izbranih krajev za Ø 1970. Vir: Po podatkih v SB 126)

Naziv artikla		januar 1958			Ø 1970
		v din			
		M	SD	KV%	KV%
Enoten kruh	kg	47,3	2,5	5,3	5,1
Svinjska mast uvožena	kg	325,1	9,8	3,0	2,6
Jabolka	kg	113,6	27,4	24,1	17,0
Jajca	kos	17,8	3,8	21,3	9,4
Prava kava	kg	2086,0	54,2	2,6	3,6

V tabeli 10.5 vidimo, da je standardni odklon za ceno prave kave SD = 54,2 din; ta je od vseh navedenih predmetov največji, stvarno pa najmanj pomemben, ker znaša komaj KV% = 2,6% od poprečne cene. Iz tabele vidimo, da sta koeficienta variacije za jabolka in jajca razmeroma izredno velika. Vzrok temu je tipično krajevni značaj teh predmetov, če jih primerjamo z drugimi tremi. Čeprav je v tabeli 10.5 za januar 1958 vključenih 35 izbranih krajev, v letu 1970 pa samo 20, moremo vseeno izvesti tudi časovno primerjavo koeficientov variacije. Medtem ko sta se koeficienta variacije za kruh in mast relativno malo spremenila, se izkaže, da je za poprečne cene leta 1970 koeficient variacije za jabolka in jajca znatno manjši kot v januarju 1958. Ne glede na časovni razmik dvajset let je ta razlika rezultat sezonskega značaja obeh predmetov. Izraz razlik v času pa je po vsej verjetnosti povečanje koeficienta variacije za pravo kavo.

MERE KONCENTRACIJE

10.33 Agregati morejo biti porazdeljeni po enotah populacije različno. Tako dobe posamezni delavci različen del skupnega fonda osebnih dohodkov. Skupna kmetijska površina je neenakomerno razdeljena med posamične kmetijske obdelovalce in imajo neki obdelovalci velike površine, drugi pa majhne. Podobno imajo od skupnega števila zaposlenih neka podjetja malo, druga pa veliko zaposlenih. Čeprav gornja lastnost agregatov pride do izraza v variiranju, vpeljemo za nakazano lastnost agregatov nov pojem - koncentracijo pojava. Če je pojav enakomerno porazdeljen po enotah, pravimo da koncentracije ni, če je pojav porazdeljen po enotah populacije tako, da ima nekaj enot razmeroma velik del agregata, večina pa le majhen del, govorimo o veliki koncentraciji pojava.

Tako opredeljen pojem koncentracije v statističnem smislu je premosorazmeren z variacijo pojava. Če pojav ne variira, koncentracije ni, velika koncentracija pa se izraža v velikem variiranju.

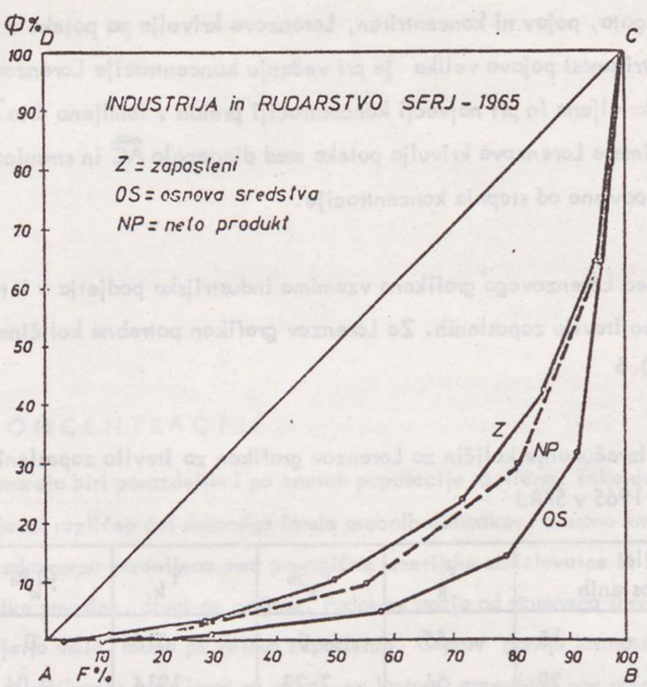
10.34 Lorenzov grafikon. Ker frekvenčna porazdelitev kaže variabilnost pojava, je iz nje posredno mogoče sklepati tudi na koncentriranost. Lepše kot iz grafičnega prikaza za frekvenčno porazdelitev pa je koncentracija pojava vidna iz Lorenzovega grafikona. Vzemimo, da imamo v frekvenčni porazdelitvi razen frekvenc f_k dane še podatke o delu agregata, ki odpade na posamezne razrede Y_k. Lorenzov grafikon narišemo tako, da v kvadrat s stranicama za kumulative relativnih frekvenc F % (abscisa) in kumulative relativnih vsot Ø % (ordinata) vnesemo točke s koordinatama (F_k%, $Ø_k$%). Če povežemo po vrsti vrisane točke, dobimo lomljeno črto, ki v loku povezuje spodnje levo oglišče kvadrata (0,0) z zgornjim desnim ogliščem (100,100). V primeru, da na vsako enoto odpade enak del agregata, pojav ni koncentriran, Lorenzova krivulja pa poteka v diagonali. Če je koncentriranost pojava velika je pri večanju koncentracije Lorenzova krivulja bolj in bolj ukrivljena in pri največji koncentraciji preide v lomljeno črto $\overline{AB} - \overline{BC}$. Za stvarne primere Lorenzova krivulja poteka med diagonalo $\overline{AC}$ in stranicama kvadrata $\overline{AB} - \overline{BC}$, odvisno od stopnje koncentracije.

10.35 Za zgled Lorenzovega grafikona vzemimo industrijska podjetja v letu 1965 v SFRJ po številu zaposlenih. Za Lorenzov grafikon potrebne količine so izračunane v tabeli 10.6

Tabela 10.6 Izračunanje količin za Lorenzov grafikon za število zaposlenih v letu 1965 v SFRJ

število zaposlenih			f_k	F_k %	Y_k	$Ø_k$ %
	-	15	55	0	516	0
16	-	29	56	2,23	1214	0,04
30	-	60	171	4,50	7979	0,13
61	-	125	407	11,44	37457	0,71
126	-	250	547	27,94	99062	3,45
251	-	500	542	50,12	192701	10,68
501	-	1000	349	72,10	246873	24,77
1001	-	2000	216	86,25	303859	42,80
2001	-		123	95,01	478814	65,01
			2466	100,00	1368475	100,00

V grafikonu 10.1 sta razen za število zaposlenih v industriji in rudarstvu narisani Lorenzovi krivulji še za osnovna sredstva in neto produkt za SFRJ v letu 1965. Ker vrisujemo v Lorenzov grafikon relativna števila, moremo v isti grafikon vrisati Lorenzove krivulje za raznovrstne pojave, ki pa morajo biti seveda v vsebinski zvezi. Iz grafikona sklepamo, da je koncentracija podjetij glede na neto produkt malenkostno višja kot za zaposlene, da pa je koncentracija osnovnih sredstev po podjetjih v industriji znatno večja kot je koncentracija zaposlenih.



Grafikon 10.1 Lorenzov grafikon za število zaposlenih, vrednost osnovnih sredstev in vrednost neto produkta v industriji in rudarstvu v SFRJ v letu 1965.

10.36 V Lorenzovem grafikonu z daljicami, ki vežejo posamezne točke, podajamo linearno interpolacijo poteka za Lorenzovo krivuljo. Zato moremo iz njega odbrati raznovrstne odnose med kumulativo relativnih vsot in frekvenc. Tako moremo iz njega oceniti nekaj pomembnih odnosov. Med drugim so pomembni odnosi, kolik del agregata ima polovica manjših enot in kolik del največjih enot v populaciji ima polovico agregata. Iz grafikona ocenimo, da ima polovica podjetij (majhnih), zaposlenih 11% od vseh zaposlenih v industriji in rudarstvu, 8,5% neto produkta in 4% vseh osnovnih sredstev. Obratno pa odberemo, da ima 11% velikih podjetij polovico vseh zaposlenih, da ima 10,5% velikih podjetij polovico neto produkta in 6% velikih podjetij polovico vseh osnovnih sredstev.

Lorenzov grafikon ima razen tega, da moremo na njem narisati Lorenzove krivulje za raznovrstne pojave tudi to prednost, da moremo na isti grafikon risati Lorenzove krivulje iz različno grupiranih podatkov. To posebno olajša mednarodne primerjave, ker ima vsaka država podatke grupirane prilagojeno svojstvenim pogojem in potrebam.

10.37 Včasih razpolagamo za manjše populacije s posameznimi podatki, ne pa s frekvenčno porazdelitvijo. V tem primeru frekvenčno porazdelitev zamenja ranžirna vrsta, v kateri so frekvence v posameznem razredu $f_k = 1$, vrednosti $Y_k = y_k$ pa posamezne vrednosti enot.

Če za proučevan pojav razpolagamo le s frekvenčno porazdelitvijo f_k, nimamo pa vsot Y_k, vsote Y_k ocenimo po obrazcu $Y_k = f_k y_k$ kot produkt frekvence f_k s sredino razreda y_k. Težava je, če je zadnji razred odprt in zanj nimamo sredine razreda. V tem primeru z analizo pojava določimo ali ocenimo pravo vrednost vsote agregata.

10.38 O psevdo-Lorenzovih grafikonih govorimo, kadar se vsote v razredih nanašajo na drug znak kot je osnovni znak frekvenčne porazdelitve.

Tako moremo za frekvenčno porazdelitev podjetij po številu zaposlenih imeti za posamezne razrede dano skupno vrednost osnovnih sredstev za podjetja v teh razredih. Podobno moremo imeti za kmetijska gospodarstva, ki so porazdeljena po velikosti površine, število živine po posameznih razredih. Formalno moremo v vseh teh primerih računati kumulative relativnih frekvenc in kumulative relativnih vsot in vrisati te podatke v grafikon, ki je podoben Lorenzovemu grafikonu, ni pa Lorenzov grafikon v pravem smislu, ker se frekvenčna porazdelitev nanaša na drug znak kot pa vsota. Zato imenujemo tovrstne grafikone psevdo-Lorenzove grafikone. Medtem ko more biti krivulja v pravem Lorenzovem grafikonu le v trikotniku $\triangle ABC$, more biti krivulja v psevdo-Lorenzovem grafikonu na celem kvadratu $\square$ ABCD. Vsekakor ima psevdo-Lorenzov grafikon svojo analitično vrednost, vendar je treba pri analizi upoštevati njegove posebnosti.

vsaka državo podatke grupirane prilagojeno svojevrstnim pogojem in potrebam.

10.37 Včasih razpolagamo za nekatere populacije z nekonzistentnimi podatki, oz. pa s frekvenčno porazdelitvijo. V tem primeru frekvenčna porazdelitev zamenja ranžirno vrsto, v kateri so frekvence v posameznem razredu $f_k = n_k$, vrednosti $Y_k = y_k$ pa ponazarjajo vrednosti enot.

Če za proučevani pojav razpolagamo le s frekvenčno porazdelitvijo f_k, izračunamo po vrsti Y_k, vsote Y_k ocenimo po obrazcu $Y_k = f_k y_k$, kot produkt frekvence f_k s sredino razreda y_k. Težava je, če je zadnji razred odprt in zanj nimamo sredine razreda. V tem primeru z analizo pojava določimo ali ocenimo pravo vrednost vsote agregata.

10.38 O psevdo-Lorenzovih grafikonih govorimo, kadar se vsote v razredih nanašajo na drug znak kot je osnovni znak frekvenčne porazdelitve.

Tako moremo za frekvenčno porazdelitev podjetij po številu zaposlenih imeti za posamezne razrede dano skupno vrednost osnovnih sredstev za podjetja v teh razredih. Podobno moremo imeti za kmetijska gospodarstva, ki so porazdeljena po velikosti površine, število živine po posameznih razredih. Formalno moremo izračunati kumulative relativnih frekvenc in kumulative relativnih vsot in vrisati te podatke v grafikon, ki je podoben Lorenzovemu grafikonu, ni pa Lorenzov grafikon v pravem smislu, ker se frekvenčna porazdelitev nanaša na drug znak kot pa vsote. Zato imenujemo tovrstne grafikone psevdo-Lorenzove grafikone. Medtem ko more biti krivulja v pravem Lorenzovem grafikonu le v trikotniku △ ABC, more biti krivulja v psevdo-Lorenzovem grafikonu na celem kvadratu □ ABCD. Vsekakor ima psevdo-Lorenzov grafikon svojo analitično vrednost, vendar je treba pri analizi upoštevati njegove posebnosti.

ENAJSTO POGLAVJE

PROUČEVANJE DINAMIKE POJAVOV-ČASOVNE VRSTE

11.1 Socialno-ekonomski pojavi niso časovno nespremenljivi. Spremembe so rezultat delovanja najrazličnejših dejavnikov, ki v tej ali oni obliki vplivajo na pojave. Sliko dinamike pojavov dobimo s časovnimi vrstami. Ena slika daje statičen prikaz. Niz slik, ki slede na filmskem platnu, pa daje videz gibanja. Enako dá en sam podatek statično sliko pojava, niz istovrstnih podatkov v enakih časovnih razmikih pa daje sliko dinamike pojava.

Časovna vrsta je niz istovrstnih podatkov, ki se nanašajo na zaporedne časovne razmike ali trenutke. Z njimi proučujemo časovni razvoj pojavov, ker prikazujejo spremembe pojavov v odvisnosti od časa. Po namenu in načinu proučevanja se razlikujejo od drugih statističnih vrst. Osnovni namen proučevanja časovnih vrst je, opazovati časovni razvoj socialno-ekonomskih pojavov, iskati njegove zakonitosti in napovedati nadaljnji razvoj. V pogojih planskega gospodarstva je proučevanje časovnih vrst važno tako pri sestavljanju plana kot pri kontroli izvajanja plana. Predvidevanje in napovedovanje razvoja ekonomskih pojavov, če poznamo zakonitosti in sedanje stanje, se je razvilo v posebno disciplino-vedo o konjunkturah, ki s proučevanjem časovnih vrst najvažnejših ekonomskih pokazovalcev skuša sklepati na nadaljnji ekonomski razvoj. Seveda to napovedovanje ne more biti popolnoma zanesljivo, ker je 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

Trenutne in razmične časovne vrste

11.2 Časovne vrste po značaju podatkov, ki jih prikazujejo, delimo na trenutne in razmične. Trenutne časovne vrste prikazujejo časovni razvoj za pojave, ki veljajo za posamezne trenutke; razmične časovne vrste pa časovni razvoj pojavov, ki so razmičnega značaja. Trenutna časovna vrsta je npr. vrsta podatkov o številu delavstva konec posameznega meseca, primer razmične časovne vrste pa je proizvodnja v podjetju po mesecih. Enako je v tabeli 11.1 število prebivalstva v SFRJ po letih sredi leta trenutna, naravni prirastek po letih pa razmična časovna vrsta.

Tabela 11.1: Število prebivalstva in naravni prirastek prebivalstva v SFRJ (Vir: SG - 1971)

Leto	Število prebivalstva sredi leta-v tisočih	Naravni prirastek v tisočih
1960	18402	250
1961	18612	255
1962	18819	226
1963	19029	238
1964	19222	220
1965	19434	238
1966	19644	240
1967	19840	216
1968	20029	208
1969	20209	194
1970	20371	179

Izvedene časovne vrste

11.3 Iz osnovnih časovnih vrst moremo izračunati več izvedenih časovnih vrst, ki pripomorejo k proučitvi dinamike pojavov. Izmed teh bomo navedli: kumulativno časovno vrsto, časovno vrsto sredin, časovno vrsto drsečih vsot in časovno vrsto drsečih sredin.

Kumulativna časovna vrsta

11.4 Kumulativno časovno vrsto izračunavamo samo za razmične vrste ekstenzivnih podatkov, ker ima samo zanje kumuliranje oziroma vsota členov smisel. Kumulativno časovno vrsto dobimo s postopnim seštevanjem členov osnovne vrste. Členi kumulativne vrste pomenijo vrednosti v razmiku od začetka proučevanega razdobja do določenega trenutka, ki se od člena do člena menja. Kumulativno časovno vrsto računamo običajno za zaključene časovne razmike kot so: mesečne vrednosti v razdobju enega leta, dnevne vrednosti v razdobju enega meseca, letne vrednosti v razdobju ene petletke itd. Kumulativne vrste se v teh primerih nanašajo na začetek petletke, začetek leta, začetek meseca.

Kumulativno časovno vrsto dobimo tako, da vzamemo vrednost prvega člena kumulativne vrste enako 0, nadaljnje člene kumulativne vrste pa izračunamo po znanem obrazcu

$$K_{k+1} = K_k + Y_k \tag{11.1}$$

K_k = kumulativa ; Y_k = vrednost v tekočem razdobju k

11.5 Za zgled izračunanja kumulativne časovne vrste vzemimo mesečno vrsto proizvodnje krogličnih ležajev v SFRJ.

Tabela 11.2: Izračunanje kumulativne časovne vrste za proizvodnjo krogličnih ležajev v SFRJ v letu 1957 (Vir: Indeks)

Mesec	Mesečna proizvodnja v tonah	Kumulativa
J	17	0 = 0
F	16	0 + 17 = 17
M	22	17 + 16 = 33
A	21	33 + 22 = 55
M	19	55 + 21 = 76
J	19	76 + 19 = 95
J	20	95 + 19 = 114
A	5	114 + 20 = 134
S	18	134 + 5 = 139
O	14	139 + 18 = 157
N	18	157 + 14 = 171
D	18	171 + 18 = 189
Skupaj		189 + 18 = 207

Šesti člen v kumulativni vrsti 95 ton pome ni, da je bilo od začetka leta 1957 do začetka meseca junija v SFRJ proizvedeno 95 ton krogličnih ležajev. Podoben pomen imajo drugi členi kumulativnih vrst. Zadnji člen v kumulativni časovni vrsti je enak vsoti vseh členov iz osnovne časovne vrste.

Vrsta sredin

11.6 Časovne vrste z velikim številom členov so dostikrat nepregledne in je osnovna dinamika zaradi prevelike dolžine vrste zabrisana. Za take vrste zmanjšamo število členov z združevanjem členov v vsote ali povprečja. Iz časovne vrste o številu tujih gostov po mesecih v razdobju 1956-1967, ki ima 144 članov, dobimo časovno vrsto z 12 členi, če seštejemo število tujih gostov po mesecih v letne vsote. Seveda ta časovna vrsta ne podaja tako natančno gibanje proučevanega pojava kot osnovna vrsta. S tem zabrišemo marsikatero značilnost pojava.

Medtem ko ima vrsta vsot smisel le, če je vrsta, ki jo prikazujemo, razmična in ekstenzivna, moremo z grupnimi poprečji skrčiti število členov za vsako časovno vrsto ne glede na značaj členov.

Člene v osnovni časovni vrsti združujemo po naravnih časovnih enotah. Tako združujemo mesečne podatke v četrtletne ali letne, dnevne podatke v tedenske ali mesečne, letne podatke pa običajno v petletja, tako da iz njih dobimo zaokrožena petletja ali desetletja, npr. petletja 1900-1904, 1905-1909, 1910-1914...1945-1949, 1950-1954.

11.7 Če izračunavamo časovno vrsto poprečij, je treba paziti, na katere trenutke se nanašajo osnovne vrednosti v časovni vrsti. Od tega je namreč odvisno, kako računamo sredine.

Če se nanašajo vrednosti na sredine osnovnih razmikov kot je to primer pri razmičnih časovnih vrstah in pri nekaterih trenutnih, računamo poprečja po obrazcu

$$\bar{Y} = \frac{1}{r}(Y_1 + Y_2 + \ldots + Y_r) \qquad (11.2)$$

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

$$\bar{Y} = \frac{1}{r}\left(\frac{1}{2}Y_0 + Y_1 + Y_2 + \ldots + Y_{r-1} + \frac{1}{2}Y_r\right) \qquad (11.3)$$

Po obrazcu 11.2 računamo na primer poprečno mesečno proizvodnjo po letih, če imamo mesečne podatke o proizvodnji, ali poprečno število prebivalstva po petletjih, če razpolagamo s številom prebivalstva po letih sredi leta. Obrazec 11.3 pa upoštevamo, če izračunavamo na primer poprečno število delavcev po letih, imamo pa podatke o številu delavcev v začetku vsakega meseca itd. Na to moramo paziti zato, da se poprečja nanašajo na zaokrožena koledarska razdobja: tedne, mesece, leta, petletja in na iste trenutke kot členi v osnovni časovni vrsti.

Vrsta drsečih vsot

11.8 Iz mesečnih podatkov o proizvodnji s seštevanjem dobimo časovno vrsto za letno proizvodnjo. Ti letni podatki se nanašajo na letne razmike koledarskih let od januarja do decembra. Pri proučevanju dinamike pa so poleg teh važni tudi drugi letni razmiki, na primer od februarja danega leta do januarja naslednjega leta, od marca do februarja naslednjega leta itd. Ti letni razmiki se med seboj delno prekrivajo, tako da je vsak razmik, če ga primerjamo s predhodnim, premaknjen za en mesec. Ti letni razmiki tako rekoč drsijo od člena do člena za en mesec. Odtod tudi ime drseče vsote. Časovna vrsta drsečih vsot ima niz lastnosti, ki ji dajo posebno analitično vrednost. Ena izmed teh je, da se v časovni vrsti drsečih vsot izravnavajo slučajni in periodični vplivi, če obsega vsota časovno periodo. O tem bomo podrobneje slišali v naslednjih poglavjih.

Zaradi zveze med zaporednimi členi v časovni vrsti drsečih vsot računamo časovno vrsto drsečih vrednosti razmeroma enostavno. Če imamo izračunan prvi člen časovne vrste drsečih vsot (dobimo ga tako, da seštejemo vrednosti od januarja do decembra za prvo leto v časovni vrsti), dobimo naslednji člen v časovni vrsti drsečih vsot tako, da januarsko vrednost od prve letne vsote odštejemo, prištejemo pa vrednost za januar naslednjega leta. Iz tako dobljenega drugega člena dobimo tretji člen časovne vrste drsečih vsot, če prištejemo februarski člen prihodnjega in odštejemo februarski člen prejšnjega leta iz osnovne časovne vrste. Posamezni členi v vrsti drsečih vsot pomenijo letno proizvodnjo v preteklem letu do začetka januarja, do začetka februarja itd.

Nakazano računanje vrste drsečih vsot ne velja samo za mesečne podatke, marveč na splošno za katero koli vrsto ekstenzivnih razmičnih podatkov. Enako iz dnevnih podatkov računamo časovno vrsto tedenskih drsečih vrednosti o številu prevoženih potnikov. Iz letnih podatkov o številu živorojenih pa moremo sestaviti petletno vrsto drsečih vsot za število živorojenih itd.

Drseče vsote v vseh primerih računamo po splošnem obrazcu:

$$S_{k+1} = S_k + Y_k - Y_{k-r} = S_k + {}_r d_k \qquad (11.4)$$

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.

11.9 Za število potnikov v zračnem prometu v SFRJ v razdobju 1955-1958 je izračunanje časovne vrste drsečih vsot nakazano v tabeli 11.3.

Tabela 11.3 Izračunanje časovne vrste drsečih vsot za število potnikov v zračnem prometu po četrtletjih v SFRJ v razdobju 1955-1958 (Vir: Indeks)

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

Iz zgleda je jasno razviden postopek:

$S_5=8,2+23,1+38,0+14,7=104,0$; $S_6=104,0+(+0,1)=104,1$; $S_7=104,1+(-3,2)=100,9$ itd.

Časovna vrsta drsečih sredin

11.10 Vrednosti členov v časovni vrsti sredin priredimo sredini razmikov, na katere se poprečja nanašajo (npr. sredini let pri letnih poprečjih, tretjemu letu petletke pri petletnih poprečjih itd.). Členi v časovni vrsti sredin so torej prirejeni samo nekaterim členom v osnovni časovni vrsti. Za analizo pa je dostikrat potrebno, da imamo prirejena poprečja vsem, ne pa samo nekaterim členom v osnovni časovni vrsti. To dosežemo s časovno vrsto drsečih sredin, ki jo izračunavamo podobno kot vrsto drsečih vsot.

Ker se morajo posamezni členi v vrsti drsečih sredin nanašati na iste trenutke kot členi v osnovni časovni vrsti, izračunamo člene vrste drsečih vsot glede na osnovno časovno vrsto na dva načina:

Če imajo razmiki, na katere se nanašajo sredine, liho število osnovnih razmikov: $r = 2i+1$, izračunavamo časovno vrsto drsečih sredin po obrazcu

$$\bar{y}_k = \frac{1}{r}(Y_{k-i} + Y_{k-i+1} + \ldots + Y_k + \ldots + Y_{k+i-1} + Y_{k+i}) = \frac{1}{r} S_{k+i+1} \qquad (11.5)$$

Tak primer je računanje tedenskih ali petletnih poprečij.

11.11 Vzemimo za zgled časovno vrsto o številu umrlih na 1000 prebivalcev v Jugoslaviji v letih 1921-1939 in zanjo izračunajmo časovno vrsto petletnih sredin. Ker računamo poprečja iz lihega števila osnovnih razmikov, uporabimo obrazec 11.5.

Iz obrazca 11.5 povzamemo, da je najbolje, če po postopku, ki smo ga že navedli, izračunamo vrsto drsečih petletnih vsot S_k, petletne vsote pa delimo s pet.

Tabela 11.4 Izračun časovne vrste drsečih sredin za število umrlih na 1000 prebivalcev v bivši Jugoslaviji v letih 1921-1939

Leto	Y_k	Y_{k-5}	$_5d_k$	S_k	$\tilde{Y}_k$
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	

11.12 Če pa sredine veljajo za sodo število osnovnih razmikov (r = 2i), računamo drseče sredine po obrazcu:

$$\bar{Y}_k = \frac{1}{r}\left(\frac{1}{2} Y_{k-i} + Y_{k-i+1} + \ldots + Y_k + \ldots + Y_{k+i-1} + \frac{1}{2} Y_{k+i}\right) \qquad (11.6)$$

Tak primer imamo pri letnih poprečjih iz mesečnih (r = 12) ali četrtletnih (r = 4) podatkov.

Če je število osnovnih razmikov sodo (r = 2i), je najprikladneje obrazec 11.6 preurediti v obliko

$$\bar{Y}_k = \frac{1}{2r}(Y_{k-i} + 2Y_{k-i+1} + \ldots + 2Y_k + \ldots + 2Y_{k+i-1} + Y_{k+i}) = \frac{1}{2r} S'_k \qquad (11.7)$$

Ker je

$$S'_{k+1} = S_{k+i+1} + S_{k+i+2} = S_{k+i} + {}_r d_{k+i} + S_{k+i+1} + {}_r d_{k+i+1} = S'_k + {}_r d_{k+i} + {}_r d_{k+i+1}$$

je ponderirano vsoto S'_k najprikladneje računati po obrazcu 11.8 tako, da prvo vsoto izračunamo, naslednje pa dobimo s postopnim prištevanjem dveh diferenc ${}_r d_k$.

11.13 Če imamo vrsto mesečnih podatkov, po pravilu računamo vrsto drsečih sredin za letna razdobja. Za ta primer je po gornjih obrazcih: $r=2i=12$; $i = 6$.

Ker začnemo s $S'_{i+1} = S'_{6+1} = S'_7$, se prvi člen drsečih sredin nanaša na 7. mesec (julij) prvega leta. Zanj računamo S'_7 po obrazcu 11.7 :

$$S_7 = (Y_1 + 2Y_2 + \ldots + 2Y_7 + \ldots + 2Y_{12} + Y_{13})$$

Y_{13} = podatek za januar drugega leta. Po obrazcu 11.8 pa dobimo dalje:

$S_8 = S_7 + {}_{12}d_{13} + {}_{12}d_{14}$: $S_9 = S_8 + {}_{12}d_{14} + {}_{12}d_{15}$ itd ..... ${}_{12}d_{13} = Y_{13} - Y_1$:

${}_{12}d_{14} = Y_{14} - Y_2$ itd.

Kot zgled vzemimo indeks industrije za gradbeni material v SFRJ. (Glej tabelo 11.51).

Razlike med ustreznimi meseci zaporednih let smo izračunali takole: Prva možna razlika je za mesec januar 1952:

${}_{12}d_{13} = Y_{J,52} - Y_{J,51} = 31-38 = -7$; za februar 1952: ${}_{12}d_{14} = Y_{F,52} - Y_{F,51} =$
$= 27-37 = -10$ itd.

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

$S'_{J,51} = 38 + 2{,}37 + 2.49 + \ldots + 2.84 + 2.71 + 2.58 + 31 = 1707.$

Nadaljnje vsote pa so izračunane, kot kaže slika 11.1

Tabela 11.5 Izračunanje časovne vrste drsečih sredin za indeks industrije gradbenega materiala v SFRJ (Ø 1956 = 100)

(Vir: SB 108)

		J	F	M	A	M	J	J	A	S	O	N	D
1951	Y	38	37	49	56	78	95	97	104	90	84	71	58
	d												
S′= 24	$\bar{Y}$							1707	1690	1667	1648	1635	1634
	$\bar{Y}$							71,1	70,4	69,5	68,7	68,1	68,1
1952	Y	31	27	36	50	71	101	124	122	112	98	71	56
	d	-7	-10	-13	-6	-7	+6	+27	+18	+22	+14	0	-2
S′= 24	$\bar{Y}$	1667	1712	1752	1788	1802	1800	1801	1810	1828	1858	1897	1936
	$\bar{Y}$	69,5	71,3	73,0	74,5	75,1	75,0	75,0	75,4	76,2	77,4	79,0	80,7
1953	Y	34	33	48	68	92	119	125	131	127	118	86	74
	d	+3	+6	+12	+18	+21	+18	+1	+9	+15	+20	+15	+18
S′= 24	$\bar{Y}$	1955	1965	1989	2024	2059	2092	2108	2095	2084	2081	2071	2062
	$\bar{Y}$	81,5	81,9	82,9	84,3	85,8	87,2	87,8	87,3	86,8	86,7	86,3	85,9
1954	Y	32	22	48	65	85	117	134	143	145	122	90	83
	d	-2	-11	0	-3	-7	-2	+9	+12	+18	+4	+4	+9
S′= 24	$\bar{Y}$	2069	2090	2120	2142	2150	2163	2195	2247	2282	2284	2292	2332
	$\bar{Y}$	86,2	87,1	88,3	89,3	89,6	90,1	91,5	93,6	95,1	95,2	95,5	97,2
1955	Y	55	51	54	61	97	145	138	144	135	121	89	81
	d	+23	+29	+6	-4	+12	+28	+4	+1	-10	-1	-1	-2
S′= 24	$\bar{Y}$	2364	2369	2360	2349	2347	2344	2339	2311	2276	2266	2256	2232
	$\bar{Y}$	98,5	98,7	98,3	97,9	97,8	97,7	97,5	96,3	94,8	94,4	94,0	93,0
1956	Y	52	26	44	61	87	131	139	157	155	154	106	89
	d	-3	-25	-10	0	-10	-14	+1	+13	+20	+33	+17	+8
S′= 24	$\bar{Y}$	2219	2233	2266	2319	2369	2394	2394	2398	2420	2445	2492	2547
	$\bar{Y}$	92,5	93,0	94,4	96,6	98,7	99,8	99,8	99,9	100,8	101,9	103,8	106,1
1957	Y	44	38	54	76	119	154	172	179	166	152	121	97
	d	-8	+12	+10	+15	+32	+23	+33	+22	+11	-2	+15	+8
S′= 24	$\bar{Y}$	2603	2658	2691	2700	2713	2736						
	$\bar{Y}$	108,5	110,8	112,1	112,5	113,0	114,0						

1707 ◇
- 7 +
-10 +
1690 ◇
-10 +
-13 +
1667 ◇
-13 +
- 6 +
1648 ◇
- 6 +
.
.
.

Slika 11.1 Izračunanje vrste drsečih vsot S'_k.

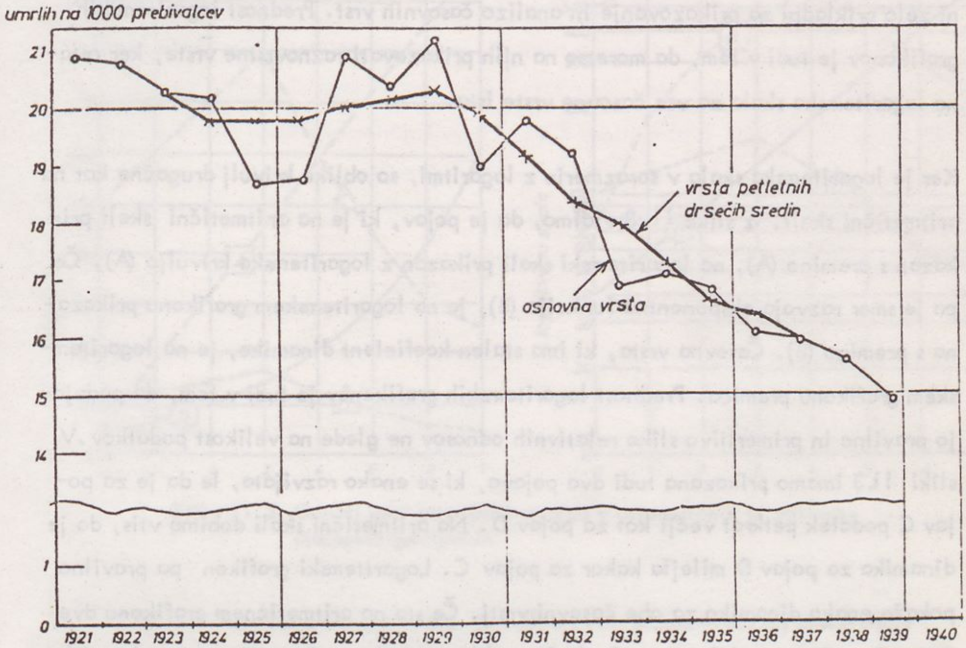
Iz obeh zgledov vidimo, da je časovna vrsta drsečih sredin krajša od osnovne časovne vrste in da začetnim in končnim i členom v osnovni časovni vrsti ne moremo prirediti drseče sredine. To je hiba časovne vrste drsečih sredin, s katero moramo računati.

GRAFIČNO PRIKAZOVANJE ČASOVNIH VRST

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

Časovne vrste prikazujemo pretežno z linijskimi grafikoni. Stolpce ali druge načine prikazovanja uporabljamo le za krajše časovne vrste.

Najenostavnejši prikaz časovne vrste je navaden linijski grafikon. V njem je količinska skala navadna aritmetična skala. Teh vrst grafikonov imamo največ. Eden izmed njih je grafikon v sliki 11.2.



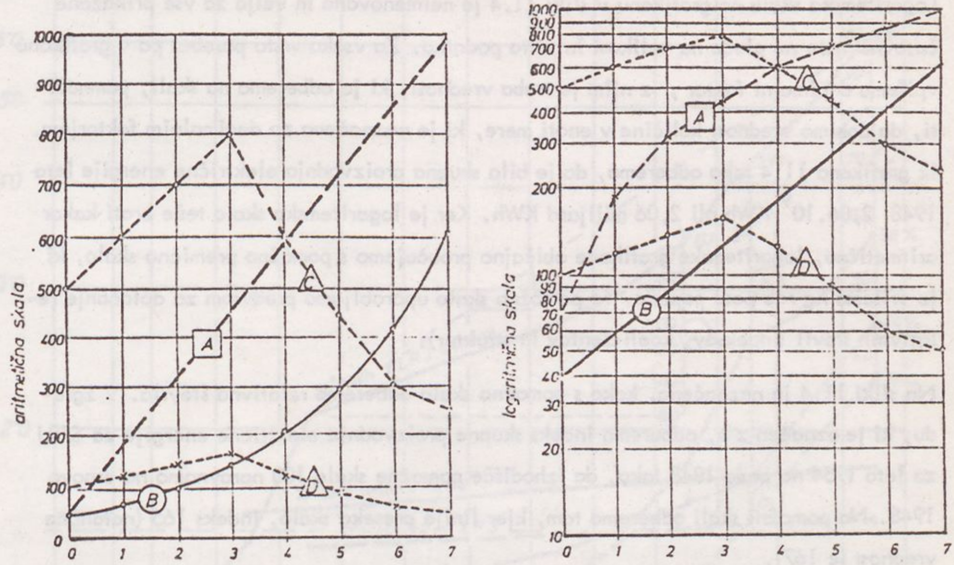
Slika 11.2 Število umrlih na tisoč prebivalcev v predvojni Jugoslaviji

Na aritmetični skali daljice med podatki prikazujejo absolutne diference med dvema podatkoma. Na linijskih grafikonih z aritmetično skalo pa moremo praviloma prikazovati samo istovrstne časovne vrste. Pri prikazovanju raznovrstnih časovnih vrst na istem grafikonu je namreč problematično razmerje med skalami.

Pollogaritemski grafikoni

11.15 Za prikazovanje dinamike pojavov je posebno prikladen grafikon, ki ima namesto aritmetične količinske skale logaritemsko skalo. Ker je razlika logaritmov enaka logaritmu kvocienta, je na logaritemskem grafikonu navpična razdalja med dvema podatkoma v razmerju z logaritmom kvocienta med podatkoma. Zato 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 aritmetični skali. Iz slike 11.3 vidimo, da je pojav, ki je na aritmetični skali prikazan s premico (A), na logaritemski skali prikazan z logaritemsko krivuljo (A). Če pa je smer razvoja eksponentna funkcija (B), je na logaritemskem grafikonu prikazana s premico (B). Časovna vrsta, ki ima stalen koeficient dinamike, je na logaritemskem grafikonu premica. Prednost logaritemskih grafikonov je tudi v tem, da podajajo pravilno in primerljivo sliko relativnih odnosov ne glede na velikost podatkov. V sliki 11.3 imamo prikazana tudi dva pojava, ki se enako razvijata, le da je za pojav C podatek petkrat večji kot za pojav D. Na aritmetični skali dobimo vtis, da je dinamika za pojav D milejša kakor za pojav C. Logaritemski grafikon pa pravilno pokaže enako dinamiko za obe časovni vrsti. Če sta na aritmetičnem grafikonu dve časovni vrsti vzporedni, pomeni, da je razlika med obema časovnima vrstama stalna. Če pa sta časovni vrsti na logaritemskem grafikonu vzporedni, pomeni, da je stalno razmerje med obema vrstama.



Slika Ι1.3 Odnosi med časovnimi vrstami v aritmetičnih in pollogaritemskih grafikonih

Ι1.16 Na logaritemskem grafikonu je v sliki Ι1.4 prikazan razvoj proizvodnje električne energije v SFRJ in SRS.

Tabela Ι1.6. Proizvodnja električne energije in število prebivalstva v SFRJ in SRS v razdobju 1946-1957 (Vir: SG 57-58)

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

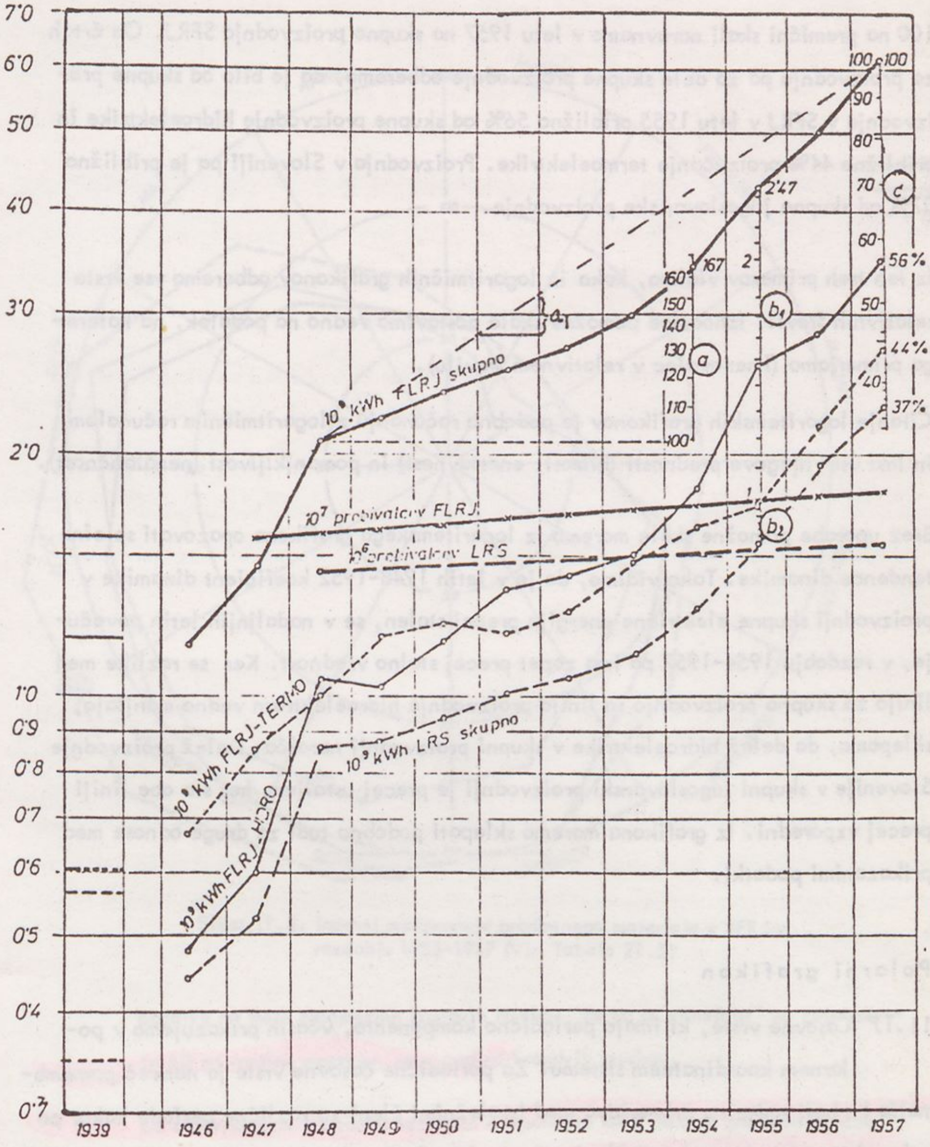
Logaritemska skala na grafikonu v sliki 11.4 je neimenovana in velja za vse prikažane časovne vrste ne glede na velikost in vrsto podatka. Za vsako vrsto posebej pa v grafikonu vpišemo decimalni faktor ; z njim je treba vrednost, ki jo odberemo na skali, pomnožiti, da dobimo vrednost količine v enoti mere, ki je naznačena za decimalnim faktorjem. Iz grafikona 11.4 tako odberemo, da je bila skupna proizvodnja električne energije leta 1948 $2,06.10^9$ KWh ali 2,06 milijard KWh. Ker je logaritemsko skalo teže brati kakor aritmetično, logaritemske grafikone običajno proučujemo s pomožno premično skalo, ki jo vrišemo na rob pasu papirja. To pomožno skalo uporabljamo predvsem za določanje relativnih števil (indeksov, koeficientov in struktur).

Na sliki 11.4 je naznačeno, kako s pomožno skalo odberemo relativna števila. V zgledu, ki je označen z a, odberemo indeks skupne proizvodnje električne energije za SFRJ za leto 1954 na bazo 1948 tako, da izhodišče pomožne skale 100 naravnamo na osnovo 1948. Na pomožni skali odberemo tam, kjer linija preseka skalo, indeks 165 (natančna vrednost je 167).

Če zvežemo točki, ki prikazujeta podatka na začetku in koncu določenega razdobja za isto časovno vrsto, letni prirastek na daljici pokaže poprečni koeficient dinamike oziroma poprečni verižni indeks, če ga merimo z ustrezno logaritemsko skalo. Za proizvodnjo električne energije v Jugoslaviji je iz grafikona (a_1) razvidno, da je poprečen koeficient dinamike v razdobju 1948-57 enak 1,14 kar ustreza poprečni stopnji rasti 14 točk.

Z b_1 je naznačeno, kako določimo koeficient proizvodnje električne energije na enega prebivalca v SFRJ v letu 1955. Ta koeficient odberemo tako, da izhodišče 1 na pomožni skali naravnamo na linijo števila prebivalstva, na skali pa pri liniji proizvodnje odčitamo ustrezno vrednost 2,47. Vrednost koeficienta pa dobimo, če to vrednost pomnožimo s kvocientom 10^9 KWh/10^7 prebivalcev ; tega dobimo iz decimalnih faktorjev, ki so vpisani ob ustreznih linijah za proizvodnjo in število prebivalstva. Koeficient je torej 247 KWh/preb. Analogen koeficient za isto leto za SR Slovenijo dobimo, če premaknemo izhodišče na pomožni skali na črto za prebivalstvo v Sloveniji, ob črti za proizvodnjo pa na pomožni skali odberemo vrednost 1,02. Po istem pravilu kot prej to vrednost pomnožimo s kvocientom 10^9 KWh/10^6 prebivalcev. Iz grafikona ocenjeni koeficient je torej 1020 KWh/preb. v letu 1955 v SR Sloveniji.

Kako s pomožno logaritemsko skalo odčitamo strukturne odstotke, je nakazano v zgledu c. Ker iščemo strukturne deleže za skupno proizvodnjo v SFRJ, izhodišče



Slika 11.4 Proizvodnja električne energije in število prebivalstva v SFRJ in SRS v razdobju 1946-1957

100 na premični skali naravnamo v letu 1957 na skupno proizvodnjo SFRJ. Ob črtah za proizvodnjo pa za dele skupne proizvodnje odberemo, da je bilo od skupne proizvodnje v SFRJ v letu 1955 približno 56% od skupne proizvodnje hidroelektrike in približno 44% proizvodnje termoelektrike. Proizvodnja v Sloveniji pa je približno 37% od skupne jugoslovanske proizvodnje.

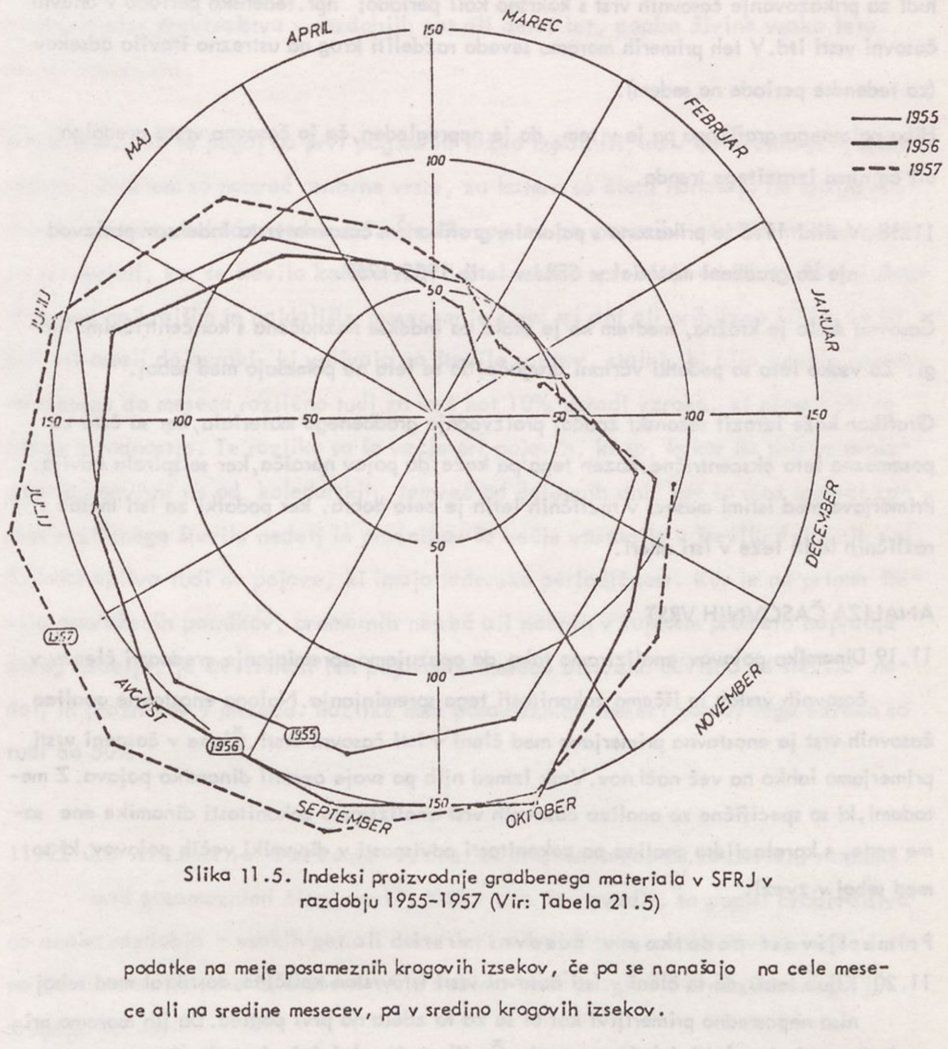
Iz teh treh primerov vidimo, kako iz logaritmičnih grafikonov odberemo vse vrste relativnih števil. Izhodišče pomožne skale postavimo vedno na podatek, na katerega primerjamo (imenovalec v relativnem številu).

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

Brez uporabe pomožne skale moremo iz logaritemskega grafikona opazovati splošne tendence dinamike. Tako vidimo, da je v letih 1948-1952 koeficient dinamike v proizvodnji skupne električne energije precej stalen, se v nadaljnjih letih povečuje, v razdobju 1954-1957 pa ima zopet precej stalno vrednost. Ker se razlike med linijo za skupno proizvodnjo in linijo proizvodnje hidroelektrarn vedno manjšajo, sklepamo, da delež hidroelektrike v skupni proizvodnji narašča. Delež proizvodnje Slovenije v skupni jugoslovanski proizvodnji je precej ustaljen, ker sta obe liniji precej vzporedni. Iz grafikona moremo sklepati podobno tudi za druge odnose med prikazanimi podatki.

Polarni grafikon

11.17 Časovne vrste, ki imajo periodično komponento, včasih prikazujemo v polarnem koordinatnem sistemu. Za periodične časovne vrste je namreč pomembnejša in bolj smiselna primerjava med istoležnimi členi za različne periode kakor pa primerjava med zaporednimi členi v časovni vrsti. S polarnim grafikonom najbolje primerjamo ustrezne člene v različnih periodah. Osnova polarnega grafikona je polarni koordinatni sistem; ta zamenja pravokotni koordinatni sistem. V polarnem grafikonu je čas znotraj ene periode nakazan s smerjo, podatek pa z oddaljenostjo točke od izhodišča polarnega koordinatnega sistema. Če imamo mesečno časovno vrsto s sezonsko variacijo, krog razdelimo v dvanajst enakih delov. Vsak izmed teh delov pomeni po en mesec. Če se podatki nanašajo na konce ali začetke mesecev, rišemo



Slika 11.5. Indeksi proizvodnje gradbenega materiala v SFRJ v razdobju 1955-1957 (Vir: Tabela 21.5)

podatke na meje posameznih krogovih izsekov, če pa se nanašajo na cele mesece ali na sredine mesecev, pa v sredino krogovih izsekov.

Za časovne vrste z izrazito sezonsko komponento kaže polarni grafikon tipično ekscentrično sliko. Če pojav narašča, dobimo na polarnem grafikonu spiralo, ki se odvija. Za pojave, ki upadajo, pa se spirala zavija proti izhodišču. Seveda polarne grafikone ne uporabljamo samo za mesečne vrste s sezonsko variacijo, temveč tudi za prikazovanje časovnih vrst s kakršno koli periodo; npr. tedensko periodo v dnevni časovni vrsti itd. V teh primerih moramo seveda razdeliti krog na ustrezno število odsekov (za tedenske periode na sedem).

Hiba polarnega grafikona pa je v tem, da je nepregleden, če je časovna vrsta predolga ali če nima izrazitega trenda.

11.18 V sliki 11.5 je prikazana s polarnim grafikonom časovna vrsta indeksov proizvodnje za gradbeni material v SFRJ v letih 1955-1957.

Časovna skala je krožna, medtem ko je skala za indekse naznačena s koncentričnimi krogi. Za vsako leto so podatki včrtani drugače, da se leta ne pomešajo med seboj.

Grafikon kaže izrazit sezonski značaj proizvodnje gradbenega materiala, ker so črte za posamezna leta ekscentrične. Razen tega pa kaže, da pojav narašča, ker se spirala odvija. Primerjava med istimi meseci v različnih letih je zelo dobra, ker podatki za isti mesec v različnih letih leže v isti smeri.

ANALIZA ČASOVNIH VRST

11.19 Dinamiko pojavov analiziramo tako, da opazujemo spreminjanje vrednosti členov v časovnih vrstah in iščemo zakonitosti tega spreminjanja. Naloga enostavne analize časovnih vrst je enostavna primerjava med členi v isti časovni vrsti. Člene v časovni vrsti primerjamo lahko na več načinov. Vsak izmed njih po svoje osvetli dinamiko pojava. Z metodami, ki so specifične za analizo časovnih vrst analiziramo zakonitosti dinamike ene same vrste, s korelacijsko analizo pa zakonitosti odvisnosti v dinamiki večih pojavov, ki so med seboj v zvezi.

Primerljivost podatkov v časovni vrsti

11.20 Kljub temu, da so členi v isti časovni vrsti istovrstne količine, dostikrat med seboj niso neposredno primerljivi kot bi se za 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.

11.21 Ker so spremembe pojava, ki ga časovna vrsta prikazuje, bistveno odvisne od časa, je zelo koristno, če so časovni razmiki med posameznimi členi enaki. Tako dosežemo, da je pri enakem delovanju dejavnikov, 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, so v določenih primerih 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 dejavniki, ki vplivajo na število rojstev, stalni, bi bilo število rojstev od meseca do meseca različno tudi za več kot 10% zaradi vzroka, ki nima nobene zveze z rodnostjo. Te razlike so še večje pri pojavih, ki so, kakor na primer proizvodnja, odvisni ne od koledarskih, temveč od delovnih dni, ker so med meseci zaradi različnega števila nedelj in praznikov še večje variacije v številu delovnih dni. Koledar vpliva tudi na pojave, ki imajo tedensko periodičnost. Ker je na primer število prevoženih potnikov, prometnih nesreč ali nočnin v tujskem prometu največje okrog nedelje, je številčnost teh pojavov v mesecu bistveno odvisno od števila nedelj in praznikov v mesecu. Razlike med posameznimi meseci zaradi tega vzroka so tudi do 50%.

11.22 Za trenutne časovne vrste si prizadevamo, da so časovni razmiki med posameznimi členi enaki. Kakor smo že navedli, so popisi prebivalstva na enaka razdobja - vsakih pet ali deset let na isti datum, popisi živine vsako leto na isti datum. Le tako iz časovne vrste dobimo hiter pregled o dinamiki enega in drugega pojava.

Vendar enakost razmikov med posameznimi členi še ni zadosten pogoj za primerljivost med členi. Vzemimo, da popisi živine ne bi bili na letne, temveč na primer sedemmesečne razmike. Popis živine bi bil na primer januarja, avgusta, marca naslednjega leta, oktobra naslednjega leta itd. Kljub temu, da so razmiki enaki, spremembe med posameznimi popisi ne bi bile primerljive, ker bi bili kritični momenti v različnih delih leta. To pa bistveno vpliva na stalež živine. Da se temu ognemo, izvajamo popise v razmikih, ki so mnogokratniki periode – v našem primeru leta.

11.23 Čeprav so razmiki med členi pri mesečnih ali četrtletnih trenutnih časovnih vrstah različni, je učinek razlik na časovno vrsto tako malenkosten, da podatkov ne popravljamo.

Pri razmičnih časovnih vrstah pa je ta učinek znaten in moramo podatke pred analizo preračunati na enake razmike. To dosežemo tako, da za osnovne podatke, ki se nanašajo na dejanska mesečna razdobja, izračunamo namesto mesečnih vrednosti poprečja na en dan, enotin mesec s 30 koledarskimi ali 25 delovnimi dnevi. Tako odstranimo učinek različne dolžine meseca. Ali preračunamo podatke na 25 ali 30 dni, je odvisno od pojava, ki ga analiziramo. Rojstva preračunamo na koledarski mesec 30 dni, enako proizvodnjo v panogah, ki imajo zvezen delovni proces; na mesec 25 dni pa preračunavamo proizvodnjo v panogah in podjetjih, v katerih je delovni proces prekinjen. Tako se najbolj prilagodimo dejanskemu številu dni v posameznih mesecih.

Popravni faktorji za posamezne vrste poprav so nakazani v tabeli 11.7.

Tabela 11.7 Popravni faktorji za mesečne časovne vrste.

Število koledarskih dni v mesecu	Popravni faktor	Število delovnih dni v mesecu	Popravni faktor
28	30/28 = 1,0714	22	25/22 = 1,1364
29	30/29 = 1,0345	23	25/23 = 1,0870
30	30/30 = 1,0000	24	25/24 = 1,0417
31	30/31 = 0,9677	25	25/25 = 1,0000
		26	25/26 = 0,9615
		27	25/27 = 0,9259

Podatke mesečnih časovnih vrst preračunamo na enotino dolžino meseca tako, da jih pomnožimo z ustreznimi popravnimi faktorji. Vsote popravljenih podatkov so brez smisla, ker so popravljeni podatki povprečje, čeprav so osnovni podatki razmičnega značaja.

11.8 Reduciranje števila rojstev v SRS v letu 1969 na enotine mesece

(Vir: MSP)

Mesec	Število rojstev Y	Število dni	Popravni faktor F	Popravljeno število rojstev $VF = Y_f$
J	2143	31	0,9677	2074
F	2275	28	1,0714	2437
M	2546	31	0,9677	2464
A	2480	30	1,0000	2480
M	2454	31	0,9677	2375
J	2190	30	1,0000	2190
J	2399	31	0,9677	2322
A	2248	31	0,9677	2175
S	2283	30	1,0000	2283
O	2388	31	0,9677	2311
N	1997	30	1,0000	1997
D	2775	31	0,9677	2685

11.24 Na velikost pojavov dostikrat vplivajo **administrativni ukrepi**, ki z vsebino proučevanega pojava nimajo neposredne zveze. Eden izmed običajnih vzrokov so upravno-teritorialne spremembe, s katerimi se spremeni geografska opredelitev pojava. Taka sprememba je lahko sprememba državnih meja, ki je posledica vojne, ali sprememba upravno-teritorialnij mej znotraj države. Različna geografska opredelitev onemogoča primerljivost podatkov v časovni vrsti. V tem primeru je potrebno podatke časovne vrste za nazaj preračunati na novo območje, kar je običajno težaven in zamuden posel. Včasih si pomagamo tako, da za čas, ko je nastopila sprememba, navedemo podatek po starem in novem stanju. Tako vidimo, kolika je ob prehodu sprememba zaradi spremembe v teritoriju.

Če časovno vrsto za nazaj ne popravimo, moramo proučiti vsak del posebej, nikakor pa ne smemo proučevati vrsto kot enotno.

Isti problem so tudi spremembe zaradi drugih administrativnih ukrepov, na primer ko je šlo za prehajanje podjetij iz ene pristojnosti v drugo itd. Tudi v teh primerih ravnamo s spremembami enako kakor pri teritorialnih spremembah.

11.25 Ker se število prebivalstva stalno spreminja, povečana proizvodnja še ne pomeni nujno povečane produktivnosti dela, povečana poraba še ne povečanje življenjske ravni itd.

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

Časovno primerljivost podatkov včasih motijo tudi spremembe v sestavi. Sprememba starostne sestave prebivalstva, ki more nastopiti v daljšem časovnem razdobju, močno vpliva na umrljivost. Primerjava koeficientov mortalitete za daljše razdobje zato lahko da napačno sliko v spremembah v umrljivosti prebivalstva. Spremenjena sestava proizvodnje tako v podjetju kot v večjih gospodarskih združbah ali v celotnem narodnem gospodarstvu odločilno vpliva na sumarne koeficiente o produktivnosti dela, o poprečnih osebnih dohodkih itd. Vpliv spremenjene sestave na določene pojave v časovni vrsti odpravimo s standardizacijo podatkov, to je s koeficienti pri stalni sestavi. O standardizaciji koeficientov je bilo več govora pri koeficientih.

11.26 Proizvodnjo, promet v zunanji ali notranji trgovini in druge pojave izražamo v vrednosti. Časovno primerjavo vrednostnih podatkov motijo spremembe v cenah, ker se vrednost proizvodnje, prometa itd. spreminja zaradi sprememb v obsegu in sprememb cen. Vpliv, ki ga povzročajo spremembe cen, odstranimo tako, da izražamo proizvodnjo, promet itd. v stalnih cenah. Vendar je ta način izredno zamuden in težko izvedljiv, ker bi ga morali upoštevati že v osnovni evidenci. Vpliv cen popravimo tudi z indeksi cen tako, da podatke po tekočih cenah delimo z ustreznim indeksom cen. Ta način je operativno veliko lažji, je pa kljub temu, da je le ocena, zadovoljiv. Le paziti je treba, da je indeks, s katerim popravljamo, mo podatke, ustrezen po vsebini in značaju.

Enostavni pokazovalci dinamike

11.27 Enostavno analiziramo dinamiko pojava s primerjavo členov v časovni vrsti.

Najlaže primerjamo člene časovne vrste z indeksi. Zato z njimi izvedemo več elementarnih pokazovalcev dinamike.

Raven Y_k je osnovna časovna vrsta. Že iz sprememb v ravni sklepamo na smer in jakost dinamike pojava.

Če raven časovne vrste izrazimo v primerjavi z nekim stalnim značilnim členom v časovni vrsti, dobimo vrsto indeksov s stalno osnovo.

$$I_{k/o} = 100\, Y_k / Y_o \tag{11.8}$$

Absolutna razlika

$$D_k = Y_k - Y_{k-1} \tag{11.9}$$

pokaže v absolutnih vrednostih spremembo pojava od člena do člena.

Temp rasti ali stopnja rasti

$$T_k = 100 \frac{Y_k - Y_{k-1}}{Y_{k-1}} = 100 \frac{D_k}{Y_{k-1}} \tag{11.10}$$

pokaže relativno razliko od člena do člena.

Koeficient dinamike

$$K_k = Y_k / Y_{k-1} \tag{11.11}$$

pokaže v obliki koeficienta relativne spremembe od člena do člena.

Enak značaj ima verižni indeks

$$I_k = 100 \cdot Y_k / Y_{k-1} \tag{11.12}$$

le, da relativno spremembo nakaže v indeksu.

Ker so vsi ti pokazovalci izpeljani iz osnovne časovne vrste, so med njimi enostavne zveze:

$$T_k = 100 \,.\, D_k/Y_{k-1} = 100\,(K_k - 1) = I_k - 100 \qquad (11.13)$$
$$I_k = 100 \,.\, K_k$$

V tabeli 21.11 je prikazano, kolikšni so posamezni pokazovalci dinamike, če pojav raste, zastane ali pada.

Tabela 11.9. Vrednosti posameznih pokazovalcev dinamike pri različnem gibanju pojava

Pokazovalec dinamike	P o j a v		
	raste	zastane	pada
Raven Y_k	$Y_{k+1} > Y_k$	$Y_{k+1} = Y_k$	$Y_{k+1} < Y_k$
Indeks s stalno osnovo	$I_{k+1/o} > I_{k/o}$	$I_{k+1/o} = I_{k/o}$	$I_{k+1/o} < I_{k/o}$
Absolutna razlika	$D_k > 0$	$D_k = 0$	$D_k < 0$
Temp rasti	$T_k > 0$	$T_k = 0$	$T_k < 0$
Koeficient dinamike	$K_k > 1$	$K_k = 1$	$K_k < 1$
Verižni indeks	$I_k > 100$	$I_k = 100$	$I_k < 100$

11.33 Za svetovno proizvodnjo boksita imamo v tabeli 21.12 izračunane vse vrste elementarnih pokazovalcev dinamike.

Tabela 11.10 Pokazovalci dinamike za svetovno proizvodnjo boksita (v tisoč tonah ; vir: SG 58)

Leto	Y_k	$I_{k/o}$	D_k	T_k	K_k	I_k
1950	7.700	100				
1951	10.300	134	+ 2600	+ 33,8	1,34	134
1952	11.900	155	+ 1600	+ 15,5	1,16	116
1953	13.000	169	+ 1100	+ 9,2	1,09	109
1954	14.850	193	+ 1850	+ 14,2	1,14	114
1955	15.350	199	+ 500	+ 3,4	1,03	103
1956	16.900	219	+ 1550	+ 10,1	1,10	110

11.28 Nakazani pokazovalci dinamike pa niso edini pokazovalci, s katerimi prikazujemo dinamiko pojava. Pogosto prikažemo dinamiko tudi z drugačnimi primerjavami. Izmed teh naj omenimo samo mesečne časovne vrste za pojave, ki so sezonskega značaja. Če je pojav sezonski, primerjava tekočega meseca s predhodnim mesecem nima logične povezave in smisla. Bolj smiselno je, da primerjamo podatek tekočega meseca s podatkom iz predhodnega leta za isti mesec. Ta dva podatka sta si sorodnejša kakor zaporedna meseca. Ti indeksi imajo premično osnovo, vendar niso verižni, ker osnova indeksov ni predhodni člen.

Enako računamo tudi indekse iz kumulativnih vrednosti dveh zaporednih let. Ta primerjava je logična, ker obe primerjani vrednosti obsegata isti del sezone v dveh zaporednih letih.

V tabeli 11.11 je prikazana standardna primerjava z indeksi iz tekočih podatkov Indeksa za julij 1971. Za proizvodnjo vseh vrst premoga so izračunani štiri vrste indeksi:

$\frac{\text{I-VII } 71}{\text{I-VII } 70}$ = indeks iz kumulative proizvodnje do julija med letoma 1971 in 1970

$\frac{\text{VII } 71}{\varnothing \quad 70}$ = indeks proizvodnje v tekočem mesecu (julij 71) s poprečjem preteklega leta;

$$\frac{VII\ 71}{VII\ 70} = \text{indeks tekočega meseca z istim mesecem preteklega leta (VII 70) in}$$

$$\frac{VII\ 71}{VI\ 71} = \text{verižni indeks.}$$

Tabela 11.11 Proizvodnja premoga v SFRJ v tisoč tonah

(Vir: Indeks VIII, 1971)

Vrsta premoga	Ø 70	I-VII 70	I-VII 71	VII 70	VI 71	VII 71	$\frac{I-VII\ 71}{I-VII\ 70}$	$\frac{VII\ 71}{Ø\ 70}$	$\frac{VII\ 71}{VII\ 70}$	$\frac{VII\ 71}{VI\ 71}$
Premog skupno	2369	15329	17539	2211	2247	2498	114	105	113	111
Črni	54	356	409	51	55	57	115	106	112	104
Rjavi	749	5147	5348	666	783	720	104	96	108	92
Lignit	1566	9826	11836	1494	1609	1721	120	110	115	107
Pogojni 4000 kal	1769	11578	13061	1638	1831	1839	113	104	112	100
Koks	109	753	757	114	107	109	101	100	96	102

Sestavine dinamike v časovnih vrstah

11.29 Časovna vrsta je številčen izraz časovnega delovanja vseh dejavnikov, ki vplivajo na pojav, ki ga prikazuje. Teh dejavnikov je veliko in se njihova jakost in učinek časovno spreminjata. Zato se členi časovne vrste spreminjajo; pravimo, da se pojav giblje.

Nemogoče je iz časovne vrste izluščiti, kolikšna je sprememba zaradi vsakega posameznega dejavnika posebej. Iz časovne vrste pa je moč razbrati skupen učinek dejavnikov, ki imajo soroden vpliv na pojav, ki ga proučujemo. Na časovni vrsti po tem vidiku opazujemo naslednje vrste sprememb:

a) **trend** – T, ki podaja osnovno smer razvoja,

b) **ciklične spremembe** – C, ki izvirajo iz dolgoročnih vzrokov,

c) **periodične spremembe** – P, ki izvirajo iz vzrokov, ki se ponavljajo na stalno razdobje – periodo,

d) iregularne spremembe - I, ki so rezultat enkratnih epizodičnih dogodkov ali rezultat stalnih slučajnih vzrokov.

Časovna vrsta ni vedno rezultat delovanja vseh sestavin. Ta ali ona sestavina ne nastopi bodisi zaradi tipa vrste ali ker ni vsebinsko pogojena. Časovna vrsta letnih podatkov ne kaže sezonskega gibanja, čeprav je pojav mogoče sezonski, marsikateri pojav nima cikličnih nihanj ali epizodičnih sprememb zaradi narave pojava. Vsak pojav pa ima neko osnovno smer razvoja. Le redki so primeri nespremenljivih pojavov. Na vsaki časovni vrsti opazimo slučajna nihanja, ki so izraz manjših vzrokov, ki nastopajo v eni ali drugi obliki.

11.30 Trend. Vsak ekonomski pojav ima časovno neko osnovno smer razvoja. Ta pa je opazna le v daljših časovnih razdobjih. Osnovno smer razvoja imenujemo trend. Če trend nakazuje razvoj za dolgo razdobje, ga včasih imenujemo sekularni trend. Trend je rezultat dejavnikov, ki pogojujejo stalen razvoj pojava. Tako je trend povečanja proizvodnje električne energije izraz tehničnega napredka, trend zniževanja umrljivosti rezultat razvoja zdravstvene službe in medicine nasploh itd. Trend proizvodnje električne energije kaže stalno naraščanje, trend gibanja umrljivosti pa stalno padanje. Trend pa more v določenem razdobju tudi spremeniti smer. Tako more proizvodnja nekega izdelka določeno razdobje naraščati, v nadaljnjem razvoju pa zaradi nadomestitve z drugim izdelkom, ki je bolj kakovosten, pada. Enako more proizvodnja v rudniku za neko razdobje naraščati, v nadaljnjem razdobju pa se zaradi osiromašenja rudnika zmanjšuje. Navedli smo že, da je konstantnih pojavov malo. Eden izmed njih je na primer spolni indeks - razmerje med številom rojstev dečkov in deklic. To razmerje ima časovno precej stalno vrednost brez težnje naraščanja ali padanja. Odkloni od konstante so le slučajni in ne kažejo težnje k časovni spremembi tega razmerja. Ta primer je prikazan v drugem poglavju v tabeli 2.1.

Zaradi drugih vplivov se dejanske vrednosti od trenda odklanjajo navzgor in navzdol, vendar v večini primerov že iz slike za časovno vrsto razberemo osnovno smer razvoja.

11.31 P e r i o d i č n a   n i h a n j a . Za veliko pojavov je značilen periodičen, zlasti sezonski značaj. Sezonski značaj marsikaterih pojavov povzroče predvsem klimatski vplivi. Tako je s klimatskimi vplivi povezana gradbena dejavnost, turizem in gostinstvo, kmetijstvo in vse druge dejavnosti, ki so odvisne od klimatskih vplivov ali posredno povezane z eno izmed navedenih. Prehrambena industrija je vezana na poljedelstvo, promet na turizem itd. Sezonski značaj pa ne izvira nujno iz klimatskih vplivov, marveč more biti vzrok tudi drugje. Tako vplivajo ustaljene navade ali prireditve na sezonski značaj pojavov. Potniški promet, poštni promet, trgovinski promet itd. je odvisen od praznikov, ki so v vsakem letu ob istem času (npr. Novo leto). Enako izzove sezonski značaj določenega pojava prireditev, ki se letno ponavlja na isti datum; tak zgled so npr. gospodarski velesejmi. Pri najrazličnejših pojavih pa opazimo periodična nihanja, za katere je perioda krajša kot leto. Tako opazujemo npr. v blagovnem prometu mesečna nihanja, v prometu tedenska ali dnevna nihanja ipd.

11.32 C i k l i č n a   n i h a n j a . V časovnih vrstah za daljša razdobja opazimo nihanja okrog trenda. Ta nihanja so več ali manj regularna, ni pa niti njihova dolžina niti oblika stalna kakor pri periodičnih nihanjih. Ta nihanja pa kljub temu, da niso popolnoma regularna, niso slučajna in so odvisna od dogajanj v preteklosti. Imenujemo jih ciklična gibanja. Ciklična gibanja so tipična za ekonomske pojave in jih zasledimo na nizu pojavov. Jakost cikličnih nihanj je za različne pojave različna. Odvisna je od tega, ali gre za pojav, ki je bolj ali manj občutljiv za spremembe v gospodarskem življenju. Zaradi povezanosti gospodarstva je med cikli za posamezne gospodarske panoge ali pojave zveza, ker se tako prosperiteta kot depresija prenaša od panoge na panogo.

11.33 I r e g u l a r n e   v a r i a c i j e . Razen navedenih treh vrst gibanj, trenda, periodičnih nihanj in cikličnih sprememb opazujemo na časovni vrsti še iregularne variacije, ki so rezultat enkratnih ali pa slučajnih vzrokov. Zaradi prekinitve toka nastane kratkoročen zastoj v prometu električne cestne železnice. Potres vpliva na niz dejavnosti. Zaradi prometne nesreče je zastoj v železniškem prometu, zaradi poplave je uničena letina, epidemija vpliva na morbiditeto in mortaliteto itd. Navedeni dogodki so e n k r a t n i v p l i v i , ki imajo za rezultat, da se za krajše razzobje pojav odkloni od rednega poteka.

Za razliko od enkratnih - epizodičnih vplivov, ki nastopijo nepričakovano, vendar za krajši čas, nastopajo stalno in v vseh pojavih s l u č a j n e v a r i a c i j e , ki so rezultat slučajnih vplivov. Slučajne variacije so manjše spremembe, ki so rezultat nepomembnejših vplivov. Slučajne variacije le redko obravnavamo posamično, ker za to bodisi ni možnosti ali potrebe.

V l o g a p o p r e č i j p r i p r o u č i t v i č a s o v n i h v r s t

11.34 Enostavno orodje, ki pomaga proučiti časovne vrste, so poprečja. Ob določenih pogojih moremo s poprečji neke sestavine v časovni vrsti odstraniti, druge pa ohraniti. Poprečja iz časovnih vrst vplivajo v eni ali drugi obliki na vse sestavine.

Če je t r e n d v razmiku, iz katerega računamo poprečje, linearen, poprečje, ki ga centriramo na sredino razmika, leži na trendu. Če je trend na odseku, iz katerega računamo poprečje, krivuljčen, je poprečje, centriramo na sredino razmika, nad trendom, če je trend konkaven in pod trendom, če je trend konveksen. Razlike so tem večje, čimbolj je trend ukrivljen. Poprečje, centrirano na obratišče, pa leži na trendu. Iz tega sklepamo, da se časovna vrsta drsečih sredin, če jo izračunamo iz trenda, na odseku, na katerem je trend linearen, sklada s trendom, če pa je trend krivuljčen, dobimo črto, ki je bolj izravnana kot trend. Ker je krivina trenda na daljših odsekih večja, je izravnavanje trenda tem večje, čim širši so razmiki, za katere izračunavamo poprečja.

Izravnavanje ni tako izdatno, če člene, iz katerih računamo poprečja, tehtamo tako, da imajo robni členi manjšo, osrednji pa večjo težo. Običajno tehtamo člene z binomskimi koeficienti (1,2,1) za poprečja iz treh členov, (1,3,3,1) za poprečja iz štirih členov, (1,4,6,4,1) za poprečja iz petih členov. Izdelani so še različni drugi načini tehtanja.

Če je zveza med sestavinami aditivna, se učinek periodičnih oziroma sezonskih vplivov v poprečju uniči, če je dolžina razdobja, za katerega računamo poprečje, enako periodi ali mnogokratniku periode.

V vrsti drsečih sredin se uniči ciklična sestavina, če so poprečja izračunana iz razdobja celotnega cikla.

Učinki slučajnih vplivov se v poprečju manjšajo, čim daljše je razdobje, iz katerega računamo poprečja. Če ta pravila izkoriščamo povezano, poprečje veliko pripomore k proučitvi časovnih vrst.

TREND

11.35 Trend proučujemo iz dveh razlogov. Samega zase proučujemo, da spoznamo smer razvoja pojava, da ga primerjamo s trendi za pojave, ki so med seboj odvisni ali pa da proučujemo vpliv trenda na druge sestavine (periodične in ciklične). Trend pa dostikrat iščemo zato, da iz odklonov stvarne časovne vrste od trenda proučujemo druge sestavine - sezonska, ciklična ali iregularna nihanja.

Metod za določanje trenda je več. Vsaka izmed njih ima za osnovo različne predpostavke in postopke. Zato je pomembno in odgovorno izbrati v vsakem posebnem primeru primerno obliko in metodo za določanje trenda. Trend določamo s prostoročnim vrisavanjem, z izravnavanjem s sredinami ali analitično s prilagajanjem analitičnih funkcij danim podatkom.

Katero metodo uporabimo za določanje trenda, je odvisno od proučevane časovne vrste in namena analize. Če potrebujemo trend zaradi proučevanja cikličnih ali sezonskih odklonov, je logično, da predpostavljamo, da je trend črta, ki poteka med realnimi vrednostmi tako, da se izenačujejo periodični in ciklični vplivi. To črto določamo običajno mehanično. Če pa potrebujemo trend za napovedovanje, iščemo za trend funkcijsko obliko, iz katere moremo z ekstrapolacijo oceniti potek trenda tudi v prihodnost.

11.36 Najvažnejši problem je izbira pravilnega tipa funkcije, ki naj predstavlja trend. Enoličnega pravila za določevanje funkcije trenda ni. Splošne smernice določajo, da mora biti funkcija, ki jo uporabimo za trend, smiselna, ne preveč zamotana.

Po sliki časovne vrste, poznavanju pojava, ki ga vrsta prikazuje in poznavanju lastnosti funkcij, ki pridejo v poštev, izberemo v vsakem določenem primeru, kateri tip funkcije je vsebinsko in tehnično najprimernejši.

Kot trend uporabljamo ali iregularno prilagojene črte ali analitične krivulje. Najenostavnejša funkcija, s katero pa v mnogih primerih zadovoljivo opišemo smer razvoja, je premica

$$T = a + bx \tag{11.14}$$

Linearen trend, katerega opisuje premica ima dva parametra: a in b, s katerima je določena osnovna smer razvoja.

11.37 Metode za določanje trenda. Omenili smo že, da imamo več metod za določanje trenda. Od njih bomo obravnavali tele:

a) prostoročno metodo,

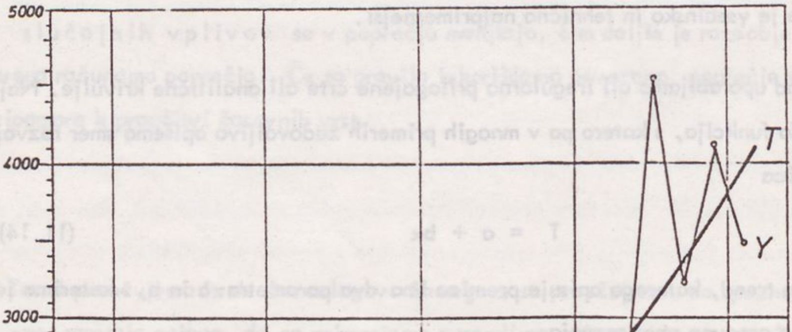
b) metodo drsečih sredin,

c) metodo najmanjših kvadratov za določanje linearnega trenda.

Prostoročna metoda

11.38 Običajno si moremo v grafikonu že iz slike časovne vrste zamisliti potek trenda. Trend v vsakem primeru poteka med realnimi vrednostmi tako, da se osnovna časovna vrsta odklanja od trenda navzgor in navzdol. Zato moremo trend včrtati v grafikonu prostoročno. Ta metoda je subjektivna in zato nima posebne analitične vrednosti. Vendar je njena prednost v enostavnosti in hitrosti. Zato jo uporabljamo predvsem za približek in za osnovo pri izbiri analitične oblike trenda.

11.39 V sliki 11.6 je narisana časovna vrsta proizvodnje cinkovega koncentrata na območju Slovenije v razdobju med obema vojnama od 1919-1941, v njej pa prostoročno včrtan trend.



Slika 11.6 Trend za proizvodnjo cinkovega koncentrata v Sloveniji med obema vojnama, določen prostoročno

Tabela 11.12 Proizvodnja cinkovega koncentrata v Sloveniji med obema vojnama v tonah (Vir: SB SRS 1952, št. 8)

Leto:	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
Proizvodnja	90	350	75	450	600	520	510	639	1143	1076	1313

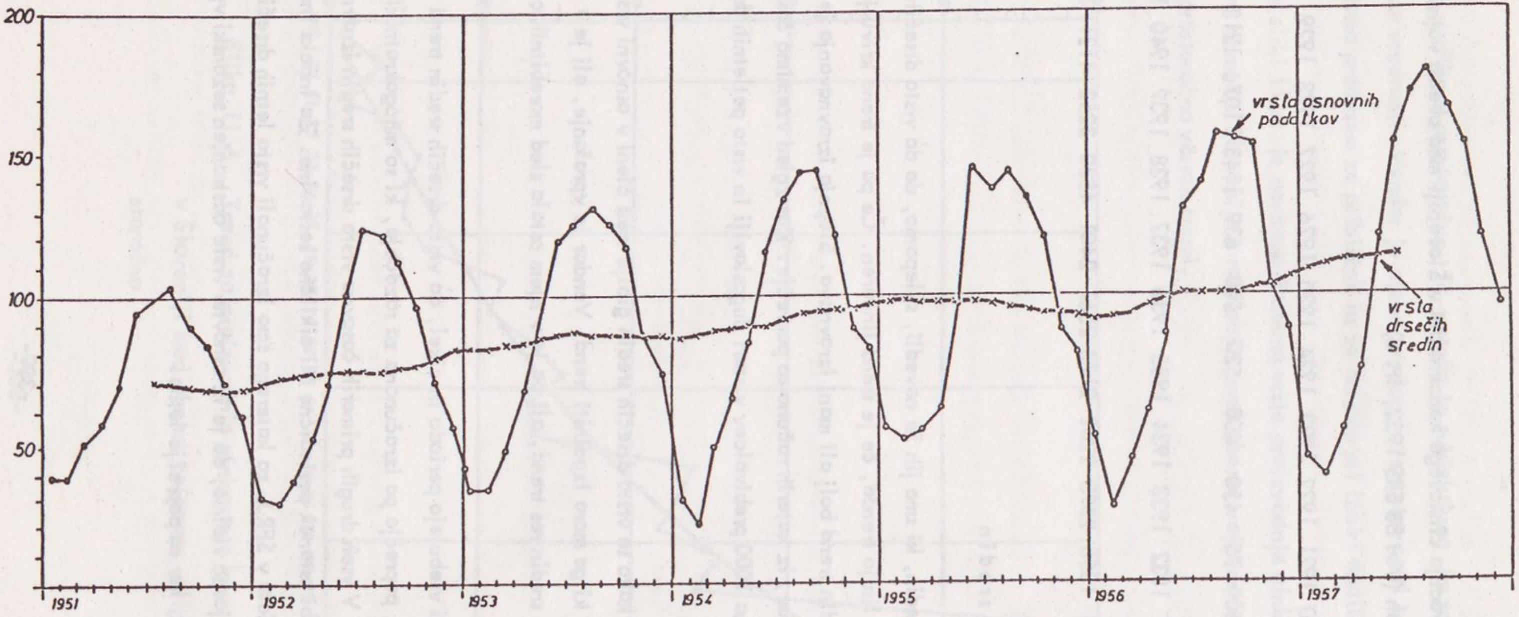
Leto:	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
Proizvodnja	1547	1301	1358	2393	2282	2173	2567	2401	4530	3252	4121	3483

Metoda drsečih sredin

11.40 Iz lastnosti sredin, ki smo jih že navedli, sklepamo, da dà vrsta drsečih sredin približno linijo trenda, če je trend linearen. Če pa je trend krivuljčen, se v vrsti drsečih sredin trend bolj ali manj izravnava. Stopnja izravnavanja je odvisna od števila členov, iz katerih računamo poprečje. Kot zgled vzemimo časovno vrsto števila umrlih na 1000 prebivalcev v stari Jugoslaviji in vrsto petletnih drsečih sredin.

Slika 11.2 pokaže, kako se vrsta drsečih sredin giblje med členi v osnovni vrsti. To pa je osnovni pogoj, ki ga mora izpolniti trend. Vendar je vprašanje, ali je v tem primeru vrsta drsečih sredin res trend, ali pa je v njem ostala sled morebitnih ciklov.

V časovnih vrstah, ki vsebujejo periodo in cikel, dà vrsta drsečih sredin trend le, če je trend linearen, poprečja pa izračunana za razdobja, ki so mnogokratniki periode oziroma cikla. V vseh drugih primerih časovna vrsta drsečih sredin izravnava trend ali pa ima v sebi ostanek periodične ali ciklične sestavine. Za indeks industrije gradbenega materiala v SFRJ, za katerega smo izračunali vrsto letnih drsečih sredin, na sliki 11.6 jasno vidimo, da je iz osnovne vrste odstranjen sezonski vpliv in slučajne variacije, ker so poprečja letna.



Slika 11.6. Indeksi industrije gradbenega materiala v SFRJ z vrisano časovno vrsto desečih sredin.

Določitev linearnega trenda po metodi najmanjših kvadratov

11.41 Transformacija časa. Členi v časovni vrsti se nanašajo na zaporedne enake časovne razmike ali trenutke. Za letne časovne vrste so členi označeni s štirimestnim številom. Če gre za mesečno časovno vrsto. pa je označevanje členov še bolj zamotano, ker je treba vsak člen označiti z letom in mesecem. Zato pri analizi časovnih vrst pa tudi v končnih rezultatih čas transformiramo na enostavnejše merjenje. V končnih rezultatih pogosto čas v časovnih vrstah ali funkcijah, ki nakazujejo zakonitosti dinamike, štejemo tako, da je izhodišče v prvem členu, razmiki med členi v časovni vrsti pa so enotina razdobja. V praktičnih primerih časovni vrsti z N členi priredimo naslednje čase: $x = 0, 1, 2, 3, \ldots (N-1)$.

Izkaže pa se, da tudi to štetje časa, kljub temu, da je v primerjavi s koledarskim označevanjem že zelo poenostavljeno, pri izračunanju pokazovalcev dinamike ni vselej najprikladnejše. Pri izračunanju ima veliko prednosti štetje, za katerega je izhodišče v sredini proučevane časovne vrste, razmik med posameznimi členi pa najmanjše celo število. To štetje časa, imenujemo ga tehnični čas, zaznamujemo pa s t, je za časovne vrste z lihim številom členov $N = 2r+1$ $t: -r = \frac{N-1}{2}$, $-r+1 \ldots\ldots -2 \; -1 \; 0 \; 1 \; 2 \; r-1, \; r = \frac{N-1}{2}$.

Če pa je število členov v časovni vrsti sodo, $N = 2r$, pa gornjim pogojem ustreza tale vrsta za tehnični čas t :

$$t: \; -N+1 \quad -N+3 \quad \ldots\ldots -3 \quad -1 \quad 1 \quad 3 \ldots N-3 \quad N-1$$

Med tehničnim časom t in časom z izhodiščem v prvem členu x, je glede na to, ali je število členov N liho ali sodo tale zveza:

$$\begin{aligned} N &= 2r+1 \; ; \quad t = x - \frac{N-1}{2} \\ N &= 2r \quad\;\; ; \quad t = 2x - (N-1) \end{aligned} \qquad (11.15)$$

11.42. L i n e a r e n   t r e n d . Po metodi najmanjših kvadratov vzamemo kot linearen trend tisto funkcijo, ki zadošča pogoju, da je vsota kvadratov odklonov stvarnih vrednosti od funkcije trenda najmanjša

$$\sum_{i=1}^{N} \left[Y_i - T(t_i) \right]^2 = \min$$

Za primer linearnega trenda $T = b_o + b_1 t$ je ta pogoj

$$\sum_{i=1}^{N} (Y_i - b_o - b_1 t_i)^2 = \min$$

Iz tega pogoja dobimo, da morata parametra b_o in b_1 zadoščati enačbama

$$\sum Y = b_o N \; ; \sum Yt = b_1 \sum t^2 \tag{11.16}$$

Parametra b_o in b_1 sta torej:

$$b_o = \frac{1}{N} \sum Y \; ; \quad b_1 = \frac{\sum Yt}{\sum t^2} \tag{11.17}$$

$\sum t^2$ izračunamo za poljuben N po obrazcih

$$\sum_1^N t^2 = \frac{1}{2} \binom{N+1}{3} \quad \text{če je } N = 2i + 1$$

$$\tag{11.18}$$

$$\sum_1^N t^2 = 2 . \binom{N+1}{3} \quad \text{če je } N = 2i$$

Za $2 \leqslant N \leqslant 30$ so po obrazcih 11.21 izračunane vrednosti $\sum t^2$ v tabeli 11.13.

Tabela 11.13 $\sum t^2$ za $2 \leqslant N \leqslant 30$

N		2	3	4	5	6	7	8	9	10
$\sum t^2$		2	2	20	10	70	28	168	60	330
N	11	12	13	14	15	16	17	18	19	20
$\sum t^2$	110	572	182	910	240	1360	408	1938	570	2660
N	21	22	23	24	25	26	27	28	29	30
$\sum t^2$	770	3542	1012	4600	1300	5850	1638	7308	2030	8990

Da bi dobili parametra b_o in b_1, je treba iz časovne vrste izračunati le $\sum Y$ in $\sum Yt$ ter uporabiti obrazca 11.17 in tabelo 11.13.

Linearen trend, transformiran na prvi člen kot izhodišče, izračunamo po obrazcih

$$T = \left[b_o - \frac{b_1(N-1)}{2}\right] + b_1 x \quad \text{za } N = 2i + 1$$

$$T = \left[b_o - b_1(N-1)\right] + 2b_1 x \quad \text{za } N = 2i \qquad (11:19)$$

11.42 Če za naravni prirastek na tisoč prebivalcev za SFRJ po nakazani metodi izračunamo linearen trend, dobimo:

Leto	Y	t	Yt	T
1955	15,5	-9	-139,5	15,18
1956	14,8	-7	-103,6	14,80
1957	13,2	-5	- 66,0	14,43
1958	14,7	-3	- 44,1	14,05
1959	13,5	-1	- 13,5	13,68
1960	13,6	1	13,6	13,30
1961	13,7	3	41,1	12,93
1962	12,0	5	60,0	12,55
1963	12,5	7	87,5	12,18
1964	11,4	9	102,6	11,80
	$\sum Y = 134,9$		$\sum Yt = -61,9$	$134,90 = \sum T$
			$\sum t^2 = 330$	

$$b_o = \frac{1}{N}\sum Y = \frac{134,9}{10} = 13,49\,; \quad b_1 = \frac{\sum Yt}{\sum t^2} = \frac{-61,9}{330} = -0,1876$$

$T = b_o + b_1 t = 13,49 - 0,1876\, t$ ali transformiran

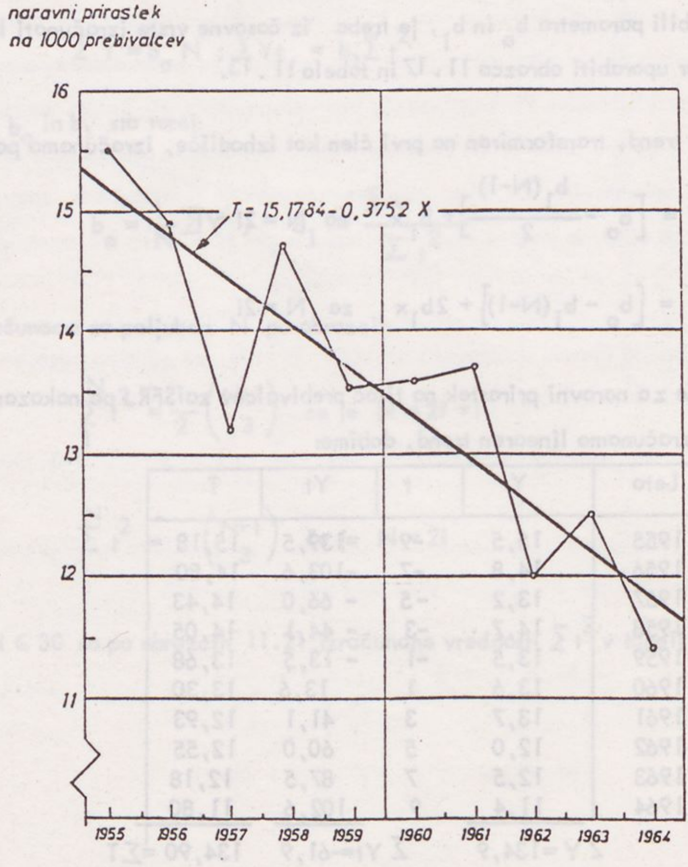
$$T = \left[b_o - b_1\,(N-1)\right] + 2\,b_1\,x = 15,1784 - 0,3752\,x$$

Ker je $\quad T = T_k - T_{k-1} = 2\,b_1$, dobimo vrsto trenda, če začetnemu trendu

$T_o = 15,1784$ kumulativno prištevano $2\;b_1 = -0,3752 = \Delta T$

$T_1 = T_o + \Delta T = 15,1784 - 0,3752 = 14,8032$

$T_2 = T_1 + \Delta T = 14,8032 - 0,3752 = 14,4280$



Slika 11.7. Trend za naravni prirastek prebivalstva za SFRJ v razdobju 1955-1964.

DVANAJSTO POGLAVJE

PROUČEVANJE ODVISNOSTI MED MNOŽIČNIMI POJAVI

12.1 V prejšnjih poglavjih smo proučevali posamezne znake samostojno, brez zveze z drugimi znaki. Te metode analize statističnih podatkov zelo podrobno obravnavajo značilnosti za posamezne znake. Vendar obsegajo samo del analize statističnih podatkov. Ne zajemajo namreč enega izmed najvažnejših problemov v analizi socialno-ekonomskih pojavov in množičnih pojavov na sploh, to je njihovo medsebojno odvisnost in povezanost. Pri proučevanju socialno-ekonomskih pojavov namreč često naletimo na problem povezanosti in odvisnosti med pojavi. Obseg proizvodnje kmetijskega obrata je odvisen od velikosti obrata, strukture osnovnih sredstev, delovne sile, klimatskih faktorjev itd. Vrednost proizvodnje industrijskega obrata je odvisna od števila delavcev, mehanizacije, produktivnosti dela, vrste proizvodnje itd. Cena je odvisna od količine blaga, ki je na trgu, osebni dohodek od kvalifikacije, službene dobe itd. V medsebojni povezavi sta tudi nepismenost moških in nepismenost žensk po občinah, starost ženina in neveste, itd. Podobnih primerov moremo našteti veliko. Vendar niso vsi pojavi med seboj povezani, čeprav bi po vsebini mogli odvisnost pričakovati. Imamo pa tudi primere nesmiselne povezave. Nesmiselno je npr. proučevati odvisnost med številom porok in množino padavin po letih, cene kmetijskih pridelkov od obolelosti za tuberkulozo itd.

FUNKCIJSKE ODVISNOSTI

12.2 O funkcijski odvisnosti med x in y v matematičnem smislu govorimo, če je dano neko pravilo zveze med neodvisno spremenljivko x in odvisno spremenljivko y , po katerem določeni vrednosti neodvisne spremenljivke x ustreza ena ali več natančno določenih vrednosti odvisne spremenljivke y. S simbolom pišemo:

$$y = f(x) \tag{12.1}$$

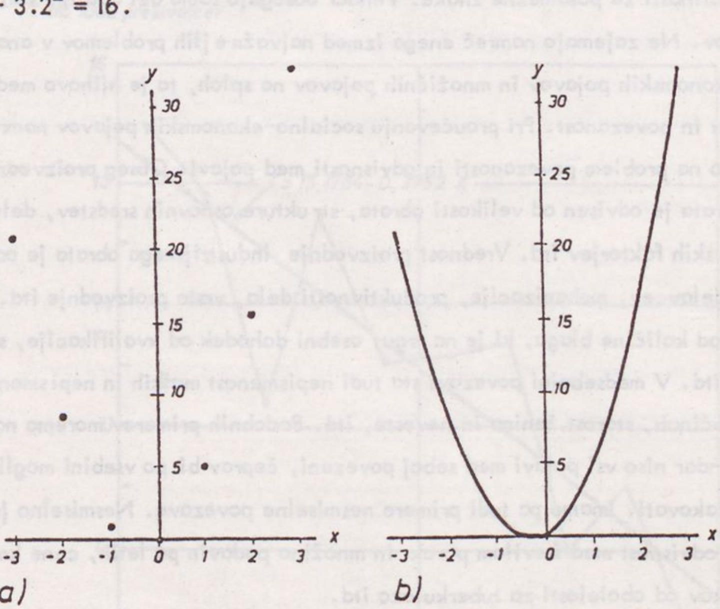
in pravimo: y je funkcija od x.

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

$$y = 2x + 3x^2 \tag{12.2}$$

x = 2 ustrezna vrednost

$y = 2.2 + 3.2^2 = 16$.



Slika 12.1 Slika funkcijske odvisnosti $y = 2x + 3x^2$

Razen z obrazcem moremo funkcijsko odvisnost med x in y nakazati tudi z nizom dvojic ustreznih vrednosti x in y. Tako je v tabeli 12.1 dana funkcijska zveza iz gornje enačbe za osem parov vrednosti x in y.

Tabela 12.1 Pari vrednosti za x in y

x	y
-3	+21
-2	+8
-1	+1
0	+0
+1	+5
+2	+16
+3	+33
+4	+56

S sistemom dvojic pa za gornji zgled ne moremo podati funkcijske zveze med x in y za vse vrednosti x. To moremo prikazati edino z enačbo med x in y.

Niz dvojic vrednosti x in y moremo prikazati tudi grafično v pravokotnem koordinatnem sistemu. Posamezno dvojico vrednosti x, y po znanem načinu prikažemo s točko v pravokotnem koordinatnem sistemu. Kolikor dvojic ustreznih vrednosti x in y imamo, toliko imamo točk v koordinatnem sistemu. Medtem ko moremo niz dvojic vrednosti grafično prikazati s sistemom točk, prikažemo funkcijo, dano z obrazcem, s krivuljo. V določenem razmiku ustreza vsaki vrednosti x funkcijska vrednost y. Slika niza dvojic iz tabele 12.1 je prikazana v sliki 12.1a, slika funkcijske odvisnosti iz obrazca 12.2 pa v sliki 12.1b.

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.

12.3 Če prenesemo pojem funkcijske odvisnosti na množične pojave, bi imela v primeru funkcijske odvisnosti med površino in proizvodnjo kmetijskega gospodarstva vsa gospodarstva z enako površino enako proizvodnjo. Vendar to ni tako. Čeprav sta površina gospodarstva in proizvodnja med seboj odvisni, imata gospodarstvi z enakima površinama le redko enako proizvodnjo. Še več. Čeprav sodimo, da ima večje gospodarstvo večjo proizvodnjo, velja ta odvisnost samo na splošno, v posameznih primerih pa more imeti tudi večje gospodarstvo manjšo proizvodnjo. Do tega pride zato, ker proizvodnja ni odvisna samo od površine, temveč še od mnogo drugih faktorjev. Poizkus, da bi odstranili vse dodatne faktorje in tako prišli do funkcijske povezave med dvema znakoma, se ne bi posrečil. Vedno ostanejo neki faktorji, katerih vpliv ne moremo odstraniti in kontrolirati in jih štejemo med slučajne faktorje. Pri množičnih pojavih moremo torej opazovati le splošno tendenco odvisnosti, v posameznih primerih pa zakonitost zaradi delovanja dodatnih - posamičnih vplivov ni nujno vidna. Zato imenujemo za razliko od funkcijskih odvisnosti te vrste odvisnosti korelacijske odvisnosti.

Proučevanje korelacijskih odvisnosti je različno od proučevanja funkcijskih odvisnosti, čeprav imata obe vrsti proučevanja svoje stične točke.

Prikazovanje korelacijskih odvisnosti

Korelacijske odvisnosti prikazujemo na enake tri načine kakor funkcijske odvisnosti:

a) v tabeli z nizom dvojic vrednosti koreliranih podatkov za vsako enoto populacije, ali v korelacijski tabeli,

b) s točkami v korelacijskem grafikonu,

c) v funkcijski obliki z regresijsko funkcijo ali črto.

12.4 Niz dvojic podatkov. Z nizom dvojic vrednosti koreliranih podatkov na splošno prikazujemo osnovne podatke pri proučevanju korelacije med pojavi. Čeprav je ta način nepregleden in iz njega še ne dobimo vtisa o zakonitosti povezave med dvema pojavoma, ga pogosto uporabljamo, ker je osnova za vsa nadaljnja proučevanja.

Ker je vir podatkov o množičnih pojavih populacija, posamezna dvojica koreliranih podatkov velja za posamezne enote proučevane populacije.

V tabeli 12.2 so prikazani osnovni podatki za proučevanje korelacije med odstotkom nepismenega prebivalstva (x) in odstotkom nepismenih žensk (y) v SFRJ v letu 1961 po okrajih.

Tabela 12.2 Odstotki nepismenega prebivalstva (x) in odstotki nepismenih žensk (y) v SFRJ v letu 1961 po okrajih

BiH		Črna Gora		Hrvatska		Makedonija		Slovenija		Srbija	
x	y	x	y	x	y	x	y	x	y	x	y
35,3	50,7	21,7	33,2	10,6	14,0	20,9	28,8	1,9	1,9	9,2	14,5
37,8	53,6			19,1	28,4	30,6	44,4	3,1	3,6	23,2	36,9
32,5	48,3			11,0	15,2	25,0	36,2	1,2	1,2	25,5	40,6
32,7	47,7			13,1	17,5	20,6	28,8	1,9	2,1	33,1	49,4
26,8	39,9			5,9	8,1	25,1	34,1			24,9	38,1
34,6	51,7			17,0	25,2	33,8	48,5			24,4	38,3
				18,1	26,4	19,7	27,7			26,0	41,5
				7,4	10,2					26,7	43,3
				7,6	10,7					25,8	38,5
										9,8	14,1
										12,1	16,8
										11,6	16,6
										8,6	11,6
										12,1	17,2
										41,1	56,0

V tabeli 12.3 so prikazani dohodki in stroški za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957.

Tabela 12.3 Dohodki v tisočih dinarjev (x) in stroški za kulturno in družbeno življenje v dinarjih (y) za 36 delavskih družin v Mariboru v novembru 1957.
(Vir: Anketa ZS SRS o življenju delavcev in nameščencev)

x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7	16,5
y	260	330	490	310	580	560	330	480	590	330	630
x	16,9	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4
y	630	650	1100	760	460	700	540	750	370	630	740
x	28,4	25,8	30,5	29,8	27,5	26,5	26,7	26,7	32,9	37,5	30,8
y	1080	630	990	1210	1100	870	640	1110	1610	1590	1070
x	31,8	38,4	35,9								
y	1600	1710	1810								

V tabeli 12.4 so podatki o odstotkih travniških in pašniških površin za okraje v SRS po stanju v letu 1952.

Tabela 12.4 Odstotki travniških in pašniških površin po okrajih v SRS po stanju leta 1952.
(Vir: Statistični bilten SRS)

Okraj	Odstotek površin pod travniki	Odstotek površin pod pašniki
Celje - mesto	22	8
Celje - okolica	3	28
Črnomelj	9	20
Gorica	5	13
Kočevje	6	15
Kranj	7	7
Krško	16	9
Ljubljana - mesto	28	3

Okraj	Odstotek površin pod	
	travniki	pašniki
Ljubljana - okolica	18	11
Ljutomer	18	6
Maribor - mesto	15	6
Maribor - okolica	15	8
Murska Sobota	21	3
Novo mesto	8	9
Postojna	10	19
Ptuj	18	10
Radovljica	5	16
Sežana	1	37
Slovenj Gradec	7	10
Šoštanj	8	9
Tolmin	5	20
Trbovlje	9	8

12.5 Iz gornjih treh primerov sklepamo še na določeno vsebinsko razliko med korelacijskimi odvisnostmi. V prvih dveh primerih je jasno, da je delež nepismenih odvisen od republike, stroški za kulturne in družbene potrebe pa od dohodkov, ne pa obratno. To so vzročne povezave oziroma odvisnosti, ker je en pojav vzrok, drugi pa posledica ali statistično, en znak faktorialen, drug pa rezultativen. Vzročne povezave pa ne zasledimo v primeru odstotka površin, ker ne moremo reči, da je visok odstotek travniških površin vzrok za nizek odstotek pašniških površin ali obratno. Kljub temu, da med njima ni vzročne odvisnosti, pa sta ta dva podatka med seboj povezana, ker na oba vplivajo isti pogoji, ki imajo v našem primeru za posledico, da je travnikov mnogo, pašnikov pa malo in obratno. Dva taka odločilna skupna faktorja sta npr. nadmorska višina in vrsta tal.

Na podobno nevzročno čisto korelacijsko povezavo naletimo npr. tudi pri proučevanju korelacije med odstotkom nepismenih žensk in moških po okrajih. Čeprav med nepismenostjo moških in žensk ni vzročne povezanosti, sta med seboj v korelaciji, ker na oba deluje isti kompleks faktorjev, od katerih je odvisna nepismenost. To so npr. število šol v okraju, kulturno prosvetna dejavnost v preteklosti itd.

Medtem ko pri vzročnih odvisnostih po pravilu proučujemo samo odvisnost posledice od vzroka, moremo pri čistih korelacijskih odvisnostih proučevati odvisnost znaka x od y in obratno, odvisnost znaka y od x.

12.6 Korelacijska tabela. Že pri majhnem številu enot je prikazovanje posamičnih dvojic koreliranih podatkov v gornji obliki zelo obširno in nepregledno. Zato za večje populacije prikazujemo korelirane podatke v kombinacijski tabeli, ki jo imenujemo korelacijska tabela. Kakor dobimo frekvenčno porazdelitev, če podatke uredimo v razrede po enem znaku, dobimo korelacijsko tabelo, če podatke uredimo v razrede po obeh koreliranih znakih hkrati. Položaj frekvenc v korelacijski tabeli zelo nazorno pokaže smer povezave med koreliranima znakoma.

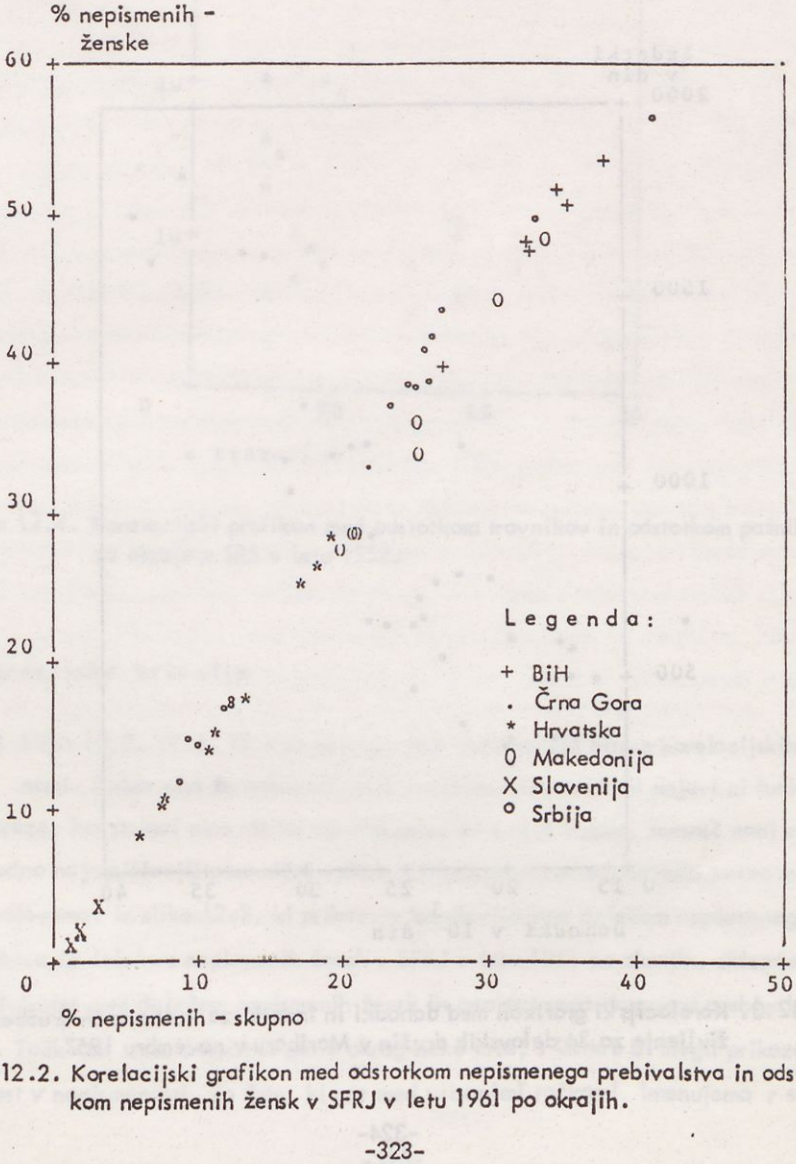
Kot zgled je v tabeli 12.5 prikazana korelacijska tabela med poprečnim neto osebnim dohodkom vseh zaposlenih (x) in poprečnim osebnim dohodkom direktorjev v letu 1963 za 205 izbranih podjetij iz ankete o neto osebnih dohodkih v letu 1963.

Tabela 12.5. Korelacijska tabela med poprečnim mesečnim dohodkom zaposlenih in poprečnim mesečnim dohodkom direktorjev za 205 v vzorec izbranih obrtnih podjetij v letu 1963 v SFRJ (v tisoč starih dinarjih).
Vir: SZS SB 369, Beograd, 1965.

Ø dohodek zaposlenih v tisoč din x / Ø dohodek direktorja v tisoč din y	-10	10-20	20-30	30-40	40-50	50-60	Skupno
20 - 40	3	18	16	3	1	-	41
40 - 60	-	19	66	17	1	-	103
60 - 80	-	-	12	20	10	7	49
80 - 100	–	–	–	4	2	3	9
100 - 120	–	–	–	1	1	1	3
Skupno	3	37	94	45	15	11	205

Podatki v tabeli 12.5 pomenijo število enot, ki ima vrednost ustreznemu kombiniranemu razredu. Tako je npr. bilo 19 podjetij, v katerih je bil poprečen osebni dohodek zaposlene ga od 10-20 tisoč din, poprečen dohodek direktorja pa med 40 in 60 tisoč din. Korelacijska tabela je torej frekvenčna porazdelitev, v kateri je populacija razdeljena v razrede po dveh znakih hkrati. Iz korelacijske tabele je jasno razvidna korelacijska odvisnost med x in y.

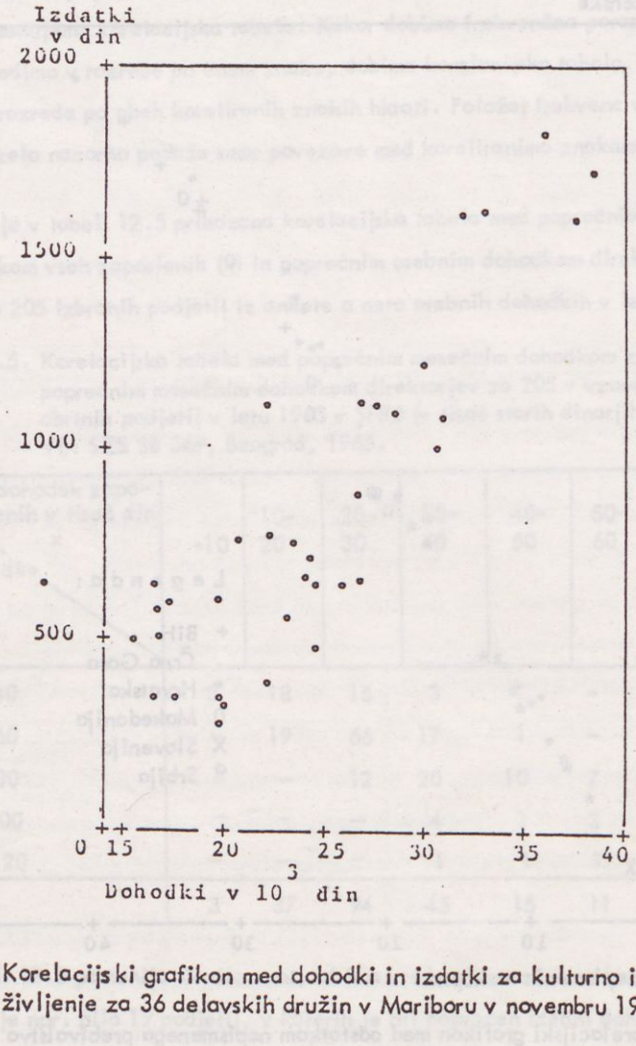
12.7 **Korelacijski grafikon.** Nazorno prikažemo korelacijo med dvema znakoma v korelacijskem grafikonu. V njem je vsaka dvojica vrednosti prikazana



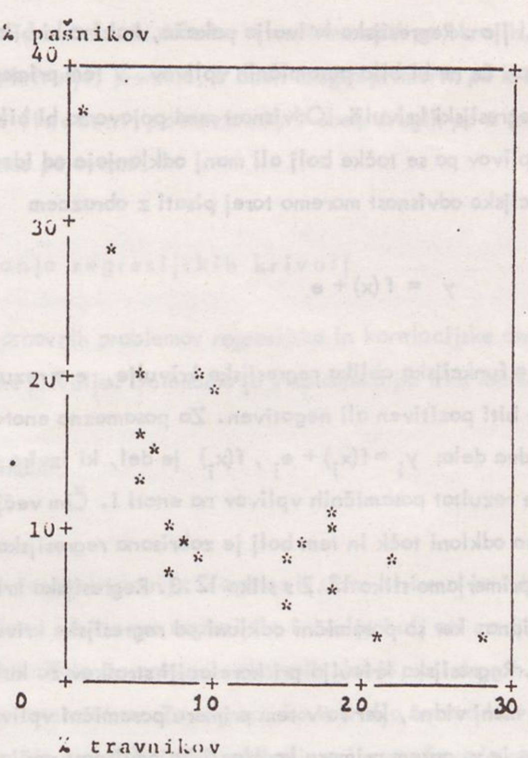
Slika 12.2. Korelacijski grafikon med odstotkom nepismenega prebivalstva in odstotkom nepismenih žensk v SFRJ v letu 1961 po okrajih.

s točko v pravokotnem koordinatnem sistemu, v katerem je na abscisi skala za znak x, na ordinati pa skala za znak y. V korelacijskem grafikonu je torej toliko točk, kolikor je enot populacije. Ker s pogledom zajamemo vse točke hkrati, je korelacija med dvema znakoma v korelacijskem grafikonu vidna neposredno.

V slikah 12.2, 12.3 in 12.4 so prikazani korelacijski grafikoni podatkov iz tabel 12.2, 12.3 in 12.4.



Slika 12.3. Korelacijski grafikon med dohodki in izdatki za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957.



Slika 12.4. Korelacijski grafikon med odstotkom travnikov in odstotkom pašnikov za okraje v SRS v letu 1952.

Regresijska krivulja

12.8 Slike 12,2, 12.3, 12.4 so zelo poučne in pokažejo bistvo korelacijskih odvisnosti. Kakor smo že nakazali, med socialno-ekonomskimi pojavi ni funkcijske zveze, ker pojavi niso nikdar odvisni samo od enega znaka, temveč nanj vplivajo vedno najrazličnejši posamični vplivi, ki motijo oziroma zabrišejo zvezo med dvema pojavoma. Iz slike 12.2, ki prikazuje korelacijo med deležem nepismenega prebivalstva in deležem nepismenih žensk v SFRJ v letu 1961 po okrajih, sklepamo, da je odvisnost med deležem nepismenih žensk in nepismenega skupnega prebivalstva velika. Točke se zelo določeno goste okrog neke črte, s katero bi mogli prikazati odvisnost v nepismenosti. To črto, ki gre med vrisanimi točkami, imenujemo regres i j s k o k r i v u l j o . Regresijska krivulja pokaže, kakšna bi bila zveza med koreliranima znakoma, če ne bi bilo posamičnih vplivov. V tem primeru bi bile točke za vse enote na regresijski krivulji. Odvisnost med pojavoma bi bila funkcijska. Zaradi posamičnih vplivov pa se točke bolj ali manj odklanjajo od idealne regresijske krivulje. Korelacijsko odvisnost moremo torej pisati z obrazcem

$$y = f(x) + e \tag{12.3}$$

Pri tem pomeni: $f(x)$ = funkcijska oblika regresijske krivulje, e = rezultat posamičnih vplivov, ki more biti pozitiven ali negativen. Za posamezno enoto moremo vrednost y razstaviti v dva dela: $y_i = f(x_i) + e_i$, $f(x_i)$ je del, ki izvira iz povezanosti med y in x, e_i pa je rezultat posamičnih vplivov na enoti i. Čim večji so posamični vplivi e, tem večji so odkloni točk in tem bolj je zabrisana regresijska krivulja. To nazorno vidimo, če primerjamo sliko 12.2 s sliko 12.3. Regresijska krivulja je v prvem primeru dobro vidna, ker so posamični odkloni od regresijske krivulje v primeru nepismenosti majhni. Regresijska krivulja pri korelaciji stroškov za kulturno in družbeno življenje pa je manj vidna, ker so v tem primeru posamični vplivi zelo veliki. Iz tega sklepamo, da je v prvem primeru korelacijska odvisnost večja kakor v drugem. Korelacijska odvisnost more torej biti v e č j a a l i m a n j š a . Največja je v skrajnem primeru funkcijske odvisnosti ($e = o$), najmanjša pa v primeru neodvisnosti ($f(x) = C$). Če sta dva pojava neodvisna, na korelacijskem grafikonu ne moremo začrtati črte, za katero bi mogli reči, da je regresijska črta. Takrat se točke brez reda goste okrog točke, ki ima za koordinati aritmetični sredini koreliranih znakov.

V prvih dveh zgledih opazimo, da je smer regresijske krivulje taka, da se veča vrednost enega znaka, če se veča vrednost drugega. Takim vrstam povezav pravimo p o z i t i v n e p o v e z a v e . Če pa se z večanjem enega znaka vrednost drugega manjša, pa pravimo, da je p o v e z a v a n e g a t i v n a . Negativna povezanost je med odstotkoma travniške in pašniške površine po okrajih, prikazana v sliki 12.4. Iz gostitve točk moremo sklepati, da je za okraje z večjim odstotkom travnikov odstotek pašnikov v splošnem manjši in obratno.

V sliki o odvisnosti proučevanih deležev o nepismenosti po okrajih, je črta, ki ponazarja regresijsko krivuljo, premica, v obeh drugih primerih pa krivulja. V prvem primeru govorimo o linearni povezavi, v obeh drugih pa o krivuljčni povezavi med dvema pojavoma.

Metode določanja regresijskih krivulj

12.9 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.

12.10 Prostoročno določanje regresijske krivulje. Iz slik 12.2, 12.3 in 12.4 vidimo, da je smer regresijske krivulje bolj ali manj vidna že iz korelacijskega grafikona. Zato moremo pri nekaterih slabše, pri drugih pa bolje, včrtati regresijsko krivuljo prostoročno. Za regresijsko krivuljo imamo krivuljo, ki jo včrtamo med točke tako, da se meglici točk najbolje prilega. Seveda ne sme biti vrisana črta preveč komplicirana, marveč čimbolj izglajena. V naših primerih regresijsko krivuljo najlaže včrtamo za pojava, prikazana v sliki 12.2, za katera je povezava največja. Zato je regresijska krivulja najbolj vidna. Težja je odločitev v drugih dveh zgledih.

Prostoročno včrtovanje regresijskih črt je prikladno, ker je izredno preprosto. Ima pa to napako, da je subjektivno. Regresijsko črto rišemo prostoročno kot osnovo za druge načine določanja regresije. V dosti primerih pa je kljub subjektivnosti ravno zaradi svoje hitrosti v primerjavi z drugimi metodami zelo koristna.

12.11 Metoda grupnih sredin. Če izhajamo iz obrazca

$$y = f(x) + e \qquad (12.4)$$

za določanje regresijske krivulje s pridom uporabimo grupne sredine. Če za enote, ki imajo iste ali ne preveč različne vrednosti x, izračunamo poprečje iz vrednosti y, dobimo:

$$\bar{y} = \bar{f}(x) + \bar{e} \doteq f(\bar{x}) \qquad (12.5)$$

Ker štejemo, da je x konstanten, je konstantna tudi vrednost $f(x)$, poprečje konstante pa konstanta. Po znanih stavkih o sredinah pa se v poprečju vpliv posamičnih ali slučajnih faktorjev sicer ne uniči, omili se pa vsekakor. Z vrsto poprečij po grupah x dobimo tako niz točk, ki so vsaj v bližini regresijske krivulje, če že niso na njej. Če te točke vrišemo v korelacijski grafikon in zvežemo z daljicami, dobimo lomljeno črto, ki je zaradi gornjih lastnosti dober približek poteka regresijske krivulje.

Po metodi grupnih sredin določamo regresijsko krivuljo torej takole:

a) Dvojice vrednosti x, y grupiramo v razrede po vrednosti x. Grupe morajo biti take, da razredi za x niso preširoki, da pa število enot v njih ni premajhno.

b) Za vsako grupo izračunamo aritmetični sredini za vrednosti x in y. Tako dobimo vrsto dvojic grupnih aritmetičnih sredin $\bar{x}_k$ in $\bar{y}_k$.

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

d) Vrisane točke grupnih sredin med seboj zvežemo z daljicami. Lomljena črta, ki jo dobimo, je približno regresijska krivulja.

18.12 Po tej metodi določimo regresijsko krivuljo za odvisnost stroškov za kulturno in družbeno življenje za 36 delavskih gospodinjstev v Mariboru. Podatki so vzeti iz tabele 12.3.

Ker je med dohodki in stroški vzročna zveza, vzemimo dohodke za neodvisno spremenljivko (x), stroške pa za odvisno spremenljivko (y). Po analizi podatkov glede na velikost dohodkov vzamemo razrede dohodkov po 5.000 din in po njih razdelimo dvojice vrednosti x in y. Tako grupirani podatki so navedeni v tabeli 12.6.

Tabela 12.6. Dohodki in izdatki za kulturno in družbeno življenje iz tabele 12.3, grupirani po višini dohodkov.

(x = dohodki v tisoč dinarjev ; y = izdatki v dinarjih)

15,0-	x	19,7	16,5	16,7	19,9	17,2	16,6	17,5	15,4	19,6	19,7	16,5	16,9
19,9	y	260	330	490	310	580	560	330	480	860	590	330	630
20,0-	x	24,0	22,0	22,2	24,5	24,2	23,1	20,6	22,1	24,5	23,4		
24,9	y	650	1100	760	460	700	540	750	370	630	740		
25,0-	x	28,4	25,8	29,8	27,5	26,5	26,7	26,7					
29,9	y	1080	630	1210	1100	870	640	1110					
30,0-	x	32,9	30,8	31,8	30,5								
34,9	y	1610	1070	1600	990								
35,0-	x	37,5	38,4	35,9									
	y	1590	1710	1810									

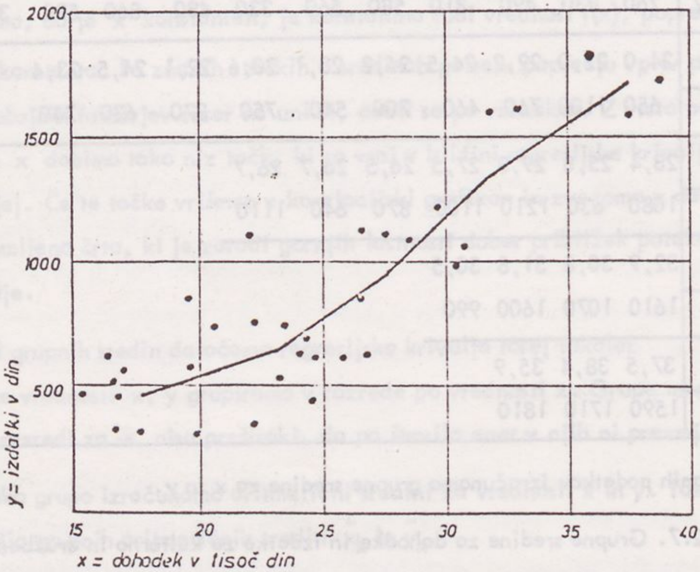
Iz grupiranih podatkov izračunamo grupne sredine za x in y.

Tabela 12.7. Grupne sredine za dohodke in izdatke za kulturno in družbeno življenje za 36 delavskih družin v Mariboru iz tabele 12.6

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

Če grupne sredine vnesemo v korelacijski grafikon, dobimo sliko 12.5. 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.

12.13 Po metodi grupnih sredin moremo določiti regresijsko črto tudi, če so podatki grupirani v korelacijski tabeli. Grupne sredine v tem primeru izračunamo po katerikoli izmed metod za izračunavanje sredin iz frekvenčnih porazdelitev. Te grupne sredine včrtamo nad sredine razrednih razmikov.



Slika 12.5. Regresijska črta med dohodki in izdatki za kulturno in družbeno življenje za 36 delavskih družin v Mariboru v novembru 1957.

Za korelacijo med poprečnimi dohodki zaposlenih in dohodki direktorjev v obrti iz tabele 12.5 je smiselno izračunati obe regresijski vrsti sredin, ker zveza ni vzročna. Ker so v korelacijski tabeli podatki že grupirani, grupiranje odpade.

Če najprej izračunamo poprečja za dohodke direktorjev za posamezne dohodkovne skupine zaposlenih, dobimo regresijsko črto, ki kaže odvisnost dohodkov direktorjev od dohodkov zaposlenih. Ker je v frekvenčni porazdelitvi dohodkov direktorjev za dohodkovno skupino do 10 tisoč din ena sama frekvenca, je ocena poprečja sredina razreda $\bar{y}_1 = 30$. Za skupino 10-20 je $\bar{y}_2 = \frac{18.30+19.50}{37} = 40.27$ tisoč din itd.

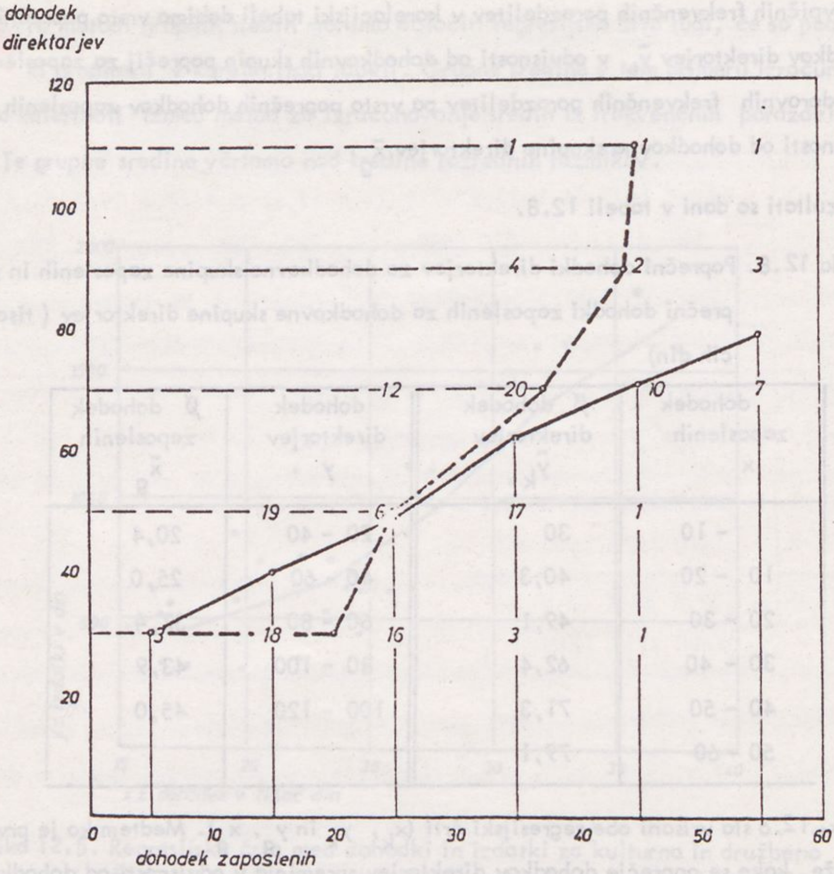
Iz navpičnih frekvenčnih porazdelitev v korelacijski tabeli dobimo vrsto poprečnih dohodkov direktorjev $\bar{y}_k$ v odvisnosti od dohodkovnih skupin poprečij za zaposlene, iz vodoravnih frekvenčnih porazdelitev pa vrsto poprečnih dohodkov zaposlenih v odvisnosti od dohodkovne skupine direktorjev $\bar{x}_g$.

Ti rezultati so dani v tabeli 12.8.

Tabela 12.8. Poprečni dohodki direktorjev za dohodkovne skupine zaposlenih in poprečni dohodki zaposlenih za dohodkovne skupine direktorjev (tisočih din)

dohodek zaposlenih x	Ø dohodek direktorjev $\bar{y}_k$	dohodek direktorjev y	Ø dohodek zaposlenih $\bar{x}_g$
- 10	30	20 - 40	20,4
10 - 20	40,3	40 - 60	25,0
20 - 30	49,1	60 - 80	37,4
30 - 40	62,4	80 - 100	43,9
40 - 50	71,3	100 - 120	45,0
50 - 60	79,1		

V sliki 12.6 sta vrisani obe regresijski črti (x_k, $\bar{y}_k$ in y_g, $\bar{x}_g$). Medtem ko je prva, ki kaže, kako se poprečje dohodkov direktorjev spreminja v odvisnosti od dohodkov zaposlenih, skoraj linearna, je druga, ki kaže obrnjene odnose, krivuljčna. Obe regresijski črti se praviloma sekata. Zaradi nazornosti so v sliki v ustreznih poljih vpisane tudi frekvence.



Slika 12.6 Regresijske črte sredin med dohodki zaposlenih in dohodki direktorjev iz tabele 18.8.

12.14 Analitična metoda. Regresijsko krivuljo moremo določiti tudi analitično. Če po analizi podatkov ugotovimo ustrezen tip funkcije

$$y' = f(x\,;\ a,\ b,\ c\ldots) \qquad (12.6)$$

ki ima več parametrov (a, b, c...), primeren za regresijsko krivuljo, je treba glede na podatke, ki jih proučujemo, poiskati vrednosti parametrov a, b, c..., tako da se regresijska krivulja danim vrednostim najbolj prilega. Kot merilo prilagojenosti krivulje vzamemo vsoto kvadratov odklonov stvarnih vrednosti y od y´

$$\sum_i (y_i - y'_i)^2 = F(a, b, c \ldots) \qquad (12.7)$$

Za regresijsko krivuljo vzamemo izmed vseh krivulj istega tipa tisto, za katero je vsota kvadratov odklonov stvarnih vrednosti y od vrednosti y´ na regresijski krivulji F(a,b,c...) najmanjša. Po tej metodi, ki jo imenujemo m e t o d o n a j m a n j š i h k v a d r a t o v , moremo določiti vrednosti parametrov za regresijsko krivuljo in tako regresijsko krivuljo samo. Kadar ugotavljamo regresijsko krivuljo analitično, uporabljamo najpogosteje metodo najmanjših kvadratov.

Mere jakosti odvisnosti

12.15 Dejavnike, ki vplivajo na določen pojav (y), moremo razdeliti v tri skupine:

a) na splošne vplive, ki so za vse enote isti.

b) na vpliv faktorja x, s katerim je y v korelaciji.

c) na ostale posamične vplive skupno s slučajnostnimi.

Tako je na primer dohodek zaposlenih odvisen od:

a) splošnih vplivov: kategorije zaposlenih, časa, za katerega velja dohodek itd.,

b) let službe, s katerimi je dohodek v korelacijski zvezi,

c) drugih, posamičnih vplivov, kot so: vestnost pri delu, šolska izobrazba, delovno mesto, itd.

Prav tako je na primer proizvodnja kmetijskega obrata odvisna od:

a) splošnih faktorjev: klimatskih razmer, specializiranosti itd.

b) velikosti gospodarstva, s katero je proizvodnja v korelacijski povezanosti in

c) od drugih, posamičnih faktorjev, ki vplivajo na proizvodnjo, kot so: razlike tal, različna produktivnost dela, različno seme itd.

Če poznamo aritmetično sredino $\bar{y}$ in regresijsko krivuljo y´ med x in y, moremo celotno vrednost y za posamezno enoto razdeliti na nakazane tri komponente, od katerih vsaka prikazuje rezultat vplivov po naštetih treh skupinah.

y je rezultat vseh dejavnikov, ki vplivajo na pojav; splošnih, koreliranega in drugih posamičnih; y´ je rezultat splošnih dejavnikov in koreliranega dejavnika x; $\bar{y}$ je rezultat samo splošnih faktorjev. Zato moremo pisati

$$y = \bar{y} + (y' - \bar{y}) + (y - y') \qquad (12.8)$$

Pri tem pomeni:

$\bar{y}$ = rezultat splošnih vplivov; $(y' - \bar{y})$ = rezultat vpliva koreliranega znaka x;

$(y - y')$ = rezultat vpliva drugih, posamičnih dejavnikov.

Tako smo uspeli razdeliti celotno vrednost y po značaju dejavnikov v tri ločene sestavine.

Če je vpliv dejavnika x velik, je odklon $(y' - \bar{y})$ velik. Če je vpliv x majhen, so vrednosti $(y' - \bar{y})$ majhne, če pa y od x ni odvisen, je $(y' - \bar{y})$ za vse enote enak nič. Za skupno merilo velikosti vpliva x na y uporabimo varianco odklonov $(y' - \bar{y})$

$$\sigma_{y'}^2 = \frac{1}{N}\sum (y' - \bar{y})^2 \qquad (12.9)$$

Enako so odkloni $(y - y')$ veliki, če so močni posamični dejavniki. Če je njihov vpliv majhen, so ti odkloni majhni, če jih ni, je $(y - y')$ enak nič. Povezava je v tem primeru funkcijska. Skupno merilo jakosti posamičnih vplivov je analogno zgornjemu merilu varianca odklonov $(y - y')$

$$\sigma_e^2 = \frac{1}{N}\sum (y - y')^2 \qquad (12.10)$$

Po stavku o razstavljanju variance na vsoto varianc po dejavnikih (glej obrazec 10.18) velja, da je skupna varianca σ_y^2 vsota dveh varianc: variance zaradi vpliva x na y $\sigma_{y'}^2$, in variance zaradi vpliva posamičnih dejavnikov σ_e^2.

Ker tako pojasnimo, da en del celotne variance izvira iz povezave y z x, imenujemo $\sigma_{y'}^2$ p o j a s n j e n o v a r i a n c o, za razliko od σ_e^2, ki izvira iz drugih, neznanih dejavnikov in jo imenujemo n e p o j a s n j e n o varianco. Z obrazcem moremo gornje napisati:

$$\sigma_y^2 = \sigma_{y'}^2 + \sigma_e^2 \tag{12.11}$$

Če to enačbo delimo z σ_y^2 dobimo

$$\frac{\sigma_{y'}^2}{\sigma_y^2} + \frac{\sigma_e^2}{\sigma_y^2} = 1 \tag{12.12}$$

V tem obrazcu kvocient $\sigma_{y'}^2 / \sigma_y^2$ pove, kolikšen del skupne variance y je pojasnjene s povezanostjo y z x, σ_e^2 / σ_y^2 pa, kolikšen del skupne variance y odpade na druge, nepojasnjene vplive.

Velikost izraza $\sigma_{y'}^2 / \sigma_y^2$ se ravna po tem, kako velik je vpliv x na y. Če y od x ni odvisen, je vrednost tega izraza nič, če pa je odvisnost funkcijska, je ta izraz največji in sicer enak 1. Vrednost kvocienta $\sigma_{y'}^2 / \sigma_y^2$ je torej med 0 in 1 ter pomeni večjo ali manjšo stopnjo vpliva x na y. Zato kvocient $\sigma_{y'}^2 / \sigma_y^2$ uporabljamo za merilo stopnje odvisnosti y od x in ga imenujemo d e t e r m i n a c i j s k i k o e f i c i e n t . Čim večji je vpliv x na y, tem večja je namreč vrednost determinacijskega koeficienta.

Čeprav je determinacijski koeficient $\sigma_{y'}^2 / \sigma_y^2$ vsebinsko zelo nazoren, saj je delež pojasnjene variance od skupne variance, dostikrat namesto njega uporabljamo kot merilo povezanosti kvadratni koren tega izraza. To merilo korelacije imenujemo i n d e k s k o r e l a c i j e . Indeks korelacije je torej razmerje med standardnim odklonom za pojasnjeno variabilnost in skupnim standardnim odklonom za y

$$I_{y.x} = \sigma_{y'} / \sigma_y \tag{12.13}$$

Če upoštevamo enačbi 12.12 in 12.13, moremo indeks korelacije pisati v obliki, ki jo splošno uporabljamo

$$I_{y.x} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}} \qquad (12.14)$$

Standardna napaka ocene

12.16 Če je povezava med x in y tesna, se točke za dejanske vrednosti x in y razmeroma malo odklanjajo od regresijske krivulje. Ako za dano enoto poznamo samo vrednost x, moremo vzeti ustrezno vrednost y´ na regresijski črti za oceno dejanske vrednosti y. To ocenjevanje ima veliko praktično vrednost. Če poznamo regresijsko krivuljo med količino padavin in hektarskim pridelkom, moremo iz količine padavin v določenem letu napovedati hektarski pridelek. Če imamo regresijsko krivuljo med oceno in stvarnim hektarskim pridelkom, moremo z njo popraviti vizuelno oceno. Če poznamo regresijsko krivuljo o odvisnosti cene od količine blaga, ki je naprodaj, moremo pri znani količini blaga oceniti ceno itd.

Ta ocena je jasno tem boljša, čim bolj je y odvisen od x. Najboljša bi bila seveda, če bi bila funkcijska odvisnost. Takrat bi se iz regresijske krivulje izračunana vrednost y´ skladala s pravo vrednostjo y. Vendar, kakor vemo, takih primerov v stvarnosti ni.

Za vsako posamezno oceno ne vemo, koliko smo pri ocenjevanju pogrešili, ali z drugimi besedami: kolikšna je razlika ocene y´ od stvarne vrednosti y. Vendar moremo izračunati standardni odklon teh napak. Napaka pri ocenjevanju izvira iz posamičnih vplivov, ki jih ne poznamo. Standardni odklon napake je torej enak standardnemu odklonu iz nepojasnjene variance. To pa moremo izračunati, če poznamo indeks korelacije $I_{y.x}$ med podatkoma y in x. Iz obrazca 18.14 namreč zlahka dobimo, da je

$$\sigma_e = \sigma_y \sqrt{1 - I_{y.x}^2} \qquad (12.15)$$

Standardni odklon nepojasnjene variance zato imenujemo tudi standardno napako ocene. Če predpostavimo, da se odkloni ocen od pravih vrednosti porazdeljujejo normalno, moremo sklepati, da 68% ocen ne bo napačnih za več kot σ_e, približno 95% ne več kot $2.\sigma_e$ in da bo le izjemoma ocena od prave vrednosti različna za več kot $3.\sigma_e$. Tako moremo s standardno napako ocene oceniti napako ocene, ki jo dobimo z regresijsko enačbo.

Koeficient $\sqrt{1 - I^2_{y.x}}$ s katerim pomnožimo standardni odklon σ_y, da dobimo standardno napako ocene σ_e, imenujemo na splošno **koeficient alienacije** ali **koeficient nepovezanosti**.

LINEARNA ODVISNOST

12.17 Pokazovalci. Dosedanja razglabljanja o regresiji in korelaciji veljajo na splošno, neodvisno od tipa regresijske krivulje. V vsakem primeru izračunamo regresijsko krivuljo, ki pokaže obliko odvisnosti med x in y, indeks korelacije, ki pokaže jakost povezanosti in standardno napako ocene, ki pokaže kvaliteto ocenjevanja z regresijsko krivuljo.

Vse te količine najenostavneje izračunamo, če je odvisnost med x in y linearna, ali drugače povedano, če je regresijska krivulja premica. Linearne odvisnosti so v praksi zelo pogoste. Tudi če je povezava krivuljčna, je na krajših odsekih premica zelo dober približek regresijske krivulje. Zato je linearna regresija za prakso zelo važna in jo najpogosteje uporabljamo.

Če upoštevamo načela metode najmanjših kvadratov, izračunavamo pokazovalce linearne regresije in korelacije po osnovnih obrazcih:

Ker imamo v splošnem dve regresijski krivulji, imamo tudi dve regresijski premici.

Prva

$$y' = M_y + b_1 (x - M_x) \tag{12.16}$$

podaja odvisnost y od x, druga

$$x' = M_x + b_2 (y - M_y) \tag{12.17}$$

pa pokaže odvisnost x od y.

V gornjih enačbah pomenita M_x in M_y aritmetični sredini za x in y, b_1 in b_2 pa sta **prvi in drugi regresijski koeficient**. Regresijska koeficienta sta smerna koeficienta obeh regresijskih premic.

Prvi regresijski koeficient b_1 pokaže, za koliko se v poprečju spremeni y, če se x poveča za enoto; drugi regresijski koeficient b_2 pa pokaže, za koliko se v poprečju spremeni x, če se y poveča za enoto. Regresijska koeficienta b_1 in b_2 računamo po obrazcih

$$b_1 = \frac{C_{xy}}{\sigma_x^2} ; \quad b_2 = \frac{C_{xy}}{\sigma_y^2} \tag{12.18}$$

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 poprečen produkt odklonov x in y od ustreznih aritmetičnih sredin M_x in M_y in jo izrazimo z obrazcem

$$C_{xy} = \frac{1}{N}\sum (x - M_x)(y - M_y) \tag{12.19}$$

Za razliko od variance more biti vrednost kovariance pozitivna ali negativna.

Indeks korelacije pri linearni korelaciji imenujemo **korelacijski koeficient**. Zaznamujemo pa ga z ρ_{xy}.

Korelacijski koeficient ρ_{xy} izračunamo po osnovnem obrazcu

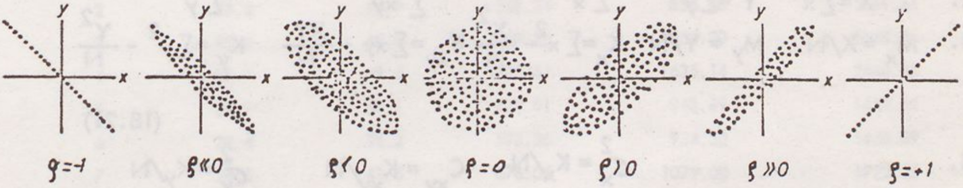
$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} \tag{12.20}$$

Vrednosti korelacijskega koeficienta leže med -1 in +1. Korelacijski koeficient je pozitiven, če je povezava med x in y pozitivna; to pomeni, da se veča tudi vrednost druge spremenljivke, če se veča vrednost prve.

Korelacijski koeficient pa je negativen, če je povezava negativna. Taka pa je povezava, če se z večanjem enega znaka vrednost drugega manjša.

Korelacijski koeficient ima vrednost -1 oziroma +1, če je povezava med x in y funkcijska in linearna, 0 pa, kadar med x in y ni linearne zveze.

Slika 18.7 nazorno kaže velikost korelacijskega koeficienta za nekaj najznačilnejših oblik linearnih odvisnosti med x in y. Te povezave so ponazorjene s korelacijskim grafikonom.



Slika 12.7. Vrednosti korelacijskega koeficienta pri različnih razmestitvah točk v korelacijskem grafikonu.

Kvadrat korelacijskega koeficienta ρ^2_{xy} imenujemo **determinacijski koeficient** za linearno odvisnost. Determinacijski koeficient sicer nima predznaka, ki bi nakazoval smer povezave, pač pa pove, kakšen del celotne variance je pojasnjen z linearno zvezo med x in y. Zato je analitično pomembnejši kakor korelacijski koeficient. Njegove vrednosti so med 0 in 1.

Računanje pokazovalcev za linearno odvisnost

12.18 Če imamo dane posamične podatke x_i, y_i $(i = 1,2..N)$ za celotno populacijo, izračunamo pokazovalce linearne regresije in korelacije po shemi:

	x	y	x^2	xy	y^2
1.	x_1	y_1	x_1^2	$x_1 y_1$	y_1^2
	x_2	y_2	x_2^2	$x_2 y_2$	y_2^2
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	x_N	y_N	x_N^2	$x_N y_N$	y_N^2
2.	$X=\sum x$	$Y=\sum y$	$\sum x^2$	$\sum xy$	$\sum y^2$

3. $M_x = X/N \quad M_y = Y/N \quad K_x = \sum x^2 - \frac{X^2}{N} \quad K_{xy} = \sum xy - \frac{XY}{N} \quad K_y = \sum y^2 - \frac{Y^2}{N}$

$$(18.21)$$

4. $\sigma_x^2 = K_x/N \qquad C_{xy} = K_{xy}/N \qquad \sigma_y^2 = K_y/N$

5. $\sigma_x = \sqrt{\sigma_x^2} \qquad \sigma_y = \sqrt{\sigma_y^2}$

6. $b_1 = \frac{C_{xy}}{\sigma_x^2} = \frac{K_{xy}}{K_x}\,; \quad \varrho_{xy}^2 = b_1 b_2 = \frac{C_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2}\,; \quad b_2 = \frac{C_{xy}}{\sigma_y^2} = \frac{K_{xy}}{K_y}$

7. $\varrho_{xy} = \frac{K_{xy}}{\sqrt{K_x K_y}}$

8. $\sigma_{e.y} = \sigma_y \sqrt{1 - \varrho_{xy}^2}\,, \sqrt{1 - \varrho_{xy}^2}\,; \quad \sigma_{e.x} = \sigma_x \sqrt{1 - \varrho_{xy}^2}$

9. $y' = M_y + b_1 (x - M_x) \qquad x' = M_x + b_2 (y - M_y)$

V shemi je nazorno nakazan potek za računanje pokazovalcev linearne odvisnosti. Pogosto posamičnih rezultatov iz točke 1 niti ne pišemo, ker na računskem stroju dobimo vsote iz točke 2 direktno

Včasih se izkaže tehnično koristno, če obračunamo pokazovalce linearne odvisnosti iz reduciranih vrednosti $u = x - x_o$ in $v = y - y_o$. Vse količine, razen $M_x = x_o + M_u$; $M_y = y_o + M_v$ so invariantne in velja $\sigma_x^2 = \sigma_u^2$; $\sigma_y^2 = \sigma_v^2$;

$$C_{xy} = C_{uv}\ ; \quad b_{1xy} = b_{1uv}\ ; \quad b_{2xy} = b_{2uv}\ ; \quad \varrho_{xy} = \varrho_{uv}$$

Zgled za izračunanje bomo podali v naslednjih odstavkih.

12.19 Za zgled računanja pokazovalcev za linearno odvisnost vzemimo podatke o odstotkih nepismenega prebivalstva (x) in odstotkih nepismenih žensk (y) v SR Srbiji v letu 1961 po okrajih iz tabele 12.2. Računanje je prikazano po shemi iz 12.18.

Tabela 12.9. Računanje pokazovalcev linearne odvisnosti med odstotkom nepismenega prebivalstva in nepismenih žensk v SR Srbiji v letu 1961 po okrajih.

i	x	y	x^2	xy	y^2
1	9.2	14.5	84.64	133.40	210.25
2	23.2	36.9	538.24	856.08	1361.61
3	25.5	40.6	650.25	1035.30	1648.36
4	33.1	49.4	1095.61	1635.14	2440.36
5	24.9	38.1	620.01	948.69	1451.61
6	24.4	38.3	595.36	934.52	1466.89
7	26.0	41.5	676.00	1079.00	1722.25
8	26.7	43.3	712.89	1156.11	1874.89
9	25.8	38.5	665.64	993.30	1482.25
10	9.8	14.1	96.04	138.18	198.81
11	12.1	16.8	146.41	203.28	282.24
12	11.6	16.6	134.56	192.56	275.56
13	8.6	11.6	73.96	99.76	134.56
14	12.1	17.2	146.41	208.12	295.84
15	44.1	56.0	1944.81	2469.60	3136.00
	$\Sigma x = 317.10$	$\Sigma y = 473.40$	$\Sigma x^2 = 8180.83$	$\Sigma xy = 12083.04$	$\Sigma y^2 = 17981.48$
	$M_x = 21.40$	$M_y = 31.56$	$K_x = 1477.34$	$K_{xy} = 2075.40$	$K_y = 3040.98$

$\sigma_x^2 = 98.49$ $\quad C_{xy} = 138.36$ $\quad \sigma_y^2 = 202.73$

$\sigma_x = 9.924$ $\quad \sigma_y = 14.238$

$b_1 = 1.405$ $\quad \rho_{xy}^2 = .959$ $\quad b_2 = .628$

$\rho_{xy} = .979$

$\sigma_{e.y} = 2.890$ $\quad \sqrt{1 - \rho_{xy}^2} = .203$ $\quad \sigma_{e.x} = 2.014$

$y' = 31.56 + 1.405\,(x - 21.40)$ $\qquad x' = 21.40 + .682\,(y - 31,56)$

$y' = 1.49 + 1.405x$ $\qquad x' = -.124 + .682y$

Iz zgleda spoznamo, da je odvisnost med deležem nepismenih za vse prebivalstvo in deležem nepismenih žensk velika, saj je determinacijski koeficient $\rho^2_{xy} = .959$. Zato z regresijsko enačbo dokaj dobro ocenjujemo oziroma napovedujemo delež nepismenih žensk, če poznamo delež nepismenega prebivalstva in obratno.

Če za zgled uporabimo regresijsko enačbo

$$y' = 1,49 + 1,405x ,$$

moremo z njo napovedati za posamezne okraje v SR Srbiji delež nepismenih žensk, če poznamo delež nepismenega prebivalstva. Za deveti okraj je delež nepismenega prebivalstva $x_9 = 25,8\%$. Iz zgornje regresijske enačbe dobimo, da je napoved oziroma ocena deleža za žensko prebivalstvo

$$y'_9 = 1,49 + 1,405 \, . \, 25,8 = 37,7\,\%$$

Ocenjena vrednost $y'_9 = 37,7\%$ je od prave vrednosti $y_9 = 38,5\%$ različna le za $y_9 - y'_9 = 38,5\% - 37,7\% = 0,8\%$.

Po mnenju ministrstva za šolstvo in šport Republike Slovenije št. 415-13/92 z dne 14. februarja 1992 je ta skripta uvrščena v tarifni razred št. 3 zakona o prometnem davku (Ur. l. RS, št.4/92), po katerem se obračunava prometni davek po 5 % stopnji.

... odvisnost med deležem nepismenih za vse prebivalstvo in deležem nepismenih žensk velika, saj je determinacijski koeficient $\rho^2_{xy} = ,959$. Zato z regresijsko enačbo dokaj dobro ocenjujemo oziroma napovedujemo delež nepismenih žensk, če poznamo delež nepismenega prebivalstva in obratno.

Če za zgled uporabimo regresijsko enačbo

$$y' = 1,49 + 1,405x ,$$

moremo z njo napovedati za posamezne okraje v SR Srbiji delež nepismenih žensk, če poznamo delež nepismenega prebivalstva. Za deveti okraj je delež nepismenega prebivalstva $x_9 = 25,8\%$. Iz zgornje regresijske enačbe dobimo, da je napoved oziroma ocena deleža za ženske prebivalstvo

$$y'_9 = 1,49 + 1,405 \cdot 25,8 = 37,7\ \%$$

Ocenjena vrednost $y'_9 = 37,7\%$ je od prave vrednosti $y_9 = 38,5\%$ različna je za $y_9 - y'_9 = 38,5\% - 37,7\% = 0,8\%$.

Po mnenju ministrstva za šolstvo in šport Republike Slovenije št. 415-13/92 z dne 14. februarja 1992 je ta skripta uvrščena v tarifni razred št. 3 zakona o prometnem davku (Ur. l. RS, št. 4/92), po katerem se obračunava prometni davek po 5 % stopnji.