

Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
Matematički odsjek

Marija Nikolić, Žana Perković, Magdalena Potočnjak, Jurica Preksavec

# Povezanost BDP-a s razvojem željezničkog prometa europskih država

SEMINARSKI RAD

Mentor:  
prof. dr. sc. Miljenko Huzak

Zagreb, 2022.

## Sadržaj

Sažetak .....	1
Uvod .....	2
Opisna statistika .....	3
1. Putnici .....	3
2. Transport dobara .....	5
3. BDP .....	7
Inferencijalna statistika .....	9
Određivanje distribucije ulaznih podataka .....	10
1. Putnici .....	11
2. Transport dobara .....	14
Linearna regresija .....	17
1. Putnici .....	17
2. Transport dobara .....	21
Zaključak .....	25

## Sažetak

Ovim seminarским radom želimo odgovoriti na pitanja vezana uz razvoj željezničkog prometa za države s područja Europe. Promatrat ćemo zasebno prijevoz putnika i transport dobara te istražiti postoji li njihova povezanost s BDP-om pojedine zemlje.

Konkretno, zanimat će nas kako je distribuirana iskorištenost željeznice za putnički prijevoz, a kako za transport robe, tj. koliko se putnika ili robe preveze godišnje u odnosu na duljinu pruge u državi.

Također, ukoliko postoji, povezanost „bogatstva“ zemlje i njenih stanovnika s iskorištenošću postojeće željezničke infrastrukture te države opisat ćemo linearnim regresijskim modelom.

## Uvod

Podatke koje koristimo u ovom seminarskom radu prikupili smo iz baza podataka Eurostata, statističkog ureda Europske unije. Eurostat je odgovoran za objavu točnih statističkih podataka o zemljama iz cijele Europe i omogućuje usporedbu država u raznim područjima. Promatramo podatke za 2019. godinu i za uzorak uzimamo slučajno odabrane države Europe.

S obzirom na razlike u površinama država, brojnosti njihovih stanovnika i duljinama njihovih pruga, želimo vrijednosti tih veličina svesti na korisno usporedive podatke. Stoga ćemo razvoj željezničkog prometa procijeniti na dva načina: prvi će se odnositi na brojnost putnika koji putuju vlakovima, a drugi na to koliko se dobara u pojedinoj zemlji prevozi željezničkim putem.

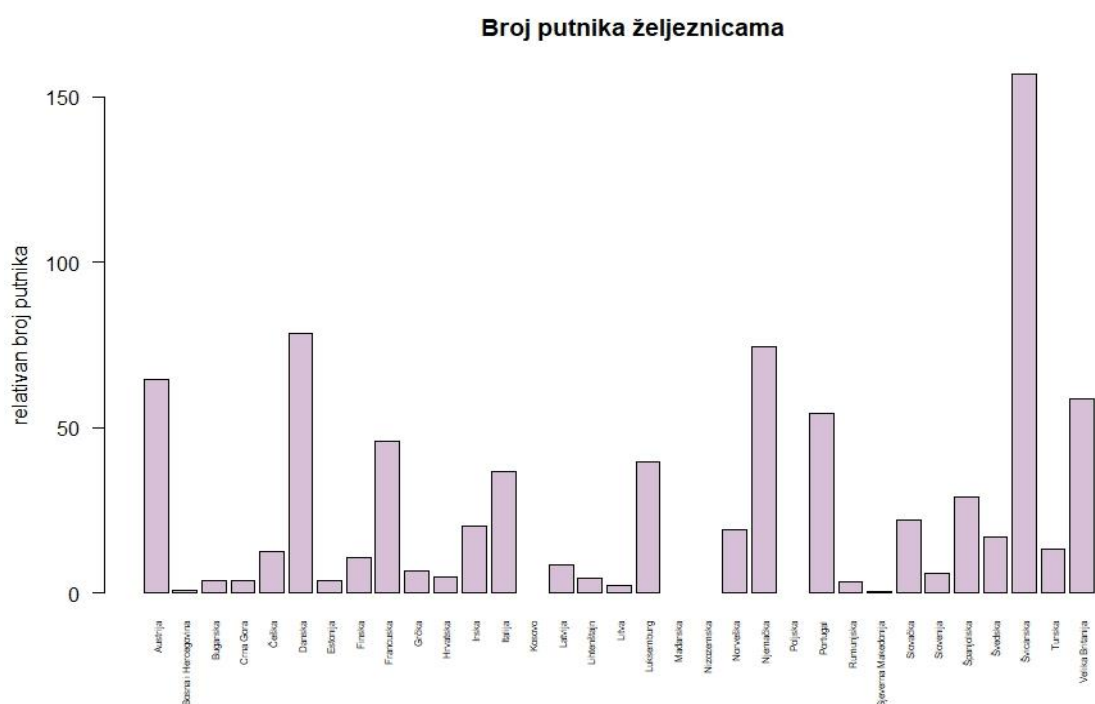
Kako bismo imali usporedive podatke, odlučili smo gledati „relativne“ podatke. Da bismo dobili „relativan“ broj putnika, podijelili smo ukupan broj putnika s ukupnom duljinom željezničke pruge, te analogno za količinu dobara prevezenih vlakovima. Duljina pruge je izražena u kilometrima, broj putnika u tisućama, količina dobara prevezenih željeznicom u tisućama tona, a BDP per capita u eurima.

## Opisna statistika

### 1. Putnici

Podatci o putnicima su bili iskazani u tisućama, a mi smo ih prilagodili tako što smo za svaku državu podijelili broj putnika s duljinom željeznice.

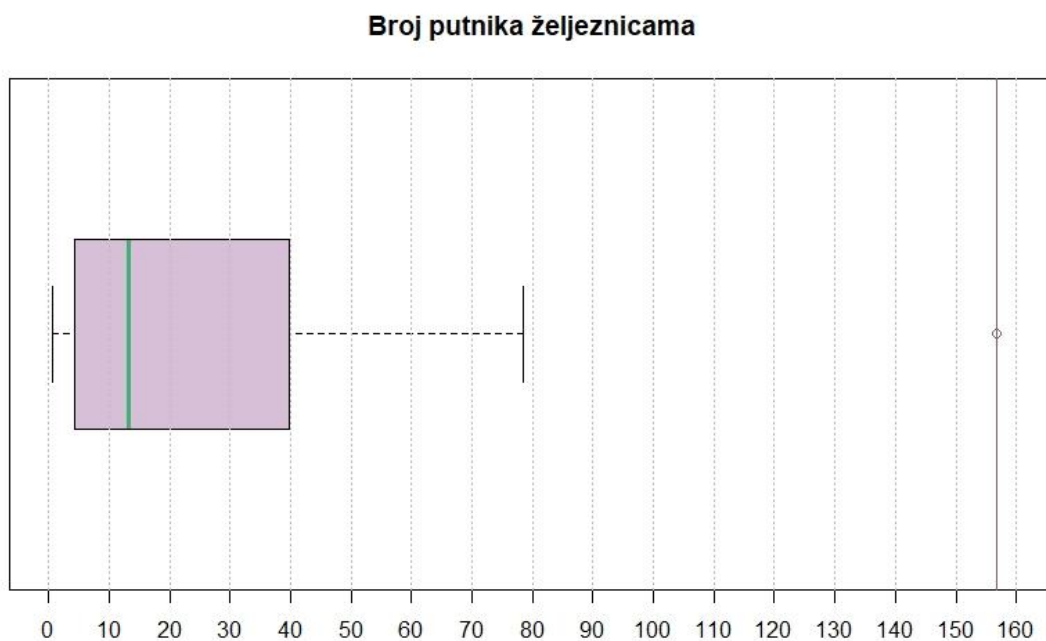
Podatke smo prikazali na više načina. Prvi od njih je stupčasti dijagram.



Slika 1. Stupčasti dijagram broja putnika po državama

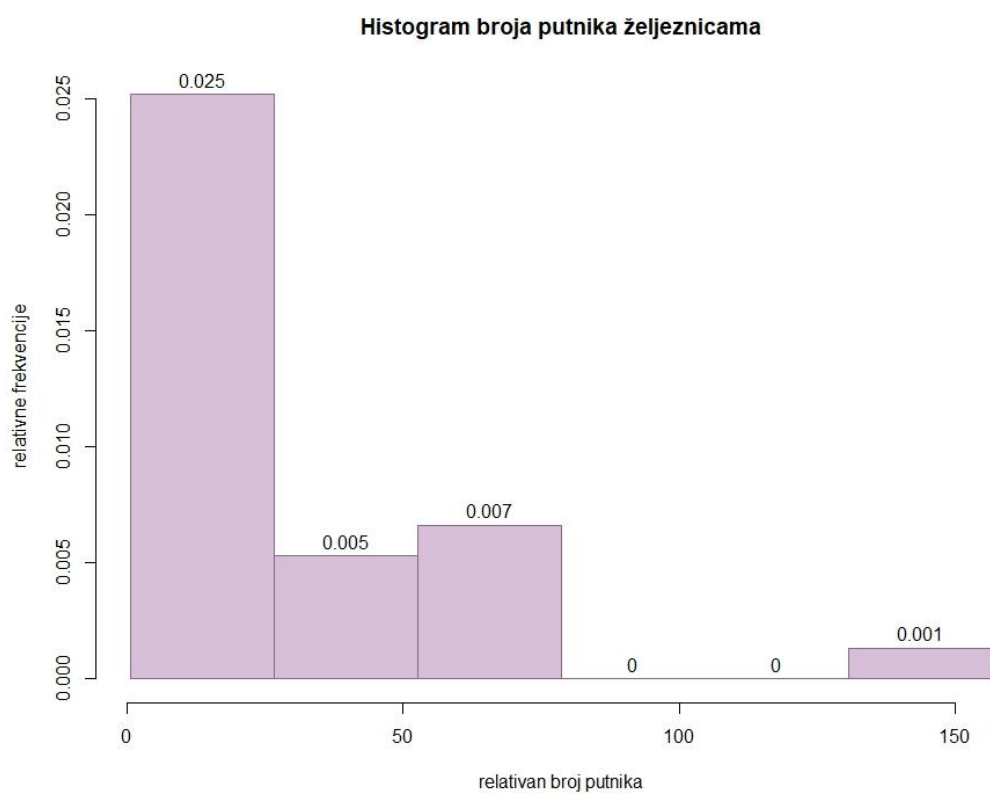
Odredili smo karakterističnu petorku i prikazali podatke pomoću dijagrama pravokutnika.

Minimum	Donji kvartil	Medijan	Gornji kvartil	Maksimum
0,60529	4,13400	13,31544	42,93743	156,74350



Slika 2. Dijagram pravokutnika broja putnika

Naposljetku smo napravili histogram broja putnika.



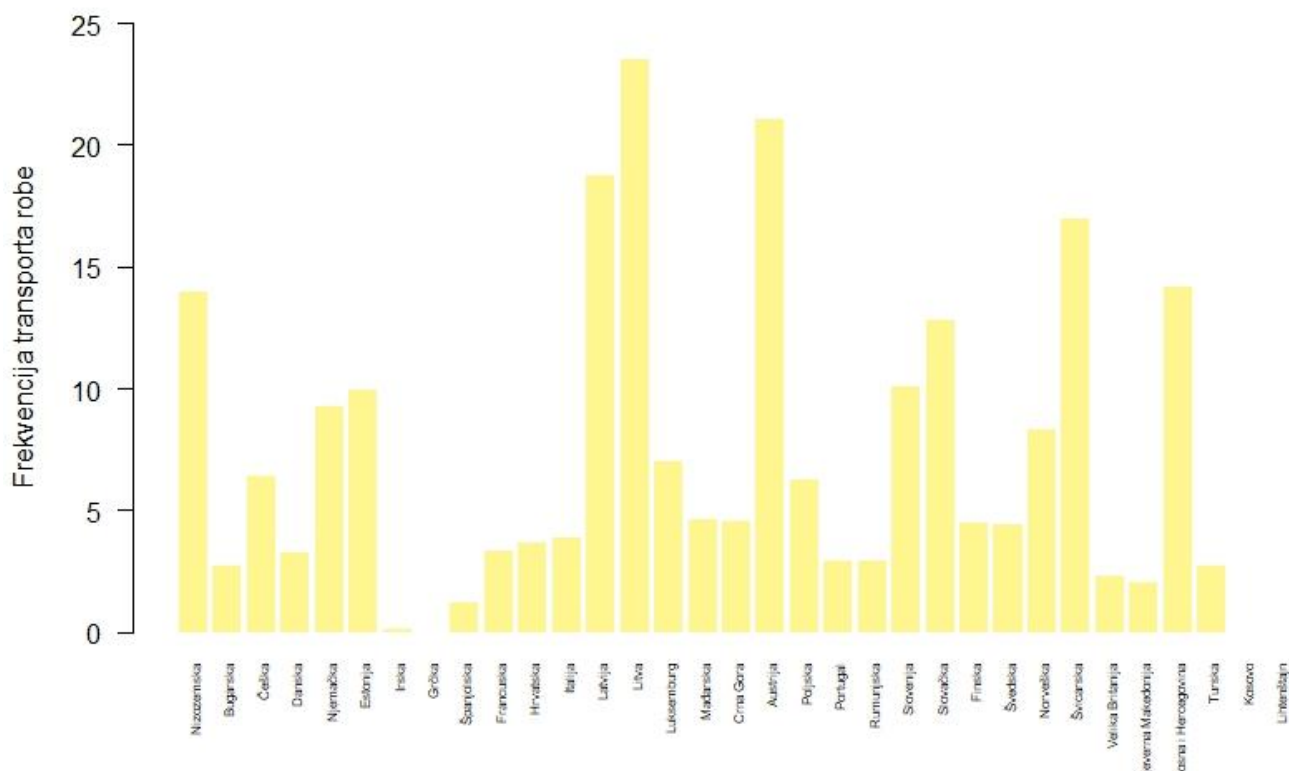
Slika 3. Histogram broja putnika

(slika ne ukazuje jasno na neku poznatu razdiobu)

## 2. Transport dobara

Podatci o količini dobara su iskazani u tisućama tona, i podijeljeni duljinom željezničke pruge.

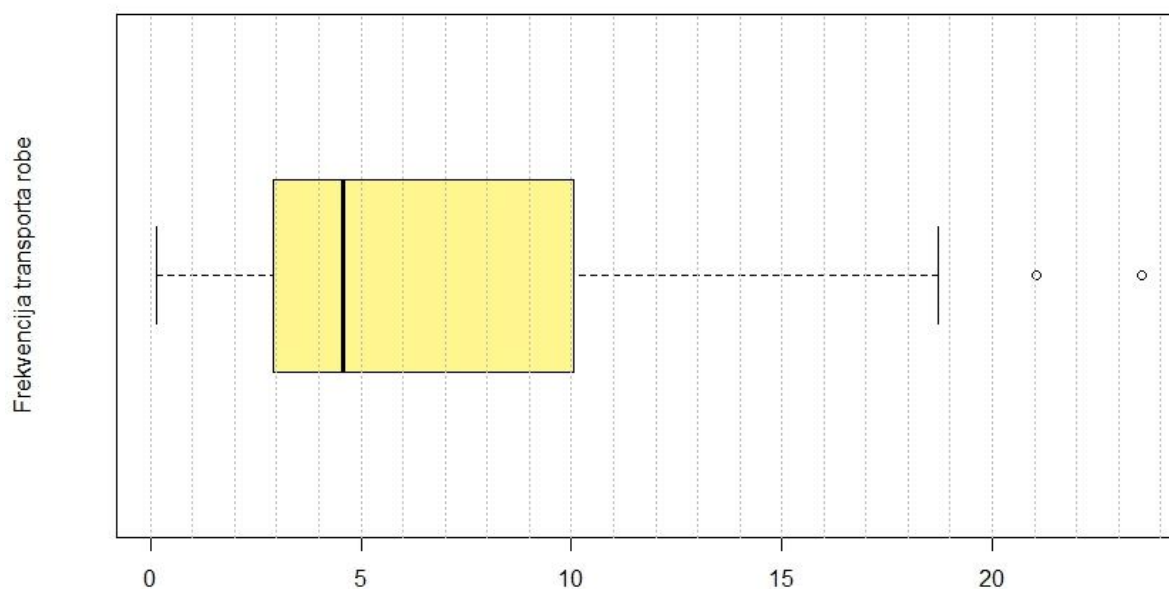
Podatke smo prikazali pomoću stupčastog dijagrama.



Slika 4. Stupčasti dijagram transporta dobara

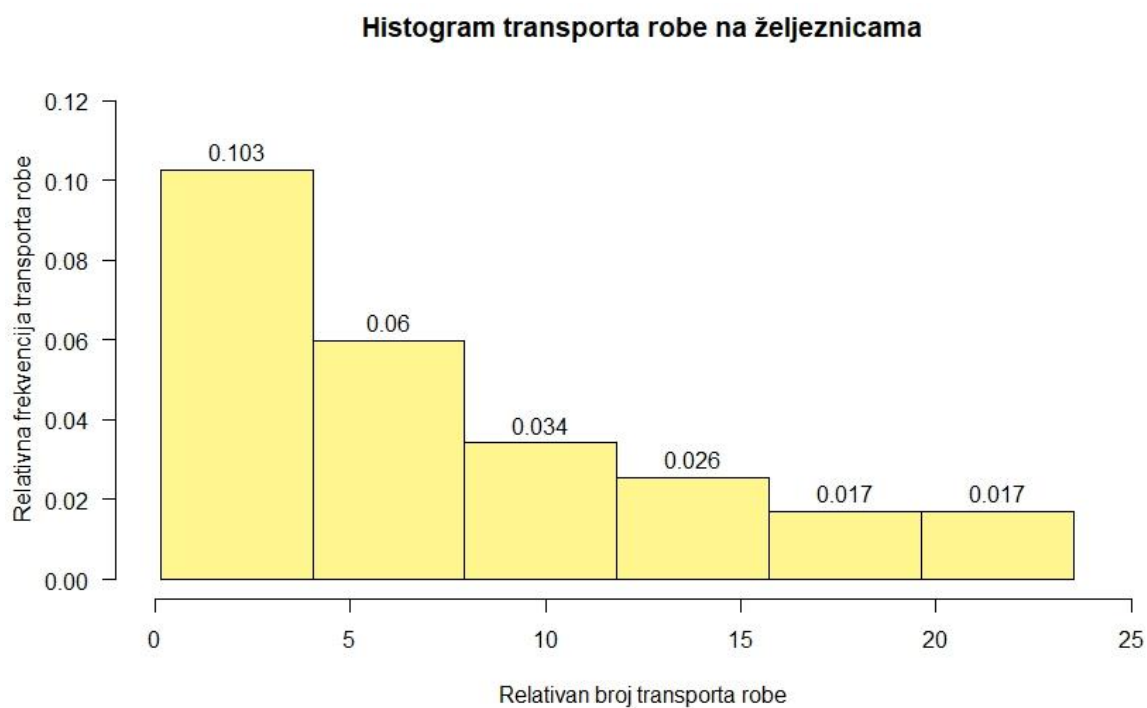
Odredili smo karakterističnu petorku i napravili dijagram pravokutnika za ove podatke.

Minimum	Donji kvartil	Medijan	Gornji kvartil	Maksimum
0.1396851	2.9228175	4.5656580	10.7393794	23.5332481



Slika 5. Dijagram pravokutnika transporta dobara

Naposljetku smo napravili histogram transporta dobara.



Slika 6. Histogram transporta dobara

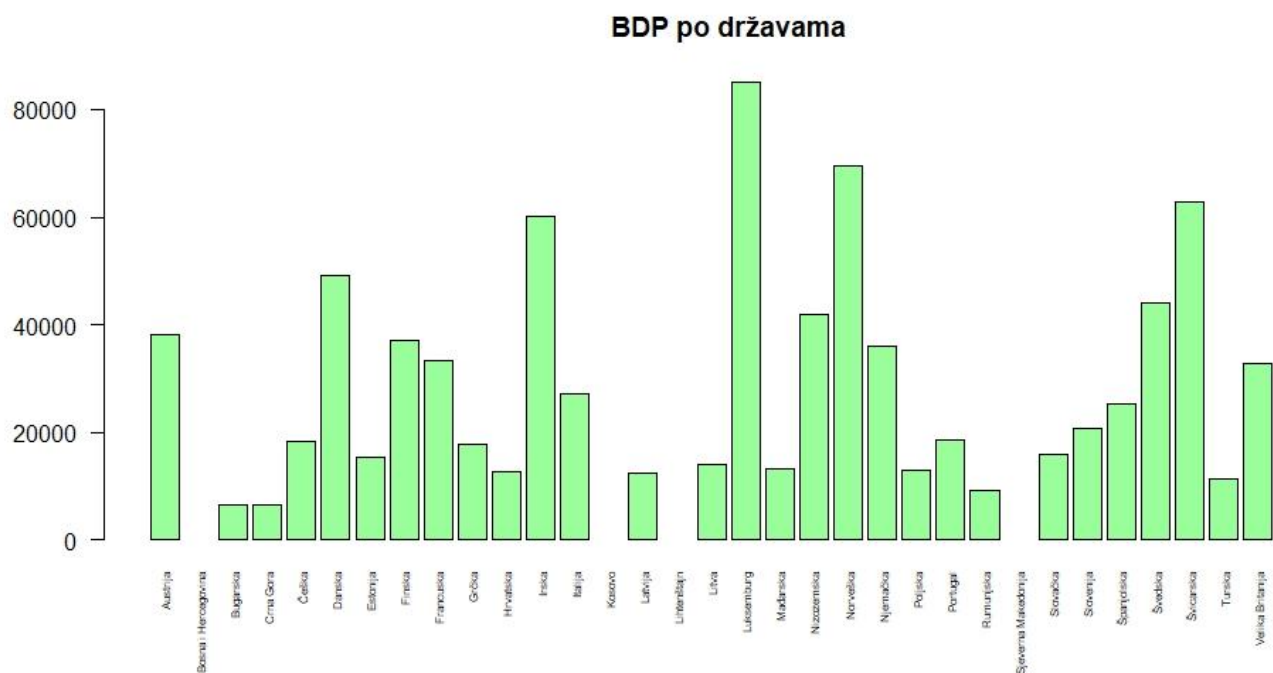
(slika ukazuje na eksponencijalnu razdiobu)



### 3. BDP

Za BDP smo uzeli podatke po glavi stanovnika.

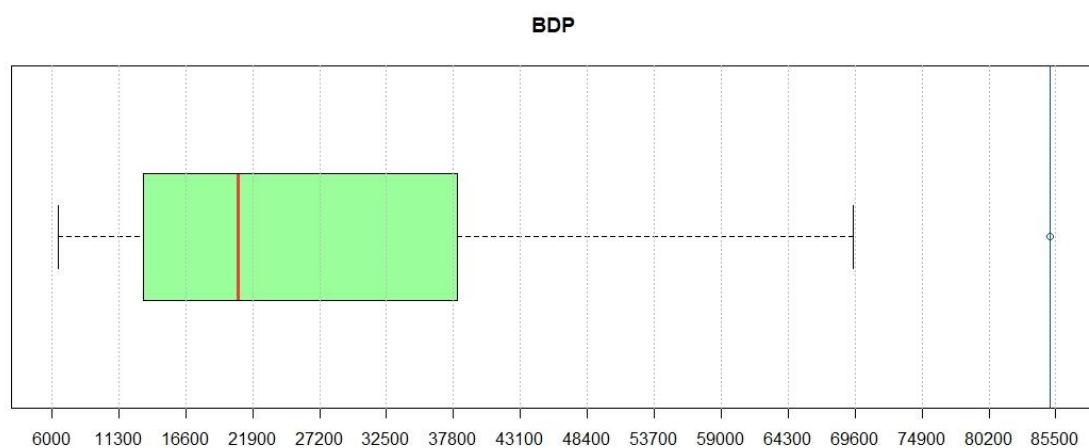
Prikazali smo podatke pomoću stupčastog dijagrama.



Slika 7. Stupčasti dijagram BDP-a

Izračunali smo karakterističnu petorku i podatke prikazali pomoću dijagrama pravokutnika.

Minimum	Donji kvartil	Medijan	Gornji kvartil	Maksimum
6480	13145	20720	40045	85030



Slika 8. Dijagram pravokutnika BDP-a

## Inferencijalna statistika

U ovom dijelu seminarskog rada ćemo, na temelju dobivenih podataka koji su prethodno prezentirani (vidi Sliku 9), pokušati odrediti njihovu razdiobu te ih opisati linearnim regresijskim modelom.

Ime države	Duljina pruge (u kilometrima)	Broj putnika (u tisućama)	Transport robe (u tisućama tona)	Relativni broj putnika (u tisućama/kilometru)	Relativni transport robe (u tis. tona/km)	BDP per capita (u eurima)
Sjeverna Makedonija	907	549	1839		2,027563	NA
Bosna i Hercegovina	965	657	13693		14,189637	NA
Crna Gora	250	986	1131	3,944000	4,524000	6480
Bugarska	5466	21329	14948	3,902122	2,734724	6630
Rumunjska	20079	69707	58808	3,471637	2,928831	9120
Turska	12370	164712	33285	13,315441	2,690784	11500
Latvija	2217	18623	41490	8,400090	18,714479	12530
Hrvatska	3945	19827	14449	5,025856	3,662611	12700
Poljska	37317	NA	233744		6,263740	13020
Mađarska	11345	NA	52270		4,607316	13270
Litva	2346	4991	55209	2,127451	23,533248	14050
Estonija	2144	8373	21341	3,905317	9,953825	15510
Slovačka	3629	80671	46413	22,229540	12,789474	15890
Grčka	2909	19599	NA	6,737367		17760
Češka	15418	193569	98804	12,554741	6,408354	18460
Portugal	3224	175511	9365	54,438896	2,904777	18670
Slovenija	2178	13259	21902	6,087695	10,056015	20720
Španjolska	21988	634954	26268	28,877297	1,194652	25200
Italija	24500	898472	94295	36,672327	3,848776	27230
Velika Britanija	31218	1836886	71053	58,840605	2,276027	32910
Francuska	27483	1265330	91957	46,040461	3,345959	33320
Njemačka	39379	2938023	364120	74,608878	9,246553	35980
Finska	8552	92801	38464	10,851380	4,497661	37150
Austrija	4877	314892	102575	64,566742	21,032397	38110
Nizozemska	3055	NA	42654		13,962029	41980
Švedska	15542	264603	68220	17,025029	4,389396	44180
Danska	2633	206601	8512	78,466008	3,232814	49270
Irska	2477	50060	346	20,209931	0,139685	60130
Švicarska	4055	635595	68658	156,743527	16,931689	62800
Norveška	4180	80402	34563	19,234928	8,268660	69490
Luksemburg	628	25016	4405	39,834395	7,014331	85030
$\Sigma$	317276	10035998	1734781	798,11	211,15	849090

Slika 9. Tablica podataka (sortirana po BDP-u), žutom bojom istaknuti su outliers

## Određivanje distribucije ulaznih podataka

Iz *Slike 3* i *Slike 6* koje prikazuju histograme broja putnika odnosno transporta dobara željeznicama, slutimo da su dobiveni podaci za putnike odnosno transport robe eksponencijalno distribuirani. To ćemo i testirati Kolmogorov-Smirnovljevim testom.

Eksponencijalna distribucija ovisi o parametru  $\lambda$  koji ćemo procijeniti metodom maksimalne vjerodostojnosti (MLE). Eksponencijalna funkcija vjerodostojnosti je oblika

$$f(t|\lambda) = \lambda e^{-(\lambda t)} \text{ za } t > 0, \text{ i } \lambda > 0$$

gdje je  $t$  broj putnika odnosno prevezene robe.

Procjena očekivane vrijednosti nepoznatog parametra  $\lambda$  pridružene funkcije vjerodostojnosti broja putnika odnosno transporta robe  $\mathbf{T}$  metodom MLE je:

$$L(\mathbf{T}|\lambda) = L(t_1, t_2, \dots, t_n) = \prod_{i=1}^n f(t_i|\lambda) = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda e^{-\lambda t_1} \cdot \lambda e^{-\lambda t_2} \dots \lambda e^{-\lambda t_n} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i} \quad |\ln$$

$$\ln L(\mathbf{T}|\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n t_i \quad \left| \frac{d}{d\lambda} \right.$$

$$\frac{d \ln L(\mathbf{T}|\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n t_i = 0$$

$$\frac{n}{\hat{\lambda}} = \sum_{i=1}^n t_i$$

Dakle, očekivana vrijednost koja maksimizira pridruženu funkciju vjerodostojnosti  $L(\mathbf{T})$  iznosi:

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n t_i}$$

gdje je  $n$  veličina uzorka na temelju kojeg procjenjujemo parametar, a  $t_i$  odgovarajući broj putnika odnosno količina prevezene robe željeznicom za pojedinu državu Europe.

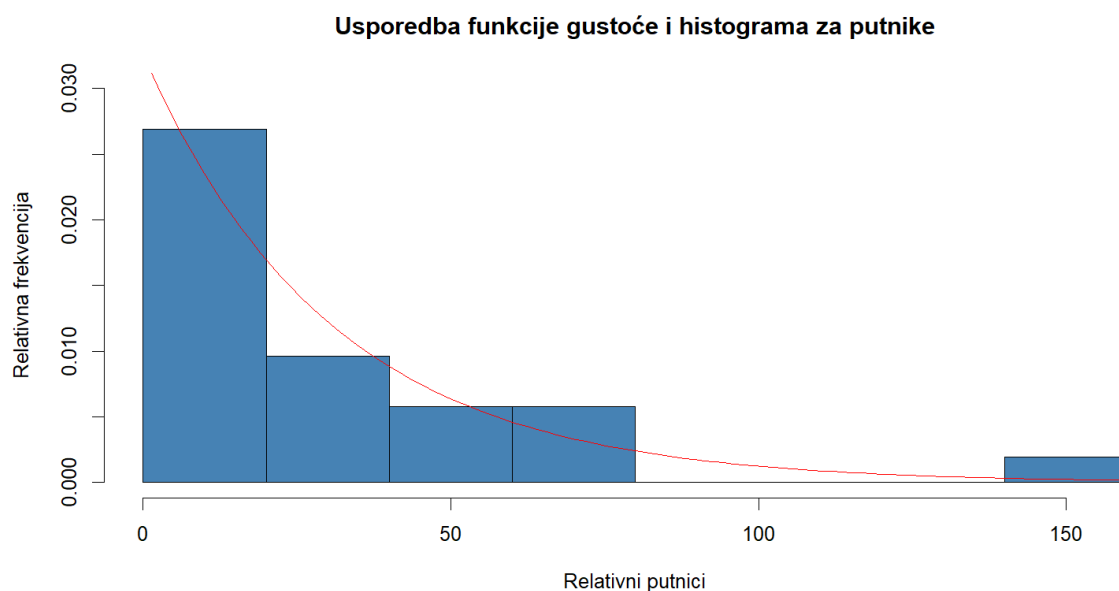
Izvor: [https://www.pfst.unist.hr/uploads/UR\\_Predavanje\\_11.pdf](https://www.pfst.unist.hr/uploads/UR_Predavanje_11.pdf)

## 1. Putnici

Metodom maksimalne vjerodostojnosti (MLE), na ranije opisan način, procjenjujemo parametar  $\lambda$  eksponencijalne distribucije za putnike na temelju dobivenog uzorka o broju putnika (u tisućama) po kilometru željezničke pruge u pojedinoj državi (stupac 5 priložene tablice sa *Slike 9*).

Veličina uzorka je  $n = 26$ , a zbroj podataka za putnike iz svih država jednak je 798.11 pa imamo:

$$\hat{\lambda} = \frac{26}{798.11} = 0.0325769$$



*Slika 10. Histogram i funkcija gustoće s procijenjenim parametrom za putnike*

### Kolmogorov-Smirnovljev test

Pripadnost podataka eksponencijalnoj distribuciji s procijenjenim parametrom testirat ćemo Kolmogorov-Smirnovljevom testom:

$$H_0: X \sim \text{Exp}(0.0325769)$$

$$H_1: \text{ne } H_0.$$

Testna statistika koju koristimo u ovom testu glasi:

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)| = \max_{1 \leq i \leq n} \{ \max\{ |\frac{i-1}{n} - F_0(x_{(i)})|, |\frac{i}{n} - F_0(x_{(i)})| \} \}$$

gdje je  $F_0(x)$  zadana sa sljedećom funkcijom distribucije  $F(x)$ :

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

pri čemu je naš  $\lambda = 0.0325769$ .

i	x(i)	i-1 / n	i/n	F0(x(i))	(i-1)/n - F0(x(i))	i/n - F0(x(i))	max
1	2.12745098	0	0.038461538	0.066958639	0.066958639	0.0284971	0.066958639
2	3.471637034	0.038461538	0.076923077	0.106934325	0.068472786	0.030011248	0.068472786
3	3.90212221	0.076923077	0.115384615	0.119371152	0.042448075	0.003986537	0.042448075
4	3.905317164	0.115384615	0.153846154	0.119462805	0.004078189	0.034383349	0.034383349
5	3.944	0.153846154	0.192307692	0.120571729	0.033274424	0.071735963	0.071735963
6	5.025855513	0.192307692	0.230769231	0.151026044	0.041281649	0.079743187	0.079743187
7	6.087695133	0.230769231	0.269230769	0.179891153	0.050878078	0.089339616	0.089339616
8	6.737366793	0.269230769	0.307692308	0.197065785	0.072164984	0.110626522	0.110626522
9	8.400090212	0.307692308	0.346153846	0.239400894	0.068291413	0.106752952	0.106752952
10	10.85137979	0.346153846	0.384615385	0.297777001	0.048376845	0.086838383	0.086838383
11	12.55474121	0.384615385	0.423076923	0.335682107	0.048933277	0.087394816	0.087394816
12	13.31544058	0.423076923	0.461538462	0.35194241	0.071134513	0.109596051	0.109596051
13	17.02502895	0.461538462	0.5	0.425711035	0.035827427	0.074288965	0.074288965
14	19.23492823	0.5	0.538461538	0.465601917	0.034398083	0.072859622	0.072859622
15	20.20993137	0.538461538	0.576923077	0.482309041	0.056152498	0.094614036	0.094614036
16	22.22953982	0.576923077	0.615384615	0.51527298	0.061650097	0.100111636	0.100111636
17	28.87729671	0.615384615	0.653846154	0.609658262	0.005726354	0.044187892	0.044187892
18	36.67232653	0.653846154	0.692307692	0.697196301	0.043350147	0.004888609	0.043350147
19	39.834395	0.692307692	0.730769231	0.726835449	0.034527757	0.003933781	0.034527757
20	46.04046138	0.730769231	0.769230769	0.776837548	0.046068318	0.007606779	0.046068318
21	54.43889578	0.769230769	0.807692308	0.830253683	0.061022914	0.022561375	0.061022914
22	58.84060478	0.807692308	0.846153846	0.852929648	0.04523734	0.006775802	0.04523734
23	64.56674185	0.846153846	0.884615385	0.877957263	0.031803417	0.006658122	0.031803417
24	74.60887783	0.884615385	0.923076923	0.912009428	0.027394043	0.011067495	0.027394043
25	78.46600836	0.923076923	0.961538462	0.922399308	0.000677615	0.039139153	0.039139153
26	156.743527	0.961538462	1	0.993941042	0.032402581	0.006058958	0.032402581
$\lambda$	<b>0.032576895</b>					d =	<b>0.110626522</b>

Slika 11. Tablica izračuna testne statistike za putnike

Iz tablice kritičnih vrijednosti za Kolmogorov-Smirnovljev test slijedi da je  $d_{0.05}(26) = 0.267$  pa je kritično područje  $C = [0.267, +\infty >$ . Kako je dobivena testna statistika  $d = 0.1106$  izvan tog intervala, zaključujemo da **ne odbacujemo  $H_0$** .

Provođenjem testa u R-u dobivamo:

Exact one-sample Kolmogorov-Smirnov test

data: Putnici

D = 0.11063, p-value = 0.8736

alternative hypothesis: two-sided

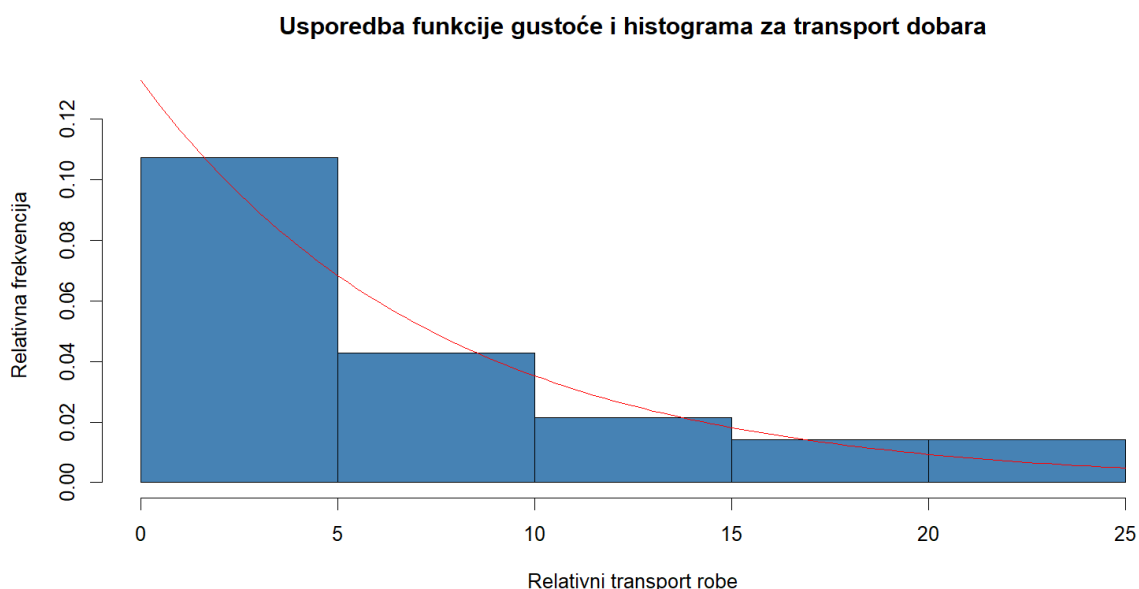
S obzirom na dobivenu p vrijednost (  $0.8736 > 0.05$  ), na razini značajnosti od  $\alpha = 0.05$  **ne odbacujemo  $H_0$**  odnosno nemamo razloga sumnjati da su podaci eksponencijalno distribuirani s parametrom  $\lambda = 0.0325769$ .

## 2. Transport dobara

Metodom maksimalne vjerodostojnosti (MLE), na ranije opisan način, procjenjujemo parametar  $\lambda$  eksponencijalne distribucije za transport dobara na temelju dobivenog uzorka o transportu robe (u tisućama tona) po kilometru željezničke pruge u pojedinoj državi (stupac 6 priložene tablice sa *Slike 9*).

Veličina uzorka je  $n = 28$ , a zbroj podataka za transport dobara iz svih država jednak je 211.15 pa imamo:

$$\hat{\lambda} = \frac{28}{211.15} = 0.1326071$$



Slika 12. Histogram i funkcija gustoće s procijenjenim parametrom za transport dobara

### Kolmogorov-Smirnovljev test

Pripadnost podataka eksponencijalnoj distribuciji s procijenjenim parametrom testirat ćemo Kolmogorov-Smirnovljevom testom:

$$H_0: X \sim \text{Exp}(0.1326071)$$

$$H_1: \text{ne } H_0.$$



Testna statistika koju koristimo u ovom testu glasi:

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)| = \max_{1 \leq i \leq n} \{ \max\{ |\frac{i-1}{n} - F_0(x_{(i)})|, |\frac{i}{n} - F_0(x_{(i)})| \} \}$$

gdje je  $F_0(x)$  zadana sa sljedećom funkcijom distribucije  $F(x)$ :

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

pri čemu je naš  $\lambda = 0.1326071$ .

i	x(i)	i-1 / n	i/n	F0(x(i))	(i-1)/n - F0(x(i))	i/n - F0(x(i))	max
1	0.139685103	0	0.035714286	0.018352501	0.018352501	0.017361785	0.018352501
2	1.194651628	0.035714286	0.071428571	0.146506407	0.110792121	0.075077835	0.110792121
3	2.276026651	0.071428571	0.107142857	0.26052396	0.189095389	0.153381103	0.189095389
4	2.690784155	0.107142857	0.142857143	0.300096267	0.19295341	0.157239124	0.19295341
5	2.734723747	0.142857143	0.178571429	0.304162487	0.161305344	0.125591059	0.161305344
6	2.904776675	0.178571429	0.214285714	0.319677983	0.141106555	0.105392269	0.141106555
7	2.928831117	0.214285714	0.25	0.321844582	0.107558868	0.071844582	0.107558868
8	3.23281428	0.25	0.285714286	0.348637264	0.098637264	0.062922978	0.098637264
9	3.34595932	0.285714286	0.321428571	0.358337128	0.072622842	0.036908557	0.072622842
10	3.6626109	0.321428571	0.357142857	0.384722529	0.063293958	0.027579672	0.063293958
11	3.84877551	0.357142857	0.392857143	0.399725586	0.042582729	0.006868443	0.042582729
12	4.389396474	0.392857143	0.428571429	0.441252509	0.048395366	0.012681081	0.048395366
13	4.497661366	0.428571429	0.464285714	0.449216865	0.020645437	0.015068849	0.020645437
14	4.524	0.464285714	0.5	0.4511372	0.013148515	0.0488628	0.0488628
15	4.607315998	0.5	0.535714286	0.457167744	0.042832256	0.078546542	0.078546542
16	6.263740386	0.535714286	0.571428571	0.56421476	0.028500475	0.007213811	0.028500475
17	6.408353872	0.571428571	0.607142857	0.57249199	0.001063419	0.034650867	0.034650867
18	7.01433121	0.607142857	0.642857143	0.605500779	0.001642078	0.037356364	0.037356364
19	8.268660287	0.642857143	0.678571429	0.66595144	0.023094297	0.012619988	0.023094297
20	9.246552731	0.678571429	0.714285714	0.706577789	0.028006361	0.007707925	0.028006361
21	9.953824627	0.714285714	0.75	0.732846206	0.018560492	0.017153794	0.018560492
22	10.05601469	0.75	0.785714286	0.736441977	0.013558023	0.049272309	0.049272309
23	12.78947368	0.785714286	0.821428571	0.816575681	0.030861395	0.004852891	0.030861395
24	13.96202946	0.821428571	0.857142857	0.842989145	0.021560574	0.014153712	0.021560574
25	16.93168927	0.857142857	0.892857143	0.894097162	0.036954305	0.001240019	0.036954305
26	18.71447903	0.892857143	0.928571429	0.916394064	0.023536921	0.012177364	0.023536921
27	21.03239697	1	1.038461538	0.938517898	0.061482102	0.099943641	0.099943641
28	23.53324808	1.038461538	1.076923077	0.955870885	0.082590654	0.121052192	0.121052192
$\lambda$	<b>0.132605388</b>					d =	<b>0.19295341</b>

Slika 13. Tablica izračuna testne statistike za transport dobara

Iz tablice kritičnih vrijednosti za Kolmogorov-Smirnovljev test slijedi da je  $d_{0.05}(28) = 0.257$  pa je kritično područje  $C = [0.257, +\infty >$ . Kako je dobivena testna statistika  $d = 0.1929$  izvan tog intervala, zaključujemo da **ne odbacujemo  $H_0$** .

Provođenjem testa u R-u dobivamo:

```
Exact one-sample Kolmogorov-Smirnov test  
data: Roba  
D = 0.19295, p-value = 0.218  
alternative hypothesis: two-sided
```

S obzirom na dobivenu p vrijednost (  $0.218 > 0.05$  ), na razini značajnosti od  $\alpha = 0.05$  **ne odbacujemo  $H_0$**  odnosno nemamo razloga sumnjati u eksponencijalnu distribuiranost podataka s parametrom  $\lambda = 0.1326071$ .

## Linearna regresija

Ispitujemo povezanost razvoja željezničkog prometa europskih država s njihovim BDP-om per capita. Želimo ju prikazati linearnim regresijskim modelom, provesti test normalnosti reziduala te izračunati pouzdane intervale.

### 1. Putnici

Razvoj željezničkog prometa prvo procjenjujemo na temelju relativnog broja putnika koji putuju vlakom po kilometru željezničke pruge u pojedinoj državi Europe. Provodimo linearnu regresiju kojom ćemo prikazati ovisnost relativnog broja putnika o BDP-u promatrane zemlje.

Pretpostavljamo da su sljedeće varijable u srednjem linearno povezane:

X -> BDP per capita (u eurima)

Y -> relativni broj putnika (u tisućama/kilometru)

Za izračun će nam biti potrebno:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{780820}{26} = 30031.54 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{798.11}{26} = 30.70$$

$$S_{XX} = \sum_{i=1}^n (x_i - \bar{x})^2 = 10989649538.46$$

$$S_{YY} = \sum_{i=1}^n (y_i - \bar{y})^2 = 30528.54$$

$$S_{XY} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 9520053.44$$

Računamo Pearsonov koeficijent korelacije:

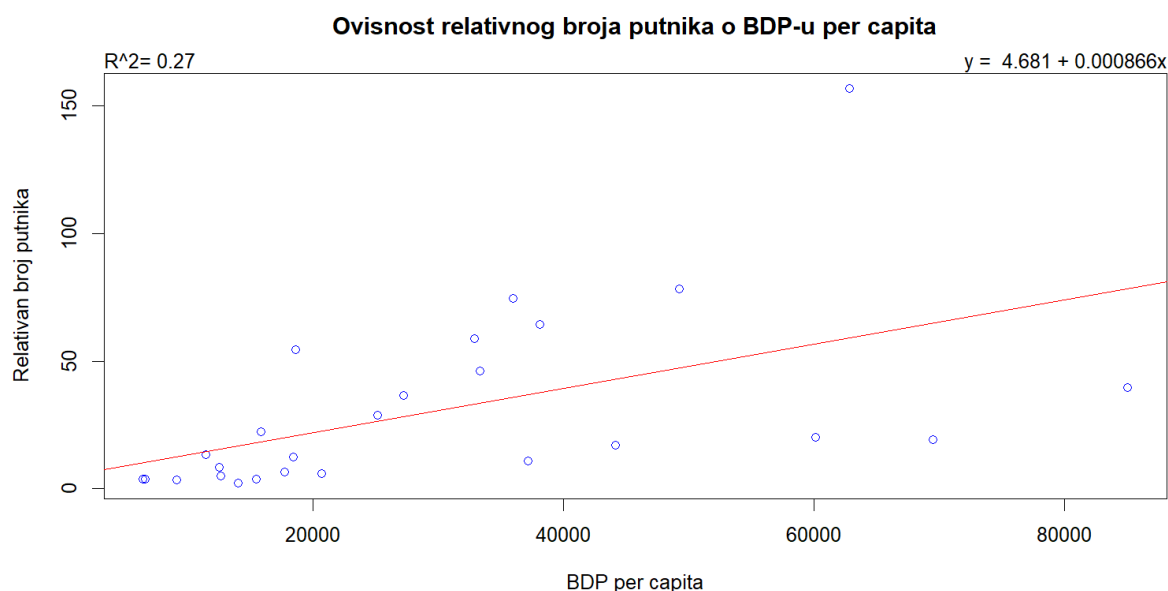
$$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{9520053.44}{\sqrt{10989649538.46 \cdot 30528.54}} = 0,519749953$$

Dobiveni koeficijent korelacije upućuje na pozitivnu koreliranost relativnog broja putnika i BDP-a zemlje. Ipak, nije dovoljno velik da bismo sa znatnom sigurnošću mogli izvršiti predikciju o pravilnosti kretanja broja putnika s obzirom na promatrani iznos BDP-a neke europske države.

Aproksimiramo pravac  $y = \hat{\alpha} + \hat{\beta}x$  metodom najmanjih kvadrata:

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{99520053.44}{10989649538.46} = 0.000866275$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 30.7 - 0.000866275 \cdot 30031.54 = 4.681$$



Slika 14. Grafikon raspršenosti BDP-a i relativnog broja putnika

Testiramo hipoteze:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Provođenjem linearne regresije u R-u dobivamo:

Call:

`lm(formula = Putnici ~ BDPputnici)`

Coefficients:

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	<b>4.681e+00</b>	1.058e+01	0.443	0.6621
BDPputnici	<b>8.663e-04</b>	2.907e-04	2.980	<b>0.0065**</b>

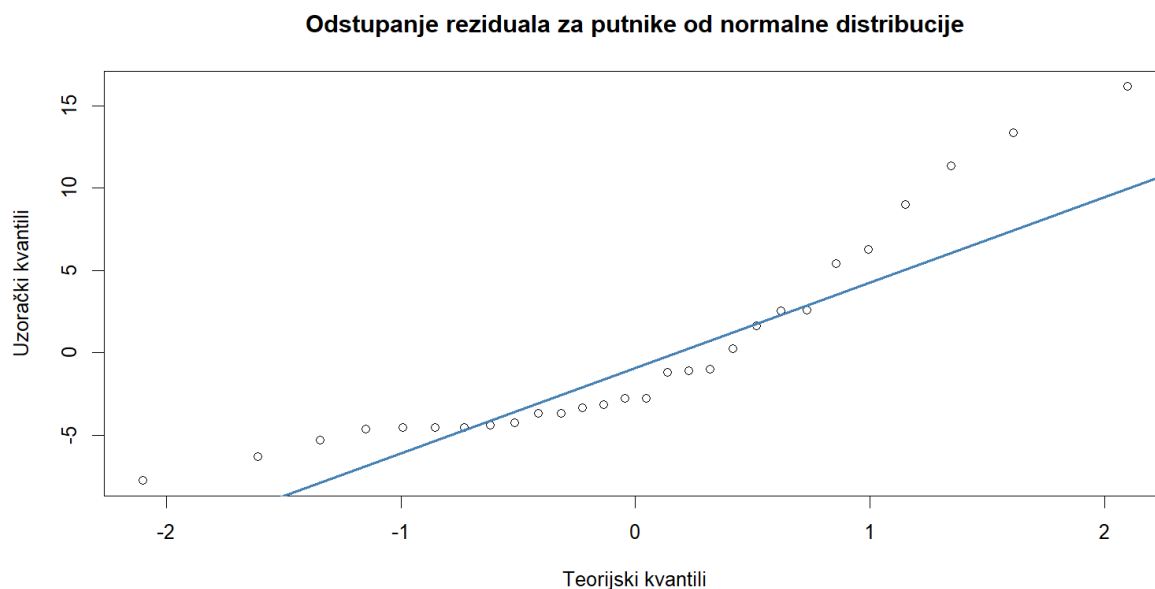
Residual standard error: 30.47 on 24 degrees of freedom

Multiple **R-squared: 0.2701**, Adjusted R-squared: 0.2397

F-statistic: 8.883 on 1 and 24 DF, **p-value: 0.006501**

S obzirom na dobivenu p vrijednost ( $0.006501 < 0.05$ ), na razini značajnosti od  $\alpha = 0.05$  **odbacujemo  $H_0$**  odnosno postoji ovisnost relativnog broja putnika o BDP-u.

### Lillieforsova inačica Kolmogorov-Smirnovljeva testa za reziduala



Slika 15. Grafikon usporedbe normalnosti reziduala linearne regresije za putnike

Provjeravamo jesu li reziduali normalno distribuirani. To ćemo testirati Lillieforsovom inačicom Kolmogorov-Smirnovljevog testa jer ne znamo očekivanje ni varijancu populacije iz koje uzorak dolazi:

$H_0$ : reziduali su normalno distribuirani

$H_1$ : ne  $H_0$ .

Provođenjem testa u R-u dobivamo:

Lilliefors (Kolmogorov-Smirnov) normality test

data: model\$residuals

D = 0.16115, p-value = 0.08066

S obzirom na dobivenu p vrijednost (  $0.08066 > 0.05$  ), na razini značajnosti od  $\alpha = 0.05$  **ne odbacujemo  $H_0$**  odnosno nemamo razloga sumnjati u normalnu distribuiranost reziduala linearne regresije za putnike.

### Intervalna procjena koeficijenta smjera pravca

Pouzdani intervali za koeficijente smjera pravca imaju  $t$  distribuciju:

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

gdje je

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}}, \quad SSE = S_{yy} - \hat{\beta}^2 S_{xx}$$

Izračunom dobivamo sljedeće 95% pouzdane intervale:

$$\alpha \in [-17.15, 26.5135]$$

$$\beta \in [0.0003, 0.00147]$$

Kako interval za  $\beta$  ne uključuje 0, možemo na razini značajnosti  $\alpha = 0.05$  dodatno potvrditi odbacivanje hipoteze  $H_0: \beta = 0$ .

Dakle, iz sveg navedenog zaključujemo da postoji linearna ovisnost relativnog broja putnika (u tisućama po kilometru željezničke pruge) o BDP-u pripadne države. Odnosno što je veći BDP promatrane zemlje, to će biti veći broj putnika koji putuju vlakom pa možemo reći da je razvoj željezničkog prometa europskih država povezan s BDP-om per capita.

## 2. Transport dobara

Razvoj željezničkog prometa ovdje procjenjujemo temeljem količine transportiranih dobara koja su prevezena putem željeznice po kilometru pruge pojedine europske države. Provest ćemo linearnu regresiju kojom želimo ispitati ovisnost relativnog broja transportirane robe o BDP-u promatrane zemlje.

Pretpostavljamo da su sljedeće varijable u srednjem linearno povezane:

X -> BDP per capita (u eurima)

Y -> relativni broj transportirane robe (u tisućama tona/kilometru)

Za izračun će nam biti potrebno:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{831330}{28} = 29690.36 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{211,15}{28} = 7.54$$

$$S_{XX} = \sum_{i=1}^n (x_i - \bar{x})^2 = 11548906896$$

$$S_{YY} = \sum_{i=1}^n (y_i - \bar{y})^2 = 1052.26$$

$$S_{XY} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 133421.66$$

Računamo Pearsonov koeficijent korelacije:

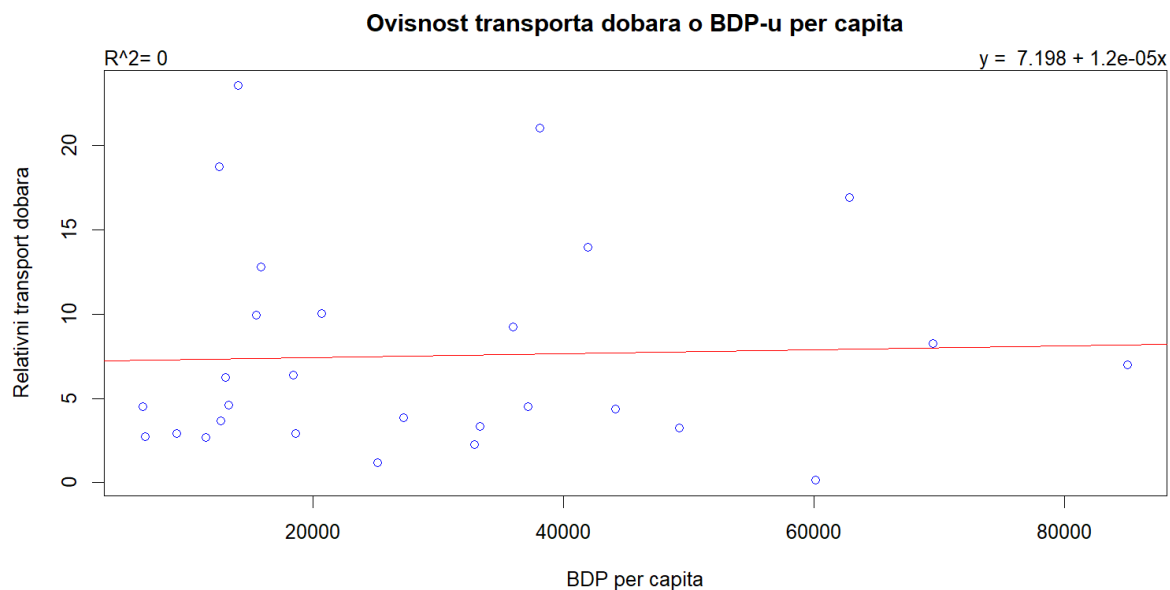
$$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{133421.66}{\sqrt{11548906896 \cdot 1052.26}} = 0.03827$$

Ovakav koeficijent korelacije ne smatramo rezultatom koji upućuje na povezanost između BDP-a neke države Europe i transporta dobara putem željeznice.

Aproksimiramo pravac  $y = \hat{\alpha} + \hat{\beta}x$  metodom najmanjih kvadrata:

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{133421.6569}{11548906896} = 0.000011553$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 7.5412 - 0.000011553 \cdot 29690,36 = 7.1982$$



Slika 16. Grafikon raspršenosti BDP-a i relativnog broja transportiranih dobara

Testiramo hipoteze:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Provođenjem linearne regresije u R-u dobivamo:

Call:

`lm(formula = Roba ~ BDProba)`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	<b>7.198e+00</b>	2.128e+00	3.383	0.00228 **
BDProba	<b>1.155e-05</b>	5.915e-05	0.195	<b>0.84668</b>

Residual standard error: 6.357 on 26 degrees of freedom

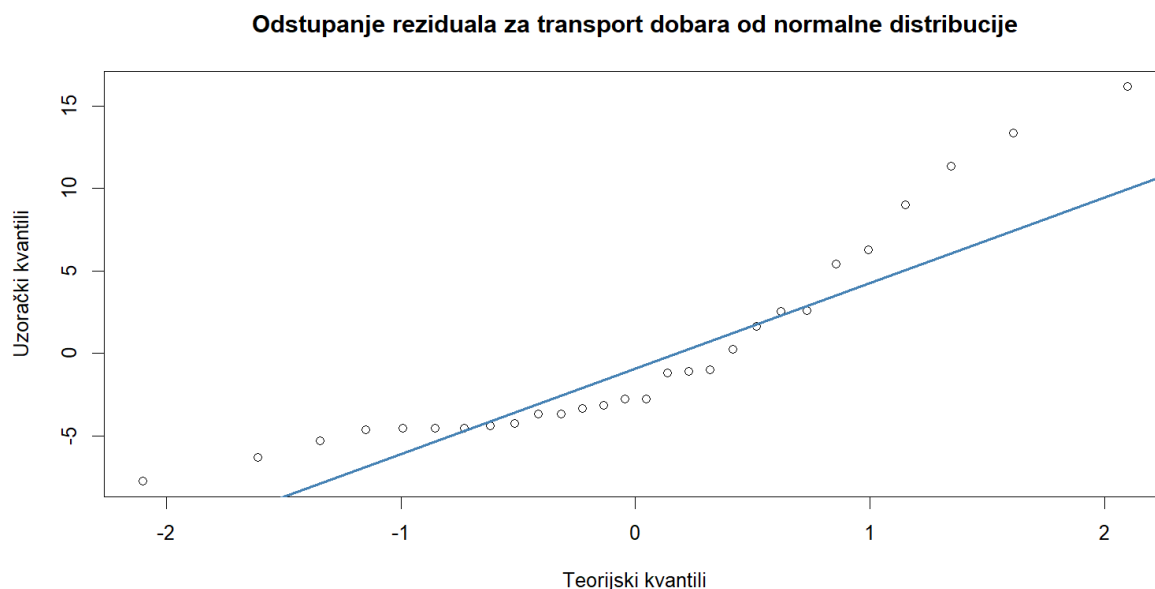
Multiple **R-squared: 0.001465**, Adjusted R-squared: -0.03694

F-statistic: 0.03814 on 1 and 26 DF, **p-value: 0.8467**

S obzirom na dobivenu p vrijednost (  $0.8467 > 0.05$  ), na razini značajnosti od  $\alpha = 0.05$  ne odbacujemo  $H_0$  odnosno ne možemo odbaciti pretpostavku da je koeficijent smjera pravca linearne regresije vezane uz transport dobara jednak 0.



### Lillieforsova inačica Kolmogorov-Smirnovljeva testa za reziduala



Slika 17. Grafikon usporedbe normalnosti reziduala linearne regresije za robu

Kako ne znamo očekivanje ni varijancu populacije iz koje uzorak dolazi, normalnost reziduala testirat ćemo Lillieforsovom inačicom Kolmogorov-Smirnovjevog testa:

$H_0$ : reziduali su normalno distribuirani

$H_1$ : ne  $H_0$ .

Provođenjem testa u R-u dobivamo:

Lilliefors (Kolmogorov-Smirnov) normality test

data: model\$residuals

D = 0.20673, p-value = 0.00346

S obzirom na dobivenu p vrijednost (  $0.00346 < 0.05$  ), na razini značajnosti od  $\alpha = 0.05$  **odbacujemo  $H_0$**  odnosno s 95%-tnom sigurnošću možemo tvrditi da reziduali nisu normalno distribuirani što znači da ovaj model nije relevantan za dobiveni uzorak.

### Intervalna procjena koeficijenta smjera pravca

Pouzdani intervali za koeficijente smjera pravca imaju  $t$  distribuciju:

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

gdje je

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}}, \quad SSE = S_{yy} - \hat{\beta}^2 S_{xx}$$

Izračunom dobivamo sljedeće 95% pouzdane intervale:

$$\alpha \in [2.8243, 11.572]$$

$$\beta \in [-0.0001, 0.0013]$$

Kako je u intervalu za  $\beta$  uključena i 0, možemo na razini značajnosti  $\alpha = 0.05$  dodatno potvrditi ne odbacivanje hipoteze  $H_0: \beta = 0$ .

Stoga, možemo zaključiti da linearna ovisnost relativnog broja transportirane robe (u tisućama tona po kilometru željezničke pruge) o BDP-u pripadne države ne postoji. Za bolju statističku relevantnost pokušali smo izbaciti outliere. Međutim, ni to nije urodilo plodom.

Drugi načini za pokazivanje ovisnosti je transformiranje podataka (npr. logaritamska transformacija, Box-Coxova transformacija, korijen, inverzna transformacija, itd.) ili promatranje drugih modela koji nisu predmet istraživanja ovog seminarskog rada.

## Zaključak

Iz razdiobi dobivenih podataka o relativnom broju putnika koji koriste željeznički prijevoz odnosno transportu dobara željeznicom, vidi se da korištenje željezničkog prometa brzo opada kad gledamo „veću iskoristivost po kilometru pruge“ bilo u putničkom ili robnom transportu.

Ipak, putnici, koji dolaze iz bogatijih zemalja Europe, koriste vlakove češće, tj. s obzirom na duljine pruga u njihovim državama, ne odbacujemo pretpostavku da oni više iskorištavaju postojeću željezničku infrastrukturu. To ne možemo reći za transport robe, koji je skoro nimalo koreliran i za koji nismo odbacili nulhipotezu o koeficijentu pravca regresije,  $\beta = 0$ .

Razvoj željezničkog prometa je koreliran s BDP-om per capita kada se on promatra pomoću relativnog broja putnika te je dobivena linearna regresija statistički značajnija od one dobivene za povezanost s transportom dobara obzirom na normalnost reziduala. Međutim, nismo dobili statistički značajnu povezanost koja bi se mogla iskoristiti za predviđanje frekvencije putničkog prijevoza odnosno transporta dobara putem željezničke infrastrukture.