Uzorak i uzoračke raspodele

Nedelja 5 - Vežbe

Dr Nikola N. Grubor

Osnovni termini

- **Populacija** (skup svih istovrsnih elemenata)
  - **Ciljna populacija** (skup elemenata za koji generalizujemo)
- **Uzoracka populacija** (populacija koja je dostupna)
- **Uzorak** (podskup osnovnog skupa (populacije) izabran na osnovu kriterijuma)

Istraživačko pitanje

Osnovni skup

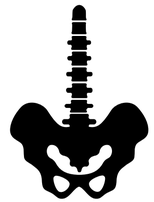
Uzorak

Jedinica posmatranja

Varijable

Istraživačko pitanje

“Da li je stepen degeneracije lumbosakruma na radiografiji povezan sa bolom?”



Ciljna grupa

Osnovni skup

Pacijenti sa lumboishialgijom

Uzorak

n pacijenata

Jedinica posmatranja

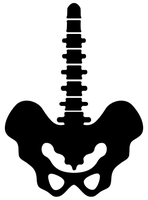
Pacijent

Varijable

Stepen degeneracije, bol

Kako da saznam
nešto o grupi pacijenata
ako ih nisam video?

“Da li je stepen degeneracije lumbosakruma na radiografiji povezan sa bolom?”



Pacijenti sa lumboishialgijom

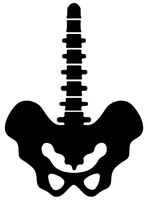
n pacijenata

Pacijent

Stepen degeneracije, bol

Ciljna populacija

“Da li je stepen degeneracije lumbosakruma na radiografiji povezan sa bolom?”



Pacijenti sa lumboishialgijom

↓

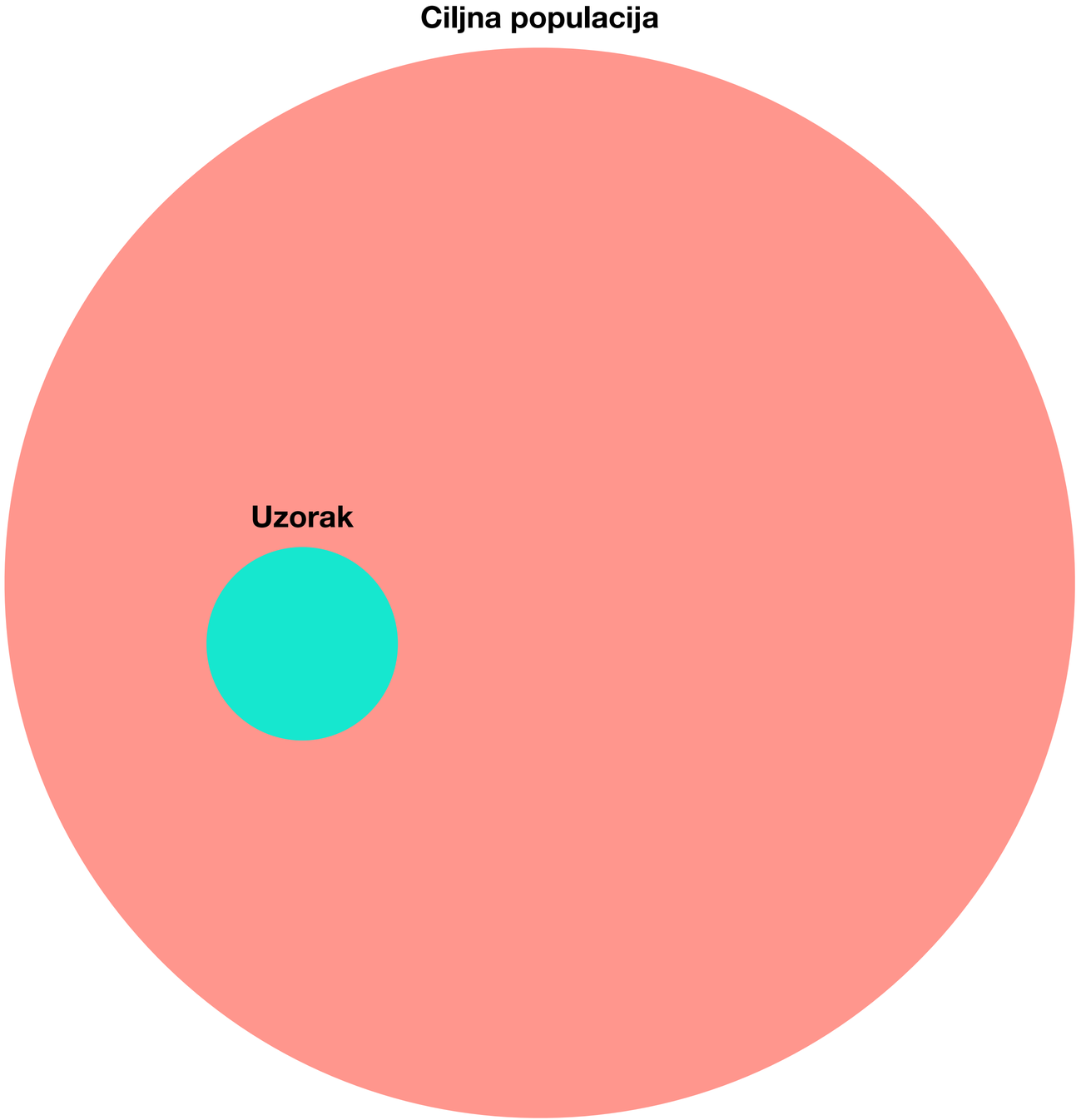
n pacijenata

↓

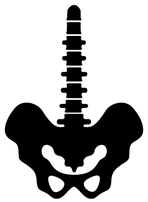
Pacijent

↓

Stepen degeneracije, bol



“Da li je stepen degeneracije lumbosakruma na radiografiji povezan sa bolom?”

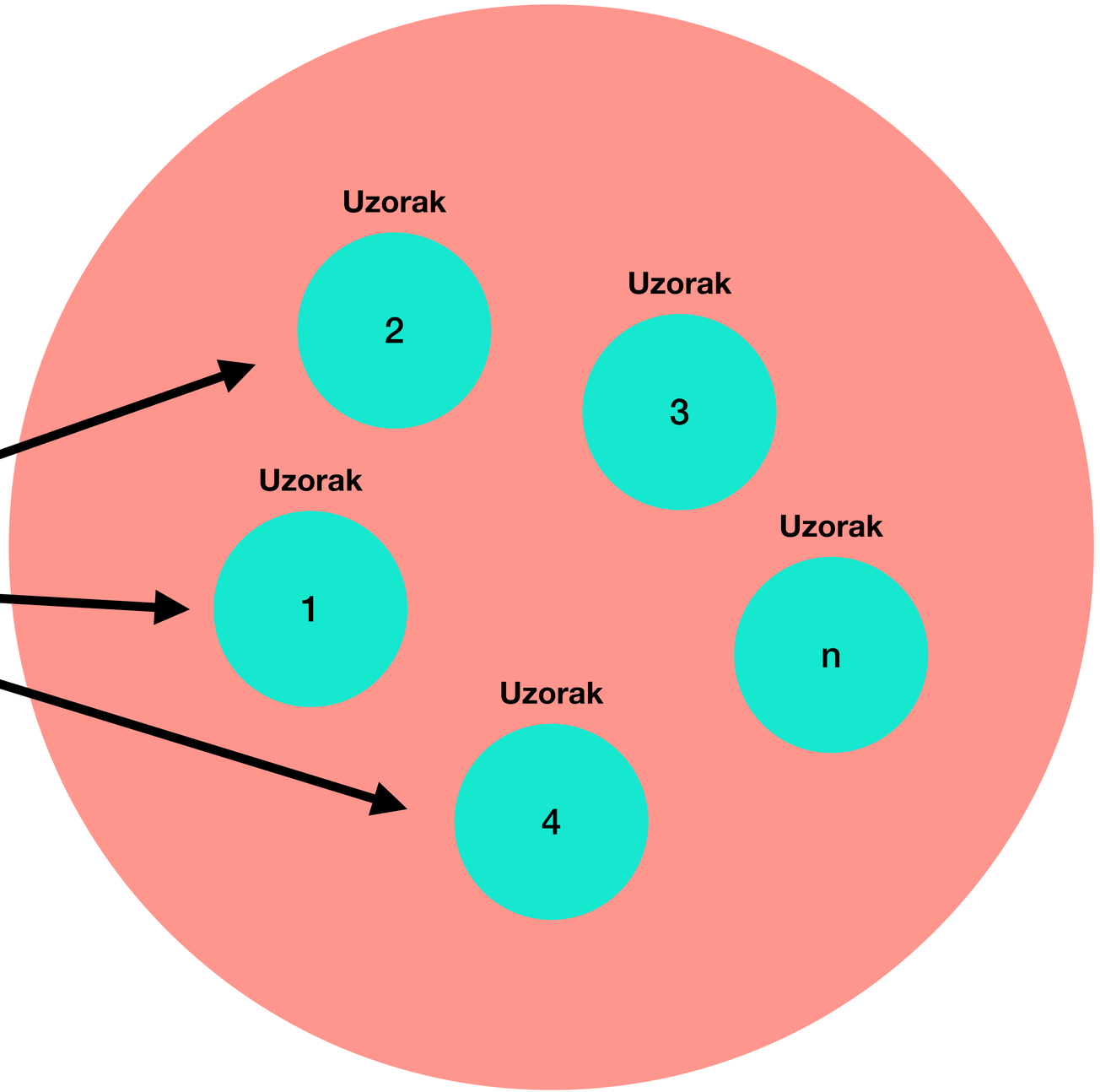


Pacijenti sa lumboishialgijom

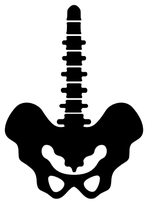
n pacijenata

Pacijent

Stepen degeneracije, bol



“Da li je stepen degeneracije lumbosakruma na radiografiji povezan sa bolom?”



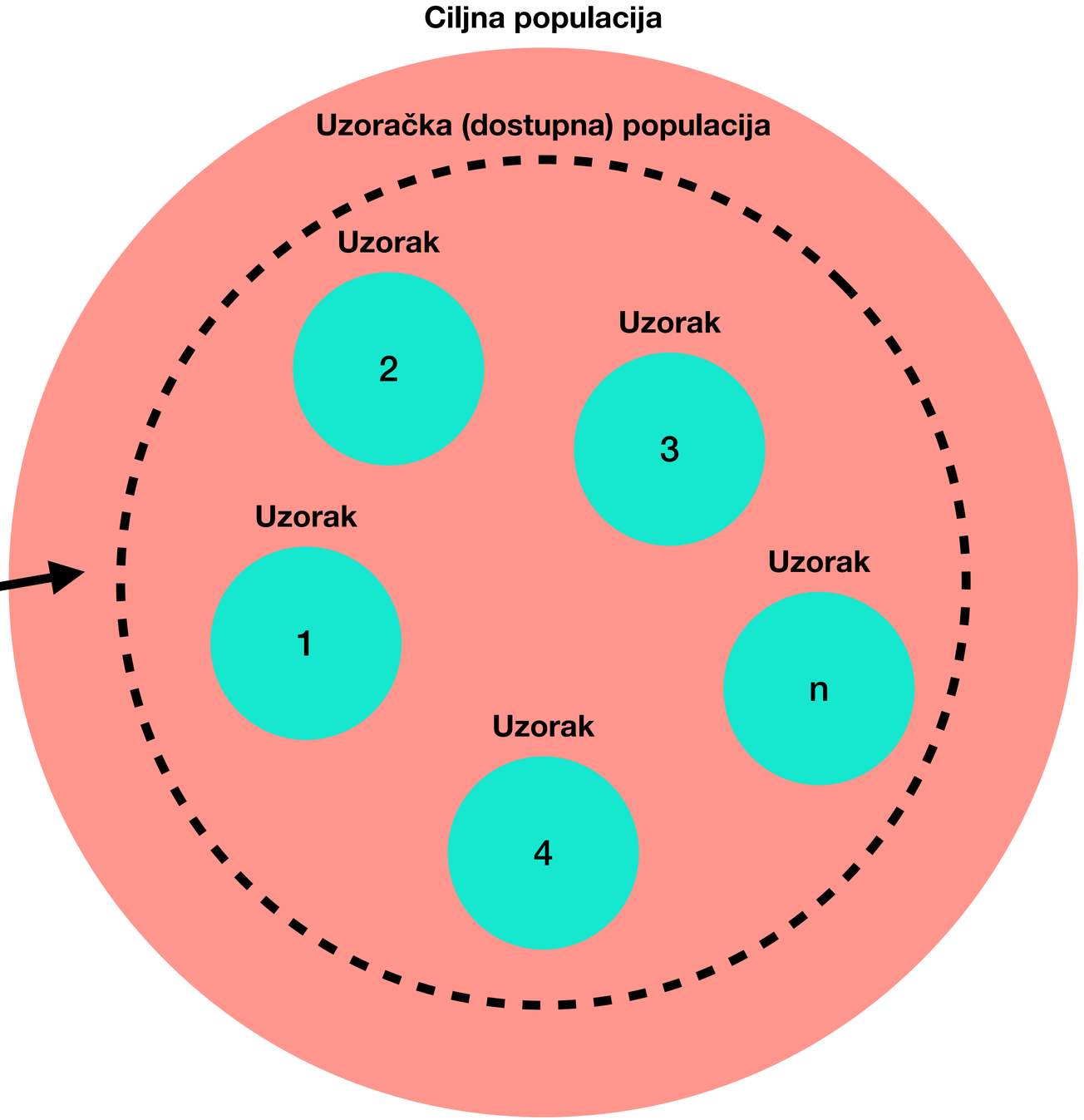
Pacijenti sa lumboishialgijom

Uzoračka (dostupna) populacija

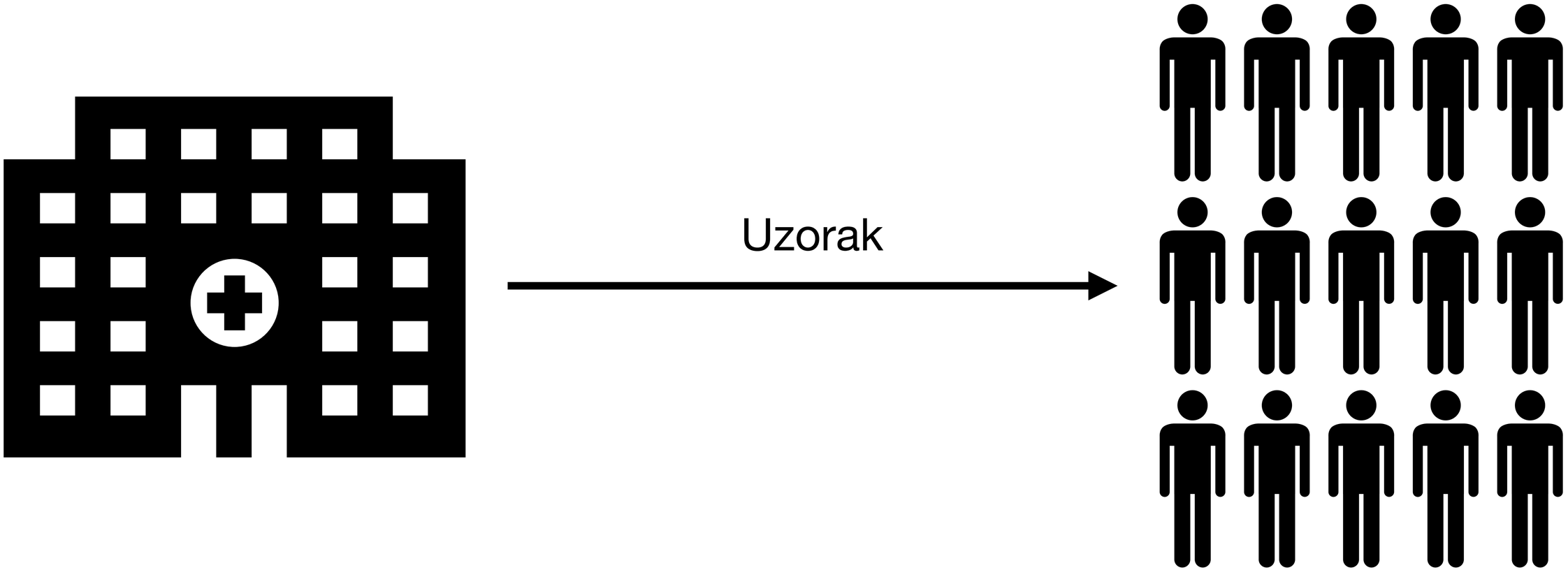
n pacijenata

Pacijent

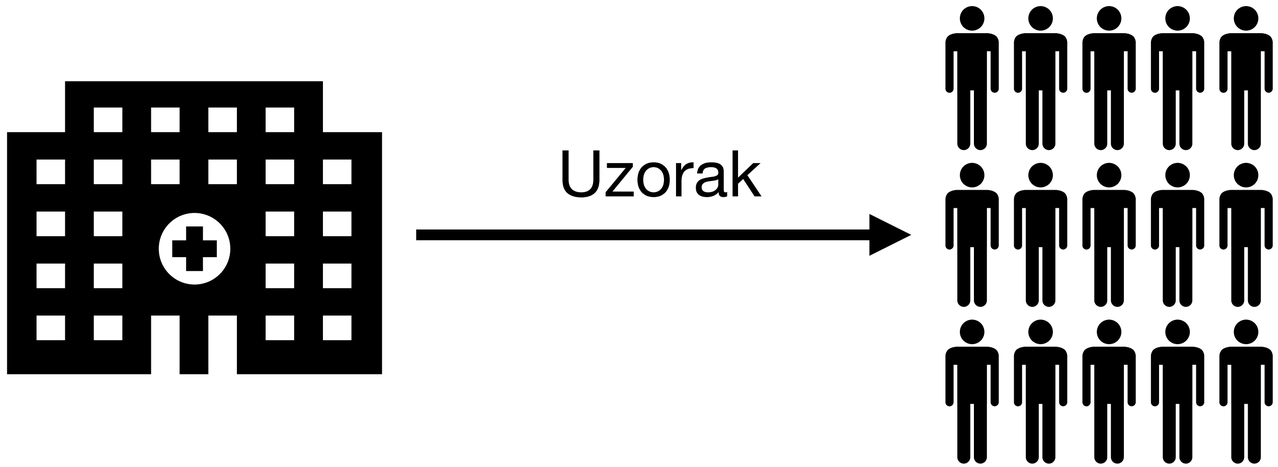
Stepen degeneracije, bol



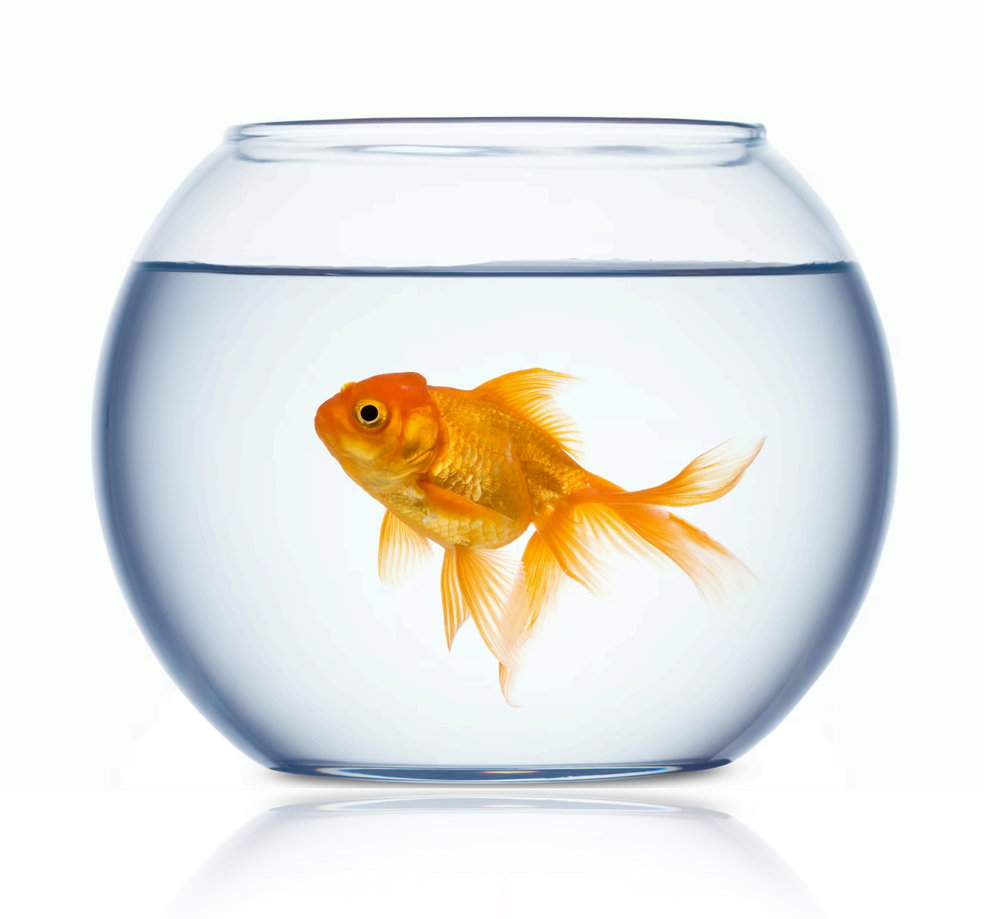
Kako se dostupna populacija razlikuje?



Kako se dostupna populacija razlikuje?

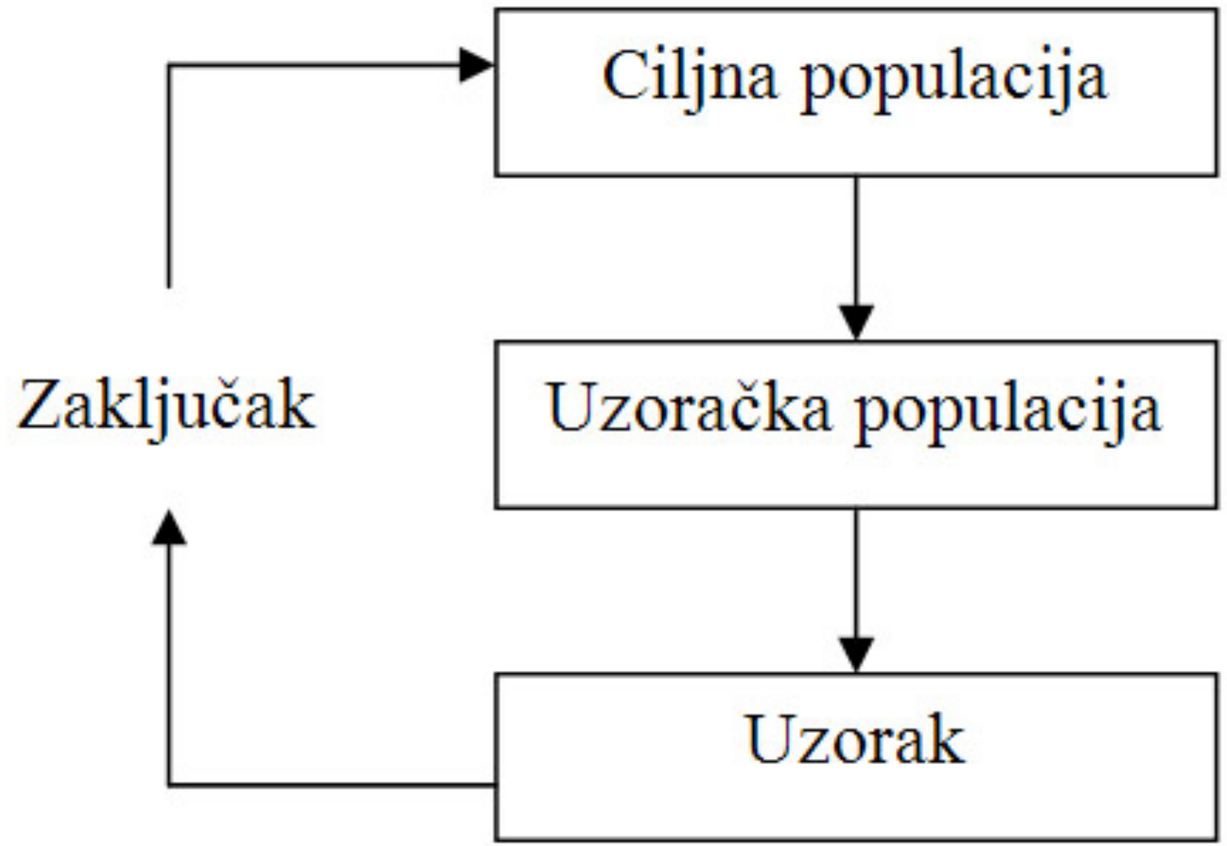


Pristrasnost?



- **Izbor jedinica posmatranja mora biti nezavisan od posmatranog obeležja.**
- **Verovatnoća odabira jedinica posmatranja mora biti unapred poznata.**

Inferenca



Zašto randomizujemo?

Težina bolesti

Novac

Lek

Motivacija

Fizički kapacitet

Lek

Uzorkovanje

Tip

- **Slučajni**
  - Tačno znamo verovatnoću da neko bude izabran
- **Neslučajni**
  - Može biti pristrasan

Uzorkovanje

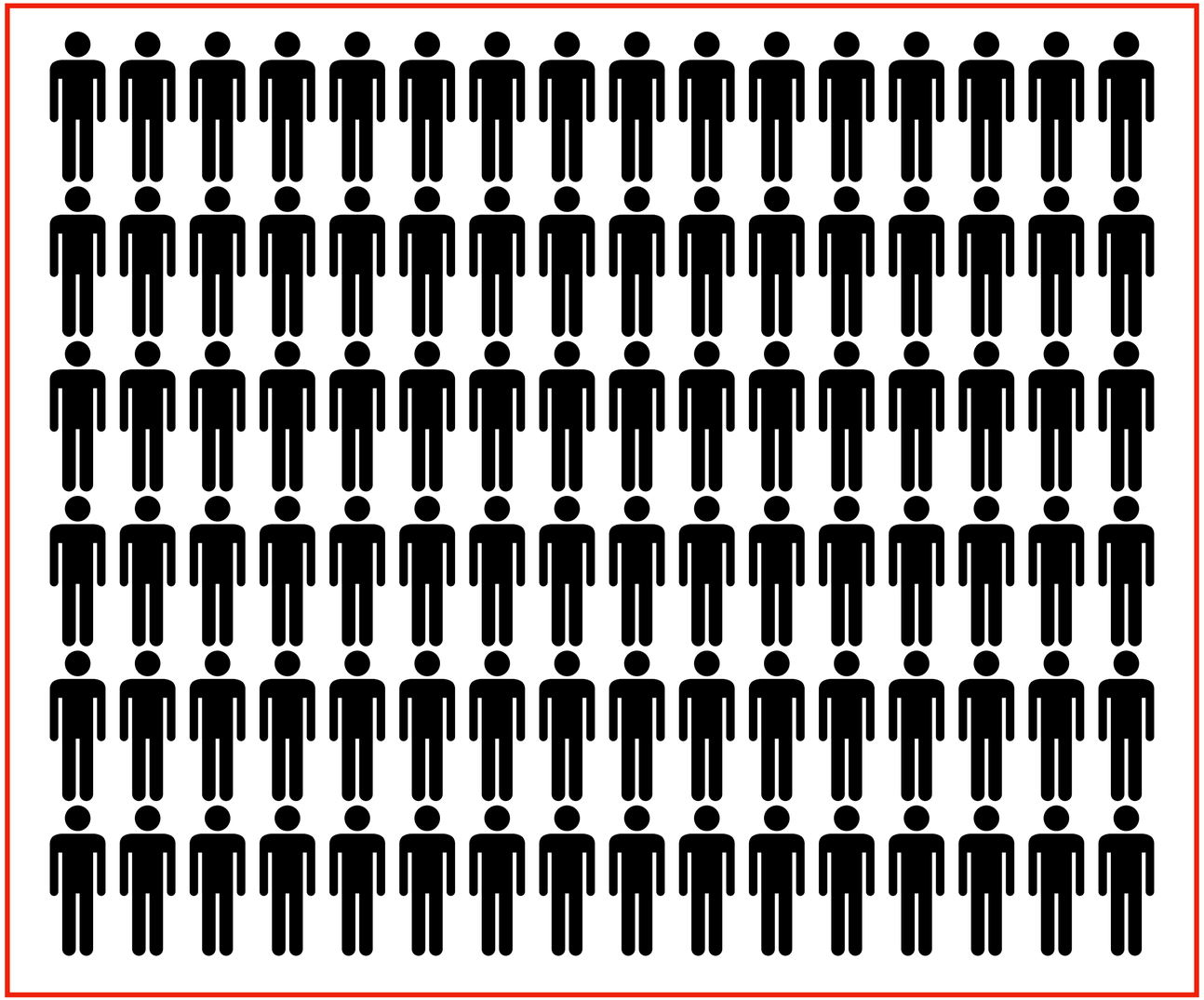
Slučajni uzorci

- Prost
- Sistematski
- Stratifikovani
- Klaster

Prost uzorak

Uzoračka okvir

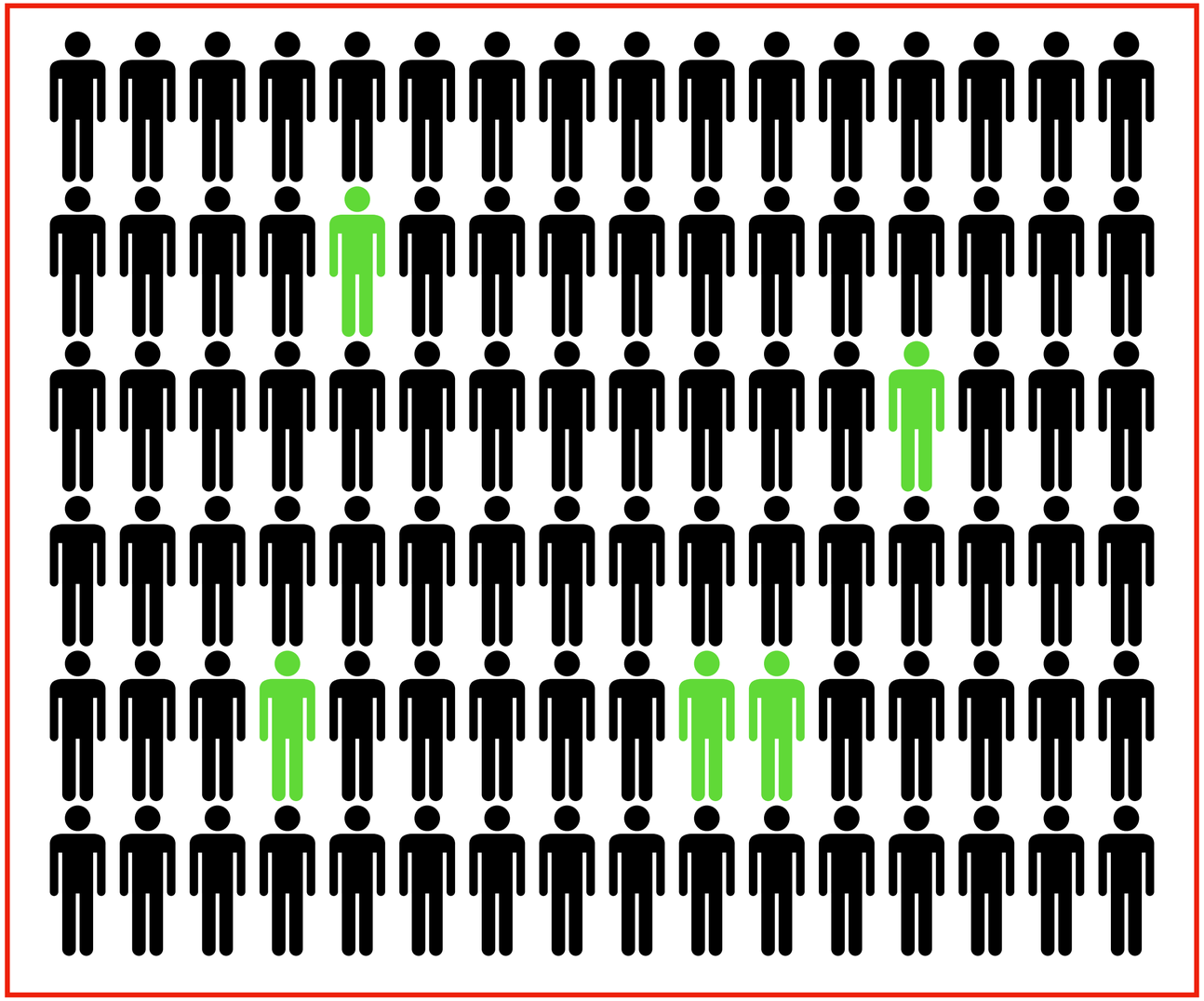
Lista svih pacijenata u bolnici



Prost uzorak

Uzorački okvir

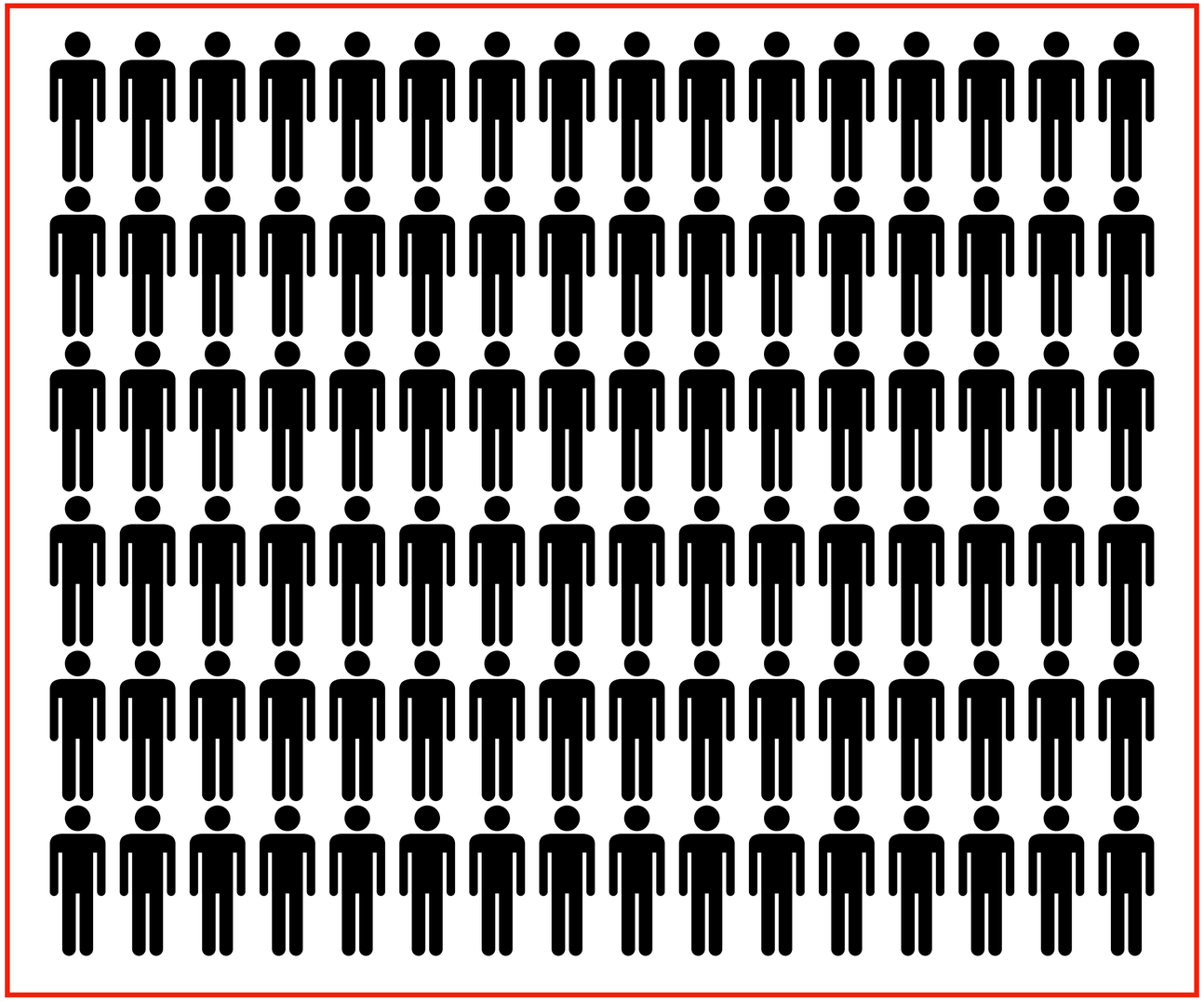
Lista svih pacijenata u bolnici



Sistematski uzorak

Uzoračka okvir

Lista svih pacijenata u bolnici

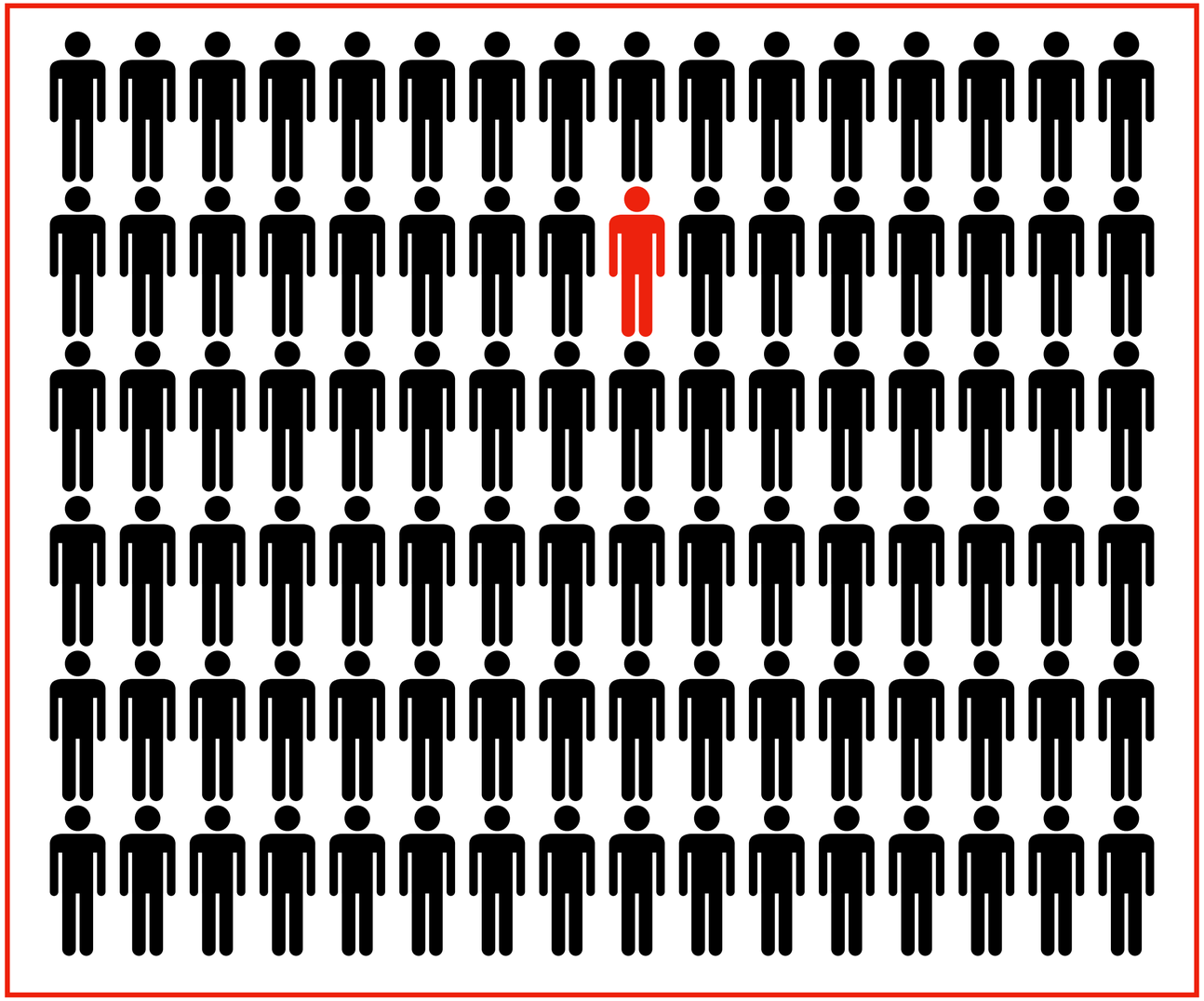


Sistematski uzorak

Uzoračkí okvir

Lista svih pacijenata u bolnici

Nasumičan početak



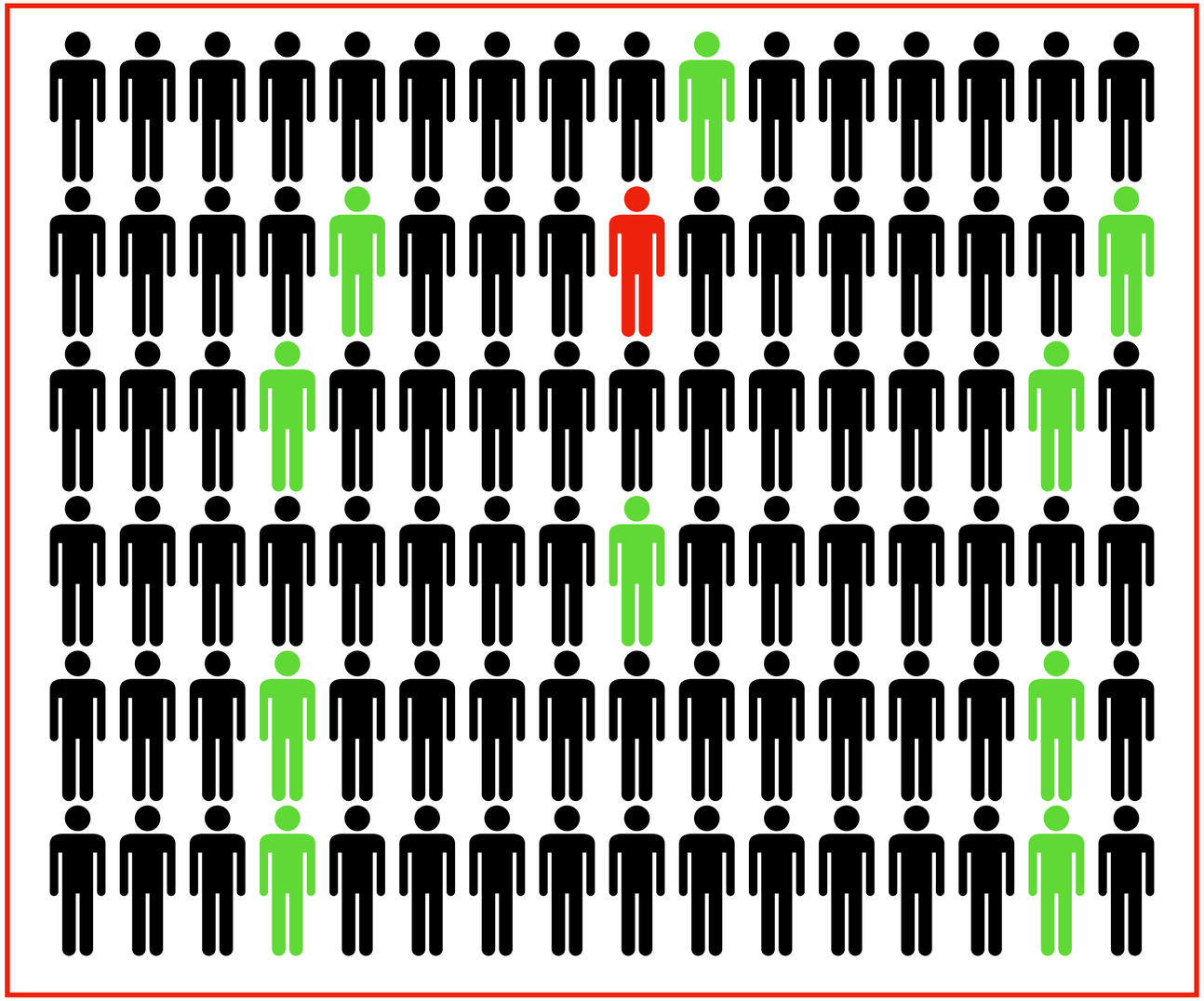
Sistematski uzorak

Uzorački okvir

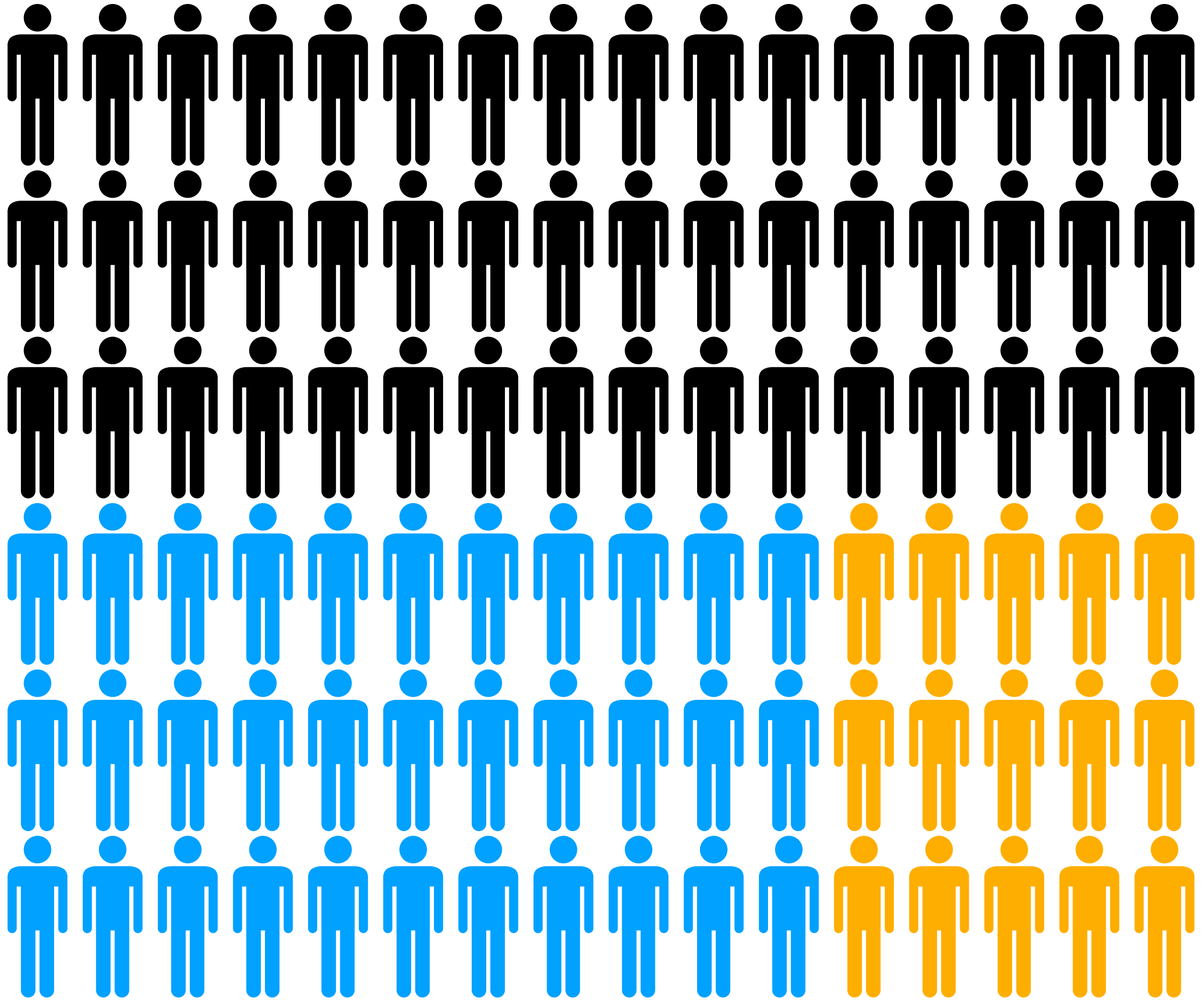
Lista svih pacijenata u bolnici

Nasumičan početak

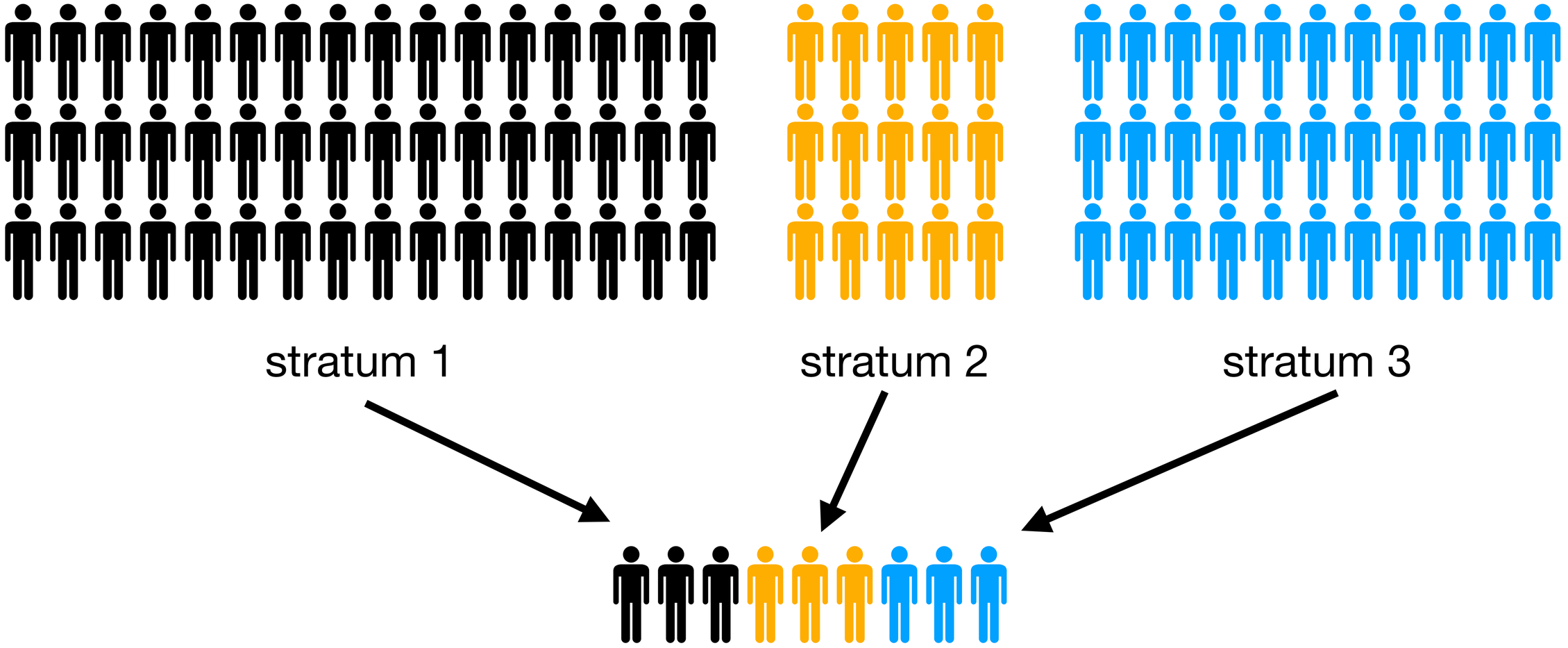
$$k = \frac{N}{n} = \frac{96}{10} \approx 10$$



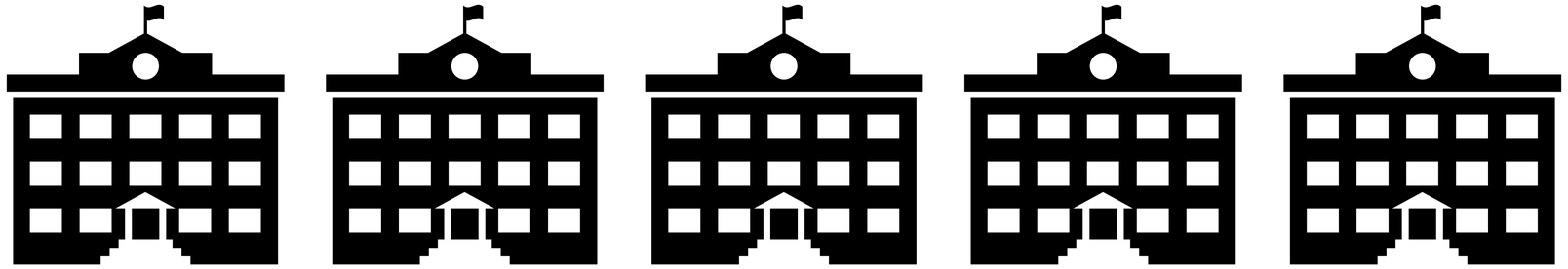
Stratifikovani uzorak



Stratifikovani uzorak

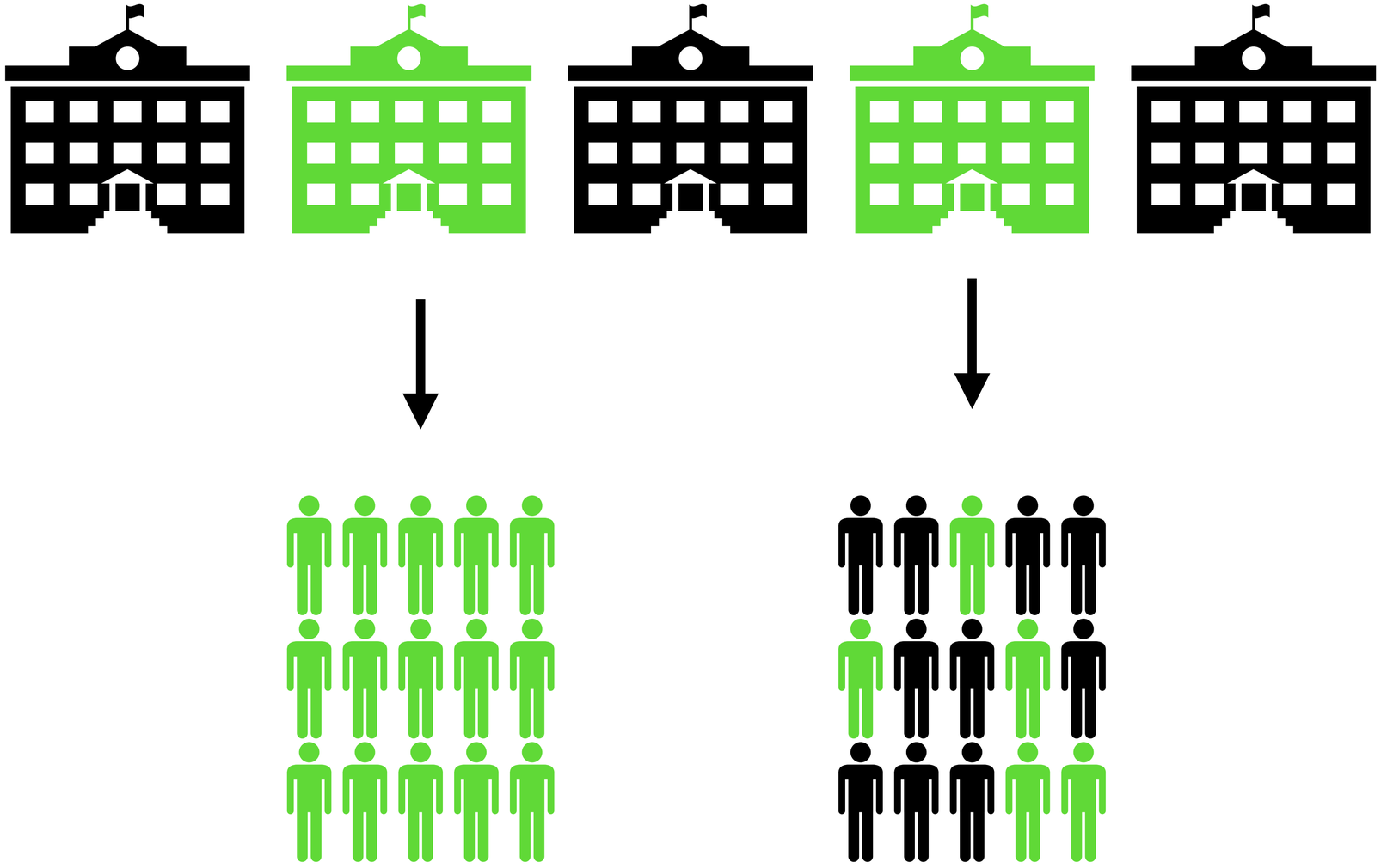


Klaster uzorak



Klaster uzorak

Biranje klastera



Prosto ili kompletno uzorkovanje

Neslučajni uzorci

- Postoji selekciona pristrasnost
- Upitna valjanost generalizacije

Neslučajni uzorci

Tip

- Prigodan
  - Pacijenti na odeljenju
- Kvota
  - Kategorije → već određena kvota se ispunjava
- Namerni
  - Pilot studija (inovacije)

Terminologija

	Populacija (parametar)	Uzorak (statistika)
Aritmetička sredina	μ	$\bar{x}$
Varijansa	σ^2	sd^2
Standardna devijacija	σ	sd
Proporcija	π	p

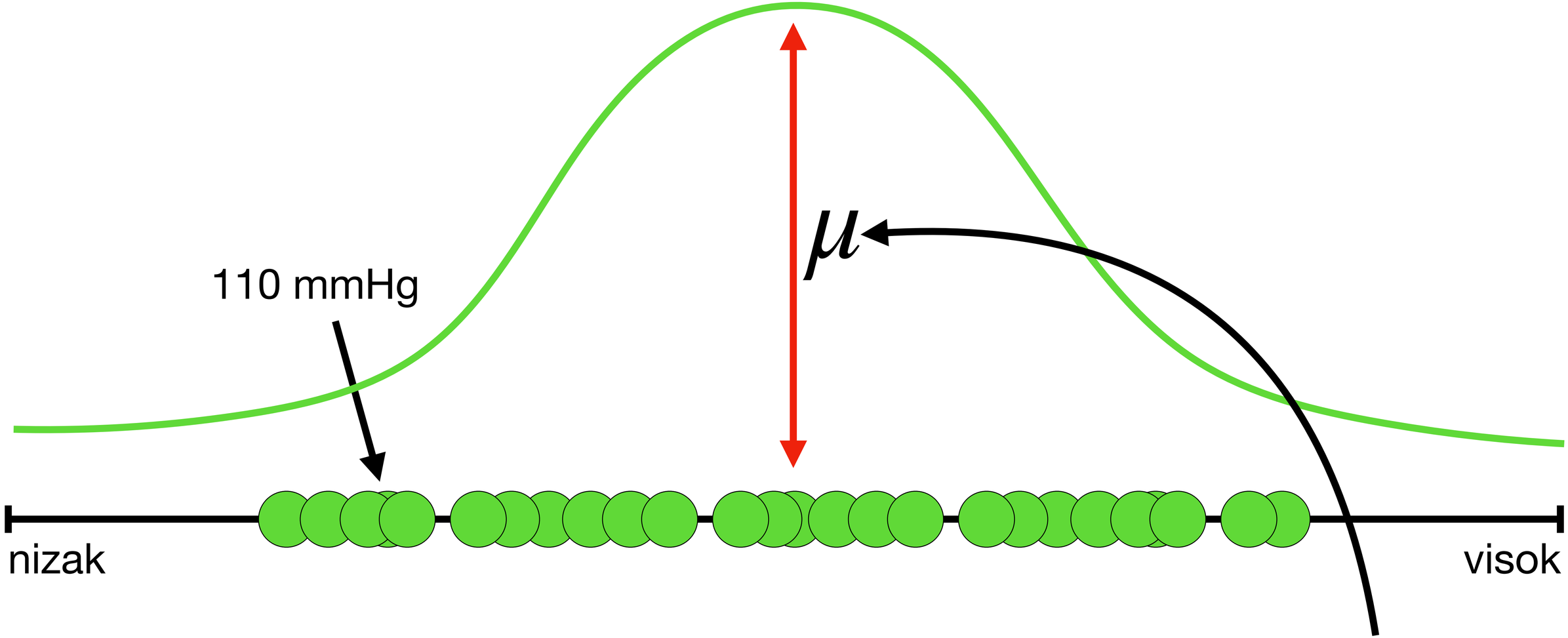
Nepoznati Procene

Iz baze “Sistolni pritisak.xlsx” odrediti prost i sistematski uzorak veličine 10.

- sample(1:99, 10, replace = F)
- sample(1:99, 1, replace = F)
  - Odrediti početnu jedinicu
  - Korak k = 99/10 = 9.9
  - Svaka izabrati svaku ~10 jedicinu

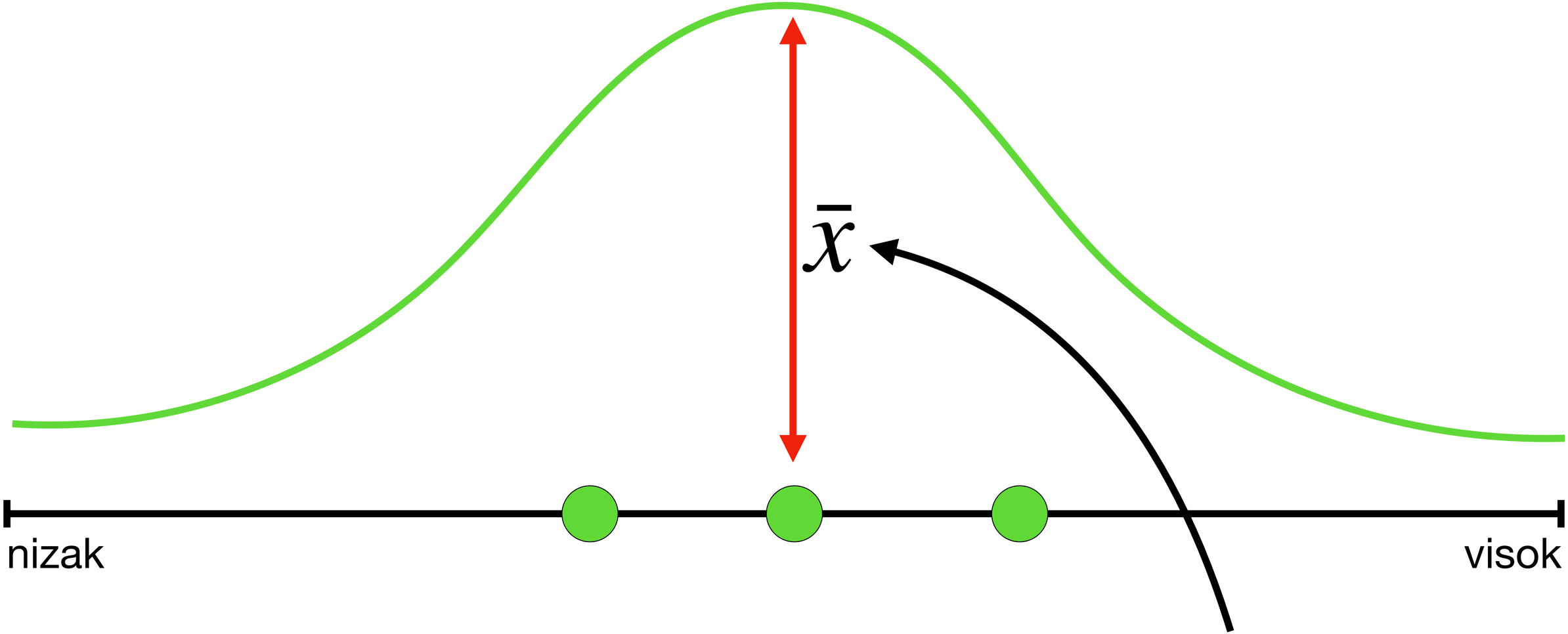
Uzoračka raspodela

Populaciona distribucija pritisaka

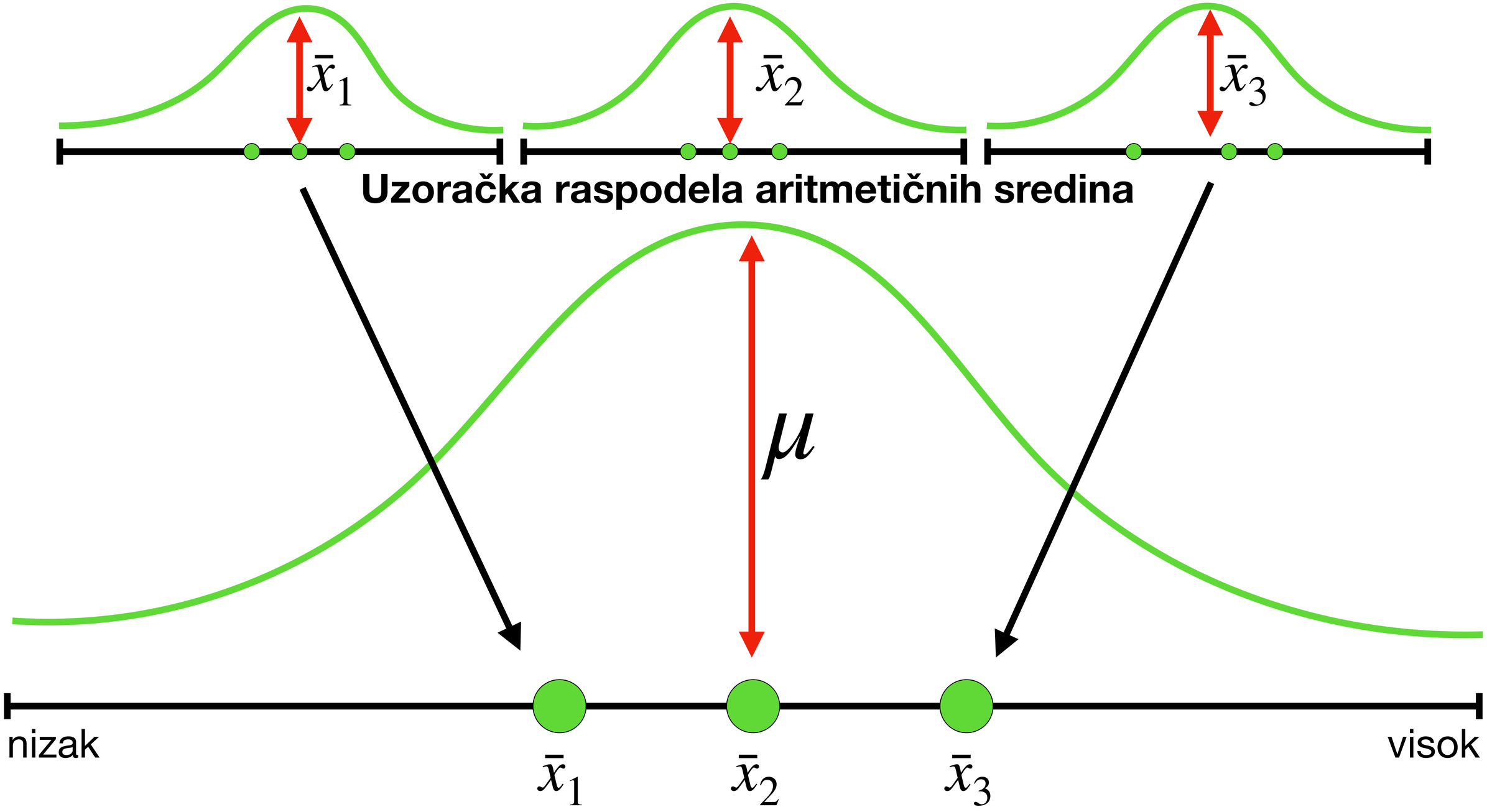


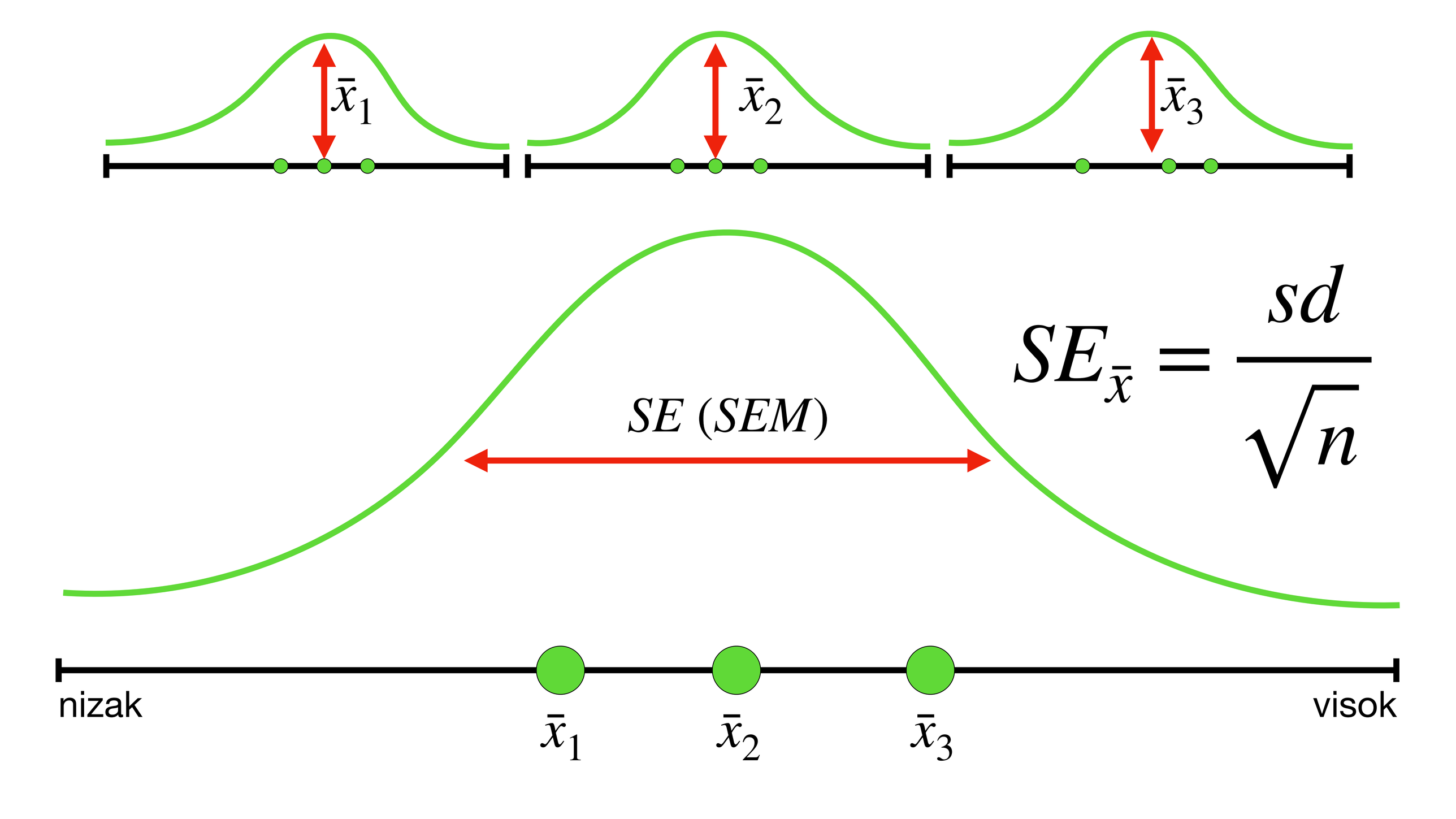
$$\text{Stvarni prosečni pritisak} = \frac{110 + 115 + 130 + 123 + \ldots + 145}{604800} = 120\,\text{mmHg}$$

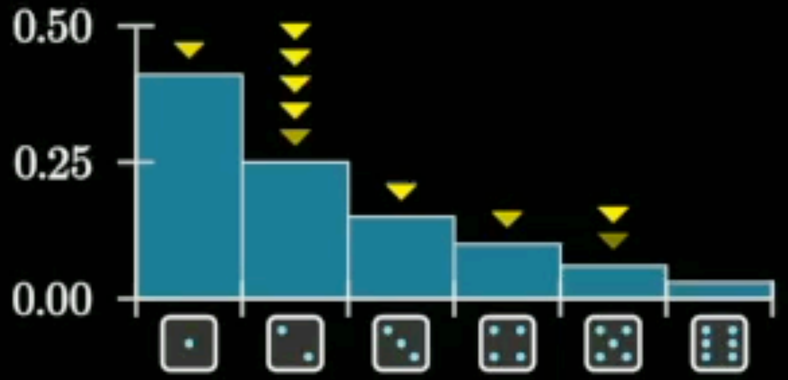
Raspodela pritisaka u jednom uzorku

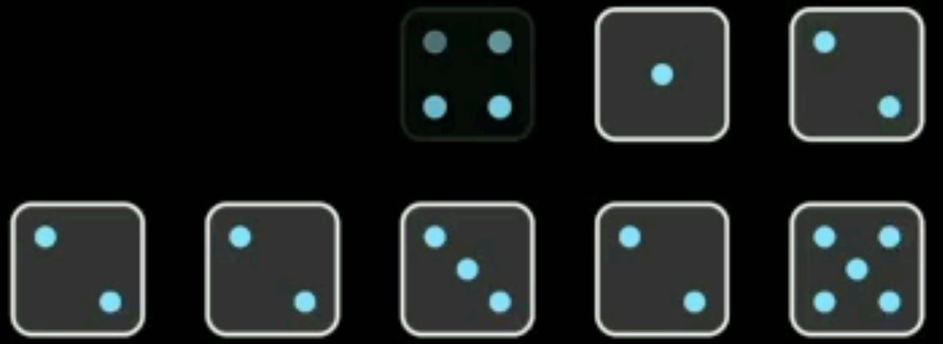


$$\text{Procenjen prosečan pritisak} = \frac{110 + 115 + 130}{3} = 118.3\,\text{mmHg}$$


$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
Uzoračka raspodela aritmetičnih sredina
μ
nizak
visok
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$


$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
$SE\ (SEM)$
$SE_{\bar{x}} = \frac{sd}{\sqrt{n}}$
nizak
visok
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$





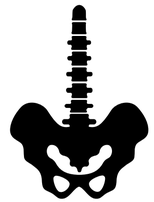
Sum = 28

Sums = 4

28

Istraživačko pitanje

“Da li je stepen degeneracije lumbosakruma na radiografiji povezan sa bolom?”



Ciljna grupa

Osnovni skup

Pacijenti sa lumboishialgijom

Uzorak

n pacijenata

Jedinica posmatranja

Pacijent

Varijable

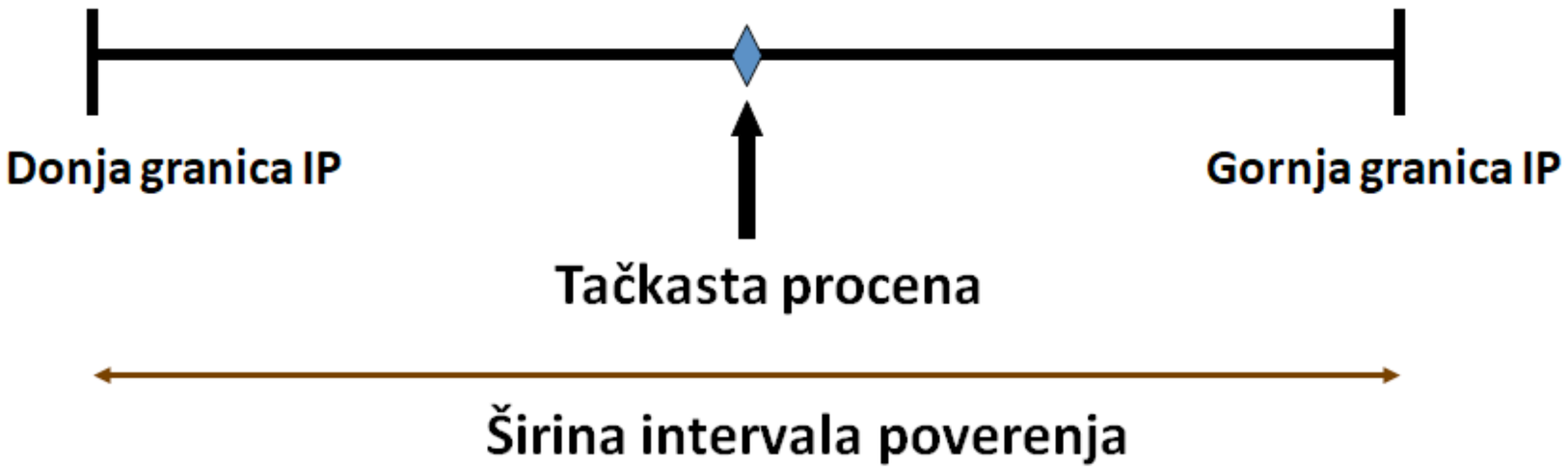
Stepen degeneracije, bol

Kako da saznam
nešto o grupi pacijenata
ako ih nisam video?

Ocenjivanje populacionih parametara na osnovu uzorka

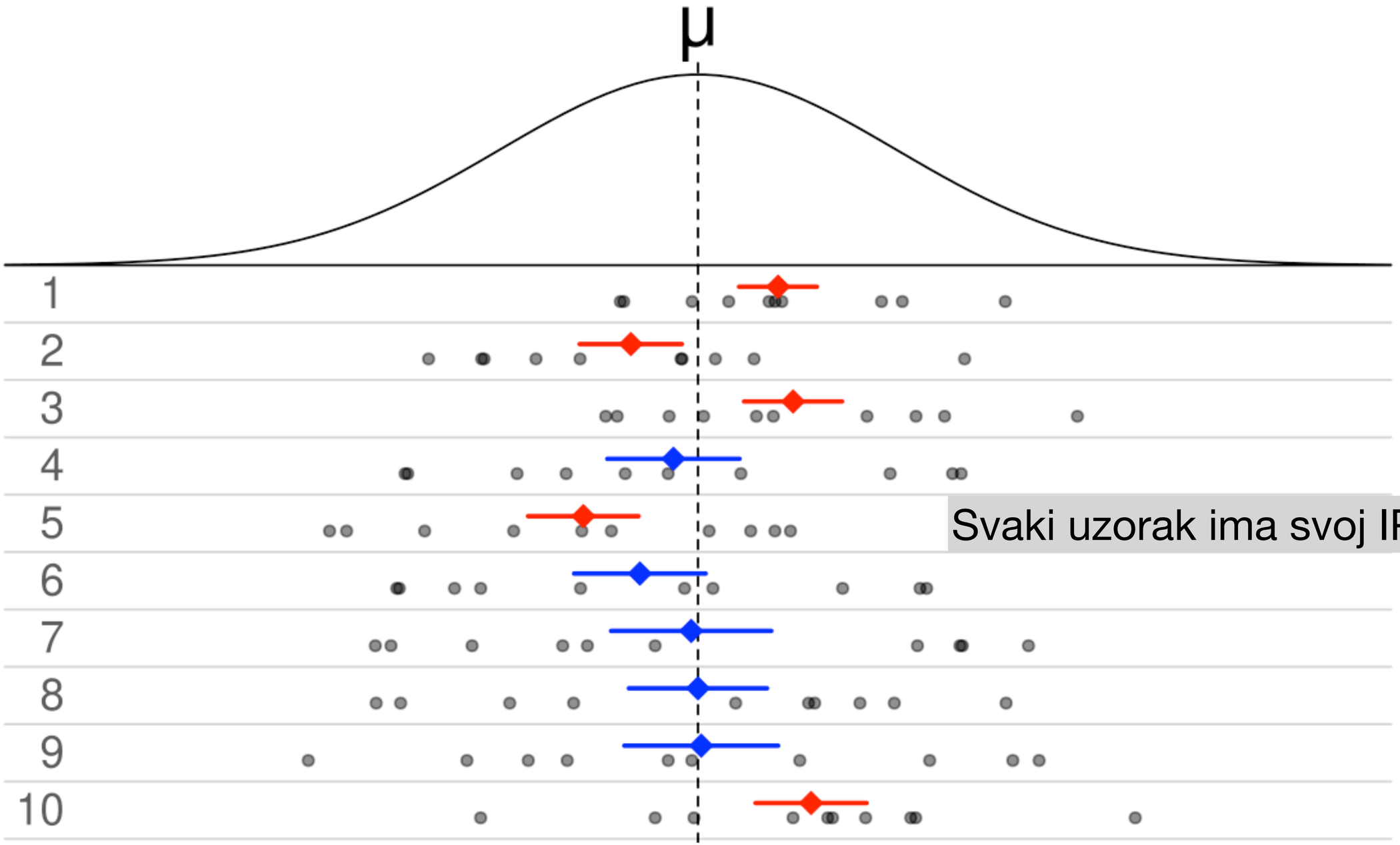
$$CI = \bar{x} \pm t \times SE$$

$$SE = \frac{sd}{\sqrt{n}}$$



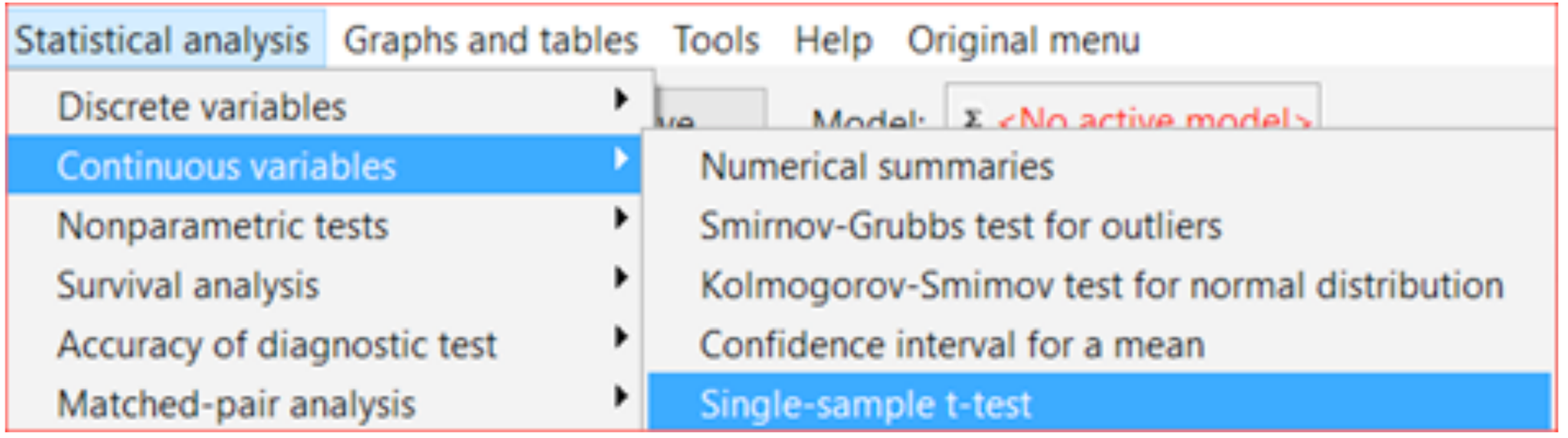
Od svih 95% intervala poverenja, 95% njih sadrži pravu populacionu vrednost.

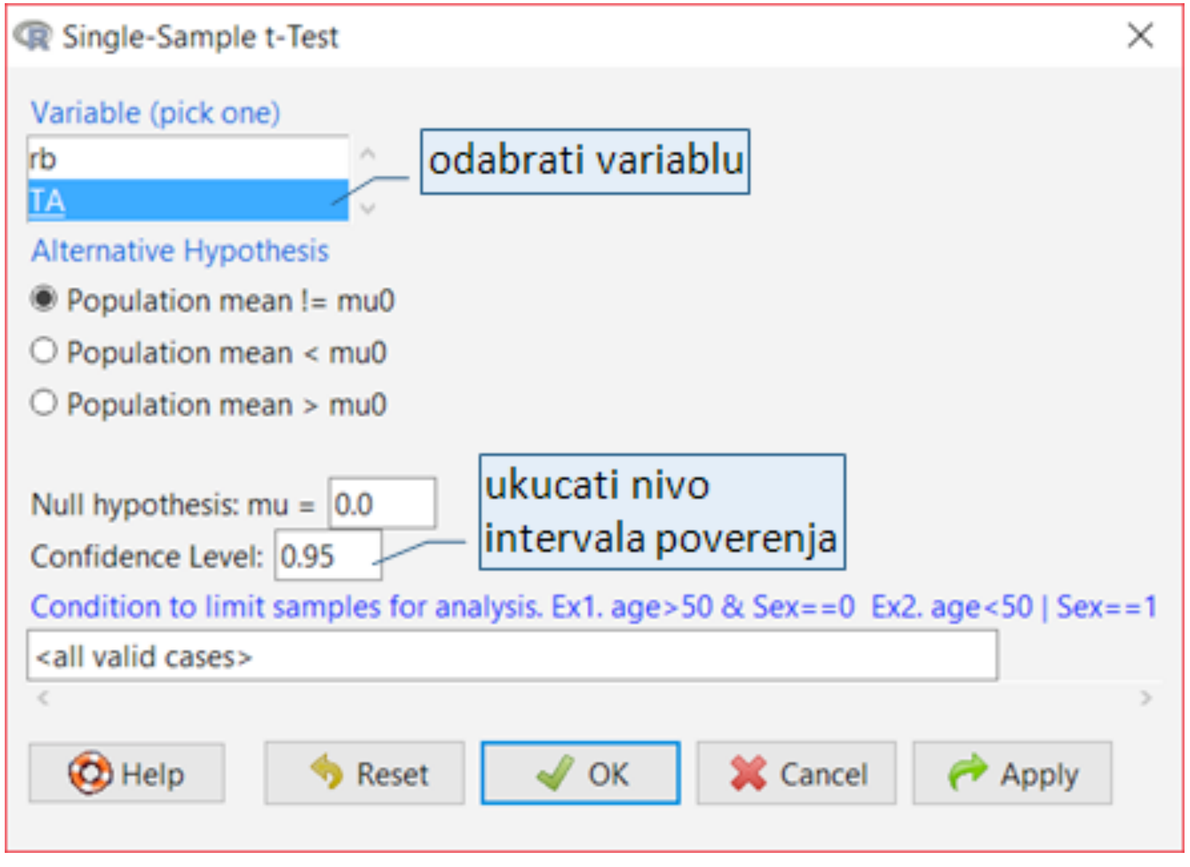
99% interval je sigurniji (ima manji rizik greške) ali je širi od 95%.



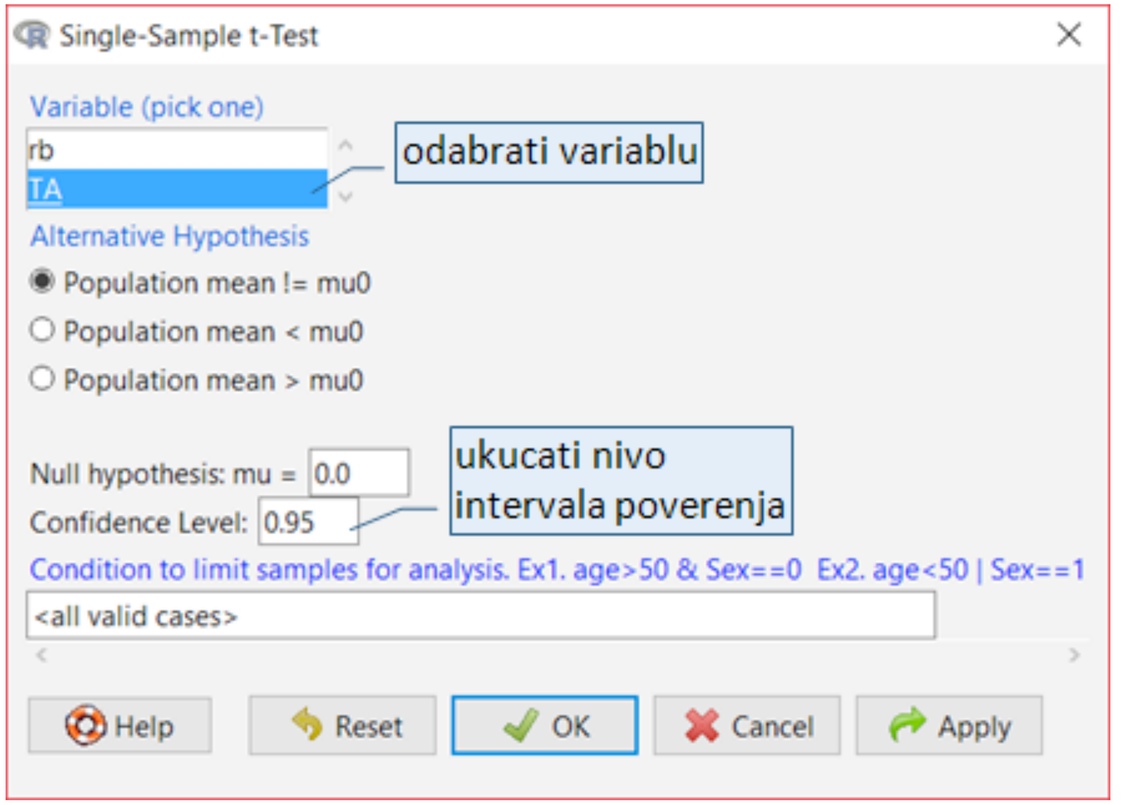
Svaki uzorak ima svoj IP.

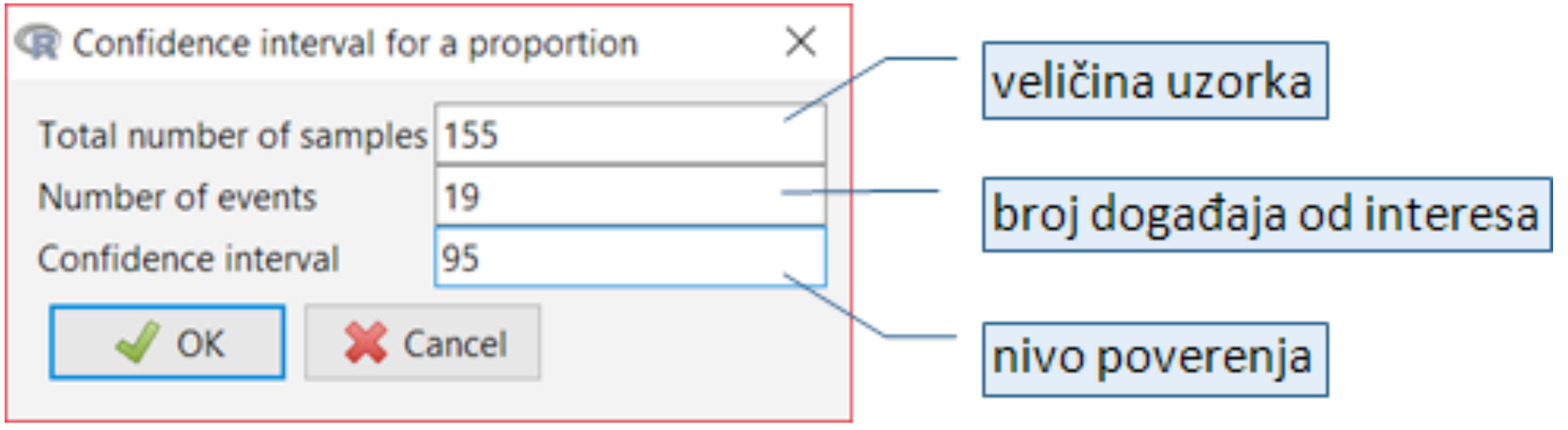
Koristeći bazu podataka Sistolna TA.xlsx odrediti 99% interval poverenja aritmetičke sredine sistolnog arterijskog pritiska ispitanika sa akutnim koronarnim sindromom.


Statistical analysis
Graphs and tables
Tools
Help
Original menu
Discrete variables
Continuous variables
Nonparametric tests
Survival analysis
Accuracy of diagnostic test
Matched-pair analysis
Numerical summaries
Smirnov-Grubbs test for outliers
Kolmogorov-Smimov test for normal distribution
Confidence interval for a mean
Single-sample t-test


Single-Sample t-Test
Variable (pick one)
rb
TA
odabrati variablu
Alternative Hypothesis
Population mean != mu0
Population mean < mu0
Population mean > mu0
Null hypothesis: mu = 0.0
Confidence Level: 0.95
ukucati nivo intervala poverenja
Condition to limit samples for analysis. Ex1. age>50 & Sex==0 Ex2. age<50 | Sex==1
<all valid cases>
Help
Reset
OK
Cancel
Apply

Na uzorku od 500 učenika u jednoj opštini nađena je anemija kod 25 učenika. Odrediti 95% interval poverenja proporcije učenika sa anemijom u toj opštini.


Single-Sample t-Test
Variable (pick one)
rb
TA
odabrati variablu
Alternative Hypothesis
Population mean != mu0
Population mean < mu0
Population mean > mu0
Null hypothesis: mu = 0.0
Confidence Level: 0.95
ukucati nivo intervala poverenja
Condition to limit samples for analysis. Ex1. age>50 & Sex==0 Ex2. age<50 | Sex==1
<all valid cases>
Help
Reset
OK
Cancel
Apply


Confidence interval for a proportion
Total number of samples
155
Number of events
19
Confidence interval
95
OK
Cancel
veličina uzorka
broj događaja od interesa
nivo poverenja