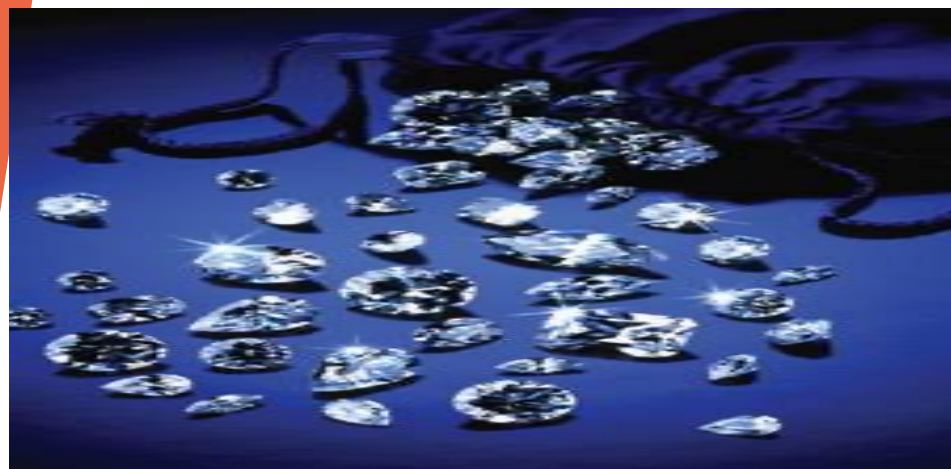


Opisivanje podataka: Mjere disperzije

Poglavlje 4

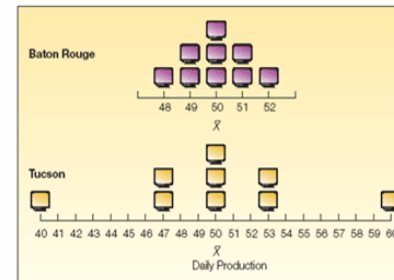
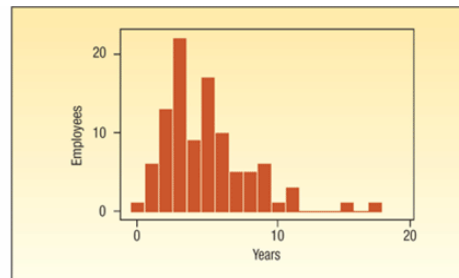


Ciljevi

1. Izračunati i interpretirati *razmak, srednje apsolutno odstupanje, varijansu i standardnu devijaciju* **negrupisanih podataka**.
2. Izračunati i interpretirati *razmak, srednje apsolutno odstupanje, varijansu i standardnu devijaciju* **grupisanih podataka**.
3. Objasniti karakteristike, upotrebu, prednosti i nedostatke svake od navedenih mjera disperzije.

MJERE DISPERZIJE UKRATKO

- ▶ Mjere centralne tendencije, kao što su aritmetička sredina i medijana, samo objašnjavaju središte podataka, međutim ne daju nam objašnjenje niti uvid u raspršenost podataka.
- ▶ Na primjer, ukoliko nam je neko rekao da je prosječna dubina rijeke 1 metar, da li biste prešli preko rijeke bez dodatnih informacija? Vjerovatno ne biste. Potrebno je da saznate nešto više o varijaciji dubine rijeke.
- ▶ Drugi razlog za proučavanje disperzije u skupu podataka je da se uporedi raspon u dvijema ili više distribucija.



- ▶ Interval varijacije (RANGE)

RANGE

Range = Largest value – Smallest value

[3-6]

- ▶ Srednje apsolutno odstupanje (MEAN DEVIATION)

MEAN DEVIATION

$$MD = \frac{\sum |X - \bar{X}|}{n}$$

[3-7]

- ▶ Varijansa (POPULATION VARIANCE)

POPULATION VARIANCE

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

[3-8]

- ▶ Standardna devijacija (POPULATION STANDARD DEVIATION)

POPULATION STANDARD DEVIATION

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

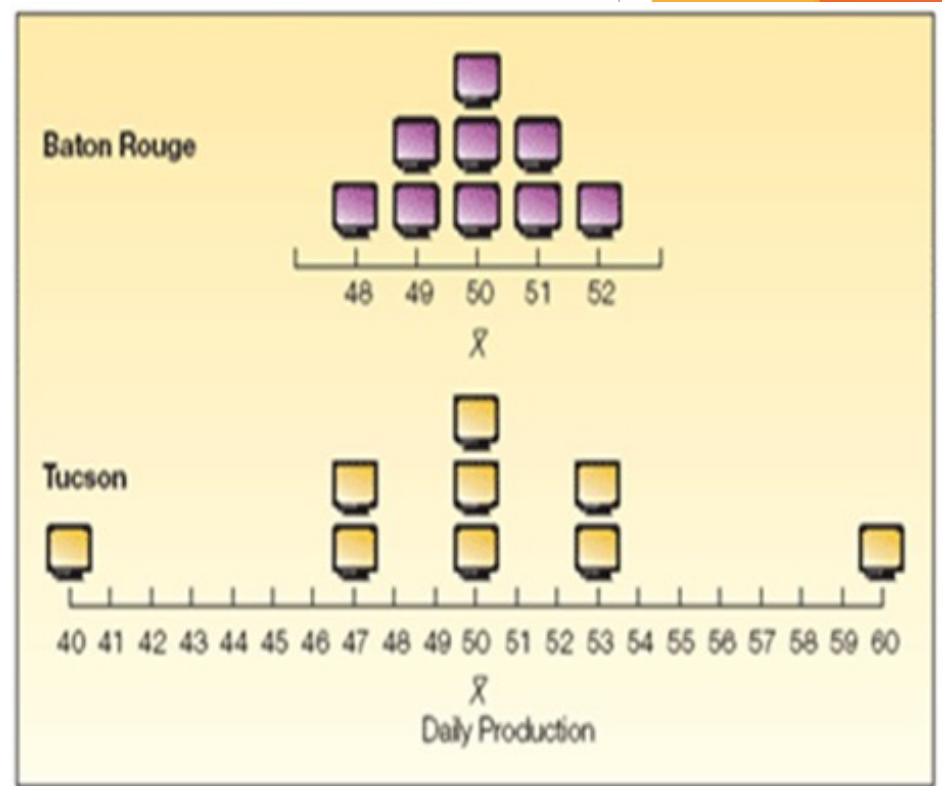
[3-9]

Interval varijacije

Za negrupisane podatke, **interval varijacije** je razlika između najveće i najmanje vrijednosti u skupu podataka.

INTERVAL VARIJACIJE = Najveća vrijednost – Najmanja vrijednost

PRIMJER 1: Izračunati interval varijacije za primjer.



Srednje apsolutno odstupanje

- **Srednje apsolutno odstupanje:** Aritmetička sredina apsolutnih vrijednosti odstupanja od aritmetičke sredine.

$$MD = \frac{\sum |X - \bar{X}|}{n}$$

- **Srednje apsolutno odstupanje** se dobija tako što se zbir apsolutnih odstupanja vrijednosti obilježja od aritmetičke sredine podijeli brojem podataka. Za izračunavanje se koristi sljedeći obrazac:

$$MD = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_i - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

PRIMJER:

Broj prodanih kafa u kafiću Štrudla između 10 i 15 sati, za uzorak od 5 dana od prošle godine, je bio: **20, 40, 50, 60, i 80**. Odredite srednje apsolutno odstupanje broja prodanih kafa.

Korak 1: Izračunati sredinu

$$\bar{x} = \frac{\sum x}{n} = \frac{20 + 40 + 50 + 60 + 80}{5} = 50$$

Korak 2: Oduzeti sredinu od svake pojedinačne vrijednosti i naći apsolutnu vrijednost tako dobijenih razlika.

Korak 3: Sabrati vrijednosti iz koraka 2, i podijeliti ih brojem podataka

$$MD = \frac{\sum |X - \bar{X}|}{n} = \frac{80}{5} = 16$$



Varijansa i standardna devijacija

Varijansa pokazuje prosjek kvadrata odstupanja pojedinačnih vrijednosti obilježja od aritmetičke sredine.

Populaciona varijansa za negrupisane podatke je aritmetička sredina kvadrata odstupanja svake pojedinačne vrijednosti od aritmetičke sredine.



$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Ako se suma kvadrata odstupanja pojedinačnih vrijednosti obilježja od aritmetičke sredine podijeli zbirom podataka N , ili $n-1$ dobija se mjera varijacije koja se u statističkoj teoriji naziva **varijansa** ili **srednje kvadratno odstupanje**.

Ona je mjera varijacije (disperzije) drugog stepena i njena se vrijednost kreće u intervalu $[0, +\infty]$, a to znači da varijansa mora uvijek biti pozitivna veličina.

PRIMJER:

Month	January	February	March	April	May	June	July	August	September	October	November	December
Citations	19	17	22	18	28	34	45	39	38	44	34	10

Broj saobraćajnih prekršaja koji su pokrenuti na sudu u Beaufort County, South Carolina, je dat narednom tabelom. Izračunati populacionu varijansu.

Korak 1: Naći sredinu

Korak 2: Naći razliku između svake opservacije i sredine, i kvadrirati razlike.

Korak 3: Sabrati sve razlike dobijene u koraku 2.

$$\mu = \frac{\sum x}{N} = \frac{19+17+\dots+34+10}{12} = \frac{348}{12} = 29$$

Korak 4: Podijeliti dobijenu sumu kvadrata brojem elemenata u populaciji.

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N} = \frac{1,488}{12} = 124$$

Month	Citations (X)	X - μ	(X - μ) ²
January	19	-10	100
February	17	-12	144
March	22	-7	49
April	18	-11	121
May	28	-1	1
June	34	5	25
July	45	16	256
August	39	10	100
September	38	9	81
October	44	15	225
November	34	5	25
December	10	-19	361
Total	348	0	1,488

- ▶ Alternativna formula za računanje populacione varijanse je:

$$\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2$$

Standardna devijacija

Varijansa je mjera disperzije (varijacije) *drugog stepena* i takvo njeno izračunavanje je nerazumljivo. Da bi se izbjegle ove nejasnoće iz varijanse se izvlači kvadratni korijen i dobija se mjera varijacije prvog stepena koja se naziva **standardna devijacija** ili **standardno odstupanje**.

- ➔ Standardna devijacija pokazuje srednju mjeru odstupanja pojedinačnih vrijednosti obilježja od aritmetičke sredine i uvijek je pozitivna vrijednost ili nula.
- ➔ Izražava se u apsolutnim vrijednostima, a mjeri se jedinicama u kojima je izraženo obilježje, odnosno aritmetička sredina.
- ➔ Ako je $\sigma^2 = 0$ i $\sigma = 0$ onda između pojedinačnih vrijednosti obilježja i aritmetičke sredine nema nikakvih odstupanja.

Standardna devijacija populacije

- ▶ Standardna devijacija populacije (σ) predstavlja kvadratni korijen populacione varijanse.
- ▶ Za **PRIMJER 5**, standardna devijacija populacije je 11.13(kvadratni korijen od 124).

Uzoračka varijansa

- ▶ Uzoračka varijansa koristi se kao ocjena populacione varijanse.

Konceptualna formula = $S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$

Formula za racunanje = $S^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}$

$S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{40}{5 - 1} = 10$ ←

PRIMJER: Dnevne plate pet honorarno zaposlenih radnika u Robnoj kući Zečanka su: **12, 20, 16, 18** i **19** eura. Kolika je uzoračka varijansa?

Dnevna plata	$x - \bar{x}$	$(x - \bar{x})^2$
12	-5	25
20	3	9
16	-1	1
18	1	1
19	2	4
85	0	40

PRIMJERI I RJEŠENJA:

Uzoračka standardna devijacija predstavlja kvadratni korijen uzoračke varijanse.

U prethodnom primjeru, uzoračka standardna devijacija = 3,16.

Primjer 1: Na slučajan način od 100 prodavnica mješovite robe izabrano je osam i popisan dnevni pazar u hiljadama EUR: 8, 10, 12, 15, 20, 29, 26 i 30. Izračunati varijansu.

●Aritmetička sredina iznosi:

$$\bar{x} = 18.75$$

●**Odgovor:** Prosječan dnevni pazar u uzorku od 8 prodavnica iznosi 18 750 EUR.

●Tada je varijansa:

$$s^2 = 76.78$$

SVOJSTVA VARIJANSE I STANDARDNE DEVIJACIJE

- Varijansa i standardna devijacija ne mogu biti negativne i jednake su nuli, samo kada su sve posmatrane vrijednosti jednake.
- Za populacije čiji su podaci približno jednaki aritmetičkoj sredini, varijansa i standardna devijacija će biti male vrijednosti.
- Za populacije čiji su podaci raspršeni, odnosno koji se više razlikuju od aritmetičke sredine, varijansa i standardna devijacija će imati veće vrijednosti..
- Varijansa koristi sve podatke u skupu.

VARIJANSA I STANDARDNA DEVIJACIJA UZORKA

SAMPLE VARIANCE $s^2 = \frac{\sum(X - \bar{X})^2}{n - 1}$ [3-10]

SAMPLE STANDARD DEVIATION $s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$ [3-11]

Gdje je :

s^2 - varijansa uzorka

X - vrijednost svakog pojedinačooog podatka u uzorku

\bar{X} - aritmetičr a sredina uzorka

n - broj podataka u uzorku

PRIMJER

Dnevnicaradnika koji rade parcijalno radno vrijeme u Okovu iznosi redom: €12, € 20, € 16, € 18 i € 19.

Izračunati uzoračku varijansu.

Hourly Wage (X)	$X - \bar{X}$	$(X - \bar{X})^2$
\$12	-\$5	25
20	3	9
16	-1	1
18	1	1
19	2	4
<u>\$85</u>	<u>0</u>	<u>40</u>

$$s^2 = \frac{\sum(X - \bar{X})^2}{n - 1} = \frac{40}{5 - 1} = 10$$

Uzoračka varijansa za grupisane podatke

- ▶ Formula za uzoračku varijansu kod grupisanih podataka, koja se koristi kao ocjena populacione varijanse je:

$$S^2 = \frac{\sum fM^2 - \frac{(\sum fM)^2}{n}}{n - 1}$$

- ▶ Gdje je f frekvencija intervala, a M je sredina intervala.