

„Od czego zależy oczekiwana długość życia człowieka?”

Najdłużej żyjącym człowiekiem, którego metryka została jednoznacznie potwierdzona, była Francuzka Jeanne Calment (1875–1997), która przeżyła 122 lata i 164 dni¹. Obecnie najstarszym człowiekiem świata jest Kane Tanaka, czyli japońska superstulatka. Kane Tanaka 2 stycznia skończyła 118 lat².



Jeanne Calment, fotografia, 1895 oraz 1995, https://pl.wikipedia.org/wiki/Jeanne_Calment

Długość ludzkiego życia od dawna była przedmiotem wielu badań. Na przykład badacze, badający szkielety znajdujące w Biskupinie wskazują, że około VIII wieku p.n.e. ludzie dożywali zazwyczaj do 35-tego roku życia³. Oczekiwana długość życia zaczęła wzrastać znacząco dopiero na początku XIX wieku we wczesnych krajach uprzemysłowionych (głównie na skutek większego dostępu do kanalizacji oraz rozwojowi medycyny), podczas gdy w pozostałych częściach świata pozostawała nadal na niskim poziomie. Doprowadziło to do bardzo dużych dysproporcji w oczekiwanej długości życia na świecie. W ciągu ostatnich dziesięcioleci ta globalna nierówność zmniejszyła się. Jednak wciąż można dostrzec duży kontrast między najmniejszą a największą oczekiwaną długością życia. W roku 2019 krajem o najniższej średniej długości życia była Republika Środkowoafrykańska z 53 latami, podczas gdy w Japonii średnia długość życia jest o ponad 30 lat dłuższa.

Ciekawostką jest fakt, że żaden kraj na świecie nie ma niższej średniej długości życia niż kraje o najwyższej średniej długości życia w 1800.⁴ Na jakiej podstawie można tak stwierdzić? Zanim przejdziemy do analizy, warto zastanowić się nad samym pojęciem oczekiwanej długości życia.

Oczekiwana dalsza długość trwania życia to wielkość statystyczna stosowana w demografii i matematyce ubezpieczeń życiowych. Miara oczekiwanej długości życia dostarcza ważnych informacji, między innymi o stanie zdrowia w danej populacji. Informacje te mogą być wykorzystane do celów aktuarialnych w planowaniu i kształtowaniu polityki. Są one również przydatne w identyfikacji populacji wysokiego ryzyka, dla których profilaktyczna opieka zdrowotna i opieka medyczna może zmniejszyć zachorowalność w ostatnich latach życia⁵.

Oczekiwana długość życia Jest równa średniej liczbie lat życia, jaka pozostała osobnikowi w danym wieku i z danej grupy. W szczególnym przypadku przy wieku równym 0 (czyli dla noworodka) oczekiwana dalsza długość trwania życia jest równa średniej długości życia w danej populacji. Zdaniem badaczy, rejestry, tablice trwania życia i modele mogą stanowić punkt wyjścia do zauważenia podstawowych zależności związanych z długością życia, jaka jest przeznaczona danym osobom.

Jednym z takich czynników, wpływających na długość życia może być na przykład alkohol. ”Kto dużo pije, ten krócej żyje”⁶ to tytuł artykułu, opublikowanego w 2018 roku przez magazyn Forbes. Artykuł wskazuje na fakt, że wypijając ponad pięć drinków w tygodniu możemy znacząco obniżyć średnią długość życia. Naukowcy podkreślają jednak, że nie trzeba całkowicie odstawić jasnego pełnego czy lampki czerwonego wina. Nawet Jeanne Calment czasami pozwalała sobie na kieliszek dobrego wina. Wystarczy zachować umiar i trzymać się wyznaczonych limitów spożycia. No właśnie, czym jest umiar? To pytanie pobudziło moje procesy myślowe do zastanowienia się, jaki jest optymalny poziom czynników, wpływających na średnią oczekiwaną długość życia.

¹ Jeanne Calment, https://pl.wikipedia.org/wiki/Jeanne_Calment

² Najstarsi ludzie, https://pl.wikipedia.org/wiki/Najstarsi_ludzie

³ Strona Internetowa Muzeum Archeologicznego w Biskupinie, ”Rezerwat Archeologiczny w Biskupinie”, <http://www.biskupin.pl/rezerwat-archeologiczny/>

⁴ Life expectancy at birth, total (years), <https://ourworldindata.org/life-expectancy>

⁵ Sidney Katz, M.D., Laurence G. Branch, Ph.D., Michael H. Branson, The New England Journal of Medicine, Active Life Expectancy, England, 17th November 1983, <https://www.nejm.org/doi/pdf/10.1056/NEJM198311173092005>

⁶ Kto dużo pije, ten krócej żyje, 15 kwiecień 2018, <https://www.forbes.pl/life/alkohol-skraca-zycie-naukowcy-sprawdzili-jak-bardzo/534vg78>,

Zastanowił mnie również fakt, że większość najstarszych ludzi na świecie pochodzi z krajów bogatych, jak Japonia oraz USA. Czy zatem istnieje związek przyczynowy między nierównością w dochodach a śmiertelnością? Czy dobrobyt rzeczywiście wydłuża ilość życia? Czy nałogi doprowadzają ludzi do przedwczesnej śmierci? Czy poziom rozwoju medycyny może skutecznie opóźnić naszą śmierć? W końcu ... Jak poziom edukacji wpływa na nasze życie? A to tylko kilka pytań, które postaram się rozwiązać, przeprowadzając analizę podstawowych czynników, wpływających na moją zmienną objaśnianą.

Oczywiście raport jest jedynie próbą zbadania grup tych czynników, pobieżną analizą. Na długość naszego życia mają wpływ tak proste do zmierzenia rzeczy, jak chociażby sytuacja ekonomiczna oraz tak trudne jak uwarunkowania genetyczne, poziom higieny oraz poziom szczęścia i zadowolenia z życia. Z powodu tego, że nie jestem w stanie opisać wszystkich tych czynników, wybrałam zmienne wpływające na długość życia pod względem **ekonomicznym, społecznym, edukacyjnym i zdrowotnym**. Każda zmienna jest **przedstawicielem** szerszego zakresu zmiennych.

Badanie oczekiwanej długości życia nie jest łatwe i po drodze możemy napotkać wiele problemów. Prawdziwym problemem jest na pewno fakt, że w przypadku najstarszych ludzi często brakuje dokumentu potwierdzającego ich rzeczywisty wiek. Kolejną trudnością jest to, że czynniki wpływające na długość życia takie, jak na przykład szczęście lub na przykład duchowość, są bardzo ciężkie do zmierzenia. Dodatkowo, dane mogą pochodzić z różnych źródeł, być przedstawione w różnych jednostkach i wymagać licznych przekształceń moim raportcie będę więc „walczyć” z różnymi problemami oraz analizować, jak mogę je wyeliminować.

2. Zbiór danych

W mojej analizie wykorzystałam dane, pochodzące z różnych ogólnodostępnych baz danych, głównie pochodzących z Data World Bank. Dane, których użyłam pochodzą z roku 2016. Zasadność takiego wyboru uzasadniam faktem, że zbiory z lat późniejszych posiadają liczne braki danych. W danych nie dokonałam podziału ze względu na płeć.

Dane zostały pierwotnie zebrane dla 218 krajów świata, po usunięciu braków danych pozostały mi 172 obserwacje.

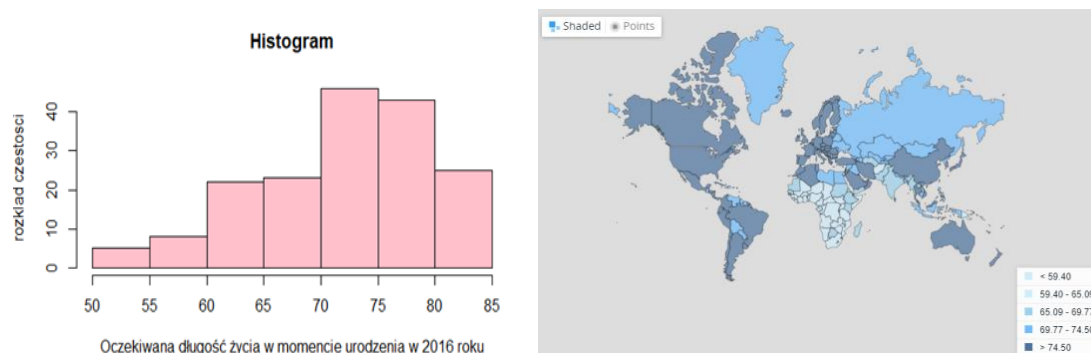
2.1. Opis zmiennych, prezentacja ich histogramów i wizualizacja na mapie

a. Oczekiwana długość życia – zmienna objaśniana

Oznaczenie w modelu: *lifeexp*

Źródło danych oraz grafiki (mapy): Data World Bank⁷

Oczekiwana długość życia w momencie urodzenia w 2016 roku dla noworodka to nic innego, jak liczba lat, które przewidują tabele umieralności w roku jego urodzenia. Na przykład średnia oczekiwana długość życia noworodka urodzonego w 2016 roku w Polsce wynosi 78 lat a średnia długość życia noworodka urodzonego w 1999 roku w Bułgarii wynosi 71 lat. Warto dodać, że oczekiwana długość życia pokazuje liczbę lat, które ma przed sobą noworodek tylko w momencie gdy wzorce umieralności pozostaną niezmienione przez całe jego życie. Miara ta jak widzimy jest więc tylko poglądowa, bo zazwyczaj zachodzą zmiany w owych wzorcach, na przykład na skutek znaczącego rozwoju medycyny. Przykładowo w 1900 roku średnia długość życia w USA wynosiła około 47 lat, zaś obecnie jest to około 77 lat. W Indiach w 1950 roku średnia długość życia wynosiła około 32 lata, a w roku 2000 osiągnęła 64 lata⁸.



Interpretując histogram oczekiwanej długości życia, zauważamy najpierw z pewnością, że w przeważającej większości oczekiwana długość życia przekracza 65 lat. Po wykresie można się spodziewać jednak, że istnieją kraje (około 15% wszystkich krajów, więc znaczna mniejszość), w których możemy spodziewać się życia krótszego niż 60 lat. Aby zobrazować tę zależność, załączam grafikę.

Na mapie widać, że zdecydowanie najniższą oczekiwaną długość życia możemy zaobserwować na kontynencie afrykańskim oraz w Indiach. Najwyższą oczekiwaną długość życia możemy zaobserwować zaś głównie w Europie, Ameryce Północnej oraz w niektórych krajach azjatyckich (między innymi Chiny i słynąca z największej liczby superstulatków Japonia).

⁷ Life expectancy at birth, total (years), <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>

⁸ Oczekiwana dalsza długość trwania życia, https://pl.wikipedia.org/wiki/Oczekiwana_dalsza_d%C5%82ugo%C5%9B%C4%87_trwania_%C5%BCyca

b. GRUPA: czynnik ekonomiczny

PKB per capita – zmienna objaśniająca

Oznaczenie w modelu: *lifeexp*

Źródło danych oraz grafiki: Data World Bank⁹

Jednostka: USD

PKB per capita to produkt krajowy brutto podzielony przez liczbę ludności. Jeśli ludzie chcą wiedzieć, które kraje są najbogatsze, zazwyczaj patrzą na wskaźniki produkcji lub dochodów. Międzynarodowy Fundusz Walutowy (MFW) prowadzi pomiar produktu krajowego brutto (PKB) na osobę, ważonego siłą nabywczą dolara. Na przykład w 2020 roku według tej miary prowadził Luksemburg, wyprzedzając Singapur i Katar.

International Monetary Fund (2020 estimates)^[5]

Rank	Country/Territory	Int\$
1	Luxembourg	112,875
2	Singapore	95,603
3	Qatar	91,897

List of countries by GDP (PPP) per capita, [https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(PPP\)_per_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita)

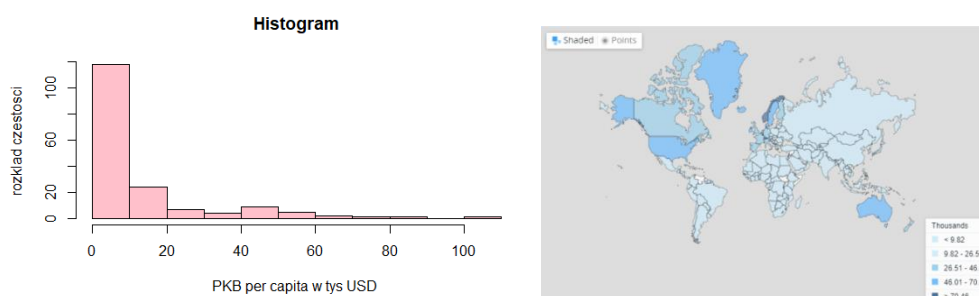
Możemy więc dostrzec pewną prawidłowość. Kraje z dużymi zasobami surowców naturalnych lub centra finansowe zajmują większość z czołowych pozycji w takim zestawieniu.

PKB jako pewna miara jest jednak daleki od ideału. Pokazuje bowiem przepływ dóbr i usług, które produkuje się w danym państwie w ciągu roku. Nie mówi natomiast absolutnie nic o bogactwie, na które składają się budynki, maszyny, surowce, paliwa, czysta woda, zdrowe ekosystemy, umiejętności, jakość edukacji i instytucje dobrego zarządzania. Nie jest to współczynnik idealny także z tego powodu, ponieważ do PKB wliczana jest na przykład szara strefa.

Nie można też mówić w sposób jednoznaczny o dochodzie na głowę, bo to przecież nie jest pensja czy jakiś przelew przychodzący na indywidualne konto. Trzeba też pamiętać, że należałoby wliczać pewne dywidendy czy elementy transferów finansowych pochodzących z zagranicy, a wówczas nie jesteśmy w stanie takich danych przełożyć na kwotę, którą otrzymuje bezpośrednio obywatel¹⁰.

Co więcej, bogate w ropę kraje, takie jak Katar, mogą nieźle funkcjonować, ale stopniowo pozbywają się swoich bogactw naturalnych. Z kolei kraje o niższych dochodach, takie jak Chiny, mogą budować infrastrukturę, z której będą czerpać pełne korzyści dopiero za kilka dekad. Więc przewidywalność tego indykatora jest daleka od ideału.

Jednak zdecydowałam się na użycie tego czynnika, ponieważ ma on wiele plusów. Jest on stosunkowo łatwy do policzenia, wiarygodny, łatwy do porównania w czasie i pomiędzy krajami, regionami. Stanowi też łatwo dostępne źródło informacji na temat gospodarki. Ponadto jest on częsty w licznych modelach ekonometrycznych, których zadaniem nie jest może najdokładniejsze przedstawienie sytuacji, lecz przedstawienie pewnego wiarygodnego trendu.



Na podstawie histogramu i grafiki można zauważać, że najwięcej krajów na świecie ma PKB per capita nie przekraczające 10 tys. Dolarów. Wyższym PKB niż 20 tys dolarów mogą się zaś pochwalić kraje głównie w Europie i Ameryce Północnej oraz Australia. To pokazuje zasadność użycia właśnie PKB per capita w mojej analizie, nie zaś PKB. Wyjaśnię to na przykładzie Chin. Wartość PKB w Chinach w 2016 r. wyniosła 11 221 836 mln dolarów, pozostawiając Chiny na 2. miejscu w rankingu PKB spośród badanych krajów. Było to z pewnością spowodowane ogromną populacją Chin. Aby uczynić więc ten miernik bardziej wiarygodnym, podzielono go przez liczbę mieszkańców. I otrzymany w ten sposób PKB per capita wyniósł już tylko 8000 dolarów na osobę, więc nie był wcale wyróżniający.

⁹ GDP per capita (current US\$), <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

¹⁰ James P. Smith, Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status, Journal Of Economic PERSPECTIVES, VOL. 13, NO. 2, SPRING 1999, (pp. 145-166)

C) GRUPA: czynnik zdrowotny

Wydatki rządu na służbę zdrowia per capita –zmienna objaśniająca

Oznaczenie w modelu: *healthexp*

Źródło danych oraz grafiki(mapa): Data World Bank¹¹

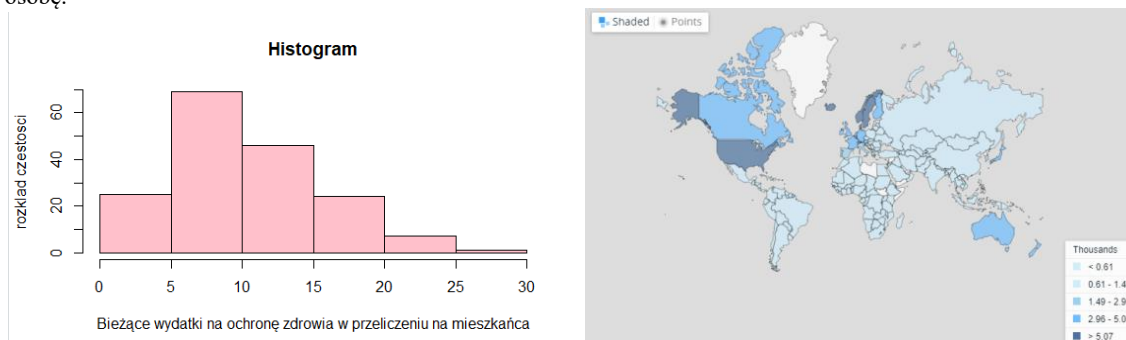
Jednostka: USD

Szacunki bieżących wydatków na ochronę zdrowia per capita obejmują wydatki na towary i usługi zdrowotne konsumowane w ciągu każdego roku na jednego mieszkańca w danym kraju.

Światowa Organizacja Zdrowia (WHO) co roku publikuje dane, dotyczące wydatków na ochronę zdrowia. Światowa Baza Danych o wydatkach na ochronę zdrowia WHO w nowej wersji stanowi źródło odniesienia dla międzynarodowych porównań wydatków na ochronę zdrowia. Wyjaśnia również mechanizmy finansowania i wprowadza nowe wymiary, które poprawiają porównywalność wydatków na ochronę zdrowia w perspektywie powszechnego zasięgu ochrony zdrowia.

Warto zaznaczyć, że wzmocnienie finansowania ochrony zdrowia jest jednym z celów zrównoważonego rozwoju. Istotne różnice w danych, dotyczących wydatków na ochronę zdrowia określają kluczowe kwestie, takie jak słabe i mocne strony oraz obszary, które wymagają inwestycji, na przykład wyrażają potrzebę sfinansowania dodatkowych placówek służby zdrowia lub dodatkowych szkoleń dla kadry medycznej. Finansowanie ochrony zdrowia ma również decydujące znaczenie dla osiągnięcia powszechnego zasięgu opieki zdrowotnej, ogólnego rozwoju społecznego a co za tym idzie wysokiej oczekiwanej długości życia

Na podstawie analizy histogramu oraz mapy, przedstawiającej ten indyktor w różnych krajach stwierdzamy, że w przeważającej części krajów wydatki na służbę zdrowia per capita są mniejsze niż 10000 dolarów amerykańskich. W niewielkiej ilości krajów wydatki te są większe. W zaledwie kilku z nich, sięgają 10-15 tysięcy dolarów amerykańskich na osobę.



d) GRUPA: czynnik demograficzny

Śmiertelność niemowląt (na 1000 żywych urodzeń) –zmienna objaśniająca

Oznaczenie w modelu: *infantmort*

Źródło danych oraz grafiki(mapa): Data World Bank¹²

Śmiertelność niemowląt to liczba zgonów niemowląt przed ukończeniem pierwszego roku życia na 1000 żywych urodzeń w danym roku.

Minusem tej zmiennej jest fakt, że dane ankietowe są często obarczone błędem, a ankiety szacujące zgony niemowląt dzieci wymagają dużych prób. Ponadto w krajach ze słabo rozwiniętą administracją i ochroną zdrowia (np. obecność szpitali), wiele zgonów noworodków nie jest nawet nigdzie zgłaszanych. Pośrednie szacunki opierają się na tabelach trwania modeli, które mogą być często nieaktualne.

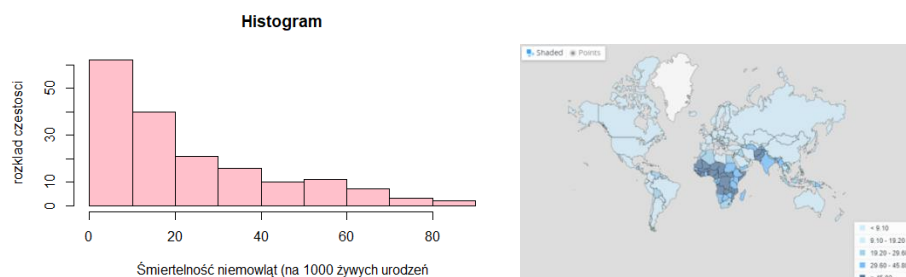
Współczynnik umieralności ma jednak wiele plusów. Jest on ważnym wskaźnikiem stanu zdrowia w danym kraju. Ponieważ dane dotyczące zachorowalności i chorobowości są często niedostępne, wskaźniki umieralności są często wykorzystywane do identyfikacji populacji narażonych na zagrożenia. Należą one również do wskaźników najczęściej wykorzystywanych do porównywania rozwoju społeczno-gospodarczego różnych krajów oraz w mierzeniu oczekiwanej długości życia.

¹¹ Current health expenditure per capita (current US\$), <https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD>

¹² Mortality rate, infant (per 1,000 live births), <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>

Światowa Organizacja Zdrowia (WHO), Bank Światowy, Wydział Ludnościowy ONZ oraz inne uniwersytety i instytuty badawcze, opracowała i przyjęła metodę statystyczną, która wykorzystuje wszystkie dostępne informacje w celu wyrównania wyżej wymienionych różnic i nieprawidłowości. Metoda ta wykorzystuje modele statystyczne w celu uzyskania najlepszego oszacowania linii trendu występowania tego zjawiska. Dlatego też wykorzystywałam właśnie te dane, dostępne również na stronie Data World Bank.

Na podstawie analizy histogramu oraz mapki widzimy, że dominuje dość mała śmiertelność niemowląt w większości państw. Współczynnik ten nie przekracza zazwyczaj 20 śmierci noworodków na 1000 żywych urodzeń. Wskaźnik ten rośnie głównie w krajach afrykańskich.



e) GRUPA: czynnik edukacyjny

Stosunek liczby uczniów do liczby nauczycieli w szkołach podstawowych – zmienna objaśniająca

Oznaczenie w modelu: *teacher*

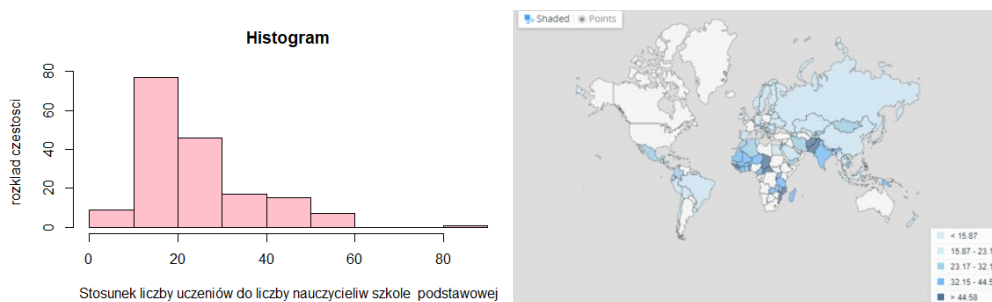
Źródło danych: Data World Bank¹³

Stosunek liczby uczniów do liczby nauczycieli w szkole podstawowej określa średnią liczbę uczniów na jednego nauczyciela w szkole podstawowej. Stosunek uczeń-nauczyciel jest często używany do porównywania jakości nauczania w różnych krajach.

Stosunek ten obliczany jest poprzez podzielenie liczby uczniów na podstawowym poziomie wykształcenia przez liczbę nauczycieli na tym samym poziomie kształcenia. Dane dotyczące wykształcenia są gromadzone przez Instytut Statystyki UNESCO na podstawie oficjalnych odpowiedzi w corocznym badaniu edukacji. Wszystkie dane są przyporządkowane do Międzynarodowej Standardowej Klasyfikacji Edukacji (ISCED) w celu zapewnienia porównywalności programów edukacyjnych na poziomie międzynarodowym.

Jednak są też minusy tej zmiennej. Na błędy w porównywalności proporcji między uczniami a nauczycielami w różnych krajach wpływają między innymi różne praktyki stosowane w poszczególnych krajach, jak między innymi klasy wielostopniowe. Co więcej, kraje słabiej rozwinięte pod kątem gromadzenia danych i administracji są narażone na różne błędy w sprawozdawczości.

Na podstawie grafik można zauważyć, że ogólny stan tego czynnika jest całkiem dobry, w większości krajów stosunek liczby uczniów do nauczycieli nie przekracza 20 osób. Jednak są takie obszary, głównie Afryka (jedne z wyższych współczynników - 40 uczniów na nauczyciela możemy zaobserwować m.in. w Zambii i Zimbabwie oraz na bliskim wschodzie, głównie w krajach dotkniętych wojnami, jak Afganistan).



¹³ Pupil-teacher ratio, primary, <https://data.worldbank.org/indicator/SE.PRM.ENRL.TC.ZS>

f) **GRUPA:** czynnik społeczny, nałogi

Odsetek dorosłych pijących alkohol – zmienna objaśniająca

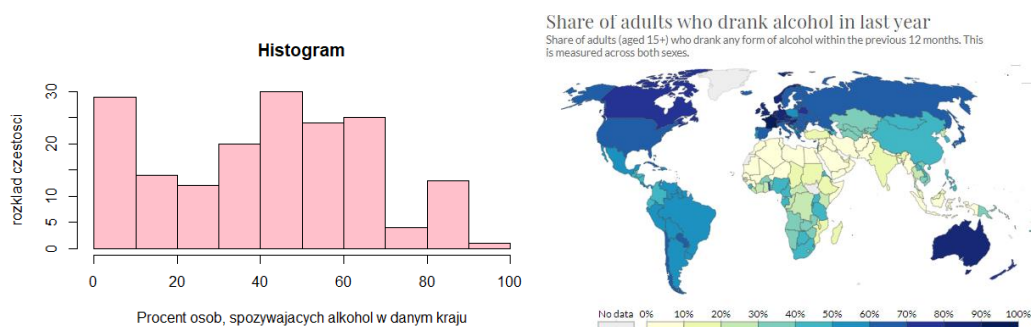
Oznaczenie w modelu: *alcohol*

Źródło danych: World Health Organisation¹⁴

Naukowo udowodniono, że spożywany w nadmiarze alkohol zwiększa ciśnienie tętnicze i powoduje niekorzystne zmiany w naczyniach krwionośnych. Alkohol obniża napięcie dolnego zwieracza przełyku. Tym sposobem sprzyja rozwojowi oraz zaostrzeniu refluksu żołądkowo – przełykowego. Globalne spożycie alkoholu jest przyczyną 2,8 mln przedwczesnych zgonów rocznie. Jednak mimo to alkohol nadal jest bardzo popularną używką w wielu krajach. Jednak nie we wszystkich. Między innymi w krajach muzułmańskich, gdzie spożycie alkoholu jest religijnie zabronione, możemy zaobserwować bardzo niski odsetek dorosłych, pijących alkohol. Są to często turyści – przykład Malediwów, gdzie odsetek jest głównie powiększony przez turystów.

Odsetek osób dorosłych pijących alkohol jest zdecydowanie najwyższy w Europie Zachodniej i Australii. Najwyższy jest we Francji: w 2010 roku blisko 95 procent dorosłych we Francji piło alkohol w poprzednim roku. Ponownie, rozpowszechnienie picia w Afryce Północnej i na Bliskim Wschodzie jest znacznie niższe niż gdzie indziej. Zwykle 5-10 procent dorosłych w tych regionach piło w ciągu poprzedniego roku, a w wielu krajach było to poniżej 5 procent. Widzimy również szczególnie wysoki poziom abstynencji alkoholowej w Afryce Północnej i na Bliskim Wschodzie. W większości krajów tego regionu ponad 80 procent (często ponad 90 procent) nigdy nie piło alkoholu.

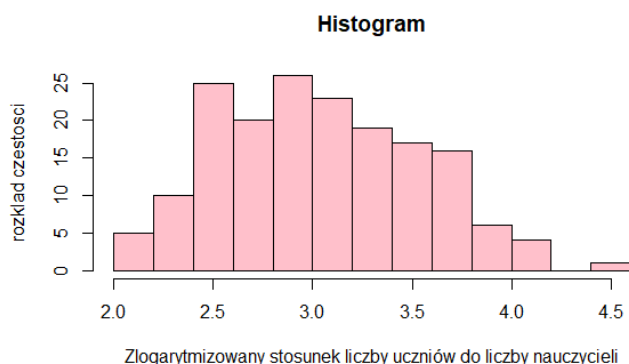
Na histogramie widać, że zazwyczaj odsetek osób, spożywających alkohol jest bardzo niski (kiedy jest to np. kraj muzułmański) lub umiarkowany (prawdopodobnie kiedy na odsetek nie wpływają względy religijne – np. w krajach europejskich często odsetek ten waha się w okolicach 40% dorosłych, spożywających alkohol).



2.2. Transformacja wybranych danych

Dokonałam tylko jednej transformacji. Zlogarytmizowałam zmienną „**Stosunek liczby uczniów do liczby nauczycieli w szkołach podstawowych**”. Dzięki temu mogę zmniejszyć skośność rozkładu, wpływ zmiennych odstających oraz będę mogła używać porównań w skali procentowej. Dodatkowo poprawi to istotność parametrów w moim modelu.

Histogram zmiennej *teacher* po jej zlogarytmizowaniu.



Tak jak wspomniałam, skośność rozkładu została wyraźnie zniwelowana.

¹⁴ Alcohol Consumption 2016, <https://www.who.int/news-room/fact-sheets/detail/alcohol>

3. Model ekonometryczny

3.1. Zapis modelu ekonometrycznego

Równanie dla modelu opisującego badane przeze mnie zjawisko:

$$lifeexp_i = \alpha_0 + \beta_1 healthexp_i + \beta_2 alcohol_i + \beta_3 GDP_i + \beta_4 infantmort_i + \beta_5 \log(teacher)_i$$

3.2. Oczekiwanie w stosunku do kierunku zależności

Na podstawie wcześniej przytoczonych histogramów i map intensywności zmiennych, oraz własnej intuicji, oczekuję, że oszacowania parametrów strukturalnych przy zmiennych *healthexp*, *GDP* będą dodatnie, natomiast przy *log(teacher)*, *infantmort*, *alcohol* będą ujemne.

4. Analiza ekonometryczna

Przyjmuję poziom istotności $\alpha=0,05$.

4.1. Oszacowania parametrów strukturalnych modelu

Wyniki na podstawie wydruku z programu R z oszacowaniami parametrów strukturalnych modelu metodą najmniejszych kwadratów:

Intercept	Estimate	p-value
Intercept	81.582435	<2e-16
Healthexp	0.144647	0.00255
alcohol	-0.017213	0.07627
GDP	0.077677	9.28e-08
Infantmort	-0.266134	<2e-16
log(teacher)	-1.804483	0.00628

Oszacowane parametry w modelu mają postać (zaokrągliłam liczby do 3 miejsc po przecinku):

$$lifeexp_i = 81.582 + 0.145 * healthexp_i - 0.017 * alcohol_i + 0.777 * GDP_i - 0.266 * infantmort_i - 1.804 * \log(teacher)_i$$

4.1.2. Istotność parametrów

Testowanie istotności parametrów => test t-studenta:

H_0 : Zmienna f_0 jest f_0 nieistotna f_0 statystycznie.

H_1 : Zmienna f_0 jest f_0 istotna f_0 statystycznie.

Na poziomie istotności $\alpha=0,05$, dla zmiennych *healthexp*, *GDP*, *infantmort* oraz *log(teacher)* wartości krytyczne są mniejsze od α . Odrzucam więc hipotezę zerową mówiącą o nieistotności parametrów, na rzecz hipotezy alternatywnej mówiącej o tym, że parametry przy zmiennych są istotne statystycznie. Jedynym wyjątkiem jest zmienna *healthexp*, która stanie się istotna dopiero na poziomie istotności $\alpha=0.1$. Jednak ze względu na dość dużą próbę (172 obserwacje) i fakt, że zmiennej mało brakuje do bycia istotną na poziomie $\alpha=0,05$, zostawię ją w modelu oraz zinterpretuję.

Dodatkowo, na podstawie wydruku z programu R, mogę sprawdzić jednoczesną istotność podzbioru zmiennych objaśniających za pomocą uogólnionego testu Walda.

H_0 : Wszystkie f_0 zmienne f_0 jest f_0 łącznie f_0 nieistotne f_0 statystycznie.

H_1 : Istnieje f_0 co f_0 najmniej f_0 jedna f_0 zmienna, która f_0 jest f_0 istotna f_0 statystycznie.

Statystyka F wynosi 304.8, wartość krytyczna jest mniejsza niż 0.05 (wynosi <2.2e-16). Na poziomie istotności 0.05 odrzucam więc hipotezę zerową, mówiącą o łącznej nieistotności na rzecz hipotezy alternatywnej mówiącej o tym, że co najmniej jedna ze zmiennych jest istotna statystycznie (co potwierdza poprzednie wnioski z testowania poszczególnych zmiennych).

4.1.2. Interpretacja parametrów

PKB per capita –zmienna objaśniająca

Wzrost PKB na jednego mieszkańca o 1 USD, wiąże się ze wzrostem oczekiwanej długości życia o 0.077 roku czyli średnio 28,105 dnia, ceteris paribus.

Wydatki rządu na służbę zdrowia per capita –zmienna objaśniająca

Wzrost wydatków rządu na służbę zdrowia na jednego mieszkańca o 1USD, zwiększa oczekiwaną długość życia średnio o 0.144647 lat ,czyli 52,7 dnia, ceteris paribus.

Śmiertelność niemowląt (na 1000 żywych urodzeń) –zmienna objaśniająca

Wzrost śmiertelności niemowląt o jedno niemowlę na 1000 żywych urodzeń, powoduje spadek oczekiwanej długości życia o około 0.266134 roku czyli 97,14 dni, ceteris paribus.

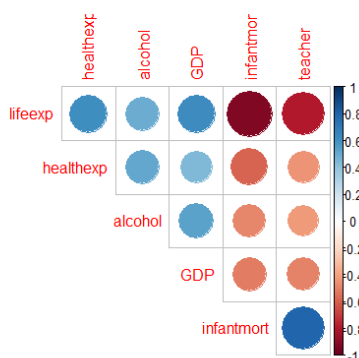
Stosunek liczby uczniów do liczby nauczycieli w szkole podstawowej –zmienna objaśniająca

Jeśli stosunek liczby uczniów do liczby nauczycieli w szkole podstawowej wzrośnie o 1%, spowoduje to spadek oczekiwanej długości życia o 0.018 lat(6,57 dni), ceteris paribus.

Odsetek dorosłych pijących alkohol –zmienna objaśniająca

Wzrost odsetka osób dorosłych, pijących alkohol na świecie o 1/100 spowoduje spadek oczekiwanej długości życia o 0.017213lat(6,28dni), ceteris paribus.

4.2 Korelogram i analiza korelacji poszczególnych zmiennych



Zdecydowałam się na przedstawienie korelogramu tuż po zinterpretowaniu parametrów rozkładu, ponieważ możemy porównać sobie wartości, które wyszły nam z modelu z korelogramem. Na pewno trzeba się zgodzić, że największa oczekiwana długość życia jest najsilniej skorelowana ze zmiennymi :infantmort oraz healthexpectancy, tyle, że w przeciwnych kierunkach- zmienna infantmort powoduje spadek oczekiwanej długości życia a zmienna healthexp jej wzrost.

Z powyższych interpretacji zmiennych w modelu wyszło, że nasze życie może wydłużyć się się nawet o około 50dni, kiedy zwiększymy wydatki rządu na służbę zdrowia na jednego mieszkańca o 1 USD, natomiast może się też skrócić statystycznie o około 100 dni ,kiedy nastąpi wzrost śmiertelności niemowląt średnio o jedno niemowlę na 1000 żywych urodzeń. Widzimy więc, jak ważną rolę pełni opieka zdrowotna w procesie przedłużania naszego życia.

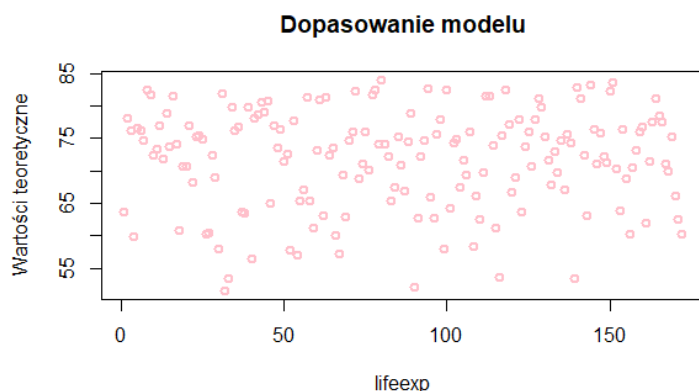
Śledząc korelogram możemy potwierdzić zasadność wyboru takich a nie innych zmiennych, ponieważ są one dość silnie skorelowane ze zmienną objaśnianą.

4.3. Dopasowanie wartości teoretycznych do obserwacji empirycznych

Do opisanie miary dopasowania wartości teoretycznych wynikających z oszacowań parametrów strukturalnych modelu ekonometrycznego do danych empirycznych, posłużę się współczynnikiem determinacji R^2 . Wynosi on 0.9018. Oznacza to, że przy danym zestawie obserwowanych danych, model tłumaczy 90,18% zmienności zmiennej objaśnianej *lifeexp*, zmiennymi objaśniającymi. Skorygowany współczynnik determinacji R^2 , który „karze” model za ilość zmiennych wynosi 89.88 %. Duża rozbieżność między R^2 a skorygowanym R^2 mogłaby sugerować, że w modelu jest zbyt dużo zmiennych nieistotnych z punktu widzenia przewidywalnego wyniku. Natomiast miała rozbieżność, czyli sytuacja w tym modelu, jest jak najbardziej poprawna.

Jednak nie wolno zapomnieć o falkie „nadmiernego dopasowania” modelu. Naszym celem jest przecież model, który trafnie odzwierciedla zależności populacyjne. A nasz model szacujemy jednak na podstawie próby, nie całej populacji. Jest więc szansa, że np. w wielu państwach, w których spożycie alkoholu jest duże, oczekiwana długość życia również może być wysoka co logicznie nie powinno mieć miejsca. Otrzymane przez nas wyniki R^2 oraz skorygowanego R^2 są dość wysokie, nawet za wysokie, biorąc pod uwagę charakter danych – dla danych przekrojowych najbardziej zadowalający byłby współczynnik $R^2 \in (0.30; 0.70)$.

Dlatego warto oprócz analizy tych współczynników, warto spojrzeć jeszcze na porównanie dopasowania teoretycznej zmiennej objaśnianej oszacowanej przez model do objaśnianej zmiennej empirycznej na wykresie.



Na podstawie wykresu widzimy, że dopasowanie nie jest idealne, niektóre wartości empiryczne nie są tożsame z teoretycznymi (np. dla oczekiwanej długości życia 20 lat). Moim zdaniem jednak dopasowanie jest na tyle satysfakcjonujące, by przeprowadzać dalsze analizy bez korygowania modelu. Dla najczęstszej oczekiwanej długości życia, występującej na świecie (w okolicach 50-70 lat), wartości teoretyczne najczęściej pokrywają się z wartościami empirycznymi.

4.4. Postaci funkcyjna modelu

Do przetestowania poprawności funkcyjnej modelu użyję testu RESET, sprawdzającego czy dodanie do modelu nowych zmiennych (np. ich wyższych potęg zmiennych) jest istotne statystycznie i poprawi dopasowanie modelu.

H_0 : Model ma poprawną specyfikację.

H_1 : Model nie ma poprawnej specyfikacji.

Wykorzystuję *RESET* test dostępny w pakiecie *lmtest*.

Statystyka **F** testu wynosi **0,59144** a krytyczny poziom istotności **0,5547**

Na poziomie istotności $\alpha=0,05$, jest brak podstaw do odrzucenia hipotezy zerowej mówiącej o tym, że model ma poprawną specyfikację.

4.5. Współliniowość zmiennych objaśniających

Współliniowość zmiennych objaśniających identyfikuję za pomocą czynnika inflacji wariancji (CIF). Testuje on, które zmienne objaśniające są powiązane z innymi, porównując sytuację faktyczną z idealną, w której wszystkie zmienne objaśniające są ze sobą nieskorelowane. Jeśli wartość CIF będzie wynosiła powyżej 10, będzie to sugerowało problem współliniowości.

Do sprawdzenia występowania współliniowości zmiennych w programie R wykorzystuję test *vif*, dostępny w pakiecie *car*.

Tabela na podstawie wydruku z programu R:

ZMIENNA	helathexp	alcohol	GDP	infantmort	log(teacher)
VIF	1.727936	1.657910	1.777511	3.303034	3.082100

Wartości VIF dla wszystkich zmiennych objaśniających są mniejsze od 10, więc problem współliniowości nie występuje w tym modelu.

4.6. Własności składnika losowego

4.6.1. Normalność składnika losowego

Normalność rozkładu składnika losowego ułatwia konstrukcję testów statystycznych przydatnych do weryfikacji modelu ekonometrycznego. Przetestuję ją za pomocą testu Jarque-Bera. Opiera się on na weryfikacji podobieństwa trzeciego i czwartego momentu rozkładu składnika losowego modelu do znanych wartości tych momentów w rozkładzie normalnym.

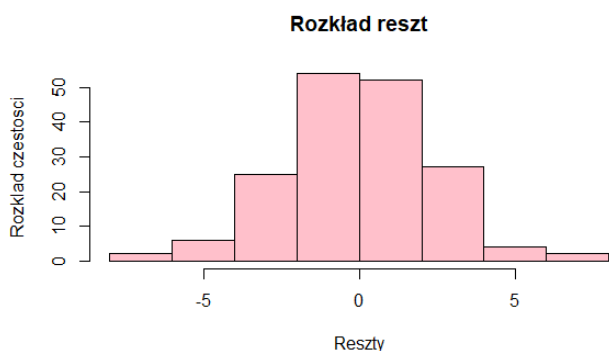
H_0 : Składnik losowy ma rozkład normalny.

H_1 : Składnik losowy nie ma rozkładu normalnego.

W programie R wykorzystuję *jarque.bera.test* dostępny w pakiecie *tseries*.

Statystyka testu chi-kwadrat wynosi 3.8221, a krytyczny poziom istotności 0,1479

Na poziomie istotności $\alpha = 0,05$ nie ma podstaw do odrzucenia hipotezy zerowej, mówiącej o normalności składnika losowego



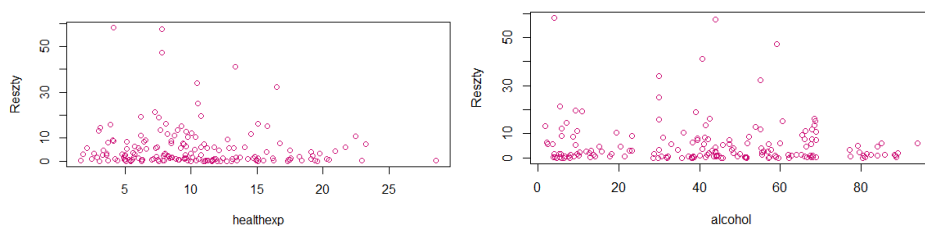
Wykres prezentujący rozkład reszt składnika losowego. Jest on zbliżony do rozkładu normalnego. Możemy więc powiedzieć (co jest zgodne z przeprowadzonym przed chwilą testem), że składnik losowy ma rozkład normalny.

4.6.2. Heteroskedastyczność

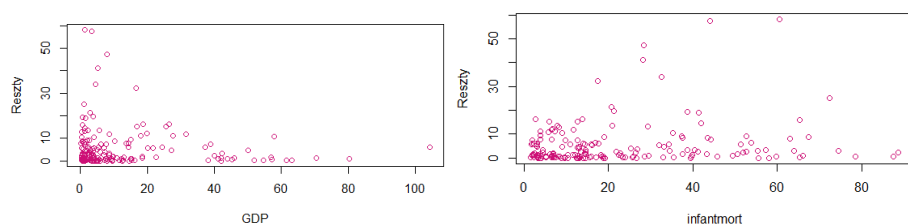
Dosyć częstym problemem towarzyszącym danym przekrojowym jest heteroskedastyczność składnika losowego, czyli sytuacja, gdy składniki losowe modelu mają różne wariancje, zaburza założenie o największej efektywności estymatora MNK.

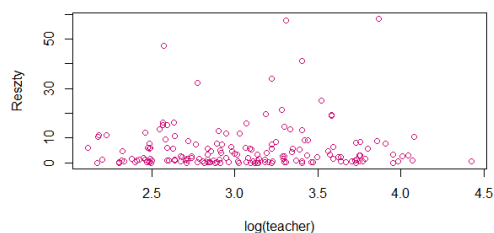
4.6.2.1. Wykresy reszt zmiennych objaśniających

Pierwszym sposobem na próbę identyfikacji problemu heteroskedastyczności składnika losowego, może być przyjrzenie się wykresom reszt zmiennych objaśniających



Reszty są zdecydowanie homoskedastyczne, czyli odchylają się od zera o taką samą wartość. Są małe wyjątki, pojedyncze punkty na wykresie odchylają się o zdecydowanie większą wartość (najwięcej przy zmiennej *infantmort*, której za chwilę dalej przyjrzymy się by dowiedzieć się czy rzeczywiście mamy w jej przypadku do czynienia z heteroskedastycznością).





4.6.2.2. Test Breuscha-Pagana

Drugim sposobem na sprawdzenie, czy w modelu występuje problem heteroskedastyczności jest test Breuscha-Pagana, który polega na oszacowaniu równania regresji, w którym zmienną objaśniającą jest kwadrat reszt podzielonych przez odchylenie standardowe.

H_0 : Składnik losowy jest homoskedastyczny.

H_1 : Składnik losowy jest heteroskedastyczny.

W programie R korzystam z *bptest*, dostępnego w pakiecie *lmtest*.

Statystyka testu chi-kwadrat wynosi 6.2588, a krytyczny poziom istotności 0,2818.

Przy poziomie istotności $\alpha = 0,05$ nie ma więc powodu do odrzucenia hipotezy zerowej, mówiącej o homoskedastyczności składnika losowego. Nie występuje więc problem heteroskedastyczności.

4.6.2.3. Test Goldfelda-Quandt

Test Gold-Quandt polega na porównaniu wariancji w dwóch podzielonych grupach. Pozwala nam on na wskazanie czy zmienna co do której mieliśmy podejrzenia, wpływa na istnienie zjawisko heteroskedastyczności składnika losowego.

H_0 : Składnik losowy jest homoskedastyczny.

H_1 : Składnik losowy jest heteroskedastyczny.

Wykorzystuję *gqtest* z pakietu *lmtest*. Tabelka na podstawie wydruku z programu R:

ZMIENNA	P-VALUE	CZY ZMIENNA WPŁYWA NA ISTNIENIE ZJAWISKA HETEROSKEDASTYCZNOŚCI?
<i>healthexp</i>	0.9732	nie
<i>alcohol</i>	0.3982	nie
<i>GDP</i>	0.7786	nie
<i>infantmort</i>	0.001842	tak
<i>teacher</i>	0.08579	nie

Wartości krytyczne p-value (mniejsze od poziomu istotności $\alpha = 0,05$) sugerują, że problem heteroskedastyczności składnika losowego może występować przez zmienną *infantmort*.

Na podstawie testu Goldfelda-Quandt został zasygnalizowany problem heteroskedastyczności składnika losowego.

Wykresy reszt oraz test sugerowały, że ten problem zróżnicowanych wariancji jest prawdopodobnie spowodowany przez **zmienną *infantmort***. W przypadku zmiennej *infantmort*, kraje z wyższą oczekiwaną długością życia charakteryzują się raczej niskim odsetkiem śmiertelności noworodków. Jednak w przypadku krajów o mniejszej oczekiwanej długości życia, odsetek ten nie ma już tak zauważalnego trendu i wykazuje większe wariancje.

4.7. Odporne estymatory macierzy wariancji-kowariancji

W modelu została wykryta heteroskedastyczność składnika losowego. Aby więc zniwelować jej wpływ, stosuję odporne błędy standardowe typu HC3, które uwzględniają tę własność składnika losowego. Stosuję tu pakiet *sandwich* dostępny w programie R.

Wyniki przedstawiam w tabelce porównawczej.

Intercept	Estimate(MNK)	p-value(MNK)	Estimate(HAC)	p-value(HAC)
Intercept	81.582435	<2e-16	81.5824353	<2e-16
Healthexp	0.144647	0.00255	0.1446474	0.0005588
alcohol	-0.017213	0.07627	-0.0172127	0.0542429
GDP	0.077677	9.28e-08	0.0776766	2.636e-09
Infantmort	-0.266134	<2e-16	-0.2661341	<2.2e-16
log(teacher)	-1.804483	0.00628	-1.8044826	0.0092936

Po zastosowaniu odpornych błędów standardowych HC3 poprawiła się bardzo istotność statystyczna zmiennej alkohol. Można praktycznie powiedzieć, że przy tak dużej ilości obserwacji jaką mamy, zmienna ta jest istotna statystycznie. Współczynniki też uległy obniżeniu ale niezauważalnie i nieznacznie

4.8. Ważona metoda najmniejszych kwadratów

Kolejnym sposobem na poradzenie sobie z problemem heteroskedastyczności składnika losowego może być ważona metoda najmniejszych kwadratów. Tym razem korzystam z estymatora FGLS. Rozważam model pomocniczy, w którym zmienną objaśnianą jest wcześniej wyznaczony z modelu wyjściowego MNK, spierwiastkowany logarytm naturalny kwadratów reszt., a zmiennymi objaśniającymi zmienne z modelu wyjściowego. Wykorzystuję wartości teoretyczne jako wagi w estymatorze WLS i przeprowadzam transformacje, dzieląc przez wagi.

Poniżej na wydruku z programu R, widać porównanie modelu (1) oszacowanego metodą najmniejszych kwadratów, oraz modelu (2) oszacowanego ważoną metodą najmniejszych kwadratów. Do porównania użyłam funkcji *stargazer*, dostępnej w pakiecie *stargazer*. Ponieważ wycinek z R dokładnie oraz w estetyczny sposób przedstawia to porównanie, umieszczę go w tym miejscu analizy.

Dependent variable:		
	l1feexp	
	(1)	(2)
healthexp	0.145*** (0.047)	0.153*** (0.042)
alcohol	-0.017* (0.010)	-0.016* (0.009)
GDP	0.078*** (0.014)	0.078*** (0.014)
infantmort	-0.266*** (0.017)	-0.273*** (0.017)
log(teacher)	-1.804*** (0.652)	-1.547** (0.640)
Constant	81.582*** (1.894)	80.801*** (1.840)
Observations	172	172
R2	0.902	0.894
Adjusted R2	0.899	0.891
Residual Std. Error (df = 166)	2.417	2.147
F Statistic (df = 5; 166)	304.808***	281.468***
Note: *p<0.1; **p<0.05; ***p<0.01		

Zastosowanie ważonej MNK spowodowało zmniejszenie oszacowań parametrów zmiennej *infantmort* oraz zwiększenie przy zmiennych *healthexp*, *alkohol*, *log(teacher)*. Dla wszystkich zmiennych objaśniających wzrósł ich wpływ na zmienną objaśnianą. Zmniejszyła się wartość współczynnika determinacji R^2 .

4.9 Endogeniczność

Endogeniczność jest zjawiskiem występującym powszechnie w ekonometrii. Endogeniczność może być wynikiem błędu pomiarowego, autoregresji, pominiętych zmiennych lub błędów w wyborze próby statystycznej. Więc warto zastanowić się, czy mamy z nią do czynienia. Szczególnie, że podczas przygotowywania danych do mojego modelu, napotkałam między innymi problem braków danych, które musiałam usunąć w programie R komendą *na.omit*, by kontynuować dalszą analizę. W wyniku tego działania straciłam około 50 zmiennych, w których brakowało danych. Dodatkowo na badane przeze mnie zjawisko wpływa całe spektrum zmiennych. Łatwo więc jest pominąć jakąś ważną i kluczową dla istnienia modelu zmienną.

Co znaczy właściwie endogeniczność? O parametrze mówi się, że jest endogeniczny, jeśli jest on skorelowany ze składnikiem losowym. W moim modelu na składnik losowy może składać się wiele rzeczy na przykład liczba zachorowań na HIV w danym państwie. Zmienna z modelu, która może być silnie skorelowana ze składnikiem losowym to np. *healthexp*.

Aby wyeliminować prawdopodobną endogeniczność, korzystam z **metody zmiennych instrumentalnych**. Najpierw należy zdecydować się na instrument, który będzie spełniał odpowiednie kryteria. Dobry, mocny instrument powinien być silnie skorelowany ze zmienną endogeniczną (a więc np. z *healthexp*) i nieskorelowany ze składnikiem losowym, a więc nie może on wpływać bezpośrednio na *lifeexp*. Wybieram instrument **immun**, czyli procent dzieci w wieku 12-23 miesięcy, które zostały zaszczepione.¹⁵

Instrument **immun** może być właśnie takim dobrym kandydatem na **mocną zmienną instrumentalną**, gdyż:

1. Jest on **silnie skorelowany ze zmienną endogeniczną** – wydatkami państwa na służbę zdrowia (*healthexp*). Wiadomo, że im więcej pieniędzy państwo przeznacza na opiekę lekarską, tym większy budżet trafi też do zagospodarowania na szczepionki dla małych dzieci.

2. Równocześnie: Procent zaszczepionych dzieci nie **jest skorelowana bezpośrednio ze składnikiem losowym**, na przykład szczepienia w żaden sposób nie mogą wpłynąć np. na ilość osób zarażonych wirusem HIV lub np. na przynależność kraju do OECD.

Jest to więc idealny kandydat na instrument do metody zmiennych instrumentalnych.

Dla pewności zbadam, czy rzeczywiście mamy do czynienia z mocnym instrumentem. Wykonam **test Walda**.

Hipotezy testu Walda:

H0: Zmienna *immun* nie jest mocnym instrumentem

H1: Zmienna *immun* jest mocnym instrumentem

Statystyka F testu Walda wyniosła 22.555, co jest większe od 10. Wykorzystując rule of thumb, można stwierdzić, że zmienna *immun* jest mocnym instrumentem.

Przeprowadzamy więc analizę metodą zmiennych instrumentalnych bez przeszkód. Czy rzeczywiście dostaliśmy pozytywny rezultat tej analizy? Czy zastosowanie w tym przypadku metody zmiennych instrumentalnych jest zasadne? Możemy użyć testu Hausmana do zbadania, czy są rzeczywiście istotne różnice między oszacowaniami w modelu MNK I i IV.

Hipotezy testu Hausmana

H0: nie ma istotnych różnic pomiędzy oszacowaniami parametrów w modelach MNK I i IV

H1: są istotne różnice....

¹⁵ Immunization, DPT (% of children ages 12-23 months), <https://data.worldbank.org/indicator/SH.IMM.IDPT?view=chart>

P-value otrzymane w teście Hausmana wyniosło 0.7. Z tego wynika, że na przyjętym poziomie istotności 5% nie ma podstaw do odrzucenia hipotezy zerowej, mówiącej o zgodności obu estymatorów. Przyjmujemy H_0 więc najprawdopodobniej nie warto jest w tym wypadku zastosować metody zmiennych instrumentalnych.

Warto dodać też, że w modelu MZI uzyskano większe wariancje, większe błędy standardowe oraz mniejszy współczynnik determinacji R^2 , niż w modelu MNK. Można więc powiedzieć, że „skutkiem ubocznym” MZI było chociaż obniżenie nieco zbyt wysokiego współczynnika determinacji R^2 .

	Dependent variable:	
	OLS (1)	instrumental variable (2)
healthexp	0.145*** (0.047)	0.901 (3.114)
alcohol	-0.017* (0.010)	-0.059 (0.173)
GDP	0.078*** (0.014)	0.051 (0.114)
infantmort	-0.266*** (0.017)	-0.171 (0.391)
log(teacher)	-1.804*** (0.652)	-2.809 (4.263)
Constant	81.582*** (1.894)	76.841*** (19.749)
observations	172	172
R2	0.902	0.750
Adjusted R2	0.899	0.742
Residual Std. Error (df = 166)	2.417	3.856
F Statistic	304.808*** (df = 5; 166)	

5. Podsumowanie

W przedstawionym raporcie przygotowałam oraz poddałam analizie model, w którym zbadałam, od czego tak naprawdę zależy średnia oczekiwana długość życia człowieka. Pod uwagę wzięłam reprezentację czterech czynników: ekonomicznego, zdrowotnego, demograficznego, edukacyjnego i społecznego.

Model prezentuje zależności w skali świata, dla 218 regionów-krajów świata. Po analizie zbioru danych widoczny był powszechnie znany podział na *bogatą północ* i *biedne południe*. W modelu zostały przedstawione grupy czynników, wpływające na zmienną objaśnianą. Aby analiza była pełniejsza, mogłabym dodać jeszcze zmienne dotyczące na przykład ekologii (zanieczyszczenie dwutlenkiem węgla) oraz poziomu higieny, przestępczości lub na przykład przynależności kraju do wspólnoty OECD. Wszystkie te zmienne rozważałam i odrzuciłam na przykład z powodu dużych braków danych dla ostatnich lat (poziom higieny i dostępności wody – gdzie dane były dostępne tylko dla 90 krajów, co zdecydowanie obniżyłoby jakość mojego modelu). Dodatkowo chciałam uporządkować moją analizę i jak najbardziej ją uprościć. Przyjęłam taki tok myślenia i założyłam, że na przykład zarówno poziom bezrobocia, jak i PKB per capita powoduje tożsame skutki ekonomiczne, wpływające na oczekiwaną długość życia.

Po stworzeniu modelu cztery z pięciu zmiennych okazały się być istotne statystycznie na poziomie istotności $\alpha = 0,05$ oraz jedna jest istotna statystycznie na poziomie istotności $\alpha = 0,1$.

Założenia co do kierunku wpływu danych zmiennych, które przyjąłam na samym początku analizy potwierdziły się. Zarówno poprawienie się jakości opieki medycznej, edukacji jak i wzrost PKB per capita w danym kraju, statystycznie wydłużają nasze życie o daną jednostkę czasu. By zobrazować ten wpływ, użyłam jednostki dni. W wyniku analizy i przeliczenia wyników naszego modelu, dowiedzieliśmy się, że nasze życie może wydłużyć lub skrócić głównie pod wpływem poziomu opieki medycznej w danym kraju. Potwierdził to fakt, że nasze życie może wydłużyć się nawet o około 50 dni, kiedy zwiększymy wydatki rządu na służbę zdrowia na jednego mieszkańca o 1 USD, natomiast może się też skrócić statystycznie o około 100 dni, kiedy nastąpi wzrost śmiertelności niemowląt średnio o jedno niemowlę na 1000 żywych urodzeń.

Model przeszedł przez liczne testy oraz weryfikacje. Na początku, analizując współczynnik determinacji R^2 oraz wykres dopasowania wartości teoretycznych do empirycznych, doszłam do wniosku, że przy danym zestawie obserwowanych danych, model tłumaczy około 90% zmienności zmiennej objaśnianej *lifeexp*, zmiennymi objaśniającymi, czyli jest to jak najbardziej zadowalający wynik. Oczywiście wzięłam pod uwagę ryzyko „nadmiernego dopasowania do danych” i wykluczyłam je, dokonując odpowiedniej analizy myślowej. W kolejnych krokach znalazłam rozwiązanie na obniżenie R^2 za pomocą zastosowania metody zmiennych instrumentalnych (był to tzw. „efekt uboczny” MZI ale warto podkreślić, że w zadowalający sposób doprowadziło to do obniżenia się R^2).

Następnie rozkład reszt przeszedł test na normalność. Model nie wykazał również problemu współliniowości, opierając się na wartościach czynnika inflacji wariancji. Największym problemem dla tego modelu jest występowanie problemu heteroskedastyczności składnika losowego, jednak wynik nie był jednoznaczny. Test Breusch-Pagana wykazały brak heteroskedastyczności, jednak w teście Goldfelda-Quandt oraz podczas analizy wykresów reszt zostało zasygnalizowane

zjawisko heteroskedastyczności dla jednej zmiennej. Najsilniej to zjawisko było spowodowane prawdopodobnie przez zmienną opisującą odsetek śmiertelność noworodków. Z tego też powodu została zastosowana ważona metoda najmniejszych kwadratów oraz odporne błędy standardowe.

Można również byłoby rozważyć oszacowanie parametrów modelu w inny sposób, np. dla konkretnego kontynentu lub jednostek administracyjnych kraju. Równocześnie, wymagałoby to doboru innych zmiennych objaśniających. Przy dokładniejszym badaniu różnic między regionami albo wewnątrz danego regionu, przydatna mogłaby być więc większa liczba zmiennych nieekonomicznych. Prawdopodobnie więc, gdyby model był opierał się na danych dokładniej zagregowanych, wnioski z niego płynące byłyby inne.

Podsumowanie zakończę cytatem. Będą to słowa Jeanne Calment, a więc osoby, której tajemnica długowieczności była najbliższa ze wszystkich ludzi, żyjących dotychczas na ziemi. Przynajmniej biorąc pod uwagę tylko fakt, że stała się najdłużej żyjącą osobą, której wiek został potwierdzony.

„Jeśli nie możesz sobie z czymś poradzić, nie przejmuj się”. Moim zdaniem to właśnie to, co niemierzalne, a więc nasze podejście do życia i ogólne szczęście (na które składa się oczywiście sytuacja ekonomiczna, zdrowotna oraz społeczna), wpływa najbardziej na nasze długie życie. Każdy student, jako początkujący ekonometryk powinien zgłębiać tajniki ekonomii, uczyć się coraz to nowych programów, ale tak jak każdy człowiek na ziemi, przede wszystkim znaleźć w życiu swoje własne szczęście.



Jeanne Calment, <https://www.dna.fr/actualite/2018/12/31/jeanne-calment-son-record-de-longevite-mis-en-cause>

6. Literatura

1. M. Gruszczyński, T. Kuszewski, M. Podgórska, *Ekonometria i badania operacyjne*, Warszawa 2009, s.53, 85, 88
2. Dr Jakub Mućk, prezentacje z wykładów oraz skrypty z ćwiczeń w programie R, z przedmiotu Metody Ekonometryczne
3. Jeanne Calment, https://pl.wikipedia.org/wiki/Jeanne_Calment
4. Najstarsi ludzie, https://pl.wikipedia.org/wiki/Najstarsi_ludzie
5. Strona Internetowa Muzeum Archeologicznego w Biskupinie, "Rezerwat Archeologiczny w Biskupinie", <http://www.biskupin.pl/rezerwat-archeologiczny/>
6. Life expectancy at birth, total (years), <https://ourworldindata.org/life-expectancy>
7. Sidney Katz, M.D., Laurence G. Branch, Ph.D., Michael H. Branson, ,The New England Journal of Medicine, Active Life Expectancy, England, 17th November 1983, <https://www.nejm.org/doi/pdf/10.1056/NEJM198311173092005>
8. Kto dużo pije, ten krócej żyje, 15 kwiecień 2018, <https://www.forbes.pl/life/alkohol-skraca-zycie-naukowcy-sprawdzili-jak-bardzo/534vg78>,
9. James P. Smith, Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status, *Journal Of Economic PERSPECTIVES*, VOL. 13, NO. 2, SPRING 1999, (pp. 145-166)
10. Dane z Data World Bank:
 - Life expectancy at birth, total (years), <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>
 - GDP per capita (current US\$), <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
 - Current health expenditure per capita (current US\$), <https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD>
 - Mortality rate, infant (per 1,000 live births), <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>
 - Pupil-teacher ratio, primary, <https://data.worldbank.org/indicator/SE.PRM.ENRL.TC.ZS>
 - Immunization, DPT (% of children ages 12-23 months), <https://data.worldbank.org/indicator/SH.IMM.IDPT?view=chart>
11. Dane z World Health Organisation:
 - Alcohol Consumption 2016, <https://www.who.int/news-room/fact-sheets/detail/alcohol>
12. Oczekiwana dalsza długość trwania życia, https://pl.wikipedia.org/wiki/Oczekiwana_dalsza_d%C5%82ugo%C5%9B%C4%87_trwania_%C5%BCycia