

# Projekt SAP

## Tema 2 - Uloga izvoza i uvoza u gospodarstvu

Pavo Matanović, Karla Baričević, Slavko Boldin

### Uvod

Na gospodarstvo neke države utječu mnogi čimbenici. Najvažniji su uvoz i izvoz. Izvozom novac "ulazi" u državu, dok uvozom on "izlazi". Upravo zato države uglavnom potiču izvoz jer smatraju da on donosi nove poslove, povećanja plaća i općenito podiže životni standard. Analogno tome, pojedine države pokušavaju ograničiti uvoz.

Analiza u ovom radu sastoji se od tri etape: deskriptivna analiza, testovi sredina i intervalne procjene, te analiza zasnovana na linearnoj regresiji i analizi varijance.

### Učitavanje podataka i deskriptivna analiza

Na početku učitavamo podatke i analiziramo kako izgledaju podaci. Dataset se sastoji od podataka o iznosima izvoza, uvoza, BDP-a i BDP-a per capita u milijunima američkih dolara za 20 država.

```
export.data = read.csv("Export_data.csv", fileEncoding="UTF-8-BOM")
# head(export.data)

import.data = read.csv("Import_data.csv", fileEncoding="UTF-8-BOM")
# head(import.data)

gdp.data = read.csv("GDP_data.csv", fileEncoding="UTF-8-BOM")
# head(gdp.data)

gdp.pc.data = read.csv("GDPpercapita_data.csv", fileEncoding="UTF-8-BOM")
# head(gdp.pc.data)
```

Sljedeći blok koda generira dataframe s brojem upisanih podataka te brojem procjena među upisanim podacima.

```
export.loc.cnt = export.data %>% group_by(LOCATION) %>%
  summarise(exp_n = n(), exp_est = sum(Flag.Codes == 'E')) %>%
  arrange(desc(exp_n), exp_est)
import.loc.cnt = import.data %>% group_by(LOCATION) %>%
  summarise(imp_n = n(), imp_est = sum(Flag.Codes == 'E')) %>%
  arrange(desc(imp_n), imp_est)
gdp.loc.cnt = gdp.data %>% group_by(LOCATION) %>%
  summarise(gdp_n = n(), gdp_est = sum(Flag.Codes == 'E')) %>%
  arrange(desc(gdp_n), gdp_est)
gdp.pc.loc.cnt = gdp.pc.data %>% group_by(LOCATION) %>%
  summarise(gdp_pc_n = n(), gdp_pc_est = sum(Flag.Codes == 'E')) %>%
  arrange(desc(gdp_pc_n), gdp_pc_est)
loc.cnt = merge(merge(export.loc.cnt, import.loc.cnt), merge(gdp.loc.cnt, gdp.pc.loc.cnt))
```

```
knitr::kable(
  head(arrange(loc.cnt,
    desc(loc.cnt[,2]), desc(loc.cnt[,4]), desc(loc.cnt[,6]), desc(loc.cnt[,8]),
    loc.cnt[,3], loc.cnt[,5], loc.cnt[,7], loc.cnt[,9]), 20),
  caption = "Broj podataka za pojedinu državu"
)
```

Table 1: Broj podataka za pojedinu državu

LOCATION	exp_n	exp_est	imp_n	imp_est	gdp_n	gdp_est	gdp_pc_n	gdp_pc_est
CAN	41	0	41	0	41	0	41	0
DNK	41	0	41	0	41	0	41	0
FRA	41	0	41	0	41	0	41	0
CHE	41	1	41	1	41	1	41	1
FIN	41	1	41	1	41	1	41	1
DEU	41	12	41	12	41	12	41	12
SWE	41	14	41	14	41	14	41	14
GBR	41	16	41	16	41	0	41	0
AUT	41	16	41	16	41	16	41	16
BEL	41	16	41	16	41	16	41	16
ESP	41	16	41	16	41	16	41	16
GRC	41	16	41	16	41	16	41	16
IRL	41	16	41	16	41	16	41	16
ITA	41	16	41	16	41	16	41	16
NLD	41	16	41	16	41	16	41	16
ISL	41	16	41	16	41	16	41	17
PRT	41	17	41	17	41	17	41	17
AUS	40	0	40	0	40	0	40	0
NOR	40	0	40	0	40	0	40	0
USA	40	0	40	0	40	0	40	0

## Odabrane drzave

Odabrali smo USA, Njemačku(DEU) i Grčku(GRC) za analizu. Vremenski period je od 1979. do 2018. u većini analiza (nedostaju podaci iz 2019. za SAD).

## Deskriptivna statistika

```
time = 1979:2018
usa = data.frame(year = 1979:2018,
  export.mln_usd = export.data$Value[export.data$LOCATION == "USA"],
  import.mln_usd = import.data$Value[import.data$LOCATION == "USA"],
  gdp.mln_usd = gdp.data$Value[gdp.data$LOCATION == "USA"],
  gdp.pc.usd_cap = gdp.pc.data$Value[gdp.pc.data$LOCATION == "USA"])
usa$net.trade = usa$export.mln_usd - usa$import.mln_usd
deu = data.frame(year = 1979:2019,
  export.mln_usd = export.data$Value[export.data$LOCATION == "DEU"],
  import.mln_usd = import.data$Value[import.data$LOCATION == "DEU"],
  gdp.mln_usd = gdp.data$Value[gdp.data$LOCATION == "DEU"],
  gdp.pc.usd_cap = gdp.pc.data$Value[gdp.pc.data$LOCATION == "DEU"])
deu$net.trade = deu$export.mln_usd - deu$import.mln_usd
grc = data.frame(year = 1979:2019,
```

```

        export.mln_usd = export.data$Value[export.data$LOCATION == "GRC"],
        import.mln_usd = import.data$Value[import.data$LOCATION == "GRC"],
        gdp.mln_usd = gdp.data$Value[gdp.data$LOCATION == "GRC"],
        gdp.pc.usd_cap = gdp.pc.data$Value[gdp.pc.data$LOCATION == "GRC"])
grc$net.trade = grc$export.mln_usd - grc$import.mln_usd

```

```

usa = usa %>% mutate(import.mln_usd,
                     import.growth = import.mln_usd - lag(import.mln_usd),
                     import.growth.p = import.growth / lag(import.mln_usd) * 100)
deu = deu %>% mutate(import.mln_usd,
                     import.growth = import.mln_usd - lag(import.mln_usd),
                     import.growth.p = import.growth / lag(import.mln_usd) * 100)
grc = grc %>% mutate(import.mln_usd,
                     import.growth = import.mln_usd - lag(import.mln_usd),
                     import.growth.p = import.growth / lag(import.mln_usd) * 100)
usa = usa %>% mutate(export.mln_usd,
                     export.growth = export.mln_usd - lag(export.mln_usd),
                     export.growth.p = export.growth / lag(export.mln_usd) * 100)
deu = deu %>% mutate(export.mln_usd,
                     export.growth = export.mln_usd - lag(export.mln_usd),
                     export.growth.p = export.growth / lag(export.mln_usd) * 100)
grc = grc %>% mutate(export.mln_usd,
                     export.growth = export.mln_usd - lag(export.mln_usd),
                     export.growth.p = export.growth / lag(export.mln_usd) * 100)
usa = usa %>% mutate(gdp.mln_usd,
                     gdp.growth = gdp.mln_usd - lag(gdp.mln_usd),
                     gdp.growth.p = gdp.growth / lag(gdp.mln_usd) * 100)
deu = deu %>% mutate(gdp.mln_usd,
                     gdp.growth = gdp.mln_usd - lag(gdp.mln_usd),
                     gdp.growth.p = gdp.growth / lag(gdp.mln_usd) * 100)
grc = grc %>% mutate(gdp.mln_usd,
                     gdp.growth = gdp.mln_usd - lag(gdp.mln_usd),
                     gdp.growth.p = gdp.growth / lag(gdp.mln_usd) * 100)
usa = usa %>% mutate(gdp.pc.usd_cap,
                     gdp.pc.growth = gdp.pc.usd_cap - lag(gdp.pc.usd_cap),
                     gdp.pc.growth.p = gdp.pc.growth / lag(gdp.pc.usd_cap) * 100)
deu = deu %>% mutate(gdp.pc.usd_cap,
                     gdp.pc.growth = gdp.pc.usd_cap - lag(gdp.pc.usd_cap),
                     gdp.pc.growth.p = gdp.pc.growth / lag(gdp.pc.usd_cap) * 100)
grc = grc %>% mutate(gdp.pc.usd_cap,
                     gdp.pc.growth = gdp.pc.usd_cap - lag(gdp.pc.usd_cap),
                     gdp.pc.growth.p = gdp.pc.growth / lag(gdp.pc.usd_cap) * 100)
# brisemo zadnju opservaciju za deu i grc jer usa nema podatke za 2019.g.
deu = deu[-nrow(deu),]
grc = grc[-nrow(grc),]

```

```

data.all = bind_rows(lapply(c("USA", "DEU", "GRC"), function (x) {
  data.frame(country=x, get(tolower(x)))
})))

```

```

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character
## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

```

```
## Warning in bind_rows(x, .id): binding character and factor vector, coercing
## into character vector
```

```
## Warning in bind_rows(x, .id): binding character and factor vector, coercing
## into character vector
```

```
data.all$country = factor(data.all$country, levels = c("USA", "DEU", "GRC"))
```

```
summary(usa)
```

```
##      year      export.mln_usd  import.mln_usd  gdp.mln_usd
## Min.   :1979   Min.    : 347872   Min.     : 366207   Min.     : 2627334
## 1st Qu.:1989   1st Qu.: 591516   1st Qu.: 682910   1st Qu.: 5540294
## Median :1998   Median :1185694   Median :1459992   Median : 9346740
## Mean   :1998   Mean    :1206334   Mean     :1503298   Mean    :10103023
## 3rd Qu.:2008   3rd Qu.:1762818   3rd Qu.:2323213   3rd Qu.:14517106
## Max.   :2018   Max.     :2416053   Max.     :3105836   Max.     :20580223
##
## gdp.pc.usd_cap  net.trade      import.growth  import.growth.p
## Min.   :11672   Min.    :-722881   Min.     :-304448   Min.     :-13.084
## 1st Qu.:22445   1st Qu.: -497748   1st Qu.:  31413   1st Qu.:  2.659
## Median :33648   Median : -274298   Median :   65420   Median :   5.288
## Mean   :34815   Mean     : -296964   Mean      : 69577   Mean      : 5.637
## 3rd Qu.:48004   3rd Qu.: -77556   3rd Qu.: 131694   3rd Qu.:  8.672
## Max.   :62853   Max.      : 19122   Max.      :265511   Max.      :24.343
##
## export.growth  export.growth.p  gdp.growth  gdp.growth.p
## Min.   :-154210   Min.    :-8.397   Min.     :-263913   Min.     :-1.794
## 1st Qu.: 25540   1st Qu.: 2.676   1st Qu.: 325520   1st Qu.: 4.095
## Median : 57629   Median : 6.614   Median : 428677   Median : 5.546
## Mean   : 53030   Mean     : 5.239   Mean      :460330   Mean      :5.447
## 3rd Qu.: 89140   3rd Qu.: 8.829   3rd Qu.: 604764   3rd Qu.: 6.360
## Max.   :204211   Max.     :16.212   Max.     :1060799   Max.     :12.240
## NA's    :1       NA's     :1       NA's      :1       NA's      :1
## gdp.pc.growth  gdp.pc.growth.p
## Min.   :-1283.1   Min.    :-2.656
## 1st Qu.: 984.6   1st Qu.: 3.252
## Median :1374.7   Median : 4.448
## Mean   :1312.3   Mean     :4.438
## 3rd Qu.:1665.6   3rd Qu.: 5.186
## Max.   :2868.6   Max.     :11.126
## NA's    :1       NA's     :1
```

```
summary(deu)
```

```
##      year      export.mln_usd  import.mln_usd  gdp.mln_usd
## Min.   :1979   Min.    : 277599   Min.     : 303002   Min.     : 736116
## 1st Qu.:1989   1st Qu.: 429192   1st Qu.: 413063   1st Qu.:1387158
## Median :1998   Median : 769614   Median : 748374   Median :2118984
## Mean   :1998   Mean     : 935370   Mean      :828751   Mean    :2287035
## 3rd Qu.:2008   3rd Qu.:1466327   3rd Qu.:1216519   3rd Qu.:3039642
## Max.   :2018   Max.     :2001818   Max.     :1740059   Max.     :4514794
##
## gdp.pc.usd_cap  net.trade      import.growth  import.growth.p
## Min.   : 9425   Min.    :-27681   Min.     :-119972   Min.     :-9.695
## 1st Qu.:17661   1st Qu.: 12404   1st Qu.:  12114   1st Qu.:  2.857
```

```
## Median :26021 Median : 35989 Median : 37491 Median : 5.185
## Mean :28294 Mean :106619 Mean : 36789 Mean : 4.674
## 3rd Qu.:37734 3rd Qu.:214854 3rd Qu.: 62588 3rd Qu.: 8.020
## Max. :54457 Max. :294551 Max. : 143844 Max. :12.871
## NA's :1 NA's :1
## export.growth export.growth.p gdp.growth gdp.growth.p
## Min. :-213503 Min. :-14.280 Min. :-85755 Min. :-2.763
## 1st Qu.: 19873 1st Qu.: 2.271 1st Qu.: 59055 1st Qu.: 3.114
## Median : 40328 Median : 5.661 Median : 79063 Median : 4.481
## Mean : 44211 Mean : 5.343 Mean : 96889 Mean : 4.792
## 3rd Qu.: 75638 3rd Qu.: 8.610 3rd Qu.:131475 3rd Qu.: 6.318
## Max. : 184671 Max. : 14.410 Max. :268773 Max. :10.569
## NA's :1 NA's :1 NA's :1 NA's :1
## gdp.pc.growth gdp.pc.growth.p
## Min. :-931.3 Min. :-2.423
## 1st Qu.: 719.1 1st Qu.: 3.090
## Median : 951.0 Median : 4.379
## Mean :1154.7 Mean : 4.630
## 3rd Qu.:1502.7 3rd Qu.: 6.190
## Max. :2880.5 Max. :10.288
## NA's :1 NA's :1
```

`summary(grc)`

```
## year export.mln_usd import.mln_usd gdp.mln_usd
## Min. :1979 Min. : 20482 Min. : 22900 Min. : 76529
## 1st Qu.:1989 1st Qu.: 28444 1st Qu.: 36651 1st Qu.:128217
## Median :1998 Median : 51170 Median : 73155 Median :196020
## Mean :1998 Mean : 55043 Mean : 69218 Mean :207226
## 3rd Qu.:2008 3rd Qu.: 80753 3rd Qu.: 96276 3rd Qu.:290500
## Max. :2018 Max. :104650 Max. :137267 Max. :341818
##
## gdp.pc.usd_cap net.trade import.growth import.growth.p
## Min. : 7933 Min. :-44429 Min. :-27939.2 Min. :-20.3539
## 1st Qu.:12593 1st Qu.: -24665 1st Qu.: 321.2 1st Qu.: 0.5161
## Median :18249 Median :-12461 Median : 1789.3 Median : 4.4176
## Mean :19320 Mean :-14175 Mean : 2041.9 Mean : 4.2218
## 3rd Qu.:26315 3rd Qu.: -3260 3rd Qu.: 4417.3 3rd Qu.: 8.6914
## Max. :30856 Max. : 2114 Max. : 18180.4 Max. : 20.1801
## NA's :1 NA's :1 NA's :1
## export.growth export.growth.p gdp.growth gdp.growth.p
## Min. :-17447.1 Min. :-18.51934 Min. :-24242 Min. :-7.337
## 1st Qu.: 52.6 1st Qu.: 0.07368 1st Qu.: 3550 1st Qu.: 2.507
## Median : 1969.5 Median : 4.08448 Median : 6619 Median : 4.241
## Mean : 2128.8 Mean : 4.51601 Mean : 6391 Mean : 3.864
## 3rd Qu.: 3815.2 3rd Qu.: 8.54993 3rd Qu.: 11152 3rd Qu.: 6.377
## Max. : 12587.3 Max. : 24.44487 Max. : 33214 Max. :11.819
## NA's :1 NA's :1 NA's :1 NA's :1
## gdp.pc.growth gdp.pc.growth.p
## Min. :-2219.0 Min. :-7.302
## 1st Qu.: 276.5 1st Qu.: 2.225
## Median : 689.4 Median : 4.321
## Mean : 574.9 Mean : 3.573
## 3rd Qu.: 975.0 3rd Qu.: 5.713
## Max. : 2937.2 Max. :11.483
```

```
## NA's :1      NA's :1
```

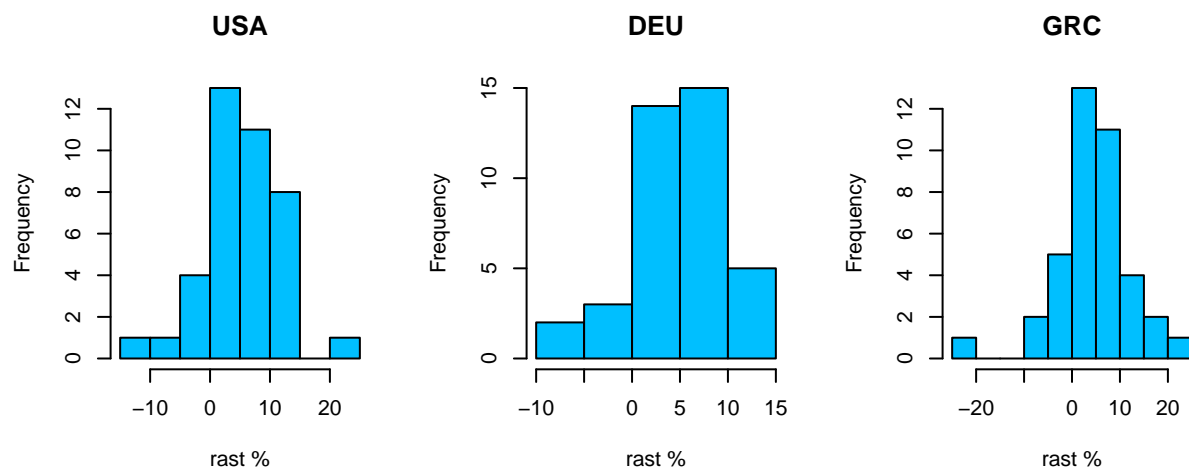
## Uvoz

Uvozi u mil. USD razlikuju se jako čak i na logaritamskoj skali. Veličine su razmjerne površini države te broju stanovnika. Za distribucije ukupnog uvoza ne možemo pretpostaviti normalnost pa nema smisla raditi parametarske testove.

Distribucije postotnog rasta izgledaju normalnije pa ćemo njih uzeti za analizu. U daljnjem tekstu ćemo za postotni rast govoriti samo rast.

```
par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))
hist(usa$import.growth.p, main="USA", xlab="rast %", col="deepskyblue")
hist(deu$import.growth.p, main="DEU", xlab="rast %", col="deepskyblue")
hist(grc$import.growth.p, main="GRC", xlab="rast %", col="deepskyblue")
mtext("Postotni rast uvoza", outer = T, cex = 1.5, font = 2)
```

## Postotni rast uvoza

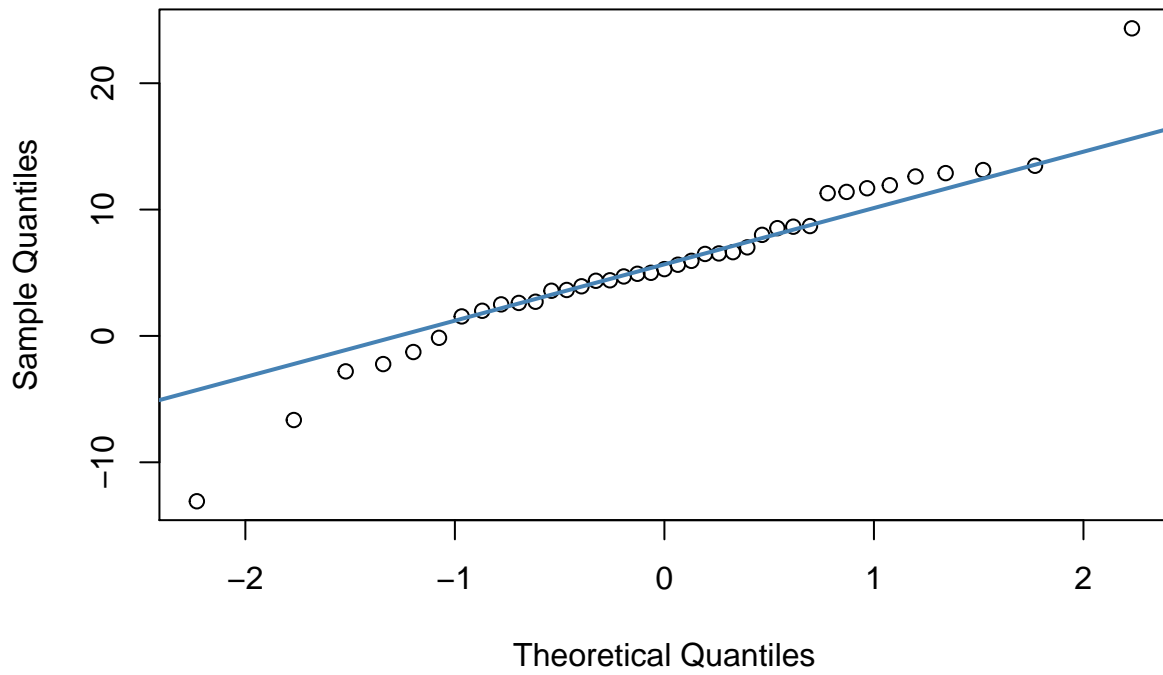


Distribucije nisu previše zakrivljene i imamo dovoljno podataka da možemo pretpostaviti normalnost distribucije.

Taj zaključak potvrđuju i qq plotovi.

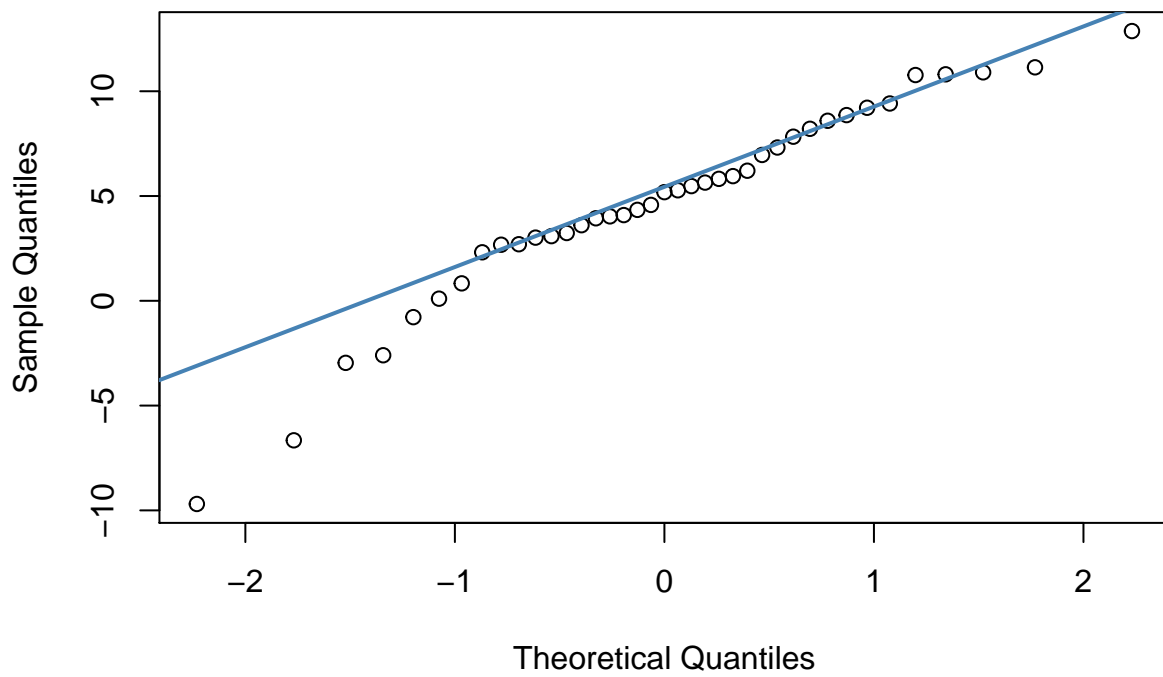
```
qqnorm(usa$import.growth.p)
qqline(usa$import.growth.p, col = "steelblue", lwd = 2)
```

Normal Q-Q Plot



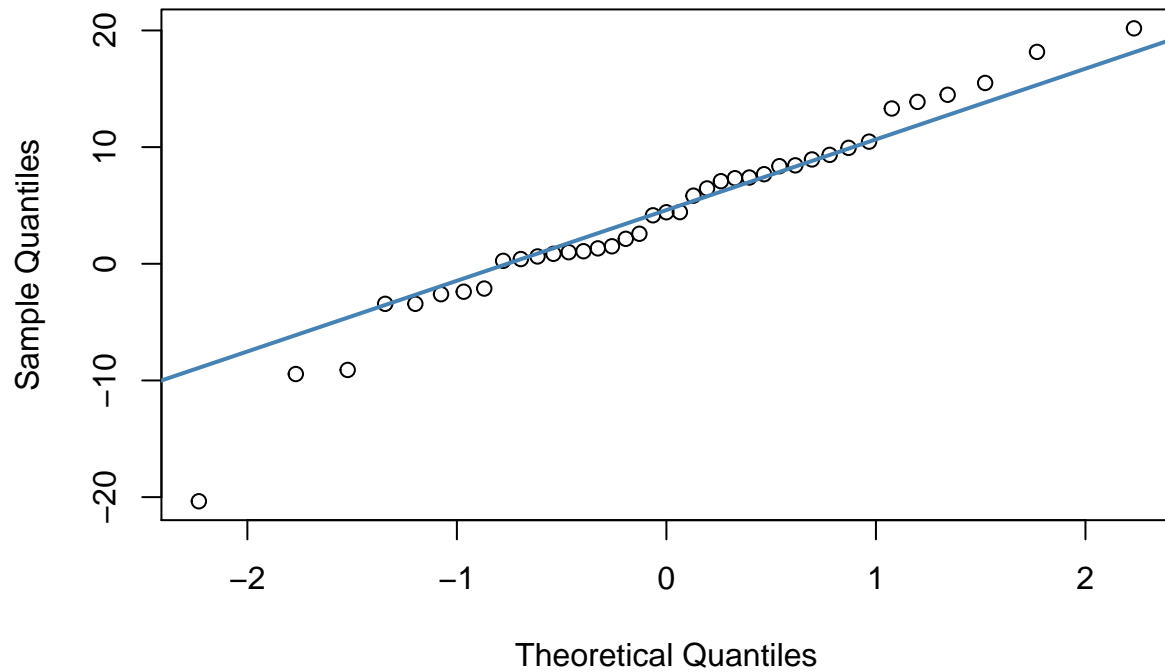
```
qqnorm(deu$import.growth.p)  
qqline(deu$import.growth.p, col = "steelblue", lwd = 2)
```

Normal Q-Q Plot



```
qqnorm(grc$import.growth.p)  
qqline(grc$import.growth.p, col = "steelblue", lwd = 2)
```

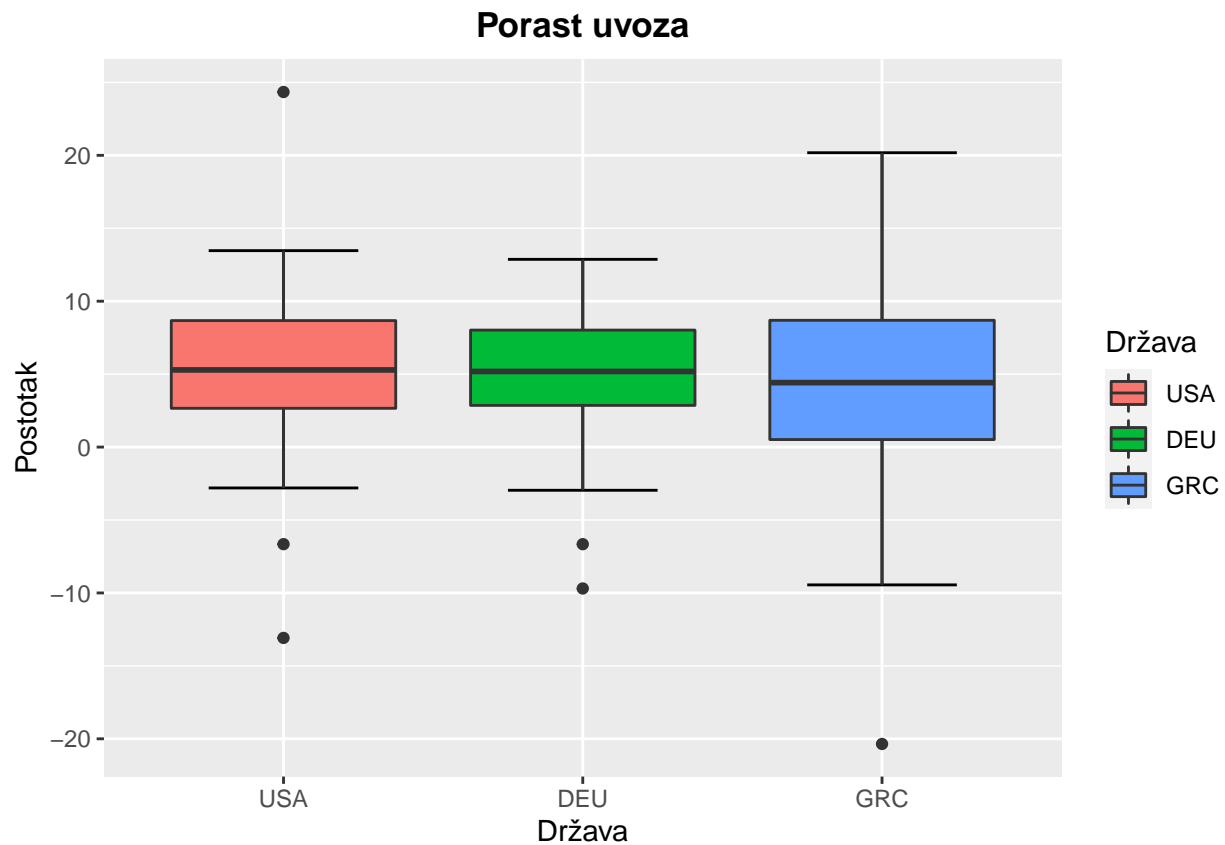
## Normal Q-Q Plot



Odstupanja na krajevima qq plota nam sugeriraju da bi distribucije mogle imati teške repove.

```
ggplot(na.omit(data.all), aes(x=country, y=import.growth.p)) +  
  stat_boxplot(geom = "errorbar", width = 0.5) +  
  geom_boxplot(aes(fill=country)) +  
  labs(title = "Porast uvoza", x="Država", y="Postotak", fill="Država") +  
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
```



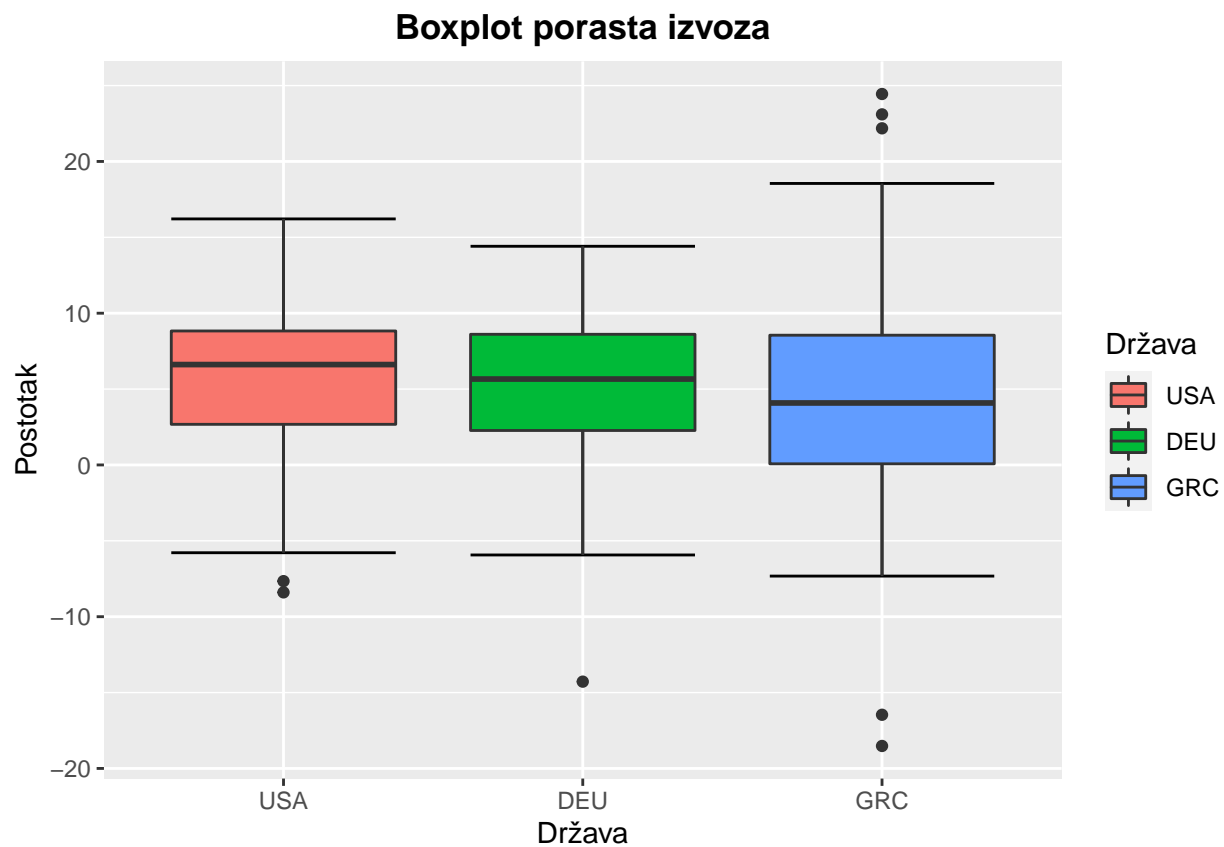


Ovaj plot pokazuje da bi varijable USA i DEU mogle imati istu sredinu.

## Izvoz

Vrijednosti izvoza po veličini su slične s uvozom pa ćemo u analizi koristiti rast godišnjeg iznosa.

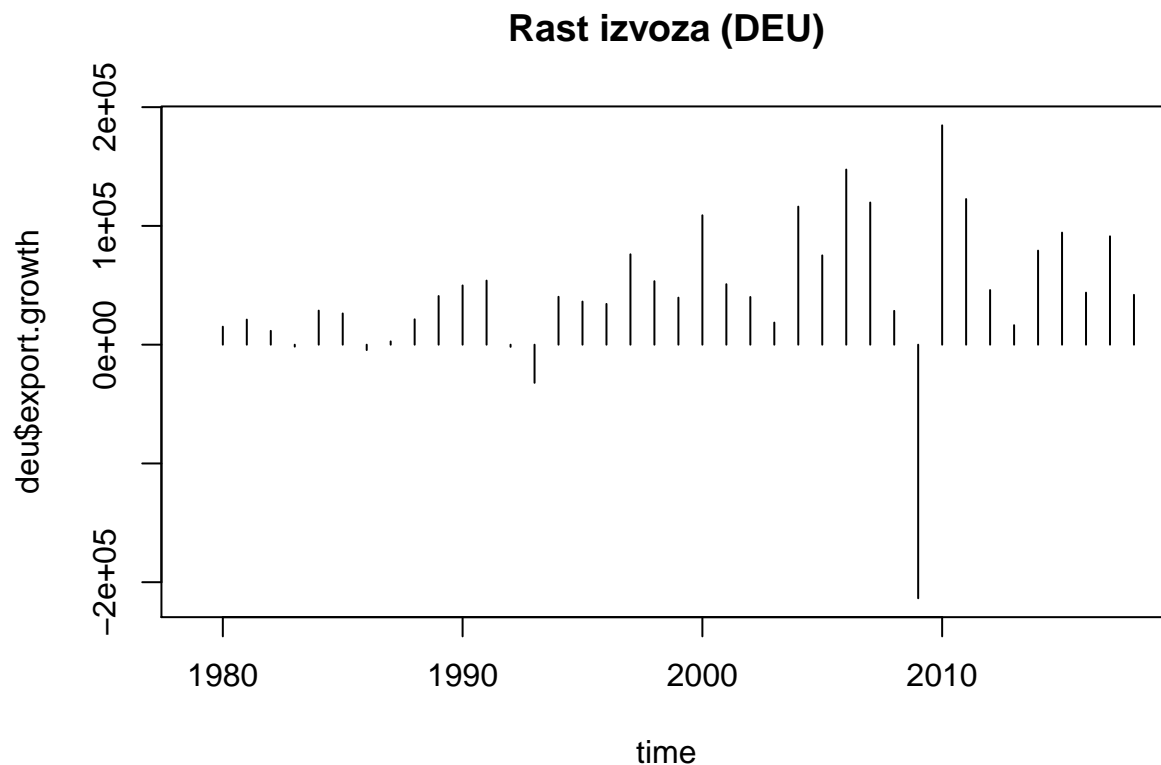
```
ggplot(na.omit(data.all), aes(x=country, y=export.growth.p)) +
  stat_boxplot(geom = "errorbar", na.rm = T) +
  geom_boxplot(aes(fill=country), na.rm = T) +
  labs(title = "Boxplot porasta izvoza", x="Država", y="Postotak", fill="Država") +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
```



Za razliku od uvoza, kod izvoza se primjećuje najveći rast SAD-a, potom Njemačke i Grčke.

Najveći broj stršećih vrijednosti primjećuje se kod Grčke u oba smjera, što upućuje na najveću gospodarsku nestabilnost od tri promatrane države.

```
plot(time, deu$export.growth, type = "h", main = "Rast izvoza (DEU)")
```

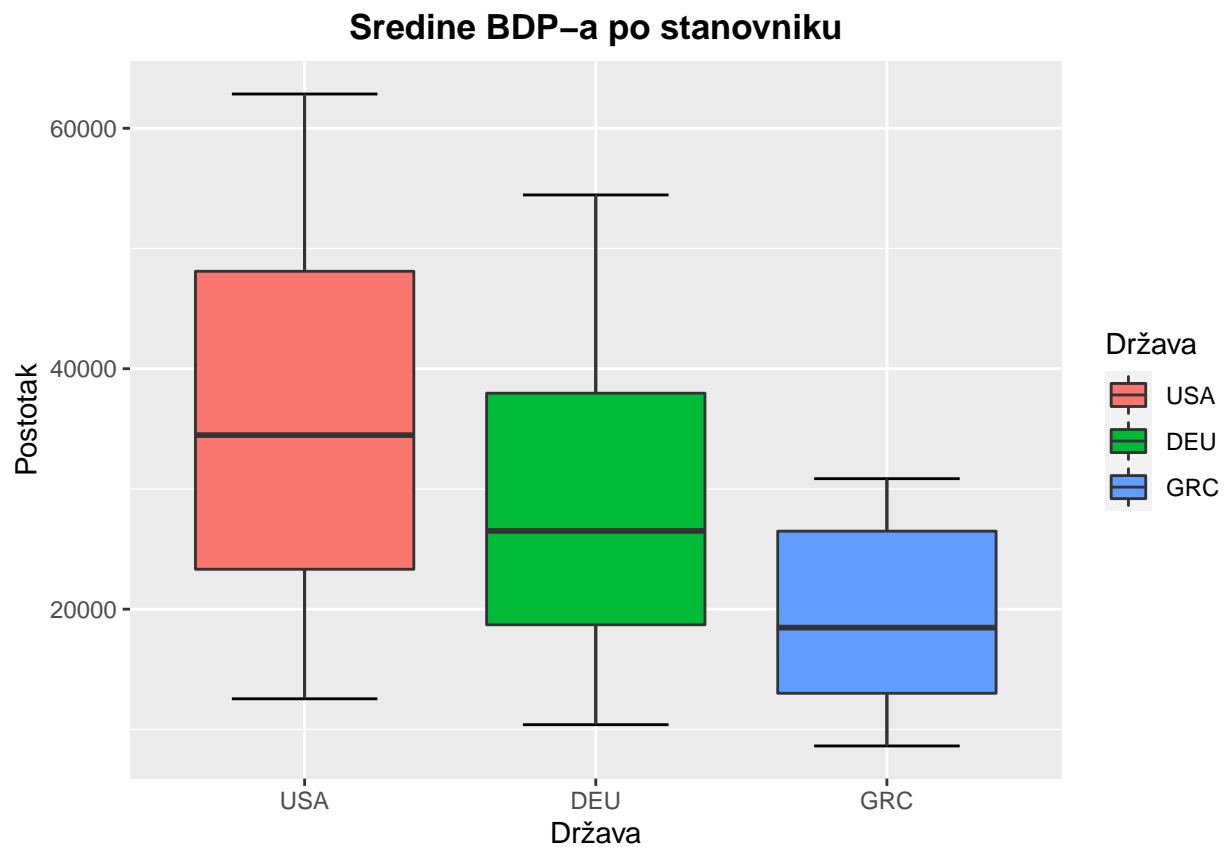


Za razliku od uvoza koji linearno raste, izvoz više “osjeća” promjene na tržištu (veće fluktuacije), npr. značajan pad izvoza 2009. godine zbog tadašnje svjetske gospodarske krize.

## BDP

BDP po stanovniku je pokazatelj razvijenosti:

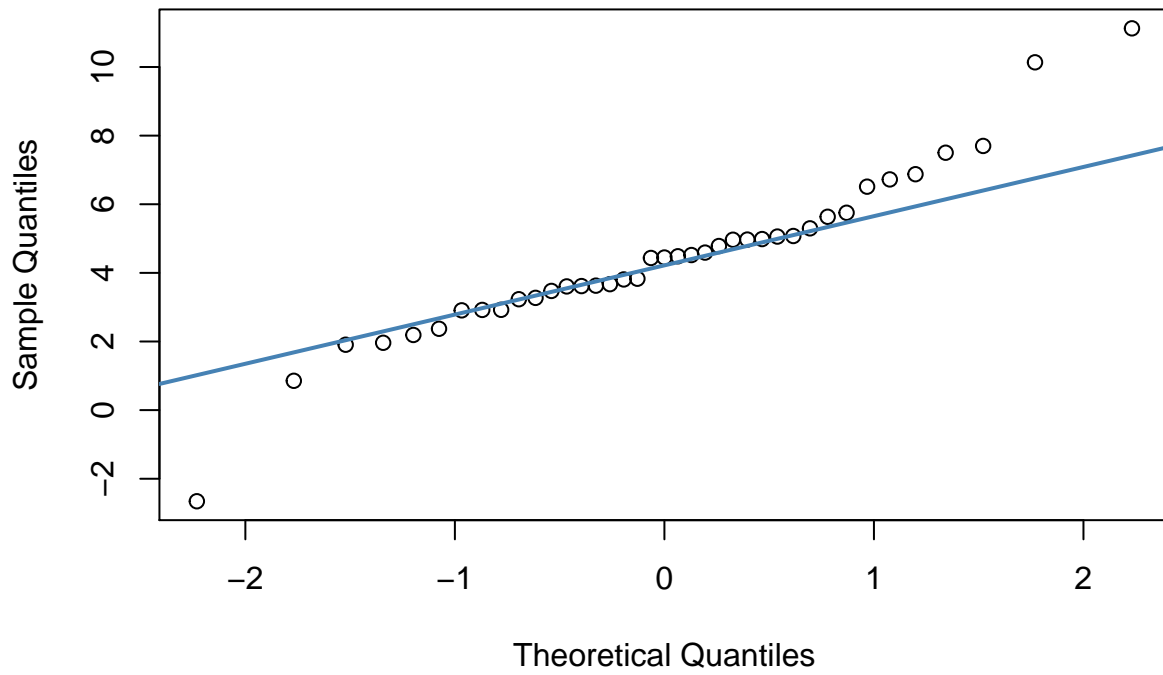
```
ggplot(na.omit(data.all), aes(x=country, y=gdp.pc.usd_cap)) +
  stat_boxplot(geom = "errorbar", width = 0.5) +
  geom_boxplot(aes(fill=country)) +
  labs(title = "Sredine BDP-a po stanovniku", x="Država", y="Postotak", fill="Država") +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
```



Kao i kod uvoza i izvoza, po BDP-u po stanovniku SAD značajno prednjači, dok je razlika između Njemačke i Grčke veća od one između SAD-a i Njemačke. Spram čistog BDP-a kod BDP-a po stanovniku podaci su normalizirani brojem stanovnika te su mjerodavniji.

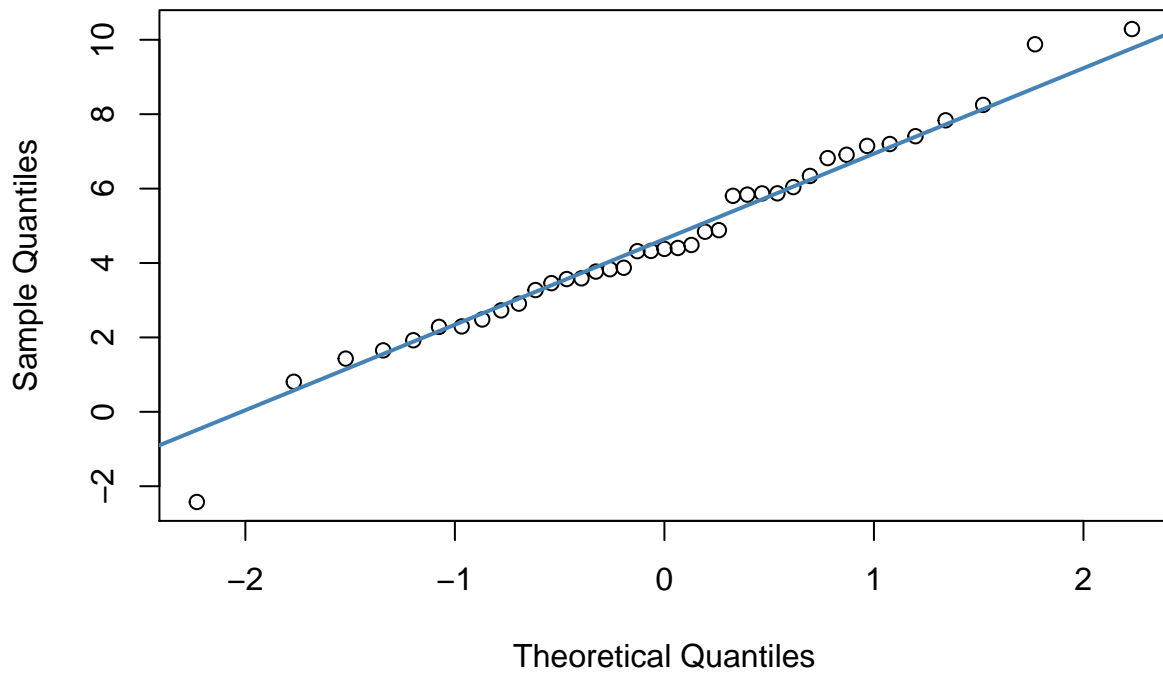
```
qqnorm(usa$gdp.pc.growth.p)
qqline(usa$gdp.pc.growth.p, col = "steelblue", lwd = 2)
```

Normal Q-Q Plot



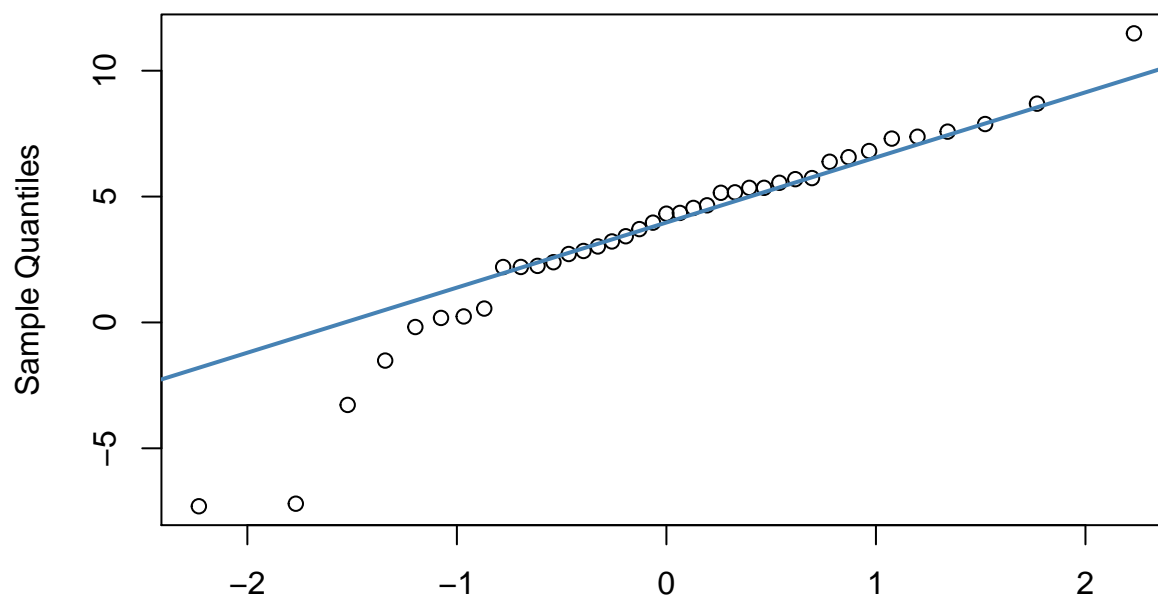
```
qqnorm(deu$gdp.pc.growth.p)  
qqline(deu$gdp.pc.growth.p, col = "steelblue", lwd = 2)
```

Normal Q-Q Plot



```
qqnorm(grc$gdp.pc.growth.p)  
qqline(grc$gdp.pc.growth.p, col = "steelblue", lwd = 2)
```

## Normal Q-Q Plot

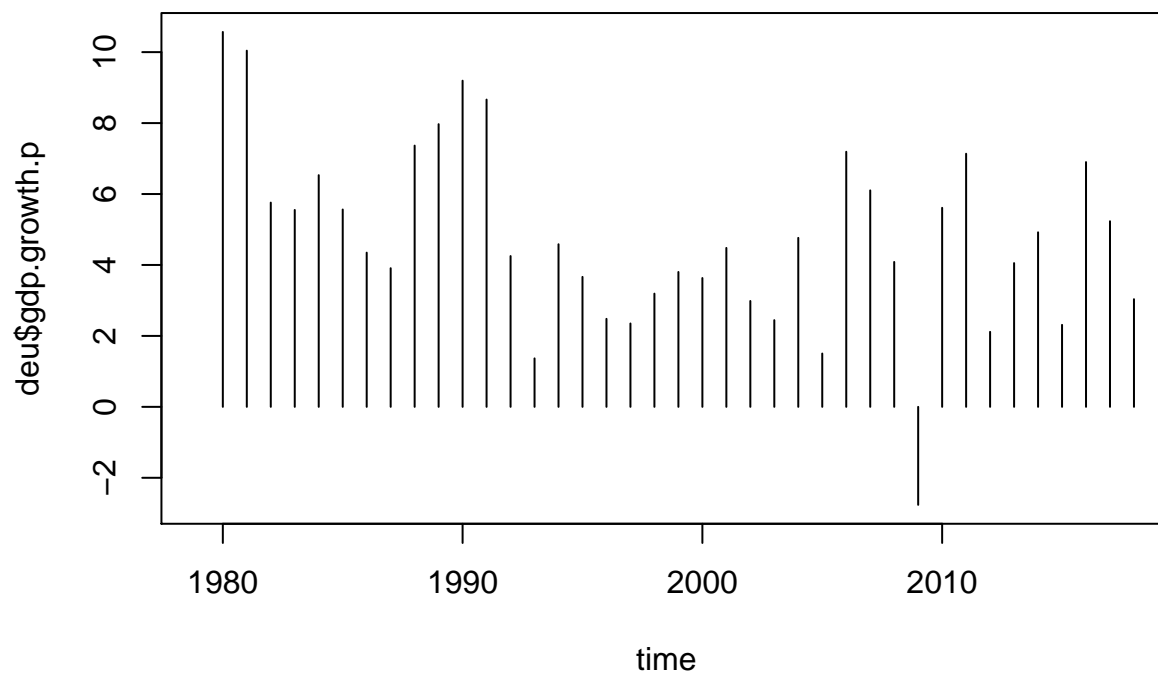


## Theoretical Quantiles

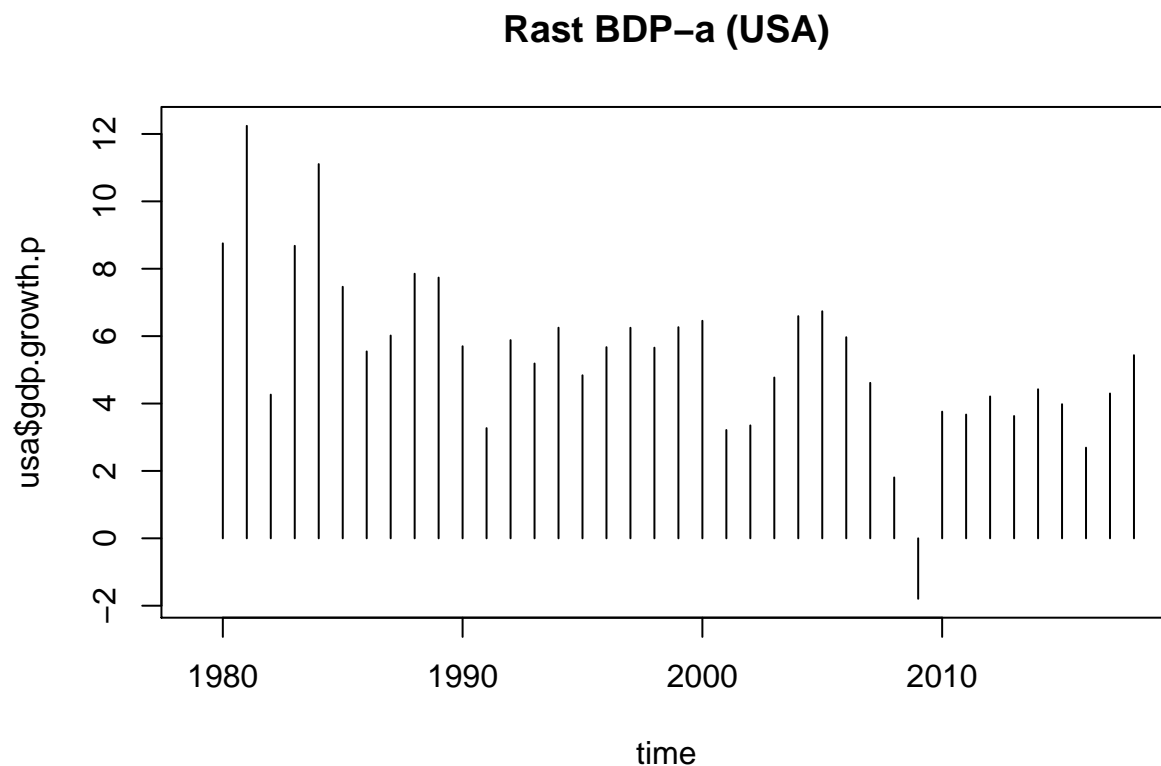
Imamo dovoljno podataka da iz gornjih prikaza možemo pretpostaviti normalnost distribucije postotnog porasta BDP-a po stanovniku.

```
plot(time, deu$gdp.growth.p, type = "h", main = "Rast BDP-a (DEU)")
```

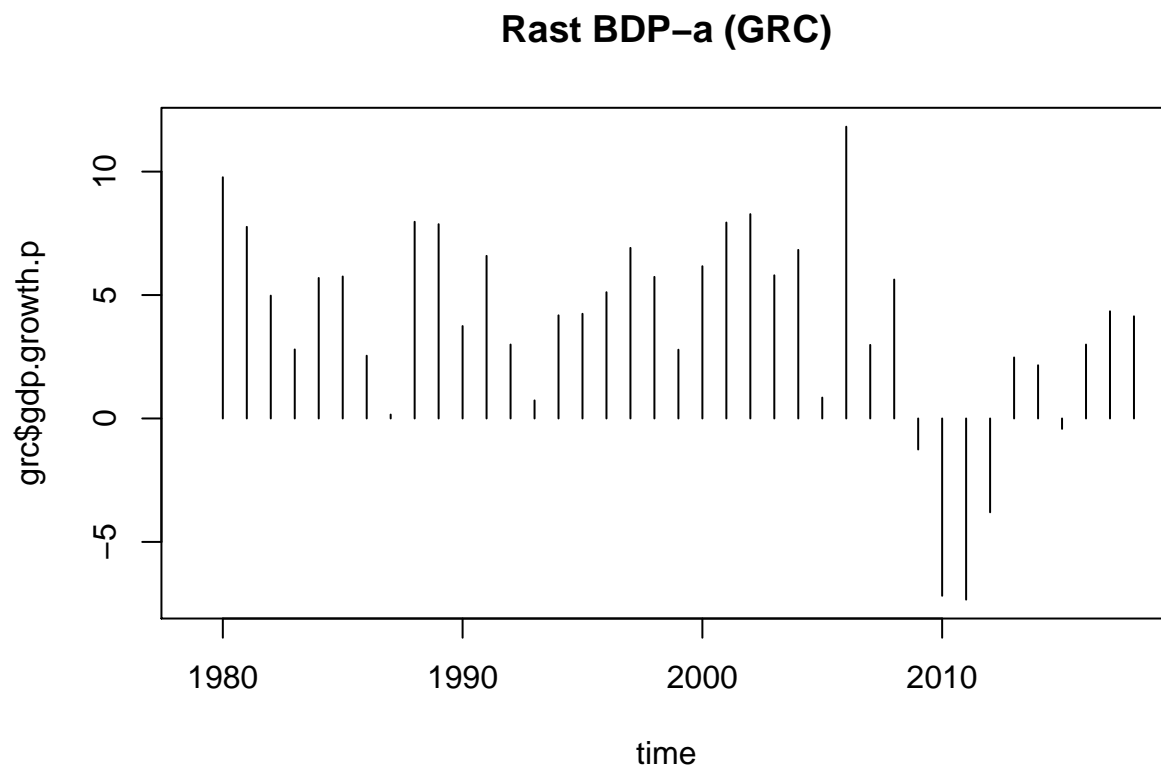
## Rast BDP-a (DEU)



```
plot(time, usa$gdp.growth.p, type="h", main = "Rast BDP-a (USA)")
```



```
plot(time, grc$gdp.growth.p, type="h", main = "Rast BDP-a (GRC)")
```



Njemačke i SAD-a je u stalnom porastu uz fluktuacije, a jedini pad BDP-a koji primjećujemo vezan je uz gospodarsku krizu 2009. godine, kada primjećujemo i značajne padove u uvozu i izvozu. BDP Grčke ima

veće fluktuacije i veća opadanja nakon 2009. godine. Rast BDP-a po stanovniku neće se puno razlikovati, a sve razlike ovisit će o promjenama u broju stanovnika.

## Testiranje hipoteza

**Pretpostavka: Rast izvoza značajno je veći od rasta uvoza za neku državu**

```
library(reshape2)
```

```
##
```

```
## Attaching package: 'reshape2'
```

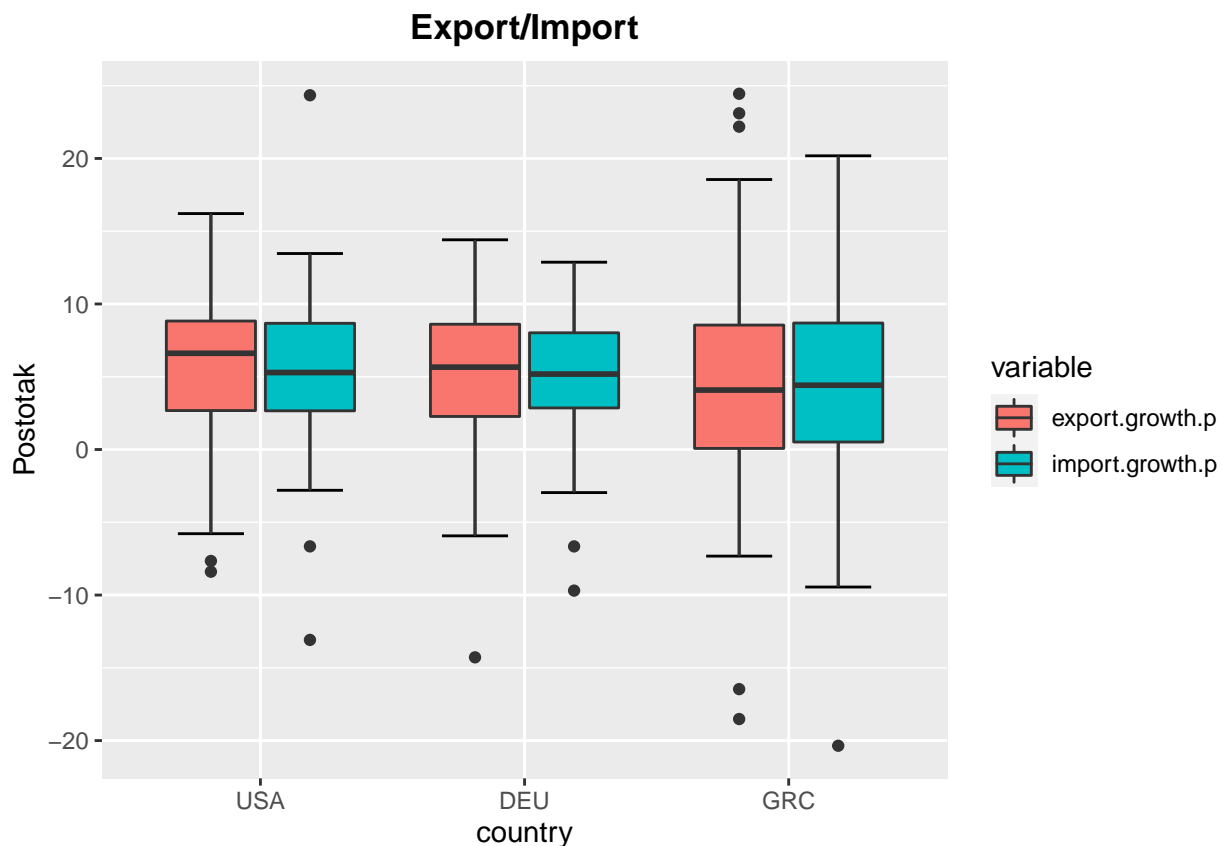
```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
## smiths
```

```
exp.imp = melt(data.all, id.vars = "country", measure.vars = c("export.growth.p", "import.growth.p"), na.rm = TRUE)
```

```
ggplot(exp.imp, aes(x = country, y = value, fill = variable)) +  
  stat_boxplot(geom = "errorbar", width = 0.5, position = position_dodge(0.75)) +  
  geom_boxplot() +  
  labs(title = "Export/Import", y = "Postotak") +  
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
```



Pogledom na gornji boxplot čini se da se uvoz i izvoz za neku državu ne razlikuju previše. Jedinu značajniju razliku vidimo za USA. Proverit ćemo je li to statistički značajno pomoću t-testa.

Prvo ćemo proveriti jednakost varijanci, ako su jednake moći ćemo koristiti inačicu t-testa sa većom snagom. Jednakost varijanci proveravamo F-testom uz razinu značajnosti  $\alpha = 0.05$ . Za F-test postavljamo sljedeće hipoteze:



$H_0$  : Omjer varijanci = 1

$H_1$  : Omjer varijanci  $\neq$  1

```
var.test(usa$export.growth.p, usa$import.growth.p, alternative = "two.sided",
         na.action = na.omit)
```

```
##
## F test to compare two variances
##
## data:  usa$export.growth.p and usa$import.growth.p
## F = 0.74877, num df = 38, denom df = 38, p-value = 0.3765
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3926431 1.4279106
## sample estimates:
## ratio of variances
##          0.7487718
```

P vrijednost testa jednakosti varijanci je veća od razine značajnosti, te pokazuje da se podaci više priklanjaju hipotezi  $H_0$ , koju ne odbacujemo.

Uz pretpostavku jednakosti varijanci postaviti ćemo hipoteze za t-test jednakosti sredina:

$H_0 : \mu_{izvoz} = \mu_{uvoz}$

$H_1 : \mu_{izvoz} > \mu_{uvoz}$

Razina značajnosti  $\alpha = 0.05$ .

```
t.test(usa$export.growth.p, usa$import.growth.p, alternative = "greater", var.equal = TRUE,
       na.action = na.omit)
```

```
##
## Two Sample t-test
##
## data:  usa$export.growth.p and usa$import.growth.p
## t = -0.29578, df = 76, p-value = 0.6159
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -2.640559      Inf
## sample estimates:
## mean of x mean of y
##  5.238851  5.637139
```

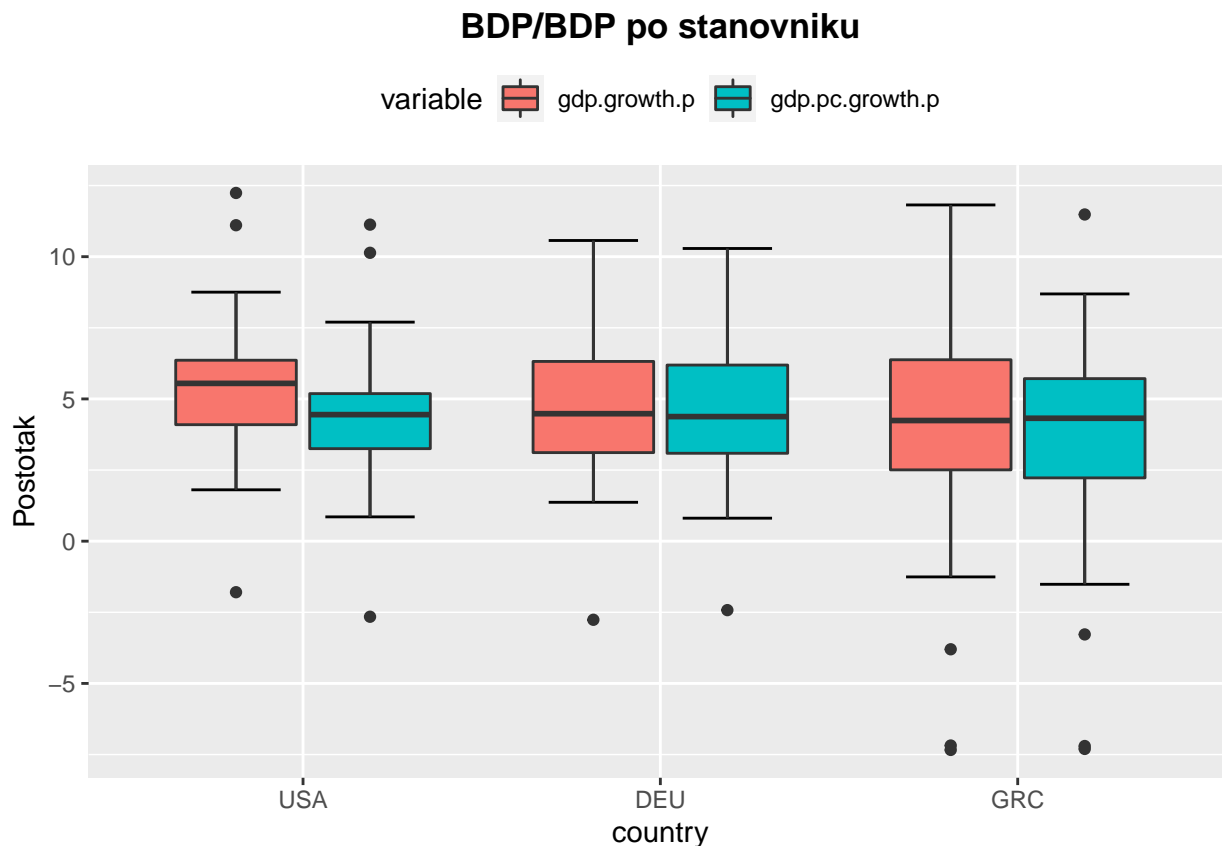
Dobivena p vrijednost testa je veća od razine značajnosti, te ne možemo odbaciti  $H_0$ .

**Zaključak:** Uz dane podatke pretpostavka nije ispunjena, tj. ne možemo pokazati da se rast izvoza neke države značajno razlikuje od rasta uvoza.

**Pretpostavka:** Rast BDP-a značajno je veći od rasta BDP-a po stanovniku neke države

```
gdp.gdppc = melt(data.all, id.vars = "country",
                 measure.vars = c("gdp.growth.p", "gdp.pc.growth.p"), na.rm = T)
ggplot(gdp.gdppc, aes(x = country, y = value, fill = variable)) +
```

```
stat_boxplot(geom = "errorbar", width = 0.5, position = position_dodge(0.75)) +
geom_boxplot() +
labs(title = "BDP/BDP po stanovniku", y="Postotak") +
theme(plot.title = element_text(hjust = 0.5, face = "bold"), legend.position = "top")
```



Iz gornjeg prikaza ne čini se kao da se BDP i BDP po stanovniku određene države pretjerano razlikuju. Jedina primjetljivija razlika koju je kod SAD-a. Provjerit ćemo statističku značajnost pomoću t-testa.

Prvo ćemo provjeriti jednakost varijanci F-testom uz razinu značajnosti  $\alpha = 0.05$ . Za F-test postavljamo sljedeće hipoteze:

$$H_0 : \text{Omjer varijanci} = 1$$

$$H_1 : \text{Omjer varijanci} \neq 1$$

```
var.test(usa$gdp.growth.p, usa$gdp.pc.growth.p, alternative = "two.sided", na.action = na.omit)
```

```
##
## F test to compare two variances
##
## data: usa$gdp.growth.p and usa$gdp.pc.growth.p
## F = 1.045, num df = 38, denom df = 38, p-value = 0.8927
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.5480025 1.9929003
## sample estimates:
## ratio of variances
## 1.045043
```

Rezultati F-testa pokazuju da se podaci više priklanjaju hipotezi  $H_0$  te ju ne odbacujemo. Dakle, uz pretpostavku jednakosti varijanci provest ćemo t-test jednakosti sredina razine značajnosti  $\alpha = 0.05$  s hipotezama:

$$H_0 : \mu_{BDP} = \mu_{BDP_{\text{stanovniku}}}$$

$$H_1 : \mu_{BDP} > \mu_{BDP_{\text{stanovniku}}}$$

```
t.test(usa$gdp.growth.p, usa$gdp.pc.growth.p, alternative = "greater", var.equal = TRUE,
       na.action = na.omit)
```

```
##
## Two Sample t-test
##
## data: usa$gdp.growth.p and usa$gdp.pc.growth.p
## t = 1.8417, df = 76, p-value = 0.03471
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.09674162      Inf
## sample estimates:
## mean of x mean of y
##  5.447202  4.438204
```

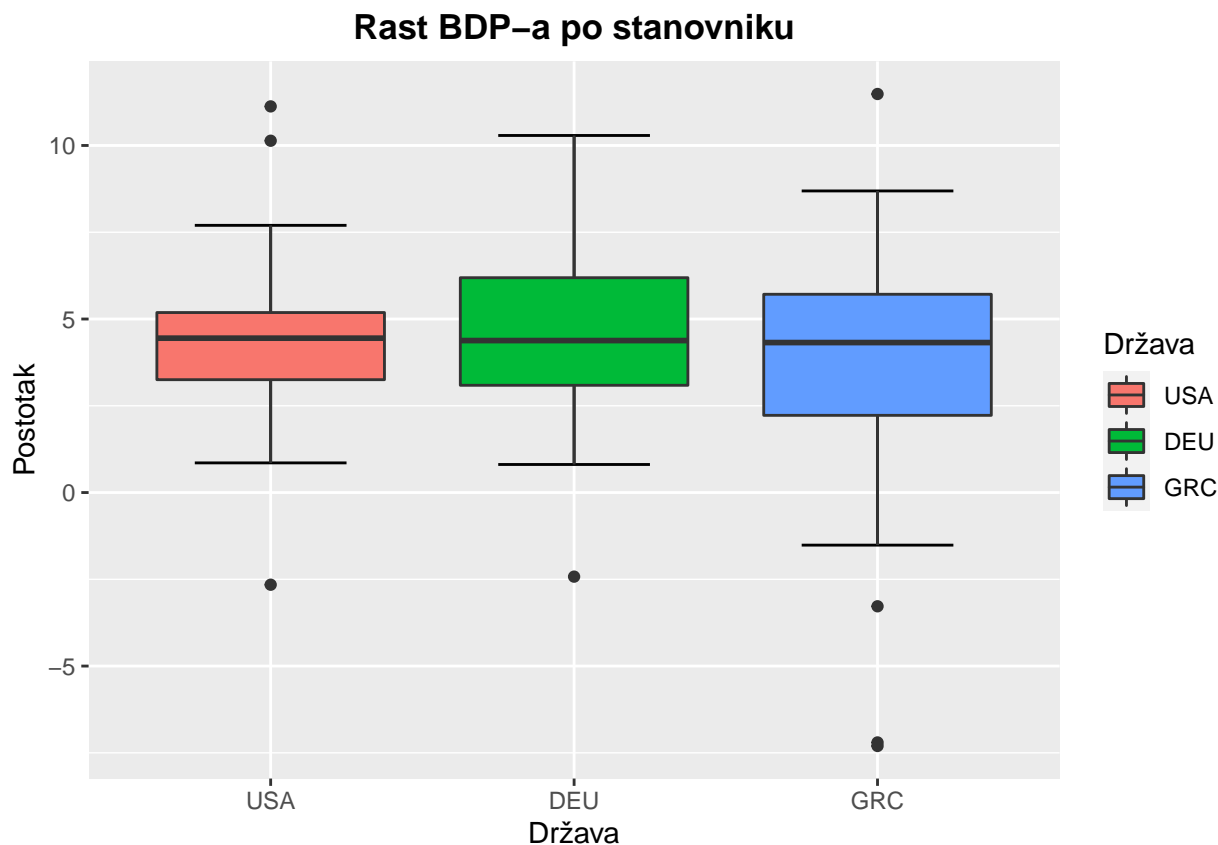
P vrijednost t-testa manja je od razine značajnosti te odbacujemo  $H_0$ .

**Zaključak:** Uz dane podatke pretpostavka je ispunjena, tj. možemo pokazati da se rast BDP-a neke države značajno razlikuje od rasta BDP-a po stanovniku. U ovom slučaju radi se o SAD-u. Interpretacija? Vidimo da se BDP i BDP po stanovniku značajnije razlikuju samo za države koje u promatranom vremenskom razdoblju imaju nekakav porast stanovništva, što je ovdje slučaj s SAD-om dok se za Njemačku i Grčku BDP i BDP po stanovniku značajno ne razlikuju. Iz ovoga zaključujemo da je BDP po stanovniku smislenije gledati u slučaju kada u promatranom razdoblju imamo porast stanovništva.

Zato u idućem testu koristimo BDP po stanovniku kao pokazatelj rasta gospodarstva.

**Pretpostavka:** Prosječni rast gospodarstva neke države značajno je veći u odnosu na druge

```
ggplot(na.omit(data.all), aes(x=country, y=gdp.pc.growth.p)) +
  stat_boxplot(geom = "errorbar", width = 0.5) +
  geom_boxplot(aes(fill=country)) +
  labs(title = "Rast BDP-a po stanovniku", x="Država", y="Postotak", fill="Država") +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
```



gornjeg prikaza se vidi da se prosječni porasti BDP-a po stanovniku ove tri države ne razlikuju previše. To možemo pokušati provjeriti ANOVA metodom. Pretpostavke ANOVA-e su nezavisnost podataka, normalna distribucija i homogenost varijanci, pa ćemo homogenost varijanci provjeriti Bartletovim testom:

$$H_0 : \sigma_{USA}^2 = \sigma_{DEU}^2 = \sigma_{GRC}^2$$

$$H_1 : \neg H_0.$$

razine značajnosti  $\alpha = 0.05$ .

```
bartlett.test(gdp.pc.growth.p ~ country, data.all)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  gdp.pc.growth.p by country
## Bartlett's K-squared = 10.982, df = 2, p-value = 0.004124
```

Dobivena p vrijednost značajno je manja od razine značajnosti što znači da se odbacuje  $H_0$  pa ne možemo koristiti ANOVA-u.

Umjesto ANOVA-e provest ćemo neparametarski test, Kruskal-Wallis test razine značajnosti  $\alpha = 0.05$  za koji pretpostavke zahtjevano parametarskim testovima ne moraju biti ispunjene. Kruskal-Wallis test slabiji je od ANOVA-e i uspoređuje medijane ali ovaj test u kombinaciji s gornjim prikazom dokazat će približnu jednakost porasta gospodarstva država.

$$H_0 : M_{USA} = M_{DEU} = M_{GRC}$$

$$H_1 : \neg H_0.$$

```
kruskal.test(gdp.pc.growth.p ~ country, data.all)
```

```
##  
##  Kruskal-Wallis rank sum test  
##  
## data:  gdp.pc.growth.p by country  
## Kruskal-Wallis chi-squared = 0.89954, df = 2, p-value = 0.6378
```

Dobivena p vrijednost veća je od razine značajnosti testa te zaključujemo da se porasti BDP-a značajno ne razlikuju, podjednaki su kao što se vidi iz boxplota. Ne odbacujemo  $H_0$ .