

Deskriptivna statistika

Vježbe III

23.10.2018.
Nemanja Batrićević

Deskriptivna statistika

- Mjere centralne tendencije
- Mjere varijacije

Centralna tendencija

- **Centralna tendencija** – mjere čije izračunavanje služi određivanju numeričke vrijednosti oko koje se rezultati grupišu
 - Aritmetička sredina – “prosjek”
 - Medijana – “središnja vrijednost”
 - Modus – “najfrekventnija vrijednost”

Centralne tendencije

- Aritmetička sredina** – suma vrijednosti podijeljena brojem vrijednosti.
 - Zavisi od ekstremnih vrijednosti
 - U slučaju agregatnih podataka koristit sredinu intervala i broj observacija
- Medijana** – središnja opservacija nakon što ih poređamo po veličini
 - Ne zavisi od ekstremnih vrijednosti
 - Uzim se središnja vrijednost (neparan broj), ili prosjek dvije središnje vrijednosti (paran)
 - $(n + 1) / 2$ je pozicija u uređenom nizu, ne vrijednost medijane
- Modus** – Vrijednost koja se najčešće pojavljuje
 - I za numeričke i atributivne (nominalne) podatke
 - Multi-modalna distribucija

Mjere Skale	Prosjek	Medijana	Modus
Racio	+	+	(+)
Intervalna	+	+	(+)
Ordinalna	(-)	+	+
Nominalna	-	-	+

Varijacija

- ❑ **Mjere varijacije** – mjere koje opisuju „rasprostranjenost“ (disperziju) podataka.
 - ❑ Raspon
 - ❑ Interkvartalni raspon
 - ❑ Varijansa
 - ❑ Standardna devijacija

Mjere varijacije

Raspon - $X_{max} - X_{min}$

- Zavisi od ekstremnih vrijednosti; distribucija nevažna; problem velikog uzorka

Varijansa – prosječno kvadratno odstupanje od prosjeka

Standardna devijacija - kvadratni korijen iz varijanse

- Najčešće korišćena mjera

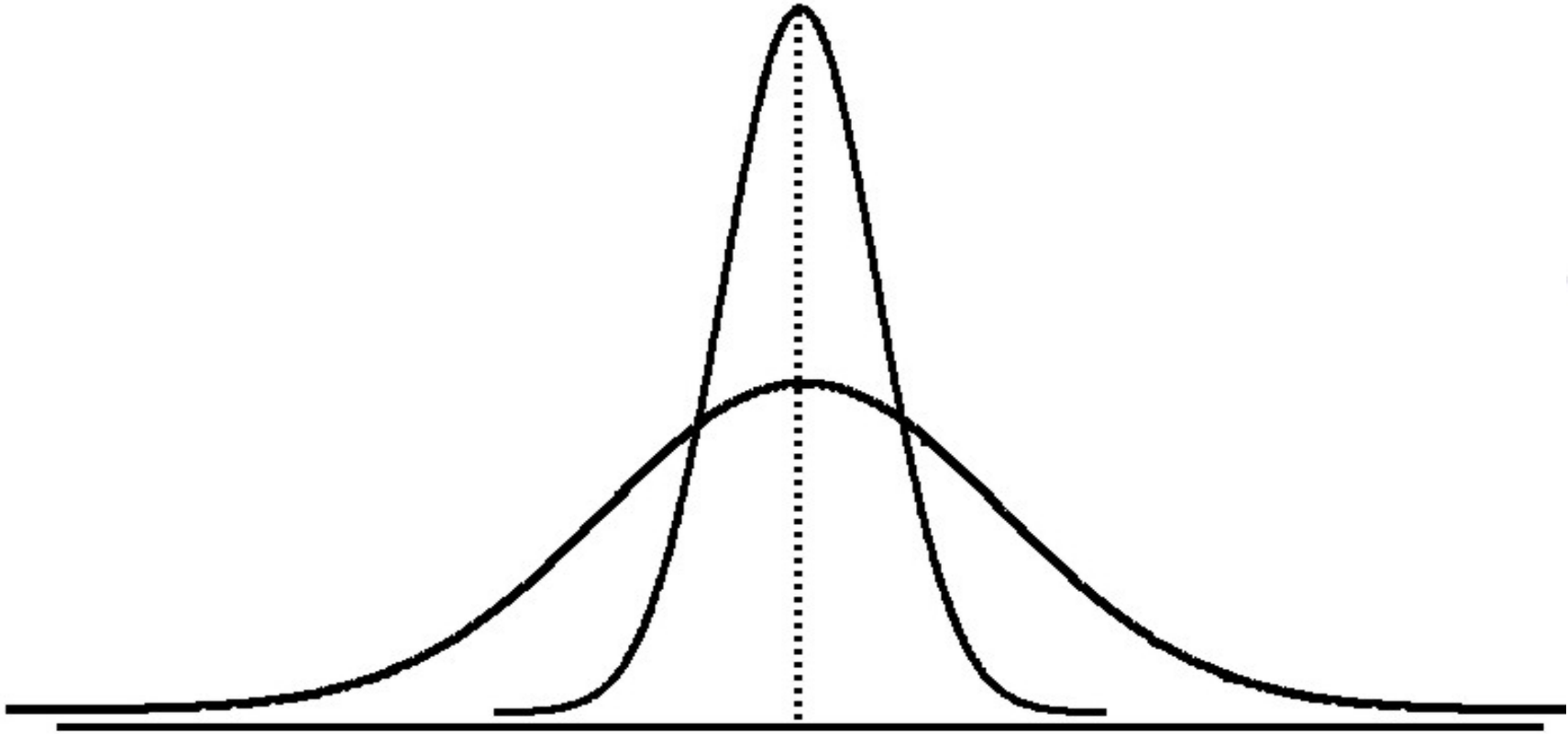
- Izražena u jedinicama kao i originalni podaci

- Empirijsko pravilo: $\pm 1SD$ (68%)

$\pm 2SD$ (95%)

$\pm 3SD$ (99.7%)

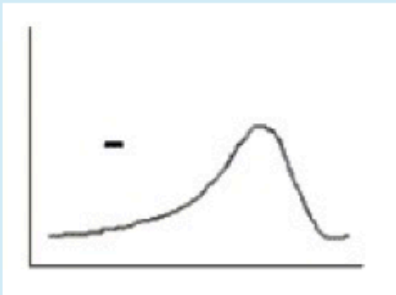
- Z- skor (standardizovano odstupanje)



Tipovi distribucije

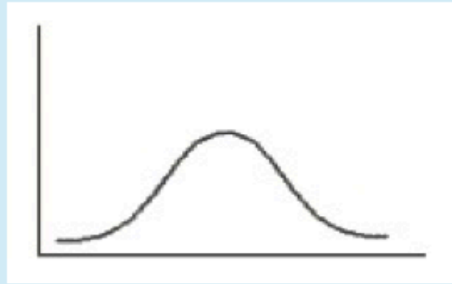
„Iskrivljena“ ulijevo

Prosjeak < Medijana



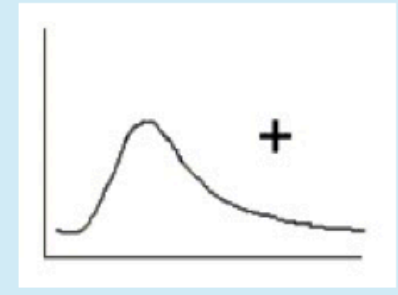
Normalna

Prosjeak = Medijana



„Iskrivljena“ udesno

Prosjeak > Medijana



Vježba II

Istraživač analizira donacije građana prema političkim partijama u zemlji X. Za populaciju od 6 partija izračunaj: aritmetičku sredinu, medijanu, modus, raspon i standardnu devijaciju.

Napomena: jedinica numeričke vrijednosti izražena je u „1.000 eura.“

Vrijednost donacija: 11, 0, 2, 9, 9, 5

Vježba II

Arit. sredina:

Medijana:

Modus:

Vježba II

Arit. sredina:

$$(11+0+2+9+9+5)/6 = 36/6 = 6$$

U prosjeku, partije u zemlji X su primile 6.000 eura donacija od strane građana.

Medijana:

11, 9, 9, 5, 2, 0; niz ima paran broj vrijednosti

Dvije središnje vrijednosti su 9 i 5. Medijanu u ovom slučaju dobijamo iz aritmetičke sredine ova dva broja.

Vrijednost medijane je: 7

Modus:

Donacija:	11	0	2	9	5
Frekvenc.:	1	1	1	2	1

Vježba III

Vrijednost donacije: 11, 0, 2, 9, 9, 5

Raspon:

Varijansa: $(SS = \sum(X - \mu)^2)/N$

Standardna devijacija: korijenovana varijansa = 4

Vježba III

Vrijednost donacije: 11, 0, 2, 9, 9, 5

Raspon:

$$11 - 0 = 11$$

Varijansa: $(SS = \Sigma(X - \mu)^2)/N$

$$\begin{aligned} SS &= (11 - 6)^2 + (0 - 6)^2 + (2 - 6)^2 + (9 - 6)^2 + (9 - 6)^2 + (5 - 6)^2 \\ &= 25 + 36 + 16 + 9 + 9 + 1 = 96 \end{aligned}$$

$$\text{Var} = 96/6 = 16$$

Standardna devijacija: korijenovana varijansa = 4

Izračunavanje stand. dev.

- ❑ **Korak 1:** Pronađi **aritmetičku sredinu**.
- ❑ **Korak 2:** Od svake pojedinačne vrijednosti oduzmi aritmetičku sredinu (**devijaciju**).
- ❑ **Korak 3:** **Kvadriraj** svaku pojedinačnu devijaciju od aritmetičke sredine.
- ❑ **Korak 4:** **Saber** sve kvadrirane devijacije.
- ❑ **Korak 5:** **Podijeli** sa brojem opservacija (veličinom populacije).
- ❑ **Korak 6:** Izračunaj **kvadratni korijen** iz dobijene vrijednosti.