

**POLITECHNIKA LUBELSKA**

**WYDZIAŁ PODSTAW TECHNIKI**

*Kierunek: MATEMATYKA*



**Praca inżynierska**

Zastosowanie modeli mieszanych w analizie rozwoju pandemii  
wywołanej wirusem Covid-19 na świecie

*The use of mixed-effects models in the analysis of the Covid-19  
pandemic in the world*

*Praca wykonana pod kierunkiem:*  
dra Dariusza Majerka

*Autor:*  
Alicja Hołowiecka  
nr albumu: 89892

**Lublin 2020**



## Spis treści

<b>Wstęp</b> . . . . .	5
<b>Rozdział 1. Teoretyczne podstawy badań własnych</b> . . . . .	7
1.1. Modele liniowe . . . . .	7
1.1.1. Metody estymacji parametrów modelu liniowego . . . . .	7
1.1.2. Badanie istotności parametrów . . . . .	8
1.2. Modele mieszane . . . . .	8
1.2.1. Metody estymacji . . . . .	9
1.2.2. Badanie istotności parametrów . . . . .	9
<b>Rozdział 2. Badania własne</b> . . . . .	11
2.1. Zbiór danych i jego wstępne przygotowanie . . . . .	11
2.2. Dyskusja wyników . . . . .	12
2.2.1. Model 1 . . . . .	12
2.2.2. Model 2 . . . . .	13
2.2.3. Model 3 . . . . .	13
2.2.4. Model 4 . . . . .	14
2.2.5. Model 5 . . . . .	15
2.2.6. Model 6 . . . . .	16
2.2.7. Model 7 . . . . .	16
2.2.8. Model 8 . . . . .	17
2.2.9. Model 9 . . . . .	18
2.2.10. Model 10 . . . . .	19
<b>Podsumowanie i wnioski</b> . . . . .	21
<b>Bibliografia</b> . . . . .	23
<b>Spis rysunków</b> . . . . .	25
<b>Spis tabel</b> . . . . .	27
<b>Załączniki</b> . . . . .	29
<b>Streszczenie (Summary)</b> . . . . .	31



## Wstęp

Pandemia choroby COVID-19 jest wydarzeniem, które wstrząsnęło całym światem w roku 2020. Właściwie nikt chyba nie może powiedzieć, że nie poczuł się dotknięty przez sytuację związaną z rozprzestrzenianiem się wirusa. Pierwsze przypadki pojawiły się pod koniec 2019 roku we wschodnich Chinach, w mieście Wuhan. Na początku 2020 roku chorowali już obywatele większości państw na świecie. Na moment pisania tej pracy, sytuacja nadal nie jest opanowana i nie wiadomo, jak się rozwinie.

Biorąc to pod uwagę, tym ważniejszy wydaje się temat poruszany w tej pracy. Wiele jednostek naukowych podejmuje próby znalezienia odpowiedniego modelu, aby przewidzieć rozwój pandemii. Przedstawione w tej pracy modele mieszane co prawda nie pozwalają na dokładną predykcję, ale są dobrym narzędziem, aby odkryć, które czynniki mają wpływ na rozwój pandemii w przeciętnym kraju.



## Rozdział 1

### Teoretyczne podstawy badań własnych

W tej części pracy przedstawimy metody matematyczne, które zostaną użyte w części praktycznej tej pracy. Zgodnie z tematem, będą to głównie modele mieszane.

#### 1.1. Modele liniowe

Na początek przypomnimy podstawowe wiadomości o modelach liniowych.

Model regresji prostej ma postać

$$y = x\beta_1 + \beta_0 + \varepsilon$$

gdzie oszacowania parametrów  $\beta_1$ ,  $\beta_0$  obliczamy następująco:

$$\hat{\beta}_1 = \frac{Cov(x, y)}{Var(x)},$$
$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1.$$

Model interpretujemy w ten sposób, że jeżeli zmienna  $x$  wzrośnie o 1, to zmienna  $y$  zmieni się o  $\beta_1$ .

##### 1.1.1. Metody estymacji parametrów modelu liniowego

1. Metoda najmniejszych kwadratów, OLS (ang. *Ordinary Least Squares*) - w metodzie tej minimalizujemy błąd kwadratowy, czyli sumę kwadratów reszt, którą oznaczamy RSS (ang. *Residual Sum of Squares*).

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Twierdzenie Gaussa-Markowa: taki estymator jest BLUE (Best Linear Unbiased Estimator), przy odpowiednich założeniach.

2. Metoda największej wiarygodności, ML (ang. *Maximum Likelihood*) polega na maksymalizacji wartości funkcji prawdopodobieństwa ze względu na  $\beta$  (w praktyce maksymalizujemy zwykle logarytm z tej funkcji)

$$\hat{\sigma}_{ML}^2 = RSS/n$$

Estymując  $\sigma^2$ , maksymalizujemy funkcję wiarygodności zarówno ze względu na  $\beta$ , jak i  $\sigma^2$ .

Estymatory uzyskane tą metodą są asymptotycznie nieobciążone.

3. Resztowa metoda największej wiarygodności, REML (ang. *Residual/Restricted Maximum Likelihood Method*) - z estymacji parametru  $\sigma^2$  usuwamy wpływ parametrów zakłócających  $\beta$ .

$$\hat{\sigma}_{REML}^2 = RSS/(n - p)$$

Estymatory uzyskane tą metodą są nieobciążone [1].

### 1.1.2. Badanie istotności parametrów

$$H_0 : \beta_i = 0$$

## 1.2. Modele mieszane

W powyżej opisanych modelach liniowych z efektami stałymi zakładamy niezależność kolejnych pomiarów, dlatego nie są to odpowiednie modele, kiedy mamy np. kilka pomiarów dla pojedynczego elementu. W takim przypadku możemy użyć modeli liniowych z efektami mieszanymi (stałymi i losowymi), które krótko nazywamy modelami mieszanymi.

Modeli mieszanych używamy w przypadku powtarzanych pomiarów bądź w przypadku hierarchicznej lub zagnieżdżonej struktury. Takie dane charakteryzują się korelacją między obserwacjami z tej samej grupy, co nie pozwala na użycie modelu liniowego z efektami stałymi. Dlatego do modelu wprowadza się czynnik losowy.

Czynnik stały jest pewnym parametrem, którego wartość estymujemy na podstawie próbki, natomiast czynnik losowy jest zmienną losową, dla której próbujemy oszacować parametry jej rozkładu [2].

Przykładową sytuacją, gdzie możemy użyć modelu mieszanego, jest badanie działania leku na grupie pacjentów, gdzie dokonujemy kilku pomiarów na danym pacjencie. W tym przypadku nie interesuje nas konkretny pacjent, ale raczej wpływ leku na przeciętnego pacjenta. Dodatkowo, traktujemy pacjentów jako losowo wybranych. Podejście modelu mieszanego będzie polegało na potraktowaniu wpływu pacjenta jako czynnik zakłócający.



Rozważamy model postaci

$$y = X\beta + Zu + \varepsilon$$

gdzie  $X$  - macierz zmiennych będących efektami stałymi,  $Z$  - macierz zmiennych będących efektami losowymi,  $\beta$  to wektor nieznanych efektów stałych,  $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_{n \times n})$  to zakłócenie losowe, a  $u \sim \mathcal{N}(0, \sigma^2 D)$  to wektor zmiennych losowych odpowiadających efektom losowym [1].

Znając  $D$ , możemy estymować parametry  $\beta$  uogólnioną metodą najmniejszych kwadratów. Do estymowania nieznanego  $D$  możemy użyć np. metodą największej wiarygodności.

### 1.2.1. Metody estymacji

Do oceny wartości parametrów modelu mieszanego można stosować metody ML (Największej Wiarygodności) oraz REML (Resztowej Największej Wiarygodności), wspomniane w tej pracy przy okazji modeli liniowych. W przypadku modeli mieszanych obydwojema metodami możemy uzyskać estymatory obciążone, ale to obciążenie jest zazwyczaj mniejsze w przypadku estymatorów uzyskanych metodą REML.

Różnica między metodą REML i ML polega na tym, że w metodzie REML najpierw usuwamy wpływ efektów stałych.

### 1.2.2. Badanie istotności parametrów

$$H_0 : \sigma_j^2 = 0$$

Te same metody co dla efektów stałych



## Rozdział 2

### Badania własne

#### 2.1. Zbiór danych i jego wstępne przygotowanie

Zbiór danych pochodzi z witryny internetowej Our World In Data [3], gdzie dane zostały zebrane z różnych źródeł, m. in. ze Światowej Organizacji Zdrowia (WHO) oraz Europejskiego Centrum ds. Zapobiegania i Kontroli Chorób (ECDC). W zbiorze znajduje się 210 krajów, dane dotyczące terytoriów międzynarodowych oraz łącznie dla całego świata. Mamy ponad 40 kolumn z różnymi parametrami - w dalszej części pracy opiszemy, które zmienne będą przez nas użyte.

W zbiorze znajdowało się wiele braków danych. Dla każdego kraju zostały usunięte dane sprzed rozpoczęcia się epidemii na jego terytorium (`total cases=0`), dni są numerowane kolejnymi liczbami całkowitymi.

Ze zbioru danych zostały usunięte wszystkie kraje o populacji poniżej miliona mieszkańców, ponieważ w większości były to nieduże wysepki, dla których dane były wybrakowane. Oprócz tego, kilka innych krajów zostało usuniętych, ponieważ mimo większej populacji, dane były niepełne.

Do formułowania hipotez i budowania modeli będziemy się posługiwać następującymi zmiennymi:

- liczba zachorowań - jest to liczba potwierdzonych przypadków koronawirusa w danym kraju od momentu rozpoczęcia epidemii. Zamiast wartości liczby zachorowań, będziemy używać liczby zachorowań na milion mieszkańców (`total cases per million`),
- liczba wykonanych testów - będziemy używać liczby wykonanych testów w przeliczeniu na tysiąc mieszkańców danego kraju (`total tests per thousand`),
- wskaźnik siły obostrzeń (`sringency index`) - wskaźnik tego, jak silne obostrzenia wprowadził rząd danego kraju. Jest to kombinacja dziewięciu innych zmiennych, m.in. zamykanie szkół, polityka wykonywania testów, ograniczenie kontaktów międzyludzkich itp. Może przyjmować wartości od 0 do 100, im większa wartość, tym silniejsze obostrzenia w danym kraju [4],

- gęstość zaludnienia (**population density**),
- PKP danego kraju na osobę (**GDP per capita**) - Produkt Krajowy Brutto, przeliczony na hipotetyczną walutę dolara międzynarodowego [5],
- część społeczeństwa żyjąca w skrajnym ubóstwie (**extreme poverty**)
- śmiertelność z powodu chorób sercowych (**cardiovasc death rate**) - stan na rok 2017
- powszechność występowania cukrzycy (**diabetes prevalence**) - odsetek populacji z cukrzycą, brane pod uwagę są osoby w wieku od 20 do 70 lat, stan na rok 2017
- oczekiwana długość życia (**life expectancy**) - kraje zostaną podzielone na kategorie ze względu na tą zmienną, wyróżnimy kraje, w których oczekiwana długość życia jest poniżej 50 lat, między 50 a 54, między 55 a 59 i tak dalej aż do grupy krajów z oczekiwaną długością życia powyżej 80 lat.

## 2.2. Dyskusja wyników

### 2.2.1. Model 1

Hipoteza 1: Wpływ kraju (efektu losowego) jest większy niż wpływ czasu (czynnika stałego) w modelu mieszanym.

Pierwszy model ma postać

```
mod <- lme(total_cases_per_million~time,
random = ~1|location,
data = covid)
```

a więc przedstawia zależność liczby zachorowań od czasu, a kraj jest efektem losowym.

Dla efektu losowego otrzymujemy następujący wynik:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3241.434 4262.032

Widać zatem, że efekt losowy jest odpowiedzialny za około 45% wariancji całego modelu.

Dla efektów stałych mamy następujący wynik:

Fixed effects: total\_cases\_per\_million ~ time

	Value	Std.Error	DF	t-value	p-value
--	-------	-----------	----	---------	---------

(Intercept)	763.1050	225.19114	51584	3.38870	7e-04
-------------	----------	-----------	-------	---------	-------

```
time          15.7823    0.25565 51584 61.73324    0e+00
```

A więc zarówno wyraz wolny, jak i współczynnik przy zmiennej Czas, są istotne statystycznie. Dodatkowo, korelacja pomiędzy liczbą zachorowań a czasem jest dodatnia, więc wraz z upływem czasu liczba zachorowań rośnie dla przeciętnego kraju.

### 2.2.2. Model 2

Hipoteza 2: Liczba wykonywanych testów na COVID-19 ma związek z liczbą zachorowań.

Drugi model to:

```
mod1 <- lme(total_cases_per_million~time+total_tests_per_thousand,
random = ~1|location,
data = covid_na)
```

Badamy tutaj, czy czas oraz liczba wykonywanych testów mają wpływ na liczbę zachorowań, jeżeli kraj traktujemy jako czynnik losowy.

Dla tego modelu otrzymujemy następujące wyniki:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3677.593 2901.592

Fixed effects: total\_cases\_per\_million ~ time + total\_tests\_per\_thousand

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-558.9119	384.0912	17724	-1.45515	0.1456
time	17.5806	0.4263	17724	41.23543	0.0000
total_tests_per_thousand	22.5888	0.3550	17724	63.63620	0.0000

Widać po pierwsze, że efekt losowy jest odpowiedzialny za ponad połowę zmienności modelu. Po drugie, widać, że oba efekty stałe są istotne statystycznie, i oba mają wpływ stymulujący na liczbę zachorowań.

### 2.2.3. Model 3

Hipoteza 3: kraje o różnej oczekiwanej długości życia różnią się liczbą zachorowań.

Trzeci model wygląda następująco:

```
mod2 <- lme(total_cases_per_million~time+age,
```

```
random=~1|location,  
data=covid)
```

Prezentuje on zależność liczby zachorowań od czasu i od oczekiwanej długości życia w danym kraju. Występuje także efekt losowy kraju.

Dla modelu trzeciego otrzymujemy następujące wyniki:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3128.64 2993.161

Fixed effects: total\_cases\_per\_million ~ time + age

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-2608.462	1280.159	35550	-2.03761	0.0416
time	28.980	0.231	35550	125.46483	0.0000
age60-64	-69.561	1430.972	145	-0.04861	0.9613
age65-69	-324.578	1477.891	145	-0.21962	0.8265
age70-74	1172.910	1406.069	145	0.83418	0.4056
age75-79	2617.920	1368.246	145	1.91334	0.0577
age80 and above	3195.777	1410.326	145	2.26598	0.0249
agebelow 55	-184.197	1898.558	145	-0.09702	0.9228

Ponownie efekt losowy odpowiada za większą część wariancji. Czas ponownie jest istotny i ma wpływ stymulujący. Dla grupy wiekowej 75-79 różnica w średniej liczbie zachorowań jest na granicy istotności statystycznej. Dopiero dla krajów o oczekiwanej długości życia powyżej 80 lat pojawia się istotna różnica - liczba zachorowań w tych krajach jest największa. W pozostałych grupach wiekowych nie można mówić o istotnych różnicach.

#### 2.2.4. Model 4

Hipoteza 4: Kraje o różnej gęstości zaludnienia różnią się liczbą zachorowań.

W czwartym modelu badamy zależność liczby zachorowań od czasu i gęstości zaludnienia, a kraj jest czynnikiem losowym.

```
mod3 <- lme(total_cases_per_million~time+population_density,  
random=~1|location,  
data = covid)
```

Wyniki są następujące:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3339.163 2993.161

Fixed effects: total\_cases\_per\_million ~ time + population\_density

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1256.0762	283.81733	35550	-4.42565	0.0000
time	28.9851	0.23098	35550	125.48973	0.0000
population_density	0.6593	0.40726	150	1.61895	0.1076

Zatem gęstość zaludnienia nie jest czynnikiem istotnie różnicującym liczbę zachorowań w krajach.

### 2.2.5. Model 5

Hipoteza 5: Kraje różniące się siłą obostrzeń mają istotne różnice w liczbie zachorowań.

Piaty model ma następującą postać:

```
covid_si <- drop_na(covid, stringency_index)
mod4 <- lme(total_cases_per_million~time+stringency_index,
random=~1|location,
data = covid_si)
```

W tym modelu sprawdzamy zależność liczby zachorowań od czasu i siły obostrzeń, kraj jest czynnikiem losowym.

Otrzymujemy następujące wyniki:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3172.804 2854.192

Fixed effects: total\_cases\_per\_million ~ time + stringency\_index

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1710.7939	269.33057	33756	-6.35202	0
time	27.9374	0.23445	33756	119.16023	0
stringency_index	10.1151	0.89377	33756	11.31725	0

Tak jak w poprzednich modelach, czynnik losowy odpowiada za największą część zmienności. Czas jest istotny statystycznie. Wskaźnik siły obostrzeń także jest istotny i ma wpływ stymulujący, co oznaczałoby, że im silniejsze obostrzenia, tym więcej zachorowań. Ta interpretacja prawdopodobnie jest niepoprawna, można się domyślać, że raczej zachodzi odwrotna zależność - w krajach z największą liczbą zachorowań są wprowadzane najsurowsze obostrzenia.

### 2.2.6. Model 6

Hipoteza 6: Kraje o różnej wysokości wskaźnika rozwoju społecznego (HDI) różnią się liczbą zachorowań.

Szósty model przedstawia zależność liczby zachorowań od czasu i wskaźnika rozwoju społecznego, a kraj jest czynnikiem losowym.

```
covid_hdi <- drop_na(covid, human_development_index)
mod5 <- lme(total_cases_per_million~time+human_development_index,
random=~1|location,
covid_hdi)
```

Z tego modelu mamy następujący wynik:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3130.083 2984.162

Fixed effects: total\_cases\_per\_million ~ time + human\_development\_index

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-6632.736	1173.8596	35117	-5.65037	0
time	28.795	0.2315	35117	124.40143	0
human_development_index	7807.594	1623.1648	148	4.81011	0

Efekt losowy jest odpowiedzialny za nieznacznie większą część wariancji niż wszystkie pozostałe czynniki. Efekty stałe są istotne i oba mają wpływ stymulujący. W krajach z wyższym wskaźnikiem rozwoju społecznego, zachorowań jest znacząco więcej.

### 2.2.7. Model 7

Hipoteza 7: Kraje o różnej wysokości odsetka śmierci z powodu chorób sercowych różnią się liczbą zachorowań.



Model siódmy wygląda następująco:

```
mod6 <- lme(total_cases_per_million~time+cardiovasc_death_rate,
random=~1|location,
data= covid)
```

i oprócz zależności liczby zachorowań od czasu zawiera także zależność od odsetka śmierci spowodowanych chorobami sercowymi. Kraj jest traktowany jako efekt losowy.

Otrzymujemy następujące podsumowanie:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 3241.687 2993.161

Fixed effects: total\_cases\_per\_million ~ time + cardiovasc\_death\_rate

	Value	Std.Error	DF	t-value	p-value
(Intercept)	830.9013	626.6099	35550	1.32603	0.1848
time	28.9831	0.2310	35550	125.48108	0.0000
cardiovasc_death_rate	-7.4612	2.1624	150	-3.45041	0.0007

Czynnik losowy zachowuje się podobnie jak we wszystkich poprzednich modelach. Oba czynniki stałe są istotne. Co ciekawe, odsetek śmierci spowodowanych chorobami serca wpływa ograniczająco na liczbę zachorowań. Może to być związane z tym, że osoby chore na serce bardziej uważają, aby się nie zarazić, tym samym zmniejszają liczbę zachorowań w danym kraju.

### 2.2.8. Model 8

Hipoteza 8: Kraje o różnej wysokości odsetka osób chorych na cukrzycę różnią się liczbą zachorowań.

```
mod7 <- lme(total_cases_per_million~time+diabetes_prevalence,
random=~1|location,
data= covid)
```

Model ten jest analogiczny do poprzedniego, z tym że zamiast chorób sercowych mamy tu odsetek chorych na cukrzycę.

Otrzymujemy następujący wynik:

Random effects:

Formula: ~1 | location

```
(Intercept) Residual
StdDev:      3173.523 2993.161
```

```
Fixed effects: total_cases_per_million ~ time + diabetes_prevalence
```

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-3320.595	566.8844	35550	-5.85762	0
time	28.984	0.2310	35550	125.48575	0
diabetes_prevalence	300.581	69.1237	150	4.34845	0

Czynnik losowy ma taką samą istotność jak poprzednio, czas także. Rozpowszechnienie cukrzycy wpływa stymulująco na liczbę zachorowań. Może to być spowodowane tym, że osoby chore na cukrzycę mają słabszy organizm i są bardziej narażone na zakażenie.

### 2.2.9. Model 9

Hipoteza 9: Kraje o różnej wysokości odsetka osób żyjących w skrajnej biedzie różnią się liczbą zachorowań.

W tym modelu sprawdzamy zależność liczby zachorowań od czasu i odsetka osób żyjących w skrajnym ubóstwie. Kraj jest czynnikiem losowym.

```
covid_ep <- drop_na(covid, extreme_poverty)
mod8 <- lme(total_cases_per_million~time+extreme_poverty,
random=~1|location,
data= covid_ep)
```

Model ten ma następujące podsumowanie:

```
Random effects:
Formula: ~1 | location
(Intercept) Residual
StdDev:      2418.049 2562.526
```

```
Fixed effects: total_cases_per_million ~ time + extreme_poverty
```

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-479.1335	282.20423	25650	-1.69783	0.0896
time	26.5944	0.23335	25650	113.96617	0.0000
extreme_poverty	-42.5325	10.97942	108	-3.87384	0.0002

Czynnik losowy jest odpowiedzialny za niecałą połowę zmienności. Oba czynniki stałe są istotne. Im większy jest w danym kraju odsetek osób żyjących w biedzie, tym wyższa liczba zachorowań (miejmy nadzieję, że nasz rząd tego nie usłyszy).

### 2.2.10. Model 10

Hipoteza 10: Kraje o różnej wysokości PKB różnią się liczbą zachorowań.

W modelu dziesiątym pojawia się zależność liczby zachorowań od czasu i PKB. Kraj jest efektem losowym.

```
covid_gdp <- drop_na(covid, gdp_per_capita)
mod9 <- lme(total_cases_per_million~time+gdp_per_capita,
random=~1|location,
data= covid_gdp)
```

Wyniki są następujące:

Random effects:

Formula: ~1 | location

(Intercept) Residual

StdDev: 2838.031 3005.475

Fixed effects: total\_cases\_per\_million ~ time + gdp\_per\_capita

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-2892.4035	322.9109	35104	-8.95728	0
time	29.2840	0.2332	35104	125.55885	0
gdp_per_capita	0.0963	0.0122	148	7.87627	0

Efekt losowy odpowiada za niecałą połowę zmienności modelu. Efekty stałe są istotne statystycznie. Im wyższe PKB danego kraju, tym wyższa liczba zachorowań.



## Podsumowanie i wnioski

Cecha	Wpływ na liczbę zachorowań
Czas	istotny, wraz z upływem czasu rośnie liczba zachorowań
Liczba wykonywanych testów na COVID-19	istotny, wraz ze wzrostem liczby testów rośnie liczba zachorowań
Oczekiwana długość życia	istotny, o ile ta wartość przekracza 80 lat, wówczas zachorowań jest więcej niż dla krajów o krótszej oczekiwanej długości życia
Gęstość zaludnienia	nieistotny
Wskaźnik siły obostrzeń	istotny, jest wysoki w krajach o dużej liczbie zachorowań
Wskaźnik rozwoju społecznego	istotny, w krajach o wysokim wskaźniku rozwoju jest więcej zachorowań
Śmiertelność z powodu chorób sercowych	istotny, im wyższy jest ten współczynnik, tym mniej zachorowań na COVID-19
Powszechność występowania cukrzycy	istotny, im wyższy jest ten współczynnik, tym więcej zachorowań na COVID-19
Część populacji żyjąca w skrajnym ubóstwie	istotny, im większa jest część mieszkańców żyjąca w biedzie, tym mniej zachorowań
PKB na osobę	istotny, im wyższe PKB, tym więcej zachorowań

**Tabela 2.1:** Porównanie istotności i wpływu różnych czynników na liczbę zachorowań w przeciętnym kraju

Indeks Akaike jest miarą utraconej informacji. Im mniejszy jest indeks Akaike, tym lepiej model wyjaśnia badane zjawisko. Widzimy z tabeli 2.2, że najmniejsze AIC występuje dla modelu nr 2. Jest to model, gdzie występuje zależność między liczbą zachorowań a liczbą wykonywanych testów. Ta zależność jest bardzo oczywista, ponie-

Nr modelu	AIC
1	1013843.9
2	335245.8
3	673611.1
4	673718.5
5	636599.2
6	665261.9
7	673706.2
8	673692.9
9	478066.0
10	665511.1

**Tabela 2.2:** Porównanie indeksów Akaike dla poszczególnych modeli

waż liczbę zachorowań zlicza się na podstawie tego, ile testów dało wynik pozytywny. Modelem z drugim najniższym AIC jest model nr 9, gdzie badamy istotność wskaźnika części populacji żyjącej w skrajnym ubóstwie.

Wszystkie modele jednoznacznie pokazują, że efekt kraju jako czynnika zakłócającego jest bardzo istotny, w wielu przypadkach bardziej niż jakikolwiek inny czynnik stały (np. czas).

Na to, co jest nazywane w tej pracy „efektem kraju”, składa się tak naprawdę wiele innych czynników, m. in. gęstość zaludnienia, sytuacja ekonomiczna danego kraju, odsetek osób z chorobami towarzyszącymi, rozkład wieku, jak również przyjęta strategia walki z koronawirusem, na którą z kolei składają się m. in. liczba wykonywanych testów, przepisy w sprawie zamykania szkół, miejsc publicznych, ograniczenie kontaktów międzyludzkich, i wiele innych.

W mojej pracy nie zajmowałam się badaniem, w jaki sposób te czynniki wpływają na wzrost lub spadek liczby zachorowań, chcę jedynie zasygnalizować, że mogą być istotne, skoro wykazany został wpływ efektu kraju na liczbę zachorowań.

## Bibliografia

- [1] Przemysław Biecek, *Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi*, Wydawnictwo Naukowe PWN, Wydanie II, Warszawa 2013
- [2] Julian J. Faraway, *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models. Second Edition*, CRC Press Taylor & Francis Group, 2016
- [3] <https://ourworldindata.org/coronavirus>
- [4] <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker> (dostęp 31.10.2020)
- [5] <https://ourworldindata.org/what-are-ppps> (dostęp 31.10.2020)





## Spis rysunków



## Spis tabel

2.1	Porównanie istotności i wpływu różnych czynników na liczbę zachorowań w przeciętnym kraju . . . . .	21
2.2	Porównanie indeksów Akaike dla poszczególnych modeli . . . . .	22



## **Załączniki**

1. Płyta CD z niniejszą pracą w wersji elektronicznej.



## Streszczenie (Summary)

### **Zastosowanie modeli mieszanych w analizie rozwoju pandemii wywołanej wirusem Covid-19 na świecie**

W tej pracy przedstawione są pojęcia związane z modelami liniowymi z efektami stałymi i losowymi. Następnie opisane są badania własne na zbiorze danych dotyczącym rozprzestrzeniania się choroby COVID-19 w różnych krajach na świecie.

### ***The use of mixed-effects models in the analysis of the Covid-19 pandemic in the world***

*In this paper, concepts related to linear models with fixed and random effects are presented. Then, our own research is described on the dataset on the spread of COVID-19 in various countries around the world.*