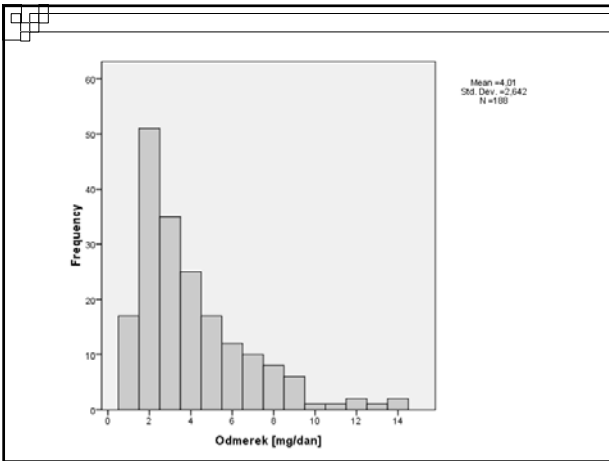
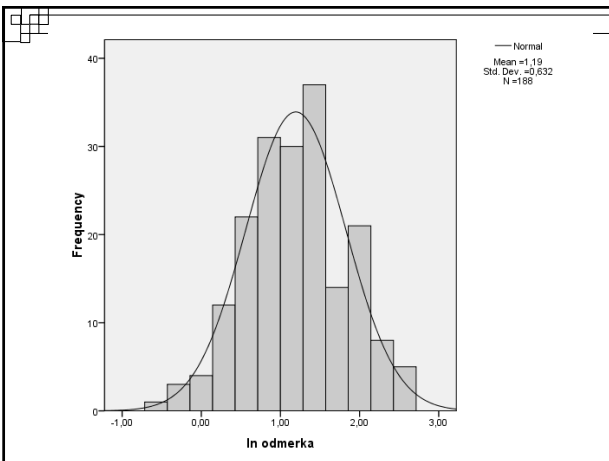


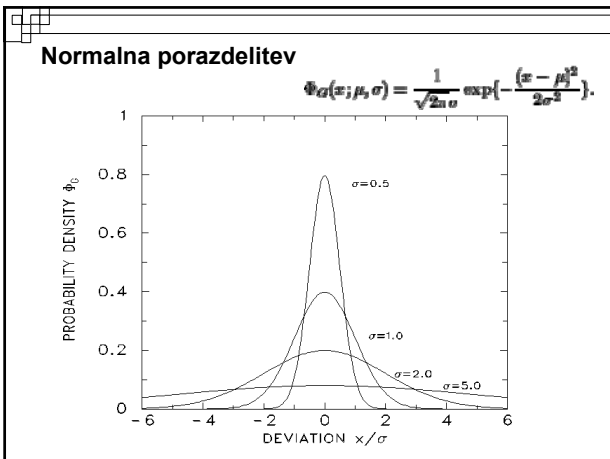


Vaja 4:
Korelacija in linearna
regresija

Igor Locatelli







Standardizacija normalne porazdelitve

- Standardizirana normalna porazdelitev:
 - Na os x nanašamo z-vrednosti
 - Srednja vrednost = 0
 - Standardni odklon = 1

- Tabelirane vrednosti integralov standardizirane krivulje (Tabela normalne porazdelitve).

Nalogi

- Izračunajte delež bolnikov, ki plazemsko koncentracijo varfarina manjšo do 1,5 mg/L oz. delež bolnikov s plazemsko koncentracijo večjo od 2,5 mg/L, če veste, da se vrednosti plazemske koncentracije varfarina porazdeljujejo po normalni porazdelitvi. (MSExcel)

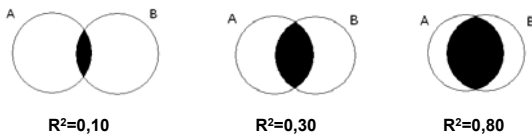
- Izračunajte delež bolnikov, ki imajo dnevni odmerek varfarina več kot 10 mg oz. manj kot 1mg, če veste, da porazdelitev dnevnih odmerkov varfarina sledi log-normalni porazdelitvi, s povprečjem 1.19 in standardno deviacijo 0.63.

Korelacija in linearna regresija

- Korelacija: opis moči povezave med dvema numeričnima spremenljivkama,
- Linearna regresija: linearna povezave odvisna spremenljivke (y) z neodvisno spremenljivko (x): $y=a \cdot x + b$; napoved vrednosti y glede na x, del variabilnosti y spremenljivke pojasnimo z x spremenljivko

Determinacijski koeficient (r^2, R^2):

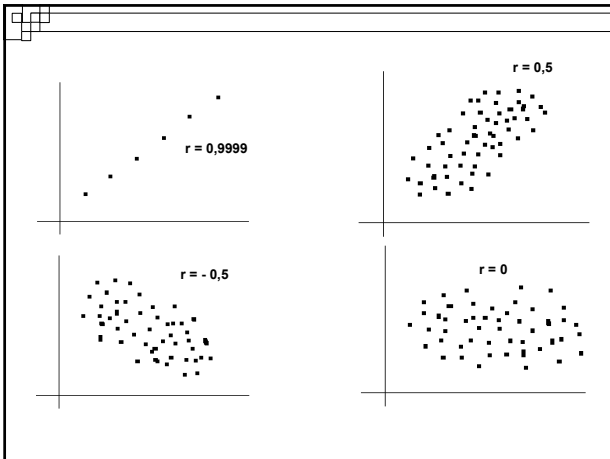
- pove, kolikšen del variance spremenljivke A pojasnimo s spremenljivko B
- meri MOČ povezave, ne pove ničesar o smeri povezave
- Vrednosti [0, 1)

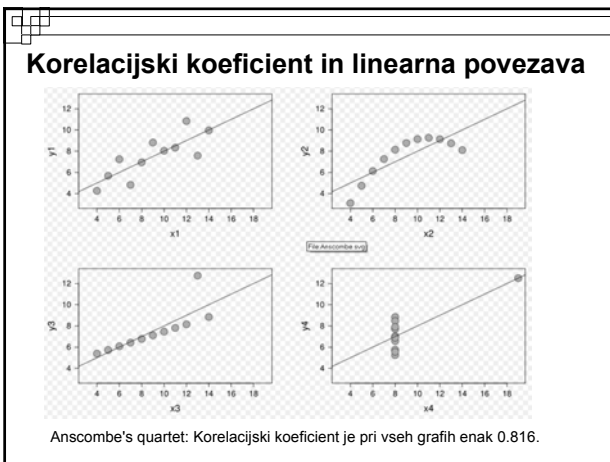


Pearsonov korelacijski koeficient (r):

- merilo za odvisnost med dvema spremenljivkama
- vrednosti (-1, 1)
- izračun: stat. program, formula

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \times \sum (Y_i - \bar{Y})^2}}$$



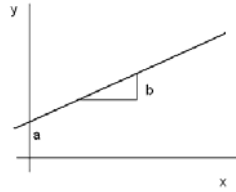


Naloga

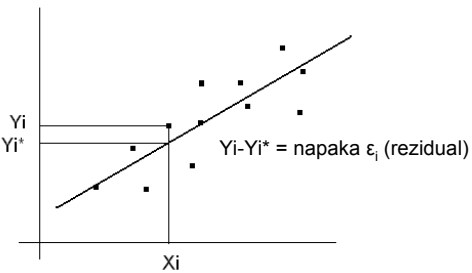
- S pomočjo Excela in SPSS-a izračunajte korelacijo med telesno višino in telesno maso.
- Izračunajte korelacijo med starostjo in telesno maso ter jo grafično predstavite.
- Merjenje korelacije, ko spremenljivke *niso* normalno porazdeljene:
 - Spearman-ova korelacija ali korelacija rangov (vrednosti pretvorimo v range, na katerih izračunamo korelacijo)

LINEARNA REGRESIJA

■ $Y = a + bX$



■ Metoda najmanjše vsote kvadratov (LSM)



LSM: $\min \sum (Y_i - Y_i^*)^2$

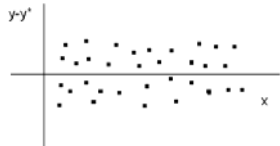
■ Izračun koeficientov linearne regresije:

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$a = \bar{Y} - b\bar{X}$$

Predpostavke linearne regresije

- Linearnost**
- Normalnost porazdelitve** spremenljivk
- Homoscedastičnost**: varianca Y se ne spreminja z vrednostjo X (reziduali niso odvisni od X)



Naloga

Določite linearno povezavo med koncentracijo analita in odzivom analizne metode.

Narišite umeritveno premico v MS Excelu.

Konc (mg/L)	Odziv
0	0
1	5.2
2	10
5	21.2
10	58.7
15	75.3
20	105.1

Naloga

Določite linearno odvisnost odmerka varfarina glede na telesno maso bolnika.

Določite linearno odvisnost odmerka varfarina glede na plazemsko koncentracijo varfarina.

Določite linearno odvisnost odmerka varfarina glede na število polimorfnih alelov CYP2C9.
