



SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
WARSAW SCHOOL OF ECONOMICS

Mikroekonometria

Badanie czynników warunkujących dzietność w rodzinie w Stanach Zjednoczonych za pomocą regresji Poissona

Justyna Zbiegień
Sebastian Krzyślak

Warszawa, maj 2021

Spis treści

Wprowadzenie	3
Opis zmiennych	3
Badania empiryczne	9
Estymacja oszacowań oraz weryfikacja modelu	9
Interpretacja wyników modelowania.....	13
Wnioski i zakończenie	14
Bibliografia.....	15
Spis tabel i rysunków	15

Wprowadzenie

Starzenie się społeczeństwa w ostatnich latach stało się tematem rezonującym przy każdej dyskusji o sytuacji ekonomicznej państw wysoko rozwiniętych. Gwałtowny wzrost technologiczny, zwłaszcza w medycynie, oraz mniejsza liczba urodzeń są decydującymi czynnikami wpływającymi na podnoszenie się średniego wieku społeczeństwa. Stany Zjednoczone jako jedno z największych gospodarek światowych posiadający wysoki współczynnik HDI nie jest wyjątkiem w tym wypadku. Starzenie się społeczeństwa w tym kraju jest zauważalne gołym okiem, w roku 2019 mediana wynosiła 38.4 lat, natomiast 20 lat wcześniej było to niewiele ponad 35 lat¹. Zauważono również znaczący spadek urodzeń, który jak wspomniano wyżej, jest jednym z dwóch głównych czynników wpływających na zmiany w strukturze wieku społeczeństwa. Właśnie ten czynnik wzięto pod zakres badań w niniejszej pracy.

W 2020 średnia liczba dzieci, którą w ciągu swojego życia posiadzie jedna Amerykanka wyniosła 1.78, porównując w 1800 roku liczba ta sięgała powyżej 7, natomiast w roku 1900 niecałe 4². To pokazuje, jak wielki jest spadek dzietności na przestrzeni tak krótkiego okresu czasu. Zdecydowano się zbadać, co wpływa i w jakim stopniu na decyzję podejmowaną przez potencjalnych rodziców dotyczącą liczby posiadanych dzieci. Skupiono się na czynnikach zewnętrznych takich, jak poziom edukacji rodziców, ich poglądy polityczne, status matrymonialny oraz kolor skóry. Zdecydowano się na zbudowanie modelu bazującego na regresji Poissona na wzór badania przeprowadzonego przez Martina Fiedera oraz Susanne Huber dla *frontiers in Psychology* zatytułowanym "Political Attitude and Fertility: Is There a Selection for the Political Extreme?".

Opis zmiennych

Dane zostały zebrane ze strony GSS (the General Social Survey) dla roku 2018. W badaniu przepytano 1039 osób w różnym wieku, o różnej płci, rasie czy wykształceniu. Poniższa tabela przedstawia wykaz wszystkich zmiennych przyjętych do dalszych badań:

Tabela 1: Opis zmiennych

Zmienna	Typ	Opis
children	Ilościowy	Liczba posiadanych dzieci
age	Ilościowy	Liczba lat respondenta
marital	Kategoryczna	Status matrymonialny respondenta
educ	Ilościowy	Najwyższy rok szkoły zakończony przez respondenta
race	Kategoryczny	Kolor skóry respondenta
polviews	Kategoryczny	Skala od 1 do 7, według której respondenta określa swoje poglądy polityczne (1 – skrajnie konserwatywne, 7 - skrajnie liberalne)
satfin	Kategoryczny	Poziom satysfakcji z sytuacji finansowej respondenta (1 - w ogóle nieusatysfakcjonowany, 2 - średnio usatysfakcjonowany, 3 - bardzo usatysfakcjonowany)
wrkstat	Binarny	Status zawodowy respondenta (1 – osoba pracująca, 0 – osoba niepracująca)
income	Kategoryczny	Zarobki całego gospodarstwa domowego wyrażone w kategorii od 1 do 12 (1 – zarobki poniżej 1000\$ rocznie, 12 – zarobki powyżej 25000\$ rocznie)

¹ Dane pochodzą z serwisu statista.com, <https://www.statista.com/statistics/241494/median-age-of-the-us-population/> [data dostępu: 06.05.2021 13:10]

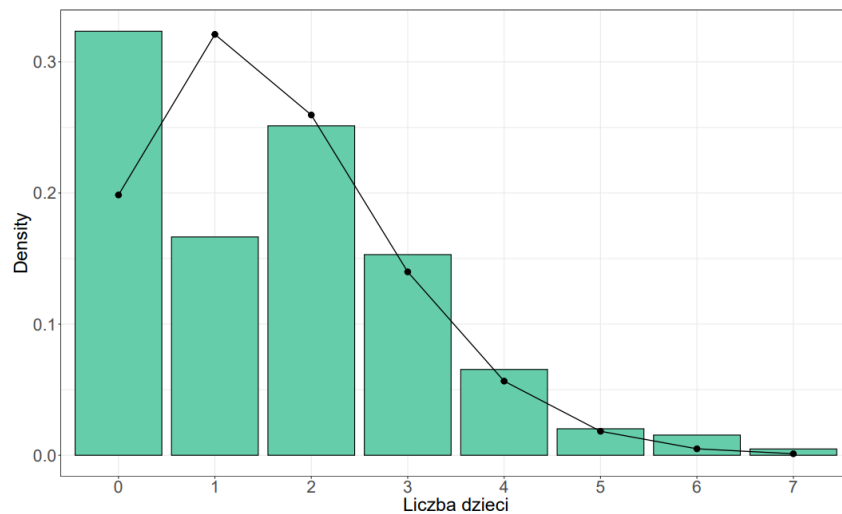
² Dane pochodzą z serwisu statista.com, <https://www.statista.com/statistics/1033027/fertility-rate-us-1800-2020/> [data dostępu: 06.05.2021 13:12]

degree	Kategoryczny	Najwyższy stopień szkolnictwa ukończony przez respondenta (1 – poniżej szkoły średniej, 2 – szkoła średnia, 3 – studia dwuletnie, 4 – I stopień studiów, 5 – II stopień studiów i wyżej)
---------------	--------------	--

Źródło: Opracowanie własne

Zmienną objaśnianą jest zmienna *children*, czyli liczba dzieci posiadanych przez respondenta w momencie badania. Średnio respondenci posiadają 1.62 dzieci, a wariancja dla tej zmiennej wyniosła 2.28, co już na tym etapie wskazało na brak możliwości estymacji modelu metodą KMNK ze względu na występowanie heteroskedastyczności. Poniższy wykres pokazuje rozkład liczby posiadanych dzieci (Rys.1). Od razu można spostrzec, że zmienna objaśniana posiada rozkład Poissona, co pokazuje linia ciągła:

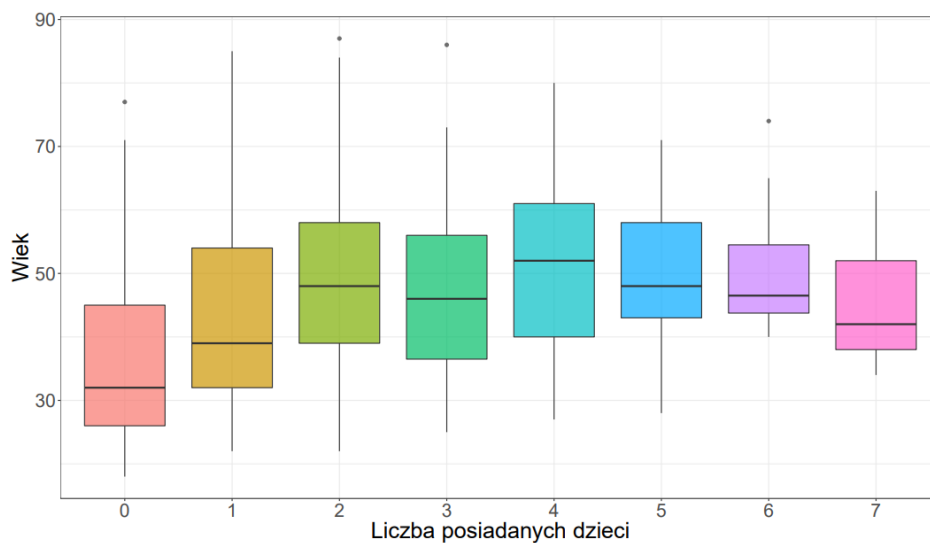
Rysunek 1: Rozkład posiadanych dzieci



Źródło: Opracowanie własne

W badaniu przepytano respondentów z przedziału wiekowego 18-87, ale średnia wieku respondentów jest równa około 44 lata. Poniższy wykres (Rys.2) przedstawia wykres pokazujący liczbę dzieci w zależności do wieku rodzica:

Rysunek 2: Liczba posiadanych dzieci w zależności od wieku respondenta



Źródło: Opracowanie własne

Jak wynika z powyższego wykresu widać tendencję wzrostową w liczbie posiadanych dzieci w zależności od wieku, co nie jest zaskoczeniem. Odpowiedzi „0” dominują wśród ludzi młodych, którzy prawdopodobnie jeszcze nie zdecydowali się na założenie rodziny.

Średnio respondenci wydają się być usatysfakcjonowani swoją sytuacją finansową (średnia ze zmiennej *satfin* jest równa w przybliżeniu 2). Poniższa tabela przedstawia średnią liczbę posiadanych dzieci w zależności od poszczególnych odpowiedzi oraz ich liczebność.

Tabela 2: Zmienna *satfin*

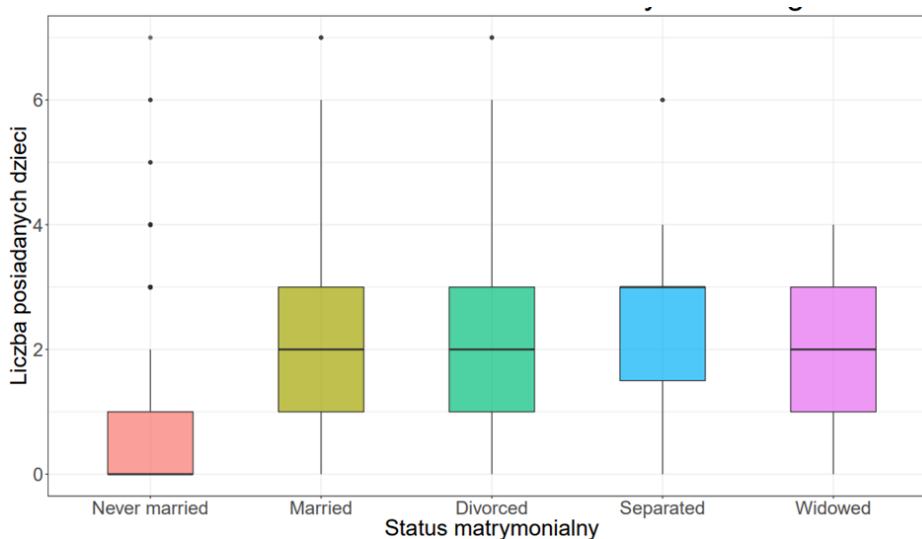
	Brak usatysfakcjonowania	Średnie usatysfakcjonowanie	Duże usatysfakcjonowanie
Liczebność odpowiedzi	259	498	282
Średnia liczba dzieci	1.76	1.59	1.53

Źródło: Opracowanie własne

Jak widać z powyższej tabeli widnieje tendencja spadkowa w posiadanej liczbie dzieci w zależności od satysfakcji z sytuacji finansowej.

Okolo 33% respondentów zaznaczyło, że nigdy nie było w związku małżeńskim i aż dwie trzecie respondentów kiedykolwiek znajdowała się w formalnym związku. Poniższy wykres (Rys.3) przedstawia liczbę posiadanych dzieci w zależności od statusu matrymonialnego:

Rysunek 3: Liczba dzieci w zależności od statusu matrymonialnego respondenta



Źródło: Opracowanie własne

Zdecydowano się przekształcić zmienną *marital* na zmienną binarną – czy kiedykolwiek osoba była w związku małżeńskim -1 lub nigdy nie była - 0. Podjęto taką decyzję, ponieważ grupy osób będących w małżeństwie, rozwiedzionych, w separacji oraz owdowiałych mają podobny rozkład, co widać na powyższym wykresie. Poniższa tabela przedstawia liczebności odpowiedzi już przegrupowaniu zmiennej:

Tabela 3: Zmienna marital

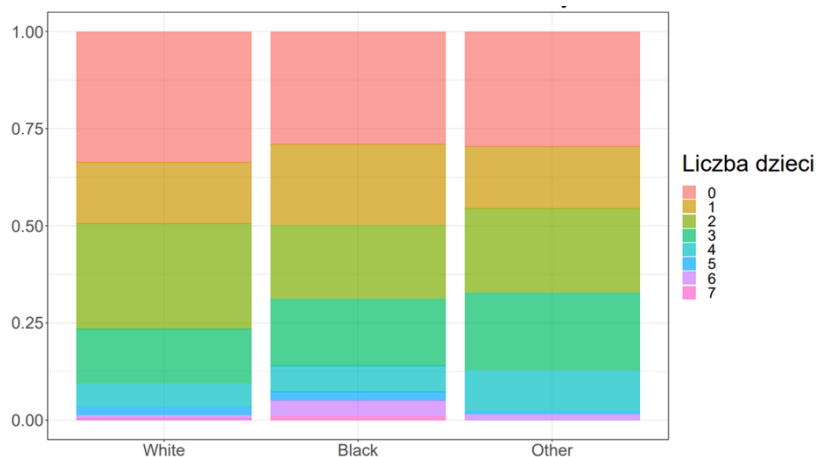
	Osoby kiedykolwiek będące w związku małżeńskim	Osoby nigdy nie będące w związku małżeńskim
Liczebność odpowiedzi	701	338
Średnia liczba dzieci	2.09	0.64

Źródło: Opracowanie własne

Jak wynika z powyższej tabeli osoby będące w małżeństwie posiadają średnio znacznie więcej dzieci niż osoby nigdy nie będące w takim związku.

Należy zwrócić również uwagę, że większość respondentów zaznaczyło swoją rasę jako białą, bo aż 71%. Druga najliczniejsza grupa w tej zmiennej to rasa czarna stanowiąca 17% odpowiedzi. Pozostałe odpowiedzi można zgrupować jako inne, które w takim wypadku stanowią około 12% respondentów. Są to wyniki odzwierciedlające rzeczywistą sytuację rasową w Stanach Zjednoczonych. Poniższy wykres (Rys.4) przedstawia liczbę posiadanych dzieci w zależności od koloru skóry:

Rysunek 4: Liczba dzieci w zależności od koloru skóry rodzica



Źródło: Opracowanie własne

Poniższa tabela przedstawia liczebności zmiennej *race* oraz średnią liczbę dzieci w zależności od koloru skóry:

Tabela 4: Zmienna race

Kolor skóry:	Biały	Czarny	Inny
Liczebność odpowiedzi	735	179	125
Średnia liczba dzieci	1.55	1.80	1.74

Źródło: Opracowanie własne

Jak wynika z powyższej tabeli kolor skóry ma wpływ na średnią liczbę dzieci, ponieważ widać znaczącą różnicę między kolorem skóry białym a innymi rasami. Osoby z białym kolorem skóry posiadają statystycznie mniej dzieci.

Zmienną budzącą największe obawy jest zmienna *income*, która przyjmuje głównie wartości równe 12 – oznacza to, że większość respondentów zarabia powyżej 25,000\$ rocznie, co nie jest dużym zaskoczeniem, ponieważ w 2018 mediana zarobków przypadających na jedno gospodarstwo domowe

wyniosła 64,324\$³. Przez brak dużej dywersyfikacji wśród odpowiedzi zmienna posiada małą zdolność dyskryminacyjną i może okazać się mało przydatna przy budowie modelu. Poniższa tabela przedstawia rozkład tej zmiennej:

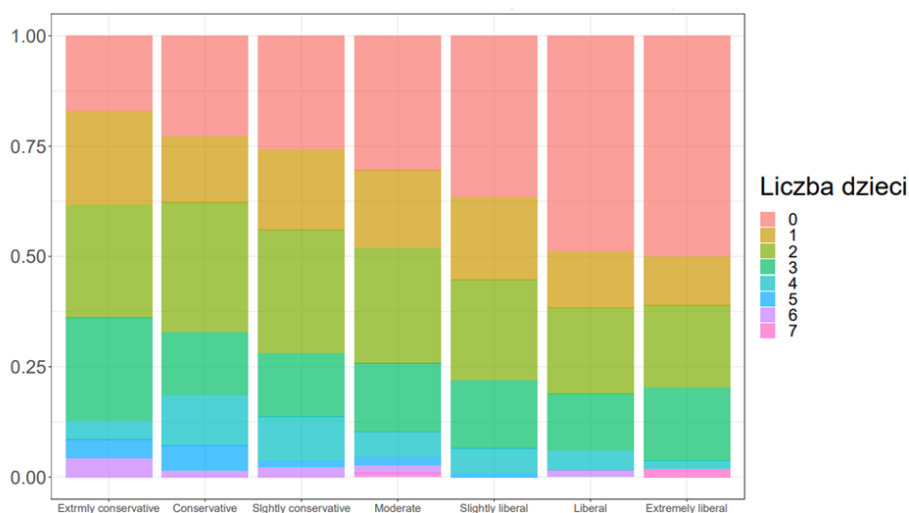
Tabela 5: Zmienna income

	1	2	3	4	5	6	7	8	9	10	11	12
	<1000\$	\$1000 - 2999	\$3000 - 3999	\$4000 - 4999	\$5000 - 5999	\$6000 - 6999	\$7000 - 7999	\$8000 - 9999	\$10000- 14999	\$15000- 19999	\$20000- 24999	>25000\$
Liczebność odpowiedzi	5	6	6	3	4	1	3	7	32	32	48	892
Średnia liczba dzieci	1	1.5	0.83	0.67	1.75	0	1.67	1.29	1.94	2.03	1.83	1.6

Źródło: Opracowanie własne

Natomiast najciekawszą zmienną jest zmienna *polviews* dotycząca poglądów politycznych. Liczne badania wykazały, że poglądy polityczne mają duży wpływ na liczbę dzieci w rodzinie. Według artykułu napisanego przez Martina Fiedera oraz Susanne Huber pt. "Political Attitude and Fertility: Is There a Selection for the Political Extreme?" skrajne poglądy (konserwatywne oraz liberalne) stymulująco działają na liczbę posiadanych dzieci, natomiast osoby o centrowych poglądach charakteryzują się posiadaniem mniejszej liczby dzieci. Poniższy wykres (Rys.5) przedstawia liczbę posiadanych dzieci w zależności od poglądów politycznych rodzica:

Rysunek 5: Liczba dzieci w zależności od poglądów politycznych rodzica



Źródło: Opracowanie własne

Jak wynika z powyższego wykresu wyniki badań Martina Fiedera oraz Susanne Huber nie potwierdziły się. Znacznie większą liczbą posiadanych dzieci odznaczają się jedynie osoby o poglądach skrajnie konserwatywnych, a osoby o poglądach skrajnie liberalnych wręcz przeciwnie – mniejszą liczbą dzieci. Poniższa tabela przedstawia rozkład zmiennej *polviews*:

³ Dane pochodzą z artykułu K. Amadeo „What is the average income in the United States?” opublikowany na stronie thebalance.com <https://www.thebalance.com/what-is-average-income-in-usa-family-household-history-3306189> [data dostępu: 08.05.2021 1:31]

Tabela 6: Zmienna polviews

	1	2	3	4	5	6	7
Liczebność odpowiedzi	47	140	132	410	123	133	54
Średnia liczba dzieci	2.06	1.99	1.78	1.66	1.37	1.17	1.19

Źródło: Opracowanie własne

Kolejną zmienną jest *wrkstat* odnosząca się do aktualnego stanu zatrudnienia respondenta. Widać, że zdecydowanie dominują odpowiedzi „osoba pracująca” – jest to około 90% odpowiedzi. Poniższa tabela przedstawia te zależności:

Tabela 7: Zmienna wrkstat

	Osoby pracujące	Osoby niepracujące
Liczebność odpowiedzi	906	132
Średnia liczba dzieci	1.62	1.59

Źródło: Opracowanie własne

Według powyższej tabeli nie ma znaczącej różnicy między osobą pracującymi i niepracującymi pod względem liczby posiadanych dzieci.

Pozostałe dwie zmienne: *educ* i *degree* odnoszą się do edukacji respondentów. Ze względu na swój zbliżony charakter mogą się wykazać dużą współliniowością i najprawdopodobniej tylko jedna z nich zostanie użyta do budowy modelu. Poniższe tabele ukazują rozkład obu zmiennych, jednak dla zmiennej *degree* nie pokazano respondentów, którzy nie ukończyli szkoły podstawowej (czyli poniżej 10 klasy) – jest takich osób 34.

Tabela 8: Zmienna degree

	10	11	12	13	14	15	16	17	18	19	20
Liczebność odpowiedzi	17	27	267	78	152	68	229	51	56	27	33
Średnia liczba dzieci	2.29	2.52	1.69	1.64	1.81	1.38	1.39	1.33	1.45	1.41	1.18

Źródło: Opracowanie własne

Jak widać według powyższej tabeli, liczba posiadanych dzieci maleje wraz ze wzrostem edukacji rodziców. Poniższa tabela przedstawia rozkład zmiennej *educ*, gdzie pominięto jedynie trzech respondentów, którzy nie ukończyli żadnego stopnia edukacji:

Tabela 9: Zmienna educ

	1	2	3	4	5
Liczebność odpowiedzi	72	490	110	241	123
Średnia liczba dzieci	1.82	1.55	1.62	1.64	1.71

Źródło: Opracowanie własne

Zmienna *educ* już nie wskazuje jednoznacznie na ujemną korelację pomiędzy stopniem edukacji a liczbą posiadanych dzieci tak, jak było to w przypadku zmiennej *degree*. Jednak podkreśla to fakt, że stopień wykształcenia rodziców ma wpływ na liczbę posiadanych przez nich dzieci.

Badania empiryczne

Niniejszy rozdział jest poświęcony wybranej metodzie modelowania, jej specyfikacji, a także motywacji, która stała za wybraniem właśnie takiego modelu. Jak wspomniano w poprzednim rozdziale zmienną objaśnianą w niniejszym badaniu jest posiadana liczba dzieci przez respondenta. Do modelowania omawianej zmiennej wykorzystano model regresji Poissona. Zdecydowano się właśnie na ten model z szeregu przyczyn. Pierwszą z nich jest sam rodzaj zmiennej – czyli dodatnia, całkowitoliczbowa, licznikowa czyli właśnie nośnik rozkładu Poissona, ponad to zmienna ma zbliżony rozkład do rozkładu Poissona, co widać na Rys.1 Kolejną przyczyną jest zakres wartości jakie zmienna przyjmuje – w analizowanym zbiorze danych od 0 do 7, ciężko sobie wyobrazić większe liczby np. 20, 50 czy 100 przy omawianiu posiadania liczby dzieci. Ostatnim powodem jest główna cecha modelu regresji Poissona czyli równomierne rozproszenie średniej oraz wariancji. W przypadku zbioru danych wykorzystanych w tym raporcie średnia liczba posiadanych dzieci wynosi 1.62, natomiast wariancja 2.28 co wskazuje lekką dyspersję. Przy takim stanie rzeczy wykorzystanie klasycznego modelu regresji liniowej byłoby niewskazane z powodu niespełnienia założenia o heteroskedastyczności składnika losowego, a także przez fakt, że prognozowane wartości mogłyby być ujemne, co w przypadku zmiennych całkowitoliczbowych dodatnich byłoby niemożliwe. Na początku został rozważony podstawowy model regresji Poissona, natomiast w dalszych rozważaniach uwzględniono również nadmierną dyspersję oszacowując ujemny model dwumianowy oraz dużą liczbę zer oszacowując model z nadmierną liczbą zer (Zero inflated Poisson model, później nazywany modelem ZIP).

Wstępna specyfikacja modelu prezentowała się następująco:

$$children_i = age_i + marital_i + educ_i + race_i + polviews_i + satfin_i + wrkstat_i + income_i + degree_i$$

Estymacja oszacowań oraz weryfikacja modelu

Proces modelowania opierał się na krokowym usuwaniu zmiennych objaśniających, które zostały uznane za nieistotne przy poziomie istotności $\alpha = 0.1$. Model wyjściowy przedstawiony w powyższej części został zredukowany o zmienne nieistotnie statystyczne przy przyjętym poziomie istotności (0.1). Ostateczna specyfikacja modelu przybrała postać:

$$children_i = age_i + marital_i + educ_i + race_i + polviews_i + satfin_i$$

Zmienne *wrkstat*, *income* oraz *degree* okazały się nie mieć statystycznie istotnego wpływu na posiadaną liczbę dzieci w badanym zbiorze danych, czego się spodziewano. Zmienna *wrkstat* prawdopodobnie została uznana za nieistotną z powodu słabej zdolności dyskryminacyjnej wynikającej z nieproporcjonalnej liczby osób pracujących (87.2%) do osób niepracujących (12.8%). W przypadku zmiennej *income* wcześniejsze obawy sprawdziły się. Do grupy z zarobkami rodziny powyżej \$25 tysięcy należy 86% respondentów, co powoduje, że zmienna podobnie jak poprzedni regresor słabo objaśnia liczbę posiadanych dzieci. Zmienna *degree* mówiąca o poziomie wykształcenia została również odrzucona, co jest prawdopodobnie spowodowane powieleniem informacji, którą niesie zmienna *educ* (liczba lat edukacji respondenta).

Na rysunku poniżej zaprezentowano oszacowania wyżej opisanego modelu:

Rysunek 6: Model 1

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5204  -1.0342  -0.1683   0.4933   4.9599

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.577184   0.181304   3.184  0.00146 **
age          0.010683   0.001950   5.478 4.30e-08 ***
marital      1.055330   0.079097  13.342 < 2e-16 ***
educ        -0.042971   0.008471  -5.073 3.92e-07 ***
raceOther   -0.246097   0.089527  -2.749  0.00598 **
raceWhite   -0.415642   0.065606  -6.335 2.37e-10 ***
polviews    -0.082872   0.017235  -4.808 1.52e-06 ***
satfin      -0.086430   0.034374  -2.514  0.01192 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1729.5  on 1038  degrees of freedom
Residual deviance: 1260.7  on 1031  degrees of freedom
AIC: 3153.3

```

Źródło: Opracowanie własne

Na początku raportu przy analizowaniu zmiennej objaśnianej (liczby posiadanych dzieci) zauważono, że wariancja jest nieco większa niż średnia (odpowiednio 2.28 i 1.62), co może sugerować nadmierne rozproszenie. W celu sprawdzenia tego efektu oszacowano ujemny model dwumianowy. Porównanie obu modeli przedstawiono poniżej:

Rysunek 7: Porównanie modeli – model 1 i ujemny model dwumianowy

	Dependent variable:	
	Poisson (1)	childs negative binomial (2)
age	0.011*** (0.002)	0.011*** (0.002)
marital	1.055*** (0.079)	1.057*** (0.080)
educ	-0.043*** (0.008)	-0.043*** (0.009)
raceOther	-0.246*** (0.090)	-0.249*** (0.091)
raceWhite	-0.416*** (0.066)	-0.418*** (0.067)
polviews	-0.083*** (0.017)	-0.083*** (0.018)
satfin	-0.086** (0.034)	-0.088** (0.035)
Constant	0.577*** (0.181)	0.583*** (0.185)
Observations	1,039	1,039
Log Likelihood	-1,568.652	-1,569.404
theta		52.680 (76.545)
Akaike Inf. Crit.	3,153.304	3,154.809
Note:	*p<0.1; **p<0.05; ***p<0.01	

Źródło: Opracowanie własne

Jak można zaobserwować oszacowania prawie się nie zmieniły, a błędy standardowe oszacowań wzrosły nieznacznie. Ponadto kryterium informacyjne Akaike'a przemawia na korzyść modelu regresji Poissona.

Równie nurtującą kwestią jest nadmierna liczba zer badanej zmiennej. Przyjmując średnią liczbę dzieci z badanego zbioru danych (1.62) prawdopodobieństwo wystąpienia 0 w rozkładzie Poissona z parametrem o tej wartości wynosi 19.79%, natomiast liczba zer w zbiorze danych stanowi aż 32.34%. Oznacza to, że w rzeczywistości wartości zerowe występują o wiele częściej, niż w zakładanym rozkładzie teoretycznym. Oszacowując taki zbiór danych za pomocą regresji Poissona lub ujemnym modelem dwumianowym narażamy się na słabą jakość modelu. W celu uwzględnienia tej obserwacji w modelowaniu oszacowano model regresji Poissona z nadmierną liczbą zer (Zero inflated Poisson model). Oszacowania części Poissona oraz części logitowej prezentują się następująco:

Rysunek 8: Zero inflated Poisson model

```
Pearson residuals:
      Min      1Q  Median      3Q      Max
-1.7212 -0.6864 -0.1444  0.4560  7.8359

Count model coefficients (poisson with log link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.920404   0.196560  4.683 2.83e-06 ***
age          0.003350   0.002188  1.531 0.125699
marital      0.787829   0.094217  8.362 < 2e-16 ***
educ        -0.032968   0.008840 -3.729 0.000192 ***
raceOther   -0.190957   0.094002 -2.031 0.042213 *
raceWhite   -0.332296   0.069523 -4.780 1.76e-06 ***
polviews    -0.061629   0.018436 -3.343 0.000829 ***
satfin      -0.066449   0.035774 -1.857 0.063242 .

Zero-inflation model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.0139     1.8646  0.544 0.586607
age         -0.3549     0.1034 -3.433 0.000597 ***
marital     -1.0186     0.5771 -1.765 0.077539 .
educ         0.4272     0.1371  3.116 0.001830 **
raceOther    0.4894     0.7852  0.623 0.533089
raceWhite    1.1204     0.6375  1.758 0.078829 .
polviews     0.4065     0.1704  2.386 0.017024 *
satfin       0.1675     0.3065  0.546 0.584726

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 29
Log-likelihood: -1513 on 16 Df
```

Źródło: Opracowanie własne

W celu porównania zwykłego modelu regresji Poissona z modelem z nadmierną liczbą zer wykorzystano test Vuong'a. Wynik testu wskazuje na podstawy do odrzucenia hipotezy zerowej i stwierdza się, że model ZIP (Zero inflated Poisson model) jest lepszy niż standardowy model regresji Poissona, co wskazuje poniższy rysunek (Rys.9):

Rysunek 9: Porównanie modeli - model ZIP i model 1

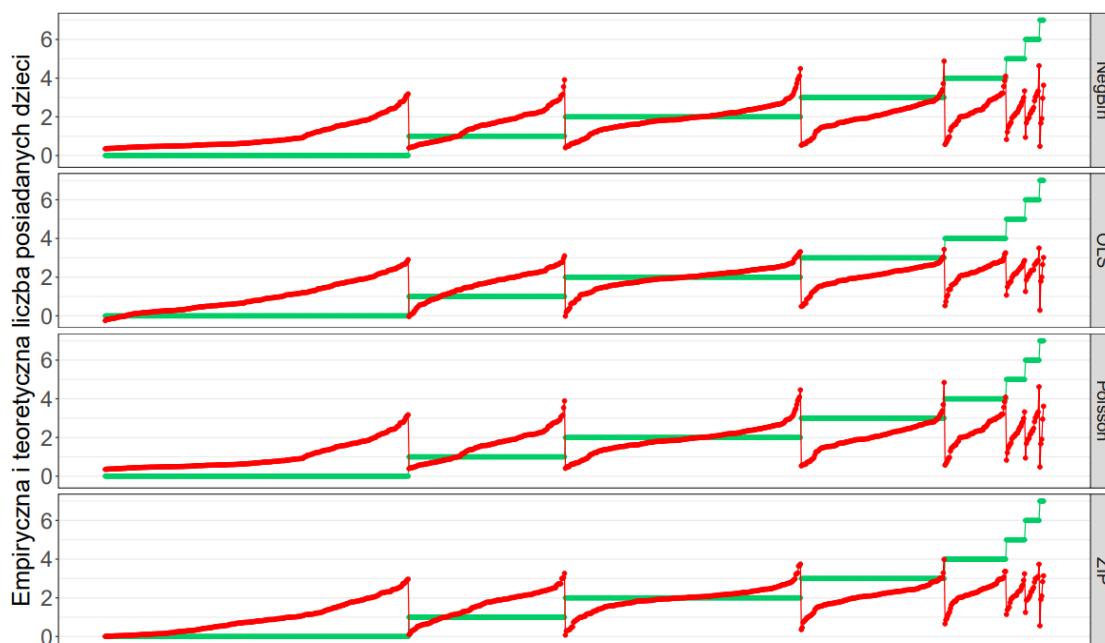
```
Vuong Non-Nested Hypothesis Test-Statistic:
(test-statistic is asymptotically distributed N(0,1) under the
null that the models are indistinguishable)
-----
      Vuong z-statistic      H_A      p-value
Raw          -5.539701 model2 > model1 1.5149e-08
AIC-corrected -4.740390 model2 > model1 1.0665e-06
BIC-corrected -2.763688 model2 > model1 0.0028576
```

Źródło: Opracowanie własne

Na podstawie powyższych porównań najlepszym modelem wydaje się być model ZIP. Lekka dyspersja średniej oraz wariancji nie okazała się mieć istotnego wpływu na oszacowania regresji Poissona w momencie porównania do modelu ujemnego dwumianowego. Model z nadmierną ilością zer okazał się być lepszym rozwiązaniem w przypadku modelowania posiadanych liczby dzieci niż podstawowy model Poissona.

Poniżej przedstawiono porównanie dopasowania wartości teoretycznych z czterech modeli: ujemny dwumianowy, klasyczny model regresji liniowej, regresja Poissona oraz model ZIP (Rys.10). Zielone punkty przedstawiają wartości empiryczne, natomiast czerwone – teoretyczne. Na podstawie wykresu można zauważyć, że modele NegBin (ujemny dwumianowy) oraz Poisson (model regresji Poissona) nie prognozują żadnych wartości zerowych, co potwierdza słuszność wyboru modelu ZIP, który prognozuje nieposiadanie żadnego dziecka. W przypadku modelu klasycznej regresji liniowej prognozy wyłapują zera, natomiast zdarza się również, że dają prognozy ujemne, które w rzeczywistości nie występują. Co więcej, można zaobserwować, że dla nieposiadania w ogóle dzieci oraz jednego dziecka model ma tendencje do przeszacowywania wartości, natomiast dla więcej niż dwójki dzieci model daje znacząco niedoszacowane prognozy liczby dzieci. Świadczyć to może o braku zmiennej definitywnie rozróżniającej braku i posiadaniu jakichkolwiek dzieci, a także o cesze determinującej dużą liczbę dzieci.

Rysunek 10: Porównanie wszystkich modeli - empiryczna i teoretyczna liczba dzieci



Źródło: Opracowanie własne

W tabeli poniżej przedstawiono również średnie błędy prognozy. Błąd średniokwadratowy (RMSE) oraz średni błąd (ME) są najmniejsze dla modelu ZIP, co potwierdza wszelkie powyższe rozważania na temat wyboru optymalnej specyfikacji modelu.

Tabela 10: Porównanie błędów prognozy

Model	RMSE	ME
NegBin	1.29	0.98
OLS	1.28	0.96
Poisson	1.29	0.98
ZIP	1.26	0.93

Źródło: Opracowanie własne

Biorąc pod uwagę wszystkie przeprowadzone testy oraz porównanie błędów prognozy ostatecznie wybrano model ZIP jako model finalny.

Interpretacja wyników modelowania

Niniejsza część raportu jest poświęcona interpretacji wyników modelu ZIP. Przeanalizowanie zostanie siła oraz kierunek oddziaływania poszczególnych determinant posiadanych liczby dzieci. Przy interpretacji oszacowań parametrów strukturalnych modelu należy zwrócić uwagę na znak przy nich stojący. Dodatni znak przy parametrze oznacza pozytywny efekt na liczbę posiadanych dzieci, natomiast przy znaku ujemnym zależność jest odwrotna.

Na podstawie przeprowadzonej analizy można stwierdzić, że czynnikami wpływającymi pozytywnie na liczbę posiadanych dzieci jest wiek oraz bycie kiedykolwiek w związku małżeńskim,. Jest to dosyć zgodne z intuicją. Natomiast zmiennymi wpływającymi na mniejszą oczekiwaną liczbę dzieci są lata edukacji, bardziej liberalne poglądy polityczne oraz satysfakcja finansowa. Efekt lat edukacji jest dosyć spodziewany, natomiast satysfakcja finansowa jest bardziej dyskusyjna. Dla osób, które są bardziej usatysfakcjonowane finansowo, zadowolenie to może wynikać właśnie z mniejszej liczby dzieci (mniejsze wydatki), natomiast osoby z większą liczbą dzieci mogą być mniej zadowolone z sytuacji finansowej właśnie z powodu dodatkowych wydatków na utrzymanie dzieci. W tabeli poniżej przedstawiono eksponenta oszacowań parametrów strukturalnych modelu.

Tabela 11: Eksponenta oszacowań parametrów modelu (Rate i Odds Ratio)

	<i>Intercept</i>	<i>Age</i>	<i>Marital</i>	<i>Educ</i>	<i>RaceOther</i>	<i>RaceWhite</i>	<i>Polviews</i>	<i>Satfin</i>
Poisson	2.51*	1.00	2.20*	0.97*	0.83*	0.72*	0.94*	0.94*
Logit	2.76	0.70*	0.36*	1.53*	1.63	3.07*	1.50*	1.18

Źródło: Opracowanie własne

Na podstawie tabeli statystycznie istotne parametry oszacowań można zinterpretować następująco:

- Dla osób, które mają jakąkolwiek szanse mieć dziecko, i kiedykolwiek były w związku małżeńskim, ceteris paribus, wartość oczekiwana liczby posiadanych dzieci wzrasta o 120%;
- Dla osób, które mają jakąkolwiek szanse mieć dziecko, biały lub inny niż biały i czarny kolor skóry, ceteris paribus, wartość oczekiwana liczby posiadanych dzieci spada o odpowiednio 28% i 17%;
- Dla osób, które mają jakąkolwiek szanse mieć dziecko, z każdą kolejną kategorią w stronę poglądów bardziej liberalnych na skali poglądów politycznych, ceteris paribus, wartość oczekiwana liczby posiadanych dzieci spada o 6%;
- Dla osób, które mają jakąkolwiek szanse mieć dziecko, z każdą kolejną kategorią w stronę większej satysfakcji finansowej na skali zadowolenia finansowego, ceteris paribus, wartość oczekiwana liczby posiadanych dzieci spada o 6%;
- Z każdym kolejnym rokiem życia, szansa na nieposiadanie żadnego dziecka maleje o 30%;
- Bycie kiedykolwiek w związku małżeńskim zmniejsza prawdopodobieństwo nieposiadania żadnego dziecka o 64%;

- Wzrost lat edukacji o rok, powoduje wzrost szansy na nieposiadanie żadnego dziecka o 53%;
- Bycie osobą o białej skórze, powoduje wzrost szansy na nieposiadanie żadnego dziecka o 207% w porównaniu do osoby o skórze czarnej;
- Każda kolejna kategoria w stronę poglądów bardziej liberalnych na skali poglądów politycznych, powoduje wzrost szansy na nieposiadanie żadnego dziecka o 50%;
- Każda kolejna kategoria w stronę większej satysfakcji finansowej na skali zadowolenia finansowego, powoduje wzrost szansy na nieposiadanie żadnego dziecka o 18%.

Wnioski i zakończenie

Badania empiryczne wykazały, że głównymi determinantami liczby posiadanych dzieci są wiek rodzica, ich status matrymonialny, poziom edukacji, kolor skóry, poglądy polityczne oraz poziom usatysfakcjonowania aktualną sytuacją finansową. W modelu jednak nie znalazły się zmienne dotyczące statusu zawodowego rodzica czy zarobków. Najskuteczniejszym modelem okazał się być model regresji Poissona z nadmierną liczbą zer - Zero Inflated Poisson model. Stało się tak, ponieważ zmienna objaśniana przyjmuje wartości od 0 do 7, przy czym wartości 0 stanowią 32% wszystkich odpowiedzi.

Najbardziej stymulująco na liczbę posiadanych dzieci okazały się działać wiek rodziców oraz ich status matrymonialny – bycie w małżeństwie znacząco podnosi szansę na posiadanie większej liczby dzieci. Natomiast wzrost edukacji czy satysfakcja z sytuacji finansowej rodziców działa demotywująco na posiadanie większej liczby dzieci. Kolor skór również okazał się mieć decydujące znaczenie – osoby z białym kolorem skóry posiadają znacznie mniej dzieci od innych ras. Może to być skorelowane z dostępem do edukacji i satysfakcją z sytuacji finansowej gospodarstwa domowego. Kolejnym interesującym faktem okazały się być poglądy polityczne. Osoby z konserwatywnymi poglądami mają szansę na posiadanie większej liczby dzieci od osób z poglądami centrowymi czy liberalnymi, a osoby z poglądami skrajnie liberalnymi posiadają średnio najmniej dzieci.

Podsumowując badanie czynników mających wpływ na liczbę dzieci okazał się być dość obszernym zagadnieniem. Zamodelowanie i zaprognozowanie takiego czynnika jest bardzo ryzykowane, ze względu na bardzo subiektywny charakter i zbyt indywidualne podejście jednostek do tematu. Wskazuje na to błąd RMSE, który wyniósł 1.26, co jest dużym odchyleniem przy zauważeniu, że 1.62 to średnia liczba posiadanych dzieci przez respondentów.

Bibliografia

B. Dibaba and G. Mitike, *"Factors Influencing Desired Family Size among Residents of Assela Town"*, Journal Women's Health Care, 2016

Martin Fieder and Susanne Huber, *"Political Attitude and Fertility: Is There a Selection for the Political Extreme?"*, frontiers in Psychology, Vienna, Austria, 2018

Khadijeh Asadi Sarvestani, Suet Leng Khoo, Nor Malina Malek, Suziana Mat Yasin, and Aliyar Ahmadi *"Factors Influencing the Desired Number of Children Among Married Women in the Reproductive Age and its Implications for Policy Making"*, Women Health Bull, 2017

Spis tabel i rysunków

Spis rysunków

Rysunek 1: Rozkład posiadanych dzieci.....	4
Rysunek 2: Liczba posiadanych dzieci w zależności od wieku respondenta	4
Rysunek 3: Liczba dzieci w zależności od statutu matrymonialnego respondenta	5
Rysunek 4: Liczba dzieci w zależności od koloru skóry rodzica	6
Rysunek 5: Liczba dzieci w zależności od poglądów politycznych rodzica	7
Rysunek 6: Model 1	10
Rysunek 7: Porównanie modeli – model 1 i ujemny model dwumianowy	10
Rysunek 8: Zero inflated Poisson model	11
Rysunek 9: Porównanie modeli - model ZIP i model 1	11
Rysunek 10: Porównanie wszystkich modeli - empiryczna i teoretyczna liczba dzieci	12

Spis tabel

Tabela 1: Opis zmiennych.....	3
Tabela 2: Zmienna satfin	5
Tabela 3: Zmienna martial.....	6
Tabela 4: Zmienna race	6
Tabela 5: Zmienna income	7
Tabela 6: Zmienna polviews	8
Tabela 7: Zmienna wrkstat	8
Tabela 8: Zmienna degree	8
Tabela 9: Zmienna educ	8
Tabela 10: Porównanie błędów prognozy.....	13
Tabela 11: Eksponenta oszacowań parametrów modelu (Rate i Odds Ratio)	13