

Covid-19: Potek širjenja do viška prvega vala v Evropski uniji

Matej Kalc

10. september 2020

1 Uvod

1.1 Motivacija

“Koronavirus je hujši kot vojna, kjer je sovražnik še vedno človek, s katerim se še vedno lahko ukvarjamo, medtem ko je kakršenkoli dogovor s smrtonosnim virusom, ki ogroža naše preživetje, nemogoč. (...)”. [1]

Tako je izjavil G. Zuccarini. Lahko bi izjavili, da je Koronavirus tretja svetovna vojna, kjer se neviden sovražnik skriva med ljudmi. Ogroža ljudem življenje, nekaterim pa ga tudi odvzame. Ljudje lahko premagamo nevidnega sovražnika, le če primerno in pravočasno ukrepamo s pravim orožjem, kot so samozavest in ukrep človeka. V taki bitki tudi študiji in analize podatkov so dobro orožje proti virusu, saj nam povejo nekaj novega o našem sovražniku. Mogoče eden izmed teh nam bo dal možnost odkritja cepiva proti virusu, toda dokler tega ne najdemo ostaja edina možnost uporaba mask, razkužil in distanca.

1.2 Cilji

Trdimo lahko, da so vse države v Evropski uniji [3] preživele prvi val Koronavirusa pred 19. julijem 2020. V seminarski nalogi bom analiziral kako se je virus širil po državah Evropske unije in predvsem interval od začetka širjenja do vrhunca prvega vala v vsaki državi, ker je ta interval najzanimivejši, saj se države prvič soočajo s takim virusom.

Cilja študija sta:

- Ugotovitev ali se je koronavirus sprožil podobno v vseh državah Evropske unije in
- Povezanost raznih faktorjev s širjenjem in fatalnostjo virusa.

Testiral bom korelacijo med spremenljivkami in izračunal 95% intervale zaupanja, saj podatki niso realni, ker v teh niso upoštevani asimptomatiki. Izbrana je stopnja značilnosti $\alpha = 0.05$. Vsak test bo vrnil p vrednost. Če bo ta manjša od 0.05, bo rezultat statistično pomemben. Podatke bom analiziral v statističnem softverju R.

1.3 Raziskave o virusu

Veliko je spletnih strani, ki analizirajo in grafično prikazujejo podatke Covid-a-19. Omenil bom tisto, ki me je motiviralo za izdelavo seminarske naloge.

Inštitut za zdravstvene meritve in vrednotenje IHME nudi spletno stran o Koronavirusu [5], kjer so grafično prikazani podatki o okuženih, mrtvih, analizah, socialni distanci ipd, ampak

najzanimivejše so projekcije v času, ki stran nudi. IHME-ove projekcije COVID-a-19 so bile razvite kot odziv na zahteve medicinske univerze v Washingtonu in drugih ameriških bolnišničnih sistemov. Napovedi kažejo povpraševanje po storitvah v bolnišnicah, dnevne in kumulativne smrti zaradi COVID-a-19, stopnje okužbe in analizah ter vpliv socialnega distanciranja, ki ga zahteva država.

1.4 Poglavlja

1. Uvod: v uvodu zastavim cilje seminarske naloge
2. Opis virusa in njegovo širjenje: opisan je virus bolj podrobno, predvsem kako je nastal in kako se širi
3. Podatki: opisani so podatki za analizo tabeli
4. Izračuni in rezultati: opravi bom analizo spremenljivk in korelacijsko analizo
5. Zaključki: odgovor na zastavljene cilje v uvodu
6. Literatura

2 Opis virusa in njegovo širjenje

COVID-19 je nalezljiva bolezen, ki jo povzroča virus SARS-CoV-2[9], virus RNA s pozitivnim občutkom razreda IV, ki je pri ljudeh nalezljiv. Kot je opisal ameriški Nacionalni zdravstveni inštitut (WHO), je naslednik virusa SARS-CoV, ki je povzročil izbruh SARS-a iz leta 2002. Taksonomsko je SARS-CoV-2 sev hudega koronavirusa, povezanega z akutnim dihalnim sindromom (SARS-CoV). Menijo, da ima zoonotski izvor in tesno genetsko podobnost s koronavirusi netopirjev[10]. Znanstveno je dokazano, da SARS-COV-2 hitro mutira zaradi tega je težko najti ustrezna zdravila. Ta vrsta virusov se tako imenuje, ker se njihovi virioni pojavljajo pod elektronskim mikroskopom kot majhne kroglice, na katerih je veliko majhnih konic, ki spominjajo na krono. Letošnji koronavirus se bistveno razlikuje od prejšnjih po inkubacijski dobi, ki lahko traja od dva do 14 dni, v tem času ne povzroča nobenih simptomov[8]. Dihalni virus se širi preko kapljice slin in sluzi okuženih ljudi. Prvi okužen Covid-a-19 je bil zaznan na Kitajskem novembra 2019. Najprej se je dihalni virus širil na Kitajskem, v Hubeju in Wuhanu. Na začetku leta 2020 se je začelo širjenje virusa po celem svetu. 11. marca 2020 je Svetovna zdravstvena organizacija WHO proglasila pandemijo. Takrat je bilo 125.000 zaznanih okuženih in 4.600 mrtvih[6]. Iz statističnih podatkov je razvidno, da do vključno 19. julija 2020 je bilo okuženih več kot 14.2 milijonov ljudi v 188 državah, od katerih 600 tisoč je mrtvih in 8.02 milijonov je ozdravelih[6]. Trdimo lahko, da je ta virus leta 2020 močno vplival na države po celem svetu.

Število zaznanih okuženih ni realno število okuženih, saj niso šteti asimptomatiki. Število teh ni znano, toda z uporabo statistike lahko si boljše predstavljamo kakšno zgleda to število. Število zaznanih mrtvih ni realno število mrtvih. Vsaka država različno beleži smrti Covid-a-19, predvsem ko pacient ima več bolezni. Drugi faktor ki vpliva, je število opravljenih testov. Višje bo število opravljenih testov višje bo zaznano število okuženih.

3 Podatki

Podatki so bili izbrani iz spleta[2][3][4][5][6]. Podatke, ki bom rabil za statistični študij, so prikazani v spodnji tabeli.

DR	MED	DPO	DPS	DVO	DVS	OV	MV	PREB	OTO
Austria	44.0	2020-02-25	2020-03-12	2020-03-27	2020-04-23	7029	52	9025715	38809
Belgium	41.4	2020-02-04	2020-03-10	2020-04-11	2020-04-12	32778	4273	11602522	103714
Bulgaria	42.7	2020-03-08	2020-03-12	2020-06-12	2020-06-06	3086	168	6943915	91083
Croatia	43.0	2020-02-25	2020-03-25	2020-04-02	2020-04-20	963	6	4101782	8110
Cyprus	36.8	2020-03-09	2020-03-25	2020-04-02	2020-03-25	320	9	1190007	8468
Czechia	42.1	2020-03-01	2020-03-23	2020-03-27	2020-04-15	2062	9	10715154	36089
Denmark	42.2	2020-02-27	2020-03-15	2020-04-08	2020-04-05	5071	203	5793679	62063
Estonia	42.7	2020-02-27	2020-03-26	2020-03-27	2020-04-03	538	1	1328655	9010
Finland	42.5	2020-01-29	2020-03-21	2020-04-05	2020-04-22	1882	25	5542713	34486
France	41.4	2020-01-24	2020-02-15	2020-04-01	2020-04-04	51477	3514	65283211	233494
Germany	47.1	2020-01-28	2020-03-09	2020-03-20	2020-04-16	18323	45	83951077	595836
Greece	44.5	2020-02-26	2020-03-12	2020-04-22	2020-04-05	2401	121	10420046	58847
Hungary	42.3	2020-03-04	2020-03-15	2020-04-10	2020-04-24	1190	77	9659639	29041
Ireland	36.8	2020-03-01	2020-03-11	2020-04-10	2020-04-26	7393	263	4953657	68922
Italy	45.5	2020-01-29	2020-02-22	2020-03-21	2020-03-28	53578	4827	60465251	239558
Latvia	43.6	2020-03-02	2020-04-04	2020-03-24	2020-04-22	180	0	1883138	8281
Lithuania	43.7	2020-02-28	2020-03-20	2020-04-04	2020-04-12	771	9	2714541	21467
Luxembourg	39.3	2020-03-01	2020-03-13	2020-03-24	2020-04-12	875	8	628614	11189
Malta	41.8	2020-03-07	2020-04-09	2020-04-08	2020-06-02	293	0	441612	14119
Netherlands	42.6	2020-02-27	2020-03-06	2020-03-24	2020-04-08	4749	213	17138553	45825
Poland	40.7	2020-03-05	2020-03-12	2020-06-05	2020-04-25	25048	1117	37850596	1006819
Portugal	42.2	2020-03-02	2020-03-17	2020-04-11	2020-04-04	15472	435	10193282	179542
Romania	41.1	2020-02-26	2020-03-23	2020-04-12	2020-05-01	5990	282	19210031	64385
Slovakia	40.5	2020-03-06	2020-04-07	2020-04-17	2020-04-16	977	8	5461415	42768
Slovenia	44.5	2020-03-04	2020-03-17	2020-03-13	2020-04-06	141	0	2079553	4228
Spain	42.7	2020-01-31	2020-03-04	2020-04-01	2020-06-20	94417	8189	46785134	466271
Sweden	41.2	2020-01-31	2020-03-15	2020-06-23	2020-04-22	58932	5122	10108080	467798

Legenda:

- DR - Ime države
- MED - Mediana starosti populacije
- DPO - Datum prvega zaznanega okuženca
- DPS - Datum prve zaznane smrti
- DVO - Datum vrhunca okuženih v prvem valu
- DVS - Datum vrhunca smrti v prvem valu
- OV - Število okuženih od prve zaznane okužbe do vrhunca okuženih v prvem valu
- MV - Število mrtvih od prve zaznane smrti do vrhunca okuženih v prvem valu
- PREB - Število prebivalcev
- OTO - Število opravljenih testov do vrhunca okužb v prvem valu

V sledečih analizah bom predvsem računal z deleži, ki vzamejo v zakup število prebivalcev ali število okuženih, omogočajo pa tudi uporabo in primerjavo podatkov iz različnih držav. Rabil bom sledeče spremenljivke:

1. Delež okuženih do vrhunca prvega vala okuženih

$$D = \frac{100 * \text{št. okuženih}}{\text{št. prebivalcev}}$$

kjer število okuženih je stolpec OV v bazi in število prebivalcev je stolpec PREB v bazi.

2. Fatalnost do vrhunca prvega vala okuženih

$$F = \frac{100 * \text{št. mrtvih}}{\text{št. okuženih}}$$

kjer število mrtvih je stolpec MV v bazi in število okuženih je stolpec OV v bazi.

3. Delež testov do vrhunca prvega vala okuženih

$$T = \frac{100 * \text{št. opravljenih testov}}{\text{št. prebivalcev}}$$

kjer število opravljenih je stolpec OTO v bazi in število prebivalcev je stolpec PREB v bazi.

4 Izračuni in rezultati

4.1 Analiza spremenljivk

Spodnje spremenljivke veljajo le za države Evropske unije. Spremenljivke so:

1. Mediana starosti (stolpec MED v bazi)
2. Število dni do vrhunca prvega vala okuženih (razlika v dnevih stolpcev DVO in DPO v bazi)
3. Število dni do vrhunca prvega vala mrtvih (razlika v dnevih stolpcev DVS in DPS v bazi)
4. Delež okuženih do vrhunca prvega vala okuženih (spremenljivka D definirana v 3. poglavju)
5. Delež mrtvih do vrhunca prvega vala okuženih (spremenljivka F definirana v 3. poglavju)
6. Delež testov do vrhunca prvega vala okuženih (spremenljivka T definirana v 3. poglavju)

Izračunal bom opisno statistiko spremenljivk, ki vsebuje minimalno in maksimalno vrednost, ter mero centra porazdelitve in mero variabilnosti porazdelitve. Simetrične porazdelitve se dobro opišejo z vzorčnim povprečjem in standardnim odklonom, medtem ko se asimetrične porazdelitve boljše opišejo z mediano in razponom. Tako da, v prvem koraku analize bom analiziral porazdelitev podatkov (normalnost, oziroma simetričnost porazdelitve). Normalnost

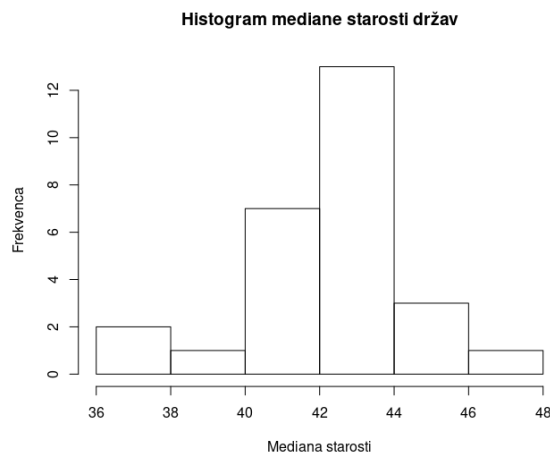
porazdelitve bom testiral s pomočjo Shapiro-Wilk testa[11], ki je eden izmed najmočnejših statističnih testov normalnosti. Ničelna domneva Shapiro-Wilkovega testa je H_0 : porazdelitev je normalna proti alternativni domnevi H_1 : porazdelitev ni normalna. Simetričnost porazdelitve bom testiral s pomočjo Miao, Gel in Gastwirthovega testa simetrije. Ničelna domneva je H_0 : porazdelitev je simetrična proti alternativni domnevi H_1 : porazdelitev ni simetrična. Testi simetrije so narejeni samo za spremenljivke, ki niso približno normalno porazdeljene. Dalje, za deleže okuženih, mrtvih in opravljenih testov vsake države bom izračunal 95% interval zaupanja za delež binomske porazdelitve

$$I_p = [\hat{p} - \Delta, \hat{p} + \Delta],$$

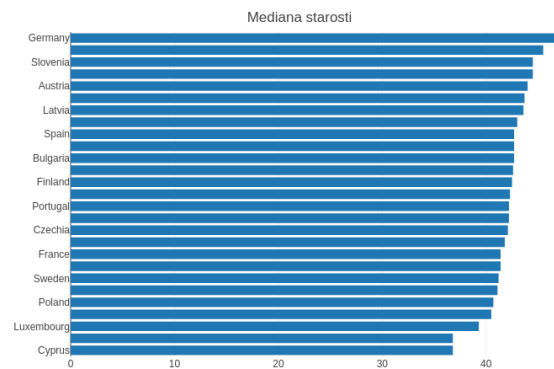
kjer je \hat{p} vzorčni delež, Δ odstopanje, ki je enako

$$\Delta = c \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}},$$

za velikost vzorca n in $c = \Phi^{-1}(0.975) = 1.96$ kvantil normalne porazdelitve. Izbrane spremenljivke prikazemo grafično s histogramom in barplotom[12]. Iz slike 1. in slike 2. opazamo, da spremenljivka mediana starosti bi lahko bila normalna ali simetrična. Nemčija ima najvišjo mediano starosti in Ciper ima najmanjšo mediano starosti.

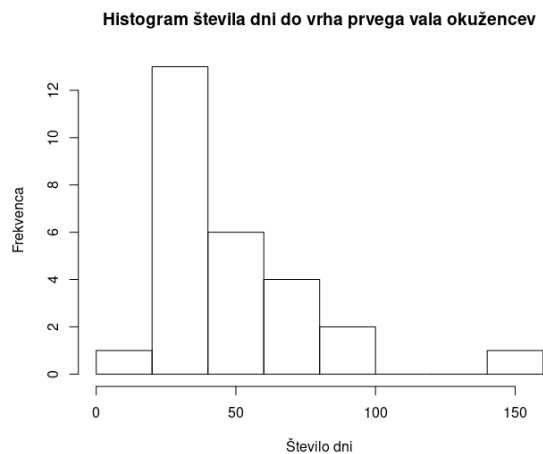


Slika 1. Histogram mediane starosti



Slika 2. Barplot mediane starosti

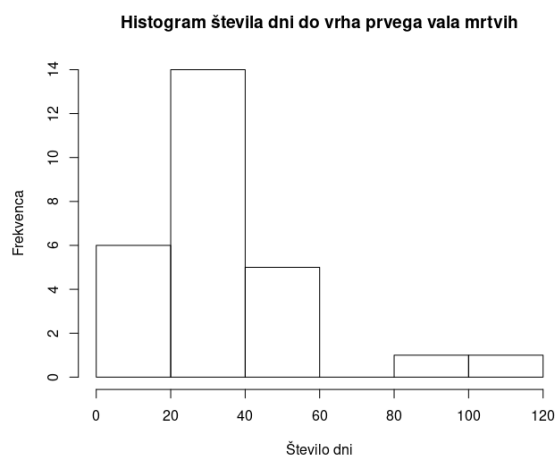
Prvi vtis iz slik je, da število dni do vrha prvega vala okuženih ni normalna ali simetrična. Naslednji računi bodo potrdili to intuicijo. Iz barplota vidimo, da obstaja velika raznolikost med državami. Sumimo, da ima spremenljivka velik razpon. Švedska je najzanimivejši primer, saj zdaleč zavzame najvišjo vrednost v grafu.



Slika 4. Barplot števila dni do vrhunca okužencev

Slika 3. Histogram števila dni do vrhunca okužencev

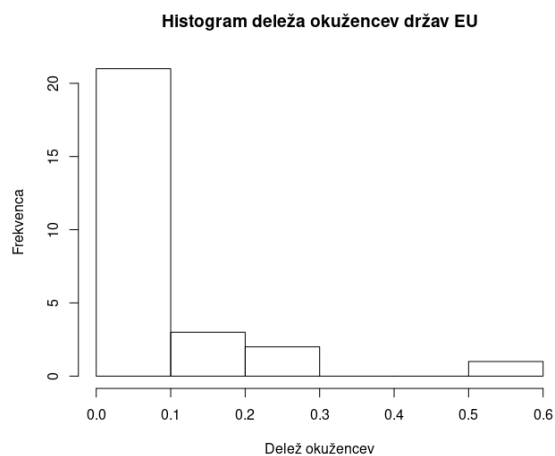
Iz histograma (slika 5.) lahko bi smatrali, da spremenljivka število dni do vrha prvega vala mrtvih je normalno porazdeljena ali vsaj simetrična. Iz barplota vidimo, da obstaja velika raznolikost med državami. Poseben primer je Ciper. Za to državo velja da je imela maksimum mrtvih kar prvi dan. V naslednjih dneh je bilo zaznanih vedno manjše število. Dogodek zglada zelo nenavaden. Zaradi raznolikosti sumimo, da ima spremenljivka velik razpon.



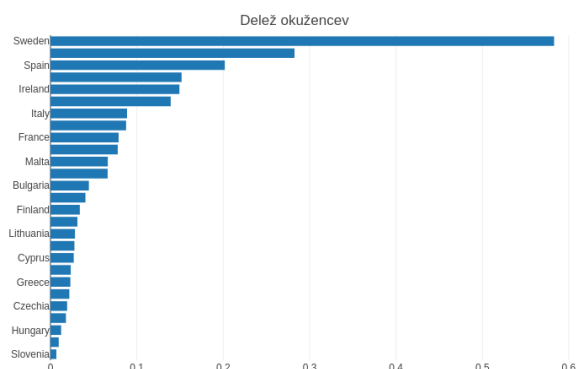
Slika 6. Barplot števila dni do vrhunca mrtvih

Slika 5. Histogram števila dni do vrhunca mrtvih

Iz slik se sluti, da spremenljivka delež okuženih ni normalna ali simetrična. Iz barplota vidimo, da obstaja velika raznolikost med državami, kljub temu da primerjamo deleže. Res je, da so ti deleži okuženih zelo majhni saj predstavljajo le izbrani časovni interval. Kot v zgornjih barplotih, tudi barplot deleža okuženih ima velik razpon podatkov. Opažamo da ima Švedska najvišji delež okuženih 0.5%.

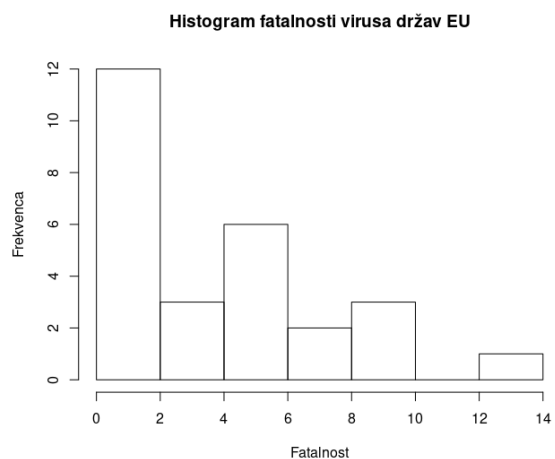


Slika 7. Histogram deleža okuženih

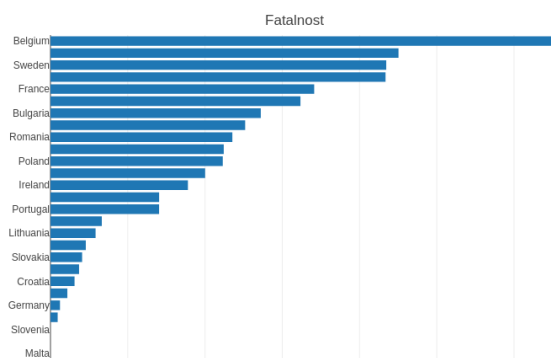


Slika 8. Barplot deleža okuženih

Kot za delež okuženih velja zelo podobno za fatalnost. Začetni vtis je, da fatalnost ni normalno porazdeljena ali simetrična. Najzanimivejši podatek je Belgija, ki ima skoraj 13% fatalnosti od začetka širjenja do vrhunca prvega vala okuženih. Med državami obstaja velika raznolikost, kar se tiče fatalnosti, saj nekatere države imajo 0% fatalnost za izbrani interval.

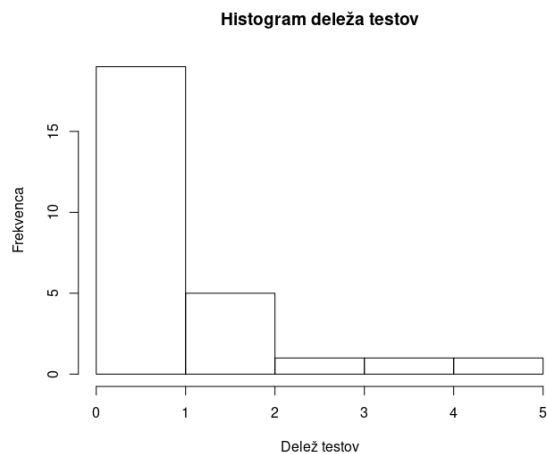


Slika 9. Histogram fatalnosti

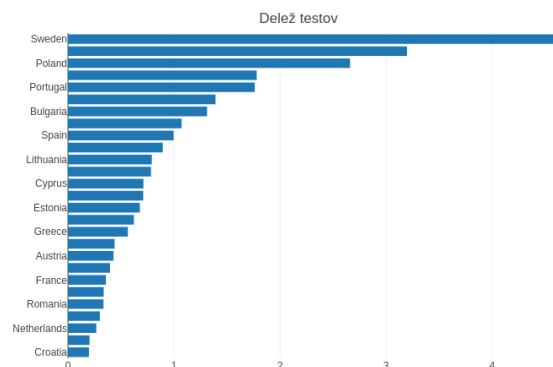


Slika 10. Barplot fatalnosti

Iz slike 11. in slike 12. opazamo, da spremenljivka delež opravljenih testov ni normalna ali simetrična. Testa normalnosti in simetrije bosta potrdili to intuicijo. Iz barplota vidimo, da obstaja velika raznolikost med državami, to pomeni so izbrane države različno obravnavale opravljanje testov.



Slika 11. Histogram deleža testov



Slika 12. Barplot deleža testov

Rezultati testa simetrije in testa normalnosti so prikazani v tabeli 1. Iz tabele 1 lahko razberemo, da 95% smo prepričani, da spremenljivke število dni do vrhunca prvega vala okuženih, število dni do vrhunca prvega vala mrtvih, delež okuženih D , fatalnost F in delež opravljenih testov T niso normalno porazdeljene. To drži, ker njihova p vrednost v stolpcu 'Test normalnosti' je manjša od $\alpha = 0.05$, zato zavrnilo ničelno hipotezo. Velja alternativna hipoteza. Izkazalo se je tudi, da spremenljivki delež okuženih D in delež opravljenih testov T nista simetrični. To drži, ker njihova p vrednost v stolpcu 'Test simetrije' je manjša od $\alpha = 0.05$, zato zavrnilo ničelno hipotezo. Velja alternativna hipoteza H_1 , torej sta asimetrični. Na podlagi zgornjih ugotovitev lahko se prikaže opisno statistiko spremenljivk v tabeli 2.

Spremenljivka	Test normalnosti	Test simetrije
Mediana starosti	$W = 0.94$ $p = 0.122$	Test st. = -0.422 $p = 0.684$
St. dni do vrhunca okuženih	$W = 0.85$ $p = 0.001$	Test st. = 2.32 $p = 0.052$
St. dni do vrhunca mrtvih	$W = 0.86$ $p = 0.002$	Test st. = 0.63 $p = 0.59$
Delež okuženih	$W = 0.62$ $p = 0.0000004$	Test st. = 3.94 $p = 0.002$
Fatalnost	$W = 0.89$ $p = 0.006$	Test st. = 1.5 $p = 0.26$
Delež testov	$W = 0.74$ $p = 0.00001$	Test st. = 2.82 $p = 0.018$

Tabela 1. Rezultati Shapiro-Wilk testa in testa simetrije

Opažamo raznolikost širjenja virusa v državah. Ker je število okuženih le vzorec, lahko

Spremenljivka	Min	Max	Povprečje/Mediana	St.odklon/Razpon
Mediana starosti	36.8	47.1	$\bar{x} = 42.18$	$s = 2.24$
St. dni do vrhunca okuženih	9	144	$M = 40$	$R = 28.5$
St. dni do vrhunca mrtvih	0	108	$M = 33$	$R = 19$
Delež okuženih	0.01%	0.58%	$M = 0.04\%$	$R = 0.06\%$
Fatalnost	0%	13.04%	$M = 2.81\%$	$R = 4.56\%$
Delež testov	0.2%	4.63%	$M = 0.71\%$	$R = 0.81\%$

Tabela 2. Opisna statistika spremenljivk

izračunamo 95% interval zaupanja za spremenljivki delež okuženih D in fatalnost F . Rezultati so prikazani v spodnjih tabelah 3 in 4. V tabelah je pomembna zgornja meja intervala. 95% interval zaupanja za delež okuženih i -te države Evropske unije je predstavljen v tabeli 3. Zgornja meja vsakega intervala je najslabši možen delež okuženih za posamezno državo. Opažamo, da zgornja meja ni drastično večja od zaznanega deleža. 95% smo prepričani, da je bil delež asimptomatikov na začetku majhen v primerjavi z zaznanim deležom okuženih.

Ime drzave	Spodnja meja intervala	Zgornja meja intervala
Austria	0.076%	0.08%
Belgium	0.279%	0.286%
Bulgaria	0.043%	0.046%
Croatia	0.022%	0.025%
Cyprus	0.024%	0.03%
Czechia	0.018%	0.02%
Denmark	0.085%	0.09%
Estonia	0.037%	0.044%
Finland	0.032%	0.036%
France	0.078%	0.08%
Germany	0.022%	0.022%
Greece	0.022%	0.024%
Hungary	0.012%	0.013%
Ireland	0.146%	0.153%
Italy	0.088%	0.089%
Latvia	0.008%	0.011%
Lithuania	0.026%	0.03%
Luxembourg	0.13%	0.149%
Malta	0.059%	0.075%
Netherlands	0.027%	0.029%
Poland	0.065%	0.067%
Portugal	0.149%	0.154%
Romania	0.03%	0.032%
Slovakia	0.017%	0.019%
Slovenia	0.006%	0.008%
Spain	0.201%	0.203%
Sweden	0.578%	0.588%

Tabela 3. 95% interval zaupanja deleža okuženih

95% interval zaupanja za fatalnost i -te države Evropske unije je predstavljen v tabeli 4. Zgornja meja vsakega intervala je najslabša možna fatalnost za posamezno državo. Tudi v tej tabeli opazamo, da zgornja meja ni drastično večja od zaznanega deleža. 95% smo prepričani, da je bila realna fatalnost na začetku majhna v primerjavi z zaznano fatalnostjo.

Ime drzave	Spodnja meja intervala	Zgornja meja intervala
Austria	0.558%	0.977%
Belgium	12.674%	13.407%
Bulgaria	4.682%	6.319%
Croatia	0.254%	1.423%
Cyprus	1.379%	5.458%
Czechia	0.213%	0.859%
Denmark	3.488%	4.589%
Estonia	0.01%	1.198%
Finland	0.88%	1.985%
France	6.611%	7.048%
Germany	0.181%	0.332%
Greece	4.215%	6.011%
Hungary	5.17%	8.059%
Ireland	3.152%	4.011%
Italy	8.769%	9.256%
Latvia	0%	2.604%
Lithuania	0.571%	2.287%
Luxembourg	0.426%	1.869%
Malta	0%	1.613%
Netherlands	3.922%	5.123%
Poland	4.209%	4.724%
Portugal	2.559%	3.087%
Romania	4.192%	5.283%
Slovakia	0.381%	1.675%
Slovenia	0%	3.306%
Spain	8.495%	8.855%
Sweden	8.466%	8.922%

Tabela 4. 95% interval zaupanja fatalnosti

4.2 Korelacijska analiza

Korelacijsko analizo bom razdelil na tri dele in sicer:

1. Mediana starosti (stolpec MED v bazi)
2. Delež testov (spremenljivka D definirana v 3. poglavju)
3. Število dni do vrha prvega vala okuženih (razlika v dnevih stolpcev DVO in DPO v bazi)

Podatke bom prikazal z razsevnimi diagrami. Izračunal bom linearno korelacijo izbranih spremenljivk z Pearsonovim koeficientom korelacije. Formula za vzorec je seldeča:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

kjer n je velikost vzorca, x_i je i -ti element spremenljivke X in \bar{x} je povprečje spremenljivke X . To se v R-ju izračuna z ukazom `cor(X, Y, method = "pearson")`, kjer sta X in Y izbrani

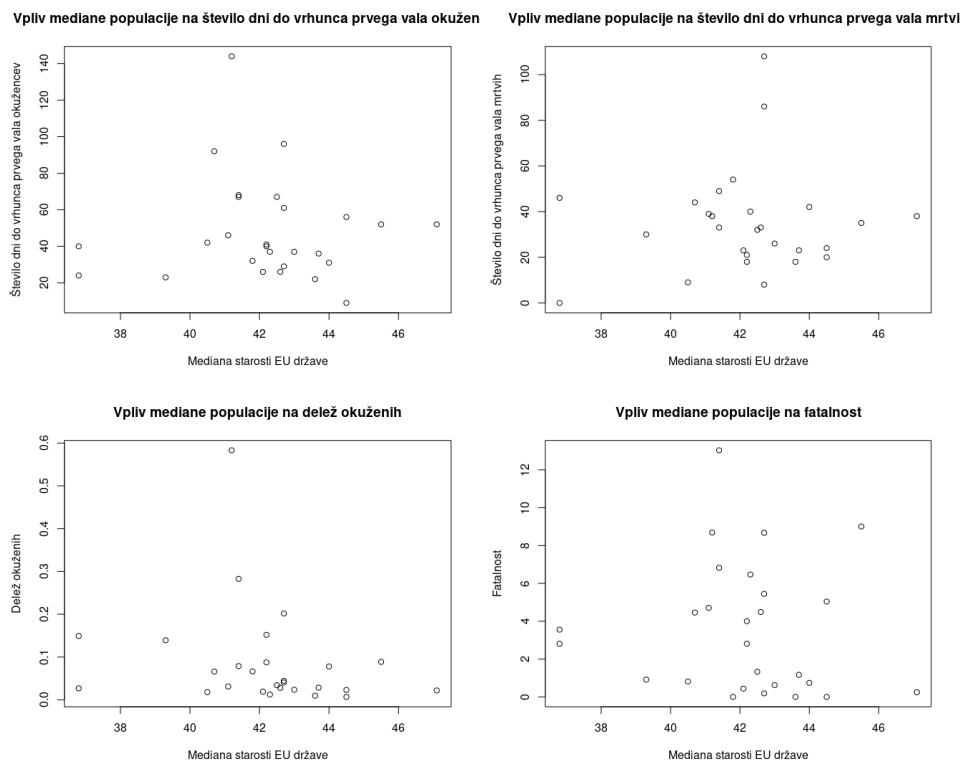
spremenljivki. Ker je izračunan vzorčni koeficient korelacije, lahko izračunamo 95% interval zaupanja za koeficiente korelacije. V R-ju se izračuna 95% interval zaupanja z ukazom `cor.test(X,Y, method = "pearson")`, kjer X in Y sta izbrani spremenljivki. Če bo korelacijski koeficient primeren, bo opravljen test moči korelacije. Testiralo se bo če je korelacijski koeficient populacije srednje močen oziroma večji od 0.3. Naj bo ničelna hipoteza H_0 : spremenljivki X in Y si nista srednje močno povezani in H_1 : spremenljivki X in Y sta srednje močno povezani. Izračunamo ga z `cor2.test(r,0.3, n)`[7], kjer n je velikost vzorca in r je linearni korelacijski koeficient spremenljivk X in Y.

4.2.1 Korelacijska analiza mediane starosti

Zanima me kako je mediana starosti povezana na druge spremenljivke in sicer na

1. število dni do vrhunca prvega vala okuženih (razlika v dnevih stolpcev DVO in DPO v bazi),
2. število dni do vrhunca prvega vala mrtvih (razlika v dnevih stolpcev DVS in DPS v bazi),
3. delež okuženih do vrhunca prvega vala okuženih (spremenljivka D definirana v 3. poglavju) in
4. delež mrtvih do vrhunca prvega vala okuženih (spremenljivka F definirana v 3. poglavju).

Podatke navedenih spremenljivk lahko prikažemo z razsevnimi diagrami. Velja, da v nobe-



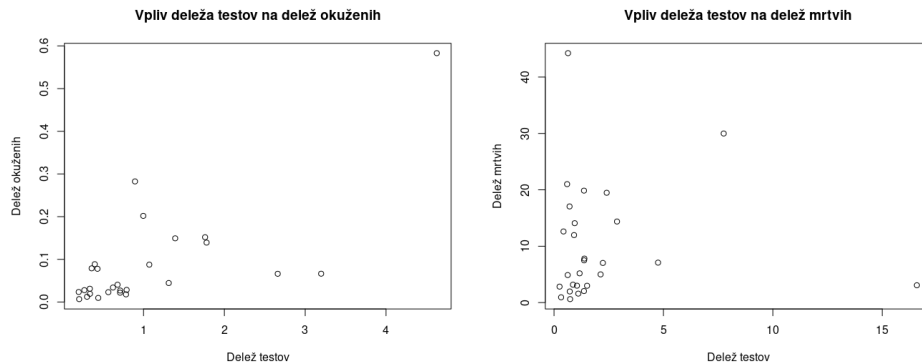
nem diagramu uspemo zaznati nek trend, toda to še ne pomeni, da so si pari spremenljivk neodvisni. Vsako korelacijo lahko izmerimo z Pearsonovim koeficientom korelacije. Rezultati računov so prikazani v tabeli 5. Noben koeficient v stolpcu 'Pearsonov koeficient' ni dovolj velik, da bi lahko smatrali, da obstaja močna povezanost med spremenljivkama. Pravzaprav ker je vsak koeficient tako blizu ničle lahko smatramo, da mediana starosti ni korelirana z ostalimi izbranimi spremenljivkami, to pa še ni gotovo. Ker je število okužencev le vzorec, ker niso šteti asimptomatiki, lahko izračunamo interval zaupanja za zgornje korelacijske koeficiente. Izračunani intervali zaupanja so prikazani v tabeli 5.

Spremenljivka	Pearsonov koeficient	Interval zaupanja
St. dni do vrhunca prvega vala okuženih	$r_1 = -0.0258$	$I_1 = [-0.4019, 0.3577]$
St. dni do vrhunca prvega vala mrtvih	$r_2 = 0.0902$	$I_2 = [-0.3001, 0.4546]$
Delež okuženih	$r_3 = -0.2203$	$I_3 = [-0.5539, 0.1743]$
Fatalnost	$r_4 = -0.0845$	$I_4 = [-0.45, 0.3053]$

Tabela 5. Rezultati korelacije mediane starosti z drugimi spremenljivkami

4.2.2 Korelacijska analiza deleža testov

Zanima me kako je delež testov povezan na druge spremenljivke in sicer na delež okuženih do vrhunca prvega vala okuženih in delež mrtvih do vrhunca prvega vala okuženih. Podatke navedenih spremenljivk lahko prikažemo z razsevnimi diagrami. Iz levega grafa lahko zaznamo nek trend. V desnem diagramu uspemo zaznati nek eksponenten trend, toda ne linearnega. Koeficient korelacije parov spremenljivk lahko izračunamo s Pearsonovim koeficientom korelacije. Rezultati so prikazani v tabeli 6.



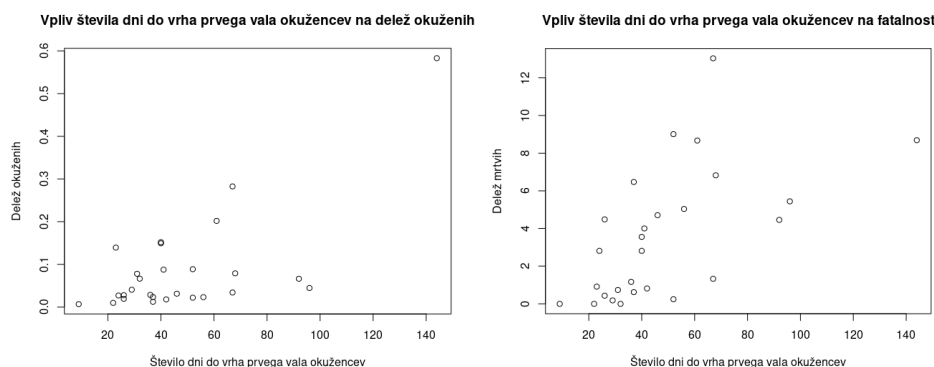
Koeficient r_1 je dovolj velik, da bi lahko smatrali, da obstaja povezanost med deležom opravljenih testov in deležom okuženih. Enako ne velja za delež opravljenih testov in fatalnostjo, kjer korelacijski koeficient r_2 je premajhen, da bi lahko sumili povezanost. Kot v razdelku 4.2.1 je število okužencev le vzorec, saj niso šteti asimptomatiki zaradi tega lahko izračunamo 95% interval zaupanja za korelacijski koeficient. Rezultati so prikazani v tabeli 7. Interval v drugi vrstici nam pove, da ne obstaja povezanost med deležom opravljenih testov in fatalnostjo. Testirati moramo srednjo korelacijo med deležom opravljenih testov in deležom zaznanih okuženih. Rezultat testa korelacije je prikazan v tabeli 6 pod stolpcem 'Test korelacije'. Ker je vrednost p vrednost manjša od $\alpha = 0.05$, lahko zavržemo ničelno hipotezo H_0 . Delež testov je srednje koreliran z deležom zaznanih okuženih.

Spremenljivka	Pearsonov koeficient	Interval zaupanja	Test korelacije
Delež okuženih	$r_1 = 0.7131$	$I_1 = [0.4568, 0.86]$	Test st. = 2.778 p = 0.003
Fatalnost	$r_2 = 0.1596$	$I_2 = [-0.2347, 0.5087]$	-

Tabela 6. Rezultati korelacije deleža opravljenih testov z drugimi spremenljivkami

4.2.3 Korelacijska analiza števila dni do vrha prvega vala okuženih

Zanima me kako je število dni do vrha prvega vala okuženih povezano z drugimi spremenljivkami in sicer na delež okuženih do vrhunca prvega vala okuženih in na delež mrtvih do vrhunca prvega vala okuženih. Podatke navedenih spremenljivk lahko prikažemo z razsevnimi diagrami. V levem diagramu lahko zaznamo nek trend. Podobno ne velja za desni diagram. Kot v zgornjih razdelkih izračunamo Pearsonov koeficient korelacije. Rezultati računov so prikazani v tabeli 7. Med številom dni do vrha prvega vala okuženih in fatalnostjo obstaja neka pobežanost, toda v manjši meri v primerjavi z drugim parom spremenljivk. Ker je



število zaznanih okuženih le vzorec, izračunamo 95% interval zaupanja za korelacijski koeficient. Izračunani intervali zaupanja so prikazani v tabeli 7 pod stolpcem 'Interval zaupanja'. Testirati moramo srednjo korelacijo med številom dni do vrha prvega vala okuženih in deležom okuženih ter fatalnostjo. Izračunani rezultati testa so prikazani v tabeli 7 pod stolpcem 'Test korelacije'. Ker je vrednost p_1 vrednost manjša od 0.05 lahko zavrnemo ničelno hipotezo. Število dni do vrha prvega vala okuženih in delež okuženih sta si vsaj srednje močno povezani. Enako velja za število dni do vrha prvega vala okuženih in fatalnost virusa. Ker je vrednost p vrednost manjša od 0.05 lahko zavrnemo ničelno hipotezo. Število dni do vrha prvega vala okuženih in fatalnost sta si vsaj srednje močno povezani.

Spremenljivka	Pearsonov koeficient	Interval zaupanja	Test korelacije
Delež okuženih	$r_1 = 0.6755$	$I_1 = [0.3976, 0.8399]$	Test st. = 2.433 p = 0.007
Fatalnost	$r_2 = 0.5856$	$I_2 = [0.2644, 0.7898]$	Test st. = 1.72 p = 0.043

Tabela 7. Rezultati korelacije št. dni do vrhunca prvega vala okuženih z drugimi spremenljivkami

5 Zaključki

Iz analize spremenljivk lahko trdimo, da število dni do vrhunca prvega vala okuženih, število dni do vrhunca prvega vala mrtvih, delež okuženih, delež mrtvih in delež testov niso normalno porazdeljene. Sklepamo lahko, da na te spremenljivke vplivajo tudi različni faktorji, kot je čas, ki ga je potrebovala posamezna država za organizacijo (maske, rokavice in socialno distanciranje), saj vsaka država je ukrepala poljubno. Velja, da Koronavirus se je širil različno med državami Evropske unije.

Iz korelacijske analize lahko smatramo, da spremenljivka mediane starosti ni linearno korelirana s številom dni do vrhunca prvega vala okuženih, številom dni do vrhunca prvega vala mrtvih, deležem okuženih do vrhunca prvega vala okuženih in deležem mrtvih do vrhunca prvega vala okuženih. Smatramo, da ne obstaja povezanost med mediano starosti in širjenjem ter fatalnostjo virusa. To pa še ni gotovo.

Korelacijska analiza deleža testov in drugih izbranih spremenljivk je privedla do zanimivih rezultatov. Izkazalo se je, da je spremenljivka deleža testov srednje močno linearno povezana z deležem okuženih, saj če država opravi več testov lahko zazna več okužencev in posledično več zaznani delež okuženih. Enako se ne izkaže za fatalnost virusa. Na fatalnost vplivajo tudi drugi faktorji, kot so zmogljivost bolnišnic in če je bolnik imel poleg te še druge bolezni.

Korelacijska analiza števila dni do vrha prvega vala okuženih in drugih izbranih spremenljivk je tudi privedla do zanimivih rezultatov. Število dni do vrha prvega vala okuženih in delež okuženih sta si srednje močno linearno povezani. Lahko zaključimo, da je smiselno, da obstaja neka pozitivna povezanost med spremenljivkama, saj v večjem številu dni se lahko zazna in zbolijo več okuženih ljudi. Podobno velja za spremenljivki število dni do vrha prvega vala okuženih in fatalnost, kjer linearna korelacija je tudi srednje močna. Smatramo lahko, da daljša je prva polovica prvega vala okuženih več bo okuženih in več bo mrtvih.

Virus se je po državah Evropske unije širil različno. Seveda na to je vplivala tudi lokacija države. Globalno gledano se je izkazalo, da mediana starosti populacije ni igrala tako pomembne vloge s širjenjem virusa v prvi polovici prvega vala okuženih. Obratno velja za delež testov in število dni do vrha prvega vala okuženih, za katere velja, da sta pozitivno linearno povezani z deležem okuženih. Iz tega sledi, da višje je število opravljenih testov višje je število zaznanih okuženih in da višje je število dni do vrha prvega vala okuženih višji je delež okuženih.

Literatura

- [1] Citat - Zuccarini,
<https://www.frasicelebri.it/frase/zuccarini-giuseppe-coronavirus-e-peggio-di-una-gue/>
- [2] List of countries by median age - Wikipedia,
https://en.wikipedia.org/wiki/List_of_countries_by_median_age
- [3] List of EU countries - Wikipedia,
https://en.wikipedia.org/wiki/European_Union
- [4] List of European countries by population - Wikipedia,
https://en.wikipedia.org/wiki/List_of_European_countries_by_population
- [5] Covid-19 - IHME,
<https://covid19.healthdata.org/>

- [6] Covid-19 - WHO,
<https://covid19.who.int/info>
- [7] Github repository - Matej Kalc,
<https://github.com/KalcMatej99/Seminarska-VS-Covid-19>
- [8] Pandemija koronavirusa: biološki, mikrobiološki in kemijski izsledki te okužbe - Martina Lizza,
https://drive.google.com/file/d/1KwhX7zzNJ0x5NnlzdY_DXD_-JdxXiGNj/view?usp=sharing
- [9] Covid-19 - Wikipedia
https://en.wikipedia.org/wiki/Coronavirus_disease_2019
- [10] SARS-CoV-2 - Wikipedia
https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2
- [11] Shapiro-Wilk test - Wikipedia
https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test
- [12] Plotly R package - Carson Sievert
<https://www.rdocumentation.org/packages/plotly>