

D E B I U T Y   S T U D E N C K I E

2 0 2 3

---

# **ZASTOSOWANIE METOD ILOŚCIOWYCH W EKONOMII I FINANSACH**

pod redakcją  
**Alicji Grześkowiak**  
**i Piotra Peterneka**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2023

Recenzja

*Katarzyna Ostasiewicz*

Redakcja wydawnicza

*Elżbieta Żurawska-Łuczyńska*

Korekta

*Katarzyna Gwizda*

Skład i łamanie

*Beata Mazur*

Projekt okładki

*Beata Dębska*

Na okładce wykorzystano zdjęcia z zasobów 123 Royalty Free

Praca opublikowana na licencji Creative Commons Uznanie autorstwa

Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0).

Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>



ISBN 978-83-67899-08-6 (wersja papierowa)

ISBN 978-83-67899-09-3 (wersja elektroniczna)

DOI: 10.15611/2023.09.3

Druk i oprawa: TOTEM

**Oktawia Golczyk**

e-mail: 181511@student.ue.wroc.pl

**Martyna Kamińska**

e-mail: 181660@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

## **Analiza dzietności i jej uwarunkowań w Polsce w 2021 roku**

DOI: 10.15611/2023.09.3.03

JEL Classification: C38, C52, J13

**Streszczenie:** Od kilku lat kraje europejskie zmagają się z dużym problemem demograficznym. Pomimo różnych programów prorodzinnych współczynnik dzietności nie wzrasta. Sytuacja ta dotyczy również Polski. Artykuł ma na celu zbadanie uwarunkowań współczynnika dzietności oraz jego kształtowania się w województwach w Polsce w 2021 roku. W analizie wykorzystano modelowanie ekonometryczne oraz analizę skupień. Okazało się, że zjawisko dzietności jest bardzo złożone i trudne do jednoznaczniego scharakteryzowania. Wybrane do analizy czynniki nie pozwoliły na stworzenie modelu, który opisywałby to zjawisko w wystarczającym stopniu. Ponadto w Polsce nadal istnieją różnice regionalne, które sprawiają, że w części województw jest dużo wyższy współczynnik dzietności niż w pozostałych. Analiza skupień zidentyfikowała, które województwa są ze sobą powiązane ze względu na badane czynniki oraz jakie klastry powstały w wyniku tego procesu.

**Słowa kluczowe:** dzietność, demografia, współczynnik dzietności, model ekonometryczny, analiza skupień

### **1. Wstęp**

W artykule podjęto temat dzietności w Polsce w 2021 roku. Problemem badawczym jest ustalenie wpływu wybranych czynników ekonomicznych i społeczno-kulturowych na analizowane zagadnienie oraz ukazanie podobieństw pomiędzy województwami ze względu na wybrane zmienne. Ponadto w pracy zostały wyznaczone skupienia metodą hierarchicznej analizy skupień. Określone pytania badawcze to:

- Jakie czynniki najbardziej kształtują współczynnik dzietności w Polsce?
- Które województwa tworzą skupienia ze względu na uwarunkowania współczynnika dzietności?
- Jakie różnice występują pomiędzy skupieniami?

Praca traktuje o współczynniku dzietności w Polsce, ponieważ jest to bardzo istotny temat. Od kilku lat kraje europejskie zmagają się z niską dzietnością. Pomimo powstawania nowych programów wspierania rodzin przez państwa współczynnik dzietności nie wzrasta, a populacja europejska stale się kurczy. Do tych krajów należy również Polska. Należy się zastanowić, w czym może leżeć przyczyna stagnacji danego współczynnika oraz które zjawiska mogą na to wpływać. W najnowszych

badaniach przeprowadzonych w drugiej połowie 2022 roku przez Centrum Badania Opinii Społecznej (CBOS) wynika, że 32% Polek w wieku 18-45 lat planuje potomstwo w bliższej lub dalszej perspektywie, a aż 68% nie ma planów prokreacyjnych. W porównaniu z 2017 rokiem, odsetek kobiet, które zamierzają mieć dzieci, zmniejszył się o 9 punktów procentowych. Potwierdza to fakt, że niska dzietność stale jest istotnym problemem.

## 2. Uwarunkowania dzietności

Jak stwierdza Kaźmierczak-Kałużna (2020), od ponad półwiecza w Polsce występuje regularny spadek współczynnika dzietności. Nabral on szczególnego tempa w latach dziewięćdziesiątych minionego wieku. Autorka zwraca uwagę na fakt, iż od trzech dekad współczynnik dzietności w naszym kraju osiąga wartość poniżej poziomu prostej zastępowalności pokoleń. Natomiast od dwóch dekad gwałtowny spadek spowodował osiągnięcie poziomu definiowanego jako skrajnie niski (poniżej 1,5). Główny Urząd Statystyczny (GUS) w 2014 roku w *Prognozie ludności na lata 2014-2050* wskazał, że rekord niskiej liczby żywych urodzeń padł w 2003 roku, kiedy na świat przyszło zaledwie 350 tys. dzieci, czyli blisko o 200 tys. mniej niż w roku 1990. Wskaźnik dzietności w 2003 roku wyniósł zaledwie 1,22. GUS zwraca również uwagę na fakt, iż Polska znajduje się na takim etapie rozwoju demograficznego, na którym nawet wzrost dzietności do poziomu zapewniającego prostą zastępowalność pokoleń w krótkim okresie nie powstrzyma zmniejszania się populacji w kraju. „Proces odbudowy demograficznej jest procesem powolnym i wymaga konsekwentnych, długofalowych działań” (GUS, 2014, s. 109).

Przyczyn niskiej dzietności w Polsce Sikorska (2021) doszukuje się w kilku ważnych aspektach społeczno-ekonomicznych. Po pierwsze, odkładanie decyzji o pierwszym dziecku spowodowane jest brakiem poczucia stabilizacji sytuacji mieszkaniowej oraz zatrudnienia obojga partnerów na rynku pracy. Poczucie niepewności wielu młodych par spowodowane różnego rodzaju „kosztami” macierzyństwa również jest przyczyną niskiej liczby urodzeń. Matki znajdują się w gorszym położeniu od ojców. Mowa tu o sytuacji na rynku pracy oraz o wciąż aktualnym wzorze „poświęcającej się” matki, z którym związana jest presja społeczna. Dodatkowo kobiety narażone są na „potrójne obciążenie” obejmujące pracę zawodową, obowiązki domowe i opiekę nad dziećmi. Autorka wskazuje, iż politykę rodzinną w Polsce niełatwo uznać za spójną, ponieważ zmieszane są w niej zarówno działania mające na celu ułatwienie łączenia pracy zawodowej z rodzicielstwem (np. ustawa żłobkowa), jak i instrumenty, które pogłębiają różnice między matkami i ojcami, w tym podział ich codziennych obowiązków i opieki nad dziećmi, np. urlop rodzicielski bez przeznaczonej części, z której może skorzystać tylko ojciec. Z kolei Kaźmierczak-Kałużna (2020) zwraca uwagę na fakt, że wprowadzony w 2016 roku program pronatalistyczny „500+” również nie przynosi oczekiwanego głównego rezultatu, jakim jest wzrost dzietności w Polsce. Co prawda w dwóch pierwszych latach działania programu

wskaźnik dzietności nieco wzrósł, przyjmując w 2017 roku wartość równą 1,453 (największą od dwudziestu lat), ale odnotowany wzrost dotyczył w szczególności dzieci urodzonych jako drugie i trzecie. W kolejnych latach faktyczne liczby urodzeń były niższe od jej zakładanych wartości, np. liczba urodzeń na 2020 rok szacowana była na 380 tys., a w rzeczywistości wyniosła 360 tys., o 20 tys. mniej niż kalkulowano (Sikorska, 2021). Bakiera (2022) jako przyczynę niskiego wskaźnika dzietności w Polsce wskazuje także coraz częstszy problem społeczny, jakim jest niepłodność. Z brakiem zdolności do posiadania dzieci zмага się aż 20-25% par, co Brachowicz (2009) określa jako epidemię niepłodności.

„To, co wydarzy się w sytuacji demograficznej Polski w przyszłości, wydarzyło się już teraz” (Sikorska, 2021, s. 10). Interpretując słowa wypowiedzi byłej wiceprezeski Polskiego Towarzystwa Demograficznego, można zdefiniować pesymistyczną tezę, że nawet efektywna polityka publiczna co najwyżej zminimalizuje zjawisko spadku liczby ludności.

### 3. Metodyka badań

Jak wskazuje Główny Urząd Statystyczny, współczynnik dzietności oznacza liczbę dzieci, które urodziłaby przeciętnie kobieta w ciągu całego okresu rozrodczego, czyli w wieku 15-49 lat, przy założeniu, że w poszczególnych fazach tego okresu rodziłaby z intensywnością obserwowaną w badanym roku, tzn. przy przyjęciu częściowych współczynników płodności z tego okresu za niezmiennie. Na potrzeby budowy modelu ekonometrycznego za zmienną objaśnianą przyjęto wcześniej opisany współczynnik dzietności ( $y$ ), a za zmienne objaśniające:  $x_1$  – stopa bezrobocia [%],  $x_2$  – liczba lekarzy w przeliczeniu na 1000 mieszkańców,  $x_3$  – oddziały żłobków w przeliczeniu na 1000 mieszkańców oraz  $x_4$  – liczba rozwodów w przeliczeniu na 1000 mieszkańców.

Niniejsza analiza została oparta na danych pobranych z GUS-u z Banku Danych Lokalnych. Wszystkie dane pochodzą z 2021 roku i są kompletne. Do badania zostały wybrane dane przekrojowe, które są rzetelne i wiarygodne. Jako obserwacje występują województwa Polski.

Jedną z najstarszych oraz najpopularniejszych metod modelowania zależności między zmienną objaśnianą a zbiorem zmiennych objaśniających jest model ekonometryczny. Jego zaletą jest m.in. łatwość w wyznaczaniu niezbędnych statystyk do estymacji i testowania (Biecek, 2011). Model to uproszczone odwzorowanie rzeczywistości oraz formalna konstrukcja, która dzięki równaniu lub układowi równań przedstawia powiązania między badanymi zjawiskami ekonomicznymi (Pawłowski, 1981). Na zmienną zależną wpływają różne czynniki, mniej lub bardziej ze sobą skorelowane. Uwzględniane w nim są zaś tylko czynniki główne, które mają istotny wpływ na opisywane zjawisko. Czynniki przypadkowe oraz słabo oddziałujące są pomijane w modelu. Jednorównaniowy model ekonometryczny można zapisać w następujący sposób:

$$Y = f(X_1, \dots, X_K; \varepsilon),$$

gdzie:  $Y$  – zmienna objaśniana, reprezentuje zjawisko modelowane;  $X_1, \dots, X_K$  – zmienne objaśniające;  $\varepsilon$  – błąd losowy;  $f$  – postać analityczna modelu;  $K$  – liczba zmiennych objaśniających (Kukuła, Goryl, Jędrzejczyk, Osiewalski i Walkosz, 2009).

Analiza skupień, nazywana też taksonomią, to zbiór metod wielowymiarowej analizy statystycznej. Odgrywa bardzo ważną rolę w różnych dziedzinach nauki, zwłaszcza w analizach zjawisk społeczno-ekonomicznych. Służy wyodrębnieniu obiektów zbioru danych w rozłączne, spójne wewnętrznie oraz zróżnicowane między sobą grupy – skupienia (Korzeniowski, 2012). Jej zasadniczymi typami algorytmów grupowania danych są algorytmy hierarchiczne i niehierarchiczne. Te pierwsze polegają na tworzeniu porządku klasyfikacji obiektów, poprzez stopniowe łączenie elementów w podgrupy. Na wstępie każdy obiekt to odrębne skupienie, lecz z każdym kolejnym krokiem zaczynają się one ze sobą łączyć, w wyniku czego na końcu powstaje jedno skupienie, zawierające wszystkie objekty. Do mierzenia odległości między skupieniami stosowane są zróżnicowane metody, m.in. Warda i najdalszego sąsiedztwa (Harańczyk, 2005). Metoda Warda wykorzystuje podejście analizy wariancji oraz ma na celu minimalizację sumy kwadratów odchyłeń wewnątrz skupień. W metodzie najdalszego sąsiedztwa odległość między skupieniami ustalana jest na podstawie największej odległości między dwoma obiektami, które należą do różnych skupień (Grajewski, 2006). Metody niehierarchiczne w przeciwieństwie do metod hierarchicznych dążą do uzyskania rozbicia, w którym każde skupienie jest niezależne od pozostałych. Przy tej metodzie znana jest *a priori* liczba skupień, której wybór ma duży wpływ na jakość uzyskanej segmentacji (Harańczyk, 2005).

## 4. Wyniki badań

### 4.1. Statystyki opisowe

W tym podpunkcie zostaną przedstawione statystyki opisowe dla badanych zmiennych. Są one zaprezentowane w tabeli 1.

Tabela 1. Statystyki opisowe

| Zmienna    | Średnia | Mediana | Kwartył dolny | Kwartył górny | Odchylenie standardowe | Współczynnik zmienności | Minimum | Maksimum |
|------------|---------|---------|---------------|---------------|------------------------|-------------------------|---------|----------|
| Dzietność  | 1,312   | 1,302   | 1,257         | 1,353         | 0,066                  | 0,050                   | 1,226   | 1,432    |
| Bezrobocie | 6,531   | 6,250   | 4,975         | 8,250         | 2,008                  | 0,307                   | 3,200   | 9,900    |
| Lekarze    | 3,898   | 4,050   | 3,343         | 4,395         | 0,770                  | 0,198                   | 2,710   | 5,070    |
| Żłobki     | 11,214  | 9,941   | 6,799         | 15,496        | 5,688                  | 0,507                   | 3,718   | 21,607   |
| Rozwody    | 1,563   | 1,565   | 1,395         | 1,705         | 0,205                  | 0,131                   | 1,220   | 1,920    |

Źródło: opracowanie własne na podstawie danych GUS.

W pierwszym kroku zbadano współczynnik zmienności dla zmiennych objaśniających w celu zweryfikowania, czy mogą one brać udział w tworzeniu modelu. Współczynnik przekroczył dla wszystkich cech wartość krytyczną  $V^* = 0,1$ . Największą zmiennością charakteryzuje się zmienna rozwody, co oznacza, że Polska jest pod tym względem mocno zróżnicowana. Dla badanych cech nie występują obserwacje odstające.

Przeciętny poziom współczynnika diety to 1,312. Najmniejsza dieta występuje w województwie opolskim. Może to wynikać z faktu, iż jest ono bardzo podatne na zmiany liczby ludności z powodu najmniejszej liczby mieszkańców zamieszkujących to województwo. Ponadto boryka się z bardzo dużą migracją, szczególnie młodych osób. Jest to jeden z najszybciej wyludniających się regionów w Europie. Największa dieta występuje w województwie pomorskim i wynosi 1,432. Zaraz za nim jest województwo wielkopolskie. Prawdopodobnie duży wpływ ma na to imigracja młodych osób do tych województw. Są to województwa wysoko rozwinięte, z bogatą ofertą pracy, dlatego też przyciągają młode osoby chcące polepszyć jakość swojego życia.

W 25% województw z najwyższą stopą bezrobocia jest ona co najmniej równa 8,25%. Do tych województw należą głównie województwa ze wschodniej ściany Polski oraz zachodniopomorskie i kujawsko-pomorskie. Nasuwa się myśl, że Polska wschodnia jest mniej rozwinięta gospodarczo od reszty Polski. Ponadto w niektórych z tych województw zatrudnienie opiera się w dużej mierze na turystyce, dlatego brakuje tam zakładów przemysłowych. Również pandemia mogła wpłynąć na podwyższenie bezrobocia w województwach turystycznych.

Zmienna lekarze, czyli średnia liczba lekarzy w przeliczeniu na 1000 mieszkańców, przeciętnie wynosi 3,898. Na tle innych państw Unii Europejskiej Polska wypada niekorzystnie. Najmniej lekarzy przeliczonych na 1000 mieszkańców jest w województwach: lubuskim, opolskim, podkarpackim i warmińsko-mazurskim. W 25% województw, które mają największą liczbę lekarzy, liczba ta jest równa co najmniej 4,395 lekarzy w przeliczeniu na 1000 osób, co również nie jest dobrym wynikiem. Może mieć to negatywny wpływ na współczynnik diety w obawie przed niewystarczającą opieką zdrowotną dzieci.

Mediana dla liczby żłobków przeliczonych na 1000 mieszkańców wynosi 9,941. Odchylenie standardowe równe jest 5,688, co oznacza, że zmienna ta przeciętnie odchyła się od średniej o tę liczbę. Najmniejsza liczba żłobków jest w województwie mazowieckim. Znaczący wpływ ma na to Warszawa, największe miasto w Polsce. Z ciągłym brakiem miejsc w żłobkach borykają się nie tylko mieszkańcy stolicy, lecz także i innych dużych miast Polski. Wpływa to znacząco na podjęcie decyzji o posiadaniu potomka, ponieważ przez niewystarczającą liczbę miejsc w żłobkach dziecko zazwyczaj musi zostać pod opieką jednego z rodziców, co często osłabia sytuację finansową rodziny.

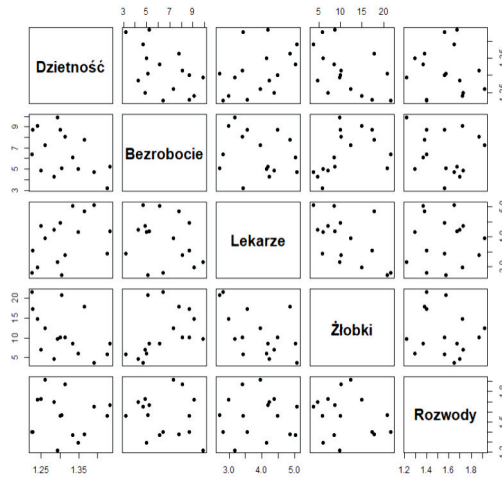
Średnia liczba rozwodów w Polsce przeliczona na 1000 mieszkańców wynosi 1,563. Najmniejsza liczba rozwodów jest w województwie podkarpackim – 1,22

oraz małopolskim – 1,29. Największa natomiast w województwie zachodniopomorskim – 1,92. Najwięcej rozwodów występuje również w województwach: kujawsko-pomorskim, warmińsko-mazurskim, śląskim i dolnośląskim. Przeciętna liczba rozwodów odchyła się od średniej o 0,2.

## 4.2. Tworzenie modelu liniowego

W tym podrozdziale zostanie podjęta próba stworzenia modelu liniowego w celu zbadania, jakie zmienne mogą mieć wpływ na kształtowanie się współczynnika dzietności w Polsce.

Pierwszym krokiem jest zbadanie korelacji pomiędzy badanymi zmiennymi oraz zmienną objaśniającą. Na rysunku 1 przedstawiony został wykres rozrzutu zmiennych, a na rysunku 2 macierz korelacji pomiędzy zmiennymi.

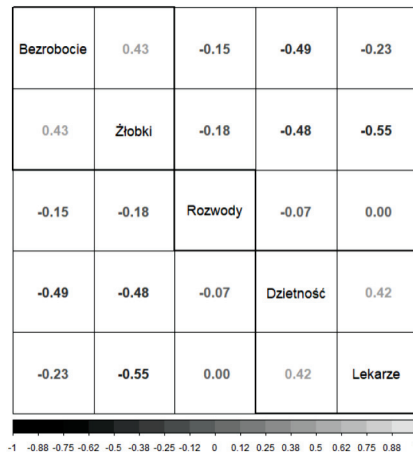


**Rysunek 1.** Wykres rozrzutu zmiennych

Źródło: opracowanie własne na podstawie danych GUS.

Na rysunku 1 nie można zobaczyć zależności liniowej pomiędzy zmiennymi, ponieważ występuje duży rozrzut punktów. Z rysunku 2 również wynika, iż żadna ze zmiennych nie jest ze sobą skorelowana w istotnym stopniu. Wszystkie korelacje są mniejsze niż 0,2. Zmienna bezrobocie jest ujemnie skorelowana ze wszystkimi zmiennymi, poza żłobkami. Oznacza to, że gdy stopa bezrobocia w Polsce jest wyższa, to wartości pozostałych badanych cech maleją. Wszystkie pozostałe zmienne również są ze sobą ujemnie skorelowane. Wyjątek stanowi korelacja dzietności z lekarzami oraz żłobków z bezrobociem. Najbardziej skorelowanymi zmiennymi są żłobki i dzietność. Najmniej skorelowana z pozostałymi cechami jest zmienna rozwody. Ze zmienną lekarze ich korelacja wynosi 0. Oznacza to, że nie jest ona istotnie skorelowana z żadną zmienną.





Rysunek 2. Macierz korelacji

Źródło: opracowanie własne na podstawie danych GUS.

W kolejnym kroku analizy zostanie zbadana korelacja badanych zmiennych ze zmienną objaśnianą, czyli dzietnością. Korelacja pomiędzy dzietnością ( $y$ ) a bezrobociem wynosi  $-0,49$ . Skorelowanie pomiędzy zmienną objaśnianą a lekarzami przyjmuje wartość  $0,42$ . Zależność pomiędzy dzietnością a żłobkami jest równa  $-0,48$ . Ostatnia badana korelacja jest między dzietnością a rozwodami i wynosi  $-0,07$ . Wszystkie zmienne objaśniające, poza zmienną lekarze, są ujemnie skorelowane ze zmienną objaśnianą. Wskazuje to na odwrotną zależność pomiędzy nimi, zatem wzrostowi wartości wybranej zmiennej towarzyszy spadek dzietności.

Następnym krokiem jest zbudowanie modelu ekonometrycznego. Model został zbudowany ze zmiennej objaśnianej  $y$  oraz wszystkich zmiennych objaśniających, czyli bezrobocia, lekarzy, żłobków i rozwodów. Otrzymano więc postać:

$$\hat{y} = 1,45 - 0,01x_1 + 0,02x_2 - 0,003x_3 - 0,06x_4$$

Badanie można traktować jako badanie całościowe, zatem pominięto testowanie istotności parametrów oraz korelacji.

Model wskazuje na zależność, że jeśli stopa bezrobocia wzrośnie o 1%, przy stałych wartościach innych zmiennych objaśniających, to współczynnik dzietności spadnie o 0,01. Pokazuje on także, że jeśli liczba lekarzy w przeliczeniu na 1000 mieszkańców wzrośnie o 1, to współczynnikowi dzietności będzie towarzyszył wzrost o 0,02, jeśli wartości innych cech nie zmienią się. Na podstawie modelu można także stwierdzić, że jeśli liczba oddziałów żłobków w przeliczeniu na 1000 mieszkańców wzrośnie o 1, to współczynnik dzietności spadnie o 0,003, przy niezmienności innych zmiennych. Następną zależnością, jaką wskazuje model, jest to, że jeśli liczba rozwodów w przeliczeniu na 1000 mieszkańców wzrośnie o 1, to współczyn-

nik dietności zmaleje o 0,06, przy stałości innych cech. Wskazuje on także zależność, że jeśli wszystkie badane objaśniające wynosiłyby 0, to współczynnik dietności przyjmowałby wartość 1,45. Współczynnik determinacji  $R^2$  wynosi 0,39, co oznacza, że tylko 39% zmiennej  $y$  jest wyjaśnione przez model. W związku z tym jego dopasowanie do zmiennej objaśnianej jest bardzo słabe i nie powinno się nim sugerować przy określaniu współczynnika dietności.

W tabeli 2 przedstawiono macierz wariancji i kowariancji.

**Tabela 2.** Macierz wariancji i kowariancji

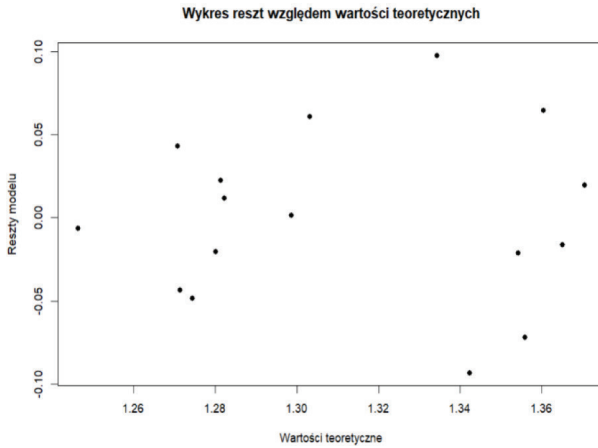
|             | Wyraz wolny | Bezrobocie | Lekarze  | Żłobki   | Rozwody  |
|-------------|-------------|------------|----------|----------|----------|
| Wyraz wolny | 0,03703     | -0,00045   | -0,00320 | -0,00033 | -0,01133 |
| Bezrobocie  | -0,00045    | 0,00007    | 0,00000  | -0,00001 | 0,00006  |
| Lekarze     | -0,00320    | 0,00000    | 0,00059  | 0,00005  | 0,00024  |
| Żłobki      | -0,01133    | 0,00006    | 0,00024  | 0,00005  | 0,00607  |
| Rozwody     | -0,00032    | -0,00001   | 0,00005  | 0,00001  | 0,00005  |

Źródło: opracowanie własne na podstawie danych GUS.

Na podstawie reszt modelu, czyli różnic pomiędzy wartościami empirycznymi i teoretycznymi, można wyznaczyć błędy szacunku parametrów modelu. Obliczając pierwiastek z wartości, które znajdują na przekątnej macierzy wariancji i kowariancji (tab. 2), otrzymano błędy  $S(a_j)$ . Błąd szacunku dla zmiennej  $x_1$  wynosi  $S(a_1) = 0,008$ , co oznacza, że wartość oceny  $a_1$  różni się przeciętnie od jej rzeczywistej wartości o 0,008. Dla zmiennej  $x_2$  błąd szacunku  $S(a_2)$  jest równy 0,024, co oznacza, że wartość oceny  $a_2$  różni się średnio od jej rzeczywistej wartości o 0,024. Dla zmiennych  $x_3$  oraz  $x_4$  błąd szacunku jest taki sam i wynosi 0,007, a więc ich wartości oceny  $a_3$  i  $a_4$  różnią się przeciętnie od ich rzeczywistych wartości o 0,007. Dla wyrazu wolnego błąd ten jest równy  $S(a_0) = 0,192$ , a więc oszacowana wartość wyrazu wolnego odchyła się od jego empirycznej wartości średnio o 0,192. Standardowy błąd modelu  $S(e)$  wynosi 0,059. Informuje on, iż wartości rzeczywiste różnią się przeciętnie od wartości teoretycznych o 0,059.

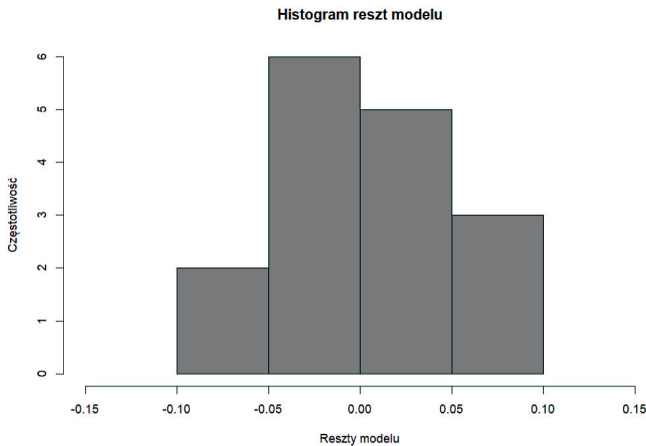
Wykres rozrzutu reszt względem wartości teoretycznych został przedstawiony na rysunku 3. Na wykresie nie można zauważyć żadnych wzorców, co wskazuje na homoskedastyczność. Punkty nie są ułożone w wykres funkcji, co świadczy o liniowości. Wszystko to świadczy na korzyść modelu.

Histogram reszt modelu przedstawiony na rysunku 4 swoim kształtem przypomina rozkład normalny. Można zatem wnioskować, że dla reszt występuje właśnie taki rozkład. Wniosek ten został sprawdzony testem Shapiro-Wilka. Hipoteza zerowa dla badanego testu zakłada, że reszty mają rozkład normalny, natomiast hipoteza alternatywna o braku takiego rozkładu.



**Rysunek 3.** Wykres reszt modelu względem wartości teoretycznych

Źródło: opracowanie własne na podstawie danych GUS.



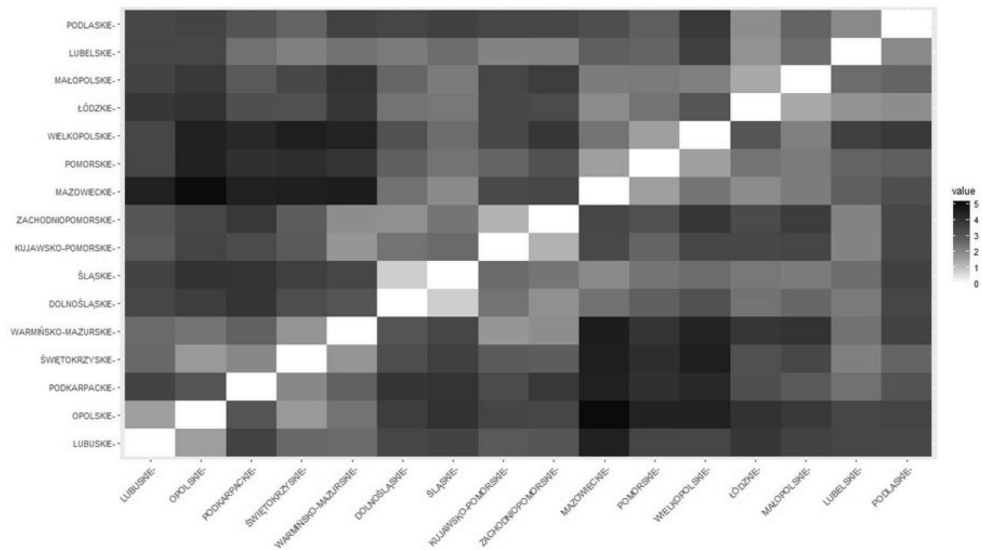
**Rysunek 4.** Histogram reszt modelu

Źródło: opracowanie własne na podstawie danych GUS.

Po wykonaniu testu p-value było równe 0,999, czyli bliskie 1. Jest ono większe niż 0,05, dlatego nie ma podstaw do odrzucenia hipotezy zerowej. Wobec tego reszty modelu mają rozkład normalny, a wynik testu potwierdził przypuszczenia oparte na podstawie histogramu.

### 4.3. Analiza skupień

Do odpowiedniego przeprowadzenia analizy skupień wykorzystano zmienne, takie jak w modelu ekonometrycznym. Aby były one porównywalne, zastosowano na początku ich standaryzację. Następnie do podziału obserwacji na grupy wyznaczono macierz odległości za pomocą miary odległości euklidesowej (rys. 5).



**Rysunek 5.** Wizualizacja macierzy odległości

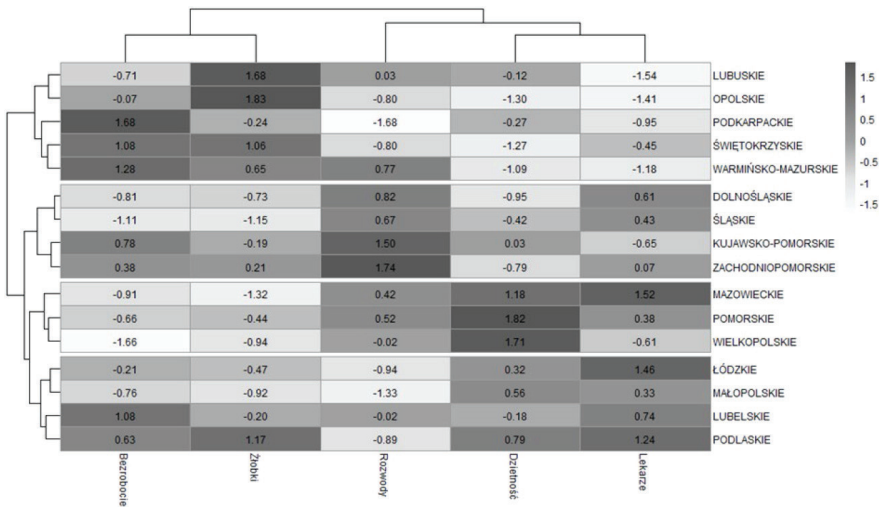
Źródło: opracowanie własne na podstawie danych GUS.

Pary obserwacji o najmniejszym podobieństwie (największej odległości) oznaczone są kolorem czarnym, a pary o największym podobieństwie (najmniejszej odległości) białym. Z rysunku 5 wynika, iż największe podobieństwo występuje np. między województwami: kujawsko-pomorskim i zachodniopomorskim, śląskim i dolnośląskim oraz małopolskim i łódzkim. Najmniejsze zaś podobieństwo zachodzi m.in. między województwem mazowieckim i świętokrzyskim, mazowieckim i opolskim oraz mazowieckim i warmińsko-mazurskim.

Przeprowadzając analizę skupień metodą Warda, wygenerowany został wykres ilustrujący hierarchiczną strukturę zbioru obiektów ze względu na zmniejszające się podobieństwo między nimi oraz standaryzowane wartości cech w obiektach (rys. 6). Województwa podzielone zostały na cztery skupienia poprzez proces kolejnego łączenia się poszczególnych regionów grupy. Zawierają one odpowiednio pięć, cztery, trzy i cztery województwa.

Do pierwszej grupy przynależą województwa: lubuskie, opolskie, podkarpackie, świętokrzyskie i warmińsko-mazurskie. Początkowo najbardziej zbliżone do siebie były województwa lubuskie i opolskie oraz warmińsko-mazurskie i świętokrzyskie, następnie zostały one połączone kolejnym wiązaniem. Analogicznie pogrupowały się pozostałe województwa przedstawione na wykresie (rys. 6). Nie są widoczne jednoznaczne różnice między skupieniami, ale w pierwszym i trzecim występuje większe zróżnicowanie pomiędzy wynikami. W pierwszym skupieniu dla większości zmiennych można zaobserwować tylko wartości ujemne. Również dla województw z tej grupy, pomimo zróżnicowanych między nimi odległości, ich wartości nie są

dotadnie. W trzecim skupieniu otrzymane wyniki dla badanych zmiennych wskazują odwrotną sytuację względem pierwszego skupienia. W grupie pierwszej zmienna diety przyjmuje wartości jedynie ujemne, a stopa bezrobocia i żłobki w większości dodatnie. Natomiast w trzecim skupieniu to zmienna diety przyjmuje wartości tylko dodatnie, a stopa bezrobocia i żłobki ujemne. Do drugiej grupy zaliczone zostały województwa: dolnośląskie, śląskie, kujawsko-pomorskie i zachodniopomorskie. Wyłącznie dla zmiennej rozwody wartości w tej grupie są dodatnie dla każdego województwa. W ostatnim skupieniu tylko dla zmiennej lekarze występują wartości dodatnie dla każdego województwa z danej grupy.



Rysunek 6. Podział na skupienia oraz standaryzowane wartości cech w obiektach

Źródło: opracowanie własne na podstawie danych GUS.

Największą dodatnią wartość ze wszystkich skupień można zaobserwować dla zmiennej żłobki i województwa opolskiego w pierwszym skupieniu. Także w tej grupie odnotowano największą ujemną wartość, która występuje dla zmiennej rozwody oraz województwa podkarpackiego.

Tabela 3. Wartości środków ciężkości dla zmiennych w skupieniach

| Nr skupienia \ Zmienna | Bezrobocie | Żłobki | Rozwody | Dieta | Lekarze |
|------------------------|------------|--------|---------|-------|---------|
| 1                      | 7,84       | 16,88  | 1,462   | 1,258 | 3,046   |
| 2                      | 6,15       | 8,584  | 1,805   | 1,277 | 3,988   |
| 3                      | 4,367      | 6,088  | 1,627   | 1,416 | 4,23    |
| 4                      | 6,9        | 10,608 | 1,4     | 1,337 | 4,623   |

Źródło: opracowanie własne na podstawie danych GUS.

Dla każdego skupienia i każdej zmiennej wyznaczono wartości środków ciężkości (tab. 3). Cechy bezrobocie oraz żłobki przyjmują największą średnią wartość dla skupienia pierwszego. Z kolei w tym skupieniu odnotowuje się najmniejsze średnie wartości dla zmiennych dietność i lekarze. Dla cechy rozwody największa wartość środka ciężkości występuje w grupie drugiej, a dla zmiennej dietność w skupieniu trzecim. Skupienie czwarte odnotowuje największą średnią wartość dla cechy lekarze.

## 5. Zakończenie

Po przeprowadzeniu analizy współczynnika dietności w Polsce w 2021 roku okazuje się, że jest to problem bardzo złożony i trudny do jednoznacznego scharakteryzowania. Największą korelację ze zmienną  $y$  ma zmienna  $x_1$ , czyli stopa bezrobocia, jednak korelacja na poziomie  $-0,49$  nie jest wysoka. Wybrane czynniki nie wpływają znacząco na badany problem, ponieważ współczynnik  $R^2$  wynosi zaledwie  $0,39$ , co oznacza, że współczynnik dietności jest wyjaśniany w  $39\%$  przez opracowany model. Z tego powodu wykonana na jego podstawie prognoza byłaby niepoprawna. Okazuje się, że trudno jest określić, które czynniki towarzyszą w istotnym stopniu współczynnikowi dietności. Pomimo wzięcia zmiennych z różnych obszarów, które wydawałoby się mają wpływ na badany współczynnik, nie opisują one podjętego tematu w sposób poprawny. Z tego powodu problematyczne jest wskazanie, jakie zmiany skutkowałyby wzrostem współczynnika dietności.

Obecnie w Europie odnotowuje się mniejszą skłonność do zawierania małżeństw, większą liczbę rozwodów i późniejszy wiek urodzenia pierwszego dziecka. W Polsce brakuje odpowiedniej opieki medycznej w czasie ciąży oraz wsparcia dla przyszłych matek. Należy też wspomnieć, że coraz więcej par ma problemy z poczęciem dziecka. Na decyzję o posiadaniu dziecka miała wpływ również pandemia, co zauważalne jest w liczbie urodzeń po jej rozpoczęciu. Zmieniły się też priorytety życiowe młodych ludzi – osiągnięcie określonego poziomu wykształcenia i stabilizacji ekonomicznej. Obecna sytuacja gospodarcza nie pomaga przy podjęciu decyzji o posiadaniu dziecka. Stale rosnąca inflacja, zła sytuacja na rynku mieszkaniowym, wysokie ceny mieszkań oraz wzrost stóp procentowych mają realny wpływ na decydowanie się na potomstwo.

Analiza skupień i wykorzystane metody pokazały, które województwa są ze sobą połączone ze względu na badane zmienne. W metodzie Warda w jednym ze skupień znalazły się województwa: wielkopolskie, pomorskie oraz mazowieckie. Wyróżniają się one dużą liczbą lekarzy przeliczonych na 1000 mieszkańców oraz wysoką wartością współczynnika dietności. Ma to wpływ na atrakcyjność tych województw, przez co więcej ludzi przyjeżdża do nich, np. w celu znalezienia pracy lub polepszenia poziomu swojego życia. Świadczy to o tym, że są one dobrym miejscem dla przyszłych rodziców. Wyłoniła się także grupa zawierająca województwa, m.in. podlaskie i lubelskie, znajdujące się na wschodzie Polski. Poziom urbanizacji jest w nich niski, a dotacje z Unii Europejskiej przeznaczane są w większości dla sezon-

wego sektora turystyki. Sytuacja gospodarcza tych województw wpływa na to, iż są one postrzegane jako mniej atrakcyjne miejsca do życia dla nowych pokoleń. W innej grupie znalazły się województwa, m.in. zachodniopomorskie oraz kujawsko-pomorskie, w których stopa bezrobocia przyjmuje wysokie wartości, co również nie jest zachęcające dla potencjalnych nowych rodzin.

## Literatura

- Bakiera, L. (2022). Styl realizacji rodzicielstwa w kontekście zmian społeczno-kulturowych i demograficznych. W: M.A. Michalski (red.), *Rodzina – wyzwania na XXI wiek* (s. 118-132). Rządowa Rada Ludności. Pobrane z <https://depot.ceon.pl/bitstream/handle/123456789/22306/3KD%20Pozna-%C5%84.pdf?sequence=1&isAllowed=y#page=119>
- Biecek, P. (2011). *Analiza danych z programem R*. Warszawa: Wydawnictwo Naukowe PWN. Pobrane z <https://libra.ibuk.pl/reader/analiza-danych-z-programem-r-modele-liniowe-z-efektami-stalymi-przemyslaw-biecek-39524>
- Brachowicz, M. (2009). Niepłodność jako niemożność doświadczenia własnego rodzicielstwa. W: S. Steuden, K. Jankowski (red.). *Psychospołeczne konteksty doświadczenia straty* (s. 25–39). Wydawnictwo KUL.
- Główny Urząd Statystyczny. (2014). *Prognoza ludności na lata 2014-2050*. Zakład Wydawnictw Statystycznych.
- Grajewski, S. (2006). Zastosowanie analizy skupień w porównawczych badaniach zdolności retencyjnych ekosystemów leśnych. *Infrastruktura i Ekologia Terenów Wiejskich*, 3(1), 155-169. Pobrane z <https://yadda.icm.edu.pl/agro/element/bwmeta1.element.agro-1c5c93d8-6dcf-42d9-879f-dc8c263a8f8>
- Harańczyk, G. (2005). *Analiza skupień na przykładzie segmentacji nowotworów*. StatSoft Polska Sp. z o.o. Pobrane 10.01.2023, z: [https://media.statsoft.pl/\\_old\\_dnn/downloads/analiza\\_skupien\\_na\\_przykladzie\\_segmentacji.pdf](https://media.statsoft.pl/_old_dnn/downloads/analiza_skupien_na_przykladzie_segmentacji.pdf)
- Kaźmierczak-Kałużna, I. (2020). Gdzie te dzieci? O pronatalistycznej (nie) efektywności programu „Rodzina 500 plus”. *Rocznik Lubuski*, 46(2), 131-144.
- Korzeniowski, J. (2012). *Metody selekcji zmiennych w analizie skupień*. Wydawnictwo Uniwersytetu Łódzkiego. <http://dx.doi.org/10.18778/7525-695-6>
- Kukuła, K., Goryl, A., Jędrzejczyk, Z., Osiewalski, J., Walkosz, A. (2009). *Wprowadzenie do ekonometrii*. Wydawnictwo Naukowe PWN. Pobrane z [https://cf-taniaksiążka.statiki.pl/images/files/317/@9788301156718\\_33,145505.pdf](https://cf-taniaksiążka.statiki.pl/images/files/317/@9788301156718_33,145505.pdf)
- Major, M. i Niezgodna, J. (2003). *Elementy Statystyki. Część I. Statystyka opisowa*. Krakowskie Towarzystwo Edukacyjne sp. z o.o. Pobrane z [https://repozytorium.ka.edu.pl/bitstream/handle/11315/28280/MAJOR\\_elementy\\_statystyki\\_2003.pdf?sequence=1&isAllowed=y](https://repozytorium.ka.edu.pl/bitstream/handle/11315/28280/MAJOR_elementy_statystyki_2003.pdf?sequence=1&isAllowed=y)
- Pawłowski, Z. (1981). Wybór modelu ekonometrycznego dla predykcji dyskryminacyjnej. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 43(4), 101-112. Pobrane z <https://repozytorium.amu.edu.pl/bitstream/10593/21164/1/011%20ZBIGNIEW%20PAW%20C5%81OWSKI%20RPEiS%2043%284%29%2C%201981.pdf>
- Ręklewski, M. (2020). *Statystyka opisowa. Teoria i przykłady*. Państwowa Uczelnia Zawodowa we Wrocławiu. Pobrane z <https://kpbc.umk.pl/Content/234634/PDF/Statystyka%20opisowa.pdf>
- Sikorska, M. (2021). *Czy zwiększenie diety w Polsce jest możliwe?* Instytut Badań Strukturalnych. Pobrane z <https://ibs.org.pl/wp-content/uploads/2022/12/Czy-zwiekszenie-dietnosci-w-Polsce-jest-mozliwe.pdf>

- Sobolewski, M. i Sokołowski, A. (2017). Grupowanie metodą k-średnich z warunkiem spójności. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (468). doi: 10.15611/pn.2017.468.22
- Zalewska, E. (2017). Zastosowanie analizy skupień i metody porządkowania liniowego w ocenie polskiego szkolnictwa wyższego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (469). doi: 10.15611/pn.2017.469.24
- Zimny, A. (2010). Statystyka opisowa. Materiały pomocnicze do ćwiczeń. Państwowa Wyższa Szkoła Zawodowa w Koninie. Pobrane z [https://depot.ceon.pl/bitstream/handle/123456789/12550/statystyka\\_opisowa.pdf?sequence=1&isAllowed=y](https://depot.ceon.pl/bitstream/handle/123456789/12550/statystyka_opisowa.pdf?sequence=1&isAllowed=y)

## Analysis of Fertility Rate and its Determinants in Poland in 2021

**Abstract:** For several years, European countries have been struggling with a significant demographic problem. Despite an implementation of various pro-family programs, the fertility rate has not increased. This situation also applies to Poland. The aim of the article is to examine the determinants of the fertility rate and its variations in the voivodeships of Poland in 2021. In the analysis were used both econometric modeling and cluster analysis. It turned out that the fertility rate is very complex and difficult to characterize unambiguously. The factors selected for analysis did not allow creation of a model that would describe this phenomenon to a sufficient degree. Moreover, regional differences are still existing in Poland, what results in a much higher fertility rate in some regions than in others. Cluster analysis identified which regions are related to each other based on the analyzed factors and which clusters emerged as a result of said process.

**Keywords:** fertility, demography, fertility rate, econometric modeling, cluster analysis