Dr Stana Barić,
Poljoprivredni fakultet, Zagreb

STATISTIČKE METODE PRIMIJENJENE U STOČARSTVU

UVOD

Istraživanja u stočarstvu imaju zadatak, kao i istraživanja na drugim područjima, da otkriju osebine i karakteristike pojava, upoznaju prirodu i njene zakonitosti, utvrde pojave te njihove uzroke i međusobne odnose. S obzirom da je stočarstvo naučno i stručno područje, koje se zasniva na živim bićima, čija je jedna od osnovnih karakteristika — varijabilnost, to je većina pojava u stočarstvu, ako ne i sve, varijabilnog karaktera. Odatle proizlazi neophodnost poznavanja statističkih metoda za svakog istraživača i stručnog radnika.

U svom radu istraživač u pravilu ne može obuhvatiti cjelokupnu populaciju jedinki jedne pojave, nego se mora poslužiti manjim brojem predstavnika populacije — uzorkom, koji na takav način postaje osnovni predmet njegovog istraživanja. U nemogućnosti, dakle, da izmjerimo ili utvrdimo osebine svih pripadnika jedne populacije služimo se manjim brojem, čijom analizom želimo utvrditi osebine — parametre populacije. Da bi se to moglo i da bi vrijednosti procjene populacije na osnovu analiza uzoraka bili što bliži parametrima, uzorak mora zadovoljavati neke zahtjeve, a i tehnika naučnog rada mora biti adekvatno postavljena.

Prvi zahtjev, kome uzorak mora udovoljavati, jest da po svom sastavu i osebinama bude reprezentant populacije. Da bi on to mogao biti, odabiranje jedinki koje će sačinjavati uzorak, mora biti izvršeno bez ikakvog subjektivnog utjecaja istraživača tj. potpunom slučajnošću. Drugi se zahtjev odnosi na veličinu uzorka, koji mora biti dovoljno velik, kako bi mogao bar donekle obuhvatiti pojavu u cijeloj njezinoj širini i intenzitetima. No uza sve to uzorci će više ili manje odstupati od parametara populacije i to za nepoznatu devijaciju, što uvjetuje stvaranje zaključaka uz stanovitu vjerojatnoću npr. 95% ili 99%. Tek na osnovu ponavljanja istih analiza i istraživanja, tj. ispitivanja većeg broja uzoraka, dobit će se tačna slika populacije s obzirom na ispitivanu pojavu.

S obzirom da se ocjena parametara populacije u stočarskim istraživanjima mora vršiti pomoću istraživanja uzoraka, to je neophodno poznavati metode procjene parametara pa ćemo se time posebno baviti.

Međutim, statističke metode nisu neophodne samo u ispitivanjima svojstava pojedinih populacija, nego su one isto toliko neophodne i u analizama eksperimentalnih rezultata kojima se služimo u otkrivanju zakonitosti pojava. Pojedine zakonitosti javljaju se maskirane osebinama, koje su nastale pod raznim utjecajima, uslijed čega dolazi do varijabilnosti koja je osnovna statistička vrijednost. Radi toga i eksperimenti moraju biti tako postavljeni, da se dobiveni podaci mogu statistički obraditi, pa je statistička metoda neophodna ne samo u analizama nego i u planiranju pokusa.

U analizi rezultata istraživanja statističke metode dobivaju naročiti značaj.

Svi dobiveni eksperimentalni podaci, bez obzira na tačnost metode njihovog skupljanja i ispravnosti rada, sadrže u sebi rezultate uvjetovane kontroliranim i nekontroliranim faktorima. Nismo, naime, u stanju fiksirati uvjete eksperimenta na takav način da potpuno isključimo djelovanje nekontroliranih faktora. Zato dolazi do različitih rezultata kod ponavljanja istih istraživanja, a opseg djelovanja tih nekontroliranih faktora može se utvrditi samo statističkim metodama, čime se zapravo omogućuje stvaranje zaključaka.

Statističke metode omogućuju i testiranje postavljenih hipoteza u eksperimentalnom radu kao i utvrđivanje odnosa između raznih pojava, pa su one neophodno sredstvo u istraživačkom radu te ćemo se u toku ovih izlaganja osvrnuti na njihovu primjenu u stočarstvu.

I. DIO

ANALIZA UZORKA

GRUPIRANJE I PRIKAZ PODATAKA

Istraživačkim radom, bilo kojeg nivoa i tipa, dobivamo određeni broj podataka i prvi problem, koji susrećemo nakon toga, jest način njihovog prikaza i pripreme za statističku obradu. Ukoliko je uzorak mali tj. ako sadrži mali broj podataka dobit će se uvid u kretanje pojave, koja se ispituje već samim prikazom pojedinih vrijednosti, te se može odmah pristupiti i odgovarajućim statističkim analizama. Osnovnu misao toga objasnit ćemo prikazom podataka jednog malog uzorka i to težine u kg 10 berksire prasadi stare 28 dana:

Težina berkshire prasadi stare 28 dana u kg:

6,95	6,53	4,92	5,90	5,30	6,12	5,60	7,10	6,80	6,20

Uvidom u navedeni prikaz možemo odmah uočiti zastupanost pojedinih težina u uzorku, tj. možemo odmah uočiti kretanje pojave koja se ispituje.

Ako uzorak sadrži veći broj podataka, koji su prikazani u nesređenom redoslijedu, podaci su nepregledni, te se bez prethodnog sređivanja i grupiranja teško dobiva uvid u kretanje i zastupanost pojedine pojave u uzorku. Da bi mogli izvršiti analizu vrijednosti pojedinih obilježja, te primijeniti metodu statističke analize, koja je često kraća i jednostavnija za grupirane podatke, moramo ih grupirati.

Nesređene podatke velikog uzorka te njihovo grupiranje prikazat ćemo pomoću podataka mliječnosti istočno-frizijskih krava u 305 dana.

Mliječnost istočno-frizijskih krava u 305 dana u kg

3856	3672	4422	3044	5639	3885	4277	3487	3208	4467
3822	4032	4001	5284	3467	3635	2804	4428	3818	5873
3831	4217	4056	4427	4066	4214	3618	3813	5722	4863
4442	3221	4681	5004	3007	4222	3424	5627	3006	4682
3853	5830	4283	3477	4024	4036	5230	4092	4247	4253
4662	4060	4432	3804	4231	5569	2681	4261	4862	3601
3485	3402	4621	5481	3266	3846	4068	3456	4055	4233
4201	5401	4015	2436	5281	4255	5462	3871	4265	3473
4606	4475	4266	5808	3616	4013	5571	4881	4085	5680
3883	3619	4657	2875	4691	5086	2878	4802	5878	4853
4031	4453	2828	3852	4801	4071	3426	4445	4405	4223
4852	4693	4009	4483	3699	3852	3868	3231	6095	3417
5663	3891	5220	3279	4069	6300+	4282	2836	4034	4235
4211	5022	4093	4809	5043	4437	4228	3607	5273	3852
3895	3665	3486	6682	5213	3875	4207	4602	3823	3690
4045	4412	4632	4230	3655	6220	6028	4412	6079	3682
4467	4250	4122							

71805 + 72315 + 71824 + 68975 + 67768 + 71596 + 66052 + 65851 + 72855 + 69077 = 698038

$\bar{x} = 4282,44$

Kako jednostavnim pregledom podataka ne možemo dobiti uvid u zastupanost pojedinih proizvodnji, nameće se potreba grupiranja. Prikazani podaci grupirat će se tako, da će se odrediti veličina razreda, ispisati razredne granice (nominalne i prave) i nakon toga će se svako obilježje upisati u odgovarajući razred. Za prikazani uzorak odredit ćemo za veličinu razreda vrijednost od 200 kg, pa ćemo grupirati podatke, kako je prikazano u tabeli 1. 1.

Tabela 1.1

Mliječnost istočno-frizijskih krava u 305 dana u kg

Nominalne granice razreda	Prave granice razreda	Sredina razreda	Frekvencija	Kumulativni niz »Manje od« (gornje granice razreda)	Kumulativni niz »Više od« (donje granice razreda)
2400—2599	2400—2600	2500	1	1	163
2600—2799	2600—2800	2700	1	2	162
2800—2999	2800—3000	2900	5	7	161
3000—3199	3000—3200	3100	3	10	156
3200—3399	3200—3400	3300	5	15	153
3400—3599	3400—3600	3500	11	26	148
3600—3799	3600—3800	3700	12	38	137
3800—3999	3800—4000	3900	19	57	125
4000—4199	4000—4200	4100	21	78	106
4200—4399	4200—4400	4300	22	100	85
4400—4599	4400—4600	4500	15	115	63
4600—4799	4600—4800	4700	10	125	48
4800—4999	4800—5000	4900	8	133	38
5000—5199	5000—5200	5100	4	137	30
5200—5399	5200—5400	5300	6	143	26
5400—5599	5400—5600	5500	5	148	20
5600—5799	5600—5800	5700	5	153	15
5800—5999	5800—6000	5900	4	157	10
6000—6199	6000—6200	6100	3	160	6
6200—6399	6200—6400	6300	2	162	3
6400—6599	6400—6600	6500	0	162	1
6600—6799	6600—6800	6700	1	163	1

U prvu kolonu unijete su nominalne vrijednosti razrednih granica i to radi tačnosti unašanja podataka. Kada bi bile unijete samo prave razredne granice, moglo bi doći do zabune prilikom grupiranja kao npr. spada li obilježje 4400 kg u razred 4200—4400 ili u razred 4400—4600.

U drugu kolonu unijete su prave razredne granice, a izračunavanjem njihove aritmetičke sredine dobiva se razredna sredina koja je unijeta u treću kolonu.

Kod grupiranja broj razreda može varirati, no mora se imati u vidu, da premali broj razreda previše uopćava podatke, dok preveliki broj razreda komplicira rad. Da li smo ispravno izvršili grupiranje može nam poslužiti kontrola iza izračunavanja standardne devijacije, koja mora biti kod precizne statističke obrade 4 puta veća od veličine razreda. Da to postignemo obično trebamo najmanje 20 razreda.

Frekvencije koje su unijete u četvrtu kolonu, dobivene su na temelju upisivanja pojedinog obilježja u razred u koji spada.

Vrijednosti kumulativnog niza »manje od« dobiju se na taj način, da se u prvi razred upiše frekvencija prvog razreda. U drugi se razred upiše vrijednost prvog razreda plus frekvencija drugog razreda, u treći razred se unosi vrijednost drugog razreda plus frekvencija trećeg razreda itd. Na taj način vrijednost kumulativnog niza »manje od« naznačuju broj obilježja koja imaju manju vrijednost od gornje granice razreda. Na primjer u petoj koloni vrijednost 57 naznačuje, da 57 grla ima manju mliječnost od 4000 kg.

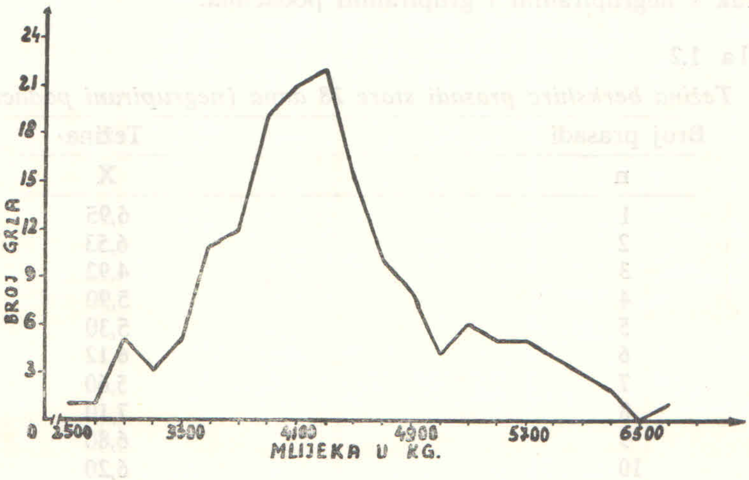
Kumulativni niz »više od« dobije se na isti način kao i kumulativni niz »manje od«, samo što se vrijednosti počinju unositi od zadnjeg razreda. U zadnji razred dolazi frekvencija zadnjeg reda. U predzadnji red dolazi frekvencija zadnjeg reda plus frekvencija predzadnjeg reda itd. Pojedine vrijednosti kumulativnog niza »više od« naznačuju broj obilježja koje imaju višu vrijednost od donje granice razreda. Npr. vrijednost u šestoj koloni u petom redu 153 naznačuje, da 153 grla imaju višu mliječnost od 3200 kg.

Podatke dobivene na temelju ankete, eksperimentalnog rada i statističke obrade, osim tabelarnog prikaza, često je vrlo korisno prikazati i grafički. Prema podacima i svrsi prikaza autor će izabrati odgovarajući način prikaza bilo linearnog, površinskog ili prostornog grafičkog prikaza. Kod toga se uvijek mora paziti da sam prikaz bude očit i da slikovito prikazuje kretanje pojave koja se ispituje.

Npr. uzorak iz tabele 1.1 prikazat ćemo grafički pomoću varijacionog poligona.

Grafikon 1.1

Mliječnost istočno-frizijskih krava u 305 dana



MJERE CENTRALNE TENDENCIJE

Distribucije obilježja, koje se ispituju u stočarskim istraživanjima, osim rijetkih izuzetaka, pokazuju tendenciju grupiranja oko srednje vrijednosti (centralna tendencija). Kako prikaz podataka treba biti što pregledniji, kao i podesan za dalju statističku obradu nastojat ćemo ispitivana obilježja prikazati pomoću reprezentativnih vrijednosti centralne tendencije, tj. pomoću srednje vrijednosti.

Od mjera centralne tendencije prikazat ćemo aritmetičku srednju vrijednost, medijan, mode i harmonijsku sredinu.

Aritmetička sredina

Aritmetička sredina naznačuje prosječnu vrijednost obilježja u uzorku ili populaciji, te reprezentira vrijednost svih obilježja iz kojih se izračunala.

Formule za izračunavanje aritmetičke sredine su slijedeće:

$$\mu = \frac{X_1 + X_2 + X_3 + \ldots + X_N}{N} = \frac{\Sigma X}{N} \quad \text{(negrupirani podaci)} \quad (1.1)$$

$$\bar{x} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{\Sigma X}{n} \quad \text{(negrupirani podaci)} \quad (1.2)$$

$$\mu = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \ldots + f_N X_N}{N} = \frac{\Sigma f X}{N} \quad \text{(grupirani podaci)} \quad (1.3)$$

$$\bar{x} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \ldots + f_n X_n}{n} = \frac{\Sigma f X}{n} \quad \text{(grupirani podaci)} \quad (1.4)$$

μ = aritmetička sredina populacije
$\bar{x}$ = aritmetička sredina uzorka
N = broj jedinica u populaciji
n = broj jedinica u uzorku
X = vrijednost pojedinog obilježja
f = frekvencija (broj obilježja u jednoj grupi)

U slijedećim primjerima prikazat će se izračunavanje aritmetičke sredine za uzorak s negrupiranim i grupiranim podacima:

Tabela 1.2

Težina berkshire prasadi stare 28 dana (negrupirani podaci)

Broj prasadi n	Težina X
1	6,95
2	6,53
3	4,92
4	5,90
5	5,30
6	6,12
7	5,60
8	7,10
9	6,80
10	6,20
Ukupno	61,42

$$\bar{x} = \frac{\Sigma X}{n} = \frac{61,42}{10} = 6,142 \text{ kg}$$

Primjer za izračunavanje aritmetičke sredine uzoraka s grupiranim obilježjima, dajemo pomoću podataka iz tabele 1.1

T a b e l a 1.3

Mliječnost istočno-frizijskih krava u 305 dana (kg) (grupirani podaci)

Količina mlijeka	Broj grla	Ukupno kg
X	f	fX
2500	1	2500
2700	1	2700
2900	5	14500
3100	3	9300
3300	5	16500
3500	11	38500
3700	12	44400
3900	19	74100
4100	21	86100
4300	22	94600
4500	15	67500
4700	10	47000
4900	8	39200
5100	4	20400
5300	6	31800
5500	5	27500
5700	5	28500
5900	4	23600
6100	3	18300
6300	2	12600
6500	0	0
6700	1	6700
Ukupno	163	706300

$$\bar{x} = \frac{\Sigma fX}{n} = \frac{706300}{163} = 4.333{,}1288 \text{ kg}$$

Ako nemamo električnog računskog stroja, aritmetičku sredinu, kao i operacije koje ćemo kasnije izložiti, često ćemo računati pomoću kodiranja.

Formula za izračunavanje aritmetičke sredine pomoću kodiranja je slijedeća:

$$\bar{x} = a + I\,\bar{d} \qquad (1.5)$$

a = jedna od razrednih sredina (izabrana po volji)

I = razredna veličina

$\bar{d} = \frac{\Sigma fd}{n}$

$d = \frac{(X - a)}{I}$

(»d« je code i u praktičnom se radu ne izračunava nego unosi u tabelu. U razredu gdje je »a«, »d« je jednak nuli, te se zatim mijenja za jedinicu u svakom slijedećem razredu. Prema tome, njegova je vrijednost 0, —1, —2, —3 itd. prema manjim vrijednostima obilježja i 0, +1, +2, +3 itd. prema većim vrijednostima obilježja).

Primjer za izračunavanje aritmetičke sredine pomoću kodiranja:

Tabela 1.4

Mliječnost istočno-frizijskih krava u 305 dana (kg)

Razredna sredina	Frekvencija	Code	Umnožak
X	f	d	fd
2500	1	—9	—9
2700	1	—8	—8
2900	5	—7	—35
3100	3	—6	—18
3300	5	—5	—25
3500	11	—4	—44
3700	12	—3	—36
3900	19	—2	—38
4100	21	—1	—21
4300	22	0	0
4500	15	1	15
4700	10	2	20
4900	8	3	24
5100	4	4	16
5300	6	5	30
5500	5	6	30
5700	5	7	35
5900	4	8	32
6100	3	9	27
6300	2	10	20
6500	0	11	0
6700	1	12	12
Ukupno	163		27

$\bar{x} = a + I\bar{d}$

$\bar{x} = 4300 + (200)\ (0{,}165644)$

$\bar{x} = 4333{,}1288$ kg

$a = 4300$

$I = 200$

$$\bar{d} = \frac{\Sigma fd}{n} = \frac{27}{163} = 0{,}165644$$

Medijan

Medijan je vrijednost obilježja, koje se nalazi u sredini distribucije frekvencija, tj. u sredini niza obilježja koja su poredana po veličini njihove vrijednosti.

Ako je broj obilježja u uzorku neparan, medijan je jednak vrijednosti obilježja $(n + 1)/2$, a ako je n paran, medijan se dobije izračunavanjem sredine između vrijednosti obilježja $n/2$ i $(n + 2)/2$.

Medijan za uzorak težine berkshirske prasadi stare 28 dana (tabela 1. 2) dobit ćemo na taj način, da obilježja poredamo po njihovoj vrijednosti, te će sredina između obilježja $n/2$ i $(n + 2)/2$ naznačavati medijan.

Težina berkshirske prasadi stare 28 dana poredane po veličini iznosi:

4,92 5,30 5,60 5,90 6,12 6,20 6,53 6,80 6,95 7,10 kg.

Vrijednost obilježja $n/2$ iznosi 6,12 kg, a vrijednost obilježja $(n + 2)/2$ iznosi 6,20 kg. Njihova sredina, koja je ujedno i medijan, jednaka je 6,16 kg.

Ako bi gornji uzorak imao još jedno obilježje (npr. 7,15 kg) medijan bi bio vrijednost jedinice $(n+ 1)/2$ tj. šestog obilježja i iznosio bi 6,20 kg.

Za grupirane podatke medijan se izračunava na taj način, da se prvo nađe medijalni razred, tj. razred u kojem se nalazi obilježje $(n + 1)/2$, ako je n neparan, ili sredina između $n/2$ i $(n + 2)/2$, ako je n paran. Kod daljnjeg izračunavanja može se primijeniti slijedeća jednadžba:

$$\text{Med} = X_L + \frac{(n_q - f_L)\, I}{f} \qquad (\ .6)$$

X_L = donja granica medijalnog razreda

n_q = broj medijalnog obilježja

f_L = vrijednost kumulativnog niza »manje od« u razredu prije medijalnog

f = frekvencija medijalnog razreda

I = veličina razreda

Izračunavanje medijana za grupirane podatke prikazat ćemo na uzorku mliječnosti istočno-frizijskih krava u 305 dana (tabela 1.1).

Iz tabele 1. 1 očitavamo slijedeće podatke:

$$n_q = 82,\ X_L = 4200,\ f_L = 78,\ f = 22,\ I = 200$$

$$\text{Med} = 4200 + \frac{(82 - 78)\, 200}{22} = 4236{,}364 \text{ kg}$$

Mode

Mode je vrijednost obilježja, koja je zastupljena s najvećom frekvencijom tj. mode jedne distribucije jest vrijednost oko koje obilježja pokazuju najjaču tendenciju grupiranja. Prema tome, može se reći, da je to najtipičnija vrijednost serije. Ako nema takve tendencije nema ni modea.

Očitavanje modea prikazat ćemo najprije na sasvim jednostavnom primjeru:

Težina 8 teladi iznosi: 39, 42, 43, 44, 44, 44, 45 i 46 kg.

Očito je, da je vrijednost modea 44 kg. U slučaju opet, kada težine 8 oteljenih teladi iznose 38, 39, 40, 41, 42, 44, 45 i 47 kg nema modea.

Kod grupiranih podataka imamo slučajeva, kada su obilježja grupirana tako, da u jednom razredu sva obilježja imaju istu vrijednost. U tom slučaju očitavanje modea je vrlo jednostavno, jer onaj razred koji ima najjaču frekvenciju naznačuje vrijednost modea.

Kod grupiranih podataka, kod kojih se u jednom razredu nalaze obilježja s različitim numeričkim vrijednostima, prvo ćemo odrediti modalni razred, tj. razred s najvećom frekvencijom, pa ćemo za izračunavanje modea primijeniti slijedeću formulu:

$$Mo = X_L + \frac{(f_2 - f_1)\, I}{(f_2 - f_1) + (f_2 - f_3)} \qquad (1.7)$$

X_L = donja granica mod. razreda

f_1 = frekvencija razreda koji prethodi modalnom razredu

f_2 = frekvencija modalnog razreda

f_3 = frekvencija razreda iza modalnog

I = veličina razreda

Izračunavanje modea prikazat ćemo na primjeru mliječnosti istočno-frizijskih krava u 305 dana. Iz tabele 1. 1 uzimamo slijedeće podatke:

$$X_L = 4200,\ f_1 = 21,\ f_2 = 22,\ f_3 = 15,\ I = 200$$

$$Mo = 4200 + \frac{(22 - 21)\ 200}{(22 - 21) + (22 - 15)} = 4.225{,}00 \text{ kg.}$$

Karakteristike aritmetičke srednje vrijednosti, medijana i modea i njihova primjena

Aritmetička srednja vrijednost u uspoređenju s medijanom i modeom najosjetljivija je srednja vrijednost, jer u njenom izračunavanju učestvuju sva obilježja. Čim se vrijednost jednog obilježja promijeni, mijenja se i aritmetička sredina.

Aritmetička srednja vrijednost ima dvije važne osebine koje daju bazu za daljnju statističku analizu. Prva važna osebina jest, da je zbroj devijacija pojedinih obilježja od aritmetičke sredine jednak nuli tj.

$$\Sigma\,(X - \bar{x}) = 0$$

Ukoliko računamo devijacije od bilo koje druge vrijednosti, a ne od aritmetičke sredine, zbroj devijacija neće biti jednak nuli.

Druga važna osebina jest, da je zbroj kvadrata devijacija pojedinih obilježja od aritmetičke sredine jednak minimumu.

$$\Sigma(X - \bar{x})^2 = \text{minimumu}$$

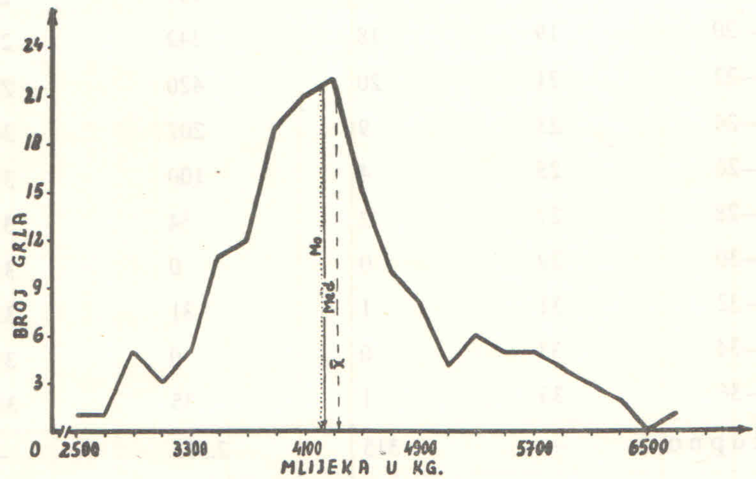
Zbroj kvadrata devijacija pojedinih obilježja od bilo koje druge vrijednosti umjesto aritmetičke sredine jeste veća vrijednost.

Medijan i mode su mnogo neosjetljivije srednje vrijednosti. Pretpostavimo da u izloženom primjeru u kojem je izračunat medijan neka obilježja iznad medijana povećaju svoje vrijednosti. Očito je, da se medijan neće promijeniti. Isto tako, ako podacima o uzorku na kojem je prikazano izračunavanje modea dodamo još nekoliko obilježja s velikim vrijednostima, mode se neće promijeniti.

U grafikonu 1.2 prikazat ćemo aritmetičku sredinu, medijan i mode za distribuciju koja ima naglašenu centralnu tendenciju i to za frekvenciju uzorka mliječnosti istočno-frizijskih krava u 305 dana (podaci iz tabele 1.1).

Grafikon 1.2

Mliječnost istočno-frizijskih krava u 305 dana sa naznačenim vrijednostima za $\bar{x}$, Med i Mo



Iz grafičkog prikaza $\bar{x}$, Med i Mo za uzorak s izraženom centralnom tendencijom (grafikon 1.2) očito je da postoje razlike između prikazanih srednjih vrijednosti, ali da one nisu velike. No sve te srednje vrijednosti, unutar svoga specifičnog značaja, reprezentiraju centralnu tendenciju.

U slijedećoj tabeli i grafikonu prikazat ćemo $\bar{x}$, Med i Mo za uzorak, čija se distribucija bitno udaljuje od simetrične.

Tabela 1.5

Ugibanje teladi od poroda do 36 dana starosti

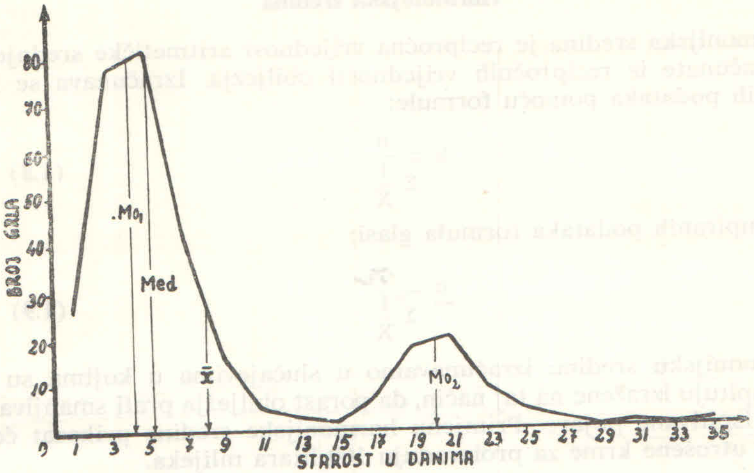
Starost u danima		Broj uginule teladi	Umnožak	Kumulativni niz »manje od«
granice razreda	sredine razreda			
	X	f	fX	
0— 2	1	26	26	26
2— 4	3	78	234	104
4— 6	5	82	410	186
6— 8	7	41	287	227
8—10	9	16	144	243
10—12	11	5	55	248
12—14	13	3	39	251
14—16	15	2	30	253
16—18	17	7	119	260
18—20	19	18	342	278
20—22	21	20	420	298
22—24	23	9	207	307
24—26	25	4	100	311
26—28	27	2	54	313
28—30	29	0	0	313
30—32	31	1	31	314
32—34	33	0	0	314
34—36	35	1	35	315
Ukupno	—	315	2.533	—

Prikazani podaci pokazuju primarnu i sekundarnu tendenciju grupiranja podataka oko stanovitih vrijednosti te ćemo izračunati osim Mo_1 i Mo_2.

$$x = 8{,}04 \text{ dana}$$
$$Med = 5{,}32 \text{ dana}$$
$$Mo_1 = 4{,}18 \text{ dana}$$
$$Mo_2 = 20{,}31 \text{ dana}$$

Grafikon 1.3

Ugibanje teladi od poroda do 36 dana starosti s naznačenim $\bar{x}$, Med i Mo



U navedenom primjeru $\bar{x}$, kao što se vidi iz grafikona 1.3, slabo reprezentira pojavu, koja se ispituje, jer starost od 8,04 dana tj. vrijednost aritmetičke sredine nije karakteristična za starost teladi koja ugiba. Med (5,32 dana), Mo_1 (4,18 dana) i Mo_2 (20,31 dana) dat će mnogo bolji opis prikazane distribucije frekvencije.

Usporedbom grafičkog prikaza $\bar{x}$, Med i Mo uzorka mliječnosti istočno-frizijskih krava u 305 dana (grafikon 1.2) sa grafičkim prikazom, $\bar{x}$, Med i Mo uzorka ugibanja teladi od poroda do 36 dana starosti (grafikon 1.3) lako uočavamo, da je u drugom uzorku veća asimetrija uzrokovala i veće udaljavanje vrijednosti $\bar{x}$ od Med i Mo.

Prilikom analize podataka, koji imaju nesimetrične distribucije, a kod kojih se pojavljuje tendencija skupljanja obilježja oko dviju vrijednosti, bit će od interesa navesti vrijednosti i drugog modea kao npr. u uzorku ugibanja teladi od poroda do 36 dana starosti ($Mo_1 = 4{,}2$ dana, $Mo_2 = 20{,}3$ dana).

Kod distribucije, koja je simetrična kao na primjer normalna distribucija, $\bar{x}$, Med i Mo se poklapaju. S povećanjem asimetrije naznačene srednje vrijednosti sve se više međusobno razlikuju. Ukoliko je asimetrija pozitivna tj. desnostrana mod $<$ med $< \bar{x}$, a ukoliko je negativna ili lijevostrana $\bar{x} <$ med. $<$ mod. Na primjer, asimetrija u primjeru mliječnosti istočno-frizijskih krava u 305 dana pozitivna je ili desnostrana (Mo $= 4.225{,}0$ kg $<$ Med $= 4.236{,}4$ kg $< \bar{x} = 4.333{,}1$ kg).

Koju ćemo sredinu upotrijebiti ovisi o podacima koje ispitujemo kao i o svrsi obrade. Najviše se primjenjuje aritmetička srednja vrijednost koja je baza za izračunavanje standardne devijacije. U slučaju kada uzorak koji se analizira pripada populaciji koja je bitno udaljena od normalne distribucije, vrijedan će nam podatak za sredinu pružiti medijan i mode, kao u prikazanom primjeru ugibanja teladi od poroda do 36 dana starosti.

Harmonijska sredina

Harmonijska sredina je recipročna vrijednost aritmetičke srednje vrijednosti izračunate iz recipročnih vrijednosti obilježja. Izračunava se kod negrupiranih podataka pomoću formule:

$$h = \frac{n}{\Sigma \frac{1}{X}} \tag{1.8}$$

a kod grupiranih podataka formula glasi:

$$h = \frac{n}{\Sigma \frac{f}{X}} \tag{1.9}$$

Harmonijsku sredinu izračunavamo u slučajevima u kojima su pojave koje se ispituju izražene na taj način, da porast obilježja prati smanjivanje vrijednosti ispitivane pojave. Primjenu harmonijske sredine prikazat ćemo na primjeru utrošene krme za proizvodnju 100 litara mlijeka.

Tabela 1.6

Utrošena krma (H. J.) za proizvodnju 100 litara mlijeka

Oznaka grla	1	2	3	4	5	6	7	8	9
HJ	145,98	113,25	97,09	84,03	75,07	69,69	66,22	63,37	61,54

Da bi prikazali značaj harmonijske sredine prvo ćemo za primjer prikazan u tabeli 1.6 razložiti tačnost i upotrebljivost aritmetičke sredine.

U gornjem primjeru aritmetička srednja vrijednost iznosi: $\bar{x} = 86{,}25$ HJ. Ukoliko bi izračunata aritmetička srednja vrijednost odgovarala zahtjevu, da reprezentira količinu HJ potrebnu za proizvodnju 100 litara mlijeka, to bi u slučaju da svako grlo utroši 86,25 HJ ($\bar{x}$) količina mlijeka, koju poizvedu sva grla, morala iznositi 900 litara (9 grla × 100). Za svako ćemo grlo primjenom pravila trojnog izračunati koliko bi to grlo proizvelo mlijeka ako dobije 86,25 HJ (kvocijent aritmetičke srednje vrijednosti uzorka i utrošene količine HJ za proizvodnju 100 litara mlijeka pojedinog grla pomnožit ćemo sa 100 t.j. $\frac{\bar{x}}{X}$ 100). Na izloženi način izračunate količine proizvedenog mlijeka od 86,25 HJ za pojedino grlo iznosi: 59,08 76,16 88,83 102,64 114,89 123,76 130,25 136,10 140,15. Zbroj je jednak 971,9 litara. Kako se izračunata suma (971,9 l) znatno razlikuje od očekivane vrijednosti (900 l), očito je, da za navedeni primjer ne odgovara aritmetička srednja vrijednost.

Ako pak primijenimo harmonijsku sredinu dobit ćemo drugačiju vrijednost. Za navedeni uzorak harmonijska sredina iznosi h = 79,87 HJ.

Da li nam odgovara harmonijska sredina ispitat ćemo na isti način kao što smo ispitali i aritmetičku srednju vrijednost tj. izračunat ćemo za svako grlo vrijednosti: $\frac{h}{X}$ 100. Ako svakom grlu damo 79,87 HJ, tada će pojedino grlo proizvesti ove količine mlijeka: 54,71 70,52 82,26 95,05 106,39 114,61 120,61 126,04 129,78. Zbroj je jednak 900 litara. Kako smo dobili očekivanu vrijednost, znači da harmonijska sredina zadovoljava i znatno bolje reprezentira svaku pojedinu varijablu.

Svaka od prikazanih srednjih vrijednosti ($\bar{x}$, Med, Mo, h) ima svoju specifičnu ulogu u tumačenju podataka, te će ovisiti o načinu prikaza rezultata i svrsi analize, koju ćemo srednju vrijednost upotrijebiti.

U slučaju potrebe rješavanja problema procjene ukupne mase ili testiranja opravdanosti hipoteza podatke ćemo iznositi na taj način, da porast vrijednosti obilježja prati i porast pojave koju ispitujemo. U našem primjeru u tom bi slučaju prikazali podatke proizvedene količine mlijeka od 100 HJ, što bi omogućilo primjenu aritmetičke srednje vrijednosti i ostalih statističkih analiza koje su s njom u vezi.

Ukoliko nam je prikladniji i svrsishodniji način izlaganja podataka na način da njihov porast naznačuje pad vrijednosti ispitivane pojave, moramo srednju vrijednost izračunati pomoću harmonijske sredine.

MJERE DISPERZIJE

Analiza uzorka ili populacije treba dati uvid u karakteristike osebina i pojava koje se ispituju. Najprije smo prikazali izračunavanje srednjih vrijednosti uzorka. Prikazom srednjih vrijednosti uopćavaju se svi ispitivani podaci i izražavaju u jednoj vrijednosti, ali se kroz nju dobija samo donekle uvid u vrijednost obilježja koja se ispituju. Dva uzorka npr. mogu imati iste srednje vrijednosti ali ako imaju različitu raspršenost i varijabilnost obilježja, to će i vrijednost svojstva koje se ispituje biti sasvim različito. Srednje vrijednosti npr. za mliječnost u 305 dana za dva uzorka (svaki uzorak od druge pasmine) iznosi 3.500 kg. Ako se npr. mliječnost u prvom uzorku kreće od 1.500 do 5.000 kg, a u drugom od 3.000 do 4.500 kg, ne možemo ih izjednačiti po kvaliteti. Kako je varijabilitet pojedinih obilježja od temeljne važnosti, mora se obzirom na utvrđivanje stupnja disperzije izvršiti odgovarajuća statistička obrada. U daljnjem razlaganju prikazat ćemo tri mjere apsolutne disperzije i to: varijacionu širinu, interkvartil i standardnu devijaciju i relativnu mjeru disperzije: varijacioni koeficijent.

Varijaciona širina

Varijaciona širina jest apsolutna mjera varijabilnosti koja naznačuje razliku između najvećeg i najmanjeg obilježja, a često se prikazuje u obliku podataka najviše i najniže vrijednosti obilježja. Izražava se u apsolutnim mjerama kao što su izražena i sama obilježja. Ona je najjednostavnija mjera za izračunavanje varijabilnosti no ne govori ništa o samom smještaju obilježja unutar granica varijacione širine, tj. da li su obilježja grupirana oko srednje vrijednosti ili su opet dosta raširena i prema granicama varijacione širine.

Varijaciona širina za uzorak težine berkshire prasadi stare 28 dana (tabela 1.2) iznosi 2,18 kg tj. od 4,92—7,10 kg. Varijaciona širina za uzorak mliječnosti istočno-frizijskih krava u 305 dana (grupirani podaci prikazani u tabeli 1.1) iznosi 4.246 tj. od 2.436 do 6.682 kg.

Interkvartil

Interkvartil je apsolutna mjera devijacije, koja naznačuje razmak u kojem se nalazi polovica svih obilježja oko medijana, a dobije se izračunavanjem diferencija između donjeg i gornjeg kvartila. Medijan je vrijednost obilježja u sredini distribucije, donji kvartil je vrijednost obilježja koji dijeli donju polovicu distribucije u dva jednaka dijela, a gornji kvartil je vrijednost obilježja koji dijeli gornju polovicu distribucije u dva jednaka dijela. Na taj način distribucija se podijeli u četiri dijela, a svaki dio sadrži 1/4 obilježja. Očito je, da interval između gornjeg i donjeg kvartila sadrži polovicu obilježja.

Prilikom izračunavanja interkvartila za grupirane podatke najprije moramo odrediti donji i gornji kvartilni razred. Donji kvartilni razred jest onaj u kojem se nalazi jedinica $n_{Q_1} = (n + 1)/4$, a u gornjem se kvartilnom razredu nalazi jedinica $n_{Q_3} = 3\,(n + 1)/4$.

Gornji i donji kvartil izračunava se po formuli:

$$Q_1 = X_L + \frac{(n_{Q_1} - f_L)\,I}{f} \qquad (1.10)$$

$$Q_3 = X_L + \frac{(n_{Q_3} - f_L)\,I}{f} \qquad (1.11)$$

X_L = vrijednost donje granice kvartilnog razreda
n_{Q_1} = broj obilježja koji naznačuje donji kvartil
n_{Q_3} = broj obilježja koji naznačuje gornji kvartil
f_L = frekvencija kumulativnog niza u razredu prije kvartilnog
f = frekvencija kvartilnog razreda
I = veličina razreda

Donji kvartil naznačujemo sa Q_1 i u formulu unosimo vrijednosti n_{Q_1}, f_L i f za donji kvartil, a gornji naznačujemo sa Q_3 i unosimo odgovarajuće vrijednosti za gornji kvartil.

Što je diferencija između gornjeg i donjeg kvartila manja ($Q_3 - Q_1$) to su obilježja više grupirana oko medijana.

Izračunavanje interkvartila prikazat ćemo na uzorku mliječnosti istočno-frizijskih krava u 305 dana (tabela 1.1).

$n_{Q_1} = 41 \quad X_L = 3.800 \quad f_L = 38 \quad f = 19 \quad I = 200$

$$Q_1 = 3.800 + \frac{(41-38)\,200}{19} = 3.831,5789 \text{ kg}$$

$n_{Q_3} = 123 \quad X_L = 4.600 \quad f_L = 115 \quad f = 10$

$$Q_3 = 4.600 + \frac{(123-115)\,200}{10} = 4.760,00 \text{ kg}$$

U ispitivanom uzorku 50% grla nalazi se unutar granica 3.831,58 kg do 4.760,00 kg. Vrijednost interkvartila jeste $Q_3 - Q_1 = 928,42$ kg.

Upotreba interkvartila u biološkim ispitivanjima je rjeđa od standardne devijacije no interkvartil nam može vrlo dobro poslužiti kao mjera varijabilnosti kod obilježja čija se distribucija bitno razlikuje od normalne distribucije. Primjer takve nesimetrične distribucije jest distribucija uzorka ugibanja teladi od poroda do 36 dana starosti (tabela 1.5), te ćemo prikazati izračunavanje vrijednosti njegovog interkvartila.

$n_{q_1} = 79 \quad X_L = 2 \quad f_L = 26 \quad f = 78$

$$Q_1 = 2 + \frac{(79-26)\,2}{78} = 3,4 \text{ dana}$$

$n_{q_3} = 237 \quad X_L = 8 \quad f_L = 227 \quad f = 16$

$$Q_3 = 8 + \frac{(237-227)\,2}{16} = 9,2 \text{ dana}$$

Interkvartil $Q_3 - Q_1 = 5,8$ dana. Znači da je 50% teladi uginulo u 5,8 dana i to između 3,4 i 9,2 dana.

Standardna devijacija

Mjera varijabilnosti, koja se u statističkoj obradi eksperimentalnog materijala najviše primjenjuje, jest standardna devijacija koja je drugi korijen iz varijance.

Varijancu populacije dobijemo ako kvadrate diferencija između pojedinih obilježja i aritmetičke sredine zbrojimo i podijelimo sa brojem obilježja:

$$\sigma^2 = \Sigma (X - \mu)^2/N \quad \text{ili skraćeno} \quad \sigma^2 = \Sigma x^2/N$$

Izračunavanje varijance temelji se na osebini aritmetičke srednje vrijednosti, da je suma kvadrata udaljenosti pojedinog obilježja od aritmetičke sredine jednaka minimumu: $\Sigma (X - \mu)^2 =$ minimum. Znači da je varijanca prosjek kvadrata udaljenosti pojedinog obilježja od aritmetičke sredine.

Kako smo napomenuli, drugi korijen iz varijance jest standardna devijacija, koja je apsolutna mjera disperzije, a temeljena je na udaljenosti svakog pojedinog obilježja od $\bar{x}$.

Standardna devijacija izračunava se za čitavu populaciju prema slijedećim jednadžbama:

$$\sigma = \sqrt{\frac{\Sigma (X - \mu)^2}{N}} \qquad \text{(negrupirani podaci)}$$

$$\sigma = \sqrt{\frac{\Sigma f (X - \mu)^2}{N}} \qquad \text{(grupirani podaci)}$$

U najvećem broju slučajeva izračunavanje standardne devijacije po izloženoj metodi vrlo je dugotrajno, a može se pojednostavniti primjenom slijedećih formula:

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2} \quad \text{(negrupirani podaci)} \quad (1.12)$$

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left(\frac{\Sigma fX}{N}\right)^2} \quad \text{(grupirani podaci)} \quad (1.13)$$

Izračunavanje standardne devijacije populacije prikazat ćemo u tabeli 1.7 na primjeru godišnje nesivosti kokoši farme A.

Tabela 1.7

Godišnja nesivost kokoši farme A

X	f	fX	fX²
155	60	9.300	1,441.500
165	92	15.180	2,504.700
175	1.035	181.125	31,696.875
185	1.087	201.095	37,202.575
195	1.133	220.935	43,082.325
205	1.318	270.190	55,388.950
215	1.533	329.595	70,862.925
225	1.639	368.775	82,974.375
235	1.501	352.735	82,892.725
245	1.331	326.095	79,893.275
255	1.042	265.710	67,756.050
265	831	220.215	58,356.975
275	603	165.825	45,601.875
285	406	115.710	32,977.350
295	326	96.170	28,370.150
305	63	19.215	5,860.575
Ukupno	14.000	3,157.870	726,863.200

$$\mu = \frac{\Sigma fX}{N} = \frac{3,157.870}{14.000} = 225,56 \text{ jaja}$$

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left(\frac{\Sigma fX}{N}\right)^2} = \sqrt{\frac{726,863.200}{14\,000} - \left(\frac{3,157\,870}{14\,000}\right)^2}$$

$$\sigma = 32,272 \text{ jaja}$$

Standardna devijacija za uzorak označava se sa »s« i ona procjenjuje standardnu devijaciju ukupne populacije.

Izračunava se prema slijedećoj formuli:

$$s = \sqrt{\frac{\Sigma (X - \bar{x})^2}{n - 1}}$$

U daljim računima označavat ćemo $(X - \bar{x}) = x$, te prema tome formula glasi:

$$s = \sqrt{\frac{\Sigma x^2}{n - 1}} \quad \text{(negrupirani podaci) (1.14)}$$

$$s = \sqrt{\frac{\Sigma f x^2}{n - 1}} \quad \text{(grupirani podaci) (1.15)}$$

Računanje brojnika odbijanjem svakog X od $\bar{x}$, često je vrlo zamorno, te ćemo prikazati jednostavniji način izračunavanja Σx^2 i $\Sigma f x^2$:

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n} \quad \text{(negrupirani podaci) (1.16)}$$

$$\Sigma f x^2 = \Sigma f X^2 - \frac{(\Sigma f X)^2}{n} \quad \text{(grupirani podaci) (1.17)}$$

U nekoliko ćemo primjera prikazati izračunavanje standardne devijacije uzorka.

Izračunavanje standardne devijacije uzorka (s) s negrupiranim podacima prikazat ćemo na podacima težine berkshire prasadi stare 28 dana. (Podaci iz tabele 1.2).

Tabela 1.8

Težina berkshire prasadi stare 28 dana

Prase broj	Težina	Težina na kvadrat
n	X	X^2
1	6,95	48,3025
2	6,53	42,6409
3	4,92	24,2064
4	5,90	34,8100
5	5,30	28,0900
6	6,12	37,4544
7	5,60	31,3600
8	7,10	50,4100
9	6,80	46,2400
10	6,20	38,4400
Ukupno	61,42	381,9542

G. 147

$$s = \sqrt{\frac{\Sigma x^2}{n-1}}$$

$$s = \sqrt{\frac{4,7126}{9}} = 0,7236 \text{ kg}$$

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$\Sigma X^2 \ . \ . \ . \ . \ . = 381,9542$$

$$\frac{(\Sigma X)^2}{n} = \frac{61,42^2}{10} = 377,2416$$

$$\Sigma x^2 = 4,7126$$

Primjer izračunavanja standardne devijacije uzorka (s) sa grupiranim podacima prikazat ćemo na uzorku mliječnosti istočno-frizijskih krava u 305 dana (podaci iz tabele 1.3).

Tabela 1.9

Mliječnost istočno-frizijskih krava u 305 dana

X	f	fX	fX²
2500	1	2500	6250000
2700	1	2700	7290000
2900	5	14500	42050000
3100	3	9300	28830000
3300	5	16500	54450000
3500	11	38500	134750000
3700	12	44400	164280000
3900	19	74100	288990000
4100	21	86100	353010000
4300	22	94600	406780000
4500	15	67500	303750000
4700	10	47000	220900000
4900	8	39200	192080000
5100	4	20400	104040000
5300	6	31800	168540000
5500	5	27500	151250000
5700	5	28500	162450000
5900	4	23600	139240000
6100	3	18300	111630000
6300	2	12600	79380000
6500	0	0	0
6700	1	6700	44890000
Ukupno	163	706300	3164830000

$$s = \sqrt{\frac{\Sigma f x^2}{n-1}}$$

$$s = \sqrt{\frac{104341104,295}{162}}$$

$$s = 802,54 \text{ kg}$$

$$\Sigma f x^2 = \Sigma f X^2 - \frac{(\Sigma f X)^2}{n}$$

$$\Sigma f X^2 = . \ . \ . \ . = 3164830000,000$$

$$\frac{(\Sigma f X)^2}{n} = \frac{706300^2}{163} = 3060488895,705$$

$$\Sigma f x^2 = 104341104,295$$

Ako nemamo prikladni računski stroj, te želimo što više pojednostaviti računanje primijenit ćemo metodu računanja pomoću kodiranja.

Računanje ćemo provesti prema slijedećoj formuli:

$$s = I\, s_d \tag{1.18}$$

I = veličina razreda

s_d = standardna devijacija kodiranog obilježja = $\sqrt{\frac{\Sigma f x^2_d}{n-1}}$

$$\Sigma f x^2_d = \Sigma f d^2 - \frac{(\Sigma f d)^2}{n}$$

A tabeli 1.10 prikazat ćemo izračunavanje standardne devijacije uzorka (s) pomoću kodiranja (podaci iz tabele 1.4)

T a b e l a 1.10

Mliječnost istočno-frizijskih krava u 305 dana

X	f	d	fd	fd^2
2500	1	— 9	— 9	81
2700	1	— 8	— 8	64
2900	5	— 7	—35	245
3100	3	— 6	—18	108
3300	5	— 5	—25	125
3500	11	— 4	—44	176
3700	12	— 3	—36	108
3900	19	— 2	—38	76
4100	21	— 1	—21	21
4300	22	0	0	0
4500	15	1	15	15
4700	10	2	20	40
4900	8	3	24	72
5100	4	4	16	64
5300	6	5	30	150
5500	5	6	30	180
5700	5	7	35	245
5900	4	8	32	256
6100	3	9	27	243
6300	2	10	20	200
6500	0	11	0	0
6700	1	12	12	144
Ukupno:	163	—	27	2613

$$s = l\, s_d \qquad \Sigma f x^2_d = \Sigma f d^2 - \frac{(\Sigma f d)^2}{n}$$

$$s_d = \sqrt{\frac{\Sigma f x^2_d}{n-1}} \qquad \Sigma f d^2 \; . \; . \; . \; . \; . \; 2613$$

$$s_d = \sqrt{\frac{2608{,}528}{162}} = 4{,}0127 \qquad \frac{(\Sigma f d)^2}{n} = \frac{27^2}{163} = \underline{\quad 4{,}4724}$$

$$\Sigma f x^2_d = 2608{,}5276$$

$$s = l\, s_d = 200\,(4{,}0127)$$

$$s = 802{,}54 \text{ kg}$$

Varijacioni koeficijent

Prilikom statističke analize obilježja jednog uzorka temeljni će podaci, osim ostalih statističkih vrijednosti, prikazati i varijabilnost ispitivanih obilježja. Kako su distribucije obilježja, koje ispitujemo u stočarskoj problematici, skoro sva vrlo blizu normalnoj distribuciji, to će se varijabilitet ispitivanog uzorka izraziti pomoću standardne devijacije (s). Standardna devijacija, koja je apsolutna mjera varijabilnosti i izražena u istim mjerama kao i aritmetička srednja vrijednost, u nekim slučajevima statističke analize neće zadovoljiti. Na teškoće analize varijabiliteta pomoću standardne devijacije naići ćemo na primjer u stanovitim komparacijama varijabiliteta. Vrijednost standardne devijacije odijeljena od ostalih podataka ne daje odgovor o veličini disperzije obilježja. Uslijed toga dvije iste standardne devijacije ne moraju naznačavati i istu disperziju i obrnuto; dvije različite vrijednosti standardne devijacije mogu naznačavati istu disperziju. Standardna devijacija težine krava npr. uvijek je veća od standardne devijacije novorođene teladi. Znači da je u svim slučajevima uvijek veća apsolutna varijabilnost težine krava od težine teladi. Iako je težina krava i teladi data u istim apsolutnim mjerama (kg) ne možemo uspoređivati njihovu varijabilnost na temelju apsolutnih mjera. Slučaj je još očitiji, tj. komparacija nemoguća, kada analiziramo varijabilitet obilježja koja nisu izražena u istim mjerama, kao npr. mliječnost krava (kg) i masnoća mlijeka (%).

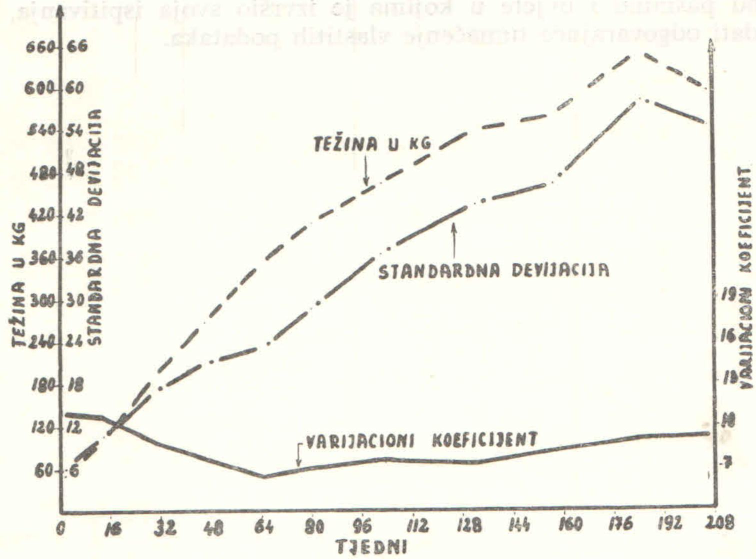
U svrhu dobivanja mjere varijabiliteta, koja je neovisna od mjera u kojima su izražena obilježja, često se upotrebljava koeficijent varijacije (C) koji je omjer standardne varijacije i aritmetičke sredine ($s/\bar{x}$). Varijacioni koeficijent se obično pomnoži sa 100, te se dobiva standardna devijacija izražena u postocima aritmetičke srednje vrijednosti: $\frac{s}{\bar{x}}\,100$ (1.19). Već je iz same formule očito, da obilježja više grupirana oko aritmetičke srednje vrijednosti imaju manji varijacioni koeficijent. Na taj način koeficijent varijacije govori o homogenosti analiziranih obilježa, bez obzira na apsolutne vrijednosti obilježja.

U mnogim se nizovima srednja vrijednost i standardna devijacija mijenjaju uglavnom usporedo, dok varijacioni koeficijent ostaje mnogo stabilniji. Znači da porast apsolutnih vrijednosti obilježja prati i usporedni porast standardne devijacije. Međutim, koeficijent je varijacije, kao što je i vidljivo iz njegove formule $\frac{s}{\bar{x}}\,100$, uslijed istovremenog povećanja brojnika i nazivnika mnogo stabilnija mjera varijacije, te ga radi te osebine uglavnom i primjenjujemo u statističkoj analizi.

Prikaz kretanja vrijednosti aritmetičke sredine, standardne devijacije i odgovarajućeg koeficijenta varijacije prikazat ćemo na podacima težine ženske simentalske teladi i junadi u razdoblju od 2 do 206 tjedana.

Grafikon 1.4

Težina, standardna devijacija i varijacioni koeficijent ženskih simentalaca u razdoblju od 2 do 206 tjedana



Grafički prikaz kretanja $\bar{x}$, s i C za 10 ženskih simentalskih grla u starosti od 2 do 206 tjedana prikazuje stalni porast $\bar{x}$, tendenciju istovremenog mijenjanja s i konstantnost C. Vrijednost varijacionog koeficijenta najveća je u prvom periodu života, što je i razumljivo obzirom na veću osjetljivost organizma u tom razdoblju na vanjske faktore.

Svaka pojava, bez obzira na vrstu i pasminu na kojoj se ispituje, ima specifične granice unutar kojih se kreće varijacioni koeficijent. Varijacioni koeficijent za tjelesne mjere npr. imat će približno iste vrijednosti kod goveda, ovaca, konja itd. Devijacije od prosječnih vrijednosti, koje se pojavljuju između raznih vrsta, između raznih pasmina ili unutar istih pasmina, naznačuju razlike za homogenost koja je nastala uslijed vanjskih faktora ili utjecajem selekcije.

Vrijednost varijacionog koeficijenta u uskoj je vezi s utjecajem vanjskih faktora na pojedino obilježje i tačnost opažanja pojava. Ukoliko je obilježje manje podložno utjecaju vanjskih faktora, utoliko će imati i manji varijacioni koeficijent. Varijacioni koeficijenat za trajanje bređosti npr. iznosi u prosjeku oko 2—5%, za tjelesne mjere 2—7%, težinu 10—15%, mliječnost 18—25% itd. Tehnika mjerenja pojave također utječe na vrijednost varijacionog koeficijenta, pa što su izmjere pojava tačnije to je i varijacioni koeficijent manji. Ukoliko dvije različite metode analiza daju iste srednje vrijednosti a različite varijacione koeficijente, sigurno je bolja ona metoda koja ima manji varijacioni koeficijent.

Vrijednost varijacionog koeficijenta izvan uobičajenih vrijednosti za ispitivanu pojavu traži posebno objašnjenje od autora, ili može izazvati sumnju u vrijednost ispitivanja (tehniku rada ili reprezentativnost uzorka).

Da bi istraživač mogao dati detaljnu analizu homogenosti, mora znati prosječnu vrijednost varijacionog koeficijenta za ispitivanu pojavu, zatim za ispitivanu pasminu i uvjete u kojima je izvršio svoja ispitivanja, kako bi mogao dati odgovarajuće tumačenje vlastitih podataka.

II. DIO

PROCJENA PARAMETARA

Prilikom ispitivanja pojedinih pojava u populaciji pomoću uzoraka, iste statističke vrijednosti od više uzoraka iz iste populacije varirat će od uzorka do uzorka, te zato one služe samo kao procjena populacije. Na taj način i vrijednosti dobivene na temelju analize svakog pojedinog uzorka služe samo za procjenu parametara koja se vrši određenim statističkim metodama.

Da bismo mogli ispravno procijeniti parametre i testirati hipoteze, neophodno je poznavati normalnu distribuciju frekvencija za velike uzorke i Studentovu »t« distribuciju za male uzorke.

Razrada matematske jednažbe za normalnu distribuciju freekvencija jedna je od najvažnijih otkrića u historiji znanosti. De Moivre (1667—1754) je razradio bazu za njenu jednadžbu i rezultate publicirao 1733. god. Na daljnjoj razradi i njenoj primjeni velikog udjela ima matematičar Gauss (1777—1855) te se ona i naziva Gaussova krivulja.

Jednadžba normalne distribucije frekvencija glasi:

$$Y = \frac{N}{\sigma \sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

Y = vrijednost ordinate koja odgovara vrijednosti apscise X
N = broj jedinica u populaciji
μ = srednja vrijednost
σ = standardna devijacija
e = baza prirodnih logaritama 2,71828
π = konstanta 3,14159
X = vrijednost numeričkog obilježja

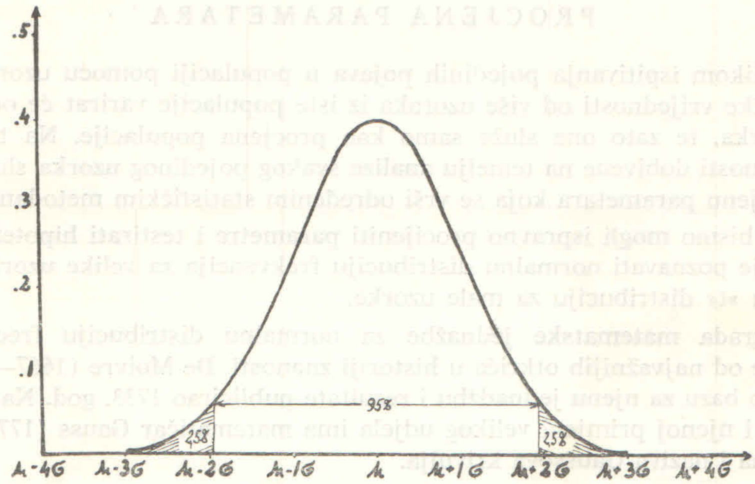
Jednadžba normalne distribucije frekvencija može se transponirati na više načina, te ćemo je prikazati u najjednostavnijem obliku, tj. kada je $N = 1$, $\sigma = 1$, a $(X - \mu) / \sigma = z$. U navedenom slučaju jednadžba normalne distribucije frekvencija glasi:

$$Y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

U grafikonu 2.1 prikazat ćemo normalnu distribuciju frekvencija kada je μ jednak nuli, σ jednaka jedinici, grafički prikaz N (0,1).

G r a f i k o n 2.1

Normalna distribucija frekvencija sa sredinom $\mu = 0$ i standardnom devijacijom $\sigma = 1$



Normalna distribucija je simetrična i potpuno određena sa dva parametra i to sa μ i σ. Aritmetička srednja vrijednost populacije (μ) nalazi se u sredini distribucije a σ je mjerilo obilježja koja su normalno distribuirana.

Unutar aritmetičke sredine populacije (μ) i stanovite udaljenosti od μ nalazi se kod normalne distribucije uvijek isti postotak ukupnih obilježja. Udaljenost pojedinog obilježja od μ (z) izražava se u jedinicama standardne devijacije koja je jednaka $z = \frac{X - \mu}{\sigma}$ (2.1). Obuhvaćeni postoci obilježja unutar naznačenih vrijednosti prikazani su u tab. 2.1: »Kumulativna normalna distribucija frekvencija«. U tabeli 2.1 pročitat ćemo npr. za udaljenost 1,96 σ od μ ($z = 1{,}96$) vrijednost 47,50%. Znači da je unutar μ i 1.96 σ obuhvaćeno 47,50% svih obilježja. Kako je normalna distribucija simetrična, to će unutar $\mu \pm 1{,}96\ \sigma$ biti obuhvaćeno 95% svih obilježja. U tabeli »Kumulativna normalna distribucija frekvencija« možemo naći za svaku udaljenost od μ izraženoj u σ jedinicama postotak obuhvaćenih obilježja. Tako se npr. unutar $\mu \pm 2{,}58\ \sigma$ nalazi 99% obilježja, a izvan $\mu \pm 3\ \sigma$ nalazi se samo 0,26% ukupnih obilježja .

Distribucije obilježja, koje se ispituju u stočarstvu, manje su ili više približne normalnoj distribuciji, te se u takvim slučajevima mogu primijeniti zakonitosti normalne distribucije.

Aritmetičke srednje vrijednosti uzoraka dobivenih iz normalno distribuirane populacije također formiraju distribuciju vrlo blizu normalnoj. Dapače, ukoliko su uzorci veći, iako su obilježja dobivena iz populacije koja nije normalno distribuirana, aritmetičke sredine također teže normalnoj distribuciji. Na taj način, kada se izračuna aritmetička srednja vrijednost uzorka

($\bar{x}$) pretpostavlja se, da se ona nalazi u okviru normalne distribucije čiju srednju vrijednost (μ) želimo procijeniti. Standardna devijacija distribucije aritmetičkih sredina uzoraka naziva se standardna greška procjene aritmetičke sredine ($\sigma_{\bar{x}}$) i izračunava se iz podataka uzorka pomoću slijedeće formule:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (2.2)$$

Prema izloženom, aritmetičke sredine uzoraka, ukoliko su izračunate iz većih uzoraka, formiraju distribuciju koja je vrlo blizu normalnoj distribuciji, te se i za njihovu distribuciju mogu primijeniti zakonitosti normalne distribucije. Tako će vrijednosti iz tabele 2.1: »Površine ispod normalne krivulje« također izražavati postotak srednjih vrijednosti uzoraka obuhvaćenih unutar intervala koji je omeđen aritmetičkom sredinom populacije (μ) i stanovitom udaljenosti od nje izraženoj u jedinicama standardne greške ($z = \frac{\bar{x} - \mu}{s_{\bar{x}}}$) (2.3). Kao što smo prije izložili, unutar $\mu \pm 1{,}96\ \sigma$ npr. nalazi se 95% obilježja, a unutar $\mu \pm 1{,}96\ \sigma_{\bar{x}}$ nalazi se 95% srednjih vrijednosti uzoraka. Unutar $\mu \pm 2{,}58\ \sigma$ nalazi se 99% obilježja, a unutar $\mu \pm 2{,}58\ \sigma_{\bar{x}}$ nalazi se 99% srednjih vrijednosti uzoraka. Od svih uzoraka dobivenih slučajnim izborom iz normalne populacije 95% će imati $z < 1{,}96$ a 99% imati će $z < 2{,}58$.

Kako smo napomenuli, zakonitosti normalne distribucije mogu se uspješno upotrijebiti prilikom procjene parametara i testiranja hipoteza, ukoliko je uzorak veći, tj. ukoliko ima više od 30 obilježja ($n > 30$). Ako su uzorci manji od 30 ($n < 30$) distribucija njihovih aritmetičkih sredina udaljavat će se od normalne distribucije i to udaljavanje bit će sve značajnije što su uzorci manji. U tom slučaju distribucija sredina uzoraka bit će također simetrična, ali za nju ne vrijede zakonitosti normalne distribucije nego se mora primijeniti »t« distribucija.

Studentovu ili »t« distribuciju razradio je W. S. Gosset (koji je pisao pod pseudonimom Student) 1908. god., a poslije ju je usavršio R. A. Fischer.

Studentova je distribucija simetrična kao i normalna distribucija. Devijacije aritmetičkih sredina uzorka od sredine populacija izražene su u jedinicama standardne greške, te se označavaju sa »t« koji iznosi: $t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$ (2.4)

Unutar intervala koji je omeđan sa μ i određene udaljenosti od μ studentova distribucija obuhvaća manji postotak ukupnih vrijednosti od normalne distribucije. Postotak vrijednosti koji je t-distribucijom obuhvaćen unutar intervala μ i stanovite udaljenosti od nje ovisi o veličini uzorka.

Na primjer u slučajevima kada je:

n	95% slučajeva obuhvaćeno je u intervalu $\mu \pm t_{.05}\, s_{\bar{x}}$	99% slučajeva obuhvaćeno je u intervalu $\mu \pm t_{.01}\, s_{\bar{x}}$
4	$\mu \pm 3{,}182\ s_{\bar{x}}$	$\mu \pm 5{,}841\ s_{\bar{x}}$
10	$\mu \pm 2{,}262\ s_{\bar{x}}$	$\mu \pm 3{,}250\ s_{\bar{x}}$
25	$\mu \pm 2{,}064\ s_{\bar{x}}$	$\mu \pm 2{,}797\ s_{\bar{x}}$
∞	$\mu \pm 1{,}96\ s_{\bar{x}}$	$\mu \pm 2{,}58\ s_{\bar{x}}$

Tabela 2.1

Površine ispod normalne krivulje
N = 10.000

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3.0	4987	4987	4987	4988	4988	4989	4989	4989	4990	4990
3.1	4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993	4993	4994	4994	4994	4994	4994	4995	4995	4995
3.3	4995	4995	4995	4996	4996	4996	4996	4996	4996	4997
3.4	4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3.6	4998	4998	4999	4999	4999	4999	4999	4999	4999	4999
3.9	4999									

Kao što se iz izloženog vidi, vrijednosti »z« i »t« kod stupnja slobode ∞ su jednake.

Vrijednosti »t« distribucije za 5% i 1% nivo signifikantnosti za razne veličine uzorka prikazane su u tabeli 2.2: »Studentova t distribucija«.

U tabeli t distribucije u predkoloni nalazi se izraz »stupanj slobode«, te kako se on često susreće u statističkim tabelama dat ćemo njegovo kratko tumačenje.

Izraz »stupanj slobode« u statistički je rječnik uveo Fischer, a upotrebljava se u više značenja koja se međusobno neznatno razlikuju. U biti stupanj slobode jednog niza varijabla jest broj vrijednosti koje se mogu označiti proizvoljno unutar specifičnosti osobina niza. Ako npr. biramo uzorak od 100 janjadi (n = 100) koji sadrži stanoviti broj muških (m) i ženskih (ž) janjadi (n = m + ž), čim proizvoljno odredimo broj ženske janjadi (ž), broj je muške janjadi fiksiran (m = n — ž). U spomenutom uzorku imamo 2 atributivna svojstva (muško, žensko) te »stupanj slobode« u tom slučaju iznosi 1, jer samo iznos jednog atributa može biti slobodno izabran. Ili npr. u slučaju da srednja vrijednost težine desetero teladi mora iznositi 100 kg, možemo slobodno izabrati devetero teladi bilo koje težine, no težina desetog teleta mora biti tačno određena. Znači da u tom slučaju »stupanj slobode« iznosi (n — 1) tj. 10 — 1.

Tabela 2.2

Studentova »t« distribucija

Stupanj slobode	P = 0,05	P = 0,01	Stupanj slobode	P = 0,05	P = 0,01
1	12,706	63,657	21	2,080	2,831
2	4,303	9,925	22	2,074	2,819
3	3,182	5,841	23	2,069	2,807
4	2,776	4,604	24	2,064	2,797
5	2,571	4,032	25	2,060	2,787
6	2,447	3,707	26	2,056	2,779
7	2,365	3,499	27	2,052	2,771
8	2,306	3,355	28	2,048	2,763
9	2,262	3,250	29	2,045	2,756
10	2,228	3,169	30	2,042	2,750
11	2,201	3,106	35	2,030	2,724
12	2,179	3,055	40	2,021	2,704
13	2,160	3,012	45	2,014	2,690
14	2,145	2,977	50	2,008	2,678
15	2,131	2,947	55	2.004	2,669
16	2,120	2,921	60	2,000	2,660
17	2,110	2,898	70	1,994	2,648
18	2,101	2,878	80	1,989	2,638
19	2,093	2,861	90	1,986	2,631
20	2,086	2,845	100	1,982	2,625
			120	1,980	2,617
			∞	1,9600	2,5758

»Stupanj slobode« prilikom očitavanja vrijednosti »t« za izračunavanje postotaka srednjih vrijednosti obuhvaćenih unutar intervala koji je omeđen μ i određene udaljenosti od μ iznosi $(n-1)$.

Prilikom različitih statističkih analiza »stupanj slobode« se mijenja, te ćemo se uvijek posebno na njega osvrnuti.

PROCJENA ARITMETIČKE SREDINE POPULACIJE POMOĆU UZORKA

Analizom uzorka koji ispitujemo, izračunamo aritmetičku srednju vrijednost $(\bar{x})$, standardnu devijaciju (s) i standardnu grešku $(s_{\bar{x}})$. Uslijed varijabilnosti obilježja očito je, da će aritmetička sredina uzorka manje ili više varirati od aritmetičke sredine ukupne mase (μ). Kod uzoraka, koji imaju više od 30 obilježja $(n > 30)$, vjerojatnost je 99% da se izračunati $\bar{x}$ nalazi unutar vrijednosti $\mu \pm 2{,}58\ s_{\bar{x}}$, a da se $\bar{x}$ nalazi unutar vrijednosti $\mu \pm 1{,}96\ s_{\bar{x}}$ vjerojatnoća je 95%. Za uzorke čiji je $n < 30$ za istu vjerojatnoću (99% ili 95%) ne mogu se upotrijebiti navedene vrijednosti 2,58 i 1,96 već vrijednosti koje se kod određenog »stupnja slobode« pročitaju u tabeli 2.2 (Studentova »t« distribucija). Međutim nama nije poznat μ već samo aritmetička srednja vrijednost uzorka, te nastojimo da na temelju $\bar{x}$ što tačnije procijenimo vrijednost aritmetičke sredine populacije.

Procjena aritmetičke sredine populacije može se vršiti pomoću intervala povjerenja, tj. određivanjem intervala u kojem se nalazi μ uz stanoviti postotak vjerojatnosti (nivo signifikantnosti). Najčešće se zadovoljavamo sa 95% intervalom povjerenja, što znači da je naša procjena ispravna u 95 slučajeva od 100. Ako je potrebna veća preciznost radimo sa 99% povjerenja, što naznačuje da je naš određeni interval u kojem se nalazi μ ispravan u 99 od 100 slučajeva.

Određivanje intervalne procjene μ na temelju uzorka u kojem je $n > 30$ vrši se pomoću slijedećeg postupka:

$$\bar{x} - z s_{\bar{x}} \leq \mu \leq \bar{x} + z s_{\bar{x}} \qquad (2.5$$

Gornji se izraz često prikazuje u skraćenom obliku:

$$\bar{x} \pm z s_{\bar{x}}$$

Određivanje intervalne procjene μ na temelju uzoraka u kojem je $n < 30$ vrši se pomoću »t« distribucije:

$$\bar{x} - t s_{\bar{x}} \leq \mu \leq \bar{x} + t s_{\bar{x}} \qquad (2\ 6)$$

$$\bar{x} \pm t s_{\bar{x}}$$

μ = aritmetička sredina populacije

$\bar{x}$ = aritmetička sredina uzorka

$s_{\bar{x}} = \frac{s}{\sqrt{n}}$ = standardna greška

z = normalizirano obilježje. Ukoliko radimo sa 95% vjerojatnosti (5% nivo signifikantnosti) z = 1,96. Kod 99% vjerojtnosti (1% nivo signifikantnosti) z = 2,58

t = vrijednost »Studentove« distribucije. Za analizirani uzorak odredi se »stupanj slobode« koji iznosi (n — 1), te se zatim u tabeli 2.2 pročita odgovarajuća vrijednost »t« distribucije već prema tome radi li se sa 5% ili 1% nivo signifikantnosti.

Procjenu aritmetičke sredine populacije pomoću uzorka prikazat ćemo na dva primjera i to na uzorku težine berkshire prasadi stare 28 dana i na uzorku mliječnosti istočno-frizijskih krava u 305 dana.

Prikaz izračunavanja 95% i 99% intervala povjerenja za uzorak težine berkshire prasadi stare 28 dana (tabela 1.2):

$$n = 10 \qquad \bar{x} = 6{,}142 \text{ kg} \qquad s = 0{,}7236 \text{ kg}$$

$$s_{\bar{x}} = s / \sqrt{n} = 0{,}7236 / \sqrt{10} = 0{,}2288 \text{ kg}$$

1. Iz tabele t-distribucije za stupanj slobode (n — 1) = 9 i nivo signifikantnosti 5% pročitamo vrijednost $t_{.05} = 2{,}262$. Za 1% nivo signifikantnosti iznosi $t_{.01} = 3{,}250$.

2. Pomnožimo: $t_{.05}\ s_{\bar{x}} = (2{,}262)\ (0{,}2288) = 0{,}5175$

$t_{.01}\ s_{\bar{x}} = (3{,}250)\ (0{,}2288) = 0{,}7436$

3. Interval povjerenja računat na 5 % nivou signifikantnosti iznosi:

$$6{,}14 - 0{,}52 = 5{,}62 \text{ kg}$$
$$6{,}14 + 0{,}52 = 6{,}66 \text{ kg}$$

Sa 95% povjerenjem možemo reći, da će izračunati interval od 5,62 kg do 6,66 kg obuhvatiti i vrijednost μ.

Interval povjerenja računat na 1 % nivou signifikantnosti iznosi:

$$6{,}14 - 0{,}74 = 5{,}40 \text{ kg}$$
$$6{,}14 + 0{,}74 = 6{,}88 \text{ kg}$$

Sa 99% povjerenjem možemo reći, da će se unutar izračunatog intervala nalaziti vrijednost aritmetičke sredine populacije tj. da je 5,40 kg $< \mu <$ 6,88 kg.

Često se interval povjerenja prikazuje u zbitijem obliku:

$$\bar{x} \pm t_{.05}\ s_{\bar{x}} = 6{,}14 \pm 0{,}52 \text{ kg}$$
$$\bar{x} \pm t_{.01}\ s_{\bar{x}} = 6{,}14 \pm 0{,}74 \text{ kg}$$

Prikaz izračunavanja 95% intervala povjerenja za uzorak mliječnosti istočno-frizijskih krava u 305 dana (podaci iz tabele 1.3):

$$n = 163 \qquad \bar{x} = 4.333{,}1288 \text{ kg} \qquad s = 802{,}544 \text{ kg}$$

$$s_{\bar{x}} = 802{,}544 / \sqrt{163} = 62{,}861 \text{ kg}$$

1. $z = 1{,}96$

2. $z\ s_{\bar{x}} = (1{,}96)\ (62{,}861) = 123{,}2075$

3. Interval povjerenja iznosi:

4.333,129 — 123,207 = 4.209,922 kg

4.333,129 + 123,207.= 4.456,336 kg

Sa 95% povjerenjem možemo kazati, da će interval od 4.210 kg do 4.456 kg obuhvatiti i vrijednost μ.

99 % interval povjerenja izračunamo na isti način samo umjesto vrijednosti z = 1,96 uzimamo vrijednost z = 2,58

1. z = 2,58

2. $z\ s_{\bar{x}} = (2{,}58)\ (62{,}861) = 162{,}1813$

3. Interval povjerenja iznosi:

4.333,129 — 162,181 = 4.170,948 kg

4.333,129 + 162,181 = 4.495,310 kg

Sa 99% povjerenjem možemo reći, da će se unutar 4.170,948 kg i 4.495,310 kg nalaziti i vrijednost μ.

PROCJENA PROPORCIJA POPULACIJE POMOĆU UZORKA

U dosadašnjem izlaganju prikazali smo neke statističke analize numeričkih obilježja, tj. obilježja koja se izražavaju brojkama (npr. mliječnost, težina), te ćemo sada dati primjenu istih statističkih analiza za atributivna obilježja, tj. za obilježja koja izražavamo riječima (npr. spol, pasmina).

Prilikom analize uzorka s numeričkim obilježjima prvo smo izračunali aritmetičku srednju vrijednost $(\bar{x})$, a kako kod atributivnih obilježja ulogu aritmetičke srednje vrijednosti ima proporcija (postotak razdijeljen na 100) to ćemo prvo prikazati izračunavanje proporcije (p) za uzorak.

Prilikom analize uzorka s atributivnim obilježljem, uvijek se može napraviti dvodioba, tj. dioba na jedinke koje imaju stanovito obilježje i sve ostale jedinke. Stado ovaca (n = 500) koje se sastoji od cigaje (27), merino (50), pramenke (225) i križanaca pramenke x merino (198), možemo podijeliti u dvije grupe: na merino ovce (50) i sve ostale (450).

Na temelju analize uzorka izračuna se proporcija »p« za jedinice koje imaju ispitivano obilježje (merino pasmina) i proporcija »q« za jedinice koje nemaju ispitivano obilježje (sve ostale pasmine).

Formule za izračunavanje proporcija p i q glase:

$$p = \frac{m}{n} \qquad (2.7)$$

$$q = \frac{n - m}{n} \quad \text{ili} \quad q = 1 - p \qquad (2.8)$$

p = proporcija jedinica koje imaju obilježje

q = proporcija jedinica koje nemaju obilježje

m = broj jedinica koje imaju obilježje

n = broj jedinica u uzorku

Za navedene podatke pasminskog sastava ovaca proporcije za merino i sve ostale pasmine iznosit će:

$n = 500$

$m = 50$

$$p = \frac{m}{n} = \frac{50}{500} = 0{,}1 \ (10\%)$$

$$q = 1 - p = 1 - 0{,}1 = 0{,}9 \ (90\%)$$

Proporcija merino ovaca iznosi 0,1 ili 10%, dok proporcija svih ostalih pasmina iznosi 0,9 ili 90%.

U mnogim slučajevima potrebno je na temelju proporcije uzorka (p) procijeniti proporciju populacije (π).

Na analogan način, kao što smo izvršili procjenu aritmetičke sredine populacije pomoću uzorka, izvršit ćemo i procjenu proporcije populacije pomoću proporcije uzorka.

Intervalnu procjenu srednje vrijednosti numeričkih obilježja populacije (μ) izračunali smo pomoću aritmetičke srednje vrijednosti uzorka ($\bar{x}$), standardne greške ($s_{\bar{x}}$), i kod većeg uzorka pomoću vrijednosti z. Prilikom intervalne procjene proporcije populacije (π), vrijednost π zauzima vrijednost μ, vrijednost s_p zauzima $s_{\bar{x}}$, a vrijednost z ostaje ista.

Proporciju uzorka (p) izračunat ćemo po formuli koja je već prikazana:

$$p = \frac{m}{n}$$

a formula za izračunavanje standardne greške procjene proporcije glasi:

$$s_p = \sqrt{\frac{p\,q}{n-1}} \quad \text{(mali uzorak)} \qquad (2.9)$$

$$s_p = \sqrt{\frac{p\,q}{n}} \quad \text{(veći uzorak)} \qquad (2.10)$$

Pomoću vrijednosti p i s_p izračuna se interval povjerenja u kojem se nalazi proporcija populacije π uz stanoviti nivo signifikantnosti i to za veći uzorak pomoću slijedećeg postupka:

$$p - z\,s_p \leq \pi \leq p + z\,s_p \qquad (2.11)$$

$$p \pm z\,s_p$$

Kao što smo razložili, kod procjene aritmetičke sredine populacije pomoću uzorka i prilikom procjene proporcije populacije pomoću uzorka, ukoliko radimo na 5% nivou signifikantnosti z = 1,96, a na 1% nivou signifikantnosti z = 2,58.

Procjenu proporcije populacije pomoću uzorka, tj. izračunavanje intervala unutar kojeg se kreće proporcija populacije uz 95% vjerojatnost, prikazat ćemo na slijedećim primjerima:

1. 200 teladi držano je u staji tipa A i do 6 mjeseci bilo je 57 oboljenja
2. 200 teladi držano je u staji tipa B i do 6 mjeseci bilo je 12 oboljenja.

Za 1. uzorak intervalnu procjenu π izvršit ćemo slijedećim redoslijedom:

1. Izračunamo proporciju p i q te standardnu grešku proporcije s_p

$$p_1 = \frac{57}{200} = 0{,}285 \ (28{,}5\%)$$

$$q_1 = 1 - p = 0{,}715 \ (71{,}5\%)$$

$$s_p = \sqrt{\frac{p_1 q_1}{n_1}} = \sqrt{\frac{(0{,}285)\ (0{,}715)}{200}}$$

$$s_p = 0{,}0319 \ (3{,}19\%)$$

2. Kako radimo uz 5% nivo signifikantnosti s_p pomnožimo sa z = 1,96

$$z\ s_p = (1{,}96)\ (0{,}0319) = 0{,}0625$$

3. Interval povjerenja uz 5% nivo signifikantnosti glasi:

$$p_1 \pm 1{,}96\ s_p = 0{,}285 \pm 0{,}0625$$

Prema tome zaključujemo, da se uz 95% vjerojatnost proporcija populacije (π) nalazi unutar granica:

$$0{,}2225 - 0{,}3475 \text{ tj. } 22{,}25\% \text{ do } 34{,}75\%$$

Znači da uz 5% nivo signifikantnosti smatramo, da bi se uz iste uvjete broj oboljele teladi ukupne populacije kretao između 22% i 35%.

Na isti način izvršit ćemo intervalnu procjenu π za 2. uzorak:

$$p_2 = \frac{12}{200} = 0{,}06 \ (6\%)$$

$$q_2 = 1 - 0{,}06 = 0{,}94 \ (94\%)$$

$$s_p = \sqrt{\frac{p_2 q_2}{n_2}} = \sqrt{\frac{(0{,}06)\ (0{,}94)}{200}}$$

$$s_p = 0{,}0167$$

$$p_2 \pm 1{,}96\ s_p = 0{,}06 \pm 0{,}0327$$

Vjerojatnost je 95% da će se proporcije populacije (π) nalaziti u intervalu 0,0273 do 0,0927 (2,73% do 9,27%) ili drugim riječima, da će uz iste uvjete postotak oboljele teladi populacije iznositi vrijednost koja se nalazi između 3% do 9 %.

III. DIO

TESTIRANJE HIPOTEZA

TESTIRANJE OPRAVDANOSTI RAZLIKA IZMEĐU DVA UZORKA

Numerička obilježja

Eksperimentalni rad može biti organiziran na taj način, da se obilježja slučajnom podjelom odijele u dvije grupe, te se zatim primijeni specifično tretiranje za svaku grupu u cilju, da se utvrde razlike između postignutih aritmetičkih sredina. Dvije grupe svinja npr. hrane se svaka sa specifičnom hranom da bi se utvrdio utjecaj pojedine hrane na dnevne priraste. Istom metodom mogu se ispitivati i svojstva pojedinih pasmina. Samo u tom slučaju razlike između grupa bit će u pasminskom sastavu, dok će tretiranje biti isto.

Uslijed uobičajene varijabilnosti obilježja možemo biti skoro potpuno sigurni, da i u slučajevima kada ne postoji razlika između dva tretiranja ili pasmina, ipak dolazi do razlika između aritmetičkih sredina dva uzorka. Znači da ćemo praktički uvijek dobiti razliku između aritmetičkih sredina dva uzorka bez obzira da li oni pripadaju istoj ili raznim populacijama.

Međutim da li se radi o uobičajenoj varijabilnosti ili je diferencija između $\bar{x}_1$ i $\bar{x}_2$ signifikantna moramo utvrditi statističkim metodama.

Ispitivanja opravdanosti između aritmetičkih sredina dva uzorka ($\bar{x}_1$ i $\bar{x}_2$) počima s postavljanjem nul hipoteze, tj. pretpostavke da nema razlike između aritmetičkih sredina populacije kojima pripadaju uzorci ($\mu_1 = \mu_2$) što možemo izraziti i na slijedeći način: $\mu_1 - \mu_2 = 0$. U tom slučaju pretpostavljamo, da su uzorci iz iste populacije, a dobivena razlika između $\bar{x}_1$ i $\bar{x}_2$ da je nastala uslijed slučajne varijabilnosti. Postavljena nul hipoteza se testira te se ili prima ili odbacuje.

Testiranje razlike između aritmetičkih sredina dva uzorka vrši se na slijedeći način:

1. za svaki uzorak izračuna se $\bar{x}$, s i $s_{\bar{x}}$;
2. na temelju izračunate standardne greške svakog uzorka ($s_{\bar{x}_1}$ i $s_{\bar{x}_2}$) izračuna se standardna greška procjene razlika između dvije sredine:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} \tag{3.1}$$

3. ako je n > od 30 treba izračunati

$$z = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} \tag{3.2}$$

ako je n < 30 izračuna se

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} \tag{3.3}$$

4. izračunati »z« ili »t« treba testirati. »z« se testira pomoću tabele 2.1: »Površine ispod normalne krivulje«. Ukoliko se računa uz 5% nivo signifikantnosti, a izračunati »z« je veći od 1,96, nul hipoteza se odbacuje tj. razlika postoji, a ukoliko je manji nul hipoteza se prima, tj. nema opravdanosti za tvrdnju da postoje signifikantne razlike između dviju aritmetičkih sredina.

Ukoliko se računa uz 1% nivo signifikantnosti a izračunati je »z« veći od 2,58 nul hipoteza se odbacuje i obrnuto.

»t« se testira pomoću tabele 2.2: »Studentova t distribucija«. Stupanj slobode = $(n_1 - 1) + (n_2 - 1)$. Kod određenog stupnja slobode pročitamo vrijednost »t« uz stanovitu signifikantnost. Ukoliko je naš izračunati »t« veći od tabelarnog, nul hipoteza se odbacuje tj. postoji razlika (uz stanovitu vjerojatnost) između uzoraka. U protivnom slučaju nemamo opravdanje za odbacivanje nul hipoteze.

Primjer testiranja razlika između aritmetičkih sredina 2 uzorka u kojima je $n > 30$ prikazat ćemo na primjeru podataka mliječnosti cigaja ovaca.

240 cigaja ovaca slučajnim izborom su podijeljene u dvije grupe na kojima se primijenila različita ishrana, te se pratio utjecaj pojedine ishrane na mliječnost. Pomoću statističke analize treba se utvrditi, da li je različita ishrana izazvala diferenciju između prosječne mliječnosti pojedinih grupa uz 5% ili 1% nivo signifikantnosti.

Tabela 3.1

Utjecaj različite ishrane na mliječnost cigaja ovce

Grupa	n	$\bar{x}$	s^2	$s_{\bar{x}}^2$
A	120	138,62	489,45	4,0787
B	120	110,50	440,55	3,6713
		Dif = 28,12		Suma = 7,7500

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} = \sqrt{7{,}7500} = 2{,}7839$$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{28{,}12}{2{,}7839} = 10{,}10 \qquad p < 0{,}01$$

Budući da je vrijednost »z« (10,10) veća ne samo od 1,96 već i od 2,58 odbacujemo nul hipotezu, te razliku između $\bar{x}_1$ i $\bar{x}_2$ ne pripisujemo slučajnoj varijabilnosti. Razlika srednjih vrijednosti grupe A i B statistički je opravdana na 1% nivou signifikantnosti.

Na podacima rezultata tova simentalskih bičića prikazat ćemo testiranje razlika aritmetičkih sredina 2 uzorka u kojima je $n < 30$.

Dva različita sistema tova ispitivali smo na dvije grupe simentalskih bičića. Na početku tova grla su bila stara 16 mjeseci, a tov je trajao 150 dana. U tabeli 3.2 prikazat ćemo testiranje opravdanosti između razlika prosječnih dnevnih prirasta dviju ispitivanih grupa.

Tabela 3.2

Utjecaj različite ishrane na prosječne dnevne priraste u tovu simentalskih bičića (dnevni prirasti izraženi u kg)

Grupa	n	$\bar{x}$	s^2	$s_{\bar{x}}^2$
A	16	1,10	0,0266	0,0017
B	13	0,99	0,0158	0,0012
		Dif = 0,11		Suma = 0,0029

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} = \sqrt{0,0029} = 0,0538$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{0,11}{0,0538} = 2,045$$

Uz stupanj slobode $(n_1 - 1) + (n_2 - 1) = 15 + 12 = 27$ u »t« distribuciji nalazimo vrijednosti $t_{.05} = 2,052$ i $t_{.01} = 2,771$. Izračunata vrijednost $t = 2,045$ nalazi se ispod vrijednosti $t_{.01}$ a vrlo je blizu t_{05}, tako da moramo biti oprezni prilikom stvaranja zaključka. U takvom će slučaju autor sa stručne strane dati svoje tumačenje te će se opredijeliti za odbacivanje ili primanje nul hipoteze uz potrebno objašnjenje.

Često se rezultati testiranja opravdanosti diferencija između dvije grupe ili više grupa izražavaju pomoću »P«. Ukoliko se nul hipoteza prima uz 5% nivo signifikantnosti rezultati se mogu prikazati izrazom $P > 0,05$, a odbacivanje nul hipoteze može se prikazati izrazom $P < 0,05$. Ukoliko se radi na 1% nivou signifikantnosti rezultati se prikazuju istom metodom samo se naznačuje, da je rađeno sa 1% nivoom signifikatnosti ($P > 0,01$, $P < 0,01$). U preciznim analizama često se naznačuje vjerojatnost i na ostalim nivoima signifikantnosti kao npr. $P = 0,04$ itd.

Pregledan je način prikaza rezultata testiranja, koji se često sreće u literaturi, pomoću zvjezdica. U tom slučaju jedna zvjezdica naznačuje, da se nul hipoteza odbacuje uz 5% nivo signifikantnosti ($P < 0,05$), ali prima se uz 1% nivo signifikantnosti ($P > 0,01$). Dvije zvjezdice naznačuju, da se nul hipoteza odbacuje na 1% nivou signifikantnosti ($P < 0,01$). Na primjer $t = 10,23^{**}$, stupanj slobode = 24 naznačava opravdanost razlike između dviju aritmetičkih sredina na 1% nivou signifikantnosti, tj. uz 99% vjerojatnost zaključujemo, da je statistički opravdana razlika između aritmetičkih sredina.

Atributivna obilježja

Eksperimentalno istraživanje atributivnih obilježja može biti organizirano u cilju da se utvrdi opravdanost između proporcija dva uzorka. Dvije se grupe teladi npr. napaja na dva različita načina. U prvoj grupi od 150 grla kod 19 su se pojavile probavne smetnje, dok je u drugoj grupi od 150 grla 10 imalo probavne smetnje. Pomoću statističkih analiza nastojimo utvrditi, da li je razlika između broja pobolijevanja u prvoj i drugoj grupi signifikantna ili se ustanovljena razlika kreće unutar uobičajenih varijacija.

Kao i kod testiranja opravdanosti razlika između dva uzorka numeričkih obilježja i ovdje će se postaviti nul hipoteza, tj. pretpostavlja se, da nema opravdane razlike između parametara ($\pi_1 = \pi_2$), i zatim će se nul hipoteza testirati.

Testiranje razlike između proporcije dva uzorka vrši se na slijedeći način:

1. za svaki uzorak izračuna se proporcija za jedinice koje imaju stanovito obilježje (p), i proporcija za jedinice koje nemaju ispitivano obilježje (q).

2. standardna greška procjene razlike proporcija glasi:

$$\sigma_{p_1 - p_2} = \sqrt{\frac{\pi_1 (1 - \pi_1)}{n_1} + \frac{\pi_2 (1 - \pi_2)}{n_2}}$$

Kako postavljamo nul hipotezu, tj. hipotezu da je $\pi_1 = \pi_2$ a $\sigma_{p_1 - p_2}$ izračunamo pomoću uzoraka, to će formula za izračunavanje standardne greške procjene razlika proporcija glasiti:

$$s_{p_1 - p_2} = \sqrt{\pi (1 - \pi) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{3.4}$$

Proporcije u populaciji nepoznate su nam, te će se pretpostaviti da je π približno jednak uravnoteženoj (vaganoj, ponderiranoj) aritmetičkoj sredini, proporcija iz dva uzorka tj.

$$\pi = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} \tag{3.5}$$

3. ako je $n > 30$ treba izračunati

$$z = \frac{p_1 - p_2}{s_{p_1 - p_2}} \tag{3.6}$$

ako je $n < 30$ treba izračunati

$$t = \frac{p_1 - p_2}{s_{p_1 - p_2}} \tag{3.7}$$

4. Izračunati »z« ili »t« testira se na isti način kao i izračunati »z« i »t« kod tretiranja opravdanosti razlika između dva uzorka numeričkih obilježja.

Testiranje razlike između proporcija 2 uzorka uz 5% nivo signifikantnosti prikazat ćemo na slijedećem primjeru:

grupa A napajana je na način »a« te je od 150 grla 19 imalo probavne smetnje;

grupa B napajana je na način »b« te je od 150 grla 10 imalo probavne smetnje

1. $p_1 = \frac{19}{150} = 0{,}1266$ (12,66%) $\quad q_1 = 1 - p_1 = 0{,}8734$ (87,34%)

$p_2 = \frac{10}{150} = 0{,}0666$ (6,66%) $\quad q_2 = 1 - p_2 = 0{,}9334$ (93,34%)

2. $$s_{p_1 - p_2} = \sqrt{\pi(1-\pi)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\pi = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{19 + 10}{150 + 150} = 0{,}0966$$

$$1 - \pi = 0{,}9034$$

$$s_{p_1 - p_2} = \sqrt{(0{,}0966)\,(0{,}9034)\left(\frac{1}{150} + \frac{1}{150}\right)} = 0{,}0340$$

3. $$z = \frac{p_1 - p_2}{s_{p_1 - p_2}} = \frac{0{,}1266 - 0{,}0666}{0{,}0340} = 1{,}76$$

Na temelju izvršenog eksperimenta i analize podataka možemo utvrditi, da ne postoji opravdana razlika između broja oboljele teladi u prvoj i drugoj grupi, jer je izračunati z (1,76) manji od kritične vrijednosti (1,96) za odbacivanje nul hipoteze uz 5% nivo signifikantnosti.

Pretpostavimo da dobiveni rezultati stručno ne zadovoljavaju i da postoji sumnja, da je do neopravdanosti razlike između proporcija došlo uslijed premalog broja grla na kojima se vršilo ispitivanje, te da se to ispitivanje ponovilo na većem broju teladi.

Prikazat ćemo rezultate ispitivanja i statističke obrade ponovljenog pokusa u kojem je veličina uzorka veća:

grupa A napajana je na način »a« i od 250 grla 31 grlo imalo je probavne smetnje,

grupa B napajana je na način »b« i od 250 grla 16 grla imalo je probavne smetnje.

1. $$p_1 = \frac{31}{250} = 0{,}1240\ (12{,}40\%) \qquad q_1 = 1 - 0{,}1240 = 0{,}8760\ (87{,}60\%).$$

$$p_2 = \frac{16}{250} = 0{,}0640\ (6{,}40\%) \qquad q_2 = 1 - 0{,}0640 = 0{,}9360\ (93{,}60\%)$$

2. $$\pi = \frac{31 + 16}{250 + 250} = 0{,}0940$$

$$1 - \pi = 0{,}906$$

$$s_{p_1 - p_2} = \sqrt{(0{,}094)\,(0{,}906)\left(\frac{1}{250} + \frac{1}{250}\right)} = 0{,}0261$$

3. $$z = \frac{0{,}1240 - 0{,}0640}{0{,}0261} = 2{,}30$$

Postotak oboljele teladi u prvom i drugom pokusu u pojedinim grupama gotovo je sasvim isti, samo što je veličina uzorka veća u drugom pokusu. Postotak oboljele teladi u prvom pokusu u grupi A iznosi 12,66%, a u drugom pokusu 12,40%. Postotak oboljele teladi u grupi B u prvom pokusu iznosi 6,66%, a u drugom pokusu 6,40%. Međutim, u drugom pokusu, uslijed većeg broja teladi na kojima se vršilo ispitivanje, dobila se opravdana razlika za proporcije grupe A i B. Znači da uz 5% nivo signifikantnosti zaključujemo, da postoji razlika između broja pobolijevanja u prvoj i drugoj grupi, što je u vezi s primjenom različite tehnike napajanja.

TESTIRANJE SIGNIFIKANTNOSTI HIPOTETIČNOG OMJERA ATRIBUTIVNIH OBILJEŽJA (Hi — kvadrat)

U stanovitim analizama atributivnih obilježja pomoću uzorka može se postaviti i hipoteza o omjeru ispitivanih atributa u populaciji. Da li se npr. cijepanja kod križanja nekih obilježja u omjeru 9:3:3:1 ili da li je spolni omjer 1:1 itd.? Eksperimentalni podaci dobiveni pomoću uzorka obično će nešto varirati od postavljenog hipotetičnog omjera te moramo utvrditi, da li su razlike odnosno odstupanja signifikantna. I u tom se slučaju postavlja nul hipoteza, tj. pretpostavlja se da nema signifikantne razlike između postavljenog hipotetičnog omjera i dobivenog omjera u ispitivanju.

Da bi protumačili način testiranja, obilježit ćemo broj individua unutar uzorka koji imaju stanoviti atribut sa f_1, f_2, f_3 . . . . fm (m = broj ispitivanih atributa), a odgovarajuće hipotetične frekvencije označavat će se sa F_1, F_2, F_3 . . . Fm.

Testiranje razlike između hipotetičnog i eksperimentom utvrđenog omjera vrši se pomoću hi - kvadrata, koji možemo prikazati slijedećom formulom:

$$\chi^2 = \frac{(f_1 - F_1)^2}{F_1} + \frac{(f_2 - F_2)^2}{F_2} + ... + \frac{(f_m - F_m)^2}{F_m}$$

U skraćenom obliku navedena formula glasi:

$$\chi^2 = \frac{\Sigma (f - F)^2}{F} \qquad (3.8)$$

Izračunavanje χ^2 prikazat ćemo na primjeru ispitivanja sekundarnog spolnog omjera kod potomaka jednog nerasta.

Od 142 potomka dobili smo 61 muško i 81 žensko prase. Postavljamo hipotezu da je spolni omjer 1:1 i da nema signifikantne razlike između istraživanjem dobivenog omjera i omjera 1:1.

χ^2 izračunamo na slijedeći način:

$f_1 = 61 \qquad F_1 = 71$

$f_2 = 81 \qquad F_2 = 71$

$$\chi^2 = \frac{(61 - 71)^2}{71} + \frac{(81 - 71)^2}{71} = 2,81$$

Testiranje nul hipoteze ili test signifikantnosti vrši se pomoću usporedbe izračunatog χ_2 i vrijednosti prikazane u tabeli 3.3: »Akumulativna distribucija hi - kvadrata« kod odgovarajućeg stupnja slobode. Stupanj slobode dobije se tako, da se od broja ispitivanih atributa odbije 1 te iznosi (m — 1).

Tabela 3.3.

Akumulativna distribucija hi-kvadrata

Stupanj slobode	P = 0,05	P = 0,01	Stupanj slobode	P = 0,05	P = 0,01
1	3,84	6,63	21	32,67	38,93
2	5,99	9,21	22	33,92	40,29
3	7,81	11,34	23	35,17	41,64
4	9,49	13,28	24	36,42	42,98
5	11,07	15,09	25	37,65	44,31
6	12,59	16,81	26	38,89	45,64
7	14,07	18,48	27	40,11	46,96
8	15,51	20,09	28	41,34	48,28
9	16,92	21,67	29	42,56	49,59
10	18,31	23,21	30	43,77	50,89
11	19,68	24,72	40	55,76	63,69
12	21,03	26,22	50	67,50	76,15
13	22,36	27,69	60	79,08	88,38
14	23,68	29,14	70	90,53	100,42
15	25,00	30,58	80	101,88	112,33
16	26,30	32,00	90	113,14	124,12
17	27,59	33,41	100	124,34	135,81
18	28,87	34,81			
19	30,14	36,19			
20	31,41	37,57			

Ukoliko je tabelarna vrijednost veća od izračunate vrijednosti χ^2, nul hipoteza se prima, dok se u protivnom slučaju odbacuje.

U našem primjeru ispitivana su dva obilježja: muško-žensko. Dakle stupanj slobode iznosi 2 — 1 = 1. U tabeli kod stupnja slobode 1 naći ćemo tabelarnu vrijednost 3,84.

Kako je tabelarna vrijednost veća od izračunatog χ_2 to nul hipotezu primamo i zaključujemo uz 5% nivo signifikantnosti, da se ne može ustvrditi da postoji signifikantna razlika između hipotetičnog omjera 1:1 i dobivenog spolnog omjera.

Izračunavanje χ^2 i testiranje hipotetičnog omjera s više od dva atributa prikazat ćemo na primjeru križanja pastuha — vrani sivac s kobilama — kulaši. U F_2 — generaciji od 112 potomaka 66 je bilo kulašastih sivaca, 18 kulaša, 25 vranih sivaca i 3 vranca. Postavlja se hipotetičan omjer tj. da se cijepanje u F_2 generaciji izvršilo u omjeru 9:3:3:1. Testiranje pomoću χ_2 izvršit ćemo na slijedeći način:

$$f_1 = 66$$
$$f_2 = 18$$
$$f_3 = 25$$
$$f_4 = 3$$

Hipotetične vrijednosti iznosit će

$$F_1 = (9/16)\,(112) = 63$$
$$F_2 = (3/16)\,(112) = 21$$
$$F_3 = (3/16)\,(112) = 21$$
$$F_4 = (1/16)\,(112) = 7$$

Uvrstimo li dobivene vrijednosti u formuli za χ^2 dobijemo da je

$$\chi^2 = \frac{(66-63)^2}{63} + \frac{(18-21)^2}{21} + \frac{(25-21)^2}{21} + \frac{(3-7)^2}{7}$$

$$\chi^2 = 3{,}616.$$

Kako su ispitivana četiri atributa, to stupanj slobode iznosi $4 - 1 = 3$. Vrijednost iz tabele 3.3: »Akumulativna distribucija χ_2« uz 5% nivo signifikantnosti iznosi 7,81. Izračunati χ^2 (3,616) manji je od tabelarne vrijednosti te zaključujemo, da razlika između omjera 9:3:3:1 i dobivenog eksperimentalnog omjera nije signifikantna.

ANALIZA VARIJANCE

Analiza varijance — jednostruka klasifikacija

U prijašnjim poglavljima prikazali smo testiranje razlika između aritmetičkih sredina dva uzorka, no kako istraživački rad često obuhvaća komparaciju 3, a i više uzoraka, to ćemo razložiti i testiranje opravdanosti razlika između više aritmetičkih sredina. I u slučaju komparacije aritmetičkih sredina više uzoraka postavit će se nul hipoteza, pa će se npr. u slučaju ispitivanja 4 uzorka postaviti hipoteza da je $\mu_1 = \mu_2 = \mu_3 = \mu_4$, a zatim se može pomoću metode analize varijance izvršiti testiranje hipoteze.

Teoretske osnove analize varijance najlakše ćemo razložiti na jednom primjeru, pa ćemo prikazati analizu varijance visine do grebena teladi stare 2 tjedna.

Četrdeset komada teladi stare dva tjedna podijeljeno je slučajnim izborom u četiri uzorka tako, da svaki uzorak sadrži 10 teladi. Telad je do starosti od 2 tjedna držana i hranjena na isti način.

U tabeli 3.4 prikazat ćemo podatke za visinu do grebena, te izračunavanje Σx^2 za svaku grupu posebno i ukupno za svu telad.

Tabela 3.4

Visina do grebena (cm) simentalske teladi stare 2 tjedna

	Grupe				Ukupno
	I	II	III	IV	
	81,5	79,0	79,0	79,0	
	76,0	77,5	78,0	76,5	
	77,5	78,5	78,0	80,0	
	78,0	83,0	77,0	76,0	
	79,0	79,0	80,0	82,0	
	78,0	77,5	77,0	78,0	
	75,0	76,5	78,0	77,0	
	78,5	74,0	77,5	80,0	
	78,0	79,0	80,0	80,0	
	77,0	78,0	76,0	75,0	
ΣX	778,5	782,0	780,5	783,5	3.124,5
$\bar{x}$	77,85	78,20	78,05	78,35	78,1125
ΣX^2	60.633,750	61,199,00	60.933,250	61.431,250	244.197,250
$(\Sigma X)^2/n$	60.606,225	61.152,40	60.918,025	61.387,225	244.062,506
Σx^2	27,525	46,600	15,225	44,025	134,744

Pomoću podataka iz tabele 3.4 može se izvršiti procjena σ^2 na tri načina:

I. izračuna se s^2 od svih obilježja kao da nisu podijeljena na 4 grupe.

$$s^2 = \frac{\Sigma x^2}{n-1} = \frac{134,744}{39} = 3,4549;$$ varijanca UKUPNO!

Vrijednost 3,455 jeste prva procjena σ^2

II. izračuna se varijanca za svaku grupu posebno, te se zatim izračuna njihova aritmetička sredina.

$$s_1^2 = \frac{27,525}{9} = 3,0583 \qquad s_2^2 = \frac{46,600}{9} = 5,1778$$

$$s_3^2 = \frac{15,225}{9} = 1,6917 \qquad s_4^2 = \frac{44,025}{9} = 4,8917$$

$$s^2 = \frac{3,0583 + 5,1778 + 1,6917 + 4,8917}{4} = 3,7048;$$ varijanca unutar grupa

Vrijednost 3,705 jeste druga procjena σ_2

III. na temelju aritmetičkih srednjih vrijednosti uzoraka
($\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, $\bar{x}_4$) izračuna se standardna greška pomoću koje se onda izračuna varijanca.

$$s_{\bar{x}}^2 = \frac{\Sigma(\bar{x} - \bar{\bar{x}})^2}{a - 1}$$

$\bar{\bar{x}}$ = aritmetička sredina uzoraka
a = broj grupa

$$s_{\bar{x}}^2 = \frac{(77,85 - 78,1125)^2 + (78,20 - 78,1125)^2 + (78,05 - 78,1125)^2 + (78,35 - 78,1125)^2}{3}$$

$s_{\bar{x}}^2 = 0,04563$

$s_{\bar{x}}^2 = \frac{s^2}{n}$ $\quad s^2 = n\, s_{\bar{x}}^2 = (10)\ (0,04563) = 0,4563$ varijanca između grupa

Treća procjena σ^2 jest vrijednost 0,4563

Dobivene podatke prikazat ćemo u slijedećoj tabeli:

Tabela 3.5.

Analiza varijance visine do grebena simentalske teladi stare 2 tjedna

Izvor varijacije	Stupanj slobode	Suma kvadrata (Σx^2)	Varijanca
ukupno	39	134,744	3,4549
između grupa	3	1,369	0,4563
unutar grupa	36	133,375	3,7048

Izračunatu varijancu na temelju podataka svih obilježja, bez obzira na diobu u grupe, unosimo u tabelu u rubriku varijanca — ukupno. Odgovarajući stupanj slobode jest broj svih obilježja smanjen za jedan (n.—1) = 39 (n. = zbroj obilježja svih grupa).

Varijancu, koja je dobivena zbrajanjem varijance pojedinih grupa i diobom sa brojem grupa, unosimo u tabelu u rubriku varijanca — unutar grupa. Stupanj slobode jest zbroj stupnjeva slobode pojedinih grupa tj. $(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + (n_4 - 1) = 36$.

Varijancu dobivenu iz izračunate standardne dreške unosimo u tabelu kod varijance između grupa a stupanj slobode iznosi broj grupa minus jedan $(a - 1) = 3$.

U tabeli zapažamo i činjenicu, da je zbroj sume kvadrata unutar grupa i između grupa jednak ukupnoj sumi kvadrata. Također je i suma stupnjeva slobode između grupa i unutar grupa jednaka ukupnom stupnju slobode.

Dioba ukupne sume kvadrata (Σx^2) i ukupnog stupnja slobode na sumu kvadrata i stupanj slobode između grupa i unutar grupa naziva se *analiza varijance.*

Navedeni način izračunavanja analize varijance iznijeli smo radi lakšeg razumijevanja teoretske osnove analize varijance no u uobičajenom izračunavanju primjenjuje se jednostavnija metoda.

Na istom primjeru prikazat ćemo izračunavanje varijance na uobičajeni način. Izračunavanje i prikaz analize varijance dat ćemo bez osvrta na već izložene račune i tabele, te ćemo opet započeti s prikazom podataka.

Tabela 3.6

Visina do grebena simentalske teladi u starosti od 2 tjedna

	Grupa			
	I	II	III	IV
	81,5	79,0	79,0	79,0
	76,0	77,5	78,0	76,5
	77,5	78,5	78,0	80,0
	78,0	83,0	77,0	76,0
	79,0	79,0	80,0	82,0
	78,0	77,5	77,0	78,0
	75,0	76,5	78,0	77,0
	78,5	74,0	77,5	80,0
	78,0	79,0	80,0	80,0
	77,0	78,0	76,0	75,0
ΣX	778,5	782,0	780,5	783,5
$\bar{x}$	77,85	78,20	78,05	78,35

Da bi se skraćenim postupkom provela analiza varijance izračuna se:

1. suma svih opažanja: $\Sigma X = 3.124{,}5$
2. korekcijski faktor: $C = (\Sigma X)^2 / a.n = 3.124{,}5^2 / 40 = 244.062{,}506$
3. ukupna suma kvadrata:

$$\Sigma X^2 - C = 81{,}5^2 + 76{,}0^2 + 77{,}5^2 \ldots + 80{,}0^2 + 75{,}0^2 - C$$
$$= 244.197{,}25 - 444.062{,}506 = 134{,}744$$

4. suma kvadrata između grupa: $\Sigma (\Sigma Xi)^2/ni - C =$

$$= (\Sigma X_1)^2/n_1 + (\Sigma X_2)^2/n_2 + (\Sigma X_3)^2/n_3 + \ldots + (\Sigma X_a)^2/n_a - C$$
$$= 778{,}5^2/10 + 782{,}0^2/10 + 780{,}5^2/10 + 783{,}5^2/10 - 244.062{,}506$$
$$= 1{,}369$$

(ΣXi = suma vrijednosti obilježja pojedinih grupa,
n_i = broj obilježja pojedinih grupa)

Tehnika izračunavanja analize varijance je ista bez obzira na broj grupa koje se kompariraju npr. 3, 4, 5 ili više grupa. U slučaju da imamo više grupa broj razlomaka u računu prikazanim pod br. 4 odgovarat će broju grupa.

Na temelju izračunate vrijednosti za »ukupnu sumu kvadrata« i »sumu kvadrata između grupa« izračuna se »suma kvadrata unutar grupa«, te odgovarajuće varijance. Vrijednosti su prikazane u tabeli br. 3.7.

Tabela 3.7.

Analiza varijance visine do grebena

Izvor varijacije	Stupanj slobode	Suma kvadrata (Σx^2)	Varijanca
ukupno	39	134,744	
između grupa	3	1,369	0,4563
unutar grupa	36	133,375	3,7048

Stupanj slobode za »Ukupno« dobije se smanjivanjem ukupnog broja obilježja za jedinicu: (n. — 1) = 40 — 1 = 39. Stupanj slobode za »Između grupa« dobije se smanjivanjem broja grupa za jedinicu: (a — 1) = (4 — 1) = 3. Stupanj slobode za »Unutar grupa« dobije se odbijanjem stupnja slobode drugog reda od prvog.

U rubriku suma kvadrata za »Ukupno« i »Između grupa« unose se vrijednosti izračunate po prikazanoj metodi u tačkama 3 i 4. Suma kvadrata za »Unutar grupa« dobije se, kao i stupanj slobode za »Unutar grupa«, odbijanjem sume kvadrata drugog reda od prvoga.

Varijance se dobiju diobom sume kvadrata s odgovarajućim stupnjem slobode.

Da li je razlika između srednjih vrijednosti grupa opravdana ili nije, utvrdit ćemo na temelju razlike varijanca »Između grupa« i »Unutar grupa«. Nul hipotezu, tj. hipotezu, da je $\mu_1 = \mu_2 = \mu_2 = \mu_4$, testirat ćemo metodom koju je razradio R. A. Fischer tj. pomoću omjera, koji se po njemu označava sa »F« i izračunava na slijedeći način:

$$F = \frac{\text{varijanca između grupa}}{\text{varijanca unutar grupa}} \tag{3.9}$$

U analizi varijance visine do grebena F iznosi:

F = 0,4563/3,7048 = 0,1232.

Da bi se utvrdilo, da li je razlika između aritmetičkih sredina grupa opravdana uz 5% nivo signifikantnosti mora se »F« (0,1232) koji smo izračunali usporediti s vrijednosti »F« iz tabele F distribucije (tabela 3.8: F distribucija). U glavi tabele nalazi se stupanj slobode f_1, a u pretkoloni stupanj slobode f_2. Stupanj slobode f_1 odgovara stupnju slobode između grupa (a — 1) a f_2 odgovara stupnju slobode unutar grupa a (n_i — 1)

Tabela 3.8.

F distribucija P = 0,05 (P = 0,01)

	f_1 stupanj slobode							
f_2	1	2	3	4	5	6	8	10
1	161 (4052)	200 (4999)	216 (5403)	225 (5625)	230 (5764)	234 (5859)	239 (5981)	242 (6056)
2	18.51 (98.49)	19.00 (99.00)	19.16 (99.17)	19.25 (99.25)	19.30 (99.30)	19.33 (99.33)	19.37 (99.36)	19.39 (99.40)
3	10.13 (34.12)	9.55 (30.82)	9.28 (29.46)	9.12 (28.71)	9.01 (28.24)	8.94 (27.91)	8.84 (27.49)	8.78 (27.23)
4	7.71 (21.20)	6.94 (18.00)	6.59 (16.69)	6.39 (15.98)	6.26 (15.52)	6.16 (15.21)	6.04 (14.80)	5.96 (14.54)
5	6.61 (16.26)	5.79 (13.27)	5.41 (12.06)	5.19 (11.39)	5.05 (10.97)	4.95 (10.67)	4.82 (10.27)	4.74 (10.05)
6	5.99 (13.74)	5.14 (10.92)	4.76 (9.78)	4.53 (9.15)	4.39 (8.75)	4.28 (8.47)	4.15 (8.10)	4.06 (7.87)
7	5.59 (12.25)	4.74 (9.55)	4.35 (8.45)	4.12 (7.85)	3.97 (7.46)	3.87 (7.19)	3.73 (6.84)	3.63 (6.62)
8	5.32 (11.26)	4.46 (8.65)	4.07 (7.59)	3.84 (7.01)	3.69 (6.63)	3.58 (6.37)	3.44 (6.03)	3.34 (5.82)
9	5.12 (10.56)	4.26 (8.02)	3.86 (6.99)	3.63 (6.42)	3.48 (6.06)	3.37 (5.80)	3.23 (5.47)	3.13 (5.26)
10	4.96 (10.04)	4.10 (7.56)	3.71 (6.55)	3.48 (5.99)	3.33 (5.64)	3.22 (5.39)	3.07 (5.06)	2.97 (4.85)
11	4.84 (9.65)	3.98 (7.20)	3.59 (6.22)	3.36 (5.67)	3.20 (5.32)	3.09 (5.07)	2.95 (4.74)	2.86 (4.54)
12	4.75 (9.33)	3.88 (6.93)	3.49 (5.95)	3.26 (5.41)	3.11 (5.06)	3.00 (4.82)	2.85 (4.50)	2.76 (4.30)
13	4.67 (9.07)	3.80 (6.70)	3.41 (5.74)	3.18 (5.20)	3.02 (4.86)	2.92 (4.62)	2.77 (4.30)	2.67 (4.10)
14	4.60 (8.86)	3.74 (6.51)	3.34 (5.56)	3.11 (5.03)	2.96 (4.69)	2.85 (4.46)	2.70 (4.14)	2.60 (3.94)
15	4.54 (8.68)	3.68 (6.36)	3.29 (5.42)	3.06 (4.89)	2.90 (4.56)	2.79 (4.32)	2.64 (4.00)	2.55 (3.80)
16	4.49 (8.53)	3.63 (6.23)	3.24 (5.29)	3.01 (4.77)	2.85 (4.44)	2.74 (4.20)	2.59 (3.89)	2.49 (3.69)
17	4.45 (8.40)	3.59 (6.11)	3.20 (5.18)	2.96 (4.67)	2.81 (4.34)	2.70 (4.10)	2.55 (3.79)	2.45 (3.59)

Tabela 3.8 — nastavak

f_2	f_1 stupanj slobode							
	1	2	3	4	5	6	8	10
18	4.41 (8.28)	3.55 (6.01)	3.16 (5.09)	2.93 (4.58)	2.77 (4.25)	2.66 (4.01)	2.51 (3.71)	2.41 (3.51)
19	4.38 (8.18)	3.52 (5.93)	3.13 (5.01)	2.90 (4.50)	2.74 (4.17)	2.63 (3.94)	2.48 (3.63)	2.38 (3.43)
20	4.35 (8.10)	3.49 (5.85)	3.10 (4.94)	2.87 (4.43)	2.71 (4.10)	2.60 (3.87)	2.45 (3.56)	2.35 (3.37)
21	4.32 (8.02)	3.47 (5.78)	3.07 (4.87)	2.84 (4.37)	2.68 (4.04)	2.57 (3.81)	2.42 (3.51)	2.32 (3.31)
22	4.30 (7.94)	3.44 (5.72)	3.05 (4.82)	2.82 (4.31)	2.66 (3.99)	2.55 (3.76)	2.40 (3.45)	2.30 (3.26)
23	4.28 (7.88)	3.42 (5.66)	3.03 (4.76)	2.80 (4.26)	2.64 (3.94)	2.53 (3.71)	2.38 (3.41)	2.28 (3.21)
24	4.26 (7.82)	3.40 (5.61)	3.01 (4.72)	2.78 (4.22)	2.62 (3.90)	2.51 (3.67)	2.36 (3.36)	2.26 (3.17)
25	4.24 (7.77)	3.38 (5.57)	2.99 (4.68)	2.76 (4.18)	2.60 (3.86)	2.49 (3.63)	2.34 (3.32)	2.24 (3.13)
26	4.22 (7.72)	3.37 (5.53)	2.98 (4.64)	2.74 (4.14)	2.59 (3.82)	2.47 (3.59)	2.32 (3.29)	2.22 (3.09)
27	4.21 (7.68)	3.35 (5.49)	2.96 (4.60)	2.73 (4.11)	2.57 (3.79)	2.46 (3.56)	2.30 (3.26)	2.20 (3.06)
28	4.20 (7.64)	3.34 (5.45)	2.95 (4.57)	2.71 (4.07)	2.56 (3.76)	2.44 (3.53)	2.29 (3.23)	2.19 (3.03)
29	4.18 (7.60)	3.33 (5.42)	2.93 (4.54)	2.70 (4.04)	2.54 (3.73)	2.43 (3.50)	2.28 (3.20)	2.18 (3.00)
30	4.17 (7.56)	3.32 (5.39)	2.92 (4.51)	2.69 (4.02)	2.53 (3.70)	2.42 (3.47)	2.27 (3.17)	2.16 (2.98)
32	4.15 (7.50)	3.30 (5.34)	2.90 (4.46)	2.67 (3.97)	2.51 (3.66)	2.40 (3.42)	2.25 (3.12)	2.14 (2.94)
34	4.13 (7.44)	3.28 (5.29)	2.88 (4.42)	2.65 (3.93)	2.49 (3.61)	2.38 (3.38)	2.23 (3.08)	2.12 (2.89)
36	4.11 (7.39)	3.26 (5.25)	2.86 (4.38)	2.63 (3.89)	2.48 (3.58)	2.36 (3.35)	2.21 (3.04)	2.10 (2.86)
38	4.10 (7.35)	3.25 (5.21)	2.85 (4.34)	2.62 (3.86)	2.46 (3.54)	2.35 (3.32)	2.19 (3.02)	2.09 (2.82)

Tabela 3.8 — nastavak

f_2	f_1 stupanj slobode 1	2	3	4	5	6	8	10
40	4.08 (7.31)	3.23 (5.18)	2.84 (4.31)	2.61 (3.83)	2.45 (3.51)	2.34 (3.29)	2.18 (2.99)	2.07 (2.80)
42	4.07 (7.27)	3.22 (5.15)	2.83 (4.29)	2.59 (3.80)	2.44 (3.49)	2.32 (3.26)	2.17 (2.96)	2.06 (2.77)
44	4.06 (7.24)	3.21 (5.12)	2.82 (4.26)	2.58 (3.78)	2.43 (3.46)	2.31 (3.24)	2.16 (2.94)	2.05 (2.75)
46	4.05 (7.21)	3.20 (5.10)	2.81 (4.24)	2.57 (3.76)	2.42 (3.44)	2.30 (3.22)	2.14 (2.92)	2.04 (2.73)
48	4.04 (7.19)	3.19 (5.08)	2.80 (4.22)	2.56 (3.74)	2.41 (3.42)	2.30 (3.20)	2.14 (2.90)	2.03 (2.71)
50	4.03 (7.17)	3.18 (5.06)	2.79 (4.20)	2.56 (3.72)	2.40 (3.41)	2.29 (3.18)	2.13 (2.88)	2.02 (2.70)
55	4.02 (7.12)	3.17 (5.01)	2.78 (4.16)	2.54 (3.68)	2.38 (3.37)	2.27 (3.15)	2.11 (2.85)	2.00 (2.66)
60	4.00 (7.08)	3.15 (4.98)	2.76 (4.13)	2.52 (3.65)	2.37 (3.34)	2.25 (3.12)	2.10 (2.82)	1.99 (2.63)
65	3.99 (7.04)	3.14 (4.95)	2.75 (4.10)	2.51 (3.62)	2.36 (3.31)	2.24 (3.09)	2.08 (2.79)	1.98 (2.61)
70	3.98 (7.01)	3.13 (4.92)	2.74 (4.08)	2.50 (3.60)	2.35 (3.29)	2.23 (3.07)	2.07 (2.77)	1.97 (2.59)
80	3.96 (6.96)	3.11 (4.88)	2.72 (4.04)	2.48 (3.56)	2.33 (3.25)	2.21 (3.04)	2.05 (2.74)	1.95 (2.55)
100	3.94 (6.90)	3.09 (4.82)	2.70 (3.98)	2.46 (3.51)	2.30 (3.20)	2.19 (2.99)	2.03 (2.69)	1.92 (2.51)
125	3.92 (6.84)	3.07 (4.78)	2.68 (3.94)	2.44 (3.47)	2.29 (3.17)	2.17 (2.95)	2.01 (2.65)	1.90 (2.47)
150	3.91 (6.81)	3.06 (4.75)	2.67 (3.91)	2.43 (3.44)	2.27 (3.14)	2.16 (2.92)	2.00 (2.62)	1.89 (2.44)
200	3.89 (6.76)	3.04 (4.71)	2.65 (3.88)	2.41 (3.41)	2.26 (3.11)	2.14 (2.90)	1.98 (2.60)	1.87 (2.41)
400	3.86 (6.70)	3.02 (4.66)	2.62 (3.83)	2.39 (3.36)	2.23 (3.06)	2.12 (2.85)	1.96 (2.55)	1.85 (2.37)
1000	3.85 (6.66)	3.00 (4.62)	2.61 (3.80)	2.38 (3.34)	2.22 (3.04)	2.10 (2.82)	1.95 (2.53)	1.84 (2.34)
∞	3.84 (6.64)	2.99 (4.60)	2.60 (3.78)	2.37 (3.32)	2.21 (3.02)	2.09 (2.80)	1.94 (2.51)	1.83 (2.32)

U prikazanom primjeru $f_1 = 3$ i $f_2 = 36$, te vrijednost iz tabele 3.8 iznosi:

$$F_{.05} = 2{,}86$$

Kako je izračunati F manji od tabelarnog, to nul hipotezu primamo i smatramo da nema signifikantne razlike između aritmetičkih sredina pojedinih grupa.

Dobiveni se rezultat morao i očekivati, jer su pojedine grupe formirane slučajnim izborom, kada je telad bila stara 2 tjedna, i do navedene dobi sva je telad bila hranjena i držana pod istim uvjetima.

Iza starosti od 2 tjedna pojedine su grupe hranjene na različiti način, te će prikaz podataka i analiza varijance u starosti od 14 tjedana dati uvid u utjecaj pojedine metode ishrane na razvoj visine do grebena.

Tabela 3.9

Visina do grebena simentalske teladi u starosti od 14 tjedana

	Grupa			
	I	II	III	IV
	98,0	95,0	94,0	90,0
	93,0	92,5	85,5	84,0
	91,5	98,0	88,0	90,0
	95,0	98,0	88,0	83,0
	96,5	96,0	88,0	86,0
	92,0	96,5	86,0	87,0
	93,5	98,0	89,5	93,0
	92,5	89,0	88,0	90,0
	92,0	96,0	90,0	90,0
	98,0	96,5	86,0	88,0
ΣX	942,0	955,5	883,0	881,0
$\bar{x}$	94,2	95,55	88,30	88,10

1. Suma svih opažanja: $\Sigma X = 3.661{,}5$

2. Korekcijski faktor: $C = (\Sigma X)^2/a\,.\,n = (3.661{,}5)^2/40 = 335.164{,}55$

3. Ukupna suma kvadrata: $\Sigma X^2 - C = 335.891{,}25 - 335.164{,}55 = 726{,}70$

4. Suma kvadrata između grupa: $\Sigma\,(\Sigma X_i)^2/n_i - C =$

$$= 942{,}0^2/10 + 955{,}5^2/10 + 883{,}0^2/10 + 881{,}0^2/10 - 335.164{,}55$$
$$= 454{,}875$$

Tabela 3.10

Analiza varijance visine do grebena simentalske teladi stare 14 tjedana

Izvor varijacije	Stupanj slobode	Suma kvadrata	Varijanca
ukupno	39	726,70	
između grupa	3	454,87	151,623
unutar grupe	36	271,83	7,5508

$F = 151{,}623/7{,}5508 = 20{,}0804$ $\quad F_{.05} = 2{,}86$

$F_{.01} = 4{,}38$

Kako je izračunati F veći od tabelarnog, to nul hipotezu odbacujemo i smatramo uz 5% i 1% nivo signifikantnosti, da je različita ishrana izvršila utjecaj na diferencijaciju u porastu visine do grebena.

Međutim, F test ne odgovara na pitanje: da li je opravdana diferencija između sredina svih grupa ili samo između nekih grupa? Da bi se odgovorilo na ovo pitanje, mora se provesti daljnje testiranje.

Testiranje opravdanosti razlika između aritmetičkih sredina više grupa može se provesti raznim metodama, a mi ćemo ovdje prikazati metodu J. S. Tukeya koju je adaptirao G. W. Snedecor.

Po navedenoj metodi izračunaju se diferencije između svih kombinacija $\bar{x}$, te ih se uspoređuje s teoretskim vrijednostima »D« i izvrši testiranje uz 5% nivo signifikantnosti:

»D« se izračunava na slijedeći način:

$$D = s_{\bar{x}}\, Q \tag{3.10}$$

Q se očitava iz tabele 3.11: vrijednosti Q za P = 0,05. Broj tretiranja jednak je broju grupa, stupanj slobode jednak je stupnju slobode unutar grupa.

$s_{\bar{x}} = s / \sqrt{n}$

s^2 = varijanca unutar grupa

n = broj obilježja u grupi

Ukoliko su diferencije između dviju aritmetičkih sredina veće od »D«, zaključuje se da je diferencija opravdana na 5% nivou signifikantnosti. U protivnom slučaju ne možemo ustvrditi da postoji signifikantna razlika.

Testiranje opravdanosti razlika između aritmetičkih sredina po Tukey — Snedecorovoj metodi prikazat ćemo na rezultatima analize varijance visine do grebena simentalske teladi stare 14 tjedana (tabela 3.9 i 3.10).

U prikazanom je primjeru broj tretiranja = 4, stupanj slobode unutar grupe = 36 te je vrijednost Q = 3,81.

$$s_{\bar{x}} = \sqrt{7{,}5508/10} = 0{,}8689$$

$$D = (3{,}81)\ (0{,}869) = 3{,}3109$$

Tabela 3.11

Vrijednost Q za P = 0,05

Stupanj slobode f	Broj tretiranja, a								
	2	3	4	5	6	7	8	9	10
1	18.0	26.7	32.8	37.2	40.5	43.1	45.4	47.3	49.1
2	6.09	8.28	9.80	10.89	11.73	12.43	13.03	13.54	13.99
3	4.50	5.88	6.83	7.51	8.04	8.47	8.85	9.18	9.46
4	3.93	5.00	5.76	6.31	6.73	7.06	7.35	7.60	7.83
5	3.61	4.54	5.18	5.64	5.99	6.28	6.52	6.74	6.93
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.35	5.59	5.80	5.99	6.15
8.	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47

U tabeli 3.12 prikazat ćemo aritmetičke sredine uzoraka i sve moguće kombinacije diferencija između srednjih vrijednosti dva uzorka. U drugu kolonu unijete su aritmetičke srednje vrijednosti uzoraka i to tačnim redoslijedom po veličini na taj način da u prvi red dođe najveća vrijednost, te zatim sve manje, dok je u zadnjem redu najmanja vrijednost. U daljnjim kolonama prikazane su diferencije između pojedinih srednjih vrijednosti.

Tabela 3.12

Diferencije između aritmetičkih sredina

Grupa	$\bar{x}$	$\bar{x}$ — 88,1	$\bar{x}$ — 88,3	$\bar{x}$ — 94,2
II	95,55	7,45	7,25	1,35
I	94,20	6,10	5,90	
III	88,30	0,20		
IV	88,10			

Od 6 diferencija 4 su veće od »D«. To su diferencije između I i III grupe, I i IV grupe, II i III grupe, II i IV grupe. Možemo zaključiti uz 5% nivo signifikantnosti, da postoji razlika između navedenih grupa. Razlika između I i II te III i IV grupe nije signifikantna.

Prilikom planiranja pokusa s više paralelnih grupa u najviše slučajeva svaka grupa ima jednaki broj obilježja. Međutim u nekim slučajevima, obzirom na pokusni plan ili uslijed neočekivanih faktora, provodimo pokus sa grupama koje imaju nejednaki broj varijabla (npr. ispitivanje porodne težine pojedinih legla krmača, slučajevi oboljenja u pojedinim grupama).

Ako imamo grupe s različitim brojem obilježja, tehnika obrade pokusa tj. analiza varijance, ista je kao što smo prikazali prilikom analize varijance sa grupama iste veličine.

Izračunavanje analize varijance za uzorke s različitim brojem obilježja prikazat ćemo na primjeru težine simentalske teladi u starosti od 14 tjedana. Četiri grupe simentalske teladi do 14 tjedana tretirane su na različite načine. Pomoću analize varijance testirat ćemo nul hipotezu, tj. pretpostavku da različita tretiranja nisu izvršila signifikantan utjecaj na prosječne težine pojedinih grupa.

Tabela 3.13

Težina simentalske teladi stare 14 tjedana

	Grupe			
	I	II	III	IV
	175	205	171	185
	162	174	163	184
	186	156	164	160
	163	170	172	207
	200	185	205	185
	177	179	159	176
	176	178	177	171
	160	174	181	183
	172	178		164
	183	162		165
		175		162
		177		177
				179
				192
ΣX	1754	2113	1392	2490
$\bar{x}$	175,40	176,08	174,00	177,85

1. $\Sigma X = 7.749$
2. $C = (\Sigma X)^2/n. = 7.749^2/44 = 1364704{,}56$
   $(n. = n_1 + n_2 + n_3 + n_4)$
3. $\Sigma X^2 - C = 1371403 - 1364704{,}56 = 6.698{,}44$
4. $\Sigma \frac{(\Sigma X_i)^2}{n_i} - C =$
   $= 1754^2/10 + 2113^2/12 + 1392^2/8 + 2490^2/14 - 1364704{,}56 = 83{,}40$

Tabela 3.14

Analiza varijance težine simentalske teladi stare 14 tjedana

Izvor varijacije	Stupanj slobode	Suma kvadrata	Varijanca
ukupno	43	6.698,44	
između grupa	3	83,40	27,800
unutar grupa	40	6.615,04	165,376

$F = 27{,}800/165{,}376 = 0{,}168 \quad F_{.05} = 2{,}84$

Na temelju izvršene analize varijance zaključujemo, da nema signifikantne razlike između aritmetičkih sredina četiri grupe teladi koja je različito tretirana do 14 tjedana starosti.

Kako smo u prikazanom primjeru analize varijance zaključili uz 5% nivo signifikatnosti, da nema opravdane razlike između pojedinih aritmetičkih sredina uzoraka, to ne moramo proširiti analizu varijance i na Tukey-Snedecorov test opravdanosti razlika između $\bar{x}$. U slučajevima odbacivanja nul hipoteze i proširenja testiranja opravdanosti razlika mora se izračunati prosječan broj obilježja u uzorku (n_0), što se može izvršiti na slijedeći način:

$$n_0 = \frac{1}{a-1}\left(n. - \frac{\Sigma n_i^2}{n.}\right) \tag{3.11}$$

n_0 = prosječan n u uzorku
a = broj grupa (uzoraka)
$n.$ = ukupan broj obilježja
n_i = broj obilježja pojedinih grupa

Prikazat ćemo npr. izračunavanje prosječnog broja obilježja za uzorak težine teladi stare 14 tjedana (tabela 3.13):

$$n_0 = \frac{1}{3}\left(44 - \frac{504}{44}\right) = 10{,}85$$

Testiranje homogenosti varijance dva uzorka

Prilikom testiranja opravdanosti razlika aritmetičkih sredina dva ili više uzoraka, varijanca je od fundamentalne važnosti. Ako izračunata varijanca uzorka (s^2) nije ispravan procjenitelj ukupne varijance (σ^2) rezultatima se smanjuje vrijednost.

Svaka bitna promjena varijance, koja je nastala uslijed tretiranja, traži specijalno tumačenje. Prilikom planiranja eksperimenta u kojemu se predviđa promjena varijance uslijed tretiranja, mora se s njom računati već prije početka pokusnog ispitivanja. Ako dođe do signifikantne promjene varijance uslijed tretiranja, a nema tumačenja promjeni varijance, procjena varijance populacije pomoću uzorka je nesigurna.

U većini slučajeva samo tretiranje ne mijenja varijancu, te se opravdanost razlika između dvije srednje vrijednosti može bez daljnjega testirati pomoću izložene metode testiranja i to pomoću vrijednosti normalne ili »t« distribucije. Međutim, u rjeđim slučajevima iza tretiranja može postojati sumnja, da je došlo do promjene varijance, te ćemo se ukratko osvrnuti na testiranje homogenosti varijance, što je u vezi i sa stvaranjem zaključaka o opravdanosti diferencije između aritmetičkih sredina dva uzorka.

Testiranje homogenosti varijanci dva uzorka, tj. ispitivanje opravdanosti razlika dviju varijanci provest ćemo pomoću F — distribucije, prikazane u tabeli 3.15, čije vrijednosti služe za testiranje na 5% nivou signifikantnosti (P = 0,05). I u ovom ćemo slučaju primijeniti nul hipotezu, tj. postavljamo hipotezu da tretiranje grupa nije proizvelo promjenu varijance.

Tabela 3.15

F distribucija

f_2 stupanj slobode za manju varijancu	f_1 stupanj slobode za veću varijancu									
	2	4	6	8	10	12	15	20	30	∞
2	39.00	39.25	39.33	39.37	39.40	39.42	39.43	39.45	39.46	39.50
3	16.04	15.10	14.74	14.54	14.42	14.34	14.25	14.17	14.08	13.90
4	10.65	9.60	9.20	8.98	8.84	8.75	8.66	8.56	8.46	8.26
5	8.43	7.39	6.98	6.76	6.62	6.52	6.43	6.33	6.23	6.02
6	7.26	6.23	5.82	5.60	5.46	5.37	5.27	5.17	5.07	4.85
7	6.54	5.52	5.12	4.90	4.76	4.67	4.57	4.47	4.36	4.14
8	6.06	5.05	4.65	4.43	4.30	4.20	4.10	4.00	3.89	3.67
9	5.71	4.72	4.32	4.10	3.96	3.87	3.77	3.67	3.56	3.33
10	5.46	4.47	4.07	3.85	3.72	3.62	3.52	3.42	3.31	3.08
12	5.10	4.12	3.73	3.51	3.37	3.28	3.18	3.07	2.96	2.72
15	4.76	3.80	3.41	3.20	3.06	2.96	2.86	2.76	2.64	2.40
20	4.46	3.51	3.13	2.91	2.77	2.68	2.57	2.46	2.35	2.09
30	4.18	3.25	2.87	2.65	2.51	2.41	2.31	2.20	2.07	1.79
60	3.93	3.01	2.63	2.41	2.27	2.17	2.06	1.94	1.82	1.48
120	3.80	2.89	2.52	2.30	2.16	2.05	1.94	1.82	1.69	1.31
∞	3.69	2.79	2.41	2.19	2.05	1.94	1.83	1.71	1.57	1.00

Metoda rada je slijedeća: za podatke koji se ispituju izračuna se slijedeći omjer:

$$F = \frac{s_1^2 \text{ (veća varijanca)}}{s_2^2 \text{ (manja varijanca)}} \qquad (3.12)$$

i testira pomoću tabele 3.15.

Stupanj slobode uzorka s većom varijancom jest $f_1 = (n_1 - 1)$, a uzorka s manjom varijancom jest $f_2 = (n_2 - 1)$. Izračunatu vrijednost F kompariramo s tabelarnom vrijednosti i ako je tabelarna vrijednost veća nul hipotezu ($\sigma_1^2 = \sigma_2^2$) primamo. U protivnom slučaju nul hipotezu odbacujemo.

Prikazat ćemo testiranje homogenosti varijance za podatke date u tabelama 3.2 i 3.1.

Varijance dobivene na temelju ispitivanja tova simentalskih bičića iznosile su:

$s_1^2 = 0{,}0266$ i $s_2^2 = 0{,}0158$
$F = 0{,}0266/0{,}0158 = 1{,}683, \quad f_1 = 15$ i $f_2 = 12$

Tabelarna vrijednost iznosi 3,18.

Prilikom ispitivanja mliječnosti cigaja ovce dobivene su slijedeće varijance:

$s_1^2 = 489{,}45$ i $s_2^2 = 440{,}55$
$F = 489{,}45/440{,}55 = 1{,}1109 \quad f_1 = 119, \quad f_2 = 119$

Tabelarna vrijednost iznosi 1,43.

U oba slučaja testiranja primamo nul hipotezu, znači da ne postoje signifikantne razlike između ispitivanih varijanca.

U slučaju kada se nul hipoteza odbacuje, tj. kada se smatra da postoji opravdana razlika između σ^2_1 i σ^2_2 mora se biti vrlo oprezan kod daljnjih testiranja. Ponovo se mora ispitati plan pokusnog rada, i nastojati, u okvirima našeg znanja, dati tumačenje razlikama između varijanca. Mora se imati na umu, da u slučaju nejednake varijance prilikom uobičajenog testiranja opravdanosti razlika između aritmetičkih sredina dva uzorka pomoću t- distribucije dolazi u većem broju slučajeva do odbacivanja nul hipoteze nego što bi se dobilo na temelju specijalnih metoda testiranja.

U slučaju da imamo stručno tumačenje za razlike u varijancama služimo se specijalnim metodama za tačnije testiranje opravdanosti razlika između aritmetičkih sredina dva uzorka. Na primjer ako je $n_1 = n_2$ a σ^2_1 nije jednaka σ^2_2 primijenit ćemo izloženu metodu izračunavanja »t«

$$t = \frac{\overline{x_1} - \overline{x_2}}{s_{\overline{x_1} - \overline{x_1}}}$$

Međutim, prilikom testiranja u tabeli t — distribucije nećemo tražiti vrijednost kod stupnja slobode 2 (n — 1) nego kod stupnja slobode (n — 1).

Ako nemamo stručnog tumačenja za neslaganje varijanca možemo upotrebiti uobičajeni t — test ali moramo biti mnogo oprezniji kod stvaranja zaključka o razlikama između aritmetičkih sredina.

Testiranje homogenosti varijance više uzoraka

Prilikom testiranja opravdanosti razlika između aritmetičkih sredina 3 i više uzoraka često će također biti potrebno provesti test homogenosti varijanca više uzoraka. Kao što smo već razložili, u većini slučajeva ne očekujemo promjenu varijance uslijed tretiranja. Ako dođe do bitne promjene između varijanca više grupa, zaključivanja moraju biti opreznija i traže opet specijalno tumačenje.

U slijedećem prikazu razložit ćemo test homogenosti varijanca po Bartlettu za uzorke s jednakim brojem obilježja. Tumačenje ćemo prikazati na primjeru, da bi praćenje metode testiranja bilo lakše.

Test se sastoji u tome, da se u odgovarajuću formulu za χ^2 uvrsti diferencija koja se dobije ako od umnoška log. prosječne varijance i broja grupa (a) odbijemo sumu log. pojedinih varijanci.

$$\chi^2 = 2{,}3026\ (n - 1)\ (a \log \overline{s}^2 - \Sigma \log s^2) \qquad (3{,}13)$$

a = broj grupa

2,3026 = konstanta ($\log_e 10$)

$\overline{s}^2$ = aritmetička sredina pojedinih s^2.

Izračunavanje homogenosti varijance prikazat ćemo na podacima iz tabele 3.6 i 3.9, tj. na visini do grebena simentalske teladi u starosti od 2 tjedna, dakle prije tretiranja i u starosti od 14 tjedana, iza tretiranja.

Tabela 3.16

Izračunavanje homogenosti varijance za uzorke jednake veličine
(Bartlettov test)
(Podaci iz tabele 3.6: Visina do grebena u starosti od 2 tjedna)

Uzorak	Suma kvadrata, Σx^2	Varijanca, s^2	$\log s^2$
1	27,525	3,0583	0,48548
2	46,600	5,1778	0,71414
3	15,225	1,6917	0,22832
4	44,025	4,8917	0,68946
		$\Sigma s^2 = 14{,}8195$	$\Sigma \log s^2 = 2{,}11740$

Sredina: $\overline{s}^2 = \Sigma s^2/a = 14.8195/4 = 3{,}7048$ $\qquad \log \overline{s}^2 = 0{,}56877$

$a \log \overline{s}^2 = 4\ (0{,}56877) = 2{,}27508$

$\Sigma \log s^2 = 2{,}11740$

Diferencija 0,15768

$$\chi^2 = 2{,}3026\,(n - 1)\,(a \cdot \log_s{}^2 - \Sigma \log s^2)$$
$$= 2{,}3026\,(9)\,(0{,}15768)$$
$$= 3{,}2677 \qquad f = a - 1 = 3$$

Izračunatu vrijednost χ^2 testiramo pomoću tabele χ^2 distribucije (tabele 3.3). Tabelarna vrijednost kod stupnja slobode 3 i uz 5% nivo signifikantnosti iznosi 7,81. U našem, dakle, slučaju možemo uz 95% vjerojatnost zaključiti, da između varijanci pojedinih grupa ne postoji signifikantna razlika.

Samo u onim slučajevima, kada izračunati χ^2 prelazi za vrlo mali iznos tabelarnu vrijednost, moramo izvršiti njegovu korekturu na slijedeći način:

$$\text{korekcijski faktor: } C = 1 + \frac{a + 1}{3a\ (n - 1)}$$

$$\text{korigirani } \chi^2 = \chi^2/C$$

Za prikazane podatke nije potrebna primjena korigiranog χ^2, te ćemo njegovo izračunavanje prikazati samo radi prikaza načina računanja.

$$C = 1 + \frac{5}{3\,(4)\,(9)} = 1{,}0463$$

$$\chi^2 = 3{,}2677/1{,}0463 = 3{,}1231$$

Tabela 3.17

Izračunavanje homogenosti varijance za visinu do grebena simentalske junadi stare 14 tjedana (Podaci iz tabele 3.9)

Uzorak	Suma kvadrata, Σx^2	Varijanca, s^2	log s^2
1	56,60	6,2888	0,79857
2	72,73	8,0805	0,90743
3	55,60	6,1777	0,79083
4	86,90	9,6555	0,98477
		$\Sigma s^2 = 30{,}2025$	$\Sigma \log s^2 = 3{,}4816$
	Sredina: $\overline{s}^2 = 30{,}2025/4 = 7{,}5506$	$\log \overline{s}^2 = 0{,}87798$	

$$a \log \overline{s}^2 = 4\,(0{,}87798) = 3{,}5119$$

$$\Sigma \log s^2 = 3{,}4816$$

Diferencija 0,0303

$$\chi^2 = 2{,}3026\,(n - 1)\,(a \log \overline{s}^2 - \Sigma \log s^2)$$

$$\chi^2 = 2{,}3026\,(9)\,(0{,}0303) = 0{,}6279 \qquad f = a - 1 = 3$$

Kako je izračunati χ^2 (0,6279) manji od tabelarne vrijednosti (7,81) možemo opet uz 5% nivo signifikantnosti zaključiti, da između ispitivanih varijanci nema opravdane razlike.

Analiza homogenosti varijance u ispitivanju utjecaja različite ishrane na visinu do grebena simentalske teladi pokazuje, da između četiri grupe teladi nije bila signifikantna razlika u homogenosti varijance ni u starosti od 2 tjedna (na početku pokusa) a ni u starosti od 14 tjedana (iza primjene različite ishrane).

Prilikom testiranja homogenosti varijance pojavljuje se mala razlika između načina testiranja homogenosti varijance za uzorke s istim i raznim brojem obilježja. Sistem rada je isti samo što je računanje χ^2 nešto komplciranije kod testiranja homogenosti varijance uzoraka s raznim brojem obilježja.

U slijedećem prikazu razložit ćemo test homogenosti varijance po Bartlettu za uzorke s nejednakim brojem obilježja i to na uzorcima težine simentalske teladi, čiji su podaci prikazani u tabeli 3.13.

Tabela 3.18

Izračunavanje homogenosti varijance za uzorke s nejednakim brojem obilježja po Bartlettovom testu (Težina simentalske teladi, podaci iz tabele 3.13)

Uzorak	Σx^2	Stupanj slobode (n — 1)	1/(n — 1)	Varijanca s^2	log s^2	(n — 1) log s^2
1	1.360,40	9	0,11111	151,15	2,17940	19,61460
2	1.580,92	11	0,09090	143,72	2,15752	23,73272
3	1.478,00	7	0,14285	211,14	2,32457	16,27199
4	2.195,72	13	0,07692	168,90	2,22763	28,95919
a = 4	$\Sigma x^2 =$ 6.615,04	$\Sigma (n-1)$ = 40	$\Sigma 1/(n-1)$ = 0,42178	$\Sigma (n-1)(\log s^2) = 88{,}57850$		

$$\bar{s}^2 = \Sigma x^2/\Sigma (n-1) = 6.615{,}04/40 = 165{,}376$$

$$(\log \bar{s}^2)\Sigma (n-1) = (2{,}21847)\ (40) = 88{,}7388$$

$$\chi^2 = 2{,}3026\,[(\log \bar{s}^2)\,\Sigma (n-1) - \Sigma (n-1)(\log s^2)]$$

$$= 2{,}3026\ (88{,}7388 - 88{,}57850) \qquad (3.14)$$

$$= 0{,}3691$$

$$f = a - 1 = 3$$

korekcijski faktor:

$$C = 1 + \frac{1}{3(a-1)}\left(\Sigma \frac{1}{n-1} - \frac{1}{\Sigma (n-1)}\right)$$

$$= 1 + \frac{1}{(3)\ (3)}\left(0{,}42178 - \frac{1}{40}\right) = 1{,}04408$$

korigirani $\chi^2 = 0{,}3691/1{,}04408 = 0{,}3535$

Iz tabele 3.3 akumulativne distribucije χ^2 kod stupnja slobode = 3 dobivamo vrijednost 7,81 sa 95% vjerojatnošću, te kako je izračunati χ^2 iz navedenog primjera niži, to smatramo da nema signifikantne razlike između varijanci pojedinih grupa. Korigiranu vrijednost χ^2 računali smo samo radi prikaza metode računanja, jer u prikazanom primjeru χ^2 leži daleko ispod kritične vrijednosti 7,81 te nije potrebno izračunavanje korigiranog χ^2.

Ukoliko se utvrdi, da je razlika između varijanci opravdana, mora stručnjak dati tumačenje te pojave. Često će nam test homogenosti biti itekako interesantan, može biti da je u nekim slučajevima interesantniji od opravdanosti između pojedinih aritmetičkih sredina.

Ukoliko netko želi, unatoč opravdane razlike između varijanca pojedinih grupa, da testira i opravdanost između aritmetičkih srednjih vrijednosti, onda ne zadovoljava izložena metoda računanja analize varijance nego se mora upotrijebiti specijalna metoda.

U primjerima analize varijance izloženima u poglavlju »Analiza varijance — jednostruka klasifikacija« izvršena je razdioba obilježja u pojedine uzorke slučajnim izborom, te se tokom tretiranja i testiranja podataka svaki uzorak promatrao kao cjelina. U stanovitim ispitivanjima može se povećati preciznost rezultata, kao i broj faktora koji se ispituju, ako se primijeni hijerarhijska klasifikacija. U tom se slučaju svaki uzorak sastoji od nekoliko podgrupa u kojima obilježja imaju stanovite zajedničke osebine po kojima su i grupirani. Svaka podgrupa može opet da se sastoji od pojedinih podgrupa itd. Dioba obilježja u pojedine uzorke i podgrupe vrši se slučajnim izborom.

Prilikom analize varijance kod hijerarhijske klasifikacije povećava se broj pojedinih komponenata prilikom diobe ukupne sume kvadrata. Kod jednostruke klasifikacije ukupna suma kvadrata dijeli se na dvije komponente (suma kvadrata između grupa i unutar grupa), dok će se kod hijerarhijske klasifikacije ukupna suma kvadrata dijeliti na više komponenata, čiji će broj ovisiti o pokusnom planu. Kroz to nastaje mogućnost šire primjene analize varijance, te se pomoću nje obuhvaća rješavanje više problema unutar eksperimentalnog rada.

Hijerarhijska klasifikacija primjenjuje se dosta često u eksperimentalnom radu, te je prikladna metoda i u genetskim ispitivanjima prilikom procjene heritabiliteta pojedinih osobina.

Primjer analize varijance za uzorke s hijerarhijskom klasifikacijom prikazat ćemo na podacima ispitivanja dnevnih prirasta potomaka 4 nerasta.

U ispitivanju je obuhvaćeno 12 krmača, te su se svakom nerastu pripustile 3 krmače, koje su opredijeljene za pojedine neraste slučajnim izborom. Iz svakog legla slučajnim su izborom izabrana 3 potomka na kojima su ispitivani prosječni dnevni prirasti. U slijedećoj tabeli prikazat ćemo rezultate ispitivanja.

Tabela 3.19

Prosječni dnevni prirasti za svinje (3 svinje iz istog legla)

Otac, i i = 1 . . . a	Majka, i j j = 1 . . b	Potomci, Xijk			X_{ij}	X_i	X
1	1	1,46	1,37	1,62	4,45		
	2	1,55	1,35	1,29	4,19		
	3	1,42	1,39	1,65	4,46	13,10	
2	1	1,33	1,13	1,11	3,57		
	2	1,32	1,60	1,40	4,32		
	3	1,30	1,42	1,31	3,93	11,82	
3	1	1,46	1,17	1,25	3,88		
	2	1,30	1,38	1,51	4,19		
	3	1,23	1,46	1,31	4,00	12,07	
4	1	1,25	1,16	1,43	3,84		
	2	1,06	1,24	1,02	3,32		
	3	1,08	1,03	1,28	3,39	10,55	47,54

Ukupan broj potomaka abn = (4) ((3) (3) = 36 potomaka
a = broj očeva, b = broj majki s istim ocem, n = broj potomaka iste majke.

Pomoću analize varijance nastojat ćemo podijeliti sumu kvadrata prema izvoru variranja, tj. obzirom na očeve, majke — isti očevi, i djeca iste majke.

$C = (\Sigma X)^2/abn = 47{,}54^2/36 = 62{,}7792$

Potomci: $\Sigma X^2ijk - C = (1{,}46^2 + 1{,}55^2 + \ldots + 1{,}28^2) - 62{,}7792 = 0{,}9430$

majke: $\Sigma X^2ij/n - C = (4{,}45^2 + 4{,}19^2 + \ldots + 3{,}39^2)/3 - 62{,}7792 = 0{,}5451$

očevi: $\Sigma X^2i/bn - C = (13{,}10^2 + 11{,}82^2 + 12{,}07^2 + 10{,}55^2)/9 - 62{,}7792 = 0{,}3663$

majke — isti očevi = majke — očevi = 0,5451 — 0,3663 = 0,1788

potomci istih majki = potomci — majke = 0,9430 — 0,5451 = 0,3979

Izračunavanje pojedinih suma kvadrata vrši se prema prije izloženom obrascu uz malo prošireni postupak. Za prikazani primjer u tabeli 3.19 izračunaju se i dvije ukupne sume kvadrata i to suma kvadrata za majke i suma kvadrata za potomke. Suma kvadrata za majke — isti očevi dobijemo odbijanjem sume kvadrata očeva od sume kvadrata majki. Sumu kvadrata za potomke istih majki dobijemo odbijanjem sume kvadrata majki od sume kvadrata potomaka.

U tabeli 3.20 prikazat ćemo izračunate vrijednosti varijance:

Tabela 3.20

Analiza varijance dnevnih prirasta svinja

Izvor varijacije	Stupanj slobode	Suma kvadrata	Varijanca
očevi	3	0,3663	0,1221
majke — isti očevi	8	0,1788	0,0223
potomci istih majki	24	0,3979	0,0165
ukupno	35	0,9430	

Varijanca »majke — isti očevi« sadrži u sebi i komponentu varijance »potomci istih majki« a varijanca »očevi« sadrži u sebi komponentu varijance »potomci istih majki« i »majke — isti očevi«. Prikaz sastavnih dijelova varijanca prikazat ćemo u tabeli 3.21. Slovima »A« označavat ćemo vrijednosti koje se odnose na očeve a slovima »B« označavat ćemo vrijednosti koje se odnose na majke.

Tabela 3.21

Upotpunjena analiza varijance prosječnih dnevnih prirasta svinja (podaci iz tabele 3.19 i 3.20)

Izvor varijacije	Stupanj slobode	Varijanca	Procijenjeni parametri
očevi	3	0,1221	$\sigma^2 + n\sigma^2_B + bn\,\sigma^2_A$
majke — isti očevi	8	0,0223	$\sigma^2 + n\sigma^2_B$
potomci istih majki	24	0,0165	σ^2

$n = 3$, $b = 3$, $a = 4$, $s^2 = 0{,}0165$ procjenjuje σ^2, $s^2_B = (0{,}0233 - 0{,}0165)/3$

$s^2_B = 0{,}0019$ procjenjuje σ^2_B, $s^2_A = (0{,}1221 - 0{,}0223)/9 = 0{,}0111$ procjenjuje σ^2_A

Postavit ćemo opet nul hipotezu obzirom na očeve ($\sigma^2_A = 0$) i majke ($\sigma^2_B = 0$) koja se može testirati na slijedeći način:

1. testiranje nul hipoteze za očeve $\sigma^2_A = 0$:

$$F = \frac{\text{varijanca »očevi«}}{\text{varijanca »majke — isti očevi«}} \quad \text{procjenjuje}$$

$$\frac{\sigma^2 + n\sigma^2_B + bn\sigma^2_A}{\sigma^2 + n\sigma^2_B}$$

$$F = \frac{0,1221}{0,0223} = 5,47 \qquad f = 3,8 \qquad F_{.05} = 4,07$$

2. testiranje nul hipoteze za majke $\sigma^2_B = 0$:

$$F = \frac{\text{varijanca »majke — isti očevi«}}{\text{varijanca »potomci istih majki«}} \quad \text{procjenjuje}$$

$$\frac{\sigma^2 + n\sigma^2_B}{\sigma^2}$$

$$F = \frac{0,0223}{0,0165} = 1,35 \qquad f = 8,24 \qquad F_{.05} = 2,36$$

Na temelju izvršene analize varijance uz 5% nivo signifikantnosti zaključujemo, da postoji razlika između nerastova obzirom na priraste potomaka, dok za majke nije utvrđena signifikantna diferencija.

Analiza varijance — dvostruka klasifikacija

Prilikom planiranja eksperimentalnog rada često su poznate neke važne osobine pokusnog materijala. Kad se u pokus npr. uzimaju praščići iz 4 legla poznato je koji praščići imaju iste roditelje ili kada se pokusni materijal sastoji od 4 pasmine kokoši, a želi se utvrditi utjecaj različitog držanja na nesivost, poznato nam je kojoj pasmini pripada pojedina kokoš. U takvim slučajevima prilikom planiranja pokusa možemo na temelju poznavanja pokusnog materijala organizirati pokusni rad na taj način, da se poveća njegova vrijednost time što ćemo analizom varijance — dvostrukom klasifikacijom izlučiti dodatni faktor varijacije čiji nam je uzrok poznat (npr. razni roditelji, pasmine).

U slučajevima, kao što smo napomenuli prilikom formiranja pojedinih uzoraka svako se grlo dodjeljuje pojedinoj grupi obzirom na dva kriterija. Prvo se uzima u obzir način tretiranja, koji se ispituje (npr. razna ishrana, razno držanje), a zatim osobine grla, koje su nam poznate i koje želimo izlučiti iz analize podataka tretiranja (npr. roditelji, pasmina). Na taj način u svakom će uzorku biti zastupan isti broj obilježja s poznatim osobinama. Dioba grla za pojedine uzorke vrši se slučajnim izborom.

U slijedećem primjeru prikazat ćemo primjenu analize varijance u obradi rezultata postavljenih na opisani način.

Plan se pokusa sastojao u tome, da se na 20 grla simetalskih krava primijeni pet različitih načina ishrane u cilju utvrđivanja njihovog djelovanja na prosječnu mliječnost u 305 dana. Kako su krave prije telenja različito tretirane u suhostajnom periodu, što znamo da utječe na iduću laktaciju, izbor se pojedinih grla u grupe vršio obzirom na ishranu kojoj će biti podvrgnute i obzirom na tretiranje u suhostaju. Pojedina grla s istim tretiranjem u suhostaju dodijeljena su slučajnim zborom u grupe za specifičnu ishranu, te svaka grupa sadrži isti broj grla koja su tretirana na isti način u suhostaju.

U slijedećoj tabeli prikazat ćemo dobivene rezultate za prosječnu mliječnost u 305 dana za pojedina grla. Različite načine ishrane označili smo slovima A, B, C, D i E, a različito tretiranje u suhostaju s brojevima I, II, III i IV.

Tabela 3.22

Prosječna dnevna mliječnost u 305 dana u kg

Način ishrane	Tretiranje u suhostaju I	II	III	IV	ΣX	$\bar{x}$
A	8,3	10,5	10,2	12,3	41,3	10,33
B	6,1	10,1	9,0	10,8	36,0	9,00
C	7,8	10,6	11,5	12,5	42,4	10,60
D	8,2	14,1	11,3	9,3	42,9	10,73
E	12,0	9,6	14,2	15,1	50,9	12,73
ΣX	42,4	54,9	56,2	60,0	213,5	
$\bar{x}$	8,48	10,98	11,24	12,0		

korekcijski faktor: $C = (\Sigma X)^2/n = 213{,}5^2/20 = 2.279{,}1125$

ukupno: $\Sigma X^2 - C = 8{,}3^2 + 6{,}1^2 + 7{,}8^2 + \ldots + 15{,}1^2 - 2.279{,}1125 = 99{,}9575$

suhostaj: $\dfrac{42{,}4^2 + 54{,}9^2 + 56{,}2^2 + 60{,}0^2}{5} - C = 34{,}9295$

ishrana: $\dfrac{41{,}3^2 + 36{,}0^2 + 42{,}4^2 + 42{,}9^2 + 50{,}9^2}{4} - C = 28{,}5547$

ostatak = greška = $99{,}9575 - (34{,}9295 + 28{,}5547) = 36{,}4733$

Izvor varijacije	Stupanj slobode	Suma kvadrata	Varijanca
suhostaj	3	34,929	11,643
ishrana	4	28,555	7,138
ostatak — greška	12	36,473	3,039
ukupno	19	99,957	

Temeljna je razlika između analiza varijance prikazanih u tabelama 3.7 i 3.22 u diobi ukupne sume kvadrata. Dok je ukupna suma kvadrata u primjeru analize varijance kod visine do grebena (tabela 3.7) dijeljena u dva dijela u primjeru prosječne dnevne mliječnosti u 305 dana, tabela 3.22 — dijeljena je u 3 dijela.

U izloženom primjeru u tabeli 3.22 obuhvaćeni su faktori suhostaj i različita ishrana. Račun se vrši na već izloženi način, jedino je izuzetak s ostatkom-greška. Suma kvadrata za »grešku« dobije se odbijanjem zbroja sume kvadrata za suhostaj i ishranu od ukupne sume kvadrata. Stupanj slobode se dobije na isti način, ali ujedno je on jednak i umnošku stupnja slobode dva ispitivana faktora (ishrana i suhostaj).

Glavna je svrha navedene diobe da se izračuna sumu kvadrata za »ostatak-grešku«, tj. varijancu »ostatka« koja se naziva i procjena eksperimentalne greške, jer time dobivamo osnovu za testiranje opravdanosti razlika između oba faktora, koji su bili temelj za diobu obilježja. Kako dozvoljavamo mogućnost da oba faktora mogu utjecati na rezultate, to niti varijanca ishrane niti suhostaja nije procjenitelj varijance populacije (σ^2). Na varijancu »greške« ishrana i suhostaj ne djeluju, te je ona procjenitelj varijance populacije i naziva se »eksperimentalna greška«.

Dioba ukupne sume kvadrata obzirom na dva faktora povećava preciznost eksperimentalnog rada.

Testiranje signifikantnosti suhostaja i ishrane izvršit ćemo opet pomoću omjera F. U oba slučaja u nazivnik će doći varijanca greške, dok će u brojnik doći varijanca za suhostaj ili za ishranu.

Za izloženi primjer iznositi će F za suhostaj:

$F = 11{,}643/3{,}039 = 3{,}83 \qquad f_1 = 3,\ f_2 = 12,\ F_{.05} = 3{,}49$

Za ishranu vrijednost F iznosi:

$F = 7{,}138/3{,}039 = 2{,}35 \qquad f_1 = 4,\ f_2 = 12,\ F_{.05} = 3{,}26$

Na temelju izvršene analize varijance možemo uz 5% nivo signifikantnosti ustvrditi, da je različito tretiranje u suhostaju izvršilo signifikantan utjecaj na iduću laktaciju. Međutim, različiti načini ishrane primijenjeni u ispitivanom slučaju nisu izvršili utjecaj koji je statistički opravdan.

Očito je, da je preciznost analize učinka različite ishrane povećana eliminiranjem utjecaja suhostaja od eksperimentalne greške.

Prilikom planiranog uspoređivanja, kao što je prikazano u tabeli 3.22., važno je da se iza izračunavanja vrijednosti F izvrši i testiranje između pojedinih aritmetičkih sredina, bez obzira na rezultate dobivene komparacijom izračunatog F s tabelarnom vrijednosti $F_{.05}$ ili $F_{.01}$. U nekim slučajevima mogu se dobiti opravdane razlike dviju ili više diferencija usprkos tome što F test nije pokazao signifikantnost.

Testiranje razlika između pojedinih vrijednosti aritmetičkih sredina izvršit ćemo na isti način kao što smo već razložili, tj. pomoću testa Tukeya koji je adaptirao Snedecor.

Testiranjem opravdanosti razlika između aritmetičkih sredina pojedinih grupa različito hranjenih dobit ćemo slijedeće vrijednosti:

$D = Q s_{\bar{x}}$ $\qquad s_{\bar{x}} = \sqrt{3,039/4} = 0,8716$ $Q = 4,51$, $a = 5$, $f = 12$

$D = (4,51)\ (0,8716) = 3,93$

Ishrana	$\bar{x}$	$\bar{x}$ — 9,00	$\bar{x}$ — 10,33	$\bar{x}$ — 10,60	$\bar{x}$ — 10,73
E	12,73	3,73	2,40	2,13	2,00
D	10,73	1,73	0,40	0,13	
C	10,60	1,60	0,27		
A	10,33	1,33			
B	9,00				

Dobiveni rezultati se slažu s prije izvršenim F testom. Nijedna diferencija između dviju srednjih vrijednosti dobivenih različitom ishranom ne prelazi kritičnu vrijednost 3,93.

Na temelju podataka utjecaja suhostaja izračunata vrijednost D iznosit će:

$D = Q s_{\bar{x}}$ $\qquad s_{\bar{x}} = \sqrt{3,039/5} = 0,7796$ $Q = 4,20$, $a = 4$; $f = 12$

$D = (4,20)\ (0,7796) = 3,27$

Tretiranje	$\bar{x}$	$\bar{x}$ — 8,48	$\bar{x}$ — 10,98	$\bar{x}$ — 11,24
IV	12,00	3,52	1,02	0,76
III	11,24	2,76	0,26	
II	10,98	2,50		
I	8,48			

Od izračunatog D (3,27) veća je jedino diferencija između srednjih vrijednosti I i IV grupe. Za navedenu diferenciju možemo uz 95% vjerojatnost zaključiti, da je signifikantna. Razlike između I-II, I-III, II-III, II-IV i III-IV grupe manje su od izračunatog »D« i nisu statistički opravdane.

Osim navedenih primjera analize varijance postoje i druge, no njihov bi prikaz prelazio granice ovog izlaganja.

1°

IV. DIO

REGRESIJA I KORELACIJA

Znatan dio istraživanja i analiza zauzima ispitivanje veza između pojava, kao i utjecaja jedne pojave na drugu. Između pojava postoji veza, ukoliko porast jedne pojave prati porast druge pojave — pozitivna veza, ili porast jedne pojave prati pad druge pojave — negativna veza. Veza među pojavama može biti različite jakosti, kao što može biti različit i oblik veze (linearni, krivolinijski). Te veze u stočarstvu nikad nemaju značaj matematskih funkcija nego su one stohastičkog karaktera.

Ispitivanje veze između pojava vrši se metodama koje su u međusobnoj uskoj vezi i to regresijskim analizama i korelacijama.

Cilj je regresijske analize, da pomoću izračunatog koeficijenta regresije i jednadžbe regresije utvrdi karakter veze između dvije ispitivane pojave, koje se međusobno nalaze u stohastičkoj vezi, te kada promjene vrijednosti jednog obilježja (neovisnog X) prati promjena drugog obilježja (ovisnog Y). Ako, dakle, veza ili ovisnost postoji, može se pomoću regresijske analize na temelju obilježja X predvidjeti obilježje Y. Pomoću regresijske jednadžbe dobiva se zapravo dinamička (pokretna) srednja vrijednost, te se mogu izračunati devijacije pojedinih eksperimentalnih vrijednosti od regresijskih vrijednosti.

Pomoću računa koleracije potvrđujemo jakost i smjer veze između obilježja, koje mogu ali ne moraju imati međusobni utjecaj. Visine braće i sestara, npr. u međusobnoj su vezi, ali visina jednih ne utječe na visinu drugih.

Izračunate regresijske vrijednosti moraju reprezentirati kretanje ispitivanih pojava te se prilikom analiza veze između dva obilježja prvo mora utvrditi smjer kretanja, tj. obliz veze. Kretanje obilježja može biti linearno, paraboličnog oblika drugog stupnja, trećeg stupnja, te u obliku eksponencijalnih krivulja ili rjeđe kakvog ostalog oblika.

U nastavku izlaganja regresija razložit ćemo regresiju linearnog i paraboličnog oblika drugog stupnja.

REGRESIJA

Linearna regresija

Ispitivanje veza između ovisnog obilježja Y i neovisnog X, ukoliko kretanje obilježja ima oblik pravca, izvršit će se pomoću linearne regresije koja se može prikazati jednadžbom pravca:

$$Y = a + bX$$

U jednadžbi linearne regresije koeficijent »a« naznačuje vrijednost ovisnog obilježja, kada je vrijednost neovisnog obilježja jednaka nuli, a koeficijent »b« naznačuje nagib pravca, tj. vrijednost za koju se promijeni obilježje Y, ako se obilježje X promijeni za jedinicu. Koeficijent »b« naziva se koeficijent regresije. Koeficijent regresije (b) može imati pozitivni ili negativni predznak, što pokazuje da je veza pozitivna ili negativna. U grafičkom prikazu ovisit će o predznaku koeficijenta regresije da li će pravac biti uzlazni ili silazni.

Izračunavanje linearne regresije, tj. izračunavanje koeficijenta a i b za jednadžbu pravca, prikazat ćemo na podacima visine do grebena desetero simentalske teladi u starosti od 2—26 tjedana.

Tabela 4.1

Regresija visine do grebena na starost

Starost u tjed.	Vis. grebena	Devijacija od sredine		Kvadrat devijacije		Produkt devijacije	Procjena	Devijacija od regresije	Kvadrat
X	Y	x	y	x^2	y^2	xy	$\hat{Y}$	$Y-\hat{Y}=d_{YX}$	$(Y-\hat{Y})^2=d^2_{YX}$
2	78	-12	-11,69	144	136,66	140,28	77,30	0,70	0,4900
4	80	-10	- 9,69	100	93,90	96,90	79,36	0,64	0,4096
6	81	- 8	- 8,69	64	75,52	69,52	81,43	-0,43	0,1849
8	83	- 6	- 6,69	36	44,76	40,14	83,49	-0,49	0,2401
10	86	- 4	- 3,69	16	13,62	14,76	85,56	0,44	0,1936
12	87	- 2	- 2,69	4	7,24	5,38	87,63	-0,63	0,3969
14	88	0	- 1,69	0	2,86	0	89,69	-1,69	2,8561
16	92	2	2,31	4	5,34	4,62	91,76	0,24	0,0576
18	95	4	5,31	16	28,20	21,24	93,82	1,18	1,3924
20	96	6	6,31	36	39,82	37,86	95,89	0,11	0,0121
22	97	8	7,31	64	53,44	58,48	97,96	-0,96	0,9216
24	100	10	10,31	100	106,30	103,10	100,02	-0,02	0,0004
26	103	12	13,31	144	177,16	159,72	102,09	0,91	0,8281
$\Sigma X = 182$	$\Sigma Y = 1166$	$\Sigma x = 0$	$\Sigma y = 0,03$	$\Sigma x^2 = 728$	$\Sigma y^2 = 784,82$	$\Sigma xy = 752,00$	$\Sigma \hat{Y} = 1166,00$	$\Sigma d_{YX} = 0,00$	$\Sigma d_{YX}^2 = 7,9834$
$\bar{x} = 14$	$\bar{y} = 89,6923$								

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{752}{728} = 1,0329 \text{ cm} \tag{4.1}$$

Koeficijenat regresije u tom je slučaju pozitivan, što pokazuje da s većom starosti (neovisno obilježje X) dolazi do povećanja visine do grebena (ovisno obilježje Y). Vrijednost koeficijenta regresije b = 1,0329 cm naznačuje, da se u prosjeku za svaki tjedan poveća visina do grebena za 1,0329 cm.

Jednadžba linearne regresije izračuna se na slijedeći način:

$$\hat{Y} - \bar{y} = b\,(X - \bar{x})$$
$$\hat{Y} = \bar{y} + b\,(X - \bar{x}) \tag{4. 2}$$

U primjeru regresije visine do grebena na starost $\bar{x}$ iznosi 14 tjedana, $\bar{y}$ = 89,6923 cm, te će jednadžba linearne regresije glasiti:

$$\hat{Y} - 89,6923 = 1,0329\,(X - 14)$$
$$\hat{Y} = 75,2317 + 1,0329\,X$$

Na temelju jednadžbe regresije može se za svaki X izračunati Y, tj. teoretske vrijednosti odnosno teoretska srednja visina do grebena u svakoj određenoj dobi. U tabeli 4. 1 prikazane su regresijske vrijednosti ($\hat{Y}$) za visinu do grebena u pojedinoj starosti (dakle pokretne srednje vrijednosti).

Prilikom procjene pojedinog obilježja nećemo njegovu vrijednost uspoređivati s aritmetičkom srednjom vrijednosti nego s očekivanom vrijednosti za određeni X, tj. s regresijskom vrijednosti. Ako bi vrijednosti za visinu do grebena u pojedinoj starosti uspoređivali s aritmetičkom sredinom, dobili bi da visina do grebena u svim tjednima do starosti od 14 tjedana iznosi manje od aritmetičke sredine, dok visina do grebena u svim tjednima iza 14 tjedana premašuje aritmetičku sredinu. Kako nikako ne možemo kazati da sve visine do grebena do 14 tjedana ne zadovoljavaju, dok iza 14 tjedana da prelaze očekivanu vrijednost, to je očito da eksperimentalne vrijednosti ne možemo uspoređivati sa $\bar{y}$ već s izračunatim regresijskim vrijednostima. Visinu do grebena npr. telića starih 18 tjedana, koja iznosi 95 cm, uspoređujemo s teoretski izračunatom vrijednosti, tj. sa $\hat{Y}$ = 93,82 cm, a ne sa sredinom koja iznosi 89,69 cm.

Isti je slučaj i prilikom međusobnog uspoređivanja dva obilježja, u konkretnom slučaju prilikom uspoređivanja visine do grebena u dva različita tjedna. U slučaju komparacije visine do grebena u starosti od 4 tjedna (80 cm) s visinom do grebena u dobi od 22 tjedna (97 cm), ukoliko bi međusobno uspoređivali njihove apsolutne vrijednosti, razlika bi iznosila 97 — 80 = 17 cm. Nikako ne bi bilo opravdano, kada bi na temelju dobivenih diferencija zaključili, da je postignuta visina do grebena u starosti od 22 tjedna u usporedbi s visinom do grebena u starosti od 4 tjedna pozitivnija i to za 17 cm. Ispravnu komparaciju dobijemo uspoređivanjem devijacija pojedinih visina do grebena od njihovih regresijskih vrijednosti.

U starosti od 4 tjedna visina do grebena iznosi 80 cm, regresijska je vrijednost 79,36 cm te devijacija iznosi +0,64 cm. U starosti od 22 tjedna visina do grebena iznosi 97 cm, regresijska je vrijednost 97,96 cm te devijacija iznosi —0,96 cm. Na temelju izračunatih devijacija možemo zaključiti, da je postignuta visina do grebena u starosti od 4 tjedna pozitivnija i to za razliku između —0,96 i 0,64, tj. za 1,6 cm. Istraživač će na temelju pojedinih devijacija dati tumačenje kretanja pojave koja se ispituje.

Do sada izvršeni račun i zaključci odnose se na uzorak koji služi kao procjena populacije. Daljnja statistička analiza treba utvrditi: da li se na temelju dobivenog »b« može stvoriti zaključak i o regresiji populacije, tj. da li je koeficijenat regresije uzorka signifikantan ili nije?

Procjenu koeficijenta regresije populacije (β) vršimo pomoću: a) intervalne procjene; i b) testa nul hipoteze koeficijenta regresije populacije, pa ćemo ih prikazati u daljem tekstu.

Prilikom razrade intervalne procjene koeficijenta regresije populacije kao i testa nul hipoteze prvo će se izračunati suma kvadrata devijacije Σdyx^2=7,9834 (tabela 4. 1) koja je baza za procjenu greške slaganja dobivenih podataka s regresijskom linijom. Sumu kvadrata devijacije od regresije podijelit ćemo stupnjem slobode, koji je jednak n-2, te ćemo dobiti varijancu devijacije od regresije:

$$s_{YX}^2 = \Sigma d_{YX}^2/(n-2) \qquad (4.\ 3)$$

$$= 7{,}9834/11 = 0{,}7258$$

Drugi korijen iz varijance devijacije od regresije jest standardna devijacija regresije:

$$s_{YX} = \sqrt{s_{YX}^2} = 0,8519$$

Izračunata vrijednost omogućuje nam izračunavanje standardne devijacije koeficijenta regresije:

$$s_b = s_{YX} / \sqrt{\Sigma x^2} \qquad (4.\ 4)$$

$$= 0,8519 / \sqrt{728} = 0,0316$$

Test signifikantnosti za b izvršit ćemo testiranjem hipoteze po kojoj je koeficijent populacije $\beta=0$, pa ćemo u tu svrhu najprije izračunati vrijednost »t« za konkretan slučaj:

$$t = b/s_b = 1,0329/0,0316 = 32,69 \qquad (4.\ 5)$$

Stupanj slobode (n-2) = 11

Izračunati »t« uspoređujemo s tabelarnim vrijednostima t — distribucije (tabela 2. 2) uz 5% i 1% nivo signifikantnosti. Stupanj slobode iznosi kako smo naveli (n-2) = 11.

$$t_{.05} = 2,201 \quad t_{.01} = 3,106$$

Kako je »t« iz prikazanog uzorka znatno veći od $t_{.05}$ i $t_{.01}$ možemo uz 5% i 1% nivo signifikantnosti utvrditi, da populacijski koeficijent regresije β nije jednak 0.

Intervalna procjena koeficijenta regresije na 5% nivou signifikantnosti vrši se na slijedeći način:

$$b - t_{.05}\, s_b \leq \beta \leq b + t_{.05}\, s_b$$

$$b \pm t_{.05}\, s_b \qquad (4.\ 6)$$

$$1,0329 \pm (2,201)\ (0,0316) = 1,0329 \pm 0,0695$$

Uz 5% nivo signifikantnosti možemo utvrditi, da se koeficijent populacije nalazi unutar granica 0,9634 do 1,1024 cm.

Ako nas ne zadovoljava 5% nivo signifikantnosti možemo intervalnu procjenu koeficijenta regresije populacije provesti na 1% nivou signifikantnosti i to izračunavanjem

$$b \pm t_{.01} s_b .$$

Prikaz veza i utjecaja neovisnog obilježja na ovisno između raznih pojava prikazat ćemo i grafički i to pomoću eksperimentalnih i regresijskih podataka. Ispitivanje veza često i prije računske analize počinjemo grafičkim prikazom, zato što nam grafički prikaz daje uvid u oblik veze (linearni, krivolinijski). Grafički prikaz eksperimentalnih i regresijskih vrijednosti daje uvid i u devijacije pojedinih eksperimentalnih vrijednosti od regresijskih.

U grafičkom prikazu na apscisu se nanose vrijednosti nezavisnog obilježja, a vrijednosti ovisnog obilježja nanosimo na os ordinatu. Tačkicama ili križićima naznačuju se eksperimentalni podaci. Ukoliko imamo izračunate vrijednosti

$\hat{Y}$ za svaku vrijednost X (kao u tabeli 4. 1) prikaz regresijske linije je jednostavan, jer uzimamo proizvoljno dvije izračunate regresijske vrijednosti, unosimo ih u grafički prikaz i spajamo linijom.

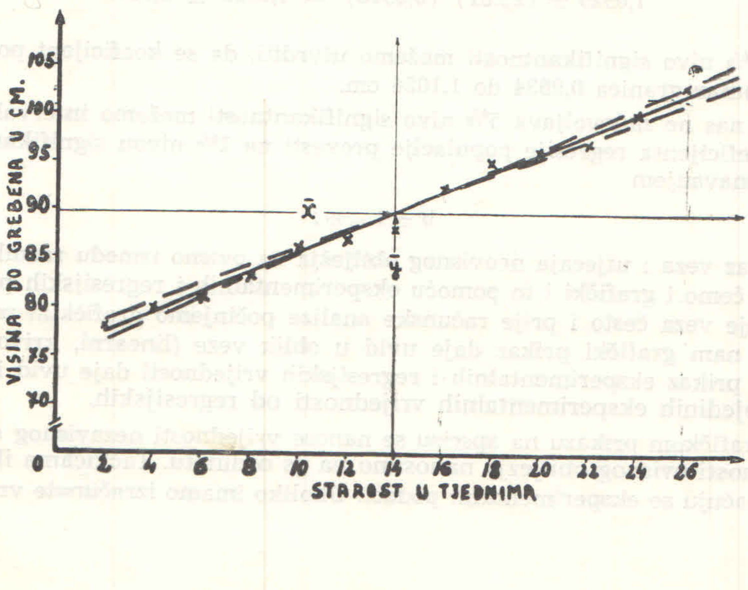
U slučajevima, kada imamo izračunatu regresijsku jednadžbu, ali nemamo izračunate pojedine regresijske vrijednosti, prikaz regresijske linije izvršit ćemo tako, da odredimo dvije tačke kroz koje prolazi regresijska linija. Prva tačka je uvijek određena s vrijednostima $\bar{x}$ i $\bar{y}$, jer linija regresije uvijek prolazi kroz tačku O' ($\bar{x}$, $\bar{y}$) koja u prikazanom slučaju iznosi: $\bar{x}$ = 14 tjedana a $\bar{y}$ = 89,69 cm. Druga tačka linije regresije dobije se uvrštavanjem jedne proizvoljne vrijednosti X u jednadžbu regresije ($\hat{Y} = a + bX$) te izračunavanjem $\hat{Y}$. U prikazanom uzorku, ako se za X izabere vrijednost od 26 tjedana i uvrsti u jednadžbu $\hat{Y} = 75{,}2317 + 1{,}0329\ X$ dobije se regresijska vrijednost $\hat{Y}$ =102,09 cm, te će druga tačka regresijskog pravca imati vrijednosti: X = 14 tjedana, $\hat{Y}$ = 102,09 cm.

U grafički prikaz može se unijeti i intervalna procjena β, tj. prikaz još dviju linija, čiji nagib ovisi o graničnim vrijednostima unutar kojih se kreće β uz stanoviti nivo signifikantnosti. Obje linije prolaze također kroz tačku O' ($\bar{x}$, $\bar{y}$), tj. kroz tačku fiksiranu s vrijednostima $\bar{X}$ = 14, $\bar{Y}$ = 89,69. Druga tačka za svaku liniju izračunat će se na taj način, da se proizvoljno izabere vrijednost X, izračuna devijacija $x = (X - \bar{x})$ koja se pomnoži s vrijednostima izračunatim prilikom intervalne procjene koeficijenta regresije populacije β, što u našem slučaju iznosi 0,9634 i 1,1024 cm. Umnožak se zbroji sa $\bar{y}$ te se dobije vrijednost visine ordinate za određenu vrijednost apscise (X).

U slijedećem grafikonu prikazat ćemo regresiju uzorka visine do grebena na starost i intervalnu procjenu β.

Grafikon 4.1:

Regresija uzorka visine do grebena na starost i intervalna procjena koeficijenta regresije



Kratka metoda računanja regresije

Regresijsku analizu možemo izvršiti i skraćenom metodom računanja, koja se primjenjuje naročito onda, kada raspolažemo s odgovarajućim računskim strojevima.

Skraćenu metodu računanja regresije prikazat ćemo na primjeru ispitivanja utjecaja razne starosti u dobi od 2 do 26 tjedana na dužinu trupa, koja se pratila na 10 simentalskih telića.

Tabela 4.2

Dužina trupa 10 simentalskih telića u starosti od 2 do 26 tjedana

Starost u tjednima	X	2	4	6	8	10	12	14	16	18	20	22	24	26
Dužina trupa u cm	Y	73	75	78	80	82	86	87	91	94	96	98	101	104

Izračunavanje regresije:

$\Sigma X = 182$	$\Sigma Y = 1145$	$n = 13$
$\bar{x} = 14$	$\bar{Y} = 88{,}08$	
$\Sigma X^2 \ldots\ldots\ldots = 3276$	$\Sigma Y^2 \ldots\ldots = 102081$	$\Sigma XY \ldots = 16976$
$(\Sigma X)^2/n = 182^2/13 = 2548$	$(\Sigma Y)^2/n = 1145^2/13 = 100848{,}076$	$(\Sigma X)(\Sigma Y)/n = 16030$
$\Sigma x^2 = 728$	$\Sigma y^2 = 1232{,}924$	$\Sigma xy = 946$

$b = \Sigma xy / \Sigma x^2 = 946/728 = 1{,}2995$ cm dužine trupa za jedan tjedan. To znači, da se svaki tjedan povećava u prosjeku dužina trupa za 1,3 cm u razdoblju između 2 i 26 tjedana.

Da bi se testirala signifikantnost koeficijenta regresije »b« postavit će se opet nul hipoteza po kojoj je koeficijent regresije populacije $\beta = 0$, te će se testiranje izvršiti na slijedeći način:

$$\Sigma d_{Yx}^2 = \Sigma y^2 - (\Sigma xy)^2 / \Sigma x^2 = 1232{,}924 - 946^2/728 = 3{,}644$$

$$s_{Yx}^2 = \Sigma d_{Yx}^2 / (n-2) = 3{,}644/11 = 0{,}3313$$

$$s_{Yx} = \sqrt{s_{Yx}^2} = \sqrt{0{,}3313} = 0{,}5756$$

$$s_b = s_{Yx}/\sqrt{\Sigma x^2} = 0{,}5756/\sqrt{728} = 0{,}0213$$

$$t = b/s_b = 1{,}2995/0{,}0213 = 61{,}01$$

Stupanj slobode iznosi (n-2) = 11 $t_{.05} = 2{,}201$ i $t_{.01} = 3.106$. Uz 5% i 1% nivo signifikantnosti 0 hipotezu odbacujemo i smatramo da je β veći od nule.

Jednadžbu regresije uzorka izračunat ćemo na izloženi način:

$$\hat{Y} - \bar{y} = b\,(X - \bar{x})$$

$$\hat{Y} - 88{,}077 = 1{,}2995\,(X - 14)$$

$$\hat{Y} = 69{,}884 + 1{,}2995\,X$$

Intervalna procjena koeficijenta regresije β glasi:

$$b \pm t_{.05}\, s_b$$
$$1{,}2995 \pm (2{,}201)\,(0{,}0213)$$
$$1{,}2995 \pm 0{,}0469$$

Koeficijent regresije populacije nalazi se uz 95 % vjerojatnost unutar granica 1,2526 cm i 1,3464 cm.

Krivolinijska regresija

Podaci dobiveni prilikom ispitivanja veza pomoću regresijske analize često se neće kretati linearno, te nam u tom slučaju neće odgovarati niti vrijednosti dobivene linearnom regresijom.

U daljnjem razlaganju osvrnut ćemo se na regresijske analize kod kojih se obilježja kreću parabolično. U prikazu eksperimentalnih i izračunatih vrijednosti obično će se pojaviti samo mali segment parabole.

U slučaju kada je kretanje podataka vrlo blizu paraboličnom obliku može se reprezentirati jednadžbom:

$$Y = a + bX + cX^2$$

Jednadžbu krivolinijske regresije izračunat ćemo prema slijedećem obrascu:

$$\hat{Y} = \bar{y} + b\,(X_1 - \bar{x}_1) + c\,(X_2 - \bar{x}_2) \qquad (4.\ 7)$$

$$b = \frac{(\Sigma x_2^2)\,(\Sigma x_1 y) - (\Sigma x_1 x_2)\,(\Sigma x_2 y)}{D}$$

$$c = \frac{(\Sigma x_1^2)\,(\Sigma x_2 y) - (\Sigma x_1 x_2)\,(\Sigma x_1 y)}{D}$$

$$D = (\Sigma x_1^2)\,(\Sigma x_2^2) - (\Sigma x_1 x_2)^2$$

$$X_2 = X_1^2$$

Kako ćemo u konkretnom slučaju izračunati krivolinijsku regresiju prikazat ćemo na podacima pojedene hrane u kg i prirastu u kg 10 grla svinja.

Tabela 4.3

Prirasti kod svinja i količina konzumirane hrane

Grlo broj	Pojedena hrana u kg X_1	Kvadrat pojedene hrane X_2 (X_1^2)	Prirasti u kg Y	Izračunata vrijednost $\hat{Y}$	Devijacija $Y - \hat{Y}$
1	163	26.569	31	32,0912	—1,0912
2	158	24.964	26	28,3005	—2,3005
3	199	39.601	42	40,5353	1,4647
4	184	33.856	42	41,0395	0,9605
5	168	28.224	37	35,2433	1,7567
6	189	35.721	39	41,5099	—2,5099
7	193	37.249	41	41,4266	—0,4266
8	160	25.600	32	29,8934	2,1066
9	180	32.400	39	40,2034	—1,2034
10	173	29.929	39	37,7570	1,2430
Suma	1767	314.113	368	368,0001	—0,0001
Sredina	176,7	31.411,3	36,8		

$$\Sigma X_1^2 = 314113{,}0 \qquad \Sigma X_1X_2 = 56172837{,}0$$
$$(\Sigma X_1)^2/n = 312228{,}9 \qquad (\Sigma X_1)(\Sigma X_2)/n = 55503767{,}1$$
$$\Sigma x_1^2 = 1884{,}1 \qquad \Sigma x_1x_2 = 669069{,}9$$

$$\Sigma X_2^2 = 10104518053{,}0 \qquad \Sigma X_2Y = 11772544{,}0$$
$$(\Sigma X_2)^2/n = 9866697676{,}9 \qquad (\Sigma X_2)(\Sigma Y)/n = 11559358{,}4$$
$$\Sigma x_2^2 = 237820376{,}1 \qquad \Sigma x_2y = 213185{,}6$$

$$\Sigma X_1Y = 65634{,}0 \qquad \Sigma Y^2 = 13802{,}0$$
$$(\Sigma X_1)(\Sigma Y)\ n = 65025{,}6 \qquad (\Sigma Y)^2/n = 13542{,}4$$
$$\Sigma x_1y = 608{,}4 \qquad \Sigma y^2 = 259{,}6$$

$$D = (\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1x_2)^2 = (1884{,}1)(237820376{,}1) - 669069{,}9^2$$
$$= 422839524{,}0$$

$$b = \frac{(\Sigma x_2^2)(\Sigma x_1y) - (\Sigma x_1x_2)(\Sigma x_1y)}{D}$$

$$= \frac{(237820376{,}1)(608{,}4) - (669069{,}9)(608{,}4)}{422839524{,}0}$$

$$b = 4{,}8573$$

$$c = \frac{(\Sigma x_1^2)(\Sigma x_2y) - (\Sigma x_1x_2)(\Sigma x_1y)}{D}$$

$$= \frac{(1884{,}1)(213185{,}6) - (669069{,}9)(608{,}4)}{422839524{,}00}$$

$$c = -0{,}01277$$

$$\hat{Y} = \bar{y} + b(X_1 - \bar{x}_1) + c(X_2 - \bar{x}_2)$$

$$\hat{Y} = 36{,}8 + 4{,}8573\ (X_1 - 176{,}7) - 0{,}01277\ (X_2 - 31411{,}3)$$

$$\hat{Y} = -420{,}3626 + 4{,}8573\ X_1 - 0{,}0128\ X_2$$

Kako je $X_2 = X_1^2$ gornju jednadžbu možemo pisati na slijedeći način:

$$\hat{Y} = -420{,}3626 + 4{,}8573\ X - 0{,}0128\ X^2$$

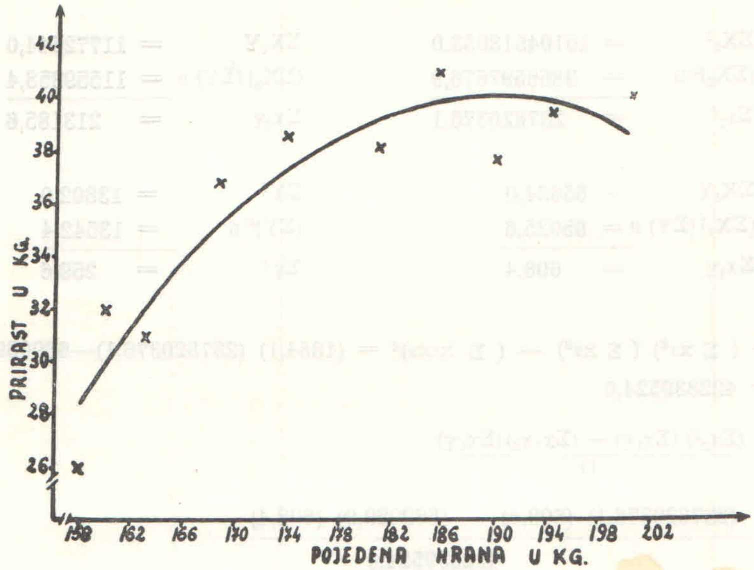
Grafički prikaz eksperimentalnih i izračunatih vrijednosti prikazat ćemo na grafikonu 4. 2.

Grafički prikaz daje uvid u kojem stupnju regresijska krivulja predstavlja kretanje ispitivanih pojava, te se na njemu mogu očitati i devijacije pojedinih eksperimentalnih vrijednosti od regresijskih, što je prikazano i u zadnjoj koloni tabele 4. 3 ($d_{YX} = Y - \hat{Y}$).

Grafikon 4.2

Regresija prirasta kod svinja na pojedenu hranu

$$\hat{Y} = -420{,}3626 + 4{,}8573\ X - 0{,}0128\ X^2$$



Test signifikantnosti skretanja od linearne regresije

Prilikom primjene krivolinijske regresije može se pomoću testa signifikantnosti skretanja od linearne regresije utvrditi, da li je opravdana primjena krivolinijske regresije ili nema signifikantne razlike između rezultata dobivenih linearnom i krivolinijskom regresijom.

Da bi se provelo testiranje, mora se izračunati »suma kvadrata devijacija od linearne regresije« $(\Sigma d_{\hat{Y}x}^2)$ i »suma kvadrata devijacije od krivolinijske regresije« $(\Sigma d^2_{\hat{Y}xx^2})$. Razlika između $\Sigma d_{\hat{Y}x}^2 - \Sigma d^2_{\hat{Y}xx^2}$ = »suma kvadrata zakrivljenosti regresije«. Suma kvadrata devijacije od krivolinijske regresije podijeli se s odgovarajućim stupnjem slobode koji u ovom slučaju iznosi: (n-3), te se dobije varijanca devijacije od krivolinijske regresije. Suma kvadrata zakrivljenosti podijeli se sa 1, jer je stupanj slobode zakrivljenosti jednak jedinici, te se dobije varijanca zakrivljenosti.

Omjer varijance zakrivljenosti regresije i varijance devijacije od krivolinijske regresije jest vrijednost »F« koju opet testiramo pomoću tabele F distribucije (tabela 3. 8). Stupanj slobode f_1 = stupanj slobode zakrivljenosti regresije, a stupanj slobode f_2 = stupanj slobode devijacije od krivolinijske regresije.

Testiranje signifikantnosti prikazat ćemo na već izloženom primjeru krivolinijske regresije prirasta kod svinja i količine pojedene hrane (tabela 4. 3).

Najprije ćemo prikazati izračunavanje sume kvadrata devijacije od linearne regresije $(\Sigma d_{\hat{Y}x}^2)$ i krivolinijske regresije $(\Sigma d^2_{\hat{Y}xx^2})$

Suma kvadrata devijacija od linearne regresije možemo izračunati pomoću formule:

$$\Sigma d_{Yx}^2 = \Sigma(Y - \hat{Y})^2$$

u kojoj je $\hat{Y}$ regresijska vrijednost linearne regresije. Međutim metoda, koja je data iza tabele 4. 2, jednostavnija je te ćemo je i primijeniti:

$$\Sigma d_{Yx}^2 = \Sigma y^2 - (\Sigma_{xY})^2 / \Sigma x^2$$

Sve vrijednosti, koje moramo uvrstiti u jednadžbu, imamo izračunate iza tabele 4. 3 te će biti:

$$\Sigma d_{Yx}^2 = 259{,}6 - (608{,}4)^2/1884{,}1 = 63{,}1399$$

Suma kvadrata devijacija od krivolinijske regresije može se izračunati pomoću formule:

$$\Sigma (Y - \hat{Y})^2$$

u kojoj je $\hat{Y}$ regresijska vrijednost krivolinijske regresije. Brži način izračunavanja bit će ako primijenimo slijedeću jednadžbu:

$$\Sigma d^2_{Yxx^2} = \Sigma y^2 - b\,\Sigma x_1 y - c\,\Sigma x_2 y$$

Iza tabele 4. 3 imamo izračunate sve vrijednosti za desnu stranu jednadžbe te dobivamo da je

$$\Sigma d^2_{Yxx^2} = 259{,}6 - (4{,}8573)\,(608{,}4) + (0{,}01277)\,(213.185{,}6)$$
$$\Sigma d^2_{Yxx^2} = 26{,}7988$$

Izračunate vrijednosti kao i izračunavanje varijance i omjera

$$F = \frac{\text{varijanca zakrivljenosti regresije}}{\text{varijanca devijacije od krivolinijske regresije}} \tag{4.8}$$

prikazat ćemo u tabeli 4. 4.

Tabela 4.4

Test signifikantnosti skretanja od linerane regresije

Izvor varijacije	Stupanj slobode	Suma kvadrata	Varijanca
devijacija od linearne regresije	8	63,1399	
devijacija od krivolinijske regresije	7	26,7988	3,8284
zakrivljenost regresije	1	36,3411	36,3411
F = 36,3411/ 3,8284 = 9,4925			

Za prikazani primjer »F« iznosi 9,4925, a vrijednost dobivena iz tabele F distribucije iznosi $F_{.05} = 5{,}59$ i $F_{.01} = 12{,}25$. Vrijednost se očitava kod stupnja slobode $f_1 = 1$ i $f_2 = 7$. Na temelju izvršene komparacije na 5% nivou signifikantnosti odbacujemo hipotezu linearne regresije, što znači da je zakrivljenost signifikantna. Međutim, uz 1% nivo signifikantnosti ne možemo utvrditi, da je zakrivljenost signifikantna ($P < 0{,}05$, $P > 0{,}01$).

Test signifikantnosti koeficijenata regresije

Na temelju jednadžbe krivolinijske regresije uzorka može se izvršiti testiranje signifikantnosti koeficijenata regresije »b« i »c« i to metodom testiranja hipoteze, da su koeficijenti jednadžbe krivolinijske regresije populacije β i γ jednaki 0.

Testiranje ćemo prikazati na primjeru prirasta kod svinja i količine konzumirane hrane (tabela 4. 3).

Za testiranje koeficijenata b i c potrebna nam je vrijednost varijance regresije ovisnog obilježja koju ćemo izračunati na slijedeći način:

$$\Sigma \widehat{Y}^2_{XX^2} = b \Sigma X_1 Y + c \Sigma X_2 Y$$

$$\Sigma d^2_{YXX^2} = \Sigma Y^2 - \Sigma \widehat{Y}^2_{XX^2}$$

$$s_{YXX}^2 = \Sigma d^2_{YXX^2}\ n - 3$$

Iza tabele 4. 3 imamo izračunate vrijednosti za b (4,8573), c (—0,01277), $\Sigma x_1 y$ (608,4), $\Sigma x_2 y$ (213185,6) i Σy^2 (259,6) te ćemo dobiti, da je varijanca regresije ovisnog obilježja jednaka:

$$\Sigma \widehat{Y}^2_{XX^2} = (4{,}8573)\ (608{,}4) + (-0{,}01277)\ (213185{,}6) = 232{,}8012$$

$$\Sigma d^2_{YXX^2} = 259{,}6 - 232{,}8012 = 26{,}7988$$

$$s^2_{YXX^2} = 26{,}7988/7 = 3{,}8284$$

Test signifikantnosti za koeficijent b (β = 0):

$$t = b/s_b \qquad (4.\ 9)$$

$t = 4{,}8573/1{,}4673$

$t = 3{,}3104$

$b = 4{,}8573$

$s_b^2 = s^2_{YXX^2}\ c_b$

$c_b = \Sigma x_2^2/D = 237820376{,}1/422839524$

$c_b = 0{,}5624$

$s^2_b = (3{,}8284)\ (0{,}5624) = 2{,}1531$

$s_b = 1{,}4673$

Test signifikantnosti za koeficijent c (γ = 0):

$$t = c/s_c \qquad (4.\ 10)$$

$t = -0{,}01277/0{,}00412$

$t = -3{,}099$

$c = -0{,}01277$

$s_c^2 = s^2_{YXX^2}\ c_c$

$c_c = \Sigma x_1^2/D = 1884{,}1/422839524$

$c_c = 0{,}00000445$

$s^2_c = (3{,}8284)\ (0{,}00000445) = 0{,}000017$

$s_c = 0{,}00412$

Stupanj slobode jednak je n—3 = 10—3 = 7, te je vrijednost $t_{.05}$ = 2,365.

Na temelju izvršenog testa odbacujemo nul hipotezu (β = 0, γ = 0) te smatramo, da su populacijski koeficijenti β i γ uz 5% vjerojatnost signifikantno različiti od nule.

Korelacija (mali uzorak)

Metodom koleracije može se ispitivati jakost veze između obilježja dobivenih slučajnim izborom iz populacije čija su obilježja normalno distribuirana. Kao što smo izložili kod regresije i kod korelacije važno je da znamo oblik veze, tako da primijenimo ispravan način analize.

Ukoliko postoji linearni oblik veze, jakost veze izrazit ćemo pomoću linearnog koeficijenta korelacije koji glasi

$$r = \frac{\Sigma x_1 x_2}{\sqrt{(\Sigma x_1^2)\ (\Sigma x_2^2)}} \tag{4.11}$$

Koeficijent korelacije može biti pozitivan i negativan, što nam naznačuje smjer veze. Negativan predznak naznačuje, da su velike vrijednosti jednog obilježja udružene s malim vrijednostima drugog obilježja, dok pozitivan predznak naznačuje, da su velike vrijednosti jednog obilježja povezane s velikim vrijednostima drugog obilježja i obrnuto.

Vrijednost koeficijenta korelacije »r« kreće se između —1 i +1, što ovisi o jačini veze obilježja u uzorku. Ukoliko nema veze između dva obilježja, r će iznositi nulu (r = 0). Što je veza jača r postaje sve veći i sve se više približava $\pm$ 1.

Izračunavanje i tumačenje linearnog korelacionog koeficijenta dat ćemo na primjeru korelacione povezanosti visine do grebena i dužine trupa 14 simentalske teladi stare 26 tjedana, čije podatke dajemo u tabeli 4.5.

Tabela 4.5

Visina do grebena i dužina trupa 14 simentalske teladi stare 26 tjedana

Tele br.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
visina do grebena	110	106	107	104	108	105	107	105	104	109	107	109	113	100
dužina trupa	108	107	108	108	113	110	111	110	106	110	107	115	113	103

Vrijednosti $\Sigma x_1 x_2$ te Σx_1^2, i Σx_2^2 koje su potrebne za koeficijent korelacije mogu se izračunati na slijedeći način:

X_1	X_2	$x_1=(X_1-\bar{x}_1)$	$x_2=(X_2-\bar{x}_2)$	x_1^2	x_2^2	$x_1 x_2$
110	108	3,29	—1,21	10,8241	1,4641	—3,9809
106	107	—0,71	—2,21	0,5041	4,8841	1,5691
107	108	0,29	—1,21	0,0841	1,4641	—0,3509
104	108	—2,71	—1,21	7,3441	1,4641	3,2791
108	113	1,29	3,79	1,6641	14,3641	4,8891
105	110	—1,71	0,79	2,9241	0,6241	—1,3509
107	111	0,29	0,79	0,0841	3,2041	0,5191
105	110	—1,71	0,79	2,9241	0,6241	—1,3509
104	106	—2,71	—3,21	7,3441	10,3041	8,6991
109	110	2,29	0,79	5,2441	0,6241	1,8091
107	107	0,29	—2,21	0,0841	4,8841	—0,6409
109	115	2,29	5,79	5,2441	33,5241	13,2591
113	113	6,29	3,79	39,5641	14,3641	23,8391
100	103	—6,71	—6,21	45,0241	38,3641	41,6691
1494	1529	0,08	0,06	128,8574	130,3574	91,8574

$n = 14 \quad \bar{x}_1 = 106{,}71 \quad \bar{x}_2 = 109{,}21$

$\Sigma x_1^2 = 128{,}8574 \qquad \Sigma x_2^2 = 130{,}3574 \qquad \Sigma x_1 x_2 = 91{,}8574$

$$r = \Sigma x_1 x_2 / \sqrt{(\Sigma x_1^2)\,(\Sigma x_2^2)} = 91{,}857 / \sqrt{(128{,}86)\,(130{,}36)} = +\,0.7087$$

Izračunati »r« odnosi se na korelaciju unutar uzorka pa se opet postavlja pitanje u kojoj se mjeri možemo na njega osloniti u procjeni korelacionog koeficijenta ukupne mase (ρ). Da bi to utvrdili testirat ćemo nul hipotezu. U tom ćemo cilju opet postaviti hipotezu da je $\rho = 0$, a zatim ćemo ispitati uz koju signifikantnost možemo ustvrditi, da na temelju izračunatog koeficijenta korelacije uzorka (r) koeficijent korelacije populacije (ρ) iznosi ili ne iznosi 0.

Tabela 4.6

Koeficijenti korelacije na 5% i 1% nivou signifikantnosti

Stupanj slobode	5%	1%	Stupanj slobode	5%	1%
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1,000	.062	.081

Testiranje se vrši pomoću tabele 4.6 »Koeficijenti korelacije na 5% i 1% nivou signifikantnosti«. Stupanj slobode iznosi n—2, te kako je u prikazanom primjeru n=14, stupanj slobode jednak je 14 — 2 = 12. Ukoliko je tabelarna vrijednost veća od našeg izračunatog »r«, nul hipoteza se odbacuje i obrnuto. Izračunati r = 0,71 veći je od tabelarne vrijednosti sa 5% nivoom signifikantnosti (0,532) i 1% nivoom signifikantnosti (0,661), te nul hipotezu odbacujemo. Uz 99% vjerojatnoću zaključujemo, da i u populaciji postoji pozitivna koleracija između visine do grebena i dužine trupa.

Korelacija (veliki uzorak)

Ukoliko nam je uzorak mali, često se neće prilikom testiranja utvrditi opravdanost korelacionog koeficijenta uz stanoviti nivo signifikantnosti unatoč tome, što u populaciji postoji korelacija. Kako prilikom svih ispitivanja, tako ćemo i kod ispitivanja korelacionog koeficijenta nastojati unutar naših mogućnosti da uzorak bude što veći.

Pomoću podataka za debljinu i nosivost vunske niti prikazat ćemo računanje korelacionog koeficijenta za velike uzorke.

Nosivost i finoća svake vunske niti unesena je u određeni red i kolonu tabele 4.7. Vrijednost f_X naznačuje zbroj frekvencija pojedinih kolona, a f_Y zbroj frekvencija pojedinih razreda. Vrijednosti za finoću i nosivost X i Y izražene su pomoću kodiranih brojeva te smo nule smjestili u razrede u kojima se nalaze aproksimativne srednje vrijednosti. Svaka vrijednost kolone $\Sigma X f_X$ dobivena je zbrajanjem umnožaka frekvencija u pojedinom redu s odgovarajućom vrijednosti u redu Code X.

$\Sigma Y f_Y$ dobivena je zbrajanjem umnožaka frekvencija u pojedinoj koloni s odgovarajućom vrijednosti u koloni Code Y. Na primjer:

$\Sigma X f_X$

1. red : $(1)\ (8) = 8$
2. red : $(1)\ (6) + (1)\ (7) + (1)\ (8) = 21$
3. red : $(3)\ (5) + (4)\ (6) + (3)\ (7) + (2)\ (8) = 76$
4. red : $(2)\ (2) + (3)\ (4) + (3)\ (5) + (2)\ (6) + (1)\ (7) = 50$

itd.

$\Sigma Y f_Y$

1. kolona : $(1)\ (-9) = -9$
2. kolona : $(2)\ (-9) = -18$
3. kolona : $(1)\ (-9) + (1)\ (-8) + (2)\ (-7) + (1)\ (-6) = -37$
4. kolona : $(3)\ (-9) + (3)\ (-8) + (2)\ (-7) + (2)\ (-6) + (1)\ (-3) = -80$

itd.

$\Sigma X^2 f_X$ i $\Sigma Y^2 f_Y$ dobije se na isti način samo je svaka vrijednost X i Y kvadrirana.

$\Sigma X^2 f_X$

1. red : $(1)\ (8^2) = 64$
2. red : $(1)\ (6^2) + (1)\ (7^2) + (1)\ (8^2) = 149$
3. red : $(3)\ (5^2) + (4)\ (6^2) + (3)\ (7^2) + (2)\ (8^2) = 494$

itd.

$\Sigma Y^2 f_Y$

1. kolona : $(1)\ (-9^2) = 81$
2. kolona : $(2)\ (-9^2) = 162$
3. kolona : $(1)\ (-9^2) + (1)\ (-8^2) + (2)\ (-7^2) + (1)\ (-6^2) = 279$

itd.

Tabela 4.7

Računanje korelacionog koeficijenta uzorka za debljinu i nosivost vunskih niti

Nosivost u gr.	Debljina u μ																				f_Y	Code Y	$\Sigma X f_x$	Produkt Y ($\Sigma X f_x$)
	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54				
50																				1	1	12	8	96
48																		1	1	1	3	11	21	231
46																	3	4	3	2	12	10	76	760
44														2		3	3	2	1		11	9	50	450
42												1		3	2	3	4	2	1	2	18	8	79	632
40													1	2	1	2	3	3	2		14	7	63	441
38												1	3	2	3	4	3	2	1	1	20	6	74	444
36											2	3	4	4	5	6	5	3	2		34	5	106	530
34											3	3	2	4	5	5	3	4	1		30	4	88	352
32										4	2	3	4	3	5	6	4	3			34	3	77	231
30									3	3	6	7	6	6	7	6	4	3			51	2	80	160
28								1	6	10	8	13	7	6	4	2	1	1			59	1	0	0
26								2	6	8	10	6	5	2	4	3	3				49	0	—4	0
24						2	2	3	4	4	8	7	6	4	3	4					47	—1	—23	23
22						3	2	3	4	4	6	5	2	3	2	1	1				36	—2	—43	86
20				1	1	3	3	5	6	4	5	4	4	2	3	2					43	—3	—74	222
18					2	4	4	4	5	6	3	6	4	4	1						43	—4	—89	356
16					2	1	3	3	5	4	3	2	1								24	—5	—72	360
14			1	2	4	4	5	2	3	3	2										26	—6	—127	762
12			2	2	3	5	2	1	1	1		1									18	—7	—104	,728
10			1	3	2	3	2	2	1												14	—8	—86	688
8	1	2	1	3	3	2	1														13	—9	—102	918
f_x	1	2	5	11	17	27	24	26	44	51	58	62	49	47	45	47	37	28	12	7	600		—2	8.470
Code X	-11	-10	—9	—8	—7	—6	—5	—4	—3	—2	—1	0	1	2	3	4	5	6	7	8				

Zadnja kolona u tabeli 4.7 $Y(\Sigma X f_x)$ dobije se umnoškom kolone Y sa $\Sigma X f_x$ a njezina je suma jednaka vrijednosti $\Sigma XY = 8.470$.

Ostala tehnika računanja ista je kao i kod prije izloženih metoda računanja.

Računanje korelacionog koeficijenta:

$\Sigma X f_x$	$= -2$	$\Sigma Y f_Y$	$= 128$		
$\Sigma X^2 f_x$	$= 9.272{,}0$	$\Sigma Y^2 f_Y$	$= 12.688{,}0$	ΣXY	$= 8\,470{,}0$
$(\Sigma X f_x)^2/n$	$= 0{,}00666$	$(\Sigma Y f_Y)^2/n$	$= 27{,}3066$	$(\Sigma X f_x)(\Sigma Y f_Y)/n$	$= 0{,}4266$
Σx^2	$= 9.271{,}99334$	Σy^2	$= 12.660{,}6934$	Σxy	$= 8.469{,}5734$

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{8.469{,}5734}{\sqrt{(9\,271{,}99)(12\,660{,}69)}} = 0{,}7817$$

Vrijednost korelacionog koeficijenta u tabeli 4.6: »Koeficijenti korelacije na 5% i 1% nivou signifikantnosti« iznosi 0,115 (stupanj slobode = 500) i 0,081 (stupanj slobode = 1.000), te nul hipotezu odbacujemo i zaključujemo uz 1% signifikantnost, da postoji korelacija između nosivosti i finoće vunskih niti.

Postoji više metoda za izračunavanje korelacionog koeficijenta velikog uzorka no sve metode zauzimaju podjednako vremena, a jasno je da daju iste rezultate.

VEZA IZMEĐU KOEFICIJENTA REGRESIJE I KOLERACIJE

Između vrijednosti dobivenih računom regresije i koleracije postoji uska veza, te ćemo se ukratko osvrnuti na jednadžbe koje nam tačno prikazuju spomenute odnose.

Označimo li u računu linearne regresije X_2 kao ovisno obilježje, a X_1 kao neovisno, koeficijent regresije iznosit će $b_{21} = \Sigma x_1x_2 / \Sigma x_1^2$. Ukoliko je X_2 neovisno a X_1 ovisno obilježje koeficijent regresije iznosi $b_{12} = \Sigma x_1x_2 / \Sigma x_2^2$. Iz izloženoga izlazi da je

$$b_{21} \,.\, b_{12} = \frac{(\Sigma x_1 x_2)^2}{\Sigma x_1^2 \; \Sigma x_2^2} = r^2$$

$$r = \sqrt{b_{21} \cdot b_{12}} \qquad (4.12)$$

U tabeli 4.5 izneseni su podaci visine do grebena i dužine trupa 14 simentalske teladi u starosti od 26 tjedana (X_1 = visina do grebena, X_2 = dužina trupa). Izračunamo li koeficijent linearne regresije za visinu do grebena na dužinu trupa dobit ćemo, da je $b_{12} = 0{,}7047$ cm. Koeficijent regresije dužine trupa na visinu do grebena iznosi $b_{21} = 0{,}7129$ cm. Na temelju koeficijenta regresije izračunat ćemo koeficijent korelacije koji iznosi:

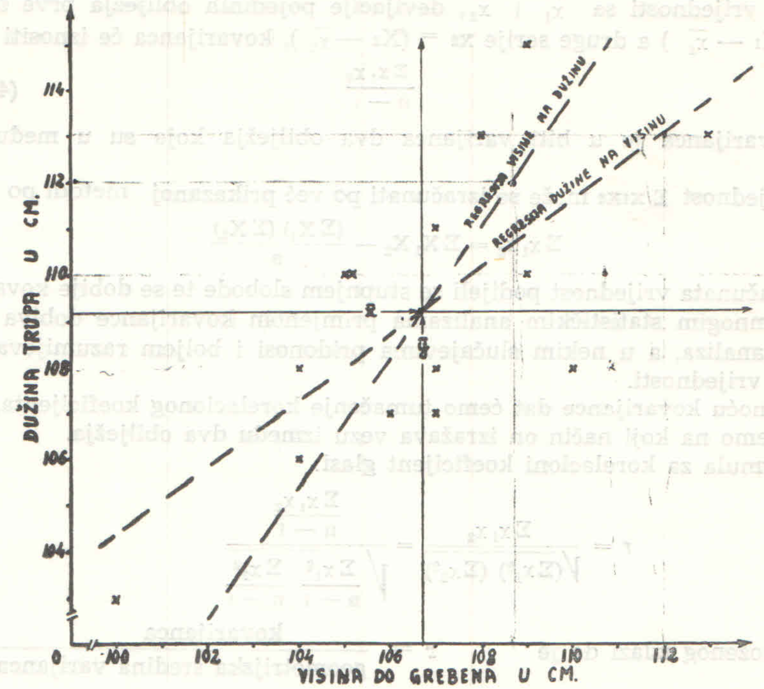
$$r = \sqrt{b_{21} \,.\, b_{12}} = \sqrt{0{,}7129 \,.\, 0{,}7047} = 0{,}7088$$

Izračunata vrijednost koeficijenta korelacije ista je kao i izračunati »r« iza tabele 4.5.

U grafikonu 4.3 prikazat ćemo podatke iz tabele 4.5 za visinu do grebena i dužinu trupa 14 simentalskih telića te regresijske jednadžbe i to regresije dužine trupa na visinu do grebena ($\hat{Y} = 33{,}1377 + 0{,}7129\ X$) i regresije visine do grebena na dužinu trupa ($\hat{Y} = 29{,}7510 + 0{,}7047\ X$).

Grafikon 4.3

Visina do grebena i dužina trupa 14 simentalske teladi u starosti od 26 tjedana ($r = 0{,}7087$)



Na grafičkom će prikazu regresijske linije biti indentične (poklapat će se) u koliko je $r = 1$, razlike će među njima biti sve veće što se vrijednost r više približava nuli, te će regresijske linije postati okomite kada vrijednost koeficijenta regresije postane nula ($r = 0$).

Vezu između koeficijenta regresije i korelacije možemo prikazati i slijedećom jednadžbom:

$$b_{21} = r\,\frac{s_Y}{s_X} \tag{4.13}$$

Za primjer izložen u tabeli 4.5 iznosit će

$$s_X = \sqrt{128{,}86/13} = 3{,}1483 \text{ a } s_Y = \sqrt{130{,}36/13} = 3{,}1665$$

$$b_{21} = 0{,}7088\ \frac{3{,}1665}{3{,}1483} = 0{,}7129$$

$$b_{12} = 0{,}7088\ \frac{3{,}1483}{3{,}1665} = 0{,}7047$$

Koeficijenti korelacije mogu se izračunati, ako su poznati koeficijenti regresije za oba obilježja. Koeficijenti regresije mogu se opet jednostavno izračunati ako su poznati koeficijenti korelacije (r) i standardne devijacije (s_Y, s_X)

KOVARIJANCA

Suma umnožaka odstupanja pojedinih obilježja dviju serija od njihovih aritmetičkih sredina podijeljena stupnjem slobode naziva se kovarijanca. Ako se sa X_1 označe obilježja prve serije, sa X_2 obilježja druge serije, odgovarajuće srednje vrijednosti sa $\bar{x}_1$ i $\bar{x}_2$, devijacije pojedinih obilježja prve serije sa $x_1 = (X_1 - \bar{x}_1)$ a druge serije $x_2 = (X_2 - \bar{x}_2)$, kovarijanca će iznositi

$$\frac{\Sigma x_1 x_2}{n - 1} \qquad (4.14)$$

Kovarijanca je u biti varijanca dva obilježja koja su u međusobnom odnosu.

Vrijednost $\Sigma x_1 x_2$ može se izračunati po već prikazanoj metodi po kojoj je

$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - \frac{(\Sigma X_1)(\Sigma X_2)}{n}$$

Izračunata vrijednost podijeli se stupnjem slobode te se dobije kovarijanca. U mnogim statističkim analizama primjenom kovarijance dobiva se preciznija analiza, a u nekim slučajevima pridonosi i boljem razumijevanju pojedinih vrijednosti.

Pomoću kovarijance dat ćemo tumačenje korelacionog koeficijenta, tj. prikazat ćemo na koji način on izražava vezu između dva obilježja.

Formula za korelacioni koeficijent glasi:

$$r = \frac{\Sigma x_1 x_2}{\sqrt{(\Sigma x_1^2)(\Sigma x_2^2)}} = \frac{\frac{\Sigma x_1 x_2}{n - 1}}{\sqrt{\frac{\Sigma x_1^2}{n - 1} \, \frac{\Sigma x_2^2}{n - 1}}}$$

te iz izloženog izlazi da je $r = \dfrac{\text{kovarijanca}}{\text{geometrijska sredina varijanca}}$

Na temelju navedene formule može se reći, da je koeficijent korelacije kvocijent dviju prosječnih varijabilnosti i to kovarijance obilježja X_1 i X_2 i geometrijske sredine varijance uzoraka s obilježjima X_1 i X_2. Znači da vrijednost koleracionog koeficijenta naznačuje mjeru u kojoj varijabilnost dva obilježja ide zajedno, tj. u kojem obimu varijabilnost jednog obilježja prati varijabilnost drugoga. Na takav način koeficijent korelacije dobiva potpuno tumačenje.

Za podatke visine do grebena i dužine trupa 14 komada simentalske teladi (tabela 4.5) izračunat ćemo kovarijancu i varijancu za oba obilježja.

Tabela 4.8 *Visina do grebena i dužina trupa 14 komada simentalske teladi stare 12 tjedana*

Tele br.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
visina do grebena, X_1	110	106	107	104	108	105	107	105	104	109	107	109	113	100
dužina trupa, X_2	108	107	108	108	113	110	111	110	106	110	107	115	113	103

$$\Sigma X_1^2 = 159560{,}00 \qquad \Sigma X_2^2 = 167119{,}00$$

$$\frac{(\Sigma X_1)^2}{n} = \frac{1494^2}{14} = 159431{,}14 \qquad \frac{(\Sigma X_2)^2}{n} = \frac{1529^2}{14} = 166988{,}64$$

$$\Sigma x_1^2 = 128{,}86 \qquad \Sigma x_2^2 = 130{,}36$$

$$s_1^2 = \frac{128{,}86}{13} = 9{,}9123 \qquad s_2^2 = \frac{130{,}36}{13} = 10{,}0276$$

$$\Sigma X_1X_2 \ldots\ldots\ldots = 163258{,}00$$

$$\frac{(\Sigma X_1)\,(\Sigma X_2)}{n} = \frac{(1494)\,(1529)}{14} = 163166{,}14$$

$$\Sigma x_1 x_2 = 91{,}86$$

$$\text{kovarijanca} = \frac{91{,}86}{13} = 7{,}0661$$

Na temelju dobivenih vrijednosti možemo izračunati koeficijent korelacije

$$r = \frac{7{,}0661}{\sqrt{(9{,}9123)\,(10{,}0276)}} = 0{,}7087$$

Dobili smo istu vrijednost kao i u računu iza tabele 4.5.

KOEFICIJENT DETERMINACIJE

Veza između koeficijenta regresije i korelacije omogućuje nam i izračunavanje koeficijenta determinacije za linearnu vezu (r^2) i krivolinijsku ($r^2_{YXX'}$). Koeficijent determinacije pokazuje u obliku proporcije dio ukupne varijabilnosti ovisnog obilježja koji je rastumačen izračunavanjem regresije.

Tumačenje koeficijenta determinacije dat ćemo pomoću podataka visine do grebena i dužine trupa 14 komada simentalske teladi prikazanih u tabeli 4.5.

Na temelju jednadžbe regresije dužine trupa na visinu do grebena ($\hat{Y} = 33{,}1377 + 0{,}7129\ X$) izračunate su regresijske vrijednosti za pojedine visine do grebena. Eksperimentalni podaci, regresijske vrijednosti te devijacije $(Y - \bar{Y})$, $(\hat{Y} - \bar{Y})$, $(Y - \hat{Y})$, prikazani su u tabeli 4.9.

Ukupne devijacije $(Y - \bar{Y})$ dobivaju se izračunavanjem diferencija pojedinog ovisnog obilježja od aritmetičke sredine i na njihove vrijednosti ne utječu regresijski podaci. Devijacije $(\hat{Y}\ \bar{Y})$ jesu diferencije pojedinih regresijskih vrijednosti od aritmetičke sredine i njihove su vrijednosti potpuno rastumačene regresijom, dok devijacije $(Y - \hat{Y})$, koje naznačuju udaljenost pojedinog obilježja od regresijske vrijednosti, nisu protumačene regresijom. Svaka devijacija $(Y - \bar{Y})$ može se podijeliti u dva dijela i to u devijaciju $(\hat{Y} - \bar{Y})$ i $(Y - \hat{Y})$, tj. $(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$.

Na prikazanom raščlanjivanju osniva se i izračunavanje dijela ukupne varijabilnosti koji je rastumačen regresijom.

U tabeli 4.9 prikazat ćemo i kvadrate devijacija, $(Y - \bar{Y})^2$, $(\hat{Y} - \bar{Y})^2$, $(Y - \hat{Y})^2$ koji služe za izračunavanje varijacije protumačene regresijom i neprotumačene regresijom.

Tabela 4.9

Izračunavanje ukupne varijacije, rastumačene varijacije i nerastumačene varijacije za regresiju dužine trupe na visinu do grebena

Visina grebena X	Dužina trupa Y	$\hat{Y}$	Devijacija			Kvadrat devijacija		
			$Y-\bar{Y}$	$\hat{Y}-\bar{Y}$	$Y-\hat{Y}$	$(Y-\bar{Y})^2$	$(\hat{Y}-\bar{Y})^2$	$(Y-\hat{Y})^2$
110	108	111,5567	—1,2142	2,3425	—3,5567	1,4743	5,4873	12,6501
106	107	108,7051	—2,2142	—0,5091	—1,7051	4,9027	0,2592	2,9074
107	108	109,4180	—1,2142	0,2038	—1,4180	1,4743	0,0415	2,0107
104	108	107,2793	—1,2142	—1,9349	0,7207	1,4743	3,7438	0,5194
108	113	110,1309	3,7858	0,9167	2,8691	14,3323	0,8403	8,2317
105	110	107,9922	0,7858	—1,2220	2,0078	0,6175	1,4933	4,0313
107	111	109,4180	1,7858	0,2038	1,5820	3,1891	0,0415	2,5027
105	110	107,9922	0,7858	—1,2220	2,0078	0,6175	1,4933	4,0313
104	106	107,2793	—3,2142	—1,9349	—1,2793	10,3311	3,7438	1,6366
109	110	110,8438	0,7858	1,6296	—0,8438	0,6175	2,6556	0,7120
107	107	109,4180	—2,2142	0,2038	—2,4180	4,9027	0,0415	5,8467
109	115	110,8438	5,7858	1,6296	4,1562	33,4755	2,6556	17,2740
113	113	113,6954	3,7858	4,4812	—0,6954	14,3323	20,0811	0,4836
100	103	104,4277	—6,2142	—4,7865	—1,4277	38,6163	22,9106	2,0383
Suma 1494	1529	1529,0004	0,0012*	0,0016*	—0,0004*	130,3574	65,4884	64,8758
Sredina 106,71	109,2142							

* U pravilu mora biti $\Sigma(Y-\bar{Y})$, $\Sigma(\hat{Y}-\bar{Y})$ i $\Sigma(Y-\hat{Y})$ jednak nuli ali uslijed korekture decimala kod računanja pojavljuju se male razlike

Dobivene podatke u tabeli 4.9 možemo pregledno prikazati na slijedeći način:

		Iznos varijacije	Postotak od ukupne varijacije
Neprotumačena devijacija . . .	$\Sigma d_{YX}^2 = \Sigma (Y - \hat{Y})^2 =$	64,87	49,766
Protumačena devijacija . . .	$\Sigma \hat{y}^2 = \Sigma (\hat{Y} - \bar{Y})^2 =$	65,48	50,234
Ukupna devijacija . . .	$\Sigma y^2 = \Sigma (Y - \bar{Y})^2 =$	130,35	100,000

Postotak devijacije, koji je rastumačen linearnom regresijom, dobit ćemo jednostavnije pomoću koeficijenta determinacije:

$$r^2 = \frac{\Sigma xy^2}{\Sigma x^2 \Sigma y^2} \quad \text{ili} \quad r^2 = 1 - \frac{\Sigma d_{YX}^2}{\Sigma y^2} \qquad (4.15)$$

Za podatke iznesene u tabeli 4.5 $r^2 = \frac{\Sigma xy^2}{\Sigma x^2 \Sigma y^2} = 0{,}7087^2$

$$r^2 = 0{,}5023$$

$$r^2 = 1 - \frac{\Sigma d_{YX}^2}{\Sigma y^2} = 1 - \frac{64{,}87}{130{,}35} = 1 - 0{,}4977 = 0{,}5023$$

U našem prikazanom primjeru od ukupne varijabilnosti neovisnog obilježja protumačeno je pomoću regresije 50,23%.

Kod krivolinijske regresije pomoću istih metoda može se izračunati dio ukupne varijabilnosti koji je rastumačen regresijom.

Ukupna devijacija može se dobiti zbrajanjem kvadrata devijacija pojedinog ovisnog obilježja od njegove aritmetičke sredine $\Sigma (Y - \bar{y})^2$. Neprotumačeni dio devijacije dobije se na taj način, da se na temelju regresijske jednadžbe izračunaju pojedine regresijske vrijednosti $\hat{Y}$ za svaku vrijednost ovisnog obilježja Y, te da se njihove diferencije kvadriraju i zbroje. $\Sigma (Y - \hat{Y})^2$. Ako ukupnu devijaciju izrazimo sa 100% izračunamo koliko postotaka otpada na neprotumačenu i protumačenu devijaciju.

Za prikazani primjer prirasta kod svinja i količine konzumirane hrane (tabele 4.3) navedene vrijednosti su već izračunate. U računu iza tabele 4.3 prikazana je vrijednost ukupne sume kvadrata, tj. ukupne devijacije (Σy^2) koja iznosi 259,60. U prikazu testa signifikantnosti skretanja od linearne regresije (tabela 4.4) neprotumačena devijacija (devijacija od krivolinijske regresije) iznosi 26,80. Protumačena devijacija izračunata je prilikom testiranja koeficijenta regresije te iznosi $\Sigma \hat{y}^2_{xx^2} = 232{,}80$. Na temelju ukupne, protumačene i neprotumačene devijacije izračunat će se postotak protumačene devijacije od ukupne devijacije:

	Iznos varijacije	Postotak od ukupne varijacije
$\Sigma d^2_{Yxx^2}$	26,80	10,32
$\Sigma \hat{y}^2_{xx^2}$	232,80	89,68
Σy^2	259,60	100

Zbroj kvadrata neprotumačene devijacije ($\Sigma d^2_{Yxx^2}$) i protumačene devijacije ($\Sigma \hat{y}^2_{xx^2}$) jednak je zbroju kvadrata ukupne devijacije (Σy^2).

Od sume kvadrata ukupne devijacije 89,68% protumačeno je regresijskom krivuljom.

Isti ćemo rezultat dobiti pomoću koeficijenta determinacije:

$$r^2{}_{YXX^2} = 1 - \frac{\Sigma\, d^2{}_{YXX^2}}{\Sigma\, y^2} = 1 - \frac{26,80}{259,50} = 0,8968 \qquad (4.16)$$

koji u obliku proporcije prikazuje dio varijabiliteta ovisnog obilježja koji je rastumačen pomoću regresije.

Ima pojava čije ćemo veze ispitati bilo pomoću regresije ili korelacije, ili ćemo upotrijebiti analizu korelacije i regresije. Međutim, ukoliko jednu varijablu možemo označiti kao ovisnu, a drugu kao neovisnu, potrebe informacije dat će nam obično analiza regresije, npr. veza između starosti i iskorištavanja hrane. Ako ispitujemo vezu između obilježja koja ne možemo označiti kao ovisno ili neovisno obilježje obično se primjenjuje korelacija, npr. veza između visine do grebena i dužine trupa.

DODATAK

SIMBOLI I FORMULE

Simboli s istim značenjem u svim formulama

μ	aritmetička sredina populacije
$\bar{x}$	aritmetička sredina uzorka
X	vrijednost pojedinog obilježja serije X
Y	vrijednost pojedinog obilježja serije Y
N	broj jedinica u populaciji
n	broj jedinica u uzorku
f	frekvencija
I	veličina razreda
d	kodirana vrijednost
Med	medijan
Mo	mode
h	harmonijska sredina
Q_1	donji kvartil
Q_3	gornji kvartil
σ	standardna devijacija populacije
s	standardna devijacija uzorka
x	$X - \bar{x}$
y	$Y - \bar{y}$
Σx^2	suma kvadrata
$s_{\bar{x}}$	standardna greška procjene aritmetičke sredine
p	proporcija u uzorku jedinica koje imaju obilježje
π	proporcija u populaciji jedinica koje imaju obilježje
q	proporcija u uzorku jedinica koje nemaju obilježje
$(1 - \pi)$	proporcija u populaciji jedinica koje nemaju obilježe
s_p	standardna greška procjene proporcije
$s_{\bar{x}_1 - \bar{x}_2}$	standardna greška procjene razliku dviju artimetičkih sredina

$s_{p_1 - p_2}$	standardna greška procjene razlika dviju proporcija
χ^2	Hi hvadrat
$\hat{Y}$	regresijska vrijednost
s_{XY}^2	varijanca linearne regresije
$s^2_{YXX^2}$	varijanca parabolične regresije
Σd^2_{YX}	$\Sigma (Y - \hat{Y})^2$, $\hat{Y}$ vrijednosti linearne regresije
$\Sigma d^2_{YXX'}$	$\Sigma (Y - \hat{Y})^2$, $\hat{Y}$ vrijednosti krivolinijske regresije
b	koeficijent regresije uzorka
β	koeficijent regresije populacije
c	koeficijent regresije uzorka
γ	koeficijent regresije populacije
r	koeficijent linearne korelacije
r^2	koeficijent determinacije (linearna korelacija)
$r^2_{YXX'}$	koeficijent determinacije (krivolinijska korelacija)
ρ	koeficijent linearne korelacije u populaciji
b_{12}	koeficijent regresije obilježja X_2 na X_1
b_{21}	koeficijent regresije obilježja X_1 na X_2

Formule

1.1	$\mu = \frac{\Sigma X}{N}$	aritmetička sredina populacije (negrupirani podaci)
1.2	$\bar{x} = \frac{\Sigma X}{n}$	aritmetička sredina uzoraka (negrupirani podaci)
1.3	$\mu = \frac{\Sigma fX}{N}$	aritmetička sredina populacije (grupirani podaci)
	X	vrijednost pojedine razredne sredine
1.4	$\bar{x} = \frac{\Sigma fX}{n}$	aritmetička sredina uzorka (grupirani podaci)
1.5	$\bar{x} = a + I\bar{d}$	aritmetička sredina uzorka računata kodiranjem
	a	jedna od razrednih sredina
	$\bar{d}$	aritmetička sredina kodiranih vrijednosti

1.6	$Med = X_L + \frac{(n_q - f_L)\, I}{f}$	medijan
	X_L	donja granica medijalnog razreda
	n_q	broj medijalnog obilježja
	f_L	vrijednost kumulativnog niza »manje od« u razredu prije medijalnog razreda
	f	frekvencija medijalnog razreda
1.7	$Mo = X_L + \frac{(f_2 - f_1)\, I}{(f_2 - f_1) + (f_2 - f_3)}$	mode
	X_L	donja granica modalnog razreda
	f_1	frekvencija razreda koji prehodi modalnom razredu
	f_2	frekvencija modalnog razreda
	f_3	frekvencija razreda iza modalnog
1.8	$h = \frac{n}{\Sigma \frac{f}{X}}$	harmonijska sredina (negrupirani podaci)
1.9	$h = \frac{n}{\Sigma \frac{f}{X}}$	harmonijska sredina (grupirani podaci)
1.10	$Q_1 = X_L + \frac{(n_{q1} - f_L)\, I}{f}$	donji kvartil
	X_L	donja granica razreda u kojoj se nalazi n_{q1}
	n_{q1}	broj obilježja koji naznačuje donji kvartil
	f_L	frekvencija kumulativnog niza u razredu prije kvartilnog razreda
	f	frekvencija kvartilnog razreda
1.11	$Q_3 = X_L + \frac{(n_{q3} - f_L)\, I}{f}$	gornji kvartil
	X_L	donja granica razreda u kojoj se nalazi n_{q3}
	n_{q3}	broj obilježja koji naznačuje gornji kvartil
	f_L	frekvencija kumulativnog niza u razredu prije kvartilnog razreda
	f	frekvencija kvartilnog razreda

1.12	$\sigma = \sqrt{\dfrac{\Sigma X^2}{N} - \left(\dfrac{\Sigma X}{N}\right)^2}$	standardna devijacija populacije (negrupirani podaci)
1.13	$\sigma = \sqrt{\dfrac{\Sigma fX^2}{N} - \left(\dfrac{\Sigma fX}{N}\right)^2}$	standardna devijacija populacije (grupirani podaci)
1.14	$s = \sqrt{\dfrac{\Sigma x^2}{n-1}}$	standardna devijacija uzorka (negrupirani podaci)
1.15	$s = \sqrt{\dfrac{\Sigma fx^2}{n-1}}$	standardna devijacija uzorka (grupirani podaci)
1.16	$\Sigma x^2 = \Sigma X^2 - \dfrac{(\Sigma X)^2}{n}$	suma kvadrata (negrupirani podaci)
1.17	$\Sigma fx^2 = \Sigma fX^2 - \dfrac{(\Sigma fX)^2}{n}$	suma kvadrata (grupirani podaci)
1.18	$s = I\, s_d$	standardna devijacija uzorka računata pomoću kodiranja
	$s_d = \sqrt{\dfrac{\Sigma fx_d{}^2}{n-1}}$	standardna devijacija kodiranih vrijednosti
	$\Sigma fx_d{}^2 = \Sigma fd^2 - \dfrac{(\Sigma fd)^2}{n}$	suma kvadrata kodiranih vrijednosti
1.19	$C = \dfrac{s}{\bar{x}} 100$	varijacioni koeficijent uzorka
2.1	$z = \dfrac{X - \mu}{\sigma}$	udaljenost pojedinog obilježja od μ izražena u σ jedinicama
2.2	$s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$	standardna greška procjene aritmetičke sredine
2.3	$z = \dfrac{\bar{x} - \mu}{s_{\bar{x}}}$	udaljenost $\bar{x}$ od μ izražena u jedinicama $s_{\bar{x}}$ $(n > 30)$
2.4	$t = \dfrac{\bar{x} - \mu}{s_{\bar{x}}}$	udaljenost $\bar{x}$ od μ izražena u jedinicama $s_{\bar{x}}$ $(n < 30)$
2.5	$\bar{x} - z\, s_{\bar{x}} \leq \mu \leq \bar{x} + z\, s_{\bar{x}}$ $\bar{x} \pm z\, s_{\bar{x}}$	intervalna procjena μ kada je $n > 30$
2.6	$\bar{x} - t\, s_{\bar{x}} \leq \mu \leq \bar{x} + t\, s_{\bar{x}}$ $\bar{x} \pm t\, s_{\bar{x}}$	intervalna procjena μ kada je $n < 30$
2.7	$p = \dfrac{m}{n}$	proporcija jedinica koje imaju obilježje
	m	broj jedinica koje imaju obilježje

2.8	$q = 1 - p$	proporcija jedinica koje nemaju obilježje
2.9	$s_p = \sqrt{\frac{pq}{n-1}}$	standardna greška procjene proporcije (mali uzorak)
2.10	$s_p = \sqrt{\frac{pq}{n}}$	standardna greška procjene proporcije (veliki uzorak)
2.11	$p - z s_p \leq \pi \leq p + z s_p$ $p \pm z\, s_p$	intervalna procjena π
3.1	$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}$	standardna greška procjene razlika dviju aritmetičkih sredina
3.2	$z = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$	diferencija aritmetičkih sredina dva uzorka izražena u jedinicama $s_{\bar{x}_1 - \bar{x}_2}$ $(n > 30)$
3.3	$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$	diferencija aritmetičkih sredina dva uzorka izražena u jedinicama $s_{\bar{x}_1 - \bar{x}_2}$ $(n < 30)$
3.4	$s_{p_1 - p_2} = \sqrt{\pi(1-\pi)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$	standardna greška procjene razlika dviju proporcija
3.5	$\pi = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$	procjena proporcija populacije izračunate iz proporcije 2 uzorka
3.6	$z = \frac{p_1 - p_2}{s_{p_1 - p_2}}$	diferencija proporcije 2 uzorka izražena u jedinicama $s_{p_1 - p_2}$ $(n > 30)$
3.7	$t = \frac{p_1 - p_2}{s_{p_1 - p_2}}$	diferencija proporcija dva uzorka izražena u jedinicama $s_{p_1 - p_2}$ $(n < 30)$
3.8	$X^2 = \frac{\Sigma (f - F)^2}{F}$	hi-kvadrat, testiranje hipotetičnog omjera
	f	broj jedinica koje imaju atribut
	F	hipotetične frekvencije
3.9	$F = \frac{\text{varijanca između grupa}}{\text{varijanca unutar grupa}}$	vrijednost F-distribucije, analiza varijance
3.10	$D = s_{\bar{x}}\, Q$	kritična vrijednost dozvoljene diferencije $\bar{x}$ dva uzorka uz 5% nivo signifikantnosti
	Q	tabelarna vrijednost

3.11 $n_0 = \frac{1}{a-1}\left(n. - \frac{\Sigma n_i^2}{n.}\right)$ — prosječan broj obilježja za jedan uzorak izračunat iz više uzoraka s nejednakim brojem obilježja.

n_0 — prosječan n u uzorku

a — broj uzoraka

$n.$ — ukupan broj obilježja

n_i — broj obilježja pojedinih grupa

3.12 $F = \frac{s_1^2 \text{ (veća varijanca)}}{s_2^2 \text{ (manja varijanca)}}$ — vrijednost F distribucije, testiranje homogenosti varijance 2 uzorka

3.13 $X^2 = 2{,}3026\,(n-1)\,(a \log \overline{s}^2 - \Sigma \log s^2)$ — hi-kvadrat, testiranje homogenosti varijance više uzoraka (uzorci s jednakim n)

2,3026 — konstanta ($\log_e 10$)

a — broj grupa

$\overline{s}^2$ — aritmetička sredina pojedinih s^2

3.14 $X^2 = 2{,}3026\,[(\log \overline{s}^2)\,\Sigma(n-1) - \Sigma(n-1)(\log s^2)]$ — hi-kvadrat, testiranje homogenosti varijance više uzoraka (uzorci sa nejednakim n)

4.1 $b = \frac{\Sigma xy}{\Sigma x^2}$ — koeficijent linearne regresije

4.2 $\hat{Y} = \overline{y} + b\,(X - \overline{x})$ — jednadžba linearne regresije

4.3 $s_{yx}^2 = \Sigma d_{yx}^2 / (n-2)$ — varijanca devijacije od regresije

4.4 $s_b = \frac{s_{yx}}{\sqrt{\Sigma x^2}}$ — standardna devijacija koeficijenta regresije

4.5 $t = \frac{b - \beta}{s_b}$ — vrijednost t-distribucije, devijacija b od β izražena u jedinicama s_b (linearna regresija)

4.6 $b - t\,s_b \leq \beta \leq b + t\,s_b$ — intervalna procjena koeficijenta regresije populacije

$b \pm t\,s_b$

4.7 $\hat{Y} = \overline{y} + b\,(X_1 - \overline{x}_1) + c\,(X_2 - \overline{x}_2)$ — jednadžba krivolinijske regresije

$X_2 = X_1^2$

$$b = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{D}$$

$$c = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{D}$$

$$D = (\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2$$

TREND

$Y_c = a + bX$

$b = \frac{\Sigma XY - \overline{X}\Sigma Y}{\Sigma X^2 - \overline{X}\Sigma X}$

$a = \overline{y} - b\overline{x}$

4.8 $F = \dfrac{\text{varijanca zakrivljenosti regresije}}{\text{varijanca devijacije od krivolinijske regresije}}$ — vrijednost F distribucije, test signifikantnosti skretanja od linearne regresije

4.9 $t = \dfrac{b - \beta}{s_b}$ — vrijednost t-distribucije, test signifikantnosti za koeficijent b (krivolinijska regresija)

$s_b^2 = s^2_{Yxx^2} \cdot c_b$

$c_b = \Sigma x_2^2 / D$

4.10 $t = \dfrac{c - \gamma}{s_c}$ — vrijednost t-distribucije, test signifikantnosti za koeficijent c (krivolinijska regresija)

$s_c^2 = s^2_{Yxx^2} \cdot c_c$

$c_c = \Sigma X_1^2 / D$

4.11 $r = \dfrac{\Sigma x_1 x_2}{\sqrt{(\Sigma x_1^2)(\Sigma x_2^2)}}$ — koeficijent linearne korelacije uzorka

4.12 $r = \sqrt{b_{21}\, b_{12}}$ — koeficijent linearne korelacije

b_{21} — koeficijent regresije obilježja X_2 na X_1

b_{12} — koeficijent regresije obilježja X_1 na X_2

4.13 $b_{21} = r\,\dfrac{s_Y}{s_X}$ — koeficijent regresije obilježja X_2 na X_1

s_Y — standardna devijacija obilježja Y

s_x — standardna devijacija obilježja X

4.14 $\dfrac{\Sigma x_1 x_2}{n - 1}$ — kovarijanca

4.15 $r^2 = 1 - \dfrac{\Sigma d_{yx}^2}{\Sigma y^2}$ — koeficijent determinacije (linearna korelacija)

Σd_{yx}^2 — neprotumačena devijacija

Σy^2 — ukupna devijacija

4.16 $r^2_{yxx^2} = 1 - \dfrac{\Sigma d^2_{yxx^2}}{\Sigma y^2}$ — koeficijent determinacije (krivolinijska korelacija)

$\Sigma d^2_{yxx^2}$ — neprotumačena devijacija

Σy^2 — ukupna devijacija

TREND

$Y_c = a + bX$

$b = \dfrac{\Sigma XY - \bar{x}\Sigma Y}{\Sigma X^2 - \bar{x}\Sigma X}$

$a = \bar{y} - b\bar{x}$

t jed.	Y vis. d. gr.	X	XY	X^2	Y_c
2	108	0	0	0	
4	111	1	111	1	
6	112	2	224	4	
8	115	3	345	9	
10	117	4	468	16	

STATISTIČKI IZRAZI*

Asimetrična distribucija

Distribucija kod koje se ne poklapaju vrijednosti aritmetičke sredine, medijana i moda.

Aritmetička sredina

Aritmetička sredina niza obilježja jednaka je njihovom zbroju podijeljenom sa brojem obilježja. Aritmetičku sredinu uzorka označavamo sa $\bar{x}$ a populacije sa grčkim slovom mi (μ).

Atributivno obilježje

Obilježja izražena opisno tj. kvalitativna karakteristika individua.

Bartlettov test

Test homogenosti varijance više uzoraka. Razradio ga je Bartlett 1937 godine.

Beta (β) koeficijent

Koeficijent u regresijskoj jednadžbi populacije ($\hat{Y} = \alpha + \beta X$, $\hat{Y} = \alpha + \beta X + \gamma X^2$).

Bimodalna distribucija

Distribucija frekvencija sa dva moda.

Centralna tendencija

Tendencija okupljanja numeričkih obilježja oko stanovite vrijednosti. Određuje se pomoću aritmetičke sredine, medijana, moda i harmonijske sredine.

Determinacija: Koeficijent determinacije

Dio ukupne varijabilnosti ovisnog obilježja (Y) koji je rastumačen utjecajem neovisnog obilježja (X) i izražen u obliku proporcije jest koeficijent determinacije.

Devijacija

Udaljenost jednog obilježja od stanovite vrijednosti (najčešće aritmetičke sredine). Devijacija se često izražava u jedinicama standardne devijacije.

---

* Neke definicije uzete su iz: Kendall and Buckland, A Dictionary of Statistical Terms

Disperzija

Stupanj raspršenosti obilježja koji se obično izražava pomoću standardne devijacije, interkvartila ili varijacione širine.

Dvostruka klasifikacija

Klasifikacija niza obilježja s obzirom na dva kriterija.

F-distribucija

Distribucija omjera varijanca dva nezavisna uzorka iz normalne populacije, koju je razradio R. A. Fisher.

F-test

Test signifikantnosti baziran na F distribuciji.

Frekvencija

Broj obilježja stanovitih svojstava, ili dio uzorka ili populacije koji pripada u stanoviti razred.

Gama (γ) koeficijent

Koeficijent u jednadžbi krivolinijske regresije populacije ($\hat{Y} = \alpha + \beta X + \gamma X^2$)

Granice povjerenja

Gornja i donja granica koja se formira pomoću »t« i zatvara interval povjerenja u kojem se kreće vrijednost parametara uz stanovitu vjerojatnost.

Greška

U statističkom izražavanju »greška« naznačuje razliku između dobivenih vrijednosti i »prave« ili »očekivane« vrijednosti. Razlike nastaju uslijed djelovanja slučajnosti.

Harmonijska sredina

Harmonijska sredina niza obilježja jest recipročna vrijednost aritmetičke sredine recipročnih vrijednosti obilježja.

Hi — kvadrat

Statistik koji je distribuiran kao hi-hvadrat (χ^2) tj. kao suma kvadrata nezavisnih standardnih normalnih varijanata.

Hi — kvadrat distribucija

Distribucija indeksa disperzije koji se označava grčkim slovom hi (χ^2). Postavio ju je Helmert (1875) a razradio Pearson (1899).

Hi — kvadrat test

Test signifikantnosti pomoću hi — kvadrata (χ^2). Upotrebljava se u raznim testiranjima no najčešće u dva i to:

a) testiranju opravdanosti razlika između dobivenih frekvencija u pojedinim grupama i hipotetičnih frekvencija,

b) testiranju opravdanosti razlika varijanca više uzoraka.

Hipoteza

Hipoteza u statistici naznačuje hipotezu koja se odnosi na parametre.

Interkvartil

Raspon između gornjeg i donjeg kvartila tj. vrijednost unutar koje se nalazi polovica ukupne frekvencije.

Interval povjerenja

Interval u kojem se uz stanovitu vjerojatnoću nalazi parametar.

Korelacija

U širem značenju korelacija naznačuje međusobnu ovisnost između kvantitativnih i kvalitativnih obilježja. U užem smislu, u kojem se češće upotrebljava, korelacija naznačuje vezu između numeričkih obilježja.

Korelacija: Koeficijent korelacije

Koeficijent korelacije jeste mjera za međuovisnost između dva obilježja. Koeficijent korelacije varira između —1 i +1. Vrijednost 0 naznačuje odsutnost korelacije, dok vrijednost +1 naznačuje potpunu pozitivnu a —1 potpunu negativnu korelaciju.

Koeficijent korelacije uzorka naznačuje se latinskim slovom »r«, a koeficijent korelacije populacije grčkim slovom ro (ρ).

Kovarijanca

Kovarijanca jest suma umnožaka odstupanja pojedinih obilježja dviju serija od njihovih aritmetičkih sredina, podijeljena stupnjem slobode.

$$\Sigma x_1 x_2 \quad n - 1$$

Kritična vrijednost

Vrijednost dobivena iz podataka uzorka odgovara stanovitom nivou signifikantnosti prilikom testiranja hipoteze. Na primjer ako je P = 0,01 onda je $t_{.01}$ kritična vrijednost za »t« na 1% nivou, a ako je P = 0,05 onda je $t_{.05}$ kritična vrijednost.

Krivolinijska regresija

Krivolinijska stohastička povezanost varijabla.

Krivolinijska korelacija

Krivolinijska korelacija izražava korelaciju obilježja kod kojih regresije nisu linearne.

Kvartil

Distribucija frekvencija može se podijeliti sa tri obilježja u četiri jednaka dijela. Vrijednost srednjeg obilježja naziva se medijan, a vrijednost gornjeg i donjeg obilježja naziva se gornji i donji kvartil. Gornji i donji kvartil dijele gornju, odnosno donju polovicu distribucije na dva jednaka dijela.

Linearna korelacija

Korelacija koja izražava vezu u slučajevima kada su odgovarajuće regresije linearne.

Linearna regresija

Veza između zavisnog (Y) i nezavisnog (X) obilježja čije kretanje ima oblik pravca te se može i izraziti jednadžbom pravca $\hat{Y} = a + bX$.

Medijan

Medijan je vrijednost obilježja koje dijeli distribuciju frekvencija u dvije polovice. Ukoliko je broj obilježja (n) neparan, medijan je jednak vrijednosti obilježja $(n+1)/2$, a ukoliko je n paran, medijan se dobije izračunavanjem sredine između vrijednosti obilježja $n/2$ i $(n+2)/2$.

Mode

Mode je vrijednost obilježja koja su zastupana s najvećom frekvencijom, tj. mode jedne distribucije jest vrijednost oko koje obilježja pokazuju najjaču tendenciju grupiranja.

Nezavisno obilježje

Izraz »nezavisno obilježje« ili »nezavisna varijabla« upotrebljava se za razlikovanje od »zavisnog obilježja« ili »zavisne varijable« u regresijskoj analizi. Kada se varijabla Y izražava kao funkcija varijable X_1, X_2, X_3 itd. uz stohastičko označavanje, X se naziva »nezavisno obilježje«.

Normalna distribucija

Kontinuirana distribucija frekvencija koju predstavlja slijedeća jednadžba:

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Jednadžbu je razradio De Moivre, a u daljnjoj razradi i primjeni velikog je udjela imao matematičar Gauss te se i naziva Gaussova krivulja.

Nul hipoteza

Navedeni izraz se odnosi na testiranje hipoteze tj. testiranje hipoteze da je stanoviti parametar jednak nuli.

Numerička obilježja

Obilježja koja se izražavaju brojkama (npr. težina, mliječnost, prirast).

Obilježje — Varijabla

Obilježje ili varijabla općenito naznačuje svaki kvantitet koji varira. Ona obilježja koja se izražavaju brojkama nazivaju se numerička obilježja. Uobičajeno je da se izraz obilježje ili varijabla primjenjuje i na karakteristike koje se ne mogu mjeriti (na primjer muško — žensko) te se u tom slučaju nazivaju atributivna obilježja.

Parametar

Parametri izražavaju distribuciju frekvencije populacije (populacijski parametri) ili stohastičke veze u populaciji (regresijski parametri populacije). Vrijednosti parametara najčešće se obilježavaju grčkim slovima kao na primjer: aritmetička sredina »mi« (μ), standardna devijacija »sigma« (σ), koeficijent regresije »alfa, beta, gama« (α, β, γ), koeficijent korelacije »ro« (ρ).

Pokretna sredina

Kada jednadžba izražava vezu između zavisnog i nezavisnog obilježja (regresija), često je poželjno da zavisno obilježje uspoređujemo sa sredinama koje dobijamo za svaku traženu vrijednost nezavisnog obilježja pomoću uvrštavanja vrijednosti nezavisnog obilježja u jednadžbu. Srednja vrijednost dobivena na izloženi način naziva se pokretna sredina.

Populacija

U statističkom rječniku izraz »populacija« upotrebljava se za svaki konačni ili beskonačni broj (skup) obilježja u kojem svaki član ima poznatu vjerojatnoću da se pojavi u uzorku. Prilikom raznih analiza mora se tačno znati definicija populacije na koju se odnose istraživanja.

Procjena

Procjena je zaključivanje o vrijednostima nepoznate populacije i to na temelju nepotpunih podataka tj. uzorka.

Procjena: Intervalna procjena

Ukoliko se izračunava interval unutar kojeg se nalazi parametar uz stanovitu vjerojatnost, postupak se naziva intervalna procjena.

Regresija

Izraz »regresija« prvi je primijenio Galton, u naznačavanju stanovitih veza prilikom ispitivanja heritabiliteta.

Pod regresijom se podrazumijeva analiza veza između dvije pojave, koje se međusobno nalaze u stohastičkoj vezi, te kada promjena vrijednosti jednog obilježja (nezavisnog X) prati promjena drugog obilježja (zavisnog Y).

Regresija: Koeficijent regresije

Koeficijent nezavisne varijable u jednadžbi regresije. Koeficijent linearne regresije naznačuje za koliko se promijeni zavisno obilježje, ako se nezavisno promijeni za jedinicu.

Reprezentativni uzorak

Uzorak čiji su članovi izabrani na taj način da je svaki član populacije imao jednaku mogućnost da bude izabran u uzorak tj. uzorak koji reprezentira populaciju.

Signifikantnost

Za stanoviti se učinik zaključuje da je signifikantan, ako je vrijednost koja se testira veća od kritične vrijednosti, što naznačuje da se odbacuje hipoteza da nema učinka.

Signifikantnost: Nivo signifikantnosti

Nivo signifikantnosti se mijenja, no najčešće se radi sa 5% i 1% nivoom signifikantnosti. Zaključak uz 5% nivo signifikantnosti znači da je naš zaključak ispravan u 95 slučajeva od 100, dok zaključak uz 1% nivo signifikantnosti znači da je zaključak ispravan u 99 slučajeva od 100.

Signifikantnost: Test signifikantnosti

Test signifikantnosti jest testiranje podataka uzoraka o hipotezi stanovitog djelovanja i to pomoću kritičnih vrijednosti.

Simetrična distribucija

Distribucija frekvencija u kojoj su jednake vrijednosti aritmetičke sredn-nje vrijednosti, medijana i moda.

Slučajni izbor

Izbor jedinica za uzorak prilikom kojeg svaki član populacije ima jednaku mogućnost da se pojavi u uzorku (Random izbor).

Standardna devijacija

Standardna je devijacija apsolutna mjera disperzije distribucije frekvencije koja se najčešće primjenjuje. Jednaka je drugom korijenu iz varijance.

Standardna greška

Standardna je greška jednaka drugom korijenu iz varijance distribucije vrijednosti »statistik« dobivenih analizom uzoraka i to podataka svih mogućih uzoraka koji se mogu odabrati u skladu sa shemom izbora.

Standardna greška procjene aritmetičke sredine

Drugi korijen iz varijance distribucije aritmetičkih sredina i to od svih mogućih uzoraka, naziva se standardna greška procjene aritmetičke sredine.

Standardna mjera

Standardnom mjerom, $(X - \mu)/\sigma$, izražava se udaljenost pojedinog obilježja (X) od aritmetičke sredine (μ) u jedinicama standardne devijacije (σ).

Statistik

Zbirna vrijednost izračunata iz uzorka opažanja, obično, ali ne i neophodno, kao procjena nekog parametra; funkcija vrijednosti uzorka.

Stohastička veza

Najjača veza između pojava jeste funkcionalna veza, tj. veza u kojoj vrijednosti jedne pojave odgovara tačno određena vrijednost druge pojave. U biološkim ispitivanjima ne nalazimo funkcionalne veze, već veze koje su labavije, tj. kod kojih vrijednostima jedne pojave ne odgovaraju tačno određene vrijednosti druge pojave.

Stupanj slobode

Izraz stupanj slobode u statistički je rječnik uveo R. A. Fisher, a upotrebljava se u više značenja koja se međusobno neznatno razlikuju. U nizu varijabla stupanj slobode jeste broj vrijednosti koje se mogu označiti proizvoljno unutar specifičnosti niza. Različite vrijednosti analiza imaju specifičan način izračunavanja stupnja slobode.

t- distribucija

Distribucija, koju je razradio W. S. Gosset (pisao je pod pseudonimom Student) 1808, god. a poslije ju je usavršio R. A. Fisher 1925 g. Studentova je distribucija simetrična kao i normalna distribucija. Devijacije aritmetičkih sredina od sredine populacije izražene su u jedinicama standardne greške, te se označavaju sa »t« koji iznosi: $t = (\bar{x} - \mu)/s_{\bar{x}}$. Unutar intervala koji je omeđen sa μ i određene udaljenosti od μ studentova distribucija obuhvaća manji postotak ukupnih vrijednosti od normalne distribucije. Postotak vrijednosti koji je t-distribucijom obuhvaćen unutar intervala μ i stanovite udaljenosti od nje ovisi o veličini uzorka.

Studentova distribucija ima vrlo veliku primjenu prilikom testiranja podataka dobivenih iz uzoraka s manjim brojem obilježja ($n<30$).

t — test

Test baziran na t— distribuciji.

Test

Testiranje hipoteze pomoću podataka iz uzoraka i kritičkih vrijednosti.

Unimodalna distribucija

Distribucija frekvencija sa jednim modom.

Uzorak

Dio populacije. Eksperimentalni se rad vrši najčešće pomoću »slučajnog uzorka«, tj. uzorka čije su jedinice dobivene slučajnim izborom.

Uzorak: Veličina uzorka

Broj obilježja u uzorku.

Varijacioni koeficijent

Predložio ga je K. Pearson 1895. godine. Izražava relativnu varijabilnost, a izračunava se diobom standardne devijacije s aritmetičkom sredinom, te se obično množi sa 100. Varijacioni koeficijent daje uvid u homogenost analiziranih obilježja bez obzira na njihove apsolutne vrijednosti.

Varijanca

Varijanca populacije je srednja vrijednost kvadrata udaljenosti pojedinih obilježja od aritmetičke sredine. Dobije se, ako se kvadrati diferencija između pojedinih obilježja od aritmetičke sredine zbroje i podijele sa brojem obilježja. Bilježi se grčkim slovom sigma ($\sigma^2 = \Sigma(X-\mu)^2/N$). Varijanca uzorka obilježava se malim latinskim slovom »s« i dobiva se diobom sume kvadrata diferencija obilježja od aritmetičke sredine sa stupnjem slobode ($s^2 = (X-\bar{x})^2/n-1$).

Varijanca: Analiza varijance

Ukupna varijacija niza obilježja, mjerena sumom kvadrata devijacije od aritmetičke sredine, može se u stanovitim slučajevima podijeliti u pojedine komponente koje ovise o uzrocima varijacije u vezi s klasifikacijom obilježja.

Ta se dioba naziva analiza varijance. U užem smislu pod analizom varijance naziva se dioba ukupne sume kvadrata devijacija od aritmetičke sredine u pojedine komponente.

Varijabla

Općenito se pod nazivom varijabla podrazumijeva svaka količina koja varira.

Varijanta

Količina koja može imati bilo koju vrijednost unutar specificiranog niza i specificirane relativne frekvencije ili vjerojatnosti.. Slučajna varijabla.

Zavisno obilježje

Izraz »zavisno obilježje« ili »zavisna varijabla« u regresijskoj analizi naznačuje razliku od »nezavisnog obilježja«. Kada varijabla Y izražava funkciju varijable X_1, X_2, X_3 itd. uz stohastičko označavanje Y se naziva zavisno obilježje.

TABLICE KVADRATA I DRUGIH KORIJENA

Tablica A: Kvadrati

Uputa za upotrebu tablice A: kvadrati

U tablici A direktno se očitavaju kvadrati jednoznamenkastih, dvoznamenkastih i troznamenkastih brojeva. Jedinice se nalaze u glavi tablice, dok se desetice i stotice nalaze u pretkoloni. Npr. kvadrat broja 9 nalazi se u koloni u čijoj se glavi nalazi broj 9, a u pretkoloni broj 0. U križanju kolone i reda

Tablica A: KVADRATI

	0	1	2	3	4
0	0	1	4	9	16
1	100	121	144	169	196
2	400	441	484	529	576
3	900	961	1024	1089	1156
4	1600	1681	1764	1849	1936
5	2500	2601	2704	2809	2916
6	3600	3721	3844	3969	4096
7	4900	5041	5184	5329	5476
8	6400	6561	6724	6889	7056
9	8100	8281	8464	8649	8836
10	10000	10201	10404	10609	10816
11	12100	12321	12544	12769	12996
12	14400	14641	14884	15129	15376
13	16900	17161	17424	17689	17956
14	19600	19881	20164	20449	20736
15	22500	22801	23104	23409	23716
16	25600	25921	26244	26569	26896
17	28900	29241	29584	29929	30276
18	32400	32761	33124	33489	33856
19	36100	36481	36864	37249	37636
20	40000	40401	40804	41209	41616
21	44100	44521	44944	45369	45796
22	48400	48841	49284	49729	50176
23	52900	53361	53824	54289	54756
24	57600	58081	58564	59049	59536
25	62500	63001	63504	64009	64516
26	67600	68121	68644	69169	69696
27	72900	73441	73984	74529	75076
28	78400	78961	79524	80089	80656
29	84100	84681	85264	85849	86436

očitamo traženu vrijednost koja iznosi 81. Kvadrat broja 29 nalazi se u koloni u čijoj se glavi nalazi brojka 9, a u pretkoloni broj 2. Kvadrat broja 829 nalazi se u koloni u čijoj se glavi nalazi broj 9, a u pretkoloni broj 82.

Kvadrat četveroznamenkastog broja može se brzo izračunati pomoću formule $(a \pm b)^2 = a^2 \pm 2\,ab + b^2$. Kvadrat broja npr. 3829 možemo izračunati na slijedeći način: $(3820 + 9)^2 = 14592400 + 68760 + 81 = 14661241$.

(Fisher and Yates: Statistical tables, New York 1957)

5	6	7	8	9
25	36	49	64	81
225	256	289	324	361
625	676	729	784	841
1225	1296	1369	1444	1521
2025	2116	2209	2304	2401
3025	3136	3249	3364	3481
4225	4356	4489	4624	4761
5625	5776	5929	6084	6241
7225	7396	7569	7744	7921
9025	9216	9409	9604	9801
11025	11236	11449	11664	11881
13225	13456	13689	13924	14161
15625	15876	16129	16384	16641
18225	18496	18769	19044	19321
21025	21316	21609	21904	22201
24025	24336	24649	24964	25281
27225	27556	27889	28224	28561
30625	30976	31329	31684	32041
34225	34596	34969	35344	35721
38025	38416	38809	39204	39601
42025	42436	42849	43264	43681
46225	46656	47089	47524	47961
50625	51076	51529	51984	52441
55225	55696	56169	56644	57121
60025	60516	61009	61504	62001
65025	65536	66049	66564	67081
70225	70756	71289	71824	72361
75625	76176	76729	77284	77841
81225	81796	82369	82944	83521
87025	87616	88209	88804	89401

Tablica A:

	0	1	2	3	4
30	90000	90601	91204	91809	92416
31	96100	96721	97344	97969	98596
32	102400	103041	103684	104329	104976
33	108900	109561	110224	110889	111556
34	115600	116281	116964	117649	118336
35	122500	123201	123904	124609	125316
36	129600	130321	131044	131769	132496
37	136900	137641	138384	139129	139876
38	144400	145161	145924	146689	147456
39	152100	152881	153664	154449	155236
40	160000	160801	161604	162409	163216
41	168100	168921	169744	170569	171396
42	176400	177241	178084	178929	179776
43	184900	185761	186624	187489	188356
44	193600	194481	195364	196249	197136
45	202500	203401	204304	205209	206116
46	211600	212521	213444	214369	215296
47	220900	221841	222784	223729	224676
48	230400	231361	232324	233289	234256
49	240100	241081	242064	243049	244036
50	250000	251001	252004	253009	254016
51	260100	261121	262144	263169	264196
52	270400	271441	272484	273529	274576
53	280900	281961	283024	284089	285156
54	291600	292681	293764	294849	295936
55	302500	303601	304704	305809	306916
56	313600	314721	315844	316969	318096
57	324900	326041	327184	328329	329476
58	336400	337561	338724	339889	341056
59	348100	349281	350464	351649	352836
60	360000	361201	362404	363609	364816
61	372100	373321	374544	375769	376996
62	384400	385641	386884	388129	389376
63	396900	398161	399424	400689	401956
64	409600	410881	412164	413449	414736

5	6	7	8	9
93025	93636	94249	94864	95481
99225	99856	100489	101124	101761
105625	106276	106929	107584	108241
112225	112896	113569	114244	114921
119025	119716	120409	121104	121801
126025	126736	127449	128164	128881
133225	133956	134689	135424	136161
140625	141376	142129	142884	143641
148225	148996	149769	150544	151321
156025	156816	157609	158404	159201
164025	164836	165649	166464	167281
172225	173056	173889	174724	175561
180625	181476	182329	183184	184041
189225	190096	190969	191844	192721
198025	198916	199809	200704	201601
207025	207936	208849	209764	210681
216225	217156	218089	219024	219961
225625	226576	227529	228484	229441
235225	236196	237169	238144	239121
245025	246016	247009	248004	249001
255025	256036	257049	258064	259081
265225	266256	267289	268324	269361
275625	276676	277729	278784	279841
286225	287296	288369	289444	290521
297025	298116	299209	300304	301401
308025	309136	310249	311364	312481
319225	320356	321489	322624	323761
330625	331776	332929	334084	335241
342225	343396	344569	345744	346921
354025	355216	356409	357604	358801
366025	367236	368449	369664	370881
378225	379456	380689	381924	383161
390625	391876	393129	394384	395641
403225	404496	405769	407044	408321
416025	417316	418609	419904	421201

Tablica A

	0	1	2	3	4
65	422500	423801	425104	426409	427716
66	435600	436921	438244	439569	440896
67	448900	450241	451584	452929	454276
68	462400	463761	465124	466489	467856
69	476100	477481	478864	480249	481636
70	490000	491401	492804	494209	495616
71	504100	505521	506944	508369	509796
72	518400	519841	521284	522729	524176
73	532900	534361	535824	537289	538756
74	547600	549081	550564	552049	553536
75	562500	564001	565504	567009	568516
76	577600	579121	580644	582169	583696
77	592900	594441	595984	597529	599076
78	608400	609961	611524	613089	614656
79	624100	625681	627264	628849	630436
80	640000	641601	643204	644809	646416
81	656100	657721	659344	660969	662596
82	672400	674041	675684	677329	678976
83	688900	690561	692224	693889	695556
84	705600	707281	708964	710649	712336
85	722500	724201	725904	727609	729316
86	739600	741321	743044	744769	746496
87	756900	758641	760384	762129	763876
88	774400	776161	777924	779689	781456
89	792100	793881	795664	797449	799236
90	810000	811801	813604	815409	817216
91	828100	829921	831744	833569	835396
92	846400	848241	850084	851929	853776
93	864900	866761	868624	870489	872356
94	883600	885481	887364	889249	891136
95	902500	904401	906304	908209	910116
96	921600	923521	925444	927369	929296
97	940900	942841	944784	946729	948676
98	960400	962361	964324	966289	968256
99	980100	982081	984064	986049	988036

5	6	7	8	9
429025	430336	431649	432964	434281
442225	443556	444889	446224	447561
455625	456976	458329	459684	461041
469225	470596	471969	473344	474721
483025	484416	485809	487204	488601
497025	498436	499849	501264	502681
511225	512656	514089	515524	516961
525625	527076	528529	529984	531441
540225	541696	543169	544644	546121
555025	556516	558009	559504	561001
570025	571536	573049	574564	576081
585225	586756	588289	589824	591361
600625	602176	603729	605284	606841
616225	617796	619369	620944	622521
632025	633616	635209	636804	638401
648025	649636	651249	652864	654481
664225	665856	667489	669124	670761
680625	682276	683929	685584	687241
697225	698896	700569	702244	703921
714025	715716	717409	719104	720801
731025	732736	734449	736164	737881
748225	749956	751689	753424	755161
765625	767376	769129	770884	772641
783225	784996	786769	788544	790321
801025	802816	804609	806404	808201
819025	820836	822649	824464	826281
837225	839056	840889	842724	844561
855625	857476	859329	861184	863041
874225	876096	877969	879844	881721
893025	894916	896809	898704	900601
912025	913936	915849	917764	919681
931225	933156	935089	937024	938961
950625	952576	954529	956484	958441
970225	972196	974169	976144	978121
990025	992016	994009	996004	998001

TABLICA B: DRUGI KORIJENI

Uputa za upotrebu tablice B: Drugi korijeni

Drugi korijeni za brojeve od 3 znamenke i za brojeve 10 puta veće u tablici B poredani su zajedno tj. paralelno. U prvom se redu nalaze drugi korijeni za brojeve s neparnim brojem znamenaka, a u drugom redu za brojeve sa parnim brojem znamenaka. Tablica je podijeljena u dva dijela: prvi dio koji služi za vađenje drugog korijena iz tri mjesta, i drugi dio koji služi za četvrto mjesto.

Prilikom vađenja drugog korijena odredit ćemo a) neparan ili paran broj znamenaka za koji vadimo drugi korijen, pa prema tome očitavamo vrijednosti u prvom redu ili drugom redu, b) prve dvije znamenke pročitamo u pretkoloni, a treću znamenku u glavi kolone i to u prvom dijelu tablica unutar brojeva 0 do 9, te u naznačenom redu i koloni očitamo traženu vrijednost, d) odredimo mjesto decimalne tačke i na taj način dobivamo drugi korijen od zadanog broja, e) ukoliko tražimo drugi korijen četveroznamenkastog broja prvo ćemo očitati tabličnu vrijednost kao da je četvrta znamenka jednaka 0. Korektura uslijed četvrte znamenke izvršit će se na taj način da će se u drugom dijelu tablice u glavi naći vrijednosti četvrte znamenke (unutar jedan do pet) i u prije određenom redu očitati vrijednost koja se pribroji pročitanoj tabličnoj vrijednosti za prve tri znamenke. Ukoliko je četvrta znamenka veća od 5, postupit ćemo na isti način, no morat ćemo u drugom dijelu tablice u dva puta očitati korekturu. Ako je npr. četvrta znamenka 9, možemo tražiti vrijednosti u koloni u čijoj se glavi nalazi 5, a zatim 4 ($9 = 5 + 4$).

Primjer 1: $\sqrt{532}$

U prvom redu (neparan broj znamenaka) kod broja 53 u pretkoloni i u koloni u čijoj se glavi nalazi broj 2 (prvi dio tablice) pročitamo broj $2306\bar{5}$. Odredimo decimalni zarez i dobijemo da je $\sqrt{532} = 23,065$.

Primjer 2: $\sqrt{53,2}$

U drugom redu (paran broj znamenaka lijevo od decimalnog zareza) kod broja 53 u pretkoloni i u koloni u čijoj se glavi nalazi broj 2 (prvi dio tablice) pročitamo broj 72938. Odredimo decimalni zarez i dobijemo da je

$\sqrt{53,2} = 7,2938$.

Primjer 3: $\sqrt{0,00532}$

Postupak je isti kao u zadatku 2, jer imamo dvije nule sa desne strane decimalne tačke, paran broj, te traženu vrijednost očitavamo u drugom redu. Razlika nastaje prilikom određivanja decimalnog zareza te je

$\sqrt{0,00532} = 0,072938$.

Primjer 4: $\sqrt{0,0532}$

Postupak je kao u zadatku 1, jer imamo jednu nulu sa desne strane decimalne tačke, neparan broj, te traženu vrijednost očitavamo u prvom redu. Razlika je u decimalnoj tački te je $\sqrt{0,0532} = 0,23065$.

Primjer 5: $\sqrt{53,27}$

Očitamo vrijednost za $\sqrt{53,2}$ kao u zadatku 2. Zatim u drugom dijelu tablice očitamo korekturu za četvrtu znamenku. Broj 7 moramo podijeliti u sastavne dijelove npr. 4 i 3 ($7 = 4 + 3$), te u drugom redu kod broja 53 u pretkoloni i u kolonama u drugom dijelu tablice u čijoj se glavi nalaze brojevi 4 i 3 pročitamo vrijednosti 27 i 21. Prije očitanoj vrijednosti (za broj 53,2) 72938 pribrojimo vrijednosti 27 i 21, što iznosi $72938 + 27 + 21 = 72986$. Odredimo decimalnu tačku, te dobijemo da je $\sqrt{53,27} = 7,2986$.

Tablica B:

DRUGI KORIJENI

(Fisher and Yates: Statistical tables, New York, 1957)

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
10	10000	10050	10100	10149	10198	10247	10296	10344	10392	10440	5	10	15	20	24
	31623	31780	31937	32094	32249	32404	32558	32711	32863	33015	15	31	46	62	77
11	10488	10536	10583	10630	10677	10724	10770	10817	10863	10909	5	9	14	19	23
	33166	33317	33466	33615	33764	33912	34059	34205	34351	34496	15	30	44	59	74
12	10954	11000	11045	11091	11136	11180	11225	11269	11314	11358	4	9	13	18	22
	34641	34785	34928	35071	35214	35355	35496	35637	35777	35917	14	28	43	57	71
13	11402	11446	11489	11533	11576	11619	11662	11705	11747	11790	4	9	13	17	22
	36056	36194	36332	36469	36606	36742	36878	37014	37148	37283	14	27	41	55	68
14	11832	11874	11916	11958	12000	12042	12083	12124	12166	12207	4	8	13	17	21
	37417	37550	37683	37815	37947	38079	38210	38341	38471	38601	13	26	39	53	66
15	12247	12288	12329	12369	12410	12450	12490	12530	12570	12610	4	8	12	16	20
	38730	38859	38987	39115	39243	39370	39497	39623	39749	39875	13	25	38	51	64
16	12649	12689	12728	12767	12806	12845	12884	12923	12961	13000	4	8	12	16	20
	40000	40125	40249	40373	40497	40620	40743	40866	40988	41110	12	25	37	49	62
17	13038	13077	13115	13153	13191	13229	13266	13304	13342	13379	4	8	11	15	19
	41231	41352	41473	41593	41713	41833	41952	42071	42190	42308	12	24	36	48	60
18	13416	13454	13491	13528	13565	13601	13638	13675	13711	13748	4	7	11	15	18
	42426	42544	42661	42778	42895	43012	43128	43243	43359	43474	12	23	35	47	58
19	13784	13820	13856	13892	13928	13964	14000	14036	14071	14107	4	7	11	14	18
	43589	43704	43818	43932	44045	44159	44272	44385	44497	44609	11	23	34	45	57
20	14142	14177	14213	14248	14283	14318	14353	14387	14422	14457	4	7	10	14	18
	44721	44833	44944	45056	45166	45277	45387	45497	45607	45717	11	22	33	44	55
21	14491	14526	14560	14595	14629	14663	14697	14731	14765	14799	3	7	10	14	17
	45826	45935	46043	46152	46260	46368	46476	46583	46690	46797	11	22	32	43	54
22	14832	14866	14900	14933	14967	15000	15033	15067	15100	15133	3	7	10	13	17
	46904	47011	47117	47223	47329	47434	47539	47645	47749	47854	11	21	32	42	53
23	15166	15199	15232	15264	15297	15330	15362	15395	15427	15460	3	7	10	13	16
	47958	48062	48166	48270	48374	48477	48580	48683	48785	48888	10	21	31	41	52

Tablica B:

DRUGI KORIJENI — nastavak

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
24	15492	15524	15556	15588	15620	15652	15684	15716	15748	15730	3	6	10	13	16
	48990	49092	49193	49295	49396	49497	49598	49699	49800	49900	10	20	30	40	51
25	15811	15843	15875	15906	15937	15969	16000	16031	16062	16093	3	6	9	13	16
	50000	50100	50200	50299	50398	50498	50596	50695	50794	50892	10	20	30	40	50
26	16125	16155	16186	16217	16248	16279	16310	16340	16371	16401	3	6	9	12	15
	50990	51088	51186	51284	51381	51478	51575	51672	51769	51865	10	19	29	39	49
27	16432	16462	16492	16523	16553	16583	16613	16643	16673	16703	3	6	9	12	15
	51962	52058	52154	52249	52345	52440	52536	52631	52726	52820	10	19	29	38	48
28	16733	16763	16793	16823	16852	16882	16912	16941	16971	17000	3	6	9	12	15
	52915	53009	53104	53198	53292	53385	53479	53572	53666	53759	9	19	28	38	47
29	17029	17059	17088	17117	17146	17176	17205	17234	17263	17292	3	6	9	12	15
	53852	53944	54037	54129	54222	54314	54406	54498	54589	54681	9	18	28	37	46
30	17321	17349	17378	17407	17436	17464	17493	17521	17550	17578	3	6	9	11	14
	54772	54863	54955	55045	55136	55227	55317	55408	55498	55588	9	18	27	36	45
31	17607	17635	17664	17692	17720	17748	17776	17804	17833	17861	3	6	8	11	14
	55678	55767	55857	55946	56036	56125	56214	56303	56391	56480	9	18	27	36	45
32	17889	17916	17944	17972	18000	18028	18055	18083	18111	18138	3	6	8	11	14
	56569	56657	56745	56833	56921	57009	57096	57184	57271	57359	9	18	26	35	44
33	18166	18193	18221	18248	18276	18303	18330	18358	18385	18412	3	5	8	11	14
	57446	57533	57619	57706	57793	57879	57966	58052	58138	58224	9	17	26	35	43
34	18439	18466	18493	18520	18547	18574	18601	18628	18655	18682	3	5	8	11	14
	58310	58395	58481	58566	58652	58737	58822	58907	58992	59076	9	17	26	34	43
35	18708	18735	18762	18788	18815	18841	18868	18894	18921	18947	3	5	8	11	13
	59161	59245	59330	59414	59498	59582	59666	59749	59833	59917	8	17	25	34	42
36	18974	19000	19026	19053	19079	19105	19131	19157	19183	19209	3	5	8	10	13
	60000	60083	60166	60249	60332	60415	60498	60581	60663	60745	8	17	25	33	41

Tablica B:

DRUGI KORIJENI — nastavak

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
37	19235	19261	19287	19313	19339	19365	19391	19416	19442	19468	3	5	8	10	13
	60828	60910	60992	61074	61156	61237	61319	61400	61482	61563	8	16	25	33	41
38	19494	19519	19545	19570	19596	19621	19647	19672	19698	19723	3	5	8	10	13
	61644	61725	61806	61887	61968	62048	62129	62209	62290	62370	8	16	24	32	40
39	19748	19774	19799	19824	19849	19875	19900	19925	19950	19975	3	5	8	10	13
	62450	62530	62610	62690	62769	62849	62929	63008	63087	63166	8	16	24	32	40
40	20000	20025	20050	20075	20100	20125	20149	20174	20199	20224	2	5	7	10	12
	63246	63325	63403	63482	63561	63640	63718	63797	63875	63953	8	16	24	31	39
41	20248	20273	20298	20322	20347	20372	20396	20421	20445	20469	2	5	7	10	12
	64031	64109	64187	64265	64343	64420	64498	64576	64653	64730	8	16	23	31	39
42	20494	20518	20543	20567	20591	20616	20640	20664	20688	20712	2	5	7	10	12
	64807	64885	64962	65038	65115	65192	65269	65345	65422	65493	8	15	23	31	38
43	20736	20761	20785	20809	20833	20857	20881	20905	20928	20952	2	5	7	10	12
	65574	65651	65727	65803	65879	65955	66030	66106	66182	66257	8	15	23	30	38
44	20976	21000	21024	21048	21071	21095	21119	21142	21166	21190	2	5	7	10	12
	66332	66408	66483	66558	66633	66708	66783	66858	66933	67007	8	15	22	30	38
45	21213	21237	21260	21284	21307	21331	21354	21378	21401	21424	2	5	7	9	12
	67082	67157	67231	67305	67380	67454	67528	67602	67676	67750	7	15	22	30	37
46	21448	21471	21494	21517	21541	21564	21587	21610	21633	21656	2	5	7	9	12
	67823	67897	67971	68044	68118	68191	68264	68337	68411	68484	7	15	22	29	37
47	21679	21703	21726	21749	21772	21794	21817	21840	21863	21886	2	5	7	9	12
	68557	68629	68702	68775	68848	68920	68993	69065	69138	69210	7	15	22	29	36
48	21909	21932	21954	21977	22000	22023	22045	22068	22091	22113	2	5	7	9	11
	69282	69354	69426	69498	69570	69642	69714	69785	69857	69929	7	14	22	29	36
49	22136	22159	22181	22204	22226	22249	22271	22293	22316	22338	2	4	7	9	11
	70000	70071	70143	70214	70285	70356	70427	70498	70569	70640	7	14	21	28	36

Tablica B:

DRUGI KORIJENI — nastavak

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
50	22361	22383	22405	22428	22450	22472	22494	22517	22539	22561	2	4	7	9	11
	70711	70781	70852	70922	70993	71063	71134	71204	71274	71344	7	14	21	28	35
51	22583	22605	22627	22650	22672	22694	22716	22738	22760	22782	2	4	7	9	11
	71414	71484	71554	71624	71694	71764	71833	71903	71972	72042	7	14	21	28	35
52	22804	22825	22847	22869	22891	22913	22935	22956	22978	23000	2	4	7	9	11
	72111	72180	72250	72319	72388	72457	72526	72595	72664	72732	7	14	21	28	34
53	23022	23043	23065	23087	23108	23130	23152	23173	23195	23216	2	4	6	9	11
	72801	72870	72938	73007	73075	73144	73212	73280	73348	73417	7	14	21	27	34
54	23238	23259	23281	23302	23324	23345	23367	23388	23409	23431	2	4	6	9	11
	73485	73553	73621	73689	73756	73824	73892	73959	74027	74095	7	14	20	27	34
55	23452	23473	23495	23516	23537	23558	23580	23601	23622	23643	2	4	6	8	11
	74162	74229	74297	74364	74431	74498	74565	74632	74699	74766	7	13	20	27	34
56	23664	23685	23707	23728	23749	23770	23791	23812	23833	23854	2	4	6	8	11
	74833	74900	74967	75033	75100	75166	75233	75299	75366	75432	7	13	20	27	33
57	23875	23896	23917	23937	23958	23979	24000	24021	24042	24062	2	4	6	8	10
	75498	75565	75631	75697	75763	75829	75895	75961	76026	76092	7	13	20	26	33
58	24083	24104	24125	24145	24166	24187	24207	24228	24249	24269	2	4	6	8	10
	76158	76223	76289	76354	76420	76485	76551	76616	76681	76746	7	13	20	26	33
59	24290	24310	24331	24352	24372	24393	24413	24434	24454	24474	2	4	6	8	10
	76811	76877	76942	77006	77071	77136	77201	77266	77330	77395	6	13	19	26	32
60	24495	24515	24536	24556	24576	24597	24617	24637	24658	24678	2	4	6	8	10
	77460	77524	77589	77653	77717	77782	77846	77910	77974	78038	6	13	19	26	32
61	24698	24718	24739	24759	24779	24799	24819	24839	24860	24880	2	4	6	8	10
	78102	78166	78230	78294	78358	78422	78486	78549	78613	78677	6	13	19	26	32
62	24900	24920	24940	24960	24980	25000	25020	25040	25060	25080	2	4	6	8	10
	78740	78804	78867	78930	78994	79057	79120	79183	79246	79310	6	13	19	25	32
63	25100	25120	25140	25159	25179	25199	25219	25239	25259	25278	2	4	6	8	10
	79373	79436	79498	79561	79624	79687	79750	79812	79875	79937	6	13	19	25	31

Tablica B:

DRUGI KORIJENI — nastavak

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
64	25298	25318	25338	25357	25377	25397	25417	25436	25456	25475	2	4	6	8	10
	80000	80062	80125	80187	80250	80312	80374	80436	80498	80561	6	12	19	25	31
65	25495	25515	25534	25554	25573	25593	25612	25632	25652	25671	2	4	6	8	10
	80623	80685	80747	80808	80870	80932	80994	81056	81117	81179	6	12	19	25	31
66	25690	25710	25729	25749	25768	25788	25807	25826	25846	25865	2	4	6	8	10
	81240	81302	81363	81425	81486	81548	81609	81670	81731	81792	6	12	18	25	31
67	25884	25904	25923	25942	25962	25981	26000	26019	26038	26058	2	4	6	8	10
	81854	81915	81976	82037	82098	82158	82219	82280	82341	82401	6	12	18	24	30
68	26077	26096	26115	26134	26153	26173	26192	26211	26230	26249	2	4	6	8	10
	82462	82523	82583	82644	82704	82765	82825	82885	82946	83006	6	12	18	24	30
69	26268	26287	26306	26325	26344	26363	26382	26401	26420	26439	2	4	6	8	10
	83066	83126	83187	83247	83307	83367	83427	83487	83546	83606	6	12	18	24	30
70	26458	26476	26495	26514	26533	26552	26571	26589	26608	26627	2	4	6	8	9
	83666	83726	83785	83845	83905	83964	84024	84083	84143	84202	6	12	18	24	30
71	26646	26665	26683	26702	26721	26739	26758	26777	26796	26814	2	4	6	7	9
	84261	84321	84380	84439	84499	84558	84617	84676	84735	84794	6	12	18	24	30
72	26833	26851	26870	26889	26907	26926	26944	26963	26981	27000	2	4	6	7	9
	84853	84912	84971	85029	85088	85147	85206	85264	85323	85381	6	12	18	23	29
73	27019	27037	27055	27074	27092	27111	27129	27148	27166	27185	2	4	6	7	9
	85440	85499	85557	85615	85674	85732	85790	85849	85907	85965	6	12	17	23	29
74	27203	27221	27240	27258	27276	27295	27313	27331	27350	27368	2	4	5	7	9
	86023	86081	86139	86197	86255	86313	86371	86429	86487	86545	6	12	17	23	29
75	27386	27404	27423	27441	27459	27477	27495	27514	27532	27550	2	4	5	7	9
	86603	86660	86718	86776	86833	86891	86948	87006	87063	87121	6	12	17	23	29
76	27568	27586	27604	27622	27641	27659	27677	27695	27713	27731	2	4	5	7	9
	87178	87235	87293	87350	87407	87464	87521	87579	87636	87693	6	11	17	23	29

Tablica B:

DRUGI KORIJENI — nastavak

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
77	27749	27767	27785	27803	27821	27839	27857	27875	27893	27911	2	4	5	7	9
	87750	87807	87864	87920	87977	88034	88091	88148	88204	88261	6	11	17	23	28
78	27928	27946	27964	27982	28000	28018	28036	28054	28071	28089	2	4	5	7	9
	88318	88374	88431	88487	88544	88600	88657	88713	88769	88826	6	11	17	23	28
79	28107	28125	28142	28160	28178	28196	28213	28231	28249	28267	2	4	5	7	9
	88882	88938	88994	89051	89107	89163	89219	89275	89331	89387	6	11	17	22	28
80	28284	28302	28320	28337	28355	28373	28390	28408	28425	28443	2	4	5	7	9
	89443	89499	89554	89610	89666	89722	89778	89833	89889	89944	6	11	17	22	28
81	28460	28478	28496	28513	28531	28548	28566	28583	28601	28618	2	4	5	7	9
	90000	90056	90111	90167	90222	90277	90333	90388	90443	90499	6	11	17	22	28
82	28636	28653	28671	28688	28705	28723	28740	28758	28775	28792	2	3	5	7	9
	90554	90609	90664	90719	90774	90830	90885	90940	90995	91049	6	11	16	22	28
83	28810	28827	28844	28862	28879	28896	28914	28931	28948	28965	2	3	5	7	9
	91104	91159	91214	91269	91324	91378	91433	91488	91542	91597	5	11	16	22	27
84	28983	29000	29017	29034	29052	29069	29086	29103	29120	29138	2	3	5	7	9
	91652	91706	91761	91815	91869	91924	91978	92033	92087	92141	5	11	16	22	27
85	29155	29172	29189	29206	29223	29240	29257	29275	29292	29309	2	3	5	7	9
	92195	92250	92304	92358	92412	92466	92520	92574	92628	92682	5	11	16	22	27
86	29326	29343	29360	29377	29394	29411	29428	29445	29462	29479	2	3	5	7	8
	92736	92790	92844	92898	92952	93005	93059	93113	93167	93220	5	11	16	22	27
87	29496	29513	29530	29547	29563	29580	29597	29614	29631	29648	2	3	5	7	8
	93274	93327	93381	93434	93488	93541	93595	93648	93702	93755	5	11	16	21	27
88	29665	29682	29698	29715	29732	29749	29766	29783	29799	29816	2	3	5	7	8
	93803	93862	93915	93968	94021	94074	94128	94181	94234	94287	5	11	16	21	27
89	29833	29850	29866	29883	29900	29917	29933	29950	29967	29983	2	3	5	7	8
	94340	94393	94446	94499	94552	94604	94657	94710	94763	94816	5	11	16	21	26

Tabela B:

DRUGI KORIJENI — nastavak

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
90	30000	30017	30033	30050	30067	30083	30100	30116	30133	30150	2	3	5	7	8
	94868	94921	94974	95026	95079	95131	95184	95237	95289	95341	5	11	16	21	26
91	30166	30183	30199	30216	30232	30249	30265	30282	30299	30315	2	3	5	7	8
	95394	95446	95499	95551	95603	95656	95708	95760	95812	95864	5	10	16	21	26
92	30332	30348	30364	30381	30397	30414	30430	30447	30463	30480	2	3	5	7	8
	95917	95969	96021	96073	96125	96177	96229	96281	96333	96385	5	10	16	21	26
93	30496	30512	30529	30545	30561	30578	30594	30610	30627	30643	2	3	5	7	8
	96437	96488	96540	96592	96644	96695	96747	96799	96850	96902	5	10	16	21	26
94	30659	30676	30692	30708	30725	30741	30757	30773	30790	30806	2	3	5	7	8
	96954	97005	97057	97108	97160	97211	97263	97314	97365	97417	5	10	15	21	26
95	30822	30838	30854	30871	30887	30903	30919	30935	30952	30968	2	3	5	6	8
	97468	97519	97570	97622	97673	97724	97775	97826	97877	97929	5	10	15	20	26
96	30984	31000	31016	31032	31048	31064	31081	31097	31113	31129	2	3	5	6	8
	97980	98031	98082	98133	98184	98234	98285	98336	98387	98438	5	10	15	20	25
97	31145	31161	31177	31193	31209	31225	31241	31257	31273	31289	2	3	5	6	8
	98489	98539	98590	98641	98691	98742	98793	98843	98894	98944	5	10	15	20	25
98	31305	31321	31337	31353	31369	31385	31401	31417	31432	31448	2	3	5	6	8
	98995	99045	99096	99146	99197	99247	99298	99348	99398	99448	5	10	15	20	25
99	31464	31480	31496	31512	31528	31544	31559	31575	31591	31607	2	3	5	6	8
	99499	99549	99599	99649	99700	99750	99800	99850	99900	99950	5	10	15	20	25

LITERATURA

1. R. L. Anderson — T. A. Bancroft: Statistical Theory in Research, New York, 1952
2. S. Barić: Analiza homogenosti uzorka i populacije u stočarskim istraživanjima pomoću varijacionog koeficijenta, »Agronomski glasnik« br. 8—9, Zagreb, 1964
3. W. G. Cochran — G. M. Cox: Experimental Designs, New York, 1956
4. E. W. Crampton — D. E. Brady — L. E. Casida — N. L. Hazel: Problems, Techniques, and Experimental Designs in Animal Investigations, Journal of Animal Science, Vol. 9, No. 4, 1950
5. F. E. Croxton — D. J. Cowden: Applied General Statistics, New Jersey, 1955
6. R. A. Fisher: Statistical Methods for Research Workers, New York, 1958
7. R. A. Fisher — F. Yates: Statistical Tables, New York. 1957
8. G. Guerrieri: Testing Hypothesis: Theoretical aspects and application Patterns, Mimeo notes, 1964
9. W. R. Harvey — J. L. Lush: Genetic Correlation Between Type and Production in Jersey Cattle, Journal of Dairy Science, Vol. XXXV, No. 3 1952
10. M. G. Kendall — W. R. Buckland: A Dictionary of Statistical Terms, London, 1960
11. A. Linden: Planen und Auswerten von Versuchen, Basel, 1959
12. J. L. Lush — L. D. Mc Gilliard: Proving Dairy Sires and Dams, Journal of Dairy Science, Vol. XXXVIII, No. 2, 1955
13. J. L. Lush: The Genetics of Populations, Mimeo notes, 1948
14. J. L. Lush: Animal Breeding Plans, Ames, IOWA, 1956
15. P. A. P. Moran: The Statistical Processes of Evolutionary Theory, Oxford. 1962
16. A. Robertson: Weighting in the Estimation of Variance Components in the Unbalanced Single Classification, Biometrics, Vol. 18, No. 3, 1962
17. A. Robertson: Statistics, Mimeo notes, 1962
18. H. L. Le Roy: Statistische Methoden als informationsquelle, Schweizerischen Landwirtschaftlischen Forschung, No. 3, 1962
19. T. Salvemini: Variability measures — The curve of Lorenz, Mimeo notes, 1964
20. T. Salvemini: Types of sampling and reliability problems, Mimeo notes, 1964
21. T. Salvemini: Depedence between two characteristics connexion and contigency, Mimeo notes, 1964
22. T. Salvemini: Regression and correlation, Mimeo notes, 1964
23. V. Serdar: Udžbenik Statistike, Zagreb, 1961

24. G. W. Snedecor: Statistical Methods Applied to Experiments in Agriculture and Biology, Ames, IOWA, 1956

25. A. Tavčar: Biometrika u poljoprivredi, Zagreb, 1946

26. A O. Toole: Elementary Practical Statistics, New York, 1964

27. V. Vranić: Vjerojatnost i statistika, Zagreb, 1958

28. A. E. Waugh: Elements of Statistical Method, New York, 1952

29. E. B. Wilson: An Introduction to scientific Research, New York, 1952

30. H. D. Young: Statistical Treatment of Experimental Data, New York, 1962

SADRŽAJ