3. OSNOVE OPISNE (DESKRIPTIVNE) STATISTIKE

3.2. SREDNJE VRIJEDNOSTI.

3.2.1. SREDNJE VRIJEDNOSTI

- Statistički skup, a samim tim i osnovni skup vrlo često znaju imati vrlo veliki opseg (npr. 3.9 milijuna podataka o jednom obilježju prikupljenih u popisu stanovništva).
- Ako je riječ o kvantitativnom obilježju, takvi podaci se obavezno grupiraju u razrede.
- Ipak, unatoč takvom grupiranju, radi kratkoće i jednostavnosti opisivanja prikupljenih podataka potrebno je imati pokazatelje koji će dovoljno dobro *reprezentirati* osnovni skup (npr. 3.9 milijuna podataka zamijeniti jednim podatkom).
- U tu se svrhu uvode tzv. *srednje vrijednosti.* One se najvećim dijelom odnose isključivo na modalitete *kvantitativnoga* obilježja koje tvore konačan *numerički niz* podataka.

3.2.1. SREDNJE VRIJEDNOSTI

- Srednje vrijednosti dijelimo na:
- **potpune** (u njihovu izračunu sudjeluju svi elementi numeričkoga niza);
- **položajne** (njihova vrijednost ovisi o položaju elemenata u *uzlazno uređenom* numeričkom nizu).
- U potpune srednje vrijednosti pripadaju *aritmetička*, *geometrijska* i *harmonijska sredina*.
- U položajne srednje vrijednosti pripadaju *mod* i *percentili*.

3.2.2. POTPUNE SREDNJE VRIJEDNOSTI

- Zbog jednostavnosti svojega izračuna, najčešće korištena potpuna srednja vrijednost je **aritmetička sredina**. Ona predstavlja prosjek svih vrijednosti elemenata numeričkoga niza. Ona je dobar reprezentant numeričkoga niza *ako taj niz ne sadrži ekstremno velike i/ili ekstremno male vrijednosti*.
- Kod opisivanja prosjeka relativnih promjena neke pojave (npr. godišnja proizvodnja čokolade) koristi se i **geometrijska sredina**. Njezina primjena je najvećim dijelom u računanju tzv. verižnih i skupnih indeksa.
- **Harmonijska sredina** primjenjuje se kad je veličina kojoj se određuje prosječna vrijednost obrnuto razmjerna veličini kojom se određuje prosječna vrijednost (npr. kad se prosječna produktivnost određuje utrošenim vremenom za izradu nekog proizvoda). Mjerna jedinica harmonijske sredine jednaka je mjernoj jedinici varijable za koju se određuje.
- Sve tri navedene vrijednosti mogu se *računati* za negrupirane podatke, odnosno *procijeniti* za grupirane podatke (podatke grupirane u razrede) (vidjeti sljedeći slide).

3.2.2. POTPUNE SREDNJE VRIJEDNOSTI

Srednja vrijednost	Izračun iz negrupiranih podataka	Izračun (procjena) iz grupiranih podataka
aritmetička sredina	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$	$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \ldots + f_k \cdot x_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{\sum_{i=1}^{k} f_i}$
geometrijska sredina	$G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n}$	$G = \sqrt[\sum_{i=1}^{n} f_i]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots\ldots \cdot x_n^{f_n}} = \sqrt[\sum_{i=1}^{n} f_i]{\prod_{i=1}^{n} x_i^{f_i}}$
harmonijska sredina	$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$	$H = \frac{f_1 + f_2 + \ldots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_n}{x_n}} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$

Napomena: U slučaju grupiranih podataka su x_i i f_i redom razredna sredina, odnosno apsolutna frekvencija i-toga razreda, za svaki $i = 1, \ldots, n$.

3.2.3. POLOŽAJNE SREDNJE VRIJEDNOSTI

- Najčešće korištena položajna srednja vrijednost je **mod**. Označava se s *Mo*. To je *modalitet* s najvećom apsolutnom/relativnom frekvencijom. Zbog toga se ta srednja vrijednost može određivati i za *kvalitativna* i za *kvantitativna* obilježja.
- **Oprez**: Mod *nije* jednak apsolutnoj frekvenciji modaliteta, nego se (samo) određuje pomoću apsolutne (ili, eventualno, relativne) frekvencije.
- Za razliku od ostalih srednjih vrijednosti, mod *ne mora postojati* i, ako postoji, *ne mora biti jedinstven*. Tako razdiobe mogu biti **bezmodalne** (nemaju nijedan mod), **unimodalne** (imaju točno jedan mod), **bimodalne** (imaju točno dva različita moda) ili **multimodalne** (imaju barem tri različita moda).
- Ako je razdioba *unimodalna*, onda je mod dobar reprezentant numeričkoga niza ako pripadna *relativna* frekvencija moda nije manja od 50%.
- Analogni kriteriji vrijede i za bimodalnu i multimodalnu razdiobu.

3.2.3. POLOŽAJNE SREDNJE VRIJEDNOSTI

- Ako su podaci grupirani u *prave* razrede, onda vrijednost moda *procjenjujemo* prema sljedećoj formuli:

$$Mo = L_i + \frac{b-a}{2 \cdot b - (a+c)} \cdot s$$

- Ovdje su:
- L_i – donja granica modalnoga razreda;
- b - *korigirana* apsolutna/relativna frekvencija modalnoga razreda (dobivena dijeljenjem originalne frekvencije i originalne razredne širine);
- a – *korigirana* apsolutna frekvencija koja neposredno prethodi frekvenciji b (dobivena analogno kao frekvencija b)
- c - *korigirana* apsolutna/relativna frekvencija koja neposredno slijedi iza frekvencije b (dobivena analogno kao frekvencija b)
- s - razredna širina modalnoga razreda.
- **Napomena**: Ako svi razredi imaju jednaku širinu, *nije* potrebno računati korigirane frekvencije (mogu se koristiti originalne frekvencije).

3.2.4. POLOŽAJNE SREDNJE VRIJEDNOSTI

- Pretpostavimo da je numerički niz *uzlazno uređen*, tj. da su svi numerički podaci poredani od najmanjega prema najvećemu.
- **Percentili** su položajne srednje vrijednosti koje dijele numerički niz na ukupno 100 *jednakobrojnih* dijelova. Ima ih ukupno 99. Označavaju se s P_k, za $k = 1, 2, \ldots, 99$.
- Za svaki $k = 1, \ldots, 99$ vrijednost P_k iz *negrupiranih* podataka računamo prema formuli:

$$P_k = \begin{cases} x_{\left\lceil \frac{k \cdot n}{100} \right\rceil}, & \text{ako } n \text{ nije djeljivo sa 100,} \\ \frac{1}{2} \cdot \left(x_{\frac{k \cdot n}{100}} + x_{\frac{k \cdot n}{100}+1} \right), & \text{ako je } n \text{ djeljivo sa 100.} \end{cases}$$

3.2.4. POLOŽAJNE SREDNJE VRIJEDNOSTI

- Iste vrijednosti iz podataka grupiranih u prave razrede *procjenjujemo* prema formuli:

$$P_k = L_1 + \frac{\frac{k}{100} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$$

- gdje su:
- L_1 – donja granica razreda kojemu pripada P_k;
- m – redni broj razreda (u uzlaznom poretku svih razreda)
- h – širina razreda kojemu pripada P_k;
- f_i – apsolutna/relativna frekvencija i-toga razreda.
- Značenje vrijednosti P_k je:
- k% elemenata niza nije veće od P_k.
- Ekvivalentno, $(100 – k)$% elemenata niza nije manje od P_k.

3.2.4. POLOŽAJNE SREDNJE VRIJEDNOSTI

- U praksi se najčešće koriste 25., 50. i 75. percentil. Oni imaju svoja posebna imena i oznake.
- 25. percentil označava se s Q_1 i naziva **prvi** ili **donji kvartil**.
- 50. percentil označava se s Q_2 ili Me, te naziva **drugi kvartil** ili **medijan**.
- 75. percentil označava se s Q_3 i naziva **treći** ili **gornji kvartil**.
- **Zadatak**: Izrecite teorijsko značenje svakoga kvartila.
- Sve tri istaknute vrijednosti mogu se izračunati iz negrupiranih podataka, te *procijeniti* iz grupiranih podataka (vidjeti sljedeći slide).

3.2.4. POLOŽAJNE SREDNJE VRIJEDNOSTI

Kvartil	Izračun iz negrupiranih podataka	Izračun (procjena) iz grupiranih podataka
Q_1	$Q_1 = \begin{cases} x_{\left\lceil \frac{n}{4} \right\rceil}, \text{ ako } n \text{ nije djeljiv s 4;} \\ \frac{1}{2} \cdot (x_{\frac{n}{4}} + x_{\frac{n}{4}+1}), \text{ ako je } n \text{ djeljiv s 4.} \end{cases}$	$Q_1 = L_1 + \dfrac{\frac{1}{4} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$
$Me =$ Q_2	$Me = Q_2 = \begin{cases} x_{\frac{n+1}{2}}, \text{ ako je } n \text{ neparan;} \\ \frac{1}{2} \cdot (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), \text{ ako je } n \text{ paran.} \end{cases}$	$Me = L_1 + \dfrac{\frac{1}{2} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$
Q_3	$Q_3 = \begin{cases} x_{\left\lceil \frac{3}{4} \cdot n \right\rceil}, \text{ ako } n \text{ nije djeljiv s 4;} \\ \frac{1}{2} \cdot (x_{\frac{3}{4} \cdot n} + x_{\frac{3}{4} \cdot n+1}), \text{ ako je } n \text{ djeljiv s 4.} \end{cases}$	$Q_3 = L_1 + \dfrac{\frac{3}{4} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$

Napomena: n je opseg osnovnoga skupa (duljina numeričkoga niza). L_1 je donja granica pravoga razreda kojemu pripada dotični kvartil, m je redni broj, a h širina toga razreda. f_i je apsolutna/ relativna frekvencija i-toga razreda.

3.2.5. NAPOMENA

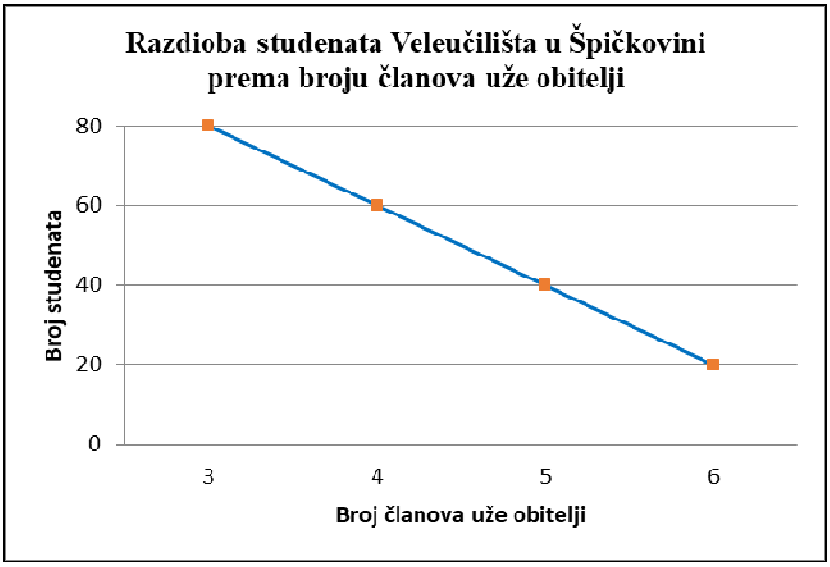
- Reprezentativnost aritmetičke sredine, medijana i kvartila utvrđuje se pomoću posebnih statističkih pokazatelja.
- Riječ je o tzv. *mjerama raspršenja* (disperzije) o kojima će biti govora u sljedećoj točki.
- Također, prilikom procjenjivanja srednjih vrijednosti iz podataka grupiranih u *neprave* (nominalne) razrede potrebno je te razrede pretvoriti u prave, pa potom provesti procjene.
- Postupak pretvorbe nominalnih razreda u prave razrede je sljedeći:
- **Korak 1.** Donja granica prvoga razreda i gornja granica posljednjega razreda ostaju nepromijenjene.
- **Korak 2.** Gornja granica i-toga pravoga razreda jednaka je aritmetičkoj sredini gornje granice i-toga nominalnoga razreda i donje granice (i+1)-voga nominalnoga razreda, za $i = 1, 2, ..., n$.
- **Korak 3.** Donja granica (i+1)-voga razreda jednaka je gornjoj granici i-toga razreda, za $i = 1, 2, ..., n–1$.

Napomena: Sve beskonačne decimalne brojeve zaokružite na dvije decimale.

Za svaku od sljedećih grafički zadanih razdioba odredite i objasnite značenje sljedećih statističkih pokazatelja:

a) aritmetička sredina;
b) mod;
c) donji kvartil;
d) medijan;
e) gornji kvartil.

1.



Slika 1.

Rješenje:

Računamo redom:

$$
\begin{aligned}
&n = 80+60+40+20 = 200,\\
&\bar{x} = \frac{80\cdot 3+60\cdot 4+40\cdot 5+20\cdot 6}{200} = 4,\\
&Mo = 3,\\
&Q_1 = \frac{1}{2}\cdot\left(x_{\frac{200}{4}} + x_{\frac{200}{4}+1}\right) =\\
&\quad = \frac{1}{2}\cdot\left(x_{50}+x_{51}\right) =\\
&\quad = \frac{1}{2}\cdot(3+3) =\\
&\quad = 3,\\
&Me = \frac{1}{2}\cdot\left(x_{\frac{200}{2}} + x_{\frac{200}{2}+1}\right) =\\
&\quad = \frac{1}{2}\cdot\left(x_{100}+x_{101}\right) =\\
&\quad = \frac{1}{2}\cdot(4+4) =\\
&\quad = 4,\\
&Q_3 = \frac{1}{2}\cdot\left(x_{\frac{3}{4}\cdot 200} + x_{\frac{3}{4}\cdot 200+1}\right) =\\
&\quad = \frac{1}{2}\cdot\left(x_{150}+x_{151}\right) =\\
&\quad = \frac{1}{2}\cdot(5+5) =\\
&\quad = 5.
\end{aligned}
$$

Interpretacije izračunanih pokazatelja su redom:

- *Prosječan broj članova obitelji po jednom studentu jednak je* 4.
- *Najčešći broj članova obitelji jednak je* 3.
- 25% *svih studenata ima obitelj s najviše* 3 *člana*.

  Budući da nijedan student nema obitelj s manje od 3 člana, ovu interpretaciju možemo zamijeniti sa:

  25% *svih studenata ima obitelj s točno* 3 *člana*.

Ekvivalentno, 75% *svih studenata ima obitelj s najmanje* 3 *člana*.

Uočite da, opet zato što nijedan student nema obitelj s manje od 3 člana, ovu interpretaciju možemo zamijeniti sa:

Svaki student ima obitelj s najmanje 3 *člana*.

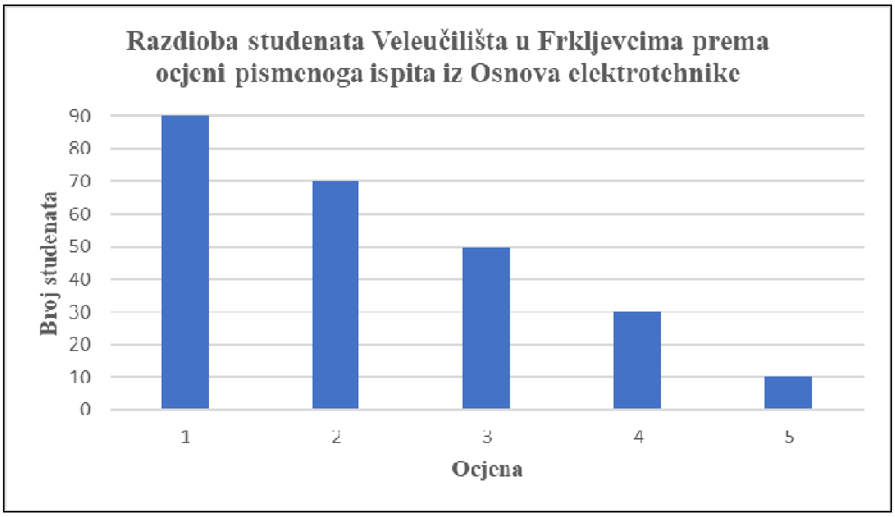
- *Polovica svih studenata ima obitelj s najviše* 4 *člana*.

  Ekvivalentno, *polovica svih studenata ima obitelj s najmanje* 4 *člana*.

- 75% *svih studenata ima obitelj s najviše* 5 *članova*.

  Ekvivalentno, 25% *svih studenata ima obitelj s najmanje* 5 *članova*.

2.


Razdioba studenata Veleučilišta u Frkljevcima prema ocjeni pismenoga ispita iz Osnova elektrotehnike
Broj studenata
0
10
20
30
40
50
60
70
80
90
1
2
3
4
5
Ocjena

Slika 2.

Rješenje:

Računamo redom:

$$
\begin{aligned}
& n = 90+70+50+30+10 = 250, \\
& \bar{x} = \frac{90\cdot 1+70\cdot 2+50\cdot 3+30\cdot 4+10\cdot 5}{250} = 2.2, \\
& Mo = 1, \\
& Q_1 = x_{\left\lceil \frac{250}{4} \right\rceil} = x_{63} = 1, \\
& Me = \frac{1}{2}\cdot\left(x_{\frac{250}{2}} + x_{\frac{250}{2}+1}\right) = \\
& \quad = \frac{1}{2}\cdot\left(x_{125}+x_{126}\right) = \\
& \quad = \frac{1}{2}\cdot(2+2) = \\
& \quad = 2, \\
& Q_3 = x_{\left\lceil \frac{3}{4}\cdot 250 \right\rceil} = x_{188} = 3.
\end{aligned}
$$

Interpretacije izračunanih pokazatelja su redom:

- *Prosjek svih ocjena po jednom studentu jednak je* 2.2.

  Kraće i nepreciznije možemo reći:

  Prosječna ocjena na ispitu je približno jednaka 2.

- *Najčešća ocjena na ispitu iznosi* 1.

- 25% *svih studenata je ocijenjeno ocjenom* 1 *ili manjom od* 1.

  Budući da ne postoji ocjena strogo manja od 1, ovu interpretaciju možemo zamijeniti s:

  25% *svih studenata je ocijenjeno ocjenom* 1.

  Ekvivalentno, 75% *svih studenata je ocijenjeno ocjenom* 1 *ili većom od* 1.

  Budući da ne postoji ocjena strogo manja od 1, ovu interpretaciju možemo zamijeniti s:

  Svaki student je ocijenjen ocjenom 1 *ili većom od* 1.

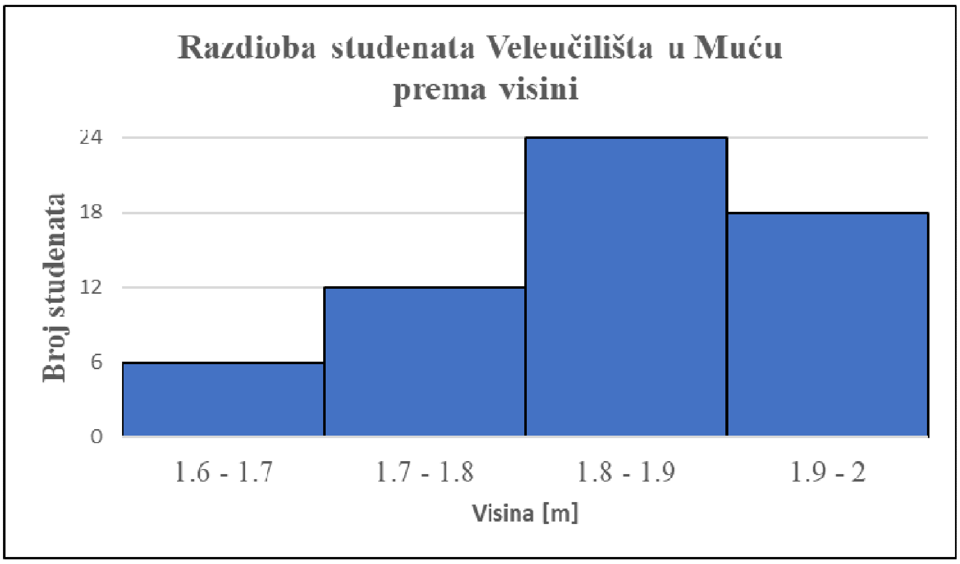
- *Polovica svih studenata je ocijenjena ocjenom* 2 *ili manjom od* 2.

  Ekvivalentno, *polovica svih studenata je ocijenjena ocjenom* 2 *ili većom od* 2.

- 75% *svih studenata je ocijenjeno ocjenom* 3 *ili manjom od* 3.

  Ekvivalentno, 25% *svih studenata je ocijenjeno ocjenom* 3 *ili većom od* 3.

3.


Razdioba studenata Veleučilišta u Muću prema visini
Broj studenata
24
18
12
6
0
1.6 - 1.7
1.7 - 1.8
1.8 - 1.9
1.9 - 2
Visina [m]

Slika 3.

Rješenje:

U ovom zadatku su podaci grupirani u razrede, pa ćemo *procijeniti* svaki pokazatelj.

Da bismo procijenili aritmetičku sredinu, najprije računamo razredne sredine:

$$
\begin{aligned}
&s_1 = \frac{1.6+1.7}{2} = 1.65,\\
&s_2 = \frac{1.7+1.8}{2} = 1.75,\\
&s_3 = \frac{1.8+1.9}{2} = 1.85,\\
&s_4 = \frac{1.9+2}{2} = 1.95,\\
&n = 6+12+24+18 = 60,\\
&\bar{x} = \frac{6\cdot 1.65 + 12\cdot 1.75 + 24\cdot 1.85 + 18\cdot 1.95}{60} = 1.84.
\end{aligned}
$$

Interpretacija ovoga pokazatelja je:

Prosjek visina svih studenata je 1.84 m.

Nadalje, primijetimo da svi pravi razredi imaju jednake širine i one iznose

$$1.7-1.6 = 1.8-1.7 = 1.9-1.8 = 2-1.9 = 0.1.$$

Zbog toga nije potrebno računati korigirane apsolutne frekvencije.

Razred s najvećom apsolutnom frekvencijom je treći razred, tj. $1.8-1.9$ čija apsolutna frekvencija iznosi 24. Taj razred je, dakle, *modalni razred*, tj. razred kojemu pripada procijenjeni mod.

Apsolutna frekvencija modalnom razredu neposredno prethodnoga razreda $(1.7-1.8)$ jednaka je 12.

Apsolutna frekvencija modalnom razredu neposredno sljedećega razreda $(1.9-2)$ jednaka je 18.

Tako je *procijenjeni* mod jednak:

$$Mo = 1.8 + \frac{24-12}{2\cdot 24-(12+18)}\cdot 0.1 \approx 1.87.$$

Interpretacija ovoga pokazatelja je:

Procijenjena najčešća visina svih studenata je 1.87 m.

U nastavku formiramo niz apsolutnih frekvencija „manje od“:

$$\begin{aligned} f_1^{<} &= 6, \\ f_2^{<} &= 6+12=18, \\ f_3^{<} &= 18+24=42, \\ f_4^{<} &= n=60. \end{aligned}$$

Za određivanje prvoga kvartila nađemo prvi član gornjega niza koji obuhvaća ukupno $\left\lceil \frac{60}{4} \right\rceil = 15$ podataka. Taj član je očito $f_2^{<} = 18$. Njemu odgovara drugi razred, odnosno razred $1.7-1.8$. Taj je razred ujedno *kvartilni razred*, tj. razred kojemu pripada procijenjeni prvi kvartil.

Apsolutna frekvencija kvartilnoga razreda $1.7-1.8$ jednaka je 12.

Član niza apsolutnih frekvencija „manje od“ koji neposredno prethodi članu $f_2^{<} = 18$ je $f_1^{<} = 6$.

Širina kvartilnoga razreda jednaka je $1.8-1.7=0.1$.

Tako slijedi da je *procijenjeni* prvi kvartil jednak:

$$Q_1 = 1.7 + \frac{\frac{1}{4}\cdot 60 - 6}{12}\cdot 0.1 = 1.775 \approx 1.78.$$

Interpretacija ovoga pokazatelja je:

25% *svih studenata nisu viši od* 1.78 m.

Ekvivalentno, 75% *svih studenata visoko je barem* 1.78 m.

Za određivanje medijana nađemo prvi član niza apsolutnih frekvencija „manje od“ koji obuhvaća ukupno $\left\lceil \frac{60}{2} \right\rceil = 30$ podataka. Taj član je očito $f_3^{<} = 42$. Njemu odgovara treći razred, odnosno razred $1.8-1.9$. Taj razred je, dakle, *medijalni razred*, tj. razred kojemu pripada procijenjeni medijan.

Apsolutna frekvencija medijalnoga razreda $1.8-1.9$ jednaka je 24.

Član niza apsolutnih frekvencija „manje od“ koji neposredno prethodi članu $f_3^{<} = 42$ je $f_2^{<} = 18$.

Širina medijalnoga razreda jednaka je $1.8-1.7=0.1$.

Tako slijedi da je *procijenjeni* medijan jednak:

$$Me=1.8+\frac{\frac{1}{2}\cdot 60-18}{24}\cdot 0.1=1.85.$$

Interpretacija ovoga pokazatelja je:

Polovica svih studenata nije viša od 1.85 m.

Ekvivalentno, *polovica svih studenata visoka je barem* 1.85 m.

Za određivanje trećega kvartila nađemo prvi član gornjega niza koji obuhvaća ukupno $\left\lceil \frac{3}{4}\cdot 60\right\rceil=45$ podataka. Taj član je očito $f_4^{<}=60$. Njemu odgovara četvrti razred, odnosno razred $1.9-2$. Taj razred je, dakle, *kvartilni razred*, tj. razred kojemu pripada procijenjeni treći kvartil.

Apsolutna frekvencija kvartilnoga razreda $1.9-2$ jednaka je 18.

Član niza apsolutnih frekvencija „manje od" koji neposredno prethodi članu $f_4^{<}=60$ je $f_3^{<}=42$.

Širina kvartilnoga razreda jednaka je $2-1.9=0.1$.

Zbog toga je *procijenjeni* treći kvartil jednak:

$$Q_3=1.9+\frac{\frac{3}{4}\cdot 60-42}{18}\cdot 0.1\approx 1.92.$$

Interpretacija ovoga pokazatelja je:

75% *svih studenata nije više od* 1.92 m.

Ekvivalentno, 25% *svih studenata visoko je barem* 1.92 m.