

3. OSNOVE OPISNE (DESKRIPTIVNE) STATISTIKE

3.3. APSOLUTNE I RELATIVNE
MJERE RASPRŠENJA (DISPERZIJE).
ČEBIŠEVljevo PRAVILO.

3.3.1. MJERE RASPRŠENJA (DISPERZIJE)

- Reprezentativnost srednjih vrijednosti potrebno je procijeniti koristeći odgovarajuće numeričke pokazatelje.
- U tu se svrhu uvode *mjere raspršenja*.
- Njihova je svrha opisati *raspršenost* elemenata *osnovnoga skupa* u odnosu na aritmetičku sredinu, odnosno medijan.
- Zbog toga razlikujemo *apsolutne* i *relativne* mjere raspršenja.
- Kao pokazatelji raspršenosti pogodnije su *relativne* mjere raspršenja.

3.3.2. APSOLUTNE MJERE RASPRŠENJA

- U apsolutne mjere raspršenja ubrajamo:
- *raspon varijacije;*
- *interpercentil;*
- *srednje apsolutno odstupanje;*
- *varijancu (disperziju);*
- *standardnu devijaciju.*
- „Mjerna jedinica” za svaku od tih mjera jednaka je mjernoj jedinici modaliteta promatranoga obilježja.

3.3.3. RASPON VARIJACIJE. INTERPERCENTIL. INTERKVARTIL.

- Raspon varijacije (oznaka: R) je razlika najveće i najmanje vrijednosti elemenata osnovnoga skupa.
- On je relativno gruba i neprecizna mjera raspršenja. *Ne procjenjuje se* za podatke grupirane u razrede.
- Interpercentil je razlika bilo kojih dvaju percentila P_k i P_l uz uvjet $k \geq l$. On predstavlja raspon varijacije središnjih $(k - l)\%$ elemenata niza.
- Ta mjera je bolja od raspona varijacije, a dodatno zahtijeva uzlazno uređivanje polaznoga niza.
- Za $k = 75$ i $l = 25$ dobiva se interkvartil (oznaka: I_q).
- On predstavlja raspon varijacije središnje polovice elemenata niza.

3.3.4. SREDNJE APSOLUTNO ODSTUPANJE. VARIJANCA. STANDARDNA DEVIJACIJA.

- U najčešće korištene *apsolutne* mjere raspršenja pripadaju *srednje apsolutno odstupanje*, *varijanca* i *standardna devijacija*.
- Sve te mjere nastoje opisati raspršenost elemenata osnovnoga skupa u odnosu na aritmetičku sredinu.
- Tradicionalno najčešće korištena mjera je standardna devijacija. Međutim, razvojem računalnih programa sve češće se koristi i srednje apsolutno odstupanje.
- Formule za izračunavanje tih veličina iz negrupiranih podataka, te *procjenu* njihove vrijednosti iz podataka grupiranih u razrede navedene su na sljedećem slideu.

3.3.4. SREDNJE APSOLUTNO ODSTUPANJE. VARIJANCA. STANDARDNA DEVIJACIJA.

Mjera raspršenja	Izračun iz negrupiranih podataka	Izračun (procjena) iz grupiranih podataka
srednje apsolutno odstupanje	$MAD_{\bar{x}} = \frac{ x_1 - \bar{x} + \dots + x_n - \bar{x} }{n} = \sum_{k=1}^n \frac{ x_k - \bar{x} }{n}$	$MAD_{\bar{x}} = \frac{f_1 \cdot x_1 - \bar{x} + \dots + f_n \cdot x_n - \bar{x} }{f_1 + \dots + f_n} = \frac{\sum_{k=1}^n f_k \cdot x_k - \bar{x} }{\sum_{k=1}^n f_k}$
varijanca	$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$	$\sigma^2 = \frac{\sum_{i=1}^k f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i \cdot x_i^2}{\sum_{i=1}^k f_i} - \bar{x}^2$
standardna devijacija	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}$	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^k f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^k f_i}} = \sqrt{\frac{\sum_{i=1}^k f_i \cdot x_i^2}{\sum_{i=1}^k f_i} - \bar{x}^2}$

Napomena: U slučaju grupiranih podataka su x_i i f_i redom razredna sredina i apsolutna frekvencija i -toga razreda.

3.3.5. NAPOMENA

- Standardna devijacija se definira na navedeni, pomalo kompliciran način jer je *zbroj svih „običnih“ odstupanja vrijednosti osnovnoga skupa od aritmetičke sredine* uvijek jednak nuli.
- Zbog toga je nužno promatrati zbroj kvadrata odstupanja elemenata osnovnoga skupa od aritmetičke sredine tih elemenata.
- Analogno razmatranje *ne* vrijedi za srednje apsolutno odstupanje.

3.3.6. RELATIVNE MJERE RASPRŠENJA

- U *relativne* mjere raspršenja ubrajamo:
- koeficijent varijacije (mjera raspršenja u odnosu na aritmetičku sredinu);
- koeficijent kvartilne devijacije (mjera raspršenja u odnosu na medijan).
- Ove mjere *nemaju mjernu jedinicu*, pa se njihove vrijednosti najčešće iskazuju u postotcima.
- One ujedno služe i kao *najbolji pokazatelji varijabiliteta elemenata osnovnoga skupa*.

3.3.7. KOEFICIJENT VARIJACIJE

- Koeficijent varijacije definira se kao *relativno srednje odstupanje* od aritmetičke sredine. Označava se s V i računa prema formuli:

$$V = \frac{\sigma}{x}$$

- Koristeći taj koeficijent, varijabilitet se obično klasificira ovako:

V [%]	0 – 10	10 – 30	30 – 50	50 – 70	≥ 70
interpretacija	vrlo slab	relativno slab	umjeren	relativno jak	jak

3.3.8. KOEFICIJENT KVARTILNE DEVIJACIJE

- Koeficijent kvartilne devijacije predstavlja *relativnu* mjeru raspršenja elemenata osnovnoga skupa u odnosu na medijan. Označava se s V_q i računa prema formuli:

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

- Koristeći taj koeficijent, varijabilitet se obično klasificira ovako:

V_q	0.0 – 0.1	0.1 – 0.2	0.2 – 0.3	0.3 – 0.5	0.5 – 1.0
<i>interpretacija</i>	vrlo slab	relativno slab	umjeren	relativno jak	jak

3.3.9. ČEBIŠEVljeVO PRAVILO

- Jedan koristan kriterij za raspršenost elemenata statističkoga niza dao je ruski matematičar Čebišev.
- Ovdje iskazujemo verziju njegova pravila koja se koristi u opisnoj statistici.

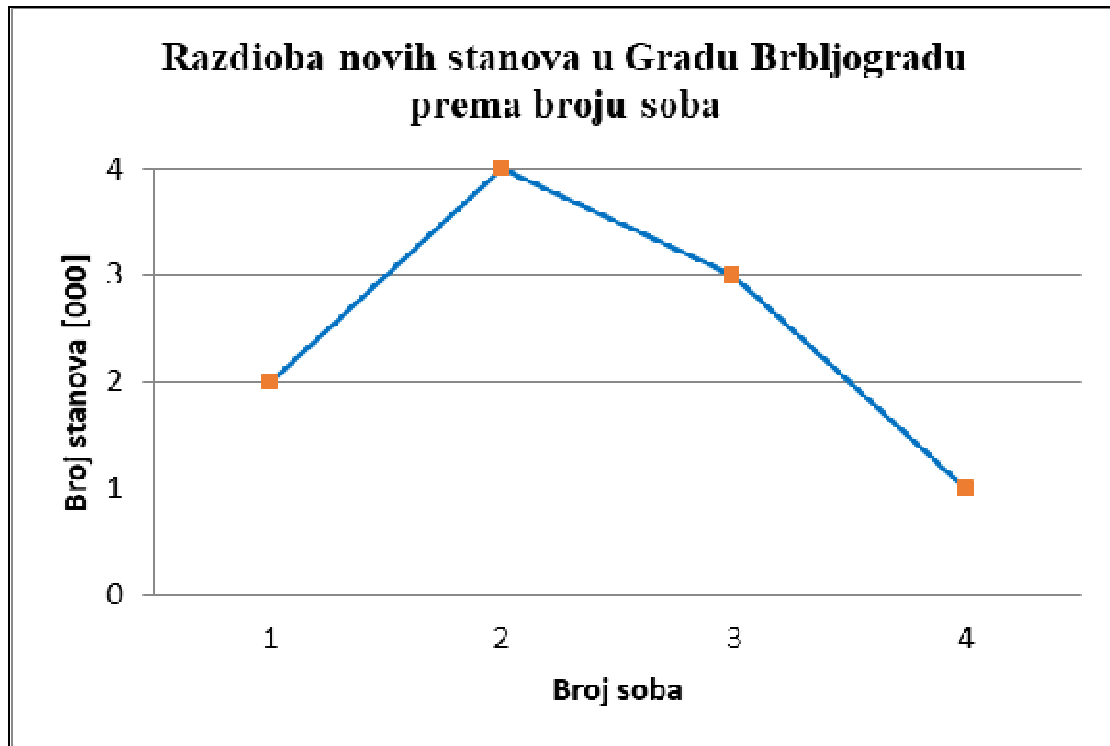
Neka je (x_n) konačan niz vrijednosti kvantitativnoga obilježja. Neka su \bar{x} i σ redom pripadna aritmetička sredina, odnosno standardna devijacija. Tada se za svaki prirodan broj $k > 1$ u segmentu $[\bar{x} - k \cdot \sigma, \bar{x} + k \cdot \sigma]$ nalazi najmanje $100 \cdot \left(1 - \frac{1}{k^2}\right)$ % članova niza (x_n) .

3.3.10. POSLJEDICE ČEBIŠEVljeVA PRAVILA

- Dobivamo ih za $k = 2$, odnosno $k = 3$.
- U segmentu $[\bar{x} - 2 \cdot \sigma, \bar{x} + 2 \cdot \sigma]$ se nalazi najmanje 75% članova niza.
- U segmentu $[\bar{x} - 3 \cdot \sigma, \bar{x} + 3 \cdot \sigma]$ se nalazi najmanje 89% članova niza.
- Posljednji kriterij se može značajno poboljšati u slučaju tzv. *normalne razdiobe*. Tada se dobiva tzv. *pravilo $3 \cdot \sigma$* (o njemu će biti riječi u drugom dijelu predmeta).

Napomena: Sve statističke pokazatelje koji su beskonačni decimalni brojevi zaokružite na dvije decimale.

1. Na slici 1. prikazana je razdioba nastanjenih stanova na području Grada Brbljograda 31. 12. 2022. prema broju soba.




Slika 1.

Izračunajte sljedeće statističke pokazatelje i objasnite njihovo značenje:

- a) raspon varijacije;
- b) interkvartil;
- c) srednje apsolutno odstupanje;
- d) varijanca;
- e) standardna devijacija;
- f) koeficijent varijacije;
- g) koeficijent kvartilne devijacije.

Na temelju rezultata **f)** i **g)** podzadataka odredite je li bolji reprezentant niza brojeva soba aritmetička sredina ili medijan.

 TEHNIČKO VELEUČILIŠTE U ZAGREBU POLYTECHNICUM ZAGABIENSE Elektrotehnički odjel	Vjerojatnost i statistika (stručni prijediplomski studij elektrotehnike)	3.3. Mjere raspšrenja - zadaci
---	---	--

Rješenje:

- a) Računamo: $R = x_{\max} - x_{\min} = 4 - 1 = 3$. Interpretacija ovoga pokazatelja je:

Razlika najmanjega i najvećega broja soba jednaka je 3.

- b) Odredimo najprije donji i gornji kvartil. Imamo redom:

$$\begin{aligned}
 n &= 2 + 4 + 3 + 1 = 10, \\
 Q_1 &= x_{\left\lceil \frac{10}{4} \right\rceil} = x_3 = 2, \\
 Q_3 &= x_{\left\lfloor \frac{3 \cdot 10}{4} \right\rfloor} = x_8 = 3, \\
 I_Q &= Q_3 - Q_1 = 3 - 2 = 1.
 \end{aligned}$$

Interpretacija ovoga pokazatelja je:

Raspon varijacije središnje polovice niza brojeva soba jednak je 1.

- c) Najprije računamo aritmetičku sredinu, a potom *MAD*. Imamo redom:

$$\begin{aligned}
 \bar{x} &= \frac{\sum_{i=1}^4 f_i \cdot x_i}{n} = \frac{2 \cdot 1 + 4 \cdot 2 + 3 \cdot 3 + 1 \cdot 4}{10} = 2.3, \\
 MAD_x &= \frac{\sum_{i=1}^4 f_i \cdot |x_i - \bar{x}|}{n} = \\
 &= \frac{2 \cdot |1 - 2.3| + 4 \cdot |2 - 2.3| + 3 \cdot |3 - 2.3| + 1 \cdot |4 - 2.3|}{10} = 0.76.
 \end{aligned}$$

Interpretacija ovoga pokazatelja je:


Srednje apsolutno odstupanje članova niza brojeva soba od njihove aritmetičke sredine iznosi 0.76.

- d) Koristeći aritmetičku sredinu izračunanu u prethodnom podzadatku, računamo:

$$\sigma^2 = \frac{\sum_{i=1}^4 f_i \cdot x_i^2}{n} - (\bar{x})^2 = \frac{2 \cdot 1^2 + 4 \cdot 2^2 + 3 \cdot 3^2 + 1 \cdot 4^2}{10} - 2.3^2 = 0.81.$$

Interpretacija ovoga pokazatelja je:

Prosjek kvadrata odstupanja članova niza brojeva soba od njihove aritmetičke sredine iznosi 0.81.

 <p>TEHNIČKO VELEUČILIŠTE U ZAGREBU POLYTECHNICUM ZAGABIENSE Elektrotehnički odjel</p>	<p>Vjerojatnost i statistika (stručni prijediplomski studij elektrotehnike)</p>	<p>3.3. Mjere raspršenja - zadaci</p>
--	--	---

e) Koristeći rezultat prethodnoga zadatka odmah dobivamo:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.81} = 0.9.$$

Interpretacija ovoga pokazatelja je:

Prosjek odstupanja članova niza brojeva soba od njihove aritmetičke sredine iznosi 0.9.

f) Koristeći pokazatelje izračunane u c) i e) podzadatku odmah imamo:

$$V = \frac{\sigma}{x} = \frac{0.9}{2.3} \approx 0.3913 = 39.13\%.$$

Interpretacija ovoga pokazatelja je:

Varijabilitet članova niza brojeva soba u odnosu na njihovu aritmetičku sredinu je umjeren.


g) Koristeći pokazatelje izračunane u b) podzadatku odmah imamo:

$$V_Q = \frac{I_Q}{Q_1 + Q_3} = \frac{1}{3 + 2} = 0.2.$$

Interpretacija ovoga pokazatelja je:

Varijabilitet članova niza brojeva soba u odnosu na njihov medijan je relativno slab.

Iz rezultata f) i g) podzadatka slijedi da je medijan bolji reprezentant niza brojeva soba u odnosu na aritmetičku sredinu.

 TEHNIČKO VELEUČILIŠTE U ZAGREBU POLYTECHNICUM ZAGABIENSE Elektrotehnički odjel	Vjerojatnost i statistika (stručni prijediplomski studij elektrotehnike)	3.3. Mjere raspšrenja - zadaci
--	---	--

2. U tablici 1. prikazana je razdioba svih pregleda koje su u Gradu Brbljogradu u 2022. godini obavili liječnici primarne zdravstvene zaštite prema dobnoj skupini pregledanih osoba.

Procijenite koeficijent varijacije i koeficijent kvartilne devijacije, pa objasnite značenje dobivenih pokazatelja.

Tablica 1. Razdioba svih pregleda koje su u Gradu Brbljogradu u 2022. godini obavili liječnici primarne zdravstvene zaštite prema dobnoj skupini pregledanih osoba

<i>Dobna skupina</i> [godina]	0 - 10	10 - 20	20 - 70	70 - 110
<i>Broj osoba</i> [000]	2	5	7	6

Izvor: Statistički ljetopis Grada Brbljograda za 2022. godinu

Rješenje:

Da bismo procijenili tražene pokazatelje, moramo procijeniti aritmetičku sredinu, standardnu devijaciju, prvi i treći kvartil. Procijenimo najprije koeficijent varijacije:

$$n = 2 + 5 + 7 + 6 = 20,$$

$$s_1 = \frac{0+10}{2} = 5, s_2 = \frac{10+20}{2} = 15, s_3 = \frac{20+70}{2} = 45, s_4 = \frac{70+110}{2} = 90,$$

$$\bar{x} = \frac{\sum_{i=1}^4 f_i \cdot s_i}{n} = \frac{2 \cdot 5 + 5 \cdot 15 + 7 \cdot 45 + 6 \cdot 90}{20} = 47,$$

$$\sigma^2 = \frac{\sum_{i=1}^4 f_i \cdot s_i^2}{n} - (\bar{x})^2 = \frac{2 \cdot 5^2 + 5 \cdot 15^2 + 7 \cdot 45^2 + 6 \cdot 90^2}{20} - 47^2 = 988.5,$$

$$\sigma = \sqrt{\sigma^2} \approx 31.44042,$$

$$V = \frac{\sigma}{\bar{x}} \approx 0.6689 = 66.89\%.$$

Interpretacija procijenjenoga pokazatelja je:


Varijabilitet niza dobi pacijenata u odnosu na njihovu aritmetičku sredinu je relativno jak.

U nastavku formiramo niz apsolutnih frekvencija „manje od“:

$$f_1^< = 2,$$

$$f_2^< = 2 + 5 = 7,$$

$$f_3^< = 7 + 7 = 14,$$

 TEHNIČKO VELEUČILIŠTE U ZAGREBU POLYTECHNICUM ZAGABIENSE Elektrotehnički odjel	Vjerojatnost i statistika (stručni prijediplomski studij elektrotehnike)	3.3. Mjere raspšrenja - zadaci
--	---	--

$$f_4^< = n = 20.$$

Za određivanje prvoga kvartila nađemo prvi član gornjega niza koji obuhvaća ukupno $\left\lceil \frac{20}{4} \right\rceil = 5$ podataka. Taj član je očito $f_2^< = 7$. Njemu odgovara drugi razred, odnosno razred 10–20.

Očitamo:

$$L_1 = 10,$$

$$n = 20$$

$$f_1^< = 2,$$

$$f_2 = 5,$$

$$h = 20 - 10 = 10.$$

Zbog toga je prvi kvartil jednak:

$$Q_1 = 10 + \frac{\frac{1}{4} \cdot 20 - 2}{5} \cdot (20 - 10) = 16.$$

Za određivanje trećega kvartila nađemo prvi član gornjega niza koji obuhvaća ukupno $\left\lceil \frac{3}{4} \cdot 20 \right\rceil = 15$ podataka. Taj član je očito $f_4^< = 20$. Njemu odgovara četvrti razred, odnosno razred 70–100.

Očitamo:

$$L_1 = 70,$$

$$n = 20$$

$$f_3^< = 14,$$


$$f_2 = 6,$$

$$h = 110 - 70 = 40.$$

Zbog toga je treći kvartil jednak:

$$Q_3 = 70 + \frac{\frac{3}{4} \cdot 20 - 14}{6} \cdot (110 - 70) = \frac{230}{3}.$$

Prema tome, koeficijent kvartilne devijacije jednak je:

 <p>TEHNIČKO VELEUČILIŠTE U ZAGREBU POLYTECHNICUM ZAGRABIENSE Elektrotehnički odjel</p>	<p>Vjerojatnost i statistika (stručni prijediplomski studij elektrotehnike)</p>	<p>3.3. Mjere raspršenja - zadaci</p>
--	--	---

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{\frac{230}{3} - 16}{\frac{230}{3} + 16} = \frac{91}{139} \approx 0.655.$$

Interpretacija ovoga pokazatelja je:

Varijabilitet članova niza dobi pacijenata u odnosu na njihov medijan je jak.

Možemo zaključiti da je u ovom slučaju aritmetička sredina bolji reprezentant niza dobi u odnosu na medijan toga niza.