

1. del: Povezanost med dvema opisnima spremenljivkama

1. Zanima nas, ali v populaciji obstaja povezanost med lastništvom živali in izbrano fakulteto.

		Fakulteta		Vsota
		VF	MF	
Živali	DA	42	63	105
	NE	2	33	35
Vsota		44	96	140

- Verjetnost, da ima študent medicine domačo žival je:

$$p_{\text{žival}|\text{MF}} =$$

- Verjetnost, da ima študent veterine domačo žival je:

$$p_{\text{žival}|\text{VF}} =$$

- Zapišite ničelno domnevo testa χ^2 (hi-kvadrat):

- Izračunajte pričakovane frekvence pod ničelno domnevo:

		Fakulteta		Vsota
		VF	MF	
Živali	DA			105
	NE			35
Vsota		44	96	140

- Izračunajte testno statistiko.

- Narišite porazdelitev testne statistike, vpišite dobljeno vrednost, izračunajte in označite kritično vrednost ter vrednost p . Ali ničelno domnevo zavrnemo?
- Zapišite vsebinski sklep.
- Kaj so bile predpostavke tega testa? Kaj lahko storimo v primeru, ko predpostavke testa niso izpolnjene?

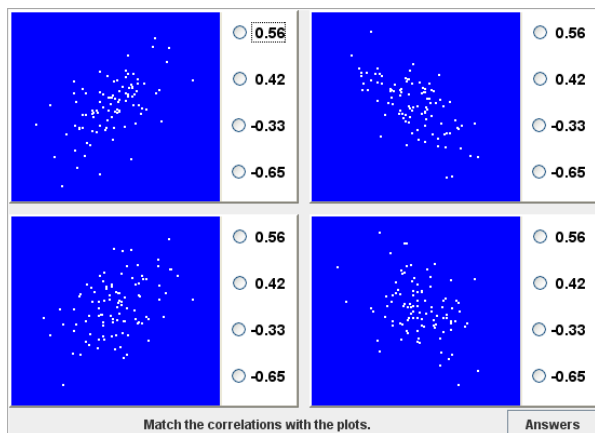
χ^2 test v R:

```
dd <- read.table("Ankete1011.txt", header = T, dec = ",", sep = "\t", fill = T)
dd$zival <- ifelse(dd$Domace_zivali == "Ne", "Ne", "Da")
```

```
chisq.test(table(dd$zival, dd$Fakulteta), cor = F) ## hi2 test brez Yatesovega popravka
chisq.test(table(dd$zival, dd$Fakulteta), cor = T) ## hi2 test z Yatesovim popravkom
fisher.test(table(dd$zival, dd$Fakulteta))      ## Fisherjev eksaktni test
```

2. del: Linearna regresija in korelacija

1. V spodnji štirih primerih ocenite korelacijski koeficient na podlagi razsevnega diagrama ¹.



Največja možna vrednost korelacijskega koeficienta je: _____

Najmanjša možna vrednost korelacijskega koeficienta je: _____

Če je korelacijski koeficient 0 pomeni, da: _____

Izračun korelacije v R:

```
x <- rnorm(100)
y <- rnorm(100)+x
cor(x, y)
cor.test(x, y)
```

2. Zaženite program R in odprite datoteko [Ankete1011.txt](#).

```
dd <- read.table("Ankete1011.txt", header = T, dec = ",", sep = "\t", fill = T)
```

- Narišite razsevni diagram, pri čemer naj bo odvisna spremenljivka teža (na ordinatni osi), neodvisna spremenljivka pa višina (na abscisni osi).

```
plot(dd$Visina, dd$Teza, xlab = "Visina", ylab = "Teza")
```

- V diagram vrišite regresijsko premico.

```
abline(a = lm(Teza ~ Visina, data = dd)$coef[1],
       b = lm(Teza ~ Visina, data = dd)$coef[2])
```

- Geometrijsko ocenite približno vrednost regresijske konstante in regresijskega koeficienta ter presodite, kolikšna se vam zdi razpršenost točk okoli regresijske premice.

- Za iste podatke izvedite linearno regresijsko analizo ter dobljeni vrednosti regresijske konstante in regresijskega koeficienta primerjajte s približno ocenjenima.

```
fit <- lm(Teza ~ Visina, data = dd)
```

¹Kratek opis korelacijskega koeficienta najdete na <http://en.wikipedia.org/wiki/Correlation>.

Interpretirajte vsako izmed vrednosti v izpisu `summary(fit)`:

- Korelacijski koeficient r in njegova interpretacija:
- Delež pojasnjene variabilnosti R^2 in njegova interpretacija:
- Kakšna je največja možna vrednost pojasnjene variabilnosti in kaj pomeni?
- Regresijska konstanta:
- Regresijski koeficient in njegova vsebinska interpretacija:
- Razložite podobnosti in razlike med regresijskim in korelacijskim koeficientom:
 - Predznak:
 - Interpretacija vrednosti 0:
 - Kateri izmed njiju je standardiziran?
- Kakšno težo pričakujete za študenta, ki ima 170 cm?
- Višina dveh študentov se razlikuje za 5 cm. Kakšna bo v povprečju razlika v njuni teži?
- Ničelna domneva:
- Izračun testne statistike:
- Kritična vrednost, vrednost p in statistični sklep:
- Vsebinski sklep:
- Izračunajte 95% interval zaupanja za regresijski koeficient in ga interpretirajte:

- Ali je smiselno interpretirati koeficient pri konstanti?
- Kaj lahko na podlagi modela poveste o teži 100 cm velikega otroka?
- Kaj so predpostavke za uporabo linearne regresije? Na podlagi razsevnega diagrama komentirajte njihovo izpolnjenost.

3. Katere spremenljivke so lahko še povezane s telesno težo poleg višine?

Oglejte si primer, kjer smo vključili spol kot neodvisno spremenljivko. Kodiranje spola: 1: moški, 2: ženski.

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	89,018	3,305		,000
	Spol	-14,599	1,884	-,551	,000

a. Dependent Variable: Teza

- Kakšno težo model napove za moškega in kakšno za žensko?
- Kako interpretiramo regresijski koeficient?
- Zapišite ničelno domnevo.
- Kaj zaključimo na podlagi rezultatov?

4. Primerjajte dobljene rezultate z rezultati testa t za dva neodvisna vzorca.

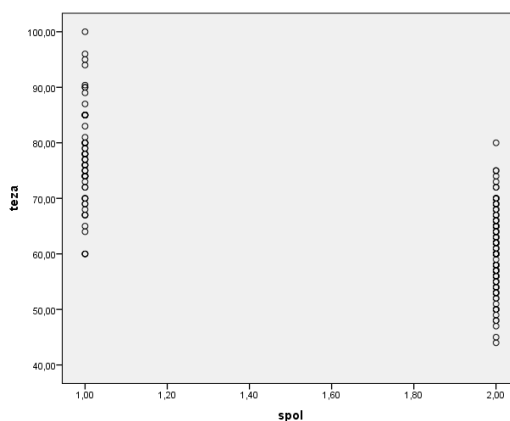
Group Statistics

	Spol	N	Mean	Std. Deviation	Std. Error Mean
Teza	moski	43	74,419	11,2594	1,7170
	zenski	97	59,820	9,8246	,9975

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Teza	Equal variances assumed	2,330	,129	7,750	138	,000	14,5990	1,8838	10,8741	18,3239
	Equal variances not assumed			7,352	71,569	,000	14,5990	1,9858	10,6401	18,5580

- Zapišite ničelno domnevo za test t . Primerjajte z ničelno domnevo pri linearni regresiji.
- Ali so rezultati obeh testov podobni?
- Primerjajte napovedani teži iz regresijskega modela s povprečnima težama v vsaki skupini.
- Vrišite premico v spodnji graf. Skozi kateri dve točki poteka?



- Ali testa odgovarjata na enako vprašanje? Ali odgovarjata enako?

5. Odprite datoteko `teza.txt`, ki vsebuje podatke o teži žensk pred in po nosečnosti. Težo pred in po porodu primerjajte najprej s parnim testom t , nato pa še z linearno regresijo. Pri vsaki analizi zapišite ničelno domnevo ter interpretirajte rezultate.
- Parni test t :
 - Ničelna domneva:
 - Vrednost p in statistični sklep:
 - Interpretacija rezultatov:
 - Linearna regresija:
 - Ničelna domneva:
 - Vrednost p in statistični sklep:
 - Interpretacija rezultatov:
 - Ali si lahko zamislite primer, ko bi eden izmed testov dal značilen rezultat, drugi pa ne?

6. Vrnite se k podatkom iz ankete ([Ankete1011.txt](#)). Ocenimo hkrati povezanost višine in spola s težo: v `lm` uporabite formulo `Teza ~ Visina + Spol`. Rezultate tega modela bomo primerjali s prejšnjim modelom `Teza ~ Visina`.

- Kaj se je zgodilo z deležem pojasnjene variabilnosti glede na prejšnji model? Komentirajte.
- Interpretirajte regresijske koeficiente in pripadajoči vrednosti p . Primerjajte z interpretacijo v prejšnjem modelu.
- Napovejte teži za 170 cm velikega moškega in za 170 cm veliko žensko. Ali sta različni ali enaki? Primerjajte s prejšnjim modelom.
- Kakšna je pričakovana razlika v teži za dva moška, katerih višina se razlikuje za 1 cm? Kaj pa za dve ženski?
- Komentirajte slednji rezultat: Kaj to pomeni? Ali je to smiselno? Kdaj je smiselno?