

# POSLOVNA STATISTIKA

---



LJILJANA CVETKOVIĆ

<http://people.dmi.uns.ac.rs/~lila/>



# Osobnosti statističkog zaključivanja

◆ *neizvesnost*



# Osobnosti statističkog zaključivanja

- ◆ *neizvesnost*
- ◆ od *uzorka* ka populaciji



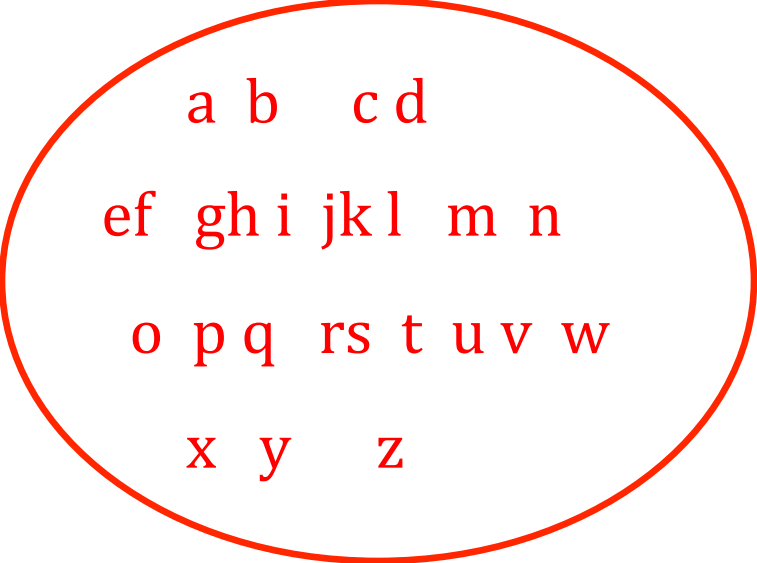
# Osobnosti statističkog zaključivanja

- ◆ *neizvesnost*
- ◆ od *uzorka* ka populaciji
- ◆ bazirano na *numeričkim* karakteristikama



# Populacija vs. uzorak

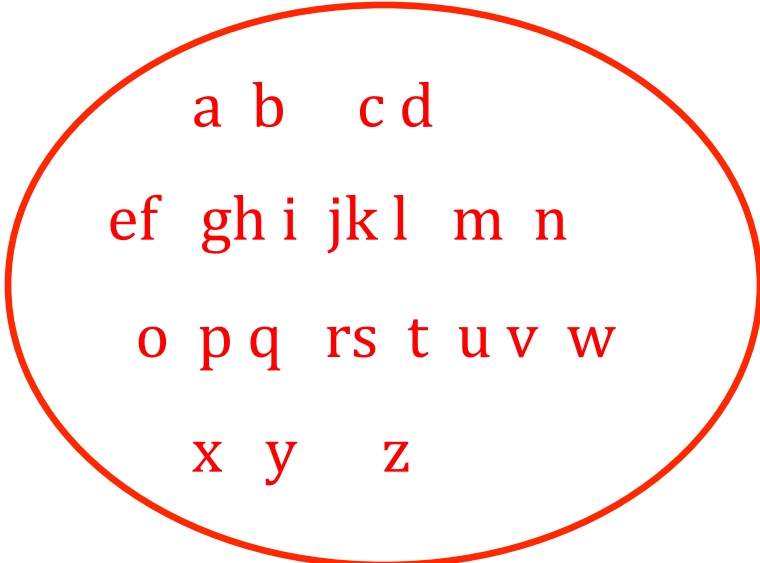
## Populacija



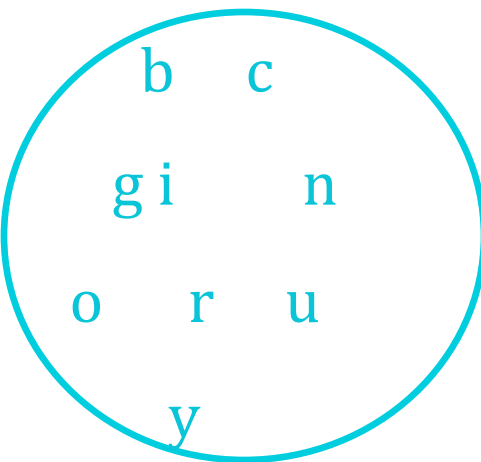


# Populacija vs. uzorak

## Populacija



## Uzorak





# Reprezentativnost uzorka

uzorak je DEO populacije  
treba da bude *reprezentativan*



# Reprezentativnost uzorka

uzorak je DEO populacije  
treba da bude *reprezentativan*

*slučajni* uzorak





# Redosled koraka

1. prikupljanje podataka
2. predstavljanje podataka
3. obrada podataka
4. prelazak na statističko zaključivanje, ako ima dovoljno podataka za to



# DESKRIPTIVNA STATISTIKA

1. prikupljanje podataka
2. predstavljanje podataka
3. obrada podataka
4. prelazak na statističko zaključivanje, ako ima dovoljno podataka za to



## DESKRIPTIVNA STATISTIKA

1. prikupljanje podataka
2. predstavljanje podataka
3. obrada podataka
4. prelazak na statističko zaključivanje, ako ima dovoljno podataka za to

## INFERENCIJALNA STATISTIKA



# Raspodele

## ◆ Empirijske

- podaci dobijeni merenjem (posmatranjem)

## ◆ Teorijske

- izvedene pomoću logičkog i matematičkog razmišljanja

**... statističko zaključivanje počinje sa jednom ili više empirijskih raspodela, a završava se pozivanjem na neku od teorijskih raspodela ...**



# Teorijske raspodele

$X$  – *slučajna promenljiva*

◆ diskretne

$$P(X = x_i) = p_i$$

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

◆ neprekidne

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

$f(x)$  – *funkcija gustine*



# Teorijske raspodele

diskretne

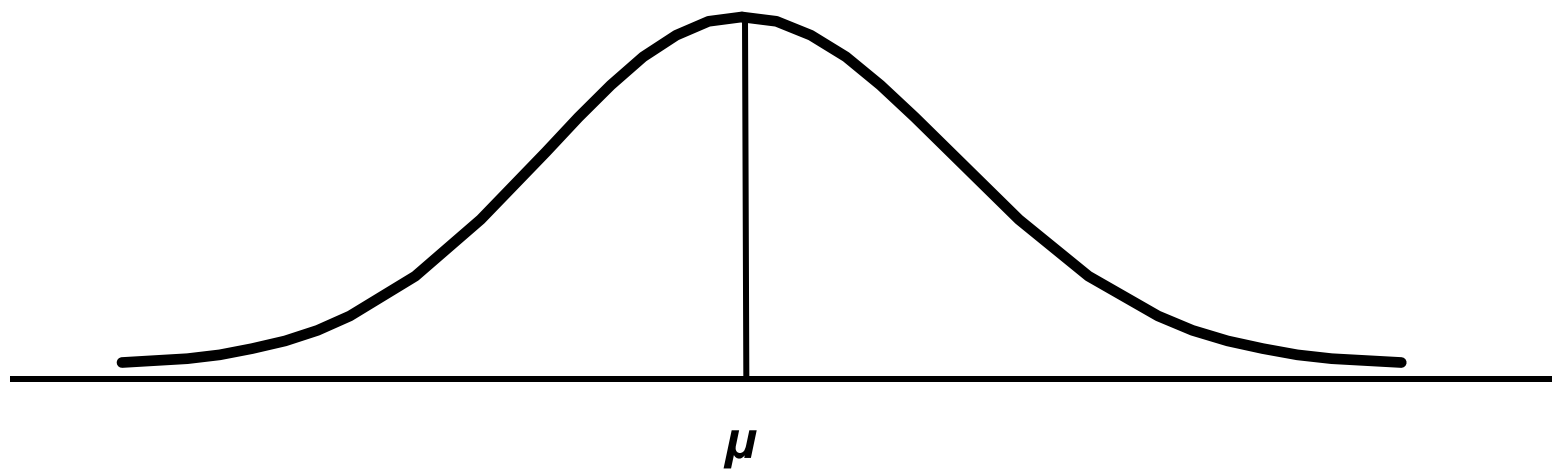
- **Binomna**
- Bernoulli-jeva
- Hipergeometrijska
- Uniformna
- Zipf-ova
- Boltzmann-ova
- Geometrijska
- Logaritamska
- Poisson-ova
- Skellam-ova
- Zeta

neprekidne

- **Normalna**
- Studentova t-raspodela
- Hi kvadrat
- F-raspodela
- Logaritamska
- Trougaona
- Uniformna
- Eksponencijalna
- Gama
- Cauchy-jeva
- Hiperbolična
- Laplace-ova



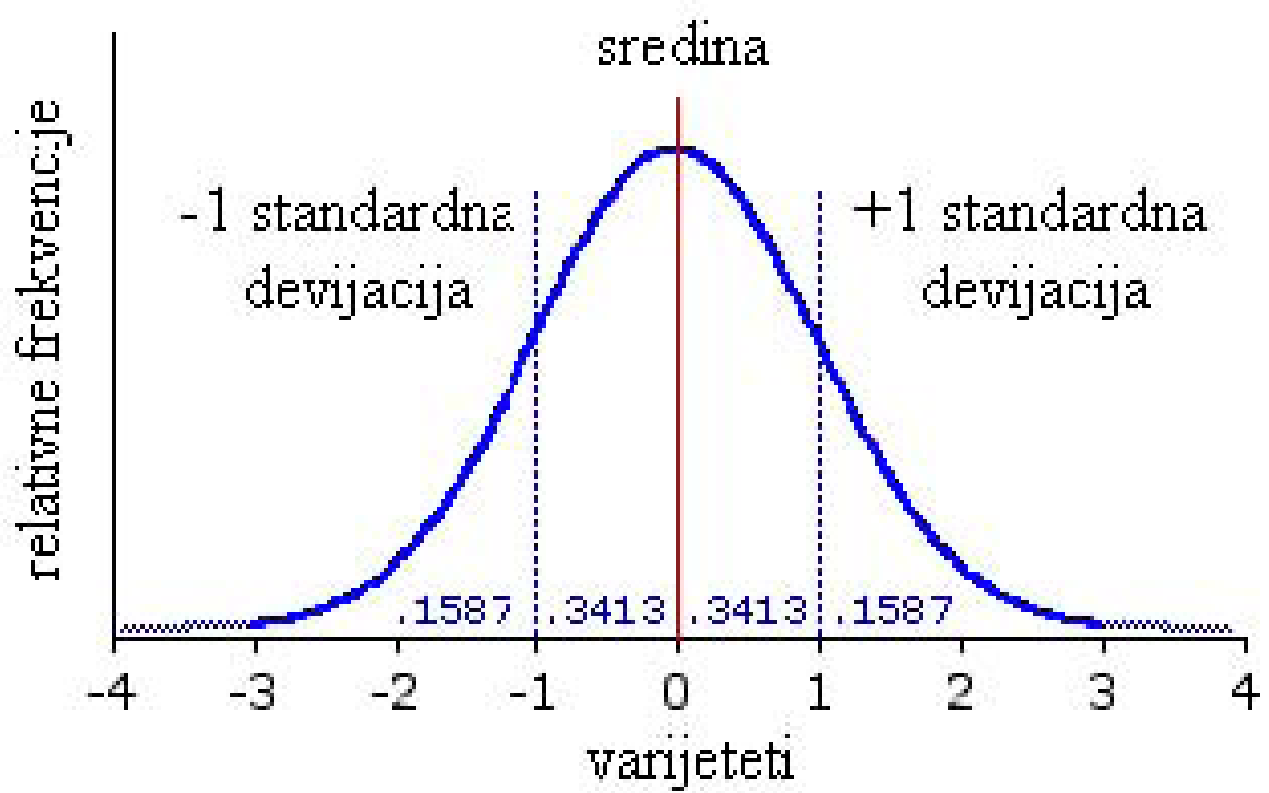
# Funkcija gustine normalne raspodele



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



# Normalna raspodela





# DESKRIPTIVNA STATISTIKA

---



LJILJANA CVETKOVIĆ

<http://people.dmi.uns.ac.rs/~lila/>



# **PRIKUPLJANJE PODATAKA**



# Metode prikupljanja podataka

**eksperiment**

**posmatranje  
merenje**

**intervju**

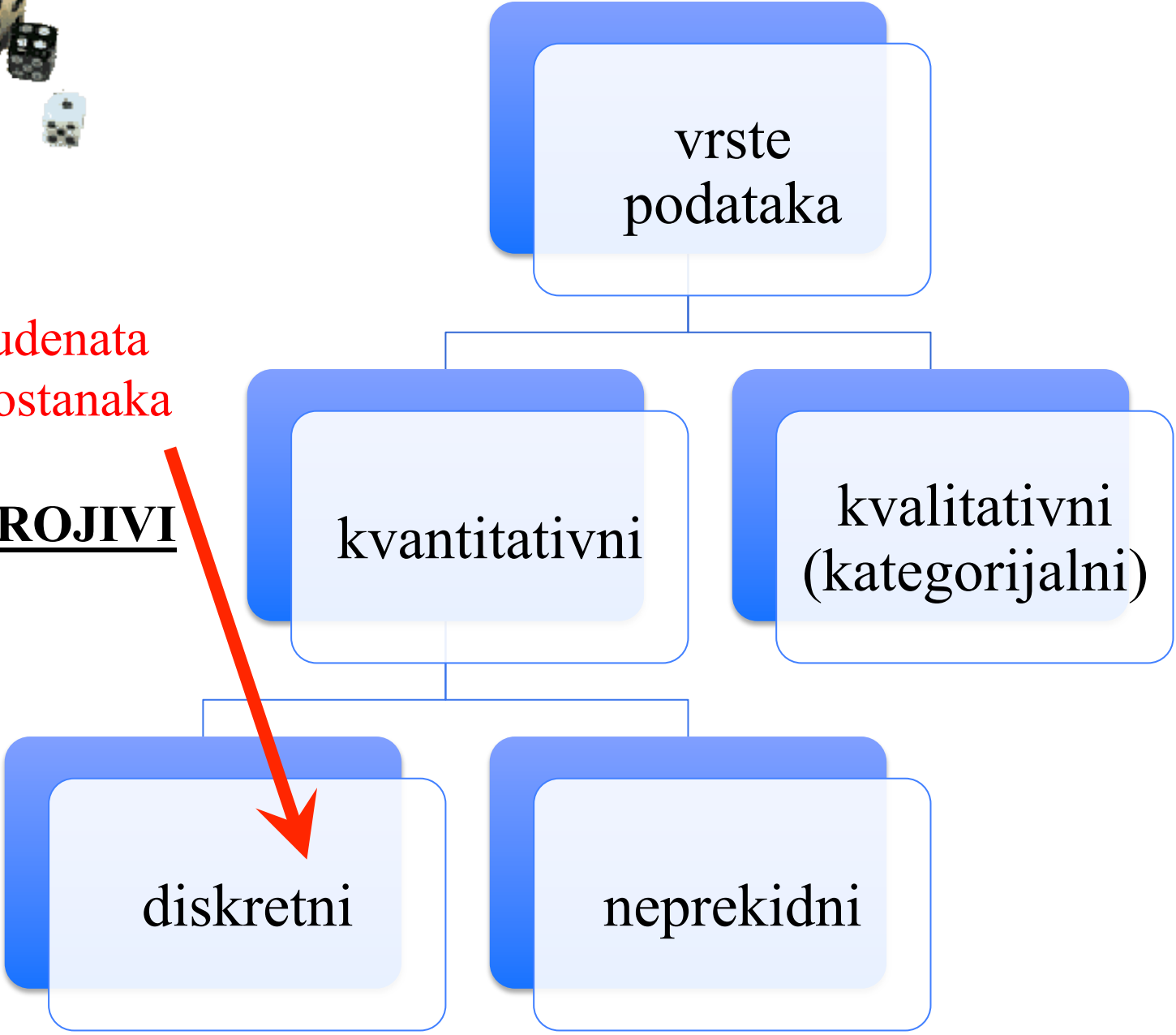
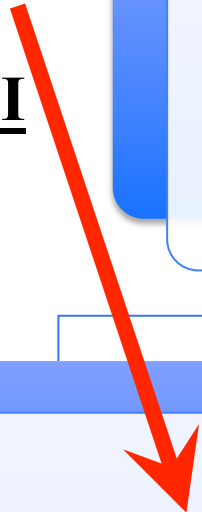
**telefonska  
anketa**

**upitnik**



broj studenata  
broj izostanaka

**PREBROJIVI**





temperatura  
težina

**IZMEREENI**





boja očiju  
bračni status

**SVRSTANI U**  
**KATEGORIJE**





# **PREDSTAVLJANJE PODATAKA**

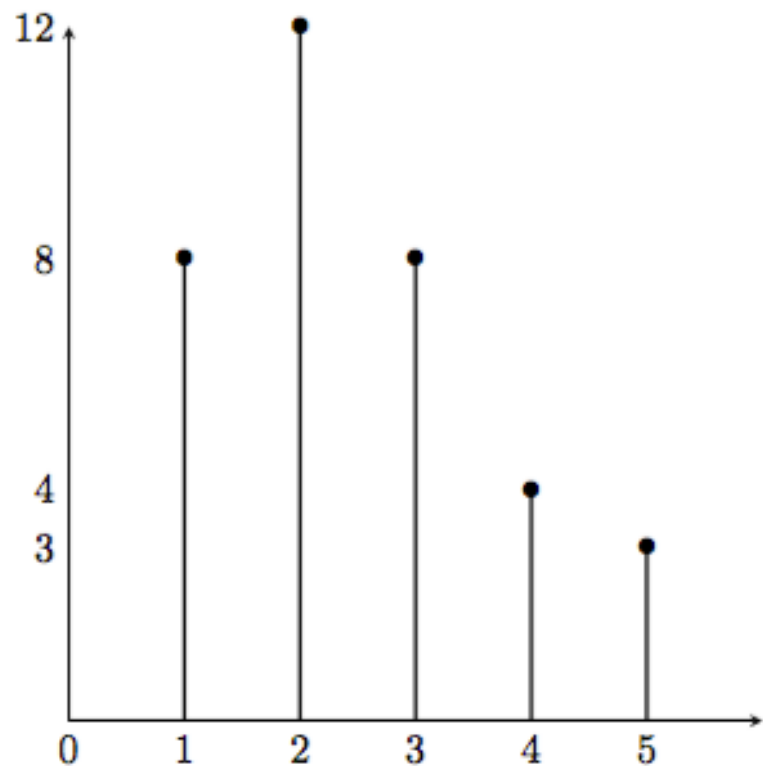


# Predstavljanje diskretnih podataka

tabelarno

$x_i$	$f_i$
1	8
2	12
3	8
4	4
5	3
$\Sigma$	35

grafički





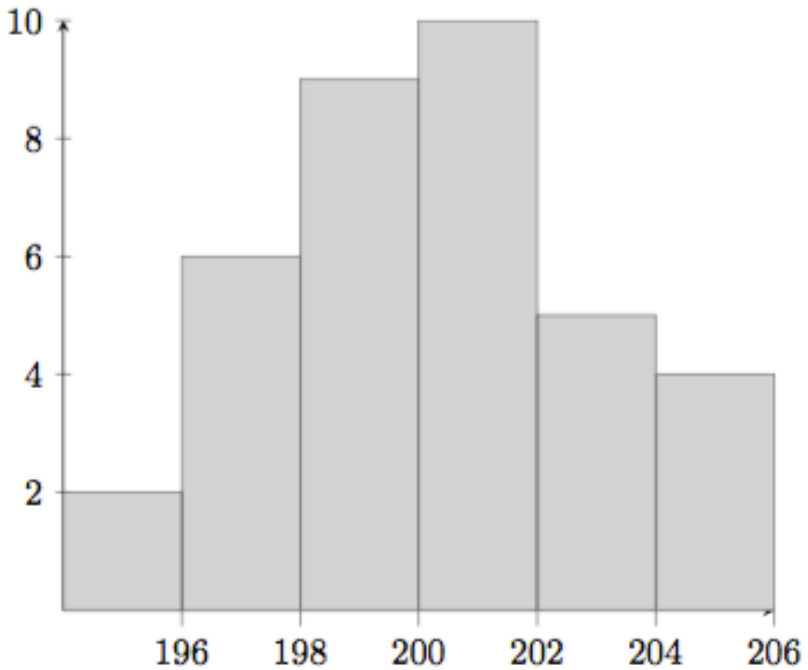


# Predstavljanje neprekidnih podataka

tabelarno

Interval	$f_i$
[ 194 , 196 )	2
[ 196 , 198 )	6
[ 198 , 200 )	9
[ 200 , 202 )	10
[ 202 , 204 )	5
[ 204 , 206 ]	5
$\Sigma$	37

grafički

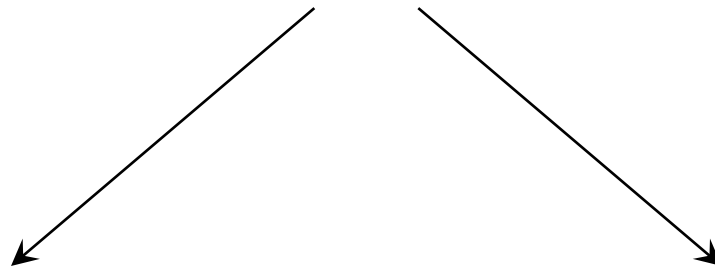




# **OBRADA PODATAKA**



# Brojne karakteristike obeležja



na populaciji

**parametri**

na uzorku

**statistike**



# Brojne karakteristike

- ◆ mere centralne tendencije
  - sredina
  - mod
  - medijana
- ◆ mere odstupanja (rasipanja)
  - varijansa
  - standardna devijacija
  - koeficijent varijacije
  - srednje apsolutno odstupanje
- ◆ mere pozicije
  - kvartili, percentili, ...



# Mere centralne tendencije

- ◆ sredina
- ◆ mod
- ◆ medijana



# MCT – sredina

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{n}$$

$x_i$  - varijeteti

$f_i$  - njihove frekvencije

$n$  - veličina uzorka



# Primer

20 studenata FABUS-a, koji su odslušali statistiku, dobilo je sledeće ocene na ispitu:



# Primer

20 studenata, koji su odslušali statistiku,  
dobilo je sledeće ocene na ispitu:

**8, 7, 7, 8, 6, 6, 9, 7, 7, 10,  
7, 8, 8, 8, 9, 10, 6, 8, 9, 9**







**8, 7, 7, 8, 6, 6, 9, 7, 7, 10,  
7, 8, 8, 8, 9, 10, 6, 8, 9, 9**

Ocene ( $x_i$ )	Frekvencije ( $f_i$ )
6	
7	
8	
9	
10	



8, 7, 7, 8, 6, 6, 9, 7, 7, 10,  
7, 8, 8, 8, 9, 10, 6, 8, 9, 9

Ocene ( $x_i$ )	Frekvencije ( $f_i$ )
6	3
7	
8	
9	
10	



**8, 7, 7, 8, 6, 6, 9, 7, 7, 10,**  
**7, 8, 8, 8, 9, 10, 6, 8, 9, 9**

Ocene ( $x_i$ )	Frekvencije ( $f_i$ )
6	3
7	5
8	
9	
10	



**8, 7, 7, 8, 6, 6, 9, 7, 7, 10,**  
**7, 8, 8, 8, 9, 10, 6, 8, 9, 9**

Ocene ( $x_i$ )	Frekvencije ( $f_i$ )
6	3
7	5
8	6
9	
10	



8, 7, 7, 8, 6, 6, 9, 7, 7, 10,  
7, 8, 8, 8, 9, 10, 6, 8, 9, 9

Ocene ( $x_i$ )	Frekvencije ( $f_i$ )
6	3
7	5
8	6
9	4
10	



8, 7, 7, 8, 6, 6, 9, 7, 7, **10**,  
7, 8, 8, 8, 9, **10**, 6, 8, 9, 9

Ocene ( $x_i$ )	Frekvencije ( $f_i$ )
6	3
7	5
8	6
9	4
10	<b>2</b>



# Statistička tabela

$x_i$	$f_i$
6	3
7	5
8	6
9	4
10	2
	20





# MCT – sredina

$x_i$	$f_i$	$x_i f_i$
6	3	18
7	5	35
8	6	48
9	4	36
10	2	20
	20	157



# MCT – sredina

$x_i$	$f_i$	$x_i f_i$
6	3	18
7	5	35
8	6	48
9	4	36
10	2	20
	20	157

$$\bar{x} = \frac{157}{20} = 7.85$$



# MCT – mod

varijetet koji ima  
najveću frekvenciju



# MCT – mod

varijetet koji ima najveću frekvenciju

$x_i$	$f_i$
6	3
7	5
8	6
9	4
10	2



# MCT – mod

varijetet koji ima najveću frekvenciju

mod = 8

$x_i$	$f_i$
6	3
7	5
8	6
9	4
10	2



# MCT – medijana

vrednost obeležja  
tačno na sredini...



# MCT – medijana

vrednost obeležja  
tačno na sredini...

$x_i$	$f_i$
6	3
7	5
8	6
9	4
10	2
	20



# MCT – medijana

**$F_i$**  - kumulativne frekvencije

$x_i$	$f_i$	$F_i$
6	3	3
7	5	8
8	6	14
9	4	
10	2	
	20	





# MCT – medijana

**$F_i$**  - kumulativne  
frekvencije

deseti i jedanaesti u  
nizu su na sredini

$$\text{medijana} = (8+8)/2 = 8$$

$x_i$	$f_i$	$F_i$
<b>6</b>	3	3
<b>7</b>	5	8
<b>8</b>	6	14
<b>9</b>	4	
<b>10</b>	2	
	20	



# Oznake...

parametar / statistika	populacija	uzorak
sredina		



# Oznake...

parametar / statistika	populacija	uzorak
sredina	$\mu$	$\bar{X}$



# Mere odstupanja

- ◆ varijansa (disperzija)
- ◆ standardna devijacija
- ◆ koeficijent varijacije
- ◆ srednje apsolutno odstupanje



# Zašto mere odstupanja?

- ◆ 4 bebe od godinu dana i 6 baka od 66 god.
- ◆ 5 žena od 39 god. i 5 muškaraca od 41 god.



# Zašto mere odstupanja?

- ◆ 4 bebe od godinu dana i 6 baka od 66 god.

$x_i$	$f_i$	$x_i f_i$
1	4	4
66	6	396
	10	400

- ◆ 5 žena od 39 god. i 5 muškaraca od 41 god.

$x_i$	$f_i$	$x_i f_i$
39	5	195
41	5	205
	10	400



# Zašto mere odstupanja?

- ◆ 4 bebe od godinu dana i 6 baka od 66 god.

$x_i$	$f_i$	$x_i f_i$
1	4	4
66	6	396
	10	400

$$\bar{x} = \frac{400}{10} = 40$$

- ◆ 5 žena od 39 god. i 5 muškaraca od 41 god.

$x_i$	$f_i$	$x_i f_i$
39	5	195
41	5	205
	10	400

$$\bar{x} = \frac{400}{10} = 40$$



# Pojedinačna odstupanja

$x_i$	$x_i - \bar{x}$	$f_i$
6		3
7		5
8		6
9		4
10		2





# Pojedinačna odstupanja

$x_i$	$x_i - \bar{x}$	$f_i$
6	- 1.85	3
7	- 0.85	5
8	0.15	6
9	1.15	4
10	2.15	2



# Pojedinačna odstupanja

$x_i$	$x_i - \bar{x}$	$f_i$	$(x_i - \bar{x}) f_i$
<b>6</b>	- 1.85	3	
<b>7</b>	- 0.85	5	
<b>8</b>	0.15	6	
<b>9</b>	1.15	4	
<b>10</b>	2.15	2	



# Pojedinačna odstupanja

$x_i$	$x_i - \bar{x}$	$f_i$	$(x_i - \bar{x}) f_i$
<b>6</b>	- 1.85	3	- 5.55
<b>7</b>	- 0.85	5	- 4.25
<b>8</b>	0.15	6	0.90
<b>9</b>	1.15	4	4.60
<b>10</b>	2.15	2	4.30



# Pojedinačna odstupanja

$x_i$	$x_i - \bar{x}$	$f_i$	$(x_i - \bar{x}) f_i$
<b>6</b>	- 1.85	3	- 5.55
<b>7</b>	- 0.85	5	- 4.25
<b>8</b>	0.15	6	0.90
<b>9</b>	1.15	4	4.60
<b>10</b>	2.15	2	4.30
			<b>0</b>



# Pojedinačna odstupanja

$x_i$	$x_i - \bar{x}$	$f_i$	$(x_i - \bar{x}) f_i$
<b>6</b>	- 1.85	3	- 5.55
<b>7</b>	- 0.85	5	- 4.25
<b>8</b>	0.15	6	0.90
<b>9</b>	1.15	4	4.60
<b>10</b>	2.15	2	4.30
			<b>0</b>





# Varijansa

$$s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{n} = \frac{\sum x_i^2 f_i}{n} - \bar{x}^2$$

$x_i$  - varijeteti

$f_i$  - njihove frekvencije

$n$  - veličina uzorka

$\bar{x}$  - sredina



# Varijansa

$x_i$	$f_i$	$x_i f_i$
6	3	18
7	5	35
8	6	48
9	4	36
10	2	20
	20	157



# Varijansa

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
6	3	18	
7	5	35	
8	6	48	
9	4	36	
10	2	20	
	20	157	





# Varijansa

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
6	3	18	108
7	5	35	245
8	6	48	384
9	4	36	324
10	2	20	200
	20	157	



# Varijansa

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
6	3	18	108
7	5	35	245
8	6	48	384
9	4	36	324
10	2	20	200
	20	157	1261



# Varijansa

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
6	3	18	108
7	5	35	245
8	6	48	384
9	4	36	324
10	2	20	200
	20	157	1261

$$s^2 = \frac{1261}{20} - 7.85^2$$



# Varijansa

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
6	3	18	108
7	5	35	245
8	6	48	384
9	4	36	324
10	2	20	200
	20	157	1261

$$s^2 = \frac{1261}{20} - 7.85^2$$

$$s^2 = 63.05 - 61.6225$$



# Varijansa

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
6	3	18	108
7	5	35	245
8	6	48	384
9	4	36	324
10	2	20	200
	20	157	1261

$$s^2 = \frac{1261}{20} - 7.85^2$$

$$s^2 = 63.05 - 61.6225$$

$$s^2 = 1.4275$$



# Standardna devijacija

standardna devijacija =  $\sqrt{\text{varijansa}}$

$$s^2 = 1.4275$$

$$s = \sqrt{1.4275} = 1.1948$$



# Oznake...

parametar / statistika	populacija	uzorak
sredina	$\mu$	$\bar{X}$
standardna devijacija	$\sigma$	S



sredina **10** standardna devijacija **1**

sredina **100** standardna devijacija **1**





prosečna teža **10 kg**

a odstupanje od nje **1 kg**

sredina **10** standardna devijacija **1**

sredina **100** standardna devijacija **1**

prosečna teža **100 kg**

a odstupanje od nje **1 kg**



10%

prosečna težina **10 kg**  
a odstupanje od nje **1 kg**

sredina **10** standardna devijacija **1**

sredina **100** standardna devijacija **1**

prosečna težina **100 kg**  
a odstupanje od nje **1 kg**

1%



---

Potrebna **i** sredina **i**  
standardna devijacija ...



# Koeficijent varijacije

$$\text{koeficijent varijacije} = \frac{\text{standardna devijacija}}{\text{sredina}}$$

$$V = \frac{1.1948}{7.85} = 0.1522$$

$$V = 0.1522 \cdot 100\% = 15.22\%$$

# INFERENCIJALNA STATISTIKA

---

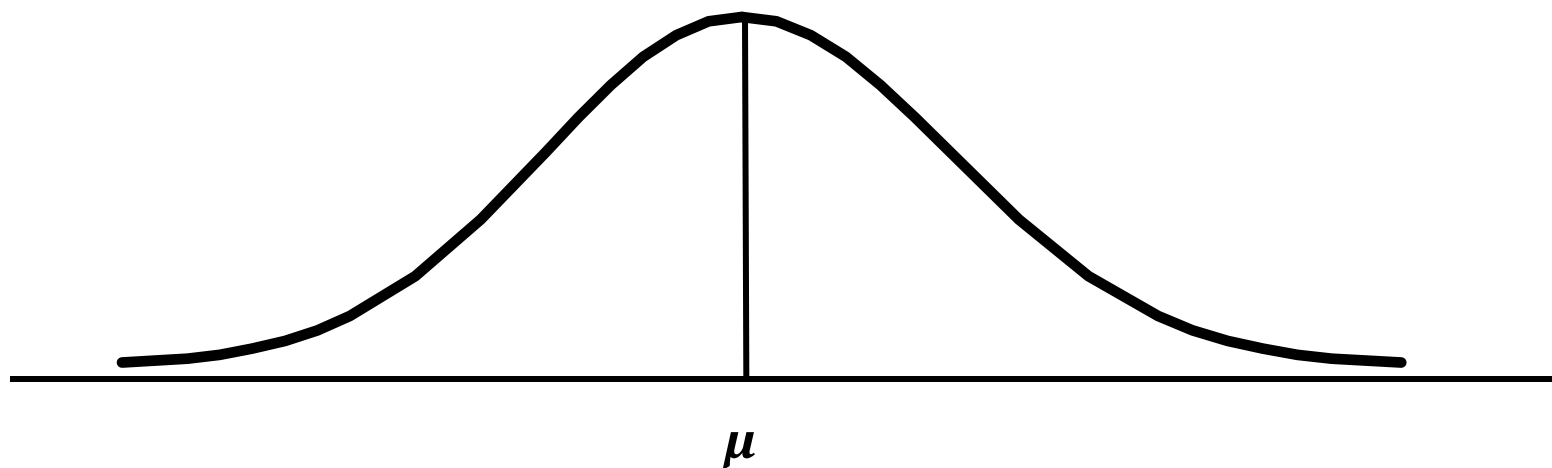


LJILJANA CVETKOVIĆ

<http://people.dmi.uns.ac.rs/~lila/>



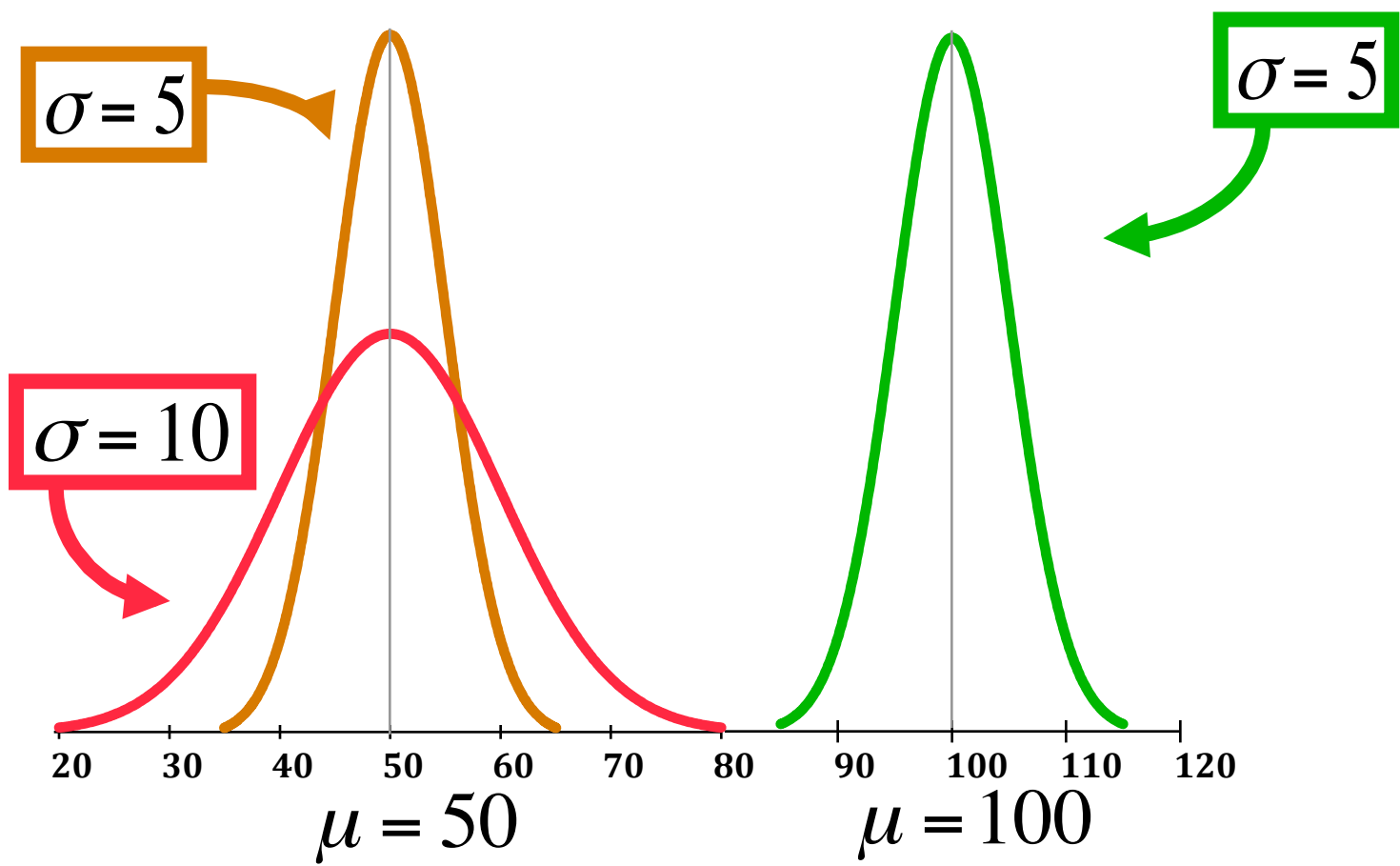
# NORMALNA RASPODELA



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



# Zvono kriva za razne $\mu$ i $\sigma$





# Empirijsko pravilo

Podaci – približno normalno distribuirani

Rastojanje od  
sredine

---

$$\mu \pm 1 \sigma$$

$$\mu \pm 2 \sigma$$

$$\mu \pm 3 \sigma$$

Procenat vrednosti unutar  
tog rastojanja

---

68,3

95,5

99,7





# Centralna granična teorema

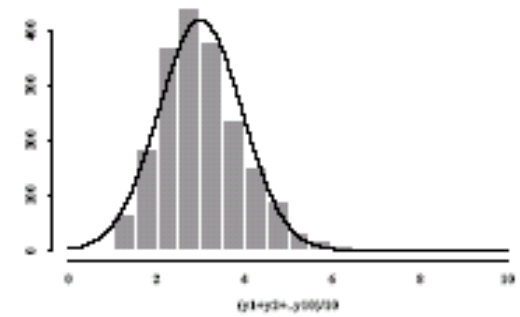
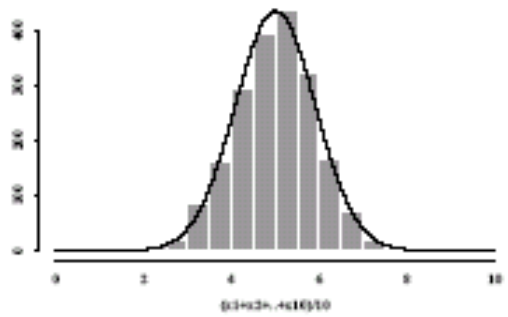
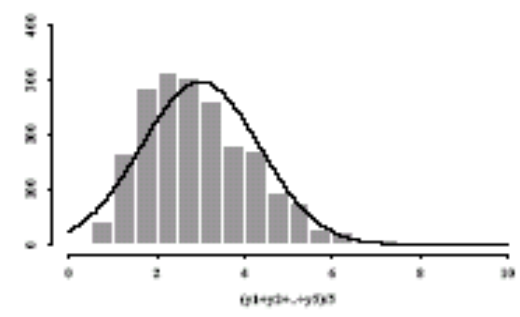
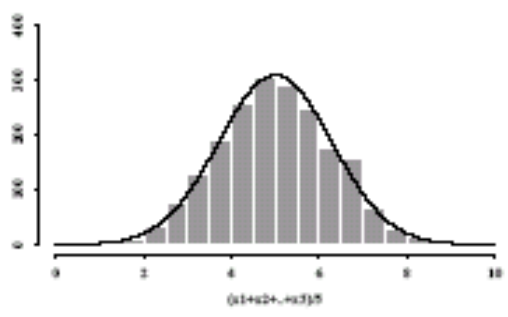
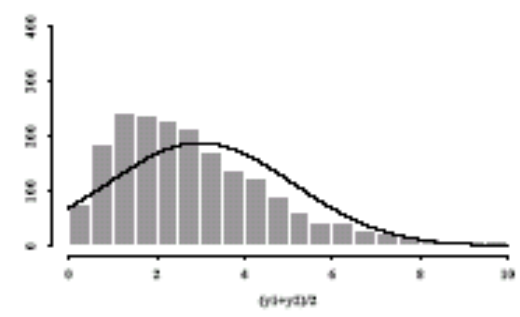
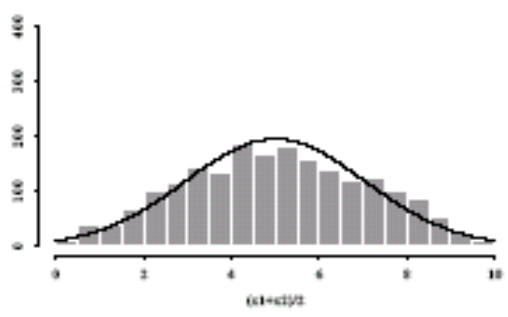
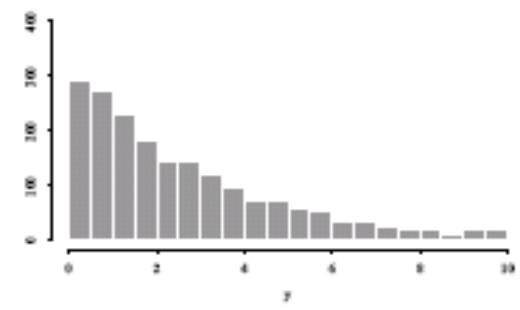
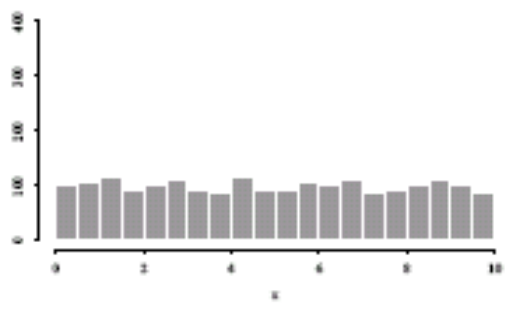
Bez obzira kakva je stvarna raspodela SP, za velike uzorke **raspodela sredina uzoraka** biće bliska normalnoj!

Dakle, možemo posmatrati SP sa raspodelom koja nije ni nalik normalnoj (čak može biti diskretna), ako uzmemo veliki broj velikih uzoraka i nacrtamo histogram sredina uzoraka, njegov oblik će ličiti na normalnu krivu.



## *Primer*

- ◆ Imamo 2 sasvim različite raspodele.  
Vršimo 2000 posmatranja
- ◆ Uzimamo uzorke dimenzije 2 iz svake populacije i računamo sredine. To ponovimo 1999 puta
- ◆ Ponovimo sve za uzorke dimenzije 5 i 10



# INTERVALI POVERENJA

---



◆ ZA SREDINU



# Osnovna ideja...

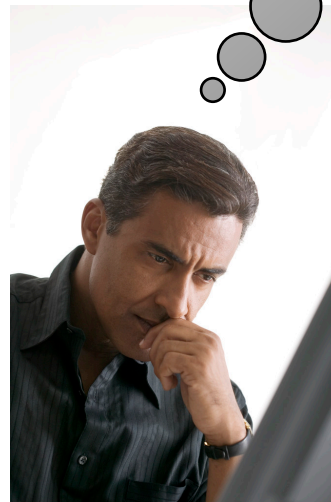
Populacija

Sredina  $\mu$  je  
nepoznata

Slučajni uzorak

Sredina  
 $\bar{x} = 50$

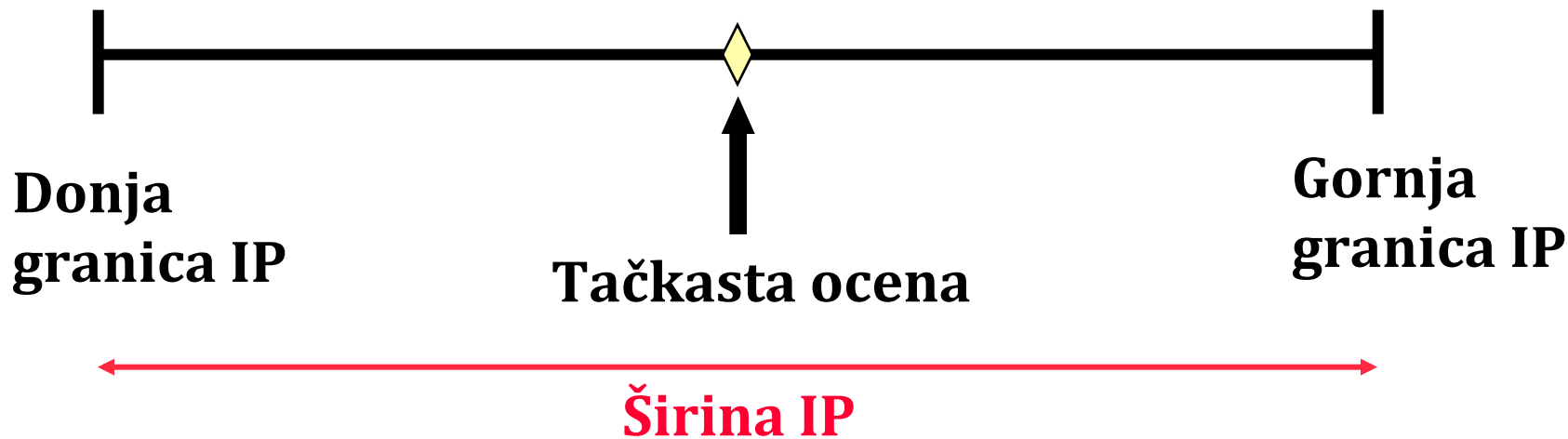
Ja sam 95%  
siguran da je  
 $\mu$  između 40  
i 60.





# Tačkaste i intervalne ocene

- ◆ Tačkasta ocena je *broj*
- ◆ Intervalna ocena daje dodatnu informaciju o *variranju*





# Tačkaste ocene

Ocenjujemo populacioni parametar ...		... pomoću uzoračke statistike (tačkaste ocene)
Sredina	$\mu$	$\bar{x}$
Varijansa	$\sigma^2$	$s^2$



# Nivo poverenja

- ◆ Verovatnoća da IP sadrži nepoznati populacioni parametar
- ◆ U procentima (manje od 100%)
- ◆ Označava se sa  $1 - \alpha$





# Nivo poverenja $1-\alpha$

- ◆ Na primer, nivo poverenja je 95%
- ◆  $1 - \alpha = 0.95$
- ◆ Interpretacija:
  - 95% svih IP koji se mogu konstruisati će sadržati tačnu vrednost nepoznatog parametra
- ◆ Svaki pojedinačni interval može a ne mora da sadrži tačan parametar



# Opšta formula za sve IP

korekcija

Tačkasta ocena  $\pm$  (Kritična vrednost) (Standardna greška)

$\bar{x}$

iz tablice

$\frac{s}{\sqrt{n-1}}$

$z_{\frac{1-\alpha}{2}}$

$t_{n-1, \alpha}$



# Računanje korekcije $\varepsilon$

$\varepsilon$

$n > 30$	$z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{n-1}}$
$n \leq 30$	$t_{n-1, \alpha} \frac{s}{\sqrt{n-1}}$



# Ako je poznata veličina populacije $N$

$$\frac{n}{N} < 0.04 = 4\%$$

$$(\bar{x} - \varepsilon, \bar{x} + \varepsilon)$$

$$\frac{n}{N} \geq 0.04 = 4\%$$

$$\kappa = \sqrt{\frac{N - n}{N - 1}}$$

$$(\bar{x} - \kappa\varepsilon, \bar{x} + \kappa\varepsilon)$$



# Naš primer...

IP za sredinu sa nivoom poverenja  $1 - \alpha = 95\%$

## Dati ili prethodno sračunati podaci:

$$n = 20, \quad \bar{x} = 7.85, \quad s = 1.1948, \quad 1 - \alpha = 0.95$$

## Izbor formule za korekciju:

$$\varepsilon = t_{n-1, \alpha} \frac{s}{\sqrt{n-1}}$$

## Računanje korekcije:

$$t_{19, 0.05} = 2.093 \quad \varepsilon = 2.093 \cdot \frac{1.1948}{\sqrt{19}} = 0.5737$$

## IP za sredinu je:

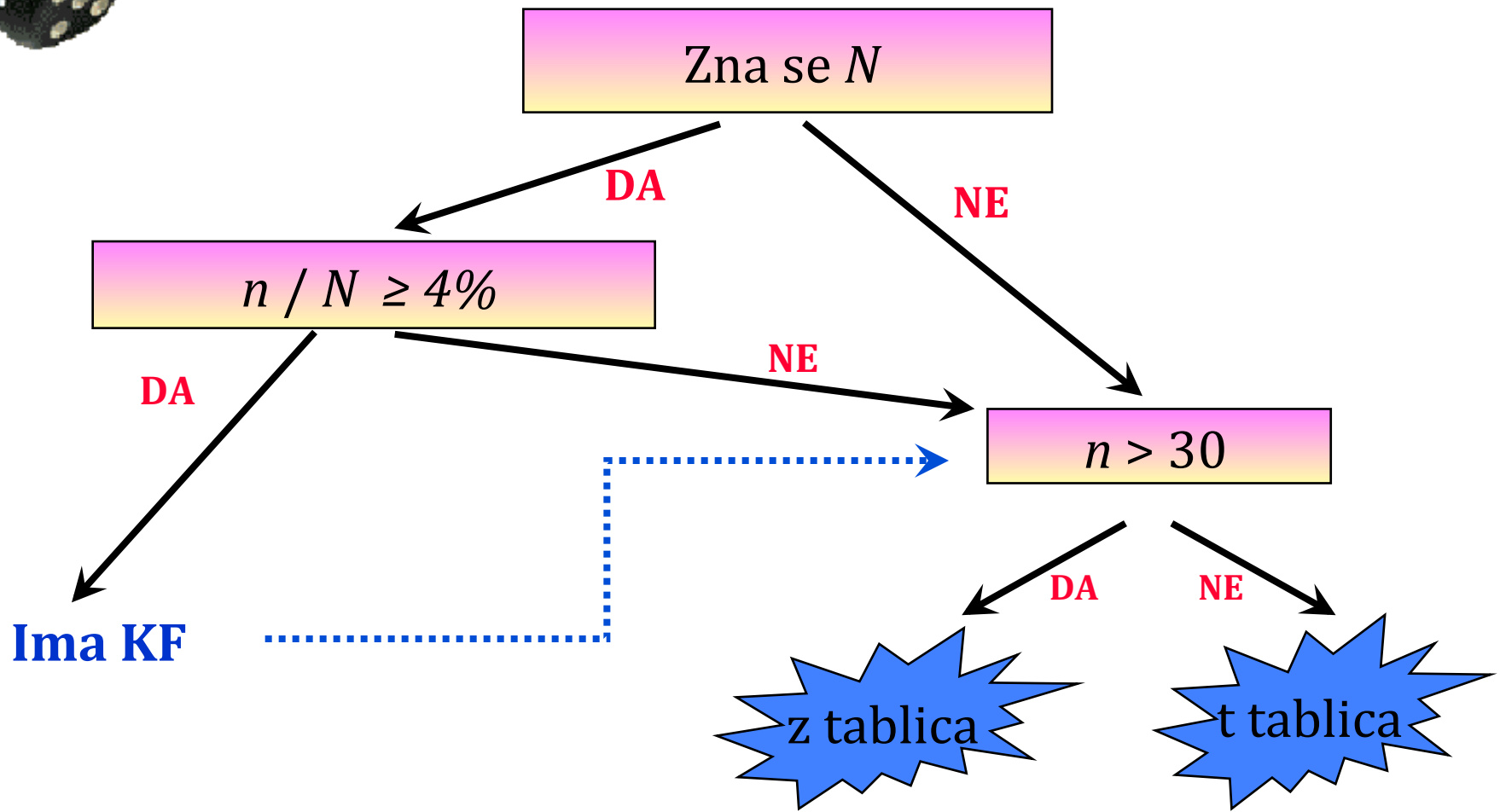
$$(81.375 - 5.711, 81.375 + 5.711) = (75.664, 87.086)$$

## Interpretacija:

Ako eksperiment ponovimo "bezbroj", tj. veliki broj puta, uzimajući nove i nove uzorke dimenzije 20 i sračunavajući IP za sredinu, u 95% slučajeva tačan parametar (sredina populacije, tj. prosečan broj poena svih studenata koji su polagali statistiku) **bi bio** neka vrednost iz dobijenog intervala.



# Algoritam



# TESTIRANJE HIPOTEZA



---

LJILJANA CVETKOVIĆ

<http://people.dmi.uns.ac.rs/~lila/>



# Šta je hipoteza?

- ◆ Hipoteza je tvrdnja (pretpostavka) o populacionom parametru:
  - populaciona sredina



**Primer: Prosečan mesečni račun za mobilni telefon u Novom Sadu je  $\mu = 420$  RSD**





# Nulta hipoteza, $H_0$

- ◆ Tvrdi nešto (numerički) što će se testirati

**Primer: Prosečan mesečni račun za mobilni telefon u Novom Sadu je  $\mu = 420$  RSD**

- ◆ Uvek se odnosi na populacioni parametar, a ne na uzoračku statistiku

$$H_0 : \mu = 420$$

$$H_0 : \bar{x} = 420$$



# Nulta hipoteza, $H_0$

- ◆ Počinje pretpostavkom da je nulta hipoteza tačna
  - Slično onome “nevin dok se ne dokaže da je kriv”
- ◆ Upućuje na **status quo**
- ◆ Može se **odbaciti** ili **ne odbaciti**



# Alternativna hipoteza, $H_A$

- ◆ Je suprotna od nulte hipoteze
- ◆ Napada **status quo**
- ◆ Najčešće je to nešto u šta istraživač veruje (i potrebno je da to potvrdi) – **istraživačeva hipoteza**



# Formulacija hipoteze

- ◆ Neko tvrdi da je prosečan mesečni račun za mobilni telefon u Novom Sadu 420 RSD. Analitičar mobilne telefonije želi da proveri ovo tvrdjenje.
- 
- ◆ Šta je odgovarajući test?



# Formulacija hipoteze

$H_0: \mu = 420$  (prosečan račun je onakav kakav se tvrdi) **status quo**

$H_A: \mu \neq 420$  (prosečan račun nije onakav kakav se tvrdi)

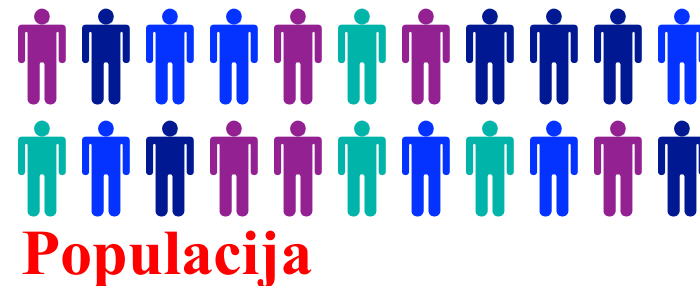
Analitičar će poverovati u tvrdnju, izuzev u slučaju da nađe dovoljno dokaza da je odbaci.



# Postupak testiranja hipoteze

**Tvrđenje: prosečna starost populacije je 50.**

**Nulta hipoteza:  $H_0: \mu = 50$**



Biramo slučajan uzorak



**Neka je prosečna starost uzorka 20:**  
 $\bar{x} = 20$

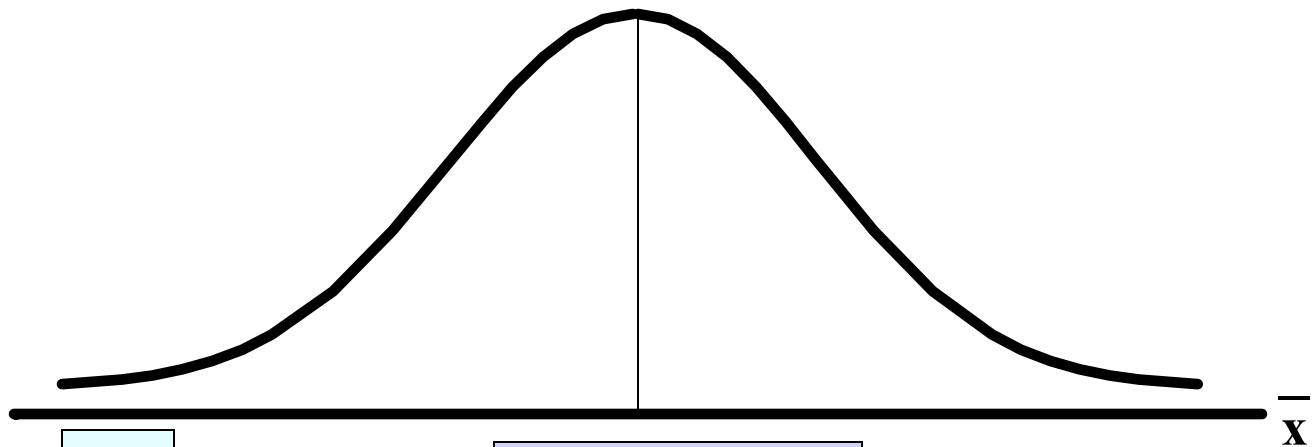
Da li je  
 $\bar{x} = 20$   
verovatno  
ako je  
 $\mu = 50?$

**Ako nije  
ODBACI  
nultu hipotezu**



# Razlog za odbacivanje $H_0$

raspodela za  $\bar{x}$



20

$\mu = 50$   
Ako je  $H_0$  tačna

Ako nije verovatno da ćemo dobiti ovoliku uzoračku sredinu...

... U stvari to je populaciona sredina...

... Onda ćemo odbaciti nultu hipotezu da je  $\mu = 50$ .



# Greške u zaključivanju

## ◆ Greška I tipa

- Odbaciti **tačnu** nultu hipotezu

Verovatnoća greške I tipa je  $\alpha$

- Zove se **nivo značajnosti** testa
- Zadaje se unapred od strane istraživača





# Greške u zaključivanju

## ◆ Greška I tipa

- Ne odbaciti **netačnu** nultu hipotezu

Verovatnoća greške II tipa je  $\beta$



# Ishodi i verovatnoće

## Mogući ishodi nakon testiranja hipoteze

**Ishod  
(Verovatnoća)**

	Činjenično stanje	
Zaključak	$H_0$ tačna	$H_0$ netačna
Ne odbaciti $H_0$	<b>Nema greške (<math>1 - \alpha</math>)</b>	<b>Greška II tipa (<math>\beta</math>)</b>
Odbaciti $H_0$	<b>Greška I tipa (<math>\alpha</math>)</b>	<b>Nema greške (<math>1 - \beta</math>)</b>



# Primer

Advokat se trudi da dokaže da određena kompanija nema fer odnos prema ženama zaposlenim u toj kompaniji, u smislu da njihovo povećanje plate ne prati prosečno povećanje plata, koje na godišnjem nivou iznosi 8%.

$$H_0 : \mu = 0.08$$

$$H_A : \mu < 0.08$$

**Greška I tipa:** Kompanija je kažnjena, iako nije imala diskriminišući odnos prema ženskim zaposlenicima

**Greška II tipa:** Dozvoljava se kompaniji da radi bez sankcija, iako ima diskriminišući odnos prema ženskim zaposlenicima



# Nivo značajnosti, $\alpha$

- ◆ Definiše vrednosti uzoračke statistike koje nisu verovatne ako je nulta hipoteza tačna
  - Definiše **kritičnu oblast** (region odbacivanja)
- ◆ Obeležava se sa  $\alpha$ 
  - Tipične vrednosti su 0.01, 0.05, 0.10
- ◆ Bira ga istraživač unapred



# Osnovni princip

Ono što nas interesuje pojednostavljeno je u vidu dva suprotstavljena tvrđenja, tj. dve hipoteze:

- ◆ nulte hipoteze  $H_0$
- ◆ alternativne hipoteze  $H_A$

Ove dve hipoteze *ne tretiraju* se na isti način!

Zaključak, nakon testiranja, može biti:

- ◆ odbaciti  $H_0$                       odbacujemo  $H_0$  u korist  $H_A$
- ◆ ne odbaciti  $H_0$                       nema dovoljno dokaza protiv  $H_0$



# Postupak

1. **korak:** Formulirati nultu hipotezu  $H_0$  i alternativnu  $H_A$
2. **korak:** Izabrati  $\alpha$  (ova vrednost treba da je mala, obično manja od 10%)
3. **korak:** Izabrati test statistiku i izračunati njenu vrednost na osnovu podataka iz uzorka
4. **korak:** Uporediti dobijenu vrednost statistike sa kritičnom vrednošću (koja zavisi od izabranog  $\alpha$ )
5. **korak:** Doneti odluku:
  - Ako test statistika "upada" u kritičnu oblast - odbaciti  $H_0$  u korist  $H_A$
  - Ako test statistika ne "upada" u kritičnu oblast - zaključiti da nema dovoljno dokaza protiv  $H_0$



# t-test za sredinu populacije

**Cilj:** Uporediti sredinu uzorka i sredinu cele populacije

**Uslovi:** Populacija mora imati neprekidnu raspodelu, ako je obim uzorka  $n < 30$

**Nulta hipoteza:**  $H_0 : \mu = \mu_0$

**Alternativna hipoteza:**  $H_A : \mu$  se značajno razlikuje od  $\mu_0$

**Statistika:** 
$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s}$$

**Kritična vrednost iz tablice:**  $t_{n-1,\alpha}$

**Zaključak:** Ako je  $|t| > t_{n-1,\alpha}$  nulta hipoteza se odbacuje



Znak =

$$\mu = \mu_0$$

sredina populacije je  
bliska broju  $\mu_0$





# Primer

Slučajno je izabrano 26 radnika jedne fabrike patika i svakom od njih izmereno je vreme (u sekundama), za koje ušnjiraju pertle

broj sekundi	broj radnika
25	4
26	4
28	6
29	5
33	4
38	3



---

Da li sa nivoom značajnosti  $\alpha = 0.05$  možemo tvrditi da je vreme šnjiranja pertli 30 sekundi?



Da li sa nivoom značajnosti  $\alpha = 0.05$  možemo tvrditi da je vreme šnjiranja pertli 30 sekundi?

$x_i$	$f_i$
25	4
26	4
28	6
29	5
33	4
38	3
	26



Da li sa nivoom značajnosti  $\alpha = 0.05$  možemo tvrditi da je vreme šnjiranja pertli 30 sekundi?

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
25	4	100	2500
26	4	104	2704
28	6	168	4704
29	5	145	4205
33	4	132	4356
38	3	114	4332
	26	763	22801
sredina	29.35		
stand. dev.	3.97		



# Test

**Uslovi:** Populacija ima neprekidnu raspodelu

**Nulta hipoteza:**  $H_0 : \mu = 30$

**Alternativna hipoteza:**  $H_A : \mu$  se značajno razlikuje od 30

**Statistika:** 
$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} = \frac{(29.35 - 30)\sqrt{26}}{3.97} = -0.835$$

**Kritična vrednost iz tablice:**  $t_{25,0.05} = 2.060$

**Zaključak:** Nije  $|t| > 2.060$ , pa se nulta hipoteza **ne odbacuje**

Dakle, sa nivoom značajnosti 0.05, *nemamo dokaza* protiv hipoteze da je prosečno vreme šnjiranja pertli 30 sekundi.



# Pristup preko p-vrednosti

- ◆ Konvertovati uzoračku statistiku ( $\bar{x}$ ) u test statistiku ( $t$ )
- ◆ Odrediti **p-vrednost** iza tablice ili kompjuterski
- ◆ Uporediti **p-vrednost** sa  $\alpha$

- Ako je  $p\text{-vrednost} < \alpha$ , odbaciti  $H_0$
- Ako je  $p\text{-vrednost} \geq \alpha$ , ne odbaciti  $H_0$



# p-vrednost

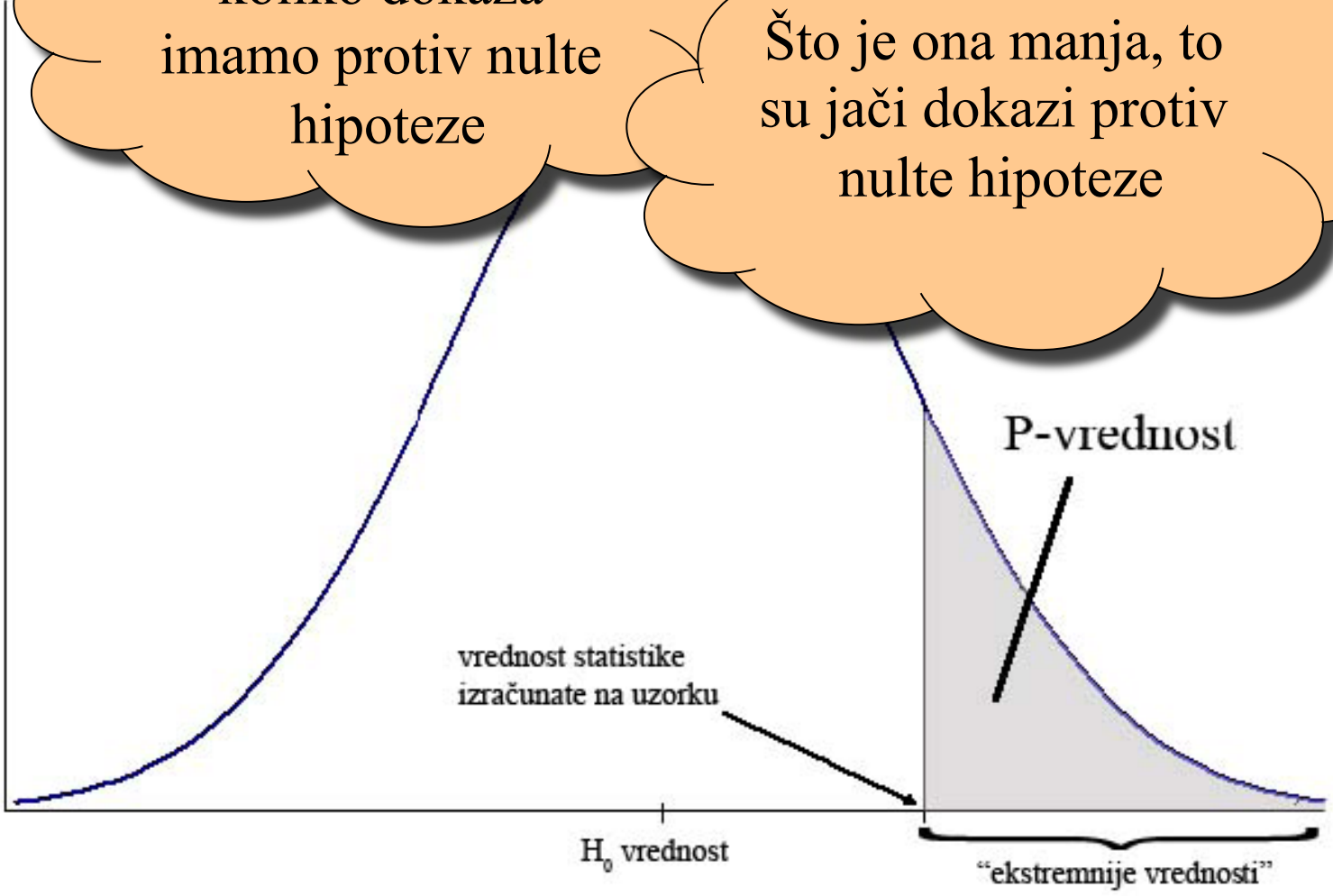
- ◆ Je verovatnoća da će test statistika biti ekstremnija nego opažena uzoračka vrednost **ako je  $H_0$  tačna**
  - Zove se još i opaženi nivo značajnosti
  - Najmanje  $\alpha$  za koje  $H_0$  može da se odbaci
  - Što je ona manja, to su jači dokazi protiv nulte hipoteze



# dnost

P-vrednost je mera koliko dokaza imamo protiv nulte hipoteze

Što je ona manja, to su jači dokazi protiv nulte hipoteze







# Interpretacija P-vrednosti

P-vrednost	Interpretacija
$P < 0.01$	veoma jak dokaz protiv $H_0$
$0.01 \leq P < 0.05$	umeren dokaz protiv $H_0$
$0.05 \leq P < 0.10$	nagoveštaj protiv $H_0$
$0.10 \leq P$	malo ili nema realnih dokaza protiv $H_0$



# Isti primer pomoću P-vrednosti

**Nulta hipoteza:**  $H_0 : \mu = 30$

**Alternativna hipoteza:**  $H_A : \mu$  se značajno razlikuje od 30

**Test statistika:**  $t = -0.835$

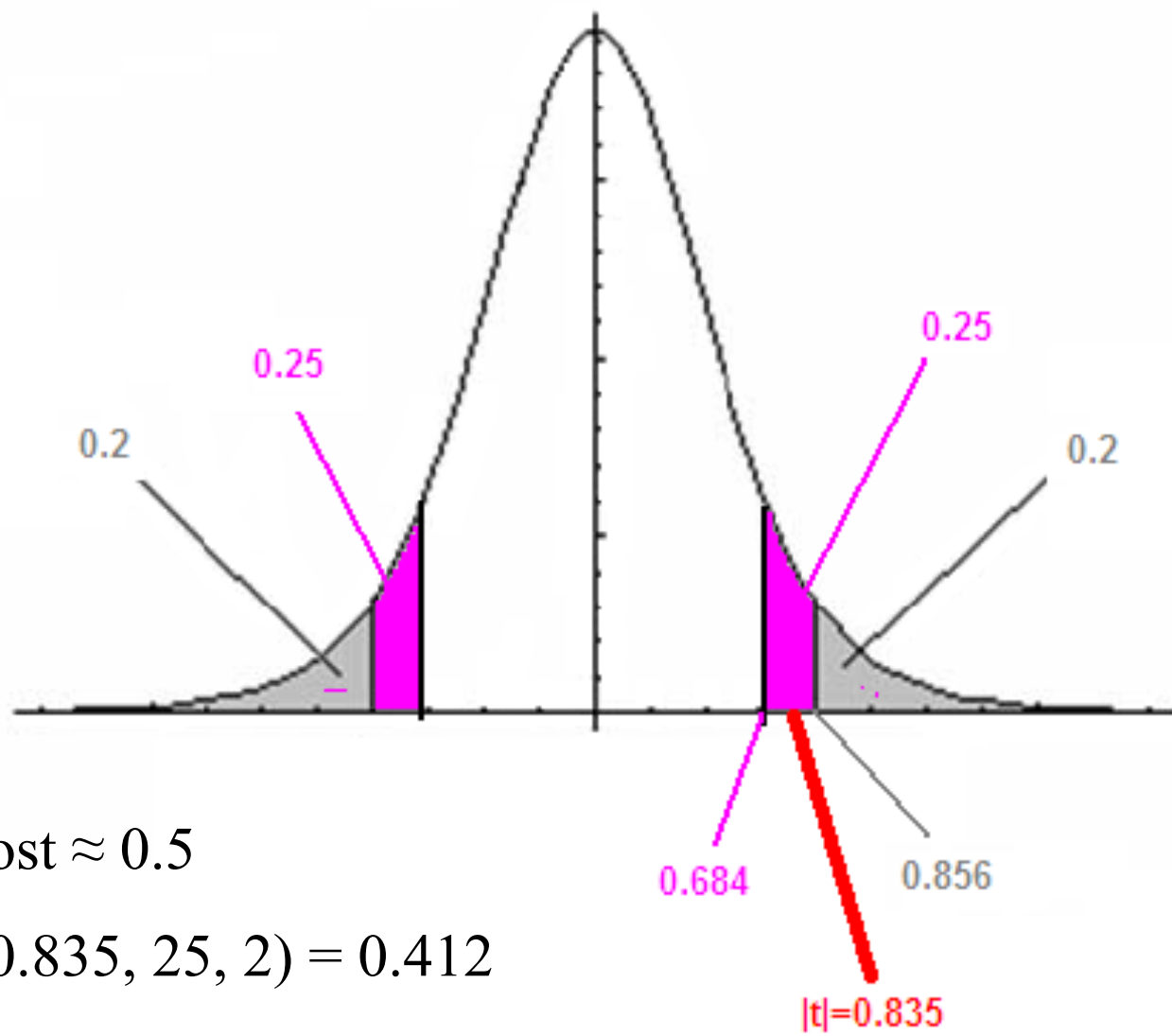
**P-vrednost:** t-tablica, 25-ta vrsta ( $n - 1 = 25$ )

$n \setminus \alpha$	0.500	0.400	0.300	0.200	0.100	0.050	...
25	0.684	0.856	1.058	1.316	1.708	2.060	...

$|t| = 0.835$  je između 0.684 i 0.856

P-vrednost je između 0.4 i 0.5 (**P-vrednost  $\approx 0.5$** )

**Zaključak:** P-vrednost (0.5) nije manja ili jednaka sa  $\alpha = 0.05$ , pa se nulta hipoteza **ne odbacuje**



P-vrednost  $\approx 0.5$

$$\text{TDIST}(0.835, 25, 2) = 0.412$$



# Pojačati zaključak...

P-vrednost	Interpretacija
$P < 0.01$	veoma jak dokaz protiv $H_0$
$0.01 \leq P < 0.05$	umeren dokaz protiv $H_0$
$0.05 \leq P < 0.10$	nagoveštaj protiv $H_0$
$0.10 \leq P$	malo ili nema realnih dokaza protiv $H_0$



# PARAMETARSKI TESTOVI

## ANOVA ANalysis Of Variances



# PARAMETARSKI TESTOVI

## ANOVA

## ANalysis Of Variances

Uveo je Fisher 1923. kao novi metod razdvajanja dejstva različitih tretmana u dobro dizajniranom naučnom eksperimentu, koga je nazvao „analiza varijanse“

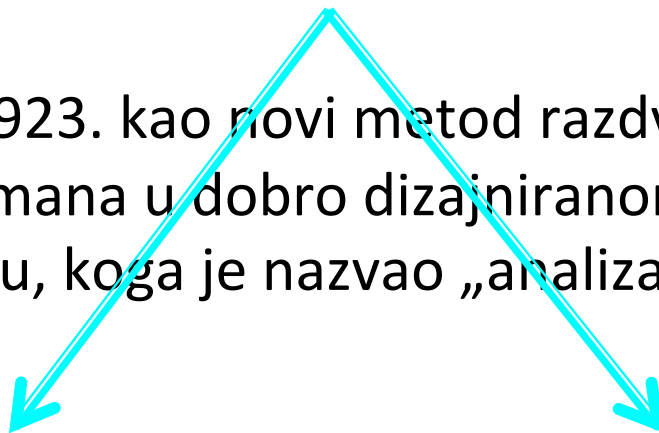


# PARAMETARSKI TESTOVI

**ANOVA**

**ANalysis Of Variances**

Uveo je Fisher 1923. kao novi metod razdvajanja dejstva različitih tretmana u dobro dizajniranom naučnom eksperimentu, koga je nazvao „analiza varijanse“



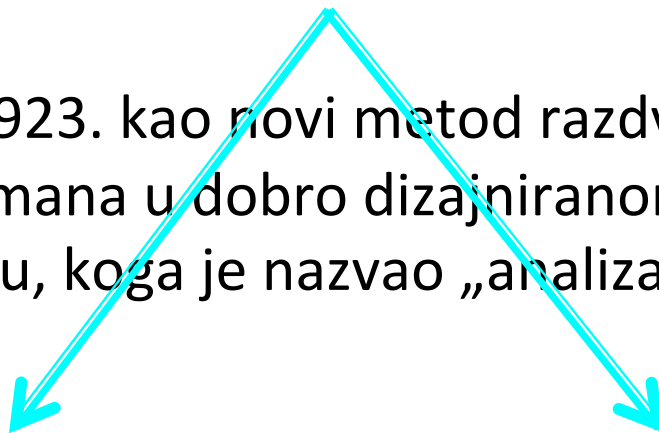


# PARAMETARSKI TESTOVI

**ANOVA**

**ANalysis Of Variances**

Uveo je Fisher 1923. kao novi metod razdvajanja dejstva različitih tretmana u dobro dizajniranom naučnom eksperimentu, koga je nazvao „analiza varijanse“



jednofaktorska

višefaktorska





# PARAMETARSKI TESTOVI

**ANOVA**

**ANalysis Of Variances**

Uveo je Fisher 1923. kao novi metod razdvajanja dejstva različitih tretmana u dobro dizajniranom naučnom eksperimentu, koga je nazvao „analiza varijanse“

**jednofaktorska**

višefaktorska



# Jednofaktorska ANOVA

... Bavi se ispitivanjem uticaja jedne nezavisne promenljive (**FAKTOR UTICAJA**) na jednu zavisnu promenljivu.



# Jednofaktorska ANOVA

... Bavi se ispitivanjem uticaja jedne nezavisne promenljive (**FAKTOR UTICAJA**) na jednu zavisnu promenljivu.

Faktor uticaja ima 3 ili više kategorija (grupa).



# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F



# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između				
unutar				
ukupno				



# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$			
unutar	$Q_u$			
ukupno	$Q$			




# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_j$	$m-1$		
unutar	$Q_u$	$mn-m$		
ukupno	$Q$	$mn-1$		





# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_j$	$m-1$	 = $\frac{\quad}{\quad}$	
unutar	$Q_u$	$mn-m$		
ukupno	$Q$	$mn-1$		



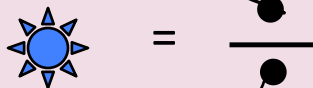


# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$	$m-1$	 = 	
unutar	$Q_u$	$mn-m$		
ukupno	$Q$	$mn-1$		





# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$	$m-1$		
unutar	$Q_u$	$mn-m$		
ukupno	$Q$	$mn-1$		





# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$	$m-1$	 = $\frac{\bullet}{\bullet}$	
unutar	$Q_u$	$mn-m$	 = $\text{—}$	
ukupno	$Q$	$mn-1$		





# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$	$m-1$	 = $\frac{\bullet}{\bullet}$	
unutar	$Q_u$	$mn-m$	 = $\frac{\bullet}{\bullet}$	
ukupno	$Q$	$mn-1$		





# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$	$m-1$	 = $\frac{\bullet}{\bullet}$	
unutar	$Q_u$	$mn-m$	 = $\frac{\bullet}{\bullet}$	
ukupno	$Q$	$mn-1$		



# ANOVA tabela

Izvor varijabilite ta	Suma kvadrata	Stepeni slobode	varijanse	F
između	$Q_i$	$m-1$	 = $\frac{\bullet}{\bullet}$	= $\frac{\text{Sun icon}}{\text{Moon icon}}$
unutar	$Q_u$	$mn-m$	 = $\frac{\bullet}{\bullet}$	
ukupno	$Q$	$mn-1$		



# PRIMER

- ◆ Radi testiranja novog sistema obuke zaposlenih, jedna firma je na slučajan način izabrala dvadeset ispitanika, koji su podeljeni u četiri grupe po pet. Prva grupa je primila obuku tipa A, druga tipa B, treća tipa C i četvrta tipa D. Zatim je tokom probnog perioda od mesec dana, merena efikasnost ispitanika u obavljanju radnih zadataka. Rezultati istraživanja su dati u tabeli:



Tip obuke		Efikasnost u %				
A	87	79	85	92	77	
B	72	71	80	73	84	
C	92	85	87	78	83	
D	70	72	70	88	75	





# PITANJE

Da li se, na nivou značajnosti od 1%, na osnovu ovog uzorka, može tvrditi da tip obuke ne utiče na efikasnost obavljanja radnih zadataka? Da li se dolazi do istog zaključka na nivou značajnosti od 5%?



# Postavljanje hipoteze



# Postavljanje hipoteze

**$H_0$ : Tip obuke ne utiče na prosečnu efikasnost  
( $\mu_1 = \mu_2 = \mu_3 = \mu_4$ )**

**$H_A$ : Tip obuke utiče na prosečnu efikasnost**



# SITAN RAČUN...

Sredine uzoraka						Zbir	Sredina
uzorak 1	87	79	85	92	77	420	84
uzorak 2	72	71	80	73	84	380	76
uzorak 3	92	85	87	78	83	425	85
uzorak 4	70	72	70	88	75	375	75
						Ukupna sredina	80



# SITAN RAČUN...

Sredine uzoraka						Zbir	Sredina
uzorak 1	87	79	85	92	77	420	84
uzorak 2	72	71	80	73	84	380	76
uzorak 3	92	85	87	78	83	425	85
uzorak 4	70	72	70	88	75	375	75
						Ukupna sredina	80

Kvadrati odstupanja od ukupne sredine						Zbir
uzorak 1	49	1	25	144	9	228
uzorak 2	64	81	0	49	16	210
uzorak 3	144	25	49	4	9	231
uzorak 4	100	64	100	64	25	353
Q						1022



# SITAN RAČUN...

Sredine uzoraka						Zbir	Sredina
uzorak 1	87	79	85	92	77	420	84
uzorak 2	72	71	80	73	84	380	76
uzorak 3	92	85	87	78	83	425	85
uzorak 4	70	72	70	88	75	375	75
						Ukupna sredina	80

Kvadrati odstupanja od ukupne sredine						Zbir
uzorak 1	49	1	25	144	9	228
uzorak 2	64	81	0	49	16	210
uzorak 3	144	25	49	4	9	231
uzorak 4	100	64	100	64	25	353
Q						1022



# SITAN RAČUN...

Sredine uzoraka						Zbir	Sredina
uzorak 1	87	79	85	92	77	420	84
uzorak 2	72	71	80	73	84	380	76
uzorak 3	92	85	87	78	83	425	85
uzorak 4	70	72	70	88	75	375	75
						Ukupna sredina	80

Kvadrati odstupanja od ukupne sredine						Zbir
uzorak 1	49	1	25	144	9	228
uzorak 2	64	81	0	49	16	210
uzorak 3	144	25	49	4	9	231
uzorak 4	100	64	100	64	25	353
Q						1022

Kvadrati odstupanja unutar uzorka						Zbir
uzorak 1	9	25	1	64	49	148
uzorak 2	16	25	16	9	64	130
uzorak 3	49	0	4	49	4	106
uzorak 4	25	9	25	169	0	228
Q <sub>u</sub>						612



# SITAN RAČUN...

Sredine uzoraka						Zbir	Sredina
uzorak 1	87	79	85	92	77	420	84
uzorak 2	72	71	80	73	84	380	76
uzorak 3	92	85	87	78	83	425	85
uzorak 4	70	72	70	88	75	375	75
						Ukupna sredina	80

Kvadrati odstupanja od ukupne sredine						Zbir
uzorak 1	49	1	25	144	9	228
uzorak 2	64	81	0	49	16	210
uzorak 3	144	25	49	4	9	231
uzorak 4	100	64	100	64	25	353
Q						1022

Kvadrati odstupanja unutar uzorka						Zbir
uzorak 1	9	25	1	64	49	148
uzorak 2	16	25	16	9	64	130
uzorak 3	49	0	4	49	4	106
uzorak 4	25	9	25	169	0	228
Q <sub>u</sub>						612





# SITAN RAČUN...

Izvor varijabiliteta	Suma kvadrata	Stepeni slobode	Varijanse	F statistika
Izmedju uzoraka	410	3	136,67	3,5730
Unutar uzoraka	612	16	38,25	
Ukupno	1022	19		



# SITAN RAČUN...

Izvor varijabiliteta	Suma kvadrata	Stepeni slobode	Varijanse	F statistika
Izmedju uzoraka	410	3	136,67	3,5730
Unutar uzoraka	612	16	38,25	
Ukupno	1022	19		



# SITAN RAČUN...

Izvor varijabiliteta	Suma kvadrata	Stepeni slobode	Varijanse	F statistika
Izmedju uzoraka	410	3	136,67	3,5730
Unutar uzoraka	612	16	38,25	
Ukupno	1022	19		

m	4
n	5
Nivo značajnosti	0,01
Kritična vrednost	5,2922



m	4
n	5
Nivo značajnosti	0,05
Kritična vrednost	3,2389





# SITAN RAČUN...

Izvor varijabiliteta	Suma kvadrata	Stepeni slobode	Varijanse	F statistika
Izmedju uzoraka	410	3	136,67	3,5730
Unutar uzoraka	612	16	38,25	
Ukupno	1022	19		

m	4
n	5
Nivo značajnosti	0,01
Kritična vrednost	5,2922



m	4
n	5
Nivo značajnosti	0,05
Kritična vrednost	3,2389





# F tablica 1% = 0.01

$n_1/n_2$	1	2	3	4	5	6	7
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70



$F_{3;16; 0.01}$

$n_1/n_2$	1	2	3	4	5	6	7
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70



$F_{3;16; 0.01}$

$n_1/n_2$	1	2	3	4	5	6	7
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70



$F_{3;16; 0.01}$

$n_1/n_2$	1	2	3	4	5	6	7
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70





# F tablica 5% = 0.05

$n_1/n_2$	1	2	3	4	5	6	7
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51



$F_{3;16; 0.05}$

$n_1/n_2$	1	2	3	4	5	6	7
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51



# ZAKLJUČAK

tablica 1%

5,29



# ZAKLJUČAK

tablica 1%

5,29

odbacujemo  $H_0$



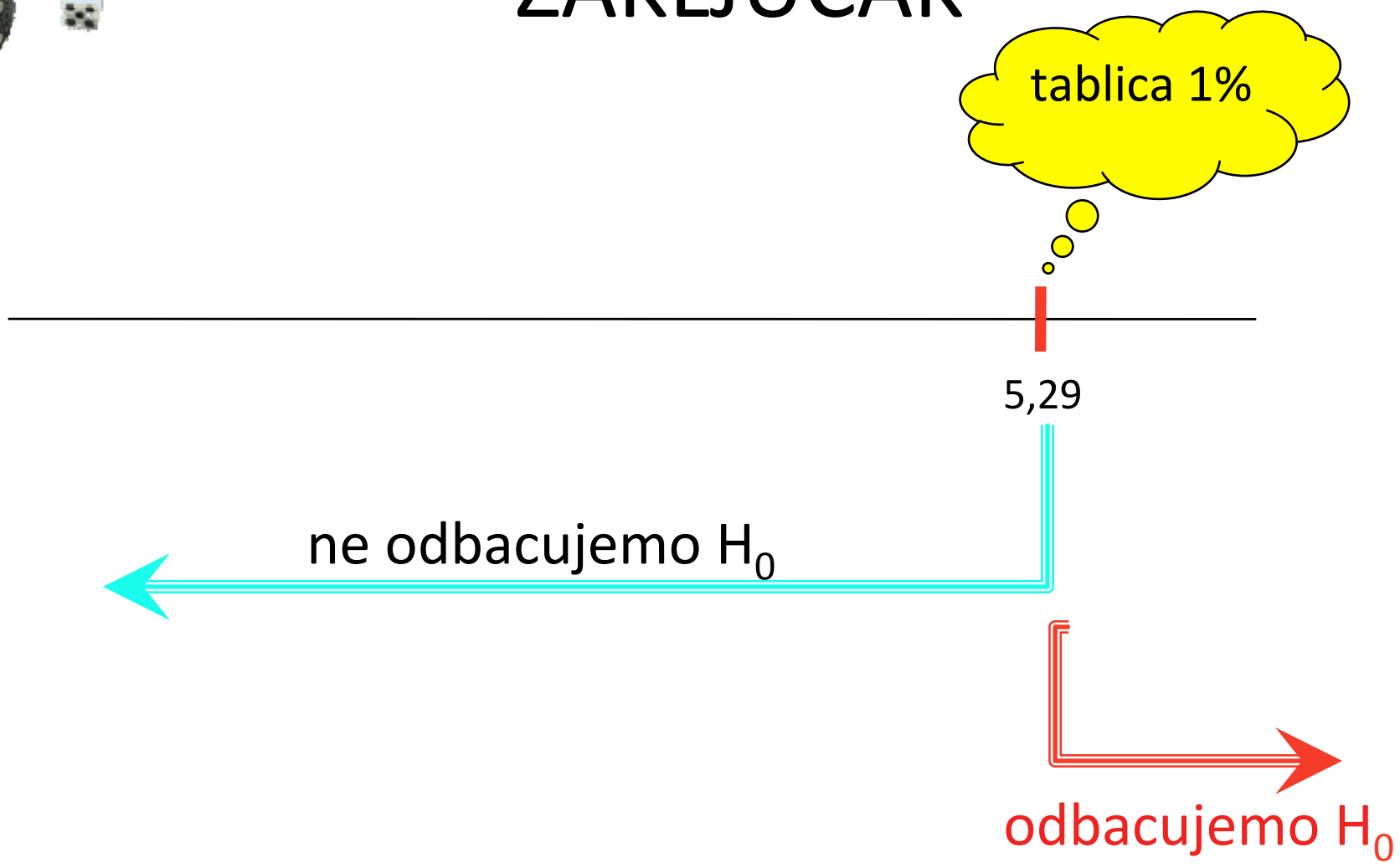
# ZAKLJUČAK

tablica 1%

5,29

ne odbacujemo  $H_0$

odbacujemo  $H_0$





# ZAKLJUČAK

sračunato

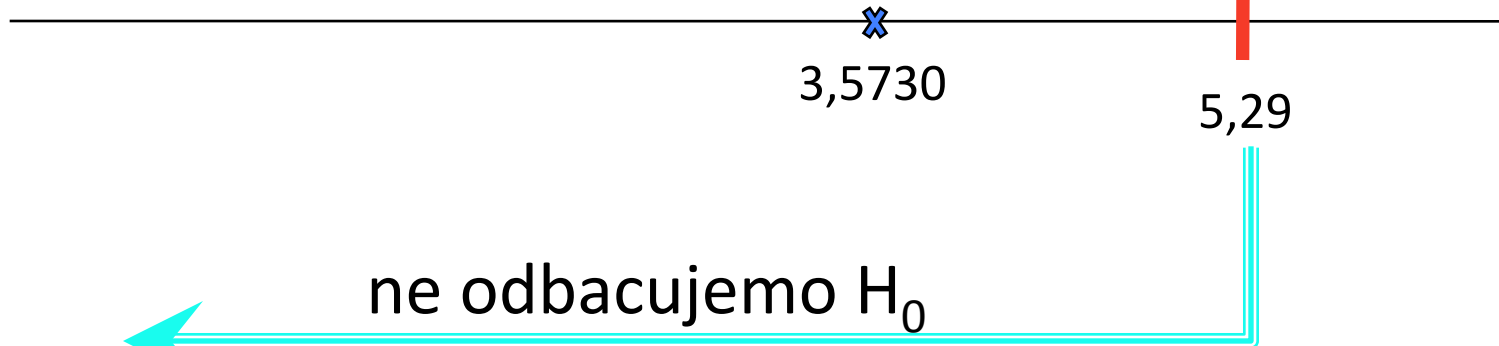
tablica 1%

3,5730

5,29

ne odbacujemo  $H_0$

odbacujemo  $H_0$





# ZAKLJUČAK

sračunato

tablica 1%



3,5730

5,29

ne odbacujemo  $H_0$

odbacujemo  $H_0$



# ZAKLJUČAK

tablica 5%



3,24





# ZAKLJUČAK

tablica 5%



3,24

ne odbacujemo  $H_0$



odbacujemo  $H_0$



# ZAKLJUČAK

tablica 5%

sračunato

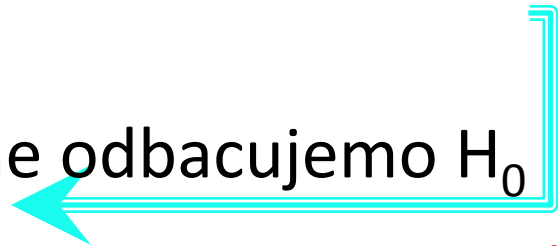


3,24

3,5730

ne odbacujemo  $H_0$

odbacujemo  $H_0$





# ZAKLJUČAK

tablica 5%

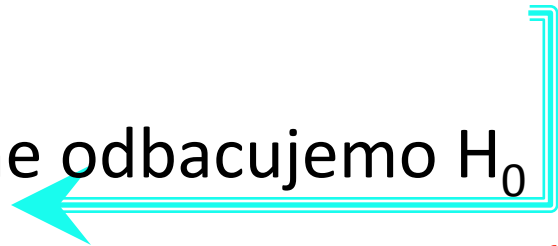
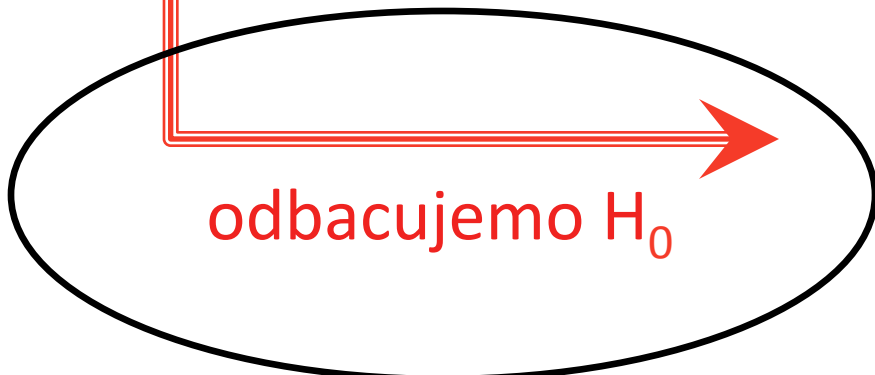
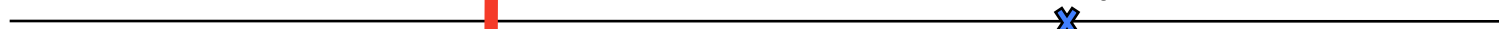
sračunato

3,24

3,5730

ne odbacujemo  $H_0$

odbacujemo  $H_0$





# ZAKLJUČAK

Izvor varijabiliteta	Suma kvadrata	Stepeni slobode	Varijanse	F statistika
Izmedju uzoraka	410	3	136,67	3,5730
Unutar uzoraka	612	16	38,25	
Ukupno	1022	19		



m	4
n	5
Nivo značajnosti	0,01
Kritična vrednost	5,2922



$H_0$  se ne odbacuje

Sa 1% nema dokaza da tip obuke utiče.

m	4
n	5
Nivo značajnosti	0,05
Kritična vrednost	3,2389



$H_0$  se odbacuje

Sa 5% ima dokaza da tip obuke utiče.

# KORELACIJA I REGRESIJA



---

LJILJANA CVETKOVIĆ

<http://people.dmi.uns.ac.rs/~lila/>



# OSNOVNI POJMOVI

## ◆ KORELACIJA

- međusobna uzajamna povezanost među pojavama (varijablama, obeležjima, faktorima)



# OSNOVNI POJMOVI

## ◆ KORELACIJA

- međusobna uzajamna povezanost među pojavama (varijablama, obeležjima, faktorima)

## ◆ VRSTE KORELACIJA

- jednostruka, višestruka
- linearna, nelinearna
- pozitivna, negativna



# OSNOVNI POJMOVI

## ◆ KORELACIJA

- međusobna uzajamna povezanost među pojavama (varijablama, obeležjima, faktorima)

## ◆ VRSTE KORELACIJA

- jednostruka, višestruka
- linearna, nelinearna
- pozitivna, negativna

## ◆ MERENJE JAČINE POVEZANOSTI

- koeficijent korelacije  $r$  (na uzorku), odn.  $\rho$  (populacija)
- koeficijent determinacije  $r^2$





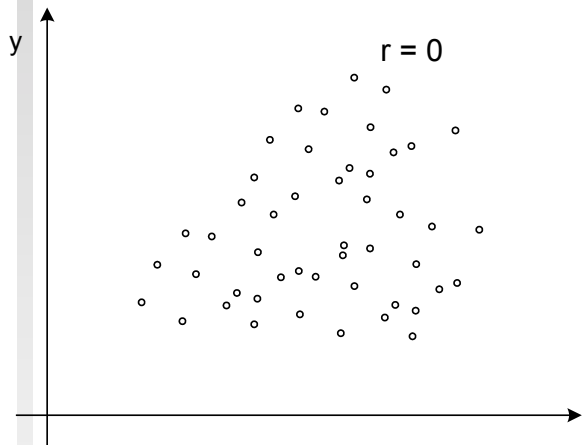
# EMPIRIJSKA RASPODELA

Grafički prikaz rasprostiranja podataka



# EMPIRIJSKA RASPODELA

Grafički prikaz rasprostiranja podataka

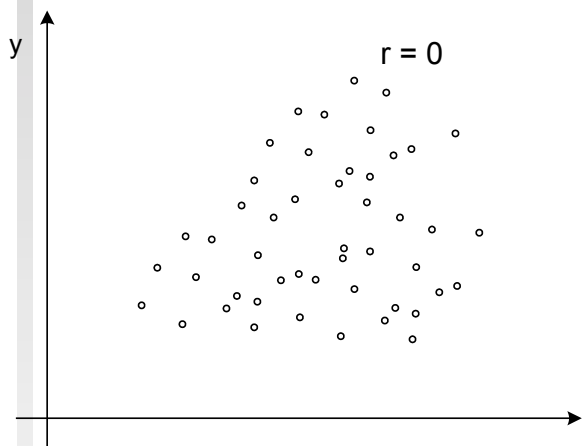


- između promenljivih  $x$  i  $y$  nema povezanosti

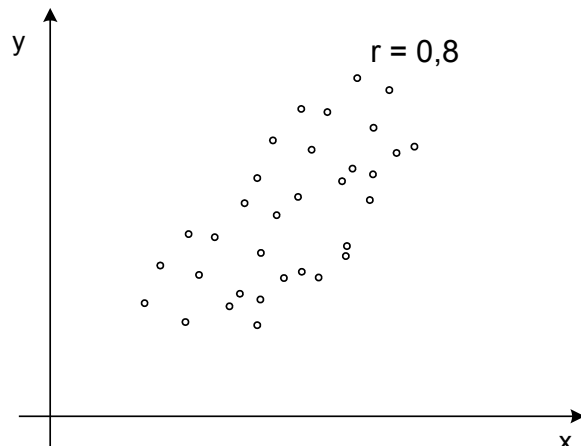


# EMPIRIJSKA RASPODELA

## Grafički prikaz rasprostiranja podataka



- između promjenljivih  $x$  i  $y$  nema povezanosti

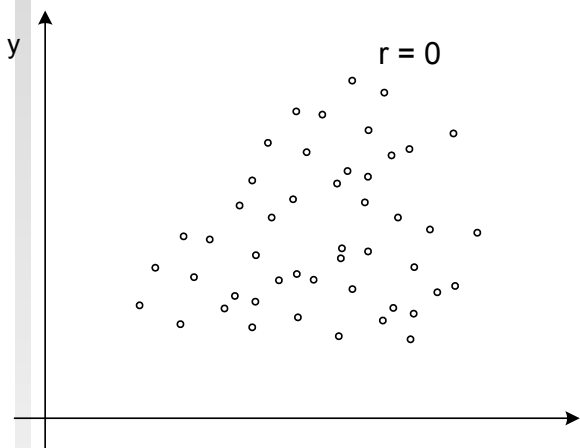


- pozitivna povezanost  
porast promjenljive  $x \rightarrow$   
porast promjenljive  $y$

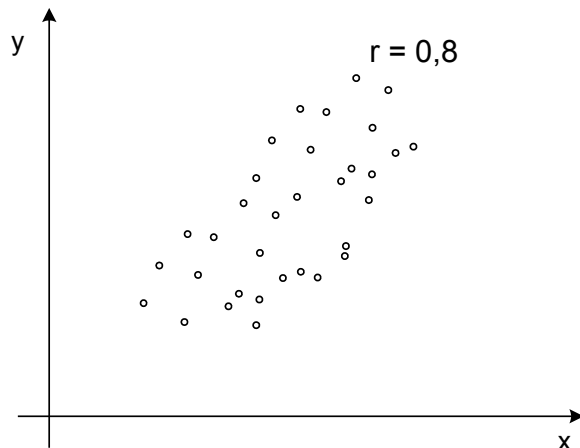


# EMPIRIJSKA RASPODELA

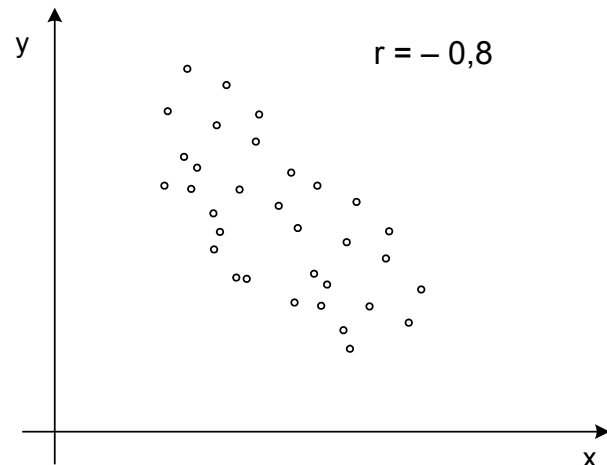
## Grafički prikaz rasprostiranja podataka



- između promenljivih  $x$  i  $y$  nema povezanosti



- pozitivna povezanost  
porast promenljive  $x \rightarrow$   
porast promenljive  $y$



- negativna povezanost  
porast promenljive  $x \rightarrow$   
smanjenje promenljive  $y$

# REGRESIJA

- izračunavanje matematičkog modela povezanosti (uticaja) promenljive  $x$  (uzroka) i promenljive  $y$  (posledice):  $y = f(x)$

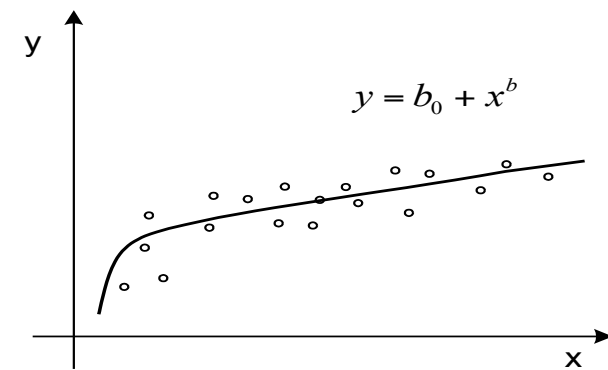
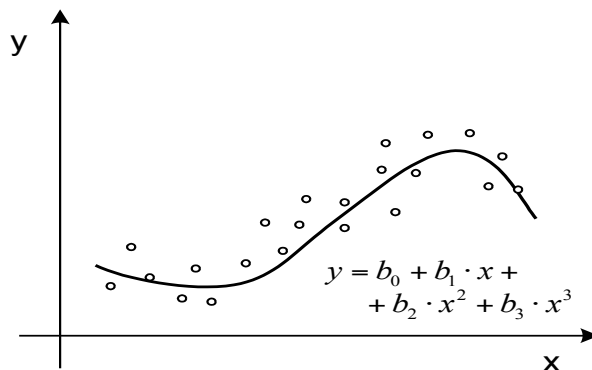
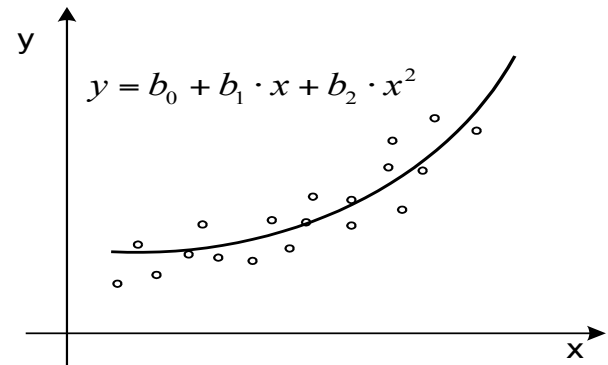
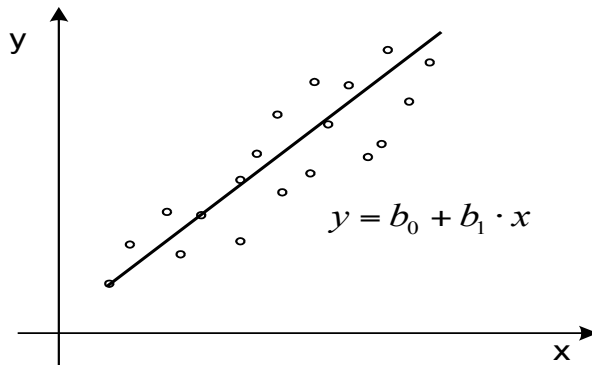
# REGRESIJA

- izračunavanje matematičkog modela povezanosti (uticaja) promenljive  $x$  (uzroka) i promenljive  $y$  (posledice):  $y = f(x)$
- oblici regresije – linearna, nelinearna



# REGRESIJA

- izračunavanje matematičkog modela povezanosti (uticaja) promenljive  $x$  (uzroka) i promenljive  $y$  (posledice):  $y = f(x)$
- oblici regresije – linearna, nelinearna





# UTVRĐIVANJE KORELACIJE

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} \quad \rightarrow \quad \text{kovarijansa izmedju x i y}$$

$$s_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad \rightarrow \quad \text{varijansa promenljive x}$$

$$s_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2 \quad \rightarrow \quad \text{varijansa promenljive y}$$

$$n \quad \rightarrow \quad \text{broj parova podataka } (x_i, y_i)$$





# UTVRĐIVANJE KORELACIJE

koeficijent korelacije (linearne):

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} \quad \rightarrow \quad \text{kovarijansa izmedju x i y}$$

$$s_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad \rightarrow \quad \text{varijansa promenljive x}$$

$$s_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2 \quad \rightarrow \quad \text{varijansa promenljive y}$$

$$n \quad \rightarrow \quad \text{broj parova podataka } (x_i, y_i)$$



# UTVRĐIVANJE KORELACIJE

koeficijent korelacije (linearne):  $r = \frac{s_{xy}}{s_x s_y}$

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} \quad \rightarrow \quad \text{kovarijansa izmedju x i y}$$

$$s_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad \rightarrow \quad \text{varijansa promenljive x}$$

$$s_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2 \quad \rightarrow \quad \text{varijansa promenljive y}$$

$n$   $\rightarrow$  broj parova podataka  $(x_i, y_i)$



# UTVRĐIVANJE KORELACIJE

koeficijent korelacije (linearne):  $r = \frac{s_{xy}}{s_x s_y}$

$$-1 \leq r \leq +1$$

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} \rightarrow \text{kovarijansa izmedju x i y}$$

$$s_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \rightarrow \text{varijansa promenljive x}$$

$$s_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2 \rightarrow \text{varijansa promenljive y}$$

$n \rightarrow$  broj parova podataka  $(x_i, y_i)$



# (PRVA) REGRESIONA PRAVA

◆ Jednačina LINEARNE regresije:

$$\hat{y} = ax + b$$



# (PRVA) REGRESIONA PRAVA

- ◆ Jednačina LINEARNE regresije:  $\hat{y} = ax + b$
- ◆ Koeficijenti  $a$  i  $b$  određuju se po “metodi najmanjih kvadrata”:

- $a$  (koeficijent pravca)  $a = \frac{s_{xy}}{s_x^2}$

- $b$  (odsečak na y osi)  $b = \bar{y} - a\bar{x}$



# (DRUGA) REGRESIONA PRAVA

- ◆ Jednačina LINEARNE regresije:  $\hat{x} = cy + d$

$$c = \frac{s_{xy}}{s_y^2}$$

$$d = \bar{x} - c\bar{y}$$