

Raport z przedmiotu Metody Ekonometryczne - Rafał Stępień 82640

Temat: Badanie ilości przeprowadzonych testów na SARS-COV-2 w państwach na świecie w zależności od wybranych zmiennych

Wstęp

W dniu pisania raportu zaczyna powoli przygasać pandemia koronawirusa, która obróciła świat do góry nogami i której skutki odczuwają społeczeństwa na całym świecie. COVID-19 jest czymś nowym dla świata medycyny, a żeby go wykryć, należy wykonać kosztowny oraz długotrwały test. Eksperci uważają, że to skuteczna diagnostyka jest kluczem do zwalczania epidemii, a żeby była ona efektywna, niezbędne jest pilne testowanie w dużych ilościach.

Chciałem w związku z tym przeprowadzić analizę, która wskazywałaby zależności między poszczególnymi zmiennymi a ilością testów na koronawirusa. Problemem, jakim zamierzam się zająć, jest zweryfikowanie, czy istnieje rzeczywista zależność między ilością testów a kondycją służby zdrowia, zaludnieniem oraz poziomem rozwoju danego kraju. Przy tym wszystkim mam zamiar także zweryfikować, czy rzeczywiście ilość wykrytych przypadków ma odzwierciedlenie w ilości testów, jak to wpływa na stopę śmiertelności oraz sprawdzić, jak kraje Unii Europejskiej radzą sobie z diagnostyką na tle reszty świata.

Opis zbioru danych oraz badanego modelu

Na początek przytoczę zbiór danych, mającego w sumie 212 obserwacji. Szkieletem tego zbioru są "zescrapowane" dane dotyczące pandemii koronawirusa ze strony Worldometers.info (z dnia 19.05.2020), dodałem jednakże również inne zmienne pochodzące z innych zbiorów danych. Omówię te dane, które zostały rozpatrzone w modelu. Kolumna *Country.Other* zawiera nazwy krajów, które zostały uwzględnione w badaniu. *Tot.Cases.1M.Pop*, *Deaths.1M.Pop* oraz *Tests.1M.Pop* to zmienne opisujące kolejno ilość przypadków, śmierci oraz testów na milion obywateli. *Population* przytacza ilość ludności w państwie. *Case.Fatality.Rate* to wskaźnik śmiertelności. Te dane są zescrapowane, ponadto dodałem zmienne z innych zbiorów, które pobierałem m.in. z danych OECD czy WHO - *GDP* to PKB per capita, *GDP for health expenditure* to odsetek PKB przeznaczany na służbę zdrowia, *CO2 emission per capita* to ilość emisji CO2 na osobę a *Life Expectancy* to średnia długość życia.

Specyfikacja modelu, gdzie zmienną objaśnianą *tests* są wartości z *Tests.1M.Pop*, to:

$$\ln(\text{tests}) = \beta_0 + \beta_1 \ln(\text{cases}) + \beta_2 \text{fatality_rate} + \beta_3 \ln(\text{GDPpercapita}) + \beta_4 \text{GDPforhealthexp} + \beta_5 \text{IsinEU} + \beta_6 \ln(\text{CO2emissionpercapita}) + \beta_7 \ln(\text{population}) + \beta_8 \text{LifeExp} + e$$

gdzie e to składnik losowy.

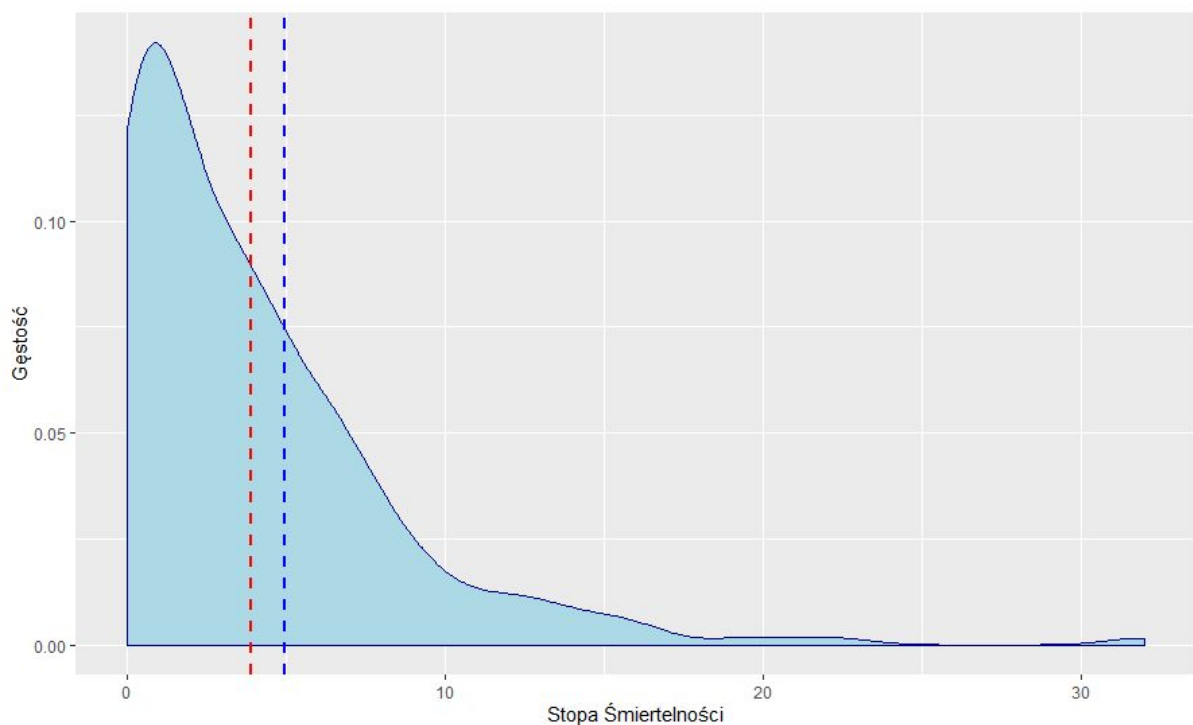
Zmienna *cases* odnosi się do *Tot.Cases.1M.Pop*, *fatality_rate* do *Case.Fatality.Rate*, *GDPpercapita* do *GDP*, *GDPforhealthexp* do *GDP for health expenditure*, *CO2emissionpercapita* do *CO2 emission per capita*, *population* do *Population* a *LifeExp* do *Life Expectancy*. Zmienna *IsinEU* to zmienna binarna, której wartość wynosi 1, gdy państwo przynależy do Unii Europejskiej, a 0, gdy nie.

Za pomocą zmiennych *GDPforhealthexp* oraz *LifeExp* będę mógł zbadać stan służby zdrowia. Te zmienne różnią się jednak o tyle, że w wypadku % PKB na służbę zdrowia trzeba wziąć pod uwagę też wysokość PKB danego państwa, w związku z tym może nie być zależności w takiej próbie. Natomiast *LifeExp* da nam pewien obiektywny miernik jakości służby zdrowia w danym kraju - idzie się spodziewać, że im lepsza służba zdrowia, tym lepsza średnia długość życia. Ta zmienna będzie też o tyle użyteczna, że da nam ogłęd, na ile struktura demograficzna państwa może korelować z ilością testów - idzie się spodziewać, że im dłuższa średnia długość życia, tym starsze społeczeństwo.

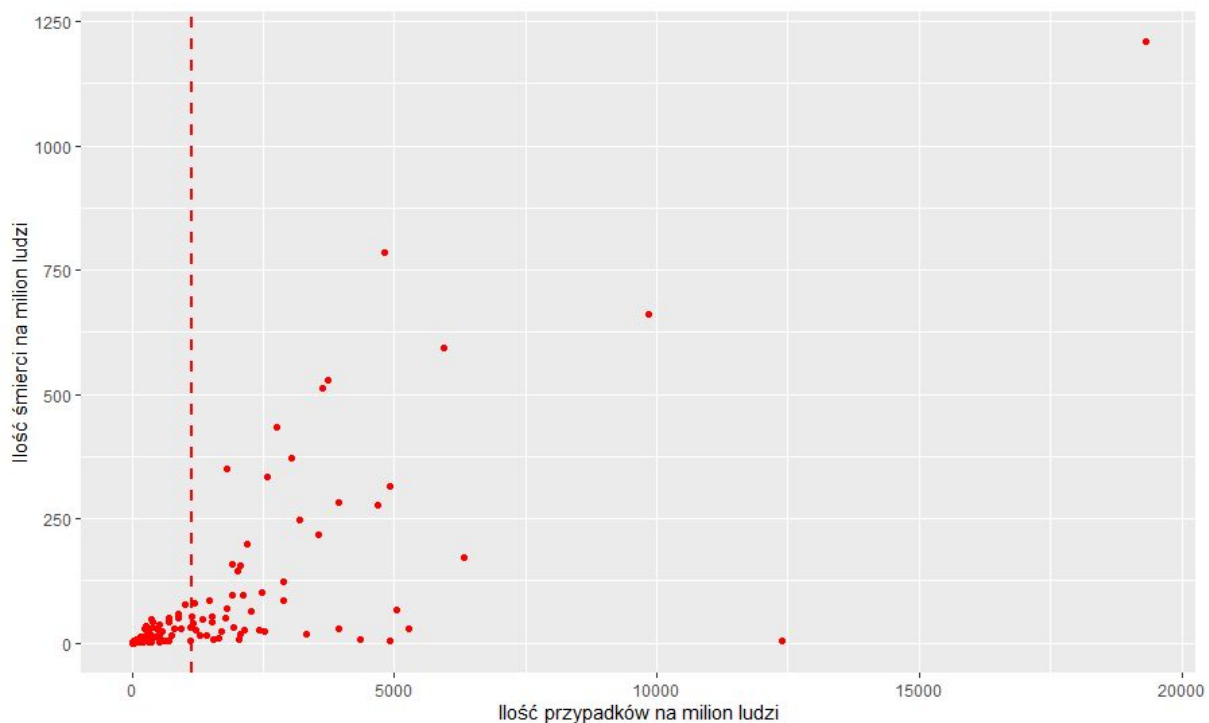
Dzięki zmiennej *population* zbadam korelację między liczbą ludności a ilością wykonywanych testów na milion mieszkańców.

Z kolei za pomocą zmiennej *CO2emissionpercapita* będę starał się odwzorować rozwój danego państwa na bazie wykazania korelacji między ilością generowanego CO2 na jedną osobę. O rozwoju państwa również może świadczyć zmienna *LifeExp*.

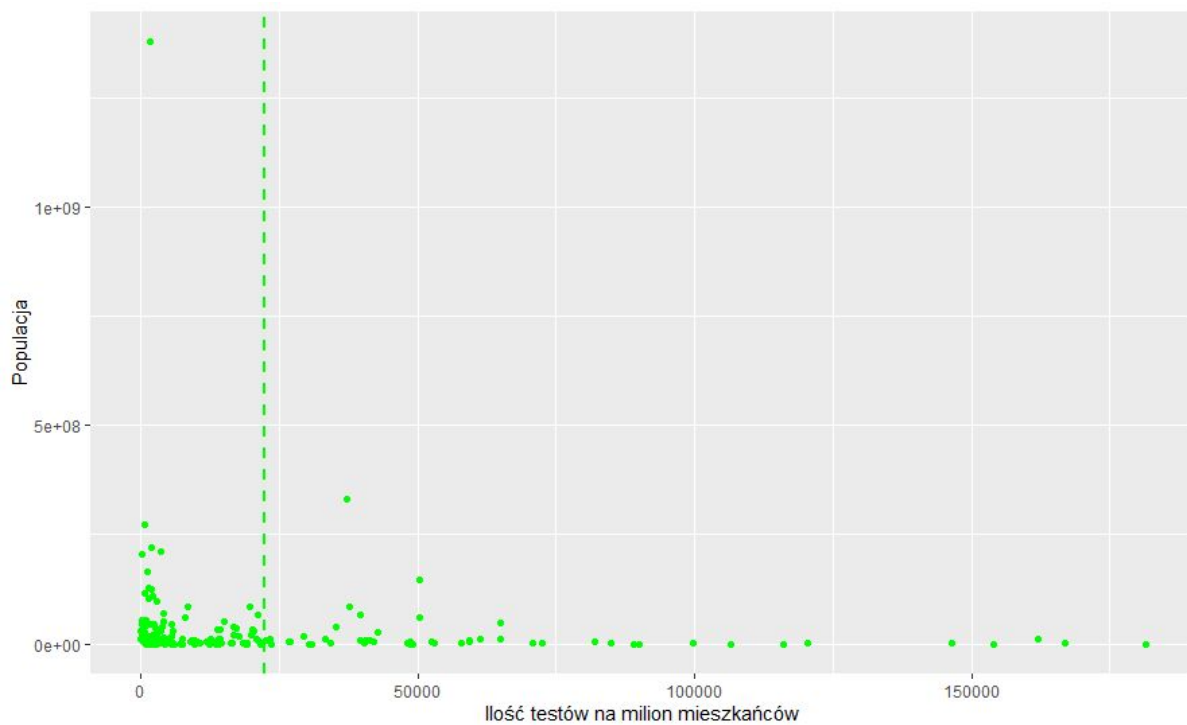
Prezentuję także wizualizację niektórych zmiennych. Tutaj mamy rozkład gęstości - ze wskazaną ogólną stopą śmiertelności (wynoszącą 4.94% z danych) na niebiesko - stopą śmiertelności w państwach. Na czerwono zaznaczona jest uśredniona wartość stóp śmiertelności w badanych państwach (wynosząca 3,9%).



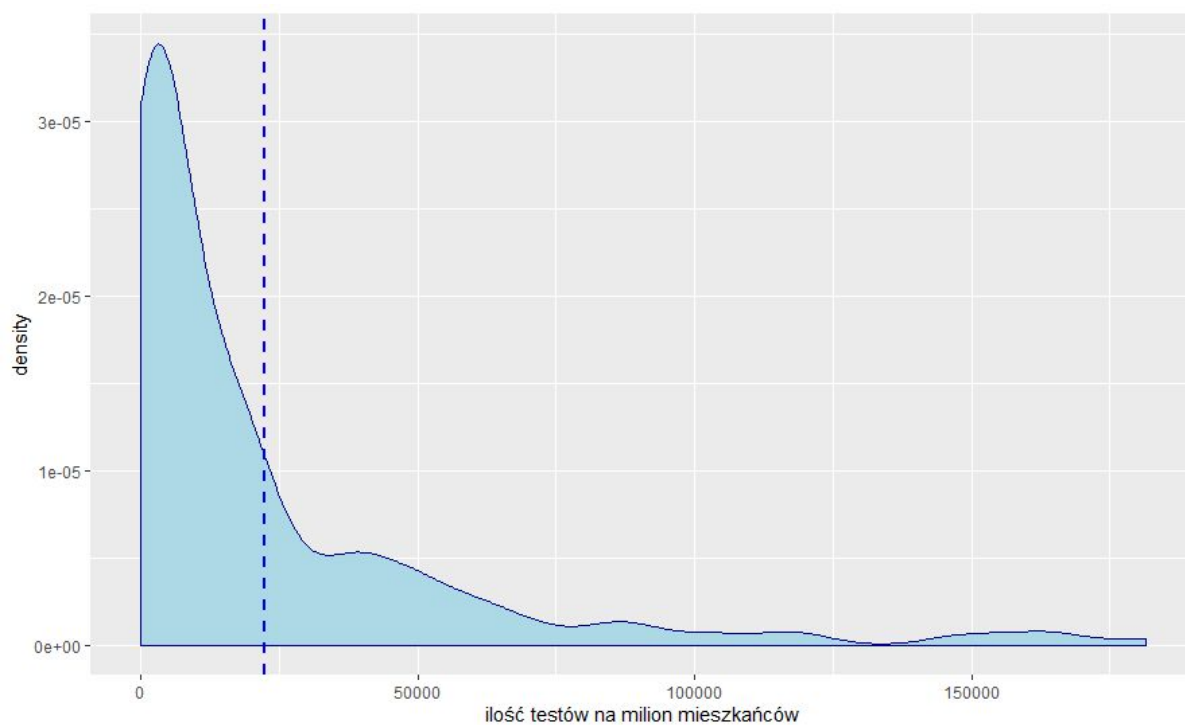
Tutaj widzimy natomiast, z jakimi liczbami mierzymy się w przypadku ilości przypadków na milion ludzi na tle ilości śmierci na milion ludzi razem ze średnią ilością przypadków na milion ludzi równą 1145, zapewne zaniżoną przez słabe warunki do diagnostyki w krajach trzeciego świata. Przypadki najbardziej skrajne to głównie mikropaństwa takie jak San Marino, Watykan czy Andora.



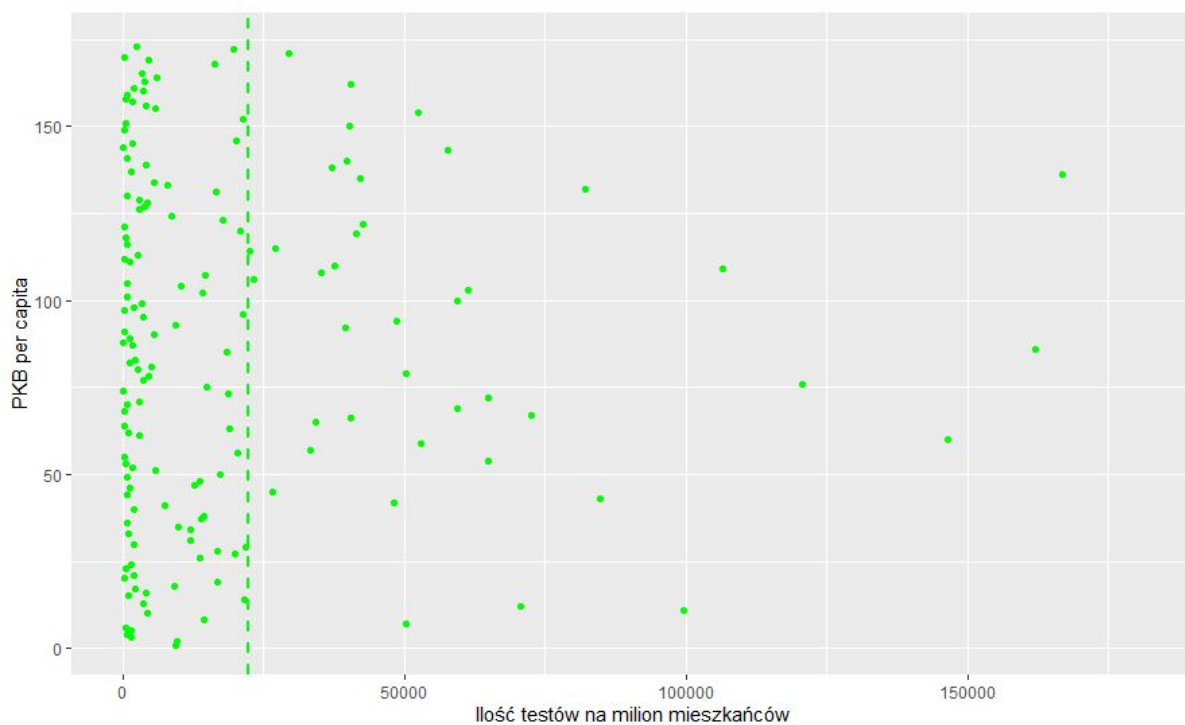
Widzimy również tutaj zależność między ilością testów na milion mieszkańców jak populacją. Wykres wskazywałby, że mniejsze państwa podejmują więcej testów. Pokazana jest również średnia ilość testów na milion mieszkańców wynosząca 22225.43.

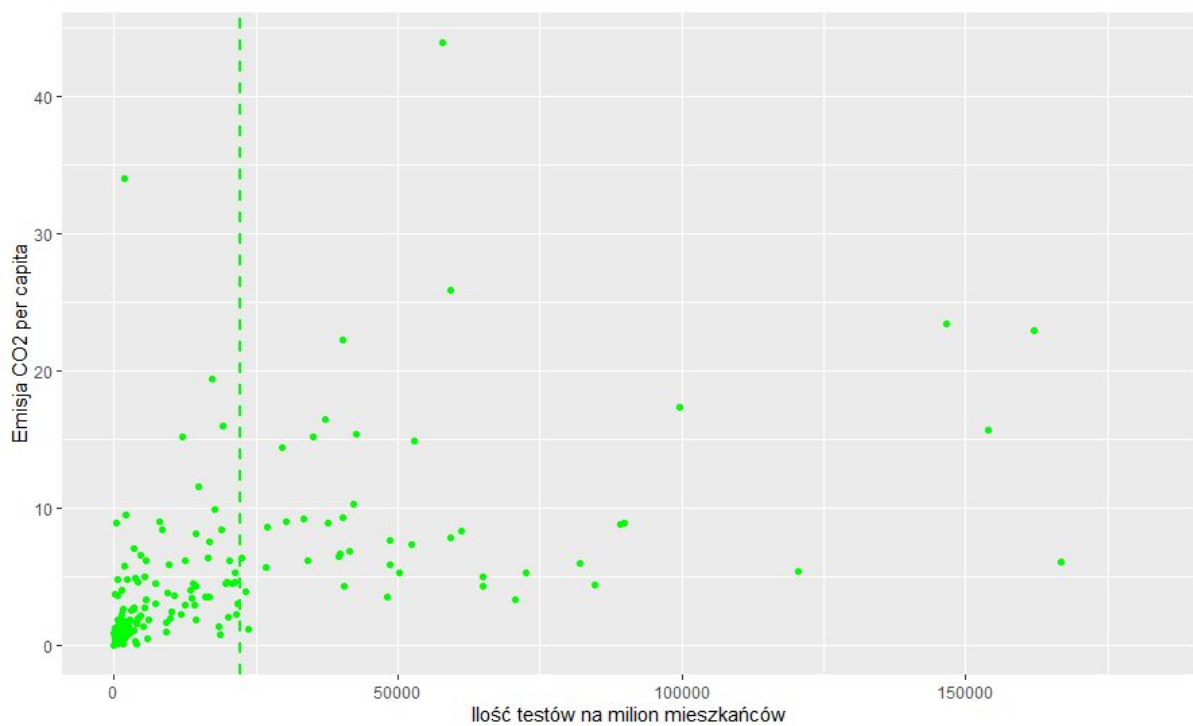
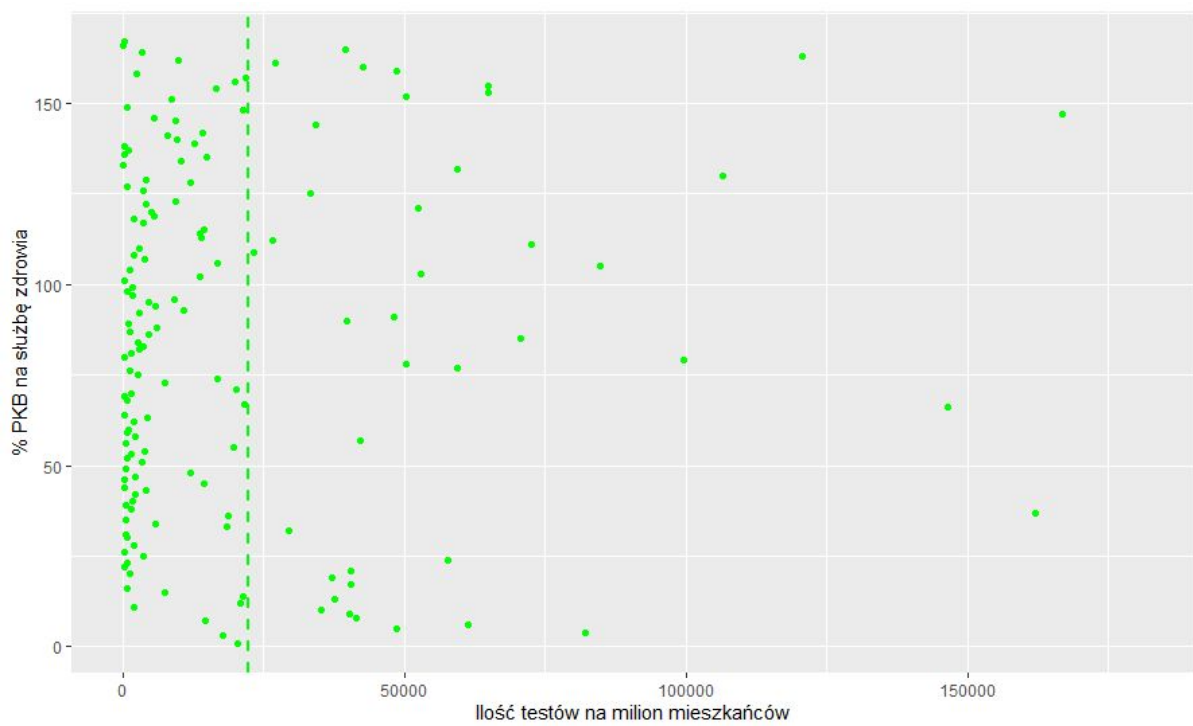


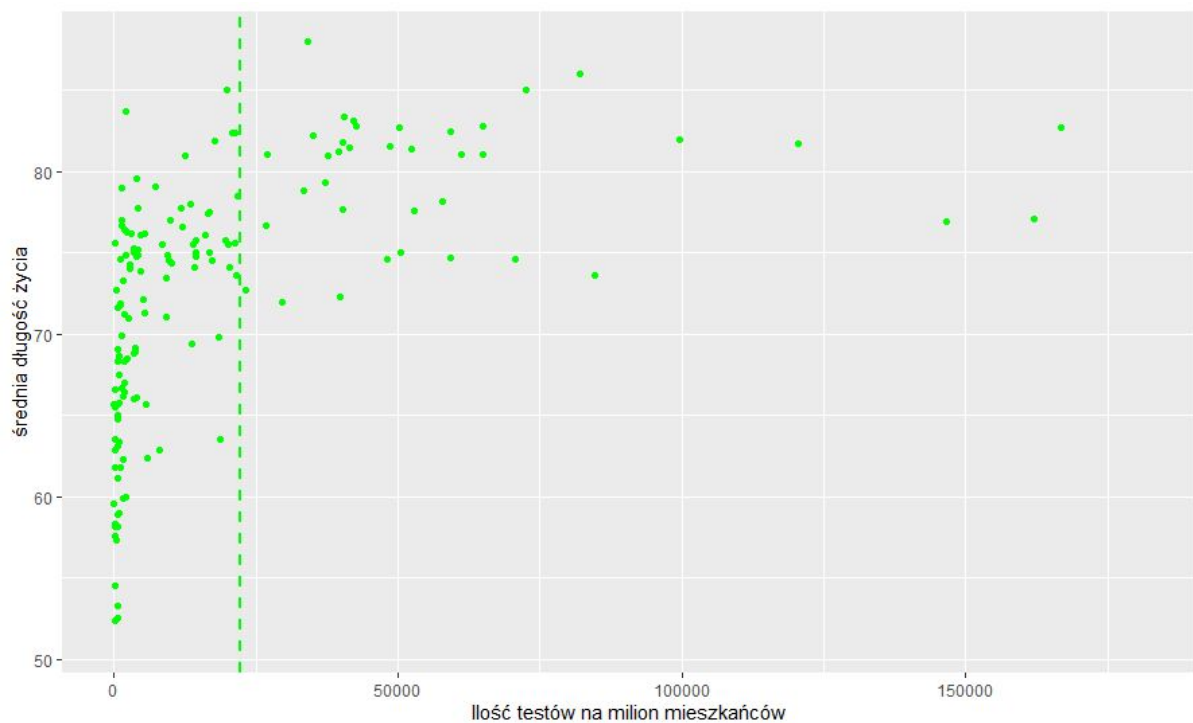
Wykres gęstości dla ilości testów:



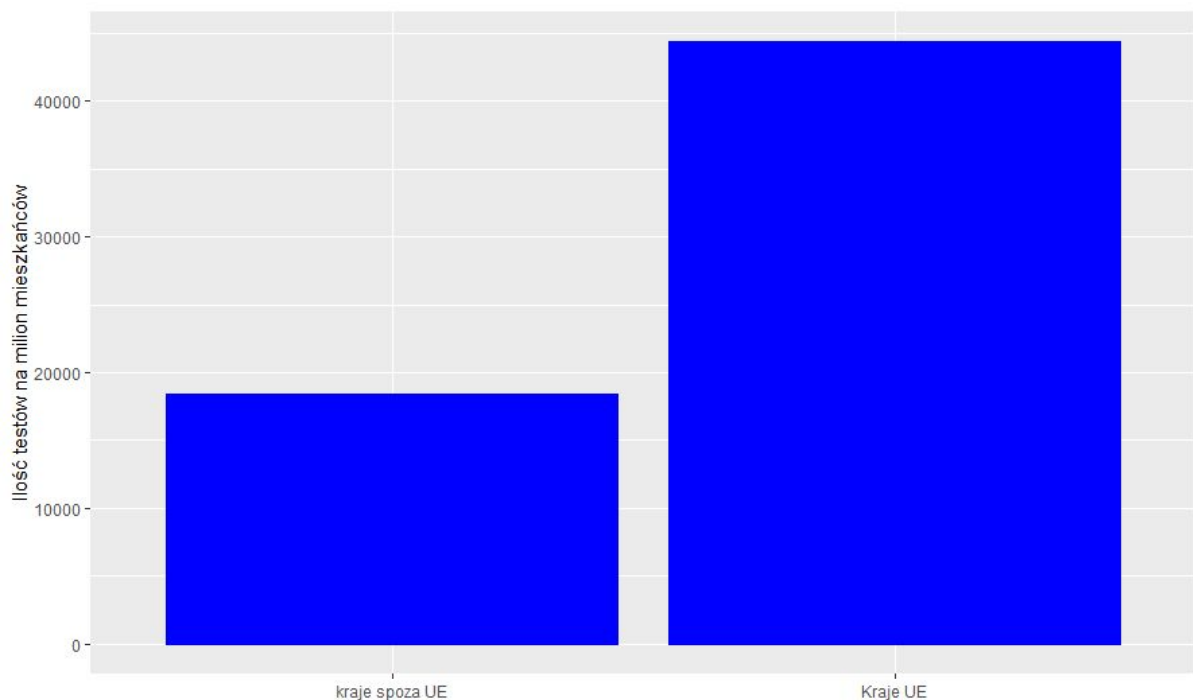
Tak to wygląda w przypadku pozostałych zmiennych:







A ten wykres prezentuje, ile kraje UE, a ile krajów spoza UE średnio wykonywały testów na koronawirusa na milion mieszkańców. Dla krajów spoza UE ta wartość to w zaokrągleniu 18441, dla krajów UE 44373.



Wydruk z programu RStudio wygląda tak:
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3431579	1.3372603	2.500	0.013664 *
log(cases)	0.2978555	0.0546421	5.451	2.43e-07 ***
fatality_rate	-0.1170544	0.0231602	-5.054	1.43e-06 ***
log(GDPpercapita)	-0.0005880	0.0798192	-0.007	0.994134
PercentGDPforhealthexp	-0.0004662	0.0016192	-0.288	0.773874
IsinEU	0.8978512	0.2301657	3.901	0.000153 ***
log(CO2emissionpercapita)	0.4016913	0.0839689	4.784	4.59e-06 ***
log(population)	-0.0838933	0.0451366	-1.859	0.065338 .
LifeExp	0.0687079	0.0176069	3.902	0.000152 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

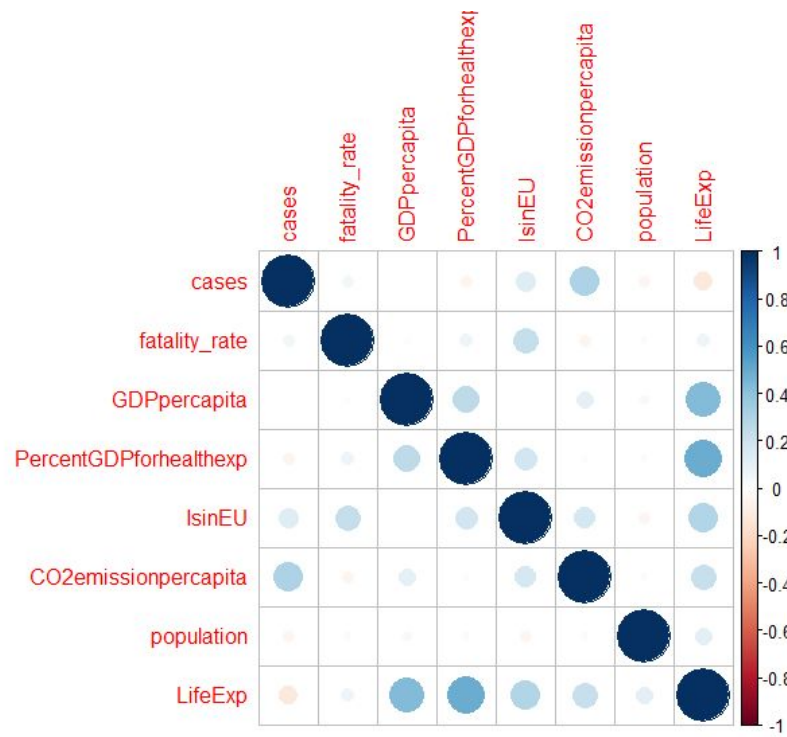
Residual standard error: 0.8873 on 130 degrees of freedom
(76 observations deleted due to missingness)

Multiple R-squared: 0.7914, Adjusted R-squared: 0.7786

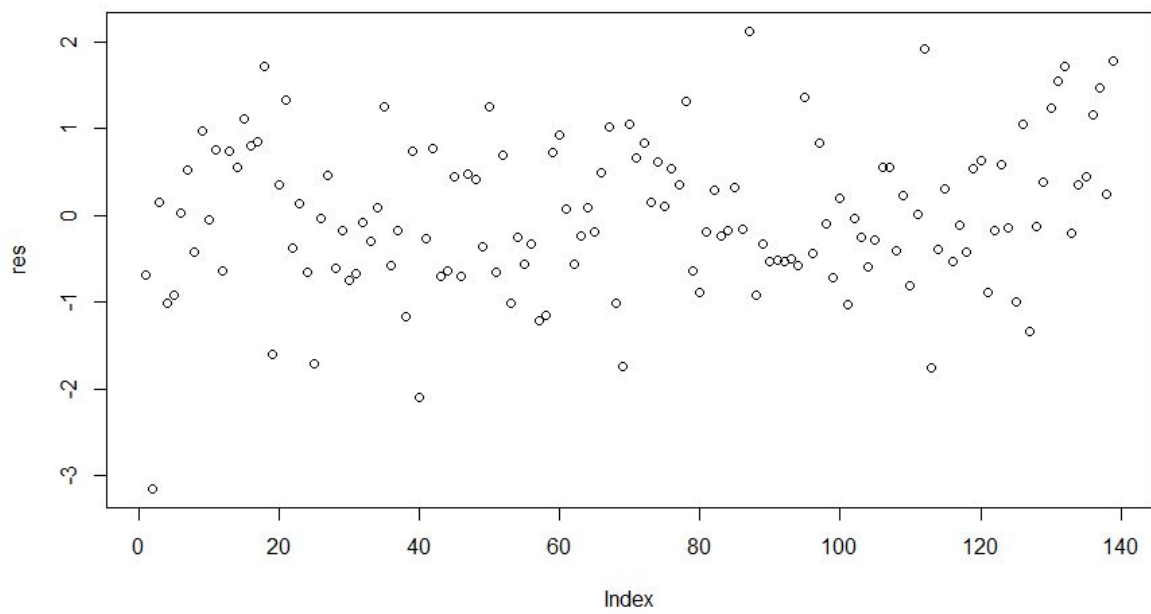
F-statistic: 61.65 on 8 and 130 DF, p-value: < 2.2e-16

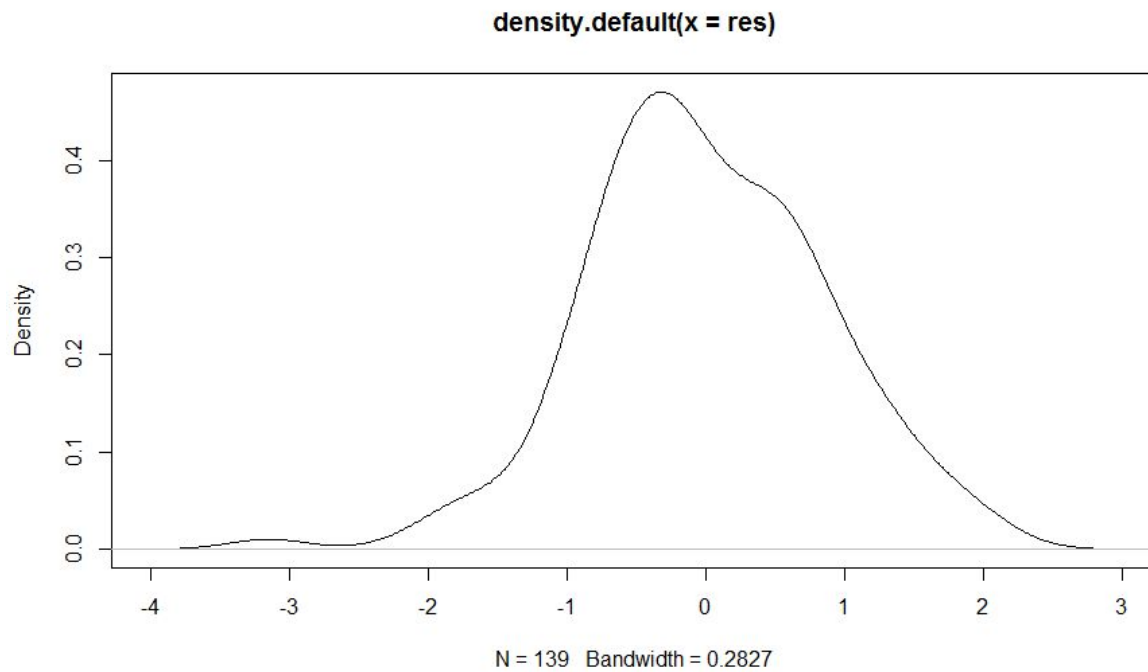
Zmienne log(GDPpercapita), PercentGDPforhealthexp oraz log(population) są, jak się okazuje, nieistotne statystycznie, więc nie poddają się interpretacji. Na poziomie istotności $\alpha=0.05$ widzimy natomiast, że ceteris paribus przy zwiększeniu się ilości testów o 1% ilość przypadków zwiększa się średnio o w przybliżeniu 0.3%, podczas gdy stopa śmiertelności maleje średnio w przybliżeniu o 0.12 punktów procentowych. Ponadto przy przyroście ilości testów o 1% zwiększa się przeciętnie ilość emisji CO2 per capita o 0.40% ceteris paribus - nie jest to kauzacja, ale pokazuje pewną korelację oraz to, że w krajach o większej emisji dwutlenku węgla per capita wykonuje się więcej testów. Tak samo przy przyroście liczby testów o 1%, średnia długość życia zwiększa się średnio w przybliżeniu o 0.07 roku ceteris paribus, można więc wnioskować, że w krajach z lepszą służbą zdrowia wykonuje się więcej testów (choć liczba ta nie jest znacząca). Ponadto, jeśli państwo przynależy do Unii Europejskiej, wykonuje ono średnio o w przybliżeniu 90% więcej testów niż w wypadku średniej, ceteris paribus. Statystyka F pokazuje, że model jest dobrze dopasowany.

Zamieszczam dodatkowo diagram macierzy korelacji tychże zmiennych:



Oraz wykres reszt modelu razem z wykresem gęstości:





Wyniki testów:

RESET:

RESET = 0.15187, df1 = 2, df2 = 128, p-value = 0.8593

Jarque-Bera:

Test Jarque-Bera:

X-squared = 2.907, df = 2, p-value = 0.2337

VIF:

log(cases)	fatality_rate	log(GDPpercapita)	PercentGDPforhealthexp
2.297092	1.233531	1.043310	1.083897
IsinEU	log(CO2emissionpercapita)	log(population)	LifeExp
1.455390	2.509390	1.086352	3.094716

Goldfeld-Quandt:

GQ = 0.75939, df1 = 61, df2 = 60, p-value = 0.8567

Breusch-Pagan:

BP = 8.4945, df = 8, p-value = 0.3867

Test White'a:

p=1.096581e-28

Durbin-Watson:

DW = 1.9584, p-value = 0.4049

Ljung-Box:

X-squared = 3.1103, df = 4, p-value = 0.5395

Breusch-Godfrey:

LM test = 3.2284, df = 4, p-value = 0.5204

Wnioski:

W modelu nie występuje problem współliniowości oraz autokorelacji. Reszty mają rozkład normalny, a specyfikacja modelu jest poprawna. Problemem natomiast jest heteroskedastyczność wykazana przez test White'a, co jest popularnym problemem dla badań przekrojowych.

Odporne błędy standardowe oraz ważona MNK:

Test White'a wskazuje na heteroskedastyczność. Aby z nią sobie poradzić, postanowiłem zastosować dwa sposoby. Pierwszym są odporne błędy standardowe. Model wygląda wówczas tak:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.34315786	1.57251222	2.1260	0.0353941 *
log(cases)	0.29785554	0.05899179	5.0491	1.467e-06 ***
fatality_rate	-0.11705437	0.02424690	-4.8276	3.815e-06 ***
log(GDPpercapita)	-0.00058798	0.06885452	-0.0085	0.9931997
PercentGDPforhealthexp	-0.00046618	0.00137322	-0.3395	0.7347972
IsinEU	0.89785118	0.19473823	4.6106	9.480e-06 ***
log(CO2emissionpercapita)	0.40169134	0.09163366	4.3837	2.385e-05 ***
log(population)	-0.08389329	0.04412965	-1.9011	0.0595071 .
LifeExp	0.06870790	0.01989325	3.4538	0.0007466 ***

Zmiany we współczynnikach są nikłe, w zaokrągleniu praktycznie zerowe, tak samo nic się nie zmienia w kontekście istotności zmiennych dla poziomu istotności $\alpha=0.05$.

Zastosowałem także ważone MNK, gdzie jako wagę użyłem wartość odwrotności PKB per capita. Wyniki prezentuję:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.638631	1.099544	2.400	0.017824 *
log(cases)	0.206685	0.051496	4.014	0.000100 ***
fatality_rate	-0.168703	0.020413	-8.264	1.41e-13 ***
log(GDPpercapita)	-0.011077	0.040940	-0.271	0.787145
PercentGDPforhealthexp	-0.001505	0.001748	-0.861	0.391018
IsinEU	0.589203	0.162174	3.633	0.000402 ***
log(CO2emissionpercapita)	0.356760	0.083948	4.250	4.05e-05 ***
log(population)	-0.118221	0.039173	-3.018	0.003063 **
LifeExp	0.098329	0.016564	5.936	2.50e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1283 on 130 degrees of freedom

(73 observations deleted due to missingness)

Multiple R-squared: 0.8109, Adjusted R-squared: 0.7992

F-statistic: 69.68 on 8 and 130 DF, p-value: < 2.2e-16

Jeśli chodzi o istotność zmiennych, istotna stała się zmienna $\log(\text{population})$, która sugeruje, że średnio na każdy przyrost o 1% ilości testów na milion mieszkańców przypada zmniejszenie się ilości mieszkańców o 0.11%, *ceteris paribus*. Na tle poprzednich oszacowań, wzrosły parametry dla średniej długości życia (do 0.09 lat na 1%), zmalały natomiast dla emisji CO2 per capita (do 0.36% na 1%), przynależności do Unii (do 59%), stopy śmiertelności (do -0.16 p.p na 1%) oraz przypadków (do 0.2% na 1%).

Endogeniczność i identyfikowalność:

Problem endogeniczności w modelu może występować w kontekście badania jakości służby zdrowia. Posłużyłem się wprowadzając miarami takimi jak % PKB na służbę zdrowia (co się okazało nieistotne statystycznie) oraz średnia długość życia, ciężko jednak ją obiektywnie zmierzyć w liczbach. Posłużyć się metodą zmiennych instrumentalnych, a jako instrument użyję zmienną *govhealth*, która przytacza wydatki na służbę zdrowia per capita skorygowane o PPP. Nakładam ten instrument na zmienną *LifeExp*.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.4874712	1.3475746	2.588	0.010770 *
$\log(\text{cases})$	0.3153981	0.0564300	5.589	1.31e-07 ***
<i>fatality_rate</i>	-0.1177938	0.0233277	-5.050	1.49e-06 ***
$\log(\text{GDPpercapita})$	0.0053664	0.0808672	0.066	0.947194
<i>PercentGDPforhealthexp</i>	-0.0006237	0.0016480	-0.378	0.705690
<i>IsinEU</i>	0.8934661	0.2315448	3.859	0.000180 ***
$\log(\text{CO2emissionpercapita})$	0.3819379	0.0866682	4.407	2.20e-05 ***
$\log(\text{population})$	-0.0921238	0.0456961	-2.016	0.045892 *
<i>LifeExp</i>	0.0674175	0.0177740	3.793	0.000228 ***

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments ($\log(\text{cases})$)	133	2	5.480e-01	0.8345
Weak instruments (<i>fatality_rate</i>)	133	2	4.748e+00	0.1897
Weak instruments ($\log(\text{GDPpercapita})$)	133	2	1.269e+00	0.5431
Weak instruments (<i>PercentGDPforhealthexp</i>)	133	2	1.222e+00	0.5566
Weak instruments (<i>IsinEU</i>)	133	2	1.109e+30	<2e-16 ***
Weak instruments ($\log(\text{CO2emissionpercapita})$)	133	2	1.765e+00	0.4311
Weak instruments ($\log(\text{population})$)	133	2	2.106e+00	0.3769
Wu-Hausman	2	126	2.988e+00	0.0539 .
Sargan	126	NA	1.307e+02	0.3692

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8949 on 128 degrees of freedom

Multiple R-Squared: 0.7885, Adjusted R-squared: 0.7753

Wald test: 60.04 on 8 and 128 DF, p-value: < 2.2e-16

Jak widać, na poziomie istotności $\alpha=0.05$ zmienne są egzogeniczne (niemniej wartość p jest bardzo bliska granicy istotności, na $\alpha=0.1$ już by była endogeniczna). Natomiast sam instrument koreluje z przynależnością do Unii Europejskiej, wskazywałoby to, że państwa Unii wydają więcej na służbę zdrowia. Pod wpływem instrumentu istotna staje się zmienna $\log(\text{population})$, która sugeruje, że średnio ceteris paribus przy przyroście ilości testów na milion mieszkańców o 1% liczba populacji maleje o 0.09%. Minimalnie zmniejszają się też oszacowania dla zmiennych $CO2emissionpercapita$, $LifeExp$ a zwiększają dla $cases$.

Założmy, że żeby odwzorować jakość służby zdrowia, tworzę model pomocniczy:

$$LifeExp = b_0 + b_1 \text{PercentGDPforhealthexp} + b_2 \text{govhealth} + b_3 \text{fatality_rate} + b_4 \text{tests} + e$$

Gdzie e to składnik losowy.

Oszacowania modelu wyglądają tak:

Call:

`lm(formula = LifeExp ~ PercentGDPforhealthexp + govhealth + fatality_rate + tests)`

Residuals:

Min	1Q	Median	3Q	Max
-18.246	-3.509	1.038	4.369	13.717

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.594e+01	1.607e+00	41.036	< 2e-16 ***
PercentGDPforhealthexp	1.585e-02	1.095e-02	1.448	0.150
govhealth	5.239e-03	1.030e-02	0.509	0.612
fatality_rate	6.592e-01	1.449e-01	4.549	1.17e-05 ***
tests	1.307e-04	1.712e-05	7.634	3.40e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.24 on 138 degrees of freedom

(69 observations deleted due to missingness)

Multiple R-squared: 0.3683, Adjusted R-squared: 0.35

F-statistic: 20.12 on 4 and 138 DF, p-value: 4.527e-13

Jak widzimy, ilość wykonanych testów oraz stopa śmiertelności są wyższe wraz ze wzrostem oczekiwanego życia. Można wywnioskować, że dzięki lepszej diagnostyce więcej śmierci wywołanych koronawirusa jest jako takowa śmierć klasyfikowane. Model jest relatywnie słabo dopasowany - R^2 wynosi 0.35. Spoza zbioru danych można zapewne zdobyć inne zmienne objaśniające stan służby zdrowia w danym kraju (np. odsetek lekarzy).

Zastosuję ten model do zbadania identyfikowalności parametrów oraz do ponownej weryfikacji endogeniczności podwójną metodą najmniejszych kwadratów. Jako instrument użyję wszystkich zmiennych egzogenicznych.

2SLS estimates for 'tests' (equation 1)

Model Formula: $\log(\text{tests}) \sim \log(\text{cases}) + \text{fatality_rate} + \log(\text{GDPpercapita}) + \text{PercentGDPforhealthexp} + \text{IsinEU} + \log(\text{CO2emissionpercapita}) + \log(\text{population}) + \text{LifeExp}$

Instruments: $\sim \log(\text{cases}) + \text{fatality_rate} + \log(\text{GDPpercapita}) + \text{PercentGDPforhealthexp} + \text{IsinEU} + \log(\text{CO2emissionpercapita}) + \log(\text{population}) + \text{govhealth}$

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	21.77192736	13.39882400	1.62491	0.106641
$\log(\text{cases})$	0.65723412	0.27261174	2.41088	0.017335 *
fatality_rate	-0.08007048	0.05044243	-1.58736	0.114898
$\log(\text{GDPpercapita})$	0.03887501	0.15023863	0.25876	0.796240
$\text{PercentGDPforhealthexp}$	0.00544285	0.00527179	1.03245	0.303808
IsinEU	1.87556329	0.81737374	2.29462	0.023383 *
$\log(\text{CO2emissionpercapita})$	1.13955924	0.55664363	2.04720	0.042685 *
$\log(\text{population})$	-0.04828654	0.08725447	-0.55340	0.580956
LifeExp	-0.24107695	0.22347630	-1.07876	0.282724

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.644474 on 128 degrees of freedom

Number of observations: 137 Degrees of Freedom: 128

SSR: 346.149924 MSE: 2.704296 Root MSE: 1.644474

Multiple R-Squared: 0.285867 Adjusted R-Squared: 0.241233

2SLS estimates for 'health' (equation 2)

Model Formula: $\text{LifeExp} \sim \text{PercentGDPforhealthexp} + \text{govhealth} + \text{fatality_rate} + \text{tests}$

Instruments: $\sim \log(\text{cases}) + \text{fatality_rate} + \log(\text{GDPpercapita}) + \text{PercentGDPforhealthexp} + \text{IsinEU} + \log(\text{CO2emissionpercapita}) + \log(\text{population}) + \text{govhealth}$

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.32372e+01	1.97199e+00	32.06762	< 2.22e-16 ***
$\text{PercentGDPforhealthexp}$	1.71738e-02	1.30602e-02	1.31497	0.19080
govhealth	3.45112e-03	1.22117e-02	0.28261	0.77792
fatality_rate	6.90589e-01	1.71170e-01	4.03452	9.2112e-05 ***
tests	2.57933e-04	2.96643e-05	8.69508	1.1990e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.278614 on 132 degrees of freedom

Number of observations: 137 Degrees of Freedom: 132

SSR: 6993.12556 MSE: 52.978224 Root MSE: 7.278614
Multiple R-Squared: 0.106331 Adjusted R-Squared: 0.07925

Obserwujemy bardzo duży spadek R^2 w podwójnej MNK - jedyne zmienne dotyczące ilości przypadków, przynależności do UE oraz emisji CO2 per capita poddają się interpretacji. W drugim równaniu natomiast również odnotowujemy spadek R^2 . Czy w takim razie służba zdrowia ma istotny wpływ na ilość testów? Najpewniej tak - niemniej potrzebowalibyśmy więcej danych, by ten temat dogłębniej zbadać.

Co z identyfikowalnością? Model spełnia warunek konieczny, jest jednak najpewniej nadmiernie identyfikowalny, bowiem różnica liczby zmiennych egzogenicznych z pierwszego równania oraz drugiego przekracza liczbę zmiennych endogenicznych.

Analizy mnożnikowej, z racji na przekrojowy charakter badania, nie stosowałem.

Podsumowanie:

Z oszacowanych modeli wynika, że na poziomie $\alpha=0.05$ zachodzi istotna korelacja między ilością testów na milion mieszkańców a ilością przypadków na milion mieszkańców, stopą śmiertelności, przynależnością kraju do Unii Europejskiej, emisji CO2 per capita oraz średnią długością życia, oraz najpewniej także ilością populacji. Co z tego idzie wywnioskować:

- im więcej testów na milion mieszkańców, tym mniejsza stopa śmiertelności, co sugeruje, że lepsza diagnostyka wpływa na zmniejszenie śmiertelności - ma to odzwierciedlenie w specyfice choroby COVID-19, która często przebiega bezobjawowo bądź z nikłymi objawami oraz w tym, że lepsza diagnostyka pozwala na bardziej efektywne działanie
- jeśli kraj przynależy do Unii Europejskiej, to wykonuje znacząco więcej testów na milion mieszkańców - jest to logiczne, kraje Unii Europejskiej na tle świata są krajami rozwiniętymi i mogą sobie pozwolić na lepszą diagnostykę
- im więcej kraj generuje CO2 per capita, tym więcej wykonuje testów na milion mieszkańców - jeśli uznać zmienną *CO2emissionpercapita* za skorelowaną z szeroko pojętą zamożnością państwa, rozwojem gospodarczym, również wydaje się to logiczne, że lepiej rozwinięte państwa będą przeprowadzać lepszą diagnostykę
- im większa średnia długość życia, tym więcej testów na milion mieszkańców - również jest to logiczne, kraje lepiej rozwinięte mają widocznie większą średnią długość życia i to one mogą sobie pozwolić na więcej testów, mają także z reguły lepiej rozwiniętą służbę zdrowia
- najprawdopodobniej im mniejsza populacja, tym więcej kraj wykonuje testów na milion mieszkańców - może to mieć uzasadnienie w tym, że np. mniejsze państwa są bardziej zorganizowane i mają lepsze dojścia do swoich obywateli (np. Islandia) oraz tym, że wiele bardzo zaludnionych państw, takich jak Indie czy Pakistan nie ma dostępu do najlepszej diagnostyki