Analiza podatko študentov biologije
3. letnik 2011/12

Andrej Blejec

19. januar 2012

Kazalo

Povzetek

Študentje biologije so v okviru predmeta Statistika pripravili zbirko podatkov o nekaterih telesnih merah sebe in staršev. V tem zapisu so prikazane nekatere možne analize.

Kazalo

```
> lfn <- "ST2011.txt"
```

Ime datoteke s podatki ST2011.txt.

1 Vnos podatkov

Podatki so bili zbrani na spletni strani predmeta *BI022 Statistika* na sistemu Moodle. Shranjeni so v datoteki ST2011.txt.

Vnos podatkov

```
> data <- read.table(file.path("../data",
+     lfn), sep = "\t", header = TRUE)
> str(data)
'data.frame':        51 obs. of  12 variables:
 $ starost: int  58 23 21 20 21 20 21 21 21 23 ...
 $ mesec  : int  7 3 4 10 1 4 10 4 7 1 ...
 $ spol   : Factor w/ 2 levels "F ženski ","M moški ": 2 1 1 1 1 1 1 1 1 2 ...
 $ masa   : int  88 54 42 60 52 56 58 77 60 95 ...
 $ visina : num  178 174 152 170 171 170 170 169 170 190 ...
 $ roke   : int  175 171 150 155 171 174 169 169 168 201 ...
 $ cevelj : int  44 40 37 39 39 40 39 40 39 46 ...
 $ lasje  : Factor w/ 2 levels "S svetli ","T temni ": 2 1 2 1 1 1 1 2 2 2 ...
 $ oci    : Factor w/ 2 levels "S svetli ","T temni ": 1 1 2 1 1 1 1 2 2 2 ...
 $ majica : Factor w/ 6 levels "L ","M ","S ",..: 1 3 6 2 2 3 2 2 3 4 ...
 $ mati   : num  150 173 157 160 173 174 155 170 155 174 ...
 $ oce    : int  180 189 159 190 183 186 179 175 189 176 ...
```

Osnovni, vnešeni podatki

```
> data
```

	starost	mesec	spol	masa	visina	roke
1	58	7	M moški	88	178.0	175
2	23	3	F ženski	54	174.0	171
3	21	4	F ženski	42	152.0	150
4	20	10	F ženski	60	170.0	155
5	21	1	F ženski	52	171.0	171
6	20	4	F ženski	56	170.0	174
7	21	10	F ženski	58	170.0	169
8	21	4	F ženski	77	169.0	169
9	21	7	F ženski	60	170.0	168
10	23	1	M moški	95	190.0	201
11	21	0	M moški	73	173.0	179
12	22	9	M moški	67	176.0	179
13	22	3	F ženski	58	173.0	172
14	22	9	F ženski	60	169.5	164
15	20	1	F ženski	55	164.0	168
16	20	9	F ženski	60	168.0	170
17	22	9	M moški	63	177.0	185
18	20	3	F ženski	61	175.0	174
19	21	9	F ženski	61	166.0	175
20	21	8	F ženski	59	165.0	162
21	20	11	F ženski	55	162.0	160
22	20	4	F ženski	52	165.0	0
23	22	6	F ženski	62	167.0	165
24	21	12	F ženski	61	173.0	170
25	21	7	F ženski	55	158.0	155
26	22	0	F ženski	92	196.0	0
27	21	4	F ženski	57	171.0	166

```
28      23     2 F ženski    65  163.0  166
29      21     3  M moški    57  168.0  168
30      21    11 F ženski    58  170.0  159
31      23     4  M moški    83  179.0    0
32      22    10  M moški    70  175.0  177
33      21     4 F ženski    60  170.0  167
34      21     5 F ženski    65  164.0    0
35      20     4 F ženski    56  172.0  171
36      22     5 F ženski    55  165.0  161
37      21     9 F ženski    66  164.0  166
38      21     6  M moški    95  194.0  182
39      21     1 F ženski    58  163.0  160
40      21     6  M moški    70  175.0  170
41      21     7 F ženski    55  170.0  171
42      22     7 F ženski    65  168.0  130
43      20    12 F ženski    64  172.0  156
44      20    11 F ženski    74  173.0  172
45      22     6 F ženski    57  161.0  150
46      21     2 F ženski    53  165.0  170
47      24     7  M moški    66  178.0  176
48      20     1 F ženski    53  165.0  165
49      23     9 F ženski    62  166.0  166
50      22     9 F ženski    52  172.0  170
51      21     6 F ženski    49  164.0    8
   cevelj     lasje       oci majica  mati oce
1      44  T temni  S svetli      L  150.0 180
2      40 S svetli  S svetli      S  173.0 189
3      37  T temni   T temni    XXS  157.0 159
4      39 S svetli  S svetli      M  160.0 190
5      39 S svetli  S svetli      M  173.0 183
6      40 S svetli  S svetli      S  174.0 186
7      39 S svetli  S svetli      M  155.0 179
8      40  T temni   T temni      M  170.0 175
9      39  T temni   T temni      S  155.0 189
10     46  T temni   T temni     XL  174.0 176
11     43  T temni  S svetli      M  168.0 174
12     42 S svetli  S svetli      S  167.0 175
13     38  T temni   T temni      M  165.0 182
14     39 S svetli  S svetli      M  175.5 176
15     40  T temni  S svetli      S  158.0 183
16     40  T temni  S svetli      M  164.0 180
17     42 S svetli   T temni      M  155.0 177
18     41  T temni  S svetli      M  175.0 183
19     41  T temni   T temni      M  165.0 181
20     38  T temni   T temni      S  162.0 175
21     37 S svetli  S svetli      S  166.0 169
22     37  T temni   T temni      S  155.0 178
23     39  T temni  S svetli      S  166.0 168
24     39  T temni  S svetli      M  161.0 184
```

```
25     37  T temni   T temni      S  178.0 170
26     49  T temni   T temni      M    0.0   0
27     38  T temni   T temni      S  165.0   0
28     39  T temni  S svetli      M  161.0 176
29     41  T temni  S svetli      M  151.0 173
30     29 S svetli   T temni      S  160.0 179
31     43  T temni   T temni      M    0.0 177
32     43 S svetli  S svetli      L  158.0 179
33     39 S svetli  S svetli      S  169.0 177
34     40 S svetli  S svetli      M  168.0 182
35     40  T temni   T temni      S  163.0 181
36     38 S svetli  S svetli      S  170.0 180
37     40  T temni   T temni      M  172.0 184
38     43  T temni  S svetli      M  164.0 184
39     38  T temni   T temni      M  159.0 177
40     43  T temni   T temni      S  160.0 182
41     39 S svetli  S svetli     XS  173.0 181
42     39 S svetli  S svetli      S  164.0 184
43     38 S svetli   T temni      M  176.0 180
44     40  T temni   T temni      M  168.0 175
45     37  T temni   T temni      L   50.0  53
46     38  T temni  S svetli      M  160.0 172
47     42  T temni   T temni      M  170.0 175
48     39 S svetli  S svetli     XS  168.0 172
49     38  T temni   T temni      M  162.0 180
50     38  T temni  S svetli      S  175.0 180
51     38 S svetli  S svetli     XS  159.0 172
```

Vrste spremenljivk

Opisne spremenljivke

```
> opisne <- which(sapply(data, "class") ==
+     "factor")
> length(opisne)
[1] 4
> names(data)[opisne]
[1] "spol"   "lasje"  "oci"    "majica"
```

Številske spremenljivke

```
> stevilske <- which(!sapply(data, "class") ==
+     "factor")
> length(stevilske)
[1] 8
> names(data)[stevilske]
[1] "starost" "mesec"   "masa"    "visina"
[5] "roke"    "cevelj"  "mati"    "oce"
```

2 Pregled osnovnih podatkov

Opisna statistika

Opisna statistika

```
> summary(data[, 1:6])
```

```
    starost          mesec              spol
 Min.   :20.00   Min.   : 0.000   F ženski :40
 1st Qu.:21.00   1st Qu.: 3.500   M moški  :11
 Median :21.00   Median : 6.000
 Mean   :21.98   Mean   : 5.922
 3rd Qu.:22.00   3rd Qu.: 9.000
 Max.   :58.00   Max.   :12.000
      masa            visina           roke
 Min.   :42.00   Min.   :152.0   Min.   :  0.0
 1st Qu.:55.50   1st Qu.:165.0   1st Qu.:160.0
 Median :60.00   Median :170.0   Median :168.0
 Mean   :62.76   Mean   :170.4   Mean   :151.5
 3rd Qu.:65.50   3rd Qu.:173.0   3rd Qu.:171.5
 Max.   :95.00   Max.   :196.0   Max.   :201.0
```

Opisna statistika

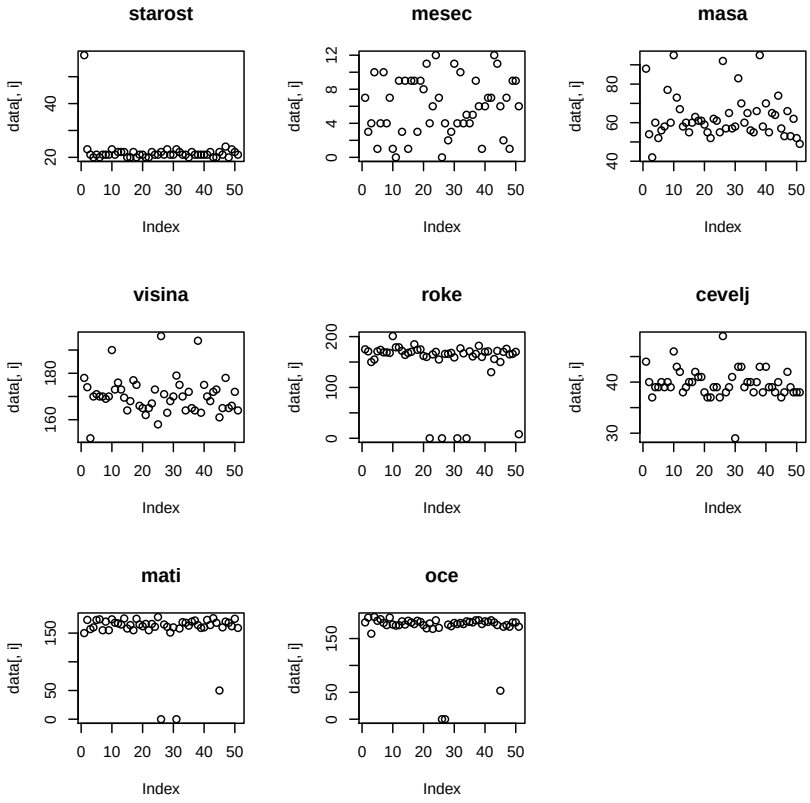
```
> summary(data[, 7:12])
```

```
     cevelj            lasje           oci
 Min.   :29.00   S svetli :19   S svetli :28
 1st Qu.:38.00   T temni  :32   T temni  :23
 Median :39.00
 Mean   :39.71
 3rd Qu.:41.00
 Max.   :49.00
  majica        mati            oce
 L   : 3   Min.   :  0.0   Min.   :  0.0
 M   :25   1st Qu.:159.0   1st Qu.:175.0
 S   :18   Median :164.0   Median :179.0
 XL  : 1   Mean   :156.2   Mean   :168.9
 XS  : 3   3rd Qu.:170.0   3rd Qu.:182.0
 XXS : 1   Max.   :178.0   Max.   :190.0
```

Grafični pregled podatkov

Grafični pregled podatkov

```
> par(mfrow = c(3, 3))
> for (i in stevilske) plot(data[, i], main = names(data)[i])
```



Pri zbiranju podatkov so manjkajoče vrednosti spremenjene v vrednost nič. Za nobeno spremenljivko taka vrednost ni smiselna, zato dobro označuje manjkajočo vrednost. Zaradi enostavnejše analize bomo enote s takimi vrednostmi pri katerikoli spremenljivki izločili.

Odstranitev enot z manjkajočimi podatki

```
> nas <- which(apply(data, 1, function(x) any(as.numeric(x) ==
+     0)))
> nas
[1] 11 22 26 27 31 34
> if (length(nas) > 0) data <- data[-nas,
+     ]
> dim(data)
[1] 45 12
```

Odstranimo še podatke nenavadno starega človeka in podatke osebe s kratko nogo.

Odstranitev podatkov

Odstranimo podatke prestarega človeka :)

```
> data <- data[data$starost < 25, ]
```

in še enega kratkorokega

```
> data <- data[data$roke > 50, ]
```

Popravimo podatke

Popravek podatkov

```
> data[, "mati"] <- data[, "mati"] + 100 *
+     (data[, "mati"] < 100)
> data[, "oce"] <- data[, "oce"] + 100 *
+     (data[, "oce"] < 100)
```

Pri prepisu podatkov iz sistema Moodle dobijo vrednosti opisnih spremenljivk na koncu nepotreben znak presledek. Odstranimo jih lahko z uporabo 'regularnih izrazov' (regular expression).

Ureditev imen nivojev

Odstranimo presledke v vrednostih

```
> x <- data[, "majica"]
> levels(x)
[1] "L "   "M "   "S "   "XL "  "XS "  "XXS "
> levels(x) <- gsub("(.*) ", "\\1", levels(x))
> levels(x)
[1] "L"   "M"   "S"   "XL"  "XS"  "XXS"
```

Neposredna sprememba imen nivojev

```
> levels(data[, "spol"]) <- c("zenski",
+     "moski")
> levels(data[, "oci"]) <- c("svetle", "temne")
> levels(data[, "lasje"]) <- c("svetli",
+     "temni")
```

Velikost majice je merjena na urejenostni lestvici, zato vrednosti uredimo po velikosti

Urejenostna merska lestvica

Spremenimo vrstni red nivojev

```
> (data[, "majica"] <- ordered(x, levels = c("XXS",
+     "XS", "S", "M", "L", "XL")))
 [1] S   XXS M   M   S   M   M   S   XL  S   M
[12] M   S   M   M   M   M   S   S   S   M   S
[23] M   M   S   L   S   S   S   M   M   M   S
[34] XS  S   M   M   L   M   M   XS  M   S
Levels: XXS < XS < S < M < L < XL
```

3 Pregled urejenih podatkov

```
> data
   starost mesec   spol masa visina roke cevelj
2       23     3 zenski   54  174.0  171     40
3       21     4 zenski   42  152.0  150     37
4       20    10 zenski   60  170.0  155     39
5       21     1 zenski   52  171.0  171     39
6       20     4 zenski   56  170.0  174     40
7       21    10 zenski   58  170.0  169     39
8       21     4 zenski   77  169.0  169     40
9       21     7 zenski   60  170.0  168     39
10      23     1  moski   95  190.0  201     46
12      22     9  moski   67  176.0  179     42
13      22     3 zenski   58  173.0  172     38
14      22     9 zenski   60  169.5  164     39
15      20     1 zenski   55  164.0  168     40
16      20     9 zenski   60  168.0  170     40
17      22     9  moski   63  177.0  185     42
18      20     3 zenski   61  175.0  174     41
19      21     9 zenski   61  166.0  175     41
20      21     8 zenski   59  165.0  162     38
21      20    11 zenski   55  162.0  160     37
23      22     6 zenski   62  167.0  165     39
24      21    12 zenski   61  173.0  170     39
25      21     7 zenski   55  158.0  155     37
28      23     2 zenski   65  163.0  166     39
29      21     3  moski   57  168.0  168     41
30      21    11 zenski   58  170.0  159     29
32      22    10  moski   70  175.0  177     43
33      21     4 zenski   60  170.0  167     39
35      20     4 zenski   56  172.0  171     40
36      22     5 zenski   55  165.0  161     38
37      21     9 zenski   66  164.0  166     40
38      21     6  moski   95  194.0  182     43
39      21     1 zenski   58  163.0  160     38
40      21     6  moski   70  175.0  170     43
41      21     7 zenski   55  170.0  171     39
42      22     7 zenski   65  168.0  130     39
43      20    12 zenski   64  172.0  156     38
44      20    11 zenski   74  173.0  172     40
45      22     6 zenski   57  161.0  150     37
46      21     2 zenski   53  165.0  170     38
47      24     7  moski   66  178.0  176     42
48      20     1 zenski   53  165.0  165     39
49      23     9 zenski   62  166.0  166     38
50      22     9 zenski   52  172.0  170     38
    lasje    oci majica  mati oce
2  svetli svetle      S 173.0 189
3   temni  temne    XXS 157.0 159
```

```
4  svetli svetle      M 160.0 190
5  svetli svetle      M 173.0 183
6  svetli svetle      S 174.0 186
7  svetli svetle      M 155.0 179
8   temni  temne      M 170.0 175
9   temni  temne      S 155.0 189
10  temni  temne     XL 174.0 176
12 svetli svetle      S 167.0 175
13  temni  temne      M 165.0 182
14 svetli svetle      M 175.5 176
15  temni svetle      S 158.0 183
16  temni svetle      M 164.0 180
17 svetli  temne      M 155.0 177
18  temni svetle      M 175.0 183
19  temni  temne      M 165.0 181
20  temni  temne      S 162.0 175
21 svetli svetle      S 166.0 169
23  temni svetle      S 166.0 168
24  temni svetle      M 161.0 184
25  temni  temne      S 178.0 170
28  temni svetle      M 161.0 176
29  temni svetle      M 151.0 173
30 svetli  temne      S 160.0 179
32 svetli svetle      L 158.0 179
33 svetli svetle      S 169.0 177
35  temni  temne      S 163.0 181
36 svetli svetle      S 170.0 180
37  temni  temne      M 172.0 184
38  temni svetle      M 164.0 184
39  temni  temne      M 159.0 177
40  temni  temne      S 160.0 182
41 svetli svetle     XS 173.0 181
42 svetli svetle      S 164.0 184
43 svetli  temne      M 176.0 180
44  temni  temne      M 168.0 175
45  temni  temne      L 150.0 153
46  temni svetle      M 160.0 172
47  temni  temne      M 170.0 175
48 svetli svetle     XS 168.0 172
49  temni  temne      M 162.0 180
50  temni svetle      S 175.0 180
```

3.1 Opisna statistika

Opisna statistika

```
> summary(data[, 1:6])
```

```
    starost          mesec           spol
 Min.   :20.00   Min.   : 1.000   zenski:35
 1st Qu.:21.00   1st Qu.: 3.500   moski : 8
 Median :21.00   Median : 7.000
 Mean   :21.26   Mean   : 6.326
 3rd Qu.:22.00   3rd Qu.: 9.000
 Max.   :24.00   Max.   :12.000
      masa            visina           roke
 Min.   :42.00   Min.   :152.0   Min.   :130.0
 1st Qu.:55.50   1st Qu.:165.0   1st Qu.:163.0
 Median :60.00   Median :170.0   Median :169.0
 Mean   :61.44   Mean   :169.7   Mean   :167.4
 3rd Qu.:64.50   3rd Qu.:173.0   3rd Qu.:171.5
 Max.   :95.00   Max.   :194.0   Max.   :201.0
```

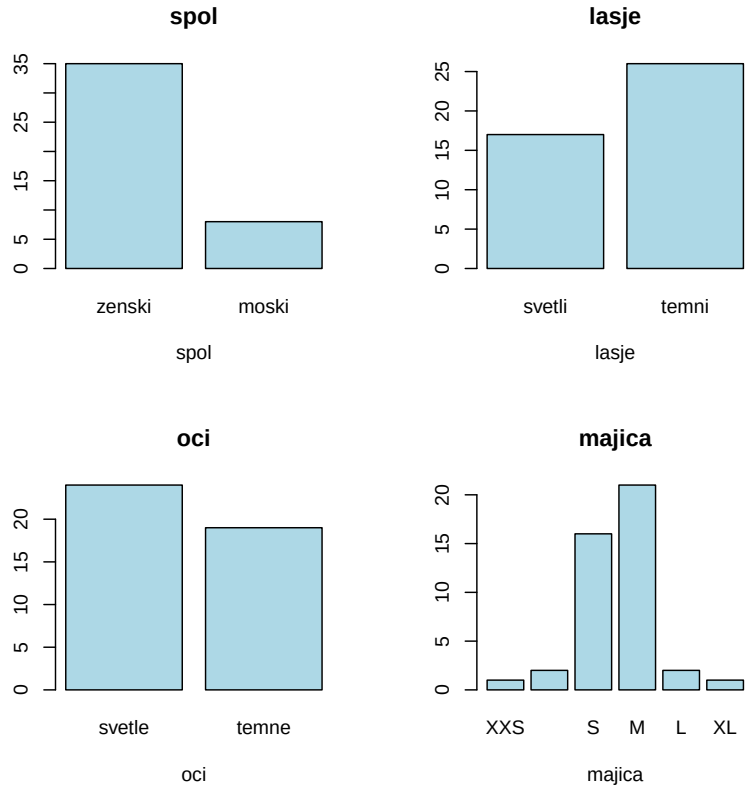
Opisna statistika

```
> summary(data[, 7:12])
     cevelj         lasje        oci     majica
 Min.   :29.00   svetli:17   svetle:24   XXS: 1
 1st Qu.:38.00   temni :26   temne :19   XS : 2
 Median :39.00                           S  :16
 Mean   :39.37                           M  :21
 3rd Qu.:40.00                           L  : 2
 Max.   :46.00                           XL : 1
      mati            oce
 Min.   :150.0   Min.   :153.0
 1st Qu.:160.0   1st Qu.:175.0
 Median :165.0   Median :179.0
 Mean   :165.2   Mean   :178.0
 3rd Qu.:171.0   3rd Qu.:182.5
 Max.   :178.0   Max.   :190.0
```

3.2 Porazdelitve podatkov

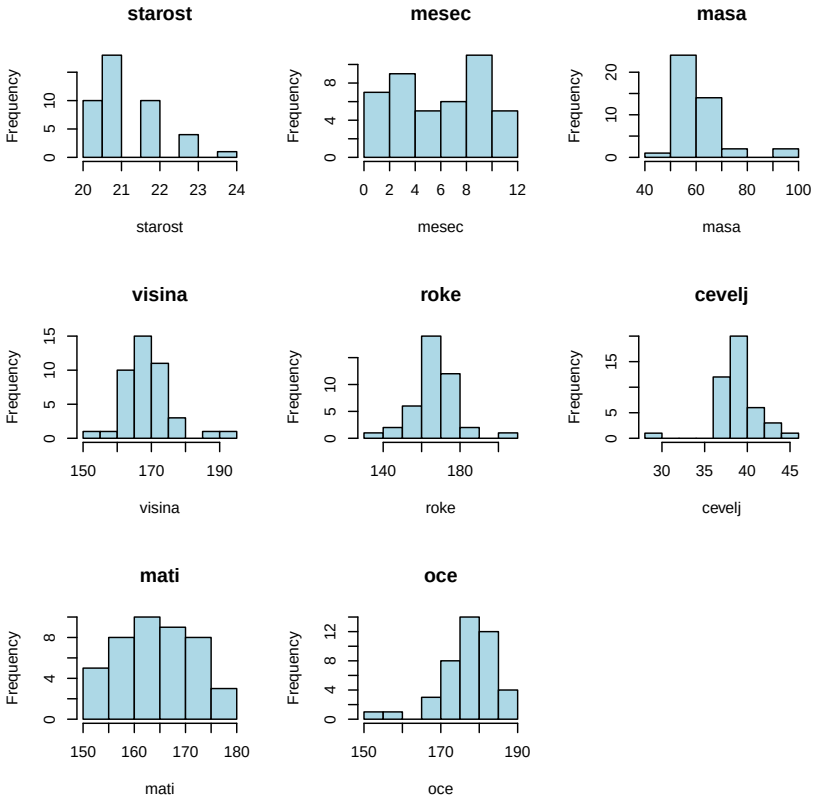
Porazdelitve opisnih spremenljivk

Porazdelitve opisnih spremenljivk



Porazdelitve številskih spremenljivk

Porazdelitve številskih spremenljivk



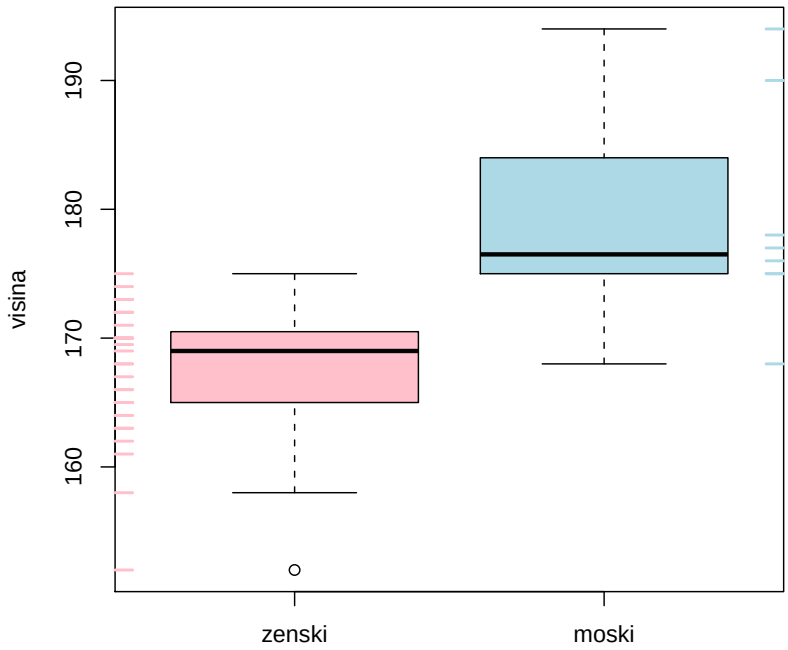
Omogočimo neposredno rabo spremenljivk.

```
> attach(data)
```

4 Testiranje hipotez

4.1 Ali so fantje večji od deklet?

Porazdelitvi



Porazdelitvi - R ukazi

```
> boxplot(visina ~ spol, col = c("pink",
+     "lightblue"), ylab = "visina")
> rug(visina[spol == "zenski"], side = 2,
+     col = "pink", lwd = 2)
> rug(visina[!spol == "zenski"], side = 4,
+     , col = "lightblue", lwd = 2)
```

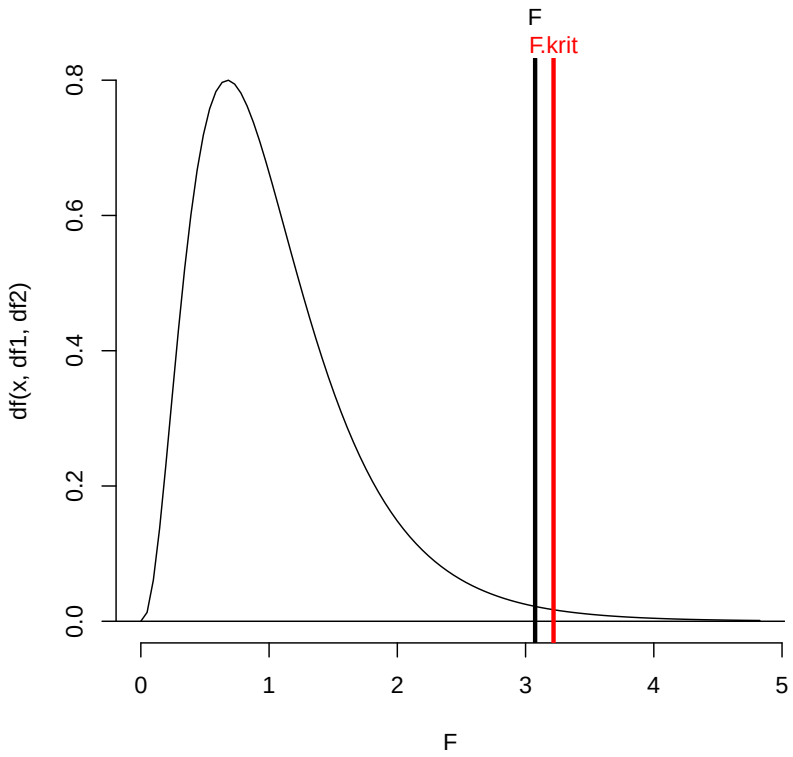
Opisni pregled

```
> y <- visina
> (n <- tapply(y, spol, length))
zenski  moski
    35      8
> (xbar <- tapply(y, spol, mean))
  zenski    moski
167.5857 179.1250
> (s <- tapply(y, spol, sd))
  zenski    moski
4.881667 8.559665
```

Ali sta varianci značilno različni?

```
> alpha <- 0.01
> (v <- sort(s^2))
  zenski    moski
23.83067 73.26786
> ns <- as.vector(n[order(s)])
> (F <- as.vector(v[2]/v[1]))
[1] 3.074519
> (df1 <- ns[2] - 1)
[1] 7
> (df2 <- ns[1] - 1)
[1] 34
> (F.krit <- qf(1 - alpha, df1, df2))
[1] 3.218154
> (p <- 1 - pf(F, df1, df2))
[1] 0.01278561
> if (F < F.krit) cat("Varianci NISTA statistično značilno različni (p =",
+     round(p, 3), ").\n") else cat("Varianci STA  statistično zančilno različni
+     alpha, ") (p =", round(p, 3), ").\n")
Varianci NISTA statistično značilno različni (p = 0.013 ).
```

Ali sta varianci značilno različni?



Kako smo narisali sliko?

```
> x <- seq(0, max(F, F.krit) * 1.5, length = 100)
> plot(x, df(x, df1, df2), type = "l", xlab = "F",
+     axes = FALSE)
> axis(1)
> axis(2)
> abline(h = 0)
> abline(v = F, lwd = 3)
> mtext("F", side = 3, line = 1, at = F)
> abline(v = F.krit, col = "red", lwd = 3)
> mtext("F.krit", side = 3, line = 0, at = F.krit,
+     col = "red")
```

Studentov t-test razlike povprečij

Hipoteze in delni rezultati

Uredimo vrstni red delnih rezultatov za test hipotez

$$H_0 : \mu_{moski} = \mu_{zenski} + \Delta$$

$$H_1 : \mu_{moski} > \mu_{zenski} + \Delta$$

```
> ord <- c("moski", "zenski")
> (xbar <- as.vector(xbar[ord]))
[1] 179.1250 167.5857
> (s <- as.vector(s[ord]))
[1] 8.559665 4.881667
> (n <- as.vector(n[ord]))
[1]  8 35
```

Stopnja tveganja in kritične vrednosti

```
> alpha <- 0.01
> delta <- 0
> (df <- n[1] + n[2] - 2)
[1] 41
> (t.krit <- qt(1 - alpha, df))
[1] 2.420803
```

Studentov t-test

```
> xbar[1] - xbar[2]
```

```
[1] 11.53929
```

```
> s2 <- ((n[1] - 1) * s[1]^2 + (n[2] - 1) *
+     s[2]^2)/(n[1] + n[2] - 2)
> (t <- (xbar[1] - xbar[2] - delta)/sqrt(s2) *
+     sqrt(n[1] * n[2]/(n[1] + n[2])))
```

```
[1] 5.18342
```

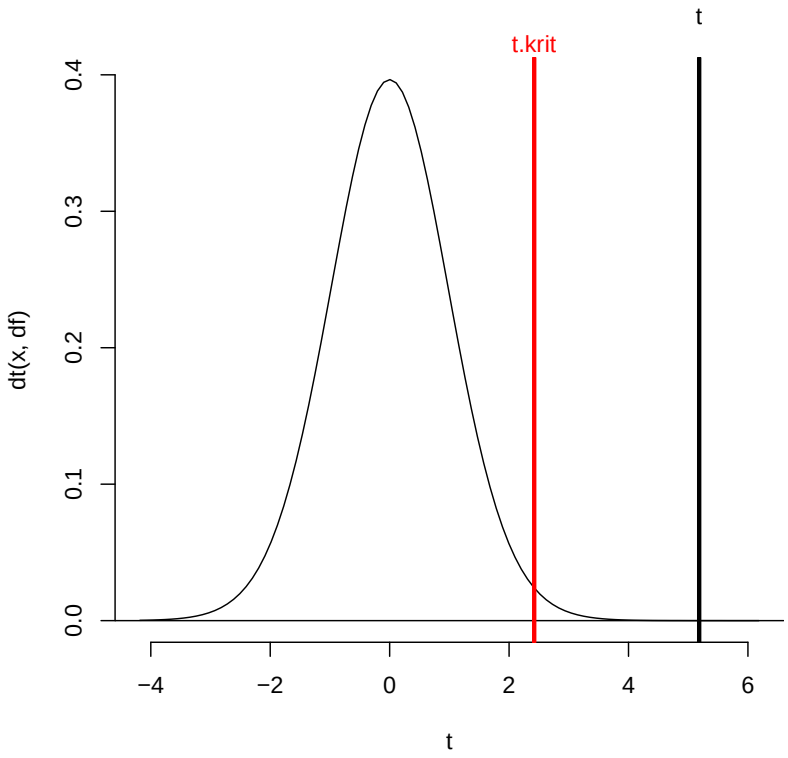
```
> (p <- 1 - pt(t, df))
```

```
[1] 3.102809e-06
```

```
> if (t < t.krit) cat("Povprečje1 NI statistično značilno večje (p =",
+     round(p, 3), ").\n") else cat("Povprečje1 JE  statistično zančilno večje (
+     alpha, ") (p =", round(p, 3), ").\n")
```

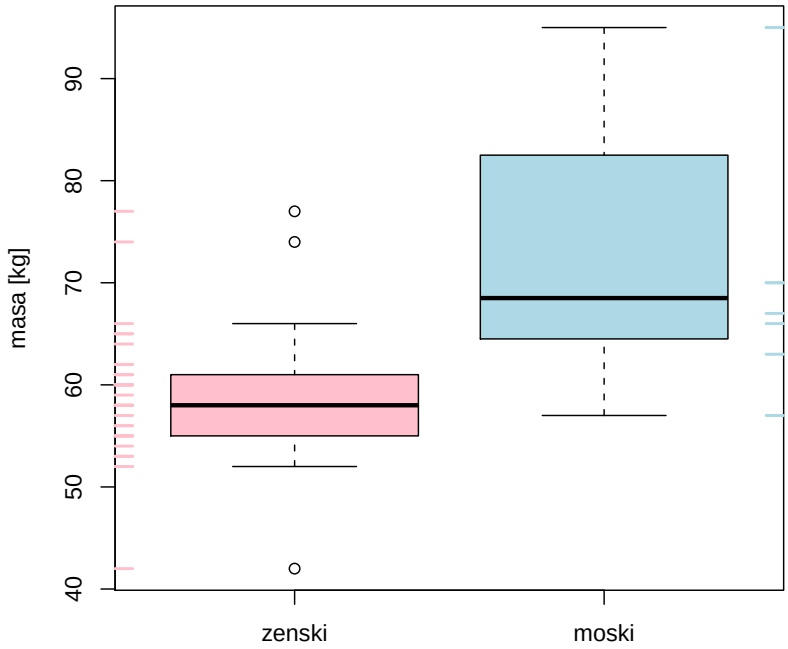
```
Povprečje1 JE  statistično zančilno večje (p < 0.01 ) (p = 0 ).
```

Slika



4.2 Ali so fantje tezji od deklet?

Porazdelitvi



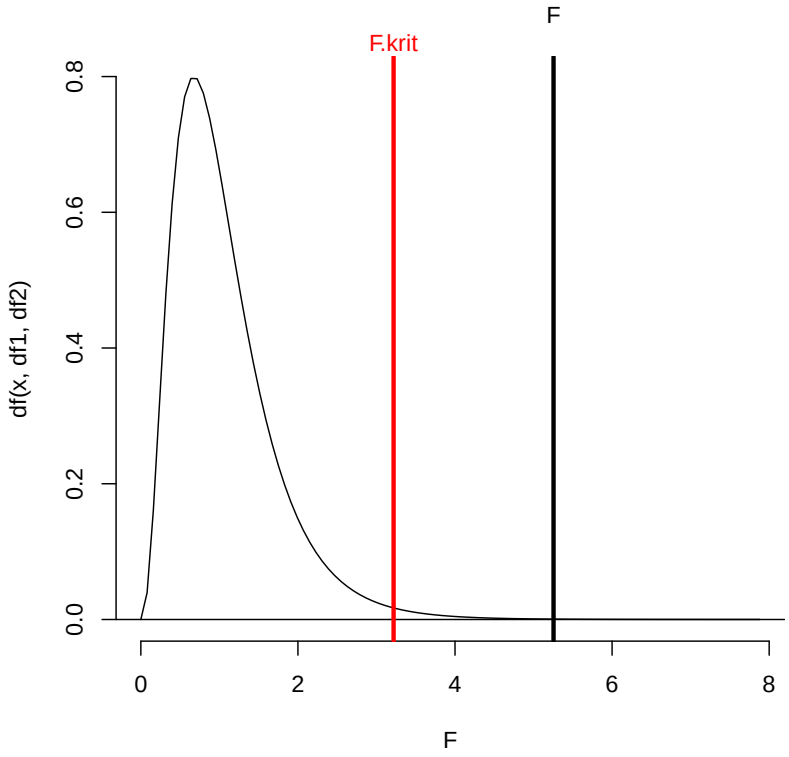
Opisni pregled

```
> (n <- tapply(y, spol, length))
zenski  moski
    35      8
> (xbar <- tapply(y, spol, mean))
  zenski   moski
58.82857 72.87500
> (s <- tapply(y, spol, sd))
   zenski    moski
 6.228425 14.277230
```

Ali sta varianci značilno različni?

```
> alpha <- 0.01
> (v <- sort(s^2))
   zenski    moski
 38.79328 203.83929
> ns <- as.vector(n[order(s)])
> (F <- as.vector(v[2]/v[1]))
[1] 5.2545
> (df1 <- ns[2] - 1)
[1] 7
> (df2 <- ns[1] - 1)
[1] 34
> (F.krit <- qf(1 - alpha, df1, df2))
[1] 3.218154
> (p <- 1 - pf(F, df1, df2))
[1] 0.0003909851
> if (F < F.krit) cat("Varianci NISTA statistično značilno različni (p =",
+     round(p, 3), ").\n") else cat("Varianci STA  statistično zančilno različni
+     alpha, ") (p =", round(p, 3), ") \n")
Varianci STA  statistično zančilno različni (p < 0.01 ) (p = 0 ).
```

Ali sta varianci značilno različni?



Kako smo narisali sliko?

```
> x <- seq(0, max(F, F.krit) * 1.5, length = 100)
> plot(x, df(x, df1, df2), type = "l", xlab = "F",
+     axes = FALSE)
> axis(1)
> axis(2)
> abline(h = 0)
> abline(v = F, lwd = 3)
> mtext("F", side = 3, line = 1, at = F)
> abline(v = F.krit, col = "red", lwd = 3)
> mtext("F.krit", side = 3, line = 0, at = F.krit,
+     col = "red")
```

Studentov t-test razlike povprečij

Hipoteze in delni rezultati

Uredimo vrstni red delnih rezultatov za test hipotez

$$H_0 : \mu_{moski} = \mu_{zenski} + \delta$$

$$H_1 : \mu_{moski} > \mu_{zenski} + \delta$$

```
> ord <- c("moski", "zenski")
> (xbar <- as.vector(xbar[ord]))
[1] 72.87500 58.82857
> (s <- as.vector(s[ord]))
[1] 14.277230  6.228425
> (n <- as.vector(n[ord]))
[1]  8 35
```

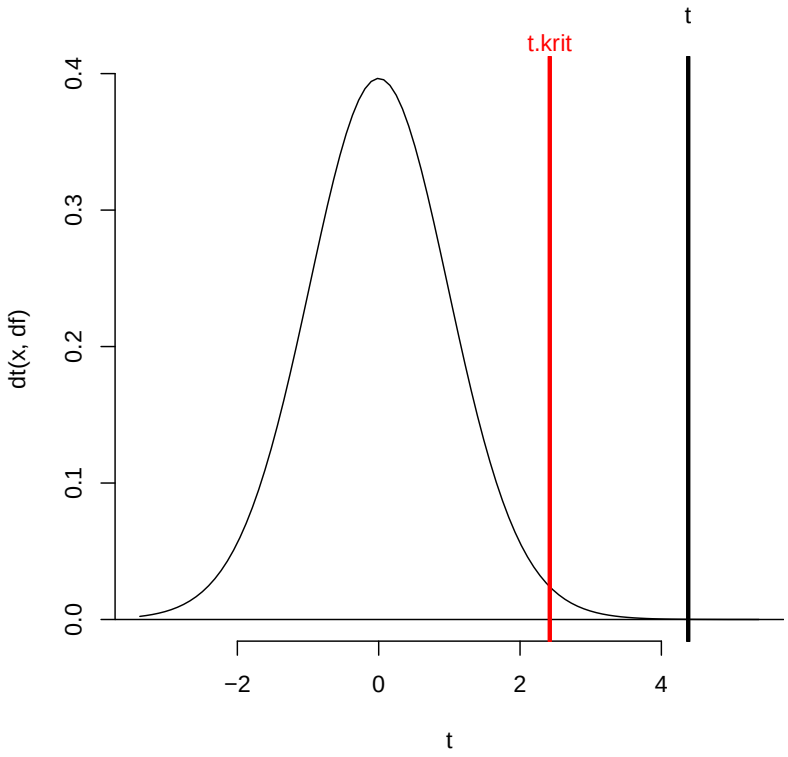
Stopnja tveganja in kritične vrednosti

```
> alpha <- 0.01
> delta <- 0
> (df <- n[1] + n[2] - 2)
[1] 41
> (t.krit <- qt(1 - alpha, df))
[1] 2.420803
```

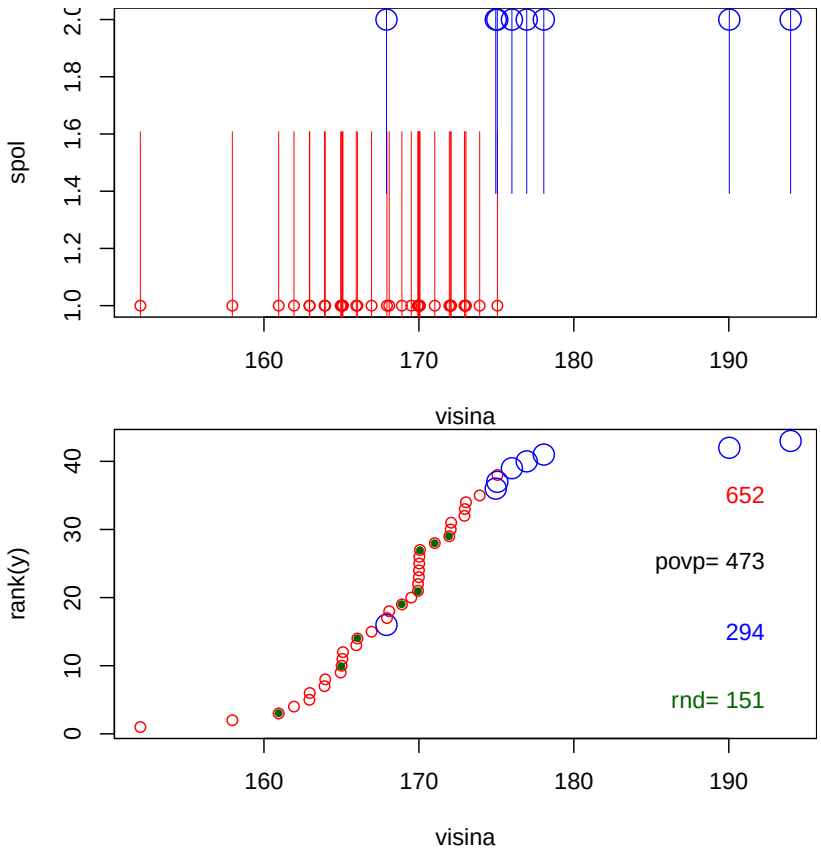
Studentov t-test

```
> xbar[1] - xbar[2]
[1] 14.04643
> s2 <- ((n[1] - 1) * s[1]^2 + (n[2] - 1) *
+     s[2]^2)/(n[1] + n[2] - 2)
> (t <- (xbar[1] - xbar[2] - delta)/sqrt(s2) *
+     sqrt(n[1] * n[2]/(n[1] + n[2])))
[1] 4.379903
> (p <- 1 - pt(t, df))
[1] 4.012688e-05
> if (t < t.krit) cat("Povprečje1 NI statistično značilno večje (p =",
+     round(p, 3), ").\n") else cat("Povprečje1 JE  statistično zančilno večje (
+     alpha, ") (p =", round(p, 3), ").\n")
Povprečje1 JE  statistično zančilno večje (p < 0.01 ) (p = 0 ).
```

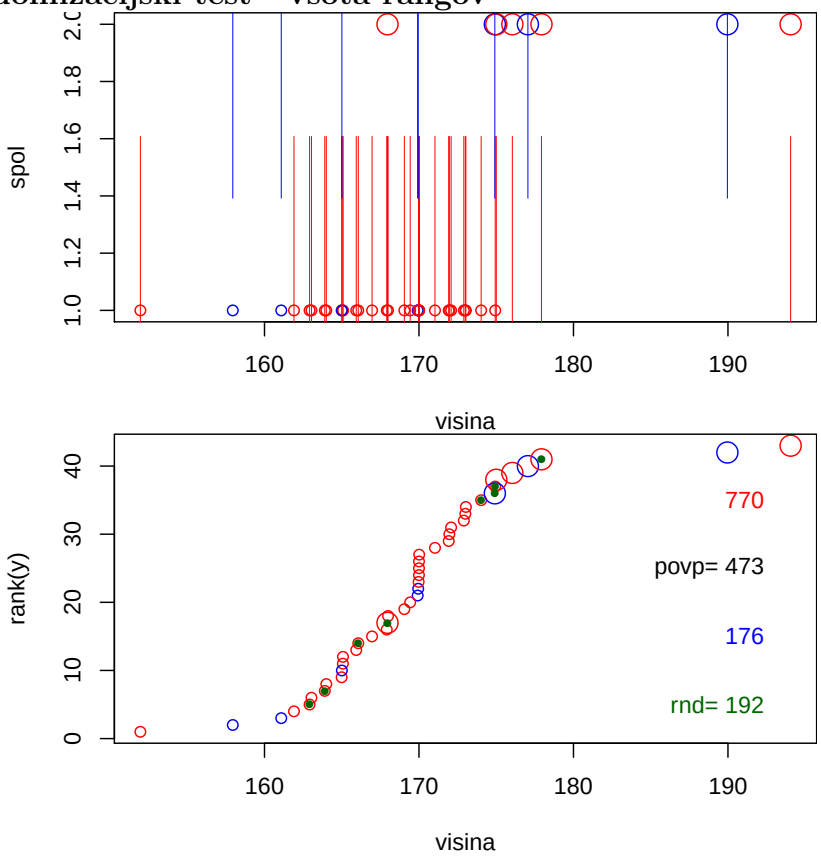
Slika



Neparametrični test - Wilcoxon



Randomizacijski test - vsota rangov



Wilcoxon test v R

```
> wilcox.test(jitter(visina) ~ spol)

        Wilcoxon rank sum test

data:  jitter(visina) by spol
W = 19, p-value = 2.296e-05
alternative hypothesis: true location shift is not equal to 0
```

5 Kontingenca

Barva las in oči

```
> f <- table(lasje, oci)
> f
        oci
lasje    svetle temnе
  svetli     14     3
  temni      10    16
> addmargins(f)
        oci
lasje    svetle temnе Sum
  svetli     14     3  17
  temni      10    16  26
  Sum        24    19  43
```

Pričakovane frekvence

```
> e <- outer(rowSums(f), colSums(f))/sum(f)
> addmargins(e)
          svetle     temnе Sum
svetli  9.488372  7.511628  17
temni  14.511628 11.488372  26
Sum    24.000000 19.000000  43
```

Razlika opaženega in pričakovanega

```
> f - e
        oci
lasje       svetle     temnе
  svetli  4.511628 -4.511628
  temni  -4.511628  4.511628
> (f - e)^2/e
        oci
lasje      svetle    temnе
  svetli 2.145235 2.709770
  temni  1.402654 1.771773
> 1.96^2
[1] 3.8416
```

Test χ^2

```
> alpha <- 0.05
> (df <- (ncol(f) - 1) * (nrow(f) - 1))
[1] 1
> (X2.krit <- qchisq(1 - alpha, df))
[1] 3.841459
> (X2 <- sum((f - e)^2/e))
[1] 8.029432
> (p <- 1 - pchisq(X2, df))
[1] 0.004602328
> if (X2 < X2.krit) cat("Spremenljivki NISTA odvisni (p =",
+     round(p, 4), ").\n") else cat("Spremenljivki STA odvisni (p <",
+     alpha, ") (p =", round(p, 4), ").\n")
Spremenljivki STA odvisni (p < 0.05 ) (p = 0.0046 ).
```

Funkcija za test v R

```
> chisq.test(lasje, oci, correct = FALSE)

        Pearson's Chi-squared test

data:  lasje and oci
X-squared = 8.0294, df = 1, p-value =
0.004602
```

5.1 Asociacija (2×2)

Barva las in oči

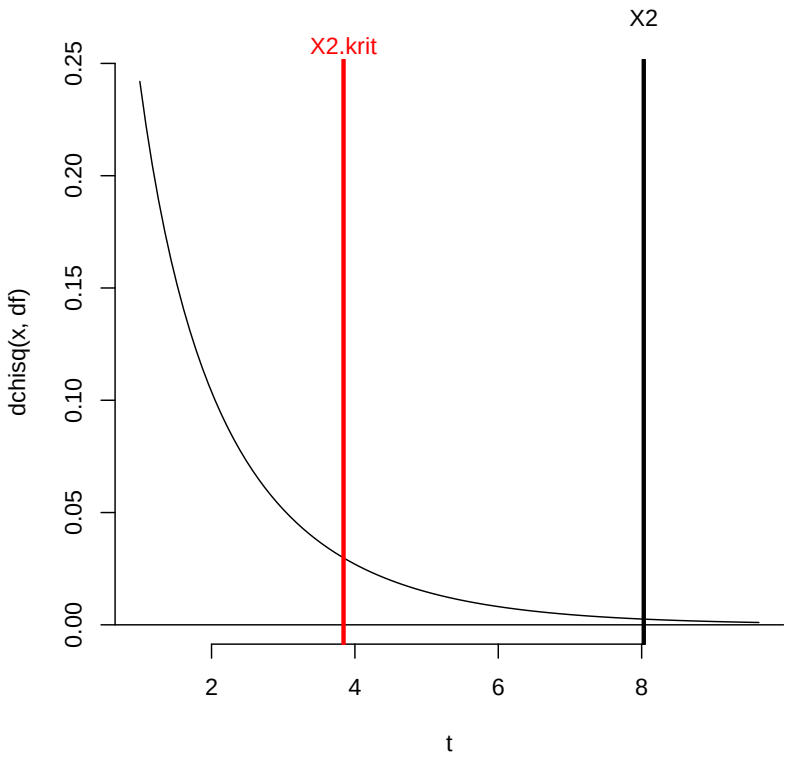
```
        oci
lasje    svetle temne Sum
  svetli     14     3  17
  temni      10    16  26
  Sum        24    19  43
```

Asociacija (2×2)

$$\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

```
> sum(f) * det(matrix(f, 2, 2))^2/prod(colSums(f),
+     rowSums(f))
[1] 8.029432
```

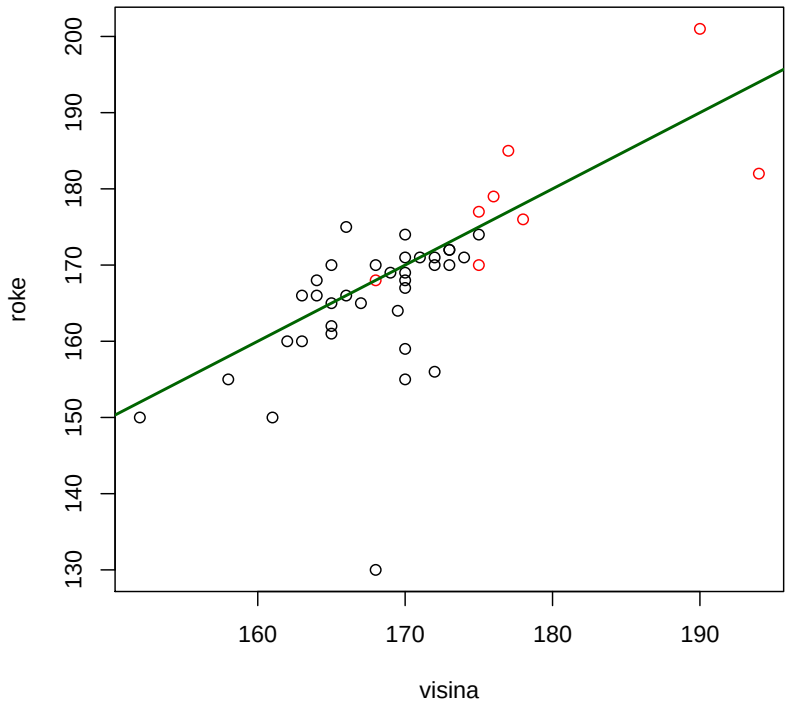
Slika



6 Korelacija in regresija

Slika Lenoarda da Vincija 'Virtruvian man' prikazuje razmerja različnih delov telesa. Višina človeka in razpon rok naj bi bila približno enaka. Če trditev drži, bi morale točke določene z višino (x) in razponom rok (y) ležati na premici $y = x$.

Virtruvian man: Roke = Višina



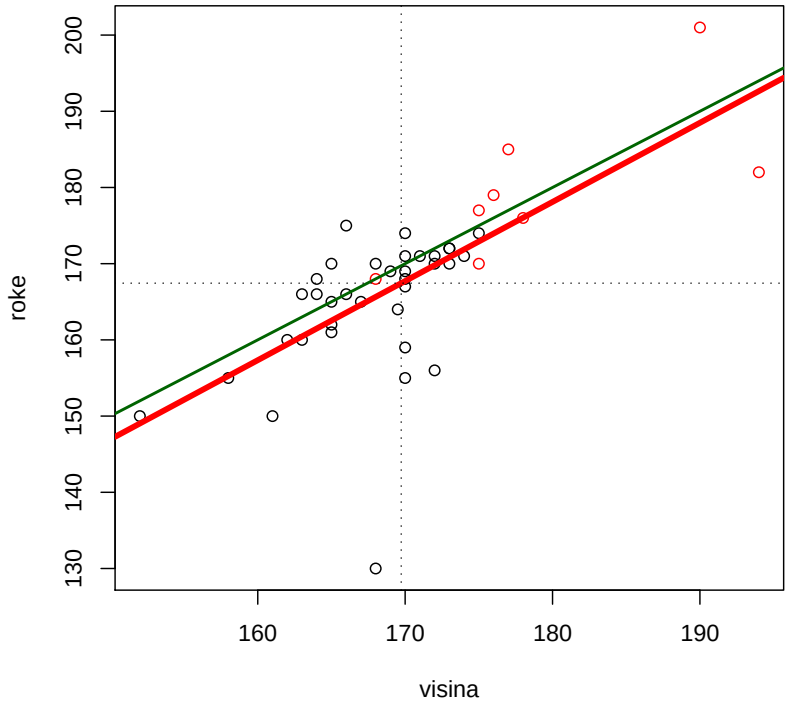
Spremenljivke

```
> select <- roke > 0
> x <- visina[select]
> y <- roke[select]
> spol <- spol[select]
> n <- length(x)
```

Osnovni račun

```
> (xbar <- mean(x))
[1] 169.7326
> (ybar <- mean(y))
[1] 167.4419
> sum((x - xbar) * (y - ybar))/(n - 1)
[1] 54.1567
> cov(x, y)
[1] 54.1567
> (r <- cov(x, y)/(sd(x) * sd(y)))
[1] 0.6906298
> r^2
[1] 0.4769695
> (b <- cov(x, y)/var(x))
[1] 1.038539
> (a <- mean(y) - b * mean(x))
[1] -8.832009
```

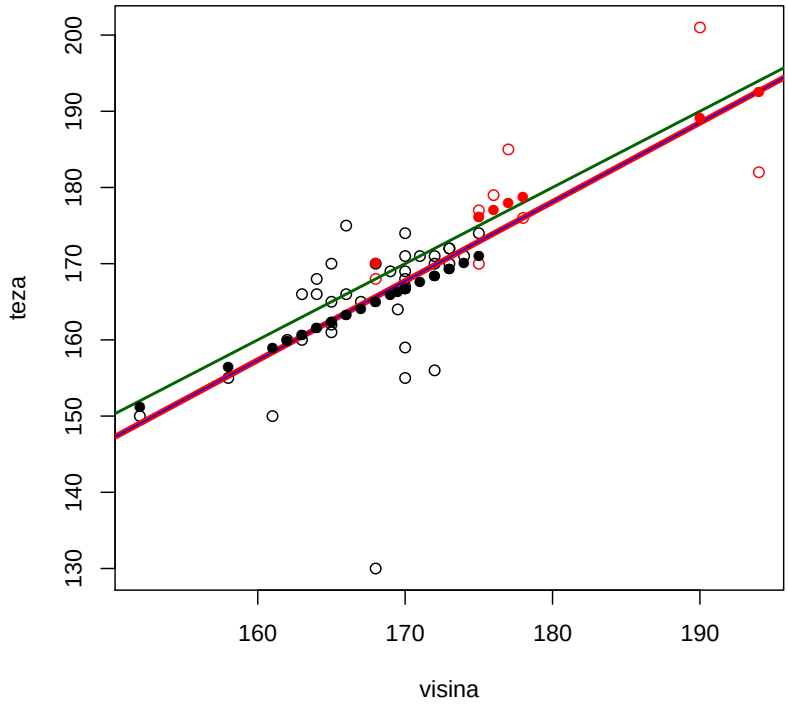
Virtruvian man: Roke = Višina



Regresija v R

```
> cor(x, y)
[1] 0.6906298

> lsfit(x, y)$coefficients

Intercept         X
-8.832009  1.038539

> lm(y ~ x)$coefficients

(Intercept)           x
  -8.832009    1.038539

> lm(y ~ x * spol)

Call:
lm(formula = y ~ x * spol)

Coefficients:
(Intercept)             x    spolmoski
  20.648518     0.859143     4.642732
x:spolmoski
   0.003153
```

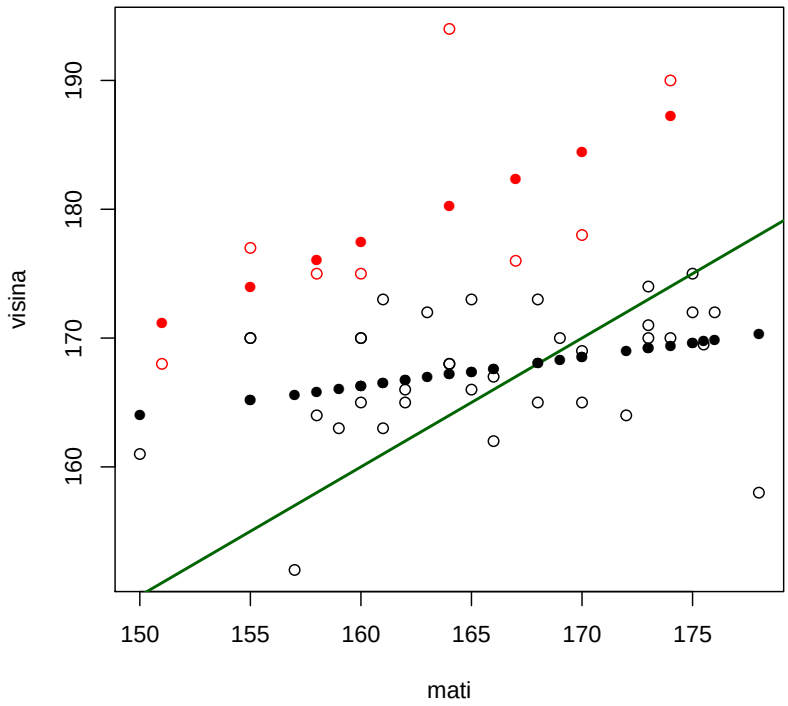
Virtruvian man: Roke = Višina



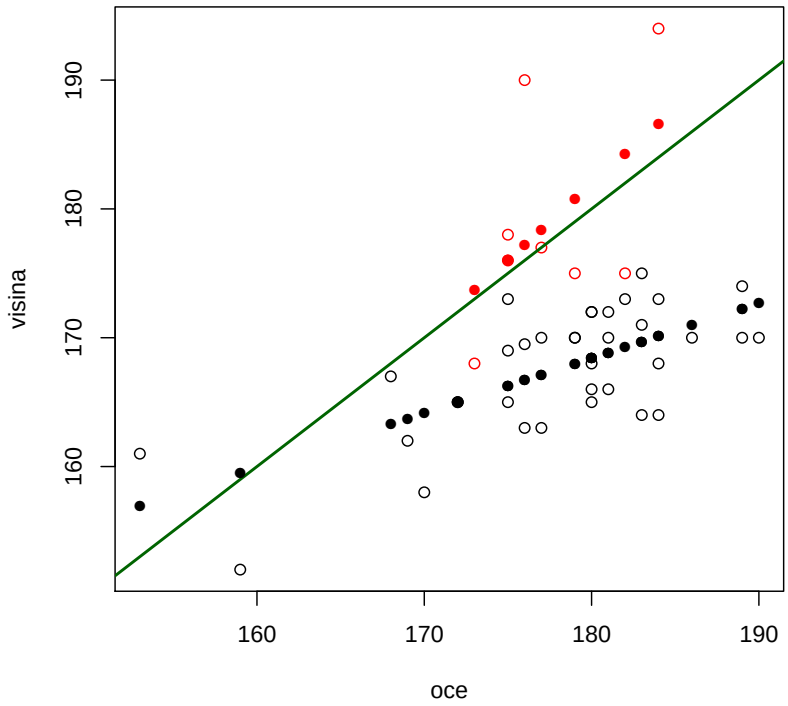
Analiza variance za linearni model

```
> anova(lm(roke ~ visina * spol))
Analysis of Variance Table

Response: roke
            Df Sum Sq Mean Sq   F value    Pr(>F)
visina       1 2362.2  2362.2   37.0845 3.887e-07
spol         1  106.1   106.1    1.6657    0.2044
visina:spol  1    0.0     0.0 4.902e-05    0.9944
Residuals   39 2484.3    63.7

visina      ***
spol
visina:spol
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

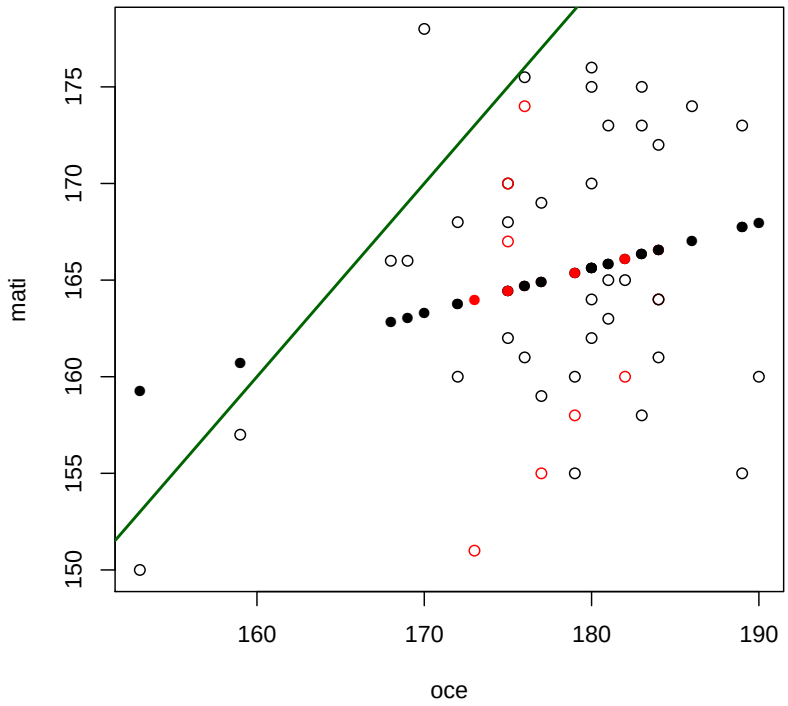
Primerjava z višino matere



Primerjava z višino očetov



Primerjava višin očetov in mam



Literatura