

### 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. 4.5 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. 4.5 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 jednak je mjerenoj 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	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$	$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_k \cdot x_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k f_i \cdot x_i}{\sum_{i=1}^k f_i}$
geometrijska sredina	$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$	$G = \sqrt[\sum_{i=1}^{f_i}]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n}} = \sqrt[\sum_{i=1}^{f_i}]{\prod_{i=1}^n x_i^{f_i}}$
harmonijska sredina	$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	$H = \frac{f_1 + f_2 + \dots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}} = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n \frac{f_i}{x_i}}$

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

### 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 absolutnoj frekvenciji modaliteta, nego se (samo) određuje pomoću apsolutne (ili, eventualno, relativne) frekvencije.
- Za razliku od ostalih srednjih vrijednosti, mod *ne mora biti jedinstven*. Tako razdiobe mogu biti **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 priпадna *relativna* frekvencija moda nije manja od 50%.
- Analogni kriteriji vrijede i za ostala dva tipa razdioba (zavisno o broju modova).

### 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* absolutna/relativna frekvencija modalnoga razreda (dobivena dijeljenjem originalne frekvencije i originalne razredne širine);
  - $a$  – *korigirana* absolutna frekvencija koja neposredno prethodi frekvenciji  $b$  (dobivena analogno kao frekvencija  $b$ )
  - $c$  - *korigirana* absolutna/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

- Prepostavimo 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  $k = 1, 2, \dots, 99$ .
- Za svaki  $k = 1, \dots, 99$  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  $P_k$ ;
- $f_i$  – absolutna/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
$Q_1$	$Q_1 = \begin{cases} x_{\lceil \frac{n}{4} \rceil}, & \text{ako } n \text{ nije djeljiv s 4;} \\ \frac{1}{2} \cdot (x_{\frac{n}{4}} + x_{\frac{n}{4}+1}), & \text{ako je } n \text{ djeljiv s 4.} \end{cases}$	$Q_1 = L_1 + \frac{\frac{1}{4} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$
$Me = Q_2$	$Me = Q_2 = \begin{cases} x_{\frac{n+1}{2}}, & \text{ako je } n \text{ neparan;} \\ \frac{1}{2} \cdot (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{ako je } n \text{ paran.} \end{cases}$	$Me = L_1 + \frac{\frac{1}{2} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$
$Q_3$	$Q_3 = \begin{cases} x_{\lceil \frac{3}{4} \cdot n \rceil}, & \text{ako } n \text{ nije djeljiv s 4;} \\ \frac{1}{2} \cdot (x_{\frac{3}{4} \cdot n} + x_{\frac{3}{4} \cdot n + 1}), & \text{ako je } n \text{ djeljiv s 4.} \end{cases}$	$Q_3 = L_1 + \frac{\frac{3}{4} \cdot n - \sum_{i=1}^{m-1} f_i}{f_m} \cdot h$

**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  $i = 1, 2, \dots, n$ .
- **Korak 3.** Donja granica  $(i+1)$ -voga razreda jednaka je gornjoj granici  $i$ -toga razreda, za  $i = 1, 2, \dots, n-1$ .