

Dinamički izvještaj

Usporedba konfiguracija laptopa

```
# Projektni zadatak - Banjavcic, Obadic
```

```
# install.packages("leaps", "MASS", "ggrepel", "nnls" )  
as.numeric.factor <- function(x) {as.numeric(levels(x))[x]}  
  
library(XML)  
library(readxl)  
laptops <- read_excel("laptops.xlsx")  
View(laptops)
```

```
# PRIPREMA PODATAKA -----
```

```
laptops$CPU.TYPE <- factor(laptops$CPU.TYPE, ordered = TRUE,  
                           levels = c("AMD", "Celeron", "Core i3", "Pentium", "Core i5", "Core i7"),  
                           labels = c("AMD", "Cel", "i3", "Pen", "i5", "i7"))  
laptops$CPU.TYPE.num <- laptops$CPU.TYPE  
levels(laptops$CPU.TYPE.num) <- c("1", "2", "3", "4", "5", "7")  
laptops$CPU.TYPE.num <- as.numeric.factor(laptops$CPU.TYPE.num)
```

```
laptops$RAM <- factor(laptops$RAM, ordered = TRUE,  
                     levels = c("4", "8", "12", "16"))  
laptops$RAM.num <- laptops$RAM  
levels(laptops$RAM.num) <- c("4", "8", "12", "16")  
laptops$RAM.num <- as.numeric.factor(laptops$RAM.num)
```

```
laptops$HDD <- as.character(laptops$HDD)  
laptops$HDD[is.na(laptops$HDD)] <- "0"  
laptops$HDD <- factor(laptops$HDD, ordered = TRUE,
```

```

        levels = c("0", "16", "32", "64", "128", "256", "500", "1000", "2000"))

laptops$HDD.num <- laptops$HDD

levels(laptops$HDD.num) <- c("0", "16", "32", "64", "128", "256", "500", "1000", "2000")

laptops$HDD.num <- as.numeric.factor(laptops$HDD.num)


laptops$SSD <- as.character(laptops$SSD)

laptops$SSD[is.na(laptops$SSD)] <- "0"

laptops$ssd <- factor(laptops$SSD, ordered = TRUE,
        levels = c("0", "32", "64", "128", "256", "512"))

laptops$ssd.num <- laptops$ssd

levels(laptops$ssd.num) <- c("0", "32", "64", "128", "256", "512")

laptops$ssd.num <- as.numeric.factor(laptops$ssd.num)


laptops$GPU <- factor(laptops$GPU, ordered = TRUE,
        levels = c("HD Graphics", "AMD", "nVidia", "UHD Graphics"))

laptops$GPU.num <- laptops$GPU

levels(laptops$GPU.num) <- c("0", "0.5", "1", "2")

laptops$GPU.num <- as.numeric.factor(laptops$GPU.num)


laptops$OS <- factor(laptops$OS, ordered = TRUE,
        levels = c("0", "1"),
        labels = c("FreeDOS", "OS"))

laptops$OS.num <- laptops$OS


# VIZUALIZACIJA PODATAKA -----

if (0) {

    library(ggplot2)

    library(ggrepel)


    windows()

```

```
ggplot(data=laptops, aes(x=CPU.TYPE,y=Cijena))+  
  geom_boxplot(fill="lightblue", color="black", notch = TRUE)+  
  geom_point(position = "jitter", color="blue", alpha=.5)+  
  geom_rug(sides="1", color="black")
```

```
windows()
```

```
ggplot(data=laptops, aes(x=RAM,y=Cijena))+  
  geom_boxplot(fill="lightblue",  
    color="black", notch = TRUE)+  
  geom_point(position = "jitter", color="blue", alpha=.5)+  
  geom_rug(sides="1", color="black")
```

```
windows()
```

```
ggplot(data=laptops, aes(x=HDD,y=Cijena))+  
  geom_boxplot(fill="cornflowerblue",  
    color="black", notch = TRUE)+  
  geom_point(position = "jitter", color="blue", alpha=.5)+  
  geom_rug(sides="1", color="black")
```

```
windows()
```

```
ggplot(data=laptops, aes(x=GPU,y=Cijena))+  
  geom_boxplot(fill="cornflowerblue",  
    color="black", notch = TRUE)+  
  geom_point(position = "jitter", color="blue", alpha=.5)+  
  geom_rug(sides="1", color="black")
```

```
windows()
```

```
ggplot(data=laptops,aes(x=Cijena, y=CPU.TYPE, color=SSD, shape=HDD))+  
  geom_point(position = "jitter", alpha=1.0)+  
  geom_rug(sides="1", color="black")+  
  geom_smooth(method = "lm", color = "red", linetype=2)+
```

```
facet_grid(~laptops$Cijena)
```

```
windows()
```

```
ggplot(data = laptops, aes(x=GPU, fill=Brend))+  
  geom_histogram(stat = "count")
```

```
windows()
```

```
ggplot(data=laptops,aes(x=Brend, y=GPU))+  
  geom_point(position = "jitter", alpha=1.0)+  
  geom_rug(sides="1", color="black")+  
  geom_smooth(method = "lm", color = "red", linetype=2)
```

```
windows()
```

```
ggplot(data=laptops, aes(x = Brend, fill=GPU))+  
  geom_histogram(stat = "count")
```

```
windows()
```

```
ggplot(data=laptops, aes(x = OS, fill=CPU.TYPE))+  
  geom_histogram(stat = "count")
```

```
}
```

```
# KORELACIJE -----
```

```
rm(laptops.num)
```

```
with(laptops, {  
  laptops.num <- data.frame(Cijena, RAM.num, HDD.num, ssd.num, GPU.num, Ekran)  
})
```

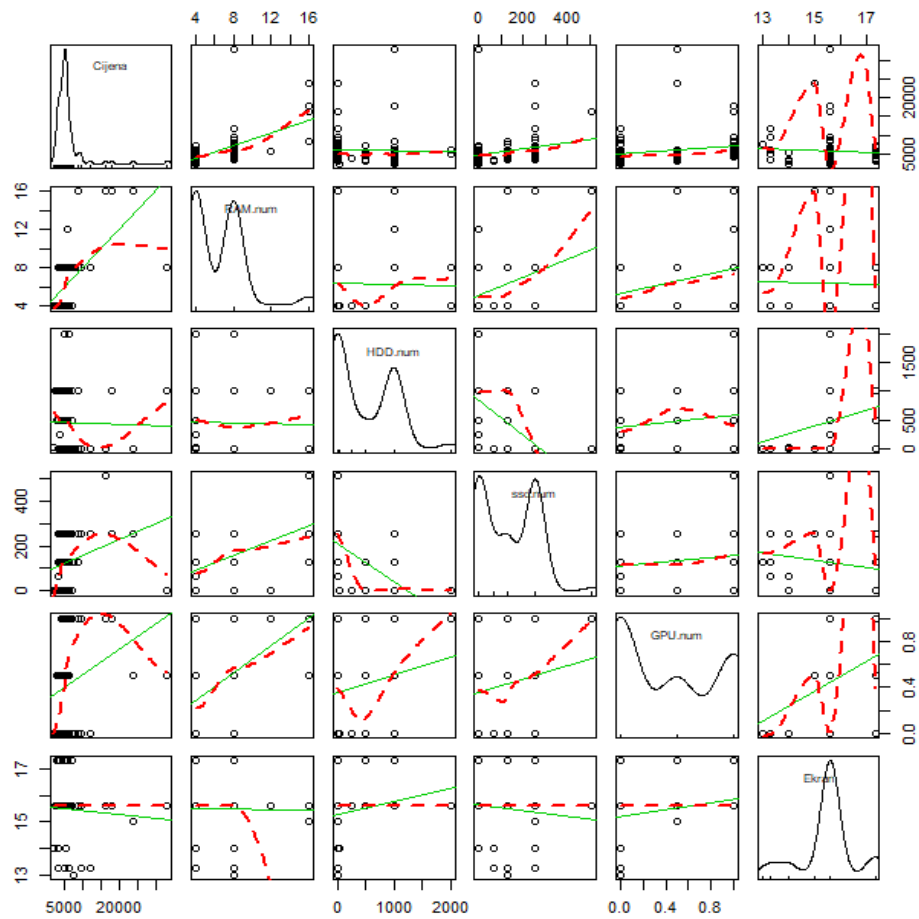
```
# Test korelacija
```

```
options(digits=2)
```

```
cor(laptops.num)
```

```
library(car)
```

```
scatterplotMatrix(laptops.num, spread=FALSE, smoother.args=list(lty=2))
```



```
# Korelogram
```

```
library(corrgram)
```

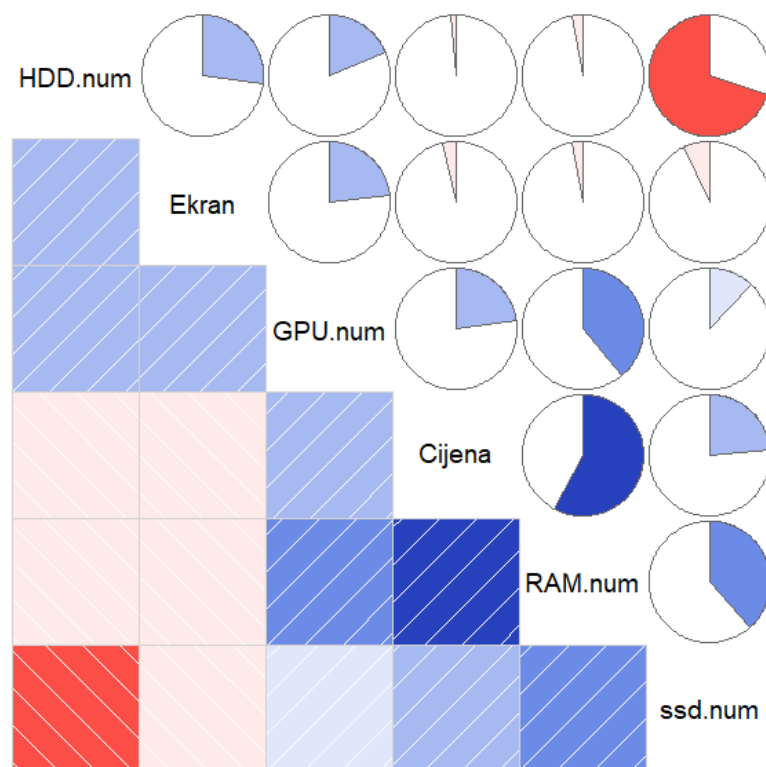
```
windows()
```

```
corrgram(laptops.num, order=TRUE,
```

```
  lower.panel = panel.shade,
```

```
  upper.panel = panel.pie,
```

```
  text.panel = panel.txt)
```



MULTIPLA REGRESIJA -----

library(MASS) #za stepwise regresiju

library(leaps) #za all-subsets metodu

Model fit - nominalne varijable

```
fit1 <- lm(Cijena ~ CPU.TYPE.num + RAM.num + HDD.num, data=laptops)
```

```
fit2 <- lm(Cijena ~ CPU.TYPE.num + RAM.num + OS.num + HDD.num, data=laptops)
```

```
fit3 <- lm(Cijena ~ CPU.TYPE.num + RAM.num + ssd.num + GPU.num + HDD.num + OS.num,
data=laptops)
```

```
summary(fit1)
```

```
summary(fit2)
```

```
summary(fit3)
```

```
stepAIC(fit3, direction = "backward")
```

```
# Model fit - numeričke varijable
```

```
fit.num <- lm(Cijena ~ RAM.num , data=laptops)
```

```
summary(fit.num)
```

```
stepAIC(fit.num, direction = "backward")
```

```
# Model fit - testiranje različitih modela
```

```
fit.test <- lm(Cijena ~ CPU.TYPE.num + RAM.num+ssd.num + HDD.num + OS.num + Ekran ,  
data=laptops)
```

```
summary(fit.test)
```

```
stepAIC(fit.test, direction = "backward")
```

```
#All-subsets metoda
```

```
leaps <- regsubsets(Cijena ~ CPU.TYPE.num +
```

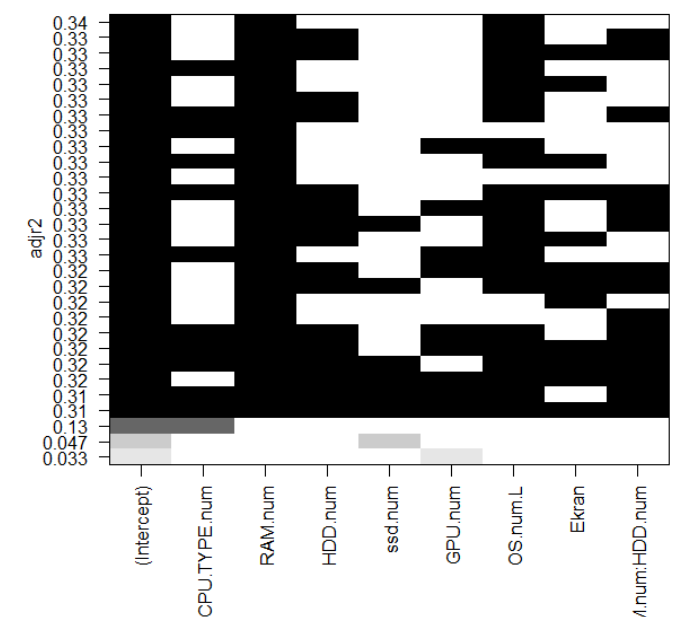
```
RAM.num*HDD.num + ssd.num +
```

```
GPU.num + OS.num +
```

```
Ekran, data=laptops, nbest=4)
```

```
windows()
```

```
plot(leaps, scale="adjr2")
```



```
fit.test2 <- lm(Cijena ~ CPU.TYPE.num +  
               RAM.num:HDD.num + ssd.num +  
               OS.num + Ekran, data=laptops)  
summary(fit.test2)
```

```
# PREDIKCIJA -----
```

```
testdata.num <- data.frame(CPU.TYPE.num = 5,  
                           RAM.num = 8,  
                           HDD.num = 1000,  
                           ssd.num = 126,  
                           GPU.num = 1,  
                           Ekran = 15.6,  
                           Tezina = 1.8)  
predict(fit.num, newdata=testdata.num)
```

```
## 1
```

```
## 7142
```

```
# DIJAGNOSTIKA REGRESIJE -----
```

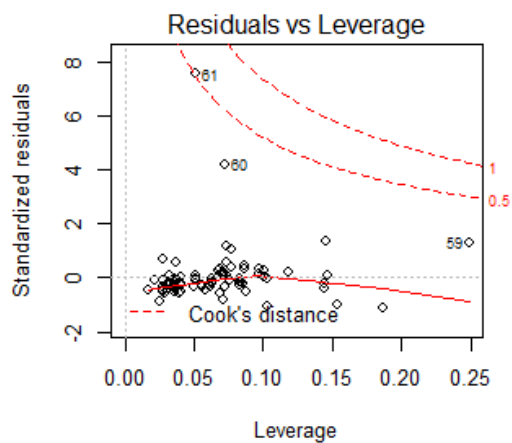
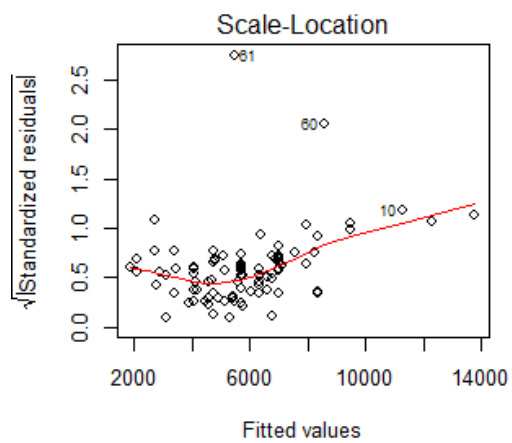
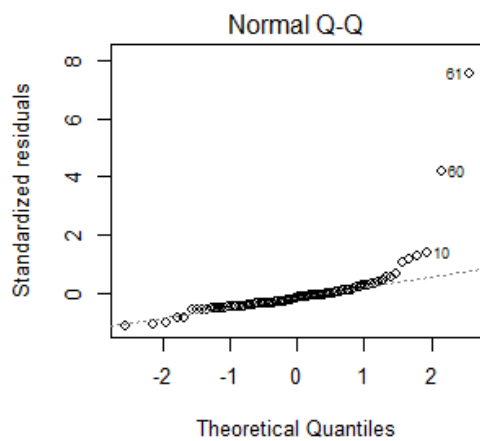
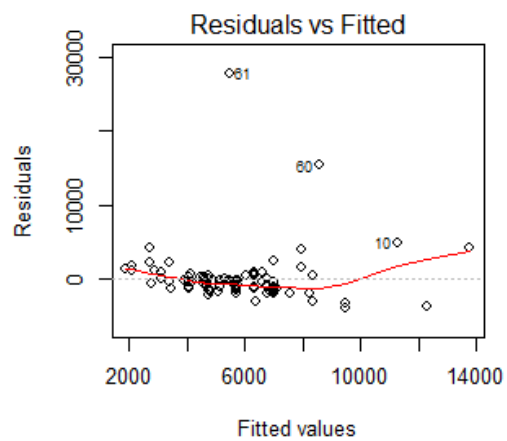
```
model <- fit.test2
```

```
# Normalnost / Linearnost / Homoskedastičnost / Nezavisnost
```

```
windows()
```

```
par(mfrow=c(2,2))
```

```
plot(model)
```

Identifikacija outlier-a

```
library(car)
```

```
outlierTest(model)
```

```
##> outlierTest(model)
```

```
# rstudent unadjusted p-value Bonferonni p
```

```
##      61    12.6      2.7e-21    2.5e-19
```

```
##      60     4.7      1.1e-05    1.0e-03
```

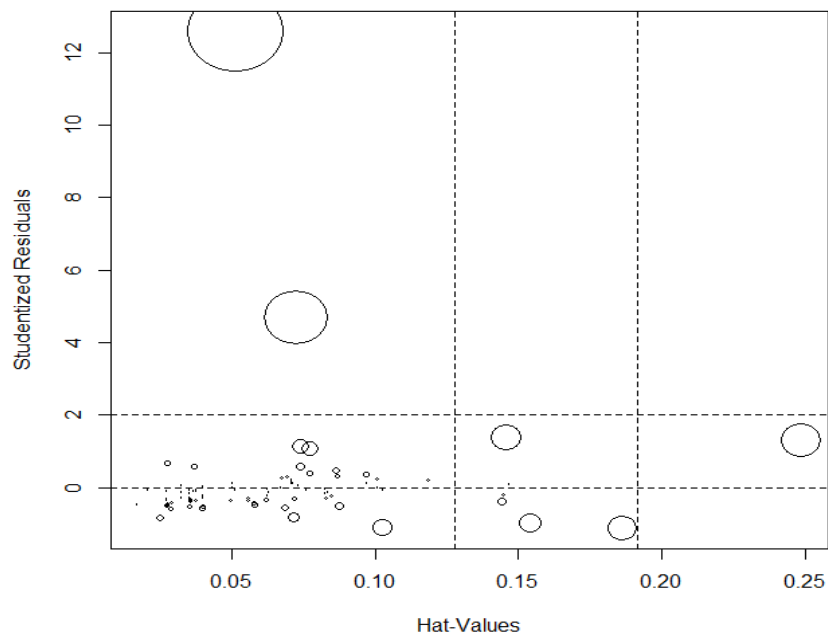
```
##>
```

```
# Identifikacija ostalih karakteristiènih toèki
```

```
library(car)
```

```
windows()
```

```
influencePlot(model, id.method = "identify")
```



```
# PRECIJENJENI I PODCIJENJENI PREMA MODELU -----
```

```
best.model <- fit.test2
```

```
fitted(best.model)
```

```
residuals(best.model)
```

```
sort(residuals(best.model)) #reziduali: podcijenjeno (<0), precijenjeno (>0)
```

```
> sort(residuals(best.model)) #reziduali: podcijenjeno (<0), precijenjeno (>0)
50      1      29      66      85      16      55      49      21      78      41      82      68      83      46      64      47      86      69      7
0      65
-3902 -3835 -3414 -3199 -3026 -2182 -2111 -2057 -2011 -1937 -1929 -1897 -1883 -1801 -1754 -1735 -1733 -1629 -1613 -158
1 -1571
36      74      19      2      11      73      18      44      84      15      20      5      52      75      13      34      77      7      9      5
3      26
-1468 -1443 -1388 -1364 -1345 -1308 -1259 -1258 -1256 -1239 -1237 -1237 -1230 -1214 -1087 -1057 -1030 -1024 -1024 -99
0 -890
22      24      27      92      42      33      17      28      54      32      56      45      76      4      48      81      67      89      23      4
3      72
-807 -781 -707 -644 -571 -522 -505 -505 -501 -430 -429 -341 -305 -304 -259 -254 -210 -202 -184 -14
7 -62
37      87      6      12      38      39      80      57      94      30      25      90      97      40      58      3      35      71      62      1
4      8
-56      26      28      47      231      244      283      424      425      434      443      460      646      746      766      902      967      1096      1126      128
1 1504
91      79      51      31      96      95      59      10      60      61
1661 2130 2155 2512 3932 4197 4223 4846 15334 27810
```

```
laptops$residuals <- residuals(best.model) #pohrana reziduala
```

```
# Top lista (od najpodcjenjenijih do najprecjenjenijih)
```

```
toplista <- order(residuals(best.model))
```

```
# Ispis
```

```
laptops[toplista] #Ispis top liste
```

```
toplista
```

```
> toplista
[1] 50 1 29 65 84 16 55 49 21 77 41 81 67 82 46 63 47 85 68 69 64 36 73 19 2 11 72 18 44 83 15 20 5 52 74 13 34 76
7 9 53 26
[43] 22 24 27 90 42 33 17 28 54 32 56 45 75 4 48 80 66 87 23 43 71 37 86 6 12 38 39 79 57 91 30 25 88 94 40 58 3 35
70 62 14 8
[85] 89 78 51 31 93 92 59 10 60 61
>
```