**ABSTRAKT**

**SŁOWA KLUCZOWE**

**WSTĘP** [cel pracy, motywacja, krótko o danych, struktura pracy]

Jednym z problemów, które dotykają państwa na świecie, w tym Polskę, jest proces intensywnego starzenia się społeczeństwa przy jednoczesnym zmniejszaniu się odsetka ludności młodszej. Zjawisko to w szczególności dotyka kraje rozwinięte i jest następstwem wystąpienia wielu czynników ekonomicznych, społecznych, cywilizacyjnych i kulturowych. Duży wpływ ma między innymi spadek tempa przyrostu naturalnego, na który wpływ ma liczba urodzeń, którą można mierzyć na przykład za pomocą współczynnika urodzeń (stosunku liczby żywych urodzeń na 1000 mieszkańców) lub współczynnika dzietności (liczba urodzeń dzieci przypadająca na jedną kobietę).

Pandemia koronawirusa COVID-19, która umownie rozpoczęła się w Polsce 20.03.2020 roku (data publikacji Rozporządzenia Ministra Zdrowia w sprawie ogłoszenia na obszarze Rzeczypospolitej Polskiej stanu epidemii) bezpośrednio wpływa na struktury demograficzne kraju. Związane jest to bezpośrednio z wysoką śmiertelnością i wzrostem liczby zgonów (Tabela 1.), a także pośrednio ze zmniejszeniem się mobilności na skutek wprowadzenia obostrzeń, których celem było ograniczenie rozprzestrzeniania się wirusa.

Tabela 1. Zestawienie liczby zgonów w Polsce w latach 2017 – 2020

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lata** | **2017** | **2018** | **2019** | **2020** |
| **Liczba zgonów** | 402 852 | 414 200 | 409 709 | 477 355 |
| **Indeks łańcuchowy (rok bazowy = rok poprzedni)** | - | 103,2% | 98,9% | 116,5% |

*Źródło:* Opracowanie własne na podstawie danych *Banku Danych Lokalnych*.

**[dlaczego istotne – konsekwencje braku dzieciorow] [jakiś lacznik z koroną, bo ten koronawius z dupska, może cos w stylu ze koronawirus jest na tyle nieprzewidywalny ze nieprzeidywalne skutki demograficzne, nie można progozowac, zaburzone prognozy sprzed lat]**

Celem pracy jest określenie istotności wybranych determinantów współczynnika urodzeń w Polsce. Analiza została stworzona w oparciu o dane dotyczące poszczególnych podregionów Polski – to jest powiatów i miast na prawach powiatu (łącznie 380 obszarów). W modelu zostały wykorzystane zmienne odnoszące się do kategorii ekonomicznych, jak i społecznych czy demograficznych.

Do stworzenia modelu ekonometrycznego wykorzystano dane z 2020 roku pochodzące z Banku Danych Lokalnych (zarządzanego przez Główny Urząd Statystyczny) oraz baz danych gromadzących statystyki dotyczące pandemii COVID-19. W procesie analizy i weryfikowania hipotez badawczych użyto Metody Najmniejszych Kwadratów (MNK).

Praca została [struktura pracy]

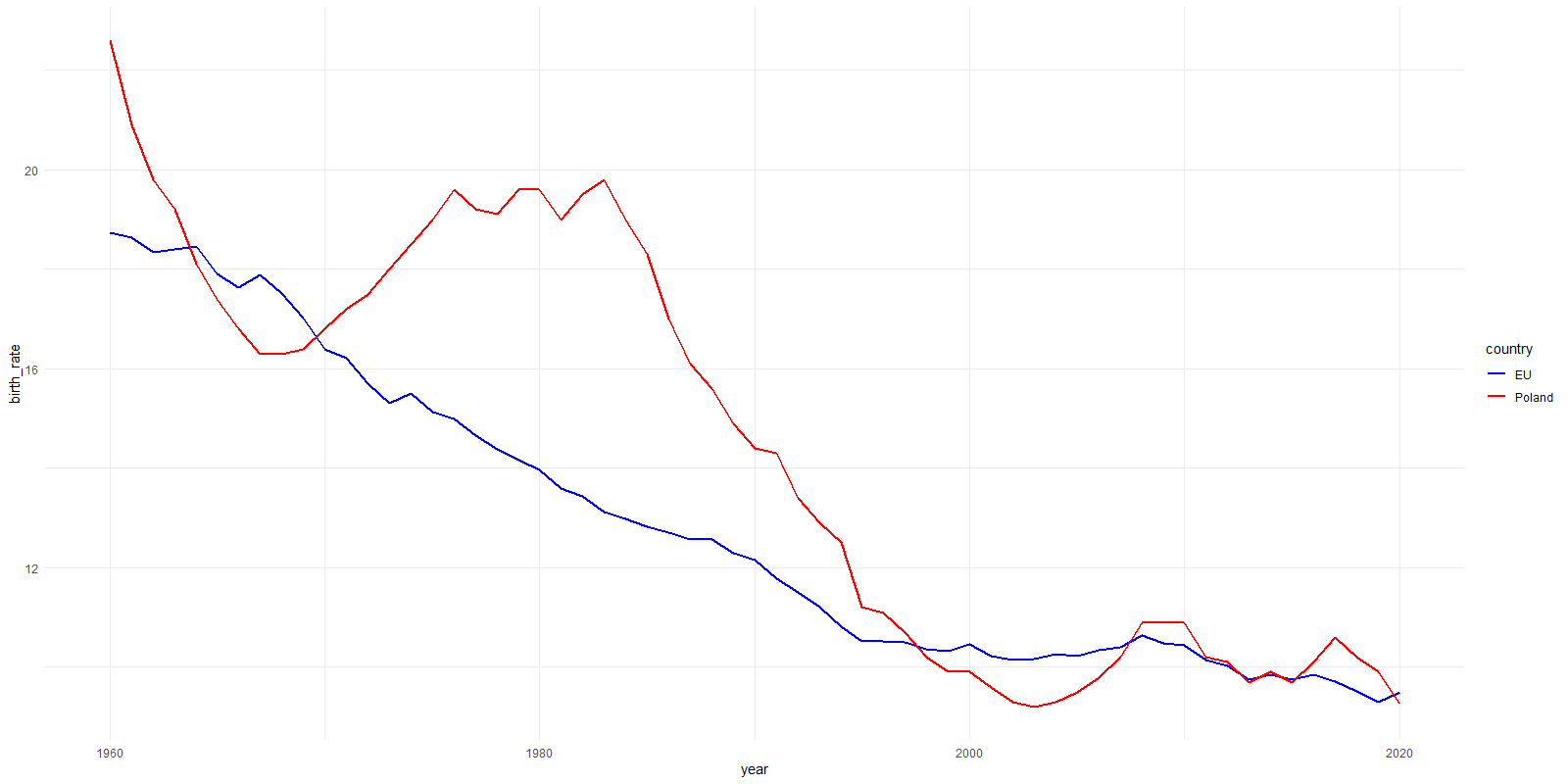
**ROZDZIAŁ I Literatura** [temat, dane, model, wnioski]

Na przełomie XIX i XX wieku zaczęły powstawać koncepcje dotyczące czynnika ekonomicznego jako stymulatora rozwoju demograficznego. Rozwijały się na płaszczyźnie mikro- i makrospołecznej (pierwsza odnosi się do jednostki, natomiast druga do gospodarstw domowych). w ujęciu makro czynnik ekonomiczny rozpatrywany jest jako teoria dobrobytu, która obejmuje zjawiska ekonomiczne (ilość i jakość dóbr i usług rozporządzana przez społeczeństwo w ramach zaspokajania swoich potrzeb) oraz pozaekonomiczne (związane z fizjologią, psychologią czy etyką). Ówcześni badacze starali się tłumaczyć jak dobrobyt wpływa na liczbę ludności. Perspektywa mikrospołeczna zainteresowała badaczy dopiero w 2. połowie XX wieku w wyniku rozwoju badań nad rodziną i gospodarstwem domowym. Celem badań było zgłębienie wiedzy na temat mechanizmu podejmowania decyzji w sprawie narodzin potomków przez małżeństwa. Na przestrzeni lat powstawały nowe teorie tłumaczące proces, które wiązały się między innymi z kalkulacją użyteczności pochodzącej z dodatkowego dziecka, poszukiwaniem czynników wpływających na częstotliwość współżycia małżonków czy stosowanie metod antykoncepcji. Wynikiem podjęcia próby analizy płodności przez ekonomistów było dostrzeżenie czynników pozabiologicznych, które wpływają na prokreację (Chromińska, 1998).

Problemem w estymowaniu zmiennych dotyczących płodności, w tym współczynnika urodzeń jest to, że w większości koncepcje mają charakter hipotez i trudno zweryfikować je empirycznie. Barierą może być rzetelność odpowiedzi jednostek w badaniach dotyczących tak intymnej sfery jak życie rodzinne. We współczesnych teoriach i analizach Chromińska (1998) dostrzega wadę w postaci pomijania wartości związanych z etyką i moralnością, co wiąże się z niepełnym obrazem oszacowań zmiennych dotyczących wydawania potomstwa na świat. W niniejszej pracy zbadano wpływ warunków ekonomicznych odgrywających dużą rolę w decyzjach dotyczących posiadania dzieci. Pominięto szereg zmiennych, których aspektami zajmuje się medycyna, takimi jak mechanizmy kontroli urodzeń czy terminacje ciąży.

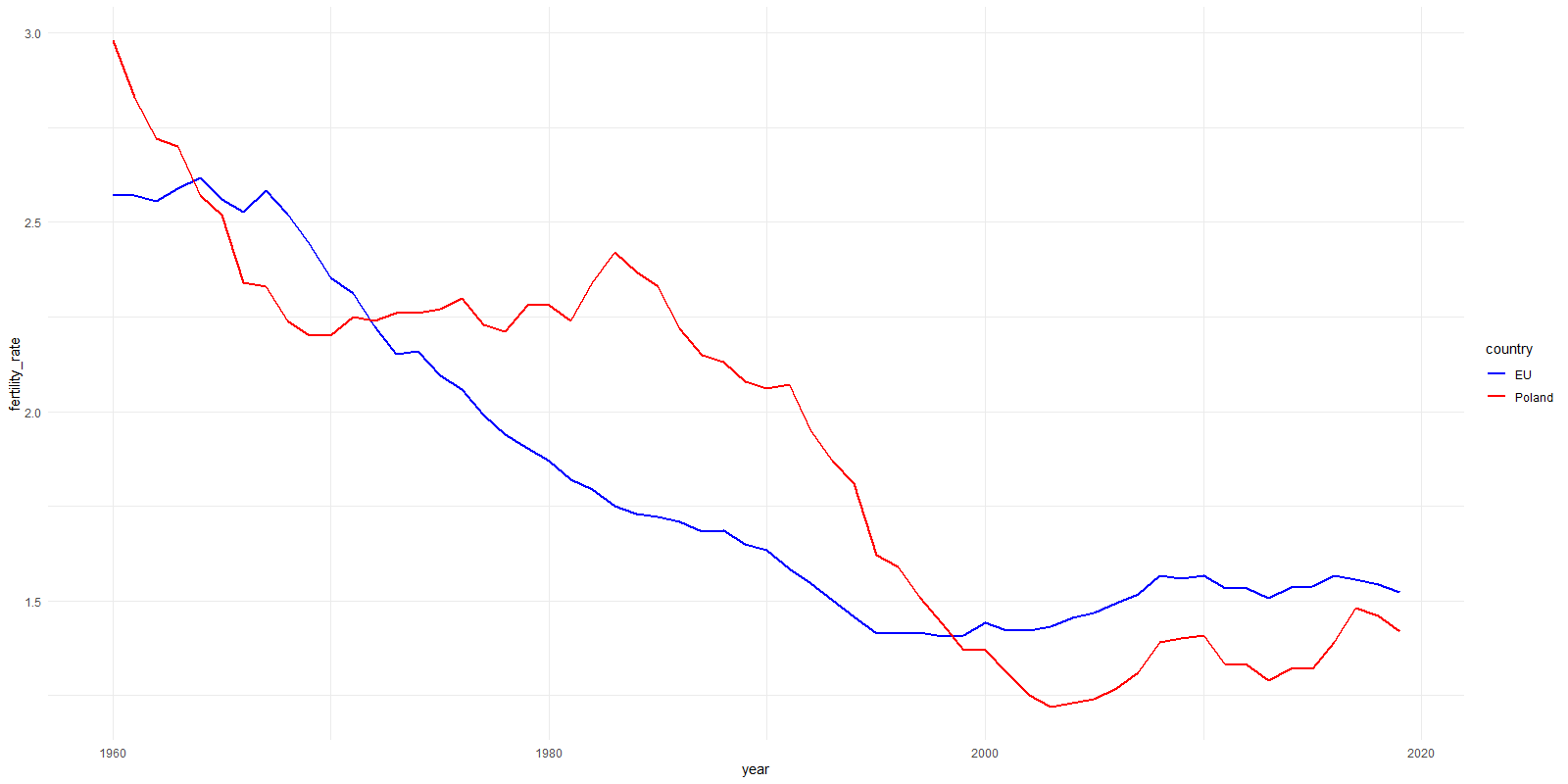
Od początku lat 90. XX wieku w Polsce możemy zaobserwować zmiany w aspekcie urodzeń. Na wykresie zmian współczynnika urodzeń (Rys. 1.) oraz współczynnika dzietności (Rys. 2.) w czasie widać wyraźny spadek, co może być skutkiem zmniejszenia się liczby nowonarodzonych dzieci. Według demografów wynika to przede wszystkim ze zmiany poglądów na temat posiadania dziecka przez młodych ludzi – decyzję o potomstwie odkładają na później (między innymi z powodu chęci osiągnięcia określonego poziomu wykształcenia i stabilizacji ekonomicznej), decydują się na mniejszą liczbę dzieci lub wolą żyć samotnie. Rezultatem tego jest zwiększenie się średniej i mediany wieku kobiet rodzących, a także zmiana struktury wykształcenia matek. Badacze wskazują, że sytuacja ludnościowa jest trudna, ponieważ niski poziom dzietności negatywnie wpływa na brak zastępowalności pokoleń (konsekwencja zmniejszenia się liczby kobiet w wieku rozrodczym) (Hrynkiewicz, Witkowski, Potrykowska, 2018).

Rys. 1. Wykres zmian współczynnika urodzeń dla Polski i państw Unii Europejskiej w latach 1960-2020.



*Źródło:* Opracowanie własne na podstawie danych *World Bank* z wykorzystaniem biblioteki *ggplot2*.

Rys. 2. Wykres zmian współczynnika dzietności dla Polski i państw Unii Europejskiej w latach 1960-2019.



*Źródło:* Opracowanie własne na podstawie danych *World Bank* z wykorzystaniem biblioteki *ggplot2*.

Prognozy przedpandemiczne wskazywały, że połączenie niskiego współczynnika urodzeń i wydłużającego się czasu trwania życia może spowodować zmniejszenie podaży pracy a także postępujące starzenie się społeczeństwa. Proces starzenia się ludności jest wyzwaniem, ponieważ wpływa na wiele sfer życia (ekonomiczną, psychologiczną, medyczną czy socjalną). Dodatkowo starość demograficzna oznacza też silną feminizację w grupach starszych wiekowo (ze względu na średnio dłuższe trwanie życia kobiet w stosunku do mężczyzn – spowodowane nawykami żywieniowymi, higienicznymi, warunkami pracy) (Hrynkiewicz, Witkowski, Potrykowska, 2018).

Z perspektywy ekonomicznej starzejące się społeczeństwo może powodować postępujące zmniejszanie się wskaźników wzrostu gospodarczego w regionach. Początkowo, wraz z malejącą liczbą urodzeń gospodarka rośnie, jednak gdy zostanie przekroczony punkt krytyczny a spadek dzietności niezahamowany to równowaga między populacją osób starszych a grupą osób w wieku produkcyjnym nie będzie mogła być utrzymana (Arcuri, 2019). Konsekwencją procesu będzie potrzeba reorganizacji sposoby finansowania wydatków publicznych, ponieważ coraz więcej osób będzie wymagało zabezpieczenia emerytalno-rentowego (Hrynkiewicz, Witkowski, Potrykowska, 2018).

Pandemia COVID-19 wywarła ogromny wpływ na struktury demograficzne państw na świecie, w tym Polski. Nieprzewidziany w dotychczasowych prognozach wzrost zgonów i recesja gospodarcza będzie odnotowywany we wskaźnikach ekonomicznych, społecznych czy demograficznych. Jeżeli chodzi o wpływ pandemii na liczbę urodzeń, to trudno wnioskować czy będzie on pozytywny czy negatywny. Porównując zmniejszenie mobilności obywateli wywołane koronawirusem do sytuacji wprowadzenia stanu wojennego w 1981 roku, to należałoby przewidywać wzrost liczby urodzeń spowodowany pozostaniem w domach razem z partnerami. Podobne wnioski zostały wyciągane na podstawie powielanych historii związanych ze skokiem urodzeń dziewięć miesięcy po zamieciach czy poważnych przerwach w dostawie prądu. Mity te zostały statystycznie obalone między innymi przez Richarda Udry’ego w 1970 roku, gdzie w artykule badał wpływ awarii zasilania w Nowym Jorku w 1965 roku na urodzenia dzieci. Wzrost współczynnika urodzeń nie został także odnotowany po historycznej zamieci śnieżnej w 1967 roku w Chicago, dlatego można odrzucić hipotezę o zwiększeniu współczynniku współżycia seksualnego prowadzącego do wzrostu liczby urodzeń w przypadku nakazów pozostania w domu (Udry, 1970).

Kryzys zdrowia publicznego wywołał także recesję gospodarczą prowadzącą do strat ekonomicznych, niepewności, braku możliwości przewidywania przyszłości i zmniejszeniu poczucia bezpieczeństwa wobec śmiertelnego wirusa.

dlaczego wspolczynnik urodzen i dzietności maleje, dlaczego to jest istotne, dlaczego jest istotne żeby to badac

pandemia a wspolczynnik urodzen

jakie modele i dane przedstawiają inni

**ROZDZIAŁ II HIPOTEZY [hipoteza główna, hipotezy pomocnicze, odwołania do lietarury kto zrobil podobnie)**

**2.1. Hipoteza główna**

H0: Pandemia COVID-19 negatywnie wpływa na liczbę urodzonych dzieci.

rozwinąć fest

**2.2 Hipotezy pomocnicze**

H1: Wzrost przeciętnego wynagrodzenia wpływa negatywnie na współczynnik urodzeń

Becker w artykule „An Economic Analysis of Fertility” stwierdza, że wzrost dochodu powinien spowodować zwiększenie ilości dzieci jak i ich jakości. Jakość dziecka definiuje jako wysokość wydatków przeznaczane na nie, to jest zapewnienie mu na przykład dostępu do prywatnej szkoły, zapisywanie na dodatkowe zajęcia z tańca czy śpiewu. Wyższa jakość oznacza wyższą użyteczność dziecka powstałą z poniesienia dodatkowych kosztów. Dodatnia korelacja między dochodem a ilością urodzeń może być spowodowana przez spadek śmiertelności wśród noworodków co bezpośrednio wpływa na zwiększenie liczby urodzeń żywych (Becker, 1960).

Z drugiej strony Becker zauważa, że postęp w dziedzinie planowania rodziny, a dokładniej w rozwoju metod antykoncepcji (kontrola narodzin nie musi sprowadzać się do braku współżycia) może negatywnie wpływać na współczynnik urodzeń. Wzrost dochodu nie gwarantuje już wzrostu liczby potomstwa, bo wraz ze zwiększeniem się dochodu rośnie wiedza na temat technik zabezpieczania się przed niechcianą ciążą oraz ich stosowalność (Becker, 1960).

Vandenbroucke (2016) zauważa negatywną tendencję jeśli chodzi o PKB per capita państw a liczbą urodzeń. Na tej podstawie można sformułować hipotezę o ujemnym wpływie wynagrodzenia na współczynnik urodzeń. Tak jak ekonomista, można to tłumaczyć tym, że mniejsze wynagrodzenie świadczy o relatywnie tańszym czasie, więc jego poświęcanie na wychowanie dziecka nie jest tak kosztowne jak w przypadku regionów bogatszych (notującym wyższe średnie wynagrodzenie).

H2: Wzrost wydatków na rodziny spowoduje wzrost współczynnika urodzeń.

John Ermisch w artykule „Economic influences on birth rate” sugeruje, że zachętą do rodzenia dzieci są wprowadzane przez państwo narzędzia polityki fiskalnej takie jak ulgi podatkowe na dzieci lub wypłaty zasiłków. Jednym z takich narzędzi są zasiłki rodzinne, które możemy interpretować jako wydatki budżetów powiatów i miast na prawach powiatu w dziale Rodzina. Wyższe wartości zasiłków (to jest wydatków) wpływają pozytywnie na liczbę rodzonych dzieci (które wpływają bezpośrednio na współczynnik urodzeń) (ermisch1988).

H3: Wyższa stopa bezrobocia ma pozytywny wpływ na współczynnik urodzeń.

Według Johna Ermischa (1988) zmiany prawdopodobieństwa znalezienia pracy mierzone za pomocą stopy bezrobocia mogą mieć wpływ na to, że posiadanie dziecka będzie mniej kosztowne.

rozwinąć fest

H4: Wzrost cen mieszkań i jednoczesny spadek ilości mieszkań na rynku ujemnie koreluje ze współczynnikiem urodzeń.

Ermisch (1988) uważa, że konkurencja na rynku mieszkaniowym zniechęca pary do posiadania dzieci. Małżeństwa, które chcą kupić mieszkanie podczas gdy popyt na nie przewyższa podaż będą negatywnie patrzeć na wizję powiększania rodziny w dobie rosnących cen zakupu domu oraz wydłużających się kolejek po mieszkania socjalne. Ekonomista jako wskaźnika presji na rynku mieszkaniowym używa średniej ceny nowego domu, w niniejszej pracy zaproponowano mierzyć ten efekt za pomocą mediany ceny za 1 m2 lokalu mieszkaniowego oraz stosunku mieszkań oddanych do użytku na 1000 ludności.

Z drugiej strony, osoby, które posiadają już własne domy w wyniku rosnących cen mieszkań stają się bardziej zamożni, co może być bodźcem do powiększania rodziny. Do innych wniosków dochodzą Black, Kolesnikova, Sanders i Taylor (2013), którzy stwierdzają, że małżeństwa żyjące w bogatych lokalizacjach (czyli w przypadku tej pracy o wyższej medianie cen za m2 lokalu mieszkaniowego) statystycznie mają mniejsze rodziny (współczynnik urodzeń jest mniejszy).

H5: Im wyższy poziom urbanizacji, tym mniejszy współczynnik urodzeń.

Demografowie odnotowują ujemną korelację między wskaźnikiem zurbanizowania danego obszaru a liczbą dzieci. Wynika to z różnic w zachowaniach demograficznych (w tym związanych z prokreacją) między wsią a miastem. W niniejszej pracy można wykorzystać współczynnik urbanizacji powiatów (i miast na prawach powiatu), aby przekonać się czy regiony o wyższym odsetku ludności zamieszkującej miasta charakteryzują się niższą skłonnością do posiadania potomstwa (badaną jako współczynnik urodzeń) (Hrynkiewicz, Witkowski, Potrykowska, 2018).

H6: Współczynnik małżeństw wpływa dodatnio na współczynnik urodzeń.

W przyczynach spadku liczby urodzeń obserwowanego w Polsce, jak i w krajach europejskich (Rys. 1.) doszukuje się związku z liczbą zawieranych małżeństw (Hrynkiewicz, Witkowski, Potrykowska, 2018).

**rozwinąć fest**

H7: Poprawa dostępu do usług prowadzonych przez żłobki pozytywnie wpływa na współczynnik urodzeń

**rozwinąć fest**

**hipotezy pomocnicze:**

- mieszkania

- ochrona zdrowia (może tez kontekst zakazu ABORCJI)

- covid Jako HIPOTEZA GŁÓWNA?

**Rozdział III Analiza danych**

**3.1. Źródło danych**

Dane przekrojowe wykorzystane w modelu ekonometrycznym dotyczą 380 powiatów i miast na prawach powiatu w Polsce i wyrażają stan na rok 2020. Do przeprowadzenia estymacji regresji liniowej za pomocą Metody Najmniejszych Kwadratów (MNK) i stworzenia wykresów i tabel znajdujących się w pracy wykorzystano język programowania R. Praca opiera się na danych pochodzących z trzech źródeł: Banku Danych Lokalnych, projektu Michała Rogalskiego – *COVID-19 w Polsce* oraz archiwalnych danych dla powiatów związanych z koronawirusem COVID-19 udostępnionych przez Ministerstwo Zdrowia.

Wskaźniki i dane dotyczące sytuacji gospodarczej, demograficznej, społecznej powiatów zostały pobrane z Banku Danych Lokalnych, za którego prowadzenie i rozwój odpowiedzialny jest Główny Urząd Statystyczny w Polsce.

Baza danych COVID-19 w Polsce stworzona przez Michała Rogalskiego na podstawie raportów Ministerstwa Zdrowia, Wojewódzkich Stacji Sanitarno-Epidemiologicznych (WSSE), Powiatowych Stacji Sanitarno-Epidemiologicznych (PSSE) i Urzędów Wojewódzki została wykorzystana do pobrania danych dotyczących sumy przypadków zakażeń koronawirusem i zgonów nim spowodowanych w powiatach. Okres sumowania zgonów jest różny dla różnych województw i zależy to od daty zaprzestania publikowania danych dotyczących pandemii przez PSSE i WSSE.

Dane archiwalne dotyczące zarażeń koronawirusem dla powiatów opublikowane przez Ministerstwo Zdrowia (MZ) posłużyły do oszacowania łącznej liczby osób przebywających na kwarantannie dla poszczególnych powiatów. W archiwum znajdują się statystyki dotyczące COVID-19 w przedziale od 24.11.2020 do 31.12.2020 roku, związane jest to ze zmianą metodologii liczenia przypadków zachorowań i zgonów spowodowanych koronawirusem przez MZ. Zostały one wykorzystane do oszacowania łącznej liczby osób przebywających na kwarantannie w 2020 roku. Sposób szacowania:

1. Na podstawie danych z okresu 23.11.2020 – 31.12.2020 wyliczono średnią dzienną liczbę osób objętych kwarantanną.

2. Zgodnie z rozporządzeniem Ministra Zdrowia, stan epidemii koronawirusa w Polsce obowiązuje od 20.03.2020 roku. Średnią dzienną liczbę osób objętych kwarantanną wyliczoną w punkcie 1. pomnożono przez liczbę dni objętych stanem epidemii (286 dni) i zaokrąglono do liczb całkowitych.

3. Jeżeli dany powiat nie został uwzględniony w zestawieniu Ministerstwa Zdrowia, to dane dotyczące kwarantann na jego obszarze zostały wyliczone na podstawie średniej dziennej liczby osób objętych kwarantanną dla całego kraju.

**3.2. Opis zmiennych**

**3.2.1. Zmienna objaśniana**

Zmienną zależną w szacowanym modelu jest współczynnik urodzeń (*birth.rate*), który określa liczbę żywych urodzeń na 1000 ludności.

**3.2.2. Zmienne objaśniające**

Wszystkie zmienne niezależne użyte w modelu są zmiennymi ciągłymi.

* *covid19\_cases* - suma przypadków zachorowań na chorobę COVID-19.
* *covid19\_deaths* – suma zgonów wywołanych chorobą COVID-19.
* *covid19\_quarantine* – suma osób przebywających na kwarantannie z powodu kontaktu z osobą chorą na COVID-19.
* *marriage\_rate* – współczynnik małżeństw określający liczbę zawartych małżeństw na 1000 ludności.
* *divorce\_rate* – współczynnik rozwodów, który jest stosunkiem liczby orzeczonych rozwodów na 1000 ludności.
* *budget\_reve\_pc* – dochody budżetów powiatów (lub miast na prawach powiatu) w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.
* *unemployment\_rate* – stopa bezrobocia rejestrowanego wyrażona w procentach.
* *women\_reproductive* – liczba kobiet w wieku reprodukcyjnym, czyli między 15. a 49. rokiem życia.
* *femininity\_ratio* – współczynnik feminizacji określający liczbę kobiet przypadającą na 100 mężczyzn.
* *avg\_salary* – średnie wynagrodzenie brutto. Zmienna wyrażona w złotych polskich.
* *women\_working* – odsetek kobiet pracujących jako stosunek liczby pracujących kobiet do liczby kobiet w danym powiecie.
* *men\_working* – odsetek mężczyzn pracujących jako stosunek liczby pracujących mężczyzn do liczby mężczyzn w powiecie.
* *median\_house\_price* – mediana cen za 1 m2 lokali mieszkaniowych sprzedanych w ramach transakcji rynkowych. Zmienna wyrażona w złotych polskich.
* *house\_ratio* – mieszkania oddane do użytkowania na 1000 ludności.
* *houses\_area\_pc* – stosunek powierzchni użytkowej mieszkań w danym powiecie do jego populacji. Zmienna wyrażona w metrach kwadratowych.
* *avg\_people\_per\_house* – średnia liczba osób przypadająca na jedno mieszkanie.
* *education\_expenditure* - wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 801 – Oświata i wychowanie w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.
* *health\_expenditure* - wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 851 – Ochrona zdrowia w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.
* *social\_expenditure* – wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 852 – Pomoc społeczna w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.
* *family\_expenditure* – wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 855 – Rodzina w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.
* *children* – liczba osób w przedziale wiekowym 0 – 2 w danym powiecie.
* *nursery\_places* – miejsca w żłobkach, oddziałach żłobków i klubach dziecięcych.
* *doctors* – personel pracujący w sektorze ochrony zdrowia na 10 tysięcy ludności.
* *urbanisation\_rate* – współczynnik urbanizacji określający udział mieszkańców miast w liczbie ludności powiatu (lub miasta na prawach powiatu).
* *bus\_stops* – liczba przystanków autobusowych (z trolejbusowymi) i tramwajowych.

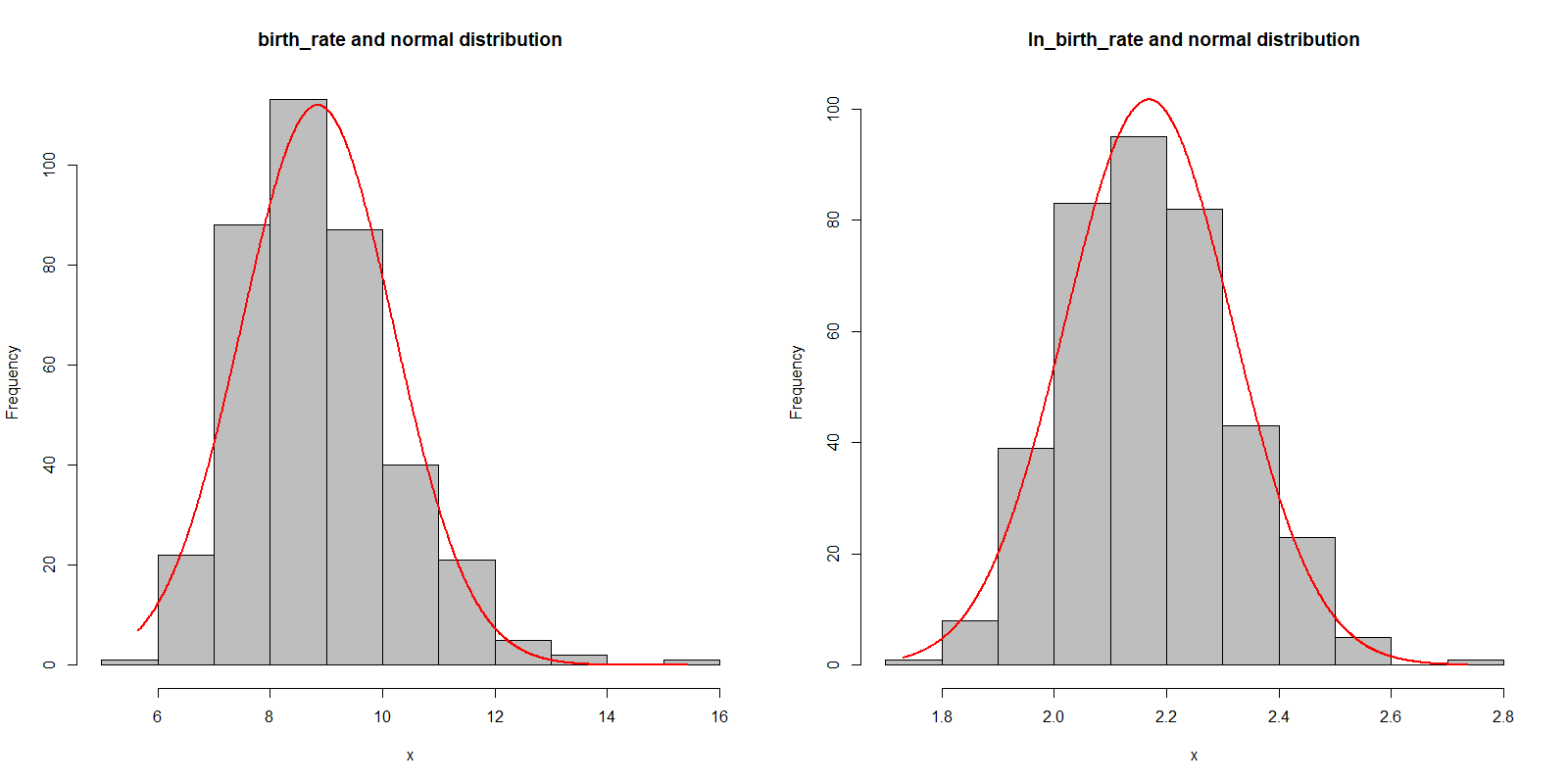
**3.3. Statystyczna analiza zmiennych**

[przyjrzyjmy się zmiennym, które zostały wykorzystane w końcowym modelu]

**3.3.1. Zmienna objaśniana – współczynnik urodzeń (birth\_rate)**

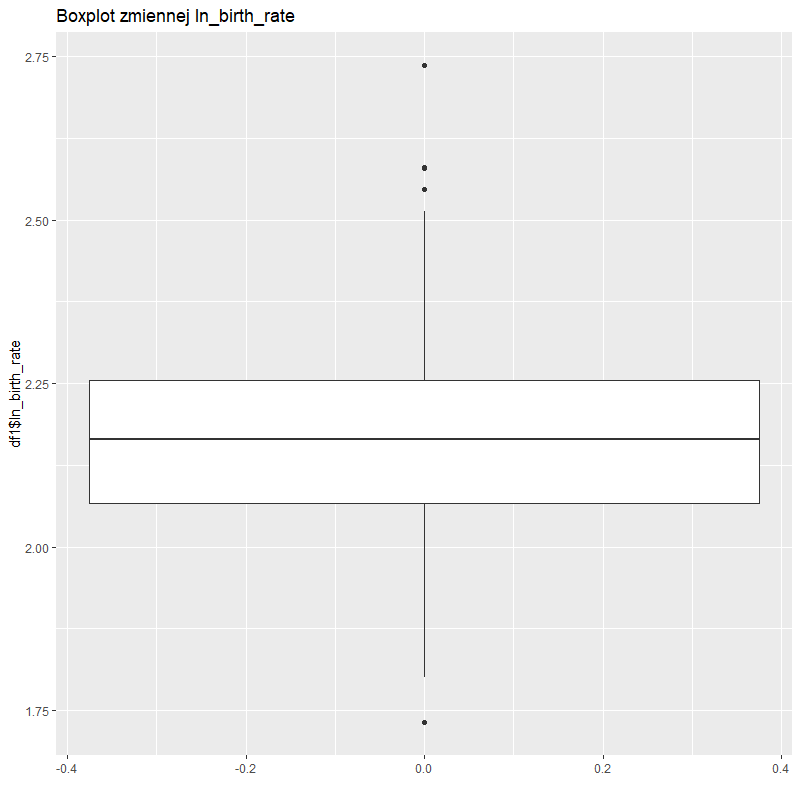
Rysunek 1. przedstawia histogramy zmiennej objaśnianej modelu ekonometrycznego w wersji podstawowej birth\_rate oraz zlogarytmowanej ln\_birth\_rate. Zmienna przyjmuje wartości od 5,45 do 15,43 (co możemy odczytać z tabeli z podstawowymi statystkami zmiennych umieszczonej w Załączniku 1.), natomiast najczęściej odnotowywane wartości znajdują się w przedziale od 8 do 9. Oba histogramy są zbliżone do rozkładu normalnego, dlatego zmienna zależna w wersji podstawowej i zlogarytmowanej została przetestowana pod względem dopasowania do rozkładu normalnego za pomocą testu Jarque-Bera. Wyniki przeprowadzonych testów wskazały, że nie ma podstaw do odrzucenia hipotezy zerowej stanowiącej o rozkładzie normalnym dla zmiennej w postaci zlogarytmowanej – ln\_birth\_rate (p-value = 0.06457). W przypadku zmiennej birth\_rate hipoteza została odrzucona. Zatem w modelu wykorzystano zmienną ln\_birth\_rate.

Rys. 1. Histogramy współczynnika urodzeń (zmienna bez logarytmu – birth\_rate i zmienna zlogarytmowana – ln\_birth\_rate)



*Źródło:* Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

Rys. 2. Wykres pudełkowy dla zmiennej opisującej współczynnik urodzeń (ln\_birth\_rate)



*Źródło:* Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *ggplot2*.

**3.3.2. Zmienne objaśniające**

1.2.1. Suma przypadków zachorowań na COVID-19 (covid19\_cases)

1.2.2. Suma zgonów spowodowanych COVID-19 (covid19\_deaths)

1.2.3. Współczynnik małżeństw (marriage\_rate)

1.2.4. Współczynnik rozwodów (divorce\_rate)

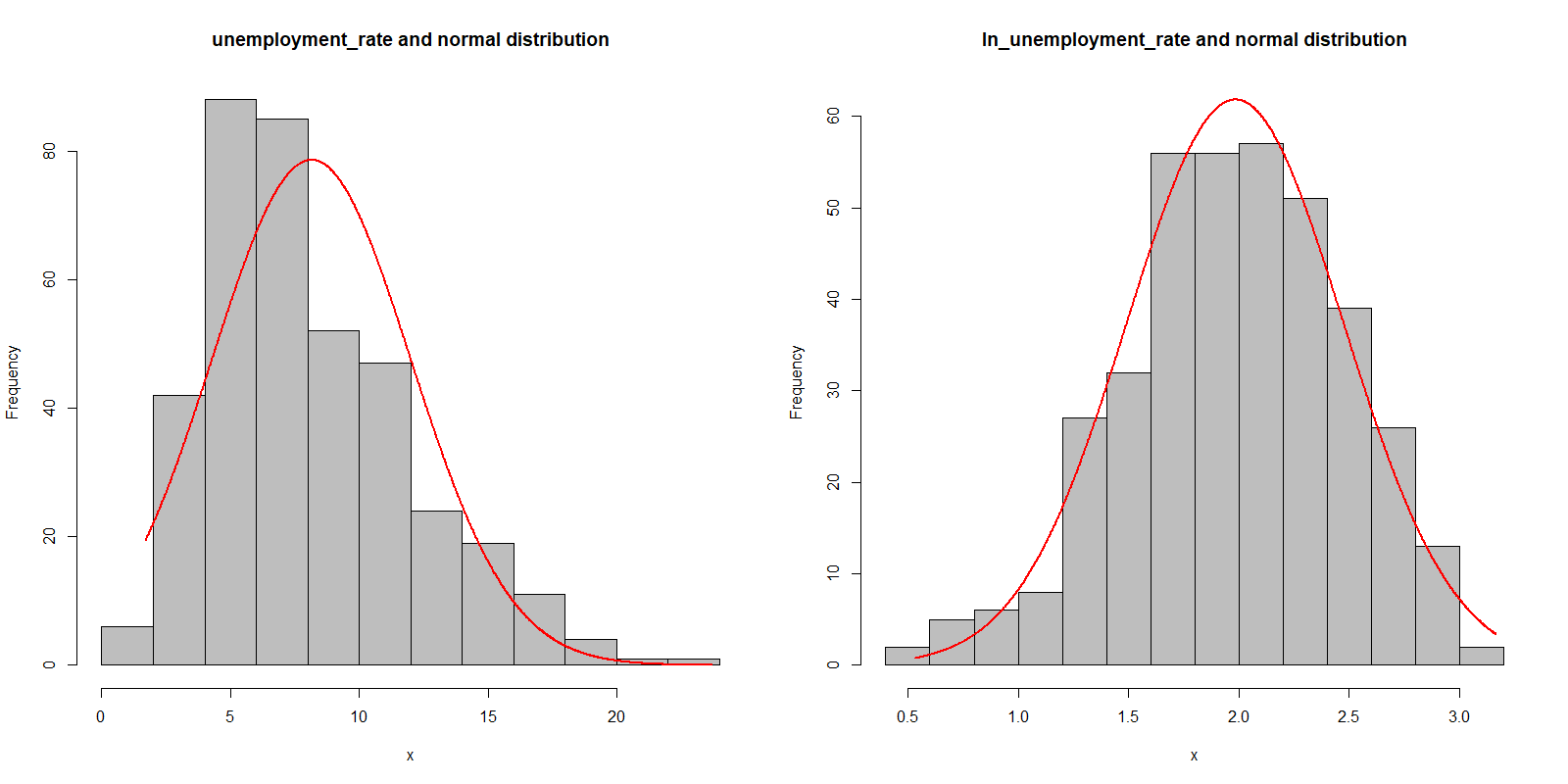
**1.2.1. Stopa bezrobocia rejestrowanego (unemployment\_rate)**

John Ermisch (1988) w swoim modelu zastosował logarytm ze zmiennej dotyczącej stopy bezrobocia wśród mężczyzn. W szacowanym modelu wykorzystano pomysł i zaproponowano logarytmowanie zmiennej mierzącej stopę bezrobocia rejestrowanego w każdym z powiatów (i miast na prawach powiatu).

Na podstawie histogramów (Rys. X.) możemy zauważyć, iż zaproponowana zlogarytmowana zmienna w większym stopniu przypomina zmienną o rozkładzie normalnym. Sprawdzając przypuszczenia za pomocą testu Jarque-Bera okazauje się, że w przypadku obu wersji odrzucamy hipotezę o normalności rozkładu, jednak dla ln\_unemployment\_rate p-value jest znacznie wyższe i prawie osiąga zakładany poziom istotności α = 0.05, ponieważ wynosi 4.807%.

Jeżeli przyjrzymy się histogramowi dla stopy bezrobocia w powiatach, to dla większości przyjmuje wartość znajdującą się w przedziale od 4 do 8 procent. Zmienna ma rozkład asymetryczny prawostronny, a jej najwyższa wartość to 23.7%.

Rys. 2. Histogram dla zmiennej opisującej stopę bezrobocia rejestrowanego w wersji podstawowej (unemployment\_rate) oraz zlogarytmowanej (ln\_unemployment\_rate)



*Źródło:* Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

1.2.1. women reproductive

**1.2.1. Średnia wartość wynagrodzenia (avg\_salary)**

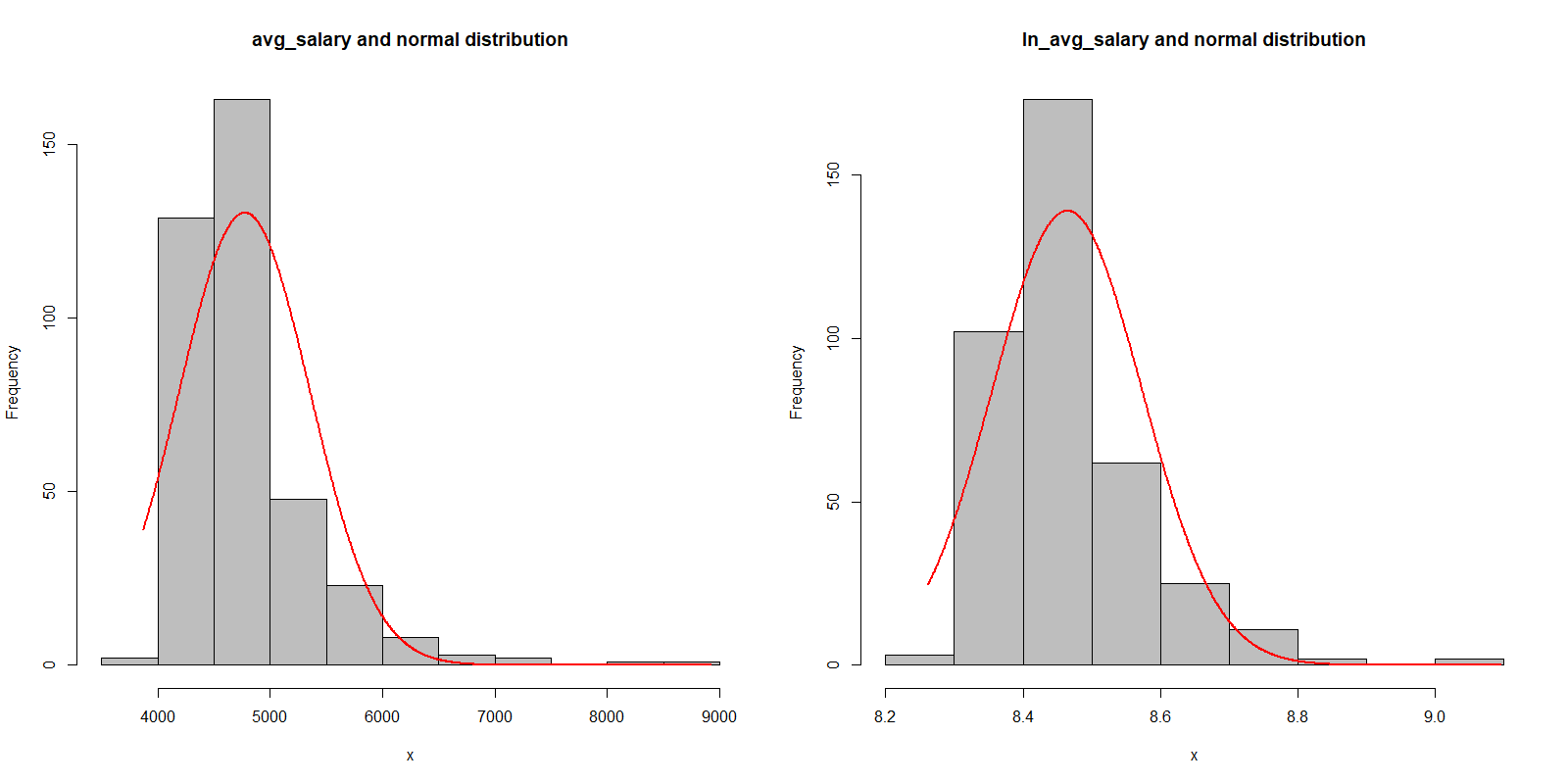
Gary Becker w artykule „An Economic Analysis of Fertility” jako jedną z determinant liczby urodzeń dzieci zaproponował wysokość dochodu. W niniejszej pracy proponuje użyć średniego wynagrodzenia jako zmiennej równoważnej zaproponowanej w artykule wymienionym powyżej.

Średnie miesięczne wynagrodzenie wynosi 4774 złotych, natomiast mediana znajduje się w przedziale od 4500 zł do 5000 zł. Na podstawie histogramu (rys. 2) można stwierdzić, że rozkład zmiennej jest asymetryczny prawostronny. Można na tej podstawie wnioskować, że mieszkańcy większości powiatów zarabiają mniej niż przeciętny Polak w 2020 roku (średnie wynagrodzenie w IV kwartale 2020 roku według Zakładu Ubezpieczeń Społecznych wynosiło 5457,98 zł).

Ermisch w artykule „Economic influences on birth rates” w swoim modelu szacującym współczynnik urodzeń zaproponował jako jedną ze zmiennych logarytm z realnej wartości zarobków mężczyzn. Inspirując się ekonomistą w modelu zastosowano logarytm ze średniej wartości wynagrodzenia.

Na podstawie histogramu (Rys. 2.) można stwierdzić, iż rozkład zmiennej avg\_salary nie jest rozkładem normalnym, co potwierdza wynik przeprowadzonego testu Jarque-Bera, który jednoznacznie stwierdza odrzucenie hipotezy zerowej o normalności rozkładu (p-value = 0). W przypadku zmiennej zlogarytmowanej, test Jarque-Bera również odrzuca hipotezę o rozkładzie normalnym, jednak histogram w większym stopniu przypomina rozkład normalny.

Rys. 2. Histogram dla zmiennej opisującej średnią wartość wynagrodzeń w wersji podstawowej (avg\_salary) oraz zlogarytmowanej (ln\_avg\_salary)



*Źródło:* Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

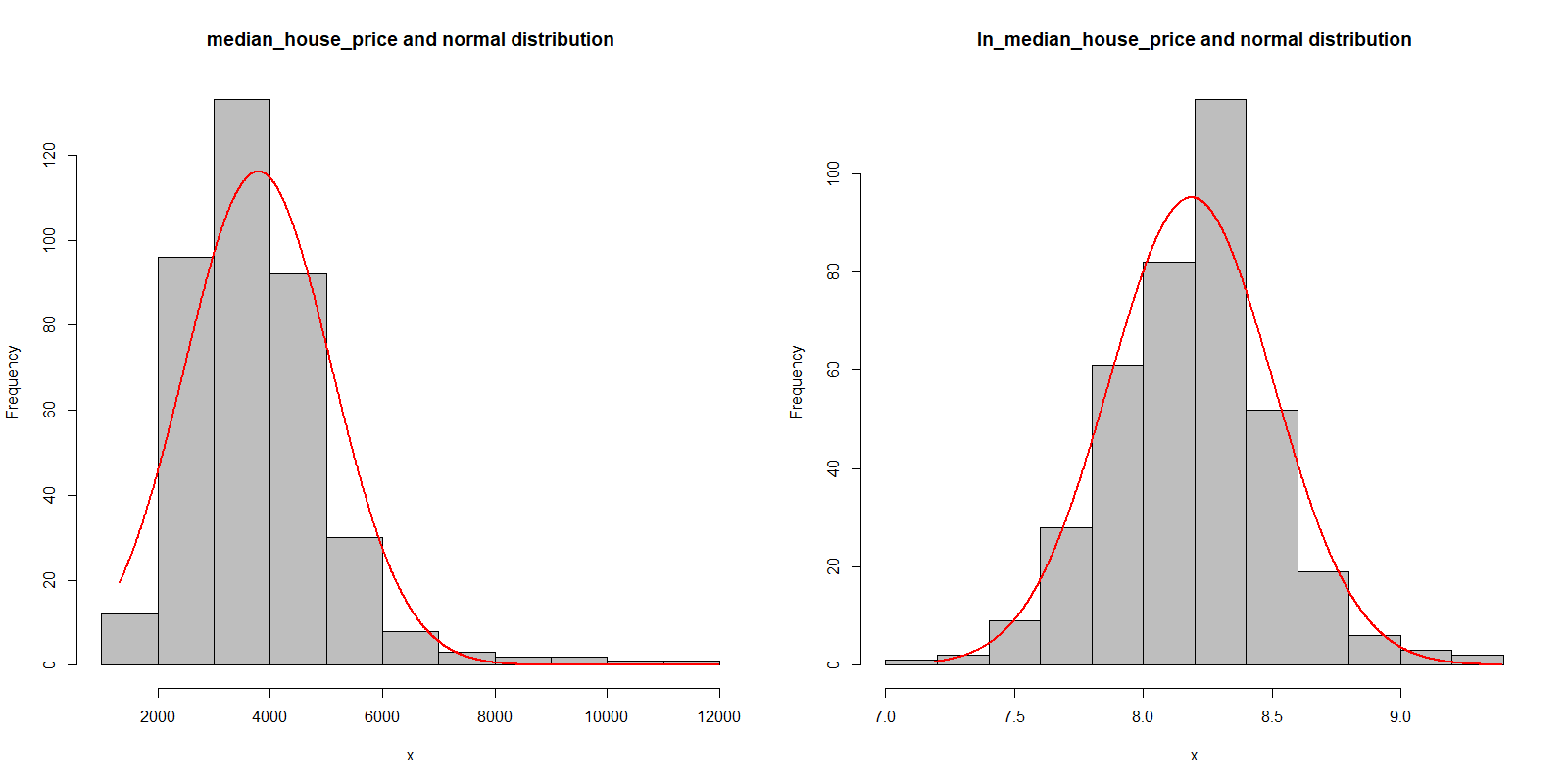
1.2.1. men working

**1.2.1. Mediana cen za 1 m2 lokali mieszkaniowych sprzedanych w ramach transakcji rynkowych (median\_house\_price)**

Na podstawie histogramu (Rys. C.) dla zmiennej median\_house\_price możemy wnioskować, że w większości powiatów najczęściej występująca cena za metr kwadratowy lokalu mieszkaniowego mieści się w przedziale od 3000 zł do 4000 zł. Średnia wartość zmiennej wynosi 3786,32 zł. Rozstrzał cenowy jeśli chodzi o mieszkania jest duży, ponieważ różnica między maksymalną a minimalną wartością median\_house\_price wynosi 10 663 złote. Najwyższe ceny notuje się w miastach, na przykład według strony muratorplus.pl przeciętna cena mieszkania w I półroczu 2020 roku wynosiła 10 800 zł za m2.

John Ermisch (1998) w swoim modelu szacującym wartość współczynnika urodzeń jako jedną ze zmiennych proponuje logarytm z realnej ceny domu. Chcąc wykorzystać myśl badacza w niniejszej pracy, zaproponowano użycie logarytmu ze zmiennej median\_house\_price w modelu. Histogram dla zmiennej zlogarytmowanej przypomina rozkład normalny. Chcąc sprawdzić przypuszczenia, przeprowadzono test na normalność rozkładu Jarque-Bera – zarówno dla median\_house\_price oraz jej zlogarytmowanej postaci odrzucono hipotezę o rozkładzie normalnym. Jednak w przypadku testu dla ln\_median\_house\_price zauważono wyższą wartość p-value (0.0002811).

Rys. 2. Histogram dla zmiennej opisującej medianę cen za 1 m2 lokali mieszkaniowych w wersji podstawowej (median\_house\_price) oraz zlogarytmowanej (ln\_median\_house\_price)



*Źródło:* Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

1.2.1. Suma przypadków zachorowań na COVID-19

1.2.1. Suma przypadków zachorowań na COVID-19

**rozdział 4. strategia empiryczna**

**4.1. metodologia badania**

METODY badania birth rates (ermisch szereg czasowy) ja uzylam mnk

**4.1. Formy funkcyjne i ich uzasadnienie**

W rozdziale x. punkcie x., gdzie przeanalizowano zmienną objaśnianą, czyli współczynnik urodzeń wskazano, że logarytm ze zmiennej birth\_rate ma rozkład normalny, dlatego szacujemy model ekonometryczny ze zmienną zależną ln\_birth rate. Postać analityczna wstępnego, pierwszego, równania regresji wygląda następująco:

Przyjmując poziom istotności α = 5%, hipoteza o łącznej nieistotności zmiennych w modelu 1. (załącznik x.) zostaje odrzucona (p-value < 0.05). Model tłumaczy w 76.96% zmienność logarytmu współczynnika urodzeń. Jednak aż 11 zmiennych jest pojedynczo nieistotnych statystycznie (p-value > 0.05), w związku z tym przeprowadzono kolejne iteracje pozwalające na osiągnięcie modelu zawierającego zmienne pojedynczo istotne statystycznie.

W kolejnym kroku dodano zmienne bedące potęgami oraz interakcjami wcześniej zastosowanych zmiennych. Do modelu dodano interakcję pomiędzy zmiennymi nursery\_places a children (chcąc przeanalizować zjawisko synergii między liczbą dzieci w wieku do 2 lat oraz liczbą miejsc w żłobkach w danym powiecie lub mieście na prawach powiatu). Kolejna interakcja zmiennych zachodzi pomiędzy education\_expenditure a children – w tym przypadku również możliwe jest zauważenie efektu wzmocnienia efektów.

Wyniki regresji modelu 2. (Załącznik 2.) poprawiły się w stosunku do wstępnej postaci, ponieważ skorygowany współczynnik determinacji wynosi 76.06% (w porównaniu z 75.53%). Iteracja ta tłumaczy 77.83% zmienności zlogarytmowanego współczynnika urodzeń. Pomimo, że wszystkie zmienne są łącznie istotne statystycznie, to 13 z nich jest pojedynczo nieistotna statystycznie, więc wnioskowanie na ich podstawie będzie obarczone błędem.

Chcemy również uwzględnić w modelu zlogarytmowane zmienne zaproponowane przez Ermischa w 1998 roku (ln\_unemployment\_rate, ln\_avg\_salary oraz ln\_median\_house\_price), z tego względu dokonano kolejnej iteracji.

Rezultaty regresji liniowej dla modelu 3 (załącznik x.) poprawiły się o 0.03 punkty procentowe w stosunku do poprzedniej iteracji. Model tłumaczy 78.11% zmienności logarytmu współczynnika wynagrodzeń, zmienne zastosowane w modelu są łącznie istotne statycznie, ale nadal wiele z nich jest pojedynczo nieistotna. W związku z tym przeprowadzono procedurę GETS (General-To-Specific), czyli zastosowano metodę „od ogółu do szczegółu” w celu usunięcia zmiennych pojedynczo nieistotnych statystycznie.

Początkowo usunięto zmienne: avg\_people\_per\_house, covid19\_quarantine, women\_working, houses\_area\_pc2, budget\_reve\_pc, houses\_area\_pc, children, nursery\_places, education\_expenditure, urbanisation\_rate, femininity\_ratio. Test hipotezy złożonej stanowiącej o łącznej nieistotności statystycznej powyższych zmiennych wykazał, że nie ma podstaw aby odrzucić hipotezę zerową (p-value = 0.2928 przy założeniu poziomu istotności α = 0.05). Test wykazywał również brak łącznej istotności statystycznej, gdy do powyżej wymienionych zmiennych dodano zmienną men\_working, jednak dla modelu 3. (załącznik x.) p-value dla tej zmiennej wynosiło 0.077266, więc postanowiono nie usuwać men\_working w tej iteracji.

W rezultacie (model 4. , załącznik x.) otrzymano model ze zmiennymi łącznie istotnymi statystycznie, jednak trzy z nich nadal były pojedynczo nieistotne. Model w porówaniu z poprzednimi wersjami stał się gorszy (skorygowany współczynnik determinacji na poziomie 76,23%), jednak nie zawiera on tak wielu zmiennych pojedynczo nieistotnych, co przekłada się na możliwość formułowania bardziej trafnych wniosków. Iteracja tłumaczy 77.29% zmienności zmiennej objaśnianej. Ze względu na to, że w modelu pozostały zmienne pojedynczo nieistotne, usunięto zmienną health\_expenditure, ponieważ miała najwyższe p-value wynoszące 7.963%.

W modelu 5. (załącznik x.) pozostaje jedna zmienna, której wartość p-value jest większa od zakładanego poziomu i wynosi 5.372%, jednak zdecydowano na pozostawienie men\_working w modelu, gdyż różnica między wartością p-value a poziomem istotności jest nieznaczna. Wersja modelu tłumaczy zmienność zmiennej zależnej w 77.1%, porównując za pomocą wartości skorygowanego R2 wersje modeli można uznać tą za najgorszą, jednak nie posiada ona zmiennych pojedynczo nieistotnych statystycznie, więc wnioskowanie będzie ułatwione.

**ROZDZIAŁ X. Diagnostyka modelu**

**5.1. Diagnostyka założeń Klasycznego Modelu Regresji Liniowej**

W celu zweryfikowania założeń Klasycznego Modelu Regresji Liniowej (KMRL) przeprowadzono testy diagnostyczne. Sprawdzenie założeń KMRL jest kluczowe, ponieważ dzięki nim możemy opierać się na poprawności estymatorów i móc na ich podstawie tworzyć wnioski (Mycielski, 2010).

**5.1.1. Założenie o prawidłowej formie funkcyjnej modelu**

Podczas weryfikacji hipotezy o liniowej formie modelu wykorzystano test RESET (Regression Specification Error Test). Testowana hipoteza zerowa oraz hipoteza alternatywa (zakładająca nieliniowość) według Mycielskiego (2010) brzmi następująco:

Test RESET dla modelu w wersji ostatecznej (Załącznik x., model 5.) wskazuje brak podstaw do odrzucenia hipotezy zerowej o poprawności formy funkcyjnej dla przyjętego poziomu istotności α = 0.05 (wartość p-value wynosi 0.4024). Założenie o liniowej formie funkcyjnej świadczy, że estymatory parametrów w modelu nie są obciążone, więc ich interpretowanie pozwala na wyciągnięcie statystycznie poprawnych wniosków.

**5.1.2. Założenie o homoskedastyczności reszt**

Weryfikację homoskedastyczności składnika losowego przeprowadzono za pomocą testu Breuscha-Pagana, którego hipoteza zerowa mówi o braku heteroskedastyczności. Dla przyjętego poziomu istotności 0.05 nie ma podstaw do odrzucenia hipotezy zerowej (p-value = 0.5237).

CO JEŚLI HETEROSKEDASTYCZNOSC

analiza ykresów

**5.1.3. Założenie o normalności rozkładu reszt ε**

Żeby sprawdzić czy możliwe jest przeprowadzenie wnioskowania statystycznego za pomocą standardowych metod przeprowadzono testowanie normalności składników losowych za pomocą testu Jarque-Bera. W powyższym teście hipoteza zerowa zakłada normalność rozkładu błędów losowych, natomiast hipoteza alternatywna jego brak (Mycielski, 2010). Testowane hipotezy:

Test Jarque-bera dla modelu w wersji ostatecznej (Załącznik x., model 5.) wskazuje na bark podstaw do odrzucenia hipotezy zerowej o normalności rozkładu reszt na poziomie istotności równym 0.05 (p-value = 0.2998). Założenie o normalności rozkładu reszt jest spełnione, dzięki temu można przeprowadzać prawidłowe wnioskowanie statystyczne w małych próbach.

analiza wykresow

**5.1.4. założenie o braku autokorelacji czynników losowych**

TESTY

CO JEŚLI AUTOKORELAJCA

**5.5. Współliniowość**

**ZAŁĄCZNIKI**

Załącznik nr 1. Podstawowe statystyki zmiennych dla początkowej wersji modelu

==========================================================================================

Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max

------------------------------------------------------------------------------------------

birth\_rate 380 8.844 1.355 5.650 7.898 9.530 15.430

covid19\_cases 380 2,422.084 3,397.959 361 1,014.8 2,654 46,889

covid19\_deaths 380 23.379 27.376 0 2.8 34 243

covid19\_quarantine 380 162,874.200 95,895.780 15,568 103,482 188,046.4 1,131,893

marriage\_rate 380 3.661 0.432 2.600 3.300 4.000 5.100

divorce\_rate 380 0.133 0.212 0.017 0.061 0.142 3.267

budget\_reve\_pc 380 2,512.856 2,473.393 811.660 1,276.720 1,764.895 14,753.110

unemployment\_rate 380 8.149 3.857 1.700 5.300 10.525 23.700

women\_reproductive 380 0.449 0.019 0.381 0.437 0.461 0.501

femininity\_ratio 380 104.897 3.899 96 102 107 119

avg\_salary 380 4,773.944 581.303 3,872.060 4,430.580 4,966.740 8,920.410

women\_working 380 0.277 0.065 0.136 0.230 0.318 0.586

men\_working 380 0.298 0.082 0.117 0.241 0.343 0.731

median\_house\_price 380 3,786.318 1,305.490 1,318 2,916.2 4,337.2 11,981

house\_ratio 380 2,007.856 8,435.287 0.100 2.400 8.125 118,023.000

houses\_area\_pc 380 28.639 3.120 22.861 26.508 30.301 44.216

avg\_people\_per\_house 380 2.784 0.374 1.758 2.566 3.045 4.005

education\_expenditure 380 1,995.955 806.579 1,077.544 1,473.283 2,324.202 6,276.602

health\_expenditure 380 50.545 118.756 6.781 17.093 44.389 1,827.289

social\_expenditure 380 360.896 180.647 133.097 237.094 434.694 1,162.852

family\_expenditure 380 1,819.767 669.389 1,173.412 1,436.539 1,723.992 3,877.993

children 380 5,005.884 6,918.297 779 2,410.2 5,296.8 105,383

nursery\_places 380 443.537 1,153.179 0 108.5 418 16,201

doctors 380 37.059 32.608 1.000 17.150 45.100 199.400

urbanisation\_rate 380 51.292 27.376 2.100 30.998 63.685 100.000

bus\_stops 380 437.532 356.981 82 235.5 540.2 4,342

------------------------------------------------------------------------------------------

Załącznik 1. Rezultaty regresji liniowej dla wstępnego modelu (model 1.)

=================================================

Dependent variable:

---------------------------

ln\_birth\_rate

-------------------------------------------------

covid19\_cases 0.00001

(0.00000)

covid19\_deaths -0.001\*\*\*

(0.0002)

covid19\_quarantine 0.00000

(0.00000)

marriage\_rate 0.093\*\*\*

(0.011)

divorce\_rate -0.462\*\*\*

(0.093)

budget\_reve\_pc 0.00000

(0.00000)

unemployment\_rate -0.004\*\*\*

(0.001)

women\_reproductive 3.009\*\*\*

(0.396)

femininity\_ratio -0.002

(0.002)

avg\_salary -0.00002\*

(0.00001)

women\_working -0.041

(0.138)

men\_working 0.194\*

(0.100)

median\_house\_price 0.00001\*\*

(0.00000)

house\_ratio -0.00000\*\*

(0.00000)

houses\_area\_pc -0.001

(0.002)

avg\_people\_per\_house -0.020

(0.027)

education\_expenditure 0.00001

(0.00002)

health\_expenditure -0.0001\*\*

(0.00004)

social\_expenditure -0.0002\*\*\*

(0.00005)

family\_expenditure 0.0001\*\*\*

(0.00002)

children 0.00002\*\*\*

(0.00000)

nursery\_places -0.00003\*\*

(0.00001)

doctors 0.001\*\*\*

(0.0002)

urbanisation\_rate 0.0002

(0.0002)

bus\_stops -0.0001\*\*\*

(0.00002)

Constant 0.688\*\*

(0.345)

-------------------------------------------------

Observations 380

R2 0.770

Adjusted R2 0.753

Residual Std. Error 0.074

F Statistic 47.289\*\*\*

=================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Załącznik 1. Rezultaty regresji liniowej dla wstępnego modelu (model 2.)

==========================================================

Dependent variable:

---------------------------

ln\_birth\_rate

----------------------------------------------------------

covid19\_cases 0.00001\*

(0.00000)

covid19\_deaths -0.001\*\*\*

(0.0002)

covid19\_quarantine 0.000

(0.00000)

marriage\_rate 0.087\*\*\*

(0.012)

divorce\_rate -0.388\*\*\*

(0.095)

budget\_reve\_pc 0.00000

(0.00000)

unemployment\_rate -0.004\*\*\*

(0.001)

women\_reproductive 2.422\*\*\*

(0.421)

femininity\_ratio -0.003

(0.002)

avg\_salary -0.00002\*\*

(0.00001)

women\_working -0.049

(0.136)

men\_working 0.200\*\*

(0.099)

median\_house\_price 0.00001\*\*\*

(0.00000)

house\_ratio -0.00001\*\*\*

(0.00000)

houses\_area\_pc -0.010

(0.015)

houses\_area\_pc\_2 0.0001

(0.0002)

avg\_people\_per\_house -0.001

(0.027)

education\_expenditure -0.00002

(0.00002)

health\_expenditure -0.0001\*

(0.00004)

social\_expenditure -0.0002\*\*\*

(0.00005)

family\_expenditure 0.0001\*\*\*

(0.00002)

children 0.00001

(0.00001)

nursery\_places -0.00001

(0.00001)

doctors 0.001\*\*

(0.0003)

urbanisation\_rate 0.0002

(0.0002)

bus\_stops -0.0001\*\*\*

(0.00002)

nurseryxchildren -0.000\*\*\*

(0.000)

education\_expenditurexchildren 0.000\*\*\*

(0.000)

Constant 1.150\*\*\*

(0.428)

----------------------------------------------------------

Observations 380

R2 0.778

Adjusted R2 0.761

Residual Std. Error 0.073

F Statistic 44.006\*\*\*

==========================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Załącznik 3. Rezultaty regresji liniowej dla iteracji 3, 4 i 5 modelu.

==================================================================

Dependent variable:

-----------------------------------

ln\_birth\_rate

(3) (4) (5)

------------------------------------------------------------------

covid19\_cases 0.00001\* 0.00001\*\* 0.00001\*\*

(0.00000) (0.00000) (0.00000)

covid19\_deaths -0.001\*\*\* -0.0005\*\*\* -0.0005\*\*\*

(0.0002) (0.0002) (0.0002)

covid19\_quarantine -0.000

(0.00000)

marriage\_rate 0.085\*\*\* 0.092\*\*\* 0.090\*\*\*

(0.012) (0.011) (0.011)

divorce\_rate -0.396\*\*\* -0.359\*\*\* -0.354\*\*\*

(0.094) (0.080) (0.080)

budget\_reve\_pc 0.00000

(0.00000)

ln\_unemployment\_rate -0.036\*\*\* -0.029\*\* -0.028\*\*

(0.013) (0.012) (0.012)

women\_reproductive 2.369\*\*\* 2.664\*\*\* 2.735\*\*\*

(0.417) (0.292) (0.290)

femininity\_ratio -0.003

(0.002)

ln\_avg\_salary -0.088\* -0.088\* -0.099\*\*

(0.049) (0.045) (0.044)

women\_working -0.053

(0.135)

men\_working 0.175\* 0.116\* 0.121\*

(0.099) (0.063) (0.063)

ln\_median\_house\_price 0.063\*\*\* 0.050\*\*\* 0.047\*\*\*

(0.017) (0.016) (0.016)

house\_ratio -0.00001\*\*\* -0.00001\*\*\* -0.00001\*\*\*

(0.00000) (0.00000) (0.00000)

houses\_area\_pc -0.008

(0.014)

houses\_area\_pc\_2 0.0001

(0.0002)

avg\_people\_per\_house 0.0003

(0.027)

education\_expenditure -0.00002

(0.00002)

health\_expenditure -0.0001\* -0.0001\*

(0.00004) (0.00003)

social\_expenditure -0.0002\*\*\* -0.0002\*\*\* -0.0002\*\*\*

(0.00005) (0.00004) (0.00004)

family\_expenditure 0.0001\*\*\* 0.0001\*\*\* 0.0001\*\*\*

(0.00002) (0.00001) (0.00001)

children 0.00001

(0.00001)

nursery\_places -0.00001

(0.00001)

doctors 0.001\*\* 0.0005\*\* 0.0005\*\*

(0.0003) (0.0002) (0.0002)

urbanisation\_rate 0.0002

(0.0002)

bus\_stops -0.0001\*\*\* -0.0001\*\*\* -0.0001\*\*\*

(0.00002) (0.00002) (0.00002)

nurseryxchildren -0.000\*\*\* -0.000\*\*\* -0.000\*\*\*

(0.000) (0.000) (0.000)

education\_expenditurexchildren 0.000\*\*\* 0.000\*\*\* 0.000\*\*\*

(0.000) (0.000) (0.000)

Constant 1.394\*\* 0.902\*\* 0.997\*\*

(0.603) (0.422) (0.420)

------------------------------------------------------------------

Observations 380 380 380

R2 0.781 0.773 0.771

Adjusted R2 0.764 0.762 0.761

Residual Std. Error 0.073 0.073 0.073

F Statistic 44.724\*\*\* 72.477\*\*\* 76.374\*\*\*

==================================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

bibliografia

<https://docs.google.com/spreadsheets/d/1Tv6jKMUYdK6ws6SxxAsHVxZbglZfisC8x_HZ1jacmBM/edit#gid=42091932>

<https://bdl.stat.gov.pl/BDL/start>

baza danych covidowa mz gov

wykres birth rate

<https://data.worldbank.org/indicator/SP.DYN.CBRT.IN?locations=EU>

wykres fertility rate

https://data.worldbank.org/indicator/SP.DYN.TFRT.IN