Ekon. fak. 17. 5. 67

DR. MARIJAN BLEJEC

STATISTIČNE METODE

Z UPORABO V TEHNIKI

(Skripta)

Izdala in založila Ekonomska fakulteta v Ljubljani
Inštitut za statistiko in operacijsko raziskovanje
Ljubljana 1967

DR. MARIJAN BLEJEC

S T A T I S T I Č N E M E T O D E

Z UPORABO V TEHNIKI

S k r i p t a

po predavanjih na seminarju iz Statističnih metod
v Železarni Jesenice v zimskem semestru 1966/1967

Izdala in založila Ekonomska fakulteta v Ljubljani
Institut za statistiko in operacijsko raziskovanje

K A Z A L O

Stran

Stran

Stran

Stran

Stran

Stran

0. ELEMENTI STATISTIČNEGA PROUČEVANJA

0. 1. Statistika

Statistiko opredelimo kot vedo, ki s kvantitativnim proučevanjem množičnih pojavov, s specifičnimi metodami odkriva zakonitosti množičnega pojavljanja in podaja kvalitativno sliko pojava.

Po tej definiciji so področje statističnega proučevanja - množični pojavi. Množični pojavi so vsi pojavi, ki v času in prostoru množično nastopajo. Tako je množičen pojav artikel, proizvajan na določenem stroju, ker nastaja v velikem številu. Enako je množičen pojav okvara v določenem sistemu proizvodnje, ker more okvara v daljšem razdobju nastopiti večkrat. Pojavov, ki v času in prostoru množično nastopajo, moremo na najrazličnejših področjih najti obilo, zato je tudi s statistiko možno proučevati najrazličnejše pojave. Tipičen primer, v katerem umetno ustvarimo pogoje množičnosti in s tem pogoje za statistično proučevanje pa je eksperimentiranje, kjer po samem planu poskuse pod danimi pogoji ponavljamo, da dobimo s tem osnovo za upravičenost in možnost uporabe statistike pri eksperimentiranju. Statistika s svojstvenimi metodami, ki v glavnem bazirajo na kvantitativnem proučevanju analizira množične pojave. V kakšni obliki s statistiko kvantitativno podajamo kvalitativno sliko pojava naj služi najenostavnejši kazalec kvalitete - delež izmeta. Delež izmeta dobimo z enostavnim kvantitativnim pristopom, sortiranjem in preštevanjem defektnih artiklov. Kljub temu, da je ta kazalec kvantitativen, pa podaja kvalitativno stran proizvodnje - kakovost proizvodnje na določenem stroju. Enako daje kazalec o poprečni produktivnosti ali o variabilnosti produktivnosti zelo pomembne kvalitativne podatke o produktivnosti dela.

Statistične metode bazirajo na kvantitativnem proučevanju specifičnih kazalcev, ki jih izračunamo iz osnovnih podatkov za posamezne elemente oziroma enote množičnega pojava. Zato se statistične metode v mnogočem oslanjajo na matematiko ali na prenos določenih pojmov iz matematike, ki so modificirani za proučevanje množičnih pojavov in manifestacij, ki se pojavljajo pri množičnih pojavih.

0. 2. Populacija, enota, znak, parameter

Statistika proučuje množične pojave. Vsak element proučevanja npr. artikel, poskus, nesreča pri delu itd., ki je predmet statističnega proučevanja imenujemo s skupnim imenom enota. Skupnost enot, ki so v konkretnem primeru predmet proučevanja oziroma raziskave imenujemo populacija. Populacijo sestavlja npr. skupnost vseh artiklov - enot proizvedenih v določenem času na določenem stroju. Enote se razlikujejo med seboj po nebroj značilnostih. Značilnosti, ki so pomembne v določenem statističnem proučevanju in jih opazujemo, imenujemo statistične znake. Tako je statistični znak kvaliteta artikla, vsebnost ogljika v jeklu v posameznih šaržah, žilavost preizkušamo za kotlovsko pločevino, temperaturo pri ohlajanju šarž ipd. Znaki so značilnosti enot, značilnosti populacij pa so parametri. Parametri so običajno izvedeni iz znakov za enote populacije. Tako je npr. poprečen odstotek izmeta v proizvodnji določenega artikla parametrer, ki je izveden iz znaka: uporabost za posamezne artikle. Enako je kazalec o variabilnosti vsebnosti po šaržah parameter za populacijo šarž iz ene peči in je izveden iz vsebnosti ogljika za posamezne šarže. Statistična deskripcija in analiza se ukvarja z izračunavanjem, primerjanjem in analiziranjem parametrov za različne populacije.

0. 3. Statistične enote

V industriji in tehniki na splošno so artikli predmet masovnega pojavljanja. Zato je eden izmed najpogostejših enot proučevanja v industriji proizveden artikel. Ker so v proizvodnem procesu možni izredni dogodki, ki vplivajo na proizvodni proces, so dostikrat enote proučevanja dogodki, kot so npr. prekinitev dela stroja zaradi določene okvare. Vse okvare na določenem stroju ali na skupini strojev moremo združiti v skupnost enot vseh okvar v določenem razdobju. Pri proučevanju delov in procesov, pri katerem proučujemo strukture dela v določenih razdobjih je posebno enota časoven moment, ki se združujejo v populacijo časovni razmak, za katerega proučujemo strukturo časa. V preskusništvu je običjana enota opazevanja dogodek ali proces, ki ga sestavimo umetno za potrebe določene raziskave. Te enote imenujemo eksperiment ali poskus.

0. 4. Statistični znaki

Enotam dajo vsebino značilnosti, ki so opisane z znaki. Znaki so po vsebini: faktorialni in rezultativni. Pri vzročni odvisnosti imajo faktorialni znaki vlogo neodvisnih spremenljivk, rezultativni pa vlogo odvisne spremenljivke. Pri statističnih raziskavah s področja eksperimentiranja, vrednosti faktorialnih znakov navadno vnaprej določamo ali kontrolirano variiramo in proučujemo vpliv faktorjev ali njihovih sprememb na rezultativen znak.

Faktorialne znake ločimo v eksperimentalnem delu na tri tehnično pomembne grupe: Medtem ko so eni faktorji opredeljivi in je možno njihove vrednosti točno bodisi zavestno določiti ali pa ugotoviti, je druga grupa faktorjev neopredeljiva in nedoločljiva in se manifestira v skupini slučajnostnih vplivov. Opredeljive faktorje najprej delimo na opredeljive pomembne in opredeljive nepomembne. Medtem ko so opredeljivi bistveni faktorji predmet raziskovanja in razlikujemo njihov vpliv na rezultativne znake, opredeljivi nepomembni znaki v bistvu motijo raziskovanje. Zato skušamo po možnosti njihov vpliv, če že ne eliminirati, vsaj držati na stalnem nivoju, da je njih vpliv konstanten.

0. 5. Numerični in atributivni znaki

Tehnično znake delimo na numerične in atributivne. Vrednosti numeričnih znakov izražamo številčno, vrednosti atributivnih pa opisno. Tako je atributiven znak vsebnost posameznih elementov v šaržah, če izražamo vsebnost v odstotkih. Numeričen znak je tudi časovna produktivnost posameznih šarž itd. Numerični znaki so ali zvezni ali nezvezni, glede na to, kako so izrazljive vrednosti znakov. Tako je npr. vsebnost posameznih elementov zvezen znak, ker more teoretično zavzeti delež elementa v šarži vse vrednosti na določenem razmaku. Nezvezen znak pa je npr. število napak na izdelku, število pulzacij preskušanca do preloma. Število možnih vrednosti za zvezne znake je neomejeno. Za praktične potrebe pa jih določamo in merimo le do določene natančnosti, zato sorodne vrednosti združujemo v razrede. Razred ima svojo spodnjo in zgornjo mejo, svojo širino razreda in sredino razreda. Spodnja meja razreda je vrednost, pod katero ni nobene vrednosti iz razreda. Zgornja meja pa je vrednost, nad katero ni nobene vrednosti iz razreda. Če zaznamujemo z $x_{k,}$ min in x_{k},max spodnjo in zgornjo mejo razreda, ~~je~~sta širina razreda i_k in srednja razreda x_k dani z obrazcema

$$i_k = x_{k,max} - x_{k,min} \quad ; \quad x_k = \frac{x_{k,min} + x_{k,max}}{2} \qquad (1)$$

Osnovne podatke zaokrožujemo na dva načina: na najbližjo in na najnižjo zaokroževano vrednost. Tako vsebuje trdnost v kp/mm2 zaokrožena na 16o po prvem načinu vse vrednosti v razmaku od 159,5 do 16o,5, po drugem načinu pa vse vrednosti od 16o,o do 161,o. Od načina osnovnega zaokroževanja je odvisno nadaljnje formiranje razredov. Če tvorimo razrede višje stopnje po pet zaokroženih vrednosti, dobimo po prvem načinu razred 159,5 do 164,5 s sredino razreda 162, po drugem načinu pa razred 16o,o do 165,o s sredino 162,5. V prvem primeru meje niso cele vrednosti, je pa cela vrednost sredina razreda, v drugem pa obratno. V splošnem pa je drugi način zaokroževanja in formiranja razredov boljši in logičnejši, čeprav ni neke vsebinske razlike med obema. Paziti moramo samo, da je znan uporabljeni način zaokroževanja zaradi nadaljnje interpretacije in računov.

Vrednosti atributivnih znakov se dajo izraziti samo opisno. ~~Tipičen~~ atributiven znak je kvaliteta artikla: uporaben, neuporaben; ravnanje žice: ročno, strojno; barva artikla itd., vrednosti za atributivne znake more biti najmanj dve. Navzgor pa število vrednosti ni omejeno. Za atributiven znak: vrsta artikla, je število raznovrstnih artiklov, ki jih more proizvajati eno podjetje, vrsta industrije ali cela industrija, zelo veliko.

Vrednosti atributivnih znakov po sorodnosti grupiramo - klasificiramo v grupe. Tehnično zelo uporabna je decimalna klasifikacija, kateri v grupe višje vrste združujemo po največ deset grup nižje vrste, da jih moremo obeležiti s ciframi od o do 9. Decimalna klasifikacija zelo nazorno prikaže pripadnost posameznih vrednosti (artiklov, surovin) itd. v grupe posameznih stopenj.

Vrednostim atributivnih znakov s samo dvema vrednostima, dostikrat za uporabo določenih metod, ki so prirejene le za analizo numeričnih znakov, pripišemo numerični vrednosti o, drugi pa 1. Tako znaku: za kvaliteto artiklov vrednosti znaka neuporaben, pripišemo vrednost 1, vrednosti uporaben pa vrednost o.

0. 6. Intervalen, ordinalen in nominalen značaj znakov

Za statistično analizo je zelo pomemben intervalen, ordinalen in nominalen značaj znakov. Intervalno lastnost imajo znaki, za katere je možno izračunati razliko med dvema vrednostima znakov $x_2 - x_1 = d$. Ordinalno lastnost imajo znaki, za katere je možno ugotoviti po določenem logičnem kriteriju vrstni red vrednosti oziroma enot, tako da je za dve poljubni vrednosti možno ugotoviti, katera je v tem logičnem redu pred drugo. Ordinalnost znakov s simboli nakažemo $x_2 > x_1$. Nominalen značaj znaka je še za eno stopnjo nižji od ordinalnega. Nominalen značaj znakov zagotavlja le možnost razlikovanja vrednosti $x_2 \neq x$. Ta lastnost daje možnost, da enote razdelimo v grupe. Medtem, ko imajo znaki z intervalnim značajem (tak značaj imajo numerični znaki) implicitno tudi ordinalen in nominalni značaj, imajo znaki, ki imajo ordinalni značaj implicitno tudi nominalen značaj. Samo nominalni značaj imajo v splošnem atributivni znaki. Poznamo pa znake, ki imajo ordinalen, nimajo pa intervalnega značaja. Tako moremo artikle razporediti po lepoti obdelave, lesku ipd., čeprav lesk in lepoto obdelave ne moremo numerično izraziti.

0. 7. Populacija

Skupnost enot, ki sestavljajo statistično populacijo in je predmet statistične analize, mora biti nedvoumno opredeljena z opredeljujočimi pogoji, ki določajo, kateri pojavi - enote spadajo v populacijo in kateri ne spadajo. Enote, ki zadoščajo opredeljujočim pogojem, so predmet opazovanja, medtem ko enote, ki tem pogojem ne zadoščajo, ne spadajo v populacijo.

Populacijo je treba opredeliti iz dveh razlogov. Z opredeljujočimi pogoji je določeno, katere enote sestavljajo populacijo, obenem pa pojasnjujemo, kaj populacija vsebuje. Opredeljujoči pogoji so do neke mere že parametri populacije, ker so informacija o populaciji, ne pa o posameznih enotah. Tako opredeljujoči pogoji da spadajo v populacijo, ingoti, proizvedeni v določenem razdobju na določeni peči pod določenimi pogoji, na eni strani opredeljujejo populacijo, po drugi strani pa dajejo karakteristike - parametre populacije.

Populacije delimo na populacije realnih enot in populacije dogodkov. Tako sestavlja populacijo skupnost pod določenimi pogoji proizvedenih artiklov, ali skupnost nesreč pri delu v določenem razdobju, skupnost prekinitev delovnega procesa itd.. Za raziskovalno delo so pomembne hipotetične populacije. Dvesto artiklov, ki smo jih proizvedli na določenem stroju v danem dnevu, predstavlja populacijo dnevne proizvodnje. Pod določenimi predpostavkami pa smatramo dnevno proizvodnjo za del artiklov iz hipotetične populacije vseh možnih artiklov, proizvedenimi pod enakimi pogoji. Hipotetične populacije ne moremo nikdar ostvariti, ker je miselna konstrukcija. Če hočemo s procentom izmeta dvesto artiklov izraziti kvaliteto proizvedene partije, je ta rezultat parameter populacije 200 artiklov. Ta procent izmeta pa je samo ocena za pravi delež izmeta v hipotetični populaciji vseh možnih artiklov, proizvedenih pod enakimi pogoji in je ocena karakteristike stroja. Še jasneje je koncept hipotetične populacije razumljiv na primeru eksperimentiranja. Trideset poskusov, ki jih izvedemo pod enakimi pogoji, sami zase ne pomenijo mnogo, dokler dobljenih rezultatov ne posplošimo, t.j. prenesemo na hipotetično populacijo vseh možnih poskusov pod enakimi pogoji.

Populacije so lahko zvezne in nezvezne. Kolobar žice moremo smatrati za populacijo. Ta populacija je zvezna in enote populacije niso vnaprej dane. Za zvezne populacije je treba enote šele definirati. Kolobar žice moremo razstaviti v populacijo s končnim številom enot, če ga razrežemo na 5 centimetrov dolge koščke žice, ki v nadaljnjem predstavljajo enoto in sestavljajo nezvezno populacijo. Moremo pa vzeti, da je kolobar sestavljen iz neomejenega števila enot, če definiramo enoto tako, da je na enodimenzionalnem traku oziroma daljici dolžine kolobarja vsaka točka lahko začetek petcentimeterskega koščka žice. Posamezne enote tako formirane populacije so definirane s točkami žice na kolobarju.

V praksi proučujemo veliko zveznih populacij, za katere je seveda treba pred proučevanjem definirati enote.

Ena izmed tipičnih zveznih populacij, ki je sicer nenavadna, vendar v raziskavah struktur procesov fundamentalna populacija, je časovni razmak. Časovni razmak delovnega časa v razdobju enega meseca predstavlja populacijo momentov. Znaki ob posameznih momentih so produktivno delo stroja ali delavca, faza dela ipd.

Za razliko od zveznih populacij, za katere je treba šele definirati enote proučevanja, so enote <u>nezveznih</u> populacij sestavljene iz diskretnih enot ali dogodkov. Tako je nezvezna populacija populacija artiklov, šarž, delavcev itd.

0. 8. Subpopulacija

Populacija je opredeljena z opredeljujočimi pogoji, s katerimi je določeno, katere pojavi so enote populacije, ki jo proučujemo in katere ne. Če opredeljujočim pogojem populacije dodamo še nov pogoj, skupnost enot, ki zadošča temu novemu pogoju, po definiciji sestavlja novo populacijo. Ker pa so vse enote nove populacije istočasno enote prvotne populacije, jo imenujemo subpopulacijo. Če je dodatni pogoj tak, da nobena enota osnovne populacije ne zadošča temu pogoju, je subpopulacija prazna. Če po različnih vrednostih določenega znaka razdelimo populacijo v več subpopulacij tako, da je vsaka enota osnovne populacije istočasno enota ene izmed subpopulacij pravimo, da je razdelitev na subpopulacije kompletna. Vzemimo kot primer dnevno proizvodnjo določenega artikla. Če opredeljujočim pogojem, da gre za proizvodnjo določenega artikla dane tovarne za določen dan pridamo nov pogoj, da je artikel uporaben, je nova populacija, ki sestoji iz uporabnih artiklov, subpopulacija osnovne populacije vseh proizvedenih artiklov.

V določeni raziskavi izvedeni poskusi pod enakimi pogoji so subpopulacija iz hipotetične populacije vseh možnih poskusov, ki bi jih izvedli pod enakimi pogoji. Stvarno izvedeni poskusi so subpopulacija hipotetične populacije, ker je osnovnim pogojem: poskusi izvedeni pod enakimi pogoji, dodan nov pogoj; stvarno izvedeni poskusi, kateremu pa vse enote hipotetične populacije ne ustrezajo. Stvarno realizirane enote iz hipotetičnih populacij so v vsakem in ne samo v tem primeru subpopulacije.

0. 9. Parametri

S parametri opisujemo značilnosti populacij. Z njimi kvantitativno podajamo kvantitativne in kvalitativne značilnosti in zakonitosti populacij. Najenostavnejše parametre dobimo z preštevanjem enot (strukture), s seštevanjem vrednosti znakov (agregati). Kot skupni kazalec, dobimo tako obseg populacije npr. število artiklov proizvedenih na stroju ali s seštevanjem vrednosti proizvodnje v danem mesecu ipd. S preštevanjem in seštevanjem po delnih populacijah, ki jih dobimo, če razdelimo populacijo po določenih znakih, dobimo vpogled v sestav populacije. Tako dá npr. seštevanje po grupah uporabnosti artiklov vpogled v sestav proizvodnje po kvaliteti.

Iz osnovnih vrednosti za posamezne enote ali iz števila enot in agregatov za delne populacije izračunavamo najrazličnejše parametre, ki kažejo na jakost individualnih vplivo (mere centralne tendence) jakost individualnih vplivov (mere variacije) jakost in zakonitosti neodvisnosti (regresijska in korelacijska analiza) in podobno. S primerjavo teh parametrov za različne populacije ali dele populacij, analiziramo odnose in vplive, ki so važni za množične pojave.

1. FREKVENČNE PORAZDELITVE

1. 1. Frekvenčna porazdelitev

Z raziskavo dobljeni podatki so nepregledni, zato jih moramo urediti. Za populacije z večjim številom enot zato urejamo vrednosti numeričnih znakov frekvenčne porazdelitve. V frekvenčni porazdelitvi so vrednosti numeričnega znaka grupirane v razrede, za vsak razred pa je v frekvenčni porazdelitvi dana frekvenca - število enot v razredu. Frekvenčna porazdelitev daje sliko^x takrat, če so širine razredov v frekvenčni porazdelitvi enake.

Frekvenčna porazdelitev je statistična vrsta, ki sestoji iz zaporedja razredov in ustreznih frekvenc. Vsak razred ima svojo spodnjo $x_{k,min}$ in svojo zgornjo $x_{k,max}$ mejo,

širino razreda $i_k = x_{k,max} - x_{k,min}$ (1),

sredino razreda $x_k = (x_{k,min} + x_{k,max})/2$ (2),

ki je reprezentant vrednosti v razredu, frekvenco f_k, ki pove, koliko enot iz populacije ima vrednosti v danem razredu k,

Relativna frekvenca f_k^o

$$f_k^o = f_k / N \qquad (3),$$

pokaže koliki del celotne populacije ima vrednosti v razredu k.

x o variabilnosti proučevanega pojava. Najbolj neposredna je ta slika

Relativno frekvenco izražamo tudi v odstotkih

$$f_k\% = 100 . f_k / N \qquad (4)$$

Gostota relativne frekvence

$$\varphi_k = f_k / N . i_k = f_k^o / N \qquad (5)$$

pove, koliki delež celotne populacije odpade na enotin razmak znaka x. Iz obrazca 5 dobimo, da je

$$f_k = N . i_k . \varphi_k \qquad (6)$$

frekvenca v danem razredu odvisna od velikosti populacije, širine razreda, ki je tehničen faktor in gostote razreda, ki je povezana z zakonitostjo pojavljanja frekvence.

V frekvenčni porazdelitvi z ozkimi razredi, je število enot v posameznih razredih majhno, na frekvence pa zato močno vplivajo slučajnostni vplivi. Zato taka frekvenčna porazdelitev, kljub temu, da dá podrobno sliko o pojavljanju posameznih vrednosti, ni pregledna.

Za frekvenčne porazdelitve s širokimi razredi je število razredov majhno, značilnosti gostitve pa zabrisane. Zato je treba pri sestavljanju frekvenčne porazdelitve paziti, da je širina in z njo število razredov v skladu z velikostjo populacije in razmakom, v katerem variirajo vrednosti populacije. Eno izmed pravil nakazuje, da je primerno število razredov K približno enako kvadratnemu korenu iz obsega populacije N

$$K = \sqrt{N} \qquad (7)$$

a) Osnovni podatki za % Si v surovem železu na II. visoki peči (Vir: Zec, Behmen Razrada sistema stat. praćenja kvaliteta u željezarama: IMIZ)

vzorci po n= 5 meritev v 26 dneh v danem mesecu

1,36	1,41	1,38	o,49	1,41	1,o8	1,13	1,25	1,36	o,94
o,94	1,97	1,22	1,17	1,64	o,85	1,27	o,94	1,74	1,17
1,72	1,22	1,31	1,33	1,6o	1,69	1,5o	o,94	1,5o	1,18
1,17	1,3o	1,31	1,18	1,13	o,91	1,o3	1,17	1,5o	1,73
o,89	1,47	1,22	o,87	1,32	1,13	1,22	1,88	o,94	o,86
o,71	1,8o	o,94	o,91	1,5o	1,5o	o,94	1,5o	1,18	1,22
o,94	1,o9	2,01	1,11	1,o3	1,o3	1,26	1,5o	o,89	1,32
1,36	1,31	1,25	1,36	1,28	1,32	1,22	1,58	o,74	1,1o
1,5o	o,94	1,47	1,46	1 69	1,32	1,32	1,33	1,17	1,5o

1,41	1,27	1,5o	1,36	1,69	o,89	o,94	o,79	o,74	1,o4
1,13	o,74	1,o8	o,94	1,6o	1,13	1,22	1,32	1,o8	o,75
1,o9	o,94	1,88	o,7o	o,94	o,47	1,23	1,13	o,99	1,o3
o,74	o,99	o,94	o,74	1,49	1,55	1,o3	1,69	1,41	1,41

Za primer formiranja frekvenčne porazdelitve vzemimo % Si v surovem železu na določeni visoki peči. Iz N = 13o vzorcev smo dobili naslednje osnovne podatke o odstotku Si: gleja.

Ker so podatki zaokroženi na eno decimalno mesto, je najmanjša širina razreda o,1, druge možne širine pa mnogokratniki od o,1 %. Če po obrazcu izračunamo optimalno število razredov za N = 13o je $K = \sqrt{130} \doteq 11$

Vrednosti variirajo med o,4 % do 2,o %, zato dobimo pri različnih širinah naslednje število razredov:

i	o,1 %	o,2 %	o,4 %
K	17	9	5

Najbolj se približamo optimalnemu številu razredov 11, če vzamemo i = o,2 %.

Zaradi primerjave so izdelane vse tri frekvenčne porazdelitve z i = o,1, i = o,2, i = o,4. Za z i = o,1 je nakazano sestavljanje frekvenčne porazdelitve s črtkanjem

Resnično dobimo pričakovano sliko. Frekvenčna porazdelitev z i = o,1 je še preveč pod vplivom slučajnih odstopanj in frekvence še ne kažejo v dovoljni meri značilnosti gostitve, frekvenčna porazdelitev z i = o,4 pa ima preširoke razrede in je zato zakonitost zabrisana. Najugodnejšo od vseh treh porazdelitev daje frekvenčna porazdelitev s širino razredov i = o,2 %, ki nakazuje sistematično večanje frekvenc do določenega maksimuma v razredu 1,2 % - 1,3 %, nato pa zakonito manjšanje frekvenc. Zanjo so razen frekvenc izračunane še relativne frekvence, prikazani z deleži in odstotnimi deleži.

b) Formiranje frekvenčne porazdelitve za i - o,1% s črtkanjem

% Si		f_k
o,4	II	2
o,5		-
o,6		-
o,7	卌 IIII	9
o,8	卌 I	6
o,9	卌 卌 卌 III	18
1,o	卌 卌 I	11
1,1	卌 卌 卌 I	16
1,2	卌 卌 卌	14
1,3	卌 卌 卌 III	18
1,4	卌 卌	9
1,5	卌 卌 II	12
1,6	卌 II	7
1,7	III	3
1,8	III	3
1,9	I	1
2,o	I	1
		n = 13o

c) Frekvenčne porazdelitve z i = o,2 % in i = o,4 %

% Si	f	f^o	$f^{\%}$
o,4 - o,5	2	,o15	1,5
o,6 - o,7	9	,o69	6,9
o,8 - o,9	24	,185	18,5
1,o - 1,1	27	,2o8	2o,8
1,2 - 1,3	32	,246	24,5
1,4 - 1,5	21	,162	16,2
1,6 - 1,7	1o	,o77	7,7
1,8 - 1,9	4	,o31	3,1
2,o - 2,1	1	,oo8	,,8
	n = 13o	1,ooo	1oo,o

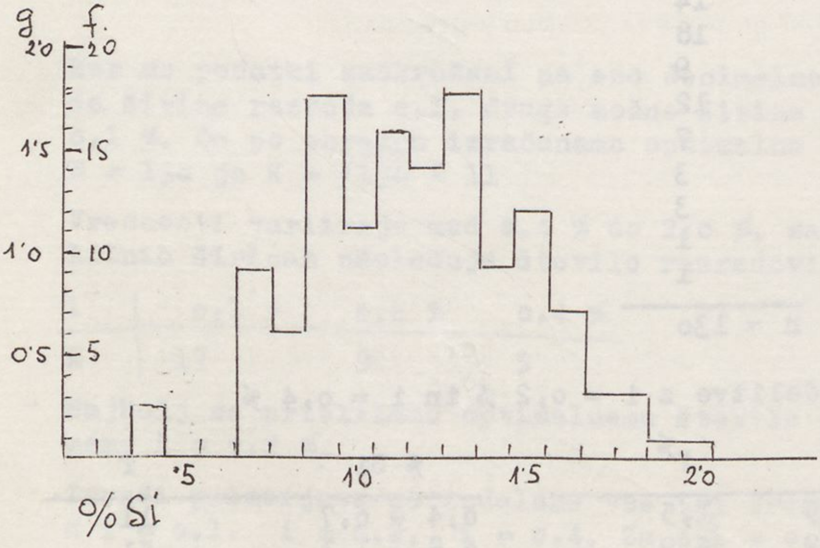
% Si	f
o,4 - o,7	11
o,8 - 1,1	51
1,2 - 1,5	53
1,6 - 1,9	14
2,o - 2,3	1
	n = 13o

1. 2. Histogram

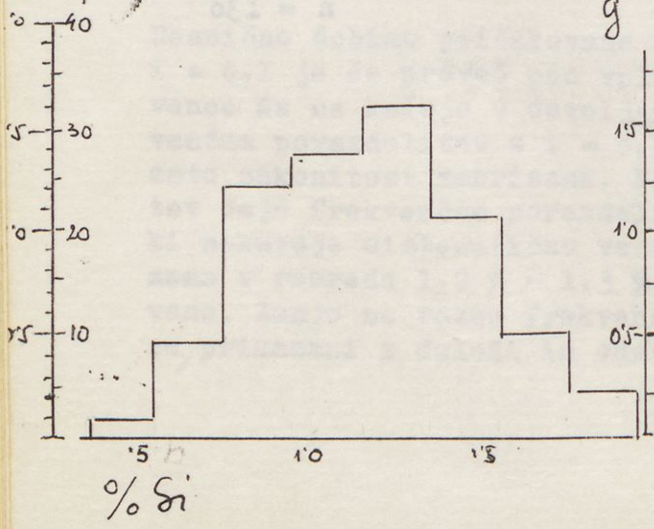
Frekvenčne porazdelitve prikazujemo s histogrami ali poligoni. S histogramom prikazujemo frekvenčno porazdelitev z nizom stolpcev tako, da frekvenco za vsak razred ponazorimo s stolpcem, ki je visok v sorazmerju s frekvenco v razredu. Ploščina pod stopničasto linijo, ki je dana s konturami stolpcev in abscisno osjo, je v sorazmerju z obsegom populacije.

Za primer so v slikah a, b, c prikazani histogrami za vse tri variante vsebnosti Si v 13o šaržah.

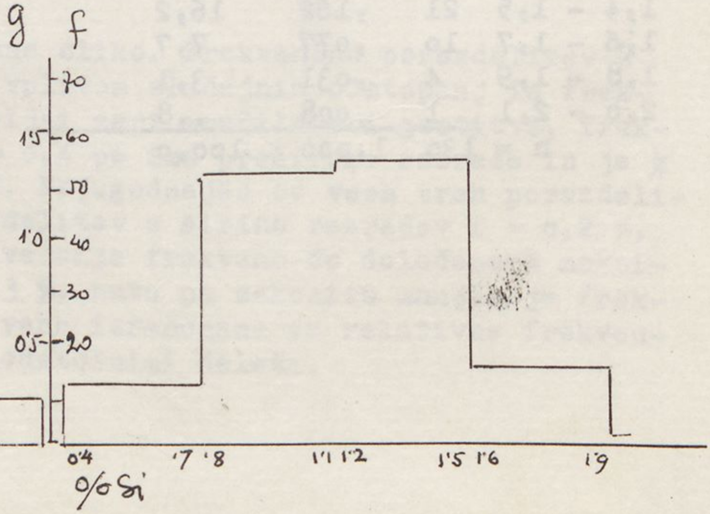
Slika 1. Vsebnost Si v odstotkih v surovem železu v 13o šaržah



a



b



c

Iz histogramov za različne širine razredov je bolj kot iz same frekvenčne porazdelitve opazno, da najboljše oriše zakonitost pojavljanja frekvenčne porazdelitve z i = o,2 %.

Višine stolpcev so proporcionalne frekvenci v posameznih razredih le, če so širine razredov enake. V drugih primerih pa je potrebno, da rišemo širine stolpcev v sorazmerju s širino razredov, višine stolpcev pa v sorazmerju z gostoto frekvence g_k. Ploščine teh stolpcev - pravokotnikov so glede na zvezo $f_k = i_k g_k$ proporcionalne frekvencam v razredih. V zgornjih slikah so sicer zaradi enake širine razredov v posameznih frekvenčnih porazdelitvah višine stolpcev proporcionalne frekvencam, moremo jih pa risati na enotno skalo gostot frekvenc. S histogrami prikazujemo frekvenčne porazdelitve absolutnih frekvenc in frekvenčne porazdelitve relativnih frekvenc.

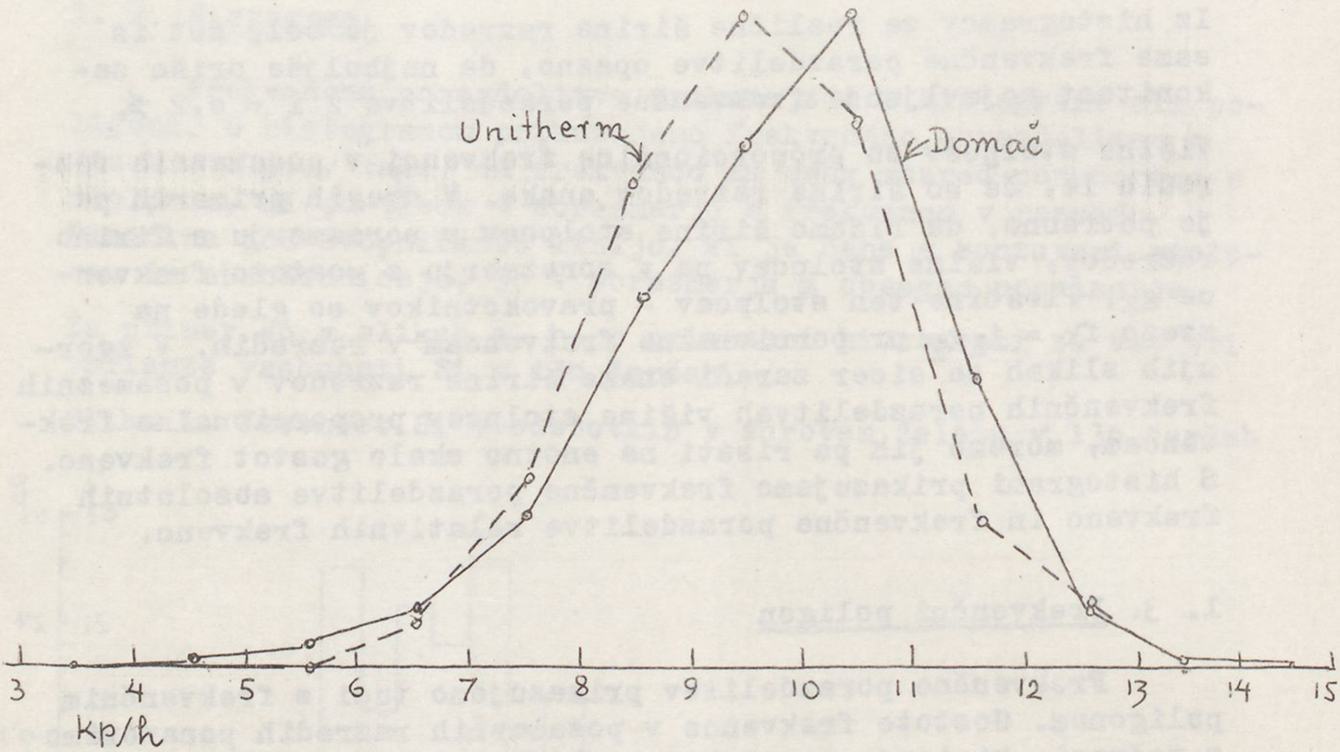
1. 3. Frekvenčni poligon

Frekvenčno porazdelitev prikazujemo tudi s frekvenčnim poligonom. Gostote frekvence v posameznih razredih ponazorimo s točkami, ki imajo za absciso sredino razreda, ordinata pa je sorazmerna gostoti frekvence ali gostoti z abscisno osjo zaključen tako, da v razredih izven porazdelitve vzamemo, da je gostota enaka o, dobimo celo boljšo sliko in predstavo o razporeditvi vrednosti kot s histogramom. Medtem ko v histogramu predpostavljamo, da je gostota frekvence v razmaku posameznega razreda konstantna, poligon realneje ponazarja stvarno gostoto na posameznih odsekih razredov.

Če so razredi enaki, moremo namesto gostote frekvence upoštevati frekvence, ker sta si proporcionalni.

Enako kot histograme, moremo tudi poligone risati za absolutne in relativne frekvenčne porazdelitve.

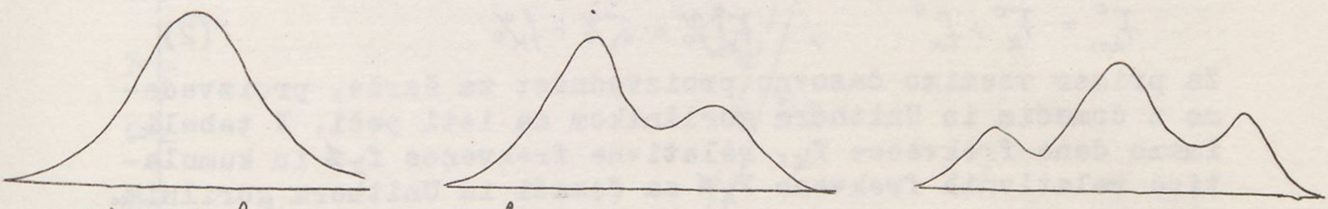
V sliki 1 sta prikazani s poligonom frekvenčni porazdelitvi frekvenc za časovno produktivnost doseženo z domačimi in Unitherm gorilniki na isti peči iz tabele v 1. 6. Da odstranimo vpliv različnega števila šarž za oba gorilnika so za primerjavo poligonov primernejši poligoni relativnih frekvenc.



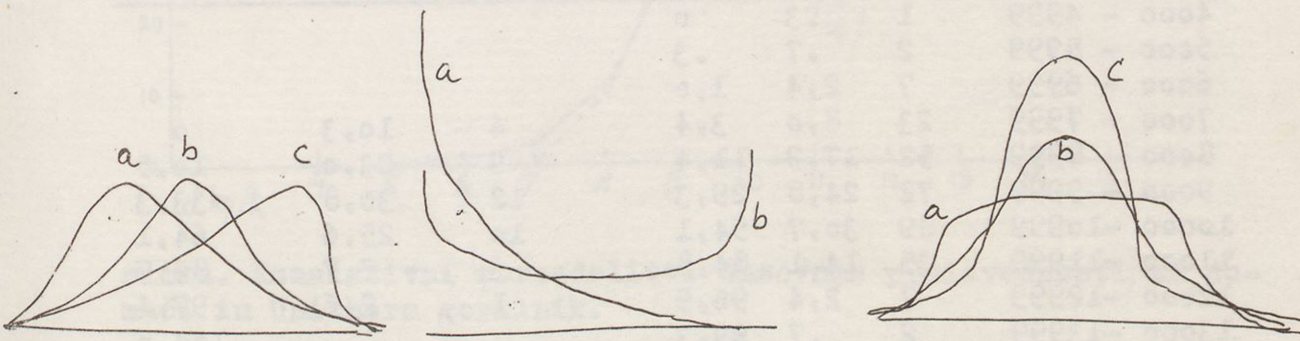
slika 1. Frekvenčni poligoni relativnih frekvenc za časovno proizvodnost domačega in Unitherm gorilnika.

1. 4. Oblike frekvenčnih porazdelitev

Zaradi različnih zakonitosti vplivanja faktorjev dobimo v konkretnih primerih različne oblike frekvenčnih porazdelitev. Tako ločimo glede na simetrijo simetrične porazdelitve od asimetričnih v levo in desno, koničaste od sploščenih, uminodalne od bimodalnih in polimodalnih, J in U porazdelitve, pravokotne in zvonaste.



unimodalna bimodalna polimodalna



a) asimetrična v desno
b) simetrična
c) asimetrična v levo

a) J-porazdelitev
b) U-porazdelitev

a) sploščena
b) normalna
c) koničasta

1\. 5. <u>Kumulativne frekvenčne porazdelitve</u>

S postopnim prištevanjem frekvenc dobimo po obrazcu

$$F_{k+1} = F_k + f_k \tag{1}$$

iz vrste frekvenc f_k, vrsto kumulativnih frekvenc F_k oziroma kumulativno frekvenčno porazdelitev. Pri tem vzamemo, da je po definiciji vrednost kumulativne frekvence v prvem razredu $F_1 = 0$.

Ker imajo relativne frekvence f_k^o oziroma $f_k\%$ enake lastnosti kot absolutne frekvence, moremo izračunavati tudi kumulative relativnih frekvenc F_k^o oziroma $F_k\%$ in to po obrazcih

$$F^o_{k+1} = F^o_k + f^o_k \quad ; \quad F_{k+1}\% = F_k\% + f_k\% \qquad (2)$$

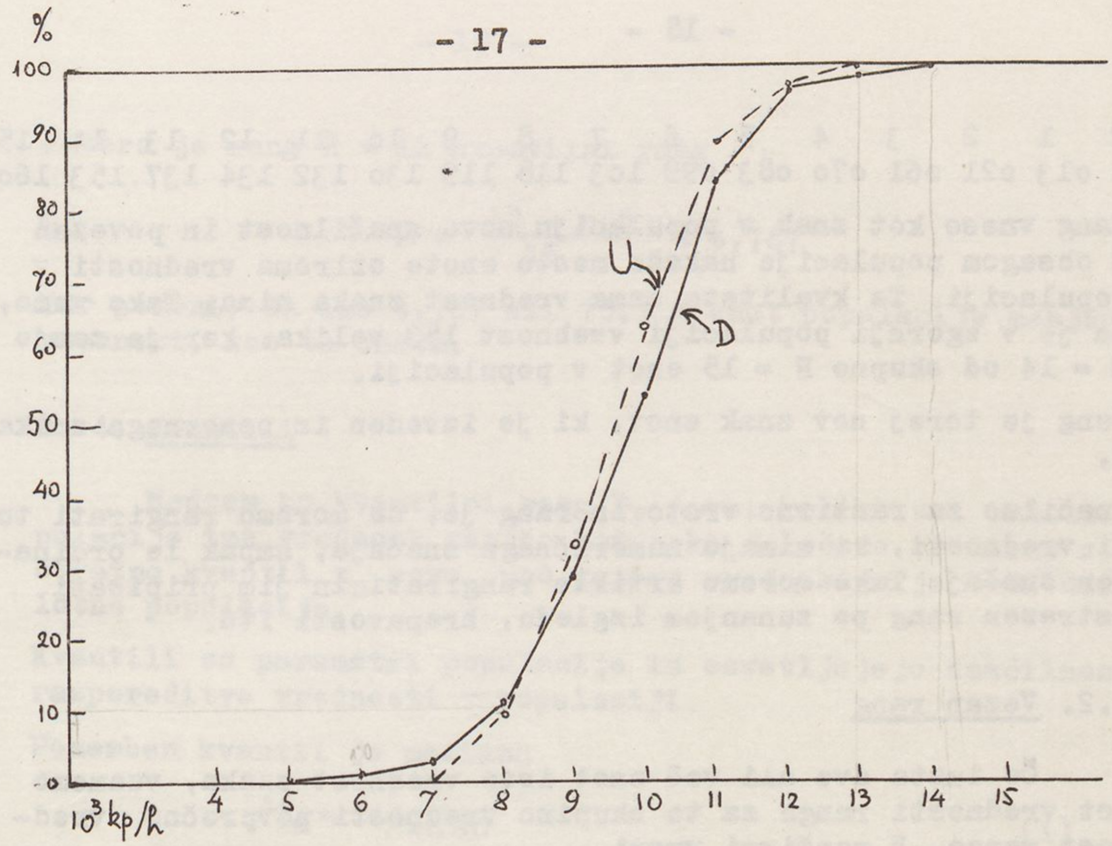
Za primer vzemimo časovno proizvodnost za šarže, proizvedene z domačim in Unitherm gorilnikom na isti peči. V tabeli imamo dane frekvence f_k, relativne frekvence $f_k\%$ in kumulative relativnih frekvenc $F_k\%$ za domači in Unitherm gorilnik.

časovna proizvodnost kp/h	gorilnik					
	domači			Unitherm		
	f_k	$f_k\%$	$F_k\%$	f_k	$f_k\%$	$F_k\%$
4000 - 4999	1	.3	0			
5000 - 5999	2	.7	.3			
6000 - 6999	7	2,4	1,0			
7000 - 7999	23	8,0	3,4	4	10,3	0
8000 - 8999	52	17,9	11,4	9	23,0	10,5
9000 - 9999	72	24,8	29,3	12	30,8	33,3
10000 -10999	89	30,7	54,1	10	25,6	64,1
11000 -11999	35	14,1	84,8	3	7,7	89,7
12000 -12999	7	2,4	96,9	1	2,6	97,4
13000 -13999	2	.7	99,3			100,0
14000 -14999			100,0			100,0
	290	100,0		39	100,0	100,0

1.6. Grafični prikaz kumulativnih frekvenčnih porazdelitev

Kumulativno porazdelitev absolutnih ali relativnih frekvenc prikažemo s točkami nad mejami razredov v oddaljenosti, ki je proporcionalna F_k ali $F_k\%$. Če te točke povežemo, dobimo naraščajočo črto, ki ponazarja kumulativno porazdelitev.

Za primer sta prikazani kumulativi relativnih frekvenc iz tabele. Čim večja je variabilnost pojava, tem položnejša je črta za kumulativo relativnih frekvenc in čim manjša je variabilnost, tem strmejša je črta. Za unimodalne zvonaste porazdelitve je kumulativa podobna idealizirani črki S.



slika. Kumulativni porazdelitvi časovnih proizvodnosti za domači in Unitherm gorilnik.

2. KVANTILI

2. 1. Rang in ranžirna vrsta

Množico numeričnih podatkov za določeno populacijo ali vzorec, urejeno po velikosti od najmanjšega do največjega, imenujemo ranžirno vrsto. Vsakemu členu v ranžirni vrsti priredimo zaporedno številko od 1 do N. Ta nov znak imenujemo rang.

Če vzamemo za primer vsebnost ogljika v N = 15 ploščah, je iz neurejenih podatkov o83 1o3 o13 132 o21 16o 119 o61 137 o7o 13o 134 o99 153 118 sestavljena ranžirna vrsta vsebnosti

R	1	2	3	4	5	6	7	8	9	1o	11	12	13	14	15
x	o13	o21	o61	o7o	o83	o99	1o3	118	119	13o	132	134	137	153	16o

Rang vnese kot znak v populacijo novo značilnost in povezan z obsegom populacije nakaže mesto enote oziroma vrednosti v populaciji. Te kvalitete sama vrednost znaka nima. Tako vemo, da je v zgornji populaciji vsebnost 153 velika, ker je zanjo R = 14 od skupno N = 15 enot v populaciji.

Rang je torej nov znak enot, ki je izveden iz osnovnega znaka x.

Značilno za ranžirno vrsto in rang je, da moremo rangirati tudi vrednosti, ki nimajo numeričnega značaja, ampak le ordinalen značaj. Tako moremo artikle rangirati in jim pripisati ustrezen rang po zunanjem izgledu, hrapavosti itd.

2.2. Vezan rang

Če imata dve ali več enot isto vrednost znaka, vzamemo kot vrednosti ranga za to skupino vrednosti povprečno vrednost ranga. V ranžirni vrsti

x 15 17 17 18 2o 22 22 22

na primer pripišemo vrednosti 17 vezan rang 2,5, ker je vrednost 17 na mestu z rangoma 2 in 3, vrednosti 22 vezan rang R = (6+7+8)/3 = 7, ker je vrednost 22 na mestu z rangi 6, 7, 8.

2.3. Kvantilni rang

Nazorneje kot rang, katerega moramo vezati s skupnim številom enot v populaciji, nakaže mesto enote v populaciji relativen rang, ki ga imenujemo kvantilni rang. Z njim razmak, na katerem se razvrste rangi, omejimo na razmak od o do 1. Zveza med rangom in kvantilnim rangom je dana prek obrazca

$$P_x = \frac{R_x - 0{,}5}{N} \qquad (1)$$

Tako je npr. za vsebnost ogljika 134 iz tabele v 2. 1, za katero je rang R = 12 kvantilni rang

$$P = \frac{12 - 0{,}5}{15} = 0{,}767$$

kar pomeni, da ima 0,767 ali 76,7 % enot populacije manjšo vsebnost, kot ta enota.

2. 4. Kvantili

Medtem ko kvantilni rang P_x pove, koliki del celotne populacije ima vrednost manjše kot neka določena vrednost x, obratno kvantil x_p pove, pod katero vrednostjo je P-ti del celotne populacije.

Kvantili so parametri populacije in osvetljujejo značilnosti razporeditve vrednosti v populaciji.

Pomemben kvantil je mediana

$$M_e = x_{P=0{\cdot}50} \tag{1}$$

ki pomeni vrednost, od katere je polovica enot populacije manjših kot $x_p = 0{,}50$ - mediana.

Podobno s kvartili

$$Q_1 = x_{P=0{\cdot}25} \quad ; \quad Q_2 = x_{P=0{,}50} \quad ; \quad Q_3 = x_{P=0{\cdot}75} \tag{2}$$

razdelimo populacijo v štiri po obsegu enake dele.

Pod Q_1 leži 25 % populacije, med Q_1 in Q_2 in Q_3 in nad Q_3 pa enako po ena četrtina celotne populacije.

Z decili

$$D_1, \; D_2, \ldots\ldots D_k \ldots\ldots\ldots D_9$$

$$D_k = x_{P=0{\cdot}1.k} \tag{3}$$

razdelimo celotno populacijo v deset po obsegu enakih delov.

Še finejše pa razdelimo populacijo s centili C_k

$$C_1 \quad C_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots C_{99}$$

$$C_k = x_{P=0{\cdot}01.k} \tag{4}$$

na sto po obsegu enakih delov.

Kvantili pomagajo opisovati značilnosti porazdelitev populacij

Meje določena srednja vrednost, kvartili opisujemo variabilnost in asimetrijo porazdelitev. Podobno vlogo imajo tudi decili. V posebne namene služi podrobna razdelitev na centile, ki igra vlogo kvantilnih rangov za vrednotenje individualnih vrednosti.

2. 5. <u>Določanje kvantilov iz frekvenčnih porazdelitev</u>

Čeprav moremo kvantile in kvantilne range določati tudi iz negrupiranih podatkov, jih v praksi največkrat določamo iz frekvenčnih porazdelitev.

Frekvenčna porazdelitev je že sama na sebi neke vrste rangžirna vrsta, ki sicer ne rangira posameznih vrednosti, temveč večje skupine vrednosti.

Iz frekvenčne porazdelitve določimo kvantilemu rang P ustrezno vrednost kvantila x_p po naslednjem postopku:

a) Za frekvenčno porazdelitev izračunamo iz frekvenc f_k kumulativo F_k

b) Iz obsega populacije N in P izračunamo ustrezen rang R_p

$$R_p = NP + 0'5 \qquad (1)$$

c) V frekvenčni porazdelitvi poiščemo med kateri vrednosti v kumulativni vrsti vrsti F_k pade vrednost ranga R_p

$$F_o < R_p < F_1 \qquad (2)$$

Razred z F_o imenujemo kvantilni razred o

d) Za kvantilni razred poiščemo spodnjo mejo razreda $x_o$min, frekvenco f_o, kumulativo F_o in širino razreda i_o.

e) Iz teh podatkov izračunamo kvantil x_p po obrazcu:

$$x_p = x_{o,min} + i_o \frac{R_p - F_o}{f_o} \qquad (3)$$

Po zgornjem postopku izračunana vrednost kvantilov je le približna vrednost, ki jo dobimo z linearno interapolacijo, če predpostavljamo enakomerno razporeditev frekvence v kvantilnem razredu.

Pomembno je, da postopek ni vezan na frekvenčno porazdelitev z enakimi širinami razredov, temveč morejo biti širine razredov poljubne.

Za primer kvantile za frekvenčno porazdelitev za trdnost patentirane jeseniške žice

kp/mm2	f_k	F_k
144 - 45	2	o
146 - 47	-	2
148 - 149	1	2
15o - 151	1	3
152 - 153	2	4
154 - 155	14	6
156 - 157	22	2o
158 - 159	41	42
16o - 161	46	83
162 - 163	21	129
164 - 165	2	15o
166 - 167	1	152
	153	153

$R_p = o,25 = N.o,25 + o,5o = 153.o,25 + o,5o = 38,75$
$R_p = o,5o = N.o,5o + o,5 = 153.o,5 + o,5 = 77$
$R_p = o,75 = N.o,75 + o,5o = 153.o,75 + o,5o = 115,25$

Če napišemo posameznim kvartilnim rangom ustrezne vrednosti v tabeli

P	R_p	f_x,omin	i_o	f_o	F_o	x_p
o,25	38,75	155,5	2	22	2o	157,2o = Q_1
o,5o	77	157,5	2	41	42	159,21 = Q_2
o,75	115,25	159,5	2	46	83	16o,9o = Q_3

Če nakažemo izračun kvartila dobimo iz obrazca (3) in prve kolone v zgornji tabeli

$$Q_1 = x_p = o,25 = 155,5 + 2. \frac{38,75 - 2o}{22} = 157,2o$$

2. 6. Določanje kvantilnih rangov iz frekvenčnih porazdelitev

Kvantilni rangi, ki ustrezajo določeni vrednosti x, določimo podobno kot kvantile z linearno interpolacijo po naslednjem postopku:

a) V frekvenčni porazdelitvi poiščemo razred, v katerem je vrednost x, za katero iščemo kvantilni rang

$$x_{o,min} < x < x_{o,max} \tag{1}$$

b) Če upoštevamo spodnjo mejo $x_{o,min}$, širino i_o, frekvenco in kumulativno frekvenco F_o za dobljeni kvantilni razred, izračunamo kvantilnemu rangu ustrezen rang R_x po obrazcu

$$R_x = F_o + f_o \frac{x_p - x_{o,min}}{i} \tag{2}$$

c) Iz R_x, ki ga dobimo iz obrazca (2) pa izračunamo kvantilni rang po znanem obrazcu

$$P_x = \frac{R_x - 0{,}5}{N} \tag{3}$$

Če se naslonimo na prejšnjo frekvenčno porazdelitev o trnosti za jeseniško patentirano žico in poiščemo kvantilni rang, ki ustreza x = 161,3.

Vrednosti x = 161,3 ustrezajo:

$x_{o,min} = 159{,}5$, $i_o = 2$; $f_o = 46$; $F_o = 83$

Dalje dobimo po obrazcih

$R_x = 83 + 46 \cdot \dfrac{161{,}3 - 159{,}5}{2} = 124{,}4 \qquad P_x = \dfrac{124{,}4 - 0{,}5}{153} = 0{,}814$

Kvantilni rang za x = 161,3 je P_x = o,814. To pomeni, da je 81 % enot populacije, ki imajo vrednost manjšo kot proučevana.

3. RELATIVNA ŠTEVILA

3. 1. Vrste relativnih števil

Statistične podatke primerjamo med seboj z relativnimi števili. Relativna števila dajo podatkom novo kvaliteto in večjo analitično vrednost in primerljivost. V glavnem ločimo glede na odnose podatkov, ki jih primerjamo: strukturne pokazovalce, indekse in koeficiente in gostote.

3. 2. Strukturni pokazovalci

Uvid v sestav populacije po določenem znaku dobimo, če izračunamo strukturne pokazovalce, ki jih dajemo bodisi v deležih, odstotkih ali promilih, odvisno od namena uporabe. Strukturni delež dobimo s primerjavo dela s celoto, t.j. s primerjavo dveh istovrstnih podatkov, od katerih eden velja za delno populacijo, drugi pa za celoto. Če primerjamo število enot, izračunamo strukturne deleže v odstotkih po obrazcu

$$P\% = 100 \cdot \frac{N_a}{N} \qquad (1)$$

pri čemer pomeni N = skupno število enot v celotni populaciji, N_a = število enot delne populacije, ali število enot, ki imajo določeno značilnost.

Tipičen primer strukturnega deleža je odstotek izmeta. Pri tem pomeni N_a = število defektnih artiklov, N = število skupno proizvedenih artiklov. Če pomeni N_{dp} = število defektnih a popravljivih artiklov, N_d = število defektnih artiklov in N = skupno število proizvedenih artiklov, moremo izračunati več strukturnih deležev. $P_{dp} = 100\frac{N_{dp}}{N}$ je strukturni delež števila defektnih a popravljivih artiklov v celotni proizvodnji:

$$P_d = \frac{N_d}{N}$$ = delež defektnih artiklov v celotni proizvodnji

$$P_{d/p} = \frac{N_{dp}}{N_d}$$ = delež defektnih a popravljivih artiklov v populaciji skupnega števila defektnih artiklov.

Iz identitete, ki jo moremo pisati kot

$$\frac{N_{dp}}{N} = \frac{N_{dp}}{N_d} \cdot \frac{N_d}{N}$$

$$P_{dp} = P_d \cdot P_{d/p} \qquad (1)$$

dobimo, da je delež defektnih, a popravljivih artiklov možno razstaviti v dve komponenti kot produkt deleža defektnih artiklov, ki je en kvalitativen pokazovalec in $P_{d/p}$ delež defektnih, a popravljivih artiklov v populaciji vseh defektnih artiklov, kar je drug kvalitativen pokazovalec.

3. 3. Statistični koeficienti

O statističnih koeficientih govorimo, kadar z relativnim številom primerjamo dvoje prirejenih, a raznovrstnih podatkov. Tipično relativno število, ki ga štejemo v to grupo, je produktivnost dela, merjena s kvocientom med proizvodnjo in vloženim delom, potrebnim za proizvodnjo ali s številom zaposlenih, ki so sodelovali pri proizvodnji itd.

Drug primer je koeficient obračanja zalog, ki ga izračunamo kot razmerje med porabo na enoto časa in poprečno zalogo ali kot recipročni pokazovalec, ki ga dobimo kot razmerje med poprečno zalogo in porabo na enoto časa. V prvem primeru je obrtljivost zalog izražena s poprečnim številom obratov v enoti časa, v drugem primeru pa s poprečno dolžino obrata.

3. 4. Indeksi

Če primerjamo dva istovrstna prirejena podatka, dobimo indekse. Po pravilu izračunavamo enostavne indekse po obrazcu

$$I_{1/o} = 1oo \cdot \frac{Y_1}{Y_o} \qquad (1)$$

o ali Y_o imenujemo osnovo ali bazo indeksa. Prednosti in pomen indeksov, ki jih ponavadi izračunavamo za celo statistično vrsto, je v tem, da omogočimo primerjavo statističnih vrst raznovrstnih podatkov ali statističnih vrst istovrstnih podatkov, ki so na različnih nivojih. Indeksne vrste imajo lastnost, da z njimi zreduciramo podatke na neimenovana števila, za katera je indeks za člen ali podatek, na katerega izračunavamo indeksno vrsto, enak 1oo.

Če vzamemo za primer časovni vrsti za odlitke iz železa in jekla v SFRJ v razdobju 1952 - 1965, dobimo naslednjo sliko:

Leto		1952	1953	1954	1955	1956	1957	1958	1959
Proiz-	J	55	67	82	98	1o6	123	141	159
vodnja	Ž	1o46o	1o666	12343	14151	13717	15o24	1717o	2o433
In-	J	1oo	122	149	178	193	224	256	289
deks	Ž	1oo	1o2	118	135	131	144	164	195

Leto		196o	1961	1962	1963	1964	1965
Proiz-	J	192	2o6	2o3	228	279	31o
vodnja	Ž	2332o	27o94	2948o	348o4	39281	41958
In-	J	349	375	369	415	5o7	564
deks	Ž	223	259	282	333	376	4o1

Z indeksi na bazo leto 1952 smo z indeksi različno velike podatke reducirali na primerljive. Iz indeksnih vrst je neposredno vidno počasnejše naraščanje proizvodnje jeklenih odlitkov.

Razen enostavnih indeksov poznamo še druge vrste indeksov. Med njimi so zelo pomembni agregatni indeksi, s katerimi ugotavljamo za količine, ki zavise od več faktorjev, kakšen je učinek delovanja enega samega faktorja. Tipičen primer agregatnega indeksa je indeks volumna proizvodnje. Z njim dosežemo z izračunavanjem vrednosti proizvodnje po stalnih cenah, da je sprememba vrednosti samo izraz sprememb količin.

4. SREDNJE VREDNOSTI

4. 1. Mere centralne tendence

Če opazujemo vrednosti homogene populacije opazimo, da vse vrednosti običajno teže k nekemu centru - središču. Ta center ni karakteristika enot ampak značilnost populacije, in je odvisen od opredeljujočih pogojev, ki so značilni za vse enote populacije. Posamezne vrednosti se zaradi individualnih vplivov sicer odklanjajo od centra, vendar v primeru, da individualni vplivi niso močni, ti odkloni niso veliki in je taka centralna vrednost reprezentant individualnih vrednosti. Statistična metodologija daje več različnih parametrov, ki morejo služiti za centralne vrednosti.

4. 2. Mediana

Kot mera centralne tendence se izkaže kot zelo primeren parameter mediana $Me = x_{0,50}$ Mediana je vrednost, od katere je polovica večjih, polovica pa manjših vrednosti. V ranžirni vrsti je mediana vrednost, ki ima rang $(N+1)/2$. Če je število vrednosti v ranžirni vrsti liho, je $(N+1)/2$ celo število in mediana direktno vrednost člena

$$Me = x_{R=(N+1)/2} \qquad (1)$$

ki je v sredini ranžirne vrste. Če pa je N sodo število vzamemo za mediano poprečje med členoma, ki sta sredini ranžirne vrste najbližja

$$Me = \frac{1}{2}\left[x_{R=N/2} + x_{R=(N+2)/2}\right] \qquad (2)$$

Iz podatkov grupiranih v frekvenčni porazdelitvi, pa izračunamo mediano kot je nakazano za izračun kvantilov.

Mediana je lahko razumljiva in logična vrednost in zato kot opisni parameter priporočljiva. Za določitev mediane je potrebno le poznavanje in obnašanje vrednosti okrog sredine populacije, zato jo moremo določati tudi iz frekvenčnih porazdelitev, ki imajo odprte razrede. Mediana je primerna srednja vrednost v primerih, če obstaja sum, da ekstremne vrednosti ne izhajajo iz homogenih populacij in so izraz nekih drugih kvalitet, ker take vrednosti nebistveno vplivajo na vrednost.

Mediana pa je le zelo neobčutljiva za spremembe vrednosti členov in se ne spremeni vse dokler neka vrednost ne preide iz ene strani mediane na drugo. Za heterogene populacije pa more biti mediana vrednost, ki ni reprezentant niti ene niti druge homogene populacije, ki jo sestavljata, niti skupne heterogene populacije.

4. 3. Modus

Modus je vrednost, ki v populaciji najpogosteje nastopa. Iz tega razloga štejemo modus med srednje vrednosti ker ima lastnosti, ki jih zahtevamo od vrednosti, ki naj reprezentira vrednosti populacije. Modus moremo določiti le za večje populacije, ker moremo samo za take populacije ugotoviti vrednost, ki se najpogosteje pojavlja ali mesto, kjer je gostota frekvence največja.

Razmeroma enostavno ocenimo modus iz frekvenčnih porazdelitev z enakimi razredi. V takem primeru je za dosti za obsežne populacije modus v razredu, ki ima največjo frekvenco. Kot prvo oceno modusa vzamemo kar sredino modalnega razreda x_0. Boljšo oceno dobimo, če upoštevamo še sosedni frekvenci: f_- razreda, ki je pod modalnim razredom in frekvenco f_+, ki je nad modalnim razredom. Če zaznamujemo z $d_{-1} = f_0 - f_{-1}$ in z $d_{+1} = f_0 - f_{+1}$, dobimo oceno modusa po obrazcu

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

Za primer frekvenčne porazdelitve trdnosti patentirane jeseniške žice dobimo, da je modalni razred 60-61 ker je zanj frekvenca največja $f_0 = 46$. Spodnja meja modalnega razreda $x_{0,min} = 159,5$, $f_{-1} = 41$ in $f_{+1} = 21$. Iz tega sledi, da je

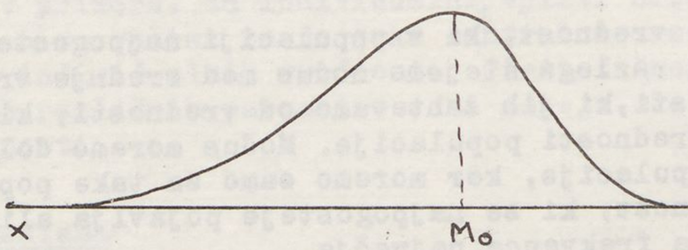
$d_{-1} = 46-41 = 5$, $d_{+1} = 46-21 = 25$,

ocena modusa pa po zgornjem obrazcu

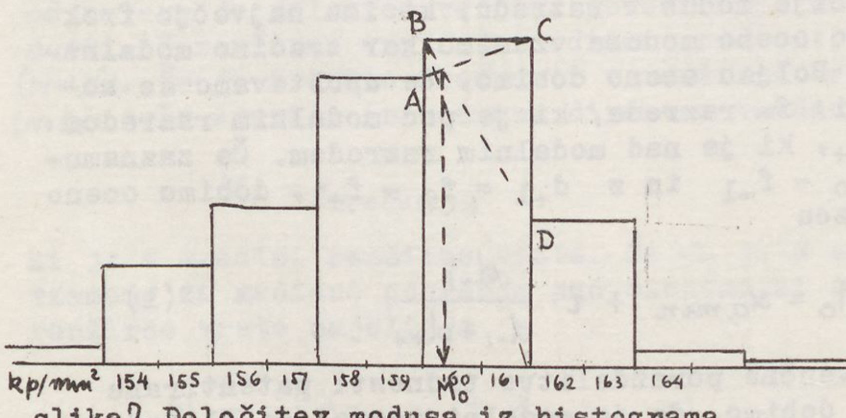
$$M_0 = 159,5 + 2 \frac{5}{5 + 25} = 159,83 \text{ kp/mm2}$$

Grafično določimo modus iz frekvenčne krivulje tako, da poiščemo projekcijo na abscisno os iz mesta, kjer je

gostota frekvence največja (slika 1) iz histograma pa tako, kot kaže slika 2)



slika 1



slika 2 Določitev modusa iz histograma

Histogram more imeti višjo frekvenco kot je v sosednjih razredih na več mestih. To je bodisi izraz slučajnostnih vplivov, ker je obseg populacije ali širina razreda premajhna. V tem primeru z združevanjem razredov v širše razrede dosežemo, da se slučajnostni vplivi eliminirajo ali pa, če gre za vzorčne podatke vzorce povečati, da dobimo stabilne frekvence.

Več lokalnih mest večje gostitve pa dobimo tudi v primeru, če je populacija, katero prikazuje frekvenčna porazdelitev heterogena. Potem je več modusov povezano z vsebino populacije in je polimodalna distribucija izraz heterogenosti osnovne populacije.

4. 4. Aritmetična sredina

Aritmetična sredina ali poprečje je M po definiciji kvocient med vsoto podatkov in številom podatkov

$$M = \frac{1}{N}(x_1 + x_2 + \ldots + x_N) = \frac{1}{N}\sum_{i=1}^{N} x_i = X/N \qquad (1)$$

pri čemer zaznamujemo aritmetično sredino z M individualne vrednosti z x_i število enot z N in vsoto podatkov z X.

Aritmetična sredina ponazarja rezultat splošnih vplivov. Če predpostavljamo, da so v individualni vrednosti rezultat splošnih vplivov M in rezultat individualnih vplivov povezani aditivno

$$x_i = M + e_i \qquad (2)$$

in če predpostavljamo, da se v sumi ali v poprečju rezultat individualnih vplivov uniči, predstavlja aritmetična sredina rezultat splošnih vplivov.

Definicija aritmetične sredine že nakazuje kako jo izračunavamo. Poprečje pokaže tudi, kakšna bi morala biti konstantna vrednost za posamezno enoto, če bi naj vsota konstantnih vrednosti bila enaka vsoti stvarnih vrednosti.

Če vzamemo za primer raztezek lo za 6 preskušancev in so njihove meritve enake: 6,45 6,2o 6,4o 6,oo 5,95 5,9o, je aritmetična sredina za teh šest preskušancev enaka

$$M = \frac{6{,}45 + 6{,}2o + \ldots\ldots + 5{,}95 + 5{,}9o}{6} = 36{,}9o/6$$

$= 6{,}15$.

Za aritmetično sredino velja, da je vsota odklonov individualnih vrednosti od aritmetične sredine enaka nič in da je vsota kvadratov odklonov individualnih vrednosti od neke konstante najmanjša, če je ta konstanta enaka aritmetični sredini.

$$\sum(x-M)=0 \quad ; \quad \sum(x-A)^2=min; \text{ če je } A=M \qquad (3)$$

4. 5. Izračunavanje skupne aritmetične sredine iz grupnih aritmetičnih sredin

Iz grupnih aritmetičnih sredin za delne populacije M_k, katerih vsaka sestoji iz n_k enot, izračunamo skupno aritmetično sredino kot tehtano aritmetično sredino po obrazcu

$$M = \frac{N_1M_1 + N_2M_2 + \dots + N_rM_r}{N_1 + N_2 + \dots + N_r} = \frac{1}{N}\sum_{k=1}^{r} N_kM_k \qquad (1)$$

Pri tem je število enot po posameznih skupinah N_k ponder. Če imamo za posamezne skupine namesto števila enot znane strukturne deleže števila enot od celotne populacije $P_k\%$ izračunamo tehtano aritmetično sredino po obrazcu

$$M = \frac{\sum P\%_k M_k}{100} \qquad (2)$$

Ker je strukturni delež števila enot z dano značilnostjo (npr. strukturni delež izmeta) aritmetična sredina atributivnega znaka, pri čemer pripišemo enoti, ki ima dano značilnost vrednost 1, enotam, ki te značilnosti nimajo pa vrednost o, izračunamo skupen poprečen delež enot z dano značilnostjo enako po obrazcu 1, le da $M_k = P^{\%}_{ak}$.

$$P_a\% = \frac{1}{N}\sum N_k P_a\%_k \qquad (3)$$

Če imamo npr. tri skupine z po N_1 = 1000, N_2 = 2000 in N_3 = 4000 artiklov, odstotek izmeta po posameznih skupinah pa je $P_1\%$ = 2,3 P_2 = 4,1 P_3 = 5,6 , je poprečen odstotek izmeta

$$P\% = \frac{1000.2,3 + 2000.4,1 + 4000.5,6}{1000 + 2000 + 4000} = 4,7\ \%$$

Nepravilen rezultat dobimo, če ne upoštevamo različne velikosti skupin in izračunamo enostavno aritmetično sredino

$$P\% = \frac{2,3 + 4,1 + 5,6}{3} = 4,0\ \%$$

4. 6. Izračunavanje aritmetične sredine iz frekvenčne porazdelitve

Poseben primer izračunavanja tehtane aritmetične sredine je izračun aritmetične sredine iz frekvenčnih porazdelitev. V frekvenčni porazdelitvi za vsak posamezen razred poznamo število enot, ne poznamo pa individualnih vrednosti niti grupne sredine razredov. Zato grupne aritmetične sredine ocenimo s sredino razreda x_k. Če te vrednosti vnesemo v obrazec 1) dobimo, da je M

$$M = \frac{1}{N}(f_1 x_1 + f_2 x_2 + \cdots + f_r x_r) = \frac{1}{N}\sum_{k=1}^{r} f_k x_k \qquad (1)$$

Tako dobljeno poprečje je le ocena pravega poprečja, ki bi ga dobili, če bi izračunali poprečje iz individualnih vrednosti, ali če bi upoštevali prave razredne sredine. Vendar je ta ocena zelo blizu pravega poprečja, če je porazdelitev simetrična ali ne preveč asimetrična. Pri zelo asimetričnih posebno pa pri J - porazdelitvah pa istosmerna sistematična pogreška v grupnih sredinah bistveno vpliva na rezultat.

4. 7. Izračunavanje aritmetične sredine iz frekvenčne porazdelitve po direktnem obrazcu

Obrazec nakazuje, kako iz frekvenčne porazdelitve izračunamo aritmetično sredino. Vsoto produktov frekvenc f_k s sredinami razredov x_k delimo s številom enot v populaciji. Kolikor je ta metoda računsko neprikladna, je splošna in jo moremo uporabiti tudi za frekvenčne porazdelitve z neenako širokimi razredi.

Za primer izračuna aritmetične sredine po direktni metodi vzemimo trdnost jeseniške patentirane žice.

kp/mm2	f_k	x_k	$f_k x_k$
144 - 145	2	144,5	289,o
146 - 147	-	146,5	-
148 - 149	1	148,5	148,5
15o - 151	1	15o,5	15o,5
152 - 153	2	152,5	3o5,o
154 - 155	14	154,5	2163,o
156 - 157	22	156,5	3443,o
158 - 159	41	158,5	6498,5
16o - 161	46	16o,5	7383,o
162 - 163	21	162,5	3412,5
164 - 165	2	164,5	392,o
166 - 167	1	166,5	166,5
	153		24288,5
	N		$\sum f_k x_k$

$$M = \frac{\sum f_k x_k}{N} = \frac{24288,5}{153} = 158,75 \text{ kp/mm2}$$

Ker je porazdelitev unimodalna ne preveč asimetrična iz frekvenčne porazdelitve izračunana vrednost nebistveno odstopa od prave aritmetične sredine M = 158,78 kp/mm2, ki jo dobimo iz individualnih podatkov.

4. 8. Izračunavanje aritmetične sredine iz frekvenčnih porazdelitev po metodi pomožnega znaka u

Če so podatki grupirani v frekvenčni porazdelitvi z enako širokimi razredi, aritmetično sredino izračunamo enostavnejše, če sredine razredov transformiramo po obrazcu

$$u_k = \frac{x_k - x_o}{i} \qquad (1)$$

pri čemer je x_o sredina poljubnega razreda nekje v sredini porazdelitve pri največjih frekvencah. Pri zgornjih pogojih so u enostavne vrednosti -5 -4 -3 -2 -1 o +1 +2 +3 +4......
Ker velja, da je aritmetična sredina linearne zveze enaka linearni zvezi aritmetičnih sredin znakov, izračunamo najprej M_u iz tega pa po obrazcu

$$M_x = x_o + i\, M_u = x_o + i\, \frac{\sum f_k u_k}{N} \qquad (2)$$

Za primer vzamemo isto frekvenčno porazdelitev, kot pri direktnem izračunu. Po metodi pomožnega znaka u dobimo

kp/mm2	f	u	fu
144 - 145	2	-6	-12
146 - 147	-	-5	-
148 - 149	1	-4	- 4
15o - 151	1	-3	- 3
152 - 153	2	-2	- 4
154 - 155	14	-1	-14
156 - 157	22	o	o
158 - 159	41	1	41
16o - 161	46	2	92
162 - 163	21	3	63
164 - 165	2	4	8
166 - 167	1	5	5

$N = 153 \quad \sum f_k u_k = +172$

$$M = x_o + i\, \frac{\sum fu}{N} = 156{,}5 + 2\, \frac{+172}{153} = 158{,}75 \text{ kp/mm2}$$

4. 9. <u>Izračunanje aritmetične sredine iz frekvenčnih porazdelitev z metodo kumulativ</u>

Zaradi posebnih lastnosti kumulativnih vrst za frekvenčne porazdelitve z enako širokimi razredi izračunati aritmetično sredino s kumulativnimi vrstami brez masovnega množenja.

Iz frekvenčne porazdelitve izračunamo po znanem postopku prvo kumulativo F_k in seštejemo člene prve kumulative, (brez člena, ki je pod črto in pomeni N) vsota členov kumulativne vrste zaznamujemo s S_1. $= F_1 + F_2 + \ldots\ldots\ldots + F_k$

b) Iz teh količin izračunamo aritmetično sredino po obrazcu

$$M_x = x_o - i\, \frac{S_1}{N} \qquad (1)$$

pri tem je razen že pojasnjenih vrednosti x_o=sredina najvišjega razreda v frekvenčni porazdelitvi.

Za primer izračunajmo za frekvenčno porazdelitev trdnosti patentirane jeseniške žice aritmetično sredino še po metodi kumulativ.

kp/mm2	f_k	F_k
144 - 145	2	o
146 - 147	-	2
148 - 149	1	2
15o - 151	1	3
152 - 153	2	4
154 - 155	14	6
156 - 157	22	2o
158 - 159	41	42
16o - 161	46	83
162 - 163	21	129
164 - 165	2	15o
166 - 167	1	152
	N =	153

S_1 = o+2+2+3+ .....+15o+152 = 593

$$M = x_o - i.\frac{S_1}{N} = 166{,}5 - 2 \cdot \frac{593}{153} = 158{,}75 \text{ kp/mm2}$$

4. lo. Harmonična sredina

Harmonična sredina H je po definiciji recipročna vrednost iz aritmetične sredine recipročnih vrednosti.

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

Če posamezne vrednosti nastopajo z različnimi utežmi w_i, izračunamo harmoničmo sredino po obrazcu

$$H = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} \qquad (2)$$

Harmonično sredino izračunavamo redkeje kot aritmetično. Dobro pa pokaže centralno tendenco takrat, če je porazdelitev asimetrična in se recipročne vrednosti porazdeljujejo simetrično pri poprečjih iz relativnih števil.

4. 11. Poprečja iz relativnih števil

Pri izračunavanju sumarnih relativnih števil iz grupnih podatkov nastopajo glede na razpoložljive podatke različne situacije.

Če je grupno relativno število r_k definirano s kvocientom dveh ekstenzivnih količin Y_k in X_k

$$r_k = \frac{Y_k}{X_k}\,, \tag{1}$$

je sumarno relativno število r, ki je po definiciji

$$\bar{r} = \frac{Y}{X} = \frac{\sum Y_k}{\sum X_k} = \frac{\sum X_k r_k}{\sum X_k} = \frac{\sum Y_k}{\sum \frac{Y_k}{r_k}} \tag{2}$$

kvocient med vsotama grupnih vrednosti Y_k in X_k ali tehtana aritmetična sredina grupnih relativnih števil r_k, če razen z grupnimi relativnimi števili razpolagamo še z vrednostmi, ki v relativnem številu nastopajo v imenovalcu. Kot tehtano harmonično sredino relativnih števil r_k izračunamo $\bar{r}$, če razen z r_k razpolagamo še z vrednostmi Y_k, ki v relativnih številih nastopajo v števcu. Za primer vzemimo poprečno hitrost žerjavov na liniji, ki sestoji iz treh odsekov. Od teh je eden dolg S_1 = 20 m, hitrost na tem odseku pa v = 11m/sek, drugi odsek je dolg Y_2 = 30 m, hitrost na tem odseku pa v_2 = 1,5 m/sek, tretji odsek ima dolžino S_3 = 15 m, hitrost na tem odseku pa je v_3 = 0,5 m/sek. Ker je hitrost po definiciji razmerje med potjo in časom, razpolagamo pa razen s hitrostmi še z dolžino odsekov, izračunamo poprečno hitrost s harmonično sredino, ker so ponderi količine, ki nastopajo v števcu

$$v = \frac{S_1 + S_2 + S_3}{\frac{S_1}{v_1} + \frac{S_2}{v_2} + \frac{S_3}{v_3}} = \frac{20 + 30 + 15}{\frac{20}{1} + \frac{30}{1,5} + \frac{15}{0,5}} = 0,929 \text{ m/sek}$$

4. 12. Geometrijska sredina

Geometrijska sredina

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \cdots\cdots x_N} \qquad (1)$$

je koren stopnje N iz produkta individualnih podatkov. Geometrijska sredina je definirana tudi tako: logaritem iz geometrijske sredine je enak aritmetični sredini iz logaritmov iz individualnih vrednosti

$$\log G = \frac{1}{N} \left(\sum \log x_i \right) \qquad (2)$$

Podobno kot za aritmetično in harmonično sredino, more biti tudi geometrijska sredina tehtana, če vsaki individualni vrednosti pripišemo nek ponder w_i. Tako dobimo

$$G = \sqrt[w]{x_1^{w_1} \cdot x_2^{w_2} \cdots\cdots\cdots x_N^{w_N}} \qquad w = w_1 + w_2 \cdots\cdots + w_N \quad (3)$$

Podobno velja tudi za logaritme

$$\log G = \frac{\sum w_i \log x_i}{\sum w_i} \qquad (4)$$

Geometrijska sredina bolje kot druge srednje vrednosti pokaže mesto centralne tendence, če so važnejši relativni kot absolutni odnosi med vrednostmi. Tak primer so posebno zelo asimetrične porazdelitve. Relativna razlika med 1 in 2 je enaka kot med 1oo in 2oo, medtem ko to ne velja za absolutne razlike.

Razen tega geometrijsko sredino uporabljamo za izračun povprečnega koeficienta dinamike.

5. MERE VARIACIJE

5. 1. Variabilnost

Pri kvantitativnem proučevanju pojavov naletimo na tipično lastnost numeričnih pojavov, na variabilnost podatkov.

Tako v proizvodnji ugotavljamo, da se karakteristike proizvodnje razlikujejo od artikla do artikla. Te razlike so izvor spreminjajočih faktorjev, ki vplivajo na proizvodnjo. Večje spremembe faktorjev se pokažejo v večji variabilnosti podatkov. Medtem ko je centralna tendenca rezultat splošnih, opredeljujočih pogojev, je variabilnost izraz delovanja individualnih vplivov, in kot so srednje vrednosti mere centralne tendence, so mere variacije numeričen izraz variabilnosti. Od vseh mer variacij so za uporabo v tehniki posebej pomembne naslednje mere variacije: variacijski razmak, popreč̌en absoluten odklon, varianca in standardni odklon.

5. 2. Variacijski odklon

Variacijski razmak

$$R = x_{max} - x_{min} \qquad (1)$$

je razlika med največjo in najmanjšo vrednostjo v opazovani populaciji. Je najenostavnejša mera variacije in ga uporabljamo v primerih, ko je treba hitro in enostavno izmeriti variabilnost nekega pojava. Variacijski razmak je zato posebno primerna mera variacije pri kontroli kvalitete proizvodnega procesa.

5. 3. Popreč̌en absoluten odklon

Mera variacije, za katero je definicija dana s samim imenom. AD je popreč̌en absoluten odklon individualnih podatkov od ustrezne srednje vrednosti. Računamo ga po obrazcu

$$AD_M = \frac{1}{N}\sum_{i=1}^{N} |x_i - M| \qquad (1)$$

kot popreč̌en absoluten odklon od aritmetične sredine, pogosteje pa kot

$$AD_{Me} = \frac{1}{N}\sum_{i=1}^{N} |x_i - Me| \qquad (2)$$

popreč̌en odklon od mediane. AD_{Me} je vsebinsko bolj upravičen, ker velja, da je popreč̌en absoluten odklon najmanjši, če odklone računamo od mediane.

Neglede na to je AD_{Me} tudi računsko enostavno izračunati po obrazcu

$$AD_{Me} = \frac{1}{N}\left(\sum_z x_i - \sum_s x_j\right) \qquad (3)$$

pri čemer pomeni $\sum_z x_i$ vsoto podatkov, ki so večji, $\sum_s x_j$ pa vsoto podatkov, ki so manjši kot mediana.

Primer. Če vzamemo za primer šest meritev za raztezek žice $\lambda_{10}\%$ ki so urejene v ranžirno vrsto

5,9o 5,95 6,oo 6,2o 6,4o 6,45

je za prve tri meritve, ki so pod mediano, vsota $\sum_s$ = 17,85, za druge tri meritve, za katere so vrednosti nad mediano pa $\sum_z$ = 19,o5.

Po zgornjem obrazcu je

$$AD_{Me} = \frac{1}{6}\,(19{,}o5 - 17{,}85) = o{,}2o$$

5. 4. <u>Varianca</u>

Varianca je mera variacije, ki je definirana kot popreček kvadratičen odklon od aritmetične sredine

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - M)^2 \qquad (1)$$

Varianca je najpomembnejša mera variacije, predvsem iz teoretičnih ozirov. Zato ima velike prednosti pred vsemi drugimi merami variacije, ki so predvsem opisnega značaja.

<u>Izračunanje variance iz individualnih podatkov</u>

Izračunanje variance po zgornjem obrazcu je neprikladno, ker je aritmetična sredina običajno decimalno število.

Enostavneje izračunavamo varianco iz negrupiranih podatkov po obrazcu

$$\sigma_x^2 = \frac{1}{N}\left[\sum u^2 - \frac{(\sum u)^2}{N}\right] = \frac{K_u}{N} \qquad (2)$$

pri čemer je $u_i = x_i - x_o$ odkloni od poljubne vrednosti, K pa izraz

$$K = \sum u^2 - \frac{(\sum u)^2}{N} \qquad (3)$$

Vzemimo za primer izračunanje variance za neto čas deset šarž: 8,oo 7,5o 6,67 7,5o 7,75, 6,75 7,5o 6,67 7,5o 6,67 7,5o 7,oo

x_i	u_i	u_i^2
8,oo	1,oo	1,oooo
7,5o	o,5o	o,25oo
6,67	-o,33	o,1o89
7,5o	o,5o	o,25oo
7,75	o,75	o,5625
6,75	o,25	o,o625
7,5o	o,5o	o,25oo
6,67	-o,33	o,1o89
7,5o	o,5o	o,25oo
7,oo	o,oo	o,oooo
	3,34	18,528
	$\sum u$	$\sum u^2$

$$K = 18,528 - \frac{3,34^2}{1o} = 18,4164$$

$$\begin{array}{r} -0,1116 \\ \hline 18,4164 \end{array}$$

$$\sigma^2 = \frac{K}{N} = \frac{18,4164}{1o} = 1,8416$$

Če računamo varianco na računski stroj se izkaže, da je najprikladnejše, da vzamemo $x_o = o$, da ni treba izračunavati u_i. V tem primeru preide obrazec (2) v

$$\sigma_x^2 = \frac{1}{N}\left[\sum x^2 - \frac{(\sum x)^2}{N}\right] \qquad (4)$$

Vsoto kvadratov in vsoto osnovnih podatkov dobimo na računskem stroju v tem primeru neposredno.

5. 5. Metoda pomožnega znaka u za izračunavanje variance

Če imamo podatke grupirane v frekvenčni porazdelitvi z enakimi razredi izračunamo varianco po naslednjem postopku:

a) Nekje v sredini frekvenčne porazdelitve vnesemo vrednost o pomožnega znaka u. Za druge razrede pa po vrsti navzdol - 1, - 2, - 3 .... in navzgor + 1, + 2, +3........

b) izračunamo vrsto produktov fu in izračunamo vsoto $\sum fu$;

c) izračunamo vrsto produktov fu^2 po postopku (fu)u in vsoto $\sum fu^2$;

d) vsoto kvadratov odklonov K izračunamo po obrazcu

$$K = \sum fu^2 - \frac{(\sum fu)^2}{N} \tag{1}$$

e) varianco izračunamo po obrazcu

$$\sigma_x^2 = \frac{i^2}{N} K \tag{2}$$

Primer: Za 153 preskusov trdnosti patentirane jeseniške žice je izračun variance po nakazani metodi naslednji

kp/mm2	f	u	fu	fu^2
144-145	2	-6	-12	72
146-147	o	-5	o	o
148-149	1	-4	-4	16
15o-151	1	-3	-3	9
152-153	2	-2	-4	8
154-155	14	-1	-14	14
156-157	22	o	o	o
158-159	41	1	41	41
16o-161	46	2	92	184
162-163	21	3	63	189
164-165	2	4	8	32
166-167	1	5	5	25
	153		+172	59o
	N		$\sum fu$	$\sum fu^2$

$$K = 59o - \frac{(+172)^2}{153} = 1o{,}3684 \; ; \quad \sigma_x^2 = \frac{2^2}{153} \cdot 396{,}64 = 10{,}3684$$

5. 6. Metoda kumulativ za izračunavanje variance

Če imamo podatke grupirane v frekvenčni porazdelitvi z enakimi razredi, enostavno izračunamo varianco tudi po naslednjih stopnjah:

a) Iz frekvenčne porazdelitve izračunamo prvo kumulativo, iz prve pa drugo kumulativo

b) izračunamo vsoti S_1 in S_2 iz členov prve in druge kumulative (S_1 se pojavi kot člen pod črto v drugi kumulativi)

c) Iz N, S_1 in S_2 izračunamo izraz K po obrazcu

$$K = 2S_2 + S_1 - \frac{S_1^2}{N} \qquad (1)$$

d) Varianco izračunamo iz dobljenih količin po obrazcu

$$\sigma_x^2 = \frac{i^2 K}{N} \qquad (2)$$

Primer: Za N= 153 preskusov trdnosti patentirane jeseniške žice izračunamo po metodi kumulativ varianco po naslednjem postopku:

kp/mm2	f_k	F_k	FF_k
144-145	2	o	
146-147	o	2	o
148-149	1	2	2
15o-151	1	3	4
152-153	2	4	7
154-155	14	6	11
156-157	22	2o	17
158-159	41	42	37

$S_1 = o+2+2+3+ \ldots .+15o+152 = 593$

$S_2 = o+2+4+ \ldots \ldots +291+441 = 1o51$

$$K = 2.1o51+593- \frac{593^2}{153} = 396{,}64$$

kp/mm^2	f_k	F_k	FF_k
16o-161	46	83	19
162-163	21	129	162
164-165	2	15o	291
166-167	1	152	441
		153	593

$$\sigma_x^2 = \frac{2^2 . 396{,}64}{153} = 1o{,}3684$$

5. 7. Sheppardova korektura

Frekvenčna porazdelitev da le približno sliko o vrednostih v populaciji. Zato je varianca, izračunana iz frekvenčnih porazdelitev le ocena prave vrednosti. Sistematična napaka, ki izvira iz tega, da so podatki grupirani, je odvisna od velikosti razreda. Korektura variance zaradi širine razreda je dana s Sheppardovo korekturo po obrazcu

$$\sigma_{cor}^2 = \sigma^2 - \frac{i^2}{12} \qquad (1)$$

Za primer N = 153 preskusov trdnosti patentirane jeseniške žice je po tem obrazcu korigirana varianca

$$\sigma^2_{cor} = 1o{,}3684 - \frac{2^2}{12} = 1o{,}o352$$

5. 8. Standardni odklon

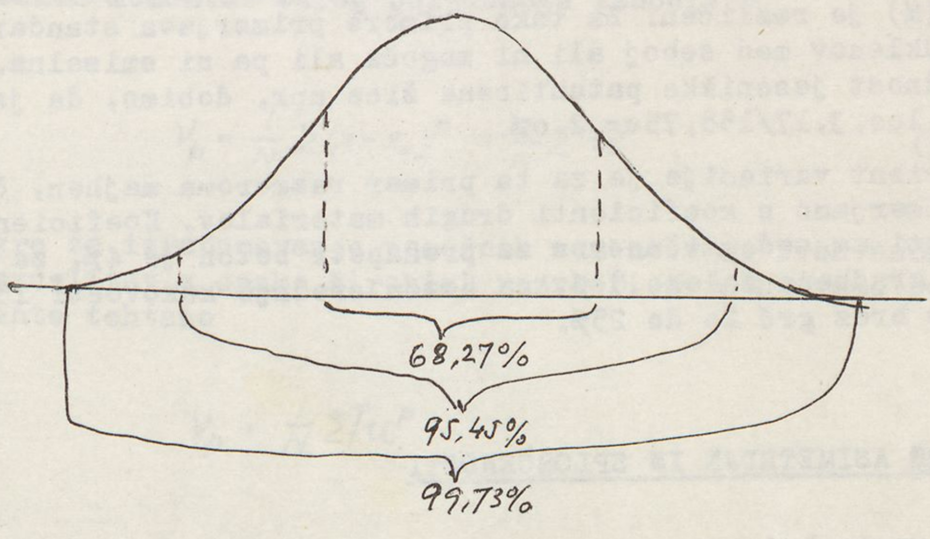
Standardni odklon je kvadratni koren iz variance

$$\sigma_x = \sqrt{\sigma_x^2} \qquad (1)$$

Izračunavamo ga kot opisni parameter variabilnosti pojava.

Razen tega je standardni odklon eden izmed bistvenih parametrov, ki so osnova za ocenjevanje s slučajnostnim vzorcem.

Standardni odklon ima isto enoto mere kot znak, za katerega ga računamo. Pravo predstavo o pomenu in velikosti standardnega odklona dobimo šele, če ga obravnavamo v zvezi z osnovno teoretično porazdelitvijo, t.j. normalno porazdelitvijo. Velja namreč, da se v razmaku od $M - \sigma$ do $M + \sigma$ nahaja ca 68 % populacije, v razmaku $M - 2\sigma$ do $M + 2\sigma$ ca 95 % v razmaku $M - 3\sigma$ do $M + 3\sigma$ pa 99,7 % vse normalno porazdeljene populacije.



Za primer trdnosti patentirane jeseniške žice je standardni odklon, izračunan iz korigirane variance

$$\sigma = \sqrt{10,0352} = 3,17 \ kp/mm^2$$

5. 9. Koeficient variacije

Koeficient variacije KV

$$KV = \frac{\sigma}{M}\ ; \quad KV\% = 100\,\frac{\sigma}{M} \qquad (1)$$

je razmerje med standardnim odklonom in aritmetično sredino in ga štejemo med relativne mere variacije. Kadar ga izračunavamo kot opisni parameter, ga navadno podajamo v odstotkih KV%. Ker je koeficient variacije neimenovano število, je zelo primeren za primerjave jakosti variacije med raznovrstnimi podatki ali tudi za istovrstne podatke, katerih nivo (M) je različen. Za take primere primerjava standardnih odklonov med seboj ali ni mogoča ali pa ni smiselna. Za trdnost jeseniške patentirane žice npr. dobimo, da je KV% = loo. 3,17/158,75 = 2,o%.

Koeficient variacije je za ta primer razmeroma majhen, če ga primerjamo s koeficienti drugih materialov. Koeficient trdnosti za sedemžično vrv za prenapeti beton je 4%, za normalno gradbeno železo 8 %, za beton srednje kakovosti 15%, za les brez grč 2o do 25%.

6. MERE ASIMETRIJE IN SPLOŠČENOSTI

6. 1. Centralni momenti

Osnovne količine za proučevanje značilnosti populacij, v posebnem primeru za proučevanje asimetrije in sploščenosti so centralni momenti.

Centralni moment stopnje p je definiran kot poprečna potenca stopnje p odklonov individualnih vrednosti od ustrezne aritmetične sredine

$$\mu_p = \frac{1}{N}\sum_{i=1}^{N}(x_i - M)^p \qquad (1)$$

Iz definicije aritmetične sredine in variance sledi, da je

$\mu_1 = 0 \quad ; \quad \mu_2 = \sigma^2$

Tretji centralni moment $\mu_3 = \frac{1}{N} \Sigma (x - M)^3$ je osnova za mero asimetrije, četrti centralni moment $\mu_4 = \frac{1}{N} \Sigma (x - M)^4$ pa za mero sploščenosti porazdelitve.

6. 2. Pomožni momenti

Iz istih razlogov kot je izračunanje variance po definicijskem obrazcu tehnično komplicirano, velja tudi za momente na splošno. Zato centralne momente izračunavamo preko pomožnih momentov okrog poljubnega izhodišča

$$\nu_p = \frac{1}{N} \Sigma (x - x_0)^p = \frac{1}{N} \Sigma u^p \qquad (1)$$

Če gre za izračunavanje pomožnih momentov iz frekvenčnih porazdelitev z enako širokimi razredi, izračunavamo pomožne momente tehtano

$$\nu_p = \frac{1}{N} \Sigma f u^p \qquad (2)$$

pri čemer je u pomožni znak $u = (x - x_0)/i$

Zveza med pomožnimi centralnimi momenti:

Iz algebraičnih zvez med potencami binomov sledijo naslednje zveze med centralnimi in pomožnimi momenti za znak u

$$\begin{aligned} \mu_{1u} &= 0 \\ \mu_{2u} &= \nu_2 - \nu_1^2 \\ \mu_{3u} &= \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 \\ \mu_{4u} &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4 \end{aligned} \qquad (3)$$

Če centralne momente izračunavamo iz negrupiranih podatkov, velja med centralnimi momenti za u in x zveza:

$$\mu_{px} = \mu_{pu} \quad (4)$$

če pa jih izračunavamo iz frekvenčnih porazdelitev, pa je

$$\mu_{px} = i^p \mu_{pu} \quad (5)$$

6. 3. Charlierjev preskus

Izračunanje pomožnih momentov iz frekvenčnih porazdelitev je računsko precej zahtevno. Da se izognemo možnim napakam, preiskusimo pravilnost izračuna izrazov $\sum fu^p$ s Charlierjevim preskusom.

Če razvijemo četrto potenco binoma

$$(1+u)^4 = 1 + 4u + 6u^2 + 4u^3 + u^4 \quad (1)$$

sledi dalje

$$\sum f(1+u)^4 = N + 4\sum fu + 6\sum fu^2 + 4\sum fu^3 + \sum fu^4 \quad (2)$$

Če izračunamo $\sum f(1+u)^4$ služi obrazec (2) za kontrolo izrazov $\sum fu^p$.

6. 4. Sheppardova korektura centralnih momentov

Če izračunamo centralne momente iz frekvenčnih porazdelitev, so po obrazcih izračunane vrednosti za centralne momente ocene pravih vrednosti, ki bi jih dobili iz negrupiranih podatkov. Vsaj delno to napako omilimo s Sheppardovo korekturo.

Ta je:

$$\mu_{2u,\,cor} = \mu_{2u} - \frac{1}{12}$$

$$\mu_{3u,\,cor} = \mu_{3u} \qquad (1)$$

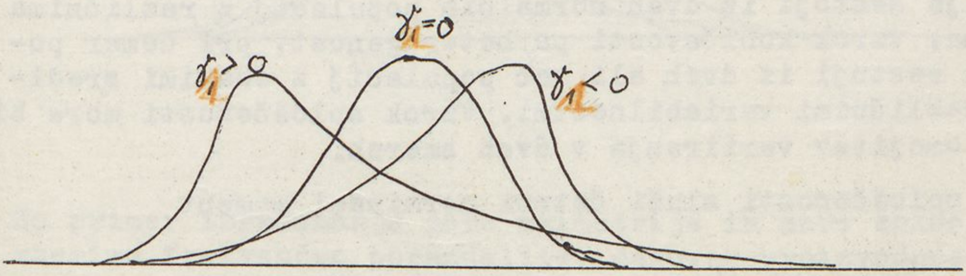
$$\mu_{4u,\,cor} = \mu_{4u} - \frac{1}{2}\mu_{2u} + \frac{7}{240}$$

6. 5. Mera asimetrije

Asimetrijo porazdelitve merimo z normiranim tretjim centralnim momentom

$$\gamma_1 = \frac{\mu_{3u\,cor}}{\sqrt[2]{\mu_{2u\,cor}^{3}}} \qquad (1)$$

Če je porazdelitev simetrična, je γ_1 enak nič, če je porazdelitev asimetrična v desno, je γ_1 večji, če pa je asimetrična v levo pa manjši od 1.



Če na pojav vplivajo samo slučajnostni faktorji in je populacija homogena, variabilnost pa ni omejena v nobeni smeri, dobimo simetrično porazdelitev. Vzrokov za asimetrijo je lahko več, ali individualni vplivi niso samo slučajnostni, če populacija ni homogena, temveč sestavljena iz več homogenih populacij, ali da ima variabilnost v eno smer omejitev. Porazdelitev določenih karakteristik proizvodnje je asimetrična, če izvira iz proizvodnje pod različnimi pogoji, ali če je poprečje tako blizu neke naravne omejitve (npr. o), da je variabilnost v eni smeri dušena.

6. 6. Sploščenost porazdelitve

Na frekvenčnih porazdelitvah opazujemo tudi pojav sploščenosti oziroma koničavosti porazdelitev. Za simetrične porazdelitve je stopnja asimetrije nič. Pri sploščenosti tega ni. Zato vzamemo kot osnovo za primerjavo sploščenosti porazdelitev normalno porazdelitev. Idealno sploščenost pripisujemo tej porazdelitvi. Vse druge porazdelitve so glede na normalno porazdelitev sploščene ali koničaste.

Določena karakteristika se porazdeljuje v normalni porazdelitvi pri idealnih pogojih, da na pojav vplivajo samo slučajnostni vplivi in ti niso dušeni v nobeno smer. Vzrok sploščenosti pa morata biti heterogenost populacije, pri čemer se populacija sestoji iz dveh normalnih populacij z različnima sredinama, vzrok koničavosti pa heterogenost, pri čemer populacija sestoji iz dveh ali več populacij z enakimi sredinami a različnimi variabilnostmi. Vzrok sploščenosti more biti tudi omejitev variiranja v dveh smereh.

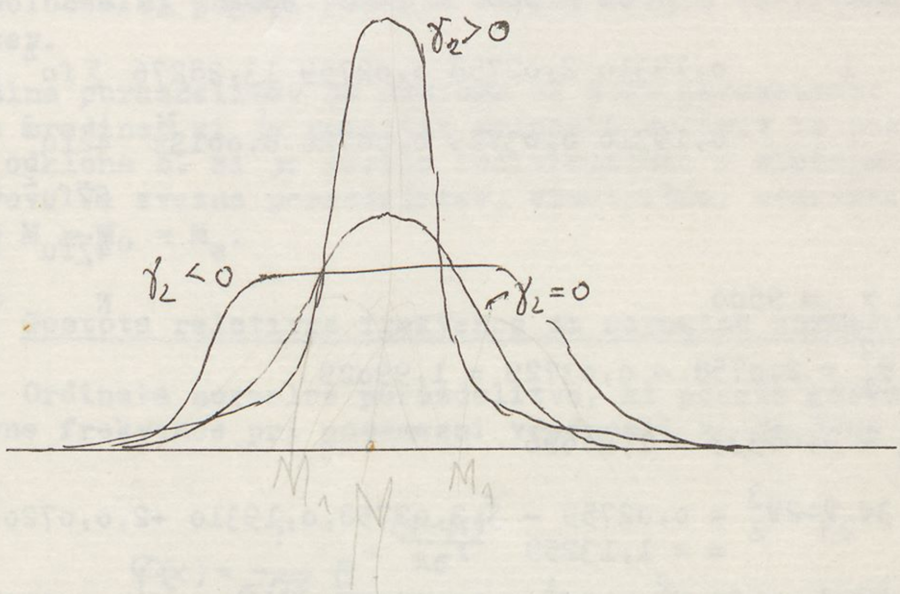
Za mero sploščenosti služi četrti normirani moment

$$\alpha_4 = \frac{\mu_{4u\,cor}}{\mu^2_{2u,cor}} \qquad (1)$$

Ker je četrti normiran moment za normalno porazdelitev enak 3, dostikrat uporabljamo za mero sploščenosti koeficient

$$\gamma_2 = \alpha_4 - 3 \qquad (2)$$

Če je porazdelitev normalno sploščena, je $\gamma_2 = 0$, za sploščene porazdelitve je γ_2 negativen za koničaste pa pozitiven.



Za primer izračunanja mere asimetrije in mere sploščenosti vzemimo frekvenčno porazdelitev časovne proizvodnosti za domači gorilnik za 29o šarž. V nakazanem primeru so izračunane vse pomožne količine , vključen Charlierjev preskus za vsote potenc in upoštevana je Sheppardova korektura. Rezultati pokažejo, da je porazdelitev asimetrična v levo, (γ_1 = -o,43o) glede na stopnjo sploščenosti pa je porazdelitev v primerjavi z normalno koničasta (γ_2 = +o,5o9).

Izračunanje M, σ, γ_1 in γ_2 za časovno proizvodnost domačega gorilnika — Charlijev preskus

kp/h	f_k	U_k	f_u	f_u^2	f_u^3	f_u^4	$(u+1)^4$	$f(u+1)^4$
4000-4999	1	-5	-5	25	-125	625	256	256
5000-5999	2	-4	-8	32	-128	512	81	162
6000-6999	7	-3	-21	63	-189	567	16	112
7000-7999	23	-2	-46	92	-184	368	1	23
8000-8999	52	-1	-52	52	- 52	52	0	0
9000-9999	72	0	0	0	0	0	1	72
10000-10999	89	+1	89	89	89	89	16	1424
11000-11999	35	+2	70	140	280	560	81	2835
12000-12999	7	+3	21	63	189	567	256	1792
13000-13999	2	+4	8	32	128	512	625	1250
$\sum fu^r$	290		+ 56	588	+ 8	3852	$\sum f(u+1)^4 =$	7926
$v_r = \frac{1}{N}\sum fu^r$	1		0,19310	2,02758	0,02759	13,28276	$\sum fu^4 =$	3852
v_1^r			0,19310	0,03729	0,00720	0,00139	$4\sum fu^3 =$	32
							$6\sum fu^2 =$	3528
							$4\sum fu =$	224
							$N =$	290
								7926

$i = 1000 : x_o = 9500$

$\mu_2 = \nu_2 - \nu_1^2 = 2,0758 - 0,03729 = 1,99029$

$\mu_{2cor} = \mu_2 - 0,08333 = 1,90696$

$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 = 0,02759 - 3.2,02758.0,19310 + 2.0,0720 =$
$= - 1,13259$

$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4 = 13,28276 - 4.0,00720.0,19310 +$
$+ 6.2,02758 . 0,03729 - 3.0,00139 =$
$= 13,72668$

$\mu_{4cor} = \mu_4 - \frac{1}{2}\mu_2 + \frac{7}{240} = 13,72668 - \frac{1}{2}.1,99029 + \frac{7}{240} = 12,76071$

$\bar{x} = x_o + i\nu_1 = 9500 + 1000.0,19310 = 9693,10$

$\sigma_{xcor} = i\sqrt{\mu_{2cor}} = 1000.\sqrt{1,90696} = 1380,93$

$KV\% = 100\,\frac{\sigma_x}{\bar{x}} = 14,25\%$ $\quad \gamma_1 = \frac{\mu_3}{\sqrt{\mu_{2cor}^3}} = \frac{-1,13259}{\sqrt{1,90696^3}} = - 0,430$

$\gamma_2 = \frac{\mu_{4.cor}}{\mu_2^2} - 3 = \frac{12,76071}{1,90696^2} - 3 = + 0,509$

7. NORMALNA PORAZDELITEV

7. 1. Normalna porazdelitev

V statistični teoriji in praksi najpomembnejša teoretična porazdelitev je normalna porazdelitev. Če na pojav razen splošnih faktorjev vplivajo le slučajnostni faktorji, ki v nobeno smer niso dušeni, se znak porazdeljuje v unimodalni, simetrični ~~porazdelitvi~~ zvonasti porazdelitvi, ki so značilnosti normalne porazdelitve.

Normalna porazdelitev je osnova vzorčenja, porazdelitev, ki je osnova za različne druge teoretične porazdelitve in v njo pod določenimi pogoji prehaja večina drugih teoretičnih porazdelitev.

Normalna porazdelitev je odvisna od dveh parametrov: aritmetične sredine M, ki je rezultat splošnih vplivov in standardnega odklona σ, ki je merilo individualnih - slučajnostnih vplivov. Je zvezna porazdelitev, simetrična, zvonasta in je zanjo $M = M_o = M_e$.

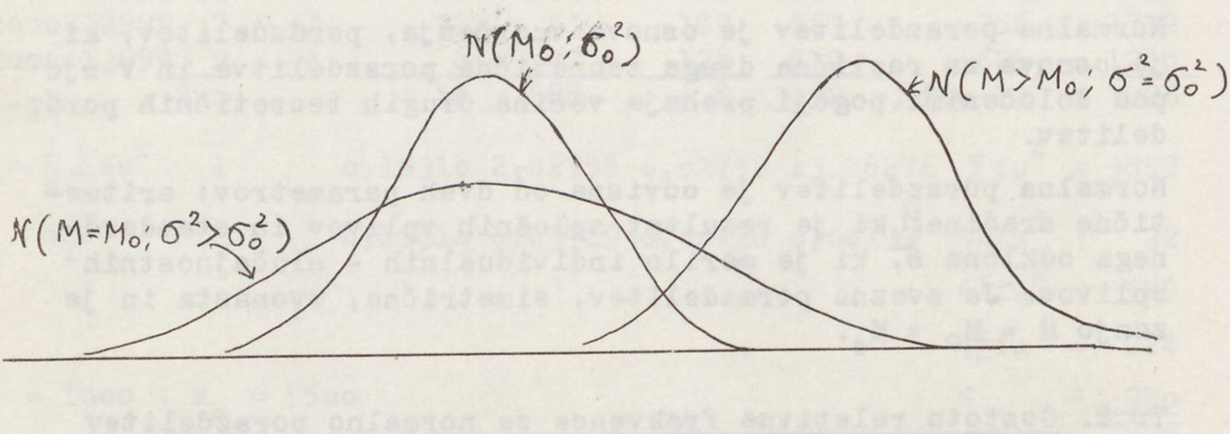
7. 2. Gostota relativne frekvence za normalno porazdelitev

Ordinata normalne porazdelitve, ki pokaže gostoto relativne frekvence pri posamezni vrednosti x, je dana z obrazcem

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} \tag{1}$$

pri čemer so: e = baza naravnega logaritma, π = Ludolfovo število, M aritmetična sredina, σ pa standardni odklon.

Normalne porazdelitve z različnimi M, a enakim standardnim odklonom so skladne, vendar pomaknjene v levo ali desno, medtem ko so normalne porazdelitve z različnimi standardnimi odkloni med seboj podobne. Na sliki je prikazano za primerjavo nekaj normalnih porazdelitev z različnimi M in σ.



slika. Normalna porazdelitev

7. 3. Kumulativna normalna porazdelitev

Integral

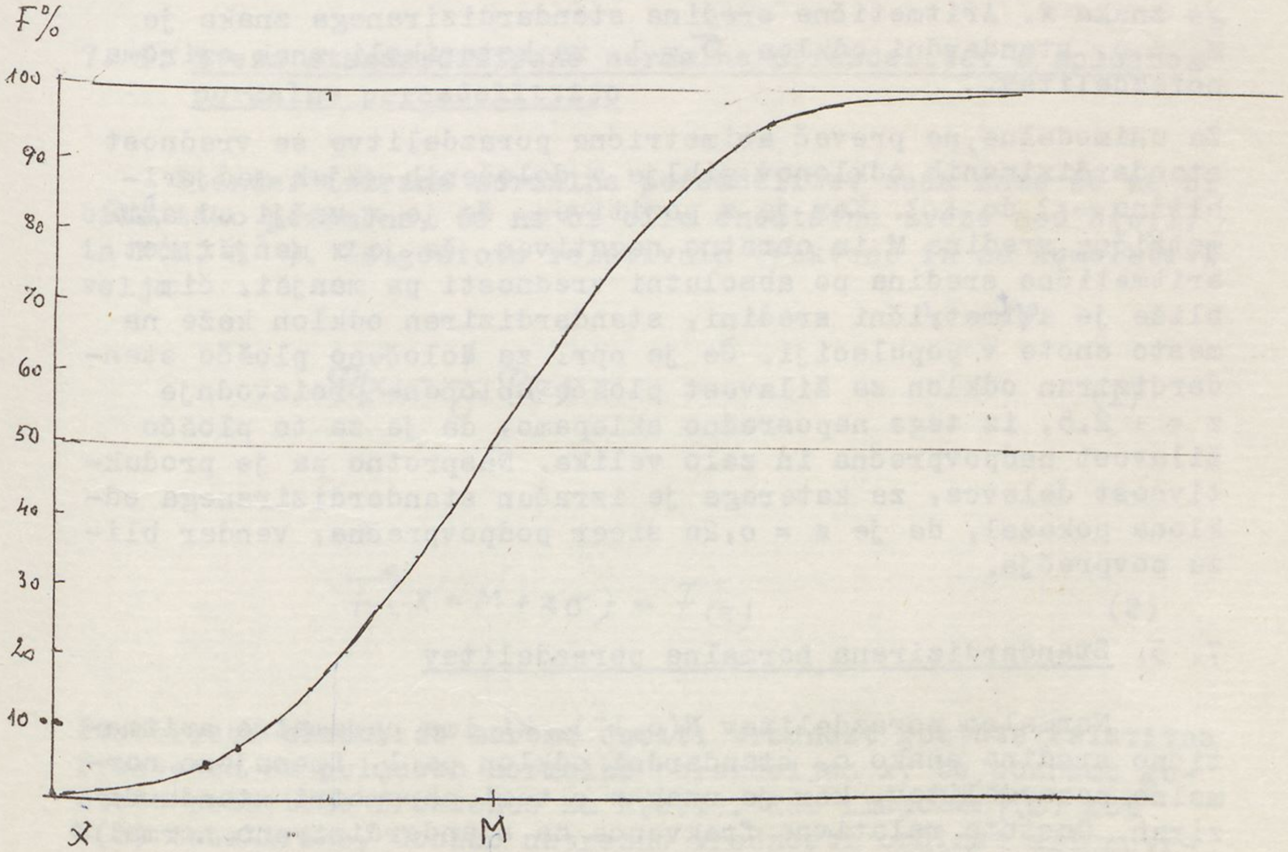
$$F^{\circ}(x) = \int_{-\infty}^{x} \varphi(x)\, dx \qquad (1)$$

je funkcija zgornje meje. Geometrijsko ponazarja ploščino lika, ki je omejen z abscisno osjo, ustreznim delom normalne krivulje in ordinato normalne porazdelitve za vrednost x. Vsebinsko zgornji integral nakazuje kumulativo relativne frekvence za normalno porazdelitev.

Iz definicje, gostote relativnih frekvenc in njenih značilnosti za normalno porazdelitev sledi, da je

$$F(x = +\infty) = 1 \qquad F(x = M) = 0{,}50 \tag{1}$$

V sliki 1 je nakazana krivulja kumulative relativnih frekvenc za normalno porazdelitev, ki ima značilno obliko črke S.



Idealna S krivulja je za normalne porazdelitve z različnimi M pomaknjena v desno ali levo, za porazdelitve z manjšim standardnim odklonom pa je strmejša kot za porazdelitve, za katere je standardni odklon večji.

7. 4. Standardiziran odklon

Standardiziran odklon z je odklon vrednosti x od ustrezne asimetrične sredine, merjen v standardnih odklonih.

$$z = \frac{x - M}{\sigma} \qquad (1)$$

Standardiziran znak z je nov znak, ki je izpeljan iz osnovnega znaka x. Aritmetična sredina standardiziranega znaka je $M_z = 0$, standardni odklon $\sigma_z = 1$, za katerikoli znak oziroma porazdelitev.

Za unimodalne, ne preveč asimetrične porazdelitve se vrednost standardiziranih odklonov giblje v določenih mejah med približno - 3 do + 3. Ker je z pozitiven, če je x večji od aritmetične sredine M in obratno negativen, če je x manjši kot aritmetična sredina po absolutni vrednosti pa manjši, čim bliže je aritmetični sredini, standardiziran odklon kaže na mesto enote v populaciji. Če je npr. za določeno ploščo standardiziran odklon za žilavost plošče določene proizvodnje z = + 2,5, iz tega neposredno sklepamo, da je za to ploščo žilavost nadpovprečna in zelo velika. Nasprotno pa je produktivnost delavca, za katerega je izračun standardiziranega odklona pokazal, da je z = 0,20 sicer podpovprečna, vendar blizu povprečja.

7. 5. Standardizirana normalna porazdelitev

Normalno porazdelitev $N(0,1^2)$, ki ima parametre aritmetično sredino enako 0, standardni odklon pa 1, imenujemo normalno porazdelitev, ker je znak x s temi parametri standardiziran. Gostota relativne frekvence za standardizirano normalno porazdelitev (SNO) je glede na zgornje parametre

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

kumulativna relativnih frekvenc pa

$$\overset{o}{F}(z) = \int_{-\infty}^{z} \varphi(z)\,dz \qquad (2)$$

7. 6. Zveza standardizirane normalne porazdelitev s splošno normalno porazdelitvijo

Standardizirana normalna porazdelitev sama zase še ne bi bila tako pomembna, če ne bi bilo enostavne zveze med N(o,1) in N(M; σ^2). Za gostoto relativnih frekvenc in za kumulativi velja

$$\varphi(x) = \frac{1}{\sigma} \varphi(z) \qquad (1)$$

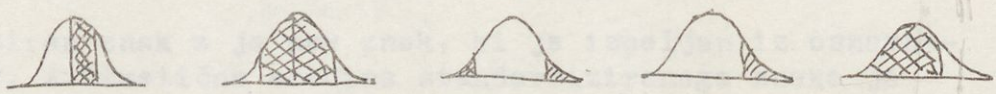
$$\overset{o}{F}(x = M + z\sigma) = \overset{o}{F}(z) \qquad (2)$$

Iz obrazca sledi, da moremo dobiti vrednost gostote relativne frekvence za poljuben normalno porazdeljen x, če poznamo gostoto relativne frekvence za N(o,1). Ker imamo *tako* $\varphi(z)$ kot F(z) tabelirane, dobimo ustrezne vrednosti ordinat ali površin pod poljubno normalno razdelitvijo enostavno tako, da za dani x in znanega M in σ izračunamo ustrezen z, preko tega pa najdemo v tablicah za N(o,1) z ustrezno vrednost.

7. 7. Tablice ordinat in površin za N(o,l)

V tabeli l imamo danih nekaj karakterističnih odnosov med z in različnimi površinami pod normalno porazdelitvijo .

Tabela l. Odnosi med z in karakterističnimi površinami

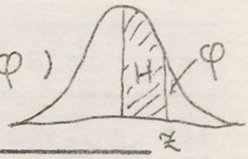


z	H	G	$\bar{G}$	P	$\bar{P}$
Ø'674	'25oo	'5ooo	'5ooo	'25oo	'75oo
1'ooo	'3413	'6826	'3174	'1587	'8413
1'645	'45oo	'9ooo	'1ooo	'o5oo	'95oo
1'96o	'475o	'95oo	'o5oo	'o25o	'975o
2'ooo	'4773	'9546	'o454	'o227	'9773
2'376	'49oo	'98oo	'o2oo	'o1oo	'99oo
2'576	'495o	'99oo	'o1oo	'oo5o	'995o
3'ooo	'4987	'9974	'oo2o	'oo13	'9987
3'o9o	'499o	'998o	'oo2o	'oo1o	'999o
3'291	'4995	'999o	'oo1o	'ooo5	'9995

Zgornja tabela vsebuje odnose med površinami in z za najznačilnejše vrednosti. Ti odnosi so važni kor orientacija in kot kritične vrednosti za sklepanje na osnovi vzorčenja.

Zveze med N(o,l) in poljubnimi normalnimi porazdelitvami pa daje podrobnejša tabela, za z v razmakih o,ol vrednosti ordinat in površin H (v razmaku od o do z).

Tabela 2: Normalna distribucija
(standardni odklon - z; površina - H; ordinata - φ)



$lo^2 z$	$lo^4 H$	$lo^4 \varphi$	$lo^2 z$	$lo^4 H$	$lo^4 \varphi$	$lo^2 z$	$lo^4 H$	$lo^4 \varphi$
oo	oooo	3989	35	1368	3752	7o	258o	3123
ol	oo4o	3989	36	14o6	3739	71	2612	3lol
o2	oo8o	3989	37	1443	3726	72	2642	3o79
o3	ol2o	3988	38	148o	3712	73	2673	3o56
o4	ol6o	3986	39	1517	3697	74	27o4	3o34
o5	o199	3984	4o	1554	3683	75	2734	3oll
o6	o239	3982	41	1591	3668	76	2764	2989
o7	o279	398o	42	1628	3653	77	2794	2966
o8	o319	3977	43	1664	3637	78	2823	2943
o9	o359	3973	44	17oo	3621	79	2852	292o
lo	o398	397o	45	1736	36o5	8o	2881	2897
11	o438	3965	46	1772	3589	81	291o	2874
12	o478	3961	47	18o8	3572	82	2939	285o
13	o517	3956	48	1844	3555	83	2967	2827
14	o557	3951	49	1879	3538	84	2996	28o3
15	o596	3945	5o	1915	3521	85	3o23	278o
16	o636	3939	51	195o	35o3	86	3o51	2756
17	o675	3932	52	1985	3485	87	3o79	2732
18	o714	3925	53	2o19	3467	88	3133	2685
19	o754	3918	54	2o54	3448	89	3133	2685
2o	o793	391o	55	2o88	3429	9o	3159	2661
21	o832	39o2	56	2123	3411	91	3186	2637
22	o871	3894	57	2157	3391	92	3212	2613
23	o91o	3885	58	219o	3372	93	3238	2589
24	o948	3876	59	2224	3352	94	3264	2565
25	o987	3867	6o	2258	3332	95	3289	2541
26	lo26	3857	61	2291	3312	96	3315	2516
27	lo64	3847	62	2324	3292	97	334o	2492
28	llo3	3836	63	2357	3271	98	3365	2468
29	1141	3825	64	2389	3251	99	3389	2444
3o	117o	3814	65	2422	323o	loo	3413	242o
31	1217	38o2	66	2454	32o9	lol	3438	2396
32	1255	379o	67	2486	3187	lo2	3461	2371
33	1293	3778	68	2518	3166	lo3	3485	2347
34	1331	3765	69	2549	3144	lo4	35o8	2323

$10^2 z$	$10^4 H$	$10^4 \varphi$	$10^2 z$	$10^4 H$	$10^4 \varphi$	$10^2 z$	$10^4 H$	$10^4 \varphi$
105	3531	2299	135	4115	1604	165	4505	1023
106	3554	2275	136	4131	1582	166	4515	1006
107	3577	2251	137	4147	1561	167	4525	0989
108	3599	2227	138	4162	1540	168	4535	0973
109	3621	2203	139	4177	1518	169	4545	0957
110	3643	2179	140	4192	1497	170	4554	0941
111	3665	2155	141	4207	1476	171	4564	0925
112	3686	2131	142	4222	1456	172	4573	0909
113	3708	2107	143	4236	1435	173	4582	0893
114	3729	2083	144	4251	1415	174	4591	0878
115	3749	2059	145	4265	1394	175	4599	0863
116	3770	2036	146	4279	1374	176	4608	0848
117	3790	2012	147	4292	1354	177	4616	0833
118	3810	1989	148	4306	1334	178	4625	0818
119	3830	1965	149	4319	1315	179	4633	0804
120	3849	1942	150	4332	1295	180	4641	0790
121	3869	1919	151	4345	1276	181	4649	0775
122	3888	1895	152	4357	1257	182	4656	0761
123	3907	1872	153	4370	1219	183	4664	0748
124	3925	1849	154	4382	1219	184	4671	0734
125	3944	1827	155	4394	1200	185	4678	0721
126	3962	1804	156	4406	1182	186	4686	0707
127	3980	1781	157	4418	1163	187	4693	0694
128	3997	1759	158	4430	1145	188	4700	0681
129	4015	1736	159	4441	1127	189	4706	0669
130	4032	1714	160	4452	1109	190	4713	0656
131	4049	1692	161	4463	1092	191	4719	0644
132	4066	1669	162	4474	1074	192	4726	0632
133	4082	1647	163	4485	1057	193	4732	0620
134	4099	1626	164	4495	1040	194	4738	0608

Tabela 2(nadaljevanje)

$lo^2 z$	$lo^4 H$	$lo^4 \varphi$	$lo^2 z$	$lo^4 H$	$lo^4 \varphi$	$lo^2 z$	$lo^4 H$	$lo^4 \varphi$
195	4744	o596	225	4873	o317	275	497o	oo91
196	475o	o584	226	4881	o31o	28o	4974	oo79
197	4756	o573	227	4884	o3o3	285	4978	oo69
198	4762	o562	228	4887	o297	29o	4981	oo6o
199	4767	o551	229	489o	o29o	295	4984	oo51
2oo	4773	o54o	23o	4893	o283	3oo	4987	oo44
2o1	4778	o529	231	4896	o277	3o5	4989	oo38
2o2	4783	o519	232	4898	o271	31o	499o	oo33
2o3	4788	o5o8	233	49o1	o264	315	4992	oo28
2o4	4793	o498	234	49o4	o258	32o	4993	oo42
2o5	4798	o488	235	49o6	o252	325	4994	oo2o
2o6	48o3	o478	236	49o9	o246	33o	4995	oo17
2o7	48o8	o468	237	4911	o241	335	4996	oo15
2o8	4812	o459	238	4913	o235	34o	4997	oo12
2o9	4817	o449	239	4916	o229	345	4997	oo1o
21o	4821	o44o	24o	4918	o224	35o	4998	ooo9
211	4826	o431	241	492o	o219	355	4998	ooo7
212	483o	o422	242	4922	o213	36o	4998	ooo6
213	4834	o413	243	4925	o2o8	365	4999	ooo5
214	4838	o4o4	244	4927	o2o3	37o	4999	ooo4
215	4842	o396	245	4929	o198	375	4999	ooo4
216	4846	o387	246	4931	o194	38o	4999	ooo3
217	485o	o379	247	4932	o189	385	4999	ooo2
218	4854	o371	248	4934	o184	39o	5ooo	ooo2
219	4857	o363	249	4936	o18o	395	5ooo	ooo2
22o	4861	o355	25o	4938	o175			
221	4865	o347	255	4946	o155			
222	4868	o339	26o	4963	o136			
223	4871	o332	265	496o	o119			
224	4875	o325	27o	4965	o1o4			

Da se izognemo decimalkam, so tabelarne vrednosti $lo^2 z$, $lo^4 H$ in $lo^4 \varphi$, kar je treba pri končnih rezultatih upoštevati.

S tablicami moremo iz danega z najti H(z) in φ(z) in obratno iz znanega H poiščemo z(H) in φ(H).

PrimerVI. $z = 1{,}79$. Iz tablic odčitamo $H = 0{,}4633$, $\varphi = 0{,}0804$.

Primer VII. $H = 0{,}397$; $z(H) = 1{,}26$; $\varphi(H) = 0{,}179$. Tabelirane so površine H v razmaku $0-^{+}z$.

7. 8. Prilagoditev normalne porazdelitve stvarni porazdelitvi

Stvarni porazdelitvi prilagodimo normalno porazdelitev tako, da poiščemo frekvenčno porazdelitev,ki ima iste osnovne parametre kot stvarna porazdelitev. Ti osnovni parametri so: obseg populacije N, aritmetična sredina M in standardni odklon σ. Če je stvarna populacija porazdeljena normalno ali ne odstopa dosti od normalnosti, se frekvence prilagojene normalne frekvenčne porazdelitve ne razlikujejo bistveno od stvarnih. Če pa stvarna porazdelitev ni normalna, so razlike v frekvencah znatne kljub temu, da se stvarna in teoretična porazdelitev ujemata v treh osnovnih parametrih: N, M, σ.

7. 9. Metoda površin za prilagoditev normalne porazdelitve stvarni frekvenčni porazdelitvi

Po metodi površin prilagodimo stvarni frekvenčni porazdelitvi normalno porazdelitev po naslednjem postopku:

a) za stvarno porazdelitev izračunamo ali ocenimo N, M in σ

b) za meje razredov $x_{k,min}$ izračunamo po obrazcu

$$z_{k,min} = \frac{x_{k,min} - M}{\sigma} \qquad (1)$$

standardizirane odklone

c) iz tablic za normalno porazdelitev poiščemo $z_{k,min}$ ustrezne površine H(z). Pri tem upoštevamo, da je $H(-z) = -H(z)$

d) da dobimo kumulativo relativnih frekvenc za prilagojeno normalno porazdelitev, dobljenim vrednostim H(z) prištejemo o,5: $F^{o}(z) = 0{,}5o + H(z)$

e) vrsto $F^{o}(z)$ pomnožimo z N. Tako dobimo vrsto kumulativnih frekvenc $F(x_{k,min})$;

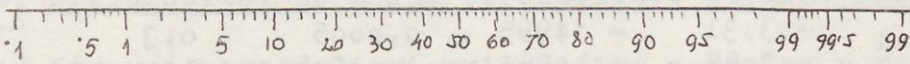
f) če poiščemo razlike med dvema zaporednima členoma v vrsti kumulativnih frekvenc, dobimo vrsto frekvenc prilagojene normalne frekvenčne porazdelitve f'.

Za primer vzemimo frekvenčno porazdelitev premerov 5oo glav zakovic. Zanjo je $\bar{x}$ = 13,426 mm, s = o,115 mm, i = o,o5 mm.

k,min	f_k	$z_{k,min}$	H(z)	$F^o(z)$	n.F^o	F = f'
						o,3
13,o45		- 3,31	- ·4995	o,ooo5	o,3	
	1					o,7
13,o95		- 2,88	- ·498o	·oo2o	1,o	
	4					2,7
13,145		- 2,44	- ·4827	·oo73	3,7	
	4					7,4
13,195		- 2,o1	- ·4778	·o222	11,1	
	18					18,o
13,245		- 1,57	- ·4418	·o582	29,1	
	38					34,5
13,295		- 1,14	·3729	·1271	63,6	
	56					57,4
13,345		- o,7o	- ·258o	·242o	121,o	
	69					75,8
13,395		- o,27	- ·1o64	·3936	196,8	
	96					87,o
13,445		o,17	+ ·o675	·5675	283,8	
	72					79,1
13,495		o,6o	+ ·2257	·7257	362,9	
	68					61,4
13,545		1,o3	+ ·3485	·8485	424,3	
	41					4o,3
13,595		1,47	+ ·4292	·9292	464,6	
	18					21,1
13,645		1,9o	+ ·4713	·9713	485,7	
	12					9,5
13,695		2,34	+ ·49o4	·99o4	495,2	
	2					3,4
13,745		2,77	+ ·4972	·9972	498,6	
	1					1,1
13,795		3,21	+ ·4993	·9993	499,7	
						o,3
	n = 5oo					5oo

7. lo. Verjetnostna skala

Ob linearno skalo za standardiziran z odklon nanešene vrednosti oziroma skala za kumulativo relativnih frekvenc $F^o(z)$ imenujemo verjetnostno skalo. Kot je razvidno, je verjetnostna skala ni linearna, temveč je okrog z = o najgostejša, razmak pa tem širši čim bolj se oddaljujemo od z = o v eno ali drugo smer. V ostalem je verjetnostna skala zaradi simetričnosti normalne porazdelitve simetrična na vrednost z = o.



slika Verjetnostna skala

7. 11. Linearno verjetnostna mreža in linearno verjetnostni grafikon

Če na absciso nanesemo linearno skalo za znak x, na ordinato pa verjetnostno skalo, dobimo linearno verjetnostno mrežo, (glej sliko). Ker je verjetnostna skala linearna v znaku z, je zaradi linearne zveze med z in x

$$z = \frac{1}{\sigma}x - \frac{M}{\sigma} \qquad (1)$$

za vsako normalno porazdelitev grafična slika za normalno porazdelitev premica, ki gre skozi točko (M,o) in ima smerni koeficient $1/\sigma$. Premica je torej tembolj strma, čim manjši je standardni odklon in tembolj položna čim večji je standardni odklon.

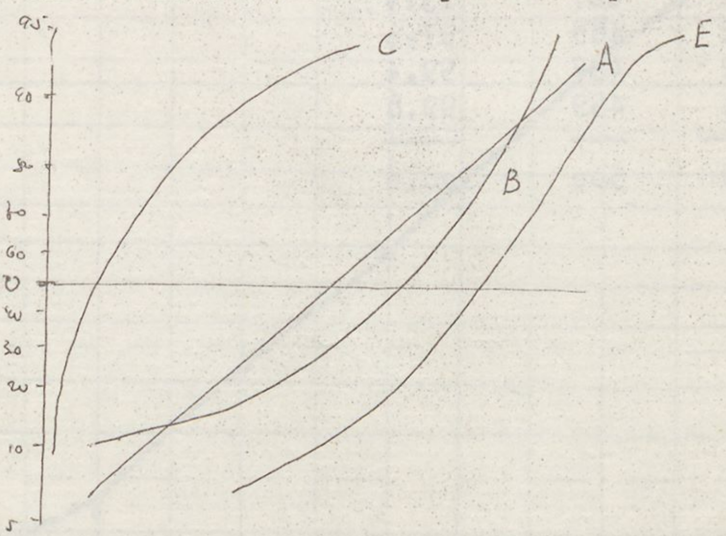
Ker je linearni z - skali prirejena verjetnostna skala F^o, je iz verjetnostnega grafikona za normalno porazdelitev, za katero je vrisana ustrezna premica, možna direktna zveza med vrednostmi x in kumulativo relativnih frekvenc (glej sliko v 7. 12.

7. 12. Kumulative relativnih frekvenc stvarnih porazdelitev v verjetnostnem grafikonu

Če narišemo kumulativo relativnih frekvenc za stvarno frekvenčno porazdelitev v verjetnostni grafikon, je narisana lomljena črta tem bolj približana premici ali potekala linearno, čimbolj je stvarna porazdelitev podobna normalni. Taki grafikoni torej služijo za analizo podobnosti oziroma odklonov stvarnih porazdelitev od normalnosti.

Ko imamo v verjetnostnem grafikonu vrisano stvarno porazdelitev moremo vrisati premico, za katero smatramo, da se stvarni črti najbolj prilagaja, ta premica predstavlja oceno slike stvarni porazdelitvi prilagojene normalne porazdelitve.

Ker je normalna porazdelitev ponazorjena v verjetnostnem grafikonu s premico, odklone od normalnosti (asimetrija, sploščenost) iz verjetnostnega grafikona lažje analiziramo, kot iz drugih prikazov. Slika kaže črte v verjetnostnem grafikonu za asimetrične in sploščene porazdelitve.

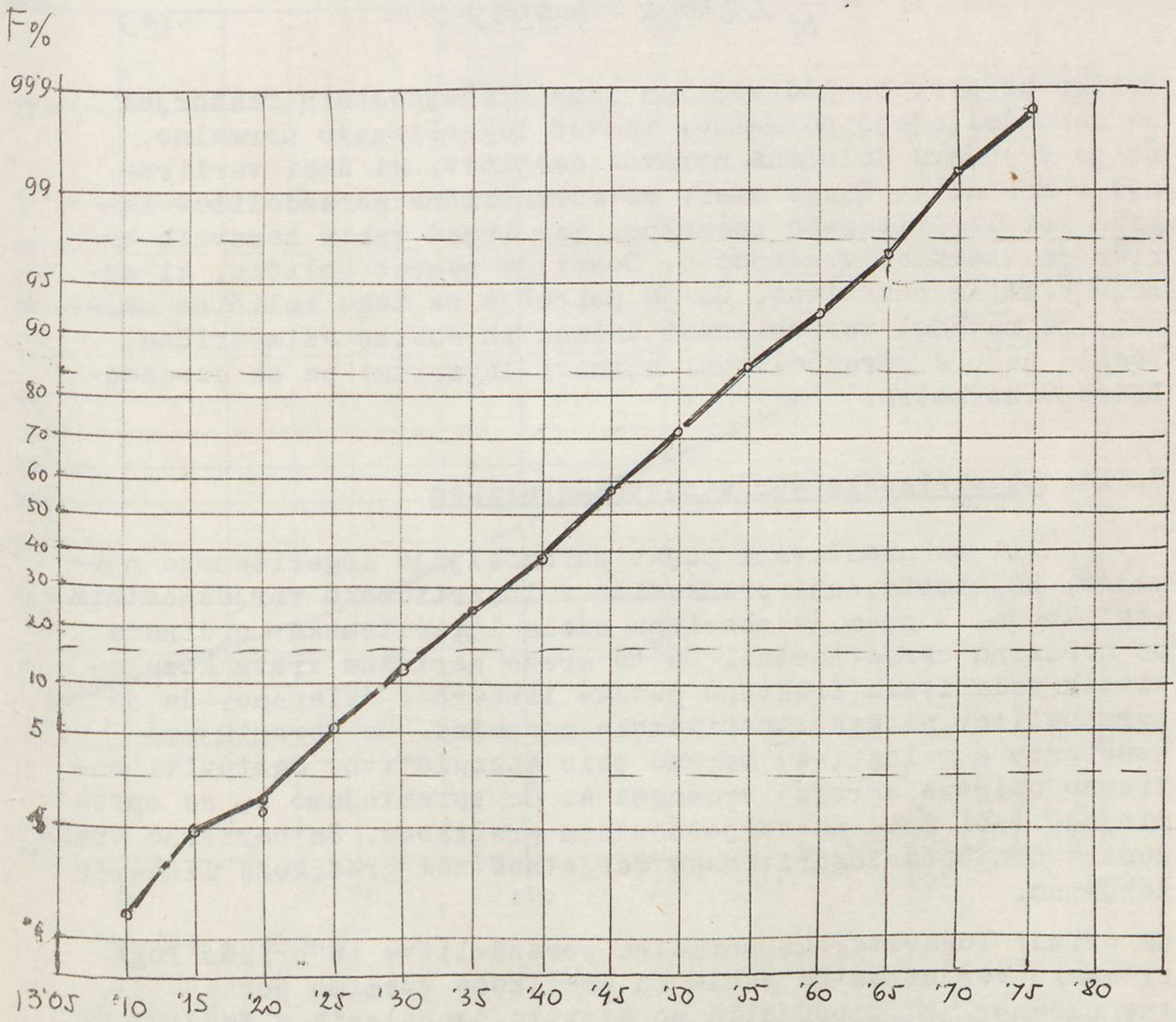


A = normalna
B = asimetrična v levo
C = asimetrična v desno
E = koničasta

Slika . Normalna, asimetrične in sploščene porazdelitve na verjetnostnem grafikonu.

Za primer uporabe verjetnostnega grafikona vzemimo frekvenčno porazdelitev N = 5oo zakovic po premeru glave: zanje je frekvenčna porazdelitev f_k, kumulativa F_k in kumulativa relativnih frekvenc Fk%, izražena v odstotkih, naslednja:

x mm	f	F	F%
13,o5-13,o9	1	o	o
13,lo-13,o4	4	1	o,2
13,15-13,19	4	5	1,o
13,2o-13,24	18	9	1,8
13,25-13,29	38	27	5,4
13,3o-13,34	56	65	13,o
13,35-13,39	69	121	24,2
13,4o-13,44	96	19o	38,o
13,45-13,49	72	286	57,2
13,5o-13,54	68	358	71,6
13,55-13,59	41	426	85,2
13,6o-13,64	18	467	93,4
13,65-13,69	12	485	97,o
13,7o-13,74	2	497	99,4
13,75-13,79	1	499	99,8
	5oo	5oo	loo,o



Normalno-verjetnostni grafikon za premere glav zakovic iz frekvenčne porazdelitve na strani 64

7. 13. Logaritemsko normalna porazdelitev

Če se porazdeljujejo v normalni porazdelitvi logx ali log(x-a) pravimo, da se x porazdeljuje logaritemsko normalno. Za tako porazdelitev je parameter M v normalni porazdelitvi enak M = logG, pri čemer je G geometrijska sredina iz x ali (x-a) varianca σ^2 pa je

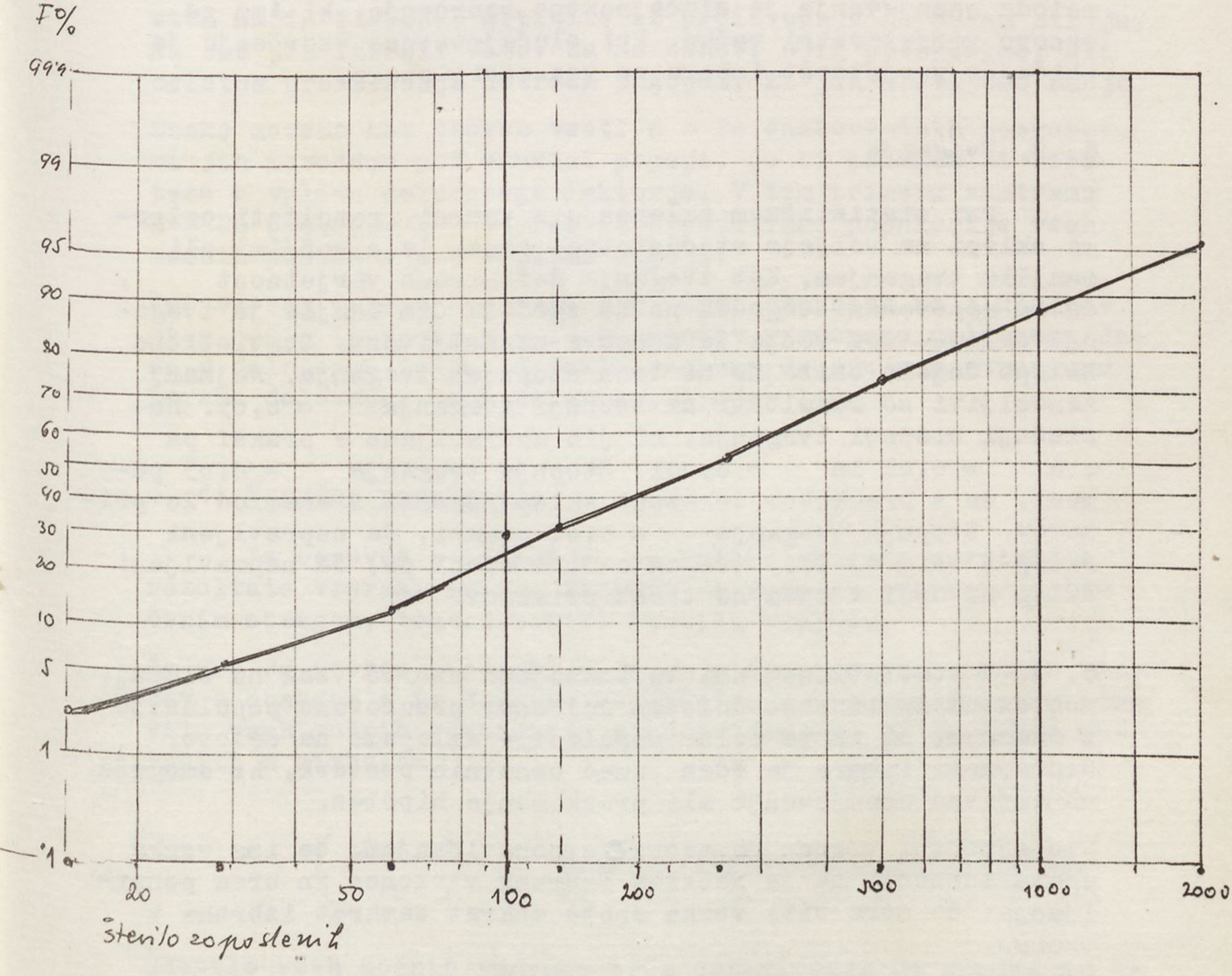
$$\sigma^2 = \frac{1}{N} \sum (\log x - \log G_x)^2 \qquad (1)$$

Veliko pojavov se pod vplivom samo slučajnostnih faktorjev ne porazdeljujejo normalno, temveč logaritemsko normalno. Če je v pojavu določena naravna omejitev, ki duši variiranje v eno ali v drugo smer, se asimetrična porazdelitev izkaže kot logaritemsko normalna. Ena izmed takih naravnih barier je vsekakor vrednost o, Dosti je namreč količin, ki morejo biti le pozitivne. Če je poprečje za take količine majhno, je navzdol variabilnost dušena in dobimo asimetrične, včasih celo J porazdelitve, njihovi logaritmi pa se porazdeljujejo normalno.

7. 14. Logaritemsko-verjetnostni grafikon

Ali se proučevani pojav porazdeljuje logaritemsko normalno, najenostavneje preskusimo z logaritemsko verjetnostnim grafikonom. V njem je abscisna skala logaritemska, ordinata pa normalno verjetnostna. Če v to mrežo narisana vrsta kumulativnih relativnih frekvenc poteka linearno, sklepamo, da je porazdelitev pojava logaritemsko normalna. Če spreminjamo konstanto a v log(x-a) moremo celo špekulativno ugotoviti barierno oziroma skrajno vrednost a. Če spreminjamo a, se spreminjajo tudi črte na verjetnostnem grafikonu. Za barierno vrednost a ima črta logaritemsko verjetnostnem grafikonu linearno tendenco.

Za primer logaritemsko normalne porazdelitve in prikaz logaritemsko verjetnostne skale in grafikona vzemimo porazdelitev industrijskih podjetij po številu zaposlenih v SFRJ v letu 1965.



Logaritmično-verjetnostni grafikon industrijskih podjetij v SFRJ v 1964 po številu zaposlenih

8. VZORČENJE

8. 1. Vzorčenje

Vzorčenje je metoda statistične indukcije-sklepanja iz dela na celoto. Z vzorčenjem ocenjujemo parametre populacije ali preskušamo hipoteze o populacijah. Objektivna metoda ocenjevanja je slučajnostno vzorčenje, ki ima za osnovo verjetnostni račun. Pri slučajnostnem vzorčenju je običajno verjetnost izbora za vse enote enaka.

8. 2. Tveganje

Pri statističnem sklepanju z vzorci, rezultati oziroma sklepi ne veljajo stoodstotno, ampak le z večjimi ali manjšim tveganjem. Kot tveganje definiramo verjetnost , da se napovedani dogodek ne bo zgodil. Čim manjše je tveganje sklepa, tem večja je njegova zanesljivost. Statistične sklepe dajemo običajno na treh stopnjah tveganja. Najmanj zanesljivi so rezultati na stopnji tveganja = o,o5. Naslednji stopnji tveganja, ki jih uporabljamo v praksi pa sta = o,ol in = o,ool. Stopnja tveganja = o,o5 pomeni, da s preskusom dokazani sklep ne drži v enem od 2o primerov. Stopnja tveganja = o,ol pomeni, da napravljeni sklep v enem od loo, tveganje = o,ool pa, da napravljeni sklep ne drži v enem od tisoč primerov.

8. 3. V statističnem smislu imenujemo vzorec vsak na slučajnosten način izbrano določen del enot proučevane populacije z namenom, da iz te delne populacije sklepamo na celoto. Slučajnost izbora je eden izmed osnovnih postavk, ki omogoča objektivno ocenjevanje ali preskušanje hipotez.

Slučajnostni vzorec je vzorec s ponavljanjem, če ima vsaka enota možnost, da je večkrat izbrana v vzorec in brez ponavljanja, če more biti vsaka enota enkrat samkrat izbrana v vzorec.

Po obsegu so vzorci mali, če število enot v vzorcu ne presega par desetin enot (ne več kot loo) in veliki, če ima vzorec nekaj sto ali tudi tisoč enot. Stroge meje med malimi in velikimi vzorci ni možno potegniti in je ta prehod zvezen. Čim večji je vzorec, tembolj veljajo zakonitosti za velike vzorce. Čim manjši je vzorec, tembolj pridejo do izraza omejitve in značilnosti malih vzorcev.

Kot vzorec moremo npr. vzeti skupnost n = 5o na slučajnosten način izbranih artiklov iz proizvodnje na nekem stroju, ki smo jih izbrali zato, da na osnovi njih preskusimo, ali celotna proizvodnja ustreza pogojem, ki jih stavljamo nanjo.

Enako moremo kot vzorec vzeti n = 2o enakovrstnih poskusov, ki jih izvedemo pod enakimi pogoji, da bi preskusili hipotezo o vplivu določenega faktorja. V tem primeru smatramo n = 2o poskusov kot vzorec iz hipotetične populacije vseh možnih poskusov pod enakimi pogoji.

Kot vzorec moremo smatrati npr. tudi skupnost na slučajnosten način izbranih n = looomomentov v razmaku delovnega časa v določenem tednu. S takim vzorcem ocenjujemo strukturo izrabe delovnega časa delavcev, strojev itd.

8. 4. Vzorčna populacija

Iz osnovne populacije, na katero skušamo posplošiti rezultate vzorca, moremo izbrati ne samo en vzorec z določenim obsegom, temveč veliko število vzorcev.

Če gre za vzorce s ponavljanjem, je število vseh možnih vzorcev z obsegom n iz osnovne populacije z obsegom N enako številu vseh možnih kombinacij s ponavljanjem. To je možno izraziti s simbolom

$$\binom{N+n-1}{n} = \frac{N.(N+1).(N+2).(N+3).\cdot.\cdot.(N+n-1)}{1.\quad 2.\quad 3.\quad 4.\ .\ .\ .\ n}$$

Število vseh možnih vzorcev brez ponavljanja je manjša in je enako številu vseh možnih kombinacij s ponavljanjem po n elementov iz populacije z N elementi. Izraženo s simbolom je to število

$$\binom{N}{n} = \frac{N.(N-1).(N-2).....(N-n+1)}{1.\ 2.\ 3.\ \ \ \ n}$$

Tako število vseh možnih vzorcev s ponavljanjem kot brez ponavljanja je že razmeroma majhne populacije in vzorcev zelo veliko. V praktičnih situacijah, kjer je populacija običajno obsežna, pa tudi vzorci imajo razmeroma veliko število enot, pa je število vseh možnih vzorcev praktično neomejeno.

Za vsak vzorec je možno za določen znak izračunati različne parametre - statistike. Tako moremo iz vzorcev po n = 5o enot izračunati odstotek defektnih artiklov, poprečen premer, poprečno žilavost plošč itd. Ker so v vsakem vzorcu izbrane različne enote, pa so tu izračunani parametri od vzorca do vzorca različni t.j. variirajo. Skupnost vseh vzorcev pa tudi skupnost vseh vrednosti določenega parametra za posamezne vzorce imenujemo vzorčno populacijo. Jasno je, da je nemogoče praktično skonstruirati vse možne vzorce in zanje izračunati določene parametre - statistike. Vendar verjetnostni račun in teorija vzorčenja dajeta zakonitosti o vzorčnih populacijah neposredno.

8. 5. Vzorčne porazdelitve

Porazdelitev določenih vrednosti vzorcev v populaciji vseh možnih vzorcev imenujemo vzorčno porazdelitev. Tako bi dobili vzorčno porazdelitev aritmetičnih sredin vzorcev, če bi sestavili iz aritmetičnih sredin za vse vzorce frekvenčno porazdelitev. Direktno sestavljanje vzorčnih porazdelitev je praktično neizvedljivo. Vzorčne porazdelitve marsikaterih vzorčnih izrazov pa je možno dobiti z uporabo verjetnostnega računa.

8. 6. Varianca in standardna pogreška vzorčnih izrazov

Za vsako vzorčno porazdelitev, ki jo poznamo, moremo izračunati varianco - Var-g .

Standardni odklon vzorčnega izraza g, ki je kvadratni koren iz variance, imenujemo standardna pogreška in jo zaznamujemo z SE(g). Standardna pogreška ocene je osnovni pokazatelj zanesljivosti ocen in osnova za izračun ocen za velike vzorce.

8. 7. Zakonitosti ocen aritmetičnih sredin iz vzorcev iz normalno porazdeljenih populacij

Če se znak x v osnovni populaciji porazdeljuje normalno $N(M,\sigma^2)$, se aritmetične sredine $\bar{x}$ v populaciji vseh možnih vzorcev porazdeljujejo normalno s parametri $E(\bar{x}) = M$ in $Var\bar{x} = \sigma^2/n$ ali če napišemo v standardni simboliki $x = N(M; \sigma^2/n)$.

Iz zgornjih zakonitosti spoznamo, da se aritmetične sredine iz vzorcev porazdeljujejo okrog prave vrednosti aritmetične sredine in da je variabilnost teh sredin tem manjša, čim manjši je standardni odklon v osnovni populaciji in čim večji je vzorec. Če je npr. aritmetična sredina za določen znak M=5o varianca v osnovni populaciji $\sigma^2 = 100$, vzorec pa ima n = loo enot, je varjanca za porazdelitev ocen vzorcev, $Var\bar{x} = 10^2/100 = 1$. Z verjetnostjo $1 - \alpha = 0{,}95$ leži ocena v razmaku $M - 1{,}96\ SE(\bar{x}) = 50 - 1{,}96.1 = 48$ do $M + 1{,}968\ E(\bar{x}) = 50 + 1{,}96.1 = 52$.

Iz numeričnega primera spoznamo, da je variabilnost že pri aritmetični sredini že pri vzorcih z n = loo zelo majhna.

Iz zakonitosti, ki slede iz verjetnostnega računa, povzamemo, da je zelo velika verjetnost, da bo iz vzorca ocenjena aritmetična sredina ležala v neposredni okolici prave aritmetične sredine.

8. 8. Zakonitosti za ocene sredin pri vzorčenju brez ponavljanja

Za enostaven slučajnosten vzorec brez ponavljanja, veljajo za porazdelitev sredin vzorcev $\bar{x}$ vse zakonitosti, kot za vzorec s ponavljanjem, le da varianca $Var\ \bar{x}$ ni enaka σ^2/n temveč

$$\text{Var } \bar{x} = \frac{S_x^2 (N - n)}{n.N} \qquad (1)$$

Pri tem je S_x^2, ki je mera variacije v osnovni populaciji, definirana z izrazom

$$S_x^2 = \frac{\sum (x-M)^2}{N-1} \qquad (2)$$

Razlika med σ_x^2 in S_x^2 je praktično nepomembna, ker je že za ne preveč velike N je razlika med N in N-1 nepomembna, pač pa je ta razlika teoretično utemeljena. Če primerjamo Var $\bar{x}$ za vzorec s ponavljanjem opazimo, da je bistvena razlika le v faktorju $(N-n)/N$. Ta faktor je tembolj različen od 1, čim večji je vzorčni delež f. Vzorčni delež f = n/N pove, kakšen del celotne populacije je vključen v vzorec. Za neomejene populacije(N = oo) je korekturni faktor 1.

Vzemimo za primer skupino N = 2ooo plošč,za katero poznamo aritmetično sredino M = 15 mm. Zanjo poznamo tudi $S_x = 2$. Za vzorčno populacijo aritmetičnih sredin, ki jih izračunamo iz vzorcev brez ponavljanja, na osnovi vzorcev z n = 2oo enotami, je po zgornjih zakonitostih $M_{\bar{x}} = 15$ in

$$\text{Var } \bar{x} = \frac{2{,}2}{2oo} \cdot \frac{2ooo - 2oo}{2ooo} = o{,}oo18$$

Kvadratni koren iz Var, SE pa je enak: $SE(\bar{x}) = \sqrt{o{,}oo18} = o{,}o425$

Če upoštevamo zakonitosti normalne porazdelitve in zakonitosti verjetnostnega računa, z verjetnostjo o,95 pričakujemo, da bo s slučajnostnim vzorčenjem brez ponavljanja izračunana ocena od prave poprečne debeline M = 15 oddaljenja manj kot $D(\bar{x})$ = 1,96,o,o425=o,o83 mm, ali relativno manj kot o,55 %: Aritmetična sredina, izračunana iz vzorca n = 2oo artiklov je torej razmeroma dobra ocena prave vrednosti poprečnega premera.

8. 9. Centralni limitni teorem

$\bar{x}$

Ne glede na to, kako se znak x v osnovni populaciji porazdeljuje, se aritmetične sredine, izračunane iz vzorcev, z večanjem obsega vzorca porazdeljujejo asimptotično normalno. To pomeni, da je distribucija sredin vzorcev porazdeljena v porazdelitvi, ki je normalni tembolj podobna, čimvečji je vzorec. Aritmetične sredine slučajnostnih vzorcev s ponavljanjem ali za vzorce iz hipotetičnih populacij porazdeljujejo asimptotično $M_{\bar{x}}$

$$\bar{x} \mathrel{\dot{=}} N(M; \sigma^2/n)$$ $E(\bar{y}) = M_y$

za aritmetične sredine slučajnostnih vzorcev brez ponavljanja pa

$$\bar{x} \mathrel{\dot{=}} N(M; \frac{S_x^2}{n} \cdot \frac{N-n}{N})$$

8. lo. Tehnika slučajnostnega izbora

Osnova vzorčenja so zakonitosti verjetnostnega računa. Te pa veljajo le pod predpostavko, da je vzorec izbran na slučajnosten način. Zato je treba pri vzorčenju zagotoviti slučajnost izbora. Ta se v večini primerov vzorčenja reducira na enako možnost izbora za vsako enoto. To enako možnost izobra moremo zagotoviti z loterijskim izborom. Iz žare, v kateri je toliko listkov, kolikor ima populacija enot izbiramo listke na slepo. Listi so oštevilčeni z zaporednimi številkami, ki ustrezajo zaporednim številkam enot v okvirju vzorčenja. V vzorec vključujemo enote, ki imajo zaporedne številke, ki so napisane na izbranih listkih.

Izbor delno poenostavimo, če namesto N listkov imamo v žari lo listkov, kock ali kroglic, ki so oštevilčene s številkami od o do 9. Štirimestno slučajnostno število npr. sestavimo če iz žare z desetimi kroglicami štirikrat zapovrstjo s ponavljanjem izberemo kroglico in zapovrstjo zapišemo številke, ki smo jih izbrali. Ta način je že toliko praktičen, da pride v poštev pri operativnem delu. Loterijski način nadomeste tablice slučajnostnih številk.

8. 11. Tablice slučajnostnih številk

Loterijski način slučajnostnega izbora nadomesti tablica slučajnostnih številk. V tablicah slučajnostnih številk so zapovrstjo zapisane številke, kot so bile npr. iz žare ali na kak drug način, ki zagotavlja slučajnost izbrane. Enkrat izbrane slučajnostne številke moremo torej uporabljati večkrat. Iz slujačnostnih številk dobimo večmestno število slučajnostno število, ki ustreza zaporednemu številu izbrane enote tako, da zgrupiramo skupine slučajnostnih številk. Če je nrp. N = 78oo enot, moramo iz tablic slučajnostnih številk sestaviti skupine po štiri slučajnostne številke. Tako moremo iz v nadaljevanju dane ilustrativnega dela sestavljene skupine po štiri zaporedne številke takole: 5354 9142 o847 5393 5418 65o5 itd. Od teh številk ne pride v poštev druga, ker je prva številka večja od 7 oziroma celotno število večje kot je število enot v populaciji.

Štirimestne grupe slučajnostnih številk moremo sestaviti tudi tako, da se pomikamo v seriji slučajnostnih številk za eno po eno:tako dobimo vrsto štirimestnih slučajnostnih številk 5354 3549 5491 4914 9142 142o 42o8 itd. Tako povečano bolje izkoristimo tablico slučajnostnih številk, ker moremo iz tablice enakega obsega dobiti več slučajnostnih številk. Tablico slučajnostnih številk moremo brati tudi od zadaj naprej, po kolonah od zgoraj navzdol ali od spodaj navzgor, v diagonali ali po kakšnem drugem vnaprej določenem sistematičnem redu, ki ne moti slučajnosti.

Poseben problem nastopi, če je število enot v populaciji tako, da je prvo v številu enot 1, 2, 3 ali 4. V tem primeru so tablice slabo izkoriščene, ker more odpasti tudi od 5o do 9o % formiranih slučajnostnih številk. Če je npr. N = 18oo, odpadejo vsa tista štirimestna števila, ki so večja kot 18oo ali z drugimi besedami, to je vsa tista, ki imajo prvo številko v štirimestni grupi večjo kot 1. Če izdelamo pravilo, da so vsa slučajnostna števila, ki imajo prvo številko sodo štejemo kot o za tiste, ki imajo liho pa 1, se slučajnostnemu principu nismo odrekli, uporabnost tablice pa smo povečali, ker so uporabne vse, tudi številke, ki se začnejo z drugimi številkami in ne samo z o in 1.

Iz operativnih števil[illegible], ki jih dobimo iz tablice slu- čajnostnih števil, pridemo do slučajnostnih števil za zgor- nji primer:

[illegible]	6812	9124	1247	4964	9540	5403	4952	[illegible]
[illegible]	o912	[illegible]	1249	o964	1540	o4o5	oo5o	[illegible]

Tako je uporabljenih vseh sedem slučajnostnih števil, medtem ko je [illegible] upoštevati samo dve.

Če je število enot v populaciji [illegible], ki se začne z dve, delimo pa [illegible], da za prvo število v slučajnostnega šte- vilu [illegible], [illegible] delimo s 3. Tako po- menijo [illegible] števila 1, 4, 7 in 2 številke 2, 5, 8 [illegible]. [illegible] vsaka šte- vilka enake možnosti, da bi bila vključena v tabor.

Za zgornjih 7 operativnih števil[illegible] v tem primeru bile slu- čajnostne številke naslednje:

[illegible]	o912	[illegible]	1249	4964	9540	6405	4o5o
[illegible]	o912	[illegible]	1248	1964	[illegible]	o4o5	1o5o

Če je prva številka [illegible] je 3, podobno, vzamemo [illegible] [illegible] 7. Za tretji razlogov [illegible] je prva številka 4, pa po enakem postopku [illegible] 5 pa [illegible] in 5 [illegible] 0, 1 [illegible] 3 in 8 [illegible] 3 in 4 in 9. [illegible]

Iz operativnih številk, ki jih dobimo iz tablice slučajnostnih številk pridemo do slučajnostnih številk za zgornji primer:

op. št.	o912	9124	1249	4964	964o	64o5	4o5o
sl. št.	o912	1124	1249	o964	164o	o4o5	oo5o

Tako so uporabnih vseh sedem slučajnostnih številk, medtem ko sta iz nekorigiranih uporabni samo dve.

Če je število enot v populaciji število, ki se začne z dve, delamo po sistemu, da za prvo številko v slučajnostnem številu vzamemo ostanek, če prvo številko delimo s 3. Tako pomenijo o številke o 3 6, 1 številke 1 4 7 in 2 številke 2 5 8. Številka devet odpade, ker drugače nima vsaka številka enake možnosti, da bi bila vključena v izbor.

Za zgornjih 7 operativnih številk bi v tem primeru bile slučajnostne številke naslednje:

op. št.	o912	9124	1249	4964	964o	64o5	4o5o
sl. št.	o912	izp.	1248	1964	izp.	o4o5	1o5o

Če je prva številka obsega populacije 3, podobno vzamemo ostanke deljenja s 4, o 4, 1,5, 2 6, 3 7. Iz istih razlogov kot zgoraj pa 8 in 9 odpadeta. Če je prva številka 4, pa po enakem postopku z deljenjem s 5 pa velja: o in 5 dasta o, 1 in 6 dasta 1,2 in 7 dasta 2, 3 in 8 dasta 3 in 4 in 9 dasta 4.

Slučajnostna števila

5354	9142	0847	5393	5416	6505	7156	5634	9703	6221
0905	6986	9396	3975	9255	0537	2479	4589	0562	5345
1420	0470	8679	2328	3939	1292	0406	5428	3789	2882
3218	9080	6604	1813	8209	7039	2086	3386	4437	3798
9697	8431	4387	0622	6893	8788	2320	9358	5904	9539
0912	4964	0502	9683	4635	2861	2876	1273	7870	2030
4636	7072	4868	0601	3894	7182	8417	2367	7032	1003
2515	4734	9878	6761	5636	2949	3979	8650	3430	0635
5964	0412	5012	2369	6461	0678	3693	2928	3740	8047
7848	1523	7904	1521	1455	7089	8094	9872	0898	7174
5192	2571	3543	0707	3434	6818	5729	8614	2498	4129
8438	8325	9886	1805	0226	2310	3675	5058	2515	2388
8166	6349	0319	5436	6838	2460	6433	0644	7428	8556
9153	8263	6504	2562	1160	1526	1816	9690	1215	9590
6051	3525	4048	0382	4224	7148	2859	6526	5340	4064
5407	2818	0520	5941	1740	5140	9844	2847	1502	0763
0469	0435	2858	7116	3297	8454	5146	9803	1694	7949
7805	8428	5745	8141	8465	4795	1895	4487	2323	1068
7294	1214	0170	9643	7891	7304	8278	2315	7139	5594
5430	2843	8903	3828	1717	6312	0384	6552	1200	7264
1017	0106	1414	9736	3886	4753	3589	3864	0073	3626
0858	1727	3020	1831	2878	2838	1319	2199	6457	5798
8396	8903	2156	1031	6182	5094	1931	9188	1672	1510
7813	4209	5295	0605	9080	6940	9657	3423	2191	3636
2712	2516	0968	7526	2176	4057	9023	2327	4311	0281
7141	7871	2878	2990	3907	8375	6005	9452	7702	9468
2418	9661	0436	1223	9708	9354	0707	4238	0756	2190
5230	6208	2742	1087	9639	6813	1963	2620	8913	7777
3517	1376	7866	6584	6381	0218	1101	3192	5965	5250
0319	0951	3976	6372	3518	1859	9038	3474	5150	3621
7526	7460	5644	8640	0643	0916	3238	0177	2592	0264
5172	1998	6030	6677	8827	7821	9933	9523	4563	7391
2023	0950	1896	4729	1789	1111	1157	0266	0438	7535
3632	3816	4575	0738	4923	4131	0819	3361	3992	6702
0761	2838	6166	8534	5353	5737	1204	2325	2036	4714

8. 12. Točkovna ocena

Točkovna ocena parametra je dana z eno samo vrednostjo, ki jo dobimo iz podatkov slučajnostnega vzorca. Točkovna ocena g je nepristranska ocena, če je aritmetična sredina ocen M(g), izračunanih iz vseh možnih vzrocev, enaka ocenjevanemu parametru G. V terminologiji verjetnostnega računa izraženo pomeni, da je g nepristranska ocena parametra G, če je matematično upanje ocen M(g), ki je slučajnostna spremenljivka, enako ocenjevanemu parametru G. V drugem primeru pa je ocena pristranska. Pristranske ocene, ki imajo lastnost, da so asimptotično nepristranske, imenujemo dosledne ocene. Asimptotična nepristranskost pomeni, da je stopnja pristranosti tem manjša, čim večji je vzorec. Ocena je torej nepristranska, če velja

$$E(g) = G \qquad (1)$$

pristranska pa je, če je

$$E(g) \neq G \qquad (2)$$

$$d = E(g) - G \qquad (3)$$

je pristranost ocene;
za dosledne ocene velja

$$\lim_{n = \infty} E(g) = G \qquad (4)$$

Iz teorije vzorčenja povzamemo, da je $\bar{x} = \frac{1}{n} S\, x$ nepristranska ocena za pravo aritmetično sredino M, ker velja $E(\bar{x}) = M$. Ocena variance, ki jo dobimo iz podatkov slučajnostnega vzorca po definicijskem obrazcu

$$\hat{s}^2 = \frac{1}{n} S(x - \bar{x})^2 \qquad (5)$$

pa je pristrana, a dosledna ocena prave vrednosti variance, ker je

$$E(\hat{s}^2) = \frac{n-1}{n} \sigma^2 \qquad (6)$$

8. 13. Intervalna ocena

Razen točkovno z eno samo vrednostjo, dajemo ocene parametrov z vzorčenjem tudi z intervalom. Intervalna ocena določenega parametra G je dana z razmakom zaupanja. Razmak zaupanja za oceno parametra G je razmak, v katerem z običajno veliko verjetnostjo pričakujemo, da se nahaja prava vrednost parametra G. Razmake zaupanja dajemo z običajnimi stopnjami tveganja, najobičajnejše na stopnji tveganja $\alpha = 0{,}05$. Če ocenjujemo razmak zaupanja dvostransko, je razmak zaupanja določen s spodnjo G_s in zgornjo mejo zaupanja G_z. V tem primeru velja verjetnostna enačba

$$P(G_s < G < G_z) = 1 - \alpha = 1 - 2P \qquad (1)$$

Včasih ocenjujemo parametre tudi enostransko tako, da iščemo samo spodnjo mejo nad katero se z veliko verjetnostjo $1 - \alpha$ nahaja ocenjevani parameter ali samo z zgornjo mejo, pod katero se nahaja prava vrednost parametra z, s tveganjem α ustrezno verjetnostjo $1-\alpha$,

Ker za ocene obravnavanih parametrov: $\bar{x}$, p%, s in kv % velja, da se porazdeljujejo normalno, če je vzorec dovolj velik, ~~v tem primeru~~ računamo intervalne ocene za te primere po splošnem obrazcu

$$g - z_{\alpha/2}\, se(g) < G < g + z_{\alpha/2}\, se(g)$$

$z_{\alpha/2}$ zavisi od stopnje tveganja α. Za običajno stopnjo tveganja $\alpha = 0{,}05$ je $z_{\alpha/2} = 1{,}96$.

8. 14. Ocenjevanje aritmetične sredine z velikimi vzorci

Ker velja centralni limitni teorem, moremo vzeti iz vzorca izračunano aritmetično sredino

$$\bar{x} = \frac{1}{n} Sx \qquad (1)$$

kot dobro točkovno oceno za M.

Intervalno oceno s tveganjem $\alpha = 0,05$ za aritmetično sredino pa dobimo iz neenačbe

$$\bar{x} - 1,96 \frac{s}{\sqrt{n}} < M < \bar{x} + 1,96 \frac{s}{\sqrt{n}} \qquad (2)$$

pri čemer je $1,96 = z_{\alpha/2}$

Primer: Iz vzorca n = 153 preskušancev trdnosti patentirane jeseniške žice smo dobili $\bar{x}$ = 158,78 kp/mm2 in s = 3,18. Po obrazcu (2) dobimo za intervalno oceno

$$158,78 - 1,96 \frac{3,18}{\sqrt{153}} < M < 158,78 + 1,96 \frac{3,18}{\sqrt{150}}$$

$$158,3 < M < 159,3$$

8. 15. Ocena razlik med aritmetičnima sredinama

Če izberemo iz dveh normalno porazdeljenih populacij $N(M_1; \sigma_1^2)$ $N(M_2; \sigma_2^2)$ vzorca z n_1 in n_2 enotami, se razlika ocen aritmetičnih sredin $\Delta \bar{x} = \bar{x}_2 - \bar{x}_1$ porazdeljujejo normalno

$$\Delta \bar{x} = \bar{x}_2 - \bar{x}_1 =: N\left(M_2 - M_1; \quad \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \qquad (1)$$

Če so vzorci veliki, velja tudi v tem primeru centralni limitni teorem in zgornje zakonitosti niso vezane na normalnost porazdelitve.

Primer. Iz dveh partij proizvodnje ocenjujemo razlike v trdnosti. Iz prve partije smo na vzorcu $n_1 = 2oo$ preskušancev dobili $\bar{x}_1 = 158$ in $s_1 = 3,18$ iz druge partije pa iz vzorca $n_2 = 4oo$ preskušancev $\bar{x}_2 = 143$ in $s_2 = 6,12$. Kolika je z $\alpha = o,o5$ najmanjša razlika v trdnosti med obema partijama. Ker vprašujemo po najmanjši razliki, je interval zaupanja enostranski in pri danem tveganju enak

$$M = \bar{x}_2 - \bar{x}_1 - z_{o,o5} se\Delta\bar{x}$$

Ker je

$$se\Delta\bar{x} = \sqrt{var\ \Delta\bar{x}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{3,18^2}{2oo} + \frac{6,21}{4oo}} =$$

$$= o,38o \qquad z_{o,o5} = 1,645$$

$$\Delta M = 158 - 143 - 1,645 \,.\, o,38o = 14,37$$

Razlika v trdnosti med obema partijama je s tveganjem $\alpha = 0,05$ najmanj 14,37 kp/mm^2

8. 16. Ocenjevanje strukturnega deleža z enostavnim slučajnostnim vzorcem

Za velike vzorce velja centralni limitni teorem tudi za strukturne deleže. Tako je

$$p\% = 1oo\,\frac{n_a}{n} \qquad (1)$$

nepristranska ocena za strukturni delež P% v populaciji. $p\%$ se porazdeljuje normalno s parametri

$$E(p\%) = P\% \qquad Var\ p\% = \frac{P\%(1oo-P\%)}{n} \qquad (2)$$

Za vzorce brez ponavljanja iz končnih populacij pa je

$$\mathrm{Var}(p\%)\quad \frac{S_a^2}{n}\,\frac{N-n}{N} \qquad (3)$$

Pri tem je $S_a^2 = P\%(1oo-P\%)\,\frac{N}{N-1}$ (4)

Če varianco ocen strukturnih deležev ocenjujemo iz vzorca, velja

$$\hat{s}_a^2 = p\%(1oo-p\%)\,\frac{n}{n-1} \qquad \mathrm{var}(p\%) = \frac{\hat{s}_a^2}{n}\,\frac{N-n}{N} \qquad (5)$$

Intervalna ocena za strukturni delež je v tem primeru s tveganjem $\alpha = o,o5$

$$p\% - 1,96\,se(p\%) < P\% < p\% + 1,96.se(p\%) \qquad (6)$$

~~Primer 1. Za določeno skupino N = 1oooo artiklov želimo oceniti odstotek artiklov kvalitete A z vzorcem n = 8oo artiklov, ki smo jih izbrali z vzorcem brez ponavljanja.~~

8. 17. Ocena razlik med strukturnima deležema

Če sta vzorca iz dveh neodvisnih populacij zadosti velika, se razlika ocen strukturnih deležev iz dveh populacij z $P_1\%$ in $P_2\%$ porazdeljuje normalno

$$\Delta p\% = p_1\% - p_2\% \doteq N(E\Delta p\% = P_1\% - P_2\%\ ;\ \mathrm{Var}\ p\% = \mathrm{Var}\ p_1\% + \mathrm{Var}\ p_2\%) \qquad (1)$$

Zato ocenjujemo razlike v strukturnih deležih za velike vzorce podobno kot razlike med aritmetičnimi sredinami.

V praktičnih primerih moremo nadomestiti prave vrednosti varianc z ocenami.

Primer. Dva stroja izdelujeta enake artikle. V dnevni proizvodnji smo na enem stroju izdelali $n_1 = 5ooo$ artiklov, od tega $n_{1a} = 25oo$ artiklov dane kvalitete, na drugem stroju pa $n_2 = 6ooo$

artiklov, od tega 24oo dane kvalitete. Če vzamemo izolirano samo dnevno proizvodnjo, je razlika v odstotkih kvalitete A

$P_1 = 5o\%$ $\quad P_2 = 4o\ \%$ $\quad \Delta P\% = 5o-4o = 1o\%.$

Če pa vzamemo dnevno proizvodnjo za reprezentanta oziroma/iz vzorec hipotetične populacije dela strojev pod istimi pogoji, pa je $\Delta p\%$ ocena razlike. Za intervalno oceno razlike izračunamo:

$$\mathrm{varp}\% = \mathrm{varp}_1\% + \mathrm{varp}_2\% = \frac{p_1\%\,(1oo-p_1\%)}{n_1} + \frac{p_2\%\,(1oo-p_2\%)}{n_2} =$$

$$= \frac{5o.5o}{5ooo} + \frac{4o.6o}{6o,oo} = o,5 + o,4 = o,9 \qquad se\,\Delta p\% = \sqrt{o,9o} = o,949$$

$$d\,\Delta\,p\% = 1,96.o949 = 1,86$$

Intervalna ocena razlike v kvaliteti dela strojev je

$$\Delta P\% = \Delta p\% \pm d\,\Delta\,p\% = 1o\% \pm 1,86\%$$

8. 18. Velikost vzorca za oceno pri dani natančnosti

Natančnost oziroma kvaliteto ocene merimo s standardno pogreško ocene SE ali pa z maksimalnim verjetnostnim odklonom D. Če je znana varianca σ^2 v osnovni populaciji, ocenimo potrebno velikost vzorca s ponavljanjem za dano absolutno natančnost ocene za aritmetično sredino D(x) po obrazcu

$$n = \left(\frac{z\ \sigma_x}{D(\bar{x})}\right)^2 \qquad (1)$$

Če pa je znan koeficient variacije in relativni maksimalni verjetni odklon $D(\bar{x})\%$ na aritmetično sredino, je n dan z obrazcem

$$n = \left(\frac{z\ KV\%}{D(\bar{x})\%}\right)^2 \qquad (2)$$

Za oceno strukturnega deleža dobimo pri dani absolutni natančnosti D(P%) za vzorec s ponavljanjem velikost vzorca

$$n = \frac{z^2 . P\% Q\%}{D(p\%)^2} \qquad (3)$$

oziroma pri relativni natančnosti D(p%)%

$$n = \frac{z^2 . Q\%}{P\% \quad D(p\%)\%^2} \qquad (4)$$

V vseh primerih pa dobimo iz potrebnega števila enot v vzorcih s ponavljanjem potrebno število enot v vzorcu brez ponavljanja n' po obrazcu

$$n' = \frac{n . N}{N + n - 1} = \frac{n}{1 + f} \qquad (5)$$

Primer 1.: Če npr. za trdnost patentirane jeseniške žice vemo, da je KV% = 2%, za oceno poprečja pa želimo, da se z tveganjem α = o,o5 ne bo odklanjala za več kot 1%, dobimo, da je treba v tem primeru vzeti približno

$$n = \left(\frac{1,96 . 2}{1} \right)^2 = 16$$

Primer 2. Za populacijo z N = loooo artiklov preskušancev ocenjujemo strukturni delež, za katerega predvidevamo, da je njegova vrednost okrog P = 7o%. V primeru, da želimo, da se ocena s tveganjem α = o,ol ne bo odklanjala od prave vrednosti za več kot D(p%) = 2%, moramo vzeti

$$n = \frac{2,58^2 . 75.25}{2^2} = 312o$$

Če gre za vzorec brez ponavljanja pa je

$$n^{1} = \frac{3120}{1 + 0,312} = 2378$$

ker je vzorčni delež $f = \frac{3120}{10000} = 0,3120$

Ker je v našem primeru vzorčni delež velik, je razlika v velikosti vzorcev z in brez ponavljanja znatna (742).

8. 19. **Standardne pogreške za standardni odklon in koeficient variacije**

Za velike vzorce je standardna pogreška za oceno standardnega odklona enaka

$$SE\ (s) = \frac{\sigma}{\sqrt{2n}} \qquad (1)$$

za oceno koeficienta variacije kv% pa

$$SE(kv\%) = \frac{KV\%}{\sqrt{2n}} \cdot \sqrt{1 + \frac{2.KV\%^2}{10^4}} = \frac{KV\%}{\sqrt{2n}} \qquad (2)$$

Standardne pogreške za zgornja parametra uporabljamo enako kot za oceno aritmetične sredine in strukturnega deleža za izračunavanje intervalnih ocen in v praktičnih primerih prave vrednosti σ in KV% zamenjamo z ocenami s, kv%.

Primer. Vzemimo, da je iz vzorca n = 153 preskušancev ocenjena poprečna trdnost patentirane jeseniške žice $\bar{x}$ = 160 kp/mm2 in standardni odklon s = 3,2 kp/mm2. Ocena za koeficient variacije je

$$kv\% = 100 \frac{s}{\bar{x}} = 100 \frac{3,2}{160} = 2\,\%$$

Po obrazcu za oceno približne vrednosti za standardne pogreške je

$$se(kv\%) = \frac{kv\%}{\sqrt{2n}} = \frac{2\%}{\sqrt{2.153}} = 0,114$$

Ocena za maksimalni verjetno~~st~~ni odklon z $\alpha = 0,05$ pa

$d(kv\%) = 1,96 . se(kv\%) = 1,96 . 0,114 = 0,224$

in intervalna ocena

$KV\% = 2\% \pm 0,224$

9. MALI VZORCI

9. 1. Teoretične porazdelitve

Pri statističnem opazovanju in preskušanju hipotez so osnova sklepanja porazdelitve vzorčnih količin. Teh pa v konkretnih primerih ne moremo sestaviti podobno kot sestavljamo frekvenčne porazdelitve iz osnovnih podatkov. Če vzorci zadoščajo določenim predpostavkam, moremo verjetnostne porazdelitve določenih vzorčnih izrazov dobiti teoretično iz postavk verjetnostnega računa.

Tako se število enot z dano značilnostjo, ali strukturni delež za to značilnost vzorcev s ponavljanjem ali vzorcev iz hipotetičnih populacij ~~enot~~ porazdeljuje v binomski porazdelitvi. Isti podatki pri vzorčenju brez ponavljanja pa se porazdeljujejo v hipergeometrični porazdelitvi. Če je pojav redek, vzorci pa zelo veliki, se število enot z dano značilnostjo bolj in bolj porazdeljujejo v Poisonvoi porazdelitvi.

Pomembna zvezna, vzorčna porazdelitev je normalna porazdelitev, ki je tudi osnova za druge teoretične porazdelitve. Najpomembnejše iz normalno porazdeljenih slučajnostnih spremenljivk izvedene teoretične oziroma verjetnostne porazdelitve so:

χ^2 porazdelitev, t - porazdelitev in F - porazdelitev.

9. 2. Stopinje prostosti

Če je n med seboj drugače neodvisnih spremenljivk vezanih na pogoj, da je aritmetična sredina enaka, od n spremenljivk svobodno variira le n-1, medtem ko je ena vezana na aritmetično sredino in je njena vrednost določena z n-1 vrednostmi. Podobno je v primeru, če imamo več skupin drugače neodvisnih podatkov, in je vsaka skupina omejena z pogojem, da je aritmetična sredina za vsako skupino konstanta. V tem primeru je število spremenljivk, ki svobodno variirajo, v celoti enako vsoti spremenljivk v vsaki skupini, zmanjšano za število aritmetičnih sredin, na katere so vezane skupine.

Število svobodno varirajočih spremenljivk imenujemo število stopinj prostosti. Število stopinj prostosti je torej v splošnem enako številu spremeljivk, zmanjšano za število omejitev oziroma pogojev, katerim mora zadostiti sistem spremenljivk.

9.3 χ^2 - Porazdelitev

Če imamo m med seboj neodvisnih slučajnostnih spremenljivk, ki se porazdeljujejo standardizirano normalno $z_i = : N(o,1)$, se vsota kvadratov teh spremenljivk

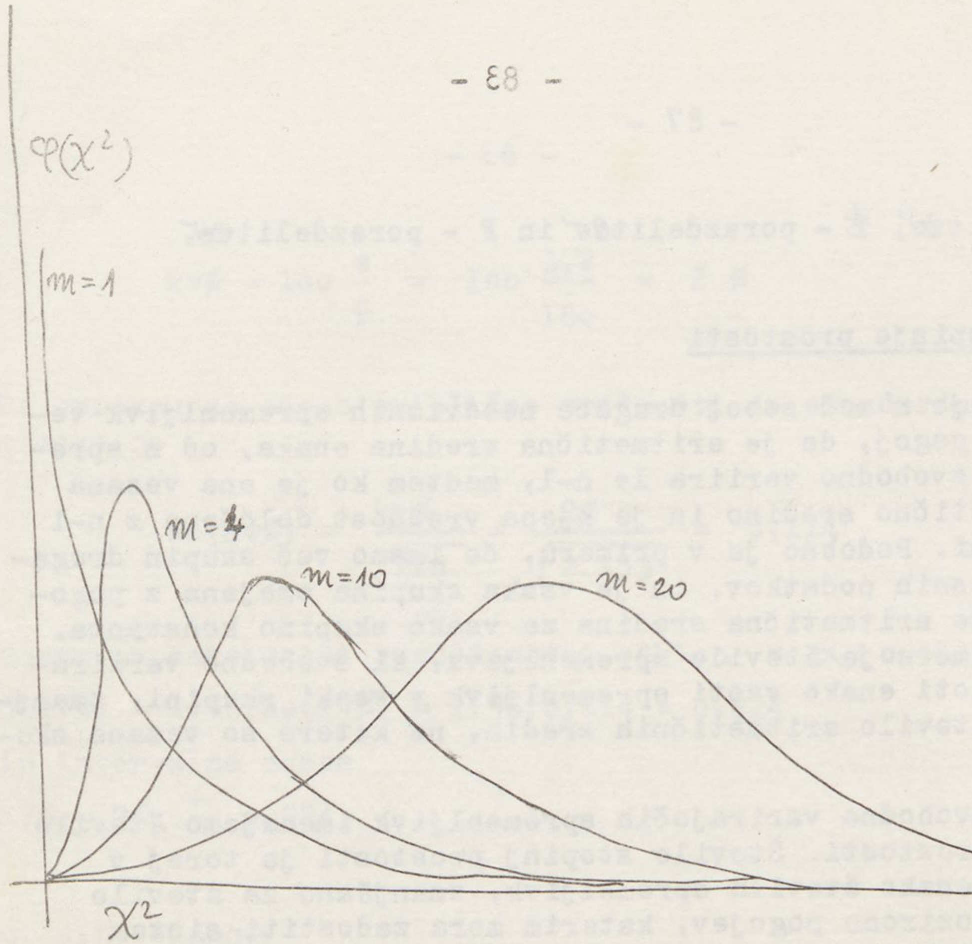
$$z_1^2 + z_2^2 + \ldots\ldots + z_m^2 = \sum_{i=1}^{m} z_i^2 = \chi^2 \qquad (1)$$

porazdeljujejo v porazdelitvi, za katero je gostota verjetnosti dana z

$$\varphi(\chi^2) = c_{\chi^2} (\chi^2)^{\frac{m-2}{2}} e^{-\frac{\chi^2}{2}} \qquad (2)$$

To porazdelitev imenujemo χ^2 porazdelitev.

χ^2 porazdelitev je v splošnem asimetrična porazdelitev. Stopnja asimetrije pa se manjša, čim večje je število stopinj prostosti (glej sliko!)



Slika . χ^2 porazdelitev

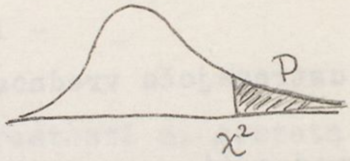
Izraz

$$\chi^2 = \frac{1}{2}\left(z + \sqrt{2m - 1}\right)^2 \qquad (3)$$

se asimptotično porazdeljuje v χ^2 porazdelitvi, če se število stopinj prostosti veča. Kot uporabno aproksimacijo moremo smatrati zgornji izraz že, če je m = 30.

V obrazcu (3) je z standardizirano normalno porazdeljena slučajnostna spremenljivka.

Tabela : χ^2 - porazdelitev



m \ P	0,99	0,95	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0002	0,004	0,46	1,07	1,64	2,71	3,84	5,41	6,64	10,83
2	0,020	0,103	1,39	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	0,115	0,35	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	0,30	0,71	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	0,55	1,14	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,52
6	0,87	1,64	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,46
7	1,24	2,17	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,32
8	1,65	2,73	7,34	9,52	11,03	13,36	15,51	18,17	20,09	26,12
9	2,09	3,32	8,34	10,66	12,24	14,68	16,92	19,68	21,67	27,88
10	2,56	3,94	9,34	11,78	13,44	15,99	18,31	21,16	23,21	29,59
11	3,05	4,58	10,34	12,90	14,63	17,28	19,28	22,62	24,72	31,26
12	3,57	5,23	11,34	14,01	15,81	18,55	21,03	24,05	26,22	32,91
13	4,11	5,89	12,34	15,12	16,98	19,81	22,36	25,47	27,69	34,53
14	4,66	6,57	13,34	16,22	18,15	21,06	23,68	26,87	29,14	36,12
15	5,23	7,26	14,34	17,32	19,31	22,31	25,00	28,28	30,58	37,70
16	5,81	7,96	15,34	18,42	20,46	23,54	26,30	29,63	32,00	39,25
17	6,41	8,67	16,34	19,51	21,62	24,77	27,59	31,00	33,41	40,79
18	7,02	9,39	17,34	20,60	22,76	25,99	28,87	32,35	34,80	42,31
19	7,63	10,12	18,34	21,69	23,90	27,20	30,14	33,69	36,19	43,82
20	8,26	10,85	19,34	22,78	25,04	28,41	31,41	35,02	37,57	45,32
21	8,90	11,59	20,34	23,86	26,17	29,62	32,67	36,34	38,93	46,80
22	9,54	12,34	21,34	24,94	27,30	30,81	33,92	37,66	40,29	48,27
23	10,20	13,09	22,34	26,02	28,43	32,01	35,17	38,97	41,64	49,73
24	10,86	13,85	23,34	27,10	29,55	33,20	36,42	40,27	42,98	51,18
25	11,52	14,61	24,34	28,17	30,68	34,38	27,65	41,57	44,31	52,62
26	12,20	15,38	25,34	29,25	31,80	35,56	38,88	42,86	45,64	54,05
27	12,88	16,15	26,34	30,32	32,91	36,74	40,11	44,14	46,96	55,48
28	13,56	16,93	27,34	31,39	34,03	37,92	41,34	45,42	48,28	56,89
29	14,26	17,71	28,34	32,46	35,14	39,09	42,56	46,69	49,59	58,30
30	14,95	18,49	29,34	33,53	36,25	40,28	43,77	47,96	50,89	59,70

Za χ^2 distribucije, ki imajo m $>$ 30, velja: $\chi_p^2 = \frac{1}{2}\left(\sqrt{2m-1} + z_p\right)^2$,

pri čemer je z standardiziran odklon normalne distribucije.

Verjetnostim P ustrezajoče vrednosti z so dane v tabeli II.

Tabela II. z-vrednosti

P	z_p
o,99	- 2,3263
o,95	- 1,6449
o,5o	o,oooo
o,3o	+ o,5244
o,2o	+ o,8416
o,1o	+ 1,2816
o,o5	+ 1,6449
o,o2	+ 2,o537
o,o1	+ 2,3263
o,oo1	+ 3,o9o2

Primer

$$\chi^2(m=85) = \frac{1}{2}\left(\sqrt{2.85-1}+1,6449\right)^2 = 1o7,24$$

V tabeli 1 so za nekatere P in stopnje m od 1 do 3o podane kritične vrednosti za χ^2 porazdelitev. Pod posameznimi tabeliranimi vrednostmi pa so v zadnji vrsti dane ustrezne vrednosti z, ki omogočajo izračun kritičnih vrednosti za m $>$ 3o po obrazcu 3.

9. 4. Studentova t-porazdelitev

Za med seboj neodvisni slučajnostni spremenljivki $z =: N(o,1)$ in $\chi^2 =: X^2(m)$ od katerih se z porazdeljuje standardizirano normalno, χ^2 pa v χ^2 porazdelitvi z m stopinjami prostosti, se izraz

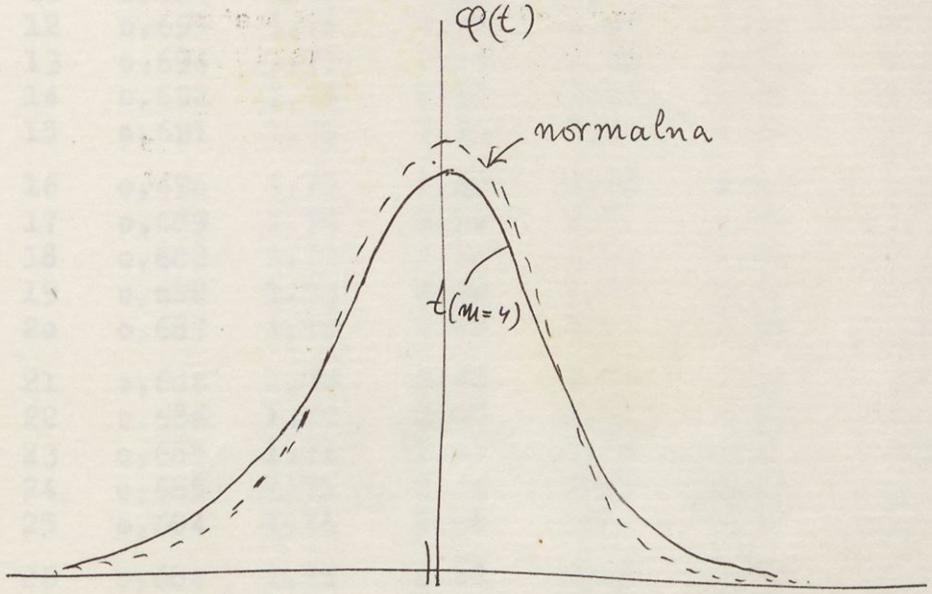
$$\frac{z}{\sqrt{\chi^2/m}} = t(m) \qquad (1)$$

porazdeljuje v porazdelitvi, ki jo imenujemo t-porazdelitev.

Zanjo je število stopinj prostosti m, gostota relativne frekvence pa je dana s funkcijo

$$\varphi(t) = \frac{C_t}{\left(1+\frac{t^2}{m}\right)^{\frac{m+1}{2}}} \qquad (2)$$

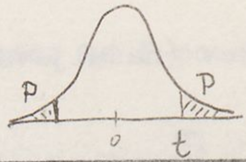
Kot kaže funkcija za gostoto verjetnosti $\varphi(t)$ in slika, je t-porazdelitev simetrična unimodalna zvonasta porazdelitev, ki se z večanjem števila stopinj prostosti asimptotično približuje standardizirano normalni porazdelitvi.



Slika 1, Studentova t-porazdelitev z vrisano asimptotično t-porazdelitvijo za m = ∞, ki sovpada z normalno porazdelitvijo.

V tabeli so dane kritične vrednosti za t-porazdelitev. Kritične vrednosti za t se z večanjem števila stopinj prostosti bolj in bolj približujejo ustreznim vrednostim za standardizirano normalno porazdelitev, čim večje je število stopinj prostosti.

Tabela 1; t-porazdelitev



P	o,25	o,o5	o,o25	o,ol	o,oo5	o,ooo5
m 2P	o,5o	o,lo	o,o5	o,o2	o,ol	o,ool
1	1,ooo	6,31	12,71	31,82	63,66	637
2	o,816	2,92	4,3o	6,96	9,92	31,6
3	o,765	2,35	3,18	4,54	5,84	12,9
4	o,741	2,13	2,78	3,75	4,6o	8,61
5	o,727	2,o2	2,57	3,36	4,o3	6,86
6	o,718	1,94	2,45	3,14	3,71	5,96
7	o,711	1,9o	2,36	3,oo	3,5o	5,4o
8	o,7o6	1,86	2,31	2,9o	3,36	5,o4
9	o,7o3	1,83	2,26	2,82	3,25	4,78
1o	o,7oo	1,81	2,23	2,76	3,17	4,59
11	o,697	1,8o	2,2o	2,72	3,11	4,44
12	o,695	1,78	2,18	2,68	3,o6	4,32
13	o,694	1,77	2,16	2,65	3,o1	4,22
14	o,692	1,76	2,14	2,62	2,98	4,14
15	o,691	1,75	2,13	2,6o	2,95	4,o7
16	o,69o	1,75	2,12	2,58	2,92	4,o2
17	o,689	1,74	2,11	2,57	2,9o	3,96
18	o,688	1,73	2,1o	2,55	2,88	3,92
19	o,688	1,73	2,o9	2,54	2,86	3,88
2o	o,687	1,72	2,o9	2,53	2,84	3,85
21	o,686	1,72	2,o8	2,52	2,83	3,82
22	o,686	1,72	2,o7	2,51	2,82	3,79
23	o,685	1,71	2,o7	2,5o	2,81	3,77
24	o,685	1,71	2,o6	2,49	2,8o	3,74
25	o,684	1,71	2,o6	2,48	2,79	3,72
26	o,684	1,71	2,o6	2,48	2,78	3,71
27	o,684	1,7o	2,o5	2,47	2,77	3,69
28	o,683	1,7o	2,o5	2,47	2,76	3,67
29	o,683	1,7o	2,o4	2,46	2,76	3,66
3o	o,683	1,7o	2,o4	2,46	2,75	3,65

Tabela 1: t-porazdelitev (nadaljevanje)

P	0,25	0,05	0,025	0,01	0,005	0,0005
2P	0,50	0,10	0,05	0,02	0,01	0,001
35	0,682	1,69	2,03	2,44	2,72	3,59
40	0,681	1,68	2,02	2,42	2,71	2,55
45	0,680	1,68	2,02	2,41	2,69	3,52
50	0,679	1,68	2,01	2,41	2,69	3,50
60	0,678	1,67	2,00	2,39	2,66	3,46
70	0,678	1,67	2,00	2,38	2,65	3,44
80	0,677	1,66	1,99	2,38	2,64	3,42
90	0,677	1,66	1,99	2,37	2,63	3,40
100	0,677	1,66	1,98	2,36	2,63	3,39
120	0,676	1,66	1,98	2,36	2,62	3,37
150	0,676	1,66	1,98	2,35	2,61	3,36
200	0,675	1,65	1,97	2,35	2,60	3,34
300	0,675	1,65	197	3,34	2,59	3,32
400	0,675	1,65	1,97	2,34	2,59	3,32
500	0,674	1,65	1,96	2,33	2,59	3,31
1000	0,674	1,65	1,96	2,33	2,58	3,30
∞	0,674	1,64	1,96	2,33	2,58	3,29

9. 5. F-porazdelitev

Če se slučajnostna spremenljivka χ_1^2 porazdeljuje v χ^2 porazdelitvi z m_1 stopinjami prostosti in χ_2^2 v χ^2 porazdelitvi z m_2 stopinjami prostosti in sta med seboj neodvisni, se

$$F = \frac{\chi_1^2/m_1}{\chi_2^2/m_2} \qquad (1)$$

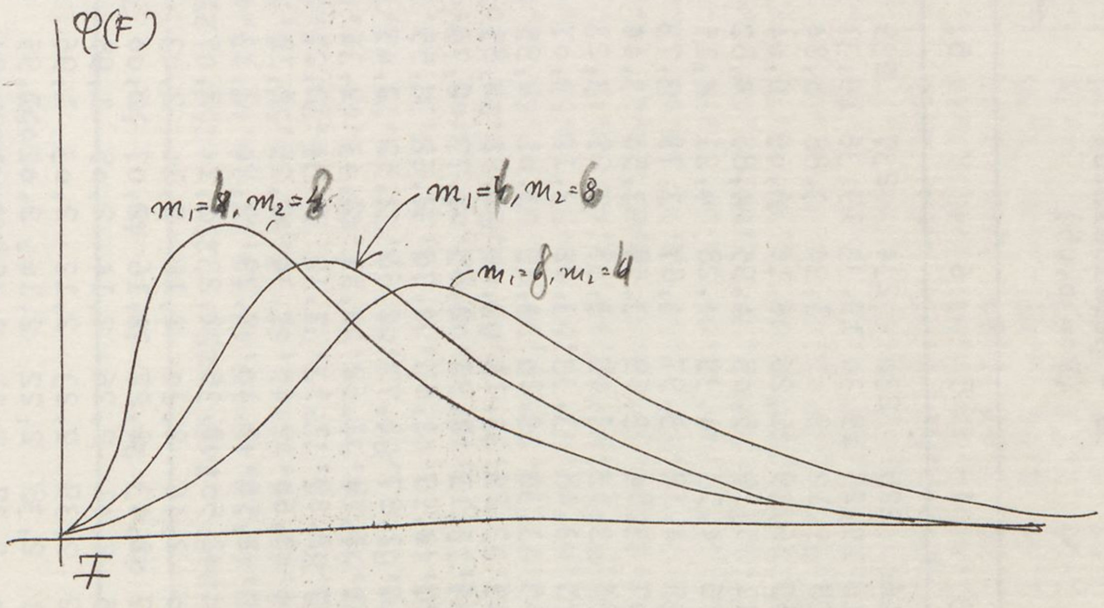
porazdeljuje v porazdelitvi, ki jo imenujemo F-porazdelitev.

Gost-ota relativne frekvence za F-porazdelitev je

$$\varphi(F) = C_F . F^{\frac{m_1 - 2}{2}} \left(1 + \frac{m_1}{m_2} F\right)^{-\frac{m_1 + m_2}{2}} \qquad (2)$$

Iz obrazca vidimo, da F-porazdelitev zavisi od dveh stopinj prostosti m_1 in m_2.

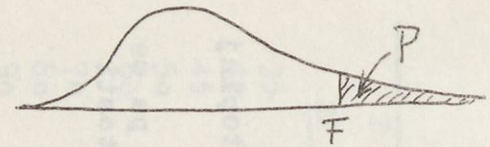
F-porazdelitev je asimetrična porazdelitev, za katero pa se stopnja asimetrije manjša, če se število stopinj prostosti veča.



Slika F-porazdelitev

Ker zavisi F-porazdelitev od dveh stopinj prostosti m_1 in m_2, je tabela kritičnih vrednosti kompleksnejša in dobimo za vsako nivo svojo tabelo.

F - porazdelitev
(P = o,o5)



m2 \ m1	1	2	3	4	5	6	7	8	9	1o	11	12
1	161	2oo	216	225	23o	234	237	239	241	242	243	244
2	18,51	19,oo	19,16	19,25	19,3o	19,33	19,36	19,37	19,38	19,39	19,4o	19,41
3	1o,13	9,55	9,28	9,12	9,o1	8,94	3,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,o9	6,o4	6,oo	5,96	5,93	5,91
5	6,61	5,97	5,41	5,19	5,o5	4,95	4,88	4,82	4,78	4,74	4,7o	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,1o	4,o6	4,o3	4,oo
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,6o	3,57
8	5,32	4,46	4,o7	3,84	3,69	3,58	3,5o	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,1o	3,o7
1o	4,96	4,1o	3,71	3,48	3,33	3,22	3,14	3,o7	3,o2	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,2o	3,o9	3,o1	2,95	2,9o	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,oo	2,92	2,85	2,8o	2,76	2,72	2,69
14	4,6o	3,74	3,34	3,11	2,96	2,85	2,77	2,7o	2,65	2,6o	2,56	2,53
17	4,45	3,59	3,2o	2,96	2,81	2,7o	2,62	2,55	2,5o	2,45	2,41	2,38
2o	4,35	3,49	3,1o	2,87	2,71	2,6o	2,52	2,45	2,4o	2,35	2,31	2,28
24	4,26	3,4o	3,o1	2,78	2,62	2,51	2,43	2,36	2,3o	2,26	2,22	2,18
3o	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,o9
4o	4,o8	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,o7	2,o4	2,oo
5o	4,o3	3,18	2,79	2,56	2,4o	2,29	2,2o	2,13	2,o7	2,o2	1,98	1,95
7o	3,98	3,13	2,74	2,5o	2,35	2,23	2,14	2,o7	2,o1	1,97	1,93	1,89
1oo	3,94	3,o9	2,7o	2,46	2,3o	2,19	2,1o	2,o3	2,97	1,92	1,88	1,85
15o	3,91	3,o6	2,67	2,43	2,27	2,16	2,o7	2,oo	1,94	1,89	1,85	1,82
2oo	3,89	3,o4	2,65	2,41	2,26	2,14	2,o5	1,98	1,92	1,87	1,83	1,8o
4oo	3,86	3,o2	2,62	2,39	2,23	2,12	2,o3	1,96	1,9o	1,85	1,81	1,78
1.ooo	3,85	3,oo	2,61	2,38	2,22	2,1o	2,o2	1,95	1,89	1,84	1,8o	1,76
∞	3,84	2,99	2,6o	2,37	2,21	2,o9	2,o1	1,94	1,88	1,83	1,79	1,75

F - porazdelitev
(P = o,o5)

m2 \ m1	14	16	2o	24	3o	4o	5o	75	1oo	2oo	5oo	∞
1	245	246	248	249	25o	251	252	253	253	254	254	254
2	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,5o	19,5o
3	8,71	8,69	8,66	8,64	8,62	8,6o	8,58	8,57	8,58	8,54	8,54	8,53
4	5,87	5,84	5,8o	5,77	5,74	5,71	5,7o	5,68	5,66	5,65	5,64	5,63
5	4,64	4,6o	4,56	4,53	4,5o	4,46	4,44	4,42	4,4o	4,38	4,37	4,36
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67
7	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	2,23
8	3,23	3,2o	3,15	3,12	3,o8	3,o5	3,o3	3,oo	2,98	2,96	2,94	2,93
9	3,o2	2,98	2,93	2,9o	2,86	2,82	2,8o	2,77	2,76	2,73	2,72	2,71
1o	2,86	2,82	2,77	2,74	2,7o	2,67	2,64	2,61	2,5o	2,56	2,55	2,54
11	2,74	2,7o	2,65	2,61	2,57	2,53	2,5o	2,47	2,45	2,42	2,41	2,4o
12	2,64	2,6o	2,54	2,5o	2,46	2,42	2,4o	2,36	2,35	2,32	2,31	2,3o
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
17	2,33	2,29	2,23	2,19	2,15	2,11	2,o8	2,o4	2,o2	1,99	1,97	1,96
2o	2,23	2,18	2,12	2,o8	2,o4	1,99	1,96	1,92	1,9o	1,87	1,85	1,84
24	2,13	2,o9	2,o2	1,98	1,94	1,89	1,86	1,82	1,8o	1,76	1,74	1,73
3o	2,o4	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62
4o	1,95	1,9o	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
5o	1,9o	1,85	1,78	1,74	1,69	1,63	1,6o	1,55	1,52	1,48	1,46	1,44
7o	1,84	1,79	1,72	1,67	1,62	1,56	1,53	1,47	1,45	1,4o	1,37	1,35
1oo	1,79	1,75	1,68	1,63	1,57	1,51	1,48	1,42	1,39	1,34	1,3o	1,28
15o	1,76	1,71	1,64	1,59	1,54	1,47	1,44	1,37	1,34	1,29	1,25	1,22
2oo	1,74	1,69	1,62	1,57	1,52	1,45	1,42	1,35	1,32	1,26	1,22	1,19
4oo	1,72	1,67	1,6o	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,16	1,13
1,ooo	1,7o	1,65	1,58	1,53	1,47	1,41	1,36	1,3o	1,26	1,19	1,13	1,o8
∞	1,69	1,64	1,57	1,52	1,46	1,4o	1,35	1,28	1,24	1,17	1,11	1,oo

F - porazdelitev
(P = o,ol)

m 2 \ ml	1	2	3	4	5	6	7	8	9	1o	11	12
1	4o52	4999	54o3	5625	5764	5859	5928	5981	6o22	6o56	6o82	61o6
2	98,49	99,o1	99,17	99,25	99,3o	99,33	99,34	99,36	99,38	99,4o	99,41	99,42
3	34,12	3o,81	29,46	28,24	27,91	27,91	27,67	27,49	27,34	27,23	27,13	27,o5
4	21,2o	18,oo	16,69	15,98	15,52	15,21	14,98	14,8o	14,66	14,54	14,45	14,37
5	16,26	13,27	12,o6	11,39	1o,97	1o,67	1o,45	1o,27	1o,15	1o,o5	9,96	9,89
6	13,74	1o,92	9,78	9,15	8,75	8,47	8,26	8,1o	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,48	7,19	7,oo	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,o1	6,63	6,37	6,19	6,o3	5,91	5,82	5,74	5,67
9	1o,56	8,o2	6,99	8,42	5,o6	5,8o	5,62	5,47	5,35	5,26	5,18	5,11
1o	1o,o4	7,56	6,55	5,99	5,64	5,39	5,21	5,o8	4,95	4,85	4,78	4,71
11	9,65	7,2o	6,22	5,67	5,32	5,o7	4,88	4,74	4,63	4,54	4,46	4,4o
12	9,33	6,93	5,95	5,41	5,o6	4,82	4,65	4,5o	4,39	4,3o	4,22	4,16
14	8,86	6,51	5,56	5,o3	4,69	4,46	4,28	4,14	4,o3	3,94	3,86	3,6o
17	8,4o	6,11	5,18	4,67	4,34	4,1o	3,93	3,79	368	3,59	3,52	3,45
2o	8,1o	5,85	4,94	4,43	4,1o	3,87	3,71	3,56	3,45	3,37	3,3o	3,23
24	7,82	5,61	4,72	4,22	3,9o	3,67	3,5o	3,36	3,25	3,17	3,o9	3,o3
3o	7,56	5,39	4,51	4,o2	3,7o	3,47	3,3o	3,17	3,o6	2,98	2,9o	2,84
4o	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,8o	2,73	2,66
5o	7,17	5,o6	4,2o	3,72	3,41	3,18	3,o2	2,88	2,78	2,7o	2,62	2,56
7o	7,o1	4,92	4,o8	3,6o	3,29	3,o7	2,91	2,77	2,67	2,59	2,51	2,45
1oo	6,9o	4,82	3,98	3,51	3,2o	2,99	2,82	2,69	2,59	2,51	2,43	2,36
15o	6,81	4,75	3,91	3,44	3,14	2,92	2,76	2,62	2,53	2,44	2,37	2,3o
2oo	6,76	4,71	3,88	3,41	3,11	2,9o	2,73	2,6o	2,5o	2,41	2,34	2,28
4oo	6,7o	4,66	3,83	3,36	3,o6	2,85	2,69	2,55	2,46	2,37	2,29	2,23
1.ooo	6,66	4,62	3,8o	3,34	3,o4	2,82	2,66	2,53	2,43	2,34	2,26	2,2o
∞	6,64	4,6o	3,78	3,32	3,o2	2,8o	2,64	2,51	2,41	2,32	2,24	2,18

F -porazdelitev
(P = o,ol)

m_2 \ m_1	14	16	2o	24	3o	4o	5o	75	1oo	2oo	5oo	∞
1	6142	6169	62o8	6234	6258	6286	63o2	6323	6334	6352	6361	6366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,5o	99,5o
3	26,92	26,83	26,69	26,6o	26,5o	26,41	26,35	26,27	26,23	26,18	26,14	26,12
4	14,24	14,15	14,o2	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46
5	9,77	9,68	9,55	9,47	9,39	9,29	9,24	9,17	9,13	9,o7	9,o4	9,o2
6	7,6o	7,52	7,39	7,31	7,23	7,14	7,o9	7,o2	6,99	6,94	6,9o	6,88
7	6,35	6,27	6,15	6,o7	5,98	5,9o	5,85	5,75	5,7o	5,7o	5,67	5,65
8	5,56	5,48	5,36	5,28	5,2o	5,11	5,o6	5,oo	4,96	4,91	4,88	4,86
9	5,oo	4,92	4,8o	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31
1o	4,6o	4,52	4,41	4,38	4,25	4,17	4,12	4,o5	4,1o	3,96	3,93	3,91
11	4,29	4,21	4,1o	4,o2	3,94	3,86	3,8o	3,74	3,7o	3,66	3,62	3,6o
12	4,o5	3,98	3,86	3,78	3,7o	3,61	3,56	3,49	3,46	3,41	3,38	3,36
14	3,7o	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,o6	3,o2	3,oo
17	3,35	3,27	3,16	3,o8	3,oo	2,92	2,86	2,79	2,76	2,7o	2,67	2,65
2o	3,13	3,o5	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21
3o	2,74	2,66	2,55	2,47	2,38	2,29	2,24	2,16	213	2,o7	2,o3	2,o1
4o	2,56	2,49	2,37	2,29	2,2o	2,11	2,o5	1,97	1,94	1,88	1,84	1,81
5o	2,46	2,39	2,26	2,18	2,1o	2,oo	1,94	1,86	1,82	1,76	1,71	1,68
7o	2,35	2,28	2,15	2,o7	1,98	1,88	1,82	1,74	1,69	1,62	1,56	1,53
1oo	2,26	2,19	2,o6	1,98	1,89	1,79	1,73	1,64	1,59	1,51	1,46	1,43
15o	2,2o	2,12	2,oo	1,91	1,83	1,72	1,66	1,56	1,51	1,43	1,37	1,33
2oo	2,17	2,o9	1,97	1,88	1,79	1,69	1,62	1,53	1,48	1,39	1,33	1,28
4oo	2,12	2,o4	1,92	1,84	1,74	1,64	1,57	1,47	1,42	1,32	1,24	1,19
1.ooo	2,o9	2,o1	1,89	1,81	1,71	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	2,o7	1,99	1,87	1,79	1,69	1,59	1,52	1,41	1,36	1,25	1,15	1,oo

9. 6. Zakonitosti aritmetičnih sredin za male vzorce

Če iz normalno porazdeljene populacije $x =: N(M, \sigma^2)$ na slučajnosten način izberemo vzorec z obsegom n, se izraz

$$\frac{\bar{x} - M}{s_x} \sqrt{n} =: t(m=n-1) \qquad s_x^2 = \frac{S(x-\bar{x})^2}{n-1} = \frac{Sx^2 - X^2/n}{n-1} \qquad (1)$$

porazdeljuje v t-porazdelitvi z $m = n-1$ stopinjami prostosti. Če iz dveh normalno porazdeljenih populacij $X_1 =: N(M_1, \sigma^2)$ in $x_2 =: N(M_2; \sigma^2)$, ki imata različni aritmetični sredini, a enaki varianci izberemo na slučajnosten način vzorca z n_1 enotami iz prve in z n_2 iz druge populacije, se izraz

$$\frac{(\bar{x}_2 - \bar{x}_1) - (M_2 - M_1)}{s_d} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} =: t(m = n_1 + n_2 - 2) \qquad (2)$$

porazdeljen v t-porazdelitvi z $m = n_1 + n_2 - 2$ stopinjami prostosti. Pri tem je

$$s_d^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{Sx_1^2 + Sx_2^2 - X_1^2/n_1 - X_2^2/n_2}{n_1 + n_2 - 2} \qquad (3)$$

9. 7. Ocenitev aritmetične sredine z malim vzorcem

Intervalno oceno za aritmetično sredino iz vzorca iz normalno porazdeljene populacije, dobimo iz obrazca z naslednjo neenačbo

$$\bar{x} - t_{\alpha/2} \frac{s_x}{\sqrt{n}} < M < \bar{x} + t_{\alpha/2} \frac{s_x}{\sqrt{n}} \qquad (1)$$

Ta neenačba je slična neenačbi za ocenjevanje aritmetične sredine z velikimi vzorci, le da je standardiziran odklon z zamenjan s kritično vrednostjo t za t-porazdelitev. Za isto tveganje α je po pravilu koeficient t večji kot ustrezni z, ker je vračunana nezanesljivost ocen za standardni odklon.

Za primer vzemimo časovno proizvodnost:

za n = 20 šarž smo dobili naslednje neto storilnosti:

9630	9985	11616	10656	9324	11382	9979	11592	9963	10740
10750	10021	9715	10635	8027	12834	8349	9875	10187	8527

Iz teh podatkov dobimo: $\bar{x}$ = 10189; s = 1166; t(m=19) = 2,09 za α = 2P = 0,05

Meje zaupanja so po obrazcu (1)

$$10189 - 2{,}09 \frac{1166}{\sqrt{20}} < M < 10189 + 2.09 \frac{1166}{\sqrt{20}}$$

$$9644 < M < 10734$$

Zaradi velike variabilnosti je dobljena ocena zelo slaba.

9. 8. Ocene razlik med asimetričnima sredinama z malimi vzorci

Če upoštevamo zakonitosti iz 9. 6, dobimo meje zaupanja za razlike aritmetičnih sredin dveh vzorcev

$$(\bar{x}_2 - \bar{x}_1) - t_{\alpha/2}^{(m)} s_d \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} < M_2 - M_1 < (\bar{x}_2 - \bar{x}_1) + t_{\alpha/2}^{(m)} s_d \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

število stopinj prostosti za določanje vrednosti t je $m = n_1 + n_2 - 2$

Vzamemo za primer razlike v kontrakciji ψ% za dva načina ravnanja žice (1 = ročno ZRKK 2 = strojno Jesenice).

Razlike v kontrakciji iščemo z vzorci po $n_1 = 6$ in $n_2 = 6$ preskušancev.

Osnovni podatki:

ročno zavod: 4o,7 42,o 42,8 39,8 4o,3 4o,7

strojno Jesenice: 42,9 43,o 43,8 44,5 44,3 45,o

Iz teh podatkov ocenimo, da je

$\bar{x}_1 = 41{,}o5 \qquad s_1^2 = 1{,}268$

$\bar{x}_2 = 43{,}92 \qquad s_2^2 = o{,}71o$

Iz teh podatkov dobimo dalje, da je

$$s_d^2 = \frac{(6-1)1{,}268 + (6-1)o{,}71o}{6 + 6 - 2} = o{,}989 \qquad s_d = \sqrt{o{,}989} = o{,}994$$

Če želimo oceniti interval zaupanja, v katerem prava vrednost razlik leži s tveganjem $\alpha = o{,}o5$, dobimo iz tablic za kritične vrednosti za t-porazdelitev, da je

$$t_{o,o5}(m = 6 + 6 - 2 = 10) = 2{,}23$$

Razmak zaupanja pa je

$$43{,}92 - 41{,}o5 - 2{,}23 \cdot o{,}994 \sqrt{\frac{6+6}{6 \cdot 6}} < M_2 - M_1 < 43{,}92 - 41{,}o5 + {}$$
$$+ 2{,}23 \cdot o{,}994 \sqrt{\frac{6+6}{6 \cdot 6}}, \quad 1{,}59 < M_2 - M_1 < 4.15\%$$

9. 9. Zakonitosti ocen varianc za male vzorce

Če se osnovna populacija porazdeljuje normalno z varianco σ^2, se izraz (1) izračunan iz slučajnostnega vzorca z n enotami porazdeljuje

$$\frac{(n-1)s^2}{\sigma^2} =: \chi^2(m=n-1) \qquad (1)$$

v χ^2 porazdelitvi z m = n = 1 stopinjami prostosti.

Če iz prve populacije, v kateri se x_1 porazdeljuje v normalni porazdelitvi z varianco σ_1^2, izberemo vzorec z n_1 enotami in iz druge populacije, v kateri se x_2 porazdeljuje v normalni porazdelitvi z varianco σ_2^2 izberemo vzorec z n_2 enotami, se izraz

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} =: F(m_1 = n_1 - 1 \,;\, m_2 = n_2 - 1) \qquad (2)$$

porazdeljuje v F porazdelitvi z $m_1 = n_1 - 1$ in $m_2 = n_2 - 1$ stopinjami prostosti. Drugi obrazec za kvocient dveh neodvisnih varianc sledi iz obrazca za definicijo F-porazdelitve neposredno, če uporabimo obrazec (1).

9. 1o. <u>Ocenjevanje variance iz standardnega odklona za male vzorce</u>

Če je osnovna populacija normalno porazdeljena, moremo zaradi zakonitosti, ki je navedena v 9. 9 skonstruirati meje zaupanja za oceno variance po navedenem obrazcu

$$\frac{(n-1)s^2}{\chi^2_p} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-p}} \qquad (1)$$

Tveganje za dobljeno intervalno oceno je $\alpha = 2P$.

Če vzamemo, da smo z vzorcem n = 25 preskušancev dobili točkovno oceno za varianco kontrakcije $s^2 = 1,1$, je po zgornji neenačbi intervalna ocena z $\alpha = 2P = 0,10$ enaka

$$\frac{(25-1).1,1}{36,42} < \sigma^2 < \frac{(25-1).1,1}{13,85}$$

$$0,72 < \sigma^2 < 1,91$$

Če za dobljeno neenačbo poiščemo kvadratni koren, dobimo oceno za intervalno oceno za standardni odklon,

$$o,85 < \sigma < 1,38$$

lo. PREISKUŠANJE HIPOTEZ

lo. 1. Statistično preskušanje hipotez

Preskušanje hipotez v statističnem smislu je preverjanje hipotez o populaciji oziroma parametrih populacije. Hipoteze preskušamo v principu s podatki iz slučajnostnega vzorca iz populacije, ki jo proučujemo. S tem je podana osnova za objektivno preskušanje hipotez. Vsi zaključki statističnega preskušanja hipotez so verjetnostni.

lo. 2. Ničelna hipoteza

Z vzorci moremo v principu hipoteze le zavračati. Zato vsaki hipotezi priredimo ustrezno negativno protihipotezo. V nadaljnjem preskušamo protihipotezo, ki jo imenujemo ničelno hipotezo. Če uspemo, da zavrnemo ničelno hipotezo, smo s tem potrdili osnovno hipotezo.

Tako npr. hipotezi, da sta dva postopka proizvodnje artiklov z različno trdnostjo, ustreza ničelna hipoteza, da razlika v postopkih ne vpliva na trdnost.
Hipotezi, da je odstotek izmeta večji od 3%, ustreza ničelna hipoteza, da je odstotek izmeta enak ali manjši kot 3%.
Hipotezi, da sta dva pojava odvisna, ustreza ničelna hipoteza, da pojava nista odvisna.

lo. 3. Kritična vrednost

Vrednost, katero slučajnostna spremenljivka u prekorači z verjetnostjo P, imenujemo kritično vrednost u_P slučajnostne spremenljivke na nivoju P. Imamo kritične vrednosti za posamezne slučajnostne spremenljivke tabelirane za verjetnosti P = o,1o P = o,o5 P = o,o1 in P = o,oo1. Kritične vrednosti za standardizirano normalno razporejeno slučajnostno spremenljivko z, za Studentovo t-porazdelitev, za χ^2 porazdelitev in F porazdelitev služijo za osnovo za preskušanje hipotez in za določanje intervalov zaupanja.

1o. 4. Kritično območje

Kritično območje je razmak vrednosti, ki so večje od kritične vrednosti.

1o. 5. Napaka prve vrste

Zaključki, ki jih napravimo s preskušanjem hipotez niso gotovi, temveč veljajo le z določenim tveganjem. Tveganje, oziroma verjetnost, da ničelna hipoteza velja, mi jo pa zavrnemo, imenujemo napako prve vrste. Napako prve vrste reguliramo oziroma določamo vnaprej. Običajne stopnje napake prve vrste, na katerih delamo sklepe so: α = o,o5, α = o,o1, α = o,oo1. Ker napaka prve vrste pri statistični kontroli sovpada s tveganjem, da zavrnemo kvaliteto, ki ustreza ničelni hipotezi, imenujemo napako prve vrste pri preskušanju ničelnih hipotez pri statistični kontroli kvalitete tveganja proizvajalca (producers risk).

1o. 6. Napaka druge vrste

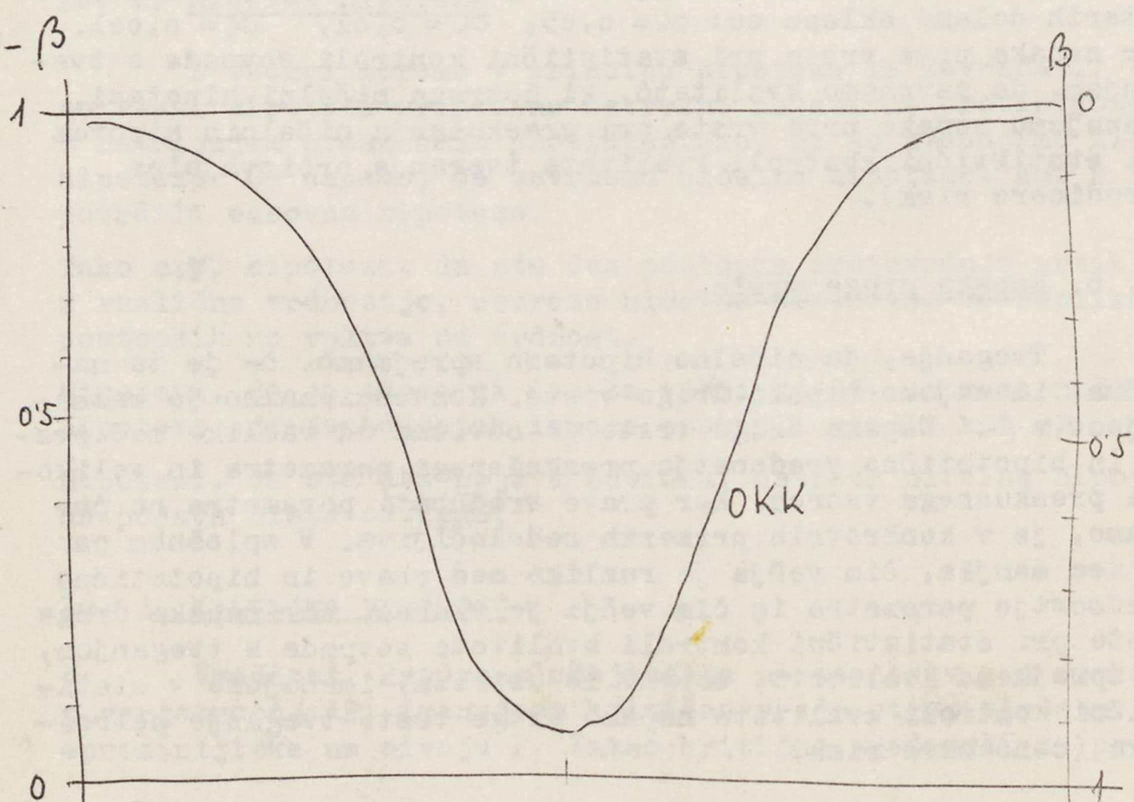
Tveganje, da ničelno hipotezo sprejmemo, če je ta napačna, imenujemo napako druge vrste. Konvencionalno jo zaznamujemo z β. Napaka druge vrste je odvisna od razlike med pravo in hipotetično vrednostjo preskušenega parametra in velikosti preskusnega vzorca. Ker prave vrednosti parametra ne poznamo, je v konkretnih primerih nedoločljiva. V splošnem pa je tem manjša, čim večja je razlika med pravo in hipotetično vrednostjo parametra in čim večji je vzorec. Ker napaka druge vrste pri statistični kontroli kvalitete sovpada s tveganjem, da sprejmemo kvaliteto, čeprav ne ustreza, imenujemo v statistični kontroli kvalitete napako druge vrste tveganje potrošnika (consumers risk).

lo. 7. Moč preskusa

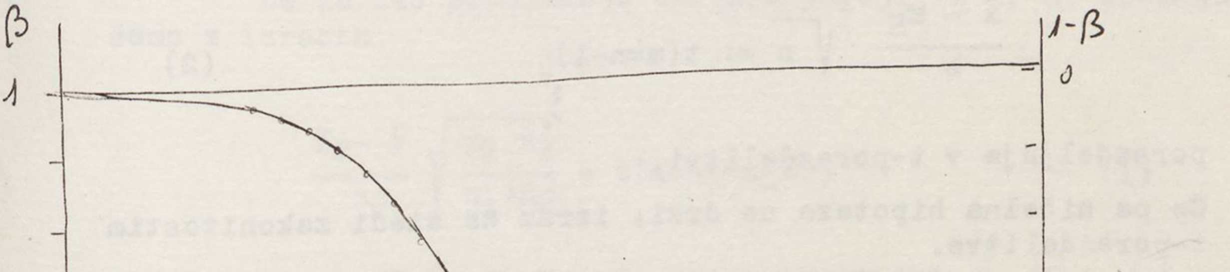
Ker je kvaliteta preskusa tem večja, s čim večjo verjetnostjo zavrnemo ničelno hipotezo, če je ta napačna, imenujemo to verjetnost moč preskusa. Ker je β napaka druge vrste, ki je v tem da ničelno hipotezo sprejmemo, če je ta napačna, je moč preskusa enaka $1-\beta$.

lo. 8. Operativna karakteristična krivulja

Krivulja, ki pokaže, kako je za konkreten plan napaka druge vrste β ali moč preskusa $1-\beta$ odvisna od prave vrednosti parametra, imenujemo operativno karakteristično krivuljo.



operativna karakteristična krivulja



Operativna karakteristična krivulja za enostranski preskus.

1o. 9. Preskušanje hipotez o aritmetični sredini

Z malimi vzorci preskušamo hipoteze o artimetičnih sredinah z uporabo zakonitosti

$$\frac{\bar{x} - M}{s}\sqrt{n} =: t(m=n-1) \qquad (1)$$

Če velja ničelna hipoteza $H_o: M = M_H$, da je prava vrednost aritmetične sredine enaka hipotetični, se tudi izraz

$$\frac{\bar{x} - M_H}{s}\sqrt{n} =: t(m=n-1) \qquad (2)$$

porazdeljuje v t-porazdelitvi.

Če pa ničelna hipoteza ne drži, izraz ne sledi zakonitostim t-porazdelitve.

Ker je majhna verjetnost, da bi po izrazu vrednost t padla v kritično območje, če ne velja ničelna hipoteza, smatramo v tem primeru, da so razlike značilne, kar pomeni, da z določeno stopinjo tveganja lahko zavržemo ničleno hipotezo ($H_o:M=M_H$) in sprejmemo osnovno hipotezo ($H_1: M \neq M_H$), da je prava vrednost različna od hipotetične.

Primer: Za primer vzemimo preskus hipoteze o trdnosti patentirane jeseniške žice. Preskušamo hipotezo, da je poprečna trdnost določene partije patentirane jeseniške žice različna od M = 186.pk/mm2. V ta namen smo vzeli vzorec n = 6 preskušancev in zanj dobili: $\bar{x} = 187,55$ in s = o,64. Da preskusimo osnovni hipotezi ustrezno ničelno hipotezo $M = M_H = 188$, vstavimo zgornje podatke v obrazec 2 in dobimo:

$$\frac{187,55 - 188}{o,64}\sqrt{6} = - 4,42 = t$$

Stopinjam prostosti m = n-1 = 6-1 = 5 ustrezne kritične vrednosti so:

$t_{(m=5)} = 2{,}57$; $t_{(m=5)} = 4{,}o3$; $t_{(m=5)} = 6{,}86$
$\alpha=2P=o{,}o5$ $\alpha=2P=o{,}o1$ $\alpha=2P=o{,}ool$

Ker je absolutna vrednost izračunanega t večja kot ustrezna kritična vrednost z α= o,ol in manjša kot kritična vrednost za α= o,ool zaključimo, da je prava vrednost za aritmetično sredino značilno različna od hipotetične M_H = 188 s tveganjem α= o,ol.

lo. lo. Preskušanje razlik med aritmetičnima sredinama

Če za dve populaciji veljajo pogoji iz 9. 6, preskušamo z izrazom

$$\frac{\bar{x}_2 - \bar{x}}{s_d} \sqrt{\frac{n_1 . n_2}{n_1 + n_2}} = t(m=n_1+n_2-2) \qquad (1)$$

ničelno hipotezo $H_o : M_1 = M_2$, da sta aritmetični sredini med seboj enaki.

Če izpade izračunana vrednost za t absolutno večja kot kritična vrednost z ustreznim tveganjem α, zaključimo, da so razlike med M_1 in M_2 značilno različne s tveganjem α.

Primer: Preskušamo razlike v poprečni trdnosti med patentirano jeseniško žico in standardno žico:

Ustrezna ničelna hipoteza je, da je $M_1 = M_2$

Da bi preskusili to ničelno hipotezo, smo vzeli n_1 = 6 preskušancev trdnosti po standardni metodi in n_2 = 6 preskušancev za patentirano jeseniško žico. Iz podatkov vzorca izračunana razlika poprečij $\bar{x}_2 - \bar{x}_1$ = 2,87 kp/mm2 po obrazcu 3 iz 9. 6 izračunani poprečni standardni odklon pa s_d = o,994.

Vzorčni izraz 1 je za te podatke

$$t = \frac{2{,}87}{o{,}994} \sqrt{\frac{6 . 6}{6+6}} = 5{,}o4$$

$m = n_1 + n_2 - 2 = 6+6-2 = 10$ stopinjam prostosti ustrezna kritična vrednost

$t(m=10) = 4{,}59,$

$\alpha = 2P = 0{,}001$

je torej manjša kot izračunana. Zato moremo sklepati, da je trdnost patentirane jeseniške žice visoko ($\alpha=0{,}001$) značilno različna od standardne.

10. 11. Preskušanje hipotez o varianci

Če proučujemo normalno porazdelitev in drži ničelna hipoteza $\sigma^2 = \sigma_H^2$ velja glede na zakonitosti iz 9, 9

$$\frac{(n-1)s^2}{\sigma_H^2} =: \chi^2(m=n-1) \qquad (1)$$

Podobno kot pri aritmetični sredini, je majhna ~~vrednost~~ verjetnost, da je vrednost vzorčnega izraza večja od ustreznih kritičnih vrednosti, če velja ničelna hipoteza. Zato v takih primerih sklepamo, da $\sigma^2 \neq \sigma_H^2$ in zaključimo, da je prava vrednost variance σ^2 značilno različna od hipotetične σ_H^2.

Primer. Pri preskušanju odpornosti materiala preskušamo hipotezo, da je varianca števila pulzacij pri dani obtežbi do preloma večja kot 4000 $H_1: \sigma^2 > 4000$ Ustrezna ničelna hipoteza je $H_0: \sigma^2 = 4000$. Dano ničelno hipotezo preskušamo z vzorcem $n = 12$ preskušancev in dobimo, da je iz vzorca ocenjena varianca $s^2 = 11116$. Če dobljene podatke vstavimo v obrazec 1 dobimo

$$\chi^2 = \frac{(12-1).11116}{4000} = 33{,}35$$

Ker je kritična vrednost za $\chi^2(m=11) = 32{,}91$ (glej tablico kritičnih vrednosti v 9.3) sklepamo, da je varianca za število pulzacij značilno večja od hipotetične s tveganjem $\alpha = 0{,}001$.

10. 12.

Preskušanje hipoteze o razlikah med variancama

Ničelno hipotezo, da je variabilnost v dveh populacijah enaka $H_o : \sigma_1^2 = \sigma_2^2$ preskušamo z dvema neodvisnima vzorcema iz dveh normalno porazdeljenih populacij z vzorčnim izrazom

$$\frac{s_1^2}{s_2^2} =: F(m_1=n-1\ ;m_2=n_2-1) \qquad (1)$$

ki sledi iz obrazca 2 v 9.9, če je $\sigma_1^2 = \sigma_2^2$

V razmerju v obrazcu vzamemo kot s_1^2 vedno večjo oceno. Za primer vzemimo proučevanje variabilnosti števila pulzacij pri različnih pritiskih.

Za vzorec n = 8 preskušancev pod prvim pritiskom smo dobili $s_1^2 = 12545$, za vzorec n = 8 preskušancev pod drugim pritiskom pa $s_2^2 = 8169$.

Če predpostavimo normalnost obeh populacij in vstavimo dobljene rezultate v obrazec 1, dobimo

$$F = \frac{12545}{8169} = 1,535$$

Iz tabele o kritičnih vrednostih za F porazdelitev v 9,5 ~~je kritična vrednost F porazdelitev v 9.5~~ je kritična vrednost $F(m_1=7;m_2=7) = 3,79$
$\alpha = 0,05$

Ker je iz podatkov vzorca izračunani F manjši kot kritična vrednost za F na stopnji tveganja $\alpha = 0,05$, smatramo, da so razlike med variancama neznačilne. To pomeni, da razlik z izvedenim preskusom nismo odkrili.

10.13. Preskušanje hipotez o frekvenčnih porazdelitvah

Hipoteze o frekvenčnih porazdelitvah preskušamo s χ^2 testom. Izraz

$$\chi^2 = \sum \frac{(f - f')^2}{f'} \qquad (1)$$

pri čemer pomenijo f stvarne, f' pa hipotetične frekvence se namreč porazdeljujejo v X^2 porazdelitvi z $m = k - p$ stopinjami prostosti, pri čemer je k število razredov v frekvenčni porazdelitvi, p pa število omejitev, ki vežejo stvarne in teoretične frekvence.

Primer: V treh partijah je bilo proizvedeno po vrsti: $n_A = 3000$, $n_B = 4000$ in $n_C = 5000$ artiklov, v posameznih partijah pa je bilo: $d_A = 60$, $d_B = 50$, $d_C = 70$ defektnih artiklov.

Če hočemo preskusiti homogenost v kvaliteti teh treh partij, je pri ničelni hipotezi enaki kvaliteti partij od skupno $d = 180$ defektnih artiklov: $d'_A = 45$, $d'_B = 60$ in $d'_C = 75$. Iz teh podatkov in stvarnih frekvenc izračunani χ^2 je

$$X^2 \frac{(60 - 45)^2}{45} + \frac{(50 - 60)^2}{60} + \frac{(70 - 75)^2}{75} = 7{,}00$$

Število stopinj prostosti je $m = 3 - 1 = 2$, ker so skupno tri frekvence v frekvenčni porazdelitvi vezane na skupno število defektnih artiklov $d = 180$.

Kritična vrednost za $\chi^2_{\alpha=0,05}$ (m=2) = 5,99, $X^2_{\alpha=0,01}$ (m=2) = 9,21, iz česar sklepamo, da so razlike v kvaliteti med tremi partijami značilno različne in sicer na nivoju $\alpha = 0{,}05$.

10.14. Preskušanje hipoteze o normalnosti frekvenčne porazdelitve

S χ^2 preskusom preskušamo tudi eno izmed zelo pomembnih hipotez: hipotezo o normalnosti frekvenčne porazdelitve. S tem preskusom preskušamo ničelno hipotezo, da je znak x v osnovi populaciji, iz katere smo na slučajnosten način izbrali vzorec, normalno porazdeljen.

Ker je stvarni porazdelitvi prilagojena normalna porazdelitev vezana na enak obseg, aritmetično sredino in standardni odklon, imamo tri omejitve, ki vežejo stvarne in teoretične frekvence. Število stopenj prostosti, ki pride v poštev pri preskušanju značilnosti razlik distribucij od normalne je torej $m = k-p = k-3$.

Za primer vzemimo frekvenčno porazdelitev trdnosti patentirane jeseniške žice za n = 153 preskušancev. To je porazdelitev, za katero smo že izračunali aritmetično sredino in varianco v odstavkih o izračunavanju teh parametrov.

x_k	f_k		f'_k		Δf_k	$(\Delta f_k)^2/f'_k$
144,5	2					
146,5	-					
148,5	1	6	o,2	7,1	-1,1	o,173
15o,5	1		1,4			
152,5	2		5,5			
154,5	14		15,7		-1,7	o,184
156,5	22		29,4		-7,4	1,86o
158,5	41		37,6		+3,4	o,3o8
16o,5	46		33,o		+13,o	5,13o
162,5	21		2o,o		1,o	o,o5o
164,5	2	3	7,8	1o,2	-7,2	5,o9o
166,5	1		2,4			
	153		153,o		0	12,795 = χ^2

Ker aproksimacija s χ^2 velja le, če so teoretične frekvence večje kot 5, smo skrajne frekvence združili v širše razrede, da je temu pogoju zadoščeno. Po združevanju ostane še k = 7 razredov. Število stopinj prostosti je po zgornjem pravilu m = k-3 = 7-3 = 4.

V tablicah za kritične vrednosti za χ^2-porazdelitev dobimo
$X^2(4) = 11,67$, $X^2(4) = 13,28$. Iz tega zaključimo, da je
$\alpha = o,o2$ $\qquad\qquad$ $\alpha = o,o1$
porazdelitev o trdnosti za patentirano jeseniško žico značilno različna od normalne in sicer s tveganjem α = o,o2. Vsebinska analiza tega zaključka pa mora odkriti vzroke nenormalnosti.

lo. 15. <u>Alternativni obrazec za izračunavanje X^2</u>

Z enostavnim preračunom iz definicijskega obrazca za χ^2

$$X^2 = \sum \frac{(f-f')^2}{f'} \qquad (1)$$

izpeljemo obrazec

$$x^2 = \sum \frac{f^2}{f'} - n \qquad (2)$$

ki je v nekaterih primerih računsko prikladnejši kot definicijski.

lo. 16. Analiza variance

Analiza variance je metoda kompleksnega preskušanja hipoteze o razlikah med aritmetičnimi sredinami za več vzorcev oziroma grup hkrati. Pri analizi variance primerjamo dve neodvisni oceni za varianco. Od teh je ena-s_k^2 ocenjena iz ocen grupnih aritmetičnih sredin, druga-s_e^2 pa je ocenjena kot aritmetična sredina grupnih varianc, če se x po grupah porazdeljuje normalno z enakimi variancami $\sigma_1^2 = \sigma_2^2 = \sigma^2$ in če velja ničelna hipoteza, da so prave aritmetične sredine med seboj enake ($M_1 = M_2 = M_3 = M_4 = M_k$) se

$$F = s_k^2/s_e^2 \qquad (1)$$

porazdeljuje v F porazdelitvi z $m_1 = k-1$ in $m_2 = n-k$ stopinjami prostosti. Če ne velja ničelna hipoteza o enakosti aritmetičnih sredin se F značilno poveča (prekorači kritično vrednost za F), kar smatramo za znak značilnih razlik med grupnimi aritmetičnimi sredinami.

Analiza variance je osnova za preskušanje hipotez o vplivih različnih faktorjev in je podlaga za posebno statistično disciplino in sicer planiranje eksperimentov.

lo. 17. Enostavna analiza variance

Kadar preskušamo značilnost razlik v učinku enega samega faktorja, govorimo o enostavni analizi variance.

Vpliv enega samega faktorja obračunamo z metodo analize variance na splošno po naslednjem postopku:

a) s ponovitvami na posameznih nivojih faktorja K in ustreznimi meritvami dobimo osnovno numerično gradivo za enostavno analizo variance x_{ki} = meritev za preskušanec i pod pogojem k,

b) Iz x_{ki} izračunamo vsote X_k za posamezne nivoje faktorja $X_k = \underset{i}{S} x_{ki}$ in skupno vsoto $X = \underset{k}{S} X_k$

c) Iz podatkov iz b izračunamo:

$$Q_{Ki} = SSx_{ki}^2$$

$$Q_K = S\frac{X_k^2}{n_k}$$

$$Q = \frac{X^2}{n}$$

Iz dobljenih podatkov obračunamo analizo variance po naslednji shemi:

Vir variacije	Vsota kvadratov	Stopinje prostosti	ocena variance	Izračunani F
Med nivoji	$K_k = Q_k - Q$	$m_k = k-1$	s_k^2	$F = s_k^2/s_e^2$
v nivoju	$K_{KI} = Q_{KI} - Q_k$	$m_e = n-k$	s_e^2	1
Skupno	$K = Q_{KI} - Q$	$m = n-1$		

Če je izračunani F večji od kritične vrednosti za $F\,[m_1 = (k-1);\ m_2 = (n-k)]$ iz tablic, so razlike v učinku faktorja K značilne. V tem primeru je smiselno analizirati diference v učinku faktorja na različnih nivojih po obrazcu

$$\Delta M = \left(\bar{x}_2 - \bar{x}_1\right) \pm t(m=n-k) . s_e \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

Primer: Na platiščih iz treh različnih šarž smo izvedli meritve trdnosti in dobili naslednje rezultate:

Šarža 1: 19 79 27 64 59 4o 48 25 51 59 73 69 58
Šarža 2: 71 91 43 45 73 49 114 54 84 67 98 81 85 59 65 66 66 6o 61
Šarža 3: 69 71 12 42 55 62 77 36 56 52 48 91 44 58

Iz teh osnovnih podatkov dobimo dalje

$n_1 = 13 \quad X_1 = Sx_{1i} = 671$

$n_2 = 19 \quad X_2 = Sx_{2i} = 1332$

$n_3 = 14 \quad X_3 = Sx_{3i} = 773$

$n = 46 \quad X = SX_k = 2776$

Iz zgornjih podatkov izračunamo po točki c sheme:

$$Q_{KI} = SSx^2_{ki} = 19+79+27+.........+44+58 = 186o14$$

$$Q_K = S\frac{X_k^2}{n_k} = \frac{671^2}{13} + \frac{1332^2}{19} + \frac{773^2}{14} = 17o695$$

$$Q = X^2/n = \frac{2776^2}{46} = 167526$$

Iz teh podatkov pa dobimo po osnovni shemi:

Vir variacije	Vsota kvadratov	Stopinje prostosti	Ocena variance	F
Med šaržami	17o695-167526=3169	3-1=2	1584	4,45[x]
Znotraj šarž	186o14-17o695=15319	46-3=43	356	1,oo
Skupno	18488	45		

Ker je dobljena vrednost F = 4,45 večja kot pa je kritična vrednost za $F(2,43_{o,o5} = 3,2$ in manjša kot je $F(2,43)_{o,o1} = 5,2$, sklepamo, da so razlike v trdnosti plastišč med šaržami značilne na nivoju $\alpha = 0,05$. Točkovne ocene aritmetičnih sredin so:

$$x_1 = X_1/n_1 = 671/13 = 51,7$$

$$x_2 = X_2/n_2 = 1332/19 = 70,1$$

$$x_3 = X_3/n_3 = 773/14 = 55,2$$

Iz ocen poprečij vidimo, da je drugi postopek dal bistveno boljšo trdnost od postopka 1 in 3 in značilnost razlik med postopki izvira iz tega dejstva. Če izračunamo še intervalno oceno za razliko med prvim in drugim postopkom, dobimo po obrazcu

$$M_2 - M_1 = 70,1 - 51,7 \pm 2,02 \cdot \sqrt{356} \sqrt{\frac{13+19}{13.19}} = 18,4 \pm 13,7$$

intervalno oceno s tveganjem $\alpha = 0,05$, ker je $t(m=n-k= 43)$

11. Statistično planiranje eksperimentov

11. 1. Cilj

Osnovni cilj eksperimentiranja je ugotavljanje učinkov različnih faktorjev na proučevani pojav. Izraz faktorjev so vrednosti faktorialnih znakov. Ti morejo biti bodisi atributivni ali faktorialni. Proučevani pojav je rezultativen znak, ki pa je po pravilu numeričen. Cilj statistične analize eksperimetna je preskušanje značilnosti vpliva posameznih faktorjev na rezultativen znak, orodje preskušanja hipotez pa je analiza variance. Statistični plan eksperimetna omogoča na eni strani odstraniti iz obračuna vpliv določljivih a nepomembnih faktorjev, ki motijo in zamegljujejo vpliv vsebinsko pomembnih faktorjev. Razen tega pa skupni učinek večih faktorjev razstavljamo na vplive posameznih faktorjev kot njihovega vzajemnega učinka.

11. 2. Osnovni elementi statističnega plana eksperimentov

Učinek določenega faktorja ali kompleksa faktorjev na določen pojav je tem bolj jasen, čim manjši je vpliv slučajnostnih ali drugih individualnih vplivov, ki motijo poskus. Pri statistično pravilno planiranem eksperimentu upoštevamo opredeljive, a za eksperiment nepomembne faktorje in jih s pravilno izvedenim eksperimentom izločimo iz obračuna.

Medtem ko vpliv opredeljivih, a vsebinsko nepomembnih faktorjev izločimo s pravilno planiranim eksperimentom, vpliv slučajnostnih faktorjev zmanjšamo s ponavljanjem poskusa pri enakih pogojih. Sumarni vpliv slučajnostnih faktorjev je za povprečja manjši, ker je varianca poprečja iz n ponavljanj enaka σ^2/n

Prav tako je razdružitev učinka večjega števila faktorjev naloga pravilnega plana eksperimenta.

Tretji bistveni element statističnega plana eksperimenta je slučajnost razporeda. S slučajnostnim razporedom faktorje , ki so sicer opredeljivi, a jih s planom ne izločimo, vključimo v sklop slučajnostnih faktorjev in s tem omogočimo nepristranske in objektivne zaključke.

11. 3. Linearni modeli plana eksperimentov

Teoretični izraz statističnega plana eksperimenta je model. Ta nakazuje, kako je rezultativen znak odvisen od vseh faktorjev, splošnih pomembnih ali nepomembnih opredeljivih faktorjev in slučajnostnih. Za računsko analizo so najprimernejši linerani modeli. V njih so učinki posameznih faktorjev vezani linearno. Vzemimo, da na pojav, katerega zunanji izraz je numerični znak, vplivajo splošni vplivi (za vse eksperimentalne enote isti) nebistveni, a opredeljiv faktor B, opredeljiva in pomembna faktorja K in L in slučajnostni vplivi e. Linearni model, ki kaže, kako je rezultativen znak odvisen od posameznih komponent, je

$$x_{bkli} = M + (B) + (K) + (L) + e_{bkli} \qquad (1)$$

pri čemer je : x_{bkli} = vrednost karakteristike, M = rezultat splošnih vplivov, (B) = rezultat opredeljivega, a nepomembnega faktorja B, (K) in (L) faktorja, katerih vpliv raziskujemo, e_{bkli} pa učinek slučajnostnih faktorjev pri individualnem poskusu.

Če vsebina pojava ne nakazuje linearne zveze med faktorji, jih skušamo na linearni model prinesti s transformacijo. Če je vsebinsko upravičena multiplikativna zveza med komponentami, tak model privedemo na aditiven z logaritmiranjem.

11. 4. Slučajnostni bloki

Najenostavnejši plan eksperimenta, pri katerem izključimo en opredeljiv, a vsebinsko nepomemben faktor, so slučajnostni bloki. S slučajnostnimi bloki odstranimo vpliv heterogenosti v eksperimentalnem materialu, heterogenosti v izvedbi poskusov, heterogenosti v produktivnosti med delavci ipd. Ta prijem v mnogih primerih doprinese k uspešnejšim rezultatom in k temu, da se standardna pogreška ocen in s tem kvaliteta zaključkov zviša, ne da bi povečali število ponovitev poskusov in s tem podražili izvedbo poskusa.

Pri čisto slučajnostnih blokih iz eksperimentalnega gradiva tvorimo skupine s čimbolj enotnimi pogoji. Število enot v skupini je enako številu nivojev pomembnega faktorja. Na vsaki taki homogeni skupini - bloku, apliciramo po en nivo faktorja, katerega raziskujemo. Te faktorje apliciramo na eksperimentalne enote na slučajnosten način, da se eventualna nehomogenost pri sistematični uporabi ne vmeša v učinek faktorja, ampak preide v slučajnostno komponento. V tako planiranem poskusu je nehomogenost med bloki eliminirana iz obračuna rezultatov poskusa, kar je cilj slučajnostnih blokov.

Model eksperimenta za slučajnostne bloke enega faktorja je:

$$x_{bp} = M + (B) + (P) + e_{bp} \qquad (1)$$

Obračun analize variance pri slučajnostnih blokih pa je naslednji:

Po izvedenem poskusu, ki smo ga izvedli v skladu z zgornjimi navodili, razpolagamo za vsak blok b, za vsak nivo faktorja s po enim podatkom x_{bp};

b) Iz osnovnih podatkov x_{bp} izračunamo vsote: $X_b = \underset{p}{S}\, x_{bp}$ in $X = S\, X_b$; $X_p = \underset{b}{S}\, x_{bp}$

c) Iz dobljenih vsot po standardnem načinu za analizo variance izračunamo količine

$$Q_{BP} = \underset{bp}{SS}\, x^2_{bp}\,;\; Q_B = \frac{1}{p}\underset{b}{S}\, X^2_b\,;\; Q_P = \frac{1}{b}\underset{p}{S}\, X^2_p\,;\; Q = \frac{1}{bp} X^2 \quad (2)$$

d) Iz teh količin obračunamo analizo variance po shemi: (3)

Vir variacije	Vsota kvadratov	Stopinje prostosti	Ocena variance	F
Bloki	$K_B = Q_E - Q$	$m_b = b-1$		
faktor P	$K_p = Q_P - Q$	$m_p = p-1$	$S^2_e = K_P/K_e$	$F = s^2_P/s^2_e$
ex.pogreška	$K_e = Q_{PB} - Q_B - Q_P + Q$	$m_e = (b-1)(p-1)$	$S^2_e = K_e/m_e$	1
Skupno	$K = Q_{BP} - Q$	$m = bp-1$		

Če se izkažejo razlike med postopki za značilne, naprej analiziramo aritmetične ocene sredin

$$\bar{x}_p = X_p/p$$

Zanje je ocena standardne pogreške $se_{\bar{x}_p} = s_e/\sqrt{p}$, ocena standardne pogreške ocen razlik med dvema aritmetičnima sredinama pa

$$se_{\Delta\bar{x}} = \sqrt{\frac{2s^2_e}{p}}$$

Te standardne pogreške služijo bodosi za preskušanje hipotez ali za izračunavanje intervalnih ocen poprečij ali razlik med poprečji.

Za primer poskusa s slučajnostnimi bloki vzemimo analizo vzdržnosti platišč, obdelovanih pri različnih žarilnih temperaturah. Kot kazalec vzdržljivosti je vzeto število pulzacij do zloma. Poskus je predvideval tri nivoje žarilne temperature $t_1 = 45o^{o}$; $t_2 = 55o^{o}$; $t_3 = 65o^{o}$.

Poskus je bil izveden v petih ponovitvah na po treh platiščih iz petih šarž, ki so vzete kot bloki. S tem skušamo eliminirati vpliv šarž na vzdržljivost in tako eliminirati en določljiv, a nepomemben faktor iz obračuna. Iz vsake šarže so bila vzeta po tri platišča, na njih pa na slučajnostno izbranih platiščih aplicirane posamezne temperature. Rezultati poskusa so naslednji:

Šarža	x_{bt}			X_b
1	t_1:335	t_3:253	t_2:9o6	1494
2	t_2:4o1	t_3:264	t_1:314	979
3	t_1:211	t_2:39o	t_3:133	734
4	t_2:634	t_1:394	t_3:336	1364
5	t_1:282	t_3:311	t_2:56o	1153

X_t $\quad X_{t_1} = 1536 \quad X_{t_2} = 2891 \quad X_{t_3} = 1297 \quad 5724 = X$

Nadalje dobimo:

$$Q_{BT} = \underset{bt}{SS}\, x_{bt}^2 = 27oo4o6$$

$$Q_B = \frac{1}{3}\, S\, X_B^2 = 6919138/3 = 23o6379$$

$$Q_T = \frac{1}{5}\, S\, X_t^2 = 12399386/5 = 2479877$$

$$Q = \frac{1}{15}\, X^2 = 32764176/15 = 2184278$$

Iz teh obračunano varianco po shemi za obračun slučajnostnih blokov

Vir variacije	Vsota kvadratov	Stopinje prostosti	Ocena variance	F
Bloki(šarže)	$K_B = 2306379-2184278 = 122101$	5-1=4		
žarilna temperatura	$K_T = 2479877-2184278 = 295599$	3-1=2	147800	12.01[xx]
ex.pogreška	$K_e = 2479877+2184278 = 98428$	(5-1)(3-1) =8	12304	1.00
Skupno	2700406-2184287= 516128	15-1=14		

Ker je kritična vrednost $F(2,8)_{0,01} = 8,65$ manjša, $F(2,8)_{0,001} = 18,49$ pa večja kot izračunani F, smtramo razlike v številu pulzacij kot značilne v odvisnosti od žarilne temperature na nivoju $\alpha = 0,01$. To opravičuje nadaljnjo analizo razlik med posameznimi temperaturami.

Iz vsot X_t dobimo poprečno število pulzacij pri posameznih žarilnih temperaturah je:

$$\bar{x}_{t_1} = 307; \quad \bar{x}_{t_2} = 964; \quad \bar{x}_{t_3} = 259$$

Maksimalni verjetnostni odklon ocen od pravih aritmetičnih sredin je

$$d_{\bar{x}_t} = t(8)_{\alpha = 0,05} \, s_e / \sqrt{b}$$

ali v našem primeru

$$d_{\bar{x}_t} = 2,306 \frac{110,9}{\sqrt{5}} = 100$$

Če želimo preskusiti značilnost razlik med $\bar{x}_{t_2}$ in $\bar{x}_{t_1}$ uporabimo obrazec, po katerem je preskusni izraz

$$t = \frac{\bar{x}_2 - \bar{x}_1}{s_e} \sqrt{\frac{b}{2}} = \frac{259 - 307}{110,9} \cdot \sqrt{\frac{5}{2}} = 0,685$$

Že brez tablic za kritične vrednosti sklepamo, da razlike niso značilne.

11. 5. Metoda parov

Poseben primer slučajnostnih blokov je metoda parov. Po tej metodi preskušamo hipoteze in ocenjujemo razlike med postopki, kadar sta v blokih samo dva postopka.

Iz splošnega linearnega modela za slučajnostne bloke

$$x_{bp} = M + (B) + (P) + e_{bp} \tag{1}$$

dobimo za prvi in drugi postopek za vsak blok

$$x_{b1} = M + (B) + P_1 + ebp$$

$$x_{b2} = M + (B) + P_2 + e_{b2} \tag{2}$$

$$d_b = x_{b2} - x_{b1} = \left(P_2 - P_1\right) + \left(e_{b2} - e_{b1}\right)$$

Razlika vrednosti obeh postopkov v blokih d_b vsebuje razliko med postopki P_2-P_1, razlike med bloki pa iz diference izginejo. Tako smo z diferencami en opredeljiv, a nepomemben faktor izločili in s tem dobili osnovo za preciznejše sklepe. Iz d_b moremo po znanih pravilih skonstruirati izraz

$$\frac{\bar{d}_b - \Delta P}{s_d} \sqrt{b} = t(m = b - 1); \quad s_d^2 = \frac{S(d_b - \bar{d})^2}{b - 1} \tag{3}$$

ki ga uporabljamo ali za ocenjevanje razlik med postopki ali za preskušanje ničelne hipoteze $H_o : \Delta P = P_2 - P_1 = o$.

Za primer vzemimo proučevanje značilnosti razlik števila pulzacij do porušitve pri dveh različnih obtežah. V ta namen smo sedem delov pla~~s~~tišč (bloki) razdelili v dva enaka dela in iz vsakega bloka za en del preskušali število pulzacij do porušitve pri eni, za drugega pa pri drugi obtežitvi. Rezultati preskusov in preskus značilnosti je naslednji:

Blok b	x_{b1}	x_{b2}	d_b	d_b^2
1	335	394	+59	3481
2	253	336	+83	6889
3	314	282	-32	1o24
4	4o1	56o	+159	25281
5	264	311	+47	22o9
6	211	253	+42	1764
7	39o	513	+123	15129
8	133	247	+114	12996
		$Sd_b =$	595	$Sd_b^2 = 68773$

Iz zgornjih podatkov dobimo, da je $\bar{d} = 595:8 = 74,4$, $s_d^2 = 35o3$ in $s_d = 59,2$.

Če preskušamo hipotezo o razlikah med poprečnim številom pulzacij pri različnih obtežbah, postavimo za ničelno hipotezo, da je $\triangle P = o$. Tako dobimo

$$t = \frac{\bar{d}}{s_d} \cdot \sqrt{n} = \frac{74,4}{59,2} \cdot \sqrt{8} = 3,56$$

Ker je izračunani t večji kot $t_{o,ol}(m = b - 1) = 3,5o$ in manjši kot je $t_{o,ool}(m=7) = 5,4o$ zaključimo, da je poprečno število pulzacij pri obeh preskušanih obtežbah s tveganjem $\alpha = o,ool$ ničelno različno.

11. 6. Faktorialni poskusi

Čeprav je že eksperiment, v katerem proučujemo z analizo variance en sam faktor, po vsebini faktorialni eksperiment, se je ustalila terminologija, da imenujemo faktorialni eksperiment tisti plan oziroma eksperiment, v katerem proučujemo vpliv večih faktorjev hkrati. Tako je "faktorialni eksperiment pxq" eksperiment, v katerega sta vključena dva vsebinsko pomembna faktorja in sicer eden na p, drugi pa na q nivojih. Kot "faktorialni eksperiment 2^4" pa npr. imenujemo eksperiment v katerem proučujemo ističasno štiri faktorje, od katerih je vsak na dveh nivojih: npr. dva različna pogoja dela, dva recepta, dve temperaturi vlivanja itd.

11. 8. Interakcija

Če je pri različnih nivojih za faktor A vpliv faktorja B na rezultat različen, pravimo, da je med faktorjem A in B interakcija. Interakcijo tolmačimo kot vzajemen vpliv dveh ali več faktorjev.

Če med vplivom temperature in vplivom staranja na žilavost kotlovske pločevine ne bi bilo interakcije, bi se žilavost spreminjala po modelu

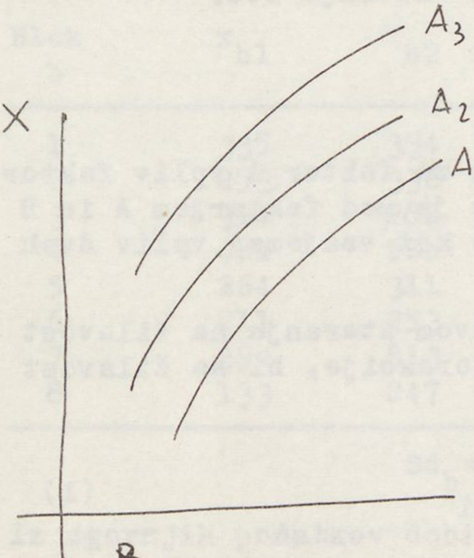
$$x_{tsi} = M + (S) + (T) + e_{tsi} \tag{1}$$

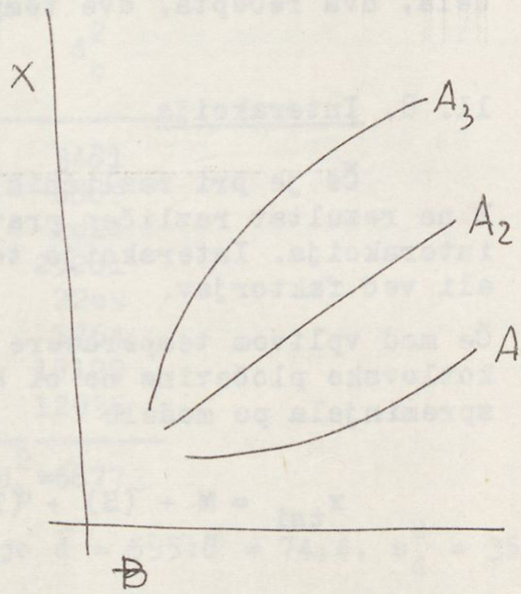
Krivulji o žilavosti v odvisnosti od temperature, bi za starano in nestarano stanje v primeru, da ni interakcije, potekali paralelno.

Če pa obstaja med staranjem in temperaturo interakcija, nastopi v modelu nova komponenta (ST) vzajemnega učinka vpliva staranja in temperature. V tem primeru ima linearni model naslednje komponente:

$$x_{tsi} = M + (S) + (T) + (ST) + e_{tsi} \tag{2}$$

Shematično moremo v sliki nakazati interakcijo med dvema faktorjema A in B v sliki





Med faktorjema A in B ni interakcije

Med faktorjema A in B je interakcija

11. 9. Faktorialni poskus pxq

Poskus, v katerem proučujemo dva faktorja, enega na p nivojih, drugega na q nivojih, imenujemo faktorialni poskus pxq. Analiza faktorialnega poskusa pxq in faktorialnih poskusov nasploh je najenostavnejša, če je število ponovitev za vsako kombinacijo nivojev faktorjev enaka.

Linearni model faktorialnega poskusa z dvema faktorjema P in Q z i ponovitvami je:

$$x_{pqi} = M + (P) + (Q) + (PQ) + e_{pqi} \qquad (1)$$

Za tak poskus potrebujemo za vsako kombinacijo PQxi ponovitev. Iz podatkov poskusa x_{pqi} z analizo variance analiziramo vpliv faktorjev po naslednjem postopku:

a) Iz osnovnih podatkov eksperimenta x_{pqi} izračunamo vsote:

$$X_{pq} = \underset{i}{S}\, x_{pqi} \quad X_p = \underset{q}{S}\, X_{pq} \quad X_q = \underset{p}{S}\, X_{pq} \quad X = \underset{p}{S}\, X_{pq} = \underset{q}{S}\, X_{pq} \qquad (2)$$

b) Iz osnovnih podatkov in vsot izračunamo pomožne količine:

$$Q_{PQI} = \underset{pqi}{SSS}\, x^2_{pqi}\,;\ Q_{PQ} = \frac{1}{i} \underset{pq}{SS}\, X^2_{pq}\,;\ Q_P = \frac{1}{iq} SX^2_p;\ Q_Q = \frac{1}{iq} \underset{q}{S}\, X^2_q$$

$$Q = \frac{1}{ipq} X^2$$

c) Iz dobljenih količin Q analiziramo varianco po naslednji standardni shemi za analizo variance za poskus pxq ~~z~~ i ponovitvami:

Vir variacije	Vsota kvadratov	Stopinje prostosti	Ocena variance	F
P	$Q_P - Q = K_P$	$m_P = p - 1$	$s_P^2 = \frac{K_P}{m_P}$	$F_P = s_P^2/s_e^2$
Q	$K_Q = Q_Q - Q$	$m_Q = q - 1$	$s_Q^2 = K_Q/m_Q$	$F_Q = s_q^2/s_e^2$
PxQ	$K_{PQ} = Q_{PQ} - Q_P - Q_Q + Q$	$m_{PQ} = (p-1)(q-1)$	$s_{pq}^2 = \frac{K_{PQ}}{m_{PQ}}$	$F_{PQ} = s_{PQ}^2/s_e^2$
eks.pogr.	$K_e = Q_{PQI} - Q_{AB}$	$m_e = pq(i-1)$	$s_e^2 = K_e/m_e$	1
Skupno	$K = Q_{PQI} - Q$	$m = pqi - 1$		

Če se razen čistih efektov posameznih faktorjev (P) in (Q) izkaže značilna tudi interakcija (PQ), naprej analiziramo poprečja $\bar{x}_{pq} = \frac{1}{i} S_i \, x_{pqi}$.

Maksimalni verjetnostni odklon teh poprečij je

$$d_{\bar{x}_{pq}} = t_{\alpha}(m=m_e)\,\frac{s_e}{\sqrt{i}}$$

maksimalni verjetnostni odklon razlike dveh poprečij pa

$$d_{\Delta\bar{x}_{pq}} = t_{\alpha}(m=m_e)\sqrt{\frac{2s_e^2}{i}}$$

Če pa je interakcija neznačilna, z njo v nadaljnem ne računamo in analiziramo poprečja

$$\bar{x}_p = \frac{1}{q} S_q X_{pq} \quad \text{in} \quad \bar{x}_q = \frac{1}{p} S_p X_{pq}$$

Za te ocene so ocene maksimalnih verjetnih odklonov

$$d_{\bar{x}_p} = t_{\alpha}(m=m_e)\,\frac{s_e}{\sqrt{i.q}} \qquad d_{\bar{x}_q} = t_{\alpha}(m=m_e)\,\frac{s_e}{\sqrt{i.p}}$$

maksimalni verjetni odkloni dveh poprečij pa:

$$d_{\Delta\bar{x}p} = t_{\alpha}(m=m_e)\sqrt{\frac{2\,s_e^2}{i.q}} \qquad d_{\Delta\bar{x}p} = t_{\alpha}(m=m_e)\sqrt{\frac{2s_e^2}{i.p}}$$

Za primer faktorialnega eksperimenta z dvema faktorjema vzemimo proučevanje trdnosti določene litine v odvisnosti od temperature in časa obdelave. Pri tem so temperature pri obdelavi na naslednjih nivojih:

T_1 = 1o% pod standardno temperaturo

T_2 = standardna temperatura

T_3 = 1o% na standardno temperaturo

T_4 = 2o% nad standardno temperaturo,

čas obdelave pa je:

$Č_1$ = 2o% krajši kod standardni čas obdelave

$Č_2$ = 1o% krajši kot standardni čas obdelave

$Č_3$ = čas standardne obdelave

$Č_4$ = 1o% daljši čas kot čas standardne obdelave

Poskus je planiran tako, da je na 32 kosih na slučajnosten način aplicirana po ena izmed 16 kombinacij med časom in temperaturo v dveh ponovitvah.

Rezultati poskusa, urejeni v tabeli, so naslednji:

	$Č_1$	$Č_2$	$Č_3$	$Č_4$	X_t
T_1	16	1o	21	42	
	1o	37	27	24	
	26	47	48	66	187
T_2	25	54	72	27	
	36	55	53	59	
	61	1o9	125	86	381
T_3	24	57	5o	57	
	32	47	53	72	
	56	1o4	1o3	129	392
T_4	44	55	63	48	
	39	49	63	56	
	83	1o4	126	1o4	417
$X_č$	226	364	4o2	385	1377 = X

Če naprej izračunamo pomožne količine, dobimo:

$$Q_{TČI} = \underset{tči}{SSS} x^2_{tči} = 68461 \qquad Q_{TČ} = \frac{1}{i} SSX^2_{tč} = 66884$$

$$Q_T = \frac{1}{ič} SX^2_t = 6346o \qquad Q_Č = \frac{1}{it} \underset{č}{S} X^2_č = 61675 \quad Q = \frac{1}{itč} X^2 = 59254$$

Po shemi za analizo variance za faktorialni poskus z dvema faktorjema sledi:

Vir variacije	Vsota kvadratov	Število stopinj prostosti	Ocena variance	F
temperatura	6346o-59254= 42o6	4 - 1 = 3	14o2	$14,22^{xxx}$
čas	61675-59254= 2421	4 - 1 = 3	8o7	$8,19^{xx}$
interakcija tempXčas	66884-6346o--61675+59254= 1oo3	(4-1)(4-1) = 9	111,4	1,13
eksperimentalna napaka	68461-66884= 1577	4.4(2-1) = 16	98,6	1,oo
SKUPNO	68461-59254= 92o7	4.4(2-1) = 31		

Če primerjamo izračunane vrednosti za F s kritičnimi vrednostmi, zaključimo, da sta faktorja temperatura in čas visoko značilna, ker so kritične vrednosti:

$F(3,16)_{o,o5} = 3,24 \quad F(3,16)_{o,ol} = 5,29$ in $F(3,16)_{o,ool} = 9,oo$.

Iz tega sledi, da se je učinek temperature pokazal značilen na nivoju α = o,ool, učinek časa obdelave pa na nivoju α= = o,o5. Ker je $F(9,16)_{o,o5} = 2,54$ znatno večji od izračunanega F iz preskusa, ni razloga za trditev, da je interakcija značilna. Zato v tem okviru smatramo, da je učinek temperature in časa na trdnost aditiven in da velja model

$$X_{pqi} = M + (P) + (Q) + e_{pqi}$$

11. 1o. Drugi plani eksperimentov

Nakazani plani so najenostavnejši plani, ki se uporabljajo pri eksperimentiranju. Razen teh imamo še druge plane, s katerimi izločujemo več opredeljivih, a nepomembnih faktorjev (latinski kvadrat, grško-latinski kvadrat, changeover poskus) poskuse z več faktorji, pri katerih pa nekatere interakcije zanemarjamo, da ne dobimo prevelikih, a zato heterogenih blokov (confounding). Inkompletne bloke uporabljamo, če je število variant ali nivojev tako veliko, da bi bil blok s takim številom elementarnih poskusov heterogen itd.

12. STATISTIČNA KONTROLA KVALITETE

12. 1. Osnova

Direktna in kompleksna uporaba metod preskušanja hipotez je statistična kontrola kvalitete (SKK). Neglede na to, da moremo vse dosedaj obravnavane metode preskušanja hipotez uporabiti neposredno pri problemih kontrole kvalitete (preskušanje hipotez o standardih, razlik v postopkih itd.), se je SKK izkristalizirala v več ali manj samostojno disciplino uporabe statističnih metod pri kontroli kvalitete v masovni proizvodnji.

SKK se je v masovni in avtomatizirani proizvodnji pokazala ne samo kot uporabna, ampak včasih tudi kot neobhodna. Kompletna ali "stoodstotna kontrola" je dostkrat neuporabna iz več razlogov. Dolgotrajnost kontrole zavira tako proizvodni proces kot prevzem. Monotonost ima za posledico propuste v kontroli. Za primere, v katerih je kontrola povezana z uničenjem artikla, pa je statistična kontrola edini izhod. Ker pravilno postavljen plan za SKK zagotavlja določeno želeno kakovost, je zaradi hitrosti, s katero jo izvajamo, ekonomičnosti, ki je zvezana z vzorčno kontrolo, možnosti izbire planov glede na potrebe in možnosti in kvalitetnosti, ki je dana s parametri plana, je SKK v večini primerov izpodrinila tako kompletno kontrolo kot tudi druge načine kontrole z neobjektivnimi osnovami.

12. 2. Vrste SKK

Čeprav so vse metode SKK izdelane predvsem za aplikacijo v masovni industrijski proizvodnji, jih s pridom uporabljamo tudi na drugih področjih: npr. kontroli pisarniškega poslovanja, kontroli procesov in sprejema v trgovini ipd.

SKK po svoji funkciji pa tudi po svojih metodah, ki jih uporabljamo, delimo v dve metodološko in vsebinsko več ali manj samostojni grupi: kontrolo prevzema in kontrolo procesa. Ti grupi imata za metodološko osnovo preskušanje hipotez, vendar se po metodoloških prijemih in po vsebini razlikujeta.

12. 3. Kontrola proizvodnega procesa

Naloga kontrole proizvodnega procesa je tekoča kontrola, s katero spremljamo ali proces teče po predpisu ali ne.

Proizvodnja oziroma karakteristike proizvoedenih artiklov variirajo iz več razlogov. Odkloni od določene srednje vrednosti morejo biti slučajnostnega izvora in izvirajo iz nehomogenosti surovin, neenakomernega delovanja stroja ali delavca in podobno. Zaradi teh slučajnostnih vzrokov, se značilnosti artiklov odklanjajo sicer od povprečja v nepredvideni smeri in iznosu, vendar v skladu z zakonitostmi in velikostjo, ki je pogojena s kvaliteto strojev, homogenostjo surovin, dela delavcev itd. Drug kompleks faktorjev, ki se more pojaviti v proizvodnem procesu so sistematični faktorji. Ti imajo za posledico sistematične stalne odklone od predpisa. Tako more zaradi sistematičnega vpliva faktorja stroj začeti delati artikle, katerih trdnost se je spremenila in ne ustreza predpisanim standardom. Podobno pa more iz nekega nekontroliranega vzroka stroj začeti izdelovati artikle z večjo heterogenostjo kot pričakujemo. Naloga kontrole procesa je, da v teku proizvodnega procesa tekoče odkriva sistematične odklone od procesa pod kontrolo in s tem daje signale za odpravo vzrokov za slabo kakovost že med proizvodnjo.

Parametri, ki jih običajno kotroliramo v teku proizvodnega procesa so: mere centralne tendence, mere variacije in število ali odstotek po predpisu neuporabnih artiklov ali artiklov z dano značilnostjo.

Osnovna mera centralne tendence, ki pride v poštev pri kontroli procesa je aritmetična sredina. Razen tega iz tehničnih razlogov pogosto uporabljamo tudi mediano. Za mero homogenosti uporabljamo pri SKK varianco ali standardni odklon, enako kot mediano pa iz tehničnih razlogov, ker je varianco razmeroma komplicirano izračunavati, uporabljamo variacijski razmak.

Pri kontroli atributivnih karakteristik proizvedenih artiklov uporabljamo kot parametre bodisi število artiklov z dano značilnostjo ali odstotek artiklov z dano značilnostjo. Najpogostejši atributiven znak pri kontroli kvalitete je uporabnost artikla, ki loči artikle v uporabne in defektne. Ta znak je običajno tipološki znak, ki je rezultat kompleksa več znakov, ki so odločilni za funkcionalnost oziroma uporabnost artiklov.

SKK vršimo tako, da iz tekoče proizvodnje občasno izbiramo po obsegu običajno majhne vzorce. Na osnovi teh preskušamo hipoteze o parametrih proizvodnje, ki jih imamo pod kontrolo. Pogostost jemanja preskusnih vzorcev zavisi od različnih momentov: škode, ki nastane, če proces ne teče v redu, od verjetnosti, da nastopijo sistematični vplivi, ki spremene proizvodni proces ipd.

SKPP mora biti tehnično prilagojena tako, da z njo čim manj motimo proizvodni proces, in da jo more opravljati tudi statistično neizkušen kontrolor. To dosežemo s kontrolnimi kartami.

12. 4. Kontrolne karte ali grafikoni

Tehnično poenostavimo preskušanje hipoteze v proizvodnem procesu s kontrolnimi kartami. Kontrolna karta je grafikon, v katerem je abscisna os po pravilu časovna os oziroma os za registracijo sukcesivnih vzorcev, ordinata pa skala za vnašanje rezultatov iz preskusnih vzorcev. Osnovne elemente kontrolne karte prikažimo na primeru kontrolne karte za aritmetično sredino.

12. 5. Kontrolna karta za aritmetično sredino

Vzemimo, da je povprečen premer v tekoči proizvodnji glav zakovic $M = 20$ mm, standardni odklon pa $\sigma = 1$ mm. Po planu tekoče kontrole od časa do časa vzemimo vzorce po $n = 9$ artiklov iz tekoče proizvodnje, pod predpostavko, da je proizvodnja pod kontrolo. V tem primeru so aritmetične sredine vzorcev z verjetnostjo $1 - \alpha = 0{,}9973$ v razmaku

$$M - 3\frac{\sigma}{\sqrt{n}} < \bar{x} < M + 3\frac{\sigma}{\sqrt{n}} \qquad (1)$$

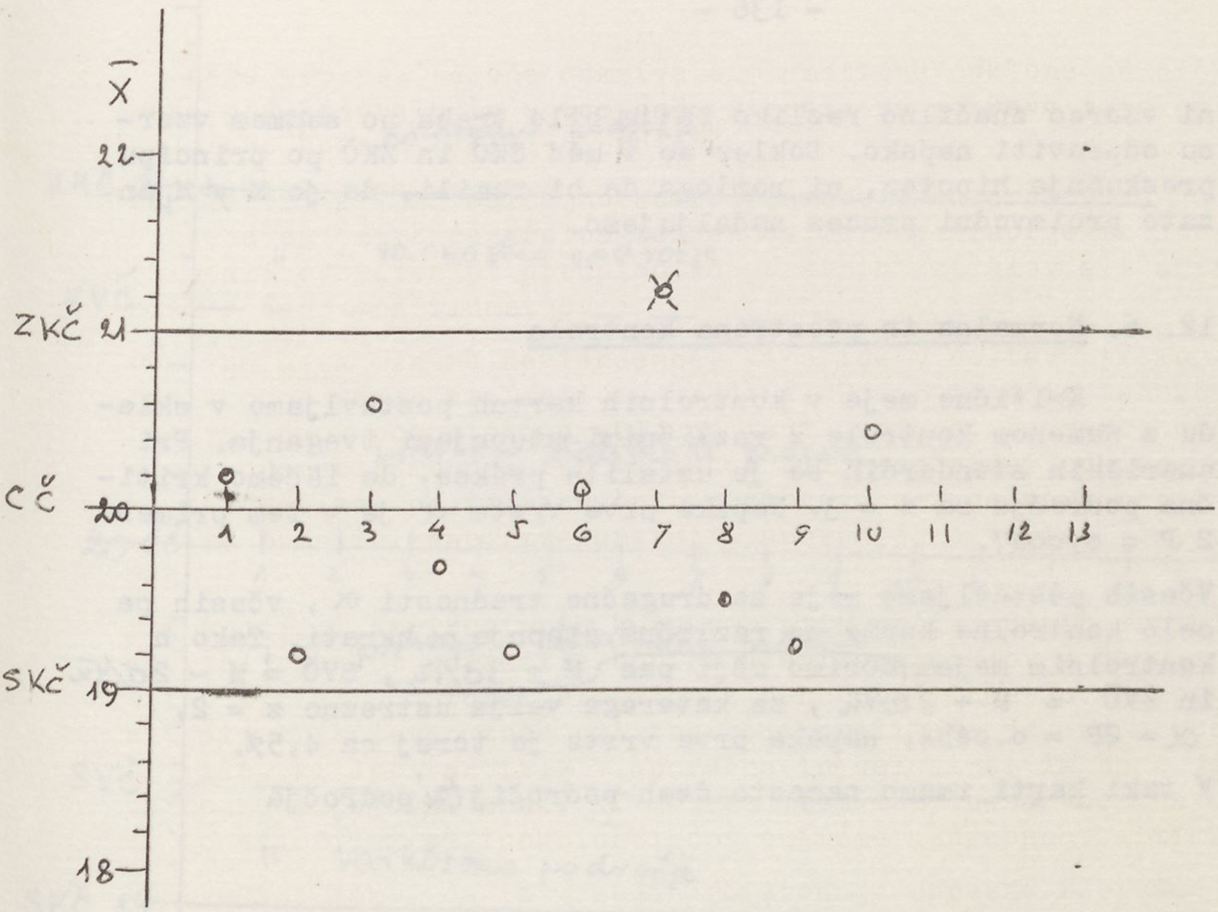
V našem primeru velja

$$20 - 3\frac{1}{\sqrt{9}} = 19 < \bar{x} < 20 + 3\frac{1}{\sqrt{9}} = 21 \qquad (2)$$

$$CČ = M = 20$$

$$SKČ = M - 3\frac{\sigma}{\sqrt{n}} = 19$$

$$ZKČ = M + 3\frac{\sigma}{\sqrt{n}} = 21$$



V kontrolno karto za aritmetično sredino (kratko $\bar{x}$ - karto), je vrisana centralna črta (CČ = M), spodnja kontrolna črta (SKČ = M - $3\sigma/\sqrt{n}$) in zgornja kontrolna črta (ZKČ = M + $3\sigma/\sqrt{n}$) Ordinatna skala je skala za aritmetične sredine vzorcev.

SKC in ZKČ sta kritični meji za preskus. Dokler prava aritmetična sredina ustreza predpisu (M = 2o in σ = 1), aritmetične sredine $\bar{x}$ iz vzorcev z verjetnostjo 1 - α = o,9973 leže v razmaku med SKČ in ZKČ. Ker je zelo majhna verjetnost (α = o,oo27), da aritmetična sredina preskusnega vzorca leži v kritičnem območju, če bi M ustrezal osnovnim pogojem, sprejmemo hipotezo, da je M različen od osnovnih pogojev in da je do razlike prišlo zaradi bistvenega vzroka. V tem primeru proizvodnjo ustavimo, da ugotovimo vzrok izjemnega odklona.

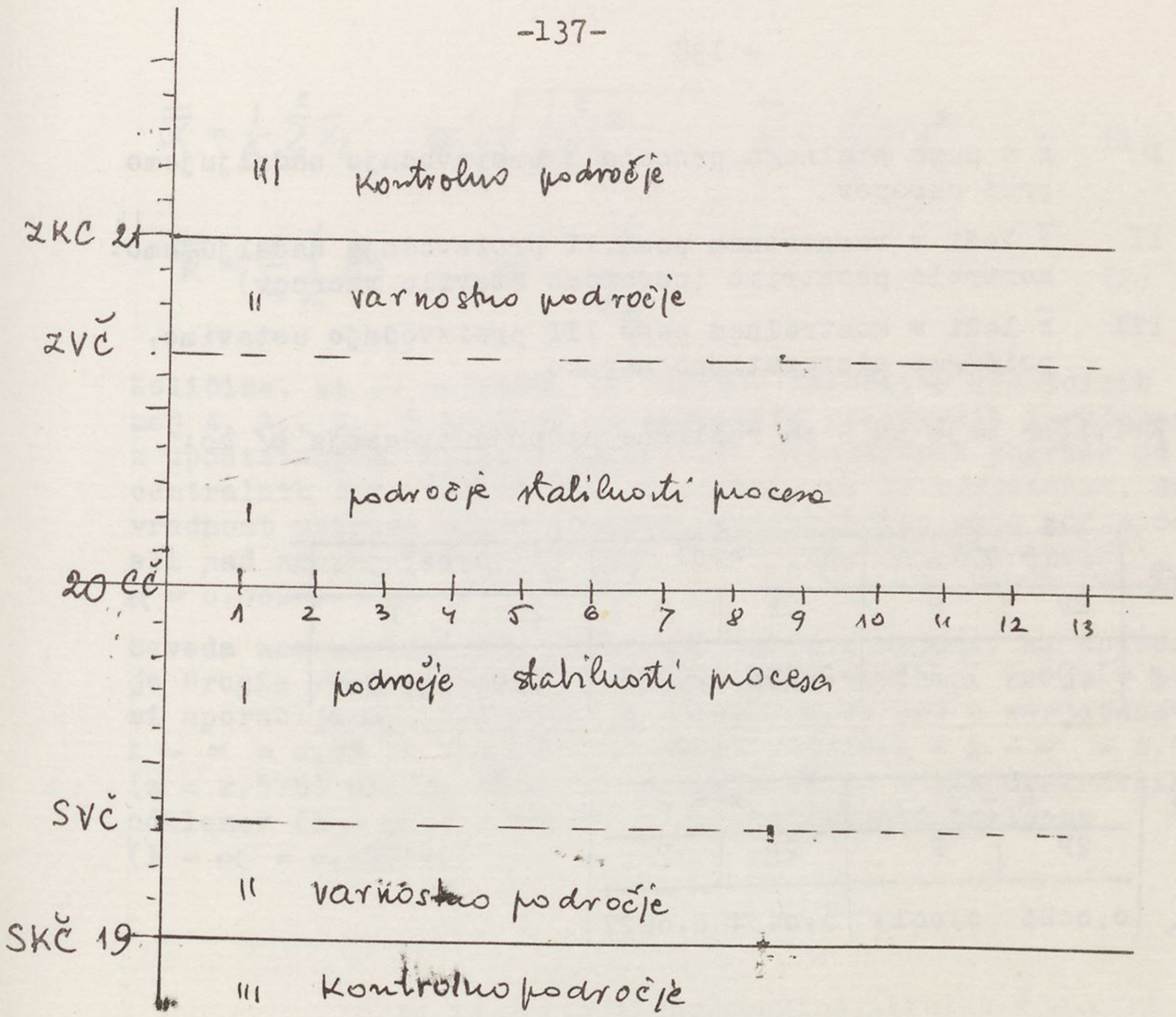
V prikazani kontrolni karti so vnešeni podatki za deset kontrolnih vzorcev. Od teh kontrolna karta kaže za sedmi kontrolni vzorec značilno razliko in je bilo treba po sedmem vzorcu odpraviti napako. Dokler so $\bar{x}$ med SKČ in ZKČ po principu preskušnja hipotez, ni razloga da bi sumili, da je $M \neq M_H$ in zato proizvodni proces nadaljujemo.

12. 6. Normalna in poostrena kontrola

Kritične meje v kontrolnih kartah postavljamo v skladu z namenom kontrole z različnimi stopnjami tveganja. Pri ameriških standardih se je ustalila praksa, da iščemo kritična področja za z = 3. Napaka prve vrste α je v tem primeru 2 P = o,oo27.

Včasih postavljamo meje za drugačne vrednosti α, včasih pa celo kontrolne karte za različne stopnje α hkrati. Tako h kontrolnim mejam (dobimo ožji pas $M \pm 3\sigma/\sqrt{n}$), SVČ = $M - 2\sigma/\sqrt{n}$ in ZVČ = $M + 2\sigma/\sqrt{n}$, za katerega velja ustrezno z = 2, $\alpha = 2P$ = o,o454, napaka prve vrste je torej ca 4,5%.

V taki karti imamo namesto dveh področij tri področja



Kontrolna karta z varnostnim pasom

Tako kontrolno karto uporabljamo in tolmačimo takole:

I $\bar{\bar{x}}$ v pasu stalnega procesa : proizvodnjo nadaljujemo brez ukrepov

II $\bar{\bar{x}}$ leži v varnostnem pasu II proizvodnjo nadaljujemo: kontrolo poostrimo (povečamo število vzorcev)

III $\bar{\bar{x}}$ leži v kontrolnem pasu III proizvodnjo ustavimo, poiščemo sistematično napako.

Kritične meje za z za različne stopnje tveganja α so:

α	0,05		0,01		0,001	
	2P	P	2P	P	2P	P
z	1,960	1,645	2,576	2,326	3,291	3,090

z	z = 3		z = 2	
	2P	P	2P	P
α	0,0026	0,0013	0,0454	0,0227

12. 7. Druge kontrolne karte za numerične znake

Razen za aritmetično sredino uporabljamo za kontrolo centralne tendence najpogosteje mediano Me, ker jo enostavno določimo. Kontrolno karto za mero centralne tendence po pravilu spremlja kontrolna karta za variabilnost. Računsko zahtevnejša je kontrola standardnega odklona σ, enostavna kontrola mere variabilnosti pa uporablja variacijski razmak R.

Parametre, potrebne za izdelavo kontrolnih kart, poznamo dostikrat vnaprej, kot standarde, karakteristike strojev itd. Tak primer smo imeli v odstavku 12. 5. Dostikrat pa potrebne parametre ocenimo iz preskusne proizvodnje, ki obsega večje število (k) zaporednih preskusnih vzorcev z n artikli iz preskusne proizvodnje. Iz teh vzorcev ocenimo parametre potrebne za izdelavo kontrolnih kart.

$$\bar{\bar{x}} = \frac{1}{k}\sum_{1}^{k} \bar{x}_l \qquad \sigma' = \sqrt{\frac{1}{k}\sum_{1}^{k} s_l^2} \qquad \bar{s}' = \frac{1}{k}\sum s_l^2 \tag{1}$$

$$\bar{R} = \frac{1}{k}\sum_{l=1}^{k} R_l \qquad \sigma' = \frac{\bar{R}}{d_2} \tag{2}$$

Količine, ki so potrebne za izračun določitve kontrolnih mej A, A_1, A_2, B in D so po ameriških standardih izračunane z upoštevanjem odklonov trikratnih standardnih pogrešk od centralnih črt. Teoretično to pomeni, da je verjetnost, da vrednost oziroma točka kontrolirane količine pade pod spodnjo ali nad zgornjo kritično mejo v kritično območje enaka α = o,oo26.

Seveda moremo izdelati kontrolne karte z mejami, ki ustrezajo drugim verjetnostnim α . Ena izmed variant, ki jo v praksi uporabljamo, upošteva dva nivoja: prvi pas z verjetnostjo $1 - \alpha$ = o,95 (z = 1,964) in drugi poostren z $1 - \alpha$ = o,99 (z = 2,576) ali za zaokrožene vrednosti z = pas dvakratnih odklonov ($1 - \alpha$ = o,9546) in pas trikratnih odklonov ($1 - \alpha$ = o,9974).

12. 8. Obrazci in tabele za izračunavanje kontrolnih linij za $\bar{x}$, me, s in R karte

Obrazci za izračunanje črt za kontrolne karte $\bar{x}$, me, s in R.

Kontrolni karti za centralno tendenco

Karta	$\bar{x}$				me
mera variabilnosti	$\sigma = \sigma'$	$\bar{s}'$	$\bar{R}$		me_R
ZKČ	$\bar{\bar{x}} + A\sigma$	$\bar{\bar{x}} + A_1\bar{s}'$	$\bar{\bar{x}} + A_2\bar{R}$	$\bar{\bar{x}} + \tilde{A}\sigma$	$\overline{me} + \tilde{A}_2 me_R$
CČ	$\bar{\bar{x}}$	$\bar{\bar{x}}$	$\bar{\bar{x}}$	$\bar{\bar{x}}$	$\overline{me}$
SKČ	$\bar{\bar{x}} - A\sigma$	$\bar{\bar{x}} - A_1\bar{s}'$	$\bar{\bar{x}} - A_1\bar{R}$	$\bar{\bar{x}} - \tilde{A}\sigma$	$\overline{me} - \tilde{A}_2 me_R$

$A = \frac{3}{\sqrt{n}}$; $A_1 = \frac{3}{C_2\sqrt{n}}$; $A_2 = \frac{3}{d_2\sqrt{n}}$; $\tilde{A} = \sqrt{\frac{\pi}{2}}\,\frac{3}{\sqrt{n}}$

Kontrolni karti za variabilnost

Karta	s		R	
mera variabilnosti	$\sigma = \sigma'$	$\bar{s}'$	$\sigma = \sigma'$	$\bar{R}$
ZKČ	$B_2\sigma$	$B_4\bar{s}''$	$D_2\sigma$	$D_4\bar{R}$
CČ	$C_2\sigma$	$\bar{s}'$	$d_2\sigma$	$\bar{R}$
SKČ	$B_1\sigma$	$B_3\bar{s}'$	$D_1\sigma$	$D_3\bar{R}$

$B_1 = C_2 - 3b_1$ $\quad B_2 = C_2 + 3b_1$ $\quad B_3 = 1 - 3\,\frac{b_1}{C_2}$ $\quad B_4 = 1 + 3\,\frac{b_1}{C_2}$

$D_1 = d_2 - 3D$ $\quad D_2 = d_2 + 3D$ $\quad D_3 = 1 - 3\,\frac{b_2}{d2}$ $\quad D_4 = 1 + 3\,\frac{b_2}{d_2}$

Tabela faktorjev za izračun kontrolnih linij pri različnem obsegu vzorcev

Velikost vzorca	Faktorji za oceno iz $\bar{s}$ in $\bar{R}$		Faktorji za kontrolne karte $\bar{x}$, me, s' in R pri danih standardih M in					
			$\bar{x}$	za me	za s'		za R	
n	c_2	d_2	A	$\tilde{A}$	B_1	B_2	D_1	D_2
2	·5642	1,128	2,121	2,658	o	1,843	o	3,686
3	·7236	1,693	1,732	1,17o	o	1,858	o	4,358
4	·7979	2,o59	1,5oo	1,88o	o	1,8o8	o	4,698
5	·84o7	2,326	1,342	1,682	o	1,756	o	4,918
6	·8686	2,534	1,225	1,535	·o26	1,711	o	5,o78
7	·8882	2,7o4	1,134	1,421	·1o5	1,672	·2o5	5,2o3
8	·9o27	2,847	1,o61	1,329	·167	1,638	·387	5,3o7
9	·9139	2,97o	1,ooo	1,253	·219	1,6o9	·546	5,394
1o	·9227	3,o78	o,949	1,189	·262	1,584	·687	5,469
12	·9359	3,258	·866	1,o85	·331	1,541	·925	5,593
15	·949o	3,472	·775	·971	·4o6	1,492	1,2o7	5,737
2o	·9619	3,735	·671	·841	·491	1,433		
25	·9696	3,931	·6oo	·725	·548	1,392		

n	Faktorji za kontrolne karte $\bar{x}$, me, s' in R pri preskusnih vrednostih $\bar{\bar{x}}$, $\overline{me}$, $\bar{s}$'in $\bar{R}$						
	$\bar{x}$		me	s'		R	
	A_1	A_2	$\hat{\tilde{A}}_2$	B_3	B_4	D_3	D_4
2	3,759	1,88o	2,232	o	3,267	o	3,268
3	2,394	1,o23	1,264	o	2,568	o	2,574
4	1,88o	·129	·828	o	2,266	o	2,282
5	1,596	·577	·112	o	2,o89	o	2,114
6	1,41o	·483	·562	·o3o	1,97o	o	2,oo4
7	1,277	·419	·519	·118	1,882	·o76	1,924
8	1,175	·373	·442	·185	1,815	·136	1,864
9	1,o94	·337	·419	·239	1,761	·184	1,816
1o	1,o28	·3o8	·368	·284	1,716	·223	1,777
12	.925	·266	·333	·354	1,646	·284	1,717
15	·817	·223	·279	·428	1,572	·348	1,652
2o	·698	·18o	·225	.51o	1,49o		
25	.619	.153	.191	.565	1,435		

12. 9. Kombinirane karte

Glede na večjo ali manjšo zamotanost izračunanja posameznih količin jemljemo v praksi tele kombinacije kontrolnih kart za raven in variabilnost: Računsko najzahtevnejša je kombinacija $\bar{\bar{x}}$-karte z s-karto, dostikrat kombiniramo $\bar{x}$-karto z R-karto, tehnično najenostavnejša pa je kombinacija me-karte z R-karto. Zadnjo kombinacijo dobimo direktno iz grafikona, v katerega vnesemo individualne podatke posameznih vzorcev.

Primer za me-R karto

Za določeno karakteristiko artiklov, ki ga proizvajamo na tekočem traku, smo se na osnovi analize primernosti in možnosti odločili, da uvedemo kontrolno karto me-R na osnovi vzorcev z n = 5 enotami. Preskusna proizvodnja 2o x 5 je dala naslednje rezultate: $\bar{\bar{x}}$ = 59,9 in $\bar{R}$ = 24.15, centralni in kontrolne črte za me-R karti, so po obrazcih:

me-karta

$$ZKČ = \bar{\bar{x}} + A_2\bar{R} = 59{,}9 + 0{,}712 .\ 24{,}15 = 77{,}1$$

$$CČ = \bar{\bar{x}} = 59{,}9 \qquad\qquad 59{,}9$$

$$SKČ = \bar{\bar{x}} - A_2\bar{R} = 59{,}9 - 0{,}712 .\ 24{,}15 = 42{,}7$$

R-karta

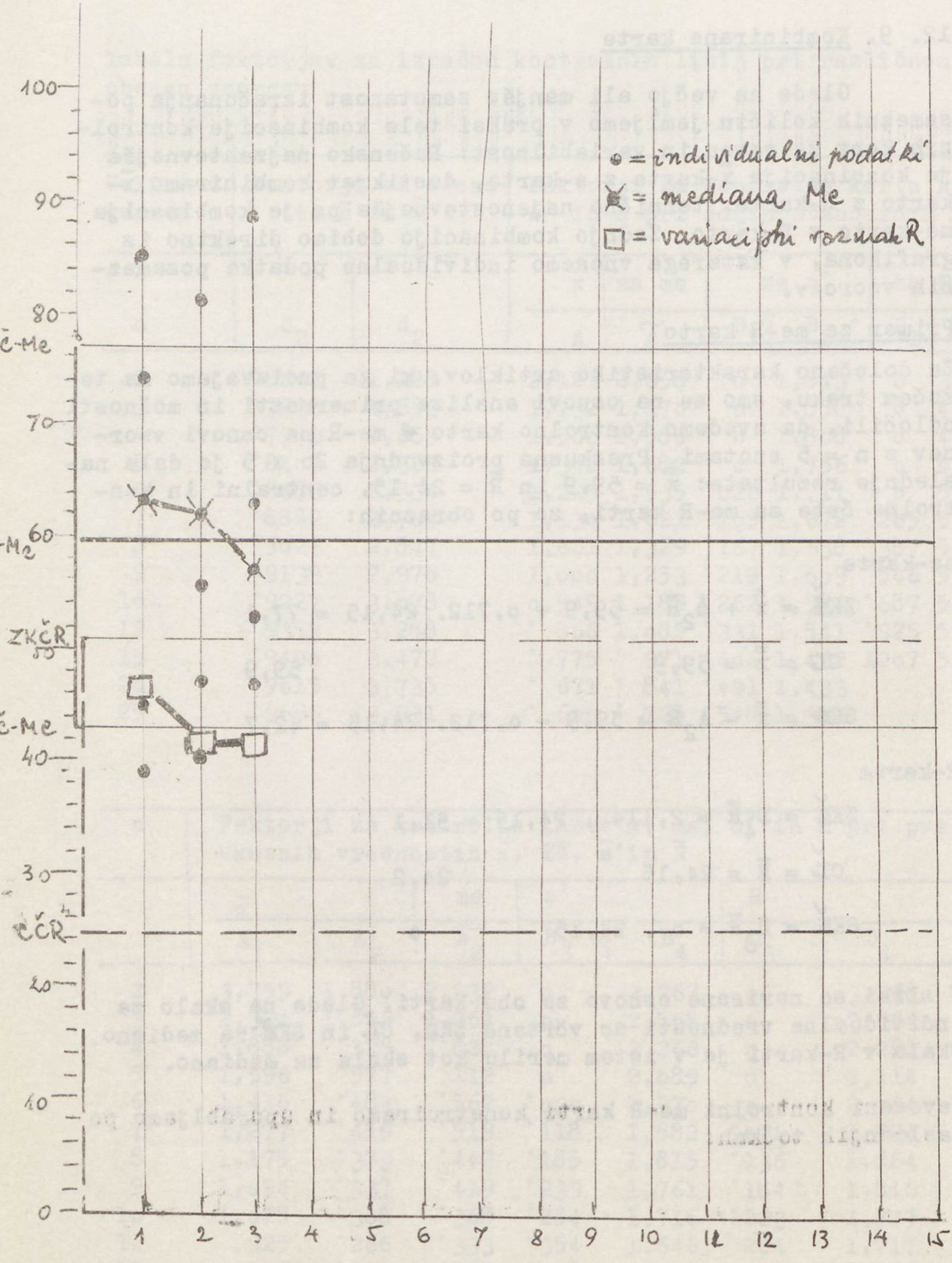
$$ZKČ = D_4\bar{R} = 2{,}114\ .\ 24{,}15 = 51{,}1$$

$$CČ = \bar{R} = 24{,}15 \qquad\qquad 24{,}2$$

$$SKČ = D_3\bar{R} = 0\ .\ 24{,}15 \qquad\qquad 0$$

V sliki so narisane osnove za obe karti. Glede na skalo za individualne vrednosti so včrtane ZKČ, CČ in SKČ za mediano. Skala v R-karti je v istem merilu kot skala za mediano.

Navedeni kontrolni me-R karti konstruiramo in uporabljamo po naslednjih točkah:



Kombinirana Me-R-karta

1. Po vrsti vnašamo v me- kontrolno karto točke za podatke kontrolnih vzorcev za pet artiklov.

2. Točka, ki je izmed petih točk v sredini, je mediana. To točko poudarimo tako, da jo prekrižamo (x).

3. Na robu pomožnega papirja napravimo črtico (osnovna črta m). Ko vskladimo osnovno črto s točko za najmanjšo vrednost, ob robu zaznamujemo mesto M, kjer leži najvišja vrednost. Razmak M - m je variacijski razmak vzorca.

4. Ta razmak vnesemo v R-karto tako, da vzorcu na ustrezno mesto na abscisni osi primaknemo osnovno črto m. pri mestu M pa ob robu papirja s krogcem, križcem ali drugim konvencionalnim znakom označimo R.

5. Če sta me in R iz vzorca v ustreznih kontrolnih pasovih, nadaljujemo s proizvodnjo, ker ni razloga, da bi sprejeli hipotezo, da se je spremenila raven ali variabilnost procesa. Proizvodni proces ustavimo v primeru, če bodisi me ali R iz kontrolnega vzorca pade izven kontrolnega pasu v kritičnega, se je zaradi bistvenega vzroka spremenil nivo (me v kritičnem območju), variabilnost (R v kritičnem območju) ali oba (če sta me in R v ustreznem kritičnem območju).

12. 10. Kontrolne karte za atributivne znake

Razen kontrolnih kart za numerične karakteristike, v praksi kontroliramo tudi atributivne znake. Ti atributi so bodisi pravi atributi ali pa izvedeni iz ene ali več numeričnih karakteristik. Tako je npr. pri kalibraciji artikel uporaben, če je premer v določenem razmaku, v nasprotnem primeru pa je neuporaben.

Ker se za majhne vzorce število enot z dano značilnostjo porazdeljuje v binomski ali Poissonovi porazdelitvi, je določanje kritičnih mej za kontrolno karto zvezano s precejšnjim računanjem, če ne razpolagamo s primernimi tabelami.

Ker se
~~Za~~ velike vzorce ~~pa se~~ d porazdeljuje aproksimativno v normalni porazdelitvi, veljajo za strukturni delež $p = \frac{d}{n}$, ~~štev~~

število defektnih artiklov d ali število redkih dogodkov c formirane kontrolne meje kot za numerične znake.

Za primer, da so dani osnovni standardi P ali α, velja za meje trikratnih standardnih odklonov

Karta	p	d	c
ZKM	$P + 3\sqrt{\frac{P(1-P)}{n}}$	$nP + 3\sqrt{nP(1-P)}$	$\alpha + 3\sqrt{\alpha}$
CL	P	nP	
SKM	$P - 3\sqrt{\frac{P(1-P)}{n}}$	$nP - 3\sqrt{nP(1-P)}$	$\alpha - 3\sqrt{\alpha}$

Če smo osnovne podatke o proizvodnem procesu ocenili s K-preskusnimi vzorci iz stabilnega proizvodnega procesa, so osnovne črte kontrolnih kart:

Karta	p	d	c
ZKČ	$\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	$\bar{d} + 3\sqrt{\bar{d}(1-\bar{p})}$	$\bar{c} + 3\sqrt{\bar{c}}$
CČ	$\bar{p}$	$\bar{d}$	$\bar{c}$
SKČ	$\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	$\bar{d} - 3\sqrt{\bar{d}(1-\bar{p})}$	$\bar{c} - 3\sqrt{\bar{c}}$

Pri tem pomeni $\bar{p} = \frac{1}{K}\sum p$ povprečen odstotek iz preskusnih vzorcev $\bar{d} = \frac{1}{2}\sum d$ povprečno število artiklov z dano značilnostjo (npr. defektnih) iz R preskusnih vzorcev stabilnega procesa; $\bar{c} = \frac{1}{2}\sum c$ poprečno število danih karakternih vzorcev iz preskusne proizvodnje. Seveda moremo kot pri vseh drugih kontrolnih kartah, ki so izdelane pod predpostavko normalnega porazdeljevanja količin faktor 3 zamenjati z drugim z, ustrezno željeni stopnji tveganja (glej tabelo v 12. 6).

12. 11. Uporaba kart p, d in c

Prvi dve od zgornjih kontrolnih kart p in d imata za osnovo zakonitosti binomske porazdelitve, tretja (c-karta) pa Poissonovo porazdelitev.

Zaradi omejitve, da zgornje kontrolne črte veljajo le za velike vzorce, dostikrat uporabljamo p-karto za ugotavljanje stabilnosti procesa iz cele proizvodnje. Celotno dnevno proizvodnjo artiklov, ki jih proizvajamo serijsko, moremo smatrati kot slučajnostni vzorec iz hipotetične populacije neomejenega števila artiklov, ki so proizvedeni pod enakimi pogoji. Če je število dnevno proizvedenih artiklov približno enako, vnesemo v zgornji obrazec za p-karto namesto n, poprečno število dnevno proizvedenih artiklov v preskusni proizvodnji $\bar{n} = \frac{1}{k}\sum n$. Če pa je število dnevno proizvedenih artiklov zelo različno, je potrebno upoštevati različne meje.

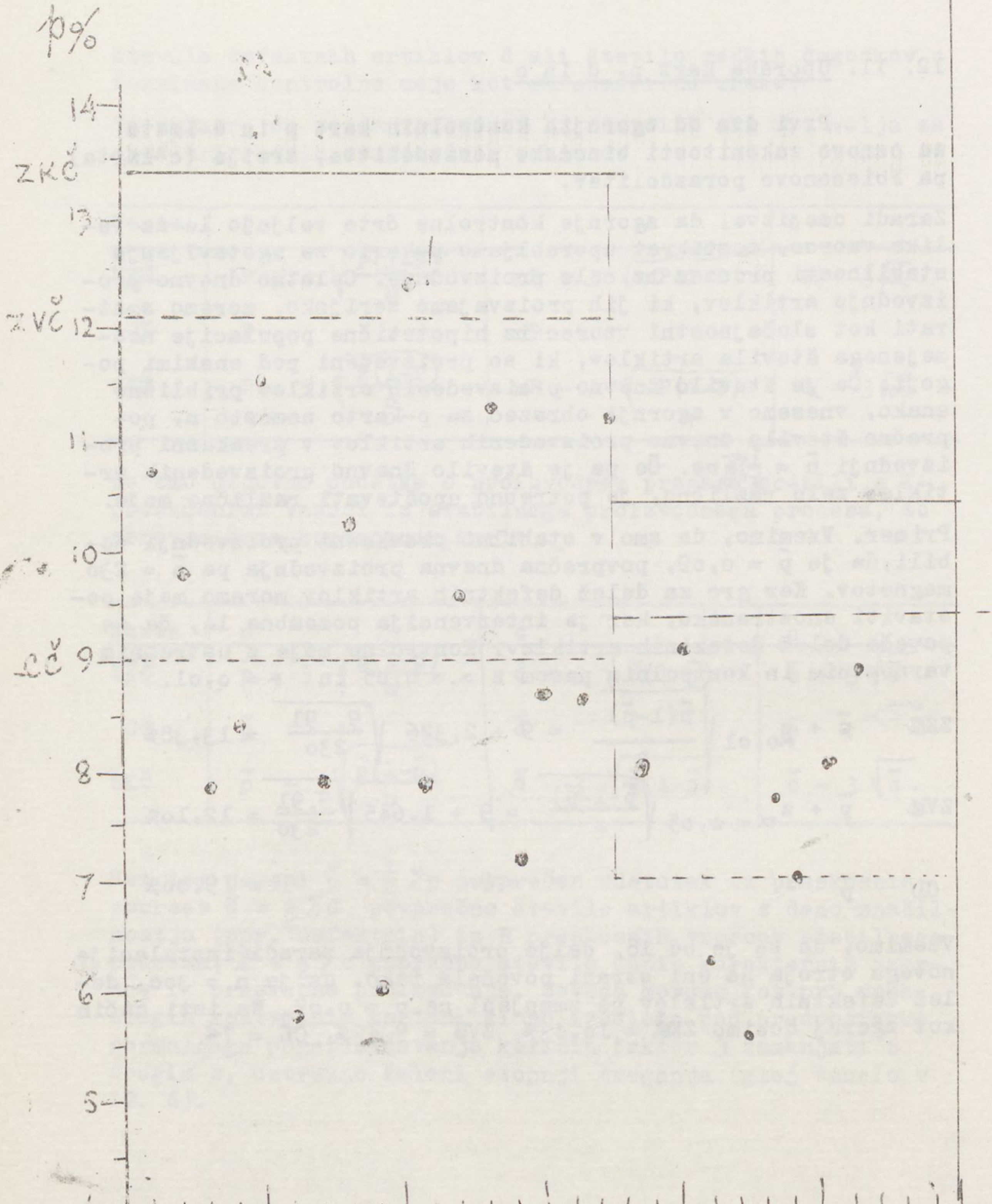
Primer. Vzemimo, da smo v stabilni preskusni proizvodnji dobili, da je $\bar{p}$ = 0,09, povprečna dnevna proizvodnja pa $\bar{n}$ = 230 magnetov. Ker gre za delež defektnih artiklov moremo meje postaviti enostransko, ker je intervencija pomembna le, če se poveča delež defektnih artiklov. Kontrolne meje z ustreznim varnostnim in kontrolnim pasom z α = 0,05 in α = 0,01.

$$\text{ZKM} \quad \bar{p} + z_{\alpha=0,01}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 9 + 2,326\sqrt{\frac{9.\ 91}{230}} = 13,38\%$$

$$\text{ZVM} \quad \bar{p} + z_{\alpha=0,05}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 9 + 1,645\sqrt{\frac{9.91}{230}} = 12,10\%$$

$$\text{CL} \quad \bar{p} \qquad = 9 \qquad = 9,00\%$$

Vzemimo, da se je od 18. dalje proizvodnja zaradi instalacije novega stroja na eni strani povečala tako, da je $\bar{n}$ = 300, delež defektnih artiklov pa zmanjšal na $\bar{p}$ = 0,07. Na isti način kot zgoraj dobimo ZKM = 10,43%, ZVM = 9,42%, CL = 7%.


p%
ZKČ
ZVČ
LČ
14
13
12
11
10
9
8
7
6
5
4
0
5
10
15
20
25
30

c-karta

c-karta za kontrolo redkih pojavov ima za osnovo Poissonovo porazdelitev. Vendar, če je osnoven parameter Poissonove porazdelitve α dovolj velik ($\alpha > 10$), moremo vzeti kot uporabno aproksimacijo, da se redki pojavi porazdeljujejo v normalni porazdelitvi z $E(x) = \alpha$ in $Var(x) = \alpha$. V tem primeru so kontrolne meje dane z:

$ZKM = \alpha + 3\sqrt{\alpha}$, $\quad CČ = \alpha \quad$ $SKM = \alpha - 3\sqrt{\alpha}$, če je dan standard in

$ZKM = \bar{c} + 3\sqrt{\bar{c}} \quad CČ = \bar{c} \quad SKM = \bar{c} - 3\sqrt{\bar{c}}$, če je parameter α ocenjen s povprečnim številom pojavov v preskusni proizvodnji.

Primerov za porabo c-karte je veliko. Običajno kontrolni vzorci niti niso vzorci v pravem smislu (na slučajnosten način izbrano določeno število artiklov iz tekoče proizvodnje), temveč more biti kot vzorec opazovanje določenega časa proizvodnje, za katerega registriramo npr. število pretrgov ali napak na stroju, včasih na slučajnosten način izbrana stran teksta in je c število napak, ali določena površina tkanine in je c število napačnih vozlov, ali kompliciran mehanizem in je c število drobnih napak, ki jih najdemo na artiklih itd.

12. 12. Statistična kontrola prevzema

Statistična kontrola prevzema je direktna uporaba statističnih metod preskušanja hipotez brez posebnih sprememb. Pri tem običajno preskušamo hipoteze o aritmetični sredini določenih karakteristik proizvodnje M, o variabilnosti teh karakteristik σ in hipoteze o strukturi proizvodnje P% po kvaliteti, uporabnosti itd.).

Te vrste kontrole vršimo v različnih fazah proizvodnje; pri prevzemanju surovin, prehodu polizdelkov iz oddelka v oddelek v istem podjetju in pri kontroli gotovega blaga.

Ker običajno kontroliramo, ali proizvodnja ustreza pogojem, ki jih stavlja kupec oziroma potrošnik, imenujemo pri kontroli kvalitete napako prve vrste, ki je v tem, da ustrezno proizvodnjo zavrnemo, tveganje proizvajalca, napako druge vrste, ki je v tem, da sprejmemo proizvodnjo, ki ne ustreza določenim pogojem pa tveganje kupca.

Kontrolo prevzema uporabljamo pri kontroli množične proizvodnje v tejle obliki. Celotno proizvodnjo razdelimo glede na pogoje proizvodnje (dnevna proizvodnja; proizvodnja proizvedena na posameznih strojih itd.) v skupine artiklov. Po vnaprej določenem planu kontrole iz posamezne skupine izberemo slučajnostno določeno število artiklov. Glede na kvaliteto teh artiklov posamezno skupino sprejmemo ali zavrnemo. Zavrnjene skupine prekontroliramo v celoti in v njih defektne artikle nadomestimo z dobrimi. Tako stvaren odstotek izmeta zmanjšamo, ker so defektni artikli v zavrnjeni skupini nadomeščeni z dobrimi. Posamezni plani statistične kontrole prevzema imajo za osnovo različne količine. Ena izmed najpogostejših je povprečen odstotek izmeta na opravljeni kontroli. Ta pa je odvisna od kvalitete proizvodov pred kontrolo. Če je kvaliteta pred kontrolo dobra, je odstotek izmeta po kontroli tudi nizek. Enako je odstotek izmeta nizek, če je kvaliteta proizvodnje pred kontrolo zelo slaba, ker v tem primeru zavrnemo razmeroma veliko število skupin, in v njih nadomestimo slabe artikle z dobrimi. Pri neki določeni kvaliteti proizvodnje pred kontrolo pa je odstotek izmeta po kontroli maksimalen. Pri vzorčnih planih je zato razen povprečnega odstotka izmeta po kontroli pomemben tudi maksimalen povprečen odstotek izmeta, ki ga dobimo pri kontroli proizvodnje, v kateri odstotek izmeta srednje velik.

12. 13. Enojni, dvojni, trojni, sekvencialni plan

Pri dani velikosti kontroliranih skupin in povprečnem odstotku izmeta po opravljeni kontroli moremo najti, pri zadostni velikosti vzorca neko celo število C_1 tako, da skupino sprejmemo, če je število defektnih artiklov v vzorcu manjše ali enako C_1 in zavrnemo, če je število slabih artiklov v vzorcu večje kot Cl. Tak plan imenujemo enojni vzorčni plan.

Pri velikih razlikah med stvarno in hipotetično kvaliteto odkrijemo razlike tudi z manjšimi vzorci. Zato se včasih obnese dvojni plan. Po tem planu v prvi fazi izberemo vzorec n_1 enot. Če je število slabih artiklov v tem vzorcu manjše ali enako C_1 skupino sprejmemo. Če je število slabih artiklov večje ali enako C_2 skupino zavrnemo. Če pa je število slabih artiklov večje kot C_1 in manjše kot C_2, izberemo dopolnilni vzorec z n_2 enotami. Ta vzorec je osnova za zavrnitev ali sprejem kot pri enostavnem planu. Če število slabih artiklov v skupnem vzorcu z $n_1 + n_2$ enotami enako ali manjše kot C_1 skupino sprejmemo, v nasprotnem primeru pa jo zavrnemo. Če pričakujemo da so med skupinami velike razlike, je ta plan bolj ekonomičen kot enojni, ker slabe skupine navadno izločimo že pri prvem vzorcu.

Podobno moremo sestaviti trojni plan, pri katerem pridemo pri nekaterih vzorcih do dokončne odločitve šele pri tretjem vzorcu. Pri tem planu vzorec n_1 enot povečamo na $n_1 + n_2$ enot, če ne dosežemo odločitve pri prvem vzorcu, na $n_1 + n_2 + n_3$ enot pa v primeru, če tudi z drugim vzorcem ne moremo skupine niti zavreči niti sprejeti.

Razširitev te ideje je sekvencialni plan, pri katerem večamo preskusni vzorec za en ali za majhno število artiklov. Pri sekvencialnem planu dodajamo nove preskusne artikle toliko časa, da pridemo do odločitve, da skupino sprejmemo ali zavrnemo. Dokler pa je število slabih artiklov tolikšno, da je situacija nedoločena, nadaljujemo z večanjem vzorca. Po določenem številu postopkov pridemo do cilja, in sicer tem prej, čim večje so razlike med stvarnim in hipotetičnim odstotkom. Sekvencialni plan je bolj zamotan kot navaden ali dvojen, ima pa to dobro lastnost, da dovede do rezultata v povprečju z manjšim številom preskusov. Zato ga uporabljamo predvsem za kontrolo proizvodov, za katere je preskus predrag.

12. 14. Kot primer vzemimo tabelo iz knjige: Freeman, Friedman, Mosteller, Wallis: Sampling Inspection. Iz priročnika je vzeta tabela za kontrolo skupin s 500-800 enotami za primer, da pričakujemo povprečni odstotek defektnih artiklov po kontroli od 1,2% do 2,2%, zgornjo limito povprečja defektnih artiklov po kontroli pa 2,5% do 3,5%.

Tabela. Plan kontrole skupin s 5oo do 8oo enotami s povprečnim odstotkom defektnih artiklov po kontroli 1,2% do 2,2% in zgornjo limito povprečja defektnih artiklov 2,5 do 3,5%

Tip vzorca	Vzorec	Velikost vzorca		C_1	C_2
		posameznega	kumulativnega		
Enojni	prvi	4o	4o	2	3
Dvojni	prvi	25	25	1	4
	drugi	5o	75	3	4
Sekvencialni	prvi	lo	lo	+	2
	drugi	lo	2o	o	3
	tretji	lo	3o	1	3
	četrti	lo	4o	1	4
	peti	lo	5o	2	4
	šesti	lo	6o	2	4
	sedmi	lo	7o	4	5

Skupino sprejmemo, če je število defektnih artiklov v vzorcu $x \leq C_1$ in zavrnemo, če je $x \geq C_2$. Vzorec nadaljujemo, če je $C_1 < x < C_2$.

\+ pomeni: skupine ne moremo sprejeti pri prvem vzorcu.

Če npr. iz skupine s 3oo enotami po dvojnem planu izberemo n_1 = 25 artiklov in med njimi najdemo 3 defektne, moramo vzorec glede na prejšnjo tabelo povečati za n2 = 5o, ker je 1 3 4! Če najdemo v skupnem vzorcu $n_1 + n_2$ = 75 enot 7 defektnih artiklov, skupino glede na prejšnjo tabelo zavrnemo.

Podobno je pri sekvencionalnem planu. Če iz skupine s N = 3oo enotami izberemo najprej lo artiklov in izmed teh ne najdemo nobenega defektnega, moramo vzeti nadaljnjih lo enot. Če v skupnem vzorcu z 2o enotami dobimo dva defektna artikla, moramo postopek ponoviti in vzorec povečati še za lo enot, ker je $0 < 2 < 3$. Če v skupnem vzorcu s 3o enotami dobimo 3 defektne artikle pa skupino po tretjem vzorcu zavrnemo, ker je $3 \geq 3$.

Podobne tabele, kot je tabela v 12. 14 so sestavljene za različne kombinacije povprečnega odstotka defektnih artiklov po kontroli in za različne velikosti skupin, tako da moremo najti tabelo za vsako kombinacijo, ki je praktično potrebna.

13. PROUČEVANJE KORELACIJSKIH ODVISNOSTI

13. 1. Funkcijske in korelacijske odvisnosti

Pri funkcijskih odvisnostih med x in y vsaki vrednosti x ustreza ena ali nekaj točno določenih vrednosti odvisne spremenljivke y. Pri proučevanju množičnih pojavov tudi zasledimo odvisnosti, ki pa so po svoji naravi različne od funkcijskih odvisnosti. Žilavost jekla je npr. odvisna od vsebnosti ogljika, časovna proizvodnost je odvisna od vlagalnega časa ipd. Vendar opazimo, da te odvisnosti niso funkcijske in more biti žilavost posameznih plošč pri isti vsebnosti ogljika različna. Enako je tudi s časovno produktivnostjo v odvisnosti od vlagalnega časa in z vsemi podobnimi primeri. Za take slučaje sicer opazimo zakonitost odvisnosti določene karakteristike od določenega faktorja, vendar to zakonitost opazimo le pri množičnem opazovanju, ker velja na splošno, ne pa v individualnih primerih. Vzrok za to je stalno prisoten vpliv individualnih faktorjev, ki jih dostikrat združimo v skupen efekt vpliva slučajnostnih faktorjev. S tem, da dodatne faktorje, ki bi motili odvisnost med proučevanima znakoma držimo na istem nivoju, sicer odstranimo vpliv nekaj faktorjev, ki so združeni v splošne pogoje poskusa. Vendar ni pri proučevanju nikdar mogoče docela odstraniti vpliv vseh faktorjev. Zato moremo za razliko od funkcijske odvisnosti, ki jo zapišemo simbolično

$$y = f(x) \qquad (1)$$

pisati korelacijsko odvisnost med x in y simbolično kot

$$y = f(x;e) \qquad (2)$$

pri čemer pomeni e individualne ali v posebnem slučajnostne vplive.

13. 2. Prikazovanje korelacijskih odvisnosti

Funkcijske odvisnosti podajamo na tri načine:

a) s pravilom $y = f(x)$, ki nakazuje, kako dobimo iz vrednosti neodvisne spremenljivke x vrednost odvisne spremenljivke y

b) z nizom dvojic ustreznih vrednosti x in y, iz katerega neposredno dobimo vrednost y, ki ustreza dani vrednosti x

c) s krivuljo v koordinatnem sistemu, v katerem vsaka točka krivulje predstavlja ustrezno dvojico vrednosti x in y.

Korelacijske odvisnosti prikazujemo podobno, čeprav je njihova narava drugačna kot narava funkcijske odvisnosti. Tako osnovne podatke prikažemo z nizom dvojic ustreznih podatkov x in y, ki se nanašata na iste enote.

Pri korelacijski odvisnosti sta namreč vrednost x in y združeni v dvojici vrednosti preko enote, na katero se nanašata.

Tako imamo npr. za povezanost med vlagalnim in neto časom za 50 šarž na peči V naslednje podatke: x = vlagalni čas, y = neto čas.

x	y	x	y	x	y	x	y
3,00	8,00	2,75	8,00	2,00	7,25	2,00	7,00
3,00	7,50	2,50	7,50	2,75	8,75	2,00	7,50
2,25	6,67	3,50	9,75	2,50	8,25	2,50	7,50
2,25	7,50	1,75	6,25	2,50	8,25	2,00	7,00
3,00	7,75	3,17	8,83	2,50	8,00	2,25	7,50
2,25	6,75	2,00	8,00	2,50	7,25	3,25	8,75
3,00	7,50	2,75	7,50	2,50	7,50	2,00	7,75
2,75	6,67	3,42	8,50	2,25	8,75	2,50	7,75
2,25	7,50	3,00	9,25	2,50	7,50	2,25	6,25
2,00	7,00	2,50	8,50	3,50	9,75	3,75	8,75
2,75	9,00	2,00	8,00	2,75	8,50	3,25	7,75
2,50	7,75	3,25	8,75	2,75	9,50		
2,00	8,00	3,50	8,25	2,25	7,00		

13.3. Korelacijska tabela

Če je število enot veliko, pri proučevanju enega samega znaka sestavimo frekvenčno porazdelitev. Ta sicer samo približno, vendar nazorno pokaže ponašanje vrednosti. Podobno je tudi pri korelaciji. Množica dvojic vrednosti je nepregledna, če je število enot veliko. Pri večjem številu enot je nepregleden tudi korelacijski grafikon. Kot pri proučevanju enega samega znaka, tudi pri proučevanju dveh znakov korelacijsko odvisnost nazorneje prikažemo, če namesto individualnih vrednosti za x in y vpeljemo razrede. Če preštejemo, koliko enot ima vrednost v posamezni kombinaciji razredov za znak x in y in podatke podamo v kombinacijski tabeli, dobimo korelacijsko tabelo. Korelacijska tabela ima vse prednosti in pomanjkljivosti frekvenčnih porazdelitev enega samega znaka. S širokimi razredi zabrišemo zakonitosti, ozki razredi pa povečajo obseg tabele, ki je zato nepregledna, razen tega pa se zaradi večjega vpliva individualnih faktorjev zabrišejo zakonitosti. Zato je izbira optimalne širine razredov za analizo zelo pomembna.

Da je možno v nadaljnji računski analizi uporabiti izdelane metode analize je priporočljivo, da so razredi za posamezen znak enako široki.

Za primer podajamo v korelacijski tabeli odnos med λ_{1o} in λ_{4o} za jeseniško patentirano žico.

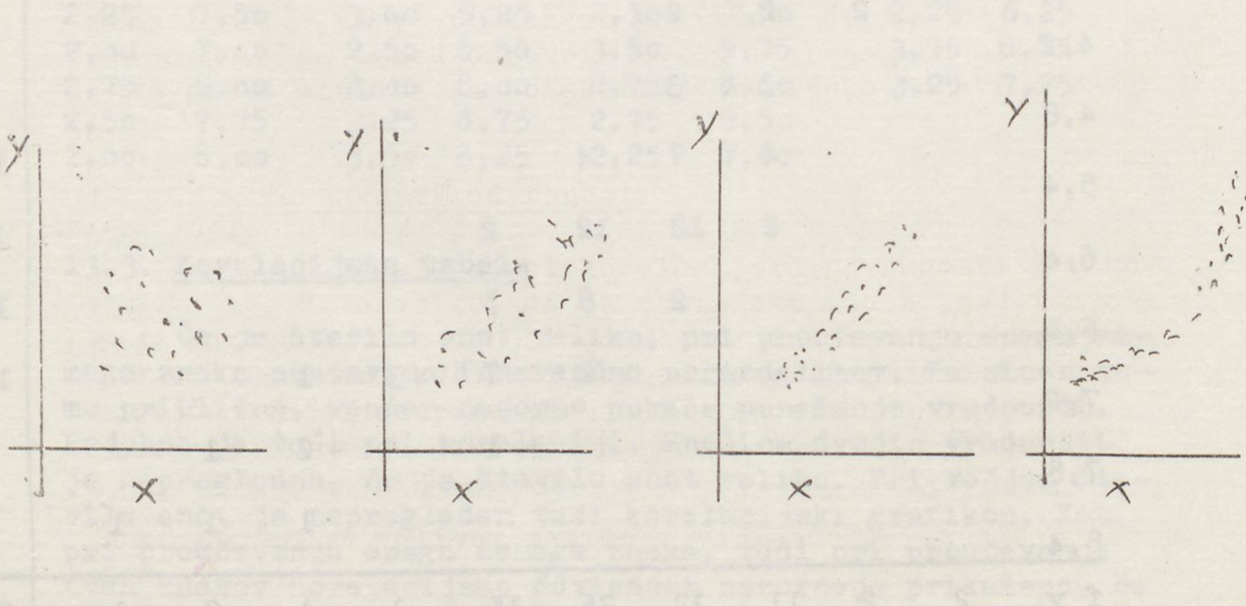
Odnos med λ_{1o} in λ_{4o} za patentirano jeseniško žico za n = 98 preskušancev

y \ x	1,o	1,4	1,8	2,2	2,6	3,o	3,4	3,8	4,2	4,6	5,o	f_x
2,4												
		1										1
3,o												
		1										1
3,6												
			2	2	2							6
4,2												
				3	3							6
4,8												
				4	7	4						15
5,4												
				2	18	12	2					34
6,o												
					2	8	5					15
6,6												
						2	7	3	1			13
7,2												
							1		2	1		4
7,8												
									1	1	1	3
8,4												
f y		2	2	11	32	26	15	3	4	2	1	98

Iz zgornje korelacijske tabele je jasno vidna pozitivna in linearna odvisnost med λ_{10} in λ_{40} za jeseniško patentirano žico. Meglico točk iz korelacijskega grafikona v tem primeru nadomesti skupnost frekvenc, ki se pomikajo v smeri pozitivne korelacije.

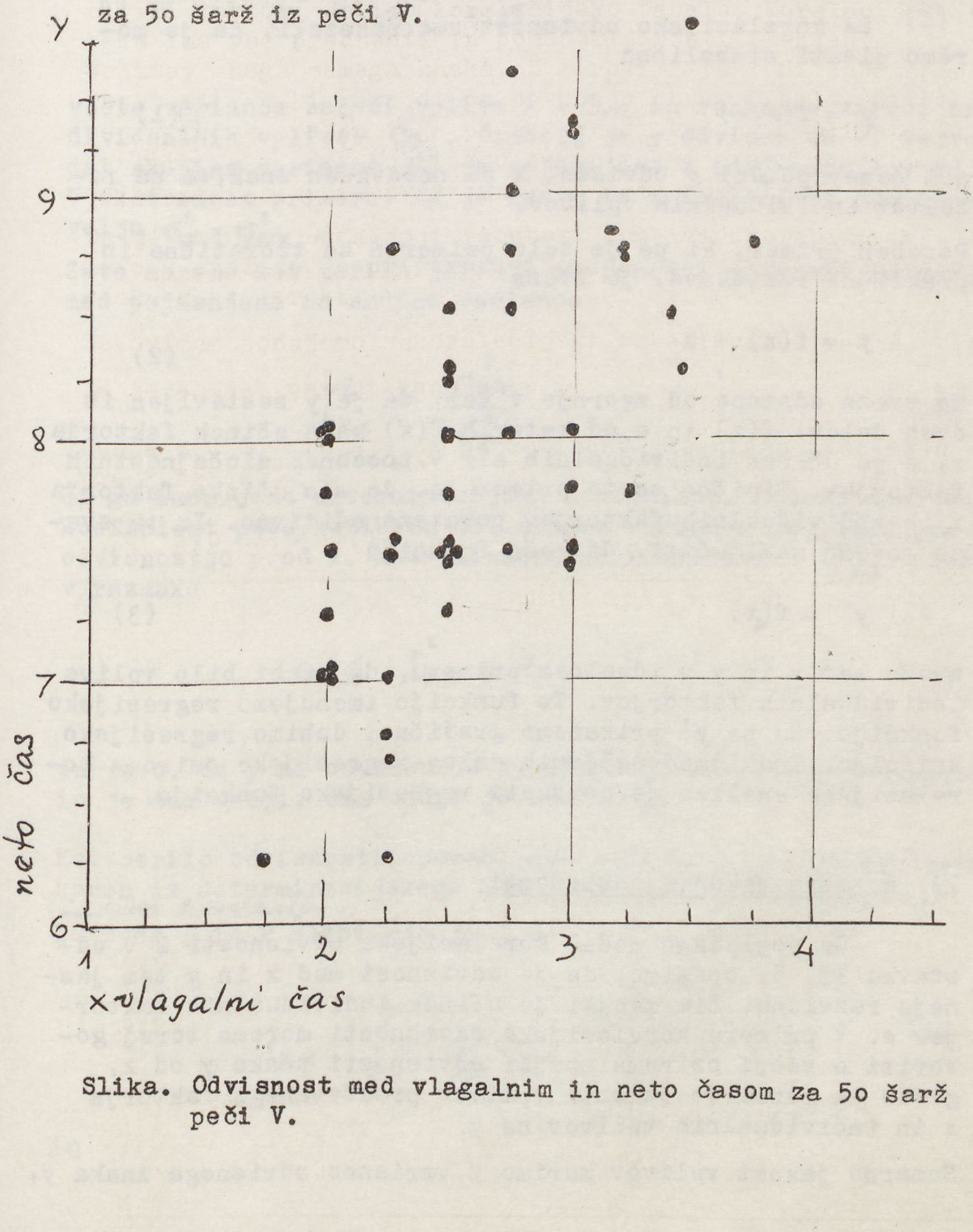
13. 4. Korelacijski grafikon

Iz množice parov podatkov, ki so osnova za korelacijsko analizo, ne moremo neposredno zaznati ali in kolika je korelacijska odvisnost in kakšen je potek regresijske krivulje. Boljši pregled nad to množico podatkov dobimo s korelacijskim grafikonom, v katerem je vsak par vrednosti (x, y) prikazan s točko v pravokotnem koordinatnem sistemu. Množica točk, od katerih vsaka predstavlja podatka za eno enoto, en poskus ali podobno, že določnejše nakazuje zakonitosti odvisnosti od x. V shematičnem prikazu iz slike a sklepamo, da med x in y ni odvisnosti, v sliki b je nakazana med x in y šibka linearna pozitivna odvisnost, v sliki c je razvidna tesnejša linearna odvisnost med x in y, slika d pa nakazuje krivuljčno korelacijsko odvisnost.



Slika. Korelacijski grafikoni

V naslednji sliki je prikazan korelacijski grafikon za korelacijsko odvisnost med vlagalnim časom in neto časom za 5o šarž iz peči V.



Slika. Odvisnost med vlagalnim in neto časom za 5o šarž peči V.

13. 5. Regresijska krivulja

Za korelacijsko odvisnost smo nakazali, da jo moremo pisati simbolično

$$y = f(x,e) \qquad (1)$$

pri čemer pomeni y odvisen, x pa neodvisen znak, e pa rezultat individualnih vplivov.

Poseben primer, ki pa je zelo primeren za teoretične in praktične raziskave, je zveza

$$y = f(x) + e \qquad (2)$$

Ta zveza odstopa od zgornje v tem, da je y sestavljen iz dveh delov: f(x) in e, od katerih f(x) kaže učinek faktorja x, e pa učinek individualnih ali v posebnem slučajnostnih faktorjev. Tipično za ta primer je, da sta učinka faktorja x in individualnih faktorjev povezana aditivno. Iz te zveze moremo zaključiti, da kaže funkcija

$$y' = f(x) \qquad (3)$$

zvezo med x in y v idealnem primeru, da ne bi bilo vpliva individualnih faktorjev. To funkcijo imenujemo regresijsko funkcijo. Če pa y' prikažemo grafično, dobimo regresijsko krivuljo. Ena izmed osnovnih nalog regresijske oziroma korelacijske analize je poiskati regresijsko funkcijo.

13. 6. Mere jakosti odvisnosti

Če pogledamo model korelacijske odvisnosti 2 v odstavku 13, 5, opazimo, da je odvisnost med x in y tem jasneje razvidna, čim manjši je učinek individualnih faktorjev e. V primeru korelacijske odvisnosti moremo torej govoriti o večji oziroma manjši odvisnosti znaka y od x, glede na razmerje jakosti vplivov proučevanega faktorja x in individualnih vplivov na y.

Sumarno jakost vplivov merimo z varianco odvisnega znaka y, ki jo moremo smatrati kot

$$\sigma_y^2 = \sigma_{y/x}^2 + \sigma_e^2 \tag{1}$$

vsoto variance zaradi vpliva x - $\sigma_{y/x}^2$ in variance zaradi individualnih vplivov σ_e^2. Čimbolj je y odvisen od x, temvečji del skupne variance σ_y^2 je pojasnjene z odvisnostjo y od x. V ekstremnem primeru, da je med y in x funkcijska odvisnost, velja $\sigma_y^2 = \sigma_{y/x}^2$.

Zato moremo kot merilo jakosti odvisnosti smatrati razmerje med pojasnjeno in skupno varianco

$$I_{y.x}^2 = \frac{\sigma_{y/x}^2}{\sigma_y^2} \tag{2}$$

ki ga imenujemo determinacijski koeficient. Determinacijski koeficient pove, kolikšen del skupne variance je pojasnjen z odvisnostjo y od x. Determinacijski koeficient $I_{y.x}^2$ leži v razmaku

$$0 \leq I_{y.x}^2 \leq 1 \tag{3}$$

in je o, če y ni odvisen od x, in 1 če je odvisnost funkcijska in je tem večji, čim večja je odvisnost.

Kot merilo odvisnosti uporabljamo tudi $I_{y.x}$, ki je kvadratni koren iz determinacijskega koeficienta in ga imenujemo indeks korelacije.
Iz zvez 1 in 2 sledi, da je

$$I_{y.x} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}} \tag{4}$$

in

$$\sigma_e = \sigma_y \sqrt{1 - r_{y,x}^2} \qquad (5)$$

σ_e, standardni odklon zaradi individualnih vplivov imenujemo standardno pogreško ocene, ker podaja kvaliteto napovedi y iz regresijske krivulje y'.

13. 7. Linearna regresija

Kadar je zveza med y in x dana z modelom

$$y = \alpha + \beta x + e \qquad (1)$$

govorimo o linearni odvisnosti med x in y, ker je regresijska funkcija

$$y' = \alpha + \beta x \qquad (2)$$

dana z linearno zvezo med x in y.

Za dan sistem dvojic vrednost x_i y_i i = 1, 2 ......N, običajno vzamemo kot regresijsko premico tisto, za katero je vsota kvadratov odklonov stvarnih vrednosti y od vrednosti na regresijski premici najmanjša

$$\sum_{i=1}^{N}(y-y')^2 = \sum(y-a-bx)^2 = \min \qquad (3)$$

Zgornji izraz je funkcija parametrov a in b in je najmanjši, če sta parcialna odvoda po a in b enaka 0. Iz tega pogoja dobimo po metodi najmanjših kvadratov sisteme normalnih enačb

$$\sum y = Na + b\sum x$$

$$\sum yx = a\sum x + b\sum x^2 \qquad (4)$$

iz njih pa regresijsko premico v obliki

$$y' = \bar{y} + b_1(x - \bar{x}) \tag{5}$$

pri čemer je b_1

$$b_1 = \frac{c_{xy}}{\sigma_x^2} \tag{6}$$

kvocient med kovarianco

$$c_{xy} = \frac{1}{N}\sum(x-\bar{x})(y-\bar{y}) \tag{7}$$

in varianco za x

$$\sigma_x^2 = \frac{1}{N}\sum(x-\bar{x})^2 \tag{8}$$

Linearno regresijo obračunamo po standardni shemi:

$$\begin{array}{ccccc}
\sum x & \sum y & \sum x^2 & \sum xy & \sum y^2 \\
\bar{x} = \frac{1}{N}\sum x & \bar{y} = \frac{1}{N}\sum y & -\frac{(\sum x)^2}{N} & -\frac{\sum x \sum y}{N} & -\frac{(\sum y)^2}{N} \\
 & & \overline{K_x} & \overline{K_{xy}} & \overline{K_y} \\
 & & \sigma_x^2 = \frac{K_x}{N} & c_{xy} = \frac{K_{xy}}{N} & \sigma_y^2 = \frac{K_y}{N}
\end{array}$$

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

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

Za primer odvisnosti med vlagalnim časom x in neto časom y dobimo po zgornji shemi:

$x = 130{,}09 \qquad y = 394{,}17 \qquad \bar{x} = \frac{1}{N}\sum x = 2{,}602$

$\bar{y} = \frac{1}{n}\sum y = 7{,}883$

$x^2 =$	350,4953	$xy =$	1038,3486	$y^2 =$	3141,6967
$-\frac{(\sum x)^2}{N}$ =	-338,4682	$-\frac{\sum x \sum y}{N}$ =	-1025,5515	$-\frac{(\sum y)^2}{N}$ =	-3107,3997
K_x =	12,0271	K_{xy} =	12,7971	K_y =	34,2970
σ_x =	0,240542	C_{xy} =	0,255942	σ_y =	685940

$$b_1 = \frac{C_{xy}}{\sigma_x^2} = 1{,}064$$

$$y^1 = \bar{y} + b_1(x - \bar{x}) = 7{,}883 + 1{,}064(x - 2{,}602) = 5{,}114 + 1{,}064\ x$$

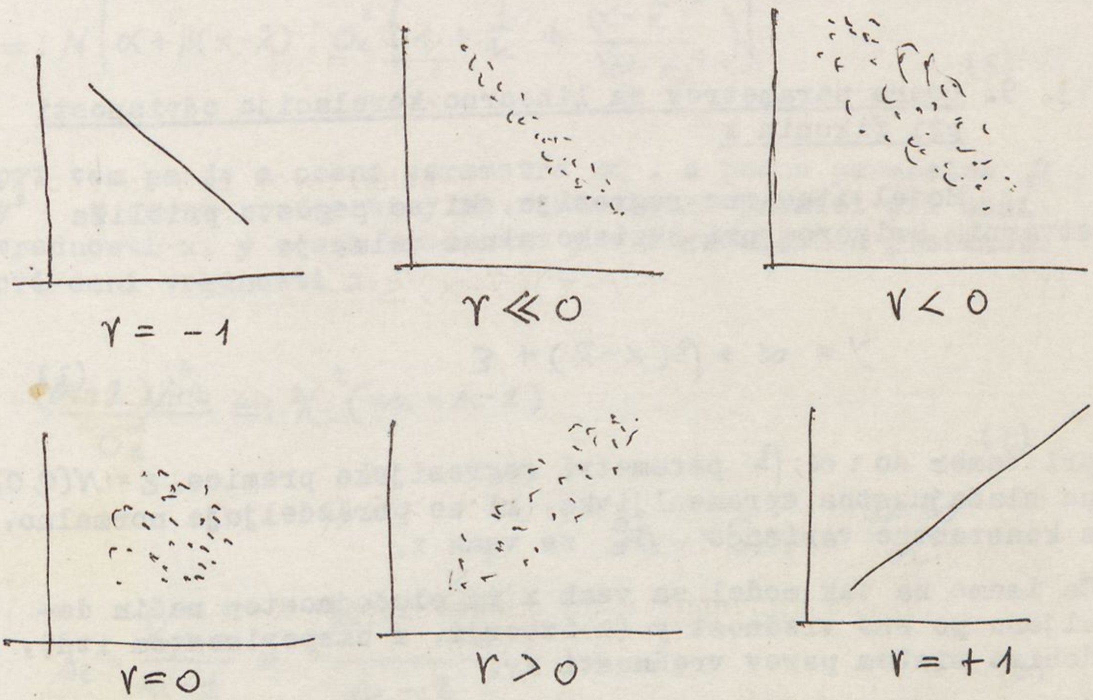
13. 8. Korelacijski koeficient

Korelacijski koeficient služi za merilo stopnje linearne odvisnosti med numeričnima znakoma x in y.

Korelacijski koeficient, ki je po definiciji kvadratni koren iz determinacijskega koeficienta za linearno odvisnost r^2, izračunamo po obrazcu

$$r_{xy} = \frac{c_{xy}}{\sigma_x\ \sigma_y}$$

Ker se vrednost korelacijskega koeficienta ravna predvsem po velikosti kovariance c_{xy}, more biti vrednost korelacijskega koeficienta pozitivna ali negativna. Pozitivni korelacijski koeficient nakazuje pozitivno, negativni pa negativno odvisnost med proučevanima pojavoma. Vrednosti korelacijskega koeficienta leže med ekstremoma r = -1 do r = +1. Korelacijski koeficient je -1 pri negativni funkcijski odvisnosti, +1 pa pri pozitivni funkcijski odvisnosti. Korelacijski koeficient je enak o, če odvisnosti ni, večjo odvisnost pa pokaže večja absolutna vrednost korelacijskega koeficienta.



Slika. r za različne korelacijske grafikone

Kvadrat iz korelacijskega koeficienta r^2_{xy} je determinacijski koeficient linearne odvisnosti, in pokaže, koliki del variabilnosti od y je pojasnjene z linearno odvisnostjo y od x.

Za naš primer (y) je korelacijski koeficient

$$r_{xy} = \frac{c_{xy}}{\sigma_x \cdot \sigma_y} = \frac{2559{,}42}{49{,}05 \cdot 82{,}82} = 0{,}630$$

in determinacijski koeficient

$$r_{xy}^2 = 0,3970$$

Z linearno odvisnostjo je pojasnjena 0,397 ali cca 40% od skupne variabilnosti.

13. 9. Ocena parametrov za linearno korelacijo odvisnosti pri fiksnih x

Model linearne regresije, ki se pogosto približa stvarnim primerom pri raziskovalnem delu, je

$$y = \alpha + \beta(x - \bar{x}) + \varepsilon \tag{1}$$

pri čemer so : α, β parametri regresijske premice, $\varepsilon =: N(0, \sigma_e^2)$ pa slučajnostna spremenljivka, ki se porazdeljuje normalno, s konstantno varianco σ_e^2 za vsak x.

Če imamo za tak model za vsak x na slučajnosten način dobljeno po eno vrednost y (ž izborom, z eksperimentom itd), dobimo sistem parov vrednosti xy.

Iz teh podatkov ocenjeni regresijski model ima obliko

$$y = a + b(x - \bar{x}) + e \tag{2}$$

Količine iz tega modela se porazdeljujejo:

$$a = \bar{y} =: N(\alpha; \frac{\sigma^2}{n}) \tag{3}$$

$$b = \frac{Sy(x-\bar{x})}{S(x-\bar{x})^2} =: N\left(\beta; \frac{\sigma_e^2}{S(x-\bar{x})^2}\right) \qquad (4)$$

$$y'(x) = a + b(x-\bar{x}) =: N\left(\alpha + \beta(x-\bar{x}); \sigma_e^2\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{S(x-\bar{x})^2}\right]\right) \qquad (5)$$

$$y =: N\left[\alpha + \beta(x-\bar{x}); \sigma_e^2\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S(x-\bar{x})^2}\right)\right] \qquad (6)$$

pri tem pa je a ocena parametra α , b ocena parametra β , y' = ocena vrednosti y na regresijski premici pri dani vrednosti x, y ocena vrednosti y v regresijskem grafikonu pri dani vrednosti x.

$$\frac{(n-2)s_e^2}{\sigma_e^2} =: \chi^2(m = n-2) \qquad (7)$$

$$K_{xy} = Sxy - \frac{Sx\,Sy}{n}$$

$$s_e = \frac{K_{y.x}}{n-2} = \frac{K_y - \frac{K_{xy}^2}{K_x}}{n-2} \qquad (8)$$

Po znanih odnosih in zvezah t, z in χ^2 porazdelitev sledi dalje

$$\frac{b-\beta}{s_e}\sqrt{S(x-\bar{x})^2} =: t(m = n-2) \qquad (9)$$

in za dva vzorca iz dveh populacij z različnima β_1 in β_2, za katera je $\sigma_{e1}^2 = \sigma_{e2}^2$

$$\frac{b_2-b_1-(\beta_2-\beta_1)}{s_d}\sqrt{\frac{K_{x_1}K_{x_2}}{K_{x_1}+K_{x_2}}} =: t\,(m=n_1+n_2-4) \qquad (10)$$

$$s_d^2=\frac{K_{y_1}-\dfrac{K_{xy_1}^2}{K_{x_1}}+K_{y_2}-\dfrac{K_{xy_2}^2}{K_{x_2}}}{n_1+n_2-4} \qquad (11)$$

Pomožni izrazi K_x K_y K_{xy} so znani izrazi, izračunani za vzorca 1 in 2.

13. lo. <u>Preskušanje hipotez o regresijskih koeficientih</u>

Iz 13. 9. privzeti izraz

$$\frac{b-\beta_H}{s_e}\sqrt{K_x} =: t\,(m=n-2)\ ;\ s_e^2=\frac{K_{y.x}}{n-2} \qquad (1)$$

uporabljamo za preskušanje značilnosti razlik stvarnega regresijskega koeficienta β od hipotetične vrednosti β_H

Podobno uporabljamo obrazec lo iz 13. 8 za preskušanje značilnosti razlik med dvema populacijama, za kateri imamo vzorca z n_1 in n_2 enotami

$$\frac{b_2-b_1}{s_d}\sqrt{\frac{K_{x_1}K_{x_2}}{K_{x_1}+K_{x_2}}} =: t\,(m=n_1+n_2-4) \qquad (2)$$

Pri ničelni hipotezi, da ni razlik med regresijskima koeficientoma namreč velja $H_0: \beta_1=\beta_2$

Po enakih principih uporabljamo tudi druge obrazce iz 13. 8 za preskušanje hipotez o parametrih za linearno regresijo in za določanje intervalnih ocen.

13. 11. Proučitev linearne regresije in korelacije iz podatkov, ki so grupirani v korelacijski tabeli

Podobno kot pri izračunanju aritmetične sredine in variance, tudi pri proučevanju korelacije iz grupiranih podatkov uporabljamo metodo pomožnih znakov. Ker pri korelaciji nastopata dva znaka (x in y), uvedemo tudi dva pomožna znaka (u in v).

Pri regresijskih premicah in merah linearne korelacije nastopajo količine $\bar{x}$ $\bar{y}$, σ_x^2, σ_y^2 in C_{xy}. Zato je količina, ki je po metodi u in v še ne znamo izračunati c_{xy}. Ker je

$$K_{uv} = \sum_{uv}\sum f_{uv} uv - \frac{U.V}{N} \qquad (1)$$

je nov le izraz $\sum_u \sum_v f_{uv} uv$. Ta izraz pa moremo izračunati na dva načina

$$\sum_v \left(\sum_u f_{uv} u\right) v = \sum_v vU_v \quad \text{ali} \quad \sum_u \left(\sum_v f_{uv} v\right) u = \sum_u uV_u \qquad (2)$$

tako, da za posamezne vrste najprej izračunamo $U_v = \sum_u f_{uv} u$, te vrednosti pa pomnožimo z ustreznimi v in seštejemo.

Podobno dobimo isto količino, če najprej izračunamo za vsak stolpec $V_u = \sum_v f_{uv} v$, te vrednosti pa pomnožimo z ustreznimi u in seštejemo. Obojni račun da kompleksno kontrolo izračuna razmeroma zamotanega postopka.

Iz korelacijske tabele izračunanih količin N, U, V, $\sum_u fu^2$ $\sum_{uv}\sum fuv$ in $\sum_v fv^2$, ki jih izračunamo na ta način, kot je prikazan v primeru, izračunamo parametre linearne regresije in korelacije po naslednji shemi

$$
\begin{array}{ccccccc}
 & & & \Sigma u^2 & \Sigma uv & \Sigma v^2 \\
N & U & V & -U^2/n & -UV/n & -V^2/n \\
\hline
 & i_x U/n & i_y V/n & K_u & K_{uv} & K_v
\end{array}
$$

$$
\frac{+ x_o}{\bar{x}} \quad \frac{+ y_o}{\bar{y}} \quad b_1 = \frac{i_y K_{uv}}{i_x K_u} \quad r_{xy}^2 = \frac{K_{uv}^2}{K_u K_v} \quad b_2 = \frac{i_x}{i_y} \frac{K_{uv}}{K_u}
$$

in dalje:

Prva regresijska premica : $y' = \bar{y} + b_1(x - \bar{x})$

Druga regresijska premica : $x' = \bar{x} + b_2(y - \bar{y})$

Za primer vzemimo odnos med λ_{10} in λ_{40} za patentirano jeseniško žico za 98 poskusov.

Odnos med x_{1o} in x_{4o} za patentirano jeseniško žico

y \ x	1,o	1,4	1,8	2,2	2,6	3,o	3,4	3,8	4,2	4,6	5,o	v	f_y	vf_y	$v^2 f_y$	U_y	vU_y
2,4	1											-1	1	-4	16	-4	16
3,o	1											-3	1	-3	9	-4	12
3,6		2	2	2								-2	6	-12	24	-12	24
4,2			3	3								-1	6	-6	6	-9	9
4,8			4	7	4							o	15	o	o	-15	o
5,4			2	18	12	2						1	34	34	34	-2o	-2o
6,o				2	8	5						2	15	3o	6o	+ 3	6
6,6					2	7	3	1				3	13	39	117	16	48
7,2						1	-	2	1			4	4	16	64	11	44
7,8								1	1	1		5	3	15	75	12	6o
8,4																	
u	-4	-3	-2	-1	o	1	2	3	4	5							
f_u	2	2	11	32	26	15	3	4	2	1			98	1o9	4o5	-22	199
uf_u	-8	-6 -	22	-32	o	15	6	12	8	5			-22 U	v	Σv^2	U	Σ uv
$u^2 f_u$	+32	+18	44	32	o	15	12	36	32	25			246 Σu^2				
V_u	-7	-4	-5	+15	34	37	9	16	9	5			1o9 V				
uV_u	28	12	1o	-15	o	37	18	48	36	25			119 Σuv				

N	U	V	$\sum u^2$	$\sum uv$	$\sum v^2$
98	-22	1o9	246	199	4o5
			- 4,9388	+ 24,4694	-121,2347
			241,o612	223,4694	283,7653

$$\sigma_u^2 = 2,4598 \quad C_{uv} = 2,28o3 \quad \sigma_v^2 = 2,8877$$

$$r^2_{xy} = o,732o \qquad r_{xy} = o,856$$

$$b_1 = \frac{C_{xy}}{\sigma_x^2} = \frac{i_y}{i_x}\frac{K_{uv}}{K_u} = \frac{o,6}{o,4}\,\frac{233,4694}{241,o612} = 1,3o95$$

$$\bar{x} = x_o + i_x \frac{U}{N} = 2,8 + o,4\,\frac{-22}{98} = 2,71o$$

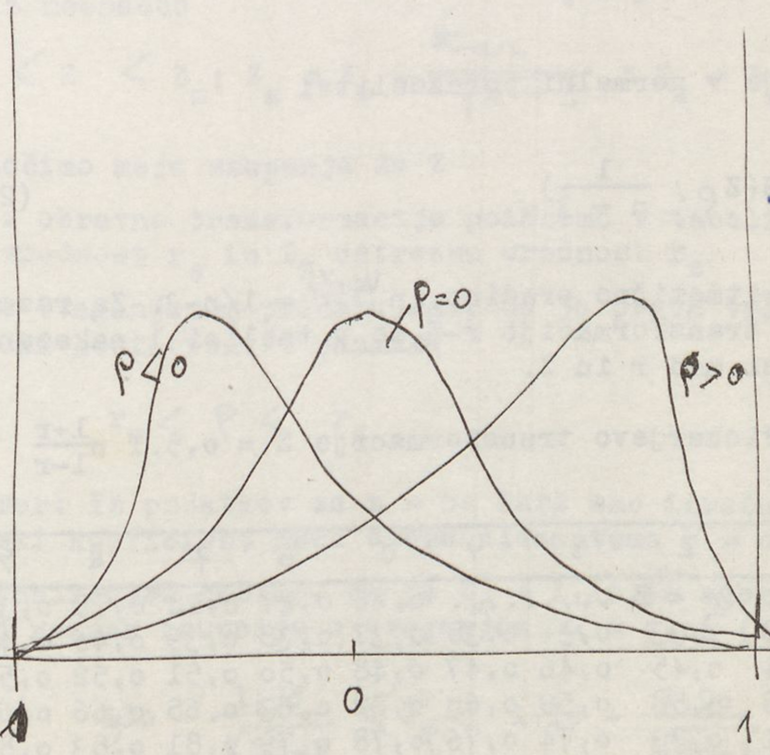
$$\bar{y} = y_o + i_y \frac{V}{n} = 5,1 + o,6\,\frac{1o9}{98} = 5,767$$

$$y' = \bar{y} + b_1 (x - \bar{x}) = 5,767 + 1,39o5\,(x - 2,71o) =$$

$$= 1,999 + 1,39o5\, x$$

13. 12. <u>Vzorčne porazdelitve korelacijskih koeficientov iz vzorcev</u>

Za primer, da je med x in y v osnovni populaciji linearna regresija, s korelacijskim koeficientom ϱ , je vzorčna porazdelitev korelacijskih koeficientov iz vzorcev z obsegom n precej zapletena. V splošnem so vzorčne porazdelitve za ocene korelacijskih koeficientov unimodalne, a simetrične. Simetrična je porazdelitev vzorčnih r le v primeru, da je ρ = o, v vseh drugih primerih pa je asimetrična. Stopnja asimetrije je tem večja, čim večji je ρ. Vzorčne porazdelitve za r iz vzorcev za nekaj vrednosti ρ so prikazane v sliki



Slika. Vzorčne porazdelitve za r pri populacijah z različnimi ρ

Za velike vzorce ($n > 200$) velja, da se r iz vzorcev porazdeljuje okrog prave vrednosti ρ približno v normalni porazdelitvi

$$r =: N\left(\rho ; \frac{(1-\rho^2)^2}{n-1}\right) \qquad (1)$$

13. 13. <u>Fisherjeva transformacija Z</u>

Vzorčna porazdelitev za r je razmeroma zapletena. R. A. Fisher pa je dokazal, da se transformirani izraz

$$Z_r = 0.5 . \ln \frac{1+r}{1-r} \qquad (1)$$

iz r porazdeljuje v normalni porazdelitvi

$$Z_r =: N\left(Z_\rho ; \frac{1}{n-3}\right) \qquad (2)$$

$EZ_r = Z_\rho$ kot aritmetično sredino in $\mathrm{Var}\,Z_r = 1/n-3$. Za razmeroma komplicirano transformacijo r-Z so v tablici 1 nakazani funkcijski odnosi med r in Z.

Tablica r-Z za Fisherjevo transformacijo $Z = 0,5 . \ln \frac{1+r}{1-r}$

r	0	1	2	3	4	5	6	7	8	9
0,20	....	....	r = Z	....	....	0,26	0,27	0,28	0,29	0,30
0,30	0,31	0,32	0,33	0,34	0,35	0,37	0,38	0,39	0,40	0,41
0,40	0,42	0,44	0,45	0,46	0,47	0,48	0,50	0,51	0,52	0,54
0,50	0,55	0,56	0,58	0,59	0,60	0,62	0,63	0,65	0,66	0,68
0,60	0,69	0,71	0,73	0,74	0,76	0,78	0,79	0,81	0,83	0,85
0,70	0,87	0,89	0,91	0,93	0,95	0,97	1,00	1,02	1,05	1,07
0,80	1,10	1,13	1,16	1,19	1,22	1,26	1,29	1,33	1,38	1,42
0,90	1,47	1,53	1,59	1,66	1,74	1,83	1,95	2,09	2,30	2,65
0,905	1,50	1,56	1,62	1,70	1,78	1,89	2,01	2,18	2,44	2,99

13. 14. Ocenjevanje korelacijskega koeficienta

Točkovna ocena za korelacijski koeficient ρ je

$$r = \frac{c_{xy}}{s_x \cdot s_y} \qquad (1)$$

izračunamo iz podatkov iz vzorca. Meje zaupanja za intervalno oceno za korelacijski koeficient pa dobimo s Fisherjevo transformacijo in znanimi zakonitostmi za normalno porazdelitev po naslednjem postopku:

a) iz podatkov vzorca z obsegom n izračunamo točkovno oceno za korelacijski koeficient r

b) iz tablice r-Z poiščemo oceni r ustrezno vrednost Z_r

c) z neenačbo

$$Z_s < Z < Z_z; \; Z_s = Z_r - \frac{z_{\alpha/2}}{\sqrt{n-3}} : Z_z = Z_r + \frac{z_{\alpha/2}}{\sqrt{n-3}} \qquad (2)$$

določimo meje zaupanja za Z

d) z obratno transformacijo poiščemo v tabeli r-Z Z_s ustrezno vrednost r_s in Z_z ustrezno vrednost r_z

e) s tveganjem α pričakujemo, da je prava vrednost za korelacijski koeficient v razmaku

$$r_s < \rho < r_z \qquad (3)$$

Primer: Iz podatkov za n = 5o šarž smo izračunali ~~med~~ korelacijski koeficient med dvema elementoma r = o,8o.

Iz tablice r-Z dobimo, da je Z_r = 1,1o. Po zgornjem postopku dobimo razmak zaupanja s tveganjem α = o,o5 (z = 1,96)

$$1{,}1o - \frac{1{,}96}{\sqrt{5o-3}} < Z_\rho < 1{,}1o + \frac{1{,}96}{\sqrt{5o-3}}$$

$$o{,}815 < Z < 1{,}385$$

Z obratno transformacijo pa dobimo iz tabele rZ meje zaupanja za

$$o{,}67 < \rho < o{,}88$$

13. 15. Preskušanje značilnosti linearne odvisnosti

Ker se v primeru, da x in y sta odvisna, izraz

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = t(m=n-2) \qquad (1)$$

porazdeljuje v t-porazdelitvi z m = n-2 stopinjami prostosti, ta izraz uporabljamo za preskušanje ničelne hipoteze o neodvisnosti. Kljub temu, da sta x in y neodvisna, so iz vzorcev izračunani r od nič različni. Razlike od nič so slučajnostne.

Vzemimo kot primer, da smo v dani raziskavi z n = 27 meritvami za x in y dobili, da je izračunani r = o,2o. Po zgornjem vzorčnem izrazu dobimo,

$$t = \frac{o{,}2o\sqrt{27-2}}{\sqrt{1-o{,}2o^2}} = +\,1{,}o2$$

Kritična vrednost $t(m=25) = 2{,}o6$
 o,o5

Ker je izračunana vrednost za t manjša kot kritična vrednost za α = o,o5 sklepamo, da so razlike neznačilne, kar pomeni, da iz vzorca ne moremo sklepati, da sta x in y odvisna.

13. 16. Preskušanje hipoteze o korelacijskem koeficientu

Če upoštevamo zakonitosti o transformiranih Fisherjevih Z vrednostih, preskušamo hipoteze o različnosti korelacijskega koeficienta od hipotetičnega po naslednjem postopku:

a) iz vzorčnih podatkov izračunamo r,

b) iz tablice r-Z ugotovimo r in ρ ustrezni vrednosti Z_r in Z_ρ

c) upoštevaje zakonitosti razporeditve Z_r izračunamo izraz

$$(z_r - Z_{\rho_H})\sqrt{n-3} = z \qquad (1)$$

d) če velja ničelna hipoteza, da je pravi korelacijski koeficient ϱ enak hipotetičnemu ϱ_H , se izraz z porazdeljuje standardizirano normalno. Značilnost razlik pa preskušamo po pravilih, ki so naznačene v odstavku o principih preskušanja hipotez.

Primer: Preskušamo ničelno hipotezo, da je $H_0 : \varrho = \varrho_H = 0,80$. V ta namen smo za n = 30 na slučajnosten način izbranih šarž ugotovili, da je r = 0,60. Iz tabele r=Z dobimo: $Z_r = 0,69$ in $Z_\varrho = 1,10$. Če vstavimo te podatke v obrazec dobimo:

$$(0,60 - 1,10)\sqrt{30 - 3} = -2,69$$

Ker je izračunani z = - 2,59 absolutno večji kot kritična vrednost $z_{\alpha=0,01}$ ~~= o,o1~~ = 2,58 sklepamo, da je prava vrednost korelacijskega koeficienta značilno različna od hipotetične $\varrho_H = 0,80$ na nivoju $\alpha = 0,01$.

13. 17. Preskušanje razlik med korelacijskimi koeficienti

Če upoštevamo zakonitosti transformiranih Z vrednosti in zakonitosti diferenc dveh normalno razporejenih slučajnostnih spremenljivk, preskušamo značilnost razlik med dvema korelacijskima koeficientoma po naslednjem postopku:

a) ničelna hipoteza, ki jo preskušamo je: $H_0 : \varrho_1 = \varrho_2$ ali $\Delta\varrho = 0$

b) iz podatkov vzorca za prvo populacijo izračunamo iz vzorca z n_1 enotami korelacijski koeficient r_1, iz podatkov vzorca za drugo populacijo z n_2 enotami pa korelacijski koeficinet r_2,

c) Iz tablice r-Z poiščemo r_1 in r_2 ustrezni vrednosti Z_1 in Z_2,

d) dobljene vrednosti vnesemo v obrazec

$$\frac{Z_1 - Z_2}{\sqrt{\dfrac{1}{n_1-3} + \dfrac{1}{n_2-3}}} = z \qquad (1)$$

e) Če velja ničelna hipoteza, da ni razlik med korelacijskima koeficientoma, se porazdeljuje standardizirano normalno, značilnost razlik pa preskušamo po pravilih za preskušanje značilnosti razlik za standardizirano normalno porazdeljene količine.

Primer. Preskušamo, ali je stopnja odvisnosti med karakteristikama x in y za proizvodnjo pod pogojema A in B značilno različna. V ta namen smo izvedli meritve za x in y za $n_1 = 103$ preskušance pod pogoji A in za $n_B = 53$ preskušancev pod pogoji B.

Dobili smo $r_A = 0{,}70$ $r_B = 0{,}80$

Iz tablic r-Z dobimo, da je $Z_A = 0{,}87$ in $Z_B = 1{,}10$. Če vnesemo ustrezne podatke v obrazec 1 dobimo

$$z = \frac{1{,}10 - 0{,}87}{\sqrt{\frac{1}{103-3} + \frac{1}{53-3}}} = 1{,}33$$

Ker je izračunani z manjši kot kritična vrednost 1,96 smatramo, da so razlike v korelacijskih koeficientih neznačilne.

13. 18. Krivuljčna regresija in korelacija

Za razliko od linerane regresije go-vorimo o krivuljčni regresiji takrat, kadar regresijska zveza med x in y ni linearna ampak krivuljčna. V tem primeru na splošno regresijsko odvisnost podamo z zvezo

$$y = g(x, \alpha, \beta, \varepsilon \ldots \varepsilon)$$

V tej splošni funkcijski obliki je x neodvisna y odvisna spremenljivka, $\alpha, \beta, \varepsilon$ so parametri regresijske zveze, kot sta pri linearni regresiji parametra α in β, ki določata konkretno regresijsko premico izmed vseh možnih premic, ε pa je rezultat slučajnostnih vplivov, za katerega predpostavljamo, da je $\varepsilon = : N(0, \sigma_\varepsilon^2)$ z enako varianco za vsak x.

13. 19. Indeks korelacije in determinacijski koeficient za krivuljčno korelacijo

Splošno merilo jakosti povezanosti med dvema znakoma oziroma pojavoma je razmerje pojasnjene variance in skupne variance. Ta koeficient na splošno imenujemo determinacijski koeficient. Ker je pojasnjena varianca razlika med skupno in nepojasnjeno - rezidualno varianco, je determinacijski koeficient za krivuljčno korelacijo enak

$$I^2_{y.x} = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_y^2} = 1 - \frac{\sigma_e^2}{\sigma_y^2} \qquad (1)$$

Kvadratni koren iz deteminacijskega koeficienta za krivuljčno korelacijo imenujemo indeks korelacije $I_{y.x}$ oziroma $I_{g(y)x}$, če je namesto y v regresijski funkciji nastopa g(y).

13. 2o. Poseben model regresijske odvisnosti

Zaradi tehničnih in teoretičnih prednosti, so posebej primerne za proučevanje regresije funkcije oblike

$$g(y) = a_o f_o(x) + a_1 f_1(x) + a_2 f_2(x) + \dots + a_r f_r(x) + e \qquad (1)$$

ali funkcije, ki jih je možno z ustrezno transformacijo prenesti nanjo. Posebnost te funkcije je v tem, da so parametri med seboj v linearni zvezi in da je rezultat slučajnostnih individualnih vplivov vezan z regresijsko funkcijo aditivno.

Model linearne regresije je poseben primer zgornjega modela: v

$$y = a + b(x-x) + e$$

je $g(y) = y$. $a_o = a$ $\quad f_o(x) = 1 \quad a_1 = b$; $f_1(x) = (x-\bar{x})$

13. 21. Določanje regresijskih krivulj po metodi najmanjših kvadratov

Po določitvi tipa krivulje, ki pride v poštev pri regresijski analizi konkretnih podatkov, je treba določiti numerične vrednosti parametrov regresijske funkcije.

Iz vsebinskega vidika smatramo za regresijsko krivuljo tisto, ki se stvarnim vrednostim najbolj prilega. Kot kriterij za prilagoditev je iz teoretičnih in praktičnih aspektov najprimerneje vzeti pogoj, da smatramo kot regresijsko krivuljo tisto, za katero je

$$S\left[g(y) - \sum_i a_i f_i(x)\right]^2 = F(a_o, a_1 \ldots\ldots a_r) = \min \qquad (1)$$

F je funkcija parametrov a_o, $a_1, \ldots\ldots a_r$. Potreben pogoj za nastop ekstrema, ki je v tem primeru minimum, je, da so parcialni odvodi funkcije F po parametrih a_i enaki nič.

Kot dobimo, če odvajamo po parametru a_o prvi pogoj

$$-2\,S\left[g(y) - a_i f_i(x)\right] f_o(x) = o \qquad (2)$$

ali v izdelani obliki

$$Sgf_o = a_o Sf_o^2 + a_1 Sf_o f_1 + \ldots\ldots + a_r Sf_o f_r \qquad (3)$$

dobimo podobno z odvajanjem dodatnih r pogojev in s tem sistem r+1 linearnih enačb z r+1 neznankami a_o, $a_1 \ldots\ldots a_r$

$$\begin{aligned} S_g f_o &= a_o Sf_o^2 + a_1 Sf_1 f_o + \ldots\ldots\ldots + a_r Sf_r f_o \\ S_g f_1 &= a_o Sf_o f_1 + a_1 Sf_1^2 + \ldots\ldots\ldots\ldots + a_r Sf_r f_1 \\ &\;\;\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4) \\ S_g f_r &= a_o Sf_o f_r + a_1 Sf_1 f_r + \ldots\ldots\ldots + a_r Sf_r^2 \end{aligned}$$

Ta sistem linearnih enačb za parametre a_i imenujemo sistem <u>normalnih enačb.</u>

Če je g(y) = y da metoda najmanjših kvadratov regresijsko funkcijo, ki se najbolje prilega vrednostim y, če pa to ni, dobimo po metodi najmanjših kvadratov funkcijo, za katero je vsota kvadratov odklonov od g(y) najmanjša.

13. 22. Standardna napaka ocene za krivuljčno regresijo standardnega tipa

Varianco odklonov stvarnih vrednosti g(y) od regresijskih vrednosti g(y') dobimo za standardni tip krivulje po obrazcu

$$\sigma_e^2 = \frac{1}{N}\left(\Sigma g^2 - a_0 \Sigma g f_0 - a_1 \Sigma g f_1 - \dots\dots a_r \Sigma g f_r\right) \quad (1)$$

parametri a_i so izračunani po metodi najmanjših kvadratov.
Če gre za vzorec n enot, dobimo nepristransko oceno variance, če namesto N vstavimo število stopinj prostosti, ki je v tem primeru n - r - 1, ker je n vrednosti vezanih na r + 1 zvez. Tako je ocena rezidualne variance

$$s_e^2 = \frac{1}{n-r-1}(Sg^2 - a_0 Sgf_0 - a_1 Sgf_1 - \dots\dots - a_r Sgf_r) \quad (2)$$

standardna napaka ocene pa je kvadratni koren iz zgornjih varianc. Standardna napaka ocene ima velik pomen že samostojno kot merilo individualnih, včasih slučajnostnih vplivov, obenem pa tudi kot kazalec zanesljivosti ocen, ki jih delamo iz regresijske krivulje o g(y) ali o y na osnovi poznavanja x.

13. 23. Regresijski model $g(y) = A + B.f(x) + e$

Regresijski model, ki je poseben primer splošnega modela (13. 20) dobimo, če vzamemo: $f_0(x) = 1$ $f_1(x) = f(x)$ $f_i(x) = 0$ $i = ,2 \dots\dots\dots r$

$$g(y) = a + b(f(x) + e \quad (1)$$

V tem modelu postavimo g(y) = Y in f(x) je X. Krivuljčno regresijo med x in y obravnavamo kot linearno regresijo med X in Y, ker je

$$Y = a + BX + e \quad (2)$$

Za ta primer velja vse, kar velja za linearno regresijo, tako za izračun parametrov, kot za izračun mere odvisnosti.

Vendar korelacijski koeficient r_{XY} linearne korelacije med X in Y ni korelacijski koeficient krivuljčne povezanosti med x in y in ga navadno pišemo kot $r_{f(x)g(y)}$.

Nakažimo nekaj primerov regresijskih funkcij zgornjega tipa ali pa funkcij, ki se dajo privesti na zgornji tip funkcije,

logaritemska funkcija: $y' = a + b\log x$ (3)

Na linearno privedemo to funkcijo s transformacijo $X = \log x$; $Y = y$

parabola: $y' = a + b\sqrt{x}$ transformacija: $X = x$ $Y = y$ (4)

hiperbola: $y' = a + b\frac{1}{x}$ transformacija: $X = \frac{1}{x}$ $Y = y$ (5)

Razen teh funkcij, ki so dane v standardni obliki direktno, moremo nekatere pomembne funkcije privesti v tako obliko. Na primer:

eksponentno funkcijo $y' = a.b^x$ (6)

prevedemo z logaritmiranjem v obliko

$$\log y' = \log a + x\log b \quad (7)$$

Parabolično funkcijo $y = ax^b$ (8)

prevedemo z logaritmiranjem v obliko

$$\log y' = \log a + b\log x \quad (9)$$

funkcijo $y = \frac{x}{a + bx}$ (1o)

prevedemo v standardno obliko, če vzamemo recipročne vrednosti

$$\frac{1}{y} = b. + a\frac{1}{x} \quad (11)$$

Parabola oblike $y' = (a+bx)^2$ (12)

dobimo standardno obliko, če izračunamo na levi in desni strani enačbe kvadratni koren

$$\sqrt{y'} = a + bx \tag{13}$$

Razen navedenih so jasno še druge funkcije, ki so bodisi že same na sebi dane v standardni obliki ali pa jih prevedemo z ustrezno transformacijo v standardno obliko.

Če rezimiramo, za take primere proučimo regresijo in korelacijo po naslednjih stopinjah:

a) Za določene tipe regresijske funkcije se odločimo bodisi na osnovi vsebinske analize ali na osnovi analize korelacijskega grafikona.

b) Če je potrebno, regresijsko funkcijo transformiramo v obliko $g(y) = A(a, b) + B(a,b).f(x)$

c) Glede na funkciji $g(y)$ in $f(x)$ prevedemo osnovne podatke v nove vrednosti $X = f(x)$ in $Y = g(y)$

d) Iz dobljenih transformiranih vrednosti X in Y določimo regresijsko premico $Y' = A + BX$ in izračunamo korelacijski koeficient r_{XY}

d) Z obratno transformacijo iz A(ab) in B(ab) poiščemo ustrezne vrednosti osnovnih parametrov a in b in vstavimo $Y' = g(y)$ in $X = f(x)$. Tako dobimo regresijsko krivuljo izraženo v osnovnih parametrih a in b in osnovnih spremenljivkah x in y

e) Korelacijski koeficient linearne korelacije med X in Y je korelacijski koeficient krivuljčne korelacije med x in y in ga pišemo $r_{f(x)g(y)}$.

13. 24. Polinom $y = a_0 + a_1x + a_2x^2 + \ldots + a_rx^r$ kot regresijska funkcija

Pogosto vzamemo kot regresijsko funkcijo kar polinom stopnje r, v katerem so funkcije f_i iz prejšnjega splošnega modela cele potence spremenljivke x. $f_i = x^i$.

Regresijski model za ta primer je:

$$y = a_o + a_1 x + a_2 x^2 + ... + a_r x^r + e \qquad (1)$$

Sistem normalnih enačb se v tem primeru poenostavi in dobimo

$$\begin{aligned}
S_y &= Na_o + a_1 Sx + a_2 Sx^2 + ... + a_r Sx^r \\
Syx &= a_o Sx + a_1 Sx^2 + a_2 Sx^3 + ... + a_r Sx^{r+1} \qquad (2) \\
Syx^2 &= a_o Sx^2 + a_1 Sx^3 + a_2 Sx^4 + ... + a_r Sx^{r+2} \\
\hline
Syx^r &= a_o Sx^r + a_1 Sx^{r+1} + a_2 Sx^{r+2} + ... + a_r Sx^{2r}
\end{aligned}$$

Vsote potenc od Sx^o do reda Sx^{2r} tvorijo simetrično matriko.

Če je x spremenljivka, za katero so dane vrednosti v enakih razmakih, dosežemo s transformacijo, s katero postavimo izhodišče za x v sredino, da so vse vsote lihih potenc Sx^{2k+1}. Tako sistem normalnih enačb znatno poenostavimo:

$$\begin{aligned}
Sy &= a_o N \qquad\qquad a_2 Sx^2 + ... a_r Sx^r \\
Syx &= \qquad a_1 Sx^2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3) \\
Syx^2 &= a_o Sx^2 + \qquad + a_2 Sx^4 + ... + a_r Sx^{r+2} \\
Syx^r &= a_o Sx^r + \qquad + a_2 SX^{r+2} + ... + a_r Sx^{2r}
\end{aligned}$$

ker sistem razpade v dva sistema linearnih enačb, ki sta med seboj neodvisna in eden vsebuje ene, drugi pa druge parametre.

13. 25. Multipla regresija

Osnovna značilnost korelacijskih odvisnosti je, da na določen pojav ne vpliva en sam faktor, ampak da je število teh faktorjev veliko. Neki od teh so opredeljivi in je njihov vpliv znaten, drugi pa niso opredeljivi in njihov vpliv opazujemo kot rezultat slučajnostnih faktorjev. Pri proučevanju odvisnosti enega pojava od določenega faktorja, druge opredeljive faktorje držimo konstantne. Vzrok v variiranju rezultativnega znaka je v tem primeru le proučevani faktor x in slučajnostni vplivi. To pa je dostikrat zaradi tehničnih ozirov težko. V tem primeru opredeljive, a nekontrolirane in neevidentirane faktorje združimo skupno s slučajnostnimi vplivi v individualne vplive. Jasno je, da je v takem primeru zakonitost povezanosti med x in y bolj zabrisana, ker na negotov rezultat vplivajo še dodatni - nekontrolirani vplivi. Razen tega pa je analiza odvisnosti od enega samega faktorja v kompleksnosti delovanja različnih faktorjev, ki morejo nastopiti, omejenega pomena. Zato z multiplo regresijo proučujemo odvisnost rezultativnega znaka y od več faktorjev oziroma faktorialnih znakov hkrati. Multiplo regresijsko odvisnost moremo napisati v splošni obliki

$$y = F(x_1, x_2, x_3 \ldots\ldots\ldots x_p, e) \qquad (1)$$

V tej obliki je zapisana korelacijska odvisnost y od x_1, x_2 ......x_p. Od funkcijske odvisnosti se zgornji izraz razlikuje v tem, da vsebuje še rezultat slučajnostnih vplivov e.

Pogosto predpostavljamo, da so vplivi opredeljenih faktorjev in individualni vplivi vezani aditivno. V tem primeru je model multiple regresije

$$y = F(x_1, x_2 \ldots\ldots\ldots\ldots x_p) + e \qquad (2)$$

13. 26. Multipla linearna regresija

Linearno odvisnost enega faktorja moremo posplošiti na več faktorjev. Multipli linearni regresijski model je:

$$y = a_o + b_1x_1 + b_2x_2 + \ldots\ldots + b_px_p + e \qquad (1)$$

regresijska ravnina v p + 1 v dimenzionalnem prostoru pa je

$$y' = a_o + a_1x_1 + a_2x_2 + \ldots\ldots\ldots + a_px_p$$

Obravnavanje multiple linearne regresije je enostavnejša, če namesto osnovnih znakov y, x_i uvedemo standardizirane znake

$$z_y = \frac{y - \bar{y}}{\sigma_y} \qquad z_i = \frac{z_i - \bar{z}_i}{\sigma_i} \qquad i = 1, 2 \ldots.p \qquad (2)$$

Regresijska ravnina je v tem primeru

$$z_y = a_o + b_1z_1 + b_2z_2 + \ldots\ldots\ldots\ldots + b_pz_p \qquad (3)$$

Opomba: parametri a_o b_k v 3 so drugi kot v 1.

13. 27. Določitev parametrov linearne regresije po metodi najmanjših kvadratov

Kot pri krivuljčni regresiji, tudi pri multipli regresiji uporabljamo za kriterij prilagoditve vsoto kvadratov odklonov stvarnih vrednosti y od vrednosti y', ki jih dobimo z regresijsko funkcijo

$$S(y - y')^2 \qquad (1)$$

in vzamemo za regresijsko ravnino tisto, ki se danim podatkom najbolje prilega.

Če upoštevamo, da za standardizirane znake na splošno velja:

$$\bar{z} = o \; ; \quad \overline{z^2} = 1 \quad \text{in} \quad \overline{z_iz_j} = r_{ij}$$

dobimo po metodi najmanjših kvadratov naslednji sistem normalnih enačb:

$$
\begin{aligned}
o &= a_o \\
r_{y1} &= b_1 r + b_2 r_{21} + b_3 r_{31} + \dots\dots\dots + b_p r_{p1} \\
r_{y2} &= b_1 r_{12} + b_2 r + b_3 r_{32} + \dots\dots\dots + b_p r_{p2} \qquad (2) \\
r_{y3} &= b_1 r_{13} + b_2 r_{23} + b_3 + \dots\dots\dots + b_p r_{p3} \\
&\vdots \\
r_{yp} &= b_1 r_{p1} + b_2 r_{p2} + b r_{p3} + \dots\dots\dots + b r_p
\end{aligned}
$$

Če zadnjim p normalnim enačbam (prva je degenerirana in samo nakaže, da gre regresijska ravnina skozi izhodišče koordinatnega sistema standardiziranih odklonov ali skozi točko, ki ima za koordinate aritmetične sredine ($\bar{y}$, $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_p$), dodamo še enačbo za regresijsko ravnino

$$z'_y = b_1 z_1 + b_2 z_2 + b_3 z_3 + \dots\dots\dots + b_p z_p \qquad (3)$$

dobimo, da je homogen sistem p + 1 linearnih enačb netrivialno rešljiv, če velja, da je determinanta

$$
\begin{vmatrix}
z'_y & z_1 & z_2 & \dots\dots & z_p \\
r_{y1} & 1 & r_{21} & \dots\dots & r_{p1} \\
r_{y2} & r_{12} & 1 & \dots\dots & r_{p2} \\
\dots & \dots & \dots & \dots & \dots \\
r_{yp} & r_{1p} & r_{2p} & \dots\dots & 1
\end{vmatrix} = o \qquad (4)
$$

enaka nič. Tako je na splošno napisana v obliki determinante regresijska ravnina. Če z D_{yi} zaznamujemo poddeterminante prve vrste, je enačba regresijske ravnine

$$z'_y D_{yy} + z_1 D_{y1} + z_2 D_{y2} + \dots + z_p D_{yp} = o \qquad (5)$$

ali eksplicitno

$$z'_y = -\frac{D_{y1}}{D_{yy}} z_1 - \frac{D_{y2}}{D_{yy}} z_2 - \ldots\ldots\ldots - \frac{D_{yp}}{D_{yy}} z_p \qquad 6$$

regresijski koeficienti b_i so torej enaki koeficientom poddeterminant

$$b_i = -\frac{D_{yi}}{D_{yy}}$$

13. 28. Determinacijski koeficient za multiplo linearno korelacijo

Z determinantami moremo izraziti determinacijski koeficient za multiplo linearno korelacijo z naslednjo zvezo

$$R^2_{y.12\ldots p} = 1 - \frac{D}{D_{yy}}$$

pri čemer je

$$D = \begin{vmatrix} 1 & r_{y1} & r_{y2} & \ldots\ldots\ldots\ldots & r_{yp} \\ r_{1y} & 1 & r_{12} & \ldots\ldots\ldots\ldots & r_{1p} \\ r_{2y} & r_{21} & 1 & \ldots\ldots\ldots\ldots & r_{2p} \\ \ldots & \ldots & \ldots & \ldots\ldots\ldots\ldots & \ldots \\ r_{py} & r_{p1} & r_{p2} & \ldots\ldots\ldots\ldots & 1 \end{vmatrix}$$

D_{yy} pa poddeterminanta, ki ustreza prvemu členu v prvi vrsti v determinanti D. Kvadratni koren iz determinacijskega koeficienta $R_{y.12\ldots\ldots p}$ pa je multipli korelacijski koeficient.

13. 29. Multipla regresija in linearna korelacija za odvisnost y od dveh faktorjev

Če upoštevamo splošne zakonitosti, je multipla regresijska ravnina za odvisnost y od dveh faktorjev dana s funkcijo

$$\frac{y' - \bar{y}}{\sigma_y} = b_{y1.2} \frac{x_1 - \bar{x}_1}{\sigma_1} + b_{y2.1} \frac{x_2 - \bar{x}_2}{\sigma_2} \qquad (1)$$

pri čemer so

$$b_{y1.2} = \frac{r_{y1} - r_{y2} r_{12}}{1 - r_{12}^2} \quad ; \quad b_{y2.1} = \frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2} \qquad (2)$$

Determinacijski koeficient $R^2_{y.12}$ je po splošnem obrazcu enak

$$R^2_{y.12} = \frac{r_{y1}^2 + r_{y2}^2 - 2 . r_{y1} r_{y2} r_{12}}{1 - r_{12}^2} \qquad (3)$$

standardna napaka ocene za y pa je

$$\sigma_{y.12} = \sigma_y \sqrt{1 - R^2_{y.12}} \qquad (4)$$

Za multiple regresije in linearne korelacije z več neodvisnimi spremenljivkami pa je najbolje izračunati ustrezne količine prek determinant.

13. 3o. Multipla parabolična regresija in korelacija

Podobno kot pri proučevanju enega znaka moremo tudi pri multipli regresiji razširiti proučevanje na nelinearne odvisnosti. Najnavadnejša je razširitev na kvadratično funkcijo. Ta ima za primer odvisnosti rezultativnega znaka y od dveh faktorialnih znakov x_1 in x_2 naslednji model

$$y = a + b_1 x_1 + b_2 x_2 + c_{11} x_1^2 + 2c_{12} x_1^2 x_2 + c_{22} x_2^2 + e \qquad (1)$$

Ta model ima šest parametrov. Če si geometrijsko ponazorimo regresijsko ploskev (če v zgornjem modelu izpustimo e), je to ploskev drugega reda.

Matrika količin, ki so potrebne za izračun parametrov kvadratične regresijske funkcije po metodi najmanjših kvadratov preko normalnih enačb je

$$
\begin{array}{ccccccc}
\underline{S_y} & \underline{N} & \underline{Sx_1} & \underline{Sx_2} & \underline{Sx_1^2} & \underline{Sx_1x_2} & \underline{Sx_2^2} \\
\underline{Syx_1} & Sx_1 & Sx_1^2 & Sx_1x_2 & \underline{Sx_1^3} & \underline{Sx_1^2x_2} & \underline{Sx_1x_2^2} \\
\underline{Syx_2} & Sx_2 & Sx_1x_2 & Sx_2^2 & Sx_1^2x_2 & Sx_1x_2^2 & \underline{Sx_2^3} \\
\underline{Syx_1^2} & Sx_1^2 & Sx_1^3 & Sx_1^2x_2 & \underline{Sx_1^4} & \underline{Sx_1^3x_2} & \underline{Sx_1^2x_2^2} \\
\underline{Syx_1x_2} & Sx_1x_2 & Sx_1^2x_2 & Sx_1x_2^2 & Sx_1^3x_2 & Sx_1^2x_2^2 & \underline{Sx_1x_2^3} \\
\underline{Syx_2^2} & Sx_2^2 & Sx_1x_2^2 & Sx_2^3 & Sx_1^2x_2^2 & Sx_1x_2^3 & \underline{Sx_2^4}
\end{array} \qquad (2)
$$

V tej matriki so vsi nepodčrtani izrazi ponovljeni in zato ni treba izračunati 6.7 = 42 členov matrike, ampak jih je med seboj različnih le polovica t. j. 21.

Analogno multipli ali krivuljčni regresiji moremo izračunati multipli determinacijski koeficient bodisi preko ustreznih determinant ali preko obrazca

$$
\sigma^2_{y.12} = \frac{1}{N}\Big[Sy^2 - aSy - b_1 Syx_1 - b_2 Syx_2 - C_{11} Syx_1^2 -
- C_{12} Syx_1^2 x_2 - C_{22} Syx_2^2 \Big] \qquad (3)
$$

Če gre za ocenjevanje z vzorcem, nadomestimo v obrazcu N s stopinjami prostosti n - 6. Za kvadratično krivuljčno regresijsko ploskev pogosto nastopi vprašanje vrednosti faktorjev x_1 in x_2, ki dajo ekstremno (minimalno ali maksimalno vrednost) za odvisno spremenljivko.

Z odvajanjem regresijske krivulje dobimo, da morata za ta primer x_1 in x_2 zadoščati pogojema

$$b_1 + 2c_{11}x_1 + c_{12}x_2 = o$$

$$b_2 + c_{12}x_1 + 2c_{22}x_2 = o$$

ki jih dobimo tako, da postavimo prve parcialne odvode po x_1 in x_2 enake o.

13. 31. Parcialna korelacija

Korelacijske odvisnosti nastopajo kompleksno. Pri analizi korelacijskih odvisnosti zato opazimo, da ni nujno, da sta med seboj v odvisnosti zaradi tega, ker vplivata eden na drugega neposredno, temveč se dostikrat zgodi, da se pokaže odvisnost med pojavoma 1 in 2 zaradi tega, ker isti faktor o ali celo sklop istih faktorjev, katere zaznamujemo kar s skupnim simbolom A, vplivajo tako na pojav 1 kot na pojav 2 in se zaradi tega pokaže posredno korelacijska odvisnost med 1 in 2. Korelacijo med pojavoma 1 in 2 brez posrednega učinka vzajemnega faktorja o imenujemo parcialno korelacijo.

Parcialno korelacijo proučujemo tako, da predhodno iz učinkov pojava 1 in 2 x_1 in x_2 izločimo vpliv faktorja o in proučujemo korelacijo med znakoma $x_{1.o}$ in $x_{2.o}$, iz katerih smo vpliv posrednega faktorja eliminirali. Podobno je v primeru, če eliminiramo posreden vpliv spleta faktorjev "A" proučujemo korelacijo med $x_{1.A}$ in $x_{2.A}$, to je iz podatkov, iz katerih smo izločili vpliv kompleksa vzajemnih faktorjev A.

13. 32. Linearni parcialni korelacijski koeficient

Ker je eliminacija vplivov razmeroma zamotana, se pri proučevanju parcialne korelacije običajno omejimo na eliminacijo* učinka faktorjev. Če iz pojavov 1 in 2, katerih numerični izraz sta znaka x_1 in x_2, eliminiramo linearen vpliv vzajemnega faktorja o, katerega numerični izraz je znak x_o, dobimo parcialni korelacijski koeficient med 1 in 2, iz katerih je eliminiran linearni učinek pojava o.

linearnega

Za tak primer je možno izračunati parcialni korelacijski koeficient $r_{12.o}$ po obrazcu

$$r_{12.o} = \frac{r_{12} - r_{o1} r_{o2}}{\sqrt{1-r^2_{o1}} \cdot \sqrt{1-r^2_{o2}}} \qquad (1)$$

iz navadnih korelacijskih koeficientov.

Po obrazcu

$$r_{12.oA} = \frac{r_{12.A} - r_{o1.A} r_{o2.A}}{1 - r^2_{o1.A} \cdot 1 - r^2_{o2.A}} \qquad (2)$$

izračunamo parcialni korelacijski koeficient med pojavoma 1 in 2, iz katerih smo eliminirali faktor o in kompleks faktorjev A, če poznamo parcialne korelacijske koeficiente $r_{12.A}$, $r_{o1.A}$ in $r_{o2.A}$ iz katerih je eliminiran samo kompleks A. Tako moremo priti postopno do parcialnih korelacijskih koeficientov, iz katerih eliminiramo poljubno število posrednih faktorjev iz navadnih korelacijskih koeficientov.

13. 33. Zveza med multiplim linearnim koeficientom $R_{y.12......p}$ in parcialnimi korelacijskimi koeficienti

Analitično pomembna je zveza med multiplim linearnim korelacijskim koeficientom $R_{y.12....p}$ in parcialnimi korelacijskimi koeficienti r_{y1}, $r_{y2.1}$, ... $r_{yp.123...(p-1)}$

$$1 - R^2_{y.12...p} = (1-r^2_{y1}).(1-r^2_{y2.1}).........(1-r^2_{yp.12...(p-1)})$$

Delež nepojasnjene variance pri multipli korelaciji je enak produktu deležev nepojasnjene variance s parcialnimi korelacijskimi koeficienti.