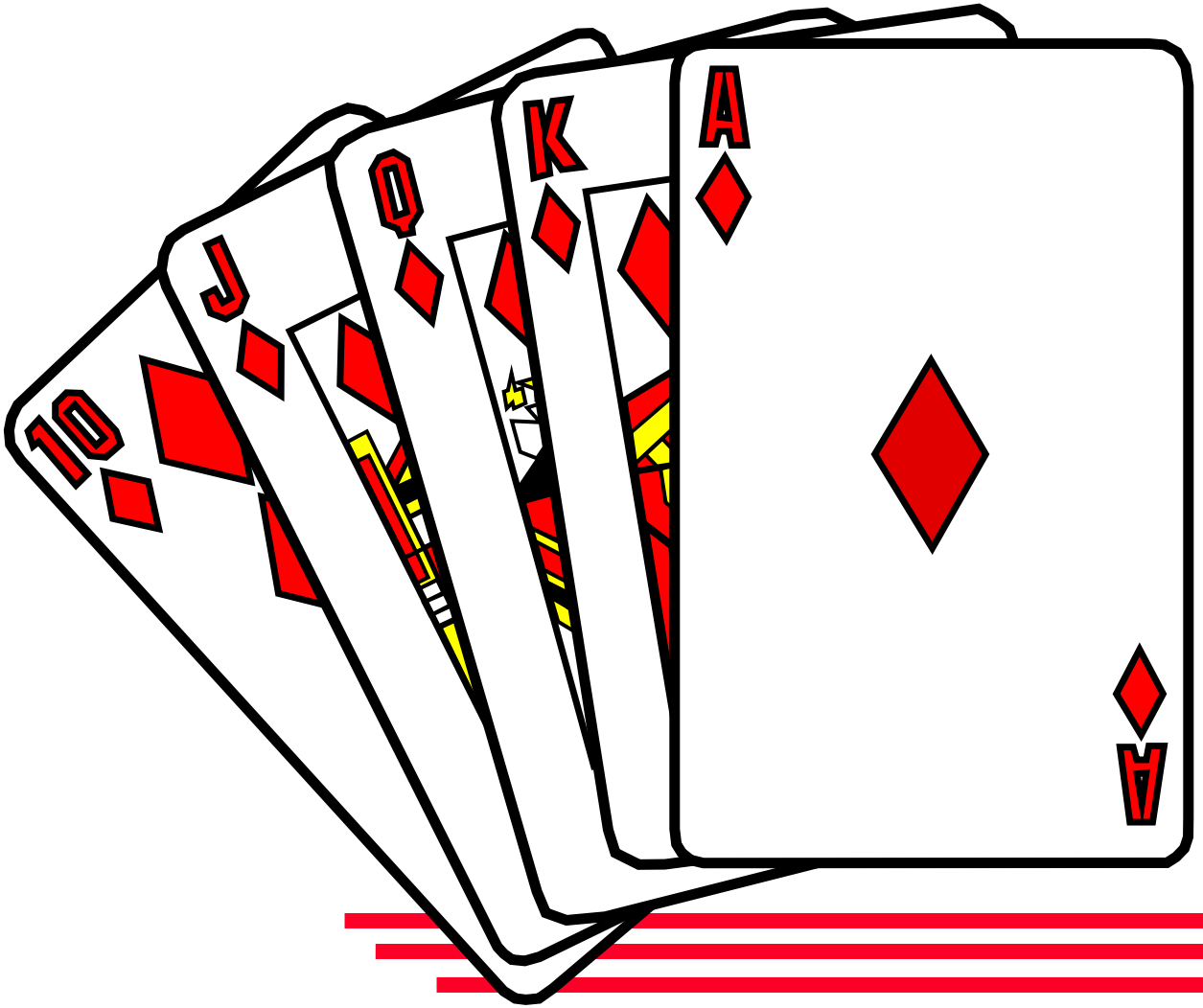
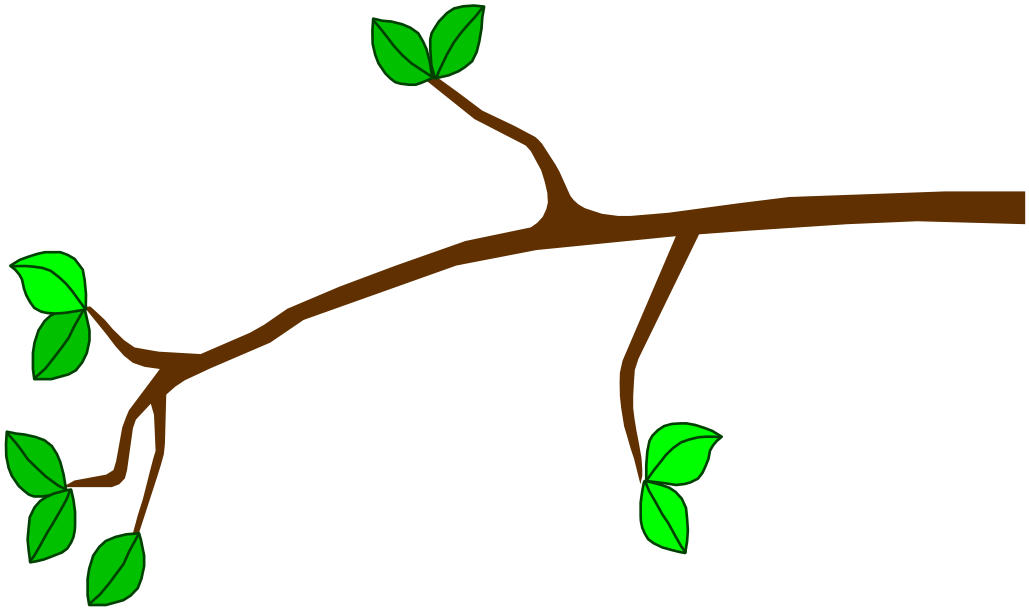
Vzorčenje





Analitična statistika je veja statistike, ki se ukvarja z uporabo vzorčnih podatkov, da bi z njimi naredili zaključek (inferenco) o populaciji.

Zakaj vzorčenje?

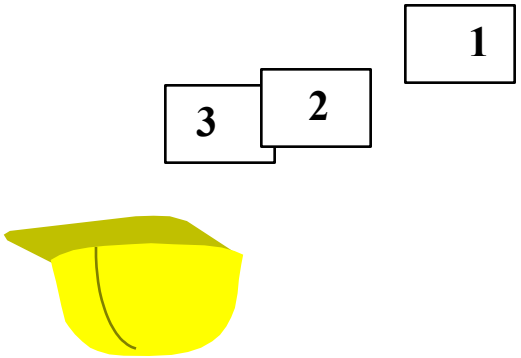
- cena
- čas
- destruktivno testiranje

Načini vzorčenja

- ocena
  - priročnost
- naključno
  - enostavno
  - deljeno
  - grozdno

Enostavno naključno vzorčenje

Pri enostavnem naključnem vzorčenju je vsak član populacije izbran/vključen z enako verjetnostjo.



Del tabele 6 stran 1096

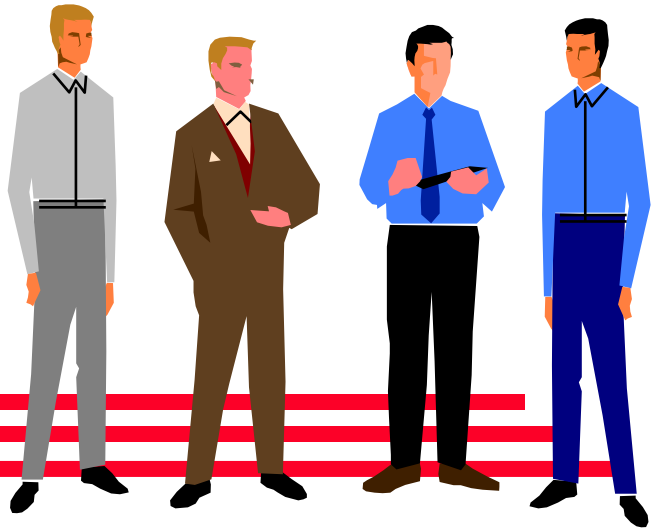
stolpec vrstica	1	2	3	4
1	10480	15011	01536	02011
2	22368	46573	25995	85393
3	24130	48360	22527	97265
4	42167	93093	06243	61680
5	37570	39975	81837	16656
6	77921	06907	11008	42751
7	99562	72905	56240	69994
8	96301	91977	05463	07972
9	89579	14342	63661	10281
10	85475	36857	53342	53988
11	28918	69578	88231	33276

Del tabele 6 stran 1096

stolpec / vrstica	1	2	3	4
1	10480	15011	01536	02011
2	22368	46573	25995	85393
3	24130	48360	22527	97265
4	42167	93093	06243	61680
5	37570	39975	81837	16656
6	77921	06907	11008	42751
7	99562	72905	56240	69994
8	96301	91977	05463	07972
9	89579	14342	63661	10281
10	85475	36857	53342	53988
11	28918	69578	88231	33276

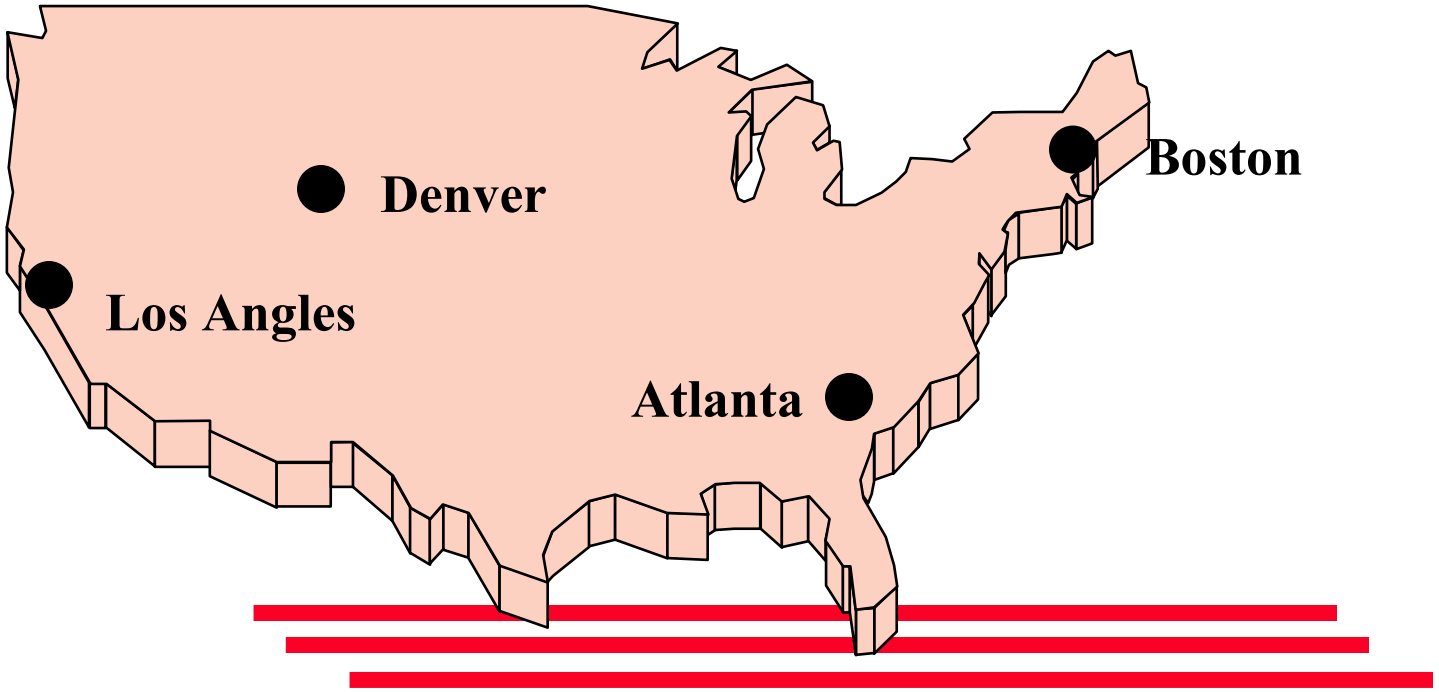
Razdeljen naključni vzorec

Razdeljen naključni vzorec dobimo tako, da razdelimo populacijo na disjunktne množice oziroma dele (razrede) in nato izberemo enostavne naključne vzorce za vsak del posebej.



Grozdno vzorčenje

Grozdno vzorčenje je enostavno naključno vzorčenje skupin ali klastrov/grozdov elementov.



Merjenje centrale tendence

- **modus**
  - **vrednost, ki se ponovi največkrat**
- **mediana**
  - **srednja vrednost**
- **povprečje**
  - **aritmetično povprečje**
- **razpon**
  - **povprečje največje in najmanjše vrednosti**

Merjenje centrale tendence

	populacija	vzorec
povprečje	μ	$\bar{Y}$, $\{\bar{y}_i\}$
mediana	τ	M, m

Populacija

1, 2, 3, 4

$n = 4$

Populacija 1, 2, 3, 4

$$\mu = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$= \frac{1+2+3+4}{4} = 2{,}5$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_i - \mu\right)^2}{n}}$$

$$= \sqrt{\frac{(1-2{,}5)^2 + (2-2{,}5)^2 + (3-2{,}5)^2 + (4-2{,}5)^2}{4}}$$

$$= \sqrt{\frac{(-1{,}5)^2 + (-0{,}5)^2 + (0{,}5)^2 + (1{,}5)^2}{4}}$$

$$= \sqrt{\frac{5}{4}} = \sqrt{1{,}25} = 1{,}118$$

Populacija 1, 2, 3, 4

$$\mu = 2{,}5$$

$$\sigma = 1{,}118$$

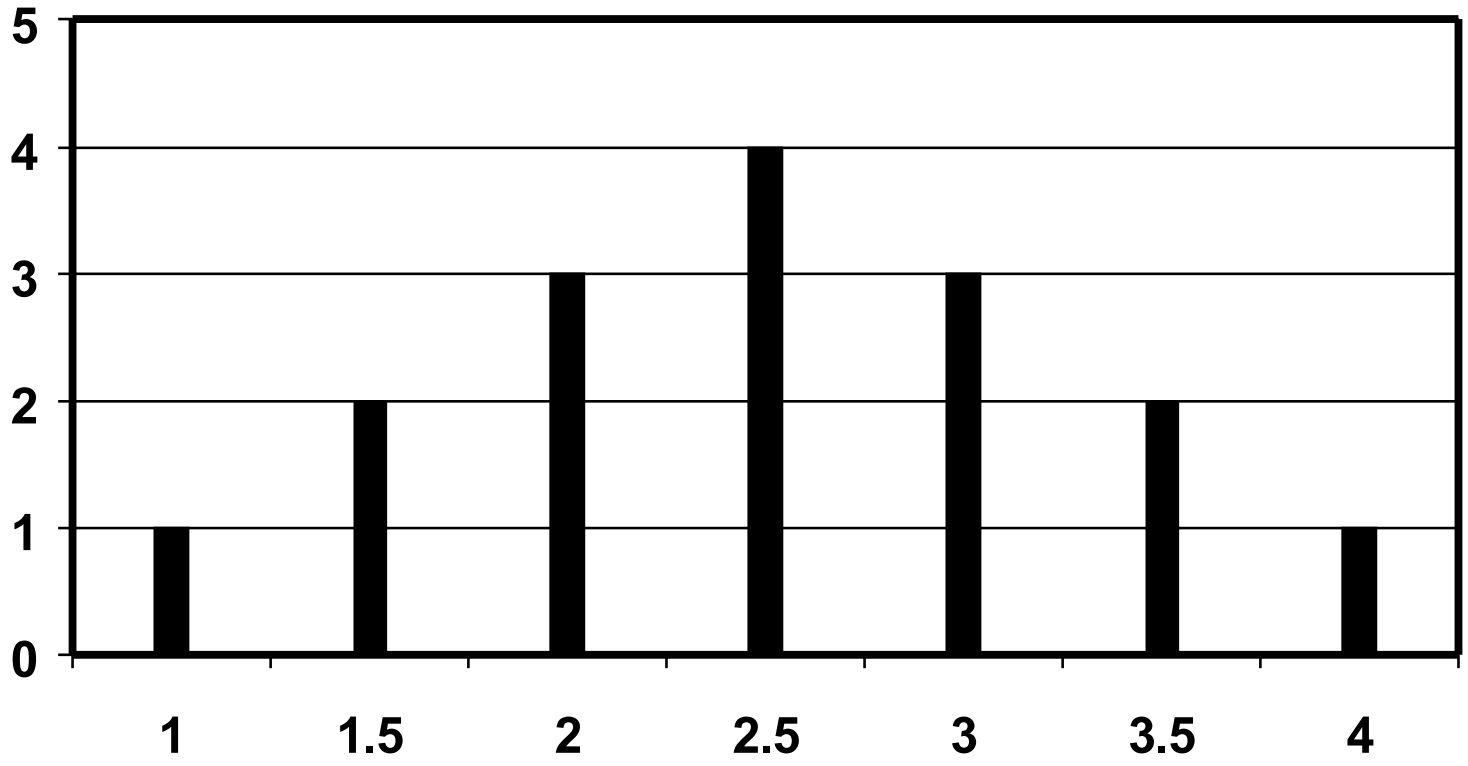
$n = 4$ **(velikost populacije)**

Vsi možni vzorci velikosti 2

1,1	2,1	3,1	4,1
1,2	2,2	3,2	4,2
1,3	2,3	3,3	4,3
1,4	2,4	3,4	4,4

vzorec	povprečje vzorca
1,1	1.0
1,2	1.5
1,3	2.0
1,4	2.5
2,1	1.5
2,2	2.0
2,3	2.5
2,4	3.0
3,1	2.0
3,2	2.5
3,3	3.0
3,4	3.5
4,1	2.5
4,2	3.0
4,3	3.5
4,4	4.0

Porazdelitev enostavnih povprečij



Populacija vzorčnih povprečij

Velikost populacije $n = 16$

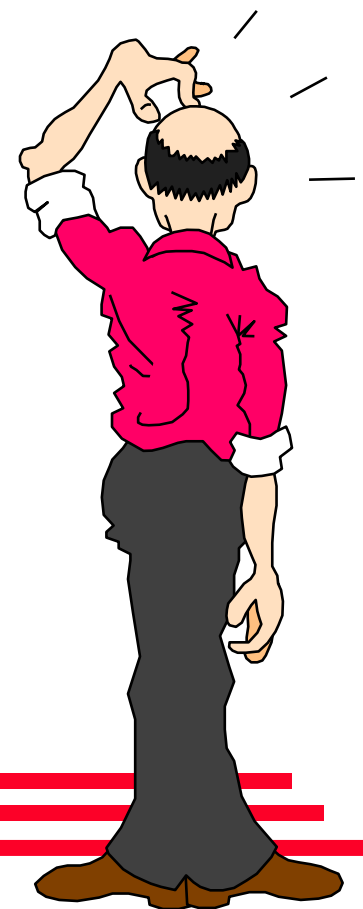
1 1,5 1,5 2 2 2

2,5 2,5 2,5 2,5 3

3 3 3,5 3,5 4

$$\mu_{\bar{y}} = [1 + 1{,}5 + 1{,}5 + 2 + 2 + 2 + 2{,}5 + 2{,}5 + 2{,}5 + 2{,}5 + 3 + 3 + 3 + 3{,}5 + 3{,}5 + 4]/16$$

¿ $40/16 = 2{,}5 = \mu$



$$\sigma_{\bar{Y}} = \sqrt{\frac{\sum_{i=1}^{n} \left(\bar{Y}_i - \mu_{\bar{Y}}\right)^2}{n}}$$

$$\left[(1-2{,}5)^2+(1{,}5-2{,}5)^2+(1{,}5-2{,}5)^2\right.$$

¿

¿

$$(2\text{-}2{,}5)^2+(2\text{-}2{,}5)^2+(2\text{-}2{,}5)^2$$

¿

$$(2{,}5\text{-}2{,}5)^2+(2{,}5\text{-}2{,}5)^2+$$

$$=\left[\text{¿}\quad(2{,}5\text{-}2{,}5)^2+(2{,}5\text{-}2{,}5)^2+\text{¿}\quad(3\text{-}2{,}5)^2+(3\text{-}2{,}5)^2+(3\text{-}2{,}5)^2+\text{¿}\quad(3{,}5\text{-}2{,}5)^2+(3{,}5\text{-}2{,}5)^2+\text{¿}\quad(4\text{-}2{,}5)^2\right]/16\Big]^{1/2}=\sqrt{\frac{10}{16}}=0{,}79$$

¿

populacija vzorčna povprečja

$\mu = 2{,}5$ $\mu_{\bar{y}} = 2{,}5 = \mu$

$\sigma = 1{,}118$ $\sigma_{\bar{y}} = 0{,}79$

$n = 4$ **(velikost populacije)**

$n = 2$ **(velikost vzorca)**

Vzorčna porazdelitev povprečja

Centralni limitni izrek

Če je naključni vzorec velikosti *n* izbran iz populacije s končnim povprečjem μ in varianco σ^2, potem je lahko, če je *n* dovolj velik, vzorčna porazdelitev povprečja $\overline{y}$ aproksimirana z gostoto normalne porazdelitve.

Vzorčna porazdelitev povprečja

Naj bo $y_1, y_2, \ldots, y_n$ **naključni vzorec, ki je sestavljen iz** n **meritev populacije s končnim povprečjem** μ **in končnim standardnim odklonom** σ**.**

Potem sta povprečje in standardni odklon vzorčne porazdelitve $\overline{y}$ **enaka**

$$\mu_{\overline{y}} = \mu \quad \text{in} \quad \sigma_{\overline{y}} = \frac{\sigma}{\sqrt{n}}.$$