

## **Metody ekonometryczne - raport zaliczeniowy**

Celem niniejszego raportu jest analiza ekonometryczna oczekiwanej długości życia w chwili narodzin na świecie. Poniższa analiza pozwoli odpowiedzieć na następujące pytania:

- Jakie czynniki ekonomiczne istotnie wpływają na oczekiwaną długość życia?
- Jakie czynniki społeczno-kulturowe istotnie wpływają na oczekiwaną długość życia?
- Czy poziom rozwoju danego kraju istotnie wpływa na oczekiwaną długość życia?
- Jak poszczególne czynniki wpływają na oczekiwaną długość życia?

### **Opis problemu ekonomicznego**

Oczekiwana długość życia w chwili urodzenia jest jednym z podstawowych wskaźników oceny stanu zdrowia społeczeństwa, świadczący o stanie zdrowia całej populacji, jakości opieki zdrowotnej czy warunków życia społeczności.

Oczekiwana długość życia w momencie urodzenia jest to liczba lat, jaką może przeżyć nowo narodzone dziecko przy założeniu, że prawdopodobieństwo zgonu w każdym roku jego życia nie zmieni się i będzie równe temu, jakie określono dla momentu urodzenia.

### **Motywacja wybranego problemu badawczego**

Poniższa analiza pozwoli stwierdzić, które czynniki istotnie wpływają na oczekiwaną długość życia, a także w jakim stanie jest zdrowie konkretnych populacji. Wyniki analizy mogą być bardzo przydatne władz krajów, które chcą poprawić jakość życia społeczeństwa. Niewątpliwie zwierzchnicy każdego państwa chcą, aby oczekiwana długość życia była na jak najwyższym poziomie.

### **Opis zbioru danych**

Zmienne, które zostały użyte w modelu ekonometrycznym zostały wybrane na podstawie raportu „A STATISTICAL ANALYSIS OF LIFE EXPECTANCY ACROSS COUNTRIES USING MULTIPLE REGRESSION” (Miranda Chen, Michael Ching, 2000), a także na podstawie własnych rozważań. Wyróżniono następujące zmienne:

Lifeexp – średnia oczekiwana długość życia w chwili urodzenia w roku 2014 (w latach); źródło: data.worldbank.org

Literacyrate – procentowy udział osób w populacji umiejących czytać i pisać w roku 2014; źródło: data.unicef.org

Homicidiesper100k – zabójstwa na 100tyś osób w roku 2014; źródło: data.worldbank.org

Electricity – konsumpcja energii elektrycznej na jedną osobę w roku 2014 (w kWh – kilowatogodzinach); źródło: data.worldbank.org

Schooling – średnia oczekiwana długość edukacji w roku 2014 (w latach); źródło: WHO data repository

HIV.AIDS – liczba zgonów dzieci w wieku 0-4 lat spowodowana HIV/AIDS na 1000 osób w 2014 roku; źródło: WHO data repository

Wateraccess – procentowy udział osób w populacji mających dostęp do wody zdatnej do użytkowania w 2014 roku; źródło: landportal data

Tuberculosis – zachorowania na gruźlicę na 100tyś osób w 2014 roku; źródło: data.worldbank.org

Inflation – roczna inflacja (w %) w 2014 roku; źródło: data.worldbank.org

GDPpercapita – PKB per capita (w dolarach amerykańskich) w 2014 roku; źródło: data.worldbank.org

Healthexppercapita – uśrednione wydatki rządowe na służbę zdrowia (w dolarach amerykańskich według parytetu siły nabywczej) w latach 2005-2014; źródło: data.worldbank.org

Fertility rate – uśredniony współczynnik płodności (urodzenia na jedną kobietę) w latach 2005-2014; źródło: data.worldbank.org

Internet – procentowy udział osób w populacji mających dostęp do internetu; źródło: data.worldbank.org

CO2 – uśrednione emisje CO2 per capita (w tonach) w latach 2005-2014; źródło: data.worldbank.org

Forest – procentowy udział zalesienia kraju w 2014 roku; źródło: data.worldbank.org

Urbanpop – liczebność populacji danego kraju w miastach w 2014 roku; źródło: data.worldbank.org

Urbanpopgrowth – uśredniony roczny wskaźnik wzrostu populacji w miastach (w %) w latach 2005-2014; źródło: data.worldbank.org

Leastdeveloped – zmienna binarna identyfikująca kraj najmniej rozwinięty (1 – kraj najmniej rozwinięty, 0 – w p.p.); źródło: UN data.

## **Transformacje oryginalnych danych oraz motywacja stojąca za ich przeprowadzeniem**

Tworząc analizę zostały wykonane następujące transformacje danych:

1. Dane dotyczące wzrostu populacji w miastach (urbanpopgrowth) zostały uśrednione, ponieważ przykładowy wzrost populacji w roku, w którym została wykonana analiza nie przełożyłby się istotnie na oczekiwaną długość życia. Aby widoczny był zauważalny efekt tej zmiennej objaśniającej należy sprawdzić, jaki był trend migrowania ludzi z terenów wiejskich do terenów miejskich.
2. Dane dotyczące wydatków na ochronę zdrowia na jednego mieszkańca (healthexppercapita) zostały uśrednione, ponieważ wyższy poziom wydatków w roku, którego dotyczyła analiza nie wpłynąłby istotnie na stan służby zdrowia. Aby zaobserwować realny wpływ należy zobaczyć na jakim poziomie utrzymywały się wydatki na przestrzeni kilku lat.
3. Dane dotyczące emisji CO2 na jednego mieszkańca zostały uśrednione, ponieważ wpływ na oczekiwaną długość życia dana zmienna objaśniająca ma przez dłuższy okres czasu. To, że emisja CO2 była niska w analizowanym roku tj. 2014 nie oznacza, że w poprzednich latach nie była na wysokim poziomie. Obniżenie emisji CO2 przez dane państwo nie wpłynie gwałtownie na poprawę zdrowia, ponieważ wysoki poziom tej zmiennej w przeszłości ma zauważalny wpływ na obecną przeżywalność.

## Problemy związane z przeprowadzeniem analizy

Celem niniejszej analizy było przeprowadzenie analizy ekonometrycznej mierzącej oczekiwaną długość życia w jak największej ilości krajów. Głównym problemem okazał się brak możliwości zebrania większości ilości danych z krajów rozwijających się i najmniej rozwiniętych, co poskutkowało wyłączeniem niektórych krajów z modelu.

Kolejny problem wynikł podczas zbierania dosyć szczegółowych danych tj. liczba lekarzy na 10 tysięcy osób, czy procentowy udział ludności z dostępem do urządzeń sanitarnych. Okazało się, że brakuje tego typu danych z większości krajów, co wymusiło wykluczenie tych zmiennych z właściwego modelu.

Początkowo został wykonany następujący model, zmienną objaśnianą była oczekiwana długość życia:

### Model1

$$\text{Lifeexp}_i = \beta_0 + \beta_1 \text{literacyrate}_i + \beta_2 \text{homicidiesper100k}_i + \beta_3 \text{electricity}_i + \beta_4 \text{Schooling}_i + \beta_5 \text{HIV.AIDS}_i + \beta_6 \text{wateraccess}_i + \beta_7 \text{tuberculosis}_i + \beta_8 \text{inflation}_i + \beta_9 \text{gdppercapita}_i + \beta_{10} \text{healthexppercapita}_i + \beta_{11} \text{fertilityrate}_i + \beta_{12} \text{internet}_i + \beta_{13} \text{CO2}_i + \beta_{14} \text{forest}_i + \beta_{15} \text{urbanpop}_i + \beta_{16} \text{urbanpopgrowth}_i + \beta_{17} \text{leastdeveloped}_i + \epsilon_i;$$

gdzie  $i$  to indeks obserwacji,  $\epsilon_i$  to składnik losowy

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	77.81497753	6.50148173	11.969	0.0000000000000014	***
literacyrate	0.00338794	0.03400918	0.100	0.9211	
homicidiesper100k	-0.01661659	0.02021888	-0.822	0.4155	
electricity	0.00001390	0.00016143	0.086	0.9318	
Schooling	0.21972199	0.21379525	1.028	0.3096	
HIV.AIDS	-0.85153588	0.47537328	-1.791	0.0800	.
wateraccess	-0.04222417	0.04536682	-0.931	0.3570	
tuberculosis	-0.01199216	0.00234107	-5.123	0.0000061110363963	***
inflation	-0.13899817	0.07811435	-1.779	0.0819	.
gdppercapita	0.00002775	0.00004920	0.564	0.5756	
healthexppercapita	0.00243630	0.00078509	3.103	0.0033	**
fertilityrate	-1.46738305	0.61763644	-2.376	0.0218	*
internet	-0.01555490	0.02440460	-0.637	0.5271	
co2	-0.18602471	0.06934368	-2.683	0.0102	*
forest	-0.00098326	0.01466697	-0.067	0.9468	
urbanpop	0.01540176	0.01992911	0.773	0.4437	
urbanpopgrowth	0.42334171	0.18705306	2.263	0.0285	*
leastdeveloped	-1.88685012	1.19084558	-1.584	0.1201	
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 1.664 on 45 degrees of freedom					
Multiple R-squared: 0.911, Adjusted R-squared: 0.8773					
F-statistic: 27.08 on 17 and 45 DF, p-value: < 2.2e-16					

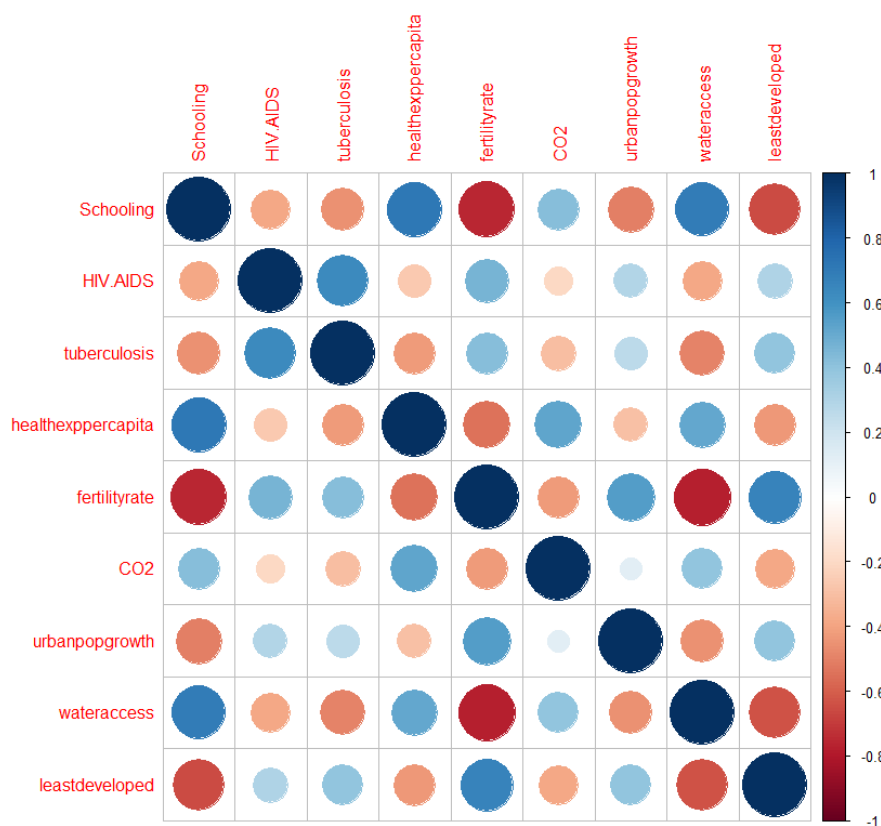
Niestety, z powodu braku danych model został wykonany dla 63 krajów. Oszacowania parametrów strukturalnych modelu pozwoliły wyciągnąć wnioski, które zmienne mogą być istotne. Bardziej szczegółowe dane są mniej dostępne i znacznie zmniejszają próbę i możliwości badania oczekiwanej

długości życia. Jednakże oszacowania parametrów strukturalnych tego modelu pomogły wyłonić zmienne najbardziej istotne, a tym samym zwiększyć próbę w rozważaniach.

## Wizualizacja dostępnych danych

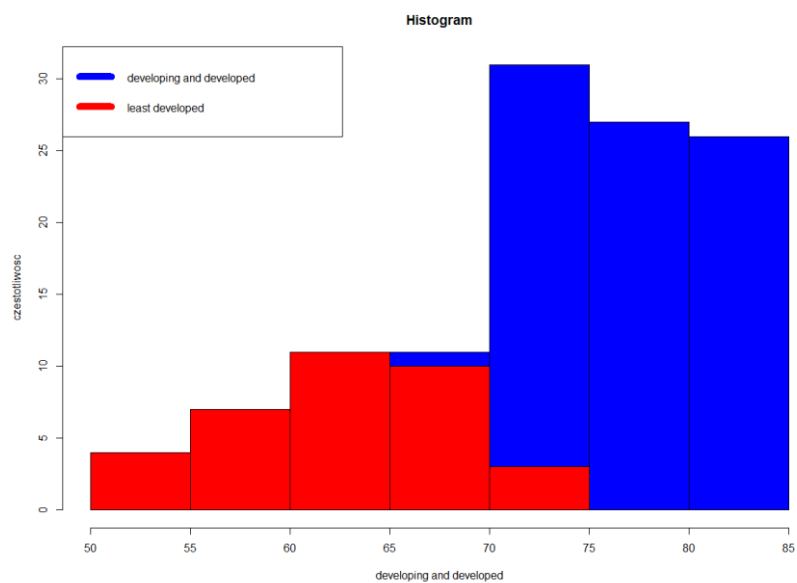
Poniższa wizualizacja danych została stworzona na podstawie danych dotyczących modelu bez większości zmiennych. Liczba obserwacji to 138. Użyty został dataset data2.

Wykres obrazujący korelację:



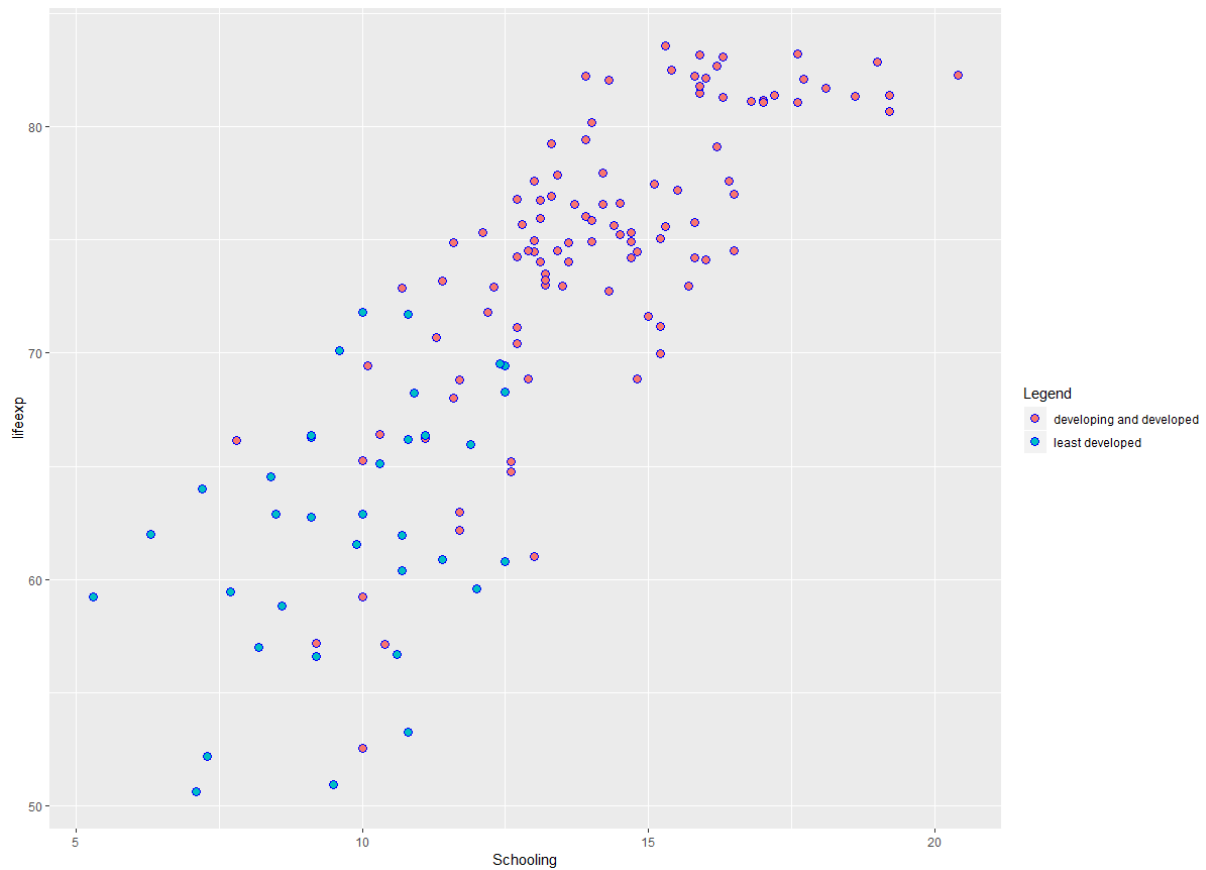
Na wykresie można zauważyć, że występuje silna ujemna korelacja pomiędzy zmiennymi fertilityrate oraz Schooling, a także fertilityrate i wateraccess. Jeśli spojrzymy na tablice korelacyjną, najbardziej skrajną wartością jest -0,77 i odnosi się do korelacji pomiędzy fertilityrate oraz wateraccess. Jest to dosyć wysoka wartość, jednak jest ona do przyjęcia.

### Kraje najmniej rozwinięte a oczekiwana długość życia:



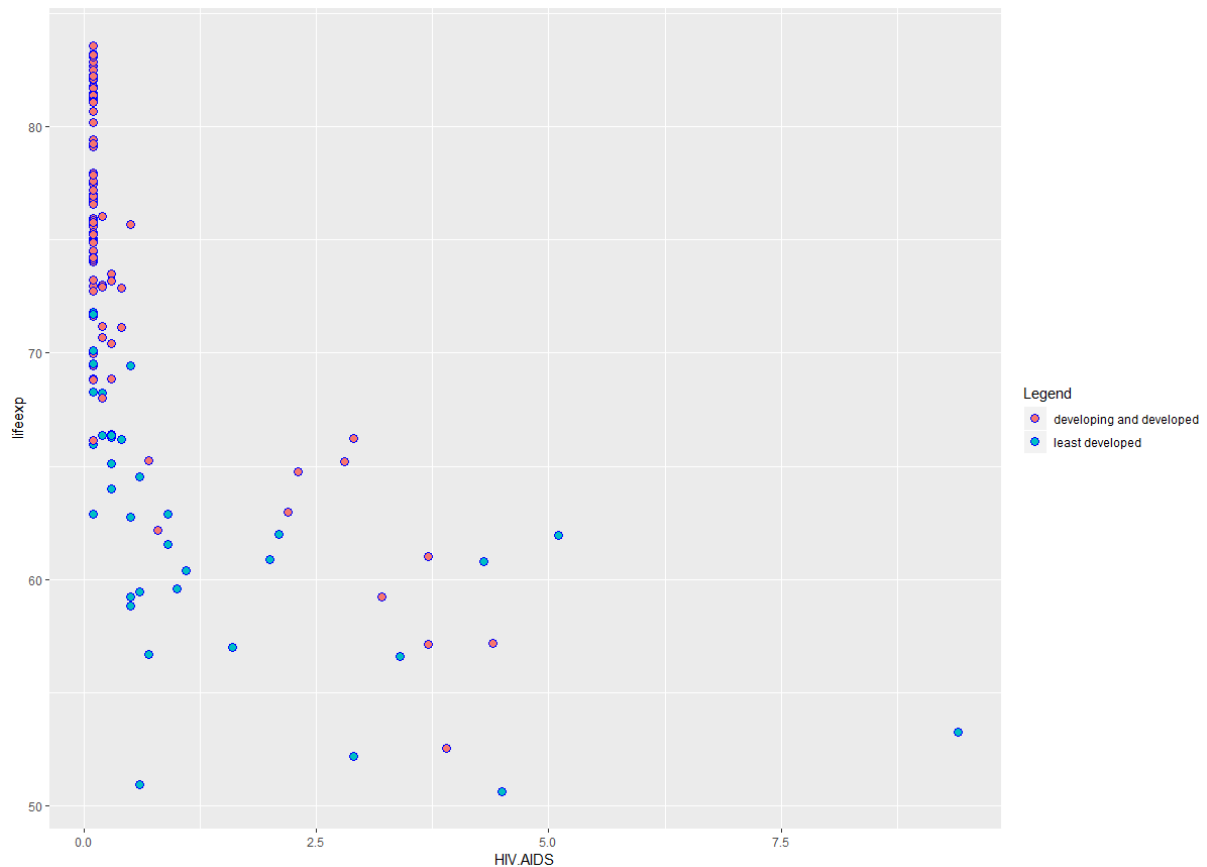
Jak widać, oczekiwana długość życia krajów najmniej rozwiniętych jest zdecydowanie niższa.

**Średnia oczekiwana długość edukacjiw krajach o różnym poziomie rozwoju oraz jej wpływ na oczekiwaną długość życia:**



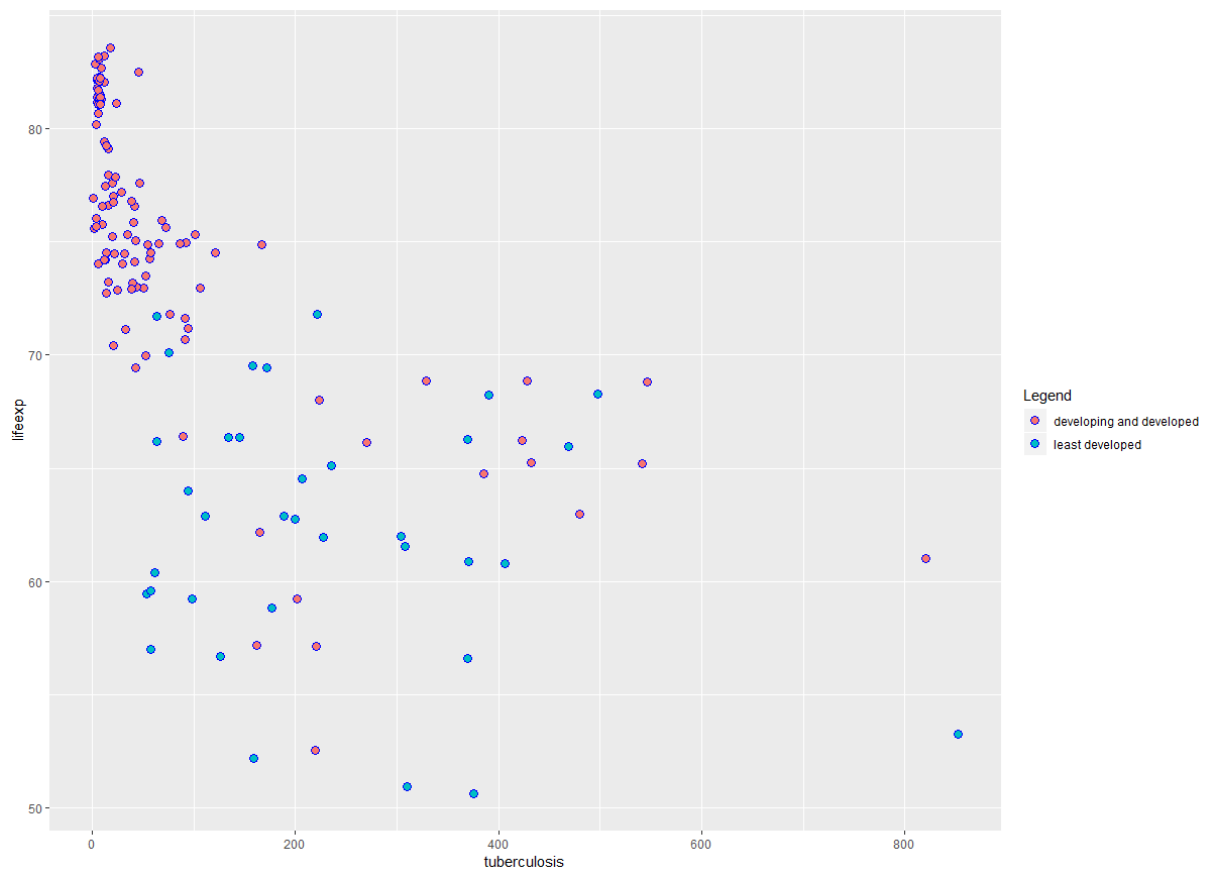
Średnia oczekiwana długość edukacja jest niższa w krajach najmniej rozwiniętych niż w krajach rozwijających się i rozwiniętych. Widać również pozytywną korelację pomiędzy oczekiwaną długością życia i średnią oczekiwaną długością edukacji.

**Umieralność na HIV/AIDS wśród dzieci (w wieku 0-4) w krajach o różnym poziomie rozwoju oraz jej wpływ na oczekiwaną długość życia:**



Na podstawie wykresu można wywnioskować ze wzrost wartości zmiennej HIV.AIDS wpływa na zmniejszenie się oczekiwanej długości życia. Widać również, że w jednym z krajów najmniej rozwiniętych występuje najwyższa umieralność na HIV/AIDS. Niestety ciężko stwierdzić, czy kraje najmniej rozwinięte mają średnio wyższy współczynnik umieralności na HIV/AIDS, ponieważ dużo krajów rozwijających się również boryka się z tym problemem.

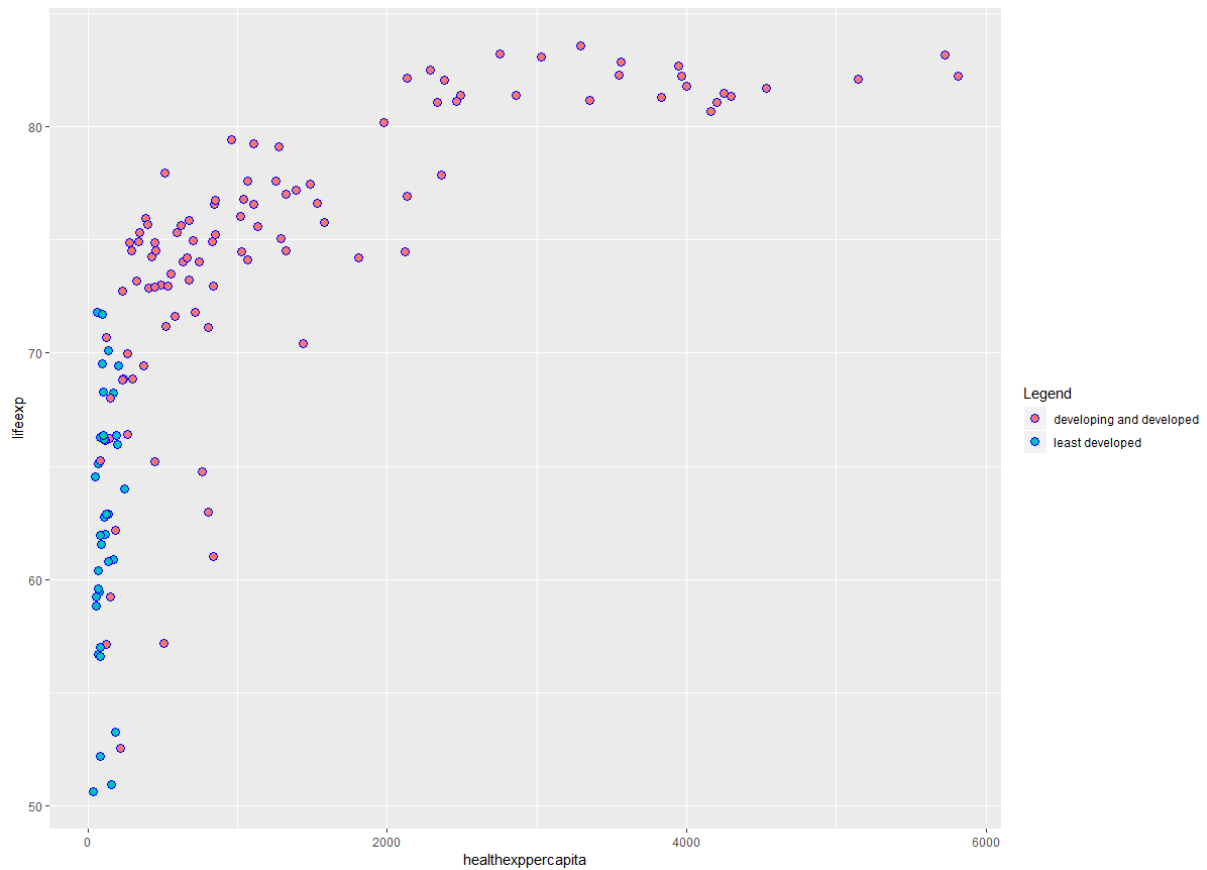
### Zachorowalność na gruźlicę:



Zachorowalność na gruźlicę wpływa na zmniejszenie się oczekiwanej długości życia. Można również wywnioskować, że w krajach najmniej rozwiniętych zachorowalność na gruźlicę jest raczej wyższa niż w krajach rozwijających się i rozwiniętych.

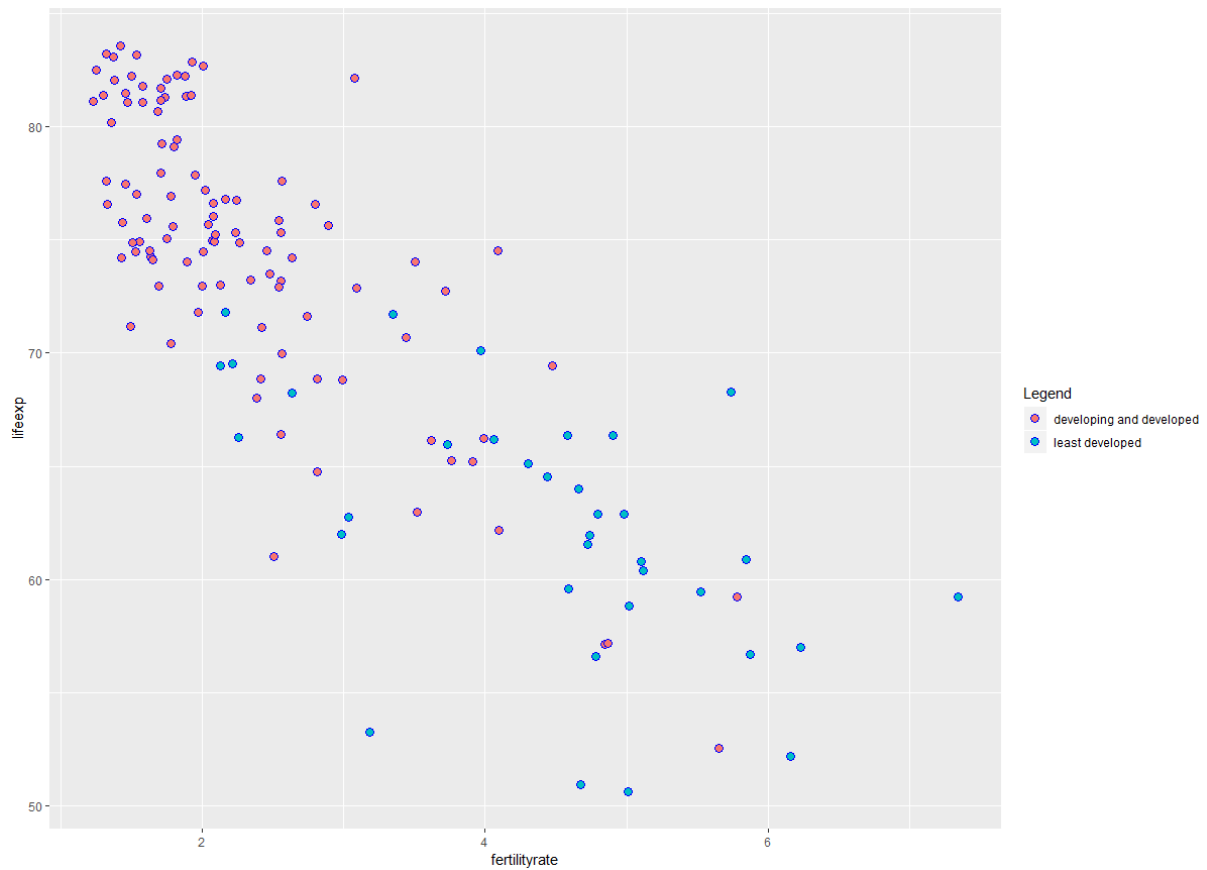


### Wydatki na ochronę zdrowia:



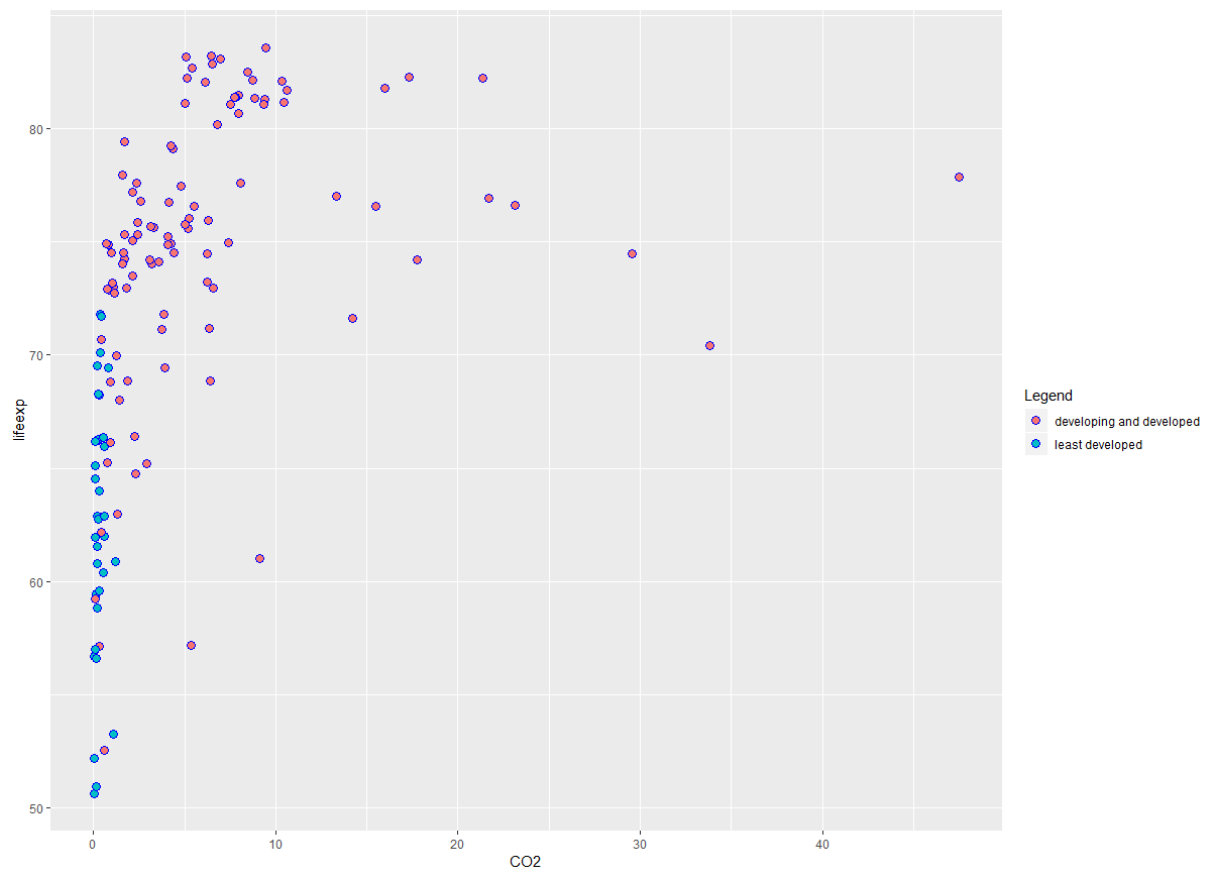
Wydatki na ochronę zdrowia per capita w krajach najmniej rozwiniętych są na bardzo niskim poziomie. Widać również pozytywną zależność pomiędzy wydatkami na ochronę zdrowia i oczekiwaną długością życia.

### Współczynnik płodności:



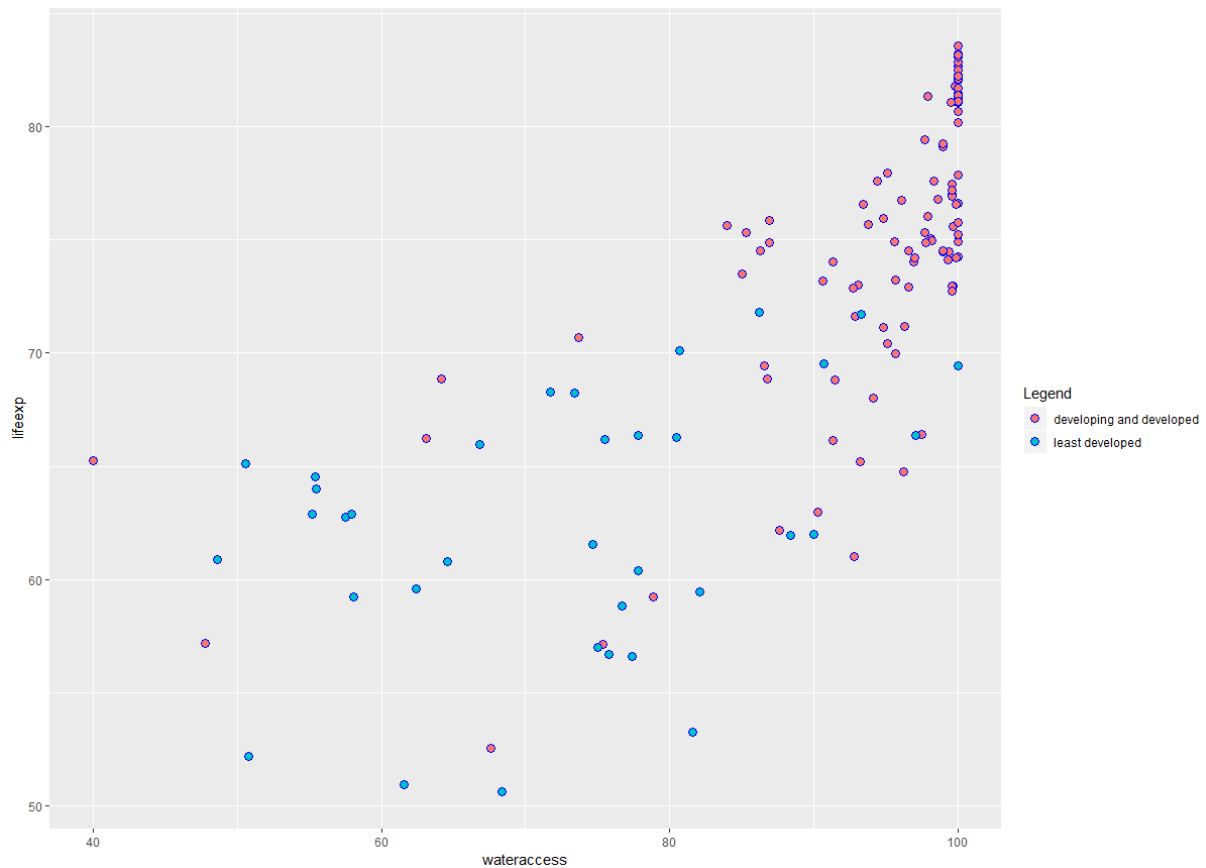
Kraje najmniej rozwinięte mają zdecydowanie wyższy współczynnik płodności niż kraje rozwijające się i rozwinięte. Występuje ujemna korelacja pomiędzy współczynnikiem płodności i oczekiwaną długością życia.

## Emisje CO2:



Emisje CO2 per capita w krajach najmniej rozwiniętych są na niskim poziomie. Można zauważyć, że emisje CO2 nie mają aż tak dużego wpływu na oczekiwaną długość życia.

### Dostęp do wody zdatnej do użytkowania:



Kraje najmniej rozwinięte mają zdecydowanie mniejszy dostęp do wody zdatnej do użytkowania. Widać pozytywną zależność pomiędzy dostępem do wody a oczekiwaną długością życia.

### Model2

W oparciu o powyższą wizualizację i analizę, a także przy uwzględnieniu problemów został stworzony następujący model:

$$\text{Lifeexp}_i = \beta_0 + \beta_1 \text{Schooling}_i + \beta_2 \text{HIV.AIDS}_i + \beta_3 \text{tuberculosis}_i + \beta_4 \text{healthexppercapita}_i + \beta_5 \text{fertilityrate}_i + \beta_6 \text{urbanpopgrowth}_i + \beta_7 \text{CO2}_i + \beta_8 \text{wateraccess}_i + \beta_9 \text{leastdeveloped}_i + \epsilon_i;$$

gdzie  $i$  to indeks obserwacji,  $\epsilon_i$  to składnik losowy.

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	65.7688540	3.4332797	19.156	< 0.00000000000000002	***
Schooling	0.5908895	0.1501799	3.935	0.000136	***
HIV.AIDS	-1.5601002	0.2382814	-6.547	0.00000000129	***
tuberculosis	-0.0058509	0.0020237	-2.891	0.004511	**
healthexppercapita	0.0015240	0.0002632	5.790	0.00000005140	***
fertilityrate	-2.1000184	0.3282672	-6.397	0.00000000272	***
urbanpopgrowth	0.4312381	0.1632133	2.642	0.009265	**
CO2	-0.1202442	0.0466633	-2.577	0.011105	*
wateraccess	0.0454963	0.0262062	1.736	0.084954	.
leastdeveloped	-1.4561433	0.7520457	-1.936	0.055042	.
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 2.625 on 128 degrees of freedom					
Multiple R-squared: 0.9043, Adjusted R-squared: 0.8976					
F-statistic: 134.4 on 9 and 128 DF, p-value: < 0.000000000000000022					

Kierunek zależności wszystkich zmiennych objaśniających wydaje się być uzasadniony i potwierdza wcześniejszą analizę. Ciężko porównać siłę zależności poszczególnych zmiennych, ponieważ są w zupełnie innej skali.

### Istotność zmiennych objaśniających

Test t-studenta

H0 – zmienna nieistotna

H1 – zmienna istotna

Mamy podstawy do odrzucenia H0 i przyjęcia H1. Decyzję podjęto przy 10% poziomie istotności.

(Powyższe wnioski będą identyczne dla wszystkich zmiennych objaśniających)

### Współczynnik determinacji R<sup>2</sup>

R<sup>2</sup> = 0,9043;

90,43% zmienności zmiennej objaśnianej jest wyjaśniana zmiennością wartości teoretycznych wynikających z modelu.

### Postać funkcyjna modelu

Aby zdiagnozować, czy postać funkcyjna modelu jest poprawna, wykonany został test reset.

Test reset:

H0 – postać funkcyjna modelu jest poprawna;

H1 – postać funkcyjna modelu jest niepoprawna.

RESET = 4.3294, df1 = 2, df2 = 126, p-value = 0.01519
---

Mamy podstawy do odrzucenia H0 i przyjęcia H1. Decyzja została podjęta przy 5% poziomie istotności.

## Współliniowość

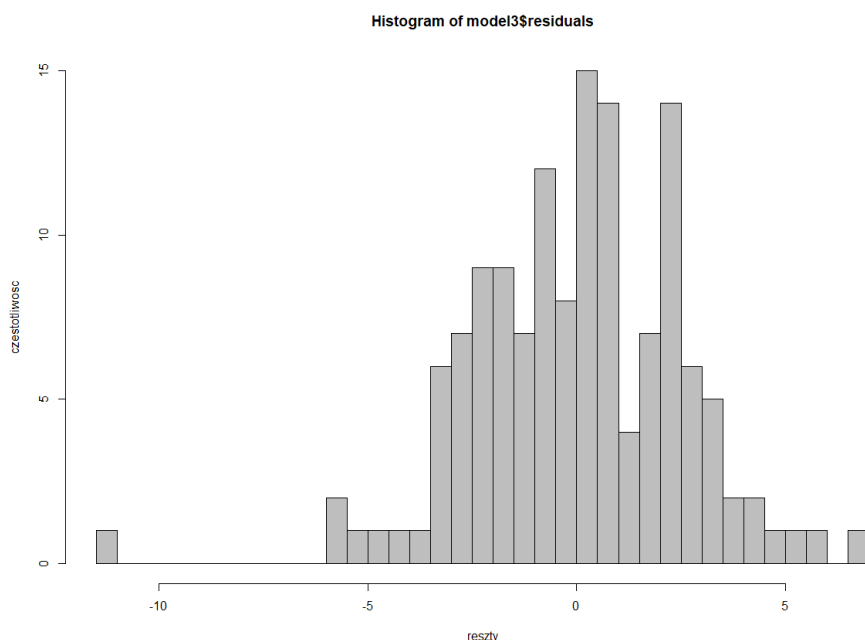
Współliniowość została zdiagnozowana za pomocą czynnika inflacji wariancji CIW.

Schooling	HIV.AIDS	tuberculosis	healthexppercapita	fertilityrate
3.876064	1.942568	2.138271	2.502190	4.175773
urbanpopgrowth	CO2	wateraccess	leastdeveloped	
2.076821	2.021450	2.963927	2.144449	

Wyniki pokazują, że w modelu nie występuje problem współliniowości zmiennych.

## Normalność składnika losowego

W celu sprawdzenia, czy składnik losowy ma rozkład normalny został wykonany histogram oraz przeprowadzony test Jarque'a-Berry.



Test Jarque'a-Berry:

H0 – składnik losowy ma rozkład normalny;

H1 – składnik losowy nie ma rozkładu normalnego

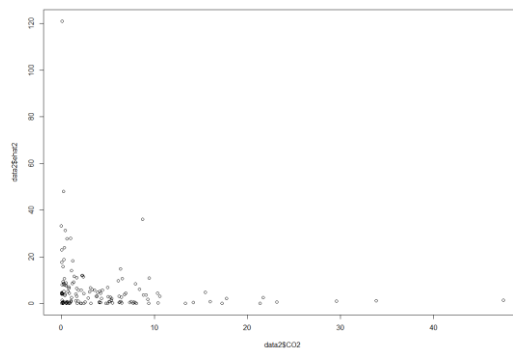
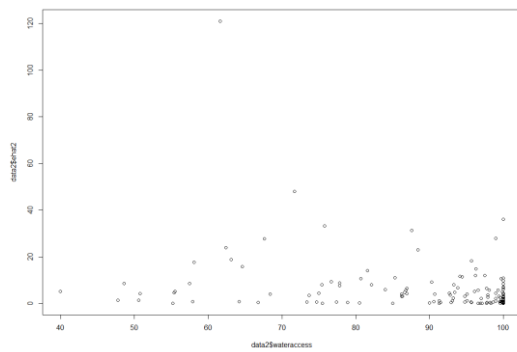
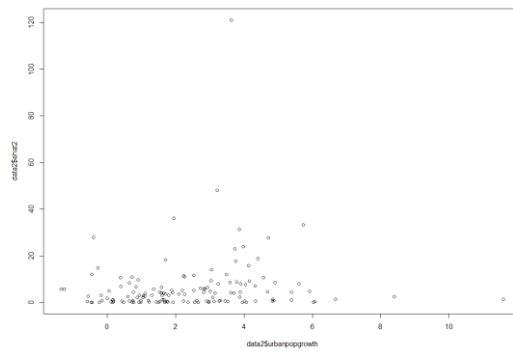
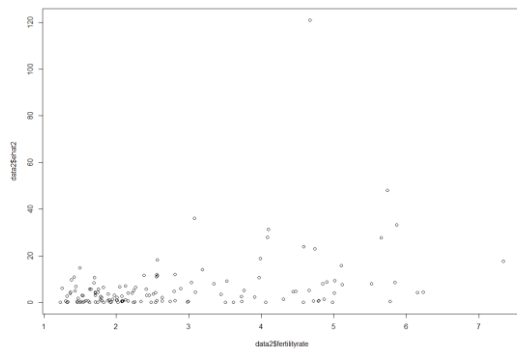
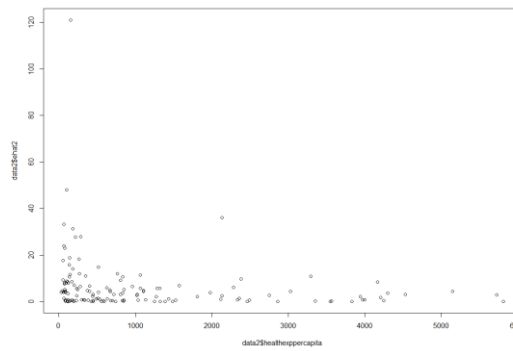
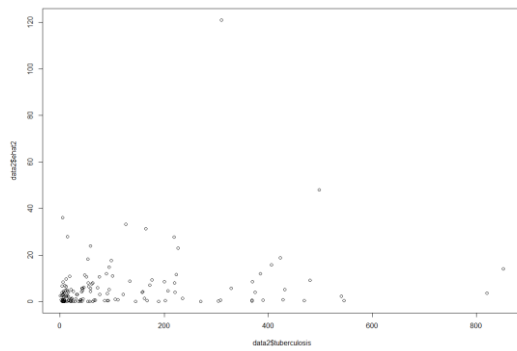
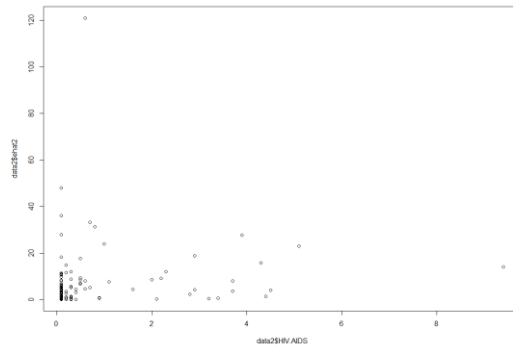
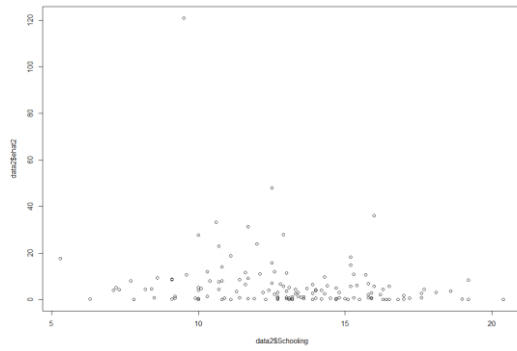
Jarque Bera Test

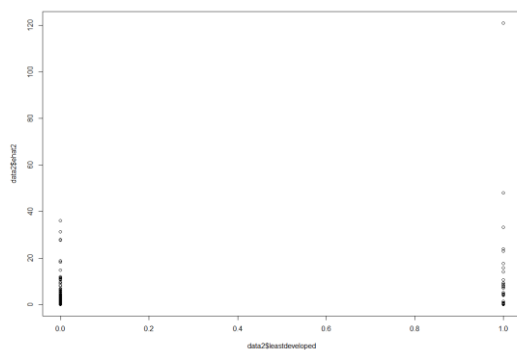
X-squared = 25.176, df = 2, p-value = 0.000003413

Zarówno narysowany histogram, jak i wyniki testu jednoznacznie wskazują, że składnik losowy nie ma rozkładu normalnego. Powoduje to utrudnione wnioskowanie dotyczące wartości składnika losowego. Jednak biorąc pod uwagę, że próba jest dosyć liczna (138 obserwacji), możemy badać własności składnika losowego.

## Diagnozowanie heteroskedastyczności składnika losowego

Na początku wykonane zostały wykresy kwadratów reszt względem zmiennych objaśniających:





Analizując wykresy można stwierdzić, że występuje heteroskedastyczność składnika losowego. Aby się upewnić, zostały wykonane następujące testy:

#### Test Breuscha-Pagana:

H0 – składnik losowy jest homoskedastyczny

H1 – składnik losowy jest heteroskedastyczny

studentized Breusch-Pagan test
data: model3
BP = 20.894, df = 9, p-value = 0.01313

Mamy podstawy do odrzucenia H0 mówiącej o homoskedastyczności i przyjęciu H1 mówiącej o heteroskedastyczności. Decyzję podjęto przy 5% poziomie istotności.

#### Test White'a:

Została wykonana regresja pomocnicza bez interakcji:

H0 – składnik losowy jest homoskedastyczny

H1 – składnik losowy jest heteroskedastyczny

Statystyka testowa: 26.45087
p-value: 0.06662851

Mamy podstawy do odrzucenia H0 mówiącej o homoskedastyczności i przyjęciu H1 mówiącej o heteroskedastyczności. Decyzję podjęto przy 10% poziomie istotności.

Niewątpliwie występuje problem heteroskedastyczności składnika losowego.

### Model3

#### Niepoprawna postać funkcyjna modelu

Aby poradzić sobie z niepoprawną postacią funkcyjną modelu została zastosowana postać log-linear:

$$\ln \text{Lifeexp}_i = \beta_0 + \beta_1 \text{Schooling}_i + \beta_2 \text{HIV.AIDS}_i + \beta_3 \text{tuberculosis}_i + \beta_4 \text{healthexppercapita}_i + \beta_5 \text{fertilityrate}_i + \beta_6 \text{urbanpopgrowth}_i + \beta_7 \text{CO2}_i + \beta_8 \text{wateraccess}_i + \beta_9 \text{leastdeveloped}_i + \epsilon_i;$$

gdzie i to indeks obserwacji,  $\epsilon_i$  to składnik losowy



Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.17868472	0.05243499	79.693	< 0.0000000000000002	***
Schooling	0.00865770	0.00229363	3.775	0.000244	***
HIV.AIDS	-0.02635383	0.00363917	-7.242	0.0000000000366	***
tuberculosis	-0.00007094	0.00003091	-2.295	0.023351	*
healthexppercapita	0.00001804	0.00000402	4.487	0.0000159147934	***
fertilityrate	-0.03139977	0.00501348	-6.263	0.0000000052729	***
urbanpopgrowth	0.00659273	0.00249269	2.645	0.009197	**
CO2	-0.00166687	0.00071267	-2.339	0.020889	*
wateraccess	0.00073828	0.00040024	1.845	0.067406	.
leastdeveloped	-0.02299959	0.01148567	-2.002	0.047348	*
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.04009 on 128 degrees of freedom					
Multiple R-squared: 0.8961, Adjusted R-squared: 0.8887					
F-statistic: 122.6 on 9 and 128 DF, p-value: < 0.00000000000000022					

Wyniki testów z modelu3 i modelu4 zostały przedstawione w tabelce testy:

	model3	model4
pvaluetestRESET	0.01519	0.07264
pvaluetestJB	0.000003413	0.00000000000000022
pvaluetestBP	0.01313	0.01552

Wyniki testu RESET z modelu4 pokazują, że nie ma podstaw do odrzucenia H0 mówiącej o poprawnej postaci funkcyjnej modelu (przy poziomie istotności 5%). Dzięki przekształceniu w pewnym stopniu pozbyto się problemu niewłaściwej specyfikacji. Jednakże wynik testu p-value nie jest na tyle zadowalający, aby twierdzić, że problem nie występuje.

Patrząc na wyniki testów na heteroskedastyczność oraz normalność reszt, można jednoznacznie stwierdzić, że problem nie został rozwiązany.

### Współliniowość

Współczynnik inflacji wariancji:

Schooling	HIV.AIDS	tuberculosis	healthexppercapita	fertilityrate
3.299274	1.566328	2.133332	2.267710	3.563803
urbanpopgrowth	CO2	wateraccess	leastdeveloped	
1.954647	1.982841	3.266598	2.362251	

Wyniki VIF są porównywalne z wynikami z modelu2. Nie ma problemu współliniowości, bo VIF<10 w każdym przypadku.

### Uwzględnienie sferyczności macierzy wariancji kowariancji składnika losowego

Aby uwzględnić sferyczność składnika losowego zostały zastosowane następujące metody:

- 1) Odporny estymator wariancji-kowariancji White'a
- 2) Ważona metoda najmniejszych kwadratów

Porównanie wyników:

### Model3

(Intercept)	4.1786847161	0.0524349858	79.6927	< 2.2e-16	***
Schooling	0.0086576964	0.0022936316	3.7747	0.0002441	***
HIV.AIDS	-0.0263538314	0.0036391686	-7.2417	0.00000000003662	***
tuberculosis	-0.0000709375	0.0000309074	-2.2952	0.0233511	*
healthexppercapita	0.0000180356	0.0000040197	4.4868	0.00001591479341	***
fertilityrate	-0.0313997661	0.0050134817	-6.2631	0.00000000527289	***
urbanpopgrowth	0.0065927342	0.0024926855	2.6448	0.0091966	**
CO2	-0.0016668704	0.0007126683	-2.3389	0.0208889	*
wateraccess	0.0007382784	0.0004002358	1.8446	0.0674062	.
leastdeveloped	-0.0229995865	0.0114856665	-2.0025	0.0473484	*

### Model3 z odpornym estymatorem wariancji-kowariancji

(Intercept)	4.1786847161	0.0601222696	69.5031	< 2.2e-16	***
Schooling	0.0086576964	0.0021903644	3.9526	0.0001271	***
HIV.AIDS	-0.0263538314	0.0053297522	-4.9447	0.0000023454532	***
tuberculosis	-0.0000709375	0.0000355376	-1.9961	0.0480422	**
healthexppercapita	0.0000180356	0.0000026678	6.7604	0.0000000004388	***
fertilityrate	-0.0313997661	0.0064007522	-4.9056	0.0000027737122	***
urbanpopgrowth	0.0065927342	0.0020733252	3.1798	0.0018488	**
CO2	-0.0016668704	0.0004013117	-4.1536	0.0000593932992	***
wateraccess	0.0007382784	0.0004128377	1.7883	0.0760930	.
leastdeveloped	-0.0229995865	0.0128484970	-1.7901	0.0758083	.

### Ważona mnk

(Intercept)	4.19895801	0.04994251	84.076	< 0.0000000000000002	***
Schooling	0.00668228	0.00180499	3.702	0.000317	***
HIV.AIDS	-0.03075521	0.00538413	-5.712	0.0000000741324	***
tuberculosis	-0.00009009	0.00003126	-2.882	0.004642	**
healthexppercapita	0.00002002	0.00000269	7.445	0.00000000000126	***
fertilityrate	-0.02817443	0.00488622	-5.766	0.0000000575609	***
urbanpopgrowth	0.00793482	0.00181043	4.383	0.0000241729151	***
CO2	-0.00197446	0.00043773	-4.511	0.0000144441431	***
wateraccess	0.00074760	0.00041072	1.820	0.071065	.
leastdeveloped	-0.03563149	0.01141792	-3.121	0.002230	**

Jak widać, największą istotność mają zmienne w modelu z wykorzystaniem ważonej mnk. Jeśli chodzi o współczynniki w modelu3 i modelu3 z odpornym estymatorem wariancji-kowariancji wyniki są porównywalne. Natomiast wyniki estymacji w modelu wnmk są znacząco inne.

### Endogeniczność

Zmienną endogeniczną jest zmienna tuberculosis (zachorowalność na gruźlicę na 100tys osób). Im większa oczekiwana długość życia tym teoretycznie większe prawdopodobieństwo zachorowania na gruźlicę. Z drugiej strony zachorowalność na gruźlicę wpływa na obniżenie się oczekiwanej długości życia. Niestety ciężko znaleźć instrument, który wpływał by na zachorowalność, a nie wpływałby na oczekiwaną długość życia

## Ostateczne oszacowanie

Model, który najlepiej opisuje badane zjawisko, to model z oszacowaniem WLS. Zmienne tego modelu są najbardziej istotne, z tego względu ostateczna analiza będzie opierała się na oszacowaniach parametrów strukturalnych modelu za pomocą ważonej metody najmniejszych kwadratów.

### WLS

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.19895801	0.04994251	84.076	< 0.0000000000000002	***
Schooling	0.00668228	0.00180499	3.702	0.000317	***
HIV.AIDS	-0.03075521	0.00538413	-5.712	0.0000000741324	***
tuberculosis	-0.00009009	0.00003126	-2.882	0.004642	**
healthexppercapita	0.00002002	0.00000269	7.445	0.0000000000126	***
fertilityrate	-0.02817443	0.00488622	-5.766	0.0000000575609	***
urbanpopgrowth	0.00793482	0.00181043	4.383	0.0000241729151	***
CO2	-0.00197446	0.00043773	-4.511	0.0000144441431	***
wateraccess	0.00074760	0.00041072	1.820	0.071065	.
leastdeveloped	-0.03563149	0.01141792	-3.121	0.002230	**
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.02764 on 128 degrees of freedom					
Multiple R-squared: 0.8926, Adjusted R-squared: 0.885					
F-statistic: 118.1 on 9 and 128 DF, p-value: < 0.00000000000000022					

## Interpretacja oszacowań:

1. Jeśli oczekiwana długość edukacji wzrośnie o 1 rok to ceteris paribus oczekiwana długość życia wzrośnie przeciętnie o 0,67%.
2. Jeśli liczba zgonów dzieci w wieku 0-4 lat spowodowana HIV/AIDS na 1000 osób wzrośnie o 1 to ceteris paribus oczekiwana długość życia spadnie przeciętnie o 3%.
3. Jeśli zachorowania na gruźlicę na 100tyś osób wzrosną o 1 to ceteris paribus oczekiwana długość życia spadnie przeciętnie o 0,009%.
4. Jeśli uśrednione wydatki rządowe na służbę zdrowia wzrosną o 1\$ dolara według parytetu siły nabywczej to ceteris paribus oczekiwana długość życia wzrośnie o 0,002%.
5. Jeśli współczynnik płodności wzrośnie o 1 to ceteris paribus oczekiwana długość życia spadnie przeciętnie o 2,8%.
6. Jeśli uśredniony roczny wskaźnik wzrostu populacji w miastach wzrośnie o 1 punkt procentowy to ceteris paribus oczekiwana długość życia wzrośnie przeciętnie o 0,79%.
7. Jeśli uśrednione emisje CO2 per capita wzrosną o 1 tonę to ceteris paribus oczekiwana długość życia spadnie przeciętnie o 0,197%.
8. Jeśli udział osób w populacji mających dostęp do wody zdatnej do użytkowania wzrośnie o 1 punkt procentowy to ceteris paribus oczekiwana długość życia wzrośnie o 0,075%.
9. Jeśli kraj jest w grupie najmniej rozwiniętych to w porównaniu do krajów rozwijających się lub rozwiniętych oczekiwana długość jest niższa o 3,56%.

## **Współczynnik determinacji $R^2$**

$R^2 = 0.8926$ ;

89,26% zmienności zmiennej objaśnianej jest wyjaśniana zmiennością wartości teoretycznych wynikających z modelu.

**Wnioski z testu RESET, Breuscha Pagana na heteroskedastyczność, Jarque-Bera na normalność składnika losowego oraz z wyników VIF nie zmieniają się.**

## **Podsumowanie**

Z powodu braku możliwości zebrania bardziej szczegółowych danych model został wykonany dla mniejszej ilości zmiennych i krajów niż początkowo założyłem. W przyszłości podobną analizę można by było kontynuować dla większej ilości krajów i większej ilości bardziej szczegółowych zmiennych.

Głównym problemem przy oszacowaniu parametrów strukturalnych modelu okazała się heteroskedastyczność składnika losowego oraz niepoprawna postać funkcyjna modelu. Powyższe problemy zostały rozwiązane za pomocą ważonej metody najmniejszych kwadratów oraz zmianą postaci funkcyjnej do log-linear.

Na podstawie modelu i analizy danych można stwierdzić, że oczekiwana długość życia jest pozytywnie skorelowana ze średnią oczekiwaną długością edukacji w kraju, wydatkami rządowymi na służbę zdrowia per capita, wskaźnikiem wzrostu populacji w miastach oraz procentowym udziałem osób w populacji mających dostęp do wody zdatnej do użytkowania, natomiast negatywnie z umieralnością dzieci w wieku 0 – 4 lat na HIV/AIDS, zachorowaniami na gruźlicę, współczynnikiem płodności, emisjami CO2 per capita oraz z przynależnością do krajów najmniej rozwiniętych.

Cieężko porównać siłę zależności poszczególnych zmiennych na oczekiwaną długość życia, ponieważ nie są w tej samej skali.

Ponadto okazało się, że zmienna wateraccess jest dużo mniej istotna niż takie zmienne jak fertilityrate czy Schooling, co wydaje się zaskakujące, ponieważ wydawałoby się, że dostęp do wody zdatnej do użytkowania ma dużo większy wpływ na oczekiwaną długość życia niż wyżej wymienione zmienne.

W modelu występuje problem endogeniczności zmiennej tuberculosis, niestety bardzo ciężko znaleźć odpowiedni instrument, który pozwoliłby uwzględnić skorelowanie tej zmiennej ze składnikiem losowym.

## **Literatura**

„A STATISTICAL ANALYSIS OF LIFE EXPECTANCY ACROSS COUNTRIES USING MULTIPLE REGRESSION”  
(Miranda Chen, Michael Ching, 2000)

[https://www.seas.upenn.edu/~ese302/Projects/Project\\_2.pdf](https://www.seas.upenn.edu/~ese302/Projects/Project_2.pdf)

Lista krajów zawrta w analizie:

1) <i>Afghanistan</i>	2) <i>Albania</i>
3) <i>Algeria</i>	4) <i>Angola</i>
5) <i>Antigua and Barbuda</i>	6) <i>Armenia</i>
7) <i>Australia</i>	8) <i>Austria</i>
9) <i>Azerbaijan</i>	10) <i>Bahrain</i>
11) <i>Bangladesh</i>	12) <i>Barbados</i>
13) <i>Belarus</i>	14) <i>Belgium</i>
15) <i>Benin</i>	16) <i>Bhutan</i>
17) <i>Bosnia and Herzegovina</i>	18) <i>Botswana</i>
19) <i>Brazil</i>	20) <i>Bulgaria</i>
21) <i>Burkina Faso</i>	22) <i>Burundi</i>
23) <i>Cambodia</i>	24) <i>Cameroon</i>
25) <i>Canada</i>	26) <i>Central African Republic</i>
27) <i>Chad</i>	28) <i>Chile</i>
29) <i>China</i>	30) <i>Colombia</i>
31) <i>Costa Rica</i>	32) <i>Croatia</i>
33) <i>Cyprus</i>	34) <i>Denmark</i>
35) <i>Djibouti</i>	36) <i>Dominican Republic</i>
37) <i>Ecuador</i>	38) <i>El Salvador</i>
39) <i>Equatorial Guinea</i>	40) <i>Estonia</i>
41) <i>Ethiopia</i>	42) <i>Fiji</i>
43) <i>Finland</i>	44) <i>France</i>
45) <i>Gabon</i>	46) <i>Georgia</i>
47) <i>Germany</i>	48) <i>Ghana</i>
49) <i>Greece</i>	50) <i>Guatemala</i>
51) <i>Guinea</i>	52) <i>Guinea-Bissau</i>
53) <i>Guyana</i>	54) <i>Haiti</i>
55) <i>Honduras</i>	56) <i>Hungary</i>
57) <i>Iceland</i>	58) <i>India</i>
59) <i>Indonesia</i>	60) <i>Iraq</i>
61) <i>Ireland</i>	62) <i>Israel</i>
63) <i>Italy</i>	64) <i>Jamaica</i>
65) <i>Japan</i>	66) <i>Jordan</i>
67) <i>Kazakhstan</i>	68) <i>Kenya</i>
69) <i>Kiribati</i>	70) <i>Kuwait</i>
71) <i>Latvia</i>	72) <i>Lebanon</i>
73) <i>Lesotho</i>	74) <i>Liberia</i>
75) <i>Lithuania</i>	76) <i>Luxembourg</i>
77) <i>Madagascar</i>	78) <i>Malawi</i>
79) <i>Malaysia</i>	80) <i>Maldives</i>
81) <i>Mali</i>	82) <i>Malta</i>
83) <i>Mauritania</i>	84) <i>Mauritius</i>
85) <i>Mexico</i>	86) <i>Mongolia</i>
87) <i>Morocco</i>	88) <i>Myanmar</i>
89) <i>Namibia</i>	90) <i>Nepal</i>
91) <i>Netherlands</i>	92) <i>New Zealand</i>
93) <i>Nicaragua</i>	94) <i>Niger</i>
95) <i>Nigeria</i>	96) <i>Norway</i>
97) <i>Oman</i>	98) <i>Pakistan</i>
99) <i>Panama</i>	100) <i>Papua New Guinea</i>

101) <i>Paraguay</i>	102) <i>Peru</i>
103) <i>Philippines</i>	104) <i>Poland</i>
105) <i>Portugal</i>	106) <i>Qatar</i>
107) <i>Romania</i>	108) <i>Rwanda</i>
109) <i>Samoa</i>	110) <i>Sao Tome and Principe</i>
111) <i>Saudi Arabia</i>	112) <i>Senegal</i>
113) <i>Seychelles</i>	114) <i>Sierra Leone</i>
115) <i>Singapore</i>	116) <i>Slovenia</i>
117) <i>Solomon Islands</i>	118) <i>South Africa</i>
119) <i>Spain</i>	120) <i>Sri Lanka</i>
121) <i>Sudan</i>	122) <i>Suriname</i>
123) <i>Sweden</i>	124) <i>Switzerland</i>
125) <i>Tajikistan</i>	126) <i>Thailand</i>
127) <i>Timor-Leste</i>	128) <i>Togo</i>
129) <i>Tonga</i>	130) <i>Trinidad and Tobago</i>
131) <i>Tunisia</i>	132) <i>Turkey</i>
133) <i>Uganda</i>	134) <i>Ukraine</i>
135) <i>United Arab Emirates</i>	136) <i>Uruguay</i>
137) <i>Vanuatu</i>	138) <i>Zambia</i>