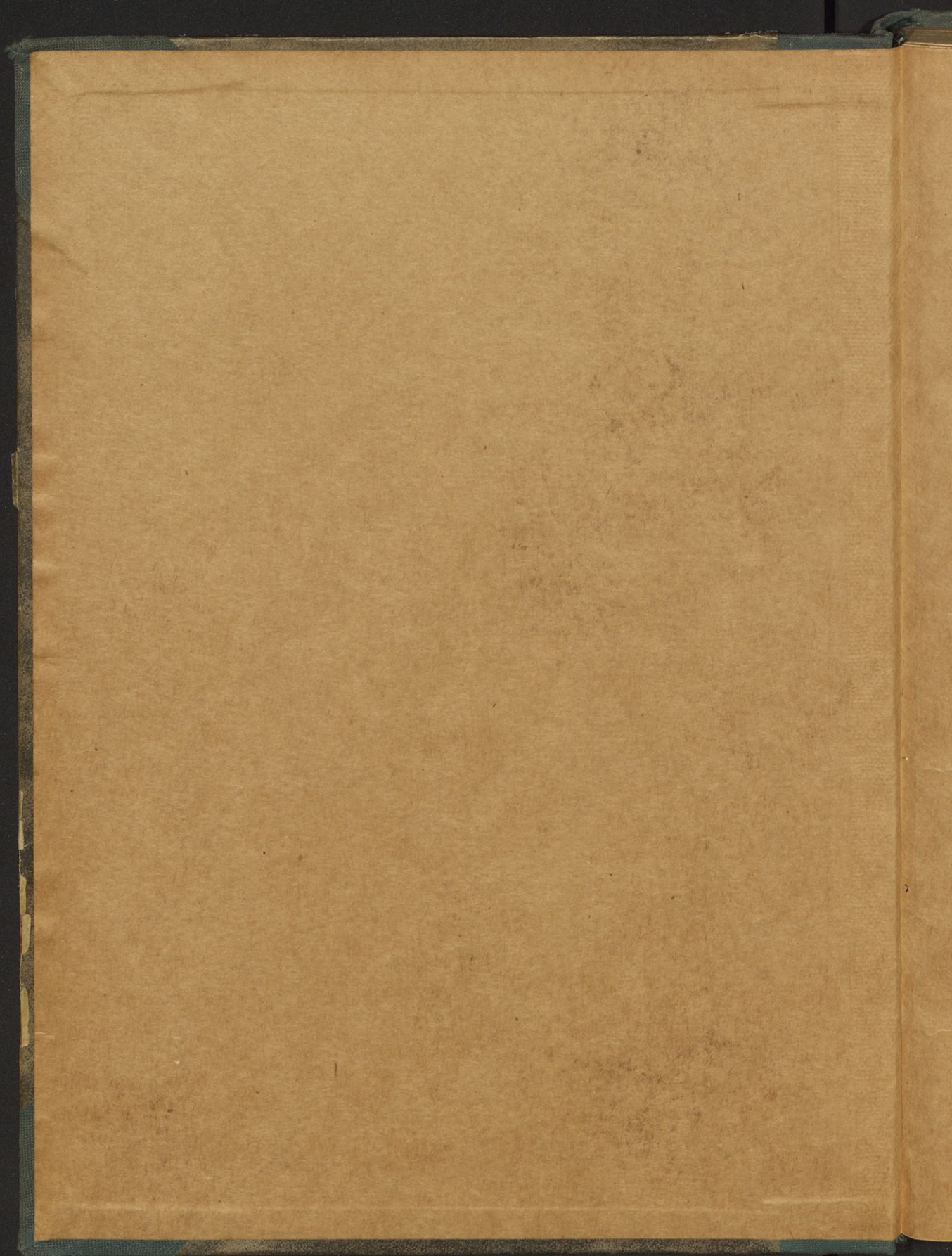


Narodna in univerzitetna knjižnica
v Ljubljani

167536



UNIVERZA V LJUBLJANI
FAKULTETA ZA AGRONOMIJO, GOZDARSTVO IN VETERINO

Dr. MARJAN BLEJEC

STATISTIČNE
METODE V GOZDARSTVU

1961

Univerza v Ljubljani
Fakulteta za agronomijo, gozdarstvo in veterino

Dr. Marjan Blejec
izr. profesor Ekonomske fakultete

STATISTIČNE METODE V GOZDARSTVU

1961
Univerzitetna založba
Ljubljana

I 167536

II 167536



769/1962

Dar: list za prouč. in kult. v Lj
20. 3. 62

Za natis odobrila: Tiskovna komisija fakultete za
agronomijo, gozdarstvo in veteri-
rino

dne 30. marca 1961

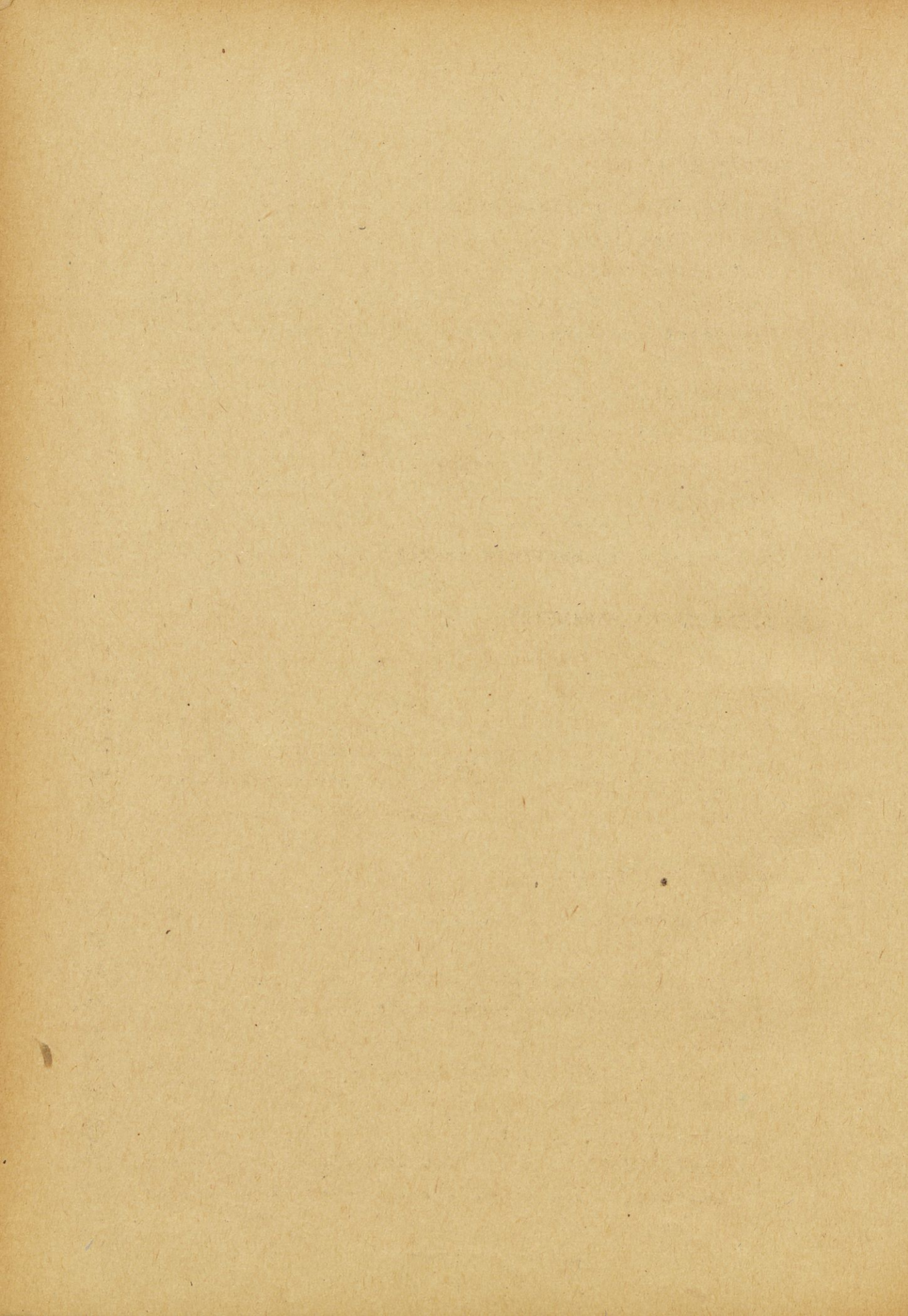
Skripta o statističnih metodah v gozdarstvu so napisana po predavanjih enosemestrskega kurza o "Metodiki raziskovalnega dela v gozdarstvu" v zimskem semestru študijskega leta 1960/61 na Gozdarskem oddelku FAGV v Ljubljani.

Po vsebini so skripta razdeljena v 11 poglavij, ki vsebujejo osnovne statistične metode, ki jih mora poznati gozdarski strokovnjak pri svojem vsakodnevnem delu. Sorazmerno obsežnejše in podrobnejše je obdelano vzorčenje, ker so v specifičnih potrebah v gozdarstvu sodobni načini ocenjevanja posebno pomembni.

Osnova za večino primerov je gradivo raziskav, ki so ga iz prijaznosti dali na razpolago IGLIS in posamezne katedre Gozdarskega oddelka FAGV, za kar se vsem zahvaljujem.

Ljubljana, september 1961

M. Blejec



1 MNOŽIČNI POJAVI

Proučevanje množičnih pojavov	1
Statistične enote	3
Statistični znaki	4
Populacije	7
Parametri	9

2 GRUPIRANJE

Zaokroževanje podatkov	10
Grupiranje vrednosti znakov	10
Načela grupiranja	11
Grupiranje numeričnih znakov	12
Grupiranje nenumeričnih znakov	15

3 ELEMENTARNI PARAMETRI

Število enot in vsote	17
Statistične vrste	17
• Frekvenčne porazdelitve	20
Sestavljanje frekvenčnih porazdelitev	20
Grafično prikazovanje frekvenčnih porazdelitev	24
Kumulativne frekvenčne porazdelitve	29

4 RELATIVNA ŠTEVILA

Vrste primerjav	33
Strukture	33
Enostavne strukturne vrste	34
Kombinirane strukturne vrste	36
• Indeksi	37
Indeksi s stalno osnovo	37
Verižni indeksi	39
Koeficienti in gostote	41
Koeficienti	41
Gostote	42

5 • KVANTILI

Ranžirna vrsta. Rang	46
Kvantilni rang	47
Kvantili	48
Izračunavanje kvantilnih rangov in kvantilov iz negrupiranih podatkov	49
Izračunavanje kvantilnih rangov in kvantilov iz frekvenčnih porazdelitev	52

6 SREDNJE VREDNOSTI

Vrste srednjih vrednosti	58
• Mediana	59
Lastnosti mediane	60
• Modus	60
Izračunavanje modusa iz frekvenčnih porazde- litv	61
Lastnosti modusa	64
• Aritmetična sredina	64
Lastnosti aritmetične sredine	65
Izračunavanje aritmetične sredine iz negrupi- ranih podatkov	66
Izračunavanje aritmetične sredine iz frekvenč- nih porazdelitev	66
Neposredna metoda	66
Metoda s pomožnim znakom u	68
Metoda kumulativ	69
Aritmetična sredina aritmetičnih sredin	71
• Harmonična sredina	73
Izračunavanje sumarnih relativnih števil	75
Kvadratična sredina	76
Geometrijska sredina	78
Zveze med srednjimi vrednostmi	80

7 MERE VARIACIJE

Vrste mer variacije	83
Variacijski razmak	84
• Kvartilni odklon	85
Povprečen absoluten odklon	86

Varianca in standardni odklon	88
Izračun variance in standardnega odklona	88
Izračun variance in standardnega odklona iz negrupiranih podatkov. Neposredna metoda	89
Metoda pomožnega znaka u	90
Izračun variance in standardnega odklona iz frekvenčnih porazdelitev	92
Metoda pomožnega znaka u	92
Metoda kumulativ	94
Sheppardov popravek	95
Zveza standardnega odklona z normalno porazde- litvijo	96
Skupna varianca	97
Relativne mere variacije	100

8 KORELACIJA

Funkcijske odvisnosti	103
Korelacijske odvisnosti	104
Vzorčne in čiste korelacijske odvisnosti	105
Prikazovanje korelacijskih odvisnosti	106
Korelacijski grafikon	106
Korelacijska tabela	108
Značilnosti korelacijskih odvisnosti	110
Določanje regresijskih krivulj	111
Prostorčna metoda	111
Metoda sredin	111
Analitična metoda za določanje regresijskih krivulj	116
Mera stopnje odvisnosti	116
Linearna regresija in korelacija	118
Izračunavanje pokazateljev za linearno regre- sijo in korelacijo	120

9 TEORETIČNE PORAZDELITVE

Pomen proučevanja teoretičnih porazdelitev	124
Normalna porazdelitev	125
Opis normalne porazdelitve	125
Standardiziran znak z	129

Standardizirana normalna porazdelitev	130
Prilagoditev normalne porazdelitve stvarni porazdelitvi	133
Verjetnostne porazdelitve	135
Pojem tveganja	136

10 VZORČENJE - OCENJEVANJE PARAMETROV

Metode delnega opazovanja	141
Vzorčenje	141
Osnovna populacija. Enota opazovanja	142
Enote vzorčenja	143
Vzorec	144
Populacija vseh možnih vzorcev	145
Teoretične vzorčne porazdelitve	146
Enostavno slučajnostno vzorčenje	147
Ocenjevanje aritmetične sredine	147
Točkovna ocena. Odklon zaupanja. Intervalna ocena. Razmak zaupanja	149
Nepriistranska ocena variance in povprečnega kvadratičnega odklona	151
Preskus zakonitosti vzorčenja na shematični populaciji	152
Ocenjevanje parametrov z enostavnim slučajnostnim vzorčenjem na splošno	158
Ocenjevanje agregata $\bar{Y}$, strukturnega deleža $P\%$ in števila enot z dano značilnostjo H	159
Tehnika izbora enostavnega slučajnostnega vzorca	163
Loterijski način	164
Tablice slučajnostnih števil	165
Primer za ocenjevanje z enostavnim slučajnostnim vzorcem	169
Določanje velikosti vzorca pri zahtevani natančnosti ocene	176
Določanje velikosti vzorca za oceno povprečja	176
Velikost vzorca pri predpisani relativni natančnosti	178
Določanje števila enot v vzorcu za ocene parametrov $\bar{Y}$, Y , $P\%$, H	179
Stratificirano vzorčenje	180

	Stran
Osnova in lastnosti	180
Razmestitev enot vzorčenja po stratumi . . .	182
Proporcionalna razmestitev	182
Optimalna razmestitev	184
Optimalna razmestitev glede na stroške . .	185
Primer za ocenjevanje s stratificiranim vzor- čenjem	188
Vzorčenje v skupinica	193
Vrste skupinic	194
Primer vzorčenja skupinic iregularnih površin.	195
Vzorčenje v pasovih	199
Primer za ocenjevanje s progami	199
Sistematično vzorčenje	200
Osnova	200
Sistematično vzorčenje kot vzorec skupinic . .	202
Sistematično vzorčenje na površinah	204
Ocenjevanje parametrov s sistematičnim vzorče- njem	205
Ocenjevanje variance za ocene parametrov . . .	205
Primer za sistematično vzorčenje	206
Vzorčenje v dveh in več stopnjah	209
Osnova	209
Prednosti in pomanjkljivosti	210
Ocenjevanje agregata in varianca za oceno agre- gata z vzorčenjem v dveh stopnjah	211
Ocenjevanje po metodi razmerij	212
Osnova in ocena agregata	212
Varianca ocene za agregat po metodi razmerij .	213
Primeri za uporabo metode razmerij	213
Primer za ocenjevanje po metodi razmerij . . .	214
Vzorčenje v dveh fazah	216
Vzorčenje v dveh fazah pri metodi razmerij . .	216
Vzorčenje v dveh fazah pri stratifikaciji . .	217
Primeri za ocenjevanje z vzorčenjem v dveh fa- zah pri stratifikaciji	218
Varianca ocene z vzorčenjem v dveh fazah . . .	219
Sestavljanje planov vzorčenja v praksi	219

11 VZORČENJE - PRESKUŠANJE HIPOTEZ

Osnove	220
Ničelna hipoteza	223
Preskušanje hipotez z malimi vzorci	225
Preskušanje hipoteze o sredini	225
Preskušanje značilnosti razlik med aritmetič- nimi sredinami	228
Preskušanje hipotez o varianci	230
Preskušanje razlik med variancama	233
Enostavna analiza variance	238
Preskušanje hipotez z veliki vzorci	241
Preskušanje značilnosti razlik za velike vzorce .	243

1. MNOŽIČNI POJAVI

Proučevanje množičnih pojavov

1.1

Če se omejimo na prirodo, spoznamo, da je sestavljena iz nebroj stvari, rastlin, živali, dogodkov in drugih pojavov. Ti so med seboj različni ali podobni, sorodni ali tuji, povezani in odvisni, vplivajo drug na drugega itd. Vsaka izmed teh rastlin, živali, dogodkov ali na kratko pojavov se v prostoru in času pojavlja v velikem številu. Vse te pojave imenujemo množične pojave.

V večini znanostih, v prirodnoslovnih pa še posebej, ima izolirano opazovanje posameznih primerov omejen pomen. Že dolgo ravno v prirodnoslovnih znanostih izkoriščamo dejstvo, da se šele pri opazovanju velikega števila sorodnih pojavov pokažejo zakonitosti, katere z individualnim opazovanjem ne moremo odkriti.

Metode proučevanja množičnih pojavov so specifične in se je razvila posebna znanstvena disciplina proučevanja množičnih pojavov - statistika. Ta se ukvarja z opisovanjem, raziskavo in analizo množičnih pojavov. Statistika se je razvila v znanost, ki s kvantitativnim proučevanjem množičnih pojavov s specifičnimi metodami odkriva zakonitosti množičnega pojavljanja in podaja kvalitativno sliko proučevanega pojava. Statistika ima torej svoj predmet proučevanja - množične pojave, specifične metode kvantitativnega proučevanja in svoje področje - odkrivanje zakonitosti množičnega pojavljanja.

Gozdarstvo je tipično področje, kjer je večina pojavov, ki jih proučujemo, množičnih. Tako je množično drevo določene drevesne vrste, ker v prostoru in času nastopa v velikem številu. Iz istega vzroka je množičen pojav določena vrsta insekta, posek drevesa, veja, list drevesa, izvrtek iz drevesa pri gozdarskih poskusih itd. Zato so statistične metode važno orodje pri proučevanju pojavov v gozdarstvu.

Pri statističnem proučevanju množičnih pojavov združujemo sorodne pojave, ki jih nameravamo proučevati, v statistično množico ali populacijo. Posamezen pojav v populaciji imenujemo statistična enota ali na kratko enota. V gozdarstvu je najobičajnejša populacija sestoj, enote te populacije pa morejo biti posamezna drevesa v proučevanem sestoju.

Populacija je določena z opredeljujočimi pogoji, ki predpisujejo, kateri pojav je enota proučevane populacije in kateri ni. Tako n. pr. so opredeljujoči pogoji za proučevani sestoj: enote populacije so vsa drevesa, ki rastejo v sestojih gozdno gospodarskega območja N.N. na dan 30. junija 1960. S temi opredeljujočimi pogoji je točno določeno, iz katerih dreves je sestavljena populacija. Tako ne spada v populacijo drevo, ki je bilo posekano 29. junija 1960, niti drevesa, ki rastejo na negozdnih površinah gozdnega gospodarskega območja N.N.

Čeprav so posamezne enote v smislu opredeljujočih pogojev med seboj sorodne, se med seboj razlikujejo v najrazličnejših značilnostih. V konkretnem proučevanju niti ne moremo niti nimamo interesa proučevati vse značilnosti enot. Značilnosti enot, ki so predmet konkretnega proučevanja, imenujemo statistične znake ali na kratko znake. Tako so v našem primeru znaki: vrsta drevesa, premer drevesa v prni višini, višina drevesa, velikost krošnje itd., če proučujemo lesno maso, ali vrsta drevesa, starost, višina, jakost napadenosti z določenim insektom itd., če proučujemo okuženost gozda itd.

Statistični znaki imajo v principu za vsako enoto različno vrednost znaka. Pravimo, da statistični znaki varirajo. Variiranje znakov je eden izmed osnovnih lastnosti znakov. V našem primeru so drevesa različne vrste, premeri dreves so različni, enako je z višino dreves in razsežnostjo krošnje, starostjo dreves pri neenodobnem sestoju, jakostjo napadenosti z insekti itd. Značilnosti enot, ki ne varirajo, so torej za vse enote iste. Te niso predmet statističnega proučevanja. Tako statistično niti ni možno niti nima smisla proučevati starost pri enodobnem sestoju, ali vrsto gozda pri čistem smrekovem sestoju, ker sta starost in vrsta za vsa drevesa ista.

Kot ima vsaka enota svoje značilnosti, jih ima tudi populacija. Pravimo jim parametri populacije. Parametri so izvedeni oziroma odvisni od vrednosti znakov enot. Tako je parameter število dreves v gozdu, število dreves dane vrste, skupen volumen vseh dreves v sestoji, povprečen premer dreves v sestoji, odstotek okuženih dreves, mere variabilnosti premera itd. Ugotavljanje in izračunavanje parametrov za populacije je osnovni cilj opisovanja populacij in eden izmed obširnih in važnih problemov statistike.

Statistične enote

1.3

Po zgornjem splošnem uvodu smatramo za statistično enoto vsak pojav, ki v času in prostoru množično nastopa in je predmet statističnega proučevanja. Po tej opredelitvi morejo biti statistične enote v gozdarstvu najrazličnejši pojavi. Najpogostejša enota je drevo. Enota pa more biti tudi veja, list ali izvrtek na drevesu, posamezen gozdni škodljivec, gozdna parcela ali gozdno gospodarstvo, posamezen poskus trdnosti lesa, če proučujemo lastnosti lesa, posek drevesa, odprema hloda iz gozda itd.

Formalno je važna delitev enot na: a) realne enote in b) dogodke. Realne enote so n.pr. drevo, gozdni škodljivec, parcela, gospodarstvo itd., ker v času in prostoru dalj časa obstajajo. Dogodki kot enote pa so posek drevesa, spravilo hlo-dov iz gozda itd., ker se v času dogode. Dogodek se dogodi v trenutku ali v razmeroma krajšem časovnem razdobju. Kot bomo spoznali kasneje, je ta delitev enot važna zaradi časovne opredelitve populacij, ki je različna za populacije realnih enot in za populacije dogodkov.

Statistične enote so bodisi: a) enostavne enote ali b) agregati. Tako je enostavna enota n.pr. drevo, če proučujemo gozd, ali list drevesa, če raziskujemo lastnosti listov. Agregati pa so skupinice enostavnih enot. Tako so agregati skupnost dreves na eni parcelici, enem gospodarstvu itd.

1.4

Statistični znaki dajo enotam vsebino. Znaki so tiste značilnosti enot, ki so predmet konkretnega proučevanja. Zato more neka značilnost pri enem proučevanju biti znak opazovanja, pri drugem pa ne. Tako širina branike pri določanju volumna v nekem sestoju ni znak, medtem ko je pri raziskavi kvalitete lesa širina branike važen znak.

1.5

Vrste statističnih znakov. Vsebinsko delimo statistične znake v te tri skupine: a) krajevne, b) časovne in c) stvarne. Stvarne znake pa naprej delimo v atributivne in numerične.

Krajevni so vsi znaki, ki določajo kraj, kjer se enota nahaja, ali kjer se je zgodil za enoto pomemben dogodek. Tako je krajevni znak mesto, kjer raste drevo, kraj, kjer se nahaja parcela ali gospodarstvo, ki ga proučujemo itd.

Časovni znaki so znaki, ki so v zvezi s časom, ki je kakorkoli pomemben za proučevano enoto. Tako je časovni znak čas, ko je bila vsajena sadika, čas poseka itd.

Vsi drugi znaki so stvarni. Glede na to, kako izražamo vrednosti znakov, stvarne znake delimo na: atributivne in numerične.

Atributivni znaki so vsi oni stvarni znaki, za katere vrednosti izražamo z besedami oziroma opisno. Tako je atributiven znak n.pr. okuženost z insekti, če vrednosti znaka izražamo z: okužen, neokužen. Enako je vrsta drevesa atributiven znak, ker posamezne vrste dreves označujemo z besedami: bukev, javor, smreka, macesen, bor itd. Atributiven znak je tudi vrsta tal, na katerem raste drevo, redka ali gosta poraslost v okolici drevesa itd.

Za numerične znake izražamo vrednosti številčno. Po tem, katere vrednosti na danem razmaku številčne premice more zavzeti numeričen znak, imamo: a) nezvezne znake, ki morejo zavzeti samo nekatere, običajno cele vrednosti in b) zvezne znake, ki morejo zavzeti vse vrednosti na danem razmaku. Tako

je nezvezen numeričen znak n.pr. število novih odganjkov v enem letu, število vej na drevesu, število dreves na parceli, število storžev na smreki itd. Zvezni znaki pa so n.pr. volumen drevesa; trdnost, merjema z utisom pri določeni sili, premer drevesa, specifična teža lesa itd. Vrednosti prvih so samo cela števila, ker drevo ne more imeti 25,1 vej, parcela ne 12 in $1/4$ dreves itd. Pač pa je volumen dreves lahko teoretično vsaka vrednost v določenih mejah.

1.6

Posebna odlika numeričnih znakov je, da se dajo enote populacije urediti po nedvoumnem vrstnem redu po vrednosti numeričnih znakov. To dopušča za časovne in numerične znake poglobljeno analizo. Te lastnosti krajevni in atributivni znaki nimajo. Vendar moremo tudi vrednosti nekaterih atributivnih znakov urediti po velikosti. Tako moremo opredeliti okuženost z vrednostmi: neokužen, malenkostno okužen, srednje okužen, precej okužen, popolnoma okužen. Enako kvaliteto lesa označujemo z: prav slaba, slaba, dobra, prav dobra in odlična. Ozadje teh atributivnih znakov je vedno nek numeričen znak, ki ga včasih niti ne moremo določiti, včasih pa ni potrebe, da ga numerično izrazimo. Tako je pri znaku okuženost osnova število insektov ali odstotek vej, ki je napaden, v osnovi kvalitete morda trdnost lesa, ki se da numerično izraziti, ali enakomernost branik, ki se da tudi na en ali drug način numerično izraziti. Za atributivne znake te vrste moremo včasih uporabiti nekatere metode analize, ki so sicer uporabne samo za numerične znake.

1.7

Za analizo je posebno važna delitev znakov na:

a) faktorialne in b) rezultatívne. Faktorialni znaki izražajo jakost faktorjev, ki vplivajo na določeno značilnost, ki je rezultat teh faktorjev. Rezultat delovanja faktorjev pa je izražen z rezultatívnim znakom. Tako so faktorialni znaki n.pr. kvaliteta tal, starost drevesa, povprečna letna temperatura, količina letnih padavin, relativna višina drevesa itd., rezultatíven znak, ki je izraz delovanja vseh teh faktorjev, pa je

širina branike v določenem letu, prirast itd. Običajno več bistvenih faktorjev vpliva na velikost rezultativnega znaka. Jasno je, da je nemogoče naenkrat vzeti v proučevanje vse faktorje, ki vplivajo na rezultativen znak, bodisi ker jih je preveč ali pa je njih delovanje tako kompleksno, da ne moremo vse upoštevati.

1.8.

Po drugem načelu delimo znake na: a) ekstenzivne in b) intenzivne. Ekstenzivni znaki nakazujejo količino in moremo vrednosti ekstenzivnih znakov seštevati v smiselno količino (n.pr. volumne posameznih dreves v volumen sestoja). Intenzivni znaki pa nakazujejo kakovost. Tako je n.pr. intenzivni znak specifična teža lesa za posamezno drevo, povprečna širina branike itd. Vsote vrednosti intenzivnih znakov nimajo vsebinske osnove.

1.9

Variiranje statističnih znakov. Vsak znak ima določeno število vrednosti. Najmanjše število vrednosti znaka je dva, ker značilnost, ki je za vse enote ena in ista, ni statistični znak, temveč kvečjemu opredeljujoč pogoj. Največje število vseh možnih vrednosti je neomejeno. Vsi zvezni znaki imajo teoretično neomejeno število možnih vrednosti, ker morejo zavzeti vse možne vrednosti na danem razmaku. Tako ima volumen dreves teoretično neomejeno možnih vrednosti, isto je z višino drevesa, s specifično težo lesa itd. Pojav, da ima teoretično vsaka enota drugo vrednost določenega znaka, imenujemo variabilnost znakov. To, da ima več enot v populaciji eno in isto vrednost za dan znak, ni v nasprotju z definicijo variiranja. Bistvo variiranja znaka je v tem, da nimajo vse enote populacije iste značilnosti, ne pa v tem, da bi morale vse enote imeti med seboj različne vrednosti danega znaka.

Variabilnost vrednosti znakov je ena izmed osnovnih značilnosti množičnih pojavov. Zato se statistika ukvarja predvsem z variabilnostjo pojavov v najrazličnejših oblikah in z najrazličnejšimi metodami. Posebno podrobno moremo proučevati variabilnost numeričnih znakov. Zato je tudi največ in najbolj

podrobnih statističnih metod, ki se bavijo z analizo numeričnih znakov.

1.10

Pojmu znaka in vrednosti znakov ustreza v matematiki pojem spremenljivke. Kot ima posamezen znak različne vrednosti, tako tudi spremenljivka zavzema različne vrednosti. Faktorialnemu znaku ustreza pojem neodvisne, rezultativnemu znaku pa pojem odvisne spremenljivke. Zaradi te ozke sorodnosti med statističnimi znaki in spremenljivkami v matematiki, ki se izkaže predvsem pri proučevanju numeričnih znakov, privzamemo iz matematike simbole za spremenljivke: x , y , z , u , w , v itd. in z njimi zaznamujemo tudi statistične znake.

Populacije

1.11

Populacija je skupnost istovrstnih pojavov, ki so opredeljeni z opredeljujočimi pogoji. Pojavi, ki so enote dane populacije, morajo zadoščati tem opredeljujočim pogojem. Populacije moramo opredeliti krajevno, časovno in stvarno. Tako je populacija: drevesa na gozdnih površinah v občini A na dan 30. junija 1960 opredeljena: krajevno (območje občine A), časovno (30. junija 1960 ob polnoči) in stvarno (vsa drevesa na površinah, ki jih v smislu definicije smatramo kot gozdne površine).

Sestoj, za katerega vzamemo, da je sestavljen iz posameznih dreves, je v gozdarstvu najpogostejša populacija. To populacijo proučujemo v različnih oblikah.

1.12

Sestoj je sestavljen iz končnega števila enot - dreves. Take populacije imenujemo končne populacije. Končna populacija je na primer tudi populacija gospodarstev, ki imajo gozd v Sloveniji 31. decembra 1960. Ta populacija je končna, ker je na ta datum število gospodarstev, ki imajo gozd, preštevna množica.

Končna populacija je tudi gozdna površina, če jo razdelimo v enote tako, da jo z vodoravnimi in navpičnimi pasovi razdelimo v kvadrate 25 m x 25 m, ki so enote populacije. Znaki teh enot so n.pr.: število dreves določene vrste na elementarnih kvadratnih površinah, sestava tal, volumen vseh dreves na osnovnih površinah, število okuženih dreves itd.

Če pa vzamemo, da je enota površine na gozdni površini vsak kvadrat 25 m x 25 m, ne glede na način, kako kvadrat leži, in se posamezni kvadrati med seboj prekrivajo, dobimo populacijo, ki ima neskončno število enot.

Za zvezne populacije, ki so take narave kot gozdna površina, za katero niso enote vnaprej dane in ločene med seboj, moremo glede na določitev enote sestaviti populacije s končnim ali neskončnim številom enot.

Tudi posamezno drevo moremo smatrati kot zvezno populacijo, ki je sestavljena iz koščkov lesa n.pr. 3 cm x 5 cm x 2 cm volumna. Drevo moremo razžagati na take ploščice, ki jih je v tem primeru končno število. Znaki teh enot so n.pr. teža, trdnost, mesto, kjer je bila ploščica v deblu, število branik itd. Iz drevesa pa dobimo neskončno populacijo koščkov lesa, če predpostavljamo, da moremo koščke lesa odrejene oblike in velikosti izrezati iz debla na katerikoli način.

1.13

Če preskušamo n.pr. vpliv impregnacije lesa na trdnost lesa, sestavimo v ta namen poskus, pri katerem vzamemo dva koščka lesa, ki sta po kvaliteti čim bolj enaka in en košček impregniramo, drugega pa ne. Na koncu poskusa merimo razliko v trdnosti med impregniranim in neimpregniranim koščkom lesa. Takih poskusov napravimo več. Če hočemo rezultate teh poskusov posplošiti, smatramo izvedene poskuse kot nekaj enot iz neskončne, umišljene - hipotetične populacije vseh možnih enakih poskusov pod enakimi pogoji. V raziskovalnem delu pogosto smatramo skupnost poskusov pod enakimi pogoji kot del iz umišljenih - hipotetičnih populacij. To delamo vselej, kadar zaključke poskusov posplošujemo.

Delne populacije. Če drevesom v populaciji: drevesa za sestoj gozdnega gospodarskega območja N.N. konec leta 1960 dodamo nov opredeljujoč pogoj: listavec, iz osnovne populacije izdvojimo novo populacijo: drevesa listavcev v sestojih gozdnega gospodarskega območja N.N. konec leta. Vsa drevesa v tej delni populaciji so enote osnovne populacije, ker zadoščajo vsem opredeljujočim pogojem iz osnovne populacije. V tem smislu so stvarno izvršeni poskusi, ki jih smatramo kot enote hipotetične populacije vseh možnih poskusov pod enakimi pogoji, tudi delna populacija, ker so to oni poskusi iz hipotetične populacije, ki so bili stvarno izvršeni.

Za analizo populacij je posebno pomembno razdeliti osnovno populacijo po danem znaku v več delnih populacij tako, da je vsaka enota osnovne populacije v eni, a v eni sami delni populaciji. Če razdelimo drevesa v sestoju v delne populacije po drevesnih vrstah, dobimo osnovo za proučevanje sestoja po drevesnih vrstah. Če razdelimo drevesa v populaciji po debelini v delne populacije po debelinskih stopnjah, dobimo osnovo za uvid v sestav po debelini dreves itd.

Parametri

Vsak numerični podatek o populaciji je parameter populacije. Ena izmed osnovnih nalog statističnega proučevanja množičnih pojavov je izračunavanje parametrov o populaciji. V nadaljevanju se bomo ukvarjali z elementarnimi parametri, ki jih dobimo z enostavnim preštevanjem ali seštevanjem podatkov za enote populacij. Iz osnovnih podatkov pa najprej podatke analiziramo tako, da izračunavamo različne izvedene parametre kot so: relativna števila, srednje vrednosti, mere variacije, mere korelacije itd.

Zaokroževanje podatkov

2.1

Numerične podatke, kot so n.pr. premer, višina, volumen in starost dreves merimo samo do določene stopnje natančnosti. Te podatke zaokrožujemo do natančnosti, ki je potrebna za praktično uporabo. Tako zaokrožujemo premere dreves na centimetre, višino dreves na decimetre ali metre, starost na leta itd.

Podatke zaokrožujemo na dva načina. Premer dreves zaokrožujemo na centimetre na najbližjo celo vrednost v cm, ali pa na najvišjo celo vrednost v cm. Po prvem načinu zaokrožimo n.pr. premer 32,7 cm na 33 cm, ker je to najbližja cela vrednost in se od prave vrednosti razlikuje le za 0,3 cm. Po drugem načinu pa zaokrožimo premer na 32 cm, ker je 32 cm najvišja vrednost v zaokroženih enotah (cm), ki ni večja kot obravnavan premer. Prednost pri prvem načinu zaokroževanja je, da je v splošnem razlika med pravo in zaokroženo vrednostjo vedno manjša kot polovica enote mere, na katero zaokrožujemo. Izkaže pa se, da je pri nadaljnji obdelavi zaokroženih podatkov bolje, če jih zaokrožujemo na najvišjo celo vrednost, čeprav je razlika med pravo in zaokroženo vrednostjo lahko tudi večja kot pa pri prvem načinu.

Ne zaokrožujemo pa samo numerične znake. Če navajamo čas pogozdovanja, je včasih zadosti, da povemo, v katerem letu je bil sestoj osnovan in ni potrebe, da povemo točno do dneva, kdaj je bilo zasajeno posamezno drevo.

Tudi krajevne podatke zaokrožujemo. Tako običajno ne navajamo točnega kraja, kjer je bilo drevo posekano, temveč zadošča, če navedemo, v katerem sestoj ali odseku je rastlo posekano drevo.

Grupiranje vrednosti znakov

2.2

Zaokroževanje, ki ga poznamo iz vsakdanjega življenja, se ujema z grupiranjem vrednosti znakov, ki ga uporablja-

mo v statistiki. Za znake z velikim številom vrednosti običajno ni potrebno oziroma je nepregledno, da navajamo podatke za vse vrednosti. V takih primerih sorodne vrednosti znakov grupiramo v grupe. Pri krajevnih, časovnih in zveznih numeričnih znakih, ki imajo neomejeno število vseh možnih vrednosti, pa je grupiranje nujno potrebno, ker je število vseh možnih osnovnih vrednosti neomejeno.

Z grupiranjem izgubimo na natančnosti, pridobimo pa na preglednosti, ki je v večini primerov statističnega proučevanja bistvena. Razen tega pa se zakonitosti in tipičnosti pri množičnih pojavih pokažejo šele, če pojav nastopa v velikem številu. To pa dosežemo s tem, da podatke ne navajamo individualno, temveč po grupah.

Načela grupiranja

2.3

Kako izvedemo grupiranje, je odvisno od vrste znaka, ki ga grupiramo, od vsebinskih razlogov in cilja statističnega proučevanja. Po pravilu ni enotnih načel in navodil za sestavljanje grup.

Za vse znake pa veljajo neka splošna načela.

Za vse vrste znakov mora biti grupiranje enolično.

To pomeni, da mora biti grupiranje izvedeno tako, da vsaka osnovna vrednost spada v eno in eno samo grupo. V tej zvezi je poseben problem mejnih vrednosti. Tako je pri časovnem grupiranju treba paziti in določiti, v katero grupo spada na primer pri grupiranju v dneve moment ob koncu enega in začetku drugega dne. Pri geografskih grupah je podoben problem pri odločitvi, kam spadajo meje med posameznimi grupami. Sporno more biti, kam spada drevo, ki raste točno na meji med dvema geografskima grupama.

Vse individualne vrednosti v eni grupi pri grupiranem znaku so izenačene in dobe neko novo grupno vrednost. Tako pri zaokroževanju štejemo, da imajo vsi premeri v grupi od 32,5 cm do 33,5 cm skupno grupno vrednost 33 cm in se ne oziroma na podrobnejše vrednosti.

Pri časovnih grupah vse trenutke enega dne združujemo v dneve in je za vse trenutke enega dneva grupna vrednost

datum itd. Enako je ime občine grupna vrednost za vse kraje občine, če grupiramo kraje v grupe po občinah.

Grupe vsakega znaka moremo naprej grupirati v grupe višjega reda, enako kot grupiramo individualne vrednosti. Grupe premerov po en centimeter združujemo v debelinske stopnje, ki so sestavljene iz petih grup po en centimeter. Tako dobimo drugo debelinsko stopnjo, če združimo premere 5 cm, 6 cm, 7 cm, 8 cm, 9 cm v novo grupo - drugo debelinsko stopnjo. Enako časovne grupe - dneve združujemo v grupe višnje stopnje - tedne, mesece, leta; geografske grupe - občine v okraje itd.

Grupe imajo torej iste lastnosti kot vrednosti znakov. Imajo svoje ime, ki jih označuje, in jih moremo grupirati kot osnovne vrednosti znaka.

Posamezne znake moremo grupirati po različnih načelih. Za vsak znak moremo sestaviti različne grupe, ker je za vsak znak več načel, po katerih izvedemo grupiranje. Omenili smo že, da je načelo grupiranja sorodnost med vrednostmi, ki so v eni grupi. Načelo sorodnosti pa je lahko različno. Tako zemljišča grupiramo po lastniškem ali vsebinskem načelu. Po prvem združujemo v grupe vsa zemljišča, ki so last istega lastnika, po drugem pa zemljišča, ki so si vsebinsko podobna, zasajena z isto vrsto drevja itd.

Enako imamo upravno teritorialno grupacijo, če je načelo geografskega grupiranja pripadnost v isto upravno teritorialno grupo, ali rajone, če je načelo geografskega grupiranja vsebinsko.

Grupiranje numeričnih znakov

2.4

Pri numeričnih znakih je merilo sorodnosti dveh vrednosti absolutna razlika med njima. Grupam za numerične znake pravimo razredi. Vsak razred je od drugega razreda razmejen z mejo razreda tako, da nobena vrednost v razredu, ki je pod mejo, ni večja, nobena vrednost v razredu, ki je nad njo, pa ni manjša kot meja razreda. Samo mejo štejemo bodisi v spodnji ali v zgornji razred. Tako je 25 cm meja med peto in šesto debelinsko stopnjo, ker vključuje peta debelinska stopnja premere

med 20 cm do 25 cm, šesta pa premere od 25 cm do 30 cm. S predpisom določimo ali spada meja (25 cm) v spodnji ali zgornji razred. To v grupaciji tudi naznačimo. Če vključimo meje razredov v spodnje razrede, debelinske stopnje označimo z:

1. deb. stop. do 5 cm
2. deb. stop. nad 5 cm do 10 cm
3. deb. stop. nad 10 cm do 15 cm
4. deb. stop. nad 15 cm do 20 cm
- itd.

Da so pa meje razredov vključene v zgornje razrede, pa je razvidno iz naslednje grupacije:

1. deb. stop. do pod 5 cm
2. deb. stop. 5 cm do pod 10 cm
3. deb. stop. 10 cm do pod 15 cm
4. deb. stop. 15 cm do pod 20 cm itd.

Ker iz grupacije:

1. deb. stop. - 5 cm
2. deb. stop. 5 cm - 10 cm
3. deb. stop. 10 cm - 15 cm
4. deb. stop. 15 cm - 20 cm
- itd.

ni razvidno, v katere razrede spadajo meje razredov, ta grupacija ni enolična in je zato pomanjkljiva.

Če zaokrožujemo premere na najvišje cele vrednosti, moremo grupacijo v debelinske stopnje pisati tudi takole:

1. deb. stop. 0 cm - 4 cm
2. deb. stop. 5 cm - 9 cm
3. deb. stop. 10 cm - 14 cm
4. deb. stop. 15 cm - 19 cm
- itd.

Da grupacija ni enolična in da ne spada vsaka izmed vrednosti v en in en sam razred, izgleda le navidezno. Premer 14,7 cm spada v 3. deb. stopnjo, ker zaokrožen premer 14 cm vključuje vse premere do pod 15 cm. Na prvi pogled pa izgleda, da se spodnja in zgornja meja dveh zaporednih razredov ne prekrivata.

Iz navedenega primera vidimo, da zaokroževanje na najbližjo celo vrednost ni primerno. Pri tem načinu zaokroževanja so namreč meje med posameznimi razredi 4,5 cm, 9,5 cm,

14,5 cm, 19,5 cm in zato iz njih ne moremo sestaviti pravih debelinskih stopenj, za katere so meje razredov 5 cm, 10 cm, 15 cm, 20 cm itd.

Vsak razred ima širino razreda, ki je razlika med zgornjo in spodnjo mejo razreda. Če zaznamujemo na splošno katerikoli razred z indeksom k , z $x_{k,min}$ spodnjo, z $x_{k,max}$ zgornjo mejo razreda, z i_k pa širino razreda, je po zgornji definiciji

$$i_k = x_{k,max} - x_{k,min} \quad (2.1)$$

Kot smo omenili, ima vsaka grupa neko grupno vrednost, ki reprezentira vse vrednosti razreda. Ta grupna vrednost je pri numeričnih znakih sredina razreda, ki jo zaznamujemo z x_k , izračunamo pa po obrazcu

$$x_k = \frac{x_{k,min} + x_{k,max}}{2} \quad (2.2)$$

Pri analizah vzamemo v vsakem primeru približek, da so vse vrednosti v razredu enake grupni vrednosti - sredini razreda, ker se nobena vrednost v razredu ne razlikuje od nje za več kot za polovico širine razreda.

2.5

Nezvezne numerične znake grupiramo po istih načelih kot zvezne. Vendar je glede na to, da so za nezvezne numerične znake vrednosti izolirane, nekaj tehničnih razlik v grupiranju.

Če vzamemo kot primer za nezvezni znak število dreves na površinah 10 a v določenem gozdnem sestoju, morejo biti grupe naslednje: 0 - 9 dreves, 10 - 19 dreves, 20 - 39 dreves, 40 in več dreves. Nepravilna je grupacija 0 - 10, 10 - 20, 20 - 40, 40 in več. Ta grupacija ni enolična, ker so meje vključene v po dva razreda.

Tudi za nezvezne razrede moremo določiti meje razredov, širine razredov in sredine razredov po istih obrazcih kot za zvezne znake, če teoretično predpostavljamo, da vsaka individualna vrednost za nezvezni znak predstavlja grupo, ki obsega razmak, ki je širok 1, individualna vrednost pa je sredina tega enotinega razmaka. Tako n.pr. vrednost 9 dreves teore-

tično predstavlja vrednosti v razmaku od 8,5 do 9,5, čeprav vemo, da je število dreves nezvezen numeričen znak. Če to upoštevamo, so pri zgornjem primeru teoretične meje 9,5, 19,5, 29,5, 39,5 dreves. Ker je n.pr. v drugem razredu spodnja meja 9,5, zgornja meja pa 19,5, izračunamo, da je širina razreda $i_2 = 19,5 - 9,5 = 10$, sredina razreda pa $x_2 = \frac{9,5 + 19,5}{2} =$

14,5. Tako izenačimo postopek grupiranja zveznih in nezveznih znakov. Zveznim vrednostim v določenem razmaku pripišemo sredino razreda, ki je reprezentant vrednosti v razmaku, nezveznim vrednostim pa pripišemo enotni razmak, ki predstavlja razmak, na katerega se nanaša posamezna nezvezna vrednost.

2.6

Kot vidimo iz zgornje grupacije dreves, morejo biti širine posameznih razredov glede na pojav, ki ga proučujemo in glede na cilj proučevanja, različne. Zadnji razred 40 in več dreves pa ima celo samo spodnjo mejo. Za take razrede, ki imajo samo spodnjo ali samo zgornjo mejo, ne moremo določiti niti sredine niti širine razreda. Imenujemo jih odprte razrede, tvorimo jih pa takrat, kadar pričakujemo, da je le nekaj enot, za katere je podatek nad ali pod določeno vrednostjo, ti podatki pa zelo variirajo.

Za formalno statistično analizo je najprimerneje, če so širine vseh razredov enake, vsi razredi pa zaprti. Iz vsebinskih razlogov pa to ni vedno najboljša.

2.7

Grupiranje numeričnih znakov. Časovne znake grupiramo v naravne časovne razmake, ure, dneve, tedne, mesece, leta, pet ali desetletja, čeprav je dolžina posameznih izmed njih različna (n.pr. pri mesecih). Vedno pa ne združujemo zaporedne časovne razmake v večje grupe. Tako n.pr. pri sezonskem proučevanju pojavov združujemo v eno grupo iste mesece za več let. Tako dobimo grupo vseh januarjev, vseh februarjev, vseh marcev itd. za več zaporednih let in proučujemo, kakšne so sezonske razlike n.pr. v sečnji. Če pa hočemo proučevati časovni razvoj sečnje v daljšem razdobju, pa združimo vse mesece posameznih

let v letna razdobja. Grupiranje je torej ozko povezano s ciljem analize.

Kot smo že omenili, se pri krajevnih grupacijah navadno oslonimo na teritorialne razdelitve, ki so izvedene v druge - običajno administrativne namene. Najmanjše teritorialne grupe, ki pridejo v poštev v gozdarstvu, so parcele, sestoji oziroma odseki. Te združujemo po različnih načelih dalje v višje grupe: oddelke, gospodarske razrede, gospodarske enote itd. Razen zgornjega grupiranja pa poznamo tudi vsebinsko krajevno grupiranje v rajone. V rajone združujemo vse kraje z istimi značilnostmi pojava. Ker je rajonizacija težaven posel, dostikrat združujemo v rajone manjše administrativne enote po pretežnosti.

Problematika grupiranja atributivnih znakov je specifična. Načelo sorodnosti je dano s kako lastnostjo, ki jo posamezna vrednost atributivnega znaka izraža. Tako moremo posamezne vrste dreves grupirati v listavce ali drevesa z mehkim in trdim lesom in podobno.

3. ELEMENTARNI PARAMETRI

3.1

Število enot in vsote. Najenostavnejši parametri o statistični populaciji so absolutni podatki, katere dobimo s preštevanjem enot v populaciji in s seštevanjem vrednosti za numerične znake. Osnovni parameter je število enot v populaciji, ki ga dobimo, če preštejemo, koliko enot sestavlja populacijo. Tako dobimo število dreves v sestoji, če je populacija, ki jo proučujemo, sestavljena iz dreves v sestoji. Število enot v populaciji dogovorno zaznamujemo vedno z N (numerus).

Preprost parameter o populaciji, ki ga dobimo s seštevanjem vrednosti znaka, je n.pr. lesna zaloga v sestoji. Tega dobimo, če volumen posameznih dreves (vrednosti znaka "volumen drevesa") seštejemo. Podobno dobimo tudi skupno temeljnico, če seštejemo temeljnice posameznih dreves v sestoji. Vsote podatkov v populaciji zaznamujemo z velikimi črkami X, Y, Z, U, V, analogno kot zaznamujemo z x, y, z, u, v znake. Glede na to velja zveza

$$Y = \sum_{i=1}^N y_i \quad (3.1)$$

$\sum_{i=1}^N$ = oznaka za seštevanje od 1 do N; y_i = individualna vrednost

3.2

Statistične vrste. Vendar ti parametri, čeprav so pomembni, povedo razmeroma malo o populaciji. Če pa preštevamo enote in seštevamo podatke po grupah določenega znaka, dobimo zelo dober uvid v sestav in osnovo za analizo populacije.

Za populacijo - "posekana drevesa v letu 1958 v FLRJ", je skupna posekana bruto lesna masa 13.296 tisoč m³. Ta podatek sam zase je pomemben, vendar dá razmeroma malo informacij o posekani lesni masi. Če pa seštejemo posekano lesno maso po drevesnih vrstah, dobimo uvid v sestav posekanega lesa po vrstah. Tako dobimo niz podatkov-parametrov, ki kompleksno prikazujejo sestav posekane lesne mase po vrstah lesa.

Tabela 3.1 Posekana bruto lesna masa po vrsti lesa
v letu 1958 v FLRJ v tisoč m³
(Vir SG 1959)

S k u p n o	13 296
Listavci skupno	9 173
od tega: hrast	1 662
bukev	5 859
ostali trdi	1 084
topol	213
ostali mehki	354
Iglavci skupno	4 123
od tega: jelka in smreka	3 787
bor	329
ostali	7

Tak niz istovrstnih podatkov, od katerih se vsak nanaša na eno izmed vrednosti ali na grupo vrednosti določenega znaka, imenujemo statistično vrsto. Ker je v našem primeru posekana lesna masa razdeljena po atributivnem znaku - vrsta dreves, imenujemo to statistično vrsto - atributivno statistično vrsto.

3.3

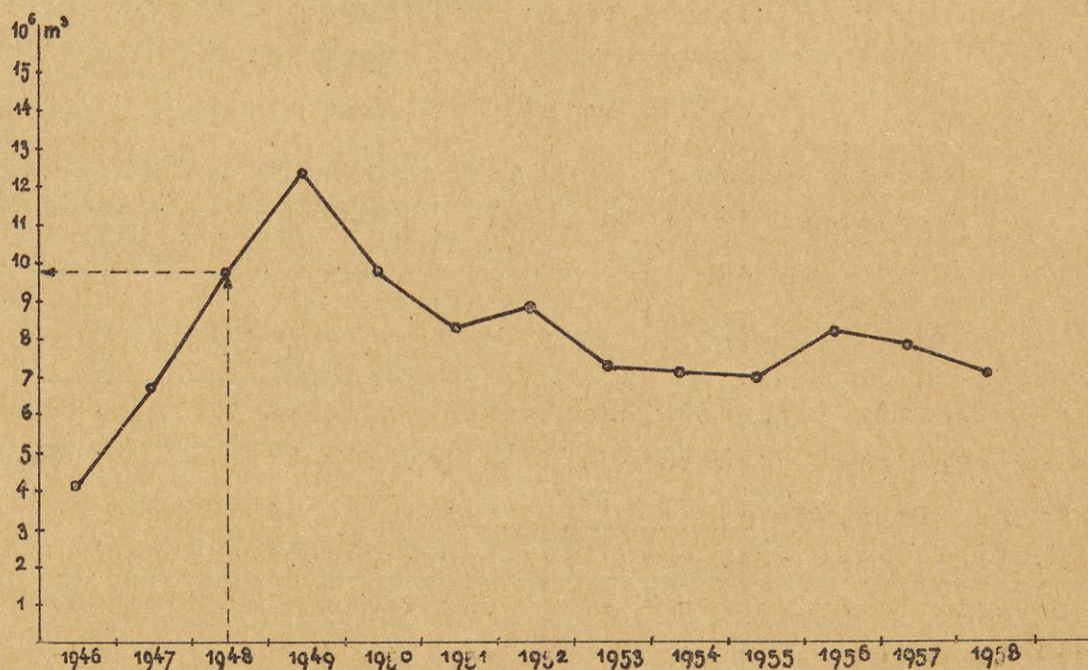
Časovne vrste. Analogno je časovna vrsta niz istovrstnih podatkov, od katerih se vsak nanaša na določen zaporeden časovni trenutek ali razmak.

Primer za časovno vrsto je izvoz gozdnih sortimentov na glavna skladišča po letih v FLRJ.

Tabela 3.2 Izvoz gozdnih sortimentov na glavna skladišča v FLRJ v razdobju 1946-1958 v tisoč m³ (Vir: SG 1959)

Leto	Lesna masa	Leto	Lesna masa
1946	4117	1952	8776
1947	6734	1953	7202
1948	9824	1954	7058
1949	12326	1955	6880
1950	9756	1956	8052
1951	8241	1957	7717
		1958	6998

Iz navedene časovne vrste moremo slediti časovni razvoj izvoza posekane lesne mase v razdobju 1946-1958. Iz nje opazimo izreden porast poseka v razdobju od leta 1946 do leta 1949, ko je posek največji (12326 tisoč m³) in ustaljen posek v naslednjem razdobju 1950-1959, kar je rezultat planiranega poseka. Nazorneje kot iz tabele moremo slediti časoven razvoj poseka iz grafičnega prikaza v sliki 3.1.



Slika 3.1

Izvoz gozdnih sortimentov na glavna skladišča v FLRJ v razdobju 1946-1958.
(Vir: SG 1959)

Z linijskim grafikonom je prikazana časovna vrsta v pravokotnem koordinatnem sistemu tako, da je abscisna os časovna os, na ordinatno os pa nanašamo podatek. V našem primeru je to obseg izvoza. Velikost podatka merimo s skalo, ki daje odnos med enoto mere (m³) in geometrijsko velikostjo - dolžino. Sistem pomožnih črt, ki olajšujejo čitanje grafikona, imenujemo mrežo grafikona. Velikost podatka je prikazana z oddaljenostjo ustrezne točke od abscisne osi. Kako odberemo na ordinatni - količinski skali, kolik je uvoz, pa je nakazano za leto 1948 v sliki. Ker se posek nanaša na celo leto, po pravilu točke nanašamo nad sredine razmakov, ki pomenijo na abscisi leta. V primeru, da v grafikonu prikazujemo časovno vrsto po-

datkov, ki se nanašajo na določene časovne momente, pa rišemo točke točno nad mesto, ki na časovni skali ustreza danemu momentu. Tak primer bi imeli, če bi grafično prikazovali časovno vrsto lesne mase stoječega lesa. Ti podatki so momentnega značaja, ker moremo dati stanje lesne zaloge za stoječi sestoj le za določen datum, oziroma trenutek.

3.4

Kombinirani podatki. Včasih dobimo osnovo za analizo, če preštevamo ali seštevamo podatke po kombinaciji dveh ali več znakov hkrati. Tako dobimo kombinirano tabelo. Primer kombinirane tabele je pregled gozdnega fonda po družbenih sektorjih in kvaliteti gozda v tabeli 3.3.

Tabela 3.3 Lesna masa gozdnega fonda v FLRJ v letu 1958 po družbenih sektorjih in kvaliteti gozda (v milijon m³) (Vir: SG 1959)

Druž. sektor	Kvali- teta	Skupno	G o z d		Grmičevje
			gojen	degradiran	
Skupno		880	771	95	14
Družbeni		700	641	51	8
Privatni		180	130	44	6

Frekvenčne porazdelitve

3.5

Za statistično analizo množičnih pojavov so posebno pomembne statistične vrste, ki za numeričen znak pokažejo, koliko enot ima vrednosti v posameznih razredih. Take statistične vrste imenujemo frekvenčne porazdelitve, ker pokažejo pogostnost pojavljanja vrednosti. Število enot v posameznem razredu imenujemo frekvenca. Dogovorno zaznamujemo v frekvenčni porazdelitvi frekvence z f , oziroma z f_k , kadar hočemo z indeksom k naznačiti, da se frekvenca nanaša na konkretne razrede v frekvenčni porazdelitvi. V frekvenčnih porazdelitvah nimamo pregleda o točnih vrednostih znaka v populaciji, ker pokaže samo koliko enot ima vrednost v posameznih razredih. Pač pa da, če je pravilno sestavljena, kompleksen in nazoren pregled o variriranju vrednosti v populaciji.

3.6

Sestavljanje frekvenčnih porazdelitev. Iz neurejene vrste individualnih podatkov za posamezne enote tehnično

sestavimo frekvenčno porazdelitev po več metodah.

Za manjše in enostavnejše populacije je prikladna metoda črtic. Po tej metodi sestavimo najprej obdelovalno tabelo, v kateri so za razrede predvidena večja polja. Ko po vrsti pregledujemo posamezne vrednosti, za vsak podatek napravimo v ustreznem razredu črtico. To ponavljamo, dokler ne izčrpamo celotne populacije. Na koncu v posameznem razredu preštajemo črtice. Število črtic v vsakem razredu je enako številu enot, ki imajo vrednosti v danem razredu. V praksi uporabljamo več načinov črtanja. Črtice moremo enostavno nizariti drugo poleg druge (//////////). Štetje črtic pa si olajšamo, če sestavljamo grupe po pet črtic tako, da s peto črtico, v vsaki grupi po pet, štiri črtice prečrtamo (////). Uporabljamo pa tudi sistem, pri katerem sestavljamo grupe po deset enot. To dosežemo s štirimi točkami za prve štiri enote. Za nadaljnje štiri enote k tem točkam narišemo stranice kvadrata, diagonali pa dopolnita deveto in deseto enoto .

3.7

Za populacijo premerov za $N=53$ dreves 27-letne marilandske topole so individualni podatki naslednji:

49 51 54 54 61 60 64 68 74 57 54 62 37 51 49 45 51 58 43 56 43
61 45 68 33 51 48 41 53 58 72 54 49 48 55 55 36 64 49 44 41 69
52 58 47 50 54 49 52 57 65 55 47.

V sliki 3.2 je po vseh treh načinih naznačeno, kako sestavimo frekvenčno porazdelitev. Za prvi podatek (49 cm) smo vrisali črtico v razred 45-49 cm, za drugega (51 cm) v razred 50-54 cm itd. do zadnjega podatka (47 cm), za katerega smo vnesli črtico v razred 45-49 cm.

PREMER	ČRTANJE			f_k
30 cm - 34 cm			°	1
35 cm - 39 cm			°	2
40 cm - 44 cm			°°	5
45 cm - 49 cm			°°°	11
50 cm - 54 cm			°°°	13
55 cm - 59 cm			°°	9
60 cm - 64 cm			°°	6
65 cm - 69 cm			°°	4
70 cm - 74 cm			°	2

53=N

Slika 3.2 Sestavljanje frekvenčne porazdelitve po metodi črtic in točk

3.8

Za večje populacije in bolj zamotane obdelave pa metoda črtic ni preveč prikladna. Zamotanost posla in velika možnost napak so hibe, ki govorijo proti uporabi te metode za večje obdelave. V takih primerih je primerneje, da podatke napišemo na individualne obdelovalne listke. Obdelovalne listke sortiramo po grupah in razredih, ustrezno s planom obdelave tako, da posamezne listke polagamo na kupe, ki ustrezajo posameznim grupam. Na koncu sortiranja preštajemo listke v posameznih razredih in dobimo zanje frekvence. Z mehaničnimi sredstvi strojne obdelave ta posel pomembno skrajšamo in mehaniziramo.

3.9

Frekvenčne porazdelitve za premere dreves. Najznačilnejši primer frekvenčnih porazdelitev v gozdarstvu so frekvenčne porazdelitve premerov dreves v prsni višini. Porazdelitve premerov za različne sestoje kažejo tipične oblike, ki so pogojene s tipom sestoja. Zato preglejmo frekvenčne porazdelitve za nekaj najznačilnejših tipov sestojev.

V tabeli 3.4 imamo šest frekvenčnih porazdelitev premerov za različne tipe sestojev. Prva in druga frekvenčna porazdelitev se nanašata na enodobna smrekova sestoja A in B na Pokljuki. Oba sestoja sta raziskovalni ploskvi IGLIS-a in obsegata vsak po en hektar površine. Starost prvega sestoja je 130 do 140 let, starost drugega sestoja pa 120 do 130 let.

Frekvenčna porazdelitev C se nanaša na prebiralni jelov-bukov sestoj na Snežniku (2 ha raziskovalne ploskve IGLIS), porazdelitev D pa prikazuje porazdelitev premerov za dvoetažni gozd površine 1 ha na Otoku, oddelek 13 b na raziskovalni ploskvi FAGV. V prvi etaži je zasajena marilandska topola s podrastom jelše v drugi etaži. Ločena frekvenčna porazdelitev za topolo in jelšo je podana v stolpcih D₁ in D₂.

Tabela 3.4 Frekvenčne porazdelitve za enodobna smrekova sestoja, za prebiralen jelkovo-bukov sestoj in dvoetažni sestoj topole in jelše.

S e s t o j	Enodobna smrekova sestoja		Prebiralni jelovbukov sestoj Snežnik	Dvoetažni sestoj O t o k		
	Pokljuka			Skupaj	topola	jelša
Debelinska stopnja	A	B	C	D	D ₁	D ₂
3 10 cm - 14 cm	-	-	257	25	-	25
4 15 cm - 19 cm	4	5	138	57	2	55
5 20 cm - 24 cm	6	34	111	41	5	36
6 25 cm - 29 cm	55	87	84	15	4	11
7 30 cm - 34 cm	118	93	77	14	13	1
8 35 cm - 39 cm	112	74	72	11	10	1
9 40 cm - 44 cm	109	42	37	13	11	2
10 45 cm - 49 cm	61	23	44	15	15	-
11 50 cm - 54 cm	28	13	35	14	14	-
12 55 cm - 59 cm	9	4	34	9	8	1
13 60 cm - 64 cm	5	2	17	13	13	-
14 65 cm - 69 cm	-	4	10	3	3	-
15 70 cm - 74 cm	-	-	3	2	2	-
16 75 cm - 79 cm	-	-	-	3	3	-
	507	381	919	235	103	132

Grafično prikazovanje frekvenčnih porazdelitev. Osnovne zakonitosti in tipičnosti za posamezne vrste sestojev, ki jih izražajo frekvenčne porazdelitve premerov, nazorneje in lepše kot iz tabele 3.4 proučimo iz grafičnega prikaza posameznih frekvenčnih porazdelitev.

Grafično prikazujemo frekvenčne porazdelitve na dva načina:

- a) s histogrami in
- b) s poligoni.

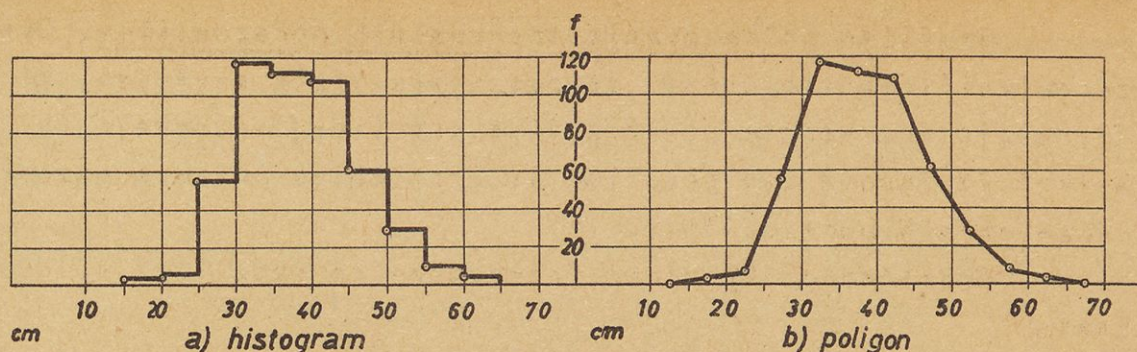
S histogramom prikažemo frekvenčno porazdelitev s stolpci. Za vsak razred je frekvenca ponazorjena s stolpcem, ki je visok v sorazmerju s frekvenco. Skupnost stolpcev za vse razrede da sliko celotne frekvenčne porazdelitve.

Frekvenčni poligon pa dobimo, če za vsak razred v frekvenčni porazdelitvi narišemo nad sredino razreda točko, ki je od abscisne osi oddaljena v sorazmerju s frekvenco, te točke pa med seboj povežemo z daljicami. Poligon nazorneje in pravilneje prikaže stvarno variiranje vrednosti znaka kot histogram.

V sliki 3.3 je za čisti smrekov sestoj A iz tabele 3.4 grafično prikazana frekvenčna porazdelitev premerov s histogramom in s poligonom. Iz slike moremo kompleksno analizirati variacijo premerov v sestoju. Največ premerov je v razredu 30 cm do 34 cm, čim bolj pa se od tega centra oddaljujemo, je število premerov manjše in manjše. Frekvence se od mesta največje gostitve zmanjšujejo na levo in desno razmeroma enakomerno in je slika frekvenčne porazdelitve na obe strani od središča precej simetrična. Taka porazdelitev frekvenc je tipična za enodobne sestoj na izenačenem terenu, kjer so splošni pogoji rasti več ali manj enaki za celo ploskev.

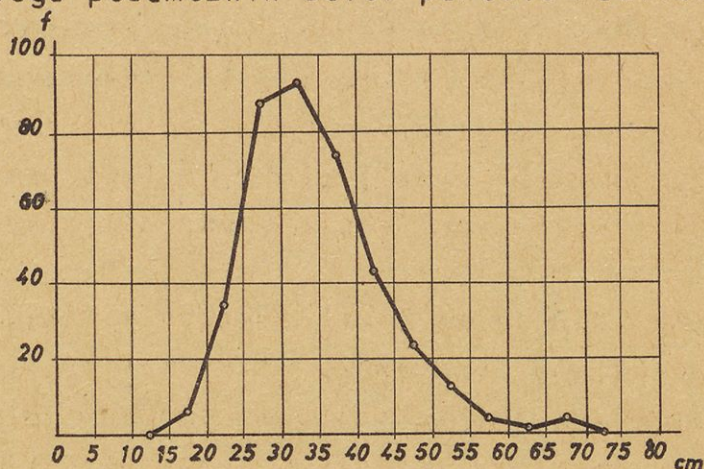
Frekvenčne porazdelitve, ki imajo tako obliko kot jo imamo v primeru sestoja A v sliki 3.3, imenujemo zaradi tipične oblike simetrične in zvonaste, ker pa imajo eno samo mesto gostitve, pa enovršne ali unimodalne.

Frekvenčna porazdelitev za čisti enodobni smrekov sestoj B na Pokljuki v sliki 3.4 kaže še vedno tipično zvonasto obliko, vendar je lahko asimetrična v desno stran, kar na-



Slika 3.3 Grafikon za frekvenčno porazdelitev enodobnega čistega smrekovega sestojaja A iz tabele 3.4

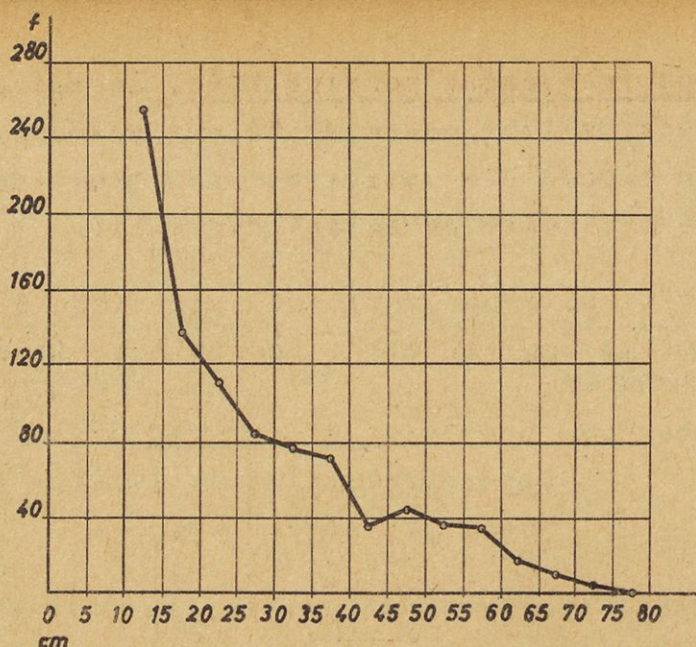
kazuje, da pogoji rasti v tem sestoju niso povsem enakomerni, da je n.pr. lega posameznih delov parcele različna.



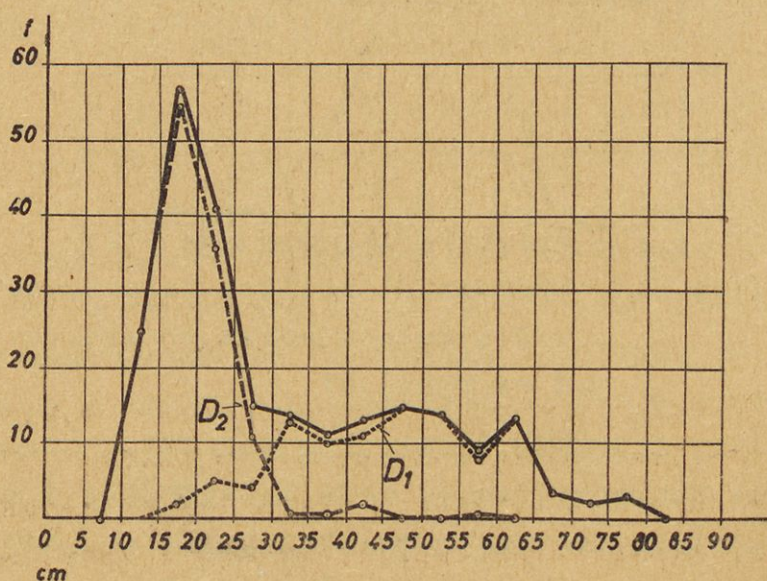
Slika 3.4 Frekvenčni poligon za premere dreves v čistem enodobnem smrekovem sestoju B na Pokljuki

Slika 3.5 kaže tipično porazdelitev premerov za prebiralni gozd. Porazdelitev je zelo asimetrična v desno, ker močno prevladujejo drevesa z majhnim premerom, ker se močnejša drevesa postopoma izsekavajo. Porazdelitev za prebiralni gozd ima tipično obliko črke J, zato take porazdelitve imenujemo kar J-porazdelitve.

Slika 3.6 ponazarja porazdelitev premerov za dvoetažni sestoj Ofok. Kot je nakazano, je sestavljena iz dveh porazdelitev: porazdelitve za nasad marilandske topole (D_1) in podrasta jelše (D_2). Porazdelitvi za posamezni etaži kažeta



Slika 3.5 Frekvenčni poligon za premerje dreves v bukovo-jelovem sestoju na Snežniku



Slika 3.6 Frekvenčni poligoni za dvoetažni sestoj Otok

tipično zvončasto obliko, skupna porazdelitev, ki je vsota frekvenc za obe etaži, pa odstopa od te lastnosti in nakazuje dva vrha (bimodalna porazdelitev), ker je sestavljena iz dveh unimodalnih porazdelitev.

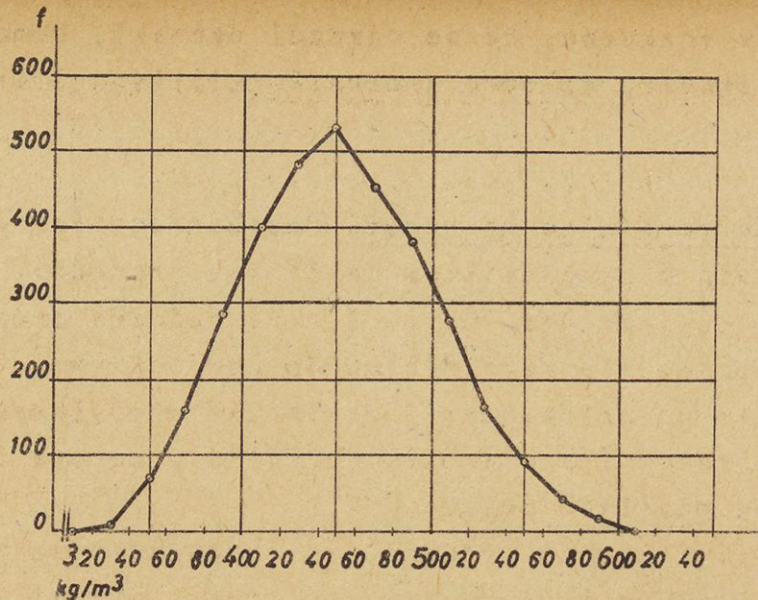
Druge frekvenčne porazdelitve. Zaradi pomembnosti frekvenčnih porazdelitev navedimo še dva primera. Frekvenčna porazdelitev v tabeli 3.5 podaja porazdelitev specifične teže lesa za $N=3376$ preskusov na zeleni duglaziji.

Tabela 3.5 Frekvenčna porazdelitev specifične teže lesa za zeleno duglazijo (Po podatkih IGLIS-a)

Teža kg/m^3	f_k
320 - pod 340	7
340 - pod 360	71
360 - pod 380	163
380 - pod 400	289
400 - pod 420	397
420 - pod 440	484
440 - pod 460	531
460 - pod 480	453
480 - pod 500	381
500 - pod 520	279
520 - pod 540	165
540 - pod 560	94
560 - pod 580	45
580 - pod 600	16
600 - pod 620	1

3376 = N

Frekvenčna porazdelitev lepo pokaže variabilnost specifične teže lesa za zeleno duglazijo. Slika 3.7 nazorno pokaže, da je porazdelitev specifičnih tež za zeleno duglazijo simetrična, zvonasta porazdelitev, da je torej močno približana tipični porazdelitvi za primer, če na pojav razen opredeljujočih pogojev vplivajo le slučajni vplivi. Take porazdelitve imenujemo normalne porazdelitve.



Slika 3.7 Porazdelitev specifičnih tež $N=3376$ preskusov za les za zeleno duglazijo

3.12

Kot zadnji primer navedimo še porazdelitev $N=1256$ kmetijskih gospodarstev v bivšem okraju Novo mesto po gozdni površini, po stanju leta 1956.

Tabela 3.6 Kmetijska gospodarstva v okraju Novo mesto po gozdni površini

Gozdna površina	f_k
0 - pod 1 ha	394
1 ha - pod 2 ha	208
2 ha - pod 4 ha	282
4 ha - pod 7 ha	246
7 ha - pod 10 ha	81
10 ha - pod 15 ha	35
15 ha - pod 20 ha	10

1256 = N

Ta frekvenčna porazdelitev kaže, da je največ gospodarstev, ki imajo malo gozda. Število gospodarstev z večjo površino je vedno manjše. Vendar ta zakonitost iz zgornje frekvenčne porazdelitve ni povsem vidna, ker so širine razredov različne, frekvence v posameznih razredih pa so razen od drugih vplivov odvisne tudi od širine razredov. Podrobnejši uvid

v porazdelitev frekvenc, če so razredi neenaki, bomo dobili v naslednjem odstavku, ko bomo spoznali relativna števila.

3.13

Problem širine razredov. Poseben problem pri sestavljanju frekvenčnih porazdelitev je širina razredov. Pri tem se namreč kosata dva momenta. Če je širina razreda premajhna, so frekvence v posameznih razredih močno pod vplivom slučajnih faktorjev. Zato ne pride dovolj do izraza množičnost pojavljanja. Če pa so razredi preširoki, je slika pregroba in zabriše osnovne težnje gostitve pojava.

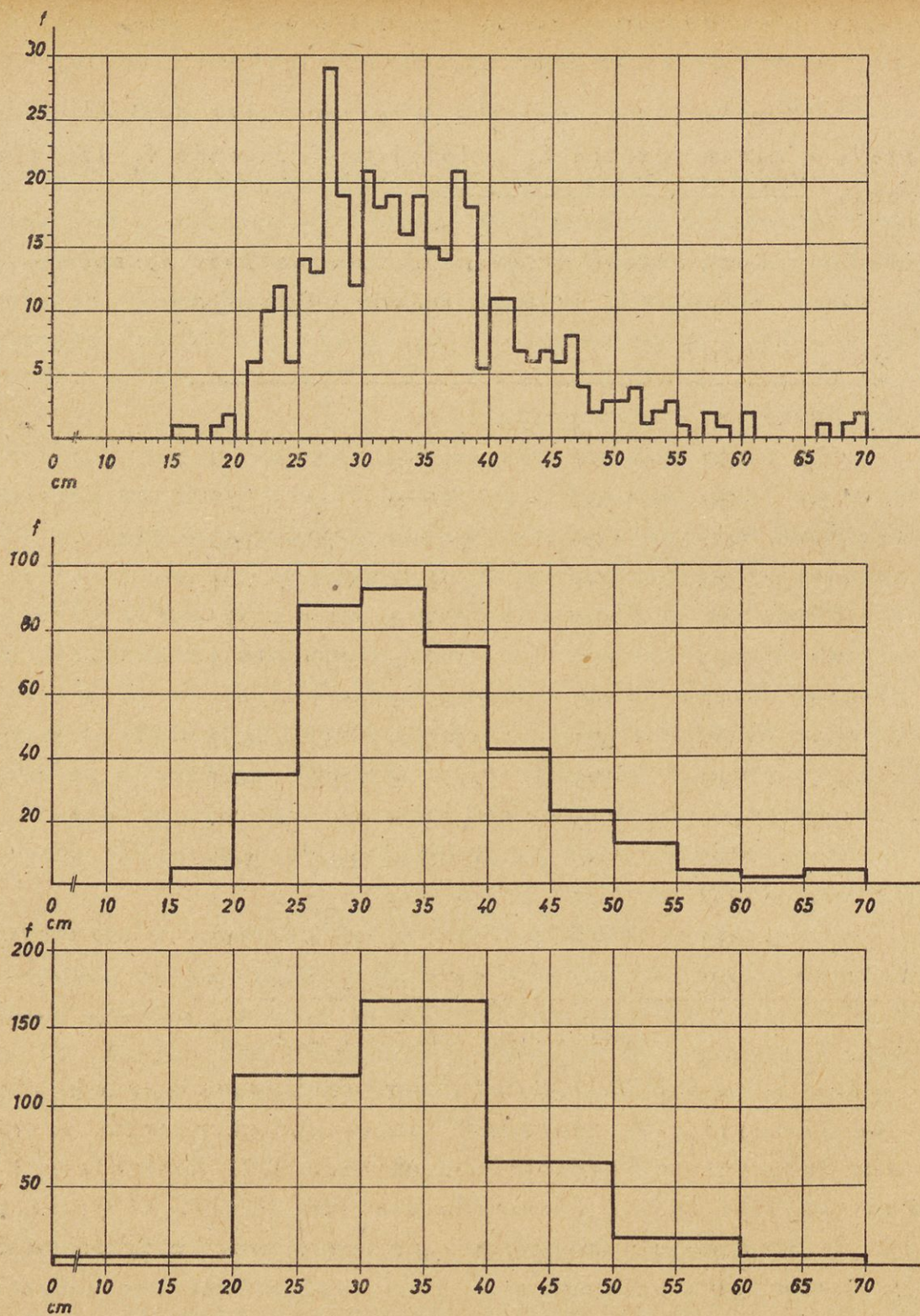
V sliki 3.8 so vrisane tri frekvenčne porazdelitve premerov za isto populacijo premerov dreves v čistem smrekovem sestoju B na Pokljuki. V prvi frekvenčni porazdelitvi so širine razredov 1 cm, v drugi 5 cm, v tretji pa 10 cm. Iz slike nazorno vidimo, da je širina razreda 1 cm v prvi porazdelitvi premajhna. Zaradi slučajnih vzrokov frekvence po razredih preveč nihajo. Širina 10 cm v tretji porazdelitvi pa je prevelika in ne pride do izraza zakonitost gostitve. Najboljša oziroma pravilna je druga frekvenčna porazdelitev, za katero so razredi 5 cm. Razredi so tako široki, da se izravnavajo rezultati individualnih vplivov, niso pa preširoki, tako da pride dobro do izraza zakonitost gostitve premerov.

Določnega pravila o velikosti razredov v posameznih primerih ni. Število razredov je lahko večje, če je populacija obsežnejša, in mora biti manjše, če je populacija manj obsežna. Običajno dobimo primerno sliko o variabilnosti pojava, če vzamemo glede na obseg populacije število razredov med deset in dvajset.

3.14

Kumulativne frekvenčne porazdelitve. Iz frekvenčne porazdelitve dobimo s postopnim seštevanjem frekvenc kumulativno frekvenčno porazdelitev. V tabeli 3.7 je za primer specifične teže lesa za zeleno duglazijo nakazano, kako iz frekvenčne porazdelitve izračunamo kumulativno frekvenčno porazdelitev. Če z F_k zaznamujemo člene v kumulativni frekvenčni porazdelitvi, iz frekvenčne porazdelitve izračunamo člene v kumulativni vrsti po obrazcu

$$F_{k+1} = F_k + f_k \quad (3.2)$$



Slika 3.8 Histogrami populacije premerov za čisti smrekov sestoj B na Pokljuki pri različnih širinah razredov

Kumulativo za naslednji razred $(k+1)$ dobimo, če kumulativi v danem razredu F_k prištejemo frekvenco f_k iz istega razreda.

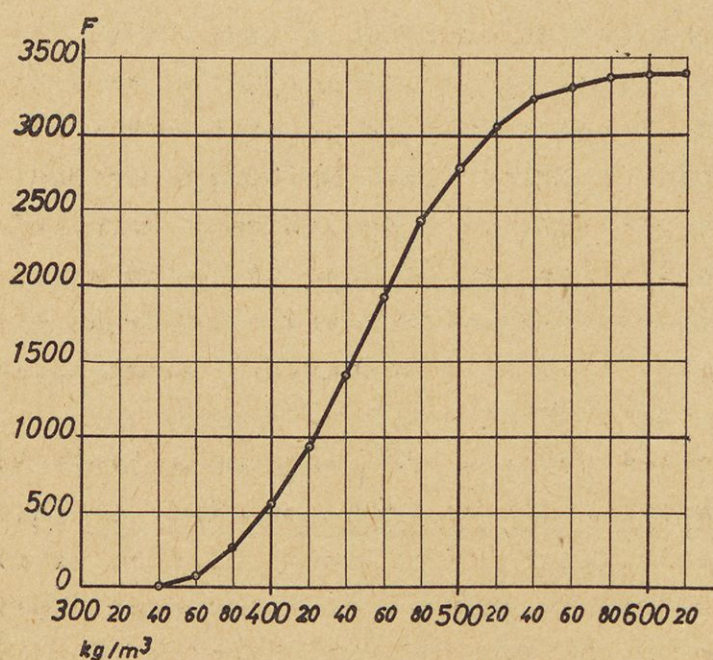
Tabela 3.7 Kumulativna frekvenčna porazdelitev za specifično težo lesa za zeleno duglazijo

Teža v kg/m^3	f_k	F_k
320 - 340	7	0
340 - 360	71	$7 = 0 + 7$
360 - 380	163	$78 = 7 + 71$
380 - 400	289	$241 = 78 + 163$
400 - 420	397	$530 = 241 + 289$
420 - 440	484	$927 = 530 + 397$
440 - 460	531	$1411 = 927 + 484$
460 - 480	453	$1942 = 1411 + 531$
480 - 500	381	$2395 = 1942 + 453$
500 - 520	279	$2776 = 2395 + 381$
520 - 540	165	$3055 = 2776 + 279$
540 - 560	94	$3220 = 3055 + 165$
560 - 580	45	$3314 = 3220 + 94$
580 - 600	16	$3359 = 3314 + 45$
600 - 620	1	$3375 = 3359 + 16$
$3376 = N = 3376 = 3375 + 1$		

Iz tabele 3.7 je razvidno, da v prvem razredu vzamemo, da je kumulativa enaka nič, za vse druge razrede pa izračunamo kumulativne vrednosti po obrazcu 3.2. Kontrola pri izračunavanju je zadnji člen v kumulativni vrsti, ki je že pod črto. Če je kumulativna vrsta izračunana pravilno, je ta člen enak obsegu populacije N .

Posamezni členi v kumulativni frekvenčni porazdelitvi povedo, koliko enot ima vrednosti, ki so manjše kot je spodnja meja ustreznega razreda. Tako na primer osmi člen v kumulativni vrsti $F_8 = 1942$ pomeni, da je bila v 1942 preskuših specifična teža manjša kot je spodnja meja ustreznega razreda, $x_{8,\min} = 460 \text{ kg/m}^3$.

Podobno kot za frekvenčne porazdelitve, da tudi za kumulativne porazdelitve grafičen prikaz kompleksno sliko, kako so kumulativne frekvence odvisne od vrednosti znaka. V grafikonu ponazorimo kumulativno frekvenčno porazdelitev tako, da v pravokotnem koordinatnem sistemu nanesimo nad meje razredov točke, ki so od abscisne osi oddaljene v sorazmerju z vrednostjo ustreznega člena kumulative. Ko zvežemo te točke med seboj, dobimo lomljeno črto, ki nazorno prikazuje kumulativno frekvenčno porazdelitev. Za unimodalne, simetrične in zvonaste frekvenčne porazdelitve ima slika kumulativne frekvenčne porazdelitve značilno obliko velike črke S. To značilnost opazimo tudi na našem primeru v sliki 3.9.



Slika 3.9 Kumulativna frekvenčna porazdelitev za specifično težo leša za zeleno duglazijo

4. RELATIVNA ŠTEVILA

4.1

Vrsta primerjav. Že v prejšnjih odstavkih smo spoznali, da en sam podatek, kljub svoji operativni vrednosti, nima posebnega analitičnega pomena. Zato podatke navajamo v statističnih vrstah, v katerih jih ne obravnavamo posamezno, temveč kot celoto. Tako dobimo vpogled v sestav populacije, če imamo podatke po grupah, uvid v variabilnost, če proučujemo frekvenčno porazdelitev kot celoto, ne pa posamezno frekvenco itd. Niz podatkov proučujemo kompleksno, če podatke med seboj primerjamo. Primerjava je najenostavnejši, obenem pa tudi najpogostejši prijem pri analizi statističnih podatkov. Če primerjamo istovrstne podatke med seboj, ugotovimo: a) da je podatek večji ali manjši od podatka, s katerim ga primerjamo, b) za koliko je podatek večji ali manjši od podatka, s katerim ga primerjamo in c) kolikokrat je podatek večji ali manjši od podatka, s katerim ga primerjamo. Najboljše primerjamo podatke na tretji način, pri katerem z deljenjem primerjanih podatkov dobimo relativna števila, ki so zelo primerno sredstvo za analizo statističnih podatkov. Relativna števila pa niso omejena samo na primerjavo istovrstnih podatkov, ampak moremo z njimi primerjati tudi raznovrstne podatke.

V splošnem poznamo tri vrste relativnih števil. O strukturah ali strukturnih deležih govorimo, če primerjamo podatek za del populacije z istovrstnim podatkom za celotno populacijo. Indekse dobimo, če primerjamo istovrstne podatke za različne populacije, ki pa so med seboj sorodne. O statističnih koeficientih in gostotah pa govorimo, če primerjamo dva raznovrstna podatka za isto populacijo.

4.2

Strukture. Od skupno $N_A = 785$ dreves v sestoju A je okuženih $H_A = 352$ dreves, od skupno $N_B = 1695$ dreves v sestoju B pa $H_B = 542$ dreves. Če hočemo primerjati okuženost v obeh sestojih, nima smisla primerjati absolutnega števila okuženih dreves H_A in H_B v obeh sestojih, ker je velikost sestojev različna. Objektivno merilo okuženosti dobimo, če izračunamo, kak-

šen del dreves v posameznem sestoju je okuženih. Ta podatek dobimo, če število okuženih dreves delimo s skupnim številom dreves v sestoju. Da izrazimo ta delež v odstotkih, dobljeni kvocient pomnožimo s 100. Za pojave, za katere so deleži zelo majhni, izražamo strukturne deleže v promilih. V takih primerih pomnožimo kvocient s 1000. Strukturni delež P^0 pove, koliki del od celote 1 ima dano značilnost, strukturni odstotek $P\%$, koliki del od celote 100, strukturni delež izražen v promilih $P\text{‰}$, pa pove, koliki del od celote 1000 ima dano značilnost. Z obrazci moremo vse tri vrste struktur izraziti takole:

$$P^0 = \frac{H}{N} \quad ; \quad P\% = 100 \frac{H}{N} \quad ; \quad P\text{‰} = 1000 \frac{H}{N} \quad (4.1)$$

Pri tem pomeni: P^0 , $P\%$ in $P\text{‰}$ so strukturni deleži, izraženi z delom od celote, v odstotkih in promilih, H = delni podatek.

V našem primeru nazorno vidimo uporabnost strukturnih deležev. Za prvi sestoj je odstotek okuženih dreves

$$P_A\% = 100 \frac{H_A}{N_A} = 100 \frac{352}{785} = 44,8 \%, \text{ odstotek okuženih dreves}$$

v drugem sestoju pa je $P_B\% = 100 \frac{H_B}{N_B} = 100 \frac{542}{1695} = 32,0 \%$.

Okuženost drugega sestoja je torej znatno nižja kot okuženost v prvem sestoju, kljub temu, da je število okuženih dreves v prvem sestoju manjše.

4.3

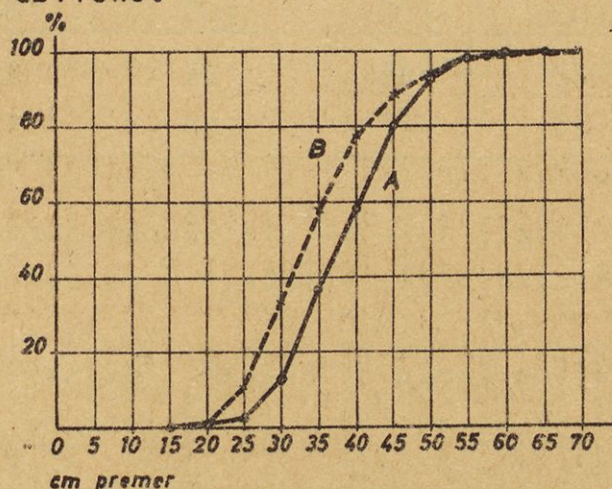
Enostavne strukturne vrste. Strukture izračunavamo vselej, kadar primerjamo sestave več populacij, za katere so podatki za celoto različno veliki. Zato strukture s pridom uporabljamo pri proučevanju frekvenčnih porazdelitev. Zaradi različnega obsega populacij frekvence v istih razredih za različne populacije med seboj niso neposredno primerljive. Pač pa so med seboj primerljivi strukturni deleži - relativne frekvence.

V tabeli 4.1 sta dani frekvenčni porazdelitvi za sestoja A in B na Pokljuki iz tabele 3.4. Zanju sta izračunani frekvenčni porazdelitvi relativnih frekvenc, ki ju moremo zelo dobro med seboj primerjati.

Tabela 4.1 Relativne frekvenčne porazdelitve za čista enodobna smrekova sestoja na Pokljuki

Debelinska stopnja		Absolutne frekvence f		Relativne frekvence f%		Kumulativne relativne frekvence F%	
		S e s t o j		S e s t o j		S e s t o j	
		A	B	A	B	A	B
4	15 cm - 19 cm	4	5	0,8	1,3	0	0
5	20 cm - 24 cm	6	34	1,2	8,9	0,8	1,3
6	25 cm - 29 cm	55	87	10,8	22,8	2,0	10,2
7	30 cm - 34 cm	118	93	23,3	24,4	12,8	33,0
8	35 cm - 39 cm	112	74	22,1	19,4	36,1	57,4
9	40 cm - 44 cm	109	42	21,5	11,0	58,2	76,8
10	45 cm - 49 cm	61	23	12,0	6,1	79,7	87,8
11	50 cm - 54 cm	28	13	5,5	3,4	91,7	93,9
12	55 cm - 59 cm	9	4	1,8	1,1	97,2	97,3
13	60 cm - 64 cm	5	2	1,0	0,5	99,0	98,4
14	65 cm - 69 cm	-	4	-	1,1	100,0	98,9
Skupno		507	381	100,0	100,0	100,0	100,0

Enako izračunamo tudi za relativne frekvence kumulativne frekvenčne porazdelitve. Iz tabele 4.1, še nazorneje pa iz slike 4.1, je razvidno, da je struktura dreves po premeru v obeh sestojih različna.



Slika 4.1 Kumulativni porazdelitvi relativnih frekvenc za čista smrekova sestoja A in B iz tabele 4.1

Kombinirane strukturne vrste. Za kombinirano razdeljene podatke proučujemo sestav s tremi vrstami struktur. V tabeli 3.3 je n.pr. lesna masa v gojenih gozdovih, ki so last družbenega sektorja (641 milj. m³), del treh populacij: a) skupne populacije (880 milj. m³), b) skupne lesne mase v gojenih gozdovih (771 milj. m³) in c) skupne lesne mase v družbenem sektorju (700 milj. m³). Analogno je z vsemi drugimi podatki v kombinacijski tabeli. Zato moremo sestaviti tri tabele struktur.

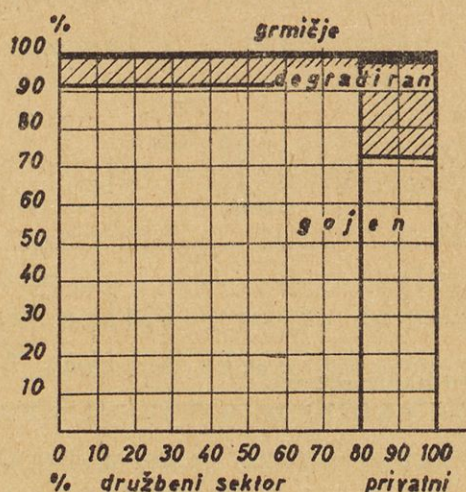
Tabela 4.2 Strukturni sestav lesnega fonda v FLRJ v letu 1958 po družbenih sektorjih in kvaliteti gozda (Vir: tab. 3.3)

	Skupno	G o z d		grmičje
		gojen	degradiran	
A Skupno	100	88	11	2
Družbeni	80	73	6	1
Privatni	20	15	5	1
B Skupno	100	88	11	2
Družbeni	100	92	7	1
Privatni	100	72	25	3
C Skupno	100	100	100	100
Družbeni	80	83	54	58
Privatni	20	17	46	42

V tabeli A je celota skupni fond, v tabeli B posamezni družbeni sektorji, v tabeli C pa ustrezne kvalitete gozda.

Katero izmed navedenih treh načinov uporabimo v konkretnem primeru za analizo, je odvisno od problema in cilja analize. V zgornjem primeru je najbolj poučna tabela B, ki prikazuje kvalitetni sestav po posameznih sektorjih. Kvalitetni sestav v družbenem sektorju je znatno boljši kot v privatnem sektorju.

Strukturni sestav po kombinaciji najlepše prikažemo v kvadratu. Če hočemo ponazoriti spremembe v kvalitetnem sestavu gozda po sektorjih, je najnazorneje, da kvadrat razdelimo v pokončne stolpce v sorazmerju z udeležbo posameznih sektorjev v celotnem fondu, dobljene stolpce pa dalje razdelimo v sorazmerju s kvalitetnim sestavom gozda iz tabele B. Po teh vodilih je izdelan grafikon v sliki 4.2.



Slika 4.2 Struktura lesnega fonda v FLRJ v letu 1958 po sestojih in kvaliteti

Medtem ko je v družbenem sektorju 92 % lesne mase v gojenih gozdovih, jo je v privatnem le 72 %. Obraten redosled pa imajo odstotki v degradiranih gozdovih in odstotki lesne mase iz grmičevja.

Indeksi

4.6

Indeksi s stalno osnovo. Omenili smo že, da dobimo indekse, če primerjamo istovrstne podatke za različne populacije. Indekse moramo torej izračunati za izvoz gozdnih sortimentov na glavna skladišča po letih v FLRJ iz tabele 3.2. Če vzamemo, da je letni izvoz na glavna skladišča samostojna populacija, so podatki o izvoženi letni lesni masi istovrstni

podatki za istovrstne populacije. Indeks $I_{1/0}$ z osnovo ali bazo o dobimo, če podatek, ki ga primerjamo (tekoči podatek Y_1), delimo s podatkom, s katerim ga primerjamo (osnovni ali bazični podatek Y_0), kvocient pa pomnožimo s sto. Z obrazcem moremo to napisati

$$I_{1/0} = 100 \frac{Y_1}{Y_0} \quad (4.2)$$

Indeks izvoza gozdnih sortimentov v letu 1958, če vzamemo za osnovo ali bazo leto 1955, dobimo po zgornjem pravilu, če izvoz v letu 1958 delimo z izvozom v letu 1955, kvocient pa pomnožimo s 100

$$I_{58/55} = 100 \frac{Y_{58}}{Y_{55}} = 100 \frac{6998}{6880} = 101,7$$

4.7

Za dano časovno vrsto običajno izračunamo celo vrsto indeksov z isto - stalno bazo ali osnovo. Tako dobimo za naš primer indeksno vrsto za izvoz s stalno bazo (leto 1939, če vsak člen v časovni vrsti primerjamo z izvozom v bazičnem letu ($Y_{39} = 7159$)).

Indekse izračunavamo največ na eno decimalko. Če so pa razlike med podatki znatne, pa najraje zaokrožujemo indekse na cele, ker so nazornejši. Pri indeksnih vrstah namreč ne gre za pretirano natančne odnose, temveč le za grob vtis o dinamiki pojava.

V tabeli 4.3 je nanizana indeksna časovna vrsta izvoza na glavna skladišča v FLRJ v razdobju 1956-1959 s stalno bazo 1939 = 100.

Tabela 4.3 Indeksna časovna vrsta izvoza lesnih sortimentov na glavna skladišča v FLRJ v razdobju 1946-1958 (leto 1939 = 100)

Leto 1946	1939	1946	1947	1948	1949	1950	1951	1952
10^3 m^3	7159	4117	6734	9824	12326	9756	8241	8776
Indeks	100	58	94	137	172	136	115	123
Leto	1953	1954	1955	1956	1957	1958		
10^3 m^3	7202	7058	6880	8052	7717	6998		
Indeks	101	99	96	112	108	98		

Indeksna vrsta z osnovo predvojnem stanjem 1939 pokaže, da smo v letih 1948-1950, posebno pa v letu 1949

($I_{49/39} = 172$) znatno presegli predvojni izvoz na glavna skladišča, da pa je v razdobju 1953 do 1958 posek na grobo precej enak kot v letu 1939. Indeks $I_{49/39} = 172$ za leto 1949 z osnovo 1939 pomeni, da je bil izvoz iz gozda v letu 1949 za 72 odstotkov večji kot v predvojnem letu 1939.

Vsebinski problem pri izračunavanju indeksov je predvsem izbira baze. Pri tem ni nekega krutega pravila, temveč se ravnamo pri izbiri baze po cilju analize. V splošnem pa izberemo za bazo člen oziroma čas, za katerega smatramo, da je stanje pojava tipično ali normalno. Lažje, kot kaj je tipično oziroma normalno, je povedati, kaj ni normalno. Zato v splošnem ne jemljemo pri primerjavah za daljša razdobja kot osnovo leta tik pred, med ali tik po vojni, leta oziroma čas izjemnih ukrepov (n.pr. v naši vrsti leto 1949), leta slabih letin itd. Dostikrat vzamemo kot bazo večletna povprečja, ker smatramo, da se v povprečju netipičnosti izravnavajo. Tako bi v našem primeru mogli vzeti za bazo povprečen letni izvoz v petletnem razdobju 1954-1958.

4.8

Verižni indeksi. Razen indeksov s stalno osnovo poznamo tudi indekse s premično osnovo. Najobičajnejši indeksi s premično osnovo so verižni indeksi. Te dobimo, če z indeksi primerjamo po dva in dva zaporedna člena. Baza pri verižnih indeksih je vedno predhodni člen. Tako dobimo verižni indeks za leto 1947, če podatek za leto 1947 primerjamo s podatkom za leto 1946, za leto 1948, če podatek za leto 1948 primerjamo s podatkom za leto 1947 itd. Verižni indeks I_k za splošen člen k izračunamo po obrazcu

$$I_k = 100 \frac{Y_k}{Y_{k-1}} \quad (4.3)$$

Vrsta verižnih indeksov pokaže spremembe pojava od člena do člena. Če se pojav ne spreminja, je vrsta verižnih indeksov enaka sto, verižni indeksi pa so med seboj enaki, vendar različni od 100, če se pojav spreminja v eksponentialni funkciji.

Za naš primer je vrsta verižnih indeksov podana v tabeli 4.4.

Tabela 4.4 Izvoz lesnih sortimentov na glavna skladišča v FLRJ v razdobju 1946 - 1958

Leto	1946	1947	1948	1949	1950	1951	1952	1953	1954
Y_k	4117	6734	9824	12326	9756	8241	8776	7202	7058
I_k		163,6	145,9	125,5	79,1	84,5	106,5	82,1	98,0
Leto	1955	1956	1957	1958					
Y_k	6880	8052	7717	6998					
I_k	97,5	117,0	95,8	90,1					

Za leto 1946 verižnega indeksa ne moremo izračunati, ker ne poznamo predhodnega člena. Iz vrste verižnih indeksov vidimo, da izvoz iz gozda prva štiri leta naglo narašča, po tem letu pa odkloni od 100, ki pomeni popolno stagnacijo, niso preveliki. Verižne indekse izračunavamo na eno decimalno, kadar so razlike od člena do člena majhne, ker bi bilo zaokroževanje na cele pregrebo.

4.9

Včasih namesto verižnih indeksov izračunavamo koeficiente dinamike K_k , ki so samo koeficienti zaporednih členov in jih ne pomnožimo s 100. Zveza med njimi je torej enostavna

$$K_k = \frac{Y_k}{Y_{k-1}} \quad ; \quad I_k = 100 K_k \quad (4.4)$$

Če pa od verižnega indeksa odštejemo 100, dobimo temp rasti T_k

$$T_k = I_k - 100 \quad (4.5)$$

ki v odstotkih pokaže, za koliko se je pojav spremenil od člena do člena.

Za leto 1947, za katero je verižni indeks $I_{47}=163,6$, je koeficient dinamike $K_{47} = 1,636$, temp rasti pa $T_{47} = +63,6$.

4.10

Koeficienti. S koeficienti primerjamo med seboj raznovrstne podatke. Primerjana podatka pa morata biti enako opredeljena. Razen tega pa mora biti primerjava vsebinsko upravičena.

Vzemimo kot primer za izračunavanje in uporabo koeficientov vrednost in težo smole, proizvedene v gozdovih v splošno družbenem sektorju v FLRJ v razdobju 1956-1958.

Tabela 4.5 Proizvodnja smole v gozdovih splošno družbenega sektorja v FLRJ v razdobju 1956-1958

	1956	1957	1958
Proizvodnja ton	1315	1678	1650
Vrednost v tisoč din	237027	215271	146888

Razmerje vrednosti in proizvodnje za proizvedeno smolo pomeni povprečno ceno smole v posameznih letih. Za leto 1956 je koeficient, ki vsebinsko pomeni povprečno ceno, če delimo vrednost s količino

$$P_{1956} = \frac{V_{1956}}{Q_{1956}} = \frac{237027}{1315} = 180 \text{ din/kg}$$

Analogno dobimo za leto 1957 $P_{1957} = \frac{215271}{1678} = 128 \text{ din/kg}$

in za leto 1958: $P_{1958} = \frac{146888}{1650} = 89 \text{ din/kg}$.

Izračunani koeficienti pokažejo, da je povprečna cena smole v letih 1956-1958 vztrajno padala, kar iz absolutnih podatkov o vrednosti in količini ni neposredno razvidno.

4.11

Pomembne koeficiente dobimo tudi s primerjavo drugih podatkov. S primerjavo posekane lesne mase s porabljenim časom za posek dobimo pokazatelje o produktivnosti dela, s primerjavo posekane lesne mase v določenem razdobju s povprečnim številom

prebivalstva v tem razdobju, pokazatelj, ki pove, kakšna je preskrbljenost prebivalstva z lesom itd. Seveda morajo biti primerjani podatki enako opredeljeni. Jasno je, da primerjamo posekano lesno maso s številom ur, ki so bile porabljene za posek tega lesa in ne drugega. Enako moramo v letu 1958 posekano lesno maso v Sloveniji primerjati s prebivalstvom na istem teritoriju in v istem času. V tem primeru nastopi problem, kako časovno enako opredeliti posekano lesno maso in prebivalstvo. Posekana lesna masa se nanaša na leto, medtem ko se število prebivalstva nanaša na določen moment. Iz momentnih podatkov za prebivalstvo dobimo podatek o prebivalstvu, ki se nanaša na celo leto, če iz več podatkov o prebivalstvu izračunamo povprečje. Če imamo podatke za sredino vsakega meseca v letu, izračunamo povprečno število prebivalstva $\bar{Y}$ tako, da podatke za vse mesece seštejemo, vsoto pa delimo z dvanajest.

$$\bar{Y} = \frac{1}{12} (Y_j + Y_f + \dots + Y_n + Y_d) \quad (4.6)$$

Če pa imamo podatke za začetke mesecev, izračunamo povprečje po obrazcu

$$\bar{Y} = \frac{1}{12} \left(\frac{1}{2} Y_j + Y_f + Y_m + \dots + Y_n + Y_d + \frac{1}{2} Y'_j \right) \quad (4.7)$$

tako, da seštejemo polovico momentnega podatka v začetku januarja, cele podatke za začetke drugih mesecev v letu in polovico podatka v začetku januarja naslednjega leta Y'_j , dobijeno vsoto pa delimo z 12. To je potrebno zato, da se izračunano povprečje nanaša točno na letni razmak.

4.12

Gostote. S primerjavo raznovrstnih podatkov pa dobimo tudi gostote. V primeru gostot primerjamo parametre, ki so vsota podatkov za populacijo, z ustreznim razmakom ali izmero za nek znak. Tako dobimo gostoto prebivalstva, če primerjamo prebivalstvo na danem teritoriju s površino tega teritorija. Enako govorimo o gostoti gozda, če primerjamo število dreves ali lesno maso stoječega gozda s površino gozda itd.

V gozdu, ki ima površino $X = 7,57$ ha, stoji $Y = 2274$ dreves.

Gostota gozda je torej

$$a = \frac{Y}{X} = \frac{2274}{7,57} = 300 \text{ dreves/ha}$$

Pri gostotah, pa tudi pri koeficientih, so smiselni tudi recipročni pokazatelji. Gostoto gozda moremo izraziti v številu dreves na enoto površine (v našem primeru na 1 ha) ali s površino, ki odpade na eno drevo. Ta drugi pokazatelj dobimo, če delimo površino s številom dreves.

$$a' = \frac{X}{Y} = \frac{7,57}{2274} = 0,333 \text{ a/drevo}$$

4.13

Med relativna števila sodi tudi prirastek na enoto časa v , ki je kvocient med prirastkom volumna ΔV in časovnim razmakom Δt

$$v = \frac{\Delta V}{\Delta t} \quad (4.8)$$

Relativni prirastek na enoto časa, ki pove, za koliko povprečno priraste v enoti časa enota volumna, pa izračunamo po obrazcu

$$v^o = \frac{\Delta V}{V \cdot \Delta t} \quad (4.9)$$

4.14

Gostote izračunavamo tudi pri proučevanju frekvenčnih porazdelitev, ki imajo neenake širine razredov. Frekvenče pri takih frekvenčnih porazdelitvah niso neposredno primerljive, ker so v bistveni meri odvisne od širine razredov, ki so različne. Vpliv različnih širin razredov odpravimo, če izračunamo za vsak razred gostoto frekvence g_k . Gostota frekvence za vsak razred pove, koliki del od skupne frekvence v razredu odpade na enoto razmaka. Gostoto frekvence izračunavamo po obrazcu

$$g_k = \frac{f_k}{i_k} \quad (4.10)$$

Ker je relativna frekvenca enaka $f_k^o = f_k/N$, izračunamo gostoto relativne frekvence g_k^o po obrazcu

$$g_k^o = \frac{f_k}{i_k \cdot N} \quad (4.11)$$

Iz tega obrazca dobimo nazaj, da je frekvenca v danem razredu f_k enaka

$$f_k = N \cdot i_k \cdot g_k^o \quad (4.12)$$

Frekvenca f_k v danem razredu je torej odvisna od velikosti populacije N , od širine razreda i_k in gostote relativne frekvence g_k^o . Medtem ko moremo za določen pojav spreminjati obseg populacije N in širine razredov, je gostota relativne frekvence najožje povezana z vsebino proučevanega pojava.

4.15

Kot primer za proučitev frekvenčne porazdelitve z neenakimi razredi vzemimo frekvenčno porazdelitev gozdnih površin za $N = 1256$ anketiranih gospodarstev v okraju Novo mesto, iz tabele 3.6.

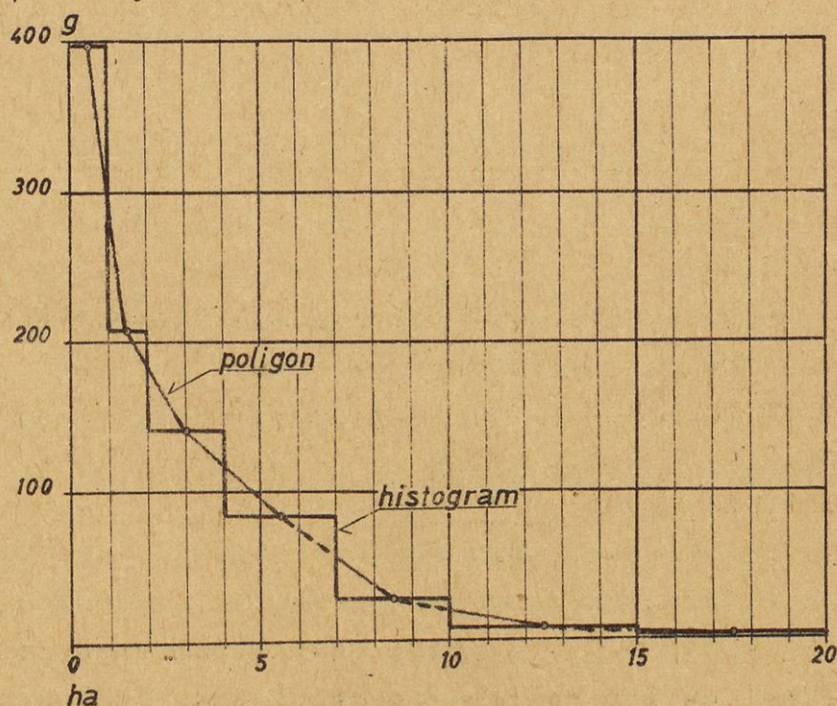
Tabela 4.6 Frekvenčna porazdelitev gozdnih površin za $N = 1256$ anketiranih gospodarstev v Novem mestu

Gozdna površina	i_k	f_k	g_k	f_k^o	g_k^o
0 - pod 1 ha	1	394	394	0,314	0,314
1 ha - pod 2 ha	1	208	208	0,166	0,166
2 ha - pod 4 ha	2	282	141	0,224	0,112
4 ha - pod 7 ha	3	246	82	0,196	0,065
7 ha - pod 10 ha	3	81	27	0,064	0,021
10 ha - pod 15 ha	5	35	7	0,028	0,006
15 ha - pod 20 ha	5	<u>10</u>	2	<u>0,008</u>	0,002
	1256			1,000	

Iz tabele 4.6 spoznamo, da frekvence ne podajajo neposredno zakonitosti o variabilnosti pojava. Frekvenca je v prvem razredu sicer najvišja, potem pade na 208, nato v tretjem razredu zopet naraste na 282 in potem polagoma pada.

V razredu od 2 ha - pod 4 ha je frekvenca izjemno velika, ker je širina razreda enkrat večja kot v predhodnem. Enako je tudi naprej, ker so razredi stalno širši. Šele, če proučimo vrsto za gostoto frekvenc g_k , spoznamo pravo tendenco zakonitosti o variranju gozdne površine po gospodarstvih. Gostota frekvenc je v prvem razredu največja, potem pa rapidno pada, če se površina veča. Iz tega zaključimo, da se kmetijska gospodarstva razporejujejo po gozdni površini v J-porazdelitvi in imamo največ gospodarstev z manjšo površino gozda, število gospodarstev z večjimi površinami gozda pa rapidno pada. Enak odnos kot med frekvenco in gostoto frekvence velja tudi za odnos med relativnimi frekvencami in gostoto relativne frekvence.

Da je frekvenca pri frekvenčnih porazdelitvah z neenakimi razredi v bistveni meri odvisna od širine razreda, moramo upoštevati tudi pri grafični ponazoritvi takih frekvenčnih porazdelitev. Zanje v histogramu prikažemo frekvence v posameznih razredih s stolpci, za katere so širine v sorazmerju s širino razreda, višine pa v sorazmerju z gostoto frekvence. V takem grafikonu je ploščina stolpcev ($i_k \cdot g_k = f_k$) proporcionalna frekvenci f_k , vsota ploščin vseh stolpcev pa ($\sum f_k = N$) obsegu populacije.



Slika 4.3 Grafični prikaz frekvenčne porazdelitve kmetijskih gospodarstev v Novem mestu po gozdni površini

5. KVANTILI

Ranžirna vrsta. Rang.

5.1

Osnovne podatke, ki jih zberemo s statističnim opazovanjem, običajno pregledno prikažemo v frekvenčni porazdelitvi. Za manjše populacije pa osnovne podatke uredimo v ranžirno vrsto. Ranžirna vrsta je niz podatkov za posamezne enote populacije, ki so urejeni po velikosti. Vsaka enota oziroma podatek ima v ranžirni vrsti svoje mesto, ki je označeno z zaporedno številko - rangom. Rang za neko enoto je torej zaporedna številka v vrsti, v kateri so enote populacije urejene po velikosti po nekem numeričnem znaku.

Vzemimo na primer podatke o premerih za 19 modelnih dreves za čist smrekov sestoj na Pokljuki. Neurejeni osnovni podatki so dani v tabeli 5.1

Tabela 5.1 Premeri za 19 modelnih dreves v čistem smrekovem sestoju na Pokljuki

18, 36, 17, 33, 28, 39, 19, 25, 24, 30, 21, 23, 32, 27, 44,
14, 31, 28, 34.

Ti podatki so nepregledni, ker so vpisani v istem vrstnem redu, kot so stala drevesa. Ranžirna vrsta teh podatkov je v tabeli 5.2.

Tabela 5.2 Ranžirna vrsta o premerih za 19 modelnih dreves v čistem smrekovem sestoju A

R	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
x	14	17	18	19	21	23	24	25	27	28	28	30	31	32	33	34	36	39	44

Ker ima vsaka enota svoj rang R, smatramo rang za statistični znak. Čeprav je v zgornjem primeru rang za posamezno drevo določen po premeru drevesa, pa rang pove nekaj drugega kot premer. Če povemo, da je premer drevesa, ki ga proučujemo, 36 cm, ne vemo nič o tem ali je to drevo glede na ostala drevesa v sestoju majhno ali veliko. Medtem pa spozhamo, da je drevo re-

lativno debelo, če povemo, da je proučevano drevo s premerom 36 cm 17. po rangi od skupno $N = 19$ dreves. Samo dvoje dreves ima namreč večji premer, medtem ko je 16 dreves tanjših.

Kvantilni rang

5.2

Hiba ranga pa je v tem, da moramo za dano enoto razen ranga R navesti še obseg populacije N , če hočemo z rangom nakazati mesto enote v populaciji, ker imajo populacije različne obsege. Tako je 17. enota po rangi v populaciji z 19 enotami razmeroma velika, medtem ko bi bila 17. enota po rangi v populaciji z $N = 500$ enotami klasificirana kot majhna, ker bi imelo 483 dreves večje premer, le 16 pa manjše kot je drevo z rangom 17. Zato je primerneje, da izražamo mesto enote v populaciji, namesto z rangom R in obsegom populacije N , s kvantilnim rangom P . Kvantilni rang v relativnem številu pove, na katerem mestu v ranžirni vrsti leži dana vrednost, ker celoto merimo z 1, ne pa z N . Od enote, za katero je kvantilni rang $P = 0,25$, je 0,25 ali 25 % vseh enot manjših, 0,75 ali 75 % pa večjih od proučevane vrednosti. Podobno velja za druge vrednosti kvantilnega ranga. Enota s kvantilnim rangom $P = 0,93$ je zelo velika, ker je 0,93 del celote ali 93 % enot z manjšimi vrednostmi. Glavna prednost kvantilnega ranga je v tem, da je mesto enote v populaciji podano z eno samo vrednostjo. Razen tega pa je to mesto določeno nazorneje, ker je v vsakem primeru, ne glede na obseg populacije, celota merjena z enoto, če so kvantilni rangi izraženi z delom od ena, ali s sto, če so kvantilni rangi izraženi v odstotkih.

5.3

V tabeli 5.2 je rang določen samo za premer, ki jih imajo posamezna drevesa. Zanje so rangi cela števila. Pojem ranga pa moremo razširiti tudi na druge vrednosti med najnižjo ($x_{\min} = 14$ cm) in najvišjo ($x_{\max} = 44$ cm) vrednostjo v populaciji. Tako štejemo, da je rang za premer $x = 37$ cm, $R(x = 37)$ med $R(x = 36) = 16$ in $R(x = 39) = 17$, ker leži vrednost $x = 37$ cm med 36 cm in 39 cm. Točen rang določimo z linearno

interpolacijo. Ker je razlika med $x = 36$ cm in $x = 39$ cm enaka 3 cm, $x = 37$ cm pa je od 36 oddaljen za 1 cm, razlika med ustreznima rangoma pa je 1, pripišemo vrednosti $x = 37$ cm rang $R = 16 \frac{1}{3}$. Enako določimo rang za poljuben premer med 14 in 44 cm. Rang je torej zvezen znak, ker more zavzeti praktično vse vrednosti na določenem razmaku. Vsakemu celemu rangi pripišemo enotni razmak, ki obsega polovico enote pod, polovico enote pa nad danim rangom. Tako pripišemo rangi $R = 1$ razmak od 0,5 do 1,5, rangi $R = 2$ razmak od 1,5 do 2,5 itd. do $R = N$ razmak od $N - 0,5$ do $N + 0,5$. Celoten ranžirni razmak sega torej od 0,5 do $N + 0,5$, ima pa širino N .

Glede na to, da obsega rang razmak od 0,5 do $N + 0,5$, kvantilni rang pa razmak od 0 do 1, je zveza med rangom R in kvantilnim rangom P dana z obrazcem

$$R = NP + 0,5 \quad (5.1)$$

in obratno

$$P = \frac{R - 0,5}{N} \quad (5.2)$$

Preskus z obrazcem 5.1 pokaže, da resnično ustreza $P = 0$, $R = 0,5$ in $P = 1$ rang $R = N + 0,5$

Za premer $x = 36$, za katerega je $R = 16$ po obrazcu 5.2, izračunamo, da je kvantilni rang $P = \frac{16 - 0,5}{19} = 0,82$. Iz tega podatka nazorno sklepamo, da je premer proučevanega drevesa tako velik, da ima 0,82 del vseh dreves ali 82 % vseh dreves manjši premer od proučevanega.

Kvantili

5.4

Z obrazcem 5.1 in 5.2 rešujemo dva po vsebini različna problema.

Če imamo dano vrednost x , moremo zanjo določiti ustrezeni kvantilni rang P_x , ki nakaže mesto te vrednosti v populaciji.

Nalogo moremo pa tudi obrniti in se vprašamo, kakšna vrednost x_p ustreza danemu kvantilnemu rangi P . Te vrednosti pa niso več karakteristike posameznih enot, ampak so parametri za populacijo. Če n.pr. za dani sestoj poiščemo premer, ki

ustreza kvantilnemu rangu $P = 0,50$, dobimo vrednost $x_{0,50}$. $x_{0,50}$ je oni premer, od katerega ima polovica dreves manjše, polovica pa večje premere. Ta vrednost karakterizira populacijo in je za mlad sesto majhna, za starejši sesto pa večja. Vrednost, s katero razdelimo populacijo v dva po obsegu enaka dela, je pomemben parameter in jo imenujemo mediana.

Na splošno imenujemo vrednosti, ki ustrezajo danim kvantilnim rangom, kvantile.

Teoretično moremo izračunati kvantile za katerokoli vrednost kvantilnega ranga. V praksi pa običajno izračunavamo kvantile, ki razdelijo populacijo na dva (mediana), na štiri (kvartil), na deset (decil) ali na sto (centil) delov.

Tako imamo tri kvartile

$$Q_1 = x_{0,25} \qquad Q_2 = x_{0,50} \qquad Q_3 = x_{0,75}$$

devet decilov

$$D_1 = x_{0,10} \qquad D_2 = x_{0,20} \dots\dots D_9 = x_{0,90}$$

in devetindevetdeset centilov

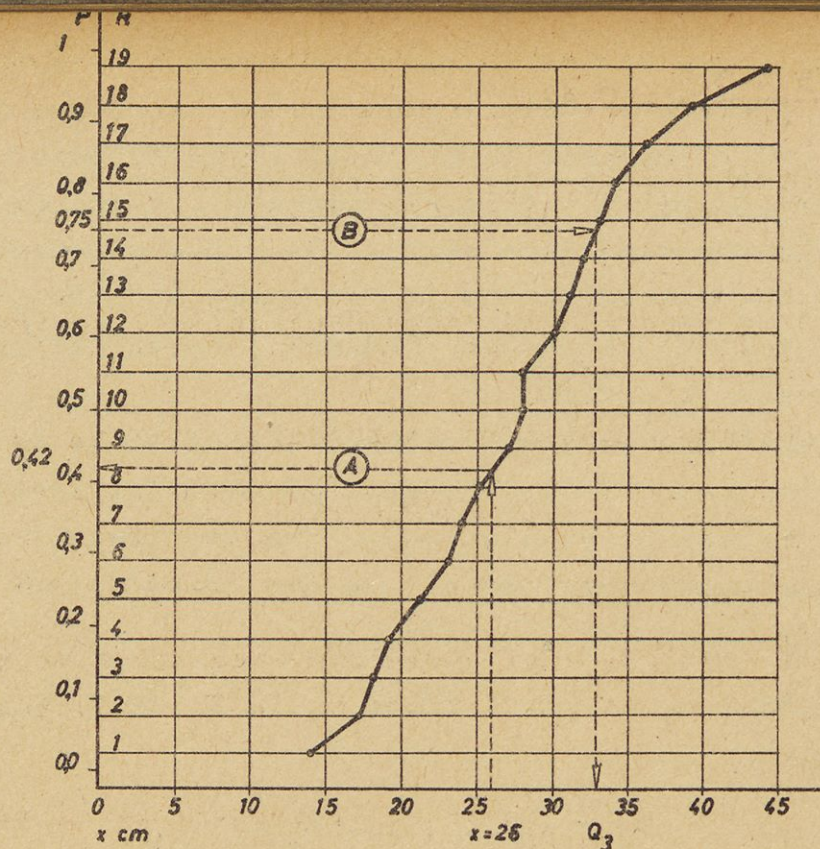
$$C_1 = x_{0,01}; \quad C_2 = x_{0,02}; \quad C_3 = x_{0,03} \dots\dots C_{99} = x_{0,99}$$

5.5

Izračunavanje kvantilnih rangov in kvantilov iz negrupiranih podatkov. Za premere dreves iz tabele 5.2 je v sliki 5.1 nazorno razvidna zveza med osnovnimi vrednostmi, rangi in kvantilnimi rangi. Lomljena črta, ki kaže zvezo med premeri in rangi oziroma kvantilnimi rangi, je stvarno slika kumulativne črte za naš primer. Iz nje ni težko grafično določiti dani vrednosti ustrezní kvantilni rang (primer A) ali danemu kvantilnemu rangi ustrezní kvantil (primer B).

Računsko pa dobimo dani vrednosti x ustrezen kvantilni rang P po naslednjem postopku:

a) V ranžirni vrsti podatkov poiščemo, med kateri zaporedni vrednosti pade dana vrednost x tako, da je $x_0 \leq x < x_1$. Vrednosti x_0 naj ustreza rang R_0 .



Slika 5.1 Odnosi med x , R in P za 19 premerov iz tabele 5.2

b) Po obrazcu

$$R_x = R_0 + \frac{x - x_0}{x_1 - x_0} \quad (5.3)$$

izračunamo rang R_x , ki ustreza vrednosti x .

Po tem obrazcu izračunam rang je enostavna linearna interpolacija med dvema celima rangoma.

c) Iz dobljenega ranga R_x izračunamo ustrezni P_x po obrazcu

$$P_x = \frac{R_x - 0,5}{N} \quad (5.4)$$

Če za primer izračunamo po tem splošnem postopku kvantilni rang za premer $x = 26$ cm, dobimo, da je $x_0 = 25$ cm, $x_1 = 27$ cm. $x_0 = 25$ cm ustreza rang $R_0 = 8$. Obseg populacije $N = 19$. Iz teh podatkov dobimo po obrazcu 5.3

$$R_{x=26} = 8 + \frac{26-25}{27-25} = 8,5$$

Po obrazcu 5.4 pa dobimo dalje, da je kvantilni rang P_x

$$P_x = \frac{8,5 - 0,5}{19} = 0,42$$

Proučevani premer $x = 26$ cm je srednje velik.

5.6

Danemu kvantilnemu rangu P pa poiščemo ustrezni kvantil x_p z obratnim postopkom tako, da:

a) Danemu P poiščemo po obrazcu

$$R_p = NP + 0,5 \quad (5.5)$$

ustrezen rang R_p .

b) V ranžirni vrsti poiščemo, med katerima celima rangoma $R_0 < R_p < R_1$ je R_p . Rangoma R_0 in R_1 ustrezata vrednosti x_0 in x_1 .

c) Iz dobljenih podatkov izračunamo ustrezni kvantil x_p z linearno interpolacijo po obrazcu

$$x_p = x_0 + (x_1 - x_0) \cdot (R_p - R_0) \quad (5.6)$$

5.6

Vzemimo, da iščemo za populacijo premerov iz tabele 5.2 kvartile. Izračun vseh treh kvartilov izvedemo hkrati, kot je navedeno v tabeli 5.3.

Tabela 5.3 Izračun kvartilov za populacijo premerov $N = 19$ smrekovih dreves iz tabele 5.2

i	P	R_p	R_0	x_0	x_1	(6)-(5)	(3)-(4)	(7)-(8)	(5)+(9)= $=Q_1$
1	2	3	4	5	6	7	8	9	10
1	0,25	5,25	5	21	23	2	0,25	0,50	21,50
2	0,50	10,00	10	28	28	0	0	0	28,00
3	0,75	14,75	14	32	33	1	0,75	0,75	32,75

Četrtnina premerov je manjša kot $Q_1 = 21,50$ cm, polovica manjša kot $Q_2 = 28,00$ cm in tri četrtnine manjših kot $Q_3 = 32,75$ cm. Nadalje je polovica premerov v razmaku med $Q_1 = 21,5$ cm

in $Q_3 = 32,75$ cm. S kvantili je populacija premerov torej že precej opisana.

5.7

Izračunavanje kvantilnih rangov in kvantilov iz frekvenčnih porazdelitev. Običajno populacije nimajo nekaj desetlin enot, temveč več sto ali tisoč. Zanje direktna pot za izračunavanje kvantilov preko ranžirne vrste ne pride v poštev, ker je preobsežna in neprikladna. Pač pa moremo zanje oceniti vrednost kvantilov, če imamo podatke urejene v frekvenčni porazdelitvi. Podobno kot so v ranžirni vrsti vse vrednosti, ki so pod določeno vrednostjo, manjše, vse vrednosti, ki so nad določeno vrednostjo, pa večje, velja tudi za frekvenčne distribucije. Vse vrednosti v razredih, ki so pod danim razredom, so manjše, vse vrednosti v razredih nad danim razredom, pa so večje. Frekvenčne porazdelitve imajo torej podobne lastnosti kot ranžirna vrsta, samo da ne veljajo za posamezne vrednosti, temveč za vse enote v posameznih razredih.

Če za frekvenčno porazdelitev izračunamo kumulativno porazdelitev in napravimo predpostavko, da je največja stvarna vrednost enot v posameznem razredu enaka zgornji meji razreda, so posamezni členi v kumulativni vrsti rangi, ki ustrezajo mejam razredov. Kumulativna porazdelitev tako nakaže range le za meje razredov, ne pa za vse vrednosti, ki se pojavljajo v populaciji. Vendar moremo kljub temu iz nje, z linearno interpolacijo, zadosti dobro oceniti range tudi za vmesne vrednosti, če predpostavljamo, da se v prvem približku vrednosti v posameznih razredih porazdeljujejo na vsem razmaku enakomerno.

V tabeli 5.4 je za frekvenčno porazdelitev specifičnih tež za zeleno duglazijo iz tabele 3.5 pod zgornjimi predpostavkami nakazana zveza med mejami razredov in členi v kumulativni vrsti.

Iz tabele 5.4 je razvidno, da je kumulativna vrsta obenem vrsta rangov za spodnje meje ustreznih razredov.

V ranžirni vrsti ima n.pr. specifična teža $x = 480 \text{ kg/m}^3$ rang $R_x = 2395$. S frekvenčno porazdelitvijo moremo določiti ustrezne range tolikim vrednostim, kolikor imamo razredov.

Tabela 5.4 Kumulativna frekvenčna porazdelitev in okrnjena ranžirna vrsta za specifično težo lesa za zeleno duglazijo

Specifična teža v kg/m^3	f_k	F_k R	x_R
320 kg - pod 340 kg	7	00	320
340 kg - pod 360 kg	71	77	340
360 kg - pod 380 kg	163	780	360
380 kg - pod 400 kg	289	241	380
400 kg - pod 420 kg	397	530	400
420 kg - pod 440 kg	484	927	420
440 kg - pod 460 kg	531	1411	440
460 kg - pod 480 kg	453	1942	460
480 kg - pod 500 kg	381	2395	480
500 kg - pod 520 kg	279	2776	500
520 kg - pod 540 kg	165	3055	520
540 kg - pod 560 kg	94	3220	540
560 kg - pod 580 kg	45	3314	560
580 kg - pod 600 kg	16	3359	580
600 kg - pod 620 kg	1	3375	600
N = 3376		3376	620

5.8

Glede na to, da dobimo range za druge vrednosti z linearno interpolacijo, moremo iz frekvenčne porazdelitve oceniti kvantilni rang P_x za vsako vrednost x po naslednjem postopku.

a) Za frekvenčno porazdelitev izračunamo kumulativno vrsto.

b) Poiščemo, v katerem razredu leži vrednost x , za katero iščemo kvantilni rang P_x . Za ta "kvantilni razred" poiščemo spodnjo mejo $x_{0,\min}$, širino razreda i_0 , frekvenco f_0 in kumulativo F_0 .

c) Iz teh podatkov izračunamo pod zgornjimi predpostavkami R_x po obrazcu

$$R_x = F_0 + f_0 \frac{x - x_{0,min}}{i_0} \quad (5.7)$$

d) Iz dobljenega R_x dobimo ustrezni kvantilni rang P_x po znanem obrazcu

$$P_x = \frac{R_x - 0,5}{N} \quad (5.8)$$

Ker je glede na to, da je obseg populacije N velik, običajno količino $0,5$ izpuščamo, ker je nebitvena.

5.9

Vzemimo kot primer, da moramo za določen les zelene duglazije oceniti kvaliteto, ki je dana s specifično težo. S preskusom ugotovimo, da je specifična teža preskušane lesa $x = 554 \text{ kg/m}^3$. Zanima nas, ali je ta specifična teža lesa glede na splošno kakovost lesa za zeleno duglazijo ugodna oziroma dobra ali ne. Odgovor na to vprašanje dobimo, če izračunamo kvantilni rang P_x , ki ustreza $x = 554 \text{ kg/m}^3$.

Po zgornjem navodilu najprej v tabeli 5.4 poiščemo, v katerem razredu v frekvenčni porazdelitvi je ta vrednost. Kvantilni razred je torej razred $540 \text{ kg} - \text{pod } 560 \text{ kg}$, količine, potrebne za izračunavanje kvantilnega ranga, pa so: $x_{0,min} = 540 \text{ kg/m}^3$; $i_0 = 20 \text{ kg/m}^3$; $f_0 = 94$; $F_0 = 3220$. Iz teh podatkov dobimo po obrazcu 5.7 najprej R_x

$$R_x = 3220 + 94 \cdot \frac{554 - 540}{20} = 3285,8$$

po obrazcu 5.8 pa dalje kvantilni rang

$$P_x = \frac{3285,8 - 0,5}{3376} = 0,973$$

Kvaliteta lesa je relativno zelo dobra in le ca 3 % lesa zelene duglazije dano kvaliteto presega.

5.10

Obratno za dan kvantilni rang P poiščemo ustrezni kvantil x_p z istimi predpostavkami po naslednjem postopku.

a) Enako kot pri iskanju kvantilnih rangov najprej izračunamo kumulativno vrsto F_k .

b) Po znanem obrazcu

$$R_p = NP + 0,5 \quad (5.9)$$

izračunamo P ustrezní rang R_p . Količino 0,5 moremo izpustiti, ker je običajno napram $N \cdot P$ nebitvena.

c) V kumulativni vrsti poiščemo, med katerima vrednostima kumulative je izračunani rang R_p ($F_0 \leq R_p < F_1$). Razred, za katerega je kumulativa F_0 , je kvantilni razred o . Zanj ugotovimo: spodnjo mejo $x_{o,min}$, širino i_o , frekvenco f_o in kumulativo F_o .

d) Kvantilnemu rangu P ustrezní kvantil x_p ocenimo po obrazcu

$$x_p = x_{o,min} + i_o \frac{R_p - F_o}{f_o} \quad (5.10)$$

5.11

Kot primer za izračunavanje kvantilov iz frekvenčnih porazdelitev sestavimo tablico decilov za specifično težo lesa zelene duglazije iz tabele 5.4! V tabeli 5.4 je že izračunana kumulativna vrsta frekvenc. Sistematičen izračun vseh kvantilov hkrati je nakazan v tabeli 5.5.

V tabeli 5.5 smo najprej po obrazcu 5.10 izračunali P ustrežajoče range R_p . Za vsak dobljeni rang smo v tabeli 5.4 poiskali, med katera zaporedna člena v kumulativni vrsti pade posamezen R_p . V stolpce (3), (4) in (5) smo vpisali ustrezne količine $x_{o,min}$, f_o in F_o za kvantilne razrede. V glavi je za stolpce (6) do (9) nakazano, kako iz teh podatkov postopoma izračunamo po obrazcu 5.10 decile.

Tablica decilov more služiti kot pripomoček za relativno klasifikacijo kakovosti lesa zelene duglazije glede na specifično težo.

Tako moremo iz decilne tablice oceniti, da je les s specifično težo 500 kg/m^3 med osmim in devetim decilom, kar pomeni, da je najmanj 80 % lesa zelene duglazije, ki ima specifično težo manjšo, in najmanj 10 % lesa zelene duglazije, ki ima večjo specifično težo. Natančna vrednost kvantilnega ranga za specifično težo $x = 500 \text{ kg/m}^3$ je 0,823. Zgornja naša izjava

Tabela 5.5 Izračun decilov za specifično težo lesa za zeleno
duglazijo

P	R_p	$x_{o,min}$	f_o	F_o	$R-F_o$	$i_o(R-F_o)$	$\frac{i_o(R-F_o)}{f_o}$	D_k
(1)	(2)	(3)	(4)	(5)	(6) = $\frac{(2)-(5)}{-(5)}$	(7) = $i_o \cdot (6)$	(8) = $\frac{(7)}{(4)}$	(9) = $=(3) +$ $+(8)$
0,1	338,1	380	289	241	97,1	1942	6,7	387
0,2	675,7	400	397	530	145,7	2914	7,3	407
0,3	1013,3	420	484	927	86,3	1726	3,6	424
0,4	1350,9	420	484	927	423,9	8478	17,5	438
0,5	1688,5	440	531	1411	277,5	5550	10,5	450
0,6	2026,1	460	453	1942	84,1	1682	3,7	464
0,7	2363,7	460	453	1942	421,7	8434	18,6	479
0,8	2701,3	480	381	2395	306,3	6126	16,1	496
0,9	3038,9	500	279	2776	262,9	5258	18,8	519

se torej sklada s točnim rezultatom, samo da je bolj groba. Točnejše vrednosti za kvantilne range dobimo iz centilnih skal. Z decili ocenimo kvantilni rang za posamezno vrednost na eno decimalko natančno. V tablici centilov pa bi dobili kvantilne range ocenjene na dve decimalki.

6.1

Premer, višina, temeljnica ali volumen za posamezno drevo so rezultat vseh faktorjev, ki vplivajo na rast in razvoj drevesa. Ker je učinek vseh faktorjev po pravilu za vsako drevo različen, imajo posamezna drevesa različne premere, različne višine, temeljnice in volumne. Če proučimo posamezne faktorje, jih moremo združiti v tri različne skupine. Proučujmo enodoben smrekov sesto, ki raste na pobočju spremenljivo skalnatega terena. Iz teh podatkov moremo klasificirati faktorje rasti tega sestoj takole. Ker je sesto enodoben in smrekov, je starost in drevesna vrsta za vsa drevesa sestoj ista. Faktorja starost in drevesna vrsta sta torej faktorja, ki vplivata na vsa drevesa enako - imenujemo jih splošne faktorje. Razen splošnih faktorjev pa na posamezno drevo vplivajo še drugi faktorji, katerih učinek se menja od drevesa do drevesa. Ker je vpliv teh faktorjev različen za posamezno enoto - drevo, jih imenujemo individualne faktorje. V našem primeru je eden izmed individualnih vplivov višinska lega drevesa, ker je sesto na pobočju. Drug individualni faktor je skalovitost, ker se spreminja od drevesa do drevesa. Oba ta faktorja moremo za posamezno drevo določiti, ker moremo za vsako drevo ugotoviti višinsko lego ali stopnjo skalovitosti terena, na katerem posamezno drevo raste. Razen tega pa na rast drevesa vplivajo še faktorji, ki jih ne moremo za vsako drevo natančno določiti, čeprav vplivajo na rast drevesa. To so mikrosestav zemlje, kakovost sadike, zasenčenost kraja itd. Te faktorje, katerih učinek ne moremo natančneje opredeliti, združujemo v skupino slučajnih faktorjev oziroma vplivov.

Če bi na pojav vplivali samo splošni vplivi, bi bila vsa drevesa enaka. Njihov premer, višina, temeljnica in volumen bi bili pogojeni le s splošnimi faktorji: starostjo, drevesno vrsto in krajem, kjer sesto raste. Zaradi individualnih vplivov pa se drevesa med seboj razlikujejo in to tem bolj, čim večji je vpliv individualnih faktorjev. Individualni faktorji vplivajo na rast dreves ugodno ali neugodno. Zato se premeri, temeljnice, višine in volumni dreves odklanjajo navzgor in

navzdol od premera, višine, temeljnice in volumna, ki bi ga drevesa imela, če bi nanj vplivali samo splošni vplivi. Premeri, višine, temeljnice ali volumni zaradi individualnih vplivov variirajo okrog nekih idealnih vrednosti, ki bi bile rezultat samo splošnih vplivov. Ker bi bile te vrednosti za vse enote populacije enake, so parametri populacije, imenujemo pa jih srednje vrednosti.

6.2

Če je vpliv individualnih faktorjev majhen ali pa če celo na enote populacije vplivajo samo splošni in slučajni vplivi, katerih učinek je običajno majhen, se posamezna drevesa med seboj malo razlikujejo. V takem homogenem sestoju je volumen vseh dreves precej izenačen. Za homogene populacije srednja vrednost volumna ni le pokazatelj splošnih vplivov. Ker se volumni za posamezna drevesa malo razlikujejo od srednje vrednosti, smatramo srednjo vrednost tudi za reprezentanta vseh vrednosti populacije. Čim večji je vpliv individualnih faktorjev, tem slabše srednja vrednost reprezentira enote v populaciji in obratno: srednja vrednost je tem boljši reprezentant za vrednosti enot v populaciji, čim manjši je vpliv individualnih oziroma slučajnih faktorjev.

Jasno je namreč, da srednji premer, srednja višina ali srednji volumen bolje reprezentira premere, višine in volumne posameznih dreves v sestoju, če je sestoj enodoben, istovrsten, na ravnini, z enakimi talnimi pogoji, kot pa če je obravnavani sestoj prebiralen ali večetažen, na področju, ki je različno skalovit. Za drugi primer srednje vrednosti niso niti približno reprezentanti razmer v populaciji. Zato nimajo velike analitične vrednosti, čeprav jih moremo formalno v vsakem primeru izračunati.

Vrste srednjih vrednosti

6.3

Parametrov, ki na en ali drug način kažejo centralno težnjo vrednosti znaka za populacijo, imamo več. Od njih bomo obravnavali naslednje:

- a) mediano
- b) modus
- c) aritmetično sredino
- d) kvadratično sredino
- e) harmonično sredino
- f) geometrijsko sredino

Mediana in modus sta srednji vrednosti, ki sta dani z lego vrednosti, medtem ko so druge štiri izračunane srednje vrednosti. Razlika srednjih vrednosti po legi v primerjavi z izračunanimi srednjimi vrednostmi je v tem, da srednje vrednosti po legi niso odvisne od vseh vrednosti, medtem ko so izračunane srednje vrednosti odvisne od vseh vrednosti v populaciji. Ta lastnost je v nekem smislu prednost, v nekem smislu pa pomanjkljivost srednjih vrednosti po legi.

Mediana

6.4

Mediana je vrednost, od katere ima polovica enot populacije manjše, polovica pa večje vrednosti. Zaradi te svoje lastnosti mediana dobro služi kot srednja vrednost, ker je v vsakem primeru vrednost, ki se vsem individualnim vrednostim dobro približuje, ker je v sredini med njimi.

Določanje mediane se sklada z določanjem kvantila $x_{0,50}$. Ker smo izračunavanje kvantilov obravnavali že v prejšnjem poglavju, postopka ne bomo ponavljali.

V prejšnjem poglavju smo izračunali mediano iz negrupiranih podatkov o premerih za 19 smrekovih dreves na Pokljuki. Za ta primer je mediana, ki se sklada z drugim kvartilom, enaka $Me = x_{0,50} = 28,00$ cm. Za specifično težo lesa za zeleno duglazijo pa smo v odstavku 5.9 izračunali, da je mediana, ki se v tem primeru sklada s petim decilom D_5 , enaka $Me = D_5 = 450 \text{ kg/m}^3$. Polovica lesa za zeleno duglazijo ima torej manjšo specifično težo kot $Me = 450 \text{ kg/m}^3$, polovica pa večjo. Oba zgornja parametra sta tipični lastnosti populacij, ne pa posameznih enot.

Lastnosti mediane. Prednost mediane je predvsem v tem, da je lahko razumljiva in zato priporočljiva kot opisni parameter. Velika prednost mediane pred izračunanimi srednjimi vrednostmi je tudi ta, da moremo mediano določiti, četudi poznamo le vrednosti, ki leže okrog sredine v ranžirni vrsti urejenih vrednosti. Ta lastnost je posebno ugodna, kadar določamo srednjo vrednost iz frekvenčnih porazdelitev z odprtimi razredi. Za frekvenčne porazdelitve z odprtimi razredi namreč običajno niti približno ne vemo, kakšne so vrednosti v odprtih razredih. Mediana je priporočljiva srednja vrednost tudi takrat, kadar so ekstremne vrednosti take, da sumimo, da so izraz nekih drugih kvalitete in zato sploh ne spadajo v populacijo.

Vsota absolutnih odklonov od neke stalne vrednosti je najmanjša, če te odklone računamo od mediane. Po tem načelu je mediana najboljši reprezentant vseh vrednosti populacije.

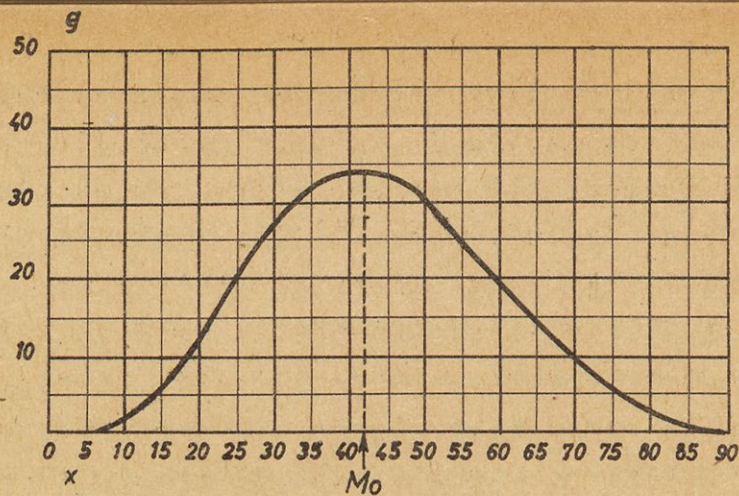
Pomanjkljivost mediane pa je v tem, da je le premalo odvisna od individualnih vrednosti. Za dve različni populaciji je mediana ista, če je le polovica vrednosti manjših kot mediana, ne glede na to, kakšne so individualne vrednosti.

Modus

6.5

Mediana pa ni vedno parameter, ki dobro reprezentira vrednosti v populaciji. Za asimetrične in polimodalne distribucije more biti mediana celo vrednost, od katere je večina vrednosti zelo različna. Zanje mediana ni reprezentant vrednosti v populaciji. Za take primere je veliko bolje, da vzamemo za srednjo vrednost tisto vrednost, okrog katere se gosti največ vrednosti v populaciji. Mesto največje gostitve vrednosti nekega znaka imenujemo najpogostejšo vrednost ali modus.

Če imamo frekvenčno krivuljo, ki ponazarja razporeditev vrednosti v populaciji, je modus enostavno poiskati. V tem primeru je modus abscisna vrednost za tisto točko na frekvenčni krivulji, za katero je ordinata, ki predstavlja gostoto frekvence, največja. To vidimo iz slike 6.1



Slika 6.1 Določanje modusa iz frekvenčne krivulje

Za populacije z majhnim številom enot in za primere, v katerih vrednosti posameznih enot niso grupirane v frekvenčni porazdelitvi, je neposredno nemogoče določiti modus. Modus moremo oceniti le za populacije z razmeroma velikim številom enot, ki so grupirane v frekvenčni porazdelitvi, ker moremo le v tem primeru izslediti mesto največje gostitve.

6.6

Izračunavanje modusa iz frekvenčnih porazdelitev.

V frekvenčni porazdelitvi z enako širokimi razredi iščemo modus v razredu, za katerega je frekvenca največja, ker predvidevamo, da je v tem razredu mesto največje gostitve. Za frekvenčne porazdelitve z različno širokimi razredi pa iščemo modus v razredu, za katerega je gostota frekvence največja. Vendar se bomo omejili le na določanje modusa za frekvenčne porazdelitve z enako širokimi razredi.

Prvi približek modusa je sredina modalnega razreda, to je razreda z največjo frekvenco. Vendar je sredina modalnega razreda modus le za simetrične frekvenčne distribucije. Za asimetrične distribucije pa se modus od sredine modalnega razreda odklanja v smer onega sosednega razreda, v katerem je frekvenca večja.

Zaznamujmo modalni razred z o , razred pred njim z $+1$, razred za modalnim razredom pa s $+1$. Ustrezne sredine razredov

so x_{-1} , x_0 , x_{+1} , ustrezne frekvence pa f_{-1} , f_0 in f_{+1} . S parabolo druge stopnje, ki gre skozi točke s koordinatami (x_{-1}, f_{-1}) , (x_0, f_0) in (x_{+1}, f_{+1}) , v tem primeru približamo frekvenčno krivuljo v območju modusa. Kot oceno modusa vzamemo absciso najvišje točke v tej paraboli. Pri tej predpostavki izračunamo drugo oceno za modus po naslednjem postopku:

a) v frekvenčni porazdelitvi poiščemo razred z največjo frekvenco - modalni razred. Modalni razred zaznamujemo z o. $x_{0,min}$ je spodnja meja, f_0 pa frekvenca v modalnem razredu. f_{-1} je frekvenca v razredu pred modalnim razredom, f_{+1} pa je frekvenca v razredu za modalnim razredom;

b) izračunamo diferenci sosednih frekvenc:

$$d_{-1} = f_0 - f_{-1}; \quad d_{+1} = f_0 - f_{+1};$$

c) modus M_0 izračunamo iz zgornjih količin po obrazcu

$$M_0 = x_{0,min} + i \frac{d_{+1}}{d_{-1} + d_{+1}} \quad (6.1)$$

6.7

Proučujmo frekvenčno porazdelitev premerov za čisti smrekov enodobni sestoj B na Pokljuki.

Proučevana frekvenčna porazdelitev je naslednja.

Tabela 6.1 Porazdelitev premerov za čisti, enodobni smrekov sestoj na Pokljuki.

Debelinska stopnja	f_k
15 cm - 19 cm	5
20 cm - 24 cm	34
25 cm - 29 cm	87
30 cm - 34 cm	93
35 cm - 39 cm	74
40 cm - 44 cm	42
45 cm - 49 cm	23
50 cm - 54 cm	13
55 cm - 59 cm	4
60 cm - 64 cm	2
65 cm - 69 cm	4
	381 = N

Iz zgornje frekvenčne porazdelitve sklepamo, da je modus v razredu 30 cm - 34 cm, ker je frekvenca v tem razredu največja (93). Prva ocena modusa je torej $M_0 = 32,5$ cm, ker se modalni razred razteza od 30 cm do pod 35 cm.

Drugo oceno modusa dobimo, če najprej iz frekvenčne porazdelitve v tabeli poiščemo potrebne količine. Modalni razred je razred 30 cm - 34 cm. Za modalni razred je spodnja meja $x_{0,min} = 30$ cm, frekvenca $f_0 = 93$, širina razreda pa $i = 5$ cm. Frekvenca razreda pred modalnim razredom je $f_{-1} = 87$, frekvenca razreda za modalnim razredom pa $f_{+1} = 74$. Iz teh podatkov dobimo, da je:

$$d_{-1} = 93 - 87 = 6; \quad d_{+1} = 93 - 74 = 19$$

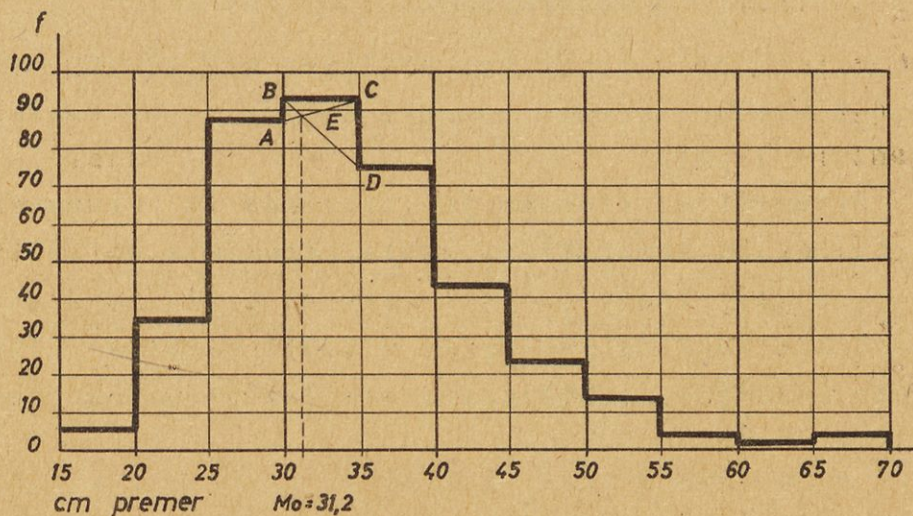
Iz teh podatkov je po obrazcu 6.1

$$M_0 = 30 + 5 \frac{6}{6 + 19} = 31,2 \text{ cm}$$

Rezultat resnično pokaže, da je modus odklonjen od sredine razreda $x_0 = 32,5$ cm navzdol v smeri večje sosedne frekvence.

6.8

Kako grafično ocenimo modus iz histograma, je nakazano v sliki 6.2. Če zvežemo točko A s C, točko B pa z D, je projekcija presečišča obeh daljic E na abscisno os - modus. Teoretično se ta rezultat sklada z rezultatom, ki ga dobimo računsko. Eventuelne razlike gredo na račun nenatančnega risanja.



Slika 6.2 Grafično določanje modusa

Lastnosti modusa. Modus je srednja vrednost, ki ima že po svoji definiciji kvalitete reprezentanta individualnih vrednosti, ker je mesto, na katerem se gosti največ vrednosti. Enako kot mediana, pa je tudi modus neobčutljiv za spremembe individualnih vrednosti. Modus ostane v dani populaciji nespremenjen vse dotlej, dokler stopnja gostitve na nekem drugem mestu ne prekorači stopnje gostitve v modusu.

Heterogene porazdelitve morejo imeti več mest gostitve. Tako imamo pri frekvenčni porazdelitvi posameznih karakteristik za drevesa v dvoetažnem sestoju običajno dva modusa. Eden izraža tipičnost za eno, drugi pa tipičnost za drugo etažo. Podobno imamo pri več-modalnih porazdelitvah še več relativnih modulusov. V vsakem takem primeru je pa običajno en modus izmed njih tisti, ki kaže mesto absolutno največje gostitve.

Aritmetična sredina

6.10

Izmed vseh srednjih vrednosti zaradi praktičnih in teoretičnih lastnosti najpogosteje uporabljamo aritmetično sredino. Aritmetična sredina $\bar{x}$ je enostavno povprečje iz vseh individualnih vrednosti za vse enote v populaciji. Dobimo jo, če vsoto vseh vrednosti $\sum x_i$ delimo s številom enot v populaciji N . Z obrazcem moremo to definicijo aritmetične sredine zapisati

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{X}{N} \quad (6.2)$$

Aritmetično sredino zaznamujemo običajno z $\bar{x}$, $\bar{y}$... Prečna črtica nad simbolom za znak pomeni, da gre za aritmetične sredine ustreznih znakov.

Aritmetična sredina je izpeljana iz predpostavke, da je individualna vrednost x_i vsota rezultata splošnih vplivov $\bar{x}$ in rezultata individualnih vplivov e_i ($x_i = \bar{x} + e_i$). Nadaljnja predpostavka pa je, da se v vsoti rezultati individualnih vplivov uničijo. Če seštejemo vse individualne vrednosti, dobimo glede na to

$$X = \sum_{i=1}^N x_i = \sum_{i=1}^N (\bar{x} + e_i) = N\bar{x} + \sum_{i=1}^N e_i = N\bar{x} \quad (6.3)$$

Iz te enačbe pa je $\bar{x} = X/N$, kar je po definiciji aritmetična sredina.

Lastnosti aritmetične sredine

6.11

Aritmetična sredina je najvažnejša izračunana srednja vrednost. Zaradi svojih lastnosti je najvažnejša in najpogostejše uporabljana srednja vrednost na sploh.

Najvažnejše lastnosti aritmetične sredine so:

a) Aritmetična sredina za linearno zvezo

$z = a + bx + cy$ med znakom x in y za isto populacijo je enaka linearni zvezi aritmetičnih sredin. V obrazcu je ta stavek naslednji:

$$\bar{z} = \overline{a + bx + cy} = a + b\bar{x} + c\bar{y} \quad (6.4)$$

Ta stavek je zlahka dokazati in izhaja neposredno iz znanega stavka, da je vsota linearnih izrazov enaka linearnemu izrazu iz vsot.

Iz obrazca neposredno sledi, da je aritmetična sredina konstante konstanta sama

$$\bar{a} = a \quad (6.5)$$

in da je aritmetična sredina produkta konstante z znakom enaka produktu konstante z aritmetično sredino znaka

$$\overline{bx} = b\bar{x} \quad (6.6)$$

b) Vsota odklonov individualnih vrednosti od aritmetične sredine

$$\sum_{i=1}^N (x_i - \bar{x}) = 0 \quad (6.7)$$

je nič.

c) Vsota kvadratov odklonov od neke konstantne vrednosti je najmanjša, če računamo odklone od aritmetične sredine

$$K = \sum_{i=1}^N (x_i - C)^2 = \text{Min, če je } C = \bar{x} \quad (6.8)$$

Ta lastnost aritmetične sredine je zelo važna in je osnova ene izmed najvažnejših mer variacije-variance.

Izračunavanje aritmetične sredine iz negrupiranih podatkov. Iz negrupiranih podatkov izračunavamo aritmetično sredino neposredno po obrazcu 6.2.

Če proučujemo pet dreves, za katere so volumni v m³ lesa 1,04 1,35 1,28 1,57 1,17, je aritmetična sredina ali povprečen volumen zanje enak

$$\bar{x} = \frac{1}{5} (1,04 + 1,35 + 1,28 + 1,57 + 1,17) = \frac{6,41}{5} = 1,282 \text{ m}^3$$

Povprečen volumen zgornjih petih dreves je $\bar{x} = 1,282 \text{ m}^3$.

To pomeni: če bi vsako izmed dreves imelo isti volumen kot je povprečen (1,282 m³), bi se skupna lesna masa teh dreves skladala z dejansko skupno lesno maso vseh petih dreves.

Izračunavanje aritmetične sredine iz frekvenčnih porazdelitev. Izračunavanje aritmetične sredine iz individualnih podatkov bi bilo za stvarne populacije, ki imajo navadno veliko število enot, neprikladno in zamudno. Frekvenčna porazdelitev daje približno sliko o vseh vrednostih v populaciji. Zato moremo iz njih oceniti aritmetično sredino. To velja predvsem za simetrične porazdelitve. Čim bolj pa je porazdelitev asimetrična, tem večja je sistematična napaka v oceni za aritmetično sredino. V skrajnosti je najslabša ocena za J-distribucije, ker so ekstremno asimetrične. Kakovost ocene za aritmetično sredino iz frekvenčne porazdelitve je jasno odvisna tudi od širine razredov v frekvenčni porazdelitvi in je tem boljša, čim manjši so razredi.

Neposredna metoda. Če predpostavljamo, da so sredine razredov x_k reprezentanti posameznih vrednosti v ustreznih razredih, je ocena za vsoto vrednosti v danem razredu produkt frekvence in sredine razreda $f_k x_k$. Ocena vsote vseh vrednosti v populaciji pa je vsota produktov frekvenc in ustreznih sredin razredov za vse razrede $\sum_k f_k x_k$. V obrazcu 6.2 za izračunavanje aritmetične sredine moremo torej vsoto $\sum_{i=1}^N x_i$ nadomestiti

z oceno vsote $\sum_{k=1}^r f_k x_k$. Tako dobimo, da je ocena za aritmetično sredino iz frekvenčne porazdelitve

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_r x_r}{f_1 + f_2 + \dots + f_r} = \frac{\sum_k f_k x_k}{\sum_k f_k} \quad (6.9)$$

Ta način izračunavanja aritmetične sredine imenujemo tehtan ali ponderiran način izračunavanja. Pri tem so frekvence teže ali ponderi.

Po tem obrazcu izračunavamo aritmetično sredino po naslednjih stopnjah:

- Za frekvenčno porazdelitev poiščemo frekvencam f_k ustrezne sredine razredov x_k .
- Izračunamo produkte frekvenc in sredin razredov $f_k x_k$.
- Izračunamo vsoto v točki b dobljenih produktov $\sum f_k x_k$.
- Vsoto $\sum f_k x_k$ delimo z vsoto frekvenc oziroma številom enot N in dobimo aritmetično sredino $\bar{x}$.

6.15

Kot primer za izračunavanje aritmetične sredine po neposredni metodi vzemimo izračun povprečnega premera za čisti smrekov sestoj A na Pokljuki iz tabele 3.4. Izračun je sistematično prikazan v tabeli 6.2.

Če za oceno kakovosti izračunane aritmetične sredine primerjamo dobljeni rezultat s pravo aritmetično sredino, izračunano iz individualnih podatkov $\bar{x}_p = 38,55$ cm, spoznamo, da je razlika resnično neznatna: $\Delta \bar{x} = 38,62 - 38,55 = 0,07$ cm ali ca 2 ‰. Ta dober rezultat je pripisati temu, da je obravnavana porazdelitev precej simetrična. Povprečje, zaokroženo na eno decimalno $\bar{x} = 38,6$ cm, se sklada s pravim rezultatom.

Tabela 6.2 Izračun povprečnega premera po neposredni metodi za čisti smrekov sestoj A na Pokljuki (iz tabele 3.4)

DS		f_k	x_k	$f_k x_k$
4	15 cm - 19 cm	4	17,5	70,0
5	20 cm - 24 cm	6	22,5	135,0
6	25 cm - 29 cm	55	27,5	1512,5
7	30 cm - 34 cm	118	32,5	3835,0
8	35 cm - 39 cm	112	37,5	4200,0
9	40 cm - 44 cm	109	42,5	4632,5
10	45 cm - 49 cm	61	47,5	2897,5
11	50 cm - 54 cm	28	52,5	1470,0
12	55 cm - 59 cm	9	57,5	517,5
13	60 cm - 64 cm	5	62,5	312,5
		$N = 507$	$19582,5 = \sum f_k x_k$	

$$\bar{x} = \frac{\sum f_k x_k}{N} = \frac{19582,5}{507} = 38,62 \text{ cm}$$

6.15

Metoda s pomožnim znakom u. Za frekvenčne porazdelitve z enakimi širinami razredov izračunavanje aritmetične sredine poenostavimo, če vpeljemo namesto osnovnega znaka x pomožni znak u , ki je z znakom x v linearni zvezi

$$x = x_0 + i \cdot u \quad (6.10)$$

Pri tem je x_0 sredina poljubnega razreda nekje v sredini frekvenčne porazdelitve, i pa širina razredov. Iz znanih lastnosti o aritmetičnih sredinah sledi, da sta tudi aritmetični sredini znakov x in u v linearni zvezi

$$\bar{x} = x_0 + i \cdot \bar{u} = x_0 + i \frac{\sum_{k=1}^r f_k u_k}{N} \quad (6.11)$$

Če proučimo, katere vrednosti za u ustrezajo sredinam razredov, s katerimi moramo pri direktni metodi pomnožiti frekvence, spoznamo, da so za sredine razredov u_k zelo enostavne vrednosti. Za razred, za katerega je sredina razreda x_0 , je $u = 0$, za druge razrede pa je po vrsti navzgor $+1, +2, +3, +4$ itd., po vrsti

navzdol pa -1, -2, -3, -4 itd. Izračun aritmetične sredine se poenostavi, ker $\bar{u}$ izračunamo enostavno, $\bar{x}$ pa dobimo po obrazcu 6.11.

6.17

V tabeli 6.3 je nakazan izračun povprečnega premera za isti sestoj kot po neposredni metodi, da s primerjavo spoznamo prednosti metode pomožnega znaka u .

Tabela 6.3 Izračun povprečnega premera po metodi s pomožnim znakom u za čisti smrekov sestoj A

DS	Deb. stopnja	f_k	u_k	$f_k u_k$
4	15 cm - 19 cm	4	-4	-16
5	20 cm - 24 cm	6	-3	-18
6	25 cm - 29 cm	55	-2	-110
7	30 cm - 34 cm	118	-1	-118
8	35 cm - 39 cm	112	0	0
9	40 cm - 44 cm	109	+1	+109
10	45 cm - 49 cm	61	+2	+122
11	50 cm - 54 cm	28	+3	+84
12	55 cm - 59 cm	9	+4	+36
13	60 cm - 64 cm	5	+5	+25

$N = 507$

$+114 = \sum fu$

V našem primeru smo vzeli izhodišče v razredu 35 cm - 39 cm, tako da je $x_0 = 37,5$, širina razreda $i = 5$ cm. Iz teh podatkov dobimo po obrazcu 6.11

$$\bar{x} = 37,5 + 5 \frac{+114}{507} = 38,62$$

Primerjava postopka po neposredni metodi s postopkom po metodi pomožnega znaka u pokaže, da smo se izognili množenju večmestnih števil, rezultata pa sta enaka.

6.18

Metoda kumulativ. Metoda kumulativ ima pred metodo pomožnega znaka še to prednost, da se izognemo sploh vsakemu množenju, ker pridemo do elementov za izračun aritmetične sre-

dine s kumulativnim seštevanjem frekvenc. Po tej metodi izračunamo aritmetično sredino po naslednjih stopnjah:

a) Za frekvenčno porazdelitev z enakimi širinami razredov izračunamo kumulativno frekvenčno porazdelitev F_k . Zadnji člen v kumulativni vrsti je N .

b) Seštejemo vrednosti vseh členov v kumulativni vrsti F_k , brez zadnjega pod črto. Tako dobimo količino A .

c) Iz dobljenih podatkov izračunamo aritmetično sredino po obrazcu

$$\bar{x} = x_0 + i \frac{A}{N} \quad (6.12)$$

pri čemer pomeni razen znanih količin x_0 sredino za zadnji razred v frekvenčni porazdelitvi.

6.19

Na istem primeru kot po prejšnjih metodah preizkusimo še metodo kumulativ.

Tabela 6.4 Izračun povprečnega premera po metodi kumulativ za čisti smrekov sestoj na Pokljuki

DS	Premer v cm	f_k	F_k	
4	15 cm - 19 cm	4	0	
5	20 cm - 24 cm	6	4	
6	25 cm - 29 cm	55	10	
7	30 cm - 34 cm	118	65	
8	35 cm - 39 cm	112	183	
9	40 cm - 44 cm	109	295	
10	45 cm - 49 cm	61	404	
11	50 cm - 54 cm	28	465	
12	55 cm - 59 cm	9	493	
13	60 cm - 64 cm	5	502	4+10+65...+493+502=
				= 2421 = A
				N = 507

Po obrazcu 6.12 je aritmetična sredina

$$\bar{x} = 62,5 - 5 \frac{2421}{507} = 38,62 \text{ cm}$$

Iz primera vidimo, da je ta metoda izmed vseh treh najprikladnejša, ~~ne~~ glede na to, da dobimo kot postranski rezultat še kumulativno vrsto, ki jo mnogokrat potrebujemo za izračunavanje drugih parametrov ali za analizo.

Aritmetična sredina aritmetičnih sredin

6.20

Če poznamo aritmetične sredine $\bar{x}_k$ in obsege populacij N_k za delne populacije, ki sestavljajo populacijo, izračunamo aritmetično sredino za celotno populacijo po obrazcu

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + \dots + N_r\bar{x}_r}{N_1 + N_2 + \dots + N_r} = \frac{\sum_k N_k \bar{x}_k}{\sum_k N_k} \quad (6.13)$$

kot tehtano aritmetično sredino iz grupnih sredin. Pri tem vzamemo število enot v delnih populacijah za pondere.

Lastnost, da moremo neposredno iz grupnih aritmetičnih sredin izračunati skupno aritmetično sredino, je velika prednost aritmetične sredine pred srednjimi vrednostmi po legi. Zanje namreč to ni mogoče. Če hočemo za skupno populacijo določiti mediano ali modus, moramo iz delnih populacij sestaviti skupno populacijo in šele iz te določiti skupno mediano ali modus.

6.21

Vzemimo za primer dvoetažni sestoj na Otoku iz tabele 3.4. Za obravnavani sestoj je povprečen premer za prvo etažo (topol) $\bar{x}_1 = 46,82$ cm, število dreves pa $N_1 = 103$. Za drugo etažo (jelša) z $N_2 = 132$ drevesi pa je povprečen premer $\bar{x}_2 = 19,70$ cm. Po obrazcu 6.13 je povprečen premer dreves v sestoju

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} = \frac{103 \cdot 46,82 + 132 \cdot 19,70}{103 + 132} = 31,59 \text{ cm}$$

Ta rezultat pa ima omejeno analitično vrednost, ker ni reprezentativen niti za prvo, niti za drugo etažo. Zato moramo za skupinska povprečja za heterogene populacije paziti, da jih pravilno tolmačimo oziroma jih moramo uporabljati s primernimi omejitvami.

6.22

Tehtan način za izračunavanje povprečij velja tudi za izračunavanje relativnih števil za populacijo iz relativnih števil za grupe. Tako računamo n.pr. skupni odstotek P iz grupnih odstotkov P_k po obrazcu

$$P = \frac{\sum N_k P_k}{\sum N_k} \quad (6.14)$$

6.23

V treh mešanih sestojih nekega gozdnega gospodarstva je število dreves in odstotek jelke po posameznih sestojih naslednje: $N_1 = 500$; $P_1 = 20 \%$, $N_2 = 1000$, $P_2 = 15 \%$ in $N_3 = 2000$, $P_3 = 10 \%$.

Povprečen odstotek jelke v sestojih gozdnega gospodarstva je po obrazcu

$$\begin{aligned} P &= \frac{N_1 P_1 + N_2 P_2 + N_3 P_3}{N_1 + N_2 + N_3} = \\ &= \frac{500 \cdot 20 + 1000 \cdot 15 + 2000 \cdot 10}{500 + 1000 + 2000} = 12,9 \% \end{aligned}$$

6.24

S tehtano aritmetično sredino izračunavamo tudi različne druge parametre iz gozdarstva.

Volumen za posamezno drevo je dan z zvezo $V = g \cdot h \cdot f$, pri čemer pomeni: V = volumen, g = temeljnica, h = višina, f = telesninski koeficient.

Če povprečno temeljnico izračunamo po obrazcu

$$\bar{g} = \frac{\sum g_i}{N} \quad (6.15)$$

povprečno višino drevesa po obrazcu

$$\bar{h} = \frac{\sum g_i h_i}{\sum g_i} \quad (6.16)$$

kot tehtano aritmetično sredino individualnih višin, pri čemer vzamemo temeljnico posameznega drevesa kot ponder,

povprečni telesninski koeficient $\bar{f}$ pa po obrazcu

$$\bar{f} = \frac{\sum g_i h_i f_i}{\sum g_i h_i} \quad (6.17)$$

kot tehtano aritmetično sredino, pri čemer vzamemo kot ponde-re produkte temeljnice in višine $g_i h_i$, dobimo, da je skupen volumen v sestoji direktno enak produktu

$$V = N \cdot \bar{g} \cdot \bar{h} \cdot \bar{f} \quad (6.18)$$

Če izračunamo povprečno temeljnico, povprečno višino in povprečni telesninski koeficient po zgornjih obrazcih, je zveza med povprečji za sestoj enaka kot za individualne vrednosti, le da pri posameznem drevesu upoštevamo individualne vrednosti, pri sestoji pa računamo s povprečji.

Harmonična sredina

6.25

Harmonična sredina je recipročna vrednost povprečja iz recipročnih vrednosti za posamezne podatke. Po tej definiciji izračunamo harmonično sredino po obrazcu

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}} \quad (6.19)$$

Tehtano harmonično sredino pa izračunamo po obrazcu

$$H = \frac{\sum_k w_k}{\sum_k \frac{w_k}{x_k}} \quad (6.20)$$

Če iz istih podatkov izračunamo aritmetično sredino in harmonično sredino, spoznamo, da so razlike lahko znatne. Vzemimo za shematičen primer pet vrednosti: 1, 2, 4, 7, 9. Za te vrednosti je aritmetična sredina

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

harmonična sredina pa

$$H = \frac{5}{\frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{7} + \frac{1}{9}} = 2,5$$

Harmonično sredino izračunavamo redkeje kot aritmetično. Kadar pa se osnovne vrednosti v populaciji porazdeljujejo asimetrično, njihovi reciproki pa simetrično, harmonična sredina bolje reprezentira vrednosti populacije kot aritmetična sredina.

6.26

Razen tega pa je harmonična sredina, ne pa aritmetična sredina upravičena pri izračunavanju nekaterih relativnih števil.

Vzemimo kot primer pet smrekovih dreves z enakim volumenom. Zanje so telesninski koeficienti: 0,38 0,36 0,41 0,35 0,40. Povprečen telesninski koeficient f izračunamo s harmonično sredino

$$f = \frac{5}{\frac{1}{0,38} + \frac{1}{0,36} + \frac{1}{0,41} + \frac{1}{0,35} + \frac{1}{0,40}} = 0,3786$$

Upravičenost zgornjega rezultata sledi iz naslednjega sklepanja. Po obrazcu 6.17 je povprečen telesninski koeficient

$$\bar{f} = \frac{\sum g_i h_i f_i}{\sum g_i h_i} = \frac{\sum V_i}{\sum V_i / f_i} = \frac{NV}{\sum 1/f_i} = \frac{N}{\sum 1/f_i} \quad (6.21)$$

Izpeljava tega obrazca se naslanja na to, da je $V = g \cdot h \cdot f$, iz tega dalje $V/f = h \cdot g$ in da je pri drevesih z enakimi volumni $\sum V_i = NV$.

6.27

Ali pri praktičnem izračunavanju povprečnih vrednosti iz relativnih števil uporabljamo aritmetično ali harmonično sredino, je odvisno od tega, katere količine so ponderi.

To moremo ugotoviti v vsakem posameznem primeru iz strukture relativnih števil.

Če vzamemo, da je za individualne enote, grupe in populacijo relativno število kvocijent dveh absolutnih podatkov, je sumarno relativno število R

$$R = \frac{Y}{X} = \frac{\sum_k Y_k}{\sum_k X_k} \quad (6.22)$$

kvocijent vsot absolutnih podatkov Y in X .

Sumarno relativno število pa izračunamo kot tehtano aritmetično sredino

$$R = \frac{\sum Y_k}{\sum X_k} = \frac{\sum X_k \frac{Y_k}{X_k}}{\sum X_k} = \frac{\sum X_k R_k}{\sum X_k} \quad (6.23)$$

kadar so ponderi absolutne količine, ki v relativnem številu nastopajo v imenovalcu.

Tehtano harmonično sredino pa uporabljamo pri izračunavanju relativnih števil, če so ponderi absolutni podatki, ki nastopajo v relativnem številu v števcu. To vidimo iz zveze

$$R = \frac{\sum Y_k}{\sum X_k} = \frac{\sum Y_k}{\sum Y_k \frac{Y_k}{X_k}} = \frac{\sum Y_k}{\sum Y_k / R_k} \quad (6.24)$$

6.28

Izračunajmo povprečno temeljnico za dvoetažni sestoj, za katerega je povprečna temeljnica v prvi etaži $\bar{g}_1 = 0,050 \text{ m}^2$, v drugi etaži pa $\bar{g}_2 = 0,075 \text{ m}^2$. Razen teh podatkov poznamo skupni temeljnici $G_1 = 50 \text{ m}^2$ in $G_2 = 150 \text{ m}^2$ za obe etaži.

Povprečna temeljnica je kvocijent med skupno temeljnico in številom dreves v sestoji $\bar{g} = G/N$. Razen povprečnih te-

meljnic za vsako etažo pa poznamo še skupno temeljnico, ki je v relativnem številu v imenovalcu. Po zgornjem pravilu v tem primeru izračunamo povprečno temeljnico za dvoetažni sestoj po obrazcu za tehtano harmonično sredino

$$\bar{g} = \frac{G_1 + G_2}{\frac{G_1}{g_1} + \frac{G_2}{g_2}} = \frac{50 + 150}{\frac{50}{0,050} + \frac{150}{0,075}} = 0,067 \text{ m}^2$$

Kvadratična sredina

6.29

Kvadratična sredina je kvadratni koren iz povprečij kvadratov iz individualnih vrednosti

$$K = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_N^2}{N}} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (6.25)$$

Kvadratično sredino samostojno redkeje uporabljamo kot druge izračunane vrednosti. Pač pa je pomembna predvsem zato, ker je najvažnejša mera variacije - standardni odklon kvadratična sredina odklonov individualnih vrednosti od aritmetične sredine

$$(\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}).$$

6.30

V gozdarstvu pa so še drugi problemi, v katerih uporabljamo kvadratične sredine.

Če iščemo, kakšen povprečen premer bi morala imeti vsa drevesa v sestoji, da bi bila skupna temeljnica za tak sestoj enaka stvarni skupni temeljnici, spoznamo, da temu pogoju ustreza kvadratična sredina iz stvarnih premerov v sestoji. To zlahka spoznamo iz naslednje zveze.

Če z x_i zaznamujemo premer za posamezno drevo v sestoji, je $\frac{\pi}{4} x_i^2$ temeljnica posameznega drevesa, $\sum_{i=1}^N \frac{\pi}{4} x_i^2$ pa vsota temeljnic vseh dreves v sestoji ali skupna temeljnica. Če s K zaznamujemo povprečen premer, je temeljnica povprečnega drevesa $\frac{\pi}{4} K^2$, skupna temeljnica vseh N dreves pa $N \frac{\pi}{4} K^2$. Zaradi zgornje zahteve, ki naj jo izpolni povprečen premer, velja

$$N \frac{\pi}{4} K^2 = \sum_{i=1}^N \frac{\pi}{4} x_i^2$$

Če to enačbo najprej delimo z $N \frac{\pi}{4}$, iz desne in leve strani pa poiščemo kvadratni koren, dobimo, da je

$$K = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

Iskani povprečni premer je torej kvadratična sredina iz individualnih premerov.

6.31

Vzemimo kot shematičen primer pet dreves. Njihovi premeri so: 38 cm, 28 cm, 32 cm, 35 cm, 41 cm. Povprečen premer, ki ustreza pogoju, da je skupna temeljnica enaka kot skupna temeljnica, če bi vseh pet dreves imelo enak povprečen premer, je enak $K = \frac{1}{5} \sqrt{(38^2 + 28^2 + 32^2 + 35^2 + 41^2)} = 35,1$ cm.

Aritmetična sredina za isti primer je $\bar{x} = 34,8$ cm. Če napravimo preskus, kateri povprečen premer ustreza zahtevi glede skupne temeljnice, dobimo:

Drevo	Premer cm	Temelnjica m ²
1	38	0,1134
2	28	0,0616
3	32	0,0804
4	35	0,0962
5	41	0,1320
		<u>0,4836 m²</u>

Skupna temeljnica je torej 0,4836 m².

Za povprečen premer $K = 35,1$ cm je temeljnica $g_K = 0,09676$ m², skupna temeljnica petih povprečnih dreves pa $5 \cdot 0,09676 = 0,4838$ m². Za aritmetično sredino premers $\bar{x} = 34,8$ cm pa dobimo, da je temeljnica povprečnega drevesa $g_{\bar{x}} = 0,09541$ m², skupna temeljnica petih povprečnih dreves pa $5 \cdot 0,09541 = 0,37705$ m².

6.32

Geometrijska sredina iz N vrednosti je N -ti koren iz produkta individualnih vrednosti. Po tej definiciji je geometrijska sredina

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N} = \sqrt[N]{\prod_{i=1}^N x_i} \quad (6.26)$$

Iz definicije za geometrijsko sredino sklepamo, da ima smisel izračunavati geometrijsko sredino le, če noben osnovni podatek ni nič ali negativen.

Izračunavanje geometrijske sredine po osnovnem obrazcu je, razen za najenostavnejše primere, nemogoč. Pač pa jo moremo razmeroma enostavno izračunati z logaritmi. Če levo in desno stran osnovnega obrazca logaritmiramo, dobimo

$$\log G = \frac{1}{N} (\log x_1 + \log x_2 + \dots + \log x_N) \quad (6.27)$$

Logaritem iz geometrijske sredine je torej povprečje iz logaritmov individualnih vrednosti.

Zaradi razmeroma zamotanega izračunavanja in težjega tolmačenja, geometrijsko sredino ne izračunavamo pogosto. V poštev pride pri izračunavanju centralne tendence za J porazdelitev, če se zanje logaritmi individualnih vrednosti porazdeljujejo v normalni porazdelitvi ali v porazdelitvi, ki je normalni podobna.

6.33

Razen tega pa je geometrijska sredina logično upravičena za izračunavanje povprečnega koeficienta dinamike pri proučevanju časovnih vrst. Vzemimo časovno vrsto, ki ima prvi člen Y_0 , zadnji člen Y_N in zaporedne koeficiente dinamike $k_1 = Y_1/Y_0$, $k_2 = Y_2/Y_1$, ..., $k_N = Y_N/Y_{N-1}$. Vprašanje je, s kakšnim stalnim povprečnim koeficientom dinamike bi se moral razvijati pojav, da bi iz začetne vrednosti Y_0 po N razdobjih bil pojav na ravni Y_N .

Glede na definicijo koeficientov dinamike in povprečnega koeficienta dinamike veljajo med zgornjimi količinami na-

slednje zveze

$$Y_N = Y_0 \cdot k_1 \cdot k_2 \cdot \dots \cdot k_N = Y_0 \cdot k \cdot k \cdot \dots \cdot k = Y_0 \cdot k^N \quad (6.28)$$

Če začetno vrednost Y_0 postopoma množimo z individualnimi koeficienti dinamike, dobimo končno vrednost Y_N . Zaradi definicije povprečnega koeficienta dinamike k pa enako dobimo končno vrednost Y_N , če začetno vrednost Y_0 N krat pomnožimo s povprečnim koeficientom dinamike k . Iz zveze (6.28) dobimo dva obrazca, po katerih izračunavamo povprečen koeficient dinamike

$$k = \sqrt[N]{Y_N / Y_0} \quad (6.29)$$

$$k = \sqrt[N]{k_1 \cdot k_2 \cdot \dots \cdot k_N} \quad (6.30)$$

Po obrazcu 6.29 je povprečen koeficient dinamike N -ti koren iz kvocienta med začetnim in končnim členom v časovni vrsti. V splošnem se stopnja korena ravna po časovnem razmaku med začetnim in končnim členom. Stopnja korena je enaka številu osnovnih razmakov med začetno in končno vrednostjo pojava, za katerega izračunavamo povprečen koeficient dinamike.

Po obrazcu 6.30 pa je povprečen koeficient dinamike geometrijska sredina iz individualnih koeficientov dinamike v razdobju, za katerega iščemo povprečje.

6.34

Vzemimo za primer razvoj površine gozdnih drevesnic v FLRJ.

Konec leta 1955 je bilo $Y_{55} = 1383$ ha drevesnic, konec leta 1959 pa $Y_{59} = 2677$ ha. Povprečen letni koeficient dinamike v razdobju 1955-1959 je po obrazcu 6.29 enak

$$k = \sqrt[4]{Y_{59} / Y_{55}} = \sqrt[4]{\frac{2677}{1383}}$$

Povprečen koeficient dinamike izračunajmo z logaritmi. Če levo in desno stran zgornje enačbe logaritmiramo, dobimo

$$\begin{aligned} \log k &= \frac{1}{4} (\log 2677 - \log 1383) = \\ &= \frac{1}{4} (3,42765 - 3,14082) = 0,07171 \end{aligned}$$

Z antilogaritmiranjem dobimo, da je $k = 1,1795$. Površina gozdnih drevesnic se je v razdobju 1955-1959 vsako leto povečala povprečno za ca 18 %.

Za isti problem poiščimo povprečen koeficient dinamike, če poznamo individualne koeficiente dinamike o razvoju drevesnic v razdobju 1955-1959. Koeficienti dinamike v tem razdobju so $k_{56} = 1,067$, $k_{57} = 1,077$, $k_{58} = 1,246$, $k_{59} = 1,352$

Po obrazcu 6.30 je povprečen koeficient dinamike

$$k = \sqrt[4]{k_{56} \cdot k_{57} \cdot k_{58} \cdot k_{59}} = \sqrt[4]{1,067 \cdot 1,077 \cdot 1,246 \cdot 1,352}$$

Tudi v tem primeru si pomagajmo z logaritmi! Če levo in desno stran logaritmiramo, dobimo

$$\begin{aligned} \log k &= \frac{1}{4} (\log 1,067 + \log 1,077 + \log 1,246 + \log 1,352) = \\ &= \frac{1}{4} (0,02816 + 0,03222 + 0,09552 + 0,13098) = \\ &= 0,07172 \end{aligned}$$

Z antilogaritmiranjem dobimo, da je $k = 1,1796$. Razlika v rezultatu izvira iz zaokroževanja.

Zveze med srednjimi vrednostmi

6.35

Vse štiri obravnavane izračunane srednje vrednosti so podobno definirane. Za vse velja, da so povprečja določenih transformiranih podatkov enaka isti transformaciji iz ustrezne sredine. Tako je aritmetična sredina povprečje iz osnovnih podatkov. Povprečje iz recipročnih podatkov je enako recipročni vrednosti iz harmonične sredine. Povprečje iz kvadratov osnovnih podatkov je enako kvadratu iz kvadratične sredine, povprečje iz logaritmov osnovnih vrednosti pa logaritmu iz geometrijske srednje vrednosti.

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum x_i \\ \frac{1}{H} &= \frac{1}{N} \sum \frac{1}{x_i} \\ K^2 &= \frac{1}{N} \sum x_i^2 \\ \log G &= \frac{1}{N} \sum \log x_i \end{aligned} \tag{6.31}$$

Če izrazimo vsako izmed navedenih sredin v eksplicitni obliki, dobimo obrazce, ki smo jih navedli kot definicije posameznih sredin.

6.36

Če iz istih podatkov izračunamo vse štiri izračunane srednje vrednosti: K , $\bar{x}$, G in H , velja, da je

$$K \geq \bar{x} \geq G \geq H \quad (6.32)$$

Preiskusimo navedeno pravilo na shematičnem primeru za dve vrednosti: 1 in 4.

$$K = \sqrt{\frac{1 + 4^2}{2}} = 2,9 ;$$

$$\bar{x} = \frac{1 + 4}{2} = 2,5 ;$$

$$G = \sqrt{1 \cdot 4} = 2 ;$$

$$H = \frac{2}{\frac{1}{1} + \frac{1}{4}} = 1,6$$

Vse štiri sredine so med seboj enako le, če so vse individualne vrednosti med seboj enake.

6.37

Med aritmetično sredino $\bar{x}$ in obema srednjima vrednostima po legi, mediano M_e in modusom M_o , veljajo tudi določeni stalni odnosi. Za simetrične porazdelitve so aritmetična sredina $\bar{x}$, modus M_o in mediana M_e med seboj enake ($\bar{x} = M_e = M_o$). Za porazdelitve, ki so asimetrične v levo, je aritmetična sredina manjša, modus pa večji kot mediana ($\bar{x} < M_e < M_o$). Za porazdelitve, ki so asimetrične v desno, pa je modus manjši, aritmetična sredina pa večja kot mediana ($M_o < M_e < \bar{x}$). Razlike so tem večje, čim večja je stopnja asimetrije.

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

$$(\bar{x} - M_o) = 3(\bar{x} - M_e) \quad (6.33)$$

Za ne preveč asimetrične porazdelitve moremo s tem obrazcem posredno izračunati modus iz negrupiranih podatkov. Če za populacijo izračunamo aritmetično sredino $\bar{x}$ in določimo

mediano Me , izračunamo iz teh dveh srednjih vrednosti modus Mo po obrazcu

$$Mo = \bar{x} - 3(\bar{x} - Me) \quad (6.34)$$

ki ga dobimo iz obrazca 6.33.

7. MERE VARIACIJE

7.1

Že v prejšnjih poglavjih smo poudarili, da je ena izmed osnovnih značilnosti statističnih znakov variabilnost. Zaradi vpliva najrazličnejših faktorjev se značilnosti pojavov od enote do enote spreminjajo. Srednja vrednost je izraz splošnih vplivov, ki z enakim učinkom vplivajo na vsako enoto. Tako je povprečna višina dreves rezultat splošnih vplivov (starosti, vrste sestoja itd.). Zaradi individualnih vplivov (individualna lega dreves, zasenčenost, kvaliteta sadike itd.) pa se višine dreves med seboj razlikujejo. Čim večji so individualni vplivi, tem večje so razlike med drevesi, oziroma odkloni višin posameznih dreves od povprečne višine. Variabilnost oziroma učinek individualnih vplivov moremo torej meriti z odkloni od sredine. Z merami variacije, s katerimi z eno količino merimo velikost variacije v populaciji, pa merimo učinek individualnih vplivov na ves sestoj, enako kot merimo s srednjimi vrednostmi učinek splošnih vplivov. Proučevanje variabilnosti v najrazličnejših oblikah je osnova dobršnega dela statističnega proučevanja.

Vrste mer variacije

7.2

Kot imamo več vrst srednjih vrednosti, imamo tudi več vrst mer variacije, od katerih vsaka ima svoje prednosti in pomanjkljivosti. Vsaka izmed njih na svoj način meri variabilnost pojava. Prednost enih je v lažjem razumevanju in določanju, prednost drugih pa v večji analitični vrednosti.

Najobičajnejše mere variacije so:

- a) variacijski razmak R
- b) kvartilni odklon Q
- c) povprečen absolutni odklon AD
- d) standardni odklon σ oziroma varianca σ^2 , iz katere je standardni odklon izpeljan.

Od navedenih mer variacije sta prvi dve dani z lego, zadnji dve pa sta izračunani iz podatkov za vse enote populaci-

je. Zaradi tega sta solidnejši in stabilnejši kot mere variacije po legi. Zaradi posebnih lastnosti, ki so v zvezi s teorijo statistike, pa je najvažnejša mera variacije varianca oziroma standardni odklon, ki je izpeljan iz variance.

Variacijski razmak

7.3

Najenostavnejša, a tudi najslabša mera variacije je variacijski razmak

$$R = x_{\max} - x_{\min} \quad (7.1)$$

ki je razlika med največjo in najmanjšo vrednostjo v populaciji.

Iz ranžirne vrste premerov za 19 modelnih smrekovih dreves v tabeli 5.2 povzamemo, da je premer najtanjšega drevesa $x_{\min} = 14$ cm, premer najdebelejšega drevesa pa $x_{\max} = 44$ cm. Variacijski razmak je torej

$$R = x_{\max} - x_{\min} = 44 - 14 = 30 \text{ cm}$$

Premeri devetnajstih modelnih dreves variirajo torej v razmaku 30 cm.

Včasih nakažemo variacijski razmak tudi tako, da povemo najmanjšo in največjo vrednost v populaciji. Tak način da večjo informacijo kot sam razmak variacije, ki ga moremo iz teh dveh podatkov neposredno izračunati.

7.4

Variacijski razmak je odvisen samo od obeh skrajnih vrednosti v populaciji. Zato je ta mera variacije zelo podvržena individualnim vplivom, ker so ekstremne vrednosti dostikrat izraz netipičnih vplivov. Razen tega pa variacijski razmak ni odvisen od razmestitve vseh drugih vrednosti v razmaku variacije, ki bistveno vplivajo na variabilnost.

Prednost variacijskega razmaka pa je v tem, da je treba zanj poznati samo obe skrajni vrednosti, kateri zlahka določimo, če so vrednosti urejene v ranžirni vrsti. Zato zaradi lahkega določanja uporabljamo variacijski razmak vselej,

kadar hočemo predvsem za manjše populacije hitro in orientacijsko oceniti variabilnost pojava.

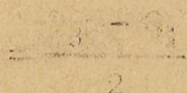
Kvartilni odklon

7.5

Hibe, ki jih ima variacijski razmak, deloma odpravimo s kvartilnim odklonom. Osnovna hiba variacijskega razmaka izvira iz tega, da je odvisen samo od skrajnih, netipičnih vrednosti. To hibo odpravimo tako, da variabilnost merimo z razmakom, v katerem niso vse, temveč samo del enot v populaciji, brez ekstremnih vrednosti. Tako moremo meriti variabilnost z razmakom med prvim in devetim decilom. V tem razmaku je 80 % vseh vrednosti in je iz njega izločeno 10 % najmanjših in 10 % največjih vrednosti.

Pogosteje kot z decilnim razmakom merimo variabilnost s kvartilnim razmakom, ki določa razmak, v katerem je 50 % osnovnih vrednosti iz populacije. Iz kvartilnega razmaka je izločena četrtna najmanjših in četrtna največjih vrednosti.

Običajno pa namesto kvartilnega razmaka $Q_3 - Q_1$ vzamemo za mero variacije kvartilni odklon



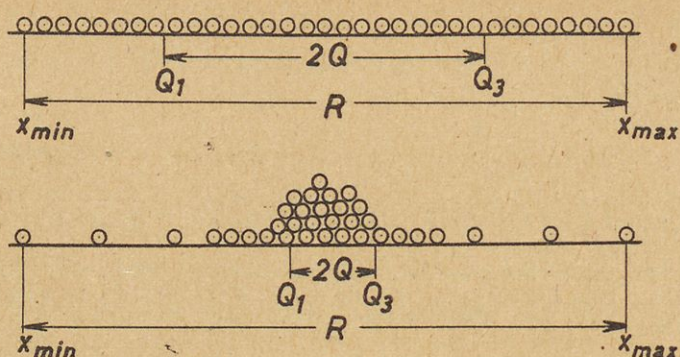
$$Q = \frac{Q_3 - Q_1}{2} \quad (7.2)$$

ki je polovica kvartilnega razmaka.

Kvartilni razmak oziroma kvartilni odklon pravilneje pokažeta variabilnost v populaciji kot variacijski razmak, ker sta določena iz stabilnega dela vrednosti v populaciji.

V sliki 7.1 je nakazano, kako je variacijski razmak neobčutljiv za različno razmestitev vrednosti v variacijskem razmaku, kako pa različna razmestitev vpliva na kvartilni razmak oziroma kvartilni odklon.

Vendar pa tudi kvartilni odklon nima vseh prednosti, ki jih imajo izračunane mere variacije, ki so odvisne od vseh vrednosti v populaciji, ki najbolj natančno merijo variabilnost.



Slika 7.1 R in Q pri različnih razmestitvah

7.6

Izračunavanje kvartilnega odklona sovpada z izračunavanjem kvartilov, ki so posebna vrsta kvartilov, katere smo obravnavali v posebnem poglavju.

Za premere 19 modelnih smrekovih dreves, za katere smo že izračunali variacijski razmak, so v tabeli 5.3 izračunani kvartili. Za to populacijo je $Q_1 = 21,50$ cm in $Q_3 = 32,75$ cm. Kvartilni odklon pa je $Q = (32,75 - 21,50)/2 = 5,62$ cm.

Povprečen absoluten odklon

7.7

Ker je višina posameznega drevesa x_i izraz delovanja individualnih in splošnih faktorjev na drevo, aritmetična sredina pa rezultat splošnih vplivov, sklepamo, da je razlika $e_i = x_i - \bar{x}$ rezultat individualnih vplivov na enoto i . Odkloni od aritmetične sredine so pozitivni ali negativni. Za merjenje jakosti učinka individualnih vplivov na enoto i pa predznak ni pomemben. Zato vzamemo za merilo jakosti individualnih vplivov na enoto i absoluten odklon individualne vrednosti od sredine $|x_i - \bar{x}|$. Za skupno merilo jakosti individualnih vplivov na vse enote pa dobro služi aritmetična sredina iz individualnih absolutnih odklonov. Povprečen absoluten odklon je torej

$$AD_{\bar{x}} = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}| \quad (7.3)$$

Povprečen absoluten odklon od aritmetične sredine je mera variacije, ki je odvisna od vseh vrednosti v populaciji in je torej občutljivejša in objektivnejša mera variabilnosti kot variacijski razmak ali kvartilni odklon.

Ni nujno, da povprečen absoluten odklon izračunamo iz odklonov od aritmetične sredine. Možno ga je izračunati iz absolutnih odklonov od katerekoli srednje vrednosti. izkaže se, da je teoretično in praktično celo bolj utemeljeno, da ga izračunamo iz odklonov od mediane. Povprečen absoluten odklon je namreč najmanjši

$$AD_{Me} = \frac{1}{N} \sum_{i=1}^N |x_i - Me| \quad (7.4)$$

če ga izračunamo iz odklonov od mediane.

7.8

Vzemimo v potrditev zgornje trditve shematičen primer petih dreves, za katere poznamo premere: 27 cm, 29 cm, 33 cm, 34 cm, 37 cm. Zanje izračunajmo povprečen absoluten odklon od aritmetične sredine $AD_{\bar{x}}$ in od mediane AD_{Me} ! Za zgornjih pet premerov je $\bar{x} = 32$ cm in $Me = 33$ cm.

Tabela 7.1 Izračun povprečnega absolutnega odklona $AD_{\bar{x}}$ in AD_{Me} za premere petih dreves (shematičen primer)

x_i	$ x_i - \bar{x} $	$ x_i - Me $
27	5	6
29	3	4
33	1	0
34	2	1
37	5	4

$$\sum |x_i - \bar{x}| = 16 \quad 15 = \sum |x_i - Me|$$

$$AD_{\bar{x}} = \frac{1}{N} \sum |x_i - \bar{x}| = 16/5 = 3,2 \text{ cm}; \quad AD_{Me} = \frac{1}{N} \sum |x_i - Me| = 15/5 = 3,0 \text{ cm}$$

Resnično je $AD_{\bar{x}} = 3,2$ cm večji kot $AD_{Me} = 3,0$ cm

7.9

Teoretično najbolj utemeljeno in najboljše merilo variabilnosti je varianca ali iz nje izpeljan standardni odklon. Analitično vrednost ima predvsem varianca, kot opisni parameter pa uporabljamo običajno standardni odklon.

Varianca, katero konvencionalno zaznamujemo s σ^2 (sigma kvadrat), je povprečen kvadratični odklon od aritmetične sredine. Iz definicije sledi obrazec

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (7.5)$$

Pri povprečnem absolutnem odklonu jakost individualnih vplivov merimo z absolutnim odklonom od sredine, pri varianci pa s kvadratom odklona od aritmetične sredine. Pri varianci torej smer učinka individualnih vplivov, ki za mero variacije ni bistvena, odpravimo s kvadriranjem.

Varianca je izražena v enoti mere, ki je kvadrat enote mere za osnovni podatek, za katerega jo izračunamo. Zato je varianca kot opisni parameter težko razumljiva in neprikladna. V osnovni enoti mere izraženo mero variacije pa dobimo, če izračunamo iz variance kvadratni koren. Ta parameter, ki ga imenujemo standardni odklon, zaznamujemo pa s σ , je torej

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (7.6)$$

Vse metode analize variabilnosti so zasnovane na analizi odnosov med variancami. Standardni odklon kot opisno merilo variabilnosti pa je posebno važen zaradi zveze z eno izmed najvažnejših teoretičnih frekvenčnih porazdelitev - z normalno porazdelitvijo.

Izračun variance in standardnega odklona

7.10

Izračunavanje variance oziroma standardnega odklona je zahtevnejše kot izračunavanje drugih mer variacije. Ker pa

pri statističnem analiziranju podatkov dnevno naletimo na izračunavanje varianc in standardnih odklonov, je praktično pomembno poznati določene tehnične olajšave, s katerimi pridemo hitreje do rezultata. Enako kot sredine moremo tudi varianco izračunati iz negrupiranih podatkov in oceniti iz grupiranih podatkov.

7.11

Izračun variance in standardnega odklona iz negrupiranih podatkov. Neposredna metoda. Če vzamemo za osnovo obrazec 7.5, izračunamo varianco po naslednjem postopku:

- a) iz osnovnih podatkov izračunamo aritmetično sredino $\bar{x}$
- b) izračunamo individualne odklone posameznih vrednosti od aritmetične sredine $(x_i - \bar{x})$
- c) dobljene individualne odklone kvadriramo $(x_i - \bar{x})^2$
- d) varianco dobimo, če izračunamo povprečje iz kvadratov individualnih odklonov od aritmetične sredine
- e) standardni odklon je kvadratni koren iz variance

$$\sigma_x^2 = \frac{1}{N} \sum (x_i - \bar{x})^2$$

$$\sigma_x = \sqrt{\sigma_x^2}$$

7.12

Za primer izračunajmo varianco za premere sedmih dreves: 31 cm, 27 cm, 35 cm, 32 cm, 29 cm, 36 cm in 32 cm. Po zgornjem pravilu izračunamo najprej aritmetično sredino

$$\bar{x} = (31+27+35+32+29+36+32) : 7 = 222 : 7 = 31,71 \text{ cm}$$

Nadaljnji postopek je nakazan v tabeli 7.2

Postopek je jasn. Edina težava je v tem, da so običajno aritmetične sredine decimalna števila. Zato so odkloni od aritmetične sredine večmestna decimalna števila, zaradi česar so težave s kvadriranjem.

Tabela 7.2 Izračun variance in standardnega odklona po neposredni metodi za premere sedmih dreves

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	31	-0,71	0,5041
2	27	-4,71	22,1841
3	35	+3,29	10,8241
4	32	+0,29	0,0841
5	29	-2,71	7,3441
6	36	+4,29	18,4041
7	32	+0,29	0,0841
222		+0,01	$59,4287 = \sum (x_i - \bar{x})^2$

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{N} = \frac{59,4287}{7} = 8,4898$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{8,4898} = 2,91 \text{ cm}$$

7.13

Metoda pomožnega znaka u. Zgornjo hibo neposredne metode odpravimo z metodo pomožnega znaka u. Zlahka moremo dokazati, da je varianca tudi

$$\sigma_x^2 = \frac{\sum u_i^2 - U^2/N}{N} \quad (7.7)$$

Pri tem je razen znanih oznak pomožni znak $u_i = x_i - x_0$ odklon individualnih vrednosti x_i od poljubne vrednosti x_0 , $U = \sum u_i$ vsota vseh vrednosti u_i , $\sum u_i^2$ pa vsota kvadratov za pomožni znak u_i za vse enote v populaciji.

S tem obrazcem izračunamo varianco po tehle točkah:

a) glede na osnovne vrednosti x_i izberemo poljubno okroglo vrednost x_0 nekje med stvarnimi vrednostmi, s čemer zmanjšamo vrednosti za pomožni znak u_i ;

b) izračunamo posamezne odklone $u_i = x_i - x_0$. Dobljene vrednosti u_i kvadriramo: u_i^2 ;

c) poiščemo vsoto: $U = \sum u_i$ in vsoto kvadratov $\sum u_i^2$;

d) dobljeni vsoti vnesemo v obrazec 7.7.

Z metodo pomožnega znaka u se izognemo zamudnemu kvadriranju in seštevanju večmestnih števil, kar je hiba neposredne metode.

Če ni moč najti neke vrednosti, ki bi mogla služiti kot pomožno izhodišče za znak u , vzamemo, da je $x_0 = 0$. V tem primeru obrazec 7.7 preide v

$$s_x^2 = \frac{\sum x_i^2 - X^2/N}{N} \quad (7.8)$$

Pri tem pomeni: $X = \sum x_i$ vsota osnovnih podatkov, $\sum x_i^2$ = vsota kvadratov osnovnih podatkov, N = število enot v populaciji.

Ta obrazec je primeren za izračunavanje variance posebno v primerih, če jo računamo z računskim strojem in kvadriranje ni poseben problem.

7.14

Za primerjavo z neposredno metodo izračunajmo po metodi pomožnega znaka u varianco za iste podatke. Izračun variance po metodi pomožnega znaka u je nakazan v tabeli 7.3. Iz pregleda osnovnih vrednosti povzamemo, da je najugodnejše, če vzamemo, da je $x_0 = 32$ cm.

Tabela 7.3 Izračun variance za premere 7 dreves po metodi pomožnega znaka u

i	x_i	$u_i = x_i - x_0$	u_i^2
1	31	-1	1
2	27	-5	25
3	35	+3	9
4	32	0	0
5	29	-3	9
6	36	+4	16
7	32	0	0

$$U = -2 \quad 60 = \sum u_i^2$$

Če vnesemo dobljene vzgledne rezultate v obrazec 7.7, dobimo

$$\sigma_x^2 = \frac{\sum u^2 - U^2/N}{N} = \frac{60 - (-2)^2/7}{7} = 8,4898$$

Dobljeni rezultat se sklada z rezultatom, ki smo ga dobili po neposredni metodi. Ker smo dobili z znatno lažjim postopkom isti rezultat, metodo pomožnega znaka v praksi uporabljamo skoro izključno.

7.15

Izračun variacije in standardnega odklona iz frekvenčnih porazdelitev. Za populacije z velikim številom enot moremo tudi varianco oceniti iz frekvenčne porazdelitve. Ocena variance iz frekvenčne porazdelitve je po obrazcu

$$\sigma_x^2 = \frac{1}{N} \sum_{k=1}^r f_k (x_k - \bar{x})^2 \quad (7.9)$$

ponderirana aritmetična sredina kvadratov odklonov sredin razredov x_k od aritmetične sredine $\bar{x}$. Pri tem so frekvence ustreznih razredov ponderi. Ocena variance po tem obrazcu je tem boljša, čim manjši so razredi, ker frekvenčna porazdelitev z ožjimi razredi verneje kaže sliko vseh vrednosti v populaciji. Iz istih razlogov kot pri negrupiranih podatkih pa je izračunavanje variance po osnovnem obrazcu 7.9 zamudno in neprikladno. Zato ga v splošnem v praksi ne uporabljamo. Pač pa v praksi uporabljamo za frekvenčne porazdelitve, ki imajo enake razrede, dve metodi: metodo pomožnega znaka u in metodo kumulativ. Pri obeh razširimo načeli, ki smo ju uporabili že pri izračunavanju aritmetične sredine iz frekvenčnih porazdelitev.

7.16

Metoda pomožnega znaka u. Pomožni znak u, ki je z osnovnim znakom x v zvezi

$$x = x_0 + i \cdot u \quad (7.10)$$

smo uporabili že pri izračunavanju aritmetične sredine iz frekvenčnih porazdelitev. Sredine razredov v frekvenčni porazdelitvi za znak u so ... -3, -2, -1, 0, +1, +2, +3 ..., pri čemer velja vrednost 0 za razred, za katerega je sredina razreda x_0 .

Po metodi pomožnega znaka u izračunamo varianco po naslednjem postopku.

- V frekvenčni porazdelitvi upeljemo pomožni znak u ... -3 -2 -1 0 +1 +2 +3 ...;
- frekvence f_k pomnožimo z ustreznimi vrednostmi u_k . Tako dobimo produkte $f_k u_k$;
- dobljene produkte $f_k u_k$ ponovno pomnožimo z ustreznimi vrednostmi u_k , da dobimo vrednosti $f_k u_k^2$;
- izračunamo vsoti $\sum f_k u_k = U$ in $\sum f_k u_k^2$;
- dobljene izraze vnesemo v obrazce

$$K = \sum f_k u_k^2 - \frac{U^2}{N} ; \quad \sigma^2 = \frac{1}{N} \cdot K \quad (7.11)$$

in dobimo varianco.

7.17

Izračunavanje variance po metodi pomožnega znaka ponaazorimo na frekvenčni porazdelitvi premerov v čistem smrekovem sestoju A na Pokljuki iz tabele 3.4. Izračun je prikazan v tabeli 7.4.

Tabela 7.4 Izračun variance in standardnega odklona po metodi pomožnega znaka u za premere dreves v čistem smrekovem sestoju A na Pokljuki

Premers	f	u	fu	fu ²
15 cm - 19 cm	4	-4	-16	64
20 cm - 24 cm	6	-3	-18	54
25 cm - 29 cm	55	-2	-110	220
30 cm - 34 cm	118	-1	-118	118
35 cm - 39 cm	112	0	0	0
40 cm - 44 cm	109	+1	+109	109
45 cm - 49 cm	61	+2	+122	244
50 cm - 54 cm	28	+3	+84	252
55 cm - 59 cm	9	+4	+36	144
60 cm - 64 cm	5	+5	+25	125

$$N=507$$

$$U = \sum fu = +114$$

$$1330 = \sum fu^2$$

Če dobljene podatke vnesemo v obrazca 7.9, dobimo:

$$K = \sum fu^2 - U^2/N = 1330 - (+114)^2/507 = 1304,3669$$

$$s^2 = i^2 \cdot K/N = 5^2 \cdot 1304,3669/507 = 64,3178$$

ocena standardnega odklona pa je

$$s = \sqrt{s^2} = \sqrt{64,3178} = 8,02 \text{ cm}$$

7.18

Metoda kumulativ. Podobno kot aritmetično sredino moremo tudi varianco izračunati z metodo kumulativ. Po metodi kumulativ izračunamo varianco tako, da:

a) Iz frekvenc f v frekvenčni porazdelitvi izračunamo enako kot pri izračunavanju aritmetične sredine prvo kumulativo F .

b) Za izračun variance pa potrebujemo še drugo kumulativno vrsto frekvenc FF . To dobimo iz prve kumulativne vrste frekvenc F , če postopek kumulativnega seštevanja ponovimo na prvi kumulativni vrsti frekvenc F .

c) Zadnji člen (pod črto) v prvi kumulativni vrsti je enak obsegu populacije N , zadnji člen v drugi kumulativni vrsti (pod črto) je količina A , vsota členov iz druge kumulativne vrste FF (brez zadnjega člena pod črto) pa je količina B .

d) Iz dobljenih podatkov dobimo varianco po obrazcih:

$$K = 2B + A - A^2/N; \quad s_x^2 = i^2 \cdot K/N \quad (7.12)$$

7.19

Prednosti postopka kumulativ najlepše ponazorimo na primeru. Zaradi kontrole vzemimo isti primer kot pri postopku pomožnega znaka u .

Dobljene pomožne rezultate iz tabele 7.5 vnesimo v obrazec 7.10.

$$K = 2 \cdot B + A - A^2/N = 2 \cdot 5222 + 2421 - 2421^2/507 = 1304,3669$$

$$s_x^2 = i^2 \cdot K/N = 5^2 \cdot 1304,3669/507 = 64,3178$$

$$s_x = \sqrt{s_x^2} = \sqrt{64,3178} = 8,02 \text{ cm}$$

Tabela 7.5 Izračun variance za premere v čistem smrekovem sestoju A na Pokljuki iz tabele 3.4

Premjer v cm	f	F	FF	
15 - 19	4	0	0	
20 - 24	6	4	4	
25 - 29	55	10	4	
30 - 34	118	65	14	
35 - 39	112	183	79	$4+14+79...+1426+1919=$ $= 5222 = B$
40 - 44	109	295	262	
45 - 49	61	404	557	
50 - 54	28	465	961	
55 - 59	9	493	1426	
60 - 64	5	502	1919	
		507	2421	
		N	A	

Rezultat je skladen z rezultatom, ki smo ga dobili po metodi pomožnega znaka u . Primerjava obeh metod pa govori v prid metode kumulativ, ker odpade za izračunavanje pomožnih količin vsako množenje. Razen tega pa dobimo kot postranski rezultat še kumulativno vrsto frekvenc, ki more služiti za analizo frekvenčne porazdelitve.

7.20

Sheppardov popravek. Ker je frekvenčna porazdelitev le približna slika vrednosti v populaciji, ki je tem natančnejša, čim ožji so razredi, je tudi varianca, izračunana iz frekvenčne porazdelitve, le ocena variance, ki jo dobimo, če jo izračunamo iz individualnih vrednosti. Za unimodalne, ne preveč asimetrične porazdelitve, se pri izračunavanju aritmetične sredine iz frekvenčnih porazdelitev učinek grupiranja v razrede v vsoti izravna. Pri izračunavanju variance iz frekvenčnih porazdelitev pa dobimo sistematično preveliko oceno. Učinek grupiranja je odvisen od širine razredov in znaša $i^2/12$. Oceno variance za unimodalne, ne preveč asimetrične porazdelitve, izračunano po prejšnjih postopkih, popravimo tako, da jo zmanjšamo za popravni člen $i^2/12$. Tako dobimo obrazec

$$\sigma_{pop}^2 = \sigma_x^2 - i^2/12$$

(7.13)

Ta popravek imenujemo Sheppardov popravek.

Čisti smrekov sestoj A na Pokljuki izpolnjuje pogoj za uporabo popravka, ker je porazdelitev unimodalna in se frekvenca v najnižjih in najvišjih razredih približujejo nič. Popravljen varianca je torej

$$\sigma_{pop}^2 = \sigma_x^2 - i^2/12 = 64,3178 - 5^2/12 = 62,2345$$

popravljen standardni odklon pa

$$\sigma_{pop} = \sqrt{\sigma_{pop}^2} = \sqrt{62,2345} = 7,89 \text{ cm}$$

Popravljen varianca je od variance $\sigma_o^2 = 62,3190$, ki je izračunana iz osnovnih, negrupiranih vrednosti, različna le za 0,141 %, medtem ko je nepopravljena ocena variance prevelika za 3,2 %.

7.21

Zveza standardnega odklona z normalno porazdelitvijo.

Variacijski razmak, kvartilni odklon in tudi povprečen absolutni odklon so mere variacije, za katere je pomen in smisel dokaj jasen. Navidezno pa je manj jasen standardni odklon, dokler ga ne obravnavamo v zvezi z lastnostmi nekih frekvenčnih porazdelitev.

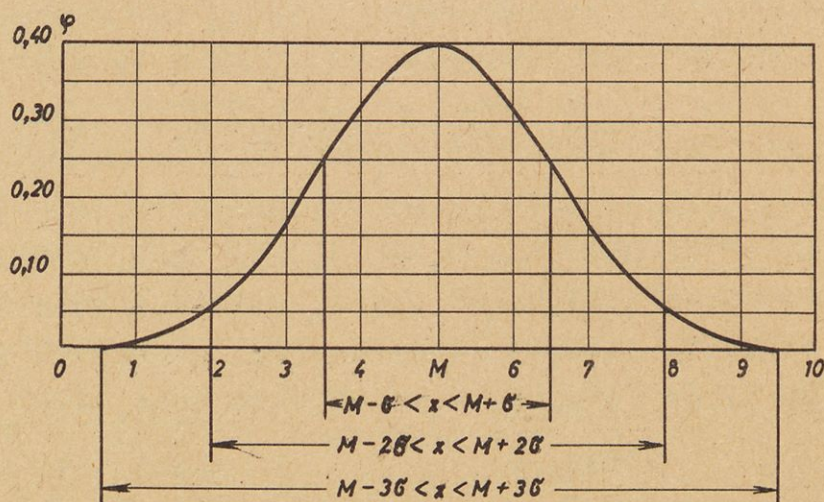
Čeprav kasneje obravnavamo normalno porazdelitev posebej, navedimo nekaj lastnosti normalne porazdelitve v zvezi s standardnim odklonom. Normalna porazdelitev je ena izmed osnovnih teoretičnih porazdelitev in se v približku pod določenimi pogoji pogosto pojavlja tudi v praksi. Vse normalne porazdelitve so si med seboj podobne, med seboj se razlikujejo samo po aritmetični sredini in standardnem odklonu. Vsaka normalna porazdelitev je določena z aritmetično sredino in standardnim odklonom. Čeprav so posamezne normalne porazdelitve glede na različne aritmetične sredine in standardne odklone različne, velja splošna zakonitost, da je v razmaku

$\bar{x} - 6$ do $\bar{x} + 6$ 68,27 % ali okroglo 2/3 vseh vrednosti; v razmaku

$\bar{x} - 2\sigma$ do $\bar{x} + 2\sigma$ 95,45 % ali okroglo 19/20 vseh vrednosti in v razmaku

$\bar{x} - 3\sigma$ do $\bar{x} + 3\sigma$ 99,73 % ali praktično vse vrednosti v populaciji.

Ti odnosi veljajo strogo za normalno porazdelitev. Približno pa veljajo tudi za unimodalne, simetrične in zvonaste stvarne porazdelitve.



Slika 7.2 Frekvence za normalno porazdelitev v razmakih $\bar{x} \pm \sigma$, $\bar{x} \pm 2\sigma$ in $\bar{x} \pm 3\sigma$

Skupna varianca

7.22

Prednost variance pred drugimi merami variacije je tudi v tem, da moremo iz podatkov o delnih populacijah izračunati skupno varianco, če poznamo za delne populacije, ki sestavljajo skupno populacijo, število enot N_k , aritmetične sredine $\bar{x}_k$ in variance σ_k^2 . Te možnosti nimamo za nobeno drugo mero variacije. Če upoštevamo osnovne obrazce o varianci, moremo dokazati, da je skupna varianca σ^2 enaka

$$\sigma_x^2 = \frac{1}{N} \sum_{k=1}^r N_k (\bar{x}_k - \bar{x})^2 + \frac{1}{N} \sum_{k=1}^r N_k \sigma_k^2 \quad (7.14).$$

Prvi člen v obrazcu je tehtan povprečen kvadratični odklon grupnih sredin $\bar{x}_k$ od skupne aritmetične sredine za celo populacijo $\bar{x}$. Drugi člen pa je tehtana aritmetična sredina iz grupnih varianc σ_k^2 .

Za primer izračunajmo skupno varianco za premere v borovem sestoju, ki je razdeljen po kakovosti v pet delov. Za ta sestoj so osnovni podatki dani v tabeli 7.6. V tej tabeli je tudi nakazan izračun za skupno varianco.

Tabela 7.6 Izračun skupne variance za premere borovega sestojja iz grupnih podatkov za pet oddelkov
(Vir: R. Frauendorfer: Planung und Durchführung von Stichprobenahmen, Wien, München 1957)

Del	N_k	$\bar{x}_k$	σ_k^2	$N_k \bar{x}_k$	$\bar{x}_k - \bar{x}$	$(\bar{x}_k - \bar{x})^2$	$N_k (\bar{x}_k - \bar{x})^2$	$N_k \sigma_k^2$
I	64	30,19	16,65	1932,16	+8,65	74,8225	4788,6400	1065,60
II	572	24,20	12,50	13842,40	+2,66	7,0756	4047,2432	7150,00
III	324	19,35	10,45	6269,40	-2,19	4,7961	1553,9364	3385,80
IV	192	16,63	10,26	3192,96	-4,91	24,1081	4628,7552	1969,92
V	50	13,00	7,04	650,00	-8,54	72,9316	3646,5800	352,00
	1202			25886,92			18665,1548	13923,32
	N			$\sum N_k \bar{x}_k$			$\sum N_k (\bar{x}_k - \bar{x})^2$	$\sum N_k \sigma_k^2$

$$\bar{x} = \frac{\sum N_k \bar{x}_k}{N} = \frac{25886,92}{1202} = 21,54$$

$$\begin{aligned} \sigma_x^2 &= \frac{1}{N} \sum N_k (\bar{x}_k - \bar{x})^2 + \frac{1}{N} \sum N_k \sigma_k^2 = \frac{18665,1548}{1202} + \frac{13923,32}{1202} = \\ &= 15,5284 + 11,5835 = 27,1119 \end{aligned}$$

Obrazec 7.14 za izračunavanje skupne variance pa ni važen le, ker moremo z njim iz grupnih podatkov izračunati skupno varianco, temveč predvsem zato, ker moremo z njim skupno varianco razdeliti v dva dela: v varianco med grupami in v varianco v grupah. Ker so grupne aritmetične sredine izraz splošnih pogojev v grupah, so razlike med grupnimi aritmetičnimi sredinami rezultat razlik v pogojih med grupami, torej rezultat vpliva faktorjev, po katerem je populacija razdeljena v grupe. Varianca med aritmetičnimi sredinami torej meri jakost faktorja, po katerem je populacija razdeljena v grupe. Od skup-

ne variance $\sigma_x^2 = 27,1119$ izvira $\sigma_x^2 = 15,5284$ variance iz razlik v kakovosti posameznih delov sestoja. Zato imenujemo ta del variance pojasnjena varianca. Če razlike v delovanju faktorjev po delih ne bi vplivale na premere dreves, bi bile povprečne debeline v vseh delih enake, varianca med grupnimi sredinami pa nič. Obratno pa so grupne variance rezultat vplivov drugih, individualnih faktorjev, ki vplivajo na premere dreves v posameznih delih. Povprečna varianca iz varianc v grupah je torej merilo jakosti individualnih nepojasnjenih vplivov v grupah populacije. Zato imenujemo to varianco nepojasnjena varianca. V skrajnem primeru, da na drevesa razen kakovosti tal ne bi vplival noben drug spreminjajoč faktor, bi bili premeri vseh dreves v posameznih grupah enaki. V tem primeru bi bile grupne variance in povprečna varianca iz njih enaka nič.

Skupno varianco moremo torej po zgornjem postopku razdeliti v del, ki izvira iz vpliva faktorjev, po katerem smo populacijo razdelili v grupe, in varianco, ki je rezultat ostalih individualnih vplivov. Zato moremo obrazec 7.12 pisati analitično v obliki

$$\sigma_{A1}^2 = \sigma_A^2 + \sigma_1^2 \quad (7.15)$$

Skupna varianca σ_{A1}^2 , ki je rezultat faktorjev A in individualnih vplivov I, je vsota variance σ_A^2 , ki izvira iz faktorja A in variance σ_1^2 , ki izvira iz vplivov individualnih faktorjev.

Tako moremo za poljuben faktor določiti jakost njegovega vpliva na proučevani podatek. Populacijo po tem faktorju razdelimo v grupe, iz njih izračunamo varianco med grupami in varianco v grupah in sklepamo: čim večji del skupne variance odpade na varianco med grupami, tem večji je vpliv proučevanega faktorja na pojav, ki ga proučujemo.

Kolik del skupne variance odpade na varianco med grupami, pa nazorneje prikažemo v relativnem številu, če postavimo skupno varianco enako 1 in izračunamo v delih od enote prispevka vsakega dela variance.

V našem primeru od skupne variance odpade $15,53/27,11 = 0,573$ ali 57,3 % na vpliv razlik v kakovosti delov. Iz tega sklepamo, da je razmeroma velik del skupne variance pojasnjen z razlikami v kakovosti delov.

7.25

Absolutne mere variacije, ki smo jih obravnavali v prejšnjih odstavkih, so primerljive le v omejenem obsegu. Nemo- goče je absolutne mere variacije primerjati za raznovrstne po- jave, čeprav so ti v vsebinski zvezi in bi bila primerjava va- riabilnosti priporočljiva. Absolutne mere variacije pa ne more- mo primerjati med seboj dostikrat niti za istovrstne pojave. Vzemimo za primer debeline dreves. Razlika 3 cm v debelini dveh dreves je za debla, katerih povprečen premer je 10 cm, zelo velika (30 %), medtem ko je za debla, katerih povprečna debelina je 60 cm, ista razlika 3 cm komaj 5 %, torej znatno manj pomembna. Enako velike absolutne razlike so relativno bolj ali manj pomembne. Absolutne mere variacije so torej primerljive le za istovrstne pojave s približno enakimi sredinami. Zato pri analiziranju pojavov dostikrat uporabljamo relativne mere variacije. Te dobimo, če primerjamo absolutno mero variacije z ustrezno srednjo vrednostjo.

Variacijski razmak najpogosteje primerjamo kar s sre- dino med najmanjšo in največjo vrednostjo

$$\frac{2(x_{\max} - x_{\min})}{x_{\max} + x_{\min}} \quad (7.16)$$

kvartilni odklon z mediano ali aritmetično sredino med obema kvartiloma

$$\frac{Q}{Me} ; \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad (7.17)$$

povprečen absoluten odklon pa ali z aritmetično sredino ali pa z mediano

$$AD_M/M \quad AD_{Me}/Me \quad (7.18)$$

odvisno od tega, katera srednja vrednost je osnova za povprečen absolutni odklon.

Ker pa je standardni odklon najsolidnejša mera varia- cije, je tudi relativna mera variacije, ki ima za osnovo stan- dardni odklon, najpomembnejša. Ta koeficient, ki ga imenujemo

koeficient variacije, je razmerje med standardnim odklonom, s katerim merimo jakost individualnih vplivov in aritmetično sredino, ki je odvisna od splošnih vplivov. Koeficient variacije KV^0 je razmerje standardnega odklona in aritmetične sredine

$$KV^0 = \frac{\sigma}{\bar{x}} \quad (7.19)$$

pogosto pa ga izražamo v odstotkih

$$KV\% = 100 \frac{\sigma}{\bar{x}} \quad (7.20)$$

S koeficientom variacije odstranimo vpliv različne ravni za primerjame pojave. Razen tega pa je KV neimenovano število, zato z njim razširimo področje primerljivosti in analize variabilnosti na populacije z različnimi srednjimi vrednostmi in na raznovrstne populacije.

7.26

Vzemimo za primer borov sesto, ki je po kakovosti tal razdeljen v pet oddelkov I, II, III, IV, V. Oddelek I je najboljši, drugi pa so po rangi vedno slabše kakovosti, do najslabšega oddelka V. V tabeli 7.6 so po vrsti vnešeni podatki o povprečnih premerih v prsni višini $\bar{x}$, o standardnih odklonih premerov σ in koeficienti variacije $KV\%$ za vseh pet delov.

Tabela 7.6 Pregled o variabilnosti premerov v prsni višini za borov sesto, razdeljen v pet delov I-V po kakovosti tal (po podatkih R. Frauendorfer: Planung und Durchführung von Stichprobenahmen, Wien-München 1957)

Parameter	D e l				
	I	II	III	IV	V
Povprečen premer $\bar{x}$ cm	30,19	24,20	19,35	16,63	13,00
Standardni odklon σ cm	4,08	3,54	3,23	3,20	2,65
Koeficient variacije $KV\%$	13,5	14,6	16,7	20,5	20,4

Analiza podatkov v tabeli 7.6 pokaže zanimivo sliko. Če analiziramo standardni odklon po oddelkih, opazimo, da se s slabšanjem kakovosti tal manjša tudi standardni odklon. To pa je navidezno v protislovju s pričakovanjem. Boljši deli sesto-

ja imajo verjetno bolj urejene in s tem tudi enotnejše pogoje rasti kot slabši deli sestoja. Če pa standardne odklone primerjamo s povprečnimi premeri, spoznamo, da je standardni odklon $\sigma_v = 2,65$ cm za najslabši del sestoja zagotovo pomembnejši kot absolutno večji standardni odklon $\sigma_l = 4,08$ cm za najboljši del gozda, ker se prvi nanaša na del gozda z majhnim povprečnim premerom, medtem ko so drevesa v najbolj kakovostnem delu v povprečju več kot enkrat debelejša. Pravilno sliko dajo v tem primeru edino relativne mere variacije. Resnično pokažejo koeficienti variacije, ki so izračunani v tretji vrsti tabele, ravno obratno oziroma pravilno sliko. Medtem ko je koeficient variacije za najboljši del najmanjši, se večajo s padanjem v kakovosti tal posameznih delov.

8.1

Do sedaj smo proučevali le posamezne znake populacij in brez zveze z drugimi znaki. Z izoliranim opazovanjem in proučevanjem enega samega znaka prodremo dosti globoko v značilnosti populacij. Vendar se pri tem ne dotaknemo važnega svojstva pojavov, to je medsebojne odvisnosti. Že pri bežnem razmišljanju spoznamo, da so posamezne značilnosti pojavov odvisne od niza faktorjev, ki vplivajo nanj. Tako je volumen drevesa odvisen od starosti, vrste drevja, lege kraja, kjer drevo raste itd. Prirastek volumna v enem letu je odvisen od starosti drevesa, drevesne vrste, povprečne letne temperature, množine padavin itd. Enako opazimo, da je tudi specifična teža lesa odvisna od niza faktorjev, ki vplivajo na drevo itd.

8.2

Funkcijske odvisnosti. Pri funkcijskih odvisnostih vsaki vrednosti neodvisne spremenljivke ustreza ena ali nekaj točno določenih vrednosti za odvisno spremenljivko. Funkcijsko zvezo med x in y izražamo s simbolom

$$y = f(x) \quad (8.1)$$

kar pomeni, da je odvisna spremenljivka y funkcija odvisne spremenljivke x . Pri tem z f na splošno naznačimo pravilo o funkcijski zvezi med x in y .

Tako je na primer z enačbo

$$y = 2 + 3x + x^2 \quad (8.2)$$

dano pravilo, s katerim moremo k vsaki vrednosti x poiskati ustrezno vrednost y .

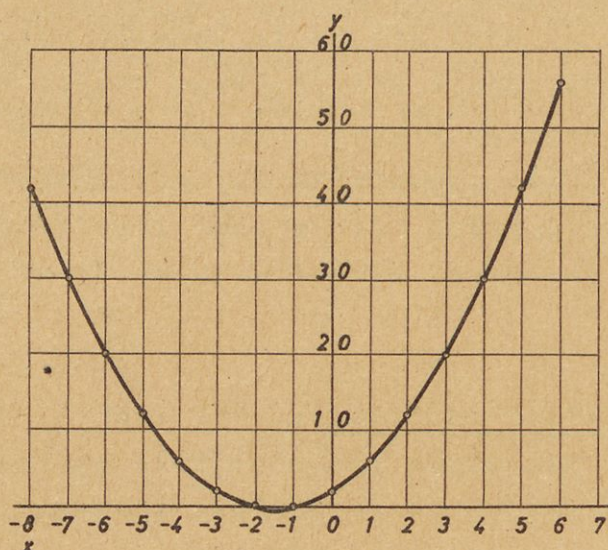
Funkcijsko odvisnost med dvema spremenljivkama moremo podati tudi s tabelo, v kateri so vpisani pari ustreznih posameznih vrednosti za x in y . Tako imamo za zgornji primer v tabeli 8.1 dane dvojice vrednosti x in y za nekaj pozitivnih, celih vrednosti za x .

Tabela 8.1 Tabela funkcije $y = 2 + 3x + x^2$

x	0	1	2	3	4	5	6
y	2	6	12	20	30	42	56

Seveda s tabelo dvojic vrednosti ne moremo vselej izčrpati vseh možnih vrednosti za neodvisno spremenljivko.

Pravilo o odvisnosti y od x pa moremo dati tudi grafično. V pravokotnem koordinatnem sistemu s točkami oziroma s krivuljami prikažemo pravilo, po katerem najdemo vrednostim neodvisne spremenljivke x ustrezne vrednosti y . Za zgornji primer je funkcija dana v sliki 8.1.



Slika 8.1 Slika funkcije $y = 2 + 3x + x^2$

Korelacijske odvisnosti

8.3

Pri opazovanju množičnih pojavov pa odvisnost v nobenem primeru ni funkcijska. Volumen danega drevesa je vsekakor odvisen od starosti, vendar ta odvisnost ni funkcijska. Za drevesa iste starosti je volumen po pravilu različen, kar ni v skladu s funkcijsko odvisnostjo. Razlog za to je v tem, da na volumen ne vpliva samo starost, temveč niz drugih, individualnih vplivov. Zato je volumen enako starih dreves različen. Čeprav v splošnem pričakujemo, da ima starejše drevo večji volumen, more zaradi individualnih vplivov imeti v posameznem pri-

meru starejše drevo tudi manjši volumen kot mlajše. V splošnem pa je volumen večji, čim starejše je drevo. Iz tega zaključimo naslednje. Zaradi individualnih vplivov, ki vplivajo na vsak pojav, v naravi ne zasledimo funkcijskih odvisnosti, kljub temu pa so pojavi med seboj odvisni. Te vrste odvisnosti imenujemo, za razliko od funkcijskih, korelacijske odvisnosti. Zakonitost korelacijske odvisnosti ne velja v vsakem posameznem primeru, temveč na splošno v množici pojavov. Zato odkrijemo korelacijske odvisnosti med pojavi le z množičnim proučevanjem pojavov.

8.4

Vzročne in čiste korelacijske odvisnosti. Pri proučevanju odvisnosti volumna posameznih dreves od starosti opazimo, da starost vpliva na volumen, ne pa obratno. V tem smislu vsebinsko tudi delimo znake na faktorialne in rezultativne. Vselej, ko je od dveh odvisnih pojavov eden faktor, drugi pa rezultat, govorimo o vzročnih korelacijskih odvisnostih. Takih primerov proučevanja imamo največ. Prirastek je vzročno povezan s količino padavin in povprečno temperaturo, višina drevesa s starostjo itd.

Imamo pa tudi korelacijske odvisnosti in zveze, ki niso vzročne. Če n.pr. opazujemo pri posameznih drevesih debelino skorje na severni in južni strani drevesa, spoznamo, da sta debelini skorje v različnih smereh v korelacijski zvezi oziroma odvisnosti. Vendar ta odvisnost ni vzročna. Ne moremo namreč trditi, da debelina skorje na južni strani vpliva na debelino skorje drevesa na severni strani, niti obratno. Kljub temu pa sta debelini skorje v naznačenih smereh v odvisnosti. To pa zato, ker na debelino skorje na južni in severni strani vplivajo pri istem drevesu isti faktorji. Rezultat teh faktorjev je posredna odvisnost skorje na severni in južni strani. V takih primerih govorimo o čistih korelacijskih odvisnostih in zvezah.

8.5

Tabela parov vrednosti. Najenostavnejši in najobičajnejši prikaz korelacijske odvisnosti med dvema pojavoma je tabela parov vrednosti za znaka x in y za posamezne enote v populaciji.

Tako imamo v tabeli 8.2 podatke o premerih (x) in povprečni debelini skorje (y) v smeri sever-jug za dvajset dreves marilandske topole z Otoka.

Tabela 8.2 Premeri in povprečna debelina skorje v smeri sever-jug za dvajset dreves marilandske topole z Otoka

Zap.štev.drevesa i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
premer v cm: x_i	40	41	45	49	47	49	49	54	52	53	52	55	56	58
deb.skorje v mm: y_i	18	20	26	22	26	22	27	25	29	32	21	28	29	30

Zap.štev.drevesa i	15	16	17	18	19	20
premer v cm: x_i	63	60	62	69	74	76
deb.skorje v mm: y_i	36	26	28	32	34	38

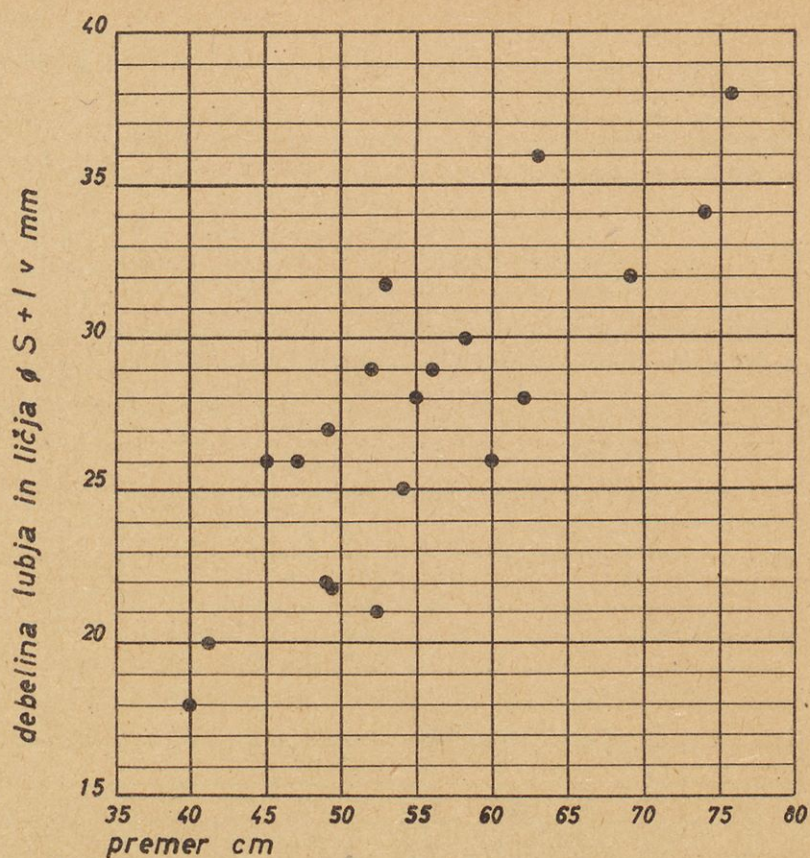
Iz zgornjih podatkov opazimo, da se v splošnem debelina skorje večja, čim večji je premer drevesa. V individualnih primerih pa ta zakonitost ne pride vselej do izraza. Tako ima n.pr. tretje drevo premer 45 cm, debelino skorje pa 26 mm, četrto drevo pa ima večji premer (49 cm), debelina skorje pa je manjša (22 mm).

8.6

Korelacijski grafikon. Zgornje podatke grafično prikažemo v korelacijskem grafikonu v pravokotnem koordinatnem sistemu tako, da par podatkov za vsako drevo ponazorimo s točko v pravokotnem koordinatnem sistemu, v katerem je premer drevesa (x_i) abscisa, povprečna debelina skorje (y_i) pa ordinata.

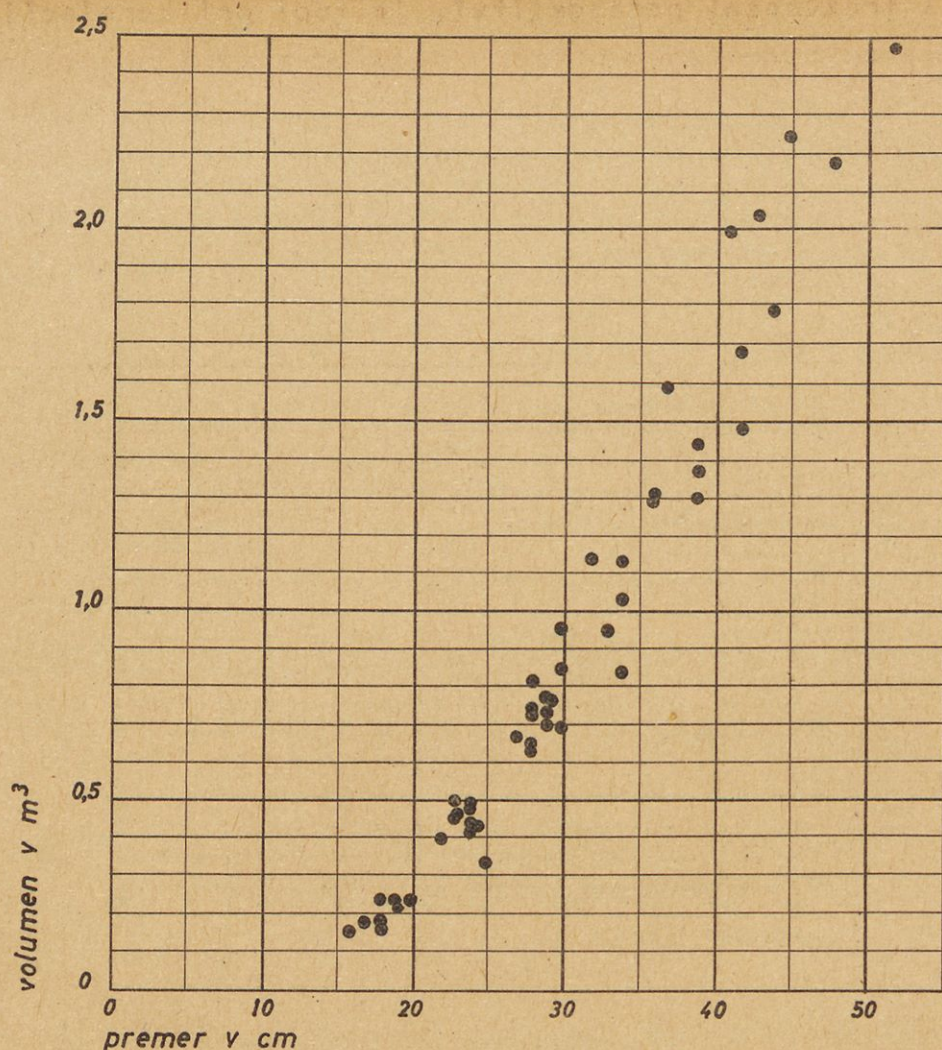
Iz gostitve točk v korelacijskem grafikonu nazorno vidimo, da velja splošna težnja, da se povprečna debelina skorje večja, če se večja premer drevesa. V splošnem so ordinate točk vedno večje, čim večje so abscise. V individualnih primerih pa

opazimo več odstopanj od tega pravila.



Slika 8.2 Korelacijski grafikon odvisnosti med premerom in povprečno debelino skorje za marilandsko topolo z Otoka

Zaradi splošne problematike je v sliki 8.3 narisan še korelacijski grafikon, ki kaže odvisnost volumna od premera za 50 modelnih smrekovih dreves sestoja na Pokljuki.



Slika 8.3 Korelacijski grafikon med premerom in volumnom za 50 modelnih smrekovih dreves poskusnega sestoja na Pokljuki

8.7

Korelacijska tabela. Pri proučevanju enega samega znaka da frekvenčna porazdelitev pregledno sliko o variriranju proučevanega znaka. Enako da kombinacijska tabela med koreliranimi znakoma pregledno sliko o korelacijski odvisnosti za velike populacije. Ta tabela, ki jo imenujemo korelacijska tabela, ima vse prednosti in hibe frekvenčnih porazdelitev, ker je v stvari frekvenčna porazdelitev po dveh znakih hkrati. Korelacijska tabela je pregledna in služi kot osnova za numerično proučevanje korelacijskih odvisnosti, je pa nenatančna in to tem bolj, čim širši so razredi bodisi po znaku x ali y. Enako

koč pri frekvenčni porazdelitvi, je tudi pri korelacijski tabeli v vsakem posameznem primeru problem z izbiro širine razredov za znak x oziroma y . Ta ne sme biti niti premajhna (tabela nepregledna), niti prevelika (zakonitosti ne pridejo do izraza).

Tabela 8.3 Korelacijska tabela med širino branike in specifično težo za zeleno duglazijo

(Vir: Katedra za tehnologijo lesa FAGV)

Teža v kg/m^3	Širina branike v mm										Skupno
	0,9	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	
600-619					1						1
580-599	1	2	6	4	3						16
560-578		10	10	20	4	1					45
540-559	7	23	28	23	13						94
520-539	13	29	51	37	29	6					165
500-519	23	48	60	73	58	11	6				279
480-499	12	67	80	94	84	32	7	5			381
460-479	10	85	110	77	89	52	19	10	1		453
440-459	13	60	100	103	99	82	54	17	2	1	531
420-439	8	21	64	102	101	84	59	34	10	1	484
400-419	5	7	22	76	84	73	70	34	20	6	397
380-399	1	5	4	18	59	93	58	34	12	5	289
360-379	7	3	5	5	15	43	39	31	12	3	163
340-359	4			3	5	6	22	17	10	4	71
320-339	1			1			3		2		7
Skupno	105	360	540	636	644	483	337	182	69	20	3376

Za primer je v tabeli 8.3 prikazana korelacijska tabela o odvisnosti specifične teže za zeleno duglazijo od širine branike za $N = 3376$ meritev.

Frekvenca 1 v razredu $600-619 \text{ kg/m}^3$ in $4,0-4,9 \text{ mm}$ za širino branike pomeni, da je v skupnem en primer, da je specifična teža enaka od 600 kg/m^3 do pod 620 kg/m^3 , širina branike za ta primer pa $4,0 \text{ mm}$ do pod $5,0 \text{ mm}$. Analogen pomen imajo tudi druge frekvence. Korelacijska tabela je zelo podobna korelacijskemu grafikonu, le da imamo namesto točk s točnimi ko-

ordinatami število enot v posameznih poljih.

8.8

Značilnosti korelacijskih odvisnosti. S primerjavo obeh korelacijskih grafiknov in korelacijske tabele napravimo pomembne zaključke o korelacijskih odvisnostih.

V obeh primerih se ordinate v meglici točk, ki opazarjajo podatke za posamezno drevo, večajo, če se večajo abscise. V obeh primerih govorimo zato o pozitivni korelacijski povezanosti, ker se en podatek večja, če se večja drugi. V splošnem se večja debelina skorje, če se večja premer. Podobno je z volumnom. Niso pa vse korelacijske odvisnosti pozitivne. Pri proučevanju specifične teže v odvisnosti od širine branike naletimo na primer na negativno korelacijsko odvisnost, ker se specifična teža v splošnem manjša, če se širina branike večja.

V primeru odvisnosti debeline skorje od premera opazimo, da se točke goste v neki določeni smeri, ki bi bila za ta primer verjetno premica. V takih primerih govorimo o linearnih odvisnostih, za razliko od primerov, kakršen je primer odvisnosti volumna od premera, v katerem smer povezanosti ni premica, temveč krivulja. V takih primerih govorimo o krivulj-
čnih odvisnostih.

Pri odvisnosti debeline skorje od premera je smer odvisnosti manj izrazita kot pri odvisnosti volumna od premerov. Če proučimo vzroke večje ali manjše izrazitosti smeri odvisnosti, spoznamo, da je vzrok temu večja ali manjša jakost individualnih vplivov. V skrajnem primeru, če individualnih vplivov ne bi bilo, bi bila odvisnost funkcijska. Vse točke bi v tem primeru ležale na neki krivulji, ki bi podajala funkcijsko odvisnost znaka y od x . Ker v vsakem primeru individualni vplivi v množičnih pojavih obstajajo, se točke od idealne smeri oziroma krivulje, ki nakazuje, kakšna bi bila zveza med x in y , če individualnih vplivov ne bi bilo, odklanjajo navzgor in navzdol. Ti odkloni so tem manjši in smer tem izrazitejša, čim manjši so individualni vplivi. Odkloni so pa tem večji oziroma smer tem bolj zabrisana, čim večji so individualni vplivi. Idealno krivuljo, ki pokaže, kakšna bi bila zveza med proučevanim pojavoma, če bi nanju ne delovali individualni vplivi, ime-

nujemo regresijsko krivuljo.

Pojava pa sta tem bolj odvisna, čim manjši so individualni vplivi oziroma čim izrazitejša je smer odvisnosti. V splošnem je stopnja odvisnosti večja ali manjša. Največja je v ekstremnem primeru pri funkcijski odvisnosti, najmanjša pa je v primeru, če pojava nista odvisna in v meglici točk ni mogoče začrtati določene smeri.

Problem proučevanja korelacijskih odvisnosti sestoji v glavnem v tem, da za proučevane pojave poiščemo regresijsko krivuljo in izračunamo koeficient stopnje odvisnosti.

Na splošno pa nimamo ene same regresijske krivulje, temveč dve. Pri čistih korelacijskih odvisnostih moremo namreč proučevati, kako je y odvisen od x ali obratno kako je x odvisen od y . Tako pri proučevanju korelacijske povezanosti med premerom in debelino lubja iščemo regresijske krivulje, ki покаžejo, kako je debelina lubja odvisna od premera ali obratno, kako je premer dreves odvisen od debeline lubja.

Določanje regresijskih krivulj

8.9

Prostoročna metoda. Najenostavnejši način za določanje regresijskih krivulj je prostoročna metoda. Po tej metodi glede na položaj meglice točk v korelacijskem grafikonu prostoročno vrišemo krivuljo, ki poteka med točkami tako, da se točkam čim bolj prilaga. Ta način je sicer subjektiven, ima pa to dobro lastnost, da je hiter in enostaven. Zato ga običajno uporabljamo pri analizah odvisnosti, da z njim ugotovimo tip krivulje ali druge lastnosti regresijskih krivulj, ki so važne za analitično proučevanje regresije.

8.10

Metoda sredin. Če simboliko za funkcijske odvisnosti prenesemo na korelacijske odvisnosti, z enačbo

$$y = f(x, I) \quad (8.3)$$

naznačimo, da je y odvisen od x in individualnih vplivov, katere izrazimo z enotnim simbolom I .

Izkaže se, da je za praktično proučevanje koristna predpostavka, da sta rezultat vpliva x in rezultat vpliva individualnih faktorjev na y v aditivni zvezi

$$y = f(x) + e(I) \quad (8.4)$$

Učinek individualnih vplivov je bodisi pozitiven ali negativen in predpostavljamo, da se v povprečju, če že ne uniči, vsaj omili. Če za dano vrednost x_k izračunamo iz vseh vrednosti y , ki tej vrednosti ustrezajo, povprečje, dobimo

$$\bar{y}_k = \overline{f(x_k) + e(I)} = f(x_k) \quad (8.5)$$

Ker je povprečje za $f(x_k)$, če ga izračunamo pri stalnem x_k , konstanta, je povprečje $\overline{f(x_k)} = f(x_k)$. To pa je ordinata $f(x_k)$ na regresijski krivulji za vrednost x_k . Ker je povprečje iz učinka individualnih vplivov po predpostavki enako nič $\overline{e(I)}=0$, zaključimo, da dobimo ordinato regresijske krivulje za vrednost x_k , če izračunamo povprečje iz vrednosti y za vse enote, ki imajo isto vrednost x_k . Če izračunamo povprečje $\bar{y}_k$ za različne vrednosti x_k , dobimo niz točk $(\bar{x}_k, \bar{y}_k)$, ki leže na oziroma v bližini regresijske krivulje. Če zvežemo te točke sredin, dobimo lomljeno črto, ki nazorno pokaže odvisnost y od x in jo smatramo za regresijsko črto.

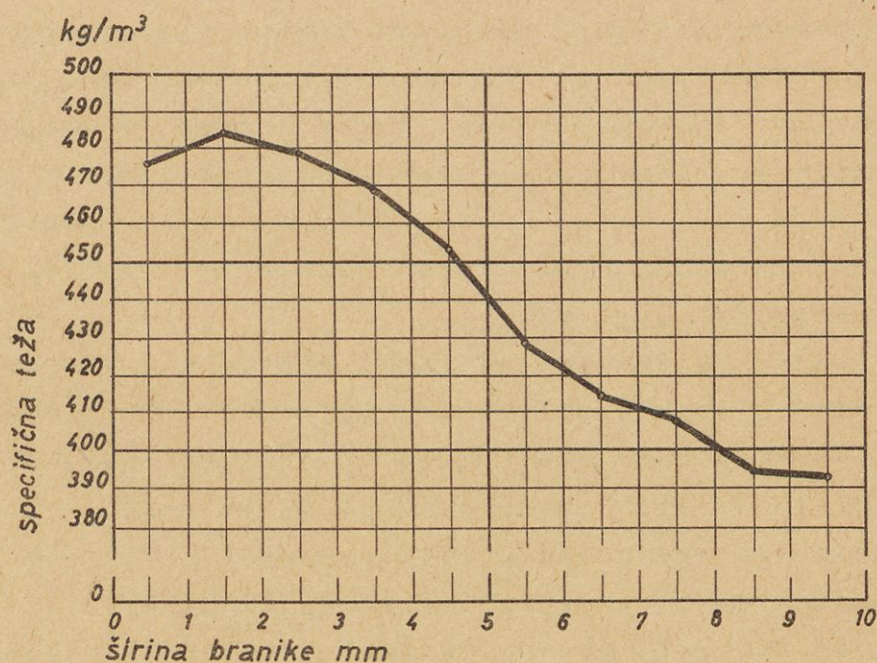
8.11

Iz korelacijske tabele 8.3 o odvisnosti specifične teže od širine branike moremo izračunati vrsto povprečnih specifičnih tež za posamezne grupe branike po 1 mm. Tako dobimo naslednjo vrsto sredin: (Glej tabelo na naslednji strani.)

Ti podatki so narisani v grafikonu v sl. 8.4. Iz slike sredin je nazorno razvidno, da je regresija krivuljčna, ker se z večanjem širine branike specifična teža najprej dvigne, nato pa v krivulji polagoma pada in sicer se spremembe v specifični teži manjšajo, čim širša je branika.

Tabela 8.4 Povprečna specifična teža pri različnih širinah
branike za les zelene duglazije.

Širina branike v mm		Povprečna te- ža v kg/m ³
Razred	Sredina razreda x_k	$\bar{y}_k$
0,0-0,9	0,5	475
1,0-1,9	1,5	484
2,0-2,9	2,5	478
3,0-3,9	3,5	468
4,0-4,9	4,5	454
5,0-5,9	5,5	428
6,0-6,9	6,5	414
7,0-7,9	7,5	407
8,0-8,9	8,5	394
9,0-9,9	9,5	392



Slika 8.4 Regresijska črta grupnih sredin za spe-
cifično težo lesa za zeleno duglazijo v
odvisnosti od širine branike

Če imamo individualne podatke, ustreza posamezni vrednosti x_i ena ali par vrednosti za y . Zato za posamezne vrednosti x ne moremo izračunati aritmetične sredine za y . V tem primeru si pomagamo tako, da zgrupiramo enote po znaku x v razrede in izračunamo za posamezne grupe povprečja $\bar{x}_k$ in $\bar{y}_k$. Kot točke na regresijski krivulji vzamemo točke s koordinatami $(\bar{x}_k, \bar{y}_k)$. Grupe ne smejo biti niti preobširne (regresijska krivulja izravnana) niti premalo zasedene (učinek individualnih vplivov se ne odstrani).

Če je število enot v populaciji majhno, dobimo po zgornjem postopku malo grup oziroma malo točk na regresijski črti.

Tako v našem primeru o odvisnosti debeline skorje od premera iz tabele 8.2 dobimo samo dve grupni sredini, če vzamemo v vsako grupo po deset dreves. V takih primerih si pomagamo z drsečimi sredinami.

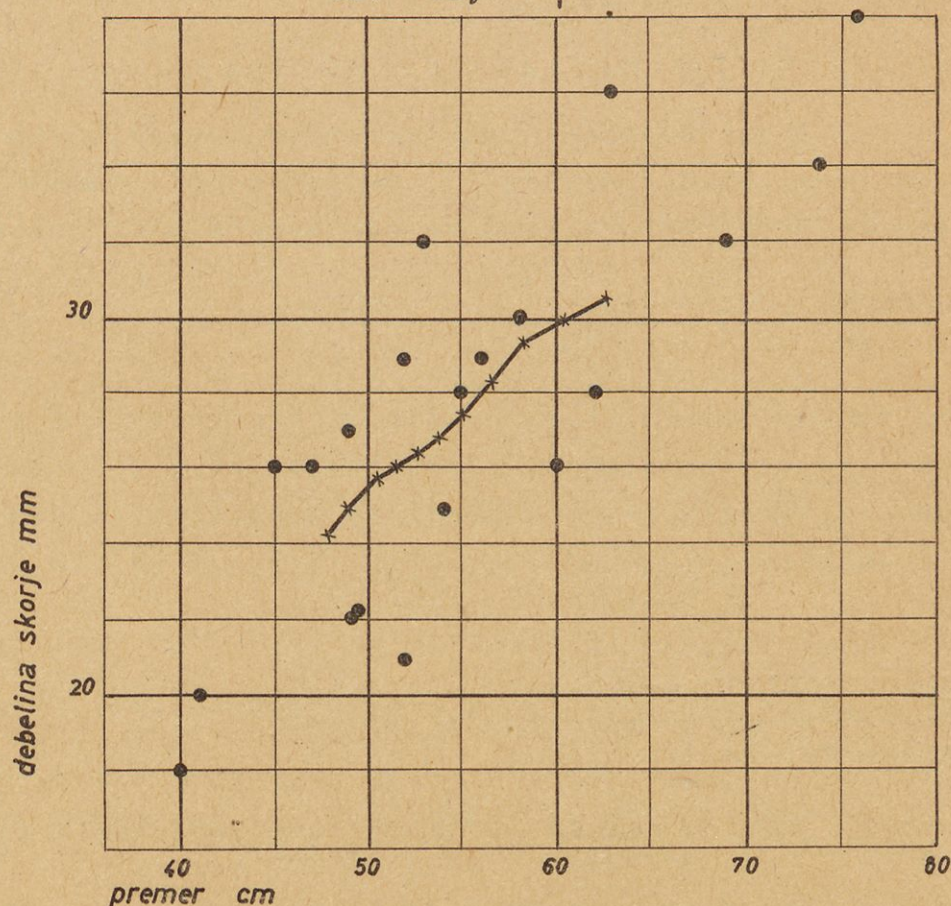
Če uredimo drevesa po velikosti premerov v ranžirno vrsto, dobimo enajst grup po deset dreves, če vzamemo vanje enote od 1 do 10, od 2 do 11, od 3 do 12 itd. do zadnje od 11 do 20. Tako dobimo namesto dveh ločenih grup 11 grup, ki se si-cer med seboj prekrivajo, dajo pa osnovo za izračun enajstih točk, ki so na ali v bližini regresijske črte. V tabeli 8.5 je za ta primer nakazano, kako izračunamo vrsti drsečih sredin za premer in debelino skorje.

Vsoto za prvih deset enot dobimo s seštevanjem prvih desetih podatkov n.pr. za premer: $S_{11} = 40 + 41 + 45 + \dots + 52 + 53 = 477$.

Vsoto podatkov od druge do enajste enote ni treba računati direktno. Ker poznamo vsoto prvih desetih podatkov, dobimo naslednjo vsoto, če enajsti podatek prištejemo in prvi podatek odštejemo od prve vsote: Torej $S_{12} = 477 + 54 - 40 = 491$. Še enostavneje pa je, da prvi vsoti prištejemo razliko med enajstim in prvim podatkom: $477 + (+14) = 491$. Iz te vsote dobimo na podoben način S_{13} , S_{14} itd. do konca vrste. Če dobljene drseče vsote po deset podatkov za premer in debeline skorje delimo z deset, dobimo enajst-člensko vrsto koordinat za sredine $\bar{x}_k$ in $\bar{y}_k$.

Tabela 8.5 Izračun koordinat za regresijsko črto drsečih sredin za odvisnost debeline skorje od premera drevesa

R_i	x_i					y_i				
1	40					18				
2	41					20				
3	45					26				
4	47					26				
5	49					22				
6	49					22				
7	49					27				
8	52					21				
9	52					29				
10	53	$x_i - 10$	d_i	S_i	$\bar{x}_i$	32	$x_i - 10$	d_i	S_i	$\bar{y}_i$
11	54	40	+14	477	47,7	25	18	+7	243	24,3
12	55	41	+14	491	49,1	28	20	+8	250	25,0
13	56	45	+11	505	50,5	29	26	+3	258	25,8
14	58	47	+11	516	51,6	30	26	+4	261	26,1
15	60	49	+11	527	52,7	26	22	+4	265	26,5
16	62	49	+13	538	53,8	28	22	+6	269	26,9
17	63	49	+14	551	55,1	36	27	+9	275	27,5
18	69	52	+17	565	56,5	32	21	+11	284	28,4
19	74	52	+22	582	58,2	34	29	+5	295	29,5
20	76	53	+23	604	60,4	38	32	+6	300	30,0
				627	62,7				306	30,6



Slika 8.5 Regresijska črta drsečih sredin za odvisnost debeline skorje od premera dreves

Dobljene drseče sredine so vnešene v korelacijski grafi-
kon. Slika 8.5 pokaže, kako sta premer in debelina skorje poziti-
vno odvisna in da je regresija precej linearna.

8.13

Analitična metoda za določanje regresijskih krivulj.

Teoretično najustreznejša je za regresijsko analizo analitič-
na metoda, pri kateri poiščemo regresijsko krivuljo po metodi
najmanjših kvadratov. Po tej metodi po predhodni analizi podat-
kov določimo najprej tip in analitično obliko funkcije

$$y' = f(x, a, b, c, \dots) \quad y' = f(x, a, b, c, \dots) \quad (8.6)$$

ki glede na stvarne podatke ustreza kot regresijska funkcija.

$a, b, c, \dots$ v obrazcu so parametri, ki določajo konkretno
funkcijo danega tipa. Po metodi najmanjših kvadratov je izmed
vseh možnih funkcij izbranega tipa regresijska funkcija tista,
za katero so vrednosti parametrov take, da zadoščajo pogoju,
da je vsota kvadratov odklonov stvarnih vrednosti za y od
ustreznih vrednosti y' najmanjša

$$\sum (y - y')^2 = \sum [y - f(x, a, b, c, \dots)]^2 = F(a, b, c, \dots) = \min \quad (8.7)$$

Po znanih pravilih za določanje ekstremov za funkcije več spre-
menljivk je izraz iz 8.7 najmanjši, če parametri zadoščajo po-
goju, da so parcialni odvodi po posameznih parametrih enaki 0.

$$\frac{\partial F(a, b, c, \dots)}{\partial a} = 0 ; \frac{\partial F(a, b, c, \dots)}{\partial b} = 0 ; \frac{\partial F(a, b, c, \dots)}{\partial c} = 0 ; \dots \quad (8.8)$$

Tako dobimo toliko enačb, kolikor parametrov ima regre-
sijska funkcija. Iz zgornjega sistema enačb, ki jih imenujemo
normalne enačbe, izračunamo vrednosti za parametre za dano re-
gresijsko funkcijo.

8.14

Mera stopnje odvisnosti. Če upoštevamo obrazce 8.4, 8.6
in 8.7 zaključimo, da je

$$\frac{1}{N} \sum (y - y')^2 = \frac{1}{N} \sum [f(x) + e(I) - f(x)]^2 = \frac{1}{N} \sum e(I)^2 = \sigma_e^2 \quad (8.9)$$

povprečen kvadratičen odklon stvarnih vrednosti za y od regresijske krivulje y' merilo jakosti individualnih vplivov I , ki razen znaka x vplivajo na znak y .

Varianco σ_e^2 imenujemo nepojasnjeno varianco, ker izvira iz individualnih vplivov, katerih ne poznamo. Nepojasnjena varianca je tem manjša, čim manjši so individualni vplivi oziroma čim večja je stopnja povezanosti in obratno. Ker je razlika med skupno varianco za y in nepojasnjeno varianco $\sigma_y^2 - \sigma_e^2$, varianca, ki je pojasnjena z odvisnostjo y od x , kvocient

$$I_{y,x}^2 = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_y^2} = 1 - \frac{\sigma_e^2}{\sigma_y^2} \quad (8.10)$$

pove, kolik del od skupne variance je pojasnjen z odvisnostjo y od x . Koeficient $I_{y,x}^2$, ki ga imenujemo determinacijski koeficient, uporabljamo za merilo jakosti odvisnosti znaka y od x . Čim večja je odvisnost, tem večji je namreč determinacijski koeficient. Pri funkcijski odvisnosti je determinacijski koeficient največji in sicer 1, ker je vsa variabilnost pojasnjena z odvisnostjo y od x . Determinacijski koeficient ima vrednosti med 0 do 1. Če pojava nista odvisna, je determinacijski koeficient 0, in je tem večji, čim večja je stopnja odvisnosti znaka y od x .

Včasih merimo jakost odvisnosti s kvadratnim korenom iz determinacijskega koeficienta

$$I_{y,x} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}} \quad (8.11)$$

Ta koeficient imenujemo na splošno indeks korelacije.

Iz obrazca 8.10 moremo izračunati standardni odklon, ki je merilo jakosti individualnih vplivov.

$$\sigma_e = \sigma_y \cdot \sqrt{1 - I_{y,x}^2} \quad (8.12)$$

Če je odvisnost y od x tesna, se stvarne vrednosti za y ne odklanjajo dosti od vrednosti y' , ki jih izračunamo iz regresijske enačbe, če poznamo za posamezne enote vrednosti x .

Zato moremo v takih primerih vzeti izračunane vrednosti y' za ocene stvarnih vrednosti y . V koliko se stvarna

vrednost y odklanja od ocene y' , ne vemo natančno. Pač pa vemo, da približno $2/3$ primerov odklonov ni večjih kot en standardni odklon, da približno $19/20$ odklonov ni večjih kot dva standardna odklona in da skoro noben odklon ni večji kot trije standardni odkloni. To povzamemo iz lastnosti standardnih odklonov iz odstavka 7.21. Zato imenujemo standardni odklon σ_e tudi standardno napako ocene, ker daje informacijo o tem, kolik je možen odklon stvarne vrednosti od ocene y' , ki jo dobimo z regresijsko krivuljo, če za dano enoto vemo, kolika je vrednost x .

Linearna regresija in korelacija

8.15

V naravi najdemo dosti pojavov, ki so povezani linearno. Za dosti pojavov pa je regresijska krivulja premici tako podobna, da moremo zanje v približku vzeti, da so povezani linearno. Tudi za krivuljčne odvisnosti moremo regresijo na krajšem razmaku v približku smatrati za linearno. Zato s proučevanjem linearne regresije rešujemo velik del problemov s področja korelacijskih odvisnosti med pojavi. Zato se bomo omejili na proučevanje linearne odvisnosti.

Če je regresijska krivulja premica, je v splošnem enačba premice

$$y' = a + b(x - \bar{x}) \quad (8.13)$$

Po metodi najmanjših kvadratov je za parametra a in b za regresijsko premico, pogoj, da je

$$\sum (y - y')^2 = \sum [y - a - b(x - \bar{x})]^2 = F(a, b) = \min \quad (8.14)$$

Če izraz 8.14 parcialno odvajamo po a in b , dobimo naslednji normalni enačbi:

$$\begin{aligned} \frac{\partial F(a, b)}{\partial a} &= -2 \sum [y - a - b(x - \bar{x})] = 0 \\ \frac{\partial F(a, b)}{\partial b} &= -2 \sum [y - a - b(x - \bar{x})](x - \bar{x}) = 0 \end{aligned} \quad (8.15)$$

Iz teh dveh normalnih enačb povzamemo, da izračunamo parametra a in b po obrazcih

$$a = \bar{y} ; \quad b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (8.16)$$

Če v obrazcu za izračunavanje parametra b , ki ga imenujemo regresijski koeficient, izraz v števcu in imenovalci delimo s številom enot N , dobimo v imenovalcu varianco za znak x , v števcu pa izraz

$$c_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{N} \quad (8.17)$$

ki ga imenujemo kovarianca. Kovarianca je po definiciji povprečen produkt odklonov znakov x in y od ustreznih aritmetičnih sredin. Če vnesemo vrednosti parametrov v obrazec (8.13) za premico, dobimo, da je

$$y' = \bar{y} + b_1(x - \bar{x}) ; \quad b_1 = \frac{c_{xy}}{\sigma_x^2} \quad (8.18)$$

Ker moremo pri čistih korelacijskih povezavah iskati razen odvisnosti y od x tudi odvisnost znaka x od y , dobimo z analognim sklepanjem poleg zgornje, ki jo imenujemo prvo regresijsko premico, še drugo regresijsko premico

$$x' = \bar{x} + b_2(y - \bar{y}) ; \quad b_2 = \frac{c_{xy}}{\sigma_y^2} \quad (8.19)$$

Medtem ko prvi regresijski koeficient b_1 pove, za koliko se v splošnem spremeni vrednost za y , če se spremeni x za enoto, pove drugi regresijski koeficient b_2 , za koliko se v splošnem spremeni x , če se y spremeni za enoto.

Determinacijski koeficient pri linearni korelaciji zaznamujemo z r_{xy}^2 , izračunamo pa ga po obrazcu

$$r_{xy}^2 = \frac{c_{xy}^2}{\sigma_x^2 \sigma_y^2} \quad (8.20)$$

Kvadratni koren iz determinacijskega koeficienta, ki ga v splošnem imenujemo indeks korelacije, imenujemo pri linearni korelaciji koeficient korelacije

$$r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y} \quad (8.21)$$

Koeficient korelacije r_{xy} ima vrednosti med -1 do $+1$. Če je korelacija pozitivna, je korelacijski koeficient pozitiven, pri negativni linearni korelaciji pa je korelacijski koeficient negativen. Absolutna vrednost korelacijskega koeficienta pa je tem večja, čim večja je stopnja odvisnosti in je ena, kadar je odvisnost linearna in funkcijska.

8.16

Izračunavanje pokazateljev za linearno regresijo in korelacijo. Če izračunavamo pokazatelje linearne korelacije, najprej izračunamo iz osnovnih podatkov za x in y količine $\bar{x}$, $\bar{y}$, σ_x^2 , σ_y^2 in c_{xy} . Vse izmed teh količin, razen kovariance, že znamo izračunavati.

Enako kot pri izračunavanju varianc iz negrupiranih podatkov se tudi pri izračunavanju pokazateljev linearne regresije in korelacije obnese metoda pomožnega znaka. Pri tem uvedemo za znak x in za znak y pomožna znaka

$$u = x - x_0 \quad ; \quad v = y - y_0 \quad (8.22)$$

tako, da od osnovnih vrednosti za znak x odštejemo neko ustrezno vrednost x_0 , od y pa ustrezno vrednost y_0 tako, da dobimo čim prikladnejše vrednosti za pomožna znaka u in v .

Iz teh pomožnih vrednosti u in v izračunamo aritmetični sredini $\bar{x}$ in $\bar{y}$, varianci σ_x^2 in σ_y^2 in kovarianco c_{xy} po obrazcih

$$\begin{aligned} \bar{x} &= x_0 + \frac{U}{N} \quad ; \quad \bar{y} = y_0 + \frac{V}{N} \\ \sigma_x^2 &= K_u / N \quad ; \quad K_u = \sum u^2 - U^2 / N \\ c_{xy} &= K_{uv} / N \quad ; \quad K_{uv} = \sum uv - UV / N \\ \sigma_y^2 &= K_v / N \quad ; \quad K_v = \sum v^2 - V^2 / N \end{aligned} \quad (8.23)$$

pri čemer je $U = \sum u$; $V = \sum v$

Iz teh vmesnih rezultatov pa izračunamo pokazatelje linearne korelacije po obrazcih 8.18 do 8.21.

Iz korelacijskega grafikona v sliki 8.2 smo ugotovili, da je debelina skorje za marilandsko topolo linearno povezana s premerom drevesa. V tabeli 8.6 so po postopku, ki je nakazan v odstavku 8.16, izračunani pokazatelji linearne regresije in korelacije po shemi, ki tehnično najbolj ustreza za izračunavanje.

Tabela 8.6 Izračun pokazateljjev linearne regresije in korelacije za povezanost debeline skorje in premera dreves za $N = 20$ dreves marilandske topole (x = premer v cm, y = debelina lubja v mm).

i	x	y	$x-50$ u	$y-30$ v	u^2	uv	v^2
1	40	18	-10	-12	100	+120	144
2	41	20	-9	-10	81	+90	100
3	45	26	-5	-4	25	+20	16
4	49	22	-1	-8	1	+8	64
5	47	26	-3	-4	9	+12	16
6	49	22	-1	-8	1	+8	64
7	49	27	-1	-3	1	+3	9
8	54	25	+4	-5	16	-20	25
9	52	29	+2	-1	4	-2	1
10	53	32	+3	+2	9	+6	4
11	52	21	+2	-9	4	-18	81
12	55	28	+5	-2	25	-10	4
13	56	29	+6	-1	36	-6	1
14	58	30	+8	0	64	0	0
15	63	36	+13	+6	169	+78	36
16	60	26	+10	-4	100	-40	16
17	62	28	+12	-2	144	-24	4
18	69	32	+19	+2	361	+38	4
19	74	34	+24	+4	576	+96	16
20	76	38	+26	+8	676	+208	64

$$\begin{aligned} U &= +104 \\ U/N &= +5,20 \\ +x_0 &= +50,00 \\ \bar{x} &= 55,20 \end{aligned}$$

$$\begin{aligned} V &= -51 \\ V/N &= -2,55 \\ +y_0 &= +30,00 \\ \bar{y} &= 27,45 \end{aligned}$$

$$\begin{aligned} \sum u^2 &= 2402 & \sum uv &= +567 & \sum v^2 &= 669 \\ -U^2/N &= -540,80 & -UV/N &= +265,20 & -V^2/N &= -130,05 \\ K_u &= 1861,20 & K_{uv} &= 832,20 & K_v &= 538,95 \\ \sigma_x^2 &= 93,06 & c_{xy} &= 41,61 & \sigma_y^2 &= 26,95 \\ \sigma_x &= 9,65 & r_{xy} &= \frac{c_{xy}}{\sigma_x \sigma_y} = 0,831 & \sigma_y &= 5,19 \end{aligned}$$

$$b_1 = \frac{c_{xy}}{\sigma_x^2} = 0,447$$

$$b_2 = \frac{c_{xy}}{\sigma_y^2} = 1,544$$

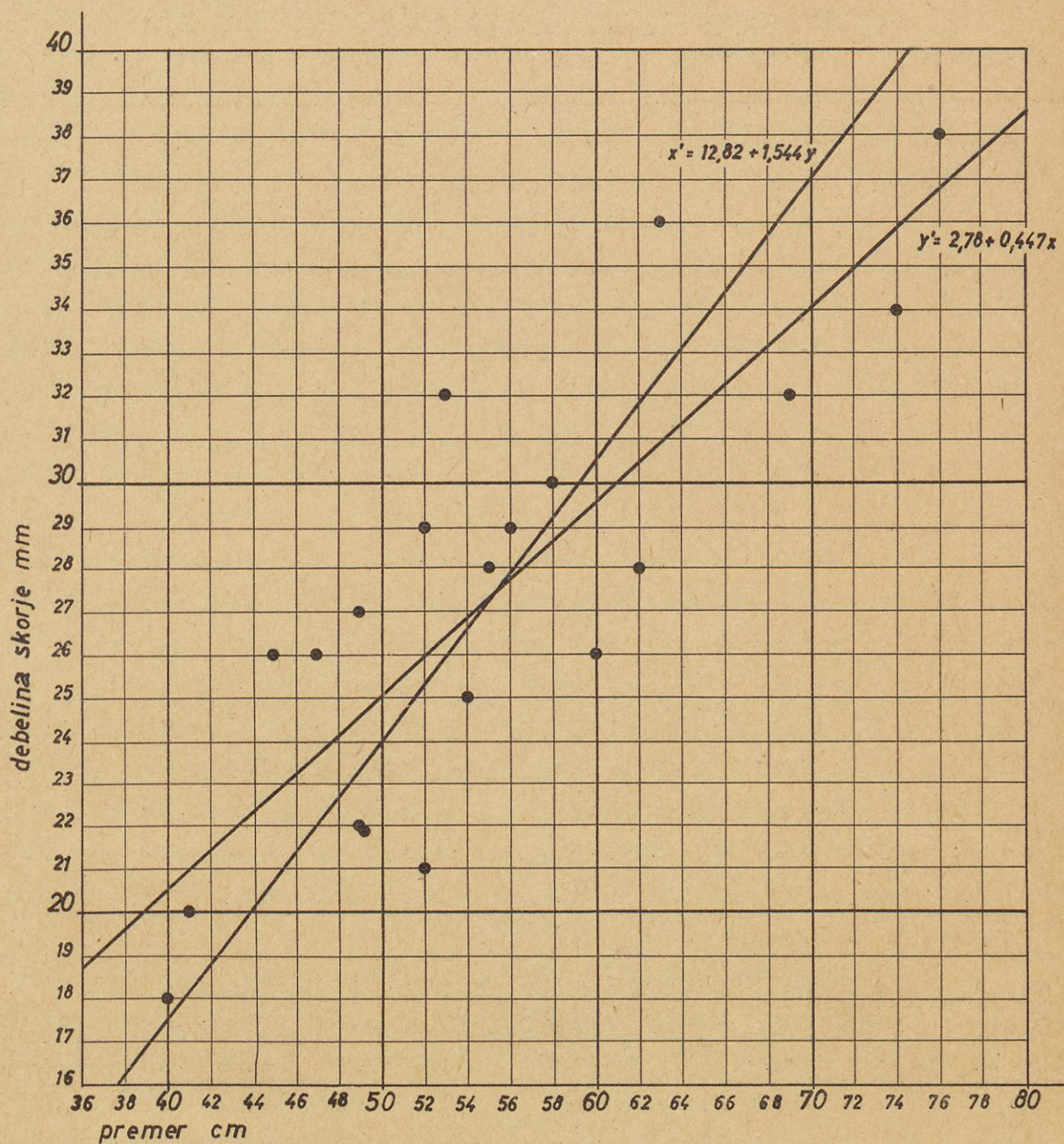
$$\sqrt{1 - r^2} = 0,556$$

$$\begin{aligned} \sigma_{ex} &= \sigma_x \sqrt{1 - r_{xy}^2} \\ &= 5,365 \end{aligned}$$

$$\begin{aligned} \sigma_{ey} &= \sigma_y \sqrt{1 - r_{xy}^2} \\ &= 2,886 \end{aligned}$$

$$\hat{y} = \bar{y} + b_1(x - \bar{x}) = 27,45 + 0,447(x - 55,20) = 2,78 + 0,447x$$

$$\hat{x} = \bar{x} + b_2(y - \bar{y}) = 55,20 + 1,544(y - 27,45) = 12,82 + 1,544y$$



Slika 8.6 Korelacijski grafikon za povezanost debeline skorje (y) in premera dreves (x) za merilandsko topolo z regresijskima premicama.

9.1

Pomen proučevanja teoretičnih porazdelitev. Do sedaj obravnavane frekvenčne porazdelitve so bile rezultat statističnega opazovanja stvarnih populacij. Stvarne frekvenčne porazdelitve so najrazličnejših oblik, odvisno od faktorjev in pogojev, v katerih se je proučevana populacija razvijala.

Razen porazdelitev, ki jih dobimo z opazovanjem stvarnih populacij, pa imamo tudi teoretične frekvenčne porazdelitve, ki jih sestavimo na osnovi logičnih predpostavk in izpeljemo matematično. Teoretične porazdelitve so važne za razvoj statistične teorije in prakse in jih uporabljamo v najrazličnejših postopkih pri statistični analizi.

Proučevanje teoretičnih porazdelitev je važno iz več vidikov:

a) Če se stvarna porazdelitev sklada s teoretično porazdelitvijo, do katere pridemo z določenimi predpostavkami, v nekaterih primerih iz tega sklepamo na stvarne pogoje, ki so identični pogojem, pod katerimi je sestavljena teoretična porazdelitev. Tako sklepamo, da je n.pr. sestoj enodoben in čist in da so nanj vplivali samo slučajni, nebitveni faktorji, če je stvarna porazdelitev premerov podobna normalni. Normalna porazdelitev je namreč teoretično izdelana pod predpostavko, da na pojav razen slučajnih vplivov ne delujejo nobeni drugi, individualni vplivi.

b) Če ne poznamo stvarne porazdelitve, s teoretičnimi porazdelitvami ugotovimo, kakšna bi bila približno stvarna porazdelitev pojava, če poznamo faktorje in predpostavke, katerim stvarna populacija oziroma enote zadoščajo. Za dan sestoj poznamo število dreves, aritmetično sredino in standardni odklon. Če predpostavljamo, da je sestoj enodoben in čist, sklepamo, da se premeri porazdeljujejo v normalni porazdelitvi. Pri tej predpostavki moremo izračunati, koliko je približno stvarno število dreves v posameznih debelinskih stopnjah, ne da poznamo individualne premere dreves.

c) Posebno pomembne pa so teoretične porazdelitve v vzorčenju, ki je posebna in najpomembnejša metoda ocenjevanja v

statistiki, in pri preizkušanju hipotez, ki je osnova za posebno statistično disciplino - planiranje eksperimentov.

Normalna porazdelitev

9.2

Normalna porazdelitev je ena izmed najvažnejših, a tudi najbolj poznanih teoretičnih porazdelitev. Veliko pojavov se porazdeljuje v unimodalni, simetrični in zvonasti obliki, kar so tipične značilnosti za normalno porazdelitev. Razen tega je normalna porazdelitev osnova za izpeljavo raznih drugih, za statistično teorijo in prakso važnih porazdelitev, kot so: t-porazdelitev, χ^2 -porazdelitev, F-porazdelitev. Pod določenimi pogoji preide v normalno porazdelitev večina teoretičnih porazdelitev: binominalna, Poissonova, hipergeometrična in vse tri zgoraj naštetе: t-porazdelitev, χ^2 -porazdelitev in F-porazdelitev.

9.3

Opis normalne porazdelitve. Normalna porazdelitev je določena z dvema parametroma: aritmetično sredino M in standardnim odklonom σ . Gostota relativne frekvence $\varphi(x)$ za normalno porazdelitev je dana s funkcijo

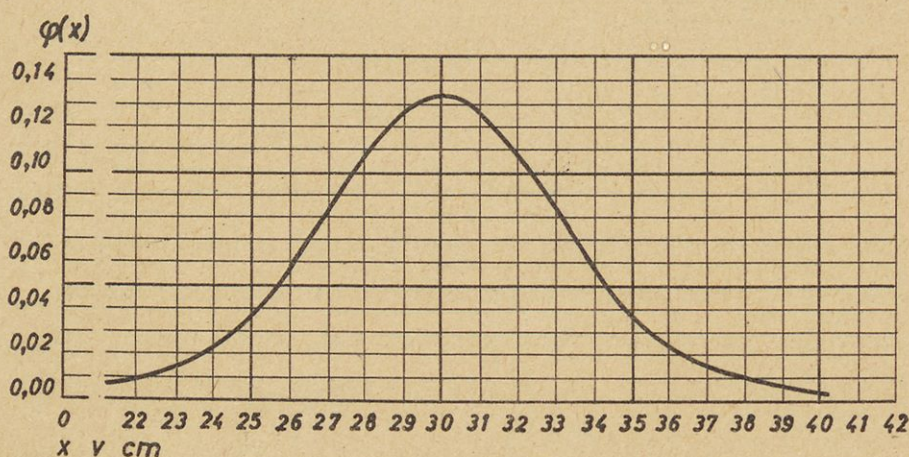
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} \quad (9.1)$$

Če narišemo zgornjo funkcijo tako, da je znak x abscisa, gostota relativne frekvence $\varphi(x)$ pa ordinata, dobimo funkcijsko sliko za normalno porazdelitev. Iz slike 9.1 vidimo, da je normalna porazdelitev unimodalna, simetrična in zvonasta. Ker je simetrična, velja, da je $M = M_e = M_o$. Med merami variacije: kvartilnim odklonom Q , povprečnim absolutnim odklonom AD in standardnim odklonom σ pa veljajo za normalno porazdelitev naslednje zveze:

$$Q = 0,675\sigma \approx \frac{2}{3}\sigma \quad \text{in} \\ AD = 0,7979\sigma \approx \frac{4}{5}\sigma$$

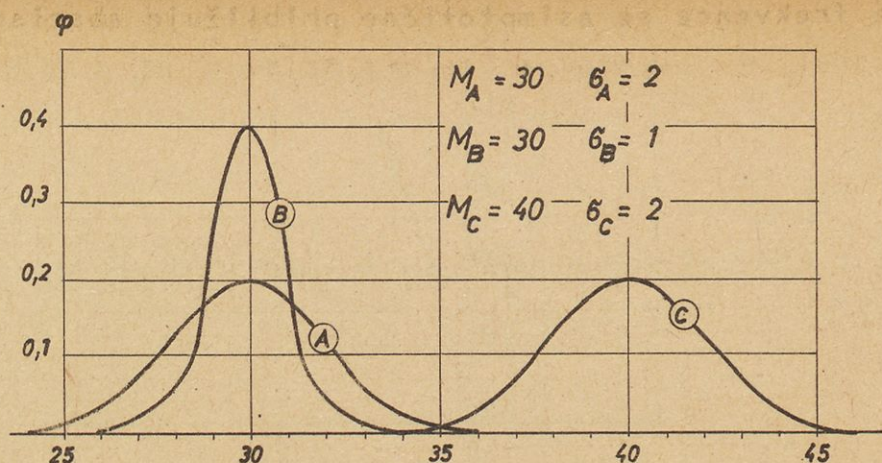
Največja gostota je za $x = M$, od tu pa gostota frekvence vztrajno pada v obe smeri, čim bolj se oddaljujemo od modusa. Gostota

relativne frekvence se asimptotično približuje abscisni osi, če se x oddaljuje od modusa čez vse meje.



Slika 9.1 Normalna porazdelitev premerov za sestoj z $\bar{Y} = 30$ cm, $\sigma = 3$ cm

Normalni porazdelitvi, ki imata enak standardni odklon in različni sredini, sta si podobni, le da sta premaknjeni ustrezno razliki med aritmetičnima sredinama. Normalni porazdelitvi, za kateri sta aritmetični sredini enaki, standardna odklona pa različna, pa se razlikujeta po tem, da se normalna porazdelitev z manjšim standardnim odklonom porazdeljuje na ožjem razmaku kot porazdelitev, za katero je standardni odklon večji. Obe pa imata vrh (modus) za isto vrednost x . Različne normalne porazdelitve so med seboj podobne. Glede na velikost aritmetične sredine pa so premaknjene v levo ali v desno, glede na velikost standardnega odklona pa se relativne frekvence porazdeljujejo na ožjem ali širšem razmaku.



Slika 9.2 Normalne porazdelitve z različnimi sredinami in različnimi standardnimi odkloni

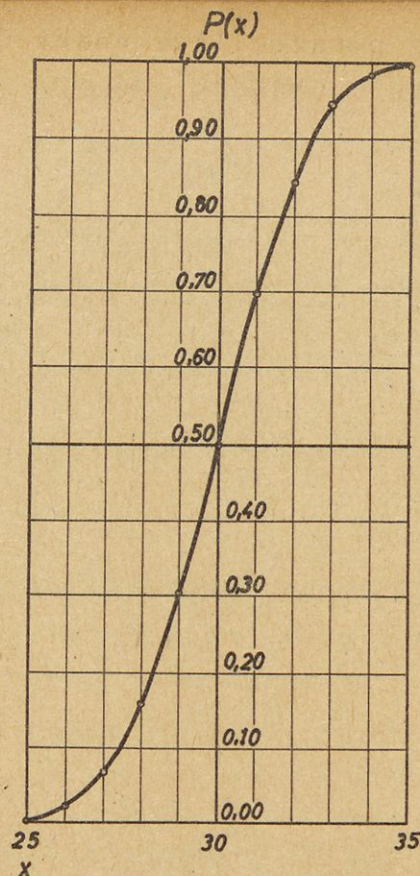
9.4

Določen integral

$$P(x) = \int_{-\infty}^x \varphi(x) dx \quad (9.2)$$

iz gostote relativne frekvence je funkcija zgornje meje x in določa kumulativno relativno frekvenco za normalno porazdelitev. $P(x)$ je monotonno naraščajoča funkcija. Grafična slika kumulativne relativne frekvence ima tipično obliko iztegnjene črke S (glej sliko 9.3). Kumulativna relativna frekvenca pove, koliki del frekvence pri normalni porazdelitvi leži v razmaku od $-\infty$ do x . Če poznamo funkcijo za kumulativno relativno frekvenco, moremo zlahka ugotoviti, kolika je relativna frekvenca v kateremkoli razmaku $x_1 \leq x \leq x_2$

$$f^o(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \varphi(x) dx = \int_{-\infty}^{x_2} \varphi(x) dx - \int_{-\infty}^{x_1} \varphi(x) dx = P(x_2) - P(x_1) \quad (9.3)$$



Slika 9.3 Kumulativna normalna porazdelitev

9.5

Za vsako normalno porazdelitev je $P(M) = 0,50$. Zaradi simetrije normalnih porazdelitev je za vsako normalno porazdelitev polovica frekvenca pod aritmetično sredino M . Nadalje je ne glede na velikost parametrov M in σ za vsako normalno porazdelitev

$$P(x = M + z\sigma) = P(z) \quad (9.4)$$

Kumulativna relativna frekvenca za $x = M + z\sigma$ je za vsako normalno porazdelitev odvisna samo od koeficienta z , nič pa od parametrov M in σ . Iz tega dalje sklepamo, da je tudi frekvenca v razmaku $M - z\sigma$ do $M + z\sigma$ pri stalnem koeficientu z enaka za vse distribucije, ne glede na velikost M in σ . To sledi iz zveze

$$\begin{aligned} f^0(M - z\sigma \leq x \leq M + z\sigma) &= P(x = M + z\sigma) - P(x = M - z\sigma) = \\ &= P(z) - P(-z) \end{aligned}$$

Tako je relativna frekvenca v razmaku enega standardnega odklona od aritmetične sredine navzdol in navzgor (od $M - \sigma$ do

$M + 6$) za vse normalne porazdelitve enaka 0,6827, v razmaku od $M - 2\sigma$ do $M + 2\sigma$ enaka 0,9545, v razmaku od $M - 3\sigma$ do $M + 3\sigma$ enaka 0,9973 itd.

9.6

Standardiziran znak z. Pomen koeficienta z najbolje spoznamo, če izhajamo iz zveze med količinami z , x , M in σ . Iz obrazca

$$x = M + z\sigma$$

dobimo, da je

$$z = \frac{x - M}{\sigma} \quad (9.5)$$

Za dano vrednost x je z , ki ga imenujemo standardiziran znak, odklon osnovne vrednosti x od aritmetične sredine M , merjen v standardnih odklonih σ . Standardiziran znak z je neimenovano število in podaja mesto vrednosti v normalni porazdelitvi. Za vse vrednosti, ki so večje kot aritmetična sredina M , je z pozitiven, in obratno, za vrednosti x , ki leže pod aritmetično sredino, je z negativen. Za približno $2/3$ (natančno 68,27 %) vseh vrednosti v normalni porazdelitvi je z torej absolutno manjši kot 1. Analogno je za približno $19/20$ (natančno 95,45 %) vseh vrednosti z absolutno manjši kot 2, za skoro vse (natančno 99,73 %) vrednosti pa z absolutno ni večji kot 3.

Vzemimo sestoj, ki se normalno porazdeljuje in ima aritmetično sredino $M = 35$ cm in $\sigma = 8$ cm. Za drevo, ki ima premer $x = 48$ cm, ne vemo, ali je glede na ta sestoj drevo z velikim ali majhnim premerom. Če pa izračunamo zanj ustrezni standardiziran znak

$$z = \frac{x - M}{\sigma} = \frac{48 - 35}{8} = + 1,625$$

spoznamo, da je drevo glede na drevesa vsega sestoja nad povprečjem in razmeroma veliko. Ustrezni z je namreč pozitiven in razmeroma velik.

Kot vidimo iz primera, s standardiziranim znakom, podobno, vendar na drug način kot s kvantilnimi rangi, določamo mesto enote v populaciji.

Standardizirana normalna porazdelitev. Če se x porazdeljuje v normalni porazdelitvi, se v normalni porazdelitvi porazdeljuje tudi standardiziran znak z , ker je z znakom x v linearni zvezi. Iz zveze

$$z = \frac{x - M}{\sigma}$$

zlahka dokažemo, da je za standardiziran znak aritmetična sredina $M_z = 0$, standardni odklon pa $\sigma_z = 1$.

Standardiziran znak z se torej porazdeljuje normalno s parametri $M_z = 0$ in $\sigma_z = 1$.

Za to normalno porazdelitev, ki jo imenujemo standardizirano normalno porazdelitev, je gostota relativne frekvenca

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (9.6)$$

kumulativna relativna frekvenca $P(z)$ pa

$$P(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \quad (9.7)$$

V tablici 9.1 imamo tabelirane podatke za gostoto relativne frekvenca za standardizirano normalno porazdelitev $\varphi(z)$ in relativne frekvenca $\Phi(z)$ v razmaku 0 do z

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}} dz \quad (9.8)$$

za pozitivne vrednosti za z . Ker je normalna porazdelitev simetrična, je $\Phi(z)$ s kumulativno relativno frekvenco v enostavni zvezi

$$P(z) = 0,50 + \Phi(z) \quad (9.9)$$

Zato iz tablic za $\Phi(z)$ po obrazcu 9.9 dobimo posredno tudi kumulativne relativne frekvenca $P(z)$. Pri tem moramo upoštevati, da je

$$\Phi(-z) = -\Phi(z) \quad (9.10)$$

Zaradi enostavne zveze med osnovnim znakom in standardiziranim znakom tablice za standardizirano normalno poraz-

Tabela 9.1 Gostota relativne frekvence za standardizirano

$$\text{normalno porazdelitev } \left(\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \right)$$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3728	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3058	3034	3011	2989	2968	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Tabela 9.2 Relativna frekvenca za standardizirano normalno

porazdelitev $\left(\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} dz \right)$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00899	00789	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41468	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45055	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46633	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49912	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

delitev uporabljamo za katerokoli normalno porazdelitev tako, da najprej vrednosti za znak x prevedemo v ustrezne vrednosti za standardizirani znak, za dobljene z pa poiščemo ustrezne vrednosti ordinat ali relativnih frekvenc.

9.8

Prilagoditev normalne porazdelitve stvarni porazdelitvi. S tablicami za normalno porazdelitev moremo rešiti problem, ki se v praksi pogosto pojavlja. K dani frekvenčni porazdelitvi večkrat prilagajamo ustrezno normalno porazdelitev, da ju v nadaljnji analizi primerjamo med seboj. Kot najbolje prilagojeno normalno porazdelitev smatramo ono, ki ima enak obseg N , enako aritmetično sredino M in enak standardni odklon σ kot stvarna porazdelitev, kateri prilagajamo normalno porazdelitev.

To nalogo rešimo s tablicami za standardizirano normalno porazdelitev, ki je dana v tabeli 9.2, takole:

a) Za stvarno frekvenčno porazdelitev, kateri hočemo prilagoditi normalno porazdelitev, poiščemo osnovne parametre: obseg populacije N , aritmetično sredino M in standardni odklon σ .

b) Meje razredov iz stvarne porazdelitve najprej prevedemo po obrazcu $z_{k,min} = (x_{k,min} - M) / \sigma$ v standardizirane vrednosti.

c) Iz tablic za relativne frekvence za standardizirano normalno porazdelitev poiščemo $\Phi(z)$, ki ustrezajo mejam razredov $z_{k,min}$.

d) Po obrazcu 9.9: $P(z) = 0,50 + \Phi(z)$ izračunamo kumulativno vrsto relativnih frekvenc $P(z)$.

e) Če $P(z)$ pomnožimo z obsegom populacije N , dobimo teoretično kumulativno frekvenčno porazdelitev F'_k .

f) Razlike dveh zaporednih členov iz vrste F'_k so teoretične frekvence f'_k za prilagojeno normalno porazdelitev.

9.9

Za primer prilagodimo normalno porazdelitev frekvenčni porazdelitvi premerov za čisti smrekov sestoje A na Pokljuki iz tabele 3.4. Sestoje ima $N = 507$ dreves. Zanj smo izračunali,

da je aritmetična sredina $M = 38,62$ cm, standardni odklon pa $\sigma = 7,89$ cm.

V tabeli 9.3 je nakazan izračun prilagojene normalne porazdelitve po postopku iz prejšnjega odstavka.

Tabela 9.3 Prilagoditev normalne porazdelitve za frekvenčno porazdelitev premerov za čisti smrekov sestoj A na Pokljuki

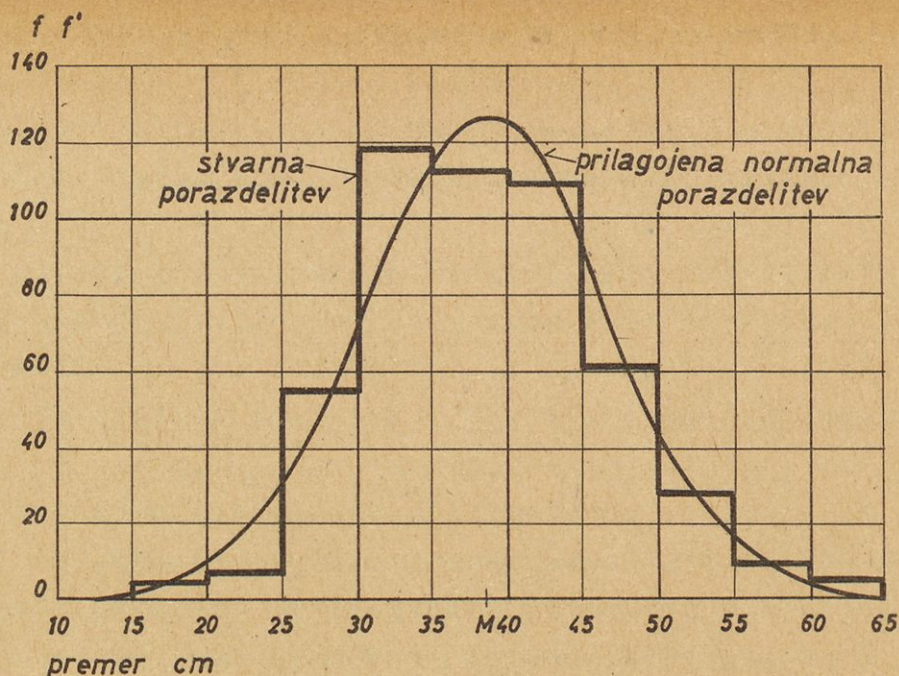
DS	Premer v cm	$x_{k,min}$	$z_{k,min}$	$\Phi(z_k)$	$P(z_k)$	F_k	f_k^*	f_k
3	14						0,7	
4	15-19	15	-2,99	-0,4986	0,0014	0,7	3,9	4
5	20-24	20	-2,36	-0,4909	0,0091	4,6	17,0	6
6	25-29	25	-1,72	-0,4573	0,0427	21,6	48,3	55
7	30-34	30	-1,09	-0,3621	0,1379	69,9	93,8	118
8	35-39	35	-0,46	-0,1772	0,3228	163,7	124,0	112
9	40-44	40	+0,17	+0,0675	0,5675	287,7	113,3	109
10	45-49	45	+0,81	+0,2910	0,7910	401,0	68,0	61
11	50-54	50	+1,44	+0,4251	0,9251	469,0	28,3	28
12	55-59	55	+2,07	+0,4808	0,9808	497,3	8,0	9
13	60-64	60	+2,71	+0,4966	0,9966	505,3	1,5	5
14	65-	65	+3,34	+0,4996	<u>0,9996</u>	<u>506,8</u>	<u>0,2</u>	<u>-</u>
					1,000	507,0	507,0	507

Za spodnjo mejo v četrti debelinski stopnji je nopr. ustrezná vrednost za $z_{4,min}$

$$z_{4,min} = \frac{x_{4,min} - M}{\sigma} = \frac{15 - 38,62}{7,89} = -2,99$$

V tabeli 9.3 so v zadnjem stolpcu vpisane še stvarne frekvence. Če stvarne frekvence primerjamo z dobljenimi teoretičnimi frekvencami za prilagojeno porazdelitev, vidimo, da razlike niso znatne. Večje razlike so edino v sedmi in osmi debelinski stopnji.

Nenavadno izgleda, da teoretične frekvence niso cela, temveč decimalna števila. Vendar je, glede na to, da so teoretične frekvence izračunane vrednosti, to umestno.



Slika 9.4 Frekvenčna porazdelitev premerov dreves v čistem smrekovem sestoj A na Pokljuki, s prilagojeno normalno porazdelitvijo

Verjetnostne porazdelitve

9.10

Predpostavljajmo, da poznamo osnove verjetnostnega računa. (Glej Rajko Jamnik, Ljubljana, 1959: Matematika za gozdarje, str.263-296.) Porazdelitev relativnih frekvenc smatramo kot verjetnostno porazdelitev za določen statistični znak, če vzamemo kot poskus v smislu verjetnostnega računa slučajnostni izbor posamezne enote populacije. Pogoji za slučajnostni izbor je, da ima vsaka enota enako možnost, da jo izberemo, kot dogodek v smislu verjetnostnega računa pa vzamemo dejstvo, da ima izbrana enota določeno vrednost znaka ali slučajnostne spremenljivke. Ta opredelitev verjetnosti se sklada z definicijo apriorne verjetnosti. Tako moremo za sestoj iz tabele 4.1 povzeti, da je verjetnost, da slučajnostno izberemo izmed $N = 507$ dreves drevo, za katerega je premer med 30 cm in 35 cm, enaka $Pr = 0,233$, ker je v tem razredu 23,3 % od vseh enot v populaciji.

Do verjetnosti pa pridemo aposteriorno tako, da ponavljamo slučajnostni izbor in iščemo relativno frekvenco po-

gostnosti, s katero izbiramo slučajnostno iz populacije drevesa s premeri 30-35 cm. Čim večkrat poskus izbora ponovimo, tem bolj se relativna frekvenca pogostnosti približuje verjetnosti, da s slučajnostnim izborom izberemo drevo s premerom 30 cm do 35 cm.

Pod istimi pogoji smatramo za verjetnostne porazdelitve tudi relativne frekvence pri teoretičnih porazdelitvah. Tablice relativnih frekvenc za teoretične porazdelitve so ob enem verjetnostne porazdelitve za določene slučajnostne spremenljivke.

Za normalno porazdelitev iz tablice o relativnih frekvencah $\Phi(z)$ n.pr. razberemo, da je verjetnost, da iz normalno porazdeljene vrednosti, slučajnostno izberemo enoto, za katero je standardiziran odklon vrednosti med $z=0$ in $z=1,00$, enaka $Pr = 0,3413$, ker je

$$Pr(0 \leq z \leq 1) = \Phi(z) = \int_0^1 \phi(z) dz = 0,3414$$

Verjetnost, da iz normalno porazdeljene populacije slučajnostno izberemo enoto, za katero je standardiziran znak v razmaku od $-1,96$ do $+1,96$, je enaka

$$Pr(-1,96 \leq z \leq +1,96) = 0,95$$

Iz tablic za relativne frekvence za normalno porazdelitev dalje sklepamo, da je verjetnost, da iz normalno porazdeljene populacije z naključnim izborom izberemo enoto, za katero je standardiziran znak z večji kot $+1,645$, enaka $Pr=0,05$. Podobno moremo določiti verjetnost za z v poljubnem razmaku.

Pojem tveganja

9.11

Popolnoma gotovih dogodkov v življenju ni veliko. Za večino dogodkov je le neka določena verjetnost, da se zgode. Vendar smatramo dogodke, za katere je velika verjetnost, da se zgode, praktično za gotove. Tako pri prečkanju ceste pričakujemo, da nas ne bo povozil avto, čeprav to ni gotovo in je neka verjetnost, da bomo povoženi. Pri prekoračenju ceste zato tvegamo, da bomo povoženi, ker je določena verjetnost, da se to

zgodí. Če vzamemo, da se bo dogodek, ki je samo verjeten, zagotovo zgodil, več ali manj tvegamo. Stopnja tveganja je enaka verjetnosti, da se dogodek, ki ga napovedujemo, ne zgodí.

Vsi zaključki, ki jih napravimo z metodami statistične indukcije, zavestno vsebujejo določeno stopnjo tveganja. Pri teh sklepih stopnjo tveganja ne samo poznamo, temveč jo moremo po želji uravnavati.

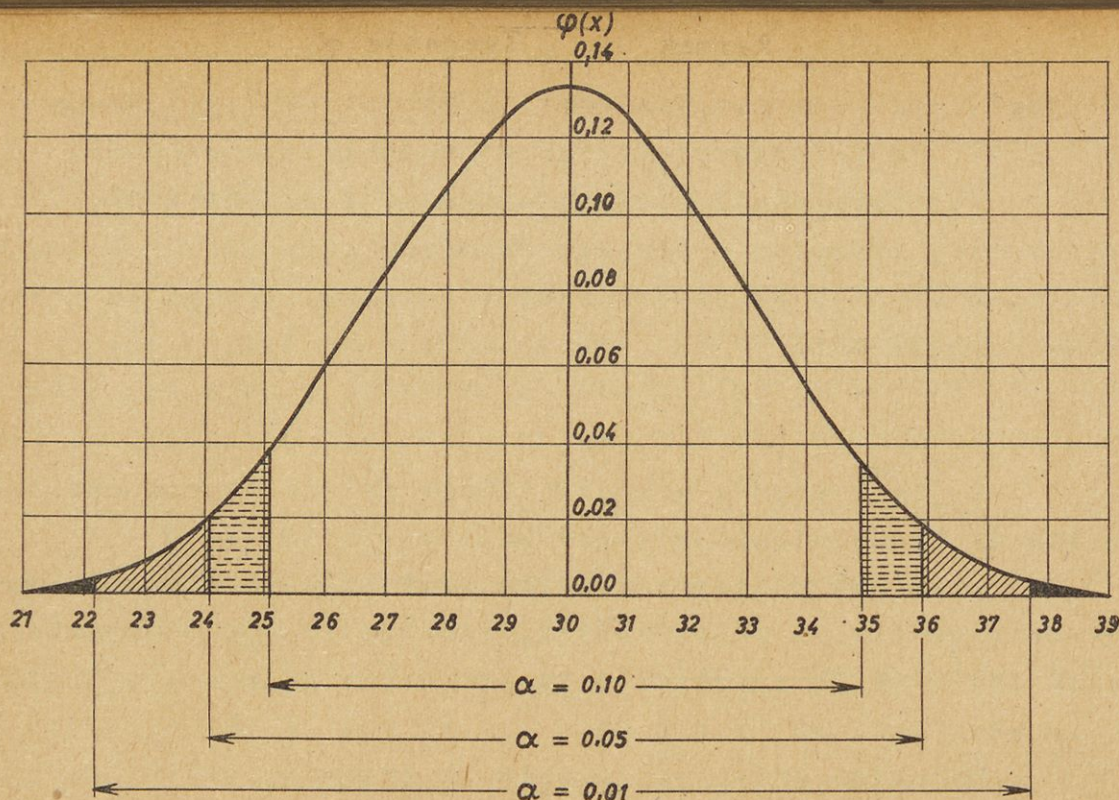
9.12

Pojem tveganja v statističnem smislu podrobneje ponazorimo na normalni porazdelitvi!

Vzemimo, da imamo idealen sestoj, v katerem se premeri dreves, merjeni v prsni višini, porazdeljujejo v normalni porazdelitvi. Za to populacijo je povprečen premer $\bar{y} = 30$ cm in standardni odklon $\sigma = 3$ cm.

Normalna krivulja relativnih frekvenc (ki je narisana v sliki 9.5) je slika o gostoti verjetnosti za slučajnostno spremenljivko – premer drevesa, če ima vsako drevo enako možnost, da ga izberemo iz populacije, ali z drugimi besedami, če drevo izberemo slučajnostno. Napravimo trditev: Če iz populacije dreves slučajnostno izberemo drevo, je premer izbranega drevesa večji kot $y_s = 24,1$ cm in manjši kot $y_z = 35,9$ cm. Kot vidimo iz slike, ta trditev ni absolutna in se more zgoditi, da slučajno izberemo drevo, ki ima ali manjši premer kot 24,1 cm ali večji premer kot 35,9 cm. V populaciji so namreč drevesa, ki imajo premere tudi izven tega razmaka. Verjetnost, da se naša trditev ne uresniči, je $\alpha = 0,05$. Trditev je napravljena s tveganjem $\alpha = 0,05$. To pomeni: če poskus slučajnostne izbire dreves ponavljamo, v povprečju v petih od sto ali v enem od dvajset poskusov zgornja trditev ne drži. Če razmak zožimo, je tveganje večje. Če n.pr. trdimo: Slučajnostno izbrano drevo ima premer v razmaku od 25,1 cm do 34,9 cm, je tveganje te trditve $\alpha = 0,10$.

Ta trditev se v povprečju v enem primeru od desetih izkaže za nepravilno. Nasprotno se stopnja tveganja zmanjša, če razmak razširimo. Tako je premer za slučajnostno izbrano drevo s tveganjem $\alpha = 0,01$ v razmaku med 22,3 cm do 37,7 cm. Če to trditev preskušamo in slučajnostno izbiramo drevesa, v



Slika 9.5 Verjetnostna porazdelitev premerov v normalno porazdeljenem sestoju z $\bar{Y} = 30$ cm in $\sigma = 3$ cm

povprečju le v enem od sto poskusov naša trditve ne drži. Zanesljivost te trditve je zelo velika. Stopnjo zanesljivosti, ki je verjetnost, da se določena trditve uresniči, in stopnjo tveganja, ki je verjetnost, da se določena trditve ne uresniči, za normalno distribuirane slučajne spremenljivke ugotovimo prek standardiziranega odklona z . Za normalno porazdelitev je namreč relativna frekvenca v razmaku $\bar{Y} - z\sigma$ do $\bar{Y} + z\sigma$ odvisna le od standardiziranega znaka z .

9.13

Ker je normalna porazdelitev ena izmed najvažnejših verjetnostnih porazdelitev, podajamo neke običajne trditve o slučajnostno izbranih enotah iz normalno porazdeljenih populacij in stopnje tveganja za te trditve.

Trditve: Če iz normalno porazdeljene populacije slučajnostno izberemo enoto, je standardizirani odklon z za vrednost x za to enoto v razmaku:

Razmak	Tveganje α
$-1,64 < z < +1,64$	0,10
$-1,96 < z < +1,96$	0,05
$-2,58 < z < +2,58$	0,01
$-3,29 < z < +3,29$	0,001
$-1,28 < z$	0,10
$-1,64 < z$	0,05
$-2,32 < z$	0,01
$-3,09 < z$	0,001
$z < +1,28$	0,10
$z < +1,64$	0,05
$z < +2,32$	0,01
$z < +3,09$	0,001

Če n.pr. trdimo, da je za vsako slučajnostno izbrano enoto iz normalno porazdeljene populacije standardiziran odklon z večji kot $-2,32$, je tveganje naše trditve $\alpha = 0,01$. To pomeni, da to trditev raziskovalec, ki jo preverja, z verjetnostjo $0,01$ ovrže.

9.14

Običajno postavljamo statistične trditve na stopnji tveganja $\alpha = 0,05$ ali za trditve, ki morajo biti iz katerega koli vzroka zanesljivejše, na stopnji $\alpha = 0,01$. Stopnja tveganja $\alpha = 0,10$ je razmeroma visoka, ker se trditev na tej stopnji tveganja izkaže kot napačna že v enem izmed desetih preskusov. Tveganje $0,001$ pa je praktično zelo visoko.

Pri trditvah, ki jih delamo s statističnimi metodami indukcije, se kosata dva momenta. Trditev, ki jo napravimo, je lahko določnejša, stopnja tveganja takih trditev pa je velika. Moremo pa trditev prirediti tako, da je stopnja tveganja poljubno majhna, vendar je trditev pri zmanjševanju tveganja bolj in bolj nedoločna. To smo uvideli že v prejšnjem odstavku. Prvi štirje razmaki za standardiziran znak so večji in večji, čim bolj stopnjo tveganja manjšamo. Če trdimo, da je standardiziran odklon za slučajnostno izbrane enote iz normalno porazdeljene populacije v razmaku med $-10,0$ do $+10,0$, v naši trditvi tveganja skoro ni. Pač pa je ta razmak tako velik, da je brez

pomena, ker da zelo nedoločeno predstavo o tem, kakšen je standardiziran znak za izbrano enoto. Nasprotno pa je trditev, da leži standardiziran odklon z za slučajnostno izbrano enoto iz normalne populacije v razmaku med $-1,0$ do $+1,0$ zelo določen. Stopnja tveganja pa je pri tem tako velika ($\alpha = 0,32$), da ta trditev nima praktične vrednosti.

9.15

Podobni problemi nastopijo pri vseh sklepih statistične indukcije, ne glede na to, kako je določena slučajnostna spremenljivka porazdeljena. Vsaka trditev in sklep ima določeno stopnjo tveganja, kar pomeni, da moremo pri ponavljanem preiskovanju trditev ovreči v skladu s stopnjo tveganja, s katero je trditev postavljena.

10.1

Metode delnega opazovanja. Metode popolnega opazovanja ne zadovoljijo več vedno večjih potreb po statističnih podatkih na nobenem področju in tudi ne v gozdarstvu. Dostikrat nimamo zadostnih finančnih in materialnih sredstev in časa, večkrat pa rezultat, ki ga iščemo, ni vreden naporov, ki bi bili potrebni, da bi ga iskali s popolnim opazovanjem. Ker so gozd in pojavi, ki so z njim v zvezi, populacije z izredno velikim številom enot, ni slučaj, da v gozdarstvu že razmeroma dolgo uporabljamo metode, s katerimi na lažji, cenejši in hitrejši način dobimo potrebne podatke, ki so, če že ne pravi rezultati, vsaj dobre in uporabne ocene. Izbor tipičnih "modelnih dreves", ki naj s svojimi značilnostmi predstavljajo celoto in na osnovi katerih sklepamo na celotno populacijo, je metoda, ki jo v gozdarstvu uporabljamo že dolgo. Vendar ima ta metoda, kljub svojim prednostim, omejeno vrednost, ker je subjektivna in zato ne vemo, v koliko je ocena izraz pravih razmer v populaciji.

Subjektivne metode ocenjevanja bolj in bolj zamenjuje objektivna in znanstveno zasnovana metoda slučajnostnega izbora ali vzorčenje. Prednosti vzorčenja pred drugimi metodami ocenjevanja so nesporne, kljub temu, da v posameznih primerih z drugimi metodami dobimo zanesljivejše ocene.

10.2

Vzorčenje. Vzorčenje je metoda statistične indukcije, s katero iz delne populacije enot, ki jih iz osnovne populacije izberemo naključno ali slučajnostno, sklepamo na celoto. Vzorčenje ima predvsem te prednosti:

Vzorčenje je objektivna metoda ocenjevanja. Napako ocene, ki izvira iz vzorčenja, moremo numerično oceniti. Razen tega jo moremo, glede na potrebe, tudi uravnavati.

V primerjavi s celotnim popisom pa se odlikuje vzorčenje po tem, da je znatno cenejše, hitrejšo in do neke mere tudi kvalitetnejše. Pri vzorčenju sicer nastopa vzorčna napaka, ki izvira iz tega, da je ocenjevani podatek rezultat vzorčenja. Nevzorčne napake, ki so rezultat nepravilnih osnovnih podatkov,

nepopolnega zajetja vseh enot itd., so pri popisih zaradi velikega obsega dela pogoste. Pri vzorčenju pa jih zmanjšamo na najmanjšo možno mero. Zaradi manjšega obsega opazovanja je namreč izbor in delo opazovalcev kvalitetnejše. Tako uspemo, da ne-
vzorčne napake z vzorčenjem zmanjšamo, na novo nastalo vzorčno napako pa moremo določiti in uravnavati. V raziskovalnem delu pa je vzorčenje edino možna metoda opazovanja in proučevanja, če proučujemo tako imenovane hipotetične populacije.

Vzorčenje ima tudi svoje hibe oziroma omejitve. Uporabno je le za velike populacije. Z vzorčenjem ne moremo dobiti podrobnih rezultatov. Zanesljivost ocen za različne vrste podatkov v istem vzorcu je različno. Ocenjevanje pojavov, ki so v populaciji redki, je razmeroma nezanesljivo. Plan za izvedbo vzorčenja in izračunavanje ocen sta bolj zamotana kot plan in izvedba za popolno opazovanje, če ne upoštevamo večjega obsega dela pri kompletnem opazovanju.

Induktivna metoda sklepanja iz dela na celoto se v vzorčenju uporablja z dvema ciljema. Z vzorčenjem ocenjujemo parametre populacije, kot so: lesna zaloga, število dreves, ki so okužena z določenim škodljivcem, povprečni premer dreves danega sestoja, variabilnost volumna dreves v sestoku itd. Z vzorci pa dostikrat tudi preskušamo hipoteze o populacijah. Tako z vzorci preskušamo hipoteze o učinkovitosti ukrepov v negi gozda, hipoteze o razlikah v rasti pod različnimi pogoji, hipoteze o sredinah, merah variacije za različne podatke iz gozdarstva.

10.3

Osnovna populacija. Enota opazovanja. Populacijo, katero proučujemo z vzorčenjem, imenujemo osnovno populacijo. Tako more biti osnovna populacija sestoj, za katerega iščemo lesno zalogo; vsa gospodarstva v Sloveniji, ki posedujejo gozd, če proučujemo ekonomiko gozdarstva v Sloveniji; posamezno drevo, če proučujemo kvalitete tega drevesa itd.

Osnovna populacija je sestavljena iz enot opazovanja. Enote opazovanja so za proučevanje vsebinsko pomembni elementi, ki sestavljajo osnovno populacijo. Če proučujemo določen sestoj, je osnovna populacija sestoj, enote opazovanja pa posamezna drevesa sestoka, ker skušamo z vzorčenjem dobiti pregled

o številčnem stanju, porazdelitvi dreves po debelini, višini, volumnu itd. Podobno so enote opazovanja posamezna gospodarstva, ki imajo gozd, če proučujemo skupnost vseh takih gospodarstev v Sloveniji. Pri določenem posebnem proučevanju morejo biti enote opazovanja vsi listi določenega drevesa, katerega proučujemo. V vseh navedenih primerih so enote opazovanja med seboj ločeni elementi, katerih število v osnovni populaciji je končno in točno določeno.

Zvezne populacije kot je n.pr. lesna masa danega drevesa, pa niso sestavljene iz nekih med seboj ločenih enot opazovanja.

Vzorčenje je edini način za proučevanje hipotetičnih osnovnih populacij. Določeno število poskusov o učinkovanju danega zaščitnega sredstva je vzorec iz neomejeno velike hipotetične populacije vseh možnih poskusov, ki bi jih izvedli pod enakimi pogoji.

10.4

Enote vzorčenja. Enote opazovanja niso vselej prikladne enote za izbiranje v vzorec. Tako je vzorčenje velikega sestojasila neprikladno, če so enote vzorčenja posamezna drevesa. Register vseh dreves, identifikacija posameznih izbranih dreves na terenu itd., je v tem primeru zamotan in obsežen posel. Zato za potrebe vzorčenja osnovno populacijo običajno razdelimo na vzorčne enote, ki so za vzorčenje prikladnejše kot osnovne enote. Zato površino sestojasila razdeljujemo v prikladnejše enote - pasove ali iregularne površinice, ki so določene z naravnimi mejami (potmi, potoki itd.). Edini namen razdelitve osnovne populacije na vzorčne enote je omogočiti in olajšati izbor. Vzorčne enote niso nujno povezane z vsebino pojava, ki ga proučujemo.

Prav posebno pride do izraza sestavljanje vzorčnih enot pri zveznih populacijah, za katere ne moremo vzeti za osnovo vzorčenja enote opazovanja, ker jih ni. Tako moremo na primer proučevati variabilnost v značilnostih lesa za določeno drevo le, če drevo razdelimo na vzorčne enote (n.pr. kvadre določene stalne izmere itd.). Površino sestojasila razdelimo na končno število vzorčnih enot, vzorčnih površin. Če pa je n.pr.

vzorčna enota kakorkoli položen kvadrat n.pr. s površino 25 a, moremo ta kvadrat na površino sestojajo položiti na neomejeno mnogo različnih načinov, ki se morejo med seboj tudi delno prekrivati. V tem primeru je to neskončna populacija vzorčnih enot.

10.5

Vzorec. Skupnost vzorčnih enot, ki jih iz osnovne populacije izberemo slučajnostno, da iz njih ocenimo podatke za osnovno populacijo, imenujemo vzorec.

Vzorec je populacija, ker ima svoje enote (vzorčne enote), te pa znake, ki so isti, kakor znaki za enote v osnovni populaciji. Kakor vsaka populacija ima tudi vzorec svoje parametre. Ti so n.pr. vsota podatkov za vse enote v vzorcu, proporci, povprečja, variance, korelacijski koeficienti itd. izračunani iz podatkov vzorca.

Število enot v vzorcu običajno zaznamujemo z n , za razliko od števila enot v osnovni populaciji, ki ga zaznamujemo z N . Ker je teorija vzorčenja različna, če je število enot v vzorcu majhno ali veliko, razlikujemo male vzorce od velikih vzorcev, glede na to, ali je število enot v vzorcu majhno ali veliko. Ostre meje med velikimi in malimi vzorci ni. Pač pa velja teorija malih vzorcev za vzorce z nekaj deset enotami, medtem ko imajo veliki vzorci običajno po več sto in tudi več tisoč enot. Medtem ko uporabljamo male vzorce predvsem pri eksperimentalnem delu in z njimi predvsem preskušamo hipoteze, z velikimi vzorci nadomeščamo popolna opazovanja in z njimi ocenjujemo parametre.

10.6

Po planu vzorčenja ima vsaka enota osnovne populacije dano verjetnost, da je vključena v vzorec. Ta verjetnost more biti za posamezne enote enaka ali različna. Pri enostavnem slučajnostnem vzorcu ima vsaka enota enako verjetnost, da je vključena v vzorec. Pri stratificiranem slučajnem izboru ima vsaka enota določenega dela - stratum enako verjetnost, da bo vključena v vzorec. Čeprav enostaven slučajen vzorec najpogostejše uporabljamo, je pri nekaterih metodah vzorčenja verjetnost izbora različna za vsako enoto. Imamo pa tudi metode delnega

opazovanja, pri katerih je izbor vseh ostalih enot v vzorcu določen s slučajnostnim izborom prve enote vzorca. Tak primer je sistematični vzorec, ki zaradi tega ne sodi čisto med slučajnostne vzorce, vendar ga zaradi njegovih posebnih prednosti zlasti v gozdarstvu pogosto uporabljamo.

Enota osnovne populacije, ki je izbrana v vzorec, more biti v nadaljevanju izbora ponovno izbrana, ali pa, glede na plan vzorčenja, nima več možnosti, da je ponovno vključena v vzorec. Glede na to govorimo o vzorčenju s ponavljanjem, kadar more biti posamezna enota osnovne populacije izbrana v vzorec večkrat in o vzorčenju brez ponavljanja, kadar more biti posamezna enota osnovne populacije izbrana v vzorec enkrat samkrat. Teorija vzorcev s ponavljanjem je nekako enostavnejša kot teorija vzorcev brez ponavljanja, vendar v praksi pogostejše uporabljamo vzorce brez ponavljanja, ker so bolj logični in dajo pri enako velikem vzorcu nekaj zanesljivejše ocene. Razlike med vzorci s ponavljanjem in vzorci brez ponavljanja pa se manjšajo, čim večja je osnovna populacija.

10.7

Populacija vseh možnih vzorcev. Vzorec, ki ga slučajnostno izberemo iz osnovne populacije, ni edini možni vzorec. Če pod enakimi pogoji izbor ponovimo, dobimo vzorec, v katerem z veliko verjetnostjo niso iste enote kot v prvem vzorcu.

Število vseh možnih različnih vzorcev je že za razmeroma majhne populacije in vzorce presenetljivo veliko. Matematično je število vseh možnih različnih vzorcev brez ponavljanja z n enotami, ki jih izberemo iz osnovne populacije, ki ima N vzorčnih enot, enako številu vseh možnih kombinacij po n elementov iz kompleksa N elementov. S simbolom izrazimo to

$$\binom{N}{n} = \frac{N(N-1)(N-2)\dots(N-n+1)}{1.2.3\dots n} \quad (10.1)$$

Število vseh slučajnostnih vzorcev s ponavljanjem pa je še večje. Če ima osnovna populacija N enot, je število vseh možnih različnih vzorcev s ponavljanjem z n enotami enako številu vseh možnih kombinacij s ponavljanjem n elementov iz osnovnega kompleksa N elementov, torej

$$\binom{N+n-1}{n} = \frac{N(N+1)(N+2)\dots(N+n-1)}{1.2.3\dots n} \quad (10.2)$$

Vzemimo za ilustracijo, kako veliko je število vseh možnih vzorcev, populacijo z $N=75$ enotami in izbirajmo vzorce z $n=25$ enotami! Že v tem primeru je število vseh možnih vzorcev brez ponavljanja 52,59 trilijonov, število vseh možnih vzorcev s ponavljanjem pa 181890 trilijonov. Že za neobičajno majhne populacije in vzorce je torej število vseh možnih vzorcev velikanško. V stvarnih primerih, v katerih je v osnovni populaciji več sto oziroma več tisoč enot in imajo tudi vzorci po več sto enot, pa je število vseh možnih vzorcev praktično neomejeno.

10.8

Če proučimo skupnost vseh možnih vzorcev, spoznamo, da ima ta skupnost vse lastnosti statističnih populacij. Vzorci so istovrstne količine, ki jih moremo šteti za enote v populaciji vseh možnih vzorcev. Enote v populaciji vseh možnih vzorcev - vzorci, imajo svoje znake, kot so povprečja, variance, relativna števila, koeficienti korelacije itd. v vzorcih. Ti znaki variirajo, ker so izračunani za vsak vzorec iz podatkov za različne enote iz osnovne populacije. Skupnost vseh možnih vzorcev je torej populacija, v kateri so posamezni vzorci enote, povprečja, proporci, variance, korelacijski koeficienti itd., izračunani iz vzorcev pa so znaki enot - vzorcev. Parametri v populaciji vseh možnih vzorcev pa so povprečja, variance itd. iz povprečij, proporcev, varianc, korelacijskih koeficientov itd. posameznih vzorcev.

10.9

Teoretične vzorčne porazdelitve. Neposredno proučevanje porazdelitev znakov in izračunavanje parametrov za populacije vseh možnih vzorcev je zaradi ogromnega števila vseh možnih vzorcev nemogoče. Praktično nemogoče je v stvarnih primerih sestaviti vse možne vzorce, za vsakega izmed njih izračunati n.pr. vzorčno povprečje, iz teh povprečij sestaviti frekvenčno porazdelitev, iz nje pa izračunati dalje aritmetično sredino in varianco iz povprečij za vse vzorce.

Izkaže pa se, da pridemo do teh količin po krajši poti s teoretičnim razglabljanjem, če zadosti podrobno poznamo osnovno populacijo. V dosti primerih moremo namreč sklepati na odnose v populaciji vseh možnih vzorcev, če poznamo osnovno populacijo. Te zakonitosti med osnovno populacijo in populacijo vseh možnih vzorcev pomaga odkriti verjetnostni račun. Prav zato, da veljajo osnovne zakonitosti med osnovno populacijo in populacijo vseh možnih vzorcev, je potrebno, da je vzorec izbran tako, da je zagotovljena naključnost izbora, ker le tedaj veljajo teoretični izsledki o medsebojnih odnosih. Te zakonitosti v populaciji vseh možnih vzorcev so za nekatere vrste osnovnih populacij in za nekatere probleme odkriti natančno, za druge pa samo približno. V obeh primerih pa jih izkoriščamo v praksi pri ocenjevanju in sklepanju iz vzorca na osnovno populacijo.

Zakonitosti v populaciji vseh možnih vzorcev so odvisne od količine, ki jo proučujemo, od osnovne populacije in od tipa vzorca. Vendar so pri velikih vzorcih te razlike vedno bolj zabrisane in so zakonitosti v populaciji vseh možnih vzorcev bolj in bolj enotne, čim večji je vzorec. Zato bomo splošna načela vzorčenja podrobno razložili na aritmetični sredini, ki je eden izmed najvažnejših in najpogostejših parametrov. Vse izsledke, ki veljajo za aritmetično sredino, pa zlahka prenesemo tudi na druge parametre.

Enostavno slučajnostno vzorčenje

Ocenjevanje aritmetične sredine

10.10

Vzorčenje, pri katerem ima vsaka enota populacije enako verjetnost, da je izbrana v vzorec, imenujemo enostavno slučajnostno vzorčenje.

Za enostavne slučajnostne vzorce, izbrane iz neomejene populacije, in za enostavne slučajnostne vzorce, ki so izbrani iz končnih populacij z izborom s ponavljanjem, veljajo za aritmetične sredine vzorcev $\bar{y}$ naslednje zakonitosti:

a) Aritmetična sredina iz povprečij $\bar{y}$ za vse možne vzorce je enaka aritmetični sredini $\bar{Y}$ v osnovni populaciji. V skladu s pojmi iz verjetnostnega računa se ta zakonitost glasi tudi takole: Matematično upanje za slučajnostno spremenljivko $\bar{y}$ slučajnostnih vzorcev je enako aritmetični sredini v populaciji $\bar{Y}$.

$$M(\bar{y}) = E(\bar{y}) = \bar{Y} \quad (10.3)$$

b) Varianca aritmetičnih sredin v slučajnostnih vzorcih $Var(\bar{y})$ je n krat manjša kot varianca znaka y v osnovni populaciji

$$Var(\bar{y}) = \frac{\sigma_y^2}{n} \quad (10.4)$$

Če iz variance aritmetičnih sredin v vzorcih izračunamo kvadratni koren, dobimo standardni odklon za aritmetične sredine v vzorcih, ki ga na splošno imenujemo standardna pogreška ocene, zaznamujemo pa ga z SE .

$$SE(\bar{y}) = \sqrt{Var(\bar{y})} = \frac{\sigma_y}{\sqrt{n}} \quad (10.5)$$

c) V praksi običajno proučujemo končne populacije z vzorci brez ponavljanja. Zanje pa veljajo za aritmetične sredine vseh vzorcev nekoliko drugačne zakonitosti.

Matematično upanje za aritmetične sredine vzorcev brez ponavljanja je še vedno enako aritmetični sredini v osnovni populaciji:

$$M(\bar{y}) = E(\bar{y}) = \bar{Y} \quad (10.6)$$

varianca za aritmetične sredine v vzorcih pa je enaka

$$Var(\bar{y}) = \frac{S_y^2}{n} \cdot \frac{N-n}{N} = \frac{S_y^2}{n} (1-f) \quad (10.7)$$

Pri tem je S_y^2 srednji kvadratični odklon za znak y , ki je izračunan po obrazcu

$$S_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1} \quad (10.8)$$

Razen neznatne spremembe, da je σ_y^2 zamenjam z nebištvono različnim izrazom S_y^2 , je v obrazcu dodan še korekturni faktor $\frac{N-n}{N} = 1-f$. Ta je tem bližji 1, čim manjši je vzorčni delež $f = n/N$, ki izraža, koliki del osnovne populacije je vključen v vzorec.

d) Ne glede na to, ali gre za enostavno vzorčenje z ali brez ponavljanja, se aritmetične sredine $\bar{y}$ iz vzorcev porazdeljujejo v normalni porazdelitvi, če se znak Y_i porazdeljuje v osnovni populaciji v normalni porazdelitvi.

Če se pa Y_i v osnovni populaciji ne porazdeljuje normalno, se aritmetične sredine $\bar{y}$ kljub temu porazdeljujejo v porazdelitvah, ki so normalni tem bolj podobne, čim večji je vzorec. Ta približek je dokaj dober že pri razmeroma majhnih vzorcih, tako da praktično vzamemo, da se aritmetične sredine iz velikih vzorcev porazdeljujejo v normalni distribuciji, ne glede na to, kakšna je porazdelitev znaka v osnovni populaciji.

Iz zgornjih zakonitosti povzamemo, da poznamo porazdelitev povprečij $\bar{y}$ v populaciji vseh možnih vzorcev, če poznamo parametre $\bar{y}$, σ^2 in S^2 v osnovni populaciji. Tako dobimo posredno porazdelitev povprečij $\bar{y}$, ki je za teorijo in prakso vzorčenja osnovnega pomena.

10.11

Točkovna ocena. Odklon zaupanja. Intervalna ocena.

Razmak zaupanja. Glede na zakonitosti v populaciji sredin v vzorcih in glede na lastnosti normalne porazdelitve v zvezi s stopnjo tveganja, na splošno postavimo trditev, da je aritmetična sredina za enostaven slučajnostni vzorec $\bar{y}$ s tveganjem $\alpha = 0,05$ v razmaku

$$\bar{Y} - 1,96 \cdot SE(\bar{y}) < \bar{y} < \bar{Y} + 1,96 \cdot SE(\bar{y})$$

Aritmetične sredine iz vzorcev se tem bolj goste okrog prave aritmetične sredine, čim manjša je standardna pogreška sredine. Ta pa je tem manjša, čim večji je vzorec.

Če iz sestoja, za katerega je $\bar{Y} = 30$ cm in $\sigma = 3$ cm izberemo enostaven slučajnostni vzorec z $n = 400$ drevesi, leži povprečje, izračunano iz vzorca, s tveganjem $\alpha = 0,05$, v razmaku

$$30 - 1,96 \frac{3}{\sqrt{400}} < \bar{y} < 30 + 1,96 \frac{3}{\sqrt{400}}$$

$$29,71 \text{ cm} < \bar{y} < 30,29 \text{ cm}$$

Povprečje iz vzorca se torej s tveganjem $\alpha = 0,05$ odklanja od pravega povprečja za največ 0,29 cm. Zato moremo vzeti povprečje iz vzorca $\bar{y}$ za oceno pravega povprečja $\bar{Y}$. To oceno imenujemo točkovno oceno, ker je dana z eno samo vrednostjo.

$$\bar{y}_{st} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = y/n \quad (10.9)$$

Količino $D(\bar{y}) = 1,96 SE(\bar{y})$ ali v splošnem

$$D(\bar{y}) = z SE(\bar{y}) \quad (10.10)$$

imenujemo odklon zaupanja. Ta pove, za koliko se z določenim tveganjem točkovna ocena za aritmetično sredino največ odklanja od prave sredine za populacijo. V našem primeru je s tveganjem $\alpha = 0,05$ odklon zaupanja $D(\bar{y}) = 0,29$ cm. S tveganjem $\alpha = 0,05$ ocena z vzorcem ni različna od prave vrednosti za več kot $D(\bar{y}) = 0,29$ cm.

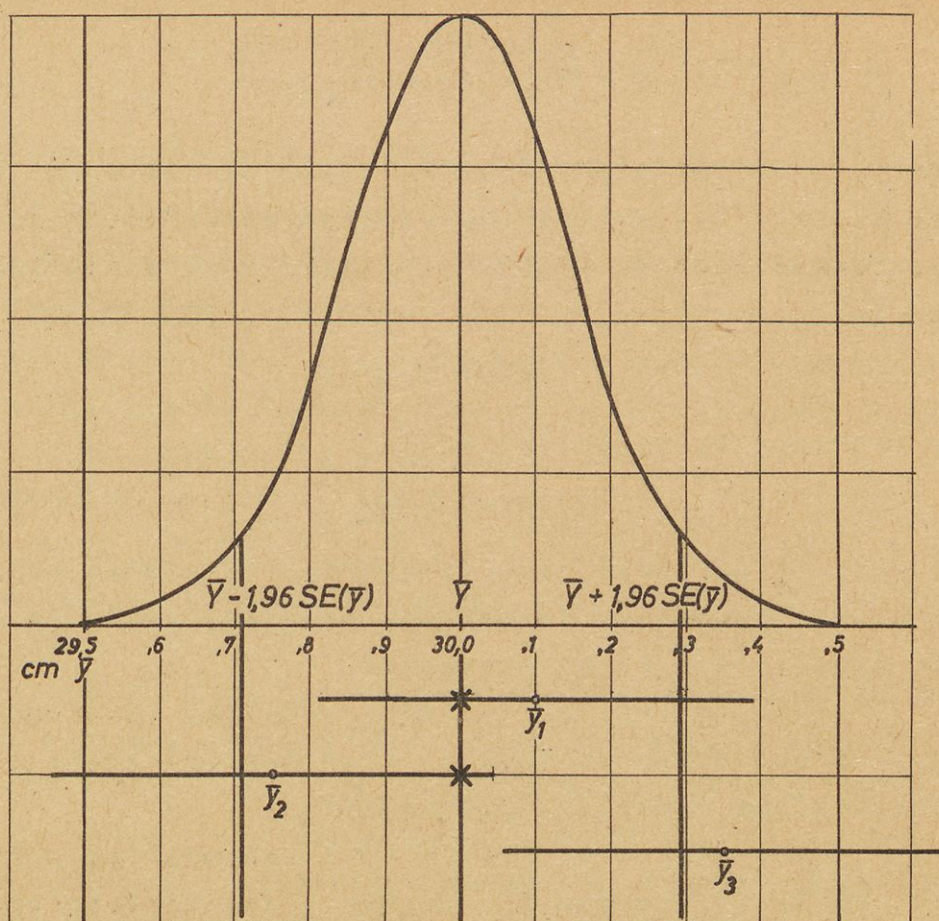
Če s tveganjem α ocena iz vzorca $\bar{y}$ ni od prave vrednosti različna za več kot $D(\bar{y})$, tudi prava vrednost povprečja $\bar{Y}$ ni od ocene $\bar{y}$ s tveganjem α različna za več kot $D(\bar{y})$. Iz tega sledi, da je slučajnostno izbrani vzorec s tveganjem $\alpha = 0,05$ tak, da je prava sredina v razmaku

$$\bar{y} - 1,96 SE(\bar{y}) < \bar{Y} < \bar{y} + 1,96 SE(\bar{y}) \quad (10.11)$$

Ta razmak imenujemo razmak zaupanja, obe meji pa spodnjo in zgornjo mejo zaupanja ocene. Ta ocena prave vrednosti povprečja je intervalna ocena, ker je dana z razmakom zaupanja, v katerem se z določenim tveganjem nahaja pravo povprečje.

V sliki 10.1 je nakazana porazdelitev sredin iz vzorcev za zgornji primer.

Iz slike vidimo, da za vzorce, za katere leže sredine v razmaku od 29,71 cm do 30,29 cm (glej $\bar{y}_1$ in $\bar{y}_2$), prava vrednost $\bar{Y}$ leži v razmakih zaupanja okrog ocen $\bar{y}$. Za ocene, ki leže izven tega razmaka, pa razmak zaupanja ne vključuje prave vrednosti $\bar{Y}$ (glej $\bar{y}_3$).



Slika 10.1 Porazdelitev sredin vzorcev s ponavljanjem z $n = 400$ enotami iz populacije premerov z $\bar{Y} = 30$ in $\sigma = 3$ cm

10.12

Nepristranska ocena variance in povprečnega kvadratnega odklona. Vendar obrazec 10.11 za praktično ocenjevanje povprečja z razmakom zaupanja nima posebnega pomena. Da izračunamo odklon zaupanja, moramo namreč poznati parameter σ_y^2 oziroma S_y^2 v osnovni populaciji. Ker pa teh parametrov ne poznamo, si pomagamo z oceno variance iz podatkov v izbranem slučajnostnem vzorcu.

Ocene variance σ_y^2 iz vzorca s ponavljanjem in oceno

povprečnega kvadratičnega odklona S_y^2 iz vzorcev brez ponavljanja ocenimo z enotnim obrazcem

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^n y_i^2 - y^2/n}{n-1} \quad (10.12)$$

To oceno imenujemo nepristransko oceno za varianco σ^2 , če je vzorec s ponavljanjem oziroma nepristransko oceno povprečnega kvadratičnega odklona S^2 , če je vzorec brez ponavljanja, ker je za vzorec s ponavljanjem

$$E(s^2) = \sigma^2 \quad (10.13a)$$

in za vzorec brez ponavljanja

$$E(s^2) = S^2 \quad (10.13b)$$

Dan izraz, ki ga izračunamo iz vzorca, je namreč nepristranska ocena nekega parametra iz osnovne populacije, če je matematično upanje tega izraza enako parametru.

Zato je tudi aritmetična sredina, izračunana iz vzorca, zaradi obrazcev 10.3 in 10.6 nepristranska ocena aritmetične sredine v osnovni populaciji.

Če v obrazcih 10.4 in 10.7 nadomestimo prave vrednosti za σ_y^2 in S_y^2 z oceno s_y^2 , dobimo oceni varianc za aritmetično sredino.

Za vzorce s ponavljanjem je ocena variance sredine enaka

$$\text{var}(\bar{y}) = \frac{s_y^2}{n} \quad (10.14)$$

za vzorce brez ponavljanja pa

$$\text{var}(\bar{y}) = \frac{s_y^2}{n} \cdot \frac{N-n}{N} = \frac{s_y^2}{n} \cdot (1-f) \quad (10.15)$$

10.13

Preskus zakonitosti vzorčenja na shematični populaciji. Zgornje zakonitosti za aritmetično sredino preskusimo na shematični populaciji z $N = 5$ enotami, iz katere izbiramo vzorce z $n = 3$ enotami! Osnovni podatki za to populacijo so: $Y_1 = 1, Y_2 = 2, Y_3 = 3, Y_4 = 4, Y_5 = 5$. Zanja je:

$$\bar{Y} = \frac{1}{5} (1+2+3+4+5) = 3$$

$$s_y^2 = \frac{1}{5} [(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2] = 2$$

$$s_y^2 = \frac{1}{5-1} [(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2] = 5/2$$

Najprej proučimo populacijo vzorcev s ponavljanjem in sestavimo sliko vseh možnih vzorcev s ponavljanjem s tremi enotami. Glede na obrazec 10.2 je število vseh možnih vzorcev s ponavljanjem enako številu kombinacij s tremi elementi s ponavljanjem iz kolektiva petih elementov, torej:

$$\binom{N+n-1}{n} = \binom{5+3-1}{3} = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} = 35.$$

Vseh teh 35 različnih vzorcev pa se ne pojavlja z enako verjetnostjo.

Medtem ko je samo ena možnost, da izberemo vzorec, v katerem v prvem, drugem in tretjem izvlečenju izberemo prvo enoto (111), so n.pr. tri različne možnosti, da izberemo vzorec, v katerem sta dvakrat prva, enkrat pa druga enota (112 121 211) in šest različnih možnosti za vzorec, v katerem je prva, druga in tretja enota (123 132 213 231 312 321). Ker je vsaka izmed teh permutacij enakomožna, vseh možnosti pa je v celoti 125, je verjetnost, da izberemo v vzorec trikrat prvo enoto $1/125$, da izberemo dvakrat prvo in enkrat drugo enoto $3/125$ in verjetnost, da izberemo v vzorec prvo, drugo in tretjo enoto, enaka $6/125$. Podobno je tudi za druge podatke in kombinacije.

V tabeli 10.1 so nanizani osnovni podatki o vsakem izmed petintridesetih različnih vzorcev s ponavljanjem z ustrezno verjetnostjo P , ocena za aritmetično sredino $\bar{y}$ in ocena variance s^2 .

Za vzorec, v katerega smo izbrali enote 113, je n.pr.

$$\bar{y} = \frac{1}{3} [1+1+3] = 5/3 \quad s^2 = \frac{1^2+1^2+3^2 - 5^2/3}{3-1} = 8/6$$

Iz zgornjih podatkov za posamezne vzorce sestavimo najprej verjetnostno porazdelitev za aritmetično sredino v vzorcih $\bar{y}$ kot slučajnostno spremenljivko. To verjetnostno po-

Tabela 10.1 Populacija vseh možnih vzorcev z $n=3$ enotami s ponavljanjem iz populacije z $N=5$ enotami. Osnovna populacija $Y: 1\ 2\ 3\ 4\ 5$

Osnovni podatki za vzorce	P	$\bar{y}$	s^2	Osnovni podatki za vzorce	P	$\bar{y}$	s^2
1 1 1	1/125	3/3	0	2 3 3	3/125	8/3	2/6
1 1 2	3/125	4/3	2/6	2 3 4	6/125	9/3	6/6
1 1 3	3/125	5/3	8/6	2 3 5	6/125	10/3	14/6
1 1 4	3/125	6/3	18/6	2 4 4	3/125	10/3	8/6
1 1 5	3/125	7/3	32/6	2 4 5	6/125	11/3	14/6
1 2 2	3/125	5/3	2/6	2 5 5	3/125	12/3	18/6
1 2 3	6/125	6/3	6/6	3 3 3	1/125	9/3	0
1 2 4	6/125	7/3	14/6	3 3 4	3/125	10/3	2/6
1 2 5	6/125	8/3	26/6	3 3 5	3/125	11/3	8/6
1 3 3	3/125	7/3	8/6	3 4 4	3/125	11/3	2/6
1 3 4	6/125	8/3	14/6	3 4 5	6/125	12/3	6/6
1 3 5	6/125	9/3	24/6	3 5 5	3/125	13/3	8/6
1 4 4	3/125	9/3	18/6	4 4 4	1/125	12/3	0
1 4 5	6/125	10/3	26/6	4 4 5	3/125	13/3	2/6
1 5 5	3/125	11/3	32/6	4 5 5	3/125	14/3	2/6
2 2 2	1/125	6/3	0	5 5 5	1/125	15/3	0
2 2 3	3/125	7/3	2/6				
2 2 4	3/125	8/3	8/6				
2 2 5	3/125	9/3	18/6				

razdelitev dobimo tako, da seštejemo verjetnosti za vse vzorce, ki imajo enako aritmetično sredino. Tako dobimo tabelo 10.2.

Iz dobljenih rezultatov spoznamo, da je res $E(\bar{y}) = 3 = \bar{Y}$.

Če pa izračunamo iz podatkov za populacijo po obrazcu 10.4 še

$$\text{Var}(\bar{y}) = \frac{6^2}{n} = \frac{2}{3}$$

spoznamo, da je tudi ta rezultat skladen z rezultatom, ki ga dobimo neposredno iz verjetnostne porazdelitve. Ostane še prekus stavka, da je ocena variance iz vzorca s^2 nepristranska ocena za varianco v osnovni populaciji 6^2 . Če podobno, kot za

Tabela 10.2 Verjetnostna porazdelitev za aritmetične sredine v vzorcih iz populacije $Y = 1, 2, 3, 4, 5$ za vzorce z $n=3$ s ponavljanjem

$\bar{y}$	P	$P \cdot \bar{y}$	$\bar{y} - E(\bar{y})$	$P \cdot (\bar{y} - E(\bar{y}))^2$
3/3	1/125	3/375	-6/3	36/1125
4/3	3/125	12/375	-5/3	75/1125
5/3	6/125	30/375	-4/3	96/1125
6/3	10/125	60/375	-3/3	90/1125
7/3	15/125	105/375	-2/3	60/1125
8/3	18/125	144/375	-1/3	18/1125
9/3	19/125	171/375	0	0
10/3	18/125	180/375	+1/3	18/1125
11/3	15/125	165/375	+2/3	60/1125
12/3	10/125	120/375	+3/3	90/1125
13/3	6/125	78/375	+4/3	96/1125
14/3	3/125	42/375	+5/3	75/1125
15/3	1/125	15/375	+6/3	36/1125

$$\frac{125}{125} = 1 \quad \frac{1125}{375} = 3 = E(\bar{y}) \quad \frac{750}{1125} = \frac{2}{3} = \text{Var}(\bar{y})$$

povprečja iz vzorcev, sestavimo verjetnostno porazdelitev še za oceno variance s^2 , dobimo tabelo 10.3

Tabela 10.3 Verjetnostna porazdelitev za ocene varianc s^2 iz vzorcev

s^2	P	$P \cdot s^2$
0	5/125	0
2/6	24/125	48/750
6/6	18/125	108/750
8/6	18/125	144/750
14/6	24/125	336/750
18/6	12/125	216/750
24/6	6/125	144/750
26/6	12/125	312/750
32/6	6/125	192/750

$$\frac{125}{125} = 1 \quad \frac{1500}{750} = 2 = E(s^2)$$

Rezultat pokaže skladnost s teorijo. Matematično upanje za varianco ocen je enako varianci v osnovni populaciji: $E(s^2) = \sigma^2 = 2$.

10.14

Napravimo za isto populacijo $Y(1\ 2\ 3\ 4\ 5)$ podoben preskus še za vzorce brez ponavljanja.

Število različnih vzorcev brez ponavljanja z $n=2$ enotami iz populacije z $N=5$ enotami je po obrazcu 10.1 enako

$$\binom{N}{n} = \binom{5}{2} = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} = 10. \text{ Če proučimo, kakšna je verjetnost za}$$

izbor posameznega vzorca brez ponavljanja, spoznamo, da moremo v vzorec brez ponavljanja izbrati iste enote na toliko različnih načinov, kolikor je permutacij iz $n=3$ elementov. V našem primeru torej na $3 = 1 \cdot 2 \cdot 3 = 6$ različnih načinov (n.pr. 123 132 213 231 312 321). Ker je skupno $5 \cdot 4 \cdot 3 = 60$ različnih možnosti za vse vzorce, je verjetnost za vsak vzorec $P = 6/60 = 1/10$. Vsak vzorec je pri vzorčenju brez ponavljanja enakoverjeten. Vsi možni vzorci za naš primer so nakazani v tabeli 10.4.

Tabela 10.4 Vzorci brez ponavljanja iz populacije $Y(1\ 2\ 3\ 4\ 5)$

Osnovni podatki za vzorec	p	$\bar{y}$	s^2
1 2 3	1/10	6/3	6/6
1 2 4	1/10	7/3	14/6
1 2 5	1/10	8/3	26/6
1 3 4	1/10	8/3	14/6
1 3 5	1/10	9/3	24/6
1 4 5	1/10	10/3	26/6
2 3 4	1/10	9/3	6/6
2 3 5	1/10	10/3	14/6
2 4 5	1/10	11/3	14/6
3 4 5	1/10	12/3	6/6

Enako kot za vzorce s ponavljanjem sestavimo tudi za vzorce brez ponavljanja verjetnostno porazdelitev za aritmetične sredine iz vzorcev in izračunajmo matematično upanje in varianco za aritmetične sredine iz vzorcev!

Tabela 10.5 Verjetnostna porazdelitev sredin $\bar{y}$ iz vzorcev brez ponavljanja

$\bar{y}$	P	$P \cdot \bar{y}$	$[\bar{y} - E(\bar{y})]$	$P[\bar{y} - E(\bar{y})]^2$
6/3	1/10	6/30	-3/3	9/90
7/3	1/10	7/30	-2/3	4/90
8/3	2/10	16/30	-1/3	2/90
9/3	2/10	18/30	0	0
10/3	2/10	20/30	+1/3	2/90
11/3	1/10	11/30	+2/3	4/90
12/3	1/10	12/30	+3/3	9/90
			90/30=3=E($\bar{y}$)	30/90=1/3=Var($\bar{y}$)

Če izračunamo $E(\bar{y})$ in $Var(\bar{y})$ posredno iz parametrov za osnovno populacijo, dobimo po obrazcih 10.6 in 10.7

$$E(\bar{y}) = \bar{Y} = 3$$

$$Var(\bar{y}) = \frac{s^2}{n} \cdot \frac{N-n}{N} = \frac{5/2}{3} \cdot \frac{5-3}{5} = \frac{1}{3}$$

Dobljena rezultata sta v skladu z rezultati, ki jih dobimo neposredno iz verjetnostne porazdelitve aritmetičnih sredin in vzorcev.

Preskusimo še stavek o nepristranosti ocene povprečnega kvadratičnega odklona s^2 . Najprej sestavimo verjetnostno porazdelitev za s^2 , tako da grupiramo rezultate iz stolpca za s^2 iz tabele 10.4

Tabela 10.6 Verjetnostna porazdelitev varianc s^2 v vzorcih

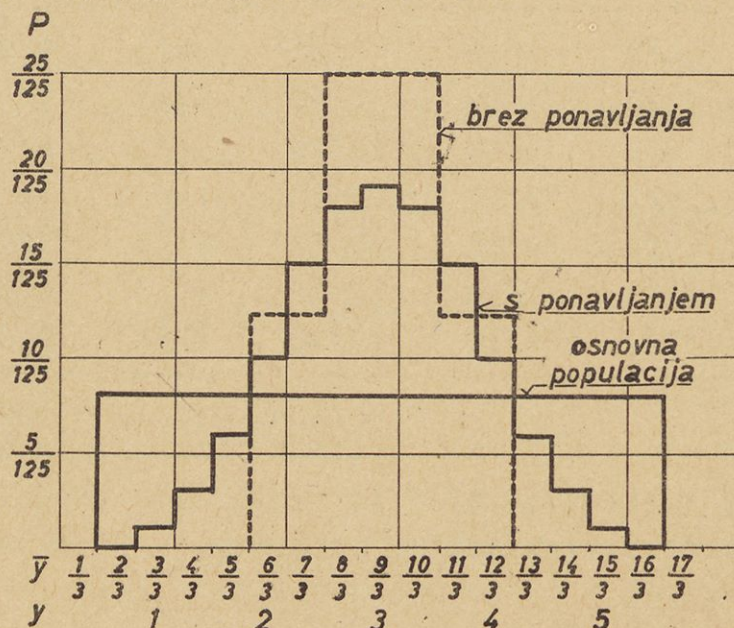
s^2	P	$P \cdot s^2$
6/6	3/10	18/60
14/6	4/10	56/60
24/6	1/10	24/60
26/6	1/10	52/60

$$\frac{10}{10} = 1 \quad \frac{150}{60} = \frac{5}{2} = E(s^2)$$

Če matematično upanje za oceno variance s^2 izračunamo posredno iz parametrov za populacijo, dobimo po obrazcu 10.13b

$$E(s^2) = s^2 = \frac{5}{2}$$

Tudi ta rezultata sta skladna.



Slika 10.2 Verjetnostni porazdelitvi za povprečja iz vzorcev z in brez ponavljanja

Iz verjetnostne porazdelitve aritmetičnih sredin vzorcev s ponavljanjem v sliki 10.2 vidimo težnjo k normalni porazdelitvi, ki smo jo navedli kot zakonitost. Čeprav porazdelitev osnovnih podatkov ni normalna in gre za izjemno majhne vzorce, se aritmetične sredine vzorcev porazdeljujejo v porazdelitvi, ki je simetrična, zvonasta, torej v porazdelitvi, ki ima iste značilnosti kot normalna. Ta težnja je manj vidna pri verjetnostni porazdelitvi sredin vzorcev brez ponavljanja, ker je vsega le deset možnih različnih vzorcev.

10.15

Ocenjevanje parametrov z enostavnim slučajnostnim vzorčenjem na splošno. Zgornji zaključki ne veljajo samo za ocenjevanje aritmetične sredine, ampak pri ocenjevanju za katerikoli drug parameter z enostavnimi velikimi vzorci. V približku se namreč za katerikoli parameter C ocene c porazdeljujejo normalno s sredino C in standardnim odklonom $SE(c)$.

Če s C zaznamujemo pravo vrednost ocenjevanega parametra, s c pa točkovno oceno tega parametra iz velikega vzorca, je odklon zaupanja za oceno

$$D(c) = z \cdot SE(c) \quad (10.16)$$

razmak zaupanja pa

$$c - z \cdot SE(c) < C < c + z \cdot SE(c) \quad (10.17)$$

s tveganjem, ki ustreza koeficientu z . Ker v praksi ne poznamo parametrov populacije, v odklonu zaupanja in razmaku zaupanja nadomestimo pravo vrednost za standardno pogreško $SE(c)$ z oceno iz vzorca, $se(c)$. Glede na to je ocena odklona zaupanja

$$d(c) = z \cdot se(c) \quad (10.18)$$

ocena za razmak zaupanja pa

$$c - z \cdot se(c) < C < c + z \cdot se(c) \quad (10.19)$$

Ocenjevanje agregata Y , strukturnega deleža $P\%$ in števila enot z dano značilnostjo H

10.16

Razen aritmetične sredine pogosto ocenjujemo vsoto vrednosti za dan znak, n.pr. skupen volumen, ali temeljnice za sestoj itd.

Med aritmetično sredino za populacijo $\bar{Y}$ in vsoto ali agregatom $Y = \sum Y_i$ je enostavna zveza

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N} \quad \text{ali} \quad \sum_{i=1}^N Y_i = Y = N \cdot \bar{Y} \quad (10.20)$$

Oceno agregata Y dobimo, če v obrazcu 10.20 nadomestimo pravo vrednost $\bar{Y}$ z oceno iz vzorca $\bar{y}$.

Ceno agregata z enostavnim slučajnim vzorcem Y_{sl} izračunamo torej po obrazcu

$$Y_{sl} = N \cdot \bar{y} = \frac{N}{n} \sum_{i=1}^n y_i = \frac{N}{n} y \quad (10.21)$$

pri čemer pomeni: y = vsota podatkov iz enostavnega slučajnega vzorca.

Ker je po znanih stavkih o varianci $\text{Var}(ax) = a^2 \text{Var}(x)$, velja, da je

$$\text{Var}(Y_{sl}) = \text{Var}(N\bar{y}) = N^2 \cdot \text{Var}(\bar{y}) = N^2 \frac{S_y^2}{n} \cdot \frac{N-n}{N} = N(N-n) \frac{S_y^2}{n} \quad (10.22)$$

Ker agregat običajno ocenjujemo z vzorčenjem brez ponavljanja, navajamo samo obrazec za ta tip vzorčenja.

Iz podatkov vzorca pa ocenjujemo varianco agregata $\text{var}(Y_{sl})$ po obrazcu, ki je prav tak kot obrazec 10.22, samo da je S^2 zamenjan z oceno s^2

$$\text{var} Y_{sl} = N(N-n) \frac{s^2}{n} \quad (10.23)$$

10.17

Pogosto ocenjujemo tudi strukturni delež P , ki ga dobimo, če število enot z dano značilnostjo H v populaciji delimo z obsegom populacije N

$$P = \frac{H}{N} \quad (10.24)$$

Oceno strukturnega deleža p dobimo iz podatkov enostavnega slučajnostnega vzorca analogno tako, da število enot z dano značilnostjo h iz vzorca delimo s skupnim številom enot v vzorcu n

$$P_{sl} = p = \frac{h}{n} \quad (10.25)$$

Varianca za oceno strukturnega deleža iz neomejene populacije ali iz vzorca s ponavljanjem je enaka

$$\text{Var}(p) = \frac{\sigma_p^2}{n} ; \quad \sigma_p^2 = P(1-P) \quad (10.26)$$

Za vzorce brez ponavljanja iz končnih populacij pa je

$$\text{Var}(p) = \frac{S_p^2}{n} \cdot \frac{N-n}{N} ; \quad S_p^2 = \frac{H(N-H)}{N(N-1)} \quad (10.27)$$

Analogno je ocena variance za ocene strukturnih deležev za vzorce iz neomejenih populacij in za vzorce s ponavljanjem iz končnih populacij

$$\text{var}(p) = \frac{s_p^2}{n} \quad (10.28)$$

za vzorce brez ponavljanja iz končnih populacij pa

$$\text{var}(p) = \frac{s_p^2}{n} \cdot \frac{N-n}{N} \quad (10.29)$$

Pri tem je izraz

$$s_p^2 = \frac{h(n-h)}{n(n-1)} \quad (10.30)$$

analogen povprečnemu kvadratnemu odklonu S_p^2 , le da je izračunan iz podatkov vzorca.

Opomba. Če strukturne deleže izražamo v odstotkih, je $P\% = 100 \cdot P$ in $p\% = 100 \cdot p$. Ker je $\text{Var}(p\%) = \text{Var}(100 \cdot p) = 100^2 \text{Var}(p)$, moramo v tem primeru vse obrazce za izračunavanje variance pomnožiti s 100^2 , če ocenjujemo strukturne deleže v odstotkih.

10.18

Število enot z dano značilnostjo H v populaciji dobimo, če

$$H = N \frac{H}{N} = NP \quad (10.31)$$

strukturni delež P pomnožimo s skupnim številom enot v populaciji N . Analogno kot pri oceni za vsoto pa dobimo oceno skupnega števila enot z dano značilnostjo H_{sl} , če v obrazcu 10.31 pravo vrednost strukturnega deleža zamenjamo z oceno p

$$H_{sl} = Np = N \frac{h}{n} \quad (10.32)$$

Podobno kot pri agregatih je varianca za oceno skupnega števila enot z dano značilnostjo za vzorce brez ponavljanja enaka

$$\text{Var}(H_{sl}) = N(N-n) \frac{S_p^2}{n} ; S_p^2 = \frac{H(N-H)}{N(N-1)} \quad (10.33)$$

$$\text{var}(H_{sl}) = N(N-n) \frac{s_p^2}{n} ; \quad s_p^2 = \frac{h(n-h)}{n(n-1)} \quad (10.34)$$

10.19

Simbolika in obrazci za ocene parametrov $\bar{Y}$, Y , P in H so sistematično nakazani v naslednjem pregledu.

	Populacija	Vzorec
Znak za seštevanje	$\sum_{i=1}^N$	$\sum_{i=1}^n$
Število enot	N	n
Število enot z dano značilnostjo	H	h
Strukturni delež	P	p
Strukturni postotek	$P\%$	$p\%$
Individualna vrednost	Y_i	y_i
Vsota za znak y	Y	y
Povprečje	$\bar{Y}$	$\bar{y}$
Povprečni kvadratni odklon	S_y^2	s_y^2
Varianca	σ_y^2	s_y^2
Povprečen kvadratni odklon za strukt. delež	S_p^2	s_p^2
Varianca strukturnega deleža	σ_p^2	s_p^2
	Prava vrednost	Ocena
Varianca	Var	var
Standardna pogreška	SE	se
Odklon zaupanja	D	d

Parametri:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i = Y/N \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = y/n$$

$$Y = \sum_{i=1}^N Y_i \quad Y_{sl} = \frac{N}{n} y$$

$$P = H/N \quad p = h/n$$

$$H \quad H_{sl} = \frac{N}{n} h$$

Variance ocen parametrov z
vzorcem s ponavljanjem:

$$\text{Var}(\bar{y}) = \frac{\sigma_y^2}{n}$$

$$\text{var}(\bar{y}) = \frac{s_y^2}{n}$$

$$\text{Var}(p) = \frac{\sigma_p^2}{n}$$

$$\text{var}(p) = \frac{s_p^2}{n}$$

Variance ocen parametrov z
vzorcem brez ponavljanja:

$$\text{Var}(\bar{y}) = \frac{s_y^2}{n} \cdot \frac{N-n}{N} \quad \text{var}(\bar{y}) = \frac{s_y^2}{n} \cdot \frac{N-n}{N}$$

$$\text{Var}(Y_{s1}) = N(N-n) \frac{s_y^2}{n}$$

$$\text{var}(Y_{s1}) = N(N-n) \frac{s_y^2}{n}$$

$$\text{Var}(p) = \frac{s_p^2}{n} \cdot \frac{N-n}{N}$$

$$\text{var}(p) = \frac{s_p^2}{n} \cdot \frac{N(N-n)}{N}$$

$$\text{Var}(H_{s1}) = N(N-n) \frac{s_p^2}{n}$$

$$\text{var}(H_{s1}) = N(N-n) \frac{s_p^2}{n}$$

$$\sigma_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N}$$

$$S_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$$

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

$$\sigma_p^2 = P(1-P)$$

$$S_p^2 = \frac{H(H-1)}{N(N-1)}$$

$$s_p^2 = \frac{h(h-1)}{n(n-1)}$$

Tehnika izbora enostavnega slučajnostnega vzorca

10.20

Prvi in glavni problem pri izvedbi enostavnega slučajnostnega vzorčenja je, kako izbrati slučajnostni vzorec, da bo zadostil osnovnemu pogoju, da ima vsaka enota vzorčenja v populaciji enako možnost, da je vključena v enostavni vzorec.

Pri vsakem sistemu vzorčenja ima posamezna enota populacije določeno, vnaprej dano verjetnost, da je vključena v vzorec. Način izbora enot mora biti tak, da zadosti temu osnovnemu pogoju, ker le v tem primeru veljajo zakonitosti vzorčenja, ki so izpeljane iz verjetnostnega računa. Pri enostavnem vzorčenju ima vsaka enota populacije enako verjetnost oziroma možnost, da je vključena v vzorec. Izbor takega vzorca imenujemo izbor brez omejitve.

Okvir vzorčenja. Tehnično izberemo enostavni slučajnostni vzorec po naslednjem postopku. Osnova slučajnostnega izbora je v vsakem primeru pregled vseh enot v populaciji. V tem pregledu, ki ga imenujemo okvir vzorčenja, je vsaka enota obvezno tako označena, da jo moremo identificirati. Pregled enot populacije, ki sestavlja okvir vzorčenja, more biti dan različno. Navadno je to spisec enot, ki so oštevilčene z zaporednimi številkami. Zelo prikladen okvir je tudi oštevilčena kartoteka enot populacije. Tudi geografska karta, v katero so vnešene enote populacije glede na geografsko lego in zaznamovane z zaporednimi številkami, more biti nazoren okvir populacije. Če so enote geografska območja (sestoj, oddelek, gospodarska enota ali podobno), je geografska karta teh območij, v kateri je vsako območje označeno z zaporedno številko, tudi primeren okvir vzorčenja. Kot okvir vzorčenja moremo uporabiti tudi vsako drugo shemo oziroma način, iz katerega je možno identificirati enote osnovne populacije.

Loterijski način. Eden izmed načinov slučajnega izbora je loterijski način. Pri loterijskem načinu imamo v žari listke z zaporednimi številkami enot. V žari je torej toliko listkov, kolikor je enot v populaciji.

Enako možnost za izbor vsake enote ustvarimo tako, da oštevilčene listke dobro premešamo in iz žare "na slepo" potegnemo listek. Enoto, ki ima izžrebano zaporedno številko, vzamemo kot enoto slučajnostnega vzorca. Postopek žrebanja enot vzorca ponavljamo vse dotlej, dokler nimamo izžrebanih ustrezno število enot n . Pri tem posamezne izžrebane listke pred izborom nove enote vračamo v žaro, če gre za izbor s ponavljanjem. Tako dosežemo, da ima ista izbrana enota možnost, da je ponovno vključena v izbor. Če pa izbiramo vzorec brez ponavljanja, izžreban listek izločimo, da ga ne moremo ponovno izbrati.

Loterijski način slučajnostnega izbora pa je posebno za večje populacije zamuden in okorel.

Isti cilj dosežemo na drug način, brez listkov za vsako enoto.

Vzemimo žaro, v kateri je deset enakih listkov, kroglic ali kock s številkami od 0 do 9. Če imamo populacijo z $N = 999$ enotami, dobimo slučajnostno zaporedno številko tako, da s trikratnim žrebanjem sestavimo tromestno slučajnostno številko. Izbor seveda vršimo s ponavljanjem. Če dobimo s prvim žrebom številko 3, z drugim 1, s tretjim žrebom pa številko 3, ta trojni žreb ustreza izžrebanj zaporedni številki 313 po prejšnjem načinu. V vzorec vključimo torej enoto, ki ima v okviru vzorčenja zaporedno številko 313. Prednost tega načina, v primerjavi s prvim, je v tem, da ni treba sestavljati obširnih žar, pomanjkljiv pa je v tem, da je treba za izbor ene same enote toliko žrebanj, kolikor mestno je število enot v populaciji N . To pa je zamuden posel.

Tablice slučajnostnih števil. Hibe loterijskega izbora odpravimo s tablicami slučajnostnih števil. Številke od 0 do 9, ki jih enkrat izberemo, zabeležimo. Te slučajnostne številke moremo uporabiti za več različnih vzorcev. Tako dobimo tablice slučajnostnih števil, v katerih so zapisane številke od 0 do 9, kot so bile po vrsti slučajnostno izbrane iz žare z desetimi številkami. En del iz tablic slučajnostnih števil iz knjige: Fisher and Yates: Statistical Tables for Biological, agricultural and medical research, je dana v tabeli 10.7.

Večmestna slučajnostna števila iz teh tablic dobimo, če združimo več slučajnostnih števil v skupine. Vzemimo prvo vrsto iz tablic slučajnostnih števil iz tabele 10.1:

03 47 43 73 86 36 96 47 36 61 46 98 63 71 itd. Zaporedne številke enot, ki jih iz populacije z $N=8324$ izberemo v slučajnostni vzorec, določimo tako, da vzamemo po vrsti skupine po štiri slučajnostne številke. Prva izbrana enota ima zaporedno številko 0347.

Skupine zaporednih štirimestnih slučajnostnih števil so: 0347 4373 8636 9647 3661 4698 6371.

Tabela 10.7 Slučajnostne številke

03	47	43	73	86	36	96	47	36	61	46	98	63	71	62	33	26	16	80	45	60	11	14	10	95
97	74	24	67	62	42	81	14	57	20	42	53	32	37	32	27	07	36	07	51	24	51	79	89	73
16	76	62	27	66	56	50	26	71	07	32	90	79	78	53	13	55	38	38	59	88	97	54	14	10
12	56	85	99	26	96	96	68	27	31	05	03	72	93	15	57	12	10	14	21	88	26	49	81	76
55	59	56	35	64	38	54	82	46	22	31	62	43	09	90	06	18	44	32	53	23	83	01	30	30
16	22	77	94	39	49	54	43	54	82	17	37	93	23	78	87	35	20	96	43	84	26	39	91	64
84	42	17	53	31	57	24	55	06	88	77	04	74	47	67	21	76	33	50	25	83	92	12	06	76
63	01	63	78	59	16	95	55	76	10	98	10	50	71	75	12	86	73	58	07	44	39	52	38	79
33	21	12	34	29	78	64	56	07	82	52	42	07	44	38	15	51	00	13	42	99	66	02	79	54
57	60	86	32	44	09	47	27	96	54	49	17	46	09	62	90	52	84	77	27	08	02	73	43	28
18	18	07	92	46	44	17	16	58	09	79	83	86	19	62	06	76	50	03	10	55	23	64	05	05
26	62	38	97	75	84	16	07	44	99	83	11	46	32	24	20	14	85	88	45	10	93	72	88	71
23	42	40	64	74	82	97	77	77	81	07	45	32	14	08	32	98	94	07	72	93	85	79	10	75
92	36	28	19	95	50	92	26	11	97	00	56	76	31	38	80	22	02	53	53	86	60	42	04	53
37	85	94	35	12	83	39	50	08	30	42	34	07	96	88	54	42	06	87	98	85	85	29	48	39
70	29	17	12	13	40	33	20	38	26	13	89	51	03	74	17	76	37	13	04	07	74	21	19	30
56	62	18	37	35	96	83	50	87	75	97	12	25	93	47	70	33	24	03	54	97	77	46	44	80
99	49	57	22	77	88	42	95	45	72	16	64	36	16	00	04	43	18	66	79	64	77	24	21	90
16	08	15	04	72	33	27	14	34	09	45	59	34	68	49	12	72	07	34	45	99	27	72	95	14
31	16	93	32	43	50	27	89	87	19	20	15	37	00	49	52	85	66	60	44	38	68	88	11	80
68	34	30	13	70	55	74	30	77	40	44	22	78	84	26	04	33	46	09	52	68	07	97	06	57
74	57	25	65	76	59	29	97	68	60	71	91	38	67	54	13	58	18	24	76	15	54	55	95	52
27	42	37	86	53	48	55	90	65	72	96	57	69	56	10	96	46	92	42	45	97	60	49	04	91
00	39	68	29	61	66	37	32	20	30	77	84	57	03	29	10	45	65	04	26	11	04	96	67	24
29	94	98	94	24	68	49	69	10	82	53	75	91	93	30	34	85	20	57	27	40	48	73	51	92
16	90	82	66	59	83	62	64	11	12	67	19	00	71	74	60	47	21	29	68	02	02	37	03	31
11	27	94	75	06	06	09	19	74	66	02	94	37	34	02	76	70	90	30	86	38	45	94	30	38
35	24	10	16	20	33	32	51	26	38	79	78	45	04	91	16	92	53	56	16	02	75	50	95	98
38	23	16	86	38	42	38	97	01	50	87	75	66	81	41	40	01	74	91	62	48	51	84	08	32
31	96	25	91	47	96	44	33	49	13	34	86	82	53	91	00	52	43	48	85	27	55	26	89	62

Tabela 10.7 Slučajnostne številke (nadaljevanje)

66	67	40	67	14	64	05	71	95	86	11	05	65	09	68	76	83	20	37	90	57	16	00	11	66
14	90	84	45	11	75	73	88	05	90	52	27	51	14	86	22	98	12	22	08	07	52	74	95	80
68	05	51	18	00	33	96	02	75	19	07	60	62	93	55	59	33	82	43	90	49	37	38	44	59
20	46	78	73	90	97	51	40	14	02	04	02	33	31	08	39	54	16	49	36	47	95	93	13	30
64	19	58	97	79	15	06	15	93	20	01	90	10	75	06	40	78	78	89	62	02	67	74	17	33
05	26	93	70	60	22	35	85	15	13	92	03	51	59	77	59	56	78	06	83	52	91	05	70	74
07	97	10	88	23	09	98	42	99	64	61	71	62	99	15	06	51	29	16	93	58	05	77	09	51
68	71	86	85	85	54	87	66	47	54	73	32	08	11	12	44	95	92	63	16	29	56	24	29	48
26	99	61	65	53	58	37	78	80	70	42	10	50	67	42	32	17	55	85	74	94	44	67	16	94
14	65	52	68	75	87	59	36	22	41	26	78	63	06	55	13	08	27	01	50	15	29	39	39	43
17	53	77	58	71	71	41	61	50	72	12	41	94	96	26	44	95	27	36	99	02	96	74	30	83
90	26	59	21	19	23	52	23	33	12	96	93	02	18	39	07	02	18	36	07	25	99	32	70	23
41	23	52	55	99	31	04	49	69	96	10	47	48	45	88	13	41	43	89	20	97	17	14	49	17
60	20	50	81	69	31	99	73	68	68	35	81	33	03	76	24	30	12	48	60	18	99	10	72	34
91	25	38	05	90	94	58	28	41	36	45	37	59	03	09	90	35	57	29	12	82	62	54	65	60
34	50	57	74	37	98	80	33	00	91	09	77	93	19	82	74	94	80	04	04	45	07	31	66	49
85	22	04	39	43	73	81	53	94	79	33	62	46	86	28	08	31	54	46	31	53	94	113	38	47
09	79	13	77	48	73	82	97	22	21	05	03	27	24	83	72	89	44	05	60	35	80	39	94	88
88	75	80	18	14	22	95	75	42	49	39	32	82	22	49	02	48	07	70	37	16	04	61	67	87
90	96	23	70	00	39	00	03	06	90	55	85	78	38	36	94	37	30	69	32	90	89	00	76	33
53	74	23	99	67	61	32	28	69	84	94	62	67	86	24	98	33	41	19	95	47	53	53	38	09
63	38	06	86	54	99	00	65	26	94	02	82	90	23	07	79	62	67	80	60	75	91	12	81	19
35	30	58	21	46	06	72	17	10	94	25	21	31	75	96	49	28	24	00	49	55	65	79	78	07
63	43	36	82	69	65	51	18	37	88	61	38	44	12	45	32	92	85	88	65	54	34	81	85	35
98	25	37	55	26	01	91	82	81	46	74	71	12	94	97	24	02	71	37	07	03	92	18	66	75
02	63	21	17	69	71	50	80	89	56	38	15	70	11	48	43	40	45	86	98	00	83	26	91	03
64	55	22	21	82	48	22	28	06	00	61	54	13	43	91	82	78	12	23	29	06	66	24	12	27
85	07	26	13	89	01	10	07	82	04	59	63	69	36	03	69	11	15	83	80	13	29	54	19	28
58	54	16	24	15	51	54	44	82	00	62	61	65	04	69	38	18	65	18	97	85	72	13	49	21
34	85	27	84	87	61	48	64	56	26	90	18	48	13	26	37	70	15	42	57	65	65	80	39	07

Prva izbrana enota ima torej zaporedno številko 347 (0 na prvem mestu odpade). Naslednja enota ima zaporedno številko 4373. Slučajnostni številki 8636 in 9647 ne prideta v poštev, ker sta večji kot je obseg populacije in ima zadnja enota v populaciji zaporedno številko 8324. Pač pa vključimo v vzorec še enote z zaporednimi številkami 3661, 4698 in 6371.

Skupine slučajnostnih zaporednih številok pa se morejo med seboj tudi delno prekrivati. Tako znatno povečamo obseg tablic slučajnostnih številok. Če za zgornji primer tvorimo štirimestne skupine slučajnostnih številok tako, da slučajnostne številke premikamo po eno mesto, dobimo te rezultate: Prva skupina 0347 ostane. Naslednjo skupino pa dobimo, če se v vrsti slučajnostnih številok premaknemo le za eno mesto. Druga skupina je torej: 3474. Naprej dobimo po istem postopku nadaljnje skupine: 4743 7437 4373 3738 itd. Namesto sedmih skupin po štiri ločene slučajnostne številke, izmed katerih jih moremo le pet uporabiti kot zaporedne številke izbranih enot, dobimo 25 skupin, od katerih je 21 manjših kot 8324. Namesto 5 enot iz istega dela tablic slučajnostnih številok izberemo 21 enot.

Po potrebi obseg tablic slučajnostnih številok povečamo tudi tako, da slučajnostne številke v tablici beremo v obratni smeri, od zgoraj navzdol ali od spodaj navzgor. V vsakem primeru dobimo nove skupine slučajnostnih številok.

10.25

V primeru, ki smo ga navedli, nekatere skupine štirimestnih slučajnostnih številok niso uporabne, ker presegajo obseg populacije. Število skupin, ki zaradi tega odpadejo, je v nekaterih primerih zelo veliko. Če je n.pr. obseg populacije $N=1832$, od 10000 različnih petmestnih skupin odpade $10000 - 1832 = 8168$ slučajnostnih števil. To pa zelo okrne uporabnost tablic.

S preprostim postopkom obseg uporabnih skupin povečamo. Od vseh štirimestnih skupin pridejo po zgornjem v konkurenco le one, ki imajo prvo številko 0 ali 1. Vse ostale izpadejo. Če pa postavimo pravilo, da štejejo kot 0 na prvem mestu štirimestnega števila vsako sodo (0, 2, 4, 6, 8), kot 1 pa vsako liho (1, 3, 5, 7, 9) slučajnostno številko, se uporabnost tablic

znatno poveča. Namesto 8168 neuporabnih skupin se število neuporabnih skupin od 8168 zniža na 840. Ker je n.pr. prva številka v skupini 5674 liha, ustreza tej slučajnostni številki enota z zaporedno številko 1674. Po prvem načinu pa je to slučajnostno število neuporabno.

Podobno reduciramo prva mesta skupin slučajnostnih števil za populacije, katerih obseg N začenja s številko 2. Številke 0, 3, 6 štejemo kot 0, slučajnostne številke 1, 4, 7 kot 1, slučajnostne številke 2, 5, 8 pa kot 2. Sistem je v tem, da je ustrezna prva številka ostanek slučajnostnega števila, če prvo slučajnostno številko od 0 do 8 delimo s tri. Opozoriti moramo, da slučajnostne skupine, ki začenjajo z 9, izpuščamo.

Analogno je pri populacijah, katerih obseg začne s številko 3, merodajen ostanek, če delimo prvo slučajnostno število s štiri, ne upoštevamo pa 8 in 9.

Če pa se število enot N začne s 4, je prva številka zaporednega števila ostanek, če delimo prvo slučajnostno številko s 5.

10.26

Primer za ocenjevanje z enostavnim slučajnostnim vzorcem. Uporabo različnih vrst vzorčenja, med njimi tudi za enostavno slučajnostno vzorčenje, bomo uporabili na shematični preskusni kvadratni ploskvi s površino 57,6 ha. Ta je razdeljena na $N = 24 \times 24 = 576$ osnovnih površin kvadratne oblike po 10 a. Kljub temu, da je primer shematičen, je populacija zadosti velika, da bomo mogli na različnih primerih oceniti kvalitete posameznih vrst vzorčenja. V skici ploskve v sliki 10.3 so nakazane osnovne površine in vanje vpisani podatki o volumnih na posameznih površinah. Ti podatki v nadaljevanju služijo za to, da z njihovo pomočjo dobimo podatke o izbranih parcelah. V stvarnem primeru dobimo za izbrane enote te podatke z opazovanjem na terenu.

Prave vrednosti parametrov, katere bomo v nadaljevanju ocenjevali, so: Število parcel, za katere je volumen večji kot 30 m^3 : $H = 320$. Odstotek parcel, na katerih je volumen večji kot 30 m^3 : $P\% = 55,6\%$. Skupen volumen sestaja: $Y = 17439 \text{ m}^3$. Povprečen volumen na osnovno parcelo : $\bar{Y} = 30,3 \text{ m}^3$.

23	21	22	20	29	33	32	31	31	27	29	32	36	29	36	30	35	43	36	37	36	37	33	36
19	24	18	22	33	24	21	28	23	34	34	28	31	32	27	38	32	35	34	33	43	34	41	43
23	20	25	21	32	29	30	25	32	29	27	34	28	36	39	30	32	37	36	37	36	41	38	33
22	22	20	22	23	23	24	28	30	24	32	30	27	31	32	29	35	33	35	35	41	35	41	38
21	22	21	20	24	33	28	29	22	32	30	30	35	30	31	32	38	34	34	34	33	34	33	38
23	18	21	22	22	31	20	25	34	26	28	35	29	36	30	31	38	37	35	37	36	35	41	36
22	21	23	21	20	28	28	24	28	24	33	31	33	31	32	39	34	33	37	33	35	37	35	41
23	20	21	23	24	23	25	29	27	30	36	25	31	37	35	32	35	36	34	36	38	34	44	33
20	25	18	25	20	29	32	21	30	26	32	27	28	34	34	39	34	37	36	33	35	37	35	42
22	19	24	22	22	30	23	30	36	25	30	38	32	33	32	36	36	33	34	39	42	35	33	38
23	23	20	23	25	32	27	33	25	26	27	30	37	35	37	33	34	39	35	34	34	34	42	35
24	21	23	19	23	22	23	25	26	24	34	37	33	33	36	35	36	35	39	33	37	35	34	40
20	23	24	22	21	20	22	22	33	34	28	29	34	35	31	36	34	40	33	37	34	36	40	35
22	19	23	25	20	21	24	33	27	27	31	37	35	34	36	35	39	36	36	35	39	40	33	38
23	21	21	23	22	19	21	26	25	31	34	34	33	37	33	34	36	33	37	39	34	34	36	34
22	19	22	20	24	22	20	34	30	34	29	34	35	32	35	35	34	37	34	35	40	35	40	38
20	23	21	21	18	25	28	26	22	27	30	35	36	35	33	39	37	35	36	34	36	34	36	33
23	20	24	22	21	20	27	33	28	26	29	31	33	34	34	34	35	33	37	36	38	40	35	38
22	21	22	19	26	22	33	26	33	30	29	33	37	33	36	39	34	39	34	36	33	34	38	35
24	18	25	21	22	20	34	38	24	27	30	29	28	36	34	34	35	34	37	35	39	38	34	36
23	22	20	22	19	21	22	28	26	24	31	29	35	35	30	34	35	36	34	37	35	33	42	33
24	22	23	21	20	23	24	32	33	31	28	31	29	31	35	35	32	37	36	33	38	38	35	38
20	21	19	26	22	21	19	26	21	24	26	29	34	28	29	37	36	39	34	35	40	35	34	35
23	22	23	21	21	24	22	20	24	21	33	27	29	34	31	30	34	35	34	40	34	45	38	35

Slika 10.3 Skica poskusne ploskve z N=576 osnovnimi parcelicami s podatki o volumnih na osnovnih ploskvah.

Z enostavnim slučajnostnim vzorcem brez ponavljanja $n=72$ osnovnih parcel na poskusni ploskvi, ocenimo:

- a) povprečni volumen na osnovno površino: $\bar{y}$
- b) skupni volumen v sestoji: Y_{sl}
- c) odstotek osnovnih parcelic v sestoji, ki imajo volumen manjši kot 30 m^3 : $p\%$
- d) število parcelic v sestoji, nakaterih je volumen manjši kot 30 m^3 : H_{sl}

Vzorčna enota je v našem primeru posamezna parcelica izmed $N=576$ osnovnih parcelic. Osnovna populacija pa je sestavljena iz vseh $N=576$ osnovnih parcelic.

Glede na to, da so osnovne enote površine, je najprikladnejši okvir shematična karta sestoja, v kateri so posamezne vzorčne enote - osnovne parcelice oštevilčene z zaporednimi številkami od 1 do 576.

Enostaven slučajnostni izbor $n=72$ osnovnih parcelic brez ponavljanja izvedemo tako, da iz tablic slučajnostnih števil vzamemo kot zaporedne številke izbranih enot vsa tromestna slučajnostna števila od 001 do 576. Tromestne slučajnostne številke sestavimo po določenem pravilu.

V našem primeru vzemimo po vrstnem redu tromestna števila iz zaporednih stolpcev v tablici slučajnostnih števil v dodatku.

Najprej vzemimo kot tromestna slučajnostna števila prvo skupino tromestnih števil (prva, druga, tretja) iz vsake vrste v tablici slučajnostnih števil v dodatku, nato drugo skupino (četrta, peta, šesta slučajnostna številka) iz vsake vrste, nato tretjo skupino itd. vse dotlej, dokler ne dobimo $n=72$ slučajnostnih tromestnih števil brez ponavljanja. $n=72$ tako ugotovljenih tromestnih slučajnostnih števil brez ponavljanja je nanizanih v tabeli 10.8. Iz te tabele je razvidno, da se je v izboru pet tromestnih števil ponovilo (dv). Izbrane parcele so v okvirju vzorca v sliki 10.3 uokvirjene. V tabeli 10.8, kakor tudi v sliki 10.4 so za izbrane parcele navedeni tudi volumni, katere smo ugotovili iz skice 10.3. Z 0 ali 1 pa je zaznamovano, da volumen ni (0) oziroma da je (1) večji kot 30 m^3 .

001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021	022	023	024
		0 22																					
025	026	027	028	029	030	031	032	033	034	035	036	037	038	039	040	041	042	043	044	045	046	047	048
							0 23			0 28												1 41	
049	050	051	052	053	054	055	056	057	058	059	060	061	062	063	064	065	066	067	068	069	070	071	072
					0 30																		
073	074	075	076	077	078	079	080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095	096
								0 24					1 31										
097	098	099	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
															1 32								
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
0 23				0 22												1 36							
145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
											1 33				1 39		1 33	1 37			1 37	1 35	
169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192
0 23		0 21									1 31												
193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216
					1 32																		
217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
0 22	0 19								0 30							1 36	1 33			1 42	1 35		1 38
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264
0 23			0 25																				
265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288
0 21								0 24		1 37	1 33						1 39						
289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312
						0 22			0 28													1 40	
313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336
			0 25			0 24				1 31	1 37								1 35		1 40		
337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
			0 23		0 19						1 34		1 37	1 33				1 39					
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384
											1 35						1 37				1 35		
385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408
									0 30														
409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432
0 23	0 20														1 34					1 38	1 40		
433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456
					0 22																		
457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
										0 29					1 34								
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504
0 22																1 36				1 33			
505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528
529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552
															1 37				1 35	1 35			
553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576
		0 23										1 34											1 35

Slika 10.4 Okvir enostavnega slučajnostnega vzorca $n=72$ parcelic z vrisanimi slučajnostno izbranimi parcelicami in vpisanimi volumni.

Tabela 10.8 Osnovni podatki o enostavnem slučajnostnem izboru vzorca z $n = 72$ parcelicami

Sluč. šte.	Volumen		Sluč. šte.	Volumen		Sluč. šte.	Volumen	
	y_i	$y_i > 30$		y_i	$y_i > 30$		y_i	$y_i > 30$
034	28	0	277	33	1	324	dv	dv
167	35	1	217	22	0	137	38	1
125	22	0	163	37	1	296	22	0
555	23	0	112	dv	dv	162	dv	dv
162	33	1	086	31	1	242	23	0
332	40	1	238	35	1	438	22	0
576	35	1	240	38	1	157	33	1
181	31	1	218	19	0	409	23	0
266	21	0	430	40	1	482	22	0
234	33	1	237	42	1	550	35	1
323	31	1	498	36	1	283	39	1
378	37	1	082	24	0	340	23	0
566	34	1	410	20	0	233	36	1
160	39	1	316	25	0	350	37	1
311	40	1	276	37	1	055	30	0
274	24	0	356	39	1	348	dv	dv
003	22	0	342	19	0	166	37	1
299	28	0	324	37	1	468	29	0
169	23	0	199	32	1	033	23	0
112	32	1	351	33	1	502	33	1
352	34	1	121	23	0	548	35	1
382	35	1	373	35	1	544	37	1
319	24	0	227	30	0	245	25	0
424	34	1	047	41	1	472	34	1
						171	21	0
						160	dv	dv
						395	30	0
						332	35	1
						429	38	1

Iz individualnih podatkov za $n=72$ izbranih parcel dobimo naslednje pomožne rezultate: število parcel, ki imajo volumen večji kot 30 m^3 , $h = 43$, vsota volumnov na izbranih parcelicah $y = \sum y_i = 2226 \text{ m}^3$. Vsota kvadratov volumnov na izbranih parcelicah $\sum y_i^2 = 71902$.

Če upoštevamo teoretične osnove za enostavni slučajnostni vzorec, dobimo naslednje rezultate:

Točkovna ocena za povprečni volumen na eno parcelo je po obrazcu 10.9

$$\bar{Y}_{s1} = y/n = 2226/72 = 30,92 \text{ m}^3$$

Iz pomožnih rezultatov dobimo dalje, da je ocena variance za osnovni znak y

$$s_y^2 = \frac{\sum_{i=1}^n y_i^2 - y^2/n}{n-1} = \frac{71902 - 2226^2/72}{72-1} = 43,40$$

Ocena variance za aritmetično sredino z vzorcem brez ponavljanja je po obrazcu 10.15

$$\text{var}(\bar{y}) = \frac{s_y^2}{n} \cdot \frac{N-n}{N} = \frac{43,40}{72} \cdot \frac{576-72}{576} = 0,527431$$

$$\text{se}(\bar{y}) = \sqrt{\text{var}(\bar{y})} = \sqrt{0,527431} = 0,726 \text{ m}^3$$

$$d(\bar{y}) = t(0,05; 72-1) \cdot \text{se}(\bar{y}) = 2,00 \cdot 0,726 = 1,45 \text{ m}^3$$

$$\bar{Y} = 30,92 \pm 1,45$$

Ker je vzorec z $n=72$ enotami v smislu vzorčenja še majhen, kot faktor, s katerim množimo oceno, da dobimo oceno odklona ni točno enak koeficientu z za normalno porazdelitev, temveč $t(\alpha, n-1)$. Ta koeficient je tem bližji vrednosti z za ustrežno tveganje α , čim večji je vzorec. To vidimo iz tabele za t porazdelitev v tabeli 11.1. V našem primeru je $t(\alpha=0,05; n-1=71)=2,00$. Za velike vzorce pa je ustrezní $z(\alpha=0,05)=1,96$. Razlika torej ni znatna.

Prava vrednost aritmetične sredine volumna na eno parcelo leži s tveganjem $\alpha=0,05$ v razmaku od $29,47 \text{ m}^3$ do $32,37 \text{ m}^3$.

Ocena aritmetične sredine volumna torej s tveganjem $\alpha=0,05$ ni od prave aritmetične sredine $\bar{Y}$ različna za več kot

$$d\%(\bar{y}) = 100 \cdot d(\bar{y})/\bar{y} = 100 \cdot 1,45/30,92 = 4,7 \%$$

Ocena, ki smo jo dobili z vzorcem, je razmeroma dobra.

Podobno ocenimo tudi skupen volumen v sestoji. Točkovna ocena skupnega volumna Y_{sl} je

$$Y_{sl} = \frac{N}{n} y = \frac{576}{72} 2226 = 17808 \text{ m}^3$$

Ocena variance za skupen volumen pa je

$$\text{var}(Y_{sl}) = N(N-n) \frac{s_y^2}{n} = 576(576-72) \frac{43,40}{72} = 174988,80$$

$$se(Y_{s1}) = \sqrt{\text{var}(Y_{s1})} = \sqrt{174988,80} = 418,32 \text{ m}^3$$

$$d(Y_{s1}) = \pm(0,0571) \cdot se(Y_{s1}) = 2,00 \cdot 418,32 = 837 \text{ m}^3$$

$$Y = 17808 \pm 837$$

Pravi skupni volumen sestoja je s tveganjem $\alpha = 0,05$ med 16971 m^3 in 18645 m^3 .

Če primerjamo dobljeni rezultat ocene z enostavnim vzorčenjem brez ponavljanja s pravim volumnom v raziskovanem sestoju $Y = 17439 \text{ m}^3$, vidimo, da je rezultat v skladu s teorijo. Prava vrednost volumna v sestoju resnično leži v razmaku zaupanja. Dobljena točkovna ocena za volumen v sestoju $Y_{s1} = 17808 \text{ m}^3$ je razmeroma dobra ocena skupnega volumna $Y = 17439 \text{ m}^3$. Razlika ocene od prave vrednosti je $17808 - 17439 = +369 \text{ m}^3$ ali v odstotkih $100 \cdot \frac{369}{17439} = 2,1 \%$

Ocena strukturnega deleža parcel z volumnom nad 30 m^3 je $p = h/n = 43/72 = 0,597$ ali v odstotkih

$$p\% = 100 \cdot 0,597 = 59,7\%$$

Da ocenimo varianco te ocene najprej iz podatkov vzorca, izračunamo oceno s_p^2

$$s_p^2 = \frac{h(n-h)}{n(n-1)} = \frac{43(72-43)}{72(72-1)} = 0,243936$$

Ocena variance za p je

$$\text{var}(p) = \frac{s_p^2}{n} \cdot \frac{N-n}{N} = \frac{0,243936}{72} \cdot \frac{576-72}{576} = 0,00296450$$

$$se(p) = \sqrt{\text{var}(p)} = \sqrt{0,00296450} = 0,0544$$

ali

$$se(p\%) = 100 \cdot se(p) = 100 \cdot 0,0544 = 5,44\%$$

$$d(p\%) = \pm(0,0571) \cdot se(p\%) = 2,00 \cdot 5,44 = 10,9\%$$

Odstotek parcel, ki imajo volumen večji kot 30 m^3 , je $P\% = 59,7\% \pm 10,9\%$. Pravi odstotek s tveganjem $\alpha = 0,05$ ni manjši kot 49% in ne večji kot 71% .

Če ocenimo še skupno število parcel, na katerih je volumen večji kot 30 m^3 , dobimo

$$H_{s1} = \frac{N}{n} \cdot h = \frac{576}{72} \cdot 43 = 344$$

$$\text{var}(H_{s1}) = N(N-n) \frac{s_p^2}{n} = 576 \cdot (576-72) \frac{0,243936}{72} = 983,55$$

$$\text{se}(H_{s1}) = \sqrt{\text{var}(H_{s1})} = \sqrt{983,55} = 31,4$$

$$d(H_{s1}) = t(0,05;71) \cdot \text{se}(H_{s1}) = 2,00 \cdot 31,4 = 62,8$$

$$H = 344 \pm 63$$

Skupno število parcel, na katerih je volumen večji kot 30 m^3 , je s tveganjem $\alpha = 0,05$ v mejah zaupanja med 281 in 407. Če za potrdilo teorije in za preskus stvarne zanesljivosti rezultata primerjamo dobljeno oceno s pravim številom parcel, ki smo ga dobili s pregledom vseh parcel, dobimo, da je pravo število parcel $H = 320$ v mejah zaupanja in da se ocena $H_{s1} = 344$ od pravega števila razlikuje za $344 - 320 = 24$ ali $100 \frac{24}{320} = 7,5\%$.

Določanje velikosti vzorca pri zahtevani natančnosti ocene

10.28

Določanje velikosti vzorca za oceno povprečja. Pri dosedanjem obravnavanju ocenjevanja z vzorčenjem je bila velikost vzorca vnaprej znana in smo pri dani velikosti vzorca ocenjevali parametre in njihovo zanesljivost.

Pri planiranju vzorcev pa običajno nastopi problem določiti velikost vzorca tako, da bo dal predpisano natančnost. Če na velikost vzorca ne polagamo pažnje, se zna zgoditi, da so ocene, ki jih dobimo, neuporabne ali pa imajo omejeno vrednost, ker so premalo zanesljive, če je vzorec premajhen. Obratno pa je ocena glede na stvarne potrebe lahko predobra in smo po nepotrebnem vzeli prevelik vzorec.

Zanesljivost ocene z enostavnim vzorcem je dana s standardno pogreško ocene SE, ali pa z odklonom zaupanja $D = z \cdot SE$,

ki pomeni največji odklon ocene od prave vrednosti pri danem tveganju α . Kot je znano, zavisi od stopnje tveganja koeficient z .

Najprej skušajmo rešiti problem o velikosti vzorca za oceno povprečja $\bar{y}$ z vzorcem s ponavljanjem. Vprašanje je, koliko enot moramo vzeti v vzorec, da se ocena povprečja $\bar{y}$ z danim tveganjem α ne odklanja od prave vrednosti za več kot $D(\bar{y})$.

Iz obrazca 10.5 in obrazca 10.6 povzamemo, da je za enostaven slučajnostni izbor s ponavljanjem

$$D(\bar{y}) = z \cdot SE(\bar{y}) = z \frac{\sigma_y}{\sqrt{n}} \quad (10.36)$$

Iz tega obrazca dobimo, da je za vzorec s ponavljanjem

$$n = \left(\frac{z \cdot \sigma_y}{D(\bar{y})} \right)^2 \quad (10.37)$$

Vzemimo našo poskusno ploskev z $N=576$ parcelicami in skušajmo določiti, kako velik vzorec s ponavljanjem moramo vzeti, da se ocena povprečja od pravega povprečnega volumna na eno parcelo s tveganjem $\alpha = 0,05$ ne bo odklanjala za več kot $D(\bar{y}) = 1 \text{ m}^3$.

Za tveganje $\alpha = 0,05$ je $z = 1,96$. Standardni odklon za volumen po parcelicah je $\sigma_y = 6,29$. Iz teh podatkov dobimo po obrazcu 10.37, da je

$$n = \left(\frac{1,96 \cdot 6,29}{1} \right)^2 = 152$$

Da z enostavnim vzorčenjem s ponavljanjem ocenjeni povprečni volumen na eno parcelo od pravega povprečja ne bo različen za več kot $D = 1 \text{ m}^3$, je treba v slučajnosten vzorec vključiti $n=152$ enot.

10.29

Med številom enot v vzorcu s ponavljanjem n in številom enot v vzorcu brez ponavljanja n' pri isti natančnosti ocene je zveza

$$n' = \frac{Nn}{N+n-1} \approx \frac{n}{1+f} \quad (10.38)$$

V našem primeru dobimo

$$n' = \frac{576.152}{576+152-1} = 120$$

Če ocenjujemo povprečje z vzorcem brez ponavljanja, dosežemo predpisano zanesljivost ocene že z vzorcem, v katerega vključimo le $n=120$ parcelic.

10.30

Velikost vzorca pri predpisani relativni natančnosti.

Zanesljivost ocene pa večkrat predpisujemo tudi relativno. Namesto absolutnega dopustnega odklona zaupanja izrazimo odklon relativno v odstotku od parametra, v našem primeru v odstotku od aritmetične sredine.

Če obrazec 10.37 ustrezno preuredimo, je

$$n = \left(\frac{z \text{ KV}\%}{D\%(\bar{y})} \right)^2 \quad (10.39)$$

V obrazcu 10.39 smo števec in imenovalec pomnožili s $100/\bar{Y}$. Tako dobimo v števcu namesto standardnega odklona σ koeficient variacije $\text{KV}\%$, v imenovalcu pa namesto absolutnega odklona zaupanja $D(\bar{y})$ relativen odklon $D\%(\bar{y})$. Iz tako dobljenega obrazca za število enot v vzorcu s ponavljanjem pa dobimo število enot v vzorcu brez ponavljanja po obrazcu 10.38.

10.31

Skušajmo določiti, s kako velikim vzorcem brez ponavljanja bi se ocena povprečnega volumna na eno parcelo v našem preskusnem sestoju s tveganjem $\alpha = 0,01$ ne odklanjala od pravega povprečja za več kot $D\%(\bar{y}) = 5\%$.

Tveganju $\alpha = 0,01$ ustreza koeficient $z = 2,576$ (glej odstavek 9.13). Ker je koeficient variacije za poskusni sestoj $\text{KV}\% = 20,8\%$, dobimo po obrazcu 10.39

$$n = \left(\frac{2,576.20,8}{5} \right)^2 = 115$$

Potrebno število enot v vzorcu brez ponavljanja pa je po obrazcu 10.38

$$n' = \frac{576 \cdot 115}{576 + 115 - 1} = 96$$

Določanje števila enot v vzorcu za ocene parametrov $\bar{Y}$, Y , $D\%$, H

10.32

Analogno dobimo obrazce za določanje števila enot v vzorcu tudi za ocene parametrov: Y , P in H . Obrazci za vse štiri osnovne parametre so dani v tabeli 10.9.

Tabela 10.9 Obrazci za določanje potrebnega števila enot pri dani zanesljivosti

Ocenjevani parameter		Število enot v vzorcu s ponavljanjem, če je maksimalni odklon predpisan	
		absolutno D	relativno D%
Aritmetična sredina	$\bar{Y}$	$n = \left(\frac{z \cdot \sigma}{D(\bar{Y})} \right)^2$	$n = \left(\frac{z \cdot KV\%}{D\%} \right)^2$
Vsota podatkov	ΣY	$n = \left(\frac{N \cdot z \cdot \sigma}{D(\Sigma Y)} \right)^2$	
Strukturni delež	$P\%$	$n = \frac{z^2 \cdot P\%(100-P\%)}{D(p\%)^2}$	$n = \left(\frac{100 \cdot z}{D\%} \right)^2 \cdot \frac{100-P\%}{P\%}$
Število enot z dano značilnostjo H		$n = \frac{z^2 H(N-H)}{D(H)^2}$	
Število enot v vzorcu brez ponavljanja		$n' = \frac{N \cdot n}{N+n-1}$	

Hiba navedenega postopka je v tem, da moramo poznati določene parametre iz osnovne populacije, če hočemo določiti potrebno število enot za planirani vzorec. Teh pa običajno nimamo. Zato si pri planiranju vzorcev pomagamo z ocenami za potrebne parametre. Te dobimo s predhodno analizo pojava, ki ga proučujemo ali iz razpoložljivih podatkov že izvršenih sorodnih raziskav. Včasih vrednosti teh parametrov ocenimo z manjšim vzorcem, katerega kasneje vključimo v glavni vzorec.

Drug problem, ki se pojavi v zvezi z določanjem velikosti vzorca, pa je v tem, da običajno z enim vzorcem ne opazujemo enega samega znaka in ne ocenjujemo enega samega parametra. Vsak izmed teh znakov pa ima drugo variabilnost, ocena pa drugo zanesljivost. Zato pri planiranju velikosti vzorca dobimo za oceno vsakega parametra drugačno potrebno velikost vzorca. Težko se je izmed različnih n odločiti za velikost vzorca, ki naj bo osnova za vzorec, s katerim dobimo podatke za ocene vseh parametrov. V tem primeru se odločimo za kompromisno srednjo velikost vzorca. Pri tem se oziramo predvsem na važnejše parametre in skušamo vzeti vzorec tako velik, da zanesljivost ocen ustreza predvsem zanje.

Stratificirano vzorčenje

Osnova in lastnosti. Obrazci o varianci ocen z enostavnim vzorčenjem pokažejo, da je zanesljivost ocene, razen od velikosti vzorca v bistveni meri odvisna od variabilnosti proučevane populacije. Za homogene populacije, za katere je variabilnost majhna, je ocena boljša, za heterogene populacije oziroma za populacije z veliko variabilnostjo, pa so ocene pri isti velikosti vzorca manj zanesljive. Medtem ko moremo velikost vzorca spreminjati in tako vplivati na kvaliteto ocene, na variabilnost neposredno ne moremo vplivati, ker je lastnost populacije.

Vendar kljub temu pri vzorčenju izkoriščamo zakonitost, da so ocene za homogene populacije zanesljivejše kot za heterogene. Če namreč uspemo heterogeno populacijo razdeliti v

homogene populacije, je variabilnost v posameznih delnih populacijah, ki jih imenujemo - stratume, manjša. S samostojnimi enostavnimi vzorci v posameznih stratumih dobimo ocene za posamezne stratume, iz teh ocen pa sestavimo ocene za populacijo. Tak način ocenjevanja je v primerjavi z enostavnim vzorcem tem uspešnejši, čim manjša je variabilnost pojava znotraj stratumov in čim večje so razlike med stratumi. Zato skušamo pri stratificiranem vzorčenju razdeliti populacijo v stratume po onem znaku oziroma znakih, ki bistveno vplivajo na proučevani pojav. Najpopularnejša je regionalna stratifikacija, pri kateri populacijo razdelimo v homogene grupe glede na geografske rajone. Tako je uspešna razdelitev sestoj na stratume po kvaliteti tal, glede na nadmorsko višino itd. Kot načelo za razdelitev na stratume pa more služiti vsak drug kriterij, ki prispeva k temu, da so razmere v stratumi čim bolj izenačene. Tako moremo v dvoetažnem sestoju vzeti kot stratume posamezni etaži. Volumen, premeri, višine drevja itd. so namreč v bistveni meri odvisni od tega, iz katere etaže je drevo. Zato je variabilnost teh značilnosti močno odvisna od etaže. Dve homogeni populaciji dobimo, če vzamemo v en stratum drevesa iz prve etaže, v drug stratum pa drevesa iz druge etaže.

Hiba stratificiranega vzorčenja pa je predvsem v tem, da moramo populacijo razmeroma dobro poznati, če hočemo izvesti stratifikacijo. Medtem ko moramo pri enostavnem vzorčenju poznati le okvir vzorčenja, t.j. spisec enot, moramo pri stratificiranem vzorčenju populacijo znatno bolje poznati, če jo hočemo razdeliti v homogene dele. Vsako enoto v osnovni populaciji je treba namreč pred izborom vzorca vključiti v ustrezno grupo - stratum.

Običajno ne proučujemo popolnoma neznanih pojavov. Zato izvedemo stratifikacijo po poznavanju o kvaliteti drevja, po poznavanju tal ali na osnovi prejšnjih pregledov sestoj.

Ker je stratifikacija običajno izvedena po nekem načelu, ki je v vsebinski zvezi s proučevanim pojavom, dobimo s stratificiranim vzorčenjem kot postranski rezultat tudi rezultate po grupah - stratumi. Zavedati pa se moramo, da so rezultati za posamezne stratume nezanesljivi, če je število enot vzorcev v posameznih stratumi majhno.

Ocenjevanje agregata s stratificiranim vzorcem. Nakazali smo že, da pri stratificiranem vzorčenju iz delnih ocen, ki jih dobimo iz samostojnih enostavnih vzorcev po stratumih, sestavimo skupno oceno za celotno populacijo.

Ocena agregata s stratificiranim vzorčenjem Y_{str} je vsota ocen agregatov v posameznih stratumih $Y_{k,sl}$

$$Y_{str} = Y_{1,sl} + Y_{2,sl} + \dots + Y_{r,sl} = \sum_{k=1}^r Y_{k,sl} \quad (10.40)$$

Varianca za oceno agregata s stratificiranim vzorčenjem $Var(Y_{str})$ pa je vsota varianc ocen agregatov po stratumih

$$Var(Y_{str}) = Var(Y_{1,sl}) + Var(Y_{2,sl}) + \dots + Var(Y_{r,sl}) = \sum_{k=1}^r Var(Y_{k,sl}) \quad (10.41)$$

Iz teh dveh obrazcev moremo z upoštevanjem zvez med $\bar{Y}$, Y , P in H , pravimi vrednostmi in ocenami razviti analogne obrazce za prave vrednosti ali ocene parametrov ali varianc za katerikoli izmed navedenih štirih parametrov.

Razmestitev enot vzorčenja po stratumih. Pri stratificiranem vzorčenju je ena izmed osnovnih nalog določiti, koliko enot od skupnega števila enot v vzorcu vzamemo v posamezen stratum. Vnaprej ta razmestitev namreč ni dana. Pri podrobnejšem študiju stratificiranega vzorčenja opazimo, da je od tega, koliko enot od skupnega števila enot v celotnem vzorcu je v posameznem stratumu, v bistveni meri odvisna zanesljivost ocene. Zato je ta problem toliko bolj pomemben.

Proporcionalna razmestitev. Najenostavnejša razmestitev, ki jo uporabljamo pri stratificiranem vzorčenju, je proporcionalna razmestitev. Pri proporcionalni razmestitvi je velikost vzorcev v posameznih stratumih sorazmerna skupnemu številu enot v stratumih. Če z N_k zaznamujemo skupno število enot, z n_k pa število enot v vzorcu v stratumu k , pri proporcionalni razmestitvi določimo n_k po obrazcu

$$n_k = a \cdot N_k ; a = n/N \quad (10.42)$$

10.38

Vzemimo naš kompleksen primer sestoj z $N = 576$ parcelicami, ki so razdeljene v tri stratum 1, 2 in 3. Osnovna razdelitev na stratum za to populacijo je nakazana v okvirju v sliki 10.5.

Zanj je v tabeli 10.10 nakazano, kako skupno $n = 72$ enot vzorca razmestimo po stratumih pri proporcionalni razmestitvi. Razen tega je v isti tabeli tudi nakazano, kolika je varianca za oceno skupnega volumna pri proporcionalni razmestitvi. Konstanta, s katero pomnožimo N_k , da dobimo n_k , je po obrazcu 10.42 $a = 72/576 = 0,125$.

Tabela 10.10 Izračun števila enot po stratumih in variance agregata (lesne zaloge) pri proporcionalni razmestitvi na prekusni ploskvi

Stratum	N_k	n_k	s_k	$\text{Var}(Y_{k,sl})$
1	144	18	3,2552	3281
2	192	24	14,5720	19585
3	240	30	6,4508	10837

$$N = 576 \quad n = 72$$

$$33703 = \text{Var}(Y_{str,p})$$

$$\text{Dalje je: } SE(Y_{str}) = \sqrt{\text{Var}(Y_{str})} = \sqrt{33703} = 183,6$$

$$SE\%(Y_{str,p}) = 100 \cdot SE(Y_{str,p}) / Y = 100 \cdot 183,6 / 17439 = 1,05\%$$

Stratificirano vzorčenje da v našem primeru torej znatno boljše rezultate kot enostavno vzorčenje. Medtem ko je relativna pogreška za oceno agregata Y z enostavnim vzorčenjem enaka $SE\%(Y_{sl}) = 2,29\%$, je relativna pogreška ocene pri stratificiranem vzorčenju s proporcionalno razmestitvijo skrčena na $SE\%(Y_{str,p}) = 1,05\%$.

Uspešnost posameznih vrst vzorčenja običajno merimo s primerjavo varianc ocen pri posameznih planih vzorčenja. To razmerje namreč približno pokaže, kolikokrat večji enostavni vzorec bi morali vzeti, če bi z njim hoteli doseči isto natanč-

nost, kot jo dosežemo z določenim tipom vzorčenja. V našem primeru je to razmerje

$$\frac{\text{Var}(Y_{sl})}{\text{Var}(Y_{str,p})} = \frac{160180}{33703} = 4,75$$

To pomeni: Če hočemo z enostavnim vzorcem doseči tako kvaliteto za oceno lesne zaloge, kot jo dosežemo s stratificiranim vzorcem s proporcionalno razmestitvijo enot s skupno $n = 72$ osnovnimi površinami, moramo v enostaven slučajen vzorec vzeti približno 4,75 krat več enot kot pri stratificiranem vzorčenju. Stratifikacija je v tem primeru zelo uspešna.

Varianco ocene agregatov v posameznih stratumih smo izračunali po znanem obrazcu 10.22 za izračunavanje variance za ocene agregatov z enostavnimi slučajnostnimi vzorci.

10.39

Optimalna razmestitev. Kljub uspešnosti proporcionalne razmestitve pri stratificiranem vzorčenju pa proporcionalna razmestitev ni najboljša. Ocena v posameznih stratumih je tem zanesljivejša, čim manjša je variabilnost v stratumu in obratno. V stratumih, ki so zelo homogeni, je potrebno za razmeroma dobro oceno vzeti sorazmerno manjše vzorce kot v variabilnejših stratumih. Iz tega zaključimo, da dobimo dovoljše razmestitev enot v stratumih, če razen števila enot v posameznih stratumih pri določanju velikosti vzorcev upoštevamo tudi variabilnost. Račun pokaže, da je optimalna razmestitev enot v stratumih dana z obrazcem

$$n_k = a \cdot N_k \cdot S_k \quad a = n / \sum n_k \cdot S_k \quad (10.43)$$

Po tem obrazcu dobimo dano število enot v vzorcu optimalno razmeščeno po stratumih, če vzamemo, da je število enot v posameznem stratumu proporcionalno številu enot N_k in standardnemu odklonu S_k v ustreznem stratumu.

Za naš primer je izračun optimalne razmestitve enot in variance ocene za tak stratificiran vzorec nakazan za oceno skupne lesne zaloge v tabeli 10.11.

Tabela 10.11 Izračun velikosti vzorcev v stratumi in variance ocene skupne lesne zaloge na preskusni ploskvi pri optimalni razmestitvi

Stratum	N_k	S_k	$N_k S_k$	n_k	$\text{Var}(Y_{k,sl})$
1	144	1,80	259,20	12	5156
2	192	3,82	733,44	33	13480
3.	240	2,54	609,60	27	12214
	576		1602,24	72	30850
	N		$\sum N_k S_k$	n	$\text{Var}(Y_{str,O})$

Število enot po stratumi dobimo po obrazcu 10.43, pri čemer je $a = n / \sum N_k S_k = 72 / 1602,24 = 0,044937$.

Osnovne podatke o N_k in S_k smo dobili iz tabele 10.9.

Dalje je:

$$SE(Y_{str,O}) = \sqrt{\text{Var}(Y_{str,O})} = \sqrt{30850} = 175,6$$

Relativna standardna pogreška pa $SE\%(Y_{str,O}) = 100 \cdot SE(Y_{str,O}) / Y = 100 \cdot 175,6 / 17439 = 1,01\%$.

Primerjava variance za stratificirano vzorčenje z optimalno razmestitvijo s stratificiranim vzorčenjem s proporcionalno razmestitvijo da

$$\text{Var}(Y_{str,p}) / \text{Var}(Y_{str,O}) = 33703 / 30850 = 1,092.$$

Da bi s proporcionalno razmestitvijo dobili enako zanesljivost kot z optimalno razmestitvijo $n=72$ enot vzorca, bi morali vzeti približno za 9% večji vzorec.

10.41

Optimalna razmestitev enot glede na stroške. Optimalna razmestitev, ki smo jo obravnavali v zgornjem odstavku, je optimalna glede na število enot. V praktičnem delu pa so pomembnejši

element pri planiranju vzorca stroški opazovanja. Zato skušamo dalje iskati optimalno razmestitev enot glede na skupne stroške opazovanja. Ta razmestitev je enaka pri optimalni razmestitvi glede na število enot, če so stroški za opazovanje ene enote v vseh stratumih enaki. Če pa so stroški za opazovanje vzorčnih enot različni, more optimalna razmestitev glede na stroške bistveno odstopati od optimalne razmestitve glede na število enot. Glede na stroške je optimalno tisto število enot v posameznih stratumih, ki da pri istih stroških najmanjšo skupno varianco in s tem najzanesljivejšo oceno.

Vzemimo enostavno funkcijo stroškov, ki upošteva samo konstantne stroške in stroške, ki so v zvezi z opazovanjem posamezne enote vzorčenja. Konstantni stroški so oni, ki niso odvisni od tega, kakšno število enot vzamemo v vzorec. To so stroški za pripravo ankete, za sestavo okvirja itd. V variabilne stroške pa vzemimo stroške, ki so v zvezi z opazovanjem posamezne enote. Pri enostavni funkciji stroškov predpostavljamo, da so stroški po posameznih stratumih proporcionalni številu enot v vzorcih po stratumih. V tem primeru so skupni stroški dani s funkcijo

$$C = C_0 + C_v = C_0 + \sum c_k n_k \quad (10.44)$$

Pri tem pomeni: C = skupni stroški, C_0 = konstantni stroški, C_v = variabilni stroški, c_k = stroški za opazovanje ene enote v stratumu; n_k = število enot v stratumu k .

Pri optimalni razmestitvi enot glede na stroške določimo število enot po posameznih stratumih po obrazcu

$$n_k = a \frac{N_k S_k}{\sqrt{c_k}} \quad ; \quad a = \frac{C_v}{\sum N_k S_k \sqrt{c_k}} \quad (10.45)$$

10.42

Za naš primer vzemimo, da so stroški za opazovanje ene osnovne parcele (potni stroški, stroški merjenja itd.) po stratumih naslednji:

$$c_1 = 3600 \text{ din}, \quad c_2 = 2500 \text{ din}, \quad c_3 = 1600 \text{ din}.$$

Stroški so različni zaradi različnega terena, velikosti teritorije sestoji itd. Če upoštevamo zgornje stroške na enoto po posameznih stratumih, znašajo spremenljivi stroški glede na število enot po stratumih pri optimalni razmestitvi iz tabele 10.11:

$$c_1 n_1 + c_2 n_2 + c_3 n_3 = 3600 \cdot 12 + 2500 \cdot 33 + 1600 \cdot 27 = 168900 \text{ din.}$$

Po nakazanem postopku določimo optimalno razmestitev enot po stratumih tako, da pri enakih spremenljivih stroških 168900 din dobimo čim manjšo varianco. Postopek izračuna optimalne razmestitve enot po stratumih glede na stroške je nakazan v tabeli 10.12.

Tabela 10.12 Določitev optimalne razmestitve glede na stroške po stratumih za enote

Stratum	N_k	S_k	c_k	$N_k S_k \sqrt{c_k}$	$\frac{N_k S_k}{\sqrt{c_k}}$	n_k	S_k^2	$\text{Var}(Y_{k,sl})$
1	144	1,80	3600	15552	4,32	10	3,2552	6300
2	192	3,82	2500	36672	14,67	32	14,5720	13989
3	240	2,54	1600	24384	15,24	33	6,4508	9712
	576			76608		75		30001
	N			$\sum N_k S_k \sqrt{c_k}$		n		$\text{Var}(Y_{str,s})$

Konstanta a , s katero pomnožimo izraze $N_k S_k / \sqrt{c_k}$, da dobimo število enot v posameznih stratumih, je

$$a = C_v / \sum N_k S_k \sqrt{c_k} = 168900 / 76608 = 2,2047$$

Iz variance za oceno agregata s stratificiranim vzorčenjem pri novi razmestitvi spoznamo, da smo za isti denar kot pri optimalni razmestitvi glede na število enot, dobili večji vzorec in boljše rezultate. Primerjava varianc za obe razmestitvi

$$\text{Var}(Y_{str,o}) / \text{Var}(Y_{str,c}) = 30850 / 30001 = 1,028$$

namreč pokaže, da bi morali v stratificiran vzorec pri optimalni razmestitvi ne glede na stroške vzeti za 2,8% več sredstev, če bi hoteli z njim dobiti enako kvalitetno oceno kot s strati-

ficiranim vzorčenjem, pri katerem upoštevamo razlike v stroških opazovanja.

10.43

Primer za ocenjevanje s stratificiranim vzorčenjem.

Omenili smo že, da pri stratificiranem vzorčenju v vsakem stratumu izvedemo samostojen, od drugih stratumov neodvisen vzorec, da ocenimo iskani parameter za vsak stratum posebej, iz ocen za posamezne stratumne pa sestavimo oceno za populacijo.

Prva naloga pri izvedbi stratificiranega vzorčenja je izdelava okvira vzorčenja. Pri stratificiranem vzorčenju je okvir sestavljen iz samostojnih seznamov enot po stratumih. V seznamu enot po posameznih stratumih ima vsak stratum svojo posebno numeracijo, ki omogoča, da pri izboru identificiramo na slučajnostni način izbrane enote. Tudi pri stratificiranem vzorčenju more biti osnova okvirja seznam, kartoteka, geografska karta, skice in podobno, le da so okvirji sestavljeni po stratumih. Vsak stratum smatramo za samostojno populacijo s samostojnim okvirjem in numeracijo.

V sliki 10.5 je nakazan okvir vzorčenja za poskusni sestoj. Ploskev je glede na kvaliteto tal razdeljena v tri stratumne: 1, 2 in 3. Vsak stratum ima v okvirju samostojno numeracijo parcelic.

Okvir vzorčenja je v našem primeru shema ploskve z vpisanimi zaporednimi številkami parcelic v posameznem stratumu.

Zaporedne številke izbranih parcel dobimo iz tablic slučajnostnih števil v dodatku. Ker je skupno število osnovnih parcelic v prvem stratumu $N_1 = 144$, v drugem $N_2 = 192$ in v tretjem $N_3 = 240$, iz tablic slučajnostnih števil dobimo zaporedne številke izbranih parcelic po naslednjem pravilu: Od začetka tablic slučajnostnih števil sestavljamo drseče trimestne slučajnostne številke tako, da tromestno skupino pomikamo za eno slučajnostno številko. Ker je število enot v prvem stratumu pod 200, dobimo iz tromestne slučajnostne številke, ki je večja kot 200, zaporedno številko izbrane parcele tako, da od nje odštevamo ustrezen mnogokratnik od 200. Enako postopamo v drugem stratumu, ker tudi v njem število enot ni večje kot 200. V tretjem stratumu je število enot $N_3 = 240$. Zanj od tromestnih slučajnostnih števil odštevamo po znanem postopku 300. Vendar

slučajnostne številke, ki so večje kot 900, v tretjem stratumu ne pridejo v poštev, ker bi sicer ne imela vsaka enota v stratumu enake možnosti, da jo izberemo in zato ocena ne bi bila nepristranska. V tabeli 10.13 je prikazan izbor vzorcev v posameznih stratumi. Poleg slučajnostne številke je v drugem stolpcu vpisana zaporedna številka parcele, ki ustreza posamezni slučajnostni številki. Iz tabele je razvidno, da nekatere slučajnostne številke ne pridejo v poštev, ker so večje kot je skupno število enot po stratumi.

V prvem stratumu dobimo na primer zaporedno številko izbrane parcele, če ostanek tromestne slučajnostne številke, ki smo jo delili z 200 ni večji kot $N_1=144$. Po tem načelu ustreza prvi tromestni slučajnostni številki 034 parcela z zaporedno številko 34, slučajnostna številka 347 ne pride v poštev, ker je ostanek $347-200=147$ večji kot $N_1=144$. Slučajnostni številki 474 ustreza parcela z zaporedno številko $474-400=74$ itd. Podobno postopamo pri drugem in pri tretjem stratumu.

V tretjem stolpcu je vpisan volumen na izbranih parcelah. Te volumne dobimo z merjenjem na izbranih parcelah, oziroma iz slike 10.3. V četrtem stolpcu je z o naznačeno, če je volumen na izbrani parceli manjši ali enak 30 m^3 , z 1 pa so označene parcele, na katerih je volumen večji kot 30 m^3 .

Izbrane parcele z vpisanimi osnovnimi podatki o volumnih so vrisane v okvirju v sliki 10.5.

Da ocenimo skupen volumen in standardno pogreško oziroma mejo zaupanja, najprej izračunamo iz osnovnih podatkov o vzorcih za vsak stratum posebej pomožne količine y_k in s_{yk}^2 ,

iz teh količin pa dalje po obrazcih iz pregleda obrazcev v odstavku 10.19,

ocene agregata - volumna $Y_{k,sl}$ in ocene varianc $\text{var}(Y_{k,sl})$ za posamezen stratum. Vsota ocen volumnov po stratumi je stratificirana ocena volumna na celi ploskvi, vsota ocen varianc pa ocena variance volumna s stratificiranim vzorcem. Analogno za oceno parcel z volumnom nad 30 m^3 najprej preštejemo za posamez-

Tabela 10.13 Osnovni podatki o stratificiranem vzorcu na eksperimentalni ploskvi

Stratum 1				Stratum 2				Stratum 3			
sl. št.	z. št.	y_i	$y_i < 30$	sl. št.	z. št.	y_i	$y_i < 30$	sl. št.	z. št.	y_i	$y_i < 30$
034	34	20	0	614	14	33	1	742	142	37	1
347	-	-	-	146	146	33	1	424	124	34	1
474	74	21	0	469	69	35	1	246	-	-	-
743	143	24	0	698	98	27	0	467	167	36	1
437	37	24	0	986	186	37	1	676	76	33	1
373	-	-	-	863	63	31	1	762	162	35	1
738	138	21	0	637	37	39	1	624	24	35	1
386	-	-	-	371	171	35	1	242	-	-	-
863	63	21	0	716	116	25	0	428	128	36	1
636	36	23	0	162	162	30	0	281	-	-	-
369	-	-	-	623	23	32	1	811	211	37	1
696	96	22	0	233	33	27	0	114	114	36	1
964	-	-	-	332	132	25	0	145	145	34	1
647	47	22	0	326	126	33	1	457	157	40	1
473	73	23	0	261	61	31	1	572	-	-	-
736	136	22	0	616	16	21	0	720	120	36	1
366	-	-	-	168	168	31	1	204	204	38	1
661	61	24	0	680	80	31	1	042	42	35	1
12 267 0				804	4	31	1	425	125	36	1
$n_1 \quad y_1 \quad h_1$				045	45	32	1	253	-	-	-
				456	56	32	1	533	233	34	1
				560	160	24	0	332	32	34	1
				601	1	29	0	323	23	35	1
				011	11	36	1	237	237	34	1
				111	111	25	0	373	73	37	1
				114	114	30	0	732	132	40	1
				141	141	22	0	322	22	33	1
				410	10	29	0	227	227	34	1
				109	109	27	0	270	-	-	-
				095	95	30	0	707	107	37	1
				959	159	28	0	073	73	dv	dv
				597	-	-	-	736	136	33	1
				977	177	29	0	360	60	36	1
				774	174	31	1	607	7	36	1
				33 991 17				27 961 27			
				$n_2 \quad y_2 \quad h_2$				$n_3 \quad y_3 \quad h_3$			

1. stratum

2. stratum

3. stratum

001	002	003	004	001 0 29	002	003	004 1 31	005	006	007	008	009	010 0 29	011 1 36	012	013	001	002	003	004	005	006	007 1 36
005	006	007	008	014 1 33	015	016 0 21	017	018	019	020	021	022	023 1 32	024	025	026	008	009	010	011	012	013	014
009	010	011	012	027	028	029	030	031	032	033 0 27	034	035	036	037 1 39	038	039	015	016	017	018	019	020	021
013	014	015	016	017	040	041	042	043	044	045 1 32	046	047	048	049	050	051	022 1 33	023 1 35	024 1 35	025	026	027	028
018	019	020	021	022	052	053	054	055	056 1 32	057	058	059	060	061 1 31	062	029	030	031	032 1 34	033	034	035	036
023	024	025	026	027	063 1 31	064	065	066	067	068	069 1 39	070	071	072	073	037	038	039	040	041	042 1 35	043	044
028	029	030	031	032	074	075	076	077	078	079	080 1 31	081	082	083	045	046	047	048	049	050	051	052	053
033	034 0 20	035	036 0 23	037 0 24	084	085	086	087	088	089	090 0 27	091	054	055	056	057	058	059	060 1 36	061	062	063	064
038	039	040	041	042	092	093	094	095 0 30	096	097	098 0 27	099	065	066	067	068	069	070	071	072	073 1 37	074	075
043	044	045	046	047 0 22	100	101	102	103	104	105	106	107 1 33	076	077	078	079	080	081	082	083	084	085	086
048	049	050	051	052	108	109 0 27	110 0 25	111	112	113	114 0 30	087	088	089	090	091	092	093	094	095	096	097	098
053	054	055	056	057	058	115 0 25	116	117	118	119	120	099	100	101	102	103	104	105	106 1 37	107	108	109	110
059	060	061 0 24	062 0 21	063	064	065	121	122	123	124	125	111	112	113	114 1 36	115	116	117	118	119 1 36	120	121	122
066	067	068	069	070	071	072	126 1 33	127	128	129	130	123	124 1 34	125 1 36	126	127	128 1 36	129	130	131	132 1 40	133	134
073	074 0 23	075 0 21	076	077	078	079	131	132 0 25	133	134	135	136 1 33	137	138	139	140	141	142 1 37	143	144	145 1 34	146	147
080	081	082	083	084	085	086	135	136	137	138	148	149	150	151	152	153	154	155	156 1 40	157	158	159	160
087	088	089	090	091	092	139	140	141 0 22	142	143	144	161	162 1 35	163	164	165	166	167 1 36	168	169	170	171	172
093	094	095 0 22	096	097	098	145	146 1 33	147	148	149	150	173	174	175	176	177	178	179	180	181	182	183	184
099	100	101	102	103	104	151	152	153	154	155	156	157	185	186	187	188	189	190	191	192	193	194	195
105	106	107	108	109	110	158	159 0 28	160 0 24	161	162 0 30	163	164	196	197	198	199	200	201	202	203	204 1 38	205	206
111	112	113	114	115	116	117	165	166	167	168 1 31	169	170	199	192 1 35	207	208	209	210	211 1 37	212	213	214	215
118	119	120	121	122	123	124	125	173 1 31	174	175	176	177	178 0 29	179	216	217	218	219	220	221	222	223	224
126	127	128	129	130	131	132	133	134	180	181	182	183	184	185	186 1 37	225	226	227 1 34	228	229	230	231	232
135	136 0 22	137	138 0 21	139	140	141	142	143 0 24	144	187	188	189	190	191	192 1 34	233	234	235	236	237 1 34	238	239	240

Slika 10.5 Okvir za stratificirano vzorčenje z vrisanimi slučajnostno izbranimi parcelami po stratumih pri optimalni razmestitvi.

ne stratume število parcel z nad 30 m^3 volumna h_k in izračunamo $s_{p,k}^2$. Iz teh količin pa dalje po obrazcih iz pregleda 10.19 izračunamo oceno števila enot z nad 30 m^3 $H_{k,sl}$ in ocene varian-
ce tega števila $\text{var}(H_{k,sl})$ za posamezen stratum. Enako kot pri agregatu je vsota ocen po stratumih enaka stratificirani oceni za celotno ploskev. Pomožne količine in osnovni izračun strati-
ficiranih ocen je nakazan v tabeli 10.14.

10.14 Ocenitev lesne zaloge in števila parcelic z lesno zalo-
go nad 30 m^3 na poskusni ploskvi s stratificiranim vzorčenjem

Stratum	N_k	n_k	y_k	$s_{y,k}^2$	$Y_{k,sl}$	$\text{var}(Y_{k,sl})$	h_k	$s_{p,k}^2$	$H_{k,sl}$	$\text{var}(H_{k,sl})$
1	144	12	267	1,84	3204	2915	0	0	0	0
2	192	33	91	17,22	5766	15930	17	0,2576	99	238,27
3	240	27	961	3,48	8542	6589	27	0	240	0
	576	72			17512	25434			339	238,27
	N	n			Y_{str}	$\text{var}(Y_{str})$			H_{str}	$\text{var}(H_{str})$

Iz varianc ocen izračunamo standardne pogreške in meje zaupa-
nja:

$$se(Y_{str}) = \sqrt{\text{Var}(Y_{str})} = \sqrt{25434} = 159,5 \text{ m}^3$$

$$d(Y_{str}) = 2,00 \cdot se(Y_{str}) = 2,00 \cdot 159,5 = 319 \text{ m}^3$$

Prava lesna zaloge s tveganjem $\alpha = 0,05$ ni manjša kot

$$Y_{min} = Y_{str} - d(Y_{str}) = 17512 - 319 = 17193 \text{ m}^3 \text{ in ne večja kot}$$

$$Y_{max} = Y_{str} + d(Y_{str}) = 17512 + 319 = 17831 \text{ m}^3.$$

Če primerjamo dobljeni rezultat s pravo lesno zalogo $Y = 17439 \text{ m}^3$, vidimo, da ta vrednost leži v mejah zaupanja. Razlika ocene od prave vrednosti $Y_{str} - Y = 17512 - 17439 = +73 \text{ m}^3$ je komaj 0,43% od skupnega volumna. Dobljena ocena je znatno boljša, kot pa smo jo dobili z enostavnim vzorcem.

Za število parcel z lesno zalogo nad 30 m^3 pa je:

$$se(H_{str}) = \sqrt{\text{var}(H_{str})} = \sqrt{238,27} = 15,44$$

$$d(H_{str}) = 2,00 \cdot se(H_{str}) = 2,00 \cdot 15,44 = 31$$

Pravo število parcel z nad 30 m^3 je s tveganjem $\alpha = 0,05$ večje kot $H_{min} = H_{str} - d(H_{str}) = 339 - 31 = 308$

in manjše kot $H_{max} = H_{str} + d(H_{str}) = 339 + 31 = 370$

Ker je stratifikacija v našem primeru zelo uspešna, se ocena $H_{str} = 339$ od pravega števila parcel z nad 30 m^3 $H = 344$ razlikuje samo za 5.

Vzorčenje v skupinicah

10.44

Hiba enostavnega vzorčenja je med drugim tudi v tem, da so enote vzorčenja po vsej ploskvi razmeščene tako, da so v posameznih primerih stroški za pregled ene same enote znatni, ker v okolici te enote ni drugih enot iz vzorca. Čeprav je to dejstvo za kvaliteto ocene ugodno, ker je vzorec bolj reprezentativen, če so enote vzorca razmeščene po vsej ploskvi, je tak vzorec sorazmerno drag. Te hibe ne odpravi niti stratificirano vzorčenje, pri katerem ta problem ostane, čeprav v okviru stratumov. Zato v določenih primerih težimo za tem, da so sosedne osnovne enote vzorca združene v skupinice. S slučajnim vzorcem v tem primeru ne izbiramo posamezne osnovne enote vzorca, temveč celotne skupinice.

Kvaliteta ocene za populacijo je predvsem odvisna od tega, kako so osnovne enote združene v skupinice, oziroma kakšne so lastnosti skupinic.

10.45

Kljub temu, da so skupinice sestavljene iz več osnovnih enot, skupinice smatramo kot samostojne enote vzorčenja. Zato je osnovno vzorčenje skupinic zelo sorodno enostavnemu vzorčenju, le da so namesto osnovnih enot enote vzorčenja skupinice.

Če z Y_{ki} zaznamujemo vrednost osnovne enote i v skupinici k , je vrednost agregata v skupinici k vsota vseh vrednosti osnovnih enot

$$Y_k = \sum_{i=1}^{N_k} Y_{ki} \quad (10.46)$$

Če je v populaciji M skupinic in s slučajnostnim vzorcem izberemo m enot, je po znanih stavkih za ocenjevanje agregata z enostavnim vzorčenjem ocena agregata z vzorčenjem v skupinicah

$$Y_{sk} = \frac{M}{m} \sum_{k=1}^m Y_k \quad (10.47)$$

varianca za oceno agregata z vzorčenjem v skupinicah pa je

$$\text{Var}(Y_{sk}) = M(M-m) \frac{S_Y^2}{m} ; S_Y^2 = \frac{\sum_{k=1}^M (Y_k - \bar{Y})^2}{M-1} ; \bar{Y} = \frac{1}{M} \sum_{k=1}^M Y_k \quad (10.48)$$

Ocena variance pa je analogno

$$\text{var } Y_{sk} = M(M-m) \frac{s_Y^2}{m} ; s_Y^2 = \frac{\sum_{k=1}^m (y_k - \bar{y})^2}{m-1} ; \bar{y} = \frac{1}{m} \sum_{k=1}^m y_k \quad (10.49)$$

Obrazci so analogni obrazcem za enostavno vzorčenje, le da se nanašajo na podatke o skupinicah. Analogni so tudi drugi obrazci za ocene varianc agregata, za ocene aritmetične sredine itd.

Iz obrazca 10.48 za izračun variance agregata z vzorčenjem v skupinicah sklepamo, da je ocena agregata z vzorčenjem v skupinicah tem uspešnejša, čim manjša je varianca med vrednostmi Y_k za posamezne skupinice. Iz tega spoznamo, da ocena z vzorčenjem v skupinicah ni odvisna od variabilnosti osnovnih podatkov v skupinicah, ampak samo od variabilnosti podatkov med skupinicami. Ta lastnost je ravno obratna kot pri stratificiranem vzorčenju, pri katerem je zanesljivost ocene odvisna samo od variabilnosti znotraj skupin-stratumov. Zato pri stratificiranem vzorčenju težimo za tem, da so razlike med stratumi čim večje, variabilnost znotraj stratumov pa čim manjša. Pri vzorčenju v skupinicah pa je obratno ocena tem boljša, čim manjša je variabilnost med skupinicami.

10.46

Vrste skupinic. Osnovne enote združujemo v skupinice po različnih načelih. Kot sledi iz prejšnjega odstavka pa pri sestavljanju skupinic težimo za tem, da združujemo enote tako,

da so razlike med skupinicami čim manjše.

Če vzamemo, da je v gozdarstvu osnovna enota drevo, združujemo drevesa v skupinice na različne načine. Skupinice morejo biti vsa drevesa na iregularnih površinah, ki so z naravnimi mejami (poti, potoki) ali kako drugače (meje parcel) lahko določljive. Pri tem težimo za tem, da tako sestavljene skupinice niso med seboj preveč različne niti po obsegu niti po številu dreves.

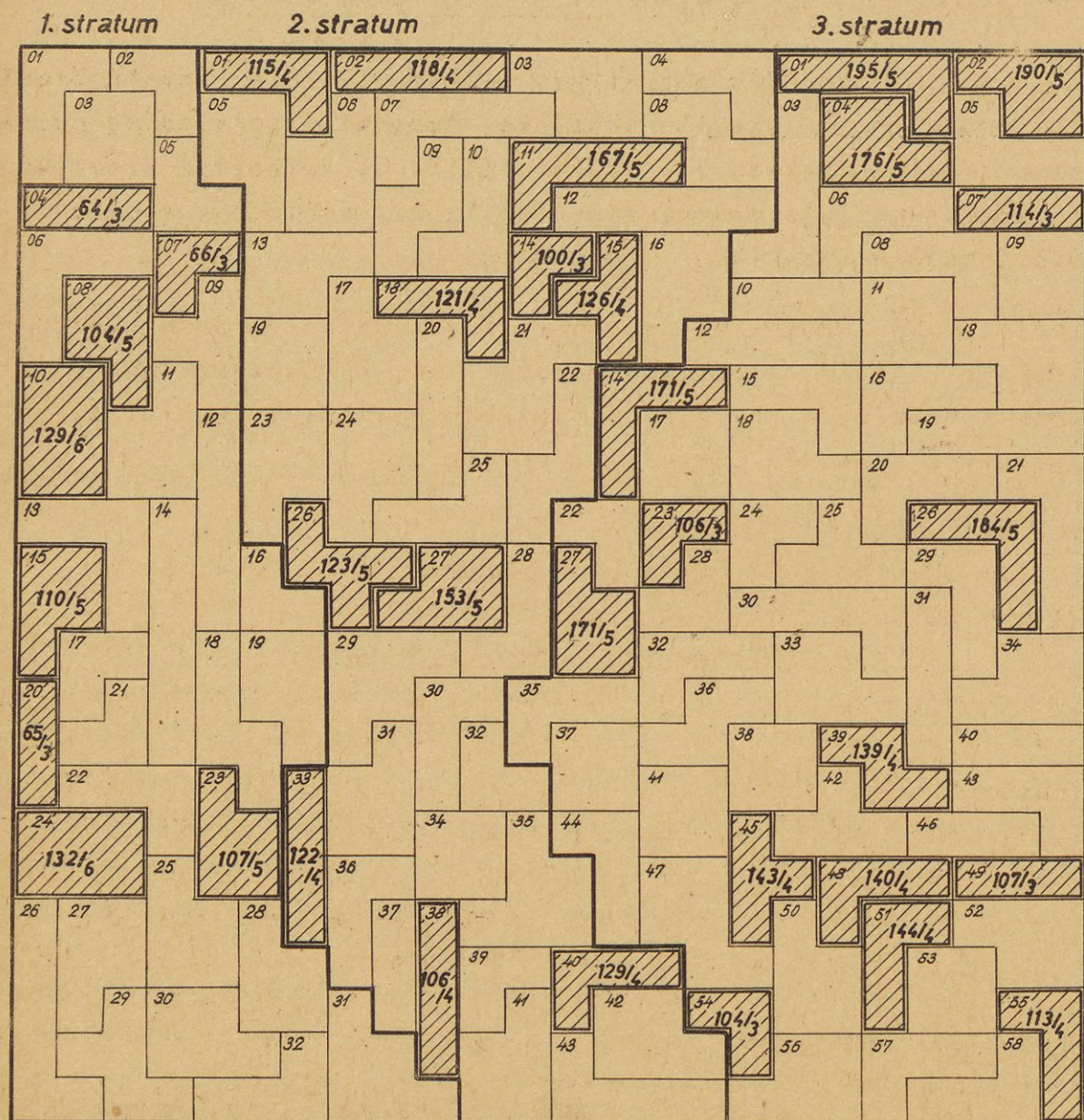
Pogoste pa so v gozdarstvu skupinice regularnih površin. Tako sestavljajo skupinice krogi z določenim radijem, kvadrati z dano stranico, pravokotniki, pasovi oziroma proge.

Posebna oblika vzorčenja v skupinicah je sistematično vzorčenje. Sistematični vzorci so po svoji naravi taki, da je v posebnih pogojih variabilnost med skupinicami tako majhna, da moremo eno samo skupinico vzeti za osnovo pri ocenjevanju parametrov populacije.

10.47

Primer vzorčenja skupinic iregularnih površin. Na poskusni ploskvi smo združili osnovne enote (kvadratne površine po 10 a) po naravnem kriteriju v iregularne površnice, ki so po obsegu različne. Vsekakor moremo kombinirati vzorčenje v skupinicah s stratifikacijo. V sliki 10.6 imamo sestavljen okvir za stratificirano vzorčenje skupinic na poskusni ploskvi. V stratumu 1 je $M_1 = 32$ skupinic, v stratumu 2 je $M_2 = 43$ skupinic, v stratumu 3 pa $M_3 = 58$ skupinic.

Ker je število skupinic v posameznih stratumih samo dvomestno število, iz tablic slučajnostnih števil sestavimo samo dvomestne slučajnostne številke. Ker je $M_1 = 32$, tablice racionalneje izkoristimo, če slučajnostne številke do 32 smatramo kot zaporedne številke izbranih parcel. Iz slučajnostnih števil od 41 do 72 pa dobimo zaporedne številke izbranih parcel, če odštejemo od njih 40. Podobno v stratumu 2 slučajnostne številke od 1 do 43 ustrezajo zaporednih številkam izbranih parcel, iz slučajnostnih števil od 51 do 93 pa dobimo zaporedne številke, če od njih odštejemo 50. Za stratum 3 pa izkoristimo kot zaporedne številke izbranih skupinic slučajnostne številke, ki so manjše kot $M_3 = 58$.



Slika 10.6 Okvir za stratificirano vzorčenje v skupinicah z vrisanimi slučajnostno izbranimi skupinicami po stratumih.

V vzorec vključimo 25% skupinic in enote vzorca razme-
stimo med stratumi proporcionalno številu skupinic. Tako dobimo,
da je število skupinic v vzorcih po stratumih:

$$m_1 = 8, \quad m_2 = 11 \quad \text{in} \quad m_3 = 15.$$

Po zgornjih pravilih s tablicami slučajnostnih števil
izberemo ustrezno število skupinic. Podatki izbora so za posa-
mezne stratumе nakazani v tabeli 10.15. Slučajnostne številke
so dvomestne slučajnostne številke iz 21. stolpca v tablici
slučajnostnih števil.

Tabela 10.15 Podatki o izboru enot v stratificiranem vzorcu
skupinic. (Sl.št.=slučajnostna številka, Z.št.=zaporedna šte-
vilka skupinice v okvirju, y_i =volumen v izbrani skupinici,
 x_i =število osnovnih parcel v skupinici.)

Stratum 1

Sl.št.	60	24	23	44	08	55	10	07
Z.št.	20	24	23	4	8	15	10	7
y_i	65	132	107	64	104	110	129	66
x_i	3	6	5	3	5	5	6	3

Stratum 2

Sl.št.	64	38	68	15	11	40	02	38	02	27	11
Z.št.	14	38	18	15	11	40	2	38	2	27	11
y_i	100	106	121	126	167	129	118	dv	dv	153	dv
x_i	3	4	4	4	5	4	4			5	

Sl.št.	51	26	83
Z.št.	1	26	33
y_i	115	123	122
x_i	4	5	4

Stratum 3

Sl.št.	26	39	02	23	27	7	54	4	48	02	45
Z.št.	26	39	2	23	27	7	54	4	48	2	45
y_i	184	139	190	106	171	114	104	176	140	dv	143
x_i	5	4	5	3	5	3	3	5	4		4

Sl.št.	51	55	14	54	49
Z.št.	51	55	14	54	49
y_i	144	143	171	dv	107
x_i	4	4	5		3

Iz podatkov v tabeli 10.15 ocenimo volumen in standardno pogreško oziroma meje zaupanja za oceno volumna enako kot za enostaven stratificiran vzorec, le da se podatki nanašajo na skupinice.

10.16 Izračun variance za stratificirano vzorčenje skupinic na poskusni ploskvi

Stratum	M_k	m_k	y_k	$Y_{k,sl}$	$s_{y.k}^2$	$var(Y_{k,sl})$
1	32	8	777	3108	805,84	77361,64
2	43	11	1380	5395	372,67	46618,62
3	58	14	2032	8418	887,21	161726,71
	153	53		16921		285705,97
	M	m		$Y_{sk,str}$		$Var(Y_{sk,str})$

Iz osnovnih podatkov po stratumih smo izračunali ocene volumnov $Y_{k,sl}$ za posamezne stratumne po obrazcu 10.47, ocene varianc za ocene agregatov pa po obrazcu 10.49. Vsota ocen za lesno zalogo po posameznih stratumih je stratificirana ocena za celoten sestoj, vsota varianc pa ocena variance za stratificirano oceno agregata.

Po zgornjih rezultatih sta meji zaupanja

$$se(Y_{sk,str}) = \sqrt{var(Y_{sk,str})} = \sqrt{285705} = 534,5 \text{ m}^3$$

odklon zaupanja s tveganjem $\alpha = 0,05$ pa je

$$d(Y_{sk,str}) = t(0,05;30) \cdot se(Y_{sk,str}) = 2,04 \cdot 534,5 = 1090 \text{ m}^3$$

Pravi volumen je torej s tveganjem $\alpha = 0,05$ v mejah $Y = 16921 \pm 1090$.

Vzorčenje v pasovih. Zelo cenjen način vzorčenja v skupinicah v gozdarstvu so pasovi ali proge. Prednost pasov pred skupinicami iregularnih oblik je v lažji tehnični opredelitvi proge in v lažjem sestavljanju okvirja. Vsebinsko pa so pasovi prikladnejši zato, ker je pri poznanem terenu možno položiti proge v taki smeri, da je variabilnost znotraj pasov velika, variabilnost med pasovi pa čim manjša. Če namreč za določen sestoj vemo, da se kvaliteta gozda spreminja v določeni smeri (n.pr. v smeri največjega padca terena), usmerimo proge v tej smeri in dobimo razdelitev populacije na enote z želenimi lastnostmi.

Proge se lahko raztezajo od meje do meje sestoja in so lahko različne dolžine, morejo pa biti tudi krajše in enako dolge.

Pri vzorčenju prog površino sestoja razdelimo po zgornjem načelu v proge, ki so enote vzorčenja. S slučajnostnim izborom izberemo ustrezno število prog, katere proučimo v celoti. Iz podatkov o izbranih progah pa ocenimo parametre in standardne pogreške za celoto po obrazcih za ocenjevanje z enostavnim vzorčenjem, pri čemer so posamezne proge vzorčne enote.

Primer za ocenjevanje s progami. Za poskusni sestoj proučimo možnost uporabe prog za ocenjevanje skupnega volumna v sestoju. Če proučimo skico poskusnega sestoja v sliki 10.3, opazimo, da se kvaliteta sestoja bolj spreminja v smeri od leve na desno, kot pa v smeri od zgoraj navzdol. Zato je primerneje, da usmerimo proge horizontalno kot pa vertikalno. Zaradi potrditve zgornjega zaključka proučimo, kakšne rezultate da vzorčenje v vertikalnih in kakšne v horizontalnih pasovih.

Vzemimo, da so proge široke toliko kot osnovne kvadratne parcele in da se raztezajo skozi vso dolžino oziroma širino sestoja. V vsako progo je torej vključenih 24 osnovnih parcelic. Zaradi možnosti primerjave kvalitet ocen, dobljenih s progami, s kvaliteto ocen po drugih metodah, vzemimo število osnovnih parcelic v vzorčenju v progah enako številu parcelic pri enostavnem vzorčenju! V vzorec vključimo 3 naključno izbrane pasove, ki imajo skupno $3 \times 24 = 72$ osnovnih parcelic.

V sliki 10.7 je narisana shema okvirja za vzorčenje v progah, če so proge orientirane navpično in shema okvirja, v katerem so proge orientirane vodoravno. Poleg oznake prog z zaporednimi številkami so v posameznih progah podatki o skupnem volumnu v posameznih pasovih. Ti podatki služijo za teoretično proučitev obeh sistemov vzorčenja.

Za navpične proge iz sheme 10.7 dobimo, da je povprečen kvadratičen odklon med progami $S_{p.n}^2 = 18630$. Ker je število prog v populaciji $M = 24$, število prog v vzorcu pa $m = 3$, je varianca za oceno volumna z vzorcem v navpičnih progah

$$\text{Var}(Y_{p.n}) = M(M-m) \frac{S_{p.n}^2}{m} = 24 \cdot (24-3) \frac{18630}{3} = 3129840$$

Nasprotno pa je vzorčenje v vodoravnih progah, ki so usmerjene v smeri intenzivne spremembe v kvaliteti sestoja, povprečen kvadratičen odklon med vodoravnimi progami $S_{p.v}^2 = 236,70$. Ker je tudi za vodoravne proge $M = 24$ in $m = 3$, je varianca za oceno lesne zaloge z vzorcem iz vodoravnih prog

$$\text{Var}(Y_{p.v}) = M \cdot (M-m) \frac{S_{p.v}^2}{m} = 24 \cdot (24-3) \frac{236,70}{3} = 39766$$

Primerjava varianc za obe oceni

$$\frac{\text{Var}(Y_{p.n})}{\text{Var}(Y_{p.v})} = \frac{3129840}{39766} = 78,7$$

pokaže, da je razlika v zanesljivosti ogromna. Varianca za oceno agregata s progami, usmerjenimi v smeri razlik v kvaliteti je skoraj devetinsedemdesetkrat manjša kot pri progah v obratni smeri.

Sistematično vzorčenje

10.50

Osnova. Čisti slučajnostni izbor s tablicami slučajnostnih števil je kljub določenim poenostavitvam še vseeno razmeroma okoren. Zato v gozdarski praksi pogosto uporabljamo sistematični izbor, ki je tehnično enostavnejši, more pa pod določenimi predpostavkami zamenjati čisti slučajnostni izbor.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
531	507	503	523	533	595	609	652	660	663	730	755	778	801	798	787	840	866	847	853	886	870	891	881

a) navpični pasovi

1	754
2	731
3	750
4	712
5	718
6	726
7	723
8	731
9	729
10	734
11	734
12	727
13	723
14	745
15	720
16	740
17	720
18	731
19	744
20	722
21	706
22	720
23	690
24	700

b) vodoravni pasovi

Slika 10.7 Shema okvirja vzorčenja v pasovih na poskusni ploskvi

Včasih pa da sistematični izbor celo boljše rezultate kot čist slučajnostni izbor.

Najenostavnejša oblika sistematičnega izbora je naslednja. Iz spiska vseh enot populacije izberemo v vzorec vsako s-to enoto, ustrezno s planirano velikostjo vzorca. s je stopinja vzorčenja in pove, vsako katero enoto po vrsti izberemo v sistematičen vzorec. Če izbiramo vzorec, za katerega je vzorčni delež 10%, vključimo v sistematični vzorec vsako deseto enoto iz spiska. Prvo enoto izberemo izmed prvih s enot s slučajnostnim vzorcem. S prvo izbrano enoto pa so vse druge enote sistematičnega vzorca dane avtomatično. Če s slučajnostnim izborom izberemo kot prvo enoto enoto z zaporedno številko 3, ima druga enota v 10% sistematičnem vzorcu zaporedno številko 13, tretja 23, četrta 33, peta 43 itd., dokler ne izčrpamo celotne populacije.

Stopinja s pri sistematičnem vzorčenju pove, vsako koliko enoto po vrsti moramo vključiti v sistematični vzorec. Stopinja pri sistematičnem vzorčenju je z vzorčnim deležem f, ki pove, koliki del od celotne populacije je vključen v vzorec, v enostavni zvezi. Stopinja sistematičnega vzorčenja s je recipročna vrednost vzorčnega deleža f. Obratno pa je vzorčni delež recipročna vrednost stopinje pri sistematičnem vzorčenju.

$$s = \frac{N}{n} ; s = \frac{1}{f} ; f = \frac{1}{s} \quad (10.50)$$

Če je vzorčni delež $f = 0,02$, pomeni, da v vzorec vključimo 2% enot populacije. Stopinja sistematičnega vzorčenja je za ta primer enaka $s = 1/0,02 = 50$. V sistematičen vzorec izberemo vsako petdeseto enoto.

Obratno je vzorčni delež za sistematične vzorce, pri katerih je stopinja vzorčenja n.pr. $s = 20$, enak $f = 1/20 = 0,05$.

Sistematično vzorčenje kot vzorec skupinic

10.51

Sistematično vzorčenje je po svoji osnovi vzorčenje v skupinicah. To razvidimo iz naslednjega sklepanja. Če je znana stopinja vzorčenja s, moremo k vsaki številki od 1 do s prirediti skupnost enot, ki imajo zaporedne številke, ki so za mno-

gokratnik od s večje kot prva številka. Tako v prvo skupino spadajo enote z zaporednimi številkami $1, s+1, 2s+1, 3s+1, 4s+1$ v drugo enoto z zaporednimi številkami $2, s+2, 2s+2, 3s+2, 4s+2$ v tretjo enoto z zaporednimi številkami $3, s+3, 2s+3, 3s+3, 4s+3$ in tako dalje do skupine s , v katero spadajo enote z zaporednimi številkami, $s, 2s, 3s, 4s, \dots$

Populacijo tako razdelimo po določenem pravilu v s skupinic.

Kot velja za vzorčenje v skupinicah na splošno, da je ocena tem boljša, čim manjša je variabilnost med skupinami, velja to tudi za sistematično vzorčenje. Iz tega izhaja niz zaključkov o kakovosti ocen s sistematičnim vzorčenjem. Predvsem je sistematično vzorčenje tehnično mnogo enostavnejše kot čisti slučajnostni vzorec in ga moremo zato izvajati neposredno na terenu. Razen tega pa že na prvi pogled izgleda, da da sistematični vzorec boljše rezultate kot čisti slučajnostni vzorec, ker je s tehniko sistematičnega izbora zagotovljeno, da so enote razporejene enakomerno po vsej populaciji, kar ni vselej primer pri čistem slučajnostnem vzorčenju.

Če je vrstni red enot v okvirju slučajnosten in ni v pojavljanju nikake zakonitosti, je sistematičen vzorec enakovreden čistemu slučajnostnemu vzorčenju, ki ima enako število enot.

Če vrstni red pojavljanja glede na proučevani znak ne moremo smatrati kot slučajnosten, ampak so enote razmeščene po naravnem vrstnem redu tako, da posamezni deli v spisku predstavljajo homogene celote, ima sistematični vzorec v sebi elemente stratificiranega vzorčenja s proporcionalno razmestitvijo. Zato dobimo z njim boljše ocene kot s čistim slučajnostnim vzorčenjem oziroma ocene, ki so ekvivalentne stratificiranemu vzorčenju. Sistematični vzorec v tem primeru predstavlja v nekem smislu samostatifikacijo, kar je vsekakor pozitiven element.

Sistematičen vzorec da v splošnem boljše ocene kot čisti slučajnostni vzorec tudi v primeru, če v okvirju vzorčenja zasledimo neko tendenco enakomernega spreminjanja pojava, ki ga proučujemo.

Sistematično vzorčenje pa daje obratno slabše ocene kot čisti slučajnostni vzorec, če še v spisku enot značilnost, ki jo proučujemo, periodično spreminja, s periodo, ki je mnogokratnik stopinje sistematičnega vzorčenja.

Sistematično vzorčenje na površinah. V prejšnjem odstavku smo navedli, kako izbiramo enote v sistematični vzorec, če je okvir vzorčenja spisek enot. Večkrat pa je v gozdarstvu okvir vzorčenja geografska karta oziroma sama ploskev ali sestoj na terenu. V teh primerih način sistematičnega izbora prilagodimo novim pogojem. Pri okvirju, ki je dan s spiskom enot, je stopinja za sistematično vzorčenje dana s tem, da povemo vsa kolika zaporedna enota v spisku je vključena v izbor. Stopinja pri sistematičnem vzorčenju na površinah pa pove, koliko je posamezna enota sistematičnega vzorca oddaljena druga od druge. Primer sistematičnega vzorčenja na površinah je naslednji sistem izbora. Vzemimo, da je vzorčna enota krog z radijem 5,64 m ali površine 1 a, stopinja sistematičnega izbora pa 50 m. Da sistematično prepredemo celotno proučevano ploskev s krogi, ploskev najprej prepredemo s sistemom vzporednih daljic, ki so oddaljene druga od druge po 50 m. Na 50 metrskem odseku prve daljice na slučajnostni način določimo središče prvega kroga. Središče drugega kroga leži na isti daljici oddaljeno od središča prvega kroga 50 m. Središče naslednjega kroga leži na isti daljici oddaljeno od središča drugega kroga 50 m itd. Ko pridemo na prvi daljici do meje ploskve in ne moremo več odmeriti 50 m, da bi izbrali središče naslednjega kroga, preidemo v obratni smeri na drugo daljico. Pri tem pa pri izboru upoštevamo ostanek v metrih iz prve daljice. Če je n.pr. središče zadnjega kroga na prvi daljici oddaljeno od meje 23 m, je središče prvega kroga na naslednji daljici od meje oddaljeno $50 - 23 = 27$ m. Postopek določanja drugih krogov v vzorcu ponavljamo po istem postopku vse dotlej, dokler ne preidemo v odsekih po 50 m vseh daljic. Da se izognemo vplivu mejnega področja na oceno, v gozdarstvu v praksi kroge pri sistematičnem vzorčenju določamo tako, da so od robov oziroma mej odmaknjeni vsaj 20-25 m.

Nakazani postopek izbora je razviden iz sheme v sliki 10.7. Sistematično vzorčenje na površinah ima pod enakimi pogoji kot sistematično vzorčenje iz spiska enot prednosti oziroma pomanjkljivosti pred čistim slučajnostnim vzorčenjem. Periodičnosti, ki so nevarne pri sistematičnem vzorčenju, se pri vzorčenju površin pojavljajo, kadar je na ploskvi več vzpored-

nih pasov spreminjajoče se kvalitete. Sistematično vzorčenje na površinah pa da boljše rezultate kot čisti slučajnostni vzorec, če je populacija nehomogena.

10.53.

Ocenjevanje parametrov s sistematičnim vzorčenjem.

S sistematičnim vzorčenjem ocenjujemo parametre z obrazci, ki so slični obrazcem za ocenjevanje z enostavnim slučajnostnim vzorčenjem. Agregat in aritmetično sredino ocenjujemo po obrazcih

$$Y_{sis} = \frac{N}{n} y = s y ; \bar{Y}_{sis} = y/n \quad (10.51)$$

strukturni delež P in število enot z dano značilnostjo H pa po obrazcih

$$P_{sis} = h/n ; H_{sis} = \frac{N}{n} h \quad (10.52)$$

pri čemer pomeni: N = število enot v populaciji; n = število enot v sistematičnem vzorcu; y = vsota podatkov v sistematičnem vzorcu; h = število enot z dano značilnostjo v sistematičnem vzorcu.

10.54

Ocenjevanje variance za ocene parametrov. Pri sistematičnem vzorčenju je populacija skupinic sestavljena iz s skupinic. Od teh je v danem sistematičnem vzorcu v vsakem primeru izbrana le ena sama. Varianca za oceno agregata za ta primer je po obrazcu 10.22 enaka

$$\text{Var}(Y_{sis}) = s(s-1) S_{y_{sis}}^2 \quad (10.53)$$

Pri tem je: $S_{y_{sis}}^2$ = povprečen kvadratični odklon med skupinami.

Ker pa je pri sistematičnem vzorčenju v vsakem primeru od s skupinic vključena v vzorec le ena, ne moremo iz nje oceniti povprečnega kvadratičnega odklona $S_{y_{sis}}^2$. To dejstvo je hiba sistematičnega vzorca.

Pod določenimi predpostavkami o vrstnem redu enot v populaciji pa moremo vseeno oceniti varianco oziroma standardno pogreško in meje zaupanja.

Če predpostavljamo, da je vrstni red enot glede na proučevano karakteristiko slučajnosten, smo nakazali, da je sistematičen vzorec po svojih kvalitetah enakovreden čistemu slučajnostnemu vzorčenju. V tem primeru moremo s sistematičnim izborom izbrane enote in podatke smatrati kot da so izbrani s slučajnostnim izborom. Za ta primer upravičeno ocenimo varianco ocene agregata po istem obrazcu kot za čisti slučajnostni vzorec po obrazcu

$$\text{var}(Y_{\text{sis}}) = N(N-n) \frac{s_y^2}{n} \quad (10.54)$$

pri čemer pomeni: N = število osnovnih enot v populaciji, n = število enot v sistematičnem vzorcu in s_y^2 = ocena variance za y iz osnovnih podatkov v sistematičnem vzorcu.

Izdelane so metode za ocenjevanje variance ocen tudi za primere, da razpored enot v okvirju ni slučajnosten. Ena od teh je dana v primeru v nadaljevanju, druge pa presegajo naš okvir.

10.55

Primer za sistematično vzorčenje. Vzemimo, da v našem poskusnem sestoji ocenjujemo lesno zalogo s sistematičnim vzorcem. Lesno zalogo ocenimo z $n = 64$ osnovnimi parcelicami, ki jih izberemo s sistematičnim vzorčenjem. Ker ploskev sestoji iz $N = 576$ osnovnih parcelic, dobimo, da s sistematičnim vzorcem opazujemo na $s = N/n = 576/64 = 9$ parcel po eno. To pomeni, da v vsaki ($\sqrt{9}=3$) tretji vrsti določimo lesno zalogo na vsaki tretji parcelici. Kot je razvidno iz sheme v sliki 10.7 pri prehodu iz ene vrste v drugo upoštevamo, da morata biti med vsako izbrano parcelo po dve parceli. Prva parcelica je izbrana s slučajnostnim vzorcem izmed $3 \times 3 = 9$ parcelic v skrajnem zgornjem levem delu ploskve. Vse druge parcele pa so s prvo parcelo določene s sistemom sistematičnega izbora. Oceno volumna dobimo po obrazcu 10.52 in je

$$Y_{\text{sis}} = \frac{N}{n} y = \frac{576}{64} 1921 = 17289 \text{ m}^3$$

Primerjava ocene s pravim volumnom v sestoji $Y = 17439$ pokaže, da je sistematično vzorčenje dalo zelo dober rezultat. Razlika je

Sistematično vzorčenje je v tem primeru enakovredno s stratificiranim vzorčenjem, ki ima n stratumov, v vsakem stratumu pa $N/n=s$ enot, izmed katerih v vsakem stratumu izberemo po eno enoto. Ker pa iz ene same enote ne moremo oceniti varianco, združimo po dva sosedna elementarna stratuma v nove stratumе z $N_k=2s$ enotami in smatramo, da smo iz vsakega vzeli v vzorec po dve enoti.

Ocena variance agregata v stratumu je dana z znanim obrazcem $\text{var}(Y_{s1}) = N(N-n) \frac{s^2}{n}$. V našem primeru je $N=2s$, $n=2$, s^2_{y1} izračunam iz dveh podatkov

$$s_y^2 = \frac{(x_1 - \frac{x_1+x_2}{2})^2 + (x_2 - \frac{x_1+x_2}{2})^2}{2-1} = \frac{(x_1 - x_2)^2}{2} = \frac{d^2}{2} \quad (10.55)$$

je polovica kvadrata difference med obema vrednostima v stratumu. Varianco ocene agregata s sistematičnim vzorcem ocenimo v tem primeru po obrazcu

$$\text{Var } Y_{sis} = \sum_k 2s(2s-2) \frac{d_k^2}{2 \cdot 2} = s(s-1) \sum_k d_k^2 \quad (10.56)$$

ker je ocena variance agregata s stratificiranim vzorčenjem enaka vsoti ocen varianc v posameznih stratumihi.

Uporabimo to metodo v našem primeru pri predpostavki, da ni v populaciji neke periodičnosti, ki bi onemogočila uporabo te metode.

Kombinirajmo prvi dve izbrani parceli z volumnoma 18 in 24, drugi dve z volumnoma 23 in 28 in tako dalje. Tako dobimo ustrezne razlike:

$$d_1=18-24=-6, \quad d_2=23-28=-5, \quad d_3=27-35=-8, \quad d_4=43-43=0 \\ d_5=34-34=0, \quad d_6=32-35=-3, \quad d_7=32-28=+4 \text{ itd. do zadnje razlike } d_{32}=34-35=-1$$

$$\sum d_k^2 = d_1^2 + d_2^2 + d_3^2 + \dots + d_{32}^2 = (-6)^2 + (-5)^2 + (-8)^2 + \dots \\ \dots + (-1)^2 = 453$$

Po obrazcu 10.56 dobimo končno

$$\text{var}(Y_{sis}) = 9(9-1) \cdot 453 = 32616$$

$$\text{se}(Y_{sis}) = \sqrt{\text{var } Y_{sis}} = \sqrt{32616} = 180,6 \text{ m}^3$$

odklon zaupanja pa:

$$d(Y_{sis}) = t(0,05;32) \text{ se}(Y_{sis}) = 2,04 \cdot 180,6 = 368$$

Ocena volumna je $Y = 17289 \pm 368$ s tveganjem $\alpha = 0,05$. To je v skladu z dejanskim stanjem, ker je ocena od prave vrednosti različna le za 150 m^3 .

Vzorčenje v dveh in več stopnjah

10.56

Osnova. Vzemimo, da z vzorčenjem v skupinicah ocenjujemo gozdno površino v privatnih gospodarstvih v FLRJ. V skupinice združimo privatna gospodarstva v istih občinah. Pri vzorčenju skupinic so vzorčne enote skupinice-občine. Če ocenjujemo skupno gozdno površino z vzorčenjem v skupinicah, izberemo enostaven slučajnostni vzorec skupinic-občin in v izbranih občinah popišemo vsa gospodarstva. Oceno za skupno gozdno površino dobimo po obrazcu 10.47 ($Y_{sk} = \frac{M}{m} \sum_{k=1}^m Y_k$). Pri tem pomeni Y_k skupno gozdno površino v posameznih izbranih občinah. Ta podatek dobimo tako, da seštejemo podatke o gozdni površini za vsa gospodarstva v posameznih izbranih občinah.

Ker pa je število gospodarstev po občinah veliko, se izkaže, da je koristno, če gozdno površino v posameznih izbranih občinah ocenimo s samostojnimi slučajnostnimi vzorci v posameznih izbranih občinah. Ocena za gozdno površino v izbrani občini k je po znanem obrazcu enaka $Y_{k,sl} = \frac{N_k}{n_k} \sum_{i=1}^{n_k} y_{ki}$. y_{ki} po-

meni gozdno površino za gospodarstvo i v občini k . Če prave vrednosti za gozdne površine v izbranih občinah v obrazcu za ocenjevanje z vzorčenjem skupinic zamenjamo z ocenami $Y_{k,sl}$, dobimo novo oceno.

Tak način vzorčenja in ocenjevanja imenujemo vzorčenje v dveh stopnjah. Pri zgornjem primeru vzorčenja v dveh stopnjah smo v prvi stopnji izbrali vzorec enot prve stopnje - občin, v drugi stopnji pa v izbranih enotah prve stopnje samostojne

vzorci enot druge stopnje - gospodarstev.

Razlika med vzorčenjem v skupinicah in vzorčenjem v dveh stopnjah je v tem, da pri vzorčenju v dveh stopnjah prave vrednosti agregatov v skupinicah zamenjamo z ocenami, ki jih dobimo z vzorci v drugi stopnji.

Če v posameznih gospodarstvih, ki jih izberemo z vzorčenjem v dveh stopnjah, ne premerimo vseh gozdnih parcel, temveč gozdne površine po posameznih izbranih gospodarstvih dalje ocenimo s samostojnimi slučajnostnimi vzorci parcel, dobimo vzorčenje v treh stopnjah. V tem primeru so enote vzorčenja v prvi stopnji občine, enote vzorčenja v drugi stopnji gospodarstva znotraj izbranih občin, enote tretje stopnje pa parcele v izbranih gospodarstvih.

10.57

Prednosti in pomanjkljivosti. Vzorčenje v več stopnjah ima osnovne kvalitete vzorčenja v skupinicah. Odpravi namreč dve osnovni hibi enostavnega slučajnostnega vzorčenja. Če izvajamo enostaven slučajnostni vzorec gospodarstev za vso Jugoslavijo, je okvir vzorčenja spisek vseh gospodarstev v Jugoslaviji. Sestavljanje takega okvira in izbor iz njega je obsežno delo. Enako je s terenskim delom pri izvedbi takega vzorčenja. Ker so posamezna gospodarstva pri enostavnem slučajnostnem vzorcu razmeščena več ali manj po vseh občinah v Jugoslaviji, je tehnična izvedba takega vzorca komplicirana in draga, ker je za anketarje zamudno obiskovanje posameznih, na slučajnostni način izbranih gospodarstev.

Tehnika vzorčenja pa se z vzorčenjem v dveh ali več stopnjah za velike populacije zelo poenostavi. Za zgornji primer je pri vzorčenju v dveh stopnjah okvir za izbor v prvi stopnji le spisek vseh občin v Jugoslaviji. Okvir za izbor v drugi stopnji pa so spiski gospodarstev le za one občine, ki so bile izbrane v prvi stopnji. Razen tega je bistveno olajšano tudi delo anketarjev. Pri vzorčenju v dveh stopnjah so gospodarstva, ki jih mora obiskati in popisati, združena le v občinah, ki smo jih izbrali v prvi stopnji. Anketar gre v posamezno občino, da pregleda več gospodarstev. Tako se terensko delo pri vzorčenju v dveh ali več stopnjah bistveno poenostavi in poceni.

Res je, da moramo pri vzorčenju v dveh ali več stopnjah za isto zanesljivost ocen pregledati več osnovnih enot kot pri enostavnem slučajnostnem vzorčenju. Vendar dobimo z vzorčenjem v dveh ali več stopnjah zaradi tehničnih prednosti pri enakih stroških zanesljivejše ocene kot z enostavnim slučajnostnim vzorčenjem.

10.58

Ocenjevanje agregata in varianca za oceno agregata z vzorčenjem v dveh stopnjah. Agregat ocenjujemo z vzorčenjem v dveh stopnjah po postopku, ki smo ga navedli v odstavku 10.56, tako da najprej ocenimo za vsako enoto prve stopnje agregat $Y_{k,sl}$, iz teh pa skupen agregat Y_{2st} .

$$Y_{2st} = \frac{M}{m} \sum_{k=1}^m Y_{k,sl} \quad ; \quad Y_{k,sl} = \frac{N_k}{n_k} \sum_{i=1}^{n_k} y_{ki} \quad (10.57)$$

Pri tem pomeni: Y_{2st} = ocena agregata z vzorčenjem v dveh stopnjah; M = število enot v prvi stopnji; m = število enot v vzorcu v prvi stopnji; $Y_{k,sl}$ = ocena agregata v enoti k iz prve stopnje z vzorcem v drugi stopnji; N_k = število enot druge stopnje v izbrani enoti k prve stopnje; n_k = število enot v vzorcu druge stopnje v enoti k prve stopnje; y_{ki} = osnovni podatek za enoto i druge stopnje v enoti k prve stopnje.

Varianca za oceno agregata z vzorčenjem v dveh stopnjah izvira iz variance vzorčenja v prvi in iz varianc vzorčenja v drugi stopnji. Izračunamo jo po obrazcu:

$$Var(Y_{2st}) = Var(Y_{sk}) + \frac{M}{m} \sum_{k=1}^m Var(Y_{k,sl}) \quad (10.58)$$

Pri tem pomeni: $Var(Y_{2st})$ = varianca za oceno agregata z vzorčenjem v dveh stopnjah; $Var(Y_{sk})$ = varianca za oceno agregata iz skupinic oziroma iz vzorca v prvi stopnji; $Var(Y_{k,sl})$ = varianca ocene agregata Y_k v enoti k iz prve stopnje s slučajnostnim vzorcem v drugi stopnji. Ti prispevki so izračunani po znanih obrazcih 10.48 in 10.22.

Oceno variance za oceno agregata z vzorčenjem v dveh stopnjah $var(Y_{2st})$ pa izračunamo po obrazcu

$$\text{var}(Y_{2st}) = \text{var}(Y_{sk}) + \frac{M^2}{m^2} \sum_{k=1}^m \text{var}(Y_{k,sl}) \quad (10.59)$$

ki je analogen obrazcu 10.58, le da so prave vrednosti zamenjane z ocenami.

Ocenjevanje po metodi razmerij

10.59

Osnova in ocena agregata. Zanesljivost ocene za določen podatek znatno povečamo z metodo razmerij, če poznamo za celotno populacijo vrednost agregata za kak znak, ki je s proučevanim podatkom v korelaciji. Če n.pr. poznamo, kolika je skupna temeljnica za celoten sestoj, zelo uspešno ocenimo skupno lesno zalogo tako, da slučajnostno izberemo vzorec dreves in zanje izmerimo volumen in temeljnico. Po metodi razmerij najprej za izbrana drevesa izračunamo, kolik volumen v povprečju odpade na enoto temeljnice. Ta koeficient je razmerje med skupnim volumenom in skupno temeljnico za vsa izbrana drevesa. Oceno skupne lesne zaloge v sestojiu pa dobimo, če to razmerje pomnožimo z znano skupno temeljnico v vsem sestojiu.

V splošnem ocenjujemo agregat po metodi razmerij po obrazcu

$$Y_{r,sl} = X \frac{y}{x} \quad (10.60)$$

pri čemer pomeni: $Y_{r,sl}$ = ocena agregata Y , izračunana iz enostavnega slučajnostnega vzorca po metodi razmerij; X = prava vrednost agregata, ki ga poznamo; $y = \sum_{i=1}^n y_i$ = vsota za znak y v enotah vzorca, $x = \sum_{i=1}^n x_i$ = vsota za znak x v enotah vzorca.

Čim tesnejša je povezava med proučevanim znakom in znakom, za katerega poznamo agregat za celoto, tem zanesljivejša je ocena.

Ocena agregata po metodi razmerij pa ni nepristranska temveč samo dosledna ocena pravega agregata. To pomeni, da aritmetična sredina vseh ocen agregatov iz vseh možnih vzorcev ni enaka pravi vrednosti agregata, pač pa se ji tem bolj približuje, čim večji je vzorec. Zato za velike vzorce to pristra-

nost zanemarimo in metodo razmerij s pridom uporabljamo pri ocenjevanju z velikimi vzorci.

10.60

Varianca ocene za agregat po metodi razmerij. Varianca za oceno agregata z metodo razmerij iz podatkov enostavnega slučajnostnega vzorca je približno

$$\text{Var}(Y_{r,sl}) = N(N-n) \frac{S_r^2}{n} \quad (10.61)$$

Pri tem je:

$$S_r^2 = \frac{\sum_{i=1}^N (Y_i - R X_i)^2}{N-1} ; \quad R = \frac{Y}{X}$$

Približek je tem boljši čim večji je vzorec.

Oceno približka za varianco $\text{var}(Y_{r,sl})$ pa izračunamo po obrazcu

$$\text{var}(Y_{r,sl}) = N(N-n) \frac{s_r^2}{n} \quad (10.62)$$

Pri tem je s_r^2 analogen izraz kot S_r^2 , samo da je izračunan iz podatkov iz vzorca.

$$s_r^2 = \frac{\sum_{i=1}^n (y_i - r x_i)^2}{n-1} = \frac{\sum y_i^2 - 2r \sum y_i x_i + r^2 \sum x_i^2}{n-1} \quad (10.63)$$

Razviti obrazec v nadaljevanju obrazca 10.63 je operativni obrazec in je prikladen za stvarno izračunavanje.

10.61

Primeri za uporabo metode razmerij v gozdarstvu. Ker običajno razpolagamo z različnimi podatki za populacijo, navedimo nekaj primerov iz gozdarstva, v katerih moremo uporabiti metodo razmerij.

Nakazali smo že, da po metodi razmerij ocenjujemo volumen, če poznamo skupno temeljnico.

Znatno izboljšamo oceno volumna tudi, če pri vzorčenju površin, pri čemer so skupinice vsa drevesa na določeni površini, poznamo površino sestoja ali skupno število dreves v sestoju.

Po metodi razmerij ocenjujemo različne podatke tudi, če razpolagamo s podatki o proučevanem pojavu v preteklosti. Če smo pred petimi leti s popolno premerbo dobili skupno temeljnico sestojaja, moremo po metodi razmerij oceniti, kolikšna je temeljnica danes.

Tudi ocene na oko korigiramo po metodi razmerij. Če n.p.r. na oko ocenimo volumen za vsa drevesa v sestoji in z enostavnim slučajnostnim vzorcem izberemo drevesa, za katera volumen določimo, moremo oceno lesne zaloge na oko popraviti po metodi razmerij.

Podobno po metodi razmerij popravljamo tudi rezultate popisov. Vzemimo, da popisujemo v nekem okraju kmetijska gospodarstva, ki imajo gozdno površino. Popis vrše popisovalci, ki se zadovolje z napovedmi gospodarjev kmetijskih gospodarstev. Če izmed vseh gospodarstev izberemo slučajnostni vzorec gospodarstev, v teh pa s podrobnim poizvedovanjem, izpiski iz katastra in drugimi dokumenti pridemo do natančnih podatkov, moremo po metodi razmerij popraviti podatke popisa.

10.62

Primer za ocenjevanje po metodi razmerij. Da prikažemo postopek ocenjevanja z metodo razmerij, vzemimo vzorčenje v skupinicah iz primera v odstavku 10.47. Volumen v posameznih skupinicah je vsekakor odvisen od velikosti skupinice. Ker poznamo skupno površino ploskve ($X=576$), moremo pri ocenjevanju volumna s skupinicami uporabiti metodo razmerij. Pri tem nas ne moti, da imamo vključeno v plan vzorčenja stratifikacijo. Za vsak stratum posebej po metodi razmerij ocenimo volumen in varianco. Vsota ocen po stratumih je stratificirana ocena volumna, vsota varianc po stratumih pa ocena variance za skupno oceno.

Ker imamo v tabeli 10.14, v kateri so vnešeni osnovni podatki o vzorčenju v skupinicah na naši poskusni ploskvi, že vse potrebne podatke o vzorcu, izračunamo stratificirano oceno volumna po metodi razmerij, kot je nakazano v tabeli 10.17.

Za posamezen stratum smo ocenili volumen po obrazcu 10.60. Stratificirano vzorčenje skupinic po metodi razmerij da

Tabela 10.17 Izračun ocene volumna po metodi razmerij s stratificiranim vzorčenjem skupinic

Stratum	x_k	y_k	x_k	$y_{k,r,sl}$
1	144	777	36	3108
2	192	1380	46	5760
3	240	2032	57	8556

$$\bar{x} = 576$$

$$17424 = Y_{r,str}$$

znatno boljši rezultat kot enostavna stratificirana ocena skupinic. Medtem ko po metodi skupinic dobimo za oceno volumna $Y_{sk,str} = 16921$, ki je od prave vrednosti različen za 518 m^3 , po metodi razmerij iz istega vzorca dobimo $Y_{r,str} = 17424 \text{ m}^3$ od prave vrednosti različen le 15 m^3 . To je delno slučajno, v veliki meri pa k izboljšanju ocene pripomore metoda razmerij. Izračun za oceno variance za stratificirano vzorčenje po metodi razmerij je nakazan v tabeli 10.18.

Tabela 10.18 Izračun variance za stratificirano vzorčenje skupinic po metodi razmerij

Stratum	N_k	n_k	y_k	x_k	$\sum y^2$	$\sum xy$	$\sum x^2$	r_k	$s_{k,r}^2$	$\text{var}(Y_{k,r,sl})$
1	32	8	777	36	81107	3738	174	21,583	115,17	11056
2	43	11	1380	46	176854	5863	196	30,000	147,40	18438
3	58	14	2032	57	306464	8589	241	35,649	277,40	50566

$$\text{var}(Y_{r,str}) = 80060$$

Prvih sedem stolpcev dobimo s preštevanjem, seštevanjem, kvadriranjem in množenjem osnovnih podatkov iz vzorca. r_k za posamezen stratum je $r_k = y_k/x_k$, $s_{k,r}^2$ za posamezen stratum izračunamo iz podatkov v prejšnjih stolpcih po obrazcu 10.63, $\text{var}(Y_{k,r})$ po obrazcu 10.62. Za prvi stratum je račun naslednji:

$$r_1 = y_1/x_1 = 777/36 = 21,583$$

po obrazcu 10.63 je

$$s_{1,r}^2 = \frac{1}{8-1} [81107 - 2 \cdot 21,583 \cdot 3738 + 21,583^2 \cdot 174] = 115,17$$

in po obrazcu 10.62 $\text{Var}(Y_{1,r,sl}) = 32(32-8) \frac{115,17}{8} = 11056$

Če primerjamo dobljeno oceno variance $\text{var}(Y_{r,str})$ z oceno variance za stratificirani vzorec skupin $\text{var}(Y_{sk,str})$ iz odstavka 10.47, dobimo

$$\frac{\text{Var}(Y_{sk,str})}{\text{Var}(Y_{r,str})} = \frac{285705}{80060} = 3,57. \text{ Metoda razmerij je v znatni meri}$$

pripomogla k izboljšavi zanesljivosti ocene.

Vzorčenje v dveh fazah

Vzorčenje v dveh fazah pri metodi razmerij

10.63

Pri ocenjevanju lesne zaloge v sestoji s pridom uporabljamo metodo razmerij, če poznamo za proučevani sestoj skupno temeljnico. Če pa skupne temeljnice ne poznamo, moremo volumen oceniti v dveh fazah po metodi, ki je metodi razmerij zelo podobna.

Analiza pokaže, da temeljnico določamo za posamezno drevo znatno lažje kot volumen. Zato skupno temeljnico za ves sestoj v prvi fazi vzorčenja ocenimo z velikim vzorcem. Tako dobimo za skupno temeljnico zanesljivo oceno $X_{1.sl}$. Ker pa sta volumen in temeljnica v tesni korelacijski odvisnosti, dobimo zadosti zanesljivo oceno za razmerje med volumenom in temeljnico $r_{2.sl}$ z majhnim vzorcem v drugi fazi vzorčenja v dveh fazah.

Iz dobljenih ocen za agregat X in razmerje R z vzorčenjem v dveh fazah sestavimo oceno za skupni volumen v sestoji po obrazcu

$$Y_{r,2f} = X_{1.sl} \cdot r_{2.sl} = \frac{N}{n_1} \cdot x_1 \cdot \frac{y_2}{x_2} \quad (10.64)$$

ki je v osnovi sličen obrazcu za ocenjevanje agregata po metodi razmerij.

V obrazcu 10.64 pomeni: $Y_{r,2f}$ = ocena agregata z vzorčenjem v dveh fazah; $X_{1.sl}$ = ocena agregata za dopolnilni znak x ; $r_{2.sl} = y_2/x_2$ = ocena razmerij $R = Y/X$ z vzorcem v drugi fazi.

x_1 = vsota podatkov za x v vzorcu v prvi fazi; x_2, y_2 = vsota podatkov za x in y v vzorcu v drugi fazi.

Zgornjo metodo vzorčenja v dveh fazah uporabljamo vselej, kadar ocenjujemo agregat za podatek, ki ga določamo razmeroma težko, obstaja pa podatek x , ki je z ocenjevanim znakom y v tesni korelacijski zvezi in ga določimo na lažji način.

Glede na to je nakazano vzorčenje v dveh fazah pri metodi razmerij priporočljivo pri kombiniranju ocen na oko s pravimi vrednostmi, ki jih dobimo z merjenjem in v drugih podobnih problemih.

Vzemimo na primer sestoj z $N = 5000$ drevesi. Z vzorcem $n = 500$ dreves smo z oceno na oko določili lesno zalogo $V_0 = 3000 \text{ m}^3$. Iz vzorca $n = 500$ dreves tega sestoja smo izbrali s slučajnostnim vzorcem $n = 50$ dreves, za katera smo določili volumne z izmero. Če vzamemo, da je lesna masa dreves v vzorcu po oceni na oko $v_0 = 36 \text{ m}^3$, s premerbo pa smo za ta drevesa dobili $v_1 = 32 \text{ m}^3$, je popravljena ocena po metodi razmerij $V_{r.2f} = V_0 \frac{v_1}{v_0} = 3000 \frac{32}{36} = 2670 \text{ m}^3$.

Vzorčenje v dveh fazah pri stratifikaciji

10.64

Z vzorčenjem v dveh fazah rešujemo tudi problem stratifikacije. Če razdelitev populacije na homogene dele ni vnaprej dana, je običajno obsežno delo, če hočemo enote populacije klasificirati v stratumе.

Za oceno agregata s stratificiranim vzorčenjem moramo po obrazcih 10.32 in 10.21

$$Y_{str} = \sum_{k=1}^r Y_{k,sl} = \sum_{k=1}^r N_k \bar{y}_k \quad (10.65)$$

poznati število enot v posameznih stratumih N_k in ocene za aritmetične sredine $\bar{y}_k$ za posamezne stratumе.

Če nimamo populacije razdeljene v stratumе, moremo število enot v posameznih stratumih oceniti z razmeroma velikim vzorcem z n enotami v prvi fazi. n enot iz vzorca v prvi fazi klasificiramo v stratumе. Tako dobimo $n_1, n_2 \dots n_r$ enot po posameznih stratumih. Če je od n enot iz vzorca v prvi stopnji, n_k enot v stratumu k , je ocena skupnega števila enot v stratumu k

$$N_{k,1.sl} = \frac{N}{n} n_k \quad (10.66)$$

Iz n_k enot, ki smo jih z vzorcem v prvi fazi dobili v stratumu k , izberemo manjši vzorec z n'_k enotami in zanj poiščemo vrednosti za znak y . Iz vzorca v drugi stopnji dobimo oceno za aritmetično sredino v stratumu k po obrazcu

$$\bar{y}'_{k,2} = \frac{1}{n'_k} \sum_{i=1}^{n'_k} y_{ki} = \frac{1}{n'_k} y'_k \quad (10.67)$$

iz podatkov vzorca v drugi fazi. Enako ocenimo aritmetične sredine v stratumih z vzorčenjem v drugi fazi tudi za druge stratumne. Ocena za agregat za stratificirano vzorčenje v dveh fazah je končno dana s skupnim obrazcem

$$Y_{str,2f} = \sum_{k=1}^r N_{k,1.sl} \cdot \bar{y}'_{k,2} = \frac{N}{n} \sum_{k=1}^r \frac{n_k}{n'_k} \cdot y'_k \quad (10.68)$$

Pri tem pomeni: $Y_{str,2f}$ = ocena agregata Y z vzorčenjem v dveh fazah s stratifikacijo; $N_{k,1.sl}$ = ocena števila enot N_k po stratumih z vzorcem v prvi fazi; $\bar{y}'_{k,2}$ = ocena povprečja $\bar{y}_k$ v stratumu k z vzorcem v drugi fazi; n = število enot v vzorcu v prvi fazi; n_k = število enot v stratumu k v vzorcu iz prve faze; n'_k = število enot v vzorcu v drugi fazi v stratumu k ; y'_k = vsota podatkov za znak y v vzorcu v drugi fazi v stratumu k .

10.64

Primer za ocenjevanje z vzorčenjem v dveh fazah pri stratifikaciji. Kot primer za uporabo vzorčenja v dveh fazah pri stratifikaciji vzemimo določanje lesne zaloge za določen sestoj. n dreves, ki smo jih izbrali iz sestoja z velikim vzorcem v prvi fazi, razdelimo po premeru v debelinske stopnje: $n_1, n_2, \dots, n_r$. Iz dreves, ki smo jih z vzorcem v prvi fazi dobili po posameznih debelinskih stopnjah, izberemo v drugi fazi v vsaki debelinski stopnji samostojen manjši vzorec z n'_k drevesi in zanje ugotovimo volumne y_{ki} . Oceno volumna za sestoj dobimo iz dobljenih podatkov po obrazcu 10.68.

Vzemimo, da na nekem obsežnem kompleksu, ki je napaden z določenim škodljivcem, ocenjujemo število okuženih dreves. Z

vzorcem v prvi fazi, v katerem smo zajeli n parcel, izbrane parcele na grobo kategoriziramo v tri stratumе: malo okužene, srednje okužene in močno okužene parcele. Iz n_1 parcel, ki smo jih iz vzorca v prvi fazi kategorizirali kot malo okužene, izberemo z vzorcem v drugi fazi n_1' parcel. Za teh n_1' parcel s štetjem ugotovimo točno število dreves, ki so okužena. Postopek ponovimo v drugem in tretjem stratumu. Iz dobljenih podatkov sestavimo oceno za celoten sestoj po obrazcu 10.68.

10.65

Varianca ocene z vzorčenjem v dveh fazah. Varianca za oceno z vzorčenjem v dveh fazah je odvisna od vzorca v prvi fazi in vzorcev v drugi fazi. Vendar iz vrednotenja varianc za ocene iz vzorcev v dveh fazah presega naš okvir.

Sestavljanje planov vzorčenja v praksi

10.66

Pri obravnavanju ocenjevanja z vzorčenjem smo navedli nekatere postopke, ki jih uporabljamo, da ocenjujemo parametre za populacijo. Tako smo navedli dvoje vrst izborov: čisti slučajnostni izbor in sistematičen izbor; več vrst vzorčenja: enostavno vzorčenje, stratificirano vzorčenje, vzorčenje v skupnicah, vzorčenje v dveh stopnjah in vzorčenje v dveh fazah; in več vrst ocenjevanj: enostavne ocene in ocene po metodi razmerij. Že pri obravnavanju posameznih metod smo včasih kombinirali po dva ali več navedenih elementov. V praktičnem izvajanju vzorčenja za ocenjevanje parametrov pa po pravilu kombiniramo najrazličnejše elemente vzorčenja tako, da dobimo s čim manjšimi stroški čim bolj zanesljivo oceno.

11. Vzorčenje - Preskušanje hipotez

11.1

Osnove. Z vzorčenjem ocenjujemo parametre populacij na dva načina: s točkovnimi in z intervalnimi ocenami. Te ocene pa dajo uporabne rezultate le v primerih, da je vzorec, ki služi kot osnova za ocenjevanje, zadosti velik. Ocene za male vzorce so običajno tako nezanesljive, da nimajo velike praktične vrednosti. Ocena, za katero je n.pr. relativna standardna pógreška 50 % ali še celo več, je praktično brez vrednosti.

V raziskovalnem delu pa z vzorčenjem rešujemo tudi druge probleme, ki so dostikrat rešljivi tudi z razmeroma majhnimi vzorci. S slučajnostnimi vzorci namreč z uspehom preskušamo določene hipoteze o populaciji oziroma populacijah. V praksi pogosto naletimo na probleme preskušanja hipotez. Z vzorčenjem moremo preskusiti, ali proizvodnja desk na žagi ustreza predpisu o povprečni širini. S statističnimi metodami preskušanja hipotez ugotavljamo, ali določeno prepariranje lesa vpliva na karakteristike lesa ali ne. Enako moremo s tehniko preskušanja hipotez odkriti ali je odstotek žagarske hlo-dovine v dveh bukovih sestojih različen ali ne. Podobno z vzorci preskušamo odstopanja okularnih ocen od stvarnih vrednosti parametrov itd. Z enostavno analizo variance, ki temelji na preskušanju hipotez o proučevani populaciji, moremo odkriti ali določen faktor vpliva na neke značilnosti populacije ali ne. Z enostavno analizo variance moremo n.pr. kompleksno ugotoviti ali več različnih postopkov dá različne rezultate. S kompliciranejšimi postopki analize variance, ki so osnova posebne discipline - planiranja eksperimentov - pa moremo istočasno analizirati in preskusiti vpliv večih faktorjev.

11.2

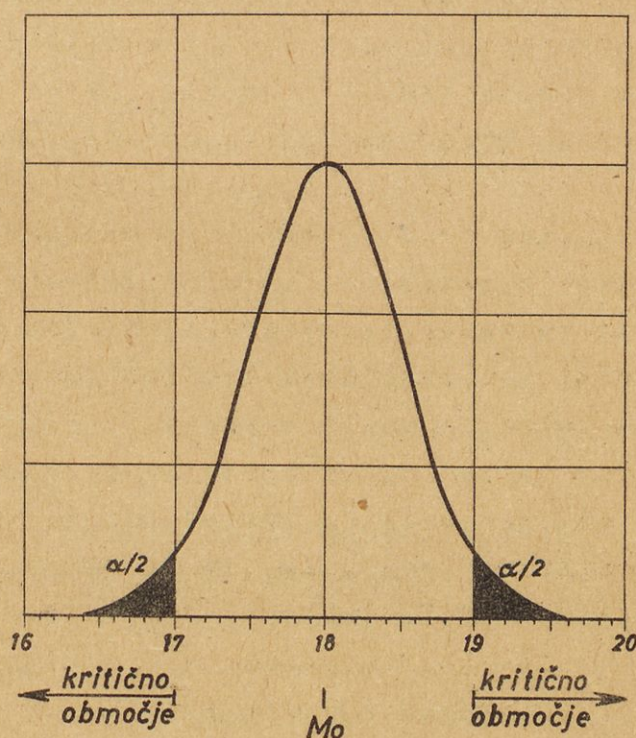
Splošno problematiko preskušanja hipotez proučimo na primeru. Proučujmo tedensko proizvodnjo desk v nekem žagarskem obratu. Vzemimo, da se širine desk porazdeljujejo v normalni porazdelitvi, za katero poznamo standardni odklon $\sigma = 3$ cm. Povprečno širino desk pa ne poznamo.

Vzemimo, da proizvajalec trdi, da je povprečna širina proizvedenih desk $M_0 = 18$ cm. Prevezemnik skuša trditev

proizvajalca preskusiti oziroma ovreči.

Hipoteza o povprečni širini desk je, da je $M_0 = 18$ cm. Z vzorcem skušamo to hipotezo potrditi ali ovreči. Vzemimo, da hipotezo preskušamo z vzorcem $n = 36$ desk, ki jih na slučajno-sten način izberemo iz osnovne populacije. Po znanih stavkih se povprečja vseh možnih vzorcev po $n = 36$ desk porazdeljujejo v normalni porazdelitvi, za katero je standardni odklon enak standardni pogreški za sredino $SE(\bar{y}) = \sigma/\sqrt{n} = 3/\sqrt{16} = 0,75$ cm. Če postavljena hipoteza drži, je aritmetična sredina normalne porazdelitve enaka hipotetični sredini $M = M_0 = 18$ cm.

Porazdelitev povprečij iz vzorcev za vse možne vzorce za ta primer je narisana v sliki 11.1. Iz slike je raz-



Slika 11.1 Porazdelitev povprečij iz vzorcev

vidno, da je velika verjetnost, da iz populacije vseh desk izberemo vzorec, za katerega se aritmetična sredina od $M = 18$ ne razlikuje mnogo, če postavljena hipoteza, da je $M = M_0 = 18$ cm, drži. V tem primeru z verjetnostjo 0,95 aritmetična

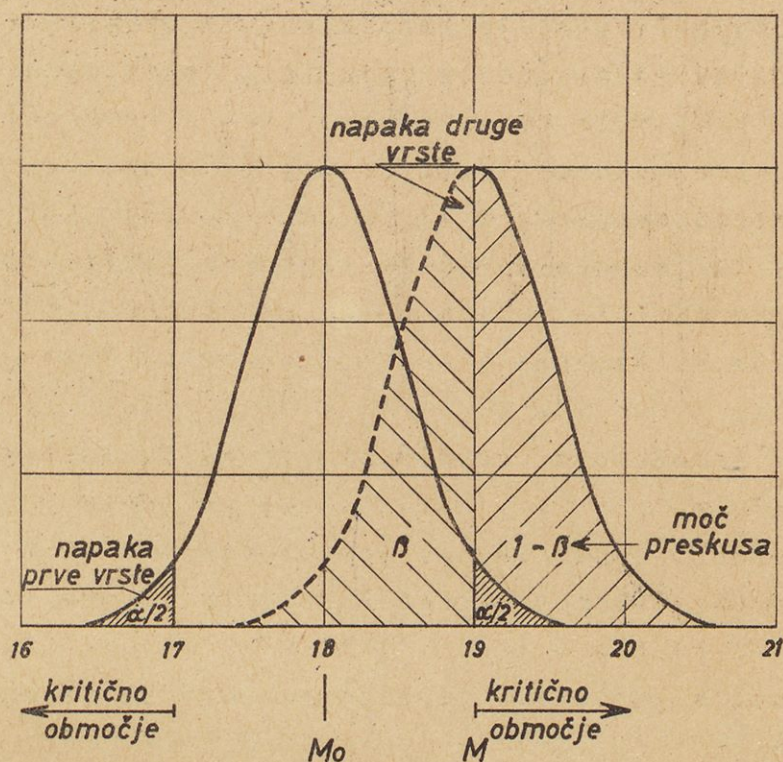
sredina preskusnega vzorca ne bo manjša kot $18 - 1,96 \cdot 0,5 = 17,02$ cm in ne večja kot $18 + 1,96 \cdot 0,5 = 18,98$ cm. Ker je razmeroma majhna verjetnost ($\alpha = 0,05$), da s preskusnim vzorcem dobimo povprečje, ki je manjše kot 17,02 ali večje kot 18,98 cm, postavimo naslednje pravilo: Hipotezo, da je širina desk enaka $M = M_0 = 18$ cm, sprejmemo, če je povprečje preskusnega vzorca v razmaku od 17,02 cm do 18,98 cm. Če pa je povprečje iz vzorca manjše kot 17,02 cm ali večje kot 18,98 cm, hipotezo, da je prava aritmetična sredina širine desk enaka $M = 18$ cm, zavrnamo in sklepamo, da povprečna širina desk ni 18 cm. Območje, v katerem hipotezo zavrnamo, imenujemo kritično območje.

Če podrobneje proučimo to pravilo, spoznamo, da obstaja možnost, da hipoteza drži, jo pa po našem pravilu zavrnamo. Z določeno, sicer majhno verjetnostjo ($\alpha = 0,05$) more povprečje iz preskusnega vzorca biti v kritičnem območju, kljub temu, da hipoteza drži. To napako, ki je v tem, da hipotezo zavrnamo, kljub temu, da hipoteza drži, imenujemo napako prve vrste.

Razen napake prve vrste pa je v tem načinu sklepanja še druga nevarnost. Vzemimo, da naša hipoteza o povprečni širini desk ne drži in je prava povprečna širina $M = 19$ cm. Aritmetične sredine iz preskusnih vzorcev se v tem primeru goste okrog $M = 19$ cm in ne okrog $M_0 = 18$ cm, ker je pravo povprečje 19 cm in ne 18 cm. Iz slike 11.2 je razvidno, v kakšni situaciji smo v tem primeru. Če ohranimo zgornje pravilo, da hipotezo $M = M_0 = 18$ cm sprejmemo, če je povprečje iz preskusnega vzorca v razmaku 17,02 cm do 18,98 cm, prava vrednost povprečja pa je $M = 19$ cm, spoznamo, da obstaja določena in sicer precejšnja verjetnost ($\beta = 0,484$), da hipotezo $M = M_0 = 18$ cm sprejmemo, čeprav ne drži. Napako, ki je v tem, da hipotezo sprejmemo, kljub temu, da je napačna, imenujemo napako druge vrste in jo zaznamujemo z β . Obratno pa je $1 - \beta$, ki jo imenujemo moč preskusa, verjetnost, da osnovno hipotezo sprejmemo, če ta drži.

Kot vidimo iz primera, je napaka druge vrste znatna in v bistveni meri odvisna od prave povprečne širine desk. Čim manjša je razlika med hipotetično in stvarno vrednostjo, tem večja je napaka druge vrste in obratno, čim večja je razlika,





Slika 11.2 Preskušanje ničelne hipoteze

lika med stvarno in hipotetično vrednostjo, tem manjša je napaka druge vrste.

Iz zgornjega spoznamo, da je po zgornjem pravilu veliko bolj tvegano hipoteze sprejemati kot pa zavračati, ker je napaka druge vrste, da hipotezo sprejmemo, čeprav ne drži, v splošnem neznana oziroma velika.

Nasprotno pa moremo kritično območje določiti tako, da je napaka prve vrste, ki je v tem, da hipotezo zavrnemo, čeprav je pravilna, poljubno majhna.

11.3

Ničelna hipoteza. Iz zgornjega izvajanja sledi, da v splošnem lahko razmeroma zanesljivo hipoteze zavračamo, medtem ko je pri sprejemanju hipotez tveganje lahko zelo veliko. Da se izognemo temu navideznemu nesoglasju, običajno postavimo hipoteze tako, da jih razmeroma zanesljivo sprejemamo in

ne zavračamo. To dosežemo z vpeljavo ničelne hipoteze.

Po tem principu vsaki hipotezi, katero preskušamo, postavimo ustrezno negativno protihipotezo, ki jo imenujemo ničelna hipoteza. V nadaljnjem preskušamo ničelno hipotezo. Če uspemo, da zavrremo ničelno hipotezo, smo istočasno sprejeli osnovno hipotezo.

Če preskušamo hipotezo, ali je povprečna širina deska na žagi različna od $M_0 = 18$ cm ($H_1 : M \neq M_0 = 18$ cm), je ustrezna ničelna hipoteza, da je povprečna širina enaka 18 cm ($H_0 : M = M_0 = 18$ cm).

Osnovni hipotezi, da je v nekem sestoju več kot 40 % furnirske hlodovine ($H_1 : P > 40 \%$), ustreza ničelna hipoteza, da je odstotek furnirske hlodovine manjši od 40 % ($H_0 : P \leq 40 \%$).

Osnovni hipotezi, da je določen postopek impregniranja učinkovit, ustreza ničelna hipoteza, da postopek ni učinkovit.

Po tem principu preskušamo hipoteze po naslednjem postopku:

- a) Osnovni hipotezi H_1 priredimo ustrezno ničelno hipotezo H_0 .
- b) Iz osnovne populacije izberemo vzorec, ki služi za preskušanje hipotez o proučevani populaciji.
- c) Za vzorčni izraz, ki ga izračunamo iz podatkov preskusnega vzorca in ničelne hipoteze, določimo kritično območje, v katerem moremo z določenim tveganjem α ničelno hipotezo zavrniti.
- d) Iz podatkov preskusnega vzorca in podatkov o ničelni hipotezi izračunamo ustrezní vzorčni izraz.
- e) Če se dobljena izračunana vrednost nahaja v kritičnem območju, ničelno hipotezo zavrremo oziroma osnovno hipotezo sprejmemo na določeni stopnji tveganja prve vrste za ničelno hipotezo. Če vrednost, ki jo dobimo, če podatke preskusnega vzorca in ničelne hipoteze vnesemo v vzorčni izraz, pade v kritično območje, pravimo, da je stvarno stanje značilno različno od ničelne hipoteze. Običajno določamo kritična območja na treh stopnjah tveganja $\alpha = 0,05$; $\alpha = 0,01$; $\alpha = 0,001$.

Preskušanje hipoteze o sredini

11.4

Ker se za vzorce, ki jih izberemo iz normalne populacije, aritmetične sredine vzorcev porazdeljujejo v normalni populaciji z aritmetično sredino, ki je enaka aritmetični sredini v osnovni populaciji M in s standardnim odklonom $SE(\bar{y}) = \sigma/\sqrt{n}$, se vzorčni izraz

$$\frac{\bar{y} - M}{\sigma/\sqrt{n}} = z \quad (11.1)$$

porazdeljuje v standardizirani normalni porazdelitvi. Zato moremo z izrazom

$$\frac{\bar{y} - M_H}{\sigma/\sqrt{n}} = z \quad (11.2)$$

preskušati ničelno hipotezo, da je prava aritmetična sredina enaka hipotetični M_H ($H_0: M = M_H$).

Če ničelna hipoteza drži in je $M_H = M$, je izračunani z z verjetnostjo 0,95 v razmaku $-1,96 < z < +1,96$, z verjetnostjo 0,99 v razmaku $-2,58 < z < +2,58$ in z verjetnostjo 0,999 v razmaku $-3,29 < z < +3,29$.

Iz tega zaključimo naslednje: a) Če je absolutna vrednost po obrazcu 11.2 izračunanega izraza z manjša kot kritična vrednost 1,96, zaključimo, da je razlika med pravo in hipotetično aritmetično sredino neznačilna. To pomeni, da prekus razlik ni odkril, ne pa da je prava aritmetična sredina enaka hipotetični.

b) Če je absolutna vrednost po obrazcu 11.2 izračunanega z večja kot kritična vrednost 1,96 in manjša kot kritična vrednost 2,58, zaključimo, da so razlike med stvarno in hipotetično sredino značilne na stopnji tveganja $\alpha = 0,05$. To pomeni, da je verjetnost 0,95, da je stvarna sredina resnično različna od hipotetične.

c) Če je po obrazcu 11.2 izračunana vrednost z večja kot 2,58 in manjša kot 3,29 s tveganjem $\alpha = 0,01$, zaključimo, da je stvarna aritmetična sredina značilno različna od hipotetične.

č) Če pa je po obrazcu 11.2 izračunani izraz absolutno večji kot 3,29, zaključimo, da so razlike med M in M_H značilno različne s tveganjem $\alpha = 0,001$.

11.5

Vzemimo kot praktičen problem nakazan primer proučevanja širine desk in preskusimo hipotezo, da je povprečna širina desk značilno različna od $M_H = 18$ cm. Ustrezna ničelna hipoteza je $H_0 : M = M_H = 18$ cm. Da preskusimo dano ničelno hipotezo, smo izbrali slučajnostni vzorec $n = 36$ desk, in zanj ugotovili, da je povprečje iz vzorca $\bar{y} = 16,5$ cm. Ker poznamo standardni odklon proizvodnje desk ($\sigma = 3$ cm), moremo izračunati po obrazcu 11.2 vzorčni izraz

$$z = \frac{\bar{y} - M_H}{\sigma/\sqrt{n}} = \frac{16,5 - 18,0}{3,0/\sqrt{36}} = -3,00$$

Ker je absolutna vrednost izračunanega vzorčnega izraza večja kot kritična meja 2,58 za tveganje $\alpha = 0,01$ in manjša kot kritična meja 3,29 za tveganje $\alpha = 0,001$, s tveganjem $\alpha = 0,01$ sklepamo, da je stvarna povprečna širina desk značilno različna od $M_H = 18$ cm.

11.6

Če ne poznamo prave vrednosti za standardni odklon v osnovni populaciji, se za preskušanje hipotez o sredini ne moremo poslužiti obrazca 11.2. V tem primeru preskušamo hipoteze o aritmetičnih sredinah za normalne populacije po obrazcu

$$\frac{\bar{y} - M}{s/\sqrt{n}} = t(m=n-1) \quad (11.3)$$

ki je zelo sličen obrazcu 11.2, le da ima namesto prave vrednosti standardnega odklona oceno s , ki jo izračunamo po znanem obrazcu za nepristransko ocenjevanje varianc

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}$$

Izraz t pa se ne porazdeljuje v standardizirani normalni porazdelitvi, temveč v distribuciji, ki je enako kot normalna porazdelitev unimodalna, simetrična in zvonasta in se

normalni porazdelitvi tembolj približuje, čim večji je preskusni vzorec. t-porazdelitev je odvisna od velikosti vzorca ali točneje od števila stopinj prostosti m , ki je z velikostjo vzorca v ozki zvezi. Pri zgornjem preskusu je število stopinj prostosti za eno manjše kot je število enot v preskusnem vzorcu: $m = n - 1$.

Glede na to, da je t-porazdelitev odvisna od stopinj prostosti m , so v tabeli 11.1 za vsako stopinjo prostosti dane kritične vrednosti za tveganje $\alpha = 0,05$, $\alpha = 0,01$ in $\alpha = 0,001$.

Tabela 11.1 t-porazdelitev za $\alpha = 0,05$, $\alpha = 0,01$ in $\alpha = 0,001$

Kritične vrednosti				Kritične vrednosti			
m	$t_{0,05}$	$t_{0,01}$	$t_{0,001}$	m	$t_{0,05}$	$t_{0,01}$	$t_{0,001}$
1	12,71	63,66	63,7	26	2,06	2,78	3,71
2	4,30	9,92	31,6	27	2,05	2,77	3,69
3	3,18	5,84	12,9	28	2,05	2,76	3,67
4	2,78	4,60	8,61	29	2,04	2,76	3,66
5	2,57	4,03	6,86	30	2,04	2,75	3,65
6	2,45	3,71	5,96	35	2,03	2,72	3,59
7	2,36	3,50	5,40	40	2,02	2,71	2,55
8	2,31	3,36	5,04	45	2,02	2,69	2,52
9	2,26	3,25	4,78	50	2,01	2,68	3,50
10	2,23	3,17	4,59	60	2,00	2,66	3,46
11	2,20	3,11	4,44	70	2,00	2,65	3,44
12	2,18	3,06	4,32	80	1,99	2,64	3,42
13	2,16	3,01	4,22	90	1,99	2,63	3,40
14	2,14	2,98	4,14	100	1,98	2,63	3,39
15	2,13	2,95	4,07	120	1,98	2,62	3,37
16	2,12	2,92	4,02	150	1,98	2,61	3,36
17	2,11	2,90	3,96	200	1,97	2,60	3,34
18	2,10	2,88	3,92	300	1,97	2,59	3,32
19	2,09	2,86	3,88	400	1,97	2,59	3,32
20	2,09	2,84	3,85	500	1,96	2,59	3,31
21	2,08	2,83	3,82	1000	1,96	2,58	3,30
22	2,07	2,82	3,79	∞	1,96	2,58	3,29
23	2,07	2,81	3,77				
24	2,06	2,80	3,74				
25	2,06	2,79	3,72				

Vzemimo za primer preskus, ali je povprečna upogibna trdnost lesa za določeno drevo za črni gaber značilno različna od $M_H = 1500 \text{ kg/cm}^2$. V ta namen smo v preskusni vzorec vzeli $n = 8$ preskušancev iz različnih delov drevesa in zanje ugotovili naslednje upogibne trdnosti: v kg/cm^2 : 1560, 1530, 1400, 1650, 1360, 1430, 1440, 1440. Iz teh podatkov dobimo:

$$y = \sum_{i=1}^n y_i = 1560 + 1530 + \dots + 1440 = 11810$$

$$\sum y_i^2 = 1560^2 + 1530^2 + \dots + 1440^2 = 1749870$$

$$\bar{y} = y/n = 11810/8 = 1476$$

$$s^2 = \frac{\sum y_i^2 - y^2/n}{n-1} = \frac{1749870 - 11810^2/8}{8-1} = 9170$$

$$s = \sqrt{s^2} = \sqrt{9170} = 95,76$$

Iz teh podatkov dobimo po obrazcu 11.3

$$t = \frac{\bar{y} - M_H}{s} \sqrt{n} = \frac{1476 - 1500}{95,76} \sqrt{8} = -1,00$$

Ker je kritična meja za t -porazdelitev za $m = 8 - 1 = 7$ stopinj prostosti za $\alpha = 0,05$ po tabeli 11.1 enaka $t_{0,05}(m=7) = 2,36$, zaključimo, da stvarna povprečna upogibna trdnost preučevanega drevesa črnega gabra ni značilno različna od hipotetične, ker je izračunani t manjši kot ustrezna kritična vrednost za $\alpha = 0,05$. Neznačilnost razlike pa ne pomeni, da razlik ni, temveč samo, da jih z izvršenim preskusom nismo odkrili.

Preskušanje značilnosti razlik med aritmetičnimi sredinami

11.8

V raziskovalnem delu se dostikrat pojavlja vprašanje, ali so povprečja za različne populacije med seboj različna ali ne. Ničelno hipotezo, da sta aritmetični sredini za dve populaciji, ki se porazdeljujeta v normalnih porazdelitvah z enakima variancama, med seboj enaki ($H_0 : M_1 = M_2$), preiskujemo z naslednjim izrazom

$$\frac{\bar{y}_1 - \bar{y}_2}{s_d} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = t(m = n_1 + n_2 - 2) \quad (11.4)$$

ki se porazdeljuje v t-porazdelitvi s številom stopinj prostosti $m = n_1 + n_2 - 2$.

Da po tem obrazcu preskusimo značilnost razlik med aritmetičnima sredinama M_1 in M_2 , moramo iz prve in druge populacije izbrati dva samostojna vzorca. Pri tem z n_1 , $\bar{y}_1$ in s_1^2 zaznamujemo število enot, oceno povprečja in oceno varian- ce za vzorec iz prve populacije, z n_2 , $\bar{y}_2$ in s_2^2 pa analogne količine iz drugega vzorca.

11.9

Če vzamemo za primer preskušanje značilnosti razlik v upogibni trdnosti med dvema drevesoma črnega gabra, poteka preskus po naslednjih stopnjah.

Za drevo 1 smo na slučajnostem način izbrali $n_1 = 8$ kosov lesa in zanj dobili naslednje rezultate o upogibni trdnosti v kg/cm^2 : 1460, 1500, 1330, 1550, 1570, 1510, 1560, 1500. Enako smo tudi za drugo drevo vzeli $n_2 = 8$ preskušancev na slučajnostno izbranih mestih in zanje izmerili upogibno trdnost v kg/cm^2 . Dobili smo naslednje rezultate: 1240, 1320, 1400, 1240, 1320, 1340, 1440, 1400. Iz teh podatkov dobimo elemente za preskus značilnosti razlik med povprečno upogibno trdnostjo po znanih obrazcih za ocenjevanje sredine in varian- ce:

$$n_1 = 8; \quad \bar{y}_1 = 1498; \quad s_1^2 = 5936$$

$$n_2 = 8; \quad \bar{y}_2 = 1337; \quad s_2^2 = 5421$$

$$s_d^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(8 - 1)5936 + (8 - 1)5421}{8 + 8 - 2} = 5678$$

$$s_d = \sqrt{5678} = 75,35$$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_d} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = \frac{1498 - 1337}{75,35} \sqrt{\frac{8 \cdot 8}{8 + 8}} = +4,41$$

Ker je izračunani $t = 4,41$ absolutno večji kot kri-

tična vrednost t za $m = n_1 + n_2 - 2 = 14$, pri tveganju $\alpha = 0,001$ [$t(m=14; \alpha=0,001)=4,14$], zelo zanesljivo ($\alpha=0,001$), sklepamo, da so razlike v upogibni trdnosti med proučevanima drevesoma značilne.

Preskušanje hipotez o varianci

11.10

Proučevanje variabilnosti v najrazličnejših oblikah zajema velik del problemov v zvezi s statističnim proučevanjem eksperimentalnih podatkov. Velika variabilnost v izdelkih je znak slabe kvalitete proizvodnje, velika variabilnost določenih značilnosti v lesu je izraz neenovitih pogojev rasti in ima posledice pri nadaljnji obdelavi lesa itd. Čim manjša variabilnost je v večini izraz boljše kvalitete surovin, dela itd.

V zvezi s proučevanjem variabilnosti je najenostavnejši problem preskušanje hipotez o varianci za določeno populacijo.

Če predpostavljamo, da smo iz populacije, ki se normalno porazdeljuje, izbrali slučajnostni vzorec n enot, se izraz

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2(m=n-1) \quad (11.5)$$

v populaciji vseh možnih vzorcev porazdeljuje v porazdelitvi, ki jo imenujemo χ^2 (hi-kvadrat porazdelitev). χ^2 -porazdelitev je podobno kot t -porazdelitev odvisna od stopinj prostosti, ki so pri zgornjem izrazu za eno manjša kot je velikost vzorca ($m = n - 1$) (glej tabelo 11.2).

Glede na število stopinj prostosti je iz podatkov slučajnostnega vzorca izračunani χ^2 s tveganjem $\alpha = 0,05$ manjši kot je kritična vrednost $\chi^2_{0,05}$ v tabeli 11.2; s tveganjem $\alpha = 0,01$ manjši kot so kritične vrednosti $\chi^2_{0,01}$ in s tveganjem $\alpha = 0,001$ manjši kot so kritične vrednosti $\chi^2_{0,001}$.

Zgornje lastnosti izkoriščamo za preskušanje hipotez o varianci. Hipotezo, da je varianca v osnovni populaciji različna od predpisane hipotetične variance σ_H^2 preskušamo z ničelno hipotezo, da je varianca za preiskovano populacijo enaka σ_H^2 ($H_0: \sigma^2 = \sigma_H^2$).

Tabela 11.2. χ^2 -porazdelitevKritične vrednosti za $\alpha = 0,05$, $\alpha = 0,01$ in $\alpha = 0,001$

m	$\chi^2_{0,05}$	$\chi^2_{0,01}$	$\chi^2_{0,001}$
1	3,84	6,64	10,93
2	5,99	9,21	13,82
3	7,82	11,34	16,27
4	9,49	13,28	18,46
5	11,07	15,09	20,52
6	12,59	16,81	22,46
7	14,07	18,48	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88
10	18,31	23,21	29,59
11	19,68	24,72	31,26
12	21,03	26,22	32,91
13	22,36	27,69	34,53
14	23,68	29,14	36,12
15	25,00	30,58	37,70
16	26,30	32,00	39,25
17	27,59	33,41	40,79
18	28,87	34,80	42,31
19	30,14	36,19	43,82
20	31,41	37,57	45,32
21	32,67	38,93	46,80
22	33,92	40,29	48,27
23	35,17	41,64	49,73
24	36,42	42,98	51,18
25	37,65	44,31	52,62
26	38,88	45,64	54,05
27	40,11	46,96	55,48
28	41,34	48,28	56,89
29	42,56	49,59	58,30
30	43,77	50,89	59,70

Podobno kot pri preskušanju hipotez o sredinah, tudi za varianco zaključimo, da prava varianca ni značilno različna od hipotetične variance σ_H^2 , če je izraz

$$\frac{(n-1)s^2}{\sigma_H^2} = \chi^2 \quad (11.6)$$

pod kritično vrednostjo $\chi_{0,05}^2$, da je prava varianca s tveganjem $\alpha = 0,05$ značilno različna od ničelne hipoteze, če je po obrazcu izračunani izraz med ustreznima vrednostima $\chi_{0,05}^2$ in $\chi_{0,01}^2$, značilen s tveganjem $\alpha = 0,01$, če je izračunani izraz med $\chi_{0,01}^2$ in $\chi_{0,001}^2$, in značilen na stopnji tveganja $\alpha = 0,001$, če je izračunani χ^2 večji kot $\chi_{0,001}^2$.

11.11

Za smrekov les preiskujemo, ali je varianca nominalne prostorninske teže smrekovega lesa določene partije večja kot $\sigma_H^2 = 30,00 \text{ kg/m}^3$. V ta namen smo izmerili nominalno prostorninsko težo za $n = 25$ preskušancev in iz njih ugotovili, da je ocena variance $s^2 = 37,91 \text{ kg/m}^2$.

Če dobljene podatke vnesemo v obrazec 11.6, dobimo

$$\chi^2 = \frac{(n-1)s^2}{\sigma_H^2} = \frac{(25-1) \cdot 37,91}{30,00} = 30,33$$

Ker $m = n-1 = 25-1 = 24$ stopinjam prostosti ustreza $\chi_{0,05}^2 = 36,42$, sklepamo, da varianca nominalne prostorninske teže ni značilno večja kot $\sigma_H^2 = 30,0 \text{ kg/m}^3$.

11.12

V tabeli 11.2 so dane kritične vrednosti za χ^2 samo do $m = 30$. Za primere, v katerih je $m > 30$, dobimo ustrezne kritične vrednosti za posamezne stopinje prostosti po obrazcih:

Za $\alpha = 0,05$ je

$$\chi_{0,05}^2 = \frac{1}{2} (\sqrt{2m-1} + 1,645)^2$$

za $\alpha = 0,01$ je

$$\chi_{0,01}^2 = \frac{1}{2} (\sqrt{2m-1} + 2,326)^2$$

za $\alpha = 0,001$ je

$$\chi_{0,001}^2 = \frac{1}{2} (\sqrt{2m-1} + 3,090)^2$$

(11.7)

Vzemimo, da smo zaradi neznačilnosti razlik v varianci v prejšnjem primeru vzorec podvojili in izmerili nominalno prostorninsko težo za skupno $n = 50$ preskušancev lesa. Nova ocena za varianco je $s^2 = 42,0$. Izračunani χ^2 je enak

$$\chi^2 = \frac{(n-1) s^2}{\sigma_H^2} = \frac{(50-1) 42,0}{30,0} = 68,60$$

Ker je $m = n - 1 = 50 - 1 = 49$ večji kot 30, izračunamo kritične vrednosti za $m = 49$ po obrazcih 11.7

$$\chi_{0,05}^2 = \frac{1}{2} (\sqrt{2m-1} + 1,645)^2 = \frac{1}{2} (\sqrt{2 \cdot 49 - 1} + 1,645)^2 = 66,11$$

$$\chi_{0,01}^2 = \frac{1}{2} (\sqrt{2m-1} + 2,326)^2 = \frac{1}{2} (\sqrt{2 \cdot 49 - 1} + 2,326)^2 = 74,12$$

$$\chi_{0,001}^2 = \frac{1}{2} (\sqrt{2m-1} + 3,090)^2 = \frac{1}{2} (\sqrt{2 \cdot 49 - 1} + 3,090)^2 = 83,71$$

Izračunani $\chi^2 = 68,60$ pade med $\chi_{0,05}^2 = 66,11$ in $\chi_{0,01}^2 = 74,12$. Iz tega sklepamo, da je varianca v nominalni prostorninski teži za proučevan les s tveganjem $\alpha = 0,05$ značilno večja kot $\sigma_H^2 = 30,00$.

Preskušanje različnosti med variancama

11.14

Ker je majhna variabilnost znak dobre kvalitete surovin, vestnega dela delavcev ali preciznega dela strojev, je velike važnosti primerjava varianc med surovinami iz različnih virov, med proizvodi, ki so bili proizvedeni pod različnimi pogoji dela, po različnih postopkih, z različnimi stroji itd. Hipoteze, da sta za dve populaciji, ki se normalno porazdeljujeta, varianci različni ($H_1 : \sigma_1^2 \neq \sigma_2^2$) preskušamo z ničelno hipotezo, da sta varianci v obeh populacijah enaki ($H_0 : \sigma_1^2 = \sigma_2^2$).

Če velja ničelna hipoteza, da sta varianci v obeh populacijah enaki, se kvocient ocen varianc iz prve s_1^2 in druge populacije s_2^2

$$\frac{s_1^2}{s_2^2} = F$$

(11.8)

(pri tem vzamemo kot prvo populacijo ono, za katero je ocena variance večja: $s_1^2 > s_2^2$) v populaciji vseh možnih vzorcev z n_1 enotami v prvi in z n_2 enotami v drugi populaciji, porazdeljuje v F-porazdelitvi. F porazdelitev je odvisna od stopinj prostosti $m_1 = n_1 - 1$, ki zavisi od števila enot v prvem vzorcu in stopinj prostosti $m_2 = n_2 - 1$, ki zavisi od števila enot v drugem vzorcu. V tabelah 11.3 a, b, c so dane m_1 in m_2 ustrežne kritične vrednosti za stopnje tveganja $\alpha = 0,05$, $\alpha = 0,01$ in $\alpha = 0,001$. Tabele vsebujejo kritične vrednosti samo za nekatere stopinje prostosti. Za vmesne vrednosti za stopinje prostosti dobimo ustrezne kritične vrednosti z linearno interpolacijo tako, da vzamemo za osnovo pri interpolaciji recipročne vrednosti iz stopinj prostosti.

11.15

Vzemimo, da preskušamo ali je kvaliteta dela na novo nabavljenega stroja značilno boljša od starega. V ta namen smo na starem stroju proizvedli $n_1 = 20$ lesenih profilov, na novem stroju pa $n_2 = 30$ lesenih profilov enake specifikacije. Varianca, izračunana iz podatkov vzorca za stari stroj, je $s_1^2 = 9,21$, varianca, izračunana iz vzorcev proizvodov na drugem stroju, pa je $s_2^2 = 3,82$.

Da preskusimo značilnost razlike v variabilnosti na starem in novem stroju, izračunamo

$$F = \frac{s_1^2}{s_2^2} = \frac{9,21}{3,82} = 2,41$$

Zgornjim vzorcem ustrezne stopinje prostosti so $m_1 = n_1 - 1 = 20 - 1 = 19$ in $m_2 = n_2 - 1 = 30 - 1 = 29$.

Ker v tabelah nimamo vrednosti, ki ustrezajo $m_1 = 19$, dobimo ustrezne kritične vrednosti z linearno interpolacijo. Za $\alpha = 0,05$ dobimo iz tabele 11.3a

$$F_{0,05} (m_1 = 12; m_2 = 29) = 2,10$$

$$F_{0,05} (m_1 = 24; m_2 = 29) = 1,90$$

TAB.11.3a F - PORAZDELITEV (KRITIČNE VREDNOSTI ZA $\alpha = 0.05 F_{0.01}$)

$m_2 \backslash m_1$	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	245.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.03	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00

TAB.11.3b F - PORAZDELITEV (KRITIČNE VREDNOSTI ZA $\alpha = 0.01$; $F_{0.05}$)

$m_2 \backslash m_1$	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5982	6106	6234	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.42	99.46	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.86	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

TAB.11.3c F - PORAZDELITEV (KRITIČNE VREDNOSTI ZA $\alpha = 0.001$ $F_{0.001}$)

m2 \ m1	1	2	3	4	5	6	8	12	24	∞
1	405284	500000	540379	562500	576405	585937	598144	610667	623497	636619
2	998.5	999.2	999.2	999.2	999.3	999.3	999.4	999.4	999.5	999.5
3	167.0	148.5	141.1	137.1	134.6	132.8	130.6	128.3	125.9	123.5
4	74.14	61.25	56.18	53.44	51.71	50.53	49.00	47.41	45.77	44.05
5	47.18	37.12	33.20	31.09	29.75	28.84	27.64	26.42	25.14	23.78
6	35.51	27.00	23.70	21.92	20.81	20.03	19.03	17.99	16.89	15.75
7	29.25	21.69	18.77	17.19	16.21	15.52	14.63	13.71	12.73	11.69
8	25.42	18.49	15.83	14.39	13.49	12.86	12.04	11.19	10.30	9.34
9	22.86	16.39	13.90	12.56	11.71	11.13	10.37	9.57	8.72	7.81
10	21.04	14.91	12.55	11.28	10.48	9.92	9.20	8.45	7.64	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.35	7.63	6.85	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	7.71	7.00	6.25	5.42
13	17.81	12.31	10.21	9.07	8.35	7.86	7.21	6.52	5.78	4.97
14	17.14	11.78	9.73	8.62	7.92	7.43	6.80	6.13	5.41	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.47	5.81	5.10	4.31
16	16.12	10.97	9.00	7.94	7.27	6.81	6.19	5.55	4.85	4.06
17	15.72	10.66	8.73	7.68	7.02	6.56	5.96	5.32	4.63	3.85
18	15.38	10.39	8.49	7.46	6.81	6.35	5.76	5.13	4.45	3.67
19	15.08	10.16	8.28	7.26	6.62	6.18	5.59	4.97	4.29	3.52
20	14.82	9.95	8.10	7.10	6.46	6.02	5.44	4.82	4.15	3.38
21	14.59	9.77	7.94	6.95	6.32	5.88	5.31	4.70	4.03	3.26
22	14.38	9.61	7.80	6.81	6.19	5.76	5.19	4.58	3.92	3.15
23	14.19	9.47	7.67	6.69	6.08	5.65	5.09	4.48	3.82	3.05
24	14.03	9.34	7.55	6.59	5.98	5.55	4.99	4.39	3.74	2.97
25	13.88	9.22	7.45	6.49	5.88	5.46	4.91	4.31	3.66	2.89
26	13.74	9.12	7.36	6.41	5.80	5.38	4.83	4.24	3.59	2.82
27	13.61	9.02	7.27	6.33	5.73	5.31	4.76	4.17	3.52	2.75
28	13.50	8.93	7.19	6.25	5.66	5.24	4.69	4.11	3.46	2.70
29	13.39	8.85	7.12	6.19	5.59	5.18	4.64	4.05	3.41	2.64
30	13.29	8.77	7.05	6.12	5.53	5.12	4.58	4.00	3.36	2.59
40	12.61	8.25	6.60	5.70	5.13	4.73	4.21	3.64	3.01	2.23
60	11.97	7.76	6.17	5.31	4.76	4.37	3.87	3.31	2.69	1.90
120	11.38	7.32	5.79	4.95	4.42	4.04	3.55	3.02	2.40	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.27	2.74	2.13	1.00

Interpolirano vrednost $F(19)$ za $m_1 = 19$ pa dobimo iz razmerja:

$$\frac{1/19 - 1/24}{1/12 - 1/24} = \frac{F(19) - 1,90}{2,10 - 1,90}$$

Iz tega razmerja dobimo, da je $F_{0,05}(m_1=19, m_2=29) = 1,95$.

Če analogno poiščemo kritično vrednost še za $\alpha = 0,01$, dobimo iz tabele 11.3b

$$F_{0,01}(m_1=12; m_2=29) = 2,87$$

$$F_{0,01}(m_1=24; m_2=29) = 2,49$$

Interpolirano vrednost $F(19)$ za $m_1=19$ dobimo analogno iz razmerja

$$\frac{1/19 - 1/24}{1/12 - 1/24} = \frac{F(19) - 2,49}{2,87 - 2,49}$$

Iz tega razmerja je $F_{0,01}(m_1=19; m_2=29) = 2,59$.

Ker je iz preskusa dobljeni $F = 2,41$ večji kot $F_{0,05}(19,29)$ in manjši kot $F_{0,01}(19,29)$, zaključimo, da novi stroj s tveganjem $\alpha = 0,05$ resnično dela kvalitetnejše kot stari.

11.16

Enostavna analiza variance. V odstavku 11.8 in 11.9 smo iz t-preskusov preskušali značilnost razlik med dvema aritmetičnima sredinama. Dostikrat pa se pojavljajo problemi, pri katerih nastopa več grupnih sredin hkrati. Izolirano preskušanje vseh mogočih parov sredin s t-preskusom je zamudno, ne glede na to, da analitično ni neoporečno.

S postopkom, ki ga imenujemo analiza variance, pa moremo kompleksno preskusiti značilnost razlik med več sredinami hkrati. Če so namreč v r populacijah, v katerih se vrednosti znaka, katerega proučujemo, porazdeljujejo v normalnih porazdelitvah, vse aritmetične sredine med seboj enake ($M_1 = M_2 = \dots M_r$) in vse variance enake ($\sigma_1^2 = \sigma_2^2 = \dots \sigma_r^2$), se izraz

$$\frac{(Q_K - Q)/(r-1)}{(Q_{KI} - Q_K)/(n-r)} = F(m_1=r-1; m_2=n-r)$$

(11.19)

$$Q_{KI} = \sum_{k=1}^r \sum_{i=1}^{n_k} y_{ki}^2 ; \quad Q_K = \sum_{k=1}^r \frac{y_k^2}{n_k} ; \quad Q = \frac{Y^2}{n}$$

porazdeljuje v F-porazdelitvi z $m_1 = r-1$ in $m_2 = n-r$ stopinjami prostosti. Pri tem pomeni n = število vseh osnovnih podatkov v celotnem preskusu; n_k = število osnovnih podatkov v grupi k ; r = število grup; y_{ki} = individualna vrednost i v grupi k ; Y_k = vsota podatkov za vse enote v grupi k ; Y = vsota vseh osnovnih podatkov v vseh grupah.

Analizo variance običajno podajamo v standardni shemi, ki je za naš enostaven primer naslednja:

Tabela 11.4 Shema enostavne analize variance

Vir variance	V. Vsočav kvadratov K	Stopinje prostosti m	Ocena variance s^2	F
Med grupami	$K_K = Q_K - Q$	$m_K = r - 1$	$s_K^2 = \frac{K_K}{m_K}$	$F = s_K^2 / s_I^2$
V grupah	$K_I = Q_{KI} - Q_K$	$m_I = n - r$	$s_I^2 = \frac{K_I}{m_I}$	1
Skupno	$K = Q_{KI} - Q$	$m_K + m_I = n - 1$		

Če postavimo kot ničelno hipotezo, da so aritmetične sredine v r populacijah med seboj enake ($H_0 : M_1 = M_2 = \dots = M_r$), moremo pri pogoju, da so variance po populacijah enake, zgornjo shemo analize variance uporabiti za preskušanje razlik med aritmetičnimi sredinami. Po standardnem načinu zaključevanja smatramo, da so razlike med sredinami za različne grupe oziroma populacije neznačilne, če je iz analize izračunani F manjši kot ustrezni $F_{0,05}(m_1=r-1; m_2=n-r)$. Razlike med grupnimi sredinami so značilne s tveganjem $\alpha = 0,05$, če je izračunani F med ustreznima vrednostima $F_{0,05}(m_1=r-1; m_2=n-r)$ in $F_{0,01}(m_1=r-1; m_2=n-r)$. Razlike med grupnimi sredinami so značilne s tveganjem $\alpha = 0,01$, če je izračunani F med $F_{0,01}(m_1=r-1; m_2=n-r)$ in $F_{0,001}(m_1=r-1; m_2=n-r)$ in značilne s tveganjem $\alpha = 0,001$, če je izračunani F večji kot $F_{0,001}(m_1=r-1; m_2=n-r)$.

11.17

Kot enostavem primer preskušanja razlik med sredinami za več populacij vzemimo preskušanje razlik v povprečni upo-

gibni trdnosti za pet dreves črnega gabra z raziskovalne ploskve v Kamniški Bistrici. Od vsakega drevesa je vzetih po osem kosov lesa na slučajnostni način in zanje izmerjena upogibna trdnost v kg/cm^2 . Osnovni podatki so dani v tabeli 11.5.

Tabela 11.5 Osnovni podatki o upogibni trdnosti za pet dreves črnega gabra z raziskovalne ploskve v Kamniški Bistrici

Drevo	Upogibne trdnosti posameznih kosov lesa v kg/cm^2									
	y_{ki}								y_k	$\bar{y}_k$
1.	1460	1500	1330	1550	1570	1510	1560	1500	11980	1498
2.	1240	1320	1400	1240	1320	1340	1440	1400	10700	1337
3.	1390	1290	1310	1310	1390	1440	1360	1260	10750	1344
4.	1530	1540	1610	1530	1290	1470	1520	1410	11900	1487
5.	1560	1530	1400	1650	1360	1430	1440	1440	11810	1476
									$Y = 57140$	

Iz osnovnih podatkov v tabeli 11.5 povzamemo, da je
 $r = 5$; $n_1 = 8$; $n_2 = 8$; $n_3 = 8$; $n_4 = 8$; $n_5 = 8$; $n = 40$

$$Q_{KI} = \sum_{k=1}^r \sum_{i=1}^{n_k} y_{ki}^2 = 1460^2 + 1500^2 + 1330^2 + \dots + 1440^2 + 1440^2 = 82069800$$

$$Q_K = \sum_{k=1}^r \frac{y_k^2}{n_k} = \frac{11980^2}{8} + \frac{10700^2}{8} + \dots + \frac{11810^2}{8} = 81832375$$

$$Q = \frac{Y^2}{n} = \frac{57140^2}{40} = 81624490$$

Če te vmesne rezultate vnesemo v shemo analize variance, dobimo tabelo 11.6.

Ker je izračunani $F = 7,66$ večji kot $F_{0,001}(4; 35) = 5,88$, sklepamo, da so razlike v povprečni upogibni trdnosti med proučevanimi petimi drevesi črnega gabra značilne s tveganjem $\alpha = 0,001$. Vrednost $F_{0,001}(4; 35)$ smo iz tabele 11.3c dobili z interpolacijo, ki je nakazana v odstavku 11.15.

Tabela 11.6 Analiza variance za upogibno trdnost med petimi drevesi

Vir variacije	Vsota kvadratov	Stopinje prostosti	Ocena variance	F
Med drevesi	Q_K 81832375 $-Q_K$ -81624490 $K_K =$ 207885	$m_K = 5-1=4$	$207885:4 =$ $= 51971 = s_K^2$	$s_K^2/s_I^2 =$ $= 7,66$
V drevesih	Q_{KI} 82069800 $-Q_K$ -81832375 $K_I =$ 237425	$m_I = 40-5=35$	$237425:35 =$ $= 6784 = s_I^2$	1
Skupno	445310	39		

Preskušanje hipotez z velikimi vzorci

11.18

Kljub temu, da z velikimi vzorci običajno ocenjujemo parametre, z njimi včasih tudi preskušamo hipoteze o parametrih proučevanih populacij. Medtem ko je, kot smo videli v prejšnjih odstavkih, preskušanje hipotez z malimi vzorci razmeroma zamotano, glede na to, da se posamezni vzorčni izrazi porazdeljujejo v različnih porazdelitvah in da je pri tem treba paziti še na stopinje prostosti, je preskušanje hipotez z velikimi vzorci razmeroma enotno. Ocene kateregakoli parametra z velikimi vzorci se namreč porazdeljujejo bolj in bolj v normalni porazdelitvi, čim večji je vzorec.

Zato se v populaciji vseh možnih vzorcev za splošen parameter C porazdeljuje izraz

$$\frac{c - C}{SE(c)} = z \quad \text{oziroma} \quad \frac{c - C}{se(c)} = z \quad (11.20)$$

v standardizirani normalni porazdelitvi. V teh obrazcih pomeni: c = ocena parametra C iz vzorca, $SE(c)$ = prava vrednost standardne pogreške; $se(c)$ = ocena standardne pogreške iz podatkov vzorca.

Zato moremo ničelno hipotezo, da je prava vrednost parametra enaka hipotetični ($H_0 : C = C_H$) preskušati z izrazom

$$\frac{c - C_H}{SE(c)} = z \quad \text{oziroma} \quad \frac{c - C_H}{se(c)} = z \quad (11.21)$$

Pri tem so kritične vrednosti za absolutne vrednosti teh izrazov $z_{0,05} = 1,96$; $z_{0,01} = 2,58$ in $z_{0,001} = 3,29$.

Način zaključevanja pa je enak kot pri malih vzorcih.

11.19

Glede na zgornje splošno pravilo preskušamo z velikimi vzorci hipoteze o najvažnejših parametrih: aritmetični sredini M , strukturnem deležu $P\%$ in standardnem odklonu σ z naslednjimi vzorčnimi izrazi:

Aritmetično sredino preskušamo z izrazom

$$\frac{\bar{y} - M_H}{\sigma} \sqrt{n} = z \quad (11.22)$$

kadar poznamo pravo vrednost za standardni odklon, ali z izrazom

$$\frac{\bar{y} - M_H}{s} \sqrt{n} = z \quad (11.23)$$

kadar ne poznamo pravega standardnega odklona, ampak ga ocenimo iz podatkov preskusnega vzorca.

Strukturni delež $P\%$ preskušamo z izrazom

$$\frac{p\% - P\%_H}{\sqrt{P\%_H (100 - P\%_H)}} \sqrt{n} = z \quad (11.24)$$

standardni odklon pa analogno z vzorčnim izrazom

$$\frac{s - \sigma_H}{\sigma_H} \sqrt{2n} = z \quad (11.25)$$

ker je standardna pogreška za standardni odklon za velike vzorce enaka $SE(s) = \sigma / \sqrt{2n}$.

Vzemimo za primer preskušanja hipotez z velikim vzorcem naslednji problem. Z okularno oceno je bilo ocenjeno, da je v nekem hrastovem sestoju 40 odstotkov furnirske hlodovine. Niželna hipoteza je, da je dejanski odstotek furnirske hlodovine enak okularni oceni ($H_0 : P\% = P\%_H = 40\%$). To ničelno hipotezo smo preskusili z vzorcem $n = 250$ dreves. V vzorcu je $h = 79$ dreves, ki ima furnirsko hlodovino. Odstotek iz vzorca je torej $p\% = 31,6\%$. Če zgornje podatke vnesemo v obrazec 11.24, dobimo

$$\frac{p\% - P\%_H}{\sqrt{P\%_H (100 - P\%_H)}} \sqrt{n} = \frac{31,6 - 40}{\sqrt{40 (100 - 40)}} \sqrt{250} = -2,71 = z$$

Ker je absolutna vrednost izračunanega vzorčnega izraza $z = 2,71$ večja kot kritična vrednost $z_{0,01} = 2,58$ in manjša kot $z_{0,001} = 3,29$, s tveganjem $\alpha = 0,01$ sklepamo, da je pravi odstotek dreves v sestoju, ki ima furnirsko hlodovino, značilno različen od okularne ocene.

Preskušanje značilnosti razlik za velike vzorce

11.21

Podobno kot za preskušanje ničelnih hipotez o velikosti parametrov veljajo tudi za preskušanje hipotez o razlikah med parametri enotna pravila, če so preskusni vzorci veliki.

Če proučujemo dve populaciji 1 in 2, preskušamo ničelno hipotezo, da vrednosti za proučevani parameter v populacijah 1 in 2 nista različni ($H_0 = C_1 = C_2$), z dvema samostojnima vzorcema iz proučevanih populacij 1 in 2. Z vzorčnim izrazom

$$\frac{c_1 - c_2}{\sqrt{SE^2(c_1) + SE^2(c_2)}} = z \quad (11.26)$$

preskušamo hipotezo o razlikah, če poznamo prave vrednosti standardnih pogrešk $SE(c_1)$ in $SE(c_2)$ za oceni parametrov c_1 iz prvega in c_2 iz drugega vzorca. Če pa pravih vrednosti standardnih pogrešk ne poznamo, preskušamo zgornjo ničelno

hipotezo z obrazcem

$$\frac{c_1 - c_2}{\sqrt{se^2(c_1) + se^2(c_2)}} = z \quad (11.27)$$

pri čemer sta $se(c_1)$ in $se(c_2)$ oceni standardnih pogrešk za oceni c_1 in c_2 , izračunani iz podatkov vzorca. Na značilnost oziroma na neznačilnost razlik v parametrih zaključujemo po znanem postopku tako, da izračunano vrednost primerjamo s kritičnimi vrednostmi $z_{0,05} = 1,96$, $z_{0,01} = 2,58$ in $z_{0,001} = 3,29$.

11.22

Če splošna obrazca 11.26 in 11.27 priredimo za preiskovanje razlik med aritmetičnimi sredinami, strukturnimi deleži in standardnimi odkloni, dobimo naslednje obrazce:

Razlike med aritmetičnimi sredinami za populaciji 1 in 2 preskušamo z obrazcem

$$\frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = z \quad (11.28)$$

ali z obrazcem

$$\frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = z \quad (11.29)$$

glede na to ali poznamo prave vrednosti za varianci v obeh populacijah (σ_1^2 in σ_2^2) ali jih ocenjujemo v vzorcih (s_1^2 in s_2^2).

n_1 in n_2 je število enot v vzorcu iz prve oziroma druge populacije, $\bar{y}_1$ in $\bar{y}_2$ pa sta oceni za aritmetični sredini, izračunani iz prvega in drugega vzorca.

Razlike med dvema strukturnima deležema za dve populaciji preskušamo z obrazcem

$$\frac{p_1\% - p_2\%}{\sqrt{\frac{p_1\%(100 - p_1\%)}{n_1 - 1} + \frac{p_2\%(100 - p_2\%)}{n_2 - 1}}} = z \quad (11.30)$$

pri čemer sta poleg znanih izrazov $p_1\%$ in $p_2\%$ oceni strukturnega deleža, izračunani iz podatkov prvega in drugega vzorca.

Razlike med standardnima odklonoma za dve populaciji

ji pa preskušamo z obrazcem

$$\frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{n_2}}} = z \quad (11.31)$$

v katerem je s_1 ocena standardnega odklona za prvo, s_2 pa ocena standardnega odklona za drugo populacijo.

11.23

Vzemimo za primer, da preskušamo značilnost razlik v premeru dreves v prsni višini za dva istodobna sestoja, ki rasteja v različnih pogojih. V ta namen smo na slučajnosten način izbrali iz prvega sestoja $n_1 = 100$ dreves in zanje izmerili premere v prsni višini. Iz podatkov vzorca smo ugotovili, da je ocena za povprečen premer $\bar{y}_1 = 42$, ocena variance pa $s_1^2 = 106$. Enako smo v drugem sestoju na slučajnosten način izbrali $n_2 = 100$ dreves in tudi zanj ocenili z vzorcem aritmetično sredino premerov $\bar{y}_2 = 38,5$ cm in ocenili varianco $s_2^2 = 90$. Če te podatke vnesemo v obrazec 11.29, dobimo

$$\frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{42 - 38,5}{\sqrt{\frac{106}{100} + \frac{90}{100}}} = +2,5 = z$$

Če dobljeno vrednost $|z| = 2,5$ primerjamo s kritičnimi vrednostmi $z_{0,05} = 1,96$, $z_{0,01} = 2,58$ in $z_{0,001} = 3,29$, zaključimo, da je razlika v povprečnem premeru v prsni višini med obema sestojema značilno različna in sicer s tveganjem $\alpha = 0,05$.

11.24

Z vzorcem $n_1 = 200$ dreves v prvem bukovem sestoju smo ocenili, da je odstotek žagarske hlodovine v prvem sestoju $p_1\% = 38\%$. Z vzorcem $n_2 = 200$ dreves v drugem bukovem sestoju pa smo ugotovili, da je ocena za odstotek žagarske hlodovine v drugem sestoju $p_2\% = 43\%$. Preskusiti je treba, ali je razlika v kvaliteti proučevanih sestojev značilna.

Če vstavimo rezultate vzorcev v obrazec 11.30, dobimo:

$$\frac{p_1\% - p_2\%}{\sqrt{\frac{p_1\%(100-p_1\%)}{n_1-1} + \frac{p_2\%(100-p_2\%)}{n_2-1}}} = \frac{38 - 43}{\sqrt{\frac{38(100-38)}{200-1} + \frac{43(100-43)}{200-1}}} = -1,02$$

Ker je absolutna vrednost izračunanega $z = 1,02$ manjša kot $z_{0,05} = 1,96$, zaključimo, da razlike v kvaliteti niso značilne.

11.25

Za smrekov sestoj 1 je iz vzorca $n_1 = 100$ dreves ocenjeni standardni odklon v premeru dreves v prsni višini $s_1 = 6$ cm. Za smrekov sestoj 2 pa je iz vzorca $n_2 = 200$ dreves ocenjeni standardni odklon za premere v prsni višini $s_2 = 8$ cm. Značilnost razlik med standardnima odklonoma preskusimo z obrazcem 11.31

$$\frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} = \frac{6 - 8}{\sqrt{\frac{6^2}{2 \cdot 100} + \frac{8^2}{2 \cdot 200}}} = -3,43 = z$$

Ker je absolutna vrednost izračunanega $z = 3,43$ večja kot $z_{0,001} = 3,29$, sklepamo, da so razlike v homogenosti visoko značilne in sicer s tveganjem $\alpha = 0,001$.



• DK 634.92:31

Naslov dela:
STATISTIČNE METODE V GOZDARSTVU

Obseg: 255 strani A-4

Naklada: 800 izvodov

Tehnika: Multilith

Izšlo septembra 1961

Avtorjev naslov:
Dr. Marjan Blejec
izr. univ. profesor
Ljubljana, Prešernova 13

Izdala
Fakulteta za agronomijo, gozdarstvo
in veterino
Univerze v Ljubljani

Založila
Univerzitetna založba
v Ljubljani

Razmnožila
Tiskarna Zavoda za statistiko
v Ljubljani

227.)

842.6
7600
129 x 200
1032003
309600

92810
1030

5160
516
1626

412
18582
5160
30960
466702
5676
43030
515408

NARODNA IN UNIVERZITETNA
KNJIŽNICA



00000020433

