3. SREDNJE VRIJEDNOSTI

(MJERE CENTRALNE TENDENCIJE)

Josipa Perkov, prof. pred.

- **Srednja vrijednost** je konstanta kojom se predstavlja niz varijabilnih podataka
  - Središnja vrijednost oko koje se gomilaju podaci – ***mjera centralne tendencije***

- Srednje vrijednosti se dijele na:

  - **<u>POTPUNE</u>** (koriste se svi podaci):

aritmetička sredina, geometrijska sredina i harmonijska sredina

  - **<u>POLOŽAJNE</u>** (vrijednost je određena položajem u nizu):

mod i medijan

- Primjena određene srednje vrijednosti uvjetovana je vrstom statističke varijable i raspoloživih podataka
- Računaju se samo za <u>varijabilne podatke iste vrste</u>

3.1. ARITMETIČKA SREDINA (AS)

- Najvažnija, najpoznatija i najviše upotrebljavana srednja vrijednost – čest naziv: *prosjek* ili *prosječna vrijednost*
- AS je omjer zbroja svih vrijednosti i broja vrijednosti numeričke varijable

- **JEDNOSTAVNA AS**
  - Primjenjuje se kod negrupiranih podataka
  - Ako numerička varijabla *X* poprima vrijednosti

  $x_1, x_2, \ldots, x_N$ aritmetička sredina $\bar{x}$ dana je izrazom:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \longrightarrow \boxed{\text{veličina u brojniku se naziva } \textbf{total}}$$

PRIMJER 1.

Za 20 zaposlenih poduzeća *X* prikupljeni su podaci o godinama starosti i uređeni po veličini. Oni su iznosili:

19	19	20	20	20	21	22	24	24	25
25	25	28	30	36	36	41	45	53	60

Total iznosi:

$19 + 19 + 20 + 20 + 20 + \ldots + 60 = 593$ godine

(ukupni broj navršenih godina starosti svih 20 radnika)

AS, tj. prosječna starost radnika iznosi $\bar{x} = \frac{593}{20} = 29.65$ godina

VAGANA (PONDERIRANA) AS

- Primjenjuje se kod grupiranih podataka, tj. kada je formirana distribucija frekvencija

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

frekvencije f_i čine **pondere** kojima se mjeri "važnost" svake pojedine vrijednosti varijable X

pojedinačni produkti $x_i f_i$ koji se zbrajaju u brojniku nazivaju se **podtotali**

- Koristi se i za računanje AS distribucije frekvencija za kontinuirana numerička obilježja u kojoj su dani razredi – vrijednost varijable X u razredu predstavlja razredna sredina

- Do istog rezultata možemo doći i korištenjem:
  - relativnih frekvencija kao pondera: $\bar{x} = \sum_{i=1}^{k} x_i p_i$
  - postotnih relativnih frekvencija kao pondera:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i P_i}{100}$$

PRIMJER 2.

Promatrano je 100 vozača koji su vozili automobil 5 godina. Proučavanjem učestalosti prometnih nezgoda tih vozača dobivena je sljedeća tabela:

Broj prometnih nezgoda	Broj vozača
0	20
1	40
2	25
3	9
4 – (7)	6

Izračunajmo prosječan broj prometnih nezgoda po jednom vozaču.

Napomena:

Postupak grupiranja kombiniranjem grupa i razreda

razred 4 – (7) naziva se interval, a procijenjena granica se stavlja u zagrade ()

Broj prometnih nezgoda	**Broj vozača** f_i	**Razredne sredine** x_i	$f_i \cdot x_i$
0	20	0	0
1	40	1	40
2	25	2	50
3	9	3	27
4 – (7)	6	5.5	33
Σ	100		150

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i} = \frac{150}{100} = 1.5 \longrightarrow$$

Prosječan broj prometnih nezgoda po jednom vozaču iznosi 1.5

- Ponekad je moguće i ekonomično izvorne vrijednosti numeričke varijable pojednostavniti smanjivanjem brojčanih vrijednosti

- TRANSFORMACIJA (KODIRANJE) polazi od izraza:

$$d_i = \frac{x_i - a}{b} \quad , \quad b \neq 0 \quad , \quad i = 1, 2, \ldots, N$$

gdje a obično predstavlja vrijednost varijable (razredne sredine) u okolini najvećih frekvencija, a kada su razredi jednakih veličina za $b \neq 0$ je prikladna veličina razreda

$$\bar{x} = a + \frac{b}{N} \sum_{i=1}^{k} f_i d_i \quad \text{ili} \quad \bar{x} = a + b \sum_{i=1}^{k} p_i d_i$$

PRIMJER 3.

Trgovačke radnje poduzeća “X” prema ostvarenom mjesečnom prometu, u 000 kn

Promet u 000 kn	Broj radnji f_i	Razredne sredine x_i	$x_i - a$	d_i	$f_i \cdot d_i$
30 – 40	2	35	-30	- 3	- 6
40 – 50	5	45	-20	- 2	- 10
50 – 60	10	55	-10	- 1	- 10
60 – 70	12	$\boxed{65}$	0	0	0
70 – 90	10	80	15	1. 5	15
90 – 110	9	100	35	3. 5	31. 5
110 – 150	2	130	65	6. 5	13
Σ	50				33. 5

***a* = 65**

razredi nisu jednake veličine, uzmimo da je npr. ***b* = 10**

$$\bar{x} = a + \frac{b}{N}\sum_{i=1}^{k} f_i d_i = 65 + \frac{10}{50}\cdot 33.5 = 71.7 \text{ tisuća kuna}$$

- Raširenost primjene AS potiče iz njezinih svojstava:

(1) zbroj odstupanja vrijednosti varijable *X* od njezine AS je jednak nuli

pojedinačne vrijednosti	distribucija frekvencija
$\sum_{i=1}^{N}\left(x_i-\bar{x}\right)=0$	$\sum_{i=1}^{k} f_i\left(x_i-\bar{x}\right)=0$

(2) zbroj kvadrata odstupanja vrijednosti varijable *X* od AS je minimalan

pojedinačne vrijednosti	distribucija frekvencija
$\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2<\sum_{i=1}^{N}\left(x_i-x_0\right)^2$	$\sum_{i=1}^{k} f_i\left(x_i-\bar{x}\right)^2<\sum_{i=1}^{k} f_i\left(x_i-x_0\right)^2$

(3) **AS uvijek se nalazi između najmanje i najveće vrijednosti varijable:**

$$x_{\min} \leq \bar{x} \leq x_{\max}$$

(4) **Ako su vrijednosti numeričke varijable jednake konstanti *C*, AS te varijable jednaka je toj konstanti:**

$$x_1 = x_2 = \ldots = x_N = C \quad , \quad \bar{x} = C$$

- Ako se raspolaže s aritmetičkim sredinama k podskupova u koje je raspoređeno N_i , $i = 1,2,\ldots,k$ elemenata i ako se podskupovi međusobno ne preklapaju, zajednička sredina za skup, tj. **aritmetička sredina aritmetičkih sredina** izračunava se pomoću izraza:

$$\overline{x} = \frac{\sum_{i=1}^{k} N_i \overline{x}_i}{\sum_{i=1}^{k} N_i}$$

PRIMJER 4.

Prosječna visina 50 studentica iznosi 172 cm, a prosječna visina 80 studenata iznosi 178 cm.

Tada je prosječna visina svih 130 studenata:

$$\overline{x} = \frac{\sum_{i=1}^{2} N_i \overline{x_i}}{\sum_{i=1}^{2} N_i} = \frac{N_1 \overline{x_1} + N_2 \overline{x_2}}{N_1 + N_2} = \frac{50 \cdot 172 + 80 \cdot 178}{50 + 80} = 175.7$$

- Vaganu AS koristimo i kod računanja prosjeka relativnih brojeva

- **Relativni brojevi koordinacije** su omjerni brojevi - nastaju diobom dviju koordinirajućih veličina (veličine koje se uspoređuju)

  dohodak po stanovniku, gustoća stanovništva,...

- Općenito se označavaju izrazom:

$$R_i = \frac{V_i}{B_i} \ , \ i = 1, 2, ..., k$$

V_i = veličina pojave koja se uspoređuje

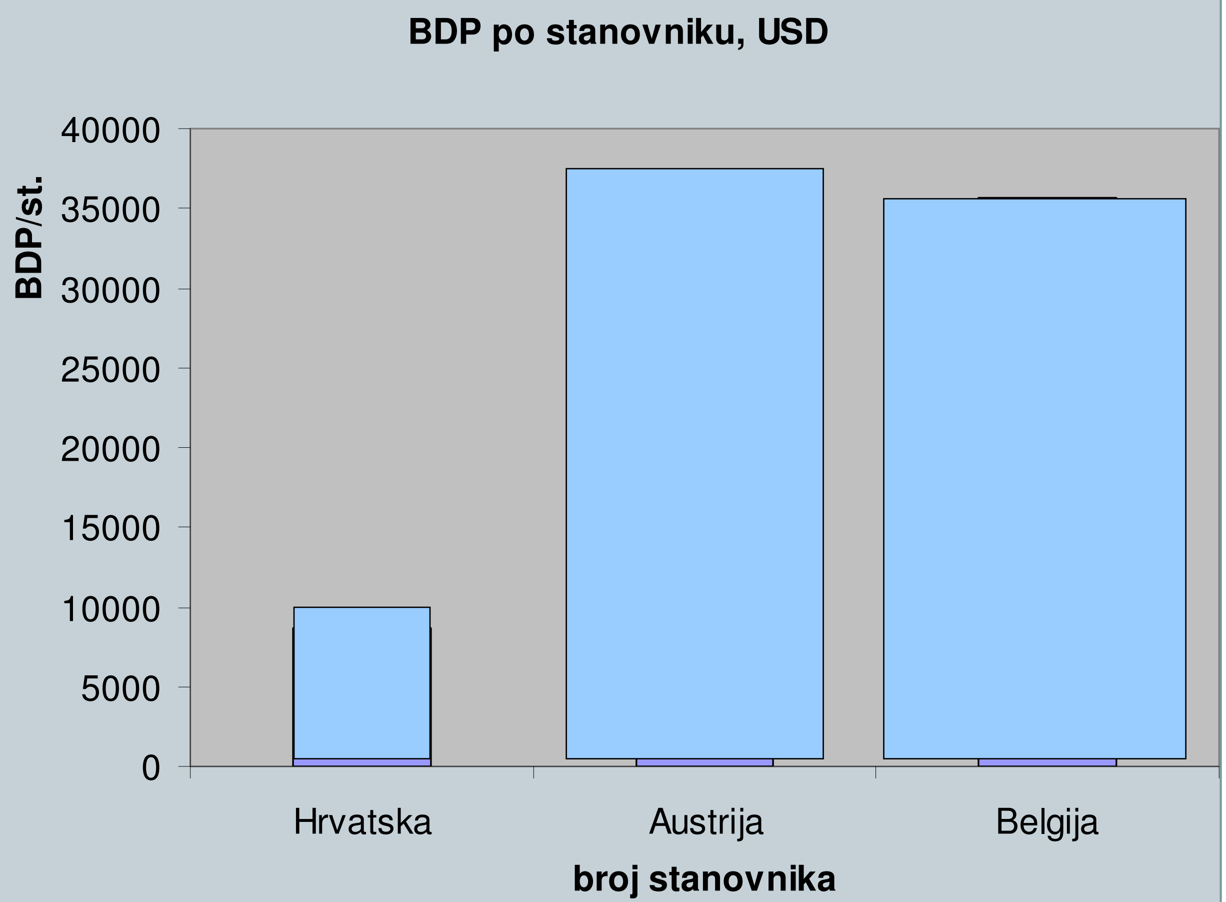
B_i = vrijednosti pojave s kojom se uspoređuje pojava u brojniku

Tabela. BDP po stanovniku u 2005. u Hrvatskoj, Austriji i Belgiji

Država	**BDP po stanovniku, USD (R_i)**	**Broj stanovnika u 000 (B_i)**
Hrvatska	8 675	4 443,9
Austrija	37 117	8 206,5
Belgija	35 712	10 445,9

Izvor: Statističke informacije 2007, DZS, Zagreb 2007.

Relativni brojevi koordinacije prikazuju se grafikonom tako da se na osi ordinata nanosi aritmetičko mjerilo za relativne brojeve koordinacije, a na os apscisa dužine proporcionalne bazama relativnih brojeva



- AS relativnih brojeva koordinacije računa se izrazom:

$$\overline{R} = \frac{\sum_{i=1}^{k} R_i B_i}{\sum_{i=1}^{k} B_i}$$

PRIMJER 5. Uvoz u RH 1999. prema području podrijetla robe i koeficijenti pokrivenosti uvoza izvozom (omjer izvoza i uvoza)

Područje podrijetla	**Uvoz u milijunima USD (B_i)**	**Pokrivenost uvoza izvozom (R_i)**	R_iB_i
Zemlje EU	4392	47.54	208795.68
Zemlje EFTA-e	200	74.00	14800.00
Ostale razvijene zemlje	583	32.42	18900.86
Zemlje u razvoju CEFTA-e	1080	53.80	58104.00
Ostale europske zemlje u razvoju	952	87.50	83300.00
Ostale zemlje u razvoju	569	77.33	44000.77
Σ	**7776**	**-**	**427901.31**

$$\overline{R} = \frac{\sum_{i=1}^{k} R_i B_i}{\sum_{i=1}^{k} B_i} = \frac{427901.31}{7776} = 55.03$$

Na svakih 100 dolara uvoza u prosjeku je 1999. dolazilo 55 dolara izvoza

3.2. GEOMETRIJSKA SREDINA (GS)

- Primjenjuje se u analizi vremenskih nizova
- GS ≤ AS

- GS (jednostavna) vrijednosti $x_1, x_2, \ldots, x_i, \ldots, x_n$ numeričke varijable X dana je izrazom:

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_i \cdot \ldots \cdot x_N} \quad , \; x_i \succ 0 \; , \; \text{za svaki } i$$

- GS (vagana) grupiranih podataka u distribuciju frekvencija dana je izrazom:

$$G = \sqrt[N]{x_1^{\,f_1} \cdot x_2^{\,f_2} \cdot \ldots \cdot x_k^{\,f_k}} \quad , \; N = \sum_{i=1}^{k} f_i$$

3.3. HARMONIJSKA SREDINA (HS)

- Primjena u izračunavanju produktivnosti rada mjerene utroškom vremena po jedinici
- HS < GS ≤ AS

pojedinačne vrijednosti	distribucija frekvencija
$H = \dfrac{N}{\sum_{i=1}^{N} \dfrac{1}{x_i}} \quad , \; x_i \neq 0$	$H = \dfrac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \dfrac{f_i}{x_i}}$

3.4. MOD

- **Mod** je najčešći oblik ili modalitet obilježja (oznaka: $\boldsymbol{M_o}$)
- određuje se i za kvalitativna i za kvantitativna obilježja
- određen je položajem u nizu pa na njega ne djeluju izrazito male ili velike vrijednosti numeričkog niza (za razliku od AS)
- ne može se odrediti ako ne postoje ***bar dvije jednake vrijednosti*** varijable
- Mod niza 1 1 1 1 $\underline{2\ 2\ 2\ 2\ 2\ 2}$ 3 3 3 je 2, tj. $M_o = 2$

PRIMJER 5.

Prikupljajući u jednom uzorku od 200 bračnih parova podatke o broju djece u obitelji, dobili bismo, npr.podatak da je 200 obitelji imalo ukupno 640 djece, što daje aritmetičku sredinu od 3.2 djeteta po obitelji. Puno bolji pokazatelj (reprezentativnija vrijednost) je najčešća vrijednost, tj. najčešći broj djece u obitelji

- Kod distribucije frekvencija diskretne numeričke varijable M_o je vrijednost numeričke varijable s najvećom frekvencijom

PRIMJER 6.

Tabela. Godišnji prosjek zaposlenih u nepoljoprivrednoj djelatnosti u RH 2006.

Vrsta djelatnosti	Broj zaposlenih
Rudarstvo i vađenje	8.844
Prerađivačka industrija	291.886
Opskrba elek. energijom, plinom i vodom	27.214
Građevinarstvo	130.375

Izvor: Statističke informacije 2007, str. 26

Maksimalna frekvencija je 291.886, pa je u ovom slučaju mod prerađivačka industrija

Kod distribucije frekvencija s razredima modalna se vrijednost aproksimira (izravno možemo identificirati samo razred u kojem se nalazi) na sljedeći način:

- Prvo treba pronaći **modalni razred** (razred s najvećom frekvencijom)

- Ako su razredi nejednakih veličina modalni razred je razred s najvećom korigiranom frekvencijom

- **Oznake:**
  - b = najveća (korigirana) frekvencija
  - a = korigirana frekvencija ispred b
  - c = korigirana frekvencija iza b
  - L_1 = donja granica modalnog razreda
  - i = veličina modalnog razreda

- Izraz za aproksimaciju moda: $$M_0 = L_1 + \frac{b-a}{(b-a)+(b-c)} \cdot i$$

PRIMJER 7.

Razredi	Frekvencije		Veličina razreda
20 – 30	2		10
30 – 40	4		10
40 – 50	8	***a***	10
50 – 60	14	***b***	10
60 – 70	9	***c***	10
70 – 80	7		10
80 – 90	5		10
90 - 100	1		10
Σ	50		-

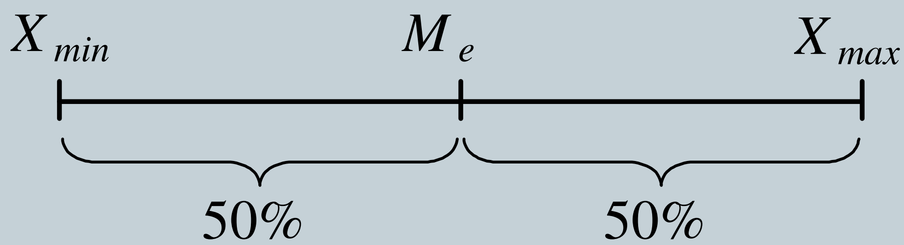
$$M_0 = L_1 + \frac{b-a}{(b-a)+(b-c)} \cdot i$$

$$M_0 = 50 + \frac{14-8}{(14-8)+(14-9)} \cdot 10$$

$$M_0 = 55.45$$

3.5. MEDIJAN

- **Medijan** je vrijednost kvantitativne varijable koja ***uređeni niz*** <u>dijeli na dva jednakobrojna dijela</u> (oznaka: $\boldsymbol{M_e}$)



  - prva polovina članova niza ima vrijednost varijable jednaku ili manju od medijana, a druga polovina članova niza ima vrijednost varijable veću od medijana

- Određen je položajem u nizu

- Medijan M_e pojedinačnih N kvantitativnih vrijednosti varijable X određuje se tako da se one prvo urede po veličini, od najmanje prema najvećoj. Ako je:
  - ***N* neparan broj:**

    M_e je vrijednost varijable središnjeg člana uređenog niza
  - ***N* paran broj:**

    M_e je poluzbroj vrijednosti varijable središnjih dvaju članova uređenog niza

PRIMJER 8.

	N paran broj	N neparan broj
Niz	4,5,6,7,8	4,5,6,7
M_e	6	5.5

- za distribuciju frekvencija s formiranim grupama koristi se kumulativni niz "manje od" – obično se za M_e uzima vrijednost varijable obilježja koje se nalazi na rednom broju $N/2$

PRIMJER 9. Plaće zaposlenika u trgovini *X*

Zaposlenici	1	2	3	4	5
Plaće	3500	3550	3600	3650	12000

Kad bismo "prosjek" računali pomoću aritmetičke sredine dobili bismo da je: $\bar{x} = \frac{3500+3550+3600+3650+12000}{5} = 5260$

što je daleko od stvarnog stanja. U tom slučaju najopravdanije je računati medijan što u ovom primjeru iznosi 3600.

PRIMJER 10. Broj pogrešnih odgovora 80 studenata na testu iz statistike

Broj pogrešnih odgovora	Broj studenata	Kumulativni niz "manje od"
0	5	5
1	7	12
2	15	27
3	19	46
4	20	66
5	10	76
6	4	80
Σ	80	-

$N = 80$, pa je medijan obilježje elemenata s rednim brojevima 40 i 41.

Prva kumulativna frekvencija, jednaka ili veća od 40, jest četvrta po redu (46). Toj grupi pripadaju i 40. i 41. student s istim brojem pogrešnim odgovora pa je $M_e = 3$

- Da bi se odredila vrijednost M_e u distribuciji frekvencija s razredima pretpostavit će se da su članovi niza u **medijalnom razredu** (razred koji sadrži član niza koji zadovoljava definiciju medijana) jednako udaljeni:

$$M_e = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^{m} f_i}{f_{med}} \cdot i$$

- **Oznake:**

$\boldsymbol{L_1}$ = donja granica medijalnog razreda

***N*/2** = polovina članova niza

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

$\boldsymbol{f}_{\mathbf{med}}$ = frekvencija medijalnog razreda

i = veličina medijalnog razreda

PRIMJER 11. Osobe prijavljene u Hrvatskom zavodu za zapošljavanje, stanje potkraj 1999.

Godine života	Broj osoba	Kumulativni niz "manje od"	Veličina razreda
15 – 20	67170	67170	5
20 – 25	48482	115652	5
25 – 30	119819	235471	5
30 – 40	82263	317734	10
40 – 50	10604	328338	10
50 – (65)	13392	341730	(15)
Σ	**341730**	**-**	

$$M_e = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^{m} f_i}{f_{med}} \cdot i$$

$$M_e = 25 + \frac{\frac{341730}{2} - 115652}{119819} \cdot 5$$

$$M_e = 27.3 \approx 27$$

PITANJA ZA USMENI DIO ISPITA:

1. Definirajte potpune srednje vrijednosti
2. Definirajte položajne srednje vrijednosti
3. Definirajte relativne brojeve koordinacije i opišite njihov grafički prikaz.