Uniwersytet Warszawski

Wydział Nauk Ekonomicznych

Jarosław Leski
Nr albumu: 411174

Czy zamożne państwa dają większe szanse na przeżycie porodu? Determinanty śmiertelności wśród noworodków.

Praca zaliczeniowa
na ćwiczenia z Ekonometrii
prowadzone przez dr Olgę Zajkowską

Warszawa, styczeń 2021 r.

Spis treści

Abstrakt

Niniejsza praca wyjaśnia determinanty śmiertelności wśród noworodków w roku 2018 na próbce danych dla państw z różnych części świata. Na podstawie informacji pochodzących z bazy danych WorldBank-u za pomocą klasycznego modelu regresji liniowej dokonano analizy wpływu, a także potwierdzono zależność między czynnikami takimi jak: PKB per capita, populacja, wydatki rządowe na ochronę zdrowia, liczba turystów, problem wirusa HIV, a poziomem śmiertelności wśród noworodków. Model wyjaśnia badane zjawisko w 83%.

Słowa kluczowe

Śmiertelność, noworodki, służba zdrowia, dane makroekonomiczne, turystyka, regresja liniowa

Rozdział I Wstęp

Śmiertelność noworodków jest niewątpliwie jednym z tych wskaźników, którego wartość państwo chciałoby utrzymywać na jak najniższym poziomie. W praktyce tak się jednak nie dzieje, a w wielu krajach odsetek ten jest daleki od ideału. Od czego jednak zależy śmiertelność wśród noworodków? Czy poziom wskaźnika jest kształtowany wyłącznie poprzez zasoby i wiedzę medyczną? Celem niniejszej pracy było wyjaśnienie śmiertelności wśród noworodków za pomocą klasycznego modelu regresji liniowej. Przy wyborze zmiennych do modelu nacisk położony został na zmienne makroekonomiczne, w tym te związane ze zdrowiem. Po dokonaniu przeglądu literatury rozszerzono dotychczasowe odkrycia naukowe o zmienne do tej pory nieuwzględniane w badaniach. Za pomocą skonstruowanego modelu KMRL za pomocą metody ‘od ogółu do szczegółu’ udało się wyjaśnić wpływ szeregu czynników na badane zjawisko. Model ujęty w pracy przeszedł pozytywnie wszystkie testy diagnostyczne, wyjaśniając rzeczywistość w 83%.

Rozdział II Przegląd literatury

Nie powinno nikogo dziwić, że zagadnienie śmiertelności wśród noworodków poruszone zostało w wielu badaniach naukowych na przestrzeni ostatnich kilkunastu lat.

Sanjay Budhdeo, Johnathan Watkins, Rifat Atun, Callum Williams, Thomas Zeltner, Mahiben Maruthappu w swoim artykule *Changes in government spending on healthcare and population mortality in the European union, 1995–2010: a cross-sectional ecological study* (2015) wykazali powiązanie wydatków na opiekę zdrowotną ze zmianami wszystkich wskaźników śmiertelności takich jak: śmiertelność noworodków, śmiertelność poporodowa, śmiertelność dzieci w wieku do 5 lat oraz śmiertelność dorosłych kobiet i mężczyzn. Ekonomiści dowiedli także, że skutki spadku poziomu wydatków na opiekę zdrowotną widoczne były przez co najmniej 5 lat we wszystkich wyżej wymienionych czynnikach, w tym w śmiertelności noworodków.

Mahiben Maruthappu, Joseph Shalhoub, Zoon Tariq, Callum Williams, Rifat Atun, Alun H. Davies, Thomas Zeltne w artykule *Unemployment, Government Healthcare Spending, and Cerebrovascular Mortality, Worldwide 1981–2009: An Ecological Study* wykazali istotną statystycznie zależność między wzrostem wydatków rządowych na opiekę zdrowia, a spadkiem liczby zgonów spowodowanych uszkodzeniem naczyń mózgowych. Zależność ta utrzymywała się przez okres dwóch lat. W badaniu naukowców nie stwierdzono istotnej statystycznie zależności pomiędzy wzrostem wydatków rządowych, a poziomem PKB per capita, inflacji, stóp procentowych czy urbanizacji.

M Maruthappu, C Williams, R Atun, P Agrawal, T Zeltner w artykule *The association between government healthcare spending and maternal mortality in the European Union, 1981–*

2010: a retrospective study wykazali na poziomie istotności $\alpha = 0{,}05$ statycznie istotną zależność między spadkiem wydatków rządowych na ochronę zdrowia, a wzrostem współczynnika śmiertelności matek. Oszacowali, że spadek GHS o 1% skutkuje zgonem 89 kobiet w krajach Unii Europejskiej, a skutki cięć budżetowych w tym sektorze miały widoczne skutki przez okres jednego roku.

We wszystkich wspomnianych wyżej artykułach, autorzy skupiali swoją uwagę na zbadaniu wpływu zmian w wydatkach rządowych na inne czynniki, głównie śmiertelności. Poniższy model odpowie natomiast na pytanie o konkretne powody takiego, a nie innego poziomu śmiertelności wśród noworodków.

Rozdział III Hipotezy badawcze

Hipoteza 1: Nominalna wartość PKB jest istotna statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 2: Poziom PKB per capita jest istotny statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 3: Populacja danego kraju jest istotna statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 4: Wydatki państwa na ochronę zdrowia są zmienną istotną statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 5: Liczba turystów odwiedzających dane państwo jest istotna statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 6: Odsetek ludzi zamieszkujących tereny miejskie jest istotny statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 7: Liczba zgonów na 1000 mieszkańców jest istotna statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 8: Zasób siły roboczej jest istotny statystycznie i ma wpływ na śmiertelność wśród noworodków

Hipoteza 9: Problem z chorobą AIDS wywołaną przez wirusa HIV w danym kraju jest zmienną istotną statystycznie i ma wpływ na śmiertelność wśród noworodków

Rozdział IV Baza danych

4.1. Przegląd bazy danych wykorzystanej w modelu

Baza danych użyta w modelu została zbudowana na podstawie danych pochodzących z WorldBank-u. Zawiera w sobie dziesięć zmiennych – jedną objaśnianą oraz dziewięć objaśniających. Udało się zebrać kompletne dane za 2018 rok dla 135 państw z różnych części świata.

4.2. Opis zmiennych

Tabela 1. Opis zmiennych

Zmienna	Opis zmiennej	Miara
Y_{infant}	Liczba zgonów noworodków na 1000 urodzeń	logarytm
X_{GDP}	Wartość nominalna PKB	logarytm
X_{GDPpc}	Wartość PKB per capita	logarytm
$X_{population}$	Populacja danego kraju	logarytm
X_{Healt_care}	Procentowy udział wydatków rządowych na ochronę zdrowia	%PKB
X_{Labor_force}	Zasób siły roboczej	wartość nominalna
X_{Death_rate}	Współczynnik śmiertelności	śmierci/1000 mieszkańców
X_{urban}	Liczba ludzi mieszkających w miastach	%populacji
$X_{tourists}$	Roczna liczba turystów	wartość nominalna
X_{HIV}	Czy występuje wirus HIV	zmienna dyskretna

Surowe dane z WorldBank-u zostały poddane następującym transformacjom:

1) Zostały usunięte obserwacje zawierające braki danych
2) Zmienna objaśniająca HIV została wykorzystana jako zmienna dyskretna, przyjmująca wartości 0 bądź 1. Wartość 0 – dla krajów posiadający mniej niż 10 zdiagnozowanych przypadków wirusa HIV na 1000 mieszkańców, 1 w przeciwnym wypadku.
3) Poszczególne zmienne (Tabela 1.) zostały zlogarytmowane. W dalszej części pracy została przedstawiona stosowna motywacja do takiego działania.

Wyjściowa wersja modelu przedstawia się zatem następująco:

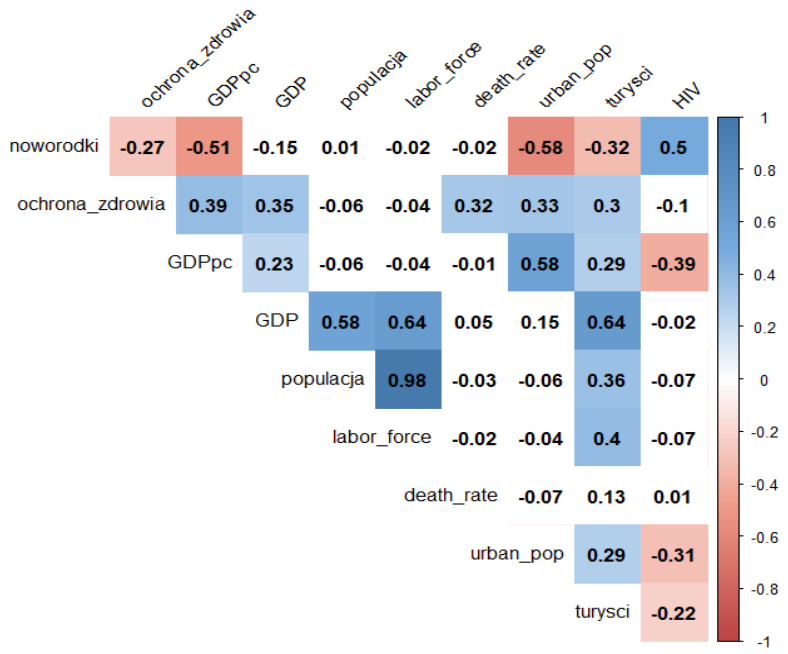
$$Y_{infant} = \beta_0 + \beta_1 * X_{GDP} + \beta_2 * X_{GDPpc} + \beta_3 * X_{pop} + \beta_4 * X_{HC} + \beta_5 * X_{LF} + \beta_6 * X_{DR} + \beta_7 * X_{urban} + \beta_8 * X_{tour} + \beta_9 * X_{HIV} + \varepsilon$$

Rozdział V Klasyczny model regresji liniowej

5.1.Macierz korelacji

Pracę nad osiągnięciem końcowej wersji statystycznie istotnego modelu została rozpoczęta od przeanalizowania macierzy korelacji. Na podstawie Wykresu 1. można zauważyć, że zmienne: populacja, labor_force oraz death_rate wykazują praktycznie zerową korelację ze zmienną objaśnianą. Dodatkowo zmienna populacja jest silnie skorelowana ze zmienną labor_force. Na tym etapie podjęta została decyzja, że zmienne labor_force oraz death_rate zostaną usunięte z modelu i nie będą brały udziału w dalszej analizie.

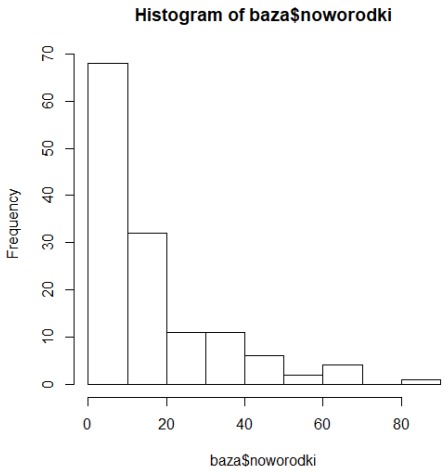
Wykres 1. Macierz korelacji zmiennych

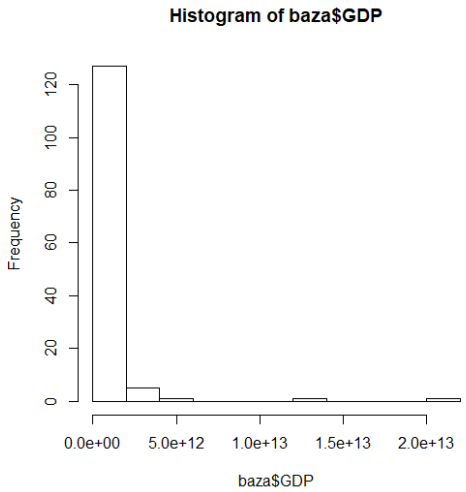


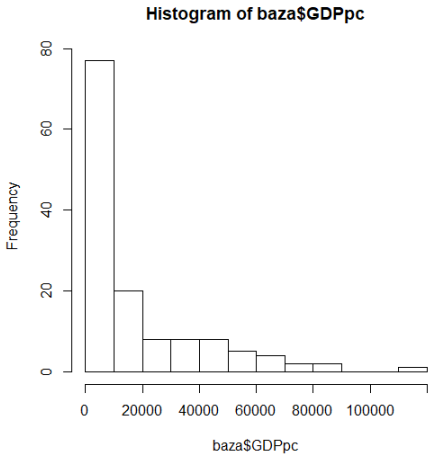
5.2. Rozkłady zmiennych w modelu

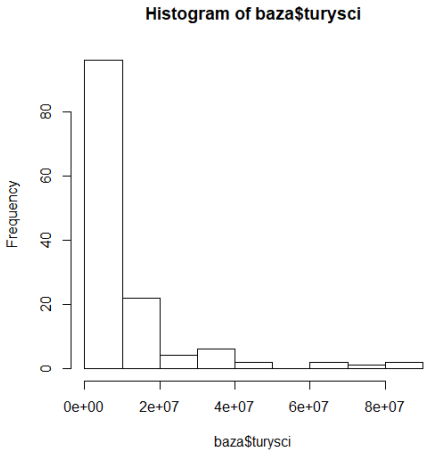
Wykres 2. przedstawia rozkłady zmiennych przed modyfikacjami. W przypadku zmiennych: noworodki, GDP, GDP per capita, turyści oraz populacja widać silną lewostronną skośność rozkładu. Na podstawie przeprowadzonej analizy graficznej została podjęta próba doprowadzenia zmiennych do rozkładu normalnego.

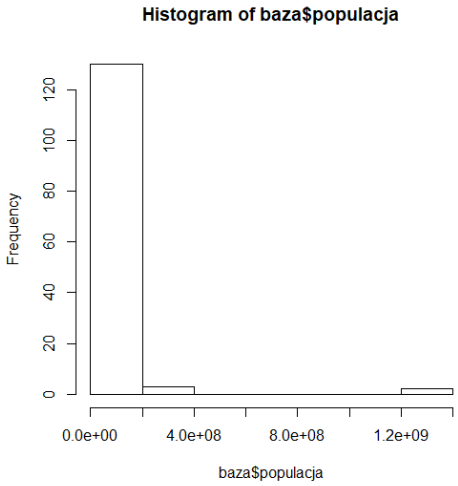
Wykres 2. Histogramy zmiennych przed przekształceniami

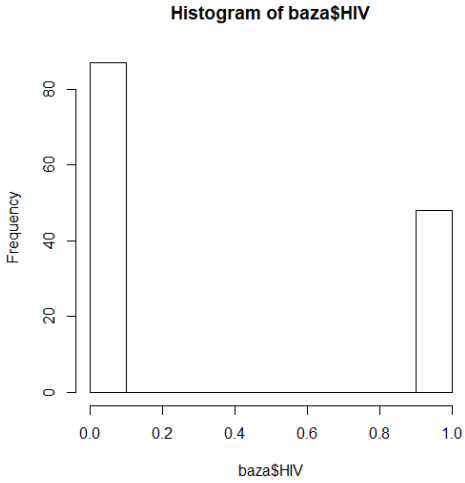


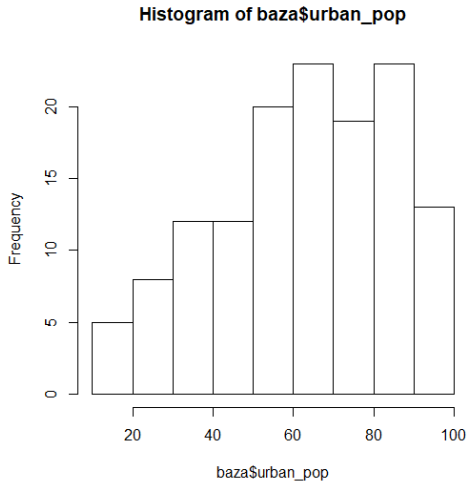


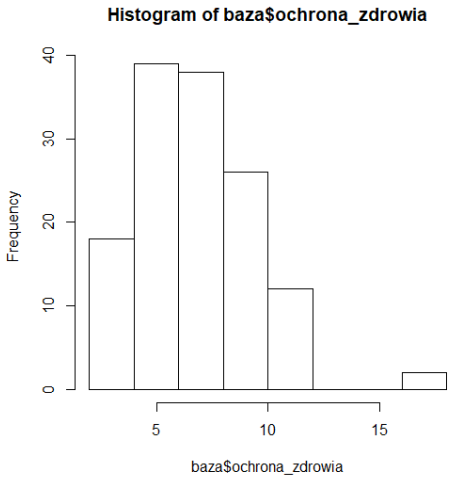












Na Wykresie 3. widać uzyskane rezultaty po zastosowaniu przekształceń. Niestety nie w każdym przypadku udało się uzyskać perfekcyjne rozkłady normalne. Widać natomiast wyraźną poprawę w stosunku do histogramów przed przekształceniami, co potwierdzają wyniki testów Shapiro-Wilka na normalność rozkładów przedstawione w Tabeli 2.

Zostały podjęte dalsze próby doprowadzenia zmiennych do postaci rozkładu normalnego, niestety bezskuteczne, stąd decyzja o pozostawieniu zmiennych zlogarytmowanych w modelu.

Wykres 3. Histogramy zmiennych po przekształceniach

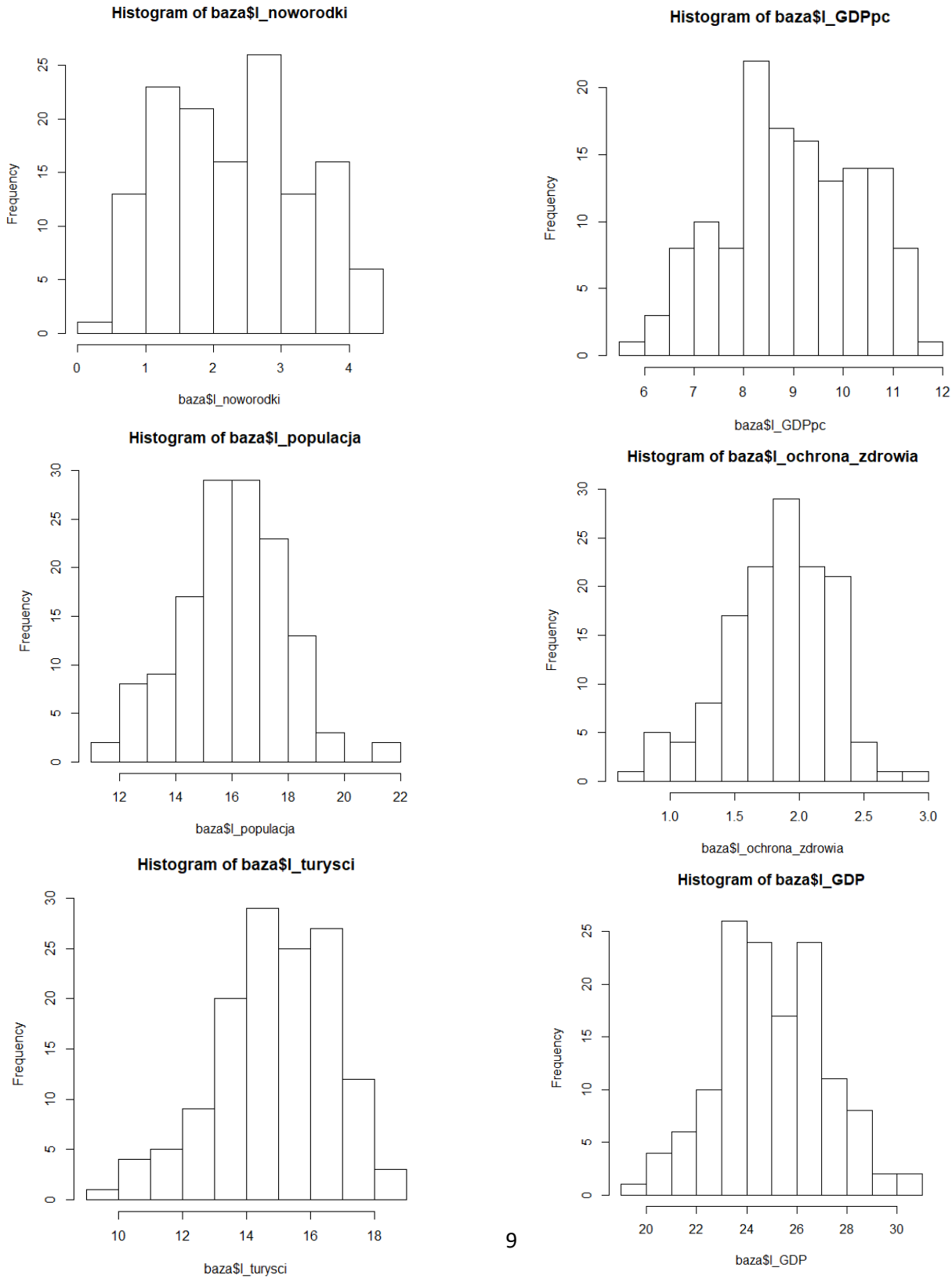


Tabela 2. Porównanie wartości statystyk testowych (test Shapiro-Wilka) dla zmiennych przekształconych

Zmienna	Przed zlogarytmowaniem	Po zlogarytmowaniu
Y_{infant}	W = 0.78992, p-value = 1.25e-12	W = 0.96996, p-value = 0.004412
X_{GDP}	W = 0.26431, p-value < 2.2e-16	W = 0.99430, p-value = 0.8703
X_{GDPpc}	W = 0.75581, p-value = 1.047e-13	W = 0.97869, p-value = 0.03255
$X_{population}$	W = 0.24085, p-value < 2.2e-16	W = 0.98972, p-value = 0.42
X_{Healt_care}	W = 0.96001, p-value = 0.0005498	W = 0.98037, p-value = 0.04858
$X_{tourists}$	W = 0.60841, p-value < 2.2e-16	W = 0.97979, p-value = 0.04226

Na tym etapie postać estymowanego modelu kształtuje się zatem następująco:

$$Y_{\log infant} = \beta_0 + \beta_1 * X_{\log GDP} + \beta_2 * X_{\log GDPpc} + \beta_3 * X_{\log pop} + \beta_4 * X_{\log HC} + \beta_5 * X_{urban} +$$
$$+\beta_6 * X_{\log\_tour} + \beta_7 * X_{HIV} + \varepsilon$$

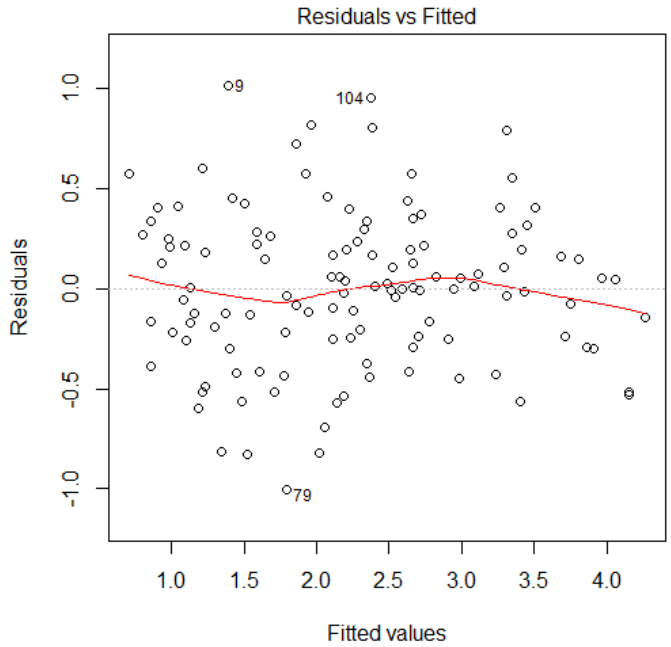
Rozdział VI Analiza obserwacji w modelu

W kolejnym kroku należy przyjrzeć się bliżej konkretnym obserwacjom w stworzonym modelu. Baza wejściowa po uwzględnieniu braku danych zawierała 135 unikalnych obserwacji. Na tym etapie na podstawie analizy odległości Cook'a zostały wyeliminowane obserwacje odstające. Łącznie wykryto 12 outlinerów, które zostały usunięte z modelu.

6.1. Reszty względem wartości dopasowanych

Jak widać na Wykresie 4. reszty w modelu są równomiernie rozłożone wokół wartości dopasowanych. Czerwona linia na wykresie jest względnie pozioma, co dobrze wróży na przyszłą weryfikację poprawności postaci funkcyjnej modelu. Na wykresie nie widać również heteroskedastyczności reszt w modelu.

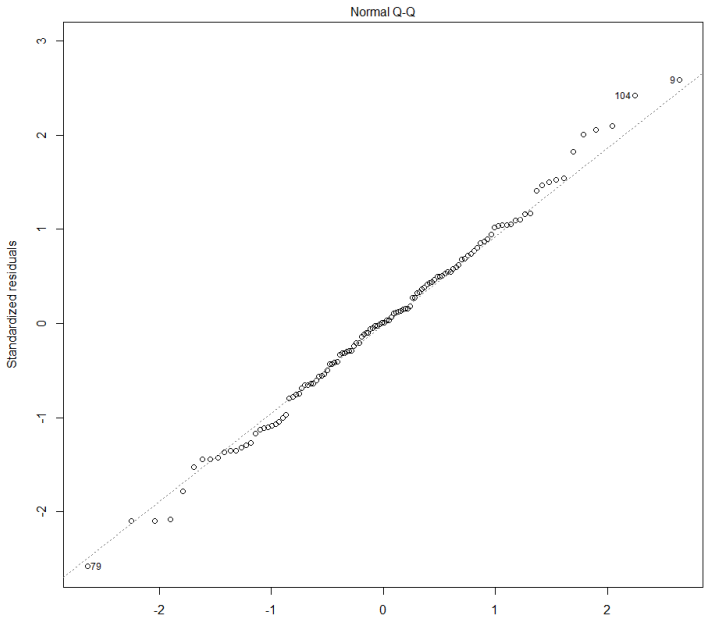
Wykres 4. Residuals vs Fitted



6.2. Kwantyle empiryczne, a kwantyle teoretyczne reszt

Na podstawie Wykresu 5. Q-Q plot można zauważyć poprawne nachylenie funkcji identycznościowej (y = x), która dowodzi zgodności pomiędzy kwantylami empirycznymi, a teoretycznymi. Niedopasowania na końcach rozkładu prawdopodobnie są spowodowane nieperfekcyjnym rozkładem normalnym, być może jest to rozkład t-studenta.

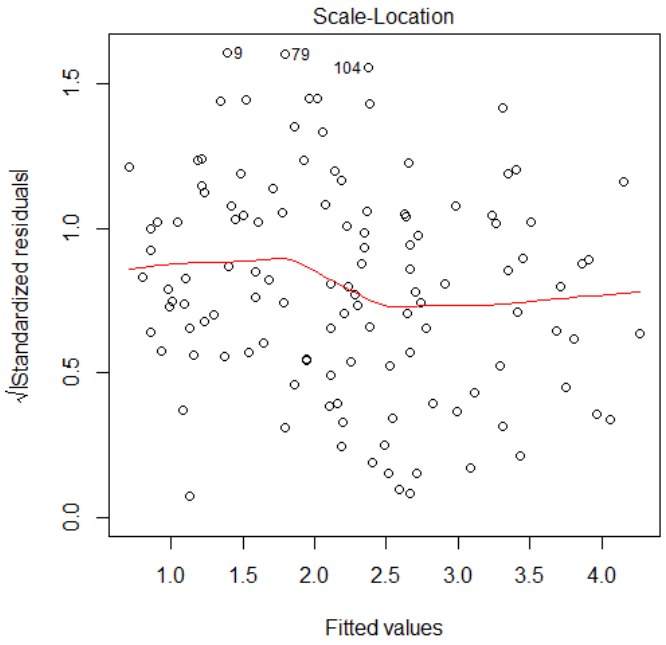
Wykres 5. Normal Q-Q plot



6.3. Scale – Location

Na podstawie Wykresu 6. Scale – Location można stwierdzić, że w modelu nie ma zależności między standaryzowanymi resztami, a dopasowanymi wartościami.

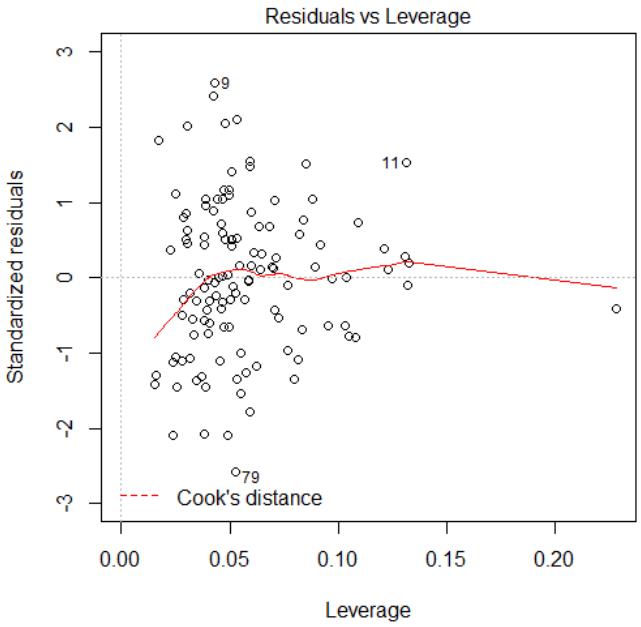
Wykres 6. Scale-Location plot



6.4. Reszty względem dźwigni

Wykres reszt względem dźwigni już po usunięciu obserwacji odstających prezentuje się jak poniżej, co kończy analizę obserwacji w modelu.

Wykres 7. Residuals vs Leverage



Rozdział VII Właściwa postać modelu – od ogółu do szczegółu

Rycina 1. prezentuje pierwsze wyniki oszacowanego modelu

Rycina 1. Oszacowanie parametrów dla pierwszej wersji modelu

	Dependent variable:
	smiertelnosc wsrod noworodkow
log GDP	21.301
	(23.962)
log GDP per capita	-21.715
	(23.953)
log populacji	-21.197
	(23.960)
log wydatkow panstwa na ochrone zdrowia	-0.519***
	(0.103)
log liczby turystow	-0.158***
	(0.040)
% ludzi mieszkajacych w miastach	0.001
	(0.003)
czy kraj ma probelm z HIV	0.225**
	(0.090)
Constant	7.523***
	(0.513)
Observations	123
R^2	0.842
Adjusted R^2	0.832
Residual Std. Error	0.400 (df = 115)
F Statistic	87.517*** (df = 7; 115)
Note:	*p<0.1; **p<0.05; ***p<0.01

P-value dla wszystkich zmiennych w modelu wynosi 2.2e-16, co pozwala odrzucić hipotezę zerową o łącznej nieistotności zmiennych w modelu. Przechodząc przez procedurę od ogółu do szczegółu, każdorazowo sprawdzając hipotezę o łącznej nieistotności parametrów, z modelu zostały usunięte zmienne nominalne PKB oraz odsetek ludzi zamieszkujący w miastach.

Ostateczne oszacowanie parametrów w modelu przedstawione zostało na Rycinie 2. w kolumnie (3)

Rycina 2. Oszacowanie parametrów dla wszystkich modeli, wydruk publikacyjny

	Dependent variable:		
	smiertelnosc wsrod noworodkow		
	(1)	(2)	(3)
log GDP	21.301	21.752	
	(23.962)	(23.829)	
log GDP per capita	-21.715	-22.156	-0.410***
	(23.953)	(23.823)	(0.044)
log populacji	-21.197	-21.647	0.104***
	(23.960)	(23.828)	(0.032)
log wydatkow panstwa na ochrone zdrowia	-0.519***	-0.518***	-0.521***
	(0.103)	(0.102)	(0.102)
log liczby turystow	-0.158***	-0.158***	-0.154***
	(0.040)	(0.039)	(0.039)
% ludzi mieszkajacych w miastach	0.001		
	(0.003)		
czy kraj ma probelm z HIV	0.225**	0.226**	0.217**
	(0.090)	(0.090)	(0.089)
Constant	7.523***	7.461***	7.489***
	(0.513)	(0.475)	(0.473)
Observations	123	123	123
R^2	0.842	0.842	0.841
Adjusted R^2	0.832	0.834	0.834
Residual Std. Error	0.400 (df = 115)	0.399 (df = 116)	0.398 (df = 117)
F Statistic	87.517*** (df = 7; 115)	102.878*** (df = 6; 116)	123.463*** (df = 5; 117)
Note:			*p<0.1; **p<0.05; ***p<0.01

Końcowa postać modelu zawiera stałą oraz 5 zmiennych istotnych statystycznie na poziomie $\alpha = 5\%$. Należy odrzucić również hipotezę zerową o łącznej nieistotności parametrów w modelu, ze względu na statystykę testową równą 123.463 i odpowiadające jej p-value 2.2e-16. Zmienność w modelu została wyjaśniona w 83% na podstawie 123 obserwacji i 117 stopni swobody.

Końcowa postać modelu prezentuje się następująco:

$$Y_{\log infant} = 7.489 - 0.41 * X_{\log GDPpc} + 0.104 * X_{\log pop} - 0.521 * X_{\log HC} - 0.154 * X_{\log tour} + \\ +0.217 * X_{HIV} + \varepsilon$$

Rozdział VIII Diagnostyka modelu

W następnym kroku należy przetestować model pod kątem prawidłowości formy funkcyjnej, homoskedastyczności reszt, autokorelacji reszt oraz normalności reszt. Wyniki przeprowadzonych testów ukazane zostały w Tabeli 3.

Tabela 3. Zestawienie testów diagnostycznych

Diagnostyka	Hipoteza zerowa	Wartość testu, odpowiadające p-value	Decyzja
Test Reset	Model posiada prawidłową formę funkcyjną	RESET = 0.55222 p-value = 0.849	Brak podstaw do odrzucenia hipotezy zerowej
Test Breusch-Pagana	Reszty są homoskedastyczne	BP = 9.7061 p-value = 0.084	Brak podstaw do odrzucenia hipotezy zerowej
Test Jarque-Bera	Reszty posiadają rozkład normalny	X-squared = 0.04812 p-value = 0.9762	Brak podstaw do odrzucenia hipotezy zerowej
Test Breusch-Godfrey'a	Nie występuje autokorelacja reszt	LM test = 1.1383 p-value = 0.286	Brak podstaw do odrzucenia hipotezy zerowej
Test Shapiro-Wilka	Reszty posiadają rozkład normalny	W = 0.99474 p-value = 0.9309	Brak podstaw do odrzucenia hipotezy zerowej

Model pozytywnie przeszedł wszystkie testy diagnostyczne. W każdym przypadku p-value przekracza przyjęty poziom istotności $\alpha = 5\%$, co nie pozwala odrzucić hipotez zerowych. Nie ma zatem podstaw do dalszej optymalizacji modelu, wszystkie założenia KMRL zostały spełnione.

Rozdział IX Interpretacja parametrów modelu

Model stworzony na potrzeby badania przyjął ostatecznie formę modelu logliniowego. Interpretacja oszacowanych parametrów przedstawiona została w Tabeli 4. poniżej.

Tabela 4. Interpretacja oszacowań parametrów w modelu

Zmienna	Oszacowanie parametru przy zmiennej	Interpretacja
Log (GDP_per_capita)	-0.410	Wzrost PKB per capita o 1% spowoduje spadek odsetka śmierci wśród noworodków o 0,41%
Log(populacja)	0.104	Wzrost populacji o 1% spowoduje wzrost odsetka śmierci wśród noworodków o 0,104%
Log (Wydatki na ochronę zdrowia %PKB)	-0.521	Wzrost wydatków rządowych na ochronę zdrowia o 1% spowoduje spadek odsetka śmierci wśród noworodków o 0,521%
Log (liczba turystów)	-0.154	Wzrost liczby turystów o 1% spowoduje spadek odsetka śmierci wśród noworodków o 0,154%
Czy kraj ma problem z HIV	0.217	Państwa, u których występuje problem z wirusem HIV mają o 21,7% wyższy odsetek śmiertelności wśród noworodków

Rozdział X Podsumowanie i wnioski płynące z modelu

Analiza determinantów śmiertelności wśród noworodków wykazała w przeważającej większości rezultaty zgodne z logiką. Każde państwo ze zbioru obserwacji cechuje się niezerową śmiertelnością wśród noworodków, o czym świadczy dodatni parametr przy stałej. Być może takie oszacowanie bierze się z braków w wiedzy medycznej, niepozwalających uratować każdego noworodka, a być może z innej zmiennej egzogenicznej nieuwzględnionej w modelu. Ujemne wartości parametrów przy zmiennych PKB per capita oraz wydatków państwa na ochronę zdrowia mierzonych jako %PKB są zgodne z przewidywaniami – pieniądze szczęścia nie dają, ale jak widać życie mogą uratować. Zastanawiające jest istotne statystycznie oszacowanie parametru stojącego przy zmiennej liczba turystów. Tworząc postać modelu spodziewano nie spodziewano się ujemnej zależności – turyści zwłaszcza w krajach słabo rozwiniętych są doskonałym narzędziem transmisji różnego typu wirusów, stąd spodziewany albo brak zależności, albo dodatnie oszacowanie parametru. Być może jest to błąd pierwszego rodzaju wynikający z analizowanej podpróbki.

Warta podkreślenia jest również niezwykle wysoka wartość skorygowanego R^2 na poziomie 83,4%. P-value dla testów diagnostycznych okazało się również bardzo wysokie, poza testem Breuscha-Pagana w każdym przypadku przekroczyło 25%. Model niewątpliwie pokazał ciekawe rezultaty, z pewnością warte powtórzenia na danych pochodzących z innego przedziału czasowego.

Bibliografia

S. Budhdeo, J. Watkins, R. Atun, C. Williams, T. Zeltner, M. Maruthappu, *Changes in government spending on healthcare and population mortality in the European union, 1995–2010: a cross-sectional ecological study* (2015)

M. Maruthappu, J. Shalhoub, Z.Tariq, C. Williams, R. Atun, A. H. Davies, T. Zeltne, *Unemployment, Government Healthcare Spending, and Cerebrovascular Mortality, Worldwide 1981–2009: An Ecological Study* (2015)

M. Maruthappu, C. Williams, R. Atun, P. Agrawal, T. Zeltner, *The association between government healthcare spending and maternal mortality in the European Union, 1981–2010: a retrospective study* (2015)

Heron, P.Melanie, *Deaths : leading causes for 2017* (2019)

C. Dollfus, M. Patetta, E. Siegel, A. W. Cross, *Infant Mortality: A Practical Approach to the Analysis of the Leading Causes of Death and Risk Factors* (1990)

C. Carne, S. Semple, A. MacLarnon, B. Majolo, *Implications of Tourist–Macaque Interactions for Disease Transmission (*2017)

M.P. Muehlenbein, J. Wallis, *Considering risks of pathogen transmission associated with primate-based tourism* (2014)

https://databank.worldbank.org/source/world-development-indicators (dostęp 13.01.2021 r.)