Kviz2

Miha Prajs

Priprava podatkov

```
my.data <- read.table("C:/Users/mihap/Code/Faks/famnit24-statistika/HW/Kviz2/data.txt",
                      header=TRUE,
                      stringsAsFactors=TRUE, sep="",
                      na.strings="NA",
                      dec=".",
                      strip.white=TRUE)
x <- 4
z <- 6
my.data <- my.data[unique(c(seq(x,
                                nrow(my.data)
                                ,by=10),
                            seq(z,
                                nrow(my.data)
                                ,by=10))),]
```

1. naloga

```
height_sample <- my.data$Height
t.test(height_sample, mu = 170)
```

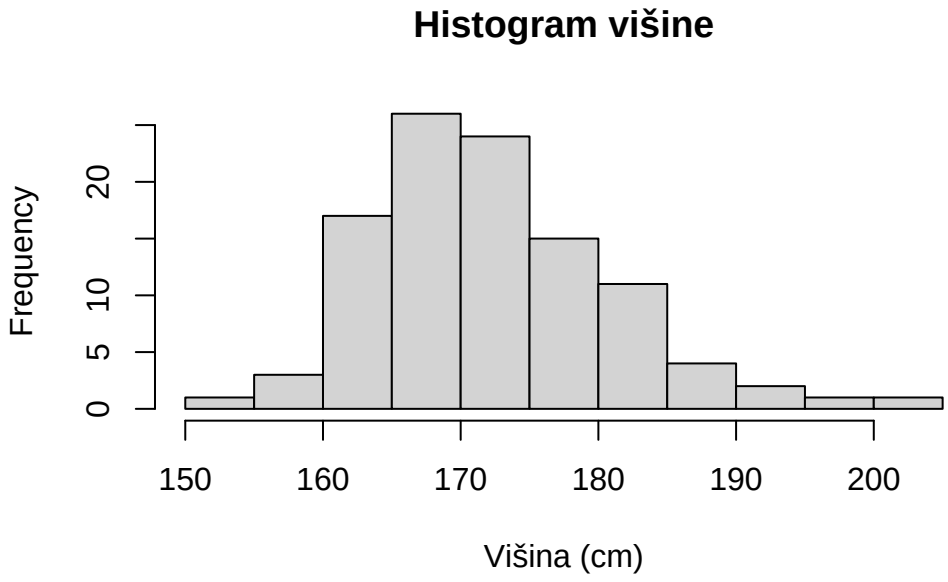
```
    One Sample t-test

data:  height_sample
t = 3.5503, df = 104, p-value = 0.0005795
alternative hypothesis: true mean is not equal to 170
```

```
95 percent confidence interval:
 171.3369 174.7202
sample estimates:
mean of x
 173.0286
```

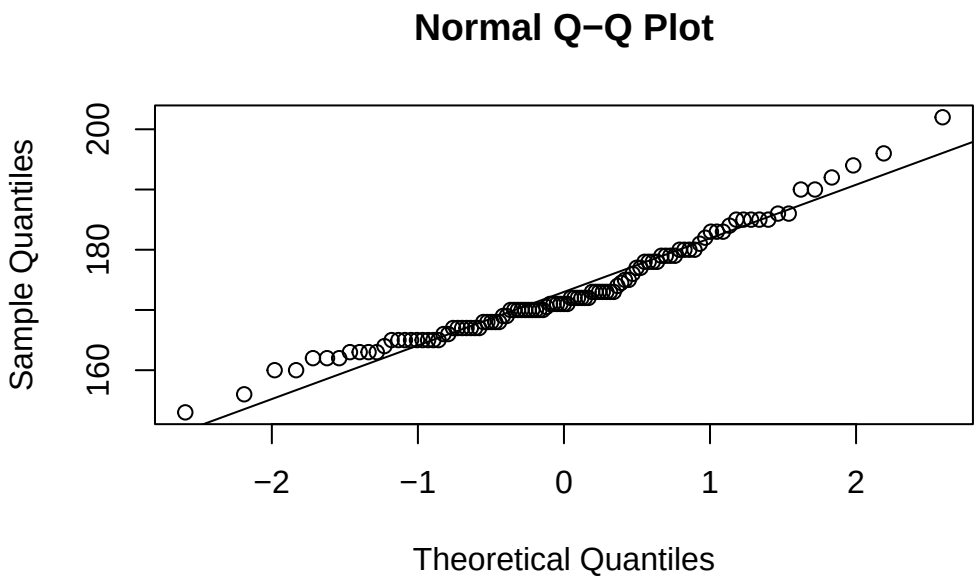
T-test pokaže, da je povprečna višina statistično značilno **različna** od 170 cm pri stopnji značilnosti = 0,05. (p-vrednost $< 0{,}05$), zato zavrnemo ničelno hipotezo in uporabimo alternativno.

2. naloga

```
hist(height_sample, main = "Histogram višine", xlab = "Višina (cm)", breaks = 10)
```



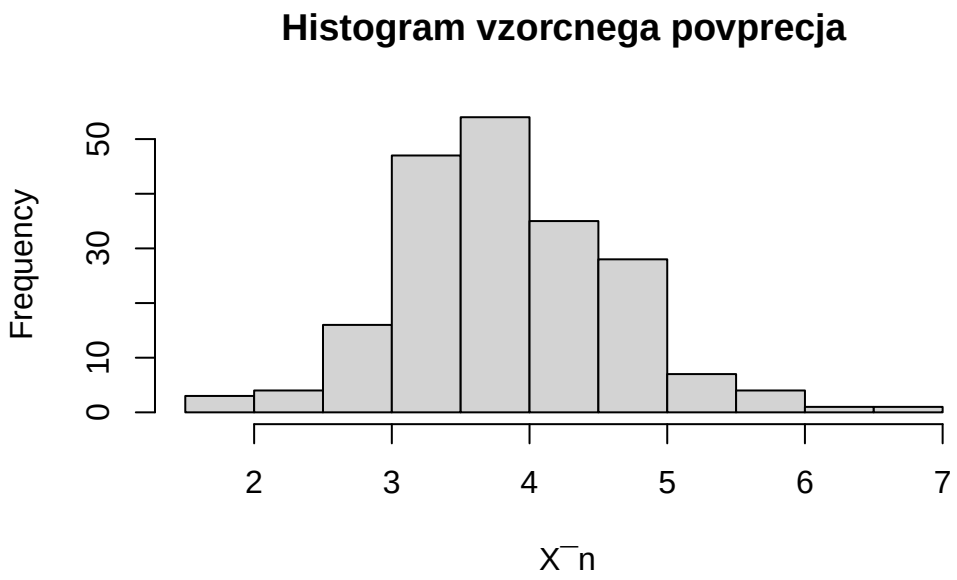

```
qqnorm(height_sample)
qqline(height_sample)
```



Histogram in Q-Q graf kažeta, da so podatki **približno normalno porazdeljeni**, kar omogoča uporabo parametričnih testov.

3. naloga

```r
n <- 30
samples <- replicate(200, mean(rgeom(n, prob = 0.2)))
hist(samples, main = "Histogram vzorčnega povprečja", xlab = "X̄n", breaks = 15)
```




```
E_Xn <- mean(samples)
Var_Xn <- var(samples)

E_Xn
```

```
[1] 3.869833
```

```
Var_Xn
```

```
[1] 0.6695544
```

Po centralnem limitnem izreku ima histogram vzorčnega povprečja obliko normalne porazdelitve.

4. naloga

```
set.seed(456)
pre_treatment <- rnorm(50, mean = 150, sd = 15)
post_treatment <- rnorm(50, mean = 140, sd = 15)

t.test(pre_treatment, post_treatment, paired = TRUE, conf.level = 0.99)
```

```
    Paired t-test

data:  pre_treatment and post_treatment
t = 3.6861, df = 49, p-value = 0.0005694
alternative hypothesis: true mean difference is not equal to 0
99 percent confidence interval:
  2.949181 18.659305
sample estimates:
mean difference
       10.80424
```

```r
library(pwr)
effect_size <- (mean(pre_treatment)-mean(post_treatment))/sd(c(pre_treatment,
                                                               post_treatment))
pwr.t.test(d = effect_size, n = 50, sig.level = 0.01, type = "paired")
```

```
     Paired t test power calculation

              n = 50
              d = 0.6765642
      sig.level = 0.01
          power = 0.9795004
    alternative = two.sided

NOTE: n is number of *pairs*
```

Statistični test kaže, da je razlika med povprečnimi vrednostmi **statistično značilna** (p-vrednost $< 0{,}01$). Moč testa je zelo dobra (power > 0.97).