Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni diplomski studij matematike
Financijska matematika i statistika

Matej Petrinović

Statistika

Seminarski rad

Analiza baze podataka siromaštva svijeta

Voditelj: prof. dr. sc. Mirta Benšić

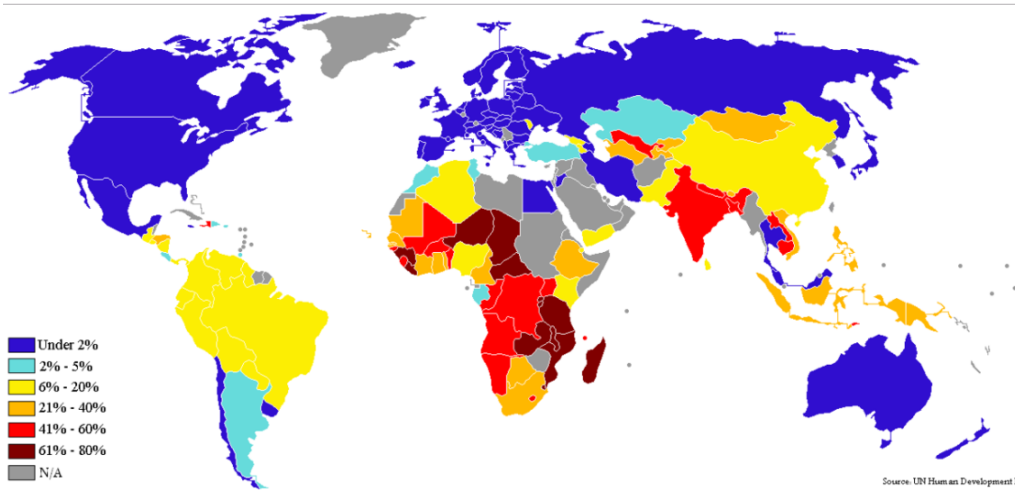
Osijek, 2018.

Sadržaj

1 Uvod

Baza podataka *Poverty* sadrži podatke o 97 zemalja svijeta. Podaci su uzeti iz knjiga UNESCO 1990 Demographic Year Book (1990), New York: United Nations i Day, A. (ed.) (1992), The Annual Register 1992, 234, London: Longmans. Za ove zemlje dani su podaci za broj rođenih, umrlih, stope smrtnosti dojenčadi, očekivane životne dobi kod muškaraca i žena, te bruto domaći proizvod (BDP).

Siromaštvo je termin koji se najčešće koristi za nedostatak osnovnih uvjeta za život. Biti siromašan znači ne imati dovoljno novaca (ili nekih drugih sredstava) za priuštiti si osnovne ljudske potrebe kao što su hrana, piće, dom, itd. Prema definiciji UN-a, siromašnima se smatraju osobe koje su odreknute načina života, komfora i dostojanstva, koji se smatraju normalnim u društvu u kojem žive. Svjetska banka kaže da je krajnje siromaštvo kada čovjek živi sa manje od jednim dolarom na dan.



Slika 1. Karta svijeta koja pokazuje broj ljudi u pojedinim državama koji žive sa manje od jednim američkog dolara na dan.

U ovom seminarskom radu analizirat ćemo opisanu bazu *Poverty*. Napravit ćemo analizu nataliteta i mortaliteta i ispitati koja stopa je veća. Analizirat ćemo očekivane dobi muškaraca i žena, te ispitati vezu između očekivane dobi i BDP-a.
Baza podataka je preuzeta sa sljedećeng linka:
`https://ww2.amstat.org/publications/jse/jse_data_archive.htm`

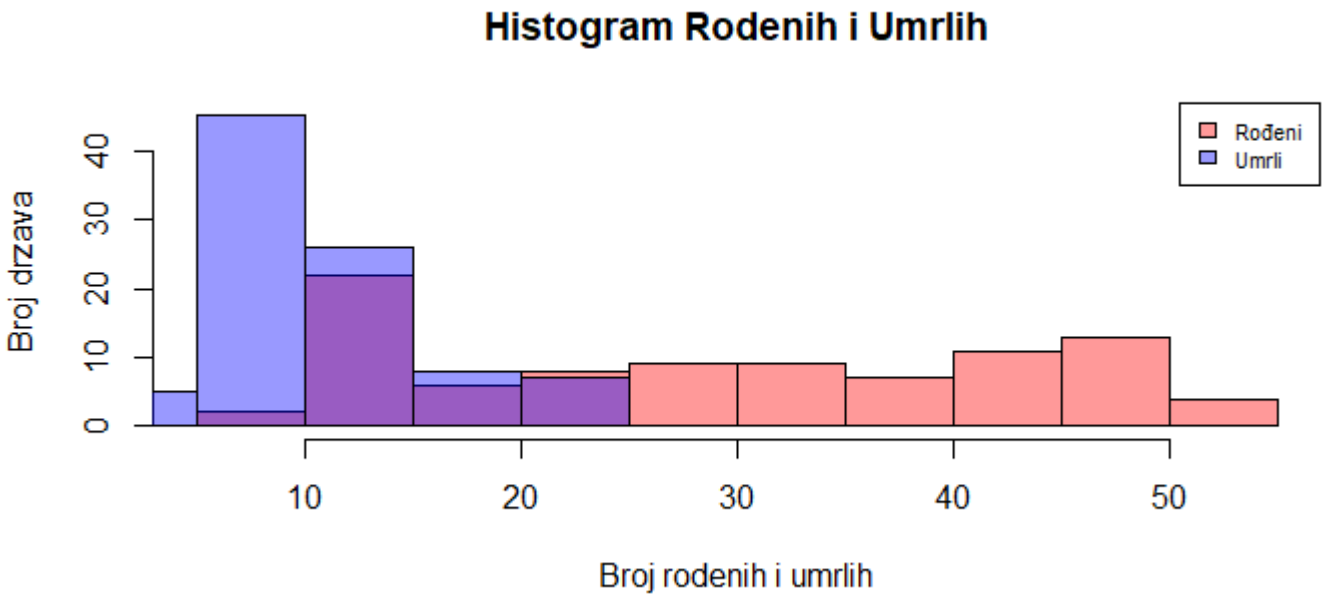
2 Deskriptivna statistika varijabli

2.1 Rođeni i Umrli

Varijable *Rođeni* i *Umrli* su numeričkog tipa i sadrže informacije o broju rođenih, odnosno umrlih ljudi na 1000 osoba. Osnovne informacije o varijablama *Rođeni* i *Umrli* mogu se vidjeti na sljedećoj tablici.

	Minumum	Donji kvartil	Medijan	Prosjek	Gornji kvartil	Maksimum
Rođeni	9.70	14.70	29.00	29.46	42.55	52.20
Umrli	2.20	7.70	9.50	10.73	12.30	25.00

Tablica 1. Deskriptivna statistika Rođenih i Umrlih



Slika 2. Histogram varijabli Rođeni i Umrli

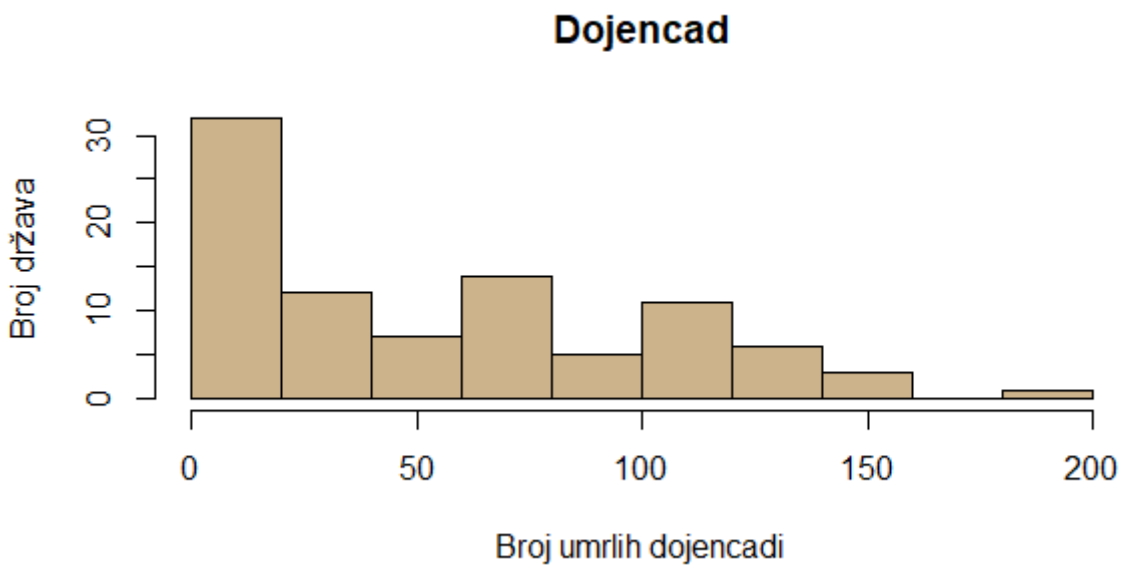
Možemo vidjeti da je broj rođenih zasigurno veći od broja umrlih. Najveći broj rođenih ljudi ima Uganda, dok najmanji broj ima Italija. Najveći broj umrlih ljudi ima Zimbabve, a najmanje ima Kuvajt.

2.2 Dojenčad

Varijabla *Dojenčad* je numeričkog tipa i sadrži informacije o broju umrle djece ispod dobi od godine dana na 1000 osoba. Osnovne informacije mogu se vidjeti u danoj tablici.

Minumum	Donji kvartil	Medijan	Prosjek	Gornji kvartil	Maksimum
4.50	13.05	43.00	55.28	86.50	181.60

Tablica 3. Deskriptivna statistika varijable Dojenčad



Slika 3. Histogram varijable Dojenčad

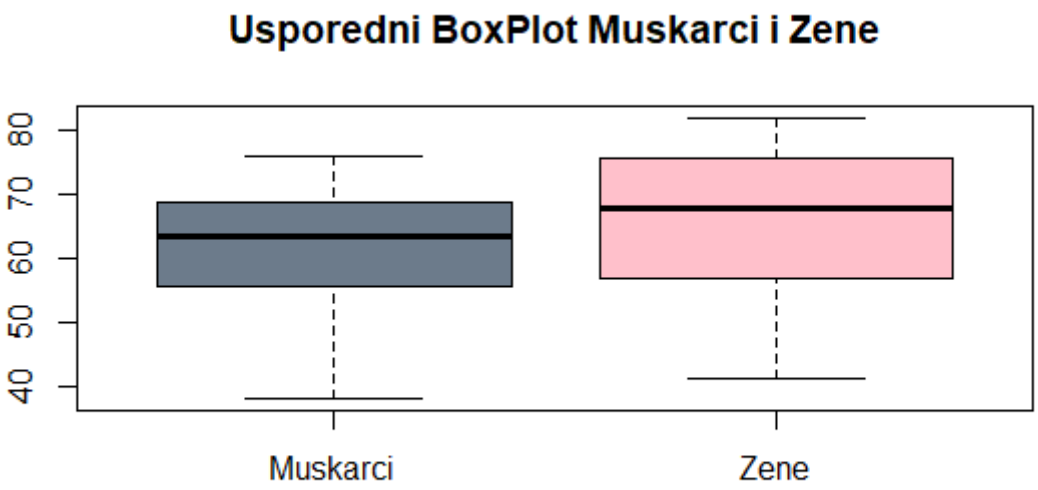
Iz histograma možemo vidjeti da veliki broj država ima mali broj umrle djece ispod godinu dana, dok mali broj država ima više od 150 umrle djece ispod godine dana. Zemlje koje imaju preko 150 umrle djece ispod godine dana su Afganistan i Sierra Leone.

2.3 Muškarci i Žene

Ove dvije varijable *Muškarci* i *Žene* su numeričke varijable koje prikazaju očekivanu dob muškaraca odnosno žena. Očekivano trajanje života statistička je mjera prosječnog vremena za koje se očekuje da će organizam živjeti na temelju godine rođenja, njezine trenutne dobi i drugih demografskih čimbenika uključujući spol. Osnove podatke možemo vidjeti u danoj tablici.

	Minumum	Donji kvartil	Medijan	Prosjek	Gornji kvartil	Maksimum
Žene	41.20	56.75	67.60	66.03	75.45	81.80
Muškarci	38.10	55.40	63.40	61.38	68.50	75.90

Tablica 4. Deskriptivna statistika varijabli Muškarci i Žene



Slika 4. Usporedni BoxPlot očekivane dobi za žene i muškarce

Iz slike kutijastih dijagrama možemo uočiti veće vrijednosti za žene nego kod muškaraca. Najveće očekivane dobi imaju stanovnici Japana, gdje je očekivana dob za muškarce 75.9, te za žene 81.8 godina.

2.4 BDP

Varijabla *BDP* opisuje bruto domaći proizvod (BDP) po osobi. Bruto domaći proizvod je makroekonomski indikator koji pokazuje vrijednost finalnih dobara i usluga proizvedenih u zemlji tijekom dane godine, izraženo u novčanim jedinicama. U našem slučaju BDP je izražen u američkim dolarima. Osnovne podatke možemo vidjeti u sljedećoj tablici.

Minumum	Donji kvartil	Medijan	Prosjek	Gornji kvartil	Maksimum
80	475	1690	5741	7325	34064

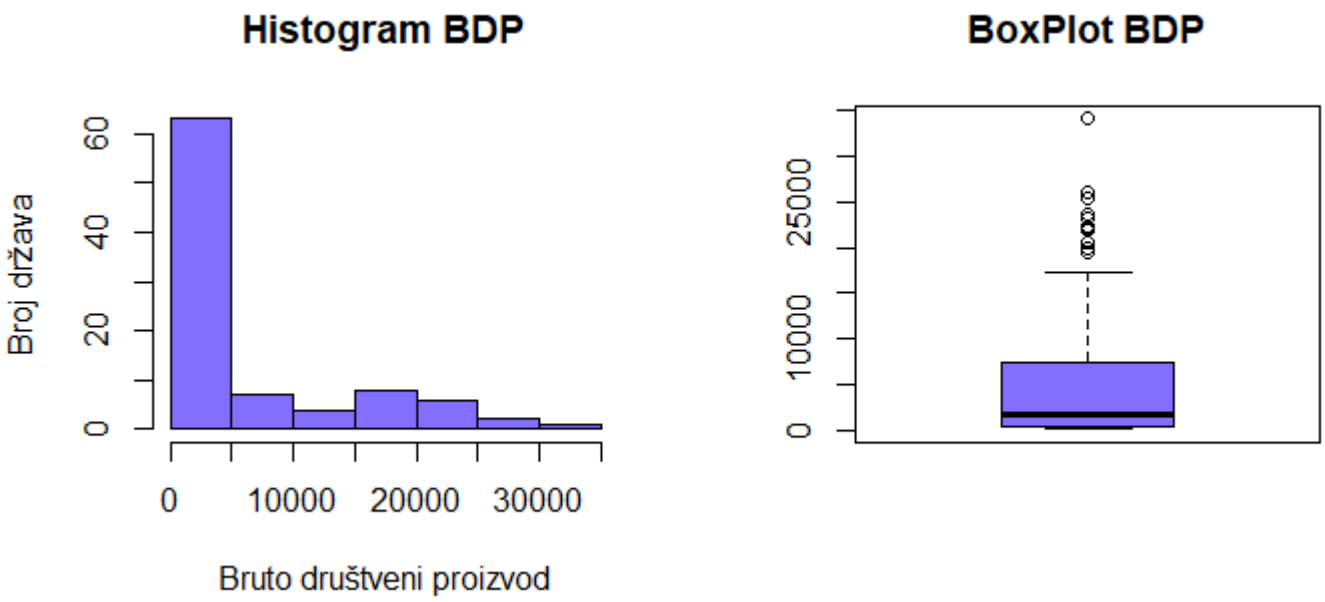
Tablica 5. Deskriptivna statistika varijable BDP

Realni BDP (oznaka Y) možemo izračunati koristeći matematičku formulu

$$Y = C + I + G + E - U$$

pri čemu je C osobna potrošnja, I nacionalne investicije, G državna potrošnja, E izvoz i U uvoz. U našem slučaju se radi o BDP-u per capita kojeg dobijemo djeljenjem realnog BDP-a sa ukupnim brojem stanovništva (ozn. S)

$$Y_{\text{per capita}} = \frac{Y}{S}$$



Slika 5. Histogram i kutijasti dijagram BDP-a

Iz histograma vidimo da s povećanjem BDP-a broj zemalja se smanjuje. Na kutijastom dijagramu uočavamo stršeće vrijednosti. Najveća stršeća vrijednost je Švicarska čiji BDP iznosi čak 34064 $, dok zemlja s najmanjim BDP-om je Monzabik koji iznosi svega 80 $.

2.5 Grupa i Ime

Varijabla *Ime* sadrži ime države, a varijabla *Grupa* je kategorijalna varijabla koja opisuje grupu kojoj država pripada i to na način:

1. Istočna Europa
2. Južna Amerika i Meksiko
3. Zapadna Europa, Sjeverna Amerika, Japan, Australija, Novi Zeland
4. Bliski Istok
5. Azija
6. Afrika

U sljedećoj tablici možemo vidjeti broj zemalja po grupama.

Grupa	1	2	3	4	5	6
Broj zemalja	9	12	19	10	14	27

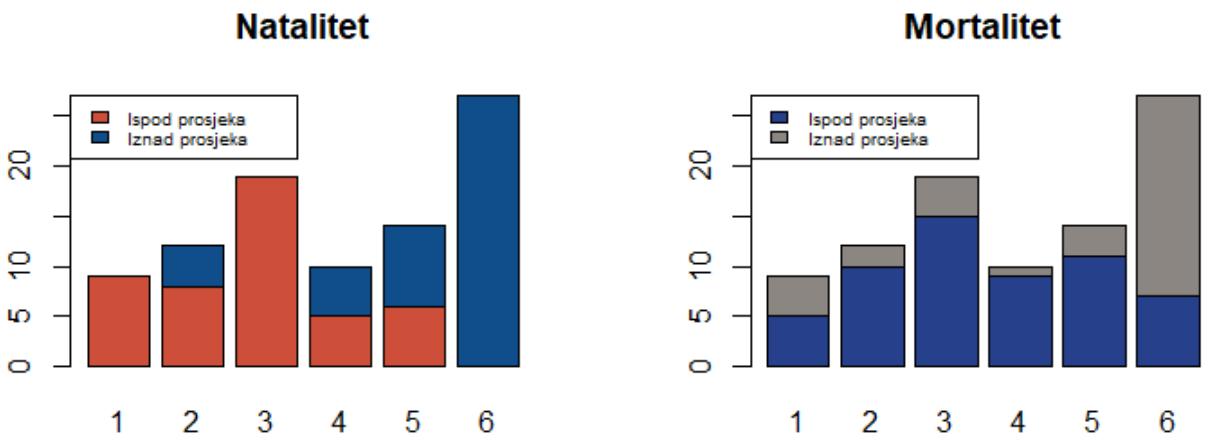
Tablica 6. Broj zemalja po grupama

3 Natelitet i mortalitet

Prije nego li krenemo s analizom podataka, definirajmo pojmove natalitet i mortalitet. Natalitet je stručni izraz kojim se koristi demografija, a označava ukupno rađanje na određenom području u određenom vremenu. Njegova visina se mjeri stavljanjem u omjer broja rođene djece − obično samo živorođene−prema ukupnom broju stanovništva. Najčešće se u razne svrhe razmatranja nataliteta i problematike uz to vezane koristi stopa nataliteta. Stopa nataliteta se obično računa na 1000 stanovnika i to uz pomoć matematičke formule

$$n = \frac{N \cdot 1000}{S}$$

pri čemu je n stopa, N broj živorođene djece i S ukupan broj stanovnika.
Mortalitet je demografski pokazatelj koji označava određeni broj smrtnih slučajeva stanovništva na temelju ukupnog broja stanovništva u određenom razdoblju (obično jedne godine) i pokazatelj je zdravstva. Mortalitet je omjer broja umrlih na prosječni broj stanovnika



Slika 6. Usporedba stupčastih dijagrama nataliteta i mortaliteta

Možemo li na razini značajnosti $\alpha = 0.05$ tvrditi da je očekivani natalitet Afrike (μ_0) veći od očekivanog nataliteta ostatka svijeta (μ_1)? Postavljamo sljedeće hipoteze:

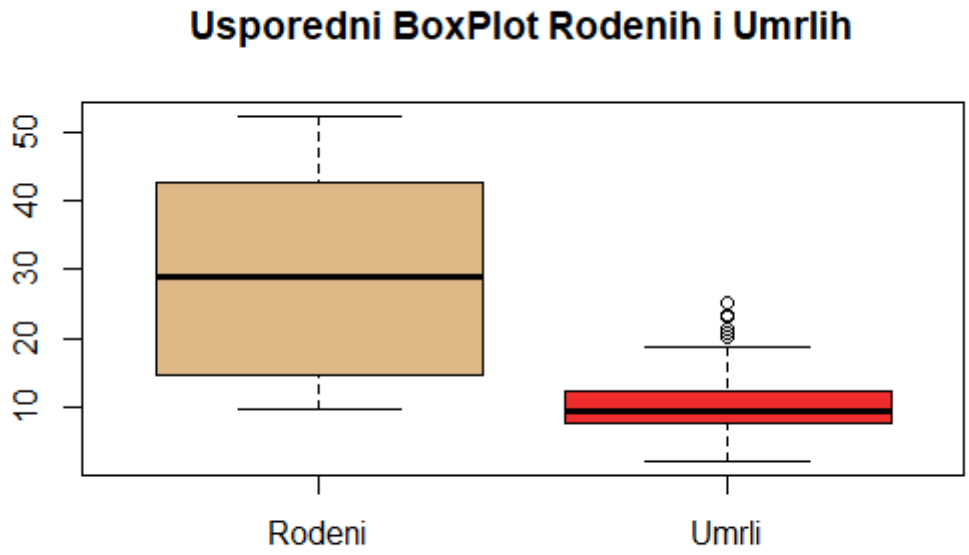
$$H_0 : \mu_0 = \mu_1$$
$$H_1 : \mu_0 > \mu_1$$

Koristeći $Welchov\, t - test$ za dva nezavisna uzorka dobivamo $p-$vrijednost manju od $2.2 \cdot 10^{-16}$, što je manje od 0.05, pa na razini značajnosti 0.05 odbacujemo H_0 te prihvaćamo H_1, tj. da je očekivana stopa nataliteta Afrike veća od očekivane stope nataliteta ostatka svijeta. Natalitet ostatka svijeta je statistički značajno manji od nataliteta Afrike. Očekivani natalitet Afrike iznosi 44.53, a dok u ostatku svijeta 23.1

Možemo li na razini značajnosti $\alpha = 0.05$ tvrditi da je očekivani mortalitet zemalja Bliskog istoka (μ_0) manji od očekivanog mortaliteta ostatka svijeta (μ_1)? Postavljamo sljedeće hipoteze:

$$H_0 : \mu_0 = \mu_1$$
$$H_1 : \mu_0 < \mu_1$$

Koristeći $Welchov\, t - test$ za dva nezavisna uzorka dobivamo $p-$vrijednost 0.0001163 što je manje od 0.05, pa na razini značajnosti 0.05 odbacujemo hipotezu H_0 te prihvaćamo H_1, tj. da je očekivani mortalitet država Bliskog manji od očekivanog mortaliteta ostalih zemala svijeta. Prikažimo usporedne kutijaste dijagrame nataliteta i mortaliteta.



Slika 7. Uspredni kutijasti dijagrami nataliteta i mortalitea

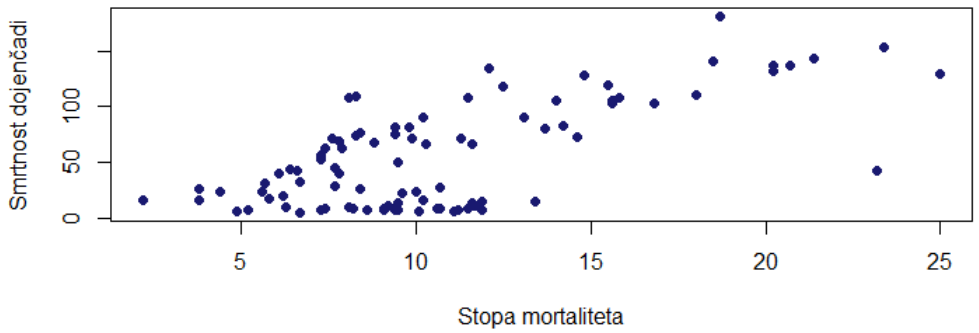
Možemo li na razini značajnosti $\alpha = 0.05$ tvrditi da je distribucija nataliteta veća od distribucije mortaliteta? Postavit ćemo sljedeće hipoteze:

$$H_0 : \text{distribucije su jednake}$$
$$H_1 : \text{distribucija nataliteta je veća}$$

Koristeći $Wilcoxov\, test$ dobijamo $p-$vrijednost $< 2.2 \cdot 10^{-16} < 0.05$, pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo H_1, tj. možemo tvrditi da je distribucija nataliteta veća.

4 Mortalitet i smrtnost dojenčadi

Možemo se zapitati raste li smrtnost djece sa rastom mortaliteta. Prikažimo točkasti graf ove dvije varijable.



Slika 9. Točkasti prikaz stope mortaliteta i smrtnosti dojenčadi

Na razini značajnosti $\alpha = 0.05$ ispitajmo postoji li rastuća veza između mortaliteta i smrtnosti dojenčadi, tj. raste li broj umrle djece sa porastom mortalteta. Postavimo hipoteze:

$$H_0 : \rho_S = 0$$
$$H_1 : \rho_S > 0$$

Korištenjem Spearmanove metode $cor - testa$ dobivamo p-vrijednost $8.696 \cdot 10^{-8} < 0.05$ i zaključujemo da postoji rastuća veza.

5 Očekivane dobi

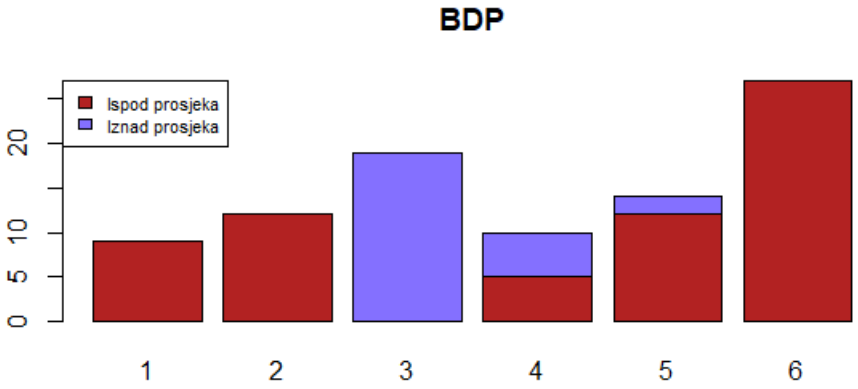
U naslovu 2.4 vidjeli smo da je prosječna očekivana dob i medijan dobi žena veći u odnosu na muškarce. Možemo li na razini značajnosti $\alpha = 0.05$ tvrditi da je dob žena veća od dobi muškaraca?

$$H_0 : \text{dobi su jednake}$$
$$H_1 : \text{dob žena je veća}$$

Koristeći $Wilcoxov\, test$ dobivamo $p-$vrijednost 0.0005939 što je manje od 0.05, pa na razini značajnosti 0.05 odbacujemo H_0 te prihvaćamo H_1, tj. možemo tvrditi da je dob žena veća.

6 Bruto domaći proizvod

Nameće nam se prirodno pitanje ovisi li BDP o grupi zemalja u kojoj se država nalazi. Pogledajmo sljedeći stupčasti dijagram na kojemu je prikazano grupe država i ispod ili iznad prosječnost za BDP.



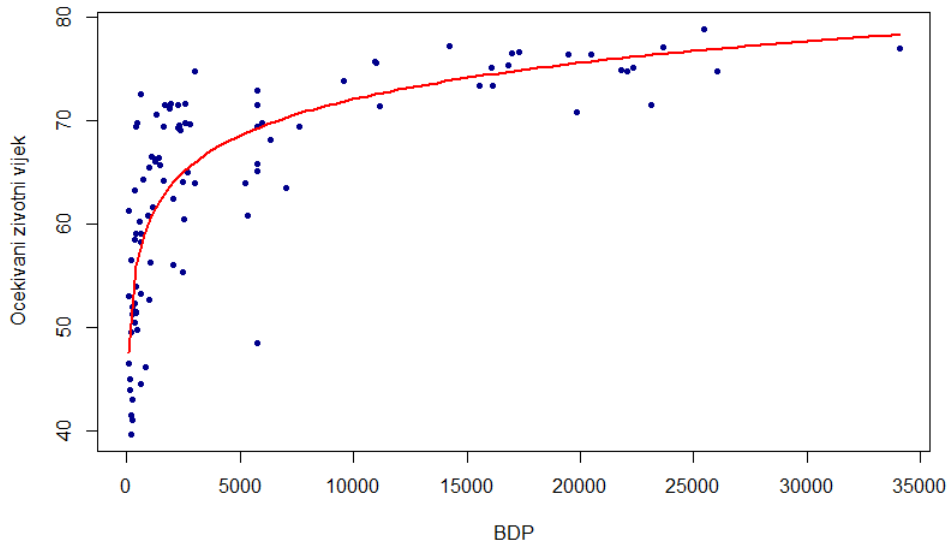
Slika 10. Stupčasti dijagram BDP-a po grupama

Iz stupčastog dijagrama možemo vidjeti da sve zemlje Zapadne Europa, Sjeverne Amerike, Japan, Australija, Novi Zeland imaju BDP iznad prosjeka. Iz grupe 4 BDP veći od prosjeka imaju Bahrein, Izrael,Kuvaj, Saudijska Arabija i UAE, dok iz grupe 5 su to Singapur i Hong Kong.

Testirajmo sada žive li ljudi u zemljama sa većim BDP-om duže, tj. testirajmo postoji li rastuća veza između očekivane dobi i BDP-a. Za potrebu ovog testa definirana je nova varijabla kao prosječna vrijednost očekivane dobi žena i muškaraca. Testirajmo sljedeće hipoteze:

$$H_0 : \rho_S = 0$$
$$H_1 : \rho_S > 0$$

Korištenjem Spearmanove metode $cor - testa$ dobivamo p-vrijednost $< 2.2 \cdot 10^{-16} < 0.05$ i zaključujemo da postoji monotona rastuća veza. Procjenjena korelacija Spearmanovom metodom iznosi $\hat{\rho} = 0.8127$ što ukazuje na snažnu rastuću vezu.



Slika 11. Točkasti prikaz BDP-a i prosječne dobi

Iz slike možemo vidjeti odnos dobi o BDP-u, te uviđamo kako zaista prosječna dob raste sa bruto domaći proizvodom tj, ljudi u bogatijim zemljama žive duže. Obzirom na jaku rastuću vezu možemo kreirati model linearne regresija oblika $Y = \alpha + \beta \log(X) + \varepsilon$, gdje je $\varepsilon \sim N(0, \sigma^2)$ greška modela. Traženi model iznosi $Y = 25.36 + 5.07 \log(X) + \varepsilon$, $\varepsilon \sim N(0, 6.18)$, te srednje kvadratna greška modela iznosi $MSE = 37.81$, a $R^2 = 63.37$. Ako uzmemo podatke za Hrvatsku, BDP per capita iznosi 15014.09 dolara, te bi po tome modelu očekivani životni vijek bio 74.11 godina, dok zapravo iznosi 78.07.

7 Zaključak

Ovim seminarskim radom, analizom baze podataka *Poverty* došli smo do zaključaka da broj rođenih (natalitet) veći od broja umrlih (mortaliteta). Isto tako uvidjeli smo da broj umlre djece ispod godine dana raste sa mortalitetom. Vidjeli smo da je očekivana dob muškaraca različita od očekivane dobi žena, tj. manja je. Analizom smo došli do zaključka da BDP ne ovisi o skupini u kojoj se zemlja nalazi, te da stanovnici bogatijih zemalja žive duže, tj. očekivana dob ljudi raste s porastom bogatsva zemlje, te je napravljen linearni model koji modelira očekivani životni vijek pomoću logaritma BDP-a.