Vaja 6:
Sklepna statistika ob uporabi statističnega programa SPSS

asist. dr. Marko Obradović
Katedra za socialno farmacijo
Fakulteta za farmacijo

Načrt dela:

1. Sklepna statistika atributivnih spremenljivk
2. Uporaba SPSS: t-test, neparametrični testi
3. ANOVA
4. Korelacija in linearna regresija

t- test, Wilcoxonov test v SPSS-u

Tabletirki – teže tablet v g:

A: 0.13, 0.15, 0.14, 0.17, 0.19, 0.20, 0.15, 0.18, 0.11, 0.17
B: 0.14, 0.19, 0.21, 0.16, 0.18, 0.17, 0.13, 0.14, 0.12, 0.20

Ali tabletirki izdelujeta tablete enake teže,
če je stopnja tveganja 0,05?

X^2 kvadrat

	Kava brez kofeina	Kava s kofeinom
budni	2	7
zaspali	8	3

ANOVA

- z-, t-test : primerjava 2 skupin
- Primerjava večih skupin hkrati?

ANOVA (Analiza varianc):

- test, s katerim ugotovimo ali med skupinami sploh obstajajo kakšne razlike;
- uporaba večkratnih primerjav namesto "globalnega testa" → večanje napake I. reda
- Razlika med skupinami obstaja → izvedba testov za primerjavo parov, da ugotovimo, kje so te razlike.

Večkratne primerjave

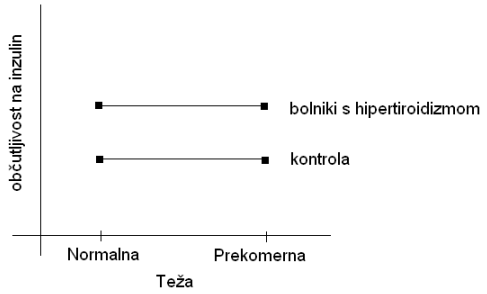
- *Post hoc* testi: kadar pokaže ANOVA signifikantne razlike
  - Bonferonni
  - Tukey-ev HSD pristop
  - Sheffé-jev pristop
  - Newman-Keuls pristop
  - Dunnett-ov pristop
  - LSD

Predpostavke ANOVA-e

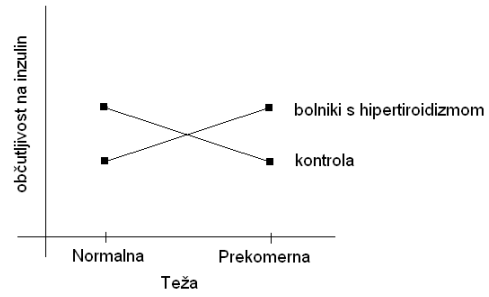
1. Homogenost varianc
   - F-test občutljiv na heterogenost varianc
2. Normalnost porazdelitve odvisne spremenljivke
   - F-test dokaj robusten na odstopanja od normalnosti
3. Neodvisnost opazovanih vrednosti

Dvofaktorska ANOVA

- Primer:
  - Gonzalo in sod. 1996: Ali hipertiroidizem oz. debelost vplivata na občutljivost na inzulin
  - 3 vprašanja:
    1. Ali obstaja razlika med osebami s hipertiroidizmom in brez le-tega?
    2. Ali obstaja razlika med osebami s prekomerno in normalno težo?
    3. Ali obstaja interakcija med prekomerno težo in hipertiroidizmom?



Interakcije NI



Interakcija JE

Neparametrični analog:

- Enosmerna ANOVA→Kruskal-Wallis test

Naloga: enofaktorska ANOVA

	Skupina 1	Skupina 2	Skupina 3
	15	14	12
	14	13	15
	14	21	19
	13	14	15
	12	14	17
	15	13	14
	12	14	15
	13	15	12
	11	12	12
	14	12	14
	11	14	14
	12	15	16
	15	14	18
	9	15	15
	14	13	13
	18	12	12
	13	15	16
	14	12	18
	13		18
	15		17
			16
			18
			17
povprečje	13,35	14,00	15,35
st.dev.	1,93	2,06	2,21

Naloga: dvofaktorska ANOVA

Ali je stopnja penetracije učinkovine odvisna od vrste mazilne podlage, odstotka dodanega emulgatorja in njune interakcije na stopnji tveganja 0,05?

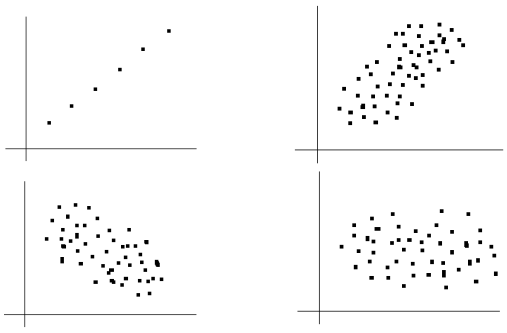
	Mazilna podlaga		
% emulg.	1	2	3
2	220	261	305
	226	245	292
	218	252	299
4	224	257	302
	231	260	310
	229	267	308
6	241	272	309
	230	259	312
	237	268	312

Uporabi:
General linear model!

KORELACIJA in LINEARNA REGRESIJA

- Sir Francis Galton: “co-relation”
- Korelacija: opis moči linearne povezave med dvema spremenljivkama
- Linearna regresija: opis povezave z linearno enačbo; napoved vrednosti *y*

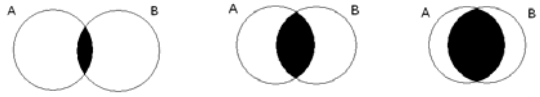
KORELACIJA



- Korelacijski koeficient **(r):**
  - merilo za velikost linearne povezave
  - od -1 do 1
  - izračun: stat. program, formula

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

- Determinacijski koeficient (r^2,R^2):
  - pove, kolikšen del variance ***y*** pojasnimo s spremenljivko ***x***
  - meri MOČ povezave, ne pove ničesar o smeri povezave
  - od 0 do 1



R^2=0,10 R^2=0,30 R^2=0,80

- t-test za korelacijo
  - H_0: $\rho=\rho_0$ ($\rho_0=0$)
  - t-test:

$$t_{eks} = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \qquad t_{tab(d.f.=N-2)}$$

  - Primer: r=0,35 ; N=92

- Fisher-jeva ***z*** transformacija
  - ugotavljanje, ali je korelacija enaka določeni specifični vrednosti, ki ni 0.
  - Primer: nova diagnostična metoda (ali je r>0,90)
  - ρ=0 → ukrivljena porazdelitev vzorčnih vrednosti → potrebna transformacija *r*

$$z(r) = \frac{1}{2} ln\left(\frac{(1+r)}{(1-r)}\right)$$

- z(r) ~ dnorm

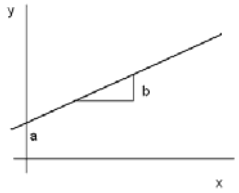
$$z_{eks} = \frac{z(r) - z(\pi)}{\sqrt{1/(n-3)}}$$

- Primer: Ugotoviti hočemo, ali je povezava med diagn. testoma signifikantno večja od 0,8. Na 300 bolnikih smo ugotovili, da je r=0,86.

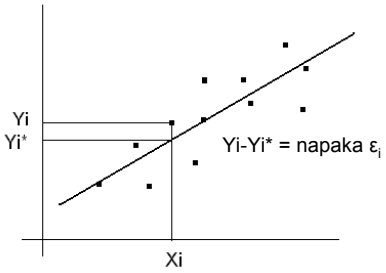
- Merjenje korelacije, ko spremenljivke ***niso*** normalno porazdeljene oz. imamo ***ordinalne*** spremenljivke:
  - Spearman-ov ρ test oz. korelacija rangov
  - vrednosti pretvorimo v range in za te izračunamo korelacijo

LINEARNA REGRESIJA

- Eno- oz. multivariatna
- Y= a+bX



- Metoda najmanjše vsote kvadratov (LSM)



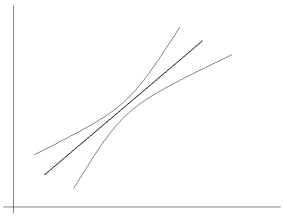
LSM: min $\sum(Y_i - Y_i^*)^2$

- Izračun koeficientov:

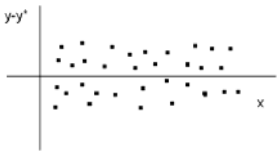
$$b = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$$

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

- Izračun signifikantnosti koeficientov *a* in *b*:
  - F-test oz. t-test
- Standardna napaka za koeficiente



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



- Kršitve predpostavk:
  - transformacija spremenljivk
  - regresija s polinomi (pri nelinearnosti)
- Multipla regresija

Naloga: Linearna regresija

Umeritvena premica
Paralelke: h1-h5

Konc (mg/L)	h1	h2	h3	h4	h5
0	0	-0,1	0,2	0,3	0
1	5,2	4,6	3,6	5,3	5,6
2	10	9,7	11,3	9,9	10,6
5	21,2	25,5	22,3	24,6	26
10	58,7	59,5	64,4	57,7	62,6
15	75,3	80,5	83,4	82,2	84,1
20	105,1	112,6	108,1	109,6	110,2