1. domača naloga

Statistika 2 in Matematična statistika

Zala Jamšek, 27122040

28. avgust 2014

1. naloga

Za analizo v nalogi sem si izbrala podatke iz datoteke 2010.dohodek.eur. Odvisna spremenljivka v teh podatkih je minimalni dohodek v EUR/mesec, neodvisna spremenljivka pa je BDP na prebivalca v EUR.

```
> #prebrani podatki iz .txt datoteke
> setwd('/Users/zala/Documents/Statistics_2/naloga1/')
> data_BDP <- read.table('2010.dohodek.eur.txt', header=TRUE, sep = ' ')
> data_BDP[1:4,]
```

```
          dohodek BDP.eur
Avstrija       NA   31100
Belgija   1387.50   29200
Bolgarija  122.71   10700
Ciper          NA   23600
```

Ker je nekaj nepopolnih podatkov, sem jih odstranila.

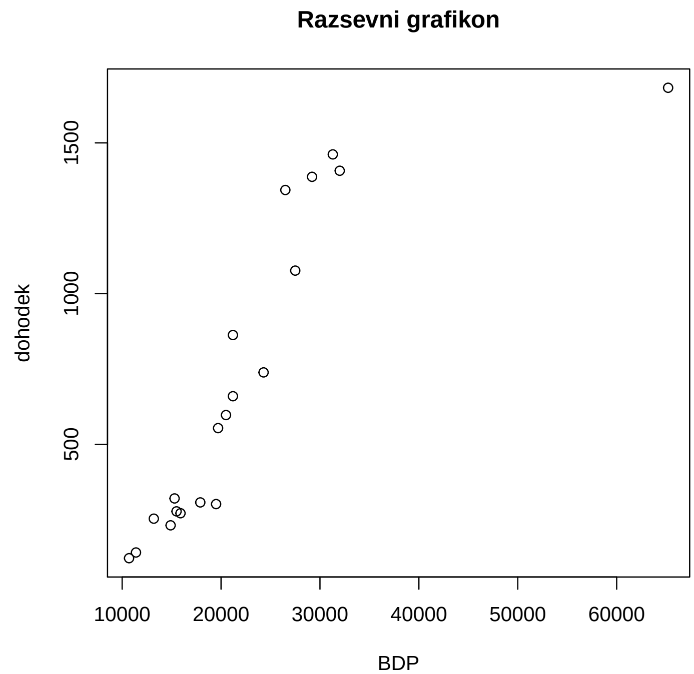
```
> #ostranjeni nepopolni podatki
> data_BDP <- na.omit(data_BDP)
> summary(data_BDP)
```

```
    dohodek          BDP.eur
 Min.   : 122.7   Min.   :10700
 1st Qu.: 276.5   1st Qu.:15450
 Median : 575.8   Median :20100
 Mean   : 700.2   Mean   :22645
 3rd Qu.:1143.3   3rd Qu.:26750
 Max.   :1682.8   Max.   :65200
```

1. **Analiza linearne povezanosti**

Razsevni grafikon ima na osi x neodvisno spremenljivko BDP na prebivalca v EUR in na y osi odvisno spremenljivko minimalni dohodek v EUR.

```
> #podatki kot vektor
> BDP <- data_BDP$BDP.eur
> dohodek <- data_BDP$dohodek
> #RAZSEVNI DIAGRAM
> plot(x=BDP, y=dohodek, main = 'Razsevni grafikon', xlab='BDP', ylab='dohodek')
```



Na grafikonu vsaka točka predstavlja eno državo. Ena država močno izstopa predvsem po velikosti BDP na prebivalca in ima poleg tega tudi največji dohodek. Sicer pa je iz grafikona razvidno, da imajo države z višjim BDP na prebivalca tudi višji dohodek. Poleg tega izgleda, da točke ležijo skoraj na premici. Bolj natančna mera za povezanost med spremenljivkama pa je korelacijski koeficient.

Pearsonov koeficient korelacije meri korelacijo (linearno povezanost) med dvema spremenljivkama in je izračunan kot razmerje med kovarianco in standardnima odklonoma:

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

$$cov(x,y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_i)(y_i - \mu_i)$$

Funkcija za izračun koeficienta je že vgrajena v R.

```
> #Pearsonov koeficient
> Pearson<- function(d){
+   pearson <-cor(d, method = 'pearson')
+   return(pearson[2,1])
+ }
> Pearson(data_BDP)
```

```
[1] 0.8388114
```

Pearsonov korelacijski koeficient je 0.8388114. Kot sem opazila že iz same oblike razsevnega grafikona je korelacija med spremenljivkama pozitivna. Pearsonov koeficient je blizu 1, kar pomeni, da je korelacija med spremenljvkama skoraj linearna, kar je videti tudi iz zame oblike grafikona, saj točke ležijo skoraj na premici.

2. Linearna regresijska analiza

Za regresijsko premico želim določiti a in b, tako da bo

$$dohodek = a + b \cdot BDP$$

Parametra a in b sta dobljena z metodo najmanjših kvadratov. V R-u se koeficienta dobi s funkcijo lm. Funkcijo $koef\_lr$, ki vrne parametra a in b sem ustvarila, ker jo bom uporabila v nadaljevanju naloge.

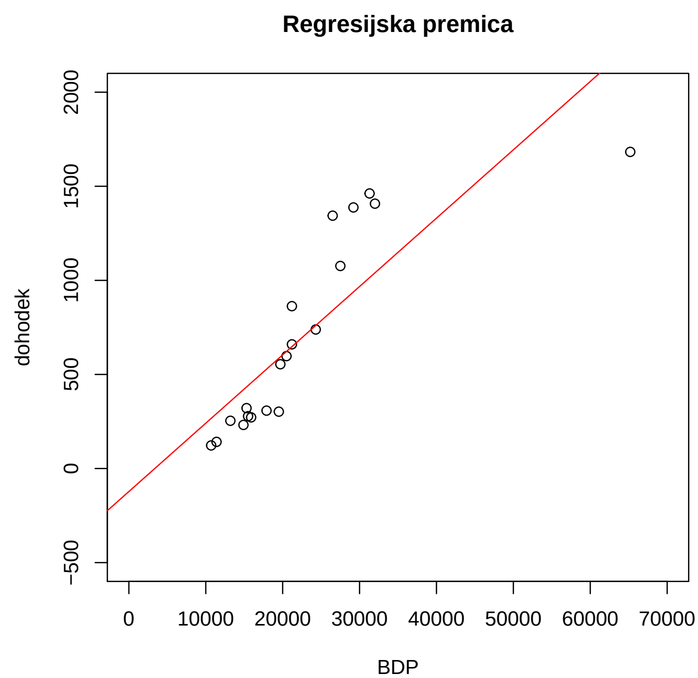
```
> #funkcija, ki vrne koeficiente linearne regresije
> koef_lr <- function(d){
+   lr <- lm(d$dohodek~d$BDP)
+   a <-lr$coefficients[1]
+   b <-lr$coefficients[2]
+   return(c(a,b))
+ }
> l_reg <- lm(dohodek~BDP)
> summary(l_reg)

Call:
lm(formula = dohodek ~ BDP)

Residuals:
    Min      1Q  Median      3Q     Max
-563.10 -167.82  -71.19  203.74  503.57

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.223e+02  1.413e+02  -0.866    0.398
BDP          3.632e-02  5.557e-03   6.537 3.83e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 287.5 on 18 degrees of freedom
Multiple R-squared:  0.7036,	Adjusted R-squared:  0.6871
F-statistic: 42.73 on 1 and 18 DF,  p-value: 3.826e-06
```

Dobimo rezultata: $a = -122,336$, $b = 0,0362$, in koeficient b je statistično značilno različen od 0. Torej je regresijska premica obike:

$$dohodek = -122,336 + 0,0362 \cdot BDP$$

To pomeni, da če se BDP na prebivalca na prebivalca poveča za eno 1 EUR, se minimalni dohodek v državi poveča za $0,0362$ EUR. Konstanta a pa pomeni, da če je BDP na prebivalca enak 0, je minimalni dohodek v državi negativen in enak $-122,336$. V razsevni grafikon je vrisana regresijska premica.

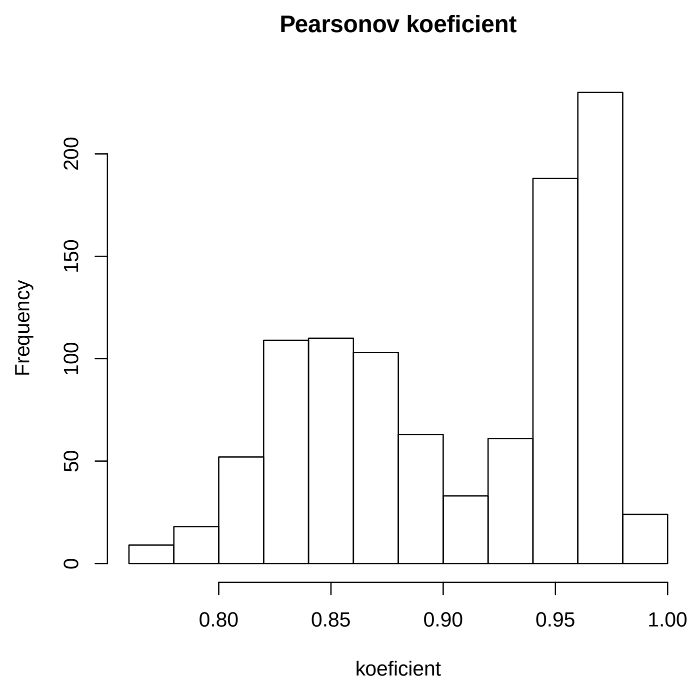
```
> #linearna regresija
> plot(x=BDP, y=dohodek,
+       xlim = c(0,70000), ylim=c(-500,2000), main = 'Regresijska premica')
> abline(l_reg, col=2)
```



3. Simulacija: Pearsonov koeficient

Za simulacijo je v vsakem od 1000 korakov naključno izberan vzorec velikosti 10 in na tem vzorcu je izračunan Pearsonov koeficient korelacije. Koeficienti so predstavljeni v histogramu.

```
> #Simulacija za Pearsonov koeficient
> pearson_vector <- numeric(0)
> for (i in 1:1000){
+   data_BDP_pearson <- data_BDP[sample(nrow(data_BDP),10, replace=FALSE),]
+   pearson_vector[i]<-Pearson(data_BDP_pearson)
+ }
> hist(pearson_vector, main = 'Pearsonov koeficient', xlab='koeficient')
```



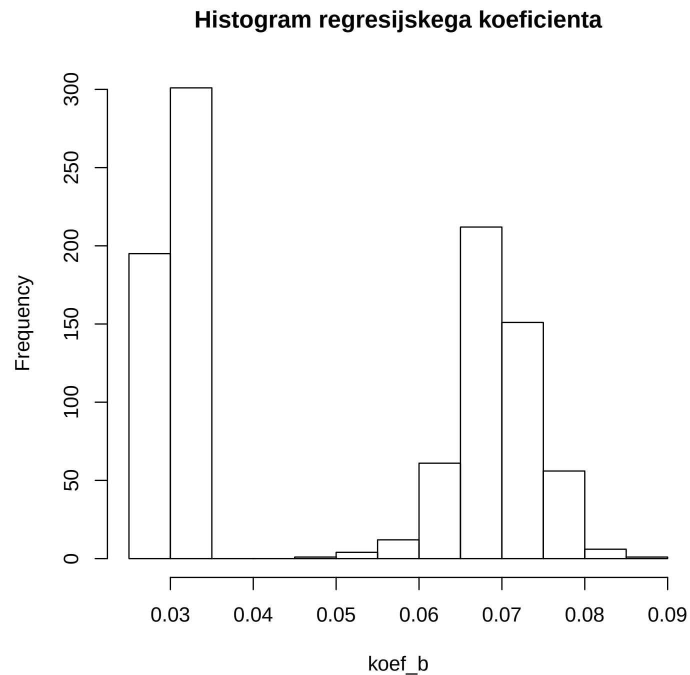
Povprečje vzočne porazdelitve in standardni odkolon:

```
> mean(pearson_vector)

[1] 0.9053758

> sd(pearson_vector)

[1] 0.05940352
```

4. Simulacija: Regresijski koeficient

V vsakem koraku simulacije je naključno izberan vzorec velikosti 10 in na tem vzorcu izračunan regresijski koeficient. Regresijski koeficienti so predstavljeni v histogramu.

```
> #Simulacija za regresijski koeficient
> koef_b <- numeric(0)
> for (i in 1:1000){
+   data_BDP_koef <- data_BDP[sample(nrow(data_BDP),10),]
+   koef_b[i] <- koef_lr(data_BDP_koef)[2]
+ }
> hist(koef_b, main='Histogram regresijskega koeficienta')
```



Na histogramu je opaziti dvojni vrh. Pri v drugem stolpcu (vrednosti 0,03-0,035) in drugi vrh v devetem okencu (vrednosti 0,065/0,07). Prvi vrh je višji od drugega. Regresijski koeficient izračunan na celotni populaciji je tudi v mejah prvega vrha.

Verjetno do drugega vrha pride zaradi naklјčno izbranega vzorca 10 držav in v primeru, ko je v vzorcu zajet tudi najbolj viden 'outlier', to zniža regresijski koeficient, v primeru ko pa ta država ni zajeta, je koeficient višji, zato se pojavi tudi drugi vrh.

5. Permutacijski test

Na vzorcu 10 naključno izbranih držav je izračuna Pearsonov koeficient

```
> perm_data <- data_BDP[sample(nrow(data_BDP),10, replace = FALSE),]
> (t_star <- Pearson(perm_data))
```

```
[1] 0.950488
```

Test hipoteze:

$$H_0 : \text{Korelacijski koeficient} = 0$$

$$H_1 : \text{Korelacijski koeficient} > 0$$

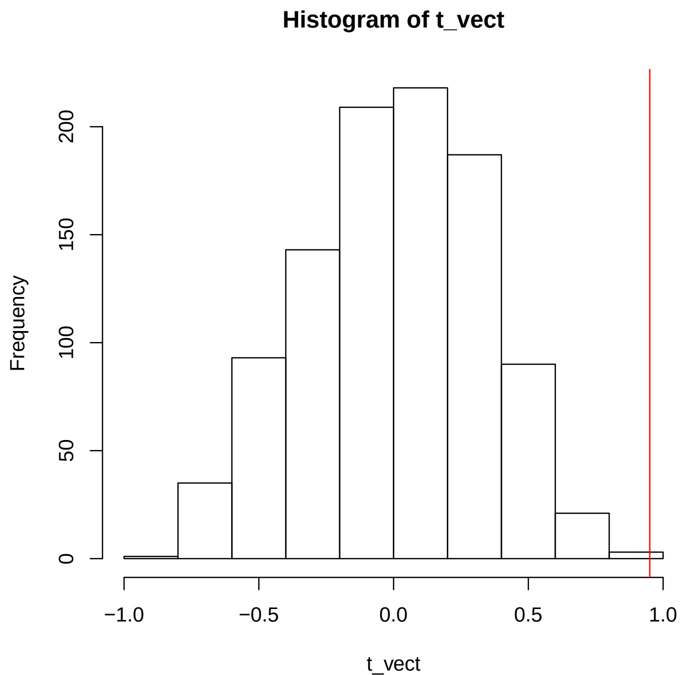
Permutacijski test je narejen z Monte Carlo simulacijo (Referenca: $http://en.wikipedia.org/wiki/Pearson\_product-moment\_correlation\_coefficient\#Use\_a\_permutation\_test$).

Osnovni vzorec vsebuje podatke 10 držav, in torej 10 parov točk (x_i, y_i). V simulaciji se v vsakem od B korakov generira nov vzorec naključnih parov točk (x_i, y_k) za $i \in \{1, ..10\}, k \in \{1, ..10\}$. V vsakem koraku se za nov vzorec izračuna Pearskonov koeficient korelacije. P- vrednost testa je izračunana tako, da

$$p = \frac{\sum_{i=1}^{B} I_{t_i > t^*}}{B}$$

kjer je t_i vrednost Pearsonovega koeficienta izračunana na vsakem koraku, t^* pa vrednost Pearsonovega koeficienta izračunana na osnovnem vzorcu.

```
> #PERMUTACIJSKI TEST
> #Monte Carlo Simulacija za Permutacijski test
>
> permutation_test <- function(data, B){
+   t_vect <- numeric(0)
+   BDP <- data$BDP
+   dohodek <- data$dohodek
+
+   #MC
+   for (i in 1:B){
+      X<-data.frame(sample(BDP, replace = FALSE),sample(dohodek, replace = FALSE))
+      t_vect[i] <- Pearson(X)
+   }
+   hist (t_vect)
+   abline(v= t_star, col = 2 )
+   p <- (sum(t_vect > t_star))/B
+   return(p)
+ }
> permutation_test(perm_data,1000)

[1] 0

>
```



Iz histograma je vidno, da so vsi izračunani koeficienti manjši od vrednosti koeficienta začetnega vzorca. Izračunana p-vrednost permutacijskega testa je enaka 0, torej lahko pri stopnji značilnosti 0 zavrnemo ničelno hipotezo, da je korelacijski koeficient enak 0.

2. naloga

1. Metoda največjega verjetja

Izračun parametrov pareto porazdelitve z metodo največjega verjetja:

$$f(x_i) = \frac{\alpha\lambda^\alpha}{x_i^{\alpha+1}}$$

$$L = \prod_{i=1}^{n} \frac{\alpha\lambda^\alpha}{x_i^{\alpha+1}} = \alpha^n \lambda^{\alpha n} \prod_{i=1}^{n} x_i^{-(\alpha+1)}$$

$$l = log(L) = nlog(\alpha) + n\alpha log(\lambda) - (\alpha+1)\sum_{i=1}^{n} log(x_i)$$

$$\frac{\partial u}{\partial \lambda} = \frac{n\alpha}{\lambda} = 0$$

Za parameter λ iskanje maksimuma z odvajanjem ni ustrezna metoda. Iz definicije Pareto porazdelitve pa je znano, da je $\lambda \leq x_i$ za vsak i. Torej je lahko parameter ocenjen iz tega pogoja:

$$\lambda \leq x_i \Rightarrow \hat{\lambda} = min(x_i)$$

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + nlog(\lambda) - \sum_{i=1}^{n} log(x_i) = 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} log(x_i) - log(\lambda)^n$$

$$\frac{\alpha}{n} = \frac{1}{\sum_{i=1}^{n} log(\frac{x_i}{\lambda})}$$

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log(\frac{x_i}{\hat{\lambda}})}$$

2. Fisherjeva matrika informacije

Za Fisherjevo informacijsko matriko sem najprej izračunala druge odvode log porazdelitvene funkcije:

$$log(f(x)) = log(\alpha) + \alpha log(\lambda) - (\alpha + 1)log(x)$$

$$\frac{\partial log(f(x))}{\partial \alpha} = \frac{1}{\alpha} + log(\lambda) - log(x)$$

$$\frac{\partial log(f(x))}{\partial \lambda} = \frac{\alpha}{\lambda}$$

$$\frac{\partial^2 log(f(x))}{\partial \alpha^2} = -\frac{1}{\alpha^2}$$

$$\frac{\partial^2 log(f(x))}{\partial \lambda^2} = -\frac{\alpha}{\lambda^2}$$

$$\frac{\partial^2 log(f(x))}{\partial \alpha \partial \lambda} = \frac{1}{\lambda}$$

Fisherjeva informacijska matrika je:

$$I(\theta) = -E[\frac{\partial^2}{\partial \theta^2} log f(x;\theta)|\theta]$$

$$I(\theta) = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{bmatrix}$$

3. Metoda momentov

Za metodo momentov uporabimo:

$$\mu_1 = E(X)$$

$$\mu_2 = E(X^2)$$

Momenti Pareto porazdelitve so:

$$\begin{aligned}
E(X) &= \int_\lambda^\infty x f(x) dx = \int_\lambda^\infty x \frac{\alpha\lambda^\alpha}{x^{\alpha+1}} dx \\
&= \alpha\lambda^\alpha \int_\lambda^\infty \frac{1}{x^\alpha} dx \\
&= \frac{\alpha\lambda^\alpha}{-(\alpha-1)} \frac{1}{x^{\alpha-1}} \Big|_\lambda^\infty \\
&= \frac{\alpha\lambda^\alpha}{-(\alpha-1)} \left(-\frac{1}{\lambda^{\alpha-1}}\right) \\
&= \frac{\alpha\lambda}{\alpha-1}
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= \int_\lambda^\infty x^2 f(x) dx = \int_\lambda^\infty x^2 \frac{\alpha\lambda^\alpha}{x^{\alpha+1}} dx \\
&= \alpha\lambda^\alpha \int_\lambda^\infty \frac{1}{x^{\alpha-1}} dx \\
&= \frac{\alpha\lambda^\alpha}{-(\alpha-2)} \frac{1}{x^{\alpha-2}} \Big|_\lambda^\infty \\
&= \frac{\alpha\lambda^\alpha}{-(\alpha-2)} \left(-\frac{1}{\lambda^{\alpha-2}}\right) \\
&= \frac{\alpha\lambda^2}{\alpha-2}
\end{aligned}$$

Potreben pogoj za obstoj momentov je torej, da je $\alpha > 2$. V obeh momentih nastopata oba parametra, zato z deljenjem eliminiramo λ

$$\begin{aligned}
\frac{\mu_1^2}{\mu_2} &= \frac{\alpha^2\lambda^2}{(\alpha-1)^2} \frac{\alpha-2}{\alpha\lambda^2} \\
&= \frac{\alpha(\alpha-2)}{(\alpha-1)^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\mu_1^2}{\mu_2} - 1 &= \frac{\alpha^2 - 2\alpha - \alpha^2 + 2\alpha - 1}{\alpha^2 - 2\alpha + 1} \\
(\alpha-1)^2 &= \frac{\mu_2}{\mu_2 - \mu_1^2} \\
\hat{\alpha} &= 1 + \sqrt{\frac{\mu_2}{\mu_2 - \mu_1^2}} \\
&= 1 + \sqrt{1 + \frac{\mu_1^2}{\mu_2 - \mu_1^2}}
\end{aligned}$$

Ker mora biti parameter α pozitiven, je upoštevan pozitivni koren.

$$\begin{aligned}\mu_1 &= \frac{\alpha\lambda}{\alpha-1}\\ \lambda &= \frac{\mu_1(\alpha-1)}{\alpha}\\ \hat{\lambda} &= \frac{\mu_1(\hat{\alpha}-1)}{\hat{\alpha}}\end{aligned}$$

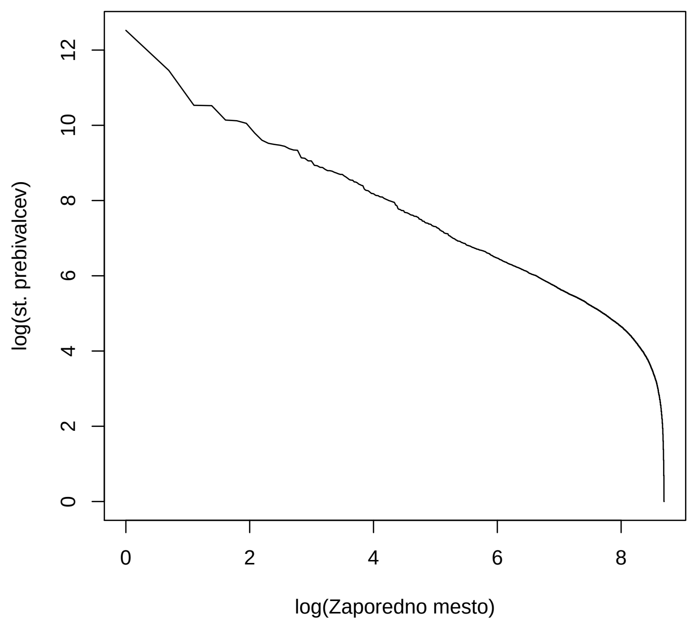
4. log-log graf

Iz datoteke podatki.nasleja.txt so prebrani podatki, urejeni po velikosti in na novo izbrani samo tisti podatki, ki imajo več ali enako kot enega prebivalca.

```
> data_naselja <- read.table('podatki.naselja.txt', sep='\t', header=TRUE)
> data_naselja[1:4,]

     naselje stevilo.prebivalcev
1 Ajdovscina                6597
2     Batuje                 339
3       Bela                  37
4       Brje                 390

> #urejeni podatki po padajocem st. prebivalcev
> data_naselja<-data_naselja[order(data_naselja$stevilo.prebivalcev, decreasing=TRUE),]
> sum(data_naselja$stevilo.prebivalcev>=1)

[1] 5973

> #novi podatki samo do 5973. nasleja, ki imajo st. prebivalcev vecje od 1
> new <- 1:sum(data_naselja$stevilo.prebivalcev>=1)
> data_naselja_new <- data_naselja[new,]
> data_naselja_new[1:4,]

       naselje stevilo.prebivalcev
2307 Ljubljana              274826
2556   Maribor               94809
297      Celje               37490
1832     Kranj               37151
```

Log log graf:

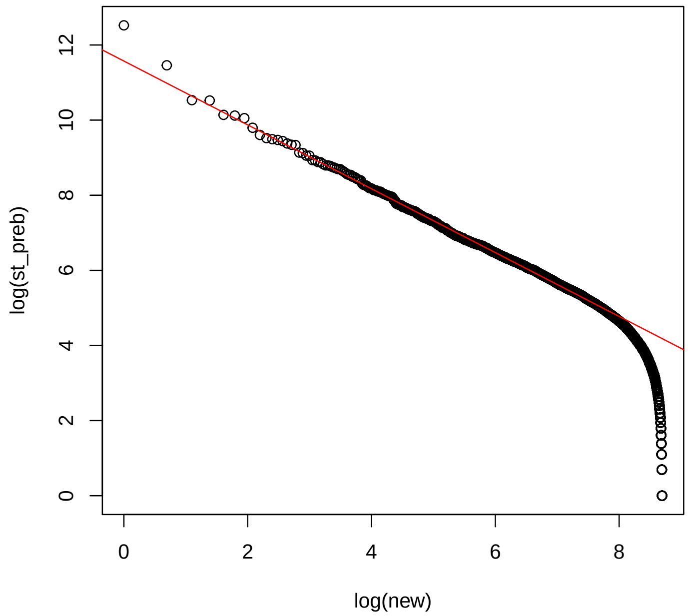
```
> st_preb <- data_naselja_new$stevilo.prebivalcev
> plot(log(new),y=log(st_preb), type='l',
+      xlab='log(Zaporedno mesto)', ylab='log(st. prebivalcev)', )
```



Opaziti je, da je na začetku log-log graf skoraj premica, kar je značilno za Pareto porazdelitev. V zadnjem delu pa točke ne ležijo več na premici.

Regresijska premica za prvih 750 naselij:

```
> n1<-750
> x<-1:n1
> st_preb_750<-(st_preb[x])
> st_preb_reg <- lm(log(st_preb_750)~log(x))
> plot(log(new),y=log(st_preb))
> abline(st_preb_reg, col=2)
```



5. Ocena parametrov: Metoda največjega verjetja

```
> l_mle <- function(data){return(min (data))}
> l1 <- l_mle(st_preb_750)
> l1

[1] 406

> alpha_mle <- function(data,n){n/sum(log(data/l_mle(data)))}
> a1 <- alpha_mle (st_preb_750,n1)
> a1

[1] 1.272517
```

Parametra po metodi največjega verjetja za prvih 750 opazovanj sta:

$$\hat{\alpha} = 1,272517$$

$$\hat{\lambda} = 406$$

6. Ocena parametrov: Metoda momentov

V nalog 3. so izraženi parametri po metodi momentov. Za izračun teh parametrov pa je treba upoštevti še naslednje:

$$\mu_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

$$Var(X) = \mu_2 - \mu_1^2 = \frac{n-1}{n}S^2$$

Potem velja:

$$\hat{\alpha} = 1 + \sqrt{1 - \frac{n}{n-1}\frac{\bar{X}}{S^2}}$$

```
> X_bar <- mean(st_preb_750)
> S <- sd(st_preb_750)
> alpha_mm <- 1+ sqrt(1 + (n1/(n1-1))*(X_bar^2)/S^2)
> alpha_mm
```

```
[1] 2.015846
```

```
> lambda_mm <- X_bar*(alpha_mm-1)/alpha_mm
> lambda_mm
```

```
[1] 986.6094
```

Ocenjena parametra sta: $\hat{\alpha} = 2,015846, \hat{\lambda} = 986,6094$. Pogoj, da je $\alpha > 2$, je izpolnjen, pojavi pa se problem, saj ni izpolnjen pogoj $\lambda > x_i$, saj je $min(x_i) = 406$. Metoda momentov torej v tem primeru ni primerna metoda za ocenjevanje parametrov.

7. Ocena parametrov na vzorcu

Novi podatki z vzorcem 100 naselij in na teh podatki izračunana parametra po metodi največjega verjetja:

```
> #7
> data_naselja_100 <- read.table('podatki.naselja.vzorec.txt', header = TRUE, sep = '\t')
> n2 <-100
> #data_naselja_vz
> st_preb_100 <- data_naselja_100$stevilo.prebivalcev
> l_100<-l_mle(st_preb_100 )
> a_100 <-alpha_mle(st_preb_100,n2)
> l_100
```

```
[1] 406
```

```
> a_100
```

```
[1] 1.391795
```

$$\hat{\alpha} = 1,391795$$
$$\hat{\lambda} = 406$$

8. Standardna napaka za α

V točki 3. sem izračunala Fisherjevo informacijo za parameter α:

$$I(\alpha) = \frac{1}{\alpha^2}$$

Od tod sledi, da je standardna napaka za parameter α:

$$Var(\alpha) = \frac{1}{nI(\alpha)}$$
$$Var(\alpha) = \frac{\alpha^2}{n}$$
$$sd(\alpha) = \sqrt{\frac{\alpha^2}{n}}$$

```
> sd_mle <- function(data, n){
+   return(sqrt(alpha_mle(data,n)^2/n))}
> sd_1 <-sd_mle(st_preb_100, n2)
> sd_1
```

```
[1] 0.1391795
```

9. Interval zaupanja za α

Iz centralnega limitnega izreka sledi:
$\sqrt{n}(\hat{\alpha} - \alpha) \to N(0, \frac{\alpha^2}{n})$, ko $n \to \infty$.

Od tod sledi, da za velike n:

$$\sqrt{n}\frac{(\hat{\alpha} - \alpha)}{\sqrt{\frac{\alpha^2}{n}}} \sim N(0,1)$$

Interval zaupanja (stopnja značilnosti je označena z δ, da ne bi prišlo do zmede z oznakami):

$$P(-\phi(1 - \frac{\delta}{2}) \leq \frac{(\hat{\alpha} - \alpha)}{\frac{sd}{\sqrt{n}}} < \phi(1 - \frac{\delta}{2})) = 1 - \delta$$
$$P(\hat{\alpha} - \phi(1 - \frac{\delta}{2})\frac{sd}{\sqrt{n}} \leq \alpha < \hat{\alpha} + \phi(1 - \frac{\delta}{2})\frac{sd}{\sqrt{n}}) = 1 - \delta$$
$$\alpha \in [\hat{\alpha} - \phi(1 - \frac{\delta}{2})\frac{sd}{\sqrt{n}}, \hat{\alpha} + \phi(1 - \frac{\delta}{2})\frac{sd}{\sqrt{n}}]$$

```
> CI <- function(n,delta,st_dev){
+   z <- qnorm(1-delta/2)
+   lb <-alpha_mle(st_preb_100,n)- (z*st_dev/sqrt(n))
+   ub <-alpha_mle(st_preb_100,n)+ (z*st_dev/sqrt(n))
+   return(c(lb,ub))
+ }
> CI(n2,0.1, sd_1)
```

```
[1] 1.368902 1.414688
```

Dobljeni interval zaupanja je:$\alpha \in [1.368902, 1.414688)$ in izračunan parameter iz točke 5 $\hat{\alpha} = 1,272517$ ne leži v tem intervalu.

10.

Bootstrap na vzorcu 100 naselij sem naredila tako, da sem na vsakem od B korakov (B je za to nalogo 1000) generirala nov vzorec velikosti 100 s ponavljanjem (bootstrap resample). Za vsak resample je izračunan parameter $\hat{\alpha}$, ki je shranjen v vekotr $T = (T_1, T_2, ...T_B)$ in $\bar{T} = \frac{1}{B}\sum_{i=1}^{B} T_i$. Nato sem naredila histogram teh parametrov. Standardna napaka parametra je izračunana po formuli:

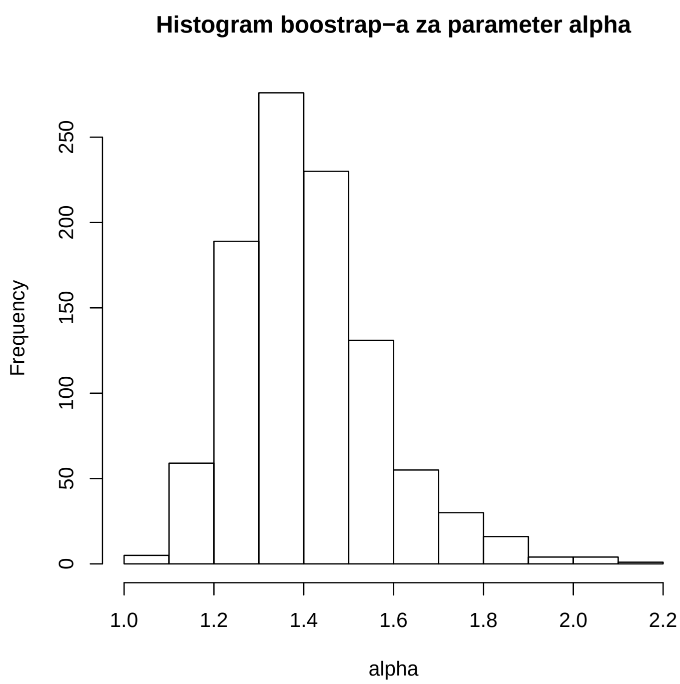
$$\hat{se}_\alpha = \sqrt{\frac{1}{B}\sum_{i=1}^{n}(T_i - \bar{T})^2}$$

Interval zaupanja pa je nato:

$$[\hat{\alpha} - \phi(1 - \frac{\alpha}{2})\hat{se}_\alpha, \hat{\alpha} + \phi(1 - \frac{\alpha}{2})\hat{se}_\alpha))$$

```
> #BOOTSTRAP
> boot <- numeric(0)
> bootstrap <-function(data, B){
+   st_preb <- data$stevilo.prebivalcev
+   n <- length(st_preb)
+   T <- numeric(0)
+   for (i in 1:B){
+     resample <- sample(st_preb,n,replace = TRUE)
+     alpha <- alpha_mle(resample,n)
+     T[i]<-alpha
+   }
+   return(T)
+ }
> #Histogram
> boot_1 <- bootstrap(data_naselja_100,1000)
> hist(boot_1, main = 'Histogram boostrap-a za parameter alpha', xlab='alpha')
> sd_boot<-function(vector){
+   B <- length(vector)
+   s <- sqrt(1/B * sum((vector - mean(vector))^2))
```

```
+   return(s)
+ }
> sum(boot_1 - mean(boot_1))/1000

[1] 8.437695e-17

> sd_2 <- sd_boot(boot_1)
> sd_2

[1] 0.1606862

> CI(n2, 0.1, sd_2)

[1] 1.365365 1.418226
```



Dobljeni interval zaupanja je v primerjavi s tistim iz točke 9. le malo širši, kar je posledica večje standardne napake v primeru, ko se ta izračunana z bootstrapom.