

# OPISNA STATISTIKA

---

---

---

---

---

---

---

---

## Cilji vaje

Načini grafičnega prikaza podatkov:

- prikaz s stolpci, krogi, trakovi, ..
- histogram, frekvenčni poligon, kvantilni diagram - boxplot, histogram s številkami )
- razsevni diagram

Načini numeričnega prikaza podatkov:

- Mere za srednje vrednosti
- Mere za razpršenost podatkov

---

---

---

---

---

---

---

---

## Spremenljivke v statistiki

**A**

<b>Atributivne</b> (opisne, kategorične)	<b>Numerične</b> (številске)
<b>Nominalne</b> (npr. krvne skupine)	<b>Zvezne</b> (npr. telesna višina)
<b>Ordinalne</b> ("koliko")	<b>Nezvezne</b> (npr. število bolnišničnih dni)

**B**

**Dihotomne** (dva možna izida, npr. živ-mrtev)

**Politomne** (več možnih izidov)

---

---

---

---

---

---

---

---

Prikaz nominalne spremenljivke

**FREKVENČNA TABELA**

genotip	f	f (%)
*1/*1	102	
*1/*2	40	
*1/*3	33	
*2/*2	3	
*2/*3	5	
*3/*3	5	

---

---

---

---

---

---

---

---

Prikaz nominalne spremenljivke

**FREKVENČNA TABELA**

genotip	f	f%
*1/*1	102	
*1/*2	40	
*1/*3	33	
*2/*2	3	
*2/*3	5	
*3/*3	5	
<b>SKUPAJ</b>	<b>188</b>	

---

---

---

---

---

---

---

---

Prikaz nominalne spremenljivke

**FREKVENČNA TABELA**

genotip	f	f%
*1/*1	102	54,25
*1/*2	40	
*1/*3	33	
*2/*2	3	
*2/*3	5	
*3/*3	5	
<b>SKUPAJ</b>	<b>188</b>	

---

---

---

---

---

---

---

---

## Prikaz nominalne spremenljivke

### FREKVENČNA TABELA

genotip	f	f%
*1/*1	102	54,25
*1/*2	40	21,28
*1/*3	33	17,55
*2/*2	3	1,60
*2/*3	5	2,66
*3/*3	5	2,66
SKUPAJ	188	100

---

---

---

---

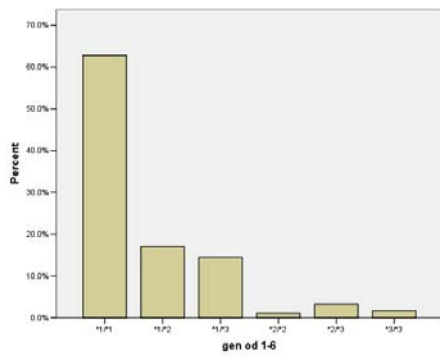
---

---

---

---

### Narišite stolpčni diagram v MS Excelu!



---

---

---

---

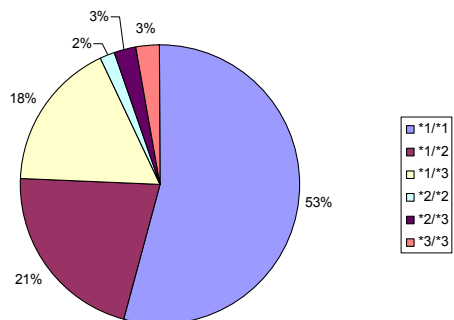
---

---

---

---

### Podatke prikažite s krožnim diagramom!



---

---

---

---

---

---

---

---

### Prikaz ordinalne spremenljivke

	f	f%	F%
<i>nezadosten</i>	3		
<i>zadosten</i>	11		
<i>dober</i>	38		
<i>prav dober</i>	17		
<i>odličen</i>	7		
<b>SKUPAJ</b>	76		

---

---

---

---

---

---

---

---

### Prikaz ordinalne spremenljivke

	f	f%	F%
<i>nezadosten</i>	3	3,9	
<i>zadosten</i>	11	14,5	
<i>dober</i>	38	50,0	
<i>prav dober</i>	17	22,4	
<i>odličen</i>	7	9,2	
<b>SKUPAJ</b>	76	100,0	

Kumulativne frekvence

---

---

---

---

---

---

---

---

### Prikaz ordinalne spremenljivke

	f	f%	F%
<i>nezadosten</i>	3	3,9	3,9
<i>zadosten</i>	11	14,5	18,4
<i>dober</i>	38	50,0	
<i>prav dober</i>	17	22,4	
<i>odličen</i>	7	9,2	
<b>SKUPAJ</b>	76	100,0	

---

---

---

---

---

---

---

---

## Prikaz ordinalne spremenljivke

	f	f%	F%
<i>nezadosten</i>	3	3,9	3,9
<i>zadosten</i>	11	14,5	18,4
<i>dober</i>	38	50,0	68,4
<i>prav dober</i>	17	22,4	90,8
<i>odličen</i>	7	9,2	100,0
<b>SKUPAJ</b>	<b>76</b>	<b>100,0</b>	<b>100,0</b>

---

---

---

---

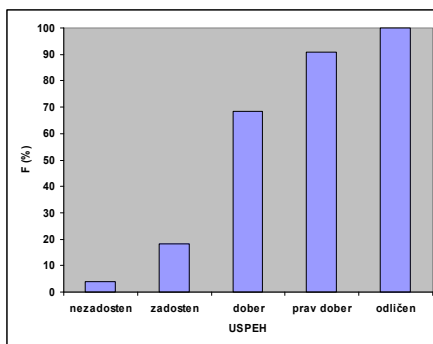
---

---

---

---

## Stolpčni diagram kumulativnih frekvenc Prikažite kumulativne frekvence s stolpci



---

---

---

---

---

---

---

---

## Urejevanje in prikaz numeričnih spremenljivk

- Ranžirna vrsta
- Določanje ranga
- Kvantili: mediana, kvartili, decili, centili

---

---

---

---

---

---

---

---

## Podatki za vajo 3

Podatki iz študije, v kateri smo proučevali vpliv različnih dejavnikov na višino odmerka varfarina (peroralni antikoagulant).

<http://www.ffa.uni-lj.si/o-fakulteti/katedre/katedra-za-socialno-farmacijo/pedagoska-dejavnost/farmaceutvska-informatika/vaje.html>

3. vaja: Podatki Excel.

Radi bi opisali plazemsko koncentracijo varfarina.

---

---

---

---

---

---

---

---

## Naloga

### 1. Naredite ranžirno vrsto koncentracij varfarina.

Razvrstite po vrsti, pripišite rang, enake vrednosti?

### 2. Izračunajte mediano:

Funkcija MEDIAN,

Liho število enot (n): m-ta največja vrednost:  $m = (n+1)/2$

Sodo število enot (n): povprečje med  $m_1$ -to in  $m_2$ -to vrednostjo

$m_1 = n/2$ ,  $m_2 = n/2 + 1$

### 3. Izračunajte prvi in tretji kvartil:

Funkcija PERCENTILE ali QUARTILE

p-percentil (1-100); 50.percentil = 2.kvartil = mediana

n-p/100 ni celo število → (k + 1)-ta največja vrednost:

k je navzdol zaokrožen n-p/100

n-p/100 je celo število → povprečje med  $m_1$ -to in  $m_2$ -to vrednostjo:

$m_1 = n \cdot p/100$ ,  $m_2 = n \cdot p/100 + 1$

---

---

---

---

---

---

---

---

vrednost	k	n=188		
MEDIAN	2.025	0.5	2.025	m1=94; m2=95
1.kvartil	1.7400	0.25	1.74	np/100=47 m1=47; m2=48
3.kvartil	2.2625	0.75	2.265	np/100=141 m1=141; m2=142
1.decil	1.2940	0.1	1.28	np/100=18.8 19
9.decil	2.5130	0.9	2.52	np/100=169.2 170
MIN	0.53			
MAX	2.98			
average	1.97			
median2	2.00	0.5		n=185 brez 3. največjih m=93

---

---

---

---

---

---

---

---

## 5 različnih načinov računanja percentilov (SPSS)

w is the sum of the weights, p is the specified percentile divided by 100, and  $X_i$  is the value of the  $i$ th case (cases are assumed to be ranked in ascending order).

HAVERAGE	Weighted average at $X(w+1)p$ . The percentile value is the weighted average of $X_i$ and $X_{i+1}$ , where $i$ is the integer part of $(w+1)p$ . This is the default if PERCENTILES is specified without a keyword.
WAVERAGE	Weighted average at $Xwp$ . The percentile value is the weighted average of $X_i$ and $X_{i+1}$ , where $i$ is the integer portion of $wp$ .
ROUND	Observation closest to $wp$ . The percentile value is $X_i$ or $X_{i+1}$ , depending upon whether $i$ or $i+1$ is "closer" to $wp$ .
EMPIRICAL	Empirical distribution function. The percentile value is $X_i$ , where $i$ is equal to $wp$ rounded up to the next integer.
AEMPIRICAL	Empirical distribution with averaging. This is equivalent to EMPIRICAL, except when $i=wp$ , in which case the percentile value is the average of $X_i$ and $X_{i+1}$ .

---

---

---

---

---

---

---

---

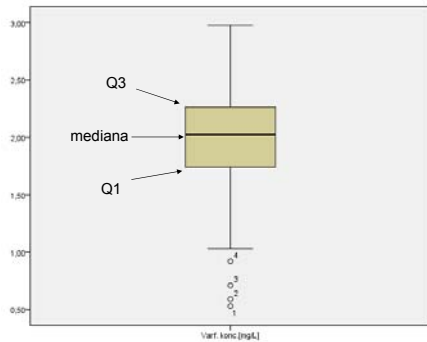
---

---

---

---

## Kvantilni diagram (box plot) – doc.file




---

---

---

---

---

---

---

---

---

---

---

---

## Histogram s številkami (Stem and leaf plot)

Varf. konc. [mg/L] Stem-and-Leaf Plot

Frequency	Stem & Leaf
4,00	Extremes (<.,92)
2,00	10 . 33
7,00	11 . 0226677
6,00	12 . 145578
5,00	13 . 01334
3,00	14 . 399
6,00	15 . 115579
9,00	16 . 133467899
15,00	17 . 002344555777899
14,00	18 . 00013344467799
20,00	19 . 00111222333445566679
18,00	20 . 00234456677789999
20,00	21 . 00011113334445678889
19,00	22 . 011123455667788999
9,00	23 . 345666899
9,00	24 . 001145788
10,00	25 . 0112334888
4,00	26 . 0148
4,00	27 . 0236
2,00	28 . 19
2,00	29 . 38

Stem width: .10  
Each leaf: 1 case(s)

---

---

---

---

---

---

---

---

---

---

---

---

### Izdelava frekvenčne distribucije

- Določanje širine oz. števila razredov

$$(m-1)j < (x_{\max} - x_{\min}) < m*j$$

število razredov = m, širina razreda = j

- Histogram koncentracije varfarina

---

---

---

---

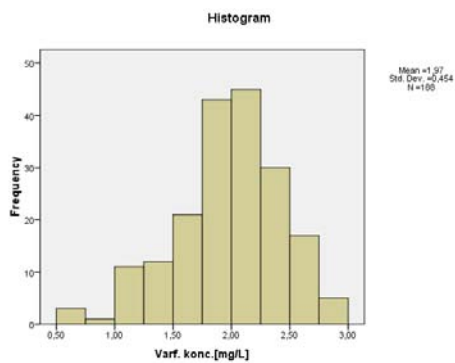
---

---

---

---

### Grafični prikaz frekvenčne porazdelitve



---

---

---

---

---

---

---

---

### NALOGA SPSS:

- Narišite histogram odmerka in logaritma odmerka varfarina.
- Kakšna spremenljivka je odmerek?
- Zakaj smo spremenljivko logaritmirali?

---

---

---

---

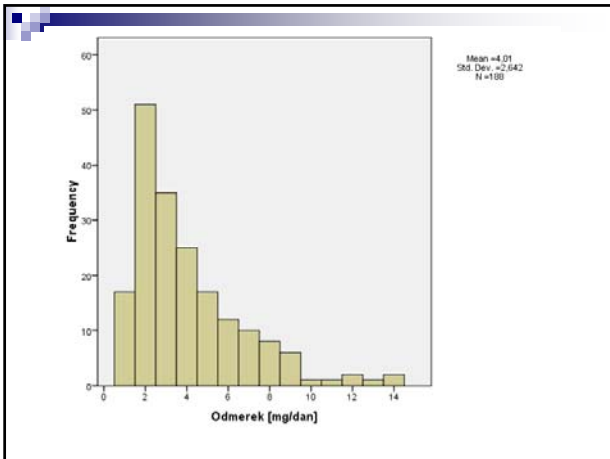
---

---

---

---






---

---

---

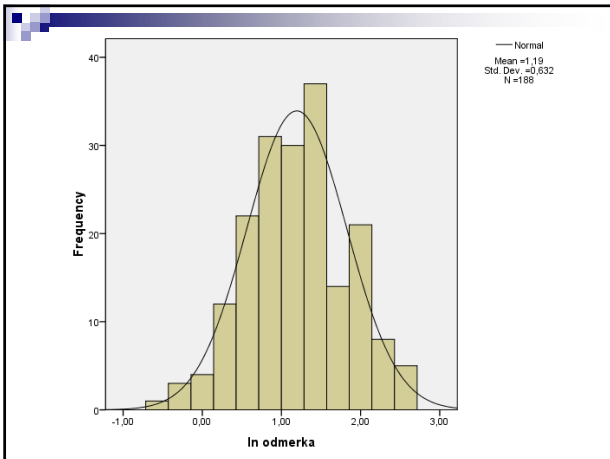
---

---

---

---

---




---

---

---

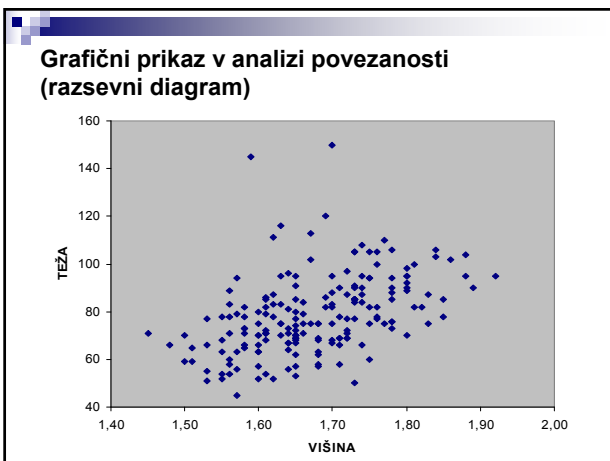
---

---

---

---

---




---

---

---

---

---

---

---

---

# Opisovanje vzorca/populacije

## 1. SREDNJE VREDNOSTI:

- Modus (najpogostejša vrednost)
- Mediana (vrednost na sredini po rangiranih vrednosti)

- Aritmetična sredina 
$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

- Tehtana aritmetična sredina 
$$M = \frac{\sum_{i=1}^k x_i N_i}{\sum_{i=1}^k N_i}$$

---

---

---

---

---

---

---

---

## 2. RAZPRŠENOST ENOT

- Variacijski razmik:  $(x_{\max} - x_{\min})$
- Decilni razmik:  $D_9 - D_1$
- Kvartilni razmik:  $Q_3 - Q_1$
- Povprečni absolutni razmik (upoštevamo vse vrednosti v množici)

$$\text{povpr. abs. odmik} = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

---

---

---

---

---

---

---

---

- Varianca  $\sigma^2$  in standardni odklon  $\sigma$

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{N}}{N}$$

- Koefficient variacije:

$$KV(RSD) = \frac{\sigma}{\mu}$$

- Relativni odklon (odklon posamezne vrednosti od povprečja)

---

---

---

---

---

---

---

---