

# UZORAK I POPULACIJA

# UZORAK I POPULACIJA

## POPULACIJA

- osnovni skup
- skup svih jedinica promatranja (entiteta) opisanih varijablama (atributima)

**UZORAK** - dio jedinica populacije (osnovnog skupa)

## TEORIJA UZORAKA

- ustanovljava svojstva populacije iz svojstava uzorka
- procjenjuje parametre populacije na temelju parametara uzorka i ocjenjuje pouzdanost te procjene

# UZORAK I POPULACIJA

## UOBIČAJENE OZNAKE

	OČJENA PARAMETRA (STATISTIKA)	PARAMETAR POPULACIJE
ARITMETIČKA SREDINA	$\bar{X}$	$\mu$
STANDARDNA DEVIJACIJA	$s$	$\sigma$
PROPORCIJA	$p$	$\pi$

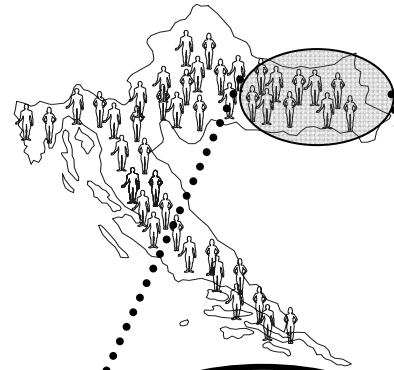
# OSNOVNI POJMOVI

Koja je skupina na koju želimo generalizirati?

Koja je populacija dostupna?

Na koji način možemo obuhvatiti populaciju?

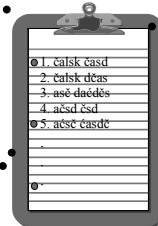
Tko je uključen u istraživanje?



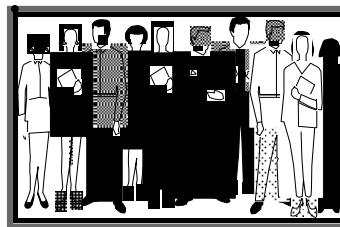
TEORETSKA  
POPULACIJA



POPULACIJA  
KOJU  
ISTRAŽUJEMO



OKVIR IZBORA



UZORAK

## UZORAK I POPULACIJA

**Kvaliteta ocjene parametara ovisi o:**

- REPREZENTATIVNOSTI UZORKA**
- ODABRANOJ VJEROJATNOSTI**

# REPREZENTATIVNI UZORAK

- uzorak koji dobro opisuje populaciju

**Na reprezentativnost uzorka utječu:**

1. Vrsta uzorka (prema metodi odabira)
2. Veličina uzorka
3. Varijabilnost promatranog obilježja

# VRSTE UZORAKA

## PROBABILISTIČKI (probability samples)

- svaka jedinica promatranja u populaciji ima jednaku vjerojatnost izbora u uzorak koja je različita od 0

## NEPROBABILISTIČKI (non-probability samples)

- vjerojatnost izbora jedinica promatranja iz populacije je različita i nepoznata (može biti i 0)

# PROBABILISTIČKI UZORCI

JEDNOSTAVNI SLUČAJNI

SUSTAVNI SLUČAJNI

SLOJEVITI (STRATIFICIRANI)

UZORAK SKUPINE

VIŠEFAZNI



# JEDNOSTAVNI SLUČAJNI UZORAK

*(simple random sample)*

## svojstva:

- *svaki element* populacije ima *jednaku šansu* da bude izabran
- *svaki uzorak* ima *jednaku šansu* da bude izabran

## način izbora:

- lutrijska metoda
- pomoću tablice slučajnih brojeva
- pomoću programske podrške koja ima funkciju generatora slučajnih brojeva

# SUSTAVNI SLUČAJNI UZORAK

## *(systematic sample)*

- jedinice koje ulaze u uzorak odabiru se po nekakvom pravilu

### **postupak:**

- numerirati jedinice populacije od 1 do N
- odrediti potrebnu veličinu uzorka (n)
- odrediti veličinu intervala  $k=N/n$
- slučajno odabrati broj između 1 i k (početna jedinica)
- uzimati svaku k-tu jedinicu

# SLOJEVITI (STRATIFICIRANI) UZORAK

- primjenjuje se u slučajevima kad je promatrano obilježje heterogeno u populaciji
- dobiva se uzimanjem jednostavnih slučajnih uzoraka iz stratumata određenih obzirom na promatrano obilježje

## postupak:

- podijeliti populaciju na disjunktne skupine od  $n_1, n_2, \dots, n_s$  jedinica, pri čemu je
$$n_1 + n_2 + \dots + n_s = N$$
- uzeti jednostavni slučajni uzorak od  $f_i = n_i/N$  jedinica iz svake skupine

# UZORAK SKUPINE (CLUSTER)

- primjenjuje se u slučajevima kada treba uzeti uzorak iz populacije koja se sastoji od skupina jedinica (ulice, popisni krugovi, škole, općine, ...)

## postupak:

- podijeliti populaciju na skupine jedinica
- jednostavnim slučajnim izborom odabrati skupine
- ispitati SVE jedinice unutar odabranih skupina

# VIŠEFAZNI UZORAK

- uzimanje "uzorka iz uzorka"
- kombinacija više metoda odabiranja uzorka

Npr. jedan od mogućih načina dobivanja uzorka iz populacije učenika osnovnih škola u Hrvatskoj:

- podijeliti osnovne škole u stratumе s obzirom na županijsku pripadnost
- iz svakog stratuma jednostavnim slučajnim izborom odabrati škole (prva faza)
- unutar odabranih škola, jednostavnim slučajnim izborom odabrati razrede (druga faza)
- unutar odabranih razreda, jednostavnim slučajnim izborom odabrati učenike (treća faza)

# NEPROBABILISTIČKI UZORCI

PRIGODNI (convenience)

UZORAK KOJI SLUŽI SVRSI (purposive)

UZORAK UDJELA (quota)

## **PRIGODNI UZORAK (convenient sample)**

- u uzorak se biraju jedinice populacije koje su "pri ruci" (npr. prolaznici, pozvani dobrovoljci, prvih 50 pacijenata u nekoj ambulanti)

## **UZORAK KOJI SLUŽI SVRSI (purposive sample)**

- u uzorak se biraju jedinice populacije koje imaju traženo svojstvo

## **UZORAK UDJELA (quota sample)**

- u uzorak se bira određeni broj jedinica odabranih dijelova populacije

U kojem od sljedećeg se koristi jednostavni slučajni uzorak:

- a) igra "Bingo",
- b) popis stanovništva,
- c) izbori za lokalnu samoupravu?

Koje metode odabira uzorka se koriste u ovim primjerima?



- a) igra "Bingo" - jednostavni slučajni uzorak
- b) popis stanovništva - ne koristi jednostavni slučajni uzorak jer SVE jedinice populacije moraju biti obuhvaćene popisom
- c) izbori za lokalnu samoupravu - neprobabilistički uzorak; na izbore izlaze oni koji to žele ("dobrovoljci")

# VELIČINA UZORKA

## *dovoljno veliki uzorak:*

- uzorak pomoću kojeg s razumnom pouzdanošću možemo prihvatiti ili odbaciti neku hipotezu i ocijeniti parametar populacije

## *ovisit će o:*

- homogenosti populacije s obzirom na promatrano obilježje
- učestalosti promatranog obilježja (obrnuto proporcionalno)

# UTJECAJ VARIJABILNOSTI

- varijabilnost je često nepoznata
- poznata, a velika varijabilnost ugrožava reprezentativnost uzorka
- utjecaj varijabilnosti se smanjuje s povećanjem uzorka

# STANDARDNA POGREŠKA

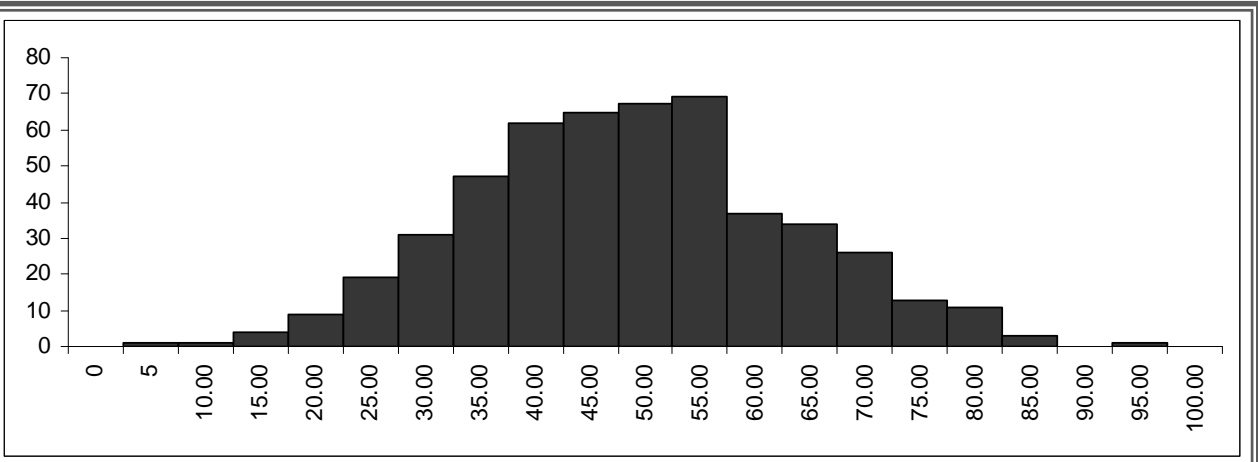
POKUS 1. Napraviti razdiobe aritmetičkih sredina 500 uzoraka veličine  $n=4$ ,  $n=20$  i  $n=50$  osnovnog skupa  $N=101$  brojeva od 0 do 100.

aritmetička sredina osnovnog skupa

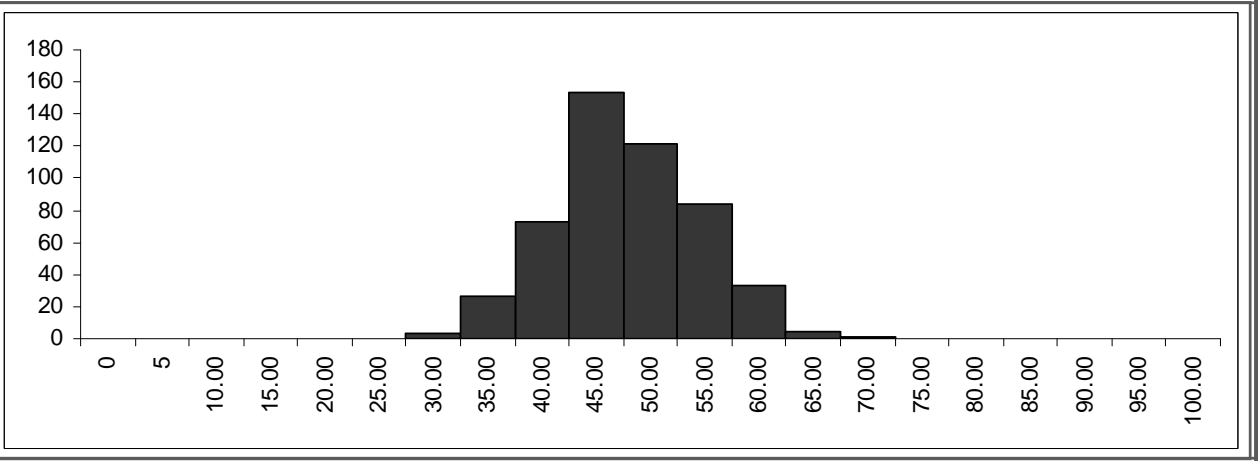
$$\mu = 50$$

standardna devijacija osnovnog skupa

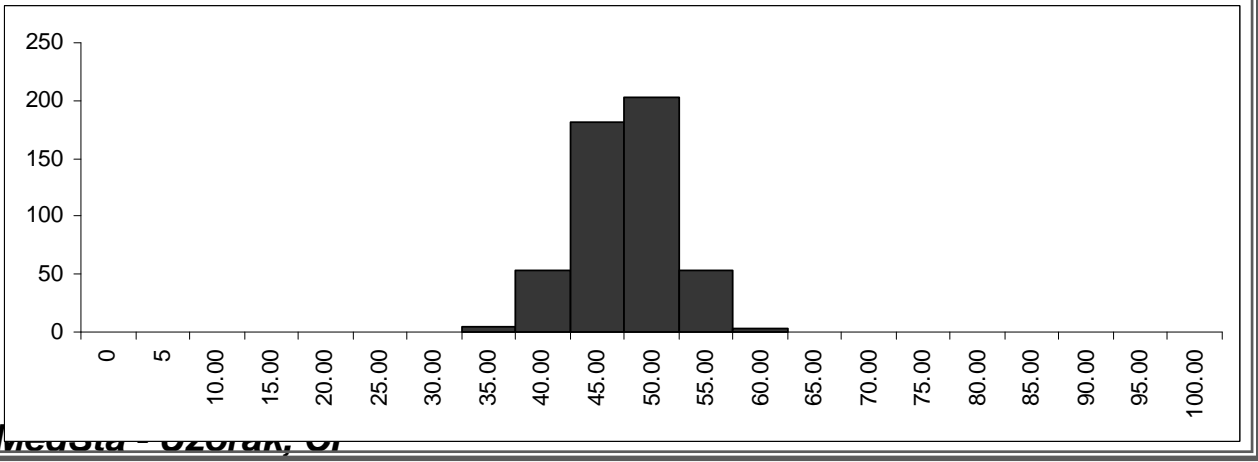
$$\sigma = 29.15$$



$n = 4$   
 $\bar{x} = 50.96$   
 $s = 14.451$



$n = 20$   
 $\bar{x} = 50.10$   
 $s = 6.639$



$n = 50$   
 $\bar{x} = 50.07$   
 $s = 4.189$

$N$  .... veličina osnovnog skupa

$n$  .... veličina slučajnih uzoraka

$\binom{N}{n}$  broj svih mogućih uzoraka veličine  $n$  uzetih iz osnovnog skupa veličine  $N$

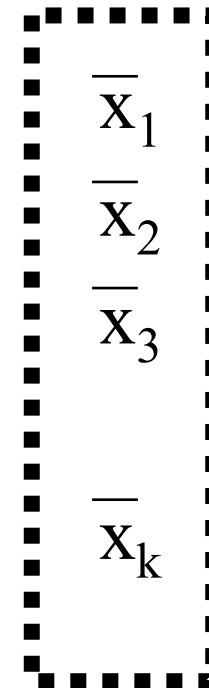
1. uzorak       $x_{11}, x_{12}, \dots, x_{1n}$       sa sredinom

2. uzorak       $x_{21}, x_{22}, \dots, x_{2n}$       sa sredinom

3. uzorak       $x_{31}, x_{32}, \dots, x_{3n}$       sa sredinom

....

k-ti uzorak       $x_{k1}, x_{k2}, \dots, x_{kn}$       sa sredinom



$\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_k$

slučajna varijabla  
(*sampling distribucija*)

**OČEKIVANJE**

$$E(\bar{X}) = \mu$$

**VARIJANCA**

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

standardna  
devijacija

$$S_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

**STANDARDNA  
POGREŠKA**  
(SE, standard  
error)



- za slučajne i dovoljno velike uzorke

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

**STANDARDNA  
POGREŠKA  
(SE, standard error)**

- standardna pogreška aritmetičke sredine (SEM, *standard error of the mean*)
- pogreška kojoj se izlažemo pri zaključivanju o populaciji na temelju uzorka

$s \uparrow \Rightarrow s_{\bar{x}} \uparrow$  *povećava se s povećanjem  
varijabilnosti obilježja*

$n \uparrow \Rightarrow s_{\bar{x}} \downarrow$  *smanjuje se s povećanjem veličine  
uzorka*

# CENTRALNI GRANIČNI TEOREM



Razdioba aritmetičkih sredina uzoraka teži normalnoj razdiobi s očekivanjem  $\mu$  i varijancom  $\sigma^2/x$   $[N(\mu, \sigma^2/x)]$  kad veličina uzorka  $n$  teži u beskonačnost.

⇒ *za dovoljno velike uzorke* razdioba aritmetičkih sredina uzoraka bit će *normalna*, bez obzira na razdiobu vrijednosti promatranog obilježja

za proporciju:

$$s_p = \sqrt{\frac{p \cdot q}{n}}$$

**STANDARDNA  
POGREŠKA  
PROPORCIJE**

$p \uparrow \Rightarrow s_{\bar{x}} \downarrow$  *smanjuje se s povećanjem homogenosti  
obilježja*

$n \uparrow \Rightarrow s_{\bar{x}} \downarrow$  *smanjuje se s povećanjem veličine  
uzorka*

# STANDARDNA POGREŠKA VS STANDARDNA DEVIJACIJA

## STANDARDNA POGREŠKA:

- procjenjuje "kvalitetu" ocjene parametra (statistike)
- velika standardna pogreška => ocjena parametra (ar. sredina, proporcija) je neprecizna

## STANDARDNA DEVIJACIJA:

- opisuje varijabilnost podataka
- velika standardna devijacija => velika varijabilnost podataka

# RASPON POUZDANOSTI

# RASPON POUZDANOSTI

- confidence interval (CI)

- uobičajeno tumačenje:

raspon unutar kojega se, s određenom vjerojatnošću, nalazi prava vrijednost (parametar) populacije

## RASPON POUZDANOSTI ARITMETIČKE SREDINE

$$\bar{X} - z \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + z \cdot s_{\bar{X}}$$

## RASPON POUZDANOSTI PROPORCIJE

$$p - z \cdot s_p \leq \Pi \leq p + z \cdot s_p$$

**z** - standardizirana vrijednost normalne raspodjele  
(ovisi o pretpostavljenoj vjerojatnosti)

# RASPON POUZDANOSTI

PRIMJER. Koliki je raspon pouzdanosti ako želimo obuhvatiti  $\mu$  sa:

a) 99% pouzdanosti

b) 95% pouzdanosti

c) 90% pouzdanosti

uz pretpostavku normalne razdiobe?

a)  $z_{0.005}=2.576\approx 2.58$

$$\bar{x} - 2.58 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 2.58 \cdot s_{\bar{x}}$$

b)  $z_{0.025}=1.96$

$$\bar{x} - 1.96 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 1.96 \cdot s_{\bar{x}}$$

c)  $z_{0.05}=1.65$

$$\bar{x} - 1.65 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 1.65 \cdot s_{\bar{x}}$$

## RASPON POUZDANOSTI

PRIMJER. Od 1000 ljudi koji su cijepljeni, 200 ih je pokazalo alergične reakcije. Koliku proporciju alergičnih očekujemo u populaciji cijepljenih uz vjerojatnost od 95%?

$$z_{0.025}=1.96$$

$$p - 1.96 \cdot s_p \leq \Pi \leq p + 1.96 \cdot s_p$$

$$p=0.20; \quad q=0.80;$$

$$s_p = \sqrt{\frac{0.2 \cdot 0.8}{1000}} = \sqrt{\frac{0.16}{1000}} = \sqrt{0.00016} = 0.0126$$

$$0.2 - 1.96 \cdot 0.0126 \leq \Pi \leq 0.2 + 1.96 \cdot 0.0126$$

$$0.2 - 0.025 \leq \Pi \leq 0.2 + 0.025$$

$$0.175 \leq \Pi \leq 0.225$$



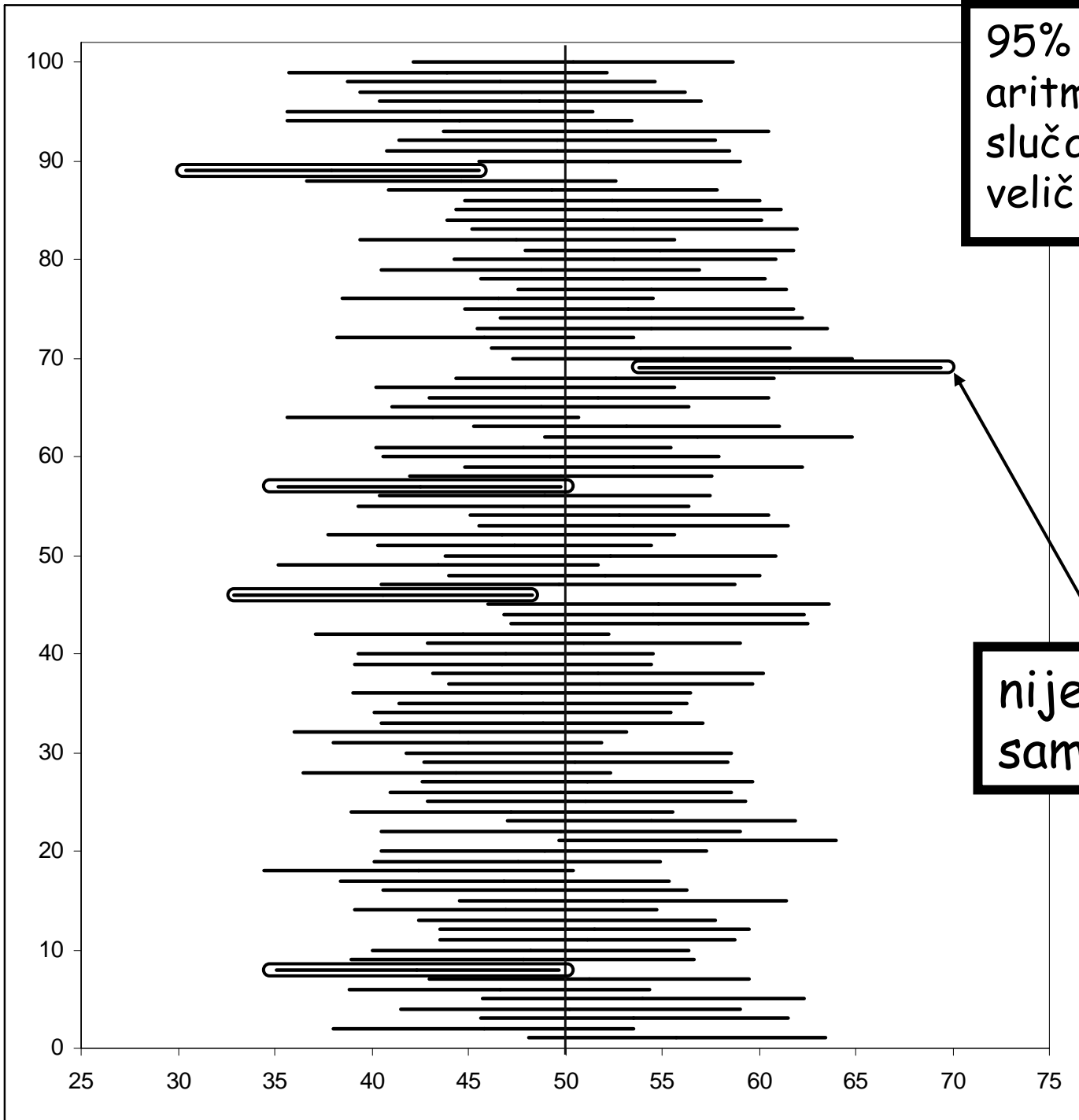
## 95% raspon pouzdanosti aritmetičke sredine izračunat iz nekog uzorka:

- *uobičajeno* se tumači kao raspon vrijednosti unutar kojeg se s 95% pouzdanosti nalazi prava vrijednost aritmetičke sredine (aritmetička sredina populacije)
- *u stvari znači* da očekujemo da 95% takvih intervala dobivenih iz uzoraka iste veličine dane populacije uključuje pravu vrijednost aritmetičke sredine

## 95% raspon pouzdanosti:

*Slučajan interval čije granice se mogu izračunati iz podataka o uzorku, takav da 95 od svakih 100 takvih intervala obuhvaća pravu vrijednost parametra koji se procjenjuje.*

- također i raspon poželjnih vrijednosti parametra populacije (prihvatljiva nul-hipoteza)



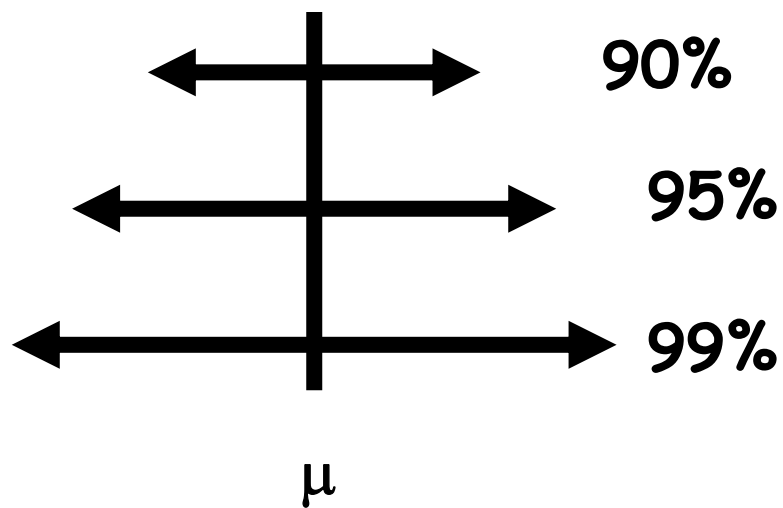
95% rasponi pouzdanosti  
aritmetičkih sredina 100  
slučajnih uzoraka  
veliĉine  $n=50$

nije "pogrešan",  
samo nema sreće

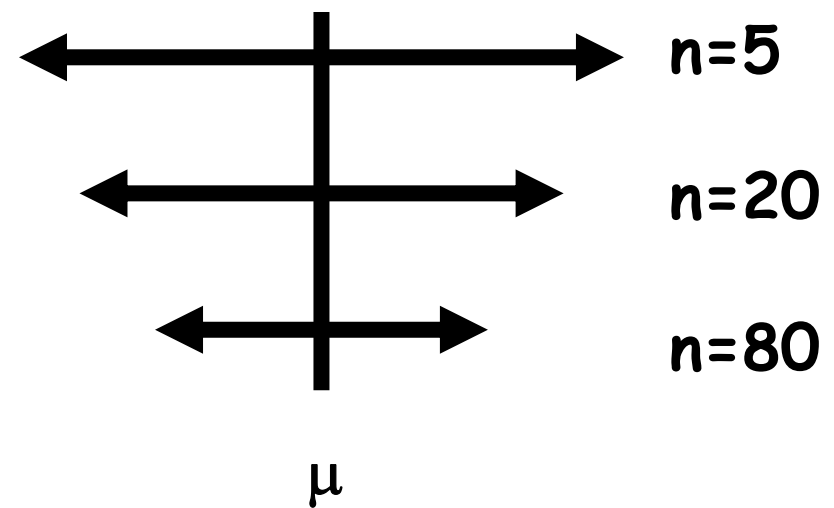
# Širina raspona pouzdanosti ovisi o:

- pretpostavljenoj vjerojatnosti
- varijabilnosti promatranog obilježja
- veličini uzorka

širi se s  
povećanjem pouzdanosti



sužava se s  
povećanjem uzorka



# POTREBNA VELIČINA UZORKA za procjenu aritmetičke sredine

Ovisit će o:

- pogrešci procjene koju ćemo tolerirati
- stupnju pouzdanosti
- pretpostavljenoj varijabilnosti

$$E = z \cdot s_{\bar{x}} = z \cdot \frac{s}{\sqrt{n}}$$

POGREŠKA PROCJENE

$$n = \left( \frac{z \cdot s}{E} \right)^2$$

POTREBNA VELIČINA  
UZORKA

PRIMJER. Koliko ispitanika treba izabrati u uzorak kako bi se procijenila prosječna starost stanovnika nekog sela u 95% rasponu pouzdanosti od 2 godine? Pretpostavlja se kako je standardna devijacija populacije 8 godina.

Koliki uzorak treba biti ako toleriramo pogrešku od najviše  $\pm 10$  mjeseci?

$$z_{0.025}=1.96$$

$$E=1$$

$$\sigma = 8$$

$$n = \left( \frac{z_{0.025} \cdot 8}{1} \right)^2 = \left( \frac{1.96 \cdot 8}{1} \right)^2 = 15.68^2 = 245.86 \approx 246$$

$$z_{0.025}=1.96$$

$$E=10/12=0.83$$

$$\sigma = 8$$

$$n = \left( \frac{z_{0.025} \cdot 8}{0.83} \right)^2 = \left( \frac{1.96 \cdot 8}{0.83} \right)^2 = 18.89^2 = 356.83 \approx 357$$

# POTREBNA VELIČINA UZORKA za procjenu proporcije

Ovisit će o:

- pogrešci procjene koju ćemo tolerirati
- stupnju pouzdanosti
- pretpostavljenoj proporciji

$$\mathbf{E = z \cdot s_p = z \cdot \sqrt{\frac{p \cdot q}{n}}}$$

POGREŠKA PROCJENE

$$\mathbf{n = \left(\frac{z}{E}\right)^2 \cdot p \cdot q}$$

POTREBNA VELIČINA UZORKA

PRIMJER. Studija provedena na Fakultetu javnog zdravstva na Harvardu utvrdila je da 19% studenata nikada ne piju alkohol. Koliki uzorak vam je potreban za procjenu proporcije studenata koji ne piju alkohol na vašem fakultetu unutar raspona od 10% uz pouzdanost od 95%, vodeći se rezultatima harvardske studije?

$$z_{0.025} = 1.96$$

$$E = 0.1/2 = 0.05$$

$$p = 0.19$$

$$n = \left( \frac{z}{E} \right)^2 \cdot p \cdot q = \left( \frac{1.96}{0.05} \right)^2 \cdot 0.19 \cdot 0.81 = 39.3^2 \cdot 0.19 \cdot 0.81 = 236.49 \approx 237$$



**p = 0.5** koristimo kada nemamo prethodnih saznanja o pretpostavljenoj proporciji

$$z_{0.025}=1.96$$

$$E=0.1/2=0.05$$

$$p = 0.5$$

$$n = \left(\frac{z}{E}\right)^2 \cdot p \cdot q = \left(\frac{1.96}{0.05}\right)^2 \cdot 0.5 \cdot 0.5 = 39.3^2 \cdot 0.25 = 384.15 \approx 385$$