

**WSPÓŁCZESNE PROBLEMY
SPOŁECZNO-EKONOMICZNE
W UJĘCIU ANALITYCZNYM**

D E B I U T Y S T U D E N C K I E

2024

**WSPÓŁCZESNE PROBLEMY
SPOŁECZNO-EKONOMICZNE
W UJĘCIU ANALITYCZNYM**

pod redakcją
Agnieszki Stanimir



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2024

Recenzja

Klaudia Przybysz

Redakcja wydawnicza

Aleksandra Śliwka

Korekta

Katarzyna Gwizda

Skład i łamanie

Małgorzata Myszowska

Projekt okładki

Beata Dębska

Na okładce wykorzystano zdjęcia z Adobe Stock i 123rf

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 978-83-67899-75-8 (wersja papierowa)

ISBN 978-83-67899-76-5 (wersja elektroniczna)

DOI: 10.15611/2024.76.5

Druk i oprawa: TOTEM

Spis treści

Wstęp	7
Dominika Perdon: Klasyfikacja krajów Unii Europejskiej ze względu na sytuację kobiet na rynku pracy	11
Aleksandra Derda: Diagnoza nierówności płacowej wśród sportowców z wykorzystaniem metod uczenia maszynowego	25
Agnieszka Krajnik: Analiza relacji między stylem życia Europejczyków i czynnikami ekonomicznymi.....	42
Żaneta Ratajczak: Zastosowanie metody wzorca rozwoju do badania wysokości wybranych dofinansowań ze środków PFRON w województwach Polski	57
Kornelia Kozaczewska: Analiza interwencji na przykładzie badania wpływu pandemii COVID-19 na dzietność za pomocą modeli SARIMA.....	69
Marta Stasiak: Badanie postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej.....	86
Martyna Kamińska: Przemoc seksualna w Polsce w latach 2013-2021: analiza skali zjawiska i skutków społeczno-ekonomicznych przemocy ze względu na płeć w kontekście UE.....	101
Julia Dymek: Badanie poziomu realizacji założeń cyberbezpieczeństwa firm i instytucji Unii Europejskiej	117
Stanisław Marek Halkiewicz, Paweł Rychlicki: Oczyszczanie wody, jej pobór i zużycie na terenach górniczych i pogórnich: analiza przestrzenna powiatów województwa śląskiego	132
Wojciech Świercz, Radosław Szostak: Modelowanie ekonometryczne kapitału inwestorów indywidualnych na rynku głównym GPW	149

Wstęp

Niniejszy zeszyt Debiutów Studenckich pod tytułem *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* składa się z dziesięciu artykułów. Prezentowane prace dotyczą bardzo zróżnicowanych obszarów koncentrujących się na jakości życia oraz problemach społeczno-ekonomicznych oraz aspektach dotyczących bezpieczeństwa jednostek, instytucji i firm. Każdy z zaprezentowanych artykułów zawiera wnikliwie przeprowadzony przegląd literatury, analizę problemu badawczego oraz wnioski ze wskazaniem kolejnych, niezbędnych w danym obszarze tematycznym, badań.

W pierwszym z publikowanych artykułów Dominika Perdon zaprezentowała swoje rozważania na temat sytuacji kobiet na rynku pracy w państwach członkowskich Unii Europejskiej. Autorka, wskazując, że równe szanse obu płci w dostępie do pracy zarobkowej niosą ze sobą wiele korzyści, właściwie przeprowadziła przegląd literatury, co dało podstawy do analizy wskaźnika zatrudnienia w Unii Europejskiej oraz klasyfikacji państw członkowskich ze względu na wybrane zmienne opisujące sytuację kobiet na rynku pracy. Praca zawiera ciekawie i wnikliwie prowadzone wnioskowanie.

Zagadnienia nierówności na rynku pracy są kontynuowane w kolejnym artykule. Aleksandra Derda zaprezentowała w swojej pracy wyniki badań prowadzonych nad nierównościami płacowymi wśród sportowców. Autorka przeprowadziła swoje analizy na podstawie wypłat koszykarzy i koszykarek z ligi NBA i WNBA. Jednak – jak słusznie zauważa – luki płacowe nie są obserwowane tylko w tej dziedzinie sportu, a ich wielkość jest znacząca. Aleksandra Derda przedstawiła również różnice między luką płacową w sporcie i w innych zawodach. Na uwagę zasługuje również trafne wykorzystanie narzędzi analitycznych z grupy metod uczenia maszynowego.

Artykuł Agnieszki Krajnik zwiera wyniki i wnioski z badań dotyczących sytuacji ekonomicznej w państwach europejskich, ale w odniesieniu do zdrowego stylu życia. Autorka trafnie wyodrębniła w pierwszej części artykułu aspekty związane ze zdrowym stylem życia, szczególny nacisk kładąc na koszty związane z ich realizacją. Dzięki temu druga część artykułu zawiera wnioski z porównania rankingów państw pod względem zdrowego stylu życia, sytuacji ekonomicznej i wydatków na ochronę zdrowia. Artykuł obfituje w ciekawe konkluzje.

Analiza wysokości wybranych dofinansowań ze środków PFRON w województwach Polski była tematem przewodnim badań przeprowadzonych przez Żanetę Ratajczak. Autorka skupiła się na rozpoznaniu sytuacji na rynku pracy osób z niepełnosprawnością i zagadnieniem ich aktywizacji zawodowej. Rozważania teoretycz-

ne dały podstawy do przeprowadzenia badania, dzięki czemu możliwe okazało się stworzenie rankingu województw, z uwzględnieniem tych najbardziej zaawansowanych w aktywizacji zawodowej wskazanej grupy społecznej. Wnioski z przeprowadzonego badania ukazują konieczność stałego monitorowania sytuacji.

Pandemia COVID-19 wywołała wiele negatywnych skutków, które najczęściej analizuje się, skupiając się na kwestiach zdrowotnych lub ekonomicznych. Kornelia Kozaczewska w swoim artykule zwróciła uwagę na inną konsekwencję pandemii, a mianowicie na występowanie relacji z poziomem dzietności. Opierając się danych GUS, autorka zbadała interwencje, wykorzystując modele SARIMA. Wnikliwie przeprowadzona analiza zaowocowała ciekawymi wnioskami.

Bardzo ważnym aspektem współczesnego świata jest również realizacja we wszystkich obszarach życia i funkcjonowania celów zrównoważonego rozwoju. Marta Stasiak zaprezentowała w swojej pracy wyniki analiz postaw i zachowań polskich turystów wobec turystyki zrównoważonej. Autorka rozpoczęła rozważania od przybliżenia idei zrównoważonego rozwoju w turystyce i oczekiwanych efektów dla lokalnych gospodarek, zachowania środowiska naturalnego, korzyści dla mieszkańców i osób przyjeżdżających. Analiza przeprowadzona na podstawie danych pierwotnych z wykorzystaniem właściwych metod pozwoliła na trafne określenie subiektywnych postaw turystów wobec nowego, niezbędnego nurtu turystycznego.

Martyna Kamińska w swoim artykule przeprowadziła ważną analizę danych dotyczących aktów przemocy seksualnej, które to dane pozyskała z Komendy Głównej i Komendy Wojewódzkiej Policji. Autorka wskazała, że przebieg i efekty takich postępowań są wstrząsające dla ofiar, a jednocześnie wywołują one skutki społeczne i ekonomiczne. Z tego względu prezentowana praca stanowi ważny głos w dyskusji o przeciwdziałaniu różnym typom przemocy i konieczności wsparcia ofiar. Należy podkreślić, że dane, którymi dysponowała autorka, dotyczą jedynie zgłoszonych przypadków, a wiele takich aktów przemocy pozostaje niezgłoszonych. Zaprezentowane w artykule wnioski wskazują jednoznacznie na znaczące konsekwencje psychologiczne, fizyczne i gospodarcze omawianych czynów.

Coraz częściej pojawiające się zagrożenia związane z cyberbezpieczeństwem stały się podstawą do przeprowadzenia badań przez Julię Dymek. Autorka skupiła się na badaniu poziomu realizacji założeń cyberbezpieczeństwa Unii Europejskiej w odniesieniu do małych i średnich przedsiębiorstw. Przeprowadzona wielowymiarowa analiza porównawcza umożliwiła pogrupowanie państw, w których sytuacja dotycząca cyberbezpieczeństwa w MŚP jest zbliżona, oraz porangowanie państw tak, by wskazać, w których z nich sytuacja jest zadowalająca, a w których cyberbezpieczeństwo nie jest jeszcze należycie gwarantowane.

Autorzy kolejnego artykułu, Stanisław Marek Halkiewicz i Paweł Rychlicki, zwrócili uwagę na narastający problem kryzysu wodnego w kontekście przemysłu wydobywczego. Autorzy trafnie zauważyli, że wodochłonność procesów wydobywczych jest jednym z ważniejszych problemów wskazanego typu przemysłu, będącego jednocześnie strategicznym ogniwem bezpieczeństwa ekonomicznego i energetycz-

nego. Korzystając z autokorelacji przestrzennej, otrzymali informację o regionach najbardziej narażonych na negatywne efekty przemysłu wydobywczego – związane zarówno z poborem i zużyciem wody, jak i oczyszczaniem ścieków.

Celem ostatniego z prezentowanych artykułów – autorstwa Wojciecha Świercza i Radosława Szostaka – było sprawdzenie, jakie czynniki wpływają na kapitał inwestowany przez inwestorów indywidualnych na Giełdzie Papierów Wartościowych w Warszawie. Praca zawiera efekty modelowania ekonometrycznego opisującego wielkość kapitału inwestorów indywidualnych na GPW oraz pozwalającego na identyfikację czynników makroekonomicznych i psychologicznych, które mogą mieć wpływ na decyzje inwestycyjne. Przedstawione przez autorów wnioski dają szansę na lepsze zrozumienie zachowań inwestycyjnych.

Agnieszka Stanimir

Dominika Perdon

e-mail: 182127@student.ue.wroc.pl

ORCID: 0009-0005-5272-649X

Uniwersytet Ekonomiczny we Wrocławiu

Klasyfikacja krajów Unii Europejskiej ze względu na sytuację kobiet na rynku pracy

DOI: 10.15611/2024.76.5.01

JEL: J16, J71, M14

© 2024 Dominika Perdon

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>

Cytuj jako: Perdon, D. (2024). Klasyfikacja krajów Unii Europejskiej ze względu na sytuację kobiet na rynku pracy. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 11-24). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Równe szanse obu płci w dostępie do pracy zarobkowej niosą ze sobą wiele korzyści – zarówno ekonomicznych, jak i społecznych. Niniejszy artykuł zawiera przegląd literatury dotyczący stereotypów, z którymi muszą mierzyć się kobiety, analizę wskaźnika zatrudnienia w krajach Unii Europejskiej oraz klasyfikację krajów Unii Europejskiej pod względem dziewięciu wybranych zmiennych dotyczących sytuacji kobiet na rynku pracy. Analizę przeprowadzono, korzystając z metody *k-średnich*. Celem artykułu było wyłonienie krajów, w których sytuacja kobiet na rynku pracy jest do siebie podobna, oraz identyfikacja zmiennych, które wypadają w tych krajach najlepiej oraz najgorzej. Analiza wykazała, że najlepsza sytuacja kobiet na rynku pracy jest w klasie, do której należą Estonia, Łotwa i Litwa. Najgorsza sytuacja kobiet na rynku pracy występuje w skupieniu, do którego należą Włochy, Grecja i Rumunia. Interesujące jest, że w skupieniu tym odnotowano najniższe, a więc najlepsze wartości luki płacowej.

Słowa kluczowe: stereotypy, luka płacowa, rynek pracy, nierówności

1. Wstęp

Sytuacja kobiet na rynku pracy w wielu krajach Unii Europejskiej różni się od sytuacji mężczyzn. Społeczeństwo zaczyna być coraz bardziej świadome nierówności wynikających z tych różnic, jednakże podejmowane do tej pory działania nie były na tyle skuteczne, aby wyeliminować obecne dysproporcje. Występuje zatem potrzeba nagłaśniania sytuacji kobiet na rynku pracy, ponieważ im rzadziej odmienne traktowanie kobiet i mężczyzn będzie postrzegane w społeczeństwie jako normalność, tym szybciej uda się wyrównać szanse obu płci w dostępie do pracy zarobkowej. O roli kampanii informacyjnych w wyrównywaniu szans kobiet na rynku w pracy wspomi-

na w swoim raporcie chociażby Polska Agencja Rozwoju Przedsiębiorczości (2023), sugerując nawet ich prowadzenie już dla młodzieży w wieku szkolnym, by wchodząc w dorosłość, była ona jak najbardziej pozbawiona stereotypów i uprzedzeń.

Równe szanse obu płci niosą wiele korzyści dla całego społeczeństwa. Dzięki nim kobiety są w stanie rozwijać się i spełniać zawodowo, czym przyczyniają się do wzrostu gospodarczego (Polski Instytut Ekonomiczny, 2023), jednocześnie stając się niezależne finansowo, mając własny wkład w budżet domowy i wyższą pozycję w społeczeństwie. Mężczyźni natomiast przestają być obciążani brzemieniem bycia głównym żywicielem rodziny. Jak wspomina Kukuczka (2023), w przeszłości mogło to się wiązać z ich nadmiernym psychicznym i emocjonalnym obciążeniem, a według Polskiego Instytutu Ekonomicznego (2023) w niektórych przypadkach powoduje to narzucenie im niepasujących ról społecznych pogarszających jakość ich życia.

Niniejszy artykuł jest rozszerzeniem badań, których wyniki zaprezentowano w opracowaniu *Analiza sytuacji kobiet na rynku pracy w krajach Unii Europejskiej* (Perdon, 2023b). Przytoczone są w nim również fragmenty pracy licencjackiej autorki, gdzie badała ona sytuację kobiet na rynku pracy (Perdon, 2023a). Jego celem jest klasyfikacja krajów Unii Europejskiej względem sytuacji kobiet na rynku pracy oraz odpowiedź na postawione pytania badawcze:

- PB1: Jakie kraje Unii Europejskiej zaliczane są do grupy, w której sytuacja kobiet na rynku pracy jest najlepsza?
- PB2: Jakie kraje Unii Europejskiej zaliczane są do grupy, w której sytuacja kobiet na rynku pracy jest najgorsza?
- PB3: Które analizowane czynniki wypadają najlepiej wśród krajów charakteryzujących się najlepszą sytuacją kobiet na rynku pracy, a które wymagają zdecydowanej poprawy?
- PB4: Które analizowane czynniki wypadają najgorzej wśród krajów charakteryzujących się najsłabszą sytuacją kobiet na rynku pracy? Czy istnieją zmienne, które w tych krajach przyjmują korzystne wartości?

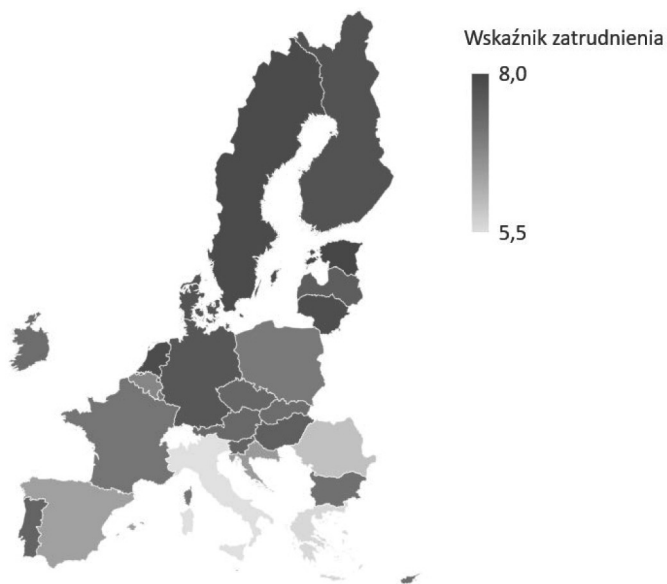
Pierwsza część artykułu zawiera omówienie wskaźnika zatrudnienia w krajach Unii Europejskiej oraz informacje o stereotypach, z którymi kobiety mierzą się na rynku pracy. Następnie przedstawiono opis danych oraz metody badawczej. Kolejny punkt zawiera analizę przedstawionych danych z wykorzystaniem metody *k*-średnich oraz omówienie otrzymanych wyników. W zakończeniu zawarto podsumowanie analizy oraz wnioski.

2. Kobiety a rynek pracy

2.1. Wskaźnik zatrudnienia

Dyrektywa Parlamentu Europejskiego (2019) wskazuje, że ogromną częścią problematyki nierówności płci jest nierówność w dziedzinie zatrudnienia, wynikająca z trudności pogodzenia ról zawodowych i rodzinnych. Odzwierciedleniem tego

problemu jest wysokość wskaźników zatrudnienia w podziale na płeć. Wskaźnik przedstawia udział, jaki w danej kategorii wiekowej mają osoby pracujące (Eurostat, 2023a). Zaprezentowano go na rys. 1, korzystając z danych dla roku 2022.



Rys. 1. Wskaźnik zatrudnienia kobiet

Źródło: opracowanie własne z wykorzystaniem danych Eurostatu (2023a).

Rysunek 1 przedstawia wysokość wskaźnika zatrudnienia kobiet w roku 2022 w podziale na kraje. Im wyższe jest natężenie koloru, tym wyższa jest wartość wskaźnika zatrudnienia. Z kartogramu wynika, że najwyższy wskaźnik zatrudnienia obserwowalny jest w rejonie Europy Północnej: w Estonii (80,4%), Szwecji (79,2%), Litwie (78,6%). Najniższy wskaźnik zatrudnienia notowany jest w rejonie Europy Południowej: we Włoszech (55%) i w Grecji (55,9%).

Tabela 1. Wartości wskaźnika zatrudnienia w 2022 r. w podziale na płeć

Kraj	Wskaźnik zatrudnienia kobiet (jako procent populacji)	Wskaźnik zatrudnienia mężczyzn (jako procent populacji)	Różnica (punkty procentowe)
1	2	3	4
Belgia	68,1	75,7	7,6
Bułgaria	71,8	79,5	7,7
Czechy	73,7	88,6	14,9
Dania	77,4	82,8	5,4
Niemcy	76,8	84,6	7,8

Tabela 1, cd.

1	2	3	4
Estonia	80,4	83,3	2,9
Irlandia	72,6	83,9	11,3
Grecja	55,9	76,9	21,0
Hiszpania	64,1	75,0	10,9
Francja	71,2	77,0	5,8
Chorwacja	65,0	74,5	9,5
Włochy	55,0	74,7	19,7
Cypr	72,1	84,2	12,1
Łotwa	75,5	78,6	3,1
Litwa	78,6	79,4	0,8
Luksemburg	71,5	78,0	6,5
Węgry	75,3	85,1	9,8
Malta	74,1	87,2	13,1
Holandia	79,0	86,9	7,9
Austria	73,4	81,2	7,8
Polska	70,2	83,1	12,9
Portugalia	74,3	80,2	5,9
Rumunia	59,1	77,7	18,6
Słowenia	74,3	81,2	6,9
Słowacja	72,6	80,7	8,1
Finlandia	77,8	79,0	1,2
Szwecja	79,2	85,0	5,8

Źródło: opracowanie własne z wykorzystaniem danych Eurostatu (2023a).

Najmniejsza przedstawiona w tabeli 1 różnica pomiędzy wskaźnikiem zatrudnienia kobiet a mężczyzn wynosi 0,8 punktu procentowego i występuje w Litwie, natomiast największą różnicę odnotowano w Grecji: aż 21 punktów procentowych. Wskazuje to na występowanie znaczących dysproporcji w wysokości wskaźnika zatrudnienia – zarówno pomiędzy płciami, jak i pomiędzy analizowanymi krajami.

Analizując szerzej dane dotyczące wskaźnika zatrudnienia, można zauważyć, że wskaźnik ten dla kobiet maleje wraz z liczbą posiadanych przez nie dzieci. Według danych z Eurostatu (2023b) w roku 2022 w grupie wieku 25-49 lat średni wskaźnik zatrudnienia w krajach Unii Europejskiej wynosił 79,7% dla kobiet bezdzietnych, 76,2% dla kobiet mających jedno dziecko, 75,3% dla kobiet mających dwójkę dzieci oraz 59,3% dla kobiet mających trójkę lub więcej dzieci.

Zygmunt (2006) wspomina, że tak wysokie różnice w aktywności zawodowej kobiet mogą wynikać w sporej części z wykonywania przez nie prac domowych, opieki nad dziećmi oraz z przekonań spowodowanych stereotypami. Zdaniem autorki kluczowe znaczenie ma tutaj również podejście pracodawców do traktowania płci jako kryterium podczas wyboru pracownika. Czasem bowiem zakładają oni, że kobieta będzie mniej dyspozycyjna oraz częściej będzie korzystać z przysługujących jej uprawnień (urlopy macierzyńskie, skrócenie czasu pracy podczas karmienia piersią, opieka nad chorującymi dziećmi), co może generować dla nich dodatkowe koszty. Podobne zdanie mają także Ratajczak i Kozłowska (2023), według których właśnie wsparcie oferowane przez pracodawcę może znacząco wyrównać szanse na rynku pracy i zaktywizować zawodowo oraz rodzicielsko oboje rodziców, co w konsekwencji przyczyni się do wyrównania istniejących dysproporcji.

2.2. Stereotypy na rynku pracy

Na rynku pracy kobiety muszą mierzyć się z wieloma stereotypami, o czym wspomina na przykład Tomaszewska (2023), zdaniem której to właśnie reprezentantom płci żeńskiej przypisywane są cechy, takie jak słabość, czy mniejsze kompetencje niż u mężczyzn, co może przekładać się na ich niedoreprezentowanie na wyższych stanowiskach. Galsanjigmed i Sekiguchi (2023) wskazują, że stereotypy narzucają określone zachowania reprezentantom różnych płci. Kobieta, która w środowisku zawodowym przyjmuje postawę asertywną, jest niejednokrotnie postrzegana jako nieprzyjazna, natomiast od mężczyzn właśnie takiej asertywności się oczekuje. Takie podejście prowadzi natomiast do rozpowszechniania uprzedzeń i większego ograniczania możliwości awansu zawodowego kobiet. Heilman i in. (2023) dodają, że wbrew powszechnej opinii stereotypy dotyczące płci zmieniły się mniej, niż wskazują na to zmiany społeczne, a także poruszają kolejny ważny problem związany ze stereotypowym postrzeganiem kobiet. Według autorek atrakcyjne kobiety rekrutowane na stanowiska w zawodach zdominowanych przez mężczyzn są często uważane za mniej wykwalifikowane.

W roku 2023 powstał raport „Niech nas usłyszą! Głos kobiet w korporacjach” (Golec i in., 2023). Jego autorki przeprowadziły kompleksowe badanie wśród ponad 4 tysięcy kobiet. Jednym z elementów badania było odniesienie się do 13 wymienionych stereotypów, dotyczących między innymi nieumiejętności opanowania emocji przez kobiety, nadawania karierom mężczyzn większego znaczenia, a także twierdzenia, że młode kobiety nie wykorzystają awansu należycie, ponieważ w niedalekiej przyszłości zostaną matkami i skorzystają z urlopu rodzicielskiego. Jedynie 0,8% ankietowanych kobiet odpowiedziało, że nigdy nie spotkało się z wymienionymi stereotypami. Jednocześnie aż 55,7% respondentek wskazało, że kobiety muszą mierzyć się z większymi oczekiwaniami stawianymi przez przełożonych niż mężczyźni pracujący na tych samych stanowiskach. Na uwagę zasługuje również fakt, że aż 91,1% ankietowanych potwierdziło, że awans i sukces zawodowy są dla nich ważne,

ale tylko 34,4% respondentek odpowiedziało, że według nich kobiety w ich branżach otrzymują takie same szanse rozwoju i awansu jak mężczyźni (Golec i in., 2023). Wyniki przywołanego raportu jasno wskazują, że w temacie równości kobiet na rynku pracy wciąż pozostaje wiele do poprawy, a dyskryminacja i dysproporcje pomiędzy płciami są mocno odczuwalne.

3. Dane i metoda analizy

3.1. Opis danych

Aby porównać kraje Unii Europejskiej pod względem sytuacji kobiet na rynku pracy, posłużono się klasyfikacją metodą k -średnich. Wykorzystano do niej następujące zmienne, które szerzej opisano w poprzedniej publikacji (Perdon, 2023b):

x_1 – luka płacowa w roku 2020 (Eurostat, 2023f); dane dla Grecji nie były dostępne, w związku z tym w tym kraju przyjęto do analizy lukę płacową z roku 2018; destymulanta;

x_2 – luka w zatrudnieniu w roku 2020 (Eurostat, 2023d); destymulanta;

x_3 – luka w zatrudnieniu w niepełnym wymiarze czasu pracy w roku 2020 (Eurostat, 2023e); destymulanta;

x_4 – udział kobiet zajmujących stanowiska w rządach w roku 2020 (Eurostat, 2023g); stymulanta;

x_5 – udział kobiet zajmujących stanowiska w parlamentach w roku 2020 (Eurostat, 2023g); stymulanta;

x_6 – procent kobiet mających wykształcenie wyższe w roku 2020 (Eurostat, 2023h); stymulanta;

x_7 – procent kobiet wśród osób pracujących na stanowiskach kierowniczych w roku 2020 (Eurostat, 2023c); stymulanta;

x_8 – procent kobiet zagrożonych ubóstwem lub wykluczeniem społecznym w roku 2020 (Eurostat, 2022b); destymulanta;

x_9 – długość życia zawodowego w roku 2020 wyrażona w latach (Eurostat, 2022a); stymulanta.

3.2. Opis metody

K -średnich to metoda umożliwiająca podział zbioru na podzbiory w taki sposób, że obiekty wśród danego podzbioru są do siebie jak najbardziej podobne i jednocześnie jak najmniej podobne do obiektów z innych podzbiorów. Algorytm tej metody jest iteracyjny, na początku oblicza się wektory średnich, następnie odległości obiektów od wektora średnich. W kolejnym kroku powstaje nowa klasyfikacja, ponieważ obiekty są na nowo przydzielane do klas na podstawie najmniejszych odległości od wyznaczonego wektora. Iteracje przestają być powtarzane, gdy nowo powstała

klasyfikacja nie różni się od poprzedniej (Kowalewski, 2006). Do wykonania analizy metodą k -średnich użyto danych po standaryzacji według wzoru:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

Podczas wykonywania tej analizy wykorzystano program Statistica v.13.3.

4. Klasyfikacja krajów metodą k -średnich

Podczas wykonywania klasyfikacji metodą k -średnich użyto danych poddanych standaryzacji. Analiza została w dużej części wykonana w programie Statistica. Wstępne centra skupień dobrano tak, aby zmaksymalizować odległości skupień. Początkowa liczba skupień została wybrana jako 6, a wstępna liczba iteracji została wyznaczona jako 10. Rozwiązanie udało się znaleźć już po 2 iteracjach.

Tabela 2 przedstawia odległości euklidesowe pomiędzy skupieniami. Poniżej przekątnej znajdują się odległości euklidesowe, a powyżej kwadraty tych wartości. Zbyt małe odległości mogłyby być przesłanką do zmniejszenia liczby skupień, jednak w tym przypadku najmniejsza odległość wynosi 0,788, a największa 1,959, a więc redukcja liczby klas nie jest konieczna.

Tabela 2. Odległości euklidesowe skupień

Numer skupienia	1	2	3	4	5	6
1	0,000	0,803	1,763	3,028	0,861	0,959
2	0,896	0,000	0,620	1,485	1,743	1,015
3	1,328	0,788	0,000	1,403	1,932	1,469
4	1,740	1,219	1,184	0,000	3,520	3,838
5	0,928	1,320	1,390	1,876	0,000	1,972
6	0,980	1,007	1,212	1,959	1,404	0,000

Źródło: opracowanie własne z wykorzystaniem programu Statistica v.13.3.

W trakcie badania przeprowadzono analizę wariancji dla każdej zmiennej. Wykazała ona, że jedynie zmienna x_7 , a więc procent kobiet wśród wszystkich osób pracujących na stanowiskach kierowniczych, nie różnicuje klas ($F = 1,746$, p -value = 0,168).

W tabeli 3 przedstawiono podział krajów na skupienia oraz przedstawiono, jakie są odległości tych krajów od środków ich skupień. Utworzone skupienia prezentują się następująco:

- skupienie pierwsze: Belgia, Dania, Hiszpania, Francja, Portugalia, Finlandia, Szwecja;
- skupienie drugie: Bułgaria, Irlandia, Cypr, Luksemburg, Polska i Słowenia;

- skupienie trzecie: Czechy, Chorwacja, Węgry, Malta, Słowacja;
- skupienie czwarte: Grecja, Włochy, Rumunia;
- skupienie piąte: Niemcy, Holandia, Austria;
- skupienie szóste: Estonia, Łotwa, Litwa.

Tabela 3. Skupienia (Sk.) oraz odległości obiektów od środków skupień (Odl.)

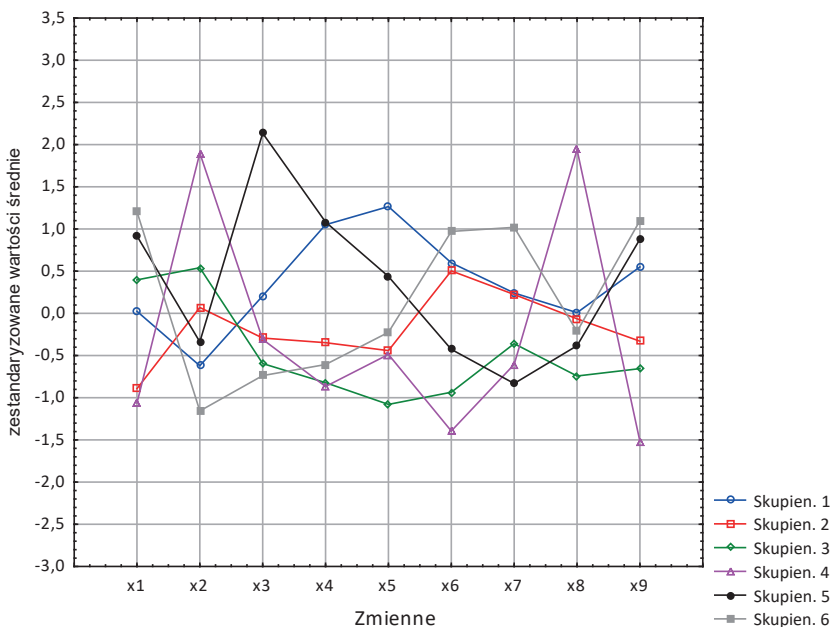
Sk.1	Odl.	Sk.2	Odl.	Sk.3	Odl.	Sk.4	Odl.	Sk.5	Odl.	Sk.6	Odl.
Belgia	0,670	Bułgaria	0,823	Czechy	0,478	Grecja	0,533	Niemcy	0,355	Estonia	0,430
Dania	0,607	Irlandia	0,474	Chorwacja	0,544	Włochy	0,659	Holandia	0,446	Łotwa	0,439
Hiszpania	0,744	Cypr	0,688	Węgry	0,526	Rumunia	0,566	Austria	0,356	Litwa	0,462
Francja	0,394	Luksemburg	0,810	Malta	0,677						
Portugalia	0,582	Polska	0,717	Słowacja	0,449						
Finlandia	0,700	Słowenia	0,690								
Szwecja	0,692										

Źródło: opracowanie własne z wykorzystaniem programu Statistica v.13.3.

Warto zaznaczyć, że gdyby w tab. 2 odległości od środków skupień były zbyt duże, byłoby to przesłanką do stworzenia kolejnego skupienia i do ponownego przeprowadzenia klasyfikacji. W tym przypadku jednak nie jest to konieczne, ponieważ żadna odległość od środka skupienia nie odstaje drastycznie od innych wartości. Należy także pamiętać, że w poprzednim etapie badania, ze względu na występujący brak danych, założono, że wartość luki płacowej dla Grecji z roku 2018 nie zmieniła się drastycznie i mogła zostać przyjęta zamiast wartości z roku 2020, która nie była dostępna. W pozostałych przypadkach wszystkie zmienne odnoszą się do roku 2020.

Rysunek 2 przedstawia graficznie średnie w podziale na skupienia. Bazując na przedstawionych informacjach, można wyciągnąć wniosek, że najlepsze średnie wartości dla luki płacowej przyjmują kraje w skupieniu czwartym (zmienna x_1 jest destymulantą). Należy jednak mieć na uwadze, że do tego skupienia należy Grecja, w przypadku której wartość luki płacowej została pobrana z roku 2018, a nie 2020, tak jak w przypadku pozostałych krajów. Najlepsze średnie wartości dla luki w zatrudnieniu mają kraje w skupieniu szóstym, gdyż ta zmienna również jest destymulantą. Widoczne są tu wysokie wartości w skupieniu czwartym. Skupienie szóste przyjmuje również najlepsze średnie wartości dla luki zatrudnienia w niepełnym wymiarze czasu pracy. Jeśli chodzi o średni udział kobiet zajmujących stanowiska w rządach, najlepiej wypada skupienie piąte. Tuż za nim plasuje się skupienie pierwsze, które z kolei przoduje pod względem odsetka kobiet zajmujących stanowiska w parlamentach. Dla obu tych zmiennych skupienia pierwsze i piąte zdecydowanie odstają od pozostałych. Pod względem średniego procenta kobiet mających wykształcenie wyższe przoduje skupienie szóste, które przyjmuje najlepsze wartości również dla średniego odsetka kobiet wśród osób zatrudnionych na stanowiskach

kierowniczych. Procent kobiet zagrożonych ubóstwem lub wykluczeniem społecznym to kolejna rozpatrywana destymulanta – pod względem średnich wartości tej zmiennej najlepiej wypadło skupienie trzecie. Wybijają się tutaj też bardzo wysokie wartości skupienia czwartego, co stawia je w bardzo niekorzystnym świetle. Średnia długość życia zawodowego najlepiej wypadła w skupieniu szóstym.



Rys. 2. Wykres średnich w podziale na skupienia

Źródło: opracowanie własne z wykorzystaniem programu Statistica.

Na rysunku 2 wyróżnia się skupienie czwarte, do którego należą Grecja, Włochy i Rumunia. Dla dziewięciu rozpatrywanych zmiennych skupienie to przyjmuje najgorsze wartości aż pięć razy. Wskazuje to na bardzo niekorzystną sytuację kobiet na rynku pracy w tych krajach. Co ciekawe, pomimo bardzo niekorzystnej sytuacji kobiet na rynku pracy sugerowanej przez pięć zmiennych, kraje te przyjmują najniższe średnie wartości luki płacowej wśród krajów Unii Europejskiej (należy mieć jednak na uwadze, że dane dla Grecji pochodzą z roku 2018). Oznacza to, że mimo iż luka płacowa jest najczęściej wspomnianym wskaźnikiem podczas oceniania sytuacji kobiet na rynku pracy, należy ostrożnie interpretować jej wartości w ogólnej ocenie. Kraje mające niskie i bardzo dobre wartości luki płacowej wypadły bardzo źle w ogólnej ocenie sytuacji kobiet na rynku pracy.

Skupienie pierwsze jest największe i składa się z Belgii, Danii, Hiszpanii, Francji, Portugalii, Finlandii i Szwecji. Grupa tych państw charakteryzuje się najwyższymi, a więc najlepszymi wartościami średniego udziału kobiet zajmujących stanowiska

w parlamentach oraz drugimi najlepszymi wartościami dla średniego udziału kobiet zajmujących stanowiska w rządach. Drugie najlepsze wartości odnotowano w tym skupieniu dla zmiennych: luka w zatrudnieniu, procent kobiet posiadających wykształcenie wyższe oraz procent kobiet zagrożonych ubóstwem. W przypadku pozostałych zmiennych wartości średnie dla tego skupienia wypadają przeciętnie.

Skupienie drugie, składające się z Bułgarii, Irlandii, Cypru, Luksemburga, Polski i Słowenii, nie zostało wspomniane ani razu, ponieważ nie ma ono wartości skrajnych – ani razu nie wypadło ani najlepiej, ani najgorzej. Warto wspomnieć, że skupienie to ma drugie najlepsze wartości pod względem luki płacowej, jednak można uznać, że sytuacja kobiet na rynku pracy w tych krajach wypada przeciętnie.

W skupieniu trzecim, na które składają się Czechy, Chorwacja, Węgry, Malta i Słowacja, również odnotowano niekorzystne wartości dla wielu zmiennych: luka w zatrudnieniu, udział kobiet zajmujących stanowiska w rządach, udział kobiet zajmujących stanowiska w parlamentach, procent kobiet posiadających wykształcenie wyższe, procent kobiet wśród osób pracujących na stanowiskach kierowniczych, długość życia zawodowego. Wyjątkiem jest procent kobiet zagrożonych ubóstwem lub wykluczeniem społecznym, gdyż pod względem tej zmiennej średnie wartości w skupieniu są najniższe, a więc najlepsze. Pod względem luki w zatrudnieniu w niepełnym wymiarze czasu skupienie to ma drugie najlepsze wartości.

Skupienie piąte tworzone jest przez Niemcy, Holandię oraz Austrię. Skupienie to jest zróżnicowane, ponieważ można w nim zaobserwować wartości zarówno bardzo dobre (najlepiej wypada pod względem odsetka kobiet zajmujących stanowiska w rządach, dobrze wypada pod względem długości życia zawodowego i odsetka kobiet zajmujących stanowiska w parlamentach), jak i bardzo niekorzystne (najwyższa średnia luka w zatrudnieniu w niepełnym wymiarze czasu, najmniejszy średni procent kobiet na stanowiskach kierowniczych, druga najwyższa średnia luka płacowa). Pod względem pozostałych zmiennych skupienie plasuje się przeciętnie.

Skupienie szóste, składające się z Estonii, Łotwy i Litwy, przyjmuje najwyższe średnie wartości pod względem luki płacowej, a więc najgorsze. Ciekawy jest jednak fakt, że pod względem innych zmiennych skupienie to wypada bardzo dobrze, ponieważ najlepsze średnie wartości ma ono dla zmiennych: luka w zatrudnieniu, luka w zatrudnieniu w niepełnym wymiarze czasu pracy, procent kobiet posiadających wykształcenie wyższe, procent kobiet wśród osób pracujących na stanowiskach kierowniczych, długość życia zawodowego. Skupienie szóste ma z kolei przeciętne wartości pod względem zmiennych: udział kobiet zajmujących stanowiska w rządach, udział kobiet zajmujących stanowiska w parlamentach oraz procent kobiet zagrożonych ubóstwem lub wykluczeniem społecznym.

Decydując się na wybór skupienia, którego kraje można uznać za najlepsze pod względem sytuacji kobiet na rynku pracy, należy podjąć decyzję, czy bardziej istotne są równe zarobki, czy pozostałe kwestie, związane między innymi z równym dostępem do stanowisk, do pracy na pełen etat oraz do wykształcenia. Jest to ciężka decyzja, ponieważ kraje mające średnie najniższe wartości luki płacowej (skupienie

czwarte) wypadły najgorzej pod względem wielu innych rozpatrywanych zmiennych, natomiast kraje mające średnie najlepsze wartości dla wielu rozpatrywanych zmiennych (skupienie szóste) wypadły najgorzej pod względem luki płacowej. Patrząc jednak na samą liczbę najgorszych i najlepszych wyników, można stwierdzić, że według przeprowadzonego badania najgorsza sytuacja kobiet na rynku pracy występuje w Grecji, Włoszech i Rumunii, natomiast najlepsza sytuacja kobiet na rynku pracy występuje w Estonii, Łotwie i Litwie.

5. Zakończenie

Niniejszy artykuł powstał w celu nagłośnienia wciąż występujących dysproporcji pomiędzy sytuacją kobiet i mężczyzn na rynku pracy oraz w celu klasyfikacji krajów Unii Europejskiej ze względu na wybrane czynniki świadczące o wspomnianych dysproporcjach.

Badania literaturowe oraz danych pozyskanych z platformy Eurostat wykazały istnienie różnych wartości wskaźnika zatrudnienia w zależności od kraju, płci oraz liczby posiadanych dzieci. Dodatkowo potwierdzono występowanie stereotypów, z którymi na rynku pracy muszą się mierzyć kobiety.

Przeprowadzona klasyfikacja krajów Unii Europejskiej z wykorzystaniem metody *k*-średnich pozwoliła dokonać podziału krajów na 6 skupień oraz odpowiedzieć na postawione pytania badawcze.

PB1: Do grupy, w której sytuacja kobiet na rynku pracy jest najlepsza, zaliczane są: Estonia, Litwa i Łotwa. Wyboru dokonano na podstawie liczby zmiennych przyjmujących w analizie najlepsze wartości.

PB2: Do grupy, w której sytuacja kobiet na rynku pracy jest najgorsza, zakwalifikowano: Grecję, Włochy i Rumunię. Wyboru dokonano na podstawie liczby zmiennych przyjmujących w analizie najgorsze wartości.

PB3: Wśród krajów charakteryzujących się najlepszą sytuacją kobiet na rynku pracy (Estonia, Litwa, Łotwa) najlepiej wypadają zmienne: luka w zatrudnieniu, luka w zatrudnieniu w niepełnym wymiarze czasu pracy, procent kobiet posiadających wykształcenie wyższe, procent kobiet wśród osób pracujących na stanowiskach kierowniczych, długość życia zawodowego. Zdecydowanej poprawy wymaga natomiast wysokość luki płacowej, która właśnie w tym skupieniu przyjęła najgorsze wartości.

PB4: Wśród krajów charakteryzujących się najłabszą sytuacją kobiet na rynku pracy (Grecja, Włochy, Rumunia) najgorzej wypadają zmienne: luka w zatrudnieniu, udział kobiet zajmujących stanowiska w rządach, procent kobiet mających wykształcenie wyższe, procent kobiet zagrożonych ubóstwem, długość życia zawodowego. Interesujący jest fakt, że skupienie to wypadło najlepiej pod względem wysokości luki płacowej (przy czym należy mieć na uwadze, że dane dla Grecji pochodzą z roku 2018, a nie 2020).

Niniejszy artykuł jest rozszerzeniem przeprowadzonych wcześniej badań, w których uzyskano nieco inne wyniki (Perdon, 2023b). Wykonane poprzednio porządkowanie liniowe metodą TOPSIS wskazało Finlandię i Szwecję jako kraje, w których sytuacja kobiet na rynku pracy była najlepsza, natomiast Chorwację, Rumunię, Grecję, Włochy i Maltę jako kraje, w których była ona najgorsza. W obu przeprowadzonych badaniach różne kraje zostały zatem wskazane jako najlepsze względem sytuacji kobiet na rynku pracy, natomiast Rumunia, Włochy i Grecja dwukrotnie zostały wskazane jako kraje z najgorszą sytuacją płci żeńskiej pod względem sytuacji zawodowej. W związku z nieco rozbieżnymi wynikami otrzymanymi przy wykorzystaniu różnych metod w kolejnym kroku może zostać przeprowadzona szersza i bardziej dogłębna klasyfikacja z użyciem większej liczby zmiennych. Analiza może zostać również powtórzona za kilka lat, aby porównać otrzymane wyniki i postępy Unii Europejskiej walczącej z nierównościami i dyskryminacją kobiet na rynku pracy.

Literatura

- Dziennik Urzędowy Unii Europejskiej. (2019). Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/1158 z dnia 20 czerwca 2019 r. w sprawie równowagi między życiem zawodowym a prywatnym rodziców i opiekunów oraz uchylająca dyrektywę Rady 2010/18/UE. *Dziennik Urzędowy Unii Europejskiej*, 188/79, 79-93.
- Eurostat. (2022a). *Duration of Working Life*. Pobrano 11 kwietnia 2023 z https://ec.europa.eu/eurostat/databrowser/view/lfsi_dwl_a/default/table?lang=en
- Eurostat. (2022b). *Persons at Risk of Poverty or Social Exclusion by Age and Sex – EU 2020 Strategy*. Pobrano 11 kwietnia 2023 z [https://ec.europa.eu/eurostat/databrowser/view/ilc_peps01\\$DV_566/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/ilc_peps01$DV_566/default/table?lang=en)
- Eurostat. (2023a). *Employment Rate by Sex*. Pobrano 18 lutego 2024 z <https://ec.europa.eu/eurostat/databrowser/view/tesem010/default/table?lang=en>
- Eurostat. (2023b). *Employment Rate of Adults by Sex, Age Groups, Educational Attainment Level, Number of Children and Age of Youngest Child (%)*. Pobrano 21 marca 2024 z https://ec.europa.eu/eurostat/databrowser/view/lfst_hheredch__custom_10514639/default/table?lang=en
- Eurostat. (2023c). *Employed Women Being in Managerial Positions by Age*. Pobrano 11 kwietnia 2023 z <https://ec.europa.eu/eurostat/databrowser/view/tqoe1c2/default/table?lang=en>
- Eurostat. (2023d). *Gender Employment Gap*. Pobrano 11 kwietnia 2023 z <https://ec.europa.eu/eurostat/databrowser/view/tesem060/default/table?lang=en>
- Eurostat. (2023e). *Gender Gap in Part-time Employment*. Pobrano 11 kwietnia 2023 z https://ec.europa.eu/eurostat/databrowser/view/tepsr_lm210/default/table?lang=en
- Eurostat. (2023f). *Gender Pay Gap in Unadjusted form by NACE Rev. 2 Activity – Structure of Earnings Survey Methodology*. Pobrano 15 kwietnia 2023 z https://ec.europa.eu/eurostat/databrowser/view/earn_gr_gpgr2/default/table?lang=en
- Eurostat. (2023g). *Seats Held by Women in National Parliaments and Governments (Source: EIGE)*. Pobrano 04.02.2023 z: https://ec.europa.eu/eurostat/databrowser/view/sdg_05_50/default/table?lang=en

- Eurostat. (2023h). *Students Enrolled in Tertiary Education by Education Level, Programme Orientation, Sex, Type of Institution and Intensity of Participation*. Pobrano 04 lutego 2023 z https://ec.europa.eu/eurostat/databrowser/view/EDUC_UOE_ENRT01_custom_4835319/default/table?lang=en
- Galsanjiimed, E. i Sekiguchi, T. (2023). Challenges Women Experience in Leadership Careers: An Integrative Review. *Merits*, 3(2), 366-389.
- Golec, A., Olszewska-Miszuris, M. i Włodarczyk, A. (2023). *Niech nas usłyszą! Głos kobiet w korporacjach*. 30% Club Poland. <https://30percentclub.org/wp-content/uploads/2023/03/Raport-30-Club-Poland-UNGICNP-Niech-nas-uslysza.pdf>
- Heilman, M. E., Caleo, S. i Manzi, F. (2024). Women at Work: Pathways from Gender Stereotypes to Gender Bias and Discrimination. *Annual Review of Organizational Psychology and Organizational Behavior*, 11, 165-192.
- Kowalewski, G. (2006). Metody klasyfikacji i porządkowania. W: A. Stanimir (red.), *Analiza danych marketingowych. Problemy, metody, przykłady* (s. 101-126). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Kukuczka, A. (2023). *Czy etyka ma płeć?: Rola instytucji finansowych w zrównoważonym rozwoju w zakresie równości płci w polskim biznesie*. Pobrano 23 marca 2024 z https://www.zbp.pl/getmedia/28adac9b-884f-40da-8ce1-449c4d17fb73/Agnieszka-Kukuczka_wyroznienie.pdf
- Perdon, D. (2023a). *Analiza krajów Unii Europejskiej względem sytuacji kobiet na rynku pracy* (praca licencjacka, niepublikowana). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- Perdon, D. (2023b). Analiza sytuacji kobiet na rynku pracy w krajach Unii Europejskiej. W: D. Teneta (red.), *Zarządzanie różnorodnością w organizacji – istota, znaczenie, wyzwania* (s. 69-81). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- Polska Agencja Rozwoju Przedsiębiorczości. [PARP]. (2023). *Rynek pracy, edukacja, kompetencje. Aktualne trendy i wyniki badań (maj 2023)*. https://www.parp.gov.pl/storage/publications/pdf/Rynek_pracy_maj-2023.pdf
- Polski Instytut Ekonomiczny. (2023). *Sytuacja kobiet w Polsce z perspektywy społeczno-ekonomicznej*. Pobrano 23 marca 2023 z https://pie.net.pl/wp-content/uploads/2023/05/PIE-Raport_Sytuacja_kobiet.pdf
- Ratajczak, Ł. P. i Kozłowska, A. (2023). Zmiana modelu ojcostwa. W stronę równoważenia ról rodzicielskich i budowania równowagi pomiędzy życiem zawodowym a rodzinnym. *Resocjalizacja Polska*, (26), 239-254.
- Tomaszewska, M. (2023). Wpływ stereotypów na postrzeganie kobiet w roli menedżera. *Edukacja Ekonomistów i Menedżerów*, 68(2), 79-94.
- Zygmunt, A. (2006). *Postulat równości płci w sferze zatrudnienia i pracy w teorii i praktyce państw Unii Europejskiej i w Polsce* (praca doktorska). Uniwersytet Śląski.

Classification of European Union Countries with Regard to the Situation of Women in the Labour Market

Abstract: Equal opportunities for both genders in accessing the employment have many benefits, both economic and social. This article provides a literature review of the stereotypes that women have to face, an analysis of the employment rate in European Union countries, and a classification of European Union countries according to nine selected variables concerning the situation of women in the labour market, using the *k*-means method. The aim of the article was to identify countries where the labour market situation is similar and to identify the variables that perform best and worst in these countries. The analysis showed that the best labour market situation for women is a class that includes Estonia,

Latvia, and Lithuania. The variables that perform the best in this class concern the employment gap, the part-time employment gap, the percentage of women with tertiary education, the percentage of women among those working in managerial positions, and the length of working life. On the other hand, the gender pay gap definitely needs improvement, since in this cluster it takes on very high values. The worst situation for women in the labour market is in the cluster that includes Italy, Greece, and Romania. It is interesting to note that this cluster has the lowest and therefore the best values of the gender pay gap.

Keywords: stereotypes, gender pay gap, labour market, inequalities

Aleksandra Derda

e-mail: 177427@student.ue.wroc.pl

ORCID: 0009-0003-6022-2238

Uniwersytet Ekonomiczny we Wrocławiu

Diagnoza nierówności płacowych wśród sportowców z wykorzystaniem metod uczenia maszynowego

DOI: 10.15611/2024.76.5.02

JEL: C5, J3, J7

© 2024 Aleksandra Derda

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>

Cytuj jako: Derda, A. (2024). Diagnoza nierówności płacowej wśród sportowców z wykorzystaniem metod uczenia maszynowego. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 25-41). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Głównym celem pracy jest udowodnienie – z wykorzystaniem metody XGBoost – występowania luki płacowej w sporcie na podstawie wypłat koszykarzy i koszykarek z lig NBA i WNBA. Sprawdzano również prawdziwość stwierdzenia, że na wysokość wypłaty wpływają cechy fizyczne, popularność oraz skuteczność na boisku w obronie i ataku zawodnika lub zawodniczki. Utworzono jeden model, którego wyniki uznano za zadowalające. Porównano dla tego modelu wpływ poszczególnych zmiennych z wykorzystaniem trzech różnych metod. Uzyskano spójne wyniki potwierdzające, że poza skutecznością w obronie wszystkie z zakładanych czynników wpływających na wysokość wypłaty okazały się istotne. Potwierdzono również występowanie dyskryminacji płacowej ze względu na płeć.

Słowa kluczowe: luka płacowa, sport, uczenie maszynowe, XGBoost

1. Wstęp

Sport od wieków jest integralną częścią społeczeństwa i kultury każdego narodu (Waśkowski, 2011). Niepodważalna jest jego rola w formowaniu społeczeństwa, kształtowaniu tożsamości narodowej czy promowaniu ludzkich wartości. Uprawiany rekreacyjnie pozytywnie wpływa szczególnie na zdrowie i relacje społeczne jednostki, natomiast zawodowo motywuje do osiągnięcia doskonałości.

Sport profesjonalny to także obszar, w którym pasja, talent czy zaangażowanie są nagradzane publicznym uznaniem i rozgłosem. Jest to widowisko, które gromadzi miliony fanów z całego świata, co przyciąga również największe firmy. Są one w stanie wyłożyć ogromne pieniądze na inwestycje w sport, by być zauważonym przez fanów wybranych dyscyplin. Sektor sportu uznaje się za globalny (Gratton, 1998), dlatego nie powinno dziwić, że branży tej towarzyszą zarówno wielkie

emocje, jak i wielkie pieniądze. Niestety, problemy, z którymi zmagać się muszą czołowi działacze i sportowcy również są znacznie trudniejsze i mają poważniejsze konsekwencje niż te obserwowane w życiu codziennym. Dotyczy to także nierówności społecznych.

Sport zawodowy jest jedną z branż najmocniej dotkniętych dyskryminacją płciową, gdyż w sporcie można zauważyć niemal każdy rodzaj dyskryminacji (Thakur, 2022). O różnych objawach dyskryminacji płciowej można przeczytać w pracy Tomaszewskiej (2004). W sporcie dyskryminacja ze względu na płeć jest znacznie szerszym i niejednokrotnie poważniejszym pojęciem. Dyskryminacja płciowa nie tylko wpływa na doświadczenie zawodowe sportowców na każdym etapie kariery zawodowej, na ich status materialny, ale także kształtuje oblicze całej branży sportowej. Jednym ze skutków występowania dyskryminacji płciowej jest również badana luka płacowa.

Celem tego badania jest udowodnienie występowania luki płacowej w sporcie na przykładzie amerykańskich profesjonalnych lig koszykarskich NBA i WNBA. Dodatkowym celem było przygotowanie modelu dobrze prognozującego wysokość wypłaty zawodnika. W badaniu poszukiwano odpowiedzi na pytania:

- Czy w amerykańskich ligach koszykarskich występuje dyskryminacja płacowa ze względu na płeć?
- Czy płeć ma największy wpływ na predykcję wysokości wypłaty?
- Czy czynniki, jakie mają wpływ na wysokość wypłaty, to cechy fizyczne zawodnika, jego popularność oraz skuteczność na boisku w obronie i ataku?

2. Luka płacowa w sporcie

Pojęcie „luka płacowa” jest bezpośrednim tłumaczeniem angielskiego zwrotu *gender pay gap*. Istnieje wiele definicji opisujących ten termin, jednak w większości z nich pojawia się wspólny element, który opisuje ją jako miarę nierówności w płacach kobiet i mężczyzn. Istnieją też inne definicje, jak np. Parlamentu Europejskiego (2023), zgodnie z którą luka płacowa ze względu na płeć to „różnica pomiędzy średnimi stawkami godzinowymi brutto, które otrzymują kobiety i mężczyźni” lub definicja GUS (2017, s. 6), według której luka płacowa to „najczęściej stosowany wskaźnik do porównań wynagrodzeń między poszczególnymi grupami osób”.

Chociaż nie występuje osobna definicja dla luki płacowej w sporcie, jest to całkowicie inne zjawisko niż luka obserwowana w sektorze biznesowym. Zjawisko to jest zauważalne w każdym aspekcie rywalizacji sportowej, szczególnie na jej najwyższym szczeblu. Różnice płacowe w sporcie bywają nieporównywalnie większe niż znane z ogólnych raportów badających lukę płacową w biznesie (Adelphi University, 2023; Parlament Europejski, 2023). Znacznie różnią się również przyczyny występowania luki płacowej między mężczyznami i kobietami o podobnych umiejętnościach.

Sytuację kobiet i mężczyzn w wybranych sportach przedstawiono w opracowaniu Adelphi University (2023). W tabeli 1 zaprezentowano średnie zarobki dla kobiet i mężczyzn w czterech dyscyplinach. Dla każdego z tych sportów analizowano zarob-

ki w dwóch ligach rozgrywek na najwyższym poziomie, po jednej dla kobiet i mężczyzn. Nazwy lig podano w nawiasach po nazwie dyscypliny, najpierw liga męska, potem liga kobieca. Wszystkie kwoty podano w dolarach. Dodatkowo przeliczono iloraz wypłat kobiet i mężczyzn, by pokazać jaką część średnich zarobków mężczyzn zarabiają średnio kobiety.

Tabela 1. Średnie zarobki kobiet i mężczyzn w wybranych dyscyplinach sportowych

Dyscyplina	Średnie roczne zarobki mężczyzn [w dolarach]	Średnie roczne zarobki kobiet [w dolarach]	Iloraz zarobków kobiet przez zarobki mężczyzn
Koszykówka (NBA i WNBA)	10 776 383	113 295	1,05%
Piłka Nożna (MLS i NWSL)	471 279	54 000	11,46%
Tenis (100 najlepszych w ATP i WTA)	1 589 024	1 039 141	65,39%
Golf (PGA i LPGA)	1 042 917	346 360	33,21%

Źródło: opracowanie własne na podstawie (Adelphi University, 2023).

Jak wynika z tab. 1, we wszystkich analizowanych dyscyplinach mężczyźni zarabiają więcej niż kobiety i – niestety – różnica ta jest zazwyczaj znacząca. Najlepszą sytuację dla kobiet można zaobserwować w tenisie ziemnym, gdzie zarabiają one średnio około 65% średnich zarobków mężczyzn. Najgorzej natomiast stosunek ten wygląda w koszykówce, gdzie kobiety otrzymują średnio tylko 1% przeciętnych wypłat mężczyzn. Różnica ta jest ogromna i pokazuje, z jak poważną dyskryminacją płacową muszą się zmagać kobiety.

3. Wprowadzenie do tematu badania i opis wykorzystanych metod

Badanie zdecydowano się przeprowadzić na podstawie danych z amerykańskich lig koszykarskich. Dyscyplina ta jest dobrym obiektem badawczym, gdyż podobnie jak w przypadku innych opisywanych wcześniej dyscyplin, w koszykówkę również początkowo mogli grać jedynie mężczyźni. National Basketball Association (NBA), czyli amerykańsko-kanadyjska męska liga koszykarska, powstała już w 1946 roku (NBA, b.d.a), podczas gdy Women's National Basketball Association (WNBA) założono dopiero w 1996 roku (WNBA, b.d.a). Te dane wskazują, że dyskryminacja ze względu na płeć jest obecna w tym sporcie od wielu lat.

Dodatkowo popularność oraz prestiż tego sportu w Stanach Zjednoczonych przyciągają nie tylko najlepszych zawodników na świecie, ale również sponsorów (KAF, 2022), którzy są w stanie przekazać duże kwoty, by uzyskać rozgłos wśród fanów tej dyscypliny, szczególnie męskich rozgrywek. Niestety, jak pokazują przedstawione we wcześniejszej części pracy badania (tab. 1), pomimo tak dużego zainteresowania, koszykówka w Stanach Zjednoczonych wciąż jest mocno dotknięta problemem luki płacowej.

3.1. Opis wykorzystanych metod

3.1.1. Analiza głównych składowych

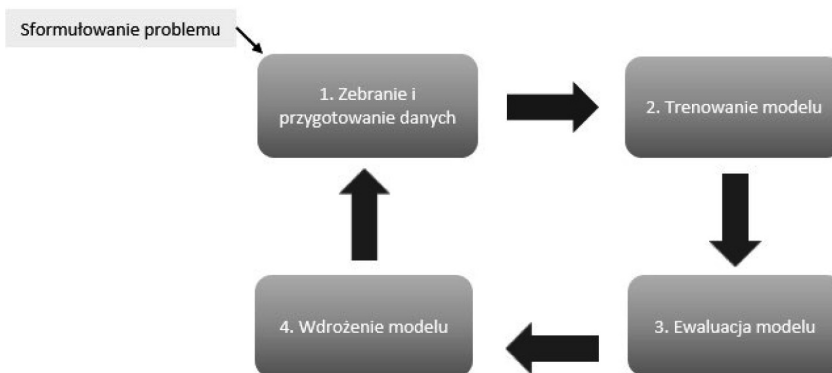
Analiza głównych składowych to jeden z algorytmów stosowanych do redukcji wymiarów, który polega na rzutowaniu danych na przestrzeni o mniejszej liczbie wymiarów przy jednoczesnym zachowaniu struktury danych w jak największym stopniu (Mamczur, 2019). Metoda ta polega na „transformacji zmiennych pierwotnych w zbiór nowych nieskorelowanych zmiennych zwanych głównymi składowymi” (Sztemberg-Lewandowska, 2017), gdzie każda z nich tworzona jest przez kombinację liniową zmiennych pierwotnych w taki sposób, by maksymalizować zmienność niewyjaśnioną przez poprzednią składową.

Dla każdej składowej można wyznaczyć wartość własną, która informuje, jaką część całkowitej zmienności wyjaśnia dana składowa. Najwięcej wariacji przedstawia zawsze pierwsza składowa, a każda następna tłumaczy niewyjaśnioną jeszcze część zmienności. Zsumowane wartości dla wszystkich składowych przedstawiają całkowitą wariację (PQStat, b.d.).

Analizę głównych składowych stosuje się przede wszystkim w celu redukcji liczby zmiennych, co pozwala zarówno na wykrycie ewentualnych prawidłowości, jak i wizualizację początkowo wielowymiarowych zbiorów danych.

3.1.2. Algorytm uczenia maszynowego

Uczenie maszynowe (*machine learning*) odnosi się do szerokiego zestawu aplikacji AI, w których komputery budują modele na podstawie wzorców rozpoznanych w zbiorach danych i wykorzystują te modele do generowania hipotez na temat świata. Takie modele mają niezliczone zastosowania w programowaniu do rozwiązywania problemów zarówno klasyfikacyjnych jak i regresyjnych (Russell i Norvig, 2020).



Rys. 1. Uproszczony przepływ implementacji modelu uczenia maszynowego

Źródło: opracowanie własne na podstawie (Sundberg i Holmström, 2024).

Stosowanie metod uczenia maszynowego wymaga ścisłego trzymania się narzuconego schematu, którego uproszczony przepływ przedstawiono na rys. 1. Widać na nim, że by osiągnąć zadowalające rozwiązanie, najczęściej trzeba przejść kilka iteracji przedstawionego procesu. Modelowanie powinno składać się z czterech etapów: zbierania i przygotowania danych, trenowania modelu, ewaluacji uzyskanego modelu oraz wdrożenia modelu do użytkowania.

Przygotowanie danych

Na etapie zbierania danych niezbędne jest wskazanie, co jest problemem, który ma być rozwiązany za pomocą modelowania. Do uzyskania dobrego jakościowo rozwiązania powinno się też prawidłowo wskazać zmienne endo- i egzogeniczne.

Część metod uczenia maszynowego wymaga również normalizacji danych. Wykorzystywana w tym badaniu metoda XGBoost oparta na drzewach decyzyjnych jest nieczuła na różnice wielkości między zmiennymi, dlatego przeprowadzanie normalizacji danych jest niepotrzebne (Filho, 2022).

Podział na zbiór uczący i testowy

Zbiór uczący wykorzystywany jest do trenowania modelu – to na podstawie tych danych wykrywane są zależności (Blog Statystyczny, 2020). Zbiór ten jest zazwyczaj najbardziej liczny.

Dane ze zbioru testowego są, jak wskazuje na to nazwa, wykorzystywane do testowania modelu. Obserwacje z tego zbioru służą do symulacji działania modelu na nowych niewykorzystywanych wcześniej danych.

Trenowanie modelu – XGBoost

W badaniu wykorzystano metodę wzmacniania gradientowego – model XGBoost. Model ten jest techniką uczenia zespołowego (*ensemble learning*), która polega na łączeniu przewidywań grup predyktorów, czyli wspólnym analizowaniu wyników wielu mniejszych modeli. Takie podejście pozwala na znaczne poprawienie wyników predykcji, istotnie zmniejsza się również wariancja głównego modelu (Data Science Team, 2020). XGBoost jest modelem sekwencyjnym, czyli modele bazowe są tworzone po kolei, a każdy następny zależy od wyników uzyskanych dla poprzedniego.

Wszystkie modele wzmacniania gradientowego, należące do uczenia zespołowego, opierają się na trzech krokach (Analytics Vidhya, 2024):

- 1) utworzenie modelu F0 do przewidywania wartości zmiennej objaśnianej; wyznaczenie reszt modelu;
- 2) utworzenie nowego modelu h1, który dopasowywany jest do reszt uzyskanych dla modelu F0;
- 3) łączenie wniosków uzyskanych dla F0 i h1, by stworzyć model F1 – poprawiona wersja F0.

Kroki te powtarza się do uzyskania minimalnej wartości wybranego wskaźnika oceny jakości modelu (takim wskaźnikiem może być MSE) lub po przejściu wskazanej liczby iteracji.

Choć XGBoost to jeden z algorytmów opartych na wzmacnianiu gradientu, znacznie różni się od standardowych modeli z tej grupy. XGBoost jest lepszy od zwykłych algorytmów wzmacniania gradientu, gdyż zamiast trenować najlepszy model na danych (jak w przypadku tradycyjnych metod), trenuje się tysiące modeli na różnych podzbiorach zbioru uczącego, a następnie głosuje się na model o najlepszej wydajności.

GridSearchCV

GridSearchCV to proces dostrajania hiperparametrów w celu określenia optymalnych wartości dla danego modelu (Great Learning Team, 2024a). Metoda ta opiera się na zautomatyzowanym testowaniu modeli ze wszystkich zaproponowanych możliwych wartości hiperparametrów. W ten sposób wyznacza się hiperparametry dla wskazanego modelu, przy których model wykazuje najlepsze wartości wybranego wskaźnika oceniającego jakość. W tej metodzie wykorzystywana jest tak zwana walidacja krzyżowa, o której przeczytać można w Great Learning Team (2024b).

Ocena jakości

W tym etapie modelowania wykorzystuje się np.: MAE, MSE, RMSE, a także współczynnik determinacji.

Skrótu RMSE (*Root Mean Squared Error*) używa się do określenia pierwiastka średniokwadratowego błędu (Hodson, 2022 oraz Chai i Draxler, 2014):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

gdzie: n – liczba obserwacji, y_i – rzeczywista wartość zmiennej objaśnianej dla i -tego obiektu, \hat{y}_i – przewidywana wartość zmiennej objaśnianej dla i -tego obiektu.

Wartość tego wskaźnika jest przedstawiana w tej samej jednostce co zmienna objaśniana, co znacznie ułatwia interpretację. Wskaźnik ten interpretuje się jako średnią oczekiwaną różnicę między wartością przewidywaną a rzeczywistą.

MAE (*Mean Absolute Error*) to wskaźnik, który oblicza się ze wzoru (Hodson, 2022; Willmott i Matsuura, 2005):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

gdzie: n – liczba obserwacji, y_i – rzeczywista wartość zmiennej objaśnianej dla i -tego obiektu, \hat{y}_i – przewidywana wartość zmiennej objaśnianej dla i -tego obiektu.

Współczynnik jest obliczany przez wyznaczenie wartości bezwzględnych błędów predykcji, co zapobiega wzajemnemu znoszeniu się błędów z dodatnim i ujemnym znakiem. Wartość tego wskaźnika to średnia wszystkich bezwzględnych błędów predykcji. Wskaźnik ten też jest przedstawiany w tej samej jednostce co zmienna objaśniana. Wartość tego wskaźnika interpretuje się jako średnią błędów predykcji.

Współczynnik determinacji R^2 wskazuje, jaka część zmienności zmiennej objaśnianej jest wyjaśniana przez model (Walesiak, 1996). Jest on obliczany zgodnie ze wzorem (Ruijie i in., 2023):

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2},$$

gdzie: y_i – rzeczywista wartość zmiennej objaśnianej dla i -tego obiektu, \hat{y}_i – przewidywana wartość zmiennej objaśnianej dla i -tego obiektu.

Współczynnik ten przyjmuje wartości z przedziału $[0,1]$. Im wartość bliższa jedności, tym lepsze dopasowanie funkcji regresji do rzeczywistych danych. Im wyższa wartość tego współczynnika, tym większa część zmienności zmiennej objaśnianej jest wyjaśniana przez model.

Istnieje jeszcze wiele innych, mniej popularnych metod ewaluacji modelu, o których przeczytać można w Hodson (2022).

W badaniu wykorzystano do oceny jakości: MAE, MSE, RMSE, a także współczynnik determinacji. RMSE (*Root Mean Squared Error*) to pierwiastek średniokwadratowy błędu (Chai i Draxler, 2014; Hodson, 2022). Wskaźnik ten interpretuje się jako średnią oczekiwaną różnicę między wartością przewidywaną a rzeczywistą.

MAE (*Mean Absolute Error*) to wskaźnik obliczany przez wyznaczenie wartości bezwzględnych błędów predykcji, co zapobiega wzajemnemu znoszeniu się błędów z dodatnim i ujemnym znakiem (Hodson, 2022; Willmott i Matsuura, 2005). Wartość tego wskaźnika interpretuje się jako średnią błędów predykcji.

Ocena wpływu i znaczenia cech

Ocenę wpływu i znaczenia cech przeprowadzono z wykorzystaniem trzech metod: metody wbudowanej w bibliotekę *sci-kit learn*, metody permutacyjnej oraz metody SHAP.

Metoda wbudowana w bibliotekę *sci-kit learn*

Metoda wbudowana w bibliotekę *sci-kit learn* dla metod opartych na drzewach decyzyjnych bazuje na zmianach wartości wybranego wskaźnika nierównomierności,

na przykład miary Giniego. Za pomocą tej metody można określić, które ze zmiennych najmocniej wpływają na zmniejszenie wartości wskazanego wskaźnika.

Wadami tej metody są: zawyżanie ważności zmiennych liczbowych i przeprowadzanie obliczeń na danych treningowych. Przez to zmienne, które nie wpływają na przewidywanie zmiennej objaśnianej, mogą zostać wskazane jako ważne (Sci-kit learn, b.d.).

Metoda permutacyjna

W odróżnieniu od metody wbudowanej metoda permutacyjna może być stosowana dla większości modeli uczenia maszynowego. W metodzie tej bada się, jak zmienia się wartość miary błędu (takiej jak MAE czy R^2) po spermutowaniu wartości jednej ze zmiennych. Tę czynność wykonuje się dla wszystkich zmiennych z osobna, a następnie ranguje się je po wpływie, jaki miały na zmianę wartości wskazanej miary błędu (Jensen, 2022).

SHAP

W tej metodzie siła wpływu poszczególnych zmiennych jest prezentowana przez wyznaczenie wartości Shapleya (1953). Za pomocą wartości SHAP można wyjaśniać wpływ poszczególnych zmiennych dla dowolnego modelu uczenia maszynowego. Przy obliczaniu tych wartości wykorzystuje się podejście oparte na teorii gier, które mierzy wkład każdego gracza w końcowy wynik. W podejściu stosowanym w uczeniu maszynowym każdej funkcji przypisuje się wartość ważności reprezentującą jej udział w wynikach modelu (All Awan, 2023). Na podstawie wartości SHAP można wyznaczyć interpretowalność zarówno lokalną, jak i globalną.

4. Wyniki badań

4.1. Zebranie i przygotowanie danych

Badaniu poddano koszykarzy grających w lidze NBA w sezonie 2022-2023 oraz koszykarki występujące w WNBA w sezonie 2023. Wybrano te sezony, by zachować aktualność wyników uzyskanych w badaniu, gdyż na moment przeprowadzania analizy są to ostatnie zakończone rozgrywki.

Do badania wybrano 3 rodzaje zmiennych: opisujące podstawowe charakterystyki zawodników, ich popularność oraz skuteczność ofensywną i defensywną. Są to zmienne:

- Plec – płeć zawodnika,
- Wiek – wiek gracza w badanym sezonie,
- Wypłata – wypłata bazowa zawodnika na badany sezon w przeliczeniu na jeden mecz w podstawowej części sezonu; jest to zmienna objaśniana,

OBS	–	liczba obserwujących na Instagramie na dzień 4 kwietnia 2024,
Min	–	średnia liczba rozegranych minut na mecz,
PTS	–	średnia liczba punktów zdobyta na mecz,
FGM	–	średnia liczba trafionych rzutów z gry na mecz,
FG%	–	średnia skuteczność prób zdobycia punktów z gry na mecz,
3PM	–	średnia liczba trafionych rzutów za trzy punkty na mecz,
3P%	–	średnia skuteczność prób rzutów za trzy punkty na mecz,
FTM	–	średnia liczba trafionych rzutów wolnych na mecz,
FT%	–	średnia skuteczność prób rzutów wolnych na mecz,
OREB	–	średnia liczba zbiórek ofensywnych na mecz,
DREB	–	średnia liczba zbiórek defensywnych na mecz,
AST	–	średnia liczba asyst na mecz,
TOV	–	średnia liczba strat na mecz,
STL	–	średnia liczba przechwytyń na mecz,
BLK	–	średnia liczba bloków na mecz,
PF	–	średnia liczba popełnionych fauli na mecz,
DD2	–	średnia liczba tak zwanych <i>double doubles</i> na mecz,
TD3	–	średnia liczba tak zwanych <i>triple doubles</i> na mecz,
GPR	–	stosunek liczby zagranych meczy do liczby wszystkich meczy w sezonie,
GWR	–	stosunek liczby wygranych meczy do liczby wszystkich zagranych meczy.

Powyższe zmienne zaczerpnięto z następujących portali: Spotrac (b.d.a; b.d.b) popularbasketbal-lers.com (2024); Instagram (2024); NBA (b.d.b) oraz WNBA (b.d.b).

Niezbędne było zapewnienie porównywalności wartości zmiennych dla kobiet i mężczyzn, gdyż większość ze zmiennych była zależna od liczby meczów w sezonie. Podstawowe rozgrywki sezonu NBA składały się z 82 spotkań, natomiast w WNBA było to jedynie 40 spotkań, dlatego zdecydowano się przedstawić każdą ze zmiennych w przeliczeniu na 1 mecz.

Wszystkie obliczenia oraz część wizualizacji wykonano w języku Python. Pozostałe wizualizacje wykonano w Excelu.

4.2. Dobór zmiennych

Doboru zmiennych zdecydowano się dokonać na podstawie korelacji Pearsona oraz współczynnika zmienności. Postanowiono, że zmienne objaśniające nie powinny być ze sobą bardziej skorelowane niż 0,8.

Wystąpiły zmienne mocno ze sobą skorelowane, z tego powodu usunięto ze zbioru potencjalnych zmiennych objaśniających zmienne PTS, FTM oraz powiązaną z nią FT%.

Zdecydowano, że do badania mogą zostać wybrane jedynie zmienne, których zmienność jest wyższa niż 10%. Wszystkie zaproponowane zmienne miały wartość współczynnika wyższą niż założone 10%, dlatego wszystkie ze zmiennych brano pod uwagę w kolejnych etapach analizy.

4.3. Wyniki XGBoost

Modelowanie przeprowadzono z wykorzystaniem metody XGBoost, gdyż jest to model oparty na drzewach decyzyjnych, który jest nieczuły na różnicę wielkości zmiennych. Na metodach opartych na drzewach możliwe jest również przeprowadzanie analizy wpływu zmiennych na predykcję, co jest niezbędne do sprawdzenia zakładanych w tym badaniu hipotez. Zaletą tej metody jest również dobre dopasowywanie się do danych. W wyborze modelu oceniano również łatwość implementacji. XGBoost jest dostępny w wielu popularnych, dobrze udokumentowanych narzędziach.

Do utworzenia modelu wykorzystano bibliotekę XGBoost dostępną w Pythonie. Chociaż rozwiązanie to umożliwiło implementację modelu z domyślnymi wartościami hiperparametrów, zdecydowano się dokonać strojenia hiperparametrów. Dokonano tego z wykorzystaniem metody przeszukiwania siatki (z angielskiego *Grid Search*). Jako minimalizowaną miarę oceny jakości modelu wybrano średni błąd kwadratowy MSE. Strojeniu poddano parametry:

- `colsample_bytree` – proporcja kolumn używanych do uczenia każdego drzewa,
- `learning_rate` – szybkość uczenia,
- `max_depth` – maksymalna głębokość pojedynczego drzewa,
- `n_estimators` – liczba drzew,
- `subsample` – proporcja próbek używanych do uczenia każdego drzewa.

Strojenie wykazało, że najlepsze wyniki można uzyskać dla modelu o hiperparametrach:

- `colsample_bytree = 0.9`,
- `learning_rate = 0.2`,
- `max_depth = 3`,
- `n_estimators = 100`,
- `subsample = 1.0`

Wartości powyżej zostały wskazane jako optymalne, dlatego wytrenowano model właśnie z tymi hiperparametrami.

4.4. Ocena jakości

Dla wytrenowanego modelu uzyskano błąd MSE w wysokości 4640102639,85, natomiast RMSE będące pierwiastkiem tej wartości wyniosło 68118,3. Oznacza to, że średnio przewidywana wartość wypłaty różni się od rzeczywistej o 68118,3 dolarów. Wartość ta jest równa niemal 74% wartości średniej zmiennej objaśnianej. Przedstawione powyżej wskaźniki mają stosunkowo wysokie wartości, co wskazuje, że model ten raczej słabo przewiduje wartość wypłaty dla poszczególnych zawodników i zawodniczek.

MAE dla tego modelu wyniosło 43394,6. Oznacza to, że średnia błędów predykcji jest równa 43394,6 dolarów. Wartość ta, choć przedstawia podobną charakterystykę do RMSE, znacznie się różni od wartości RMSE. Ta duża różnica sugeruje, że w zbiorze danych występują obserwacje, które znacząco różnią się między sobą.

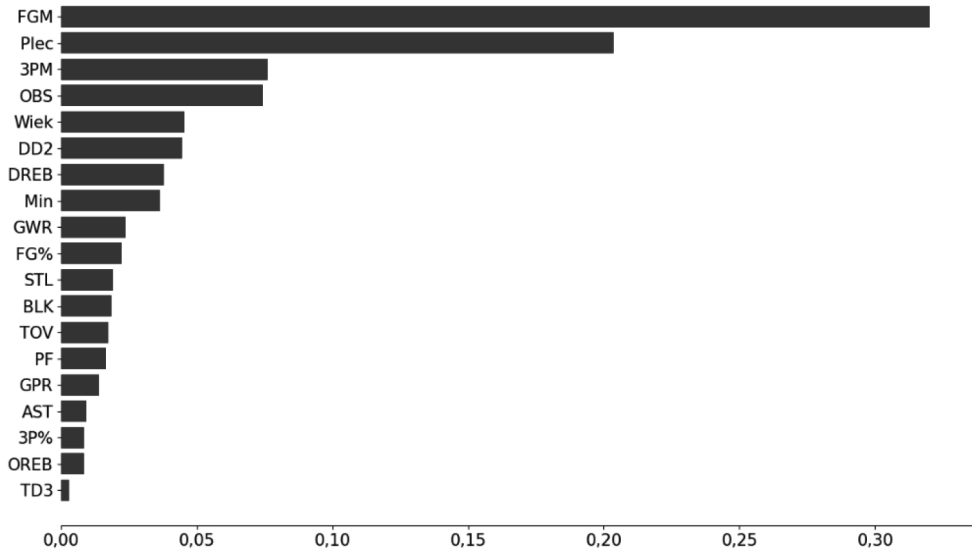
Współczynnik determinacji wyniósł 0,72, co wskazuje, że 72% wariacji jest wyjaśniana przez model. Biorąc pod uwagę, że zjawisko to jest niełatwe do przewidywania, a wysokość wypłaty często jest zależna od innych, nieuwzględnionych w tym badaniu czynników, można uznać, że zmienność tej zmiennej jest dosyć dobrze reprezentowana przez model. Trzeba pamiętać, że w rzeczywistości wysokość wypłaty jest ustalana najczęściej na podstawie osiągnięć zawodników z poprzednich lat, które nie były badane w tej analizie. Kontrakty są również ustalane przed rozpoczęciem sezonu i obowiązują kilka lat, przez co wysokość wypłaty może nie odzwierciedlać obecnej formy zawodnika. Wysokość wypłaty zapisanej w umowie między zespołem a zawodnikiem niejednokrotnie zależy też od umiejętności negocjacyjnych czy sympatii międzyludzkiej. Wszystkie te czynniki sprawiają, że zmienną Wypłata trudno jest przewidzieć za pomocą metod statystycznych, więc uzyskane dla tego modelu wyniki można uznać za zadowalające.

Wyjaśnialność globalna

By wykręć globalny wpływ poszczególnych zmiennych na predykcję modelu, zdecydowano się porównać 3 różne sposoby wyznaczenia znaczenia cech: sposób wbudowany w bibliotekę *sci-kit learn* oparty na zmianach wskaźnika nierównomierności, metodę permutacyjną oraz metodę opartą na współczynniku Shapleya.

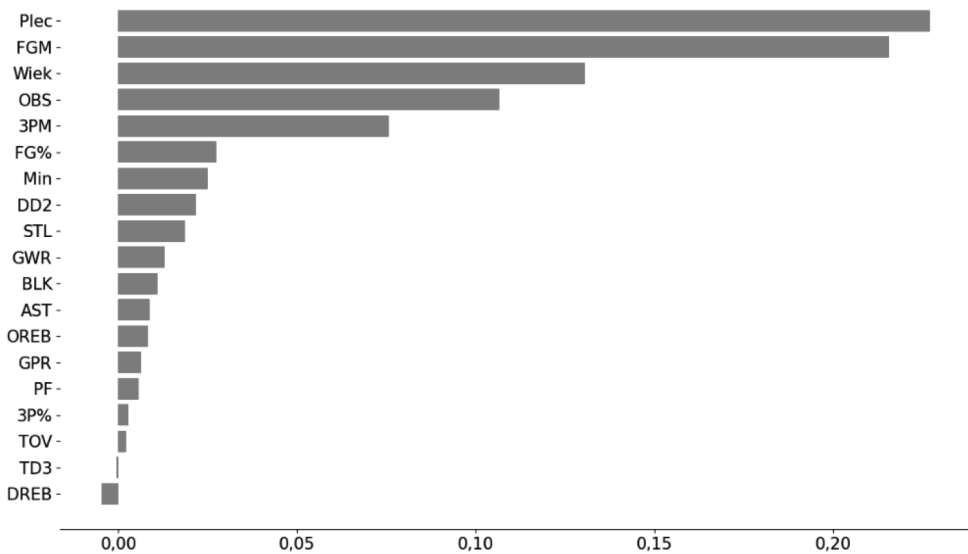
Analizując powyższe wykresy można zauważyć, że pierwsze 5 cech we wszystkich metodach jest takie same, różnią się jedynie kolejnością (rys. 2-4). Do tych cech należą: Plec, FGM (średnia liczba zdobytych punktów na mecz), Wiek, 3PM (średnia liczba udanych rzutów za 3 punkty na mecz) oraz OBS (liczba obserwujących na Instagramie). Dwie spośród tych cech są wskazywane jako ważniejsze od pozostałych: Plec i FGM. Pozostałe wybrane zmienne nie wpływają znacząco na wysokość wypłaty.

Na wykresie (rys. 5) przedstawiono wpływ poszczególnych zmiennych na wartości zmiennej objaśnianej dla każdej z obserwacji. Cechy są przedstawione osobno wierszami. Dla łatwiejszego odbioru analizy zmienne posortowano od najbardziej wpływowej do najmniej znaczącej. Kolorami zaznaczono wartości dla każdej zmiennej indywidualnie, na niebiesko zaznaczono niskie wartości każdej ze zmiennych, na różowo wysokie. Im obserwacja położona jest bliżej lewej strony wykresu, tym niższa wartość Shapleya, czyli tym bardziej wpływa ona na obniżenie wypłaty, im bardziej na prawo – tym wyższa wypłata. Jak widać na wykresie, najbardziej wpływową zmienną jest ponownie płęć. Wysokim wartościom tej zmiennej (w tym przypadku jest to wartość 1 odpowiadająca kobietom) odpowiadają niskie wartości wypłaty, co oznacza, że kobiety otrzymują niższe wynagrodzenie niż mężczyźni. Kolejną wpływową zmienną jest FGM. Wysokie wartości tej zmiennej odpowiadają wysokim i bardzo wysokim wypłatom, co sugeruje, że żeby otrzymywać wysoką wypłatę, trzeba też zdobywać dużą liczbę punktów. Wiersz dla zmiennej OBS pokazuje, że duża liczba obserwujących na Instagramie przekłada się na wyższe wynagrodzenie. Można jednak zauważyć, że niskie wartości dla tej zmiennej również



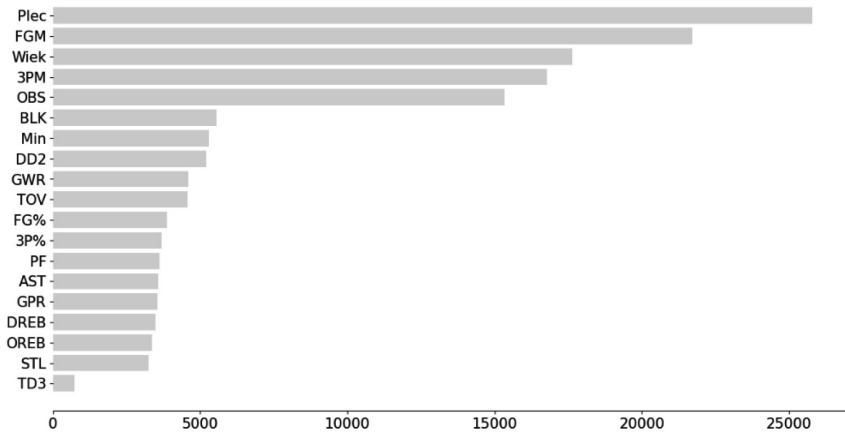
Rys. 2. Wpływ zmiennych na predykcję modelu – metoda wbudowana w bibliotekę *sci-kit learn*

Źródło: opracowanie własne.



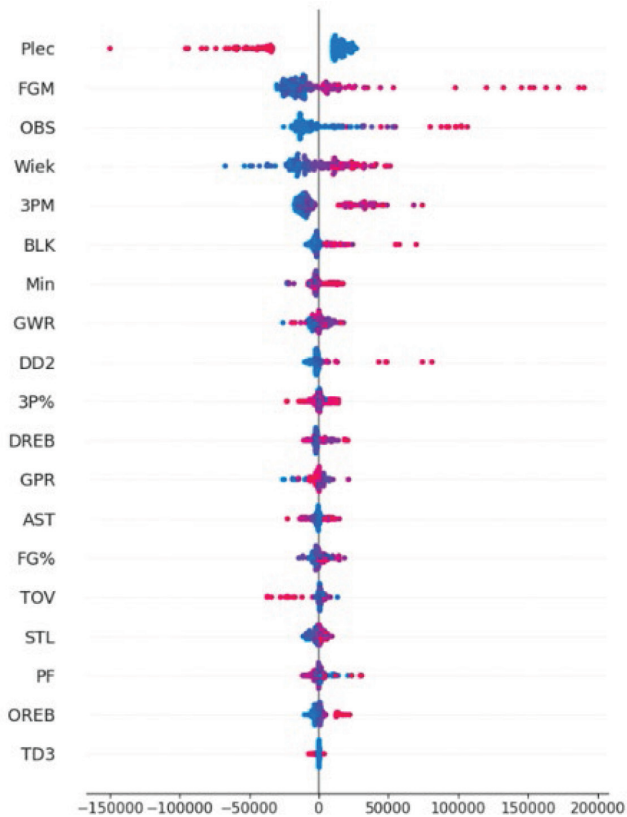
Rys. 3. Wpływ zmiennych na predykcję modelu – metoda permutacyjna

Źródło: opracowanie własne.



Rys. 4. Wpływ zmiennych na predykcję modelu – współczynnik Shapleya

Źródło: opracowanie własne.



Rys. 5. Wpływ zmiennych na wartość predykcji w podziale na obserwacje

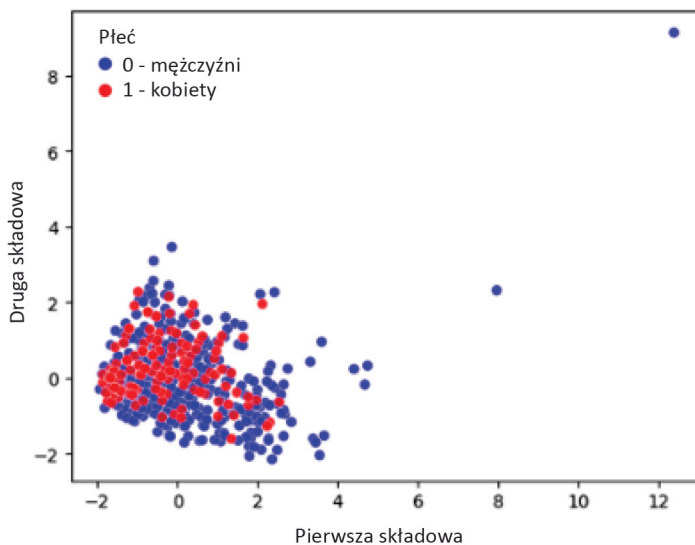
Źródło: opracowanie własne.

mogą oznaczać wynagrodzenie ponad średnią. Analizując wiersz dla wieku, zauważono, że występuje wyraźna tendencja, że im starszy zawodnik, tym wyższą wypłatę otrzymuje. Wiersz dla zmiennej 3PM pokazuje, że większość obserwacji dla tej zmiennej ma niskie wartości, które odpowiadają niskim wypłatom dla tych zawodników. Sportowcy, którzy podczas meczu średnio trafiają więcej za trzy punkty, otrzymują zazwyczaj wyższe wynagrodzenie. Pozostałe zmienne zostały wskazane przez poprzednie analizy jako mniej istotne, można jednak zauważyć tendencję, że zawodnicy, którzy przyjmują dla tych zmiennych wartości uznawane za korzystne i są one znacząco różne od średniej, najczęściej również otrzymują wyższą wypłatę.

Podsumowując wyniki analizy wpływu poszczególnych zmiennych, można zauważyć, że wszystkie 3 badane charakterystyki mają wpływ na wypłatę zawodnika: jego cechy personalne (zmienna Plec i Wiek), popularność (OBS) oraz wyniki uzyskiwane na boisku w kategorii statystyk ofensywnych (FGM i 3PM).

4.5. Szczegółowa analiza zależności

Zdecydowano się zobrazować obserwacje w przestrzeni dwuwymiarowej z uwzględnieniem jedynie zmiennych, które miały największy wpływ na predykcję modelu. W tym celu wykorzystano metodę analizy głównych składowych. Porównano, jak w grupach kobiet i mężczyzn ułożone są punkty ze względu na zmienne Wiek, OBS, FGM i 3PM. Takie przedstawienie pozwala określić, czy wartości tych zmiennych są znacząco różne między kobietami i mężczyznami.



Rys. 6. Wizualizacja dwóch pierwszych składowych głównych

Źródło: opracowanie własne.

Na rysunku 6 przedstawiono dwie pierwsze składowe dla najbardziej wpływowych zmiennych. Wyjaśniają one 70% zmienności zmiennej objaśnianej. Na czerwono zaznaczono obserwacje kobiet, a na niebiesko mężczyzn. Można zauważyć, że występuje kilka obserwacji odstających, z czego dwie z nich są wartościami ekstremalnymi. Większość tych obserwacji dotyczy mężczyzn, a tylko jedna kobiet. Pomijając wartości odstające, należy stwierdzić, że niewidoczna jest różnica między grupami kobiet i mężczyzn. Pokazuje to, że przedstawiciele obu płci nie różnią się znacząco między sobą ze względu na analizowane zmienne. Pozwala to na wysunięcie wniosku, że tak ogromna różnica w zarobkach między kobietami i mężczyznami nie wynika z ich skuteczności na boisku czy popularności w mediach społecznościowych, a właśnie jedynie z innej płci.

5. Zakończenie

Celem zaprezentowanego badania było udowodnienie, że w koszykówce występuje dyskryminacja. W tym celu stworzono model regresyjny, który w dobry sposób przewidywałby na podstawie dostarczonych danych, jakiej wysokości wypłatę powinien otrzymywać zawodnik z amerykańskich lig koszykarskich. Na podstawie wpływu zmiennych na predykcję otrzymanego modelu udowodniono, że to, jakiej płci jest badana osoba, szczególnie mocno wpływa na wysokość wypłaty. Jak wykazały szczegółowe badania, fakt, że badana obserwacja jest charakterystyką kobiety, silnie wpływa na obniżenie przewidywanej wartości. Zauważono również, że kobiety, które prezentowały wartości poszczególnych zmiennych na podobnym poziomie co mężczyźni, otrzymywały nieporównywalnie niższe wynagrodzenie. Oznacza to, że kobiety nie otrzymują odpowiedniego wynagrodzenia za wykonywaną pracę.

W badaniu wykazano, że ze względu na analizowane zmienne występuje dyskryminacja płacowa w amerykańskich ligach koszykarskich. Należy jednak pamiętać, że choć wyniki badania są jednoznaczne, mogą wystąpić czynniki, które nie zostały uwzględnione w analizie, a które mocno wpływają na to zróżnicowanie.

Wszystkie trzy badane aspekty, odnośnie do których podejrzewano, że mogą mieć wpływ na wysokość wypłaty, czyli charakterystyki zawodników, ich popularność oraz skuteczność na boisku, miały istotny wpływ na przewidywaną wartość płacy. Nieistotne okazały się jedynie zmienne, które reprezentowały skuteczność zawodników w obronie. Może to wynikać z faktu, że gracze defensywni są uznawani za mniej efektywnych.

Literatura

- Adelphi University. (2023). *Male vs Female Professional Sports Salary Comparison*. Pobrano 13 marca 2024 z <https://online.adelphi.edu/articles/male-female-sports-salary/>
- All Awan, A. (2023). *An Introduction to SHAP Values and Machine Learning Interpretability*. Datacamp. Pobrano 5 maja 2024 z <https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability>

- Analytics Vidhya. (2024). *Introduction to XGBoost Algorithm in Machine Learning*. Pobrano 5 maja 2024 z <https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/>
- Blog Statystyczny. (2020). *Grupa ucząca, walidacyjna i testowa*. Pobrano 22 kwietnia 2024 z <https://statystyczny.pl/grupa-uczaca-walidacyjna-i-testowa/>
- Chai, T. i Draxler, R. R. (2014). Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments against Avoiding RMSE in the Literature. *Geosci. Model Dev.* 7(3), 1247-1250, <https://doi.org/10.5194/gmd-7-1247-2014>
- Data Science Team. (2020). *XGBoost*. Pobrano 3 maja 2024 z <https://datascience.eu/pl/programowanie-komputerowe/xgboost/>
- Filho, M. (2022). *Does XGBoost Need Feature Scaling Or Normalization?* ? Forecastegy. Pobrano 29 kwietnia 2024 z <https://forecastegy.com/posts/does-xgboost-need-feature-scaling-or-normalization/>
- Gratton, C. (1998). The Economic Importance of Modern Sport. *Culture, Sport, Society*, 1(1), 101-117. <https://doi.org/10.1080/14610989808721803>
- Great Learning Team. (2024a). *Hyperparameter Tuning with GridSearchCV*. Pobrano 3 maja 2024 z <https://www.mygreatlearning.com/blog/gridsearchcv/>
- Great Learning Team. (2024b). *What is Cross Validation in Machine Learning? Types of Cross Validation*. Pobrano 3 maja 2024 z <https://www.mygreatlearning.com/blog/cross-validation/>
- GUS. (2017). *Wskaźniki jakości pracy*. Pobrano 2 lutego 2024 z https://stat.gov.pl/files/gfx/portalinformacyjny/pl/defaultaktualnosci/5821/10/2/1/wskazniki_jakosci_pracy_21_03_2018_pl.pdf
- Hodson, T. O. (2022). Root-mean-square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geoscientific Model Development*, 15(14), 5481-5487. DOI: <https://doi.org/10.5194/gmd-15-5481-2022>
- Instagram. (2024, 4 kwietnia). *Followers*. Pobrano 4 kwietnia 2024 z <https://www.instagram.com/>
- Jensen, T. (2022). *Feature Importance for Any Model using Permutation*. Medium. Pobrano 3 maja 2024 z https://medium.com/@T_Jen/feature-importance-for-any-model-using-permutation-7997b7287aa
- KAF. (2022). *The World's Top Basketball Leagues*. Medium. Pobrano 4 kwietnia 2024 z <https://medium.com/krause-house-dao/the-worlds-top-basketball-leagues-3b634c5e426c>
- Mamczur, M. (2019). *Na czym polega analiza składowych głównych (PCA)?* Pobrano 21 kwietnia 2024 z <https://miroslawmamczur.pl/na-czym-polega-analiza-skladowych-glownych-pca/>
- NBA. (b.d.a). *NBA Advanced Stats*. Pobrano 4 kwietnia 2024 z <https://www.nba.com/stats/players/traditional?PerMode=PerGame&sort=PTS&dir=-1&Season=2022-23>
- NBA. (b.d.b). *NBA History*. Pobrano 4 kwietnia 2024 z <https://www.nba.com/history>
- Parlament Europejski. (2023, 5 kwietnia). *Luka płacowa między kobietami a mężczyznami: definicja i przyczyny*. Pobrano 4 lutego 2024 z <https://www.europarl.europa.eu/news/pl/headlines/society/20200109STO69925/luka-placowa-miedzy-kobietami-a-mezczyznami-definicja-i-przyczyny>
- Popularbasketballers.com. (2024, 4 kwietnia). *NBA Players Ranked by Instagram Followers*. Pobrano 4 kwietnia 2024 z <https://www.popularbasketballers.com/>
- PQStat. (b.d.). *Analiza składowych głównych*. Pobrano 21 kwietnia 2024 z https://pqstat.pl/?mod_f=test_pca
- Ruijie, S., Xiangjie, L., Yanxia, W., Shiyin, Z., Xingcan, C. i Xuejing, L. (2023). Machine Learning Regression Algorithms to Predict Short Term Efficacy after Anti VEGF Treatment in Diabetic Macular Edema Based on Real World Data. *Scientific Reports*, 13(18746). DOI: <https://doi.org/10.1038/s41598-023-46021-2>
- Russell, S. i Norvig, P. (2020). *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson.

- Sci-kit learn. (b.d.). *Permutation Importance vs Random Forest Feature Importance (MDI)*. Pobrano 3 maja 2024 z https://scikit-learn.org/stable/auto_examples/inspection/plot_permutation_importance.html#sphx-glr-auto-examples-inspection-plot-permutation-importance-py
- Shapley, L. S. (1953). *A Value for n-Person Games. Contributions to the Theory of Games II* (s. 307-317). Princeton.
- Spotrac. (b.d.a). *NBA Player Earnings*. Pobrano 2 kwietnia 2024 z <https://www.spotrac.com/nba/rankings/2022-23/base/>
- Spotrac. (b.d.b). *WNBA Salary Rankings*. Pobrano 2 kwietnia 2024 z <https://www.spotrac.com/wnba/rankings/2023/base/>
- Sundberg, L. i Holmström, J. (2024). Teaching Tip. Using No-Code AI to Teach Machine Learning In Higher Education. *Journal of Information Systems Education*, 35(1), 56-66. <https://doi.org/10.62273/CYPL2902>
- Sztemberg-Lewandowska, M. (2017). Analiza niezależnych głównych składowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 468, 222-229. <https://doi.org/10.15611/pn.2017.468.23>
- Thakur, A. (2022). Gender Discrimination in Sports. *International Journal of Law Management & Humanities*, 5(3), 210-219. <https://doi.org/10.1000/IJLMH.113056>
- Tomaszewska, J. (2004). Dyskryminacja ze względu na płeć. *Dyskryminacja ze względu na płeć i jej przeciwdziałanie*. Pobrano 15 marca 2024 z https://streswpracy.pl/wp-content/uploads/2022/bzp/DYSKryminacja_ze_wzgledu_na_plec_podr.pdf
- Walesiak, M. (1996). *Metody analizy danych marketingowych*. Państwowe Wydawnictwo Naukowe.
- Waśkowski, Z. (2011). Integracyjna rola sportu we współczesnym świecie. *Ekonomiczne Problemy Usług*, (78), 23-32.
- Willmott, C. J. i Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) Over the root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.*, (30), 79-82, <https://doi.org/10.3354/cr030079>
- WNBA. (b.d.a). *History*. Pobrano 4 kwietnia 2024 z <https://www.wnba.com/history>
- WNBA. (b.d.b). *WNBA Stats*. Pobrano 4 kwietnia 2024 z <https://stats.wnba.com/players/traditional/?sort=PTS&dir=-1&Season=2023&SeasonType=Regular%20Season>

Diagnosis of Pay Inequality among Athletes Using Machine Learning Methods

Abstract: The main goal of the work is to prove the existence of a wage gap in sports using the XGBoost method based on the salaries of basketball players from the NBA and WNBA leagues. The truthfulness of the statement that the amount of pay is influenced by the physical characteristics, popularity, and effectiveness on the pitch in defense and attack of a player was also checked. One model was created, the results of which were considered satisfactory. The influence of individual variables was compared for this model using three different methods. Obtained results confirmed that, apart from defensive effectiveness, all of the assumed factors influencing the amount of pay are significant. The occurrence of gender-based pay discrimination was also confirmed.

Keywords: gender pay gap, sport, machine learning, XGBoost

Agnieszka Krajnik

e-mail: 181792@student.ue.wroc.pl

ORCID: 0009-0000-2628-2331

Uniwersytet Ekonomiczny we Wrocławiu

Analiza relacji między stylem życia Europejczyków i czynnikami ekonomicznymi

DOI: 10.15611/2024.76.5.03

JEL: C38, E24, H51, I3, I12

© 2024 Agnieszka Krajnik

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na

Cytuj jako: Krajnik, A. (2024). Analiza relacji między stylem życia Europejczyków i czynnikami ekonomicznymi. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 42-56). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Praca dotyczy zagadnienia zdrowego stylu życia w kontekście sytuacji ekonomicznej w krajach europejskich. Celem artykułu jest sprawdzenie zależności między prowadzeniem zdrowego stylu życia a czynnikami ekonomicznymi i wydatkami na ochronę zdrowia w państwach Unii Europejskiej. W pierwszej części artykułu omówiono elementy składające się na zdrowy styl życia i podjęto temat kosztów związanych z prowadzeniem zdrowego oraz niezdrowego stylu życia. W drugiej części opisano badanie polegające na utworzeniu trzech rankingów i klasyfikacji krajów członkowskich UE osobno pod względem sytuacji ekonomicznej, wydatków na ochronę zdrowia w przeliczeniu na osobę i prowadzenia zdrowego stylu życia przez mieszkańców w 2019 roku. Następnie obliczono miary zależności między rankingami. Badanie wykazało, że w krajach o lepszych warunkach ekonomicznych ludzie zazwyczaj prowadzą zdrowszy styl życia. Jednak nie zawsze tak jest – na przykład w Luksemburgu, który jest bogatym krajem, mieszkańcy nie żyją zbyt zdrowo.

Słowa kluczowe: zdrowy styl życia, zdrowie, sytuacja ekonomiczna, ochrona zdrowia, Unia Europejska

1. Wstęp

Styl życia jest jednym z elementów opisujących życie człowieka. Obecnie coraz częściej zwraca się uwagę na zagadnienie zdrowego stylu życia, które jest istotne przede wszystkim ze względu na zdrowotne konsekwencje prowadzenia niezdrowego stylu życia. Choroby cywilizacyjne spowodowane antyzdrowotnymi zachowaniami przyczyniają się do ogromnych kosztów leczenia ponoszonych przez państwo oraz jednostki. Interesujące wydaje się zatem określenie, co warunkuje wybór określonego stylu życia przez człowieka.

Głównym celem artykułu jest analiza stylu życia w odniesieniu do sytuacji ekonomicznej krajów Unii Europejskiej. Postanowiono odpowiedzieć na następujące

pytanie badawcze: czy pomiędzy prowadzeniem zdrowego stylu życia a sytuacją ekonomiczną lub wydatkami na ochronę zdrowia w państwach Unii Europejskiej występuje zależność?

W pierwszej części artykułu opisano elementy składające się na zdrowy styl życia. Następnie skupiono uwagę na kosztach związanych z prowadzeniem zdrowego oraz niezdrowego stylu życia.

Druga część artykułu jest opisem przeprowadzonego badania dotyczącego rozpoznawania zależności między zdrowym stylem życia a czynnikami ekonomicznymi i wydatkami na ochronę zdrowia w państwach Unii Europejskiej w 2019 roku. Wykonano porządkowanie liniowe, a następnie utworzono rankingi oraz klasyfikacje krajów Unii Europejskiej oddzielnie ze względu na czynniki ekonomiczne, wydatki na ochronę zdrowia w przeliczeniu na osobę i prowadzenie zdrowego stylu życia. W kolejnym kroku obliczono współczynniki korelacji między rankingami. Sprawdzo- no, w których państwach mieszkańcy prowadzili najzdrowszy styl życia i czy można zauważyć zależność między wybieraniem zdrowego stylu życia a warunkami ekono- micznymi w kraju.

2. Zdrowy styl życia

2.1. Definicja zdrowego stylu życia

Zgodnie z raportem ministra zdrowia i opieki społecznej Kanady Marca Lalonde'a opublikowanym w 1974 roku stan zdrowia człowieka zależy aż w 53% od prowadzo- nego stylu życia (Woźniak i in., 2015c). Siciński (2002, s. 22-23) definiuje styl życia jako „zespół codziennych zachowań (sposób <postępowania>, aktywność życiowa) specyficzny dla danej zbiorowości lub jednostki”. Styl życia jest zatem określany przez nawyki, codzienne wybory i powtarzające się zachowania, które razem tworzą pewien obraz życia danej osoby (Giddens, 2001).

Na styl życia człowieka składa się wiele elementów, nawyki żywieniowe, styl ubioru czy sposób spędzenia czasu wolnego. Zdrowy styl życia to taki, który w du- żym stopniu składa się z zachowań prozdrowotnych, czyli takich, które pozwala- ją utrzymać dobry stan zdrowia, zapobiegać chorobom i pomagać w zdrowieniu. Osoby prowadzące zdrowy styl życia unikają zachowań niekorzystnych dla zdrowia (Heszen i Sęk, 2007; Woźniak i in., 2015a). Ostrowska (1999, s. 14) zwraca uwagę, że „prozdrowotne style życia w opracowaniach medycznych sprowadzają się bowiem prawie wyłącznie do określonego sposobu odżywiania się, unikania tytoniu i alkoholu, aktywności fizycznej i – w nowszych pracach – umiejętności zwalczania stresu.”

Odpowiednie odżywianie się jest niezwykle ważne dla prawidłowego funkcjo- nowania organizmu i utrzymania zdrowej sylwetki. Zgodnie z zaleceniami Światowej Organizacji Zdrowia dorosły człowiek powinien spożywać około 2000 kalorii dzien- nie, z czego tłuszcz powinien stanowić maksymalnie 30%, a cukry wolne nie więcej niż 10%. Dieta powinna składać się z różnorodnych pokarmów, a w codziennej die-

cie powinno znaleźć się przynajmniej 400 g (pięć porcji) warzyw i owoców. Ważne jest również, by nie spożywać więcej niż 5 g soli dziennie (WHO, 2020).

Dzięki regularnej aktywności ruchowej człowiek pozostaje w dobrej kondycji fizycznej oraz samopoczuciu. Ćwiczenia fizyczne umożliwiają spalenie spożytych kalorii oraz zredukowanie nadmiaru tkanki tłuszczowej. Zalecane jest, aby osoby zdrowe w wieku 18-65 lat wykonywały ćwiczenia fizyczne o umiarkowanym natężeniu przez co najmniej 30 minut 5 razy w tygodniu. Najkorzystniejsze dla zdrowia są ćwiczenia aerobowe, czyli ciągłe ćwiczenia wytrzymałościowe, angażujące duże grupy mięśni, np. takie jak spacer, jazda na rowerze, pływanie, taniec czy gry zespołowe (Siwiński i Rasińska, 2015; Woźniak i in., 2015b).

Spożywanie dużych ilości alkoholu, palenie papierosów, branie narkotyków oraz nadużywanie leków bardzo niekorzystnie wpływa na zdrowie. Negatywny wpływ na funkcjonowanie organizmu człowieka ma również długotrwały stres. Unikanie używek oraz umiejętne radzenie sobie ze stresem to jedno z elementów zdrowego stylu życia (Kłos, 2014).

Dla pozostania w dobrym zdrowiu ważny jest również odpoczynek, w tym odpowiednia ilość snu. Według National Sleep Foundation dorosła osoba potrzebuje co najmniej siedem godzin snu dziennie (Newsom, 2023). Istotną rolę odgrywa także higiena fizyczna i psychiczna oraz zachowania profilaktyczne i wykonywanie badań kontrolnych (Heszen i Sęk, 2007).

2.2. Koszty związane z prowadzeniem zdrowego stylu życia

Wiele osób uważa, że zdrowy styl życia wymaga wysokich nakładów finansowych (Ostrowska, 1999). Produkty zdrowe, wysokiej jakości lub ekologiczne często są droższe od innych (Mucha i Mucha, 2020). Jednak warto zauważyć, że samodzielne przygotowanie zdrowych posiłków zwykle nie wiąże się z aż tak dużymi kosztami w porównaniu z zakupem gotowych dań lub spożywaniem posiłków w restauracji (Dopierała, 2022). Uprawianie aktywności fizycznej również nie musi się przyczyniać do zwiększonych wydatków. Aby prowadzić zdrowy styl życia, nie trzeba chodzić na siłownię ani zajęcia sportowe. Wystarczy regularne chodzenie na spacer czy ćwiczenie w domu.

Często nie zwraca się również uwagi, że zdrowy styl życia może prowadzić do znacznych oszczędności. Jednym z elementów zdrowego stylu życia jest rezygnacja z używek. Ze względu na to, że papierosy i alkohol są drogimi produktami, ograniczenie lub rezygnacja z ich spożycia pozytywnie wpłynie na domowy budżet.

2.3. Koszty związane z niezdrowym stylem życia

Prowadzenie niezdrowego stylu życia może bardzo negatywnie wpłynąć na stan zdrowia człowieka i przyczynić się do rozwinięcia różnego rodzaju chorób. Badania pokazują, że liczne choroby cywilizacyjne, takie jak nowotwory, cukrzyca, otyłość, choroby układu krążenia oraz niektóre choroby układu pokarmowego, a także za-

burzenia psychiczne są „ściśle związane z warunkami życia, środowiskiem, określonymi zachowaniami ludzi, wzorami konsumpcji – paleniem tytoniu i piciem alkoholu, nadużywaniem leków, zbyt obfitym lub niewłaściwie dobranym jedzeniem, skażeniem środowiska, brakiem aktywności fizycznej, trudnościami w uporaniu się ze stresem” (Ostrowska, 1999, s. 9). Sama niewłaściwa dieta ma związek z około 30-50 jednostkami chorobowymi (Ostrowska, 1999).

Według badania Global Burden of Disease (Globalne Obciążenie Chorobami) przeprowadzonego przez Institute for Health Metrics and Evaluation (IHME) do najczęstszych czynników ryzyka śmierci należą wysokie ciśnienie krwi, palenie tytoniu, otyłość, wysokie stężenie cukru we krwi oraz spożywanie alkoholu (Ritchie i in., 2018). Niezdrowy styl życia jest zatem związany nie tylko z pogorszeniem stanu zdrowia, ale może także prowadzić do przedwczesnej śmierci.

Choroby spowodowane niezdrowym stylem życia generują bardzo duże koszty bezpośrednie, czyli związane z leczeniem. Trzeba także zwrócić uwagę na tzw. koszty pośrednie, związane z okresową lub długotrwałą nieobecnością w pracy oraz zmniejszoną wydajnością pracy – tutaj straty ponoszą zarówno pracownicy, pracodawcy jak i państwo oraz gospodarka. Koszty pośrednie są równe wartości utraczonego PKB lub niewypracowanym zarobkom pracownika w okresie, gdy osoba nie pracuje z powodu choroby. W przypadku przedwczesnej śmierci koszty pośrednie oblicza się w ten sam sposób, z tym że w okresie od wieku, w którym nastąpił zgon pracownika do wieku emerytalnego (Rottermund i Knapik, 2015).

W Wielkiej Brytanii przeprowadzono badanie mające na celu określenie kosztów związanych z antyzdrowotnymi zachowaniami. Wyniki były następujące: nieodpowiednia dieta generowała koszt równy 5,8 mld funtów rocznie, brak aktywności fizycznej – 0,9 mld funtów, picie alkoholu – 3,3 mld funtów, palenie tytoniu – 3,3 mld funtów, otyłość – 5 mld funtów (Breuer, 2014).

Prowadzenie zdrowego stylu życia pozwoli zaoszczędzić pieniądze, które zostałyby wydane na leczenie, w tym zakup leków. Zdrowsze społeczeństwo to również mniejsze koszty dla państwa i przedsiębiorców zatrudniających pracowników.

Aby zapobiegać chorobom przewlekłym i umożliwić ludziom utrzymanie dobrego zdrowia, stosuje się działania prewencyjne i promocję zdrowia. W tym celu propaguje się zachowania prozdrowotne i zwiększa świadomość ludzi w zakresie zdrowia (Kurpas i in., 2015). W Unii Europejskiej na prewencję przeznaczają się jedynie około 3% środków z budżetu na ochronę zdrowia (European Commission, 2019).

3. Opis danych i metod badawczych

Badanie dotyczy relacji między czynnikami ekonomicznymi a zdrowym stylem życia w wymiarze krajowym. Celem analizy było sprawdzenie, czy lepsze warunki ekonomiczne i gospodarcze w kraju wiązały się z praktykowaniem zdrowszych zachowań przez obywateli, a także ustalenie, w jakiej relacji pozostają państwowe wydatki na ochronę zdrowia ze stylem życia mieszkańców danego państwa.

W analizie uwzględniono wyłącznie państwa członkowskie Unii Europejskiej. Ze względu na różnego rodzaju dofinansowania ze środków unijnych warunki ekonomiczne krajów nienależących do Unii Europejskiej nie są porównywalne z tymi w państwach członkowskich. W badaniu wykorzystano dane z 2019 roku dla 27 krajów. Z powodu dużych braków danych z badania wyłączono Wielką Brytanię, pomimo że w badanym roku kraj ten wciąż był członkiem Unii Europejskiej. Wykorzystane zmienne podzielono na trzy kategorie: czynniki ekonomiczne, wydatki na ochronę zdrowia oraz dane związane ze zdrowym stylem życia.

Do kategorii czynników ekonomicznych zaliczono produkt krajowy brutto w przeliczeniu na jednego mieszkańca – PKB *per capita* [wyrażony w jednostce PPS (*Purchasing Power Standard* – standard siły nabywczej)], stopę bezrobocia wśród osób od 15. do 74. roku życia [%] oraz odsetek osób zagrożonych ubóstwem [%], czyli takich, których dochód po uwzględnieniu transferów socjalnych jest niższy od 60% mediany dochodu w kraju. Są to zmienne opisujące ogólną sytuację gospodarczą kraju oraz warunki ekonomiczne jego mieszkańców.

Drugą kategorię tworzą jedynie bieżące wydatki na ochronę zdrowia w przeliczeniu na osobę [PPS/os], uwzględniające wszystkie wydatki na dobra i usługi opieki zdrowotnej świadczone osobom indywidualnym oraz zbiorowe usługi opieki zdrowotnej. Zdecydowano się na odłączenie tej zmiennej od innych danych ekonomicznych ze względu na to, że jest to jedyna zmienna ekonomiczna związana ze zdrowiem. Dodatkowo wydatki na ochronę zdrowia w przeliczeniu na osobę nie mają zbyt dużego bezpośredniego wpływu na ekonomiczną jakość życia obywateli, a są zależne od sytuacji gospodarczej państwa.

Prowadzenie zdrowego stylu życia w tym badaniu rozumiane jest jako regularne uprawianie aktywności fizycznej, spożywanie co najmniej 5 zalecanych porcji warzyw i owoców, niekonsumowanie zbyt dużych ilości cukru i innych słodzików, utrzymywanie prawidłowej masy ciała, niepicie zbyt często alkoholu oraz niepalenie papierosów. W związku z tym w kategorii zmiennych związanych z prowadzeniem zdrowego stylu życia uwzględniono odsetek osób uprawiających prozdrowotną aerobową (niezwiązaną z pracą) aktywność fizyczną przez co najmniej 150 minut tygodniowo [%], gdzie aktywność fizyczna obejmuje różnego rodzaju sporty i jazdę na rowerze w celu dotarcia do jakiegoś miejsca; odsetek osób spożywających co najmniej 5 porcji warzyw i owoców dziennie [%]; roczne spożycie cukru i innego rodzaju słodzików w przeliczeniu na osobę [kg/os/rok], mierzone jako suma cukru, miodu i innego rodzaju słodzików, takich jak np. syrop klonowy lub chemicznie uzyskana fruktoza, dostępnych mieszkańcom kraju do spożycia w czystej postaci lub w przetworzonych produktach; odsetek osób mających wskaźnik BMI w normie [%], gdzie BMI jest współczynnikiem wykorzystywanym do oceny prawidłowości masy ciała dorosłego człowieka, liczonym jako iloraz masy ciała [kg] i kwadratu wzrostu [m²], a zakres normy BMI mieści się w przedziale od 18,5 do 24,99 [kg/m²]; odsetek osób pijących codziennie alkohol [%] oraz odsetek osób palących codziennie papierosy (wytwarzane i skręcane ręcznie) [%]. Dane dotyczą osób od 15. roku życia.

Dane na temat spożycia cukru i innych słodzików pochodzą ze strony Organizacji Narodów Zjednoczonych do spraw Wyżywienia i Rolnictwa (FAO), natomiast reszta zmiennych z bazy danych Eurostat. Wszystkie wykorzystane dane były reprezentatywne. Wystąpiły jedynie trzy braki danych, które przed przystąpieniem do analizy uzupełniono, wykorzystując metodę średniej z sąsiednich obserwacji¹.

Na każdej grupie zmiennych wykonano porządkowanie liniowe metodą wzorca rozwoju². Przed przystąpieniem do porządkowania wszystkie zmienne znormalizowano i ustalono charakter zmiennych. Do stymulant zaliczono: PKB *per capita*, medianę dochodu, wydatki na ochronę zdrowia w przeliczeniu na osobę, odsetek osób uprawiających aerobową (niezwiązaną z pracą) aktywność fizyczną przez co najmniej 150 minut tygodniowo, odsetek osób spożywających co najmniej 5 porcji warzyw i owoców dziennie oraz odsetek osób mających wskaźnik BMI w normie. Destymulantami były: stopa bezrobocia, odsetek osób zagrożonych ubóstwem, roczne spożycie cukru i innego rodzaju słodzików w przeliczeniu na osobę, odsetek osób pijących codziennie alkohol oraz odsetek osób palących codziennie papierosy.

Na podstawie malejących wartości miar rozwoju utworzono rankingi krajów. Następnie państwa podzielono na cztery klasy przy wykorzystaniu metody odchyłeń standardowych³. W celu analizy zależności pomiędzy rankingami dla czynników ekonomicznych i wydatków na ochronę zdrowia oraz rankingiem dotyczącym prowadzenia zdrowego stylu życia obliczono współczynniki korelacji rang Spearmana oraz Kendalla⁴. Aby zbadać podobieństwo między utworzonymi klasami, obliczono indeks Randa⁵.

4. Wyniki analizy

4.1. Porządkowanie krajów Unii Europejskiej ze względu na zdrowy styl życia, czynniki ekonomiczne i wydatki na ochronę zdrowia

Przeprowadzono porządkowanie liniowe metodą wzorca rozwoju osobno dla kategorii czynników ekonomicznych i zmiennych związanych ze zdrowym stylem życia. W kolejnym kroku utworzono rankingi krajów na podstawie malejących wartości miar rozwoju. Wartości wydatków na ochronę zdrowia w przeliczeniu na osobę zstandardyzowano, a następnie utworzono ranking na podstawie malejących wartości

¹ Opis metody uzupełniania danych można znaleźć w *Podręczniku użytkownika IBM SPSS Statistics 28 – System podstawowy* (2021).

² Porządkowanie liniowe metodą wzorca rozwoju jest opisane w rozdziale Kowalewskiego w książce pod redakcją Stanimir (2006b).

³ Metodę podziału na klasy opisano w pracy Pośpiecha (2016).

⁴ Współczynniki korelacji rang Spearmana i Kendalla są opisane w rozdziale Kowalewskiego w książce pod redakcją Stanimir (2006a).

⁵ Szczegóły na temat indeksu Randa można znaleźć w jego pracy (1971).

zmiennej znormalizowanej. Rankingi 27 krajów Unii Europejskiej dla wszystkich kategorii zmiennych przedstawiono w tab. 1.

Tabela 1. Rankingi 27 krajów Unii Europejskiej ze względu na czynniki ekonomiczne, wydatki na ochronę zdrowia i zdrowy styl życia w 2019 roku

Kraj	Miejsce w rankingu ze względu na:			Kraj	Miejsce w rankingu ze względu na:		
	czynniki ekonomiczne	wydatki na ochronę zdrowia	zdrowy styl życia		czynniki ekonomiczne	wydatki na ochronę zdrowia	zdrowy styl życia
Luksemburg	1	5	16	Włochy	15	12	12
Austria	2	3	10	Polska	16	21	17
Dania	3	6	2	Portugalia	17	15	23
Holandia	4	2	1	Słowacja	18	22	20
Niemcy	5	1	9	Litwa	19	18	11
Belgia	6	7	7	Estonia	20	19	8
Finlandia	7	10	5	Węgry	21	23	18
Irlandia	8	9	3	Chorwacja	22	25	25
Szwecja	9	4	4	Hiszpania	23	13	13
Francja	10	8	6	Łotwa	24	24	22
Cypr	11	17	15	Rumunia	25	26	24
Słowenia	12	16	14	Bułgaria	26	27	27
Malta	13	11	26	Grecja	27	20	19
Czechy	14	14	21				

Źródło: opracowanie własne z wykorzystaniem programu MS Excel.

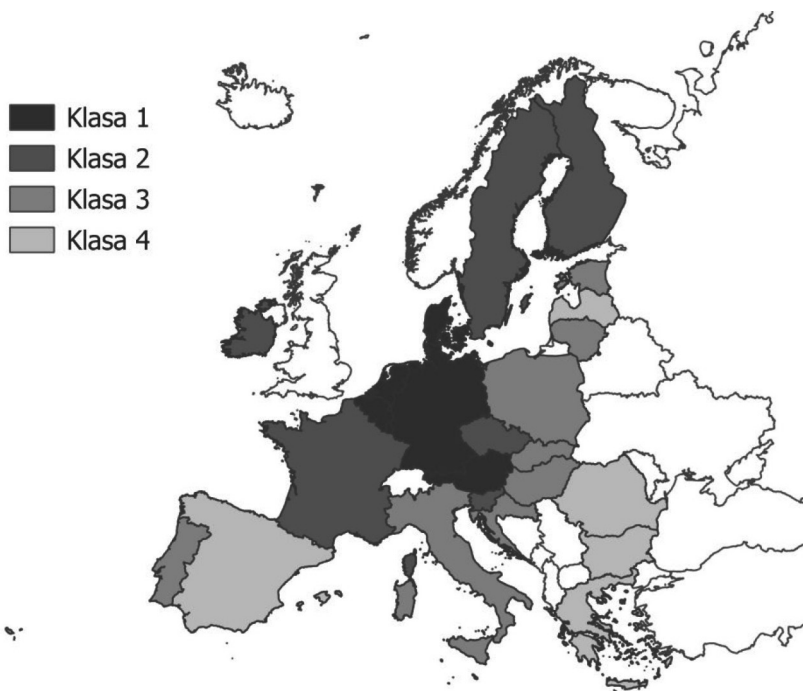
Na podstawie tab. 1 można zauważyć, że na pierwszym miejscu w rankingu pod względem czynników ekonomicznych znalazł się Luksemburg. Oznacza to, że w tym kraju są najlepsze warunki ekonomiczne w Unii Europejskiej. Następne miejsca zajęły Austria, Dania, Holandia i Niemcy. Na końcu rankingu widać Rumunię, Bułgarię oraz Grecję i Rumunię. Polska znalazła się na 16. miejscu.

W rankingu pod względem wysokości wydatków na ochronę zdrowia w przeliczeniu na osobę w 2019 roku przodowały Niemcy. Na kolejnych miejscach znalazły się Holandia, Austria, Szwecja i Luksemburg. Na końcu rankingu znalazły się Chorwacja, Rumunia i Bułgaria. Polska uplasowała się na 21. miejscu.

Na pierwszym miejscu w rankingu pod względem prowadzenia zdrowego stylu życia była Holandia. Kolejne miejsca zajęły: Dania, Irlandia, Szwecja oraz Finlandia. Na ostatnich miejscach znalazły się Chorwacja, Malta i Bułgaria. Polska zajęła 17. miejsce.

W następnym kroku dokonano podziału krajów na cztery klasy ze względu na wartości miar rozwoju, a w przypadku wydatków na ochronę zdrowia – ze względu

na wartości znormalizowane zmiennej. Graficzne przedstawienie klasyfikacji krajów ze względu na czynniki ekonomiczne pokazano na rys. 1.



Rys. 1. Graficzne przedstawienie podziału krajów na 4 klasy ze względu na wartości miar rozwoju dla czynników ekonomicznych w 2019 r.

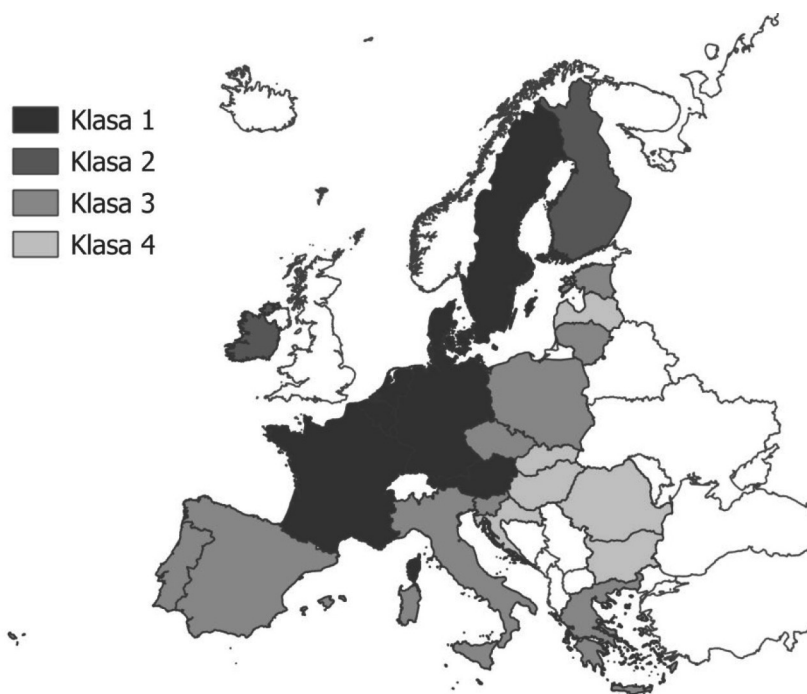
Źródło: opracowanie własne z wykorzystaniem programu QGIS.

Na rysunku 1 widać, że wszystkie państwa zaklasyfikowane do klasy o najlepszych wartościach czynników ekonomicznych leżą w centralnej części Europy. Są to Luksemburg, Austria, Dania, Holandia, Niemcy i Belgia. Klasa druga jest najbardziej zróżnicowana geograficznie. Przyporządkowano do niej Finlandię, Irlandię, Szwecję, Francję, Cypr, Słowenię, Maltę oraz Czechy. Państwa należące do klasy trzeciej są położone w południowej i wschodniej części kontynentu. Są to Estonia, Litwa, Polska, Słowacja, Węgry, Chorwacja, Włochy i Portugalia. Do klasy o najgorszej sytuacji ekonomicznej trafiły głównie kraje znajdujące się w Europie Południowo-Wschodniej (Rumunia, Bułgaria i Grecja) oraz Hiszpania i Łotwa.

Graficzną prezentację klasyfikacji krajów ze względu na wartości znormalizowanej zmiennej dotyczącej wydatków na ochronę zdrowia w przeliczeniu na osobę pokazano na rys. 2.

Na rysunku 2 widać, że do najlepszej klasy pod względem wydatków na ochronę zdrowia w przeliczeniu na osobę w 2019 roku przyporządkowano Niemcy, Holandię,

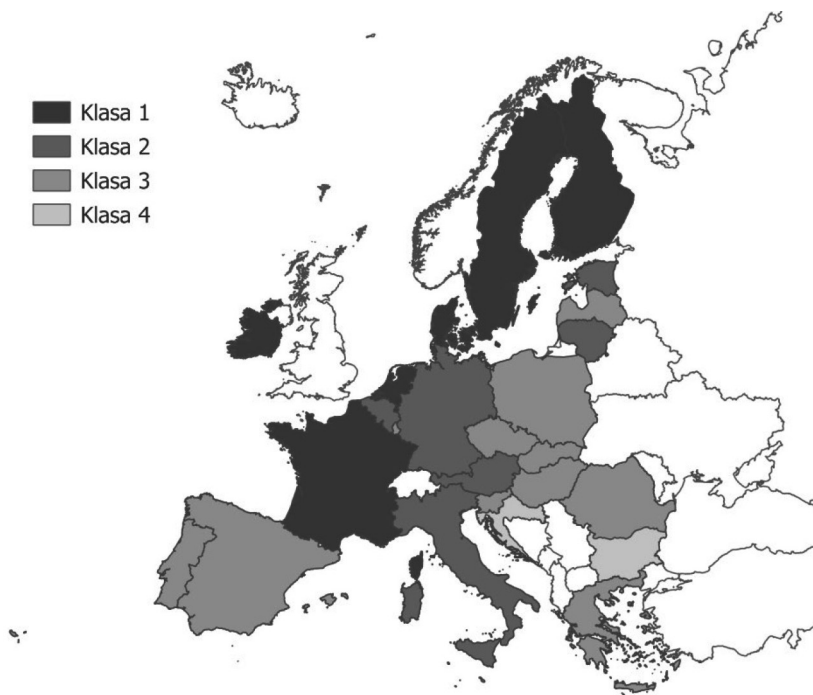
Szwecję, Austrię, Luksemburg, Francję, Danię i Belgię, a więc kraje położone w środkowo-zachodniej i północnej części Europy. Do drugiej klasy trafiły jedynie Irlandia, Finlandia i Malta. Są to państwa znajdujące się w różnych częściach kontynentu. W trzeciej, największej klasie znalazły się kraje Europy Środkowo-Wschodniej (Polska, Czechy, Litwa, Estonia), a także południowej (Słowenia, Włochy, Grecja i Cypr) i południowo-zachodniej (Hiszpania i Portugalia). Do klasy o najniższych wydatkach na ochronę zdrowia na osobę trafiły Słowacja, Węgry, Łotwa, Chorwacja, Rumunia oraz Bułgaria, czyli głównie kraje leżące we wschodniej części kontynentu.



Rys. 2. Graficzne przedstawienie podziału krajów na 4 klasy ze względu na znormalizowaną wartość wydatków na ochronę zdrowia w przeliczeniu na osobę w 2019 r.

Źródło: opracowanie własne z wykorzystaniem programu QGIS.

Graficzne przedstawienie podziału krajów na cztery klasy ze względu na wartości miar rozwoju dla danych dotyczących zdrowego stylu życia zaprezentowano na rys. 3. Wynika z niego, że krajami, których mieszkańcy charakteryzują się najzdrowszym stylem życia, są Holandia, Dania, Irlandia, Szwecja, Finlandia oraz Francja. Są to państwa leżące głównie w północnej części Europy – wyjątkiem jest tutaj Francja. Do drugiej klasy należą Belgia, Niemcy, Austria i Włochy – kraje znajdujące się w centralnej części kontynentu, a także Litwa i Estonia. Do trzeciej, największej klasy przyporządkowano Hiszpanię, Słowenię, Cypr, Luksemburg, Polskę, Węgry, Grecję,



Rys. 3. Graficzne przedstawienie podziału krajów na 4 klasy ze względu na wartości miar rozwoju dla danych dotyczących zdrowego stylu życia z 2019 r.

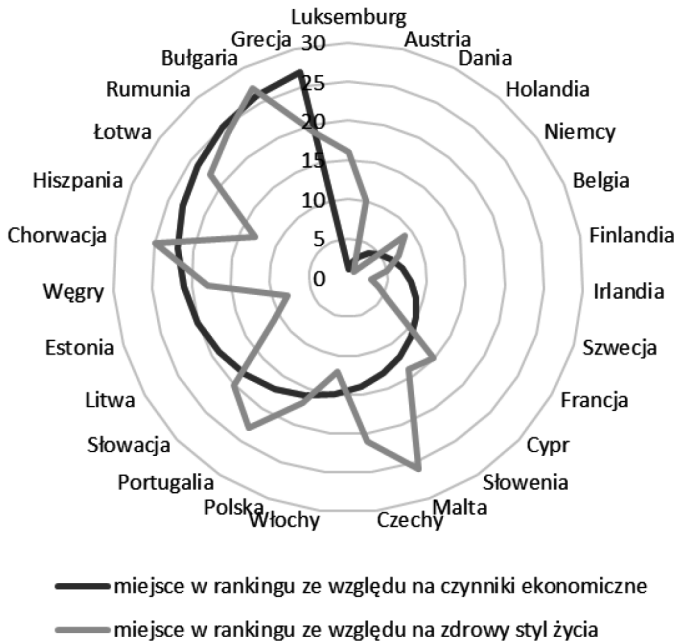
Źródło: opracowanie własne z wykorzystaniem programu QGIS.

Słowację, Czechy, Łotwę, Portugalię oraz Rumunię, a więc państwa położone głównie w środkowo-wschodniej części Europy. Do najgorszej klasy pod względem zdrowego stylu życia zaklasyfikowano kraje Europy Południowej, takie jak Chorwacja, Malta i Bułgaria.

4.2. Zależność między zdrowym stylem życia a sytuacją ekonomiczną oraz wydatkami na ochronę zdrowia

W celu sprawdzenia podobieństwa między rankingami obliczono miary zależności. Współczynnik korelacji rang Spearmana między rankingami dotyczącymi sytuacji ekonomicznej i zdrowego stylu życia w 2019 roku jest równy 0,67, natomiast współczynnik tau Kendalla wynosi jedynie 0,47. Można zatem stwierdzić, że pomiędzy tymi rankingami występuje dodatnia, umiarkowana zależność. Indeks Randa dla klasyfikacji utworzonych na podstawie wyżej wymienionych rankingów jest równy 0,64, a więc oba podziały są do siebie podobne w umiarkowanym stopniu. W związku z tym można uznać, że w państwach bogatszych, o lepszej sytuacji gospodarczej przeciętnie ludzie prowadzą zdrowszy styl życia. Należy jednak zauważyć, że współ-

czynniki korelacji nie informują o przyczynowości, nie można zatem stwierdzić, czy lepsza sytuacja finansowa obywateli powoduje, że wybierają oni zdrowszy styl życia. Nie ma jednak wątpliwości, że między tymi cechami występuje dodatnia zależność, choć nie we wszystkich państwach ją zaobserwowano. Na rysunku 4 przedstawiono różnice pomiędzy miejscami zajmowanymi przez wszystkie kraje w wyżej wymienionych rankingach.



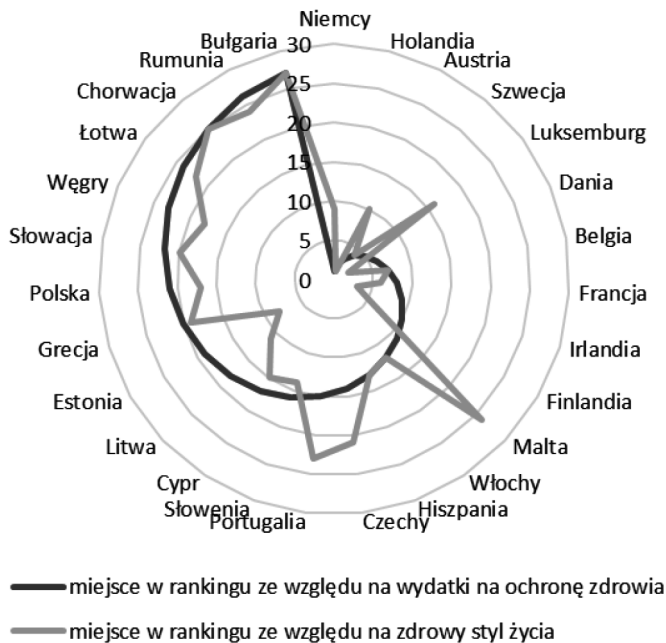
Rys. 4. Porównanie rankingów ze względu na sytuację ekonomiczną oraz zdrowy styl życia w 2019 r.

Źródło: opracowanie własne z wykorzystaniem programu MS Excel.

Na rysunku 4 widać, że największe różnice między miejscami w rankingach ze względu na sytuację ekonomiczną oraz zdrowy styl życia w 2019 roku zaobserwowano dla Luksemburga, Malty, Estonii, Hiszpanii, Austrii, Grecji i Litwy. Różnica między pozycjami w rankingach dla Luksemburga wynosi aż 15 miejsc. Lepsze miejsca w rankingu ze względu na sytuację ekonomiczną zajęło 13 krajów, a 14 krajów – w rankingu dla zdrowego stylu życia. Żaden kraj nie znalazł się na tym samym miejscu w obu rankingach.

W następnej kolejności obliczono korelacje między miejscami w rankingach dla krajów Unii Europejskiej ze względu na wydatki na ochronę zdrowia w przeliczeniu na osobę oraz zdrowy styl życia. Współczynnik korelacji rang Spearmana pomiędzy rankingami wynosi 0,73, a współczynnik tau Kendalla jest równy 0,56, a więc między tymi rankingami również występuje dodatnia zależność, o co najmniej umiarkowa-

nej sile. Indeks Randa dla klasyfikacji utworzonych na podstawie tych rankingów jest równy 0,69, co świadczy o dość wysokim podobieństwie obu podziałów. Zatem w krajach o wyższych wydatkach na ochronę zdrowia na osobę przeciętnie obywatele prowadzą zdrowszy styl życia. Jest to ciekawa obserwacja, ponieważ ze względu na to, że leczenie chorób wywołanych niezdrowym stylem życia jest bardzo kosztowne, można by przypuszczać, że w krajach, w których mieszkańcy nie prowadzą zbyt zdrowego stylu życia, wydatki na ochronę zdrowia będą wyższe. Warto jednak mieć na uwadze, że wysokość wydatków na ochronę zdrowia jest silnie skorelowana z innymi czynnikami ekonomicznymi, takimi jak PKB *per capita* i mediana dochodu w kraju. Oznacza to, że w bogatszych państwach na służbę zdrowia przeznaczają się więcej nakładów, bez względu na styl życia obywateli. Na rysunku 5 przedstawiono różnice pomiędzy rankingami ze względu na wydatki na ochronę zdrowia w przeliczeniu na osobę oraz zdrowy styl życia.



Rys. 5. Porównanie rankingów ze względu na wydatki na ochronę zdrowia w przeliczeniu na osobę oraz zdrowy styl życia w 2019 r.

Źródło: opracowanie własne z wykorzystaniem programu MS Excel.

Z rysunku 5 wynika, że największa różnica między miejscami w rankingach ze względu na wydatki na ochronę zdrowia oraz zdrowy styl życia w 2019 roku dotyczyła Malty i wynosiła aż 15 miejsc. Duże różnice między pozycjami w obu rankingach zaobserwowano również dla Luksemburga (11 miejsc), Estonii (11 miejsc), Portu-

galii (8 miejsc) i Niemiec (8 miejsc). Tylko 6 krajów znalazło się na lepszych miejscach w rankingu dla wydatków na ochronę zdrowia, a 15 krajów – w rankingu dla zdrowego stylu życia. Co ciekawe, aż 6 państw (Szwecja, Belgia, Włochy, Hiszpania, Chorwacja oraz Bułgaria) w obu rankingach zajęło te same miejsca.

Warto wspomnieć także, że klasyfikacje wykonane na podstawie wszystkich rankingów pokazały również przestrzenną prawidłowość. Państwa położone w północnej i środkowo-zachodniej części Europy charakteryzowały się najlepszą sytuacją ekonomiczną i wysokimi wydatkami na ochronę zdrowia, a ich mieszkańcy prowadzili stosunkowo zdrowy styl życia. Kraje Europy Południowo-Wschodniej wypadły najgorzej we wszystkich rankingach.

5. Zakończenie

Celem badania było sprawdzenie zależności między prowadzeniem zdrowego stylu życia a czynnikami ekonomicznymi oraz wydatkami na ochronę zdrowia w państwach Unii Europejskiej. Udało się ustalić, że pomiędzy wyżej wymienionymi elementami występuje dodatnia zależność o umiarkowanej sile. Oznacza to, że w krajach o lepszej sytuacji ekonomicznej lub o wyższych wydatkach na ochronę zdrowia w przeliczeniu na osobę przeciętnie ludzie prowadzili zdrowszy styl życia. Interesująca wydaje się sytuacja w Luksemburgu, który zajął pierwsze miejsce w rankingu ze względu na czynniki ekonomiczne, natomiast wypadł dość słabo w rankingu ze względu na zdrowy styl życia. Być może należałoby przeanalizować, co powoduje, że mieszkańcy tego kraju nie prowadzą zdrowego stylu życia, pomimo sprzyjającej sytuacji ekonomicznej.

Warto zwrócić uwagę również na to, że w badaniu uwzględniono całość wydatków na ochronę zdrowia w kraju, która jest silnie skorelowana z innymi czynnikami ekonomicznymi, takimi jak PKB *per capita* czy mediana dochodu. Aby lepiej zbadać relację między wydatkami na ochronę zdrowia a prowadzeniem zdrowego stylu życia, można byłoby uwzględnić same koszty leczenia chorób cywilizacyjnych mających podłoże w niezdrowym stylu życia. Ciekawe mogłoby być również wzięcie pod uwagę nakładów na prewencję i działania informacyjne związane z propagowaniem zdrowego stylu życia, co pokazałoby, czy edukowanie w tym zakresie przynosi oczekiwane skutki.

Literatura

- Breuer, C. (2014). *The Future of Health and Fitness. A Plan for Getting Europe Active by 2025*. BlackBoxPublishers.
- Dopierała, R. (2022). Kontrola życia poprzez jedzenie na przykładzie dobrowolnej prostoty. *Acta Universitatis Lodzianis. Folia Sociologica*, (83), 57-73. <https://doi.org/10.18778/0208-600X.83.03>
- European Commission. (2019). *State of Health in the EU. Companion Report 2019*. Publications Office of the European Union. <https://doi.org/10.2875/85326>.

- Eurostat. (2022a, 11 listopada). *At-risk-of-poverty Rate by Poverty Threshold, Age and Sex – EU-SILC and ECHP surveys* [zbiór danych]. Pobrano 13 grudnia 2022 z https://ec.europa.eu/eurostat/databrowser/view/ILC_LIO2_custom_4146643/default/table
- Eurostat. (2022b, 24 października). *Body Mass Index (BMI) by Sex, Age and Educational Attainment Level* [zbiór danych]. Pobrano 13 grudnia 2022 z [https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_BM1E\\$DEFAULTVIEW/default/table](https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_BM1E$DEFAULTVIEW/default/table)
- Eurostat. (2022c, 4 kwietnia). *Daily Consumption of Fruit and Vegetables by Sex, Age and Educational Attainment Level* [zbiór danych]. Pobrano 13 grudnia 2022 z https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_FV3E_custom_4145855/default/table
- Eurostat. (2022d, 4 kwietnia). *Daily Smokers of Cigarettes by Sex, Age and Educational Attainment Level* [zbiór danych]. Pobrano 13 grudnia 2022 z https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_SK3E_custom_4145789/default/table
- Eurostat. (2022e, 1 kwietnia). *Frequency of Alcohol Consumption by Sex, Age and Educational Attainment Level* [zbiór danych]. Pobrano 28 marca 2023 z [https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_AL1E\\$DEFAULTVIEW/default/table](https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_AL1E$DEFAULTVIEW/default/table)
- Eurostat. (2022f, 1 grudnia). *Health Care Expenditure by Function* [zbiór danych]. Pobrano 13 grudnia 2022 z https://ec.europa.eu/eurostat/databrowser/view/HLTH_SHA11_HC_custom_6316794/default/table?lang=en
- Eurostat. (2022g, 8 grudnia). *Main GDP Aggregates Per Capita* [zbiór danych]. Pobrano 13 grudnia 2022 z [https://ec.europa.eu/eurostat/databrowser/view/NAMA_10_PC\\$DV_648/default/table](https://ec.europa.eu/eurostat/databrowser/view/NAMA_10_PC$DV_648/default/table)
- Eurostat. (2022h, 24 listopada). *Mean and Median Income by Age and Sex – EU-SILC and ECHP Surveys* [zbiór danych]. Pobrano 13 grudnia 2022 z https://ec.europa.eu/eurostat/databrowser/view/ILC_DIO3_custom_4146868/default/table
- Eurostat. (2022i, 1 czerwca). *Time Spent on Health-enhancing (Non-work-related) Aerobic Physical Activity by Sex, Age and Educational Attainment Level* [zbiór danych]. Pobrano 28 marca 2023 z [https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_PE2E\\$DEFAULTVIEW/default/table](https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_PE2E$DEFAULTVIEW/default/table)
- Eurostat. (2022j, 16 listopada). *Unemployment Rates by Sex, Age and Educational Attainment Level (%)* [zbiór danych]. Pobrano 13 grudnia 2022 z https://ec.europa.eu/eurostat/databrowser/view/LFSA_URGAED__custom_4146437/default/table
- FAO. (2023, 17 marca). *Sugar (Raw Equivalent), Sweeteners, Other, Honey* [zbiór danych]. Pobrano 3 kwietnia 2023, z <https://www.fao.org/faostat/en/#data/FBS>
- Giddens, A. (2001). *Nowoczesność i tożsamość. „Ja” i społeczeństwo w epoce późnej nowoczesności*. Wydawnictwo Naukowe PWN.
- Heszen, I. i Sęk, H. (2007). *Psychologia zdrowia*. Wydawnictwo Naukowe PWN.
- IBM. (2021). *Podręcznik użytkownika IBM SPSS Statistics 28 – System podstawowy*. Pobrano 23 kwietnia 2023 z https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/pl/IBM_SPSS_Statistics_Core_System_User_Guide.pdf
- Kłós, J. (2014). *Zjawisko zdrowego stylu życia we współczesnym społeczeństwie polskim* [rozprawa doktorska, Uniwersytet Medyczny im. Karola Marcinkowskiego w Poznaniu]. Wielkopolska Biblioteka Cyfrowa.
- Kowalewski, G. (2006a). Metody analizy rozkładu wielu zmiennych. W: A. Stanimir (red.), *Analiza danych marketingowych. Problemy, metody, przykłady* (s. 81-98). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Kowalewski, G. (2006b). Metody klasyfikacji i porządkowania. W: A. Stanimir (red.), *Analiza danych marketingowych. Problemy, metody, przykłady* (s. 101-126). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Kurpas, D., Kern, J., Jacquet, J., Randall-Smith, J. i Mroczek, B. (2015). Programy promocji zdrowia i profilaktyki chorób – przykłady z Europy i USA. *Family Medicine & Primary Care Review*, 2, 152-156.

- Mucha, B. i Mucha, M. (2020). Czynniki determinujące zdrowy styl życia w opinii młodych konsumentów. *Ekonomia – Wrocław Economic Review*, 26(4), 21-35. <https://doi.org/10.19195/2658-1310.26.4.2>
- Newsom, R. (2023, 3 marca). *Diet, Exercise, and Sleep*. Sleep Foundation. Pobrano 13 marca 2023 z <https://www.sleepfoundation.org/physical-health/diet-exercise-sleep#references-184386>.
- Ostrowska, A. (1999). *Styl życia a zdrowie*. IFiS PAN.
- Pośpiech, E. (2016). Analiza porównawcza wybranych metod grupowania spółek giełdowych. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (297), 153-165.
- Rand, W. (1971). Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336), 846-850. <https://doi.org/10.2307/2284239>
- Ritchie, H., Spooner, F. i Roser, M. (2018). *Causes of Death*. Our World in Data. Pobrano 4 marca 2023 z <https://ourworldindata.org/causes-of-death>
- Rottermund, J. i Knapik, A. (2015). Ekonomiczny aspekt braku sprawności ruchowej i aktywności fizycznej. W: L. Niebrój (red.), *Health Sciences: Studies for and about Healthcare/ Nauki o zdrowiu: badania dla i o ochronie zdrowia* (s. 79-86). Lulu Press.
- Siciński, A. (2002). *Styl życia, kultura, wybór. Szkice*. IFiS PAN.
- Siwiński, W. i Rasińska, R. (2015). Aktywność fizyczna jako zasadniczy cel stylu życia i zdrowia człowieka. *Pielęgniarstwo Polskie*, 2(56), 181-189.
- World Health Organization [WHO]. (29.04.2020). *Healthy Diet*. Pobrano 6 maja 2023 z <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>
- Woźniak, M., Brukwicka, I., Kopański, Z., Kollár, R., Kollárová, M. i Bajger, B. (2015a). Zdrowie i kultura zdrowotna. *Journal of Clinical Healthcare*, (4), 10-14.
- Woźniak, M., Brukwicka, I., Kopański, Z., Kollár, R., Kollárová, M. i Bajger, B. (2015b). Zdrowie jako wypadkowa różnych czynników. *Journal of Clinical Healthcare*, (4), 15-20.
- Woźniak, M., Brukwicka, I., Kopański, Z., Kollár, R., Kollárová, M. i Bajger, B. (2015c). Zdrowie jednostki i zbiorowości. *Journal of Clinical Healthcare*, (4), 1-3.

Analysis of the Relationship between the Lifestyle of Europeans and Economic Factors

Abstract: The study concerns the issue of a healthy lifestyle in the context of the economic situation in European countries. The aim of the article is to check the relationship between leading a healthy lifestyle, economic factors, and health care expenditure in European Union countries. The first part of the work discussed the elements of a healthy lifestyle and the costs associated with leading a healthy and unhealthy lifestyle. The second part described the creation of three rankings and classification of EU member countries separately in terms of economic situation, health care expenditure per person, and maintaining a healthy lifestyle among residents in 2019. Then, measures of the relationship between the rankings were calculated. The study showed that usually in countries with better economic conditions, people lead a healthier lifestyle. However, there are some exceptions to this rule, for example Luxembourg is a wealthy country, but its citizens do not live a very healthy lifestyle.

Keywords: healthy lifestyle, health, economic situation, health care, European Union

Żaneta Ratajczak

e-mail: 182219@student.ue.wroc.pl

ORCID: 0009-0009-1073-7675

Uniwersytet Ekonomiczny we Wrocławiu

Zastosowanie metody wzorca rozwoju do badania wysokości wybranych dofinansowań ze środków PFRON w województwach Polski

DOI: 10.15611/2024.76.5.04

JEL: C38, I14, J14

© 2024 Żaneta Ratajczak

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>

Cytuj jako: Ratajczak, Ż. (2024). Zastosowanie metody wzorca rozwoju do badania wysokości wybranych dofinansowań ze środków PFRON w województwach Polski. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 57-68). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Tematem niniejszego artykułu jest sytuacja osób z niepełnosprawnością w Polsce w 2021 roku, ze szczególnym uwzględnieniem ich aktywizacji zawodowej. W artykule opisano definicję osoby z niepełnosprawnością oraz rehabilitacji, zwracając uwagę na pojęcie aktywizacji zawodowej. Wyjaśniono, czym jest Państwowy Fundusz Rehabilitacji Osób Niepełnosprawnych [PFRON] i przedstawiono, jaki jest jego wpływ na zatrudnienie osób z niepełnosprawnością na przestrzeni lat. Celem badania było stworzenie rankingu województw pod względem wsparcia aktywizacji zawodowej osób z niepełnosprawnością. By to osiągnąć, wykorzystano metodę wzorca rozwoju, uwzględniając zmienne związane z dofinansowaniami do aktywizacji zawodowej osób z niepełnosprawnością. Przeprowadzone badanie wykazało, że w większości województw Polski wsparcie aktywizacji zawodowej osób z niepełnosprawnością jest na zadowalającym poziomie.

Słowa kluczowe: osoba z niepełnosprawnością, aktywizacja zawodowa, PFRON, metoda wzorca rozwoju

1. Wstęp

W niniejszym artykule podjęto tematykę osób z niepełnosprawnością. Jest to temat aktualny, a nadal często pomijany. W dostępnych źródłach jest mało danych, które są niezbędne do przeprowadzenia ciekawych i przydatnych analiz statystycznych w tej tematyce. Okazało się, że instytucje nigdy nie zbierały istotnych i szczegółowych informacji dotyczących osób z niepełnosprawnością.

Przez wzgląd na rosnącą liczbę osób z niepełnosprawnością należy przeprowadzać jak najwięcej analiz w tej dziedzinie, aby poznać specyfikę życia tej grupy społecznej. Głównym celem badania, którego wyniki zaprezentowano w artykule, było

zbadanie, czy województwa Polski różnią się między sobą pod kątem aktywizacji zawodowej, z uwzględnieniem wysokości wybranych dofinansowań.

W artykule opisano definicję osoby z niepełnosprawnością, biorąc pod uwagę różne aspekty tego pojęcia. Przybliżono również pojęcie rehabilitacji, zwracając szczególną uwagę na jej nieodłączny element – aktywizację zawodową. Co więcej, przedstawiono, czym jest i czym się zajmuje Państwowy Fundusz Rehabilitacji Osób Niepełnosprawnych (PFRON) oraz zaprezentowano efekty jego działania dotyczące sytuacji zawodowej osób z niepełnosprawnością.

Dofinansowania ze środków PFRON wspierają osoby z niepełnosprawnością w wielu obszarach, jednak szczególna uwaga poświęcana jest aktywizacji zawodowej. Poniższa analiza powstała po to, aby zbadać sytuację w Polsce pod kątem wybranych dofinansowań wspierających aktywizację zawodową i uzyskać odpowiedzi na zadane pytania badawcze:

- Czy istnieją województwa, w których struktura wysokości analizowanych dofinansowań dotyczących aktywizacji zawodowej jest większa od pozostałych?
- Czy są województwa, gdzie w niskim stopniu wspiera się aktywizację zawodową zamieszkujących je osób z niepełnosprawnością?

W celu uzyskania odpowiedzi na pytania badawcze został zaprezentowany ranking oraz grupowanie państw uzyskane na podstawie wyników metody wzorca rozwoju.

2. Przegląd literatury

Bardzo trudne okazuje się znalezienie definicji niepełnosprawności. Piocha i Nadolna (2009) twierdzą, że nie jest możliwe uniwersalne opisanie każdej jednostki z dysfunkcjami organizmu. Jest to spowodowane popularnością tego terminu w różnych dziedzinach nauki – każda z nich ustanawia indywidualną definicję, która najlepiej opisuje to, na czym najbardziej chcą się skupić (Paluszkiewicz, 2015). Szczególnie precyzyjne są definicje odnoszące się do poszczególnych kategorii. Różnicują jednostki na podstawie zaawansowania dysfunkcji i aspektów społeczno-demograficznych (Piocha i Nadolna, 2009). Najbardziej znana w Polsce prawna definicja niepełnosprawności mówi, że jest to stan, w którym osoba – czasowo lub na stałe – nie może – całościowo lub częściowo – wykonywać czynności dnia codziennego i wypełniać ról społecznych ze względu na ograniczenia związane z funkcjonowaniem organizmu (Paluszkiewicz, 2015).

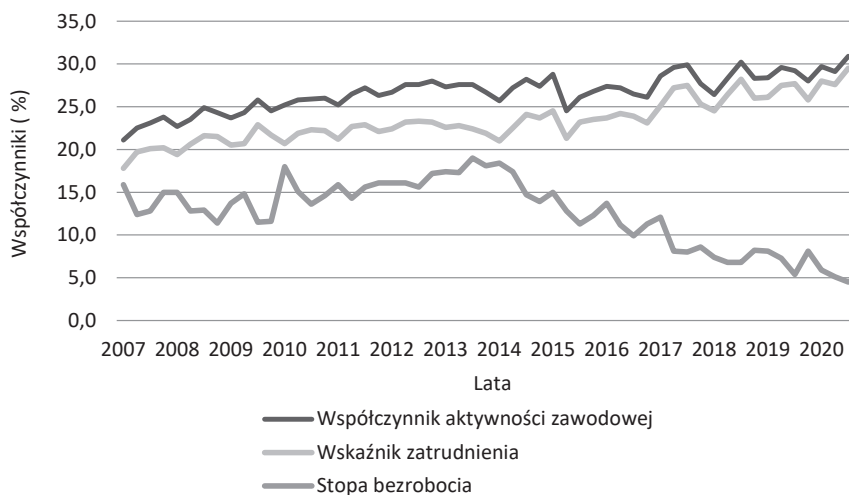
Zwiększenie się liczby osób z niepełnosprawnością wydaje się nieuniknione w przyszłości i jest możliwe do zaobserwowania na przestrzeni lat. Bardzo ważne jest, aby zapewnić im poczucie bezpieczeństwa i odpowiednie warunki do poprawy jakości ich życia. Z tym procesem bezpośrednio związane jest pojęcie rehabilitacji (Marszałek, 2009). Definicja tego terminu wywodzi się z języka łacińskiego. Jest to połączenie słowa *habilitas*, które w dokładnym tłumaczeniu znaczy „zręczność”, z prefiksem *re*, który oznacza powrót do czegoś, co w przeszłości zostało stracone

(Zajac, 2021, s. 145). Ostrowska (2009) twierdzi, że rehabilitacja oznacza ogół czynności, które prowadzą do maksymalnego odzyskania sprawności w ramach możliwości, jakie dana jednostka posiada. Oznacza to, że nie zawsze osoba z niepełnosprawnością jest w stanie powrócić do stanu sprzed nabycia niepełnosprawności, ale może dążyć do maksymalnej samodzielności w zakresie granic, jakie wyznacza choroba lub uszkodzenie.

Nieodłącznym elementem rehabilitacji jest aktywizacja zawodowa. Jest ona niezbędna, gdyż znalezienie odpowiedniej pracy sprawia osobom z niepełnosprawnością bardzo dużo trudności. Aktywizacja zawodowa dotyczy takich etapów, jak (Barczyński, 2013; Kryńska, 2013):

- wybór zawodu – ustalenie mocnych i słabych stron, zdolności i predyspozycji, następnie doradztwo w zakresie szkoleń i wyboru kierunku zawodowego;
- przygotowanie do pracy – pomoc w uzyskaniu dostępu do odpowiedniego miejsca mającego w ofercie wybrane kursy, dobranie odpowiednich metod szkolenia do osoby niepełnosprawnej oraz wsparcie w całym procesie;
- zatrudnienie – odnalezienie pracodawcy, który wyraża chęć zatrudnienia osoby niepełnosprawnej, posiada odpowiednie stanowisko; pomoc w dostosowaniu miejsca pracy, etatu do możliwości jednostki oraz doborze obowiązków;
- adaptacja – wsparcie w początkowym okresie w nowej pracy, zaznajomienie ze współpracownikami i całym środowiskiem w danej firmie, przedstawienie obowiązków, w razie potrzeby umożliwienie wzięcia udziału w dodatkowych szkoleniach oraz monitorowanie całego procesu.

W Polsce aktywizacją zawodową zajmuje się głównie Państwowy Fundusz Rehabilitacji Osób Niepełnosprawnych (PFRON). Jest to fundusz celowy, który powstał na mocy ustawy z dnia 27 sierpnia 1997 r. o rehabilitacji zawodowej i społecznej oraz zatrudnieniu osób niepełnosprawnych (Dz.U. 1997 nr 123 poz. 776). Jego nadrzędną misją jest pomoc w życiu społecznym i zawodowym osobom mającym dowolne orzeczenia o stopniu niepełnosprawności. Dąży do usuwania barier towarzyszących osobom z niepełnosprawnością w życiu codziennym. Fundusz zajmuje się rozdysponowywaniem środków na rehabilitację (Bartniczak, 2012; Giermanowska i Raćław, 2014; Niepełnosprawni.pl, 2015; PFRON, b.d.; WatchDogPfron, b.d.). O dofinansowania z PFRON mogą się ubiegać zarówno pracodawcy i osoby indywidualne, jak i organizacje pozarządowe oraz fundacje. Dofinansowania mają pomóc w zapewnieniu wsparcia w różnych formach aktywizacji zawodowej i społecznej oraz w zakresie rehabilitacji. Pracodawcy mogą się ubiegać między innymi o dofinansowanie do wynagrodzeń i zwrot kosztów dostosowywania stanowiska pracy. Osoby indywidualne mogą między innymi brać udział w szkoleniach, stażach, korzystać z dofinansowań do likwidacji barier architektonicznych, zakupu sprzętu – zarówno rehabilitacyjnego, jak i takiego, który jest niezbędny do prawidłowego funkcjonowania. Mogą również skorzystać ze wsparcia na rozpoczęcie własnej działalności gospodarczej lub rolniczej. Możliwe jest uzyskanie zwrotu kosztów składek na ubezpieczenia społeczne przez rolników i przedsiębiorców (Rogowska, 2021).



Rys. 1. Wykres radarowy miar rozwoju dla województw Polski

Źródło: opracowanie własne na podstawie danych Biura Pełnomocnika Rządu do Spraw Osób Niepełnosprawnych (2022).

Na rysunku 1 zostały zaprezentowane dane dotyczące współczynnika aktywności zawodowej, wskaźnika zatrudnienia oraz stopy bezrobocia obserwowane w Polsce w latach 2007-2020 (dane kwartalne). Widoczne jest, że występuje nieznaczny wzrost współczynnika aktywności zawodowej oraz wskaźnika zatrudnienia, jednak wskaźniki te nadal są niewielkie, pomimo iż PFRON szczególną uwagę poświęca aktywizacji zawodowej. Można również zauważyć, że oba te wskaźniki rosną i spadają proporcjonalnie, co świadczy o ścisłym związku między nimi. Liczba osób z niepełnosprawnością w wieku produkcyjnym, które są zatrudnione, zależy od liczby osób niepełnosprawnych aktywnych zawodowo. Możliwe jest stwierdzenie, że w latach 2007-2015 liczba osób aktywnych zawodowo, które nie pracowały, była stała. Liczba tych osób w ciągu kolejnych kilku lat malała, a wskaźnik zatrudnienia przyjmował wartości coraz bliższe współczynnikowi aktywności zawodowej, co jest pożądaną zmianą. Potwierdza to zauważalny znaczny spadek stopy bezrobocia od 2014 roku. Na tej podstawie możliwe jest stwierdzenie, że sytuacja osób z niepełnosprawnością na rynku pracy znacząco się poprawiła od 2007 roku. Zmiany tych wskaźników świadczą o tym, że działania tego funduszu przynoszą pozytywne efekty.

3. Dane i metody analizy

Niniejsza analiza została przeprowadzona na podstawie wybranych dofinansowań, które wspomagały aktywizację zawodową osób z niepełnosprawnością w 2021 roku w poszczególnych województwach Polski. W badaniu zostały użyte następujące zmienne (PFRON, 2022):

- x_1 – dofinansowanie kosztów działania zakładów aktywizacji zawodowej w 2021 roku – w przeliczeniu na jeden zakład aktywizacji zawodowej;
- x_2 – zwrot kosztów dostosowania i wyposażenia stanowiska pracy dla osoby z niepełnosprawnością w 2021 roku – w przeliczeniu na jedno dostosowane i wyposażone stanowisko;
- x_3 – jednorazowa pomoc finansowa na podjęcie działalności przez osoby z niepełnosprawnością w 2021 roku – w przeliczeniu na jednego beneficjenta korzystającego z dofinansowania;
- x_4 – wydatki na realizację w 2021 roku pilotażowego programu Rehabilitacja 25 plus – w przeliczeniu na jednego beneficjenta korzystającego z programu;
- x_5 – zwrot wydatków na instrumenty finansowe i usługi na rynku pracy w 2021 roku – w przeliczeniu na jedną osobę z niepełnosprawnością, która otrzymała dofinansowanie;
- x_6 – wydatki na realizację zadania Dofinansowanie kosztów szkolenia, o których mowa w art. 18 ustawy o języku migowym i innych środkach komunikowania się w 2021 roku na jednego beneficjenta korzystającego z dofinansowania.

W tabeli 1 zaprezentowano wartości zmiennych w województwach Polski użyte w przeprowadzonych analizach.

Tabela 1. Wartości zmiennych

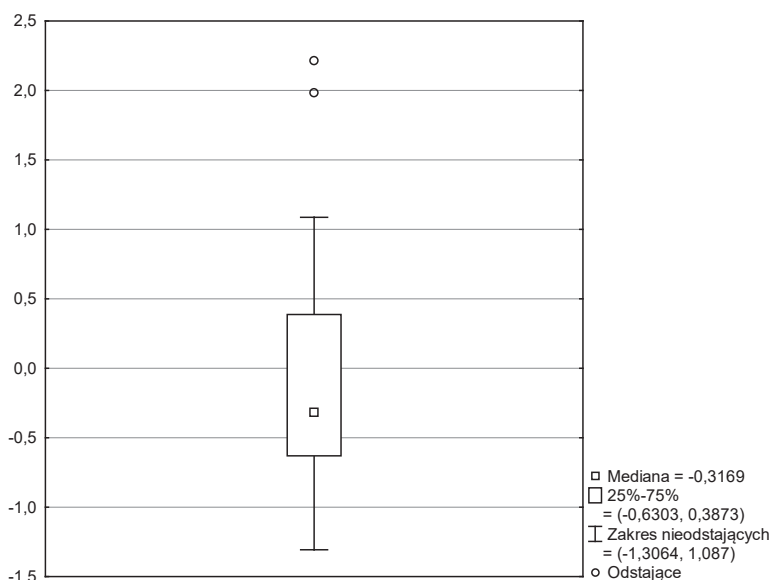
Województwo	x_1	x_2	x_3	x_4	x_5	x_6
Dolnośląskie	2 903 083,63	46 924,00	41 571,15	26 500,00	8 149,63	970,90
Kujawsko-pomorskie	2 788 060,22	59 281,81	48 020,85	19 401,25	9 335,61	845,01
Lubelskie	2 879 693,88	47 155,72	39 619,63	22 796,56	8 864,32	638,59
Lubuskie	1 489 870,50	33 851,71	26 673,86	16 417,65	8 986,00	664,15
Łódzkie	2 271 808,00	59 040,08	45 960,42	26 957,45	3 683,73	860,76
Małopolskie	2 524 362,55	41 898,60	46 415,22	29 868,42	6 352,04	790,49
Mazowieckie	2 304 283,89	66 125,00	50 943,46	24 183,28	7 470,23	1 065,08
Opolskie	1 974 615,80	72 065,00	34 862,00	29 633,33	9 454,38	933,60
Podkarpackie	4 315 123,31	44 547,41	40 587,72	23 396,98	8 043,93	414,26
Podlaskie	2 527 943,80	74 428,38	52 309,53	28 021,07	6 746,57	665,64
Pomorskie	5 648 467,00	48 148,77	38 109,83	27 500,00	7 000,29	862,10
Śląskie	3 298 985,21	51 391,03	42 953,68	23 857,28	7 302,17	869,53
Świętokrzyskie	2 399 263,67	48 486,69	45 178,45	25 781,43	7 335,38	636,32
Warmińsko-mazurskie	2 131 869,11	55 450,27	48 619,05	29 999,99	7 486,59	953,79
Wielkopolskie	3 679 338,89	48 897,31	34 664,92	26 821,38	6 258,10	705,29
Zachodniopomorskie	5 375 781,33	50 881,93	55 074,00	26 965,85	5 329,98	944,40

Źródło: opracowanie własne na podstawie danych z Państwowego Funduszu Rehabilitacji Osób Niepełnosprawnych (2022, s. 10, 14, 22, 26, 27-29).

Metoda wzorca rozwoju jest jedną z metod wielowymiarowej analizy porównawczej. Celem tej metody jest stworzenie rankingu, w którym można odczytać obiekty najlepsze bądź najgorsze pod względem analizowanych cech. Polega to na „określeniu odległości taksonomicznej od wzorca, którym jest (najczęściej sztuczna) jednostka posiadająca najkorzystniejsze wartości dla każdej z cech” (Namyślak, 2015, s. 43). Wszystkie zmienne powinny mieć charakter stymulant lub destymulant, a ponadto zaleca się, aby były one znormalizowane. W prowadzonej analizie zmienne zestandaryzowano. Sposoby normalizacji zmiennych są opisane przez Walesiaka (2014). Dla każdego obiektu w wyniku obliczeń otrzymuje się unormowane miary rozwoju (najczęściej z przedziału $[0,1]$), na których podstawie tworzony jest ranking. Im wyższe wartości miary rozwoju, tym lepsza jest pozycja danego obiektu pod kątem badanego zjawiska (Namyślak, 2015; Stec, 2008). Procedura metody wzorca rozwoju jest szczegółowo opisana przez Kowalewskiego (2006). Otrzymane miary rozwoju sklasyfikowano, tworząc cztery grupy na podstawie średniej i odchylenia standardowego wyznaczonych z miar rozwoju dla województw (Namyślak, 2015).

4. Wyniki porządkowania liniowego

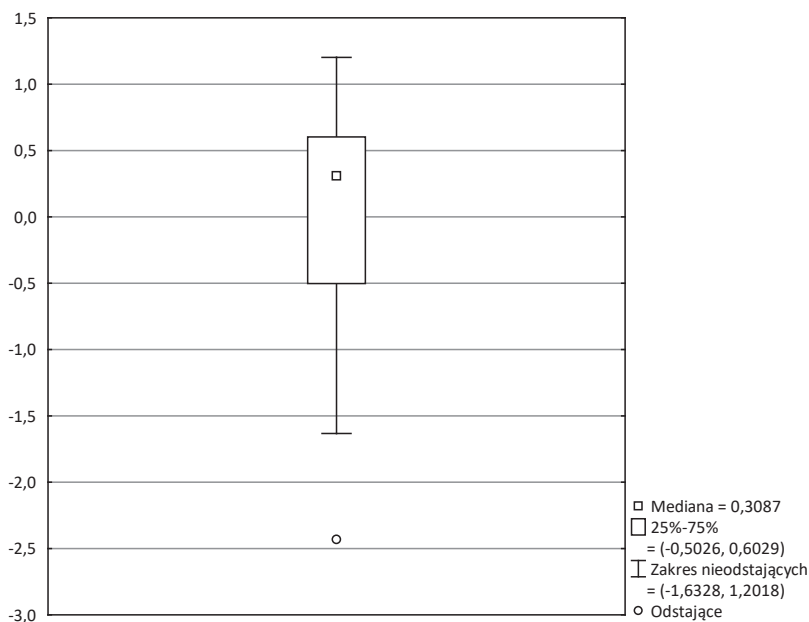
Na rysunkach 2 i 3 przedstawiono wykresy pudełkowe zmiennych, dla których występowały obserwacje odstające. Dla pozostałych zmiennych nie zaobserwowano obserwacji odstających.



Rys. 2. Wykres pudełkowy dla zmiennej x_1

Źródło: opracowanie własne z wykorzystaniem programu Excel.

Na rysunku 2 zaprezentowano wykres pudełkowy dla zmiennej x_1 – dofinansowanie kosztów działania zakładów aktywizacji zawodowej w 2021 roku w przeliczeniu na jeden zakład aktywizacji zawodowej. Na jego podstawie można stwierdzić, że dla analizowanej zmiennej występują dwie obserwacje odstające, które znajdują się powyżej górnego wąsa, i są to województwa zachodniopomorskie i pomorskie.



Rys. 3. Wykres pudełkowy dla zmiennej x_4

Źródło: opracowanie własne z wykorzystaniem programu Excel.

Na rysunku 3 został zaprezentowany wykres pudełkowy dla zmiennej x_4 – wydatki na realizację pilotażowego programu Rehabilitacja 25 plus w 2021 roku na jednego beneficjenta korzystającego z programu. Na jego podstawie można stwierdzić, że dla analizowanej zmiennej występuje jedna obserwacja odstająca, która znajduje się poniżej dolnego wąsa, i jest to wartość odnotowana w województwie lubuskim.

Tak jak wspomniano wcześniej, do przeprowadzenia wielowymiarowej analizy porównawczej wykorzystano metodę wzorca rozwoju. Zostały w niej użyte następujące zmienne: $x_1, x_2, x_3, x_4, x_5, x_6$. Wszystkie zmienne są stymulantami, nie było zatem konieczności ujednolicania charakteru zmiennych. Zmienne zostały zestandaryzowane. Miary rozwoju oraz ranking województw zaprezentowano w tab. 2.

Tabela 2. Ranking i miary rozwoju wyliczone na podstawie metody wzorca rozwoju

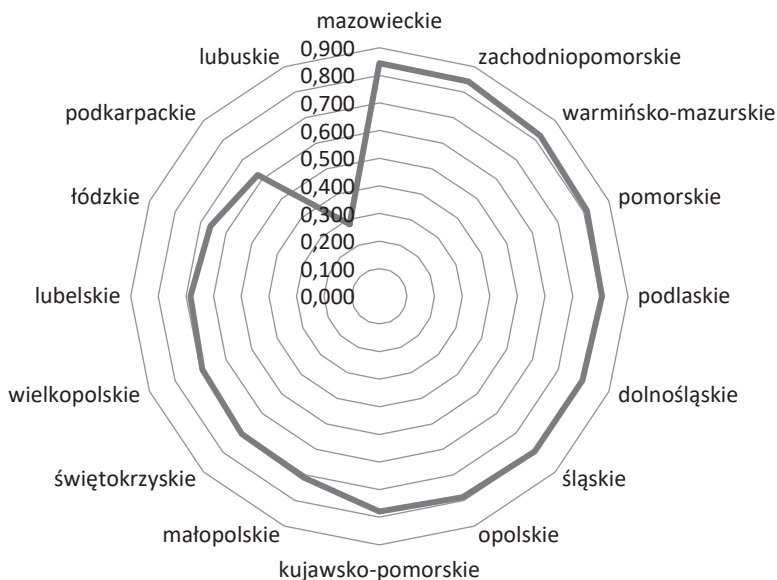
Województwo	Miary rozwoju	Ranking
Mazowieckie	0,844	1
Zachodniopomorskie	0,841	2
Warmińsko-mazurskie	0,823	3
Pomorskie	0,814	4
Podlaskie	0,806	5
Dolnośląskie	0,796	6
Śląskie	0,795	7
Opolskie	0,787	8
Kujawsko-pomorskie	0,779	9
Małopolskie	0,711	10
Świętokrzyskie	0,704	11
Wielkopolskie	0,694	12
Lubelskie	0,684	13
Łódzkie	0,662	14
Podkarpackie	0,622	15
Lubuskie	0,282	16

Źródło: opracowanie własne z wykorzystaniem programu Excel.

Tabela 2 przedstawia ranking stworzony na podstawie miar rozwoju policzonej na podstawie metody wzorca rozwoju. Na podstawie stworzonego rankingu można stwierdzić, że najlepszym województwem pod względem wysokości wybranych dofinansowań ze środków PFRON jest województwo mazowieckie. Można wysunąć wnioski, że w tym województwie PFRON aktywizacji zawodowej osób niepełnosprawnych poświęca szczególną uwagę. Na ostatnim miejscu przedstawionego rankingu znajduje się województwo lubuskie – w tym województwie dofinansowania ze środków PFRON są najniższe na tle innych województw Polski. Województwo lubuskie jest obserwacją odstającą dla zmiennej x_4 , dodatkowo dla zmiennych x_1, x_2, x_3 wartości minimalne również dotyczą tego województwa. Województwa zachodniopomorskie i pomorskie są obserwacjami odstającymi dla zmiennej x_1 , obserwacje te znajdują się powyżej górnego wąsa na wykresie pudełkowym na rys. 2, co wpływa na ich wysoką pozycję w rankingu. Z tabeli 2 można również odczytać, że miary rozwoju dla większości województw oscylowały wokół wartości 0,8, co świadczy o tym, że wielkości analizowanych dofinansowań są podobne w większości z nich.

Rysunek 4, na którym zaprezentowano wykres radarowy przedstawiający miary rozwoju, potwierdza wnioski płynące z rankingu w tab. 2. Widać, że województwa kujawsko-pomorskie, mazowieckie, opolskie, podlaskie, pomorskie, śląskie, warmińsko-mazurskie, zachodniopomorskie i dolnośląskie są podobne pod względem

wysokości analizowanych dofinansowań ze środków PFRON w 2021 roku. Na tle wszystkich województw najgorzej plasuje się województwo lubuskie – na rys. 4 wiadać, że różnica wyliczonej miary rozwoju między nim a innymi województwami jest znacząca.



Rys. 4. Wykres radarowy miar rozwoju dla województw Polski

Źródło: opracowanie własne z wykorzystaniem programu Excel.

Tabela 3. Klasyfikacja wyników porządkowania

Klasa	Zaklasyfikowane województwa
I	brak
II	mazowieckie, zachodniopomorskie, warmińsko-mazurskie, pomorskie, podlaskie, dolnośląskie, śląskie, opolskie, kujawsko-pomorskie
III	małopolskie, świętokrzyskie, wielkopolskie, lubelskie, łódzkie, podkarpackie
IV	lubuskie

Źródło: opracowanie własne z wykorzystaniem programu Excel.

W tabeli 3 przedstawiono klasyfikację wyników porządkowania, gdzie klasa I oznacza najlepszą pod względem analizowanych zmiennych, a IV najgorszą. Można z niej odczytać, że nie ma żadnego województwa, które zostało zakwalifikowane do klasy I, tj. osiągającego wartości wyższe niż średnia miar rozwoju plus odchylenie standardowe miar rozwoju. Nawet miara rozwoju województwa mazowieckiego, które w rankingu zajmuje pierwszą pozycję, nie była na tyle duża, aby móc je zakla-

syfikować do najlepszej klasy. Zdecydowana większość, bo aż dziewięć województw zostało zaklasyfikowanych do klasy II – w tych województwach wielkości wybranych do analizy dofinansowań są zadowalające. Do ostatniej, najgorszej klasy zostało zaklasyfikowane jedno województwo – lubuskie, czyli to, które w rankingu zajmowało ostatnią pozycję.



Rys. 5. Wykres radarowy miar rozwoju dla województw Polski

Źródło: opracowanie własne z wykorzystaniem programu Excel.

Na rysunku 5 przedstawiono klasyfikację wyników porządkowania analizowanych zmiennych na mapie Polski. Można na niej zauważyć grupowanie się poszczególnych województw należących do tych samych klas. Widać, że do klasy drugiej w całości należy północna, północno-wschodnia i południowo-zachodnia część Polski oraz województwa mazowieckie i kujawsko-pomorskie. Do klasy trzeciej należy południowo-wschodnia część oraz województwa łódzkie i wielkopolskie znajdujące się w centralnej części kraju. Do najgorszej klasy pod względem analizowanych dofinansowań należy znajdujące się w zachodniej części Polski województwo lubuskie.

5. Zakończenie

Celem badania było sprawdzenie zróżnicowania wysokości wybranych dofinansowań w województwach. Cel ten w pełni zrealizowano, dodatkowo przybliżając zagadnienie niepełnosprawności oraz rehabilitacji, ze zwróceniem uwagi na pojęcie

aktywizacji zawodowej. Rozpoznano różnice w kwestiach wsparcia osób z niepełnosprawnością w województwach Polski. Pod względem analizowanych zmiennych województwa, które najbardziej wspierają aktywizację zawodową, to m.in. mazowieckie, zachodniopomorskie i warmińsko-mazurskie, a te, które najmniej, to m.in. podkarpackie, łódzkie i lubuskie.

Zaprezentowana analiza nie wyczerpuje tematu aktywizacji zawodowej osób z niepełnosprawnością i badania w tym zakresie powinny być kontynuowane. Ciekawą propozycją na rozszerzenie tego badania jest przeanalizowanie sytuacji osób z niepełnosprawnością związanej z aktywizacją zawodową Polski na tle krajów Unii Europejskiej. Niestety, aktualnie taka analiza nie jest możliwa ze względu na brak dokładnych danych w tej tematyce. Interesujących wniosków mogłoby dostarczyć porównanie działania funduszu PFRON z innymi podobnymi instytucjami w pozostałych krajach Unii Europejskiej.

Literatura

- Barczyński, A. (2013). Aktywizacja zawodowa w kontekście wieku osób niepełnosprawnych. *Problemy Profesjologii*, (2), 127-144.
- Bartniczak, B. (2012). Państwowy fundusz rehabilitacji osób niepełnosprawnych jako podmiot udzielający pomocy publicznej. *Nauki o Finansach*, 2(11), 11-21.
- Biuro Pełnomocnika Rządu do Spraw Osób Niepełnosprawnych. (2022). *Dane na podstawie Badania Aktywności Ekonomicznej Ludności (BAEL)*. Pobrano 2 lutego 2023 z <https://niepelnosprawni.gov.pl/p,81,bael>
- Giermanowska, E. i Raclaw, M. (2014). Niepełnosprawni absolwenci szkół wyższych – nowy potencjał zawodowy czy narastający problem społeczny. *Acta Universitatis Lodzianensis. Folia Sociologica*, 50, 7-26.
- Kowalewski, G. (2006). Metody klasyfikacji i porządkowania. W: A. Stanimir (red.), *Analiza danych marketingowych. Problemy metody przykłady* (s. 101-126). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Kryńska, E. (2013). *Wyniki dotychczasowych badań dotyczących aktywności zawodowej osób niepełnosprawnych*. IPIPS.
- Marszałek, L. (2009). Systemy wsparcia społecznego w procesie rehabilitacji osób niepełnosprawnych. *Seminare. Poszukiwania Naukowe*, (26), 201-213.
- Namyślak, B. (2015). Zastosowanie metody wzorca rozwoju Hellwiga do badania sektora kultury w miastach wojewódzkich. *Wiadomości Statystyczne. The Polish Statistician*, 60(3), 36-51.
- Niepełnosprawni.pl. (2015). *Informator PFRON: Czym jest Fundusz?* Pobrano 2 lutego 2023 z <http://www.niepelnosprawni.pl/ledge/x/276194;jsessionid=4690FBEC6936F3DDF8A467AB77C58842>
- Ostrowska, A. (2009). Niepełnosprawność, rehabilitacja i integracja społeczna osób niepełnosprawnych. W: A. Ostrowska (red.), *Socjologia medycyny: podejmowane problemy, kategorie analizy* (s. 145-171). Wydawnictwo IFiS PAN.
- Paluszkiwicz, M. (2015). Prawne pojęcie niepełnosprawności. *Studia Prawno-Ekonomiczne*, 95, 77-98.
- Państwowy Fundusz Rehabilitacji Osób Niepełnosprawnych [PFRON]. (b.d.). *Co robimy?* Pobrano 2 lutego 2023 z <https://www.pfron.org.pl/o-funduszu/co-robimy/>
- Państwowy Fundusz Rehabilitacji Osób Niepełnosprawnych [PFRON]. (2022). *Sprawozdanie Zarządu PFRON z realizacji Planu Rzeczowo-finansowego Państwowego Funduszu Rehabilitacji Osób Niepełnosprawnych w 2021 roku*.

- Piocha, S. i Nadolna, E. (2009). Zatrudnienie osób niepełnosprawnych jako forma rehabilitacji zawodowej. *Zeszyty Naukowe Wydziału Nauk Ekonomicznych*, 1(13), 127-140.
- Rogowska, D. (2021). Wokół wybranych zagadnień aktywności zawodowej osób niepełnosprawnych. *Zeszyty Naukowe Wydziału Nauk Ekonomicznych*, 1(25), 151-170.
- Stec, M. (2008). Ranking poziomu rozwoju krajów Unii Europejskiej. *Gospodarka Narodowa. The Polish Journal of Economics*, 225(7-8), 99-118.
- Ustawa z dnia 27 sierpnia 1997 r. o rehabilitacji zawodowej i społecznej oraz zatrudnianiu osób niepełnosprawnych (Dz. U. z 1997 r. Nr 123, poz. 776)
- Walesiak, M. (2014). Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej. *Przegląd Statystyczny*, 61(4), 363-372.
- WatchDogPfron. (b.d.). *Czym jest PFRON*. Pobrano 2 lutego 2023 z <http://www.watchdogpfron.pl/czym-jest-pfron/>
- Zając, P. (2021). Formy rehabilitacji zawodowej osób z niepełnosprawnościami finansowane ze środków PFRON, realizowane przez samorządy w latach 2010-2020. *Polonia Journal*, (14), 142-164.

Application of the Development Pattern Method to Study the Amount of Selected Subsidies from PFRON Funds in the Voivodeships in Poland

Abstract: The subject of this article is the situation of disabled people in Poland in 2021, with a special focus on professional activation. The article describes the definition of disabled people and rehabilitation, with particular emphasis on professional activation. It explains what State Fund for Rehabilitation of Disabled Persons [PFRON] is and shows what its impact is on the employment of disabled people over the years. The purpose of the article was to create a ranking of voivodeships in terms of support for the vocational activation of people with disabilities. In this regard, a development pattern method was performed using variables related to subsidies for professional activation of disabled people.

Keywords: disabled person, professional activation, PFRON, development pattern method

Kornelia Kozaczewska

e-mail: kozaczewskakornelia@gmail.com

ORCID: 0009-0002-1084-0284

Uniwersytet Ekonomiczny we Wrocławiu

Analiza interwencji na przykładzie badania wpływu pandemii COVID-19 na dzietność za pomocą modeli SARIMA

DOI: 10.15611/2024.76.5.05

JEL: C22, I1, J13

© 2024 Kornelia Kozaczewska

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>

Cytuj jako: Kozaczewska, K. (2024). Analiza interwencji na przykładzie badania wpływu pandemii COVID-19 na dzietność za pomocą modeli SARIMA. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 69-85). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Wystąpienie pandemii mogło różnie wpłynąć na liczbę urodzeń – zależało to od wielu czynników społecznych, ekonomicznych i zdrowotnych. Problemem badawczym w niniejszym artykule jest określenie, w jakim stopniu pandemia COVID-19 wpłynęła na dzietność w Polsce. Efekty takiego zdarzenia mogą być trwałe, ale mogą też mieć krótkoterminowe skutki. W pracy dokonano analizy wpływu pandemii COVID-19 na dzietność w Polsce, wykorzystując do tego modele SARIMA. Dane dotyczące liczby urodzeń pochodzą z Głównego Urzędu Statystycznego (GUS). Model I lepiej przewidywał liczbę urodzeń, co sugeruje, że COVID-19 nie wpłynął tak mocno na wcześniej podjęte decyzje reprodukcyjne. Jednocześnie dla modelu II wartości zaobserwowane były wyższe niż prognozowane. Badanie należałoby powtórzyć, wykorzystując bardziej skomplikowane metody, np. transmitancję operatorową (*transfer function*). Pełne skutki wpływu pandemii COVID-19 na dzietność mogą być jeszcze nieznanne, co sugeruje potrzebę dalszej analizy.

Słowa kluczowe: dzietność, liczba urodzeń, pandemia, COVID-19, modelowanie, SARIMA, prognozowanie, analiza interwencji, szeregi czasowe

1. Wstęp

Pandemia COVID-19, która wybuchła na początku 2020 roku, wywarła ogromny wpływ na wiele aspektów życia społecznego i gospodarczego na całym świecie. Dotychczasowe badania pokazują, że doprowadziła ona do spadku liczby urodzeń pod koniec 2020 roku i na początku roku 2021 (Aassve i in., 2021; Sobotka i in., 2021), ale ten trend odwrócił się w niektórych krajach w drugiej połowie 2021 roku (UNFPA, 2021). Reakcje społeczeństwa na zmieniające się warunki mają istotny wpływ na

kierunek i dynamikę badanych zjawisk nie tylko w krótkim, ale również w długim okresie. Na kształtowanie się zjawisk społecznych ma wpływ szereg różnorodnych czynników. Dodatkowo pandemia COVID-19 komplikuje analizę, wprowadzając zaburzenia, które mogą skutkować zmianą dynamiki i trendów dzietności.

Badania sugerują, że normy dotyczące rodzicielstwa, polityki łączenia pracy z rodzicielstwem oraz bezrobocie mają wpływ na wybory kobiet w zakresie płodności, całkowitej wielkości rodziny oraz różnicę między nimi (Beaujouan i Berghammer, 2019). Natomiast inne badania podkreślają istnienie zawitych zależności między ekonomią i kulturą w kwestii dzietności, gdzie większy wpływ na normy społeczne mają warunki ekonomiczne niż odwrotnie (Snopkowski i Kaplan, 2015).

Przyczynowość jest szczególnie trudna do zbadania w naukach społecznych ze względu na skomplikowane relacje między zmiennymi. W takim przypadku badanie wpływu zdarzeń jest bardziej zasadne niż badanie bezpośredniej przyczynowości (Morrison, 2012). Przeszłość szeregu czasowego może pomóc w wyjaśnieniu jego bieżącej wartości i w ten sposób uchwycić jego endogeniczną dynamikę. Zmiany, które wpływają na szereg, mogą mieć charakter endogeniczny lub egzogeniczny – ten wpływ może być tymczasowy lub długotrwały (Vasileiadou i Vliegenthart, 2014). W takich sytuacjach często korzysta się ze skomplikowanych modeli, które mimo wszystko nie są skuteczniejsze od prostych modeli prognostycznych (Armstrong i Green, 2015).

Problemem badawczym jest ocena wpływu pandemii COVID-19 na trendy urodzeń przy użyciu narzędzi analizy szeregów czasowych oraz wybranych modeli prognostycznych. Analiza ma na celu zrozumienie, jak kształtują się trendy demograficzne w obliczu trudnych i jedynych w swoim rodzaju okoliczności, jakie przyniosła pandemia.

Artykuł składa się z sześciu punktów. W punkcie drugim opisano dzietność oraz zjawiska z nią związane, czynniki, które na nią wpływają, oraz rolę, jaką odegrała pandemia COVID-19. W punkcie trzecim zaprezentowano dane. W czwartym omówiono analizę interwencji, modele SARIMA oraz szacowanie i metody oceny jakości modeli, a w piątym przedstawiono wyniki analizy liczby urodzeń w Polsce. Zakończenie zawiera podsumowanie rozważań przeprowadzonych w artykule.

2. Dzietność

2.1. Współczynnik dzietności

Współczynnik dzietności teoretycznej TFR (*Total Fertility Rate*) określa liczbę urodzonych dzieci przypadających na jedną kobietę w wieku rozrodczym (15-49 lat). W Polsce od lat 90. wskaźnik ten przyjmuje wartości, poniżej których gwarantowana jest prosta zastępowalność pokoleń. Niekorzystne proporcje między liczbą dzieci, dorosłych i osób starszych wpływają na organizację systemu emerytalnego, rynku pracy i ogólnego funkcjonowania społeczeństwa. Te proporcje są możliwe do odwrócenia właśnie poprzez zwiększenie liczby urodzeń (Kotowska, 2019).

W 2019 roku tylko dwie osoby na sto nie wyrażały chęci posiadania dzieci (CBOS, 2019), ale w 2023 roku takich osób było już 8%. Jednocześnie jedynie 30% ma tyle dzieci, ile chciałoby mieć, a 57% ma mniej dzieci, niż by chciało (CBOS, 2023). Zjawisko to oznacza to, że część planów prokreacyjnych nie jest realizowana. Takie rozbieżności określa się mianem *fertility gap* – jest to różnica między faktyczną a pożądaną liczbą dzieci, która wynika z szeregu powiązanych ze sobą czynników.

2.2. Czynniki wpływające na dzietność

Czynniki, które mają wpływ na liczbę urodzeń, to między innymi czynniki kulturowe, które są związane z dynamiką w związku oraz oczekiwaniami wobec partnera i rodziny. Badania pokazują, że egalitarny podział obowiązków domowych istotnie wpływa na podejmowanie decyzji o posiadaniu potomstwa (Okun, 2016). Natomiast tradycyjne poglądy nie muszą oznaczać większej liczby dzieci – w rodzinach Norwegów o nowoczesnych poglądach średnia liczba dzieci nie była mniejsza niż w rodzinach Polaków o tradycyjnych poglądach (Kocot-Górecka, 2015), gdzie częściej obserwuje się model rodziny z podwójnym obciążeniem kobiet (*dual earner – double burden of women*) (Matysiak, 2006).

Dodatkowo na podejmowanie decyzji reprodukcyjnych wpływ mają czynniki polityczno-prawne związane ze wsparciem na rynku pracy oraz polityką rodzinną, a które są silnie związane z sytuacją zawodową matek. Ponad dwie trzecie kobiet uważa, że powiększenie rodziny wpłynie negatywnie na ich sytuację na rynku pracy, natomiast prawie połowa mężczyzn uważa, że taka decyzja wpłynie pozytywnie na ich sytuację. Najczęściej wymieniane czynniki, które pomagają kobietom w powrocie do pracy, to elastyczne godziny pracy, wsparcie w zapewnieniu opieki nad dziećmi, a także możliwość wykonywania jej zdalnie (Ministerstwo Rodziny i Polityki Społecznej, 2021). Praca z domu pomaga kobietom zmniejszyć koszty i czas związany z pracą (np. oszczędzanie czasu na dojazdach) (Osiewalska i in., 2022). Te czynniki są powiązane z czynnikami kulturowymi – często w sytuacji, gdy kobieta wraca do pracy po urlopie macierzyńskim, opieka nad dzieckiem jest powierzana rodzinie (Styrc, 2007). Elastyczny czas pracy oraz lepsze możliwości opieki poza domem nad dziećmi w wieku do 7 lat są najbardziej preferowanymi rozwiązaniami problemów wynikających z łączenia rodzicielstwa i pracy zawodowej (Strzelecki i in., 2013).

2.3. Pandemia COVID-19 a dzietność

Widoczny jest wpływ pandemii na czynniki ekonomiczne, które z kolei oddziałują na dzietność. W pierwszych miesiącach pandemii wiele osób zostało zwolnionych z pracy, a grupy, które zostały najbardziej poszkodowane, to ludzie młodzi oraz kobiety (Smit i in., 2023). W Polsce stanowią one większość pracowników w sektorach, które zostały najsilniej dotknięte pandemią (Krause, 2021).

Również rodzicielstwo zostało dotknięte konsekwencjami pandemii. To głównie kobiety przejmowały nowe obowiązki, które spadały na rodzinę (np. pomoc dzieciom w nauce w trakcie godzin szkolnych) (Hays Poland, 2021). Na świecie kobiety na niepłatną pracę (zajmowanie się rodziną) oraz na wykonywanie obowiązków domowych poświęcają aż trzy razy więcej czasu niż mężczyźni (Committee for the Coordination of Statistical Activities, 2020). Poczucie odpowiedzialności za wszelkie skutki, które przyniosła pandemia, było również przyczyną niepewności oraz strachu. Stwierdzono też, że to kobiety częściej miały negatywne wizje przyszłości związane z pandemią (Drozdowski i in., 2020) – co wynika z silnych powiązań między czynnikami kulturowymi oraz społecznymi.

Pandemia wpłynęła również na czynniki zdrowotne – kobiety oraz ludzie młodzi deklaruowali wzrost postrzeganego u siebie stresu (Kowal i in., 2020). Pomimo osłabienia restrykcji poziom stresu u kobiet nie był istotnie różny od tego odczuwanego na początku pandemii (Rokosz i in., 2021). Wydłużający się kryzys wyniszcza organizm – najbardziej na ten stan są narażeni ludzie młodzi, którzy jeszcze nie mają wypracowanych mechanizmów radzenia sobie z trudnymi sytuacjami (Markiewicz, 2019). Najczęściej wymieniane obawy związane z pandemią to niepewna sytuacja w kraju, brak stabilności zatrudnienia i strach przed utratą pracy, a także utrudniony dostęp do lekarzy. Wspominano również o restrykcjach w szpitalach, a zwłaszcza na oddziałach położniczych (Ministerstwo Rodziny i Polityki Społecznej, 2021).

3. Opis danych

Do analizy został wykorzystany szereg czasowy liczby urodzeń żywych (tys.) w Polsce. Dane były kompletne. Pobrano je z „Biuletynu Statystycznego” nr 11/2023 Głównego Urzędu Statystycznego (GUS)¹, a dokładniej z tablicy 7. *Stan i ruch naturalny ludności*. Rejestrowane dane są miesięczne, podane w tysiącach. Do analizy został użyty okres od stycznia 2013 roku do października 2022 roku włącznie.

4. Metodyka

4.1. Analiza interwencji

Aby ocenić wpływ czynników zewnętrznych na szereg czasowy, można wykorzystać analizę interwencji ITSA (*Intervention Time Series Analysis*). Jest to metoda badawcza służąca do oceny wpływu jednorazowego zdarzenia na zachowanie szeregu. Interwencje mogą przyjmować różne formy, takie jak zmiany polityki, wydarzenia społeczne czy kryzysy ekonomiczne, czas wystąpienia (jednorazowa lub ciągła) oraz mieć różną skalę – od lokalnej po globalną.

¹ Dane do pobrania ze strony GUS: <https://stat.gov.pl/obszary-tematyczne/inne-opracowania/informacje-o-sytuacji-spoeczno-gospodarczej/biuletyn-statystyczny-nr-112023,4,144.html>.

Analiza interwencji znalazła zastosowanie do zbadania rozkładu radioaktywności po katastrofie w Czarnobylu oraz jego wpływu na łańcuch pokarmowy człowieka i procesy gospodarcze (Hultkrantz i Olsson, 1997; Smith i in., 2000) czy do sprawdzenia siły wpływu dostępności heroiny w Australii w latach 2001-2002 (Degenhardt i in., 2005).

Jednym ze sposobów przeprowadzenia analizy interwencji jest analiza różnic między okresem przed interwencją a okresem po interwencji. Skutki takiego zdarzenia można oszacować, stosując modele do określenia wzrostów lub spadków średniego poziomu lub zmiany trendu (Gilmour i in., 2006). W tym celu utworzono dwa modele dla szeregu czasowego. Pierwszy opiera się na danych zebranych przed wystąpieniem interwencji i na jego podstawie zaprognozowano kolejne 12 miesięcy po interwencji. Jego celem jest analiza zachowania szeregu czasowego i próba określenia jego kierunku, gdyby interwencja nigdy nie wystąpiła. Drugi model uwzględnia dane z 12 miesięcy po wystąpieniu interwencji. Na podstawie tych danych również wykonano prognozę na kolejne 12 miesięcy. Celem drugiego modelu jest zbadanie ewentualnych zmian w zachowaniu szeregu czasowego spowodowanych interwencją. Oba modele oraz ich rezultaty zostały porównane ze sobą, co umożliwi ocenę wpływu interwencji na badany szereg czasowy.

Stosując metodę ITSA, należy mieć na uwadze dokładne zdefiniowanie nieplanowanego zdarzenia. Data wystąpienia interwencji ma kluczowe znaczenie – bez jasnej daty zakres możliwych modeli poszerza się, co może prowadzić do nieścisłości.

4.2. Modele SARIMA

Modele ARIMA są uważane za najbardziej ogólny rodzaj analizy szeregów czasowych (Boef i Keele, 2008) i nadają się szczególnie dla szeregów czasowych o długości większej niż 50 obserwacji (Yanovitzky i VanLear, 2008).

Model ARIMA składa się z trzech elementów: składnika AR (p) – procesu autoregresyjnego, którego każda wartość jest kombinacją liniową poprzednich wartości, składnika MA (q) – procesu średniej ruchomej, który działa podobnie jak proces AR, a także uwzględniają stopień integracji – I (d). Stopień integracji odnosi się do liczby różnicowania szeregu potrzebnego do uzyskania jego stacjonarności.

Końcowy model, uwzględniający opóźnienia, można zapisać w następujący sposób:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

gdzie: y'_t – zróżnicowany szereg, $y_{t-1} + \dots + y_{t-p}$ – wartości szeregu w czasie od $t-1$ do $t-p$, ϕ – parametr siły wpływu poprzedniej wartości na wartość bieżącą (dla składnika AR), θ – parametr siły wpływu starszych zaburzeń szeregu na bieżącą wartość (dla składnika MA), ε_t – składnik losowy w czasie t .

Rozszerzeniem modelu ARIMA (p, d, q) jest model SARIMA, zapisywany jako ARIMA $(p, d, q)(P, D, Q)_m$, gdzie dodatkowo jest brany pod uwagę składnik sezonowy S oraz liczba obserwacji w ciągu roku – m . Część sezonowa modelu zawiera takie same elementy jak niesezonowa, ale obejmuje one sezonowe opóźnienia. W modelu SARIMA sezonowe i niesezonowe składniki są przez siebie mnożone:

$$\begin{aligned} (1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^m - \dots - \Phi_P B^{mP}) \\ (1 - B)^d (1 - B^m)^D y_t = (1 + \theta_1 B + \dots + \theta_q B^q) \\ (1 + \Theta_1 B^m + \dots + \Theta_Q B^{mQ}) \varepsilon_t, \end{aligned}$$

gdzie: ϕ_p – parametr AR dla składowych niesezonowych, Φ_p – parametr AR dla składowych sezonowych, θ_q – parametr MA dla składowych niesezonowych, Θ_Q – parametr MA dla składowych sezonowych, B – operator różnicowy (zwykle jest to opóźnienie o jeden okres).

Wartości parametrów modeli ARIMA można określić, analizując wykresy funkcji autokorelacji ACF (*Autocorrelation Function*) i funkcji autokorelacji cząstkowej PACF (*Partial Autocorrelation Function*). ACF i PACF są również przydatne do analizy struktury szeregu czasowego.

4.3. Szacowanie oraz ocena jakości modelu

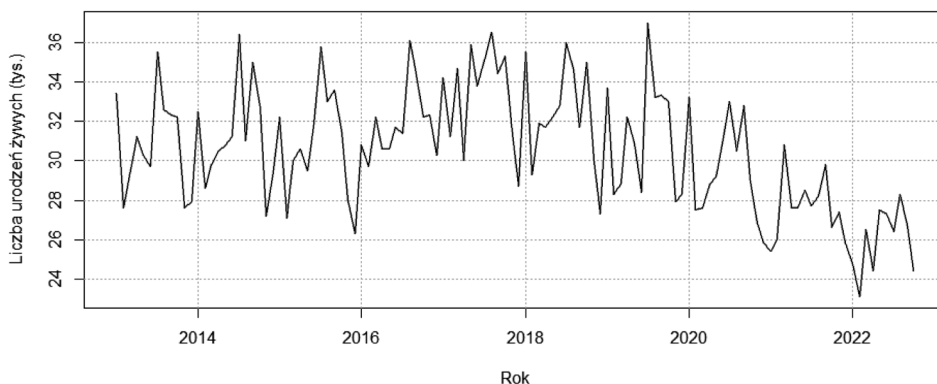
Modele zostały oszacowane za pomocą metody największej wiarygodności (przy wykorzystaniu funkcji `auto.arima()` dla modeli SARIMA oraz `ets()` dla modeli ETS). Porównywane są ze sobą kryteria informacyjne – im niższa wartość, tym lepsze dopasowanie modelu do danych. Ostatecznie wybierany jest model z najniższym AIC (*Akaike's Information Criterion* – kryterium informacyjne Akaike). Zarówno AIC (Akaike, 1973), jak i BIC (*Bayesian Information Criterion* – kryterium bayesowskie Schwartz'a) (Schwarz, 1978) są najważniejszymi kryteriami doboru modeli w ekonometrii i statystyce (Hossain, 1998). Często sprawdza się również AIC_c (AIC z poprawką) (Hurvich i Tsai, 1989), które jest szczególnie polecane dla małych prób.

W celu oceny jakości modelu porównuje się zbiór testowy (*test set*) (w tym przypadku 12 ostatnich obserwacji szeregu) z prognozą wykonaną na podstawie pozostałych obserwacji – zbioru uczącego (*training set*), za pomocą błędów *ex post*. Taka metoda umożliwia ocenę predykcyjną modelu. Te same miary można stosować, aby sprawdzić, czy model dobrze wykonał prognozę. Błędy *ex post* opierają się na resztach modelu, czyli różnicy między wartością empiryczną a teoretyczną. Oceny jakości predykcji modelu dokonuje się poprzez porównanie ze sobą błędów *ex post*. Niższa wartość błędu oznacza dokładniejszą prognozę.

5. Liczba urodzeń w Polsce

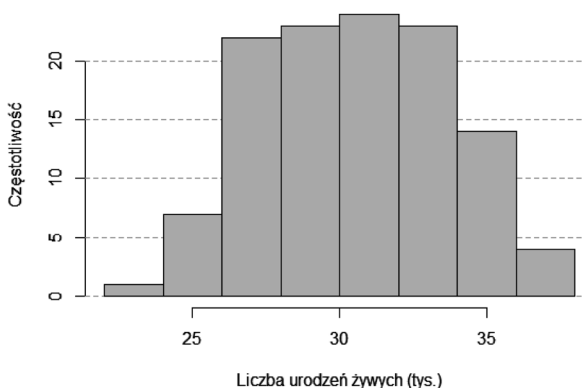
5.1. Charakterystyka liczby urodzeń przed pandemią i po jej zakończeniu

Rysunki 1 oraz 2 przedstawiają wykres oraz histogram liczby urodzeń żywych. Od 2016 do 2018 roku występuje widoczny wzrost liczby urodzeń, po którym następuje spadek – przez okres pandemii aż do 2022 roku (rys. 1). Najczęściej występowały wartości od 30 do 32 tys. urodzeń. Dane są dość symetryczne, z lekką lewostronną asymetrią (rys. 2).



Rys. 1. Wykres liniowy liczby urodzeń żywych w Polsce (w tys.)

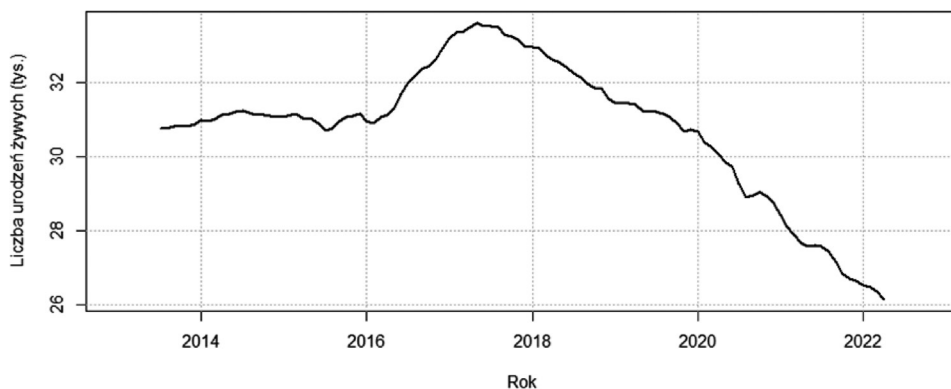
Źródło: obliczenia własne na podstawie danych GUS.



Rys. 2. Histogram liczby urodzeń żywych w Polsce (w tys.)

Źródło: obliczenia własne na podstawie danych GUS.

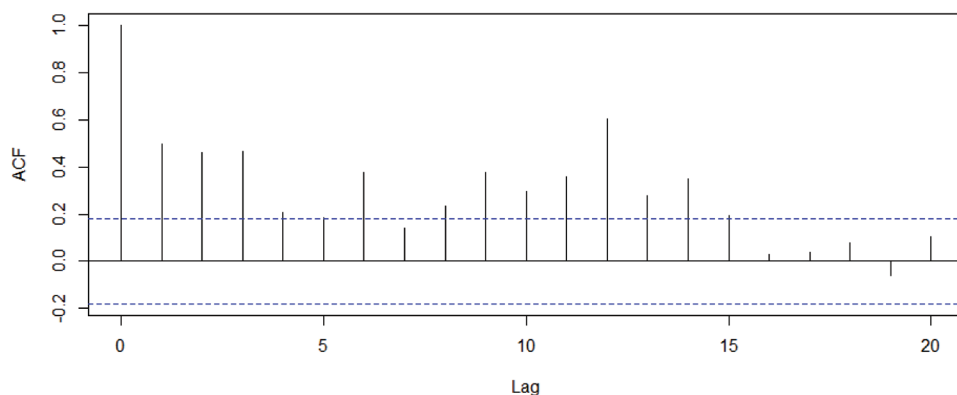
Trend liczby urodzeń był spadkowy już przed 2018 rokiem. Przed 2016 rokiem trend był stabilny. Okres wzrostu został zarejestrowany od roku 2016 do czwartego kwartału 2017 roku (rys. 3).



Rys. 3. Wykres linii trendu liczby urodzeń żywych w Polsce (w tys.)

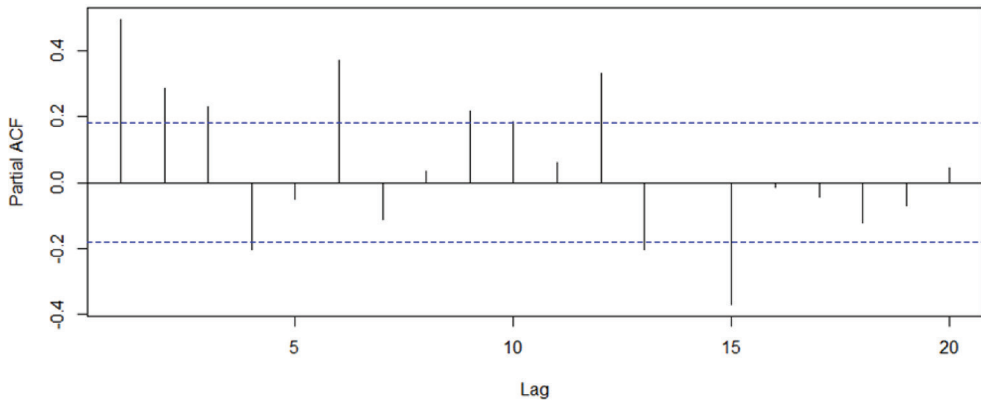
Źródło: obliczenia własne na podstawie danych GUS.

W zdecydowanej większości opóźnień występuje dodatnia korelacja. Największa wartość jest przyjmowana dla 12-tego opóźnienia, co wskazuje na sezonowość roczną (rys. 4). Korelacje między obserwacją a konkretnymi opóźnieniami są zarówno dodatnie, jak i ujemne. Największa ujemna korelacja jest dla opóźnienia równego 1 oraz 15, natomiast największa dodatnia korelacja jest dla 1., 6. oraz 12. miesiąca (rys. 5).



Rys. 4. Wykres ACF liczby urodzeń żywych w Polsce (w tys.)

Źródło: obliczenia własne na podstawie danych GUS.



Rys. 5. Wykres PACF liczby urodzeń żywych w Polsce (w tys.)

Źródło: obliczenia własne na podstawie danych GUS.

5.2. Modele urodzeń przed pandemią i po pandemii

W pierwszej kolejności zostały wykonane testy ADF (Dickey-Fuller = $-4,9286$, Lag order = 4, p -value = 0,01) oraz KPSS (KPSS Level = 0,1792, Truncation lag parameter = 3, p -value = 0,1) dla okresu do 9 miesięcy przed pandemią, które to potwierdziły stacjonarność danych. W przypadku drugiego badanego okresu, uwzględniającego pandemię, test na istnienie pierwiastka jednostkowego potwierdza stacjonarność danych (Dickey-Fuller = $-4,3499$, Lag order = 4, p -value = 0,01), jednak test na trendostacjonarność danych odrzuca hipotezę zerową mówiącą o stacjonarności (KPSS Level = 0,44548, Truncation lag parameter = 4, p -value = 0,05755). Ze względu na niskie p -value dla testu KPSS w obu przypadkach dokonano jednokrotnego zróżnicowania danych, aby usunąć niestacjonarność.

Z powodu występowania autokorelacji danych oraz sezonowości w obu przypadkach dopasowano model z autoregresyjnym sezonowym składnikiem różnicowania uwzględniający dwa opóźnienia oraz z autoregresyjnym niesezonowym składnikiem różnicowania uwzględniającym dwa sezony opóźnień, gdzie okres sezonowy wynosi 12 miesięcy i można go zapisać jako $ARIMA(2,1,0)(2,1,0)_{12}$. Współczynniki dla modelu I są nieznacznie mniejsze niż dla modelu II, co oznacza, że ma on mocniejsze autoregresyjne oraz sezonowe składniki niż model I. Estymacja współczynników jest bardziej precyzyjna dla modelu II, na co wskazują niższe błędy standardowe niż w przypadku błędów modelu I. Model II ma natomiast słabsze autoregresyjne i sezonowe składniki niż model I – wynika to prawdopodobnie ze zmiany trendu oraz wzorców, które sprawiają, że model jest mniej skomplikowany (tab. 1).

Tabela 1. Współczynniki oraz błędy standardowe modeli I oraz II

Model	ϕ_1	ϕ_2	Φ_1	Φ_2
Model I	-0,98	-0,65	-0,52	-0,38
	(0,091)	(0,089)	(0,114)	(0,112)
Model II	-0,89	-0,64	-0,56	-0,33
	(0,089)	(0,086)	(0,125)	(0,112)

Źródło: obliczenia własne na podstawie danych GUS.

Model I można przedstawić za pomocą wzoru modelu SARIMA (składniki niesezonowe oraz sezonowe MA są równe 0):

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B)(1 - B^{12})y_t = \varepsilon_t.$$

Dla uproszczenia $(1 - B)(1 - B^{12})y_t$ oznaczone jest jako z_t :

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})z_t = \varepsilon_t.$$

Po przekształceniu z_t wyraża się jako:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \Phi_1 z_{t-12} - \phi_1 \Phi_1 z_{t-13} - \phi_2 \Phi_1 z_{t-14} + \Phi_2 z_{t-24} - \phi_1 \Phi_2 z_{t-25} - \phi_2 \Phi_2 z_{t-26} + \varepsilon_t.$$

Podstawiając współczynniki parametrów, otrzymuje się wyrażenie:

$$z_t = -0,98z_{t-1} - 0,65z_{t-2} - 0,52z_{t-12} - (-0,98)(-0,52)z_{t-13} - (-0,65)(-0,52)z_{t-14} - 0,38z_{t-24} - (-0,98)(-0,38)z_{t-25} - (-0,65)(-0,38)z_{t-26} + \varepsilon_t.$$

Ostateczna postać modelu I ma następującą postać:

$$z_t = -0,98z_{t-1} - 0,65z_{t-2} - 0,52z_{t-12} - 0,51z_{t-13} - 0,34z_{t-14} - 0,38z_{t-24} - 0,37z_{t-25} - 0,25z_{t-26} + \varepsilon_t.$$

Model II również można przedstawić za pomocą wzoru modelu SARIMA (składniki niesezonowe oraz sezonowe MA są równe 0). Podstawiając współczynniki parametrów, otrzymuje się:

$$z_t = -0,89z_{t-1} - 0,64z_{t-2} - 0,56z_{t-12} - (-0,89)(-0,56)z_{t-13} - (-0,64)(-0,56)z_{t-14} - 0,33z_{t-24} - (-0,89)(-0,33)z_{t-25} - (-0,64)(-0,33)z_{t-26} + \varepsilon_t.$$

Ostateczna postać modelu II to:

$$z_t = -0,89z_{t-1} - 0,64z_{t-2} - 0,56z_{t-12} - 0,5z_{t-13} - 0,36z_{t-14} - \\ - 0,33z_{t-24} - 0,3z_{t-25} - 0,21z_{t-26} + \varepsilon_t.$$

Model I dopasowuje się lepiej do danych ze zbioru uczącego (niższa wartość log-likelihood od modelu II) oraz ma mniejszą estymowaną wariancję błędów (mniejsza σ^2). Kryteria informacyjne (AIC, AIC_c oraz BIC) wskazują na lepsze dopasowanie do danych modelu I niż modelu II (tab. 2).

Tabela 2. Dopasowanie modelu oraz kryteria informacyjne modelu I oraz II

Model	log-likelihood	σ^2	AIC	AIC _c	BIC
Model I	-157,23	2,74	324,46	325,26	336,44
Model II	-190,73	3,45	391,45	392,14	404,12

Źródło: obliczenia własne na podstawie danych GUS.

Na podstawie błędów *ex post* zbioru uczącego można stwierdzić, że model I wypada lepiej od modelu II i ma tendencję do przewidywania wartości bliższych rzeczywistym wartościom (tab. 3).

Tabela 3. Błędy prognozy modeli I oraz II na podstawie zbioru uczącego

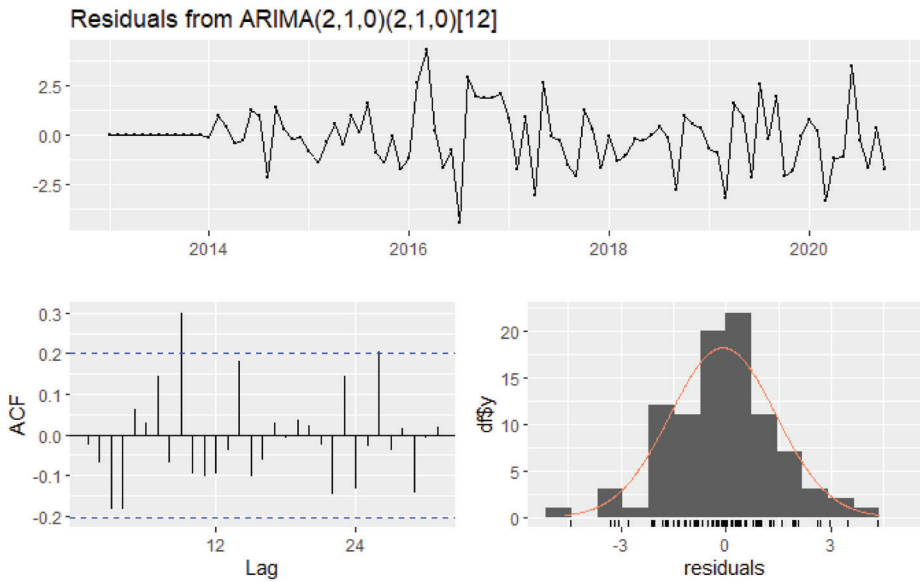
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Model I	-0,08	1,50	1,08	-0,41	3,44	0,63	-0,02
Model II	-0,09	1,70	1,22	-0,45	3,97	0,66	0,01

Źródło: obliczenia własne na podstawie danych GUS.

Rozkład reszt dla modelu II wskazuje na grubsze ogony, niż ma to miejsce w przypadku modelu I (rys. 6 oraz 7). Test Shapiro-Wilka dla modelu I potwierdza normalność rozkładu reszt ($W = 0,9798$, $p\text{-value} = 0,1545$), ale dla modelu II założenie to nie jest spełnione ($W = 0,97497$, $p\text{-value} = 0,04233$), co może wskazywać na słabe dopasowanie modelu do danych. W obu przypadkach nie występuje heteroskedastyczność reszt (rys. 6 oraz 7).

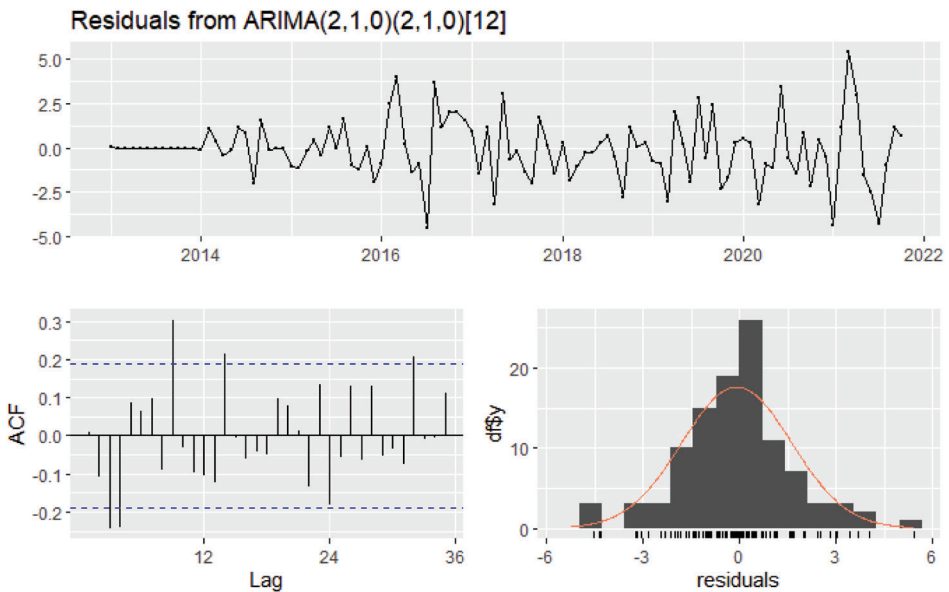
Dla obu modeli ACF wskazuje na występowanie autokorelacji reszt (rys. 6 oraz 7), która może prowadzić do niedokładnych prognoz. Potwierdza to test Ljung-Boxa dla modelu I ($Q^* = 28,741$, $df = 15$, $p\text{-value} = 0,01737$) oraz modelu II ($Q^* = 41,989$, $df = 17$, $p\text{-value} = 0,0006731$).

Na podstawie obu modeli wykonano prognozę dla dwóch okresów: od listopada 2020 roku do października 2021 roku włącznie, na podstawie danych od początku badanego okresu (styczeń 2013 roku) dla modelu I, oraz na kolejne 12 miesięcy (od listopada 2021 roku do października 2022 roku włącznie) na podstawie danych od stycznia 2013 roku do listopada 2021 roku dla modelu II.



Rys. 6. Wykres liniowy, wykres funkcji autokorelacji oraz histogram reszt modelu I

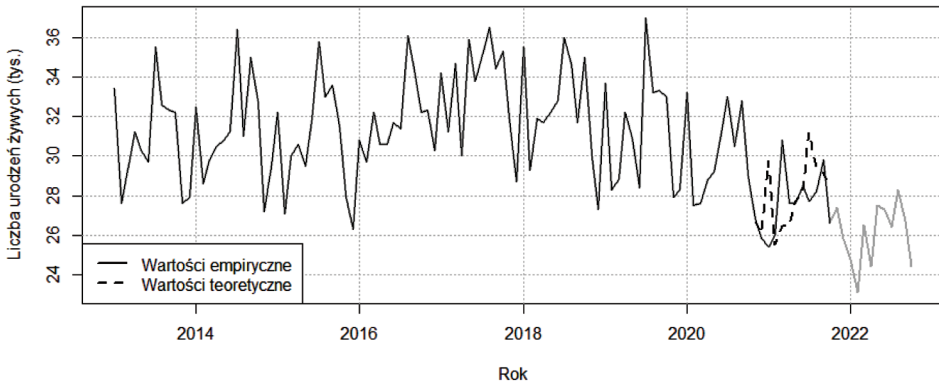
Źródło: obliczenia własne na podstawie danych GUS.



Rys. 7. Wykres liniowy, wykres funkcji autokorelacji oraz histogram reszt modelu II

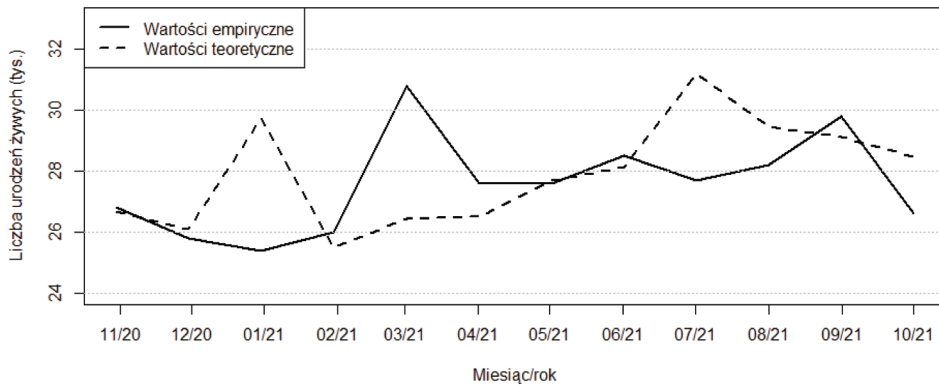
Źródło: obliczenia własne na podstawie danych GUS.

W przypadku modelu I nie ma znaczącej różnicy między wartościami teoretycznymi a wartościami zaobserwowanymi. Dane empiryczne wydają się opóźnione o 3 okresy względem prognozy (rys. 8 oraz 9). Równanie linii trendu dla wartości zaobserwowanych jest równe $y_t = 0,1923x_t + 26,317$. Natomiast dla prognozy jest to $y_t = 0,2796x_t + 26,109$.



Rys. 8. Wykres liniowy dla prognozy modelu I oraz dla danych zaobserwowanych

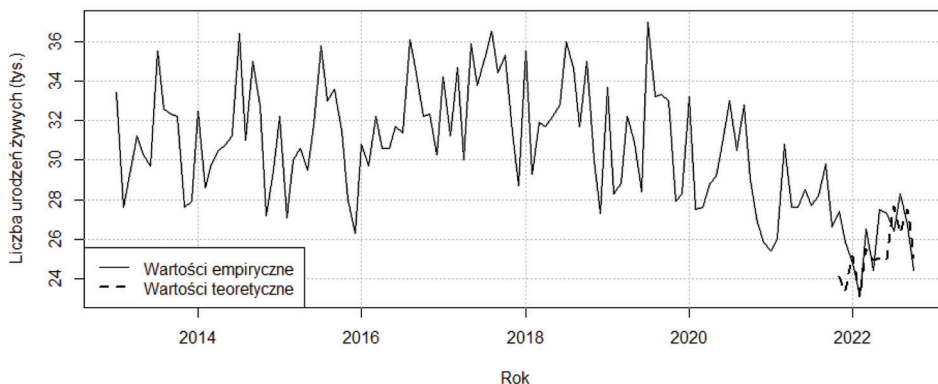
Źródło: obliczenia własne na podstawie danych GUS.



Rys. 9. Wykres liniowy dla prognozy modelu I

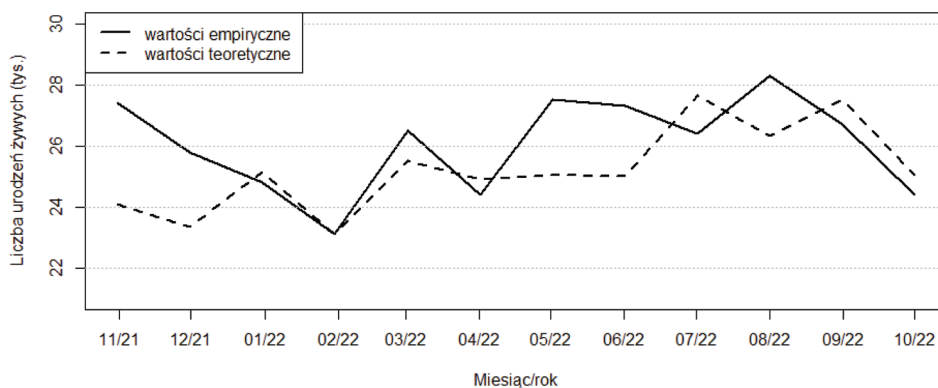
Źródło: obliczenia własne na podstawie danych GUS.

Zdecydowanie gorzej zostały zaprognozowane wartości na podstawie modelu II (rys. 10 oraz 11). Równanie linii trendu dla wartości empirycznych wynosi $y_t = 0,0755x_t + 25,559$. Natomiast dla wartości z prognozy jest to $y_t = 0,2719x_t + 23,455$.



Rys. 10. Wykres liniowy dla prognozy modelu II oraz dla danych zaobserwowanych

Źródło: obliczenia własne na podstawie danych GUS.



Rys. 11. Wykres liniowy dla prognozy modelu II

Źródło: obliczenia własne na podstawie danych GUS.

Błędy *ex post* wskazane na podstawie danych empirycznych dla modelu II są nieznacznie lepsze, niż dla modelu I (tab. 4). Model I przeszacowywał wartości, a model II nie doszacowywał.

Tabela 4. Błędy *ex post* modeli I oraz II

Model	ME	RMSE	MAE	MPE	MAPE
Model I	-0,36	2,18	1,54	-1,58	5,56
Model II	0,83	1,74	1,42	2,98	5,31

Źródło: obliczenia własne na podstawie danych GUS.

6. Zakończenie

Na podstawie analizy można dojść do wniosków wskazujących na to, że modele SARIMA dobrze przewidziały średnią liczbę urodzeń, jednak model I lepiej poradził sobie z prognozowaniem niż model II, który uwzględniał rok pandemii w Polsce. Co ciekawe, w przypadku modelu I dane empiryczne wydają się opóźnione o 3 okresy względem prognozy. Może to oznaczać, że plany prokreacyjne są podejmowane z wyprzedzeniem i pandemia COVID-19 nie zaburzyła planów osób, które wcześniej przygotowały się na posiadanie dzieci. Reakcja społeczeństwa w pierwszych miesiącach pandemii była mniej istotna dla trendów demograficznych, niż można było przypuszczać. Dodatkowo zaobserwowane wartości wyższe niż prognozowane modelem II mogą wskazywać na spełnienie odroczonej planów prokreacyjnych.

Mimo że modele nie uwzględniały wystąpienia wojny w Ukrainie, to wcześniejsze informacje o możliwej agresji Rosji mogły z wyprzedzeniem wpływać na reakcje społeczne. Jest to jedna z wielu trudności analizowania i prognozowania złożonych zjawisk – wpływ społecznych czynników na badane trendy jest trudny do oszacowania.

Dodatkową trudność sprawiło wybranie dokładnego momentu wystąpienia interwencji, co ma wpływ na jakość analizy. Badanie należałoby powtórzyć, wykorzystując bardziej skomplikowane metody, np. transmitancję operatorową (*transfer function*).

Pojawienie się pandemii COVID-19 wywarło istotny wpływ na wiele aspektów życia społecznego i gospodarczego, jednak jej pełne skutki mogą być jeszcze nieznanymi. Chaos i niepewność, które wprowadziła pandemia, mogły być czynnikami zapalnymi łańcucha szeregu zdarzeń, które wpływają na urodzenia w Polsce, a które będą wpływać jeszcze przez lata.

Literatura

- Aassve, A., Cavalli, N., Mencarini, L., Plach, S. i Sanders, S. (2021). Early Assessment of the Relationship between the COVID-19 Pandemic and Births in High-income Countries. *Proceedings of the National Academy of Sciences of the United States of America*, 118(36). <https://doi.org/10.1073/pnas.2105709118>
- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. W: B. N. Petrov i F. Csaki (red.), *Second International Symposium on Information Theory*. Akademiai Kiado.
- Armstrong, J. S. i Green, K. (2015). Simple Versus Complex Forecasting: The Evidence. *Journal of Business Research*, 68(8). <https://doi.org/10.1016/j.jbusres.2015.03.026>
- Beaujouan, E. i Berghammer, C. (2019). The Gap between Lifetime Fertility Intentions and Completed Fertility in Europe and the United States: A Cohort Approach. *Population Research and Policy Review*, (38), 507-535. <https://doi.org/10.1007/s11113-019-09516-3>
- Boef, S. D. i Keele, L. (2008). Taking Time Seriously. *American Journal of Political Science*, (52), 184-200.
- CBOS. (2019). *Preferowane i realizowane modele życia rodzinnego*. Pobrano 4 stycznia 2024 z https://www.cbos.pl/SPISKOM.POL/2019/K_046_19.PDF

- CBOS. (2023). *Bariery zamierzeń prokreacyjnych*. Pobrane 4 stycznia 2024 z https://www.cbos.pl/SPISKOM.POL/2023/K_087_23.PDF
- Committee for the Coordination of Statistical Activities. (2020). *How COVID-19 Is Changing the World: A Statistical Perspective*.
- Degenhardt, L., Day, C., Dietze, P., Pointer, S., Conroy, E., Collins, L. i Hall, W. (2005). Effects of a Sustained Heroin Shortage in three Australian States. *Addiction*, 100(7), 908-920.
- Drozdowski, R., Frąckowiak, M., Krajewski, M., Kubacka, M., Luczys, P., Modrzyk, A., Rogowski, Ł., Rura, P. i Stamm, A. (2020). *Życie codzienne w czasach pandemii. Raport z drugiego etapu badań*. Pobrane 9 stycznia 2024 z https://socjologia.amu.edu.pl/images/pliki/dokumenty/Do_pobrania/Zycie_codzienne_w_czasach_pandemii_Raport_z_drugiego_etapu_badan_wersja_skrocona.pdf
- Gilmour, S., Degenhardt, L., Hall, W. i Day, C. (2006). Using Intervention Time Series Analyses to Assess the Effects of Imperfectly Identifiable Natural Events: a General Method and Example. *BMC Medical Research Methodology*, 6(16) <https://doi.org/10.1186/1471-2288-6-16>
- Hays Poland. (2021). *Kobiety na rynku pracy 2021*. Pobrano 10 lutego 2024 z <https://www.hays.pl/kobiety>
- Hossain, M. (1998). AIC and BIC – The Two Competitive Information Criteria for Model Selection in Economics and Statistics. *The Jahangirnagar Review, Part II: Social Science*, 19, 133-140.
- Hultkrantz, L. i Olsson, C. (1997). Chernobyl Effects on Domestic and Inbound Tourism in Sweden – A Time Series Analysis. *Environmental and Resource Economics*, (9), 239-258.
- Hurvich, C. M. i Tsai, C. L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76, 297-307.
- Kocot-Górecka, K. (2015). Kulturowe role płci i dzietność w Polsce i Norwegii. *Studia Demograficzne*, 2(168), 61-88. <https://doi.org/10.33119/SD.2015.2.4>
- Kotowska, I. E. (2019). Uwagi o urodzeniach i niskiej dzietności w Polsce oraz polityce rodzinnej wspierającej prokreację. *Studia Demograficzne*, 2(176), 11–29. <https://doi.org/10.33119/SD.2019.2.1>
- Kowal, M., Karwowski, M., Coll-Martín, T., Ikizer, G., Rasmussen, J., Lieberoth, A., Eichel, K., Studzińska, A., Koszałkowska, K., Najmussaqib, A., Pankowski, D. i Ahmed, O. (2020). Who is the Most Stressed During the COVID-19 Pandemic? Data from 26 Countries and Areas. *Applied Psychology: Health and Well-Being*, 12(4), 946-966, <https://doi.org/10.1111/aphw.12234>
- Krause, E. (2021). Sytuacja kobiet na rynku pracy w czasie pandemii. *Edukacja Ustawiczna Dorosłych*, 1(112), 159-176. <https://doi.org/10.34866/6gaf-2m32>
- Markiewicz, R. (2019). Kryzys w kryzysie. Psychologiczny wymiar pandemii. *Zdrowie Publiczne*, 129, 138-140. <https://doi.org/10.2478/pjph-2019-0031>
- Matysiak, A. (2006). Sharing Professional and Household Duties within the Polish Couples – Preferences and Actual Choices. *Studia Demograficzne*, 1(147), 122-153.
- Ministerstwo Rodziny i Polityki Społecznej. (2021). *Badania warunków do poprawy kondycji demograficznej Polski*. Pobrano 20 stycznia 2024 z <https://www.gov.pl/attachment/b0c2fb5e-6a0c-4928-92e2-f74ee621ed1b>
- Morrison, K. (2012). Searching for Causality in the Wrong Places. *International Journal of Social Research Methodology*, (15), 15-30.
- Okun, B. S. (2016). An Investigation of the Unexpectedly High Fertility of Secular, Native-born Jews in Israel. *Population Studies*, 70(2), 239-257. <https://doi.org/10.1080/00324728.2016.1195913>
- Osiewalska, B., Matysiak, A. i Kurowska, A. (2022) When Are Women Who Work from Home More Likely to Have Children? *Working Papers*, 13(389). Faculty of Economic Sciences, University of Warsaw.
- Rokosz, M., Poprawa, R., Tabiś, K., Barański, M. i Lewandowska, B. (2021). Analiza zmian w zakresie zaspokojenia i frustracji podstawowych potrzeb psychologicznych oraz stresu i radzenia sobie z nim w pierwszych miesiącach pandemii COVID-19 w 2020 roku wśród kobiet. *Polskie Forum Psychologiczne*, 26(3), 330-348. <https://doi.org/10.34767/PFP.2021.03.06>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6, 461-464.

- Smit, J., Nacer, E., Sikorski, A., Godard, C. i Magdziarz, W. (2023) Consequences for Workers. W: K. Polluveer (red.), *Social and Economic Consequences of COVID-19* (s. 116-136). Policy Department for Economic, Scientific and Quality of Life Policies.
- Smith, J. T., Clarke, R. T. i Saxén, R. (2000). Time-Dependent Behaviour of Radiocaesium: A New Method to Compare the Mobility of Weapons Test and Chernobyl Derived Fallout. *Journal of Environmental Radioactivity*, (49), 65-83. [http://dx.doi.org/10.1016/S0265-931X\(99\)00088-0](http://dx.doi.org/10.1016/S0265-931X(99)00088-0)
- Snopkowski, K. i Kaplan, H. (2015). Fertility Theory: Theory of Intergenerational Wealth Flows. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (s. 35-39). <https://doi.org/10.1016/B978-0-08-097086-8.31074-1>
- Sobotka, T., Jasilioniene, A., Zeman, K., Nemeth, L., Winkler-Dworak, M., Galarza, A. A., Brzozowska, Z. i Jdanov, D. (2021). Booms, Busts and Trend Reversals? Shifts in Births and Fertility Rates across the Highly Developed Countries during the COVID-19 Pandemic [prezentacja]. *Pandemic Babies? The Covid-19 Pandemic and Its Impact on Fertility and Family Dynamics*. MPIDR.
- Strzelecki, P., Saczuk, K. i Kotowska, I. E. (2013). Godzenie posiadania dzieci i pracy zawodowej – opinie o niektórych rozwiązaniach polityki społecznej. W: J. Czapiński, T. Panek (red.), *Diagnoza Społeczna 2013. Warunki i jakość życia Polaków. Raport*. RMS.
- Styrc, M. (2007). Transfery finansowe i opieki w nieformalnych sieciach wsparcia. W: I.E. Kotowska, U. Sztanderska, I. Wójcicka (red.), *Aktywność zawodowa i edukacyjna a obowiązki rodzinne w Polsce w świetle badań empirycznych*. Scholar.
- UNFPA. (2021). *How Will the COVID-19 Pandemic Affect Births? Technical Brief*. Pobrano 5 stycznia 2024 z <https://www.unfpa.org/sites/default/files/pub-pdf/How%20will%20the%20COVID-19%20pandemic%20affect%20births.pdf>
- Vasileiadou, E. i Vliegenthart, R. (2014). Studying Dynamic Social Processes with ARIMA Modeling. *International Journal of Social Research Methodology*, 17(6), 693-708. <https://doi.org/10.1080/13645579.2013.816257>
- Yanovitzky, I. i VanLear, A. (2008). Time Series Analysis: Traditional and Contemporary Approaches. W: A. F. Hayes, M. D. Slater, L. B. Snyder (red.), *The Sage Sourcebook of Advanced Data Analysis Methods for communication research* (s. 89-124). Sage Publications, Inc. <https://doi.org/10.4135/9781452272054.n4>

Intervention Time Series Analysis Based on the Example of Examining the Impact of the COVID-19 Pandemic on Fertility Using SARIMA Models

Abstract: The pandemic could have affected birth rates differently, depending on many social, economic, and health factors. The research problem is to determine the extent to which the COVID-19 pandemic has affected fertility in Poland. The effects of such an event can be lasting, but they can also be short-term. The work analyses the impact of the COVID-19 pandemic on fertility in Poland, using SARIMA models. Data on the number of births comes from the Central Statistical Office (GUS). Model I predicted the number of births better, suggesting that COVID-19 has not impacted that much previously made reproductive decisions. At the same time, for model II the observed values were higher than predicted. The test should be repeated using more complicated methods, e.g. transfer function. The full effects of the COVID-19 pandemic on fertility may not yet be known, suggesting the need for further analysis.

Keywords: fertility, number of births, pandemic, COVID-19, modeling, SARIMA, forecasting, intervention analysis, time series

Marta Stasiak

e-mail: 182375@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

Badanie postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej

DOI: 10.15611/2024.76.5.06

JEL: C38, Q5, Z3

© 2024 Marta Stasiak

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>

Cytuj jako: Stasiak, M. (2024). Badanie postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 86-100). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Dynamiczny rozwój turystyki niesie ze sobą wiele konsekwencji, które negatywnie wpływają na środowisko przyrodnicze oraz społeczność lokalną. W artykule przedstawiono ideę wprowadzenia zrównoważonego rozwoju w sektorze turystycznym, która prowadzi do wsparcia lokalnej gospodarki przy jednoczesnym zachowaniu dobrostanu środowiska naturalnego. Głównym celem jest zbadanie postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej. Dane pierwotne zostały zebrane w 2023 roku za pomocą kwestionariusza internetowego. Do analizy wyników zastosowano analizę korespondencji wielu zmiennych, która wykazała różnicowanie grup badanych osób wobec pytań dotyczących częstotliwości, kierunku i determinant podróŜowania. Analiza czynnikowa pozwoliła na wyodrębnienie siedmiu czynników wskazujących struktury łączące zmienne związane z postawami i zachowaniami podróżujących Polaków.

Słowa kluczowe: koncepcja turystyki zrównoważonej, zrównoważony rozwój w turystyce, zrównoważone podróŜowanie, środowisko przyrodnicze, społeczność lokalna

1. Wstęp

Współcześnie jednym z głównych problemów występujących na całym świecie jest pogarszający się stan środowiska przyrodniczego. Dynamiczny rozwój turystyki powodujący masowy napływ turystów do różnych destynacji, z jednej strony, przyczynia się do degradacji środowiska, a z drugiej – przynosi wiele trudności społecznościom lokalnym. Rozwiązaniem tej sytuacji może być wprowadzanie koncepcji turystyki zrównoważonej, która jest szansą dla rozwoju sektora turystycznego w związku z wieloma korzyściami płynącymi zarówno dla lokalnej gospodarki, jak i środowiska przyrodniczego.

Głównym celem artykułu jest zbadanie postaw i zachowań polskich turystów wobec założeń koncepcji turystyki zrównoważonej. Warto rozważyć, w jaki sposób turystyka wpływa na środowisko przyrodnicze oraz osoby zamieszkujące obszary turystyczne? Jakie są główne zamierzenia koncepcji turystyki zrównoważonej? Czy podróżujący w celach turystycznych Polacy postępują według zasad podróżowania w sposób zrównoważony?

W artykule zawarto opis negatywnego wpływu turystyki na środowisko przyrodnicze i społeczność lokalną oraz wprowadzono pojęcie, cel oraz zasady koncepcji turystyki zrównoważonej. Zaprezentowano opis przeprowadzonego badania, charakterystykę respondentów oraz wybór metody statystycznej. Przedstawione zostały również wyniki badania dotyczącego podróżujących Polaków za pomocą wielowymiarowej analizy korespondencji oraz analizy czynnikowej.

2. Koncepcja turystyki zrównoważonej

2.1. Negatywny wpływ turystyki na środowisko przyrodnicze i społeczność lokalną

Obecnie turystyka pełni istotną rolę w funkcjonowaniu gospodarek na całym świecie. Stanowi ważne źródło koniunktury dla wielu rodzajów działalności gospodarczej, generuje rozwój handlu i infrastruktury oraz przyczynia się do tworzenia dużej liczby miejsc pracy (Alejski, 1999). Z jednej strony, branża turystyczna przynosi wiele korzyści ekonomicznych, natomiast z drugiej strony obserwowany obecnie dynamiczny rozwój turystyki niesie ze sobą wiele zagrożeń.

Jednym z nich jest wpływ na środowisko przyrodnicze z powodu coraz większego, masowego napływu turystów do miejsc charakteryzujących się dużą liczbą walorów przyrodniczych i kulturowych. W 1999 roku The International Council for Local Environmental Initiatives stwierdziło, że negatywne skutki dla środowiska w związku z konsumpcją turystyczną występują na każdym jej etapie: „w miejscu zamieszkania jest to nabywanie dóbr potrzebnych w związku z podróżą, następnie transport między miejscem zamieszkania a miejscem docelowym, wreszcie korzystanie z usług i środowiska miejsca odwiedzanego” (Dziedzic, 2008, s. 48).

Brajcich (2020) podkreśliła, że rozwój środków transportu oraz wzrost stopy życiowej, który skutkował zwiększoną mobilnością społeczeństwa, w latach 2005-2016 przyczynił się do wzrostu emisji gazów cieplarnianych związanych z transportem w turystyce o ponad 60%. W odniesieniu do zakwaterowania należy wspomnieć, że główną przyczyną degradacji środowiska jest emisja CO₂, ale również zużycie energii, wody i innych surowców. Zanieczyszczenie wód również dotyczy wszelkiego rodzaju obiektów turystycznych z powodu zrzutu niebezpiecznych ścieków do gruntu i wód (Kapera, 2018). Kolejnymi problemami występującymi przede wszystkim w obiektach noclegowych świadczących usługi gastronomiczne jest kwestia generowania dużej ilości odpadów oraz marnowania żywności (Brajcich, 2020). Należy również dodać, jak wielkie

szkody dla środowiska przyrodniczego powoduje rozwój infrastruktury turystycznej. Są to między innymi nadmierna konsumpcja zasobów naturalnych, erozja gleby, utrata naturalnych siedlisk i różnorodności biologicznej. Ponadto budowa nowych obiektów turystycznych wiąże się z redukcją terenów zielonych oraz wycinką drzew, której konsekwencją jest duże zanieczyszczenie powietrza z powodu emisji dwutlenku węgla do atmosfery (Baloch i in., 2023). Działania turystów podczas pobytu w miejscach turystycznych także przyczyniają się do degradacji środowiska. Odpady, takie jak plastikowe i szklane butelki oraz puszki, są nieodłącznym elementem na szlakach turystycznych, plażach oraz w zbiornikach wodnych (Zaręba, 2020).

Warto podkreślić, że turystyka oddziałuje nie tylko na środowisko przyrodnicze, ale również środowisko społeczno-kulturowe. Negatywne aspekty dla miejscowej ludności przejawiają się według Zaręby (2020) między innymi w zaniku autentyczności krajobrazu (np. zabytkowej architektury) i kultury regionalnej (np. tradycji, wierzeń, obyczajów), pogorszeniu jakości życia lokalnych mieszkańców, nadaniu kulturze miejscowej komercyjnego charakteru oraz akulturacji, czyli przejmowaniu wzorców zachowań od przyjezdnych. Ponadto warto wspomnieć o niszczeniu i dewastacji materialnych dóbr kultury przez turystów. Istotnym problemem są również stale dokonywane duże inwestycje turystyczne, które przyczyniają się do przejmowania znacznej części dochodów z turystyki przez międzynarodowe koncerny (Zaręba, 2020).

2.2. Pojęcie turystyki zrównoważonej

Dynamiczny rozwój turystyki w ciągu lat skutkujący dużymi szkodami dla środowiska przyrodniczego i społeczno-kulturowego stał się wyzwaniem dla współczesnego społeczeństwa. Konieczne było podjęcie natychmiastowych kroków i działań na rzecz ochrony przyrody. Coraz bardziej popularna i rozpowszechniana na całym świecie stała się idea zrównoważonego rozwoju, czyli wzajemna zależność pomiędzy przyrodą, społeczeństwem a gospodarką, która zapewnia zaspokojenie najważniejszych potrzeb współczesnych pokoleń (Kapera, 2018). Wdrażanie zasad zrównoważonego rozwoju w działach gospodarki zapoczątkowało wdrażanie zrównoważenia również w turystyce. W opracowaniu UNWTO Development Programme (2018, s. 15) powołano się na definicję Światowej Organizacji Turystyki z 2017 roku, zgodnie z którą turystyka zrównoważona to „turystyka, która w pełni uwzględnia jej obecne i przyszłe skutki gospodarcze, społeczne i środowiskowe, zaspokajając potrzeby odwiedzających, branży, środowiska i społeczności lokalnych”.

2.3. Cele i zasady turystyki zrównoważonej

Osiągnięcie stanu równowagi pomiędzy podmiotami zaangażowanymi w rozwój turystyki a środowiskiem przyrodniczym powinno opierać się na postawieniu celów oraz przestrzeganiu konkretnych zasad. Istotne jest ich zaangażowanie oraz aktywne działanie. Głównymi kryteriami i zasadami turystyki zrównoważonej opracowanymi w 2017 roku przez Światową Organizację Turystyki są (Nowacki i in., 2017):

- korzystanie z zasobów środowiska w sposób optymalny;
- przyczynianie się do zachowania dziedzictwa naturalnego i różnorodności biologicznej;
- traktowanie z szacunkiem tradycyjnych wartości społeczności lokalnej;
- wspieranie lokalnego dziedzictwa kulturowego;
- przyczynianie się do międzykulturowego zrozumienia i tolerancji;
- zapewnienie korzyści społeczno-gospodarczych wszystkim interesariuszom;
- przyczynianie się do zmniejszenia ubóstwa na świecie.

Istotne jest również wdrażanie zasad koncepcji turystyki zrównoważonej przez turystów podczas podróży. Powinno to mieć miejsce między innymi podczas wyboru środka transportu, formy turystyki, rodzaju zakwaterowania oraz bazy gastronomicznej. Szczególnie ważne jest zastosowanie działań proekologicznych w trakcie podróży, takich jak segregowanie śmieci, oszczędzanie wody i energii, korzystanie z wielorazowych produktów. Kluczowa dla realizacji wymienionych zasad turystyki zrównoważonej jest również postawa lokalnej społeczności, która powinna wykazać się dużą aktywnością i zaangażowaniem w podejmowaniu inicjatyw związanych z rozwojem turystyki w danym regionie (Durydiwka i in., 2010).

Należy podkreślić, że cele zrównoważonego rozwoju opierają się na trzech głównych aspektach: ekonomicznym, ekologicznym oraz społecznym. Meyer (2011, s. 110), powołując się na opracowanie Niezgody z 2006 roku, przedstawiła trzy podstawowe priorytety rozwoju przemysłu turystycznego w sposób zrównoważony:

- „cel ekonomiczny, który dotyczy zapewnienia dobrobytu ekonomicznego ludności miejscowej oraz utrzymania i optymalnego wykorzystania infrastruktury turystycznej;
- cel ekologiczny, który porusza kwestię zachowania naturalnych zasobów dla potrzeb turystycznych oraz redukcji emisji zanieczyszczeń spowodowanych przez turystykę;
- cel społeczny, który bazuje na osiągnięciach zadowalających możliwości zatrudnienia w sektorze turystycznym, zapewnieniu wypoczynku miejscowym i gościom, ochronie tożsamości kulturowej ludności miejscowej oraz wzroście partycypacji lokalnych mieszkańców w polityce turystycznej”.

3. Badanie postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej

3.1. Cel oraz pytania badawcze

Celem badania była ocena postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej. Przede wszystkim warto zbadać, czy turyści mieszkający w Polsce podróżują w sposób przyjazny dla środowiska oraz jaki mają stosunek do uprawiania turystyki zgodnej z celami zrównoważonego rozwoju. Dodatkowo, opierając się na odpowiedziach na pytania metryczkowe, należy przeanalizować często-

tliwość podróży badanych oraz najczęściej wskazywane przez nich kierunki. Warto uwzględnić również determinanty wpływające na kierunek podróży w podziale na wiek respondentów. Ponadto należy zbadać, na które pytania znajdujące się w kwestionariuszu ankietowani odpowiadają podobnie. Powiązanie zmiennych pozwoli ustalić schemat odpowiedzi badanych polskich turystów na pytania dotyczące częstotliwości wykonywania określonych działań i wyrażania opinii na dany temat.

3.2. Charakterystyka respondentów

Dane zostały zebrane za pomocą kwestionariusza utworzonego w Internecie. Ankieta została rozesłana na grupy podróżnicze na portalach społecznościowych. Badanie przeprowadzono w pierwszym kwartale 2023 roku wśród polskich turystów. Grupa ankietowanych liczyła 269 osób.

W badaniu wzięło udział 77,7% kobiet oraz 21,56% mężczyzn. Osoby, które nie chciały odpowiadać na pytanie o płeć, stanowią 0,74% wszystkich respondentów. Większość badanych była w wieku 40-49 lat (31,97%). Spośród ankietowanych 30,48% było w wieku 18-29 lat, 15,99% w wieku 30-39 lat oraz 13,75% w wieku 50-59 lat. Osoby starsze w wieku 60 lat i więcej stanowiły 7,81% wszystkich badanych turystów. Większość ankietowanych miało wykształcenie wyższe (69,14%). Wykształcenie średnie miało 27,14% respondentów, a zawodowe 2,97%. Natomiast zarówno osoby z wykształceniem innym, jak i podstawowym stanowiły 0,37% wszystkich badanych. W badaniu uczestniczyło 23% respondentów, których miejscem zamieszkania jest miasto od 50 do 100 tysięcy mieszkańców, 22% ankietowanych wskazało miasto powyżej 500 tysięcy mieszkańców, natomiast 20% miasto od 100 do 500 tysięcy mieszkańców. Spośród badanych 19% mieszkało na wsi, a 16% w mieście do 50 tysięcy mieszkańców.

4. Wyniki analizy i wnioski

4.1. Analiza korespondencji

Przeprowadzono wielowymiarową analizę korespondencji¹ dla 267 respondentów. W postępowaniu pominięto dwóch polskich turystów, którzy nie odpowiedzieli na pytanie o płeć. W analizie wykorzystano sześć zmiennych:

- płeć: P1 – kobieta, P2 – mężczyzna;
- wiek: W1 – 18-29 lat, W2 – 30-39 lat, W3 – 40-49 lat, W4 – 50-59 lat, W5 – 60 lat i więcej;
- miejsce zamieszkania: M1 – wieś, M2 – miasto do 50 tys. mieszkańców, M3 – miasto od 50 do 100 tys. mieszkańców, M4 – miasto od 100 do 500 tys. mieszkańców, M5 – miasto powyżej 500 tys. mieszkańców;

¹ Szczegółowy opis algorytmu analizy korespondencji można znaleźć w książce Stanimir (2006).

- częstotliwość podróżowania: I1 – raz na kilka lat, I2 – raz w roku, I3 – 2 razy w roku, I4 – raz w miesiącu, I5 – kilka razy w roku;
- kierunek podróży: K1 – jedno docelowe miejsce w kraju, K2 – jedno docelowe miejsce za granicą, K3 – wycieczki objazdowe po kraju, jedno zakwaterowanie, K4 – wycieczki objazdowe po kraju, kilka zakwaterowań, K5 – wycieczki objazdowe za granicą, jedno zakwaterowanie, K6 – wycieczki objazdowe za granicą, kilka zakwaterowań, K7 – inny;
- czynnik wpływający na kierunek podróży: C1 – atrakcje turystyczne, C2 – klimat, C3 – odległość od miejsca zamieszkania, C4 – środek transportu, C5 – bezpieczeństwo, C6 – cena wycieczki, C7 – dobrze rozwinięta baza noclegowa, gastro-nomiczna, C8 – opinie na temat miejsca, C9 – poprzednie doświadczenia.

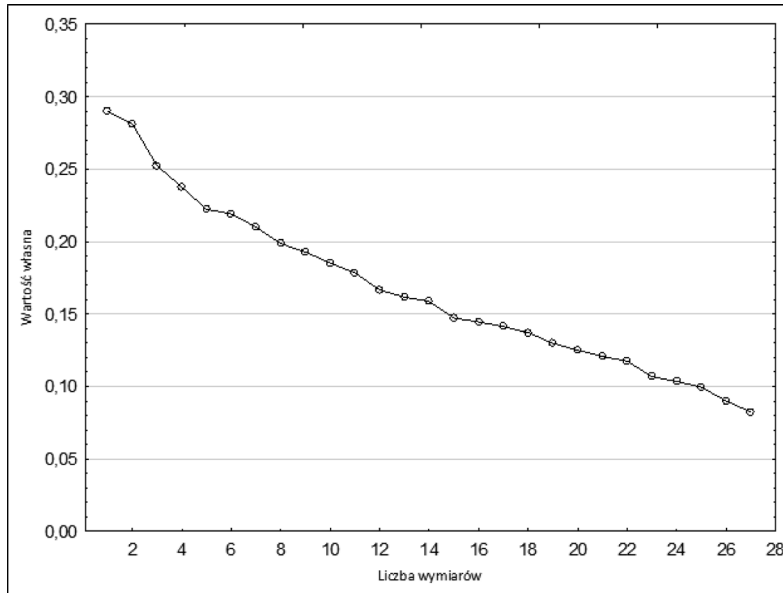
Celem analizy było zbadanie współwystąpienia kategorii wymienionych cech. Połączenie pytań metryczkowych z pytaniami dotyczącymi podróżowania w sposób zrównoważony pozwoli stwierdzić, która grupa badanych osób częściej uprawia turystykę zrównoważoną.

Łączna liczba kategorii wszystkich analizowanych cech wyniosła 33. W związku z tym rzeczywista przestrzeń rzutowania jest 27-wymiarowa. Za pomocą programu Statistica v13.3 obliczono wartości własne oraz stopień wyjaśniania inercji całkowitej dla wspomnianej przestrzeni. Należy wskazać, że inercja całkowita jest równa 4,5. Jakość prezentacji w przestrzeni jednowymiarowej wynosi 6,44%, zatem 6,44% informacji o powiązaniach kategorii analizowanych zmiennych jest prezentowane w tej przestrzeni. Stopień wyjaśniania inercji całkowitej w przestrzeni dwuwymiarowej jest równy 12,69%. Przyrost wyjaśniania rzeczywistych współwystąpień kategorii cech po kolejnym zwiększeniu wymiaru rzutowania nie jest tak znaczący. W związku z tym wyniki analizy należy przedstawić w przestrzeni dwuwymiarowej.

Na podstawie przebiegu krzywej wartości własnych (zob. rys. 1) należy wywnioskować, że prezentacja współwystąpień analizowanych cech powinna zostać wykonana w przestrzeni dwuwymiarowej, ponieważ miejscem zagięcia (uskoku) jest druga wartość własna.

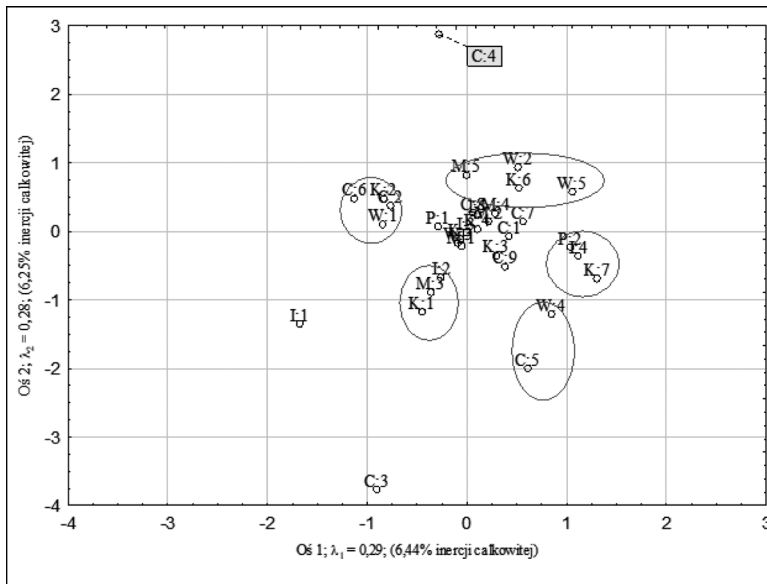
Przedstawiono wyniki analizy korespondencji dla 33 analizowanych kategorii zmiennych (zob. rys. 2). Należy wywnioskować, że kategorie cech, takie jak C4 (środek transportu jako czynnik decydujący o kierunku podróży), I1 (podróżowanie raz na kilka lat) oraz C3 (odległość od miejsca zamieszkania jako czynnik decydujący o kierunku podróży), są odrębne, nie łączą się z innymi kategoriami, ponieważ są odmiennie oceniane przez respondentów.

Na podstawie rozrzutu punktów na rys. 2 należy wskazać, że ankietowani w wieku 18-29 lat często wybierali jedno docelowe miejsce za granicą jako kierunek podróży, a czynnikami, które miały na to wpływ, były klimat oraz cena wycieczki. Obserwując grupę punktów z kategoriami K6, M5, W2 i W5, należy stwierdzić, że badani turyści w wieku 30-39 lat, 60 lat i więcej oraz zamieszkujący w mieście powyżej 500 tysięcy mieszkańców najczęściej jeżdżą na wycieczki objazdowe do innego kraju z możliwością kilku zakwaterowań. Współwystąpienie kategorii I4, K7 oraz P2



Rys. 1. Graficzna prezentacja kryterium łokcia (K=27)

Źródło: opracowanie własne w programie Statistica v13.3.



Rys. 2. Graficzna prezentacja wyników analizy korespondencji zmiennych demograficznych oraz częstotliwości, kierunku i determinant podróży

Źródło: opracowanie własne w programie Statistica v13.3.

wskazuje, że mężczyźni podróżują przeciętnie raz w miesiącu, a ich kierunki podróży są zazwyczaj różne. Należy również zaznaczyć, że ankietowani, którzy wybierają jako kierunek podróży wycieczki objazdowe po kraju z jednym zakwaterowaniem, sugerują się atrakcjami turystycznymi, dobrze rozwiniętą bazą noclegową i gastro-nomiczną w miejscu destynacji oraz poprzednimi doświadczeniami. Głównym czynnikiem wpływającym na kierunek podróży badanych osób w wieku 50-59 lat jest bezpieczeństwo. Ponadto badani polscy turyści zamieszkujący w mieście od 50 do 100 tysięcy mieszkańców najczęściej podróżują średnio raz w roku, wybierając jedno docelowe miejsce w kraju (zob. rys. 2).

Analiza położenia punktów prowadzi również do wniosku, że kategorie cech, które znajdują się blisko początku układu współrzędnych, nie różnicują respondentów (zob. rys. 2).

4.2. Analiza czynnikowa

Wykonano analizę czynnikową² dla 38 zmiennych uwzględniających pięciostopniową skalę Likerta. Celem była redukcja wymiaru danych oraz wskazanie struktur ukrytych łączących poszczególne zmienne. Analizę przeprowadzono dla wszystkich respondentów.

Opisano zmienne wzięte do analizy czynnikowej (zob. tab. 1). Zmienne od q1 do q6 określają częstotliwość wykonywanych działań i zmierzone są na skali Likerta, gdzie 1 oznacza nigdy, a 5 bardzo często. Natomiast zmienne q7-q12 dotyczą wyrażania opinii. Wówczas 1 na skali Likerta jest przypisana wariantowi zdecydowanie nie, natomiast 5 wariantowi zdecydowanie tak.

Tabela 1. Opis zmiennych użytych do analizy czynnikowej

Zmienna	Nazwa zmiennej	Zmienna	Nazwa zmiennej
1	2	3	4
q1-a	Częstotliwość wyboru usług biura podróży	q4-d	Częstotliwość zapoznawania się z obostrzeniami oraz przepisami panującymi w danym miejscu przed podróżą
q1-b	Częstotliwość wyboru usług agenta turystycznego	q5-a	Częstotliwość wybierania się do lokalnych restauracji
q1-c	Częstotliwość dokonywania rezerwacji przez platformy rezerwacyjne	q5-b	Częstotliwość wybierania się do restauracji tematycznych
q1-d	Częstotliwość dokonywania rezerwacji osobistej	q5-c	Częstotliwość wybierania się do sieci <i>fast food</i>

² Szczegółowy opis analizy czynnikowej można znaleźć w książce Sztemberg-Lewandowskiej (2008).

Tabela 1, cd.

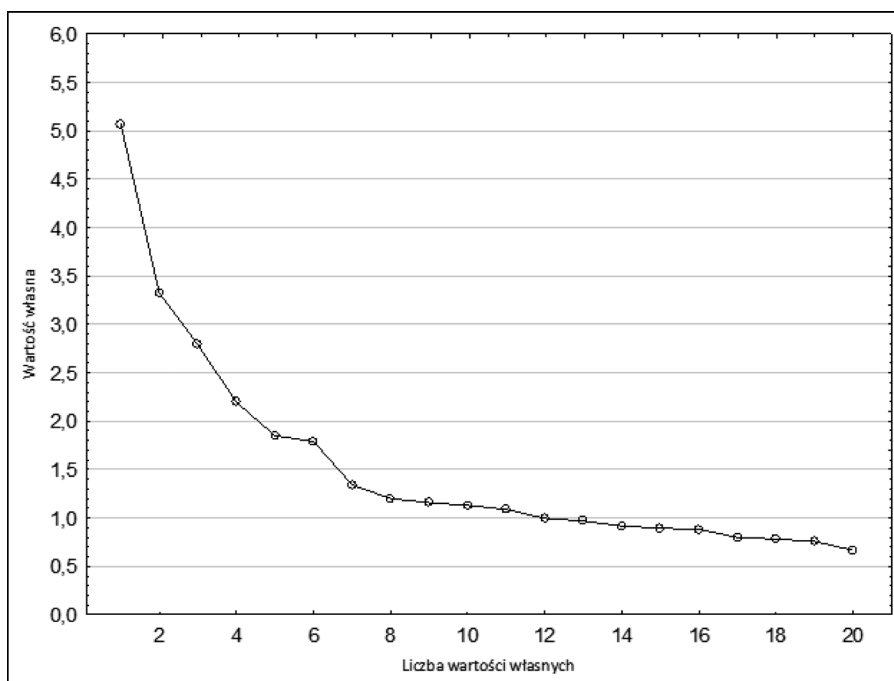
1	2	3	4
q2-a	Częstotliwość podróży samolotem	q5-d	Częstotliwość korzystania z pakietu <i>all inclusive</i>
q2-b	Częstotliwość podróży pociągiem	q6-a	Częstotliwość segregowania w trakcie podróży śmieci
q2-c	Częstotliwość podróży statkiem	q6-b	Częstotliwość korzystania w trakcie podróży z transportu publicznego
q2-d	Częstotliwość podróży samochodem osobowym	q6-c	Częstotliwość odwiedzania w trakcie podróży mniej znanych miejsc
q2-e	Częstotliwość podróży autokarem	q6-d	Częstotliwość próbowania i kupowania w trakcie podróży lokalnie tworzonych produktów
q2-f	Częstotliwość podróży kamperem	q6-e	Częstotliwość korzystania w trakcie podróży z elektronicznych biletów i map
q3-a	Częstotliwość wyboru hotelu	q6-f	Częstotliwość korzystania w trakcie podróży z wielorazowych butelek do picia
q3-b	Częstotliwość wyboru apartamentu	q6-g	Częstotliwość oszczędzania w trakcie podróży wody i energii
q3-c	Częstotliwość wyboru hostelu/ motelu	q6-h	Częstotliwość odwiedzania w trakcie podróży miejsc, w których główną atrakcją są zwierzęta
q3-d	Częstotliwość wyboru pensjonatu, domu agroturystycznego	q7	Zwracanie uwagi na działania obiektu noclegowego na rzecz ochrony środowiska
q3-e	Częstotliwość korzystania z pola biwakowego, kempingu	q8	Skłonność do podróżowania poza szczytem sezonu turystycznego
q3-f	Częstotliwość korzystania z <i>couchsurfingu</i>	q9	Skłonność do wprowadzenia zasad <i>zero waste</i> w trakcie podróży
q4-a	Częstotliwość kupowania nowych ubrań na wyjazd	q10	Odwiedzanie znanych i zatłoczonych miejsc aniżeli mniejszych lokalnych miejscowości
q4-b	Częstotliwość kupowania nowych produktów kosmetycznych na wyjazd	q11	Podróżowanie do tych samych destynacji aniżeli zwiedzanie nowych miejsc
q4-c	Częstotliwość zapoznawania się z miejscowymi zwyczajami i tradycjami przed podróżą	q12	Podróżowanie bardziej odpowiedzialnie, wspieranie lokalnych społeczności oraz szanowanie bardziej przyrody po pandemii COVID-19

Źródło: opracowanie własne.

Wykonano test Barletta oraz zastosowano kryterium KMO w celu zbadania, czy analiza czynnikowa powinna zostać przeprowadzona na zebranych danych. P-value w teście Barletta jest równe 0, w związku z tym hipoteza zerowa o braku korelacji pomiędzy zmiennymi powinna zostać odrzucona. Należy stwierdzić, że pomiędzy zmiennymi zachodzą istotne korelacje. Ogólne kryterium KMO wyniosło 0,715, co oznacza, że analiza czynnikowa jest metodą właściwie wybraną do analizy zebranych danych. Wartość wskaźnika KMO jest mniejsza od 0,5 dla zmiennej q2-e, a dla pozostałych zmiennych otrzymane wartości są dopuszczalne.

Za pomocą metody głównych składowych³ określono oraz wyznaczono liczbę czynników wspólnych. Z powodu dużej liczby zmiennych zredukowano wymiar do 10 czynników. Ze względu na to, że wartości własne były wyższe od jedności dla wszystkich dziesięciu wartości własnych, niemożliwe było podjęcie decyzji dotyczącej wyboru liczby czynników głównych.

Przedstawiono przebieg krzywej wartości własnych w celu wskazania liczby istotnych czynników głównych (zob. rys. 3). Należy wywnioskować, że odpowiednia liczba czynników głównych wynosi 7, ponieważ siódma wartość własna znajduje się w miejscu zagięcia (uskoku).



Rys. 3. Graficzna prezentacja kryterium łokcia (K=20)

Źródło: opracowanie własne w programie Statistica v13.3.

³ Szczegółowy opis metody głównych składowych można znaleźć w artykule Górniaka (1998).

Zaprezentowano ładunki czynnikowe otrzymane po zastosowaniu rotacji varimax znormalizowanej (zob. tab. 2). Przyjęto poziom istotności ładunków czynnikowych równy 0,5.

Tabela 2. Ładunki czynnikowe

Zmienna	Czynnik 1	Czynnik 2	Czynnik 3	Czynnik 4	Czynnik 5	Czynnik 6	Czynnik 7
q1-a	-0,382715581	0,55807832	0,040228755	0,245989834	-0,224973796	-0,139922411	0,346722218
q1-b	-0,081304424	0,498881934	-0,012151779	0,065489481	0,106763747	0,043811749	0,046824615
q1-c	0,756024605	-0,132485354	-0,09123188	0,186563356	0,046546666	0,075617105	-0,067257548
q1-d	0,037308997	0,096611454	0,059608848	-0,517161501	0,324316137	-0,108942699	0,08548014
q2-a	0,146345417	0,196391733	-0,066871933	0,731599466	-0,124094607	0,100504365	0,072506123
q2-b	0,268875678	-0,169753974	0,086896236	-0,069228909	-0,106987501	0,626408145	-0,037844817
q2-c	0,207607496	0,452805198	0,02661817	0,147459105	0,193745018	0,273378698	-0,054832481
q2-d	0,196007039	0,060473984	0,074597497	-0,521875819	0,248361694	-0,445931505	0,128547216
q2-e	-0,164768277	0,119630865	0,152519194	-0,098711691	-0,082328351	0,620675753	0,178667828
q2-f	0,005322656	0,393180615	-0,048988588	0,005072566	0,520415767	-0,05183118	-0,040409111
q3-a	0,209355725	0,428396752	-0,081929063	0,279875871	-0,414774045	-0,198056622	0,301520608
q3-b	0,610660676	0,088297748	0,147870919	0,017883437	-0,047937329	0,022504458	-0,039873491
q3-c	0,181488645	0,133045889	-0,12577867	0,183079991	0,322987416	0,507109603	-0,007724242
q3-d	0,03623406	-0,026881045	0,147058412	-0,335353346	0,351308962	0,124514722	0,075491236
q3-e	0,117128782	0,10729101	-0,004875976	0,001649016	0,667694929	-0,021078454	-0,079330134
q3-f	-0,0082908	-0,032960569	-0,077747839	0,032557144	0,319935467	0,34706243	0,06549102
q4-a	0,112809615	0,113214071	-0,069511919	-0,082632949	-0,079627813	-0,017957052	0,813250252
q4-b	0,05147312	0,011581166	-0,056381946	0,019600888	-0,055311047	0,078589502	0,84870875
q4-c	0,297984014	0,100271505	0,239159214	0,421151253	0,078295814	0,141446882	0,248369852
q4-d	0,320141425	0,125850606	0,293141111	0,387928666	0,081689921	0,121057081	0,244604222
q5-a	0,649172806	-0,072139662	0,06536323	0,105081723	0,105259682	-0,032073763	0,269354169
q5-b	0,607972812	0,145054655	0,052062079	-0,192123126	-0,082490038	0,152396437	0,128479094
q5-c	0,279089597	0,266547552	-0,230778254	-0,385800041	-0,201308791	0,377519741	-0,158291048
q5-d	-0,245816405	0,548340583	5,85112E-05	0,017063602	-0,255937928	-0,182230697	0,350244294
q6-a	0,031166472	-0,124307067	0,646331957	-0,102240594	0,150207286	0,059726555	0,041637748
q6-b	0,348539496	-0,09293204	0,132374187	0,287014509	0,039972949	0,579837599	-0,073906693
q6-c	0,233827042	0,156449858	0,429783529	0,1915371	0,281513633	0,229838147	0,108439674
q6-d	0,246847155	0,020305224	0,418009636	0,285941173	0,165640428	0,059204202	0,400664047
q6-e	0,664436493	-0,161455196	0,168102866	0,172346527	0,182428763	0,067198897	0,037076437
q6-f	0,191582114	-0,084733749	0,478154176	0,262301956	0,364894145	0,23989423	-0,07864766
q6-g	0,257368592	-0,059879313	0,563185494	0,073546301	0,225506691	0,098772911	-0,036758941
q6-h	0,099499738	0,473397275	0,198499411	-0,15454291	0,132997656	-0,051820465	-0,064739585
q7	-0,028224016	0,073198549	0,744487656	-0,136792883	-0,089575566	0,075656635	0,024212749
q8	0,070588091	-0,236637924	0,117249394	0,341583264	0,08446294	-0,141498754	-0,163922617
q9	0,048570742	0,061831577	0,635347163	0,07082517	-0,047255614	-0,208448579	-0,083267657
q10	0,059288069	0,085176119	-0,292492274	0,143913552	-0,423284034	0,160657182	0,146084379
q11	-0,085312018	-0,081108491	0,041151444	-0,545583946	-0,118836425	0,033842824	-0,042337259
q12	-0,009222373	0,141496302	0,625455645	-0,063957098	-0,258245271	-0,025074716	-0,073393025

Źródło: opracowanie własne przygotowane za pomocą programu Statistica v13.3.

Należy wywnioskować, że 24 zmienne mają istotne ładunki czynnikowe. Zmienne q1-b, q2-c, q3-a, q3-d, q3-f, q4-c, q4-d, q5-c, q6-c, q6-d, q6-f, q6-h, q8 oraz q10 należy pominąć, ponieważ nie są powiązane z żadnym czynnikiem (zob. tab. 2).

Pierwszy czynnik główny jest wynikiem redukcji zmiennych: częstotliwość dokonywania rezerwacji przez platformy rezerwacyjne, wybór apartamentu, wybieranie się do lokalnych restauracji, restauracji tematycznych oraz częstotliwość korzystania z elektronicznych biletów i map w trakcie podróży (zob. tab. 2). Wymienione zmienne są ze sobą dodatnio skorelowane. Zatem czynnik pierwszy może dotyczyć przygotowywania się do podróży przy pomocy technologii cyfrowych.

Drugi czynnik główny tworzą zmienne dodatnio ze sobą skorelowane, takie jak częstotliwość wyboru usług biura podróży oraz korzystania z pakietu *all inclusive* (zob. tab. 2). Czynnik można scharakteryzować jako ścisłe powiązanie pomiędzy wyborem biura podróży jako sposobu rezerwacji zakwaterowania oraz korzystania w trakcie wyjazdu turystycznego z opłaconego bufetu w hotelu (*all inclusive*).

Trzeci czynnik związany jest ze zmiennymi, takimi jak częstotliwość segregowania śmieci, oszczędzania wody i energii w trakcie podróży, zwracanie uwagi na działania obiektu noclegowego na rzecz ochrony środowiska, skłonność do wprowadzenia zasad *zero waste* w trakcie podróży oraz podróżowanie bardziej odpowiedzialnie po pandemii COVID-19 (zob. tab. 2). Czynnik może być powiązany z wykonywaniem działań związanych z ochroną środowiska zarówno w życiu codziennym, jak i podczas podróży.

Czwarty czynnik jest wynikiem redukcji zmiennych, takich jak częstotliwość dokonywania rezerwacji osobistej, podróży samolotem, samochodem osobowym oraz podróżowanie do tych samych destynacji aniżeli zwiedzanie nowych miejsc (zob. tab. 2). Im więcej badanych podróżuje samolotem, tym rzadziej korzysta z samochodu osobowego oraz dokonuje rezerwacji osobistej, ale częściej zwiedza nowe miejsca. Czynnik może być związany z wyborem środka transportu oraz takich samych destynacji przez osoby dokonujące rezerwacji osobistej.

Piąty czynnik charakteryzują zmienne dodatnio skorelowane, takie jak częstotliwość podróży kamperem oraz korzystanie z pola biwakowego lub kempingu (zob. tab. 2). Wobec tego czynnik dotyczy wybieranego rodzaju zakwaterowania przez podróżujących kamperem.

Czynnik szósty jest tworzony przez zmienne dodatnio ze sobą skorelowane, takie jak częstotliwość podróży pociągiem, autokarem, korzystanie z transportu publicznego w trakcie podróży oraz częstotliwość wyboru hostelu lub motelu (zob. tab. 2). W związku z tym może być scharakteryzowany jako wybór rodzaju zakwaterowania przez osoby poruszające się niskoemisyjnymi środkami transportu.

Siódmy czynnik związany jest ze zmiennymi dotyczącymi częstotliwości kupowania nowych ubrań i produktów kosmetycznych przed podróżą, które są ze sobą dodatnio skorelowane (zob. tab. 2). Należy stwierdzić, że czynnik łączy działania wykonywane przed wyjazdem w celach turystycznych.

4.3. Podsumowanie badania

Przeprowadzone badanie pozwoliło zrealizować ustalony cel dotyczący oceny postaw oraz zachowań polskich turystów wobec koncepcji turystyki zrównoważonej. Widoczne jest duże zróżnicowanie w odpowiedziach na pytania dotyczące podróżowania w sposób zrównoważony. Należy podkreślić, że nie jest możliwe wskazanie, która z grup częściej uprawia turystykę zgodną ze środowiskiem przyrodniczym i społeczno-kulturowym.

Analiza korespondencji uwidoczniała różnice w odpowiedziach na pytania dotyczące częstotliwości podróżowania, kierunku podróży oraz czynników wpływających na kierunek podróży z uwzględnieniem płci, wieku i miejsca zamieszkania respondentów. Dla osób w wieku 18-29 lat duże znaczenie podczas wyboru kierunku podróży ma klimat i cena, natomiast dla osób w wieku 50-59 lat bezpieczeństwo. Osoby, które często wybierają wycieczki objazdowe do innego kraju z możliwością kilku zakwaterowań, mieszkają w mieście powyżej 500 tysięcy mieszkańców i są w wieku 30-39 lat lub 60 lat i więcej, natomiast osoby wybierające jedno docelowe miejsce w kraju mieszkają w mieście od 50 do 100 tysięcy mieszkańców.

Analiza czynnikowa umożliwiła przedstawienie zależności w postawach i zachowaniach badanych turystów w trakcie podróży. Wyodrębnione czynniki wskazały powiązania w pytaniach dotyczących przygotowania do podróży za pomocą technologii cyfrowych, dodatniej zależności między wyborem biura podróży i pakietu *all inclusive*, wykonywanych działań proekologicznych oraz nastawienia do ekologii, wyboru samolotu do zwiedzania nowych miejsc, wyboru pola biwakowego lub kempingu przez osoby podróżujące kamperem, wyboru hostelu lub motelu przy korzystaniu z niskoemisyjnych środków transportu oraz dodatniej zależności pomiędzy kupowaniem nowych ubrań i produktów kosmetycznych przed podróżą.

Przeprowadzone badanie nie jest reprezentatywne między innymi ze względu na małą liczebność ankietowanych. Nie można uogólnić otrzymanych wyników na wszystkich turystów z Polski. Należy podkreślić, że warto przeprowadzić ankietę na znacznie większej liczbie osób starszych, aby porównać postawy i zachowania respondentów w podziale na kategorie wieku oraz zbadać również inne aspekty dotyczące postępowania zgodnie z zasadami zrównoważonego podróżowania.

5. Zakończenie

Zaprezentowane negatywne skutki dla środowiska przyrodniczego i społeczno-kulturowego wywołane dynamicznym rozwojem turystyki w ostatnich latach pozwoliły na przedstawienie skali światowego problemu. Idea wdrażania zrównoważonego rozwoju w turystyce rozbudza nadzieję na przyszłość, która może przynieść korzyści zarówno dla środowiska, jak i dla lokalnej społeczności. Istotne jest, aby stale uświadamiać, a także zachęcać społeczeństwo do zmiany niektórych nawyków w trakcie podróży. Warto wskazywać, jakie działania na rzecz ochrony środowiska należy podejmować.

Przedstawione w artykule wyniki analizy korespondencji i analizy czynnikowej wskazały, że badani polscy turyści częściowo postępują zgodnie z zasadami turystyki zrównoważonej. W odpowiedziach ankietowanych widoczny jest duży potencjał i chęci do realizowania turystyki przyjaznej dla środowiska oraz społeczności lokalnej. Wyróżnione zostały postawy i zachowania respondentów, które łączą się ze sobą. Pozwala to na znalezienie obszarów, w których konieczna jest zmiana niektórych przyzwyczajzeń oraz rutyn w celu poprawy stanu środowiska i jakości życia miejscowej ludności.

Należałoby zbadać oraz porównać postawy i zachowania w kolejnych latach, aby zauważyć, czy podróżujący Polacy bardziej troszczą się o środowisko przyrodnicze i bardziej wspierają środowisko społeczno-kulturowe podczas wyjazdów w celach turystycznych. Dodatkowo powinno się przeprowadzić analizę wyboru przez ankietowanych ekologicznych biur podróży oraz obiektów noclegowych z certyfikatami zrównoważonego rozwoju. Takie badanie pozwoliłoby szerzej zbadać stosunek turystów do koncepcji turystyki zrównoważonej.

Zrealizowano główny cel artykułu oraz poszczególne cele badawcze. Warto podkreślić, że przeprowadzonego badania nie można uogólnić na całą populację, ponieważ nie jest reprezentatywne, między innymi z powodu nielosowego doboru próby badawczej i jej małej liczebności. Skonstruowany kwestionariusz oraz metody statystyczne wykorzystane w badaniu stanowią adekwatny zbiór narzędzi do dalszej analizy postaw i zachowań polskich turystów wobec koncepcji turystyki zrównoważonej.

Literatura

- Alejski, W. (1999). *Turystyka w obliczu wyzwań XXI wieku*. F.H-U Albis.
- Baloch, Q. B., Shah, S. N., Iqbal, N., Sheeraz, M., Asadullah, M., Mahar, S. i Khan, A. U. (2023). Impact of Tourism Development upon Environmental Sustainability: A Suggested Framework for Sustainable Ecotourism. *Environmental Science and Pollution Research*, 30(3), 5917-5930. <https://doi.org/10.1007/s11356-022-22496-w>
- Brajcich, K. (2020, 8 października). *Carbon Footprint of Tourism*. Sustainable Travel International. <https://sustainabletravel.org/issues/carbon-footprint-tourism/>
- Dziedzic, E. (2008). Ekonomiczne aspekty zrównoważonego rozwoju turystyki. W: S. Wodejko (red.), *Zrównoważony rozwój turystyki* (s. 45-59). Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.
- Durydziwka, M., Kowalczyk, A. i Kulczyk, S. (2010). Definicja i zakres pojęcia „turystyka zrównoważona”. W: A. Kowalczyk (red.), *Turystyka zrównoważona* (s. 21-27). Wydawnictwo Naukowe PWN.
- Górniak, J. (1998). Analiza czynnikowa i analiza głównych składowych. *Ask: Research and Methods*, 7(1), 83-102.
- Kapera, I. (2018). *Rozwój zrównoważony turystyki. Problemy przyrodnicze, społeczne i gospodarcze na przykładzie Polski*. Oficyna Wydawnicza AFM.
- Meyer, B. (2011). Zrównoważony rozwój turystyki i rekreacji. W: A. Panasiuk (red.), *Ekonomika turystyki i rekreacji* (s. 103-111). Wydawnictwo Naukowe PWN.
- Nowacki, M., Awedyk, M., Kowalczyk-Anioł, J., Królikowska K., Pstrocka-Rak, M. i Rak, G. (2017). Strategie rozwoju turystyki w Polsce w kontekście zrównoważonego rozwoju. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (473), 381-389. <https://doi.org/10.15611/pn.2017.473.35>

- Stanimir, A. (2006). Analiza korespondencji. W: A. Stanimir (red.), *Analiza danych marketingowych. Problemy, metody, przykłady* (s. 192-207). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Sztemberg-Lewandowska, M. (2008). *Analiza czynnikowa w badaniach marketingowych*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- World Tourism Organization and United Nations Development Programme [UNWTO Development Programme]. (2018). *Tourism and the Sustainable Development Goals – Journey to 2030*. UNWTO. <https://doi.org/10.18111/9789284419401>.
- Zaręba, D. (2020). *Ekoturystyka* (wydanie 4). Wydawnictwo Naukowe PWN.

Study on Attitudes and Behaviour of Polish Tourists towards the Concept of Sustainable Tourism

Abstract: The dynamic development of tourism has many consequences that negatively affect the natural environment and the local community. The article presents the idea of introducing sustainable development in the tourism sector, which leads to supporting the local economy while preserving the well-being of the natural environment. The main aim is to study the attitudes and behaviour of Polish tourists towards the concept of sustainable tourism. Primary data was collected in 2023 using an online questionnaire. A multivariate correspondence analysis was used to analyse the results, which showed the differentiation of the groups of people surveyed in relation to questions on the frequency, direction and determinants of travel. Factor analysis identified seven factors indicating structures linking individual variables related to the attitudes and behaviours of Polish travellers.

Keywords: concept of sustainable tourism, sustainability in tourism, sustainable travel, natural environment, local community

Martyna Kamińska

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

ORCID: 0009-0003-7223-2409

Uniwersytet Ekonomiczny we Wrocławiu

**Przemoc seksualna w Polsce
w latach 2013-2021: analiza skali
zjawiska i skutków społeczno-
-ekonomicznych przemocy
ze względu na płeć w kontekście UE**

DOI: 10.15611/2024.76.5.07

JEL: I19, I39, J16, Z13

© 2024 Martyna Kamińska

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>

Cytuj jako: Kamińska, M. (2024). Przemoc seksualna w Polsce w latach 2013-2021: analiza skali zjawiska i skutków społeczno-ekonomicznych przemocy ze względu na płeć w kontekście UE. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 101-116). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: W artykule dokonano analizy zjawiska przemocy seksualnej w Polsce w latach 2013-2021 oraz skutków społeczno-ekonomicznych przemocy ze względu na płeć dla kraju i Unii Europejskiej. Zbadano zgłoszone przypadki przestępstw seksualnych w Polsce, opierając się na danych udostępnionych przez Komendę Główną Policji i komendy wojewódzkie. Skupiono się na charakterystyce demograficznej ofiar przemocy seksualnej. Badanie przeprowadzono za pomocą analizy wskaźnikowej. Pozwoliło to na wyciągnięcie ważnych wniosków. Okazuje się, iż ponad 80% ofiar to kobiety. W ciągu badanych lat wszczęto 57 tysięcy postępowań związanych z przemocą seksualną, jednak wiele przypadków pozostaje niezgłoszonych. Wskaźnik wykrycia przestępstw w skali kraju wyniósł 124,8%, co oznacza, że Policja ujawniła więcej przestępstw, niż pierwotnie przewidywano. Najwyższa skuteczność wykrywania wystąpiła w Szczecinie (99,2%) i Gdańsku (94,5%), gdzie również zidentyfikowano ogrom przypadków pornografii. W artykule podkreślono także społeczno-ekonomiczne skutki przemocy, zwracając uwagę na konsekwencje psychologiczne, fizyczne oraz gospodarcze dla krajów UE.

Słowa kluczowe: przemoc, przemoc seksualna, przemoc ze względu na płeć, analiza zjawiska przemocy, skutki przemocy

1. Wstęp

W artykule podjęto temat przemocy seksualnej w Polsce w latach 2013-2021 oraz zdiagnozowano jej skutki społeczno-ekonomiczne zarówno dla Polski, jak i dla Unii Europejskiej. Przygotowano analizę zgłoszonych przestępstw seksualnych w Polsce. Na potrzeby analizy zwrócono się z prośbą do Komendy Głównej Policji [KGP] i komend wojewódzkich policji [KWP] o udostępnienie danych dotyczących przestępstw prze-

ciwko wolności seksualnej i obyczajności (rozdział XXV Kodeksu karnego [k.k.] z lat 2013-2021, w rozbiu na: jednostki administracyjne poszczególnych KWP, artykuły i paragrafy XXV rozdziału Kodeksu karnego, liczbę i płeć poszkodowanych, postępowania wszczęte oraz przestępstwa stwierdzone i wykryte. Dostęp do wymienionych danych możliwy jest jedynie po złożeniu wniosku o dostęp do informacji publicznej. Po złożeniu i zaakceptowaniu wniosku wygenerowane zostały dane statystyczne z Krajowego Systemu Informacyjnego Policji [KSIP] dla poszczególnych jednostek. Dla Polski i każdej z siedemnastu komend wojewódzkich policji w Polsce w latach 2013-2021 wyliczono i zbadano wskaźniki odnoszące się do wymienionych wyżej danych. Pytanie badawcze dla tej analizy brzmi: Jak kształtuje się charakterystyka demograficzna ofiar przemocy seksualnej w rejestrze przestępstw w komendach wojewódzkich policji w Polsce w latach 2013-2021? Artykuł składa się z sześciu punktów. W kolejnym przedstawiono zjawisko i przyczyny przemocy. Punkt trzeci poświęcony jest prezentacji danych, a czwarty wynikom analizy. Następny dotyczy skutków społeczno-ekonomicznych przemocy, a ostatni – szósty – stanowi zakończenie.

2. Zjawisko i przyczyny przemocy

2.1. Definicja przemocy

Zdaniem Borowskiego (b.d., s. 2) „przemoc odziedziczyliśmy po naszych zwierzęcych przodkach”. Zasadniczą różnicą jest to, że zwierzęta stosują przemoc jedynie w dwóch przypadkach – gdy czują się zagrożone lub w celu zdobycia pożywienia. Można zatem powiedzieć, że jest to racjonalny sposób wykorzystywania przemocy. Natomiast ludźmi często kierują zupełnie inne bodźce. Mają skłonności do nadużywania siły.

Przemoc jest zjawiskiem wszechobecnym, dotyka osób z różnych grup społecznych, niezależnie od wykształcenia, wieku czy majątku. Podanie jednej definicji przemocy jest problematyczne, ponieważ wyróżnić można szereg kryteriów określających przemoc, m.in. jej rodzaj, skutki oraz intencje oprawcy (Adamowska-Kałwa, 2018; Jedlecka, 2017). Nie jest ona ani pojęciem klinicznym, ani naukowym. Jest to pojęcie polityczne, co skutkuje zmianą jego postrzegania, wraz ze zmianą sytuacji społecznej (Borowski, b.d.). Ogólnopolskie Pogotowie dla Ofiar Przemocy w Rodzinie (b.d.a) podaje, że przemoc jest związana z zamierzonym działaniem wykorzystującym wyraźną przewagę nad drugą osobą, naruszając tym samym jej prawa i dobra osobiste oraz przyczyniając się do jej cierpienia i szkody na tle psychicznym i fizycznym. Określana jest także jako cicha choroba (Malgesini i in., 2019), która początkowo jest niedostrzegana, a z biegiem czasu niesie coraz większe szkody. Przemoc kojarzy się najczęściej z brutalnością, okrucieństwem oraz przestępczością (Młyński, 2012). Przede wszystkim jednak należy pamiętać, iż przemoc to nie tylko przestępstwo, ale również definitywne pogwałcenie praw człowieka.

Ćwikła (2018) zwraca uwagę na przenikający się znaczeniowo z przemocą czyn, którym jest agresja. Jak podaje literatura przedmiotu, nie należy postrzegać ich tak samo. Autorka wskazuje, że podstawową różnicą między dwoma terminami jest to, iż sama złość nie musi być odbierana jako przemoc. Z kolei agresji zawsze towarzyszy zrównoważona siła pomiędzy agresorem a ofiarą, co jest równoznaczne z tym, że osoba pokrzywdzona ma możliwość skutecznej obrony przed oprawcą.

Przemoc ze względu na płeć, zwana również przemocą płciową, jest jednym z rodzajów przemocy, która jest stosowana w celu podporządkowania, kontrolowania lub zranienia osoby ze względu na jej płeć lub orientację seksualną. Jej źródło doszukuje się w nierównościach między kobietami a mężczyznami (Malgesini i in., 2019). Przemoc wobec kobiet często stosowana jest jako jej termin zamienny, ale nie jest jej synonimem. Deklaracja Organizacji Narodów Zjednoczonych z 1993 roku o eliminacji przemocy wobec kobiet definiuje ją jako podkategorię *Gender – based Violence* [GBV] (Terry, 2007).

Przemoc wobec kobiet (*Violence Against Women*, [VAW]) może przybierać wiele różnych form, m.in. takich jak: gwałt, przemoc domowa, okaleczenie żeńskich narządów, przymusowe małżeństwa, handel kobietami oraz ich wyzyskiwanie. Najczęściej zaś uwidacznia się jako przemoc fizyczna, psychiczna, ekonomiczna, seksualna, a także jako połączenie kilku z wymienionych rodzajów (Malgesini i in., 2019). Wiele kultur ma swoje wierzenia, normy i instytucje społeczne, które legitymizują, a tym samym utrwalają przyzwolenie na przemoc wobec kobiet. Te same akty przemocy, które zostałyby ukarane, gdyby skierowano je do pracodawcy, sąsiada lub znajomego, często zostają bezkarne, kiedy mężczyźni stosują je wobec kobiet, zwłaszcza w rodzinie (Violence, W.I.G.B., 1999).

2.2. Przyczyny przemocy

Do wystąpienia przemocy przyczynia się wiele powiązanych ze sobą czynników. Może ona wynikać np. z traumatycznych przeżyć i międzypokoleniowego przekazu wzorców przemocy. Osoby, które doświadczyły jej w dzieciństwie, są bardziej skłonne do powielania określonych ról w rodzinie, w tym agresji (Decyk, 2022). U źródeł przemocy leżą przede wszystkim ideologie kulturowe i tradycyjny podział ról w społeczeństwie. Poprzez uwarunkowania historyczne przyjęło się, iż głową rodziny powinien być mężczyzna. Helios i Jedlecka (2017) oraz Decyk (2022) zgodnie podkreślają, że w systemie patriarchalnym mężczyzna kontroluje rodzinę i wymusza bezwarunkowe posłuszeństwo, kobieta zaś odpowiedzialna jest za zadania w sferze prywatnej. Kobiety często nie są świadome swojej podległej roli w stosunku do mężczyzny, więc trudno im zmienić wpojony w procesie socjalizacji sposób myślenia o swojej niższej pozycji i wartości oraz odczuć z tym związanych (Zimoń, 2014).

Innym czynnikiem skorelowanym z przemocą jest niski status socjoekonomiczny, w tym ubóstwo i bezrobocie. Osoba bezrobotna w strachu przed oprawcą nie szuka zatrudnienia, a ze względu na brak środków do życia musi na nim polegać.

W ten sposób tworzy się błędne koło zależności finansowej. Kobieta boi się zakończyć krzywdzącą relację, ponieważ pojawia się obawa o byt – swój i swoich dzieci, strach przed przegraną walką w sądzie o opiekę nad nimi oraz przed bezdomnością. Natomiast kiedy kobieta pracuje, a jej partner jest bezrobotny, niekiedy dochodzi do sytuacji, w których agresja ze strony mężczyzny jest wynikiem narastającej frustracji. W końcu stereotypowo to mężczyzna powinien być głównym żywicielem rodziny. Jeśli nie jest w stanie spełnić tej funkcji, pojawia się właśnie agresja (Malgesini i in., 2019).

Duch-Krzystoszek (2009) wskazuje na panujące w społeczeństwie przekonanie, że jedynie mężczyźni, którzy nadużywają alkoholu, są w stanie dopuścić się przemocy wobec swojej rodziny. Naukowcy dawno już stwierdzili, że istnieje korelacja między alkoholem i przemocą. Sprzyja on występowaniu agresji, co z kolei prowadzi do wzrostu spożywania alkoholu także przez ofiary (Fronczyk i Sacewicz, 1999), m.in. w celu ucieczki od bólu – fizycznego czy psychicznego. Nie można natomiast zakładać, że przemoc jest stosowana przez sprawców tylko po spożyciu przez nich alkoholu. Niekiedy przejawia się ona np. jako źródło osobistych nagród. Agresywni i niedowartościowani ludzie również pragną zaznać sukcesu. Niestety czasami ich nagrodą, samodzielnie zdobytą, jest znęcanie się nad najbliższymi. Odbierają zadany ból (niekoniecznie fizyczny) jako honorarium (Decyk, 2022). Wśród sprawców przemocy znajdują się nie tylko bezrobotni i alkoholicy, ale także urzędnicy, lekarze, nauczyciele czy dyrektorzy.

Przemoc towarzyszyć mogą zaburzenia osobowości i psychopatia. Cechuje się ona antyspołeczną agresją, prowadząc do działań mających na celu skrzywdzenie drugiej osoby (Pastwa-Wojciechowska, 2008). Brutalne przestępstwa, m.in. z wykorzystaniem seksualnym, dokonane bez emocji, najczęściej przypisywane są jednostkom psychopatycznym (Gulla, 2020).

Brak dostępu do odpowiedniej edukacji seksualnej skutkuje zmniejszoną świadomością i może prowadzić do niezrozumienia oraz nieposzanowania przestrzeni drugiej osoby, wymuszeń (seks bez zgody obu stron) czy brutalności podczas stosunku seksualnego, a co się z tym bezpośrednio wiąże, może być jednym z czynników prowadzących do stosowania przemocy seksualnej (Czerwiński, 2019).

2.3. Przemoc seksualna

Przemoc seksualna to każdy rodzaj zachowania, który przyczynia się do niechciane go kontaktu w sferze seksualnej. Narusza najbardziej intymną i osobistą sferę życia. Najistotniejszą w niej kwestią nie jest sam seks, ale zaakcentowanie i odzwierciedlenie władzy przez sprawcę, który pokazuje, kto jest silniejszy, a kto powinien ulec (Jedlecka, 2017). Jest to szczególny rodzaj przemocy, który dogłębnie poniża jej ofiarę i odbiera jej godność. Centrum Praw Kobiet zalicza do niej m.in.: niestosowne zwroty o charakterze seksualnym do drugiej osoby, naruszenie nietykalności osobistej poprzez dotykanie miejsc intymnych (nie tylko narządów płciowych, ale także

np. piersi kobiet), zmuszanie do oglądania treści pornograficznych bezpośrednio lub za pośrednictwem np. telefonu (tzw. sexting), szantażowanie lub wymuszanie pieniędzy pod groźbą udostępnienia intymnych zdjęć ofiary (tzw. sextortion) oraz napaść seksualną. Najokrutniejszą formą przemocy seksualnej jest gwałt. Definiowany jest jako zmuszenie do odbycia niechcianego stosunku płciowego lub innego aktu seksualnego bez zgody ofiary. Mogą mu towarzyszyć takie zachowania, jak: przytrzymywanie za kończyny, przygniatanie i duszenie, które mają na celu obezwładnienie ofiary. Nie jest on tożsamy jedynie z biciem czy gwałtownym zachowaniem. Wtedy mowa już o szczególnie nieludzkim oraz okrutnym zachowaniu sprawcy (Polanowski, 2002). Natarczywe naleganie do wysyłania sobie intymnych zdjęć czy nadużycie w formie ściągnięcia prezerwatywy podczas stosunku bez zgody osoby współżyjącej to również przemoc seksualna, której wiele osób nie jest świadomych. Małżeństwo oraz inne formy relacji pod żadnym pozorem nie usprawiedliwiają stosowania przemocy seksualnej (Stoch i Teutsch, 2019). Prawdziwa liczba popełnionych przestępstw na tle seksualnym jest często zaniżana, ponieważ ich ofiary nie zdają sobie sprawy z powagi zaistniałych zdarzeń. Każde działanie mające na celu zaspokojenie potrzeb seksualnych oprawcy wbrew woli ofiary to gwałt (Polanowski, 2002). Ofiara nigdy nie jest mu winna. Gwałt jest okrutnym przestępstwem, niezależnie od tego, czy ofiarą jest młoda dziewczyna, dziewczica czy staruszka. Ofiarą gwałtu może być również osoba świadcząca usługi seksualne, jeśli nie wyraża na nie lub na daną ich formę zgody (Błoch-Gnych, 2011).

Pomimo że ten rodzaj przemocy jest potępiany w społeczeństwie i wywołuje szczególnie oburzenie, to wciąż obserwuje się tendencję wzrostową w liczbie popełnianych przestępstw na tle seksualnym (Jedlecka, 2017). W marcu 2024 roku cała Polska została wstrząśnięta brutalnym wydarzeniem, które miało miejsce w Warszawie. 25-letnia Lizaweta z Białorusi padła ofiarą bestialskiego gwałtu i po kilku dniach walki lekarzy o jej życie zmarła. Przechodnie, którzy byli świadkami tej okrutnej zbrodni, nie zareagowali. To tragiczne wydarzenie skłoniło do refleksji polską dziennikarkę, Staśko (2024), która zadała pytanie: „Czy musimy umrzeć, abyście nas usłyszeli? Czy musimy umrzeć, żebyście uznali gwałt za zbrodnię?”. To pytanie stawia przed nami fundamentalne dylematy moralne, które wymagają pilnej refleksji oraz podjęcia skutecznych działań. Jak podają Helios i Jedlecka (2017), mężczyźni znacznie częściej gwałcą, niż bywają gwałceni. Kobieta jako gwałcicielka to przypadek incydentalny. Autorki zwracają uwagę na niebagatelną różnicę między płcią żeńską i męską. Kobiety najczęściej padają ofiarą gwałtu, a mężczyźni najczęściej tworzą prawo. Kobiety próbują na wszelki sposób uniknąć napaści seksualnych, mężczyźni nie biorą ich nawet praktycznie pod uwagę. Na pytanie, jaka jest prawdziwa skala przemocy seksualnej, mogą odpowiedzieć jedynie badania ilościowe, których forma wiąże się z wieloma ograniczeniami po stronie respondentów. Przemoc seksualna uważana jest za temat drażliwy, krępujący oraz nieodpowiedni do publicznej dyskusji. Dla wielu wciąż stanowi temat tabu, przez co wyniki przeprowadzanych badań oraz zgłaszane przestępstwa są niedoszacowane, co z kolei prowadzi do powsta-

wania rozbieżności. Na przykład badania Agencji Praw Podstawowych (FRA) z 2012 roku pokazują, że w Polsce skala tej formy przemocy jest względnie niska, ponieważ jedynie 19% kobiet doświadczyło przemocy seksualnej lub fizycznej. W Szwecji ten sam wskaźnik wyniósł 46%. Z kolei Fundacja STER w 2015 roku przeprowadziła badanie wśród kobiet, z którego wynikało, iż skala przemocy seksualnej w Polsce jest ogromna, bowiem aż 87,6% kobiet padło ofiarą różnych form molestowania seksualnego, które w polskim prawie karnym nie są nawet zdefiniowane, a nieliczne, m.in. groźby i nękanie, można znaleźć zaledwie w kilku artykułach Kodeksu karnego. Najwięcej ankietowanych – 77,5% – doświadczyło molestowania seksualnego w formie nieprzyzwoitych dowcipów i rozmów o podtekście seksualnym. W przywołanych badaniach zastosowano odmienny sposób zadawania pytań, dobór prób oraz poprzedzono je szkoleniem respondentów, w wyniku czego porównywanie ich wyników jest utrudnione (Babińska i in., 2018).

Z tym rodzajem przemocy można spotkać się w każdym miejscu: w domu, pracy, komunikacji miejskiej, klubie, na ulicy, a także w kościele. Szacuje się, iż odsetek niezgłaszanych zdarzeń przemocy seksualnej oscyluje w granicach 80% (Polanowski, 2002). Na brak raportowania większości przypadków przemocy seksualnej wpływa nie tylko wstyd i strach, ale także możliwość spotkania się z brakiem zrozumienia ze strony policji czy sądu. Niejednokrotnie zgłaszanie przestępstwa oraz przesłuchania przez policję/sąd stają się kolejną traumą dla poszkodowanych.

Przygotowując artykuł, autorka przeprowadziła wywiad z młodą wrocławianką, która padła ofiarą napaści seksualnej. Opowiadając swoją historię, kobieta zwróciła uwagę na fakt, że jedno z pierwszych pytań, jakie usłyszała od funkcjonariuszy, którzy przyjechali na miejsce zdarzenia, dotyczyło jej ubioru. Kiedy ofiara zaproponowała policjantom, by sprawdzili monitoring, usłyszała, że „ma nie bawić się w dektetywa”. Z kolei sędzia, po wysłuchaniu zeznań ofiary, poinstruował ją jedynie, aby nie wracała sama po zmroku. Sprawa została umorzona, a ofiara, zamiast otrzymać wsparcie, została z dodatkową traumą. Z powodu tak wyglądających realiów systemu składania zeznań i szukania sprawcy wiele ofiar przemocy nawet nie próbuje prosić o pomoc.

3. Prezentacja danych

Jak podaje Urząd Statystyczny w Lublinie (b.d.), wskaźnik to zmienna, która może przyjmować różne wartości. W dziedzinie statystyki wykorzystuje się wskaźniki liczbowe jako narzędzia do pomiaru określonych cech rzeczywistych. Te liczby statystyczne reprezentują wartość danej cechy i wynikają z przyporządkowania konkretnym obiektom pewnych wartości liczbowych jako ich miary. Wskaźniki statystyczne są zazwyczaj wyrażane jako parametry opisowe, często w formie stosunków liczbowych (Czyż, 2016).

Do przeprowadzenia analizy wykorzystano dane pozyskane poprzez złożenie wniosku do komend wojewódzkich policji w Polsce. Badane zmienne obejmowały

liczbę postępowań wszczętych, przestępstw stwierdzonych oraz liczbę przestępstw wykrytych dotyczących przemocy seksualnej w Polsce w latach 2013-2021. Postępowanie wszczęte definiowane jest przez KGP jako każde postępowanie wszczęte przez jednostkę organizacyjną Policji. Może zarówno być efektem zawiadomienia złożonego przez osobę pokrzywdzoną, instytucję np. prokuratury, jak i być wynikiem prowadzonej własnej pracy operacyjnej Policji. Jest nim każde zdarzenie, co do którego zachodzi podejrzenie, iż jest przestępstwem. Przestępstwo stwierdzone w zakończonym postępowaniu przygotowawczym odnotowane jest jako złamanie prawa i przyporządkowuje mu się konkretny artykuł i paragraf Kodeksu karnego (k.k.). Przestępstwo wykryte jest to stwierdzone przestępstwo, w którym ustalono przynajmniej jednego podejrzanego i udowodniono mu popełnienie przestępstwa. Postępowania mogą zakończyć się skierowaniem do sądu aktu oskarżenia (czyny wykryte) i są one w statystyce wykazane w rubryce „przestępstwa stwierdzone”. Może być ich znacznie więcej niż wszczętych postępowań, gdyż w trakcie wszczętego postępowania policja może ujawnić szereg przestępstw różnego typu. Sprawca może działać na szkodę wielu osób, a czyn na szkodę każdej z nich liczony jest jako osobny zarzut i tym samym czyn stwierdzony. Nastąpić może także umorzenie postępowania z powodu niewykrycia sprawcy przestępstwa. Ten czyn również odnotowany będzie w rubryce „czyny stwierdzone”, ale wpłynie ujemnie na rubrykę „wykrywalność”. Umorzenie wobec braku znamion przestępstwa bądź ze względu na brak dostatecznych dowodów na popełnienie przestępstwa skutkuje brakiem przyporządkowania tych czynów jako czyny stwierdzone. To sprawy, które zostały wszczęte, ale po przeprowadzeniu postępowania dowodowego nie znaleziono potwierdzenia zaistnienia przestępstwa lub materiał był zbyt słaby na skonstruowanie zarzutów karnych. Zaliczają się one zatem jedynie do postępowań wszczętych.

W skład XXV rozdziału k.k. wchodzi artykuły 197-204 odnoszące się do zgwałcenia lub wymuszenia czynności seksualnej, obcowania płciowego z osobą małoletnią lub doprowadzenia jej do poddania się czynnościom seksualnym. Uwzględnione są tam także seksualne wykorzystanie bezradności bądź niepoczytalności innej osoby, kazirodztwo, prezentacja oraz rozpowszechnianie pornografii. Część rozdziału dotyczy również seksualnego nadużycia stosunku zależności lub krytycznego położenia, zmuszania do uprawiania prostytucji, stręczycielstwa, sutenerstwa i kuplerstwa.

Oznaczenie płci osoby pokrzywdzonej podczas rejestracji przestępstwa nie było obligatoryjne do 2016 roku. Rejestrujący miał do wyboru opcje: kobieta, mężczyzna i nieznana. W przypadku, kiedy nie wybrał żadnej z wymienionych, system przypisywał automatycznie opcję „nie dotyczy”. Wartość ta ze względu na sposób gromadzenia danych może występować również po 2016 roku, ponieważ przestępstwa zarejestrowane w latach wcześniejszych wykazywane są jako przestępstwa stwierdzone dopiero w momencie ich zakończenia.

4. Wyniki analizy

Prezentując wyniki przeprowadzonych analiz, w pierwszej kolejności skupiono się na liczbie wszczętych postępowań dotyczących przemocy seksualnej (tab. 1).

Tabela 1. Dane statystyczne dotyczące postępowań wszczętych oraz stwierdzonych i wykrytych przez Komendy Wojewódzkie Policji w Polsce w latach 2013-2021 przestępstw przeciwko wolności seksualnej i obyczajności

Komenda wojewódzka/stożeczna	Postępowania		Przestępstwa		% S/W	% Wk/W	% Wk/S	% S/P
	wszczęte (W)	stwierdzone (S)	wykryte (Wk)					
Polska	57 080	71 253	65 609	124,8	114,9	92,1	181,3	
Bydgoszcz	3016	2 192	1 940	72,7	64,3	88,5	111,1	
Białystok	1 337	1 000	880	74,8	65,8	88,0	110,0	
Gdańsk	4 742	8 852	8 364	186,7	176,4	94,5	266,6	
Gorzów Wielkopolski	2 051	1 427	1 270	69,6	61,9	89,0	123,0	
Katowice	7 709	5 977	5 251	77,5	68,1	87,9	121,1	
Kielce	1 332	1 162	1 031	87,2	77,4	88,7	129,0	
Lublin	2 900	2 092	1 833	72,1	63,2	87,6	116,3	
Łódź	3 338	2 310	1 980	69,2	59,3	85,7	112,4	
Olsztyn	2 828	1 841	1 583	65,1	56,0	86,0	111,6	
Opole	1 403	983	875	70,1	62,4	89,0	118,0	
Poznań	4 415	3 203	2 856	72,5	64,7	89,2	120,0	
Radom	2 689	2 234	2 016	83,1	75,0	90,2	133,0	
Rzeszów	2 088	1 489	1 279	71,3	61,3	85,9	124,1	
Szczecin	2 923	24 924	24 730	852,7	846,0	99,2	444,5	
Warszawa	4 684	3 008	2 333	64,2	49,8	77,6	132,8	
Wrocław	5 449	3 745	3 113	68,7	57,1	83,1	111,4	
Kraków	4 176	4 814	4 275	115,3	102,4	88,8	161,9	

Źródło: opracowanie własne na podstawie danych z Komendy Głównej Policji oraz komend wojewódzkich policji w Polsce (dane są dostępne tylko na podstawie złożonego wniosku o informacji publicznej).

Z tabeli 1 wynika, że w całej Polsce w latach 2013-2021 wszczęto ich aż 57 tysięcy. Warto podkreślić, iż zdecydowana większość jej przypadków nie jest zgłaszana. Wskaźnik przestępstw stwierdzonych względem postępowań wszczętych wyniósł w skali kraju 124,8%. Świadczy to o ujawnieniu przez Policję większej liczby przestępstw, niż początkowo zakładano w trakcie wszczętych postępowań. Wskaźnik ten przekroczył również wartość 100% dla trzech KWP: w Szczecinie, Gdańsku i w Krakowie.

wie. Dla KWP w Szczecinie jego wartość osiągnęła poziom 852,7%. Wynika to między innymi z faktu, iż w 2019 roku szczecińska Policja zatrzymała 50-letniego mężczyznę, który miał blisko 12 tysięcy plików elektronicznych zawierających dziecięcą pornografię. Sprawca działał na szkodę wielu ofiar, więc usłyszał zarzut odnośnie do każdej z nich (Staszak, 2019). Najniższą wartość odnotowano dla Komendy Stołecznej w Warszawie (64,2%) i w KWP w Olsztynie (65,1%). Następnie wskaźnik przestępstw wykrytych względem postępowań wszczętych wyniósł dla Polski 114,9%. Nie licząc KWP w Gdańsku, Szczecinie i Krakowie, należy stwierdzić, że dla reszty komend odnotowano jego średnią wartość na poziomie 63%. Najniższe jego wartości to 49,8% dla Komendy Stołecznej w Warszawie, 56% dla KWP w Olsztynie i 57,1% dla KWP we Wrocławiu. Średnio tyle procent zgłoszonych zdarzeń przemocy seksualnej w latach 2013-2021 zostało wykrytych na terenie tych komend. Kolejnym obliczonym wskaźnikiem w analizie jest wskaźnik wykrywalności – iloraz liczby przestępstw wykrytych przez liczbę przestępstw stwierdzonych. Wartość tego wskaźnika dla całej Polski osiągnęła poziom równy 92,1%. We wszystkich komendach wojewódzkich jego wartość była wyższa niż 75%. Największą wykrywalność przestępstw na tle seksualnym odnotowuje KWP w Szczecinie (99,2%) oraz w Gdańsku (94,5%). Na terenie obu tych komend wykryto wysoką liczbę przestępstw dotyczących prezentacji i rozpowszechniania pornografii (art. 202 k.k.). Najniższą zaś wartość tego wskaźnika zaobserwowano dla Komendy Stołecznej w Warszawie (77,6%) i KWP we Wrocławiu (83,1%). Analizując dane z tab. 1, można wnioskować, że na jednego pokrzywdzonego przypadło średnio 1,81 przestępstwa stwierdzonego w skali kraju. Dla KWP w Szczecinie przypadło średnio aż 4,44 przestępstwa na jednego poszkodowanego. KWP w Gdańsku odnotowała wartość tego wskaźnika równą 2,66, a KWP w Krakowie na poziomie 1,61. Najmniejszą jego wartość odnotowała KWP w Białymstoku – 1,1. We wszystkich komendach na jednego pokrzywdzonego przypadło średnio więcej niż jedno przestępstwo.

Tabela 2. Udział procentowy liczby postępowań wszczętych na podstawie danego aktu prawnego względem ich ogólnej liczby w XXV rozdziale k.k. w latach 2013-2021

Zgwałcenie (art. 197 § 1-4)	37
Pornografia (art. 202 § 1-4)	9

Źródło: opracowanie własne na podstawie danych z KGP.

Tabela 3. Udział procentowy liczby przestępstw stwierdzonych na podstawie danego aktu prawnego względem ich ogólnej liczby w XXV rozdziale k.k. w latach 2013-2021

Zgwałcenie (art. 197 § 1-4)	17
Pornografia (art. 202 § 1-4)	52

Źródło: opracowanie własne na podstawie danych z KGP.

Dane zawarte w tab. 2 wskazują, że sam artykuł 197 k.k., który uwzględnia nie tylko zgwałcenie ofiary lub uśiłowanie gwałtu, lecz również każde doprowadzenie pokrzywdzonego do poddania się lub wykonania niechcianej czynności seksualnej, stanowił 37% wszystkich wszczętych postępowań. Ten artykuł odnosi się do brutalnej formy przemocy seksualnej, dlatego najczęściej właśnie ofiary tej formy przemo-

cy decydują się na zgłoszenie zaistniałego przestępstwa. Udział postępowań wszczętych dotyczących pornografii (art. 202 k.k.) wyniósł 9% względem ich ogólnej liczby z XXV rozdziału k.k. w latach 2013-2021. Natomiast w przestępstwach stwierdzonych stanowił on już 52% (tab. 3). Na podstawie tych danych można wywnioskować, iż w trakcie postępowań wszczętych dotyczących pornografii stwierdzane były znacznie wyższe liczby przestępstw dotyczących artykułu 202. Udział badanego wskaźnika dla artykułu k.k. dotyczącego zgwałcenia wyniósł zaledwie 17%. Oznacza to, że w porównaniu z wartością z tab. 2 ze zgłoszonych zdarzeń tej formy przemocy seksualnej ich znacznie mniejsza część została uznana za przestępstwo. Przyczyną może być fakt, iż zgromadzone dowody nie były wystarczające do udowodnienia zaistniałego czynu i skonstruowania zarzutów karnych.

Tabela 4. Dane statystyczne dotyczące pokrzywdzonych w postępowaniach wszczętych oraz przestępstwach stwierdzonych i wykrytych przeciwko wolności seksualnej i obyczajności przez komendy wojewódzkie policji w Polsce w latach 2013-2021

Komenda wojewódzka/ Stożeczna	Pokrzywdzeni (P)	Płeć pokrzywdzonego				Udział procentowy kobiet	Udział procentowy mężczyzn
		kobieta	mężczyzna	nieznana	nie dotyczy		
Polska	39 294	29 123	5120	4892	159	74,1	13,0
Bydgoszcz	1 973	1 666	275	27	5	84,4	13,9
Białystok	909	729	167	11	2	80,2	18,4
Gdańsk	3 320	2 473	565	261	21	74,5	17,0
Gorzów Wielkopolski	1 160	945	191	21	3	81,5	16,5
Katowice	4 937	4 105	708	110	14	83,1	14,3
Kielce	901	811	69	16	5	90,0	7,7
Lublin	1 799	1 604	146	41	8	89,2	8,1
Łódź	2 056	1 770	259	25	2	86,1	12,6
Olsztyn	1 650	1 392	211	38	9	84,4	12,8
Opole	833	671	135	22	5	80,6	16,2
Poznań	2 669	2 144	436	56	33	80,3	16,3
Radom	1 680	1 395	250	23	12	83,0	14,9
Rzeszów	1 200	1 001	175	17	7	83,4	14,6
Szczecin	5 607	1 510	272	3814	11	26,9	4,9
Warszawa	2 265	1 749	317	199	0	77,2	14,0
Wrocław	3 361	2 695	551	106	9	80,2	16,4
Kraków	2 974	2 463	393	105	13	82,8	13,2

Źródło: opracowanie własne na podstawie danych z Komendy Głównej Policji oraz komend wojewódzkich policji w Polsce.

Rezultaty przeprowadzonego badania dotyczącego przestępstw na tle seksualnym w Polsce i w poszczególnych KWP (tab. 4 zał. 1) wskazują, że zdecydowaną większość poszkodowanych stanowią kobiety. Ich udział w ogólnej liczbie ofiar przekroczył 80% dla większości komend. Udział procentowy kobiet w ogólnej liczbie poszkodowanych w całej Polsce w latach 2013-2021 wyniósł 74,1%. Wartość tego samego wskaźnika dla mężczyzn była równa 13%. Procentowo najwięcej kobiet padło ofiarą przestępstwa na tle seksualnym na terenie podległym KWP w Kielcach (90%), gdzie udział procentowy mężczyzn jako ofiar wyniósł 7,7%. Dla KWP w Szczecinie odnotowano zaledwie 26,9% poszkodowanych kobiet i 4,9% mężczyzn. W tym przypadku aż w 68% przypisano ofiarom opcję płci nieznanej, co wpłynęło na zaniżenie wyników dla płci żeńskiej i męskiej.

5. Skutki społeczno-ekonomiczne spowodowane przez przemoc

5.1. Skutki społeczne

Przemoc ze względu na płeć jest jednym z najpoważniejszych problemów społecznych, który ma negatywny wpływ na zdrowie i dobrostan jednostek, rodzin, społeczności oraz państw. Niesie za sobą zarówno natychmiastowe, jak i długotrwałe konsekwencje dla zdrowia fizycznego i psychicznego ofiar przemocy i jej świadków. Ma negatywny wpływ na społeczeństwo na wielu płaszczyznach. W mediach przemoc ze względu na płeć poruszana jest najczęściej jedynie w kontekście kobietobójstwa, a takie przypadki, jak samobójstwa, śmierć okołoporodowa czy brak opieki medycznej, spowodowane przemocą, są często pomijane. O konsekwencjach innych niż śmierć ofiary, np. problemach ze zdrowiem fizycznym, psychicznym i reprodukcyjnym, w ogóle się nie wspomina (Malgesini i in., 2019).

Poniżanie, krytykowanie czy bicie skutkuje u ofiar utratą poczucia własnej wartości, bezpieczeństwa, szacunku i zaufania do samego siebie. Takie czyny mogą również skłaniać do myśli samobójczych i niestety w niektórych przypadkach również do ich realizacji. Ofiary przemocy nie umieją lub boją się prosić o pomoc, a winy doszukują się w sobie. Ogólnopolskie Pogotowie dla Ofiar Przemocy w Rodzinie (b.d.b) podaje następujące skutki dla poszczególnych form przemocy. W przemocy fizycznej występuje uszkodzenie ciała oraz choroby psychiczne wynikające z traumatycznych przeżyć ofiary. W przypadku przemocy psychicznej są one szczególnie nasilone. Towarzyszy im również całkowite uzależnienie ofiary od sprawcy, destrukcja poczucia własnej wartości i godności oraz odizolowanie od zewnętrznych źródeł wsparcia. Przemoc ekonomiczna prowadzi do całkowitej zależności finansowej od oprawcy oraz braku zaspokojenia podstawowych potrzeb.

W wyniku przemocy seksualnej dochodzi do obrażeń fizycznych i psychicznych, co narusza intymność i podważa poczucie własnej wartości oraz poziom zaufania. Ofiary tego rodzaju przemocy często izolują się społecznie, odczuwając wzmożoną niechęć do płci sprawcy i obawiając się powtórzenia sytuacji. Strach przed opusz-

zeniem domu oraz traumatyczne wspomnienia towarzyszą ofierze przez długi czas. Nowicka (2016) zauważa, że skutki przemocy ze względu na płeć są długotrwałe. Ofiary, nawet po latach, nadal cierpią z powodu przeżytych traum, co ma negatywny wpływ zarówno na ich życie osobiste, jak i na społeczeństwo. Często doświadczają wstydu i poczucia winy, a także mogą odczuwać złość i żal wobec sprawców i osób, które nie udzieliły pomocy. Świadkowie przemocy również mogą doświadczać poczucia winy i zastraszania, a milczenie społeczne tylko utrwała bierne podejście do problemu.

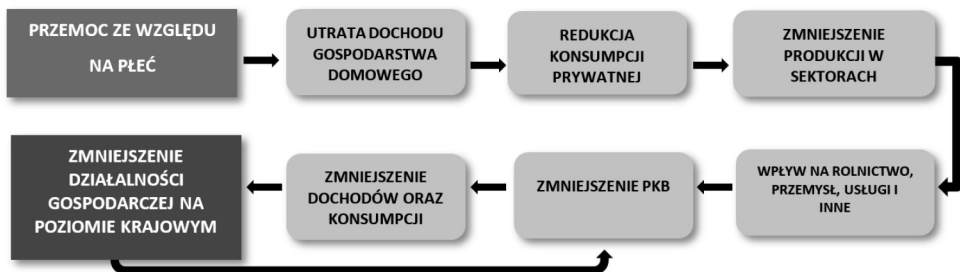
Kolejnym newralgicznym skutkiem przemocy ze względu na płeć jest jej negatywny wpływ na dzieci, które również są ofiarami lub świadkami przemocy. Badania Agencji Praw Podstawowych Unii Europejskiej (2014) wykazały, że aż 73% kobiet doświadczających przemocy potwierdziło, że ich dzieci były świadkami tych zdarzeń. Problemy z dzieciństwa mają długoterminowy wpływ na życie dorosłe, kształtując schematy zachowań i relacje interpersonalne. Dowodem tego są badania Emery'ego (2006), który wykazał, że dzieci będące świadkami przemocy rodziców mogą doświadczać trudności w zachowaniu i zaburzeń osobowości przez całe życie. Wzorce wychowania przekazywane w dzieciństwie mogą prowadzić do agresywnego zachowania lub stosowania przemocy w przyszłości. Ofiary lub świadkowie przemocy często doświadczają problemów zarówno fizycznych, jak i psychicznych, a także mogą mieć trudności w nauce oraz prowadzić do zaburzeń kognitywnych (Malgesini i in., 2019).

Nie jest jednak tak, że każdy sprawca przemocy miał z nią styczność w okresie dzieciństwa i nie każdy, kto jej doświadczył, stosuje przemoc wobec innych osób. Nie jest to związek przyczynowo-skutkowy, lecz pewna zależność występująca z wysokim prawdopodobieństwem. Z tego powodu tak ważne jest, aby zwalczać wszelkie formy przemocy i nie dopuszczać do jej powielania przez kolejne pokolenia. Malgesini i in. (2019) wskazują, że wciąż istnieją społeczeństwa, dla których przemoc to forma tradycji kulturowych oraz sugeruje, iż w UE takimi krajami są m.in. Malta, Rumunia oraz Bułgaria.

5.2. Skutki ekonomiczne

Przemoc ze względu na płeć to nie tylko poważny problem społeczny i humanitarny, który ma wpływ na zdrowie i bezpieczeństwo ofiar, ale także ogromny ciężar ekonomiczny dla całego społeczeństwa. Choć trudno jest oszacować koszty tej przemocy, jest to niezbędne do opracowania skutecznych strategii jej zwalczania. Według danych World Bank (2019), przemoc wobec kobiet może kosztować niektóre kraje nawet 3,7% ich PKB – ponad dwa razy więcej, niż większość rządów wydaje na edukację. W Unii Europejskiej koszty te między 2009 a 2017 rokiem wynosiły średnio 1,9% PKB. Europejski Instytut ds. Równości Kobiet i Mężczyzn (EIGE) szacuje, że przemoc wobec kobiet kosztuje UE 289 miliardów euro rocznie (Shreeves, 2021). W 2019 roku koszty przemocy ze względu na płeć dla krajów UE wyniosły 366 miliardów euro, z czego większość to koszty przemocy przeciwko kobietom (EIGE, b.d.).

Obejmują one m.in. straty produktywności ofiar w pracy, koszty leczenia fizycznych i psychicznych obrażeń, wydatki na system sprawiedliwości oraz koszt pobytu w specjalistycznych ośrodkach dla ofiar. Nie należy zapominać również o indywidualnych kosztach poszkodowanych oraz wszelkich wydatkach mających na celu zwalczanie przemocy i edukowanie o niej. Przemoc ze względu na płeć nie ma ani społecznych, ani ekonomicznych barier. Dotyka osób zarówno z biednych, jak i zamożnych krajów. Według Europejskiej Federacji Krajowych Organizacji Pracujących z Ludźmi Bezdomnymi, przemoc jest jedną z głównych przyczyn bezdomności kobiet, podczas gdy u mężczyzn dominują kłopoty osobiste lub ograniczenia systemu pomocy społecznej. Uciekając przed przemocą, kobiety często tracą swoje dobra osobiste i stają się bezdomne. Przemoc jest także jedną z głównych przyczyn ubóstwa (Malgesini i in., 2019). Ofiary często po ucieczce są zmuszone do podjęcia nisko płatnej pracy, aby zapewnić sobie i swoim dzieciom podstawowe środki do życia. Niektóre z nich, z powodu obrażeń fizycznych lub psychicznych, mogą być całkowicie niezdolne do pracy, co przyczynia się do wzrostu bezrobocia. Na rysunku 1 przedstawiony został efekt mnożnikowy ilustrujący wpływ i przebieg GBV na gospodarkę. Poprzez zubożenie i zmniejszenie konsumpcji zmniejsza się produkcja w różnych sektorach, a następnie PKB, które ma wpływ na dochody obywateli. Z tego powodu przemoc to problem nie tylko społeczny, ale również ekonomiczny.



Rys. 1. Efekt mnożnikowy przemocy ze względu na płeć (GBV) na gospodarkę

Źródło: opracowanie własne na podstawie Wspólnota Narodów (2022).

Współcześnie, w dużej mierze przez pandemię, popularny i często stosowany jest model pracy zdalnej. Niezaprzeczalnie dla wielu osób jest to wygodniejsze rozwiązanie, a dla niektórych stanowi ucieczkę przed molestowaniem seksualnym lub nieprzyzwoitym zachowaniem w miejscu pracy. Kobiety często zmuszone są pracować z domu, aby czuć się bezpieczniej, co może prowadzić do zmiany lub porzucenia pracy. Przemoc ze względu na płeć może również obniżyć inwestycje i rozwój gospodarczy, powodując brak zaufania do rynków i inwestycji. Ofiary przemocy często nie mogą wykorzystać swojego potencjału, co prowadzi do utraty talentów i umiejętności oraz ogranicza rozwój osobisty. Ta strata jest zjawiskiem szkodliwym dla rozwoju gospodarczego, a także dobrobytu społeczeństwa.

6. Zakończenie

Przedmiotem rozważań w artykule była przemoc ze względu na płeć. Jest ona naganą i nieakceptowalną w żadnej formie. Każda osoba ma prawo do poszanowania swojego ciała oraz godności, a GBV jest naruszeniem tych praw. Problem ten dotyka osób na całym świecie, w tym obywateli Unii Europejskiej.

Poszukując odpowiedzi na pytanie badawcze dotyczące charakterystyki demograficznej ofiar przemocy seksualnej w Polsce, na podstawie zgromadzonych danych zidentyfikowano wiele interesujących zagadnień badanego problemu. Udział procentowy kobiet w liczbie ofiar przemocy seksualnej ogółem w Polsce w latach 2013-2021 przekroczył ponad 80% dla większości komend wojewódzkich policji. Badanie przestępstw na tle seksualnym w Polsce wykazało, iż w latach 2013-2021 wszczęto 57 tysięcy postępowań dotyczących przemocy seksualnej. Jednak zdecydowana część jej przypadków nie jest zgłaszana z powodu wstydu, strachu czy obawy przed złym potraktowaniem w czasie procesu sądowego. Stwierdzono, że wskaźnik wykrycia przestępstw w porównaniu do wszczętych postępowań wyniósł 124,8% w skali kraju. Oznacza to, że policja ujawniła większą liczbę przestępstw, niż pierwotnie przewidywano na etapie rozpoczętych śledztw. Warto zauważyć, że wskaźnik ten przekroczył również wartość 100% w trzech Komendach Wojewódzkich Policji: w Szczecinie (gdzie wskaźnik ten osiągnął poziom aż 852,7%!), Gdańsku i Krakowie. Najmniejszą wartość wskaźnika przestępstw wykrytych względem postępowań wszczętych odnotowano dla Komendy Stołecznej w Warszawie – 49,8%. Wyniki wskazały, że Komenda Wojewódzka Policji w Szczecinie oraz Komenda Wojewódzka Policji w Gdańsku odnotowały najwyższą skuteczność w wykrywaniu przestępstw o charakterze seksualnym, odpowiednio 99,2% i 94,5%. W obu tych jednostkach policyjnych zidentyfikowano znaczną liczbę przestępstw związanych z prezentacją i rozpowszechnianiem pornografii.

W pracy poddano diagnozie skutki społeczne i ekonomiczne przemocy ze względu na płeć, zwracając szczególną uwagę na powikłania psychologiczne i fizyczne ofiar przemocy oraz straty gospodarcze krajów UE spowodowanych GBV.

Literatura

- Adamowska-Kałwa, A. (2018). Przemoc w rodzinie. Analiza podstawowych pojęć na podstawie współczesnej literatury. *Studia Teologiczno-Historyczne Śląska Opolskiego*, 38(1), 265-276. [https://doi.org/10.25167/RTSO/38\(2018\)1/265-276](https://doi.org/10.25167/RTSO/38(2018)1/265-276).
- Babińska, M., Bulska, D., Kwaśniewska, A., i Winiewski, M. (2018). *Doświadczenie molestowania wśród studentek i studentów. Analiza i zalecenia*. Biuro Rzecznika Praw Obywatelskich. <https://bip.brpo.gov.pl/sites/default/files/Do%20C5%9Bwiadczenie%20molestowania%20w%20C5%9Br%20C3%B3d%20studentek%20i%20student%20C3%B3w%20C2%202018.pdf>
- Bloch-Gnych, A. (2011). Usługi seksualne. Czy można zgwałcić prostytutkę? *Niebieska Linia*, 1(72), 19.
- Borowski, M. (b.d.). *Przemoc w rodzinie*. Pobrano 22 listopada 2022 z <http://www.korzan.edu.pl/konferencja/referaty/borowski.pdf>
- Czerwiński, M. (2019, marzec 13). *Mamy prawo do edukacji seksualnej!* Amnesty International. Pobrano 28 grudnia 2022 z <https://amnesty.org.pl/mamy-prawo-do-edukacji-seksualnej/>.

- Czyż, T. (2016). Metoda wskaźnikowa w geografii społeczno-ekonomicznej. *Rozwój Regionalny i Polityka Regionalna*, 34, 9-19.
- Ćwikła, M. (2018). Czynniki determinujące agresję i przemoc – podobieństwa i różnice. W: E. Żywucka-Kozłowska, A. Szczechowicz i K. Rzeczkowska (red.), *Problemy współczesnego świata. Zagadnienia wybrane. KPP Monografie* (s. 27-37). http://www.uwm.edu.pl/kpp/files/numery_kpp/kpp_monografie_problemy_wspolczesnego_swiate.pdf#page=33
- Decyk, M. (2022, sierpień 22). *Przyczyny i skutki przemocy w rodzinie*. Czas prawnika. <https://www.facebook.com/czasprawnika/>
- Duch-Krzystoszek, D. (2009). Alkohol i przemoc wobec kobiet – komentarz socjologa o wzajemnych związkach. *Alcohol and Drug Abuse*, 21. http://fnp.ipin.edu.pl/archiwum/2009/materialy/AiN_1-2009.pdf#page=21.
- Emery, R. E. (2006). *The Truth about Children and Divorce: Dealing with the Emotions so You and Your Children Can Thrive*. Penguin.
- European Institute for Gender Equality [EIGE]. (b.d.). A Lack of Evidence to Assess Violence against Women. Pobrano 10 marca 2023 z <https://eige.europa.eu/gender-equality-index/2021/domain/violence>
- Fronczyk, K. i Sacewicz, M. (1999). Alkohol, przemoc i agresja. *Alkohol i Nauka*, (1). <https://psychologia.edu.pl/czytelnia/51-alkohol-i-nauka/251-alkohol-przemoc-i-agresja.html>.
- Gulla, B. (2020). Zaburzenia osobowości. W: B. Gulla, *Gniew, agresja i przemoc w wybranych zaburzeniach psychopatologicznych* (s. 51). Uniwersytet Jagielloński, Biblioteka Jagiellońska. https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/254498/gulla_gniew_agresja_i_przemoc_2020.pdf?sequence=3&isAllowed=y
- Helios, J. i Jedlecka, W. (2017). Współczesne oblicza przemocy. Zagadnienia wybrane. *Prace Naukowe Wydziału Prawa, Administracji i Ekonomii Uniwersytetu Wrocławskiego. Seria: e-Monografie*, 99. https://www.repozytorium.uni.wroc.pl/Content/80015/Wspolczesne_oblicza_przemocy.pdf
- Jedlecka, W. (2017). Formy i rodzaje przemocy. W: M. Sadowski, A. Szychalska, K. Sadowa (red.), *Przemoc w prawie i polityce* (s. 13-31). Wydawnictwo Wydziału Prawa, Administracji i Ekonomii Uniwersytetu Wrocławskiego oraz Wydawnictwo Beta-Druk. https://www.bibliotekacyfrowa.pl/Content/89196/PDF/Wroclawskie_Studia_Erazmianskie_2017_tom_11.pdf#page=14.
- Malgiesini, G., Sforza, L.C., i Babović, M. (2019). *Ubóstwo i przemoc ze względu na płeć w Europie*. Europejska Sieć Przeciw Ubóstwu (EAPN). https://www.eapn.org.pl/eapn/uploads/2020/04/eapn-pl_przemoc-ze-wzgledu-na-plec-w-kontekscie-ubostwa.pdf.
- Młyński, J. (2012). Przemoc w rodzinie – skala zjawiska, ofiary przemocy i formy pomocy pracowników socjalnych. *Studia Socialia Cracoviensia*, 4, 2(7). <https://doi.org/10.15633/ssc.416>.
- Nowicka, K. (2016, styczeń 8). *Jakie są konsekwencje przemocy – czyli o tym, czego doświadczają agresor, ofiara i świadek*. Dompokoju. Pobrano 28 stycznia 2023 z <https://www.dompokoju.org/2016/01/08/jakie-sa-konsekwencje-przemocy-czyli-o-tym-czego-doswiadczaja-agresor-ofiara-i-swiadek/>
- Ogólnopolskie Pogotowie dla Ofiar Przemocy w rodzinie. (b.d.a). *Co to jest przemoc??* Niebieska Linia. Pobrano 22 listopada 2022 z <https://www.niebieskalinia.info/index.php/przemoc-w-rodzinie/6-co-to-jest-przemoc>
- Ogólnopolskie Pogotowie dla Ofiar Przemocy w Rodzinie. (b.d.b). *Rodzaje przemocy*. Niebieska Linia. Pobrano 29 grudnia 2022 z <https://www.niebieskalinia.info/index.php/przemoc-w-rodzinie/8-rodzaj-przemocy>
- Pastwa-Wojciechowska, B. (2008). Samoregulacja i samokontrola a psychopatyczne zaburzenia osobowości. *Czasopismo Psychologiczne*, 14(2), 247-254. <http://www.czasopismopsychologiczne.pl/files/articles/2008-14-samoregulacja-i-samokontrola-a-psychopatyczne-zaburzenia-osobowoci.pdf>
- Polanowski, J. (2002). *Przemoc seksualna*. Niebieska Linia, 1. <https://psychologia.edu.pl/czytelnia/59-niebieska-linia/478-przemoc-seksualna.html>.
- Rozdział XXV. Przepięstwa przeciwko wolności seksualnej i obyczajności. Ustawa z dnia 6 czerwca 1997 roku – Kodeks karny (Dz. U. z 2024 r., poz. 17)

- Shreeves, R. (2021). *Violence against women in the EU: State of play* (Informator polityczny). European Parliamentary Research Service. Pobrano 28.01.2023, z: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698809/EPRS_BRI\(2021\)698809_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698809/EPRS_BRI(2021)698809_EN.pdf).
- Staszak, A. (2019). *Szczecin: Miał na komputerze prawie 12 tys. plików z dziecięcą pornografią. Stanie przed sądem*. Onet Szczecin. Pobrano 28.02.2023, z: <https://wiadomosci.onet.pl/szczecin/szczecin-mial-na-komputerze-prawie-12-tys-plikow-z-dziecieca-pornografia-stanie-przed/jmpp1n6>.
- Staśko, M. (2024). *Maja Staśko: Czy musimy umrzeć, byście nas usłyszeli? Czy musimy umrzeć, żebyście uznali gwałt za zbrodnię?* Wysokieobcasy.pl. Pobrano 18 marca 2024 z <https://www.wysokieobcasy.pl/wysokie-obcasy/7,100865,30757614,maja-stasko-czy-musimy-umrzec-byscie-nas-uslyszeli-czy-musimy.html>
- Stoch, M. i Teutsch, A. (2019). *Bez zgody na przemoc*. Fundacja Autonomia, Ośrodek Badań nad Mediami Uniwersytetu Pedagogicznego im. KEN w Krakowie. <https://autonomia.org.pl/wp-content/uploads/2019/09/AUT-BEZ-ZGODY-NA-PRZEMOC.pdf>
- Terry, G. (2007). *Gender-based Violence*. Oxfam GB. <https://oxfamilibrary.openrepository.com/bitstream/handle/10546/115394/bk-gender-based-violence-300907-en.pdf?sequence=5&isAllowed=y>.
- Urząd Statystyczny w Lublinie. (b.d.). *Wskaźniki i ich podstawowe cechy*. Pobrano 18 marca 2024 z https://lublin.stat.gov.pl/files/gfx/lublin/pl/defaultstronaopisowa/1244/1/1/wskazniki_i_ich_podstawowe_cechy.pdf
- Violence, W.I.G.B. (1999). Ending violence against women. *Issues in World Health*, 11, 1-44.
- World Bank. (2019, 25września). *Gender-Based Violence (Violence Against Women and Girls)*. Pobrano 10 marca 2023 z <https://www.worldbank.org/en/topic/socialsustainability/brief/violence-against-women-and-girls>
- Wspólnota Narodów. (2022). *Measuring the Economic Costs of Violence Against Women and Girls. Facilitator's Guide*. Sekretariat Wspólnoty Narodów. https://production-new-commonwealth-files.s3.eu-west-2.amazonaws.com/s3fs-public/2022-06/Facilitator%E2%80%99s%20Guide%20to%20Measuring%20the%20Economic%20Costs%20of%20VAWG_Jun22.pdf?VersionId=3Pbab6tqp3yilwCqQFctYRcrvI0M7W40.
- Zimoń, K. (2014). Klucz do przeciwdziałania przemocy. *Niebieska Linia*, 4(93), 29-30. <https://docplayer.pl/17741005-4-93-2014-algorytmy-dla-oswiaty-i-ochrony-zdrowia-dobro-dziecka-a-prawo-rodzicow-do-rozvodu-alimenty-obowiazek-i-egzekwowanie.html>

Sexual Violence in Poland in 2013-2021: Analysis of the Scale of the Phenomenon and the Socio-economic Impact of Gender-based Violence in the EU Context

Abstract: The article examines sexual violence in Poland from 2013 to 2021 and its socio-economic impact on the country and the EU. Cases of sexual crimes were analysed using data from the Main Police Headquarters and Provincial Police Headquarters. Over 80% of victims were found to be women. Despite initiating 57,000 proceedings, many cases go unreported due to various concerns. The crime detection rate at the national level was 124.8%, indicating that the police revealed more crimes than originally anticipated. The effectiveness of detecting sexual violence crimes was highest in the Provincial Police Headquarters in Szczecin (99.2%) and Gdańsk (94.5%), where a significant number of cases related to pornography were also identified. The article also underscores the social and economic consequences of gender-based violence, highlighting the psychological, physical, and economic repercussions for EU countries.

Keywords: violence, sexual violence, gender-based violence, analysis of the phenomenon of violence, consequences of violence

Julia Dymek

e-mail: 181441@student.ue.wroc.pl

ORCID: 0009-0008-0600-8552

Uniwersytet Ekonomiczny we Wrocławiu

Badanie poziomu realizacji założeń cyberbezpieczeństwa firm i instytucji Unii Europejskiej

DOI: 10.15611/2024.76.5.08

JE: C38, K24

© 2024 Julia Dymek

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>

Cytuj jako: Dymek, J. (2024). Badanie poziomu realizacji założeń cyberbezpieczeństwa firm i instytucji Unii Europejskiej. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 117-131). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Cyberbezpieczeństwo jest bardzo ważnym tematem, o którym coraz częściej się mówi oraz słyszy. Zyskało ono na znaczeniu wraz z rozwojem technologicznym. Celem artykułu jest zbadanie sytuacji w obszarze cyberbezpieczeństwa w małych i średnich przedsiębiorstwach krajów Unii Europejskiej oraz omówienie form zabezpieczenia firm przed cyberzagrożeniami. W tym celu przeprowadzono dwutorową analizę za pomocą klasyfikacji hierarchicznej oraz porządkowania liniowego ważoną metodą sum standaryzowanych. Zaczepnięte dane pochodzą z badania przeprowadzonego przez Ipsos European Public Affairs. Analiza i wynikające z niej wnioski potwierdziły hipotezy o tym, że bogatsze i bardziej rozwinięte kraje znajdują się w jednym skupieniu oraz państwa Europy Zachodniej i Północnej będą plasowały się wysoko w utworzonych rankingach.

Słowa kluczowe: cyberbezpieczeństwo, cyberprzestępstwo, łańcuch bloków, klasyfikacja hierarchiczna, porządkowanie liniowe

1. Wstęp

Rozwój technologiczny oraz rewolucja cyfrowa są nieodłącznymi elementami współczesnego świata. Wskutek pandemii COVID-19 pojawiło się więcej możliwości załatwienia spraw służbowych oraz prywatnych przez Internet. Jednakże wraz z postępem technologicznym zaczęło zyskiwać na znaczeniu cyberbezpieczeństwo. Zjawisko z nim powiązane to cyberprzestępstwo. Szafranek (2021) na podstawie przeprowadzonych badań i opublikowanych przez CERT Polska raportów stwierdził, iż w latach 2010-2020 można zauważyć wzrost liczby zgłaszanych cyberprzestępstw. Najczęstsze rodzaje przeprowadzanych ataków to między innymi: włamania do sieci wewnętrznej Intranet, phishing, ataki na serwery, kradzieże tożsamości, spam czy ataki szkodliwego oprogramowania (Maciejowski, 2004, za: Szafranek 2021).

W związku z tym w niniejszym artykule postanowiono zgłębić temat związany z cyberbezpieczeństwem. Za cele obrano zbadanie w głównej mierze tego, jak wygląda sytuacja w małych i średnich przedsiębiorstwach krajów Unii Europejskiej w tym obszarze oraz w jaki sposób można zabezpieczyć firmę przed cyberzagrożeniami i wynikającymi z nich skutkami.

Przeprowadzono analizy z wykorzystaniem klasyfikacji hierarchicznej oraz porządkowania liniowego. Wskazane metody wybrano, by określić, które kraje Unii Europejskiej są podobne do siebie pod względem badanych zmiennych odzwierciedlających sytuację panującą w małych i średnich przedsiębiorstwach w obszarze cyberbezpieczeństwa. Przystępując do badania, postawiono następujące hipotezy badawcze:

- *H1*: bogatsze i bardziej rozwinięte kraje stanowią spójną grupę pod względem cyberbezpieczeństwa,
- *H2*: kraje Europy Zachodniej i Północnej plasują się wysoko w rankingach ze względu na analizowane zmienne.

Do badania wzięto pod uwagę zmienne dotyczące:

- poziomu obaw pracowników wobec cyberataków,
- podejmowanych działań, jak na przykład przeprowadzenie szkolenia wśród zatrudnionych,
- tego, w jakim stopniu są oni zorientowani w kwestii niebezpieczeństw wynikających z cyberprzestępczości.

2. Cyberbezpieczeństwo – definicja i inne jego aspekty

Dział cyberbezpieczeństwa Departamentu Bezpieczeństwa Wewnętrznego w Stanach Zjednoczonych opisuje cyberbezpieczeństwo jako pewien proces, zdolność, stan, który przyczynia się do tego, iż wszelkie informacje i systemy komunikacyjne oraz wszystkie dane w nich zawarte są zabezpieczone w razie zniszczenia, modyfikacji czy nieprawego wykorzystania przez osoby trzecie (Bay, 2016).

W cyberbezpieczeństwie dużą rolę odgrywa sformułowane w 1965 roku przez Gordona Moore'a, współzałożyciela firmy Intel, prawo Moore'a (Cunningham, 2021). Dzięki obserwacjom stale rozwijających się trendów zdołał on oszacować tempo rozwoju nowoczesnej cyfrowej rewolucji. Należy również zaznaczyć, iż firma Intel (b.d.) stwierdziła, że prawo to uzyskało miano złotej zasady przemysłu elektronicznego. Pierwotnie odnosiło się ono do tempa podwajania się liczby tranzystorów w układach scalonych. Jednakże wykazuje też, że zapotrzebowanie na wzrost mocy obliczeniowej zmienia się z biegiem czasu zgodnie z trendem wykładniczym (Cunningham, 2021). Moc obliczeniowa dotyczy wydajności danego komputera. Zgodnie z fizycznym wzorem na moc można dowiedzieć się, jaką ilość pracy wykonał komputer w określonym czasie (Ghosh, 2012). Cunningham (2021) stwierdził, iż w związku z postępującym rozwojem technologii występuje coraz więcej zagrożeń w cyberprzestrzeni.

Cyberprzestępcy atakują przedsiębiorstwa zarówno małe, średnie, jak i duże. Warto przytoczyć incydent dotyczący firmy, która zajmuje się marketingiem oraz tworzeniem marki i zatrudnia 25 osób. Jeden z pracowników padł ofiarą phishingu. Wskutek tego cyberprzestępcy przejęli jego konto mailowe i wysłali z niego wiadomości do klientów. Jedne z nich miały na celu dalsze przeprowadzanie ataków phishingowych, a pozostałe dostarczały fałszywy komunikat o tym, iż firma zmieniła dane dotyczące konta bankowego, na które należy przelewać należności. Jednakże podane w wiadomości konto należało do przestępców. Jedno z przedsiębiorstw, które korzystało z usług opisywanej organizacji, wykryło nieprawidłowość i zgłosiło, że otrzymane e-maile są nieprawdziwe. W ostateczności klient zerwał umowę i wszystkie nadchodzące przelewy pieniężne opiewające nawet na 300 tysięcy euro zostały unieważnione (European Union Agency for Cybersecurity, 2021).

Celem cyberprzestępców stały się małe firmy, ponieważ zatrudniają niewielką liczbę pracowników. Dodatkowo osoby pełniące ważne funkcje w firmie oraz administratorzy są bardzo często przepracowani, co wpływa negatywnie na ich koncentrację i może skutkować brakiem zachowania ostrożności w podejmowaniu decyzji. Ponadto z racji tego, iż przedsiębiorstwa te korzystają ze starszych rozwiązań technologicznych, niż są aktualnie dostępne, firmowe sieci nie są w odpowiedni sposób skonfigurowane i zabezpieczone (Cunningham, 2021). Ważny jest też fakt, iż małe i średnie firmy niejednokrotnie nawiązują współpracę z większymi przedsiębiorstwami przykładowo stają się one ich dostawcami, partnerami czy też podwykonawcami podczas realizacji konkretnego procesu (Cunningham, 2021; Williamson, 2014, za Choi i Allison, 2017). W związku z tym cyberprzestępcy, włamując się na konta mniejszych przedsiębiorstw, mogą również uzyskać nieautoryzowany dostęp do poufnych informacji czy danych dotyczących większych firm.

Warto również wspomnieć o kosztach związanych z cyberbezpieczeństwem ponoszonych przez przedsiębiorstwo. Można wyróżnić koszty pośrednie oraz bezpośrednie. Koszty pośrednie to takie, których wartość można przewidzieć i dotyczą one zapobiegania atakom ze strony hakerów. Zaliczane są do nich między innymi: szkolenia pracowników z zakresu cyberbezpieczeństwa, ubezpieczenie od ryzyka cybernetycznego, doradztwo specjalistyczne, wynagrodzenia pracowników dbających o cyberbezpieczeństwo, zakup licencji na programy antywirusowe oraz koszty związane z bezpieczeństwem danych. Koszty bezpośrednie wiążą się z naprawą szkód wywołanych cyberprzestępczością. Do tej kategorii kosztów należą: koszty związane z ochroną konsumenta w momencie wycieku danych osobowych, opłata za usługi prawne, koszt odzyskania utraconych danych, naruszenie wizerunku, utrata środków pieniężnych w wyniku włamania na konto bankowe firmy oraz utrata własności intelektualnej (Antczak, 2020).

Organizacje są zmuszone szukać nowych rozwiązań, które pozwolą zabezpieczyć firmę przed atakami cyberprzestępców (Szafranek, 2021). Niezmiernie istotną kwestią jest technologia *blockchain*. Jest to łańcuch bloków, w których zapisywane są różnego rodzaju dane, między innymi dotyczące transakcji handlowych (Bambara

i Allen, 2018). Dane są niezmiennie, nieodwracalne i zdecentralizowane, czyli rozproszone nawet po całym świecie (Mathew, 2019). Nie są one gromadzone na jednym komputerze, tylko na wielu urządzeniach. Każdy blok, oprócz tego, że zawiera dane, charakterystyczny dla niego hash, to ma również hash bloku poprzedniego. (Barbara i Allen, 2018; Mathew, 2019;). To oznacza, że dysponuje również informacjami o bloku poprzednim. Hash można określić jako pewien ciąg znaków, dzięki któremu można zaszyfrować ważne komunikaty. Jest to zaleta technologii *blockchain*, ponieważ w przypadku, gdy osoba trzecia będzie chciała zmodyfikować dane znajdujące się w danym bloku, to spowoduje, iż zostanie zaburzona integralność łańcucha (Mathew, 2019). Warto także scharakteryzować dwa systemy, które mogą pomóc w walce z cyberprzestępczością. Pierwszy z nich to system wykrywania włamań *Intrusion Detection System* (IDS). Jest on zaliczany do pasywnych systemów bezpieczeństwa, ponieważ nie podejmuje on żadnych czynności celem ochrony przed nieprawym dostaniem się osób trzecich czy złośliwych oprogramowań do systemów (Choi i Allison, 2017). Stosuje się go do obserwowania aktywności w sieci. Drugi również ważny system to system zapobiegania włamaniom – *Intrusion Prevention System* (IPS). Jego zadaniem jest zidentyfikowanie zagrożeń w sieci oraz podejmowanie kroków, aby ochronić sieć (Chakraborty, 2013).

3. Metodyka badań

3.1. Charakterystyka wykorzystanych zmiennych

Zbiór danych, z którego skorzystano w przeprowadzonej analizie, powstał na podstawie badania, które zostało zlecone przez Komisję Europejską i Dyрекcyję Generalną do spraw Polityki Regionalnej i Miejskiej. Wskazane badanie było koordynowane przez Dyрекcyję Generalną do spraw Komunikacji. Przeprowadzone zostało w formie wywiadów telefonicznych między 26 listopada 2021 r. a 17 grudnia 2021 r. przez Ipsos European Public Affairs. Zbadano małe i średnie przedsiębiorstwa w krajach Unii Europejskiej, w których liczba pracowników jest mniejsza niż 250. Firmy te działają w następujących obszarach: produkcyjnym, usługowym, sprzedaży detalicznej oraz przemysłowym. Rozmowy telefoniczne prowadzone były z najważniejszymi osobami w przedsiębiorstwie, między innymi z: głównym menadżerem, dyrektorem finansowym, prezesem zarządu czy menadżerem sprzedaży i marketingu (European Commission, 2022).

Do przeprowadzenia analizy wzięto pod uwagę następujące zmienne:

- X_1 – procent firm, w których pracownicy są w ogóle niepoinformowani o zagrożeniach związanych z cyberprzestępczością,
- X_2 – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, jest bardzo zaniepokojona wirusami, *spyware* lub *malware* podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,

- X_3 – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, jest bardzo zaniepokojona phishingiem, przejęciem konta lub podszywaniem się podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,
- X_4 – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, jest bardzo zaniepokojona ransomware podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,
- X_5 – procent firm, w których w ciągu ostatnich 12 miesięcy firma nie przeprowadziła szkolenia pracowników lub nie podniosła świadomości na temat zagrożeń związanych z cyberprzestępczością,
- X_6 – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, jest bardzo zaniepokojona zhakowaniem internetowego konta bankowego firmy podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,
- X_7 – procent firm, w których pracownicy są bardzo dobrze poinformowani o zagrożeniach związanych z cyberprzestępczością,
- X_8 – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, w ogóle nie jest zaniepokojona wirusami, *spyware* lub *malware* podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,
- X_9 – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, w ogóle nie jest zaniepokojona phishingiem, przejęciem konta lub podszywaniem się podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,
- X_{10} – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, w ogóle nie jest zaniepokojona *ransomware* podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej,
- X_{11} – procent firm, w których w ciągu ostatnich 12 miesięcy firma przeprowadziła szkolenie pracowników lub podniosła świadomość na temat zagrożeń związanych z cyberprzestępczością,
- X_{12} – procent firm, w których najważniejsza osoba, pełniąca funkcje decyzyjne, w ogóle nie jest zaniepokojona zhakowaniem internetowego konta bankowego firmy podczas korzystania z Internetu, wykonując działania związane z biznesem, takie jak sprzedaż towarów lub usługi bankowości internetowej.

W przeprowadzonym badaniu skorzystano z ankiety, w której zastosowano skalę Likerta. W związku z tym w niektórych przypadkach można było wybrać jedną z następujących odpowiedzi: bardzo zaniepokojony, nieco zaniepokojony, w ogóle nie zaniepokojony bądź nie wiem. Były również pytania, na które należało odpowiedzieć „tak” bądź „nie”. W związku z powyższym przeprowadzono dwutorową analizę zgromadzonych danych. Utworzone zostały dwa ich zbiory, które przed-

stawiają zupełnie skrajne sytuacje dotyczące cyberbezpieczeństwa. W pierwszym zbiorze wzięto pod uwagę zmienne od X_1 do X_6 , które są destymulantami, czyli wyższe wartości wskazują, iż poziom analizowanego zjawiska, czyli sytuacji, jaka panuje w firmach odnośnie do cyberbezpieczeństwa, jest gorszy (Stanimir, 2006). W drugim rozpatrywano cechy od X_7 do X_{12} , które są stymulantami, czyli przyjmowanie wyższych wartości ma pozytywny wpływ na zjawisko, które podlega badaniu (Stanimir, 2006).

3.2. Metody analityczne zastosowane w badaniu

Na samym początku dokonano wstępnej analizy danych z wykorzystaniem statystyk opisowych. Za pomocą współczynnika zmienności określono, czy dana zmienna jest *quasi*-stałą, czy też nie. Jeżeli przyjmuje on wartość powyżej 10%, to można uznać, iż jest on istotny i dana zmienna wykazuje zróżnicowanie (Ostasiewicz i in., 2006).

W następnej kolejności dokonano analizy skupień, korzystając z klasyfikacji hierarchicznej. Pozwala ona pogrupować badane obiekty, które są homogeniczne, czyli do siebie podobne, pod względem zmiennych uwzględnionych w analizie (Koleda, 2009, za: Prus i Król 2017). W związku z tym różnice między obiektami, które są w jednej grupie, są mniejsze niż pomiędzy obiektami pochodzącymi z różnych skupień. Aby skorzystać z tej metody, należy na samym początku znormalizować dane, np. za pomocą standaryzacji. Bardzo ważnym etapem jest utworzenie macierzy odległości pomiędzy obiektami. W celu jej sporządzenia skorzystano z odległości euklidesowej (Statsoft, b.d.). Na jej podstawie wybiera się parę obiektów, pomiędzy którymi występuje najmniejsza odległość (Balicki, 2013). To od nich rozpocznie się budowanie dendrogramu, czyli diagramu drzewa (Balicki, 2013). Finalnie otrzymuje się jedno skupienie. W celu policzenia odległości między skupieniami, które w kolejnych krokach należy połączyć, można skorzystać z następujących metod: najbliższego sąsiada, najdalszego sąsiada, mediany, centroidalnej, inaczej środka ciężkości, czy Warda (Balicki, 2013). W momencie, gdy wszystkie obiekty zostaną już połączone, należy odpowiednio podzielić diagram drzewa.

Na końcu przeprowadzono porządkowanie liniowe za pomocą ważonej metody sum standaryzowanych, która pozwala uszeregować obiekty. W wyniku tego można określić, w których z nich badane zjawisko jest lepiej oceniane, a w których gorzej. Istotną rolę odgrywa tutaj normalizacja zmiennych poprzez standaryzację oraz ujednolicenie ich charakteru. Muszą być one stymulantami. Ponadto należy pamiętać, iż wagi muszą przyjmować wartości większe od zera i sumować się do jedynki. Wyznaczone miary rozwoju należy posortować od największej do najmniejszej, ponieważ wyższe wartości wskazują na to, że zjawisko, które jest poddawane analizom, odznacza się osiągnięciem wyższego poziomu (Stanimir, 2006).

4. Wyniki badań

4.1. Statystyki opisowe

Początkowo zbadano statystyki opisowe dla pierwszego zbioru zmiennych od X_1 do X_6 , które znajdują się w tab. 1.

Tabela 1. Statystyki opisowe dla pierwszej grupy danych, gdzie zmienne są destymulantami

Zmienna	Liczba ważnych obserwacji	Średnia	Minimum	Maksimum	Odchylenie standardowe	Współczynnik zmienności
X_1	27	8,074	2,000	20,000	4,057	50,242
X_2	27	24,889	9,000	63,000	11,908	47,844
X_3	27	24,444	7,000	71,000	13,414	54,877
X_4	27	18,148	7,000	62,000	10,683	58,867
X_5	27	77,000	58,000	90,000	8,005	10,396
X_6	27	26,778	7,000	72,000	14,996	56,001

Źródło: opracowanie własne z wykorzystaniem programów Statistica 13 i Excel.

Na podstawie sporządzonej tab. 1 należy stwierdzić, iż nie występują braki danych oraz współczynnik zmienności dla każdej ze zmiennych jest istotny. Maksymalne wartości dla zmiennych X_2 , X_3 , X_4 , oraz X_6 odnotowano dla Hiszpanii. W przypadku cechy X_1 maksimum zaobserwowano dla Francji, a minimum dla Estonii, w której najniższe wartości dostrzeżono także dla zmiennych: X_3 oraz X_6 . Warto również wspomnieć o sytuacji w Danii, gdzie odnotowano najniższe wartości dla cech: X_2 , X_3 i X_4 . Zważywszy na to, że przeprowadzanie szkoleń w firmach jest istotne, warto przyrzeć się średniej arytmetycznej dla zmiennej X_5 . Jej wartość to 77%. Jest to nadal dość duży procent przedsiębiorstw. Jeśli chodzi o obawy związane z poszczególnymi cyberprzestępstwami, to średnie dla zmiennych X_2 , X_3 , X_4 , X_6 oscylują między 18% a 26% (tab. 1). Ponadto za pomocą wykresów skrzynkowych zbadano występowanie obserwacji odstających. Dla zmiennej X_1 występuje tylko jedna obserwacja odstająca, którą jest Francja. Dla zmiennych X_2 , X_3 oraz X_6 to Hiszpania jako jedyna jest obserwacją odstającą. W przypadku zmiennej X_4 występuje jedna obserwacja ekstremalna, którą jest Hiszpania. W przypadku zmiennej X_5 nie odnotowano obserwacji odstających.

Następnie wyznaczono statystyki opisowe dla drugiego zbioru zmiennych. Zostały one przedstawione w tab. 2. Na jej podstawie należy stwierdzić, iż nie występują braki danych oraz współczynnik zmienności dla każdej ze zmiennych jest istotny. Minimalne wartości dla zmiennych: X_8 , X_9 , X_{10} , X_{12} odnotowano dla Hiszpanii. Dla Portugalii najniższe wyniki w porównaniu z pozostałymi obiektami zaobserwowano dla zmiennych: X_7 oraz X_8 . Dla cech X_8 , X_9 i X_{12} maksymalne wartości, a zarazem lepsze od pozostałych analizowanych krajów stwierdzono dla Danii. Dla zmiennej X_{11} najwyższa

Tabela 2. Statystyki opisowe dla drugiej grupy danych, gdzie zmienne są stymulantami

Zmienna	Liczba ważnych obserwacji	Średnia	Minimum	Maksimum	Odchylenie standardowe	Współczynnik zmienności
X_7	27	16,111	6,000	38,000	7,165	44,471
X_8	27	29,222	9,000	45,000	8,911	30,495
X_9	27	32,556	7,000	52,000	11,026	33,867
X_{10}	27	40,000	7,000	59,000	13,229	33,072
X_{11}	27	20,556	8,000	40,000	7,366	35,834
X_{12}	27	34,407	6,000	57,000	12,401	36,042

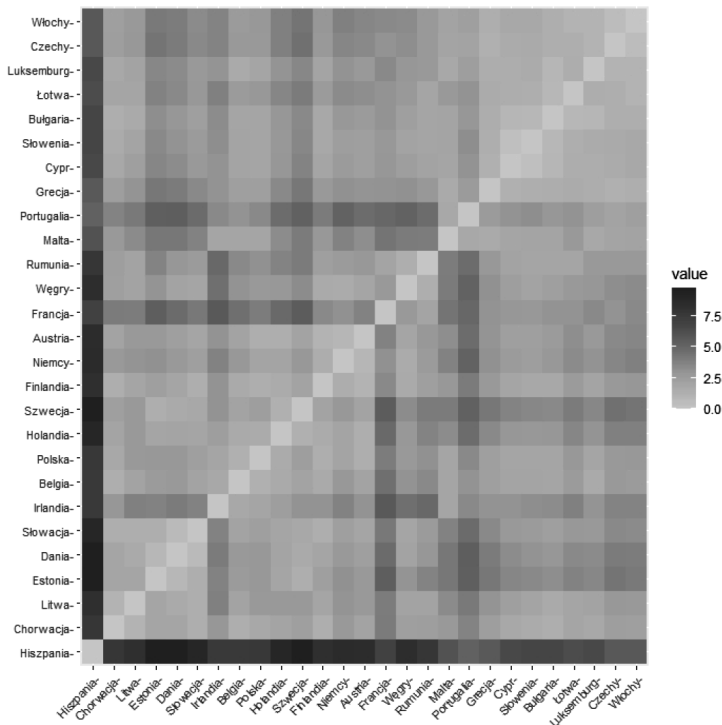
Źródło: opracowanie własne z wykorzystaniem programów Statistica 13 i Excela.

wartość dotyczyła Irlandii. Warto zwrócić uwagę na cechę X_7 , która dotyczy procentu firm, w których pracownicy są bardzo dobrze poinformowani o zagrożeniach związanych z cyberprzestępczością, gdyż wartość średniej arytmetycznej jest dosyć niska i ukształtowała się na poziomie około 16%. Podobna sytuacja dotyczy przeprowadzania szkoleń, czyli zmiennej X_{11} , dla której średnia wyniosła około 21% (tab. 2). Na podstawie sporządzonego wykresu skrzynkowego zauważono, iż dla zmiennej X_7 występuje jedna obserwacja odstająca i jest nią Irlandia. W przypadku pozostałych zmiennych nie odnotowano obserwacji odstających.

4.2. Analiza skupień – klasyfikacja hierarchiczna

Początkowo analizę przeprowadzono dla pierwszego zbioru zmiennych od X_1 do X_6 . Celem przeprowadzenia klasyfikacji hierarchicznej najpierw znormalizowano zmienne za pomocą standaryzacji. Następnie utworzono macierz odległości euklidesowych pomiędzy państwami Unii Europejskiej, która stanowi bazę wykorzystywanej metody analizy (Kassambara i Mundt, 2020) (rys. 1).

Na rysunku 1 ciemniejszy kolor świadczy o tym, że odległości między krajami są większe. Im jaśniejszy kolor, tym mniejsza odległość między obiektami, czyli istnieje między nimi większe podobieństwo. Można zauważyć, iż w zasadzie największe odległości występują pomiędzy poszczególnymi krajami a Hiszpanią. Zatem już na samym początku można wyciągnąć wniosek, iż obserwacje badanych zmiennych w Hiszpanii różnią się od pozostałych państw. Kolejnym etapem było utworzenie wykresu przebiegu aglomeracji. Umożliwia on podjęcie decyzji w kwestii wyboru odpowiedniej liczby klas, na które należałoby podzielić badane obiekty, czyli państwa. Na podstawie tego wykresu stwierdzono, iż należałoby przeciąć 23. wiązanie. Jednakże w ramach dokładniejszej analizy posłużono się także indeksem Grabińskiego i wyznaczono maksimum lokalne wskaźnika q_i , które wskazało miejsce podziału diagramu drzewa (Grabiński, 1992). Pierwsze maksimum lokalne wskaźnika q_i równe 2,019 wskazuje, iż należałoby uciąć drugie wiązanie. W wyniku tego wyłoniłoby się

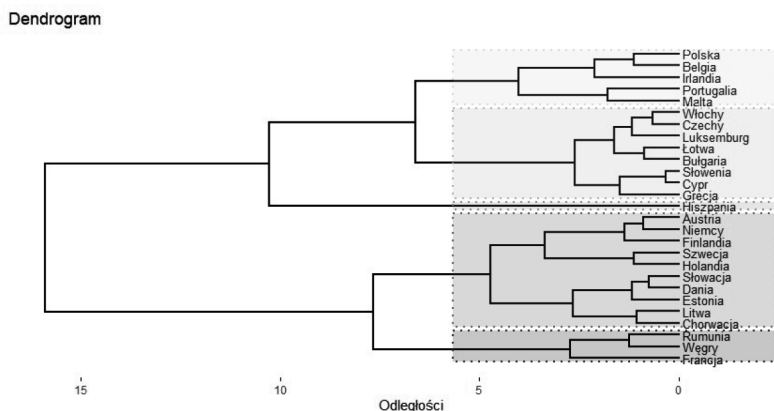


Rys. 1. Macierz odległości dla pierwszej grupy danych, gdzie zmienne są destymulantami

Źródło: opracowanie własne z wykorzystaniem programu RStudio

aż 26 grup państw. Na podstawie drugiego maksimum, które wynosi 1,545, ucięto by wiązanie 26., czyli ostatnie. Skutkowałoby to powstaniem tylko dwóch klas. Z racji tego, że dwa pierwsze maksima lokalne są skrajnymi wartościami, czyli występują dla drugiego i ostatniego wiązania, postanowiono wyznaczyć trzecie, którego wartość wynosi 1,398. Zgodnie z tym potwierdził się wniosek wyciągnięty na podstawie analizy wykresu przebiegu aglomeracji. Wobec powyższego w ostateczności dokonano podziału obiektów na 5 klas, co wiązało się z ucięciem wiązania 23. W następnej kolejności zgodnie z wyciągniętymi wcześniej wnioskami skonstruowano dendrogram, czyli diagram drzewa (Kassambara i Mundt, 2020) (rys. 2).

Zgodnie z rysunkiem 2 pierwsze skupienie to klasa jednoelementowa tworzona przez Hiszpanię. Do drugiego skupienia trafiły: Włochy, Czechy, Luksemburg, Łotwa, Bułgaria, Słowenia, Cypr oraz Grecja. W trzecim skupieniu znalazły się: Rumunia, Węgry oraz Francja. W czwartej grupie można wyróżnić następujące kraje: Polskę, Belgię, Irlandię, Portugalię oraz Maltę. Natomiast do ostatniej klasy zaliczono: Austrię, Niemcy, Finlandię, Szwecję, Holandię, Słowację, Danię, Estonię, Litwę i Chorwację (rys. 2).



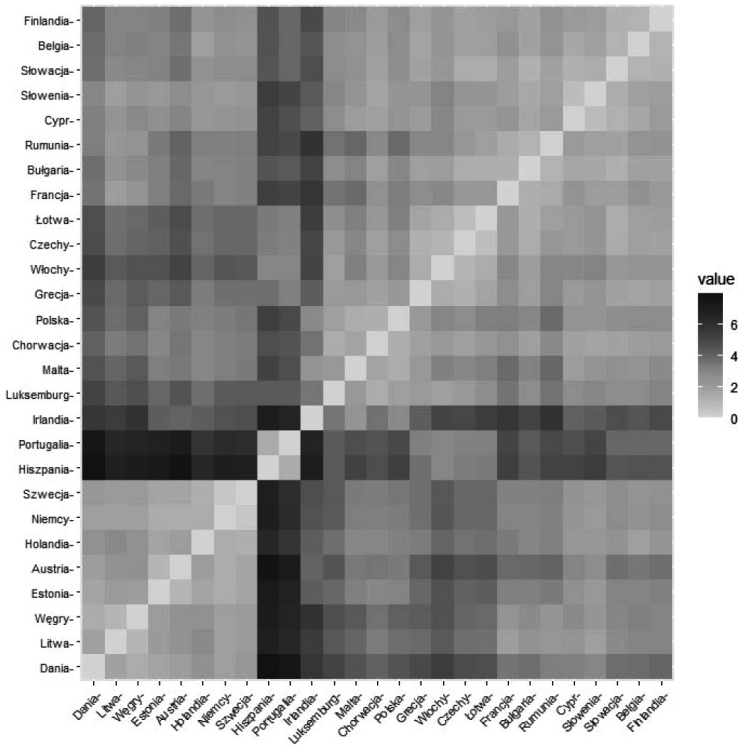
Rys. 2. Dendrogram dla pierwszej grupy danych, gdzie zmienne są destymulantami

Źródło: opracowanie własne z wykorzystaniem programu RStudio.

W celu przeprowadzenia drugiej klasyfikacji hierarchicznej opierającej się na drugim zbiorze danych dla zmiennych od X_7 do X_{12} również na samym początku znormalizowano zmienne poprzez ich standaryzację. Następnie sporządzono macierz odległości euklidesowych pomiędzy badanymi obiektami (rys. 3).

Na podstawie rysunku 3 zauważono, iż zarówno Hiszpania, jak i Portugalia są do siebie bardzo podobne, ale najmniej podobne do innych państw ze względu na badane zmienne, gdyż odległości pomiędzy nimi a pozostałymi krajami są znaczące. Wyznaczona została również najmniejsza odległość, która dotyczy Szwecji i Niemiec. Właśnie te dwa kraje zostaną jako pierwsze połączone i od nich rozpocznie się konstrukcja diagramu drzewa, czyli dendrogramu. Następną bardzo ważną kwestią było ustalenie liczby klas, na które należy podzielić państwa Unii Europejskiej. Wobec tego posłużono się wykresem przebiegu aglomeracji. W kroku 24. zauważalny był dość duży skok, jeśli chodzi o długość wiązania. Zatem zgodnie z tym wykresem można było podejrzewać, iż to w tym miejscu należałoby zakończyć łączenie diagramu drzewa. Zanim podjęto ostateczną decyzję, zbadano także indeks Grabińskiego. Pierwsze maksimum lokalne wskaźnika q_i równe 1,909 wskazuje, iż należałoby uciąć 24. wiązanie. W wyniku tego powstaną 4 klasy. Zatem zgadza się to z wnioskiem wyciągniętym na podstawie wykresu przebiegu aglomeracji. W następnej kolejności zgodnie z wyciągniętymi wcześniej wnioskami skonstruowano dendrogram, czyli diagram drzewa (Kassambara i Mundt, 2020) (rys 4).

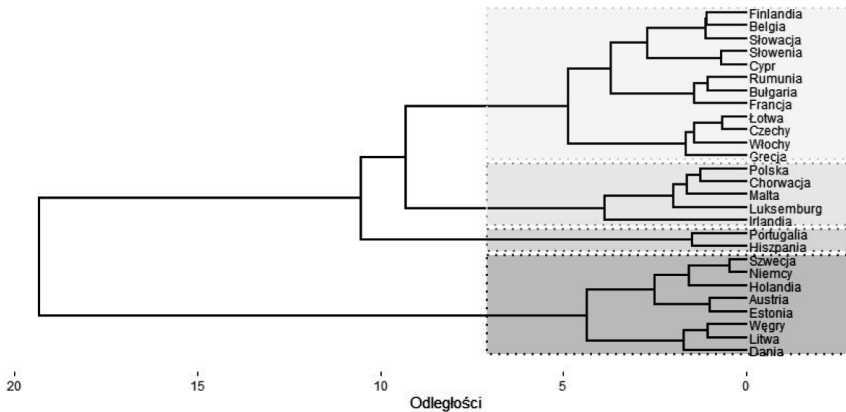
Na rysunku 4 widać, że pierwsze skupienie tworzą Portugalia oraz Hiszpania. Do drugiej klasy należą: Polska, Chorwacja, Malta, Luksemburg i Irlandia. W trzecim skupieniu znalazły się: Szwecja, Niemcy, Holandia, Austria, Estonia, Węgry, Litwa oraz Dania. Natomiast Finlandia, Belgia, Słowacja, Słowenia, Cypr, Rumunia, Bułgaria, Francja, Łotwa, Czechy, Włochy oraz Grecja pogrupowały się razem, tworząc ostatecznie ze skupień (rys. 4).



Rys. 3. Macierz odległości dla drugiej grupy danych, gdzie zmienne są stymulantami

Źródło: opracowanie własne z wykorzystaniem programu RStudio.

Dendrogram



Rys. 4. Dendrogram dla drugiej grupy danych, gdzie zmienne są stymulantami

Źródło: opracowanie własne z wykorzystaniem programu RStudio.

W celu porównania obu przeprowadzonych klasyfikacji hierarchicznych wyznaczono współczynnik Randa. Jego wartość jest równa 0,724. Oznacza to, iż zgodność utworzonych podziałów analizowanych obiektów jest dość wysoka. Ponadto powstałe skupienia są relatywnie zbliżone do siebie. W związku z tym państwa Unii Europejskiej pogrupowały się podobnie.

4.3. Porządkowanie liniowe – ważona metoda sum standaryzowanych

Następnym krokiem niniejszego badania było skonstruowanie rankingu dla obu grup danych za pomocą metody porządkowania liniowego, jaką jest ważona metoda sum standaryzowanych. Na początku zmiennym zostały przypisane odpowiednio wagi. Zmienną, która dotyczy przeprowadzenia szkoleń i podnoszenia świadomości na temat zagrożeń związanych z cyberprzestępczością, uznano za najważniejszą w porównaniu z pozostałymi cechami. Wynika to z faktu, iż organizowanie kursów przyczynia się do tego, że pracownicy i dyrektorzy zdobywają kompetencje i poszerzają wiedzę w określonym obszarze. Zatem zmiennej tej nadano wagę równą 0,25. Cechę dotyczącą stopnia poinformowania pracowników o cyberzagrożeniach uznano za drugą co do istotności i przyznano jej wagę 0,19. Uznano, iż ma ona większe znaczenie niż pozostałe, ponieważ poinformowani pracownicy mają wiedzę na temat tego, jakie występują cyberprzestępstwa i jak w razie ich wystąpienia prawidłowo postępować. Następne cztery zmienne zostały uporządkowane na podstawie tego, ile procent małych i średnich przedsiębiorstw w Unii Europejskiej w ciągu 2021 roku doświadczyło poszczególnych cyberprzestępstw. W przypadku wirusów, *spyware* oraz *malware* odsetek ten był największy, gdyż wyniósł 14%. Natomiast na drugim miejscu znalazł się cyberatak typu phishing – z wynikiem 11%. Jeśli chodzi o *ransomware*, zjawisko to dotyczyło około 4% firm w Unii Europejskiej. Wobec powyższego zmienne dotyczące obaw związanych z wirusami, *spyware*, *malware* oraz z phishingiem otrzymały takie same wagi równe 0,17. Zmiennej, która opisuje obawy dotyczące *ransomware*, przypisano wagę 0,12. Tym samym ostatnia ze zmiennych otrzymała wagę 0,10.

Najpierw utworzono ranking na podstawie pierwszej grupy danych. Z racji tego, iż wszystkie zmienne mają charakter destymulant, przed wykonaniem porządkowania przekształcono je na stymulanty. Następnie znormalizowano zmienne za pomocą standaryzacji. W ten sposób po odpowiednich obliczeniach i uwzględnieniu wag wyznaczono miary rozwoju i uporządkowano je od największej do najmniejszej. Zgodnie z powstałym rankingiem najgorzej wypada Hiszpania. Natomiast na pierwszym miejscu plasuje się Estonia. Polska zajmuje jedenastą pozycję. Pierwsze cztery miejsca zajmują następujące państwa: Estonia, Szwecja, Dania oraz Holandia. Ponadto pogrupowały się one również razem w obu przeprowadzonych wcześniej analizach skupień.

Drugie porządkowanie dotyczyło drugiej grupy danych. W tym przypadku zmienne były stymulantami, więc zostały one najpierw zestandaryzowane. Następnie wy-

znaczono miary rozwoju. Na podstawie rankingu zauważono, iż powtórnie ostatnie miejsce zajmuje Hiszpania. Natomiast najlepiej zjawisko badane oceniono w Irlandii. Miejsca od drugiego do siódmego zajmują ponownie kraje, które znajdowały się w jednym skupieniu dla obu analiz skupień. Są to kolejno: Austria, Estonia, Dania, Niemcy, Szwecja oraz Holandia. Polska uplasowała się na dziewiątej pozycji. W porównaniu do pierwszego utworzonego rankingu znalazła się dwa miejsca wyżej.

Za pomocą współczynnika korelacji rang Spearmana zanalizowano także zależność między dwoma powstałymi rankingami. Współczynnik zawiera się w przedziale od -1 do 1 (Stanimir, 2006). Wartość tego współczynnika jest równa $0,82$. Świadczy to o tym, że występuje silna korelacja i oba utworzone uporządkowania można uznać za stosunkowo zgodne.

5. Podsumowanie

Przeprowadzona analiza pozwoliła wyciągnąć wiele interesujących wniosków. Po przeprowadzeniu klasyfikacji hierarchicznych zarówno dla pierwszego, jak i drugiego zbioru danych do jednej grupy trafiły między innymi następujące państwa: Niemcy, Estonia, Szwecja, Holandia, Austria oraz Dania. Zatem potwierdziła się hipoteza, iż w jednej klasie znajdują się kraje, które zaliczane są do bogatych i bardziej rozwiniętych. Ciekawe jest to, że Estonia nie znalazła się w skupieniu z krajami Europy Środkowo-Wschodniej. W dodatku uplasowała się na pierwszym oraz trzecim miejscu w sporządzonych rankingach. Interesujący jest również fakt, iż od samego początku przeprowadzanej analizy Hiszpanię uznano za obserwację odstającą. W pierwszej klasyfikacji hierarchicznej utworzyła grupę jednoelementową, natomiast w drugiej klasyfikacji połączyła się w jedno skupienie z Portugalią. W przypadku rankingów utworzonych za pomocą ważonej metody sum standaryzowanych zauważono, iż pierwsze miejsca zajmują kraje Europy Zachodniej oraz Północnej. W związku z tym druga z hipotez również się potwierdziła i kraje te plasują się wysoko w rankingach ze względu na analizowane zmienne.

W niniejszym artykule poruszono zaledwie kilka kwestii związanych z cyberbezpieczeństwem. W związku z tym sugeruje się prowadzenie dalszych badań w celu poszerzenia wiedzy. Ważne jest też zgłębienie tematu cyberwojny i jej wpływu na państwo. Jako przykład można podać cyberataki na Ukrainę w 2017 roku oraz na Estonię w 2007 roku. Zaletą jednak jest to, że udostępniane jest coraz więcej raportów z wynikami badań na temat zarejestrowanych incydentów dotyczących cyberprzestępczości, świadomości społeczeństw w zakresie cyberbezpieczeństwa czy wydatków z nim związanych.

Literatura

- Antczak, J. (2020). Costs of Cyber – Security in a Business Entity. *Education of Economists and Managers*, 55(1), 81-93. <https://doi.org/10.33119/EEiM.2020.55.6>
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Bambara, J. J. i Allen, P. R. (2018). *Blockchain a Practical Guide to Developing Business, Law, and Technology Solutions*. Mc Graw Hill Education.
- Bay, M. (2016). What Is Cybersecurity? In Search of an Encompassing Definition for the Post – Snowden Era. *French Journal for Media Research*, (6).
- Chakraborty, N. (2013). Intrusion Detection System and Intrusion Prevention System: A Comparative Study. *International Journal of Computing and Business Research*, 4(2).
- Choi, Y.B. i Allison, G. D. (2017). Intrusion Prevention and Detection in Small to Medium – Sized Enterprises. *SAIS 2017 Proceedings*, (11).
- Cunningham, Ch. (2021). *Wojny w cyberprzestrzeni: koncepcje, strategię i taktyki, dzięki którym przetrwasz i ocalisz swoją organizację*. Helion.
- European Commission. (2022). *Flash Eurobarometer 496 SMEs and Cybercrime – November-December 2021. Report*. (Fieldwork 26/11-17/12/2021). Pobrano 14 marca 2024 z <https://europa.eu/eurobarometer/surveys/detail/2280>
- European Union Agency for Cybersecurity. (2021). *Cybersecurity for SMES – Challenges and Recommendations*. Pobrano 14 marca 2024 z <https://www.enisa.europa.eu/publications/enisa-report-cybersecurity-for-smes>
- Ghosh, A. (2012, 3 maja). *Performance of Computer: What it Actually Means*. The Customize Windows. Pobrano 5 marca 2024 z <https://thecustomizewindows.com/2012/05/performance-of-computer-what-it-actually-means/>
- Główny Urząd Statystyczny [GUS]. (2016). *Jakość życia w Polsce*. Główny Urząd Statystyczny.
- Grabiński, T. (1992). *Metody taksonometrii*. Wydawnictwo Akademii Ekonomicznej w Krakowie.
- Intel. (b.d.). *Ponad 50 lat prawa Moore’a*. Pobrano 24 października 2022 z <https://www.intel.pl/content/www/pl/pl/innovation/leadership/overview.html>
- JEL Classification System. (2023.) Pobrano 10 maja 2024 [https://www.aeaweb.org/econlit/jelCodes.php#:~:text=EconLit%](https://www.aeaweb.org/econlit/jelCodes.php#:~:text=EconLit%20)
- Kassambara, A. i Mundt, F. (2020). *Extract and Visualize the Results of Multivariate Data Analyses – Package Factoextra*. Pobrano z <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf>
- Mathew, A. R. (2019). Cyber Security through Blockchain Technology. *International Journal of Engineering and Advanced Technology*, 9(1), 3821-3824, <https://doi.org/10.35940/ijeat.A9836.109119>
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Prus, B. i Król, K. (2017). Ocena zastosowania wybranych metod taksonomicznych do klasyfikacji zjawisk społeczno-gospodarczych. *Acta Scientiarum Polonorum Formatio Circumiectus*, 16(2), 179-197. Pobrano z <https://acta.urk.edu.pl/pdf-102528-34000?filename=OCENA%20ZASTOSOWANIA.pdf>
- Stanimir, A. (red.). (2006). *Analiza danych marketingowych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- StatSoft. (b.d.). *Analiza skupień*. Pobrano 2 stycznia 2024 z https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstcluan.html
- Szafranek, D. (2021). Wpływ rozwoju cyberprzestępczości na funkcjonowanie współczesnych organizacji. *Nowoczesne Systemy Zarządzania*, 16(4), 43-54, DOI: <http://dx.doi.org/10.37055/nsz/147080>.

Research on the Level of Implementation of Cybersecurity Objectives of European Union Companies and Institutions

Abstract: Cybersecurity is a very important topic that is being talked about and heard more and more often. It has gained importance with technological development. The aim of the article is to examine the situation in the area of cybersecurity in small- and medium-sized enterprises in the European Union and to discuss the forms of protecting companies against cyber threats. For this purpose, a two-way analysis was carried out using hierarchical classification and linear ordering using the standardised sum method. The data comes from a study conducted by Ipsos European Public Affairs. The analysis and the resulting conclusions confirmed the hypotheses that richer and more developed countries would be grouped together and that Western and Northern European countries would rank high in the rankings.

Keywords: cybersecurity, cybercrime, blockchain, hierarchical classification, linear ordering

Stanisław Marek Halkiewicz

e-mail: 184031@student.ue.wroc.pl

ORCID: 0009-0000-7344-7522

Uniwersytet Ekonomiczny we Wrocławiu

Paweł Rychlicki

e-mail: 184100@student.ue.wroc.pl

ORCID: 0009-0005-6267-6457

Uniwersytet Ekonomiczny we Wrocławiu

Oczyszczanie wody, jej pobór i zużycie na terenach górniczych i pogórniczych: analiza przestrzenna powiatów województwa śląskiego

DOI: 10.15611/2024.76.5.09

JEL: C21, Q24, Q53, Q57, R12

© 2024 Stanisław Marek Halkiewicz, Paweł Rychlicki

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>

Cytuj jako: Halkiewicz, S. M. i Rychlicki, P. (2024). Oczyszczanie wody, jej pobór i zużycie na terenach górniczych i pogórniczych: analiza przestrzenna powiatów województwa śląskiego. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 132-148). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Przemysł wydobywczy w Polsce jest elementem strategicznych zasobów gospodarki, zapewniającym bezpieczeństwo ekonomiczne i energetyczne. Problem stanowić może jednak wodochłonność procesów wydobywczych, szczególnie w XXI wieku, kiedy to coraz więcej mówi się o kryzysie wodnym. W niniejszej pracy poruszono problem szeroko rozumianej gospodarki wodnej w powiatach województwa śląskiego, będącego zagłębiem górnictwa węglowego. Omówiono, jak podchodzi się do tego zagadnienia w literaturze. Następnie wykonano analizę autokorelacji wybranych zmiennych opisujących sposoby oszczędzania wody w powiatach w całej Polsce. Wykazano, że powiaty w województwie śląskim w porównaniu z resztą kraju wyróżniają się pod względem zużycia i poboru wody, ale również wykazują istotną korelację typu Low-Low i High-Low pod względem stopnia oczyszczania wody. Następnie dokonano analizy statystycznej zmiennych w powiatach województwa śląskiego. Wskazała ona, że obszary wydobywcze lub powydobywcze charakteryzują się wyraźnie niższym udziałem ścieków oczyszczonych oraz wyższym zużyciem i poborem wody na potrzeby przemysłowe. Oznacza to gorszą sytuację gospodarki wodnej dla obszarów związanych z górnictwem.

Słowa kluczowe: górnictwo, Śląsk, autokorelacja przestrzenna, gospodarka wodna, ekologia

1. Wstęp

W niniejszej pracy postanowiono podjąć temat stanu środowiska i jego ochrony na terenach objętych wpływami przemysłu wydobywczego lub o charakterze powydobywczym, ze szczególnym uwzględnieniem górnictwa węgla kamiennego. Celem analizy jest sprawdzenie, czy istnieje zależność między obecnością przemysłu wydobywczego – w perspektywie teraźniejszej lub historycznej – a:

- udziałem ścieków oczyszczonych w ściekach wymagających oczyszczenia,
- poborem wody na potrzeby przemysłu na km²,
- zużyciem wody na potrzeby przemysłu na km².

Postawiono następujące pytania badawcze:

- Czy ośrodki przemysłu wydobywczego zużywają istotnie większe ilości wody niż powiaty, w których takie ośrodki nie są istotnym elementem charakterystyki? (T1)
- Czy ośrodki przemysłu wydobywczego wpływają na zwiększenie poboru wody w powiatach ościennych? (T2)
- Czy ośrodki przemysłu wydobywczego dążą do redukcji negatywnego oddziaływania na stan środowiska poprzez oczyszczanie ścieków przemysłowych? (T3)

Jako obszar, na którym przeprowadzona została analiza, wybrano województwo śląskie, a rok badania ustawiono na 2022. Było to motywowane dostępnością danych. Województwo śląskie jest powszechnie znane jako zagłębie przemysłu węglowego, czyli największej gałęzi przemysłu wydobywczego w Polsce.

1.1. Ochrona środowiska na terenach górniczych i pogórnich

Ochrona terenów górniczych w Polsce od dawna jest obiektem zainteresowań ustawodawców (Dekret z dnia 6 maja 1953 r.) i badaczy (Duży i in., 2018; Kowalik i in., 2001; Lipiński, 1979; Popiołek, 2009). W literaturze przedmiotu autorzy skupiają się najczęściej na zagadnieniach szkód górniczych związanych ze zniekształceniem terenu i zagrożeniami sejsmicznymi – zapobieganiu im (Gracia i in., 2023; Jian, 2020; Misa, 2015), ich wpływem na otoczenie (Pała, 2023), unormowaniami prawnymi (Lipiński, 1979; Mika i Kaszowska, 2015), oceną szkód (Deliormanli, 2012) czy metodami modelowania ryzyka (Kwiatkiewicz, 2010; Misa, 2015; Ścigała, 2008). Przemysł wydobywczy jednak stwarza zagrożenie ze względu nie tylko na ryzyko zapaści, ale również na eksploatację zasobów i zanieczyszczanie środowiska naturalnego, co stwarza poważne zagrożenie dla fauny i flory, ale również dla społeczności zamieszkującej tereny, dla której obecność przemysłu wydobywczego jest (lub była) czynnikiem istotnie kształtującym charakterystykę gospodarki środowiskowej. Badacze rozpoznają również ten problem, czego dowodzi już ponad ćwierćwieczny cykl corocznych konferencji poświęconych ochronie środowiska na terenach górniczych i pogórnich (Duży i in., 2018). Wybrane problemy ogólnej ochrony środowiska na terenach górniczych opisywał na przykład Strzałkowski (2007), a analizę dla konkretnego przypadku lokalnego przeprowadzali m.in. Dźwigoł i in. (2004), Nietrzeba-Marcinonis (2010) czy Saini i in. (2016).

2. Metodyka badania

2.1. Dobór zmiennych

Wśród czynników środowiskowych najbardziej narażonych na negatywne skutki eksploatacji górniczej znajdują się ścieki przemysłowe. W procesach górniczych używa się bardzo dużych ilości wody. Jest ona wykorzystywana przy wspomaganiu oczyszczania przekopów, chłodzeniu maszyn wydobywczych czy przy czynnościach sanitarnych (Lévy i in., 2006). Jednym z naturalnych następstw wodochłonności tych procesów jest drenaż naturalnych zasobów wody w okolicy (Adeeyo i in., 2023; Gunson i in., 2012;) ale również zanieczyszczenie, często metalami ciężkimi lub toksycznymi chemikaliami, co może prowadzić do wyjąławiania gleby i zamierania fauny i flory akwenów (Adeeyo i in., 2023). Przykłady negatywnego wpływu wodochłonności przemysłu wydobywczego na zasoby wody na zagłębiu górnośląskim opisują Kidawa i Molenda (2021). W swojej pracy autorzy ci poruszają temat degradacji zalewów na tamtym terenie, będącej bezpośrednim następstwem drenażu wody na potrzeby tamtejszych kopalni.

W drodze ku zielonemu górnictwu podejmowane są – we współpracy przedstawicieli praktyki i nauki – próby minimalizacji zużycia wody w procesach wydobywczych (Gunson i in., 2012; Miller i in., 2022), w tym opracowywania rozwiązań technologicznych (Araya i in., 2021; Firmani, 2024), lub też wprowadzenia cyrkularnej gospodarki wodnej w kopalniach (Miller i in., 2022; Hamraoui i in., 2024) czy też ponownego użycia wody, np. do irygacji pól uprawnych (Miller i in., 2022).

Wyżej wymienione techniki są przedmiotem badań lub eksperymentalnych wdrożeń. Najczęstszym sposobem utylizacji zanieczyszczonych wód pokopalnianych nadal pozostają te konwencjonalne – wypuszczenie do wód gruntowych lub naziemnych. Dlatego też jako metrykę ograniczenia zanieczyszczeń wodnych pochodzenia górniczego wyznaczono udział ścieków komunalnych i oczyszczonych w ściekach wymagających oczyszczenia. Jest to zmienna względna, wyrażona w procentach.

Jako miary wodochłonności przemysłu wydobywczego wybrano pobór i zużycie wody na potrzeby przemysłu na km². Jak wspomniano wyżej, szkodliwość ekologiczna obszarów przemysłu wydobywczego objawia się nie tylko poprzez samo zanieczyszczenie wody, ale też poprzez wysokie jej zużycie, które powoduje drenaż pobliskich zbiorników naturalnych i sztucznych. W związku z tym kopalnie mogą prowadzić działania prośrodowiskowe dwojako. Oprócz oczyszczania zużytej wody, ośrodki te mogą ograniczać swoje zużycie. Pobór natomiast nie jest równoważny zużyciu, ponieważ często pobór odbywa się nie na terenie powiatu, w którym ośrodek przemysłowy się znajduje, a w powiatach sąsiadujących. Stąd można podejrzewać, że sam pobór wody nie będzie konkluzywny w odpowiedzi na pytania badawcze.

2.2. Autokorelacja przestrzenna

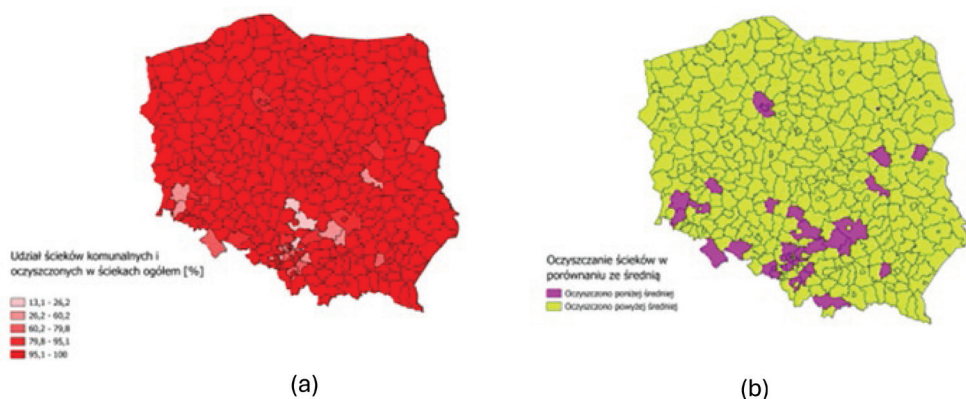
Autokorelacja przestrzenna to zjawisko występujące w danych przestrzennych, które polega na tym, że wartości obserwowane w jednym miejscu są skorelowane z wartościami w miejscach sąsiednich. Innymi słowy, jeśli dane w jednym obszarze mają tendencję do podobieństwa do danych w sąsiednich obszarach, to mówimy o występowaniu autokorelacji przestrzennej. Autokorelację nazywamy dodatnią (pozytywną), gdy w obszarach sąsiadujących wartości są do siebie podobne (wysokie koło wysokich, niskie obok niskich), lub ujemną (negatywną), gdy z obszarem o wysokich wartościach sąsiadują obszary o niskich wartościach (Cliff i Ord, 1981). Aby zbadać autokorelację przestrzenną, stosuje się różne metody, z których jedną z najpopularniejszych jest współczynnik autokorelacji przestrzennej. Jedną z podstawowych, a zarazem najbardziej znanych miar tego współczynnika jest indeks *I* Morana, który wykorzystuje odległości między obserwacjami oraz wartości obserwowane w tych miejscach (Jackson i in., 2010). Wyższe wartości bezwzględne indeksu Morana wskazują na silniejszą autokorelację przestrzenną. W lokalnej perspektywie możemy liczyć lokalne statystyki autokorelacji LISA (z ang. *Local Indicators of Spatial Association*), które pozwalają identyfikować lokalne skupiska podobieństw lub dyspersji w danych przestrzennych. Dzięki badaniu autokorelacji przestrzennej możliwe jest lepsze zrozumienie wzorców występujących w danych przestrzennych. Testowanie istotności statystycznej globalnej statystyki Morana odbywa się najczęściej za pomocą testów iteracyjnych (Xiong i in., 2019).

3. Analiza przestrzenna wybranych zjawisk dla Polski

3.1. Rozłożenie zjawisk w Polsce

Przed przystąpieniem do szczegółowego omówienia badanych zmiennych na terenie województwa śląskiego należy zbadać, jak omawiane zjawiska prezentują się dla całego kraju. Kartogram pokazany na rysunku 1a przedstawia rozkład udziału oczyszczanych ścieków komunalnych i przemysłowych w stosunku do całkowitej ilości ścieków do oczyszczenia [%]. Wyznaczenie klas odbyło się za pomocą metody podziału naturalnego Jenksa, polegającej na minimalizacji wariancji wewnątrz klas i maksymalizacji wariancji między klasami (Jiang, 2011). Wybrano ten sposób klasyfikacji ze względu na charakter zmiennej – bardzo duża część obserwacji ma zbliżone wartości oraz występują naturalne minimum i maksimum wartości (0 to minimum i 100 to maksimum). Użycie tej metody podziału pozwala na uniknięcie powstawania teoretycznych pustych klas dla obiektów o wartościach powyżej lub poniżej możliwych maksimum i minimum. Widać z rozkładu cechy, że zdecydowana większość powiatów w kraju należy do najwyższej klasy, w której między 95 a 100% ścieków jest oczyszczanych. W takim rozkładzie mocno wyróżniają się powiaty o najniższym procencie, znajdujące się w największej liczbie w województwie ślą-

skim. W najgorszej klasie znajdują się powiat częstochowski, będziński i pajęczański z łódzkiego. Jednak w województwie śląskim w porównaniu z całym krajem jest najwięcej powiatów z klas poniżej 80% oczyszczania ścieków. Na rysunku 1b zaprezentowano mapę z zaznaczonymi powiatami, w których oczyszczono mniejszy odsetek ścieków niż średnio w całym kraju. Zdecydowana większość takich powiatów znajduje się w południowej części kraju, z mocną koncentracją ponownie w województwie śląskim. Poniżej średniej wypadają również pojedyncze powiaty z województw dolnośląskiego, łódzkiego, małopolskiego, mazowieckiego i kujawsko-pomorskiego. W większości są to powiaty wokół dużych miast, gęstej aglomeracji lub na terenach, gdzie znajdują się zakłady przemysłowe lub kopalnie.



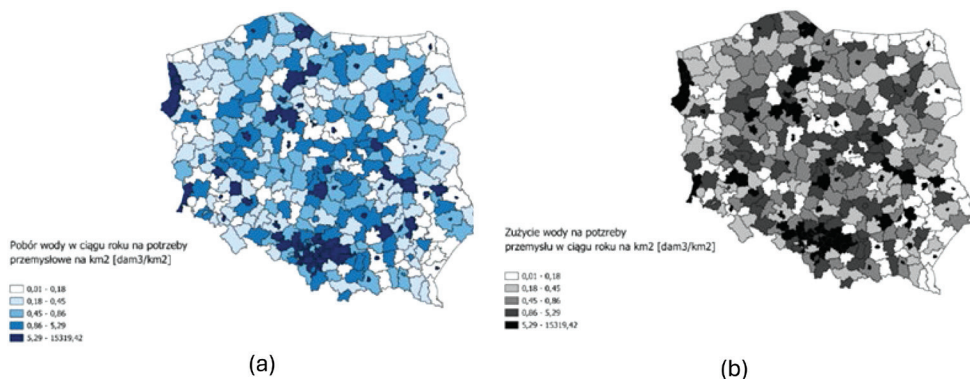
Rys. 1. Udział ścieków komunalnych i oczyszczonych w ściekach ogółem [%]:

(a) jako wartości indywidualne; (b) w porównaniu ze średnią w Polsce

Źródło: Bank Danych Lokalnych (BDL), obliczenia wykonane za pomocą programu QGIS.

Dla dwóch pozostałych zmiennych wyznaczono klasy według kryterium kwantylowego. Przyjęto równe rozpiętości wartości kwantyli. O takim wyborze podziału klas zdecydował rozkład wartości dla obu zmiennych, obie mają szeroki rozstęp wartości przyjmowanych dla obserwacji. Jednocześnie to podejście pozwala uniknąć tworzenia klas pustych pomiędzy klasami z obserwacjami oraz niweluje powstawanie klas jedno- lub kilkuelementowych z obserwacjami o najbardziej skrajnych wartościach. Rysunki 2a i 2b przedstawiają odpowiednio kartogram dla poboru rocznego wody w $\frac{dam^3}{km^2}$ i zużycia rocznego wody wyrażonego również w $\frac{dam^3}{km^2}$.

Porównanie rysunków 2a i 2b wskazuje, że więcej powiatów charakteryzuje się ponadprzeciętnym zużyciem wody, a ponadprzeciętnym poborem. Wynika to zapewne z tego, że występują powiaty o bardzo niskim własnym poborze, które sprowadzają wodę z powiatów sąsiadujących, co potwierdzają powiaty o niskim zużyciu i bardzo wysokim poborze. Przykładem tych pierwszą są powiat lubiński, miasto



Rys. 2. Kartogram z podziałem na powiaty przedstawiający: (a) pobór roczny wody w $\frac{dam^3}{km^2}$; (b) zużycie roczne wody w $\frac{dam^3}{km^2}$

Źródło: BDL, obliczenia wykonane za pomocą programu QGIS.

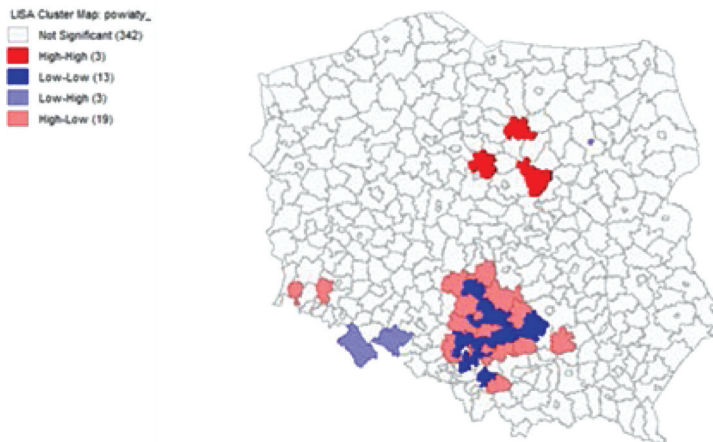
Dąbrowa Górnicza czy Jastrzębie-Zdrój – obszary związane z górnictwem i przemysłem. Natomiast przykładami powiatów z grupy o wysokim poborze i niskim zużyciu są np. Włocławek, powiat łódzki i Tarnów. Obszary o najwyższym poborze wody najęściej koncentrują się w województwach śląskim i małopolskim, mniejsze koncentracje występują na Kujawach, Pomorzu i w Zachodniopomorskiem. Ogólnie obszary wschodnie charakteryzują się mniejszym poborem wody na potrzeby przemysłu. W przypadku zużycia także obserwuje się jeszcze mocniejszą koncentrację powiatów o wysokim zużyciu na Śląsku i w Małopolsce, gdzie występują licznie powiaty z najwyższych klas. W porównaniu do poboru przy zużyciu pojawiają się obszary koncentracji w centrum kraju, w województwach łódzkim i mazowieckim. Intensywniej również zużycie skupia się na Kujawach. Ponownie przede wszystkim wschodnia część kraju należy głównie do klas o najniższych wartościach.

3.2. Autokorelacja od udziału ścieków komunalnych i oczyszczonych w ściekach ogółem (w %)

W celu zbadania statystycznej prawidłowości przestrzennej obliczono współczynniki autokorelacji Morana globalne i lokalne dla każdej ze omawianych zmiennych. W tym celu na początek wyznaczono macierze wag przestrzennych między środkami geograficznymi sąsiadów, których ustalono metodą Queen. Zaczynając od udziału ścieków komunalnych i oczyszczonych w ściekach ogółem [%], otrzymano współczynnik $I = 0,181$, co sugeruje dodatnią autokorelację między sąsiadami – czyli sąsiedzi obiektu o wysokiej (niskiej) wartości zmiennej również przeciętnie mają wysoką (niską) jej wartość. Sprawdzenie istotności odbyło się z pomocą permutacyjnego testu istotności, według którego $pseudo p = 0,0003$. Informuje to o istotności

otrzymanej statystyki Morana, a zatem wśród powiatów występuje statystycznie istotna dodatnia korelacja przestrzenna dodatnia poziomu oczyszczonych ścieków.

Patrząc na statystyki lokalne o istotnym znaczeniu (przy poziomie istotności 0,05) (rys. 3), można stwierdzić, że największa koncentracja powiatów z istotnymi lokalnymi statystykami autokorelacji występuje na Śląsku i na granicy z województwami łódzkim oraz małopolskim.



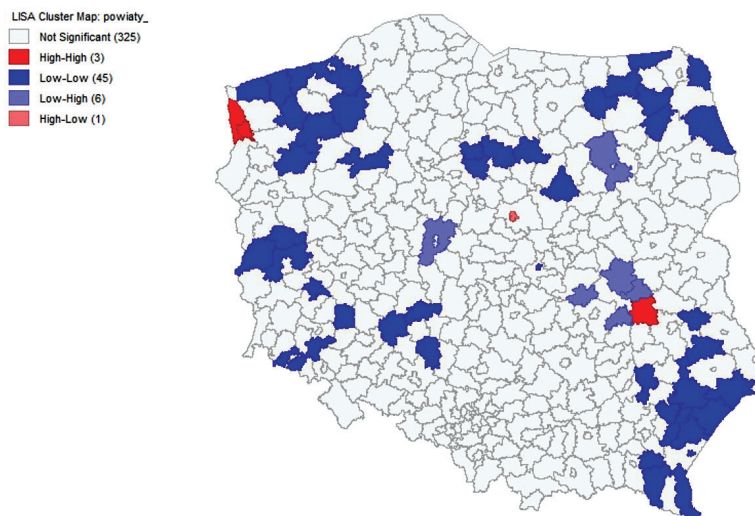
Rys. 3. Kartogram charakterystyki istotnych statystycznie autokorelacji lokalnych udziału ścieków oczyszczonych w ściekach ogólnie

Źródło: BDL, wizualizacje i obliczenia przeprowadzone za pomocą programu QGIS.

Na rysunku 3 można zauważyć, że w głównej mierze są to powiaty w grupie Low-Low, czyli o niskich własnych wartościach cechy i o średnio niskich u sąsiadów. Natomiast równie często na tym obszarze występują powiaty w grupie High-Low, które same mają wysokie wartości cechy, natomiast ich sąsiedzi średnio niskie. Występują pojedyncze powiaty o istotnej statystyce lokalnej w reszcie kraju, ale są to pojedyncze przypadki.

3.3. Autokorelacja poboru wody na potrzeby przemysłowe na km²

Analiza wskazała na bardzo bliską 0 wartość statystyki globalnej Morana, wynoszącą zaledwie $-0,002$. Sugeruje to autokorelację ujemną, jednak sama wielkość podaje w wątpliwość jej statystyczną istotność. Wątpliwości te potwierdza przeprowadzony test permutacyjny na istotność tej statystyki, wskazując *pseudo p* = 0,13, co przy przyjętym poziomie istotności 0,05 oznacza brak statystycznej istotności globalnej statystyki Morana – czyli występuje brak globalnej autokorelacji przestrzennej poboru wody na potrzeby przemysłowe.



Rys. 4. Kartogramy przedstawiające charakterystykę autokorelacji

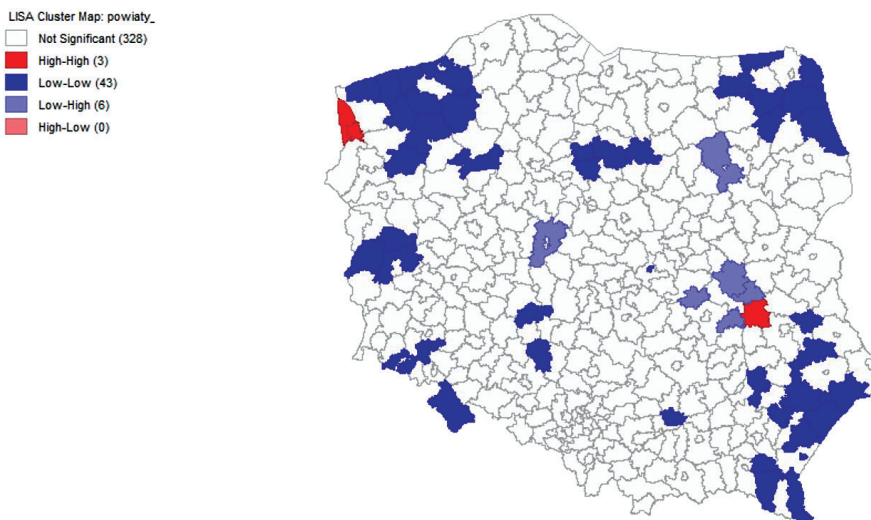
Źródło: BDL, wizualizacje i obliczenia przeprowadzone za pomocą programu QGIS.

Analiza lokalnych statystyk autokorelacji przestrzennej wskazuje, że występują lokalnie istotne autokorelacje (rys. 4). Zdecydowanie najwięcej powiatów o istotnej lokalnej statystyce Morana trafia do grupy Low-Low, czyli o niskich własnych i niskich średnio wartościach cechy dla sąsiadów. Powiaty, w których taka zależność występuje, to głównie obszary na obrzeżach państwa, głównie południowo-wschodnia i północno-zachodnia część. W centralnej części kraju tylko pojedyncze powiaty okazują się mieć istotną statystycznie lokalną autokorelację przestrzenną. Występują pojedyncze powiaty o istotnych statystykach lokalnych, nienależące do klasy Low-Low. W większości należą wówczas do klasy Low-High, jednak nie koncentrują się w żadnej części kraju – wyjątek stanowią okolice powiatu ryckiego, gdzie występują 3 takie powiaty.

3.4. Autokorelacja zużycia wody na potrzeby przemysłowe na km²

Dla zużycia wody na potrzeby przemysłu wyniki analizy autokorelacji okazały się być podobne do wyników dla poboru. Statystyka globalna Morana $I = -0,002$, a odpowiadające *pseudo p* = 0,1297 – co ponownie oznacza brak istotności statystyki dla przyjętego alfa 0,05, zatem nie występuje przestrzenna autokorelacja dla zużycia wody na potrzeby przemysłowe w skali Polski.

Lokalnie odnotowuje się nieco większe zagęszczenie powiatów o istotnej statystyce lokalnej w północno-wschodniej i północno-zachodniej Polsce w porównaniu z rozkładem analogicznej statystyki dla poboru wody (rys. 5). Jednocześnie wiele



Rys. 5. Kartogramy przedstawiające: (a) istotność autokorelacji przestrzennej; (b) charakterystykę autokorelacji

Źródło: BDL, wizualizacje i obliczenia przeprowadzone za pomocą programu QGIS.

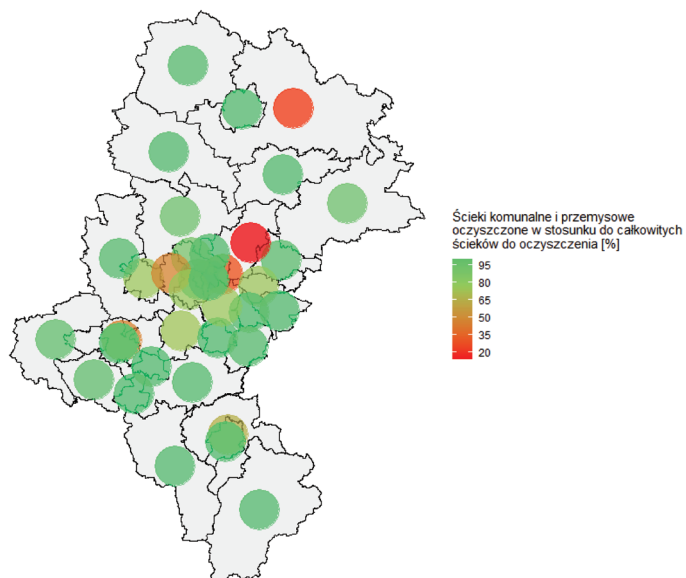
mniej powiatów ma statystycznie istotną autokorelację lokalną w zestawieniu z poborem. Ponownie – zdecydowana większość obiektów o istotnej statystyce trafia do klasy Low-Low; skoncentrowane są one się przy granicach państwa.

4. Szczegółowa analiza eksploracyjna zmiennych w województwie śląskim

4.1. Udział ścieków oczyszczonych w ściekach wymagających oczyszczenia [%]

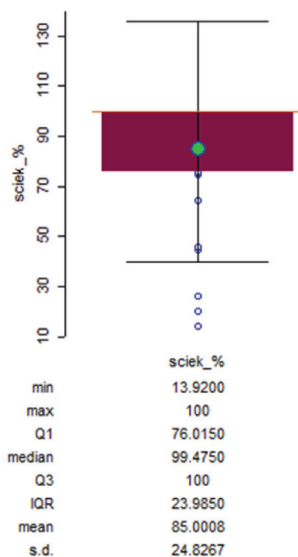
Chcąc zaprezentować rozkład analizowanych cech, posłużono się kartodiagramami. Dla poziomu oczyszczonych ścieków wykorzystano kartodiagram słupkowy (rys. 6), na którym widać, w jakich powiatach oczyszczono najmniejszy procent ścieków. Na rysunku 7 przedstawiono odpowiadający tej zmiennej wykres pudełkowy wraz ze statystykami opisowymi.

Mediana na poziomie 99,475% pozwala wnioskować, że większość z powiatów oczyszcza prawie wszystkie ścieki przemysłowe i komunalne, które tego wymagają. Średnia niższa od mediany wskazuje natomiast na asymetrię prawostronną rozkładu zmiennej. Asymetria ta powodowana jest przez obserwacje odstające. Są to powiaty częstochowski, będziński, rybnicki i miasta Siemianowice Śląskie oraz Zabrze. W tych powiatach oczyszczono poniżej 50% ścieków, a w będzińskim i częstochowskim



Rys. 6. Kartodiagram przedstawiający udział ścieków oczyszczonych w ściekach wymagających oczyszczenia (w %)

Źródło: BDL, obliczenia i wizualizacja wykonane w środowisku R.



Rys. 7. Wykres pudełkowy oraz statystyki opisowe dla zmiennej udziału ścieków oczyszczonych w ściekach wymagających oczyszczenia (w%)

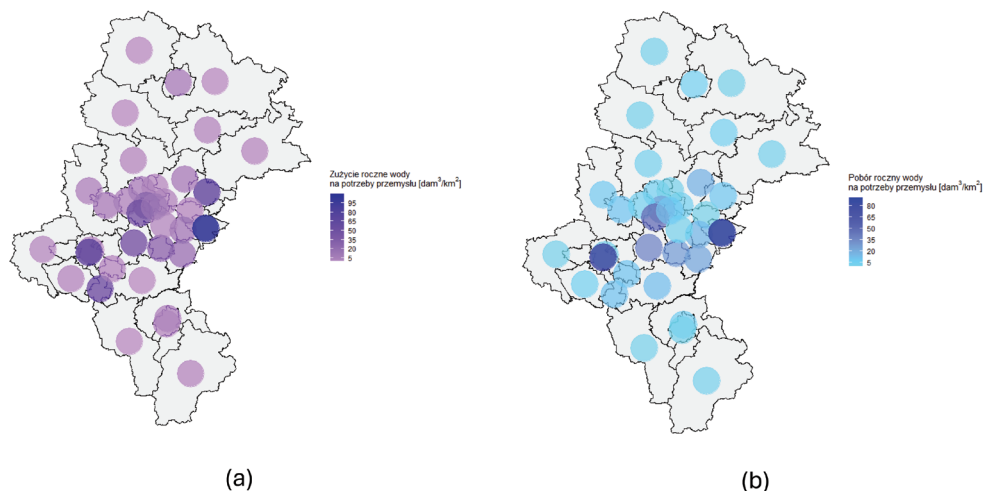
Źródło: BDL, obliczenia i wizualizacja wykonane za pomocą programu Geoda.

nawet poniżej 20%. Wskazuje to na bardzo słabe możliwości lokalnych oczyszczalni lub ścieki będące poza systemem kanalizacyjnym – na przykład pochodzenia przemysłowego, co może powodować wylewanie ścieków do wód lub wymuszać odprowadzanie ich do oczyszczenia w innym powiecie.

Warto zauważyć, że Siemianowice Śląskie (kopalnię zamknięto tam w latach dziewięćdziesiątych) mają na obrzeżach hutę oraz rozwinięty przemysł na potrzeby wojska. W Zabrzej jednak dawniej działało aż 10 kopalni, z czego do dzisiaj wydobycie prowadzi prywatna kopalnia SilTech, która wykorzystuje architekturę i tereny odkrywkowe wcześniejszych zabrzańskich kopalni. W powiecie rybnickim mieści się Rybnicki Okręg Węglowy. Teren Będzina i okolic również historycznie był miejscem ponad 12 zakładów kopalnianych. Dzisiaj na terenie powiatu będzińskiego działają zakłady wydobywcze dolomitu, piasku i ilitu. Również na terenie powiatu częstochowskiego działają zakłady piaskowe, chociaż historycznie i kulturowo nie należy ona do Śląska.

4.2. Pobór i zużycie wody na potrzeby przemysłowe na km^2

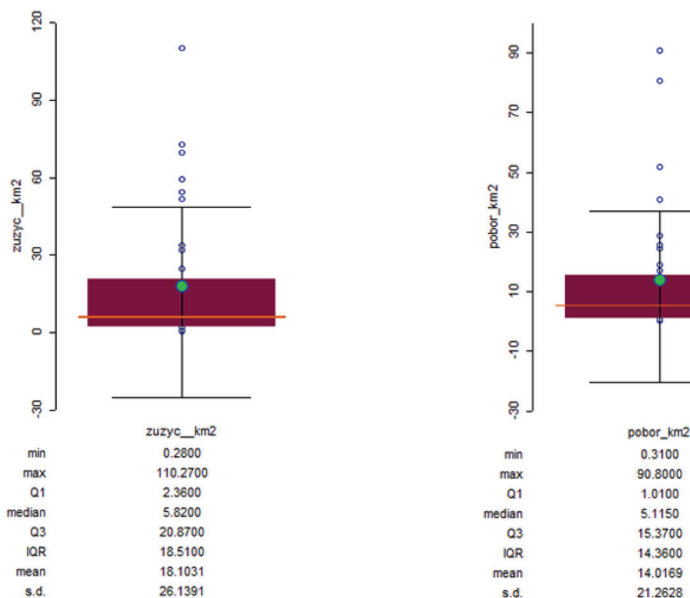
W celu wizualizacji rozkładu zmiennych dotyczących poboru i zużycia wody na potrzeby przemysłowe użyto kartodiagramów tekstowych, na których pokazano, ile wody zużyto (rys. 8a) lub pobrano (rys. 8b) $\frac{\text{dam}^3}{\text{km}^2}$ w 2022 roku na potrzeby przemysłowe.



Rys. 8. Kartodiagram tekstowy przedstawiający: (a) zużycie wody na potrzeby przemysłu w $\frac{\text{dam}^3}{\text{km}^2}$; (b) pobór wody na potrzeby przemysłu w $\frac{\text{dam}^3}{\text{km}^2}$

Źródło: BDL, obliczenia i wizualizacje przeprowadzone w środowisku R.

Na rysunkach 9a i 9b przedstawiono wykresy pudełkowe dla omawianych zmiennych. W województwie śląskim przeważają powiaty o zużyciu od 0 do 16 $\frac{dam^3}{km^2}$ (26 powiatów). Różnice pomiędzy medianami dla obu zmiennych wskazują na skrajnie silną asymetrię lewostronną. Powodowana jest ona m.in. przez obecność wielu obserwacji – nie tylko odstających, ale i ekstremalnych. Jedną z nich stanowi miasto Rybnik, gdzie w 2022 roku zużyto 110,3 $\frac{dam^3}{km^2}$. Rozkład poboru wody wygląda podobnie. Największym poborem, oprócz Rybnika, charakteryzuje się również powiat miasto Jaworzno, które kiedyś było miejscem poboru węgla z dużego kompleksu Kopalni Węgla Kamiennego Jaworzno, a więc można powiedzieć, że powiat miasto Jaworzno jest miastem o charakterze powydobywczym. Jak już wcześniej wspomniano, w Rybniku odbywa się wydobycie węgla kamiennego. Razem z okalającymi go mniejszymi miejscowościami i ośrodkami przemysłu wydobywczego miasto to tworzy Rybnicki Okręg Węglowy. Również Jaworzno jest prężnie działającym terenem górniczo-energetycznym. Działa tam Zakład Górniczo-Energetyczny „Sobieski” Jaworzno III, będący kompleksem pod zarządem spółki TAURON. Należy pamiętać, że obie z tych zmiennych są destymulantami.



(a)

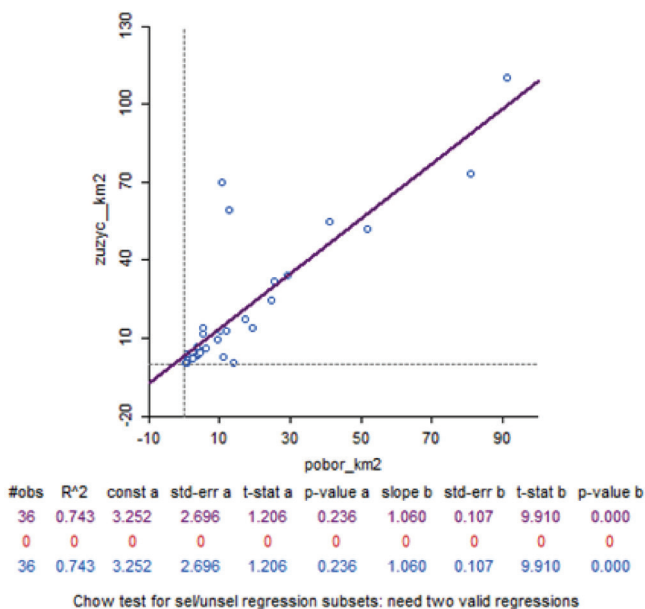
(b)

Rys. 9. Wykres pudełkowy przedstawiający rozkład zmiennej: (a) zużycie wody na potrzeby przemysłu w $\frac{dam^3}{km^2}$; (b) pobór wody na potrzeby przemysłu w $\frac{dam^3}{km^2}$

Źródło: BDL, obliczenia i wizualizacje przeprowadzone w programie Geoda.

4.3. Badanie korelacji poboru i zużycia wody na potrzeby przemysłu

Ocena wizualna pozwala na podejrzenia, że istotnie skorelowane są ze sobą zmienne opisujące pobór i zużycie wody na potrzeby przemysłowe w $\frac{\text{dam}^3}{\text{km}^2}$. Na rysunku 10 przedstawiono wykres rozrzutu dla tych zmiennych.



Rys. 10. Wykres rozrzutu z zaznaczoną dopasowaną prostą regresji

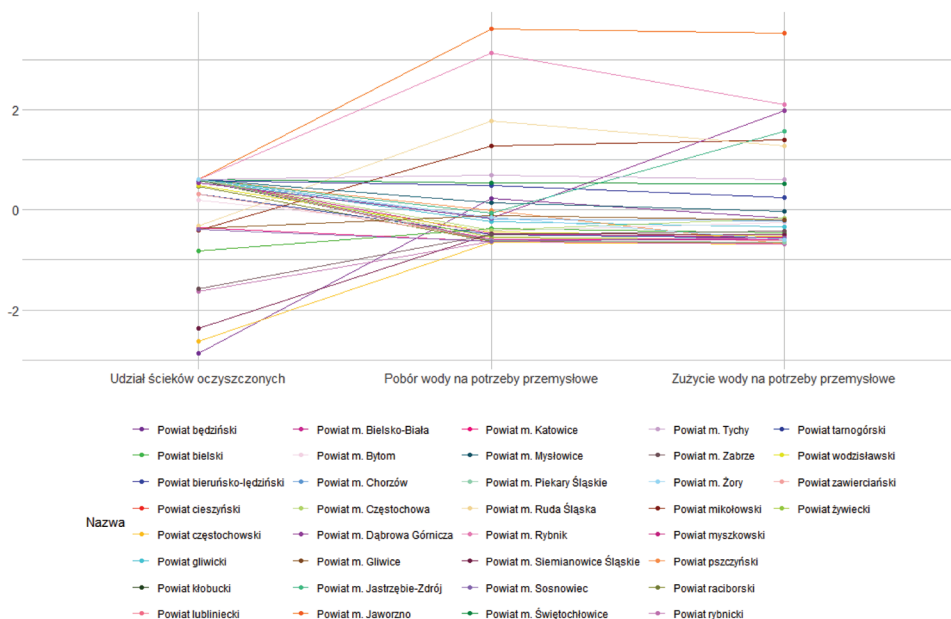
Źródło: BDL, obliczenia i wizualizacje przeprowadzone w programie Geoda.

Prosta regresji jest dobrze dopasowana (rys. 10), co ma silną podstawę merytoryczną, jako że pobór wody i jej zużycie idą w parze. Najbardziej interesującymi wnioskami z wykresu są obserwacje odstające – o stosunkowo wysokim zużyciu wody w stosunku do jej poboru. Są to dwa powiaty miejskie – Dąbrowa Górnicza oraz Jastrzębie-Zdrój. W obu przypadkach są to miasta o charakterze kopalnianym. Wykres ten potwierdza przypuszczenie zaproponowane w rozdziale 3, że kopalnie w tych miastach pobierają wodę z punktów poboru znajdujących się w powiatach ościennych.

4.4. Wykres zmiennych równoległych

Sporządzono wykres zmiennych równoległych w celu bliższego zbadania, czy powiaty o charakterze górniczym mają inną charakterystykę od pozostałych powiatów (rys. 11). Można wizualnie ocenić, że istnieją obserwacje, które nie podążają takim

samym schematem jak pozostałe powiaty. W większości obserwacji zmienna opisująca pobór wody (destymulanta) przyjmuje wartości stosunkowo niskie, dotycząca oczyszczania ścieków (stymulanta) stosunkowo wysokie, a opisująca zużycie wody (destymulanta) znowu stosunkowo niskie. Taką sytuację możemy opisać jako pożądaną, patrząc na charakterystykę zmiennych.



Rys. 11. Wykres zmiennych równoległych (dane standaryzowane)

Źródło: BDL, obliczenia i wizualizacje przeprowadzone w środowisku R.

Wśród anomalii najwyraźniej zarysowuje się powiat będziński, w którym pobór i zużycie są stosunkowo wysokie, a procent oczyszczonych ścieków stosunkowo niski. Jak wspomniano wcześniej, powiat będziński jest obszarem przemysłu wydobywczego. Do obserwacji mających wysoki zarówno pobór wody, jak i jego zużycie, ale przy jednoczesnym wysokim procencie oczyszczonych ścieków, można zaliczyć powiat miasto Jaworzno, powiat miasto Ruda Śląska i powiat miasto Rybnik. Ten ostatni przypadek jest szczególnie interesujący, gdyż jest tam wyraźnie niższe zużycie niż pobór wody. Może być to spowodowane tym, że kopalnie znajdujące się w różnych powiatach Rybnickiego Okręgu Węglowego pobierają wodę właśnie z sieci wodnej miasta Rybnik. Pozostałe z anomalii mają podobny pobór wody, jak również jej zużycie, co można wyjaśnić ich przemysłowo-miejskim charakterem.

5. Podsumowanie i wnioski końcowe

Przeprowadzone badanie udowadnia, że istnieje zależność między obecnością przemysłu wydobywczego a udziałami ścieków oczyszczonych w ściekach wymagających oczyszczenia. Wielokrotnie cecha ta okazywała się wyraźnie gorsza (niższa) dla obszarów o charakterystyce wydobywczej lub powydobywczej w porównaniu z powiatami niezwiązanymi z górnictwem. Przykładami takich obszarów są między innymi powiat miasto Jastrzębie-Zdrój, powiat będziński czy powiat rybnicki. Oznacza to negatywną odpowiedź na ostatnie z postawionych pytań badawczych (T3), mówiącej o dążeniu powiatów charakteryzujących się obecnością przemysłu górniczego do mitygacji swojego negatywnego wpływu na czystość wód. Relacja między obszarami wydobywczymi a poborem i zużyciem wody nie jest jednoznacznie widoczna dla wszystkich powiatów o charakterze górniczym, jednak można zauważyć, że jedne z najwyższych wartości tych zmiennych występują w powydobywczym Jaworznie oraz w Rybniku, będącym centrum Rybnickiego Okręgu Węglowego. Lokalne statystyki autokorelacji spektralnej sugerują brak takiej relacji. Z drugiej strony, patrząc tylko na wartości poboru i zużycia wody, można stwierdzić, że tereny wydobywcze znalazły się w zdecydowanej czołówce wielkości dla obu zmiennych. Ta niejednoznaczność sugeruje potrzebę przeprowadzenia badania inną metodą w celu dokładniejszej weryfikacji tej relacji. Jednocześnie można raczej odrzucić tezę (T2) o istotnym wpływie charakterystyki górniczej powiatów na zwiększenie poboru wody w powiatach ościennych.

Odpowiedziano twierdząco na pytanie badawcze o większym poborze i zużyciu wody na potrzeby przemysłowe w powiatach o wydobywczej lub powydobywczej charakterystyce (T1). Można to zaobserwować zarówno na kartogramach (rys. 2a i 2b), jak i na wykresie zmiennych równoległych na rysunku 11.

Bibliografia

- Adeeyo, A. O., Bello, O. S., Agboola, O. S., Adeeyo, R. O., Oyetade, J. A., Alabi, M. A., Edokpayi, J. N. i Makungo, R. (2023). Recovery of Precious Metals from Processed Wastewater: Conventional Techniques Nexus Advanced and Pragmatic Alternatives. *Journal Of Water Reuse And Desalination*, 13(2), 134-161. <https://doi.org/10.2166/wrd.2023.068>
- Araya, N., Ramírez, Y., Cisternas, L. A. i Kraslawski, A. (2021). Use of Real Options to Enhance Water-energy Nexus in Mine Tailings Management. *Applied Energy*, 303(1), 117626. <https://doi.org/10.1016/j.apenergy.2021.117626>
- Cliff, A.D. i Ord, J. K. (1981). *Spatial Processes: Models & Applications*. Pion.
- Dekret z dnia 6 maja 1953 r. Prawo Górnicze (Dz. U. z 1978 r. Nr 4, poz. 12)
- Deliormanli, A. H. (2012). Assessment of Coal Mining Subsidence Damage in Soma-turkey Coalfield Using FEM. *Archives of Mining Sciences*, 57(1), 179-192. <https://doi.org/10.2478/v10267-012-0013-6>
- Dobrowolski, G., Lipiński, A., Mikosz, R. i Radecki, G. (2018). *Gospodarowanie geologicznymi zasobami środowiska w świetle zasady zrównoważonego rozwoju. Zagadnienia prawne*. Wydawnictwo Uniwersytetu Śląskiego. <https://doi.org/10.31261/MG.2018.0001>

- Duży, S., Pluta, I., Sosna, A., Fuchs, R. i Kopiec, W. (2018). 25 Konferencji poświęconych ochronie środowiska na terenach górniczych i pogórnich w Zachodnim Subregionie Województwa Śląskiego. W: S. Duży (red.), *Ochrona środowiska na terenach górniczych i pogórnich kopalń w Subregionie Zachodnim Województwa Śląskiego* (XXV Konferencja Naukowo-Techniczna, s. 7-19). Stowarzyszenie Inżynierów i Techników Górnictwa. <https://omega.polsl.pl/info/article/PSL059adf829d8c43098ab9e4868fee8d6f/>
- Dźwigoł, H., Chudek, M. i Sapic'kij, K. F. (2004). *Ochrona środowiska w Górnośląskim i Donieckim Zagłębiu Węglowym*. Wydawnictwo Politechniki Śląskiej. https://www.researchgate.net/publication/344262710_Ochrona_srodowiska_w_Gornoslaskim_i_Donieckim_Zaglebiu_Weglowym
- Firmani, G. (2024). Software Development of an Integrated Water Management Optimisation Model: The Roy Hill Mine Case Study Application. *Mine Water and Environment*, 43(1), 41-52. <https://doi.org/10.1007/s10230-023-00967-x>
- Gunson, A. J., Klein, B., Veiga, M. i Dunbar, S. (2012). Reducing Mine Water Requirements. *Journal of Cleaner Production*, 21(1), 71-82. <https://doi.org/10.1016/j.jclepro.2011.08.020>
- Gracia, A., Torrijo F. J., Garzón-Roca, J., Pérez-Picallo, M. i Alonso-Pandavenes, O. (2023). Identification and Mitigation of Subsidence and Collapse Hazards in Karstic Areas: A Case Study in Alcalá de Ebro (Spain). *Applied Sciences*, 13(9), 5687. <https://doi.org/10.3390/app13095687>
- Hamraoui, L., Bergani, A., Ettoumi, M., Aboulaich, A. Taha, Y., Khalil, A. Neculita, C. M. i Benzazoua, M. (2024). Towards a Circular Economy in the Mining Industry: Possible Solutions for Water Recovery through Advanced Mineral Tailings Dewatering. *Minerals*, 14(3), artykuł 319. <https://doi.org/10.3390/min14030319>
- Jackson, M. C., Huang, L., Xie, Q. i Tiwari, R. C. (2010). A Modified Version of Moran's I. *International Journal of Health Geographics*, 9(1), artykuł 33. <https://doi.org/10.1186/1476-072X-9-33>
- Jian, Y. (2020). Solutions to Some Mine Subsidence Research Challenges [rozprawa doktorska, West Virginia University]. <https://researchrepository.wvu.edu/etd/7948>
- Jiang, B. (2011). *A New Classification Scheme for Data with a Heavy-tailed Distribution*. University of Galve.
- Kidawa, J. i Molenda, T. (2021). Degradation of Dam Reservoirs under the Influence of Mining Subsidence in Upper Silesian Coal Basin, South Poland. *Lakes & Reservoirs: Research & Management*, 26, artykuł e12388. <https://doi.org/10.1111/lre.12388>
- Kowalik, S., Lis, M. i Lis, A. (2001). Ochrona oraz rekultywacja terenów górniczych na przykładzie Jastrzębskiej Spółki Węglowej S.A. (materiały pokonferencyjne VI Dni Miernictwa i Ochrony Terenów Górniczych, s. 211-218). Polski Komitet Międzynarodowego Towarzystwa Miernictwa Górniczego. https://www.researchgate.net/publication/306012422_OCHRONA_ORAZ_REKULTYWACJA_TERENOW_GORNICZYCH_NA_PRZYKLADZIE_JASTRZEBSKIEJ_SPOLKI_WEGLOWEJ_SA
- Kwiatek, J. (2010). Szkody górnicze w obiektach budowlanych jako zjawisko losowe. *Bezpieczeństwo Pracy i Ochrona Środowiska w Górnictwie*, (4). https://www.wug.gov.pl/wydawnictwa/miesiecznik_wug/212/1609
- Lévy, V., Fabre, R., Boebel, B. i Hertle, C. (2006). Water Use in the Mining Industry – Threats and Opportunities. W: *Proceedings of Water in Mining 2006*. AIMM.
- Lipiński, A. (1979). Z problematyki prawnej ochrony terenów górniczych. *Prawne Problemy Górnictwa*, 3(1), 22-39. [https://bazhum.muzhp.pl/media/files/Prawne_Problemy_Gornictwa/Prawne_Problemy_Gornictwa-r1979-t3-s22-39/Prawne_Problemy_Gornictwa-r1979-t3-s22-39.pdf](https://bazhum.muzhp.pl/media/files/Prawne_Problemy_Gornictwa/Prawne_Problemy_Gornictwa-r1979-t3/Prawne_Problemy_Gornictwa-r1979-t3-s22-39/Prawne_Problemy_Gornictwa-r1979-t3-s22-39.pdf)
- Mika, W. i Kaszowska, O. (2015). Kryteria dopuszczania eksploatacji górniczej pod terenami zabudowanymi. *Przegląd Górniczy*, (3), 44-49. <https://bibliotekanauki.pl/articles/166576.pdf>
- Miller, K. D., Bentley, M. J., Ryan, J. N., Linden, K. G., Larison, C., Kienzle, B. A., Katz, L. E., Wilson, A. M., Cox, J. T., Kurup, P., van Allsburg, K. M., McCall, J., Macknick, J. E., Talmadge, M. S., Miara, A., Sitterley, K. A., Evans, A., Thirumaran, K., Malhotra, M. i Chellam, S. (2022). Mine Water Use,

- Treatment and Reuse in the United States: A Look at Current Industry Practices and Select Case Studies. *ACS ES&T Engineering*, 2(3), 391-408. <https://doi.org/10.1021/acsesteng.1c00244>
- Misa, R. (2015). *Metody ograniczenia wpływu eksploatacji podziemnej na obiekty budowlane poprzez zastosowanie rozwiązań geotechnicznych* [rozprawa doktorska, Akademia Górniczo-Hutnicza im. Stanisława Staszka w Krakowie]. Biblioteka Główna AGH. <https://winntbg.bg.agh.edu.pl/rozprawy2/11013/full11013.pdf>
- Nietrzeba-Marcinonis, J. (2010). Rekultywacja biologiczna gruntów pogórnich w PGE KWB Turów SA. *Górnictwo i Geoinżynieria*, 34(4), 435-443. https://journals.bg.agh.edu.pl/GORNICtwo/2010-04/GG_2010_4_32.pdf
- Pala, P. (2023). *Wpływ deformacji powierzchni powstałych w wyniku działalności górnictwa węgla kamiennego na wartość gruntów przeznaczonych pod zabudowę na terenie górnośląskiego zagłębia węglowego* [rozprawa doktorska, Uniwersytet Śląski]. <https://fbc.pionier.net.pl/id/oai:www.sbc.org.pl:745456>
- Popiołek, E. (2009). *Ochrona terenów górniczych*. Wydawnictwo AGH.
- Saini, V., Gupta, R. P. i Arora, M. K. (2016). Environmental IMPACT STUDIES in Coalfields in India: A Case Study from Jharia Coal-Field. *Renewable and Sustainable Energy Reviews*, 53(1), 1222-1239. <https://doi.org/10.1016/j.rser.2015.09.072>
- Strzałkowski, P. (2007). *Ochrona środowiska na terenach górniczych: wybrane problemy*. Wydawnictwo Politechniki Śląskiej. <https://delibra.bg.polsl.pl/publication/81291>
- Ścigała, R. (2008). *Komputerowe wspomaganie prognozowania deformacji górotworu i powierzchni wywołanych podziemną eksploatacją górniczą*. Wydawnictwo Politechniki Śląskiej. <https://delibra.bg.polsl.pl/dlibra/doccontent?id=73592>
- Xiong, Y., Bingham, D., Braun, W. J. i Hu, X. J. (2019). Moran's I Statistic-Based Nonparametric Test with Spatio-Temporal Observations. *Journal of Nonparametric Statistics*, 31(1), 244-267. <https://doi.org/10.1080/10485252.2018.1550197>

Water Treatment, Abstraction and Consumption in Mining and Post-mining Areas: Spatial Analysis of Counties in the Silesian Voivodeship

Abstract: The mining industry in Poland is a strategic resource that ensures both economic and energy security. However, the water intensity of mining processes can be a significant issue, particularly when one considers the growing discourse surrounding the impending water crisis that is expected to emerge in the 21st century. This paper will address the challenge of water management in the Silesian Voivodeship, which is a coal mining basin. The prevailing approaches to this challenge will be examined. Subsequently, an autocorrelation analysis of selected variables describing methods for water conservation will be conducted across Polish counties. The results will demonstrate that, in terms of water consumption and abstraction, the Silesian Voivodeship districts stand out among the rest of Poland. Furthermore, the Silesian Voivodeship counties will be observed to exhibit a notable degree of low-low and high-low correlation in terms of the degree of water purification. This will be followed by a statistical analysis of the variables in the districts in the Silesian Voivodeship. The analysis carried out indicated that mining or post-mining areas are characterised by a significantly lower proportion of treated wastewater and higher water consumption and abstraction for industrial purposes. This implies a worse water management situation for mining-related areas.

Keywords: mining, Silesia, spatial autocorrelation, water management, ecology

Wojciech Świercz

e-mail: 187223@ue.wroc.pl

ORCID: 0009-0002-1906-842X

Radosław Szostak

e-mail: 187234@ue.wroc.pl

ORCID: 0009-0006-0156-7297

Uniwersytet Ekonomiczny we Wrocławiu

Modelowanie ekonometryczne kapitału inwestorów indywidualnych na rynku głównym GPW

DOI: 10.15611/2024.76.5.10

JEL: C510

© 2024 Wojciech Świercz, Radosław Szostak

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na

Cytuj jako: Świercz, W. i Szostak, R. (2024). Modelowanie ekonometryczne kapitału inwestorów indywidualnych na rynku głównym GPW. W: A. Stanimir (red.), *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym* (s. 149-164). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Artykuł prezentuje badanie, którego celem było sprawdzenie, jakie czynniki wpływają na kapitał inwestowany przez inwestorów indywidualnych na Giełdzie Papierów Wartościowych w Warszawie. Analizy zostały oparte na Ogólnopolskim Badaniu Inwestorów z 2023 roku, danych z BDL (GUS), podstawowych statystykach z GPW i innych publikacjach. W artykule podjęto próbę stworzenia modelu opisującego wielkość kapitału inwestorów indywidualnych na GPW oraz pozwalającego na identyfikację czynników makroekonomicznych i psychologicznych, które mogą mieć wpływ na decyzje inwestycyjne. Analiza danych wykazała, że kapitał inwestorów indywidualnych na GPW jest silnie zależny od obrotów sesyjnych oraz średniej ceny indeksu WIG20. Nie można jednoznacznie potwierdzić poprawności stworzonego modelu z powodu stosunkowo niskiego współczynnika determinacji oraz występowania autokorelacji. Istotną dziedziną, która przyczyni się do jednoznacznego orzeczenia o poprawności analiz, są finanse behawioralne.

Słowa kluczowe: giełda papierów wartościowych, inwestor indywidualny, kapitał inwestycyjny, rynek kapitałowy, klasyczna metoda najmniejszych kwadratów

1. Wstęp

Przeciętny inwestor indywidualny jako jednostka ma ograniczony wpływ na rynki finansowe ze względu na swoje niewielkie środki pieniężne. Dlatego powszechnie za ważnych graczy na rynku uznaje się instytucje, które mogą swoimi olbrzymimi zasobami gotówki lub szeroką gamą operacji oddziaływać na rynek. Banki centralne,

państwowe rządy i organy regulacyjne czy też instytucje finansowe, takie jak banki, fundusze inwestycyjne, zakłady ubezpieczeniowe, agencje ratingowe i domy maklerskie, rządzą światem finansów od początków ich powstawania. Najstarszą giełdą, na której pierwsza firma wyemitowała swoje akcje, była giełda amsterdamska założona w XVII wieku. Jak pisze Iwuć (2020), Holenderska Kompania Wschodnioindyjska, utworzona w 1602 roku, była pierwszą firmą, w którą mogło zainwestować wielu średnio zamożnych kupców. Emisja akcji pozwoliła na pozyskanie kapitału niezbędnego do budowy okrętów, na których sprowadzano złoto, srebro, miedź oraz tekstylia z różnych zakątków świata. Akcje mogły być zakupione przez inwestorów, którzy nie mieli tak wielkich środków pieniężnych jak Izabela Kastylijska, królowa Hiszpanii, która sfinansowała drogie i ryzykowne wyprawy Kolumba.

Lowenstein (2008) w szóstej edycji książki Grahama i Dodda *Security Analysis* napisał, że inwestorzy jeszcze w pierwszej połowie XX wieku działali na całkowicie innym rynku niż tym, który widzimy dzisiaj. Kryzysy gospodarcze były wtedy zdecydowanie bardziej poważne w odróżnieniu od dzisiejszych, na ogół łagodnych recesji. Giełda nowojorska była wtedy ekskluzywnym klubem, a inwestowanie było dostępne tylko dla inwestorów instytucjonalnych i bogatych biznesmenów. W dzisiejszych czasach inwestorzy mają dostęp do szerszego spektrum instrumentów finansowych, które często można kupić ułamkowo. Powoduje to, że na giełdzie rosnące znaczenie zaczęli mieć inwestorzy prywatni. W dzisiejszych czasach, wraz z rozwojem idei kapitalistycznej oraz zwiększaniem się świadomości i zamożności osób prywatnych, to właśnie inwestorzy indywidualni coraz częściej lokują swoje środki na giełdach. Duży udział inwestorów indywidualnych niesie za sobą wiele skutków – zarówno pozytywnych, jak i negatywnych. Inwestorzy mogą dokonać inwestycji w różne instrumenty finansowe i zminimalizować ryzyko portfela. Osoby prywatne są też bardziej elastyczne, bo mają większą swobodę w podejmowaniu decyzji inwestycyjnych i mogą szybciej reagować na zmiany rynkowe. Według Dahlquista, Martineza oraz Söderlinda (2017) poprzez aktywnie zarządzany portfel inwestor może osiągać ponadprzeciętne stopy zwrotu. Następną zaletą jest to, że aktywny udział w rynku edukuje inwestora i zwiększa jego umiejętności. Niestety działalność każdego inwestora na rynku wiąże się z ryzykiem finansowym, zwłaszcza w przypadku braku planu inwestycyjnego. Inwestorzy indywidualni pod wpływem emocji mogą podejmować działania nieprzemyślane, co prowadzi do potencjalnych strat finansowych. Dodatkowo, odwołując się do artykułu Ayoub'a i Baławiego (2022), należy stwierdzić, że istnieje ryzyko wystąpienia zjawiska owczego pędu, które ma wpływ na kryzysy finansowe i odgrywa dużą rolę podczas kreowania się baniek finansowych. Wysoki udział inwestorów nieinstytucjonalnych może także wpływać na niestabilność rynku. Zdaniem Lo i in. (2022) rynki finansowe mocno zareagowały na szok wywołany wybuchem wojny w Ukrainie, obniżając ceny wielu aktywów. Jeszcze silniejszy efekt został zaobserwowany dla zmienności aktywów, która znacznie się zwiększyła. Sugeruje to, że niepewność na rynku może wpływać na decyzje inwestorów indywidualnych. W przeciwieństwie do podmiotów instytucjonalnych indywidualni działacze

rynkowy nie ma dostępu do specjalistycznych informacji i analiz, co może skutkować złymi decyzjami. Wraz ze wzrostem liczby inwestorów indywidualnych na rynku rośnie też konkurencja, co sprawia, że osiągnięcie nadzwyczajnych wyników na giełdzie staje się trudniejsze. Aby zbadać, co wpływa na zagregowaną wielkość kapitału alokowanego przez inwestorów indywidualnych na polskiej giełdzie, w artykule zaprezentowano wyniki analizy determinant, tworząc liniowy model ekonometryczny. Dobierając zmienne do modelu, zastosowano metodę grafową, a dla oszacowanego modelu wykonano testy weryfikujące jego poprawność.

W kolejnej części artykułu przedstawiono wyniki przeglądu literatury. Punkt trzeci poświęcono prezentacji danych i wykorzystanych metod. W kolejnym opisano proces, a także wyniki szacowania oraz weryfikacji modelu. W punkcie piątym zawarto wnioski z przeprowadzonej analizy, a w ostatnim podsumowanie przeprowadzonego badania oraz końcowe uwagi.

2. Przegląd literatury

Analiza, której wyniki przedstawiono w artykule, ma na celu odkrycie czynników, które mogą wpłynąć na inwestorów indywidualnych oraz kapitał przez nich inwestowany. Model stworzony na podstawie badania może pomóc lepiej zrozumieć, co ma rzeczywiście znaczący wpływ na inwestora. Model może być użyteczny od strony praktycznej dla: spółek – dla których może wskazać, czy raportowane przez nie wyniki oraz ich forma są dla inwestorów istotne; GPW – przez możliwość dostosowania oferty dla inwestorów, oraz do analiz, a także samych inwestorów – model może pomóc w odkryciu, czym tak naprawdę inwestorzy się kierują, i zmusić ich do samorefleksji na temat strategii inwestowania. Postawiono następujące pytania badawcze:

- Czy możliwe jest stworzenie modelu opisującego kapitał inwestorów indywidualnych na rynku głównym GPW na przestrzeni lat?
- Czy czynniki uznane za ważne w dotychczasowych badaniach okażą się mieć znaczący wpływ?
- Czy na podstawie modelu można stworzyć predykcję co do lat kolejnych?

Jakubiak i in. (1999) przeprowadzili obszerne badanie wpływu różnych czynników na stopę oszczędności Polaków oraz domostw i przedsiębiorstw z krajów będących członkami OECD. Za pomocą ważonej metody najmniejszych kwadratów oszacowali model, z którego wynika, że podniesienie stopy wzrostu PKB *per capita* wpływa istotnie na wzrost stopy oszczędności Polaków. Autorzy artykułu udowodnili, że oszczędności i inwestycje inwestorów są silnie skorelowane.

Według Shillera (2000), media odgrywają kluczową rolę zarówno w ustawianiu sceny dla ruchów na rynku, jak i w samodzielnym wywoływaniu tych ruchów. Wpływ społeczny i interakcje z innymi ludźmi mogą prowadzić do irracjonalnych zachowań inwestorów, którzy pod wpływem owczego pędu mogą popełniać błędy w sposób stadny. Oprócz tego, jak podają Ayoub i Balawi (2022), „owczy pęd” może wynikać także z podążania inwestorów za obecnymi trendami na rynku. Z tego po-

wodu w prezentowanym artykule zdecydowano się opóźnić o jeden okres dane dotyczące zwrotu indeksu S&P500 oraz WIG20. Wraz ze wzrostem stóp zwrotu indeksu autorzy artykułu spodziewają się dla indeksu S&P500 odpływu kapitału na rzecz rynków amerykańskich, a dla indeksu WIG20 wraz z rosnącą stopą zwrotu oczekują zwiększenia napływu kapitału na rynek główny.

Czynniki definiujące zachowania inwestorów indywidualnych stanowią główne zagadnienie rozwijającej się dziedziny finansów, zwanej finansami behawioralnymi (Ritter, 2003). Nauka ta skupia się głównie na indywidualnych cechach psychologicznych, które kształtują powszechne praktyki finansowe i inwestycyjne. W prezentowanym badaniu skorzystano ze zmiennej „dostęp Polaków do Internetu”, ponieważ może mieć ona kluczowy wpływ na lokowanie pieniędzy przez inwestorów indywidualnych na warszawskiej Giełdzie Papierów Wartościowych.

Hassan Al-Tamimi (2006) przeprowadził badanie mające na celu zidentyfikowanie czynników wpływających na decyzje inwestorów. Z 343 ankietowanych inwestorów indywidualnych z Dubaju i Abu Zabi 53,9% odpowiedziało, że płynność rynkowa jest ważnym czynnikiem wpływającym na ich wybory. Spośród ankietowanych 52,6% stwierdziło, że istotnymi czynnikami są przeszłe wyniki spółek, a 46,4% odpowiedziało, że ważne jest wypłacanie dywidend. Spośród ankietowanych osób 35,5% stwierdziło, że swoje wybory opierają na ostatnio zaobserwowanej zmianie cen akcji. Na podstawie przeprowadzonego badania w artykule uwzględniono dane dotyczące średniej stopy dywidend oraz wartości obrotów sesyjnych. Obie zmienne powinny mieć dodatni wpływ na wielkość kapitału inwestorów indywidualnych.

Gümüş, Koç i Agalarova (2013) w swoim badaniu udowodnili, że psychologiczne i demograficzne czynniki mają wpływ na decyzje inwestorów indywidualnych. Badanie przeprowadzono na próbie 384 mieszkańców Turcji i Azerbejdżanu. Wyniki wskazały na to, że na decyzje inwestycyjne mają wpływ zarobki, edukacja, status małżeński, wiek i płeć inwestorów. Na tej podstawie w artykule uwzględniono dane dotyczące średnich zarobków, wraz ze wzrostem których spodziewano się zwiększenia oszczędności inwestorów indywidualnych, a co za tym idzie – inwestowanego przez nich kapitału.

Według badania przeprowadzonego przez Wasilewskiego i Juszczyka (2015) najważniejszymi czynnikami wpływającymi na decyzje inwestycyjne indywidualnych inwestorów są trendy na globalnych rynkach kapitałowych, czyli na głównych giełdach papierów wartościowych, oraz wyniki finansowe przedsiębiorstw. Badanie ugruntowało przekonanie autorów prezentowanego artykułu co do konieczności wykorzystania w prezentowanym badaniu zmiennych dotyczących zwrotu indeksów. Oprócz nich uwzględniono także średnią cenę indeksu WIG20 oraz średnie wskaźniki cena/zysk dla spółek z GPW.

Stowarzyszenie Inwestorów Indywidualnych (Masłowski, 2023) między lipcem a październikiem 2023 roku przeprowadziło ogólnopolskie badania inwestorów obejmujące 4546 inwestorów. Jego celem było stworzenie profilu statystycznego polskiego inwestora indywidualnego. Badanie wykazało, że statystyczny polski in-

westor indywidualny to mężczyzna w wieku około 40 lat, mający wyższe wykształcenie, inwestujący głównie na polskiej giełdzie, który od około 5 lat inwestuje po godzinach w celu odłożenia na emeryturę, dywersyfikacji przychodów i ochrony przed inflacją. Jeśli 81,5% inwestorów z próby znajduje się w przedziale wieku od 26 do 55 lat, to zmiany w ciągu lat w tym przedziale wiekowym mogą wpłynąć na wielkość udziału inwestorów indywidualnych na polskiej giełdzie.

3. Dane i metody analizy

Zgodnie z wnioskami zaprezentowanymi w poprzednim punkcie w prezentowanej analizie wykorzystano następujące zmienne:

- Y – kapitał inwestorów indywidualnych zainwestowany w akcje na rynku głównym GPW w mln zł;
- x_1 – liczba absolwentów studiów magisterskich w danym roku;
- x_2 – procentowy dostęp do Internetu w Polsce;
- x_3 – średnie zarobki w zł;
- x_4 – liczba spółek notowanych na GPW;
- x_5 – zwrot rok do roku indeksu S&P500;
- x_6 – zwrot rok do roku indeksu WIG20;
- x_7 – średnia cena indeksu WIG20 w zł;
- x_8 – wskaźnik C/Z uśredniony dla wszystkich spółek notowanych na GPW;
- x_9 – PKB *per capita* w zł;
- x_{10} – średnia stopa dywidendy w %;
- x_{11} – ludność w wieku 25-54 lata jako odsetek populacji Polski;
- x_{12} – średnia liczba transakcji na sesję;
- x_{13} – wartość obrotów sesyjnych w mln zł;
- x_{14} – PKB w mld dol.;
- x_{15} – procentowa zmiana inflacji rok do roku.

Zmienne pochodzą z baz danych GPW, GUS-BDL, The Global Economy, ZUS, Slickcharts oraz The World Bank. Dokładne źródła podano w spisie literatury. Dane dotyczą lat 2002-2022.

Do stworzenia modelu ekonometrycznego autorzy wykorzystali klasyczną metodę najmniejszych kwadratów. Jest ona oparta na następujących założeniach Gaussa-Markova (Dziechciarz, 2003):

- model jest niezmienniczy ze względu na obserwacje;
- model jest liniowy względem parametrów;
- zmienne objaśniane są nielosowe i ustalone w powtarzalnych próbach;
- nie występuje współliniowość między zmiennymi objaśniającymi;
- reszty przyjmują rozkład normalny, a wartość oczekiwana (nadzieja matematyczna) reszt jest równa 0;
- składnik losowy ma stałą wariancję, nie występuje autokorelacja;
- estymacja może odbywać się tylko na podstawie informacji zawartych w próbie.

Aby sprawdzić, czy założenia są spełnione, a także czy model jest poprawnie skonstruowany, przeprowadzono szereg testów statystycznych, które są następujące (Dziechciarz, 2003):

- test Shapiro-Wilka, który sprawdza, czy błędy składnika losowego mają rozkład normalny;
- test Hellwiga, który także sprawdza normalność rozkładu reszt;
- test losowości serii, sprawdzający, czy reszty modelu są losowe, użyty w celu oceny trafności wyboru postaci analitycznej modelu;
- test *t*-Studenta pozwalający na dobór takich zmiennych objaśniających, które istotnie wpływają na zmienną objaśnianą;
- test Goldfelda-Quandta, badający, czy wariancja reszt modelu jest homoskedastyczna, tzn. stała;
- test White'a, badający, czy wariancja jest heteroskedastyczna;
- test symetrii składnika losowego, sprawdzający, czy liczba reszt *in plus* jest dostatecznie równa liczbie reszt *in minus*;
- test Durbina-Watsona, sprawdzający, czy w modelu występuje autokorelacja pierwszego rzędu.

4. Szacowanie i weryfikacja modelu

4.1. Dobór zmiennych

Zmienne $x_1 - x_{15}$ stanowią zbiór potencjalnych zmiennych objaśniających. Aby dobrać zmienne do modelu, porównano wynik uzyskany po zastosowaniu metody grafowej z wynikiem otrzymanym na podstawie metody analizy współczynników korelacji na poziomie istotności 0,05. W przypadku rozbieżnych wyników postaci modelu wybrano na podstawie przesłanek wynikających z literatury – zdecydowano się na taki, którego zmienne były lepiej ugruntowane merytorycznie oraz poparte literaturą.

4.1.1. Współczynnik zmienności

Aby zmienne miały dostateczną zmienność, przyjęto minimalny poziom współczynnika zmienności równy 20% ze względu na wysoką zmienność zjawiska objaśnianego, a także często obserwowaną wysoką zmienność w przypadku danych dotyczących zjawisk finansowych. Tabela 1 przedstawia zmienność poszczególnych zmiennych.

Tabela 1. Współczynnik zmienności zmiennych (%)

Zmienna	y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
V	47	17	15	30	22	154	376	20
Zmienna	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
V	73	35	27	0,01	55	36	26	111

* V – współczynnik zmienności; wartości oznaczone na szaro nie spełniają ustalonych założeń.

Źródło: opracowanie własne.

Opierając się na współczynniku zmienności, odrzucono zmienne x_1 , x_2 oraz x_{11} . Nie spełniły one wyznaczonego kryterium – ich zmienność statystyczna jest niewystarczająca, aby mogły znaleźć się w modelu.

4.1.2. Dobór zmiennych metodą analizy współczynników korelacji

Na poziomie istotności 0,05 minimalna korelacja zmiennej objaśnianej ze zmienną objaśniającą musi być większa niż $r^* = 0,44376$. Wartość r^* została obliczona na podstawie wzoru (Dziechciarz, 2003):

$$r^* = \sqrt{\frac{t^2}{t^2 + n - 2}},$$

gdzie: n – liczebność próby, t – wartość statystyki odczytana z tablic t -Studenta dla poziomu istotności α i $n - 2$ stopni swobody.

W tabeli 2 przedstawiono poszczególne współczynniki korelacji Pearsona między zmienną objaśnianą a zmiennymi objaśniającymi.

Tabela 2. Współczynniki korelacji liniowej Pearsona

Zmienna	X_3	X_4	X_5	X_6	X_7	X_8
y	0,370057	0,162715	0,262305	0,016706	0,435583	-0,17419
Zmienna	X_9	X_{10}	X_{12}	X_{13}	X_{14}	X_{15}
y	0,374189	0,148871	0,67104	0,729543	0,439187	0,545878

* Na szaro oznaczono wartości nieosiągające dostatecznego poziomu.

Źródło: opracowanie własne.

Na podstawie współczynników korelacji odrzucono zmienne x_3 , x_4 , x_5 , x_6 , x_7 , x_8 , x_9 , x_{10} , x_{14} . Największą korelację ze zmienną objaśnianą ma zmienna x_{13} – 0,729. Następnie sprawdzono korelację zmiennej x_{13} ze zmiennymi x_{12} , x_{14} oraz x_{15} . Korelacja musi być mniejsza niż r^* .

Tabela 3. Współczynniki korelacji między zmienną x_{13} a zmiennymi x_{12} , x_{14} , x_{15}

Zmienna	X_{12}	X_{14}	X_{15}
X_{13}	0,880458	0,887346	0,501323

* Na szaro oznaczono wartości niespełniające kryterium.

Źródło: opracowanie własne.

Na podstawie analizy współczynników korelacji zmiennej x_{13} ze zmiennymi będącymi jeszcze kandydatkami do modelu odrzucono zmienne x_{12} , x_{14} oraz x_{15} . Z wykorzystaniem metody analizy współczynników korelacji do modelu dobrano zmienną x_{13} .

4.1.3. Dobór zmiennych metodą grafową

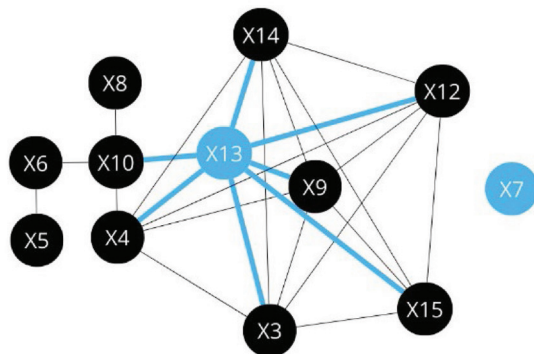
W kolejnym kroku prowadzonej analizy dobór zmiennych został przeprowadzony za pomocą metody grafowej z kryterium r^* . Wartość kryterium to 0,44376, jak przy metodzie analizy współczynników korelacji. Kryterium zostało obliczone w ten sam sposób. Na jego podstawie zbudowana została macierz R' .

Tabela 4. Macierz R'

Zmienne	x3	x4	x5	x6	x7	x8	x9	x10	x12	x13	x14	x15
x3	0	0,674219	0	0	0	0	0,998427	0	0,922051	0,785902	0,895559	0,615811
x4	0,674219	0	0	0	0	0	0,676506	0,569573	0,581659	0,745457	0,800781	0
x5	0	0	0	0,583619	0	0	0	0	0	0	0	0
x6	0	0	0,583619	0	0	0	0	0,48749	0	0	0	0
x7	0	0	0	0	0	0	0	0	0	0	0	0
x8	0	0	0	0	0	0	0	0,48358	0	0	0	0
x9	0,998427	0,676506	0	0	0	0	0	0	0,919189	0,792011	0,899375	0,620194
x10	0	0,569573	0	0,48749	0	0,48358	0	0	0	0	0	0
x12	0,922051	0,581659	0	0	0	0	0,919189	0	0	0,880458	0,837831	0,681386
x13	0,785902	0,745457	0	0	0	0	0,792011	0	0,880458	0	0,887346	0,501323
x14	0,895559	0,800781	0	0	0	0	0,899375	0	0,837831	0,887346	0	0,535907
x15	0,615811	0	0	0	0	0	0,620194	0	0,681386	0,501323	0,535907	0

Źródło: opracowanie własne.

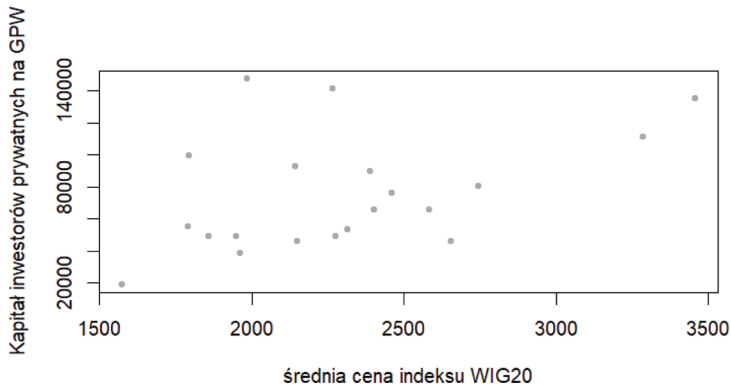
Opierając się na macierzy R' zbudowano graf połączeń.



Rys. 1. Graf połączeń

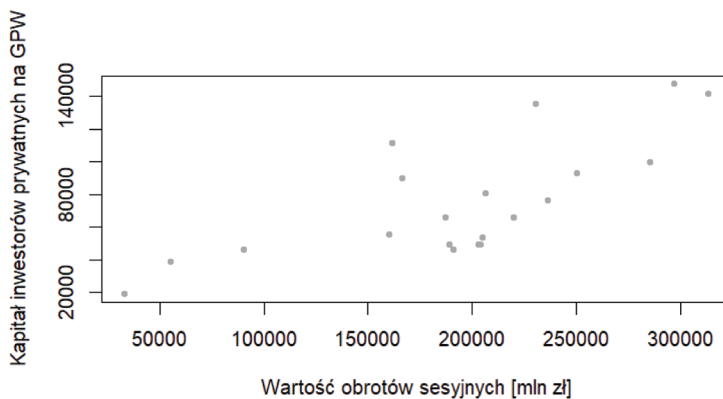
Źródło: opracowanie własne.

Powstały dwa grafy. Na podstawie grafu złożonego z większej liczby zmiennych stwierdzono, że największą liczbą połączeń cechują się wierzchołki dla zmiennych $x_3, x_4, x_9, x_{12}, x_{13}, x_{14}$. Z tych zmiennych największą korelację ze zmienną objaśnianą wykazuje zmienna x_{13} . To ona zostanie wybrana do modelu. Drugi graf składa się tylko ze zmiennej x_7 , z tego powodu na pewno trafi ona do modelu. Na podstawie metody grafowej do modelu wybrano zmiennę x_7 oraz x_{13} . Zależności zmiennej endogenicznej od zmiennych x_7 oraz x_{13} można zaobserwować na rys. 2 oraz 3.



Rys. 2. Zależność Y od x_7

Źródło: opracowanie własne.



Rys. 3. Zależność Y od x_{13}

Źródło: opracowanie własne.

Korelacja liniowa między zmiennymi przedstawiona na rys. 3 jest silniejsza niż ta na rys. 2. W przyszłych badaniach należy sprawdzić występowanie innych korelacji niż liniowa między Y a x_7 . Na rysunku 3 widać występowanie korelacji liniowej z występowaniem wartości odstających, szczególnie dla poziomów wartości obrotów sesyjnych wyższych niż 150 000.

4.1.4. Podsumowanie doboru zmiennych

Ostatecznie do modelu wybrano zmienne dobrane na podstawie metody grafowej. Przesłanki wynikające z literatury wskazują, że zmienna x_7 jest zmienną istotną, co dodatkowo potwierdza fakt, że WIG20 jest indeksem zawierającym 20 spółek

z największą kapitalizacją, a co za tym idzie – są to spółki najbardziej popularne wśród inwestorów zarówno prywatnych, jak i instytucjonalnych. O ich popularności świadczą dzienne wolumeny obrotu oraz liczba transakcji. Oprócz tego skorygowany współczynnik determinacji jest znacznie większy dla modelu stworzonego na podstawie metody grafowej. Odpowiednio współczynnik ten wynosi 62% dla doboru metodą grafową oraz 51% dla doboru metodą analizy współczynników korelacji. Wstępna postać modelu wygląda następująco:

$$y = \alpha_0 + \alpha_1 x_7 + \alpha_2 x_{13} + \varepsilon,$$

gdzie: ε – reszty modelu; α_1, α_2 – parametry strukturalne modelu

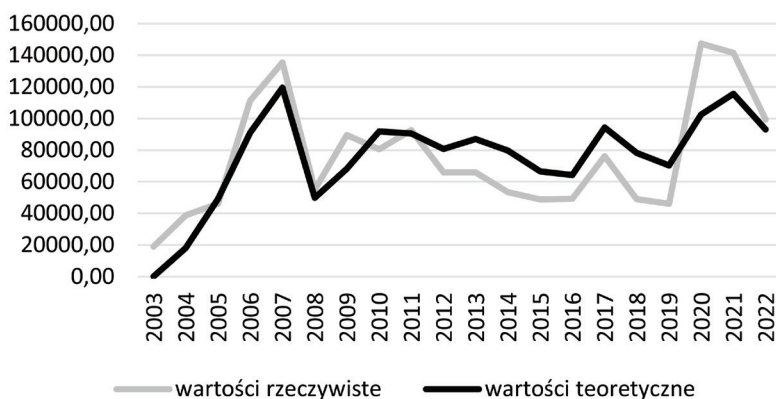
4.2. Oszacowanie modelu oraz test t-Studenta

Po oszacowaniu modelu przybiera on następującą postać:

$$\hat{y} = -54147,14 + 27,17x_7 + 0,35x_{13}.$$

Współczynnik R^2 wynosi 0,661. Oznacza to, że zmienność zmiennej objaśnianej jest w 66,1% wyjaśniana przez zmienne objaśniające. Aby sprawdzić, czy zmienne x_7 oraz x_{13} mają istotny wpływ na zmienną objaśnianą, przeprowadzono test t-Studenta. Przy poziomie istotności 0,05 obie zmienne okazały się mieć istotny wpływ na zmienną Y . Wartości p -value wyniosły odpowiednio 0,02 dla statystyki t parametru przy zmiennej x_7 oraz 0,001 dla statystyki t parametru przy zmiennej x_{13} .

Na rysunku 4 przedstawiono wartości rzeczywiste i wartości teoretyczne zmiennej objaśnianej.



Rys. 4. Wartości teoretyczne i rzeczywiste

Źródło: opracowanie własne.

Jak widać na rys. 4, wartości teoretyczne dość dobrze odzwierciedlają wartości rzeczywiste, a wartości teoretyczne na koniec badanego okresu cechują się mniejszymi wahaniami w porównaniu z wartościami rzeczywistymi.

4.3. Testy weryfikujące poprawność modelu

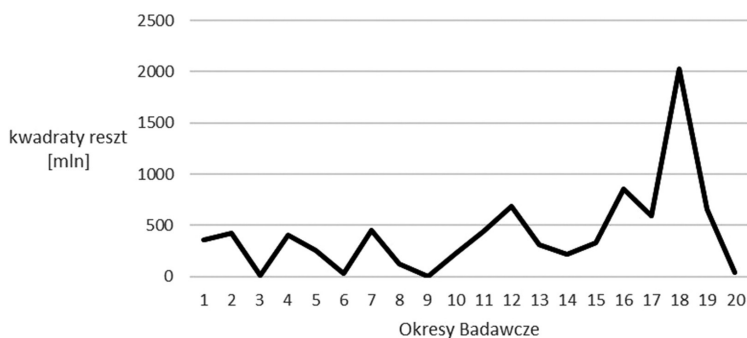
Poziom istotności został wybrany w celu minimalizacji ryzyka popełnienia błędu pierwszego rodzaju.

Przeprowadzono test Hellwiga, który weryfikuje jedno z założeń metody KMNK o rozkładzie normalnym błędów składnika losowego. Przy liczbie obserwacji 20 oraz przy poziomie istotności równym 0,05 odczytano z tablic wartości obszaru krytycznego, który wyniósł $\langle 4; 9 \rangle$. Liczba pustych cel wynosi 7. Jako że liczba cel pustych znajduje się w obszarze krytycznym, można stwierdzić, że model spełnia założenie o normalności rozkładu składnika losowego.

Aby upewnić się co do normalności rozkładu składnika losowego, przeprowadzono również test Shapiro-Wilka przy poziomie istotności 0,05. Zakres krytyczny testu to $\langle 0; 0,905 \rangle$. Statystyka testu wyniosła 0,936, więc można stwierdzić, że nie ma podstaw do odrzucenia hipotezy o normalności składnika losowego.

Test serii Walda-Wolfowitza weryfikuje hipotezę o losowości rozkładu odchyleń losowych modelu. Test ma na celu ocenę trafności wyboru postaci analitycznej modelu. Liczba serii wyniosła 7, liczba błędów większych od zera – 10, podobnie jak mniejszych niż zero – również 10. Przy poziomie istotności równym 0,05 liczba serii zawiera się w obszarze krytycznym (odczytanym z tablic) $\langle 6; 15 \rangle$, więc można stwierdzić, że rozkład odchyleń losowych jest losowy.

Aby zweryfikować, czy model ma heteroskedastyczną wariancję, reszt przeprowadzono test Goldfelda-Quandt. W tym celu podzielono kwadraty reszt na 2 równe podgrupy i przy poziomie istotności równym 0,05 okazało się, że wartość teoretyczna testu jest większa niż wartość statystyki testu. Na tej podstawie stwierdzono, że występuje homoskedastyczność wariancji reszt, co jest zjawiskiem pozytywnym.



Rys. 5. Kwadraty reszt

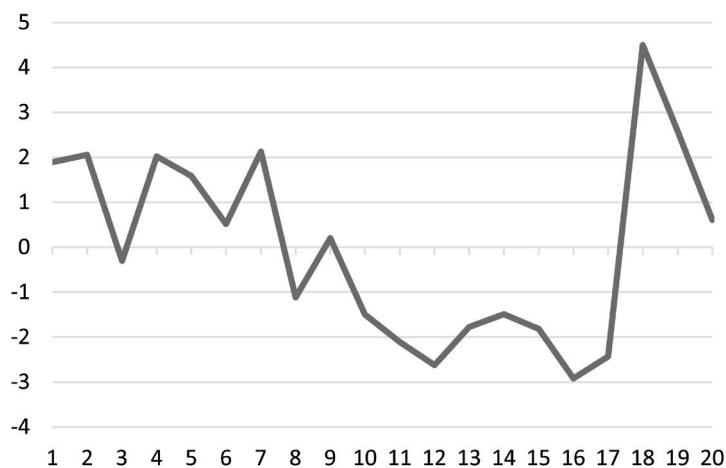
Źródło: opracowanie własne.

Aby sprawdzić, czy wariancja reszt jest heteroskedastyczna, przeprowadzono test White'a. Obszar odrzucenia hipotezy 0 mówiącej, że wariancja reszt nie jest heteroskedastyczna, zawierał się w przedziale $(11,0705; +\infty)$ przy poziomie istotności 0,05. Wartość statystyki testowej wyniosła 13,21. Należy więc odrzucić hipotezę 0. Wariancja reszt jest heteroskedastyczna.

Jak widać, test Goldfelda-Quandt oraz test White'a nie dały tego samego rezultatu. Dla przeprowadzonego badania przyjęty został wynik testu Goldfelda-Quandt ze względu na jego przystosowanie do małych prób (Cichocki, 2012).

Aby sprawdzić, czy składniki losowe są symetryczne, na poziomie istotności 0,05 przeprowadzono test symetrii składnika losowego. Wartość reszt *in plus* wyniosła 10, co przy próbie $n = 20$ dało wartość statystyki testowej równą 0. Obszar odrzucenia hipotezy 0 mówiącej o symetrii składnika losowego zawierał się w przedziale $(2,09; +\infty)$. Nie można więc odrzucić hipotezy 0 mówiącej, że składnik losowy jest symetryczny.

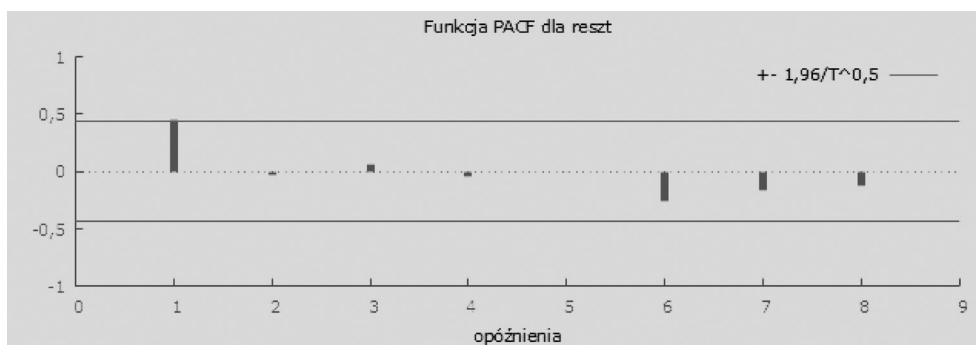
Przeprowadzono test Durбина-Watsona sprawdzający, czy między resztami występuje autokorelacja pierwszego rzędu. Statystyka testowa wyniosła 1,05. Przy poziomie istotności 0,05 statystyka testowa wpadła do przedziału $(0; 1,1)$. Na podstawie testu Durбина-Watsona należy odrzucić hipotezę 0 na rzecz hipotezy 1, mówiącej o występowaniu autokorelacji reszt rzędu pierwszego.



Rys. 6. Rozkład reszt

Źródło: opracowanie własne.

Na podstawie wykresu rozkładu reszt można się spodziewać występowania dodatniej autokorelacji. Wykres rozkładu reszt nie przypomina typowych wykresów białego szumu, co oznacza, że można podejrzewać występowanie autokorelacji i brak losowości reszt (Strawiński, 2005).



Rys. 7. Korelogram

Źródło: opracowanie własne.

Na podstawie korelogramu można zauważyć występowanie autokorelacji pierwszego rzędu. Potwierdza to wnioski wyciągnięte z wykresu rozkładu reszt oraz testu Durбина-Watsona.

4.4. Podsumowanie oceny analitycznej modelu

Na podstawie przeprowadzonych testów można zauważyć, że model spełnia wszystkie założenia Gaussa-Markova, oprócz założeń o braku autokorelacji reszt, gdyż test Durбина-Watsona wykazał występowanie autokorelacji. Model opisuje zmienność y na poziomie 66,1%, co jest wynikiem nie w pełni zadowalającym ze względu na istotność w analizie danych finansowych nawet najmniejszych zróżnicowań. Wzrost zmiennej x_7 o jedną jednostkę, czyli wzrost średniej ceny indeksu WIG20 o złotówkę, powoduje wzrost kapitału inwestorów indywidualnych o 27,17 jednostek przy założeniach *ceteris paribus*. Natomiast przyrost zmiennej x_{13} , czyli wartości obrotów sesyjnych o jedną jednostkę, prowadzi do zwiększenia y teoretycznego o 0,35 jednostek przy założeniach *ceteris paribus*. Ze względu na występowanie autokorelacji pierwszego rzędu nie została przeprowadzona predykcja.

5. Wnioski

Stworzenie modelu opisującego kapitał zainwestowany przez inwestorów indywidualnych na rynku głównym GPW wymaga dalszego dopracowania. Jako potencjalne zmienne do dalszych badań należy rozważyć średnią cenę indeksu WIG20 oraz wartość obrotów sesyjnych w milionach zł. Możliwa jest potrzeba poszukania kolejnych zmiennych lub próba integracji finansów behawioralnych w tworzenie modelu. Stosunkowo niski współczynnik determinacji oraz występowanie autokorelacji nie pozwala na pozostawienie stworzonego modelu jako ostatecznego. Podejrzenia

co do przyczyny występowania autokorelacji dotyczą psychologii podejmowania decyzji przez inwestorów indywidualnych oraz dużej zależności badanego zjawiska i czynników na niego wpływających od zjawisk gospodarczych, takich jak inercja zjawisk gospodarczych. Na decyzję inwestorów mogą mieć wpływ trendy, które występują na rynku, a także sentyment rynkowy, tj. obecne nastawienie wszystkich uczestników rynku kapitałowego do rynku, sektora lub spółek. Sentyment rynkowy jest określany jako optymistyczny w przypadku wzrostu cen aktywów i pesymistyczny w przypadku spadków cen (Smith, 2023). Oprócz tego należy wziąć pod uwagę trendy samych aktywów, takie jak zmiana z ryzykowniejszych inwestycji w akcje na obligacje lub od niedawna rosnącą popularność kryptowalut, które mogą odciągać z rynku akcji znaczną część kapitału. Nie jest możliwe udzielenie odpowiedzi na zaistniałe pytania badawcze, konieczne jest przeprowadzenie dalszych badań oraz próba zastosowania innych modeli.

6. Zakończenie

Zaprezentowana w artykule próba stworzenia modelu opisującego kształtowanie się kapitału inwestorów indywidualnych na rynku głównym GPW zakończyła się niepowodzeniem. Jednak na podstawie powyższego badania także można wyciągnąć ważne wnioski co do istotności średniej ceny indeksu WIG20 oraz wartości obrotów sesyjnych jako zmiennych objaśniających, które przydadzą się w kolejnych próbach stworzenia modelu. Trzeba także zwrócić uwagę na istotność finansów behawioralnych w badanym zjawisku, a tym samym nieilościową postać niektórych czynników.

Dynamika zmian na świecie utrudnia stworzenie modelu. Ponadto sytuacja komplikuje się ze względu na powstawanie nowych, nieznanych inwestorom w przeszłości aktywów. W następnej pracy badawczej autorzy podejmą się stworzenia modelu na podstawie innych, być może bardziej odpowiednich, metod.

Literatura

- Al-Tamimi, H. A. H. (2006). Factors Influencing Individual Investor Behaviour: An Empirical Study of the UAE Financial Markets. *The Business Review, Cambridge*, 5(2), 225-232.
- Ayoub, A., i Balawi, A. (2022). Herd Behavior and its Effect on the Stock Market: An Economic Perspective. *Quality Access to Success*, 23(188). <https://doi.org/10.47750/QAS/23.188.38>
- Cichocki, S. i Nehrebecka, N. (2012). *Testy diagnostyczne. Problemy z danymi*. Pobrano z <http://www.ekonometria.wne.uw.edu.pl/uploads/Main/Wyklad%2003.01.2012.pdf>
- Dahlquist, M., Martinez, J. V. i Söderlind, P. (2017). Individual Investor Activity and Performance. *The Review of Financial Studies*, 30(3). <https://doi.org/10.1093/rfs/hhw093>
- Giełda Papierów Wartościowych. (2023). *Udział inwestorów w obrotach giełdowych – dane zagregowane*. <https://www.gpw.pl/analizy>

- Giełda Papierów Wartościowych. (2024). *Podstawowe statystyki GPW*. <https://www.gpw.pl/podstawowe-statystyki-gpw>
- Główny Urząd Statystyczny. Baza Danych Lokalnych. (2024). *Absolwenci studiów na poziomie magisterskim ogółem. Produkt krajowy brutto na 1 mieszkańca. Ludność wg grup wieku i płci. Roczne wskaźniki cen towarów i usług konsumpcyjnych od 1950 roku*.
- Graham, B. i Dodd, D. L. (2008). *Security Analysis 6th Edition Updated with New Commentary by Seth A. Klarman, James Grant, Bruce Greenwald, and Others*. Mc Graw Hill. <https://pdfcoffee.com/security-analysis-sixth-edition-foreword-by-warren-buffett-pdf-free.html>
- Gümüş, F. B., Koç, M. i Agalarova, M. (2013). A Study on the Identification of Effective Demographic and Psychological Factors on the Investment Decisions of Individual Investors: Turkey and Azerbaijan. *Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(6), 71-93.
- Iwuć, M. (2020). *Finansowa Forteca*. FBO Marcin Iwuć.
- Jakubiak, M., Kaczorowski, P., Kowalski, P., Liberda, B., Siwińska, J. i Tokarski, T. (1999). Determinanty oszczędzania w Polsce. *Centrum Analiz Społeczno-Ekonomicznych*, (28).
- Lo, G.-D., Marcelin, I., Bassène, T. i Sène, B. (2022). The Russo-Ukrainian War and Financial Markets: The Role of Dependence on Russian Commodities. *Finance Research Letters*, (50), 103194.
- Lowenstein, R. (2008). *Introduction. Security Analysis* (6th ed.).
- Masłowski, M. (2023). *Stowarzyszenie Inwestorów Indywidualnych. 21 edycja Ogólnopolskie Badanie Inwestorów. Długi termin, ETF-y i zagranica, czyli w co i jak inwestuje polski inwestor w 2023 r.* Pobrano 20 lutego 2024 z <https://www.sii.org.pl/16885/aktualnosci/badania-i-rankingi/ogolnopolskie-badanie-inwestorow-2023-wyniki.html>
- Ritter, J. R. (2003). Behavioral Finance. *Pacific-Basin Finance Journal*, (11), 429-437. [https://doi.org/10.1016/S0927-538X\(03\)00048-9](https://doi.org/10.1016/S0927-538X(03)00048-9)
- Shiller, R. C. (2000). Irrational Exuberance. *Philosophy and Public Policy Quarterly*, (20), 18-23.
- Slickcharts. (2024). *S&P 500 Total Returns by Year*. Pobrano z <https://www.slickcharts.com/sp500/returns>
- Smith, T. (2023). *What Is Market Sentiment? Definition, Indicator Types, and Example*. Investopedia. Pobrano 22 lutego 2024 z <https://www.investopedia.com/terms/m/market-sentiment.asp>
- Strawiński, P. (2005). *Notatki do ćwiczeń z ekonometrii. 2.2 Autokorelacja*. Pobrano 22 lutego 2024 z <https://coin.wne.uw.edu.pl/pstrawinski/notatki/auto.pdf>
- The Global Economy. (2023). *Poland, Internet Users, Percent of Population*. Pobrano z https://www.theglobaleconomy.com/Poland/Internet_users/
- The World Bank. (2023). *GDP (current US\$)*. Pobrano z <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>
- Wasilewski, M. i Juszczyk, M. (2015). Czynniki kształtujące decyzje inwestorów na rynku kapitałowym. *Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 855: Finanse, Rynki Finansowe, Ubezpieczenia*, 74(1).
- Zakład Ubezpieczeń Społecznych. (2023). *Przeciętne wynagrodzenie od 1950 r.* Pobrano z <https://www.zus.pl/baza-wiedzy/skladki-wskazniki-odsetki-wskazniki/przecietne-wynagrodzenie-w-latach>

Econometric Modelling of Individual Investors' Capital on GPW Main Market

Abstract: The article presents a study aimed at examining the factors that influence the capital invested by individual investors on the Warsaw Stock Exchange (WSE). The analyses were based on, among others, the 2023 National Investor Survey, data from the Local Data Bank (GUS), basic statistics from the WSE, and other publications. The article attempts to create a model that describes the volume of individual investors' capital on the WSE and identifies macroeconomic and psychological factors that may influence investment decisions. The data analysis showed that the capital of individual investors on the WSE is strongly dependent on session turnovers and the average price of the WIG20 index. However, the accuracy of the created model cannot be definitively confirmed due to the relatively low coefficient of determination and the presence of autocorrelation. Behavioral finance is a crucial field that will contribute to a clear conclusion on the correctness of the analyses.

Keywords: stock exchange, individual investor, seed capital, capital market, Least Squares Method