

Projekt iz kolegija Statistička analiza podataka

KGB (Ivo Karamatić, Vjeran Grozdanić, Jelena Bratulić)

20. svibnja 2020.

Uvod

U sklopu projekta iz kolegija Statistička analiza podataka promatrat ćemo povezanost i utjecaj izvoza i uvoza na gospodarstvo pojedine države. Za sam rad i analize koristit ćemo podatke Organizacije za ekonomsku suradnju i razvoj (OECD) koji su javno dostupni na službenim stranicama organizacije. Skup podataka sastoji se od niza podataka o izvozu, uvozu, bruto domaćem proizvodu te bruto domaćem proizvodu po glavi stanovnika za razdoblje od 1979.godine do 2019.godine. Za daljnje analize promatrat ćemo detaljnije četiri države: Australiju, Koreju, Norvešku i Sjedinjene Američke Države.

Učitavanje i transformacija podataka

Podaci se nalaze u zasebnim csv datotekama. Dodatno ćemo u analizama koristiti i izvedene varijable iz prethodno spomenutih podataka pa je potrebno izračunati vrijednosti rasta varijabli. Godišnji rast BDP-a mjerimo prema sljedećoj formuli: $rastBDP = \frac{BDP_N - BDP_{N-1}}{BDP_{N-1}}$. Varijabla BDP_N predstavlja vrijednost BDP-a u tekućoj godini, a varijabla BDP_{N-1} predstavlja vrijednost BDP-a u prethodnoj godini. Dijeljenje s vrijednošću iz prethodne godine nužno je zbog inflacije. Kada bismo riječima trebali objasniti značenje ove varijable, okvirna definicija mogla bi glasiti: Godišnji rast BDP-a je razlika između BDP-a u tekućoj godini i BDP-a u prethodnoj godini prilagođena monetarnoj situaciji iz prethodne godine. Analogno vrijedi i za rast izvoza, uvoza te BDP-a po glavi stanovnika.

```
rm(list=ls()) # cisti sve lokalne varijable

export.data <- read_csv("../data/Export_data.csv")

## Parsed with column specification:
## cols(
##   LOCATION = col_character(),
##   INDICATOR = col_character(),
##   SUBJECT = col_character(),
##   MEASURE = col_character(),
##   FREQUENCY = col_character(),
##   TIME = col_double(),
##   Value = col_double(),
##   `Flag Codes` = col_character()
## )

import.data <- read_csv("../data/Import_data.csv")

## Parsed with column specification:
## cols(
##   LOCATION = col_character(),
##   INDICATOR = col_character(),
##   SUBJECT = col_character(),
##   MEASURE = col_character(),
```

```

##   FREQUENCY = col_character(),
##   TIME = col_double(),
##   Value = col_double(),
##   `Flag Codes` = col_character()
## )

gdp.data <- read_csv("../data/GDP_data.csv")

## Parsed with column specification:
## cols(
##   LOCATION = col_character(),
##   INDICATOR = col_character(),
##   SUBJECT = col_character(),
##   MEASURE = col_character(),
##   FREQUENCY = col_character(),
##   TIME = col_double(),
##   Value = col_double(),
##   `Flag Codes` = col_character()
## )

gdp.per.capita.data <- read_csv("../data/GDPpercapita_data.csv")

## Parsed with column specification:
## cols(
##   LOCATION = col_character(),
##   INDICATOR = col_character(),
##   SUBJECT = col_character(),
##   MEASURE = col_character(),
##   FREQUENCY = col_character(),
##   TIME = col_double(),
##   Value = col_double(),
##   `Flag Codes` = col_character()
## )

# sve drzave u podacima vec imaju iste godine, 1979 - 2018

export.data %>% filter(LOCATION=="AUS" | LOCATION=="NOR" | LOCATION=="USA" | LOCATION=="KOR") %>% select(LOCATION, TIME, Value)
import.data %>% filter(LOCATION=="AUS" | LOCATION=="NOR" | LOCATION=="USA" | LOCATION=="KOR") %>% select(LOCATION, TIME, Value)
gdp.data %>% filter(LOCATION=="AUS" | LOCATION=="NOR" | LOCATION=="USA" | LOCATION=="KOR") %>% select(LOCATION, TIME, Value)
gdp.per.capita.data %>% filter(LOCATION=="AUS" | LOCATION=="NOR" | LOCATION=="USA" | LOCATION=="KOR") %>% select(LOCATION, TIME, Value)

# spajanje svih podataka u jednu tablicu

sap.data <- inner_join(export.data.filter, import.data.filter, by = c("LOCATION" = "LOCATION", "TIME" = "TIME"))
sap.data <- inner_join(sap.data, gdp.data.filter, by = c("LOCATION" = "LOCATION", "TIME" = "TIME"))
sap.data <- inner_join(sap.data, gdp.per.capita.data.filter, by = c("LOCATION" = "LOCATION", "TIME" = "TIME"))

rm(list=setdiff(ls(), "sap.data")) # cisti sve varijable osim sap.data koja ce biti koristena za daljnje radove

# izracun rasta

sap.data <- mutate(sap.data, EXPORT_GROWTH = ifelse(LOCATION == lag(LOCATION), (EXPORT - lag(EXPORT)) / lag(EXPORT), 1))
sap.data <- mutate(sap.data, IMPORT_GROWTH = ifelse(LOCATION == lag(LOCATION), (IMPORT - lag(IMPORT)) / lag(IMPORT), 1))
sap.data <- mutate(sap.data, GDP_GROWTH = ifelse(LOCATION == lag(LOCATION), (GDP - lag(GDP)) / lag(GDP), 1))
sap.data <- mutate(sap.data, GDP_PER_CAPITA_GROWTH = ifelse(LOCATION == lag(LOCATION), (GDP_PER_CAPITA - lag(GDP_PER_CAPITA)) / lag(GDP_PER_CAPITA), 1))

```

```

# izracun prosjeka i standardne devijacije

sap.data.summary <- sap.data %>% group_by(LOCATION) %>% summarise(EXPORT_MEAN=mean(EXPORT, na.rm=T), EXI

# izracun dodatnih varijabli za testiranja

aus_test_data = sap.data %>% filter(LOCATION=="AUS")
kor_test_data = sap.data %>% filter(LOCATION=="KOR")
nor_test_data = sap.data %>% filter(LOCATION=="NOR")
usa_test_data = sap.data %>% filter(LOCATION=="USA")

e = rep("E", 39)
i = rep("I", 39)
f = rep("F", 39)
s = rep("S", 39)

aus_ei_test = data.frame(na.omit(aus_test_data$EXPORT_GROWTH), e)
ii = data.frame(na.omit(aus_test_data$IMPORT_GROWTH), i)
names(ii) = names(aus_ei_test)
aus_ei_test = rbind(aus_ei_test, ii)

kor_ei_test = data.frame(na.omit(kor_test_data$EXPORT_GROWTH), e)
ii = data.frame(na.omit(kor_test_data$IMPORT_GROWTH), i)
names(ii) = names(kor_ei_test)
kor_ei_test = rbind(kor_ei_test, ii)

nor_ei_test = data.frame(na.omit(nor_test_data$EXPORT_GROWTH), e)
ii = data.frame(na.omit(nor_test_data$IMPORT_GROWTH), i)
names(ii) = names(nor_ei_test)
nor_ei_test = rbind(nor_ei_test, ii)

usa_ei_test = data.frame(na.omit(usa_test_data$EXPORT_GROWTH), e)
ii = data.frame(na.omit(usa_test_data$IMPORT_GROWTH), i)
names(ii) = names(usa_ei_test)
usa_ei_test = rbind(usa_ei_test, ii)

aus_gdp_test = data.frame(na.omit(aus_test_data$GDP_GROWTH), e)
ii = data.frame(na.omit(aus_test_data$GDP_PER_CAPITA_GROWTH), i)
names(ii) = names(aus_gdp_test)
aus_gdp_test = rbind(aus_gdp_test, ii)

kor_gdp_test = data.frame(na.omit(kor_test_data$GDP_GROWTH), e)
ii = data.frame(na.omit(kor_test_data$GDP_PER_CAPITA_GROWTH), i)
names(ii) = names(kor_gdp_test)
kor_gdp_test = rbind(kor_gdp_test, ii)

nor_gdp_test = data.frame(na.omit(nor_test_data$GDP_GROWTH), e)
ii = data.frame(na.omit(nor_test_data$GDP_PER_CAPITA_GROWTH), i)
names(ii) = names(nor_gdp_test)
nor_gdp_test = rbind(nor_gdp_test, ii)

usa_gdp_test = data.frame(na.omit(usa_test_data$GDP_GROWTH), e)
ii = data.frame(na.omit(usa_test_data$GDP_PER_CAPITA_GROWTH), i)

```

```

names(ii) = names(usa_gdp_test)
usa_gdp_test = rbind(usa_gdp_test, ii)

kor_nor_test = data.frame(na.omit(kor_test_data$GDP_GROWTH), f)
ii = data.frame(na.omit(nor_test_data$GDP_GROWTH), s)
names(ii) = names(kor_nor_test)
kor_nor_test = rbind(kor_nor_test, ii)

kor_usa_test = data.frame(na.omit(kor_test_data$GDP_GROWTH), f)
ii = data.frame(na.omit(usa_test_data$GDP_GROWTH), s)
names(ii) = names(kor_usa_test)
kor_usa_test = rbind(kor_usa_test, ii)

nor_usa_test = data.frame(na.omit(nor_test_data$GDP_GROWTH), f)
ii = data.frame(na.omit(usa_test_data$GDP_GROWTH), s)
names(ii) = names(nor_usa_test)
nor_usa_test = rbind(nor_usa_test, ii)

gdp_g_all_states = data.frame(na.omit(aus_test_data$GDP_GROWTH), e)
ii = data.frame(na.omit(kor_test_data$GDP_GROWTH), i)
names(ii) = names(gdp_g_all_states)
gdp_g_all_states = rbind(gdp_g_all_states, ii)
ii = data.frame(na.omit(nor_test_data$GDP_GROWTH), f)
names(ii) = names(gdp_g_all_states)
gdp_g_all_states = rbind(gdp_g_all_states, ii)
ii = data.frame(na.omit(usa_test_data$GDP_GROWTH), s)
names(ii) = names(gdp_g_all_states)
gdp_g_all_states = rbind(gdp_g_all_states, ii)

colnames(aus_ei_test) = c("NUMBERS", "MARKER")
colnames(kor_ei_test) = c("NUMBERS", "MARKER")
colnames(nor_ei_test) = c("NUMBERS", "MARKER")
colnames(usa_ei_test) = c("NUMBERS", "MARKER")
colnames(kor_nor_test) = c("NUMBERS", "MARKER")
colnames(kor_usa_test) = c("NUMBERS", "MARKER")
colnames(nor_usa_test) = c("NUMBERS", "MARKER")
colnames(gdp_g_all_states) = c("NUMBERS", "MARKER")

e = rep("E", 40)
i = rep("I", 40)
f = rep("F", 40)
s = rep("S", 40)

aus2_ei_test = data.frame(na.omit(aus_test_data$EXPORT), e)
ii = data.frame(na.omit(aus_test_data$IMPORT), i)
names(ii) = names(aus2_ei_test)
aus2_ei_test = rbind(aus2_ei_test, ii)

kor2_ei_test = data.frame(na.omit(kor_test_data$EXPORT), e)
ii = data.frame(na.omit(kor_test_data$IMPORT), i)
names(ii) = names(kor2_ei_test)
kor2_ei_test = rbind(kor2_ei_test, ii)

```

```

nor2_ei_test = data.frame(na.omit(nor_test_data$EXPORT), e)
ii = data.frame(na.omit(nor_test_data$IMPORT), i)
names(ii) = names(nor2_ei_test)
nor2_ei_test = rbind(nor2_ei_test, ii)

usa2_ei_test = data.frame(na.omit(usa_test_data$EXPORT), e)
ii = data.frame(na.omit(usa_test_data$IMPORT), i)
names(ii) = names(usa2_ei_test)
usa2_ei_test = rbind(usa2_ei_test, ii)

colnames(aus2_ei_test) = c("NUMBERS", "MARKER")
colnames(kor2_ei_test) = c("NUMBERS", "MARKER")
colnames(nor2_ei_test) = c("NUMBERS", "MARKER")
colnames(usa2_ei_test) = c("NUMBERS", "MARKER")

colnames(aus_gdp_test) = c("NUMBERS", "MARKER")
colnames(kor_gdp_test) = c("NUMBERS", "MARKER")
colnames(nor_gdp_test) = c("NUMBERS", "MARKER")
colnames(usa_gdp_test) = c("NUMBERS", "MARKER")

```

Pogled s ekonomskog aspekta

U ekonomiji, snaga gospodarstva neke države mjeri se uglavnom BDP-om (Bruto domaćim proizvodom). To je ekonomski faktor koji se najčešće računa prema sljedećoj formuli: $Y = C + I + G + X$. U upravo navedenoj formuli Y označava BDP, C označava osobnu potrošnju, I označava investicije, G označava državnu potrošnju, a X označava neto izvoz. Neto izvoz je definiran sljedećom formulom: $X = \text{ukupanIzvoz} - \text{ukupanUvoz}$. Dakle, možemo zaključiti da matematička veza između uvoza neke države, izvoza neke države i njenog gospodarstva svakako postoji.

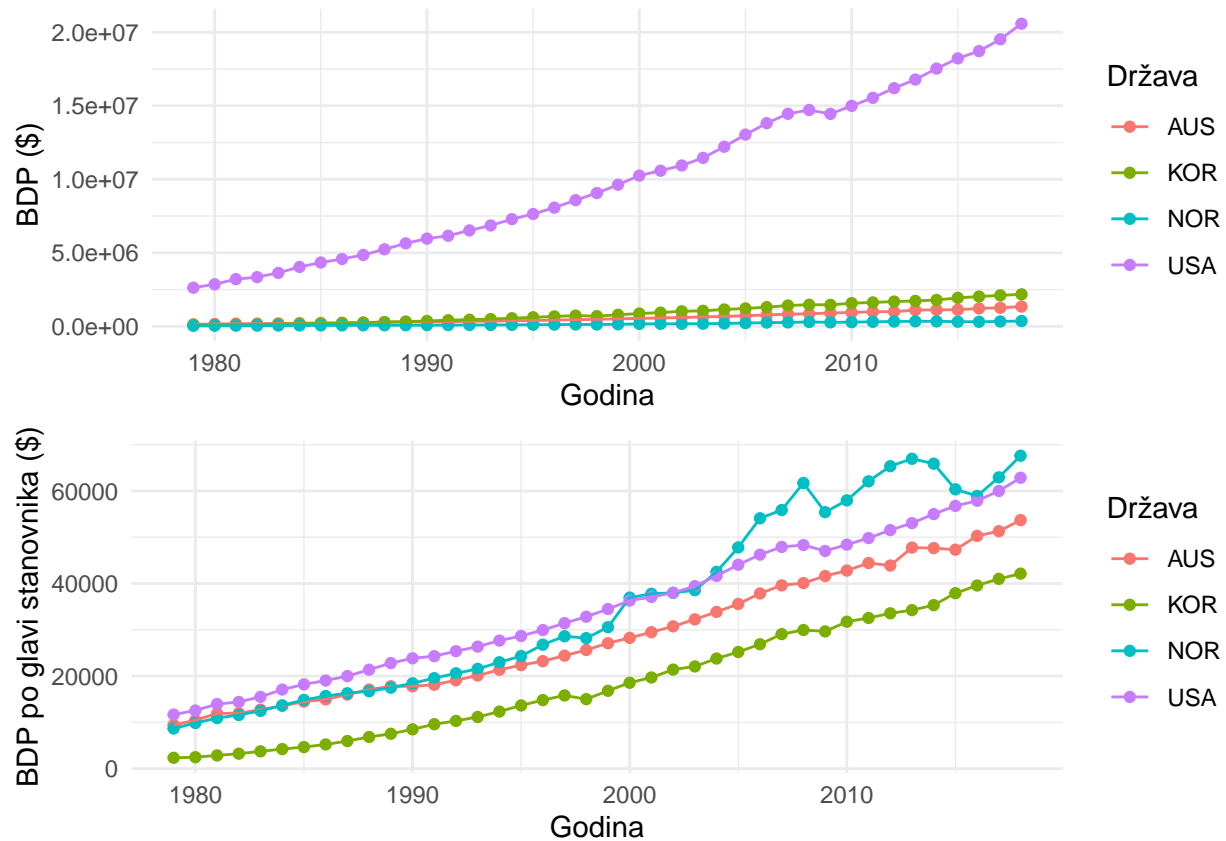
S obzirom da na BDP, osim uvoza i izvoza, utječu i potrošačke navike društva, logično je za očekivati da će ukupni BDP kod država s većim brojem stanovnika biti uglavnom veći te da bismo, kako bismo imali dobru predodžbu gospodarske snage neke države u odnosu na njeno stanovništvo, trebali uspoređivati BDP po glavi stanovnika. Pogledajmo malu usporedbu između SAD-a, Južne Koreje, Norveške i Australije. Na prvom grafu imamo vrijednosti BDP-a u periodu od 1979. do 2018. godine. Gledajući samo taj graf, moglo bi se zaključiti da je osoba u SAD-u nemjerljivo bogatija i ekonomski prosperitetnija u odnosu na ljude u preostalim zemljama. Međutim, pogledamo li drugi graf koji prikazuje BDP po glavi stanovnika u istom vremenskom periodu, vidimo da su ekonomije ostalih zemalja ne samo usporedive, već u nekoliko navrata i bolje od ekonomije SAD-a.

```

p1 = ggplot(sap.data, aes(TIME, colour = LOCATION)) + geom_point(aes(y = GDP)) + geom_line(aes(y = GDP))

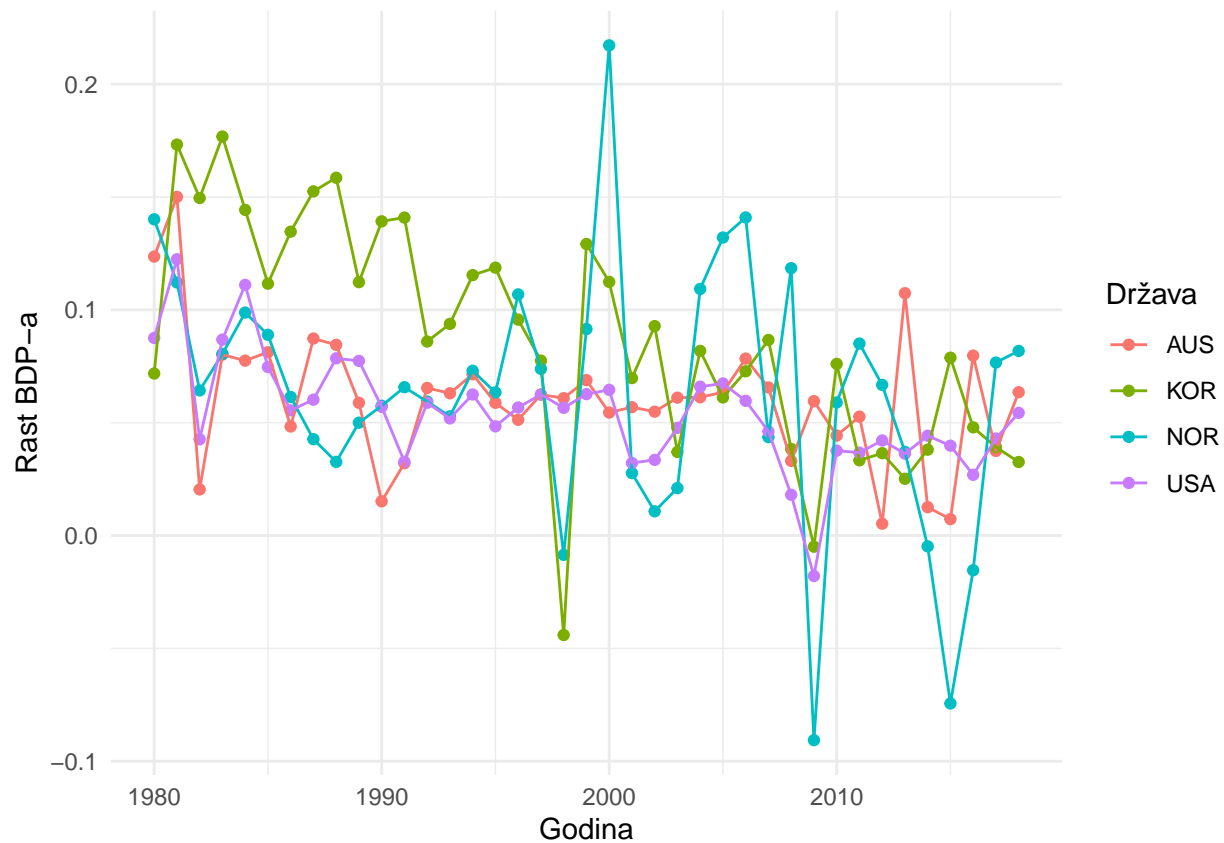
p2 = ggplot(sap.data, aes(TIME, colour = LOCATION)) + geom_point(aes(y = GDP_PER_CAPITA)) + geom_line(aes(y = GDP_PER_CAPITA))
plot_grid(p1, p2, ncol= 1)

```



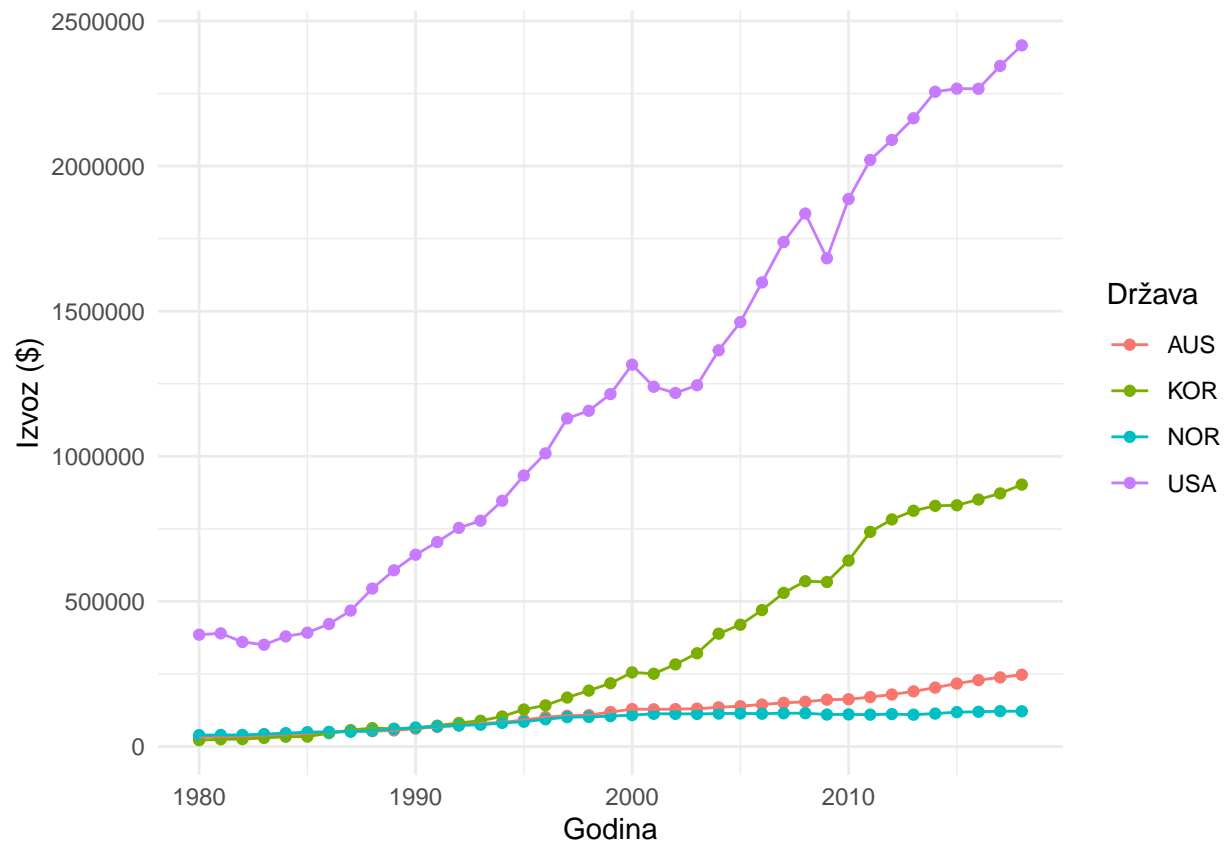
Nadalje, možemo vizualizirati i kretanje gospodarskog rasta kroz navedne godine. Definiciju gospodarskog rasta dali smo u prethodnom odijelku, no spomenimo samo kako je gospodarski rast jedan od najvažnih pokazatelja stabilnosti i jačine gospodarstva neke države. Na sljedećem grafu pogledajmo kako se rast BDP-a u naše 4 države kretao kroz vrijeme.

```
ggplot(na.omit(sap.data), aes(TIME, colour = LOCATION)) + geom_point(aes(y = GDP_GROWTH)) + geom_line(aes(y = GDP_GROWTH))
```

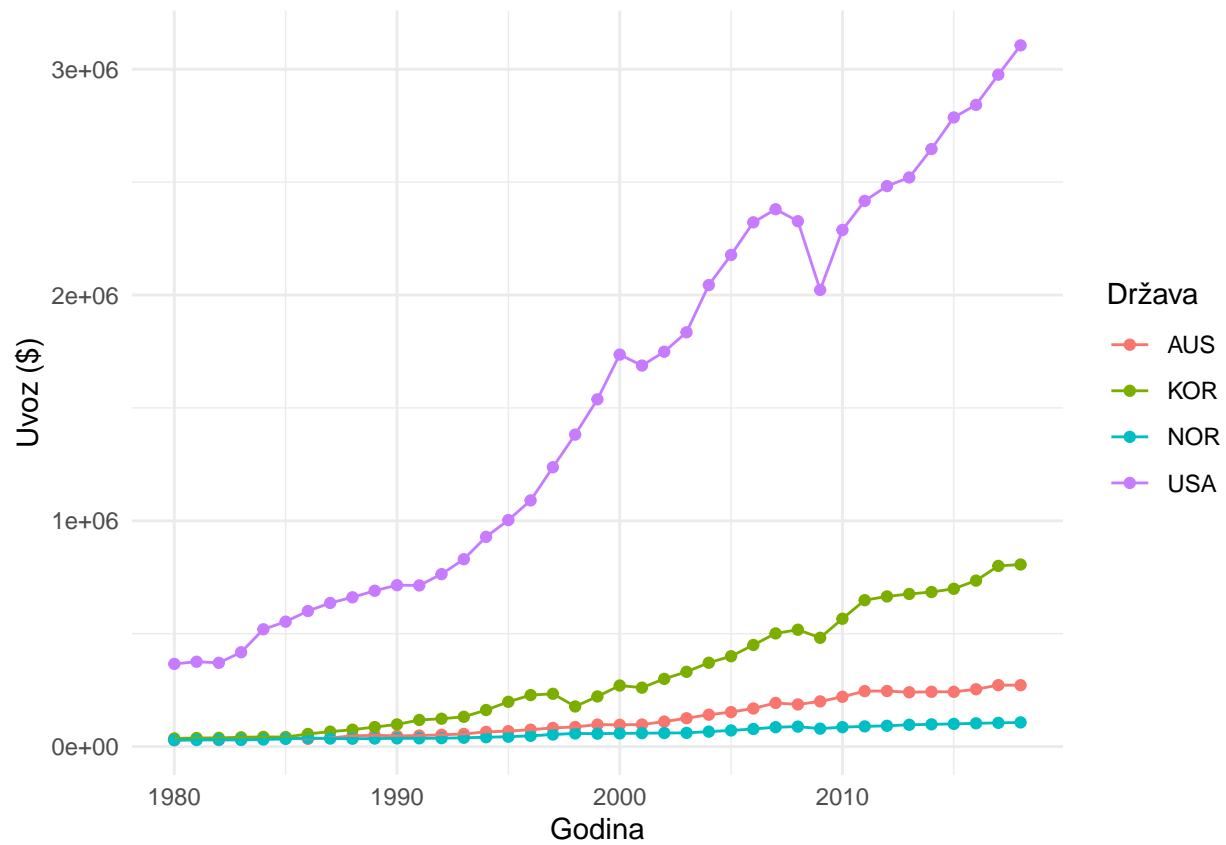


Za kraj, pogledajmo još grafove za uvoz, izvoz, rast izvoza i neto izvoz. Vidimo da je SAD najnižu vrijednost neto izvoza doživio u drugoj polovici prvog desetljeća 20. stoljeća što je ujedno vrijeme koje je prethodilo posljednjoj velikoj ekonomskoj krizi. Također, vidimo osjetan pad uvoza i izvoza SAD-a 2009. godine u vrijeme upravo navedene krize.

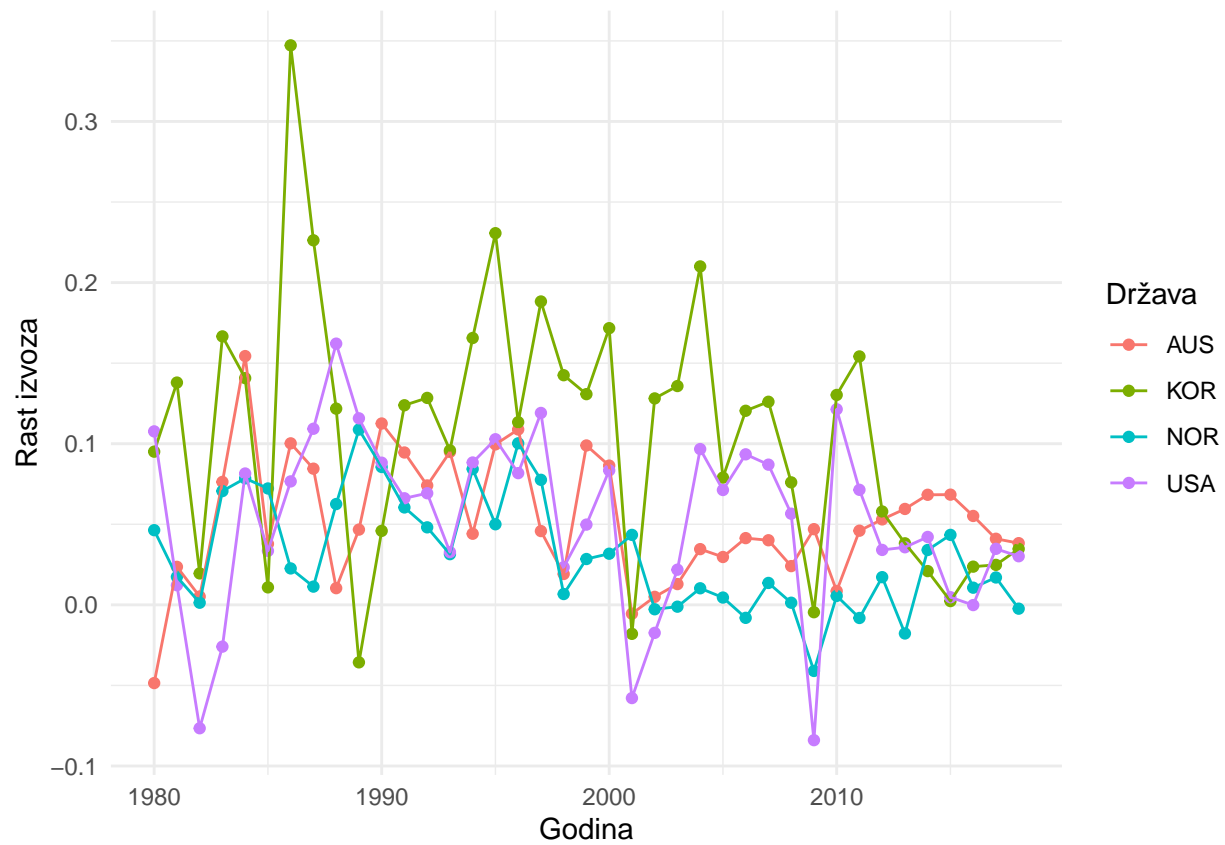
```
ggplot(na.omit(sap.data), aes(TIME, colour = LOCATION)) + geom_point(aes(y = EXPORT)) + geom_line(aes(y
```



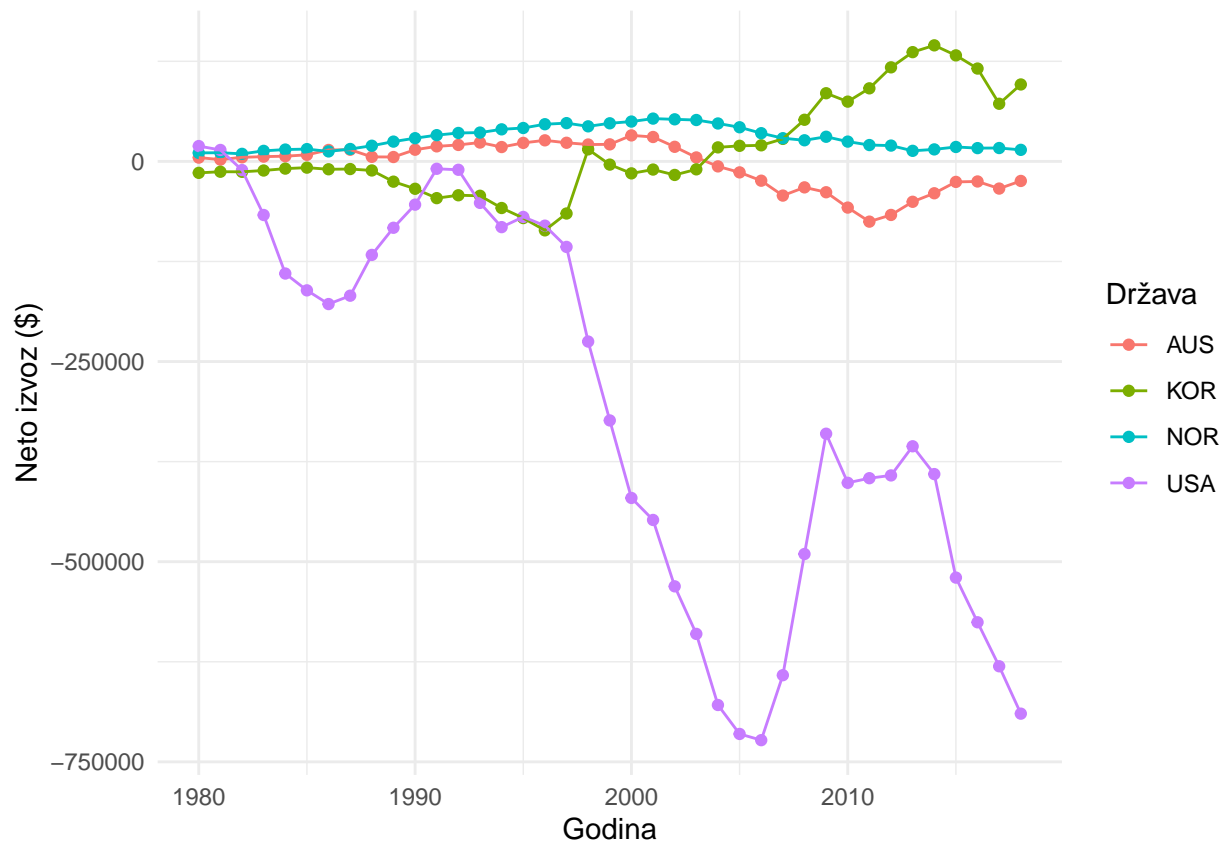
```
ggplot(na.omit(sap.data), aes(TIME, colour = LOCATION)) + geom_point(aes(y =IMPORT)) + geom_line(aes(y =
```

```
ggplot(na.omit(sap.data), aes(TIME, colour = LOCATION)) + geom_point(aes(y = EXPORT_GROWTH)) + geom_line(aes(y = EXPORT_GROWTH))
```



```
ggplot(na.omit(sap.data), aes(TIME, colour = LOCATION)) + geom_point(aes(y = EXPORT - IMPORT)) + geom_l
```

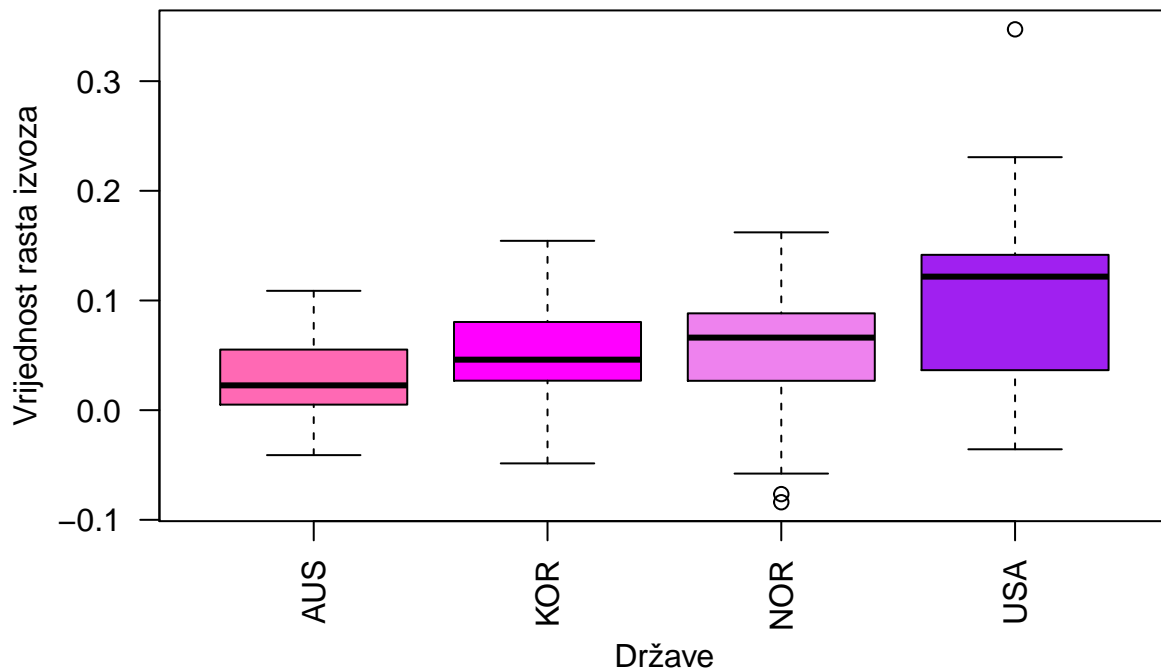


Stršeće vrijednosti

Kao što smo već spomenuli, rast BDP-a, izvoza i uvoza dobri su pokazatelji stabilnosti države i jačine njezinog gospodarstva. To je vidljivo i analiziranjem stršećih vrijednosti pomoću Box-plot grafa. Neka vrijednost je stršeća ukoliko se od središnje vrijednosti razlikuje za više od 3σ .

```
boxplot(nor_test_data$EXPORT_GROWTH, aus_test_data$EXPORT_GROWTH, usa_test_data$EXPORT_GROWTH, kor_test_data$EXPORT_GROWTH,
main = "Box-plot rasta izvoza",
at = c(1,2,3,4),
names = c("AUS", "KOR", "NOR", "USA"),
las = 2,
col = c("hot pink","magenta", "violet", "purple"),
border = "black",
xlab = "Države",
ylab = 'Vrijednost rasta izvoza')
```

Box-plot rasta izvoza

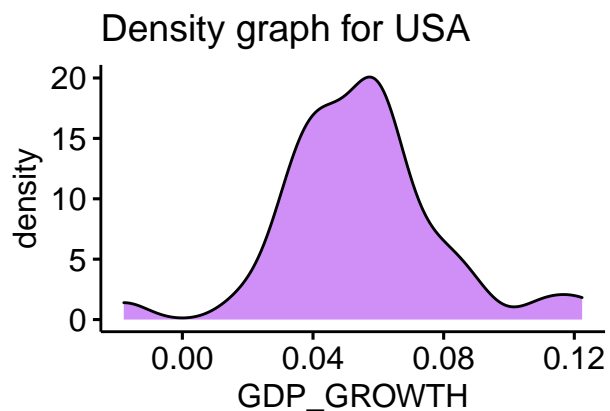
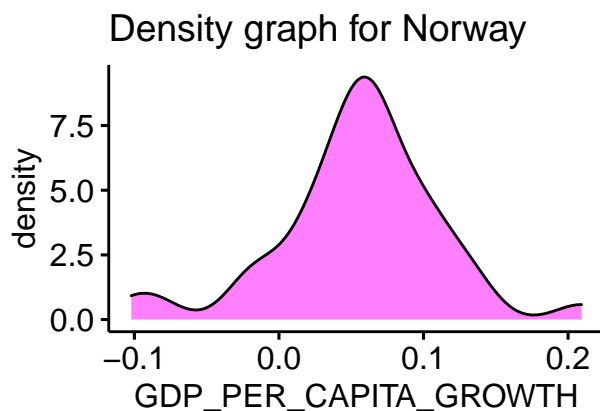
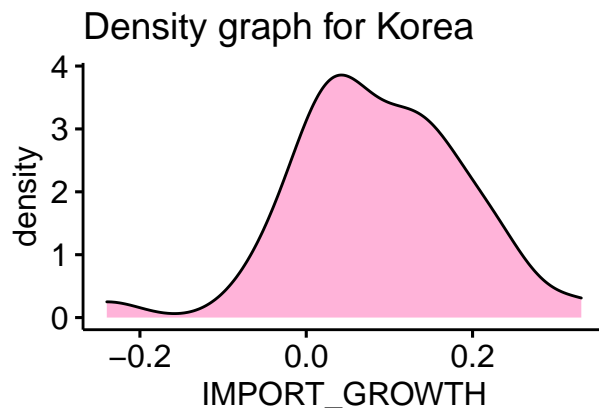
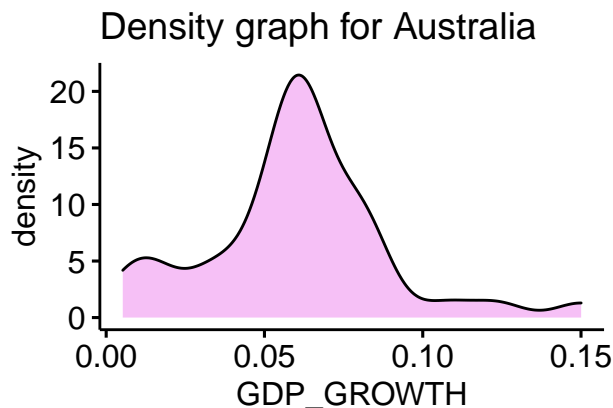


U ovom slučaju, strešće vrijednosti nam ukazuju na nepravilnosti u gospodarstvu neke države. Kod Južne Koreje i SAD-a su vidljive stršće vrijednosti koje se mogu objasniti sljedećim događajima. Južna Koreja je 1986. godine započela trgovati sa SAD-om što je u državi povećalo izvoz, a samim time možemo reći da se dogodio i veći prtok novaca nego prije. Nadalje, SAD je imala dva pada, odnosno 1982. godine te 2009. godine kada je cijeli svijet zahvatila ekonomska kriza, što se i vidljivo osjetilo u gospodarstvu SAD-a.

Grafičke provjere podataka

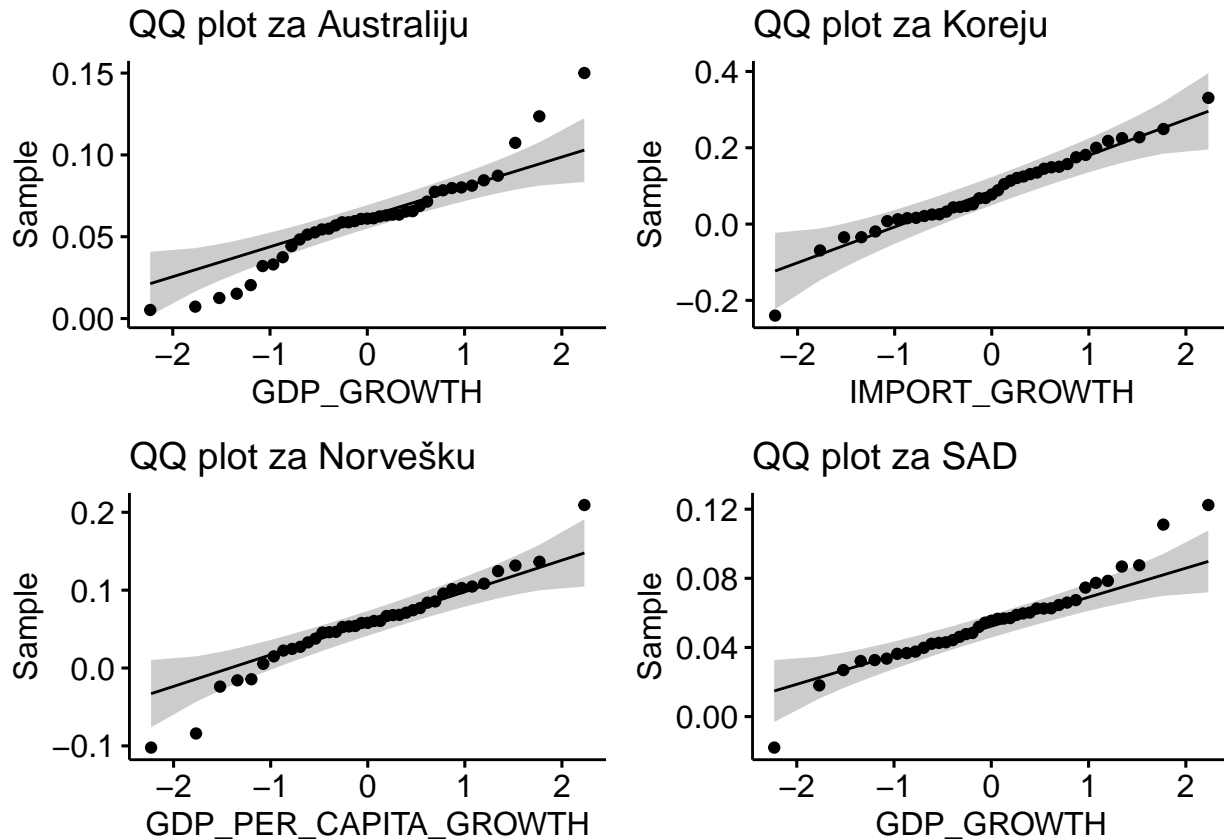
Kako bismo mogli u daljnjem radu detaljnije analizirati podatke te testirati određene hipoteze, potrebno je provjeriti imaju li svi podaci normalnu razdiobu kako bismo znali trebali li koristiti parametarske ili neparametarske testove. Iako prema graničnom centralnom teoremu, za uzorak veći od 30, distribucija nije bitna jer se može aproksimirati normalnom, želimo biti konzistentni i provjeriti distribuciju. Provjeru normalnosti, odnosno normalne distribucije neke populacije možemo provjeriti jednostavnim grafičkim prikazom. U nastavku su prikazani izabrani grafovi distribucija za rast BDP-a, rast BDP-a po glavi stanovništva, rast izvoza te rast uvoza.

```
p1 = ggdensity(na.omit(aus_test_data$GDP_GROWTH), main = "Density graph for Australia", xlab = "GDP_GROWTH")
p2 = ggdensity(na.omit(kor_test_data$IMPORT_GROWTH), main = "Density graph for Korea", xlab = "IMPORT_GROWTH")
p3 = ggdensity(na.omit(nor_test_data$GDP_PER_CAPITA_GROWTH), main = "Density graph for Norway", xlab = "GDP_PER_CAPITA_GROWTH")
p4 = ggdensity(na.omit(usa_test_data$GDP_GROWTH), main = "Density graph for USA", xlab = "GDP_GROWTH")
plot_grid(p1, p2, p3, p4)
```



Iz samih grafičkih prikaza distribucije populacije ne možemo s velikom sigurnošću odrediti ima li neka populacija normalnu razdiobu ili ne. Iako svi grafovi podsjećaju na normalnu distribuciju, većina njih ima dodatne teške repove koji remete distribuciju. Kako bismo s većom sigurnošću mogli odrediti distribuciju, koristit ćemo QQ grafove te dodatno i Shapiro test.

```
p1 = ggqqplot(na.omit(aus_test_data$GDP_GROWTH), main = "QQ plot za Australiju", xlab = "GDP_GROWTH")
p2 = ggqqplot(na.omit(kor_test_data$IMPORT_GROWTH), main = "QQ plot za Koreju", xlab = "IMPORT_GROWTH")
p3 = ggqqplot(na.omit(nor_test_data$GDP_PER_CAPITA_GROWTH), main = "QQ plot za Norvešku", xlab = "GDP_P")
p4 = ggqqplot(na.omit(usa_test_data$GDP_GROWTH), main = "QQ plot za SAD", xlab = "GDP_GROWTH")
plot_grid(p1, p2, p3, p4)
```



Shapiro testom možemo uz postavljenu razinu značajnosti od 5% biti sigurni ima li neka populacija normalnu razdiobu ili ne. Nulta hipoteza kod Shapiro testova je da uzorak ima normalnu distribuciju, dok je alternativna H_1 hipoteza kako ju nema. Hipotezu H_0 možemo odbaciti ukoliko je p-vrijednost testa manja od razine značajnosti, dakle manja od 5%.

```
shapiro.test(aus_test_data$GDP_GROWTH)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  aus_test_data$GDP_GROWTH
## W = 0.93703, p-value = 0.03038
```

```
shapiro.test(kor_test_data$IMPORT_GROWTH) # --> normalna
```

```
##
##  Shapiro-Wilk normality test
##
## data:  kor_test_data$IMPORT_GROWTH
## W = 0.96901, p-value = 0.35
```

```
shapiro.test(nor_test_data$GDP_PER_CAPITA_GROWTH) # --> normalna
```

```
##
##  Shapiro-Wilk normality test
##
## data:  nor_test_data$GDP_PER_CAPITA_GROWTH
## W = 0.95423, p-value = 0.1142
```

```
shapiro.test(usa_test_data$GDP_GROWTH) # --> normalna
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: usa_test_data$GDP_GROWTH  
## W = 0.94814, p-value = 0.07113
```

Provedeni testovi te grafički prikazi pokazuju nam da nemaju svi uzorci normalnu distribuciju. Zbog toga ćemo za testiranje pojedinih hipoteza morati koristiti alternativne neparametarske testove. Osim prikazanih Shapiro testova, odbacivanje nulte hipoteze dobiveno je još i za rast BDP-a per capita u Australiji te SAD-u.

Testovi

Kako bismo ispitali pojedine hipoteze koje su se i prirodno nametnule promatrajući grafove prikazane na početku rada, koristimo različite testove. Podaci koji su vezani za istu državu su zavisni jer se promatraju kroz isto vremensko razdoblje za istu državu.

Testovi homogenosti varijanci

Kako bismo mogli kasnije koristiti određenu vrstu testova, kao i linearnu regresiju, moramo prethodno odrediti odnose varijanci pojedinih populacija. Varijance uzoraka su nam nepoznate, ne svejedno možemo odrediti jesu li one značajno iste. To se može postići F testom čija je nulta hipoteza da su varijance populacije homogene, odnosno iste, a alternativna hipoteza je da su one različite. Nulta hipoteza se odbacuje ukoliko je p-vrijednost testa manja od razine značajnosti.

```
# usporedba varijanci rasta uvoza i izvoza za jednu državu, ako je p-vrijednost manja od 0.05 onda nisu  
var.test(NUMBERS ~ MARKER, data=aus_ei_test) # nisu iste
```

```
##  
## F test to compare two variances  
##  
## data: NUMBERS by MARKER  
## F = 0.31435, num df = 38, denom df = 38, p-value = 0.0005562  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.1648372 0.5994573  
## sample estimates:  
## ratio of variances  
## 0.3143451
```

```
var.test(NUMBERS ~ MARKER, data=kor_ei_test) # iste su
```

```
##  
## F test to compare two variances  
##  
## data: NUMBERS by MARKER  
## F = 0.58482, num df = 38, denom df = 38, p-value = 0.1025  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.3066714 1.1152606  
## sample estimates:  
## ratio of variances  
## 0.5848235
```

```
var.test(NUMBERS ~ MARKER, data=nor_ei_test) # iste su
```

```
##
```

```

## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 0.56163, num df = 38, denom df = 38, p-value = 0.07934
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2945114 1.0710388
## sample estimates:
## ratio of variances
##      0.5616344
var.test(NUMBERS ~ MARKER, data=usa_ei_test) # iste su

##
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## 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
#usporedba varijanci rasta bdpa za dvije države, ako je p-vrijednost manja od 0.05 onda nisu iste
var.test(NUMBERS ~ MARKER, data=kor_nor_test) # iste su

##
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 0.7938, num df = 38, denom df = 38, p-value = 0.48
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.416255 1.513779
## sample estimates:
## ratio of variances
##      0.7937998
var.test(NUMBERS ~ MARKER, data=kor_usa_test) # nisu iste

##
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 4.2346, num df = 38, denom df = 38, p-value = 2.153e-05
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  2.220552 8.075400
## sample estimates:
## ratio of variances
##      4.234601
var.test(NUMBERS ~ MARKER, data=nor_usa_test) # nisu iste

##

```



```

## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 5.3346, num df = 38, denom df = 38, p-value = 1.125e-06
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  2.79737 10.17309
## sample estimates:
## ratio of variances
##      5.334596

#usporedba varijanci uvoza i izvoza za jednu državu, ako je p-vrijednost manja od 0.05 onda nisu iste
var.test(NUMBERS ~ MARKER, data=aus2_ei_test) # nisu iste

##
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 0.55082, num df = 39, denom df = 39, p-value = 0.06623
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2913265 1.0414392
## sample estimates:
## ratio of variances
##      0.5508165

var.test(NUMBERS ~ MARKER, data=kor2_ei_test) # iste su

##
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 1.4668, num df = 39, denom df = 39, p-value = 0.236
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7757971 2.7733336
## sample estimates:
## ratio of variances
##      1.466814

var.test(NUMBERS ~ MARKER, data=nor2_ei_test) # iste su

##
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 1.2813, num df = 39, denom df = 39, p-value = 0.4425
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6776659 2.4225325
## sample estimates:
## ratio of variances
##      1.281276

var.test(NUMBERS ~ MARKER, data=usa2_ei_test) # iste su

##

```

```
## F test to compare two variances
##
## data:  NUMBERS by MARKER
## F = 0.59474, num df = 39, denom df = 39, p-value = 0.1089
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3145569 1.1244836
## sample estimates:
## ratio of variances
##      0.5947386
```

Ukoliko testiramo više varijanci odjednom, moramo koristiti testove koji podržavaju viševarijabilno istodobno testiranje. Ukoliko uzorci koje testiramo imaju normalnu razdiobu možemo koristiti Bartlettov test. Nulta hipoteza je kako su sve srednje vrijednosti iste, dok je alternativna hipoteza kako nisu iste. Ukoliko je p-vrijednost manja od razine značajnosti od 5%, nulta hipoteza se odbacuje.

Međutim, naši uzorci nemaju svi normalnu razdiobu pa koristimo Leveneov test. Nulta hipoteza govori kako su sve varijance homogene, dok alternativna hipoteza govori da nisu. Ukoliko je p-vrijednost testa manja od razine značajnosti testa od 5%, nulta hipoteza se odbacuje što znači da nije zadovoljena homogenost varijanci.

```
leveneTest(NUMBERS ~ MARKER, data=gdp_g_all_states)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value    Pr(>F)
## group  3  7.3523 0.0001237 ***
##      152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Kruskalov test

Za početak, ispitujemo je li rast gospodarstva značajno jednak za sve države. Koristit ćemo Kruskalov test koji je zapravo neparametarska alternativa za ANOVA testiranje. Budući da smo ranije pokazali kako nemaju sve populacije normalnu razdiobu te homogenost varijanci, ne možemo u svim slučajevima koristiti parametarske testove koji imaju određene pretpostavke. Nadalje, budući da testiramo jednakost većeg broja populacije, ne možemo iterativno uspoređivati sve kombinacije i to testirati jer to dovodi do povećanja greške.

```
kruskal.test(GDP_GROWTH ~ LOCATION, data = sap.data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  GDP_GROWTH by LOCATION
## Kruskal-Wallis chi-squared = 13.693, df = 3, p-value = 0.003354
```

Dakle, budući da je p-vrijednost manja od razine značajnosti od 5%, odbacujemo nultu hipotezu koja je tvrdila kako su srednje vrijednosti svih populacija jednake te prihvaćamo alternativnu hipotezu H1 koja kaže da su barem dvije srednje vrijednosti različite.

T testovi

Nadalje, ispitali smo i odnos rasta BDP-a između država. Testiranje nećemo provesti za sve slučajeve, nego ćemo odabrati određene slučajeve, odnosno kombinacije države za ispitivanje. Promatrajući vizualizirane podatke na početku rada, možemo odabrati Koreju, Norvešku i SAD za testiranje. Budući da smo u prethodnim odijeljima pokazali kako rast BDP-a ih država ima normalnu distribuciju, možemo koristiti t-test koji koristi t-statistiku:

$$T = \frac{Z\sqrt{n_1 + n_2 - 2}}{W} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_X\sqrt{1/n_1 + 1/n_2}}$$

Nulta hipoteza u t-testu jest da je rast BDP-a države A manji ili jednak od rasta BDP-a države B, dok je alternativna H1 hipoteza da je rast BDP-a države A veći od rasta BDP-a države B. Ukoliko je p-vrijednost manja od razine značajnosti od 5%, nulta hipoteza se odbacuje i prihvaća se alternativna hipoteza.

```
t.test(kor_test_data$GDP_GROWTH, nor_test_data$GDP_GROWTH, alternative = 'greater', var.equal = T)
```

```
##
## Two Sample t-test
##
## data: kor_test_data$GDP_GROWTH and nor_test_data$GDP_GROWTH
## t = 1.9978, df = 76, p-value = 0.02466
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 0.004030151 Inf
## sample estimates:
## mean of x mean of y
## 0.08698283 0.06277968
```

```
t.test(kor_test_data$GDP_GROWTH, usa_test_data$GDP_GROWTH, alternative = 'greater', var.equal = F)
```

```
##
## Welch Two Sample t-test
##
## data: kor_test_data$GDP_GROWTH and usa_test_data$GDP_GROWTH
## t = 3.6283, df = 54.999, p-value = 0.0003129
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 0.01751998 Inf
## sample estimates:
## mean of x mean of y
## 0.08698283 0.05447202
```

```
t.test(nor_test_data$GDP_GROWTH, usa_test_data$GDP_GROWTH, alternative = 'greater', var.equal = F)
```

```
##
## Welch Two Sample t-test
##
## data: nor_test_data$GDP_GROWTH and usa_test_data$GDP_GROWTH
## t = 0.84283, df = 51.763, p-value = 0.2016
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -0.008200879 Inf
## sample estimates:
## mean of x mean of y
## 0.06277968 0.05447202
```

Ispitali smo također i povezanost između izvoza i uvoza pojedine države. Budući da su vrijednosti izvoza i uvoza promatrani kroz isto razdoblje i za istu državu, podaci su međusobno zavisni i potrebno je koristiti upareni t-test. Ostale pretpostavke t-testa (normalna distribucija i homogenost varijanci) su prethodno ispitani te su zadovoljeni. Nulta hipoteza kod ovog testiranja jest da je rast izvoza manji ili jednak od rasta izvoza, dok je alternativna hipoteza da je rast izvoza veći od rasta uvoza. Ukoliko je p-vrijednost testa manja od razine značajnosti od 5%, nultu hipotezu odbacujemo.

```
t.test(aus_test_data$EXPORT_GROWTH, aus_test_data$IMPORT_GROWTH, alternative = 'greater', paired = T, var.equal = F)
```

```
##
## Paired t-test
```

```
##
## data:  aus_test_data$EXPORT_GROWTH and aus_test_data$IMPORT_GROWTH
## t = -0.89125, df = 38, p-value = 0.8108
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.03489003      Inf
## sample estimates:
## mean of the differences
##                -0.0120657

t.test(kor_test_data$EXPORT_GROWTH, kor_test_data$IMPORT_GROWTH, alternative = 'greater', paired = T, v

##
## Paired t-test
##
## data:  kor_test_data$EXPORT_GROWTH and kor_test_data$IMPORT_GROWTH
## t = 1.2705, df = 38, p-value = 0.1058
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.006033397      Inf
## sample estimates:
## mean of the differences
##                0.01845319

t.test(nor_test_data$EXPORT_GROWTH, nor_test_data$IMPORT_GROWTH, alternative = 'greater', paired = T, v

##
## Paired t-test
##
## data:  nor_test_data$EXPORT_GROWTH and nor_test_data$IMPORT_GROWTH
## t = -0.57161, df = 38, p-value = 0.7145
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.0196199      Inf
## sample estimates:
## mean of the differences
##               -0.004967719

t.test(usa_test_data$EXPORT_GROWTH, usa_test_data$IMPORT_GROWTH, alternative = 'greater', paired = T, v

##
## Paired t-test
##
## data:  usa_test_data$EXPORT_GROWTH and usa_test_data$IMPORT_GROWTH
## t = -0.38863, df = 38, p-value = 0.6501
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.02126128      Inf
## sample estimates:
## mean of the differences
##               -0.00398288
```

Testiranjem smo dobili kako pri razini značajnosti of 5% ne možemo odbaciti nultu hipotezu da je rast izvoza manji ili jednak od rasta uvoza.

I naposljetku, ispitujemo povezanost rasta BDP-a i BDP-a po glavi stanovnika. Budući da smo ranije pokazali kako ove populacije nemaju normalnu razdiobu, trebali bismo koristiti neparametarske testove. No, budući

da koristimo uzorak veličine 39, po graničnom centralnom teoremu (CGT) možemo zanemariti nepostojanje normalne distribucije i koristiti aproksimiranje na normalnu distribuciju. Dakle, možemo provesti upareni t test budući da su podaci međusobno zavisni. Dodatno kao provjeru provodimo i njegovu neparametarsku alternativu, odnosno Wilcoxonov test.

```
t.test(aus_test_data$GDP_GROWTH, aus_test_data$GDP_PER_CAPITA_GROWTH, var.equal = F, paired = T)
```

```
##
## Paired t-test
##
## data: aus_test_data$GDP_GROWTH and aus_test_data$GDP_PER_CAPITA_GROWTH
## t = 29.563, df = 38, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.01349020 0.01547354
## sample estimates:
## mean of the differences
## 0.01448187
```

```
wilcox.test(aus_test_data$GDP_GROWTH, aus_test_data$GDP_PER_CAPITA_GROWTH, paired=T)
```

```
##
## Wilcoxon signed rank test
##
## data: aus_test_data$GDP_GROWTH and aus_test_data$GDP_PER_CAPITA_GROWTH
## V = 780, p-value = 3.638e-12
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(kor_test_data$GDP_GROWTH, kor_test_data$GDP_PER_CAPITA_GROWTH, paired = T)
```

```
##
## Paired t-test
##
## data: kor_test_data$GDP_GROWTH and kor_test_data$GDP_PER_CAPITA_GROWTH
## t = 13.625, df = 38, p-value = 3.322e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.007615458 0.010273276
## sample estimates:
## mean of the differences
## 0.008944367
```

```
t.test(nor_test_data$GDP_GROWTH, nor_test_data$GDP_PER_CAPITA_GROWTH, paired = T)
```

```
##
## Paired t-test
##
## data: nor_test_data$GDP_GROWTH and nor_test_data$GDP_PER_CAPITA_GROWTH
## t = 13.686, df = 38, p-value = 2.888e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.006106251 0.008226328
## sample estimates:
## mean of the differences
## 0.007166289
```

```
t.test(usa_test_data$GDP_GROWTH, usa_test_data$GDP_PER_CAPITA_GROWTH, paired = T, var.equal = F)

##
## Paired t-test
##
## data: usa_test_data$GDP_GROWTH and usa_test_data$GDP_PER_CAPITA_GROWTH
## t = 30.743, df = 38, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.009425557 0.010754400
## sample estimates:
## mean of the differences
## 0.01008998

wilcox.test(usa_test_data$GDP_GROWTH, usa_test_data$GDP_PER_CAPITA_GROWTH, paired = T)

##
## Wilcoxon signed rank test
##
## data: usa_test_data$GDP_GROWTH and usa_test_data$GDP_PER_CAPITA_GROWTH
## V = 780, p-value = 3.638e-12
## alternative hypothesis: true location shift is not equal to 0
```

Zaključujemo kako u pravilu ne vrijedi da su rast BDP-a i rast BDP-a po glavi stanovništva značajno isti.

Linearna regresija

Iako smo spomenuli kako se BDP izračunava po formuli u kojoj ovisi o 4 različite varijable, testirali smo različite modele koji su modelirali vezu između uvoza i izvoza te BDP-a.

Isti modeli nisu davali dobru aproksimaciju za sve države. Za svaku državu odabrali jedan ili dva modela za koje smo dobili dobre rezultate. Te iste modele smo i prikazali grafički (crne točkice - stvarna vrijednost, plava linija - fittana vrijednost). Dodatno smo još testirali i normalnost reziduala putem QQ grafa i Kolmogorov-Smirnovljev testa. Sama funkcija modela linearne regresije već ima implementirane podatke o vrijednosti F statistike koja nam govori o signifikantnosti modela, ali i o vrijednosti koeficijenta determinancije. Nulta hipoteza Kolmogorov-Smirnovljeva testa govori kako je distribucija uzorka normalna distribucija. S druge strane, alternativna H1 hipoteza govori kako distribucija uzorka nije normalna. Uz nivo značajnosti od 5%, nultu hipotezu odbacujemo ukoliko je p-vrijednost testa manja od 5%.

Australija

```
lm.data <- sap.data %>% filter(LOCATION=="AUS")

lm1 <- lm(GDP ~ EXPORT + IMPORT, lm.data)
summary(lm1)

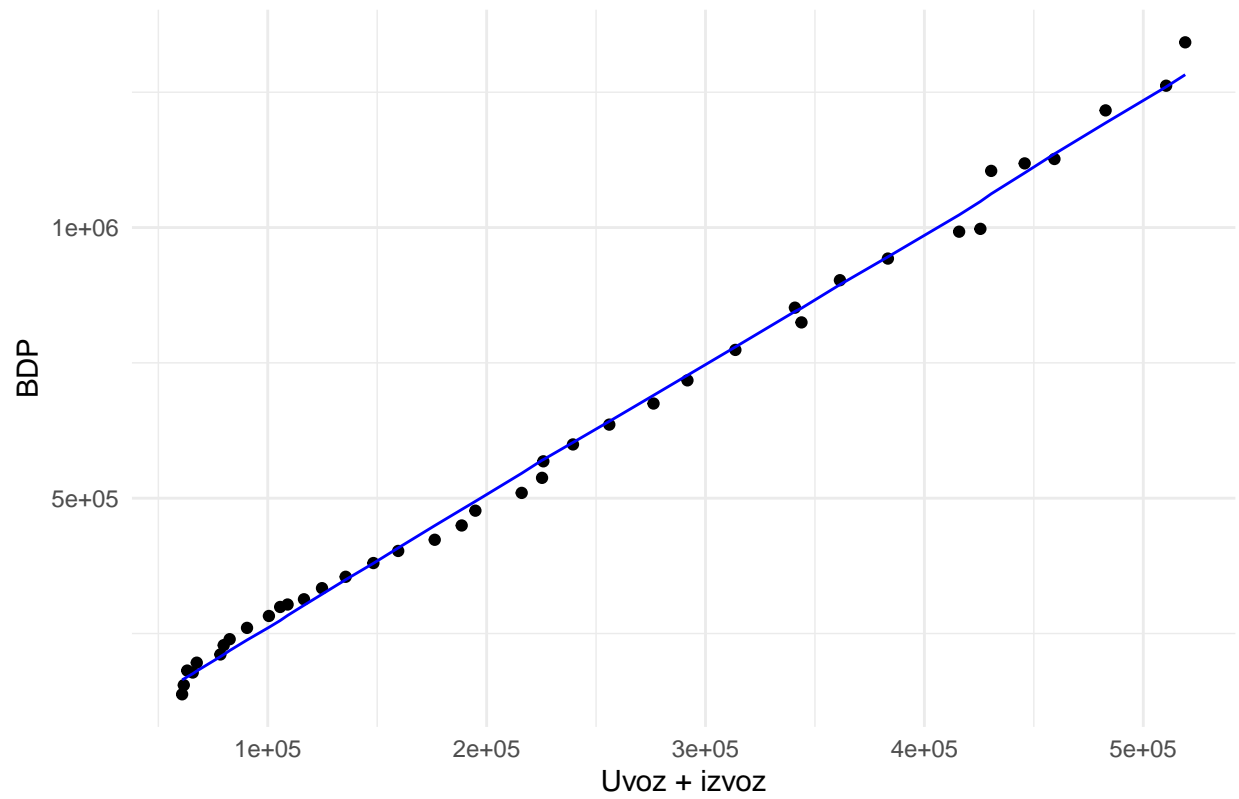
##
## Call:
## lm(formula = GDP ~ EXPORT + IMPORT, data = lm.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50573 -12320   1293  15875  59758
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.464e+04  9.195e+03  1.593    0.12
```

```
## EXPORT      2.570e+00  2.281e-01  11.267 1.61e-13 ***
## IMPORT      2.325e+00  1.693e-01  13.733 4.23e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22990 on 37 degrees of freedom
## Multiple R-squared:  0.9961, Adjusted R-squared:  0.9959
## F-statistic:  4687 on 2 and 37 DF,  p-value: < 2.2e-16
lm2 <- lm(GDP ~ EXPORT + IMPORT + I(EXPORT^2), lm.data)
summary(lm2)
```

```
##
## Call:
## lm(formula = GDP ~ EXPORT + IMPORT + I(EXPORT^2), data = lm.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42735 -14265   4343   13170   44817
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.028e+04  1.220e+04   4.119 0.000213 ***
## EXPORT      1.884e+00  2.652e-01   7.104  2.4e-08 ***
## IMPORT      2.241e+00  1.465e-01  15.297 < 2e-16 ***
## I(EXPORT^2) 3.119e-06  8.170e-07   3.818 0.000511 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19660 on 36 degrees of freedom
## Multiple R-squared:  0.9972, Adjusted R-squared:  0.997
## F-statistic:  4276 on 3 and 36 DF,  p-value: < 2.2e-16
```

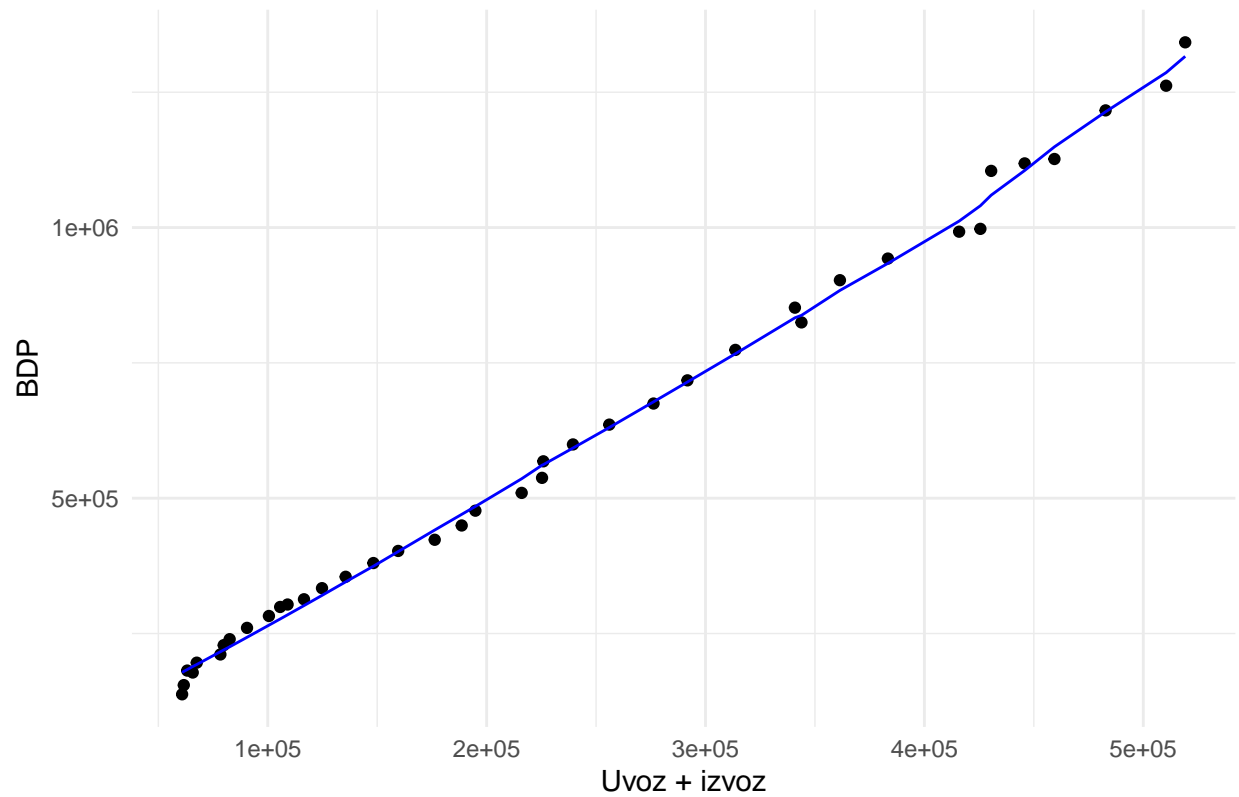
```
ggplot(data = aus_test_data) + geom_point(aes(x = EXPORT + IMPORT, y = GDP)) + geom_line(aes(x = EXPORT
```

Australija – prvi model



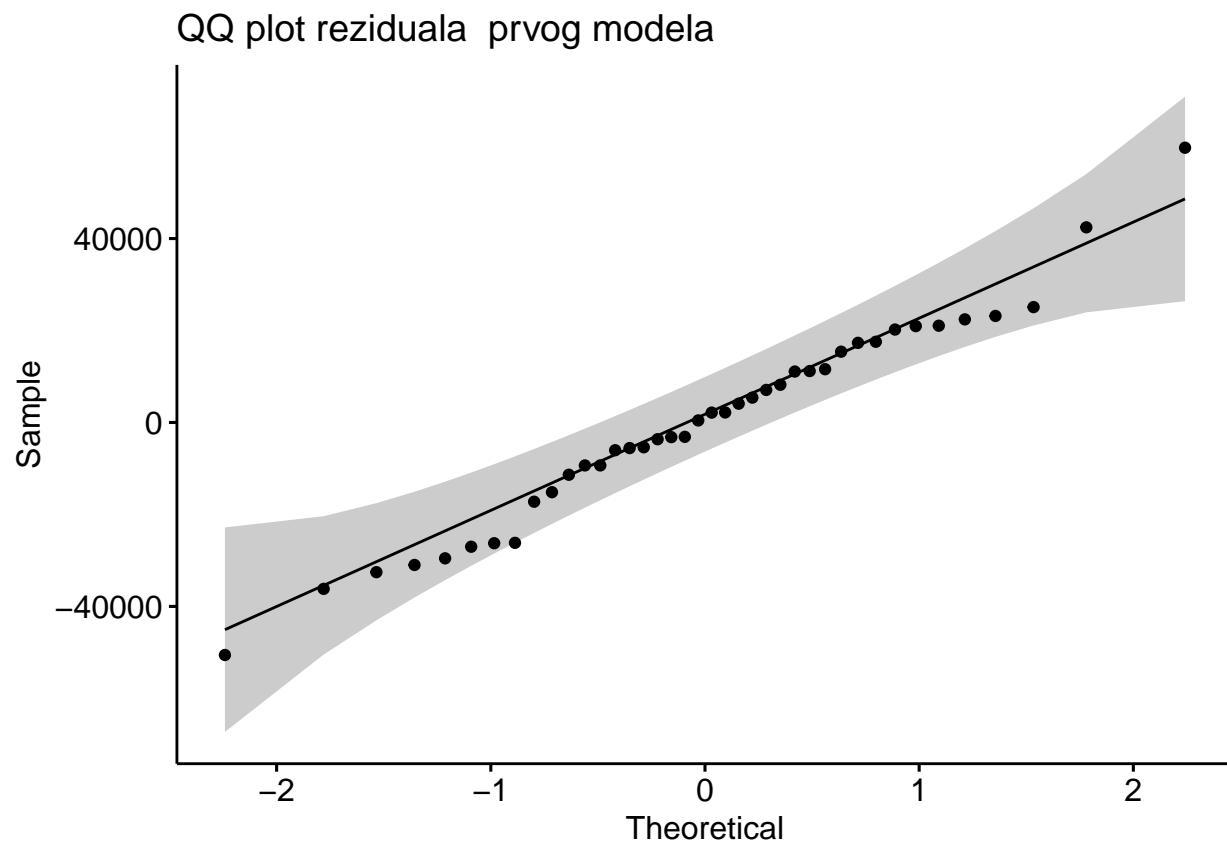
```
ggplot(data = aus_test_data) + geom_point(aes(x = EXPORT + IMPORT, y = GDP)) + geom_line(aes(x = EXPORT
```


Australija – drugi model



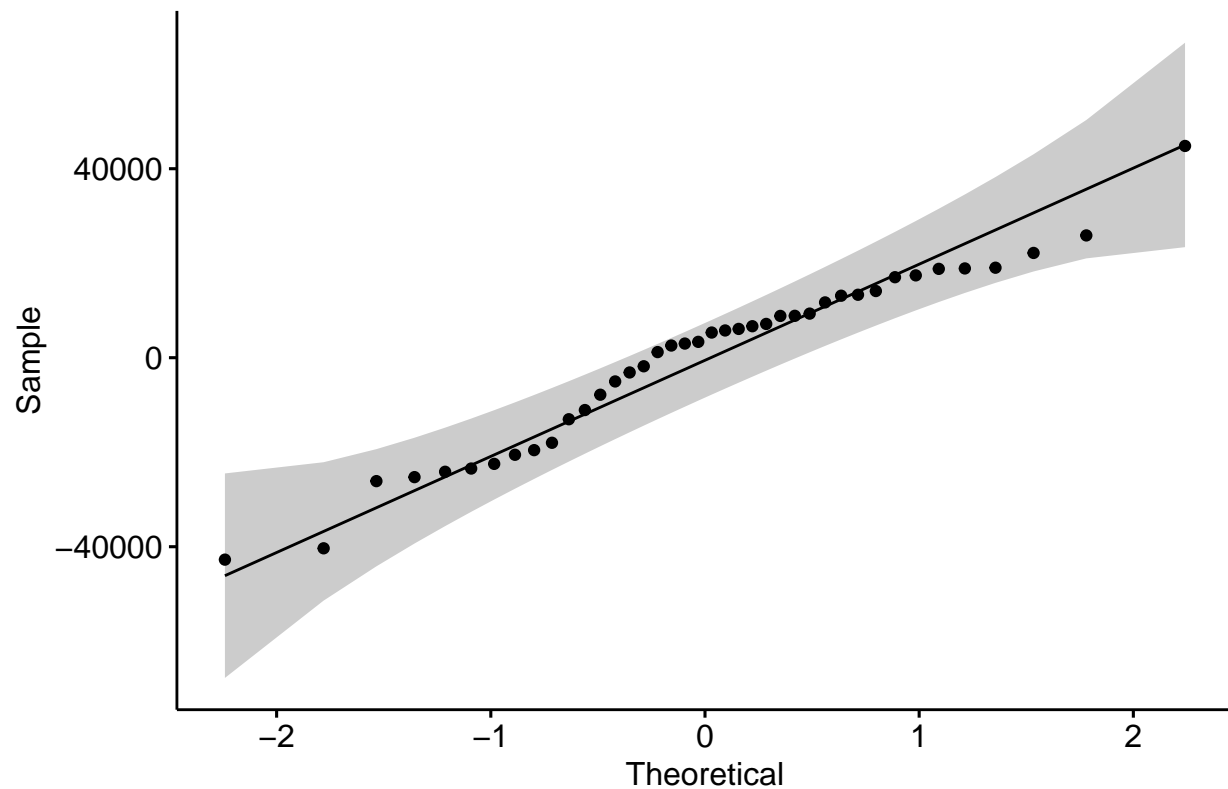
Provjerit ćemo i normalnost reziduala preko QQ plotta i Kolmogorov-Smirnovljeva testa.

```
ggqqplot(lm1$residuals, main = 'QQ plot reziduala prvog modela')
```



```
ggqqplot(lm2$residuals, main = 'QQ plot reziduala drugog modela')
```

QQ plot reziduala drugog modela



```
ks.test(rstandard(lm1), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: rstandard(lm1)
## D = 0.081257, p-value = 0.935
## alternative hypothesis: two-sided
```

```
ks.test(rstandard(lm2), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: rstandard(lm2)
## D = 0.13312, p-value = 0.4397
## alternative hypothesis: two-sided
```

Južna Koreja

```
lm.data <- sap.data %>% filter(LOCATION=="KOR")
```

```
lm1 <- lm(GDP ~ EXPORT + IMPORT + I(EXPORT^2), lm.data)
summary(lm1)
```

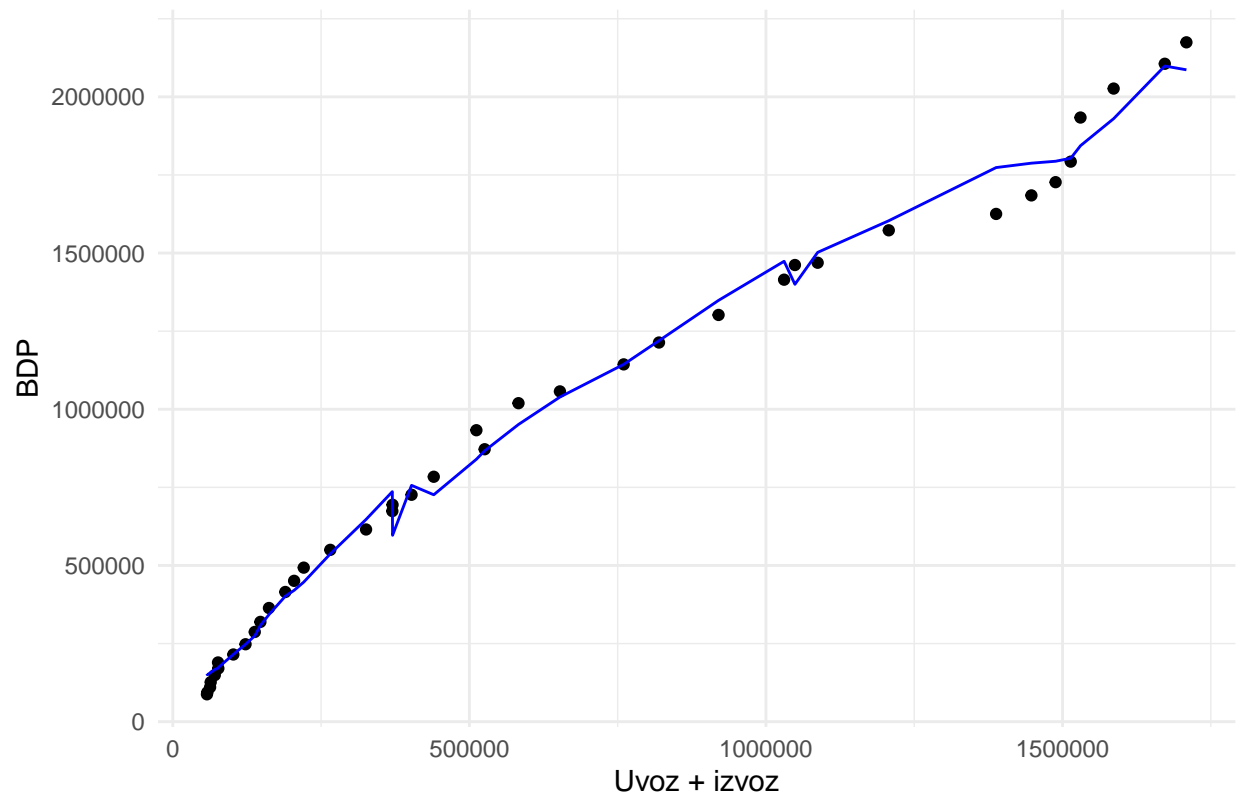
```
##
## Call:
## lm(formula = GDP ~ EXPORT + IMPORT + I(EXPORT^2), data = lm.data)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -148341  -32457       67   23416   97773
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.732e+04  2.124e+04   1.757  0.0874 .
## EXPORT       3.805e-01  3.806e-01   1.000  0.3241
## IMPORT       2.895e+00  3.733e-01   7.757 3.43e-09 ***
## I(EXPORT^2) -7.723e-07  1.576e-07  -4.899 2.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58140 on 36 degrees of freedom
## Multiple R-squared:  0.9927, Adjusted R-squared:  0.9921
## F-statistic: 1637 on 3 and 36 DF,  p-value: < 2.2e-16
lm2 <- lm(GDP ~ IMPORT + I(IMPORT^2), lm.data)
summary(lm2)
```

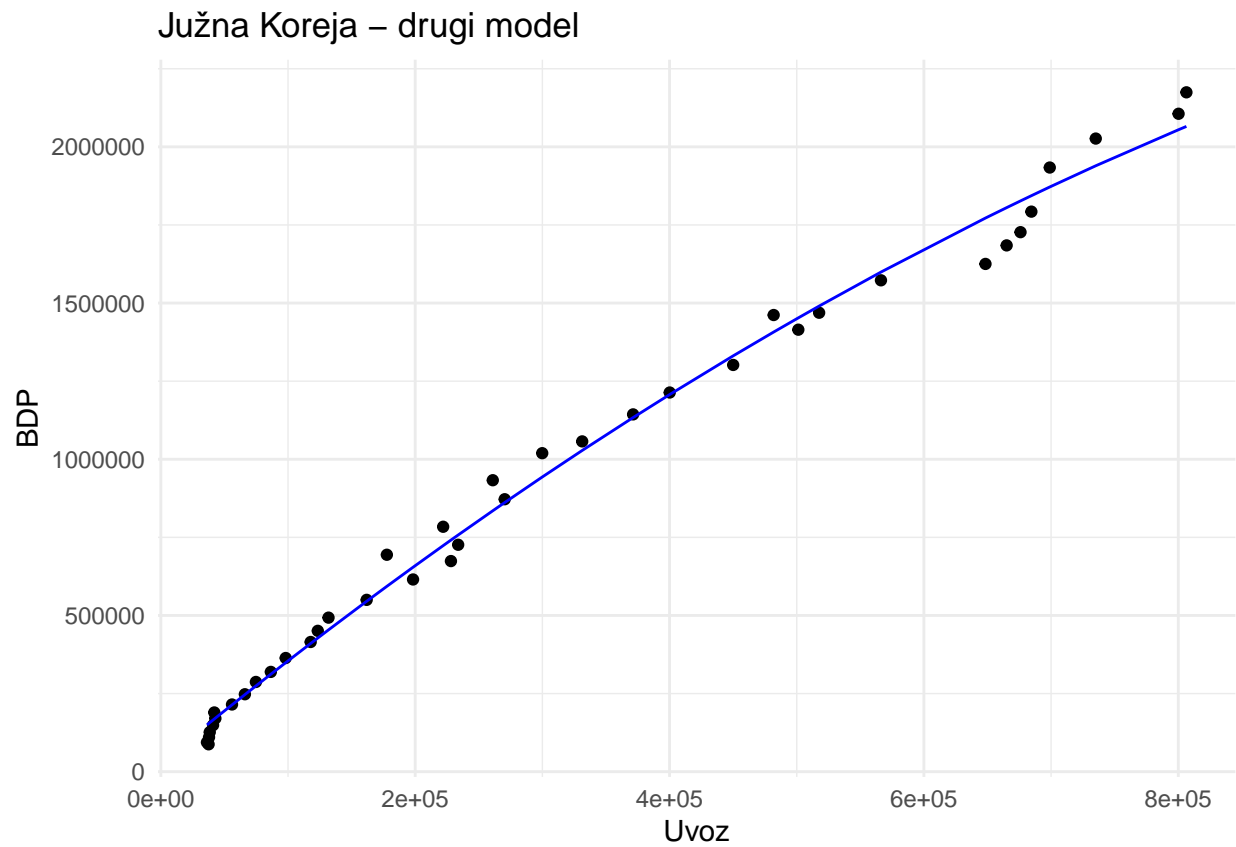
```
##
## Call:
## lm(formula = GDP ~ IMPORT + I(IMPORT^2), data = lm.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -146808  -32986    4556   31585  109123
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.933e+04  2.150e+04   1.364   0.181
## IMPORT       3.356e+00  1.534e-01  21.881 < 2e-16 ***
## I(IMPORT^2) -1.031e-06  1.937e-07  -5.323 5.18e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59780 on 37 degrees of freedom
## Multiple R-squared:  0.9921, Adjusted R-squared:  0.9917
## F-statistic: 2321 on 2 and 37 DF,  p-value: < 2.2e-16
```

```
ggplot(data = kor_test_data) + geom_point(aes(x = EXPORT + IMPORT, y = GDP)) + geom_line(aes(x = EXPORT
```

Južna Koreja – prvi model

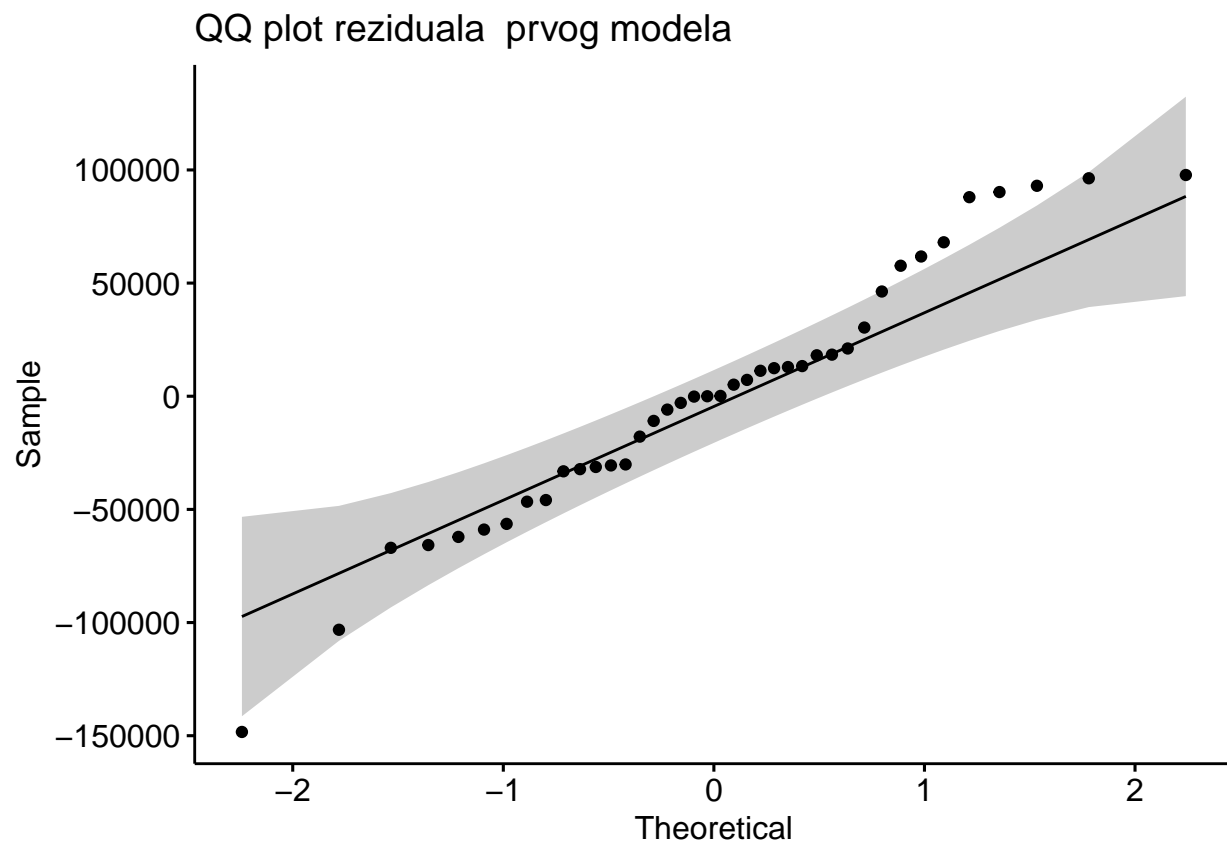


```
ggplot(data = kor_test_data) + geom_point(aes(x = IMPORT, y = GDP)) + geom_line(aes(x = IMPORT, y = lm2
```



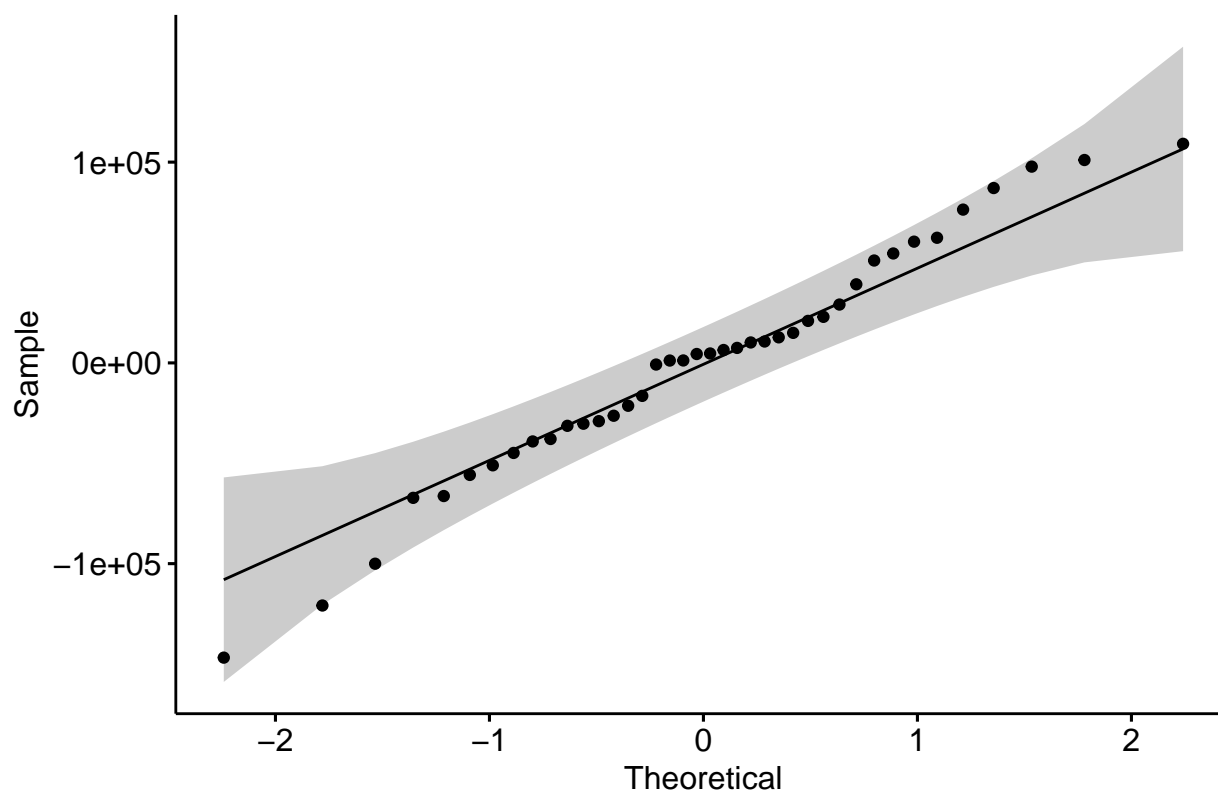
Provjera normalnosti reziduala za izvedene modele.

```
ggqqplot(lm1$residuals, main = 'QQ plot reziduala prvog modela')
```



```
ggqqplot(lm2$residuals, main = 'QQ plot reziduala drugog modela')
```

QQ plot reziduala drugog modela



```
ks.test(rstandard(lm1), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: rstandard(lm1)
## D = 0.10472, p-value = 0.7335
## alternative hypothesis: two-sided
```

```
ks.test(rstandard(lm2), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: rstandard(lm2)
## D = 0.094273, p-value = 0.8366
## alternative hypothesis: two-sided
```

Norveška

```
lm.data <- sap.data %>% filter(LOCATION=="NOR")

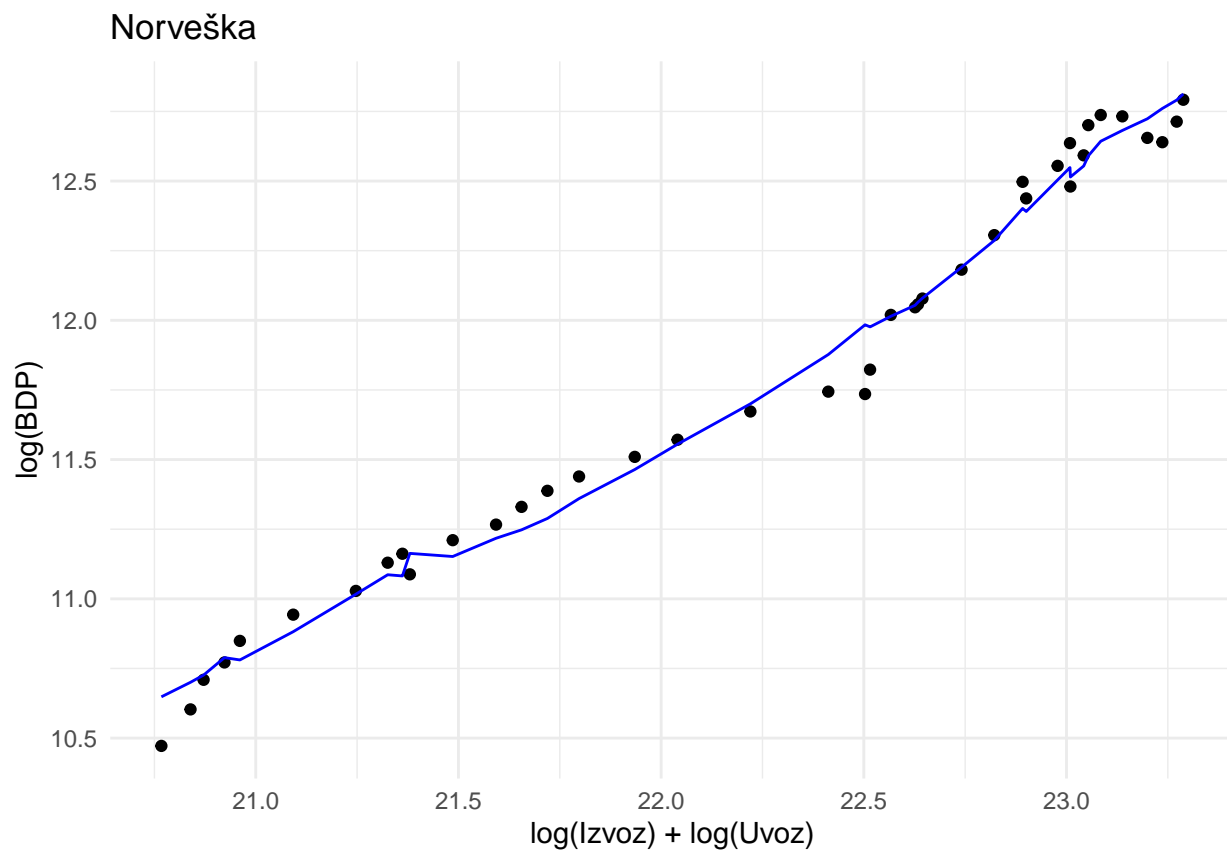
lm1 <- lm(log(GDP) ~ log(EXPORT) + log(IMPORT), lm.data)
summary(lm1)
```

```
##
## Call:
## lm(formula = log(GDP) ~ log(EXPORT) + log(IMPORT), data = lm.data)
```



```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24857 -0.02910  0.01391  0.05935  0.11026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.49757    0.39074  -16.629  < 2e-16 ***
## log(EXPORT)    0.42592    0.08076   5.274 6.04e-06 ***
## log(IMPORT)    1.23687    0.07218  17.135  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08586 on 37 degrees of freedom
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.9857
## F-statistic: 1343 on 2 and 37 DF,  p-value: < 2.2e-16
```

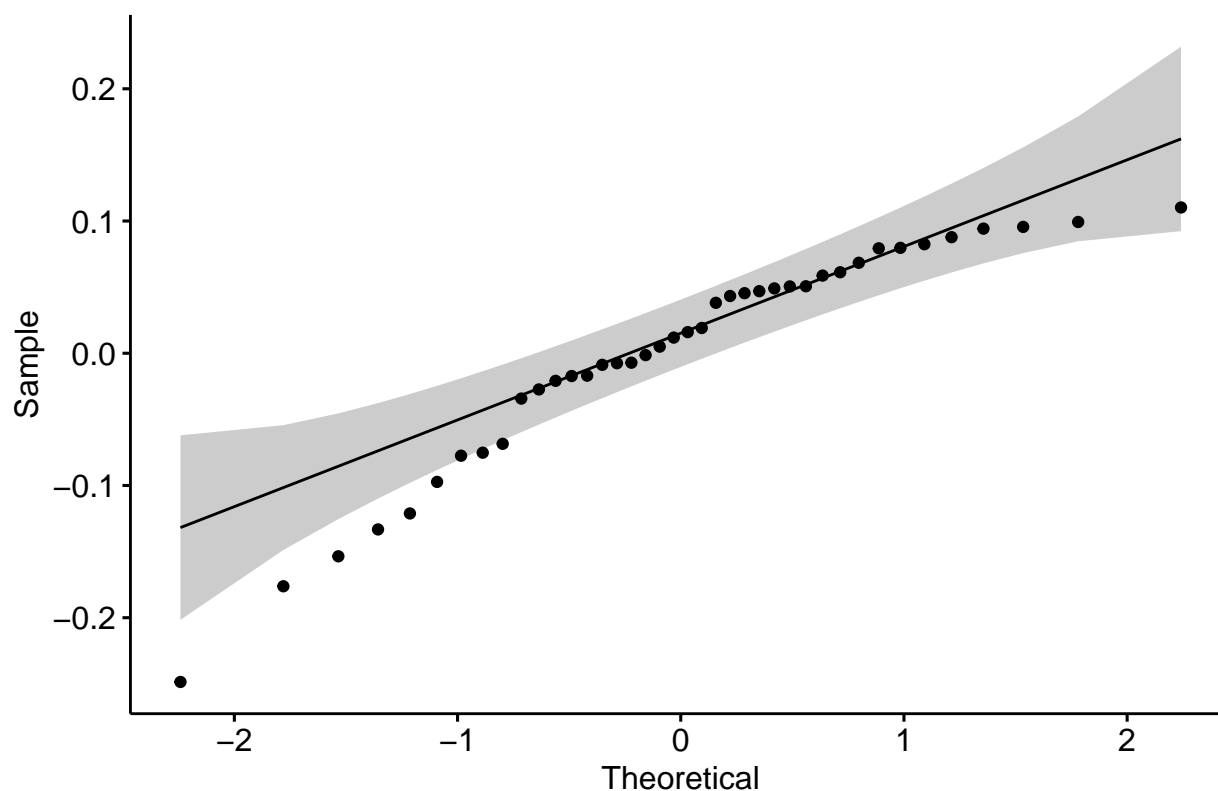
```
ggplot(data = nor_test_data) + geom_point(aes(x = log(EXPORT) + log(IMPORT), y = log(GDP))) + geom_line
```



Provjera normalnosti reziduala za izvedene modele.

```
ggqqplot(lm1$residuals, main = 'QQ plot reziduala modela za Norvešku')
```

QQ plot reziduala modela za Norvešku



```
ks.test(rstandard(lm1), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data:  rstandard(lm1)
## D = 0.12647, p-value = 0.5044
## alternative hypothesis: two-sided
```

Sjedinjene Američke Države

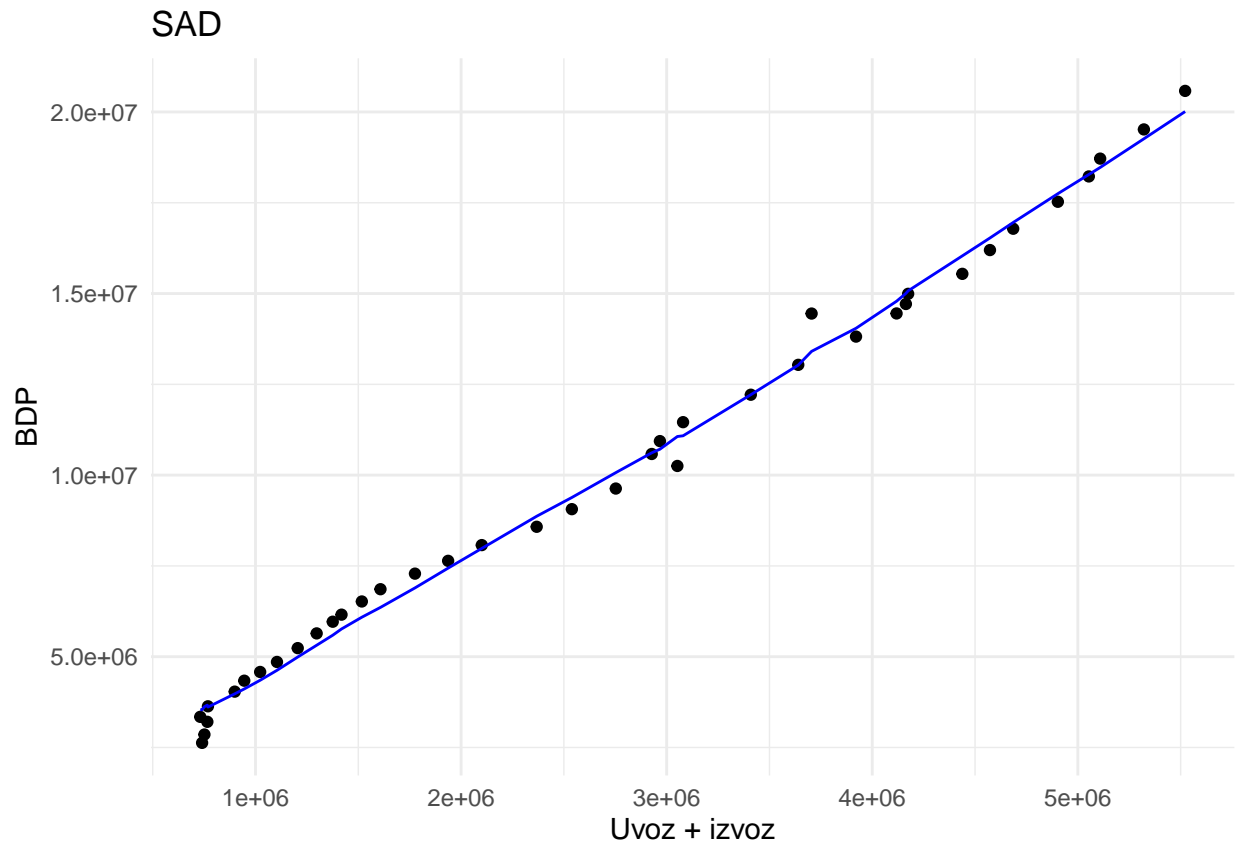
```
lm.data <- sap.data %>% filter(LOCATION=="USA")
```

```
lm1 <- lm(GDP ~ EXPORT + IMPORT + I(IMPORT^2), lm.data)
summary(lm1)
```

```
##
## Call:
## lm(formula = GDP ~ EXPORT + IMPORT + I(IMPORT^2), data = lm.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -911413 -290719  14381  249133 1040924
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.234e+06  2.596e+05  4.753 3.18e-05 ***
## EXPORT      3.976e+00  5.890e-01  6.751 6.99e-08 ***
```

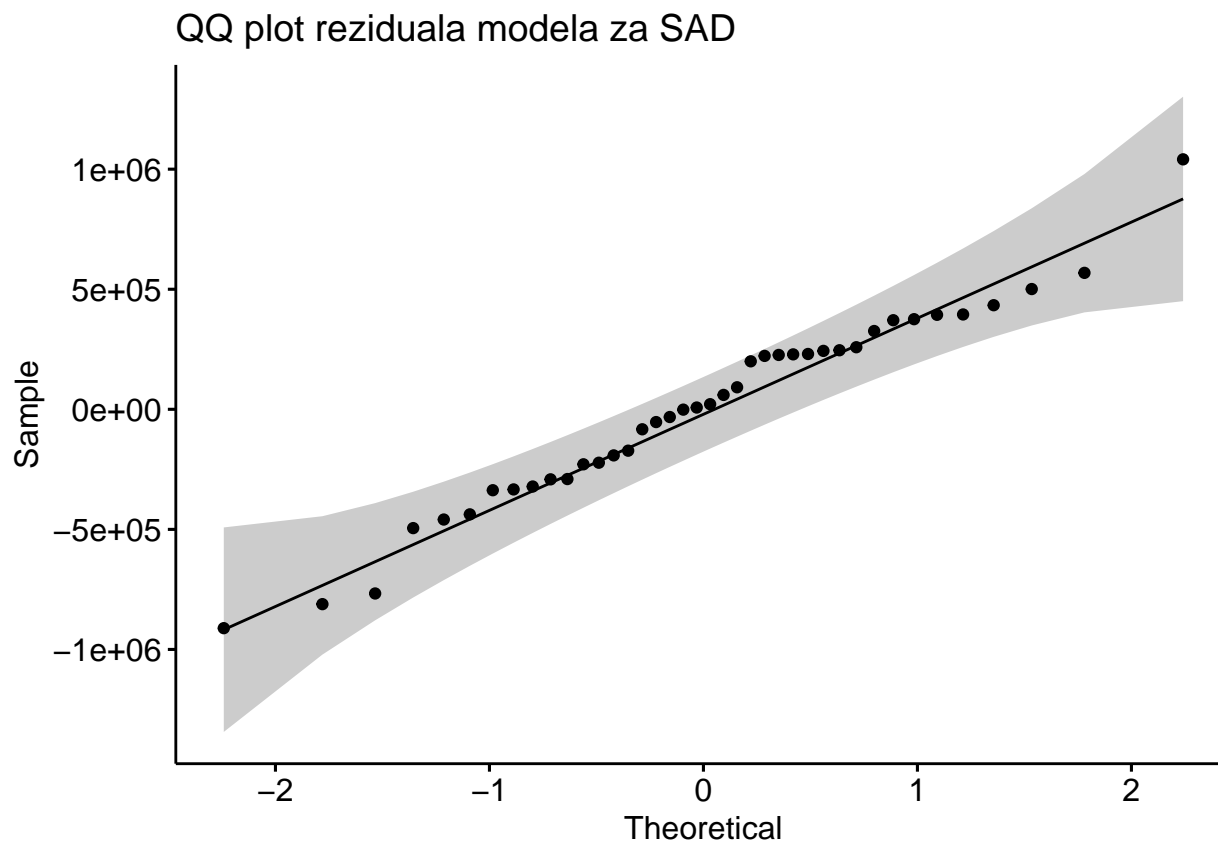
```
## IMPORT      2.261e+00  5.313e-01  4.256 0.000142 ***
## I(IMPORT^2) 2.226e-07  1.198e-07  1.859 0.071251 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 421200 on 36 degrees of freedom
## Multiple R-squared:  0.9942, Adjusted R-squared:  0.9938
## F-statistic: 2075 on 3 and 36 DF,  p-value: < 2.2e-16
```

```
ggplot(data = usa_test_data) + geom_point(aes(x = EXPORT + IMPORT, y = GDP)) + geom_line(aes(x = EXPORT
```



Provjera normalnosti reziduala za izvedene modele.

```
ggqqplot(lm1$residuals, main = 'QQ plot reziduala modela za SAD')
```



```
ks.test(rstandard(lm1), 'pnorm')
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data:  rstandard(lm1)
## D = 0.11437, p-value = 0.6308
## alternative hypothesis: two-sided
```

Zaključak

Gospodarstvo pojedine države određeno je iznosom uvoza i izvoza. Bruto domaći proizvod je pokazatelj jačine gospodarstva neke države, no nema smisla upoređivati apsolutne vrijedi cjelokupnog BDP-a jer možemo dobiti krivu sliku o jačini gospodarstva. BDP se zato najčešće prikazuje po glavi stanovnika. Nadalje, pokazali smo kako se skoro za sve ispitane države BDP može aproksimirati linearnom regresijom pri čemu su regresori vrijednosti uvoza i izvoza te države. Konačno, pokazali smo kako s razinom značajnosti od 5% ne možemo pokazati jednakost rasta BDP-a i BDP-a per capita za sve ispitane države, kao i jednakost rasta izvoza i uvoza ispitanih država.

U izradi ovog rada uvelike nam je pomogao ovaj motivirajući video: <https://www.youtube.com/watch?v=BvWefB4NGGI>