

1. del: Linearna regresija in korelacija

- Ugibanje korelacijskega koeficienta na podlagi razsevnega diagrama:
 - Uporabite dejavne spletne strani **Guessing Correlations** (v spletni brskalnik vtipkajte URL <http://www.stat.uiuc.edu/courses/stat100/java/GCApplet/GCAppletFrame.html> ali pa pridite do njih preko domače strani IBMI po poti Povezave→Statistics→Teaching)!
 - Nalogo ponovite trikrat (uporabite gumb New Plots)!
- Vaje z dejavnimi spletnimi stranmi **Illuminations Lessons: Investigating Linear Relationships: The Regression Line and Correlation** (http://illuminations.nctm.org/index_o.aspx?id=135):
 - Sledite povezavi Lesson 2: The Effects of Outliers. Odgovorite na **vprašanja a in b v poročilu!** Pri uporabi interaktivnega delovnega področja (koordinatni sistem z gumbi in okenci za določanje osi, izris regresijske premice in izbris točk) upoštevajte navodila, izpisana v rdeči barvi:
 - ob kazalniku miši so vseskozi izpisane trenutne koordinate;
 - točko dodate s klikom na levi gumb miši;
 - CTRL+klik na obstoječo točko (kadar je izbrana, se pojavi okoli kazalnika krogec) jo izbriše;
 - SHIFT+klik na obstoječo točko in držanje levega gumba miši pritisnjene točko premika.
 - Odgovorite na **vprašanja c, d in e** v poročilu!
- Zaženite program SPSS in odprite datoteko C:\Statistika\Dodipl\RegX.
 - S postopkom Graphs → Scatter... (privzeta možnost Simple → gumb Define) narišite razsevni diagram, pri čemer naj bo **Y** odvisna spremenljivka (na ordinatni osi), **X** pa neodvisna spremenljivka (na abscisni osi)! V diagram vrišite regresijsko premico (z dvoklikom v diagram odprite **Chart Editor**; s klikom na eno od točk označite podatkovni niz; dodajte regresijsko premico s klikom na ikono Add fit line ali preko menija Chart → Add Chart Element → Fit Line at Total → gumb Close)! Geometrijsko ocenite približno vrednost regresijske konstante in regresijskega koeficienta ter presodite, kolikšna se vam zdi razpršenost točk okoli regresijske premice!
 - S postopkom Analyze → Regression → Linear... za iste podatke izvedite linearno regresijsko analizo (v okence Dependent spravite spremenljivko **Y**, v okence Independent(s) pa **X**) ter dobljeni vrednosti regresijske konstante in regresijskega koeficienta primerjajte s približno ocenjenima! Zapišite enačbo regresijske premice ter interpretirajte korelacijski koeficient, regresijski koeficient in test statistične značilnosti regresijskega koeficienta!
- Odprite datoteko PorX in preučite možnost premo sorazmerne napovedi porodne teže otroka na podlagi starosti matere, t.j. povezanost med spremenljivko **STAROST** kot prediktorjem in spremenljivko **PTO** kot kriterijem v linearnem modelu. Uporabite enak postopek (razsevni diagram, približna ocena parametrov, linearna regresija in korelacija) kot pri prejšnji nalogi!

2. del: Preizkušanje povezanosti opisnih spremenljivk

5. Preučujemo povezanost med kajenjem matere med nosečnostjo in nizko porodno težo otroka (glede na mejo 2500g je izračunana spremenljivka **NPTO**: 0=ne, t.j. normalna porodna teža; 1=da, t.j. nizka porodna teža).
- Izdelajte kontingenčno tabelo (postopek Analyze → Descriptive Statistics → Crosstabs...; spremenljivka **kajenje** naj določa vrstice, **NPTO** pa stolpce), ki naj poleg opaženih frekvenc vsebuje tudi pričakovane frekvence (gumb Cells...)! Presodite, ali so podatki ustrezni za uporabo testa hi-kvadrat! V zvezi s tem premislite, kateri dve možnosti imamo na razpolago, če želimo analizirati povezanost dveh opisnih spremenljivk s testom hi-kvadrat, dani podatki pa ne ustrezajo zahtevi o majhnem deležu celic z nizko pričakovano frekvenco!
 - Odprite datoteko 500PorX! Ponovite postopek izdelave kontingenčne tabele s spremenljivkama **KAJ** (0=ne, t.j. mati ni kadila; 1=da, t.j. mati je kadila) in **NPTO**, pri čemer naj bo v celicah izpisana tudi relativna frekvenca kategorij **NPTO** za dano vrednost **KAJ** (t.j. vrstični delež), za vsako celico pa nato ročno izračunajte razliko med opaženo in pričakovano frekvenco! Hkrati izvedite tudi test hi-kvadrat (gumb Statistics... → opcija Chi-square)! Interpretirajte dobljene rezultate!

2. a) Pritisnite *CLEAR*; narišite točke, ki kažejo naraščanje z leve na desno; pritisnite *SHOW GRAPH*.

- Ali se premica dobro prilega točkam? _____
- Ali enačba regresijske premice kaže pozitiven naklon? _____

b) Dodajte točko, ki odstopa od trenda.

- Kaj se zgodi z regresijsko premico? Razložite!

- Zgrabite osamelca in ga vlecite naokoli! Kako se pri tem spreminja regresijska premica?

c) Izdelajte razsevni diagram, ki ima točno določeno vrednost korelacijskega koeficienta:

- $r = 1$
- $r = -1$
- $r = 0$

d) V levem spodnjem kotu koordinatne ravnine narišite oblak točk, ki ne kažejo nikakršnega trenda. Potem narišite eno točko v desnem zgornjem kotu.

- Ali dobljeni razsevni diagram kaže linearno povezanost? _____
- Je vrednost r blizu ena? _____

e) Ali visoka vrednost r nujno pomeni, da so podatki linearni? _____

Ali vrednost r blizu nič vedno pomeni, da podatki niso linearni? _____

3. Približna ocena na podlagi razsevnega diagrama:

$$a \approx \underline{\hspace{2cm}}$$

$$b \approx \underline{\hspace{2cm}}$$

razpršenost točk okoli regresijske premice (obkrožite): majhna zmerna velika

Rezultati linearne regresije:

$$a = \underline{\hspace{2cm}}$$

$$b = \underline{\hspace{2cm}}$$

enačba regresijske premice: _____

interpretacija regresijskega koeficienta: _____

$$r = \underline{\hspace{2cm}}$$

interpretacija korelacijskega koeficienta: _____

$$SE_b = \underline{\hspace{2cm}}$$

$$t = \underline{\hspace{2cm}}$$

$$p = \underline{\hspace{2cm}}$$

interpretacija testa statistične značilnosti regresijskega koeficienta: _____
