1. DESKRIPTIVNA STATISTIČKA ANALIZA

1.1. Uređivanje i prikazivanje podataka

Dobijeni podaci o vrednostima obeležja koje imaju ispitivani elementi populacije (ili uzorka) predstavljaju niz neuređenih numeričkih podataka.

Postupak njihovog uređivanja predstavlja njihovo predstavljanje po veličini od najmanje do najveće vrednosti. Tako uređen niz vrednosti posmatranog obeležja predstavlja **uređenu statističku seriju**, čiji ćemo *i*-ti član obeležavati sa x_i, i obično ih nazivamo **negrupisanim podacima**.

Prikazivanje uređenih podataka može biti **tabelarno** (u vidu tabela) i **grafičko** (u vidu grafikona).

Da bi se uređeni podaci prikazali tabelarno ili grafički, neophodno je prvo ih **grupisati**, odnosno odrediti **klase** po kojima će se raspoređivati elementi populacije (ili uzorka). U zavisnosti od osobina obeležja koje se ispituje, klase mogu biti **diskretne** (ispitivana obeležja su diskretna) ili **intervalne** (ispitivana obeležja su neprekidna).

Diskretna klasa, kod **diskretnih obeležja**, određuje onu grupu članova populacije koja ima **istu** vrednost obeležja, dok **intervalna klasa,** kod **neprekidnih obeležja**, određuje onu grupu članova populacije koja ima vrednosti unutar **istog**, **unapred određenog intervala** vrednosti obeležja.

Klasni intervali intervalnih klasa obično su iste širine (u analizama koje će biti sprovedene u ovom udžbeniku, **klasni intervali moraju biti iste širine**).

Donja i gornja granica klasnog intervala moraju biti **jasno i nedvosmisleno određene i moraju očuvati kontinualnu prirodu vrednosti** posmatranog obeležja, odnosno moraju biti u vidu poluotvorenih intervala koji se ne preklapaju (čiji je presek prazan skup), i koji prekrivaju čitavu oblast mogućih posmatranih vrednosti obeležja (čija je unija čitav interval mogućih vrednosti obeležja).

Dakle, kod diskretnih obeležja, klase su određene vrednostima obeležja koje se pojavljuju u uređenoj seriji, dok kod neprekidnih obeležja, klase (njihov broj, odnosno klasni interval) određuje osoba koja vrši statističku analizu.

Broj klasa nadalje ćemo obeležavati sa ***K.***

Ne postoji jasno definisano pravilo po kome bi se trebalo rukovoditi prilikom određivanja broja intervalnih klasa, već određeni broj preporuka.

Neke od njih su:

a) po Sturgesovoj formuli $K \approx 1 + 3.3\log N$

b)

$$\begin{aligned}
&N \in [40,60) \Rightarrow K \in [6,8) \\
&N \in [60,100) \Rightarrow K \in [7,10) \\
&N \in [100,200) \Rightarrow K \in [9,12) \\
&N \in [200,500) \Rightarrow K \in [12,17) \\
&N \rangle 500 \Rightarrow K = 21
\end{aligned}$$

c)

za bilo koje N važi $K \in [5,15]$.

Predstavnik *i*-te klase (obeležavaćemo ga sa $x_i^{;}$) kod diskretnih klasa predstavlja **vrednost obeležja koje određuje *i*-tu klasu**, dok kod neprekidnih klasa predstavlja **sredinu *i*-tog klasnog intervala**.

Širina *i*-tog klasnog intervala (obeležavaćemo je sa Δ_i) predstavlja razliku između početka (i+1)-og i početka i-tog klasnog intervala. U analizama koje će biti sprovedene u ovom udžbeniku, intervalne klase moraju biti izabrane tako da su im širine međusobno iste, pa ćemo **širinu klasnog intervala obeležavati sa Δ.**

Tabelarno, odnosno grafičko prikazivanje uređenih podataka, u stvari, predstavlja prikazivanje određenog broja pojavljivanja vrednosti datog obeležja u ispitivanoj populaciji (uzorku), to jest prikazivanje sledećih frekvencija:

- **Apsolutna frekvencija *i*-te klase – obeležićemo je sa f_i –** koja predstavlja ukupan broj članova populacije (uzorka) koji imaju vrednost obeležja određenu ***i***-tom klasom (*i = 1, 2, ..., K*).

  Očigledno je da važi $\sum_{i=1}^{K} f_i = N$.

- **Relativna frekvencija *i*-te klase – obeležićemo je sa p_i, –** koja predstavlja relativan broj članova populacije (uzorka) koji imaju vrednost obeležja određenu ***i***-tom klasom (*i = 1,2,...,K*), to jest

  $p_i = \frac{f_i}{N}$.

  Očigledno je da važi $\sum_{i=1}^{K} p_i = 1$.

- **Kumulativna frekvencija *i*-te klase – obeležićemo je sa K_i – određena sa „≤" ukoliko se radi o diskretnim klasama ili intervalnim klasama čija je gornja granica zatvorena, ili sa „<" ukoliko se radi o intervalnim klasama čija je gornja granica otvorena.** Ova kumulativna frekvencija predstavlja ukupan broj članova populacije (uzorka) koji imaju vrednost obeležja manju ili jednaku od vrednosti obeležja ***i***-te diskretne klase, ili od vrednosti gornje granice ***i***-te intervalne klase čija je gornja granica zatvorena, odnosno predstavlja ukupan broj članova populacije (uzorka) koji imaju vrednost obeležja manju od vrednosti gornje granice ***i***-te intervalne klase čija je gornja granica otvorena (*i = 1, 2, ..., K*).

  Očigledno je da važi $K_K = N$.

- **Kumulativna frekvencija *i*-te klase – obeležićemo je sa W_i – određena sa „≥" ukoliko se radi o diskretnim klasama ili intervalnim klasama čija je donja granica zatvorena, ili sa „>" ukoliko se radi o intervalnim klasama čija je donja granica otvorena.** Ova kumulativna frekvencija predstavlja ukupan broj članova populacije (uzorka) koji imaju vrednost obeležja veću ili jednaku od vrednosti obeležja ***i***-te diskretne klase, ili od vrednosti donje granice ***i***-te intervalne klase čija je donja granica zatvorena, odnosno predstavlja ukupan broj članova populacije (uzorka) koji imaju vrednost obeležja veću od vrednosti donje granice ***i***-te intervalne klase čija je donja granica otvorena ($i = 1, 2, ..., K$).

Očigledno je da važi $W_1 = N$.

Naravno, sve gore navedeno važi i za uzorak, uz činjenicu da je tada $N=n$.

Grafički prikaz frekvencija daje se u vidu grafikona u pravouglom Dekartovom koordinatnom sistemu (x, y), gde se na x-osu nanose vrednosti obeležja i-te klase, a na y-osu vrednost frekvencije koja se predstavlja (recimo f_i).

Kod diskretnih obeležja (diskretnih klasa) taj grafički prikaz je u vidu izlomljenih pravih linija koje spajaju tačke (x_i^*, f_i) (ukoliko predstavljamo apsolutnu frekvenciju f_i) i naziva se **poligon.**

Kod neprekidnih obeležja (intervalnih klasa) taj grafički prikaz je u vidu pravougaonika čija je širina jednaka širini klasnog intervala, a visina nad i-tim klasnim intervalom jednaka vrednosti i-te frekvencije koja se predstavlja (recimo f_i).

Primer 1.1.1. Anketirana je populacija od 50 studenata o broju položenih ispita i telesnoj težini. Dobijeni su sledeći rezultati:

a) broj položenih ispita:

7	4	12	3	7	8	6	5	9	9	10
11	6	7	8	6	9	4	5	5	7	3
9	8	6	8	7	6	8	9	6	7	4
10	11	11	12	6	7	7	8	4	10	11
4	12	6	7	8	9					

b) težina (u kg):

57	84	112	83	77	68	96	105	90	69	102
71	68	72	78	76	89	74	55	85	87	73
89	78	67	68	74	69	80	79	66	67	64
104	110	91	92	68	75	77	82	64	101	91
64	62	65	73	81	91					

Prikazati tabelarno i grafički apsolutnu frekvenciju, relativnu frekvenciju, kumulativne frekvencije određene sa „manje do jednako od" i „veće do jednako od", za oba obeležja.

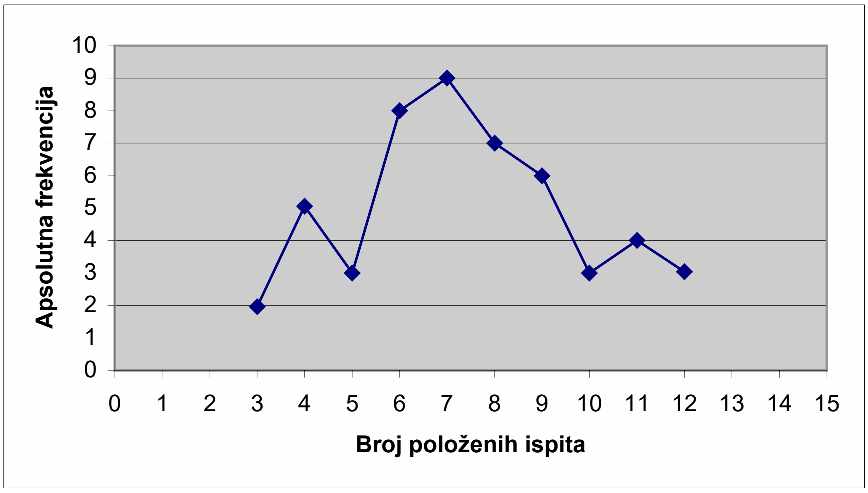
Rešenje:

a)
Tabelarni prikaz dat je u Tabeli 1.1.1.a.

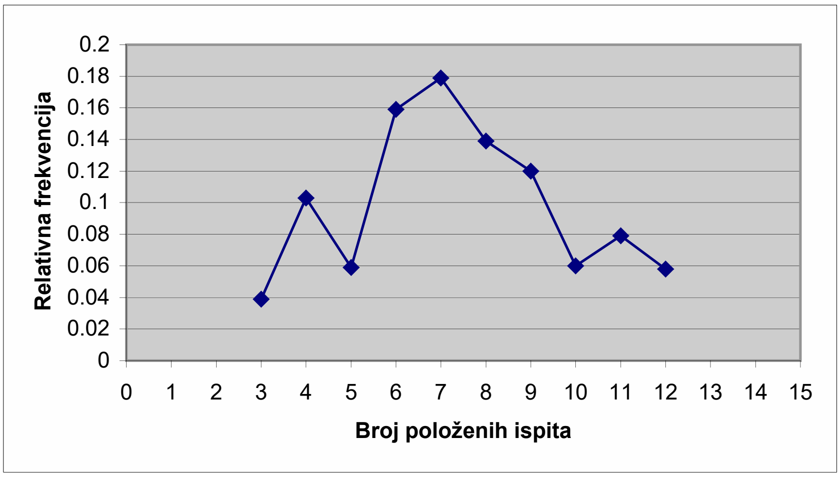
Broj položenih ispita	*Apsolutna frekvencija f_i*	*Relativna frekvencija p_i*	*Kumulativna frekvencija ≤ K_i*	*Kumulativna frekvencija ≥ W_i*
3	2	2/50	2	50
4	5	5/50	7	48
5	3	3/50	10	43
6	8	8/50	18	40
7	9	9/50	27	32
8	7	7/50	34	23
9	6	6/50	40	16
10	3	3/50	43	10
11	4	4/50	47	7
12	3	3/50	50	3
	$\sum f_i=50$	$\sum p_i=1$		

Tabela 1.1.1.a.

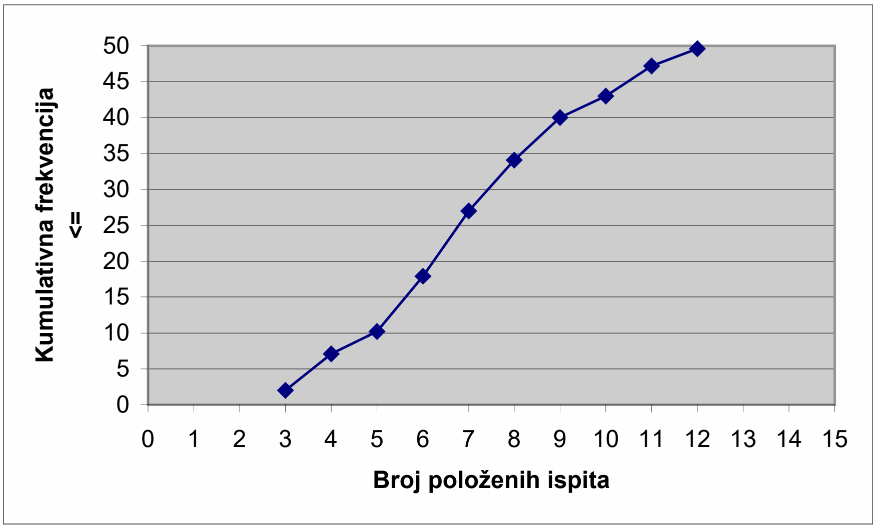
Grafički prikaz apsolutne, relativne i kumulativnih frekvencija dat je na sledećim slikama.



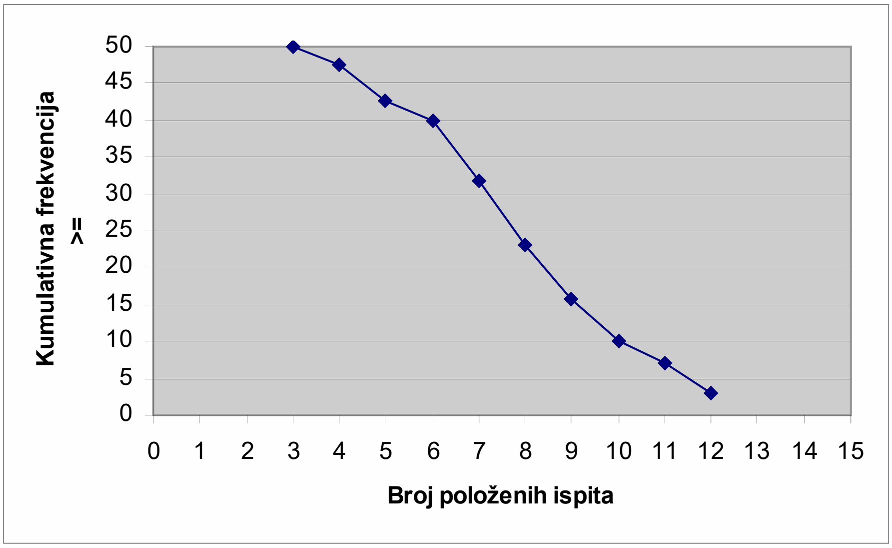
Slika 1.1.1.a. Grafički prikaz apsolutne frekvencije



Slika 1.1.1.b. Grafički prikaz relativne frekvencije



1.1.1.c. Grafički prikaz kumulativne frekvencije (≤)



1.1.1.d. Grafički prikaz kumulativne frekvencije (≥)

b)

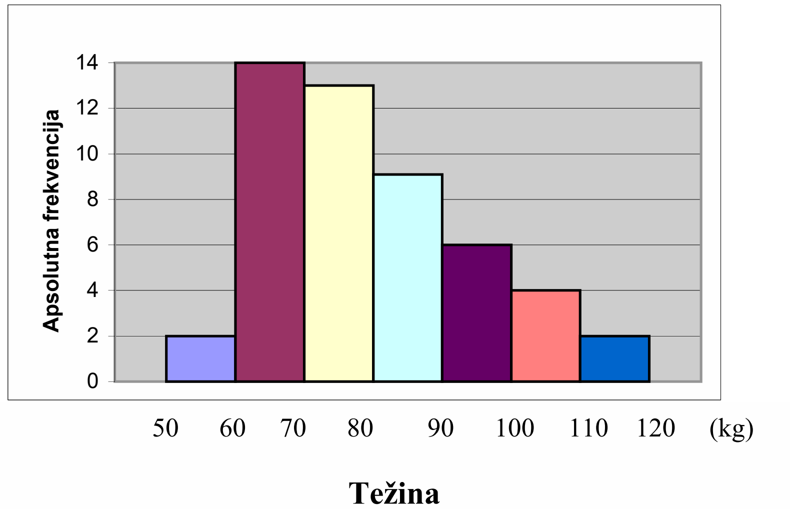
Imajući u vidu preporuke date u ovom udžbeniku o grupisanju neprekidnih vrednosti obeležja u intervalne klase i činjenicu da je najmanja težina u populaciji 55 kg, a najveća 112 kg, podatke o težini iz ove populacije ćemo grupisati u sedam klasa širine 10 kg, počev od 50 kg.

Tabelarni prikaz je dat u Tabeli 1.1.1.b.

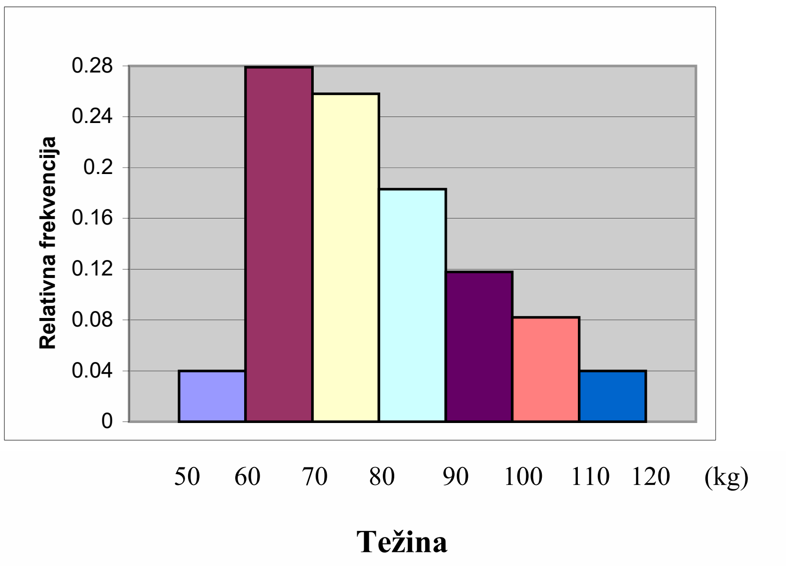
Interval težine (kg)	Apsolutna frekvencija f_i	Relativna frekvencija p_i	Kumulativna frekvencija < K_i	Kumulativna frekvencija ≥ W_i
[50 -60)	2	2/50	2	50
[60 -70)	14	14/50	16	48
[70 -80)	13	13/50	29	34
[80 -90)	9	9/50	38	21
[90 -100)	6	6/50	44	12
[100 -110)	4	4/50	48	6
[110 -120)	2	2/50	50	2
	$\sum f_i=50$	$\sum p_i=1$		

Tabela 1.1.1.b.

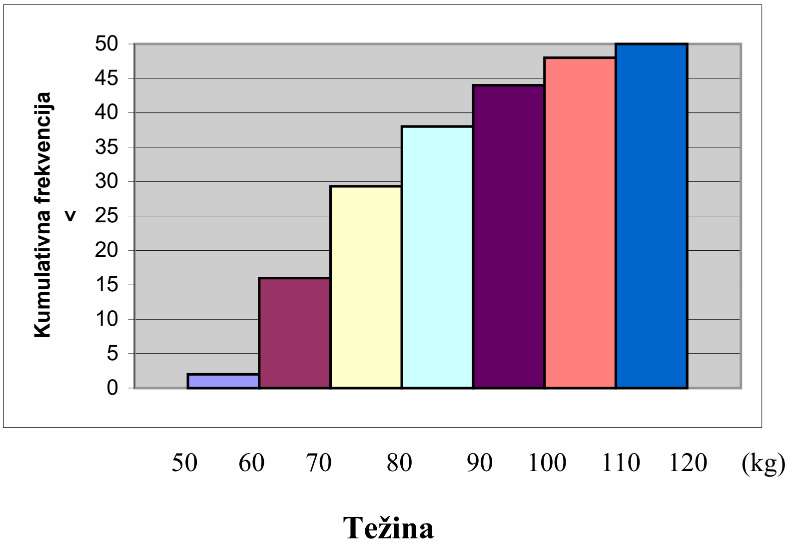
Grafički prikaz apsolutne, relativne i kumulativnih frekvencija dat je na sledećim slikama.



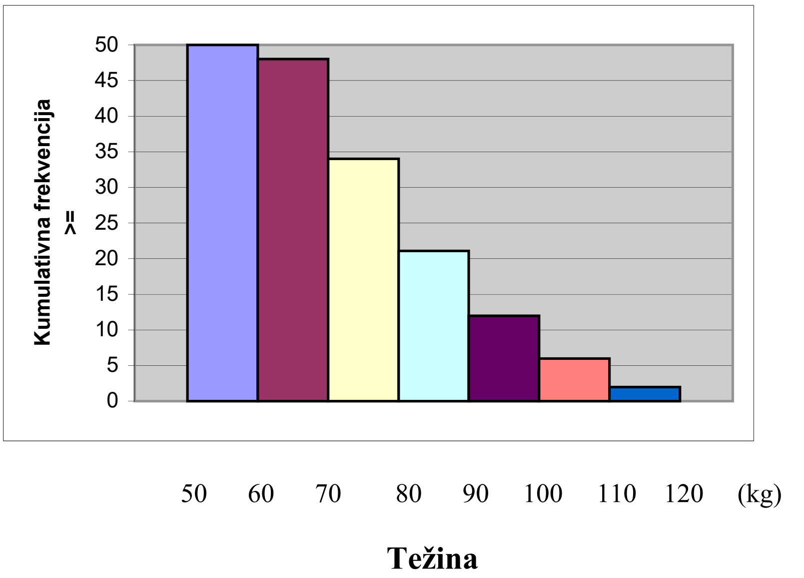
Slika 1.1.1.e. Grafički prikaz apsolutne frekvencije



Slika 1.1.1.f. Grafički prikaz relativne frekvencije



Slika 1.1.1.g. Grafički prikaz kumulativne frekvencije (<)



Slika 1.1.1.h. Grafički prikaz kumulativne frekvencije (≥)

1.2. Izračunavanje parametara raspodele obeležja

Uređivanje i prikazivanje podataka analizirano u prethodnom odeljku 1.1, samo je priprema podataka za njihovu dalju obradu. Naime, da bi se izveli određeni zaključci o raspodeli ispitivanog obeležja u populaciji, ili u uzorku, pa kasnije na osnovu toga i u celoj populaciji, neophodno je izračunati neke veličine (parametre, pokazatelje, mere) koji na neki način prezentuju ponašanje raspodeljenosti obeležja u posmatranoj populaciji.

Uopšteno, ti parametri se mogu podeliti u tri osnovne grupe:

- parametri srednje vrednosti (mere srednje vrednosti ili centralne tendencije);
- parametri varijabiliteta (mere disperzije);
- parametri oblika rasporeda (mere oblika rasporeda).

Neki od parametara koje ćemo izračunavati isto će se označavati i za populaciju i za uzorak, dok će neki imati različite oznake u zavisnosti od toga da li se radi o populaciji ili uzorku.

Takođe, neke od formula po kojima će se računati parametri, biće identične i za populaciju i za uzorak, dok će se neke razlikovati.

U ovom trenutku čitalac treba da obrati pažnju na te razlike i da ih prihvati bez dodatnih objašnjenja, uz napomenu da će se u kasnijim delovima udžbenika adekvatno razjasniti potreba za ovim razlikama.

1.2.1. Parametri srednje vrednosti (mere srednje vrednosti)

Parametre srednje vrednosti možemo podeliti na **izračunate** i **pozicione**.

Izračunati su:

- aritmetička sredina,
- geometrijska sredina,
- harmonijska sredina.

Pozicioni su:

- modus,
- medijana.

Zavisnost prirode promene ispitivanog obeležja nam diktira koji ćemo od gore pomenuta tri izračunata parametra srednje vrednosti da koristimo kao relevantnu meru srednje vrednosti datog obeležja u posmatranoj populaciji (uzorku).

Kada je priroda promene posmatranog obeležja **linearno zavisna**, koristićemo **aritmetičku sredinu** kao meru srednje vrednosti. Linearna zavisnost promene ogleda se u tome da je ukupna grupna vrednost obeležja jednaka aritmetičkom zbiru obeležja svakog člana grupe ponaosob (visina, težina, broj prodatih automobila, broj poena na ispitu...). Ovakva priroda promene je najzastupljenija u svakodnevnom životu.

Kada je priroda promene posmatranog obeležja **direktno proporcionalna**, koristićemo **geometrijsku sredinu** kao meru srednje vrednosti. Direktna proporcionalnost promene ogleda se u tome da je ukupna grupna vrednost obeležja jednaka proizvodu obeležja svakog člana grupe ponaosob (procenat...).

Kada je priroda promene posmatranog obeležja **obrnuto proporcionalna**, koristićemo **harmonijsku sredinu** kao meru srednje vrednosti. Obrnuta proporcionalnost promene ogleda se u tome da se grupna vrednost obeležja smanjuje kada se povećava broj članova grupe i obrnuto (vreme završetka nekog posla u odnosu na broj ljudi koji ga istovremeno rade i sl.).

- Aritmetička sredina. Aritmetičku sredinu bilo kojih N brojeva $x_1, x_2, ..., x_N$ obeležavamo sa $\overline{X}$ i izračunavamo po formuli:

$$\overline{X} = \frac{x_1 + x_2 + ... + x_N}{N}.$$

Od sada pa nadalje u ovom udžbeniku, **aritmetičku sredinu populacije** ćemo obeležavati sa μ, dok ćemo **aritmetičku sredinu uzorka** obeležavati sa $\overline{X}$. Kada podatke nismo grupisali u klase važi:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \qquad \overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Kada smo podatke grupisali u klase važi:

$$\mu = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{'}}{N} \qquad \overline{X} = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{'}}{n}.$$

Imajući u vidu da je predstavnik diskretne klase $x_i^{'}$ jednak vrednosti obeležja koje imaju svi članovi te klase (u i-toj klasi f_i članova ima vrednost x_i koja je jednaka $x_i^{'}$), zaključujemo da je aritmetička sredina u istoj seriji podataka diskretnih obeležja identična bez obzira na to da li su podaci grupisani ili ne, dok za neprekidna obeležja to ne možemo da tvrdimo jer je kod grupisanih neprekidnih obeležja predstavnik i-te klase $x_i^{'}$, sredina i-tog klasnog intervala, a sama raspodela obeležja u i-toj klasi ne mora biti takva da je njihova aritmetička sredina baš sredina tog klasnog intervala.

Primer 1.2.1.1. Za podatke iz **Primera 1.1.1.** izračunati aritmetičku sredinu broja položenih ispita i težine anketiranih studenata

a) negrupišući podatke,
b) grupišući podatke.

Rešenje:

i)

Ako ne grupišemo podatke, za aritmetičku sredinu broja položenih ispita dobijamo:

$$\mu = \frac{\sum_{i=1}^{50} x_i}{50} = 7.44$$

a za aritmetičku sredinu težine studenata dobijamo:

$$\mu = \frac{\sum_{i=1}^{50} x_i}{50} = 79.26\,kg.$$

ii)

Ako grupišemo podatke, za aritmetičku sredinu broja položenih ispita dobijamo:

$$\mu = \frac{\sum_{i=1}^{10} f_i \cdot x_i^{'}}{N} =$$

$$= \frac{2\cdot 3 + 5\cdot 4 + 3\cdot 5 + 8\cdot 6 + 9\cdot 7 + 7\cdot 8 + 6\cdot 9 + 3\cdot 10 + 4\cdot 11 + 3\cdot 12}{50} = 7.44$$

naravno, istu vrednost kao i kada nismo grupisali podatke, jer je obeležje diskretnog tipa.

Ako grupišemo podatke, za aritmetičku sredinu težine studenata dobijamo:

$$\mu = \frac{\sum_{i=1}^{7} f_i \cdot x_i^{'}}{N} =$$

$$= \frac{2\cdot 55 + 14\cdot 65 + 13\cdot 75 + 9\cdot 85 + 6\cdot 95 + 4\cdot 105 + 2\cdot 115}{50} = 79.6\,kg.$$

Kao što smo i očekivali, aritmetička sredina grupisanih podataka se razlikuje od aritmetičke sredine negrupisanih jer je težina obeležje neprekidnog tipa.

Aritmetička sredina se ponekad naziva i moment prvog reda.

Generalno, **moment reda *k*** označavamo sa $\boldsymbol{m_k}$ i izračunavamo po formuli:

$$m_k = \frac{\sum_{i=1}^{N} x_i^{\,k}}{N}$$ za negrupisane podatke,

$$m_k = \frac{\sum_{i=1}^{K} f_i \cdot x_i^k}{N}$$ za grupisane podatke.

- Aritmetička sredina aritmetičkih sredina. Neka je dato m različitih populacija (ili uzorka), i neka je broj članova svake od ovih populacija (uzorka) N_i (odnosno n_i), i neka je aritmetička sredina svake od ovih populacija (uzorka) μ_i (odnosno $\overline{X}_i$), tada je, ako sve ove populacije (uzorke) spojimo u jednu, aritmetička sredina novodobijene populacije (uzorka) jednaka:

$$\overline{\mu} = \frac{\sum_{i=1}^{m} N_i \cdot \mu_i}{\sum_{i=1}^{m} N_i} \qquad \overline{X} = \frac{\sum_{i=1}^{m} n_i \cdot \overline{X}_i}{\sum_{i=1}^{m} n_i}.$$

Primer 1.2.1.2. Prosečna visina svih 220 devojaka na prvoj godini studija je 168 cm, a prosečna visina svih180 muškaraca na prvoj godini je 182 cm. Kolika je prosečna visina studenta prve godine?

Rešenje:

Prosečna visina studenta na prvoj godini studija je

$$\overline{\mu} = \frac{\sum_{i=1}^{m} N_i \cdot \mu_i}{\sum_{i=1}^{m} N_i} = \frac{220 \cdot 168 + 180 \cdot 182}{220 + 180} = 174.3\,cm.$$

- Geometrijska sredina. Geometrijsku sredinu bilo kojih N brojeva $x_1, x_2, ...,$ x_N obeležavamo sa G i izračunavamo po formuli:

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot ... \cdot x_N}\ .$$

Kada želimo da iskažemo geometrijsku sredinu kao meru srednje vrednosti određenog obeležja u populaciji (ili uzorku), onda važi:

za negrupisane podatke $G = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_N}$,

za grupisane podatke $G = \sqrt[K]{x_1'^{f_1} \cdot x_2'^{f_2} \cdot \ldots \cdot x_K'^{f_K}}$,

gde je:

G - reprezent geometrijske sredine koji pravilno predstavlja direktno proporcionalnu prirodu promene datog obeležja;

x_i - reprezent vrednosti obeležja koje pravilno predstavlja direktno proporcionalnu prirodu promene datog obeležja (i = 1, 2, ..., N);

x_i' - reprezent predstavnika i-te klase koji pravilno predstavlja direktno proporcionalnu prirodu promene datog obeležja (i = 1, 2, ..., K).

Formule za geometrijsku sredinu uzorka su identične uz uslov *N*=*n.*

Primer 1.2.1.3. U poslednje četiri godine cena neke robe se menjala na sledeći način: prve godine povećala se za 10%, druge se povećala za 12%, treće godine se smanjila za 5% i četvrte godine se smanjila za 17%. Kakva je prosečna procentualna godišnja promena cene ove robe u ovom periodu?

Rešenje:

Ukoliko sa ***p*** *predstavimo decimalni zapis procentualne promene od* ***p%***,
odnosno $p = \dfrac{p\%}{100}$, *onda je*

(1+ p) pravilni reprezent porasta od p%, a
(1 - p) pravilni reprezent smanjenja od p%,

pa je reprezent geometrijske sredine procentualne godišnje promene cena u ovom periodu

$$G = \sqrt[4]{1{,}1 \cdot 1{,}12 \cdot 0{,}95 \cdot 0{,}83} = 0{,}99278 .$$

Zaključujemo da se u ovom četvorogodišnjem periodu ovakvim promenama cena robe prosečno godišnje smanjivala 0,722%.

- Harmonijska sredina. *Harmonijsku sredinu bilo kojih* N *brojeva* $x_1, x_2, \ldots, x_N$ *obeležavamo sa* H *i izračunavamo po formuli:*

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}} .$$

Kada želimo da iskažemo harmonijsku sredinu kao meru srednje vrednosti određenog obeležja u populaciji (ili uzorku), onda važi:

za negrupisane podatke: $$H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

za grupisane podatke: $$H = \frac{N}{\sum_{i=1}^{K} \frac{f_i}{x_i'}} .$$

Formule za harmonijsku sredinu uzorka su identične uz uslov *N=n.*

Primer 1.2.1.4. Osoba A uradi jedan posao radeći sam za 10 časova. Osoba B za 12 časova, a osoba C za 9 časova. Ako su osobe A, B i C iz iste grupacije ljudi, koliko je srednje vreme potrebno da jedan čovek iz te grupacije sam uradi taj posao?

Rešenje:

Pošto je vreme za koje bi se završio taj posao obrnuto proporcionalno broju ljudi iz te grupacije koji bi ga istovremeno obavljali (što više ljudi istovremeno radi na tom poslu, to će vreme izrade posla biti manje), to je srednje vreme potrebno da jedan čovek iz te grupacije sam uradi taj posao jednako:

$$H = \frac{3}{\frac{1}{10} + \frac{1}{12} + \frac{1}{9}}$$ *časova=10,19 časova.*

Između harmonijske, geometrijske i harmonijske sredine izračunate nad istim skupom brojeva, važi sledeća nejednačina:

$$H \leq G \leq \overline{X}.$$

- Modus. Za negrupisane podatke modus (***M₀***) je vrednost obeležja koje se najčešće javlja u seriji.

Za grupisane podatke diskretnog tipa, modus ($\boldsymbol{M_0}$) je takođe vrednost obeležja koje se najčešće javlja u seriji, što u ovom slučaju predstavlja vrednost klase čija je apsolutna frekvencija najveća.
Za grupisane podatke neprekidnog tipa modus ($\boldsymbol{M_0}$) nalazimo tako što prvo pronađemo modalnu klasu, kao klasu čija je apsolutna frekvencija najveća. Zatim modus izračunavamo po formuli:

$$M_0 = L_{M_0} + \frac{f_{M_0} - f_{M_0 - 1}}{f_{M_0} - f_{M_0 - 1} + f_{M_0} - f_{M_0 + 1}} \cdot \Delta \ ,$$

gde je:

L_{M_0} - početak modalne klase,

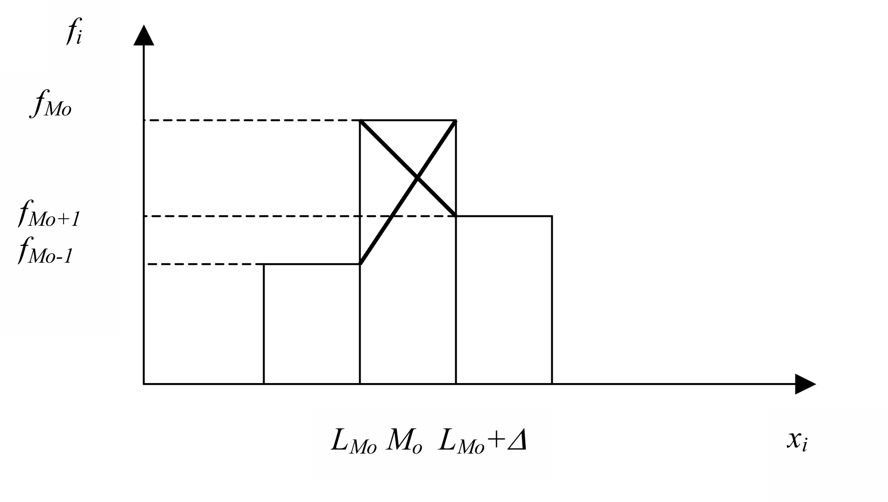
f_{M_0} - apsolutna frekvencija modalne klase,

$f_{M_0 - 1}$ - apsolutna frekvencija klase pre modalne,

$f_{M_0 + 1}$ - apsolutna frekvencija klase posle modalne,

Δ - širina klasnog intervala.

Vrednost modusa neprekidnih podataka grupisanih u klase može se dobiti i grafičkim putem kao apscisa tačke preseka prave koja je određena tačkama $\left(L_{M_0}, f_{M_0 - 1}\right)$ i $\left(L_{M_0} + \Delta, f_{M_0}\right)$ i prave određene tačkama $\left(L_{M_0}, f_{M_0}\right)$ i $\left(L_{M_0} + \Delta, f_{M_0 + 1}\right)$ (*Slika 1.2.1*).



Slika 1.2.1.

Interesantno je napomenuti da ako bi se konstruisala parabola koja prolazi kroz tri sredine gornjih strana pravougaonika, apscisa maksimuma te parabole bi se poklopila sa modusom.

Ukoliko se svaki podatak u seriji pojavljuje isti broj puta, odnosno ukoliko su sve apsolutne frekvencije međusobno jednake, onda modus ne postoji.

Ukoliko postoji samo jedan modus, kažemo da je serija, odnosno raspodela unimodalna. Ukoliko postoje dva modusa, onda kažemo da je serija, odnosno raspodela bimodalna, a ako postoji više od dva modusa, onda seriju, odnosno raspodelu proglašavamo polimodalnom.

Primer 1.2.1.5. Za podatke iz **Primera 1.1.1.** izračunati modus:

a) broja položenih ispita,
b) težine anketiranih studenata.

Rešenje:
a)
Kako je najfrekventnija klasa klasa u kojoj se nalaze članovi populacije koji imaju sedam položenih ispita, to je modus ove unimodalne raspodele $M_o=7$.

b)

Najfrekventnija klasa, odnosno modalna klasa, je klasa [60 -700) kg, pa je modus

$$M_o = L_{M_o} + \frac{f_{M_o} - f_{M_o-1}}{f_{M_o} - f_{M_o-1} + f_{M_o} - f_{M_o+1}} \cdot \Delta = 60 + \frac{14-2}{14-2+14-13} \cdot 10 \approx 69{,}23kg.$$

- Medijana. Za negrupisane podatke medijana (***M***M_e) je vrednost obeležja koje se nalazi u sredini uređene statističke serije. Dakle, broj članova populacije (uzorka) koji su manji od M_e te populacije (tog uzorka) jednak je broju članova te populacije (tog uzorka) koji su veći od M_e te populacije (tog uzorka).

Ukoliko je broj članova uređene statističke serije neparan, $N = 2 \cdot k + 1$, tada je medijana jednaka vrednosti $\left[\frac{N+1}{2}\right]$-og člana te serije, a ukoliko je broj članova uređene statističke serije paran, $N = 2 \cdot k$, tada je medijana jednaka aritmetičkoj sredini $\left[\frac{N}{2}\right]$-og i $\left[\frac{N}{2}+1\right]$-og člana te serije.

Odnosno, za negrupisane podatke važi:

$$N = 2 \cdot k + 1 \quad \Rightarrow \quad M_e = x_{\left[\frac{N+1}{2}\right]}$$

$$N = 2 \cdot k \quad \Rightarrow \quad M_e = \frac{x_{\left[\frac{N}{2}\right]} + x_{\left[\frac{N}{2}+1\right]}}{2}.$$

Za grupisane podatke diskretnog tipa, ***M_e*** takođe predstavlja vrednost obeležja koje se nalazi u sredini uređene statističke serije. Izračunavanje vrednosti $\boldsymbol{M_e}$ je u ovom slučaju isto kao i za negrupisane podatke, samo se vrednosti $\left[\frac{N+1}{2}\right]$-og člana (ukoliko je $N = 2 \cdot k + 1$), odnosno $\left[\frac{N}{2}\right]$-og i $\left[\frac{N}{2}+1\right]$-og člana (ukoliko je $N = 2 \cdot k$), nalaze iz tabelarne predstave uređene serije.

Za grupisane podatke neprekidnog tipa, $\boldsymbol{M_e}$ je vrednost koja deli histogram na dva dela jednakih površina. Naravno da i tada važi da je broj članova populacije (uzorka) koji su manji od tako određene $\boldsymbol{M_e}$ jednak broju članova te populacije (tog uzorka) koji su veći od $\boldsymbol{M_e}$.

Da bismo odredili $\boldsymbol{M_e}$ za grupisane podatke neprekidnog tipa, potrebno je prvo da pronađemo medijalnu klasu $M_{e,kl.}$

Medijalnu klasu $M_{e,kl.}$ određujemo iz uslova:

$$\sum_{i=1}^{M_{e,kl}-1} f_i \le \frac{N}{2} \langle \sum_{i=1}^{M_{e,kl}} f_i \, .$$

Medijanu $\boldsymbol{M_e}$ određujemo iz jednačine

$$M_e = L_{M_{e,kl}} + \frac{\frac{N}{2} - \sum_{i=1}^{M_{e,kl}-1} f_i}{f_{M_{e,kl}}} \cdot \Delta \, ,$$

gde je:

$L_{M_{e,kl}}$ - početak medijalne klase,
$f_{M_{0,kl}}$ - apsolutna frekvencija medijalne klase,
N - broj članova populacije,
f_i - apsolutna frekvencija i-te klase,
Δ - širina klasnog intervala.

Formule za medijanu $\boldsymbol{M_e}$ uzorka su identične uz uslov N=n.

Primer 1.2.1.6. Za podatke iz **Primera 1.1.1.** izračunati medijanu:

a) broja položenih ispita,
b) težine anketiranih studenata.

Rešenje:

a)

Populacija sadrži paran broj članova, N=50, pa medijanu nalazimo kao aritmetičku sredinu 25. i 26. člana serije uređene u rastući poredak.

Analizirajući kumulativnu frekvenciju K_i , vidimo da se i 25. i 26. član serije uređene u rastući poredak nalaze u klasi u kojoj se nalaze članovi populacije koji su položili sedam ispitai, pa je medijana ove raspodele

$$M_e = \frac{X_{25} + X_{26}}{2} = \frac{7+7}{2} = 7.$$

b)

Kako je $\frac{N}{2} = \frac{50}{2} = 25$ *, to je medijalna klasa [70 -80) kg, pa je medijana*

$$M_e = L_{M_{e,kl}} + \frac{\frac{N}{2} - \sum_{i=1}^{M_{e,kl}-1} f_i}{f_{M_{e,kl}}} \cdot \Delta = 70 + \frac{25-16}{13} \cdot 10 \approx 76{,}92\,kg.$$

1.2.2. Parametri varijabiliteta (mere disperzije)

Parametre varijabiliteta možemo podeliti na **apsolutne** i **relativne**.

Apsolutni su:

- interval varijacije,
- interkvartilna razlika,
- srednje apsolutno odstupanje,
- srednje kvadratno odstupanje (varijansa),
- standardna devijacija.

Relativni su:

- koeficijent varijacije,
- normalizovano standardno odstupanje.

- Interval varijacije. Interval varijacije (***I_v***) je interval vrednosti obeležja u kome se nalaze vrednosti obeležja svih članova populacije (uzorka).

Kod negrupisanih podataka, kao i kod grupisanih podataka diskretnog tipa, izračunava se kao razlika najveće i najmanje vrednosti obeležja u ispitivanoj populaciji (uzorku):

$I_v = x_{\max} - x_{\min}$.

Kod grupisanih podataka neprekidnog tipa, izračunava se kao razlika gornje granice poslednje klase i donje granice prve klase.

Primer 1.2.2.1. Za podatke iz **Primera 1.1.1.** izračunati interval varijacije:

c) broja položenih ispita
d) težine anketiranih studenata.

Rešenje:

a)

Interval varijacije je $I_v = x_{\max} - x_{\min} = 12 - 3 = 9$ *položenih ispita.*

b)

Interval varijacije je $I_v = x_{\max} - x_{\min} = 120 - 50 = 70\,kg.$

- Interkvartilna razlika. Pre nego što definišemo interkvartilnu razliku, definisaćemo sledeće pojmove: **kvartil, decil** i **percentil.**

- Kvartili su brojevi, koje ćemo obeležavati sa $Q_i \quad (i = 1,...,3)$, sa osobinom da *i*-ti kvartil (***Q****i*) deli interval varijacije u odnosu $i : 4 - i$.
To znači da Q_i predstavlja vrednost obeležja od koje $i \cdot 25\%$ elemenata populacije (uzorka) ima vrednost manju ili jednaku od Q_i. Očigledno da je $Q_2 = M_e$.

Za negrupisane podatke, kao i za grupisane podatke diskretnog tipa, *i*-ti kvartil izračunavamo po formuli:

$$Q_i = x_{\left[i\cdot\frac{N-1}{4}+1\right]} + \left(\left(i\cdot\frac{N-1}{4}+1\right) - \left[i\cdot\frac{N-1}{4}+1\right]\right)\cdot\left(x_{\left[i\cdot\frac{N-1}{4}+2\right]} - x_{\left[i\cdot\frac{N-1}{4}+1\right]}\right)$$

$i = 1,...,3$

gde je sa $[\alpha]$ označen ceo deo broja α.

U suštini, ova formula znači sledeće:

ukoliko je $\left(i\cdot\frac{N-1}{4}+1\right)$ ceo broj, tada je vrednost *i*-tog kvartila jednaka vrednosti $\left(i\cdot\frac{N-1}{4}+1\right)$-og člana uređene statističke serije.

A ukoliko $\left(i\cdot\frac{N-1}{4}+1\right)$ nije ceo broj, tada se vrednost *i*-tog kvartila dobija linearnom interpolacijom vrednosti $\left(i\cdot\frac{N-1}{4}+1\right)$-og i $\left(i\cdot\frac{N-1}{4}+2\right)$-og člana uređene statističke serije.

Da bismo odredili vrednost *i*-tog kvartila (***Q**_i*) za grupisane podatke neprekidnog tipa, neophodno je prvo odrediti klasu *i*-tog kvartila $Q_{i,kl}$, iz sledećeg uslova:

$$\sum_{i=1}^{Q_{i,kl}-1} f_i \leq i \cdot \frac{N}{4} \langle \sum_{i=1}^{Q_{i,kl}} f_i \qquad i=1,...,3.$$

Vrednost $\boldsymbol{Q_i}$ i=1,...,3 određujemo iz jednačine

$$Q_i = L_{Q_{i,kl}} + \frac{i \cdot \frac{N}{4} - \sum_{i=1}^{Q_{i,kl}-1} f_i}{f_{Q_{i,kl}}} \cdot \Delta$$

gde je:

$L_{Q_{i,kl}}$ - početak klase *i*-tog kvartila,
$f_{Q_{i,kl}}$ - apsolutna frekvencija klase *i*-tog kvartila,
N - broj članova populacije,
f_i - apsolutna frekvencija *i*-te klase,
Δ - širina klasnog intervala.

Formule za Q_i uzorka su identične uz uslov $N=n$.

- Decili su brojevi, koje ćemo obeležavati sa $D_i \quad (i = 1,...,9)$, sa osobinom da *i*-ti decil ($\boldsymbol{D_i}$) deli interval varijacije u odnosu $i:10-i$.
To znači da D_i predstavlja vrednost obeležja od koje $i \cdot 10\%$ elemenata populacije (uzorka) ima vrednost manju ili jednaku od D_i. Očigledno da je $D_5 = M_e$.

Za negrupisane podatke, kao i za grupisane podatke diskretnog tipa, *i*-ti decil izračunavamo po formuli:

$$D_i = x_{\left[i \cdot \frac{N-1}{10}+1\right]} + \left(\left(i \cdot \frac{N-1}{10}+1\right) - \left[i \cdot \frac{N-1}{10}+1\right]\right) \cdot \left(x_{\left[i \cdot \frac{N-1}{10}+2\right]} - x_{\left[i \cdot \frac{N-1}{10}+1\right]}\right)$$

$i = 1,...,9$

gde je sa $[\alpha]$ označen ceo deo broja α .

U suštini, ova formula znači sledeće:

ukoliko je $\left(i\cdot\frac{N-1}{10}+1\right)$ ceo broj, tada je vrednost i-tog decila jednaka vrednosti $\left(i\cdot\frac{N-1}{10}+1\right)$-og člana uređene statističke serije.

A ukoliko $\left(i\cdot\frac{N-1}{10}+1\right)$ nije ceo broj, tada se vrednost i-tog decila dobija linearnom interpolacijom vrednosti $\left(i\cdot\frac{N-1}{10}+1\right)$-og i $\left(i\cdot\frac{N-1}{10}+2\right)$-og člana uređene statističke serije.

Da bismo odredili vrednost *i*-tog decila (***D**$_i$*) za grupisane podatke neprekidnog tipa, neophodno je prvo odrediti klasu *i*-tog decila $D_{i,kl}$, iz sledećeg uslova:

$$\sum_{i=1}^{D_{i,kl}-1} f_i \le i\cdot\frac{N}{10} \langle \sum_{i=1}^{D_{i,kl}} f_i \qquad i=1,...,9.$$

Vrednost $\boldsymbol{D_i}$ i=1,...,9 određujemo iz jednačine

$$D_i = L_{D_{i,kl}} + \frac{i\cdot\frac{N}{10} - \sum_{i=1}^{D_{i,kl}-1} f_i}{f_{D_{i,kl}}}\cdot\Delta \ ,$$

gde je:

$L_{D_{i,kl}}$ - početak klase *i*-tog decila,
$f_{D_{i,kl}}$ - apsolutna frekvencija klase *i*-tog decila,
N - broj članova populacije,
f_i - apsolutna frekvencija *i*-te klase,
Δ - širina klasnog intervala.

Formule za D_i uzorka su identične uz uslov $N=n$.

- Percentili su brojevi, koje ćemo obeležavati sa $P_i \quad (i=1,...,99)$, sa osobinom da i-ti percentil (***P$_i$***) deli interval varijacije u odnosu $i:100-i$. To znači da P_i predstavlja vrednost obeležja od koje $i\cdot 1\%$ elemenata populacije (uzorka) ima vrednost manju ili jednaku od P_i. Očigledno da je $P_{50}=M_e$.

Za negrupisane podatke, kao i za grupisane podatke diskretnog tipa, i-ti percentil izračunavamo po formuli:

$$P_i = x_{\left[i\cdot\frac{N-1}{100}+1\right]} + \left(\left(i\cdot\frac{N-1}{100}+1\right)-\left[i\cdot\frac{N-1}{100}+1\right]\right)\cdot\left(x_{\left[i\cdot\frac{N-1}{100}+2\right]} - x_{\left[i\cdot\frac{N-1}{100}+1\right]}\right)$$

$i=1,...,99$

gde je sa $[\alpha]$ označen ceo deo broja α.

U suštini, ova formula znači sledeće:

ukoliko je $\left(i\cdot\frac{N-1}{100}+1\right)$ ceo broj, tada je vrednost i-tog percentila jednaka vrednosti $\left(i\cdot\frac{N-1}{100}+1\right)$-og člana uređene statističke serije.

A ukoliko $\left(i\cdot\frac{N-1}{100}+1\right)$ nije ceo broj, tada se vrednost i-tog percentila dobija linearnom interpolacijom vrednosti $\left(i\cdot\frac{N-1}{100}+1\right)$-og i $\left(i\cdot\frac{N-1}{100}+2\right)$-og člana uređene statističke serije.

Da bismo odredili vrednost i-tog percentila (***P$_i$***) za grupisane podatke neprekidnog tipa, neophodno je prvo odrediti klasu i-tog percentila $P_{i,kl}$, iz sledećeg uslova:

$$\sum_{i=1}^{P_{i,kl}-1} f_i \le i\cdot\frac{N}{100}\langle\sum_{i=1}^{P_{i,kl}} f_i \qquad i=1,...,99.$$

Vrednost ***P$_i$*** i=1,...,99 određujemo iz jednačine

$$P_i = L_{P_{i,kl}} + \frac{i \cdot \frac{N}{100} - \sum_{i=1}^{P_{i,kl}-1} f_i}{f_{P_{i,kl}}} \cdot \Delta \ ,$$

gde je:

$L_{P_{i,kl}}$ - početak klase *i*-tog percentila,

$f_{P_{i,kl}}$ - apsolutna frekvencija klase *i*-tog percentila,

N - broj članova populacije,

f_i - apsolutna frekvencija *i*-te klase,

Δ - širina klasnog intervala.

Formule za P_i uzorka su identične uz uslov $N=n$.

Interkvartilna razlika I_q predstavlja interval vrednosti obeležja u kome se nalazi 50% središnih vrednosti uređene serije, i izračunava se po formuli

$$I_q = Q_3 - Q_1 = P_{75} - P_{25} \, .$$

Ona isključuje 25% podataka sa najnižim vrednostima i 25% podataka sa najvišim vrednostima.

U nekim literaturama se često definiše i pojam **polukvartilne razlike (semikvartilna razlika)** i ona je polovina interkvartilne razlike, odnosno važi

$$\textbf{semikvartilna razlika} = \frac{I_q}{2} \, .$$

Primer 1.2.2.2. Za podatke iz **Primera 1.1.1.** izračunati interkvartilnu razliku, treći decil i trideset peti percentil:

a) broja položenih ispita,
b) težine anketiranih studenata.

Rešenje:

Kako je N=50, važi sledeće:

a)

$$Q_1 = x_{\left[1\cdot\frac{50-1}{4}+1\right]} + \left(\left(1\cdot\frac{50-1}{4}+1\right) - \left[1\cdot\frac{50-1}{4}+1\right]\right)\cdot\left(x_{\left[1\cdot\frac{50-1}{4}+2\right]} - x_{\left[1\cdot\frac{50-1}{4}+1\right]}\right) \Rightarrow$$

$$Q_1 = x_{[13.25]} + ((13.25) - [13.25])\cdot(x_{[14.25]} - x_{[13.25]}) \Rightarrow$$
$$Q_1 = x_{13} + 0.25\cdot(x_{14} - x_{13}) = 6 + 0.25(6-6) = 6$$

$$Q_3 = x_{\left[3\cdot\frac{50-1}{4}+1\right]} + \left(\left(3\cdot\frac{50-1}{4}+1\right) - \left[3\cdot\frac{50-1}{4}+1\right]\right)\cdot\left(x_{\left[3\cdot\frac{50-1}{4}+2\right]} - x_{\left[3\cdot\frac{50-1}{4}+1\right]}\right) \Rightarrow$$

$$Q_3 = x_{[37.75]} + ((37.75) - [37.75])\cdot(x_{[38.75]} - x_{[37.75]}) \Rightarrow$$
$$Q_3 = x_{37} + 0.75\cdot(x_{38} - x_{37}) = 9 + 0.75\cdot(9-9) = 9$$

pa je interkvartilna razlika $I_q = Q_3 - Q_1 = 9 - 6 = 3$ *položena ispita.*

Treći decil iznosi:

$$D_3 = x_{\left[3\cdot\frac{50-1}{10}+1\right]} + \left(\left(3\cdot\frac{50-1}{10}+1\right) - \left[3\cdot\frac{50-1}{10}+1\right]\right)\cdot\left(x_{\left[3\cdot\frac{50-1}{10}+2\right]} - x_{\left[3\cdot\frac{50-1}{10}+1\right]}\right) \Rightarrow$$
$$D_3 = x_{[15.7]} + ((15.7) - [15.7])\cdot(x_{[16.7]} - x_{[15.7]}) =$$
$$= x_{15} + (0.7)\cdot(x_{16} - x_{15}) = 6 + 0.7\cdot(6-6) = 6.$$

Trideset peti percentil iznosi:

$$P_{35} = x_{\left[35\cdot\frac{50-1}{100}+1\right]} + \left(\left(35\cdot\frac{50-1}{100}+1\right) - \left[35\cdot\frac{50-1}{100}+1\right]\right)\cdot\left(x_{\left[35\cdot\frac{50-1}{100}+2\right]} - x_{\left[35\cdot\frac{50-1}{100}+1\right]}\right) \Rightarrow$$
$$P_{35} = x_{[18.15]} + ((18.15) - [18.15])\cdot(x_{[19.15]} - x_{[18.15]}) =$$
$$= x_{18} + 0.15\cdot(x_{19} - x_{18}) = 6 + 0.15\cdot(7-6) = 6.15.$$

b)

Kako je $\frac{N}{4}=\frac{50}{4}=12{,}5$ *to je klasa prvog kvartila [60 -70) kg, pa je prvi kvartil* Q_1

$$Q_1 = L_{Q_{1,kl}} + \frac{\frac{N}{4} - \sum_{i=1}^{Q_{1,kl}-1} f_i}{f_{Q_{1,kl}}} \cdot \Delta = 60 + \frac{12{,}5-2}{14} \cdot 10 = 67.5\,kg\,.$$

Kako je $\frac{3\cdot N}{4}=\frac{3\cdot 50}{4}=37{,}5$ *to je klasa trećeg kvartila [80 -90) kg, pa je treći kvartil* Q_3

$$Q_3 = L_{Q_{,kl3}} + \frac{\frac{3\cdot N}{4} - \sum_{i=1}^{Q_{3,kl}-1} f_i}{f_{Q_{3,kl}}} \cdot \Delta = 80 + \frac{37{,}5-29}{9} \cdot 10 \approx 89.44\,kg\,.$$

Interkvartilna razlika je $I_q = Q_3 - Q_1 = 89.44 - 67.5 = 21.94\,kg.$

Kako je $\frac{3\cdot N}{10}=\frac{3\cdot 50}{10}=15$ *to je klasa trećeg decila [60 -70) kg, pa je treći decil* D_3

$$D_3 = L_{D_{3,kl}} + \frac{\frac{3\cdot N}{10} - \sum_{i=1}^{D_{3,kl}-1} f_i}{f_{D_{3,kl}}} \cdot \Delta = 60 + \frac{15-2}{14} \cdot 10 \approx 69.29\,kg.$$

Kako je $\frac{35\cdot N}{100}=\frac{35\cdot 50}{100}=17.5$ *to je klasa trideset petog percentila [70 -80) kg, pa je trideset peti percentil* P_{35}

$$P_{35} = L_{P_{i,kl}} + \frac{35\cdot\frac{N}{100} - \sum_{i=1}^{P_{i,kl}-1} f_i}{f_{P_{i,kl}}} \cdot \Delta = 70 + \frac{17.5-16}{13} \approx 70.12\,kg.$$

- Srednje apsolutno odstupanje

Srednje apsolutno odstupanje (***AD***) možemo odrediti u odnosu na aritmetičku sredinu, modus i medijanu.

Srednje apsolutno odstupanje od aritmetičke sredine označavamo sa $AD(\mu)$ ukoliko je u pitanju populacija, odnosno sa $AD(\overline{x})$ ukoliko je u pitanju uzorak.

Za negrupisane podatke se izračunava po formuli:

$$AD(\mu) = \frac{\sum_{i=1}^{N} |x_i - \mu|}{N} \quad \text{za populaciju,}$$

$$AD(\overline{x}) = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n} \quad \text{za uzorak,}$$

dok se za grupisane podatke izračunava po formuli:

$$AD(\mu) = \frac{\sum_{i=1}^{K} f_i \cdot |x_i^{;} - \mu|}{N} \quad \text{za populaciju,}$$

$$AD(\overline{x}) = \frac{\sum_{i=1}^{K} f_i \cdot |x_i^{;} - \overline{x}|}{n} \quad \text{za uzorak.}$$

Srednje apsolutno odstupanje od modusa označavamo sa $AD(M_o)$ i za negrupisane podatke populacije se izračunava po formuli:

$$AD(M_o) = \frac{\sum_{i=1}^{N} |x_i - M_o|}{N}$$

dok se za grupisane podatke izračunava po formuli:

$$AD(M_o) = \frac{\sum_{i=1}^{K} f_i \cdot |x_i^{'} - M_o|}{N} \quad .$$

Formule za $AD(M_o)$ uzorka su identične uz uslov $N=n$.

Srednje apsolutno odstupanje od medijane označavamo sa $AD(M_e)$ i za negrupisane podatke populacije se izračunava po formuli:

$$AD(M_e) = \frac{\sum_{i=1}^{N} |x_i - M_e|}{N}$$

dok se za grupisane podatke izračunava po formuli:

$$AD(M_o) = \frac{\sum_{i=1}^{K} f_i \cdot |x_i^{'} - M_e|}{N} \quad .$$

Formule za $AD(M_e)$ uzorka su identične uz uslov $N=n$.

Interesantno je napomenuti da zbir $\dfrac{\sum_{i=1}^{N} |x_i - A|}{N}$ ima minimum kada je $A = M_e$, odnosno za negrupisane podatke je srednje apsolutno odstupanje od medijane najmanje.

Primer 1.2.2.3. Za podatke iz **Primera 1.1.1.** izračunati srednje apsolutno odstupanje od aritmetičke sredine

a) broja položenih ispita,
b) težine anketiranih studenata.

Podatke analizirati kao grupisane.

Rešenje:

a)

Srednje apsolutno odstupanje broja položenih ispita od aritmetičke sredine je

$$AD(\mu)=\frac{\sum_{i=1}^{10} f_i\cdot\left|x_i-\mu\right|}{N}=$$

$$=\frac{2\cdot|3-7.44|+5\cdot|4-7.44|+3\cdot|5-7.44|+8\cdot|6-7.44|+9\cdot|7-7.44|}{50}+$$

$$\frac{7\cdot|8-7.44|+6\cdot|9-7.44|+3\cdot|10-7.44|+4\cdot|11-7.44|+3\cdot|12-7.44|}{50}=$$

$$=\frac{2\cdot 4.44+5\cdot 3.44+3\cdot 2.44+8\cdot 1.44+9\cdot 0.44+7\cdot 0.56+6\cdot 1.56}{50}+$$

$$\frac{3\cdot 2.56+4\cdot 3.56+3\cdot 4.56}{50}=1{,}9552$$

b)

Srednje apsolutno odstupanje težine od aritmetičke sredine je

$$AD(\mu)=\frac{\sum_{i=1}^{7} f_i\cdot\left|x_i-\mu\right|}{N}=$$

$$=\frac{2\cdot|55-79.6|+14\cdot|65-79.6|+13\cdot|75-79.6|+9\cdot|85-79.6|+6\cdot|95-79.6|}{50}+$$

$$\frac{4\cdot|105-79.6|+2\cdot|115-79.6|}{50}=$$

$$=\frac{2\cdot 24.6+14\cdot 14.6+13\cdot 4.6+9\cdot 5.4+6\cdot 15.4+4\cdot 25.4+2\cdot 35.4}{50}=12.536\,kg.$$

- Srednje kvadratno odstupanje (varijansa)

Srednje kvadratno odstupanje (varijansu) ćemo posebno izračunavati i obeležavati u zavisnosti od toga da li računamo varijansu populacije ili uzorka. Razlozi zbog kojih se način izračunavanja varijanse populacije razlikuje od načina izračunavanja varijanse uzorka, biće objašnjen u kasnijim poglavljima ovog udžbenika.

- Varijansa populacije

Varijansu populacije obeležavamo sa σ^2 i izračunavamo po formuli:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$ za negrupisane podatke,

$$\sigma^2 = \frac{\sum_{i=1}^{K} f_i \cdot (x_i' - \mu)^2}{N}$$ za grupisane podatke.

Imajući u vidu da važi sledeće:

$$\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{\sum_{i=1}^{N}\left(x_i^2 - 2 \cdot x_i \cdot \mu + \mu^2\right)}{N} = \frac{\sum_{i=1}^{N} x_i^2}{N} - 2 \cdot \mu \cdot \frac{\sum_{i=1}^{N} x_i}{N} + \frac{\sum_{i=1}^{N} \mu^2}{N} =$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - 2 \cdot \mu \cdot \mu + \mu^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2$$

i

$$\frac{\sum_{i=1}^{K} f_i \cdot (x_i^{`} - \mu)^2}{N} = \frac{\sum_{i=1}^{K} f_i \cdot \left(x_i^{`2} - 2 \cdot x_i^{`} \cdot \mu + \mu^2\right)}{N} =$$

$$= \frac{\sum_{i=1}^{K} f_i \cdot x_i^{`2}}{N} - 2 \cdot \mu \cdot \frac{\sum_{i=1}^{K} f_i \cdot x_i^{`}}{N} + \frac{\sum_{i=1}^{K} f_i \cdot \mu^2}{N} =$$

$$= \frac{\sum_{i=1}^{K} f_i \cdot x_i^{`2}}{N} - 2 \cdot \mu \cdot \mu + \mu^2 = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{`2}}{N} - \mu^2$$

varijansa populacije se može izračunavati i po sledećim (tzv. radnim) formulama:

$$\sigma^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2$$ za negrupisane podatke,

$$\sigma^2 = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{`2}}{N} - \mu^2$$ za grupisane podatke.

-Varijansa uzorka

Način izračunavanja i obeležavanja varijanse uzorka zavisi od toga da li je aritmetička sredina populacije (μ), iz koje je uzorak uzet, poznata ili ne.

U slučaju kada je aritmetička sredina populacije **(μ) poznata**, varijansu uzorka obeležavamo sa s^2 i izračunavamo po formuli:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \mu^2$$ za negrupisane podatke,

$$s^2 = \frac{\sum_{i=1}^{K} f_i \cdot (x_i^{`} - \mu)^2}{n} = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{`2}}{n} - \mu^2 \qquad \text{za grupisane podatke.}$$

U slučaju kada je aritmetička sredina populacije **(μ) nepoznata**, varijansu uzorka obeležavamo sa s_c^2 i izračunavamo po formuli:

$$s_c^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2}{n-1} \qquad \text{za negrupisane podatke,}$$

$$s_c^2 = \frac{\sum_{i=1}^{K} f_i \cdot \left(x_i^{`} - \overline{x} \right)^2}{n-1} \qquad \text{za grupisane podatke,}$$

gde je sa $\overline{x}$ obeležena aritmetička sredina ispitivanog uzorka.

Imajući u vidu da važi sledeće:

$$\frac{\sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2}{n-1} = \frac{\sum_{i=1}^{n} \left(x_i^2 - 2 \cdot x_i \cdot \overline{x} + \overline{x}^2 \right)}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - 2 \cdot \overline{x} \cdot \frac{\sum_{i=1}^{n} x_i}{n-1} + \frac{\sum_{i=1}^{n} \overline{x}^2}{n-1} =$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n-1} - 2 \cdot \overline{x} \cdot \frac{n \cdot \overline{x}}{n-1} + \frac{n \cdot \overline{x}^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2}{n-1}$$

i

$$\frac{\sum_{i=1}^{K} f_i \cdot \left(x_i^{\cdot} - \bar{x}\right)^2}{n-1} = \frac{\sum_{i=1}^{K} f_i \cdot \left(x_i^{\cdot 2} - 2 \cdot x_i^{\cdot} \cdot \bar{x} + \bar{x}^2\right)}{n-1} =$$

$$= \frac{\sum_{i=1}^{K} f_i \cdot x_i^{\cdot 2}}{n-1} - 2 \cdot \bar{x} \cdot \frac{\sum_{i=1}^{K} f_i \cdot x_i^{\cdot}}{n-1} + \frac{\sum_{i=1}^{K} f_i \cdot \bar{x}^2}{n-1} =$$

$$= \frac{\sum_{i=1}^{K} f_i \cdot x_i^{\cdot 2}}{n-1} - 2 \cdot \bar{x} \cdot \frac{n \cdot \bar{x}}{n-1} + \frac{n \cdot \bar{x}^2}{n-1} = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{\cdot 2} - n \cdot \bar{x}^2}{n-1}$$

varijansa uzorka se, u slučaju ne poznavanja aritmetičke sredine populacije, može izračunavati i po sledećim formulama:

$$s_c^2 = \frac{\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2}{n-1}$$ za negrupisane podatke,

$$s_c^2 = \frac{\sum_{i=1}^{K} f_i \cdot x_i^{\cdot 2} - n \cdot \bar{x}^2}{n-1}$$ za grupisane podatke.

Primer 1.2.2.4. Za podatke iz **Primera 1.1.1.** izračunati varijansu:

a) broja položenih ispita,
b) težine anketiranih studenata.

Podatke analizirati kao grupisane i smatrati da se radi o populaciji.

Rešenje:

a)

Varijansa broja položenih ispita je

$$\sigma^2 = \frac{\sum_{i=1}^{10} f_i \cdot \left(x_i' - \mu\right)^2}{N} =$$

$$= \frac{2\cdot(3-7.44)^2 + 5\cdot(4-7.44)^2 + 3\cdot(5-7.44)^2 + 8\cdot(6-7.44)^2 + 9\cdot(7-7.44)^2}{50} +$$

$$\frac{7\cdot(8-7.44)^2 + 6\cdot(9-7.44)^2 + 3\cdot(10-7.44)^2 + 4\cdot(11-7.44)^2 + 3\cdot(12-7.44)^2}{50} =$$

$$= \frac{2\cdot(-4.44)^2 + 5\cdot(-3.44)^2 + 3\cdot(-2.44)^2 + 8\cdot(-1.44)^2 + 9\cdot(-0.44)^2}{50} +$$

$$\frac{7\cdot 0.56^2 + 6\cdot 1.56^2 + 3\cdot 2.56^2 + 4\cdot 3.56^2 + 3\cdot 4.56^2}{50} = 5.6864$$

b)

Varijansa težine je

$$\sigma^2 = \frac{\sum_{i=1}^{7} f_i \cdot \left(x_i' - \mu\right)^2}{N} =$$

$$= \frac{2\cdot(55-79.6)^2 + 14\cdot(65-79.6)^2 + 13\cdot(75-79.6)^2 + 9\cdot(85-79.6)^2}{50} +$$

$$\frac{6\cdot(95-79.6)^2 + 4\cdot(105-79.6)^2 + 2\cdot(115-79.6)^2}{50} =$$

$$= \frac{2\cdot(-24.6)^2 + 14\cdot(-14.6)^2 + 13\cdot(-4.6)^2 + 9\cdot 5.4^2}{50} +$$

$$\frac{6\cdot 15.4^2 + 4\cdot 25.4^2 + 2\cdot 35.4^2}{50} = 224{,}84.$$

- Standardna devijacija

Standardna devijacija je kvadratni koren varijanse i obeležava se i izračunava po sledećim formulama:

- standardna devijacija populacije (σ)

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

- standardna devijacija uzorka kada je ***μ* poznato** (s)

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

- standardna devijacija uzorka kada ***μ* nije poznato** (s_c)

$$s_c = \sqrt{s_c^2}\ .$$

Primer 1.2.2.5. Za podatke iz **Primera 1.1.1.** izračunati standardnu devijaciju:

a) broja položenih ispita,
b) težine anketiranih studenata.

Podatke analizirati kao grupisane i smatrati da se radi o populaciji.

Rešenje:

Imajući u vidu rešenje Primera 1.2.2.4. dobijamo:

a)

standardna devijacija broja položenih ispita iznosi

$$\sigma = \sqrt{\sigma^2} = \sqrt{5.6864} \approx 2.38$$

b)

standardna devijacija težine studenata iznosi

$$\sigma = \sqrt{\sigma^2} = \sqrt{224.84} \approx 14.99\,kg.$$

Aritmetička sredina i standardna devijacija u nekoj meri određuju raspored vrednosti obeležja analiziranih podataka u populaciji (uzorku). Naime, na osnovu **Čebiševljeve teoreme**, koju ovde nećemo prezentovati, definisano je sledeće pravilo, poznato kao **Čebiševljevo pravilo**:

u $\pm k$ standardnih devijacija oko aritmetičke sredine nalazi se **bar** $\left(1-\frac{1}{k^2}\right)\cdot 100\%$ vrednosti obeležja podataka analizirane populacije (uzorka).

Ovo pravilo važi za $k>1$, $k \in R$.

Dakle:

- bar $\left(1-\frac{1}{2^2}\right)\cdot 100\% = 75\%$ podataka ima vrednost obeležja u intervalu ± 2 standardne devijacije oko aritmetičke sredine;
- bar $\left(1-\frac{1}{3^2}\right)\cdot 100\% = 88{,}89\%$ podataka ima vrednost obeležja u intervalu ± 3 standardne devijacije oko aritmetičke sredine;
- bar $\left(1-\frac{1}{4^2}\right)\cdot 100\% = 93{,}75\%$ podataka ima vrednost obeležja u intervalu ± 4 standardne devijacije oko aritmetičke sredine.

- Koeficijent varijacije

Koeficijent varijacije obeležavamo sa C_v i izračunavamo po sledećim formulama:

- koeficijent varijacije populacije

$$C_v = \frac{\sigma}{\mu}$$

- koeficijent varijacije uzorka kada je ***μ poznato***

$$C_v = \frac{s}{\overline{x}}$$

- koeficijent varijacije uzorka kada ***μ nije poznato***

$$C_v = \frac{s_c}{\overline{x}}.$$

Kada koeficijent varijacije izražavamo u procentima, obeležavamo ga sa *C_v%* i izračunavamo po formuli

$$C_v\% = C_v \cdot 100\%.$$

Koeficijent varijacije je relativni parametar koji ukazuje na nivo homogenosti posmatranog obeležja u populacija (uzorku). Što je koeficijent varijacije veći, to je homogenost manja, odnosno populacija (uzorak) je nestabilnija u aspektu vrednosti posmatranog obeležja. Pomoću ovog parametra može se upoređivati homogenost različitih obeležja u različitim populacijama (uzorcima). Ovaj parametar, dakle, opisuje svojstvo čitave populacije, pa se često kaže da ima integralni karakter.

Primer 1.2.2.6. Za podatke iz **Primera 1.1.1.** izračunati koeficijent varijacije:

a) broja položenih ispita,
b) težine anketiranih studenata.

Podatke analizirati kao grupisane i smatrati da se radi o populaciji.

Rešenje:

a)

Koeficijent varijacije broja položenih ispita iznosi

$$C_v = \frac{\sigma}{\mu} = \frac{2.38}{7.44} \approx 0.32$$
$$C_v\% = 32\%$$

b)
Koeficijent varijacije težine iznosi

$$C_v = \frac{\sigma}{\mu} = \frac{14.99}{79.6} \approx 0.188$$
$$C_v\% = 18.8\%$$

Raspodeljenost broja položenih ispita ovih studenata nestabilnija je od raspodeljenosti njihovih težina.

- Normalizovano standardno odstupanje

Normalizovano standardno odstupanje je relativni parametar koji ukazuje na odstupanje određenog (*i*-tog) člana populacije (uzorka) u odnosu na aritmetičku sredinu te populacije (uzorka), imajući u vidu i stepen stabilnosti vrednosti obeležja u toj populaciji (uzorka). Pomoću ovog parametra mogu se upoređivati individualni članovi iz različitih populacija.

Normalizovano standardno odstupanje *i*-tog člana, čija je vrednost obeležja x_i, obeležavamo sa Z_i i izračunavamo po sledećim formulama:

- Z_i *i*-tog člana populacije

$$Z_i = \frac{x_i - \mu}{\sigma},$$

- Z_i *i*-tog člana uzorka kada je ***μ poznato***

$$Z_i = \frac{x_i - \overline{x}}{s},$$

- Z_i *i*-tog člana kada ***μ nije poznato***

$$Z_i = \frac{x_i - \overline{x}}{s_c}.$$

Primer 1.2.2.7. Student koji pripada populaciji iz **Primera 1.1.1.** ima 75 kg. i položio je 7 ispita. Naći normalizovano odstupanje:

a) broja položenih ispita,
b) težine

ovog studenta.

Rešenje:

a)

Normalizovano odstupanje broja položenih ispita ovog studenta iznosi

$$Z_i = \frac{x_i - \mu}{\sigma} = \frac{7 - 7.44}{2.38} \approx -0.185 .$$

b)

Normalizovano odstupanje težine ovog studenta iznosi

$$Z_i = \frac{x_i - \mu}{\sigma} = \frac{75 - 79.6}{14.99} \approx -0.307 .$$

Ovaj student je i u jednom i u drugom slučaju ispod proseka svoje populacije, ali više odstupa u težini nego u položenim ispitima.