OPISNA STATISTIKA

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

Andrijana Rebekić

Sveučilište Josipa Jurja Strossmayera u Osijeku
Poljoprivredni fakultet u Osijeku

Opisna statistika

Autorica
doc. dr. sc. Andrijana Rebekić

Recenzenti
doc. dr. sc. Domagoj Šimić
Poljoprivredni institut Osijek

prof. dr. sc. Zoran Škrtić
Poljoprivredni fakultet Sveučilišta u Osijeku

Lektorica
Isidora Vujošević, prof.

Izdavač
Poljoprivredni fakultet
Sveučilište Josipa Jurja Strossmayera u Osijeku
Vladimira Preloga 1, HR-31000 Osijek, Hrvatska

ISBN 978-953-7871-69-7

Osijek, 2017.

It is the mark of a truly intelligent person to be moved by statistics.

(George Bernard Shaw)

Predgovor

Ovaj priručnik namijenjen je prvenstveno studentima diplomskih i poslijediplomskih studija koji slušaju module „Biometrika“ ili „Ekonometrija“ na Poljoprivrednome fakultetu u Osijeku.

U priručniku su obrađene nastavne cjeline koje su sastavni dio vježbi iz navedenih modula. Svaka nastavna cjelina obrađena je kroz:

i. kratak teorijski uvod u nastavnu cjelinu
ii. primjere zadataka s prikazanim postupkom rješavanja
iii. zadatke za vježbu.

Na temelju dosadašnjeg iskustva u radu sa studentima, uočila sam da velik dio vremena na vježbama odlazi na zapisivanje i prepisivanje zadataka te je svrha ovog priručnika olakšati studentima praćenje vježbi.

Dio priručnika u kojemu su zadaci za vježbu koncipiran je po uzoru na radnu bilježnicu. Uz primjer zadataka za rješavanje ostavljen je prazan prostor za njegovo rješavanje. To omogućava studentima da po završetku vježbi imaju na istome mjestu teoriju, postupak rješavanja i primjere za vježbu, što bi trebalo olakšati savladavanje i usvajanje gradiva.

Osim toga, priručnik *Opisna statistika* može poslužiti i studentima drugih studija za savladavanje osnovnih metoda i principa opisne statistike.

Andrijana Rebekić

Sadržaj

1. Podjela statističkih metoda i vrste obilježja

I. Podjela statističkih metoda

Statistika je znanstvena disciplina koja se bavi prikupljanjem, uređivanjem i obradom podataka te tumačenjem dobivenih rezultata. S obzirom na tematiku kojom se bavi, statistika se dijeli na deskriptivnu i inferencijalnu statistiku.

Deskriptivna statistika (opisna statistika) obuhvaća postupke kojima se prikupljeni podaci uređuju, a ispitivana populacija (N) opisuje se na temelju svih vrijednosti u osnovnom skupu i to pomoću numeričkih i grafičkih modela. Prema tome, deskriptivna statistika podrazumijeva:

- grupiranje i grafičko prikazivanje podataka
- izračunavanje brojčanih pokazatelja koji izražavaju karakteristike promatrane pojave (mjere centralne tendencije, mjere varijabilnosti itd.) u cilju opisivanja populacije (N)
- dobiveni rezultati odnose se samo na analizirani skup podataka i ne poopćavaju se.

Inferencijalna statistika (analitička statistika) obuhvaća postupke kojima se na temelju rezultata iz uzorka – podskupa populacije (n) – donose zaključci o populaciji s određenom stupnjem sigurnosti. Metode inferencijalne statistike temelje se na teoriji vjerojatnosti.

Osim deskriptivne i inferencijalne statistike, može se izdvojiti dio statistike koji se bavi osmišljavanjem i planiranjem pokusa te načinima prikupljanja podataka nazvanim **eksperimentalni dizajn.**

II. Vrste obilježja

Varijable su **obilježja** koja promatramo na **jedinicama promatranja.**

Na primjer, jedinica promatranja može biti biljka pšenice, a obilježja odnosno varijable koje promatramo: visina biljke, masa klasa, masa zrna po klasu, broj zrna po klasu, urod, hektolitarska masa zrna itd.

U tom slučaju, jedinica promatranja (jedna biljka pšenice) pripada **statističkom skupu** (populaciji), koji mora biti ujednačen. U populaciji (statističkom skupu) jedinice promatranja mogu se razlikovati samo prema obilježju promatranja, odnosno ispitivanoj varijabli.

S obzirom na svojstvo koje prikazuju, statistička obilježja (varijable) najčešće se dijele na kvalitativne (kategorijske) i kvantitativne (numeričke) varijable.

i. **Kvalitativne (kategorijske) varijable** najčešće se izražavaju pojmovno, a unutar skupine kategorijskih varijabli razlikujemo dvije **mjerne skale**:

- <u>Nominalna mjerna skala</u>
  - ne postoji skala vrijednosti
  - nema boljeg ili lošijeg modusa
  - vrijednosti su iskazane pojmovno
  - pojmovno iskazanim vrijednostima često se, radi olakšavanja obrade podataka, pridružuju brojevi.

  Primjeri varijabli izraženih pomoću nominalne mjerne skale:
  - krvne grupe (0, A, B, AB)
  - vrste žitarica (pšenica, ječam, zob)
  - boja dlake (crna, smeđa, bijela, siva, crvena)
  - gravidnost (da, ne)
  - zdravstveni status (zdrav, bolestan).

- <u>Ordinalna mjerna skala</u>
  - postoji skala kvalitete, ali precizno kvantificiranje nije moguće
  - najčešće se radi o nekom obliku ocjenjivanja
  - vrijednosti mogu biti iskazane pojmovno ili brojčano.

  Primjeri varijabli izraženih pomoću ordinalne mjerne skale:
  - rezultat na testu (loš, dobar, izvrstan)
  - plodnost (niska, srednja, visoka)
  - ocjene od 1 do 5.

ii. **Kvantitativne (numeričke) varijable** izražavaju se brojevima pogodnim za različite matematičko-statističke postupke. S obzirom na to jesu li razlike između brojeva konstantne ili nekonstantne, razlikujemo omjernu i intervalnu mjernu skalu.

- Intervalna mjerna skala
  - vrijednosti se označavaju brojčano
  - jednake razlike brojeva na skali predstavljaju jednake razlike mjerenog svojstva
  - temperaturne skale primjer su intervalne skale (0 °C znači da je izmjerena temperatura 0, a ne da temperatura nije izmjerena ili je nije bilo).

- Omjerna mjerna skala
  - vrijednosti se označavaju brojčano te se nazivaju vrijednostima numeričke varijable
  - jednake razlike brojeva na skali predstavljaju jednake razlike mjerenog svojstva
  - nula (0) predstavlja nepostojanje mjerenog svojstva.

S obzirom na broj vrijednosti koji numeričke varijable mogu poprimiti, dijele se na diskontinuirane i kontinuirane numeričke varijable.

- Diskontinuirana (diskretna) numerička varijabla – iskazuje se cijelim brojem i može prikazati prebrojivo mnogo vrijednosti, npr.:

  - broj potomaka u leglu
  - broj snesenih jaja
  - broj studenata po godini
  - broj zrna po klasu.

- Kontinuirana numerička varijabla – iskazuje se teoretski beskonačnim brojevima što ovisi o mogućnostima mjerenja, npr.:

  - količina pomuženog mlijeka (l)
  - tjelesna masa (kg, g)
  - visina (cm, m)
  - koncentracija metabolita u krvi i urinu (mg kg-1)
  - koncentracija elementa u tkivu (μg kg-1)
  - urod (t/ha)
  - masa klasa (g).

RIJEŠI ZADATKE I ODGOVORI NA PITANJA!

1. Zaokruži točne odgovore!

a)	populacija predstavlja skup svih mogućih jedinica promatranja	Da	Ne
b)	uzorak je podskup populacije	Da	Ne
c)	varijabla je promatrano obilježje čije se vrijednosti izražavaju isključivo numerički	Da	Ne
d)	kvalitativne varijable se uglavnom izražavaju pojmovno	Da	Ne
e)	statističke metode kojima se na temelju rezultata iz uzorka zaključuje o populaciji pripadaju skupini inferencijalne statistike	Da	Ne

2. Navedeno je deset varijabli (obilježja promatranja). Napišite koje od tih varijabli su nominalne (n), ordinalne (o), kontinuirane (k) ili diskontinuirane (d).

a)	urod zrna pšenice (t/ha)	
b)	broj zrna po klipu kukuruza	
c)	boja zrna pšenice (izražena ocjenama od 1 do 3)	
d)	prisutnost osja	
e)	zemlja porijekla kultivara pšenice	
f)	slaganje s tvrdnjom (u potpunosti se ne slažem – u potpunosti se slažem)	
g)	uspjeh na ispitu	
h)	sadržaj škroba u zrnu (%)	
i)	postotak uginulih štetnika 7 dana nakon aplikacije pesticida	
j)	broj uginulih štetnika 7 dana nakon aplikacije pesticida	
k)	opseg ploda luka	

Rješenja se nalaze na 68. stranici!

2. Grupiranje podataka, apsolutna i relativna učestalost

I. Grupiranje podataka

Zašto se podaci grupiraju?

Podaci se grupiraju kako bi se povećala preglednost podataka. Prilikom grupiranja podataka treba paziti na zadovoljavanje principa preglednosti i principa iscrpnosti.

Prema **principu preglednosti**, podatke treba grupirati u što manje grupa (razreda), čime se povećava preglednost podataka, dok prema **principu preciznosti**, podatke treba grupirati u što više grupa (razreda) kako se ne bi smanjila preciznost.

U pravilu, podatke bi trebalo grupirati u 5 – 20 razreda, a razredi bi trebali biti iste veličine (kada je to moguće).

Na temelju **Sturgesova pravila**, moguće je odrediti odgovarajući broj razreda za svaku statističku seriju, a određuje se na temelju izraza:

$$k = 1 + 3{,}322(\log_{10} n)$$

gdje je:

k – broj razreda

n – veličina uzorka.

Podaci se u razrede razvrstavaju prema obilježju i njegovim modalitetima te prema principima:

- iscrpnosti – svaki se podatak mora razvrstati
- isključivosti – jedan podatak može biti član samo jedne grupe (razreda).

Numerički podaci se grupiraju tako da se jedinice koje imaju istu ili sličnu vrijednost obilježja (varijable) svrstavaju u istu grupu te se na taj način formiraju razredi:

- svaki razred obuhvaća određeni raspon numeričkog obilježja ograničen donjom ($L1_i$) i gornjom granicom ($L2_i$) razreda, tako da je: $L1_i < x_i \leq L2_i$
- kontinuirana obilježja – $L1_i \leq x_i < L2_i$
- diskontinuirana obilježja – $L1_i \leq x_i \leq L2_i$
- raspon vrijednosti obilježja obuhvaćen razredom naziva se veličina razreda

- **veličina razreda** je razlika između donje granice razreda i donje granice prethodnog razreda
- s obzirom na način formiranja granica, razlikujemo nominalne, prave i precizne granice (Tablica 1.).

Tablica 1. Primjer nominalnih, pravih i preciznih granica razreda

Nominalne granice		Prave granice		Precizne granice	
L_{1i}	L_{2i}	L_{1i}	L_{2i}	L_{1i}	L_{2i}
1	50	1	51	1,00	49,99
51	100	51	101	50,00	99,99
101	150	101	151	100,00	149,99
151	200	151	201	150,00	199,99

II. Apsolutna i relativna učestalost

Primjerice, imamo niz koji se sastoji od sljedećih numeričkih diskontinuiranih vrijednosti:

X_i: 1 2 3 1 1 5 3 5 4 2 2 5 10 11 2 26

Ako te vrijednosti složimo u rastuću statističku seriju (od najmanje do najveće vrijednosti – to se zove uređena statistička serija), dobijemo sljedeći niz:

X_i: 1 1 1 2 2 2 2 3 3 4 5 5 5 10 11 26

Tako uređene podatke možemo, na primjer, svrstati u tri razreda, s tim da su u jednom razredu vrijednosti do 9, u drugom vrijednosti od 10 do 19, a u trećem razredu vrijednosti od 20 do 29. Svrstavanjem vrijednosti u pripadajuće razrede te prebrojavanjem vrijednosti u svakom pojedinom razredu, utvrđujemo učestalost (frekvenciju) (Tablica 2.).

Tablica 2. Grupiranje podataka u tri razreda i utvrđivanje učestalosti

	0 – 9	10 – 19	20 – 29
	1, 1 ,1, 2, 2, 2, 2, 3, 3, 4, 5, 5, 5,	10, 11,	26
Učestalost	**13**	**2**	**1**

Nadalje, isti niz podataka može se grupirati i prema drugačijim kriterijima, recimo u razrede veličine 3 (Tablica 3.) ili se može na temelju Sturgesova pravila izračunati optimalan broj razreda za taj uzorak te bi u ovom slučaju optimalan broj razreda iznosio:

$$k = 1 + 3{,}322\ (\log_{10} n) = 1 + 3{,}322\ (\log_{10} 16) = 5$$

Tablica 3. Primjer formiranja razreda veličine 3 te izračunavanja apsolutne i relativne učestalosti

Razredi	Apsolutna učestalost (f_i)	Relativna učestalost (%)	Apsolutna kumulativna učestalost	Relativna kumulativna učestalost
1 – 3	9	= (9/16) x 100 = 56,25	9	56,25
4 – 6	4	= (4/16) x 100 = 25	= 9 + 4 = 13	= 56,25 + 25 = 81,25
7 – 9	0	= (0/16) x 100 = 0	= 13+ 0 = 13	= 81,25 + 0 = 81,25
10 – 12	2	= (2/16) x 100 = 12,5	= 13+ 2 = 15	= 81,25 + 12,5 = 93,75
13 – 15	0	= (0/16) x 100 = 0	= 15 + 0 = 15	= 93,75 + 0 = 93,75
16 – 18	0	= (0/16) x 100 = 0	= 15 + 0 = 15	= 93,75 + 0 = 93,75
19 – 21	0	= (0/16) x 100 = 0	= 15 + 0 = 15	= 93,75 + 0 = 93,75
22 – 24	0	= (0/16) x 100 = 0	= 15 + 0 = 15	= 93,75 + 0 = 93,75
25 – 27	1	= (1/16) x 100 = 6,25	= 15 + 1 = 16	= 93,75 + 6,25 = 100
Suma	16	100 %		

Formiranjem razreda te svrstavanjem podataka u razrede u koje pripadaju s obzirom na svoju vrijednost, izračunava se **učestalost** odnosno **frekvencija (f_i).**

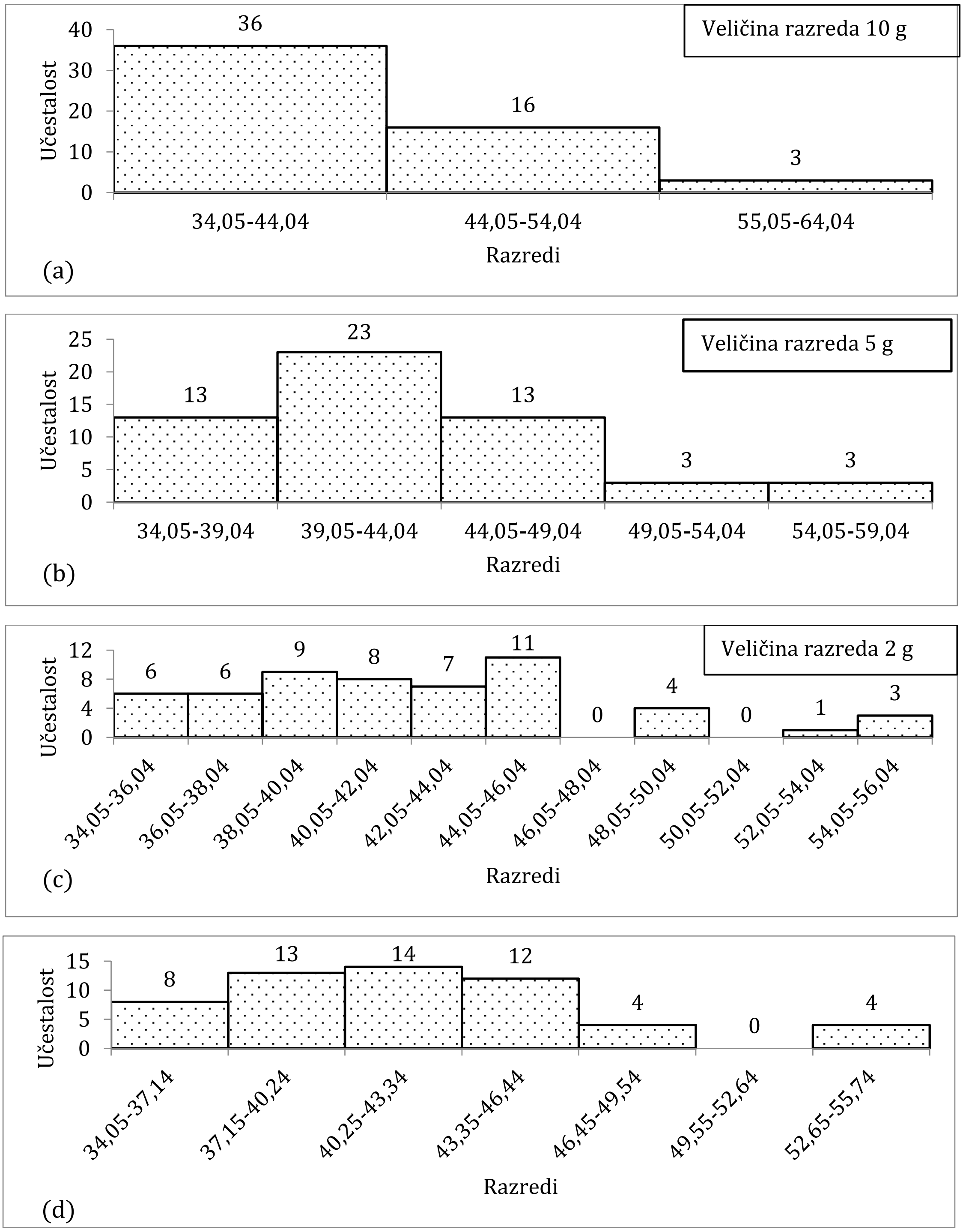
Učestalost ili frekvencija predstavlja broj vrijednosti koje pripadaju pojedinom razredu i kada je izražena u broju jedinica promatranja govorimo o **apsolutnoj učestalosti.**

Na temelju apsolutne učestalosti i veličine uzorka (ukupan broj podataka) izračunava se **relativna učestalost**, koja predstavlja postotni udio svakog razreda u ukupnom uzorku.

Na temelju vrijednosti apsolutne i relativne učestalosti izračunavaju se **apsolutna kumulativna** odnosno **relativna kumulativna učestalost.**

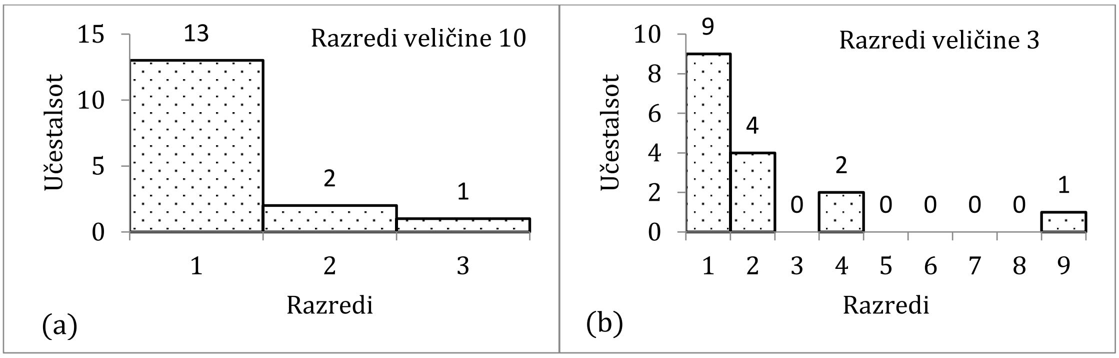
Učestalost se grafički najčešće prikazuje pomoću **histograma učestalosti**.

Na temelju iste statističke serije koju čine numeričke diskontinuirane vrijednosti kreirana su četiri histograma učestalosti. Za prva tri histograma proizvoljno je određen broj, odnosno veličina razreda, dok je u primjeru (d) pomoću Sturgesova pravila izračunan optimalan broj razreda (7) za navedeni niz podataka. Iz prikazanih primjera razvidno je kako odabir veličine, odnosno broja razreda utječe na vizualni dojam o raspodjeli ispitivanih vrijednosti.



Grafikon 1. Primjeri histograma učestalosti za isti set numeričkih diskontinuiranih vrijednosti (n = 55) kreiranih na temelju razreda različitih veličina

U nastavku su prikazana dva histograma učestalosti (Grafikon 2.) izrađena na temelju istih diskontinuiranih podataka, pri čemu su podaci grupirani u razrede veličine 10 (a) i razrede veličine 3 (b).



Grafikon 2. Histogrami učestalosti za isti set podataka kreirani na temelju podataka svrstanih u razrede veličine 10 (a) i razrede veličine (3)

Pomoću Sturgesova pravila izračunan je optimalan broj razreda za navedenu statističku seriju. Na temelju tih vrijednosti nacrtajte histogram te komentirajte razlike između histograma!

Primjer 1A.

Primjer formiranja razreda zadane veličine

U nastavku su prikazane vrijednosti mase 1000 zrna (g) 55 sorata ozime pšenice:

X_i :	34,05	34,06	34,26	35,31	35,46	35,77	36,56	36,57	37,56	37,68	37,80	37,88
	38,51	39,22	39,25	39,27	39,32	39,46	39,56	39,74	40,00	40,49	40,86	40,90
	41,12	41,42	41,59	41,64	41,83	42,19	42,27	42,58	42,60	42,80	43,14	43,58
	44,17	44,33	44,34	44,47	44,55	44,93	45,00	45,08	45,33	45,65	45,66	48,06
	48,38	49,11	49,43	53,27	54,32	54,33	55,35					

A. Na temelju prikazanih podataka potrebno je:

a) odgovoriti što je ispitivano numeričko obilježje (varijabla)?

b) odgovoriti kojeg je tipa ispitivana varijabla?

c) formirati razrede veličine 5 g.

Razredi zadane veličine formiraju se tako da se u statističkoj seriji utvrdi najniža izmjerena vrijednost (minimum). U seriji iz primjera 1A, najniža izmjerena vrijednost je 34,05 g te se ta vrijednost uzima kao donja granica prvog razreda.
Zadana veličina razreda je 5 g, što znači da razlika između donje granice prvog i donje granice drugog razreda treba iznositi 5 te je prema tome donja granica drugog razreda 39,05 g (34,05 + 5).
Prema istom pravilu izračunavaju se donje granice svih sljedećih razreda (39,05 + 5; 44,05 + 5; 49,05 + 5).
Gornja granica prvog i donja granica drugog razreda trebaju biti formirane tako da se ne preklapaju i tako da između njih nema „praznog prostora“. Na isti način određuju se gornje granice ostalih razreda. U Tablici 4. prikazani su formirani razredi i apsolutna učestalost za svaki pojedini razred.

Tablica 4. Formiranje razreda i utvrđivanje apsolutne učestalosti

Razredi	Apsolutna učestalost (f_i)
34,05 – 39,04	13
39,05 – 44,04	23
44,05 – 49,04	13
49,05- 54,04	3
54,05 – 59,04	3
Ukupno	55

d) izračunati relativnu učestalost

Relativna učestalost predstavlja postotni udio pojedinog razreda u ukupnom uzorku, a računa se ovako:

Relativna učestalost (%) = (apsolutna učestalost (f_i) / veličina uzorka) x 100

Prema tome, relativna učestalost prvog razreda je: (13/55) x 100 = 23,63 %, a relativne učestalosti ostalih razreda prikazane su u Tablici 5.

Tablica 5. Prikaz formiranih razreda, apsolutne i relativne učestalosti

Razredi	Apsolutna učestalost (f_i)	**Relativna učestalost (%)**
34,05 – 39,04	13	23,63
39,05 – 44,04	23	41,82
44,05 – 49,04	13	23,63
49,05- 54,04	3	5,46
54,05 – 59,04	3	5,46
Ukupno	55	100 %

e) izračunati apsolutnu kumulaciju iznad i apsolutnu kumulaciju ispod i odgovoriti na pitanja:

   i. koliko sorata pšenice ima masu 1000 zrna manju od 49,05 g?

   ii. koliko sorata pšenice ima masu 1000 zrna veću od 44,05 g?

Apsolutna kumulacija ispod izračunava se tako da se zbrajaju vrijednosti apsolutne učestalosti od prvog prema posljednjem razredu, dok se apsolutna kumulacija iznad izračunava zbrajanjem apsolutnih učestalosti, od posljednjeg prema prvom razredu (Tablica 6.).

Tablica 6. Prikaz formiranih razreda, apsolutne i relativne učestalosti te apsolutne kumulacije

Razredi	Apsolutna učestalost (f_i)	Relativna učestalost (%)	**Apsolutna kumulacija**	
			Ispod	**Iznad**
34,05 – 39,04	13	23,63	13	55
39,05 – 44,04	23	41,82	36	42
44,05 – 49,04	13	23,63	**49**	**19**
49,05- 54,04	3	5,46	52	6
54,05 – 59,04	3	5,46	55	3
Ukupno	55	100 %		

Nakon što je izračunana kumulacija „ispod“, može se utvrditi koliko sorata pšenice u ispitivanom uzorku ima masu 1000 zrna manju od 49,05 g. To se utvrđuje tako da se iz tablice iščita vrijednost iz stupca „Apsolutna akumulacija ispod“ i to u redu koji obuhvaća sve sorte koje su imale masu 1000 zrna nižu od 49,05, a to su sorte koje pripadaju u prva tri razreda.
Njihova suma učestalosti iznosi 49, što znači da je 49 od ukupno 55 sorata pšenice imalo masu 1000 zrna nižu od 49,05 g.
Na isti način, ali iz stupca „Apsolutna kumulacija iznad“ iščita se koliko sorata pšenice ima masu 1000 zrna veću od 44,05 g. To su sve sorte koje ulaze u razred čija je donja granica 44,05, i sve više razrede, što znači da te sorte imaju masu 1000 zrna veću od 44,05, a u ovom primjeru ima ukupno 19 takvih sorata.

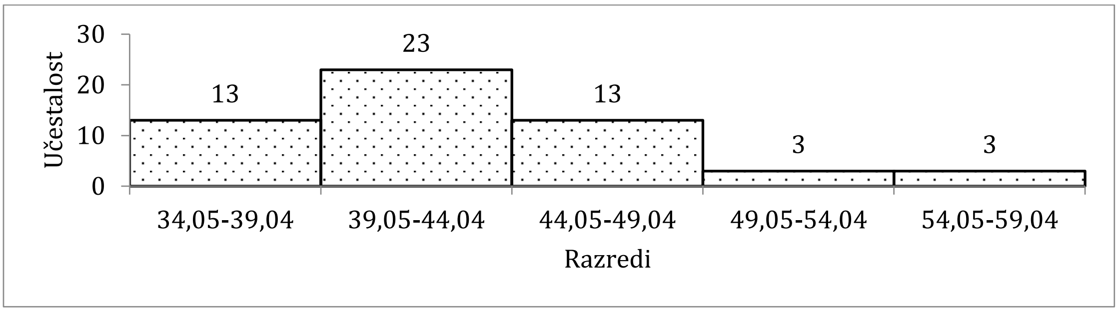
e) izračunati relativnu kumulaciju iznad i relativnu kumulaciju ispod i odgovoriti na pitanja:
   i. koliko sorata, u postotku, ima masu 1000 zrna manju od 49,05 g?
   ii. koliko sorata, u postotku, ima masu 1000 zrna veću od 44,05 g?

Relativna kumulacija ispod i iznad računaju se jednako kao apsolutne kumulacije ispod i iznad, s tim da se zbrajaju vrijednosti relativne učestalosti. Nakon što se izračunaju relativne kumulacije ispod i iznad, iz tablice se može vidjeti da 89,10 % ispitivanih sorata pšenice ima masu 1000 zrna manju od 49,05 g, dok 34,55 % ispitivanih sorata pšenice ima masu 1000 zrna veću od 44,05 g (Tablica 7.).

Tablica 7. Prikaz formiranih razreda, apsolutne i relativne učestalosti te apsolutne kumulacije i relativne kumulacije (ispod i iznad)

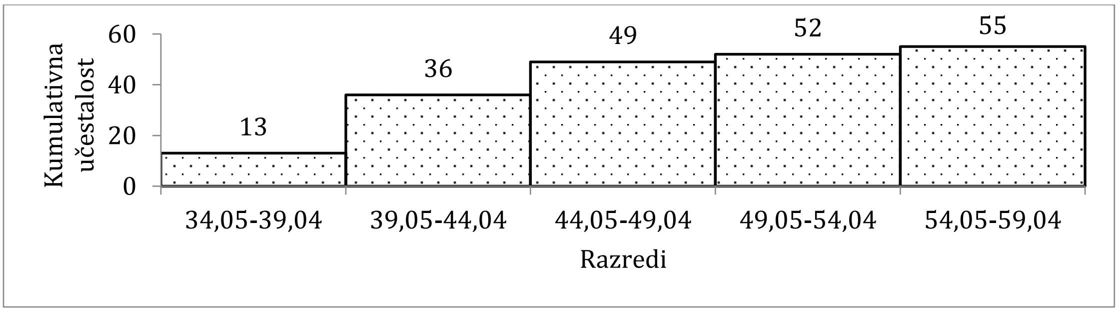
Razredi	Apsolutna učestalost (f_i)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
34,05 – 39,04	13	23,63	13	55	23,64	100
39,05 – 44,04	23	41,82	36	42	65,45	76,36
44,05 – 49,04	13	23,63	**49**	**19**	**89,10**	**34,55**
49,05- 54,04	3	5,46	52	6	94,55	10,90
54,05 – 59,04	3	5,46	55	3	100	5,46
Ukupno	55	100 %				

f) nacrtati histogram učestalosti

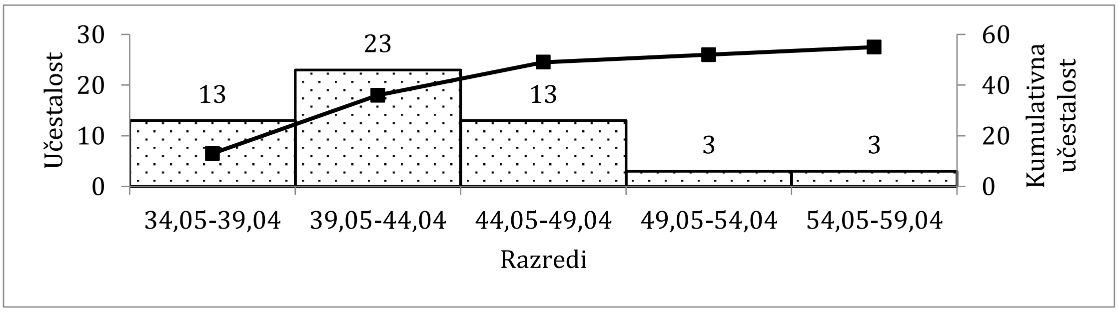


Grafikon 3. Histogram učestalosti mase 1000 zrna (g) 55 sorata ozime pšenice

g) nacrtati histogram kumulativne učestalosti



Grafikon 4. Histogram kumulativne učestalosti mase 1000 zrna (g) 55 sorata ozime pšenice



Grafikon 5. Primjer prikazivanja apsolutne učestalosti i apsolutne kumulativne učestalosti na istom grafikonu

Primjer 1B.

Primjer formiranja zadanog broja razreda

B. Na temelju istih podataka (stranica 10) potrebno je formirati šest (6) razreda te:

a) odgovoriti koliko sorata pšenice ima masu 1000 zrna manju od 44,85 g?
b) odgovoriti koliko sorata pšenice ima masu 1000 zrna veću od 44,85 g?
c) odgovoriti koliko sorata pšenice, u postotku, ima masu 1000 zrna manju od 41,25 g?
d) odgovoriti koliko sorata pšenice, u postotku, ima masu 1000 zrna veću od 52,04 g?
e) odgovoriti koliki je udio sorata (izraziti rezultat u postotku) s masom 1000 zrna većom od 41,25 g, a manjom od 44,84 g?
f) grafički prikazati apsolutnu kumulativnu učestalost.

Grupiranje podataka u šest (6) razreda

Da bismo niz podataka podijelili u zadani broj razreda, potrebno je izračunati veličinu razreda.

Veličina razreda je razlika između donje granice razreda i donje granice prethodnog razreda koja se računa tako da se prvo izračuna raspon varijacije (RV), koji predstavlja razliku između najviše i najniže vrijednosti u nizu.
Zatim se raspon varijacije podijeli s brojem razreda koje želimo formirati, a rezultat koji dobijemo predstavlja veličinu razreda, kako je prikazano:

$$\text{Veličina razreda} = (x_{max} - x_{min}) \,/\, \text{zadani broj razreda}$$

Prema tome, veličina razreda za zadanu seriju je:

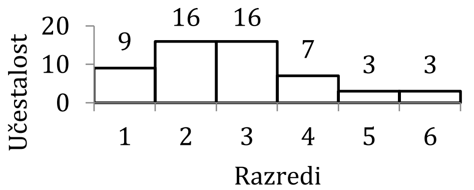
Veličina razreda = (55,35 – 34,05)/6 =
= 21,3/6 = **3,55** ≈ **3,60 g**

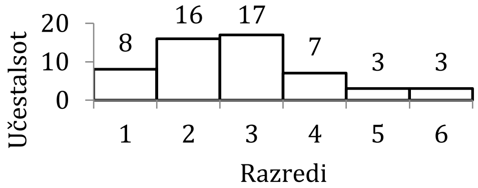
veličina razreda

U nastavku su prikazani primjeri formiranja razreda na temelju zadanog niza podataka s različitom donjom granicom prvog razreda te s različitim veličinama razreda.

Tablica 8. Formiranje šest razreda, veličina razreda 3,55 g

Razredi	f_i	Razredi	f_i
34,05 – 37,59	9	34,00 – 37,54	8
37,60 – 41,14	16	37,55 – 41,09	16
41,15 – 44,69	16	41,10 – 44,64	17
44,70 – 48,24	7	44,65 – 48,19	7
48,25 – 51,79	3	48,20 – 51,74	3
51,80 – 55,34	3	51,75 – 55,29	3
	54		54

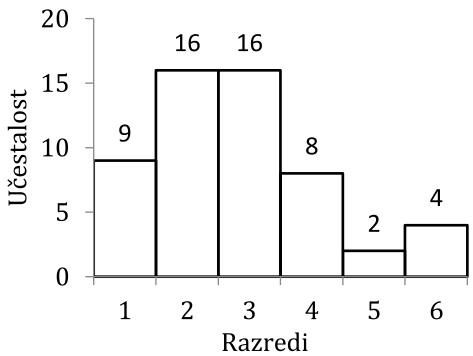




Najveća vrijednost mjerenja u nizu je 55,35 g. Formiranjem šest razreda, veličine razreda 3,55 g gornja granica zadnjeg (šestog) razreda je niža od najviše vrijednosti mjerenja (55,35 g) te ta maksimalna vrijednost ne može biti uvrštena ni u jedan razred. Zbog toga je ukupna suma učestalosti 54, iako je u nizu bilo 55 mjerenja (n = 55).
Ovako formirani razredi ne zadovoljavaju kriterij iscrpnosti, odnosno princip prema kojemu svaka vrijednost mjerenja **mora** biti svrstana u jedan od razreda.

Tablica 9. Formiranje šest razreda, veličina razreda 3,60 g

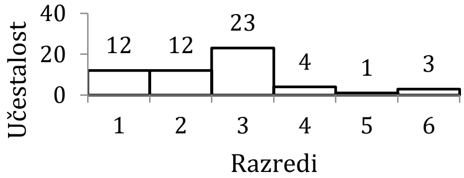
Razredi	f_i	Razredi	f_i
34,05 – 37,64	9	34,00 – 37,59	9
37,65 – 41,24	16	37,60 – 41,19	16
41,25 – 44,84	16	41,20 – 44,79	16
44,85 – 48,44	8	44,80 – 48,39	8
48,45 – 52,04	2	48,40 – 51,99	2
52,05 – 55,64	4	52,00 – 55,59	4
	55		55

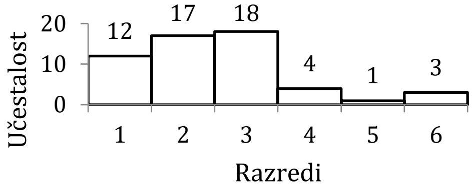


Formiranjem razreda veličine 3,60 g ispunjeni su kriteriji iscrpnosti i isključivosti te je u oba slučaja, neovisno o donjoj granici prvog razreda, utvrđena jednaka raspodjela učestalosti ispitivanih podataka.

Tablica 10. Formiranje šest razreda, veličina razreda 4,00 g

Razredi	f_i	Razredi	f_i
34,05 – 38,04	12	34,00 – 37,99	12
38,05 – 41,04	12	38,00 – 41,99	17
42,05 – 46,04	23	42,00 – 45,99	18
46,05 – 50,04	4	46,00 – 49,99	4
50,05 – 54,04	1	50,00 – 53,99	1
54,05 – 58,04	3	54,00 – 57,99	3
	55		55





Formiranjem šest razreda veličine razreda 4,00 g, također je zadovoljen princip iscrpnosti, no u ovisnosti o donjoj granici prvog razreda, utvrđena je različita učestalost vrijednosti po razredima.

Izračunavanje relativne učestalosti te apsolutne i relativne kumulacije

U Tablici 11. prikazane su vrijednosti učestalosti (apsolutne i relativne) te apsolutne i relativne kumulacije, koje su neophodne da bi se moglo odgovoriti na postavljena pitanja.

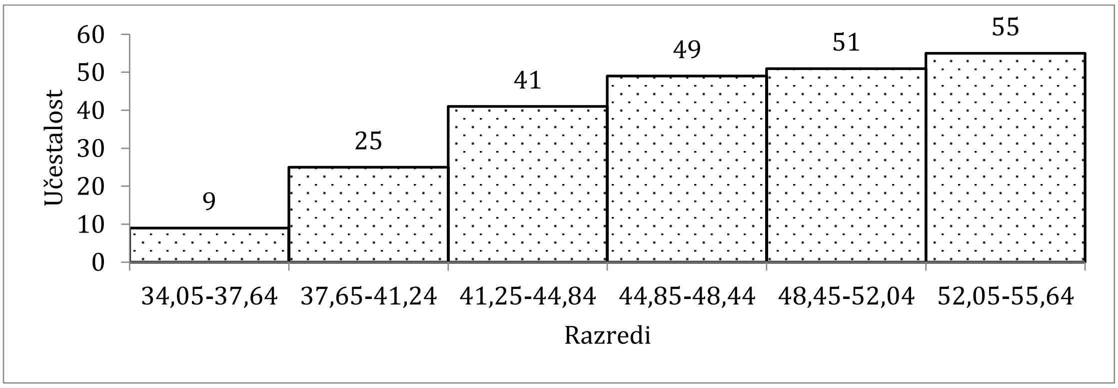
Tablica 11. Razredi, učestalosti i kumulacija za zadanu seriju podataka

Razredi	Učestalost (fi)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
34,05 – 37,64	9	16,36	9	55	16,36	100
37,65 – 41,24	16	29,09	25	46	c) **45,45**	83,64
41,25 – 44,84	e) 16	29,09	a) 41	30	74,54	54,55
44,85 – 48,44	8	14,55	49	b) 14	89,09	25,46
48,45 – 52,04	2	3,64	51	6	92,73	10,91
52,05 – 55,64	4	7,27	55	4	100	d) 7,27
Ukupno	55	100,00				

Odgovori:

a) 41 sorta pšenice ima masu 1000 zrna nižu od 44,85 g
b) 14 sorata pšenice ima masu 1000 zrna višu od 44,85 g
c) 45,45 % ispitivanih sorata pšenice ima masu 1000 zrna nižu od 41,25 g
d) 7,27 % ispitivanih sorata pšenice ima masu 1000 zrna višu od 52,05 g
e) 24 sorte imaju masu 1000 zrna veću od 41,25 g, a manju od 48,44 g, to je 43,63 %; (24/55) x 100 = 43,63 %

f) histogram kumulativne učestalosti



Grafikon 6. Histogram kumulativne učestalosti

RIJEŠI ZADATKE I ODGOVORI NA PITANJA!

1. Proveden je poljski pokus u kojemu su ispitivane razlike između sorata pšenice u duljini stabljike (cm). Na temelju prikupljenih podataka, potrebno je:

   a) formirati razrede veličine 4 cm
   b) izračunati apsolutnu i relativnu raspodjelu učestalosti
   c) izračunati apsolutnu i relativnu kumulaciju iznad i kumulaciju ispod
   d) grafički prikazati apsolutnu raspodjelu učestalosti
   e) odgovoriti koliko sorata ima duljinu stabljike manju od 45,7 cm.

Prilikom računanja veličine razreda i granice razreda zaokružiti na jednu decimalu.

Redni broj	Sorta	Duljina stabljike (cm)	Uređena statistička serija
1.	Ana	35,9	
2.	Janica	36,5	
3.	Seka	37,3	
4.	Dana	39,9	
5.	Korana	42,3	
6.	Elvira	39,4	
7.	Ružica	38,6	
8.	Divana	36,6	
9.	Ilirija	37,9	
10.	Jana	35,0	
11.	Katarina	33,7	
12.	Srpanjka	36,0	
13.	Magdalena	52,4	
14.	Nirvana	52,1	
15.	Libelula	45,3	
16.	Pipi	46,1	
17.	Renata	50,7	
18.	Sana	51,1	
19.	Osječka 20	49,3	
20.	U1	52,4	

Razredi	Učestalost (f_i)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
Ukupno						

d) grafički prikaz apsolutne raspodjele učestalosti!

e) koliko sorata ima duljinu stabljike manju od 45,7 cm?

Rješenje zadatka nalazi se na 68. stranici!

2. Proveden je poljski pokus u kojemu su ispitivane razlike između sorata pšenice u visini biljke (cm). Na temelju prikupljenih podataka potrebno je:

a) formirati 5 razreda
b) izračunati apsolutnu i relativnu raspodjelu učestalosti
c) izračunati kumulaciju iznad i kumulaciju ispod
d) grafički prikazati kumulativnu raspodjelu učestalosti
e) odgovoriti koliko sorata ima visinu biljke veću od 51,1, a manju od 55,7 cm
f) odgovoriti koliki je udio sorata pšenice koje imaju visinu biljke nižu od 55,7 cm u ispitivanom uzorku.

Prilikom računanja veličine razreda i granice razreda zaokružiti na jednu decimalu.

Redni broj	Sorta	Visina biljke (cm)
1.	Ana	42,2
2.	Janica	43,5
3.	Seka	41,5
4.	Dana	44,4
5.	Korana	47,3
6.	Elvira	44,2
7.	Ružica	43,0
8.	Divana	41,1
9.	Ilirija	42,5
10.	Jana	39,2
11.	Katarina	37,7
12.	Srpanjka	39,1
13.	Magdalena	58,5
14.	Nirvana	57,5
15.	Libelula	49,9
16.	Pipi	51,0
17.	Renata	55,6
18.	Sana	56,6
19.	Osječka 20	55,4
20.	U1	60,1

Uređena statistička serija

Razredi	Učestalost (f_i)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
Ukupno	20	100,00 %				

d) grafički prikaz kumulativne raspodjele učestalosti!

e) koliko sorata ima visinu biljke veću od 51,1, a manju od 55,7 cm?

f) koliki je udio sorata pšenice koje imaju visinu biljke nižu od 55,7 cm u ispitivanom uzorku?

Rješenje zadatka nalazi se na 68. stranici!

Na području Osječko–baranjske županije provedeno je anketiranje na slučajno odabranom uzorku od 100 ispitanika starosti od 18 do 65 godina. Ispitanici su s obzirom na dob grupirani u razrede veličine 10 godina.

Razredi	Učestalost (fi)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
18 – 27	26					
28 – 37	27					
38 – 47	12					
48 – 57	19					
58 – 67	16					
Ukupno						

a) kolika je veličina uzorka?

b) što je ispitivana varijabla?

c) što je jedinica promatranja?

d) što je osnovni skup – populacija?

e) grafički prikazati raspodjelu ispitanika po dobi!

f) utvrditi koja dobna skupina je najzastupljenija u ispitivanom uzorku i u kojem postotku?

g) koliki je udio najmanje zastupljene dobne skupine u ispitivanom uzorku?

h) koliki je udio ispitanika starijih od 48, a mlađih od 58 godina u ispitivanom uzorku?

Rješenje zadatka nalazi se na 69. stranici!

3. Mjere centralne tendencije – srednje vrijednosti

Srednja vrijednost je izračunana vrijednost oko koje se gomilaju podaci na temelju kojih je izračunana. U praksi se često koristi te služi za poopćavanje podataka, odnosno za prikazivanje niza varijabilnih podataka jednim brojem. S obzirom na način računanja, razlikujemo izračunane ili potpune te položajne srednje vrijednosti.

Izračunane srednje vrijednosti računaju se na temelju svih vrijednosti u statističkoj seriji te su osjetljive na ekstremno niske ili ekstremno visoke vrijednosti. Ovoj skupini srednjih vrijednosti pripadaju:

- aritmetička sredina
- geometrijska sredina
- harmonijska sredina.

Položajne srednje vrijednosti određene su svojim položajem u uređenoj statističkoj seriji i nisu osjetljive na ekstremne vrijednosti. U ovu skupinu srednjih vrijednosti pripadaju:

- modus (mod)
- medijana.

I. Izračunane srednje vrijednosti

i. Aritmetička sredina

Aritmetička sredina je najpoznatija i najčešće korištena potpuna mjera centralne tendencije. Često se koristi i u svakodnevnom životu te se za nju koriste nazivi „srednja vrijednost“, „prosječna vrijednost“ ili „prosjek“.

Uobičajeno se označava sa $\bar{x}$ (x potez), a izražava se u jedinici mjerenja numeričke varijable. Aritmetičku sredinu može se definirati kao omjer zbroja svih vrijednosti numeričke varijable i ukupnog broja vrijednosti numeričke varijable.

Kada se aritmetička sredina računa za podatke koji nisu grupirani, odnosno kada znamo pojedinačne vrijednosti numeričke varijable, govorimo o *jednostavnoj aritmetičkoj sredini*, koja se računa na temelju formule:

$$\overline{X} \equiv \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

gdje je:

X_i – vrijednost jedinice promatranja

n – broj članova statističke serije, veličina uzorka.

Primjer 1.

Primjer izračunavanja jednostavne aritmetičke sredine

Na temelju podataka iz Tablice 12., izračunati jednostavnu aritmetičku sredinu (prosjek) otkupne cijene mlijeka (kn/l) u Republici Hrvatskoj u protekle tri godine!

Tablica 12. Otkupne cijene mlijeka u Republici Hrvatskoj od 2014. do 2016. godine

Godina	Mjeseci kroz godinu											
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
2014.	2,44	2,45	2,43	2,45	2,43	2,44	2,44	2,44	2,38	3,38	2,36	2,36
2015.	2,36	2,36	2,36	2,36	2,36	2,78	2,67	2,49	2,99	2,18	2,15	2,01
2016.	2,02	2,03	2,00	2,03	2,03	1,80	1,80	1,90	1,90	2,10	2,10	2,10

$$\overline{X} \equiv \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_i}{n} =$$

$$= \frac{2{,}44 + 2{,}45 + 2{,}43 + 2{,}45 + \cdots + 2{,}10}{n} = = \frac{82{,}88}{36} = 2{,}30 \text{ kn/l}$$

Prosječna otkupna cijena mlijeka u Republici Hrvatskoj u protekle tri godine (2014. – 2016.) iznosi 2,30 kn/l.

U slučajevima kada imamo distribuciju frekvencija, odnosno kada su podaci podijeljeni u razrede gdje se svaka vrijednost numeričke varijable pojavljuje s određenom učestalošću ili kada sve vrijednosti obilježja nemaju istu važnost za izračunavanje aritmetičke sredine, koristi se formula za *složenu, ponderiranu ili vaganu aritmetičku sredinu:*

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum f_i x_i}{\sum f_i}$$

gdje je:

x_i – vrijednost jedinice promatranja

f_i – učestalost (frekvencija)

n – broj članova statističke serije, veličina uzorka.

Primjer 2.

Primjer izračunavanja složene aritmetičke sredine

Proveden je pokus u kojemu je utvrđen broj zrna po klasu za 200 sorata pšenice (Tablica 13.). Na temelju prikazanih podataka potrebno je:

a) utvrditi kojeg je tipa ispitivana varijabla?
b) odrediti veličinu razreda?
c) izračunati koliki je prosječan broj zrna po klasu pšenice u ispitivanom uzorku?
d) utvrditi koja je skupina sorata pšenice najzastupljenija u ispitivanom uzorku s obzirom na broj zrna po klasu?
e) grafički (histogramom) prikazati apsolutnu raspodjelu učestalosti te utvrditi u koji od ponuđenih razreda bi ulazila aritmetička sredina uzorka te s kolikim je udjelom taj razred zastupljen u ispitivanom uzorku.
f) izračunati kolika je zastupljenost (%) sorata koje su imale više od 43, a manje od 47 zrna po klasu?

Tablica 13. Raspodjela učestalosti za obilježje broja zrna po klasu u 200 sorata pšenice

Razredi	Učestalost (f_i)
18 – 22	2
23 – 27	3
28 – 32	5
33 – 37	17
38 – 42	37
43 – 47	37
48 – 52	48
53 – 57	31
58 – 62	18
63 – 67	2
Ukupno	200

Složena aritmetička sredina računa se prema formuli:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

U navedenoj formuli x_i je vrijednost jedinice promatranja. S obzirom na to da su podaci raspoređeni u razrede, potrebno je izračunati **sredinu razreda** (x_i), koja predstavlja vrijednost jedinica promatranja određenog razreda, a računa se kao prosjek donje i gornje granice razreda prema formuli:

$$x_i = \frac{L_{1i} + L_{2i}}{2}$$

gdje je:

x_i – vrijednost jedinice promatranja

L_{1i} – donja granica razreda

L_{2i} – gornja granica razreda.

Tako izračunana sredina razreda množi se s pripadajućom učestalošću (f_i), čime se dobiju pojedinačni produkti – podtotali za svaki razred pojedinačno. Zbrajanjem podtotala dobije se total, koji se uvrštava u brojnik formule za složenu (ponderiranu) aritmetičku sredinu (Tablica 14.).

Tablica 14. Primjer izračunavanja sredine razreda (x_i) i podtotala (f_ix_i)

Razredi	Učestalost (f_i)	Sredina razreda (x_i)	Podtotali (f_ix_i)
18 – 22	2	= (18 + 22)/2 = 20	= 2 x 20 = 40
23 – 27	3	= (23 + 27)/2 = 25	= 3 x 25 =75
28 – 32	5	= (28 + 32)/2 = 30	= 5 x 30 =150
33 – 37	17	= (33 + 37)/2 = 35	= 17 x 35 = 595
38 – 42	37	= (38 + 42)/2 = 40	= 37 x 40 = 1480
43 – 47	37	= (43 + 47)/2 = 45	= 37 x 45 = 1665
48 – 52	48	= (48 + 52)/2 = 50	= 48 x 50 = 2400
53 – 57	31	= (53 + 57)/2 = 55	= 31 x 55 = 1705
58 – 62	18	= (58 + 62)/2 = 60	= 18 x 60 = 1080
63 – 67	2	= (63 + 67)/2 = 65	= 2 x 65 = 130
Ukupno	$\sum (f_i)$ 200		$\sum(f_ix_i)$ 9320

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{9320}{200} = 46{,}6 \text{ zrna po klasu}$$

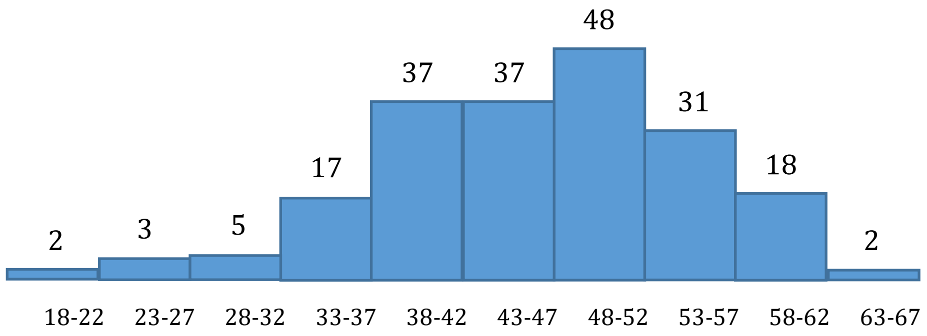
a) ispitivana varijabla (broj zrna po klasu) numerička je kvantitativna diskontinuirana varijabla

b) ispitivani podaci svrstani su u razrede veličine 5 zrna

c) u ispitivanom uzorku utvrđena je aritmetička sredina odnosno prosječan broj zrna po klasu od 46,6 zrna

d) s obzirom na broj zrna po klasu, u ispitivanom uzorku je najzastupljenija skupina sorata koje su imale broj zrna po klasu između 48 i 52 zrna

e) izračunana aritmetička sredina od 46,6 zrna po klasu ulazila bi u razred koji obuhvaća sorte koje su imale 43 do 47 zrna po klasu i taj razred je zastupljen s 18,5 % u ispitivanom uzorku.



Histogram učestalosti broja zrna po klasu 200 sorata ozime pšenice

f) udio razreda (43 – 47) u uzorku = (37/200) x 100 = 18,5 %

ii. Harmonijska sredina

Harmonijska sredina koristi se za izračunavanje sredine relativnih brojeva s istim brojnicima, npr. za izračunavanje prosječne brzine savladavanja određene radnje ili za izračunavanje produktivnosti rada. Izražava se u jedinici mjerenja numeričke varijable.

Kada su nam poznate pojedinačne vrijednosti numeričke varijable, izračunavamo *jednostavnu harmonijsku sredinu*. Računa se prema formuli:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

gdje je:

x_i – vrijednost jedinice promatranja

n – broj članova statističke serije, veličina uzorka.

Prilikom računanja *ponderirane harmonijske sredine*, kao ponderi pojedinih vrijednosti varijable uzimaju se učestalosti ili njima proporcionalne veličine, a računa se prema formuli:

$$H = \frac{f_1 + f_2 + \cdots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

gdje je:

x_i – vrijednost jedinice promatranja

f_i – učestalost.

Primjer 3.

Izračunavanje jednostavne harmonijske sredine

Na istraživanom području ima ukupno 750 km kanala uz oranice koji se održavaju strojnom košnjom. Kosi se pomoću šest različitih tipova kosilica od kojih svaka ima optimalnu brzinu rada (6,0 km/h, 6,5 km/h, 7,0 km/h, 7,5 km/h, 8,0 km/h i 8,5 km/h), a radnici su dužni pridržavati se preporučenih brzina košnje. Svakom kosilicom pokošeno je 125 km. Treba izračunati:

a) koja je prosječna brzina košnje (km/h)
b) koliko je radnih sati potrebno da bi se pokosilo svih 750 km.

S obzirom na to da je pomoću svih šest kosilica pokošena jednaka duljina ruba kanala (125 km), za izračunavanje prosječne brzine košnje upotrebljavamo formulu za jednostavnu harmonijsku sredinu:

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{6}{\frac{1}{6} + \frac{1}{6{,}5} + \frac{1}{7} + \frac{1}{7{,}5} + \frac{1}{8} + \frac{1}{8{,}5}} = \frac{6}{0{,}84} = 7{,}14$$

a) prosječna brzina košnje je 7,14 km/h
b) s obzirom na to da je prosječna brzina košnje 7,14 km/h, znači da je za košnju svih 750 km potrebno 750/7,14 = 105,04 radnih sati, odnosno 13,13 osmosatnih radnih dana.

Primjer 4.

Izračunavanje složene harmonijske sredine

Pretpostavimo da tijekom košnje nisu sve kosilice jednako korištene, već da je 300 km pokošeno kosilicom koja radi brzinom od 6,0 km/h, 256 km je pokošeno kosilicom koja radi brzinom od 6,5 km/h, 62 km je pokošeno kosilicom koja radi brzinom od 7,0 km/h, 48 km je pokošeno kosilicom koja radi brzinom od 7,5 km/h, 14 km je pokošeno kosilicom koja radi brzinom od

8,00 km/h, a preostalih 70 km je pokošeno kosilicom koja radi brzinom od 8,5 km/h. Potrebno je izračunati:

a) koja je prosječna brzina košnje (km/ha)
b) koje je vrijeme (radni sati) potrebno za košnju svih 750 km.

$$H = \frac{\sum f_i}{\sum \frac{f_i}{x_i}} = \frac{300 + 256 + 62 + 48 + 14 + 70}{\frac{300}{6} + \frac{256}{6,5} + \frac{62}{7} + \frac{48}{7,5} + \frac{14}{8} + \frac{70}{8,5}} = \frac{750}{114,62} = 6,54$$

a) prosječna brzina košnje je 6,54 km/h

b) za košnju svih 750 km potrebno je 14,33 radnih dana odnosno 114,62 sata.

II. Položajne srednje vrijednosti

i. Mod (modus)

Mod je najčešće puta postignuta vrijednost u nizu mjerenja.

S obzirom na to da nije izračunana nego položajna srednja vrijednost, mod se može koristiti i za kvalitativne (kategoričke) i kvantitativne (numeričke) varijable.

Da bi se mod mogao odrediti, moraju postojati barem dvije jednake vrijednosti varijable. Kod grupiranih obilježja modom se može smatrati vrijednost sredine razreda koja ima najveću učestalost, no to nije precizno te se za grupirane podatke mod računa prema formuli:

$$M_{o=} L_1 + \left[\frac{(b-a)}{(b-a)+(b-c)} \times i\right]$$

gdje je:

L_1 – donja granica modalne grupe (modalnog razreda)

a – učestalost razreda koji prethodi modalnom razredu

b – učestalost modalnog razreda

c – učestalost razreda koji slijedi nakon modalnog razreda

i – veličina razreda.

Primjer 5.

Određivanje moda za nominalne varijable

U Tablici 15. prikazane su kategorije poljoprivrednih površina s obzirom na način korištenja u Republici Hrvatskoj u 2014. godini (Statistički ljetopis 2015.). Podaci su uređeni u rastući niz s obzirom na njihov udio u ukupnim poljoprivrednim površinama u Republici Hrvatskoj.

Na temelju prikazanih podataka možemo zaključiti da je modalna kategorija „Oranice i vrtovi" jer su najzastupljenije, odnosno imaju najveću učestalost u odnosu na ostale kategorije poljoprivrednih površina. Kod nominalnih varijabli, mod se iščitava iz tablice podataka i on predstavlja najučestaliju kategoriju podataka.

Tablica 15. Kategorije poljoprivrednih površina u Republici Hrvatskoj u 2014. godini

Kategorija poljoprivredne površine	**Prosječna površina (ha)**	**Udio u ukupnoj površini (%)**
Rasadnici	221	0,01
Košaračka vrba i božićna drvca	407	0,03
Povrtnjaci	2 150	0,14
Maslinici	19 082	1,26
Vinogradi	26 164	1,73
Voćnjaci	31 724	2,10
Livade i trajni travnjaci	618 070	40,96
Oranice i vrtovi	811 067	53,75
Ukupno	1 508 885	100,00

Primjer 6.

Određivanje moda za ordinalne varijable

U slučaju ordinalnih varijabli, kao što su ocjene od 1 do 5 na Likertovoj skali, također možemo odrediti mod.

Na primjer: provedena je anketa (n = 1166), u kojoj jedno pitanje glasi:

> Kako biste ocijenili svoj stav o tvrdnji da prehrambeni proizvodi koji sadrže GMO organizme nisu štetni za zdravlje (1 – uopće se ne slažem, 2 – djelomično se ne slažem, 3 – nemam mišljenje, 4 – slažem se, 5 – potpuno se slažem)?

Tablica 16. Raspodjela učestalosti odgovora na postavljeno pitanje

Odgovori	**Učestalost**	**Udio (%)**
Uopće se ne slažem (1)	285	24,44
Djelomično se ne slažem (2)	198	16,98
Nemam mišljenje (3)	523	44,85
Slažem se (4)	101	8,66
Potpuno se slažem (5)	59	5,06
Ukupno	1166	100

Na temelju prikazanih podataka možemo zaključiti da je mod, odnosno modalna kategorija odgovor „Nemam mišljenje“, odnosno najveći broj ispitanika, njih 523, tvrdi da nemaju mišljenje o tome je li hrana koja sadrži GMO organizme štetna za zdravlje.

Primjer 7.

Izračunavanje moda za grupirane kvantitativne varijable

Na temelju raspodjele učestalosti duljine klasova pšenice (cm), treba izračunati mod!

Razredi	Učestalost	
5,21 – 5,80	5	
5,81 – 6,40	21	Grupa prije modalne grupe
6,41 – 7,00	102	Modalna grupa
7,01 – 7,60	48	Grupa poslije modalne grupe
7,61 – 8,20	23	
8,21 – 8,80	1	
Ukupno	200	

Mod za grupirane podatke računa se na temelju formule:

$$M_{o}= L_1 + \frac{(b-a)}{(b-a)+(b-c)} \times i$$

Modalna grupa ima najveću učestalost!

gdje je:

L_1 – donja granica modalne grupe (modalnog razreda)

a – učestalost razreda koji prethodi modalnom razredu

b – učestalost modalnog razreda

c – učestalost razreda koji slijedi nakon modalnog razreda

i – veličina razreda.

Izračunavanje veličine razreda prikazano je na stranici 9!

$$M_{o}= L_1 + \left[\frac{(b-a)}{(b-a)+(b-c)} \times i\right] = 6{,}41 + \left[\frac{(102-21)}{(102-21)+(102-48)} \text{ x } 0{,}60\right] =$$

$$= 6{,}41 + 0{,}36 = 6{,}77$$

U ispitivanom uzorku (n = 200) najčešće puta postignuta vrijednost mjerenja je 6,77 cm.

ii. Medijana

Medijana je centralna vrijednost raspodjele učestalosti. Nalazi se u sredini sređene serije podataka te ju dijeli na dva jednaka dijela.

S obzirom na to da se za njezino računanje ne uzimaju u obzir sve vrijednosti mjerenja, koristi se za varijable s ekstremnim vrijednostima.

Prilikom računanja medijane na temelju sljedećih formula zapravo izračunavamo mjesto u uređenoj statističkoj seriji na kojemu se nalazi medijana, a ne samu medijanu.

Kod *neparne sređene serije podataka* medijana je srednji član, a izračunava se na temelju formule:

$$M_e = \frac{N+1}{2}$$

Kod *parne sređene serije podataka* za medijanu se određuju dva člana niza, na temelju kojih se izračunava mjesto na kojem se nalazi medijana:

$$1.\,\text{član: } M_e = \frac{N}{2}$$

$$2.\,\text{član: } M_e = \frac{N}{2} + 1$$

$$M_e = \frac{1.\,\text{član} + 2.\,\text{član}}{2}$$

Za podatke koji su grupirani u razrede medijana se računa pomoću sljedeće formule:

$$M_e = L_1 + \left[\frac{\frac{N}{2} - \sum_{i=1}^{m} f_i}{f_{med}}\right] * i$$

gdje je:

L_1 – donja granica medijalnog razreda

$\frac{N}{2}$ – polovina članova niza

$\sum_{i=1}^{m} f_i$ – zbroj svih frekvencija do medijalnog razreda

f_{med} = frekvencija medijalnog razreda

i = veličina medijalnog razreda.

Primjer 8.

Izračunavanje medijane neparne serije podataka

Izračunati medijanu za sljedeći neparni niz!

Mjesto u nizu	1.	2.	3.	4.	5.
x_i	2	4	6	8	10

$$M_e = \frac{N+1}{2} = \frac{5+1}{2} = 3$$

Medijana se nalazi na trećem mjestu u uređenoj statističkoj seriji. Medijana je 6, što znači da je 50 % vrijednosti u nizu manje, a 50 % vrijednosti u nizu veće je od 6.

Primjer 9.

Izračunavanje medijane parne serije podataka

Izračunati medijanu za sljedeći parni niz!

Mjesto u nizu	1.	2.	3.	4.	5.	6.
x_i	2	3	5	6	8	10

1. član: $M_e = \frac{N}{2} = \frac{6}{2} = 3$

2. član: $M_e = \frac{N}{2} + 1 = \frac{6}{2} + 1 = 4$

Medijana se nalazi između 3. i 4. mjesta u uređenoj statističkoj seriji. Da bismo izračunali koji je to broj, koristimo sljedeću formulu:

$$M_e = \frac{1.\,\text{član} + 2.\,\text{član}}{2} = \frac{5+6}{2} = 5{,}5$$

Medijana je 5,5 što znači da je 50 % vrijednosti u nizu niže, a 50 % vrijednosti u nizu više od 5,5.

Primjer 10.

Izračunavanje medijane za podatke grupirane u razrede

Na temelju podataka grupiranih u razrede i s već unaprijed utvrđenom učestalošću i izračunanom apsolutnom kumulacijom, potrebno je izračunati medijanu.

Tablica 17. Podaci na temelju kojih treba izračunati medijanu

Razredi	Učestalost (f_i)	Kumulacija
18 – 22	2	2
23 – 27	3	5
28 – 32	5	10
33 – 37	17	27
38 – 42	37	64
43 – 47	37	101
48 – 52	48	149
53 – 57	31	180
58 – 62	18	198
63 – 67	2	200
Ukupno	200	

$$M_e = L_1 + \left[\frac{\frac{N}{2} - \sum_{i=1}^{m} f_i}{f_{med}}\right] * i =$$

$$= 43 + \left[\frac{\frac{200}{2} - 64}{37}\right] * 5 = 47{,}68 \sim 48$$

Medijana je 48, znači da je 50 % vrijednosti niže, a 50 % vrijednosti više od 48.

RIJEŠI ZADATKE I ODGOVORI NA PITANJA!

1. Zaokruži točne odgovore!

a)	modus i medijana su mjere srednje vrijednosti	Da	Ne
b)	aritmetička sredina je jako osjetljiva na ekstremne vrijednosti	Da	Ne
c)	aritmetička sredina, modus i medijana su približno jednake vrijednosti kod normalno distribuiranih podataka	Da	Ne

2. Što je vrijednost 543 u sljedećem nizu podataka: 170, 249, 523, 543, 572, 689, 1050?

a) aritmetička sredina
b) medijana
c) modus

3. Koja je aritmetička sredina sljedećeg niza podataka: 1501, 1736, 1930, 1176, 446, 428, 768, 861?

a) 1105,75
b) 1018,5
c) 428
d) 1930
e) 8

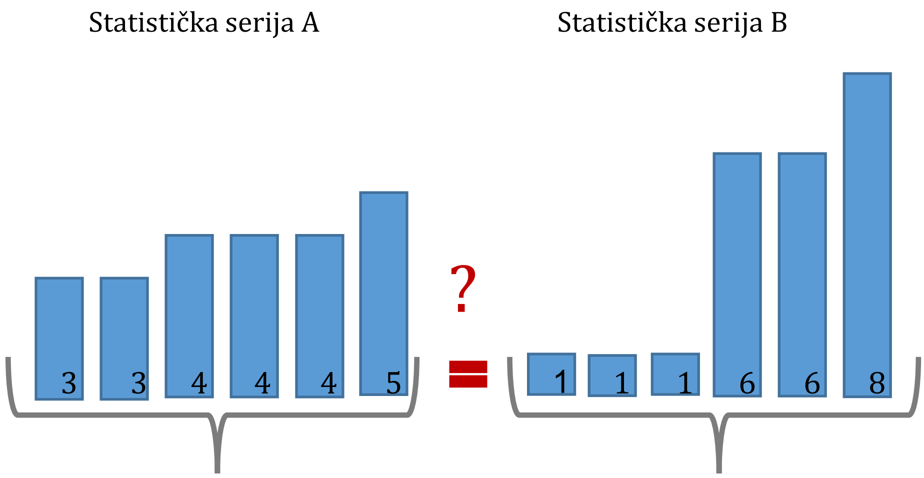
4. Imamo set podataka o naseljenosti pojedinih županija u Republici Hrvatskoj. Odgovori su:

a. Osječko-baranjska županija
b. Splitsko-dalmatinska županija
c. Zagrebačka županija
d. Vukovarsko-srijemska županija

a) kojeg je tipa ispitivana varijabla?
b) koju od mjera centralne tendencije bismo trebali koristiti za ovu vrstu varijabli?

Rješenje zadatka je na 69. stranici!

5. Na temelju podataka prikazanih u sljedećem grafičkom prikazu, komentirajte vrijednosti aritmetičke sredine, moda i medijane te usporedite njihove vrijednosti u dva seta podataka.



	Statistička serija A	Statistička serija B
Veličina uzorka	6	6
Aritmetička sredina	3,83	3,83
Mod	4	1
Medijana	4	3,5

6. S pokusnih parcelica površine 6,5 m^2 uzeti su uzorci od po 50 nasumično izabranih biljaka devet sorata pšenice. Klasovi su odrezani sa stabljike, izvagani te je za svaku sortu izračunana prosječna masa klasa prikazana u tablici.

Na temelju navedenih podataka potrebno je:

a) izračunati aritmetičku sredinu mase klasova navedenih sorata pšenice i odgovoriti na sljedeća pitanja:
b) kolika je veličina uzorka?
c) što je ispitivana varijabla, a što jedinica promatranja?
d) koliki je udio sorata s masom klasa nižom od 1 g u ispitivanom uzorku?

Sorta	Masa klasa (g)
Tena	1,43
Atena	1,18
Matea	0,93
Njivka	1,67
Srpanjka	2,39
Sana	1,10
Elvira	1,39
Žitarka	1,84
Super žitarka	1,21

a) aritmetička sredina mase klasova navedenih sorata pšenice

b) kolika je veličina uzorka?

c) što je ispitivana varijabla, a što jedinica promatranja?

d) koliki je udio sorata s masom klasa nižom od 1 g u ispitivanom uzorku?

Rješenje zadatka je na 69. stranici!

7. Firma zapošljava 18 djelatnika od kojih su 3 pomoćna djelatnika bez završene škole (NKV), 9 djelatnika sa završenom srednjom školom (SSS), 4 djelatnika sa završenim stručnim studijem (VS) i 2 djelatnika sa završenim sveučilišnim studijem (VSS). Plaće djelatnika variraju s obzirom na stupanj stručne spreme i radni staž te se kreću u sljedećim rasponima:

6350,00 – 8250,00 kn za VSS
4450,00 – 6000,00 kn za VŠ
3200,00 – 4100,00 kn za SSS
2300,00 – 3125,00 kn za NKV.

Na temelju priloženih podataka treba izračunati prosječnu plaću u firmi!

Raspon plaća – razredi	**Učestalost (f_i)**	**Sredina razreda (x_i)**	**Podtotali ($f_i x_i$)**
Ukupno			

Prosječna plaća u firmi iznosi ____________________________ .

Rješenje zadatka je na 70. stranici!

8. Na odjelu za pakiranje jagoda, u prvoj smjeni, četiri radnika pakiraju jagode u posudice veličine 0,5 kg. Radnici ne rade jednakom brzinom.

- prvi (1) radnik za 8 sati rada napuni 400 posudica
- drugi (2) radnik za 8 sati rada napuni 450 posudica
- treći (3) radnik za 8 sati rada napuni 300 posudica
- četvrti (4) radnik za 8 sati rada napuni 370 posudica

Odgovoriti na pitanja:

a) koliki je prosječan broj napunjenih posudica po radniku u prvoj smjeni?

b) koliko je prosječno vrijeme (minuta) punjenja jedne posudice jagodama u prvoj smjeni?

Rješenje zadatka je na 70. stranici!

9. Na odjelu za pakiranje jagoda, u drugoj smjeni, 5 radnika pakira jagode u posudice veličine 0,5 kg. Radnici ne rade jednakom brzinom.

- prvi (1) radnik za 8 sati rada napuni 300 posudica
- drugi (2) radnik za 8 sati rada napuni 330 posudica
- treći (3) radnik za 8 sati rada napuni 500 posudica
- četvrti (4) radnik za 8 sati rada napuni 420 posudica
- peti (5) radnik za 8 sati rada napuni 400 posudica

a) koliki je prosječan broj napunjenih posudica po radniku u drugoj smjeni?

b) koliko je prosječno vrijeme (min) punjenja jedne posudice jagodama u drugoj smjeni?

c) koliko prosječno posudica napune jagodama obje smjene zajedno u jednom danu?

d) koliko je prosječno vrijeme (minuta) punjenja jedne posudice jagodama zajedno u obje smjene?

Rješenje zadatka je na 70. stranici!

10. Proveden je poljski pokus u kojemu je jedan od ciljeva istraživanja bio utvrditi broj zrna po klasu ispitivanih sorata pšenice. U istraživanje je bilo uključeno 200 sorata pšenice. Na temelju podataka prikazanih u tablici treba utvrditi koja je vrijednost mjerenja postignuta najviše puta. Na 27. stranici je izračunana aritmetička sredina za isti set podataka. Usporedi i komentiraj vrijednosti moda i aritmetičke sredine!

Razredi	Učestalost (f_i)
18 – 22	2
23 – 27	3
28 – 32	5
33 – 37	17
38 – 42	37
43 – 47	37
48 – 52	48
53 – 57	31
58 – 62	18
63 – 67	2
Ukupno	200

a) izračunavanje moda

b) usporedba i komentar izračunane aritmetičke sredine i moda

Rješenje zadatka je na 70. stranici!

11. Na temelju podataka o prosječnom urodu kukuruza (t/ha) u Republici Hrvatskoj u razdoblju od 2000. do 2010. godine, izračunati aritmetičku sredinu, mod i medijanu.

Godina	Urod (t/ha)
2000.	4,10
2001.	5,70
2002.	6,40
2003.	4,20
2004.	6,30
2005.	6,90
2006.	6,50
2007.	4,90
2008.	8,00
2009.	7,40
2010.	7,00

a) aritmetička sredina

b) mod

c) medijana

Rješenje zadatka je na 71. stranici!

12. Pšenica je uz kukuruz najvažnija žitarica u Republici Hrvatskoj. Na temelju podataka o proizvodnji pšenice u Republici Hrvatskoj u razdoblju od 1986. do 2010. godine, uređenih u jednostavnu statističku seriju, izračunati:

a) aritmetičku sredinu

b) mod

c) medijanu.

Zatim podatke svrstati u 6 razreda te na temelju grupiranih podataka također izračunati

e) aritmetičku sredinu

f) mod

g) medijanu

Usporediti i komentirati mjere centralne tendencije izračunane na temelju jednostavne statističke serije i na temelju podataka svrstanih u razrede!

Godina	Ukupna proizvodnja pšenice u tis. tona
2003.	506
1999.	558
2005.	602
1992.	658
2010.	681
1996.	741
1994.	750
2004.	801
2006.	805
2001.	812
2007.	812
2002.	823
1997.	834
2008.	858
2000.	865
1995.	877
1993.	887
2009.	936
1998.	1020
1986.	1078
1987.	1274
1989.	1288
1988.	1434
1990.	1602
1991.	1996

a) jednostavna aritmetička sredina

b) mod

c) medijana

d) izračunavanje veličine razreda i formiranje razreda

Razredi	Učestalost (f_i)	Sredina razreda (x_i)	Podtotali	Apsolutna kumulacija ispod
Ukupno				

e) složena aritmetička sredina

f) mod

g) medijana

h) usporedba i komentar izračunanih mjera centralne tendencije

Rješenje zadatka je na 71. stranici

4. Mjere varijacije

Mjerama varijacije opisuje se varijabilnost vrijednosti unutar ispitivane serije podataka. S obzirom na jedinicu u kojoj se izražavaju, razlikujemo apsolutne (razmak varijacije, interkvartilna razlika, srednje apsolutno odstupanje, varijanca, standardna devijacija) mjere varijacije, koje se izražavaju u jedinici mjerenja i relativne mjere varijacije (koeficijent varijacije), koje se izražavaju u postotku. Izbor odgovarajuće mjere varijacije ovisi o tipu varijable.

I. Razmak varijacije (raspon) – RV

Razmak varijacije je apsolutna mjera varijacije koja se izražava u jedinicama mjere obilježja. Pomoću razmaka varijacije izražava se stupanj raspršenosti numeričke varijable, a sama vrijednost razmaka varijacije predstavlja razliku između najviše i najniže vrijednosti obilježja u statističkoj seriji.

Ako u nizu nema raspršenosti (disperzije), razmak varijacije je jednak nuli, a povećanjem raspršenosti povećava se i raspon varijacije.

Za negrupirane podatke raspon varijacije računa se prema formuli:

$$RV = x_{max} - x_{min}$$

gdje je:

x_{max} – najveća vrijednost mjerenja

x_{min} – najmanja vrijednost mjerenja.

Kod podataka koji su grupirani (svrstani u razrede) raspon varijacije procjenjuje se na temelju razlike gornje granice posljednjeg razreda i donje granice prvog razreda, što se smatra nepouzdanim.

S obzirom na to da se prilikom računanja raspona u obzir uzimaju samo dvije (ekstremne) vrijednosti mjerenja, raspon se smatra nepotpunom mjerom varijacije.

Primjer 1.

Izračunavanje razmaka varijacije

U nastavku su prikazane vrijednosti za proteklih 15 godina iz kojih se vidi broj dana u kojima je temperatura tla na dubini od 5 cm bila niža od -10° . Potrebno je izračunati razmak varijacije!

X_i : 2 2 2 2 3 3 3 4 5 5 5 10 11 14 14

$RV = x_{max} - x_{min} = 26 - 2 = 24$

Prema tome, maksimalna razlika između ispitivanih godina u broju dana u kojima je temperatura tla na dubini od 5 cm bila niža od 10 °C iznosi 24 dana.

II. Interkvartilna razlika (I_{QR})

Da bismo razumjeli i znali izračunati interkvartilnu razliku, prije toga se trebamo upoznati s pojmom kvantila.

Kvantili su vrijednosti brojčanog obilježja koji statistički niz uređen po veličini dijele na q jednakih dijelova. Ovisno o tome na koliko dijelova kvantili dijele po veličini uređen statistički niz, razlikujemo:

- **percentile –** dijele uređeni statistički niz na sto jednakih dijelova
- **decile –** dijele uređeni statistički niz na deset jednakih dijelova
- **kvartile** – dijele uređeni statistički niz na četiri jednaka dijela.

i. Percentili

- dijele uređen statistički niz na sto jednakih dijelova
- računaju se na temelju formule:

$$i = \frac{P}{100}(n+1)$$

gdje je:

i – pozicija traženog percentila

P – traženi percentil

n – ukupan broj rezultata u distribuciji (veličina uzorka).

Primjer 2.

Izračunavanje percentila

Na temelju vrijednosti prikazanih u uređenoj statističkoj seriji, odredi 36. percentil!

1	4	4	4	6	7	7	9	11	11	11	11	12	15	16
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.

Mjesto u uređenoj statističkoj seriji

$$i = \frac{P}{100}(n+1) = \frac{36}{100}(15+1) = 0{,}36\ x\ 16 = 5{,}76$$

36. percentil nalazi se na 5,76. mjestu, odnosno **između 5. i 6. mjesta** u uređenoj statističkoj seriji.

- da bismo točno znali koja je to vrijednost, odnosno da bismo mogli odrediti koliko je vrijednosti niže ili više od 36. percentila, moramo izračunati koja se to vrijednost nalazi između 5. i 6. mjesta u uređenoj statističkoj seriji

$$P_{36} = (5.\ \text{mjesto} + 6.\ \text{mjesto})/2 = (6+7)/2 = 6{,}5$$

ili

$$P_{36} = 6 + 0{,}76\ (7-6) = 6 + 0{,}76 = 6{,}76$$

- prema tome, 36. percentil nalazi se između 5. i 6. mjesta u uređenoj statističkoj seriji i iznosi 6,5
- odnosno 36 % vrijednosti u nizu je niže od 6,5, dok je 64 % vrijednosti u nizu više od 6,5.

ii. Kvartili i interkvartilna razlika

Kvartili su kvantili koji dijele uređeni statistički niz na četiri jednaka dijela, pa prema tome razlikujemo tri kvartila: prvi ili donji kvartil, drugi ili srednji kvartil i treći ili gornji kvartil.

1. *prvi ili donji kvartil* – dijeli uređenu statističku seriju na dva dijela, s tim da je 25 % vrijednosti u uređenoj statističkoj seriji niže ili jednako prvom kvartilu, dok je 75 % vrijednosti u statističkoj seriji više od vrijednosti prvoga kvartila, a računa se na temelju izraza:

$$Q_1 = \frac{n+1}{4}$$

gdje je:

Q_1 – mjesto u uređenoj statističkoj seriji na kojem se nalazi prvi kvartil

n – ukupan broj članova serije.

2. *drugi ili srednji kvartil* dijeli uređenu statističku seriju na dva jednaka dijela, s tim da je 50 % vrijednosti u uređenoj statističkoj seriji niže, a 50 % vrijednosti više od vrijednosti drugog kvartila – prema tome, drugi kvartil je *medijana*

$$Q_2 = \frac{2(n+1)}{4}$$

gdje je:

Q_2 – mjesto u uređenoj statističkoj seriji na kojem se nalazi prvi kvartil

n – ukupan broj članova serije.

3. *treći ili gornji kvartil* dijeli uređenu statističku seriju na dva dijela, s tim da je 75 % vrijednosti u uređenoj statističkoj seriji niže ili jednako trećem kvartilu, dok je 25 % vrijednosti u statističkoj seriji više od vrijednosti trećega kvartila

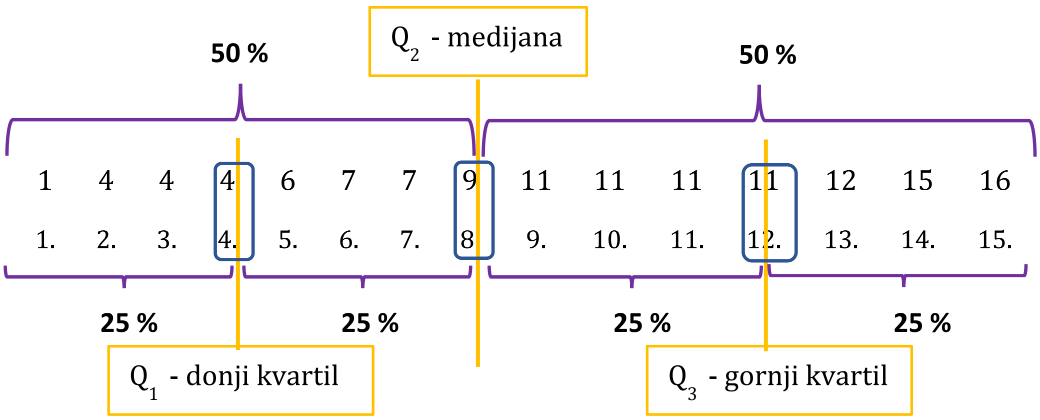
$$Q_3 = \frac{3\,(n+1)}{4}$$

gdje je:

Q_3 – mjesto u uređenoj statističkoj seriji na kojem se nalazi prvi kvartil

n – ukupan broj članova serije.

Grafički prikaz kvartila slijedi na Slici 1.



Slika 1. Položaj donjeg, srednjeg i gornjega kvartila u uređenoj statističkoj seriji

Primjer 3.

Računanje kvartila i interkvartilne razlike

Na temelju sljedećeg niza izračunati donji i gornji kvartil!

1	4	4	4	6	7	7	9	11	11	11	11	12	15	16
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.

Donji ili prvi kvartil

$$Q_1 = \frac{n+1}{4} = \frac{15+1}{4} = 4$$

Prvi kvartil nalazi se na 4. mjestu u uređenoj statističkoj seriji. Na četvrtome mjestu u uređenoj statističkoj seriji nalazi se broj 4, prema tome 25 % vrijednosti u nizu je niže od 4, dok je 75 % izmjerenih vrijednosti veće od 4.

Gornji ili treći kvartil

$$Q_3 = \frac{3\,(n+1)}{4} = \frac{3\,(15+1)}{4} = \frac{48}{4} = 12$$

Treći kvartil nalazi se na 12. mjestu u uređenoj statističkoj seriji. Prema tome, 75 % vrijednosti u nizu je niže od 11, dok je 25 % izmjerenih vrijednosti veće od 11.

Vrijednosti prvog i trećega kvartila potrebne su za izračunavanje **interkvartilne razlike.**

Interkvartilna razlika je razlika između trećeg i prvoga kvartila te obuhvaća 50 % središnjih vrijednosti, odnosno prikazuje varijabilnost 50 % središnjih vrijednosti. Izražava se u mjernim jedinicama ispitivane varijable, a izračunava se na temelju izraza:

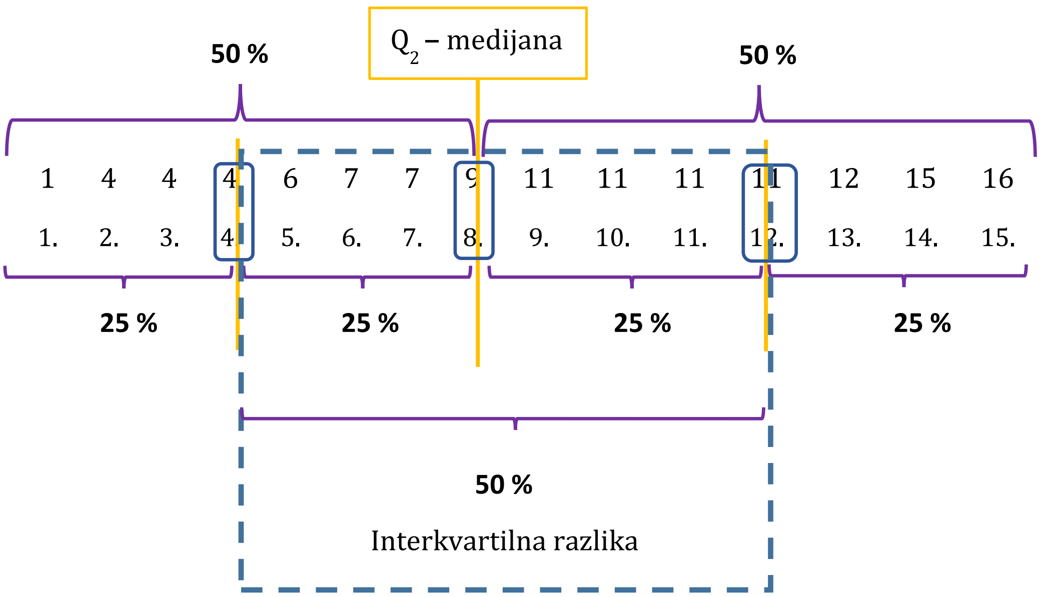
$$I_{QR} = Q_3 - Q_1$$

gdje je:

I_{QR} – interkvartilna razlika

Q_3 – treći kvartil

Q_1 – prvi kvartil.



Slika 2. Prikaz interkvartilne razlike u uređenoj statističkoj seriji

$$I_{QR} = Q_3 - Q_1 = 11 - 4 = 7$$

Na temelju interkvartilne razlike zaključujemo da se 50 % vrijednosti u nizu nalazi u rasponu od 7 mjernih jedinica.

Usporedbom interkvartilne razlike i raspona, koji u ovom slučaju iznosi 15, možemo zaključiti da je interkvartilna razlika mjera varijacije koju je dobro proučiti uz raspon, jer se prilikom njezina računanja izostavljaju krajnje – ekstremne vrijednosti korištene prilikom računanja raspona. Uz to, treba imati na umu da se za računanje interkvartilne razlike u obzir uzimaju samo dvije vrijednosti iz niza te prema tome ni interkvartilna razlika, kao ni raspon nije potpuna mjera varijacije.

Osim interkvartilne razlike, koja je apsolutna mjera varijacije, možemo računati i **koeficijent kvartilne devijacije**, koji je relativni pokazatelj varijacije, a također se računa na temelju prvog i trećega kvartila.
S obzirom na to da je koeficijent kvartilne devijacije relativna mjera varijabiliteta, omogućuje usporedbu varijabli izraženih u različitim mjernim jedinicama. Računa se na temelju izraza:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Za prethodni primjer možemo izračunati koeficijent kvartilne devijacije, kako slijedi:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{11 - 4}{11 + 4} = \frac{7}{15} = 0{,}47$$

Koeficijent kvartilne devijacije je 0,47, odnosno 47 %.

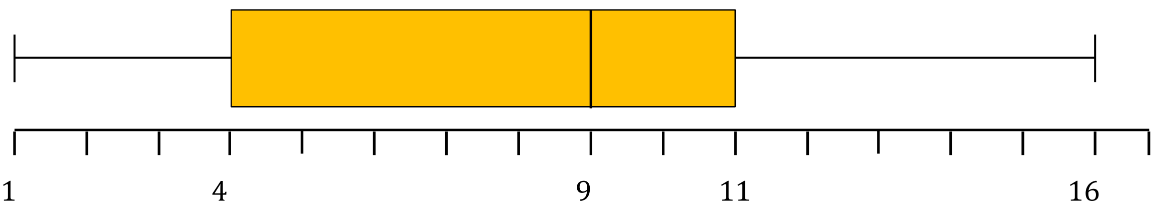
Za grafičko prikazivanje raspršenosti vrijednosti koristi se **dijagram pravokutnika** (*box*-plot odnosno *box i whisker* plot dijagram – B-P dijagram).

Prilikom prikazivanja, B-P dijagram može biti položen okomito ili vodoravno. Sastoji se od pravokutnika, čija je duljina određena interkvartilnim rasponom. Odnosno, donja linija pravokutnika označava prvi, a gornja linija treći kvartil. Vodoravna linija unutar pravokutnika označava medijanu. Uz medijanu unutar pravokutnika može biti prikazana i aritmetička sredina, koja se onda prikazuje određenim simbolom. Iz pravokutnika (kutije – *box*) na obje strane izlazi „brk“ (*whisker*) koji najčešće označavaju:

- minimum i maksimum ispitivane serije
- unutarnje međe – najniža vrijednost koja se nalazi unutar $Q_1 - 1{,}5\ I_{QR}$ odnosno najviša vrijednost koja se nalazi unutar $Q_3 + 1{,}5\ I_{QR}$
- vanjske međe – najniža vrijednost koja se nalazi unutar $Q_1 - 3\ I_{QR}$ odnosno najviša vrijednost koja se nalazi unutar $Q_3 + 3\ I_{QR}$
- 9. percentil i 91. percentil
- 2. percentil i 98. percentil.

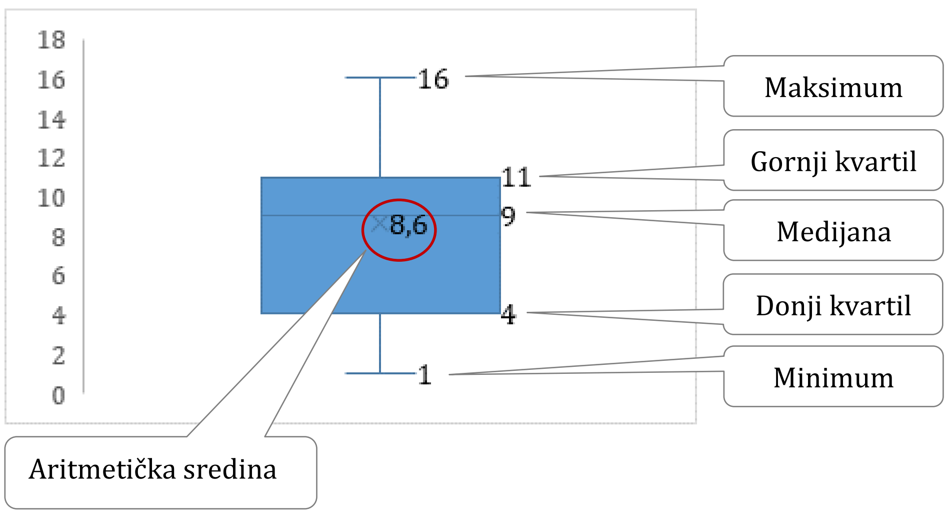
U nastavku su prikazani B-P dijagrami za podatke iz Primjera 2. (stranica 50), nacrtani s *whiskerima* koji prikazuju:

11. *Box*-plot dijagram, *whiskeri* prikazuju minimalnu i maksimalnu vrijednost ispitivane serije podataka

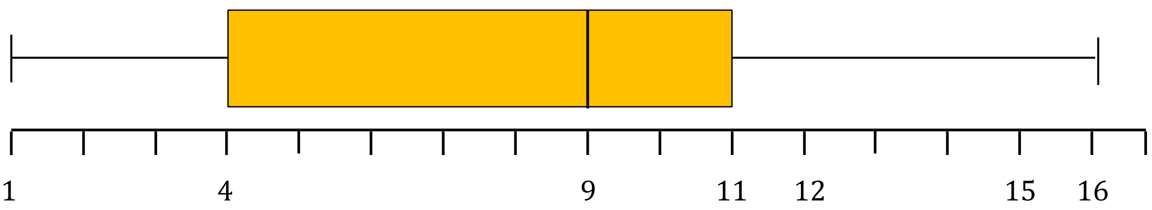


- minimum – 1
- prvi kvartil – 4
- medijana – 9
- treći kvartil – 11
- maksimum – 16.

Isti dijagram napravljen pomoću programa Excel izgleda ovako:



12. *Box*-plot dijagram, *whiskeri* prikazuju raspon koji obuhvaća vrijednosti koje se nalaze unutar unutarnje međe



- prvi kvartil – 4
- medijana – 9

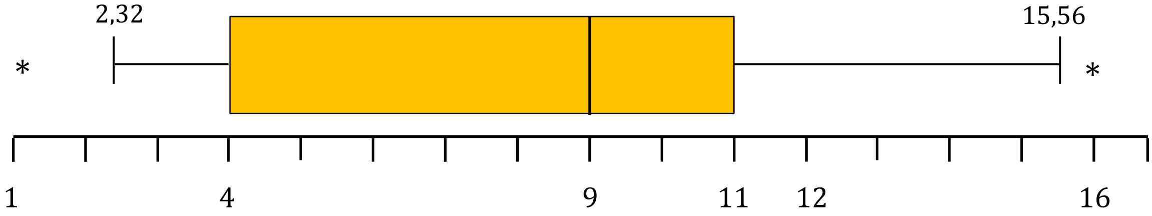
- treći kvartil – 11.

Donja granica unutarnje međe (Q_1 – 1,5 I_{QR}) je -6,5.
S obzirom na to da je najniža vrijednost mjerenja 1 (minimum), lijevi „brk" će sezati do spomenute minimalne vrijednosti.
Gornja granica unutarnje međe iznosi 21,5 a desni „brk" će biti nacrtan do vrijednosti 16, jer je to najviša vrijednost mjerenja ispitivanog svojstva.

Vrijednosti koje se nalaze između unutarnje i vanjske međe označavaju se istim znakom, npr. zvjezdicama, dok se vrijednosti koje su smještene izvan vanjskih međa označavaju drugačijim simbolima (npr. kružićima), da bi se mogle razlikovati od vrijednosti koje su unutar vanjskih, ali izvan unutarnjih međa.
Za ovaj niz podataka donja vanjska međa iznosi -17, a gornja vanjska međa 32. Da smo, recimo, u nizu vrijednosti imali mjerenja koja su veća od 32, tada bi te vrijednosti u B-P dijagramu bile označene kružićem (o) i te vrijednosti predstavljaju **atipične vrijednosti –** ***outliere***.

Outlieri su vrijednosti koje se izdvajaju u seriji podataka i treba obratiti posebnu pažnju na njih. Potrebno je utvrditi radi li se o pogrešno upisanim podacima, o podacima koji pripadaju nekoj drugoj populaciji ili o podacima koji pripadaju ispitivanoj populaciji, ali su iz nekog razloga različiti. Vrijednosti koje se u B-P dijagramu nalaze između unutarnjih i vanjskih međa su „potencijalni" *outlieri* te na njih također treba obratiti pažnju.

13. *Box*-plot dijagram, *whiskeri* prikazuju raspon koji obuhvaća vrijednosti koje se nalaze između 9. i 91. percentila



- prvi kvartil – 4
- medijana – 9
- drugi kvartil – 11
- 9. percentil – 2,32
- 91. percentil – 15,56.

III. Varijanca, standardna devijacija i koeficijent varijacije

i. Varijanca i standardna devijacija

Varijanca i standardna devijacija su najkorištenije mjere varijacije. Smatraju se potpunim mjerama varijacije jer se prilikom njihova računanja u obzir uzimaju sve vrijednosti obilježja.

Varijanca i standardna devijacija ostaju nepromijenjene ako se sve vrijednosti obilježja promijene za jedan stalni iznos (Tablica 18., Primjer 2.).

Ako se sve vrijednosti obilježja uvećaju ili umanje za određeni iznos, za isti će se iznos povećati ili smanjiti varijanca i standardna devijacija (Tablica18., Primjer 3.)

Tablica 18. Primjeri aritmetičke sredine, standardne devijacije i koeficijenta varijacije

	Primjer 1	Primjer 2	Primjer 3
	2	2 + 2 = 4	2 x 2 = 4
	3	3 + 2 = 5	3 x 2 = 6
	4	4 + 2 = 6	4 x 2 = 8
	3	3 + 2 = 5	3 x 2 = 6
	4	4 + 2 = 6	4 x 2 = 8
	3	3 + 2 = 5	3 x 2 = 6
	2	2 + 2 = 4	2 x 2 = 4
	4	4 + 2 = 6	4 x 2 = 8
	3	3 + 2 =5	3 x 2 =6
	5	5 + 2 = 7	5 x 2 = 10
Aritmetička sredina	3,30	5,30	6,60
Standardna devijacija	0,95	0,95	1,90
Koeficijent varijacije	28,75	17,90	28,75

Varijanca je prosjek kvadrata odstupanja vrijednosti obilježja od njegove aritmetičke sredine te se za pojedinačne vrijednosti računa prema formuli:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

- a za distribuciju frekvencija

$$\sigma^2 = \frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i}$$

S obzirom na to da je izražena u kvadratima odstupanja od aritmetičke sredine, varijanca nije najpogodnija mjera za interpretaciju te se stoga raspršenost vrijednosti obilježja najčešće izražava pomoću standardne devijacije.

Standardna devijacija je izražena u jedinici mjerenja obilježja, a predstavlja drugi korijen iz varijance te se za pojedinačne vrijednosti obilježja računa prema formuli:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

- a za distribuciju frekvencija

$$\sigma = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i}}$$

Prilikom računanja standardne devijacije u obzir se uzimaju sve vrijednosti obilježja te se ona smatra potpunom mjerom varijacije. S obzirom na to da se prilikom njezina računanja u obzir uzimaju sve vrijednosti iz statističke serije, jako je osjetljiva na ekstremne vrijednosti i *outliere*. Ako se na temelju varijance ili standardne devijacije izračunane na temelju vrijednosti uzorka želi procijeniti varijabilnost obilježja u populaciji, potrebno je napraviti nepristranu procjenu, što znači da se u nazivniku formule za računanje varijance odnosno standardne devijacije od ukupne veličine uzorka (n) oduzme 1.

Za usporedbu varijabilnosti raznoimenih vrijednosti obilježja koristi se koeficijent varijacije.

ii. <u>Koeficijent varijacije</u>

Koeficijent varijacije je relativna mjera varijacije. Koristi se za usporedbu varijabilnosti vrijednosti obilježja različitih skupova.

Izračunava se prema formuli:

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

Primjer 4.

Izračunavanje varijance, standardne devijacije i koeficijenta varijacije za pojedinačne vrijednosti obilježja

Na temelju sljedećeg niza izračunati varijancu, standardnu devijaciju i koeficijent varijacije!

1	4	4	4	6	7	7	9	11	11	11	11	12	15	16
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.

Tablica 19. Koraci u izračunavanju standardne devijacije

Redni broj	Vrijednost obilježja (x_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1.	1	= 1 – 8,6 = -7,6	= $(-7,6)^2$ = 57,76
2.	4	= 4 – 8,6 = -4,6	= $(-4,6)^2$ = 21,16
3.	4	= 4 – 8,6 = -4,6	= $(-4,6)^2$ = 21,16
4.	4	-4,6	21,16
5.	6	-2,6	6,76
6.	7	-1,6	2,56
7.	7	-1,6	2,56
8.	9	0,4	0,16
9.	11	2,4	5,76
10.	11	2,4	5,76
11.	11	2,4	5,76
12.	11	2,4	5,76
13.	12	3,4	11,56
14.	15	6,4	40,96
15.	16	7,4	54,76
Suma	129	0	$\boxed{263,6}$

Ovo je $\sum(x_i - \bar{x})^2$ i uvrštava se u brojnik formule za standardnu devijaciju.

- prvo je potrebno izračunati aritmetičku sredinu na temelju formule za jednostavnu aritmetičku sredinu

$$\bar{x} = \frac{\sum x_i}{n} = \frac{129}{15} = 8,6$$

- nakon toga se izračunava razlika između svake vrijednosti obilježja promatranja i aritmetičke sredine $(x_i - \bar{x})$, kako bismo utvrdili odstupanje svake vrijednosti mjerenja od aritmetičke sredine (treći stupac u Tablici 19.)

- kvadriranjem tako izračunanih vrijednosti te izračunavanjem njihove sume dobijemo vrijednost koja odgovara izrazu $\sum(x_i - \bar{x})^2$ (četvrti stupac u Tablici 19.) te se ta vrijednost uvrštava u brojnik formule za standardnu devijaciju

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{263}{15}} = \sqrt{17{,}53} = 4{,}19$$

Kada imamo izračunanu aritmetičku sredinu i standardnu devijaciju, tada rezultat možemo izraziti u obliku aritmetička sredina $\pm$ standardna devijacija, što bi u ovom primjeru bilo 8,6 $\pm$ 4,19.

Standardna devijacija je 4,19, što znači da je prosječno odstupanje vrijednosti obilježja od aritmetičke sredine 4,19 mjernih jedinica.

Izračunavanje varijance!

Varijanca se računa na temelju iste formule kao i standardna devijacija samo što se ne korjenuje cijeli izraz.

$$\sigma = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{263}{15} = 17{,}53$$

Ako već imamo izračunanu standardnu devijaciju, onda varijancu možemo izračunati tako da vrijednost standardne devijacije kvadriramo, npr.

$$\sigma = 4{,}19$$
$$\sigma^2 = 4{,}19^2 = 17{,}53$$

ii. Koeficijent varijacije

Koeficijent varijacije je relativna mjera varijacije i koristi se za usporedbu varijabilnosti različitih obilježja.

Koeficijent varijacije je postotni omjer standardne devijacije i aritmetičke sredine te prema tome za njegovo izračunavanje moramo imati izračunanu aritmetičku sredinu i standardnu devijaciju, koje se onda uvrštavaju u sljedeću formulu:

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{4,19}{8,6}\, x\, 100 = 48,72$$

Prema tome, u ispitivanom uzorku koeficijent varijacije iznosi 48,72 %.

Primjer 5.

Izračunavanje varijance, standardne devijacije i koeficijenta varijacije za raspodjelu učestalosti

Na temelju vrijednosti obilježja grupiranih u razrede izračunajte varijancu, standardnu devijaciju i koeficijent varijance!

Tablica 20. Primjer i koraci u izračunavanju složene standardne devijacije i varijance

Razredi	**Učestalost (f_i)**	**Sredina razreda (x_i)**	**Podtotali ($f_i x_i$)**	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
18 – 22	2	20	40	-26,6	707,56	1415,12
23 – 27	3	25	75	-21,6	466,56	1399,68
28 – 32	5	30	150	-16,6	275,56	1377,80
33 – 37	17	35	595	-11,6	134,56	2287,52
38 – 42	37	40	1480	-6,6	43,56	1611,72
43 – 47	37	45	1665	-1,6	2,56	94,72
48 – 52	48	50	2400	3,4	11,56	554,88
53 – 57	31	55	1705	8,4	70,56	2187,36
58 – 62	18	60	1080	13,4	179,56	3232,08
63 – 67	2	65	130	18,4	338,56	677,12
Ukupno	200		9320		2230,6	14838,00

Izračunavanje aritmetičke sredine

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{9320}{200} = 46,6 \text{ zrna po klasu}$$

Formulom za ponderiranu aritmetičku sredinu izračunana je prosječna masa zrna po klasu, koja iznosi 46,6 zrna po klasu.

Nakon toga od svake sredine razreda (x_i) oduzima se vrijednost aritmetičke sredine $(x_i - \bar{x})$ te se ta vrijednost kvadrira i pomnoži s pripadajućom učestalošću. Suma tih vrijednosti uvrštava se u brojnik formule za složenu standardnu devijaciju.

Izračunavanje standardne devijacije

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} = \sqrt{\frac{14838,00}{200}} = \sqrt{74,19} = 8,61$$

Standardna devijacija iznosi 8,61 zrno po klasu te ona predstavlja prosječno odstupanje vrijednosti u statističkoj seriji od aritmetičke sredine.

Izračunavanje varijance

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i} = \frac{14838{,}00}{200} = 74{,}19$$

Izračunavanje koeficijenta varijacije

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{8{,}61}{46{,}6} x\ 100 = 18{,}48\ \%$$

U ispitivanom uzorku prosječna aritmetička sredina iznosi 46,6 $\pm$ 8,61 standardna devijacija. Relativna varijabilnost uzorka iznosi 18,48 %.

RIJEŠI ZADATKE I ODGOVORI NA PITANJA!

1. Odgovori na sljedeća pitanja!

a)	standardna devijacija osjetljiva je na ekstremne vrijednosti	Da	Ne
b)	dva niza podataka koja imaju jednaku aritmetičku sredinu imat će i jednaku standardnu devijaciju	Da	Ne
c)	dva niza podataka mogu imati istu varijancu i standardnu devijaciju	Da	Ne
d)	varijancu je lakše interpretirati nego standardnu devijaciju	Da	Ne
e)	koeficijent varijacije izražava se u jedinici mjerenja obilježja	Da	Ne

2. U kojem od sljedećih nizova podataka nema varijabilnosti?

a) 3, 6, 7, 7, 7
b) 0, 1, 0, 1, 0
c) 4, 4, 4, 4, 4
d) 1, 1, 1, 1, 0

3. U kojem od sljedećih nizova podataka je najveća varijabilnost?

a) 2, 4, 6, 8, 10
b) 1, 3, 5, 7, 9
c) 2, 4, 5, 4, 13
d) 2, 4, 5, 4, 8

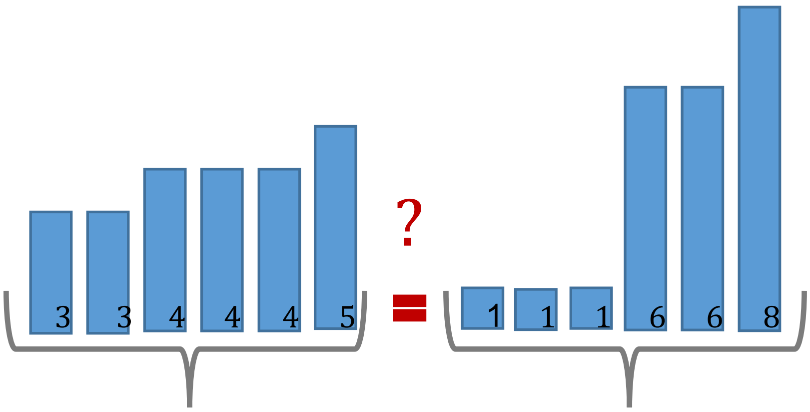
4. U koja dva niza je ista standardna devijacija?

1. 2, 2, 3, 5, 6
2. 1, 4, 6, 10, 12
3. 3, 3, 4, 6, 7
4. 3, 4, 5, 6, 7

Rješenje zadataka je na 71. stranici!

5. Na temelju podataka prikazanih u sljedećem grafičkom prikazu, komentirajte vrijednosti mjera varijacije u dva seta podataka. Što možemo zaključiti na temelju tih vrijednosti?

Statistička serija A Statistička serija B



	Statistička serija A	Statistička serija B
Veličina uzorka	6	6
Aritmetička sredina	3,83	3,83
Mod	4	1
Medijana	4	3,5
Raspon	2	7
Standardna devijacija	0,75	3,19
Koeficijent varijacije	19,65	83,25

Rješenje zadataka je na 71. stranici!

6. Proveden je poljski pokus u kojemu je utvrđena duljina klasa u 200 sorata ozime pšenice. Prikupljeni podaci grupirani su u šest razreda. Na temelju prikazanih podataka potrebno je izračunati standardnu devijaciju, varijancu i koeficijent varijacije.

Razredi	Učestalost (f_i)
5,20 – 5,80	5
5,81 – 6,40	21
6,41 – 7,00	102
7,01 – 7,60	48
7,61 – 8,20	23
8,21 – 8,80	1
Ukupno	200

a) izračunavanje varijance i standardne devijacije

b) izračunavanje koeficijenta varijacije

Rješenje zadataka je na 71. stranici!

7. Od 2000. do 2009. godine praćen je prosječan prinos pšenice (t/ha) u Osječko-baranjskoj županiji. Izračunajte koeficijent varijacije!

Godina	Urod (t/ha)
2007.	3,3
2008.	4,7
2009.	4,4
2010.	4,6
2011.	3,2
2012.	4,9
2013.	4,1
2014.	4,6
2015.	4,6
2016.	5,5

a) koeficijent varijacije

Rješenje zadataka je na 72. stranici!

8. Proveden je poljski pokus u kojemu je jedan od ciljeva istraživanja bio utvrditi broj zrna po klasu ispitivanih sorata pšenice. U istraživanje je bilo uključeno 200 sorata pšenice. Na temelju podataka prikazanih u tablici treba utvrditi kolika je standardna devijacija. Na 27. stranici je izračunana aritmetička sredina za isti set podataka.

Razredi	Učestalost (f_i)
18 – 22	2
23 – 27	3
28 – 32	5
33 – 37	17
38 – 42	37
43 – 47	37
48 – 52	48
53 – 57	31
58 – 62	18
63 – 67	2
Ukupno	200

a) izračunavanje standardne devijacije

Razredi	Učestalost (f_i)					
18 – 22	2					
23 – 27	3					
28 – 32	5					
33 – 37	17					
38 – 42	37					
43 – 47	37					
48 – 52	48					
53 – 57	31					
58 – 62	18					
63 – 67	2					
Ukupno	200					

Rješenje zadataka je na 72. stranici!

9. Pšenica je uz kukuruz najvažnija žitarica u Republici Hrvatskoj. Podatke o proizvodnji pšenice u Republici Hrvatskoj u razdoblju od 1986. do 2010. godine grupirati u razrede veličine 300 000 t te izračunati standardnu devijaciju i koeficijent varijacije.

Godina	Ukupna proizvodnja pšenice u tis. tona
2003.	506
1999.	558
2005.	602
1992.	658
2010.	681
1996.	741
1994.	750
2004.	801
2006.	805
2001.	812
2007.	812
2002.	823
1997.	834
2008.	858
2000.	865
1995.	877
1993.	887
2009.	936
1998.	1020
1986.	1078
1987.	1274
1989.	1288
1988.	1434
1990.	1602
1991.	1996

Razredi	Učestalost (f_i)					
Ukupno						

Rješenje zadataka je na 72. stranici!

5. Rješenja

1. Podjela statističkih metoda i vrste obilježja

1. zadatak - a) da, b) da, c) ne, d) da, e) da
2. zadatak - a) kontinuirana, b) diskontinuirana, c) ordinalna, d) nominalna, e) nominalna, f) ordinalna, g) ordinalna, h) kontinuirana, i) kontinuirana, j) diskontinuirana, k) kontinuirana

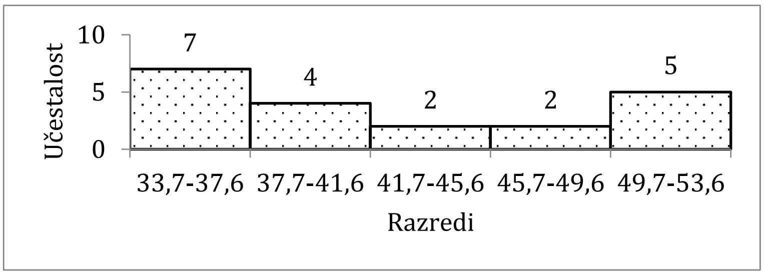
2. Grupiranje podataka, apsolutna i relativna učestalost

1. zadatak

Razredi	Učestalost (fi)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
33,7 – 37,6	7	35	7	20	35	100
37,7 – 41,6	4	20	11	13	55	65
41,7 – 45,6	2	10	13	9	65	45
45,7 – 49,6	2	10	15	7	75	35
49,7 - 53,6	5	25	20	5	100	25
Ukupno	20	100 %				

e) histogram apsolutne raspodjele učestalosti

f) 13 od ukupno 20 ispitivanih sorata ima duljinu stabljike manju od 45,7 cm.



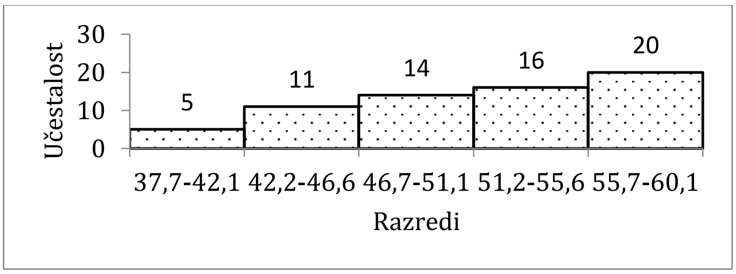
2. zadatak

Razredi	Učestalost (fi)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
37,7 – 42,1	5	25	5	20	25	100
42,2 – 46,6	6	30	11	15	55	75
46,7 – 51,1	3	15	14	9	70	45
51,2 – 55,6	2	10	16	6	80	30
55,7 – 60,1	4	20	20	4	100	20
Ukupno	20	100,00 %				

d) histogram kumulativne raspodjele učestalosti

e) dvije sorte pšenice imaju visinu biljke višu od 51,1 i nižu od 55,7 cm

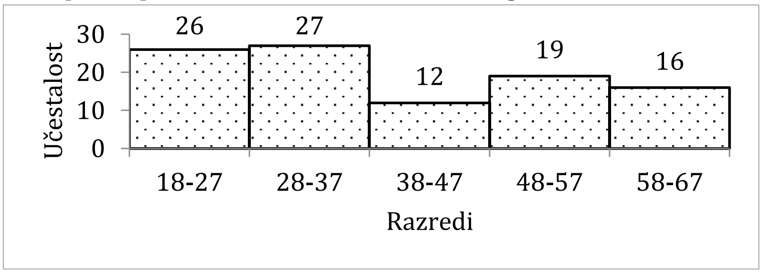
f) u ispitivanom uzorku 80 % ispitivanih sorata ima visinu biljke nižu od 55,7 cm



3. zadatak

Razredi	Učestalost (fi)	Relativna učestalost (%)	Apsolutna kumulacija		Relativna kumulacija	
			Ispod	Iznad	Ispod	Iznad
18 – 27	26	26	26	100	26	100
28 – 37	27	27	53	74	53	74
38 – 47	12	12	65	47	65	47
48 – 57	19	19	84	35	84	35
58 – 67	16	16	100	16	100	16
Ukupno	100	100%				

a) veličina uzorka n = 100
b) ispitivana varijabla je starosna dob ispitanika izražena u godinama
c) jedinica promatranja je jedan ispitanik, individua
d) populacija je 100 ispitanika starosne dobi 18 do 65 godina, koji žive na području Osječko-baranjske županije
e) u ispitivanom uzorku najzastupljenija je skupina ispitanika starosne dobi 28 do 37 godina i to s 27 %
f) u ispitivanju su bili najmanje zastupljeni ispitanici dobne skupine 38 do 47 godina i to s 12 %
g) u istraživanje je bilo uključeno 19 % ispitanika starijih od 48, a mlađih od 58 godina
h) grafički prikazati apsolutnu raspodjelu učestalosti ispitanika po dobi



3. Mjere centralne tendencije – srednje vrijednosti

1. zadatak – a) da, b) da, c) da
2. zadatak – b) medijan
3. zadatak – a) aritmetička sredina je 1105,75
4. zadatak - a) nominalna varijabla, b) mod
5. zadatak - radi se o uzorcima iste veličine (n = 6), iste aritmetičke sredine (3,83), ali različitih vrijednosti moda i medijane. U statističkoj seriji A, aritmetička sredina, mod i medijana su sličnih vrijednosti što ukazuje na moguću normalnu raspodjelu podataka. U statističkoj seriji B medijana je također približno slična aritmetičkoj sredini, dok je mod različit i iznosi 1. Mod pokazuje da je u drugom uzorku najčešće postignuta vrijednost mjerenja 1, a različitost aritmetičke sredine, medijane i moda ukazuje na to da ispitivana varijabla vjerojatno ne slijedi normalnu raspodjelu podataka.

6. zadatak

a) aritmetička sredina, odnosno prosječna masa klasa ispitivanog uzorka je 1,46 g
b) veličina uzorka, n = 9
c) ispitivana varijabla je masa klasa izražena u gramima, a jedinica promatranja je sorta pšenice
d) u ispitivanom uzorku je udio sorata s masom klasa nižom od 1 g 11 %

7. zadatak

Raspon plaća	Učestalost (f_i)	Sredina razreda (x_i)	Podtotali ($f_i x_i$)
2300,00 – 3125,00	3	2712,50	8137,50
3200.00 – 4100.00	9	3650.00	32850.00
4450,00 – 6000,00	4	5225,00	20900,00
6350,00 – 8250,00	2	7300,00	14600,00
Ukupno	18		76487,50

$$\bar{x} = \frac{70487,5}{18} = 4249,31$$

Prosječna plaća u firmi iznosi 4249,31 kn mjesečno.

8. zadatak

a) koliki je prosječan broj napunjenih posudica po radniku, u prvoj smjeni?

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{4}{\frac{1}{400} + \frac{1}{450} + \frac{1}{300} + \frac{1}{370}} =$$

$$= \frac{4}{0,0025 + 0,0022 + 0,0033 + 0,0027} = \frac{4}{0,0107} = 373,82$$

U prosjeku, u prvoj smjeni jedan radnik napuni 373,82 posudice jagodama tijekom radnog dana.

b) koliko je prosječno vrijeme (min) punjenja jedne posudice jagodama u prvoj smjeni?

- 1. radnik: 8 sati rada 400 posudica
  8 sati x 60 min = 480 minuta/400 posudica = 1,20 minuta po posudici
- 2. radnik: 8 sati rada 450 posudica
  8 sati x 60 min = 480 minuta/450 posudica = 1,07 minuta po posudici
- 3. radnik: 8 sati rada 300 posudica
  8 sati x 60 min = 480 minuta/300 posudica = 1,60 minuta po posudici
- 4. radnik: 8 sati rada 370 posudica
  8 sati x 60 min = 480 minuta/370 posudica = 1,29 minuta po posudici

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{4}{\frac{1}{1,20} + \frac{1}{1,07} + \frac{1}{1,60} + \frac{1}{1,29}} = \frac{4}{5,16} = 0,78$$

U prvoj smjeni, prosječno vrijeme punjenja jedne posudice jagodama je 0,78 minuta.

9. zadatak

a) koliki je prosječan broj napunjenih posudica po radniku, u drugoj smjeni?

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{5}{\frac{1}{300} + \frac{1}{330} + \frac{1}{500} + \frac{1}{420} + \frac{1}{400}} = \frac{5}{0,0132} = 378,78$$

U prosjeku, u drugoj smjeni jedan radnik napuni 378,78 posudica jagodama tijekom jednog radnog dana.

b) koliko je prosječno vrijeme (min) punjenja jedne posudice jagodama?

- 1. radnik: 8 sati rada 300 posudica
  8 sati x 60 min = 480 minuta/300 posudica = 1,60 minuta po posudici
- 2. radnik: 8 sati rada 330 posudica
  8 sati x 60 min = 480 minuta/330 posudica = 1,45 minuta po posudici
- 3. radnik: 8 sati rada 500 posudica
  8 sati x 60 min = 480 minuta/500 posudica = 0,96 minuta po posudici
- 4. radnik: 8 sati rada 420 posudica
  8 sati x 60 min = 480 minuta/420 posudica = 1,14 minuta po posudici
- 5. radnik: 8 sati rada 400 posudica
  8 sati x 60 min = 480 minuta/400 posudica = 1,20 minuta po posudici

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{5}{\frac{1}{1{,}60} + \frac{1}{1{,}45} + \frac{1}{0{,}96} + \frac{1}{1{,}14} + \frac{1}{1{,}20}} = \frac{5}{6{,}35} = 0{,}79$$

U prvoj smjeni, prosječno vrijeme punjenja jedne posudice jagodama je 0,79 minuta.

c) koliko prosječno posudica napune jagodama <u>obje smjene zajedno</u> u jednom danu?

1. smjena: 400 + 450 + 300 + 370 = 1520 posudica
2. smjena: 300 + 330 + 500 + 420 + 400 = 1950 posudica

$$H = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{4+5}{\frac{4}{1520} + \frac{5}{1950}} = \frac{9}{0{,}052} = 1730{,}8$$

U jednom danu obje smjene zajedno u prosjeku jagodama napune 1730,8 posudica.

d) koliko je prosječno vrijeme (minuta) punjenja jedne posudice jagodama zajedno u obje smjene?

$$H = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{4+5}{\frac{4}{0{,}78} + \frac{5}{0{,}79}} = \frac{9}{11{,}46} = 0{,}79$$

Prosječno vrijeme potrebno za punjenje jedne posudice jagodama u obje smjene zajedno je 0,79 minuta.

10. zadatak

Razredi	Učestalost
18 – 22	2
23 – 27	3
28 – 32	5
33 – 37	17
38 – 42	37
43 – 47	37
48 – 52	48
53 – 57	31
58 – 62	18
63 – 67	2
Ukupno	200

$$M_o = 48 + \left[\frac{(48-37)}{(48-37)+(48-31)} \times 5\right] =$$

Vrijednost mjerenja koja je u ispitivanom uzorku najčešće puta postignuta je 49,96 zrna po klasu.

11. zadatak

a) aritmetička sredina je 6,13 t/ha
b) u ispitivanoj statističkoj seriji niti jedna vrijednost se ne ponavlja dva ili više puta, odnosno nema učestalosti pa samim time nema ni moda
c) medijana je 6,40 t/ ha i nalazi se na 6. mjestu u uređenoj statističkoj seriji

12. zadatak

Razredi	Učestalost (f_i)	Sredina razreda (x_i)	Podtotali	Apsolutna kumulacija ispod
506 – 754	7	630	4410	7
755 – 1003	11	879	9669	18
1004 – 1252	2	1128	2256	20
1253 – 1501	3	1377	4131	23
1502 – 1750	1	1626	1626	24
1751 – 1999	1	1875	1875	25
Ukupno	25		23967	7

a) aritmetička sredina izračunana na temelju jednostavne serije podataka je 939 920 t
b) moda nema
c) medijana izračunana na temelju jednostavne serije podataka je 834 000 t
d) složena aritmetička sredina je 958 680 t
e) mod izračunan na temelju grupiranih podataka iznosi 831 615 t
f) medijana izračunana na temelju grupiranih podataka iznosi 830 583 t

4. Mjere varijacije

1. zadatak – a) da, b) ne, c) da, d) ne, e) ne
2. zadatak – c) 4, 4, 4, 4, 4
3. zadatak – c) 2, 4, 5, 4, 13
4. zadatak – a) 2, 2, 3, 5, 6 i d) 3, 3, 4, 6, 7
5. zadatak – na temelju izračunanih vrijednosti razmaka varijacije, standardne devijacije i koeficijenta varijacije koji su mnogo veći u statističkoj seriji B, možemo zaključiti da je statistička serija B puno varijabilnija nego statistička serija A.

6. zadatak

Razredi	Učestalost (f_i)	Sredina razreda	Ponderi ($f_i x_i$)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
5,20 – 5,80	5	5,5	27,5	-1,4	1,96	9,8
5,81 – 6,40	21	6,1	128,1	-0,8	0,64	13,44
6,41 – 7,00	102	6,7	683,4	-0,2	0,04	4,08
7,01 – 7,60	48	7,3	350,4	0,4	0,16	7,68
7,61 – 8,20	23	7,9	181,7	1	1	23
8,21 – 8,80	1	8,5	8,5	1,6	2,56	2,56
Ukupno	200		1379,6			60,56

Aritmetička sredina duljine klasa u ispitivanih sorata pšenice iznosi 6,9 cm.

Standardna devijacija je 0,55 cm

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} = \sqrt{\frac{60,56}{200}} = 0,55$$

Varijanca je 0,30 cm

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i} = \frac{60,56}{200} = 0,30$$

Koeficijent varijacije je 7,97 %

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{0,55}{6,9} x\, 100 = 7,97\ \%$$

7. zadatak

Godina	Urod (t/ha)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
2000.	3,3	-1,09	1,19
2001.	4,7	0,31	0,10
2002.	4,4	0,01	0,00
2003.	4,6	0,21	0,04
2004.	3,2	-1,19	1,42
2005.	4,9	0,51	0,26
2006.	4,1	-0,29	0,08
2007.	4,6	0,21	0,04
2008.	4,6	0,21	0,04
2009.	5,5	1,11	1,23
Σ			4,41

$$\bar{x} = \frac{43,9}{10} = 4,39$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{4,41}{10}} = 0,66$$

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{0,66}{4,39} x\, 100 = 15,12\ \%$$

Koeficijent varijacije za prosječan urod pšenice (t/ha) u Osječko-baranjskoj županiji iznosi 15,12 %.

8. zadatak

Razredi	Učestalost (f_i)	Sredina razreda (x_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
18 – 22	2	20	-26,6	707,56	1415,12
23 – 27	3	25	-21,6	466,56	1399,68
28 – 32	5	30	-16,6	275,56	1377,80
33 – 37	17	35	-11,6	134,56	2287,52
38 – 42	37	40	-6,6	43,56	1611,72
43 – 47	37	45	-1,6	2,56	94,72
48 – 52	48	50	3,4	11,56	554,88
53 – 57	31	55	8,4	70,56	2187,36
58 – 62	18	60	13,4	179,56	3232,08
63 – 67	2	65	18,4	338,56	677,12
Ukupno	200				14838,00

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{14838,00}{200}} = 8,61$$

Standardna devijacija za svojstvo broj zrna po klasu iznosi 8,61 zrno.

9. Zadatak

Razredi	Učestalost (f_i)	Sredina razreda (x_i)	(x_i f_i)	($x_i - \bar{x}$)	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
506 – 805	9	655,5	5899,5	-300	90000	810000
806 – 1105	11	955,5	10510,5	0	0	0
1106 – 1405	2	1255,5	2511	300	90000	180000
1406 – 1705	2	1555,5	3111	600	360000	720000
1706 – 2005	1	1855,5	1855,5	900	810000	810000
Ukupno	25		23887,5			2520000

Standardna devijacija je 317 490 t.

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{2520000}{25}} = 317\ 490$$

Koeficijent varijacije je 33,23 %.

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{317\ 490}{955\ 500}\ x\ 100 = 33{,}23\ \%$$

6. Prilog

Zadaci za vježbu!

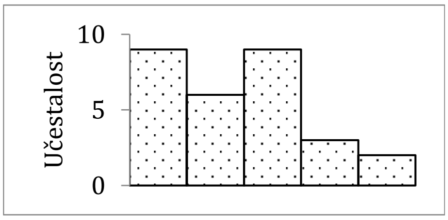
1. Proveden je poljski pokus u kojemu je na 29 sorata ozime pšenice ispitivana duljina klasa (cm).

a) na temelju izmjerenih vrijednosti duljine klasa potrebno je formirati razrede veličine 0,7 cm te izračunati aritmetičku sredinu, standardnu devijaciju, koeficijent varijacije, medijanu (na temelju negrupiranih i grupiranih podataka) te 38. percentil! Apsolutnu raspodjelu učestalosti grafički prikazati pomoću histograma!

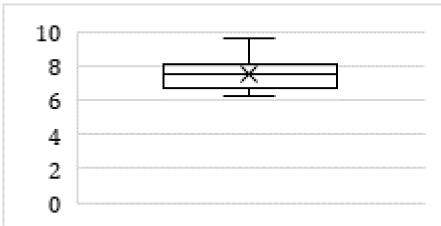
b) na temelju izmjerenih vrijednosti duljine klasa potrebno je formirati 6 razreda te izračunati aritmetičku sredinu, standardnu devijaciju, koeficijent varijacije, mod i interkvartilnu razliku! Ispitivane podatke grafički prikazati pomoću B-P dijagrama!

Sorta	Zemlja porijekla	Duljina klasa (cm)
Barbara	HR	6,25
Ana	HR	6,35
Demetra	HR	6,35
Libelula	IT	6,45
BC Elvira	HR	6,60
Lucija	HR	6,60
Divana	HR	6,70
Edison	AUT	6,80
Bastide	AUT	6,90
Anđelka	HR	7,05
Felix	HR	7,30
GK Kalasz	HUN	7,30
Lela	HR	7,35
Katarina	HR	7,50
Janica	HR	7,55
Adriana	HR	7,65
MV Emesse	HUN	7,70
Aida	HR	7,75
Golubica	HR	7,80
MV Magvas	HUN	7,80
Alka	HR	8,00
MV Magdalena	HUN	8,05
Bezostaja	HR	8,20
Ilirija	HR	8,20
Eurofit	AUT	8,80
Dekan	DEU	8,85
Ludwig	AUT	8,95
Eurojet	AUT	9,05
Antonius	AUT	9,60

a)
- aritmetička sredina – 7,58 cm
- standardna devijacija – 0,85 cm
- koeficijent varijacije – 11,24 %
- medijana (negrupirani) – 7,55 cm
- medijana (grupirani) – 7,50 cm
- 38. percentil – 7,3 cm



b)
- aritmetička sredina – 7,55 cm
- standardna devijacija – 0,87 cm
- koeficijent varijacije – 11,54 %
- mod (negrupirani) – 6,35, 6,60, 7,30, 7,80, 8,20,
- mod (grupirani) – 6,66 cm
- interkvartilna razlika – 1,37



2. Proveden je poljski pokus u kojemu je na 28 sorata ozime pšenice utvrđena masa klasa (g).

a) na temelju izmjerenih vrijednosti duljine klasa potrebno je formirati razrede veličine 0,5 g te izračunati aritmetičku sredinu, standardnu devijaciju, koeficijent varijacije, mod, medijanu (na temelju negrupiranih i grupiranih podataka) i 80. percentil

b) na temelju izmjerenih vrijednosti duljine klasa potrebno je formirati 4 razreda te izračunati aritmetičku sredinu, standardnu devijaciju, koeficijent varijacije, mod, raspon i 4. percentil

c) na temelju izmjerenih vrijednosti duljine klasa potrebno je formirati 5 razreda te izračunati aritmetičku sredinu, standardnu devijaciju, koeficijent varijacije, mod, raspon i 4. percentil

d) za obilježje „zemlja podrijetla" izračunajte mod te utvrdite koliki je udio hrvatskih sorata u ispitivanom uzorku

Sorta	Zemlja podrijetla	Masa klasa (g)
Njivka	HR	0,92
Zlatna dolina	HR	1,19
U1	HR	1,25
Super žitarka	HR	1,26
Slavonija	HR	1,44
SW Maxi	AUT	1,51
Osječka crvenka	HR	1,52
Valerius	AUT	1,54
Mihaela	HR	1,54
Žitarka	HR	1,58
Srpanjka	HR	1,68
Sana	HR	1,70
Seka	HR	1,70
MV Magdalena	HUN	1,71
Renan	FRA	1,72
MV Emesse	HUN	1,85
Zlata	HR	1,91
Lucija	HR	2,00
Pipi	HR	2,00
Panonija	HR	2,10
Soissons	FRA	2,15
Libelula	IT	2,18
MV Magvas	HUN	2,21
Ludwig	AUT	2,22
Renata	HR	2,23
Patria	HR	2,63
Ružica	HR	2,64
MV Mambo	HUN	2,86

a)
- aritmetička sredina – 1,84 g
- standardna devijacija – 0,43 g
- koeficijent varijacije – 23,22 %
- mod – 1,82 g
- medijana (negrupirani) – 1,72 g
- medijana (grupirani) – 1,73 g
- 80. percentil – 2,215 cm

b)
- aritmetička sredina – 1,84 g
- standardna devijacija – 0,42 g
- koeficijent varijacije – 22,90 %
- mod – 1,65 g
- raspon – 1,94 g
- 4. percentil – 1,06 cm

c)
- aritmetička sredina – 1,85 g
- standardna devijacija – 0,47 g
- koeficijent varijacije – 25,50 %
- mod – 1,83 g
- interkvartilna razlika – 0,64 cm
- 14. percentil – 1,35 cm

d)
- najveći broj ispitivanih sorata, 18 od ukupno 28, podrijetlom je iz Hrvatske
- udio hrvatskih sorata je 64,3 %

Literatura

1. Gogala, Z. (2001.). Osnove statistike. Zagreb: Sinergija.
2. Horvat, D. (2005.). Biometrika u poljoprivredi. Osijek: Sveučilište Josipa Jurja Strossmayera u Osijeku, Poljoprivredni fakultet u Osijeku.
3. Lane, D. M. (2013). Introduction to Statistics: An Interactive eBook. Poveznica na knjigu.
4. Lang, T. (2004.). Twenty Statistical Errors Even YOU Can Find in Biomedical Research Articles. Croatian Medical Journal, 45 (4) 361-370
5. Madsen, B. S. (2016.). Statistics for Non-Statisticians – 2 ed. Berlin: Springer – Verlag.
6. McCollough, C., Van Atta, L. (1963.). Statistical concepts: A program for self-instructions. Mc Graw-Hill, Inc.
7. Salkind, N. J. (2011.). Statistics for People Who (Think they) Hate Statistics -4th ed. SAGE Publications, Inc.
8. Šošić, I., Serdar, V. (2000.) Uvod u statistiku. Zagreb: Školska knjiga.
9. Vasilj, Đ. (2000.). Biometrika. Zagreb: Sveučilište u Zagrebu.

Kazalo