

Prirodoslovno-matematički fakultet Sveučilišta u Zagrebu

Očekivana životna dob žena u ovisnosti o BDP-u država svijeta

Seminarski rad

Zagreb, 16. svibnja 2022.

MENTORI: Miljenko Huzak, Nikolina Milinčević
Dunja Prelog, Petra Salaj, Maja Soldo, Josipa Šimić

UVOD

U ovom seminarskom radu analiziramo podatke očekivane životne dobi žena u pojedinim državama, podatke BDP-a po glavi stanovnika (u nastavku BDP) tih država te ispitujeemo njihovu povezanost. Pretpostavljamo da države s većim BDP-om imaju veću očekivanu životnu dob žena. Postoje brojni razlozi koji motiviraju takvu pretpostavku; adekvatna zdravstvena skrb, razvijenost farmaceutske industrije, kvalitetna socijalna pomoć i općenito veći životni standard koji imaju države s većim BDP-om sve su čimbenici poboljšanja kvalitete življenja i povećanja očekivanog životnog vijeka stanovnika tih država. Jesu li te karakteristike zaista povezane i koliko, predmet je našeg istraživanja.

Radimo s dvije varijable; očekivana životna dob žena i BDP (po glavi stanovnika). Podatke o očekivanoj životnoj dobi žena preuzeli smo sa stranice Gapminder¹, a podatke o BDP-u sa stranice Svjetske banke².

Države svijeta dijele se u četiri kategorije razvijenosti obzirom na BDP; ekonomije s niskim dohotkom, ekonomije s nižim srednjim dohotkom, ekonomije s višim srednjim dohotkom i višim dohotkom.³ Zbog mnogobrojnosti država svijeta, države koje promatramo smo podijelili u tri skupine. U prvoj skupini su države s BDP-om manjim od 4096\$, u drugoj skupini su države s BDP-om između 4096 i 12695\$, dok su u trećoj skupini države s više od 12695\$. Iz svake skupine slučajnim odabirom smo izabrali njih 30 te smo na taj način dobili uzorak od 90 država čije podatke koristimo za istraživanje. Uzorak analiziramo opisnom statistikom i statističkim testovima, a računanje i grafičko prikazivanje podataka provodimo u programskom jeziku R.

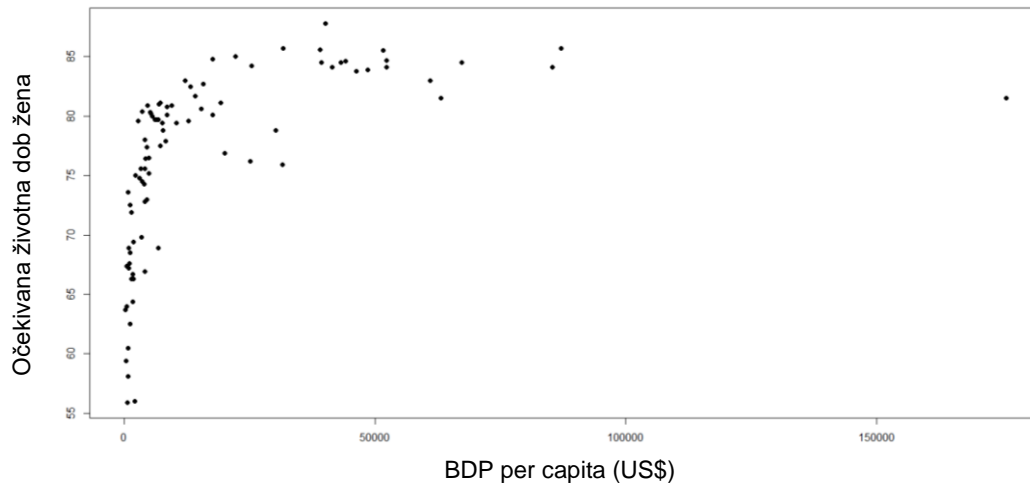
Država	BDP	Dob	Država	BDP	Dob
Albanija	5246.1	80.3	Kuba	9477.9	80.9
Angola	1776.2	64.4	Latvija	17736.5	80.1
Argentina	8579.0	80.1	Lesoto	875.4	58.1
Aruba	30253.3	78.8	Libanon	4649.5	80.9
Australija	51680.3	85.5	Lihtenštajn	175813.9	81.5
Austrija	48588.7	83.9	Luksemburg	39403.1	84.5
Azerbajdžan	4221.4	75.6	Maldivi	6924.1	81.0
Bahami	25194.0	76.2	Mali	862.5	60.5
Barbados	15373.9	80.6	Mauritanija	1702.0	66.7
Bjelorusija	6424.2	79.7	Meksiko	8329.3	77.9
Belize	4115.2	78.0	Mongolija	4061.0	74.3
Bolivija	3133.1	74.8	Namibija	4179.3	66.9
Brazil	6796.8	79.7	Nepal	1155.1	72.5
Burundi	239.0	63.7	Niger	567.7	64.0
Crna Gora	7677.2	79.4	Nigerija	2097.1	56.0
Čad	659.3	55.9	Nizozemska	52396.0	84.1
Čile	13231.7	82.5	Norveška	67329.7	84.5
Danska	61063.3	83.0	Novi Zeland	41441.5	84.1
Dominikanska Republika	7268.2	77.5	Njemačka	46252.7	83.8
Džibuti	3425.5	69.8	Pakistan	1188.9	68.5
Egipat	3569.2	74.5	Panama	12509.8	81.9
Ekvador	5600.4	80.0	Paragvaj	5001.1	76.5
Etiopija	936.3	68.9	Peru	6126.9	79.7
Filipini	3298.8	75.6	Poljska	15742.5	82.7
Francuska	39037.1	85.6	Portugal	22194.6	85.0
Gabon	6881.7	68.9	Rumunjska	12915.2	79.6
Grčka	17647.2	84.8	SAD	63206.5	81.5
Gvatemala	4603.3	77.4	Saudijska Arabija	20110.3	76.9
Gvineja	1194.0	62.5	Slovačka Republika	19266.5	81.1
Hrvatska	14132.5	81.7	Slovenija	25489.5	84.2
Irak	4145.9	72.8	Solomonski Otoci	2250.6	75.0
Irska	85422.5	84.1	Somalija	438.3	59.4
Istočni Timor	1442.7	71.9	Srbija	7730.7	78.8
Italija	31770.0	85.7	Sudan	486.4	67.4
Izrael	44177.6	84.6	Surinam	4916.6	75.2
Japan	40193.3	87.8	Šri Lanka	3680.7	80.4
Jordan	4282.8	76.4	Švedska	52274.4	84.7
Kanada	43258.3	84.5	Švicarska	87100.4	85.7
Kenija	1878.6	69.4	Tadžikistan	859.1	73.6
Kina	10434.8	79.4	Tajland	7186.9	81.1
Kolumbija	5334.6	80.2	Tonga	4624.8	73.0
Kongo Republika	1846.1	66.3	Tanzanija	1076.5	67.6
Komori	1420.7	66.3	Turska	8536.4	80.8
Koreja Republika	31597.5	75.9	Vijetnam	2785.7	79.6
Kostarika	12140.9	83.0	Zambija	985.1	67.2

Tablica 1. Statistički podaci za 2020.godinu

OPIŠNA STATISTIKA

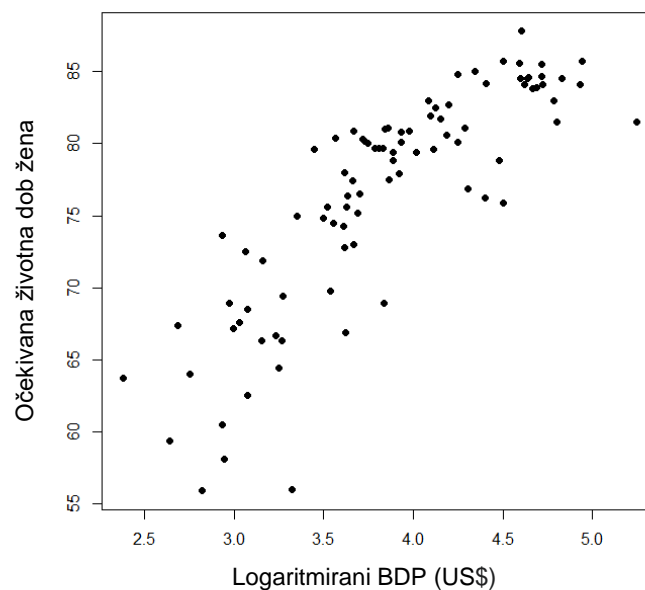
Za početak, analizirat ćemo podatke koristeći opisnu statistiku. Pomoću nje se upoznajemo s uzorkom te preglednije prikazujemo prikupljene podatke.

Pogledajmo prvo koordinatni prikaz podataka, da vidimo postoji li ikakva funkcijska ovisnost varijabli koje promatramo.



Slika 1. Koordinatni prikaz promatranih varijabli

Iz koordinatnog prikaza varijabli na Slici 1. vidimo da se podaci ponašaju približno logaritamski, stoga ćemo logaritmirati podatke da dobijemo linearni model pomoću kojeg možemo proučavati linearnu regresiju.



Slika 2. Koordinatni prikaz transformiranih podataka

Na Slici 2. vidimo da s logaritmiranim podacima o BDP-u grafički prikaz promatranih varijabli pobliže poprima oblik pravca. Nakon analize podataka, taj pravac ćemo i modelirati.

Analiza podataka

Za svaku varijablu izračunat ćemo karakterističnu petorku, aritmetičku sredinu i standardnu devijaciju te nacrtati histogram i dijagram pravokutnika.

Prvo analiziramo očekivanu životnu dob žena.

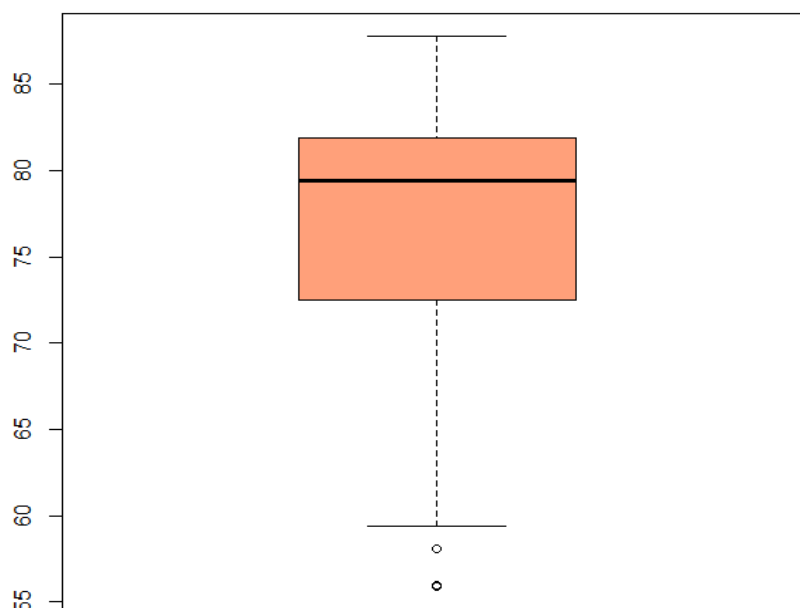
Karakteristična petorka

Najmanja vrijednost	Donji kvartil	Medijan	Gornji kvartil	Najveća vrijednost
55.90	72.35	79.40	82.05	87.80

Aritmetička sredina podataka je 76.52 godine, što je vrlo blizu medijanu. To nam ukazuje na dobru raspoređenost podataka, kao i na malu količinu podataka koji odstupaju od većine, što predočava i histogram na Slici 4. Iz danih podataka vidimo da u približno pola država svijeta očekivana životna dob žena iznosi 79.4 godine, što je i blizu očekivanju u Hrvatskoj (81.7 godina). Budući da Hrvatska ima relativno kvalitetnu zdravstvenu i socijalnu skrb, takav ishod je smislen i podržava početnu pretpostavku.

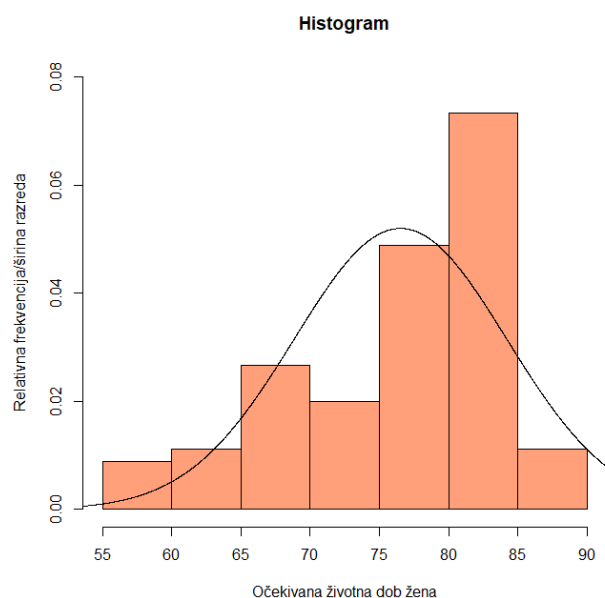
S druge strane, najmanju očekivanu životnu dob žena ima afrička država Čad, a iznosi samo 55.9 godina. Čad je donedavno bila jedna od najsiromašnijih zemalja (te je i dalje zemlja trećeg svijeta) s vrlo niskim BDP-om. Stanovništvo živi u vrlo lošim životnim uvjetima; samo 8,83% ima pristup struji, zdravstvena skrb nije usustavljena, građevinska infrastruktura je slabo razvijena i dr. Svi ovi razlozi nameću pretpostavku da je životni vijek nizak.

Najveću očekivanu dob žena ima Japan, koji je poznat po svom zdravom načinu prehrane i učinkovitom zdravstvenom sustavu. Japan je država s visokim BDP-om, što itekako utječe na opremljenost bolnica i općenitu kvalitetu življenja.



Slika 3. Dijagram pravokutnika za očekivanu životnu dob žena

Iz dijagrama pravokutnika na Slici 3. uočavamo da su dvije države s najmanjom očekivanom životnom dobi žena (Čad i Lesoto) u stvari outlieri, tj. jako su udaljeni od većine ostalih podataka, što ima smisla, budući da se radi o vrlo nerazvijenim zemljama čija se očekivana životna dob žena znatno razlikuje od razvijenijih zemalja.



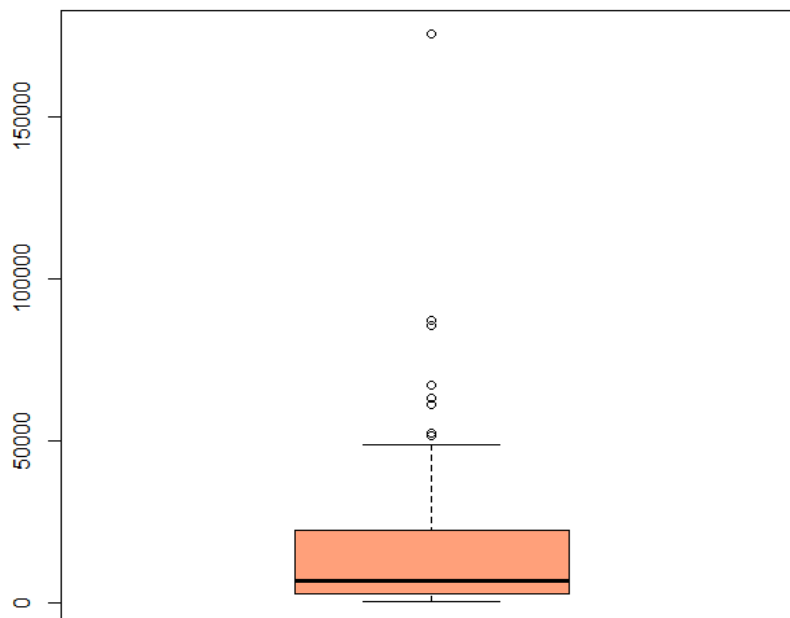
Slika 4. Histogram frekvencija za očekivanu životnu dob žena

Sada analiziramo BDP.

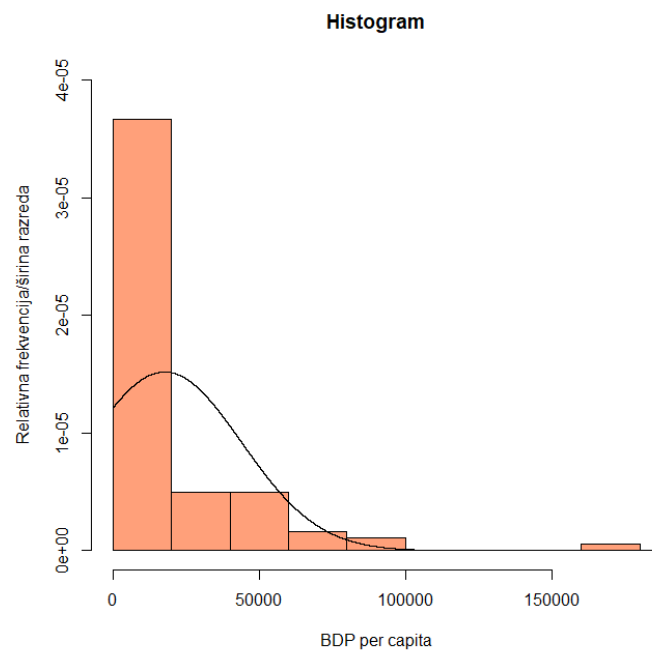
Karakteristična petorka

Najmanja vrijednost	Donji kvartil	Medijan	Gornji kvartil	Najveća vrijednost
239.000	2651.925	6839.250	22944.450	175813.900

Aritmetička sredina podataka je 17723.69 američkih dolara, što je bliže gornjem kvartilu. To ukazuje na nepovoljnu raspoređenost podataka; točnije, država s visokim BDP-om ima malo i one značajno kvare prosjek (što vidimo s donjeg dijagrama pravokutnika i histograma na Slikama 5. i 6.). Iz karakteristične petorke vidimo da je raspon BDP-a poprilično velik, tj. najveća vrijednost je čak 735 puta veća od najmanje vrijednosti. Budući da je gornji kvartil također zamjetno manji od najveće vrijednosti, zaključujemo da je država s vrlo visokim BDP-om samo nekoliko. Najmanja vrijednost pripada afričkoj državi Burundi, a slijede ju ostale afričke zemlje u kojima je životni standard također znatno niži.



Slika 5. Dijagram pravokutnika za BDP



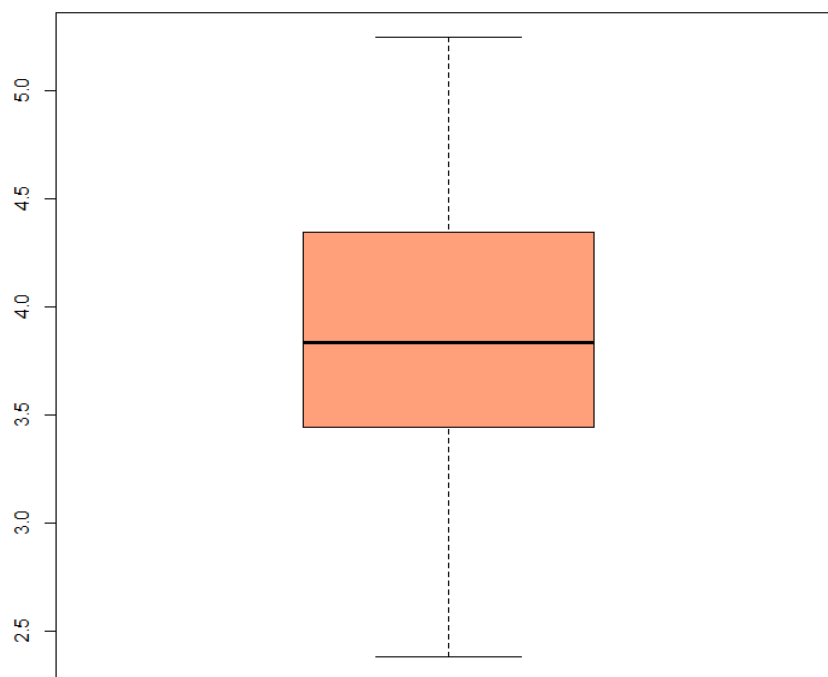
Slika 6. Histogram frekvencija za BDP per capita

Na kraju analiziramo i logaritmirani BDP.

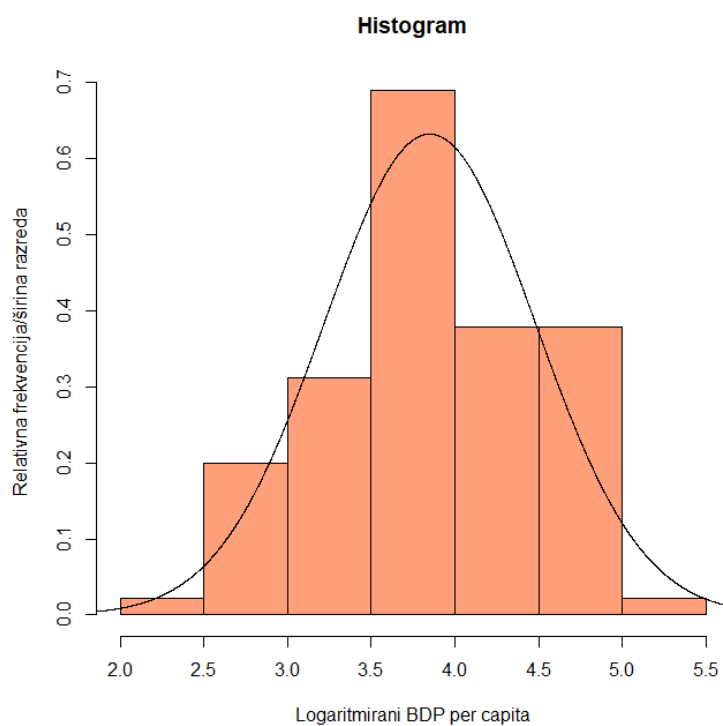
Karakteristična petorka

Najmanja vrijednost	Donji kvartil	Medijan	Gornji kvartil	Najveća vrijednost
2.378398	3.421775	3.835000	4.360010	5.245053

Ovdje vidimo da su podaci ujednačeniji, na što nam ukazuje činjenica da je aritmetička sredina, u iznosu od 3.849918, vrlo blizu medijana. Pogledajmo dijagram pravokutnika i histogram za bolju predodžbu.



Slika 7. Dijagram pravokutnika za logaritmirani BDP per capita



Slika 8. Histogram frekvencija za logaritmirani BDP per capita

Sada na dijagramu pravokutnika (Slika 7.) nema više outliera, a histogram (Slika 8.) izgleda znatno ujednačenije, gotovo s naznakama normalne razdiobe. Konkretno, rubnih (vrlo siromašnih i bogatih) zemalja ima malo, a najmnogobrojnije su srednje bogate države.

To se uvelike razlikuje od histograma nelogaritmiranih podataka, na kojem je siromašnih država nerazmjerno puno.

INFERENCIJALNA STATISTIKA: HIPOTEZE, STATISTIČKI TESTOVI I REZULTATI

Kako bismo ispitali stupanj linearne povezanosti varijabli, koristimo Pearsonov koeficijent korelacije. Računanjem u R-u dobivamo da Pearsonov koeficijent između BDP-a i očekivane životne dobi žena iznosi 0.5131627, što ukazuje na pozitivnu koreliranost promatranih varijabli. Ipak, zbog nezanemarive udaljenosti dobivenog koeficijenta od broja 1, kojim je definirana jaka koreliranost, promotrit ćemo i Pearsonov koeficijent između logaritmiranog BDP-a i očekivane životne dobi žena. Računanjem u R-u dobivamo da traženi koeficijent iznosi 0.8384299, što daje znatno veću koreliranost podataka. Iz tog razloga, za daljnja razmatranja koristit ćemo logaritmirane podatke o BDP-u kako bismo dobili jasniji model.

Neka je slučajna varijabla X logaritmirani BDP po glavi stanovnika, a Y očekivana životna dob žena. Uz pretpostavku da je slučajni vektor (X, Y) normalno distribuiran (kasnije ćemo objasniti zašto ova pretpostavka ima smisla), provodimo test koreliranosti na nivou značajnosti od 5%. Pearsonov koeficijent korelacije je statistika

$$R = \frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}}$$

Testiramo hipoteze:

$H_0: \rho = 0$ (varijable X i Y su nekorelirane)

$H_1: \rho > 0$ (X i Y su pozitivno korelirane)

Testna statistika uz pretpostavku H_0 dana je s $T = \frac{R}{\sqrt{1-R^2}} \sqrt{n-2}$ i ima Studentovu t-razdiobu s $n-2$ stupnja slobode. Kritično područje je $C = [t_{0.05}(n-2), +\infty >$, gdje smo iz tablice Studentove t-razdiobe očitali vrijednost $t_{0.05}(88)$.

Realizacija testne statistike iznosi $t = \frac{0.8384299}{\sqrt{1-0.8384299^2}} \sqrt{88} = 14.4313 \in C$, zbog čega odbacujemo H_0 u korist H_1 . Zaključujemo, koreliranost je pozitivna na razini značajnosti 5%.

Koristimo metodu linearne regresije kako bismo ispitali linearnu zavisnost logaritmiranog BDP-a po glavi stanovnika i očekivane životne dobi žena u nekoj državi. Zatim određujemo 95%-tne intervale pouzdanosti za parametre regresije te provodimo test značajnosti linearnog regresijskog modela.

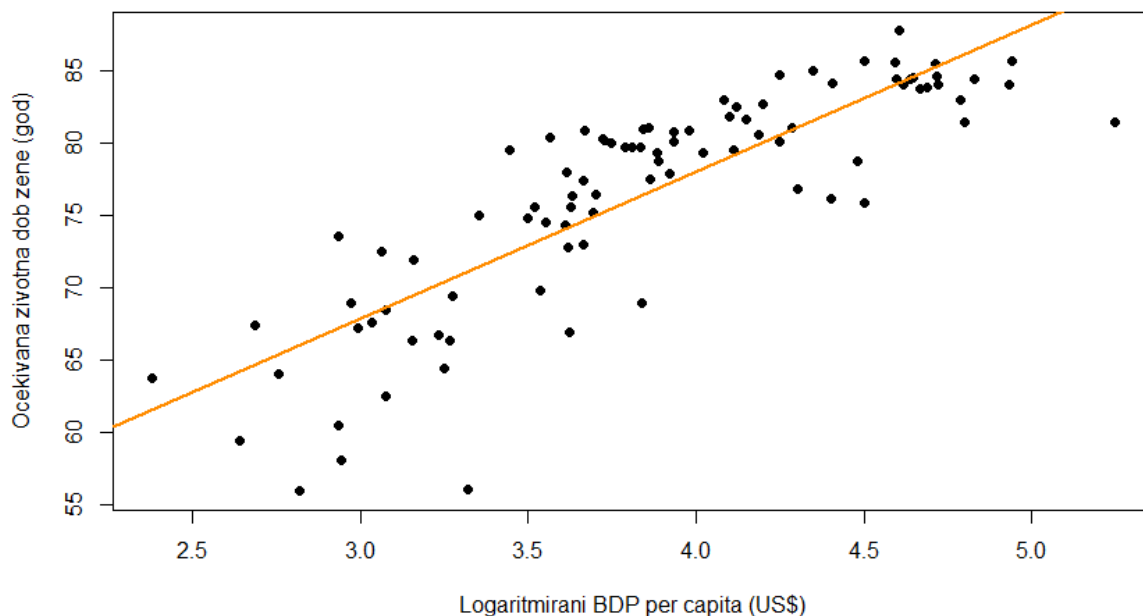
Početna pretpostavka našeg modela je linearna povezanost između očekivane vrijednosti životne dobi žena i logaritmiranog BDP-a po glavi stanovnika neke države, odnosno $E[Y|x] = \alpha + \beta x$. Za stvarne vrijednosti varijable Y ne zahtijevamo da nužno leže na pravcu pa u model uračunavamo slučajnu pogrešku ε uz pretpostavku da je ona normalno distribuirana s očekivanjem 0. Dakle, model je oblika $Y = \alpha + \beta x + \varepsilon$.

Nepristrane procjenitelje $\hat{\alpha}$ i $\hat{\beta}$ za parametre regresije računamo, primijenjujući metodu najmanjih kvadrata, pomoću formula $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ i $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$, gdje su \bar{x} i \bar{y} redom aritmetičke sredine podataka iz uzorka za varijable X i Y , a $S_{XX} = \sum_{i=1}^n x_i^2 - n\bar{x}^2$, $S_{XY} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$ za $n = 90$.

Provođenjem ovog postupka u R-u, za procjenitelje parametara dobivamo

$$\hat{\alpha} = 37.296 \text{ i } \hat{\beta} = 10.1897.$$

Dakle, pravac koji modelira podatke uzorka je $y = 37.296 + 10.189x$ i prikazan je na Slici 9.

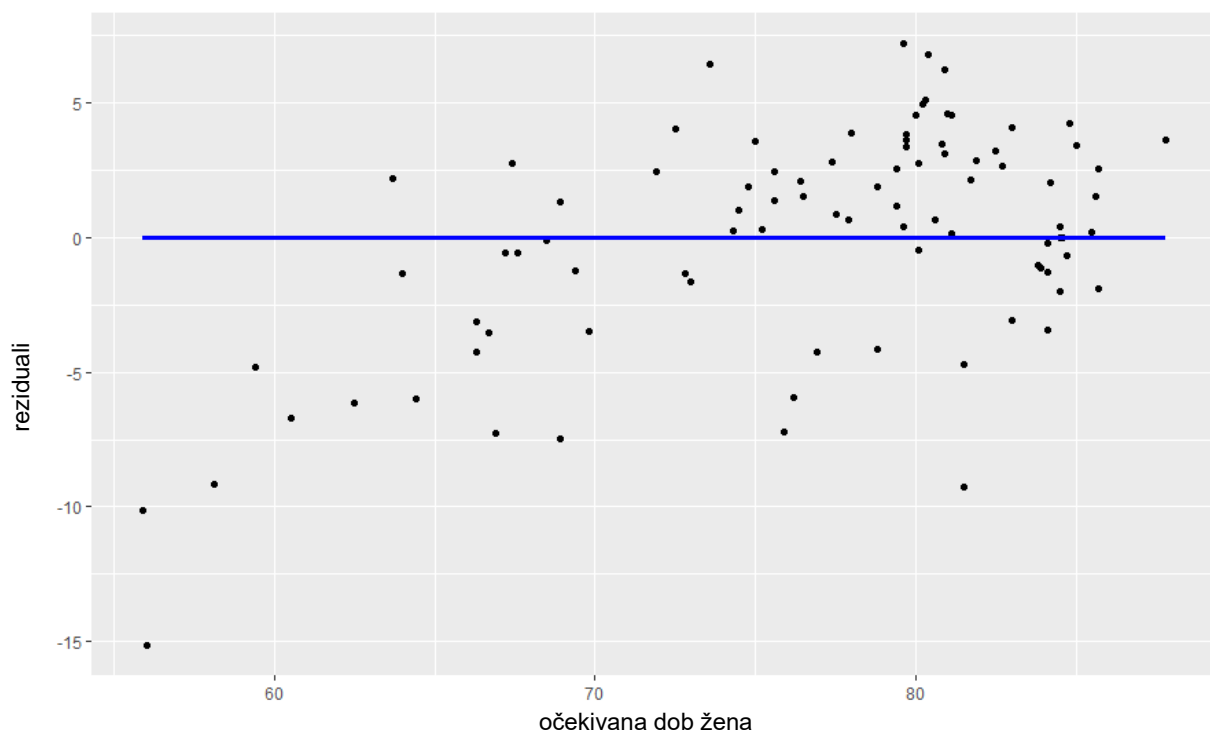


Slika 9. Koordinatni prikaz transformiranih podataka zajedno s regresijskim pravcem

Za daljnju analizu, testiramo homogenost i normalnu distribuiranost reziduala. Pomoću R-a računamo reziduale naše linearne regresije (Slika 10.) te prikazujemo očekivane vrijednosti varijable Y i reziduale (Slika 11.).

```
Residuals:
    Min       1Q   Median       3Q      Max
-15.1395  -1.8566   0.6515   2.8387   7.2041
```

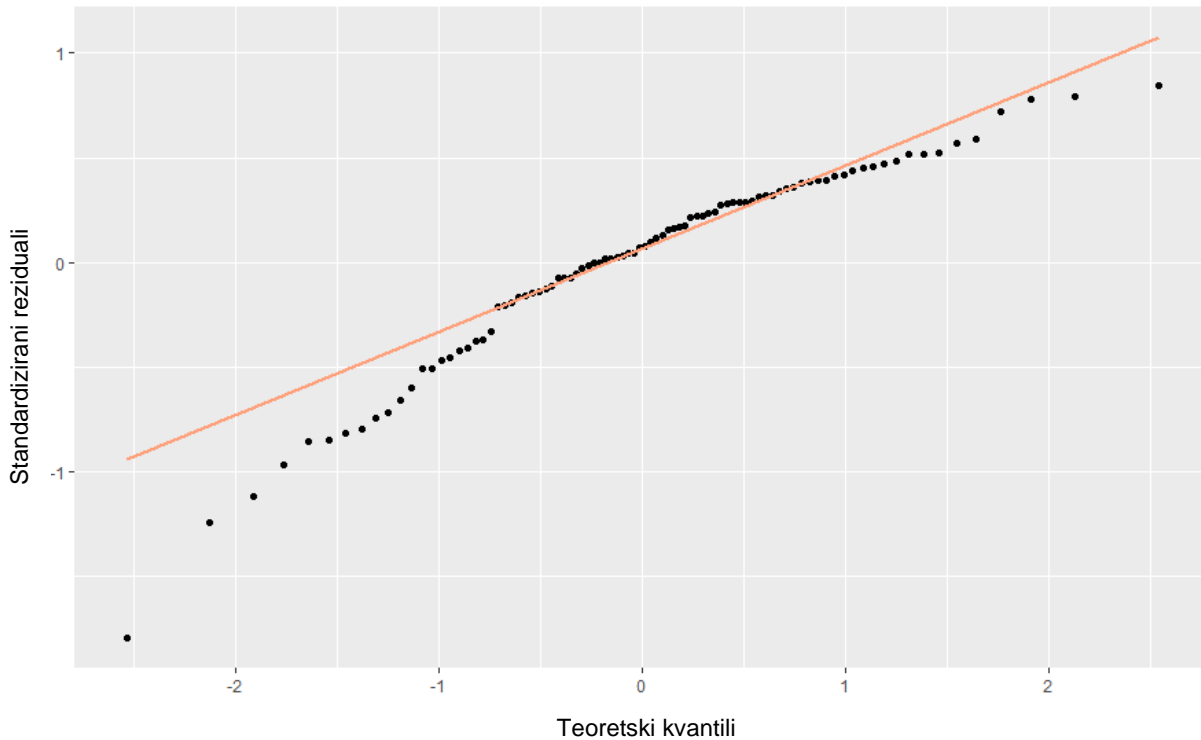
Slika 10. Snimka zaslona dobivenih reziduala linearnom regresijom



Slika 11. Grafički prikaz odnosa reziduala i predviđenih vrijednosti za varijablu Y koristeći dobiveni model

Na Slici 11. opažamo raspršenost podataka bez značajnog grupiranja oko neke vrijednosti ili područja, što ukazuje na linearnost podataka. Uvjet homogenosti reziduala ispitujemo promatrajući njihov raspon za različite očekivane vrijednosti te uočavamo da je zadovoljen za očekivanu vrijednost veću od 60, dok je za manje možda narušen.

Nadalje, za testiranje normalnosti reziduala crtamo graf kvantila (Q-Q plot) prikazan na Slici 12. Na njemu su uspoređeni kvantili standardiziranih reziduala naših podataka i teoretski kvantili jedinične normalne distribucije. Na Slici 12. vidljivo je da dobivene točke većinom uz manje oscilacije leže pravcu, što ukazuje na normalnu distribuiranost reziduala.



Slika 12. Q-Q graf standardiziranih reziduala

Sada ćemo konstruirati 95% pouzdane intervale za parametre regresije α i β . Uz pretpostavku o normalnoj distribuiranosti slučajne greške slijedi da sljedeće statistike imaju Studentovu t-distribuciju s $n-2$ stupnja slobode

$$\frac{\hat{\alpha} - \alpha}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}}} \sim t(n-2), \quad \frac{\hat{\beta} - \beta}{\hat{\sigma} \sqrt{\frac{1}{s_{xx}}}} \sim t(n-2)$$

gdje je $\hat{\sigma} = \sqrt{\frac{SSE}{n-2}}$, $SSE = S_{YY} - \hat{\beta}^2 S_{XX}$.

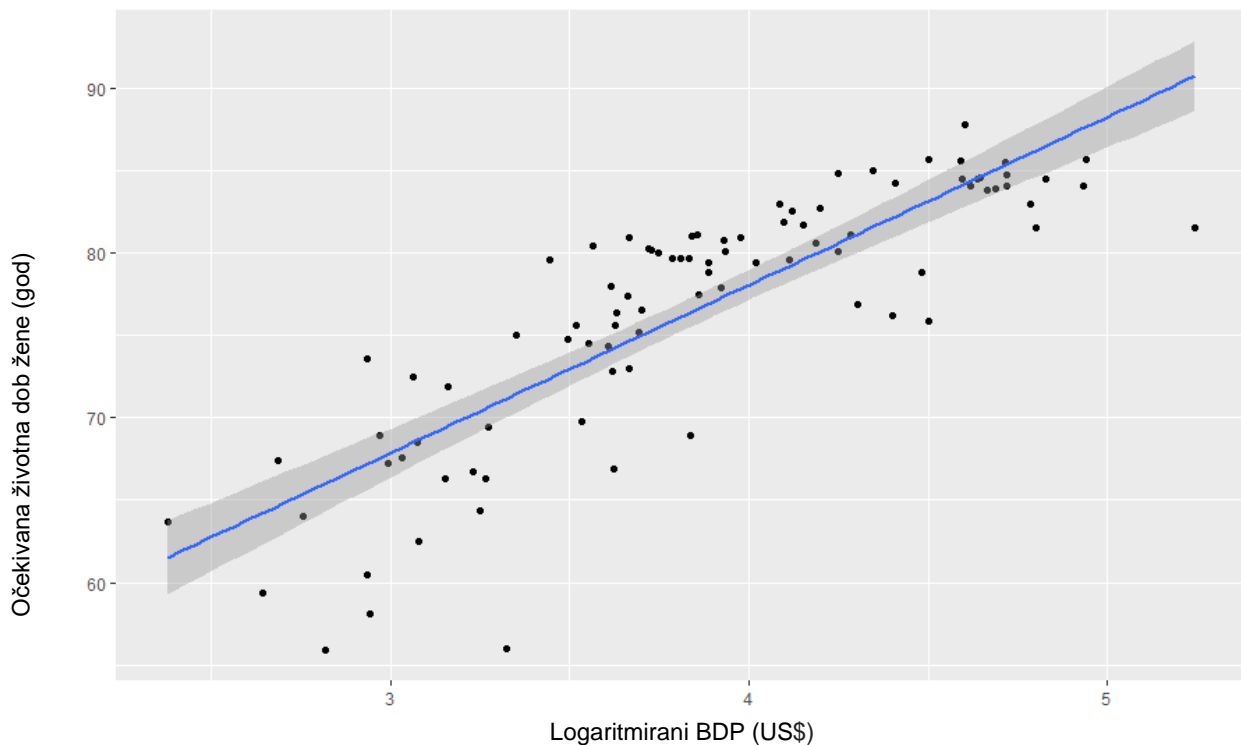
95% pouzdan interval uz razinu značajnosti $\gamma = 0.05$ za α dan je s

$$\left[\hat{\alpha} - t_{\frac{\gamma}{2}}(n-2) \cdot \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}}, \hat{\alpha} + t_{\frac{\gamma}{2}}(n-2) \cdot \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}} \right]$$

te za β s

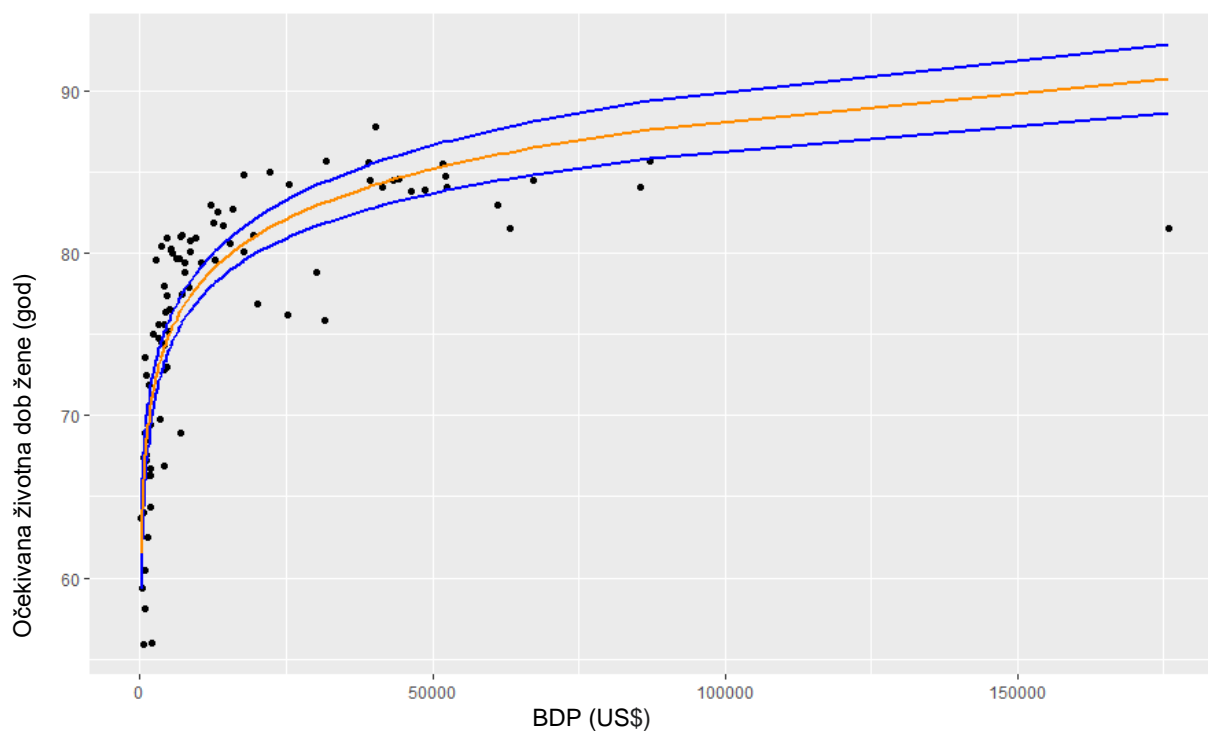
$$\left[\hat{\beta} - t_{\gamma/2}(n-2) \cdot \hat{\sigma} \sqrt{\frac{1}{s_{xx}}}, \hat{\beta} + t_{\gamma/2}(n-2) \cdot \hat{\sigma} \sqrt{\frac{1}{s_{xx}}} \right]$$

Dakle, dobivamo da je jedna procjena za 95% pouzdani interval za parametar α jednaka [31.822945, 42.76908], a za parametar β je [8.785763, 11.59192], što grafički vidimo na Slici 13.



Slika 13. Koordinatni prikaz varijabli zajedno s regresijskim pravcem i pouzdanim intervalom za $E[Y|x]$

Radi usporedbe, slično možemo napraviti s regularnim BDP-om. Neka je sada slučajni vektor Z regularni BDP. Koristimo supstituciju $z = e^x$ te promatramo očekivanu životnu dob žena i BDP. U R-u crtamo graf funkcijske ovisnosti naših varijabli i granice 95% pouzdanog intervala za $E[Y|z]$, što je prikazano na Slici 14.



Slika 14. Koordinatni prikaz varijabli zajedno s grafom njihove funkcijske ovisnosti i pouzdanim intervalom za $E[Y|z]$

Provest ćemo test značajnosti linearnog regresijskog modela na nivou 0.05 kako bismo se uvjerali da on ima smisla u našem slučaju. Testiramo hipoteze:

$H_0: \beta = 0$ (nema linearne ovisnosti očekivane životne dobi žena i logaritmiranog BDP-a)

$H_1: \beta \neq 0$

Ranije smo odredili da je za parametar β jedna procjena 95% pouzdanog intervala interval $[8.785763, 11.59192]$. Budući da $0 \notin [8.785763, 11.59192]$, odbacujemo H_0 u korist H_1 na nivou značajnosti 0.05, odnosno sa sigurnošću 95% tvrdimo da očekivana životna dob žena linearno ovisi o logaritmiranoj vrijednosti BDP-a.

Na kraju, provodimo Lillieforsov test (inačicu Kolmogorov-Smirnovljevog testa) kako bismo ispitali imaju li naše varijable normalnu distribuciju. Prvo ćemo ga provesti na podacima za logaritmirani BDP.

$$H_0: X \sim N(\mu, \sigma^2)$$

$$H_1: ne H_0$$

```
> lillie.test(log)

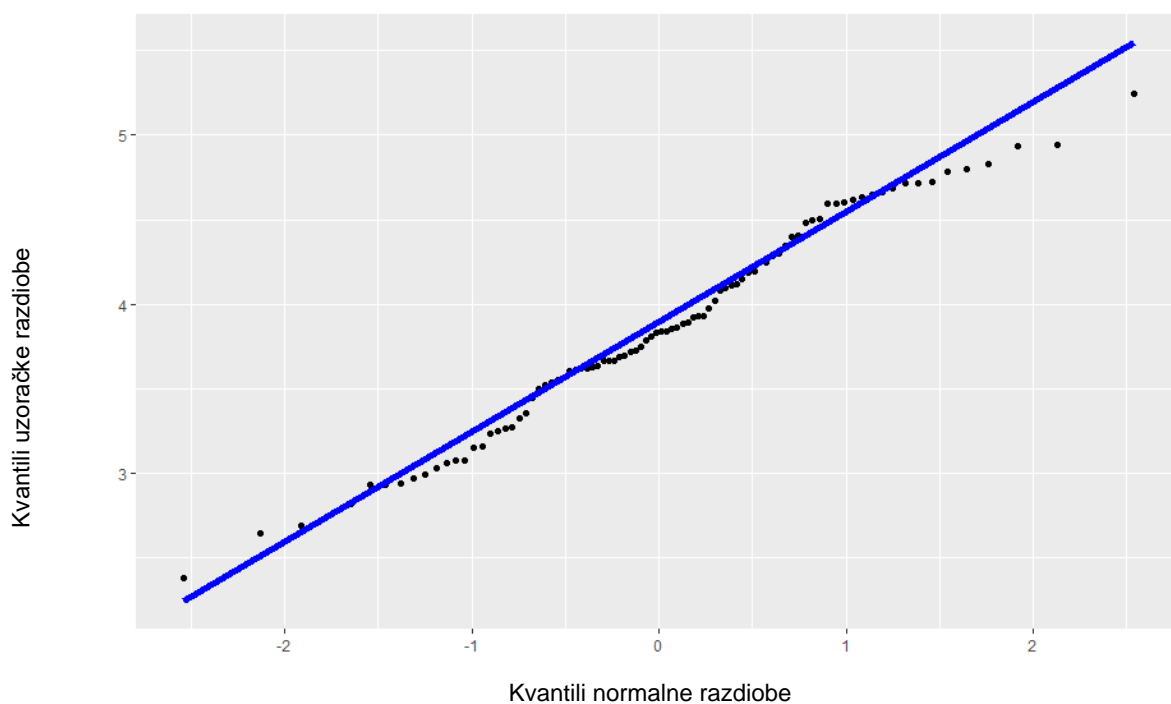
Lilliefors (Kolmogorov-Smirnov) normality test

data: log
D = 0.068847, p-value = 0.3665
```

Slika 15. Snimka zaslona rezultata Lillieforsovog testa za logaritmirani BDP pc

Prema rezultatima sa Slike 15., p -vrijednost jednaka je $0.3665 \gg 0.05$ pa ne odbacujemo pretpostavku H_0 na nivou značajnosti od 5%, odnosno nemamo dovoljno dokaza da bismo odbacili pretpostavku da su logaritmirane vrijednosti BDP-a normalno distribuirane.

Smislenost rezultata Lillieforsovog testa potvrđuje i Q-Q (kvantil-kvantil graf) koji prikazuje korelaciju podataka odabrane varijable i normalne distribucije. U idealnom slučaju, kada se teorijska distribucija (u našem slučaju normalna) poklapa sa stvarnom distribucijom podataka, kvantili uzoračke razdiobe i teorijske razdiobe se poklapaju i leže na istom pravcu. Iz Slike 16. vidimo da smo za logaritmirani BDP dosta blizu idealnom slučaju.



Slika 16. Q-Q graf logaritmiranog BDP-a

Sada provodimo i Lillieforsov test za očekivanu životnu dob žena.

$$H_0: Y \sim N(\mu, \sigma^2)$$

$$H_1: ne H_0$$

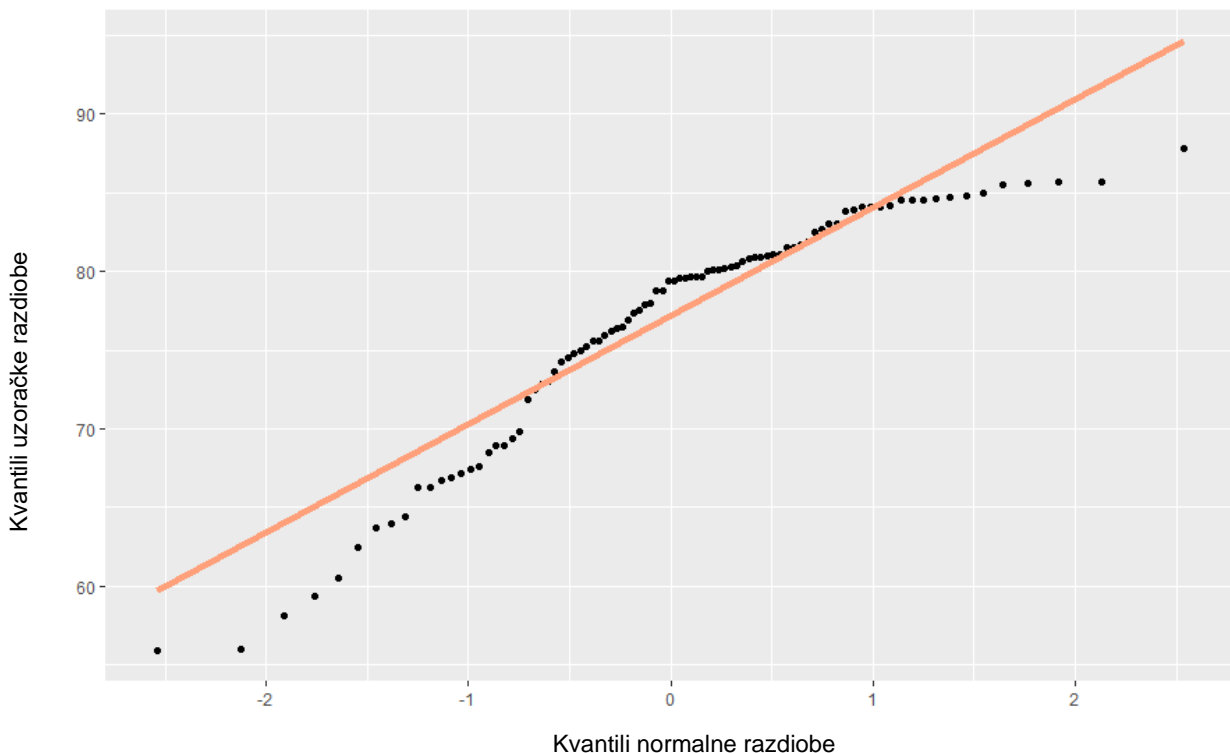
```
> lillie.test(ocekivana)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  ocekivana
D = 0.15734, p-value = 9.475e-06
```

Slika 17. Snimka zaslona rezultata Lillieforsovog testa za očekivanu životnu dob žena

Prema rezultatima sa Slike 17., p -vrijednost iznosi $0.000009475 \ll 0.05$, što znači da odbacujemo pretpostavku H_0 u korist pretpostavke H_1 na nivou značajnosti od 5%. Mogući razlog zbog kojeg dolazi do odbacivanja normalnosti varijable Y je to što je ona zavisna varijabla, pa se time njena distribucija mijenja za svaku vrijednost od X odnosno $Y_i \sim N(\mu_i, \sigma_i^2)$ $i=1, \dots, 90$.



Slika 18. Q-Q graf očekivane životne dobi žena

Iz Slike 18. također vidimo da očekivana životna dob žena nije potpuno normalno distribuirana, ali ima linearan trend, iz čega je imalo smisla u testu koreliranosti pretpostaviti da je (X, Y) ,

gdje je X logaritmirani BDP, a Y očekivana životna dob žena, normalno distribuiran slučajni vektor. Također, ako je uzorak veći od 30 podataka (u našem slučaju 90), tada po Centralnom graničnom teoremu možemo pretpostaviti da je on normalno distribuiran.

ZAKLJUČAK

Primjenom raznih statističkih metoda, analizirali smo povezanost očekivane životne dobi žena i BDP-a pojedinih država te smo napokon u mogućnosti donijeti zaključke o našoj početnoj pretpostavci.

Provedbom Lillieforsovog testa zaključujemo da su, na nivou značajnosti od 5%, logaritmirani podaci za BDP razmatranih država normalno distribuirani. S druge strane, isti test je odbacio normalnu distribuiranost očekivane dobi žena. Međutim, Centralni granični teorem i dalje osigurava smislenost provođenja testa koreliranosti, budući da smo promatrali uzorak od 90 mjerenja.

Iz tog razloga, ranije provedeni test koreliranosti povlači pozitivnu koreliranost logaritmiranog BDP-a i očekivane životne dobi žena, na razini značajnosti od 5%.

Nadalje, ispitali smo i funkcijsku ovisnost između promatranih varijabli, te smo ju najbolje moguće aproksimirali regresijskim pravcem $y = 37.296 + 10.189x$. Testiranjem značajnosti ovog modela na nivou od 5%, uspijevamo zaključiti da su varijable zaista približno linearno zavisne.

Dakle, statistička analiza je pokazala da su promatrane varijable uistinu korelirane. Štoviše, ispitali smo i jače svojstvo – linearnu zavisnost logaritmiranih vrijednosti BDP-a i očekivane životne dobi (do na normalno distribuiranu grešku modela), te sada sa sigurnošću od 95% možemo tvrditi da što je veći BDP pojedine države (jer je $x \rightarrow \ln(x)$ strogo rastuća funkcija na $(2,6)$), to je i očekivana životna dob žena iste države veća, što podržava polaznu pretpostavku.

LITERATURA

1 <https://www.gapminder.org/data/>

2 <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

3 <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519>

A. Mimica, M. Ninčević, *Statistika primjeri i zadaci*, Zagreb, 2010.

Ž. Pauše, *Uvod u matematičku statistiku*, Školska knjiga, Zagreb, 1993.