Sadržaj

1. POPULACIJA, DISTRIBUCIJA FREKVENCIJA 1

1.1. Varijable i tipovi varijabli 3
1.2. Populacija i uzorak 4
1.3. Distribucija frekvencija 6

2. MJERILA KOJA OPISUJU DISTRIBUCIJU 11

2.1. Mjerila centralne tendencije ili sredine 13
2.1.1. Aritmetička sredina ($\overline{x}$) 14
2.1.2. Medijana (*Med*) 18
2.1.3. Modus (*Mod*) 19
2.2. Mjerila varijabilnosti ili disperzije 21
2.2.1. Varijacijska širina ili raspon (*vš*) 23
2.2.2. Standardna devijacija (s) i varijanca (s^2) 24
2.2.3. Varijacijski koeficijent (*cv*) 29
2.3. Mjerila centralne tendencije i disperzije uzoraka i parametri populacije 31

3. DISTRIBUCIJA VJEROJATNOSTI I NEKE VAŽNIJE TEORETSKE DISTRIBUCIJE 33

3.1. Binominalna distribucija 36
3.2. Teoretska normalna distribucija frekvencija 39
3.2.1. Položaj pojedine varijante u distribuciji; z-vrijednosti 44

4. PROCJENA PARAMETARA POPULACIJE PREKO VRIJEDNOSTI IZ UZORKA 47

4.1. Granice pouzdanosti srednje vrijednosti 52
4.2. Studentova "t-distribucija" 54

5. NULTA HIPOTEZA I TESTIRANJE NULTE HIPOTEZE 61

5.1. Testiranje nulte hipoteze o razlici između prosječnih vrijednosti 64
5.2. t-test 66
5.2.1. Nezavisni uzorci 74
5.2.2. Zavisni uzorci 77

6. ANALIZA VARIJABILNOSTI 81

6.1. F-distribucija i F-test 83
6.2. Analiza varijance (Anova) 87

7. REGRESIJA I KORELACIJA 99

7.1. Regresija 102
7.2. Korelacija 107

8. PLANIRANJE I IZVOĐENJE EKSPERIMENATA 121
8.1. Kako proučiti problem? 123
8.2. Principi planiranja pokusa 129

9. PLANOVI ILI SHEME POKUSA 137
9.1. Potpuno slučajni raspored 140
9.2. Slučajni blokni raspored 154
9.3. Latinski kvadrat 166
9.4. Latinski pravokutnik 177

10. VIŠEFAKTORIJALNI POKUSI 187
10.1. Dvofaktorijalni pokusi 202
10.2. Trofaktorijalni pokusi 226

11. POKUSI S RAZDIJELJENIM PARCELAMA 245
11.1. Split-plot 248
11.2. Split-blok ili strip-plot 260
11.3. Split-split plot 272

12. POKUSI PONOVLJENI U VREMENU ILI PROSTORU 283

LITERATURA 293

TABLICE 295
Tablica A: Površina ispod normalne krivulje 297
Tablica B: Granične vrijednosti za *t* 298
Tablica C1: Granične vrijednosti za F (P=0.05) 299
Tablica C2: Granične vrijednosti za F (P=0.01) 302

POJMOVNIK 305

Đurđica Vasilj: „Biometrika i eksperimentiranje u bilonogojstvu“, Hrvatsko agronomsko društvo, Zagreb, 2000.

1 Varijable, populacija, distribucija frekvencija

2

3

4

5

6

7

1.1. Varijable i tipovi varijabli
1.2. Populacija i uzorak
1.3. Distribucija frekvencija

1.1. Varijable i tipovi varijabli

Da bi se neki fenomen proučio, potrebno je prvenstveno raspolagati s mnoštvom informacija o različitim relevantnim svojstvima. Kako su biološka svojstva pod jakim utjecajem i unutarnjih i vanjskih činilaca, to su podložna promjenama, tj. većem ili manjem variranju. Zato ih se i naziva **varijablama**. Nemoguće bi ih bilo nabrojiti sve, jer svaka disciplina ima specifične grupe varijabli kao: morfološke izmjere, koncentracije kemijskih sastojaka, različita fizikalna očitanja na instrumentima, frekvencija genotipova, itd.

Ipak, sve vrste varijabli s kojima se poljoprivredni stručnjaci kao predstavnici jedne od bioloških struka susreću, moguće je svrstati u sljedeće skupine:

1. Mjerne varijable
2. Rangovi
3. Obilježja ili atributi
4. Izvedene varijable

Mjerne varijable su, kako im već samo ime kaže, ona svojstva koja se na bilo koji način mogu izmjeriti i brojčano izraziti. U toj skupini razlikujemo ***kontinuirane*** i ***nekontinuirane*** mjerne varijable.

Kontinuiranim varijablama zovemo sva ona svojstva koja se daju izraziti decimalnim brojem (dakle, između dvije susjedne vrijednosti moguće je mnoštvo drugih, ovisno o preciznosti mjerila). Takva su, na primjer, svojstva: težina ploda (25.5 g), duljina klasa (10.7 cm), promjer cvijeta (25.0 mm), gomolja (3.2 cm) ili visina stabljike (1.75 m), prinos (prirod) (4.5 dt/ha), sadržaj šećera ili kiseline u plodu (soku)(18.8 % ili 6.7 %), itd., itd.

Nekontinuirane varijable, koje se često nazivaju i *diskretne* varijable, predstavljaju svojstva koja se izražavaju isključivo cijelim brojem. Takva su na primjer: broj cvjetova na biljci (npr. 5), broj zrna u klasu

(60) ili klipu (360), pjege ili mrlje na listu ili plodu (8), biljke jedne biljne vrste u nekoj biljnoj zajednici (25), broj mahuna u etaži (20) ili zrna u mahuni (8), itd.

Rangovima se izražavaju one varijable koje nije moguće direktno mjeriti, ali se mogu svrstati u određeni redoslijed po veličini (sa nejednako velikim intervalima). Na primjer intenzitet zaraženosti ili naoblaka i sl. izražavaju se u skalama 0-5 ili 0-9 i sl.

One varijable koje se ne daju ni mjeriti ni rangirati, nego samo opisati kvalitativno, zovu se **atributi ili obilježja** (boja, spol i sl.).

U nekim slučajevima se i obilježja mogu pretvoriti u mjerne varijable. Ako se na primjer boja izrazi duljinom spektra, ona postaje mjerna varijabla.

Izvedene ili derivirane varijable su ona svojstva koja se najčešće izražavaju pomoću dvije ili više neovisno izmjerenih mjernih varijabli. Postoje različiti načini za to, no najčešći su oblici *omjeri* i *postoci*. Omjer stavlja u odnos dvije varijable i taj kvocijent predstavlja deriviranu varijablu. Treba imati na umu da je on neimenovan broj, tj. nije izražen u nikakvim mjernim jedinicama. Postoci su također na neki način omjeri. Od izvedenih varijabli u našoj se struci često služimo indeksima. Primjer za to je indeks lisne površine, koji se dobije stavljajući u odnos dimenzije lista. Omjeri često mogu biti jedini ispravan put u interpretiranju izvjesnih bioloških problema (nasljeđivanje oblika ploda nekih vrsta ispravno je proučavati jedino u vidu proučavanja omjera visine i opsega, a ne na temelju neovisnih izmjera dimenzija).

1.2. Populacija i uzorak

Podaci u biološkim istraživanjima obično predstavljaju opažene (izmjerene) vrijednosti na svakoj pojedinoj najmanjoj jedinki neke cjeline.

Mjerimo li duljinu svakog klipa jednog hibrida kukuruza ili opseg svakog ploda neke voćne vrste, zapravo vršimo opažanja. Ako pri tome obuhvatimo sve moguće individue neke cjeline, govorimo o **populaciji**.

Mjerimo li pak samo dio svih mogućih individua neke cjeline, provodimo opažanja na **uzorku**. Uzorak je dakle samo dio populacije. Populacija može biti *manja* ili *veća, konačna* ili *beskonačna*. Primjer za beskonačnu populaciju su, recimo, sve biljke neke sorte pšenice ili svi ljudi na svijetu (svi predstavnici roda Homo sapiens) i sl. Konačnu populaciju bi predstavljale, recimo, biljke jednog hibrida kukuruza posijane na našem pokusnom polju, ili svi studenti Agronomskog fakulteta i sl. U biološkim istraživanjima se polazi od postavke da su populacije vrlo velike, skoro beskonačne, a ujedno podvrgnute djelovanju niza činilaca koji uzrokuju raznolikost u njima. Ta različitost individua u uzorku ili populaciji zove se **varijabilnost**, a svaka jedinka **varijanta** (koju ćemo označavati s "x").

Zbog raznolikosti između pojedinih individua, nemoguće je uzorak potpuno poistovjetiti s populacijom. Upravo zbog varijabilnosti, moguće je da svojstva jednog uzorka budu sličnija populaciji kojoj uzorci pripadaju nego svojstva drugog uzorka. Koji od dva uzorka bolje predstavlja populaciju, to se, dakako, ne zna. Ipak se međutim, zahvaljujući izvjesnim statističkim postupcima, može na osnovi uzoraka zaključivati o karakteristikama populacija, odnosno s određenim postotkom sigurnosti utvrditi koliko se vrijednosti dobivene iz uzoraka podudaraju s vrijednostima populacije.

Često se nailazi na termine "*mali*" i "*veliki uzorak*". Teško je reći što je mali a koliki je velik uzorak, iako se najčešće pod malim uzorkom podrazumijeva onaj koji sadrži do 50 osnovnih podataka ili varijanata.

Veličina uzorka predstavljena je dakle brojem varijanata u uzorku, a bilježit ćemo je s n.

Upravo zbog velikih ili beskonačno velikih populacija prisiljeni smo raditi s uzorcima, te na temelju proučavanja uzoraka zaključivati o populacijama.

Kako je uzorak samo dio populacije - cjeline, to vrijednosti koje se eventualno izračunaju iz uzorka, mogu biti manje ili više slične pravim vrijednostima populacije. Određeni, statistički postupak pomaže da se utvrdi koliko se dobivene vrijednosti za uzorak, podudaraju s pravim vrijednostima populacije iz koje je uzorak uzet.

Najvažnije je pri tom osigurati dobar uzorak, tj. takav uzorak koji će dobro predstavljati populaciju, a to znači sadržavati sve tipične varijante populacije iz koje je uzet. Takav se uzorak zove **reprezentativni uzorak**. Doći do reprezentativnog uzorka moguće je samo ako se svakoj varijanti populacije da ista prilika da bude uključena u uzorak. To se osigurava *slučajnim uzimanjem varijanata iz populacije*, a primjenom objektivnog postupka, kao što je izvlačenje brojevima označenih ceduljica ili žetona iz šešira ili kutije, korištenjem tablice slučajnih brojeva i sl. U protivnom se dogodi da se samo pojedini individui izaberu. Naime, subjektivnost ili pristranost je uvijek moguća. Ako recimo iz vagona plodova ili klipova treba osigurati dobar uzorak, uvijek postoji opasnost da uzimamo ljepše plodove ili klipove. To se može izbjeći izdvajanjem svakog 10-og ili 15-og ploda ili klipa koji redom dođe pod ruku.

Ako se recimo iz jednog velikog jata pilića želi dobiti uzorak od dvadesetak, bilo bi pogrešno otvoriti vrata i pustiti ih da izlaze, te prvih dvadesetak uzeti kao uzorak. To najvjerojatnije ne bi bio reprezentativan uzorak, jer je za očekivati da oni snažniji, temperamentniji, nagrnu na vrata prvi.

Ovo su samo dva primjera koji ukazuju na potreban oprez pri osiguravanju reprezentativnog uzorka.

Veličina uzorka ovisi prvenstveno o različitosti ili varijabilnosti svojstva koje se pruočava. Naime, ako je varijabilnost mala (varijante međusobno slične), dovoljan je i manji uzorak. Ako je pak varijabilnost svojstva velika (dakle individui se međusobno jako razlikuju), potrebno je osigurati što je moguće veći uzorak kako bi se u njemu našli svi predstavnici cjeline (populacije).

Dobar uzorak je jedino onaj koji je dovoljno reprezentativan, a to znači slučajan i dovoljno velik.

1.3. Distribucija frekvencija

Ako se neku varijablu proučava na varijantama uzorka, opažene vrijednosti (izmjere) moguće je na naki način srediti radi boljeg uvida u ponašanje te varijable.

PRIMJER 1.3.1.

Kao primjer neka posluže vrijednosti dvadeset varijanata jedne nekontinuirane mjerne varijable i to:

76	72	73	75	66	68	70	71	71	73
70	71	69	70	69	69	70	67	72	74

Imamo dakle $x_1, x_2, \dots, x_{20}$, odnosno veličina uzorka $n = 20$.

Lako je uočiti da se neke varijante pojavljuju češće od drugih. Prikažu li se podaci tog uzorka na malo drugačiji način, poredavši ih rastućim redom i s naznakom koliko se puta pojavljuju, to je još očitije:

vrijednost varijante	66	67	68	69	70	71	72	73	74	75	76
frekvencija (f)	1	1	1	3	4	3	2	2	1	1	1

Učestalost pojavljivanja pojedinih varijanata u uzorku zove se **frekvencija** (f), pa je tako dobivena **razdioba** ili **distribucija frekvencija** u uzorku. Kako je ovo očito mjerna varijabla, to je tzv. **kvantitativna distribucija frekvencija**.

Distribucija frekvencija kod varijabli koje smo nazvali obilježjima, zove se **kvalitativna distribucija frekvencija**. Primjer za takvu distribuciju bila bi zastupljenost pripadnika pojedinih biljnih vrsta na određenoj površini neke biljne zajednice.

PRIMJER 1.3.2.

Kako će se kvantitativna distribucija frekvencija najčešće susretati, treba je malo podrobnije opisati. Poslužimo se podacima o duljini klipa kukuruza (u cm) izmjerenim na uzorku od 50 klipova (n = 50):

13.5	16.0	17.0	17.5	20.0	21.0	18.5	15.0	12.0	14.0
12.5	18.0	15.5	17.5	14.0	16.0	18.0	19.0	19.5	18.5
14.5	20.5	18.5	16.0	15.5	15.5	16.5	17.5	17.5	17.0
17.5	18.0	17.5	13.0	18.5	18.0	18.0	17.0	17.5	16.5
16.0	19.0	18.0	16.5	16.0	16.5	17.0	17.5	16.5	17.0

Podaci ovog uzorka mogu se srediti tako, da se najprije utvrdi u kojim se granicama izmjerene vrijednosti kreću. Najkraći izmjereni klip je bio 12.0 cm, a najdulji 21.0 cm. Razlika između najveće i najmanje varijante u uzorku (najduljeg i najkraćeg izmjerenog klipa) je **raspon** ili **varijacijska širina** ($vš$). U ovom primjeru

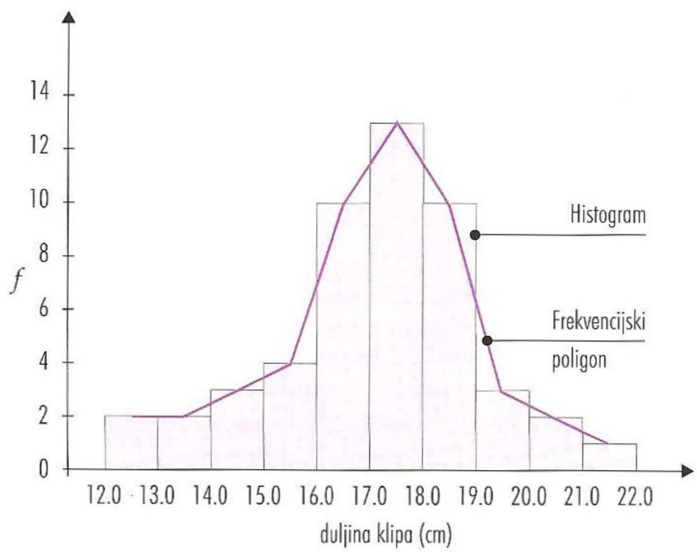
$$vš \; = \; 21.0 - 12.0 \; = \; 9.0 \text{ cm}.$$

Ako se varijacijsku širinu podijeli na nekoliko grupa (odsječaka), pa sve podatke unese u grupe ovisno o njihovim vrijednostima, dobiva se **varijacijski red** za ovu varijablu.

vrijednost varijante	12	13	14	15	16	17	18	19	20	21	22
	//	//	///	////	///// /////	///// ///// ///	///// /////	///	//	/	
frekvencija (f)	2	2	3	4	10	13	10	3	2	1	

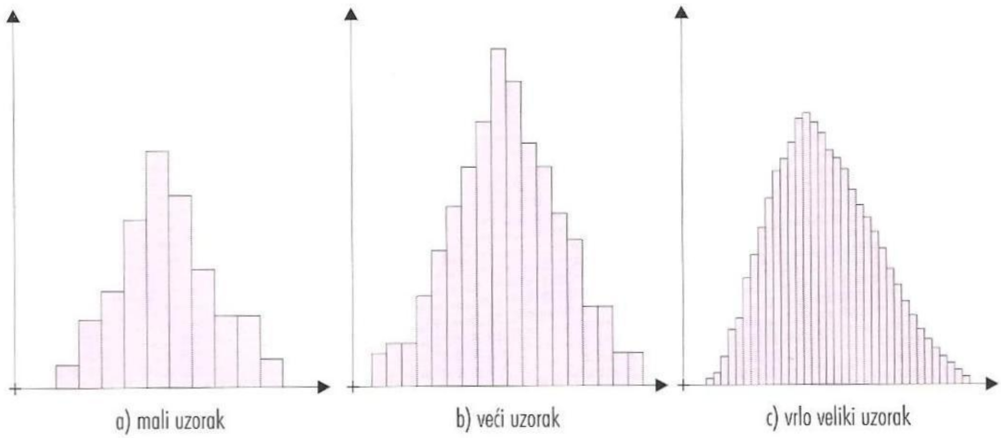
Ove grupe (odsječci ili dijelovi) nazivaju se ***razredima varijacijskog reda***, a raspon između gornje i donje granice razreda ***razredni razmak*** ili ***razredni interval*** ili ***areal*** (a). Tako smo za ovaj primjer konstruirali varijacijski red s 10 razreda, s razrednim razmakom ili arealom a = 1.0 cm.

Ovaj bi se varijacijski red mogao prikazati i grafički tako da se iznad odgovarajućih vrijednosti varijable na apscisi, na ordinatu nanesu frekvencije u svakom razredu, tj.



SLIKA 1.
Histogram i frekvencijski poligon varijanata uzorka za duljinu klipa (n=50)

Takav prikaz distribucije frekvencija zove se **histogram**, odnosno **frekvencijski poligon** (ako se spoje sredine pojedinih razreda). Kako



SLIKA 2.
Izgled histograma ovisno o veličini uzorka

raste broj izmjerenih varijanata, tako histogram poprima sve pravilniji oblik koji se kod vrlo velikog n-a može aproksimirati krivuljom.

U proučavanju nekog svojstva korisno je najprije konstruirati frekvencijski poligon, iz kojeg se može lako i brzo utvrditi da li se radi o simetričnoj ili nekoj "iskrivljenoj" distribuciji. Ova posljednja može pokazivati npr. da u uzorku (populaciji) postoji selekcija za ili protiv izvjesnih individua (koje pripadaju u "iskrivljeni" dio distribucije) ili je asimetričnost uzrokovana primijenjenom skalom mjerenja.

Postoji nekoliko karakterističnih oblika distribucije frekvencija. O najvažnijima od njih bit će riječi u jednom od sljedećih poglavlja.

1

2 Mjerila koja opisuju distribuciju

3

4

5

6

7

2.1. Mjerila centralne tendencije ili sredine
2.2. Mjerila varijabilnosti ili disperzije
2.3. Mjerila centralne tendencije i disperzije uzoraka i parametri populacije

Već histogram i frekvencijski poligon mogu dati grubu sliku distribucije varijanata u uzorku. Međutim, da bi se neka distribucija mogla detaljnije opisati, potrebno je imati i neke numeričke pokazatelje.

Iz navedenih primjera distribucija frekvencija, lako se može uočiti da se veliki broj varijanata grupira oko sredine (ili ako je distribucija frekvencija prikazana histogramom ili poligonom frekvencija, u srednjim stupcima), a manji dio prema krajevima.

U svakom uzorku (populaciji) djeluju činioci koji uvjetuju grupi-ranje varijanata oko sredine, kao i oni koji uvjetuju "raspršivanje" unutar varijacijske širine.

Grupiranje varijanata u sredini distribucije zove se **centralna tendencija**, a raspršivanje prema krajevima **disperzija**.

Posljedice ovih procesa daju se izmjeriti procjenom parametara-mjerila centralne tendencije i mjerila disperzije.

2.1. Mjerila centralne tendencije ili sredine

To su zapravo pokazatelji sredine distribucije, dakle prosječne ili srednje vrijednosti. Najčešće se primjenjuju: **aritmetička sredina**, **medijana**, **modus**, te **harmonijska** i **geometrijska sredina**.

2.1.1. Aritmetička sredina ($\bar{x}$)

Aritmetička sredina ($\bar{x}$) je prosječna vrijednost svih varijanata. Računa se zbrajanjem svih podataka i dijeljenjem dobivene sume s ukupnim brojem varijanata u uzorku (n). To je najčešće korišteno mjerilo od svih navedenih mjerila sredine, pa smo u svakodnevnom žargonu skloni jedino nju zvati srednjom vrijednosti ili prosjekom.

Na ovom ćemo mjestu objasniti neke simbole i pojednostaviti ih radi lakšeg korištenja.

Svaku varijantu u uzorku (dakle pojedinačna opažanja u uzorku) označit ćemo dogovorno sa x_i. To je dakle i-ti podatak uzorka. Stoga se uzorak od recimo 8 varijanata može pisati kao

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$$

Ili općenito,

$$x_1, x_2, \dots, x_n$$

Zbrajanje (sumiranje) svih opažanja u uzorku označava se:

$$\sum_{i=1}^{i=n} x_i \text{ , ili}$$

$$\sum_{i=1}^{i=n} x_i = x_1 + x_2 + \dots + x_n$$

$\sum$ je veliko slovo "sigma" grčke abecede, a označava zbroj svih podataka ispred kojih stoji (i = 1 znači da podatke treba zbrojiti od prvog, a $i = n$ do zadnjeg).

Tako na primjer

$$\sum_{i=3}^{i=8} x_i$$

označava zbroj svih podataka od trećeg do osmog.

Za praktičara (a to znači za nematematičara), pogodno je zbog jednostavnosti ispustiti ostale oznake, te sumu svih podataka uzorka označavati jednostavno kao $\sum x$, čime se automatski podrazumijeva da treba zbrojiti sve podatke (od 1 do n, tj. od prvog do posljednjeg).

$\bar{x}$ ćemo upotrebljavati kao oznaku za aritmetičku sredinu. Napi-šemo li dakle aritmetičku sredinu pomoću dogovorenih simbola, tada je ona

$$\bar{x} = \frac{\sum x}{n}$$

Tako bi aritmetička sredina za primjer 1.3.1. o podacima jedne mjerne nekontinuirane varijable bila

$$\bar{x} = \frac{76 + \ldots + 74}{20} = \frac{1416}{20} = 70.8$$

Za primjer 1.3.2., kojim je predstavljen uzorak kontinuirane varijable (duljina klipa kukuruza)

$$\bar{x} = \frac{13.5 + 16.0 + \ldots + 17.0}{50} = \frac{843.5}{50} = 16.87 \text{ cm}$$

Ako su, međutim, osnovni podaci sređeni tako da se znaju i frekvencije pojedinih varijanata, aritmetička sredina se izračunava po formuli

$$\bar{x} = \frac{\sum fx}{n}$$

gdje $\sum fx$ predstavlja zbroj umnožaka svih vrijednosti varijanata (x) i pripadajućih frekvencija (f).

Za primjer 1.3.1., to bi bilo

$$\bar{x} = \frac{(66 \text{x} 1) + (67 \text{x} 1) + \ldots + (75 \text{x} 1) + (76 \text{x} 1)}{20} = \frac{1416}{20} = 70.8$$

U primjeru 1.3.2., gdje se radi o uzorku podataka kontinuirane varijable grupiranim u razrede varijacijskog reda, frekvencija u svakom razredu znači broj varijanata čije su vrijednosti unutar donje i gornje granice svakog razreda.

vrijednost varijante	12	13	14	15	16	17	18	19	20	21	22
f	2	2	3	4	10	13	10	3	2	1	
x'	−5	−4	−3	−2	−1		+1	+2	+3	+4	
fx'	−10	−8	−9	−8	−10		10	6	6	4	

$$\sum fx' = -45 + 26 = -19$$

Tako, recimo, frekvencija 2 u razredu od 12.0 do 13.0 cm znači da su u uzorku dva klipa čija duljina je 12.0 cm ili više, a manja od

13.0 cm, itd., za svaki razred. Zato se svaki razred varijacijskog reda može predstaviti samo vrijednošću sredine razreda.

Iz uzorka varijanata sređenih u ovakav varijacijski red aritmetička sredina se računa kao:

$$\bar{x} = \bar{x}_{pr} + \left(a \cdot \frac{\sum fx'}{n} \right)$$

$\bar{x}_{pr}$ je provizorna, aproksimativna ili privremena vrijednost aritmetičke sredine

x' je intervalna vrijednost razreda (s obzirom na razred u kojem je $\bar{x}_{pr}$, a koji je označen nulom)

Aproksimativnu ili privremenu aritmetičku sredinu odredimo "od oka" u razredu u kojem mislimo da će biti prava. To je za očekivati u razredu s najvećom frekvencijom. U primjeru 1.3.2. to je razred od 17.0 do 18.0 cm; njegova je vrijednost 17.5 cm.

$\bar{x}_{pr}$=17.5 cm

a =1.0 cm

$\sum fx' = -19$

Prava aritmetička sredina je dakle

$$\bar{x} = 17.5 + \left(1.0 \cdot \frac{-19}{50} \right) = 17.12 \text{ cm}$$

Ovo su samo različiti načini računanja aritmetičke sredine, prilagođeni načinu na koji su sređeni podaci uzorka. Dakako, čim je uzorak veći, raspoređenost varijanata unutar varijacijske širine je pravilnija. Uz dovoljan broj razreda u varijacijskom redu, vrijednosti izračunane na različite načine, sve manje će se razlikovati.

Aritmetička sredina ili prosjek uzorka predstavlja sredinu svih opaženih vrijednosti. Svojstva aritmetičke sredine su:

- To je vrijednost od koje je polovica varijanata u uzorku manja, a polovica veća. Drugim riječima, odstupanje varijanata od aritmetičke sredine je jednako i u + i - smislu, odnosno suma odstupanja svih varijanata uzorka od aritmetičke sredine je nula, tj.

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

- Aritmetička sredina je ona vrijednost, od koje je suma kvadratnih odstupanja varijanta u uzorku najmanja, tj.

$$\Sigma(x-\bar{x})^2 = \min$$

- Uzme li se bilo koja druga varijanta u uzorku i utvrdi ova suma kvadratnih odstupanja svih ostalih vrijednosti uzoraka od nje, taj će iznos uvijek biti veći nego za $\bar{x}$.

Potkrijepimo ove karakteristike aritmetičke sredine dokazom na primjeru 1.3.1. s n = 20 varijanata, mjerne nekontinuirane varijable, čiju aritmetičku sredinu smo već izračunali ($\bar{x}$ = 70.8)

x	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-70)$	$(x-70)^2$
76	5.2	27.04	6	36
72	1.2	1.44	2	4
73	2.2	4.84	3	9
75	4.2	17.64	5	25
66	–4.8	23.04	–4	16
68	–2.8	7.84	–2	4
70	–0.8	0.64	0	0
71	0.2	0.04	1	1
71	0.2	0.04	1	1
73	2.2	4.84	3	9
70	–0.8	0.64	0	0
71	0.2	0.04	1	1
69	–1.8	3.24	–1	1
70	–0.8	0.64	0	0
69	–1.8	3.24	–1	1
69	–1.8	3.24	–1	1
70	–0.8	0.64	0	0
67	–3.8	14.44	–3	9
72	1.2	1.44	2	4
74	3.2	10.24	4	16

$\Sigma(x-\bar{x}) = +20.0-20.0 = 0$

$\Sigma(x-\bar{x})^2 = 125.20$

$\Sigma(x-70) = +28.0-12.0 = 16$

$\Sigma(x-70)^2 = 138.00$

Dokazane su obje karakteristike aritmetičke sredine. Naime, razvidno je da je $\Sigma(x-\bar{x}) = 0$ i $\Sigma(x-\bar{x})^2 = \min$, jer ako se uzme bilo koja druga vrijednost umjesto aritmetičke sredine to više ne stoji.

Kad smo na primjer umjesto $\bar{x}$ uzeli vrijednost 70, dobili smo $\Sigma(x-70) = 16$ (a ne nula), a $\Sigma(x-70)^2 = 138.00$ (a to je više nego 125,20 dobiveno uz $\bar{x}$).

Kod računanja aritmetičke sredine uzimaju se u obzir brojčane vrijednosti svih varijanata u uzorku. Zbog toga ponekad treba dobro paziti, jer će aritmetička sredina slabo predstavljati uzorak koji sadrži neke ekstremne vrijednosti. Npr.: u uzorku 5, 4, 6, 5, 5,

$$\bar{x} = \frac{25}{5} = 5$$

Zamijenimo li samo jednu varijantu, pa imamo 5, 4, 6, 5, 15, vrijednost aritmetičke sredine će se bitno promijeniti. Naime

$$\bar{x} = \frac{35}{5} = 7$$

U takvim slučajevima (naravno, samo ako to ima smisla s obzirom na veličinu uzorka i način na koji su varijante raspoređene) postoje bolja mjerila sredine nego što je $\bar{x}$.

2.1.2. Medijana (*Med*)

Za mnoga svojstva (varijable) u biologiji su karakteristične nesimetrične distribucije, s manjim brojem ekstremnih varijanata. U takvim je distribucijama medijana vrlo pogodno mjerilo sredine.

Medijana (medias = sredina) je vrijednost srednje varijante u uzorku varijanata poredanih rastućim redom.

Kao svaka srednja vrijednost, medijana dijeli uzorak na dva dijela s jednakim brojem varijanata. Brojčane vrijednosti pojedine varijante nemaju nikakvog utjecaja na medijanu. Tako je u ovom primjeru medijana u oba slučaja ostala ista, iako se u uzorku dogodila promjena. Naime:

4, 5, 5, 5, 6 $\quad Med = 5$

4, 5, 5, 6, 15 $\quad Med = 5$

Ili u prethodnim primjerima:

Za prvi (1.3.1.) s n = 20 (ako varijante poredamo rastućim redom) to će biti vrijednost između desete i jedanaeste varijante

66 67 68 69 69 69 70 70 70 <u>70 71</u> 71 71 72 72 73 73 74 75 76

$$Med = 70.5$$

Ako su varijante uzorka predstavljene varijacijskim redom, tada se medijana računa na tome prilagođen način. Uzmimo za to primjer 1.3.2. s uzorkom od n = 50 (duljina klipa kukuruza). Medijana će biti vrijednost između vrijednosti 25-e i 26-e varijante uzorka. One se nalaze u razredu 17.0 do 18.0 cm. Do donje granice tog razreda je 21 varijanta (2+2+3+4+10), a trebamo naći vrijednost 25. i 26. U tom razredu čiji areal je 1.0 cm je 13 varijanata;

vrijednost jedne je $\frac{1.0}{13} = 0.077$

vrijednost 25-e varijante je stoga vrijednost 21. plus 4 puta vrijednost jedne, a to je 17.0 + 4 × 0.077 = 17.308;

isto tako vrijednost 26-e varijante je 17.0 + 5 × 0.077 = 17.385, pa je

$$Med = \frac{17.308 + 17.385}{2} = 17.346 \text{ cm}$$

U mnogim biološkim problemima, kad se moraju uzeti u obzir i vrijednosti ekstremnih varijanata jer pripadaju uzorku, medijana je često najispravnije mjerilo sredine.

2.1.3. Modus (*Mod*)

Modus je vrijednost one varijante koja se u uzorku najčešće pojavljuje. Ili, to je vrijednost varijante (ili razreda ako se radi o varijacijskom redu) s najvećom frekvencijom.

Dakako, modus ima svoj smisao u dovoljno velikom uzorku. U primjeru s duljinom klipa kukuruza, *Mod* je vrijednost razreda s najvećom frekvencijom, dakle *Mod* = 17.5 cm.

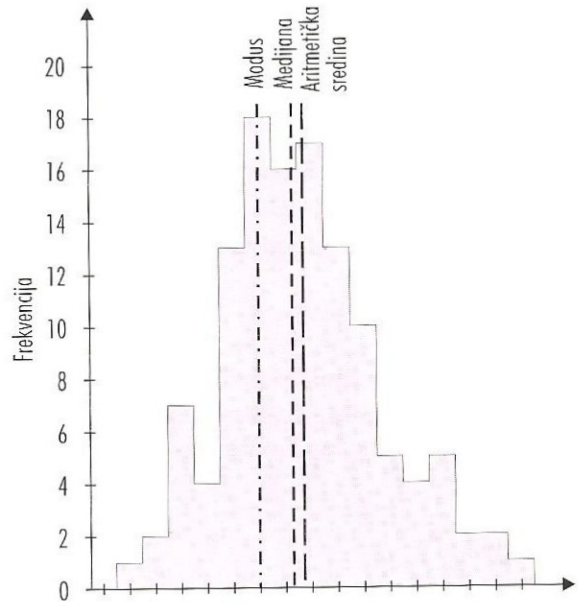
U distribuciji frekvencija to je ona vrijednost koja pripada vrhu frekvencijskog poligona odnosno krivulje. Stoga kažemo da distribucije mogu biti s jednim vrhom ili *unimodalne*, s dva *bimodalne* itd.

■

Sve ove opisane vrijednosti su mjerila sredine, odnosno to su srednje vrijednosti. Znači, njima je moguće locirati sredinu distribucije. Kad bi se histogram izrezao iz kartona, ili pak napravio model histograma, srednja vrijednost bi bila u točki u kojoj on balansira.

Ostaje međutim, pitanje, kad je koja od opisanih srednjih vrijednosti pogodnija za određivanje prave sredine.

Kod simetrične distribucije frekvencija aritmetička sredina, *Med* i *Mod* padaju, dakako, u istu točku (imaju iste vrijednosti). To, međutim, nije slučaj kod nesimetričnih distribucija.



SLIKA 3.
Aritmetička sredina ($\overline{x}$), medijana (Med) i modus (Mod) u nesimetričnoj distribuciji

Odnos između ova tri mjerila sredine značajno ovisi o tipu asimetričnosti distribucije. Kako je svako ovo mjerilo sredine različito definirano može se i različito, ili točnije rečeno u različitim uvjetima koristiti. Općenito, kad su podaci simetrično raspoređeni, potpuno je svejedno koje se od ovih mjerila sredine koristi. Numeričke će im vrijednosti biti vrlo slične ili jednake.

U praksi se rijetko koristi modus, jer nerijetko kod malih uzoraka i ne postoji.

U slučaju nesimetričnih distribucija varijanata, najpogodnija je medijana jer leži najbliže najvećoj koncentraciji varijanata uzoraka.

Evo nekoliko primjera koji ilustriraju ispravno korištenje ovih različitih srednjih vrijednosti.

Kad slušamo radio izvještaj s tržnice, čujemo informacije ovog tipa: na zagrebačkim tržnicama jutros se salata prodaje u prosjeku po 10 kuna za kilogram. To zapravo znači da najveći broj prodavača prodaje salatu po toj cijeni, a to je *modus*.

Primjena aritmetičke sredine za računanje prosječnog osobnog dohodka iz relativno malog i nesimetričnog uzorka (a nesimetričnost kod ovakvog fenomena gotovo u pravilu postoji, bez obzira kako dobar uzorak bio) ne bi bila opravdana. Uzmimo primjer podataka o osobnim dohocima desetorice ljudi (izraženih u stotinama kuna):

$$x_1 = 18 \quad x_2 = 26 \quad x_3 = 62 \quad x_4 = 13 \quad x_5 = 80$$

$$x_6 = 28 \quad x_7 = 26 \quad x_8 = 58 \quad x_9 = 28 \quad x_{10} = 32$$

Aritmetička sredina bi bila

$$\bar{x} = \frac{371}{10} = 37.1$$

a medijana, (vrijednost srednje varijante u nizu varijanata poredanih rastućim redom)

$$13 \quad 18 \quad 26 \quad 26 \quad \underline{28 \quad 28} \quad 32 \quad 58 \quad 62 \quad 80$$

$$Med = 28$$

S obzirom da je po definiciji srednja vrijednost ona vrijednost koja locira sredinu distribucije, od koje je dakle jednak broj većih kao i manjih varijanata, u ovom primjeru očito medijana ima prednost.

2.2. Mjerila varijabilnosti ili disperzije

Procesi disperzije u uzorku ili populaciji uvjetuju različitost među jedinkama, a ta se različitost varijanata zove **varijabilnost**. Kao što su procesi centralne tendencije razlogom za grupiranje varijanata oko sredine, tako su i procesi disperzije ili "raspršivanja" varijanata razlogom za različitost u uzorku ili populaciji. Tu "raspršenost", različitost ili disperziju varijanata u uzorku, moguće je također mjeriti. Da bi se lakše razumjelo što se pod tim "mjeriti varijabilnost" podrazumijeva, poslužimo se primjerom s dva uzorka (A i B) svaki sa po 5 varijanata.

Uzorak A				
5	6	7	8	9

$$Med_A = 7$$

$$\bar{x}_A = \frac{5+6+7+8+9}{5} = 7$$

Uzorak B				
1	4	7	10	13

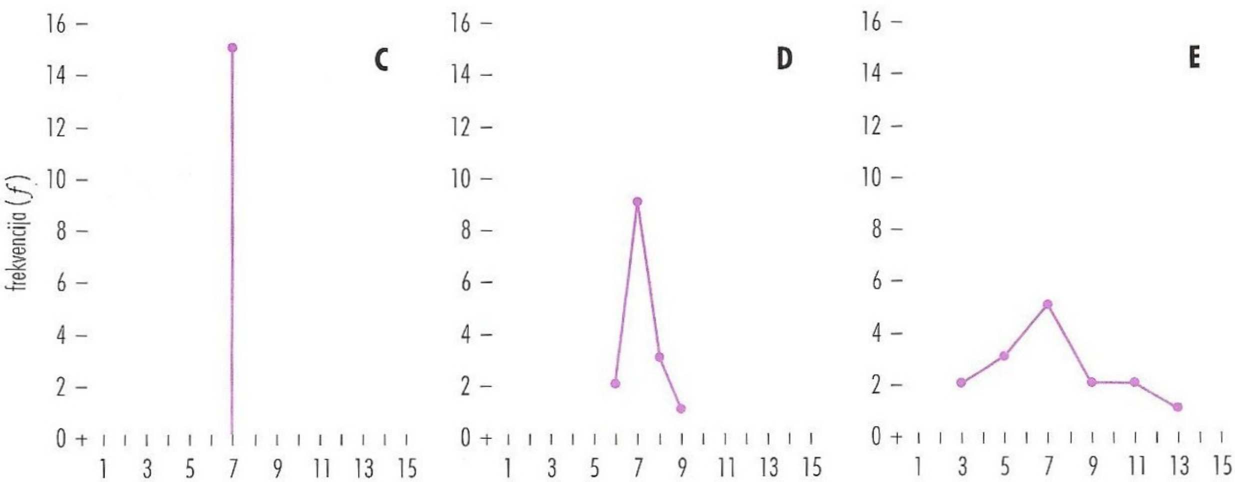
$$Med_B = 7$$

$$\bar{x}_A = \frac{1+4+7+10+13}{5} = 7$$

Oba uzorka su iste veličine ($n_A = 5$ i $n_B = 5$), a imaju i iste aritmetričke sredine i medijane. Međutim, očito je veća različitost među varijantama uzorka B. One su i međusobno više različite nego varijante uzorka A, a i od aritmetičke sredine je odstupanje veće u uzorku B.

Uzmimo i primjer s tri jednako velika uzorka (n = 15), gotovo istih prosječnih vrijednosti (7.0, 7.2, odnosno 7.3). Već samo iz njihovih grafičkih prikaza (slika 4) možemo dobiti dojam o varijabilnosti.

Uzorak C	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
Uzorak D	6	6	7	7	7	7	7	7	7	7	7	8	8	8	9
Uzorak E	3	3	5	5	5	7	7	7	7	7	9	9	11	11	13



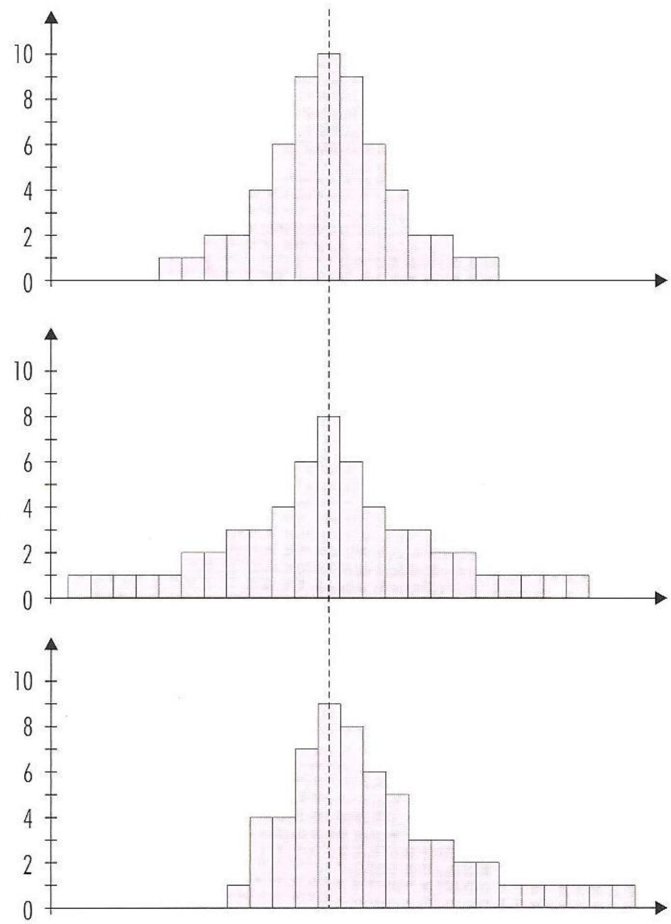
SLIKA 4.
Frekvencijski poligoni uzoraka različitih varijabilnosti

U uzorku C sve su varijante iste, dakle nema varijabilnosti.

Varijante u uzorku D su tako grupirane oko sredine, da je frekvencijski poligon "uzak i šiljast", a varijante uzorka E su različitije pa je poligon "širi i manje šiljast".

Općenito, što je manja koncentracija varijanata oko sredine, što je veća različitost varijanata, to je frekvencijski poligon širi, to je varijabilnost veća i obratno (slika 5).

Od mjerila kojima je moguće definirati i mjeriti tu različitost ili varijabilnost varijanata unutar uzorka ili populacije, opisat ćemo ***apsolutna mjerila***: **varijacijsku širinu**, **standardnu devijaciju** i **varijancu**, te **varijacijski koeficijent** - ***relativno mjerilo varijabilnosti***.



SLIKA 5.
Distribucije frekvencija istog n-a, istih aritmetičkih sredina, ali različitih varijabilnosti

Apsolutna mjerila varijabilnosti izražavaju se u onim mjernim jedinicama u kojima su izražena svojstva o kojima se radi. Nerijetko se, međutim, želi uspoređivati varijabilnost različitih uzoraka (istih ili različitih varijabli). Pri tome se mora koristiti neko relativno mjerilo.

2.2.1. Varijacijska širina ili raspon ($vš$)

Najjednostavniji način da se dobije uvid u varijabilnost uzorka je naći raspon vrijednosti varijanata, odnosno **razliku između najveće i najmanje vrijednosti u uzorku** ili **varijacijsku širinu** ($vš$):

$$vš = x_{max} - x_{min}$$

U slučaju uzorka A ona je 4 (tj. 9 - 5), a za uzorak B iznosi 12 (tj. 13-1). Dakle varijacijska širina uzorka B, koja je veća od varijacijske širine uzorka A ($vš_B > vš_A$), govori o većoj različitosti varijanata u uzorku B, odnosno manjem rasponu (raspršenosti, disperziji) nego u uzorku A. Isto vrijedi i za uzorke E i D ($vš_E > vš_D$).

S obzirom da se varijacijska širina temelji samo na vrijednostima krajnjih (ekstremnih) varijanata, i ništa ne kaže o načinu variranja varijanata unutar tih krajnjih vrijednosti, ona u većini slučajeva može biti samo grubo ili tek orijentacijsko mjerilo varijabilnosti. To se može bolje razumjeti usporedbom vrijednosti sljedeća dva uzorka.

Uzorak I	*				*	**	***	**	*				*
Vrijednost varijanata	3	4	5	6	7	8	9	10	11	12	13	14	15
Uzorak II	*	*	*	*	*	*	*	*	*	*	*	*	*

* predstavljaju frekvencije pojedinih vrijednosti u uzorku

Kako oba uzorka imaju iste ekstremne vrijednosti (3 i 15), varijacijske su im širine iste (12). Očito je međutim, da je način variranja unutar uzorka I različit od onog unutar uzorka II. Varijante uzorka II su jednolično raspoređene unutar varijacijske širine, dok je najveći broj varijanata uzorka I grupiran unutar vrijednosti 7 i 11. Upravo tu različitost u načinu "raspršivanja" varijanata u uzorku, nije moguće opisati varijacijskom širinom. Ona, naime, može biti velika (jer se vrlo razlikuju samo krajnje varijante), a da varijabilnost ipak nije velika (jer su ostale varijante vrlo slične). Upravo u tome i jest slabost varijacijske širine.

2.2.2. Standardna devijacija (s) i varijanca (s^2)

To su najvažnija i najbolja mjerila varijabilnosti i mjerila koja se najčešće koriste. Ispravno mjerilo disperzije mora obuhvatiti sve varijante i izraziti njihovu udaljenost od sredine.

Uzmimo na primjer uzorak od 7 varijanata (n = 7)

18 20 20 16 16 18 18

Aritmetička sredina $\overline{x} = 18$. Odstupanje svake varijante od te srednje vrijednosti može se označiti kao $x - \overline{x}$, pa su odstupanja svih varijanata sljedeća:

(18-18), (20-18), (20-18), (16-18), (16-18), (18-18), (18-18), ili

0, +2, +2, -2, -2, 0, 0

Nezgoda je u tome što je zbroj svih ovih odstupanja nula (što je uostalom svojstvo aritmetičke sredine).

Zbog toga se uzima prosječno odstupanje, a to je zbroj apsolutnih vrijednosti odstupanja varijanata od srednje vrijednosti, podijeljen s ukupnim brojem varijanata

$$\text{prosječno odstupanje} = \frac{\sum \left| x - \overline{x} \right|}{n}$$

Za navedeni primjer to bi bilo:

$$\frac{0+2+2+2+2+0+0}{7} = 1.14$$

što znači da varijante odstupaju od srednje vrijednosti u prosjeku za 1.14.

Ovo prosječno odstupanje varijanata od sredine može se koristiti kod uspoređivanja različitih grupa podataka (uzoraka), da se ustanovi da li je veća varijabilnost oko srednje vrijednosti u jednoj, ili u drugoj grupi. Prosječno odstupanje se u pravilu koristi onda kad se kao mjerilo sredine uzima medijana.

Međutim, najvažnija mjerila varijabilnosti su ipak **varijanca** (s^2) i **standardna devijacija** (s).

Uzimaju u obzir sve varijante distribucije i definiraju odstupanje svake varijante od sredine distribucije.

Varijanca (s^2) **je prosječno kvadratno odstupanje varijanata od srednje vrijednosti uzorka**, tj.

$$s^2 = \frac{\sum (x - \overline{x})^2}{n-1}$$

a **standardna devijacija** (s) **je pozitivna vrijednost drugog korijena iz varijance**

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x - \overline{x})^2}{n-1}}$$

U ovim formulama (n-1) se koristi kao divizor, a naziva se i **broj slobodnih varijanata**. Kasnije će se to pobliže obrazložiti (str. 31).

Da se izračuna standardna devijacija i varijanca za ranije naveden primjer, treba najprije odstupanje svake varijante od aritmetičke sredine kvadrirati, zbrojiti ih i podijeliti s brojem slobodnih varijanata, tj.

x	$x-\bar{x}$	$(x-\bar{x})^2$	x^2
18	0	0	324
20	+2	4	400
20	+2	4	400
16	–2	4	256
16	–2	4	256
18	0	0	324
18	0	0	324
$\sum x=126$	$\sum(x-\bar{x})=0$	$\sum(x-\bar{x})^2=16$	$\sum x^2=2284$

$$\bar{x}=18$$

$$s^2=\frac{16}{6}=2.67$$

$$s=\sqrt{2.67}=1.63$$

Izraz $\sum(x-\bar{x})^2$ je zbroj ili **suma kvadrata**, pa se stoga bilježi i u teoretskim i u praktičnim tekstovima sa SS (od engleskog *Sum of Squares*).

U tom smislu bi se varijancu i standardnu devijaciju moglo pisati:

$$s^2=\frac{SS}{n-1} \quad \text{odnosno} \quad s=\sqrt{\frac{SS}{n-1}}$$

Za praktično i brzo računanje standardne devijacije i varijance, koristi se i formula za SS

$$SS=\sum x^2-\frac{(\sum x)^2}{n}$$

pa je

$$s^2=\frac{\sum x^2-\frac{(\sum x)^2}{n}}{n-1}$$

Često je puno praktičnija, a rezultat je dakako isti. Pri tom najprije treba izračunati zbroj kvadratnih vrijednosti svih varijanata (Σx^2), odnosno kvadrirati zbroj svih varijanata $(\Sigma x)^2$.

$$s^2 = \frac{2284 - \frac{126^2}{7}}{7-1} = \frac{16}{6} = 2.67$$

$$s = \sqrt{2.67} = 1.63$$

(što dakako predstavlja isti rezultat kao prije, samo dobiven na drugi način).

Ako je uzorak sređen u varijacijski red, varijanca i standardna devijacija se izračunaju po formuli

$$s^2 = \frac{\sum fx'^2 - \frac{(\sum fx')^2}{n}}{n-1}$$

$$s = \sqrt{\frac{\sum fx'^2 - \frac{(\sum fx')^2}{n}}{n-1}}$$

Tako bi za primjere 1.3.1. i 1.3.2. uzorka nekontinuirane varijable s $n = 20$ i uzorka kontinuirane varijable s $n = 50$, varijacijske širine ($v\check{s}$), standardne devijacije (s) i varijance (s^2) bile:

■

$v\check{s}$ = 76 - 66 = 10

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

Kako je izraz u brojniku $\sum (x - \bar{x})^2$ već izračunan (str. 17) i iznosi 125.2, to je

$$s^2 = \frac{125.2}{20-1} = 6.59$$

$$s = \sqrt{6.59} = 2.57$$

Ako pak koristimo drugi način računanja, tj.

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

tada je

$$\sum x^2 = 76^2 + 72^2 + ... + 74^2 = 100378, \text{ a } \sum x = 1416$$

$$s^2 = \frac{100378 - \frac{1416^2}{20}}{20-1} = \frac{100378 - 100252}{19} = 6.59$$

$$s = \sqrt{6.59} = 2.57$$

■

$$vš = 21.0 - 12.0 = 9.0 \text{ cm}$$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

$$\sum x^2 = 13.5^2 + 16.0^2 + ... + 17.0^2 = 14412.25, \text{ a } \sum x = 843.5$$

pa je $s^2 = \frac{14412.25 - \frac{843.5^2}{50}}{50-1} = \frac{14412.25 - 14229.845}{49} = 3.722 \text{ cm}$

$$s = \sqrt{3.73} = 1.93 \text{ cm}$$

Da bismo mogli primijeniti formulu za računanje s^2 i s prilagođenu varijacijskom redu, prikažimo ga ponovno i dodajmo potrebne izraze za računanje ovih mjerila varijabilnosti:

	12.0–13.0	13.0–14.0	14.0–15.0	15.0–16.0	16.0–17.0	17.0–18.0	18.0–19.0	19.0–20.0	20.0–21.0	21.0–22.0
f	2	2	3	4	10	13	10	3	2	1
x'	–5	–4	–3	–2	–1		+1	+2	+3	+4
x'^2	25	16	9	4	1		1	4	9	16
fx'^2	50	32	27	16	10		10	12	18	16

$\sum f = 50$ (=n),

$\sum fx'^2 = 191$

$$s^2 = \frac{\sum fx'^2 - \frac{(\sum fx')^2}{n}}{n-1} = \frac{191 - \frac{(-19)^2}{50}}{49} = 3.75 \text{ cm}$$

$s = \sqrt{3.75} = 1.94$ cm

Za primijetiti je da se ovim načinom čini mala pogreška (to manja što je uzorak veći!) jer se u računanju zanemaruje razred s najvećom frekvencijom (nulti razred). No i u ovom uzorku sa 50 varijanata razlika u odnosu na onaj prethodni način računanja je minimalna.

Već iz formula za računanje varijance je vidljivo da što su varijante različitije, čim je dakle veća njihova varijabilnost, veće će biti i njihove pojedinačne razlike u odnosu na srednju vrijednost. Iz toga proizlaze i veće kvadratne vrijednosti tih razlika, odnosno veća varijanca.

U slučaju da su sve varijante uzorka potpuno istovjetne, varijabilnosti dakle nema, varijanca je nula.

2.2.3. Varijacijski koeficijent (*cv*)

Ako se želi uspoređivati varijabilnosti različitih uzoraka, potrebno je to činiti nekim mjerilom koje sadrži u sebi i procjenu sredine i procjenu disperzije.

Takvo **relativno mjerilo varijabilnosti je varijacijski koeficijent (*cv*), koji daje standardnu devijaciju u dijelovima srednje vrijednosti a izražava se u** %.

$$cv = \frac{100 \cdot s}{\bar{x}}$$

Kod usporedbe varijabilnosti dva ili više uzoraka učinila bi se velika pogreška kad bi se o jačini varijabilnosti zaključivalo iz vrijednosti varijanci ili standardnih devijacija, iako je i to u određenim slučajevima dopustivo. Naime, kad se uspoređuje varijabilnost uzoraka s istim srednjim vrijednostima to je moguće i na osnovi varijanci ili standardnih devijacija. Ako se, pak, radi o uzorcima s različitim srednjim vrijednostima, tada se kao mjerilo mora koristiti varijacijski koeficijent, jer on izražava standardnu devijaciju u dijelovima srednje vrijednosti.

Dakle, ako su na primjer srednje vrijednosti i standardne devijacije dva uzorka

$$\bar{x}_A = 150 \qquad \bar{x}_B = 150$$

$$s_A = 28 \qquad s_B = 31$$

iz $s_A < s_B$ može se ustvrditi da je uzorak B varijabilniji nego uzorak A. Isti zaključak bi se dobio i pomoću varijacijskih koeficijenata:

$$cv_A = \frac{100 \cdot 28}{150} = 18.7\%, \qquad cv_B = \frac{100 \cdot 31}{150} = 20.7\%$$

$$cv_A < cv_B$$

Upravo je zato uspoređivanje varijabilnosti dvaju uzoraka koji imaju iste aritmetičke sredine, moguće učiniti već na temelju njihovih standardnih devijacija ili varijanci, odnosno nije potrebno računati varijacijske koeficijente (jer oni će ionako zbog istog nazivnika - jednakih aritmetičkih sredina, biti u istom odnosu).

Ako se pak radi o uzorcima različitih srednjih vrijednosti, jedino ispravno je u usporedbi koristiti varijacijski koeficijent:

$$\bar{x}_A = 150 \qquad \bar{x}_B = 185$$

$$s_A = 28 \qquad s_B = 31$$

$$cv_A = \frac{100 \cdot 28}{150} = 18.7\%, \qquad cv_B = \frac{100 \cdot 31}{185} = 16.7\%$$

$$cv_A > cv_B$$

U usporedbi temeljenoj na standardnim devijacijama, dobio bi se potpuno suprotan i neispravan zaključak o varijabilnosti ova dva uzorka (jer $s_A < s_B$).

Varijacijski koeficijent ne ovisi o skali mjerenja (budući da je kvocijent dviju procjena mjerenih u istoj skali), pa je vrlo upotrebljiv i koristan za komparaciju varijabilnosti uzoraka varijabli izraženih i u različitim skalama i mjernim jedinicama. Tako, na primjer, iz vrijednosti varijacijskih koeficijenata, temeljenim na izmjerama visine i težine jedne velike skupine ljudi, a koji iznose

$$cv_{visine} = 3.6\% \qquad cv_{težine} = 7.8\%$$

smijemo zaključiti da težina više varira od visine, bez obzira što je jedna varijabla izražena u jednim (cm), a druga u drugim (kg) mjernim jedinicama.

2.3. Mjerila centralne tendencije i disperzije uzoraka i parametri populacije

Sve do sada razmatrali smo mjerila centralne tendencije i disperzije u uzorku, bez mnogo objašnjavanja o tome što ona predstavljaju.

Već je ranije spomenuto, da gotovo nikad u našoj struci nismo u stanju analizirati populacije (zbog njihove veličine), pa koristimo veće ili manje, ali uvijek slučajne i reprezentativne uzorke iz tih populacija, analiziramo ih i donosimo zaključke. Pri tome nas zapravo uvijek zanima saznati o populacijama iz kojih su uzorci uzeti. Stoga, kad računamo prosječnu vrijednost uzorka, zapravo želimo saznati pravu prosječnu vrijednost cjeline kojoj taj uzorak pripada - populacije. Zato mjerila centralne tendencije i disperzije dobivena iz uzoraka zovemo **procjenama pravih parametara populacija** kojima uzorci pripadaju.

Dogovorno se za oznake tih mjerila u uzorcima koriste latinska slova, a u populacijama grčka. Srednja vrijednost uzorka $\bar{x}$ je dakle procjena prave prosječne vrijednosti populacije μ (mi).

uzorak	populacija
$\bar{x}$	μ (mi)
s	σ (sigma)
s^2	σ^2

Ovdje treba naglasiti da varijanca uzorka ne bi bila dobra procjena varijance populacije kad bi se računala kao

$$s^2 = \frac{SS}{n}$$

Matematičari - statističari su dokazali da bi u tom slučaju uvijek bila manja od varijance populacije. Da bi se to izbjeglo, suma kvadrata se dijeli s brojem varijanata u uzorku umanjenim za jedan ili (n - 1), a taj se naziva **broj slobodnih varijanata** ili **broj stupnjeva slobode.** Tako se postiže da varijanca uzorka postaje vjerodostojnije mjerilo varijance populacije. Svakako, čim je uzorak veći, manja će biti razlika dijeli li se s n ili s (n - 1). Kad raspolažemo s podacima cijele populacije, što je zaista izuzetno rijetko, opravdano je dijeliti s n jer tada ne procjenjujemo nego računamo varijancu populacije (σ^2). Stoga i govorimo o procjenama parametara populacije, preko vrijednosti tih parametara u uzorku.

■

1

2

3 Distribucija vjerojatnosti i neke važnije teoretske distribucije

4

5

6

7

3.1. Binominalna distribucija

3.2. Teoretska normalna distribucija frekvencija

Kako sve pojave variraju pod utjecajem brojnih poznatih uzroka koji su često izvan naše kontrole, razumljivo je upitati se kako možemo očekivati da će opažanja koja o tim pojavama provodimo imati nekog smisla.

Na sreću, kod velikog broja svojstava postoji određeni stupanj zakonitosti u variranju.

Ako se konstruiraju frekvencijski poligoni iz podataka o: broju zrna u svakoj od, recimo, 500 mahuna soje, ili o težini 300 izvaganih plodova neke sorte jabuke, ili o prinosu zrna u 50 klasova pšenice ili 100 klipova kukuruza, ili o broju cvjetova na biljci kod ukrasnog bilja i slično, svi će ti grafički prikazi imati nešto zajedničko. Naime, najviše točke svih tih frekvencijskih poligona nalazit će se negdje oko sredine, gdje je i frekvencija varijanata najveća, a blago aproksimirati apscisu prema ekstremima, zbog sve manje i manje frekvencije (to se već moglo primijetiti na histogramu uzorka n = 50 o duljini klipa kukuruza, str. 8).

Ako je broj opaženih podataka dovoljno velik, frekvencijski poligoni najvećeg broja bioloških svojstava prelaze u krivulje zvonolikog oblika.

Da bismo bolje razumjeli raspodjelu ili distribuciju varijanata, zadržat ćemo se nakratko na nekim važnijim teoretskim distribu-cijama frekvencija.

3.1. Binominalna distribucija

Već je spomenuto da velik broj svojstava u biološkim strukama predstavljaju tzv. kontinuirane varijable. Treba se ipak malo zadržati na distribucijama frekvencija i prisjetiti se osnovnih pojmova vjerojatnosti.

Poslužimo se u tu svrhu jednostavnim primjerom iz naše struke.

Uzmimo jednu biljnu populaciju u kojoj smo opažanjem utvrdili da je omjer biljaka inficiranih nekom bolešću i zdravih individua 4% : 96%, što također možemo pisati kao 0.04 : 0.96. Sve individue populacije možemo dakle prikazati kao: 0.96 + 0.04 = 1.0 ili 96% + 4% = 100. To znači da ako iz ove populacije uzmemo uzorak od 100 biljaka, za očekivati je da će biti 96 zdravih, a 4 zaražene biljke. Međutim, može se dogoditi da u uzorku bude nešto više ili nešto manje od 4 inficirane biljke. Ako jako povećavamo uzorak (recimo na 500 ili 1000 i više biljaka) taj će se broj sve manje razlikovati od 0,04. Može se reći, da je **vjerojatnost pojavljivanja inficiranog individua u uzorku 0.04 a zdravog 0.96**. Ako se vjero-jatnost zdravih biljaka označi p a inficiranih q, to se može napisati kao:

$$p + q = 0.96 + 0.04 = 1.0$$

Pogledajmo što možemo očekivati ako iz populacije potpuno slučajno izdvojimo uzorak koji se sastoji od samo dvije biljke (n = 2).

Te će biljke biti:

ili a) *obje zdrave biljke*

ili b) *1 zdrava, 1 inficirana*
(i to na 2 načina: ili je prva zdrava, a druga inficirana, ili obratno; no to je ista kombinacija)

ili c) *obje inficirane biljke.*

Zato se to može pisati kao

$$p^2 + 2pq + q^2$$

Ovo se može izraziti i u vidu vjerojatnosti svake od ovih mogućnosti:

a) $p^2 = 0.96^2 = 0.9216$

b) $2pq = 2 \cdot 96 \cdot 0.04 = 0.0768$

c) $q^2 = 0.04^2 = 0.0016$

Ako se uzorak sastoji od 3 biljke (n = 3), tada su mogućnosti ove:

ili a) *sve 3 zdrave biljke*

ili b) *2 zdrave, 1 inficirana*
(što se može dobiti na 3 načina i to:
zdrava, zdrava, inficirana;
zdrava, inficirana, zdrava, te
inficirana, zdrava, zdrava)

ili c) *2 inficirane, 1 zdrava*
(što se može dobiti na 3 načina i to:
inficirana, inficirana, zdrava;
inficirana, zdrava, inficirana, te
zdrava, inficirana, inficirana)

ili d) *sve 3 inficirane*

odnosno

$$p^3 + 3p^2q + 3pq^2 + q^3$$

Vjerojatnost svake kombinacije je:

a) $p^3 = 0.96^3 = 0.88474$

b) $3p^2q = 3 \cdot 0.96^2 \cdot 0.04 = 0.1106$

c) $3pq^2 = 3 \cdot 0.96 \cdot 0.04^2 = 0.0046$

d) $q^3 = 0.04^3 = 0.000064$

Povećamo li uzorak, tada za n = 4 imamo: 4 zdrave, 4 puta kombinaciju 3 zdrave 1 inficirana, 6 puta kombinaciju 2 zdrave i 2 inficirane, 4 puta kombinaciju 1 zdrava i 3 inficirane, te konačno sve 4 inficirane. To ćemo napisati kao:

$$p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

odnosno uz n = 5 to je

$$p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5.$$

Ovaj nas primjer sada već svakako podsjeća na dobro nam znani binom $(p+q)^n$. Pišemo li ove slučajeve u obliku binoma, vodeći računa o veličini uzorka, tada imamo za:

$$
\begin{array}{lll}
n = 1 & (p+q)^1 & = p + q \\
n = 2 & (p+q)^2 & = p^2 + 2pq + q^2 \\
n = 3 & (p+q)^3 & = p^3 + 3p^2q + 3pq^2 + q^3 \\
n = 4 & (p+q)^4 & = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4 \\
n = 5 & (p+q)^5 & = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5 \\
 & . & \\
\text{itd.} & . & \\
 & . &
\end{array}
$$

do konačnog oblika: $(p + q)^n = p^n + np^{(n-1)}q + ... + npq^{(n-1)} + q^n$

Koeficijenti za razvijanje ovog binoma dobiju se pomoću Pascal-ovog trokuta, tj.

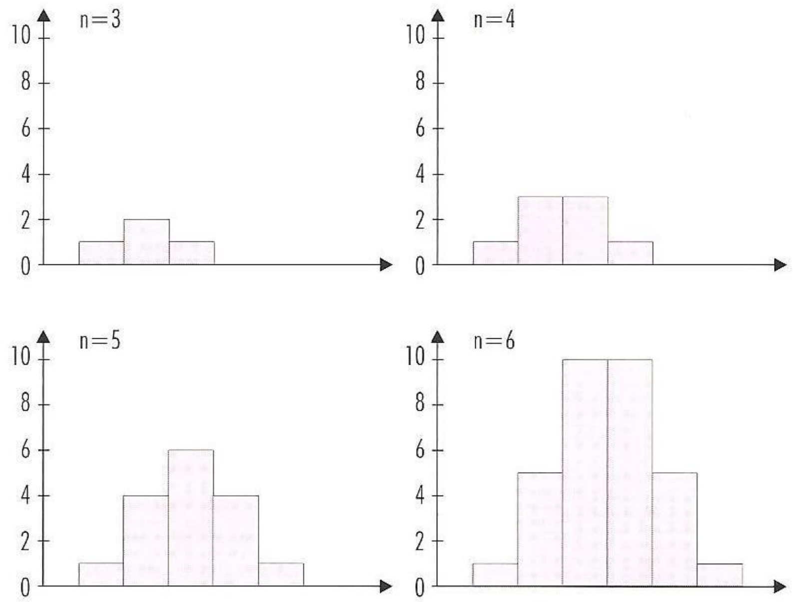
n											
1					1		1				
2				1		2		1			
3			1		3		3		1		
4		1		4		6		4		1	
5	1		5		10		10		5		1

Tako nije teško utvrditi vjerojatnost svakog pojedinog slučaja (kombinacije). Kolika bi bila npr. vjerojatnost da će od 5 biljaka u uzorku (n = 5), 4 biti zdrave a samo 1 inficirana? Tu dakle treba izračunati vjerojatnost pojavljivanja kombinacije p^4q, a ta je

$$5p^4q = 5 \times 0.96^4 \times 0.04 = 0.1699 \text{ ili oko } 17\%.$$

Treba uočiti da se u ovom primjeru radi o nekontinuiranoj varijabli koja se pojavljuje u samo dvije mogućnosti tj. ili zdravi ili inficirani indvidui. Kako broj varijanata u uzorku raste, tako je moguće odrediti sve više i više stupaca histograma, dok se na koncu uz vrlo povećani ***n*** nazire i krivulja. Tako se dobiva teoretska distribucija frekvencija ili distribucija vjerojatnosti varijabli koje karakteriziraju dvije kategorije (crno-bijelo, zdravo-inficirano, živo-uginulo i sl.), a koja je poznata kao **binominalna distribucija**.

■



SLIKA 6.
Nastanak binominalne distribucije

3.2. Teoretska normalna distribucija frekvencija

Međutim, najveći broj svojstava koja se u biologiji proučavaju predstavljaju tzv. kontinuirane varijable (težina ploda, dimenzije, intenzitet zaraženosti bolešću ili štetnikom, sadržaj biogenih elemenata u nekom tipu tla, količina šećera u soku ili plodu i sl.). Zato ćemo se više zadržati na ***distribuciji kontinuiranih varijabli***.

Distribucija frekvencija varijanata takvih varijabli predstavljena je kontinuiranom krivuljom. Visinu krivulje uvjetuje koncentracija varijanata u svakoj točki apscise, odnosno frekvencija. Kako je kod kontinuiranih varijabli moguća bilo koja vrijednost unutar dviju granica, u nekom rasponu moguće je utvrditi samo frekvenciju.

Tako, na primjer, u varijacijskom redu za duljinu klipa kukuruza (str. 8) u razredu od 17.0 do 18.0 cm frekvencija 13 znači da je u uzorku 13 klipova duljine između 17.0 i 18.0 cm. To znači da ima klipova dugih na primjer 17.6, ili 17.8, ili ... Stoga se kod sređivanja podataka u varijacijski red i prikazivanja podataka u vidu raspodjele (distribucije) frekvencija, proizvoljno izabire broj razreda i veličinu razrednog razmaka. Na taj se način može odrediti samo frekvencija varijanata koje padaju unutar granica svakog razreda, ali ne i točne vrijednosti svake od njih.

Vratimo se na čas binominalnoj distribuciji. Za binominalnu distribuciju je, naime, karakteristično da je svojstvena varijablama koje se pojavljuju u samo dvije mogućnosti (recimo p i q).

Ako zamislimo čitav dijapazon mogućnosti u kojima se neka varijabla može pojavljivati (dakle uz $p + q$ još i $+ r + ... + z$), tada i polinom $(p + q + r + ... + z)^n$ poprima izgled pravilne zvonolike distribucije. Takva zvonolika distribucija u kojoj su varijante najjače zgusnute oko sredine, a sve manje i manje ih ima prema krajevima zove se **normalna distribucija**. **Dakle, ako n postaje jako velik, tj. $n \to\infty$ binominalna distribucija poprima izgled pravilne zvonolike krivulje, odnosno prelazi u teoretsku normalnu distribuciju frekvencija**. To je distribucija koja se najviše koristi, a zove se još i **normalna distribucija vjerojatnosti** ili jednostavno **normalna distribucija** ili **normalna krivulja**.

Prema matematičarima koji su je otkrili, te razradili jednadžbu krivulje koja predstavlja tu distribuciju, ona se zove i Gaussova distribucija (Karl F. Gauss, 1777.-1855.), de Moivre-ova distribucija (Abraham de Moivre, 1667.-1745.) ili Laplace-ova (Pierre S. de Laplace, 1749.-1827.) distribucija.

Dok binominalna distribucija proizlazi iz vrijednosti varijabli izraženih isključivo cijelim brojevima, za normalnu distribuciju je karakteristično da sadrži sve moguće vrijednosti duž horizontalne skale.

Zato se za ovu distribuciju i kaže da je to **kontinuirana distribucija vjerojatnosti**.

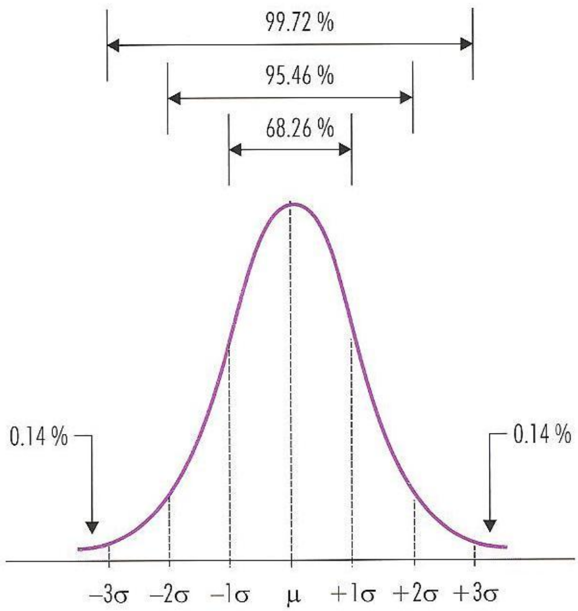
Teoretska normalna distribucija predstavljena je zvonolikom simetričnom krivuljom koja blago aproksimira apscisu ali je nikada ne dotiče, a može se opisati jednadžbom

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Osim dvije matematičke konstante (π = 3,14159 i e = 2,71828) u ovoj jednadžbi su još i dva parametra μ i σ za koje već znamo da predstavljaju: μ - mjerilo sredine i σ - mjerilo raspršenosti (disperzije).

Dakle, teoretska normalna distribucija frekvencija u kojoj $n \to \infty$ je potpuno definirana s dva parametra tj. μ i σ. Oni opisuju oblik i lokaciju distribucije.

Kako je normalna krivulja simetrična, sve vrijednosti koje lociraju sredinu padaju u istu točku. Srednja vrijednost teoretske normalne distribucije μ = 0 (*Med* = 0, *Mod* = 0), **a standardna devijacija s = 1.0 i varijanca s^2 = 1.0.**



SLIKA 7.
Normalna distribucija

Za teoretsku, normalnu distribuciju izračunato je koliki se postotak od ukupnog broja varijanata nalazi u određenim intervalima, tj.

u intervalu $\mu \pm 1\ \sigma$ nalazi se 68.26% varijanata
u intervalu $\mu \pm 2\ \sigma$ nalazi se 95.46% varijanata
u intervalu $\mu \pm 3\ \sigma$ nalazi se 99.72% varijanata

Ili, izraženo na drugi način:

50% varijanta je u intervalu $\mu \pm 0.674\ \sigma$
95% varijanta je u intervalu $\mu \pm 1.960\ \sigma$
99% varijanta je u intervalu $\mu \pm 2.576\ \sigma$

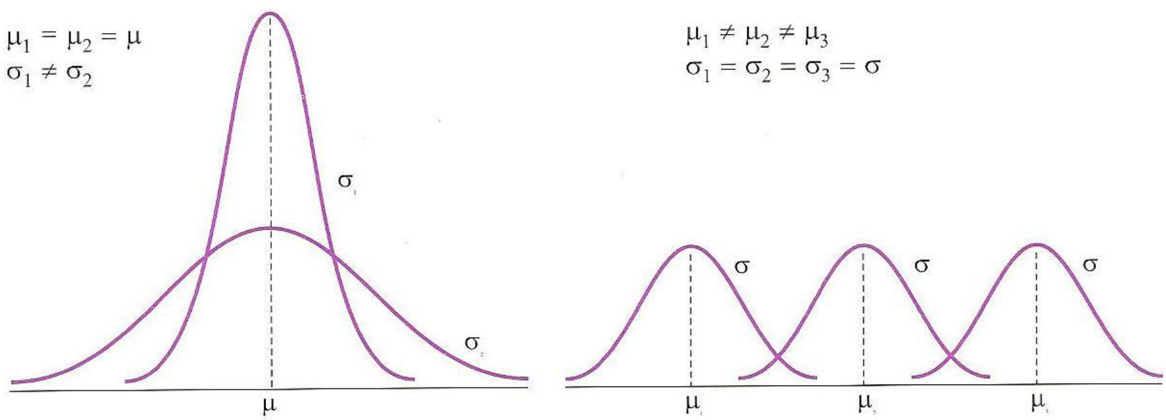
To se može pročitati i kao:

50% varijanata je veće od μ - 0.674 σ, a manje od μ + 0.674 σ
95% varijanata je veće od μ - 1.960 σ, a manje od μ + 1.960 σ
99% varijanata je veće od μ - 2.576 σ, a manje od μ + 2.576 σ

Na zakonitostima koje u njoj vladaju temelje se sve osnovne statističke metode i razrađene su mnoge tablice što služe pri provjeravanju (testiranju) različitih postavki o svojstvima koja se proučavaju, a čije distribucije varijanata u uzorcima slijede normalnu distribuciju.

Vrlo velik broj svojstava (varijabli) je raspoređen prema normalnoj distribuciji, pa je zato važno poznavati sve njezine zakonitosti.

Dakako, ovisno o vrijednostima za μ i σ moguć je beskonačno veliki broj normalnih krivulja. Distribucije mogu, dakle, vrlo različito izgledati, no opisane zakonitosti uvijek vrijede.



SLIKA 8.

Normalne distribucije s različitim vrijednostima za μ i σ

Ako uzorak ili populacija neke varijable slijedi normalnu distribuciju dozvoljeno je primijeniti statističke postupke koji se zasnivaju na tim zakonitostima. To znači da je u distribuciji varijanata varijable koja slijedi normalnu distribuciju, moguće uz određenu vjerojatnost ustvrditi koje varijante uopće pripadaju toj distribuciji (tom uzorku), a koje ne. Moglo bi se reći da je, za sve one varijante koje se nalaze izvan intervala obuhvaćenog u $\bar{x} \pm 3s$, vrlo mala vjerojatnost da pripadaju uzoraku čija srednja vrijednost je $\bar{x}$, a standardna devijacija s. To je, međutim, vrlo strogo jer ne zaboravimo da ovo $\pm 3s$ vrijedi za vrlo, vrlo veliki n.

Može se uočiti, da se u normalnoj distribuciji unutar $\mu \pm 3\ \sigma$ nalaze gotovo sve varijante (99,72%). Njihova je zgusnutost oko sredine, pa se skoro 2/3 svih varijanata nalazi u intervalu $\mu \pm 1\ \sigma$.

Teoretski se normalna distribucija frekvencija proteže od $-\infty$ do $+\infty$ na apscisi, koja predstavlja vrijednosti varijable. To znači da varijabla koja je raspoređena prema normalnoj distribuciji može poprimiti bilo koje vrijednosti. Međutim, vrlo male ili vrlo velike vrijednosti koje padaju izvan intervala prosjek plus ili prosjek minus tri standardne devijacije vrlo su malo vjerojatne. Odnosno, vjerojatnost da pripadaju u tu distribuciju je vrlo mala (0.28% ili 0.14 + 0.14).

Stoga te zakonitosti možemo izraziti i na ovaj način:

- vrlo je velika vjerojatnost (99%) da varijante veće od μ - 2.576σ i manje od μ + 2.576 σ pripadaju takvoj distribuciji
- za varijante veće od μ +2.576 σ i varijante μ -2.576 σ vjerojatnost da pripadaju toj distribuciji vrlo je mala.

Dakle vrlo pouzdano možemo tvrditi da one varijante koje su izvan intervala $\mu \pm 3\ \sigma$ ne pripadaju toj distribuciji.

Zato se opisani intervali zovu **intervali pouzdanosti** ili **intervali povjerenja** s **granicama pouzdanosti**

Granice pouzdanosti za pojedine intervale su tada:

gornja granica puzdanosti $\mu + 0.674\ \sigma$
za interval u kojem je 50% svih varijanata
donja granica puzdanosti $\mu - 0.674\ \sigma$

gornja granica puzdanosti $\mu + 1.960\ \sigma$
za interval u kojem je 95% svih varijanata
donja granica puzdanosti $\mu - 1.960\ \sigma$

gornja granica puzdanosti $\mu + 2.576\ \sigma$
za interval u kojem je 99% svih varijanata
donja granica puzdanosti $\mu - 2.576\ \sigma$

O njima ćemo još i detaljnije raspravljati u sljedećem poglavlju.

Može se slobodno reći da je **teoretska normalna distribucija osnova primijenjene statistike**.

Ako bismo to doslovno primijenili na uzorak u primjeru za duljinu klipa, gdje je $\bar{x}$ = 16.87 cm, a standardna devijacija s = 1.93 cm, mogli bismo zaključiti da je 99.72% klipova duljih od 11.08 cm (a to je $\bar{x}$ - 3s) a kraćih od 22.66 cm (to je $\bar{x}$ + 3s). Od preostalih 0.28% (t.j. 100 - 99.72) klipova, 0.14% ih je duljih od 22.66 cm, a 0.14% kraćih od 11.08 cm (nalaze se u onom zatamnjenom dijelu distribucije na slici 7.).

3.2.1. Položaj pojedine varijante u distribuciji; z-vrijednosti

Ako se neka varijanta distribucije nalazi točno na 1, 2 ili 3 standardne devijacije lijevo ili desno od sredine, tada točno znamo koliko je % varijanata većih ili manjih od nje.

Međutim, što ako je na nekom drugom mjestu u distribuciji?

Položaj bilo koje varijante u distribuciji je moguće odrediti pomoću z-vrijednosti. To je udaljenost varijante od srednje vrijednosti, izražena u dijelovima standardne devijacije:

$$z = \frac{x - \bar{x}}{s}$$

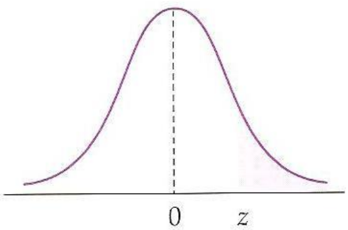
Ako za neku varijantu izračunamo z vrijednost, na primjer, 1.28, to znači da se ta varijanta nalazi na 1.28 standardne devijacije desno od srednje vrijednosti. Jednako tako vrijednost z = –0.94 znači da se određena varijanta nalazi na 0.94 standardne devijacije ispod prosjeka, tj. lijevo od srednje vrijednosti (predznak!).

Iz tablice površine ispod normalne krivulje (slika 9. je dio tablice A u dodatku) moguće je na temelju z-vrijednosti očitati koliko se % varijanata u distribuciji nalazi između određene varijante pretvorene u z-vrijednost i bližeg kraja krivulje.

Površina ispod cijele krivulje je 100% (u tablicama 100% = 1.0). S obzirom da je normalna krivulja distribucija vjerojatnosti, to iz tablica *očitane vrijednosti znače površinu ispod krivulje, a ujedno i* vjerojatnost varijanata koje su veće ili manje od zadane varijante.

Kako koristiti tablicu?

U prvom stupcu tablice se nalaze z-vrijednosti izražene s jednom decimalom, a drugu decimalu pronalazimo u prvom redu na vrhu tablice.



SLIKA 9.
Dio tablice A u dodatku

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681

Tako recima za $z = 1.28$ očitamo vrijednost na mjestu gdje se sijeku: redak koji počinje sa 1.2 i stupac 0.08, što je 0.1003. To znači da je 10,03% varijanata većih od varijante čija je z vrijednost 1.28.

Isto tako iz $z = -0.94$ očitamo vrijednost 0.1736 koja kaže da je 17.36% varijanata manjih od varijante izražene u $z = -0.94$.

Negativna vrijednost z znači da tablična vrijednost govori o % varijanata u dijelu krivulje lijevo od $\overline{x}$ (manje od $\overline{x}$).

U našem primjeru s duljinom klipa (iz $n = 50$) gdje je srednja vrijednost bila $\overline{x} = 17.12$ cm a standardna devijacija $s = 1.94$ cm, za klip duljine recimo $x = 15.0$ cm to bi značilo:

$$z = \frac{15.0 - 17.12}{1.94} = -1.093$$

Iz tablice očitana vrijednost je 0.1379, a znači da je 13.79% od svih 50 klipova manjih od $x = 15.0$ cm. To ujedno znači da je 86.21% klipova koji su dulji od 15.0 cm (jer 1.0 - 0.1357 = 0.8621 = 86.21%).

Na taj način moguće je pronaći položaj svake (bilo koje) varijante u distribuciji.

■

1
2
3

4 Procjena parametara populacije preko vrijednosti iz uzorka

5
6
7

4.1. Granice pouzdanosti srednje vrijednosti
4.2. Studentova "t-distribucija"

Kako je već više puta naglašeno, populacije (zbog njihove veličine ili čak beskonačnosti) proučavamo preko reprezentativnih i dovoljno velikih uzoraka. Iz svake populacije je moguće izdvojiti velik broj istih ili različito velikih uzoraka. Ako su varijante populacije normalno raspoređene, tada će i uzorci uzeti iz te populacije slijediti normalnu distribuciju.

Kao što su varijante u populaciji raspoređene oko srednje vrijednosti populacije (μ), tako isto su varijante svakog uzorka normalno distribuirane oko aritmetičke sredine $\overline{x}$ uzorka (slika 10).

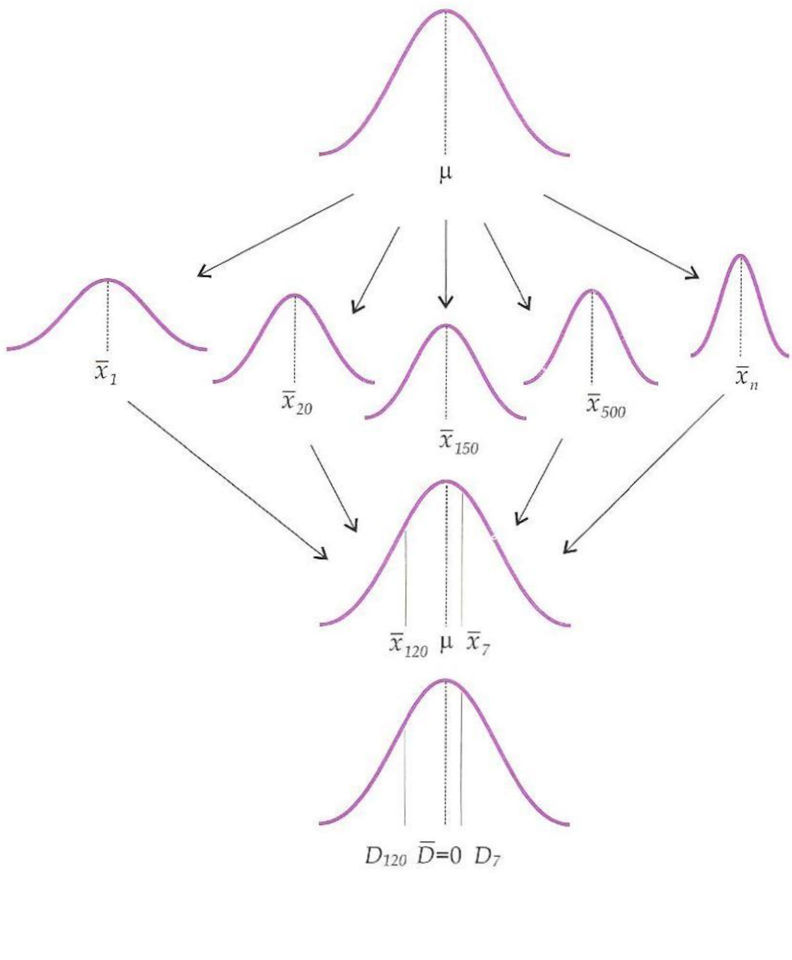
SLIKA 10.

Distribucija frekvencija u populaciji s prosječnom vrijednosti μ i standardnom devijacijom σ

Distribucije varijanata uzoraka s prosječnim vrijednostima $\overline{x}_i$ i standardnim devijacijama s_i

Distribucija prosječnih vrijednosti uzoraka ($\overline{x}_i$) oko μ populacije, sa standardnom devijacijom-standardnom pogreškom srednje vrijednosti ($s_{\overline{x}}$)

Distribucije razlika $D_i = \overline{x} - \mu$ oko prosječne razlike ($\overline{D} = 0$) sa standardnom devijacijom-standardnom pogreškom razlike (s_D)



Ako se iz svih prosječnih vrijednosti uzoraka ($\overline{x}_1 ... \overline{x}_n$) konstruira distribucija, tada se opet dobije normalna distribucija prosječnih vrijednosti uzoraka oko prave prosječne vrijednosti populacije (μ) iz koje su uzorci uzeti.

Iz slike je uočljivo da je prosječna vrijednost, na primjer, uzorka 7 ($\overline{x}_7$) bliža μ populacije iz koje su uzorci uzeti, nego $\overline{x}_{120}$ koja se više razlikuje od μ. Ili razlika μ - $\overline{x}_7$ je manja od razlike μ -$\overline{x}_{120}$.

Sve one prosječne vrijednosti uzoraka koje su gotovo iste ili jednake kao μ populacije, razlikuju se od nje minimalno ili ništa, pa su razlike blizu nuli ili nula. I razlike između prosječne vrijednosti populacije i prosječnih vrijednosti uzoraka ($D_i = \overline{x}_i - \mu$) također su normalno distribuirane oko prosječne razlike nula ($\overline{D} = 0$).

U svakoj toj distribuciji odstupanje varijanata od prosjeka distribucije mjeri se standardnom devijacijom. Standardna devijacija u distribuciji varijanata u populaciji je σ (sigma), u distribuciji varijanata uzorka to je s, u distribuciji prosječnih vrijednosti uzoraka ($\overline{x}_i$) oko μ populacije to je $s_{\overline{x}}$ (standardna pogreška srednje vri-jednosti), a u distribuciji razlika (D_i) to je standardna pogreška razlike (s_D).

Kako su sve ovo normalne distribucije, to za standardnu devijaciju svake distribucije vrijede opisane zakonitosti normalne distribucije.

Te zakonitosti se koriste u donošenju statističkih zaključaka.

Ako je varijabla normalno distribuirana, za sve varijante populacije koje se nalaze izvan $\mu \pm 3\sigma$, mala je vjerojatnost da uopće pripadaju u tu populaciju (jer 99.72% svih varijanata je unutar tih granica, a samo ostatak od 0.28% izvan).

Analogno tome u distribuciji prosječnih vrijednosti uzoraka ($\overline{x}_i$) oko prosječne vrijednosti populacije (μ), moguće je definirati odstupanja u vidu standardne devijacije i varijance.

Matematičari su izračunali tzv. očekivanu vrijednost varijance i standardne devijacije prosječne vrijednosti:

$$\sigma_{\overline{x}}^2 = \frac{\sigma^2}{n}, \qquad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Iz formule je očito da je standardna devijacija prosječnih vrijednosti uzoraka, oko μ populacije, funkcija standardne devijacije i veličine uzorka. Što je uzorak veći, to manja će biti standardna devijacija prosječne vrijednosti.

Kako se povećava veličina uzorka, standardna devijacija prosječnih vrijednosti postaje zanemarivo mala. To je logično. Naime, vrlo veliki uzorci koji sadrže vrlo velik broj varijanata, dat će manje varijabilne procjene srednje vrijednosti nego oni bazirani na malom broju varijanata (veća je naime mogućnost da su u velikom uzorku zastupljene sve tipične varijante iz populacije).

Radeći s uzorcima, ne znamo pravu standardnu devijaciju populacije iz koje su uzorci uzeti (σ). Nju možemo samo procijeniti pomoću standardne devijacije uzorka (s).

Isto tako, obično nemamo velik broj uzoraka iz kojih bi pojedinačno procjenjivali standardnu devijaciju populacije kojoj pripadaju, nego analiziramo samo jedan uzorak. **Zato će standardna devijacija prosječne vrijednosti biti:**

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

U uzorku se mogu procijeniti različiti parametri kao: prosječna vrijednost, medijana, modus, standardna devijacija, varijanca, varijacijska širina i dr. Ako se ovi parametri procijene iz svakog od velikog broja uzoraka uzetih iz neke populacije, moguće je dobiti distribucije frekvencija ovih parametara oko pravih parametara populacije, odnosno za svaki od njih (baš kao što smo to učinili za $\bar{x}$) izračunati odstupanje od njegove prave vrijednosti u populaciji pomoću standardne devijacije. Standardne devijacije ovih parametara zovu se **standardne pogreške**. Tako je $s_{\bar{x}}$ **standardna pogreška srednje vrijednosti**.

Dakle, radeći s uzorcima mogu se više ili manje točno procijeniti pravi parametri populacije. Parametri uzorka su stoga samo njihove "približne" vrijednosti. Svakako, procjena neke vrijednosti bit će to bolja, što je uzorak na temelju kojeg ju procjenjujemo veći i što je svojstvo manje varijabilno.

Kad ne bi bilo varijabilnosti tada bi već jedna varijanta (jer ostale su u tom slučaju iste) bila dovoljna. Međutim svojstva s kojima se u biološkim istraživanjima susrećemo (koja proučavamo, pratimo, opažamo, mjerimo) imaju karakterističnu varijabilnost. Zato, što neka pojava ili svojstvo više varira, to se većoj pogrešci izlažemo kad na temelju parametara uzorka zaključujemo o pravim vrijednostima populacije. Isto tako, većoj se pogrešci izlažemo radeći s malim, u odnosu na velike uzorke.

Na varijabilnost nekog svojstva ne može se utjecati. Ali zato možemo povećati broj varijanata u uzorku (broj mjerenja, opažanja, n).

Iz formule za standardnu pogrešku srednje vrijednosti to se lako može uočiti, jer

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

Računski: što je brojnik manji a nazivnik veći, kvocijent je manji. To znači da je *standardna pogreška srednje vrijednosti to manja (dakle srednja vrijednost uzorka je to bliže pravoj srednjoj vrijednosti populacije) što je varijabilnost manja i što je veći broj mjerenja.* Ali, pogreška ne opada proporcionalno broju mjerenja, već korijenskoj vrijednosti tog broja.

Kako $n \to \infty$ tako $s_{\bar{x}} \to 0$. To ujedno znači, da kako broj mjerenja raste, tako je srednja vrijednost uzorka ($\bar{x}$) sve bolja procjena prave srednje vrijednosti populacije (μ). Drugim riječima, **standardna pogreška srednje vrijednosti nije ništa drugo nego standardna devijacija u distribuciji prosječnih vrijednosti uzoraka** $\bar{x}_i$ **oko prave prosječne vrijednosti populacije** μ (slika 10).

Isto tako se u distribuciji razlika oko prosječne razlike $\bar{D} = 0$ odstupanje svih razlika D_i od prosjeka $\bar{D} = 0$ mjeri standardnom devijacijom, koja se u ovoj distribuciji zove **standardna pogreška razlike** (s_D).

Stoga i za standardnu pogrešku srednje vrijednosti ($s_{\bar{x}}$) i za standardnu pogrešku razlike (s_D) vrijede iste zakonitosti kao i za standardnu devijaciju.

4.1. Granice pouzdanosti srednje vrijednosti

Razmatrajući odnos između parametara uzorka i analognih pravih vrijednosti populacije, nismo se zadržavali na tome koliko su ti, iz uzorka procijenjeni parametri pouzdani - kako dobro predstavljaju parametre populacije.

Naravno, ono što stvarno želimo znati, to su vrijednosti μ, σ, σ^2. Međutim, to je praktički nemoguće (osim kad bi se uzimao beskonačno velik broj uzoraka iz jedne populacije). Zato treba znati ocijeniti pouzdanost parametara uzoraka s obzirom na parametre populacije.

Uzmimo za primjer jedan neuobičajen slučaj da razmotrimo populaciju s poznatom srednjom vrijednosti i standardnom devijacijom.

Srednju vrijednost uzorka veličine n označavamo s $\overline{x}$. Očekivana standardna pogreška srednje vrijednosti je

$$s_x = \frac{\sigma}{\sqrt{n}}$$

Kako su prosječne vrijednosti uzoraka normalno distribuirane oko μ populacije, to područje 1.96 $\sigma_{\overline{x}}$ ispod μ do 1.96 $\sigma_{\overline{x}}$ iznad μ uključuje 95%, a ono 2.576 $\sigma_{\overline{x}}$ ispod μ i 2.576 iznad μ 99% svih prosječnih vrijednosti uzoraka veličine n.

To se može pisati i kao:

$$P\left\{\overline{x} - 1.96\sigma_{\overline{x}} \leq \mu \leq \overline{x} + 1.96\sigma_{\overline{x}}\right\} = 0.95$$

$$P\left\{\overline{x} - 2.576\sigma_{x} \leq \mu \leq \overline{x} + 2.576\sigma_{x}\right\} = 0.99$$

To znači 95%-tnu, odnosno 99%-tnu vjerojatost da prosječna vrijednost uzorka, ne odstupa od prave μ populacije kojoj uzorak pripada za više od $\pm$1.96 $\sigma_{\overline{x}}$, odnosno $\pm$2.576 $\sigma_{\overline{x}}$.

Govorimo, zapravo, o granicama pouzdanosti, pa ove dvije vrijednosti zovemo **donjom i gornjom granicom pouzdanosti srednje vrijednosti,** a interval između njih **intervalom povjerenja. One ujedno znače 5% i 1% vjerojatnosti pogreške, a to se označava kao P=5% ili P=0.05, odnosno P=1% ili P=0.01 (P- od probabilitet).**

Iako se ovo odnosilo na uzorak uzet iz populacije poznate μ i σ, analogno se može protegnuti na uzorke uzete iz populacije nepoznatih parametara, ali uz uvjet da slijede normalnu distribuciju i da su uzorci veliki. U takvim slučajevima uzima se standardna devijacija uzorka (s) za procjenjivanje standardne pogreške srednje vrijednosti ($s_{\overline{x}}$).

Pomoću odnosa između $\overline{x}$ i pogreške srednje vrijednosti ($s_{\overline{x}}$) koja je standardna devijacija u distribuciji srednjih vrijednosti uzoraka oko prave srednje vrijednosti populacije kojoj uzorci pripadaju (μ), moguće je ustvrditi koliko srednja vrijednost uzorka odstupa od prave srednje vrijednosti populacije iz koje je uzorak uzet.

Ako se to primijeni u primjeru s duljinom klipa kukuruza gdje je n = 50, $\overline{x}$ = 16.87, a s = 1.93 cm standardna pogreška srednje vrijednosti bi bila:

$$s_x = \frac{1.93}{\sqrt{50}} = 0.273 \text{ cm}$$

To znači, da srednja vrijednost procijenjena iz uzorka ne odstupa od prave srednje vrijednosti populacije više od ±0.819 cm (tj. $\pm 3s_{\bar{x}}$ ili ±0.819). Prava prosječna vrijednost populacije kojoj uzorak pri-pada je negdje između 16.05 cm (to je $\bar{x} - 3s_x$) i 17.69 cm (a to je $\bar{x} + 3s_x$).

No, opet treba naglasiti da ovo nije potpuno točno, jer ove granice pouzdanosti vrijede za vrlo veliki n. Kolike su one za ovakav uzorak gdje je n = 50 saznat ćemo kroz sljedeće razmatranje.

4.2. Studentova "t-distribucija"

Sve distribucije koje smo do sada razmatrali bile su normalne distribucije varijanata oko svojih prosječnih vrijednosti, uz standardne devijacije - mjerila odstupanja njihovih varijanata od prosječne vrijednosti. Naime, i

x_1 distribucija varijanata u populaciji oko μ populacije sa standardnom devijacijom σ

$\vdots$

$x_{n \to \infty}$

i

x_1 distribucija varijanata u uzorku oko $\bar{x}$ uzorka sa standardnom devijacijom s

$\vdots$

x_n

i

$\bar{x}_1$ distribucija prosječnih vrijednosti uzoraka ($\bar{x}_i$) oko μ populacije sa standardnom devijacijom $s_{\bar{x}}$

$\vdots$

$\bar{x}_n$

i

$\bar{x}_1 - \mu = D_i$ distribucija razlika prosječnih vrijednosti uzoraka i μ populacije sa standardnom devijacijom s_D

$\vdots$

$\bar{x}_n - \mu = D_n$

su normalne distribucije.

Ako bi se razlike izrazile u dijelovima standardne devijacije tj.

$$\frac{\overline{x}_1 - \mu}{\sigma} = z_1$$
$$\vdots$$
$$\frac{\overline{x}_n - \mu}{\sigma} = z_n$$

i to bi bila normalna distribucija.

Međutim, praktički nikad ne znamo standardnu devijaciju populacije, nego samo standardnu devijaciju uzorka, pa ovi izrazi prelaze u oblike $\frac{\overline{x}_1 - \mu}{s / \sqrt{n}}$ tj.

$$\frac{\overline{x}_1 - \mu}{s / \sqrt{n}} = t_1$$ zovu se **t-faktori** i više nisu normalno distribuirani, nego čine **t-distribuciju** ili **Studentovu distribuciju**

$$\vdots$$
$$\frac{\overline{x}_n - \mu}{s / \sqrt{n}} = t_n$$

t-faktor je dakle odnos razlike i standardne pogreške razlike, pa kako je $s_x = \frac{s}{\sqrt{n}}$ to je

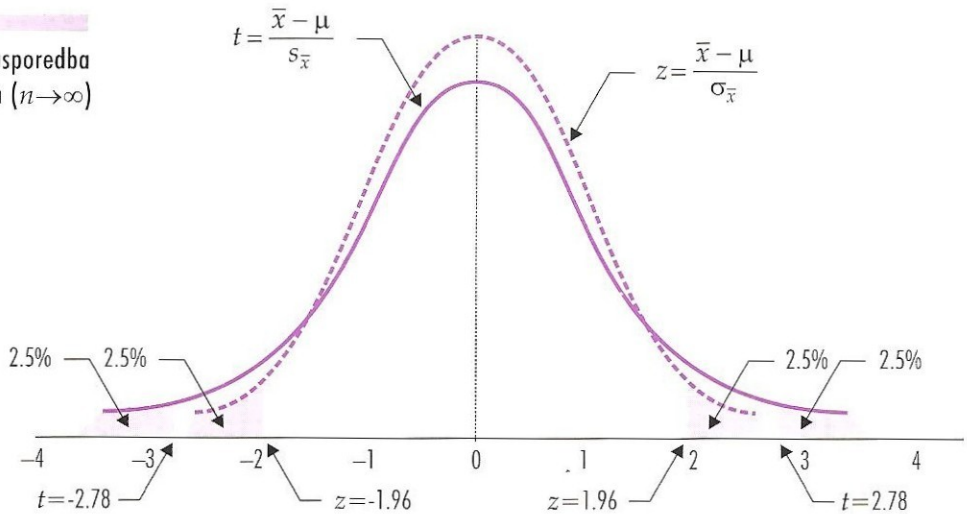
$$\frac{\overline{x} - \mu}{s_{\overline{x}}} = t$$

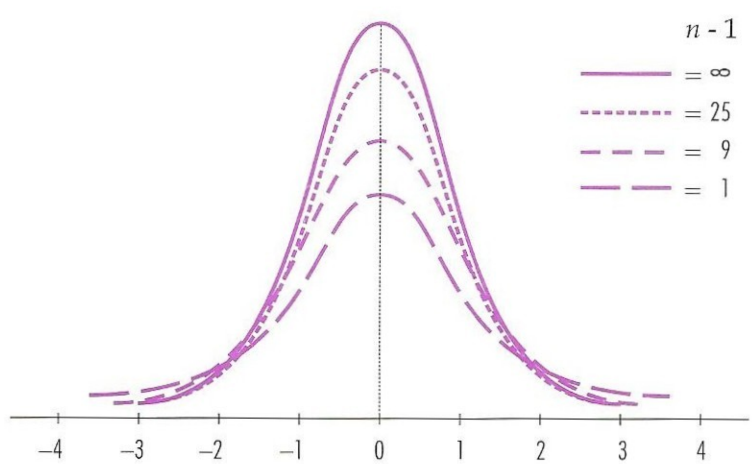
t-distribuciju ili **Studentovu distribuciju** je otkrio William Sealy Gosset 1908. (pod pseudonimom Student), a usavršio R. A. Fisher 1926. Ova je distribucija označila revoluciju u statistici malih uzoraka.

Kao i normalna distribucija, t-distribucija je simetrična i proteže se od -∞ do +∞. Od normalne distribucije se razlikuje i po tome što može poprimiti različite oblike ovisno o n odnosno broju slobodnih varijanata (n-1). Moglo bi se reći da se t-distribucija razlikuje od normalne distribucije po tome što je krivulja t-distribucije uvijek nižeg vrha i šira prema krajevima nego krivulja normalne distribucije. Što je n manji to je različitost veća, a kako n raste, Studentova t-distribucija poprima oblik sve bliži normalnoj distribuciji, pa kad

n dosegne beskonačno, t-distribucija postaje normalna distribucija. Kod normalne distribucije točno se zna da se u intervalu $\mu \pm 1.96\sigma$ nalazi 95% od svih varijanata, a interval $\mu \pm 2.576\sigma$ sadrži 99% varijanata. Što je uzorak manji t-distribucija je šira prema krajevima, pa će se tako mijenjati intervali i postoci varijanata u njima, odnosno granice pouzdanosti (slike 12 i 13).

SLIKA 11.
t-distribucija (za n=5) i usporedba s normalnom distribucijom ($n \to \infty$)

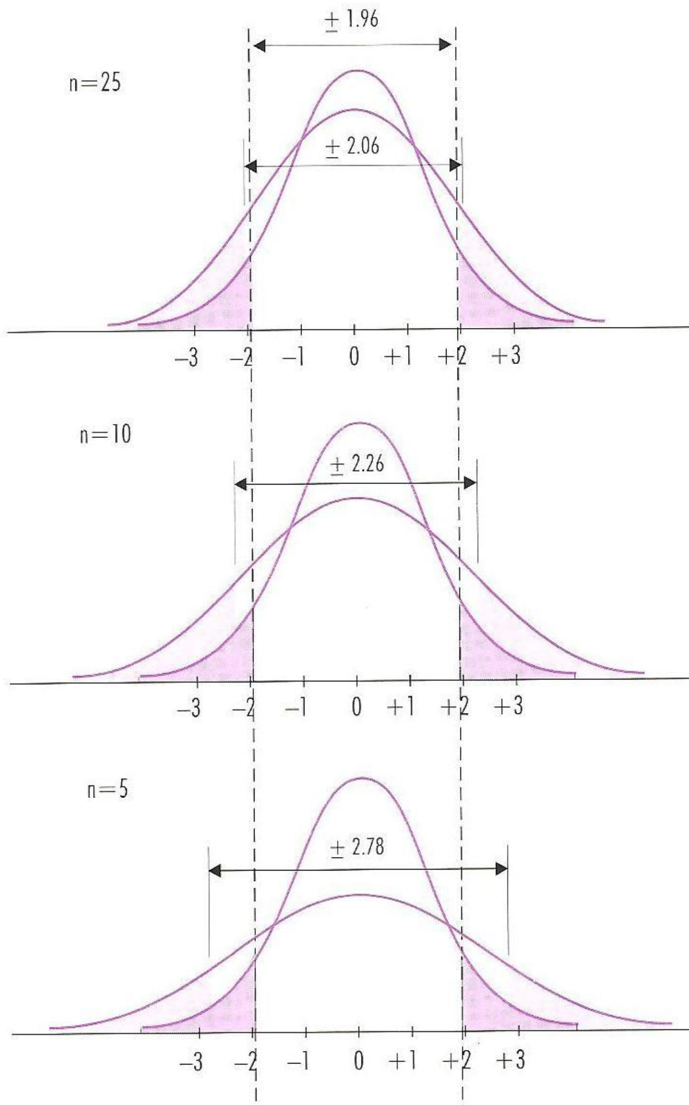




SLIKA 12.
Različite t-distribucije, ovisno o broju slobodnih varijanata (n-1)

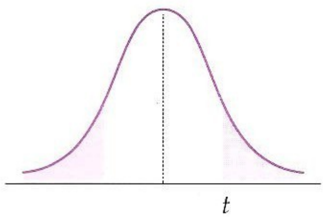
Tako u slučaju da je n=5, 95% varijanata je tek unutar intervala od $\overline{x}$ ± 2.78s. Analogna vrijednost kod teoretske normalne distribucije je ± 1.96s (slika 11).

Dakle, t vrijednosti su ovisne o broju varijanata u uzorku ili točnije o broju slobodnih varijanata ili o stupnjevima slobode (n-1), pa je t-distribucija osnova za utvrđivanje granica pouzdanosti kod malih uzoraka.



SLIKA 13.
95 %-tne granice pouzdanosti u normalnoj distribuciji ($n \rightarrow \infty$) i u distribucijama s n=5, n=10 i n=25

Zato su razrađene **t tablice** (na slici 14 je dio tablice B iz dodatka), iz kojih se za svaku veličinu uzorka, iz n-1 može očitati pripadna vrijednost t, jer bi za mali uzorak vrijednosti 1.96 i 2.576, kako vidimo, bile prestroge i neispravne.



n-1	P			
	0.10	0.05	0.01	0.001
1	6.31	12.71	63.66	632.60
2	2.92	4.30	9.92	31.60
3	2.35	3.18	5.84	12.92
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.90	2.36	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
-				
-				
-				
21	1.72	2.08	2.83	3.82
22	1.72	2.07	2.82	3.79
23	1.71	2.07	2.81	3.77
24	1.71	2.06	2.80	3.74
25	1.71	2.06	2.79	3.72

SLIKA 14.
Dio tablice B u dodatku

Uzmimo primjer distribucija sa slike 13, za n=5, n=10, n=25 i $n \to \infty$. Iz t tablica očitamo t faktore za svaku, a uz vjerojatnost P=0.05 (analogno bi postupili i za P = 0.01, ali radi jasnosće slike zadržat ćemo se samo na P=0.05). Oni su: 2.78 (za 5–1), 2.26 (za 10–1), 2.06 (za 25–1) i 1.96 (za $n \to \infty$).

To znači da je 0.95 ili 95% površine unutar vrijednosti ± 2.78, ± 2.26, ±2.06 i ±1.96 u svakoj od ovih distribucija. Isti, dakle, postotak od ukupnog broja varijanata nalazi se unutar to širih granica, što je n manji.

Ili, na drugi način rečeno, samo 0,05 ili 5% (odnosno 2,5% sa svakog kraja distribucije) je izvan vrijednosti ± 2.78, odnosno ± 2.26, odnosno ± 2.06, odnosno ± 1.96.

Stoga se za ***male uzorke određenog n*-a** (koji je uvijek manji od onog u teoretskoj normalnoj distribuciji gdje $n \to \infty$) ***uzete iz populacije normalno raspoređenih varijanata, granice pouzdanosti srednje vrijednosti računaju kao***

$$\bar{x} - ts_x \qquad \text{i} \qquad \bar{x} + ts_x$$

pri čemu se vrijednost t očita iz tablica uz odgovarajuću vjerojatnost pogreške (P). Obično se uzima da 5% i 1% vjerojatnosti znače dovoljnu pouzdanost (jer je to 95 odnosno 99 postotna sigurnost). Za naš primjer uzorka za duljinu klipa kukuruza granice pouzdanosti srednje vrijednosti izračunale bi se iz:

$\bar{x} = 16.87$	iz tablica očitan t-faktor (iz n-1= 49),
$s = 1.93$	a uz dvije razine sigurnosti iznosi
$s_{\bar{x}} = 0.273$	$t_{0.05} = 2.01$
$n = 50$	$t_{0.01} = 2.68$

donja granica pouzdanosti ($\bar{x} - ts_{\bar{x}}$)

za $P = 0.05$ iznosi 16.87 - (2.01 · 0.273) = 16.87 - 0.549 = 16.32

za $P = 0.01$ iznosi 16.87 - (2.68 · 0.273) = 16.87 - 0.732 = 16.14

a gornja granica pouzdanosti ($\bar{x} \pm ts_{\bar{x}}$):

za $P = 0.05$ iznosi 16.87 + (2.01 · 0.273) = 16.87 + 0.549 = 17.42

za $P = 0.01$ iznosi 16.87 + (2.68 · 0.273) = 16.87 + 0.732 = 17.60

To znači da se može s 99% sigurnosti ili uz $P = 0.01$ vjerojatnosti pogreške (što naravno uključuje i 95% sigurnosti ili $P = 0.05$) tvrditi da prosječna vrijednost procijenjena iz uzorka ne odstupa od prave prosječne vrijednosti populacije više od ±0.732 cm.

Ili: s 99% sigurnošću možemo tvrditi da se prava prosječna vrijednost populacije iz koje je uzorak uzet nalazi između 16.14 i 17.60 cm. tj.

$$\bar{x} - t_{0.01} \cdot s_x \leq \mu \leq \bar{x} - t_{0.01} \cdot s_x$$

Ili: uz vjerojatnost $P = 0.01$ da griješimo (a to znači da se u 100 slučajeva može samo jedanput dogoditi da to nije istina) možemo reći da prava prosječna duljina klipa populacije kojoj uzorak pripada nije manja od 16.14 cm. niti veća od 17.60 cm.

Ili: $\bar{x}$ je pouzdana uz vjerojatnost P = 0.01 (dakle uz 99%-tnu sigurnost. koja dakako uključuje i 95%-tnu).

Treba primijetiti: u prethodnom smo poglavlju (str. 54). računajući granice pouzdanosti srednje vrijednosti za ovaj isti primjer primijenili kriterij $\pm 3s_{\bar{x}}$ i naglasili da nije potpuno ispravan jer vrijedi za vrlo veliki n. Zato su tako dobivene granice pouzdanosti srednje vrijednosti bile 16.05 i 17.69 cm. Kako sad znamo. ispravno izračunane vrijednosti koje uzimaju u obzir veličinu uzorka su 16.14 i 17.60 cm.

Tako ih uvijek i treba računati.

■

5 Nulta hipoteza i testiranje nulte hipoteze

5.1. Testiranje nulte hipoteze o razlici između prosječnih vrijednosti

5.2. t-test

Najčešća primjena statistike u biološkim istraživanjima je provjeravanje ili testiranje različitih znanstvenih postavki - hipoteza.

Naime, eksperimentalni rezultati često nisu potpuno jasni, pa stoga zahtijevaju primjenu statističkih postupaka koji pomažu pri odlučivanju između alternativnih postavki.

Statističkim testovima se ispituju eksperimentalni podaci i na osnovi očekivane distribucije dolazi se do odluke prihvatiti ili odbaciti pretpostavku (hipotezu).

Proučavajući izvjesnu pojavu, istraživač ima određenu pretpostavku o njoj.

Recimo, križajući rajčicu crvenog i žutog ploda, poznavajući izvjesne zakonitosti pri nasljeđivanju, istraživač u potomstvu očekuje crvene i žute plodove u omjeru 3:1. Ako je, recimo, od ukupno dobivenih 400 plodova bilo 250 crvenih a 150 žutih, što će zaključiti?

Ili liječnik, koji pretpostavlja da od neke bolesti podjedako obolijevaju oba spola, od 900 oboljelih bilo je 480 muškaraca i 420 žena. Kakav on sud iz ovih podataka može donijeti o svojoj pretpostavki?

Da se odgovori na ovakva pitanja, potrebno je imati izvjesna mjerila odstupanja uzorka od populacije.

Oba istraživača iz gornjih primjera imaju svoje pretpostavke (hipoteze) da će se rezultati ponašati u omjeru 3:1, odnosno 1:1. Oni, dakle, pretpostavljaju da će se praktični rezultati slagati s njihovom hipotezom, odnosno da neće od nje odstupati.

To je tzv. **nulta hipoteza** ili pretpostavka da nema razlike između pretpostavljene (hipotetične) i u uzorku realizirane situacije. Najčešće se ta nulta hipoteza označava s H_0.

Mogli bismo općenito reći da je **nulta hipoteza svaka moguća pretpostavka koju provjeravamo, a ovo "nulta" se koristi zbog toga što je smisao "anulirati", poništiti, smatrati podudarnim**.

Da li je nulta hipoteza istinita ili nije, može se utvrditi statističkim provjeravanjem njezine istinitosti, odnosno **testiranjem nulte hipoteze**. Testiranjem H_0 može se dokazati da je istinita - pa govorimo o ***prihvaćanju nulte hipoteze***, ili pak dokazati da ona nije istinita, što znači ***odbacivanje nulte hipoteze***.

Pitanje je uz koju pouzdanost, odnosno vjerojatnost pogreške bi se smjelo prihvatiti ili pak odbaciti H_0.

Valja opet naglasiti da se u statistici primijenjenoj na većinu problema u poljoprivrednim istraživanjima dogovorno najčešće dopušta vjerojatnost pogreške najviše od 0.05 (P = 0.05 ili P = 5%), a to znači najmanje 95% sigurnosti. Govorimo o ***razini značajnosti, opravdanosti*** *ili* ***signifikantnosti***. Testirajući H_0 možemo je, dakle, odbaciti uz P = 5% (P = 0.05) što znači od 100 slučajeva 5 puta pogriješiti, ili P = 1% (P = 0.01) čime se dozvoljava samo 1 pogreška na 100 slučajeva (tj. 95% i 99% sigurnosti, opravdanosti ili signifikantnosti).

P = 0.05 i P = 0.01 je doduše dogovor, međutim na istraživaču je da odluči koju će vjerojatnost pogreške tolerirati. Usput, statistički programi (software) koji se danas koriste, omogućavaju da istraživač odluči uz koju će vjerojatnost pogreške prihvatiti ili odbaciti H_0.

Odbacivanje nulte hipoteze uz određenu vjerojatnost P zapravo definira ***statističku opravdanost, značajnost*** ili ***signifikantnost***.

Testirati se može i svaki iz uzorka procijenjeni parametar (bila to prosječna vrijednost, standardna devijacija, varijanca, varijacijski koeficijent ili dr.) i utvrditi koliko je on dobar procjenitelj pravog parametra populacije; znači utvrditi njegovu opravdanost.

5.1. Testiranje nulte hipoteze o razlici između prosječnih vrijednosti

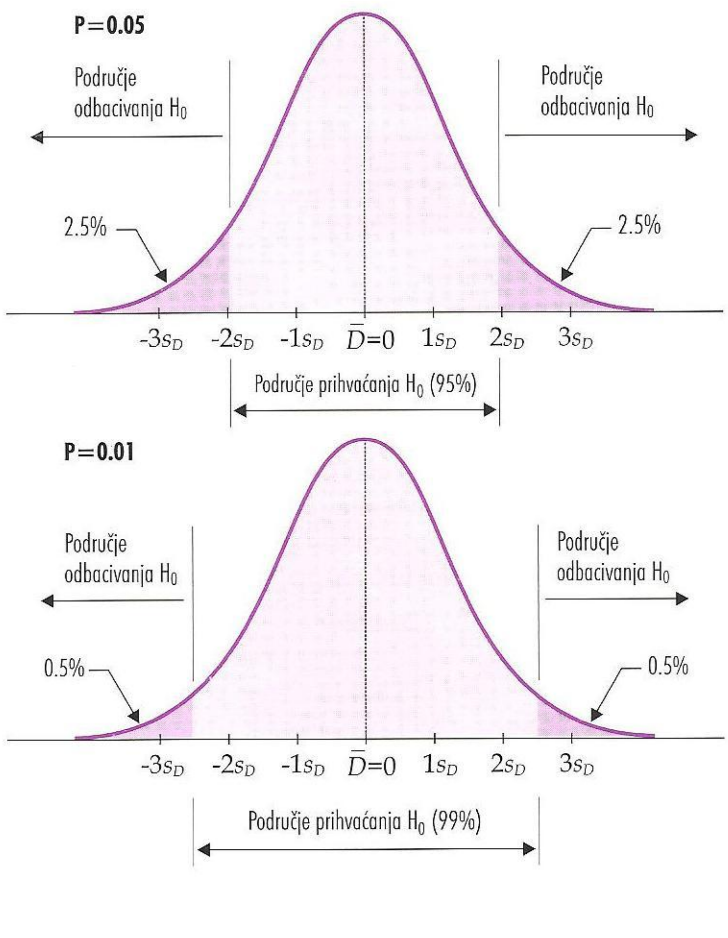
Pođimo opet od populacije, s brojnim uzorcima koji se iz nje mogu izdvojiti i s odstupanjima prosječnih vrijednosti uzoraka ($\overline{x}_i$) od prosječne vrijednosti populacije (μ) iz koje su uzorci uzeti (slika 10). Kako je i distribucija razlika oko prosječne razlike ($\overline{D} = 0$) normalna distribucija, i u njoj je moguće izraziti varijabilnost pomoću stadardne devijacije (slika 15).

Standardna devijacija u distribuciji razlika između prosječnih vrijednosti uzoraka i prosječne vrijednosti populacije je **standardna pogreška razlike** ili s_D.

I za nju, rekli smo, vrijede iste zakonitosti koje vrijede u normalnoj distribuciji. Naime 99.72% svih varijanata (dakle razlika D_i) nalazi se u intervalu ± 3 standardne devijacije (s_D) od sredine distribucije (ovdje je to $\overline{D} = 0$). Ili, s prosječnom vrijednosti ove distribucije ($\overline{D} = 0$) može se predstaviti 99.72% svih varijanata (dakle onih koje su ili veće od $\overline{D} - 3s_D$ ili manje od $\overline{D} + 3s_D$). Ili, vrlo je mala vjerojatnost da razlike koje su izvan ovog intervala pripadaju toj distribuciji, čija prosječna vrijednost je nula.

Stoga je za bilo koju razliku D moguće ustvrditi gdje se u takvoj distribuciji nalazi. Praktično, ako je unutar $\overline{D} \pm 1.96s_D$ odnosno $\overline{D} \pm 2.576s_D$ može se predstaviti prosjekom ove distribucije (a koji je nula), može se smatrati zanemarivo malom. Za takvu razliku kažemo da pada u **područje prihvaćanja nulte hipoteze** (i to uz 95 ili 99% sigurnost) ili **nije signifikantna**.

Obratno, svaka ona razlika koja je izvan intervala $\overline{D} \pm 1.96s_D$ odnosno $\overline{D} \pm 2.576s_D$ kažemo da je u **području odbacivanja nulte hipoteze** (također uz P = 0.05 ili P = 0.01), nije zanemariva, nego opravdana, signifikantna.



SLIKA 15.
Područja prihvaćanja H_0 i odbacivanja H_0

Primijenimo li ovo na bilo koje dvije prosječne vrijednosti dvaju uzoraka i razliku između njih, moramo imati na umu da se distribucije mijenjanju i ovisno o varijabilnosti (izraženoj sa s) i ovisno o veličini uzorka (n).

Ako prosječnu vrijednost jednog uzorka veličine n_1 označimo $\bar{x}_1$, pripadajuću standardnu devijaciju s_1, standardnu pogrešku $s_{\bar{x}_1}$, a isti ti parametri u drugom uzorku veličine n_2 neka su $\bar{x}_2$, s_2 i $s_{\bar{x}_2}$, tada je $D = \bar{x}_1 - \bar{x}_2$, a standardna pogreška razlike s_D će se izračunati kao korijenska vrijednost iz sume kvadratnih vrijednosti standardnih pogrešaka ovih dviju srednjih vrijednosti:

$$s_D = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}$$

Kako je standardna pogreška srednje vrijednosti $s_x = \frac{s}{\sqrt{n}}$, to je standardna pogreška razlike između dvije srednje vrijednosti:

$$s_D = \sqrt{\left(\frac{s_1}{\sqrt{n_1}}\right)^2 + \left(\frac{s_2}{\sqrt{n_2}}\right)^2}$$

$$s_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

5.2. t-test

5.2.1. Nezavisni uzorci

Nulta hipoteza o razlici između dvije prosječne vrijednosti je pretpostavka da je razlika takva da se može zanemariti, dakle da su te dvije prosječne vrijednosti vrlo, vrlo slične, da se razlikuju za zanemariv iznos, da uzorci iz kojih su ove prosječne vrijednosti izračunane pripadaju istoj populaciji.

To pišemo

$$H_0 : \bar{x}_1 = \bar{x}_2 \quad \text{ili} \quad H_0 : D = 0$$

Provjeravanje istinitosti H_0 o razlici prosječnih vrijednosti ilustrirati ćemo primjerima potpuno nezavisnih uzoraka i to najprije samo slikom i odnosom tazlike i njezine standardne pogreške.

GRAFIČKI

PRIMJER 5.2.1.1.

$\bar{x}_1 = 28.3$ $\bar{x}_2 = 25.0$
$s_1 = 3.1$ $s_2 = 3.9$
$n_1 = 70$ $n_2 = 65$

Razlika između ovih srednjih vrijednosti je:

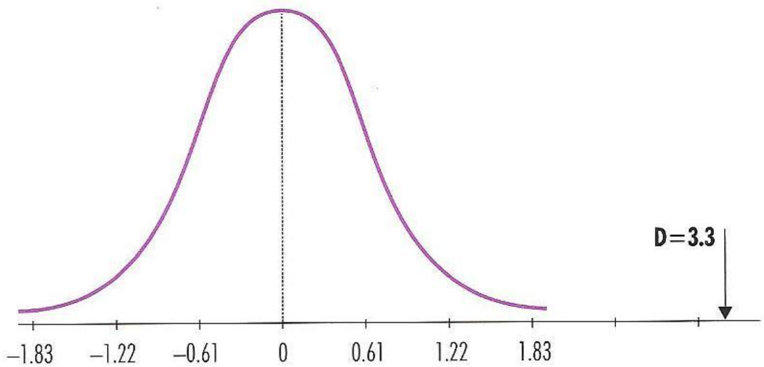
$$D = |\bar{x}_1 - \bar{x}_2| = 3.3$$

Kako saznati o toj razlici D? Nulta hipoteza je pretpostavka da je D zanemariva, t.j. $H_0 : D = 0$.

Istinitost H_0 moguće je provjeriti pomoću prije opisanih odnosa. Naime,

$$s_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$s_D = \sqrt{\frac{3.1^2}{70} + \frac{3.9^2}{65}} = 0.609$$



SLIKA 16.
Razlika $D = |\bar{x}_1 - \bar{x}_2|$ u području odbacivanja H_0

Razlika **D** = 3.3 nalazi se izvan intervala u kojem su sve one razlike (njih 99.72%) koje se mogu poistovjetiti s prosjekom ove distribucije, dakle zanemariti. Razlika se nalazi u području odbacivanja nulte hipoteze, ***pa zaključujemo: nulta hipoteza o razlici D = 3.3 nije bila istinita, razlika je značajna ili opravdana ili signifikantna (što su istoznačnice).***

PRIMJER 5.2.1.2.

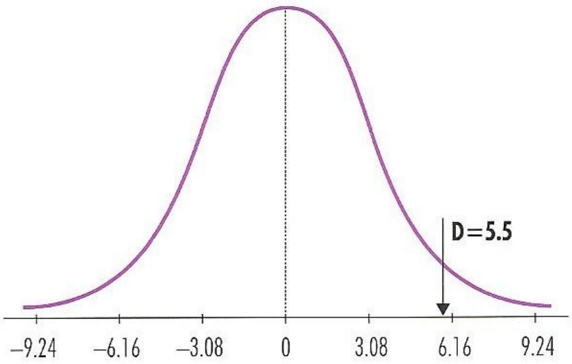
$$\bar{x}_1 = 49.5 \qquad \bar{x}_2 = 55.0$$
$$s_1 = 9.0 \qquad s_2 = 10.0$$
$$n_1 = 18 \qquad n_2 = 20$$

Razlika između ovih srednjih vrijednosti je:

$$D = |\bar{x}_1 - \bar{x}_2| = 5.5$$

$$s_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$s_D = \sqrt{\frac{9.0^2}{18} + \frac{10.0^2}{20}} = 3.082$$



SLIKA 17.
Razlika $D = |\bar{x}_1 - \bar{x}_2|$ u području prihvaćanja H_0

Razlika D = 5.5 pada dakle u područje prihvaćanja nulte hipoteze. To znači da je možemo smatrati zanemarivo malom, odnosno zaključiti da nije značajna (ili nije opravdana; ili nije signifikantna), spada, naime, u distribuciju razlika čija je prosječna vrijednost nula.

Kad bi, recimo, u prvom primjeru razlika bila $D = 1.5$ (dakle pala bi između $2s_D$ i $3s_D$), to bi značilo da je opravdana ili značajna uz 95% sigurnosti ili P = 0.05 vjerojatnost pogreške.

Jednako tako bi u drugom primjeru, ako bi razlika bila D=7.0, odbacili H_0 o njoj, ali opet samo uz 95% sigurnost (ili P = 0.05).

Ova dva primjera mogla bi pomoći razumijevanju vrlo općenitog zaključka: ako je razlika (D) između bilo koje dvije prosječne vrijednosti bar 2 odnosno 3 puta (ili točnije 1.96 i 2.576) veća od svoje standardne pogreške (s_D) ona će pasti u područje odbacivanja H_0 pa ju se može smatrati značajnom, opravdanom ili signifikantnom i to uz vjerojatnost od P = 0.05 odnosno P = 0.01.

Međutim, opet treba naglasiti da to vrijedi za velike uzorke, gdje zbog zakonitosti normalne distribucije 95%, odnosno 99% svih varijanata pada baš u intervale 1.96 s , odnosno 2.576 s.

ODNOS t_{exp} i t_{tabl}

Kod malih uzoraka, čije distribucije jače odstupaju od normalne, ovisno o njihovoj veličini (n), uvijek je potrebno posebno utvrditi koliko najmanje puta razlika mora biti veća od svoje pogreške da bi bila značajna ili signifikantna.

Faktor koji kaže gdje je zapravo granica područja prihvaćanja i odbacivanja nulte hipoteze o određenoj razlici, je ***t*** faktor. On se očita iz *t-tablica* (ovisno o veličini uzorka) koje su konstruirane za sve moguće vrijednosti *n-a*, pa ga se zove **tabličnim *t* faktorom (t_{tabl})**, za razliku od onog kojeg se izračuna iz podataka, a kojeg zovemo **eksperimentalnim *t* faktorom (t_{exp})**.

Eksperimentalni *t* faktor (t_{exp}) je broj koji kaže koliko je puta razlika (D) veća od svoje pogreške (s_D), a tablični *t* faktor (t_{tabl}) znači koliko najmanje puta mora razlika (D) biti veća od pogreške (s_D) da bi odbacili H_0 o toj razlici.

$$t_{exp} = \frac{D}{s_D}$$

t_{tabl} se očita iz t-tablica, iz $(n_1\text{-}1) + (n_2\text{-}1)$, tj. zbroja slobodnih varijanata uzoraka iz kojih su izračunane $\overline{x}_1$ i $\overline{x}_2$ koje se uspoređuju.

Usporedbom t_{exp} i t_{tabl} moguće je vrlo jednostavno i brzo računski provjeriti istinitost nulte hipoteze o razlici između bilo koje dvije prosječne vrijednosti.

Ako je

$$t_{exp} < t_{tabl} \rightarrow \text{prihvaća se } H_0 \text{ o razlici}$$

a ako je

$$t_{exp} > t_{tabl} \rightarrow \text{odbacuje se } H_0 \text{ o razlici}$$

Ovaj postupak testiranja razlike između dvije prosječne vrijednosti, poznat kao **Studentov t-test** ili jednostavno **t-test**, jedan je od najvažnijih testova u primijenjenoj statistici.

Primijenimo t test na prethodna dva primjera.

Za primjer 5.2.1.1., gdje je

$\overline{x}_1$ izračunana iz n_1 = 70 varijanata
$\overline{x}_2$ izračunana iz n_2 = 65 varijanata
$D = |\overline{x}_1 - \overline{x}_2| = 3.3$ $\quad s_D = 0.609$

nultu hipotezu o razlici testirat ćemo tako da najprije izračunamo t_{exp}:

$$t_{exp} = \frac{D}{s_D}$$

$$t_{exp} = \frac{3.3}{0.609} = 5.42$$

To znači da je razlika (D) pet puta veća od svoje pogreške (s_D). Tablični t faktor ćemo očitati iz $(n_1 - 1) + (n_2 - 1) = 69 + 64 = 133$ i to već uz uobičajeni $P = 0.05$ i $P = 0.01$. Kako u tablicama nema 133 uzeti ćemo 100, pa je

$$t_{P=5\%} = 1.98 \qquad t_{P=1\%} = 2.63$$

Uspoređujući t_{exp} i t_{tabl} zaključujemo: kako je, dakle, $t_{exp} > t_{tabl}$, *odbacujemo* H_0, *razlika je značajna, opravdana ili signifikantna*. To tvrdimo s 99% sigurnosti ili uz vjerojatnost $P = 0.01$ (što, dakako, uključuje i 95% sigurnosti ili vjerojatnosti $P = 0.05$).

U primjeru 5.2.1.2. radi se o razlici $D = 5.5$, $s_D = 3.082$, $n_1 = 18$ i $n_2 = 20$, t_{exp} će biti:

$$t_{exp} = \frac{5.5}{3.082} = 1.78\ ,$$

a t_{tabl} treba očitati iz $(n_1 - 1) + (n_2 - 1) = 17 + 19 = 36$

U tablici, međutim, opet nema vrijednosti 36 pa možemo postupiti tako da: ili očitamo t_{tabl} za prvu bližu vrijednost broja slobodnih varijanata (obično strožu), ili interpoliranjem izračunati željenu točnu vrijednost. Odlučiti ćemo se, kao i u prethodnom primjeru, za prvu mogućnost pa t_{tabl} očitati za 30 slobodnih varijanata. (U slučaju da se t_{exp} vrlo malo razlikuje od t_{tabl} očitanog za prvu bližu vrijednost slobodnih varijanata, dobro je interpolirati. U našem primjeru to nije potrebno, jer je $t_{exp} = 5.42$ puno veći, a $t_{exp} = 1.78$ je puno manji i od ovih strožih vrijednosti).

$$t_{P=5\%} = 2.04 \qquad t_{P=1\%} = 2.75$$

Kako je $t_{exp} < t_{tabl}$ *prihvaćamo* H_0 *i zaključujemo da razlika nije značajna, nije opravdana, nije signifikantna*. Ona je, dakle, zanemariva.

ODNOS D i LSD

Međutim, t test se može provesti i tako, da se razlika (D) usporedi s minimalnom ili graničnom razlikom.

Naime, kako je

$$t = \frac{D}{s_D}$$

tada je

$$D = t \cdot s_D$$

Ako se za vrijednost t uzme ona granična, tj. t_{tabl}, tada ovaj izraz postaje granična vrijednost tj. **najmanja opravdana razlika ili najmanja signifikantna razlika**, koja se u statistici označava *LSD* (od engl. ***L****east* ***S****ignificant* ***D****ifference*).

LSD je, naime, najmanja razlika koja mora postojati između dvije srednje vrijednosti, da bi odbacili H_0 o razlici među njima, a te dvije srednje vrijednosti smatrali značajno različitima.

Dakle,

$$LSD = t \cdot s_D$$

Kako t_{tabl} možemo očitati uz dvije razine vjerojatnosti pogreške (P = 0.05 P = 0.01), to i najmanju signifikantnu razliku možemo izračunati uz manju ili veću sigurnost. To je

$$LSD_{P=5\%} = t_{P=5\%} \cdot s_D$$

$$LSD_{P=1\%} = t_{P=1\%} \cdot s_D$$

I usporedbom razlike između dvije prosječne vrijednosti (D) i najmanje značajne razlike (LSD) provodimo t test, odnosno testiramo razliku (D) i to:

ako je $D < LSD \rightarrow$ prihvaćamo H_0, to znači da razlika D nije značajna, nije opravdana, nije signifikantna; ona je slučajna, dakle, zanemariva

ako je $D > LSD \rightarrow$ odbacujemo H_0, to znači da je razlika D značajna, opravdana ili signifikantna

Konačno, ako i ovaj način testiranja razlike između dvije srednje vrijednosti primijenimo na prethodne primjere, tada će za primjer 5.2.1.1. (gdje je $D = 3.3$, $s_D = 0.609$, $t_{P=5\%} = 1.98$, $t_{P=1\%} = 2.63$) najmanja signifikantna razlika iznositi

$$LSD_{P=5\%} = 1.98 \cdot 0.609 = 1.205$$

$$LSD_{P=1\%} = 2.63 \cdot 0.609 = 1.602$$

Kako je D = 3.3 veća od $LSD_{P=1\%}$ odbacujemo H_0, zaključujemo da razlika nije zanemariva nego značajna, opravdana ili signifikantna, a sve to uz P = 1%, što znači 99% sigurnosti (koja, dakako, uključuje P = 5% odnosno 95% sigurnosti). Kažemo razlika je signifikantna uz P = 1%, a to se obično označava kao D^{**} (kad bi D bila veća samo od $LSD_{P=5\%}$ označili bismo to s jednom zvjezdicom tj. D^{*}). Za ovaj primjer, dakle,

$$D = 3.3^{**}$$

Za primjer 5.2.1.2 (gdje je D=5.5, s_D = 3.082, $t_{P=5\%}$ = 2.04, $t_{P=1\%}$ = 2.75) najmanja značajna razlika je

$$LSD_{P=5\%} = 2.04 \cdot 3.082 = 6.287$$

$$LSD_{P=1\%} = 2.75 \cdot 3.082 = 8.475$$

S obzirom da je D = 5.5 manja od LSD, to prihvaćamo H_0 o njoj, razlika je slučajna, dakle nije značajna, nije opravdana, nije signifikantna. Često se to označi i sa ns, pa je

$$D = 5.5 \ \text{ns}$$

■

Rezimirajmo: pretpostavka o razlici između bilo koje dvije prosječne vrijednosti može i ne mora biti istinita. Stoga je postupak slijedeći:

H_0 o razlici između $\bar{x}_1 - \bar{x}_2$

$H_0 : \bar{x}_1 = \bar{x}_2$ ili $H_0 : D = 0$

↓

Provjeriti istinitosti H_0 ili testirati H_0

↓

t-test

grafički (iz odnosa D i s_D)

usporedbom t_{exp} i t_{tabl}

usporedbom D i LSD

Prihvaćanje H_0 je dokaz da je H_0 bila istinita Razlika nije opravdana, nije značajna, nije signifikantna	Odbacivanje H_0 je dokaz da H_0 nije bila istinita Razlika je opravdana, značajna, signifikantna

uz vjerojatnost pogreške P=0.05 i P=0.01

Na dva primjera pokazali smo testiranje H_0 o razlici između srednjih vrijednosti na sva tri načina.

Dakako da smo uvijek dobili iste zaključke, što mora i biti jer je to isti test izveden na različite načine. Svaki od njih je ispravan, no grafički je pokazan samo radi boljeg razumijevanja, a ostala dva koriste se ovisno o praktičnosti (ponekad je pogodnije usporediti t_{exp} i t_{tabl}, a nekad D i LSD).

Često, međutim, broj varijanata jednog i drugog uzorka čije se prosječne vrijednosti uspoređuju nije isti, a dakako i njihove varijance mogu biti iste ili različite. Tada se za računanje s_D koriste formule:

a) ako je $n_1 \neq n_2$ $\quad s_1^2 \neq s_2^2$

$$s_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

b) ako je $n_1 = n_2 = n$ $\quad s_1^2 \neq s_2^2$

$$s_D = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}} = \sqrt{\frac{s_1^2 + s_2^2}{n}}$$

c) ako je $n_1 \neq n_2$ $\quad s_1^2 = s_2^2 = s^2$

$$s_D = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} = \sqrt{s^2 \frac{n_1 + n_2}{n_1 \cdot n_2}}$$

d) ako je $n_1 = n_2 = n$ $\quad s_1^2 = s_2^2 = s^2$

$$s_D = \sqrt{\frac{s^2}{n} + \frac{s^2}{n}} = \sqrt{\frac{2s^2}{n}}$$

PRIMJER 5.2.1.3. $\quad n_1 = n_2 = n \quad s_1^2 \neq s_2^2$

U proizvodnji gladiola najviše su cijenjeni cvatovi tzv. prve klase. Podatke o broju takvih cvatova (na istoj jedinici površine) kod dvije sorte, Oscar i Peter Pears, analizirat ćemo u smislu testiranja nulte hipoteze o razlici između prosječnog broja prvoklasnih cvatova ovih sorata. Pretpostavka je da se ove dvije sorte ne razlikuju u tom svojstvu (da imaju zanemarivo malo različit broj cvatova prve klase), dakle

$$H_0 : \overline{x}_0 = \overline{x}_{PP} \text{ ili } H_0 : D = 0$$

Broj parcele	Oscar	Peter Pears
1	34	39
2	29	32
3	31	35
4	36	35
5	32	35
6	36	33
7	32	25
8	29	30
9	29	30
10	30	36
11	25	39
12	32	33
	$\sum x_O = 375$	$\sum x_{PP} = 402$
	$\overline{x}_O = \frac{375}{12} = 31.25$	$\overline{x}_{PP} = \frac{402}{12} = 33.50$

$$D = \overline{x}_{PP} - \overline{x}_O = 33.50 - 31.25$$

$$D = 2.25 \text{ cvatova}$$

Kako su ova dva uzorka iste veličine (t.j. $n_O = n_{PP} = n$) za računanje s_D koristimo formulu

$$s_D = \sqrt{\frac{s_O^2 + s_{PP}^2}{n}}$$

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}$$

$$s_O^2 = \frac{34^2 + 29^2 + ... + 32^2 - \frac{375^2}{12}}{12-1} = \frac{110.25}{11} = 10.02$$

$$s_{PP}^2 = \frac{39^2 + 32^2 + ... + 33^2 - \frac{402^2}{12}}{12-1} = \frac{173.00}{11} = 15.72$$

$$s_D = \sqrt{\frac{10.02 + 15.720}{12}} = \sqrt{\frac{25.74}{12}} = 1.46$$

$$t_{exp} = \frac{D}{s_D} = \frac{2.25}{1,46} = 1.54$$

Tablični se t faktor očita iz $(n_O - 1) + (n_{PP} - 1) = 22$

$t_{P=5\%} = 2.07$

$t_{P=1\%} = 2.82$

$t_{exp} < t_{tabl} \rightarrow$ prihvaća se H_0

$D = 2.25$ ns

Zaključujemo, da se sorte Oscar i Peter Pears ne razlikuju značajno u prosječnom broju prvoklasnih cvatova. Razlika nije opravdana, nije značajna, nije signifikantna.

PRIMJER 5.2.1.4. $n_1 = n_2 = n \qquad s_1^2 \neq s_2^2$

U usjevu duhana bila je primijećena zaraženost jednom nematodom. Da se vidi utječe li ona na rast biljaka, izmjerena je visina biljke u uzorcima nezaraženih i uzorcima zaraženih biljaka. Iz tog mnoštva podataka o visini biljaka, uzet ćemo 10 iz uzorka zaraženih i 8 iz uzorka nezaraženih biljaka i provesti t-test za testiranje hipoteze o razlici u visini biljaka. Naša je, dakle, pretpostavka (nulta hipoteza) da je razlika u visini zaraženih u odnosu na nezaražene biljke zanemariva, odnosno

$$H_0 : \overline{x}_Z = \overline{x}_{NZ} \text{ ili } H_0 : D = 0$$

Visina zaraženih biljaka (cm)	Visina nezaraženih biljaka (cm)
51	99
52	66
49	88
62	91
71	64
57	85
78	99
65	93
68	
59	
$\sum x_Z = 612$	$\sum x_{NZ} = 685$
$n_Z = 10$	$n_{NZ} = 8$
$\bar{x}_Z = 61.20$	$\bar{x}_{NZ} = 85.62$

$D = 24.42$ cm

Zbog toga što je $n_Z \neq n_{NZ}$ za računanje s_D koristimo formulu

$$s_D = \sqrt{\frac{s_Z^2}{n_Z} + \frac{s_{NZ}^2}{n_{NZ}}}$$

$$s_Z^2 = \frac{51^2 + 52^2 + \ldots + 59^2 - \dfrac{612^2}{10}}{10-1} = \frac{799.6}{9} = 88.84$$

$$s_{NZ}^2 = \frac{99^2 + 66^2 + \ldots + 93^2 - \dfrac{685^2}{8}}{8-1} = \frac{1299.9}{7} = 185.70$$

$$s_D = \sqrt{\frac{88.84}{10} + \frac{185.70}{8}} = \sqrt{32.096} = 5.66 \text{ cm}$$

$$t_{exp} = \frac{24.42}{5.66} = 4.31$$

Tablični t faktor će se očitati iz $(n_Z - 1) + (n_{NZ} - 1) = 16$

$t_{P=5\%} = 2.12$

$t_{P=1\%} = 2.92$

$t_{exp} > t_{tabl} \rightarrow$ odbacuje se H_0

$D = 24.42$ cm **

To znači da su nematodom zaražene biljke bile značajno niže.

Ili: visina nezaraženih biljaka bila je signifikantno veća od onih zaraženih.

*Ili: razlika u visini nezaraženih i zaraženih biljaka je signifikantna (uz P=0.01, što označavamo sa **).*

5.2.2. Zavisni uzorci

U svim primjerima do sada razmatrali smo načine testiranja razlika između prosječnih vrijednosti dva potpuno nezavisna uzorka. Međutim, često puta imamo uzorke koji su na neki način vezani (biološki, prostorno, vremenski i sl.), pa se zato i t-test provodi na specifičan način. Tu se opažanja pojavljuju u parovima, na primjer: vrijednost nekog svojstva prije i poslije tretmana, broj listova iznad i ispod klipa kukuruza, prinos istih sorata na 2 različita lokaliteta i sl. ***Zato se mogu utvrditi razlike unutar svakog para i analizirati uzorak razlika, a zatim testiranjem prosječne razlike utvrditi da li se uzorci značajno ili zanemarivo malo razlikuju.***

PRIMJER 5.2.2.1.

Da se utvrdi učinkovitost dva herbicida (H_1 i H_2) u suzbijanju korova na šećernoj repi, na površini zasijanoj jednom sortom šećerne repe na 9 parcela su primijenjena oba herbicida i to tako da se na polovicu svake parcele aplicirao jedan, a na drugu polovicu drugi herbicid. Učinkovitost ovih pripravaka pratilo se preko količine (mase) preostalih korova na svakoj polovici svake parcelice. Ovo je očito slučaj vezanih uzoraka (ista sorta, polovica parcele tretirana jednim, a polovica drugim herbicidom).

Nulta hipoteza je pretpostavka da je razlika u učinkovitosti ovih herbicida zanemariva - nije značajna.

Kako se radi o zavisnim, vezanim uzorcima, analizirat će se uzorak razlika među parovima i testirati nulta hipoteza o prosječnoj razlici:

$$H_0 : \bar{d} = 0$$

Parcelica broj	Masa preostalih korova (u kg/m)[2] nakon primjene H₁	H₂	$x_1 - x_2 = d_i$	$(d_i - \bar{d})$	$(d_i - \bar{d})^2$
1	1.62	1.50	0.12	0.07	0.0049
2	1.60	1.42	0.18	0.13	0.0169
3	1.64	1.55	0.09	0.04	0.0016
4	1.45	1.49	−0.04	0.01	0.0001
5	1.52	1.70	−0.18	−0.13	0.0169
6	1.65	1.40	0.25	0.20	0.0400
7	1.48	1.60	−0.12	−0.07	0.0049
8	1.50	1.45	0.05	0.00	0.0000
9	1.65	1.55	0.10	0.50	0.2500
	$\bar{x}_{H_1} = 1.568$	$\bar{x}_{H_2} = 1.518$ $D = 0.05$	$\sum d_i = 0.45$ $\bar{d} = \frac{0.45}{9} = 0.05$		$\sum (d_i - \bar{d})^2 = 0.3353$

Naime, na svakoj parcelici čija je jedna polovica tretirana s H_1, a druga s H_2, najprije se utvrdi razlika ($d = x_1 - x_2$) u masi korova unutar svakog para (između polovica parcelice). Iz svih tih razlika izračuna se prosječna razlika $\bar{d}$ (vodeći računa o predznacima!), a potom odstupanja svake razlike od prosječne razlike $(d_i - \bar{d})$, te suma njihovih kvadratnih vrijednosti, tj. $\sum (d_i - \bar{d})^2$. Iz toga je moguće provesti t-test za testiranje H_0 o prosječnoj razlici $\bar{d} = 0.05$ (koja je ista kao i razlika $D = \bar{x}_1 - \bar{x}_2$).

$$H_0 : \bar{d} = 0$$

$$t_{exp} = \frac{\bar{d}}{s_{\bar{d}}}$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

$$s_{\overline{d}} = \frac{\sqrt{\frac{\sum(d_i - \overline{d})^2}{n-1}}}{\sqrt{n}} = \sqrt{\frac{\sum(d_i - \overline{d})^2}{n(n-1)}} = \sqrt{\frac{0.3353}{9 \cdot 8}} = 0.068 \text{ kg/m}^2$$

$$t_{exp} = \frac{0.05}{0.068} = 0.735$$

Kod zavisnih uzoraka tablični t faktor očita se iz n–1 parova. U našem primjeru dakle, iz $n-1 = 9 - 1 = 8$, $t_{P=0.05} = 2.31$ i $t_{P=0.01} = 3.36$.

Kako je $t_{exp} < t_{tabl}$ → prihvaćamo H_0, te zaključujemo:

prosječna razlika nije opravdana, nije značajna ili nije signifikantna:

$$\overline{d} = 0.05 \text{ kg/m}^2 \text{ ns}$$

Praktično, to znači da su ovi herbicidi bili podjednako učinkoviti u suzbijanju korova kod šećerne repe.

Naglasimo ponovno: važno je uočiti da se **kod zavisnih uzoraka testira nulta hipoteza o prosječnoj razlici među parovima ($\overline{d}$), analizirajući uzorak razlika (d_i) i njihovog odstupanja od prosječne razlike ($\overline{d}$).**

Jednako tako treba upamtiti da se **tablični t faktor očita iz n–1 parova**.

■

1
2
3
4
5

6 Analiza varijabilnosti

7

6.1. F-distribucija i F-test
6.2. Analiza varijance (Anova)

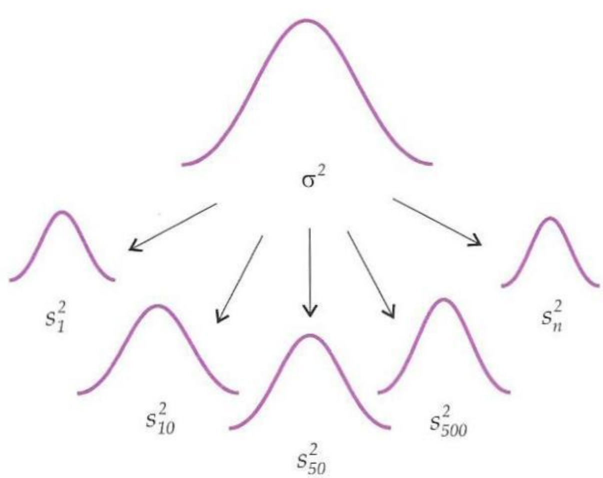
6.1. F-distribucija i F-test

Vratimo se opet na populaciju, iz koje je moguće napraviti bezbroj uzoraka, a na temelju svakog od njih procijeniti parametre populacije (μ, σ, σ^2).

O procjeni prosječne vrijednosti populacije (μ) preko $\overline{x}$ uzorka, već je bilo riječi. Zaustavimo se na varijanci (slika 18).

Svaki od brojnih, iz populacije uzetih uzoraka, ima svoju varijancu ($s_1^2, \ldots, s_n^2$). Varijanca svakog uzorka procjenjuje dakako istu vrijednost - varijancu populacije iz koje su uzorci uzeti (σ^2).

Ovisno o varijabilnosti populacije, veličini i reprezentativnosti uzoraka, njihove će varijance biti više ili manje dobre procjene varijance populacije iz koje su uzeti (σ^2).



SLIKA 18.

Varijance uzoraka ($s_1^2, \ldots, s_n^2$) uzetih iz populacije s varijancom σ^2

Varijanca svakog onog uzorka koja je vrlo dobra procjena varijance populacije, razlikovat će se od nje vrlo, vrlo malo. Zato, ako bi se načinili omjeri između varijanci tih uzoraka i varijance populacije iz koje su oni uzeti, ti omjeri bi bili oko 1.0 , jer ako je

$$s_1^2 = \sigma^2, \ldots, s_n^2 = \sigma^2$$

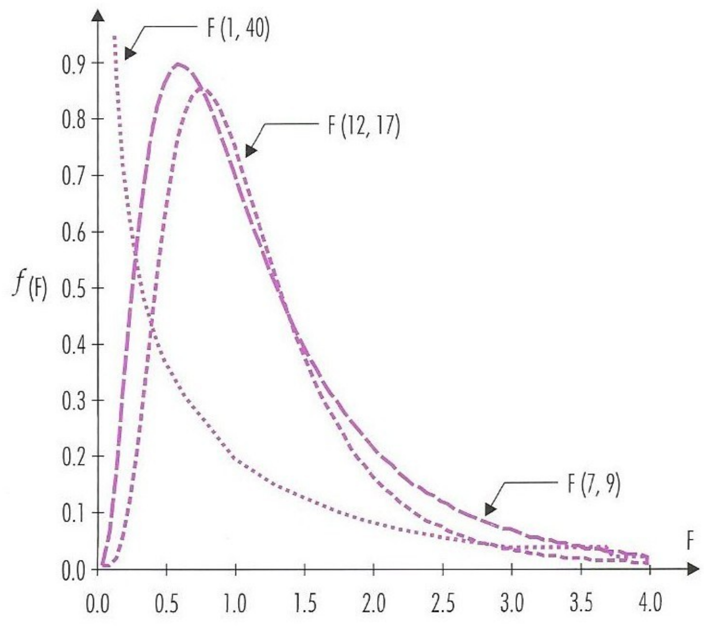
tada je

$$\frac{s_1^2}{\sigma^2} = 1.0, \ldots, \frac{s_n^2}{\sigma^2} = 1.0$$

Isto tako, ako se stave u odnos bilo koje dvije varijance uzoraka, moguće je dobiti bezbroj ovakvih omjera, koji će doduše ovisiti o veličini varijanci, ali prosjek svih tih omjera će biti (uz dovoljno veliki n) vrlo blizu 1.0. **Omjer između dvije varijance naziva se F-faktorom**.

$$F = \frac{s_1^2}{s_2^2}$$

Statističari su izračunali očekivanu distribuciju vrijednosti takvih omjera i u čast sir Ronalda Fishera nazvali je **F distribucijom**. **F distribucija** je teoretska distribucija vjerojatnosti, krivulja je asimetrična i različito izgleda, ovisno o n_1 i n_2 (slika 19).

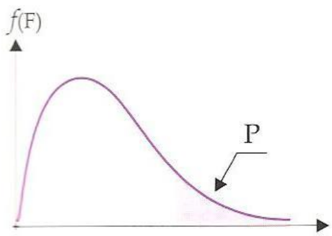


SLIKA 19.
Različite F distribucije

Ovu je distribuciju moguće opisati kompliciranom matematičkom funkcijom, čime se ovdje nećemo baviti. Za razliku od t-distribucije, **oblik F-distribucije je određen s dvije vrijednosti, to jest sa slobodnim varijantama koje pripadaju dvjema varijancama koje su u omjeru ili** $(n_1 - 1)$ i $(n_2 - 1)$.

Stoga za svaki mogući par vrijednosti postoji druga F distribuicija.

Zbog tog su razrađenene **F-tablice** u kojima je, za svaki par vrijednosti slobodnih varijanata dvije pripadajuće varijance, moguće očitati F faktor uz određenu vjerojatnost pogreške (slika 20).



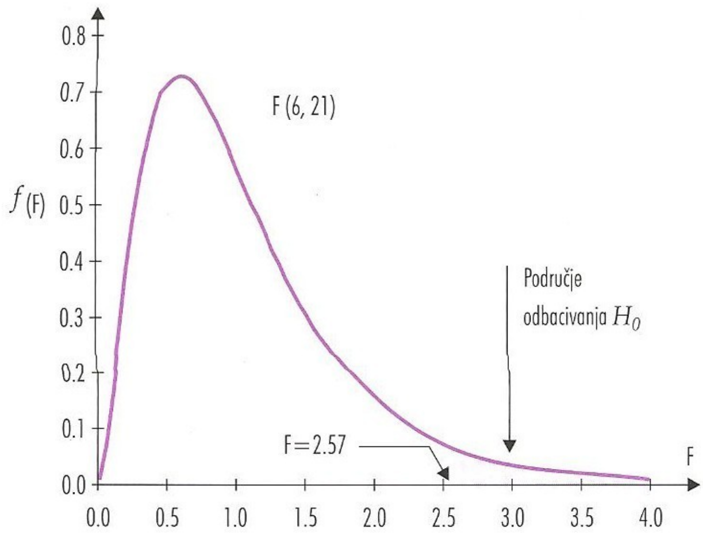
SLIKA 20.
Dio tablice C u dodatku

n-1 nazivnika	*n*-1 brojnika						
	30	40	50	60	75	100	∞
1	250.00	251.00	252.00	252.00	253.00	253.00	254.00
2	19.46	19.47	19.47	19.48	19.48	19.49	19.50
3	8.62	8.59	8.58	8.57	8.57	8.56	8.53
4	5.74	5.72	5.70	5.69	5.68	5.66	5.63
5	4.50	4.46	4.44	4.43	4.42	4.40	4.36
6	3.81	3.77	3.75	3.74	3.72	3.71	3.67
7	3.38	3.34	3.32	3.30	3.29	3.28	3.23
8	3.08	3.04	3.03	3.00	3.00	2.98	2.93
9	2.86	2.82	2.80	2.79	2.77	2.76	2.71
10	2.70	2.66	2.64	2.62	2.61	2.59	2.54
-							
-							
-							
26	1.90	1.85	1.82	1.80	1.78	1.76	1.69
27	1.88	1.84	1.80	1.78	1.76	1.74	1.67
28	1.87	1.82	1.78	1.77	1.75	1.72	1.65
29	1.85	1.80	1.77	1.75	1.73	1.71	1.64
30	1.84	1.79	1.76	1.74	1.72	1.69	1.62
40	1.74	1.69	1.66	1.64	1.61	1.59	1.51
50	1.69	1.63	1.60	1.58	1.55	1.52	1.44

I ovdje govorimo o vrijednosti koju je moguće izračunati, a to je **eksperimentalni F faktor** (F_{exp}) i faktoru očitanom iz tablica ovisno o n_1 i n_2, a to je **tablični F faktor** (F_{tabl}).

Eksperimentalni F faktor je broj koji pokazuje koliko je puta jedna varijanca veća od druge.

Tablični F faktor je kritična vrijednost ili broj koji pokazuje koliko najmanje puta mora jedna varijanca biti veće od druge da bi odbacili nultu hipotezu o njima.



SLIKA 21.
F-distribucija
(n_1-1=6 i n_2-1=21)

Nulta hipoteza o dvije varijance (H_0) je pretpostavka da su dvije varijance iste, da pripadaju istoj distribuciji, ili

$$H_0 : s_1^2 = s_2^2$$

Postupak testiranja nulte hipoteze o dvije varijance zove se **F-test**.

Usporedbom eksperimentalnog i tabličnog F faktora moguće je, dakle, testirati nultu hipotezu o varijancama.

Ako je

$F_{exp} < F_{tabl} \rightarrow$ prihvaća se H_0 (nema dovoljno argumenata da bi se odbacila pretpostavka o jednakosti varijanci)

a ako je

$F_{exp} > F_{tabl} \rightarrow$ odbacuje se H_0 (podaci pružaju dovoljno sigurnosti za tvrdnju da se varijance razlikuju)

U slučaju prihvaćanja H_0 govorimo o nesignifikantnom, a u slučaju odbacivanja H_0 o signifikantnom F testu.

■

Ako su varijance, na primjer: s_1^2 = 3.6, s_2^2 = 2,8 , a dobivene su iz uzoraka veličine n_1 = 30, n_2 = 50 tada je

$$F_{exp} = \frac{3.6}{2.8} = 1.286 \text{ ns}$$

F_{tabl} se očita iz n_1-1 = 29 i n_2-1 = 49, no kako u tablicama nema ni 29 niti 49, očitati ćemo za prve najbliže (strože) vrijednosti, tj. 30 i 50:

$$F_{P=0,05} = 1.69$$
$$F_{P=0,01} = 2.10$$

Kako je $F_{exp} < F_{tabl}$ prihvaćamo H_0 i zaključujemo da se ove dvije varijance ne razlikuju, odnosno pripadaju istoj distribuciji.

Ili, drugi slučaj, gdje je s_1^2 = 43.6, s_2^2 = 5.2, n_1 = 3, n_2 = 10

$$F_{exp} = 8.38 \ ^{**}$$

Iz n_1 - 1 = 2 i n_2 - 1 = 9 očita se

$$F_{P=0,05} = 4.26$$
$$F_{P=0,01} = 8.02$$

S obzirom da je $F_{exp} > F_{tabl}$ *odbacujemo H_0 o ove dvije varijance i zaključujemo da pripadaju distribucijama različitih varijabilnosti.*

6.2. Analiza varijance (ANOVA)

Ovim do sada opisanim postupcima moguće je, dakle, utvrditi pripadaju li dva uzorka (a onda se to naravno odnosi i na njihove srednje vrijednosti) istoj populaciji, ili se toliko razlikuju da možemo zaključiti da su dobiveni iz različitih populacija.

Kad su u pitanju samo dva uzorka, odgovor je moguć na temelju **t-testa**. No, što ako imamo više prosječnih vrijednosti dobivenih iz više grupa podataka? Kako postuputi ako treba usporediti prosječni broj prvoklasnih cvatova 4 sorte gladiola (a ne dvije kako smo imali u primjeru 5.2.1.3.) i utvrditi jesu li razlike značajne ili ne? To bismo, doduše, mogli postići primjenom t-testa uspoređujući dvije i dvije prosječne vrijednosti. Međutim, kad se radi o više prosječnih vrijednosti (uzoraka ili grupa u uzorku) uspoređivanje svake srednje vrijednosti sa svakom, nije dozvoljeno. Naime, kad bi se i uspoređivalo tako bilo bi to dugotrajno, a što je najvažnije, izgubilo bi se na pouzdanosti.

t-test vrijedi samo za slučajne usporedbe, a usporedba recimo, najveće i najmanje srednje vrijednosti, nije slučajna. Usporedbom dvije i dvije prosječne vrijednosti gubi se i na preciznosti računanja varijance koja je uvjetovana varijabilnošću svih grupa, a ne samo onih dviju koje uspoređujemo.

U tu svrhu je Ronald A. Fisher razradio postupak, poznat kao **analiza varijance**, ili **Fisherova analiza varijance** za koju se udomaćila skraćenica **ANOVA**. Moglo bi se reći da je ona više nego tehnika za statističku analizu. Poznavanje i razumijevanje analize *varijance omogućava uvid u način variranja prirodnih pojava, u* Prirodu samu, što je vjerojatno još od veće važnosti nego metoda kao takva.

Upravo zbog toga, ako se može govoriti o ljepoti statističkih metoda, onda analiza varijance sigurno posjeduje više od bilo koje druge.

Tim postupkom se ukupna varijabilnost rastavlja na varijabilnost između prosječnih vrijednosti grupa i na varijabilnost unutar grupa.

Ova metoda odgovara na pitanje je li signifikantno više varijabilnosti između srednjih vrijednosti grupa ili između varijanata unutar grupa.

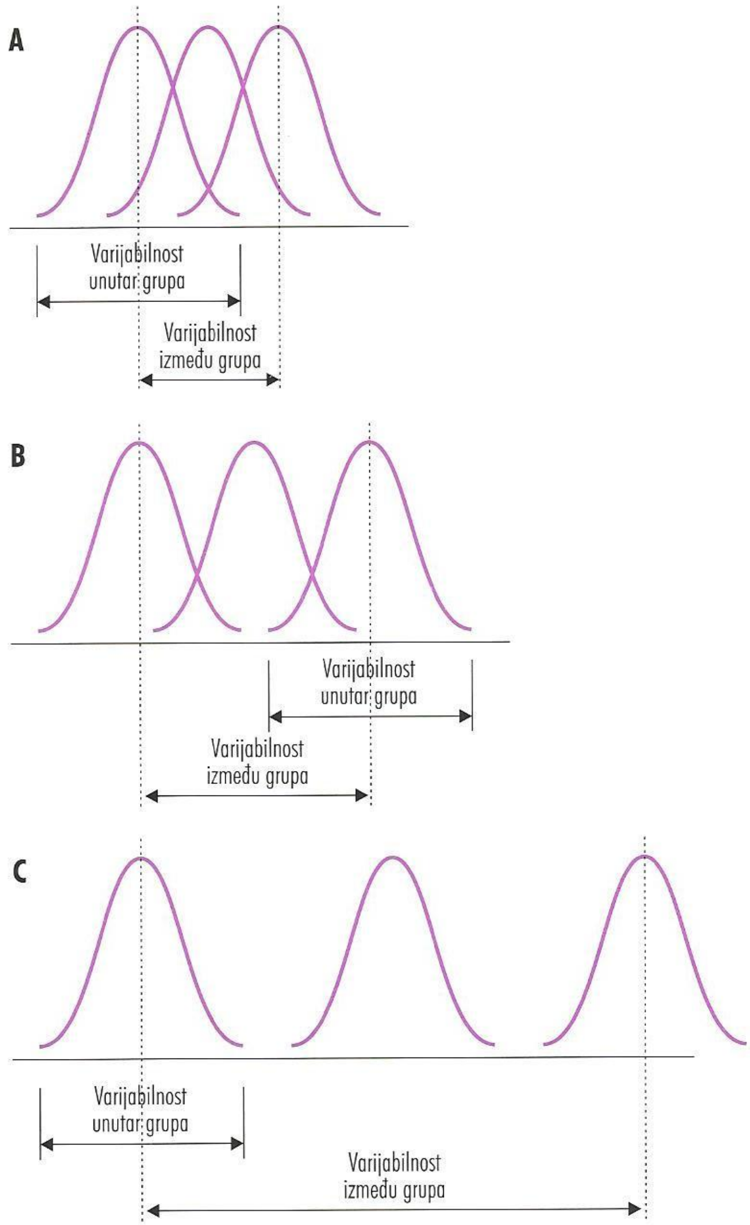
Iako se metoda zove analiza varijance, ona je zapravo postupak za proučavanje varijabilnosti između prosječnih vrijednosti, koju se mjeri varijancom. U tu svrhu se u postupku analize varijance primjenjuje test za testiranje varijanci i test za testiranje razlika između prosječnih vrijednosti (F-test i t-test).

Pretpostavke koje moraju biti zadovoljene za analizu varijance su: da su uzorci iz populacije uzeti slučajno, da su pogreške neovisne i normalno distribuirane, i da imaju iste (homogene) varijance.

Kako smo već opetovano naglašavali, populacije uglavnom proučavamo preko uzoraka. Pa, pokušajmo na uzorcima razumjeti i naučiti analizu varijance. Uzmimo da se uzorak sastoji od više grupa. Kako saznati (i, dakako, uz dovoljnu pouzdanost) da te grupe pripadaju istom uzorku?

Ako na najjednostavniji način prikažemo uzorak koji se sastoji od više grupa, a različitost između varijanata unutar grupa i različitost između srednjih vrijednosti grupa označimo "strelicama", tada su mogući (opet, razmatrajući na najjednostavniji, slikovit način) slučajevi kako prikazuje slika 22.

Uspoređujući duljine strelica kojima je označen raspon između najmanje i najveće prosječne vrijednosti grupa (između grupa) i najmanje i najveće vrijednosti u svakoj grupi (unutar grupa), mogli bismo reći da je u slučaju A varijabilnost između grupa manja od



SLIKA 22.
Varijabilnost između grupa u odnosu na varijabilnost unutar grupa

varijabilnosti unutar grupa, a obratno za B gdje je varijabilnost između grupa veća od varijabilnosti unutar grupa.

Ako je varijabilnost između grupa signifikantno veća od varijabilnosti unutar grupa, tada grupe ne pripadaju istoj distribuciji. To, međutim ne mora značiti da se sve varijante jedne grupe razlikuju od svih varijanata druge grupe (ima preklapanja kao u B).

Ali, slučaj C ilustrira situaciju gdje se baš sve varijante jedne razlikuju od svih varijanata druge grupe, sve u drugoj od svih varijanata u trećoj itd.

Tehnika analize varijance omogućava nam da to računski izrazimo i utvrdimo odnose između tih varijabilnosti. I ne samo to. Analizom varijance, statističkim postupkom koji objedinjava sve ono o čemu je do sada bilo riječi, moguće je doznati da li grupe pripadaju istom uzorku ili možda samo neke od njih, a neke ne, itd. Jer, *analiza varijance je metoda kojom se definira ukupna varijabilnost i raščlanjuje (analizira) na dijelove uvjetovane različitim činiocima.*

Strelice na slici kojom smo ilustrirali moguće slučajeve, dakako, treba zamijeniti pravim mjerilom varijabilnosti - varijancom.

Uzmimo jedan hipotetičan uzorak veličine n = 18, koji neka se sastoji od dvije grupe A i B.

Uzorak	
grupa A	grupa B
3	9
4	10
4	10
5	11
5	11
5	11
6	12
6	12
7	13
$\sum x_A = 45$	$\sum x_B = 99$
$n_A = 9$	$n_B = 9$
$\bar{x}_A = 5.0$	$\bar{x}_B = 11.0$
$\sum x = 144$	$\bar{x} = 5.0$

Provesti analizu varijance znači računski definirati ukupnu varijabilnost ovog uzorka, raščlaniti je na varijabilnost unutar grupa i varijabilnost između grupa i testirati hipotezu da grupe pripadaju istom uzorku, odnosno da se njihove prosječne vrijednosti ne razlikuju značajno.

$$H_0 : \bar{x}_A = \bar{x}_B \quad \text{ili} \quad H_0 : D = 0$$

Svaka varijanta ovog uzorka u određenom iznosu odstupa od prosječne vrijednosti grupe u kojoj se nalazi, ali isto tako za izvjestan iznos odstupa i od prosječne vrijednosti cijelog uzorka.

Isto tako i prosječna vrijednost svake grupe odstupa od prosječne vrijednosti cijelog uzorka.

Ta odstupanje je moguće izraziti sumama kvadratnih odstupanja (SS).

Ukupna varijabilnost (tj. odstupanja svake varijante ovog uzorka od prosječne vrijednosti uzorka) izražena kao SS je:

$$SS_{\text{Ukupno}} = (3 - 8)^2 + ... + (13 - 8)^2 = 186$$

Varijabilnost unutar grupa (a to je odstupanje svake varijante u grupi od srednje vrijednosti grupe) moguće je izračunati kao:

$$\begin{aligned} SS_{\text{Unutar grupa}} &= SS_{\text{Unutar grupe A}} + SS_{\text{Unutar grupe B}} \\ &= \Sigma(x-\bar{x}_A)^2 + \Sigma(x-\bar{x}_B)^2 \\ &= (3 - 5)^2 + ...+(7 - 5)^2 + (9 - 11)^2+...+(13 - 11) \\ &= 12 + 12 \\ &= 24 \end{aligned}$$

Varijabilnost između grupa moguće je izraziti odstupanjem prosjeka svake grupe od prosjeka uzorka, s time, da se svaka varijanta svake grupe predstavi svojom srednjom vrijednošću

$$SS_{\text{Između grupa}} = 9 \cdot (5 - 8)^2 + 9 \cdot (11 - 8)^2 = 162$$

SS se može izračunati (kako smo već u početku pokazali) i na drugi način:

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

Tako izračunane pojedine SS vrijednosti su:

$$SS_{\text{Ukupno}} = 3^2 + ... + 13^2 - \frac{144^2}{18} = 186$$

$$SS_{\text{Između grupa}} = \frac{45^2}{9} + ... + \frac{99^2}{9} - \frac{144^2}{18} = 162$$

$$\begin{aligned} SS_{\text{Unutar grupa}} &= (3^2 + ... + 7^2) - \frac{45^2}{9} + (9^2 + ... + 13^2) - \frac{99^2}{9} \\ &= (237 - 225) + (1101 - 1089) = 24 \end{aligned}$$

Razvidno je da su to iste vrijednosti za pojedine SS, samo je drugi način računanja. Kako je i jednostavniji, gotovo uvijek se koristi.

Još nešto treba uočiti: ukupna varijabilnost izražena kao SS_{Ukupno} iznosi 186, a to je upravo zbroj varijabilnosti između grupa i varijabilnosti unutar grupa ($SS_{Između\ grupa} + SS_{Unutar\ grupa} = 162 + 24 = 186$). To znači da u buduće možemo izračunati ovu treću sumu kvadratnih odstupanja kao razliku između $SS_{Ukupno} - SS_{Između\ grupa} = SS_{Unutar\ grupa}$ i tako još olakšati i skratiti postupak.

Ne smijemo zaboraviti da je varijanca prosječno kvadratno odstupanje $\left(s^2 = \dfrac{SS}{n-1}\right)$, pa iz ovih SS vrijednosti tek treba izračunati pripadajuće varijance. Stoga se obično postupak nastavlja konstrukcijom tzv. **tablice ANOVA**:

Tablica analize varijance (Tablica ANOVA)

Izvori varijabilnosti	n-1	SS	$s^2=\dfrac{SS}{n-1}$	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	(18 - 1) = 17	186	10.94			
Između grupa	(2 - 1) = 1	162	162.00	$\dfrac{162.00}{1.5} = 108.0^{**}$	4.49	8.53
Unutar grupa	(17 - 1) = 16	24	1.50			

Da bi se iz izračunanih SS vrijednosti došlo do varijanci, potrebno je znati broj slobodnih varijanata za svaki SS. Kako je analiziran uzorak od 18 varijanata, broj slobodnih varijanata koji pripada SS_{Ukupno} je 17.

Dvije su grupe u uzorku (A i B) pa je za $SS_{Između\ grupa}$ broj slobodnih varijanata 2 - 1 = 1.

Na $SS_{Unutar\ grupa}$ pripada suma slobodnih varijanata grupe A i grupe B tj. (9 - 1) + (9 - 1) = 16. No baš kao što se i taj SS može dobiti kao razlika između prva dva, tako se i pripadajući broj slobodnih varijanata dobije iz 17 - 1 = 16. Dakako da će se to praktično i koristiti.

Pripadajuće varijance su tada kvocijenti SS i broja slobodnih varijanata koje pripadaju svakom SS-u.

Vratimo se sad na sliku 22 koja ilustrativno daje usporedbu varijabilnosti između i unutar grupa. U prvom slučaju je strelica koja prikazuje varijabilnost između grupa manja od strelice koja znači varijabilnost unutar grupa. U drugom i trećem slučaju je obratno: strelica koja prikazuje varijabilnost između grupa, veća je od one koja prikazuje varijabilnost unutar grupa.

Sad imamo mogućnost te varijabilnosti, brojčano izražene u vidu varijanci, staviti u odnos, primijeniti postupak za testiranje nulte hipoteze o varijancama (F-test) i ustanoviti pripadaju li one istom ili različitim uzorcima.

Kvocijent **varijance između grupa** i **varijance unutar grupa** je

$$F_{exp} = \frac{s^2 \text{ između grupa}}{s^2 \text{ unutar grupa}}$$

F testom možemo testirati H_0 o varijancama i utvrditi da li je varijanca između grupa značajno veća od varijance unutar grupa ili ne.

Ako:

- F nije signifikantan (jer $F_{exp} < F_{tabl}$) zaključujemo da grupe (koliko god ih bilo) pripadaju istoj distribuciji pa su i njihove prosječne vrijednosti zanemarivo različite; prihvaćamo H_0 o te dvije varijance. Razlike između prosječnih vrijednosti grupa mogu se pripisati pogreški uzorka.
- F je signifikantan (jer $F_{exp} > F_{tabl}$) odbacujući H_0 zaključujemo da varijance (grupe) ne pripadaju istom uzorku (distribuciji). Dakle, varijabilnost između grupa je dovoljno veća od varijabilnosti unutar grupa da podupire tezu o različitosti prosjeka grupa.

To, međutim, ne mora značiti da se sve grupe međusobno razlikuju, ali to svakako znači da su bar dvije signifikantno različite.

Za naš hipotetičan primjer: F_{tabl} očitan iz broja slobodnih varijanata varijance koja je u brojniku i broja slobodnih varijanata varijance u nazivniku (znači iz 1 i 16) iznosi

$$F_{P=5\%} = 4.49$$

$$F_{P=1\%} = 8.53$$

Usporedbom eksperimentalnog i tabličnog F-faktora, budući da je $F_{exp} > F_{tabl}$, odbacujemo H_0 o varijancama i zaključujemo da one ne pripadaju istom uzorku. Dakle ove dvije grupe (A i B) u našem primjeru ne spadaju u isti uzorak. One se opravdano ili signifikantno razlikuju. S obzirom da se radi o samo dvije grupe, to posredno znači da su i njihove srednje vrijednosti signifikantno različite.

Naime, u slučaju dvije grupe F-test je ekvivalentan t-testu, jer je $F = t^2$. U ovom primjeru gdje je $D = \bar{x}_B - \bar{x}_A = 6$, varijanca grupe A jednaka je varijanci grupe B, tj.

$$s_A^2 = s_B^2 = \frac{SS}{n-1} = \frac{12}{8} = 1.5$$

$$s_D = \sqrt{\frac{1.5}{9} + \frac{1.5}{9}} = 0.57735$$

$$t = \sqrt{\frac{6}{0.57735}} = 10.39231$$

Kako je $F = t^2 = 108.00$, što smo dobili i u tablici ANOVA kao kvocijent varijance između grupa i varijance unutar grupa, razvidna je istovjetnost ova dva testa. Razlika je samo u tome što t-test može služiti samo za usporedbu dvije srednje vrijednosti, dok se F test može primijeniti za dvije ili više.

Opravdan ili signifikantan F-test varijance između grupa, označava se ovisno o razini (nivou) značajnosti s jednom (uz $P = 5\%$) ili dvije (uz $P = 1\%$) zvjezdice uz F-faktor (F* ili F**, odnosno signifikantan ili visoko signifikantan).

Ovo je **osnovni model analize varijance**, koji se samo računski mijenja ovisno o broju grupa u uzorku koji se analizira, odnosno o broju varijanata u svakoj toj grupi (koji može biti isti ili različit).

Uzmimo još jedan primjer kojim ćemo pokazati kako se provede kompletna analiza uzorka s više grupa, nejednakog broja varijanata u svakoj grupi. Postupak je sljedeći: nultom hipotezom pretpostavljamo da ove grupe pripadaju istoj distribuciji, odnosno

$$H_0 : \bar{x}_A = \bar{x}_B = \bar{x}_C$$

Uzorak		
Grupa A	Grupa B	Grupa C
75	68	59
82	72	61
87	73	58
80	79	59
83		60
85		
78		

$\sum x_A = 570$	$\sum x_B = 292$	$\sum x_C = 297$	$\sum x = 1159$
$n_A = 7$	$n_B = 4$	$n_C = 5$	$n_{\text{ukupno}} = 16$
$\bar{x}_A = 81.43$	$\bar{x}_B = 73.0$	$\bar{x}_C = 59.4$	

$$SS_{\text{ukupno}} = 75^2 + \ldots + 60^2 - \frac{1159^2}{16} = 1585.94$$

$$SS_{\text{između grupa}} = \frac{570^2}{7} + \frac{292^2}{4} + \frac{297^2}{5} - \frac{1159^2}{16} = 1417.03$$

$$SS_{\text{unutar grupa}} = 1585.94 - 1417.03 = 168.91$$

Tablica analize varijance (Tablica ANOVA)

Izvori varijabilnosti	n-1	SS	$s^2 = \frac{SS}{n-1}$	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	16 - 1 = 15	1585.94				
Između grupa	3-1=2	1417.03	708.515	54.5**	3.8	6.7
Unutar grupa	15-2=13	168.91	12.993			

F-test je, dakle, opravdan uz P=1%. To znači da se radi o grupama signifikantno različitih varijabilnosti koje ne pripadaju istom uzorku, grupama čije se prosječne vrijednosti razlikuju.

Međutim, kada se uzorak sastoji od više grupa, opravdani F ne mora značiti da se sve prosječne vrijednosti grupa međusobno razlikuju. Razlikovati se mogu samo neke, a neke pripadati istom uzorku. Ali, kako smo već naglasili, signifikantan F-faktor znak je da se bar neke razlikuju. Da se to utvrdi, poslužimo se t-testom i uspoređivanjem prosječnih vrijednosti grupa, odnosno njihove razlike s graničnom ili najmanjom signifikantnom razlikom (*LSD*). Prisjetimo se, ona se izračuna iz

$$LSD = t \cdot s_D$$

t-faktor se očita iz tablica iz sume slobodnih varijanata unutar grupa (a to je u ovom slučaju 13) i iznosi

$$t_{p=5\%} = 2.16$$

$$t_{p=1\%} = 3.01$$

s_D se izračuna iz $s_D = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Kako se analizom varijance procijeni zajednička varijanca unutar grupa, to je $s_1^2 = s_2^2 = s^2$, pa je kao što smo već pokazali (str. 73)

$$s_D = \sqrt{s^2 \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}$$

Ne zaboravimo, grupe A, B i C u našem primjeru imaju različiti broj varijanata ($n_A = 7$, $n_B = 4$, $n_C = 5$) pa će se za svaku usporedbu prosječnih vrijednosti (tj. testiranje nulte hipoteze o njihovoj razlici) morati izračunati posebna s_D (odnosno *LSD*).

Naime, usporedit ćemo $\bar{x}_A$ sa $\bar{x}_B$, $\bar{x}_A$ sa $\bar{x}_C$ i $\bar{x}_B$ sa $\bar{x}_C$.

Za usporedbu $\bar{x}_A$ i $\bar{x}_B$ koje se razlikuju za $D = 8.43$ (tj. $\bar{x}_A - \bar{x}_B = 81.43 - 73.0 = 8.43$) standardna pogreška će biti

$$s_D = \sqrt{12.993 \cdot \frac{7+4}{7 \cdot 4}} = 2.26$$

Iz tabličnog *t*-faktora ($t_{P=5\%}$=2.16, $t_{P=1\%}$=3.01) izračunat ćemo najmanju signifikantnu razliku (*LSD*)

$$LSD_{P=5\%} = 2.16 \cdot 2.26 = 4.88$$

$$LSD_{P=1\%} = 3.01 \cdot 2.26 = 6.80$$

Usporedimo li razliku *D* s *LSD* *zaključujemo kako je* $D > LSD_{P=1\%}$ *(što uključuje i* $LSD_{P=5\%}$*) odbacujemo nultu hipotezu o njoj; razlika,* dakle, *nije* *slučajna, nije zanemariva, ona je opravdana, značajna ili signifikantna i to uz 99% sigurnosti (ili samo je 1% vjerojatnosti pogreške da bi mogla biti slučajna).*

$D = 8.43^{**}$

Za usporedbu $\bar{x}_A$ i $\bar{x}_C$ gdje je $D = 22.03$ (t.j. 81.43 - 59.4)

$$s_D = \sqrt{12.993 \cdot \frac{7+5}{7 \cdot 5}} = 2.11$$

$$LSD_{P=5\%} = 2.16 \cdot 2.11 = 4.56$$

$$LSD_{P=1\%} = 3.01 \cdot 2.11 = 6.35$$

Kako je D > LSD, to odbacujemo nultu hipotezu i zaključujemo da je razlika opravdana ili signifikantna. Kako je to uz P=1% možemo to naznačiti kao

$D = 22.03^{**}$

Za usporedbu $\bar{x}_B$ sa $\bar{x}_C$ čija razlika D=13.6 (t.j.73.0 - 59.4)

$$s_D = \sqrt{12.993 \cdot \frac{4+5}{4 \cdot 5}} = 2.42$$

$$LSD_{P=5\%} = 2.16 \cdot 2.42 = 5.23$$

$$LSD_{P=1\%} = 3.01 \cdot 2.42 = 7.28$$

*Razlika D je također signifikantna i to uz P=1% ili D = 13.6***

■

S ovih nekoliko primjera kroz koje smo željeli objasniti analizu varijance, ujedno smo objedinili sva dosadašnja znanja o varijabilnosti, parametrima kojima je moguće izraziti varijabilnost, nultoj hipotezi o razlici između prosječnih vrijednosti i nultoj hipotezi o varijancama, te primjeni t i F testova za testiranje tih hipoteza.

To su ujedno i one najosnovnije statističke metode koje ćemo koristiti u analiziranju eksperimentalnih podataka, u problemima koji će se obrađivati u narednim poglavljima. Metode su poznate kao **parametrijske statističke metode** i njima se analiziraju podaci mjernih varijabli (kojih je i najviše).

Za sve one varijable koje se ne mogu mjeriti nego se izražavaju u relativnim vrijednostima (rangovima i sl.) primjenjuju se **neparametrijske metode** kojima se uspoređuju distribucije a ne parametri.

■

7 Regresija i korelacija

7.1. Regresija
7.2. Korelacija

Do sada je bilo govora samo o jednom svojstvu ili varijabli, načinima na koje se ona može opisati (procjenom mjerila varijabilnosti i sredine i sl.), testirati razlike itd.

Međutim, svojstva i pojave su gotovo uvijek u međusobnoj vezi i zavisnosti, a često na jednom individuumu mjerimo i proučavamo dva ili više svojstava, želeći otkriti povezanost među njima i izraziti prirodu odnosa među njima.

I u svakodnevnom životu kad koristimo uzrečice:

"Kako siješ, tako ćeš i žeti"

"Milo za drago"

"Kako došlo, tako prošlo"

"Tko visoko leti, nisko pada"

u svakoj spominjemo dva svojstva ili varijable, koje su na neki način povezane i zavisne jedna o drugoj. Promjena jedne, uvjetuje da se i druga promijeni na određeni način.

To nas dovodi do pojmova **regresije** i **korelacije**.

Regresijom opisujemo odnos između dvije varijable izražavajući jednu u vidu linearne (ili kompleksnije) funkcije druge. Regresija govori o promjeni jedne varijable povezanoj s jediničnom promjenom druge.

Korelacijom se procjenjuje stupanj do kojeg dvije varijable zajednički variraju. Korelacija se odnosi na jačinu i smjer povezanosti dviju varijabli koje su u tom odnosu.

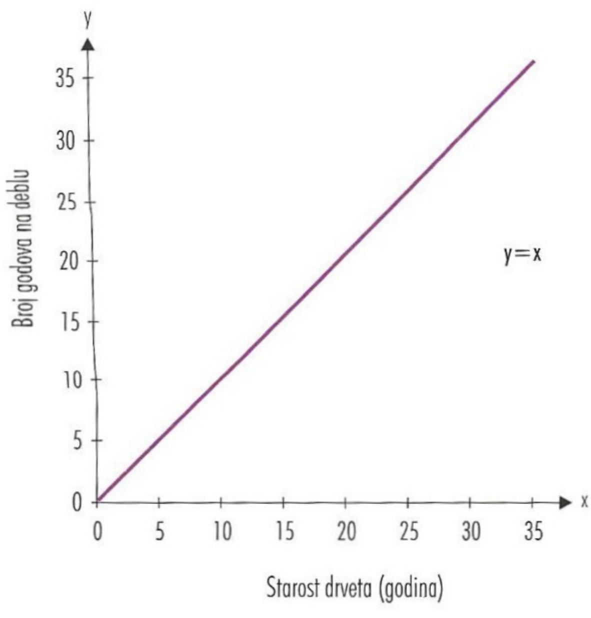
7.1. Regresija

Regresija, dakle, predstavlja funkcionalni odnos dvije varijable, te opisuje način na koji su one povezane. Prisjetimo se samo osnovnog znanja o funkcijama. Funkcija je matematički odnos, koji omogućuje "predskazati" koje vrijednosti jedne varijable korespondiraju zadanim vrijednostima druge varijable. Takav funkcionalni odnos može se najjednostavnije napisati kao

$$y = f(x)$$

Uobičajeno je varijablu čija promjena zavisi o drugoj zvati **zavisnom varijablom**, a onu koja je uzrok promjeni ove prve **nezavisnom varijablom**.

Najjednostavniji tip regresije slijedi jednadžbu $y = x$, a ilustriran je na slici 23. Pokazuje odnos između broja godova (y) i starosti drveta (x): koliko godina, toliko godova

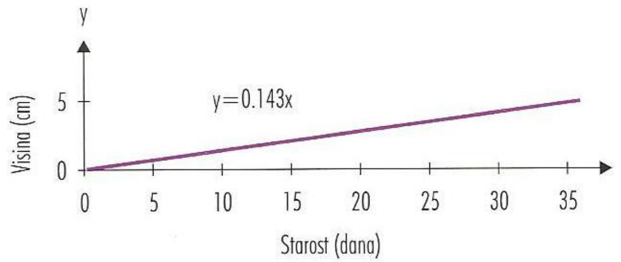


SLIKA 23.
Odnos starosti drveta (x) i broja godova na deblu (y)

Razvidno je da odnosom $y = x$ možemo "predvidjeti" vrijednost y-a za određenu vrijednost x-a (na primjer, ako je drvo staro 21 godinu, tada ima $y = x = 21$ god).

Vrijednost y-a ovisi o vrijednosti x-a, pa se najčešće nezavisna varijabla (x) smatra uzrokom, a zavisna (y) posljedicom.

Funkcionalni odnos, na primjer, između starosti biljke u danima (x) i pripadajuće visine u cm (y) izražen kao $y = 1/7x$ ili $y = 0.143x$ prikazan je na slici 24.



SLIKA 24.
Odnos starosti biljke (x) i visine biljke (y)

To znači nakon 5 dana, biljka će biti visine 0,71 cm jer

$$y = \frac{1}{7}x$$

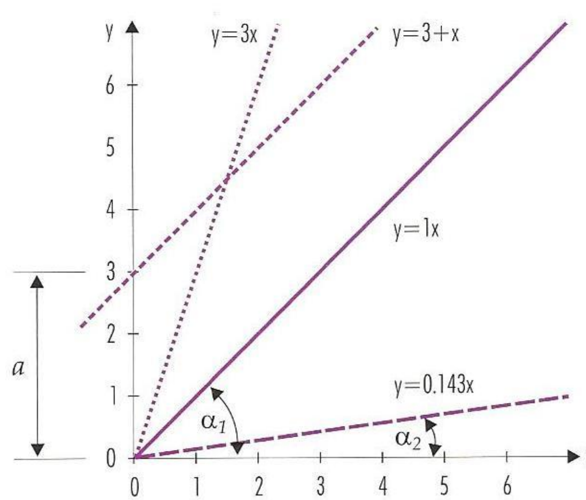
$$y = \frac{1}{7} \cdot 5 = 0.71$$

nakon 10 dana 1.43 cm jer

$$y = \frac{1}{7} \cdot 10 = 1.43$$

15 dana stara biljka ima visinu 2.14 cm itd.

Koeficijent b je razlogom za veći ili manji nagib pravca. Što je vrijednost b veća, to je pravac strmiji (slika 25). Regresijski koeficijenti za funkcije prikazane na slici su: b=1 (za y=x i y=3+x), b=3 (za y=3x) i b=0.143x (za y=0.143x).



SLIKA 25.
Pravci različitih nagiba ovisno o koeficijentu smjera (b)

Pravac se može izraziti funkcijom

$$y = a + bx$$

Pri tome je a odsječak na ordinati ili točka u kojoj pravac siječe os y (a = 0 ako pravac prolazi kroz ishodište); b je koeficijent smjera pravca, odnosno tangens kuta što ga pravac zatvara s osi x, a to je kvocijent nasuprotne i priležeće katete

Ako za funkciju y = x izračunamo tangens kuta α, za recimo $x = 2$ ili $x = 5$ ili $x = 7$ (vrijednosti priležećih kateta) i pripadajućih vrijednosti za nasuprotnu katetu y (a te su iste kao x-evi, jer je $y = x$), tada je

$$\operatorname{tg}\alpha_1 = \frac{2}{2} = 1.0$$

$$\operatorname{tg}\alpha_1 = \frac{5}{5} = 1.0$$

$$\operatorname{tg}\alpha_1 = \frac{7}{7} = 1.0$$

Dakle, tg $\alpha_1 = 1$, odnosno $b = 1$.

Za funkciju y = 1/7 x iz vrijednosti za

$x = 2 \rightarrow y = 2/7 = 0.28$

$x = 5 \rightarrow y = 5/7 = 0.71$

$x = 7 \rightarrow y = 7/7 = 1.00$

pa je $\operatorname{tg}\alpha_2 = \frac{0.28}{2} = 0.143$

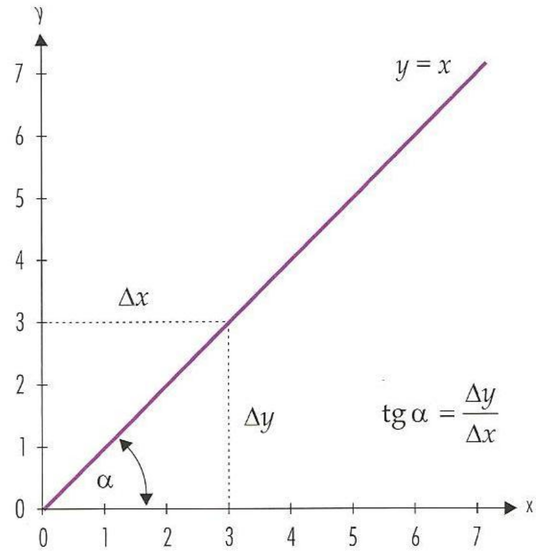
$$\operatorname{tg}\alpha_2 = \frac{0.71}{5} = 0.143$$

$$\operatorname{tg}\alpha_2 = \frac{1.00}{7} = 0.143$$

Dakle tg $\alpha_2 = 0.143$, odnosno $b = 0.143$.

Dakako da je za vrijednost x-a i y-a u bilo kojoj točki ista vrijednost za tangens kuta što ga pravac zatvara s osi x. Općenito, za bilo koju promjenu vrijednosti x i y varijable (tj. Δx i Δy) tangens kuta je:

$$\operatorname{tg}\alpha = \frac{\Delta y}{\Delta x}$$



SLIKA 26.
Koeficijent smjera pravca je tangens kuta α

Uz jediničnu promjenu vrijednosti x-a, dakle uz Δx=1,0 tangens kuta će biti jednak promjeni vrijednosti y-a tj. Δy, jer

$$\mathrm{tg}\,\alpha = \frac{\Delta y}{1} = \Delta y$$

To konačno znači da je

$$b = \mathrm{tg}\ \alpha = \Delta y$$

U biometrici se b zove **regresijskim koeficijentom**. **Regresijski koeficijent (*b*) je, dakle, iznos za koji se promijeni zavisna varijabla (*y*) uz jediničnu promjenu nezavisne varijable** (x).

Funkciju $y = a + bx$ zovemo **regresijskom jednadžbom.**

Ona je za

prvi primjer $y = 0 + 1x = x$

a drugi $y = 0 + 0.143x = 0.143x$

Oba pravca prolaze kroz ishodište, pa je $a = 0$.

Iz regresijske jednadžbe moguće je za bilo koju vrijednost x-a izračunati pripadajuću vrijednost y-a, koju se bilježi kao $\hat{y}$.

$$\hat{y} = a + bx$$

Kako pravac mora prolaziti kroz točku $\mathrm{T}(\bar{x}, \bar{y})$, to možemo pisati

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

odakle je

$$a = \bar{y} - b\bar{x}$$

Uvrstimo li ***a*** u regresijsku jednadžbu dobivamo ***jednadžbu linearne regresije***

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

Regresijski koeficijent (***b***) se računa kao

$$b = \frac{SP_{xy}}{SS_x}$$

SS_x je suma kvadratnih odstupanja za x varijablu

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

a SP_{xy} zovemo **sumom produkata** ovih dviju koreliranih varijabli.

Svaki SS možemo pisati i kao:

$$SS_x = \Sigma(x \cdot x) - \frac{\Sigma x \cdot \Sigma x}{n}$$

Zamijenimo li u tom izrazu jedan ***x*** s ***y*** dobivamo formulu za računanje SP_{xy}:

$$SP_{xy} = \Sigma(x \cdot y) - \frac{\Sigma x \cdot \Sigma y}{n}$$

Prisjetimo se da je varijanca mjerilo varijabilnosti jednog svojstva ili računski, to je prosječno kvadratno odstupanje varijanata od srednje vrijednosti $(s^2 = \frac{SS}{n-1})$. Analogno tome **zajedničko mjerilo varijabilnosti dvaju svojstava koja su u korelaciji je kovarijanca** (Cov).

$$Cov = \frac{SP_{xy}}{n-1}$$

Za razliku od varijance koja je uvijek pozitivan broj (jer je prosječno kvadratno odstupanje), kovarijanca može biti i negativna (jer produkti mogu biti negativni!)

7.2. Korelacija

Regresijom se opisuje način povezanosti dvije varijable, a **korelacijom otkrivamo u kom smjeru i do kojeg stupnja dvije varijable zajednički variraju - kovariraju.**

I u svakodnevnom životu i u struci susrećemo se s pojmom korelacije, *znali to* ili ne.

Svaki grafikon je, zapravo, prikaz korelacijske povezanosti dvije varijable i svaki od primjera povezanosti: visine i težine čovjeka, visine biljke i broja listova na biljci, količine gnojiva i prinosa, broja cvjetova i broja zametnutih plodova, učinkovitosti pesticida i prinosa, predstavlja odnos između dva svojstva u kojem porastom vrijednosti jednog rastu i vrijednosti drugog (viša osoba ima i veću težinu, što je veći broj cvjetova bit će i veći broj plodova, itd.).

U primjerima: dob u čovjeka i broj zdravih zubi, prinos i kvaliteta, sadržaj šećera u soku i sadržaj kiselina, % bjelančevine i % škroba u zrnu, radi se o drugačijoj povezanosti svojstava. Naime, većim vrijednostima jednog pripadaju manje vrijednosti drugog svojstva (starija osoba ima manje zdravih zubi, što veći prinos slabija je kvaliteta, što više šećera to manje kiseline u soku, itd.)

Sve su to primjeri korelacijske povezanosti dvije varijable, s time da prva skupina predstavlja **pozitivnu korelaciju**, a druga **negativnu korelaciju**.

Pozitivna korelacija je takva povezanost svojstava, kod koje porastom vrijednosti jednog (obično x) rastu i pripadajuće vrijednosti drugog svojstva (y).

Kod **negativne korelacije** porast vrijednosti jedne varijable, popraćen je padom pripadajućih vrijednosti druge.

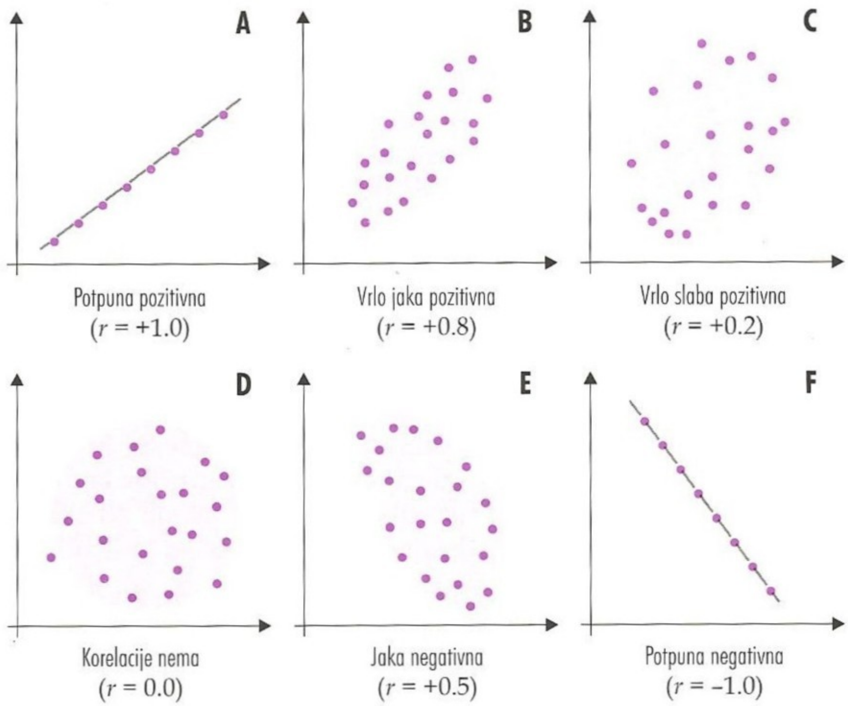
Kod korelacije nas interesira stupanj (jačina) povezanosti među varijablama. Stoga teoretski i nije bitno koja je nezavisna (x) a koja zavisna (y), iako se mogu označiti po logici stvari (gnojidba kao x, a prinos kao zavisna-y; dob-x, a broj zdravih zubi-y).

Koliko jaka i kakva je povezanost među varijablama procjenjujemo **korelacijskim koeficijentom** (r).

$$r = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}}$$

Korelacijski koeficijent je parametar koji govori o jačini i smjeru (načinu) povezanosti među svojstvima.

Na slici 27 su ilustrirane linearne korelacije različitih smjerova i različite jačine.



SLIKA 27.
Korelacije različitih smjerova i jačina

Točkice predstavljaju parove vrijednosti x i y varijable. Varijabilnost i jednog i drugog svojstva razlogom je njihove raspršenosti (disperzije).

Ponekad istoj vrijednosti x-a pripada veća, nekad manja vrijednost y-a. Tako već graf ukazuje na bliskost veze ili na jačinu povezanosti dvije varijable.

U slučaju kad s linearnim porastom vrijednosti jedne varijable rastu i vrijednosti druge varijable, a pri tom svakoj vrijednosti jedne varijable korespondira samo jedna vrijednost druge varijable, govorimo o potpunoj pozitivnoj korelaciji (A). Analogno tome, ako se linearnim porastom vrijednosti jedne varijable linearno smanjuju vrijednosti druge varijable i to tako da svakoj vrijednosti jedne pripada samo jedna vrijednost druge varijable, radi se o potpunoj negativnoj korelaciji (F). Ako se pak linearnom promjenom vrijednosti jedne varijable linearno mijanjaju i vrijednosti druge varijable, ali tako da jednoj vrijednosti jedne varijable pripada više različitih vrijednosti druge varijable korelacija je (bilo pozitivna, bilo negativna) slabija ili jača (B, C i E) ili je nema (D).

Dakle: ***korelacija može po smjeru biti pozitivna ili negativna, a po jačini od vrlo slabe ili nikakve do vrlo jake i potpune.*** Tako korelacijski koeficijent (r) može poprimiti vrijednost od nula do jedan, ili točnije s obzirom i na smjer

$$-1.0 \leq r \leq +1.0$$

Roemer i Orphal su izradili tablicu iz koje se, ovisno o vrijednosti r-a, može očitati jačina korelacije.

Roemer-Orphal-ova tablica

korelacijski koeficijent (r)	jačina korelacije
0.00 – 0.10	nema
0.10 – 0.25	vrlo slaba
0.25 – 0.40	slaba
0.40 – 0.50	srednja
0.50 – 0.75	jaka
0.75 – 0.90	vrlo jaka
0.90 – 1.00	potpuna

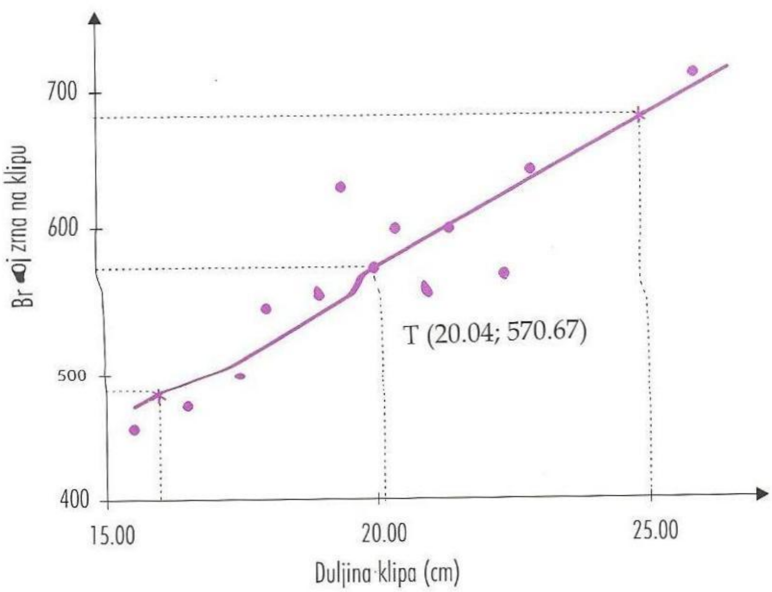
Stoga, da bi se iz podataka o vrijednostima x i y lako uočilo o kakvoj se korelaciji radi, najčešće je prvi korak grafički prikazati parove vrijednosti x i y, čiji smjer i disperzija već ukazuju na tip korelacije.

PRIMJER 7.2.1.

Cjelovitu regresijsku i korelacijsku analizu prikazat ćemo na primjeru duljine klipa i broja zrna na klipu kukuruza. Za primjer ćemo koristiti samo mali dio (n = 12) jednog velikog reprezentativnog uzorka. Kako smo već više puta naglašavali, i za korelacijsku analizu potrebno je, dakako, uvijek nastojati osigurati čim veći uzorak!

Osnovni podaci za ova svojstva u našem primjeru su:

Klip broj	Duljina klipa (u cm) x	Broj zrna na klipu y
1	16.5	470
2	15.5	453
3	21.0	560
4	26.0	713
5	21.5	600
6	18.0	552
7	19.5	630
8	23.0	643
9	22.5	567
10	19.0	560
11	20.5	600
12	17.5	500
$n = 12$	$\sum x = 240.5$ $\bar{x} = 20.04$	$\sum y = 6848$ $\bar{y} = 570.67$



SLIKA 28.
Korelacija i regresijski pravac za duljinu klipa (cm) i broj zrna na klipu

Iz ovog grafičkog prikaza (slika 28) naslućuje se pozitivna korelacija (dulji klip–veći broj zrna). No, utvrdimo to računanjem korela-cijskog koeficijenta:

$$r = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}}$$

$$SS_x = \sum(x^2) - \frac{(\sum x)^2}{n}$$

$$SS_x = 16.5^2 + \ldots + 17.5^2 - \frac{240.5^2}{12} = 98.73$$

$$SS_y = \sum(y^2) - \frac{(\sum y)^2}{n}$$

$$SS_y = 472^2 + \ldots + 500^2 - \frac{6848^2}{12} = 60294.7$$

$$SP_{xy} = \sum(x \cdot y) - \frac{\sum x \cdot \sum y}{n}$$

$$SP_{xy} = (16.5 \cdot 470) + \ldots + (7.5 \cdot 500) - \frac{240.5 \cdot 6848}{12} = 2186.67$$

$$r = \frac{2186.67}{\sqrt{98.73 \cdot 60294.7}} = 0.896$$

Iz Roemer Orphal-ove tablice, ova vrijednost korelacijskog koeficijenta ukazuje na postojanje vrlo jake pozitivne korelacije između duljine klipa i broja zrna na klipu.

Kako je korelacijski koeficijent stupanj do kojeg dvije varijable zajednički variraju ne izražava se u nikakvim mjernim jedinicama, već je neimenovan broj.

No s kolikom pouzdanošću možemo na temelju dobivenog r-a tvrditi da se radi o vrlo jakoj pozitivnoj korelaciji između duljine klipa i broja zrna na klipu? Da bismo saznali, dakle, da li je taj odnos samo posljedica slučaja ili je značajan, treba testirati opravdanost korelacijskog koeficijenta, dakle **testirati nultu hipotezu o r-u**, ili **testirati hipotezu da se r ne razlikuje značajno od nule**, ili

$$H_0 : r = 0$$

Za testiranje ove H_0 primijenjuje se t-test (računanjem t_{exp} i usporedbom s t_{tabl}):

$$t_{exp} = \frac{r}{s_r}$$

$$s_r = \sqrt{\frac{1-r^2}{n-2}}$$

$$t_{exp} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \sqrt{\frac{r^2(n-2)}{1-r^2}}$$

U ovom primjeru t_{exp} iznosi

$$t_{exp} = \sqrt{\frac{0.896^2(12-2)}{1-0.896^2}} = 6.38$$

***Tablični faktor t očita se na temelju n - 2 slobodnih varijanata (radi se o parovima i dvije varijable)* tj. n - 2 = 10, dakle**

$t_{P=5\%} = 2.23 \qquad t_{P=1\%} = 3.17$

Kako je $t_{exp} > t_{tabl}$ odbacujemo nultu hipotezu i zaključujemo da se radi o visoko signifikantnoj vrlo jakoj pozitivnoj korelaciji između duljine klipa i broja zrna na klipu, što pišemo kao

$r = 0.896^{**}$

Ovu vezu ćemo opisati i regresijom.

Ako se radi o linearnom odnosu, tada ga je moguće opisati regresijskim pravcem i to onim pravcem od kojeg su odstupanja pojedinih točaka (parova x i y) najmanja. Iz slike 28 je vidljivo da s obzirom na disperziju parova x i y vrijednosti pravac ne može prolaziti kroz sve točke, no prolazi kroz T ($\bar{x}$, $\bar{y}$).

Regresijska jednadžba glasi:

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

Sve vrijednosti u njoj već imamo, osim regresijskog koeficijenta (***b***) On se izračuna iz

$$b = \frac{SP_{xy}}{SS_x} = \frac{2186.67}{98.73} = 22.15 \text{ zrna}$$

Praktično, dakle, regresijski koeficijent (b) je broj koji pokazuje za koliko se promijeni vrijednost zavisne varijable (y) ako se

nezavisna varijabla (x) promijeni za jedinicu u kojoj je mjerena. Izražava se u mjernim jedinicama u kojima je mjerena zavisna varijabla. Može poprimiti bilo koju vrijednost ovisno o promjeni y varijable. Dakle, vrijednosti b-a nemaju ograničenja (kao što je to kod r-a).

To u ovom primjeru, znači, promijeni li se duljina klipa za 1 cm, broj zrna će se promijeniti za 22 (klip 1 cm dulji imat će 22 zrna više).

Uvrštavanjem vrijednosti u jednadžbu linearne regresije, dobivamo

$$\hat{y} = 570.67 + 22.15\,(x - 20.04)$$

$$\hat{y} = 126.78 + 22.15\,x$$

iz čega je moguće za bilo koju vrijednost x-a izračunati (predskazati) pripadajući y.

Kod negativne korelacije, regresijski koeficijent će, dakako, imati negativan predznak, što znači da se s porastom vrijednosti x varijable za 1 jedinicu (u kojoj je mjerena) y varijabla smanjuje za iznos regresijskog koeficijenta izražen u jedinicama y varijable.

Ako smo izračunali, na primjer, da je b = -0.5 kg, to znači: poveća li se vrijednost x varijable za jedinicu u kojoj je mjerena, pripadajuća vrijednost y varijable će se smanjiti za 0.5 kg.

Još nešto valja uočiti: korelacijski i regresijski koeficijent uvijek su istog predznaka. Razlog je tome način povezanosti x i y varijable, a i računski to mora biti tako jer je u brojniku formule za računanje i r-a i b-a suma produkata (SP_{xy}) koja može imati + ili - predznak.

No, ***i regresijski koeficijent treba testirati, odnosno provjeriti nultu hipotezu o njemu***

$$H_0 : b = 0$$

Evo kako:

$$t_{exp} = \frac{b}{s_b} \qquad s_b = \sqrt{\frac{SS_y - \frac{SP_{xy}^2}{SS_x}}{SS_x(n-2)}}$$

$$t_{exp} = \frac{22.15}{3.3} = 6.7 \qquad s_b = \sqrt{\frac{60294.7 - \frac{2186.7^2}{98.73}}{98.73\,(12-2)}} = 3.305$$

Tablični t faktor, očitan iz n - 2 = 10 iznosi

$$t_{P=5\%} = 2.23 \qquad t_{P=1\%} = 3.17$$

Kako je $t_{exp} > t_{tabl}$ to znači da je regresijski koeficijent opravdan, značajan ili signifikantan uz P = 1%, što se označava kao

$$b = 22.1^{**}$$

S 99% sigurnosti možemo, dakle, ustvrditi, da se radi o vrlo jakoj pozitivnoj korelaciji između duljine klipa i broja zrna na klipu, a da će se broj zrna na klipu povećati za 22 zrna ako se duljina klipa poveća za 1.0 cm.

Regresijska nam jednadžba omogućuje ovu korelacijsku vezu prikazati pravcem od kojeg su odstupanja najmanja, a to je regresijski pravac. U našem primjeru regresijski pravac je

$$\hat{y} = 126.78 + 22.15x$$

Da bi se nacrtao pravac regresije za ovaj primjer izračunati ćemo dvije točke, odnosno procijeniti pripadajuće vrijednosti y za x = 16.0 cm i x= 25.0 cm, uvrštavanjem vrijednosti x-a u regresijsku jednadžbu

$$\hat{y}_{x=16} = 126.78 + 22.15 \cdot 16.0 = 481.18$$

$$\hat{y}_{x=25} = 126.78 + 22.15 \cdot 25.0 = 680.53$$

Na slici 28, gdje smo grafički prikazali povezanosti x i y varijable (iz n = 12), ucrtan je pomoću ovih vrijednosti i regresijski pravac, koji dakako, prolazi kroz točku T(20.04; 570.67).

Za bilo koju duljinu klipa (koju nismo ni izmjerili) moguće je tako izračunati pripadajući broj zrna na klipu. Na primjer, zanima nas koliki broj zrna bi imao klip dugačak 24.0 cm? U regresijsku jednadžbu za x uvrstimo 24.0 pa dobijemo:

$$\hat{y} = 126.78 + 22.15 \cdot 24.0 = 658.4$$

Klip te duljine imat će, dakle, 658 zrna.

■

U prirodi gotovo i nema svojstva koje bi bilo neovisno o drugim svojstvima. Ona su obično povezana i zavisna. Zato je poznavanje regresije i korelacije izuzetno važno za otkrivanje i objašnjavanje takvih veza.

Kad se regresijom utvrdi uzročno posljedična veza dvije varijable, moguće je preko regresijske jednadžbe procijeniti vrijednost y za bilo koju vrijednost x-a. To se često koristi u situacijama kad je za jednu varijablu potrebno dugotrajno ili skupo mjerenje. Poznavanjem odnosa s drugom varijablom, regresijom je moguće procijeniti njezine vrijednosti (računski, dakle, bez mjerenja). Zbog toga je regresija neobično korisna metoda "predskazivanja" u svim prognostičkim službama.

■

Raspravili smo, dakle, o linearnoj korelaciji i regresiji, odnosno linearnom korelacijskom koeficijentu r (koji izražava jačinu povezanosti dvije varijable) i regresijskom koeficijentu b (koji izražava promjenu zavisne varijable, uzrokovanu jediničnom promjenom nezavisne - dakle pokazuje kako su zavisne).

Naravno da uvijek treba provjeriti pouzdanost r-a i b-a, znači testirati nultu hipotezu $H_0 : r = 0$ odnosno $H_0 : b = 0$

Kako su korelacija i regresija jedno od područja primjenjene statistike u kojima se vrlo često čine pogreške u smislu pogrešne interpretacije, to treba naglasiti nekoliko najvažnijih principa koji se ne smiju zaboraviti ili zanemariti.

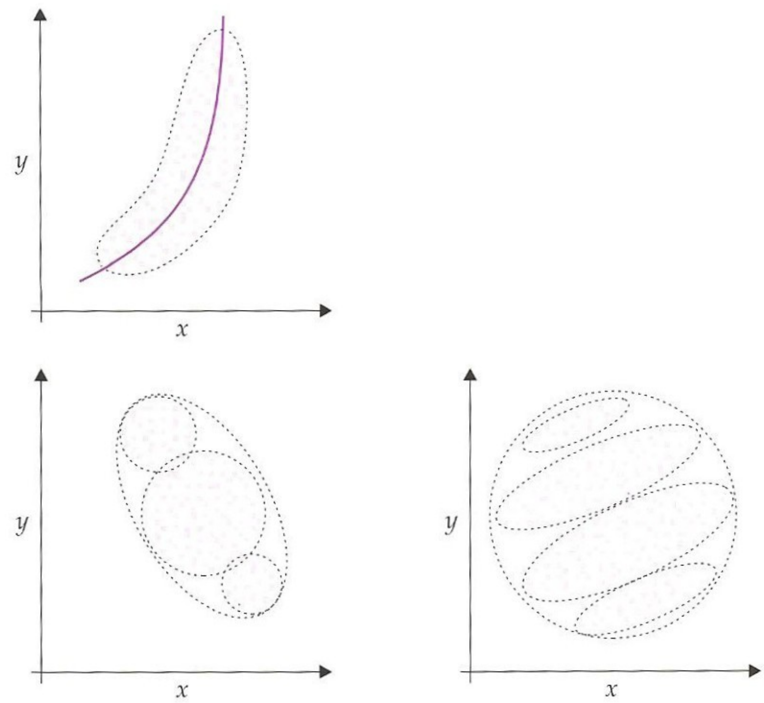
Kad se govori o korelacijskom koeficijentu, ne smije se zaboraviti da se radi o **koeficijentu linearne korelacije**. Isto tako u definiciji korelacije ništa ne ukazuje na to da se u vezi dvaju svojstava (varijabli) radi o uzročno-posljedičnom odnosu.

Dobivena vrlo niska korelacija ne znači da korelacije između dva svojstva nema, baš isto tako kao što se može izračunati vrlo visoki korelacijski koeficijent, iako se radi o dvije varijable koje nisu u nikakvoj uzročno-posljedičnoj vezi, a često ni korelaciji uopće.

U prvom slučaju razlog za krivi zaključak može biti u tome što se eventualno ne radi o linearnoj korelaciji, a procijeni se pomoću koeficijenta linearne korelacije (r).

U drugom pak slučaju, jedan od razloga za pogrešnu interpretaciju je taj, što je korelacijski koeficijent samo mjerilo jačine povezanosti dvaju svojstava, ali ne govori ništa o tome koja je varijabla uzrok, a koja posljedica (o prirodi te povezanosti). To kaže regresija.

Isto tako se do pogrešaka u interpretaciji dolazi ako se zanemari način na koji su varijable povezane (slika 29).



SLIKA 29.
Mogući razlozi pogrešne analize i interpretacije korelacije

U prvom primjeru radi se o korelacijskoj vezi koja nije linearna. Ako se to zanemari pa se ili cjelovita ili djelomična opisuje pomoću koeficijenta linearne korelacije, postupa se potpuno pogrešno. Ta je veza možda samo u nekom manjem dijelu linearna (samo za neke parove vrijednosti x i y varijable, što treba posebno naglasiti), ali u cijelosti se radi o nelinearnoj korelaciji.

U drugom primjeru očito se radi o nekoliko grupa, unutar kojih nema nikakve korelacije između varijabli. Zanemarivši to i uzevši sve kao cjelinu, može se doći do zaključka da se radi o vrlo jakoj negativnoj korelaciji, što je potpuno pogrešno.

U trećem primjeru prikazan je slučaj, gdje se ako se zanemare grupe unutar kojih postoji jaka pozitivna korelacija (a isto je moguće u slučaju negativne korelacije unutar grupa), računajući korelacijski koeficijent može zaključiti da korelacije između x i y varijable uopće nema. To je, naravno, kao i u ranijim primjerima, potpuno pogrešan zaključak. Isto tako je vrlo opasno zaključivati o korelaciji izvan intervala u kojima su varijable mjerene.

U našem primjeru korelacije između duljine klipa kukuruza i broja zrna na klipu, konstatirali smo vrlo jaku pozitivnu korelaciju. Međutim, o njoj smijemo tvrditi samo za one klipove čija duljina je u granicama opaženih-izmjerenih vrijednosti tj. između 15.5 i 26.0 cm.

Zato uz korelacijski koeficijent r uvijek treba navesti i granice x i *y varijable unutar kojih vrijedi takva korelacija. Kakva bi korelacija izvan tih granica bila, to ne možemo reći. Ako se želi utvrditi u kakvom su odnosu dvije varijable izvan granica u kojima su vršena mjerenja, moraju se, svakako, izvršiti i mjerenja u onom području u kojem su vrijednosti koje nas interesiraju. Svakako izbjegavati ekstrapoliranje!*

I još nešto, iako se čini vrlo trivijalnim: korelacijski koeficijent na bazi samo dva para podataka bit će uvijek +1 ili -1, ali to je, naravno, besmislica.

Radi li se o linearnoj ili nekoj od mogućih nelinearnih veza, najlakše je uočiti ako se podaci urede samo grafički (prikažu točkama u grafu koji izražava vezu između *x* i *y* svojstva). Već takav jednostavan prikaz podataka često može reći mnogo više nego bilo kakva statistička analiza primijenjena na podatke na koje se ne bi smjela primijeniti. Stoga bi preporuka u analizi korelacije bila: najprije grafički prikazati osnovne podatke, a onda odlučiti što dalje. Na taj se način može uočiti i izbjeći mnoge pogreške u analiziranju i interpretaciji korelacijske veze.

Ako su dvije varijable povezane na taj način da se radi o nelinearnoj vezi, tada je moguće postupiti na više načina: ili pokušati doći do linearnog odnosa (a to se postiže transformiranjem podataka), ili pak postupnim ispitivanjem utvrditi koja krivulja najbolje aproksimira podatke, pa ići u analizu takve nelinearne veze.

Nelinearna povezanost može slijediti različite krivulje, no ako se podaci transformiraju u, recimo, logaritamsku skalu, može se doći do linearnog odnosa, pa je moguće primijeniti koeficijent linearne korelacije.

Na taj je način promjenom skale moguće nelinearnu korelaciju pretvoriti u linearnu. Često se to postiže već samo transformiranjem vrijednosti zavisne (*y*) varijable, a ponekad je to potrebno učiniti za obje.

■

Do sada je bilo riječi o odnosu između dvije varijable bez obzira na ostale varijable koje mogu istovremeno varirati. Taj se odnos zove **jednostavna** ili **ukupna korelacija**. No, treba makar kratko spomenuti **parcijalnu i multiplu korelaciju**. Naime, **parcijalna korelacija je odnos između dvije varijable u kojem se jednu ili više varijabli drži konstantnima**.

Ako se radi o, recimo, tri svojstva (x, y i z), moguća su tri jednostavna korelacijska koeficijenta i to: r_{xy}, r_{xz} i r_{yz}.

Parcijalni korelacijski koeficijenti se računaju po formulama:

$$r_{xy/z} = \frac{r_{xy} - r_{xz} \cdot r_{zy}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

$$r_{xy/z} = \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{(1-r_{xy}^2)(1-r_{zy}^2)}}$$

$$r_{xy/z} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1-r_{xy}^2)(1-r_{xz}^2)}}$$

Ilustrirati ih možemo primjerom: neka je x-duljina klipa, y-masa zrna na klipu a z-broj zrna na klipu. Jednostavni korelacijski koeficijenti su:

$$r_{xy} = 0.67 \qquad r_{xz} = 0.54 \qquad r_{yz} = 0.86$$

Da bi se u vezi između duljine klipa (x) i mase zrna na klipu (y) eliminirao utjecaj broja zrna na klipu (z), izračunat ćemo parcijalni korelacijski koeficijent $r_{xy/z}$:

$$r_{xy/z} = \frac{0.67 - 0.54 \cdot 0.86}{\sqrt{(1-0.54^2)(1-0.86^2)}} = \frac{0.2056}{\sqrt{0.9688}} = 0.21$$

Koeficijent parcijalne korelacije između duljine klipa i mase zrna na klipu, uz odstranjen utjecaj broja zrna na klipu ($r_{xy/z}$), je manji od r_{xy}.

To znači da broj zrna na klipu ometa pravu sliku korelacije između duljine klipa i mase zrna na klipu u smislu povećanja vrijednosti korelacijskog koeficijenta. Izdvoji li se njegov utjecaj dobije se ispravnija slika o vezi između ta dva svojstva.

Parcijalna korelacija ne uključuje pojavu nezavisne i zavisne varijabe. Ona je samo mjerilo međusobnih odnosa svojstava. Međutim, **multipla korelacija** i **regresija** govore o povezanosti i zavisnosti više varijabli. Naime, zavisna varijabla (y) može ovisiti o više nezavisnih (x_1, x_2, ... , x_n). Taj odnos se može izraziti jednadžbom multiple regresije:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

Analogno korelaciji i regresiji između dvije varijable, u slučaju, recimo, s 2 nezavisne i jednom zavisnom varijablom, odstupanja će biti najmanja ako je

$$a = \overline{y} - b_1 - b_1\overline{x}_1 - b_2\overline{x}_2$$

pa je onda regresijska jednadžba

$$\hat{y} = \overline{y} + b_1(x_1 - \overline{x}_1) + b_2(x_2 - \overline{x}_2)$$

Regresijski koeficijenti b_1 i b_2 se izračunaju po formulama:

$$b_1 = \frac{SS_{x_2} \cdot SP_{x_1y} - SP_{x_1x_2} \cdot SP_{x_2y}}{SS_{x_1} \cdot SS_{x_2} - (SP_{x_1x_2})^2}$$

$$b_2 = \frac{SS_{x_1} \cdot SP_{x_2y} - SP_{x_1x_2} \cdot SP_{x_1y}}{SS_{x_1} \cdot SS_{x_2} - (SP_{x_1x_2})^2}$$

Do svih vrijednosti za sume kvadrata (*SS*) i sume produkata (*SP*) lako je doći iz opaženih podataka, pa se za bilo koje vrijednosti nezavisnih varijabli može izračunati pripadajuća vrijednost zavisne varijable.

Postupak je potpuno analogan onom u slučaju dvije varijable (x i y).

■

8 Planiranje i izvođenje eksperimenata

10
10
11
12

8.1. Kako proučiti problem?
8.2. Principi planiranja pokusa

8.1. Kako proučiti problem?

U nastojanju da dođe do novih saznanja, proučavajući različite probleme i fenomene, poljoprivredni stručnjak se neprestano susreće s raznolikošću materijala na kojem temelji svoja istraživanja.

Kad je varijabilnost mala, moguće je i bez dugotrajnog i skupog rada, već samo na temelju većeg broja informacija ili podataka donositi zaključke i preporuke.

Ako je pak variranje unutar opažanja na istom materijalu veliko (a takvo najčešće u biološkim materijalima jest), potrebno je primijeniti takve postupke koji će osigurati objektivno i pouzdano vrednovanje i zaključivanje.

Općenito, u zaključivanju su poznata dva pristupa.

Ako se na osnovi poznavanja nekih općih principa utvrdi što će se dogoditi u specifičnim uvjetima (dakle iz općeg se zaključuje o specifičnom), govorimo o ***dedukciji*** ili ***deduktivnom zaključivanju***.

Drugi pristup je upravo obratan, jer se iz niza specifičnih slučajeva dolazi do općeg zaključka, generalizira se. To je ***indukcija*** ili ***induktivno zaključivanje***, a temelji se na određenoj količini informacija - podataka.

I induktivno i deduktivno zaključivanje je u znanstveno-istraživačkom radu neizbježno.

Spomenimo samo neke probleme s kojima se poljoprivredni stručnjak zaokuplja:

- interesira ga hoće li uvođenje nekog novog agrotehničkog zahvata dati bolji učinak od do tada korištenog
- je li nova sorta ili hibrid prinosniji od dosadašnjih
- u kojem roku sijati ili saditi da se postigne optimalni rezultat
- koji pesticid primijeniti u suzbijanju različitih štetočinja
- koje količine gnojiva treba primijeniti i kojih formulacija
- koji je optimalan sklop (broj biljaka po jedinici površine) za pojedine kulture itd., itd.

Da bi se odgovorilo na ta i brojna druga slična pitanja, potrebno je provesti **pokus** ili **eksperiment**.

Pokus ili eksperiment je planirano istraživanje čiji je cilj otkrivanje novog ili dokazivanje (ili pobijanje) rezultata prethodnih istraživanja.

Sastoji se od niza postupaka koji osiguravaju vjerodostojnost i pouzdanost odgovora na pretpostavke ili hipoteze, od kojih se polazi iz osobnih saznanja ili iskustva drugih.

Dokazati, potvrditi ili opovrgnuti hipoteze moguće je samo eksperimentiranjem i uz primjenu odgovarajućih statističkih postupaka. Ali, zbog uvijek prisutne varijabilnosti, to nije moguće sa 100% sigurnošću. Uvijek treba računati na određenu razinu nesigurnosti, nepouzdanosti, na određenu pogrešku.

Uzmimo najjednostavniji slučaj agronoma, koji želi usporediti prinos jedne nove sorte s prinosom standardne sorte čija svojstva su poznata i provjerena. Na dvije jednako velike parcele jednu do druge sije sortu A i sortu B, zatim mjeri prinose jedne i druge i zaključuje o boljoj (lošijoj) sorti. Smatrajući da je razlika među njima uvjetovana isključivo svojstvima sorata, čini nedopustivi propust. To, naime, sasvim sigurno nije istina, jer brojni drugi čimbenici (tlo, vlaga, bolesti, štetnici i dr.) također utječu na prinos (a dakako i druga svojstva) svake sorte.

Zato bi se vrednovanje ovih sorata moralo provesti pomoću postupka kojim se mogu odvojiti različitosti među sortama (uvjetovane njihovim genotipovima) od svih drugih uvjetovanih brojnim izvorima varijabilnosti.

Sorte A i B mogle bi se smatrati različito prinosnima samo ako je opažena razlika u prinosu veća nego bi bila da su obje parcele zasijane istom sortom.

Naime, razlika između potpuno istovjetno tretiranih parcela (u ovom slučaju razlika u prinosu između parcela zasijanih istom sotrom), zove se **pogreška pokusa** ili **eksperimentalna pogreška**.

To je, zapravo, varijabilnost unutar istovjetnih eksperimentalnih jedinica (parcela u ovom primjeru).

Pogreška pokusa je osnova za zaključivanje je li opažena razlika stvarna ili samo slučajna.

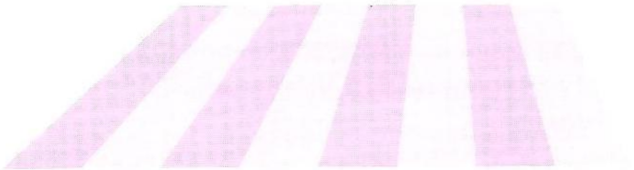
Upravo stoga zaključci i odgovori ne mogu biti apsolutni, te se generalizirati može samo uz određenu vjerojatnost pogreške koja se pri tome čini.

Zato eksperiment treba kreirati i provesti tako da postoji mogućnost mjerenja eksperimentalne pogreške.

Međutim, pogrešku pokusa ne može se procijeniti ako se ne raspolaže s bar dvije istovjetne eksperimentalne jedinice. Zato takve istovjetne jedinice pokusa treba ponavljati kako bi se dobile informacije na temelju kojih se može doći do mjerila pogreške pokusa.

U slučaju ove dvije sorte A i B, trebalo bi svaku posijati na bar dvije ili više parcela. No, nije niti to dovoljno! Ima još mnogo drugih čimbenika i razloga o kojima pri tom treba voditi računa.

Agronom može svaku ovu sortu ponovljeno sijati na, recimo, četiri parcele, kao:



A B A B A B A B

Ako je, međutim, samo plodnost tla različita i smanjuje se od lijeva prema desno, prinos će uvijek biti u korist sorte A već samo zbog plodnosti tla, pa bi na taj način usporedba ovih sorata bila opet pristrana. Dakle, dio različitosti sorata A i B bit će uvjetovan različitom plodnosti tla a ne samo sortnim razlikama.

Ispravnim eksperimentiranjem to se mora izbjeći i osigurati objektivnu, pouzdanu procjenu i vrednovanje rezultata pokusa.

Općenito, u pokusima se učinci procjenjuju ponavljanjem istovjetnih eksperimentalnih jedinica ili **članova pokusa** ili **tretiranja**.

Član pokusa ili tretiranje je bilo koji, od istraživača (stručnjaka) nametnuti zahvat, čiji se učinci prate i vrednuju u pokusu.

S obzirom na problematiku koja se proučava, cilj i eksperimentalni materijal, govorimo o ***sortnim pokusima*** (proučavaju se sorte, hibridi, linije ili općenito genotipovi), ***agrotehničkim pokusima*** (ispituju učinke različitih agrotehničkih zahvata kao: jesensko u odnosu na zimsko oranje, različiti načini obrade tla i sl.), ***gnojidbenim pokusima*** (rješavaju pitanja npr. optimalne kombinacije gnojiva, doze, načina i vremena primjene i sl.), ***pokusi s pesticidima*** (daju odgovore o efikasnosti pojedinih pesticida, koncentracija, načina aplikacije itd.), ***pokusi pomoću kojih se vrednuju različiti meliorativni zahvati*** itd.

Moglo bi se nabrajati još mnogo tipova eksperimenata, ovisno o problemima koji se pokusima rješavaju, a o tome ovise i članovi pokusa ili tretiranja.

- Tako bi agrotehnički pokus u kojem bi se proučavao učinak, recimo, tri načina obrade tla, imao tri člana ili tretiranja.
- Primjenom 6 kombinacija gnojiva i praćenjem njihovog utjecaja na prinos neke kulture, radilo bi se o pokusu sa 6 članova.
- U pokusu s dvije sorte pšenice, govorimo o dva člana ili tretiranja.
- *Sortni pokus s 15 hibrida suncokreta imao bi dakle 15 tretiranja.*
- Pokus u kojem bi se proučavao utjecaj 3 različite koncentracije nekog pesticida, imao bi 3 člana pokusa.

Pokusi se mogu izvoditi na otvorenom (polje, voćnjak, vinograd), ali i u zatvorenom prostoru (laboratorij, staklenik, plastenik, skladište, podrum i sl.).

No, bez obzira radi li se o otvorenom ili zatvorenom prostoru, za eksperiment je potrebno osigurati takve uvjete, koji će omogućiti **kontrolu** i **procjenu pogreške pokusa**.

Kontrola pogreške pokusa osigurava se

- **Grupiranjem članova u blokove**
- **Izborom medija za pokus**
- **Analizom eksperimentalnih podataka**

S obzirom da mogućnost otkrivanja postojećih razlika među članovima pokusa raste kako se pogreška pokusa smanjuje, dobar eksperiment karakteriziraju svi oni uvjeti koji minimaliziraju pogrešku.

Najvažnije je grupirati članove pokusa u grupe ili blokove, unutar kojih su uvjeti maksimalno slični. Na taj je način moguće mjeriti različitost među blokovima i tu varijabilnost izdvojiti iz pogreške pokusa.

To grupiranje tretiranja u blokove naročito je važno u poljskim pokusima.

Ispravna tehnika izbora medija za pokus također je važna. Za sve tipove pokusa apsolutno je bitno da svi ostali činioci, osim onih koje nameće istraživač, a to su članovi ili tretiranja, budu maksimalno ujednačeni (uniformni).

Tako je, na primjer, u sortnom pokusu (gdje su sorte tretiranja) nužno, da svi mogući čimbenici budu za sve sorte isti (tlo, gnojidba, gustoća sklopa, agrotehnika, zaštita, te bezbroj drugih). Znači sve parcele pokusa moraju biti uniformne.

Dakako da je nemoguće apsolutno zadovoljiti ovim uvjetima, ali poznavanjem mogućih uzroka varijabilnosti (izvora varijabilnosti) moguće ih je ili izbjeći ili eliminirati i tako svim tretiranjima (članovima) pokusa osigurati maksimalno homogene uvjete.

Izvori varijabilnosti, naročito u poljskim pokusima mogu biti:

- različitost među individuima (biljkama, stablima, parcelama i dr.)
- klimatski i općenito okolinski uvjeti
- heterogenost tla (plodnost, nagibi, mikrodepresije i sl.)
- kultura koja je prethodila
- gnojidba prethodne godine
- rezidui pesticida, gnojiva itd.

Dakako, da je **samo ispravnom i primjerenom analizom eksperimentalnih podataka moguće pravilno procijeniti pogrešku pokusa, koja je osnova za prihvaćanje ili odbacivanje postavke od koje se krenulo u eksperiment**.

Procjena pogreške pokusa postiže se

- **Repetiranjem članova**
- **Randomiziranjem članova**

Ne smije se zaboraviti da su rezultati pokusa u najstrožem smislu riječi primjenljivi samo na uvjete koji su isti ili slični onima u kojima je eksperiment proveden.

Stoga pokus treba ponoviti i vremenski i prostorno da bi se rezultati mogli generalizirati za širi raspon okoline (biti primjenjivi u različitim uvjetima).

Ponovo naglasimo i ne zaboravimo da pogreška pokusa mjeri varijabilnost između opažanja istog tretiranja, a moguće ju je mjeriti samo ako je svako tretiranje ponovljeno više puta u pokusu.

Višekratno ponavljanje članova u pokusu zove se **repetiranje**.

Pri tome je važno obuhvatiti cjelovit medij u kojem se pokus izvodi, odnosno omogućiti da se svaki član nađe na svakom dijelu pokusnog medija. **To se postiže potpuno slučajnim raspoređi-vanjem članova pokusa ili *randomiziranjem*.**

Repeticija ili ponavljanje je dio pokusa koji sadrži sve potpuno slučajno raspoređene članove pokusa.

Najčešće je repeticija isto što i blok, iako ponekad jedan blok može sadržavati i više repeticija. Repetiranjem i randomiziranjem članova pokusa osigurava se mogućnost procjene pogreške pokusa.

U pripremi uvjeta za eksperiment kao i tijekom izvođenja pokusa, treba učiniti sve što je moguće da bi se pogreška pokusa smanjila na minimum.

Poljoprivrednog stručnjaka, agronoma, trajno zaokupljaju fenomeni, problemi, pitanja struke, koje istražuje i proučava najčešće s konačnom namjerom da rezultati budu preporučljivi i primjenjivi u praksi.

Pri tome pokusi ili eksperimenti predstavljaju integralni nezamjenjiv dio svakog istraživanja.

Kad se pristupa proučavanju nekog problema potrebno je slijediti ove etape:

- **Definirati problem i formulirati postavke ili hipoteze i cilj** (odnosno pitanja na koja se želi odgovoriti) - treba dobro razumjeti problematiku koju se želi proučiti (što znači detaljno se upoznati sa svim što je već učinjeno na istoj ili sličnoj), jer samo tako je moguće oblikovati pitanja čiji odgovori vode rješenju problema.
- **Odlučiti o načinu provjere ispravnosti hipoteze (testiranja).**
- **Planirati i izvesti eksperiment, uz prikupljanje informacija i podataka iz pokusa.**
- **Odabrati postupak analize eksperimentalnih podataka**, čiji će rezultati omogućiti da se provjeri hipoteza i dođe do odgovora.

Posebno ćemo se zadržati na ovom trećem - planiranju i izvođenju eksperimenta ili pokusa.

Principi planiranja pokusa imaju korijene prvenstveno u logici znanstvene metodologije.

8.2. Principi planiranja pokusa

B. J. Winer u uvodu svoje knjige "Statistical Principles in Experimental Design" (1971.) kaže: "Planiranje eksperimenata može se usporediti s radom arhitekta koji odlučuje hoće li se graditi veći ili manji objekt i kakav. Osnovne zahtjeve postavlja vlasnik, a na arhitektu je da zahtjevima udovolji. On može predložiti nekoliko nacrta-planova koji svi udovoljavaju zahtjevima. Neki će, međutim, biti skuplji, neki imaju određene prednosti pred drugima i slično.

U planiranju pokusa ulogu arhitekta ima stručnjak za postavljanje pokusa i analiziranje rezultata pokusa - biometričar, a ulogu vlasnika projekta, istraživač - specijalist u određenom dijelu struke. Nerijetko je oboje utjelovljeno u jednoj osobi.

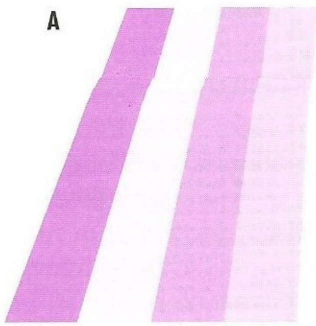
Bez dobro planiranog eksperimenta, niti potencijalno plodonosne hipoteze ne mogu se testirati sa zadovoljavajućom pouzdanosti."

Je li u istraživanju potreban eksperiment ili nije, ovisi o problemu koji se proučava kao i cilju, svrsi istraživanja. Ima problema koji se mogu riješiti i pitanja na koja se može odgovoriti samo opisivanjem pojedinih individua, mjerenjem većeg broja uzoraka i sl., a neki pak problemi (a takvih je u agronomskoj struci više) zahtijevaju kreiranje kontroliranih uvjeta i eksperimentiranje, što se postiže izvođenjem dobro planiranih pokusa.

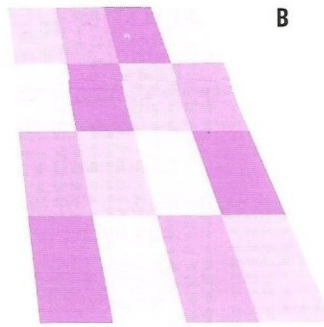
Bez obzira na tip eksperimenta, treba poduzeti sve ono što će svakom članu pokusa osigurati jednaku priliku da manifestira svoj učinak, a to prvenstveno znači da ih treba po mediju (bilo to polje, laboratorij, staklenik, skladište i sl.) randomizirati i repetirati (slučajno rasporediti i ponavljati).

Uzmimo, na primjer, određenu površinu na kojoj treba proučiti učinke 4 člana. Pridržavajući se spomenutih uvjeta, mogli bismo ih rasporediti tako da svaki zauzme po četvrtinu površine (A), pa potpuno slučajnim načinom ponovljeno zabilježiti informacije o njihovom učinku. Pri tome se zadovoljavaju uvjeti slučajnog rasporeda i ponavljanja.

A



B



Još bi bolje postupili kad bi tu istu površinu iskoristili na drugi način (B), grupirajući članove pokusa u repeticije (blokove) unutar kojih su isti uvjeti za svakog člana. Tako je svaki član četiri puta ponovljen, a njihov raspored unutar svakog ponavljanja potpuno je slučajan. Učinak svakog člana na taj se način prati po cijeloj površini, a opažanja i mjerenja se mogu bilježiti na svakoj eksperimentalnoj jedinici koju zovemo **osnovnom parcelom**.

Osnovna parcela u pokusu je, dakle, najmanja jedinica pokusa na koju je primijenjen pojedini član ili tretiranje.

Ovisno o tipu eksperimenta, problemu i ciljevima, osnovna parcela će ponekad morati biti veća, a ponekad će i manja biti dostatna.

U sortnim pokusima s kulturama gustog sklopa (kakve su npr. žitarice) veličina osnovne parcele je najčešće 5 - 10 m^2; s kukuruzom, suncokretom, i duhanom nešto veća tj. 15 - 20 - 25 m^2.

U *pokusima u voćarstvu i vinogradarstvu* osnovnu parcelu može predstavljati 1 do više stabala ili grmova, 3 i više trsova (čokota) i sl.

U *pokusima s pesticidima* osnovne parcele moraju biti veće zbog "preklapanja" tretiranja (pogotovo ako se aplicira strojem).

Gnojidbeni i agrotehnički pokusi zahtijevaju izuzetno veliku osnovnu parcelu (od 20 m^2 do nerijetko i 500 m^2), zbog opasnosti od preklapanja djelovanja, ispiranja, kao i zbog radnih zahvata poljoprivrednih strojeva.

U *laboratorijskim uvjetima*, osnovnu parcelu mogu predstavljati petrijevka, cilindar, uzorak za kemijsku ili drugu analizu itd.

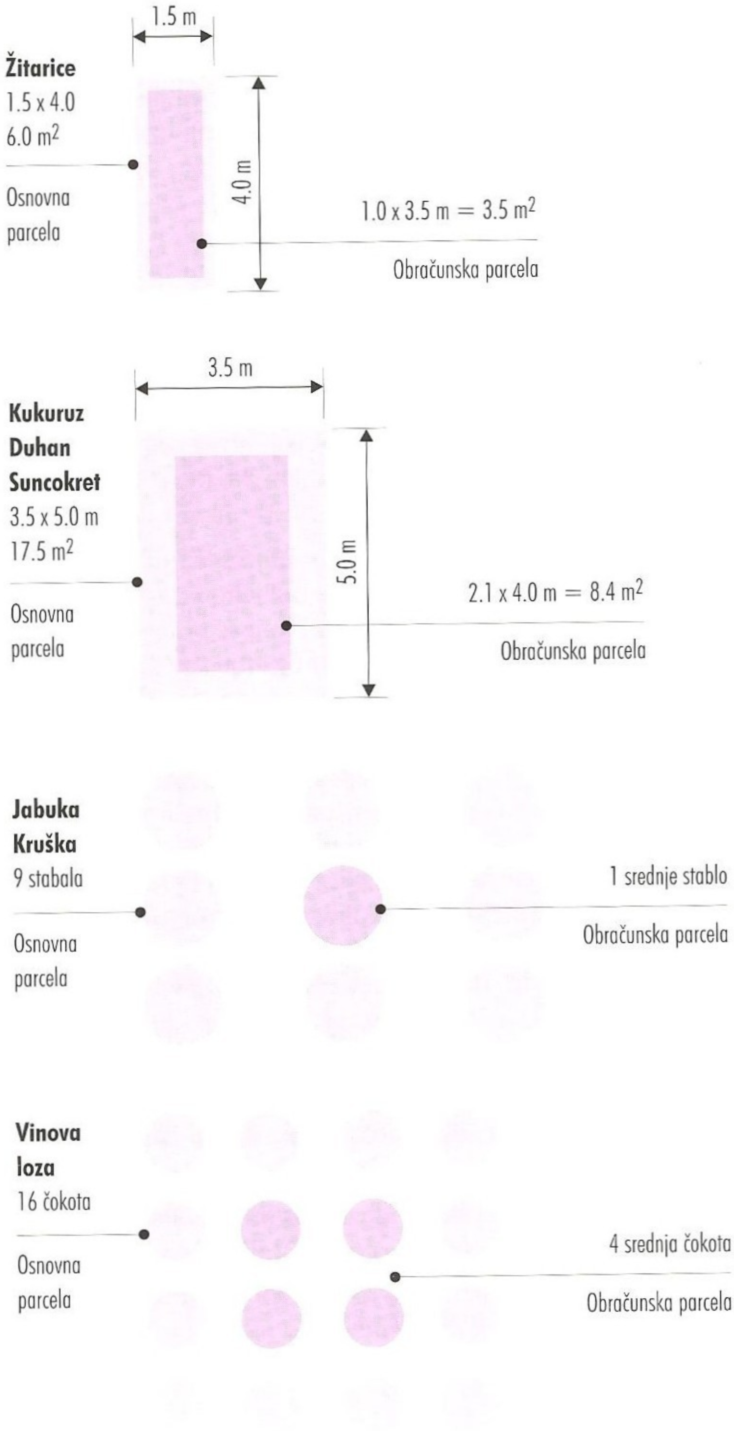
Iako, dakle, osnovna parcela u pokusu može biti različito velika ovisno o nabrojenim razlozima, općenito je važno da ona bude dovoljno velika da može osigurati vjerodostojnu informaciju o učinku člana ili tretiranja koje je na njoj primijenjeno.

Informacije su najčešće numeričke (izmjere, odvage i sl.), iako mogu biti i u obliku različito izraženih zabilješki o opažanjima, ali to moraju biti oni podaci koji se odnose isključivo na učinak člana ili tretiranja, bez eventualnog utjecaja susjednih parcela. Ti okolni utjecaji mogu biti uvjetovani zasjenjivanjem, kompeticijom, preklapanjem tretiranja (gnojiva, pesticida, regulatora rasta, agrotehničkih zahvata i sl.) itd.

Zato je najispravnije učinak svakog člana pokusa pratiti i zabilježiti samo s unutarnjeg dijela osnovne parcele, oslobođenog mogućih utjecaja "susjeda". To se obično postiže prethodnim odstranjivanjem rubova osnovne parcele (rubnih redova ili rubnih i čeonih biljaka, stabala, čokota i dr.). Taj **unutarnji dio osnovne parcele** (oslobođen mogućih utjecaja sa strane) **s kojeg se prikupljaju podaci o članovima tretiranja zove se obračunska parcela pokusa**.

S tim u svezi treba reći da se optimalnom veličinom osnovne parcele u pokusu (o kojoj onda ovisi i veličina obračunske parcele) smatra ona koja omogućuje smanjenje varijabilnosti uvjetovane nekontroliranim učincima na najmanju moguću mjeru.

Međutim, prevelika osnovna parcela može biti razlogom predimenzioniranosti cijelog pokusa, pa je onda upitna ujednačenosti uvjeta. Velika je vjerojatnost da se povećanjem pokusa na veliku površinu dolazi do problema heterogenosti medija, što dakako nije dobro.



SLIKA 30.
Nekoliko primjera osnovne i obračunske parcele

homogenosti uvjeta za pokus (što je medij za pokus heterogeniji, to je više repeticija potrebno).

6. **Osigurati kontrolu utjecaja susjednih parcela** - To se obično rješava eliminiranjem rubnih dijelova parcele.
7. **Detaljno odrediti koje podatke i kako prikupiti** - Jer oni će biti osnovica za procjenu učinaka tretiranja i testiranje različitih hipoteza.
8. **Skicirati postupak statističke analize podataka i prikaza rezultata**. Ispisati sve izvore varijabilnosti i pripadajuće slobodne varijante (*n* - 1) u analizi varijance, pa usporediti s ciljem eksperimenta da se već prije izvođenja pokusa vidi hoće li se pokusom doći do željenih odgovora.

Tek nakon svih ovih koraka moguće je

Provesti pokus

↓

Sakupiti podatke o učincima članova

↓

Analizirati podatke pokusa

↓

Odgovoriti na pitanje hipoteze i interpretirati rezultate

Kod izvođenja pokusa bitno je koristiti sve postupke koji eliminiraju subjektivnost, a isto tako odmah provjeravati podatke i opažanja za koje se čini da na bilo koji način odstupaju od onih koji se uopće mogu očekivati.

Osim svega, cjeloviti pokus treba izvesti u jednom danu (pod istim *uvjetima*).

Za ilustraciju, osnovne, odnosno obračunske parcele za različite biljne vrste i različite pokuse mogu izgledati kao na slici 30.

Opetovano smo naglašavali da je glavni cilj u pokusnom radu pogrešku pokusa svesti na najmanju moguću mjeru. Pri tome osnovna parcela, slučajnost raspoređivanja članova i ponavljanje članova u pokusu imaju najvažniju ulogu.

Sve ovo mogli bismo sažeti upitom:

Kako planirati eksperiment?

Dobro isplaniran pokus već je pola uspjeha. Zato bi obavezno uvijek trebalo slijediti ove korake:

1. **Definirati cilj** - To znači formulirati pitanja na koja se želi odgovoriti, hipoteze koje će se testirati, učinke (efekte) koji će se procijeniti. Cilj(evi) se mora(ju) detaljno ispisati, jer to istraživaču omogućuje bolje planirati eksperiment. Ako je više ciljeva, treba ih poredati prioritetnim redoslijedom (po važnosti).
2. **Izabrati tretiranja ili članove** - Pažljiv izbor tretiranja važan je kako u postizanju odgovora na postavljena pitanja, tako i u povećanju preciznosti eksperimenta. Često se čini pogreška izborom previše sličnih tretiranja, tretiranja koja se premalo razlikuju, pa se različitost njihovih učinaka ne može niti otkriti (doze, koncentracije, formulacije pripravaka, gnojiva, genotipova i dr.). Jednako tako nezgodnu i neželjenu posljedicu uzrokuje i izbor prerazličitih članova. Zato izboru članova treba pridavati izuzetnu pažnju.
3. **Izabrati eksperimentalni materijal** - Kod izbora materijala za pokus, treba imati na umu cilj(eve) eksperimenta i populaciju o kojoj se traže zaključci. Izabrani materijal mora dobro predstavljati populaciju na kojoj će se testirati tretiranja (ista starost i kondicija stabala voćne vrste, usjeva zaraženog bolešću ili štetnikom itd.).
4. **Izabrati eksperimentalni plan** - To je zapravo raspored članova; treba izabrati takav raspored članova pokusa koji će osigurati željenu preciznost.
5. **Izabrati osnovnu (obračunsku) parcelu i odlučiti o broju repeticija** - Odlučiti o veličini osnovne parcele (o čemu će ovisiti i veličina obračunske parcele) i o tome koliko će se puta članovi pokusa ponoviti. Broj repeticija u pokusu uglavnom ovisi o broju članova (što je manje članova, to više treba repeticija) i o

Treba izbjegavati sve one razloge koji bi mogli uvjetovati neželjenu varijabilnost. Zato treba u poljskim uvjetima sijati (saditi) okomito na smjer obrade, izbjegavati površinu na kojoj je prethodne godine bio gnojidbeni pokus ili pokus s pesticidima (zbog eventualnih rezidua), ne ići s gnojidbenim pokusom na površinu na kojoj je prethodne godine bila, recimo, soja (zbog ostataka N hranjiva), a repetirati i randomizirati, jer to su "alfa i omega" u eksperimentiranju.

Analiza podataka pokusa mora uvijek slijediti način na koji je pokus izveden - dakle plan pokusa ili raspored tretiranja ili shemu pokusa.

Rezultate pokusa treba interpretirati u svjetlu uvjeta eksperimenta, testiranih hipoteza ali i u odnosu na prethodna saznanja.

Ne treba pod svaku cijenu srljati u zaključivanje kad su rezultati iole "neobični" ili "sumnjivi". U takvom slučaju istraživanje se nastavlja i pokus ponavlja.

Treba upamtiti: nema negativnog rezultata pokusa. Ako je nulta hipoteza prihvaćena (uz zadovoljavajuću sigurnost), to samo znači da zaista nema značajnih različitosti među tretiranjima koja su se kroz eksperiment proučavala.

Statistička analiza eksperimentalnih podataka je nezamjenjiva karika u lancu istraživanje-eksperimentiranje-zaključivanje. Istraživaču pomaže kreirati i analizirati pokus, a jednako tako i valorizirati rezultate.

Naravno da svaki istraživač (a pogotovo svaki praktičar) ne mora biti kompletan biometričar, no mora se čvrsto pridržavati " **3R** ":

Randomizirati

Repetirati

Razgovarati s biometričarom

8

9 Planovi ili sheme pokusa

10

11

12

9.1. Potpuno slučajni raspored
9.2. Slučajni blokni raspored
9.3. Latinski kvadrat
9.4. Latinski pravokutnik

U izvođenju pokusa najvažnije je, dakle, osigurati mogućnost kontrole i ispravne procjene pogreške pokusa, koja sadrži svu slučajnu varijabilnost uvjetovanu različitim činiocima. Zato treba nastojati definirati čim više dijelova varijabilnosti i znati njihove izvore (uzroke).

To se postiže različitim načinima raspoređivanja, grupiranja i ponavljanja članova pokusa, poznatim kao **planovi** ili **sheme pokusa**.

Mnoštvo je različitih shema ili rasporeda tretiranja, a razlikuju se po tome kako su grupirane osnovne parcele na koje će se rasporediti članovi, kao i po restrikcijama koje se u pojedinom planu nameću. Iz toga proizlaze prednosti i mane različitih shema, kao i mogućnost definiranja više ili manje dijelova varijabilnosti, odnosno u konačnici različito velika pogreška pokusa.

Cilj je minimalna pogreška pokusa, pa je stoga potrebno sve ovo imati na umu svaki puta prilikom izbora sheme pokusa (plana, rasporeda).

Različite sheme pružaju različite mogućnosti analize eksperimentalnih podataka i sukladno tome rezultiraju različito velikom pogreškom pokusa. S tim u vezi svakako treba upamtiti:

Analizu pokusnih podataka **određuje** **Plan ili shema pokusa**

Podaci se moraju analizirati isključivo sukladno shemi po kojoj je pokus izveden.

U principu se analiza pokusnih podataka provodi primjenom analize varijance, odnosno testiranjem hipoteze o varijabilnosti članova pokusa (tretiranja) i razlika između njihovih prosječnih vrijednosti:

$$H_0 : \overline{x}_A = \overline{x}_B = \overline{x}_C = ... = \overline{x}_n$$

$$Test : F_{exp} = \frac{s^2 tretiranja}{s^2 pogreške}$$

(F_{tabl} baziran na *n-1* tretiranja i *n-1* pogreške)

Usporedba prosječnih vrijednosti tretiranja i testiranje razlika

$$\text{t-test} \rightarrow LSD = t_{\text{tabl}} \cdot s_D$$

(t_{tabl} baziran na *n-1* pogreške)

Specifičnosti ovisne o različitim shemama pokusa, pokazat ćemo na primjerima koji slijede.

9.1. Potpuno slučajni raspored

To je najjednostavnijt plan pokusa, odnosno raspored tretiranja na osnovne eksperimentalne jedinice. Ne zahtijeva nikakve posebne uvjete. Svaka osnovna parcela ima istu priliku da primi bilo koje tretiranje (člana), a tretiranja čak i ne moraju biti zastupljena isti broj puta.

Kod takvog rasporeda tretiranja u pokusu, razlike između eksperimentalnih jedinica (osnovnih parcelica) moraju se smatrati pogreškom pokusa. Naime, ta shema ne zahtijeva grupiranje članova pokusa u repeticije, pa iz toga proizlazi i velika pogreška pokusa!

Ipak, potpuno slučajni raspored bila bi zadovoljavajuća shema za homogene uvjete kakvi su, na primjer, uvjeti laboratorija ili staklenika, u kojima je relativno lako kontrolirati okolinu.

U poljskim pokusima taj se plan rijetko koristi i trebalo bi ga izbjegavati, osim kada se podaci sakupljaju na velikim proizvodnim površinama.

Pretpostavimo da se želi eksperimentirati s 3 člana ili tretiranja A, B i C, a svako da se u pokusu pojavljuje 4 puta. To znači trebalo bi osigurati 3 x 4 = 12 osnovnih parcela u pokusu. Na tih 12 parcela treba potpuno slučajno rasporediti članove pokusa. Da se izbjegne subjektivnost, najbolje je primijeniti neku potpuno objektivnu metodu. Najčešće je praktično ispisati 12 ceduljica (4 puta svaki od članova A, B i C), izmiješati ih i raspoređivati članove onim redom kojim se ceduljice uzimaju. To je onda ujedno i jedan od mogućih slučajnih rasporeda:

1 A	2 A	3 B	4 C
5 A	6 B	7 C	8 B
9 C	10 B	11 A	12 C

Podaci o učinku tretiranja prikupljeni na osnovnim parcelama podvrgnu se statističkoj analizi primjenom analize varijance (ANOVA) i testova za testiranje hipoteza (F i t-test).

Kako ovakav potpuno slučajni raspored omogućuje isključivo definiranje ukupne varijabilnosti i varijabilnosti između članova, to sav preostali dio varijabilnosti predstavlja pogrešku pokusa, koja je zbog toga vrlo velika.

Shematski:

$SS_{\text{Između članova}}$ **(tretiranja)**

SS_{Ukupno}

$SS_{\text{Unutar članova}}$ **(ostatak, pogreška)**

Pripadajući broj slobodnih varijanata za svaki dio varijabilnosti:

Izvori varijabilnosti	n - 1
Ukupno	n - 1
Tretiranja	$n_T - 1$
Pogreška	$n - n_T$

PRIMJER 9.1.1

U želji da se prouči utječe li ishrana gusjenica različitim travama na fertilnost ženki gubara (*Hypogimna morio* L.) proveden je laboratorijski pokus. Gusjenicama su ponuđene četiri različite trave: *Carex divulsa* (C), *Juncus conglomeratus* (J), *Lolium perenae* (L) i mješavina sve tri (M). To su, dakle, članovi pokusa, pa ovaj pokus ima 4 tretiranja.

Pretpostavka ili nulta hipoteza je, da različita ishrana gusjenica neće značajno utjecati na fertilnost ženki ovog štetnika. Pokus je izveden u laboratoriju, tako da su u po 7 cilindara gusjenice hranjene svakom od ove 4 trave. To znači, pokus je imao 28 osnovnih eksperimentalnih jedinica (7 cilindara svake od 4 trave), potpuno slučajno raspoređenih na stalku u laboratoriju.

U svakom cilindru bio je isti broj gusjenica, a broj odloženih jaja smatra se pokazateljem fertilnosti ženki. Evo tih podataka za svako tretiranje i svaki cilindar:

114 J	140 M	90 L	253 J	210 M	270 L	206 C
97 L	178 C	313 J	173 L	354 M	255 C	112 J
300 C	124 M	62 L	199 M	130 J	193 L	273 M
221 M	168 C	211 L	157 J	142 C	198 J	296 C

Ako ove podatke sredimo tako da ih svrstamo po travama, oni izgledaju ovako:

Cilindar broj	*Carex divulsa* C	*Juncus conglomeratus* J	*Lolium perenae* L	Mješavina C + J + L M	
1	206	114	90	140	
2	178	253	270	210	
3	255	313	173	354	
4	300	112	97	124	
5	168	130	62	199	
6	142	157	193	273	
7	296	198	211	221	
$\sum x_{\text{Trave}}$	1545	1277	1096	1521	$\sum x = 5439$
$\bar{x}_{\text{Trave}}$	220.7	182.4	156.6	217.3	

Čini se da je broj odloženih jaja različit, ovisno o ishrani gusjenica različitim travama. No, jesu li te vrijednosti značajno različite ili su razlike među njima možda samo slučajne, pokazat će detaljna analiza podataka.

Analizu pokusnih podataka provest ćemo primjenom analize varijance (ANOVA), znači izračunati sume kvadratnih odstupanja za pojedine dijelove varijabilnosti uvjetovane različitim izvorima (SS), pripadajuće slobodne varijante (n-1) i varijance (s^2), tj.

		n - 1
SS_{Ukupno}	$\rightarrow$	$(4 \cdot 7) - 1 = 27$
$SS_{\text{Između grupa}}$	$\rightarrow$	$4 - 1 = 3$
$SS_{\text{Unutar grupa}}$	$\rightarrow$	$4 \cdot (7 - 1) = 24$ (ili $27 - 3 = 24$)

Iz tih će vrijednosti biti moguće izračunati pojedine varijance, jer prisjetimo se: varijanca je prosječno kvadratno odstupanje varijanata od prosjeka $\left(s^2 = \dfrac{SS}{n-1}\right)$

Dakle:

$$SS_{\text{Ukupno}} = 206^2 + \ldots + 221^2 - \frac{5439^2}{28} = 150209.25$$

$$SS_{\text{Između trava}} = \frac{1545^2 + \ldots + 1521^2}{7} - \frac{5439^2}{28} = 19532.96$$

$$SS_{\text{Unutar trava}} = SS_{\text{Ukupno}} - SS_{\text{Između trava}} =$$

$$= 150209.25 - 19532.96 = 130676.28$$

Kod računanja $SS_{\text{Između trava}}$ dijeli se sa 7 jer je svaka vrijednost u brojniku dobivena iz 7 osnovnih podataka tj. 1545 je zbroj odloženih jaja u 7 cilindara, itd.

Iz dobivenih ćemo vrijednosti konstruirati tablicu analize varijance i izračunati pojedine varijance i njihove odnose.

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl}	
					P=5%	P=1%
Ukupno	27	150209.25				
Između trava	3	19532.96	6510.99	1.20	3.01	4.72
Unutar trava (pogreška)	24	130676.29	5444.84			

Pretpostavku ili nultu hipotezu o broju odloženih jaja ženki, ovisno o ishrani gusjenica, provjerit ćemo tako da najprije provedemo F-test varijance između trava. Tablični F-faktor očitat ćemo iz broja slobodnih varijanata varijance koju se testira (a to je s^2 između trava) i slobodnih varijanata varijance s kojom se testira (a to je varijanca pogreške ili ostatka).

F_{tabl} očita se dakle iz 3 i 24, što je

$F_{P=5\%} = 3.01$

$F_{P=1\%} = 4.72$

Budući da je $F_{exp} < F_{tabl}$ prihvaćamo nultu hipotezu. To znači da se ovim eksperimentom nisu utvrdile nikakve značajne, opravdane ili signifikantne razlike u broju odloženih jaja ženki koje bi uvjetovala ishrana gusjenica različitim travama. Razlike su samo slučajne.

PRIMJER 9.1.2.

U uzorcima mošta triju sorata: Traminca, Burgundca i Graševine, u laboratoriju je određena ukupna kiselost izražena u g/l (kao vinska kiselina).

Uzorci su potpuno nasumice analizirani pa tako ovaj pokus predstavlja potpuno slučajni raspored članova ili tretiranja (u ovom primjeru to su 3 sorte).

Pretpostavka ili nulta hipoteza koju se želi provjeriti jest, da nema opravdanih razlika u ukupnoj kiselosti mošta ovih triju sorata.

Radi se, dakle, o pokusu čiji su članovi bili potpuno slučajno raspoređeni, a 31 osnovni podatak (12 + 9 + 10) podvrgnut ćemo analizi varijance, tj. izračunati

		$n - 1$
SS_{Ukupno}	→	(12 + 9 + 10) - 1 = 30
$SS_{\text{Između sorata}}$	→	3 - 1 = 2
$SS_{\text{Pogreške}}$	→	30 - 2 = 28

Podaci o ukupnoj kiselosti mošta (g/l) ovih sorata (na različitom broju uzoraka) su:

Uzorak	Traminac	Burgundac	Graševina	
1	8.81	9.56	9.37	
2	9.56	10.12	10.50	
3	8.69	11.44	9.75	
4	8.06	10.43	10.87	
5	8.25	10.50	10.90	
6	9.00	10.69	9.37	
7	9.00	10.62	8.44	
8	8.25	10.31	9.00	
9	8.44	10.75	8.81	
10	8.00		10.90	
11	7.80			
12	9.00			
$\sum x$	102.86	94.42	97.91	$\sum x = 295.19$
n	12	9	10	
$\bar{x}$	8.57	10.49	9.79	

$$SS_{\text{Ukupno}} = 8.81^2 + \dots + 10.90^2 - \frac{295.19^2}{31} = 32.981$$

$$SS_{\text{Između sorata}} = \frac{102.86^2}{12} + \frac{94.92^2}{9} + \frac{97.91^2}{10} - \frac{295.19^2}{31} = 20.014$$

$$SS_{\text{Pogreške}} = 32.981 - 20.014 = 12.967$$

Konstruirat ćemo tablicu ANOVA s izračunatim vrijednostima za *SS*-ove, odrediti broj slobodnih varijanata (n - *1*) za svaki dio varijabilnosti, izračunati varijance i testirati hipotezu od koje se krenulo u pokus (da među sortama nema razlika u ukupnoj kiselosti mošta).

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	F_{exp}	F_{tabl}	
					P=5%	P=1%
Ukupno	30	32.981				
Između sorata	2	20.014	10.007	21.6**	3.34	5.45
Pogreške	28	12.967	0.463			

Tablični F-faktor očita se iz 2 (to je broj slobodnih varijanata varijance između sorata) i 28 (a to su slobodne varijante varijance unutar sorata, odnosno varijance pogreške) i iznosi

$$F_{P=5\%} = 3.34$$

$$F_{P=1\%} = 5.45$$

Kako je $F_{exp} > F_{tabl}$ za P = 0.01 (a to uključuje i P = 0.05) odbacujemo nultu hipotezu i uz 99 % sigurnosti zaključujemo da se ove tri sorte značajno razlikuju po ukupnoj kiselosti mošta. No F-test ne pokazuje koja se od koje i koliko razlikuje.

Za testiranje razlika između prosječne ukupne kiselosti mošta ovih sorata, primijenit ćemo t-test. To znači izračunati najmanju ili graničnu razliku (*LSD*) za koju se moraju razlikovati dvije prosječne vrijednosti da bi ih smatrali opravdano različitima.

Ne zaboravimo, u ovom slučaju imamo različit broj podataka za svaku sortu ($n_T = 12$, $n_B = 9$, $n_G = 10$) pa to znači da s_D treba izračunati za svaku usporedbu posebno.

$$LSD = t \cdot s_D$$

s_D ćemo izračunati, a t očitati iz tablica iz n - 1 varijance pogreške jer to je varijanca unutar grupa (vidi osnovnu ANOVA na str 92).

Dakle, tablični t faktor iz $n - 1 = 28$ je

$$t_{P=5\%} = 2.05$$

$$t_{P=1\%} = 2.76$$

Usporedit ćemo međusobno prosječnu ukupnu kiselost mošta sorata Traminac ($\bar{x}_T$), Burgundac ($\bar{x}_B$) i Graševina ($\bar{x}_G$), odnosno testirati razlike (D) među njima.

Usporedba:

$$\bar{x}_B - \bar{x}_T = 10.49 - 8.57$$

$$D = 1.92 \text{ g/l}$$

$$s_D = \sqrt{s^2 \cdot \frac{n_B + n_T}{n_B \cdot n_T}} \quad \text{(slučaj C na str. 73)}$$

$$s_D = \sqrt{0.463 \cdot \frac{9+12}{9 \cdot 12}} = \sqrt{0.463 \cdot \frac{21}{108}} = 0.3 \text{ g/l}$$

Uvrstimo li očitani tablični t faktor i izračunani s_D u formulu za najmanju značajnu razliku (*LSD*) dobijemo:

$$LSD_{P=5\%} = 2.05 \cdot 0.3 = 0.615 \text{ g/l}$$

$$LSD_{P=1\%} = 2.76 \cdot 0.3 = 0.828 \text{ g/l}$$

Kako je

$$D > LSD_{P=1\%}$$

zaključujemo: razlika u ukupnoj kiselosti mošta je opravdana ili signifikantna uz P = 1% (što dakako uključuje i P = 5%). Drugim riječima, s 99% sigurnošću možemo reći da sorta Burgundac ima signifikantno višu ukupnu kiselost mošta od sorte Traminac. Ili: razlika je visoko signifikantna, što označavamo kao

$$D = 1.92 \text{ g/l}^{**}$$

Usporedba:

$$\bar{x}_B - \bar{x}_G = 10.49 - 9.79$$

$$D = 0.70 \text{ g/l}$$

$$s_D = \sqrt{0.463 \cdot \frac{9+10}{9 \cdot 10}} = \sqrt{0.463 \cdot \frac{19}{90}} = 0.31$$

$$LSD_{P=5\%} = 2.05 \cdot 0.31 = 0.635 \text{ g/l}$$

$$LSD_{P=1\%} = 2.76 \cdot 0.31 = 0.856 \text{ g/l}$$

$$D > LSD_{P=5\%}$$

Zaključujemo: razlika je signifikantna uz P = 5%, što znači da s 95% sigurnošću (koja je, kako smo već spominjali, u našoj struci dogovoreno dostatna) možemo tvrditi da Burgundac ima višu ukupnu kiselost mošta i od sorte Graševine. Zato ovu razliku označimo kao

$$D = 0.70 \text{ g/l}^{*}$$

Usporedba:

$$\bar{x}_G - \bar{x}_T = 9.79 - 8.57$$

$$D = 1.22 \text{ g/l}$$

$$s_D = \sqrt{0.463 \cdot \frac{10+12}{10 \cdot 12}} = \sqrt{0.463 \cdot \frac{22}{120}} = 0.29$$

$$LSD_{P=5\%} = 2.05 \cdot 0.29 = 0.59 \text{ g/l}$$

$$LSD_{P=1\%} = 2.76 \cdot 0.29 = 0.80 \text{ g/l}$$

$$D > LSD_{1=1\%}$$

Signifikantna razlika D = 1.22 g/l** dozvoljava ovaj zaključak: mošt Graševine ima značajno višu ukupnu kiselost od mošta Traminca.

Dakle, analizom podataka ovog laboratorijskog pokusa najprije smo odbacili nultu hipotezu kojom se pretpostavljalo da se sorte ne razlikuju po ukupnoj kiselosti mošta (za što nam je argument opravdani F-test varijance između sorata). Usporedbom prosječnih vrijednosti sorata primjenom t-testa utvrdili smo da opravdano najvišu ukupnu kiselost mošta ima Burgundac, od njega značajno nižu Graševina i Traminac, ali Graševina signifikantno višu od Traminca. Ili,

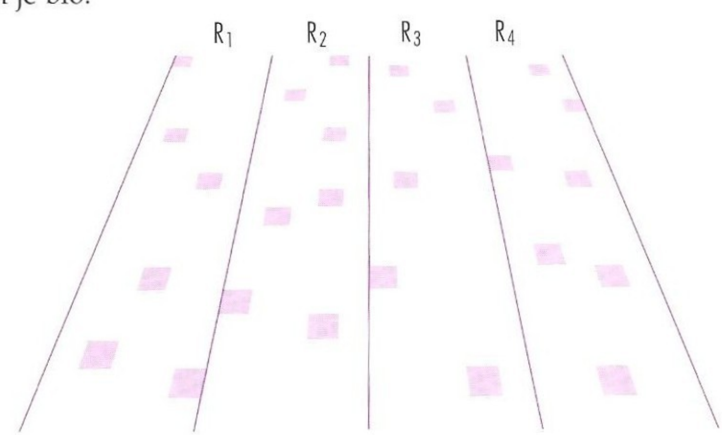
$\bar{x}_B$		$\bar{x}_G$		$\bar{x}_T$
10.49		9.79		8.57
	0.70*		1.22**	
		1.92**		

PRIMJER 9.1.3.

Cilj ovog eksperimenta bio je provjeriti pretpostavku da rok sjetve neće značajno utjecati na prinos šećerne repe.

Pokus je izveden na proizvodnim površinama gdje se u svakom od 4 uzastopna roka (R_1 = rana sjetva, R_2 = srednje rana, R_3 = kasna, R_4 = vrlo kasna) zasijala površina od 500 m^2. Pokus, dakle, ima 4 člana. U vrijeme vađenja repe, izvagani su prinosi s potpuno nasumičnih parcela (veličine 5,0 m^2), s površina sijanih u različitim rokovima. Tako je dobiven različit broj podataka po rokovima (6 u R_1, 7 u R_2, 5 u R_3, 7 u R_4) s ukupno 25 osnovnih parcela.

Plan je bio:



Prikupljene podatke o prinosu (x_1, ..., x_{25}) analizirat ćemo pomoću analize varijance, vodeći računa da članovi pokusa (rokovi sjetve) imaju različiti n.

Podaci o prinosu (kg/5.0 m^2) u različitim rokovima sjetve su:

Parcelice	R_1	R_2	R_3	R_4	
1	22.3	18.3	17.2	14.9	
2	21.8	18.4	17.2	12.6	
3	19.7	18.5	17.9	13.1	
4	21.2	21.5	18.8	14.4	
5	20.0	17.3	16.7	12.4	
6	22.2	18.0		12.2	
7		18.6		12.0	
$\sum x_{Roka}$	127.2	130.6	87.8	91.6	$\sum x = 437.2$
n	6	7	5	7	$n = 25$
$\bar{x}_{Roka}$	21.20	18.66	17.56	13.08	

Procijeniti će se sljedeći dijelovi varijabilnosti, uz pripadajuće slobodne varijante:

		n - 1
SS_{Ukupno}	$\rightarrow$	(6 + 7 + 5 + 7) - 1 = 24
$SS_{\text{Između rokova}}$	$\rightarrow$	4 - 1 = 3
$SS_{\text{Pogreške}}$	$\rightarrow$	24 - 3 = 21

ANOVA:

$$SS_{\text{Ukupno}} = 22.3^2 + \ldots + 12.0^2 - \frac{437.2^2}{25} = 255.106$$

$$SS_{\text{Između rokova}} = \frac{127.2^2}{6} + \frac{130.6^2}{7} + \frac{67.8^2}{5} + \frac{91.6}{7} - \frac{437.2^2}{25}$$

$$= 227.929$$

$$SS_{\text{Pogreške}} = 255.106 - 227.929 = 27.177$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl}	
					P=5%	P=1%
Ukupno	24	255.106				
Između rokova	3	227.929	75.976	58.71**	3.07	4.87
Pogreška	21	27.177	1.294			

Signifikantan F-test (jer je $F_{exp} > F_{tabl}$) znači da odbacujemo nultu hipotezu da rok sjetve nema značajnog utjecaja na prinos šećerne repe.

Zaključujemo, naprotiv, da su prinosi značajno različiti ovisno o tome kad je obavljena sjetva. Da bismo došli do konačnog odgovora na pitanje: kada sijati i koliko je eventualno moguće kasniti sa sjetvom, usporedit ćemo prosječne prinose postignute na površi-nama sijanim u ova četiri roka.

To znači provest ćemo t-test.

Tablični t-faktor, očitan iz n - 1 = 21, iznosi

$$t_{P=5\%} = 2.07$$

$$t_{P=1\%} = 2.83$$

a s obzirom na različiti n u svakom roku sjetve, opet će trebati izračunati s_D za svaku usporedbu posebno.

Iz prosječnih vrijednosti, čini se da se najveći prinos šećerne repe postiže u prvom roku sjetve, a sa svakim daljnjim rokom je manji. Da li je to smanjenje prinosa iz roka u rok značajno ili nije, utvrdit ćemo njihovim usporedbama.

Usporedba:

$$\overline{x}_{R_1} - \overline{x}_{R_2} = 21.20 - 18.66$$

$$D = 2.54 \text{ kg}/5.0 \text{ m}^2$$

$$s_D = \sqrt{1.294 \cdot \frac{6+7}{6 \cdot 7}} = \sqrt{1.294 \cdot \frac{13}{42}} = 0.63 \text{ kg}$$

$$LSD_{P=5\%} = 2.08 \cdot 0.63 = 3.31 \text{ kg}/5.0 \text{ m}^2$$

$$LSD_{P=1\%} = 2.83 \cdot 0.63 = 1.78 \text{ kg}/5.0 \text{ m}^2$$

Razlika je veća od granične, dakle visoko signifikantna. To znači da se prinos značajno smanji ako se sije u nešto kasnijem roku (tj. R_2), ili

$$D = 2.54 \text{ kg}/5.0 \text{ m}^2 \, ^{**}$$

Iz prosječnih vrijednosti je vidljivo da prinos pada što se kasnije sije ($\bar{x}_{R_1} > \bar{x}_{R_2} > \bar{x}_{R_3} > \bar{x}_{R_4}$), no ipak ćemo napraviti i druge usporedbe.

Usporedba:

$$\bar{x}_{R_1} - \bar{x}_{R_3} = 21.20 - 17.56$$

$$D = 3.64 \text{ kg}/5.0 \text{ m}^2$$

$$s_D = \sqrt{1.294 \cdot \frac{6+5}{6 \cdot 5}} = \sqrt{1.294 \cdot \frac{11}{30}} = 0.69$$

$$LSD_{P=5\%} = 2.08 \cdot 0.69 = 1.43 \text{ kg}$$

$$LSD_{P=1\%} = 2.83 \cdot 0.69 = 1.95 \text{ kg}$$

$$D > LSD_{P=1\%} \quad \text{dakle} \quad D = 3.64 \text{ kg}/5.0 \text{ m}^{2**}$$

Usporedba:

$$\bar{x}_{R_1} - \bar{x}_{R_4} = 21.20 - 13.08$$

$$D = 8.12 \text{ kg}/5.0 \text{ m}^2$$

$LSD_{P=5\%} = 1.31$ kg (*LSD* je isti kao kod usporedbe

$LSD_{P=1\%} = 1.78$ kg R_1 i R_2, jer $n_{R_1} = 6$, $n_{R_4} = 7$)

$$D > LSD_{P=1\%} \quad \text{dakle} \quad D = 8.12 \text{ kg}/5.0 \text{ m}^2 \, ^{**}$$

Testirajući razlike između prinosa postignutog u najranijem roku sjetve (R_1) i onih u kasnijim rokovima možemo konačno zaključiti: prinosi šećerne repe se signifikantno smanjuju ako se sjetvom kasni.

No često se u praksi dogodi da se zbog vremenskih uvjeta sjetva ipak ne može obaviti u ovom prvom roku kad se postižu najveći prinosi.

Pa da vidimo, ako se već kasni, koliko se sjetva smije odgoditi.

Usporedba:

$$\bar{x}_{R_2} - \bar{x}_{R_3} = 18.66 - 17.56$$

$$D = 1.10 \text{ kg}/5.0 \text{ m}^2$$

$$s_D = \sqrt{1.294 \cdot \frac{7+5}{7 \cdot 5}} = \sqrt{1.294 \cdot \frac{12}{35}} = 0.67 \text{ kg}$$

$$LSD_{P=5\%} = 2.08 \cdot 0.67 = 1.39 \text{ kg/5.0 m}^2$$

$$LSD_{P=1\%} = 2.83 \cdot 0.67 = 1.90 \text{ kg/5.0 m}^2$$

$$D < LSD_{P=1\%} \quad \text{dakle} \quad D = 1.10 \text{ kg/5.0 m}^2 \text{ ns}$$

Razlika u prinosu s parcela sijanih u roku R_2 i prinosa postignutog u još kasnijoj sjetvi R_3 manja je od najmanje značajne razlike, a to znači da nije opravdana, nije signifikantna (što smo označili s *n.s.*). Praktično, to znači da ako se i zakasni sa sjetvom, može se sijati i u trećem roku, jer u odnosu na sjetvu u roku R_2 prinos se neće značajno smanjiti.

Usporedba:

$$\bar{x}_{R_2} - \bar{x}_{R_4} = 18.66 - 13.08$$

$$D = 5.58 \text{ kg/5.0 m}^2$$

$$s_D = \sqrt{\frac{2 \cdot s^2}{n}}$$

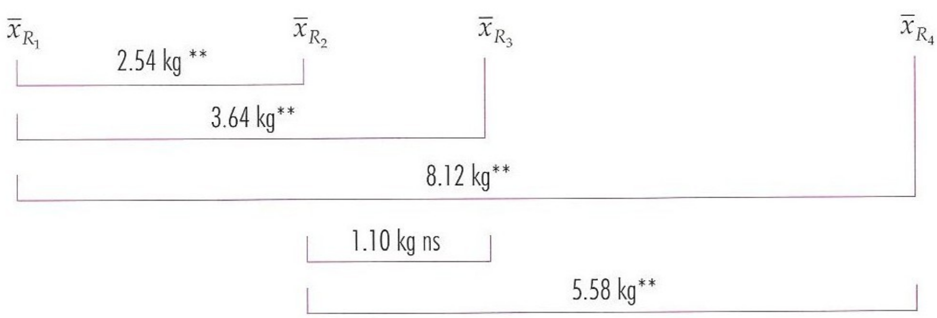
$$s_D = \sqrt{\frac{2 \cdot 1.294}{7}} = 0.61 \text{ kg}$$

$$LSD_{P=5\%} = 2.08 \cdot 0.61 = 1.27 \text{ kg/5.0 m}^2$$

$$LSD_{P=1\%} = 2.83 \cdot 0.61 = 1.73 \text{ kg/5.0 m}^2$$

$$D > LSD_{P=1\%} \quad \text{dakle} \quad D = 5.58 \text{ kg/5.0 m}^2 \text{ **}$$

Ova visoko signifikantna razlika (uz P = 1%) znači da se još kasnijom sjetvom (u vrlo kasnom roku, R_4) prinos značajno smanji, pa ako se već sjetva mora odgoditi, to je dopušteno samo do trećeg roka.



Sve su ovo bili primjeri pokusa izvedenih po najjednostavnijoj mogućoj shemi - potpuno slučajnom rasporedu tretiranja ili članova pokusa, koja ima više nedostataka nego prednosti.

Ako se o prednostima uopće može govoriti, one su u jednostavnosti - kako u izvođenju eksperimenta tako i u analizi eksperimentalnih podataka (pri čemu svi članovi i ne moraju imati isti broj podataka).

Ovaj način analize pokusnih podataka može se vrlo uspješno i učinkovito primijeniti u eksperimentiranju na proizvodnim površinama. Potrebno je samo prikupiti što više podataka o svakom tretiranju (potpuno nasumce sa cijele površine).

Posebno treba naglasiti: podaci pokusa izvedenog po potpuno slučajnom rasporedu ne bi se niti smjeli analizirati (niti F-test provoditi) ako na varijancu pogreške pokusa otpada *n* - 1 koji je manji od 6.

Jednako tako je važno znati da nesignifikantni F-test na kraju ANOVA ukazuje na nemogućnost da se eksperimentom otkriju razlike među članovima.

Razlogom tome mogu biti ili izuzetno male razlike među tretiranjima, ili vrlo velika pogreška pokusa, ili oboje.

Zato, kad god je F-test neopravdan, treba dobro preispitati veličinu pogreške pokusa i razlike među tretiranjima. Ako su obje velike, trebalo bi pokus ponoviti i poduzeti sve da se pogreška pokusa smanji kako bi se, ako ih ima, otkrile razlike među tretiranjima. Suprotno tome, ako su obje vrijednosti male, razlike među tretiranjima su vjerojatno premale da bi se o njima uopće trebalo voditi računa. U takvom slučaju ne treba pokus ponavljati.

U našim primjerima, gdje su F_{exp} faktori bili mnogo veći od tabličnih F-faktora čak uz P = 0.01 (F_{exp} 58.71 > F_{tabl} 4.87 i F_{exp} 21.61 > F_{tabl} 5.45) vjerojatnost da bi razlike među članovima pokusa bile samo slučajne manja je od 1 na 100 slučajeva. Dakle, sigurnost zaključaka o onim razlikama koje smo u primjerima testirali je 99%.

Treba se podsjetiti i da je u ovim primjerima primijenjena takva analiza kojom je iz ukupne varijabilnosti bilo moguće izdvojiti samo onaj dio koji je uzrokovan različitošću među članovima pokusa. Sav ostali nedefinirani dio varijabilnosti ulazi u ostatak ili pogrešku pokusa. Stoga se ovakvu analizu varijance obično zove **jednosmjerna analiza varijance (jednosmjerna ANOVA)**.

■

9.2. Slučajni blokni raspored

To je vjerojatno jedna od najčešće korištenih shema pokusa u poljoprivrednim istraživanjima.

U njoj su članovi pokusa slučajno raspoređeni i ponovljeni tako da tvore repeticije ili blokove.

Prisjetimo se: repeticija je onaj dio pokusa koji sadrži randomizirana sva tretiranja pokusa. To znači da je broj osnovnih pokusnih jedinica u svakoj repeticiji ili bloku jednak broju članova, a svaki se član u repeticiji pojavljuje samo jedanput.

Više repeticija može činiti blok, no najčešće je ipak repeticija isto što i blok, pa otud i naziv ovog plana slučajni blokni raspored.

Opetujmo: u eksperimentiranju je najvažnije reducirati pogrešku pokusa, eliminiranjem udjela poznatih uzroka varijabilnosti među eksperimentalnim jedinicama (osnovnim parcelicama). To se postiže upravo grupiranjem članova u repeticije ili blokove, unutar kojih uvjeti moraju biti maksimalno ujednačeni. Svrha je minimalizirati varijabilnost među parcelicama unutar repeticije, a maksimalizirati varijabilnost između repeticija. Na taj način u pogrešci pokusa ostaje samo varijabilnost unutar repeticija, pa je tako pogreška manja, a rezultati pouzdaniji.

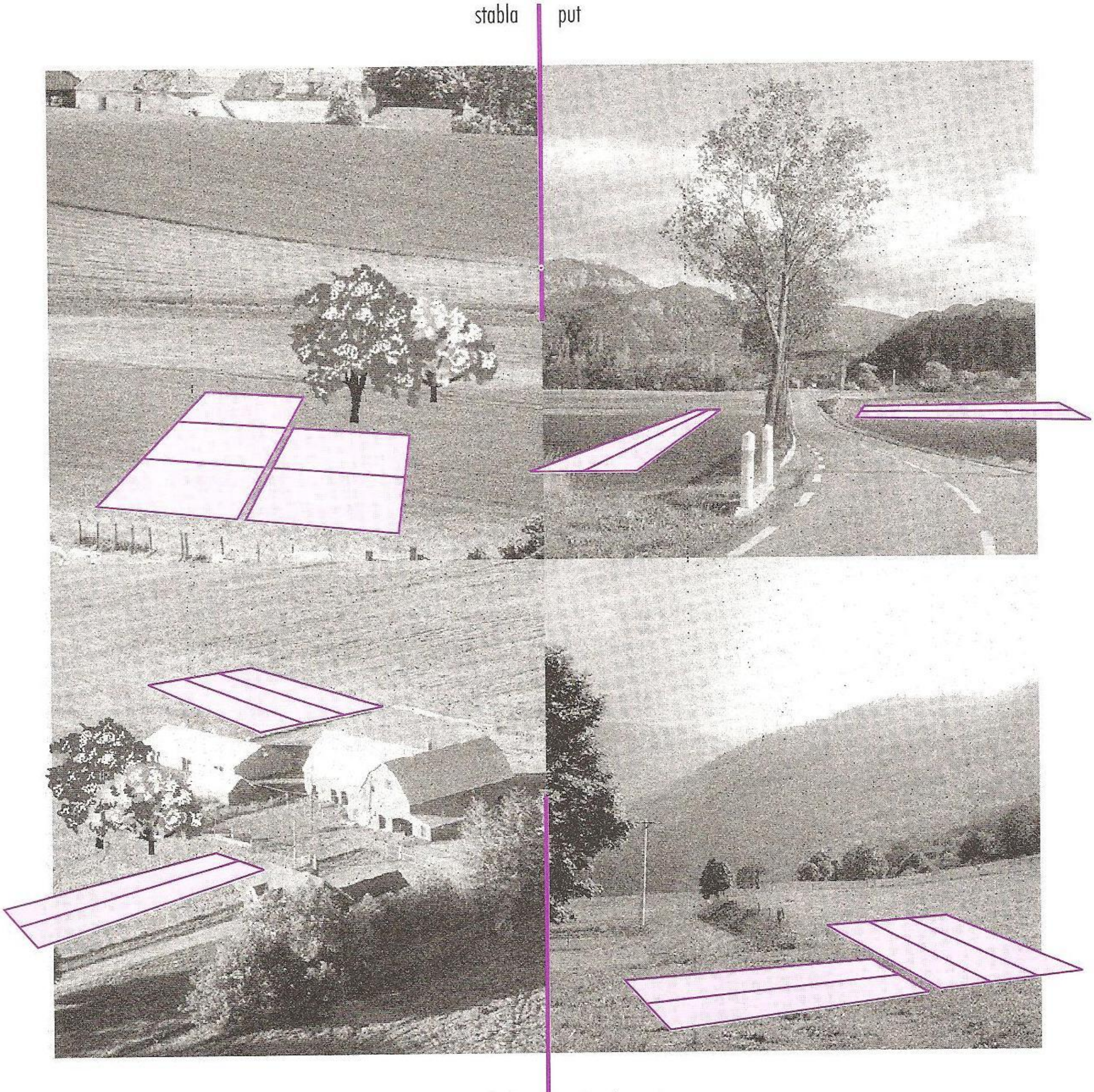
Postizanje ujednačenosti uvjeta unutar repeticija odnosi se kako na razmještaj ponavljanja u pokusu, tako i na sve zahvate koji se tijekom pokusa izvode.

Često su razlogom za posebnu pažnju pri razmještanju ponavljanja u pokusu i heterogenost tla, i smjer gradijenta plodnosti, i nagib terena, i smjer migracije insekata i dr.

Zato se blokovi ili ponavljanja mogu različito razmještati. Bitno je osigurati najveću moguću ujednačenost uvjeta za sve članove unutar ponavljanja, a različitost između ponavljanja.

Da bi se to postiglo, nerijetko repeticije mogu biti čak na određenoj prostornoj udaljenosti, a ponekad i nepravilnog oblika.

Na slici 31 je nekoliko ilustracija:



SLIKA 31.
Mogući razmještaji repeticija u pokusu

Kad su ponavljanja formirana, dobro se tijekom cijelog eksperimentiranja prema njima odnositi kao prema cjelinama.

Najbolje je da se sjetva, sadnja, obrada, različiti zahvati, uzimanje uzoraka, berba, žetva i dr. za cijeli pokus izvedu istovremeno.

Ako se svi spomenuti i drugi zahvati i ne mogu završiti isti dan za cijeli pokus, započeta repeticija se mora dovršiti.

Jednako tako, ako te operacije ili zahvate obavlja više osoba, svakog bi trebalo zadužiti za jedno ponavljanje (kako bi način i kriterij za sve članove u ponavljanju bili isti).

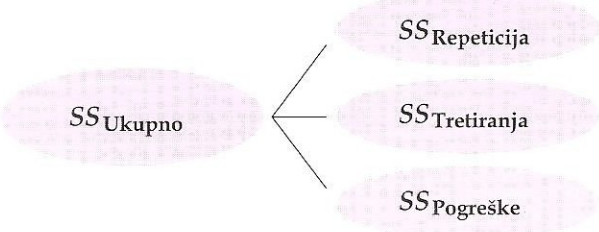
Ponovimo: ključ uspjeha plana pokusa je minimalizirati varija-bilnost unutar repeticija, a maksimizirati varijabilnost između repeticija.

Slučajni blokni raspored predstavlja fleksibilnu shemu koja nema posebnih restrikcija osim što svako tretiranje treba ponavljati isti broj puta. Nije čak niti nužno, iako je poželjno, da sva ponavljanja budu istog oblika.

Analiza pokusnih podataka provodi se primjenom takve analize, kojom je moguće iz ukupne varijabilnosti izdvojiti još i dio varijabilnosti uvjetovan različitošću između ponavljanja, što nije bilo moguće kod prve opisane sheme - potpuno slučajnog rasporeda.

Naime, u analizi varijance koju smo tamo primijenili, iz ukupne varijance izdvojili smo samo varijabilnost između tretiranja, a sav preostali dio je predstavljao pogrešku pokusa. Zato se takva analiza naziva **jednosmjerna ANOVA**.

Za razliku od nje, plan slučajnog bloknog rasporeda omogućuje **dvosmjernu ANOVA** (iz ukupne varijance izdvaja se varijanca među tretiranjima, varijanca među ponavljanjima, a onaj nedefi-nirani dio je ostatak ili pogreška pokusa).



Slobodne varijante za svaki izvor varijabilnosti:

		n - 1
SS_{Ukupno}	$\rightarrow$	$n - 1$
$SS_{Repeticija}$	$\rightarrow$	$n_r - 1$
$SS_{Tretiranja}$	$\rightarrow$	$n_T - 1$
$SS_{Pogreške}$	$\rightarrow$	$(n_r - 1)(n_T - 1)$

SS pogreške pokusa predstavlja zapravo varijancu unutar eksperimentalnih jedinica, dobivenu izdvajanjem dijelova uvjetovanih različitošću repeticija i različitošću tretiranja iz ukupne varijabilnosti.

Zato je ona kod slučajnog bloknog rasporeda manja u odnosu na potpuno slučajni raspored.

Još jedna prednost ovog plana pokusa je u tome, što se cijelo tretiranje ili cijela repeticija mogu isključiti iz analize ako se iz nekih razloga "izgube" informacije (propale biljke, uništeni dijelovi pokusa i sl.).

PRIMJER 9.2.1.

U Poreštini je 1980. godine podignut eksperimentalni nasad s sortimentom ranih jabuka (S_1=Quinte, S_2=Vista Bella, S_3=Manet, S_4=Red Melba, S_5=Julyred, S_6=Jersey, S_7=Scotia, S_8=James Grieve Lired, S_9=Akane, S_{10}=Alkmene, S_{11}=Discovery, S_{12}=Early Blaze).

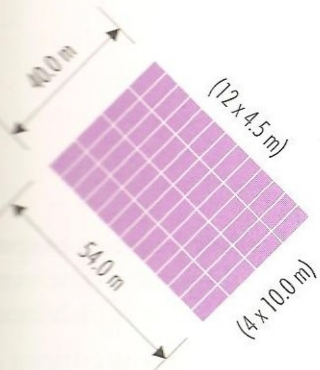
Pokus je postavljen po planu slučajnog bloknog rasporeda u 4 repeticije.

Svaka sorta je u svakoj repeticiji bila zastupljena s 5 stabala (osnovna parcela), pa je tako pokusom bilo obuhvaćeno ukupno 240 stabala (12 sorata x 5 stabala x 4 repeticije).

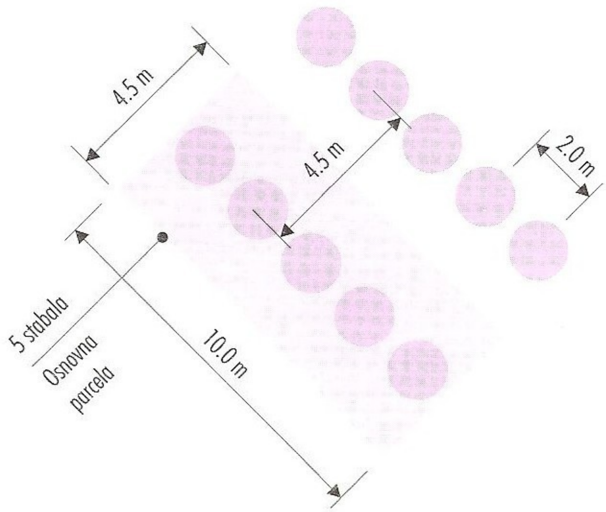
Razmak sadnje je bio 4.5 m (između redova) x 2.0 m (unutar redova), što znači da je osnovna parcela bila veličine 45,0 m^2, a veličina pokusa 54.0 m (ili 12 sorata x 4.5 m) x 40.0 m (ili 5 stabala x 2.0 m x 4 repeticije) = 2160.0 m^2.

Proučavana su brojna morfološka i gospodarska svojstva. Za ilustraciju analize podataka ovako izvedenog pokusa, uzet ćemo jedno od njih: promjer debla mjeren u petoj vegetacijskoj godini (na srednjoj voćki na parceli, na 30 cm iznad mjesta cijepljenja sorte na podlogu MM 106).

Pretpostavka (nulta hipoteza) je bila da se petogodišnja stabla ovih sorata neće razlikovati u promjeru debla.



Veličina pokusa 2160.0 m^2



Plan pokusa i podaci o promjeru debla (cm):

I	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
	6.2	7.9	6.6	5.9	6.9	6.7	6.3	6.3	7.1	6.1	7.3	5.2
II	S_{12}	S_{10}	S_1	S_5	S_7	S_9	S_3	S_{11}	S_8	S_4	S_6	S_2
	6.1	6.7	6.6	6.5	6.1	6.5	6.0	6.0	6.6	6.5	7.3	8.1
III	S_5	S_8	S_4	S_9	S_6	S_1	S_{10}	S_2	S_7	S_{12}	S_3	S_{11}
	7.3	7.0	5.7	7.0	7.3	6.3	6.4	8.9	7.1	5.2	6.8	5.7
IV	S_7	S_{11}	S_8	S_2	S_{12}	S_3	S_6	S_4	S_{10}	S_1	S_9	S_5
	6.9	6.1	7.0	8.5	5.7	6.8	7.0	6.2	6.9	8.0	7.9	8.0

Kako bismo provjerili nultu hipotezu, provest ćemo analizu ovih pokusnih podataka, primjenom dvosmjerne analize varijance. Definirat ćemo ove dijelove varijabilnosti i pripadajuće slobodne varijante:

		$n - 1$
SS_{Ukupno}	$\rightarrow$	$(12 \cdot 4) - 1 = 27$
$SS_{\text{Repeticija}}$	$\rightarrow$	$4 - 1 = 3$
SS_{Sorata}	$\rightarrow$	$12 - 1 = 11$
$SS_{\text{Pogreške}}$	$\rightarrow$	$47 - 3 - 11 = 33$

Podaci svrstani po sortama i ponavljanjima

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	$\sum x_{\text{Repeticije}}$
I	6.2	7.9	6.6	5.9	6.9	6.7	6.3	6.3	7.1	6.1	7.3	5.2	78.5
II	6.6	8.1	6.0	6.5	6.5	7.3	6.1	6.6	6.5	6.5	7.3	8.1	82.1
III	6.3	8.9	6.8	5.7	7.3	7.3	7.1	7.0	7.0	5.2	6.8	5.7	81.1
IV	8.0	8.5	6.8	6.2	8.0	7.0	6.9	7.0	7.9	8.0	7.9	8.0	90.2
$\sum x_{\text{Sorte}}$	27.1	33.4	26.2	24.3	28.7	28.3	26.4	26.9	28.5	25.8	29.3	27.0	$\sum x$ =331.9
$\bar{x}_{\text{Sorte}}$	8.77	8.35	6.55	6.07	7.17	7.07	6.60	6.72	7.12	6.45	7.32	6.75	

$$SS_{\text{Ukupno}} = 6.2^2 + ... + 8.0^2 - \frac{331.9^2}{48} = 32.98$$

$$SS_{\text{Repeticija}} = \frac{78.5^2 + ... + 90.2^2}{12} - \frac{331.9^2}{48} = 6.37$$

$$SS_{\text{Sorata}} = \frac{27.1^2 + ... + 27.0^2}{4} - \frac{331.9^2}{48} = 14.41$$

$$SS_{\text{Pogreške}} = SS_{\text{Ukupno}} - SS_{\text{Repeticija}} - SS_{\text{Sorata}}$$

$$= 32.98 - 6.37 - 14.41 = 12.20$$

Kod računanja $SS_{\text{Repeticija}}$ dijeli se s 12 jer je svaki podatak u brojniku suma svih 12 sorata u pojedinoj repeticiji. Jednako tako kod SS_{Sorata} svaki je podatak suma jedne sorte u sve 4 repeticije.

Iz tih izračunanih suma kvadratnih odstupanja (SS-ova) i pripadajućih slobodnih varijanata (n - 1) izračunaju se varijance (s^2).

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno	47	32.98				
Repeticije	3	6.37				
Sorte	11	14.41	1.309	3.55**	2.16	2.98
Pogreška	33	12.20	0.369			

Tablični F faktor očita se iz slobodnih varijanata varijance u brojniku (s^2 sorata) tj. n - 1 = 11 i varijance u nazivniku (s^2 pogreške) tj. n - 1 = 33. S obzirom da u tablici nema baš tih vrijednosti, očitali smo za n - 1 = 10 i n - 1 = 30 (što je nešto strožije).

F-test je visoko signifikantan ($F_{exp} > F_{tabl\ P=0.01}$), što znači da odbacujemo nultu hipotezu i zaključujemo da se promjeri debla petogodišnjih stabala ovih ranih sorata jabuka međusobno značajno razlikuju. No, u pokusu je bilo 12 sorata, pa signifikantan F-test ne može dati konačan odgovor: koja se od koje sorte razlikuje, a koja eventualno ne.

Naime, opravdan F-test samo znači da signifikantnih razlika ima. To nikako ne znači da su sve razlike opravdane. Jer, moguće su brojne usporedbe: sorta najvećeg promjera debla u odnosu na ostale, svaka sorta sa svakom i sl.

Zato nakon opravdanog F-testa treba provesti t-test za uspoređivanje prosječnih vrijednosti.

Kriterij je najmanja značajna razlika (*LSD*)

$$LSD = t \cdot s_D$$

$$s_D = \sqrt{\frac{2 \cdot 0.369}{4}} = 0.429 \text{ cm}$$

t_{tabl} očitan za n - 1 = 30

$$t_{P=0.05} = 2.042$$

$$t_{P=0.01} = 2.750$$

$$LSD_{P=5\%} = 2.042 \cdot 0.429 = 0.876 \text{ cm}$$

$$LSD_{P=1\%} = 2.750 \cdot 0.429 = 1.180 \text{ cm}$$

Znači: ako se promjer debla bilo kojih dviju sorata razlikuje bar za 0.876 cm, možemo ih smatrati signifikantno različitima u tom svojstvu (odnosno visokosignifikantno ako je razlika veća od 1.18 cm).

Na narednoj tablici su prikazane sve moguće razlike, s oznakom razine signifikantnosti.

Značajno je najveći promjer debla petogodišnjih stabala sorte Vista Bella (S_2). Sve razlike u odnosu na ostale sorte su, naime, signifikantne (*) ili visokosignifikantne (**).

Ostale sorte se međusobno ne razlikuju signifikantno, osim što je promjer debla sorata S_{11} , S_5 , S_9 i S_6 značajno veći od sorte najtanjeg debla, sorte Red Melba (S_4).

		S_{11} 7.32	S_5 7.17	S_9 7.12	S_6 7.07	S_1 6.77	S_{12} 6.75	S_8 6.72	S_7 6.60	S_3 6.55	S_{10} 6.45	S_4 6.07
S_2	8.35	1.03*	1.18*	1.23**	1.28**	1.58**	1.60**	1.63**	1.75**	1.80**	1.90**	2.28**
S_{11}	7.32		0.15	0.20	0.25	0.55	0.57	0.60	0.72	0.77	0.87	1.25**
S_5	7.17			0.05	0.10	0.40	0.42	0.45	0.57	0.62	0.72	1.10*
S_9	7.12				0.05	0.35	0.37	0.40	0.52	0.57	0.67	1.05*
S_6	7.07					0.30	0.32	0.35	0.47	0.52	0.62	1.00*
S_1	6.77						0.02	0.05	0.17	0.22	0.32	0.70
S_{12}	6.75							0.03	0.15	0.20	0.30	0.68
S_8	6.72								0.12	0.17	0.27	0.65
S_7	6.60									0.05	0.15	0.53
S_3	6.55										0.10	0.48
S_{10}	6.45											0.38

Opravdanosti razlika među prosječnim vrijednostima mogu se prikazati i tako, da se sve prosječne vrijednosti koje nisu signifikantno različite označe istim slovom.

U ovom primjeru, to je:

S_2	8.35	A			A		
S_{11}	7.32	B			A		
S_5	7.17	B	C		A	B	C
S_9	7.12	B	C			B	C
S_6	7.07	B	C			B	C
S_1	6.77	B	C	D		B	C
S_{12}	6.75	B	C	D		B	C
S_8	6.72	B	C	D		B	C
S_7	6.60	B	C	D		B	C
S_3	6.55	B	C	D		B	C
S_{10}	6.45		C	D		B	C
S_4	6.07			D			C
		uz P=5%			uz P=1%		

PRIMJER 9.2.2.

U problematici suzbijanja krumpirove zlatice, proučavala se učinkovitost 8 različitih pripravaka (P_2, P_3, ... , P_9), a uključio se i standardni insekticid (P_1).

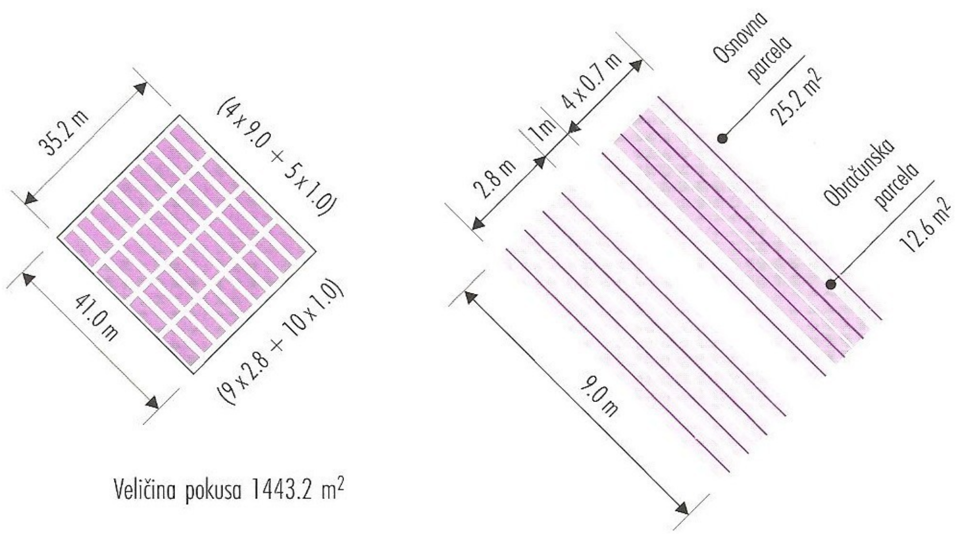
Pretpostavka ili hipoteza je bila, da se učinkovitost ovih pripravaka neće značajno razlikovati u odnosu na standardni, a niti međusobno.

Proveden je poljski pokus u kojem su se sa svih 9 pripravaka tretirale ličinke krumpirove zlatice (I. i II. stadija). Primijenjena je shema slučajnog bloknog rasporeda s 4 ponavljanja.

Prije tretiranja izbrojene su ličinke na svakoj osnovnoj parceli, a isto tako 3 dana nakon tretiranja, pa se tako mogla izračunati učinkovitost svakog od ovih 9 pripravaka (izražena u %).

Osnovna parcela je bila veličine 25.2 m^2 (tj. 4 reda duljine 9.0 m s razmakom od po 0.7 m ili 2.8 m x 9.0 m = 25.2 m^2).

Obračunsku parcelu predstavljala su dva unutarnja reda, a to je 12.6m^2 (1.4 m x 9.0 m).



Razmak između osnovnih parcela u pokusu i između repeticija bio je po 1.0 m.

Veličina pokusa bila je 1443.2 m^2 tj. 41.0 m (4 m x 9 m plus 5 m staza) 35.2 m (9 m x 2.8 m plus 10.0 m staza)

Plan pokusa i podaci o postotku učinkovitosti svakog pripravka u svim ponavljanjima:

I	P_4	P_5	P_3	P_7	P_8	P_2	P_6	P_9	P_1
	65.02	91.92	97.24	66.24	91.56	85.61	49.38	76.98	75.17
II	P_3	P_2	P_8	P_6	P_1	P_9	P_7	P_4	P_5
	92.94	92.69	98.05	88.34	88.62	95.76	47.81	76.36	95.34
III	P_8	P_1	P_9	P_2	P_4	P_3	P_5	P_7	P_6
	99.37	60.13	84.65	66.24	64.11	84.79	82.56	20.41	75.23
IV	P_5	P_6	P_7	P_8	P_9	P_1	P_3	P_2	P_4
	83.56	84.52	67.82	86.26	78.85	76.71	91.56	92.96	83.60

Podaci svrstani po pripravcima:

Repeticija	Pripravci									
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	$\sum x_{Rep}$
I	75.17	85.61	97.24	65.02	91.92	49.38	66.24	91.56	76.98	699.12
II	88.62	92.69	92.94	76.36	95.34	88.34	47.81	98.05	95.76	775.91
III	60.13	66.24	84.79	64.11	82.56	75.23	20.41	99.37	84.65	637.49
IV	76.71	92.96	91.56	83.60	83.56	84.52	67.82	86.26	78.85	745.84
$\sum x_{Pripravka}$	300.63	337.50	366.53	289.09	352.38	297.47	202.28	375.24	336.24	$\sum x = 2858.36$
$\bar{x}_{Pripravka}$	75.16	84.37	91.63	72.27	88.34	74.37	50.57	93.81	84.06	

ANOVA:

$$SS_{Ukupno} = 75.17^2 + \ldots + 78.85^2 - \frac{2858.36^2}{36} = 9690.485$$

$$SS_{Repeticija} = \frac{699.12^2 + \ldots + 745.84^2}{9} - \frac{2858.36^2}{36} = 1213.382$$

$$SS_{Pripravaka} = \frac{300.63^2 + \ldots + 336.24^2}{4} - \frac{2858.36^2}{36} = 5636.226$$

$$SS_{Pogreške} = 9690.485 - 1213.382 - 5636.226 = 2840.877$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno	35	9690.485				
Repeticije	3	1213.382				
Pripravci	8	5636.226	704.528	5.95**	2.36	3.36
Pogreška	24	2840.877	118.369			

Visoko signifikantan F-test varijance pripravaka upućuje na postojanje značajnih razlika u učinkovitosti ovih pripravaka, odnosno na odbacivanje nulte hipoteze.

Međutim, treba opet naglasiti da opravdani F-test nikako ne znači da sve moguće razlike moraju biti opravdane (iako mogu), ali on je znak da sigurno bar neka(e) je(su). Stoga treba provesti t-test i utvrditi opravdanost.

$$LSD = t \cdot s_D$$

$$s_D = \sqrt{\frac{2 \cdot 118.369}{4}} = 7.693\%$$

Tablični t-faktor, očitan iz n - 1 pogreške (24) je

$$t_{P=0.05} = 2.06$$

$$t_{P=0.01} = 2.80$$

a najmanja signifikantna razlika je

$$LSD_{P=5\%} = 2.06 \cdot 7.693 = 15.85\ \%$$

$$LSD_{P=1\%} = 2.80 \cdot 7.693 = 21.54\ \%$$

Dakle, pripravak čija je efikasnost bar za 15.85 % (ili 21.54 %) veća od učinkovitosti standardnog, značajno je učinkovitiji u suzbijanju ovog štetnika uz P = 5 % (ili P = 1 %).

Pa usporedimo li sve sa standardnim:

Apsolutne razlike | D |,%

75.16 (P_1) - 84.37 (P_2) = 9.21 ns

75.16 (P_1) - 91.63 (P_3) = 16.47 *

75.16 (P_1) - 72.27 (P_4) = 2.89 ns

75.16 (P_1) - 88.34 (P_5) = 13.18 ns

$75.16\ (P_1) - 74.37\ (P_6) = 0.79$ ns

$75.16\ (P_1) - 50.57\ (P_7) = 24.59$ **

$75.16\ (P_1) - 93.81\ (P_8) = 18.65$ *

$75.16\ (P_1) - 84.06\ (P_9) = 8.90$ ns

Iz ovog možemo zaključiti:

- efikasnost pripravaka označenih s P_2, P_4, P_5, P_6 i P_9 podjednaka je efikasnosti standardnog pripravka P1 (jer razlike nisu opravdane)
- pripravak s oznakom P_7 značajno je manje učinkovit od standardnog pripravka u suzbijanju ovog štetnika
- pripravci P_3 i P_8 signifikantno su učinkovitiji od standardnog pripravka P_1, pa bi se mogli preporučiti kao efikasniji u suzbijanju krumpirove zlatice (i to podjednako, jer se po učinkovitosti međusobno ne razlikuju značajno; $P_8 - P_3 = 2.18$).

■

9.3. Latinski kvadrat

To je shema pokusa u kojoj se tretiranja raspoređuju tako da se istovremeno randomiziraju u dva smjera: redove i stupce. Pri tome se svako tretiranje i u svakom redu i u svakom stupcu javlja samo jedanput.

Na taj se način kreiraju dvovrsne ili dvosmjerne repeticije: redovi su "vodoravne" repeticije, a stupci predstavljaju "okomite" repeticije.

Svaki red ili svaki stupac je cjelovito ponavljanje (repeticija).

Time se omogućuje kontrola i procjena varijabilnosti uvjetovane različitošću između redova i između stupaca, odnosno izdvajanje tih dijelova varijabilnosti iz pogreške pokusa. Tako pogreška pokusa postaje manja u odnosu na pogrešku kod slučajnog bloknog rasporeda.

Ove dvosmjerne repeticije predstavljaju i određeno ograničenje u primjeni sheme latinskog kvadrata. Naime, da bi se ta shema mogla izvesti, nužno je da broj ponavljanja (i vodoravnih i okomitih) bude jednak broju članova pokusa.

Dakle, **uvjet za latinski kvadrat je koliko tretiranja, toliko ponavljanja u pokusu**. Stoga ova shema postaje nepraktična ili neizvediva kad je broj članova u pokusu velik.

Naime, veliki broj članova, a time i repeticija, može predimenzionirati pokus, pa postaje upitna homogenost uvjeta za sva tretiranja uključena u pokus.

Jednako tako kad je broj članova u pokusu malen, slobodne varijante pogreške postaju premale za realno i pouzdano testiranje.

Latinski kvadrat bi se zato mogao preporučiti za pokuse s 5 do 8 ili maksimalno 10 tretiranja, što dakako uvjetuje 5-8-10 ponavljanja.

Zbog randomiziranja članova pokusa u dvovrsna (dvosmjerna) ponavljanja, za izradu ovog plana trebalo bi više truda i vremena, pa se preporuča korištenje već gotovih shema. Evo nekih od njih, s različitim brojem tretiranja (n_T), odnosno vodoravnih i okomitih ponavljanja.

$n_T = 5$

	I	II	III	IV	V
I	1	2	3	4	5
II	5	1	4	2	3
III	4	3	1	5	2
IV	2	4	5	3	1
V	3	5	2	1	4

	I	II	III	IV	V
I	3	4	2	5	1
II	5	3	1	2	4
III	2	5	4	1	3
IV	4	1	5	3	2
V	1	2	3	4	5

$n_T = 6$

	I	II	III	IV	V	VI
I	1	2	3	4	5	6
II	3	4	1	2	6	5
III	5	1	6	3	2	4
IV	2	5	4	6	1	3
V	4	6	2	5	3	1
VI	6	3	5	1	4	2

	I	II	III	IV	V	VI
I	4	3	5	1	6	2
II	2	5	6	3	4	1
III	6	1	4	2	3	5
IV	5	4	2	6	1	3
V	3	6	1	5	2	4
VI	1	2	3	4	5	6

$n_T = 7$

	I	II	III	IV	V	VI	VII
I	1	2	3	4	5	6	7
II	7	5	2	6	1	4	3
III	3	6	1	5	2	7	4
IV	5	4	7	2	6	3	1
V	4	1	6	3	7	2	5
VI	6	7	4	1	3	5	2
VII	2	3	5	7	4	1	6

	I	II	III	IV	V	VI	VII
I	5	4	6	2	7	1	3
II	4	6	2	7	3	5	1
III	7	5	4	1	2	3	6
IV	6	1	7	3	4	2	5
V	3	7	5	6	1	4	2
VI	2	3	1	5	6	7	4
VII	1	2	3	4	5	6	7

$n_T = 8$

	I	II	III	IV	V	VI	VII	VIII
I	5	6	1	8	3	7	2	4
II	6	7	2	5	1	8	4	3
III	7	4	8	2	6	3	5	1
IV	8	1	4	6	2	5	3	7
V	3	5	6	7	4	1	8	2
VI	2	8	5	3	7	4	1	6
VII	4	3	7	1	8	2	6	5
VIII	1	2	3	4	5	6	7	8

	I	II	III	IV	V	VI	VII	VIII
I	1	2	3	4	5	6	7	8
II	3	5	7	2	4	8	1	6
III	8	6	4	1	7	2	3	5
IV	2	3	8	5	6	1	4	7
V	4	1	6	7	3	5	8	2
VI	5	7	2	8	1	4	6	3
VII	6	8	1	3	2	7	5	4
VIII	7	4	5	6	8	3	2	1

Latinski kvadrat, shema s dvosmjernim repeticijama, omogućuje analiziranje varijabilnosti kroz procjenu ovih suma kvadratnih odstupanja (*SS*):

SS_{Ukupno} →
- $SS_{\text{Vodoravnih repeticija}}$
- $SS_{\text{Okomitih repeticija}}$
- $SS_{\text{Tretiranja}}$
- $SS_{\text{Pogreške}}$

Slobodne varijante za svaki izvor varijabilnosti, a s obzirom da je jednak broj vodoravnih repeticija, okomitih repeticija i tretiranja (to znači $n_{rv} = n_{ro} = n_T$), je:

		n - 1
SS_{Ukupno}	→	$n_T^2 - 1$
$SS_{\text{Vodoravne repceticije}}$	→	$n_T - 1$
$SS_{\text{Okomite repeticije}}$	→	$n_T - 1$
$SS_{\text{Tretiranja}}$	→	$n_T - 1$
$SS_{\text{Pogreške}}$	→	$(n_T - 1)(n_T - 2)$

Razvidno je da se u odnosu na slučajni blokni raspored iz pogreške pokusa izdvaja još jedan dio varijabilnosti, čime se pogreška smanjuje.

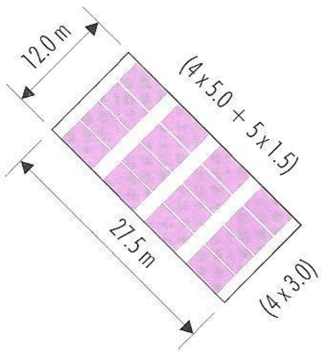
PRIMJER 9.3.1.

U vrlo ujednačenim uvjetima eksperimentalnog polja, proveden je pokus s 4 sorte jagoda: 3 nove sorte (S_1, S_2 i S_3) i jednom sortom do tada korištenom u proizvodnji, a zadovoljavajućeg prinosa (S_4).

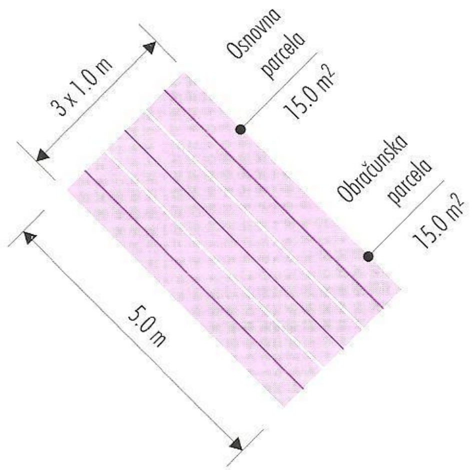
Pošlo se od pretpostavke da se prinosi ovih sorata neće značajno razlikovati od prinosa sorte S_4.

Primijenjena je bila shema latinskog kvadrata. To znači da je pokus imao 4 tretiranja, 4 vodoravne i 4 okomite repeticije, a ukupno 16 osnovnih parcela.

Osnovna parcela (ujedno i obračunska) bila je veličine 10.5 m^2 (3 reda 5.0 m duljine, s 0.3 m razmacima unutar reda i 1.0 m razmakom između redova).



Veličina pokusa 412.5 m^2



Između parcelica nije bilo nikakvih posebnih razmaka (jer je razmak od 1.0 m između redova dovoljan za kretanje), a između ponavljanja bile su staze širine 1.5 m.

Tako je veličina pokusnog polja bila 412.5 m^2

Plan pokusa i prinosi u kg/15.0 m^2:

	I.	II.	III.	IV.	$\sum x_{\text{Vod. rep.}}$
	S_1	S_2	S_3	S_4	
I.	29.1	20.0	47.0	36.0	132.1
	S_3	S_1	S_4	S_2	
II.	45.0	22.1	41.1	17.4	125.6
	S_4	S_3	S_2	S_1	
III.	29.0	38.0	32.2	39.0	138.2
	S_2	S_4	S_1	S_3	
IV.	21.3	42.0	26.3	40.1	129.7
$\sum x_{\text{Okom. rep.}}$	124.4	122.1	146.6	132.5	$\sum x = 525.6$

Iako je svaka repeticija istovjetnog sadržaja (sve sorte), ipak je zamjetna različitost prinosa, kako između vodoravnih tako i između okomitih repeticija.

Podatke o prinosu ćemo analizirati na isti način kao kod slučajnog bloknog rasporeda, s time što će se dodatno procijeniti varijabilnost uvjetovana različitošću okomitih repeticija.

Dijelovi varijabilnosti s pripadajućim slobodnim varijantama:

		n - 1
SS_{Ukupno}	$\rightarrow$	$4 \cdot 4 - 1 = 15$
$SS_{\text{Vodoravne repceticije}}$	$\rightarrow$	$4 - 1 = 3$
$SS_{\text{Okomite repeticije}}$	$\rightarrow$	$4 - 1 = 3$
SS_{Sorte}	$\rightarrow$	$4 - 1 = 3$
$SS_{\text{Pogreške}}$	$\rightarrow$	$15 - 9 = 6$

Da bismo mogli provesti cjelovitu ANOVA, podatke o prinosima sredit ćemo po sortama:

	S_1	S_2	S_3	S_4	
	29.1	20.0	47.0	36.0	
	22.1	17.4	45.0	41.0	
	39.0	32.2	38.0	29.0	
	26.3	21.3	40.1	42.0	
$\sum x_{\text{Sorte}}$	116.5	90.9	170.1	148.1	$\sum x = 525.6$
$\bar{x}_{\text{Sorte}}$	29.12	22.72	42.52	37.02	

ANOVA:

$$SS_{\text{Ukupno}} = 29.1^2 + \ldots + 40.1^2 - \frac{525.6^2}{16} = 1351.460$$

$$SS_{\text{Vodoravnih rep.}} = \frac{132.1^2 + \ldots + 129.7^2}{4} - \frac{525.6^2}{16} = 20.815$$

$$SS_{\text{Okomitih rep.}} = \frac{124.4^2 + \ldots + 132.5^2}{4} - \frac{525.6^2}{16} = 91.935$$

$$SS_{\text{Sorata}} = \frac{116.5^2 + \ldots + 148.1^2}{4} - \frac{525.6^2}{16} = 909.710$$

$$SS_{\text{Pogreške}} = SS_{\text{Ukupno}} - SS_{\text{Vod. rep.}} - SS_{\text{Okom. rep.}} - SS_{\text{Sorata}}$$
$$= 1351.460 - 20.815 - 91.935 - 909.710 = 329.000$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	$F_{exp.}$	$F_{tabl.}$	
					P=5%	P=1%
Ukupno	15	1351.460				
Vod. repet.	3	20.815				
Okom. repet.	3	91.935				
Sorte	3	909.710	303.237	5.53*	4.76	9.78
Pogreška	6	329.000	54.833			

Tablični je F-faktor očitan iz n - 1 sorte (3) i n - 1 pogreške (6), a manji je od F_{exp} (uz P=0.05), što znači da sa 95% sigurnošću možemo odbaciti hipotezu da se prinosi ovih sorata jagoda ne razlikuju značajno.

To ne mora značiti da su sve razlike signifikantne. To samo znači da opravdanih razlika u prinosu ima. No koje su to (a koje možda ne) treba ustanoviti pomoću t-testa.

Tablični t-faktor očitan iz n - 1 pogreške (6) je

$$t_{P=5\%} = 2.45$$
$$t_{P=1\%} = 3.71$$

s_D treba izračunati, pa je

$$s_D = \sqrt{\frac{2 \cdot 54.833}{4}} = 5.236 \text{ kg}$$

Iz toga je najmanja signifikantna razlika

$$LSD_{P=5\%} = 2.45 \cdot 5.236 = 12.83 \text{ kg}/15.0 \text{ m}^2$$
$$LSD_{P=1\%} = 3.71 \cdot 5.236 = 19.42 \text{ kg}/15.0 \text{ m}^2$$

To znači, ako se prinosi ovih novih sorata jagoda razlikuju od prinosa sorte S_4 za bar 12.83 kg (odnosno 19.42 kg) po parceli, možemo ih smatrati statistički opravdanima, značajno ili signifi-kantno različitima (odnosno visokosignifikantno ako su razlike bar 19.42 kg ili veće).

Usporedimo ih:

	Razlike |D|, kg/15.0 m^2
37.02 (S_4) - 29.12 (S_1) =	7.9 ns
37.02 (S_4) - 22.72 (S_2) =	14.3 *
37.02 (S_4) - 42.52 (S_3) =	5.5 ns

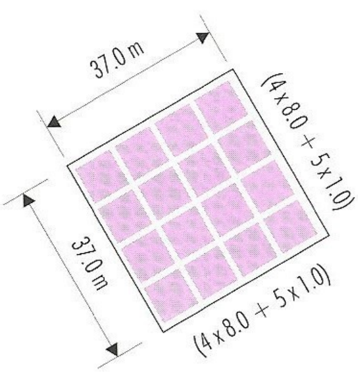
Konačno možemo zaključiti: sortu S_4, do tada korištenu u proizvodnji kao sortu zadovoljavajućeg prinosa ni jedna od ove tri nove, pokusom provjerene sorte, ne može nadomjestiti. Naime, sorte S_1 i S_2 manjeg su prinosa (S_2 čak opravdano manjeg), a prinos sorte S_3 je nešto veći ali ta razlika nije opravdana, nije statistički značajna, nije signifikantna; ona je samo slučajna.

PRIMJER 9.3.2.

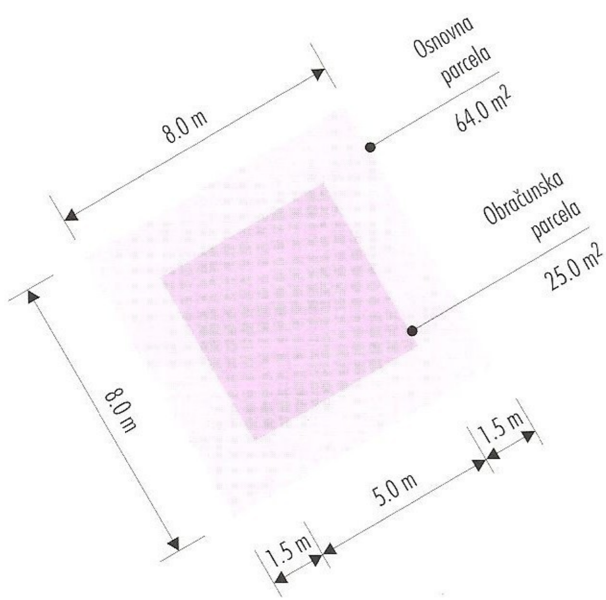
Kroz gnojidbeni pokus željelo se utvrditi kojim od četiri različita dušična gnojiva N_1, N_2, N_3 ili N_4 (a uz istu količinu P_2O_5) je moguće postići najveći prinos sijena.

Pokus je izveden po shemi latinskog kvadrata. Veličina osnovne parcelice je bila 64.0 m^2 (8.0 m x 8.0 m). a obračunske 25.0 m^2 (nakon što su se sa svake strane odstranili rubovi od po 1.5 m.

Između parcelica u repeticiji kao i između repeticija bile su staze širine 1.0 m.



Veličina pokusa 1369.0 m^2



Veličina pokusa bila je 1369.0 m^2 (tj. 4 x 8 m = 32.0 m + 5 staza po 1.0 m = 37.0 m, odnosno 37.0 m x 37.0 m = 1369.0 m^2)

Plan pokusa i podaci o prinosima sijena u kg po obračunskoj parcelici

	I.	II.	III.	IV.	$\sum x_{\text{Vod. rep.}}$
I.	N_1 17.0	N_2 15.0	N_4 19.5	N_3 26.0	77.5
II.	N_3 14.3	N_1 21.2	N_3 22.4	N_2 17.5	75.4
III.	N_3 25.0	N_4 13.2	N_2 16.7	N_1 22.3	77.2
IV.	N_2 13.8	N_3 26.0	N_1 24.0	N_4 17.4	81.2
$\sum x_{\text{Okom. rep.}}$	70.1	75.4	82.6	83.2	$\sum x = 311.3$

Prinosi po gnojidbama:

	N_1	N_2	N_3	N_4	
	17.0	15.0	26.0	19.5	
	21.2	17.5	22.4	14.3	
	22.3	16.7	25.0	13.2	
	24.0	13.8	26.0	17.4	
$\sum x_{\text{Gnojidbe}}$	84.5	63.0	99.4	64.4	$\sum x = 311.3$
$\bar{x}_{\text{Gnojidbe}}$	21.12	15.75	24.85	16.10	

ANOVA:

$$SS_{\text{Ukupno}} = 17.0^2 + ... + 17.4^2 - \frac{311.3^2}{16} = 296.079$$

$$SS_{\text{Vodoravnih rep.}} = \frac{77.5^2 + ... + 81.2^2}{4} - \frac{311.3^2}{16} = 4.442$$

$$SS_{\text{Okomitih rep.}} = \frac{70.1^2 + ... + 83.2^2}{4} - \frac{311.3^2}{16} = 29.312$$

$$SS_{\text{Gnojidbi}} = \frac{84.5^2 + ... + 64,4^2}{4} - \frac{311,3^2}{16} = 227.512$$

$$SS_{\text{Pogreške}} = 296.079 - 4.442 - 29.312 - 227.512 = 34.813$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl}	
					P=5%	P=1%
Ukupno	15	296.079				
Vod. repet.	3	4.442				
Okom. repet.	3	29.312				
Gnojidbe	3	227.512	75.837	13.07**	4.76	9.78
Pogreška	6	34.813	5.802			

Signifikantan F-test gnojidbe upućuje na provođenje t-testa u svrhu uspoređivanja prosječnih vrijednosti.

Tablični t faktor očitan iz n - 1 pogreške (6) je

$$t_{P=5\%} = 2.45$$

$$t_{P=1\%} = 3.71$$

$$s_D = \sqrt{\frac{2 \cdot 5.802}{4}} = 1.703 \text{ kg}$$

Najmanja značajna razlika je

$$LSD_{P=5\%} = 2.45 \cdot 1.703 = 4.17 \text{ kg}/25.0 \text{ m}^2$$

$$LSD_{P=1\%} = 3.71 \cdot 1.703 = 6.32 \text{ kg}/25.0 \text{ m}^2$$

Iz vrijednosti prosječnih prinosa čini se da je gnojidbom s oznakom N_3 postignut najveći prinos sijena. No, usporedimo ih i utvrdimo opravdanost razlika:

	Razlike |D|, kg/25.0 m²
24.85 (N_3) - 21.12 (N_1) =	3.73 ns
24.85 (N_3) - 15.75 (N_2) =	9.10 **
24.85 (N_3) - 16.10 (N_4) =	8.75 **

Dakle, primjenom dušične gnojidbe označene kao N_3 postižu se zanemarivo malo veći prinosi u odnosu na gnojidbu N_1, a visoko signifikantno veći od prinosa dobivenih uz varijante N_2 i N_4.

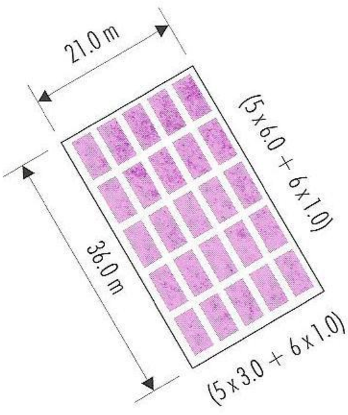
Praktično gledano, ako je kombinacija gnojiva N_3 skuplja od N_1, odlučit ćemo se za N_1, jer je prinos zanemarivo malo niži (razlika nije opravdana).

■

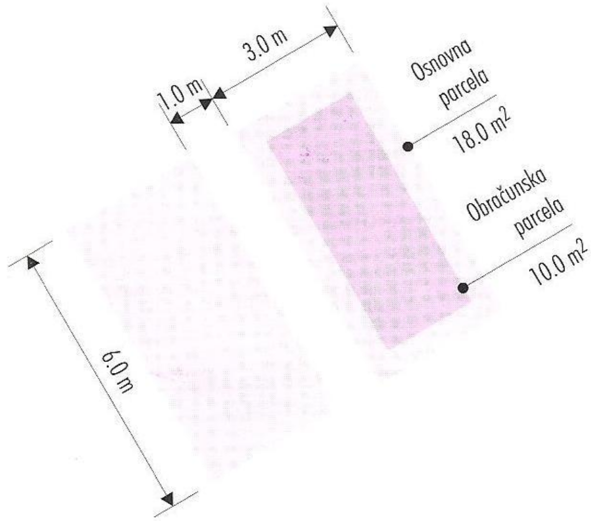
PRIMJER 9.3.3.

Plan latinskog kvadrata primijenjen je u pokusu s 5 sorata jarog stočnog graška (G_1, G_2, G_3, G_4 i G_5).

Nulta hipoteza od koje se krenulo u eksperiment, bila je pretpostavka da su prinosi suhog zrna s 14% vlage ovih sorata podjednaki.



Veličina pokusa 756.0 m²



Veličina osnovne parcele je bila 18.0 m² (3.0 m x 6.0 m), a obračunske parcele 10.0 m² (nakon otkosa po 0.5 m sa svih strana to je 2.0 m x 5.0 m). Između osnovnih parcela kao i između repeticija ostavljene su staze od po 1.0 m, pa je tako veličina pokusa bila 756.0 m² (21.0 m x 36.0 m).

Raspored sorata u pokusu (plan) i podaci o prinosima suhog zrna u kg na osnovnoj parceli:

	I.	II.	III.	IV.	V.	$\sum x_{\text{Vod. rep.}}$
	G_2	G_4	G_5	G_3	G_1	
I.	4.00	4.93	1.01	4.95	2.78	17.67
	G_1	G_2	G_3	G_4	G_5	
II.	0.87	1.50	1.20	1.82	1.01	6.40
	G_4	G_3	G_1	G_5	G_2	
III.	4.38	4.86	0.55	5.07	3.85	18.71
	G_5	G_1	G_4	G_2	G_3	
IV.	1.30	0.71	1.06	1.59	1.08	5.74
	G_3	G_5	G_2	G_1	G_4	
V.	0.87	4.96	4.98	3.02	4.09	17.92
$\sum x_{\text{Okom. rep.}}$	11.42	16.96	8.80	16.45	12.81	$\sum x = 66.44$

Podaci svrstani po sortama:

	G_1	G_2	G_3	G_4	G_5	
	2.78	4.00	4.95	4.93	1.01	
	0.87	1.50	1.20	1.82	1.01	
	0.55	3.85	4.86	4.38	5.07	
	0.71	1.59	1.08	1.06	1.30	
	3.02	4.98	0.87	4.09	4.96	
$\sum x_{Sorte}$	7.93	15.92	12.96	16.28	13.35	$\sum x = 66.44$
$\bar{x}_{Sorte}$	1.59	3.18	2.59	3.26	2.67	

ANOVA:

$$SS_{\text{Ukupno}} = 4.00^2 + \ldots + 4.09^2 - \frac{66.44^2}{25} = 72.4178$$

$$SS_{\text{Vodoravnih rep.}} = \frac{17.67^2 + \ldots + 17.92^2}{5} - \frac{66.44^2}{25} = 34.8944$$

$$SS_{\text{Okomitih rep.}} = \frac{11.42^2 + \ldots + 12.81^2}{5} - \frac{66.44^2}{25} = 9.4684$$

$$SS_{\text{Sorata}} = \frac{7.93^2 + \ldots + 13.35^2}{5} - \frac{66.44^2}{25} = 8.9398$$

$$SS_{\text{Pogreške}} = 72.4178 - 34.8944 - 9.4684 - 8.9398 = 19.1152$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	24	72.4178				
Vod. repet.	4	34.8944				
Okom. repet.	4	9.4684				
Sorte	4	8.9398	2.2349	1.40 ns	3.26	5.41
Pogreška	12	19.1152	1.5929			

Ovaj nesignifikantni F-test upućuje na zaključak: pretpostavka je bila istinita, prihvaća se nulta hipoteza, što praktično znači da nisu utvrđene opravdane razlike u prinosima zrna (s 14 % vlage) sorata koje su bile u pokusu. Postojeće razlike u prinosima zrna ovih sorata, moraju se stoga smatrati samo slučajnima.

9.4. Latinski pravokutnik

Kad je broj članova koje želimo uključiti u pokus velik, primjenom latinskog kvadrata, a time i potrebe velikog broja ponavljanja, povećavala bi se varijabilnost koju se ne bi moglo kontrolirati niti točno procijeniti.

Ipak, rješenje je u shemi latinskog pravokutnika, rasporedu članova koji je na neki način kombinacija slučajnog bloknog rasporeda i latinskog kvadrata.

Tretiranja se grupiraju u onoliko grupa koliko ima vodoravnih repeticija, a grupe se randomiziraju unutar svake okomite repeticije. To je moguće samo ako je broj članova djeljiv brojem ponavljanja (što je uvjet za latinski pravokutnik).

Ako, na primjer, u pokus želimo uključiti 8 članova, odlučili bismo se za 4 repeticije (što je moguće jer je 8 djeljivo sa 4), pa bi tako dijeleći tretiranja s 4 dobili 4 grupe od po 2 člana i zatim ih randomizirali u oba smjera. To se može pisati i kao 4 x (4 x 2).

	I.		II.		III.		IV.	
I.	1	2	3	4	5	6	7	8
II.	5	8	6	7	2	4	1	3
III.	6	3	8	2	1	7	5	4
IV.	4	7	5	1	3	8	2	6

Razvidno je da su u svakoj vodoravnoj i svakoj okomitoj repeticiji svi članovi i da se nikad niti jedan ne ponavlja. Tako će biti moguće pokusne podatke analizirati potpuno analogno planu latinskog kvadrata.

Evo još nekoliko mogućih planova, ovisno o broju tretiranja (n_T)

$n_T = 10 \rightarrow 5 \times (5 \times 2)$

	I.		II.		III.		IV.		V.	
I.	1	2	3	4	5	6	7	8	9	10
II.	8	5	9	7	2	4	10	3	1	6
III.	9	7	8	1	3	10	4	6	5	2
IV.	6	3	5	10	1	7	9	2	8	4
V.	4	10	6	2	8	9	1	5	7	3

$n_T = 12 \rightarrow 4 \times (4 \times 3)$

	I.			II.			III.			IV.		
I.	1	3	2	6	4	5	7	8	12	10	9	11
II.	6	4	5	11	8	2	10	3	9	1	12	7
III.	9	7	8	10	12	1	6	2	11	4	3	5
IV.	12	11	10	9	7	3	5	4	1	2	8	6

$n_T = 15 \rightarrow 5 \times (5 \times 3)$

	I.			II.			III.			IV.			V.		
I.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
II.	4	13	11	9	15	7	1	6	12	8	2	14	5	10	3
III.	9	14	7	10	8	12	15	2	5	13	1	3	4	11	6
IV.	8	5	12	11	2	13	3	14	10	6	15	4	9	1	7
V.	10	6	15	14	1	3	11	13	4	9	5	7	2	12	8

$n_T = 20 \rightarrow 5 \times (5 \times 4)$

	I.				II.				III.				IV.				V.			
I.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
II.	13	5	16	11	17	2	9	15	19	3	6	14	18	1	20	10	7	4	12	8
III.	14	18	8	6	10	16	1	12	13	20	2	15	7	19	4	17	9	11	3	5
IV.	15	7	17	19	4	20	11	14	18	1	16	5	8	12	3	9	6	13	2	10
V	20	10	12	9	13	19	3	18	7	17	4	8	6	11	2	5	14	16	1	15

Analiza podataka pokusa izvedenog po shemi latinskog pravokutnika potpuno je istovjetna analizi po latinskom kvadratu.

PRIMJER 9.4.1.

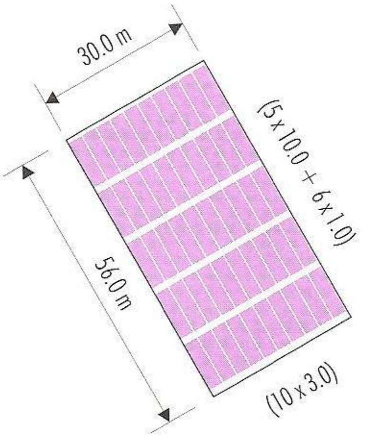
Kako bi se provjerila gospodarska vrijednost 9 novostvorenih hibrida duhana (H_1, H_2, H_3, H_4, ... , H_9), proveden je poljski pokus u kojem je osim ovih hibrida bio i standard (H_{10}), do tada u proizvodnji najbolji hibrid.

Tijekom pokusa i nakon berbi praćena su mnoga, za proizvodnju važna svojstva, a među njima prinos osušenog lista.

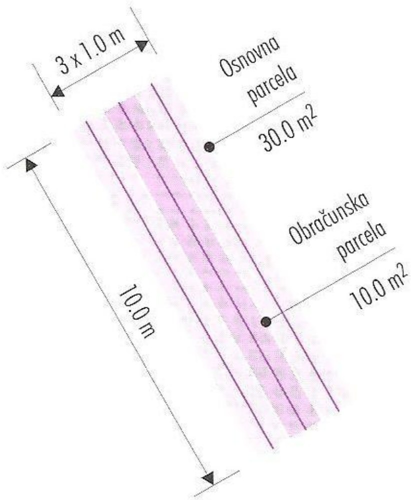
Krenulo se od pretpostavke da prinos lista ovih novostvorenih hibrida neće značajno nadmašiti standard.

Pokus je izveden po shemi latinskog pravokutnika s 5 ponavljanja. Naime, broj članova (hibrida) je djeljiv s 5 (tako se dobije 5 grupa po 2 člana) pa je moguće kreirati vodoravne i okomite repeticije.

Veličina osnovne parcele je bila 30.0 m^2 (tj. 3 reda s razmakom po 1.0 m, duljine 10.0 m), a obračunske 10.0 m^2 (1 unutarnji red).



Veličina pokusa 1680.0 m^2



Nije bilo razmaka između okomitih repeticija niti između osnovnih parcela, a između vodoravnih repeticija bile su staze od 1.0 m.

Tako je veličina pokusa bila 1680.0 m^2, 10 hibrida x 3.0 m = 30.0 m, te 5 repeticija po 10.0 m plus 6 staza od 1.0 m je 56.0 m ili 30.0 m x 56.0 m = 1680.0 m^2.

Plan pokusa i podaci o prinosu osušenog lista (t/ha):

	I.		II.		III.		IV.		V.		$\sum x_{\text{Vod. rep.}}$
	H_8	H_5	H_9	H_7	H_2	H_4	H_{10}	H_3	H_1	H_6	
I.	3.11	1.97	2.97	2.27	2.31	2.51	2.19	2.51	2.41	2.68	24.93
	H_9	H_7	H_8	H_1	H_3	H_{10}	H_4	H_6	H_5	H_2	
II.	2.67	2.79	1.94	1.85	2.18	2.48	3.03	2.77	2.35	2.05	24.11
	H_6	H_3	H_5	H_{10}	H_1	H_7	H_9	H_2	H_8	H_4	
III.	2.50	1.80	2.46	1.90	1.84	2.36	2.30	1.78	2.66	2.08	21.68
	H_4	H_{10}	H_6	H_2	H_8	H_9	H_1	H_5	H_7	H_3	
IV.	1.69	1.69	1.41	1.50	1.91	1.63	1.76	2.33	1.89	1.58	17.39
	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	
V.	1.80	1.60	1.80	1.70	2.00	2.00	1.90	1.70	1.70	1.80	18.00
$\sum x_{\text{Okom. rep.}}$	21.62		19.80		21.22		22.27		21.20		$\sum x = 106.11$

Analizom podataka moći će se procijeniti ovi dijelovi varijabilnosti (*SS-ovi*) i pripadajuće slobodne varijante (n-1):

		$n - 1$
SS_{Ukupno}	→	(10 x 5) - 1 = 49
$SS_{\text{Vodoravnih repeticija}}$	→	5 - 1 = 4
$SS_{\text{Okomitih repeticija}}$	→	5 - 1 = 4
SS_{Hibrida}	→	10 - 1 = 9
$SS_{\text{Pogreške}}$	→	49 - 4 - 4 - 9 = 32

Ako se podaci o prinosu srede po hibridima:

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	
	2.41	2.31	2.51	2.51	1.97	2.68	2.27	3.11	2.97	2.19	
	1.85	2.05	2.18	3.03	2.35	2.77	2.79	1.94	2.67	2.48	
	1.84	1.78	1.80	2.08	2.46	2.50	2.36	2.66	2.30	1.90	
	1.76	1.50	1.58	1.69	2.33	1.41	1.89	1.91	1.63	1.69	
	1.80	1.60	1.80	1.70	2.00	2.00	1.90	1.70	1.70	1.80	
$\sum x_{\text{Hibrida}}$	9.66	9.24	9.87	11.01	11.11	11.36	11.21	11.32	11.27	10.06	$\sum x = 106.11$
$\bar{x}_{\text{Hibrida}}$	1.932	1.848	1.974	2.202	2.222	2.272	2.242	2.264	2.254	2.012	

ANOVA:

$$SS_{\text{Ukupno}} = 3.11^2 + \ldots + 1.80^2 - \frac{106.11^2}{50} = 9.0212$$

$$SS_{\text{Vodoravnih rep.}} = \frac{24.93^2 + \ldots + 18.00^2}{10} - \frac{106.11^2}{50} = 4.7365$$

$$SS_{\text{Okomitih rep.}} = \frac{21.62^2 + \ldots + 21.20^2}{10} - \frac{106.11^2}{50} = 0.3279$$

$$SS_{\text{Hibrida}} = \frac{9.66^2 + \ldots + 10.06^2}{5} - \frac{106.11^2}{50} = 1.1803$$

$$SS_{\text{Pogreške}} = 9.0212 - 4.7365 - 0.3279 - 1.1803 = 2.7765$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	$F_{exp.}$	$F_{tabl.}$	
					P=5%	P=1%
Ukupno	49	9.0212				
Vodoravne repet.	4	4.7365				
Okomite repet.	4	0.3279				
Hibridi	9	1.1803	0.1311	1.51 ns	2.21	3.07
Pogreška	32	2.7765	0.0867			

Tablični F-faktor očitan je za 9 i 30 (a ne 32 koliko je n - 1 pogreške; tako je dostupno u tablicama, no to je još stroži kriterij).

Nesignifikantni F-test dozvoljava nam prihvatiti nultu hipotezu, odnosno zaključiti da se u ovom pokusu nije ustanovila značajna različitost u prinosima suhog lista ovih deset uključenih hibrida duhana. Razlike među njima su neopravdane, nesignifikantne, zanemarive. Prinos niti jednog novog hibrida duhana nije opravdano veći od standarda (H_{10})

PRIMJER 9.4.2.

Provedena su istraživanja o utjecaju gnojidbe crvenog radiča različitim varijantama N, P i K gnojiva u kojima je među brojnim kemijskim i gospodarskim svojstvima bilo interesantno odrediti i količinu N u suhoj tvari biljke.

U tu je svrhu izveden pokus s 8 varijanata gnojidbe ($G_1, G_2, \ldots, G_8$) po shemi latinskog pravokutnika u 4 ponavljanja.

Varijante gnojidbe su proizašle iz četiri razine gnojidbe dušikom i bez gnojidbe s N tj. 0, 50, 100, 150 i 200 kg N/ha, iz 100 kg P_2O_5/ha i bez P, te dvije razine kalija 100 i 200 kg/ha i bez K. One su:

$G_1 = N_0\ P_0\ K_0$ $\qquad G_5 = N_{150}\ P_{100}\ K_{100}$

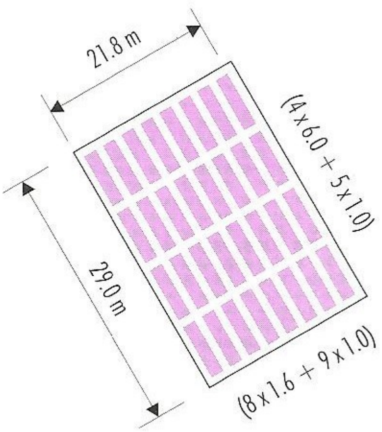
$G_2 = N_0\ P_{100}\ K_{100}$ $\qquad G_6 = N_{200}\ P_{100}\ K_{100}$

$G_3 = N_{50}\ P_{100}\ K_{100}$ $\qquad G_7 = N_{100}\ P_{100}\ K_0$

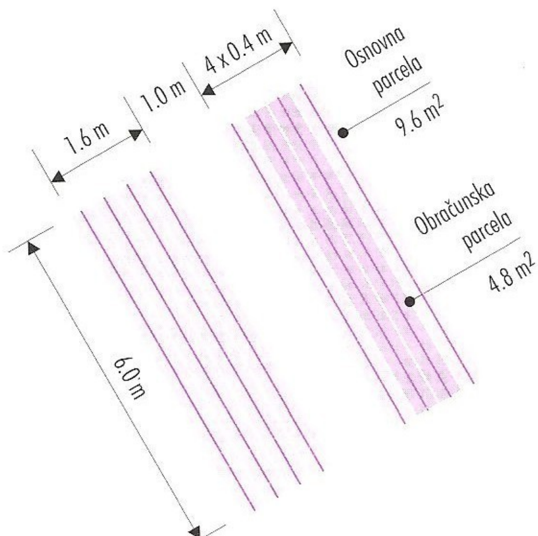
$G_4 = N_{100}\ P_{100}\ K_{100}$ $\qquad G_8 = N_{200}\ P_{100}\ K_{200}$

S obzirom da se nije moglo pretpostaviti da li će i kako ove varijante gnojidbe utjecati na postotak N u biljci radiča, pošlo se od pretpostavke da neće biti nikakvih značajnih razlika.

Osnovna parcela je bila veličine 1.6 m x 6.0 m = 9.6 m^2 (4 reda s razmakom 0.4 m. duljine 6.0 m), a obračunska 4.8 m^2 (dva unutarnja reda).



Veličina pokusa 654.0 m^2



Između osnovnih parcelica i između repeticija bile su staze po 1.0 m širine. Tako je veličina pokusa bila 21.8 x 30.0 m = 654.0 m^2

Plan pokusa i podaci o sadržaju N u biljci (u %):

	I.		II.		III.		IV.		$\sum x_{\text{Vod. rep.}}$
	G_7	G_8	G_1	G_3	G_4	G_5	G_2	G_6	
I.	4.13	3.95	3.84	3.57	3.40	3.89	3.90	4.50	31.18
	G_5	G_6	G_2	G_4	G_1	G_7	G_3	G_8	
II.	4.70	4.75	4.15	4.07	3.78	4.77	4.27	2.82	33.31
	G_3	G_4	G_6	G_7	G_2	G_8	G_1	G_5	
III.	4.24	3.90	3.84	4.12	3.80	2.81	3.93	4.39	31.03
	G_1	G_2	G_5	G_8	G_3	G_6	G_4	G_7	
IV.	3.96	4.43	4.66	4.18	4.25	5.29	3.98	4.49	35.24
$\sum x_{\text{Okom. rep.}}$	34.06		32.43		31.99		32.28		$\sum x = 130.76$

Podaci o sadržaju N u biljci poredani po pojedinim varijantama gnojidbe:

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	
	3.84	3.90	3.57	3.40	3.89	4.50	4.13	3.95	
	3.78	4.15	4.27	4.07	4.70	4.75	4.77	2.82	
	3.93	3.80	4.24	3.90	4.39	3.84	4.12	2.81	
	3.96	4.43	4.25	3.98	4.66	5.29	4.49	4.18	
$\sum x_{\text{Gnojidbe}}$	15.51	16.28	16.33	15.35	17.64	18.38	17.51	13.76	$\sum x = 130.76$
$\bar{x}_{\text{Gnojidbe}}$	3.877	4.070	4.080	3.837	4.410	4.590	4.377	3.440	

ANOVA

$$SS_{\text{Ukupno}} = 3.13^2 + \ldots + 4.49^2 - \frac{130.76^2}{32} = 8.15275$$

$$SS_{\text{Vodoravnih rep.}} = \frac{31.18^2 + \ldots + 35.24^2}{8} - \frac{130.76^2}{32} = 1.49032$$

$$SS_{\text{Okomitih rep.}} = \frac{34.06^2 + \ldots + 32.28^2}{8} - \frac{130.76^2}{32} = 0.32532$$

$$SS_{\text{Gnojidbe}} = .\frac{15{,}51^2 + \ldots + 13.76^2}{4} - \frac{130.76^2}{32} = 3.88735$$

$$SS_{\text{Pogreške}} = SS_{\text{Ukupno}} - SS_{\text{Vod. rep.}} - SS_{\text{Okom. rep.}} - SS_{\text{Gnojidbe}} =$$

$$= 2.44975$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	31	8.15275				
Vodoravne repet.	3	1.49032				
Okomite repet.	3	0.32532				
Gnojidbe	7	3.88735	0.5553	4.08 **	2.58	3.84
Pogreška	18	2.44975	0.1361			

Razvidno je da se nakon signifikantnog F-testa odbacuje nulta hipoteza. To, naime, znači da ima značajnih različitosti u sadržaju N u radiču, uvjetovanih različitim varijantama gnojidbe. Koje su to, utvrdit ćemo t-testom.

Tablični t-faktor (očitan iz 18 = n - 1 pogreške) je

$$t_{P=5\%} = 2.10$$
$$t_{P=1\%} = 2.88$$
$$s_D = \sqrt{\frac{2 \cdot 0.1361}{4}} = 0.2609\ \%$$

Najmanje značajne razlike su

$$LSD_{P=5\%} = 2.10 \cdot 0.2609 = 0.55\ \%$$
$$LSD_{P=1\%} = 2.88 \cdot 0.2609 = 0.75\ \%$$

Iz prosječnih se vrijednosti čini da je varijanta gnojidbe G_6 (a to je $N_{200}\ P_{100}\ K_{100}$) uzrokovala i najveći sadržaj N u biljci. No. to tek treba utvrditi usporedbom prosječnih vrijednosti i usporedbom razlika s najmanjom značajnom razlikom:

	Razlike |D|, %
4.59 (G_6) - 3.88 (G_1) =	0.71 *
4.59 (G_6) - 4.07 (G_2) =	0.52 ns
4.59 (G_6) - 4.08 (G_3) =	0.51 ns
4.59 (G_6) - 3.84 (G_4) =	0.75 *
4.59 (G_6) - 4.41 (G_5) =	0.18 ns
4.59 (G_6) - 4.38 (G_7) =	0.21 ns
4.59 (G_6) - 3.44 (G_8) =	1.15 **

Dakle, uz gnojidbu s najvećom količinom N, a uz 100 kg P_2O_5/ha i 100 kg K_2O/ha (G_6) utvrđen je signifikantno veći sadržaj dušika u biljci radiča nego s $N_{100}\ P_{100}\ K_{100}$ (G_4), ili s $N_{200}\ P_{100}\ K_{200}$ (G_8) ili bez gnojidbe (G_1), Ostale varijante gnojidbe rezultirale su podjednakim sadržajem N u biljci radiča (među njima nema opravdanih razlika).

■

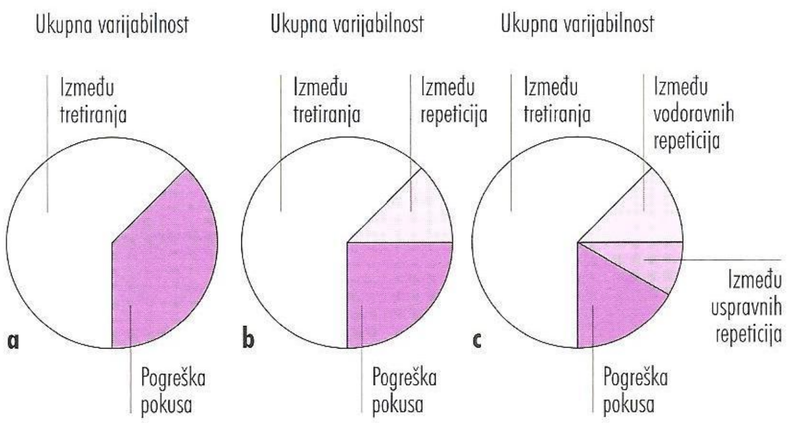
Kroz sve do sada razrađene primjere prikazali smo sheme koje se u poljoprivrednim istraživanjima najčešće koriste.

Opetovano smo naglašavali kako je najvažniji cilj u planiranju i izvođenju eksperimenta (čija je svrha provjeriti hipotezu i odgovoriti na postavljena pitanja) kontrola i procjena pogreške eksperimenta i njezino minimaliziranje.

To će se uspješnije ili manje uspješno postizati primjenom različitih shema prema kojima se provodi pokus i analiziraju eksperimentalni podaci.

Primjerima koje smo do sada naveli i opisivali, pošli smo od najjednostavnijeg rasporeda tretiranja (potpuno slučajnog rasporeda) u kojem se ukupna varijabilnost mogla raščlaniti samo na dva dijela (tretiranja i pogrešku), preko slučajnog bloknog rasporeda kojim je bilo moguće definirati jedan dio više (tretiranja, repeticije, pogreška), do latinskog kvadrata i latinskog pravokutnika u kojima je zbog dvostrukog grupiranja članova definiran još jedan dio varijabilnosti (tretiranja, vodoravne repeticije, okomite repeticije, pogreška).

Prikazano shematski, ako se krugom predstavi ukupna varijabilnost iz koje se, ovisno o planu pokusa, izdvajaju definirani dijelovi (kružni isječci), preostali nedefinirani dio - pogreška, razvidno je, postaje sve manji.



Potpuno slučajni raspored — Slučajni blokni raspored — Latinski kvadrat ili latinski pravokutnik

Time se treba rukovoditi kod izbora sheme po kojoj će se izvesti eksperiment, pa kad god je moguće izabrati onaj plan kojim se osigurava najmanja pogreška pokusa.

Ipak, moramo imati na umu da to nije uvijek moguće izvesti. Naime, ponekad broj tretiranja, veličina osnovne parcele (pa time i veličina pokusa) ili radni zahvati poljoprivrednih strojeva indirektno određuju jedino moguću shemu pokusa u danim okolnostima.

Prije nego završimo s ovim poglavljem treba naglasiti i ovo: svi primjeri do sada, opisujući osobitosti i način primjene planova pokusa koji se najčešće u bilinogojstvu koriste, imali su nešto zajedničko. Naime, u svakom od tih primjera članovi pokusa ili tretiranja bili su na neki način varijacije određenog faktora: **4 različite trave** u ishrani gusjenica gubara, **3 sorte** vinove loze, **4 roka sjetve** šećerne repe, **12 sorata** jabuka, **9 insekticida** u suzbijanju krumpirove zlatice, **4 sorte** jagoda, **4 varijante gnojidbe** livada, **8 kombinacija gnojidbe** radiča itd.

U svakom od ovih primjera proučavao se utjecaj jednog faktora predstavljenog doduše s više ili manje mogućnosti (4 trave, 3 sorte, 4 roka sjetve, 9 insekticida, 8 varijanata gnojidbe itd.).

Stoga govorimo o **jednofaktorijalnim pokusima, pokusima u kojima se proučava učinak jednog faktora (kroz manji ili veći broj njegovih gradacija) na svojstvo(a) od interesa**.

Međutim, u poljoprivredi gdje su pojave i svojstva izuzetno povezani i zavisni, nerijetko se ne možemo zadovoljiti s tako jednostavnom situacijom. Interesira nas mnogo više, na primjer: kada sijati ili saditi, koliko sjemena koristiti, da li primijeniti gnojidbu s osnovnim elementima i koju, da li dodati mikroelemente, treba li navodnjavati, kada i kako, koje je najpogodnije vrijeme berbe ili žetve, itd. itd.

Pokuse kojima možemo dobiti odgovore na ovakva i slična ali kompleksnija pitanja, detaljno ćemo obraditi u sljedećem poglavlju.

■

10 Višefaktorijalni pokusi

10.1. Dvofaktorijalni pokusi
10.2. Trofaktorijalni pokusi

U poljoprivredi je vrlo važno proučiti i što bolje razumjeti ponašanje biljke pod utjecajem okolinskih uvjeta koji se mijenjaju sa ili bez utjecaja čovjeka.

Stručnjaka zanima, na primjer, kako će različite sorte reagirati posiju li se u različitim rokovima, da li će uzgojni oblik neke voćne vrste uvjetovati isti učinak kod različitih sorata, koja je optimalna gustoća sklopa za postizanje maksimalnog prinosa različitih hibrida kukuruza, koji pesticid, kada i kako aplicirati, koju koncentraciju hormonskih preparata primijeniti za ožiljavanje reznica različitih biljnih vrsta, koliko gnojiva, kada i kako primjenjivati, itd. itd.

Za razliku od dosadašnjih, ovi se primjeri odnose na dva i više faktora čiji se učinci žele pratiti i proučiti.

Pokusi koji omogućavaju istovremeno uključivanje dvaju ili više faktora i proučavanje njihovih učinaka, zovu se višefaktorijalni pokusi.

Prisjetimo se onih primjera jednofaktorijalnih pokusa koje smo razrađivali u prethodnom poglavlju, a kroz koje se nastojalo doći do odgovora na pitanja:

1. utječe li ishrana gusjenica s 4 različite trave (I_1, I_2, I_3, I_4) na fertilnost ženki gubara
2. razlikuju li se tri sorte grožđa (S_1, S_2, S_3) po sadržaju ukupnih kiselina
3. utječu li rokovi sjetve (R_1, R_2, R_3, R_4) na prinos šećerne repe
4. razlikuju li se promjeri debla kod stabala 12 sorata jabuka (S_1, S_2, ... , S_{12})
5. da li 9 pripravaka (P_1, P_2, ... , P_9) ima istu učinkovitost u suzbijanju krumpirove zlatice
6. razlikuju li se prinosi 4 sorte jagoda (S_1, S_2, S_3, S_4)
7. kakav je učinak 4 različite varijante gnojidbe (G_1, G_2, G_3, G_4) na prinos sijena
8. jesu li prinosi 5 sorata jarog stočnog graška (S_1, S_2, S_3, S_4, S_5) različiti
9. da li se 10 novih hibrida duhana (H_1, H_2, ... , H_{10}) razlikuje po prinosu suhog lista
10. kako 8 varijanata gnojidbe (G_1, G_2, ... , G_8) utječe na sadržaj N u biljci radiča.

Sve su to itekako zanimljiva pitanja, no poljoprivrednog stručnjaka nerijetko zanima i više od toga. Tako bismo svaki od navedenih problema mogli proširiti dodatnim zanimljivim faktorima. Naime:

1. uz ishranu gusjenica različitim travama (I_1, I_2, I_3, I_4) na fertilnost ženki gubara mogla bi utjecati i temperatura (T_1 - niža, T_2 - viša)
2. na sadržaj ukupnih kiselina kod 3 sorte grožđa (S_1, S_2, S_3) mogla bi utjecati i različita gnojidba (G_1, G_2, G_3)
3. utječe li rok sjetve (R_1, R_2, R_3, R_4) i gustoća sjetve (G_1, G_2, G_3) na prinos dvije različite sorte (S_1, S_2)
4. hoće li se promjer debla kod 12 sorata (S_1, S_2, ... , S_{12}) mijenjati ako se one prate kroz 2 uzgojna oblika (O_1, O_2)
5. da li će učinkovitost 9 pripravaka (P, P_2, . . . , P_9) za suzbijanje krumpirove zlatice ovisiti o primijenjenim koncentracijama (K_1, K_2, K_3) i načinu primjene (N_1 - prskanje, N_2 - zaprašivanje)
6. hoće li prinosi 4 sorte jagoda (S_1, S_2, S_3, S_4) biti uvjetovani uzgojem na crnoj polietilenskoj foliji (F_1 - bez folije, F_2 - s folijom)

7. kakav će biti učinak 4 varijante gnojidbe (G_1, G_2, G_3, G_4) na prinos sijena ako se unesu u tlo u dva roka (R_1, R_2)
8. kakvi će biti prinosi 5 sorata jarog stočnog graška (S_1, S_2, ... , S_5) ako se one siju svaka u dva roka (R_1, R_2) i pri tome primijene 3 varijante gnojidbe (G_1, G_2, G_3)
9. hoće li na prinos suhog lista 10 novih hibrida duhana (H_1, H_2, ... , H_{10}) utjecati razmak redova (R_1 - uži, R_2 - širi)
10. da li će osim varijanata gnojidbe (G_1, G_2, ... , G_8) na sadržaj N u biljci radiča utjecati i doze gnojiva (D_1 - niža, D_2 - viša) i vrijeme primjene (V_1 - ranije, V_2 - kasnije)

Proširivši naše zahtjeve, uključili smo dodatne faktore, pa tako u pojedinim slučajevima imamo:

1. ishrana (I_1, I_2, I_3, I_4) i
   temperatura (T_1 - niža, T_2 - viša)
2. sorta (S_1, S_2, S_3)
   gnojidba (G_1, G_2, G_3)
3. sorta (S_1, S_2)
   gustoća sjetve (G_1, G_2, G_3)
   rok sjetve (R_1, R_2, R_3, R_4)
4. sorta (S_1, S_2, ... , S_{12})
   uzgojni oblik (O_1, O_2)
5. pripravak (P_1, P_2, . . . , P_9)
   koncentracija (K_1, K_2 K_3)
   način primjene (N_1 - prskanje, N_2 - zaprašivanje)
6. sorta (S_1, S_2, S_3, S_4)
   folija (F_1 - bez, F_2 - s folijom)
7. gnojidba (G_1, G_2, G_3, G_4)
   rokovi primjene (R_1, R_2)
8. sorta (S_1, S_2, . . . , S_5)
   rok sjetve (R_1, R_2)
   gnojidba (G_1, G_2, G_3)
9. hibridi (H_1, H_2, ... , H_{10})
   razmak redova (R_1 - uži, R_2 - širi)
10. gnojidba (G_1, G_2, ... , G_8)
    doza (D_1 - niža, D_2 - viša)
    vrijeme primjene (V_1 - ranije, V_2 - kasnije)

U svim ovim primjerima je vidljivo da je svaki faktor zastupljen s manjim ili većim brojem mogućnosti koje nazivamo **stepenice** ili **gradacije** ili **razine faktora**.

Tako bismo svaki primjer mogli pisati kao:

	faktori i stepenice faktora	broj kombinacija	broj interakcija
1.	4 x 2	8	1
2.	3 x 3	9	1
3.	2 x 3 x 4	24	4
4.	12 x 2	24	1
5.	9 x 3 x 2	54	4
6.	4 x 2	8	1
7.	4 x 2	8	1
8.	5 x 2 x 3	30	4
9.	10 x 2	20	1
10.	8 x 2 x 2	32	4

U svim ovim slučajevima radi se o više faktora čiji se učinci mogu u pokusu istovremeno proučiti.

Jednako tako je moguće proučiti i sve kombinacije koje proizlaze iz faktora i njihovih stepenica.

Kombinacije predstavljaju umnožak gradacija (stepenica) u pokus uključenih faktora.

Broj kombinacija u višefaktorijalnom pokusu ovisi, dakle, o broju faktora i njihovih stepenica.

Spomenimo još da se u slučajevima kad faktori imaju isti broj gradacija, koriste i takve oznake gdje eksponent predstavlja broj faktora, a osnovica broj stepenica tih faktora. Na primjer:

3 x 3 može se pisati kao 3^2, a znači da su 2 faktora (što označava eksponent) svaki sa po 3 stepenice

2 x 2 x 4 ili $2^2 \times 4$ znači 3 su faktora od kojih prva dva imaju po dvije, a treći četiri stepenice

U višefaktorijalnim pokusima kombinacije su članovi ili tretiranja pokusa.

Tako u našem prvom primjeru s ishranom gusjenica gubara različitim travama i na različitoj temperaturi, imamo 8 članova pokusa – 8 mogućih kombinacija koje proizlaze iz stepenica faktora.

To su:

I_1T_1 I_2T_1 I_3T_1 I_4T_1

I_2T_1 I_2T_2 I_3T_2 I_4T_2

U drugom primjeru sa 3 sorte grožđa i 3 gnojidbe (dakle 3^2) kombinacije su:

S_1G_1 S_1G_2 S_1G_3

S_2G_1 S_2G_2 S_2G_3

S_3G_1 S_3G_2 S_3G_3

Za zadnji primjer gnojidbe radiča s 9 varijanata gnojiva, primijenjenih u po 2 doze i u različito vrijeme (dakle 8 x 2^2) moguće kombinacije su:

1. $G_1D_1V_1$
2. $G_1D_1V_2$
3. $G_1D_2V_1$
4. $G_1D_2V_2$

.
.
.

32. $G_8D_2V_2$

U višefaktorijalni pokus moraju se dakako uključiti sve moguće kombinacije faktora i njihovih stepenica.

Zbog toga takvi pokusi omogućavaju proučavanje i testiranje djelovanja svih faktora, ali jednako tako i njihovog zajedničkog djelovanja koje se očituje kroz kombinacije.

Naime, faktori mogu djelovati potpuno nezavisno, ali i zajednički, što zovemo interakcijom. **Interakcija je zajedničko zavisno djelovanje dvaju ili više faktora**.

U višefaktorijalnom pokusu su zato, ovisno o broju faktora, moguće interakcije dva, tri i više faktora.

Broj interakcija ovisi isključivo o broju faktora (bez obzira koliko stepenica svaki faktor imao). Zato u svakom dvofaktorijalnom pokusu (s faktorima A i B, a bez obzira na njihove stepenice) postoji samo jedna interakcija koju označavamo A x B. Takvu situaciju imamo u našim primjerima 1, 2, 6, 7 i 9.

Interakcija dva faktora zove se jednostruka interakcija ili interakcija prvog reda.

Zajedničko djelovanje tri faktora zovemo dvostrukom interakcijom ili interakcijom drugog reda, a označava se kao A x B x C. (u primjerima 3, 4, 5, 8 i 10)

Kako je, dakle, broj interakcija uvjetovan brojem faktora:

- u dvofaktorijalnom pokusu ima:
  1 jednostruka interakcija:
  A x B
- u trofaktorijalnom pokusu ima:
  3 jednostruke interakcije:
  A x B, A x C, B x C
  1 dvostruka interakcija:
  A x B x C
- u četverofaktorijalnom pokusu ima:
  6 jednostrukih interakcija:
  A x B, A x C, A x D, B x C, B x D, C x D
  4 dvostruke interakcije:
  A x B x C, A x B x D, A x C x D, B x C x D
  1 trostruka interakcija:
  A x B x C x D

Najvažnija prednost višefaktorijalnih pokusa u odnosu na jednofaktorijalne je mogućnost otkrivanja i testiranja interakcija. U višefaktorijalnom pokusu može se procijeniti i testirati učinke svakog faktora posebice, ali isto tako i njihove zajedničke učinke ili interakcije.

Interakcija se, naime, javlja onda, kad je rezultat djelovanja jednog (ili više) faktora modificiran djelovanjem drugog (drugih) faktora.

Zato jednim višefaktorijalnim pokusom možemo otkriti jednostavne učinke svakog faktora, glavne učinke svakog faktora, te njihove interakcije.

Jednostavni učinci (efekti) su razlike između stepenica jednog faktora, u svakoj stepenici drugog faktora.

Pod **glavnim učinkom** faktora podrazumijevamo promjene izazvane stepenicama pojedinog faktora. To je zapravo prosjek jednostavnih učinaka.

Uzmimo primjer 2^2 pokusa (faktori A i B svaki s po 2 stepenice). Neka su četiri moguće kombinacije predstavljene vrijednostima:

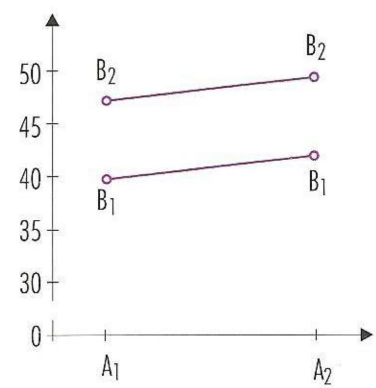
	B_1	B_2	
A_1	40	47	+7
A_2	42	49	+7
	+2	+2	

Jednostavni učinci faktora B
($B_1 - B_2$ unutar A_1 i $B_1 - B_2$ unutar A_2)

Jednostavni učinci faktora A
($A_1 - A_2$ unutar B_1 i $A_1 - A_2$ unutar B_2)

Interakcija, zapravo, opisuje razliku u učinku s obzirom na promjene stepenica jednog faktora unutar stepenice drugog faktora.

Grafički bi se ona za ovaj primjer mogla prikazati kao



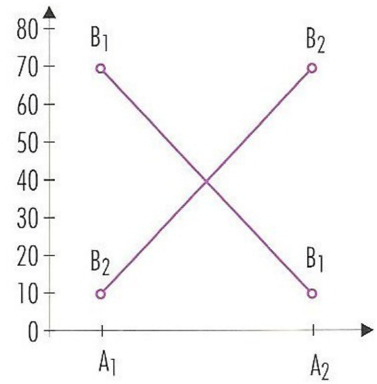
U svakoj stepenici faktora A ista je promjena vrijednosti s obzirom na stepenice faktora B. I u A_1 i u A_2 vrijednost od B_1do B_2 se povećala za 7. Dakle promjene uvjetovane gradacijama jednog faktora istovjetne su za obje stepenice drugog faktora.

Slična je situacija i obratno: u svakoj stepenici faktora B razlike između A_1 i A_2 su 2.

Ovo je, zapravo, slučaj kad interakcije nema, pa je grafički predstavljena paralelnim linijama.

Kad bismo, recimo, imali ove vrijednosti kombinacija

	B_1	B_2	
A_1	70	10	−60
A_2	10	70	+60
	−60	+60	

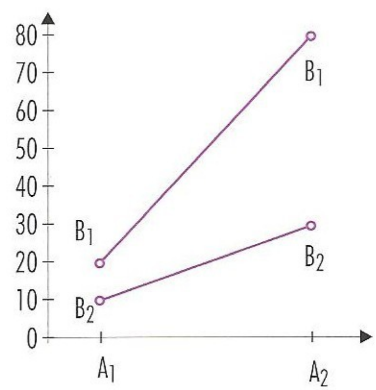


U A_1, prvoj stepenici faktora A vrijednost se od B_1 do B_2 smanjila za 60, a u drugoj stepenici A_2 povećala za isti iznos.

Taj primjer ilustrira slučaj potpune interakcije s razlikom u smjeru.

Ako su vrijednosti kombinacija

	B_1	B_2	
A_1	20	10	−10
A_2	80	30	−50
	+60	+20	



različit je iznos smanjenja ili povećanja u nekoj stepenici faktora A, s obzirom na stepenice faktora B (u A_1 razlika između B_1 i B_2 je 10, a u A_2 ona je 50). Tu se radi o **interakciji s promjenom iznosa**.

Da zaključimo: **među faktorima može, dakle, postojati interakcija s obzirom na smjer i iznos promjene**.

Faktorijalni pokusi pružaju tako mogućnost otkrivanja i testiranja djelovanja svakog od uključenih faktora, a isto tako i njihovih interakcija. Pri tom treba naglasiti da čak ako interakcije nema, višefaktorijalnim pokusom ne gubi se ništa od informacija o faktorima.

Signifikantna interakcija znači da faktori ne djeluju neovisno, već učinak jednog ovisi o drugom.

U takvom bi slučaju odvojeni jednofaktorijalni pokusi s pojedinim faktorima dali pogrešnu informaciju.

Ako interakcija nije opravdana, znači da faktori djeluju nezavisno jedan o drugom, pa bi se i pojedinačnim jednofaktorijalnim pokusima došlo do jednako točne interpretacije. Tako se ne gubi na pouzdanosti o djelovanju faktora, a dobiva se na vremenu.

Kad govorimo o interakciji, treba naglasiti da se povećanjem broja faktora uvelike povećava broj mogućih interakcija (s 2 faktora jedna je interakcija, s 3 faktora već su 4 interakcije, a 4 faktora rezultiraju s 11 mogućih interakcija, itd.). Time se ujedno povećava opasnost od "prekrivanja" djelovanja i nemogućnosti ispravnog otkrivanja interakcije, a napose njihove interpretacije.

Iz tog razloga korisna praktična preporuka bi bila: u pokusu se ograničiti na najviše 3 faktora.

Vratimo se opet na članove višefaktorijalnog pokusa, a to su kombinacije.

Kako broj kombinacija ovisi o broju faktora i broju stepenica faktora, to pri njihovom izboru treba biti oprezan.

Već smo kod jednofaktorijalnih pokusa naglašavali da ako je broj tretiranja vrlo velik, postoji opasnost da pokus postane preglomazan, a tada je upitna homogenost uvjeta za sve članove. Isto, dakako, vrijedi i kod višefaktorijalnih pokusa.

Ovisno o broju kombinacija (=članova = tretiranja) i veličini osnovne parcele, odabire se najpogodnija shema pokusa.

I za višefaktorijalne pokuse vrijede iste one sheme koje se primjenjuju kod jednofaktorijalnih pokusa (slučajni raspored, slučajni blokni raspored, latinski kvadrat, latinski pravokutnik).

Tako bi kod onih navedenih 10 primjera, s obzirom na broj kombinacija mogli primijeniti, recimo:

- slučajni blokni raspored ili latinski kvadrat → za primjere 1, 2 i 6
- slučajni blokni raspored (s recimo 4 repeticije) → za sve ostale primjere
- latinski pravokutnik (s 4 repeticije, jer je u svim tim primjerima broj članova - kombinacija djeljiv s 4) → za primjere 3, 4, 8 i 9

Uvijek ćemo odabrati onaj raspored ili plan pokusa koji osigurava najmanju pogrešku pokusa.

I konačno o analizi podataka višefaktorijalnog pokusa, koja bi se mogla općenito opisati u smislu testiranja utjecaja faktora i interakcije:

Testiranje prosječnih vrijednosti kombinacija (tretiranja u pokusu):

H_0 : nema razlika između prosječnih vrijednosti kombinacija, tj.

$$H_0 : \overline{x}_{komb.1} = \overline{x}_{komb.2} = \ldots = \overline{x}_{komb.n}$$

$$Test : F_{exp} = \frac{s^2_{kombinacija}}{s^2_{pogreške}}$$

Odbacivanje $H_0 : F_{exp} > F_{tab}$ ($F_{tab.}$ baziran na n-1 kombinacija i n-1 pogreške)

Testiranje interakcija:

H_0 : faktori djeluju nezavisno - nema interakcija

$$Test : F_{exp} = \frac{s^2_{interakcije}}{s^2_{pogreške}}$$

Odbacivanje $H_0 : F_{exp} > F_{tabl}$ (F_{tabl} baziran na n-1 interakcije i n-1 pogreške)

Usporedba prosječnih vrijednosti kombinacija : t-test $\rightarrow$ $LSD = t_{tabl} \cdot s_D$
(t_{tabl} baziran na n-1 pogreške)

Testiranje svakog faktora:

$$H_0 : \overline{x}_{stepenice\,1} = \overline{x}_{stepenice\,2} = \ldots = \overline{x}_{stepenice\,n}$$

$$Test : F_{exp} = \frac{s^2_{faktora}}{s^2_{pogreške}}$$

Odbacivanje $H_0 : F_{exp} > F_{tabl}$ (F_{tabl} baziran na n-1 pojedinog faktora i n-1 pogreške)

Usporedba $\overline{x}_{stepenica}$ svakog faktora : t-test $\rightarrow$ $LSD = t_{tabl} \cdot s_D$
(t_{tabl} baziran na n-1 pogreške s_D se izračuna na osnovi broja repeticija i stepenica faktora)

U principu se analiza provodi u dva dijela (dvije etape): prvi je istovjetan analizi jednofaktorijalnog pokusa, a u drugom se tretiranja (a to su sada kombinacije) raščlanjuju na faktore i interakcije.

Dijelovi varijabilnosti koji se definiraju u pojedinom dijelu analize, nulte hipoteze koje se testiraju i testovi koji se koriste:

	I dio		**II dio**
		$SS_{repeticije}$	
Dijelovi	SS_{ukupno}	$SS_{kombinacija}$	$SS_{faktora:}$
varijabilnosti		$SS_{pogreške}$	A
			B
			C
			.
			.
			.
			$SS_{interakcija:}$
			A x B
			A x C
			B x C

Općenita shema analize podataka bilo kojeg (svakog) višefaktorijalnog pokusa izgledala bi:

I. dio

SS_{Ukupno}			
$SS_{Repeticija}$			
$SS_{Kombinacija}$	F-test $_{kombinacija}$	Ako F ns →	nije utvrđena opravdana različitost među kombinacijama, dakle nema signifikantnog djelovanja faktora
$SS_{Pogreške}$			
		Ako F * →	prelazi se na II. dio analize

II. dio

SS_{Faktora}:

A B	F-test svakog faktora	F ns → nema značajnog djelovanja faktora F * → t-testom treba testirati razlike među prosječnim vrijednostima svakog faktora

$SS_{\text{Interakcija}}$:

A x B A x C	F-test svake interakcije	F ns → faktori djeluju neovisno F * → t-testom treba testirati razlike između prosječnih vrijednosti kombinacija faktora koji su u interakciji

U I. dijelu analize ukupna se varijabilnost dakle, ovisno o izvorima, raščlanjuje na repeticije (jedno ili dvosmjerne), tretiranja (kombinacije), definira se pogreška pokusa i provede se F-test kombinacija.

Signifikantni F-test kombinacija znači da neki od faktora (ne nužno svi, iako je i to moguće) ili samostalno ili zajednički s drugim(a) uvjetuje značajne razlike u svojstvu koje se proučava. Koji su to faktori i koje interakcije, saznaje se u II. dijelu analize, kad se kombinacije raščlanjuju na faktore i interakcije i provede F-test za svaki faktor i svaku interakciju.

Za one faktore za koje je F-test opravdan (a to znači njihovo je djelovanje značajno) dalje se provodi t-test kojim se usporede prosječne vrijednosti stepenica faktora, npr. $\bar{x}_{A_1} - \bar{x}_{A_2}$ ili $\bar{x}_{B_1} - \bar{x}_{B_2}$ itd. Svaki faktor ima onoliko prosječnih vrijednosti, koliko ima stepenica. To je, dakako, nepotrebno u slučaju da faktor ima samo dvije gradacije, jer tada već F-test testira tu jednu jedinu razliku među njima.

Neopravdan F-test faktora je dokaz da djelovanja faktora nije značajno.

Svaka opravdana interakcija znači da faktori ne djeluju neovisno, nego zavisno i zajednički.

Zbog opisanih mogućnosti, posebno treba naglasiti interpretaciju rezultata analize višefaktorijalnih pokusa. Naime, kad se utvrdi signifikantna interakcija, glavna djelovanja pojedinih faktora koji su u nju uključeni gotovo da nemaju smisla, bez obzira što su eventualno njihovi F-testovi opravdani. To je zbog toga što se razlike između stepenica jednog faktora mijenjaju od jedne do druge stepenice drugog, pa bi prosjek dobiven iz takvih vrijednosti imao malo smisla. U takvom slučaju najviše smisla ima usporedba prosječnih vrijednosti onih kombinacija preko kojih se manifestira određena opravdana interakcija.

Kad faktori djeluju neovisno (kad interakcije nema) ispravno je uspoređivati prosječne vrijednosti stepenica svakog faktora i testirati razlke među njima.

Kao i do sada, preko različitih primjera ćemo nastojati razumjeti i naučiti o faktorijalnim pokusima, analizi podataka takvih pokusa i interpretaciji rezultata analize.

10.1. Dvofaktorijalni pokusi

Ako faktore označimo s A i B, a broj njihovih stepenica s n_A i n_B, broj repeticija s n_r, tada je općenita shema analize dvofaktorijalnog pokusa:

	Izvori varijabilnosti	$n - 1$
I.	Ukupno	$n_A \cdot n_B \cdot n_r - 1$
	Repeticije	$n_r - 1$
	Kombinacije	$n_A \cdot n_B - 1$
	Pogreška	$(n_r - 1)(n_A \cdot n_B - 1)$
II.	Faktori:	
	A	$n_A - 1$
	B	$n_B - 1$
	Interakcija A x B	$(n_A - 1)(n_B - 1)$

Nakon signifikantnih F testova faktora i interakcije, za provođenje t-testa izračunat će se standardne pogreške srednje vrijednosti po formulama:

$$\text{za faktor A} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_B \cdot n_r}}$$

$$\text{za faktor B} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_A \cdot n_r}}$$

$$\text{za interakciju A x B} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r}}$$

PRIMJER 10.1.1. **2 x 2 (2^2); slučajni blokni raspored)**

Proveden je pokus s dvije sorte grožđa, Graševinom i Moslavcem, uzgajanima u kordoncu s 1.0 (R_1) i 2.0 m (R_2) razmaka u redu, a s ciljem utvrditi utječu li ovi faktori (sorta, razmak) na prinos grožđa.

Iz faktora i njihovih stepenica proizlaze 4 kombinacije:

$GR_1 \quad GR_2 \quad MR_1 \quad MR_2$

Pokus s ova 4 tretiranja bio je postavljen po shemi slučajnog bloknog rasporeda u 6 ponavljanja.

Plan pokusa i podaci o prinosu grožđa (dt/ha):

I.	GR_1	GR_2	MR_1	MR_2
	85.99	70.97	99.66	90.46
II.	GR_2	MR_1	MR_2	GR_1
	91.30	113.99	74.80	94.32
III.	MR_2	MR_1	GR_1	GR_2
	118.95	140.32	74.99	84.13
IV.	MR_1	GR_1	GR_2	MR_2
	178.31	70.99	119.62	144.77
V.	MR_2	GR_2	MR_1	GR_1
	88.30	85.63	135.32	110.25
VI.	GR_1	MR_1	MR_2	GR_2
	115.00	120.00	162.00	125.00

Prije analize praktično je podatke srediti po kombinacijama i repeticijama:

	Kombinacije				
Repeticije	GR_1	GR_2	MR_1	MR_2	$\sum x$ Repeticije
I.	85.99	70.97	99.66	90.46	347.08
II.	94.32	91.30	113.99	74.80	374.41
III.	74.99	84.13	140.32	118.95	418.39
IV.	70.99	119.62	178.31	144.77	513.69
V.	110.25	85.63	135.32	88.30	419.50
VI.	115.00	125.00	120.00	162.00	522.00
$\sum x$ Kombinacije	551.54	576.65	787.60	679.28	$\sum x = 2595.07$

I. dio ANOVA

$$SS_{\text{Ukupno}} = 85.99^2 + \ldots + 162.00^2 - \frac{2595.07^2}{24} = 19498.79$$

$$SS_{\text{Repeticija}} = \frac{347.08^2 + \ldots + 522.00^2}{4} - \frac{2595.07^2}{24} = 6410.29$$

$$SS_{\text{Kombinacija}} = \frac{551.54^2 + ... + 679.28^2}{6} - \frac{2595.07^2}{24} = 5809.93$$

$$SS_{\text{Pogreške}} = SS_{\text{Ukupno}} - SS_{\text{Repeticija}} - SS_{\text{Kombinacija}} = 7278.57$$

Konstruirat ćemo tablicu ANOVA-e, u nju unijeti dobivene vrijednosti, te odrediti pripadajući broj slobodnih varijanata, izračunati varijance i provesti F-test kombinacija.

Tablica ANOVA

Izvori varijabilnosti	$n-1$	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	(6·4)–1=23	19498.79				
Repeticije	6–1=5	6410.29				
Kombinacije	4–1=3	5809.93	1936.64	3.99**	3.25	5.42
Pogreška	23–8=15	7278.57	485.24			
Faktori:						
Sorta	2–1=1	4779.62	4779.62	9.85**	4.54	8.68
Razmak	2–1=1	288.50	288.50	0.59 ns	4.54	8.68
Interakcija S x R	1·1=1	741.82	741.82	1.53 ns	4.54	8.68

Visokosignifikantan F-test kombinacija znači da ima opravdanih razlika u prinosu grožđa, no da li je razlogom za to sorta ili razmak sadnje ili njihovo zajedničko djelovanje, tek treba utvrditi raščlanjivanjem kombinacija.

II. dio ANOVA

Da bi se moglo raščlaniti kombinacije, potrebno je konstruirati tzv. dvosmjernu tablicu kombinacija, prema stepenicama svakog od faktora.

Sorta - Razmak

	R_1	R_2	$\sum x$ Sorte	$\bar{x}$
Graševina	551.54	576.65	1128.19	94.02
Moslavac	787.60	679.28	1466.88	122.24
$\sum x$ Razmaka	1339.14	1255.93	$\sum x$ = 2595.07	
$\bar{x}$ Razmaka	111.59	104.66		

Različitost kombinacija uvjetovana je glavnim djelovanjem svakog faktora kao i njihovim zajedničkim djelovanjem - interakcijom.

Zato se $SS_{\text{kombinacija}}$ (koju smo izračunali u I. dijelu analize) može sastojati samo od onog dijela varijabilnosti koji je uvjetovan djelovanjima faktora (dakle SS_{Sorte} i SS_{Razmaka}) i njihovom interakcijom (SS_{SxR}).

U tom smislu treba raščlaniti kombinacije, pri čemu pomaže dvosmjerna tablica. Iz nje, dakle, računamo:

$$SS_{\text{Sorte}} = \frac{1128.19^2 + 1466.88^2}{n_r \cdot n_R} - \frac{2595.07^2}{24} = 4779.62$$

$$SS_{\text{Razmaka}} = \frac{1339.14^2 + 1255.93^2}{n_r \cdot n_S} - \frac{2595.07^2}{24} = 288.50$$

$$SS_{\text{Interakcije SxR}} = SS_{\text{Kombinacija}} - SS_S - SS_R =$$

$$= 5809.93 - 4779.62 - 288.49 = 741.82$$

(Dijelili smo s 12 jer je broj u svakoj stepenici i faktora S i faktora R zapravo zbroj od 2 x 6 osnovnih podataka)

Izračunane vrijednosti za *SS*-ove unesemo u tablicu ANOVA-e, kako bismo izračunali varijance i proveli F-test za svaki faktor i interakciju. Baš kao što smo $SS_{\text{kombinacija}}$ podijelili na SS_S, SS_R i SS_{SxR}, isto tako je i sa slobodnim varijantama kombinacija (3) koje se dijele na faktore i interakciju (1 + 1 + 1 = 3).

Nakon provedenih F-testova možemo zaključiti:

- interakcija S x R nije signifikantna, što znači da faktori djeluju neovisno jedan o drugome.
- Prinos grožđa ne razlikuje se značajno, ovisno o tome da li je razmak bio 1.0 ili 2.0 m (djelovanje faktora R nije opravdano).
- Signifikantan F-test faktora S znači da prinosi grožđa značajno ovise o sorti.

Kako su u pokus bile uključene samo dvije sorte (Graševina i Moslavac), to signifikantan F-test ujedno znači da je razlika u prosječnom prinosu grožđa ove dvije sorte opravdana, značajna ili signifikantna (u korist sorte Moslavac).

(Podsjetimo se: kad su u pitanju samo dvije prosječne vrijednosti koje se uspoređuju, među njima je jedna jedina razlika, te nakon opravdanog F-testa nije potrebno provoditi t-test, jer već F-test kaže sve o toj razlici).

PRIMJER 10.1.2.

3 x 3 (3^2); slučajni blokni raspored

U proizvodnji gospodarski važnih kultura iznimno je korisno znati sve razloge koji bi mogli uvjetovati što veći prinos. Tako je i kod graha mahunara, kod kojeg prinos tehnološki zrelih mahuna može ovisiti među ostalim i o sorti, odnosno gustoći sjetve. Zato je proveden eksperiment, u kojem su bile tri sorte: Favorit (F), Procesor (P) i Top crop (T) sijane svaka na tri razmaka sjetve (R_1 = 60x4.2 cm, R_2 = 30x8.3 cm, R_3 = 15x16,7 cm,) da se utvrdi da li se prinosi tehnološki zrelih mahuna razlikuju ovisno o sorti, o gustoći sjetve ili pak o njihovom zajedničkom utjecaju (interakciji).

Devet mogućih kombinacija sijano je u pokus po planu slučajnog bloknog rasporeda u 5 ponavljanja.

FR_1	FR_2	FR_3
PR_1	PR_2	PR_3
TR_1	TR_2	TR_3

Plan pokusa i prinosi u dt/ha:

I.	FR_1	TR_1	PR_1	TR_3	FR_2	PR_3	TR_2	FR_3	PR_2
	43.5	87.3	86.7	78.0	122.3	75.0	91.7	99.8	118.3
II.	PR_2	TR_2	TR_1	FR_1	PR_3	FR_3	PR_1	TR_3	FR_2
	96.2	86.5	74.2	86.3	84.3	97.0	93.7	95.7	110.7
III.	TR_3	FR_2	TR_2	PR_1	FR_3	PR_2	FR_1	TR_1	PR_3
	57.5	114.7	102.8	91.4	88.7	81.7	79.5	84.8	107.2
IV.	FR_3	PR_2	TR_1	FR_1	PR_3	TR_3	FR_2	TR_2	PR_1
	97.0	95.2	84.5	93.3	68.5	68.3	94.2	92.2	96.3
V.	PR_3	TR_2	FR_2	PR_1	TR_1	FR_3	PR_2	TR_3	FR_1
	64.7	93.2	94.7	83.3	87.7	94.3	87.3	77.7	83.2

Podaci svrstani po kombinacijama i repeticijama:

Repeticije	FR_1	FR_2	FR_3	PR_1	PR_2	PR_3	TR_1	TR_2	TR_3	$\sum x$ Repeticije
I.	43.5	122.3	99.8	86.7	118.3	75.0	87.3	91.7	78.0	807.6
II.	86.3	110.7	97.0	93.7	96.2	84.3	74.2	86.5	95.7	824.6
III.	79.5	114.7	88.7	91.4	81.7	107.2	84.8	102.8	57.5	808.3
IV.	93.3	94.2	97.0	96.3	95.2	68.5	84.5	92.2	68.3	789.5
V.	83.2	94.7	94.3	83.3	87.3	64.7	87 7	93.2	77.7	766.1
$\sum x$ Kombinacije	385.8	536.6	476.8	451.4	478.7	399.7	418.5	466.4	377.2	$\sum x = 3991.1$

I. dio ANOVA

$$SS_{\text{Ukupno}} = 43.5^2 + \ldots + 77.7^2 - \frac{3911.1^2}{45} = 9679.876$$

$$SS_{\text{Repeticija}} = \frac{807.6^2 + \ldots + 766.1^2}{9} - \frac{3911.1^2}{45} = 213.825$$

$$SS_{\text{Kombinacija}} = \frac{385.8^2 + \ldots + 377.2^2}{5} - \frac{3911.1^2}{45} = 4374.152$$

$$SS_{\text{Pogreške}} = SS_{\text{Ukupno}} - SS_{\text{Repeticija}} - SS_{\text{Kombinacija}} = 5091.899$$

(u izrazima $SS_{\text{Repeticija}}$ i $SS_{\text{Kombinacija}}$ dijeli se s 9 odnosno s 5, jer svaku repeticiju predstavlja zbroj osnovnih podataka svih 9 kombinacija u njoj, a svaku kombinaciju zbroj podataka kroz svih 5 repeticija).

U tablicu ANOVA unosimo dobivene vrijednosti, odredimo pripadajući broj slobodnih varijanata, izračunamo varijance i provedemo F-test kombinacija.

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	44	9679.876				
Repeticije	4	213.825				
Kombinacije	8	4374.152	546.77	3.44 **	2.27	3.17
Pogreška	32	5091.899	159.12			
Faktori:						
Sorta	2	626.579	313.29	1.97 ns	3.32	5.39
Razmak	2	2290.311	1145.15	7.20 **	3.32	5.39
Interakcija S x R	4	1457.262	364.31	2.29 ns	2.69	4.02

F-test kombinacija je signifikantan, što upućuje na raščlanjivanje kombinacija na faktore i interakciju.

Prelazimo, dakle, na II. dio ANOVA uz prethodno konstruiranje dvosmjerne tablice:

Sorta - Razmak

	R_1	R_2	R_3	$\sum x$ Sorte	$\overline{x}$
F	385.8	536.6	476.8	1399.2	93.28
P	451.4	478.7	399.7	1329.8	88.65
T	418.5	466.4	377.2	1262.1	84.14
$\sum x$ Razmaka	1255.7	1481.7	1253.7	$\sum x = 3991.1$	
$\overline{x}$ Razmaka	83.71	98.78	83.58		

$$SS_{\text{Sorte}} = \frac{1399.2^2 + 1329.8^2 + 1262.1^2}{5 \cdot 3} - \frac{3991.1^2}{45} = 626.579$$

$$SS_{\text{Razmaka}} = \frac{1255.7^2 + 1481.7^2 + 1253.7^2}{5 \cdot 3} - \frac{3991.1^2}{45} = 2290.311$$

$$SS_{\text{Interakcije SxR}} = SS_{\text{Kombinacija}} - SS_S - SS_R =$$

$$= 4374.152 - 626.579 - 2290.311 = 741.82$$

Izračunane *SS* vrijednosti stavimo u tablicu ANOVA, pridodamo pripadajući broj slobodnih varijanata za svaki faktor i interakciju (raščlanimo, zapravo, slobodne varijante kombinacija), izračunamo varijance, provedemo F-test za faktore i interakciju.

Rezultati provedenih F-testova dopuštaju nam zaključiti:

- F-test za faktor Sorta i interakciju S x R nije signifikantan. To znači da faktori ne djeluju zavisno jedan o drugom (interakcije nema), a Sorta kao faktor ne utječe na različitost prinosa. Favorit ima, doduše, najveći prosječni prinos, nešto manji Processor, a Top crop najmanji, ali sve su te razlike neopravdane.
- Razmak sjetve je, međutim, faktor o kojem značajno ovisi prinos. Iz prosječnih prinosa po razmacima sjetve čini se da je s gustoćom sjetve R_2 postignut najveći prinos mahuna, no da li je to tako treba ustanoviti t-testom, testirajući razlike između ovih prosječnih vrijednosti.

t-test za faktor Razmak sjetve

$$LSD = t \cdot s_D$$

$$s_D = \sqrt{\frac{2 \cdot 159.12}{5 \cdot 3}} = 4.606 \text{ (iz formula na str. 201)}$$

$t_{P=5\%} = 2.04$ (očitan iz 32 slobodne varijante pogreške)

$t_{P=1\%} = 2.75$

$$LSD_{P=5\%} = 2.04 \cdot 4.606 = 9.40 \text{ dt/ha}$$

$$LSD_{P=1\%} = 2.75 \cdot 4.606 = 12.67 \text{ dt/ha}$$

Usporedimo li prosječne prinose postignute u različitim gustoćama sjetve:

$$98.78\ (R_2) - 83.71\ (R_1) = 15.07^{**} \text{ dt/ha}$$

$$98.78\ (R_2) - 83.58\ (R_3) = 15.20^{**} \text{ dt/ha}$$

zaključujemo da je visoko signifikantno najveći prinos tehnološki zrelih mahuna graha postignut s razmakom sjetve označenim kao R_2. To bi, dakle, trebalo preporučiti za praksu i proizvodnju.

PRIMJER 10.1.3.

2 x 4; slučajni blokni raspored

Proveden je pokus s dvije sorte mrkve: Amsterdamska (A) i Nantes (N), sijane svaka u četiri gustoće sjetve ovisno o količini sjemena sijanog po hektaru: G_1 (4.0 kg/ha). G_2 (8.0 kg/ha), G_3 (12.0 kg/ha) i G_4 (18.0 kg/ha).

Cilj je bio utvrditi ima li među sortama kakvih razlika u prinosu korijena, utječu li različite gustoće sjetve na različitost prinosa, odnosno jesu li ovi faktori neovisni ili postoji interakcija među njima.

Osam kombinacija ovog 2 x 4 pokusa je:

$AG_1 \quad AG_2 \quad AG_3 \quad AG_4$

$NG_1 \quad NG_2 \quad NG_3 \quad NG_4$

Pri izvođenju pokusa korištena je shema slučajnog bloknog rasporeda s 5 ponavljanja.

Plan pokusa i podaci o prinosu korijena (u kg/5.0 m^2):

I.	AG_1 11.4	AG_2 14.6	AG_3 15.6	AG_4 16.1	NG_1 17.8	NG_2 18.6	NG_3 19.6	NG_4 24.1
II.	NG_2 21.4	AG_3 9.1	NG_3 23.1	AG_1 4.1	AG_4 23.4	NG_4 24.2	AG_2 5.8	NG_1 16.1
III.	NG_4 25.2	NG_1 22.1	AG_1 9.5	NG_2 17.3	AG_2 14.9	AG_4 24.8	NG_3 29.4	AG_3 21.1
IV.	AG_4 19.3	NG_4 16.0	NG_1 25.0	NG_3 28.4	AG_1 8.6	AG_3 24.0	AG_2 9.9	NG_2 26.0
V.	NG_4 28.3	NG_3 27.9	AG_2 14.6	AG_3 20.7	NG_1 25.0	NG_2 25.2	AG_4 24.3	AG_1 9.0

Podaci sređeni po kombinacijama i repeticijama

Repeticije	AG_1	AG_2	AG_3	AG_4	NG_1	NG_2	NG_3	NG_4	$\sum x$ Repeticije
I.	11.4	14.6	15.6	16.1	17.8	18.6	19.6	24.1	137.8
II.	4.1	5.8	9.1	23.4	16.1	21.4	23.1	24.2	127.2
III.	9.5	14.9	21.1	24.8	22.1	17.3	29.4	25.2	164.3
IV.	8.6	9.9	24.0	19.3	25.0	26.0	28.4	16.0	157.2
V.	9.0	14.6	20.7	24.3	25.0	25.2	27.9	28.3	175.0
$\sum x$ Kombinacije	42.6	59.8	90.5	107.9	106.0	108.5	128.4	117.8	$\sum x = 761.5$

I. dio ANOVA

$$SS_{Ukupno} = 11.4^2 + ... + 28.3^2 - \frac{761.5^2}{40} = 1789.75$$

$$SS_{Repeticija} = \frac{137.8^2 + ... + 175.0^2}{8} - \frac{761.5^2}{40} = 190.44$$

$$SS_{Kombinacija} = \frac{42.6^2 + ... + 117.8^2}{5} - \frac{761.5^2}{40} = 1221.96$$

$$SS_{Pogreške} = 1789.75 - 190.44 - 1221.96 = 377.35$$

Uvrstimo dobivene vrijednosti u tablicu ANOVA-e s pripadajućim brojem slobodnih varijanata, pa izračunajmo varijance i provedimo F-test kombinacija.

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	39	1789.75				
Repeticije	4	190.44				
Kombinacije	7	1221.96	174.56	12.95 **	2.36	3.36
Pogreška	28	377.35	13.48			
Faktori:						
Sorte	1	639.20	639.20	47.43 **	4.20	7.64
Gustoće	3	429.39	143.13	10.82 **	2.95	4.57
Interakcija S x G	3	153.37	51.12	3.79 *	2.95	4.57

F-test kombinacija je signifikantan, pa treba nastaviti analizu i utvrditi koji od faktora ili interakcija su razlogom za to:

II. dio ANOVA

Gustoća - Sorta

	A	N	$\sum x_{\text{Gustoće}}$	$\bar{x}_{\text{Gustoće}}$
G_1	42.6	106.0	148.6	14.86
G_2	59.8	108.5	168.3	16.83
G_3	90.5	128.4	218.9	21.89
G_4	107.9	117.8	225.7	22.57
$\sum x_{\text{Sorte}}$	300.8	460.7	$\sum x = 761.5$	
$\bar{x}_{\text{Sorte}}$	15.04	23.03		

$$SS_{\text{Sorte}} = \frac{300.8^2 + 460.7^2}{5 \cdot 4} - \frac{761.5^2}{40} = 639.20$$

$$SS_{\text{Gustoće}} = \frac{148.6^2 + ... + 225.7^2}{5 \cdot 2} - \frac{761.5^2}{40} = 429.39$$

$$SS_{\text{SxG}} = 1221.96 - 639.20 - 429.39 = 153.37$$

Svi F-testovi u II. dijelu analize su opravdani. To znači da faktori djeluju zavisno jedan o drugom (interakcija je opravdana), odnosno da će kod ovih sorata biti različit učinak količine sjemena korištenog za sjetvu. Zbog toga bi trebalo usporediti prosječne vrijednosti kombinacija.

Tablica prosječnih vrijednosti S x G kombinacija koje se dobiju dijeleći sume svake kombinacije u dvosmjernoj tablici s 5 (jer je svaka kombinacija 5 puta ponovljena):

	A	N
G_1	8.52	21.20
G_2	11.96	21.70
G_3	18.10	25.68
G_4	21.58	23.56

Treba izračunati LSD, kriterij za usporedbu ovih prosjeka. Svaka je kombinacija ponovljena onoliko puta koliko je repeticija, pa će zato s_D biti:

$$s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{5}} = \sqrt{\frac{2 \cdot 13.48}{5}} = 2.32$$

t_{tabl} se očita iz 28 (slobodne varijante pogreške) pa je:

$$LSD_{P=5\%} = 2.05 \cdot 2.32 = 4.76 \text{ kg}$$

$$LSD_{P=1\%} = 2.76 \cdot 2.32 = 6.40 \text{ kg}$$

Prema prosječnim vrijednostima kombinacija, čini se da je najveći prosječni prinos korijena mrkve postignut sa sortom Nantes uz 12 kg sjemena po hektaru (kombinacija NG_3 = 25.68), pa uspoređeno s drugim kombinacijama:

	Razlike, (kg)
25.68 (NG_3) - 8.52 (AG_1) =	17.16 **
25.68 (NG_3) - 11.96 (AG_2) =	13.72 **
25.68 (NG_3) - 18.10 (AG_3) =	7.58 **
25.68 (NG_3) - 21.58 (AG_4) =	4.10 ns
25.68 (NG_3) - 21.20 (NG_1) =	4.48 ns
25.68 (NG_3) - 21.70 (NG_2) =	3.98 ns
25.68 (NG_3) - 23.56 (NG_4) =	2.12 ns

To znači da je prinos sorte Nantes, ako se sije s 12.0 kg sjemena po hektaru, značajno veći od prinosa Amsterdamske (sijane s 4.0 ili 8.0 ili 12.0 kg/ha), a zanemarivo malo različit od prinosa koji se postiže sa sortom Amsterdamska s najvećom količinom sjemena (18.0 kg/ha).

Prinos sorte Nantes je uz sjetvu manjih količina sjemena doduše nešto manji, no te razlike nisu opravdane.

Konačno bi se moglo zaključiti da za postizanje najvećeg prinosa korijena treba sortu Amsterdamska sijati s 18.0 kg sjemena po hektaru, a Nantes s bilo kojom od ovih količina sjemena, jer se kod nje nikakvo značajno povećanje prinosa korijena ne osigurava većom količinom sjemena po hektaru, a to dakako daje prednost najmanjoj (i uz to najjeftinijoj) količini sjemena od 4.0 kg/ha.

Iako faktori uključeni u pokus djeluju zavisno jedan o drugom, i djelovanje svakog od njih je značajno.

Iz prosječnih vrijednosti (u dvosmjernoj su tablici) i opravdanog F-testa zaključujemo da je prinos korijena sorte Nantes (23.03 kg) značajno veći od Amsterdamske (15.04 kg).

Što se faktora G tiče, čini se da se s 18.0 kg sjemena po ha (G_4) postižu najveći prinosi (22.57 kg). Da bismo to provjerili provest ćemo t-test, pri čemu je

$$s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_S}} = \sqrt{\frac{2 \cdot 13.48}{5 \cdot 2}} = 1.64$$

$$LSD_{P=5\%} = 2.05 \cdot 1.64 = 3.36 \text{ kg}$$

$$LSD_{P=1\%} = 2.76 \cdot 1.64 = 4.53 \text{ kg}$$

Razlike, (kg)

$22.57\ (G_4) - 14.86\ (G_1) = 7.71$ **

$22.57\ (G_4) - 16.83\ (G_2) = 5.74$ **

$22.57\ (G_4) - 21.89\ (G_3) = 0.68$ ns

Govoreći samo o faktoru Gustoća, mogli bismo zaključiti da za značajno veći prinos treba sijati 12.0 ili 18.0 kilograma sjemena po ha (G_3 ili G_4), što može dati prednost jeftinijoj varijanti (12.0/ha) jer se s njom postiže zanemarivo malo manji prinos nego s 18.0 kg/ha (G_4).

PRIMJER 10.1.4. **4 x 3; slučajni blokni raspored**

Kako bi se proučilo utječe li gustoća sjetve na prinos različitih sorata kukuruza za silažu, provedena su istraživanja s četiri sorte (Arizona, Iowa, Wisconsin i Vukovarski) i svaka je sijana u 3 gustoće sklopa ovisno o razmaku sjetve (S_1 = 70 cm x 70 cm, S_2 = 70 cm x 20 cm, S_3 = 50 cm x 30 cm).

U ovom je, dakle, pokusu bilo 12 kombinacija:

AS_1	IS_1	WS_1	VS_1
AS_2	IS_2	WS_2	VS_2
AS_3	IS_3	WS_3	VS_3

a pokus je izveden po planu slučajnog bloknog rasporeda u 6 ponavljanja.

Plan pokusa i podaci o prinosu (u dt/ha):

I.	AS_1	AS_2	AS_3	IS_1	IS_2	IS_3	WS_1	WS_2	WS_3	VS_1	VS_2	VS_3
	472	536	608	463	502	558	485	496	547	482	634	587
II.	IS_1	WS_1	WS_2	IS_3	AS_1	WS_3	VS_1	AS_3	VS_2	IS_2	VS_3	AS_2
	467	488	490	486	439	651	494	614	474	419	813	555
III.	WS_3	IS_3	AS_3	IS_2	WS_2	VS_2	AS_2	VS_3	AS_1	WS_1	IS_1	VS_1
	687	466	583	503	480	513	546	520	600	592	624	476
IV.	WS_1	WS_2	AS_2	WS_3	AS_1	VS_1	IS_1	IS_3	VS_2	AS_3	VS_3	IS_2
	608	444	485	593	512	473	400	466	451	620	626	407
V.	AS_3	VS_1	VS_3	IS_3	VS_2	WS_1	IS_2	AS_2	WS_2	AS_1	WS_3	IS_1
	540	412	460	450	589	418	456	611	573	473	542	372
VI.	VS_2	AS_2	IS_1	WS_2	IS_2	AS_3	VS_3	AS_1	VS_1	IS_3	WS_1	WS_3
	480	553	412	462	407	488	493	469	535	422	370	496

Isti podaci sređeni po kombinacijama i repeticijama:

Repeticije	AS_1	AS_2	AS_3	IS_1	IS_2	IS_3	WS_1	WS_2	WS_3	VS_1	VS_2	VS_3	$\sum x$ Repeticije
I.	472	536	608	463	502	558	485	496	547	482	634	587	6370
II.	439	555	614	467	419	486	488	490	651	494	474	813	6390
III.	600	546	583	624	503	466	592	480	687	476	513	520	6590
IV.	512	485	620	400	407	466	608	444	593	473	451	626	6085
V.	473	611	540	372	456	450	418	573	542	412	589	460	5896
VI.	469	553	488	412	407	422	370	462	496	535	480	493	5587
$\sum x$ Kombinacije	2965	3286	3453	2738	2694	2848	2961	2945	3516	2872	3147	3499	$\sum x = 36918$

I. dio ANOVA

$$SS_{\text{Ukupno}} = 472^2 + \ldots + 493^2 - \frac{36918^2}{72} = 454919.50$$

$$SS_{\text{Repeticija}} = \frac{6370^2 + \ldots + 5587^2}{12} - \frac{36918^2}{72} = 57104.67$$

$$SS_{\text{Kombinacija}} = \frac{2965^2 + \ldots + 3499^2}{6} - \frac{36918^2}{72} = 159909.16$$

$$SS_{\text{Pogreške}} = 454919.50 - 57404.67 - 159909.16 = 237905.67$$

Kad se ove vrijednosti unesu u tablicu ANOVA-e uz pripadajući broj slobodnih varijanata i izračunanih varijanci, moguće je testirati varijancu kombinacija:

Tablica ANOVA

Izvori varijabilnosti	$n-1$	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	71	454919.50				
Repeticije	5	57104.67				
Kombinacije	11	159909.16	14537.20	3.36 **	2.02	2.70
Pogreška	55	237905.67	4325.56			
Faktori:						
Sorta	3	69086.83	23028.94	5.32 **	2.79	4.20
Sklop	2	69608.33	34804.17	8.05 **	3.18	5.06
Interakcija S x R	6	21214.00	3535.67	0.82 ns	2.29	3.18

Statistički opravdan F-test kombinacija ukazuje na značajne razlike u prinosu među kombinacijama, no jesu li te razlike uvjetovane sortom ili gustoćom sklopa ili njihovom interakcijom, saznat ćemo raščlanjivanjem kombinacija.

II. dio ANOVA

Sorta - Sklop

	S_1	S_2	S_3	$\sum x_{Sorte}$	$\bar{x}_{Sorte}$
A	2965	3286	3453	9704	539.11
I	2738	2694	2848	8280	460.00
W	2961	2945	3516	9422	523.44
V	2872	3141	3499	9512	528.44
$\sum x_{Sklopa}$	11536	12066	13316	$\sum x = 36918$	
$\bar{x}_{Sklopa}$	480.67	502.75	554.83		

$$SS_{Sorte} = \frac{9704^2 + ... + 9512^2}{6 \cdot 3} - \frac{36918^2}{72} = 69086.83$$

$$SS_{Sklopa} = \frac{11536^2 + ... + 13316^2}{6 \cdot 4} - \frac{36918^2}{72} = 69608.33$$

$$SS_{SxR} = 159909.16 - 69086.83 - 69608.33 = 21214.00$$

Interakcija nije signifikantna, a to znači da faktori nezavisno utječu na prinos. Opravdani F-testovi za oba faktora znak su njihovog signifikantnog učinka. Kako oba faktora imaju više stepenica, to za usporedbu njihovih prosječnih vrijednosti treba provesti t-test.

t-test za faktor Sorta

$$s_D = \sqrt{\frac{2 \cdot 4325.56}{6 \cdot 3}} = 21.92$$

tablični t očitan iz $n - 1 = 50$ (jer u tablici nema 55) je $t_{P=5\%} = 2.01$ i $t_{P=1\%} = 2.68$

$$LSD_{P=5\%} = 21.92 \cdot 2.01 = 44.06 \text{ dt/ha}$$
$$LSD_{P=1\%} = 21.92 \cdot 2.68 = 58.74 \text{ dt/ha}$$

Prema vrijednostima prosječnih prinosa, čini se da je sorta Arizona najprinosnija.

	Razlike, (dt/ha)
539.11 (A) - 460.00 (I) =	79.11 **
539.11 (A) - 523.44 (W) =	15.67
539.11 (A) - 528.44 (V) =	10.67

Nakon usporedbe prosječnih prinosa ostale tri sorte sa Arizonom i nakon provedenog t-testa, vidljivo je da je jedino sorta Iowa opravdano nižeg prinosa, dok su prinosi sorata Wisconsin i Vukovarski nešto niži (ali ne signifikantno).

Dakle, svejedno je koju od ove tri sorte koristiti (jer će prinos biti zanemarivo malo manji od prinosa sorte Arizona), ali ne bi trebalo sijati sortu Iowa zbog njezinog statistički opravdano manjeg prinosa.

t-test za faktor Sklop

$$s_D = \sqrt{\frac{2 \cdot 4325.56}{6 \cdot 4}} = 18.98$$

$$LSD_{P=5\%} = 18.98 \cdot 2.01 = 38.15 \text{ dt/ha}$$
$$LSD_{P=1\%} = 18.98 \cdot 2.68 = 50.87 \text{ dt/ha}$$

Usporedit ćemo prosječne prinose postignute u različitoj gustoći sklopa.

	Razlike, (dt/ha)
554.83 (S_3) - 480.67 (S_1) =	74.16 **
554.83 (S_3) - 502.75 (S_2) =	52.08 **

To znači da se opravdano najveći prinos kukuruza za silažu postiže najgušćom sjetvom, tj. sjetvom na razmak od 50 cm x 30 cm (S_3).

PRIMJER 10.1.5.

2 x 3; latinski kvadrat

Primjena inhibitora rasta može u cvjećarstvu izazvati zanimljive učinke (promjena veličine ili boje cvijeta ili listova i sl.)

Zato se izveo eksperiment s dvije sorte pelargonija Cherie (C) i *Sprinter (S), koje su tretirane s jednim inhibitorom rasta* primijenjenim u dvije koncentracije (K_1 i K_2), a u pokus su uključene i netretirane biljke (K_0).

Iz toga proizlazi 6 kombinacija (CK_0, CK_1, CK_2, SK_0, SK_1 i SK_2), koje su u pokus raspoređene po planu latinskog kvadrata (što dakako zahtijeva 6 dvosmjernih ponavljanja).

Cilj je pokusa bio saznati da li se ove sorte razlikuju po veličini cvata, hoće li eventualno različite koncentracije primijenjenog inhibitora uvjetovati kakve razlike u veličini cvata, odnosno da li će one izazvati iste ili različite učinke kod obje sorte (dakle, ima li interakcije).

Plan pokusa i podaci o promjeru cvata (u cm):

Vodoravne repeticije	Okomite repeticije						$\sum x$ Vod. rep.
	I	II	III	IV	V	VI	
I.	CK_0	CK_2	CK_1	SK_0	SK_2	SK_1	
	7.3	7.4	8.9	9.7	8.6	8.5	50.4
II.	CK_1	SK_1	CK_0	SK_2	CK_2	SK_0	
	7.8	8.8	8.5	8.8	6.8	9.3	50.0
III.	SK_2	SK_0	CK_2	SK_2	CK_0	CK_1	
	8.3	9.1	7.9	8.1	7.8	7.8	49.0
IV.	SK_1	CK_0	SK_0	CK_2	CK_1	SK_2	
	8.7	9.7	9.3	7.6	7.8	9.1	52.2
V.	CK_2	SK_2	SK_1	CK_1	SK_0	CK_0	
	7.6	8.3	8.8	8.9	8.8	8.0	50.4
VI.	SK_0	CK_1	SK_2	CK_0	SK_1	CK_2	
	9.1	8.3	8.6	8.0	8.6	8.5	51.1
$\sum x$ Okom. rep.	48.8	51.6	52.0	51.1	48.4	51.2	$\sum x = 303.1$

Isti podaci sređeni po kombinacijama:

Repeticije	CK_0	CK_1	CK_2	SK_0	SK_1	SK_2	
I.	7.3	8.9	7.4	9.7	8.5	8.6	
II.	8.5	7.8	6.8	9.3	8.8	8.8	
III.	7.8	7.8	7.9	9.1	8.1	8.3	
IV.	9.7	7.8	7.6	9.3	8.7	9.1	
V.	8.0	8.9	7.6	8.8	8.8	8.3	
VI.	8.0	8.3	8.5	9.1	8.6	8.6	
$\sum x_{\text{Kombinacije}}$	49.3	49.5	45.8	55.3	51.5	51.7	$\sum x = 303.1$

I. dio ANOVA

$$SS_{\text{Ukupno}} = 7.3^2 + \ldots + 8.5^2 - \frac{303.1^2}{36} = 16.0164$$

$$SS_{\text{Vodoravnih repeticija}} = \frac{50.4^2 + \ldots + 51.1^2}{6} - \frac{303.1^2}{36} = 0.9614$$

$$SS_{\text{Okomitih repeticija}} = \frac{48.8^2 + \ldots + 51.2^2}{6} - \frac{303.1^2}{36} = 1.9347$$

$$SS_{\text{Kombinacija}} = \frac{49.3^2 + \ldots + 51.6^2}{6} - \frac{303.1^2}{36} = 7.2314$$

$$SS_{\text{Pogreške}} = 16.0164 - 0.9614 - 1.9347 - 7.2314 = 5.8889$$

Unesu li se ove vrijednosti u tablicu ANOVA-e, iz izračunanih varijanci moguće je testirati kombinacije. U ovom je slučaju F-test kombinacija statistički opravdan, a to je znak da se ide u II. dio ANOVA.

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	F_{exp}	F_{tabl}	
					P=5%	P=1%
Ukupno	35	16.0164				
Vodoravne rep.	5	0.9614				
Okomite rep.	5	1.9347				
Kombinacije	5	7.2314	1.4463	4.91 **	2.71	4.10
Pogreška	20	5.8889	0.2944			
Sorte	1	5.3669	5.3669	18.23 **	4.35	8.10
Inhibitori	2	1.6106	0.8053	2.74 *	3.40	5.85
Interakcija S x I	2	0.2539	0.1270	0.43 ns	3.40	5.85

II. dio ANOVA

Dvosmjerna tablica: Sorta - Inhibitor

	K_0	K_1	K_2	$\sum x_{\text{Sorte}}$	$\bar{x}_{\text{Sorte}}$
C	49.3	48.0	47.3	144.6	8.03
S	55.3	51.6	51.6	158.5	8.80
$\sum x_{\text{Inhibitora}}$	104.6	99.6	98.9	$\sum x = 303.1$	
$\bar{x}_{\text{Inhibitora}}$	8.72	8.30	8.24		

$$SS_{\text{Sorte}} = \frac{144.6^2 + 158.5^2}{6 \cdot 3} - \frac{303.1^2}{36} = 5.3669$$

$$SS_{\text{Inhibitora}} = \frac{104.6^2 + 99.6^2 + 98.9^2}{6 \cdot 2} - \frac{303.1^2}{36} = 1.6106$$

$$SS_{\text{SxI}} = SS_{\text{Kombinacija}} - SS_{\text{Sorte}} - SS_{\text{Inhibitora}} =$$

$$= 7.2314 - 5.3669 - 1.6106 = 0.2539$$

Nakon provedenih F-testova u II. dijelu analize, razvidno je da faktori djeluju nezavisno jedan o drugom, da nema interakcije (interakcija nije signifikantna).

Signifikantni F-testovi za oba faktora znače:

- Sorte Cherie i Sprinter su značajno različitog promjera cvata. Razlika od 0.77 cm (tj. 8.80 - 8.30) je opravdana, značajna, signifikantna; Sprinter ima značajno veći cvat od Cherie.
- Primjena inhibitora rasta uvjetovala je značajne promjene u veličini cvata. Kako se iz prosječnih vrijednosti čini, primijenjeni inhibitor rasta uvjetovao je smanjenje promjera cvata. Naime, najveći promjer cvata imale su netretirane biljke (K_0). Da li je to baš tako vidjet ćemo nakon testa.

t-test za faktor Inhibitor

$$s_D = \sqrt{\frac{2 \cdot 0.2393}{6 \cdot 2}} = 0.1997$$

tablični t očitan iz n - 1 pogreške tj. 20

$t_{\text{tabl } P=5\%} = 2.09$

$t_{\text{tabl } P=1\%} = 2.84$

$$LSD_{P=5\%} = 2.09 \cdot 0.1997 = 0.42 \text{ cm}$$

$$LSD_{P=1\%} = 2.84 \cdot 0.1997 = 0.57 \text{ cm}$$

Usporedimo li prosječne vrijednosti:

	Razlike, (cm)
8.72 (K_0) – 8.42 (K_1) =	0.30 ns
8.72 (K_0) – 8.12 (K_2) =	0.60 **

Znači najveći su se cvatovi dobili na biljkama koje nisu bile tretirane inhibitorom rasta (K_0). Primjena inhibitora rasta u koncentraciji označenoj kao K_1 uvjetuje nešto manji, ali ne značajno manji cvat; međutim, ako se ovaj inhibitor koristi u koncentraciji K_2 cvat se signifikantno smanji.

PRIMJER 10.1.6.

4 x 3; latinski pravokutnik

Važan preduvjet punog sklopa, a ovaj opet dobrog prinosa šećerne repe je osiguranje dobre klijavosti sjemena. Sjeme se, međutim, mora tretirati pesticidima, a oni mogu utjecati na klijavost.Na klijavost može djelovati i namakanje sjemena u vodi prije sjetve.

Da se prouče ovi mogući učinci, izveden je eksperiment: klupka poliploidne šećerne repe zaprašena su različitim dozama jednog pesticida (D_0 = nezaprašeno, D_1 = 500, D_2 = 600 i D_3 = 800 g na 100 kg sjemena) uz različito vrijeme namakanja (V_0 = nemočeno, V_1 = močeno pola sata i V_2 = močeno 1 sat).

Iz ovih, u pokus uključenih faktora, nastaje 12 kombinacija:

D_0V_0 – nezaprašeno, nemočeno
D_0V_1 – nezaprašeno, močeno pola sata
D_0V_2 – nezaprašeno, močeno 1 sat
D_1V_0 – zaprašeno s 500 g/100 kg, nemočeno
D_1V_1 – zaprašeno s 500 g/100 kg, močeno pola sata
D_1V_2 – zaprašeno s 500 g/100 kg, močeno 1 sat
D_2V_0 – zaprašeno s 600 g/100 kg, nemočeno
D_2V_1 – zaprašeno s 600 g/100 kg, močeno pola sata
D_2V_2 – zaprašeno s 600 g/100 kg, močeno 1 sat
D_3V_0 – zaprašeno s 800 g/100 kg, nemočeno
D_3V_1 – zaprašeno s 800 g/100 kg, močeno pola sata
D_3V_2 – zaprašeno s 800 g/100 kg, močeno 1 sat

Naveli smo sve ove kombinacije kako bismo još jedanput naglasili da u faktorijalnom pokusu moraju biti zastupljene baš sve kombinacije, dakle i ove tzv. "nulte" (s D_0 - nezaprašeno ili V_0 - nemočeno). Samo tako je višefaktorijalan pokus kompletan, sa svim mogućim kombinacijama koje proizlaze iz faktora i svih njihovih stepenica iz kojih se onda mogu otkriti i testirati djelovanja svih faktora i interakcija.

Jednako tako se valja opet podsjetiti, da su sve sheme koje smo navodili kod jednofaktorijalnih pokusa, primjenjive i kod višefaktorijalnih.

Zato je za ovaj pokus korištena shema latinskog pravokutnika s 4 repeticije (broj tretiranja - kombinacija je djeljiv s brojem ponavljanja).

Plan pokusa i podaci o postotku klijavosti:

Vodoravne repeticije	Okomite repeticije												$\sum x$ Vod. rep.
	I.			II.			III.			IV.			
I.	D_0V_0	D_2V_0	D_1V_0	D_1V_1	D_3V_0	D_0V_1	D_2V_1	D_3V_1	D_3V_2	D_1V_2	D_0V_2	D_2V_2	852.6
	54.7	69.0	66.7	77.3	67.0	56.3	76.3	80.3	79.3	75.0	72.0	78.7	
II.	D_1V_1	D_3V_0	D_0V_1	D_2V_2	D_3V_1	D_1V_0	D_1V_2	D_2V_0	D_0V_2	D_0V_0	D_3V_2	D_2V_1	847.1
	75.7	66.0	59.7	78.8	75.0	63.3	74.3	69.3	69.7	57.3	77.0	81.0	
III.	D_0V_2	D_2V_1	D_3V_1	D_1V_2	D_3V_2	D_0V_0	D_1V_1	D_1V_0	D_2V_2	D_3V_0	D_2V_0	D_0V_1	875.0
	72.0	83.3	77.3	77.0	74.7	59.0	83.0	68.3	79.0	66.7	69.0	65.7	
IV.	D_3V_2	D_2V_2	D_1V_2	D_0V_2	D_2V_1	D_2V_0	D_0V_1	D_3V_0	D_0V_0	D_1V_0	D_3V_1	D_1V_1	859.1
	73.7	77.7	78.7	72.0	76.3	66.7	63.0	65.0	61.0	69.0	78.3	77.7	
$\sum x$ Okom. rep.		854.5			843.4			868.5			867.4		$\sum x$=3433.8

Podaci po kombinacijama:

	D_0V_0	D_0V_1	D_0V_2	D_1V_0	D_1V_1	D_1V_2	D_2V_0	D_2V_1	D_2V_2	D_3V_0	D_3V_1	D_3V_2
I.	54.7	56.3	72.0	66.7	77.3	75.0	69.0	76.3	78.7	67.0	80.3	79.3
II.	57.3	59.7	69.7	63.3	75.7	74.3	69.3	81.0	78.8	66.0	75.0	77.0
III.	59.0	65.7	72.0	68.3	83.0	77.0	69.0	83.3	79.0	66.7	77.3	74.7
IV.	61.0	63.0	72.0	69.0	77.7	78.7	66.7	76.3	77.7	65.0	78.3	73.7
$\sum x$ Kombinacija	232.0	244.7	285.7	267.3	313.7	305.0	274.0	316.9	314.2	264.7	310.9	304.7
$\bar{x}$ Kombinacija	58.0	61.2	71.4	66.8	78.4	76.2	68.5	79.2	78.5	66.2	77.7	76.2

I. dio ANOVA

$$SS_{\text{Ukupno}} = 54.7^2 + \ldots + 77.7^2 - \frac{3433.8^2}{48} = 2569.00$$

$$SS_{\text{Vodoravnih repeticija}} = \frac{852.6^2 + \ldots + 859.1^2}{12} - \frac{3433.8^2}{48} = 36.45$$

$$SS_{\text{Okomitih repeticija}} = \frac{854.5^2 + \ldots + 867.4^2}{12} - \frac{3433.8^2}{48} = 35.27$$

$$SS_{\text{Kombinacija}} = \frac{232.0^2 + \ldots + 304.7^2}{4} - \frac{3433.8^2}{48} = 2353.60$$

$$SS_{\text{Pogreške}} = 2569.00 - 36.45 - 35.27 - 2353.60 = 143.68$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	47	2569.00				
Vodoravne rep.	3	36.45				
Okomite rep.	3	35.27				
Kombinacije	11	2353.60	213.96	44.67 **	2.16	2.98
Pogreška	30	143.68	4.79			
Doze	3	1053.00	351.10	73.30 **	2.92	4.51
Vrijeme	2	1082.00	541.20	112.98 **	3.32	5.39
Interakcija D x V	6	218.60	36.43	7.60 **	2.42	3.47

Nakon izračunatih varijanci i provedenog F-testa kombinacija (tablica ANOVA) imamo dokaza (visokosignifikantan F-test) da među kombinacijama ima različitosti u postotku klijavosti.

Koji od faktora ili njihova interakcija su razlogom za to, saznat ćemo raščlanjivanjem kombinacija u drugom dijelu analize.

II. dio ANOVA

Doza - Vrijeme namakanja

	V_0	V_1	V_2	$\sum x_{Doze}$	$\bar{x}_{Doze}$
D_0	232.0	244.7	285.7	762.4	63.53
D_1	267.3	313.7	305.0	886.0	73.83
D_2	274.0	316.9	314.2	905.1	75.42
D_3	264.7	310.9	304.7	880.3	73.36
$\sum x_{Vremena}$	1038.0	1186.2	1209.6	$\sum x = 3433.8$	
$\bar{x}_{Vremena}$	64.87	74.14	75.56		

$$SS_{\text{Doza}} = \frac{762.4^2 + ... + 880.3^2}{4 \cdot 3} - \frac{3433.8^2}{48} = 1053.00$$

$$SS_{\text{Vremena namakanja}} = \frac{1038.0^2 + 1186.2^2 + 1209.6^2}{4 \cdot 4} - \frac{3433.8^2}{48} =$$

$$= 1082.00$$

$$SS_{\text{DxV}} = 2353.60 - 1053.00 - 1082.00 = 218.60$$

Iz ovih vrijednosti za *SS*-ove i pripadajućih slobodnih varijanata izračunane su varijance i provedeni F testovi za faktore i interakciju (tablica ANOVA-e). Svi su visokosignifikantni, što znači:

- Statistički je opravdan i utjecaj različitih doza pesticida (D) i vremena močenja sjemena (V) na klijavost sjemena šećerne repe.
- Međutim, faktori ne djeluju nezavisno, već zajednički - među njima postoji interakcija.

t-test za faktor Doza

$$s_D = \sqrt{\frac{2 \cdot 4.79}{4 \cdot 3}} = 0.893$$

tablični t faktor treba očitati iz n - 1 pogreške, dakle iz 30 - 1 = 29

$$t_{\text{tabl}_{P=5\%}} = 2.04$$

$$t_{\text{tabl}_{P=1\%}} = 2.75$$

$$LSD_{P=5\%} = 2.04 \cdot 0.893 = 1.82\ \%$$

$$LSD_{P=1\%} = 2.75 \cdot 0.893 = 2.45\ \%$$

Iz prosječnih vrijednosti (u dvosmjernoj tablici) je vidljivo da je najniži postotak klijavosti netretiranog sjemena (D_0), a primjenom pesticida u dozi D_2 klijavost je najveća. Usporedimo:

	Razlike, (%)
63.53 (D_0) - 73.83 (D_1) =	- 10.30 **
63.53 (D_0) - 75.42 (D_2) =	- 11.89 **
63.53 (D_0) - 73.36 (D_3) =	- 9.83 **

Visokosignifikantne razlike znače da je opravdana primjena pesticida, jer se u odnosu na netretirano sjeme, postotak klijavosti signifikantno poveća. No, da li to vrijedi jednako za sve tri upotrijebljene doze?

75.42 (D_2) - 73.83 (D_1) = 1.59 ns

75.42 (D_2) - 73.36 (D_3) = 2.06 *

Konačno bismo nakon ovih usporedbi mogli zaključiti: za bolju klijavost opravdano je tretirati sjeme pesticidom, svejedno je da li primijeniti dozu D_1 ili D_2 (razlika u postotku klijavosti nije značajna) ali ne dozu D_3 jer se klijavost značajno smanjuje.

t-test za faktor Vrijeme namakanja

$$s_D = \sqrt{\frac{2 \cdot 4.79}{4 \cdot 4}} = 0.774$$

$$LSD_{P=5\%} = 2.04 \cdot 0.774 = 1.58\ \%$$

$$LSD_{P=1\%} = 2.75 \cdot 0.774 = 2.13\ \%$$

Prosječni postotak klijavosti najmanji je izgleda kod nemočenog sjemena (V_0), no da utvrdimo da li je to i opravdano, usporedimo vrijednosti:

	Razlike, (%)
64.87 (V_0) - 74.14 (V_1) =	- 9.27 **
64.87 (V_0) - 75.56 (V_2) =	- 10.69 **

Statistički opravdane razlike u postotku klijavosti močenog i nenamakanog sjemena dozvoljavaju zaključiti da je sjeme opravdano namakati, jer je klijavost takvog sjemena veća.

Kako razlika u postotku klijavosti sjemena močenog pola sata i jedan sat iznosi 1.42 % (75.56 - 74.14 = 1.42) i nije statistički opravdana, to znači da je dostatno sjeme namakati pola sata da bi se klijavost značajno povećala.

Međutim, najvažnije što smo utvrdili jest da je djelovanje ova dva proučena faktora zavisno, da postoji opravdana interakcija. Zbog toga ima najviše smisla testirati razlike među kombinacijama jer se kroz njih manifestira to zajedničko i zavisno djelovanje faktora.

t-test za interakciju

$$s_D = \sqrt{\frac{2 \cdot 4.79}{4}} = 1.547$$

$$LSD_{P=5\%} = 2.04 \cdot 1.547 = 3.15\ \%$$

$$LSD_{P=1\%} = 2.75 \cdot 1.547 = 4.25\ \%$$

Prosječne vrijednosti DxV kombinacija

	V_0	V_1	V_2
D_0	58.0	61.2	71.4
D_1	66.8	78.4	76.2
D_2	68.5	79.2	78.5
D_3	66.2	77.7	76.2

Svaka vrijednost koja je manja od 76.05 odnosno 74.95 (a to je najveća vrijednost 79.2 minus $LSD_{P=5\%}$ ili $LSD_{P=1\%}$) znači signifikantno ili visokosignifikantno manji postotak klijavosti.

Kako se prosječne vrijednosti koje su bojom otisnute na tablici ne razlikuju značajno, to znači da se svakom od tih kombinacija postiže podjednako povećanje postotka klijavosti.

Kad je tako, praktično je, dakako, odabrati najjeftiniju i najbržu, a to znači preporuku: zaprašiti sjeme s 500 grama na 100 kg sjemena i sjeme namakati pola sata (dakle kombinacija D_1V_1) i time osigurati opravdano veću klijavost, pa u konačnici i prinos.

10.2. Trofaktorijalni pokusi

Kako sam naziv ukazuje, to su pokusi koji omogućuju istovremeno proučavanje i testiranje djelovanja tri faktora i njihovih interakcija. Naime, uz glavna djelovanja faktora A, B i C, u trofaktorijalnom pokusu su još i četiri interakcije i to: tri jednostruke ili interakcije prvog reda (AxB, AxC i BxC) i jedna dvostruka interakcija ili interakcija drugog reda, u koju su uključena sva tri faktora (AxBxC).

Princip izbora sheme pokusa u trofaktorijalnim pokusima je isti kao u svim pokusima uopće, a odlučujuću ulogu pri tome ima broj kombinacija koji proizlazi iz broja faktora i njihovih stepenica.

Općenita shema analize trofaktorijalnog pokusa:

Izvori varijabilnosti		n - 1
I.	Ukupno	$n_A \cdot n_B \cdot n_C \cdot n_r - 1$
	Repeticije	$n_r - 1$
	Kombinacije	$n_A \cdot n_B \cdot n_C - 1$
	Pogreška	$(n_r - 1)\ (n_A \cdot n_B \cdot n_C - 1)$
II.	Faktori:	
	A	$n_A - 1$
	B	$n_B - 1$
	C	$n_C - 1$
	Interakcije:	
	AxB	$(n_A - 1)\ (n_B - 1)$
	AxC	$(n_A - 1)\ (n_C - 1)$
	BxC	$(n_B - 1)\ (n_C - 1)$
	AxBxC	$(n_A - 1)\ (n_B - 1)\ (n_C - 1)$

Za provođenje t-testova za faktore i interakcije standardna pogreška razlike (s_D) se računa po formulama:

$$\text{za faktor A} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_B \cdot n_C}}$$

$$\text{za faktor B} \rightarrow {}_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_A \cdot n_C}}$$

$$\text{za faktor C} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_A \cdot n_B}}$$

za interakcije:

$$\text{A x B} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_C}}$$

$$\text{A x C} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_B}}$$

$$\text{B x C} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r \cdot n_A}}$$

$$\text{A x B x C} \rightarrow s_D = \sqrt{\frac{2 \cdot s^2 pogreške}{n_r}}$$

PRIMJER 10.2.1.

3 x 3 x 2 (3^2 x 2); slučajni blokni raspored

Kako bi se proučilo da li i kako različiti faktori utječu na prinos plodova po stablu, proveden je pokus s tri sorte jabuka (S_1 = Idared, S_2 = Jonagold, S_3 = Elstar), svaka cijepljena na tri podloge i međupodloge (P_1 = M27/25 na MM106, P_2 = M27/50 na MM106, P_3 = M_9), a sađena na dvije različite dubine (D_1 = duboka sadnja, D_2 = plitka sadnja).

Kombinacije ova tri faktora (Sorta, Podloga, Dubina) su:

1) $S_1P_1D_1$	7) $S_2P_1D_1$	13) $S_3P_1D_1$
2) $S_1P_1D_2$	8) $S_2P_1D_2$	14) $S_3P_1D_2$
3) $S_1P_2D_1$	9) $S_2P_2D_1$	15) $S_3P_2D_1$
4) $S_1P_2D_2$	10) $S_2P_2D_2$	16) $S_3P_2D_2$
5) $S_1P_3D_1$	11) $S_2P_3D_1$	17) $S_3P_3D_1$
6) $S_1P_3D_2$	12) $S_2P_3D_2$	18) $S_3P_3D_2$

Pokus je izveden po shemi slučajnog bloknog rasporeda u 4 ponavljanja.

Svaka kombinacija je u svakoj repeticiji bila predstavljena sa 10 stabala. Podaci o prinosu plodova jabuke u kg po stablu, sređeni po kombinacijama su:

Repeticije	Kombinacije																		$\sum x_{\text{Repeticije}}$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
I.	4.0	4.3	8.8	5.8	3.0	6.3	10.7	6.7	7.5	3.0	12.1	9.3	8.7	6.0	9.8	6.4	10.4	9.3	132.1
II.	3.9	3.2	2.0	6.0	4.5	3.2	12.5	4.4	6.4	10.0	13.4	6.7	20.1	8.6	11.1	10.8	18.0	12.8	157.6
III.	3.5	2.0	4.8	2.5	4.5	4.5	12.2	5.1	9.6	8.0	9.7	4.0	5.0	3.9	5.5	6.1	8.8	6.7	106.4
IV.	8.3	3.7	4.6	2.7	3.7	7.9	5.5	5.7	7.0	8.1	10.0	11.9	8.0	10.4	6.9	7.3	15.4	13.4	140.5
$\sum x_{\text{Kombinacije}}$	19.7	13.2	20.2	17.0	15.7	21.9	40.9	21.9	30.5	29.1	45.2	31.9	41.8	28.9	33.3	30.6	52.6	42.2	536.6

I. dio ANOVA

$$SS_{\text{Ukupno}} = 4.0^2 + \ldots + 13.4^2 - \frac{536.6^2}{72} = 991.5994$$

$$SS_{\text{Repeticija}} = \frac{132.1^2 + \ldots + 140.5^2}{18} - \frac{536.6^2}{72} = 75.8050$$

$$SS_{\text{Rombinacija}} = \frac{19.7^2 + \ldots + 42.2^2}{4} - \frac{536.6^2}{72} = 538.5643$$

$$SS_{\text{Pogreške}} = 991.5994 - 75.8050 - 538.5643 = 377.2300$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	71	991.5994				
Repeticije	3	75.8050				
Kombinacije	17	538.5643	31.6802	4.28 **	1.90	2.39
Pogreške	51	377.2300	7.3967			
Faktori						
S	2	335.1686	167.5843	22.66 **	3.18	5.06
P	2	59.3269	29.6635	4.01*	3.18	5.06
D	1	55.4755	55.4755	7.50 **	4.03	7.17
Interakcije						
S x P	4	29.9889	7.4972	1.01 ns	2.56	3.72
S x D	2	20.5219	10.2610	1.39 ns	3.18	5.06
P x D	2	20.9453	10.4726	1.42 ns	3.18	5.06
S x P x D	4	17.1372	4.2843	0.58 ns	2.56	3.72

Iz dobivenih SS vrijednosti i pripadajućih slobodnih varijanata u tablici ANOVA su izračunane varijance i proveden F-test kombinacija. Signifikantan *F-test* kombinacija upućuje na zaključak da različitosti među kombinacijama ima, no koji od faktora ili koja interakcija je tome razlog otkriti ćemo u II dijelu analize, u kojem ćemo raščlaniti $SS_{\text{Kombinacija}}$ na faktore i interakcije. Da bi to mogli, treba konstruirati tri dvosmjerne tablice (naime, toliko njih koliko je jednostrukih interakcija).

Sorta - Podloga

	P_1	P_2	P_3	$\sum x_{\text{Sorte}}$	$\bar{x}_{\text{Sorte}}$
S_1	32.9	37.2	37.6	107.7	4.49
S_2	62.8	59.6	77.1	199.5	8.31
S_3	70.7	63.9	94.8	229.4	9.56
$\sum x_{\text{Podloge}}$	166.4	160.7	209.5	$\sum x = 536.6$	
$\bar{x}_{\text{Podloge}}$	6.93	6.69	8.73		

Sorta - Dubina

	D_1	D_2	$\sum x_{\text{Sorte}}$
S_1	55.6	52.1	107.7
S_2	116.6	82.9	199.5
S_3	127.7	101.7	229.4
$\sum x_{\text{Dubine}}$	299.9	236.7	$\sum x = 536.6$
$\bar{x}_{\text{Dubine}}$	8.33	6.57	

Podloge - Dubine

	D_1	D_2	$\sum x_{\text{Podloge}}$
P_1	102.4	64.0	166.4
P_2	84.0	76.7	160.7
P_3	113.5	96.0	209.5
$\sum x_{\text{Dubine}}$	299.9	236.7	$\sum x = 536.6$

Iz prve dvosmjerne tablice moguće je izračunati SS_{Sorte}, SS_{Podloge} i SS_{SxP}

$$SS_{\text{Sorte}} = \frac{107.7^2 + 199.5^2 + 229.4^2}{4 \cdot 3 \cdot 2} - \frac{536.6^2}{72} = 335.1686$$

$$SS_{\text{Podloge}} = \frac{166.4^2 + 160.7^2 + 209.5^2}{4 \cdot 3 \cdot 2} - \frac{536.6^2}{72} = 59.3269$$

$$SS_{\text{SxP}} = SS_{\text{SxP kombinacija}} - SS_{\text{Sorte}} - SS_{\text{Podloge}} =$$

$$= \frac{32.9^2 + \ldots + 94.8^2}{4 \cdot 2} - \frac{536.6^2}{72} - 335.1686 - 59.3269 =$$

$$= 29.9889$$

Druga dvosmjerna tablica omogućuje izračunati SS_{Sorte} (što smo već gore izračunali), SS_{Dubine} i $SS_{\text{Interakcije:}}$

$$SS_{\text{Dubine}} = \frac{299{,}9^2 + 236{,}7^2}{4 \cdot 3 \cdot 3} - \frac{536.6^2}{72} = 55.4755$$

$$SS_{\text{SxD}} = SS_{\text{SxD kombinacija}} - SS_{\text{Sorte}} - SS_{\text{Dubine}} =$$

$$= \frac{55.6^2 + \ldots + 101.7^2}{4 \cdot 3} - \frac{536.6^2}{72} - 335.1686 - 55.4755 =$$

$$= 20.5219$$

Konačno, iz treće dvosmjerne tablice moguće je računanje $SS_{Podloge}$ (već izračunano), SS_{Dubine} (također već imamo) i SS_{PxD} što ćemo izračunati:

$$SS_{PxD} = SS_{PxD\ kombinacija} - SS_P - SS_D =$$

$$= \frac{102.4^2 + ... + 96.0^2}{4 \cdot 3} - \frac{536.6^2}{72} - 59.3269 - 55.4755 = 20.9453$$

Zadnja, dvostruka interakcija se dobije tako da se od $SS_{Kombinacija}$ odbije SS svih faktora i svih interakcija:

$$SS_{SxPxD} = SS_{Kombinacija} - SS_{Sorte} - SS_{Podloge} - SS_{Dubine} - SS_{SxP} -$$

$$- SS_{SxD} - SS_{PxD} =$$

$$= 538.5643 - 335.1686 - 59.3269 - 55.4755 - 29.9889 - 20.5219 -$$

$$- 20.9453 = 17.1372$$

Sve ove izračunate SS vrijednosti uvrstimo u drugi dio tablice ANOVA, za svaki faktor i interakciju odredimo pripadajući n-1, izračunamo varijance i za svaki faktor i svaku interakciju provedemo F-test.

Ne zaboravimo: tablični F faktori se očitavaju iz n-1 faktora ili interakcija i n-1 pogreške (znači 1 i 51, 2 i 51 i 4 i 51; u tablici nema vrijednosti 51 pa ćemo uzeti 50).

Na temelju rezultata F-testa za faktore i interakcije, zaključujemo:

- nesignifikantne interakcije znače da faktori djeluju nezavisno jedan o drugom
- opravdani F testovi za sva tri proučavana faktora, dokaz su da prinos po stablu značajno ovisi i o sorti i o podlozi i o dubini sadnje. Faktori, Sorta i Podloga imaju po tri stepenice (3 sorte, 3 podloge), pa je za testiranje razlika između prosječnih vrijednosti stepenica svakog faktora, potrebno provesti t-test.

t-test za faktor Sorta

$$LSD = t \cdot s_D$$

$t_{P=5\%}$ = 2.01 (očitan iz 50, jer je broj slobodnih

$t_{P=1\%}$ = 2.68 varijanata pogreške 51)

$$s_D = \sqrt{\frac{2 \cdot 7.3967}{4 \cdot 3 \cdot 2}} = 0.7851 \text{ kg}$$

$$LSD_{P=5\%} = 2.01 \cdot 0.7851 = 1.57 \text{ kg}$$

$$LSD_{P=1\%} = 2.68 \cdot 0.7851 = 2.10 \text{ kg}$$

Usporede li se prosječni prinosi sorata (izračunani su i navedeni u prvoj dvosmjernoj tablici) vidljivo je da sorta Idared (S_1) ima značajno najmanji prinos, a sorte Jonagold (S_2) i Elstar (S_3) imaju podjednak prinos po stablu. Razlika od 1.25 kg je zanemariva (nije signifikantna).

4.49 (S_1) 8.31 (S_2) 9.56 (S_3)

D=3.82** D=1.25 ns

D=5.07**

t-test za faktor Podloga

Najmanja signifikantna razlika (LSD) će biti ista kao i kod faktora Sorta jer se s_D izračuna na isti način (obzirom na isti broj stepenica ova dva faktora)

$$s_D = \sqrt{\frac{2 \cdot 7.3967}{4 \cdot 3 \cdot 2}} = 0.7851 \text{ kg}$$

$$LSD_{P=5\%} = 1.57 \text{ kg}$$

$$LSD_{P=1\%} = 2.10 \text{ kg}$$

Prosječni prinosi po stablu uvjetovani podlogom i međupodlogom navedeni su također u prvoj dvosmjernoj tablici.

Usporedimo ih:

6.93 (P_1) 6.69 (P_2) 8.73 (P_3)

D=0.24 ns D=2.04*

D=1.8*

To znači, siginifikantno najveći prinos jabuka dobiven je na podlozi P_3 (a to je M_9).

Treći faktor - **dubina sadnje** imao je dvije gradacije, pa je već visokosignifikantan F-test dokaz da je razlika između prosječnog prinosa dublje sađenih voćaka jabuke i onih sađenih pliće, opravdana (prosječne vrijednosti su u drugoj po redu dvosmjernoj tablici)

$$8.33\ (D_1) - 6.57\ (D_2) = 1.76^{**}$$

Dakle, dubljom sadnjom postiže se veći prinos po stablu.

PRIMJER 10.2.2.

2 x 4 x 2; latinski pravokutnik

Kroz eksperiment proveden sa dvije sorte soje, nastojao se dobiti odgovor na pitanje o možebitnom pozitivnom ili negativnom utjecaju tretiranja sjemena soje fungicidima, kao i mogućnosti poticanja razvoja većeg broja korjenovih kvržica (važnih u fiksaciji dušika) a time i povećanja prinosa inokulacijom sjemena bakterijom Rhizobium japonicum (ili novije Bradyrhisobium japonicum).

U tu svrhu su u pokus uključeni ovi faktori i njihove stepenice: faktor Sorta (Maksimirka i Srećka), faktor Fungicid (F_0 - sjeme koje nije tretirano fungicidom, uz ono tretiriano s tri različita fungicida F_1, F_2 i F_3), te faktor Rhizobium (Rh_0- neinokulirano sjeme, Rh_1- sjeme inokulirano ovom bakterijom).

Kombinacije tih faktora su

1) MF_0Rh_0 9) SF_0Rh_0

2) MF_0Rh_1 10) SF_0Rh_1

3) MF_1Rh_0 11) SF_1Rh_0

4) MF_1Rh_1 12) SF_1Rh_1

5) MF_2Rh_0 13) SF_2Rh_0

6) MF_2Rh_1 14) SF_2Rh_1

7) MF_3Rh_0 15) SF_3Rh_0

8) MF_3Rh_1 16) SF_3Rh_1

U pokusu su one bile raspoređene po planu latinskog pravokutnika u četiri ponavljanja.

Među mnogim drugim pokazateljima mogućih učinaka u pokus uključenih faktora, praćen je i broj korjenovih kvržica na kojem ćemo se zadržati u ovoj analizi.

Raspored kombinacija u pokusu i podaci o broju korjenovih kvržica za svaku kombinaciju u svakoj repeticiji:

Vodoravne repeticije	Okomite repeticije																$\sum x$ Vod. rep.
	I.				II.				III.				IV.				
I.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	371
	16	18	20	17	18	17	20	8	48	36	32	39	13	19	33	27	
II.	8	5	6	7	16	13	15	14	1	4	3	2	10	12	11	9	325
	17	19	14	16	23	29	18	28	13	26	16	13	22	22	25	24	
III.	11	9	12	10	4	1	2	3	14	16	13	15	8	5	6	7	336
	29	32	31	27	14	14	19	17	20	27	26	16	10	18	11	25	
IV.	13	16	15	14	12	10	11	9	8	5	7	6	4	2	3	1	316
	30	25	19	29	21	23	35	26	12	8	13	15	21	16	14	9	
$\sum x$ Okom. rep.	359				330				350				309				$\sum x = 1348$

Podaci sređeni po kombinacijama:

MF_0Rh_0	MF_0Rh_1	MF_1Rh_0	MF_1Rh_1	MF_2Rh_0	MF_2Rh_1	MF_3Rh_0	MF_3Rh_1	SF_0Rh_0	SF_0Rh_1	SF_1Rh_0	SF_1Rh_1	SF_2Rh_0	SF_2Rh_1	SF_3Rh_0	SF_3Rh_1
16	18	20	17	18	17	20	8	48	36	32	29	13	19	33	27
13	13	16	26	19	14	16	17	24	22	25	22	29	28	18	23
14	19	17	14	18	11	25	10	32	27	29	31	26	20	16	27
9	16	14	21	8	15	13	12	26	23	35	21	30	29	19	25
52	66	67	78	63	57	74	47	130	108	121	103	98	96	86	102

I. dio ANOVA

$$SS_{\text{Ukupno}} = 16^2 + \ldots + 9^2 - \frac{1348^2}{64} = 3787.750$$

$$SS_{\text{Vodoravnih repeticija}} = \frac{371^2 + \ldots + 316^2}{16} - \frac{1348^2}{64} = 108.875$$

$$SS_{\text{Okomitih repeticija}} = \frac{359^2 + \ldots + 309^2}{16} - \frac{1348^2}{64} = 92.875$$

$$SS_{\text{Kombinacija}} = \frac{52^2 + \ldots + 102^2}{4} - \frac{1348^2}{64} = 2350.25$$

$$SS_{\text{Pogreške}} = 3787.750 - 108.875 - 92.875 - 2350.25 = 1235.75$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno	63	3787.750				
Vodoravne repeticije	3	108.875				
Okomite repeticije	3	92.875				
Kombinacije	15	2350.250	156.6833	5.32 **	1.90	2.49
Pogreška	42	1235.750	29.4226			
Faktori:						
S	1	1806.250	1806.2500	61.39 **	4.08	7.31
F	3	168.625	56.2083	1.91 ns	2.84	4.31
Rh	1	18.0625	18.0625	0.61 ns	4.08	7.31
Interakcije:						
S x F	3	106.625	35.5417	1.21 ns	2.84	4.31
S x Rh	1	5.0625	5.0625	0.17 ns	4.08	7.31
F x Rh	3	0.5625	0.1875	0.01 ns	2.84	4.31
S x F x Rh	3	245.0625	81.6875	2.78 ns	2.84	4.31

F-test kombinacija na kraju I dijela analize znači da među njima ima značajnih razlika u broju korjenovih kvržica.

U drugom dijelu analize će se, raščlanjivanjem $SS_{\text{Kombinacija}}$, otkriti opravdanost djelovanja svakog pojedinog faktora, kao i njihovog kombiniranog djelovanja kroz interakcije.

U tu svrhu se podatke treba svrstati u dvosmjerne tablice (opet: onoliko njih koliko jednostrukih interakcija), iz kojih je moguće izračunati SS svakog faktora i svake interakcije

Sorta - Fungicid

	F_0	F_1	F_2	F_3	$\sum x_{\text{Sorte}}$	$\bar{x}_{\text{Sorte}}$
M	118	145	120	121	504	15.75
S	238	224	194	188	844	26.37
$\sum x_{\text{Fungicida}}$	356	369	314	309	$\sum x = 1348$	
$\bar{x}_{\text{Fungicida}}$	22.25	23.06	19.62	19.31		

Sorta - Rhizobium

	Rh_0	Rh_1	$\sum x$ Sorte
M	256	248	504
S	435	409	844
$\sum x$ Rhizobiuma	691	657	$\sum x = 1348$
$\bar{x}$ Rhizobiuma	21.59	20.53	

Rhizobium - Fungicid

	F_0	F_1	F_2	F_3	$\sum x$ Rhizobiuma
Rh_0	182	188	161	160	691
Rh_1	174	181	153	149	657
$\sum x$ Fungicida	356	369	314	309	$\sum x = 1348$

$$SS_{\text{Sorte}} = \frac{504^2 + 844^2}{4 \cdot 2 \cdot 4} - \frac{1348^2}{64} = 1806.250$$

$$SS_{\text{Fungicida}} = \frac{356^2 + \ldots + 309^2}{4 \cdot 2 \cdot 2} - \frac{1348^2}{64} = 168.625$$

$$SS_{SxF} = SS_{\text{SxF kombinacija}} - SS_{\text{Sorte}} - SS_{\text{Fungicida}} =$$

$$= \frac{118^2 + \ldots + 188^2}{4 \cdot 2} - \frac{1348^2}{64} - 1806.25 - 168.625 =$$

$$= 106.625$$

$$SS_{\text{Rhizobium}} = \frac{691^2 + 657^2}{4 \cdot 2 \cdot 4} - \frac{1348^2}{64} = 18.0625$$

$$SS_{\text{SxRh}} = SS_{\text{SxRh Komibinacija}} - SS_{\text{Sorte}} - SS_{\text{Rhizobiuma}} =$$

$$= \frac{256^2 + \ldots + 409^2}{4 \cdot 4} - \frac{1348^2}{64} - 1806.25 - 18.0625 =$$

$$= 5.0625$$

$$SS_{\text{FxRh}} = SS_{\text{FxRh Komibinacija}} - SS_{\text{Fungicida}} - SS_{\text{Rhizobiuma}} =$$

$$= \frac{182^2 + \ldots + 149^2}{4 \cdot 2} - \frac{1348^2}{64} - 168.625 - 18.0625 =$$

$$= 0.5625$$

$$SS_{SxFxRh} = SS_{Kombinacija} - SS_S - SS_F - SS_{Rh} - SS_{SxF} - SS_{SxRh} - SS_{FxRh}$$
$$= 2350.25 - 1806.25 - 168.625 - 18.0625 - 106.625 -$$
$$- 5.0625 - 0.5625 = 245.0625$$

Na temelju ovih vrijednosti u tablici ANOVA su izračunane varijance i provedeni F-testovi, pa možemo zaključiti:

- Na broj korjenovih kvržica soje značajan utjecaj je imao samo faktor Sorta. Niti fungicidi, niti inokulacija bakterijom nisu uvjetovali opravdane razlike u broju korjenovih kvržica. One razlike koje su utvrđene (a prosječne vrijednosti za svaku stepenicu svakog faktora su u dvosmjernim tablicama) samo su slučajne, zanemarive, nesignifikantne. Jednako tako nema opravdanih interakcija, a to znači da niti zajednička djelovanja proučavanih faktora nisu značajna.
- Kako su u pokus bile uključene samo dvije sorte, to već na temelju visokosignifikantnog F-testa možemo zaključiti da je razlika u broju korjenovih kvržica između sorata Srećka i Maksimirka opravdana:

$$\bar{x}_{Srećka} - \bar{x}_{Maksimirka} = 26.37 - 15.75 = 10.62^{**}$$

Kod sorte Srećka razvio se dakle opravdano veći broj korjenovih kvržica.

■

PRIMJER 10.2.3.

2 x 4 x 4 ($2x4^2$); slučajni blokni raspored

Da bi se saznalo kako se mijenja sadržaj dušika u biljci duhana uz gnoijidbu s različitim količinama dušika, izveden je jedan trofaktorijalni pokus.

Sorte Čulinec i Burley gnojene su sa 0 ili 100 ili 200 ili 300 kg N/ha, pa su se 48, 67, 98 i 125 dana nakon sadnje uzimali uzorci za analizu sadržaja dušika u biljci.

To znači pokus je imao ova tri faktora: Sorta (S_1 = Čulinec, S_2 = Berley), Gnojidba sa N (N_1 = 0, N_2 = 100, N_3 = 200 i N_4 = 300 kg N/ha) i Rok uzimanja uzoraka (R_1 = 48, R_2 = 67, R_3 = 98 i R_4 = 125 dana nakon sadnje) a rezultirajuće 32 kombinacije su:

1) $S_1N_1R_1$	17) $S_2N_1R_1$
2) $S_1N_1R_2$	18) $S_2N_1R_2$
3) $S_1N_1R_3$	19) $S_2N_1R_3$
4) $S_1N_1R_4$	20) $S_2N_1R_4$
5) $S_1N_2R_1$	21) $S_2N_2R_1$
6) $S_1N_2R_2$	22) $S_2N_2R_2$
7) $S_1N_2R_3$	23) $S_2N_2R_3$
8) $S_1N_2R_4$	24) $S_2N_2R_4$
9) $S_1N_3R_1$	25) $S_2N_3R_1$
10) $S_1N_3R_2$	26) $S_2N_3R_2$
11) $S_1N_3R_3$	27) $S_2N_3R_3$
12) $S_1N_3R_4$	28) $S_2N_3R_4$
13) $S_1N_4R_1$	29) $S_2N_4R_1$
14) $S_1N_4R_2$	30) $S_2N_4R_2$
15) $S_1N_4R_3$	31) $S_2N_4R_3$
16) $S_1N_4R_4$	32) $S_2N_4R_4$

Izveden je bio prema slučajnom bloknom rasporedu sa tri ponavljanja.

Podaci o sadržaju N u biljci (u %) po kombinacijama i ponavljanjima su:

Sorte	$N_{gnojidbe}$	Rokovi	Kombinacije	Repeticije I	Repeticije II	Repeticije III	$\sum x_{Kombinacije}$
S_1	N_1	R_1	1	2.83	2.33	2.47	7.63
		R_2	2	1.92	1.86	2.09	5.87
		R_3	3	1.92	1.12	1.40	4.44
		R_4	4	1.36	1.53	1.07	3.96
	N_2	R_1	5	3.92	3.43	2.80	10.15
		R_2	6	2.49	2.43	1.75	6.67
		R_3	7	1.69	0.93	1.46	4.08
		R_4	8	1.63	1.75	1.63	5.01
	N_3	R_1	9	3.59	3.39	3.52	10.50
		R_2	10	2.15	2.32	2.02	6.49
		R_3	11	1.93	2.01	2.09	6.03
		R_4	12	1.98	1.88	1.49	5.35
	N_4	R_1	13	2.97	3.68	3.92	10.57
		R_2	14	2.75	2.59	2.84	8.18
		R_3	15	2.21	2.04	2.17	6.42
		R_4	16	2.02	2.22	1.74	5.98
S_2	N_1	R_1	17	2.26	2.59	2.07	6.92
		R_2	18	1.47	1.92	1.94	5.33
		R_3	19	1.67	1.53	1.98	5.18
		R_4	20	1.23	1.79	1.67	4.69
	N_2	R_1	21	3.37	2.80	2.70	8.87
		R_2	22	2.06	2.16	1.81	6.03
		R_3	23	2.11	2.20	1.52	5.83
		R_4	24	1.65	1.65	2.05	5.35
	N_3	R_1	25	3.13	3.20	2.85	9.18
		R_2	26	2.45	2.12	2.32	6.89
		R_3	27	2.15	2.13	1.86	6.14
		R_4	28	1.80	1.67	1.89	5.36
	N_4	R_1	29	3.26	3.62	3.24	10.12
		R_2	30	2.30	2.65	2.31	7.26
		R_3	31	1.94	2.25	2.09	6.28
		R_4	32	1.89	1.93	2.37	6.19
$\sum x_{Repeticije}$				72.10	71.72	69.13	$\sum x = 212.95$

I. dio ANOVA

$$SS_{\text{Ukupno}} = 2.83^2 + \ldots + 2.37^2 - \frac{212.95^2}{96} = 40.240$$

$$SS_{\text{Repeticija}} = \frac{72.10^2 + 71.72^2 + 69.13^2}{32} - \frac{212.95^2}{96} = 0.163$$

$$SS_{\text{Kombinacija}} = \frac{7.63^2 + \ldots + 6.19^2}{3} - \frac{212.95^2}{96} = 35.667$$

$$SS_{\text{Pogreška}} = 40.240 - 0.163 - 35.667 = 4.410$$

Tablica ANOVA

Izvori varijabilnosti	$n-1$	SS	s^2	F_{exp}	F_{tabl} P=5%	F_{tabl} P=1%
Ukupno	95	40.240				
Repeticije	2	0.163				
Kombinacije	31	35.667	1.1505	16.20 **	1.65	2.03
Pogreška	62	4.410	0.0710			
Faktori:						
Sorta	1	0.030	0.030	0.42 ns	4.00	7.08
N gnojidba	3	6.420	2.140	30.14 **	2.76	4.13
Rok uzimanja uzoraka	3	26.489	8.830	124.37 **	2.76	4.13
Interakcije:						
S x N	3	0.070	0.023	0.32 ns	2.76	4.13
S x R	3	1.001	0.334	4.70**	2.76	4.13
N x R	9	1.070	0.119	1.68 ns	2.04	2.72
S x N x R	9	0.587	0.065	0.91 ns	2.04	2.72

Rezultat F-testa u prvom dijelu analize dokazuje da se kombinacije značajno razlikuju po sadržaju dušika u biljci.

U drugom dijelu analize saznati ćemo o utjecaju pojedinih faktora i njihovim interakcijama.

No, najprije konstruirajmo dvosmjerne tablice.

Sorta - N gnojidba

	N_1	N_2	N_3	N_4	$\sum x$ Sorte	$\bar{x}$ Sorte
S_1	21.90	25.91	28.37	31.15	107.33	2.236
S_2	22.12	26.08	27.57	29.85	105.62	2.200
$\sum x$ Ngnojidbe	44.02	51.99	55.94	61.00	$\sum x = 212.95$	
$\bar{x}$ Ngnojidbe	1.834	2.166	2.331	2.542		

Sorta - Rok uzimanja uzoraka

	R_1	R_2	R_3	R_4	$\sum x$ Sorte
S_1	38.85	27.21	20.97	20.30	107.33
S_2	35.09	25.51	23.43	21.59	105.62
$\sum x$ Roka	73.94	52.72	44.40	41.89	$\sum x = 212.95$
$\bar{x}$ Roka	3.081	2.197	1.850	1.745	

N gnojidba - Rok uzimanja uzoraka

	R_1	R_2	R_3	R_4	$\sum x$ Ngnojidbe
N_1	14.55	11.20	9.62	8.65	44.02
N_2	19.02	12.70	9.91	10.36	51.99
N_3	19.68	13.38	12.17	10.71	55.94
N_4	20.69	15.44	12.70	12.17	61.00
$\sum x$ Roka	73.94	52.72	44.40	41.89	$\sum x = 212.95$

Iz dvosmjernih tablica računajmo redom *SS* - ove:

$$SS_{\text{Sorte}} = \frac{107.33^2 + 105.62^2}{3 \cdot 4 \cdot 4} - \frac{212.95^2}{96} = 0.030$$

$$SS_{\text{Ngnojidbe}} = \frac{44.02^2 + ... + 61.00^2}{3 \cdot 2 \cdot 4} - \frac{212.95^2}{96} = 6.420$$

$$SS_{\text{SxN}} = SS_{\text{SxN kombinacija}} - SS_{\text{S}} - SS_{\text{N}} =$$

$$= \frac{21.9^2 + ... + 29.85^2}{3 \cdot 4} - \frac{212.95^2}{96} - 0.030 - 6.420 = 0.070$$

$$SS_{\text{Roka}} = \frac{73.94^2 + ... + 41.89^2}{3 \cdot 2 \cdot 4} - \frac{212.95^2}{96} = 26.489$$

$$SS_{SxR} = SS_{SxR\ kombinacija} - SS_S - SS_R =$$

$$= \frac{38.85^2 + ... + 21.59^2}{3 \cdot 4} - \frac{212.95^2}{96} - 0.030 - 26.489 = 1.001$$

$$SS_{NxR} = SS_{NxR\ kombinacija} - SS_N - SS_R =$$

$$= \frac{14.55^2 + ... + 12.7^2}{3 \cdot 2} - \frac{212.95^2}{96} - 6.420 - 26.489 = 1.070$$

$$SS_{SxNxR} = SS_{Kombinacija} - SS_S - SS_N - SS_R - SS_{SxN} - SS_{SxR} - SS_{NxR}$$

$$= 35.667 - 0.030 - 6.420 - 26.489 - 0.070 - 1.001 - 1.070$$

$$= 0.587$$

Nakon što su iz ovih SS vrijednosti izračunane varijance i provedeni F-testovi za sve faktore i interakcije zaključujemo:

- Sorte Čulinec i Burley se ne razlikuju opravdano po sadržaju dušika u biljci (F-test nesignifikantan - razlika je samo slučajna).
- Značajno se međutim mijenja postotak dušika u biljci ovisno o gradaciji primijenjenog dušika kao i o roku u kojem su uzimani uzorci. Provođenjem t-testa za ove faktore, ustanoviti ćemo kako.
- Od zajedničkih djelovanja opravdana je samo interakcija Sorta x Rok. To znači da je sadržaj dušika u biljci značajno uvjetovan kombiniranim djelovanjem ova dva faktora. Opravdanost razlika između njihovih kombinacija ustanoviti će se t-testom.

  Neopravdane interakcije znače da faktori djeluju nezavisno.

t-test za faktor N gnojidba

$$LSD = t \cdot s_D$$

$t_{P=5\%} = 2.01$ (t_{tab} očitan iz n-1=60)

$t_{P=1\%} = 2.68$

$$s_D = \sqrt{\frac{2 \cdot 0.071}{3 \cdot 2 \cdot 4}} = 0.0769$$

$$LSD_{P=5\%} = 2.00 \cdot 0.0769 = 0.154\ \%$$

$$LSD_{P=1\%} = 2.66 \cdot 0.0769 = 0.204\ \%$$

Iz prosječnih vrijednosti po gradacijama gnojidbe dušikom (prva dvosmjerna tablica) razvidno je da sadržaj dušika u biljci raste kako rastu dodane količine N/ha.

Usporedimo ih i testirajmo razlike među njima:

1.834 (N_1)	0.332**	0.497**	0.708**			
2.166 (N_2)				0.165**	0.376**	
2.331 (N_3)						0.211**
2.542 (N_4)						

Opravdano najmanji sadržaj dušika u biljci je u biljaka kojima nije dodan dušik (N_1). Kako količina dodanog dušika raste. tako se njegov sadržaj u biljci signifikantno povećava.

t-test za faktor Rok uzimanja uzorka

Za usporedbu prosječnih vrijednosti po rokovima, vrijediti će iste granične razlike (jer je s_D isti - izračuna se iz $s_D = \sqrt{\frac{2 \cdot 0.071}{3 \cdot 2 \cdot 4}}$ jer faktori N i R imaju isti broj stepenica)

3.081 (R_1)	0.884**	1.231**	1.336**			
2.197 (R_2)				0.347**	0.425**	
1.850 (R_3)						0.105 ns
1.745 (R_4)						

Uspoređujući prosječne vrijednosti i testirajući razlike između njih, zaključujemo da je sadržaj dušika u biljci bio opravdano najveći 48 dana nakon sadnje (tj. u prvom roku uzimanja uzoraka), da signifikantno opada u svakom sljedećem roku, da bi 125 dana nakon sadnje (R_4) bio gotovo upola manji u odnosu na prvi rok.

t-test za kombinacije faktora Sorta-Rok

Interakcija ovih faktora je signifikantna, pa je potrebno usporediti prosječne vrijednosti njihovih kombinacija (dobiju se iz druge po redu dvosmjerne tablice, dijeleći svaki broj sa 12 tj. 3 repeticije x 4 stepenice faktora N):

	R_1		R_2		R_3		R_4	
		**		**		ns		
S_1 (Čulinec)	3.237		2.267		1.747		1.692	
	**		ns		ns			ns
S_2 (Barley)	2.924		2.126		1.952		1.799	
		**		ns		ns		

$$s_D = \sqrt{\frac{2 \cdot 0.071}{3 \cdot 4}} = 0.109$$

$$LSD_{P=5\%} = 2.00 \cdot 0.109 = 0.218\ \%$$

$$LSD_{P=1\%} = 2.66 \cdot 0.109 = 0.290\ \%$$

Opravdana interakcija znači da se sadržaj dušika u biljci mijenja zbog zavisnosti djelovanja faktora, odnosno zbog kombinacije utjecaja sorte i roka uzimanja uzoraka.

U tablici prosječnih vrijednosti SxR kombinacija naznačeni su rezultati t-testa (opravdanost ili neopravdanost razlika) iz čega proizlazi da je u prvom roku (R_1 = 48 dana nakon sadnje) sadržaj dušika u biljci značajno niži u sorte Burley, dok u slijedećim rokovima razlike među sortama nisu opravdane.

Obje sorte imaju najveći sadržaj dušika u biljci u prvom roku, međutim kod sorte Burley on značajno padne već do roka od 69 dana nakon sadnje pa dalje gotovo stagnira, a kod sorte Čulinec se signifikantno smanjuje od roka do roka sve do 98 dana nakon sadnje, a zatim stagnira.

8
9
10

11 Pokusi s razdijeljenim parcelama

12

11.1. Split-plot
11.2. Split-blok ili strip-plot
11.3. Split-split plot

Slučajni blokni raspored je najčešće primjenjivana shema kod višefaktorijalnih pokusa. Kod nje se u svim ponavljanjima sve kombinacije faktora potpuno slučajno i ravnopravno raspoređuju po osnovnim parcelama.

Međutim, u izvođenju pokusa često se nailazi na poteškoće u jednakom radnomiziranju svih kombinacija. Jer, ili eksperimentalni materijal ili neki zahvati i operacije (obrada, sjetva, sadnja, radni zahvat stroja i sl.) koji su u eksperimentu nužni, otežavaju održati sve kombinacije faktora ravnopravnima.

Agronoma, recimo, zanima proučiti učinak obrade tla i gnojidbe dušikom. Stepenice faktora obrada zahtijevaju velike parcele (zbog primjene stroja), a razine faktora gnojidbe se apliciraju na manjim parcelama (čak i ručno).

Ili, pokusi s rokovima sjetve (sadnje) kao jednim od faktora, lakše će se izvesti ako se parcele koje se istovremeno siju (sade) grupiraju zajedno.

Ovi primjeri predstavljaju probleme kod kojih su za učinkovitu primjenu dva faktora potrebne različito velike parcele. Zato je dobro imati takav plan pokusa u kojem se stepenice jednog faktora mogu primijeniti na veće, a stepenice drugog na manje parcele. Upravo tim zahtjevima udovoljavaju sheme poznate pod nazivom **split-plot** ili **planovi s razdijeljenim parcelama**, u nekoliko varijacija.

11.1. Split-plot

Sam naziv split-plot doslovno znači dijeliti (split) parcelu (plot).

To je specifična shema za dvofaktorijalne pokuse, kod koje se stepenice jednog faktora raspoređuju na veće parcele, a unutar njih se na manjim parcelama rasporede stepenice drugog faktora, pri čemu se unutar svake repeticije sve moraju randomizirati.

Veće parcele se obično zovu cijele ili glavne parcele (engl. main plot), a manje podparcele (engl. subplot), pa je otud u praktičnoj primjeni uobičajeno zvati prvi faktor glavnim faktorom, a drugi podfaktorom (to je vjerojatno proizišlo isključivo zbog prijevoda, jer procjena učinkovitosti niti jednog od ova dva faktora nije manje važna).

Evo kako bi izgledao postupak izrade plana pokusa prema split-splot rasporedu za recimo 4 x 3 pokus sa 4 ponavljanja:

- svaku repeticiju se najprije podijeli na 4 parcele na koje se rasporede stepenice glavnog faktora (A)

A_1	A_2	A_3	A_4

- dalje se svaka ta parcela podijeli na 3 podparcele na koje dolaze stepenice podfaktora (B)

	A_1			A_2			A_3			A_4	
B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3

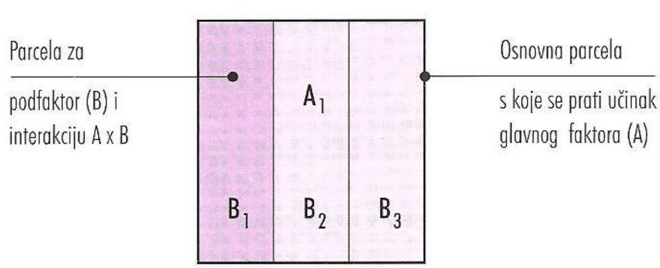
U svakoj se repeticiji i stepenice glavnog faktora i stepenice podfaktora moraju randomizirati. Tako bi konačno shema ovog pokusa bila:

I		A_1			A_2			A_3			A_4	
	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
II		A_4			A_3			A_2			A_1	
	B_3	B_1	B_2	B_1	B_3	B_2	B_1	B_3	B_2	B_2	B_3	B_1
III		A_2			A_4			A_1			A_3	
	B_2	B_3	B_1	B_2	B_3	B_1	B_1	B_3	B_2	B_3	B_2	B_1
IV		A_3			A_1			A_4			A_2	
	B_3	B_2	B_1	B_2	B_3	B_1	B_1	B_3	B_2	B_2	B_3	B_1

Razvidno je da se svaka stepenica glavnog faktora (A) u repeticiji javlja samo jedanput i to na većim parcelama, a svaka stepenica podfaktora (B) je u svakoj stepenici glavnog faktora ali na manjim parcelama.

Općenito: u split-plot shemi gradacije glavnog faktora zastupljene su u svakoj repeticiji samo jedanput, a gradacije podfaktora onoliko puta koliko glavni faktor ima stepenica.

Obavezno ih je slučajno rasporediti, ali treba i spomenuti da može jedan faktor biti raspoređen, recimo, po slučajnom bloknom rasporedu, a drugi po latinskom kvadratu, što ne predstavlja nikakav dodatni problem niti kod izvođenja pokusa, niti kod analize pokusnih podataka. Sve to ima za posljedicu da se djelovanja ovih faktora procjenjuju s različito velikih parcela i na osnovi različitog broj podataka.



Zbog različito velikih parcela u analizi podataka takvog pokusa treba procijeniti i dvije pogreške, primjerene svakom faktoru.

Ova shema obično žrtvuje preciznost procjene efekata faktora na većim parcelama, a povećava preciznost procjene efekata faktora na podparcelama i interakcije.

Upravo zbog različito velikih parcela i različite preciznosti mjerenja efekata dva faktora, izuzetno je važno paziti u koji položaj će se koji faktor staviti. Kod odlučivanja za split-plot shemu treba se rukovoditi time da

- Faktor čiji učinak se želi preciznije procijeniti rasporedi se na podparcele. Tako će, recimo, oplemenjivač staviti sorte koje želi proučiti u položaj podfaktora, a razine gnojidbe kao glavni faktor. Stručnjaku u proizvodnji će biti važnije preciznije procijeniti djelovanje različitih razina gnojidbe, pa će taj faktor staviti u poziciju podfaktora, a sorte kao glavni faktor.
- Ako se već zna o jačini djelovanja faktora tada se onaj faktor čije je djelovanje jače stavi na veće parcele, a onaj slabijeg djelovanja na podparcele. Tako se povećava mogućnost za otkrivanje razlika između stepenica faktora slabijeg djelovanja.
- Nekad je pak zbog razloga koji uvjetuju potrebu velikih parcela (kakve su u pokusima u kojima se proučava djelovanje različitih gnojidbi ili meliorativnih zahvata, u kojima se koriste poljoprivredni strojevi velikog radnog zahvata) nužno određeni faktor staviti kao glavni faktor.

Obzirom da u split-plotu faktori nisu ravnopravni, analiza podataka takvih pokusa mora biti specifična. Provodi se u više dijelova, kako bi se procijenile pogreške primjerene tako neravnopravnim faktorima. Najprije se procijeni ukupna varijabilnost cijelog pokusa, zatim se raščlanjuje na dio uvjetovan glavnim faktorom i dio zbog podfaktora i interakcije. Tako se uzima u obzir i različito velike parcele i različiti broj podataka za svaki faktor, pa se procjenjuju dvije pogreške : pogreška (a) za testiranje glavnog faktora i pogreška (b) kojom se testira djelovanje podfaktora i interakcije.

Izvori varijabilnosti	n - 1
Ukupno za cijeli pokus	$n_A \cdot n_B \cdot n_r - 1$
Ukupno za glavni faktor	$n_A \cdot n_r - 1$
Repeticije	$n_r - 1$
Glavni faktor(A)	$n_A - 1$
Pogreška (a)	$(n_r - 1)\ (n_A - 1)$
Ukupno za podfaktor	$n_A \cdot n_r\ (n_B - 1)$
Podfaktor (B)(A x B)	$n_B - 1$
Interakcija	$(n_A - 1)\ (n_B - 1)$
Pogreška (b)	$n_A\ (n_r - 1)\ (n_B - 1)$

Različite pogreške pokusa koje pripadaju faktorima A i B, razlogom su za posebnost u računanju standardne pogreške razlike (s_D) kod provođenja t-testa za usporedbu različitih prosječnih vrijednosti.

Za glavni faktor (A) $$s_D = \sqrt{\frac{2 \cdot s^2 pogreške\,(\mathrm{a})}{n_r \cdot n_B}}$$

Za podfaktor (B) $$s_D = \sqrt{\frac{2 \cdot s^2 pogreške\,(\mathrm{b})}{n_r \cdot n_A}}$$

Za interakciju, tj. kombinacije glavnog faktora i podfaktora, ovisno o stepenicama:

Za kombinacije istih stepenica faktora A, a različitih faktora B (npr. A_1B_1 - A_1B_2) $$s_D = \sqrt{\frac{2 s^2 pogreške\,(\mathrm{b})}{n_r}}$$

Za kombinacije istih ili različitih stepenica i faktora A i faktora B (npr. A_2B_1 - A_1B_2 ili A_1B_1 - A_2B_2) $$s_D = \sqrt{\frac{2\left[(n_B - 1) s_b^2 + s_a^2\right]}{n_r \cdot n_B}}$$

U ovim zadnjim usporedbama komplicira se računanje granične vrijednosti t-faktora. Ako se sa t_a i t_b označe tablični t-faktori očitani iz slobodnih varijanata pogrešaka a i b, t-faktor za takve usporedbe se računa

$$t = \frac{(n_B - 1)s_b^2 \cdot t_b + s_a^2 t_a}{(n_B - 1)s_b^2 + s_a^2}$$

Ipak, za praktičnu primjenu za uspoređivanje svih mogućih kombinacija, moglo bi se dopustiti računanje pogreške razlike, kao, $s_D = \sqrt{\frac{2s^2 pogreške\,(\mathrm{b})}{n_r}}$ jer je u najvećoj većini slučajeva ovako računana vrijednost t-faktora ipak zadovoljavajuća (Cochran & Cox).

PRIMJER 11.1.1.

U poljoprivrednoj proizvodnji uopće, jedno od najvažnijih pitanja je kada sijati ili saditi da se postigne najbolji prinos ili prirod. Tako je u zadarskom području proizvodnje kupusa bilo interesantno proučiti sorte: Yeldina (S_1), Bewama (S_2) i Hidena (S_3) sađene 20.7. (R_1), 10.8. (R_2) i 1.9. (R_3) sa konačnim ciiljem: proizvodnji preporučiti najprinosniju sortu i najpogodnije vrijeme sadnje. Zato je proveden pokus u kojem su ova dva faktora bila raspoređena po split-plot shemi u 5 ponavljanja. Razmaci sadnje su bili 0,70 m red od reda i 0,40 m u redu, veličina osnovne parcelice je bila 16.8 m^2 (tj 5 redova od po 12 biljaka), a veličina obračunske 8.4 m^2 (3 unutarnja reda sa po 10 biljaka u redu).

Plan pokusa:

I		R_1			R_2			R_3	
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
II		R_2			R_3			R_1	
	S_3	S_1	S_2	S_2	S_3	S_1	S_2	S_3	S_1
III		R_3			R_2			R_1	
	S_3	S_2	S_1	S_1	S_2	S_3	S_3	S_1	S_2
IV		R_1			R_3			R_2	
	S_2	S_3	S_1	S_2	S_3	S_1	S_1	S_2	S_3
V		R_1			R_3			R_2	
	S_3	S_2	S_1	S_3	S_2	S_1	S_3	S_2	S_1

Podaci o prinosu po kombinacijama (u kg/8.4 m^2)

Kombinacije	I	II	II	IV	V	$\sum x_{\text{Kombinacije}}$
R_1S_1	61.8	52.7	68.1	70.0	68.1	320.7
R_1S_2	56.8	52.6	64.3	54.6	50.5	278.8
R_1S_3	62.5	58.8	51.7	55.9	59.2	288.1
R_2S_1	37.7	36.8	40.6	37.3	42.3	194.7
R_2S_2	34.7	31.2	35.4	38.7	32.4	172.4
R_2S_3	*39.8*	*34.4*	*34.8*	*34.0*	*37.8*	*180.8*
R_3S_1	29.5	29.2	32.6	31.7	33.9	156.9
R_3S_2	18.8	17.3	22.2	20.7	21.7	100.7
R_3S_3	33.2	29.9	29.1	29.0	29.0	150.2
$\sum x_{\text{Repeticije}}$	374.8	342.9	378.8	371.9	374.9	$\sum x = 1843.3$

Prije nego se počne s analizom, podatke treba srediti tako da se mogu procijenit dijelovi varijabilnosti uvjetovani različitim izvorima.

Dvosmjerne tablice:

Glavni faktor – Repeticije

	R_1	R_2	R_3	
I.	181.1	112.2	81.5	
II.	164.1	102.4	76.4	
III.	184.1	110.8	83.9	
IV.	180.5	110.0	81.4	
V.	177.8	112.5	84.6	
$\sum x_{\text{Razmaka}}$	887.6	547.9	407.8	$\sum x = 1843.3$
$\bar{x}_{\text{Razmaka}}$	59.17	36.53	27.19	

Glavni faktor - Podfaktor

	S_1	S_2	S_3	
R_1	320.7	278.8	288.1	
R_2	194.7	172.4	180.8	
R_3	156.9	100.7	150.2	
$\sum x_{\text{Sorte}}$	672.3	551.9	619.1	$\sum x = 1843.3$
$\bar{x}_{\text{Sorte}}$	44.82	36.79	41.27	

ANOVA

$$SS_{\text{Ukupno za cijeli pokus}} = 61.8^2 + \ldots + 29.0^2 - \frac{1843.3^2}{45} = 9246.66$$

$$SS_{\text{Ukupno za glavni faktor R}} = \frac{181.1^2 + \ldots + 84.6^2}{3} - \frac{1843.3^2}{45} = 8234.65$$

$$SS_{\text{Repeticija}} = \frac{374.8^2 + \ldots + 374.9^2}{9} - \frac{1843.3^2}{45} = 94.84$$

$$SS_R = \frac{887.6^2 + 547.9^2 + 407.8^2}{5 \cdot 3} - \frac{1843.3^2}{45} = 8116.27$$

$$SS_{\text{Pogreške(a)}} = 8243.65 - 94.84 - 8116.27 = 23.54$$

$$SS_{\text{Ukupno za podfaktor S}} = SS_{\text{Ukupno za cijeli pokus}} - SS_{\text{Ukupno za glavni faktor R}} =$$
$$= 9246.66 - 8234.65 = 1012.01$$

$$SS_S = \frac{672.3^2 + 551.9^2 + 619.1^2}{5 \cdot 3} - \frac{1843.3^2}{45} = 485.38$$

$$SS_{\text{Interakcije RxS}} = SS_{\text{RxS kombinacija}} - SS_R - SS_S =$$
$$= \frac{320.7^2 + ... + 150.2^2}{5} - \frac{1843.3^2}{45} - 8116.27 - 485.38 =$$
$$= 135.92$$

$$SS_{\text{Pogreške (b)}} = SS_{\text{Ukupno za podfaktor S}} - SS_S - SS_{R \times S} =$$
$$= 1012.01 - 485.38 - 135.92 = 390.71$$

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno za cijeli pokus	44	9246.66				
Ukupno za glavni faktor	14	8234.65				
Repeticije	4	94.84				
R	2	8116.27	4058.13	1380.32 **	4.46	8.65
Pogreška (a)	8	23.54	2.94			
Ukupno za podfaktor S	30	1012.01				
S	2	485.38	242.69	14.91 **	3.40	5.61
R x S	4	135.92	33.98	2.09 ns	2.78	4.22
Pogreška (b)	24	390.71	16.28			

Opravdanost djelovanja glavnog faktora (rok) testirana je s pogreškom glavnog faktora (s_a^2), a podfaktora i interakcije s pogreškom podfaktora (s_b^2).

Interakcija RxS nije signifikantna, znači rok sadnje i sorta utječu na prinos kupusa potpuno nezavisno, a utjecaj svakog od njih je visoko signifikantan.

Za usporedbu prosječnih prinosa treba provesti t-test

t-test za glavni faktor (Rok sadnje)

$$s_D = \sqrt{\frac{2 \cdot 2.94}{5 \cdot 3}} = 0.626$$

$LSD_{P=5\%} = 2.31 \cdot 0.626 = 1.45\ \text{kg}/8.4\ \text{m}^2$ $\quad t_{tabl}$ očitan iz

$LSD_{P=1\%} = 3.36 \cdot 0.626 = 2.10\ \text{kg}/8.4\ \text{m}^2$ $\quad n - 1$ varijance pogreške (a)

Usporede li se prosječni prinosi postignuti u različitim rokovima sadnje, zaključuje se da je za postizanje opravdano najvećeg prinosa kupus potrebno saditi u prvom roku, jer kašnjenjem sa sadnjom prinos se značajno smanjuje

59.17 ($\bar{x}_{R_1}$) 36.53 ($\bar{x}_{R_2}$) 8.73 ($\bar{x}_{R_3}$)

22.64** 9.34**

t-test za podfaktor (Sorta)

$$s_D = \sqrt{\frac{2 \cdot 16.28}{5 \cdot 3}} = 1.473$$

$LSD_{P=5\%} = 2.06 \cdot 1.473 = 3.03\ \text{kg}/8{\cdot}4\ \text{m}^2$ $\quad t_{tabl}$ očitan iz

$LSD_{P=1\%} = 2.80 \cdot 1.473 = 4.12\ \text{kg}/8{\cdot}4\ \text{m}^2$ $\quad n - 1$ varijance pogreške (b)

44.82 ($\bar{x}_{S_1}$) 36.79 ($\bar{x}_{S_2}$) 41.27 ($\bar{x}_{S_3}$)

8.03** 4.48**

3.55**

Yeldina (S_1) je dakle sorta koju bi se zbog signifikano većeg prinosa od ostalih s vrlo velikom sigurnosti moglo preporučiti za proizvodnju.

■

PRIMJER 11.1.2.

Posađen je pokusni nasad s ranim, po prirodu vrlo različitim sortama jabuka (S_1 = Stark Earliest, S_2 = Discovery, S_3 = Julyred, S_4 = Red Melba i S_5 = Vista Bella) cijepljenim na podloge: $P_1 = M_1$, $P_2 = A_2$, P_3 = MM 106 i P_4 = H 535 a s ciljem da se otkrije možebitni utjecaj podloge i njene interakcije sa sortom na prirod jabuka.

Korišten je split-plot raspored sa 4 ponavljanja, razmak sadnje u pokusnom voćnjaku je bio 5,5 m x 4,0 m.

Shema pokusa:

S_1				S_2				S_3				S_4				S_5			
P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4
S_4				S_5				S_1				S_2				S_3			
P_4	P_1	P_3	P_2	P_3	P_4	P_2	P_1	P_2	P_4	P_1	P_3	P_1	P_4	P_2	P_3	P_2	P_4	P_1	P_3
S_3				S_1				S_5				S_4				S_2			
P_1	P_3	P_4	P_2	P_2	P_4	P_1	P_3	P_3	P_1	P_4	P_2	P_2	P_4	P_1	P_3	P_2	P_1	P_3	P_4
S_5				S_3				S_1				S_2				S_4			
P_3	P_4	P_2	P_1	P_4	P_3	P_1	P_2	P_4	P_2	P_3	P_1	P_4	P_1	P_3	P_2	P_3	P_2	P_4	P_1

Podaci o prirodu (u kg/stablo) po kombinacijama

Kombinacije	I	II	II	IV	$\sum x_{Kombinacije}$
S_1P_1	19.7	21.2	17.5	18.7	77.1
S_1P_2	20.5	14.5	13.5	15.2	63.7
S_1P_3	23.8	22.3	17.3	13.3	76.7
S_1P_4	22.2	14.5	15.7	20.0	72.4
S_2P_1	3.8	5.7	3.3	5.7	18.5
S_2P_2	5.0	4.8	4.0	5.3	19.1
S_2P_3	4.2	7.0	5.0	4.5	20.7
S_2P_4	3.3	3.0	2.7	3.8	12.8
S_3P_1	41.2	40.8	33.7	43.2	158.9
S_3P_2	39.5	31.2	31.3	32.5	134.5
S_3P_3	27.2	45.3	34.8	37.2	144.5
S_3P_4	32.3	35.5	34.7	35.0	137.5
S_4P_1	4.0	5.3	3.7	7.0	20.0
S_4P_2	5.7	4.2	5.7	4.7	20.3
S_4P_3	3.5	5.7	3.7	7.0	19.9
S_4P_4	3.7	6.3	5.0	4.3	19.3
S_5P_1	23.3	22.3	25.0	20.3	90.9
S_5P_2	11.0	21.7	14.2	27.2	74.1
S_5P_3	23.8	22.5	24.2	19.2	89.7
S_5P_4	17.3	39.5	21.8	12.8	91.4
$\sum x_{Repeticije}$	335.0	373.3	316.8	336.9	$\sum x = 1362.0$

Dvosmjerne tablice:
Glavni faktor (S) - Repeticije

Repeticije	S_1	S_2	S_3	S_4	S_5	
I.	86.2	16.3	140.2	16.9	75.4	
II.	72.5	20.5	152.8	21.5	106.0	
III.	64.0	15.0	134.5	18.1	85.2	
IV.	67.2	19.3	147.9	23.0	79.5	
$\sum x_{Sorte}$	289.9	71.4	575.4	79.5	346.1	$\sum x = 1362.0$
$\bar{x}_{Sorte}$	18.12	4.44	35.96	4.97	21.63	

Glavni faktor - Podfaktor

	P_1	P_2	P_3	P_4	$\sum x_{\text{Sorte}}$
S_1	77.1	63.7	76.7	72.4	289.9
S_2	18.5	19.1	20.7	12.8	71.1
S_3	158.9	134.5	144.5	137.5	575.4
S_4	20.0	20.3	19.9	19.3	79.5
S_5	90.9	74.1	89.7	91.4	346.1
$\sum x_{\text{Podloge}}$	365.4	311.7	351.5	333.4	$\sum x = 1362.0$
$\bar{x}_{\text{Podloge}}$	18.27	15.58	17.57	16.67	

ANOVA

$$SS_{\text{Ukupno za cijeli pokus}} = 19.7^2 + \ldots + 12.8^2 - \frac{1362.0^2}{80} = 12179.75$$

$$SS_{\text{Ukupno za glavni faktor S}} = \frac{86.2^2 + \ldots + 79.5^2}{4} - \frac{1362.0^2}{80} = \\ = 11225.81$$

$$SS_{\text{Repeticija}} = \frac{335.0^2 + \ldots + 336.9^2}{20} - \frac{1362.0^2}{80} = 84.04$$

$$SS_{\text{S}} = \frac{289.9^2 + \ldots + 346.1^2}{4 \cdot 4} - \frac{1362.0^2}{80} = 10954.94$$

$$SS_{\text{Pogreške (a)}} = 11225.81 - 84.04 - 10954.94 = 186.83$$

$$SS_{\text{Ukupno za podfaktor P}} = 12179.75 - 11225.81 = 953.94$$

$$SS_{\text{P}} = \frac{365.4^2 + \ldots + 333.4^2}{4 \cdot 5} - \frac{1362.0^2}{80} = 81.04$$

$$SS_{\text{SxP}} = \frac{77.1^2 + \ldots + 91.4^2}{4} - \frac{1362.0^2}{80} - 10954.94 - 81.04 = 97.56$$

$$SS_{\text{Pogreške (b)}} = 953.94 - 81.04 - 97.56 = 775.34$$

Tablica ANOVA

Izvori varijabilnosti	n - 1	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno za cijeli pokus	79	12179.75				
Ukupno za glavni faktor S	19	11225.81				
Repeticije	3	84.04				
S	4	10954.94	2738.73	175.90 **	3.26	5.41
Pogreška (a)	12	186.83	15.57			
Ukupno za podfaktor P	60	953.94				
P	3	81.04	27.01	1.57 ns	2.84	4.31
S x P	12	97.56	8.13	0.47 ns	2.00	2.66
Pogreška (b)	45	775.34	17.23			

Rezultati F-testova pokazuju da samo sorta značajno utječe na prirod (znači sorte su opravdano različitih priroda; no to je već bilo poznato i nije bio cilj to dokazati)

Kad bi trebalo, t-test za faktor S bi se proveo na način:

$$s_D = \sqrt{\frac{2 \cdot 15.57}{4 \cdot 4}} = 1.295$$

$LSD_{P=5\%} = 2.18 \cdot 1.395 = 3.04$ kg/stablu $\quad t_{tabl}$ očitan iz
$LSD_{P=1\%} = 3.06 \cdot 1.395 = 4.27$ kg/stablu $\quad n - 1$ varijance pogreške (a)

Niti utjecaj podloge niti interakcija nisu statistički opravdani. To znači da se prirodi ovih sorata cijepljenih na različite podloge zanemarivo malo (neopravdano) razlikuju.

■

11.2. Split-blok ili strip-plot

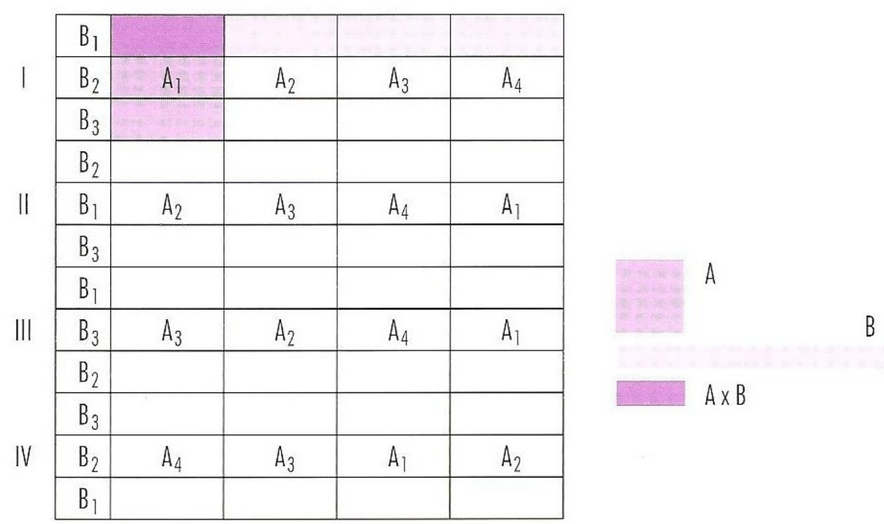
To je također shema za dvofaktorijalne pokuse, u kojoj se gradacije jednog faktora raspoređuju u trake (=strip) kroz svaku repeticiju ili blok. U svakoj repeticiji moraju stepenice i jednog i drugog faktora biti potpuno slučajno raspoređene.

To je raspored koji omogućuje uključivanje faktora koje bi bilo teško primijeniti na malim parcelama (npr. pokus s različitim načinima predsjetvene obrade i nekoliko gradacija N-gnojidbe, uz pretpostavku da se oboje izvodi strojem).

Split-blok shema žrtvuje preciznost procjene učinaka faktora, s povećanjem preciznosti za interakciju. Interakcija se odredi i bolje nego primjenom split-plota.

Kako je uz to i broj slobodnih varijanata za procjenu pogrešaka za faktore obično malen, split-blok shemu ne bi trebalo koristiti osim ako to neizbježno nalažu uvjeti izvedbe eksperimenta (potrebe velikih parcela za faktore, ili korištenje poljoprivrednih strojeva u primjeni gradacija oba faktora i sl.).

Za isti onaj primjer 4 x 3 u 4 repeticije koji smo uzeli objašnjavajući split-plot shemu, split-blok shema bi bila:



Zbog specifičnosti u dijeljenju parcela, kod ovog se plana pokusa radi sa tri različite veličine osnovnih parcela: one na kojima se prati djelovanje faktora A, druge sa kojih se otkriva učinak faktora B i treće za interakciju A x B.

Stoga je i statistička analiza nešto kompliciranija jer se procjenjuju tri pogreške: posebno za svaki faktor i posebno za interakciju.

Izvori varijabilnosti	$n - 1$
Ukupno za cijeli pokus	$n_A \cdot n_B \cdot n_r - 1$
Ukupno za glavni faktor A	$n_A \cdot n_r - 1$
Repeticije	$n_r - 1$
A	$n_A - 1$
Pogreška (a)	$(n_A - 1)\ (n_r - 1)$
Ukupno za podfaktor B	$n_B \cdot n_r - 1$
Repeticije	$n_r - 1$
B	$n_B - 1$
Pogreška (b)	$(n_B - 1)\ (n_r - 1)$
Interakcija A x B	$(n_A - 1)\ (n_B - 1)$
Pogreška (c)	$(n_r - 1)\ (n_A - 1)\ (n_B - 1)$

Pogreške razlike, potrebne za provođenje t testa za uspoređivanje prosječnih vrijednosti, računaju se po formulama:

Za faktor A $$s_D = \sqrt{\frac{2 \cdot s^2{}_{pogreške\ (a)}}{n_r \cdot n_B}}$$

Za faktor B $$s_D = \sqrt{\frac{2 \cdot s^2{}_{pogreške\ (b)}}{n_r \cdot n_A}}$$

Za interakciju:

Za usporedbu kombinacija istih stepenica faktora B, a različitih za faktor A (npr. A_1B_1-A_2B_1) $$s_D = \sqrt{\frac{2\left[(n_B - 1)s_c^2 + s_a^2\right]}{n_r \cdot n_B}}$$

Za usporedbu kombinacija istih stepenica faktora A, a različitih za faktor B (npr. A_1B_1-A_1B_2)

$$s_D = \sqrt{\frac{2\left[(n_A - 1)s_c^2 + s_b^2\right]}{n_r \cdot n_A}}$$

Ipak, treba opet naglasiti i ne zaboraviti da je split-blok shema dobra uglavnom za interakciju (rijetko je kad broj slobodnih varijanata pogrešaka za faktore A i B dovoljno velik da bi testovi bili pouzdani).

PRIMJER 11.2.1.

Često se zbog nepovoljnih prilika u jesen, sa sjetvom pšenice mora kasniti. Kako će na to reagirati pojedine sorte, istražilo se u pokusu sa 5 sorata: Super Zlatna (S_1), Baranjka (S_2), Sanja (S_3), Zg 6569/76 (S_4) i Zg 471/80 (S_5), koje su se sijale u ranom (između 1. i 5. listopada), optimalnom (između 15. i 20 listopada), zakašnjelom (između 05. i 10. studenog) i vrlo, vrlo kasnom roku (od 20. do 25. studenog). Radi se, dakle, o 5 x 4 pokusu.

Glavni cilj pokusa je bio otkriti kako će pojedine sorte reagirati na rokove sjetve tj. učinke interakcije sorte i roka na prinos zrna. Zato se koristio split-blok raspored sa 6 ponavljanja. Obračunska parcela je bila veličine 7.15 m^2.

Shema pokusa i podaci o prinosu zrna s 13% vlage
(u kg/parceli 7.15 m^2):

	S_1		5.20		5.18		4.48		4.76
	S_2		5.93		5.39		5.07		5.30
I	S_3	R_1	5.63	R_2	5.83	R_3	5.59	R_4	5.23
	S_4		5.66		5.77		5.65		5.44
	S_5		5.38		5.77		5.35		5.59
	S_2		5.48		5.47		5.06		5.60
	S_4		5.76		5.56		4.94		5.44
II	S_1	R_2	6.00	R_4	3.99	R_3	5.80	R_1	5.56
	S_5		5.98		5.91		5.40		5.56
	S_3		5.55		5.72		5.12		5.88
	S_1		5.38		5.46		4.99		6.05
	S_4		5.61		5.76		5.41		6.10
III	S_2	R_4	4.90	R_1	5.61	R_3	5.91	R_2	6.03
	S_5		5.94		5.84		5.46		6.35
	S_3		6.04		5.96		5.22		6.09
	S_1		5.72		5.23		5.16		5.82
	S_5		5.91		5.38		5.51		6.17
IV	S_3	R_2	6.06	R_3	5.61	R_4	5.77	R_1	5.80
	S_4		5.66		5.22		5.51		5.76
	S_2		5.96		5.12		5.28		5.64
	S_3		5.79		5.80		5.39		6.26
	S_5		5.89		5.50		5.50		6.11
V	S_1	R_1	5.99	R_3	5.22	R_4	3.85	R_2	6.06
	S_2		5.81		4.72		5.46		5.93
	S_4		5.90		5.50		5.50		6.30
	S_4		5.06		4.94		5.76		5.83
	S_1		5.15		5.37		6.15		5.68
VI	S_5	R_3	6.06	R_4	5.31	R_2	6.24	R_1	5.56
	S_3		5.39		5.90		5.93		5.71
	S_2		5.18		4.91		5.86		5.36

Dvosmjerne tablice
Rokovi - Repeticije

	I	II	III	IV	V	VI	$\sum x_R$
R_1	27.80	28.04	28.63	29.19	29.38	28.14	171.18
R_2	27.94	28.77	30.62	29.31	30.66	29.94	177.24
R_3	26.14	26.32	26.99	26.56	26.74	26.84	159.59
R_4	26.32	26.65	27.87	27.23	25.70	26.43	160.20
$\sum x_{\text{Repeticije}}$	108.20	109.78	114.11	112.29	112.48	111.35	$\sum x = 66821$

Sorte - Repeticije

	I	II	III	IV	V	VI	$\sum x_S$
S_1	19.62	23.35	21.88	21.93	21.12	22.35	128.25
S_2	21.69	21.61	22.45	22.00	21.92	21.31	130.98
S_3	22.28	22.27	23.31	23.24	23.24	22.93	137.27
S_4	22.52	21.70	22.88	22.15	23.20	21.59	134.04
S_5	22.09	22.85	23.59	22.97	23.00	23.17	137.67

Rokovi - Sorte

	S_1	S_2	S_3	S_4	S_5
R_1	33.71	33.95	34.77	34.35	34.40
R_2	35.16	34.65	35.72	35.35	36.36
R_3	30.87	31.06	32.73	31.78	33.15
R_4	28.51	31.32	34.05	32.56	33.76

ANOVA

$$SS_{\text{Ukupno za cijeli pokus}} = 5.20^2 + \ldots + 5.31^2 - \frac{668.21^2}{120} = 21.4376$$

$$SS_{\text{Ukupno za faktor R}} = \frac{27.80^2 + \ldots + 26.43^2}{5} - \frac{668.21^2}{120} = 9.7030$$

$$SS_{\text{Repeticija}} = \frac{108.20^2 + \ldots + 111.35^2}{20} - \frac{668.21^2}{120} = 1.1082$$

$$SS_{\text{R}} = \frac{171.18^2 + \ldots + 160.20^2}{6 \cdot 5} - \frac{668.21^2}{120} = 7.4489$$

$$SS_{\text{Pogreške (a)}} = 9.7030 - 1.1082 - 7.4489 = 1.1459$$

$$SS_{\text{Ukupno za faktor S}} = \frac{19.62^2 + \ldots + 23.17^2}{4} - \frac{668.21^2}{120} = 5.2077$$

$$SS_{\text{Repeticija}} = 1.1082$$

$$SS_S = \frac{128.25^2 + \ldots + 137.67^2}{6 \cdot 4} - \frac{668.21^2}{120} = 2.7377$$

$$SS_{\text{Pogreške (b)}} = 5.2077 - 1.1082 - 2.7377 = 1.3618$$

$$SS_{RxS} = \frac{33.71^2 + \ldots + 33.76^2}{6} - \frac{668.21^2}{120} - 7.4489 - 2.7377 = = 1.6952$$

$$SS_{\text{Pogreške (c)}} = 21.4376 - 9.7030 - 2.7377 - 1.3618 - 1.6952 = = 5.9399$$

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno za cijeli pokus	4·5·6 – 1 = 119	21.4376				
Ukupno za R	4·6 – 1 = 23	9.7030				
Repeticije	6 – 1 = 5	1.1082				
R	4 – 1 = 3	7.4489	2.4830	32.5 **	3.29	5.42
Pogreška (a)	23–5–3 = 15	1.1459	0.0764			
Ukupno za S	5·6 – 1 = 29	5.2077				
Repeticije	6 – 1 = 5	1.1082				
S	5 – 1 = 4	2.7377	0.6844	10.0 **	2.87	4.43
Pogreška (b)	29–5–4 = 20	1.3618	0.0681			
RxS	3·4 = 12	1.6952	0.1413	1.4 ns	1.92	2.50
Pogreška (c)	119–23–24–12 = 60	5.9399	0.0990			

Split-blok raspored, koji je izuzetno dobra shema za procjenu interakcije, za ovaj je pokus odabrana baš zbog toga što je bilo najvažnije doći do saznanja hoće li i kako različito vrijeme sjetve utjecati na prinos sorata pšenice uključenih u pokus.

Iz nesignifikantnog F-testa interakcije, mora se zaključiti da se ovim eksperimentom nije uspjelo utvrditi zavisnost djelovanja faktora, tj. opravdana različitost prinosa sorata sijanih u različitim rokovima.

U tablici prosječnih prinosa (kombinacija, sorata i rokova) zamjetne su razlike među kombinacijama, no nisu statistički značajne već slučajne.

	S_1	S_2	S_3	S_4	S_5	$\overline{x}_{Roka}$
R_1	5.62	5.66	5.79	5.72	5.73	5.71
R_2	5.86	5.77	5.95	5.89	6.06	5.91
R_3	5.14	5.18	5.45	5.30	5.52	5.32
R_4	4.75	5.22	5.67	5.43	5.63	5.34
$\overline{x}_{Sorte}$	5.34	5.46	5.72	5.58	5.74	

Unatoč slabosti split-bloka u procjeni efekata faktora, u ovom pokusu ćemo nakon opravdanih F-testova za faktore ipak provesti i t-testove (da se pokaže kako, a i obzirom na dovoljno velik broj slobodnih varijanata i pogreške (a) i pogreške (b).

t-test za faktor Rok

$$s_D = \sqrt{\frac{2 \cdot 0.0764}{6 \cdot 5}} = 0.071$$

$LSD_{P=5\%} = 2.13 \cdot 0.071 = 0.15\ kg/7.15\ m^2$ t_{tabl} očitan iz
$LSD_{P=1\%} = 2.95 \cdot 0.071 = 0.21\ kg/7.15\ m^2$ n-1 pogreške (a)

I u ovom pokusu je potvrđeno da je optimalno vrijeme sjetve pšenice 15. do 20 listopad (R_2). Naime, u tom je roku postignut signifikantno najveći prinos. Jer:

$$5.91\ (\overline{x}_{R_2}) - 0.15\ (LSD_{P=5\%}) = 5.76\ kg/7.15\ m^2$$

$$5.91\ (\overline{x}_{R_2}) - 0.21\ (LSD_{P=1\%}) = 5.70\ kg/7.15\ m^2$$

Prosječni prinosi manji od 5.76 ili 5.70 kg/parceli 7.15 m^2 statistički su opravdano manji od prinosa u R_2 (uz P=5% ili P=1%).

t-test za faktor Sorta

$$s_D = \sqrt{\frac{2 \cdot 0.0681}{6 \cdot 4}} = 0.075$$

$LSD_{P=5\%} = 2.09 \cdot 0.075 = 0.16$ kg/7.15 m^2 t_{tabl} očitan iz

$LSD_{P=1\%} = 2.84 \cdot 0.075 = 0.21$ kg/7.15 m^2 n-1 pogreške (b)

$5.74\ (\bar{x}_{S_5}) - 0.16\ (LSD_{P=5\%}) = 5.58$ kg/7.15 m^2

$5.74\ (\bar{x}_{S_5}) - 0.21\ (LSD_{P=1\%}) = 5.53$ kg/7.15 m^2

Apsolutno najveći prinos zrna je postigla sorta S_5 (tj. Zg 471/80), no ne i signifikantno veći od prinosa sorata S_3 (Sanja) i S_4 (Zg 6569/76). Statistički značajno manji prinos od S_5 imale su sorte S_1 (Super zlatna) i Baranjka (S_2).

PRIMJER 11.2.2.

Tri sorte krušaka (Boskova bočica, Viljamovka i Trevuška) cijepljene su na tri međupodloge (Hardenpontova, Gelertova i Carska), pa je bilo važno proučiti ima li interakcija sorte i međupodloge utjecaja na bujnost i ostala svojstva. Upravo zbog toga se kod sadnje eksperimentalnog voćnjaka odlučilo za split blok shemu sa tri ponavljanja.

Shema pokusa i podaci o duljini jednogodišnjih izboja (m):

	I.			II.			III.			
	H	G	C	H	G	C	H	G	C	$\sum x$ Vod. rep.
I.	Boskova bočica			Viljamovka			Trevuška			
	30.9	33.4	40.7	60.5	46.6	52.4	55.8	25.7	26.5	372.5
II.	Viljamovka			Trevuška			Boskova bočica			
	57.1	52.4	39.6	61.7	17.5	26.9	19.2	24.1	40.1	338.6
III.	Trevuška			Boskova bočica			Viljamovka			
	20.1	40.9	18.4	29.6	38.2	15.8	45.7	33.9	41.0	283.6
$\sum x$ Okom. rep.		333.5			349.2			312.0		$\sum x = 994.7$

U ovoj shemi treba uočiti da:

- tri su vrste parcela različitih po veličini
- sorte su raspoređene po planu latinskog kvadrata, a međupodloge u trakama kroz okomite repeticije za sorte (to znači repeticije za međupodloge su okomite repeticije za sorte).

O tome treba naravno voditi računa u analizi podataka ovog pokusa.

Dvosmjerne tablice:
Vodoravne repeticije - Okomite repeticije

	I	II	III	$\sum x$ Vod. rep.
I	BB	V	T	
	105.0	159.5	108.0	372.5
II	V	T	BB	
	149.1	106.1	83.4	338.6
III	T	BB	V	
	79.4	83.6	120.6	283.6
$\sum x$ Okom. rep.	333.5	349.2	312.0	$\sum x = 994.7$

Sorte - Međupodloge

	C	G	H	$\sum x$ Sorte
BB	96.6	95.7	79.7	272.0
V	133.0	132.9	163.3	429.2
T	71.8	84.1	137.6	293.5
$\sum x$ Međupodloge	301.4	312.7	380.6	$\sum x = 994.7$

Međupodloge - Okomite repeticije

	I	II	III	$\sum x$ Međupodloge
C	98.7	95.1	107.6	301.4
G	126.7	102.3	83.7	312.7
H	108.1	151.8	120.7	380.6
$\sum x$ Okom. rep.	333.5	349.2	312.0	$\sum x = 994.7$

ANOVA

$$SS_{\text{Ukupno za cijeli pokus}} = 30.9^2 + ... + 41.0^2 - \frac{994.7^2}{27} = 5022.99$$

$$SS_{\text{Ukupno za faktor S}} = \frac{105.0^2 + ... + 120.6^2}{3} - \frac{994.7^2}{27} = 2158.02$$

$$SS_{\text{Vodoravnih repeticija}} = \frac{372.5^2 + 338.6^2 + 283.6^2}{9} - \frac{994.7^2}{27} =$$

$$= 447.31$$

$$SS_{\text{Okomitih repeticija}} = \frac{333.5^2 + 349.2^2 + 312.0^2}{9} - \frac{994.7^2}{27} = 77.50$$

$$SS_S = \frac{272.0^2 + 429.2^2 + 293.5^2}{9} - \frac{994.7^2}{27} = 1614.39$$

$$SS_{\text{Pogreške (a)}} = 2158.02 - 447.31 - 77.50 - 1614.39 = 18.82$$

$$SS_{\text{Ukupno za faktor M}} = \frac{98.7^2 + ... + 120.7^2}{3} - \frac{994.7^2}{27} = 1082.74$$

$$SS_{\text{Repeticija}} = SS_{\text{Okomitih repeticija}} = 77.50$$

$$SS_M = \frac{301.4^2 + 312.7 + 380.6^2}{9} - \frac{994.7^2}{27} = 407.80$$

$$SS_{\text{Pogreške (b)}} = 1082.74 - 77.50 - 407.80 = 597.44$$

$$SS_{\text{Interakcije SxM}} = SS_{\text{SxM kombinacija}} - SS_S - SS_M =$$

$$= \frac{96.6^2 + ... + 137.6^2}{3} - \frac{994.7^2}{27} - 1614.39 - 407.80 = 673.07$$

$$SS_{\text{Pogreške (c)}} = 5022.99 - 2158.02 - 407.80 - 597.44 - 673.27 =$$

$$=1186.66$$

Tablica ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno za cijeli pokus	26	5022.99				
Ukupno za faktor S	8	2158.02				
Vodoravne repeticije	2	447.31				
Okomite repeticije	2	77.50				
S	2	1614.39	807.19	85.78 *	19.0	99.0
Pogreška (a)	2	18.82	9.41			
Ukupno za faktor M	8	1082.74				
Repeticije (=okomite)	2	77.50				
M	2	407.80	203.90	1.36 ns	6.94	18.0
Pogreška (b)	4	597.43	149.36			
S x M	4	673.07	168.27	1.13 ns	3.84	7.0
Pogreške (c)	8	1186.67	148.33			

Rezultati ANOVA-e pokazuju da interakcija nije signifikantna što znači da međupodloge u kombinaciji sa sortama nemaju nikakvog značajnog utjecaja na bujnost izraženu duljinom jednogodišnjih izboja. Bez obzira na opravdanost, F-test za sorte nema nikakav realan značaj jer je broj slobodnih varijanata varijance pogreške puno premalen. Osim toga, pokus je i izveden po ovoj shemi isključivo zbog interesa za interakciju.

■

USPOREDBA SHEMA

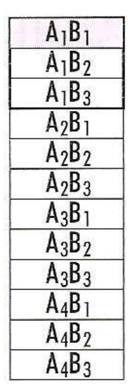
Nakon što smo proučili primjenu, prednosti i nedostatke različitih shema po kojima se mogu izvesti dvofaktorijalni pokusi, način statističke analize podataka, testiranja opravdanosti učinaka faktora i interakcija, korisno bi bilo učiniti usporedbu. Uzmimo opet onaj 4x3 primjer u 4 ponavljanja, pa napravimo pregled dijelova varijabilnosti koji se mogu definirati, i način testiranja opravdanosti djelovanja pojedinih faktora i interakcije, ovisno o korištenoj shemi pokusa.

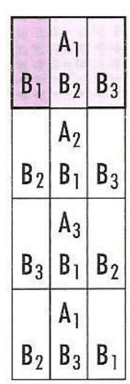
Kod slučajnog bloknog rasporeda sve kombinacije faktora zastupljene su u repeticiji jedamput i sve na parcelicama jednake

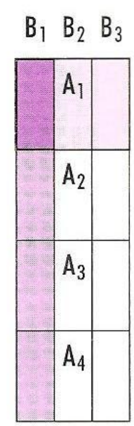
Slučajni blokni raspored	
Ukupno	47
Repeticije	3
Kombinacije	11
Pogreška	33
A	3
B	2
AxB	6

Split plot	
Ukupno	47
Ukupno za glavni faktor	15
Repeticije	3
A	3
Pogreška (a)	9
Ukupno za podfaktor	32
B	2
AxB	6
Pogreška (b)	24

Split blok	
Ukupno	47
Ukupno za glavni faktor	15
Repeticije	3
A	3
Pogreška (a)	9
Ukupno za podfaktor	11
Repeticije	3
B	2
Pogreška (b)	6
AxB	6
Pogreška (c)	18







veličine. Zbog toga se i F-test za oba faktora i interakciju provodi s jedinstvenom pogreškom pokusa.

Za split-plot shemu su karakteristične dvovrsne parcele pa su i dvije pogreške pokusa: s onom za veće parcele testira se opravdanost djelovanja glavnog faktora, a s pogreškom s manjih parcela testira se podfaktor i interakcija.

Ako su stepenice podfaktora raspoređene u trake (pojaseve) kroz svaku repeticiju kao što je slučaj kod split-blok rasporeda, tada se javljaju tri vrste različito velikih parcela, a otud se dolazi i do tri pogreške u pokusu. Svaka od njih je primjerena faktorima i interakciji.

11.3. Split-split plot

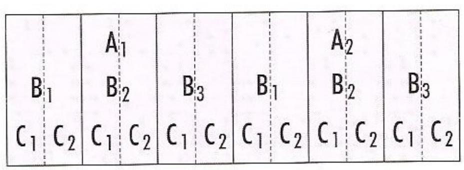
Već i sam naziv ove sheme upućuje na proširenje na tri faktora (A, B i C) i uzastopno dijeljenje parcela obzirom na gradacije tih faktora.

Split-split- plot raspored stoga daje tri razine preciznosti u procjeni efekata faktora: najmanja je za prvi faktor (glavni faktor), a najveća za treći (podpodfaktor) i interakcije s trećim faktorom.

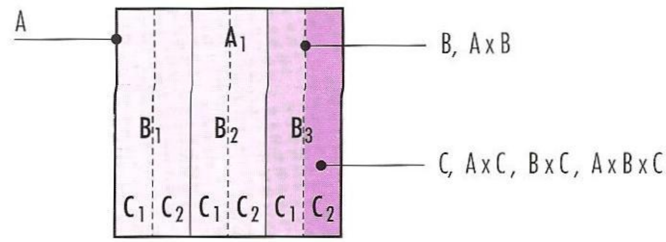
Shema se izvodi tako da se svako ponavljanje najprije podijeli na onoliko dijelova koliko glavni faktor ima gradacija, zatim se svaka ta stepenica razdijeli na onoliko dijelova koliko podfaktor ima stepenica i, konačno, unutar svake stepenice podfaktora dolaze stepenice podpodfaktora.

Dakako, u svakoj repeticiji se stepenice svakog faktora moraju slučajno rasporediti.

Za recimo 2x3x2 pokus jedna bi repeticija izgledala ovako:



Tako se stvaraju tri različite veličine parcela, pri čemu se s najveće prati djelovanje glavnog faktora (A), sa podaparcele učinak podfaktora B i interakcije AxB, a s najmanje parcele podpodfaktora C i svih interakcija s podpodfaktorom: AxC, BxC i AxBxC.



Zato je i analiza podataka takvog pokusa kompleksnija:

Izvori varijabilnosti	$n - 1$
Ukupno za cijeli pokus	$n_A \cdot n_B \cdot n_C \cdot n_r - 1$
Ukupno za glavni faktor	$n_A \cdot n_r - 1$
Repeticije	$n_r - 1$
Glavni faktor A	$n_A - 1$
Pogreška (a)	$(n_A - 1)(n_r - 1)$
Ukupno za podfaktor	$n_A \cdot n_r (n_B - 1)$
Podfaktor B	$n_B - 1$
AxB	$(n_A - 1)(n_B - 1)$
Pogreška (b)	$n_A (n_B - 1)(n_r - 1)$
Ukupno za podpodfaktor	$n_A \cdot n_B \cdot n_r (n_C - 1)$
Podfaktor C	$n_C - 1$
AxC	$(n_A - 1)(n_C - 1)$
BxC	$(n_B - 1)(n_C - 1)$
AxBxC	$(n_A - 1)(n_B - 1)(n_C - 1)$
Pogreška (c)	$n_A \cdot n_B (n_r - 1)(n_C - 1)$

Odnos faktora u split-split-plotu, veličine parcela i procijenjene pogreške, nameću potrebu i za kompliciranim računanjem standardne pogreške razlike za usporedbe pojedinih prosječnih vrijednosti u koje je uključen faktor C:

Prosječne vrijednosti koje se uspoređuju	s_D
$\bar{x}_{C_1} - \bar{x}_{C_2}$	$\sqrt{\frac{2 \cdot s^2 pogreške\,(c)}{n_r \cdot n_A \cdot n_B}}$
$\bar{x}_{A_1C_1} - \bar{x}_{A_2C_2}$	$\sqrt{\frac{2 \cdot s_c^2}{n_r \cdot n_B}}$
$\bar{x}_{B_2C_2} - \bar{x}_{B_2C_1}$	$\sqrt{\frac{2 \cdot s_c^2}{n_r \cdot n_A}}$
$\bar{x}_{A_2B_2C_2} - \bar{x}_{A_2B_2C_1}$	$\sqrt{\frac{2 \cdot s_c^2}{n_r}}$
$\bar{x}_{B_2C_2} - \bar{x}_{B_1C_2}$ ili $\bar{x}_{B_2C_2} - \bar{x}_{B_1C_1}$	$\sqrt{\frac{2\left[(n_C - 1)s_c^2 + s_b^2\right]}{n_r \cdot n_A \cdot n_C}}$
$\bar{x}_{A_2B_2C_2} - \bar{x}_{A_2B_1C_2}$	$\sqrt{\frac{2\left[(n_C - 1)s_c^2 + s_b^2\right]}{n_r \cdot n_C}}$
$\bar{x}_{A_2C_2} - \bar{x}_{A_1C_2}$ ili $\bar{x}_{A_2C_2} - \bar{x}_{A_1C_1}$	$\sqrt{\frac{2\left[(n_C - 1)s_c^2 + s_a^2\right]}{n_r \cdot n_B \cdot n_C}}$
$\bar{x}_{A_2B_2C_2} - \bar{x}_{A_1B_2C_2}$	$\sqrt{\frac{2\left[(n_B - 1)s_c^2 + (n_B - 1)s_b^2 + s_a^2\right]}{n_r \cdot n_C \cdot n_B}}$

Razvidno je koliko treba biti oprezan kod uspoređivanja i testiranja razlika između pojedinih prosječnih vrijednosti kombinacija faktora. Samo mali propust u tome, dovodi do upitnih zaključaka.

U svojoj sam praksi primjetila da se ovaj raspored koristi, a da se očito ne znaju slabosti, opasnosti i posljedice najčešće pogrešne interpretacije rezultata. Razlog primjene je najvjerojatnije taj, što se čini praktičnim primjenjivati stepenice pojedinih faktora redom: na velike parcele, unutar njih manje i unutar na one najmanje.

Zbog toga ćemo ga ipak ilustrirati primjerom.

PRIMJER 11.3.1.

Svrha ovog istraživanja je bila proučiti utjecaj gnojidbe dušikom i primjene CCC (Cycocel - regulator rasta koji utječe na skraćivanje stabljike) na prinos nekih sorata pšenice.

Izveden je 4 x 3 x 2 pokus po split-split plot shemi, sa 6 repeticija. Glavni faktor su bile sorte, podfaktor gnojidbe dušikom, a podpodfaktor primjena CCC.

Veličina osnovne parcele je bila 7.7 m^2, a obračunske 7.15 m^2.

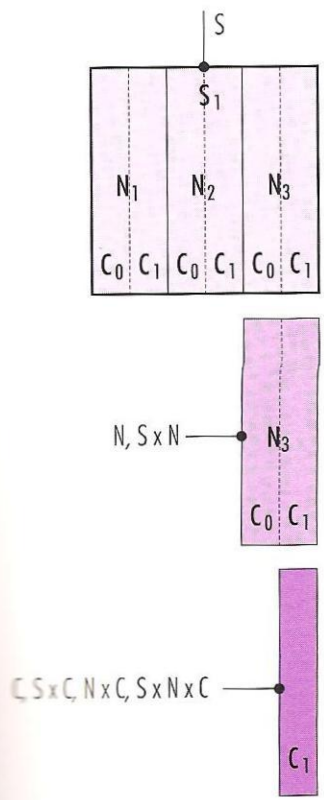
Uključene su bile ove gradacije pojedinih faktora:

Sorte: Superzlatna (S_1), Zg 471 (S_2), Sana (S_3) i Sivka (S_4).

Gnojidbe dušikom: 136 kg N/ha (N_1), 166 kg N/ha (N_2) i 196 kg N/ha (N_3)

Cycocel: bez CCC (C_0) i tretirano sa CCC (C_1).

Vidljive su tri različite veličine parcela s kojih se prate učinci pojedinih faktora i interakcija. Naime:



Plan pokusa:

I	S_1 N_1 N_1 N_1 C_0 C_1 C_0 C_1 C_0 C_1	S_2 N_2 N_2 N_3 C_0 C_1 C_0 C_1 C_0 C_1	S_3 N_3 N_3 N_4 C_0 C_1 C_0 C_1 C_0 C_1	S_4 N_4 N_4 N_4 C_0 C_1 C_0 C_1 C_0 C_1
II	S_3 N_1 N_1 N_1 C_1 C_0 C_0 C_1 C_1 C_0	S_4 N_2 N_1 N_3 C_1 C_0 C_0 C_1 C_0 C_1	S_1 N_3 N_3 N_4 C_1 C_0 C_0 C_1 C_1 C_0	S_2 N_4 N_4 N_4 C_0 C_1 C_1 C_0 C_0 C_1
III	S_4 N_1 N_1 N_1 C_0 C_1 C_1 C_0 C_0 C_1	S_3 N_2 N_2 N_3 C_1 C_0 C_1 C_0 C_0 C_1	S_2 N_3 N_3 N_4 C_0 C_1 C_1 C_0 C_0 C_1	S_1 N_4 N_4 N_4 C_1 C_0 C_1 C_0 C_0 C_1
IV	S_2 N_1 N_1 N_1 C_1 C_0 C_0 C_1 C_1 C_0	S_1 N_2 N_2 N_3 C_0 C_1 C_1 C_0 C_1 C_0	S_4 N_3 N_3 N_4 C_0 C_1 C_0 C_1 C_1 C_0	S_3 N_4 N_4 N_4 C_1 C_0 C_0 C_1 C_0 C_1
V	S_4 N_1 N_1 N_1 C_1 C_0 C_1 C_0 C_0 C_1	S_3 N_2 N_2 N_3 C_0 C_1 C_0 C_1 C_1 C_0	S_2 N_3 N_3 N_4 C_1 C_0 C_0 C_1 C_1 C_0	S_1 N_4 N_4 N_4 C_0 C_1 C_1 C_0 C_0 C_1
VI	S_3 N_1 N_1 N_1 C_1 C_0 C_0 C_1 C_1 C_0	S_1 N_2 N_2 N_3 C_0 C_1 C_1 C_0 C_0 C_1	S_4 N_3 N_3 N_4 C_0 C_1 C_1 C_0 C_0 C_1	S_2 N_4 N_4 N_4 C_1 C_0 C_0 C_1 C_1 C_0

Prinosi zrna s 14% vlage u kg/parceli 7.15 m^2.

Kombinacije			I	II	III	IV	V	VI	$\sum x$ Kombinacije
S_1	N_1	C_0	4.76	4.52	3.94	5.43	4.71	4.53	27.89
		C_1	4.22	4.87	4.02	5.07	3.58	4.50	26.26
	N_2	C_0	5.29	5.76	4.66	5.21	5.67	4.63	31.22
		C_1	5.41	4.51	4.79	5.48	3.79	5.00	28.98
	N_3	C_0	4.60	5.31	5.25	5.55	5.91	4.99	31.61
		C_1	5.09	3.77	5.52	5.41	4.65	5.12	29.56
S_2	N_1	C_0	5.42	4.85	4.32	5.20	4.91	5.26	29.96
		C_1	5.11	4.28	4.87	4.86	3.99	5.37	28.48
	N_2	C_0	6.05	5.91	5.37	5.17	5.88	5.52	33.90
		C_1	6.07	5.79	5.69	5.41	4.31	5.78	33.05
	N_3	C_0	5.81	5.15	6.14	5.58	6.15	5.68	34.51
		C_1	4.83	4.56	6.22	5.79	4.96	6.01	32.37
S_3	N_1	C_0	5.44	5.09	5.18	6.26	5.71	6.30	33.98
		C_1	4.84	5.76	4.86	5.03	4.84	6.01	31.34
	N_2	C_0	6.79	6.51	5.75	5.89	5.45	6.24	36.63
		C_1	5.58	5.41	5.55	5.17	4.52	6.02	32.25
	N_3	C_0	6.21	6.03	6.13	6.77	6.28	6.15	37.57
		C_1	6.01	4.71	6.37	5.65	6.01	6.02	34.77
S_4	N_1	C_0	4.81	4.60	3.03	4.88	4.51	5.03	26.86
		C_1	3.82	2.06	2.97	4.12	4.07	5.04	22.08
	N_2	C_0	5.03	5.61	3.98	4.46	5.74	4.96	29.78
		C_1	4.75	4.50	4.39	4.27	4.35	4.95	27.21
	N_3	C_0	4.91	5.33	4.94	4.54	5.17	5.06	29.95
		C_1	4.33	3.91	4.80	4.90	5.18	4.70	27.82
$\sum x$ Repeticije			125.18	118.80	118.74	126.10	120.34	128.87	$\sum x = 738.03$

U dvosmjernim tablicama koje slijede, osnovni podaci su svrstani tako da omogućuju procjenu pojedinih dijelova varijabilnosti.

	I	II	III	IV	V	VI	$\sum x_{\text{Sorte}}$
S_1	29.37	28.74	28.18	32.15	28.31	28.77	175.52
S_2	33.29	30.54	32.61	32.01	30.20	33.62	192.27
S_3	34.87	33.51	33.84	34.77	32.81	36.74	206.54
S_4	27.65	26.01	24.11	27.17	29.02	29.74	163.70
$\sum x_{\text{Repeticije}}$	125.18	118.80	118.74	126.10	120.34	128.87	$\sum x =$ 738.03

		S_1	S_2	S_3	S_4
	N_1	8.98	10.53	10.28	8.63
I	N_2	10.70	12.12	12.37	9.78
	N_3	9.69	10.64	12.22	9.24
	N_1	9.39	9.13	10.85	6.66
II	N_2	10.27	11.70	11.92	10.11
	N_3	9.08	9.71	10.74	9.24
	N_1	7.96	9.19	10.04	6.00
III	N_2	9.45	11.06	11.30	8.37
	N_3	10.77	12.36	12.50	9.74
	N_1	10.50	10.06	11.29	9.00
IV	N_2	10.69	10.58	11.06	8.73
	N_3	10.96	11.37	12.42	9.49
	N_1	8.29	8.90	10.55	8.58
V	N_2	9.46	10.19	9.97	10.09
	N_3	10.56	11.11	12.29	10.35
	N_1	9.03	10.63	12.31	10.07
VI	N_2	9.63	11.30	12.26	9.91
	N_3	10.11	11.69	12.17	9.76

	N_1	N_2	N_3	$\sum x_S$	$\bar{x}_S$
S_1	54.15	60.20	61.17	175.52	4.87
S_2	58.44	66.95	66.88	192.27	5.34
S_3	65.32	68.88	72.34	206.54	5.74
S_4	48.94	56.99	57.77	163.70	4.55
$\sum x_N$	226.85	253.02	258.16	$\sum x = 738.03$	
$\bar{x}_N$	4.73	5.27	5.38		

	C_0	C_1	$\sum x_S$
S_1	90.72	84.80	175.52
S_2	98.37	93.90	192.27
S_3	108.18	98.36	206.54
S_4	86.59	77.11	163.70
$\sum x_C$	383.86	354.17	$\sum x = 738.03$
$\bar{x}_C$	5.33	4.92	

	C_0	C_1	$\sum x_N$
N_1	118.69	108.16	226.85
N_2	131.53	121.49	253.02
N_3	133.64	124.52	258.16
$\sum x_C$	383.86	354.17	$\sum x = 738.03$

ANOVA

$$SS_{\text{Ukupno za cijeli pokus}} = 4.76^2 + \ldots + 4.70^2 - \frac{738.03^2}{144} = 85.1284$$

$$SS_{\text{Ukupno za glavni faktor S}} = \frac{29.37^2 + \ldots + 29.74^2}{3 \cdot 2} - \frac{738.03^2}{144} =$$

$$= 37.9810$$

$$SS_{\text{Repeticija}} = \frac{125.18^2 + \ldots + 128.87^2}{4 \cdot 3 \cdot 2} - \frac{738.03^2}{144} = 3.8201$$

$$SS_S = \frac{175.52^2 + ... + 163.70^2}{6 \cdot 3 \cdot 2} - \frac{738.03^2}{144} = 29.4281$$

$$SS_{\text{Pogreške (a)}} = 37.9810 - 3.8201 - 29.4281 = 4.7328$$

$$SS_{\text{Ukupno za podfaktor N}} = \frac{8.98^2 + ... + 9.76^2}{2} - \frac{738.03^2}{144} - $$
$$- SS_{\text{Repeticija}} - SS_S = 25.3253$$

$$SS_N = \frac{226.85^2 + 253.82^2 + 258.16^2}{6 \cdot 4 \cdot 2} - \frac{738.03^2}{144} = 11.7472$$

$$SS_{SxN} = \frac{54.15^2 + ... + 57.77^2}{6 \cdot 2} - \frac{738.03^2}{144} - 29.4281 - 11.7472 =$$
$$= 0.6913$$

$$SS_{\text{Pogreške (b)}} = 25.3253 - 11.7472 - 0.6913 = 12.8868$$

$$SS_{\text{Ukupno za podpodfaktor C}} = 85.1284 - 37.9810 - 25.3253 = 21.8219$$

$$SS_C = \frac{383.86^2 + 354.17^2}{6 \cdot 4 \cdot 3} - \frac{738.03^2}{144} = 6.1215$$

$$SS_{SxC} = \frac{90.72^2 + ... + 77.11^2}{6 \cdot 3} - \frac{738.03^2}{144} - 29.4281 - 6.1215 =$$
$$= 0.5821$$

$$SS_{NxC} = \frac{118.96^2 + ... + 124.52^2}{6 \cdot 4} - \frac{738.03^2}{144} - 11.7472 - 6.1215 =$$
$$= 0.0213$$

$$SS_{SxNxC} = SS_{\text{Svih kombinacija}} - SS_S - SS_N - SS_C - SS_{SxN} -$$
$$- SS_{SxC} - SS_{NxC} = \frac{27.89^2 + ... + 27.82^2}{6} - \frac{738.03^2}{144} - 29.4281 -$$
$$- 11.7472 - 0.6913 - 6.1215 - 0.5821 - 0.0213 = 0.5545$$

$$SS_{\text{Pogreške (c)}} = 21.8219 - 6.1215 - 0.5821 - 0.0213 - 0.5545 = 14.5425$$

Tablica ANOVA

Izvori varijabilnosti	$n-1$	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno za cijeli pokus	143	85.1284				
Ukupno za glavni faktor (S)	23	37.9810				
Repeticije	5	3.8201				
S	3	29.4281	9.8094	12.84 **	3.29	5.42
Pogreška (a)	15	4.7328	0.7640			
Ukupno za podfaktor (N)	48	25.3253				
N	2	11.7472	5.8736	18.23 **	3.23	5.18
SxN	6	0.6913	0.1152	0.36 ns	2.34	3.29
Pogreška (b)	40	12.8868	0.3222			
Ukupno za pod-podfaktor (C)	72	21.8219				
C	1	6.1215	6.1215	25.25 **	4.00	7.08
SxC	3	0.5821	0.1940	0.80 ns	2.76	4.13
NxC	2	0.0213	0.0107	0.04 ns	3.15	4.98
SxNxC	6	0.5545	0.0924	0.38 ns	2.25	3.12
Pogreška (c)	60	14.5425	0.2424			

Rezultati analize varijance dozvoljavaju zaključak:

- Prinos zrna značajno ovisi i o sorti i o gnojidbi i o primjeni CCC.
- Interakcije nisu opravdane, što znači da su djelovanja sva tri faktora neovisna jedno o drugom.

U daljnjoj analizi nikad ne smijemo zanemariti činjenicu da se ovom shemom najmanje precizno procjenjuje djelovanje glavnog faktora, nešto preciznije podfaktora, a najbolje podpodfaktora i interakcija s podpodfaktorom.

Dovoljno velik broj slobodnih varijanata pogrešaka a i b u ovom slučaju nam ipak dopušta provesti i t-testove za glavni faktor i podfaktor.

t-test za usporedbu prosječnih prinosa sorata

$$s_D = \sqrt{\frac{2 \cdot 0.764}{6 \cdot 3 \cdot 2}} = 0.206 \quad t_{tabl} \text{ očitan iz n-1 pogreške (a)}$$

$$LSD_{P=5\%} = 2.13 \cdot 0.206 = 0.44 \text{ kg/parceli}$$

$$LSD_{P=1\%} = 2.95 \cdot 0.206 = 0.61 \text{ kg/parceli}$$

Prinosi sorata Sana (S_3) i Zg 471 (S_2) se međusobno ne razlikuju značajno, a imaju signifikantno veći prinos zrna od ostale dvije sorte (S_1 - Super zlatne i S_4 - Sivke) koje su opet međusobno zanemarivo malo različitog prinosa.

5.74 (S_3) 5.34 (S_2) 4.87 (S_1) 4.55 (S_4)

0.4 ns

0.87**

1.19**

t-test za faktor N gnojidbu

$$s_D = \sqrt{\frac{2 \cdot 0.3222}{6 \cdot 4 \cdot 2}} = 0.1159$$

$LSD_{P=5\%} = 2.02 \cdot 0.1159 = 0.23$ kg/parceli t_{tabl} očitan iz n-1 pogreške (b)

$$LSD_{P=1\%} = 2.70 \cdot 0.1159 = 0.31 \text{ kg/parceli}$$

4.73 (N_1) 5.27 (N_2) 5.38 (N_3)

0.54** 0.11 ns

Gnojidbom s povećanom količinom dušika od 136 kg/ha (N_1) na 166 kg/ha (N_2) postignut je značajno veći prinos pšenice za 755 kg/ha (jer 0.54 kg/parceli 7.15 m^2 = 755.24 kg/ha). Daljnjim povećavanjem količine dodanog dušika prinos se, međutim, ne povećava značajno.

Za faktor CCC jer je zastupljen u dvije gradacije nije potrebno provoditi t-test. Signifikantan F-test navodi na zaključak da je primjena ovog regulatora rasta koji snižava stabljiku, uvjetovala signifikantno smanjenje prinosa (0.41 kg/parceli ili 573.42 kg/ha). Jer:

$$5.33\ (C_0) - 4.92\ (C_1) = 0.41^{**}$$

8
9
10
11

12 Pokusi ponovljeni u vremenu ili prostoru

Pokusi ponovljeni u vremenu ili prostoru (višegodišnji pokusi ili pokusi na više lokacija)

Kakva će biti vrijednost koju manifestira biljka, usjev ili nasad, ovisi i o genotipu i o okolini i o njihovoj interakciji. Genotip i neke faktore okoline (kao što su: količina gnojiva, gustoća usjeva, kontrola zaraze) stručnjak može kontrolirati. Ali ostali okolinski čimbenici kao: broj sunčanih sati, padaline, neka svojstva tla i sl. u principu su fiksni, teško ih je mijenjati i gotovo nemoguće kontrolirali, a jednako su važni ako ne i važniji od onih prvih. Bitno je moći procijeniti njihove utjecaje. Kako je za očekivati da se nekontrolirani faktori mijenjaju s vremenom i od lokacije do lokacije i da su te promjene mjerljive, to njihov utjecaj na biljku može biti vrednovan. Zato istraživač mora ponavljati eksperiment, da bi mogao procijeniti i one utjecaje koji su izvan njegove kontrole.

U bilinogojstvu se najčešće koristi i prostorno i vremensko ponavljanje eksperimenta, tj. na više lokacija, ili tijekom nekoliko godina, ili oboje. Za to se tako ponovljeni pokusi zovu i **serija pokusa**.

Svaka serija predstavlja jedinstveni problem kako za stručnjaka (istraživača i praktičara), tako isto i za biometričara i to podjednako kod planiranja i u analizi podataka takvih pokusa.

Kako analizirati podatke serije pokusa?

Pretpostavke koje moraju biti zadovoljene za zajedničku kombiniranu analizu su:

- Svi pokusi moraju biti iste veličine i strukture.
- Lokacije ili godine trebale bi predstavljati slučajan uzorak niza mogućih. Mnogo je, međutim, razloga zbog kojih se u praksi rijetko može ostvariti slučajnost lokacija, a pogotovo godina (2 ili 3 godine teško se može smatrati slučajnima u nekom razdoblju).
- Varijance pogrešaka serije pokusa moraju biti homogene.

Homogenost varijanci pogreške u vrlo praktičnom i jednostavnom smislu testira se F-testom na uobičajeni način.

Ako se radi o dvije lokacije ili godine

$$F = \frac{veća\ s^2\ pogreške}{manja\ s^2\ pogreške}$$

Ako je pokus ponovljen na više lokacija ili tijekom više godina, također se može primijeniti F-test s time da se F_{exp} izračuna kao omjer najveće i najmanje varijance pogreške pokusa

$$F = \frac{najveća\ s^2\ pogreške}{najmanja\ s^2\ pogreške}$$

pri čemu se F_{tabl} očita iz broja lokacija (n_L) i broja slobodnih varijanata varijance pogreške na lokaciji.

Još jedan praktičan način na koji se može ocijeniti mogu li se podaci serije pokusa zajednički analizirati, je računanje varijacijskih koeficijenata svakog pokusa (svake lokacije ili svake godine), na osnovi varijance pogreške i prosječne vrijednosti pokusa

$$cv = \frac{100\sqrt{s^2 pogreške}}{\bar{x}}$$

Lokacija, čiji je cv < 20% može se uključiti u zajedničku analizu, a ona lokacija na kojoj je cv > 20% izbacuje se - ne ulazi u kombiniranu analizu.

Ovo su praktični i brzi načini testiranja homogenosti varijanci pogreške.

Međutim, precizan i osjetljiv test homogenosti varijanci je Bartlett-ov test, baziran na prirodnom logaritmu varijance pogreške (Cochran & Cox, Snedecor & Cochran i drugi).

Analiza serije pokusa provodila bi se, dakle, na ovaj način:

1. ANOVA za svaki pokus (lokaciju, godinu)

Izvori varijabilnosti	n - 1
Ukupno	$n_T \cdot n_r - 1$
Repeticije	$n_r - 1$
Tretiranja	$n_T - 1$
Pogreška	$(n_T - 1)(n_r - 1)$

2 Testirati homogenost varijanci pogrešaka pokusa na *lokacijama ili godinama, pa ako su homogene ići u zajedničku* kombiniranu analizu.

3 Zajednička ANOVA se provodi na uobičajeni način, samo treba naglasiti da se u takvoj analizi $SS_{Repeticija}$ dobije kao zbroj $SS_{Repeticija}$ sa svih pokusa. Jednako tako je u kombiniranoj analizi zajednička $SS_{Pogreške}$ suma $SS_{Pogrešaka}$ svih pokusa, a zajednička s^2 pogreške je zapravo prosječna pogreška svih pokusa.

Tablica kombinirane ANOVA

Izvori varijabilnosti	$n - 1$
Ukupno	$n_L \cdot n_T \cdot n_r - 1$
Repeticije u lokacijama	$n_L \cdot (n_r - 1)$
Lokacije	$n_L - 1$
Tretiranja	$n_T - 1$
L x T	$(n_L - 1)(n_T - 1)$
Pogreška	$n_L (n_T - 1)(n_r - 1)$

Ako varijance pogreške nisu homogene, zajednička analiza je moguća samo iz prosječnih vrijednosti tretiranja u pojedinim pokusima, ali tada se gubi mogućnost otkrivanja interakcije, jer je ona u tom slučaju dio varijabilnosti s kojom se testiraju efekti tretiranja i lokacije (ili godine). Naime,

Izvori varijabilnosti	$n - 1$
Ukupno	$n_L \cdot n_T - 1$
Lokacije	$n_L - 1$
Tretiranja	$n_T - 1$
L x T	$(n_L - 1)(n_T - 1)$

Treba uočiti da kombinirana analiza serije pokusa uključuje dodatni faktor ili faktore (lokaciju ili godinu, ili oboje), što na neki način komplicira problem posebno kod serije višefaktorijalnih pokusa.

Naime, efekti nekih faktora će biti podjednaki od lokacije do lokacije (ili od godine do godine) dok će se efekti drugih faktora mijenjati. Isto vrijedi i za njihove interakcije. Stoga je kod analize serije pokusa potreban poseban oprez, jer propusti i pogreške mogu biti razlogom potpuno neispravne interpretacije tako kompleksne analize.

Cjelovitu analizu ćemo pokazati na jednom najjednostavnijem primjeru.

PRIMJER 12.1.

Sortni pokus sa 6 novostvorenih sorata i jednom standardnom sortom ozime uljane repice izveden je po shemi slučajnog bloknog rasporeda sa 5 ponavljanja u Osijeku 1997/98 i 1998/99. Na podacima o prinosu sjemena u dt/ha provesti ćemo cjelovitu analizu ovog pokusa ponovljenog tijekom dvije godine.

Podaci o prinosu sjemena (dt/ha) po sortama, repeticijama i godinama:

		S_1	S_2	S_3	S_4	S_5	S_6	S_7	$\sum x$ Repeticije
1997/98	I	24.43	31.84	25.95	35.97	18.60	15.33	20.20	172.32
	II	22.07	35.76	31.68	35.15	21.96	15.72	22.62	184.96
	III	18.92	26.54	21.28	26.00	25.95	18.87	24.64	162.20
	IV	28.14	35.11	27.42	27.38	24.59	20.83	21.00	184.47
	V	18.52	40.83	28.65	35.97	26.51	16.90	24.48	191.86
$\sum x_S$		112.08	170.08	134.98	160.47	117.61	87.65	112.94	$\sum x_{97/98} = 895.81$
1998/99	I	21.76	24.96	25.92	32.88	25.60	27.20	29.12	187.44
	II	29.60	25.68	28.00	34.56	25.52	28.64	25.60	197.60
	III	21.92	31.76	29.76	32.56	33.52	26.64	22.72	198.88
	IV	25.76	26.32	25.68	38.16	34.24	32.16	22.16	204.48
	V	28.96	34.48	31.76	38.88	33.52	32.48	29.76	229.84
$\sum x_{Sorte}$		128.00	143.20	141.12	177.04	152.40	147.12	129.36	$\sum x_{97/98} = 1018.24$
$\sum x_{Sorte}$		240.08	313.28	276.10	337.51	270.01	234.77	242.30	$\sum x = 1914.05$

ANOVA za svaku pojedinu godinu:

1997/98

$$SS_{\text{Ukupno}} = 24.43^2 + ... + 24.48^2 - \frac{895.81^2}{35} = 1430.86$$

$$SS_{\text{Repeticija}} = \frac{172.32^2 + ... + 191.86^2}{7} - \frac{895.81^2}{35} = 79.65$$

$$SS_{\text{Sorata}} = \frac{112.08^2 + ... + 112.94^2}{5} - \frac{895.81^2}{35} = 1018.01$$

$$SS_{\text{Pogreške}} = 1430.86 - 79.65 - 1018.01 = 333.20$$

1998/99

$$SS_{\text{Ukupno}} = 21.76^2 + ... + 29.76^2 - \frac{1018.24^2}{35} = 676.54$$

$$SS_{\text{Repeticija}} = \frac{187.44^2 + ... + 229.84^2}{7} - \frac{1018.24^2}{35} = 144.10$$

$$SS_{\text{Sorata}} = \frac{128.00^2 + ... + 129.36^2}{5} - \frac{1018.24^2}{35} = 327.24$$

$$SS_{\text{Pogreške}} = 676.54 - 144.10 - 327.24 = 205.20$$

Tablica ANOVA

		1997/1998			1998/1999			$F_{tabl.}$	
Izvori varijabilnosti	$n - 1$	SS	s^2	$F_{exp.}$	SS	s^2	$F_{exp.}$	P=5%	P=1%
Ukupno	34	1430.86			676.54				
Repeticije	4	79.65			144.10				
Sorte	6	1018.01	169.67	12.22**	327.24	54.54	6.388**	2.51	3.67
Pogreška	24	333.20	13.88		205.20	8.55			

F-testom ćemo testirati homogenost varijanci pogreške, kako bismo ustanovili možemo li provesti zajedničku analizu za obje godine.

$$F_{exp} = \frac{13.88}{8.55} = 1.62$$

F_{tabl} očitan iz slobodnih varijanata ovih varijanci (24) je $F_{P=5\%} = 1.98$, $F_{P=1\%} = 2.66$.

$F_{exp} < F_{tabl}$ što znači da su varijance homogene, pa je moguće ići u zajedničku analizu za obje godine. Isto potvrđuju i varijacijski koeficijenti koji su u obje godine manji od 20%.

Iz $s^2_{pogreške\,97/98}$ (13.88), $s^2_{pogreške\,98/99}$ (8.55) prosječnih vrijednosti $\bar{x}_{97/98} = \frac{895.81}{35} = 25.59$ i $\bar{x}_{98/99} = \frac{1018.24}{35} = 29.09$ varijacijski koeficijenti su:

$$cv_{97/98} = \frac{100}{25.59} = 10.05\%$$

$$cv_{98/99} = \frac{100}{29.09} = 14.56\%$$

Kombinirana ANOVA

$$SS_{\text{Ukupno}} = 24.43^2 + ... + 29.76^2 - \frac{1914.05^2}{70} = 2321.53$$

$$SS_{\text{Repeticija}} = SS_{\text{repeticija 1997/98}} + SS_{\text{repeticija 1998/99}} = 223.75$$

$$SS_{\text{Sorata}} = \frac{240.08^2 + ... + 242.30^2}{2 \cdot 5} - \frac{1914.05^2}{70} = 928.90$$

$$SS_{\text{Godina}} = \frac{895.81^2 + 1018.24^2}{7 \cdot 5} - \frac{1914.05^2}{70} = 214.13$$

$$SS_{\text{SxG}} = \frac{112.08^2 + ... + 129.36^2}{5} - \frac{1914.05^2}{70} - 928.90 - 214.13 =$$

$$= 416.36$$

$$SS_{\text{Pogreške}} = 2321.53 - 223.75 - 928.90 - 214.13 - 416.36 = 538.39$$

(a to je zapravo suma $SS_{\text{Pogreške '98}} + SS_{\text{Pogreške '99}}$)

Tablica kombinirane ANOVA

Izvori varijabilnosti	$n - 1$	SS	s^2	$F_{exp.}$	$F_{tabl.}$ P=5%	$F_{tabl.}$ P=1%
Ukupno	69	2321.53				
Repeticije u godinama	8	223.75				
Godine	1	214.13	214.13			
Sorte	6	928.90	154.82			
SxG	6	416.36	69.39	6.18**	2.29	3.18
Pogreška	48	538.39	11.22			

U ovakvoj kombiniranoj analizi najprije se testira interakcija. Ako je interakcija signifikantna (kakav je slučaj u ovom primjeru) to znači da se tretiranja različito ponašaju od godine do godine (ili od lokacije do lokacije). Zato testiranje tretiranja ne bi imalo smisla.

Ako se radi o pokusu ponovljenom na više lokacija, nakon opravdanog F-testa interakcije ipak se mogu potražiti izvjesne interesantne lokacije za koje se donose određene preporuke.

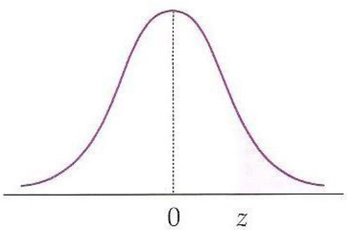
U slučaju nesignifikantne interakcije, zbroje se varijanca interakcije i varijanca pogreške i ta suma služi za testiranje tretiranja.

Više i detaljnije o pokusima ponovljenim na više lokacija i godina može se naći u Cochran & Cox, Petersen, Gomez & Gomez, Snedecor & Cochran.

■

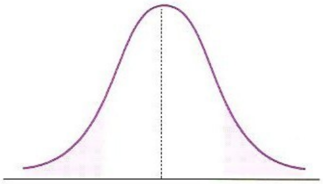
TABLICE

Tablica A	Površina ispod normalne krivulje
Tablica B	Granične vrijednosti za t
Tablica C1	Granične vrijednosti za F (P=0.05)
Tablica C2	Granične vrijednosti za F (P=0.01)



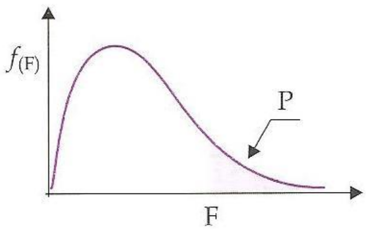
Tablica A: Površina ispod normalne krivulje

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	*.3821*	*.3783*	*.3745*	*.3707*	*.3669*	*.3632*	*.3594*	*.3557*	*.3520*	*.3483*
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	*.0019*	*.0018*	*.0018*	*.0017*	*.0016*	*.0016*	*.0015*	*.0015*	*.0014*	*.0014*
3.0	.0014	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.9	.0000									



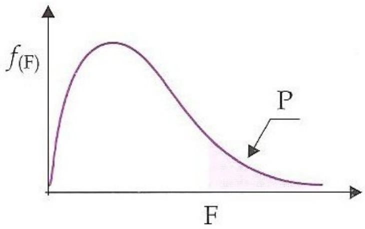
Tablica B: Granične vrijednosti za t

n-1	P			
	0.10	0.05	0.01	0.001
1	6.31	12.71	63.66	632.60
2	2.92	4.30	9.92	31.60
3	2.35	3.18	5.84	12.92
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.90	2.36	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
11	1.80	2.20	3.11	4.44
12	1.78	2.18	3.06	4.32
13	1.77	2.16	3.01	4.22
14	1.76	2.14	2.98	4.14
15	1.75	2.13	2.95	4.07
16	1.74	2.12	2.92	4.02
17	1.74	2.11	2.90	3.96
18	1.73	2.10	2.88	3.92
19	1.73	2.09	2.86	3.88
20	1.72	2.09	2.84	3.85
21	1.72	2.08	2.83	3.82
22	1.72	2.07	2.82	3.79
23	1.71	2.07	2.81	3.77
24	1.71	2.06	2.80	3.74
25	1.71	2.06	2.79	3.72
26	1.71	2.06	2.78	3.71
27	1.70	2.05	2.77	3.69
28	1.70	2.05	2.76	3.67
29	1.70	2.05	2.76	3.66
30	1.70	2.04	2.75	3.65
40	1.68	2.02	2.70	3.55
50	1.68	2.01	2.68	3.50
60	1.67	2.00	2.66	3.46
100	1.66	1.98	2.63	3.39
∞	1.64	1.96	2.58	3.29



Tablica C1: Granične vrijednosti za F (P=0.05)

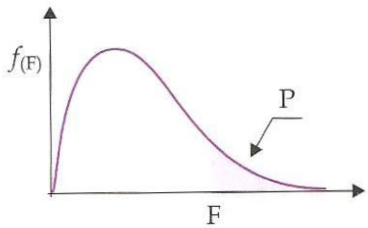
n-1 nazivnika	*n*-1 brojnika						
	1	2	3	4	5	6	7
1	161.00	200.00	216.00	225.00	230.00	234.00	237.00
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01
12	4.75	3.88	3.49	3.26	3.11	3.00	2.91
13	4.67	3.80	3.41	3.18	3.02	2.92	2.83
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54
20	4.35	3.40	3.10	2.87	2.71	2.60	2.51
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42
25	4.24	3.38	2.99	2.76	2.60	2.49	2.40
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39
27	4.21	3.35	2.96	2.37	2.57	2.46	2.37
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05
∞	3.84	3.00	2.60	2.37	2.21	2.09	2.01



Tablica C1: Granične vrijednosti za F (P=0.05)

n-1 nazivnika	n-1 brojnika						
	8	9	10	12	16	20	24
1	239.00	241.00	242.00	244.00	246.00	248.00	249.00
2	19.37	19.38	19.40	19.41	19.43	19.44	19.45
3	8.84	8.81	8.78	8.74	8.69	8.66	8.64
4	6.04	6.00	5.96	5.91	5.84	5.80	5.77
5	4.82	4.77	4.74	4.68	4.60	4.56	4.53
6	4.15	4.10	4.00	4.06	3.92	3.87	3.84
7	3.73	3.68	3.64	3.57	3.49	3.44	3.41
8	3.44	3.39	3.53	3.28	3.20	3.15	3.12
9	3.23	3.18	3.14	3.07	2.98	2.94	2.90
10	3.07	3.02	2.98	2.91	2.82	2.77	2.74
11	2.95	2.90	2.85	2.79	2.70	2.65	2.61
12	2.85	2.80	2.75	2.69	2.60	2.54	2.50
13	2.77	2.71	2.67	2.60	2.51	2.46	2.42
14	2.70	2.65	2.60	2.53	2.44	2.39	2.35
15	2.64	2.59	2.54	2.48	2.39	2.33	2.29
16	2.59	2.54	2.49	2.42	2.33	2.28	2.24
17	2.55	2.49	2.45	2.38	2.29	2.23	2.19
18	2.51	2.46	2.41	2.34	2.25	2.19	2.15
19	2.48	2.42	2.38	2.31	2.21	2.16	2.11
20	2.45	2.39	2.35	2.28	2.18	2.12	2.08
21	2.42	2.37	2.32	2.25	2.15	2.10	2.05
22	2.40	2.34	2.30	2.23	2.13	2.07	2.03
23	2.37	2.32	2.27	2.20	2.10	2.05	2.00
24	2.36	2.30	2.25	2.18	2.09	2.03	1.98
25	2.34	2.28	2.24	2.16	2.06	2.01	1.96
26	2.32	2.27	2.22	2.15	2.05	1.99	1.95
27	2.30	2.25	2.20	2.13	2.03	1.97	1.93
28	2.29	2.24	2.19	2.12	2.02	1.96	1.91
29	2.28	2.22	2.18	2.10	2.00	1.94	1.90
30	2.27	2.21	2.16	2.09	1.99	1.93	1.89
40	2.18	2.12	2.08	2.00	1.90	1.84	1.79
50	2.13	2.07	2.02	1.95	1.85	1.78	1.74
60	2.10	2.04	1.99	1.92	1.81	1.75	1.70
70	2.07	2.01	1.97	1.89	1.79	1.72	1.67
80	2.05	1.99	1.95	1.88	1.77	1.70	1.65
100	2.03	1.97	1.92	1.85	1.75	1.68	1.63
150	2.00	1.94	1.89	1.82	1.71	1.64	1.59
200	1.98	1.92	1.87	1.80	1.69	1.62	1.57
∞	1.94	1.88	1.83	1.75	1.64	1.57	1.52

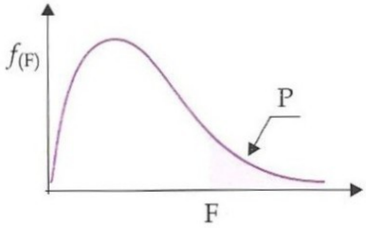
... nastavak



Tablica C1: Granične vrijednosti za F (P=0.05)

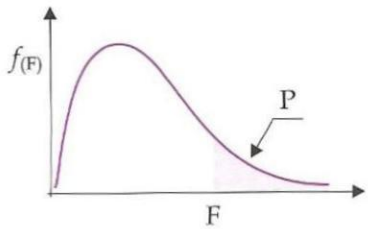
n-1 nazivnika	*n*-1 brojnika						
	30	40	50	60	75	100	∞
1	250.00	251.00	252.00	252.00	253.00	253.00	254.00
2	19.46	19.47	19.47	19.48	19.48	19.49	19.50
3	8.62	8.59	8.58	8.57	8.57	8.56	8.53
4	5.74	5.72	5.70	5.69	5.68	5.66	5.63
5	4.50	4.46	4.44	4.43	4.42	4.40	4.36
6	3.81	3.77	3.75	3.74	3.72	3.71	3.67
7	3.38	3.34	3.32	3.30	3.29	3.28	3.23
8	3.08	3.04	3.03	3.00	3.00	2.98	2.93
9	2.86	2.82	2.80	2.79	2.77	2.76	2.71
10	2.70	2.66	2.64	2.62	2.61	2.59	2.54
11	2.57	2.53	2.50	2.49	2.47	2.45	2.40
12	2.47	2.42	2.40	2.38	2.36	2.36	2.30
13	2.38	2.34	2.32	2.30	2.28	2.26	2.21
14	2.31	2.27	2.24	2.22	2.21	2.19	2.13
15	2.25	2.20	2.18	2.16	2.15	2.12	2.07
16	2.19	2.15	2.13	2.10	2.09	2.07	2.01
17	2.15	2.10	2.08	2.06	2.04	2.02	1.96
18	2.11	2.06	2.04	2.02	2.00	1.98	1.92
19	2.07	2.03	2.00	1.98	1.96	1.94	1.88
20	2.04	1.99	1.96	1.95	1.92	1.90	1.84
21	2.01	1.96	1.93	1.92	1.89	1.87	1.81
22	1.98	1.94	1.91	1.89	1.87	1.84	1.78
23	1.96	1.91	1.88	1.86	1.84	1.82	1.76
24	1.94	1.89	1.86	1.84	1.82	1.80	1.73
25	1.92	1.87	1.84	1.82	1.80	1.77	1.71
26	1.90	1.85	1.82	1.80	1.78	1.76	1.69
27	1.88	1.84	1.80	1.78	1.76	1.74	1.67
28	1.87	1.82	1.78	1.77	1.75	1.72	1.65
29	1.85	1.80	1.77	1.75	1.73	1.71	1.64
30	1.84	1.79	1.76	1.74	1.72	1.69	1.62
40	1.74	1.69	1.66	1.64	1.61	1.59	1.51
50	1.69	1.63	1.60	1.58	1.55	1.52	1.44
60	1.65	1.59	1.56	1.53	1.50	1.48	1.39
70	1.62	1.56	1.53	1.50	1.47	1.45	1.35
80	1.60	1.54	1.51	1.48	1.45	1.42	1.32
100	1.57	1.51	1.48	1.45	1.42	1.39	1.28
150	1.54	1.47	1.44	1.41	1.37	1.34	1.22
200	1.52	1.45	1.42	1.39	1.35	1.32	1.19
∞	1.46	1.39	1.35	1.32	1.28	1.24	1.00

... nastavak



Tablica C2: Granične vrijednosti za F (P=0.01)

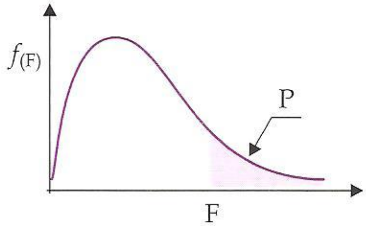
n-1 nazivnika	n-1 brojnika						
	1	2	3	4	5	6	7
1	4052.00	4999.00	5403.00	5625.00	7564.00	5859.00	5928.00
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20
11	9.65	7.20	6.22	5.67	5.32	5.07	4.89
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64
13	9.07	6.70	5.74	5.20	4.86	4.62	4.44
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93
18	8.28	6.01	5.09	4.58	4.25	4.01	3.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.76
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64
22	7.94	5.72	4.82	4.31	3.99	3.76	3.59
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12
50	7.17	5.06	4.20	3.72	3.41	3.18	3.02
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95
70	7.01	4.92	4.08	3.60	3.29	3.07	2.91
80	6.96	4.88	4.04	3.56	3.25	3.04	2.87
100	6.90	4.82	3.98	3.51	3.20	2.99	2.82
150	6.81	4.75	3.91	3.44	3.14	2.92	2.76
200	6.76	4.71	3.88	3.41	3.11	2.90	2.73
∞	6.63	4.60	3.78	3.32	3.02	2.80	2.64



Tablica C2: Granične vrijednosti za F (P=0.01)

n-1 nazivnika	n-1 brojnika						
	8	9	10	12	16	20	24
1	5982.00	6062.00	6056.00	6106.00	6169.00	6209.00	6235.00
2	99.36	99.39	99.40	99.42	99.44	99.45	99.46
3	27.49	27.34	27.23	27.05	26.83	26.69	26.60
4	14.80	14.66	14.55	14.37	14.15	14.02	13.93
5	10.29	10.16	10.05	9.89	9.68	9.55	9.47
6	8.10	7.98	7.87	7.72	7.52	7.40	7.31
7	6.84	6.72	6.62	6.47	6.27	6.16	6.07
8	6.03	5.91	5.81	5.67	5.48	5.36	5.28
9	5.47	5.35	5.26	5.11	4.92	4.81	4.73
10	5.06	4.94	4.85	4.71	4.52	4.40	4.33
11	4.74	4.63	4.54	4.40	4.21	4.10	4.02
12	4.50	4.39	4.30	4.16	3.98	5.86	3.78
13	4.30	4.19	4.10	3.96	3.78	3.66	3.59
14	4.14	4.03	3.94	3.80	3.62	3.50	3.43
15	4.00	3.89	3.80	3.67	3.48	3.37	3.29
16	3.89	3.78	3.69	3.55	3.37	3.26	3.18
17	3.79	3.68	3.59	3.45	3.27	3.16	3.08
18	3.71	3.60	3.51	3.37	3.19	3.08	3.00
19	3.63	3.52	3.43	3.30	3.12	3.00	2.92
20	3.56	3.46	3.37	3.23	3.05	2.94	2.86
21	3.51	3.40	3.31	3.17	2.99	2.88	2.80
22	3.45	3.36	3.26	3.12	2.94	2.83	2.75
23	3.41	3.30	3.21	3.07	2.89	2.78	2.70
24	3.36	3.26	3.17	3.03	2.85	2.74	2.66
25	3.32	3.22	3.13	2.99	2.81	2.70	2.62
26	3.29	3.18	3.09	2.96	2.77	2.66	2.58
27	3.26	3.15	3.06	2.93	2.74	2.63	2.55
28	3.23	3.12	3.03	2.90	2.71	2.60	2.52
29	3.20	3.09	3.00	2.87	2.68	2.57	2.49
30	3.17	3.07	2.98	2.84	2.66	2.55	2.47
40	2.99	2.89	2.80	2.66	2.49	2.37	2.29
50	2.88	2.78	2.70	2.56	2.39	2.26	2.18
60	2.82	2.72	2.63	2.50	2.32	2.20	2.12
70	2.77	2.67	2.59	2.45	2.28	2.15	2.07
80	2.74	2.64	2.55	2.41	2.24	2.11	2.03
100	2.69	2.59	2.51	2.36	2.19	2.06	1.98
150	2.62	2.53	2.44	2.30	2.12	2.00	1.91
200	2.60	2.50	2.41	2.28	2.09	1.97	1.88
∞	2.51	2.41	2.32	2.18	1.99	1.88	1.79

... nastavak



Tablica C2: Granične vrijednosti za F (P=0.01)

n-1 nazivnika	n-1 brojnika						
	30	40	50	60	75	100	∞
1	6261.00	6287.00	6302.00	6313.00	6323.00	6334.00	6366.00
2	99.47	99.47	99.48	99.48	99.49	99.49	99.50
3	26.50	26.41	26.35	26.32	26.27	26.23	26.12
4	13.84	13.74	13.69	13.65	13.61	13.57	13.46
5	9.38	9.29	9.24	9.20	9.17	9.13	9.02
6	7.23	7.14	7.09	7.06	7.02	6.99	6.88
7	5.99	5.91	5.85	5.82	5.78	5.75	5.65
8	5.20	5.12	5.06	5.03	5.00	4.96	4.86
9	4.65	4.57	4.51	4.48	4.45	4.41	4.31
10	4.25	4.16	4.12	4.08	4.05	4.01	3.91
11	3.94	3.86	3.80	3.78	3.74	3.70	3.60
12	3.70	3.62	3.56	3.54	3.49	3.46	3.36
13	3.51	3.42	3.37	3.34	3.30	3.27	3.16
14	3.35	3.27	3.21	3.18	3.14	3.11	3.00
15	3.21	3.13	3.07	3.05	3.00	2.97	2.87
16	3.10	3.02	2.96	2.93	2.89	2.86	2.75
17	3.00	2.92	2.86	2.83	2.79	2.76	2.65
18	2.91	2.84	2.78	2.75	2.71	2.68	2.57
19	2.84	2.76	2.70	2.67	2.63	2.60	2.49
20	2.78	2.69	2.63	2.61	2.56	2.53	2.42
21	2.72	2.64	2.58	2.55	2.51	2.47	2.36
22	2.67	2.58	2.53	2.50	2.46	2.42	2.31
23	2.62	2.54	2.48	2.45	2.41	2.37	2.26
24	2.58	2.49	2.44	2.40	2.36	2.33	2.21
25	2.54	2.45	2.40	2.36	2.32	2.29	2.17
26	2.50	2.42	2.36	2.33	2.28	2.25	2.13
27	2.47	2.38	2.33	2.29	2.25	2.21	2.10
28	2.44	2.35	2.30	2.26	2.22	2.18	2.06
29	2.41	2.32	2.27	2.23	2.19	2.15	2.03
30	2.39	2.30	2.24	2.21	2.16	2.13	2.01
40	2.20	2.11	2.05	2.02	1.97	1.94	1.80
50	2.10	2.00	1.94	1.91	1.86	1.82	1.68
60	2.03	1.94	1.87	1.84	1.79	1.74	1.60
70	1.98	1.88	1.82	1.79	1.74	1.69	1.53
80	1.94	1.84	1.78	1.74	1.70	1.65	1.49
100	1.89	1.79	1.73	1.69	1.64	1.59	1.43
150	1.83	1.72	1.66	1.62	1.56	1.51	1.33
200	1.79	1.69	1.62	1.58	1.53	1.48	1.28
∞	1.70	1.59	1.52	1.47	1.41	1.36	1.00

... nastavak