

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

Matej Kalc

17. avgust 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. Zanima me kako so se ljudje odzvali na epidemijo in katere države so bile najboljše in katere najslabše organizirane za preprečevanje okužbe. Ker je epidemija še v teku, bom kot vzorec izbral države Evropske unije, ker se je v teh epidemija sprožila približno sočasno.

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. Analiziral bom 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.

Cilji študija sta:

- Ugotovitev ali se je koronavirus sprožil podobno v vseh državah Evropske unije in
- Vpliv raznih faktorjev na širjenje in fatalnost 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 Koronavirus [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-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: opravljal bom analizo spremenljivk in korelacijo 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. SARS-CoV-2 je virus RNA s pozitivnim občutkom razreda IV, ki je nalezljiv pri ljudeh. Kot je opisal ameriški Nacionalni zdravstveni inštitut (WHO), je naslednik virusa SARS-CoV-1, ki je povzročil izbruh SARS od leta 2002 do leta 2004. Taksonomsko je SARS-CoV-2 sev hudega koronavirusa, povezanega z akutnim respiratornim sindromom (SARSr-CoV). Menijo, da ima zoonotski izvor in ima tesno genetsko podobnost s koronavirusi netopirjev, kar kaže na to, da izvira iz virusa, ki se prenaša z netopirjem. Zaenkrat še ni dokazov, ki bi povezali vmesni rezervoar za živali, na primer pangolin, z njegovo uvedbo na ljudeh. Virus kaže malo genetske raznolikosti, kar kaže na to, da se bo verjetno zgodil dogodek prelivanja, ki bi ljudem uvedel SARS-CoV-2, konec leta 2019. Dihalni virus se širi preko kapljice sline 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 6.540 zaznanih okuženih in 355 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. Drugi faktor ki vpliva, je število opravljenih testov. Višje bo število opravljenih testov višji bo zaznani delež okuženih.

3 Podatki

Podatki so bili izbrani iz spleta[2][3][4][5][6][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 zazanega 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 s deleži, ki vzamejo v zakup število prebivalcev ali število okuženih, omogočajo 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 (definirano v 3. poglavju)
5. Delež mrtvih do vrhunca prvega vala okuženih
6. Delež testov do vrhunca prvega vala okuženih

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 en od 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

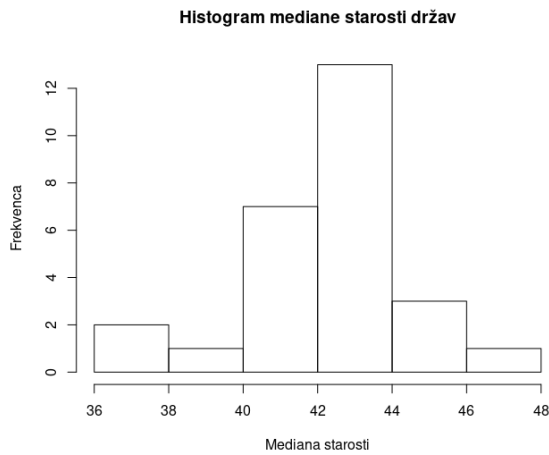
$$I_p = [\hat{p} - \Delta, \hat{p} + \Delta],$$

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

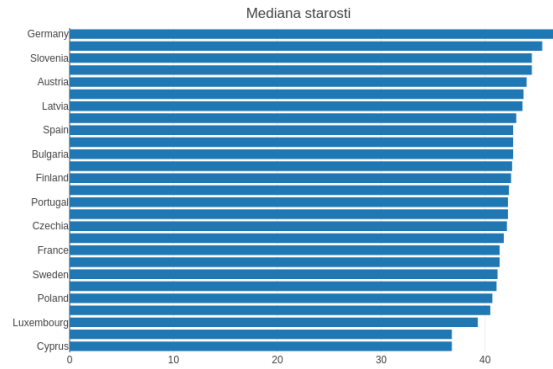
$$\Delta = c\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}},$$

za velikost vzorca n in $c = \Phi^{-1}(0.975) = 1.96$ kvantil normalne porazdelitve.

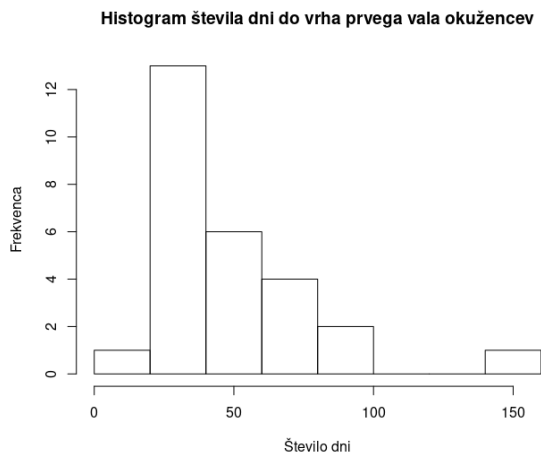
Izbrane spremenljivke prikažemo grafično s histogramom in barplotom.



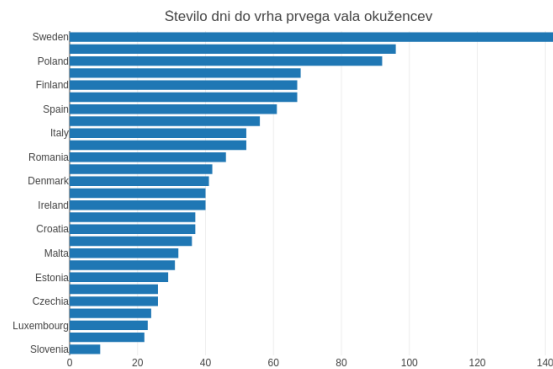
Slika 1. Histogram



Slika 2. Barplot



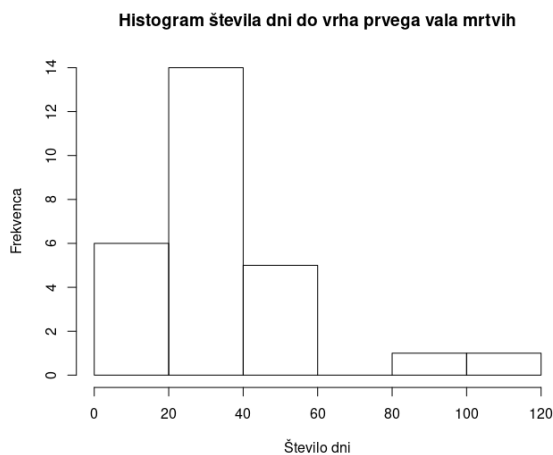
Slika 3. Histogram



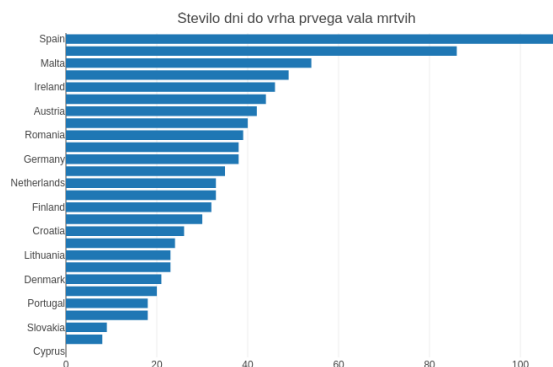
Slika 4. Barplot

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.

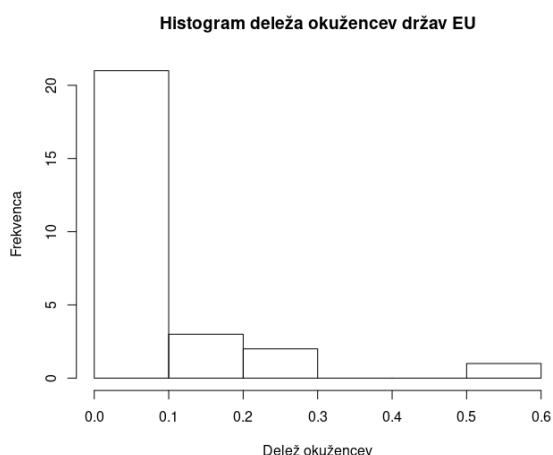
Iz slike 3. in slike 4. opazamo, da spremenljivka število dni do vrha prvega vala okuženih ni normalna ali simetrična. Naslednji računi bodo potrdili to intuicijo. Iz barplota vidimo, da obstaja velika raznolikost med državami. Sumimo, da ima spremenljivka velik razpon.



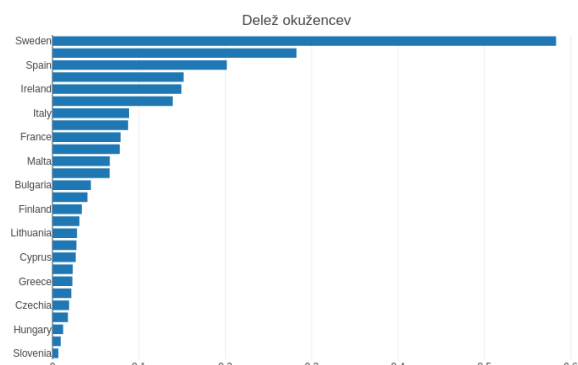
Slika 5. Histogram



Slika 6. Barplot



Slika 7. Histogram

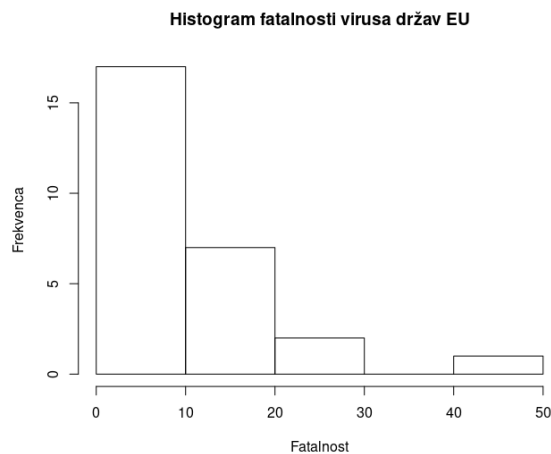


Slika 8. Barplot

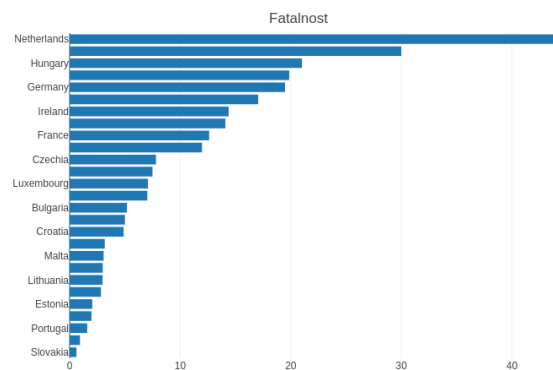
Iz slike 5. in slike 6. opazamo, da spremenljivka število dni do vrha prvega vala okuženih ni normalna ali simetrična. Računi bodo potrdili to intuicijo. Iz barplota vidimo, da obstaja velika raznolikost med državami. Poseben primer je Ciper. Za to državo velja da je imela maksimum mrtvih kar prvi dan. V naslednjih dneh je bilo zaznanih vedno manjše število. Zaradi raznolikosti sumimo, da ima spremenljivka velik razpon.

Iz slike 7. in slike 8. opazamo, da spremenljivka delež okuženih ni normalna ali simetrična. Testi bodo potrdili to intuicijo. 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 interval.

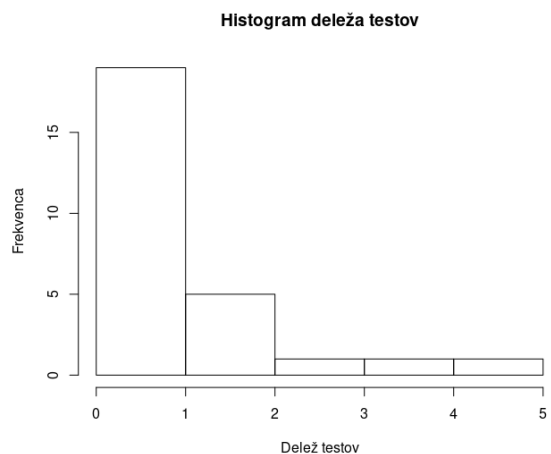
Iz slike 9. in slike 10. opazamo, da spremenljivka fatalnost F ni normalna ali simetrična. Testi bodo potrdili to intuicijo. Iz barplota vidimo, da obstaja velika raznolikost med državami, kljub temu da primerjamo deleže. Res je, da so ti deleži okuženih zelo majhni saj



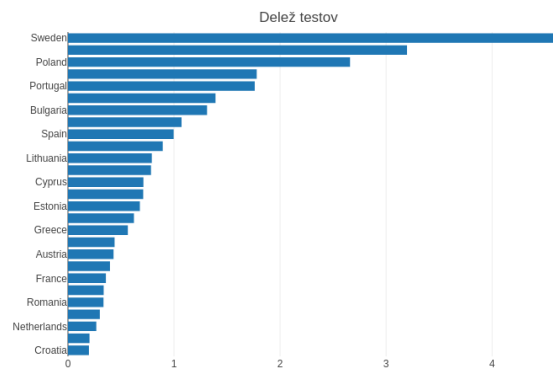
Slika 9. Histogram



Slika 10. Barplot



Slika 11. Histogram



Slika 12. Barplot

predstavljajo le izbrani interval.

Iz slike 11. in slike 12. opazamo, da spremenljivka fatalnost F ni normalna ali simetrična. Testi bodo potrdili to intuicijo. 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 interval.

Rezultati so prikazani v tabeli 1 in tabeli 2 v poglavju 4.3. Iz tabele 1 lahko razberemo, da z verjetnostjo 95% 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 porazdeljeni. 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 prikažejo podatki o spremenljivkah v tabeli 2. Opazamo raznolikost širjenja virusa v državah.

Ker je število okuženih le vzorec, lahko izračunamo 95% interval zaupanja za spremenljivki delež okuženih D in fatalnost F . Rezultati so prikazani v tabelah 3 in 4 v poglavju 4.3. V tabelah je pomembna zgornja intervala.

Tabela 3 Z verjetnostjo 95% lahko trdimo da realno število okuženih i -te države v i -ti vrstici pripada intervalu v i -ti vrstici. Zgornja meja vsakega intervala je najslabši možen delež okuženih za posamezno državo. Opažamo, da zgornja meja ni drastično večja od zaznanega deleža. Z verjetnostjo 95% lahko trdimo, da je bil delež asimptomatikov na začetku majhen v primerjavi z zaznanim deležom okuženih.

Tabela 4 Z verjetnostjo 95% lahko trdimo da realna fatalnost i -te države v i -ti vrstici pripada intervalu v i -ti vrstici. Zgornja meja vsakega intervala je najslabša možna fatalnost za posamezno državo. Tudi v tej tabeli opažamo, da zgornja meja ni drastično večja od zaznanega deleža. Z verjetnostjo 95% lahko trdimo, da je bila realna fatalnost na začetku majhna v primerjavi z zaznano fatalnostjo.

4.2 Korelacijska analiza

Korelacijsko analizo bom razdelil na tri dele in sicer:

1. Mediana starosti (stolpec MED v bazi)
2. Delež testov
3. Število dni do vrha prvega vala okuženih

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

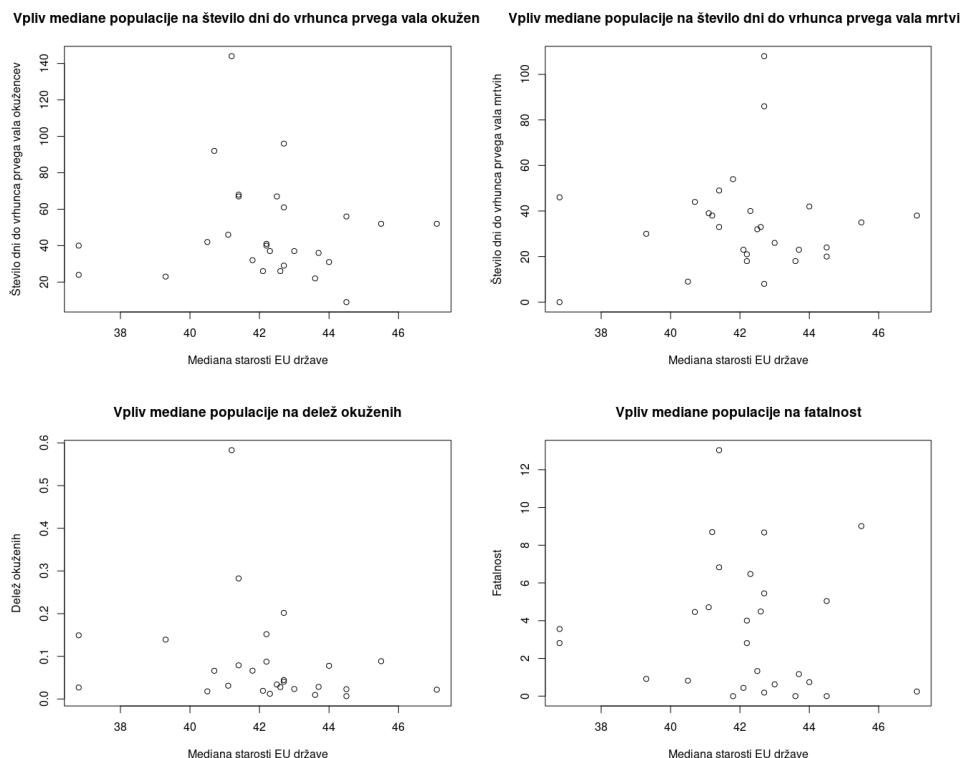
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

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 korelacijski koeficient le vzorec, lahko izračunamo 95% intervalom 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 vplivala na druge spremenljivke in sicer na

1. število dni do vrhunca prvega vala okuženih,
2. število dni do vrhunca prvega vala mrtvih,



3. delež okuženih do vrhunca prvega vala okuženih in

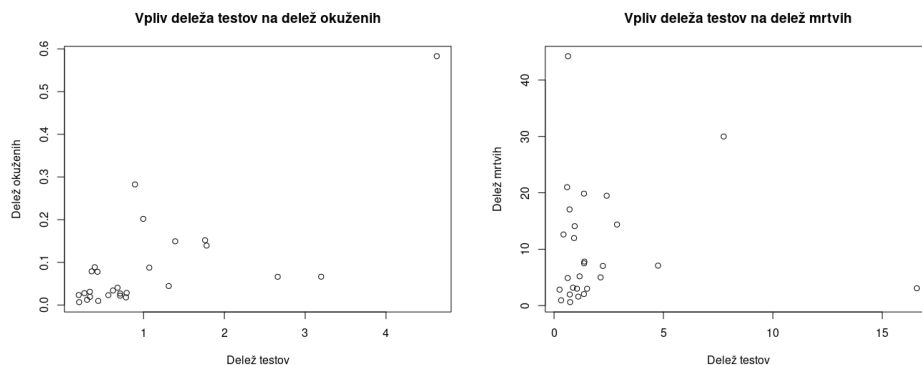
4. delež mrtvih do vrhunca prvega vala okuženih.

Podatke navedenih spremenljivk lahko prikažemo z razsvenim grafom.

Velja, da v nobenem 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 v poglavju 4.3. 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 spremenljivka M mediana starosti ne vpliva na ostale izbrane spremenljivke, to pa še ni gotovo. Ker je število okuženecv le vzorec, ker niso šteti asimptomatiki, lahko izračunamo interval zaupanja za zgornje korelacijske koeficiente. Izračunani intervali zaupanja so prikazani v tabeli 5 v poglavju 4.3. Da bi bila spremenljivka M v korelaciji z neko drugo izbrano spremenljivko, bi moral biti koeficient korelacije močno pozitiven oziroma večji od 0.7 ali močno negativen oziroma manjši od -0.7. Nobena vrednost v intervalih zaupanja ne zadošča pogoju zaradi tega lahko smatramo z verjetnostjo 95%, da mediana starosti ne vpliva nobeno drugo izbrano spremenljivko.

4.2.2 Korelacijska analiza deleža testov

Zanima me kako je delež testov vpliva 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 razsvenim grafom. 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 v poglavju 4.3.

Koeficient r_1 je dovolj velik, da bi lahko smatrali, da obstaja močna povezanost med deležom opravljenih testov in deležom okuženih. Lahko trdimo, da spremenljivki sta si odvisni. Enako ne velja za delež opravljenih testov in fatalnostjo, kjer korelacijski koeficient r_2 je premajhen, da bi lahko sumili močno povezanost. Kot v razdelku 4.2.1 je število okuženecv le vzorec, saj niso šteti asimptomatiki zaradi tega lahko izračunamo 95% interval zaupanja za korelacijski koeficient. Rezultati so prikazani v tabeli 6 v poglavju 4.3. Interval v drugi vrstici nam pove, da ne obstaja z verjetnostjo 95% močna 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 koreliran z deležom zaznanih okuženih vsaj srednje močno.

4.2.3 Korelacijska analiza števila dni do vrha prvega vala okuženih

Zanima me kako je število dni do vrha prvega vala okuženih povezano na druge spremenljivke in sicer na delež okuženih do vrhunca prvega vala okuženi in na delež mrtvih do vrhunca prvega vala okuženih. Podatke navedenih spremenljivk lahko prikažemo z razsvenimi 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 v poglavju 4.3. Trdimo lahko, da obstaja neka linearna povezanost med številom dni do vrha prvega vala okuženih in deležom okuženih, saj njihov korelacijski koeficient r_1 se bliža vrednosti 0.7. Med številom dni do vrha prvega vala okuženih in fatalnostjo obstaja neka povezanost, 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 zavržemo 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 zavržemo ničelno hipotezo. Število dni do vrha prvega vala okuženih in fatalnost sta si vsaj srednje močno povezani.



5 Zaključki

Iz analize spremenljivk lahko trdimo, da spremenljivke število dni do vrhunca prvega vala okuženih, števila dni do vrhunca prvega vala mrtvih, delež okuženih, delež mrtvih in delež testov niso normalno porazdeljene. Sklepamo lahko, da na te spremenljivke vplivajo drugi različni faktorji, kot je čas, ki je potrebovala posamezna država za regulacije (maske, rokavice in socialno distanciranje), saj vsaka država je ukrepala poljubno.

Iz korelacijske analize lahko smatramo, da spremenljivka mediane starosti ni linearno korelirana z številom dni do vrhunca prvega vala okuženih, številom dni do vrhunca prvega vala mrtvih, deležom okuženih do vrhunca prvega vala okuženih, deležom mrtvih do vrhunca prvega vala okuženih. Noben koeficient korelacije ni pozitivno močen ali negativno močen, pravzaprav so vsi zelo blizu ničli, razen za delež okuženih. Podobno velja za izračunane intervale zaupanja zaradi tega se sluti, da spremenljivka mediana starosti ne vpliva na ostale izbrane spremenljivke.

Korelacijska analiza deleža testov in drugih izbranih spremenljivk je privedla do zanimivih rezultatov. Izkaže se, da je spremenljivka deleža testov srednje močno linearno povezana z deležom okuženih, saj če država opravi več testov lahko zazna več okužencev in večja 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 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 povezanost med spremenljivkama, saj v večjem številu dni se lahko zazna in zboli več okuženih ljudi. Podobno velja za spremenljivki število dni do vrha prvega vala okuženih in fatalnost, kjer linearna korelacija obstaja je srednje močnejša. Smatramo lahko, da daljša je prva polovica prvega vala okuženih več bo okuženih in posledično več bo mrtvih.

Globalno gledano se je izkazalo, da mediana starosti populacije ni igrala tako pomembne vloge v širjenjem virusa v prvi polovici prvega vala okuženih. Obratno velja za delež testov in število dni do vrha prvega vala okuženih. Za katerih velja, da sta pozitivno linearno povezani z deležom okuženih. Iz tega sledi, da višje je število opravljenih testov višje je število zaznanih okuženih in višje je število dni do vrha prvega vala okuženih višji je delež okuženih.

6 Literatura

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] Lawstat R package - Vyacheslav Lyubchich
<https://www.rdocumentation.org/packages/lawstat/versions/3.4>
- [11] Shapiro-Wilk test - Wikipedia
https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test

6.1 Tabele

Legenda

- SDO - Število dni do vrhunca prvega vala okuženih
- SDM - Število dni do vrhunca prvega vala mrtvih
- D - Delež okuženih
- F - Fatalnost
- T - Delež opravljenih testov

Spremenljivka	Test normalnosti	Test simetrije
Mediana starosti	$W = 0.94$ $p = 0.122$	Test st. = -0.422 $p = 0.684$
SDO	$W = 0.85$ $p = 0.001$	Test st. = 2.32 $p = 0.052$
SDMh	$W = 0.86$ $p = 0.002$	Test st. = 0.63 $p = 0.59$
D	$W = 0.62$ $p = 0.0000004$	Test st. = 3.94 $p = 0.002$
F	$W = 0.89$ $p = 0.006$	Test st. = 1.5 $p = 0.26$
T	$W = 0.74$ $p = 0.00001$	Test st. = 2.82 $p = 0.018$

Slika 13. Tabela 1

Spremenljivka	Min	Max	Povprecje/Mediana	St.odklon/Razpon
Mediana starosti	36.8	47.1	$\bar{x} = 42.18$	$s = 2.24$
SDO	9	144	$M = 40$	$R = 28.5$
SDM	0	108	$M = 33$	$R = 19$
D	0.01%	0.58%	$M = 0.04\%$	$R = 0.06\%$
F	0%	13.04%	$M = 2.81\%$	$R = 4.56\%$
T	0.2%	4.63%	$M = 0.71\%$	$R = 0.81\%$

Slika 14. Tabela 2

Spremenljivka	Pearsonov koeficient	Interval zaupanja
SDO	$r_1 = -0.0258$	[-0.4019, 0.3577]
SDM	$r_2 = -0.0845$	[-0.4546, 0.3053]
D	$r_3 = -0.2203$	[-0.5539, 0.1743]
F	$r_4 = -0.0845$	[-0.45, 0.3053]

Slika 15. Tabela 5

Spremenljivka	Pearsonov koeficient	Interval zaupanja	Test korelacije
D	$r_1 = 0.7131$	[0.4568, 0.86]	Test st. = 2.778 $p = 0.003$
F	$r_2 = 0.1596$	[-0.2347, 0.5087]	-

Slika 16. Tabela 6

Spremenljivka	Pearsonov koeficient	Interval zaupanja	Test korelacije
D	$r_1 = 0.6755$	[0.3976, 0.8399]	Test st. = 2.433 $p = 0.007$
F	$r_2 = 0.5856$	[0.2644, 0.7898]	Test st. = 1.72 $p = 0.043$

Slika 17. Tabela 7

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%

Slika 18. Tabela 3

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%

Slika 19. Tabela 4