

**Joanna Kisielińska, Bolesław Borkowski
Katarzyna Czech, Anna Górka, Grzegorz Koszela
Monika Krawiec, Joanna Landmesser, Luiza Ochnio
Łukasz Pietrych, Robert Pietrzykowski
Ewa Wasilewska, Monika Zielińska-Sitkiewicz**

Wielowymiarowa analiza danych w ekonomice rolnictwa



**Wydawnictwo SGGW
Warszawa 2021**

© Copyright by Wydawnictwo SGGW, Warszawa 2021
Wydanie I

Recenzenci: prof. dr hab. Karol Kukuła

dr hab. inż. Stanisław Gędek – prof. PRz

Projekt okładki i strony tytułowej – Violetta Kaska
Grafika na okładce – https://ml4a.github.io/demos/f_mnist_1layer/
Opracowanie redakcyjne – Anna Dołomisiewicz

ISBN 978-83-8237-041-6

Wydawnictwo SGGW
ul. Nowoursynowska 161, 02-787 Warszawa
tel. 22 593 55 20
e-mail: wydawnictwo@sggw.edu.pl
www.wydawnictwosggw.pl

Druk: Libra-Print, al. Legionów 114B, 18-400 Łomża

Spis treści

Wstęp	5
1. Regresja wieloraka	9
1.1. Opis metody – modele regresyjne	9
1.2. Zastosowania regresji wielorakiej	13
Literatura	24
2. Regresja panelowa	29
2.1. Opis metod	31
2.2. Zastosowania metod regresji panelowej	37
Literatura	43
3. Analiza wariacji	45
3.1. Opis metod	45
3.1.1. Jednoczynnikowa i dwuczynnikowa analiza wariacji	45
3.1.2. Wielowymiarowa analiza wariacji	50
3.2. Zastosowania analizy wariacji	51
Literatura	57
4. Wzorcowe metody klasyfikacji (grupowania) obiektów	61
4.1. Opis metod	61
4.1.1. Analiza dyskryminacyjna	61
4.1.2. Modele prawdopodobieństwa	65
4.1.3. Sieci neuronowe	67
4.2. Zastosowania wzorcowych metod klasyfikacji	68
Literatura	76
5. Bezwzorcowe metody klasyfikacji (grupowania) obiektów	81
5.1. Opis metod	81
5.1.1. Hierarchiczne metody aglomeracyjne i deglomeracyjne	83
5.1.2. Metody obszarowe	86
5.1.3. Metody optymalizujące wstępny podział obiektów	88
5.1.4. Weryfikacja uzyskanej klasyfikacji	90
5.2. Zastosowania metod klasyfikacji bezwzorcowej	91
Literatura	99
6. Przestrzenna analiza taksonomiczna	105
6.1. Opis metod	106
6.2. Zastosowania metod przestrzennej analizy taksonomicznej	116
Literatura	121

7. Metody porządkowania liniowego	125
7.1. Opis metod	125
7.1.1. Dobór i selekcja zmiennych diagnostycznych	126
7.1.2. Normalizacja zmiennych	126
7.1.3. Wyznaczenie zmiennej lub zmiennych syntetycznych	128
7.1.4. Wybór rankingu ostatecznego	132
7.1.5. Podział zbiorowości na klasy	133
7.2. Zastosowania metod porządkowania liniowego	135
Literatura	143
8. Metoda głównych składowych oraz analiza czynnikowa	147
8.1. Opis metod	147
8.1.1. Metoda głównych składowych	147
8.1.2. Analiza czynnikowa	149
8.1.3. Porównanie analizy czynnikowej z metodą głównych składowych	154
8.2. Zastosowania metod głównych składowych i analizy czynnikowej	154
Literatura	160
9. Analiza struktury	163
9.1. Opis metod	164
9.1.1. Analiza koncentracji	168
9.2. Zastosowania metod	170
Literatura	177

Wstęp

Rozwój społeczno-gospodarczy skutkuje pojawianiem się w rzeczywistości gospodarczej wielu aktywnych podmiotów. Podmioty te są ze sobą powiązane, tworząc złożony system wzajemnych zależności. Powiązania te w coraz większym stopniu nie ograniczają się do skali jednego, kilku czy kilkunastu krajów, ale mają zakres globalny. Rzeczywistość gospodarcza staje się coraz bardziej złożona, czego skutkiem jest między innymi generowanie mnogości informacji. Analiza tych informacji może przyczynić się do zmniejszenia ryzyka związanego z działalnością gospodarczą. Poprzez usystematyzowanie czy zsyntetyzowanie danych można dostarczyć osobom zarządzającym na różnych szczeblach narzędzi ułatwiających podejmowanie racjonalnych decyzji ekonomicznych.

Zmiany systemów gospodarczych, w jakich funkcjonują jednostki, nie omijają rolnictwa. Wstąpienie w 2004 roku Polski do Unii Europejskiej zapoczątkowało jego integrację ze strukturami europejskimi, co skutkowało wieloma zmianami w jego funkcjonowaniu. Najważniejsze z nich to postępująca koncentracja i specjalizacja. Tradycyjne gospodarstwa rolnicze prowadzące różne rodzaje działalności w niewielkiej skali albo wycofują się z działalności towarowej, albo zamieniają się w wyspecjalizowane dynamicznie rozwijające się przedsiębiorstwa. Sprawne zarządzanie nimi w coraz większym stopniu wymaga wsparcia, tak jak w przypadku podmiotów działających w innych branżach.

Zamiarem autorów niniejszej monografii jest dostarczenie informacji o grupie metod umożliwiających analizy złożonych i nawet bardzo dużych zbiorów danych. Są to metody zaliczane do grupy metod wielowymiarowej analizy danych (WAD). Wielowymiarowa analiza danych zajmuje się badaniem zbiorów obiektów opisanych wieloma cechami. Aktualnie zakres ten może zostać rozszerzony o dodatkową obserwację tych obiektów w czasie.

Każdy rozdział monografii jest poświęcony jednej metodzie i stanowi kompletną całość obejmującą część teoretyczną (przeznaczenie metody wraz z jej opisem), praktyczną (jej zastosowanie do rozwiązania problemów z zakresu ekonomiki rolnictwa) oraz wykaz literatury. Czytelnik, który jest zainteresowany konkretną metodą, swoje poszukiwania może ograniczyć do jednego rozdziału bez potrzeby zagłębiania się w rozdziały pozostałe. Z tego też względu może się zdarzyć, że niektóre treści będą się powtarzać, choć zespół autorski starał się tego unikać. Przegląd zastosowań ma na celu pomoc w zakresie aplikacji danej metody do określonego problemu badawczego. Jak wspomniano, część praktyczna ogranicza się do zastosowań metod WAD w ekonomice rolnictwa. Nie oznacza to

jednak, że czytelnik zainteresowany innymi obszarami gospodarki nie może znaleźć tu potrzebnych informacji. Wręcz przeciwnie – problemy często są bardzo podobne, choć dotyczą innych podmiotów, zjawisk czy zdarzeń.

Monografię podzielono na dziewięć rozdziałów. Dwa pierwsze poświęcono modelom regresyjnym. Metody te stosuje się do określenia postaci matematycznych zależności między zjawiskiem, którego zmiany mają być wyjaśnione (określone jako zmienna zależna), a zjawiskami czy wielkościami, które zjawisko to kształtują (zmiennie niezależne). W modelach regresyjnych zakłada się, że zmiennie w nich uwzględnione są zmiennymi ilościowymi. W pierwszym rozdziale opisano przede wszystkim liniowy model regresji wielorakiej, w drugim rozdziale zaś modele panelowe. Modele panelowe są szacowane na podstawie danych przekrojowo-czasowych. Omówiono tu modele statyczne z efektami ustalonymi i efektami losowymi – ich estymację oraz weryfikację statystyczną.

W dwóch kolejnych rozdziałach opisano modele, które stanowią pewne uogólnienie modeli regresyjnych na modele ze zmiennymi jakościowymi. Jeśli zmienna zależna jest zmienną ilościową, a zmienna/zmiennie niezależne są jakościowe, do analizy danych stosujemy analizę wariancji przedstawioną w trzecim rozdziale. Analiza wariancji pozwala wychwycić (zobrazować w postaci modelu stochastycznego) różnice w wartościach badanego zjawiska dla różnych poziomów czynników, danych jako zmiennie jakościowe. Jeśli natomiast zmienna zależna jest jakościowa, a zmiennie niezależne ilościowe, tworzone są modele dyskryminacyjne omówione w czwartym rozdziale. Zmienna jakościowa reprezentuje wówczas klasy, jakie występują w badanej zbiorowości. Zadaniem oszacowanego modelu dyskryminacyjnego jest określenie, do której z klas występujących w zbiorowości należy obiekt, dla którego jest on stosowany. Analiza dyskryminacji nazywana jest metodą klasyfikacji wzorcowej. Wynika to z tego, że jej zadaniem jest przypisanie badanego obiektu do jednej z klas, których liczba i rodzaje są znane.

Piąty rozdział zawiera omówienie metod umożliwiających bezwzorcową klasyfikację obiektów. Metody te nazywane są także analizą skupień. W tym przypadku poszukiwany jest podział zbiorowości na klasy zawierające obiekty podobne. Ten rodzaj klasyfikacji nazywa się klasyfikacją bezwzorcową, ponieważ ani liczba, ani charakterystyka klas nie jest znana.

Szósty rozdział, poświęcony przestrzennej analizie taksonomicznej, nawiązuje do piątego rozdziału. Jego celem jest pogrupowanie obiektów, ale w tym wypadku obiektów przestrzennych.

W siódmym rozdziale przedstawiono metody porządkowania liniowego, które stosuje się do analiz zbiorów obiektów ze względu na pewne nieobserwowane zjawisko złożone określone zestawem zmiennych. Metody te umożliwiają sporządzenie rankingu obiektów, a także podział zbiorowości na klasy.

Ósmy rozdział poświęcono metodom w pewnym sensie pomocniczym w stosunku do innych metod WAD, czyli metodzie głównych składowych i analizie czynnikowej. Obydwie metody służą transformacji wejściowego zestawu zmiennych opisujących obiekty. W metodzie głównych składowych chodzi o uzyskanie transformowanych zmiennych, które kolejno wyjaśniają, jak największą część zmienności zmiennych wejściowych. Analiza czynnikowa polega zaś na tym, aby uzyskane w wyniku transformacji zmienne nie były ze sobą skorelowane. Uzyskane w wyniku transformacji zestawy zmiennych lub ich podzbiory mogą być następnie podstawą prowadzenia analiz metodami WAD.

W dziewiątym rozdziale omówiono zagadnienia związane z analizą struktur badanych zbiorowości. Poza wskaźnikami struktury omówiono miary ich podobieństwa, zróżnicowania oraz natężenia ich zmian. Przedstawiono także miary koncentracji rozumianej jako nierównomierny podział zjawiska w zbiorowości.

W czasach najnowszych, ze względu na dynamiczny rozwój technik komputerowych, badacze dysponują coraz szerszym dostępem do informacji pochodzących z rzeczywistych dziedzin działalności człowieka. Odnosi się to również do gospodarki, w tym także rolnictwa. Stwarza to nowe możliwości prowadzenia badań naukowych z zakresu między innymi ekonomii, zarządzania oraz szeroko rozumianych nauk społecznych. Ponadto proponowane są coraz bardziej zaawansowane i wyrafinowane metody pozwalające na analizę dostępnych danych. Umożliwiają one opracowywanie charakterystyk czy modeli, często bardzo złożonych zbiorowości, pozwalających na coraz głębsze zrozumienie zasad ich funkcjonowania. Przedstawione w niniejszej pracy metody wpisują się w to zapotrzebowanie. Autorzy mają nadzieję, że będzie ona stanowić ułatwienie i pomoc w zakresie wyboru właściwej metody wielowymiarowej analizy danych, czyli takiej, która pozwoli rozwiązywać problemy stojące przed badaczami.

1. Regresja wieloraka

Analiza regresji znajduje szerokie zastosowanie w naukach ekonomiczno-społecznych. Badanie zależności przyczynowo-skutkowych między obserwowanymi zjawiskami jest jednym z głównych celów prowadzenia badań empirycznych w wielu dziedzinach, między innymi w ekonomice rolnictwa. Znajduje to swoje odzwierciedlenie w bogatym piśmiennictwie. Jako przykłady można wymienić następujące prace: Grabiński, Wydymus i Zeliaś [1982], Nowak [1984, 1990], Sen i Srivastava [1990], Zieliński [1998], Dziechciarz (red.) [2002], Welfe [2003], Gruszczyński i Podgórska (red.) [2004], Walesiak i Gatnar (red.) [2004, 2009], Welfe i Welfe [2004], Maddala [2006], Osińska (red.) [2007], Kukuła (red.) i in. [2009], Kisielińska [2012], Borkowski, Dudek i Szczesny [2017].

Podstawowym zadaniem regresji wielorakiej jest ilościowe ujęcie związków między zmienną zależną (objaśnianą, endogeniczną, kryterialną czy regresantem) a wieloma zmiennymi niezależnymi (objaśniającymi, egzogenicznymi, predyktorami czy regresorami).

1.1. Opis metody – modele regresyjne

Modele regresji wielorakiej służą do przewidywania wartości zmiennej zależnej na podstawie wartości jednej lub wielu zmiennych niezależnych. Modele te zapisywane są w postaci funkcji matematycznej f , określającej analityczny sposób przyporządkowania konkretnym wartościom zmiennych niezależnych wartości zmiennej zależnej z uwzględnieniem błędu losowego.

Zależność stochastyczną można zapisać jako:

$$y = f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon, \quad (1.1)$$

gdzie y to zmienna zależna, $\mathbf{x}^T = [x_1, x_2, \dots, x_k]$ to wektor zmiennych niezależnych, $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_m]$ to wektor parametrów strukturalnych modelu, a ε to składnik losowy.

Jeśli w modelu występuje jedna zmienna zależna (jak w formule (1.1)), model nazywamy jednorównaniowym i do takich modeli ograniczone zostaną rozważania. Jeśli funkcja f jest funkcją liniową, model nazywamy modelem liniowym. Jeśli f nie jest funkcją liniową, model nazywamy nieliniowym.

Wyznaczenie postaci związku między zmiennymi niezależnymi \mathbf{x} i zmienną zależną y polega na doborze oszacowań parametrów strukturalnych modelu, tak aby jak najlepiej odwzorować tę zależność. Model oszacowany zapisuje się jako:

$$\hat{y} = f(\mathbf{x}, \mathbf{b}), \quad (1.2)$$

gdzie \hat{y} to oszacowanie zmiennej zależnej, a \mathbf{b} to wektor oszacowań parametrów strukturalnych modelu.

Model szacowany jest na podstawie n elementowej próby losowej, która obejmuje zestaw k wartości zmiennych niezależnych i odpowiadającą im wartość zmiennej zależnej. Dla i -tej obserwacji model zależności stochastycznej i model oszacowany są następujące:

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad \hat{y}_i = f(\mathbf{x}_i, \mathbf{b}). \quad (1.3)$$

Wektor oszacowań parametrów strukturalnych modelu dobiera się tak, aby rozbieżności między zaobserwowanymi wartościami zmiennej zależnej y_i a ich oszacowaniami \hat{y}_i były jak najmniejsze. Różnicę między tymi wielkościami nazywamy błędem: $e_i = y_i - \hat{y}_i$, który jest oszacowaniem składnika losowego ε_i dla i -tej obserwacji. Jeśli jako łączną miarę wszystkich błędów przyjmiemy sumę kwadratów błędu (ang. *sum square of error* – SSE), to metodę wyznaczania oszacowań parametrów nazywamy metodą najmniejszych kwadratów (MNK). Problem określenia modelu (1.2) sprowadza się do rozwiązania następującego zadania:

$$\min_{\mathbf{b}} (\text{SSE} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2). \quad (1.4)$$

Zadanie takie można w pewnych przypadkach rozwiązać analitycznie z warunku koniecznego istnienia ekstremum funkcji wielu zmiennych (SSE jest funkcją m parametrów). W ekstremum wszystkie pochodne cząstkowe są równe 0. Daje to układ m równań z m niewiadomymi. Jeśli uzyskany układ równań jest liniowy, jego rozwiązanie jest proste. Jeśli natomiast jest nieliniowy, rozwiązanie jego jest możliwe jedynie w pewnych przypadkach. Jeśli nie można znaleźć rozwiązania analitycznego problemu (1.4), można rozwiązać go numerycznie, wykorzystując w tym celu algorytmy optymalizacji nieliniowej.

Warunek konieczny istnienia ekstremum funkcji wielu zmiennych prowadzi do liniowego układu równań, jeśli funkcja f w modelach (1.3) jest funkcją liniową. Oszacowany model regresji liniowej z k zmiennymi niezależnymi przyjmuje wówczas postać:

$$\hat{y}_i = b_0 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + \dots + b_k \cdot x_{ki}. \quad (1.5)$$

Można udowodnić, że wektor $\mathbf{b}^T = [b_0, b_1, \dots, b_k]$, stanowiący rozwiązanie zadania (1.4) w przypadku liniowego modelu regresji wielorakiej (1.5), określa następująca formuła macierzowa:

$$\mathbf{b} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}, \quad (1.6)$$

gdzie \mathbf{y} to wektor n obserwacji zmiennej zależnej $\mathbf{y}^T = [y_1, y_2, \dots, y_n]$, a \mathbf{X} to następująca macierz zawierająca obserwacje k zmiennych niezależnych:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}, \quad (1.7)$$

w której pierwsza kolumna zawierająca same jedynki odpowiada wyrazowi wolnemu modelu.

Aby estymatory parametrów strukturalnych modelu miały pewne wymagane własności statystyczne [Borkowski, Dudek, Szczesny 2017], powinny być spełnione następujące założenia:

- 1) zmienne niezależne są nielosowe,
- 2) wartości oczekiwane składników losowych ε_i są równe 0,
- 3) wariancje składników losowych ε_i są stałe (homoskedastyczność składnika losowego),
- 4) składniki losowe ε_i i ε_j są od siebie niezależne dla $i \neq j$, gdzie $i, j = 1, 2, \dots, n$ (brak autokorelacji składnika losowego),
- 5) każdy ze składników losowych ε_i ma rozkład normalny,
- 6) wielkość próby powinna znacznie przekraczać liczbę szacowanych parametrów,
- 7) między wektorami wartości zmiennych niezależnych nie istnieje zależność liniowa.

Założenia te są potrzebne do wyznaczenia estymatorów parametrów strukturalnych oraz, aby do weryfikacji modelu i jego elementów można było zastosować odpowiednie testy statystyczne. W podanej literaturze można znaleźć informacje o testach umożliwiających weryfikację tych założeń oraz o postępowaniu w przypadku ich niespełnienia [Gruszczyński, Podgórska (red.) 2004; Maddala 2006; Borkowski, Dudek, Szczesny 2017].

Weryfikacja statystyczna liniowego modelu regresyjnego obejmuje weryfikację hipotezy o istnieniu lub braku zależności między zmienną zależną i przynajmniej jedną zmienną niezależną oraz hipotez dotyczących istotności poszczególnych jego parametrów.

Formułuje się następującą parę hipotez statystycznych:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0, \quad (1.8)$$

$$H_1: \exists j, \beta_j \neq 0.$$

Konsekwencją prawdziwości H_0 jest brak zależności liniowej między zmienną zależną y i zmiennymi niezależnymi \mathbf{x} , co oznacza, że model należy odrzucić. Konsekwencją prawdziwości H_1 jest istnienie zależności liniowej między zmienną zależną i przynajmniej jedną niezależną.

Sprawdzianem testu jest następująca statystyka:

$$F_{k, n-k-1} = \frac{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{k}}{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1}}, \quad (1.9)$$

gdzie jest $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Jeśli spełnione są założenia MNK, statystyka (1.9) ma rozkład F z liczbą stopni swobody licznika k i mianownika $n - k - 1$. Omawiany test ma prawostronny obszar krytyczny.

Jeśli odrzucona zostanie hipoteza H_0 (1.8), należy zweryfikować hipotezy dotyczące istotności wszystkich parametrów strukturalnych modelu. Dla każdego $j = 0, 1, \dots, k$ weryfikowane są następujące pary hipotez statystycznych:

$$H_0: \beta_j = 0, \quad (1.10)$$

$$H_1: \beta_j \neq 0.$$

Konsekwencją prawdziwości H_0 jest konieczność usunięcia wyrazu wolnego z modelu, jeśli $j = 0$ lub zmiennej j -tej, jeśli $j \neq 0$.

Sprawdzianami testów są następujące statystyki:

$$T_{n-k-1} = \frac{b_j}{s(b_j)}. \quad (1.11)$$

Jeśli spełnione są założenia MNK dla każdego j , statystyki (1.11) mają rozkład T z liczbą stopni swobody $n - k - 1$. Test ten ma dwustronny obszar krytyczny.

Jeśli dla pewnego j nie uda się odrzucić H_0 , konieczne jest ponowne szacowanie modelu bez elementu j . Może się zdarzyć, że w modelu należy usunąć więcej niż jeden element i wówczas wskazane jest zastosowanie procedury tzw. regresji krokowej dostępnej w pakietach statystycznych. Regresję krokową prowadzi się w przód lub wstecz. W wariancie w przód do modelu kolejno wprowadzane są zmienne niezależne i pozostawiane te, których współczynniki kierunkowe są istotnie różne od zera. W regresji krokowej wstecz startuje się od wszystkich potencjalnych zmiennych niezależnych i kolejno eliminuje nieistotne.

Prowadząc weryfikację modelu regresyjnego, należy pamiętać o jego merytorycznej poprawności. Chodzi o to, aby jego wskazania zgadzały się z tym, czego należy w danych warunkach oczekiwać.

Ważnym zagadnieniem w analizie regresji jest ocena jakości modelu regresyjnego. Jako jej najbardziej popularną miarę stosuje się współczynnik determinacji $R^2 \in (0, 1]$, który interpretowany jest jako udział zmienności zmiennej zależnej wyjaśnionej przez model. Współczynnik R^2 oblicza się jako:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (1.12)$$

Im większa wartość R^2 , tym lepszy jest model (jeśli R^2 jest równy 1 między zmiennymi jest zależność deterministyczna). Miara ta wykazuje jednak pewne wady w przypadku modeli z różną liczbą zmiennych niezależnych. Jej wartość rośnie zawsze wraz z włączaniem do modelu kolejnych zmiennych niezależnych [Stanisz 2007], nawet jeśli nie mają one żadnego związku ze zmienną zależną.

W takiej sytuacji do oceny jakości modelu należy stosować skorygowany współczynnik determinacji, który uwzględnia jego stopień złożoności [Kisielińska 2012] i określony jest jako:

$$\bar{R}^2 = 1 - \frac{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1}}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}. \quad (1.13)$$

Teoria związana z szacowaniem i weryfikowaniem liniowych modeli regresyjnych jest dobrze opracowana, nie może być jednak zastosowana dla wszystkich modeli nieliniowych.

Nieliniowe modele regresyjne dzielone są na dwie grupy: modele linearyzowalne oraz ściśle nieliniowe. Modele linearyzowane sprowadza się do modeli liniowych poprzez odpowiednie transformacje zmiennych i ewentualnie parametrów [Kisielińska 2012]. Modele ściśle nieliniowych natomiast zlinearyzować się nie da. Parametrów tych modeli poszukuje się metodami optymalizacji nieliniowej [Kisielińska 2012]. Najczęściej stosowane linearyzowalne modele nieliniowe to: wielomianowe, hiperboliczne, potęgowe, wykładnicze i logarytmiczne.

Warto jeszcze wspomnieć, że poza podziałem na modele liniowe i nieliniowe są jeszcze inne kryteria klasyfikacji modeli regresyjnych [Gruszczński, Podgórska (red.) 2004]. Jako najważniejsze dla modeli jednorównaniowych można uznać ogólnopoznawcze oraz czasowe kryterium.

Według kryterium ogólnopoznawczego modele regresyjne można podzielić na:

- **modele przyczynowo-opisowe**, które wyrażają związki przyczynowo-skutkowe między zmienną zależną i zmiennymi niezależnymi,
- **modele symptomatyczne**, w których rolę zmiennych niezależnych odgrywają zmienne skorelowane ze zmienną zależną, ale nie wyrażają one źródeł zmienności tej zmiennej.

Z punktu widzenia roli czynnika czasu można wyróżnić:

- **modele statyczne**, w których związki występują w tej samej jednostce czasu lub nie ma znaczenia, z jakiego okresu pochodzą dane,
- **modele dynamiczne** uwzględniające czynnik czasu w formie opóźnień lub zmiennej czasowej.

1.2. Zastosowania regresji wielorakiej

Modele regresji wielorakiej znajdują w naukach ekonomicznych szerokie spektrum zastosowań. Można je odnaleźć w pracach poświęconych analizom ilościowym zmiennych reprezentujących na przykład czynniki produkcji, wskaźniki finansowe, czynniki dochodowości i wydajności lub inne parametry gospodarcze. Przykłady zastosowań omówionych metod przedstawiono w tabeli 1.1.

W większości badań skonstruowane modele wykorzystywano do zdefiniowania czynników kształtujących opisywane zjawisko oraz do określenia siły i kierunku ich wpływu. Zjawiska ekonomiczne są procesami złożonymi, które kształtuje mnogość czynników. W pracach zaprezentowanych w tabeli 1.1 autorzy niejednokrotnie badali dużą liczbę cech kandydatek, aby w końcowym etapie analiz dokonać redukcji do zestawu zmiennych niezależnych istotnie kształtujących poziom zmiennej kryterialnej. Eliminację cech słabo determinujących poziom zmiennej zależnej przeprowadzano w większości przypadków metodą krokową.

Analiza rodzajów zmiennych zależnych w prezentowanych modelach pozwala na potwierdzenie szerokiej różnorodności badanych zjawisk z zakresu ekonomiki rolnictwa. Najwięcej modeli dotyczyło problematyki związanej z efektywnością działania gospodarstw rolnych i przedsiębiorstw sektora spożywczego oraz z oceną rozwoju społeczno-ekonomicznego gospodarstw lub regionów rolniczych (dwadzieścia dziewięć publikacji w tabeli 1.1). W piętnastu publikacjach badano wielkość produkcji, a dziesięć dotyczyło rozważań na temat popytu, poziomu cen i sprzedaży produkcji rolnej. Dochód rolniczy wraz z wartością dodaną w rolnictwie wyjaśniano w dziewięciu artykułach, koszty i nakłady w ośmiu.

Modele konstruowano przede wszystkim dla gospodarstw rolniczych (31), a następnie dla przedsiębiorstw spożywczych (12), makroregionów, województw i gmin (11) oraz spółdzielni (2) i gospodarstw agroturystycznych (2).

Funkcje regresji prostej wykorzystano w ośmiu publikacjach, w których jedna zmienna dała podstawy do określenia wartości zmiennej kryterialnej, a modele prawidłowo zweryfikowano pod względem merytorycznym i statystycznym. Znalazły one zastosowanie w ocenie zjawiska konwergencji w województwach w Polsce [Niewiadomski 2009] i koncentracji monopolistycznej w przemyśle spożywczym [Kufel, Hamulczuk 2015], a także w badaniu atrakcyjności wiejskiej przestrzeni rekreacyjnej [Ciepiela, Sosnowski 2011]. Ponadto analizowano zależności między cenami [Bodnar, Galchynska, Maciejczak 2020], kosztami ubezpieczeń [Zielińska 2018], zadłużeniem gmin wiejskich [Wichowska 2019] oraz wskaźnikami ekonomiczno-finansowymi wyznaczanymi dla przedsiębiorstw spożywczych [Wasilewska 2011] bądź województw [Rzeszutko 2015].

Modele liniowej regresji wielorakiej zastosowano w siedemnastu artykułach poświęconych problematyce oceny efektywności działań podmiotów rolniczych oraz ewaluacji rozwoju społeczno-gospodarczego regionów. Podstawą konstruowanych modeli były liczby zmienne wahające się od dwóch do dwunastu predyktorów. Dla gospodarstw rolnych identyfikowano między innymi czynniki efektywności operacyjnej [Kulawik, Płonka 2014] i determinujące sytuację finansową [Średzińska

2019], zmiany surowcowe [Pawlewicz, Szamrowski 2014] i ograniczenia produkcji [Tanimonure, Yewande, Fayemi 2020], aktualne problemy bezpieczeństwa żywnościowego [Salau 2020] oraz zadłużenia [Kata 2020]. Przedsiębiorstwa przemysłu spożywczego analizowano pod kątem rentowności [Bieniasz, Gołaś 2012], stopy zwrotu z kapitału własnego [Adamenko, Oliynyk, Wasilewski 2015], udziału kapitału obrotowego w aktywach [Hodun, Kuś 2014], determinant efektywności ekonomicznej [Wasilewski, Żurakowska-Sawa 2017] oraz płynności finansowej [Czerwińska-Kayzer, Florek 2018]. W badaniu regionów zajmowano się identyfikacją i oceną wpływu czynników charakteryzujących poziom rozwoju społeczno-gospodarczego na szczeblu gmin [Paluch, Sroka 2013] oraz województw, gdzie wykorzystano model regresji liniowej ważonej geograficznie [Chrzanowska, Drejerska 2016]. Ponadto dla państw Unii Europejskiej (UE) przeprowadzono analizy związku między stanem rozwoju gospodarczego obszarów wiejskich a poziomem długu publicznego [Siudek 2014] oraz wpływu indeksów efektów i nakładów na lukę cenową w rolnictwie [Czyżewski, Matuszczak 2016]. Oceniano również wpływ nakładów na działalność badawczo-rozwojową w dziedzinie biotechnologii [Matras-Bolibok 2015] oraz identyfikowano czynniki warunkujące zmiany w handlu targowiskowym [Hamulczuk 2016].

Liniowe funkcje regresji wielokrotnej stosowano również w zagadnieniach dotyczących dochodu i wartości dodanej w rolnictwie oraz w analizach ponoszonych kosztów i nakładów w obszarach działalności rolnej. Poruszano problematykę porównania dochodów, składników majątkowych i źródeł jego finansowania [Ryś-Jurek 2009; Sass 2014] oraz wpływu zmiennych generujących dochód i obrazujących sukces ekonomiczny [Sroka 2012]. Identyfikowano czynniki kształtujące zmienność dochodów rolniczych w Grecji [Papaelias, Panagou, Hyz 2014] oraz w polskich gospodarstwach rolnych [Wasąg 2015; Wysokiński 2016]. Sprawdzano, w jakim stopniu ubezpieczenia upraw w Polsce [Kobus 2015] oraz rozwój bankowego systemu finansowego na Ukrainie [Wasilewski, Oliynyk, Adamenko 2016] stabilizują dochód i determinują wartość dodaną gospodarstw rolnych. Ponadto rozpoznawano czynniki kształtujące koszty produkcji jaj [Sytchevnik, Golovkov 2010], badano skalę nakładów i rodzaje inwestycji w innowacje w przedsiębiorstwach sektora spożywczego [Domańska, Kijek 2015] oraz w gospodarstwach rolnych w Polsce [Kusz, Gędek 2015].

Rozmiary produkcji rolnej, jej wydajność oraz sprzedaż, a także determinanty cen produktów rolnych również modelowano liniową regresją wieloraką. Analizowano zmiany i rozwój spożycia mleka wraz z jego przetworami [Kubicová, Kádeková 2013; Kubicová, Dobák, Kádeková 2014; Bórawski, Grzybowska-Brzezińska, Dunn 2020], badano wartość produkcji i sprzedaży artykułów

mleczarskich [Bórawski 2015; Juszczak 2016; Bórawski, Zalewski 2018] oraz oceniano możliwości oddziaływania na poziom cen wyrobów sektora mlecznego [Juszczak 2011; Juszczak, Balina 2015].

Modele potęgowe i wykładnicze użyto w dwudziestu publikacjach. W artykułach z wykorzystaniem funkcji Cobba-Douglasa skupiono się na analizie produktywności, dochodowości i opłacalności substytucji czynników produkcji rolnej. Badano rentowność determinant produkcji mleczarskiej [Adamski 2009; Kubik 2014], rozważano wydajność czynników produkcji w przemyśle spożywczym [Łukiewska 2014a, 2014b] oraz w gospodarstwach towarowych [Nowak, Wójcik, Krukowski 2015], nastawionych na produkcję zwierzęcą [Kubik, Wójcik 2015] i roślinną [Zieliński 2014; Niezgoda 2015; Omotesho, Douvi, Olaghere 2020]. Na poziomie makroregionów i państw oceniano elastyczność związków między czynnikami produkcji [Nowak 2015] oraz analizowano szacunki wskaźnika całkowitej produktywności rolnictwa (ang. *total factor productivity* – TFP) [Grochowska, Mańko 2014]. Ponadto modele potęgowe zastosowano do opisu zależności makroekonomicznych między dochodami rolniczymi a czynnikami ich wzrostu [Wiśniewska, Ryś-Jurek 2010]. Funkcji wykładniczych użyto w badaniu związków rozwoju sektora finansowego i rolniczego na Ukrainie [Wasilewski, Oliynyk, Adamenko 2016], w rozpoznaniu znaczenia determinant kształtujących procesy reprodukcji majątku [Grzelak 2016] oraz w analizie wpływu zmian światowych cen cukru na ceny krajowe [Szajner 2019].

Modele regresji wielomianowej wykorzystano w czterech artykułach omawiających problematykę maksymalizacji zysku ze sprzedaży sałaty masłowej [Kozioł, Zieliński 2009], sprawdzających czynniki wpływające na wskaźnik reprodukcji środków trwałych [Sobczyński 2009] oraz analizujących rentowność przedsiębiorstw rolniczych na Ukrainie [Adamenko, Oliynyk, Wasilewski 2015].

Modele hiperboliczne i ich pewne modyfikacje zastosowano w analizie makroekonomicznej stopy nadwyżki ekonomicznej [Czyżewski, Majchrzak 2015] oraz w badaniu wsparcia finansowego wspólnej polityki rolnej na dochody sektora rolnego w Polsce [Buks, Pietrzykowski 2015].

Podsumowując, należy zauważyć, że modele regresji wielorakiej stają się docenianym przez ekonomistów narzędziem analizy, a szeroka dostępność przyjaznego dla użytkownika oprogramowania pozwala na konstruowanie coraz bardziej złożonych modeli.

Tabela 1.1. Przykłady zastosowań analizy regresji w ekonomice rolnictwa

Autor	Problem	Obiekty	Zmienna zależna	Zmienne niezależne	Metoda
Dec, Pawlak, Poczta 2008	wyznaczenie czynników determinujących stopień wyżywienia ludności wraz z ilościowym określeniem siły ich oddziaływania w regionach świata dotkniętych problemem niedożywienia	5 wyodrębnionych regionów świata	stopień niedożywienia ludności	8 zmiennych	modele: regresja liniowa
Adamski 2009	analiza dochodowości materialnych czynników produkcji w polskich gospodarstwach mlecznych	gospodarstwa rolne mleczne	dochód z gospodarstwa rolniczego	3 zmienne	model: funkcja produkcji Cobba-Douglasa
Kozioł, Zieliński 2009	analiza problemu maksymalizacji zysku ze sprzedaży sałaty masłowej oraz oszacowanie optymalnego terminu sadzenia sałaty w przypadku gdy nie są spełnione założenia KMNK	sałata	dzień zbioru sałaty oraz zysk dla danej masy m główki sałaty	2 zmienne	modele: regresja wielomianowa
Niewiadomski 2009	ocena zjawiska konwergencji między województwami w Polsce, analiza wpływu dynamiki wartości dodanej brutto na relatywny indeks wzrostu w odniesieniu do kategorii osób pracujących	województwa	relatywny indeks wzrostu	wartość dodana brutto	model: regresja prosta
Ryś-Jurek 2009	porównanie produkcji, dochodów, składników majątku i źródeł jego finansowania w indywidualnych gospodarstwach rolnych w Polsce oraz w innych krajach UE oraz badanie relacji między dochodem, produkcją a zapasami w gospodarstwie	gospodarstwa rolne	dochód z gospodarstwa rolnego	2, 3 zmienne	modele: regresja liniowa
Sobczyński 2009	analiza czynników wpływających na wskaźnik reprodukcji środków trwałych w gospodarstwach rolnych z 3 największych klas wielkości ekonomicznej	gospodarstwa rolne	wskaźnik odnowienia środków trwałych	2 zmienne	modele: regresja wielomianowa
Niewiadomski 2010	ocena czynników determinujących rozwój gospodarstw agroturystycznych	gospodarstwa agroturystyczne	liczba gospodarstw agroturystycznych	od 4 do 10 zmiennych	modele: regresja liniowa
Sytchevnik, Golovkov 2010	analiza wyników produkcji i sprzedaży jaj w fermach drobiarskich oraz identyfikacja czynników kształtujących koszty produkcji jaj kurzych na Białorusi	jaja kurze	koszt produkcji jaj	5 zmiennych	model: regresja liniowa
Wiśniewska, Ryś-Jurek 2010	opis najważniejszych zależności makroekonomicznych między dochodami rolniczymi a czynnikami ich wzrostu w sektorze gospodarstw domowych	gospodarstwa rolne	dochody do dyspozycji brutto	4 zmienne	model: funkcja potęgowa
Ciepiela, Sosnowski 2011	badanie zależności między atrakcyjnością wiejskiej przestrzeni rekreacyjnej a wielkością agroturystycznej bazy noclegowej i jej wykorzystaniem w regionie siedleckim	gospodarstwa agroturystyczne	liczba gospodarstw agroturystycznych; liczba miejsc noclegowych; stopa wykorzystania obiektów noclegowych	wartości wskaźnika wiejskiej przestrzeni rekreacyjnej	modele: regresja prosta

Tabela 1.1 (cd.)

Autor	Problem	Obiekty	Zmienna zależna	Zmienne niezależne	Metoda
Juszczak 2011	ocena możliwości oddziaływania na poziom cen mleka surowego poprzez ceny kluczowych artykułów pochodnych będących powszechnym przedmiotem handlu międzynarodowego	mleko	cena mleka surowego	3 zmienne	model: regresja liniowa
Wasilewska 2011	opracowanie modelu szybkiego testu płynności finansowej (wskaźnika szybkiego) dla grupy przedsiębiorstw rolniczych	przedsiębiorstwa rolnicze	wartość szybkiego wskaźnika płynności	różnica cykli należności i zobowiązań	model: regresja prosta
Bieniasz, Gołaś 2012	weryfikacja związków przyczynowo-skutkowych między wskaźnikami gospodarowania zapasami a wynikami finansowymi przedsiębiorstw przemysłu spożywczego	przedsiębiorstwa przemysłu spożywczego	rentowność	od 5 do 7 zmiennych	modele: regresja liniowa
Sroka 2012	analiza wpływu poszczególnych zmiennych na kategorie obrazujące sukces ekonomiczny gospodarstw rolnych	gospodarstwa rolne	dochód z gospodarstwa rolniczego; wartość majątku gospodarstwa	od 4 do 6 zmiennych	modele: regresja liniowa
Kubicová, Kádekóvá 2013	modele popytu na przetwory mleczne opracowane na podstawie ceny detalicznej produktu mlecznego i dochodu dyspozycyjnego w gospodarstwach domowych pracowników, osób pracujących na własny rachunek, a także rencistów	przetwory mleczne	popyt	5 zmiennych	modele: regresja liniowa popytu
Paluch, Sroka 2013	identyfikacja oraz ocena wpływu czynników charakteryzujących poziom rozwoju społeczno-gospodarczego na zrównoważenie środowiskowe gmin wiejskich województwa małopolskiego	gminy wiejskie	syntetyczny wskaźnik zrównoważonego rozwoju	6 zmiennych	model: regresja liniowa
Grochowska, Mańko 2014	przegląd badań pokazujących szacunki wskaźnika całkowitej produktywności rolnictwa (<i>total factor productivity</i> – TFP) w skali globalnej i w wybranych krajach	gospodarstwa rolne	wartość produkcji ogółem	3 zmienne	modele: funkcja produkcji Cobba-Douglasa
Hodun, Kuś 2014	analiza wpływu zmian badanych wskaźników finansowych na wskaźnik udziału kapitału obrotowego w aktywach ogółem m.in. w przedsiębiorstwach sektora spożywczego	przedsiębiorstwa sektora spożywczego	wskaźnik udziału kapitału obrotowego w aktywach ogółem	10 zmiennych	model: regresja liniowa
Kubicová, Dobák, Kádekóvá 2014	ocena zmian i rozwoju spożycia mleka i przetworów mlecznych w ostatnich latach od przystąpienia Słowacji do UE	mleko i przetwory mleczne	wielkość sprzedaży mleka i jego przetworów	2 zmienne	modele: regresja liniowa popytu
Kubik 2014	ocena substytucji czynników wytwórczych i jej wpływu na kształtowanie się kosztów całkowitych produkcji mleka w gospodarstwach towarowych	gospodarstwa rolne towarowe	wartość produkcji ogółem	3 zmienne	model: funkcja produkcji Cobba-Douglasa

Kulawik, Płonka 2014	zidentyfikowanie czynników, które wpływają na efektywność operacyjną gospodarstw rodzinnych	gospodarstwa rolne rodzinne	miernik efektywności ekonomiczno-finansowej	od 4 do 12 zmiennych	modele: regresja liniowa
Łukiewska 2014a	określenie poziomu oraz tempa zmian produktywności cząstkowej i całkowitej, a także próba identyfikacji czynników wzrostu produkcji w przemyśle spożywczym i jego branżach w Polsce	przedsiębiorstwa przemysłu spożywczego	wartość produkcji sprzedanej	3 zmienne	model: funkcja produkcji Cobba-Douglasa
Łukiewska 2014b	ocena poziomu wybranych czynników i ich wpływu na produktywność przemysłu spożywczego na tle przetwórstwa	przedsiębiorstwa sektora spożywczego	produkcja; wskaźniki produktywności dla pracy i kapitału	4 zmienne	modele: funkcja produkcji Cobba-Douglasa i regresja liniowa
Papaelias, Panagou, Hyz 2014	analiza zmian efektywności produkcji rolnej Grecji w latach 1911-2011 oraz oszacowanie trendów, które pojawiły się po wejściu Grecji do UE	gospodarstwa rolne w Grecji	dochody rolnicze	6 zmiennych	modele: regresja liniowa i logarytmiczno-liniowa
Pawlewicz, Szamrowski 2014	prezentacja stanu rolnictwa ekologicznego oraz możliwych zmian, jakie nastąpią na rynku ekologicznych surowców żywnościowych w nowej perspektywie finansowania	gospodarstwa rolne ekologiczne	powierzchnia ekologicznych użytków rolnych	2 zmienne	model: regresja liniowa
Sass 2014	badanie zależności między poniesionymi nakładami inwestycyjnymi w okresie przedakcesyjnym i po Polski do UE a zmianami w zasobach ziemi i majątku produkcyjnym gospodarstw rolnych	gospodarstwa rolne	wartość majątku ogółem	2 zmienne	model: regresja liniowa
Siudek 2014	określenie związku między poziomem rozwoju gospodarczego obszarów wiejskich a poziomem długu publicznego w krajach UE	kraje UE	poziom rządowego długu publicznego w Polsce; poziom rządowego długu publicznego w Grecji	4 zmienne	modele: regresja liniowa
Zieliński 2014	analiza porównawcza potencjału produkcyjnego, organizacji i struktury produkcji, efektywności ekonomicznej oraz nasilenia inwestycyjnego w określeniu przyczyn dysproporcji w efektywności produkcji rolniczej gospodarstw specjalizujących się w uprawach polowych	gospodarstwa rolne	wartość produkcji ogółem powiększona o dopłaty i subwencje budżetowe	4 zmienne	model: funkcja produkcji Cobba-Douglasa
Adamenko, Oliynyk, Wasilewski 2015	prezentacja modeli finansowania bazujących na zadłużeniu i kapitale własnym w podejściu operacyjnym dla dużych i średnich przedsiębiorstw rolniczych prowadzących działalność na Ukrainie	przedsiębiorstwa rolnicze na Ukrainie	ROE i aktywa ogółem	2 zmienne	modele: regresja liniowa i regresja wielomianowa
Bórawski 2015	rozpoznanie oraz ocena rozmiarów handlu zagranicznego mlekiem i produktami mleczarskimi, a także identyfikacja czynników mających wpływ na saldo bilansu handlowego artykułami mleczarskimi w Polsce	przetwory mleczne	saldo handlu zagranicznego artykułów mleczarskich	5 zmiennych	modele: regresja liniowa

Tabela 1.1 (cd.)

Autor	Problem	Obiekty	Zmienna zależna	Zmienne niezależne	Metoda
Buks, Pietrzykowski 2015	określenie oddziaływania wsparcia finansowego wspólnej polityki rolnej na dochody w polskim sektorze rolnym	przedsiębiorstwa rolnicze	udział procentowy dopłat w dochodzie przedsiębiorcy rolnego	dochód przedsiębiorcy rolnego	model: regresja hiperboliczno-wykładnicza
Czyżewski, Majchrzak 2015	analiza makroekonomicznej funkcji dochodów (nadwyżki) w rolnictwie	sektor rolniczy	stopa nadwyżki ekonomicznej przy braku subsydiów	4 zmienne	model: regresja hiperboliczna
Domańska, Kijek 2015	identyfikacja wraz z badaniem skali i rodzaju inwestycji w innowacje w próbie producentów żywności z województwa lubelskiego	przedsiębiorstwa sektora spożywczego	wysokość nakładów na badania i rozwój	4 zmienne	model: regresja liniowa
Juszczak, Balina 2015	analiza kluczowych czynników kształtujących poziom cen mleka po likwidacji systemu kwotowego oraz wskazanie pożądanych kierunków inwestycyjnych związanych z eksportem artykułów mleczarskich	ceny mleka	cena surowego mleka	3 zmienne	modele: regresja liniowa
Kobus 2015	sprawdzenie w jakim stopniu ubezpieczenia upraw stabilizują dochód gospodarstw rolniczych wraz z wykazaniem znaczenia doboru metod statystycznych	gospodarstwa rolne	wartość dodana brutto (WDB)	6 zmiennych	modele: regresja liniowa
Kubik, Wójcik 2015	ocena substytucji nakładów pracy ludzkiej kapitałem w towarowych gospodarstwach rolnych typów „krowy mleczne” i „zwierzęta ziarnożerne” ukierunkowanych na produkcję zwierzęcą w Polsce	gospodarstwa rolne towarowe	produkcja rolnicza w gospodarstwach	2 zmienne	modele: funkcja produkcji Cobba-Douglasa
Kufel, Hamulczuk 2015	ocena zmian koncentracji na tle zmian w wywieranej sile rynkowej (zdolność firmy do podnoszenia ceny powyżej kosztu krańcowego) w polskim przemyśle spożywczym	przedsiębiorstwa przemysłu spożywczego	indeks Lerner (marża monopolistyczna)	udział dużych przedsiębiorstw w wartości produkcji sprzedanej przemysłu spożywczego	modele: regresja prosta
Kusz, Gędek 2015	analiza wpływu czynników egzogenicznych i endogenicznych na poziom nakładów inwestycyjnych w rolnictwie w Polsce	gospodarstwa rolne	poziom nakładów inwestycyjnych w rolnictwie	11 zmiennych	model: regresja liniowa
Matras-Bolibok 2015	ocena wpływu nakładów na działalność B+R w dziedzinie biotechnologii na innowacyjność biogospodarki	nakłady na działalność B+R	liczba patentów biotechnologicznych udzielonych w roku t	2 zmienne	model: regresja liniowa

Nieżgoda 2015	ocena ekonomicznej opłacalności substytucji czynników zmiennych, czyli nawozów mineralnych środkami ochrony roślin w procesie produkcji towarowych gospodarstw rolnych w Polsce	gospodarstwa rolne towarowe	produkcja roślinna	3 zmienne	model: funkcja produkcji Cobba-Douglasa
Nowak 2015	ocena elastyczności związków między czynnikami produkcji (ziemi, pracy i kapitału) a uzyskanymi w wyniku ich zastosowania dochodami w gospodarstwach rolnych czterech makroregionów FADN	makroregiony	produkcja (wartość dodana netto)	3 zmienne	modele: funkcja produkcji Cobba-Douglasa
Nowak, Wójcik, Krukowski 2015	ocena elastyczności związków między czynnikami produkcji w badaniu produktywności towarowych gospodarstwach rolnych	gospodarstwa rolne	produkcja całkowita	3 zmienne	model: funkcja produkcji Cobba-Douglasa
Rzeszutko 2015	wskazanie zmian w zakresie stopnia wykorzystania potencjału produkcyjnego rolnictwa w poszczególnych województwach w Polsce w okesie członkostwa w UE	województwa	syntetyczna miara rzeczywistej produktywności czynników produkcji	wartości syntetycznego miernika poziomu potencjału produkcyjnego	model: regresja prosta
Wasąg 2015	określenie wpływu wybranych czynników na efektywność wykorzystania funduszy w gospodarstwach rodzinnych uczestniczących w programach pomocowych UE	gospodarstwa rolne rodzinne	produkcja towarowa; dochód; nakłady pracy; wskaźniki technicznego uzbrojenia pracy	od 3 do 6 zmiennych	modele: regresja liniowa
Chrzanowska, Drejerska 2016	analiza poziomu rozwoju społeczno-gospodarczego województw w Polsce	województwa	PKB <i>per capita</i> według parytetu siły nabywczej	3 zmienne	model: regresja liniowa ważona geograficznie
Czyżewski, Kryszak 2016	identyfikacja i porównanie wpływu podstawowych kategorii ekonomicznych na rachunek dochodów w rolnictwie poszczególnych krajów UE	kraje UE	całkowita wartość produkcji sektora rolniczego	3 zmienne	modele: funkcja produkcji Cobba-Douglasa
Czyżewski, Matuszczak 2016	oszacowanie długoterminowych modeli regresji luki cenowej w rolnictwie dla wybranych krajów europejskich reprezentujących różne struktury agrarne	kraje UE	luka cenowa	indeksy efektów i nakładów (od 5 do 11)	modele: regresja liniowa
Grzelak 2016	identyfikacja znaczenia determinant kształtujących procesy reprodukcji majątku (majątku trwałego z wyłączeniem ziemi) w gospodarstwach rolnych w Polsce prowadzących rachunkowość rolną	gospodarstwa rolne	wskaźnik reprodukcji	4 zmienne	modele: regresja wykładnicza
Hamulczuk 2016	wskazanie czynników warunkujących zmiany handlu targowiskowego w Polsce	targowiska w Polsce	powierzchnia sprzedażowa targowisk w Polsce	4 i 5 zmiennych	modele: regresja liniowa
Juszczyk 2016	wpływ innowacji produktowych oraz kosztów związanych z ochroną środowiska na produkcję spółdzielni mleczarskich, mierzoną wartością sprzedaży	spółdzielnie mleczarskie	wartość sprzedaży	od 2 do 5 zmiennych	modele: regresja liniowa

Tabela 1.1 (cd.)

Autor	Problem	Obiekty	Zmienna zależna	Zmienne niezależne	Metoda
Mekonnen i in. 2016	analiza sposobu postrzegania ochrony gleby i wody (SWC) przez rolników ze wschodniego Hararghe w Etiopii	rolnicy z Etiopii	miara postrzegania ochrony gleby i wody (SWC)	12 zmiennych	model: regresja liniowa (GLM)
Utnik-Banaś, Żmija 2016	określenie wpływu wybranych czynników na kształtowanie się poziomu cen żywca kurcząt brojlerów	brojlery	cena żywca brojlerów	8 zmiennych	model: regresja liniowa
Wasilewski, Oliynyk, Adamenko 2016	badanie relacji między wartością dodaną na pracującego w rolnictwie a rozwojem bankowego systemu finansowego na Ukrainie w porównaniu do sytuacji w Polsce	sektor rolniczy i sektor bankowy	wartość dodana na pracownika w rolnictwie	2 zmienne	modele: regresja liniowa i wykładnicza
Wysokiński 2016	identyfikacja czynników wpływających na dochodowość z produkcji mleka w rodzinnych gospodarstwach rolnych	gospodarstwa rolne rodzinne mleczne	dochód z rodzinnego gospodarstwa rolnego	9 i 10 zmiennych	modele: regresja liniowa
Kusz, Misiak 2017	ocena zmian wydajności pracy w rolnictwie w Polsce w aspekcie zmian technicznego uzbrojenia pracy oraz postępu technicznego	gospodarstwa rolne	wydajność pracy	2 zmienne	model: funkcja produkcji Cobba-Douglasa
Wasilewski, Żurakowska-Sawa 2017	określenie uwarunkowań z zakresu otoczenia makroekonomicznego, jak i charakterystyk mikroekonomicznych przedsiębiorstw determinujących zmiany efektywności ekonomicznej spółek giełdowych znajdujących się w fazie wprowadzenia	przedsiębiorstwa sektora spożywczego	syntetyczna miara efektywności ekonomicznej	3 zmienne	model: regresja liniowa
Bórawski, Zalewski 2018	określenie wielkości produkcji mleka na tle pozostałych krajów UE oraz rozpoznanie zmian w produkcji mleka wraz z identyfikacją czynników kształtujących produkcję mleka	mleko	produkcja mleka	4 zmienne	model: regresja liniowa
Czerwińska-Kayzer, Florek 2018	wskazanie czynników kształtujących płynność finansową w przedsiębiorstwach produkujących pasze dla zwierząt oraz określenie siły i kierunku ich wpływu	przedsiębiorstwa rolnicze	płynność finansowa	6 wskaźników finansowych	modele: regresja liniowa
Zielińska 2018	ocena związku między poziomem odprowadzanych składek na ubezpieczenie wypadkowe a poziomem kosztów na działalność profilaktyczną w zakresie bezpieczeństwa i higieny pracy w przedsiębiorstwach spożywczych w województwie łódzkim	przedsiębiorstwa sektora spożywczego	koszty działalności profilaktycznej ogółem	wysokość składki na ubezpieczenie wypadkowe ogółem	modele: regresja prosta
Szajner 2019	ocena wpływu zmian światowych cen cukru na ceny krajowe	cukier	krajowe ceny zbytu i eksportowe cukru	światowe ceny cukru białego	model: regresja wykładnicza

Średzińska 2019	określenie czynników determinujących sytuację finansową gospodarstw rolnych o różnej wielkości ekonomicznej w krajach Europy Środkowo-Wschodniej i UE-15	gospodarstwa rolne	syntetyczny wskaźnik sytuacji finansowej gospodarstw	od 5 do 8 zmiennych	modele: regresja liniowa
Wichowska 2019	ocena poziomu zadłużenia gmin wiejskich województwa warmińsko-mazurskiego na tle innych gmin oraz określenie czynników determinujących ich dług	gminy wiejskie	całkowite, długoterminowe i krótkoterminowe zadłużenie <i>per capita</i>	3 zmienne	modele: regresja prosta
Bodnar, Galchynska, Maciejczak 2020	badanie zależności cenowych towarów rolnych od innych towarów, których ceny kształtują się na rynkach światowych, ze szczególnym uwzględnieniem roli ich zmienności	towary rolne	ceny na rynku wewnętrznym dla odpowiedniego rodzaju upraw	cena na giełdach światowych odpowiedniego rodzaju uprawy	modele: regresja prosta
Bórawski, Grzybowska-Brzezińska, Dunn 2020	przedstawienie poziomu i zmian spożycia mleka oraz jego przetworów w UE	mleko i przetwory mleczne	spożycie mleka na mieszkańca	4 zmienne	model: regresja liniowa
Kata 2020	analiza dynamiki oraz struktury zobowiązań kredytowych rolników wobec banków w Polsce	gospodarstwa rolne	poziom zadłużenia rolników w bankach	4 zmienne	model: regresja liniowa
Omotesho, Douvi, Olaghere 2020	oszacowanie opłacalności i poziomu prawności technicznej upraw cebuli wraz z identyfikacją czynników wpływających na wydajność techniczną w produkcji cebuli	gospodarstwa rolne produkujące cebulę w Nigerii	produkcja; efekt nieefektywności technicznej	5 i 6 zmiennych	modele: funkcja produkcji Cobba-Douglasa i regresja liniowa
Salau 2020	pomiar stanu bezpieczeństwa żywnościowego i ocena wpływu bezpieczeństwa żywnościowego na technologie zrównoważonego zarządzania gruntami (SLM) w Nigerii	gospodarstwa rolne	ocena technologii zrównoważonego zarządzania gruntami (SLM)	5 zmiennych	model: regresja liniowa
Tanimonure, Yewande, Fayemi 2020	badanie aktualnych problemów ekonomicznych i izolowanych ograniczeń związanych z produkcją przypraw w stanie Ondo w Nigerii	gospodarstwa rolne	rentowność produkcji przypraw	9 zmiennych	model: regresja liniowa

Źródło: opracowanie własne

Literatura

- Aczel A.D. (2005) Statystyka w zarządzaniu. Wydawnictwo Naukowe PWN, Warszawa.
- Adamenko V., Oliynyk O., Wasilewski M. (2015) Modele finansowania kapitałem własnym i obcym przedsiębiorstw rolniczych na Ukrainie: wyzwania i perspektywy. *Zarządzanie Finansami i Rachunkowość*, 3(3), 93-106.
- Adamski M. (2009) Dochodowość materialnych czynników produkcji w gospodarstwach produkujących mleko w latach 2004-2006. *Journal of Agribusiness and Rural Development*, 3(13), 5-12.
- Bieniasz A., Gołaś, Z. (2012) Efektywność gospodarowania zapasami w przemyśle spożywczym. *Zagadnienia Ekonomiki Rolnej*, 332(3), 68-88.
- Bodnar O., Galchynska J., Maciejczak M. (2020) Price interdependence of agricultural commodities from Ukraine and world markets. *Acta Scientiarum Polonorum, Oeconomia*, 19(4), 15-22.
- Borkowski B., Dudek H., Szczesny W. (2017) *Ekonometria: wybrane zagadnienia*. Wydawnictwo Naukowe PWN, Warszawa.
- Bórawski P. (2015) Tendencje w handlu zagranicznym polskiego sektora mleczarskiego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 15(30)(1), 7-20.
- Bórawski P., Grzybowska-Brzezińska M., Dunn J.W. (2020) Dynamics of dairy products consumption in Poland against the background of the EU. *Journal of Agribusiness and Rural Development*, 2(56), 145-154.
- Bórawski P., Zalewski K. (2018) Czynniki kształtujące produkcję mleka w Polsce na tle UE. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(33)(3), 36-48.
- Buks J., Pietrzykowski R. (2015) Relacje między dochodem rolniczym a dopłatami unijnymi w Polsce. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(3), 67-71.
- Chrzanowska M., Drejerska N. (2016) Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej. *Taksonomia 27: Klasyfikacja i analiza danych – teoria i zastosowania. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 427, 58-65.
- Ciepiela G., Sosnowski J. (2011) Wpływ atrakcyjności wiejskiej przestrzeni rekreacyjnej na działalność agroturystyczną gospodarstw rolnych regionu siedleckiego. *Więś i Rolnictwo*, 1(150), 134-146.
- Czerwińska-Kayzer D., Florek J. (2018) Determinanty płynności finansowej przedsiębiorstw produkujących pasze dla zwierząt. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 20(6), 41-48.
- Czyżewski A., Kryszak Ł. (2016) Relacje głównych składowych determinujących rachunki dochodów rolniczych w krajach o różnym stażu członkowskim w Unii Europejskiej. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(4), 17-31.
- Czyżewski B., Majchrzak A. (2015) Związek dochodów, cen i produktywności w rolnictwie w Polsce – ujęcie makroekonomiczne. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(2), 26-31.
- Czyżewski B., Matuszczak A. (2016) Determinanty nożyc cen w rolnictwie krajów Unii Europejskiej o zróżnicowanej strukturze agrarnej. *Więś i Rolnictwo*, 3(172), 7-40.
- Dec M., Pawlak K., Poczta W. (2008) Determinanty sytuacji żywnościowej ludności świata. *Więś i Rolnictwo*, 2(139), 9-25.
- Domańska K., Kijek T. (2015) Inwestycje w innowacje w przedsiębiorstwach spożywczych zlokalizowanych w województwie lubelskim. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 109, 175-186.
- Dziechciarz J. (red.) (2002) *Ekonometria. Metody, przykłady, zadania*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Geweke J.F., Horowitz J.L., Pesaran M.H. (2006) *Econometrics: A Bird's Eye View*. IZA Discussion Paper Series, 2458, 3-4.
- Grabiński T., Wydymus S., Zeliaś A. (1982) *Metody doboru zmiennych w modelach ekonometrycznych*. PWN, Warszawa.

- Grochowska R., Mańko S. (2014) Produktywność gospodarstw rolnych w Polsce na tle innych krajów. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 14(29)(1), 25-33.
- Gruszczyński M., Podgórska M. (red.) (2004) *Ekonometria*. Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa.
- Grzelak A. (2016) The importance of selected determinants of assets reproduction in agricultural holdings in Poland. *Acta Scientiarum Polonorum, Oeconomia*, 15(2), 49-61.
- Hamulczuk M. (2016) Czynniki warunkujące kierunki zmian handlu targowiskowego w Polsce. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(2), 69-77.
- Hodun M., Kuś A. (2014) Determinanty kapitału obrotowego w przedsiębiorstwach przemysłowych – ujęcie modelowe. *Zarządzanie Finansami i Rachunkowość*, 2(2), 91-104.
- Juszczak S. (2011) Wpływ cen podstawowych produktów mleczarskich na ceny mleka surowego w Nowej Zelandii. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 11(26)(2), 56-66.
- Juszczak S. (2016) Znaczenie innowacji produktowych i kosztów ochrony środowiska w polskich spółdzielniach mleczarskich. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 18(4), 124-130.
- Juszczak S., Balina R. (2015) Financing of Investments in Dairy Cooperatives to Stabilize the Milk Market after the Liquidation of the Quota System. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(2), 107-111.
- Kata R. (2020) Zadłużenie rolników w Polsce w aspekcie przemian strukturalnych i koniunktury w rolnictwie. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 23(72), 33-44.
- Kisielińska J. (2012) *Podstawy ekonometrii w Excelu*. Wydawnictwo SGGW, Warszawa.
- Kisielińska J., Stańko S. (2009) Wielowymiarowa analiza danych w ekonomice rolnictwa. *Roczniki Nauk Rolniczych, Seria G*, 96(2), 63-76.
- Kobus P. (2015) Wpływ ubezpieczeń rolniczych na stabilność dochodową gospodarstw rolnych. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(6), 116-120.
- Koziół D., Zieliński W. (2009) Prognoza optymalnego terminu sadzenia sałaty w uprawach szklarniowych. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 9(24)(4), 83-88.
- Kubicová L., Dobák D., Kádekóvá Z. (2014) Trends in consumption of milk and dairy products in Slovakia after EU accession. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 12(61), 90-97.
- Kubicová L., Kádekóvá Z. (2013) Impact of consumer prices and cash income on consumption of dairy products. *Acta Scientiarum Polonorum, Oeconomia*, 12(3), 61-71.
- Kubik R., Wójcik E. (2015) Substytucyjność czynników produkcji w towarowych gospodarstwach rolnych ukierunkowanych na produkcję zwierzęcą. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(3), 225-230.
- Kufel J., Hamulczuk M. (2015) Koncentracja a wywieranie siły rynkowej w polskim przemyśle spożywczym. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(5), 158-163.
- Kukuła K. (red.), Goryl A., Jędrzejczyk Z., Osiewalski J., Walkosz A. (2009) *Wprowadzenie do ekonometrii*. Wydawnictwo Naukowe PWN, Warszawa.
- Kulawik J., Płonka R. (2014) Subsydia i efektywność ekonomiczno-finansowa a typ produkcyjny gospodarstw osób fizycznych. *Zagadnienia Ekonomiki Rolnej*, 340(3), 3-19.
- Kusz D., Gędek S. (2015) Egzogeniczne i endogeniczne uwarunkowania inwestycji w rolnictwie w Polsce. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(3), 237-242.
- Kusz D., Misiak T. (2017) Wpływ technicznego uzbrojenia pracy i postępu technicznego na wydajność pracy w rolnictwie. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 19(2), 145-150.
- Łukiewska K. (2014a) Produktywność w przemyśle spożywczym w Polsce – poziom i międzybranżowe zróżnicowanie. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 101(1), 53-63.

- Łukiewska K. (2014b) Wpływ wybranych czynników w Polsce na produktywność zasobów w przetwórstwie przemysłowym ze szczególnym uwzględnieniem przetwórstwa żywności. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 108, 145-158.
- Maddala G.S. (2006) *Ekonometria*. Wydawnictwo Naukowe PWN, Warszawa.
- Matras-Bolibok A. (2015) The Impact of R&D Expenditures on Bioeconomy Innovativeness. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(6), 180-184.
- Mekonnen H., Kebede K., Hasen M., Tegegne B. (2016) Farmer's Perception of Soil and Water Conservation Practices in Eastern Hararghe, Ethiopia. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(4), 232-239.
- Niewiadomski K. (2009) Ocena konwergencji rolnictwa w Polsce w latach 1998-2005. *Więś i Rolnictwo*, 3(144), 49-62.
- Niewiadomski K. (2010) Czynniki różnicujące rozwój gospodarstw agroturystycznych w Polsce. *Więś i Rolnictwo*, 2(147), 199-210.
- Niezgoda D. (2015) Opłacalność substytucji nawozów mineralnych środkami ochrony roślin w procesie produkcji roślinnej. *Zagadnienia Ekonomiki Rolnej*, 344(3), 3-18.
- Nowak E. (1984) *Problemy doboru zmiennych do modelu ekonometrycznego*. PWN, Warszawa.
- Nowak E. (1990) *Problem informacji w modelowaniu ekonometrycznym*. PWN, Warszawa.
- Nowak A. (2015) Zróżnicowanie regionalne elastyczności produkcji w towarowych gospodarstwach rolnych w Polsce. *Journal of Agribusiness and Rural Development*, 1(35), 75-82.
- Nowak A., Wójcik E., Krukowski A. (2015) The changes in productivity of production factors in commercial farms in Poland in 2004 and 2012. *Acta Scientiarum Polonorum, Oeconomia*, 14(3), 105-115.
- Omotesho O.A., Douvi C.P., Olaghere I.L. (2020) Economic viability of onion production in the commune of Grand-Popo, Republic of Benin. *Journal of Agribusiness and Rural Development*, 3(57), 289-297.
- Osińska M. (red.) (2007) *Ekonometria współczesna*. Wydawnictwo „Dom Organizatora”, Toruń.
- Paluch Ł., Sroka W. (2013) Socio-economic and environmental determinants of sustainable development of rural communes in Małopolska province. *Acta Scientiarum Polonorum, Oeconomia*, 12(2), 65-76.
- Papaelias T., Panagou V., Hyz A. (2014) The developments and trends of the agricultural sector in Greece. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 14(29)(4), 108-115.
- Pawlewicz A., Szamrowski P. (2014) Funkcjonowanie i rozwój rynku ekologicznych surowców żywnościowych w nowej perspektywie finansowej w latach 2014-2020. *Więś i Rolnictwo*, 3(164), 175-188.
- Ryś-Jurek R. (2009) The output, incomes and assets-capital relations in the individual farms. *Journal of Agribusiness and Rural Development*, 1(11), 177-188.
- Rzeszutko A. (2015) Regionalne zróżnicowanie wykorzystania potencjału produkcyjnego w rolnictwie polskim w warunkach wspólnej polityki rolnej. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 102(1), 46-58.
- Salau S.A. (2020) The influence of food security on Sustainable Land Management technologies of farming households in Kwara State, Nigeria. *Journal of Agribusiness and Rural Development*, 3(57), 309-315.
- Sass R. (2014) Wpływ inwestycji przed i po akcesji Polski do Unii Europejskiej na potencjał produkcyjny gospodarstw rolnych. *Journal of Agribusiness and Rural Development*, 2(32), 155-170.
- Sen A.K., Srivastava M.S. (1990) *Regression analysis: theory, methods and applications*. Springer, New York.
- Siudek T. (2014) Dług publiczny a rozwój gospodarczy obszarów wiejskich w krajach Unii Europejskiej. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 101(1), 15-30.
- Sobczyński T. (2009) Wpływ wielkości ekonomicznej gospodarstw rolniczych UE na ich możliwości rozwojowe. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 9(24)(4), 159-168.
- Sroka W. (2012) Zastosowanie modeli przyczynowo-skutkowych do oceny czynników sukcesu gospodarstw rolniczych – porównanie wybranych metod. *Zagadnienia Ekonomiki Rolnej*, 332(3), 28-42.
- Stanisz A. (2007) *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe*. Wydawnictwo Statsoft Polska, Kraków.

- Sytchevnik A., Golovkov V. (2010) Prospects of table eggs production in Belarus. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 10(25)(3), 106-113.
- Szajner P. (2019) Wahania cykliczne na światowym rynku cukru. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 19(34)(2), 186-195.
- Średzińska J. (2019) Determinants of the financial situation of farms in European Union countries grouped by economic size. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 21(4), 523-532.
- Tanimonure V.A., Yewande R.M., Fayemi O.G. (2020) Exploring economic opportunities in some selected spices farming enterprise in Ondo State, Nigeria. *Journal of Agribusiness and Rural Development*, 1(55), 85-92.
- Utnik-Banaś K., Żmija J. (2016) Wpływ wybranych czynników na cenę żywca brojlerów w latach 1995-2015. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(2), 344-352.
- Walesiak M., Gatnar E. (red.) (2004) *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Walesiak M., Gatnar E. (red.) (2009) *Statystyczna analiza danych z wykorzystaniem programu R*. Wydawnictwo Naukowe PWN, Warszawa.
- Wasąg Z. (2015) Wpływ wybranych czynników na efektywność wykorzystania funduszy Unii Europejskiej w gospodarstwach rodzinnych. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(4), 336-341.
- Wasilewska E. (2011) Model szybkiego testu płynności finansowej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 89, 109-122.
- Wasilewski M., Oliynyk O., Adamenko V. (2016) Financial System and Agricultural Growth: Evidence from Poland and Ukraine. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 15(30)(4), 204-214.
- Wasilewski M., Żurakowska-Sawa J. (2017) Uwarunkowania efektywności ekonomicznej przedsiębiorstw przemysłowych w fazie wprowadzenia ich cyklu życia. *Zarządzanie Finansami i Rachunkowość*, 5(3), 101-116.
- Welfe A. (2003) *Ekonometria*. PWE, Warszawa.
- Welfe W., Welfe A. (2004) *Ekonometria stosowana*. PWE, Warszawa.
- Wichowska A. (2019) Determinants of debt in rural municipalities on the example of the Warmińsko-Mazurskie Voivodeship. *Acta Scientiarum Polonorum, Oeconomia*, 18(4), 121-130.
- Wiśniewska J., Ryś-Jurek R. (2010) Agregaty dochodowe w rolnictwie indywidualnym w Polsce w latach 1990-2008. *Journal of Agribusiness and Rural Development*, 4(18), 107-118.
- Wysokiński M. (2016) Czynniki różnicujące dochodowość produkcji mleka w Polsce w specyficznych latach dla rolnictwa. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 18(6), 228-233.
- Zielińska A. (2018) Ocena zależności pomiędzy składką na ubezpieczenie wypadkowe a profilaktyką BHP w przedsiębiorstwach przemysłu spożywczego. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 104(3), 106-114.
- Zieliński W. (1998) *Analiza regresji*. Fundacja „Rozwój SGGW”, Warszawa.
- Zieliński M. (2014) Efektywność funkcjonowania gospodarstw specjalizujących się w uprawach polowych w latach 2005-2011. *Annals of The Polish Association of Agricultural and Agribusiness Economists*, 16(1), 250-256.

2. Regresja panelowa

Do metod wielowymiarowej analizy danych można zaliczyć również regresję panelową wykorzystującą pewien wektor zmiennych niezależnych. Terminem dane panelowe określa się zbiory danych powstających z połączenia szeregów czasowych obserwacji dla jednostek przekrojowych. Tego typu dane mają zatem dwa wymiary: przekrojowy (n) i czasowy (t). Wspomniany przekrój jest ważny, ponieważ każdy szereg czasowy powinien mieć informację o wartości pewnej zmiennej dla tej samej jednostki tylko w różnych okresach [Gruszczyński (red.) 2012]. Przykładem może być badanie ankietowe, które jest kierowane w okresie pięciu lat do tej samej grupy gospodarstw domowych. Gospodarstwa mają pewien identyfikator i , który umożliwia przedstawienie poziomu zjawiska dla tego samego obiektu (gospodarstwa) w okresie pięciu lat¹.

W literaturze przedmiotu dokonano podziału danych panelowych na mikropanele i makropanele. Mikropanele to dane, które charakteryzują się dużą liczbą obserwacji przekrojowych oraz niewielką liczbą obserwacji opisujących poszczególne obiekty w czasie, a makropanele są zbiorami danych, które cechuje mniejsza liczba obserwacji przekrojowych oraz dłuższy szereg czasowy [Dańska-Borsiak 2011]. Przykładem danych panelowych mogą być dane zebrane w wyniku badania nakładów inwestycyjnych na środki trwałe w województwach w kolejnych latach, badania dochodu rolniczego w wytypowanej grupie gospodarstw rolnych w kolejnych latach lub badanie liczby odwiedzin w punktach użyteczności publicznej przez użytkowników danej sieci telefonicznej w kolejnych dniach.

Zbiór danych otrzymany w wyniku pierwszego badania to przykład makropanelu i są to zazwyczaj dane wtórne zawarte na przykład w rocznikach statystycznych. Jednostkami statystycznymi w takich badaniach są zazwyczaj kraje, jednostki terytorialne, działy gospodarki, a informacje w nich zawarte dotyczą zazwyczaj wskaźników makroekonomicznych (np. PKB, inflacja, rentowność). Kolejne dwa przykłady badań pozwalają uzyskać tzw. mikropanele. Jednostkami statystycznymi w tego typu badaniach są zazwyczaj gospodarstwa domowe, pracownicy przedsiębiorstwa, zaś informacje zbierane są zazwyczaj w badaniu pierwotnym, które jest prowadzone przez dłuższy okres (nie jest jednorazowe). W ramach tego typu badań można zebrać informacje przykładowo o produkcji zwierzęcej gospodarstw rolnych czy dochodach pojedynczych osób. Jednym z najdłużej prowadzonych badań mikropanelowych jest amerykańskie PSID (ang. *panel study of income dynamics*). Badanie

¹ Uwaga: stosowanie metod ekonometrycznych opartych na danych panelowych jest możliwe również wtedy, gdy wymiary, w których identyfikowana jest obserwacja, są inne niż „czas” i „przestrzeń”. Przykładowo badanie pracowników w poszczególnych oddziałach korporacji (pierwszy wymiar) zatrudnionych na poszczególnych grupach stanowisk, które można uporządkować rosnąco (drugi wymiar).

rozpoczęło się w 1968 roku na reprezentatywnej w skali kraju próbie ponad 18 000 osób z 5000 rodzin w USA. Informacje o tych osobach i ich potomkach są gromadzone w sposób ciągły i dotyczą takich zmiennych jak zatrudnienie, dochód, wydatki, zdrowie, edukacja i wiele innych zmiennych². Polskim odpowiednikiem tego badania mogło być badanie Diagnoza Społeczna. Zgodnie z pierwotnym zamysłem badania w ramach Diagnozy Społecznej miały charakter panelowy w odstępach kilkuletnich (pierwszy pomiar odbył się w 2000 roku, następny trzy lata później, kolejne trzy pomiary odbyły się w odstępach dwuletnich)³. Poza tym w wielu krajach badania panelowe prowadzą narodowe urzędy statystyczne – w Polsce jest to Badanie Budżetów Gospodarstw Domowych.

Zainteresowanie metodami analiz danych panelowych utrzymuje się stale na wysokim poziomie⁴. Wynika to przede wszystkim z przewagi tego typu danych nad jednowymiarowymi szeregami czasowymi lub danymi przekrojowymi. Do podstawowych zalet danych panelowych zalicza się [Gruszczyński (red.) 2012]:

- Możliwość dysponowania większą liczbą dostępnych obserwacji: jest to szczególnie ważne w przypadku makropaneli, w których liczba dostępnych obserwacji jest ograniczona. Ponadto długie szeregi czasowe prezentujące zmienne ekonomiczne mogą wiązać się z występowaniem niestacjonarności i niestabilności w czasie. Uwzględnienie w badaniu drugiego wymiaru (przekroju) pozwala zastosować krótsze szeregi redukujące problemy typowe dla analizy szeregów czasowych, ale bez utraty cennych informacji.
- Możliwość uzyskania dodatkowych informacji o badanej zbiorowości: śledzenie tej samej grupy jednostek w pewnym okresie pozwala określić dynamikę procesu na poziomie mikro. Przykładowo w pewnym kraju zaobserwowano wzrost dochodu rolniczego w ciągu trzech lat o 20%, w związku z tym niech celem badawczym będzie odpowiedź na pytanie, jakie składowe wyniki finansowego miały na ten wzrost największy wpływ. Badanie tej samej grupy gospodarstw w ciągu trzech lat pozwala prześledzić, jak zmieniły się poszczególne składowe dochodu i odpowiedzieć na sformułowane pytanie badawcze. Takich danych nie można uzyskać na podstawie trzech różnych fal badania.
- Wykorzystanie panelu umożliwia usunięcie obciążenia estymatora w związku z pominięciem ważnego czynnika (na skutek błędu pominiętych zmiennych). Na tę kwestię po raz pierwszy zwrócił uwagę Mundlak [1961], rozważając funkcję produkcji Cobba-Douglasa w postaci $y_i = \mu + x_i^T \beta + \varepsilon_i$, gdzie y_i jest logarytmem wielkości produkcji, x_i wektorem logarytmów nakładów na czynniki produkcji, ε_i składnikiem losowym, a μ stałą. Mundlak zauważył istnienie tzw. efektu umiejętności menadżera (poziom zysku zależy od umiejętności zarządzania)

² Panel Study of Income Dynamics: <https://psidonline.isr.umich.edu> (dostęp 30.03.2021).

³ Diagnoza Społeczna: J. Czapiński, T. Panek (red.): www.diagnoza.com (dostęp 30.03.2021).

⁴ Ponad 16 000 wyników wyszukiwań dla *panel model* w bazie Web of Sciences w naukach ekonomicznych.

oraz zaproponował uwzględnienie tego efektu w modelu. Nie byłoby jednak to możliwe w przypadku danych przekrojowych, ponieważ efektów umiejętności menadżera jest tyle, ile firm w badaniu. Efekt ten staje się więc częścią składnika losowego, co ostatecznie skutkuje korelacją x_i z ε_i . Efekt umiejętności menadżera jest skorelowany z x_i , ponieważ wysokie umiejętności w zakresie zarządzania procesem produkcji przekładają się na wysoką efektywność zarządzania zasobami produkcyjnymi i wyższym y_i . W krótkim okresie, w wyniku osiągnięcia zysku, firma zwiększy inwestycje na czynnik x_i . W konsekwencji stosowany estymator będzie niezgodny. W przypadku dysponowania danymi panelowymi możliwe jest wprowadzenie do modelu efektu umiejętności menadżera w postaci m_i : $y_{it} = m_i + x_{it}^T \beta + \varepsilon_{it}$ (dla $i = 1, 2, \dots, n$ oraz $t = 1, 2, \dots, T$). W takim przypadku dla każdej firmy dysponuje się informacjami z wielu okresów, a zatem został uwzględniony stały w czasie, powiązany z pozostałymi zmiennymi i mający wpływ na zmienną objaśnianą efekt umiejętności menadżera.

Wady związane z wykorzystaniem danych panelowych mają charakter przede wszystkim techniczny. Zebranie pełnych mikropaneli wiąże się często z koniecznością prowadzenia czasochłonnego i kosztownego badania. W przypadku obserwowania przez kilka lat tej samej grupy jednostek statystycznych nie da się uniknąć problemu braku danych (np. na skutek odmowy uczestnictwa w dalszych badaniach czy śmierci). W takich sytuacjach otrzymuje się panel niezbilansowany, czyli zbiór danych panelowych, w którym nie dysponuje się informacjami o wartościach wszystkich zmiennych dla wszystkich jednostek we wszystkich okresach badania. Ten problem występuje w znacznie mniejszym stopniu w przypadku makropaneli.

2.1. Opis metod

Modele panelowe można podzielić na statyczne i dynamiczne. Modele dynamiczne są autoregresyjne, wszystkie pozostałe traktowane są jako statyczne. Najczęściej stosowanym modelem statycznym jest tzw. model jednokierunkowy mający postać⁵:

$$y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \quad (2.1)$$

gdzie $i = 1, \dots, n$ to numer obserwacji, $t = 1, \dots, T$ czas, y_{it} zmienna zależna, \mathbf{x}_{it} wektor zmiennych niezależnych, ε_{it} składnik losowy, $\boldsymbol{\beta}$ wektor parametrów modelu, α_i to efekty indywidualne.

Efekt indywidualny obejmuje stałe w czasie charakterystyki danej jednostki statystycznej, które wpływają na zmienną zależną, ale które nie zostały uwzględnione w wektorze zmiennych niezależnych \mathbf{x}_{it} . Efektem indywidualnym może być wpływ menadżera na poziom zysku zapisany wcześniej jako m_i . Należy jednak zauważyć, że jeżeli istnieją inne, stałe w czasie i nieuwzględnione

⁵ Zastosowane w pracy oznaczenia przyjęto za [Gruszczyński (red.) 2012].

w wektorze \mathbf{x}_{it} charakterystyki, to będą one uwzględnione w tym samym α_i , a ich podział nie będzie możliwy [Gruszczyński (red.) 2012].

W przypadku zastosowania modelu nieuwzględniającego żadnych efektów indywidualnych należy założyć homogeniczność populacji oraz, że różnice zachodzące między rzeczywistymi wartościami zmiennej zależnej a wartościami teoretycznymi są jedynie następstwem występowania składnika losowego [Geise 2013]. Wspomniane założenia są w praktyce trudne do spełnienia ze względu na niejednorodność badanych populacji. Zalecane jest zatem uwzględnienie w modelu różnic między poszczególnymi jednostkami w postaci efektów indywidualnych. W związku z tym wprowadzono rozróżnienie na efekty ustalone (ang. *fixed effects*) oraz efekty losowe (ang. *random effects*).

Modele z efektami stałymi (ustalonymi). Modele z efektami stałymi (ustalonymi) (ang. *fixed effects regression models*) stosuje się w przypadku stwierdzenia występowania efektów indywidualnych. Zakłada się przy tym występowanie nieznanymi (nieobserwowalnymi), ale stałymi w czasie różnic między obiektami. Model taki przyjmuje postać podaną w równaniu (2.1). Ewentualny alternatywny zapis to:

$$y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \sum_{j=1}^n \alpha_j d_{ij} + \varepsilon_{it}, \quad (2.2)$$

gdzie d_{ij} to zmienna zero-jedynkowa przyjmująca wartość 1 dla $i = j$ oraz 0 dla $i \neq j$. Składniki losowe mają niezależne i identyczne rozkłady, czyli $\varepsilon_{it} \sim IID(0; \sigma_\varepsilon^2)$.

Zapis (2.2) oznacza, że efekty indywidualne mogą być traktowane jako wyrazy wolne w poszczególnych szeregach czasowych, ale są różne dla poszczególnych obiektów, czyli mają charakter stałego w czasie parametru, ale różnego dla poszczególnych i . Ponadto omawiany model nie uwzględnia w swojej postaci wyrazu wolnego, ponieważ doprowadziłoby to ścisłej współliniowości ze zmiennymi d_{ij} , co stanowi pewne ograniczenie modelu [Gruszczyński (red.) 2012].

Parametry modelu można szacować przy użyciu klasycznej metody najmniejszych kwadratów (estymator KMNK)⁶. Estymacja modelu w postaci (2.2) może być jednak obliczeniowo utrudniona, ponieważ liczba zmiennych objaśniających wynosi $k + n$, gdzie k oznacza liczbę zmiennych objaśniających w wektorze \mathbf{x}_{it} (ograniczenie jest szczególnie istotne w przypadku mikropaneli, w których n jest duże). W takiej sytuacji stosuje się tzw. transformację wewnątrzgrupową. Niech \bar{y}_i oznacza średnią arytmetyczną wartości zmiennej zależnej $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$ (analogicznie można zdefiniować $\bar{\mathbf{x}}_i$ oraz $\bar{\varepsilon}_i$). Równanie (2.1) można zapisać dla i -tej jednostki w każdym okresie (tj. w pierwszym, drugim itd.). Następnie dodając je do siebie stronami i dzieląc przez liczbę okresów T , otrzyma się równanie:

⁶ W literaturze zastosowany estymator nazywany jest *least squares dummy variables* (LSDV). Słowo *dummy* oznacza zmienną binarną.

$$\bar{y}_i = \bar{\mathbf{x}}_i^T \boldsymbol{\beta} + \alpha_i + \bar{\varepsilon}_i, \quad (2.3)$$

przy czym w przypadku parametru α_i należy pamiętać o jego stałości w czasie.

Transformacji wewnątrzgrupowej dokonuje się, odejmując równanie (2.3) od (2.1), co można zapisać następująco:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T \boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i). \quad (2.4)$$

W powyższym równaniu wyeliminowane zostały już efekty indywidualne. Następnie dla każdej obserwacji można wprowadzić oznaczenia: $\tilde{y}_{it} = y_{it} - \bar{y}_i$ oraz $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, a następnie oszacować parametry strukturalne metodą KMNK modelu, w którym zmienną zależną będzie \tilde{y}_{it} , a wektorem zmiennych niezależnych $\tilde{\mathbf{x}}_{it}$. Estymator uzyskany KMNK ma postać:

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i)^T. \quad (2.5)$$

W ostatnim kroku możliwe jest oszacowanie efektów indywidualnych w analogiczny sposób, jak szacowane są wyrazy w modelu KMNK:

$$\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i^T \hat{\boldsymbol{\beta}}_{FE}. \quad (2.6)$$

Estymator ten nazywany jest estymatorem wewnątrzgrupowym (ang. *within groups*) lub estymatorem *fixed effects*. Oszacowania parametrów w wektorze $\boldsymbol{\beta}$ są takie same jak przy użyciu estymatora LSDV stosowanego przy pominięciu procedury transformacji wewnątrzgrupowej. Estymator zapisany we wzorze (2.5) ma, przy spełnieniu odpowiednich założeń⁷, własności takie jak zgodność, nieobciążoność i efektywność.

Modele z efektami losowymi. Modele z efektami losowymi (ang. *random effects*) o postaci jednokierunkowej można zapisać następująco:

$$y_{it} = \mu + \mathbf{x}_{it}^T \boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \quad (2.7)$$

gdzie μ oznacza wyraz wolny. Uwzględnienie wyrazu wolnego w tym przypadku implikuje różnice między równaniami (2.1) i (2.7).

W modelu z efektami losowymi każdemu obiektowi przypisana jest zmienna losowa, której realizacja odpowiada za efekt indywidualny w danym okresie. Zatem w modelu tym efekty indywidualne nie są stałe w całym okresie. W konsekwencji efekty te nie są traktowane jak parametry i nie szacuje się ich wartości. Efekty indywidualne można interpretować w tym przypadku jako indywidualne składniki losowe, poszerzają zatem część stochastyczną modelu (w modelu *fixed* efekty

⁷ Oprócz wspomnianego wcześniej niezależnego i identycznego rozkładu składnika losowego konieczne jest założenie o ścisłej egzogeniczności zmiennych niezależnych, czyli brak związku między zmiennymi objaśnianymi a składnikiem losowym dla danej jednostki. Do modelu nie może też zostać włączona opóźniona zmienna objaśniana.

indywidualne stanowią część deterministyczną modelu). Jeżeli do modelu wprowadzi się łączny składnik losowy $v_i = \alpha_i + \varepsilon_i$, czyli zmienną losową zdefiniowaną jako sumę indywidualnego składnika losowego i białego szumu, to równanie będzie miało postać:

$$y_{it} = \mu + \mathbf{x}_{it}^T \boldsymbol{\beta} + v_i, \quad (2.8)$$

gdzie μ oznacza wyraz wolny.

W przeciwieństwie do modeli z efektami ustalonymi uwzględnienie w tym przypadku wartości stałych w czasie dla i -tego obiektu jest możliwe, ponieważ nie spowoduje pojawienia się ścisłej współliniowości ze zmiennymi zero-jedynkowymi. Jest to zaleta modeli typu *random effects* [Gruszczyński (red.) 2012].

W przypadku modeli z efektami losowymi, podobnie jak dla wcześniejszego przypadku, zakłada się, że $\varepsilon_{it} \sim IID(0; \sigma_\varepsilon^2)$, oraz przyjmuje się ścisłą egzogeniczność zmiennych niezależnych zgodnie ze wzorem $E(x_{it} \cdot \varepsilon_{it}) = 0$. Ponadto konieczne jest przyjęcie pewnych dodatkowych założeń, które są konsekwencją przeniesienia efektów indywidualnych do części stochastycznej modelu:

- dla każdego i -tego obiektu zachodzi $\alpha_i \sim IID(0; \sigma_\alpha^2)$ (jest to założenie dotyczące rozkładu efektów indywidualnych),
- niezależność α_i od x_{jt} dla dowolnego i oraz t (założenia pozwala na uniknięcie problemu endogeniczności zmiennych objaśniających),
- niezależność efektów indywidualnych od składników losowych, czyli $E(\varepsilon_{it} \cdot \alpha_j) = 0$, dla dowolnego i oraz t .

Estymator KMNK zastosowany do szacowania parametrów modelu (2.7) byłby zgodny i nieobciążony, ale nie byłby efektywny (wynika to z braku łącznego składnika losowego – wariancja efektów indywidualnych jest różna od zera). W takiej sytuacji lepszym rozwiązaniem jest zastosowanie estymatora uogólnionej metody najmniejszych kwadratów (UMNK) [Gruszczyński (red.) 2012].

Po odpowiednich przekształceniach estymator ma postać:

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T + \varphi T \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \right)^{-1} \times \\ \times \left(\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i)^T + \varphi T \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(y_i - \bar{y}) \right), \quad (2.9)$$

gdzie:

$$\varphi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}. \quad (2.10)$$

Można zauważyć, że przy $\varphi = 0$ estymator (2.9) jest taki sam jak stosowany przy modelach z ustalonymi efektami – wzór (2.5). Dla $\varphi > 0$ estymator UMNK jest bardziej efektywny niż estymator *fixed effects*. Przy $\varphi = 1$ staje się estymatorem KMNK. W szacowaniu parametrów w przypadku

random effects ważne jest również założenie o niezależności zmiennych objaśniających i efektów indywidualnych. Efekty te są częścią składnika losowego, a zatem zależność α_i i x_{jt} będzie prowadzić do utraty zgodności estymatora. Szczególnie założenie to nie byłoby spełnione, gdyby do modelu włączono opóźnioną zmienną zależną – estymator FE, tak jak i estymator RE, nie może być stosowany do estymacji modeli dynamicznych [Gruszczyński (red.) 2012].

Weryfikacja liniowych modeli statycznych dla danych panelowych. Metody weryfikacji modeli panelowych można podzielić na metody stosowane w analizie modeli opartych na danych przekrojowych lub szeregach czasowych oraz metody specyficzne dla modeli panelowych i niestosowane w innych sytuacjach. Druga grupa metod ma na celu ocenę zasadności uwzględnienia w modelu efektów indywidualnych w formie losowej lub ustalonej. W literaturze podaje się trzy najczęściej stosowane testy weryfikujące słuszność uwzględnienia danych efektów. Pierwszym z nich jest test Walda, który stosuje się do testowania istotności efektów ustalonych [Geise 2013]. Na podstawie testu Hausmana dokonuje się wyboru między estymatorem efektów stałych a estymatorem losowych efektów indywidualnych [Hausman 1978]. Testowanie efektów losowych odbywa się natomiast przy wykorzystaniu testu Breuscha-Pagana [Breusch, Pagan 1980].

W przypadku braku konieczności uwzględniania w modelu efektów indywidualnych model (2.1) sprowadza się do postaci:

$$y_{it} = \mu + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}. \quad (2.11)$$

Parametry tego modelu mogą być szacowane za pomocą KMNK, ponieważ zbiór danych jest traktowany nie jako panel, ale jako zbiór $n \cdot T$ obserwacji przekrojowych (tym samym pomijany jest efekt indywidualny).

Procedura weryfikacji istotności efektów indywidualnych (ustalonych), czyli **test Walda**, polega na zbadaniu słuszności hipotezy zerowej, zgodnie z którą brak jest zróżnicowania, a statystyka testowa oparta jest na różnicy współczynników determinacji w modelach (2.1) i (2.11) oszacowanych na podstawie tych samych danych:

$$H_0: \forall_{i,t} \alpha_{it} = \alpha,$$

$$H_1: \exists_{i,t} \alpha_{it} \neq \alpha,$$

a statystyka testowa przyjmuje postać⁸:

$$F = \frac{(R_1^2 - R_0^2)/(n-1)}{(1-R_1^2)/(nT-n-k)}. \quad (2.12)$$

Hipotezę zerową należy odrzucić, jeśli różnica $R_1^2 - R_0^2$ będzie względnie wysoka, czyli model z efektami indywidualnymi (ustalonymi) będzie znacznie lepiej dopasowany do danych niż model ze stałym wyrazem.

⁸ Najczęściej podaje się statystykę F (test χ^2 ma nieco gorsze własności dla małych prób).

Inaczej przebiega weryfikacja istotności efektów indywidualnych z efektami losowymi. Statystyczna weryfikacja słuszności ich uwzględnienia polega na zbadaniu prawdziwości hipotezy zerowej:

$$H_0: \sigma_\alpha^2 = 0$$

wobec hipotezy alternatywnej:

$$H_1: \sigma_\alpha^2 \neq 0.$$

Weryfikacji hipotezy dokonuje się na podstawie statystyki **testu Breuscha-Pagana**:

$$\chi = \frac{nT}{2(T-1)} \left[\sum_{i=1}^n (\sum_{t=1}^T \hat{v}_{it})^2 / \sum_{i=1}^n \sum_{t=1}^T \hat{v}_{it}^2 - 1 \right]. \quad (2.13)$$

Statystyka ta ma rozkład chi-kwadrat z jednym stopniem swobody. Brak podstaw do odrzucenia H_0 wskazuje na słuszność zastąpienia modelu z efektami losowymi (2.7) przez podstawowy model bez efektów indywidualnych (2.11).

Ostatnim z szerzej stosowanych testów jest **test Hausmana**, który służy do odpowiedzi na pytanie, które z efektów – stałe czy losowe – powinny być w danym przypadku uwzględnione. W teście tym zakłada się, że jeżeli spełnione będzie założenie o niezależności zmiennych obserwowalnych od efektów indywidualnych, estymator *random effects* będzie zgodny i bardziej efektywny niż estymator *fixed effects*. Jeśli spełnione są tylko założenia potrzebne dla zastosowania estymatora *fixed effects* oraz efekty indywidualne i zmienne objaśniające nie są niezależne, to estymator *random effects* traci zgodność i będzie dawał inne wyniki. Hipotezy można zapisać następująco:

H_0 : Założenie niezależności jest spełnione, obydwa estymatory są nieobciążone, a estymator *random effects* jest bardziej efektywny.

H_1 : Założenie nie jest spełnione, estymator *fixed effects* jest nieobciążony zaś *random effects* jest obciążony, lub nastąpił błąd specyfikacji modelu.

Test sprowadza się zatem do oszacowania modelu za pomocą obydwu estymatorów oraz zbadania różnic między tymi oszacowaniami. Test jest oparty na statystyce zdefiniowanej następującym wzorem:

$$W = (\hat{\beta}_{FE} - \hat{\beta}_{RE})^T \cdot [\hat{V}(\hat{\beta}_{FE}) - \hat{V}(\hat{\beta}_{RE})]^{-1} \cdot (\hat{\beta}_{FE} - \hat{\beta}_{RE}), \quad (2.14)$$

gdzie $\hat{V}(\hat{\beta}_{FE})$ oraz $\hat{V}(\hat{\beta}_{RE})$ są estymatorami macierzy wariancji–kowariancji oszacowań wektorów parametrów. Statystyka testowa ma rozkład chi-kwadrat z liczbą stopni swobody równą liczbie parametrów szacowanych w obydwu modelach (bez wyrazu wolnego, który nie jest szacowany w modelu *fixed effects*). Brak różnicy bądź ich niewielka wartość skłania do wniosku o braku podstaw do odrzucenia H_0 (finalnie preferowany będzie estymator *random effects*), natomiast znaczące różnice powodują odrzucenie H_0 na korzyść H_1 . Podsumowanie omówionych testów zamieszczono w tabeli 2.1.

Tabela 2.1. Weryfikacja liniowych modeli statycznych dla danych panelowych – wybrane testy

Model bez efektów indywidualnych	Model z efektami ustalonymi	Model z efektami losowymi
Test Walda <i>H₀: brak różnicowania efektów indywidualnych</i>		
Test Breuscha-Pagana <i>H₀: efekt indywidualny jest stały dla wszystkich jednostek i można go zastąpić wyrazem wolnym</i>		Test Breuscha-Pagana <i>H₀: efekt indywidualny jest stały dla wszystkich jednostek i można go zastąpić wyrazem wolnym</i>
	Test Hausmana <i>H₀: obydwa estymatory są nieobciążone, a estymator <i>random effects</i> jest bardziej efektywny</i>	

Źródło: opracowanie własne

Należy zaznaczyć, że statycznych modeli liniowych jest znacznie więcej niż omówiono w niniejszym opracowaniu. Często zasadnym rozwiązaniem jest wprowadzenie do modelu nie tylko efektów indywidualnych, ale też efektów czasowych, tak rozbudowany model nazywany jest modelem dwukierunkowym. Efekty czasowe są stałe dla wszystkich jednostek w danym okresie, ale różnią się między okresami. Efekty czasowe również dzielą się na ustalone i losowe. Ważną grupę modeli stanowią również modele dynamiczne, uwzględniające w swojej postaci jako zmienną objaśniającą $y_{i,t-1}$ [Gołaś 2019]. Inną kategorią modeli danych panelowych są też modele zmiennych dwumianowych czy jakościowe modele panelowe [Kisielińska 2008]. Właściwe są również inne metody estymacji tego typu modeli.

2.2. Zastosowania metod regresji panelowej

W literaturze przedmiotu można znaleźć wiele zastosowań regresji panelowej w obszarze ekonomiki rolnictwa. W zakresie zastosowania makropaneli dominują badania, w których wymiarem przekrojowym były państwa Unii Europejskiej (UE). Dla przykładu Staniszewski i Czyżewski [2016] analizowali oddziaływanie wybranych zmiennych strukturalnych w krajach UE w latach 2005, 2007, 2010 oraz 2012 na produktywność i dochodowość podstawowych czynników wytwórczych – pracy, kapitału i ziemi. W większości przypadków (w zależności od zmiennej objaśnianej) przeprowadzone testy wskazywały, że właściwa okazała się postać modelu regresji panelowej o stałych efektach ustalonych.

Makropanele dla krajów UE są częstym zastosowaniem, przy czym niektórzy autorzy dokonywali podziału na kraje „nowej” i „starej” UE. Takie podejście zastosowali w swojej pracy Wasilewska i Pietrych [2018], w której zbadali oni relację między starzeniem się społeczeństwa krajów UE a wzrostem gospodarczym, przy czym kraje pogrupowano zgodnie z okresem przystąpienia do Unii. Znane są również modele oszacowane tylko dla państw wysoko rozwiniętych (Japonii, USA) czy krajów UE [Czyżewski, Kułyk 2017]. Wyniki estymacji modelu regresji panelowej potwierdziły

w tym przypadku istotną rolę uwarunkowań makroekonomicznych w utrzymywaniu procesów rozwojowych rolnictwa w tych gospodarkach.

Bereźnicka [2018a] w celu rozpoznania sytuacji ekonomicznej gospodarstw funkcjonujących w krajach UE oraz wskazania czynników, które oddziaływały na nadwyżkę finansową, prezentuje jeszcze inny podział makropanelu. Ze względu na duże zróżnicowanie (w systemie FADN) siły ekonomicznej gospodarstw w poszczególnych krajach UE zdecydowała się na zastosowanie metody kwartyli do utworzenia jednorodnych grup, które były podstawą do budowy modeli. Poniżej pierwszego kwartyla znalazły się kraje z gospodarstwami słabymi ekonomicznie, natomiast powyżej trzeciego kwartyla – kraje z gospodarstwami silnymi ekonomicznie.

Inny podział badań w zakresie ekonomiki rolnictwa, jaki można zastosować, to podział uwzględniający przedmiot badania. W tym przypadku spora część badań koncentruje się na kwestii rentowności poszczególnych typów produkcji. Przykład zastosowania dla produkcji pszenicy można znaleźć w pracy [Gołaś 2012a] – wynik badań wskazywały, że na zmiany rentowności najsilniej wpływały ceny i plony pszenicy oraz koszty nawożenia i koszty energii. Podobnym przykładem zastosowanie makropaneli było modelowanie zmienności poziomu rentowności produkcji mleka w krajach UE [Gołaś 2017b]. Z punktu widzenia oszacowanych parametrów modeli regresji zmienność rentowności produkcji mleka uwarunkowana była w głównej mierze przez powierzchnię paszową, wielkość stada krów, wydajność mleczną krów, ceny mleka oraz koszty energii i koszty wynagrodzeń, natomiast w relatywnie mniejszym stopniu przez wydajność mleczną krów oraz koszty pasz. Kolejnym przykładem zastosowania makropaneli może być badanie źródeł rentowności kapitału własnego z podziałem na kraje o silnych oraz słabych ekonomicznie gospodarstwach rolnych [Bereźnicka 2018b]. W tym badaniu pominięto sytuację ścisłej egzogeniczności oraz występowania zmiennych, których wartości są stałe w czasie dla wszystkich jednostek (krajów) panelu.

Badaniom mogą podlegać również czynniki kształtujące ceny surowców w poszczególnych krajach UE. Przykładowo Olipra [2019] przeanalizował ceny mleka. Wyniki testu Hausmana wskazały, że specyfikacja efektów losowych jest bardziej odpowiednia niż efekty stałe. Najprawdopodobniej odzwierciedla to to, że w dzisiejszych czasach czynniki specyficzne dla danego kraju, takie jak warunki klimatyczne, są mniej ważne.

Zalety modeli dla danych panelowych podkreśla również Kufel [2014] w badaniu wysokości i zmienności marż w przemyśle spożywczym krajów UE. Modele panelowe zostały wybrane ze względu na możliwość badania zależności między typem zmiennych wybranych do analizy (między zmiennymi świadczącymi o koniunkturze a zmienną świadczącą o poziomie marż). Ponadto autor zauważa, że ze względu na ograniczoną liczbę zmiennych objaśniających, zastosowanie zwykłego modelu regresji skutkowałoby pominięciem istotnych zmiennych objaśniających i obciążeniem estymatorów.

Analiza regresji panelowej była również przeprowadzana do badania, które programy WPR mają wpływ na dochody z gospodarstwa rolnego w przeliczeniu na jednostkę pracy rodziny w poszczególnych klasach wielkości ekonomicznej [Guth, Smędzik-Ambroży 2017]. Na podstawie testu Hausmana stwierdzono, że model z efektami losowymi okazał się być odpowiedni, co oznacza, że nie występują istotne statystycznie różnice między wpływem poszczególnych grup subsydiów na dochody z gospodarstwa rolnego w przeliczeniu na jednostkę pracy rodziny w poszczególnych krajach UE. Znaczenie finansowego wsparcia na alokację zasobów w rolnictwie zostało udowodnione za pomocą modeli panelowych w pracy Czyżewskiego i Kułyka [2014]. Współzależności między zmianami w finansowym wsparciu rolnictwa a przepływem mobilnych zasobów w rolnictwie były bardziej widoczne w powiązaniu z czynnikiem ziemi (niższa mobilność tego zasobu niż w przypadku zasobów pracy wymaga kompensacji poprzez finansowe wsparcie rolnictwa). Kolejnym przykładem podziału makropanelu na grupy państw w badaniu wpływu WPR na dochody jest opracowanie Czyżewskiego i Poczta-Wajdy [2017]. W celu sprawdzenia, czy członkostwo w UE i korzystanie ze środków WPR pomagają zredukować relatywną lukę dochodową w rolnictwie, oszacowano trzy modele regresji panelowej: dla wszystkich, „starych” i „nowych” krajów członkowskich.

Prowadzone były również liczne badania w zakresie wpływu rolnictwa na środowisko. Część z nich uwzględniała determinanty ekonomiczne. Czyżewski i Kryszak [2017] poddali weryfikacji hipotezę, że produktywność czynników wytwórczych oraz charakterystyki opisujące intensywność gospodarowania, inaczej wpływają na emisję gazów cieplarnianych w zależności od typu rolnictwa. Pod uwagę wzięto cechy w ujęciu zarówno mikro- (wynagrodzenie czynników wytwórczych, nakłady związane z intensywnością gospodarowania), jak i makroekonomicznym (PKB *per capita*, przeciętna wielkość gospodarstwa rolnego, poziom wsparcia sektora rolnego). Modele panelowe były również wykorzystywane do badania związków między wynikami ekonomicznymi gospodarstw rolnych a ich oddziaływaniem na środowisko [Grzelak 2018]. Estymując model panelowy z wykorzystaniem uśrednionych wartości zmiennych na podstawie danych FADN, dla każdego z krajów UE stwierdzono, że korzystniejszej sytuacji ekonomicznej badanej grupy gospodarstw rolnych towarzyszy silniejsze negatywne oddziaływanie na środowisko.

Innym badaniem z zakresu makropaneli może być analiza oparta na danych FAO dotyczących podaży kalorii na mieszkańca danego kraju i jej podział na produkty pochodzenia zwierzęcego i roślinnego [Dudek, Krawiec, Koszela 2015]. W tabeli 2.2 zaprezentowano przykłady zastosowań modeli panelowych (z podziałem na mikro- i makropanele).

Przekładem zastosowania mikropaneli jest badanie panelowe polskich gospodarstw rodzinnych [Halamska, Lamarche, Maurel 2003]. Badania panelowe przeprowadzone w latach 1988 i 2001 dokumentują zmiany, jakie zaszły w rolnictwie rodzinnym w Polsce. W pracy autorzy skupili również uwagę na pożytkach płynących z badań panelowych. Inny rodzaj badania gospodarstw domowych

prezentuje Dobrowolska [2012]. Jest ono przykładem na to, że modele regresji panelowej mogą być stosowane do badania redystrybucyjnych efektów funkcjonowania podatków pośrednich. Autorka zwróciła uwagę, że zaletą wykorzystania regresji panelowej do badania redystrybucyjnych konsekwencji funkcjonowania podatków konsumpcyjnych jest to, że pozwalają one wskazać typy gospodarstw domowych najbardziej obciążonych poszczególnymi rodzajami podatków pośrednich.

Mikropanele znajdują również zastosowanie dla danych FADN. Czekaj [2011] w swoim badaniu porównał różne modele ekonometryczne (uogólniona regresja, regresja o stałych i zmiennych efektach), które mogą być stosowane w szacowaniu modeli inwestycyjnych na podstawie danych panelowych dotyczących polskich gospodarstwach rolnych, zbieranych w systemie rachunkowości rolnej Polski FADN. Według Grzelaka i Wiktorowicz [2009] krótkie szeregi czasowe uniemożliwiają analizę klasycznymi metodami (w tym KMNK). Skłoniło to autorów do podjęcia próby opisu związku między wysokością dopłat a dochodem osiąganym przez gospodarstwa indywidualne przy wykorzystaniu próby przekrojowo-czasowej (dane obejmowały cztery lata, każda ze zmiennych rejestrowana była dla siedmiu obiektów, tj. gospodarstw rolnych, wyróżnionych według typu działalności). Dla tego typu danych MNK nie powinna być stosowana z uwagi na ryzyko wystąpienia heteroskedastyczności. Problem ten jest zminimalizowany w modelach panelowych, co skłania do stosowania tego podejścia w modelowaniu dochodów gospodarstw.

Kolejny przykład zastosowania mikropaneli można odnaleźć w publikacji Stępnia, Smędzik-Ambroży i Guth [2017]. Analizowane dane były połączonymi danymi przekrojowymi (dla klas wielkości ekonomicznej) i danymi dotyczącymi szeregów czasowych (2004-2013), dlatego też zbadano zależność dochodu netto z gospodarstwa rolnego w przeliczeniu na jednostkę pracy rodziny od różnych grup subsydiów, przy użyciu regresji panelowej.

Z zastosowaniem mikropanelu badano również relacje między rentownością a płynnością finansową przedsiębiorstw sektora przetwórstwa spożywczego. Autorzy w tym przypadku zwracali również uwagę, że istotnymi ograniczeniami diagnostycznym mogą być krótki okres badawczy (cztery lata) oraz wysokie rozproszenie wskaźników płynności [Jaworski, Czerwonka, Mądra-Sawicka 2018]. Z kolei Franc-Dąbrowska [2009] wykorzystując mikropanele dla przedsiębiorstw rolniczych, stwierdza, że uzyskane wyniki wskazują na przewagę modeli panelowych o stałych efektach, nad modelami panelowymi budowanymi za pomocą estymacji KMNK oraz modeli panelowych o zmiennych efektach dla badanej grupy przedsiębiorstw.

Podsumowując, modele panelowe mają szerokie zastosowanie w analizach w obszarze ekonomiki rolnictwa. W literaturze przeważają modele dla makropaneli, co częściowo wynika z dostępności danych. Analizy panelowe znalazły zastosowanie w takich obszarach jak badanie rentowności firm z branży rolno-spożywczej, badanie czynników wpływających na wynik finansowych gospodarstw rolnych czy badanie wpływu redystrybucji wsparcia finansowego na dochody rolnicze.

Tabela 2.2. Przykłady zastosowań modeli panelowych w ekonomice rolnictwa

Autor	Problem	Obiekty	Metoda
Gołaś 2017a	wielowymiarowa analiza rentowności produkcji pszenicy	kraje UE, okres: 2007-2013	makropanele, model z efektami stałymi i losowymi
Kufel 2014	oszacowanie wysokości i zmienności marż w przemyśle spożywczym krajów UE oraz poszukiwanie zależności między poziomem marż w przemyśle spożywczym państw UE a koniunkturą makroekonomiczną i sektorową w tym okresie z zastosowaniem regresji panelowej	kraje UE, okres: 1995-2010	makropanele, model o losowych i ustalonych efektach
Czyżewski, Kułyk 2014	wskazanie określonych prawidłowości co do wielkości zasobów alokowanych w rolnictwie i związkach tego z finansowym wsparciem rolnictwa	badaniami objęto trzydzieści państw, w tym ugrupowanie integracyjne UE-15 potraktowane jako jeden; okres: 1986- 2012	makropanele, model o ustalonych efektach
Dudek, Krawiec, Koszela 2015	ocena zmian spożycia żywności w UE w odniesieniu do ich wpływu na środowisko	kraje UE (podział na kraje „starej” Unii i „nowej” Unii), okres: 2004-2011	makropanele, model o losowych efektach
Staniszewski, Czyżewski, 2016	wpływ zmiennych strukturalnych na produktywność i dochodowość podstawowych czynników wytwórczych	kraje UE, okres: 2005, 2007, 2010, 2013	makropanele, model o losowych ustalonych efektach i z brakiem efektów indywidualnych
Guth, Smeździk-Ambroży, 2017	ocena wpływu subsydiów wspólnej polityki rolnej na dochody gospodarstw mlecznych FADN	kraje UE (z podziałem na kraje UE-10 i UE-15), okres: 2004-2013	makropanele, model o losowych efektach
Grzelak 2018	zbadanie związków między wynikami ekonomicznymi gospodarstw rolnych a ich oddziaływaniem na środowisko	kraje UE (uśrednione wyniki gospodarstw rolnych prowadzących rachunkowość rolną), okres: 2004-2015	makropanele, model o ustalonych efektach
Gołaś 2017b	uwarunkowania rentowności produkcji gospodarstw rolnych wyspecjalizowanych w produkcji mleka (modelowania zmienności poziomu rentowności)	kraje UE, okres: 2007-2013	makropanele (model o ustalonych efektach)
Czyżewski, Poczta-Wajda 2017	sprawdzenie, czy członkostwo w UE i korzystanie ze środków WPR pomagają zredukować relatywną lukę dochodową w rolnictwie oraz zbadanie, które czynniki wpływają na tę lukę i w jaki sposób	kraje UE (z podziałem na „starych” i „nowych” członków), okres: 1995–2015	makropanele, model o losowych i ustalonych efektach
Czyżewski, Kułyk 2017	ocena wpływu otoczenia makroekonomicznego na rozwój sektora rolnego w krajach wysokorozwiniętych	państwa wysokorozwinięte: Japonia, UE oraz US, okres: 1990-2016	makropanele, modele z ustalonymi efektami
Czyżewski, Kryszak 2017	badanie, w jaki sposób cechy ekonomiczne poszczególnych typów rolnictwa wpływają na emisję zanieczyszczeń gazowych	40 krajów świata, okres: 1995-2009	makropanele, model o ustalonych efektach
Wasilewska, Pietrych 2018	relacja między starzeniem się społeczeństwa krajów UE a wzrostem gospodarczym	kraje UE (podział na kraje „starej” Unii i „nowej” Unii), okres: 1996-2016	makropanele, model o losowych i ustalonych efektach
Firlej, Kubala 2018	oceny zachodzących zmian w cenach ziemi rolnej w Polsce na tle UE oraz określenie zmiennych mogących mieć silny wpływ na kształtujący się średni poziom cen gruntów rolnych w krajach należących do UE	kraje UE, okres: 2005-2015	makropanele, model o losowych efektach

Tabela 2.2 (cd.)

Autor	Problem	Obiekty	Metoda
Bereźnicka 2018a	rozpoznanie sytuacji ekonomicznej gospodarstw funkcjonujących w krajach UE z uwzględnieniem dopłat do działalności operacyjnej, a także wskazanie czynników, które oddziaływały na nadwyżkę finansową	kraje UE (na podstawie bazy FADN – zastosowano podział krajów ze względu na siłę ekonomiczną gospodarstw), okres: 2009-2015	makropanele, model o ustalonych efektach
Bereźnicka 2018b	źródła rentowności kapitału własnego w krajach o silnych oraz słabych ekonomicznie gospodarstwach rolnych	kraje UE, okres: 2009-2015	makropanele, model o ustalonych efektach
Olipra 2019	weryfikacja hipotezy zgodnie, z którą sezonowość nadal odgrywa ważną rolę w kształtowaniu cen skupu mleka w UE	kraje UE, okres: 1977-2017	makropanele, model o losowych efektach
Halamska, Lamarche, Maurel 2003	analiza zmian, jakie zaszły w rolnictwie rodzinnym w Polsce	gospodarstwa rodzinne, okres: 1988-2001	mikropanel
Grzelak, Wiktorowicz 2009	przedstawienie poziomu i struktury wsparcia publicznego udostępnionego polskiemu rolnictwu i obszarom wiejskim oraz próba oceny wpływu tego wsparcia na dochodowość gospodarstw w Polsce	gospodarstwa indywidualne (dane rachunkowości rolniczej prowadzonej w ramach FADN), okres: 2004-2007	mikropanele, model o ustalonych efektach
Franc-Dąbrowska 2009	determinanty kształtujące przychody ze sprzedaży w przedsiębiorstwach rolniczych	67 przedsiębiorstw rolniczych dobranych nieprobabilistycznie, okres: 2001-2007	model o losowych ustalonych efektach i z brakiem efektów indywidualnych
Czekaj 2011	badanie rynku kredytów w rolnictwie w Polsce po przystąpieniu do UE i wprowadzeniu wspólnej polityki rolnej (WPR)	zbilansowany panel gospodarstw, które w sposób nieprzerwany prowadziły rachunkowość w systemie polskiego FADN, okres: 2004-2007	mikropanele, model o ustalonych i losowych efektach
Dobrowolska 2012	analiza ekonomicznych konsekwencji opodatkowania polskich gospodarstw domowych podatkiem VAT w latach 1995–2009	próba gospodarstw domowych analizowanych przez GUS w ramach badań budżetów gospodarstw domowych, okres: 1995-2009	mikropanele, modele jedno i dwuczynnikowe
Stępień, Smędzik-Ambroży, Guth 2017	określenie wpływu WPR na poziom zrównoważenia ekonomiczno-społecznego gospodarstw rolnych w Polsce	gospodarstwa rolne (na podstawie bazy FADN), okres: 2004-2013	mikropanele, model o ustalonych efektach
Jaworski, Czerwonka, Mądra-Sawicka 2018	rozpoznanie relacji między rentownością a płynnością finansową przedsiębiorstw sektora przetwórstwa spożywczego	1046 podmiotów z branży spożywczej, okres: 2012-2015	mikropanele, model panelowy KMNK

Źródło: opracowanie własne

Literatura

- Bereźnicka J. (2018a) Dopłaty do działalności operacyjnej a stabilność finansowa gospodarstw rolnych w krajach Unii Europejskiej. *Wieś i Rolnictwo*, 3(180), 113-135.
- Bereźnicka J. (2018b) Źródła rentowności kapitału własnego w zróżnicowanym ekonomicznie rolnictwie krajów Unii Europejskiej. *Zagadnienia Ekonomiki Rolnej*, 356(3), 76-93.
- Breusch T.S., Pagan A.R. (1980) The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1), 239-253.
- Czapiński J., Panek T. (red.) *Diagnoza Społeczna*, www.diagnoza.com (dostęp 30.03.2021).
- Czekaj T. (2011) Zachowania inwestycyjne polskich gospodarstw rolnych. *Zagadnienia Ekonomiki Rolnej*, 4, 102-113.
- Czyżewski B., Kryszak Ł. (2017) Wpływ typów rolnictwa na emisję gazów cieplarnianych. *Wieś i Rolnictwo*, 1(174), 99-122.
- Czyżewski A., Kułyk P. (2014) Relacja ziemia-praca w warunkach finansowego wsparcia rolnictwa na przykładzie wybranych krajów świata i Unii Europejskiej-15 po 1986 r. *Problemy Rolnictwa Światowego*, 14(29), 2, 31-42.
- Czyżewski A., Kułyk P. (2017) Wpływ zmian w otoczeniu makroekonomicznym na rozwój rolnictwa w krajach wysokorozwiniętych w długim okresie. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 491, Współczesna ekonomia a rozwój zrównoważony, 86-96.
- Czyżewski B., Poczta-Wajda A. (2017) Wpływ polityki i rynku na relatywną depriwację dochodową czynnika pracy w rolnictwie. *Wieś i Rolnictwo*, 3(176), 53-70.
- Dańska-Borsiak B. (2011) *Dynamiczne modele panelowe w badaniach ekonomicznych*. Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Dobrowolska B. (2012) Zastosowanie modeli panelowych do badania redystrybucyjnych efektów opodatkowania konsumpcji polskich gospodarstw domowych podatkiem VAT w latach 1995-2009. *Annales Universitatis Mariae Curie-Skłodowska, Sectio H – Oeconomia*, 46(4), 85-94.
- Dudek H., Krawiec M., Koszela G. (2015) Are Changes in Food Consumption in the European Union Environmentally Friendly? *Scientific Journal Warsaw University of Life Sciences – SGGW, Problems of World Agriculture*, 15(30), 4, 33-40.
- Firlej K., Kubala S. (2018) Ceny ziemi rolnej w Polsce na tle Unii Europejskiej. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, 3(975), 159-171.
- Franc-Dąbrowska J. (2009) Praktyczne zastosowanie wybranych modeli panelowych do oceny sytuacji finansowej przedsiębiorstw rolniczych. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 76, 31-40.
- Geise A. (2013) Przestrzenno-czasowe modelowanie zmienności produkcji w sektorach mikro-, małych, średnich i dużych przedsiębiorstwach w Polsce. *Przegląd Statystyczny*, 60(2), 269-282.
- Grzelak A. (2018) Związki pomiędzy sytuacją ekonomiczną a oddziaływaniem środowiskowym gospodarstw rolnych w krajach Unii Europejskiej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 20(3), 59-64.
- Grzelak M., Wiktorowicz J. (2009) Ocena wsparcia publicznego rolnictwa w Polsce – wybrane zagadnienia. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 7(22), 21-31.
- Gołaś Z. (2017a) Rentowność produkcji pszenicy zwyczajnej w UE. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(32)(2), 29-40.
- Gołaś Z. (2017b) Uwarunkowania rentowności produkcji mleka w gospodarstwach mlecznych krajów Unii Europejskiej. *Zagadnienia Ekonomiki Rolnej*, 352(3), 19-40.
- Gołaś Z. (2019) Impact of stock management on the profitability of milk processing companies. *Annals of the Polish Association of Agricultural and Agrobusiness Economists*, 21(3), 82-92.
- Gruszczyński M. (red.) (2012) *Mikroekonometria. Modele i metody analizy danych indywidualnych*. Wyd. II. Wolters Kluwer Polska, Warszawa.

- Guth M., Smędzik-Ambroży K. (2017) Wpływ subsydiów wspólnej polityki rolnej na dochody gospodarstw mlecznych FADN w krajach Unii Europejskiej w latach 2004-2013. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego, 17(32), 1, 53-62.
- Halamska M., Lamarche H., Maurel M-C. (2003) Rolnictwo rodzinne w transformacji postkomunistycznej. Anatomia zmiany. Instytut Rozwoju Wsi i Rolnictwa PAN, Warszawa.
- Hausman J. (1978) Specification tests in econometrics. *Econometrica*, 46, 1251-1271.
- Jaworski J., Czerwonka L., Mądra-Sawicka M. (2018) Zależność między rentownością a płynnością finansową w sektorze przetwórstwa spożywczego w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 20(1), 58-63.
- Kisielińska J. (2008) Panelowe klasyfikacyjne modele upadłości ekonomicznej gospodarstw rolniczych. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego, 4(19), 259-269.
- Kufel J. (2014) Marże w przemyśle spożywczym krajów Unii Europejskiej a wahania koniunktury gospodarczej. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego, 14(29), 3, 130-139.
- Mundlak Y. (1961) Empirical Production Function Free of Management Bias. *Journal of Farm Economics*, 43(1), 44-56.
- Olipra J. (2019) Change in Seasonality Pattern of EU Farmgate Milk Price. *Scientific Journal Warsaw University of Life Sciences – SGGW, Problems of World Agriculture*, 19(34)(3), 75-84.
- Panel Study of Income Dynamics: <https://psidonline.isr.umich.edu> (dostęp 30.03.2021).
- Staniszewsk J., Czyżewski A. (2016) Zastosowanie regresji panelowej dla oceny produktywności i dochodowości w rolnictwie krajów UE po 2005 r. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(3), 7-21.
- Stępień S., Smędzik-Ambroży K., Guth M. (2017) Oddziaływanie Wspólnej Polityki Rolnej na zrównoważenie ekonomiczno-społeczne gospodarstw rolnych na przykładzie Polski. *Więś i Rolnictwo*, 4(177), 39-58.
- Wasilewska E., Pietrych Ł. (2018) Starzenie się społeczeństwa a wzrost gospodarczy w krajach Unii Europejskiej. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego, 18(33)(4), 481-492.

3. Analiza wariancji

Analiza wariancji (ang. *analysis of variance* – ANOVA) to technika statystyczna używana do sprawdzania, czy średnie dwóch lub więcej grup znacząco się od siebie różnią. ANOVA sprawdza wpływ jednego lub więcej czynników, porównując średnie z różnych próbek. Analiza wariancji ma szerokie zastosowanie w wielu dyscyplinach, w tym między innymi w rolnictwie oraz ekonomii i finansach. Modele ANOVA należą do klasy modeli liniowych odpowiednich do modelowania ciągłej zmiennej objaśnianej względem jednej lub kilku jakościowych zmiennych objaśniających, zwanych ogólnie czynnikami, które są mierzone na nominalnej lub porządkowej skali pomiaru. Rozszerzeniem jednowymiarowej analizy wariancji jest wielowymiarowa analiza wariancji (ang. *multivariate analysis of variance* – MANOVA). Procedurę MANOVA stosuje się w sytuacji, gdy zmiana kombinacji czynników wyjaśnia zmiany nie tylko jednej zmiennej objaśnianej, a kilku zmiennych objaśnianych jednocześnie. Metodę MANOVA wykorzystuje się zatem do przeprowadzenia testu na różnicę średnich między centroidami kilku populacji [Aczel 2005].

3.1. Opis metod

3.1.1. Jednoczynnikowa i dwuczynnikowa analiza wariancji

Analiza wariancji to parametryczna technika statystyczna używana do porównywania średnich wartości wybranych zbiorów danych. Metoda ta została wprowadzona przez Fishera [Fisher, Mackenzie 1923; Fisher 1925]. Początkowo analiza wariancji była stosowana przede wszystkim do rozwiązywania problemów w badaniach doświadczalnych w rolnictwie. Obecnie analizę wariancji wykorzystuje się w badaniach z zakresu psychologii, biologii, medycyny, ekonomii, finansów, politologii i innych [Rutherford 2011].

Analiza wariancji służy do określenia, czy istnieją istotne statystycznie różnice między średnimi wielu grup obserwacji. W analizie tej wykorzystuje się testy oparte na ilorazach wariancji. Jednoczynnikowa analiza wariancji dotyczy sytuacji, w której badamy wpływ jednego czynnika (zmiennej jakościowej) na zmienną zależną. Wówczas ogólna postać modelu ANOVA dla zmiennej losowej Y , której i -tą obserwację z j -tej grupy oznaczamy jako y_{ji} , jest następująca:

$$y_{ji} = \mu + \alpha_j + \varepsilon_{ji}, \quad (3.1)$$

gdzie μ oznacza średnią w całej populacji, α_j to odchylenie od μ spowodowane działaniem czynnika A na poziomie j ($j = 1, \dots, p$), ε_{ji} to odchylenie losowe o rozkładzie normalnym związane z i -tą obserwacją ($i = 1, \dots, n_j$) dla j -tego poziomu czynnika A .

Jednoczynnikowy model ANOVA można przedstawić także w postaci:

$$y_{ji} = \mu_j + \varepsilon_{ji}, \quad (3.2)$$

gdzie μ_j oznacza średnią w grupie j .

Analiza wariancji to parametryczne narzędzie, które opiera się na kilku podstawowych założeniach:

- analizowana zmienna Y jest zmienną ilościową,
- każda z p niezależnych populacji ma rozkład normalny $N(\mu_j, \sigma_j)$,
- średnie (μ_j) w p populacjach mogą, ale nie muszą być równe,
- wariancje w p populacjach są takie same ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = \sigma^2$).

Hipoteza zerowa zakłada, że wszystkie poziomy czynnika w jednakowy sposób wpływają na zmienną zależną Y . Oznacza to, że wszystkie średnie w p grupach są takie same. Hipotezy w teście ANOVA można zapisać następująco:

$$H_0: \forall j \alpha_j = 0, \quad (3.3)$$

$$H_1: \exists j \alpha_j \neq 0$$

lub

$$H_0: \forall j, l \mu_j = \mu_l, \quad (3.4)$$

$$H_1: \exists j, l \mu_j \neq \mu_l,$$

gdzie $j = 1, \dots, p$ oraz $l = 1, \dots, p$.

Łączne zróżnicowanie zmiennej zależnej Y (ang. *total sum of squares* – SST) jest sumą zróżnicowania międzygrupowego spowodowanego działaniem czynnika (ang. *sum of squares for treatment* – SSTR) oraz wewnątrzgrupowego spowodowanego czynnikami losowymi (ang. *sum of squares for errors* – SSE).

$$SST = SSTR + SSE, \quad (3.5)$$

$$\sum_{j=1}^p \sum_{i=1}^{n_j} (y_{ji} - \bar{y})^2 = \sum_{j=1}^p n_j (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^p \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2, \quad (3.6)$$

gdzie \bar{y} jest średnią ogólną z wszystkich wyników obserwacji traktowanych jako jeden zbiór danych, \bar{y}_j jest średnią wszystkich wyników obserwacji przy poziomie j czynnika A .

Zgodnie z podstawową zasadą ANOVA w sytuacji gdy średnie w grupach nie są równe, to przeciętne odchylenie losowe jest stosunkowo małe w porównaniu z przeciętnym odchyleniem α_j spowodowanym działaniem czynnika [Aczel 2005].

Statystyka testowa ma rozkład F o liczbach stopni swobody $p - 1$ w liczniku i $n - p$ w mianowniku, gdzie $n = n_1 + \dots + n_p$ oznacza liczebność próby, a p liczbę grup zmiennej losowej Y .

$$F_{(p-1, n-p)} = \frac{SSTR/p-1}{SSE/n-p}. \quad (3.7)$$

Statystyka F przyjmuje tym większe wartości, im większe jest zróżnicowanie międzygrupowe spowodowane działaniem wybranego czynnika w porównaniu do zróżnicowania wewnątrzgrupowego spowodowanego czynnikami losowymi. Obszar krytyczny testu F jest obszarem prawostronnym. W przypadku odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej można stwierdzić, że co najmniej dwie średnie w grupach różnią się od siebie, czyli czynnik A wywiera istotny wpływ na zmienną zależną y_{ji} .

Wieloczynnikowa analiza wariancji bada wpływ kilku czynników na zmienną zależną [Field 2005]. Ogólną postać modelu ANOVA dla dwóch czynników można zapisać następująco:

$$y_{jki} = \mu + \alpha_j + \beta_k + (\alpha, \beta)_{jk} + \varepsilon_{jki}, \quad (3.8)$$

gdzie μ oznacza średnią w całej populacji, α_j jest efektem działania czynnika A na poziomie j ($j = 1, \dots, p$), β_k jest efektem działania czynnika B na poziomie k ($k = 1, \dots, s$), $(\alpha, \beta)_{jk}$ to efekt interakcji czynników A oraz B , ε_{jki} to odchylenie losowe o rozkładzie normalnym związane z i -tą obserwacją ($i = 1, \dots, n_{jk}$) j -tego poziomu czynnika A oraz k -tego poziomu czynnika B . W przypadku dwuczynnikowej analizy wariancji (3.8) weryfikacji podlegają następujące hipotezy:

$$H_0: \forall j \alpha_j = 0, \quad (3.9)$$

$$H_1: \exists j \alpha_j \neq 0,$$

$$H_0: \forall k \beta_k = 0, \quad (3.10)$$

$$H_1: \exists k \beta_k \neq 0,$$

$$H_0: \forall jk (\alpha, \beta)_{jk} = 0, \quad (3.11)$$

$$H_1: \exists jk (\alpha, \beta)_{jk} \neq 0.$$

Hipoteza zerowa (3.9) jest prawdziwa wtedy i tylko wtedy, gdy nie występują różnice między średnimi zmiennej objaśnianej wynikające z wpływu czynnika A na j poziomach. Hipoteza zerowa (3.10) jest prawdziwa wtedy i tylko wtedy, gdy nie występują różnice między średnimi zmiennej objaśnianej wynikające z wpływu czynnika B na k poziomach. Hipoteza zerowa (3.11) jest prawdziwa wtedy i tylko wtedy, gdy nie ma powiązania między efektami oddziaływania czynników A i B , czyli wtedy gdy te efekty są addytywne [Aczel 2005].

W wieloczynnikowej analizie wariancji suma zróżnicowania międzygrupowego spowodowanego działaniem czynników (SSTR) obejmuje składniki odpowiadające działaniu poszczególnych

czynników oraz interakcji między czynnikami. Równania (3.12), (3.13), (3.14) przedstawiają podział całkowitej sumy kwadratów odchyłeń na składniki w przypadku wariancji dwuczynnikowej.

$$SST = \underbrace{SSA + SSB + SS(AB)}_{SSTR} + SSE, \quad (3.12)$$

$$\underbrace{\sum_{j=1}^p \sum_{k=1}^s \sum_{i=1}^{n_{jk}} (y_{jki} - \bar{y})^2}_{SST} = \underbrace{\sum_{j=1}^p \sum_{k=1}^s (\bar{y}_{jk} - \bar{y})^2}_{SSTR} + \underbrace{\sum_{j=1}^p \sum_{k=1}^s \sum_{i=1}^{n_{jk}} (y_{jki} - \bar{y}_{jk})^2}_{SSE}, \quad (3.13)$$

$$\underbrace{\sum_{j=1}^p \sum_{k=1}^s (\bar{y}_{jk} - \bar{y})^2}_{SSTR} = \underbrace{s \sum_{j=1}^p (\bar{y}_j - \bar{y})^2}_{SSA} + \underbrace{p \sum_{k=1}^s (\bar{y}_k - \bar{y})^2}_{SSB} + \underbrace{\sum_{j=1}^p \sum_{k=1}^s (\bar{y}_{jk} - \bar{y}_j - \bar{y}_k + \bar{y})^2}_{SS(AB)}, \quad (3.14)$$

gdzie SSA to suma kwadratów odchyłeń wynikająca z działania czynnika A , SSB to suma kwadratów odchyłeń wynikająca z działania czynnika B , $SS(AB)$ to suma kwadratów odchyłeń wynikająca z interakcji czynników A i B , \bar{y} jest średnią ogólną z wszystkich wyników obserwacji traktowanych jako jeden zbiór danych, y_{jki} jest i -tym wynikiem obserwacji przy poziomie j czynnika A oraz k czynnika B , \bar{y}_j jest średnią wszystkich wyników obserwacji przy poziomie j czynnika A , \bar{x}_k jest średnią wszystkich wyników obserwacji przy poziomie k czynnika B , \bar{x}_{jk} jest średnią wszystkich wyników obserwacji przy poziomie j czynnika A oraz k czynnika B .

Iloraz F dla każdej z testowanych hipotez, (3.9), (3.10), (3.11), jest ilorazem średniego kwadratu odchyłeń spowodowanych działaniem czynników oraz średniego kwadratu błędów:

$$F_{(p-1, ps(n-1))} = \frac{SSA/p-1}{SSE/ps(n-1)}, \quad (3.15)$$

$$F_{(p-1, ps(n-1))} = \frac{SSA/p-1}{SSE/ps(n-1)}, \quad (3.16)$$

$$F_{((p-1)(s-1), ps(n-1))} = \frac{SS(AB)/(p-1)(s-1)}{SSE/ps(n-1)}. \quad (3.17)$$

W celu weryfikacji hipotezy (3.9) dotyczącej wpływu czynnika A na zmienną zależną wykorzystuje się statystykę testową, która ma rozkład F o $(p - 1)$ stopniach swobody w liczniku i $ps(n - 1)$ stopniach swobody w mianowniku (3.15). Analogicznie przy sprawdzeniu czy występują efekty swoiste czynnika B (hipoteza 3.10), statystyką testową jest iloraz F (3.16). Przy weryfikacji występowania interakcji (hipoteza 3.11) sprawdzianem jest iloraz F (3.17).

Wyniki testu F wskazują, że istnieją co najmniej dwie średnie, które różnią się istotnie od siebie, nie wiadomo jednak których grup to dokładnie dotyczy. Z tego też względu analiza wariancji jest na ogół uzupełniana o tzw. testy *post-hoc*, zwane także testami porównań parami czy też porównań wielokrotnych. Testy te dzielą się na dwie podstawowe grupy, czyli testy dla równych wariancji i testy dla nierównych wariancji. Jednym z najstarszych testów *post-hoc* jest wprowadzony

przez Fishera [1935] test najmniejszych istotnych różnic (NIR) (*Fisher's least significant difference* – LSD). Test ten polega na przeprowadzeniu testu t-Studenta dla każdej porównywanej pary średnich. Procedurę w teście NIR dla wariancji jednoczynnikowej można opisać następująco:

- Wybranie pary średnich \bar{y}_j i \bar{y}_l gdzie $j = 1, \dots, p$, $l = 1, \dots, p$ oraz $j \neq l$.
- Wyznaczenie najmniej istotnej różnicy (NIR), jaką należy przyjąć, aby uznać, że średnie w populacjach (μ_j, μ_l) są istotnie różne:

$$F_{(p-1, ps(n-1))} = \frac{SSA/p-1}{SSE/ps(n-1)}. \quad (3.18)$$

- Porównanie wszystkich par średnich z próby. Jeżeli $[\bar{y}_j - \bar{y}_l] \geq \text{NIR}$, to uznaje się, że średnie są istotnie różne.
- Ze względu na kumulację prawdopodobieństw popełnienia błędu pierwszego rodzaju test NIR wykonywany jest zazwyczaj tylko pomocniczo w celu potwierdzenia wyników innych testów *post-hoc*.

Jednym z najpopularniejszych testów *post-hoc* jest test uczciwie istotnych różnic (ang. *honestly significant difference* – HSD) Tukeya [Tukey 1953]. Pozwala on na porównanie dowolnej pary średnich przy ustalonym dla wszystkich porównań poziomie istotności. Procedura testu HSD Tukeya jest analogiczna do procedury testu NIR. Testy te różnią się obliczoną wartością krytyczną. W teście najmniejszych istotnych różnic była to statystyka F (3.18), w teście uczciwie istotnych różnic Tukeya statystyka T :

$$T = q_{p, n-p, \alpha} \sqrt{\frac{SSE}{n-p} \left(\frac{1}{n_j} \right)}, \quad (3.19)$$

gdzie $q_{p, n-p, \alpha}$ jest odpowiednim kwantylem studentyzowanego rozstępu przy p i $n - p$ stopniach swobody oraz poziomie istotności α .

W przypadku kiedy grupy nie są równoliczne, można zastosować test *post-hoc* Tukeya-Kramera [Kramer 1956; Hayter 1984; Driscoll 1996] i wówczas zamiast statystyki T wykorzystuje się statystykę TK :

$$TK = q_{p, n-p, \alpha} \sqrt{\frac{SSE}{n-p} \frac{1}{2} \left(\frac{1}{n_j} + \frac{1}{n_l} \right)}. \quad (3.20)$$

Innymi testami *post-hoc* dla równych wariancji stosowanymi w przypadku nierównolicznych grup są testy Bonferroniego [Miller 1966] i Scheffého [Scheffé 1953, 1977]. Test Bonferroniego jest bardzo zbliżony do testu najmniejszych istotnych różnic, jednak w porównaniu do testu NIR pozwala on na porównanie dowolnej pary średnich przy poziomie istotności α/D , gdzie D oznacza liczbę wykonanych porównań. Test Scheffého jest uznawany za jeden z najbardziej konserwatywnych. W jego przypadku liczba grup różniących się istotnie od siebie będzie zazwyczaj mniejsza niż w przypadku innych testów [Schwertman, Carter 1995]. Stoline [1981] rekomenduje stosowanie

testu *post-hoc* Tukeya-Kramera spośród opisanych powyżej testów. Podkreśla on jednak, że test ten nie jest odpowiedni w przypadku występowania heteroskedastyczności. Do najbardziej popularnych testów porównań parami dla nierównych wariancji zalicza się między innymi test Gamesa-Howella [Games, Howell 1976] oraz test Gabriela [Gabriel 1978]. Przeglądu testów *post-hoc* dokonali między innymi Carmer i Swanson [1973], Jaccard, Becker i Wood [1984] oraz Demirhan i in. [2010].

3.1.2. Wielowymiarowa analiza wariancji

Wielowymiarową analizę wariancji (ang. *multivariate analysis of variance* – MANOVA) stosuje się do jednoczesnego badania powiązań między jedną lub wieloma zmiennymi niezależnymi typu jakościowego oraz dwoma lub większą liczbą zmiennych zależnych typu ilościowego. Koncepcja MANOVA została wprowadzona przez Wilksa [Wilks 1932]. W przypadku wielowymiarowej analizy wariancji weryfikowana hipoteza dotyczy efektów grupowych dla kombinacji badanych zmiennych zależnych. W przeciwieństwie do jednowymiarowej analizy ANOVA, gdzie porównuje się średnie między grupami, w analizie MANOVA porównuje się wektory średnich (centroidy grup) dla dwóch lub większej liczby zmiennych zależnych.

Niech n_j stanowi liczbę obiektów w j -tej grupie, gdzie $n = n_1 + \dots + n_p$. Jednoczynnikowy model MANOVA można przedstawić w postaci:

$$\mathbf{y}_{ji} = \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_{ji}, \quad (3.21)$$

gdzie \mathbf{y}_{ji} jest wektorem zmiennych zależnych, $\boldsymbol{\mu}_j$ jest wektorem wartości średnich w j -tej grupie dla r zmiennych zależnych, $\boldsymbol{\varepsilon}_{ji}$ to wektor błędów. W analizie MANOVA weryfikacji podlegają następujące hipotezy:

$$H_0: \forall j, l \boldsymbol{\mu}_j = \boldsymbol{\mu}_l, \quad (3.22)$$

$$H_1: \exists j, l \boldsymbol{\mu}_j \neq \boldsymbol{\mu}_l,$$

gdzie $j = 1, \dots, p$; $l = 1, \dots, p$; $\boldsymbol{\mu}_j^T = [\mu_{j1} \ \mu_{j2} \ \dots \ \mu_{jr}]$. Wektor $\boldsymbol{\mu}_j^T$ stanowi transponowaną macierz $\boldsymbol{\mu}_j$ wartości średnich w j -tej grupie dla r badanych zmiennych (centroida j -tej populacji). MANOVA jest zatem testem równości wielowymiarowych wartości średnich p populacji. Założenia wielowymiarowej analizy wariancji stanowią rozszerzenie założeń jednowymiarowej analizy wariancji. W MANOVA przyjmuje się, że:

- każda z p niezależnych populacji ma wielowymiarowy rozkład normalny,
- wszystkie p populacje mają takie same macierze wariancji/kowariancji.

W przypadku wielowymiarowej analizy wariancji zależność (3.5) zapisuje się jako:

$$\mathbf{S}_T = \mathbf{S}_A + \mathbf{S}_e, \quad (3.23)$$

gdzie \mathbf{S}_T to macierz całkowitej sumy kwadratów i iloczynów mieszanych, \mathbf{S}_A to macierz sumy kwadratów i iloczynów mieszanych między p grupami, \mathbf{S}_e to macierz kwadratów i iloczynów mieszanych dla błędów losowych.

Popularną statystyką testową wykorzystywaną w wielowymiarowej analizie wariancji jest statystyka lambda Wilksa (Λ):

$$\Lambda = \frac{|\mathbf{S}_e|}{|\mathbf{S}_T|}, \quad (3.24)$$

gdzie $|\mathbf{S}_e|$ to wyznacznik macierzy \mathbf{S}_e , $|\mathbf{S}_T|$ to wyznacznik macierzy \mathbf{S}_T . Wyznacznik macierzy wewnątrzgrupowej sum kwadratów i iloczynów mieszanych dla p grup ($|\mathbf{S}_e|$) obrazuje zmienność wewnątrzgrupową. Wyznacznik całkowitej macierzy sum kwadratów oznacza zmienność całkowitą ($|\mathbf{S}_T|$). Gdy wektory średnich w p populacjach nie są równe, to wówczas zmienność wewnątrzgrupowa będzie mała w porównaniu ze zmiennością całkowitą. Z powyższego wynika, że współczynnik lambda Wilksa (Λ) będzie przyjmował małe wartości w sytuacji, gdy hipoteza zerowa (3.22) będzie odrzucona. Z uwagi na to, że rozkład statystyki lambda Wilksa (Λ) jest skomplikowany, na ogół dokonuje się transformacji statystyki (Λ) w statystyki mające rozkład F [Aczel 2005].

Poniżej przedstawiono równanie macierzy całkowitej sum kwadratów i iloczynów mieszanych dla wielowymiarowej dwuczynnikowej analizy wariancji:

$$\mathbf{S}_T = \mathbf{S}_A + \mathbf{S}_B + \mathbf{S}_{AB} + \mathbf{S}_e. \quad (3.25)$$

W przypadku wielowymiarowej dwuczynnikowej analizy wariancji mianownik w statystyce lambda Wilksa (Λ) przyjmuje różne wartości w zależności od tego, jaka hipoteza zerowa podlega weryfikacji. Współczynnik lambda Wilksa (Λ) dla hipotezy zerowej dotyczącej efektu głównego działania czynnika A będzie zawierał w mianowniku wyznacznik sumy macierzy \mathbf{S}_A oraz \mathbf{S}_e ($|\mathbf{S}_A + \mathbf{S}_e|$). Wyznacznik $|\mathbf{S}_B + \mathbf{S}_e|$ służy do obliczania wartości współczynnika lambda Wilksa (Λ) dla efektu głównego czynnika B, natomiast wyznacznik $|\mathbf{S}_{AB} + \mathbf{S}_e|$ dla efektu interakcji obu czynników. Wśród innych testów stosowanych w wielowymiarowej analizie wariancji wyróżnia się między innymi test Pillai i Barletta [Barlett 1939; Pillai 1955], test Hotellinga i Lawleya [Lawley 1938, 1939; Hotelling 1951] oraz test Roya [1945]. Przeglądu testów wykorzystywanych w MANOVA dokonali między innymi Pillai i Jayachandran [1967], Finch [2005], Van Aelst i Willems [2011], Finch i French [2013].

3.2. Zastosowania analizy wariancji

Zaproponowana przez brytyjskiego statystyka i genetyka Ronalda Fishera analiza wariancji znalazła swoje szerokie zastosowanie w doświadczałnictwie rolniczym [Fisher 1923; Mead, Riley 1981; Anderson, Liberta 1986; Martiniello 1988; Andow, Hidaka 1989; Anderson, Domsch 1990; Maxwell 1995; Mądry 1996; Pietrzykowski 2005; Śmiałowski i in. 2011]. Jednowymiarowa i wielowymiarowa analiza wariancji wykorzystywana jest w literaturze także do badania kwestii z zakresu makroekonomii, ekonomii behawioralnej, finansów przedsiębiorstw, bankowości, finansów publicznych.

Schultz i Snedecor [1933] jako jedni z pierwszych pokazali zastosowanie jednowymiarowej analizy wariancji w ekonomii. Tan, Koh i Low [1997] przy użyciu analizy wariancji wykazali, że średnie wartości wskaźników finansowych przedsiębiorstw z różnych branż istotnie różnią się od siebie. Dobrowolska [2016] na podstawie jednoczynnikowej analizy wariancji oraz testu *post-hoc* Bonferro-niego dokonała oceny zróżnicowania obciążeń polskich gospodarstwa domowych podatkami pośrednimi. Wykazała ona, że podatki pośrednie w Polsce relatywnie najbardziej obciążają konsumpcję gospodarstw o najniższych przychodach. Ślusarczyk i Grondys [2018] zastosowali analizę wariancji do oceny koncepcji specjalnych stref ekonomicznych z uwzględnieniem kryteriów ekonomicznych, społecznych i środowiskowych. Wykazali oni, że gminy funkcjonujące na terenie specjalnych stref ekonomicznych istotnie różnią się od pozostałych gmin pod względem analizowanego zakresu. Wielechowski, Czech i Grzęda [2020] przy użyciu analizy wariancji dokonali oceny wpływu pandemii COVID-19 na zmiany mobilności ludzi w transporcie publicznym w Polsce.

Zastosowanie wielowymiarowej analizy wariancji w badaniach z zakresu ekonomii i finansów zaprezentowali między innymi Bracker i Pearson [1986], Meric, Leveen i Meric [1991], Lopez-Gracia i Aybar-Arias [2000]. Bracker i Pearson [1986] na podstawie analizy MANOVA wykazali istnienie istotnych statystycznie różnic między danymi dotyczącymi wyników finansowych firm, które wdrażają planowanie strategiczne i tych, które tego nie robią. Meric, Leveen i Meric [1991] zajmowali się problemem przejęć w sektorze bankowym. Na podstawie wielowymiarowej analizy wariancji pokazali oni, że banki o podobnej wielkości różnią się istotnie w zakresie podstawowych charakterystyk finansowych w zależności od roli, jaką odgrywają w procesie przejęcia. Lopez-Gracia i Aybar-Arias [2000] za pomocą analizy MANOVA przedstawili wpływ wybranych ekonomicznych czynników na strategię finansowania małych i średnich przedsiębiorstw. Wykazali oni, że wielkość przedsiębiorstwa wpływa na strategię samofinansowania przedsiębiorstw, natomiast sektor biznesowy wpływa na krótkoterminową politykę finansową.

Jednowymiarowa oraz wielowymiarowa analiza wariancji stosowana jest także w badaniach związanych z ekonomicznymi uwarunkowaniami rozwoju sektora rolno-spożywczego. W tabeli 3.1 przedstawiono wybrane pozycje literatury, w których wykorzystano analizę wariancji do badań z zakresu ekonomiki rolnictwa.

W literaturze przedmiotu z zakresu ekonomiki rolnictwa analiza wariancji wykorzystywana jest przede wszystkim do badania społeczno-ekonomicznego zróżnicowania gospodarstw rolnych, gospodarstw domowych, przedsiębiorstw, jednostek samorządu terytorialnego oraz państw. Zróżnicowanie gospodarstw rolnych badali między innymi Bratka i Praulins [2009], Czyżewski i Matuszczak [2009] oraz Hampł [2020]. Bratka i Praulins [2009] zastosowali dwuczynnikową analizę wariancji do badania zróżnicowania sytuacji finansowej polskich i łotewskich gospodarstw rolniczych. Wyniki przeprowadzonej przez nich analizy porównawczej wskazują na istnienie istotnych różnic między

relacjami zadłużenia do kapitału własnego oraz między sumą zobowiązań wśród badanych gospodarstw rolniczych. Czyżewski i Matuszczak [2009] na podstawie wielowymiarowej analizy wariancji wskazali na istotne zróżnicowanie alokacji czynników wytwórczych w kluczowych typach produkcyjnych w Polsce. Ich zdaniem zróżnicowanie to spowodowane jest oddziaływaniem struktury instytucjonalnej, w której funkcjonuje gospodarstwo rolne. Hampl [2020] przy użyciu wielowymiarowej analizy wariancji badał zróżnicowanie wyników finansowych gospodarstw rolnych ekologicznych i konwencjonalnych w Czechach. Jego zdaniem wybrany system rolniczy ma istotny wpływ na kształtowanie się rentowności sprzedaży, wybranych wskaźników kosztów oraz wskaźnika rotacji majątku.

Makhura, Kirsten i Mathye [1999] oraz Gutkowska i Piekut [2016] zastosowali analizę wariancji do badania wybranych zagadnień ekonomicznych dotyczących gospodarstw domowych. Makhura Kirsten i Mathye [1999] na podstawie wielowymiarowej analizy wariancji badali powiązania między warunkami panującymi na rynkach rolnych a zaangażowaniem w działalność pozarolniczą. Ich zdaniem udział w pozarolniczej działalności gospodarczej jest odpowiedzią na nieadekwatność działalności rolniczej do generowania wystarczających dochodów i zapewnienia bezpieczeństwa żywnościowego dla gospodarstwa domowego. Gutkowska i Piekut [2016] na podstawie analizy wariancji dokonały oceny zmian wydatków na usługi gastronomiczne w polskich gospodarstwach domowych. W badaniu ilościową zmienną zależną były wydatki na usługi gastronomiczne, natomiast jakościową zmienną niezależną były socjogeograficzne i ekonomiczne czynniki takie jak poziom wykształcenia głowy gospodarstwa domowego czy lokalizacja gospodarstwa domowego.

Badaniem zróżnicowania jednostek samorządu terytorialnego zajmowali się między innymi Standar [2013] i Sroka [2018]. Standar [2013] przy użyciu jednoczynnikowej analizy wariancji oraz testu *post-hoc* Tukeya badał zróżnicowanie sytuacji finansowej gmin wiejskich i miejskich w województwie wielkopolskim. Zgodnie z wynikami jego badań ośrodki wiejskie charakteryzują się mniejszą samodzielnością finansową i najwyższą dynamiką zadłużenia. Sroka [2018] zajmował się określeniem skali konwersji użytków rolnych. Wykazał on, że odległość gminy od rdzenia obszaru metropolitalnego istotnie różnicuje przebieg procesów konwersji, a dobre warunki przyrodnicze produkcji rolnej nie ograniczają procesu wyłączenia ich z produkcji.

Analiza wariancji stosowana jest także do badania zagadnień makroekonomicznych, w których jakościowe czynniki społeczno-ekonomiczne wpływają na zróżnicowanie wskaźników makroekonomicznych w danym państwie lub grupie państw. Czech i Imbeah [2019] wykorzystali jednoczynnikową analizę wariancji do przedstawienia wpływu uruchomienia komercyjnej produkcji i eksportu ropy naftowej w Ghanie na sytuację gospodarczą tego kraju. Wykazali oni, że komercyjna produkcja ropy naftowej spowodowała znaczący spadek udziału sektora rolnego w wartości dodanej brutto Ghany. Dodatkowo Czech i Imbeah [2019] pokazali, że rozwój branży wydobywczej przyczynił się do spadku średniego udziału eksportu produktów rolno-spożywczych w relacji do całkowitego

eksportu towarów. Badaniem krajów bogatych w ropę naftową zajmowali się także Niftiyev i Czech [2020]. Na podstawie jednoczynnikowej analizy wariancji wskazali oni na istotne zróżnicowanie wartości eksportu warzyw w Azerbejdżanie, w zależności od aktualnego stopnia uzależnienia kraju od branży wydobywczej.

W przypadku niespełnienia podstawowych założeń analizy wariancji w literaturze przedmiotu stosowany jest często jej nieparametryczny odpowiednik tzw. test Kruskala-Wallisa [Kruskal 1952; Kruskal, Wallis 1952]. Test ten stosuje się do weryfikacji hipotezy zakładającej, że mediany w badanych populacjach są równe. W literaturze przedmiotu z zakresu ekonomiki rolnictwa test Kruskala-Wallisa zastosowali między innymi Goraj i Mańko [2011], Domańska i Felczak [2014], Kuś i Pawlik [2016], Matyja [2017a], Nowacki i Fandrejewska [2017], Średzińska [2018]. Goraj i Mańko [2011] zajmowali się badaniem czynników różnicujących koszty pracy najemnej i czynsz dzierżawny w gospodarstwach rolniczych prowadzących książki rachunkowe w ramach polskiego FADN. Domańska i Felczak [2014] na podstawie testu Kruskala-Wallisa wykazali, że przedział wielkości ekonomicznej gospodarstw rolnych istotnie różnicuje wartość wydatków inwestycyjnych ponoszonych przez zarządzających gospodarstwami. Matyja [2017a] w wyniku analizy przeprowadzonej przy użyciu testu Kruskala-Wallisa stwierdziła, że istnieje istotne zróżnicowanie w rentowności spółdzielni rolniczych i rentowności innych form rolniczej działalności gospodarczej w Polsce. Kuś i Pawlik [2016] przy użyciu testu Kruskala-Wallisa wykazały, że strategia gospodarowania kapitałem obrotowym istotnie różnicuje wybrane miary finansowe przedsiębiorstw z sektora spożywczego w Polsce. Średzińska [2018] badała zróżnicowanie dochodów uzyskiwanych przez rolników w gospodarstwach rolnych krajów Unii Europejskiej (UE) z różnych klas wielkości ekonomicznej. Wykazała ona, że wielkość ekonomiczna wpływa na dochodowość pracy własnej, szczególnie w gospodarstwach większych ekonomicznie.

Podsumowując, analiza wariancji ma szerokie zastosowanie w różnych dyscyplinach, w tym między innymi w rolnictwie oraz ekonomii i finansach. W literaturze przedmiotu z zakresu ekonomiki rolnictwa jednowymiarowa i wielowymiarowa analiza wariancji stosowana jest do oceny zróżnicowania gospodarstw rolnych, przedsiębiorstw z sektora rolno-spożywczego, jednostek samorządu terytorialnego, państw pod względem czynników ekonomiczno-społecznych, w tym w dużej mierze pod względem osiągniętych przez nich wyników finansowych. W sytuacji, gdy nie są spełnione podstawowe założenia analizy wariancji, stosowany jest zazwyczaj test Kruskala-Wallisa, który stanowi nieparametryczny odpowiednik jednoczynnikowej analizy wariancji ANOVA.

Tabela 3.1. Przykłady zastosowań analizy wariancji w ekonomice rolnictwa

Autor	Problem	Obiekty	Metoda
Maxwell 1995	rolnictwo miejskie jako alternatywna strategia bezpieczeństwa żywnościowego	gospodarstwa domowe	jednoczynnikowa analiza wariancji
Makhura, Kirsten, Mathye 1999	powiązania między warunkami panującymi na rynkach rolnych a zaangażowaniem w działalność pozarolniczą	gospodarstwa domowe	wielowymiarowa analiza wariancji
Gellynck, Vermeire, Viaene 2006	rola sieci regionalnych w podnoszeniu kompetencji w zakresie innowacji na przykładzie przedsiębiorstw z branży spożywczej	przedsiębiorstwa	jednoczynnikowa analiza wariancji
Bratka, Praulins 2009	zróżnicowanie polskich i łotewskich gospodarstw rolnych ze względu na ich typ i wielkość ekonomiczną	gospodarstwa rolne	dwuczynnikowa analiza wariancji
Czyżewski, Matuszczak 2009	alokacja zasobów w indywidualnych gospodarstwach rolnych	gospodarstwo rolne	wielowymiarowa analiza wariancji
Kronen, Vunisea 2009	wpływ połowów na bezpieczeństwo żywnościowe	rybacy	jednoczynnikowa analiza wariancji
Ali, Kapoor, Moorthy 2010	zachowanie konsumentów na rynku produktów żywnościowych w Indiach	konsumenci	jednoczynnikowa analiza wariancji
Kühne i in. 2010	postrzeganie przez konsumentów innowacji w tradycyjnych produktach spożywczych w Europie	konsumenci	nieparametryczny test rang Kruskala-Wallis
Goraj, Mańko 2011	czynniki różnicujące koszty pracy najemnej i czynsz dzierżawny w gospodarstwach rolnych	gospodarstwa rolne	nieparametryczny test rang Kruskala-Wallis
Schiefer i in. 2013	czynniki determinujące rentowność przedsiębiorstw przemysłu spożywczego	przedsiębiorstwa	jednoczynnikowa analiza wariancji
Standar 2013	zróżnicowanie sytuacji finansowej gmin wiejskich i miejskich w województwie wielkopolskim	gminy	jednoczynnikowa analiza wariancji, HSD Tukeya <i>post-hoc</i> test
Domańska, Felczak 2014	czynniki różnicujące wydatki inwestycyjne w gospodarstwach rolnych z regionu Mazowsza i Podlasia	gospodarstwa rolne	nieparametryczny test rang Kruskala-Wallis
Lewandowska 2014	finansowe i rynkowe bariery innowacyjności przedsiębiorstw z branży spożywczej	przedsiębiorstwa	jednoczynnikowa analiza wariancji, Bonferroni <i>post-hoc</i> test
Gutkowska, Piekut 2016	czynniki determinujące zmiany wydatków na usługi gastronomiczne w polskich gospodarstwach domowych	gospodarstwa domowe	jednoczynnikowa analiza wariancji, Scheffe <i>post-hoc</i> test
Kuś, Pawlik 2016	strategia gospodarowania kapitałem obrotowym jako czynnik różnicujący wyniki finansowe spółek z branży spożywczej	przedsiębiorstwa	nieparametryczny test rang Kruskala-Wallis
Matyja 2017a	ocena zróżnicowania rentowności spółdzielni rolniczych i innych form rolniczej działalności gospodarczej w Polsce	spółdzielnie rolnicze, gospodarstwa rolne	nieparametryczny test rang Kruskala-Wallis
Matyja 2017b	zasoby pracy w polskich gospodarstwach z perspektywy pracowniczych spółdzielni rolniczych	spółdzielnie rolnicze	nieparametryczny test rang Kruskala-Wallis

Tabela 3.1 (cd.)

Autor	Problem	Obiekty	Metoda
Nowacki, Fandrejewska 2017	postawy Polaków wobec zjawiska dopasowania kulturowego przekazów reklamowych przedsiębiorstw zagranicznych na polskim rynku artykułów żywnościowych	konsumenci	nieparametryczny test rang Kruskala-Wallisa
Utnik-Banaś 2017	zmienność cen mięsa kurcząt brojlerów w krajach UE	państwa	jednoczynnikowa analiza wariancji
Sroka 2018	konwersja gruntów rolnych na cele pozarolnicze	gminy	jednoczynnikowa analiza wariancji, HSD Tukeya <i>post-hoc</i> test
Średzińska 2018	zróźnicowanie poziomu dochodu rolników w gospodarstwach z różnych klas wielkości ekonomicznej	gospodarstwa rolne	nieparametryczny test rang Kruskala-Wallisa
Czech, Imbeah 2019	ocena wpływu rozwoju branży wydobywczej w Ghanie na sektor rolny tego kraju	państwo	jednoczynnikowa analiza wariancji, HSD Tukeya <i>post-hoc</i> test
Majka, Zając 2019	czynniki kształtujące rozwój pozarolniczej działalności	gminy	jednoczynnikowa analiza wariancji
Blekking i in. 2020	formy zatrudnienia a bezpieczeństwo żywnościowe w miastach Afryki Subsaharyjskiej	gospodarstwa domowe	jednoczynnikowa analiza wariancji
de Paulo Farias, de Araújo 2020	wpływ pandemii COVID-19 na podaż żywności w centrach dystrybucji w wybranych regionach Brazylii	regiony	jednoczynnikowa analiza wariancji, HSD Tukeya <i>post-hoc</i> test
HAMPL 2020	zróźnicowanie wyników finansowych gospodarstw rolnych ekologicznych i konwencjonalnych w Czechach	gospodarstwa rolne	wielowymiarowa analiza wariancji
Niftiyev, Czech 2020	ocena wpływu branży wydobywczej na eksport warzyw w Azerbejdżanie	państwo	jednoczynnikowa analiza wariancji, HSD Tukeya <i>post-hoc</i> test

Źródło: opracowanie własne

Literatura

- Aczel A.D. (2005) Statystyka w zarządzaniu. Wydawnictwo Naukowe PWN, Warszawa.
- Ali J., Kapoor S., Moorthy J. (2010) Buying behaviour of consumers for food products in an emerging economy. *British Food Journal*, 112(2), 109-124.
- Anderson R.C., Liberta A.E. (1986) Occurrence of fungal-inhibiting *Pseudomonas* on caryopses of *Tripsacum dactyloides* L. and its implication for seed survival and agriculture application. *Journal of Applied Bacteriology*, 61(3), 195-199.
- Anderson T.H., Domsch K.H. (1990) Application of eco-physiological quotients (qCO_2 and qD) on microbial biomasses from soils of different cropping histories. *Soil Biology and Biochemistry*, 22(2), 251-255.
- Andow D.A., Hidaka K. (1989) Experimental natural history of sustainable agriculture: syndromes of production. *Agriculture, Ecosystems & Environment*, 27(1-4), 447-462.
- Bartlett M.S. (1939) A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(2), 180-185.
- Blekking J., Waldman K., Tuholske C., Evans T. (2020) Formal/informal employment and urban food security in Sub-Saharan Africa. *Applied Geography*, 114, 102-131.
- Bracker J.Y., Pearson J.N. (1986) Planning and financial performance of small, mature firms. *Strategic Management Journal*, 7(6), 503-522.
- Bratka V., Praulins A. (2009) Diversity of farm indebtedness in Latvia and Poland: a comparative study. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 6(21), 10-25.
- Carmer S.G., Swanson M.R. (1973) An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association*, 68(341), 66-74.
- Czech K., Imbeah N. (2019) Crude oil export of Ghana and its impact on the economy. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 21(4), 54-63.
- Czyżewski B., Matuszczak A. (2009) Alokacja zasobów w indywidualnych gospodarstwach rolnych w Polsce w świetle ich związków instytucjonalnych. *Więś i Rolnictwo*, 3(144), 29-48.
- Demirhan H., Dolgun N.A., Demirhan Y.P., Dolgun M.Ö. (2010) Performance of some multiple comparison tests under heteroscedasticity and dependency. *Journal of Statistical Computation and Simulation*, 80(10), 1083-1100.
- Dobrowolska B. (2016) Analiza wariancji – narzędzie do badań opodatkowania konsumpcji. *Annales Universitatis Mariae Curie-Skłodowska, Sectio H – Oeconomia*, 50(4), 69-78.
- Domańska T., Felczak T. (2014) Czynniki różnicujące wydatki inwestycyjne w gospodarstwach rolniczych. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 11(60), 46-55.
- Driscoll W.C. (1996). Robustness of the ANOVA and Tukey-Kramer statistical tests. *Computers & Industrial Engineering*, 31(1-2), 265-268.
- Field A. (2005) *Discovering statistics using SPSS*. SAGE Publications, London.
- Finch H. (2005) Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology*, 1(1), 27-38.
- Finch H., French B. (2013) A Monte Carlo comparison of robust MANOVA test statistics. *Journal of Modern Applied Statistical Methods*, 12(2), 35-81.
- Fisher R.A. (1925) *Statistical methods for research workers*. Oliver & Boyd, Edinburgh.
- Fisher R.A. (1935) *The design of experiment*. Oliver & Boyd, Edinburgh.
- Fisher R.A., Mackenzie W.A. (1923) Studies in crop variation. The manorial response of different potato varieties. *Journal of Agricultural Science*, 13, 311-320.
- Gabriel K.R. (1978) Power differences between pairwise multiple comparisons: comment. *Journal of the American Statistical Association*, 73(363), 485-487.

- Games P.A., Howell J.F. (1976) Pairwise multiple comparison procedures with unequal N's and/or variances: a Monte Carlo study. *Journal of Educational Statistics*, 1(2), 113-125.
- Gellynck X., Vermeire B., Viaene J. (2006) Innovation in the food sector: Regional networks and internationalisation. *Journal on Chain and Network Science*, 6(1), 21-30.
- Goraj L., Mańko S. (2011) Model szacowania pełnych kosztów działalności gospodarstw rolnych. *Zagadnienia Ekonomiki Rolnej*, 3(328), 28-58.
- Gutkowska K., Piekut M. (2016) Korzystanie z usług gastronomicznych przez Polaków. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 16(65), 15-24.
- HAMPL F. (2020) A statistical analysis of the financial performance of organic and conventional farms in the Czech Republic with respect to their size. *Agricultural Economics*, 66(1), 1-9.
- Hayter A.J. (1984) A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *The Annals of Statistics*, 12(1), 61-75.
- Hotelling H. (1951) A generalized T test and measure of multivariate dispersion, [w:] *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. The Regents of the University of California, California.
- Jaccard J., Becker M.A., Wood G. (1984) Pairwise multiple comparison procedures: a review. *Psychological Bulletin*, 96(3), 589-596.
- Kramer C.Y. (1956) Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 12(3), 307-310.
- Kronen M., Vunisea A. (2009) Fishing impact and food security—Gender differences in finfisheries across Pacific Island countries and cultural groups. *SPC Women in Fisheries Information Bulletin*, 19, 3-10.
- Kruskal W.H. (1952) A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, 23(4), 525-540.
- Kruskal W.H., Wallis W.A. (1952) Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621.
- Kühne B., Vanhonacker F., Gellynck X., Verbeke W. (2010) Innovation in traditional food products in Europe: Do sector innovation activities match consumers' acceptance? *Food Quality and Preference*, 21(6), 629-638.
- Kuś A., Pawlik M. (2016) Wybrane miary finansowe a strategia gospodarowania kapitałem obrotowym na przykładzie przedsiębiorstw przemysłu spożywczego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 18(3), 217-222.
- Lawley D.N. (1938) A generalization of Fisher's z test. *Biometrika*, 30(1/2), 180-187.
- Lawley D.N. (1939) A Correction to "A Generalization of Fisher's z Test". *Biometrika*, 30(3/4), 467-469.
- Lewandowska M. (2014) Innovation barriers and international competitiveness of enterprises from Polish food processing industry: Research results. *Acta Scientiarum Polonorum, Oeconomia*, 13(4), 103-113.
- Lopez-Gracia J., Aybar-Arias C. (2000) An empirical approach to the financial behaviour of small and medium sized companies. *Small Business Economics*, 14(1), 55-63.
- Majka A., Zając D. (2019) Factors Influencing the Development of Non-Agricultural Business Activities in Rural Eastern Poland. *Acta Scientiarum Polonorum, Oeconomia*, 18(1), 43-52.
- Makhura M.T., Kirsten J.F., Mathye M.M. (1999) The response of nonfarm income to conditions in agricultural markets: an application of multivariate analysis of variance (MANOVA). *Agrekon*, 38(4), 594-602.
- Martiniello P. (1988) Development of a database computer management system for retrieval on varietal field evaluation and plant breeding information in agriculture. *Computers and Electronics in Agriculture*, 2(3), 183-192.
- Matyja M. (2017a) A Comparative Study of Profitability of Agricultural Cooperatives in Poland and Around the World. *Problems of World Agriculture/Problemy Rolnictwa Światowego*, 17(4), 218-227.
- Matyja M. (2017b) Polish farm labour force from the perspective of agricultural worker cooperatives. *Więś i Rolnictwo*, 176(3), 71-95.
- Mądry W. (1996) Doświadczalnictwo rolnicze. Planowanie doświadczeń czynnikowych i analiza wyników. Fundacja „Rozwój SGGW”, Warszawa.

- Maxwell D.G. (1995) Alternative food security strategy: A household analysis of urban agriculture in Kampala. *World Development*, 23(10), 1669-1681.
- Mead R., Riley J. (1981) A review of statistical ideas relevant to intercropping research. *Journal of the Royal Statistical Society: Series A*, 144(4), 462-487.
- Meric G., Leveen S.S., Meric I. (1991) The financial characteristics of commercial banks involved in interstate acquisitions. *Financial Review*, 26(1), 75-90.
- Miller R.G. (1966) *Simultaneous Statistical Inference*. McGraw-Hill, New York.
- Niftiyev I., Czech K. (2020) Dutch Disease Perspective on Vegetable Exports in the Azerbaijan Economy. *Journal of Applied Economic Sciences*, 15, 4(70), 813-827.
- Nowacki R., Fandrejewska A. (2017) Uwarunkowania kulturowe reklamy jako narzędzia komunikacji marketingowej przedsiębiorstw zagranicznych na polskim rynku artykułów żywnościowych. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 117, 137-150.
- Paulo Farias D. de, Araújo F.F. de (2020) Will COVID-19 affect food supply in distribution centers of Brazilian regions affected by the pandemic? *Trends in Food Science & Technology*, 103, 361-366.
- Pietrzykowski R. (2005) Wykorzystanie nowej wielowymiarowej metody statystycznej do badania zmienności somaklonalnej na przykładzie żyta ozimego (*Secale cereale* L.). *Zeszyty Problemowe Postępów Nauk Rolniczych*, 497, 495-502.
- Pillai K.C.S. (1955) Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, 26(1), 117-121.
- Pillai K.S., Jayachandran K. (1967) Power comparisons of tests of two multivariate hypotheses based on four criteria. *Biometrika*, 54(1-2), 195-210.
- Roy S.N. (1945) The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null-hypothesis. *Sankhyā: The Indian Journal of Statistics*, 7(2), 133-158.
- Rutherford A. (2011) *ANOVA and ANCOVA: a GLM approach*. John Wiley & Sons, New Jersey.
- Scheffé H. (1953) A method for judging all contrasts in the analysis of variance. *Biometrika*, 40(1-2), 87-110.
- Scheffé H. (1977) A Note on a Reformulation of the S-Method of Multiple Comparison. *Journal of the American Statistical Association*, 72(357), 143-144.
- Schiefer J., Hirsch S., Hartmann M., Gschwandtner A. (2013) Industry, firm, year and country effects on profitability in EU food processing. *School of Economics Discussion Papers*, 1309, 1-47.
- Schultz T.W., Snedecor G.W. (1933) Analysis of variance as an effective method of handling the time element in certain economic statistics. *Journal of the American Statistical Association*, 28(181), 14-30.
- Schwertman N.C., Carter N.J. (1995) A more practical Scheffe-type multiple comparison procedure for commonly encountered numbers of comparisons. *Journal of Statistical Computation and Simulation*, 53(3-4), 181-196.
- Sroka W. (2018) Conversion of agricultural land to non-agricultural purposes in selected Polish metropolitan areas. *Acta Scientiarum Polonorum, Oeconomia*, 17(2), 97-107.
- Standar A. (2013) Analiza wskaźnikowa sytuacji finansowej gmin województwa wielkopolskiego. *Journal of Agribusiness and Rural Development*, 1(27), 219-232.
- Stoline M.R. (1981) The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. *The American Statistician*, 35(3), 134-141.
- Ślusarczyk B., Grondys K. (2018) The concept of sustainable development in the functioning of municipalities belonging to special economic zones in Poland. *Sustainability*, 10(7), 1-20.
- Śmiałowski T., Bogacka M., Nita Z., Witkowski E. (2011) Wykorzystanie wieloczynnikowej analizy wariancji do oceny przetrzymywania wybranych rodów pszenicy ozimej. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 259, 51-61.
- Średzińska J. (2018) Zróżnicowanie poziomu dochodów rolników w gospodarstwach z różnych klas wielkości ekonomicznej w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 20(69), 215-223.

- Tan P.M., Koh H.C., Low L.C. (1997) Stability of financial ratios: A study of listed companies in Singapore. *Asian Review of Accounting*, 5(1), 19-39.
- Tukey J.W. (1953) The problem of multiple comparisons. Princeton University, Princeton.
- Utnik-Banaś K. (2017) Zmienność cen mięsa kurcząt brojlerów w krajach Unii Europejskiej w latach 2007-2016. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(2), 287-297.
- Van Aelst S., Willems G. (2011) Robust and efficient one-way MANOVA tests. *Journal of the American Statistical Association*, 106(494), 706-718.
- Wielechowski M., Czech K., Grzęda Ł. (2020) Decline in Mobility: Public Transport in Poland in the time of the COVID-19 Pandemic. *Economies*, 8(4), 78, 1-24.
- Wilks S.S. (1932) Certain generalizations in the analysis of variance. *Biometrika*, 24(3/4), 471-494.

4. Wzorcowe metody klasyfikacji (grupowania) obiektów

Pojęcie klasyfikacji zazwyczaj rozumie się jako systematyczny podział przedmiotów lub zjawisk na klasy, działy czy poddziały, dokonywany według określonej zasady. W sensie teoriomnogościowym klasyfikacja oznacza podział zupełny danego zbioru na pewną liczbę rozłącznych podzbiorów. Przy tym przedmiotem klasyfikacji są zbiory obserwacji (obiektów) opisanych zazwyczaj wieloma cechami – zarówno mierzalnymi (ilościowymi), jak i niemierzalnymi (jakościowymi). Podziału dokonuje się na podstawie relacji podobieństwa, a otrzymane podzbiory nazywa się klasami abstrakcji, klasami podobieństwa lub jednorodności [Ostasiewicz (red.) 1999].

W literaturze przedmiotu zamiast określenia klasyfikacja używa się zamiennie następujących terminów: grupowanie, podział, delimitacja, taksonomia numeryczna, taksonometria, analiza skupień czy identyfikacja. Wynika to z tego, iż metody klasyfikacji znajdują zastosowania w wielu obszarach badawczych, na przykład w biologii, botanice, psychologii, antropologii, ekonomii, matematyce czy zoologii. W efekcie mamy do czynienia z dużą różnorodnością terminologiczną.

Podstawowy podział metod klasyfikacji pozwala rozróżnić wzorcową i bezwzorcową klasyfikację obiektów. To, co je różni, to fakt, że w klasyfikacji bezwzorcowej (w przeciwieństwie do wzorcowej) nie jest znana przynależność obiektów do klas. Zatem chcąc przeprowadzić wzorcową klasyfikację obiektów (nazywaną również rozpoznawaniem z nauczycielem), przede wszystkim należy założyć, że populację można podzielić na części nazywane klasami. Przy tym najczęściej analizuje się przypadki dwóch klas, na przykład: czy pacjent jest chory – czy nie, czy klient spłaci kredyt – czy nie, czy firma jest w dobrej kondycji finansowej – czy nie itd. Model klasyfikacyjny, który został opracowany na podstawie zbioru obiektów, o których wiadomo, do jakiej klasy należą, może być wykorzystany do klasyfikacji obiektów o nieznanym przynależności.

4.1. Opis metod

Najczęściej stosowane metody klasyfikacji wzorcowej to [Kisielińska, Stańko 2009]:

- 1) liniowa analiza dyskryminacyjna,
- 2) modele prawdopodobieństwa,
- 3) sieci neuronowe.

4.1.1. Analiza dyskryminacyjna

Analiza dyskryminacyjna stwarza możliwości konstrukcji formuły matematycznej, identyfikującej przynależność obiektów do jednej z kilku wyróżnionych grup, przy czym najprostszym przypadkiem jest podział na dwie grupy. Konstrukcja reguły przynależności opiera się na wielowymia-

rowej informacji o każdej badanej jednostce (obiekcie) i jej przynależności do określonej grupy. W analizie dyskryminacji definiuje się zmienną grupującą, nazywaną zmienną zależną, której wartości pozwalają na zdefiniowanie poszczególnych grup. Definiuje się również zmienne niezależne, które charakteryzują poszczególne obiekty, należące do każdej z wyróżnionych grup. Reguła przynależności jest formułą matematyczną zbudowaną ze zmiennych niezależnych, zwanych również zmiennymi diagnostycznymi, ponieważ dysponując skonstruowaną wcześniej regułą przynależności, można jedynie na ich podstawie rozstrzygnąć o przynależności obiektu do określonej grupy [Rószkiewicz 2002].

Formalnie w zagadnieniu dyskryminacji, jeśli dysponujemy obserwacją (wektorem) \mathbf{x} , zadanie polega na „odgadnięciu”, czy obserwacja pochodzi z rozkładu o gęstości $f(\mathbf{x}, \theta_1)$, czy $f(\mathbf{x}, \theta_2)$. Kryterium klasyfikacji jest następujące: obserwację należy zaliczyć do pierwszej grupy (z parametrem θ_1), jeśli

$$\frac{f(\mathbf{x}, \theta_1)}{f(\mathbf{x}, \theta_2)} > 1. \quad (4.1)$$

W przeciwnym razie obserwacja należy do drugiej grupy (z parametrem θ_2).

W szczególnym przypadku kiedy obie gęstości są normalne (k -wymiarowe):

$$f(\mathbf{x}, \mu_i) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_i)^T \Sigma^{-1} (\mathbf{x} - \mu_i) \right] \quad (4.2)$$

dla $i = 1, 2$ iloraz gęstości przyjmuje postać:

$$\exp \left[(\mu_1 - \mu_2)^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) \right] \quad (4.3)$$

i jest większy od jedności, jeśli kryterium dyskryminacyjne

$$KD = (\mu_1 - \mu_2)^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) > 0, \quad (4.4)$$

gdzie μ_1 i μ_2 to średnie, a Σ jest macierzą wariancji–kowariancji [Dziechciarz (red.) 2002].

Na podstawie wielowymiarowej informacji o każdym z obserwowanych obiektów oraz o jego przynależności do określonej grupy konstruuje się funkcje nazywane funkcjami dyskryminacji, które przekształcają wielowymiarową informację o badanym obiekcie w skalar. Funkcje te wyznacza się w taki sposób, aby maksymalizować zróżnicowanie ich wartości między grupami względem zróżnicowania wewnątrz grup. Poszukuje się więc takiego rozwiązania, aby wewnątrz istniejących grup jednostek uzyskać jak największą jednorodność wartości funkcji przynależności, a między grupami – jak największą heterogeniczność [Rószkiewicz 2002].

W analizie dyskryminacyjnej dąży się więc do utworzenia kombinacji liniowej zmiennych niezależnych, która najlepiej rozdziela dwie (lub więcej) grupy określone *a priori*. Zadaniem analizy dyskryminacyjnej jest budowa funkcji dyskryminacyjnej postaci:

$$y = a_1 x_1 + a_2 x_2 + \dots + a_k x_k, \quad (4.5)$$

gdzie y jest wartością funkcji, jeśli wartości czynników wynosiły odpowiednio: x_1, x_2, \dots, x_k . Ponadto zakłada się, że wszystkie rozpatrywane zmienne mają rozkład normalny.

Wektor kolumnowy współczynników w równaniu (4.5) wyraża się wzorem:

$$\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \quad (4.6)$$

gdzie $\bar{\mathbf{x}}_1$ to średnia z pierwszej grupy obserwacji, $\bar{\mathbf{x}}_2$ średnia z drugiej grupy obserwacji (nazywane też centroidami klas), a \mathbf{S} to macierz wariancji i kowariancji dla całej próby dana wzorem:

$$\mathbf{S} = \frac{1}{n_1+n_2-2} \cdot \left[\sum_{i=1}^{n_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2) \cdot (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)^T + \sum_{i=1}^{n_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) \cdot (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)^T \right]. \quad (4.7)$$

Funkcja dyskryminacyjna liniowa ma postać:

$$\mathbf{y} = \mathbf{a}^T \mathbf{x}, \quad (4.8)$$

gdzie T oznacza transpozycję macierzy.

Jeżeli obliczymy wartości funkcji dyskryminacyjnej dla wielkości $\bar{\mathbf{x}}_1$ oraz $\bar{\mathbf{x}}_2$ otrzymanych z próby i oznaczymy $\bar{\mathbf{y}}_1 = \mathbf{a}^T \bar{\mathbf{x}}_1$ i $\bar{\mathbf{y}}_2 = \mathbf{a}^T \bar{\mathbf{x}}_2$, to okaże się, że punktem leżącym „pośrodku” między punktami $\bar{\mathbf{y}}_1$ oraz $\bar{\mathbf{y}}_2$ jest punkt:

$$\frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2). \quad (4.9)$$

Jest to średnia obliczona z liczb $\bar{\mathbf{y}}_1$ i $\bar{\mathbf{y}}_2$. Stanowi to podstawę do przyjęcia założenia, że daną obserwację należy zaliczyć do grupy obserwacji „zadowolających”, jeśli miara dyskryminacyjna jest dla niej większa niż średnia (4.9). Obserwację, dla której funkcja dyskryminacyjna ma wartość mniejszą od średniej (4.9), należy zaś zakwalifikować do drugiej grupy. Kryterium decyzyjne dla nowej obserwacji \mathbf{x} można zapisać następująco:

$$KD = \mathbf{a}^T \mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (4.10)$$

i jeżeli $KD > 0$, to obserwację klasyfikujemy do pierwszej grupy, a jeśli $KD < 0$, to do drugiej grupy [Dziechciarz (red.) 2002].

Skonstruowany model dyskryminacyjny powinien gwarantować istotność różnicy między średnimi wartościami funkcji dyskryminacyjnej dla badanych populacji. Do zbadania tej hipotezy można wykorzystać następujący test (nazywany testem Hotelinga):

$$F = \frac{n_1 \cdot n_2 \cdot (n_1 + n_2 - k - 1)}{(n_1 + n_2) \cdot (n_1 + n_2 - 2) \cdot k} \cdot D^2, \quad (4.11)$$

gdzie n_1 i n_2 są liczebnościami odpowiednio pierwszej i drugiej klasy, a D^2 jest odległością Mahalanobisa między centroidami klas:

$$D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \cdot \mathbf{S}^{-1} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \quad (4.12)$$

Statystyka (4.11) ma rozkład F o k i $(n_1 + n_2 - k - 1)$ stopniach swobody [Kisielińska 2008].

Ponadto do oceny funkcji dyskryminacyjnych wykorzystywany jest współczynnik λ Wilksa:

$$\lambda = \frac{\det(\mathbf{S})}{\det(\mathbf{T})}, \quad (4.13)$$

gdzie \mathbf{S} jest macierzą wariancji–kowariancji wewnątrzgrupowej, natomiast \mathbf{T} jest łączną macierzą wariancji–kowariancji. Jeśli jako \mathbf{A} oznaczona zostanie macierz wariancji–kowariancji międzygrupowej, łączna macierz wariancji–kowariancji \mathbf{T} jest równa $\mathbf{S} + \mathbf{A}$.

Współczynnik Wilksa informuje o tym, jaka część zmienności funkcji dyskryminacji nie jest wyjaśniana różnicami między grupami, stąd jego wartości mieszczą się w przedziale od 0 do 1. Im mniejsza wartość statystyki λ , tym poprawniejsza jest formuła dyskryminacyjna. W procedurze analizy dyskryminacji weryfikuje się zgodność współczynnika Wilksa z wartością 1. Jeśli brak podstaw do odrzucenia takiej hipotezy, to klasyfikacja metodą analizy dyskryminacji nie daje istotnych statystycznie wyników. Jeśli natomiast można uznać, iż współczynnik Wilksa istotnie różni się od 1, klasyfikację można uznać za istotną statystycznie. Miarą rozbieżności wartości współczynnika Wilksa od wartości 1 jest statystyka χ^2 , obliczana według wzoru:

$$\chi^2 = -\left(n - \frac{s+k}{2}\right) \ln\left(\frac{1}{1+\lambda}\right), \quad (4.14)$$

gdzie n to liczba obserwacji, s liczba grup ($s = 2$), k liczba zmiennych diagnostycznych, λ wartość współczynnika Wilksa. Gdy obliczona wartość statystyki χ^2 przekracza wartość krytyczną, odczytaną z tablic rozkładu chi-kwadrat dla danego poziomu istotności α oraz $\nu = k$ stopni swobody, wyniki klasyfikacji metodą analizy dyskryminacji można uznać za statystycznie istotne [Rószkiewicz 2002].

Współczynnik Wilksa pozwala ocenić jakość klasyfikacji metodą analizy dyskryminacji, lecz nie dostarcza informacji o rolach, jakie odgrywają poszczególne zmienne. Aby taką informację uzyskać, wyznacza się funkcje dyskryminacyjne, pomijając pojedyncze zmienne. Dla takich zredukowanych modeli także oblicza się statystyki Wilksa, oznaczając je jako λ_j , przy czym j jest tu indeksem wyeliminowanej zmiennej. Jeżeli rola jej nie jest duża, to różnica między λ_j i λ jest niewielka. Jako miarę znaczenia cechy w modelu dyskryminacyjnym wykorzystuje się cząstkowy współczynnik λ Wilksa LC_j , będący ilorazem λ_j i λ . Wartości LC_j będą zawsze mniejsze lub równe 1, gdyż wyeliminowanie zmiennej może albo pogorszyć liniowy model dyskryminacyjny, albo go nie zmienić, a na pewno nie może go poprawić. Gdy zmienna nie ma dużej mocy dyskryminacyjnej, pominięcie jej nie wpływa na jakość modelu i wówczas cząstkowy współczynnik Wilksa jest bliski jedności. Im mniejsza jest wartość LC_j dla danej zmiennej, tym większy jest jej udział w dyskryminowaniu klas.

Na podstawie cząstkowych współczynników λ Wilksa wyznacza się dla cech statystyki F_j według wzoru:

$$F_j = \frac{(1-LC_j)/(k-1)}{LC_j/(n-s-k)}. \quad (4.15)$$

Im F_j jest większe, tym większa jest moc dyskryminacyjna cechy. Oczywiście w funkcji dyskryminacyjnej powinno się uwzględniać tylko te cechy, które w istotny sposób przyczyniają się do rozróżniania klas. Włączenie do modelu zmiennych o małych mocach dyskryminacyjnych nie podnosi jego jakości, a tylko go komplikuje. Podstawą eliminacji cech w modelach dyskryminacyjnych są właśnie wartości F_j [Kisielińska 2008].

4.1.2. Modele prawdopodobieństwa

Probabilistyczne metody estymacji modelu są stosowane wtedy, gdy zmienna objaśniana ma charakter dychotomiczny, a więc jest zmienną dwumianową, na przykład zero-jedynkową. Zmienna dychotomiczna przyjmuje jedną z dwóch wartości, na przykład zgodnie z regułą:

$$y_i = \begin{cases} 1, & \text{zdarzenie} \\ 0, & \text{brak zdarzenia} \end{cases} \quad (4.16)$$

W kontekście ekonomicznym wystąpienie zdarzenia można rozumieć bardzo szeroko. Może to być na przykład posiadanie danej cechy przez jakiś obiekt, podjęcie określonej decyzji, zakup pewnego produktu, odniesienie sukcesu rynkowego przez firmę czy zwrot kredytu. Wartości zmiennej objaśnianej y_i są kształtowane przez realizacje zmiennej objaśniającej x_i , co prowadzi do liniowego modelu regresji z jedną zmienną objaśniającą:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad (4.17)$$

Gdy zaś liczba zmiennych niezależnych jest większa niż 1, model przyjmuje postać:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (4.18)$$

z założeniem, że $E(\varepsilon_i) = 0$. Wówczas:

$$E(y_i) = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki}, \quad (4.19)$$

a to znaczy, że prawdopodobieństwo

$$p_i = F(\mathbf{x}_i \boldsymbol{\beta}) = \mathbf{x}_i \boldsymbol{\beta}. \quad (4.20)$$

Jest to liniowy model prawdopodobieństwa (LMP), w którym prawdopodobieństwo p_i jest liniową funkcją parametrów modelu. Szacując ten model, w istocie szacuje się p_i , co znaczy, że oszacowana wartość zmiennej y_i jest oceną prawdopodobieństwa [Gruszczynski, Kluza, Winek 2003].

Można jednak wykazać, że składnik systematyczny modeli (4.17) i (4.18) może przyjmować wartości różne od 0 i 1. W związku z tym wartości teoretyczne \hat{y}_i (to znaczy oszacowanie p_i), otrzymane na ich podstawie, mogą być większe od 1, mniejsze od 0 lub należeć do przedziału $[0; 1]$, podczas gdy y_i , jak na to wskazuje zależność (4.16), powinno być równe 1 lub 0. Z tego powodu wprowadza się przekształcenia odwzorowujące skalę parametru p z przedziału $[0; 1]$ na przedział $(-\infty; +\infty)$. Do najczęściej stosowanych metod transformacji należą przekształcenia logitowe i probitowe.

Przekształcenie logitowe jest definiowane następująco:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right), \quad (4.21)$$

gdzie $\text{logit}(p)$ to wartość logitu dla danego p , a p to wartość prawdopodobieństwa lub częstość występowania określonego zdarzenia w próbie [Gatnar, Walesiak (red.) 2004].

Logit jest określany jako logarytm ilorazu szans. Określa on relatywną możliwość wystąpienia zdarzenia. Transformacja ta ma następujące własności [Ostasiewicz (red.) 1999]:

$$p \in (0; 1) \Leftrightarrow \text{logit}(p) \in (-\infty; +\infty),$$

$$p \rightarrow 0 \Leftrightarrow \text{logit}(p) \rightarrow -\infty,$$

$$p \rightarrow 1 \Leftrightarrow \text{logit}(p) \rightarrow +\infty.$$

Wykres funkcji logit jest symetryczny względem punktu $(0,5; 0)$, a dla $p \in (0,2; 0,8)$ funkcja ta jest prawie liniowa.

Stosując transformację logitową do danych dwumianowych, opisanych liniowym modelem regresji (4.18), otrzymujemy logitowy model regresji postaci:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i. \quad (4.22)$$

Rozwiązując równanie (4.22) względem p_i , otrzymujemy funkcję logistyczną, umożliwiającą obliczenie wartości prawdopodobieństwa p_i postaci:

$$p_i = \frac{1}{1 + \exp(-x_i \beta)} \quad (4.23)$$

lub równoważnie:

$$p_i = \frac{\exp(x_i \beta)}{1 + \exp(x_i \beta)}. \quad (4.24)$$

W modelu probitowym nieobserwowalna wielkość p_i jest wyznaczana z wartości dystrybuanty rozkładu normalnego i wynosi:

$$p_i = \int_{-\infty}^{x_i \beta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \quad (4.25)$$

Między parametrami β w modelu logitowym i probitowym zachodzi relacja:

$$\beta_{\text{logit}} = 1,6 \beta_{\text{probit}}. \quad (4.26)$$

W praktyce wykorzystuje się tylko jedną z tych metod, ponieważ wszystkie miary dopasowania modelu są podobne [Kufel 2007].

Do estymacji parametrów modelu logitowego i probitowego stosuje się zwykle metodę największej wiarygodności. Estymatory parametrów, uzyskane tą metodą, mają asymptotyczny rozkład normalny i są asymptotycznie najefektywniejsze. Ocena dobroci modelu obejmuje ocenę wiarygodności uzyskanych wyników. Miarą wiarygodności jest tu ujemny dwukrotny logarytm funkcji wiarygodności $(-2 \ln L)$, który jest tym mniejszy, im większa jest wiarygodność wyników. Dla modelu doskonałego, którego wiarygodność wynosi 1, mamy: $-2 \ln L = 0$. Poprawność rozwiązania jest oceniana za pomocą różnicy między ujemnym dwukrotnym logarytmem wiarygodności dla modelu zawierającego tylko stałą i dla oszacowanego modelu. Statystyka ilorazu wiarygodności może być zapisana jako:

$$\chi^2 = -2(\ln L_R - \ln L_{UR}). \quad (4.27)$$

Dla dużych prób statystyka ta ma rozkład zbliżony do rozkładu χ^2 o liczbie stopni swobody k . Duża wartość statystyki, przekraczająca wartość krytyczną odczytaną z tablic, wskazuje na istotność otrzymanego rozwiązania. Dodatkowo oceniana jest istotność poszczególnych współczynników regresji logistycznej. Podstawą oceny jest statystyka χ^2 dana wzorem:

$$\chi^2 = \left(\frac{b_j}{S(b_j)} \right)^2, \quad (4.28)$$

gdzie b_j to oszacowana na podstawie danych empirycznych wartość współczynnika regresji logistycznej, zaś $S(b_j)$ to standardowy błąd szacunku współczynnika regresji logistycznej. Jest to statystyka Walda. Im większe przyjmuje ona wartości, tym mocniejsze są podstawy do uznania oszacowania współczynnika za rozwiązanie istotne. Punktem odniesienia jest wartość krytyczna odczytana z tablic rozkładu chi-kwadrat dla danego poziomu istotności i jednego stopnia swobody [Rószkiewicz 2002].

Gdy zmienna objaśniana przyjmuje tylko dwie wartości, nie jest uzasadnione korzystanie z konwencjonalnych miar dopasowania typu R^2 . Dla modeli zmiennych jakościowych wykorzystuje się między innymi [Maddala 2006]:

- kwadrat współczynnika korelacji między y oraz \hat{y} ,
- miarę opartą na ilorazie wiarygodności, daną wzorem:

$$R^2 = 1 - \left(\frac{L_R}{L_{UR}} \right)^{2/n}, \quad (4.29)$$

- pseudo R^2 :

$$pseudo R^2 = \frac{L_{UR}^{2/n} - L_R^{2/n}}{(1 - L_R^{2/n}) L_{UR}^{2/n}}, \quad (4.30)$$

- R^2 McFaddena:

$$R_{McFadden}^2 = 1 - \frac{\log L_{UR}}{\log L_R}, \quad (4.31)$$

- zliczeniowy R^2 :

$$zliczeniowy R^2 = \frac{\text{liczba trafnych prognoz}}{\text{łączna liczba obserwacji}}. \quad (4.32)$$

Uogólnieniem modelu logitowego dla danych binarnych (regresji logistycznej) jest wielomianowy model logitowy, który może być stosowany wówczas, gdy zmienna objaśniana przyjmuje w sposób dyskretny wartości ze zbioru liczącego więcej niż dwie kategorie. Z kolei uogólnieniem wielomianowego modelu logitowego jest, zaproponowany przez McFaddena [1974], warunkowy model logitowy [Gatnar, Walesiak (red.) 2004].

4.1.3. Sieci neuronowe

Sieci neuronowe także można wykorzystać do zadania klasyfikacji wzorcowej, podobnie jak dla zadania regresyjnego, zakładając jedynie dwie wartości zmiennej zależnej: 0 i 1. Takie podejście jest analogiczne do liniowej analizy dyskryminacyjnej. Skonstruowanie sieciowego modelu

prawdopodobieństwa wymaga ograniczenia zakresu zmian jej wyjścia do przedziału $[0; 1]$, co można zrealizować, wprowadzając neurony sigmoidalne na wyjściu sieci [Kisielińska, Stańko 2009].

Empiryczna ocena wyników klasyfikacji. Generalnie skuteczność modeli dyskryminacyjnych ocenia się na podstawie przeprowadzonych, za ich pomocą, klasyfikacji zbioru jednostek o liczebności n o znanej przynależności do klas. W zbiorze tym jest n_0 jednostek należących do klasy oznaczonej umownie 0 oraz n_1 jednostek należących do klasy oznaczonej umownie 1. Zbiór ten można podzielić na cztery podzbiory:

- 1) jednostki należące do klasy 0 i zaklasyfikowane do klasy 0 (o liczebności n_{00}),
- 2) jednostki należące do klasy 0, a zaklasyfikowane do klasy 1 (o liczebności n_{01}),
- 3) jednostki należące do klasy 1, lecz zaklasyfikowane do klasy 0 (o liczebności n_{10}),
- 4) jednostki należące do klasy 1 i zaklasyfikowane do klasy 1 (o liczebności n_{11}).

Jednostek rozpoznanych poprawnie jest więc $n_{00} + n_{11}$, zaś błędnie: $n_{01} + n_{10}$. Liczebności, należące do poszczególnych podzbiorów, pozwalają wyznaczyć współczynniki empirycznej trafności (jakości) klasyfikacji:

- globalny (łączy) odsetek poprawnych klasyfikacji (określany w pracy Maddali [2006] jako zliczeniowy R^2):

$$w_t = \frac{n_{00} + n_{11}}{n}, \quad (4.33)$$

- odsetek poprawnych klasyfikacji elementów klasy 0:

$$w_{t0} = \frac{n_{00}}{n_0}, \quad (4.34)$$

- odsetek poprawnej klasyfikacji elementów klasy 1:

$$w_{t1} = \frac{n_{11}}{n_1}. \quad (4.35)$$

Łączne miary trafności klasyfikacji są miarami dobrymi, jeśli klasyfikacja jest symetryczna, a jest – jeśli w obydwu klasach uzyskano taki sam udział poprawnych klasyfikacji, a więc kiedy $w_{t0} = w_{t1}$. Z kolei gdy $w_{t0} \neq w_{t1}$, klasyfikacja jest asymetryczna. Asymetria rozpoznawalności jest niekorzystna, a działania zmierzające do jej zmniejszenia, polegające na zmianie wartości rozgraniczającej klasy (tzw. kalibracja modelu), zawsze zmniejszają łączny wskaźnik trafności klasyfikacji. Ponadto w przypadku klasyfikacji asymetrycznej możliwe jest uzyskanie bardzo dużego udziału poprawnych identyfikacji, nawet jeśli jedna z klas nie jest w ogóle rozpoznawana. Zatem ocena klasyfikacji musi uwzględniać nie tylko miary łączne, ale również oddzielnie dla klas [Kisielińska, Waszkowski 2015].

4.2. Zastosowania wzorcowych metod klasyfikacji

Obszar zastosowań modeli klasyfikacyjnych, zarówno funkcyjnych, jak i sieciowych, obejmuje szeroki zakres zagadnień z dziedziny ekonomii i finansów. Wśród nich można wymienić następujące:

- 1) ocena i prognozowanie sytuacji gospodarczej,
- 2) analiza kształtowania się struktur rynkowych,
- 3) ocena kondycji finansowej firm,
- 4) prognozowanie upadłości,
- 5) ocena wiarygodności kredytobiorców,
- 6) analiza fundamentalna i portfelowa,
- 7) inne.

Jednym z najstarszych obszarów badań były prace nad modelami klasyfikującymi firmy. Rozpoczął je Altman [1968], który jako jeden z pierwszych w 1968 roku za pomocą analizy dyskryminacyjnej badał bankructwa firm produkcyjnych, podczas gdy na przykład Meyer i Pifer [1970] za pomocą liniowego modelu prawdopodobieństwa analizowali wypłacalność banków. Badania kontynuowane były następnie przez wielu autorów z różnych krajów. W polskim piśmiennictwie także można znaleźć liczne pozycje literatury z tego obszaru, między innymi wskazujące na niską przydatność w polskich warunkach modeli stworzonych na potrzeby rozwiniętych gospodarek. W efekcie powstały autorskie propozycje modeli prognozujących upadłość polskich firm (np.: Hadasik [1998], Wierzba [2000], Prusak [2005], Mączyńska i Zawadzki [2006], Hamrol i Chodakowski [2008], Juszczyk [2010], Wardzińska [2012] czy Jagiełło [2013]).

Inny istotny obszar zastosowań modeli klasyfikacyjnych to analiza fundamentalna i portfelowa. Analiza dyskryminacyjna z jednej strony umożliwia klasyfikację oraz podział spółek na atrakcyjne i nieatrakcyjne z punktu widzenia przyjętego kryterium, z drugiej zaś umożliwia dokonanie doboru spółek do portfela papierów wartościowych. Jedną z pierwszych prac, traktujących o zastosowaniu tej metody na polskim rynku kapitałowym, była publikacja Tarczyńskiego [1996]. Z badań przeprowadzonych w Polsce w kolejnych latach wynika, że zastosowanie metod wielowymiarowej analizy porównawczej, w tym analizy dyskryminacyjnej, prowadzi do znacznego zwiększenia efektywności inwestycji (np.: Gruszczyński [2002], Łuniewska [2003], Gierałtowska [2004], Kisielińska i Skórnik-Pokarowska [2006], Łuniewska i Tarczyński [2006], Gierałtowska [2008], Lach [2013], Waszkowski [2013], Zygmunt i Szewczyk [2013] czy Letkowski [2014]).

Metody klasyfikacji wzorcowej znajdują też zastosowania w ekonomice rolnictwa. W tabeli 4.1 zestawiono prace stanowiące przykłady badań, w których w większości wykorzystywano analizę dyskryminacyjną i modele prawdopodobieństwa (sporadycznie – sieci neuronowe). Pozycje te można pogrupować według obszarów tematycznych. Liczną grupę stanowią prace wykorzystujące, znane z literatury, modele dyskryminacyjne do oceny kondycji finansowej, przewidywania kryzysu finansowego lub ryzyka bankructwa przedsiębiorstw branży spożywczej, rolnej i hotelarsko-turystycznej. Na przykład wykorzystaniem modeli analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw przemysłu mięsnego zajęli się Sołoma i Plesiewicz [2011], Wysocki i Kozera [2012] oraz Zielińska-Chmielewska [2015], podejmując próbę weryfikacji skuteczności wybranych (polskich

i zagranicznych) modeli dyskryminujących w prognozowaniu zagrożenia upadłością przedsiębiorstw tej branży w Polsce. Przy tym Wysocki i Kozera [2012] dodatkowo przedstawili propozycję własnego modelu dyskryminacyjnego oceny ryzyka upadłości przedsiębiorstw *stricte* dla przemysłu mięsnego.

Zagrożenie upadłością przedsiębiorstw przemysłu spożywczego w Polsce było przedmiotem badań Firleja, Bargieł i Szymańskiego [2014] oraz Zielińskiej-Sitkiewicz [2016]. Autorzy ci uwzględnili 26 spółek z indeksu WIG-Spożywczy [Firlej i in. 2014] oraz 15 firm sektora spożywczego notowanych na rynku głównym GPW w Warszawie [Zielińska-Sitkiewicz 2016], dla których szacowali modele dyskryminacyjne, zaproponowane przez polskich autorów i porównywali otrzymane wyniki. Podobne podejście, ale tylko do oceny kondycji finansowej Zakładów Tłuszczowych Kruszwica S.A., zastosowali Wasilewski i Domańska [2012] (przy czym uwzględnili oni również model Altmana) oraz Janik [2018], który analizował skuteczność wybranych polskich modeli dyskryminacyjnych na przykładzie spółki Alma Market S.A. Zavrtoniy i Bilyk [2017] do oceny kondycji finansowej 10 ukraińskich firm produkujących olej słonecznikowy zastosowali model Tereschenko rekomendowany przez ukraińskie Ministerstwo Finansów w rozporządzeniu z 14.07.2016 roku. Na tej podstawie wskazali zmienną o największym znaczeniu, określoną przez nich jako autonomia finansowa.

Wykorzystaniem modeli dyskryminacyjnych w przewidywaniu kryzysu finansowego przedsiębiorstw agrobiznesowych zajęli się Zuba [2011] i Żelazowska-Przewłoka [2012a, 2012b]. Zuba [2011] weryfikowała przydatność polskich modeli dyskryminacyjnych do oceny sytuacji finansowej 41 spółdzielni mleczarskich w Polsce, podczas gdy Żelazowska-Przewłoka [2012a, 2012b] wykorzystwała je do prognozowania zagrożenia finansowego przedsiębiorstw agrobiznesowych w województwie świętokrzyskim. Góralski, Pietrzak i Jędralski [2012] zastosowali polskie i zagraniczne modele dyskryminacyjne do oceny kondycji ekonomiczno-finansowej 19 spółek Agencji Nieruchomości Rolnych. Noga, Adamowicz i Jakubowski [2014] weryfikowali zaś skuteczność polskich modeli dyskryminacyjnych w ocenie sytuacji finansowej przedsiębiorstw sektora leśno-drzewnego, obejmując badaniem 98 polskich podmiotów. Balina i Pochopień [2012] sprawdzali efektywność modeli dyskryminacyjnych dla przedsiębiorstw handlu hurtowego żywnością, napojami i wyrobami tytoniowymi, a Tomczak [2018] dla firm przetwórstwa spożywczego. W tych 2 pracach uwzględniono szeroki przegląd polskich i zagranicznych modeli dyskryminacyjnych (odpowiednio 26 i 34), oceniając ich skuteczność w warunkach polskiego rynku. Z kolei Maciąg i Boboła [2015] zastosowały tylko 2 modele dyskryminacyjne (Altmana oraz Mączyńskiej i Zawadzkiego) do oceny możliwości kontynuowania działalności na przykładzie siedmiu wybranych spółek giełdowych z branży gastronomicznej i hotelarsko-turystycznej. W przedstawionych pracach najczęściej wykorzystywano model Mączyńskiej i Zawadzkiego [2006] oraz model Hadasik [1998].

Zastosowaniem analizy dyskryminacyjnej do prognozowania sytuacji finansowej gospodarstw rolniczych zajmowała się Kisielińska [2004, 2006, 2008], między innymi wyznaczając funkcje dyskryminacyjne, wykorzystujące szeroki zakres wskaźników finansowych, w których kryterium

klasyfikacyjne zostało oparte na dochodzie rolniczym. Ryś-Jurek [2008a] stosowała analizę dyskryminacyjną do pomiaru i oceny dochodowości typów produkcyjnych według wielkości ekonomicznej, opartą na danych źródłowych z bazy FADN, obejmujących 615 typów produkcyjnych według wielkości ekonomicznej z krajów UE. Z kolei Zaród [2009] przedstawiła wykorzystanie analizy dyskryminacyjnej do podziału województwa zachodniopomorskiego na regiony przydatności rolniczej. Ta sama autorka wykorzystwała też analizę dyskryminacyjną do klasyfikacji polskich miast na podstawie warunków środowiskowych i komunikacyjnych [Zaród 2020].

Modele prawdopodobieństwa (najczęściej model logitowy, rzadziej – probitowy) również znalazły zastosowanie w szeroko rozumianej ekonomice rolnictwa. W analizach, których przedmiotem były gospodarstwa rolne, model logitowy wykorzystwała na przykład Ryś-Jurek [2005] do badania struktur podmiotowych w rolnictwie, dążąc do uzyskania takiego modelu klasyfikującego gospodarstwa, który będzie można wykorzystać do identyfikacji efektywnych lub nieefektywnych struktur podmiotowych. Ta sama autorka wykorzystwała modele logitowe i probitowe do diagnozowania poziomu rozwoju indywidualnych gospodarstw rolnych [Ryś-Jurek 2008b]. Z kolei Dudek [2010] posłużył się modelami logitowymi do określenia wpływu różnych czynników na zamiar likwidacji gospodarstw indywidualnych. Ostromęcki, Zajac i Mantaj [2015] badali zaś uwarunkowania i stopień dofinansowania gospodarstw rolnych z dochodów uzyskiwanych przez rolników z prowadzonej przez nich pozarolniczej działalności gospodarczej. Strzelecka, Zawadzka i Kurdyś-Kujawska [2018] analizowały wpływ czynników endogenicznych na prawdopodobieństwo dzierżawy gruntów przez gospodarstwa rolne w Polsce.

Innym przykładem zastosowania modelu logitowego jest praca Sompolskiej-Rzechuły i Świłtyka [2016], w której autorzy wykorzystali tę postać funkcji do określenia czynników wpływających na prawdopodobieństwo poprawy przychodów gospodarstw rolnych, specjalizujących się w produkcji mleka. Grzegorzewska [2012] stosowała zaś model logitowy do prognozowania zagrożenia sytuacji finansowej spółek hodowli roślin i zwierząt, podczas gdy Gołaś i Kurzawa [2014] wykorzystali uporządkowany model logitowy w analizie rentowności branż przemysłu spożywczego. Uporządkowany model logitowy zastosowali również Bieniasz i in. [2014] do identyfikacji czynników wpływających na kondycję finansową gmin wiejskich w Polsce oraz Gołaś [2014] do analizy zróżnicowania dochodowości pracy w rolnictwie krajów UE.

Osobną grupę stanowią prace wykorzystujące modele logitowe w badaniach o charakterze społeczno-ekonomicznym. Na przykład Dudek [2009] zajmował się problemem sukcesji w gospodarstwach rolników w wieku przedemerytalnym. Drejerska [2010] analizowała uwarunkowania aktywności zawodowej ludności wiejskiej z perspektywy teorii „swoich” i „obcych”. Kmiec [2017] badała zaś uwarunkowania aktywności edukacyjnej ludności wiejskiej w wieku produkcyjnym.

Roszkowska-Mądra i Mańkowski [2010] analizowali determinanty decyzji rolników o korzystaniu z funduszy UE i kredytów na działalność rolniczą. Przedmiotem badań Katy i Waleni [2015]

było wykluczenie finansowe rolników i przedsiębiorców wiejskich. Kołodziejczyk i Wysocki [2016] oceniali szanse i prawdopodobieństwo zmiany stanu aktywności ekonomicznej ludności wiejskiej w zależności od jej wybranych cech demograficzno-społecznych. Sączewska-Piotrowska [2016] analizowała ryzyko wejścia do sfery ubóstwa w odniesieniu do miejskich i wiejskich gospodarstw domowych. Zawadzka i Strzelecka [2020] skoncentrowały się zaś na identyfikacji cech społeczno-ekonomicznych wiejskich gospodarstw domowych wpływających na prawdopodobieństwo uzyskania wyższego przeciętnego miesięcznego dochodu netto *per capita*. Kurdyś-Kujawska i Sompolska-Rzechuła [2019] badały determinanty popytu na ubezpieczenie na życie wśród rolników z regionu północno-zachodniej Polski. Z kolei Strzelecka i Zawadzka [2020] zajęły się identyfikacją czynników wpływających na skłonność do oszczędzania wiejskich gospodarstw domowych Pomorza Środkowego.

Modele logitowe i probitowe znalazły zastosowanie w badaniach różnych aspektów aktywności innowacyjnej przemysłu spożywczego w Polsce, na przykład w kontekście determinant tej aktywności [Świadek 2012], powiązań z dostawcami [Świadek, Szopik-Depczyńska 2015], odbiorców [Szopik-Depczyńska, Świadek 2015], powiązań z dostawcami i odbiorcami w łańcuchu dostaw [Dzikowski 2015] czy międzynarodowej konkurencyjności [Lewandowska 2014].

W ostatnich latach modele logitowe i probitowe były też szeroko stosowane w analizach regionalnych problemów rolnictwa w krajach kontynentu afrykańskiego. Na przykład Maponya i in. [2018] badali społeczno-ekonomiczne czynniki wpływające na udział małych przedsiębiorstw ogrodniczych w oficjalnych rynkach zbytu w rejonie Przylądka Wschodniego (Afryka Południowa). Kubwimana [2020] badał percepcję ryzyka w działalności biznesowej sprzedawców warzyw (głównie kapusty i marchwi) produkowanych na obszarach wulkanicznych Prowincji Północnej Rwandy. Mdoda [2020] weryfikował czynniki wpływające na świadomość zmian klimatycznych i wybór strategii adaptacyjnych na przykładzie Przylądka Wschodniego w Afryce Południowej. Sekele i in. [2020] analizowali czynniki wpływające na udział drobnych rolników w lokalnym rynku drobiu w prowincji Limpopo (Afryka Południowa). Adams, Osei-Amponsah i Jumpah [2020] badali determinanty, ograniczenia i szanse drobnych farmerów w północnej Ghanie w dostępie do czynników produkcji. Sechube, Belete i Hlongwane [2020] analizowali wpływ dostępu do maszyn rolniczych na efektywność produkcji kukurydzy na przykładzie drobnych rolników z prowincji Mpumalanga w Afryce Południowej.

Niewątpliwie przedstawiony przegląd literatury pokazuje duże możliwości wykorzystania wzorcowych metod klasyfikacji, głównie analizy dyskryminacyjnej i modeli prawdopodobieństwa, w analizie wielu problemów, wpisujących się w szeroko rozumianą ekonomikę rolnictwa i gospodarki żywnościowej.

Tabela 4.1. Przykłady zastosowań wzorcowych metod klasyfikacji w ekonomice rolnictwa

Autor	Problem	Obiekty	Metoda
Kisielińska 2004, 2006, 2008	budowa funkcyjnych i sieciowych modeli dyskryminacyjnych do klasyfikacji gospodarstw rolniczych umożliwiających prognozowanie sytuacji finansowej	gospodarstwa rolne	analiza dyskryminacyjna*, sieci neuronowe
Ryś-Jurek 2005	badanie struktur podmiotowych w rolnictwie	gospodarstwa rolne	logit
Ryś-Jurek 2008a	zastosowanie analizy dyskryminacyjnej do pomiaru dochodowości typów produkcyjnych według wielkości ekonomicznej	gospodarstwa rolne	analiza dyskryminacyjna*
Ryś-Jurek 2008b	diagnoza poziomu rozwoju indywidualnych gospodarstw rolnych	gospodarstwa rolne	logit, probit
Dudek 2009	określenie czynników sprzyjających sukcesji w gospodarstwach rolników w wieku przedemerytalnym	gospodarstwa rolne	logit
Zaród 2009	podział województwa zachodniopomorskiego na rejony przydatności rolniczej	gminy	analiza dyskryminacyjna*
Drejska 2010	weryfikacja założeń teorii „swoich” i „obcych” w kontekście aktywności zawodowej ludności wiejskiej	wiejskie gospodarstwa domowe	logit
Dudek 2010	określenie wpływu różnych czynników na zamiar likwidacji gospodarstw indywidualnych	gospodarstwa rolne	logit
Roszkowska-Mądra, Mańkowski 2010	określenie zależności prawdopodobieństwa podjęcia decyzji przez rolników o korzystaniu z funduszy unijnych i kredytów od wielu zmiennych przyczynowych	właściciele indywidualnych gospodarstw rolnych	logit
Sołoma, Plesiewicz 2011	ocena przydatności klasycznych modeli analizy dyskryminacyjnej do prognozowania upadłości polskich przedsiębiorstw branży mięsnej	spółki	modele dyskryminacyjne
Zuba 2011	ocena skuteczności analizy dyskryminacyjnej w przewidywaniu kryzysu finansowego przedsiębiorstw w Polsce	spółdzielnie mleczarskie	modele dyskryminacyjne
Balina, Pochopień 2012	weryfikacja skuteczności modeli prognozowania zagrożenia bankructwem przedsiębiorstw handlu hurtowego żywnością, napojami i wyrobami tytoniowymi w Polsce	spółki z o.o.	modele dyskryminacyjne
Góralski, Pietrzak, Jędralski 2012	określenie kondycji finansowej i zagrożenia upadłością spółek ANR prowadzących produkcję i hodowlę zwierząt gospodarskich	spółki objęte nadzorem właścicielskim ANR	modele dyskryminacyjne
Grzegorzewska 2012	prognoza zagrożenia sytuacji finansowej spółek hodowli roślin i zwierząt	spółki objęte nadzorem właścicielskim ANR	logit
Świadek 2012	determinanty aktywności innowacyjnej przemysłu spożywczego w Polsce	przedsiębiorstwa	probit
Wasilewski, Domańska 2012	pomiar zagrożenia upadłością z wykorzystaniem wybranych modeli dyskryminacyjnych przedsiębiorstwa ZT Kruszwica S.A.	spółka ZT Kruszwica S.A.	modele dyskryminacyjne
Wysocki, Kozera 2012	ocena sytuacji ekonomiczno-finansowej i zagrożenia upadłością przedsiębiorstw przemysłu mięsnego	spółki z o.o., jawne i akcyjne	analiza dyskryminacyjna*
Żelazowska-Przewłoka 2012a	prognozowanie zagrożenia finansowego wybranych przedsiębiorstw agrobiznesu w województwie świętokrzyskim	przedsiębiorstwa	modele dyskryminacyjne

Tabela 4.1 (cd.)

Autor	Problem	Obiekty	Metoda
Żelazowska-Przewłoka 2012b	rozpoznanie sytuacji ekonomicznej wybranych przedsiębiorstw agrobiznesu w województwie świętokrzyskim	przedsiębiorstwa	modele dyskryminacyjne
Firlej, Bargieł, Szymański 2014	wyłonienie przedsiębiorstw przemysłu spożywczego zagrożonych upadłością i przesłanek tej sytuacji	spółki indeksu WIG-Spożywczy	modele dyskryminacyjne
Bieniasz, Gołaś, Łuczak 2014	identyfikacja czynników wpływających na kondycję finansową gmin wiejskich w Polsce	gminy wiejskie	logit
Gołaś 2014	analiza zróżnicowania dochodowości pracy w rolnictwie krajów UE	kraje UE	logit
Gołaś, Kurzawa 2014	analiza rentowności branż przemysłu spożywczego	branże przemysłu spożywczego	logit
Lewandowska 2014	ocena wpływu barier innowacji na sprawność innowacyjną i międzynarodową konkurencyjność przedsiębiorstw polskiego przemysłu spożywczego	średnie i duże przedsiębiorstwa przetwórstwa spożywczego	logit
Noga, Adamowicz, Jakubowski 2014	ocena skuteczności modeli predykcyjnych w ocenie sytuacji finansowej przedsiębiorstw sektora leśno-drzewnego	przedsiębiorstwa drzewne	modele dyskryminacyjne
Dzikowski 2015	wpływ powiązań z dostawcami i odbiorcami w łańcuchu dostaw na aktywność innowacyjną przemysłu spożywczego w zachodniej Polsce	przedsiębiorstwa	probit
Kata, Walenia 2015	ocena skali i powodów wykluczenia finansowego rolników i drobnych przedsiębiorców na wsi	gospodarstwa indywidualne	logit
Maciąg, Bobola 2015	ocena możliwości kontynuowania działalności na przykładzie wybranych spółek branży gastronomicznej i hotelarsko-turystycznej	spółki	modele dyskryminacyjne
Ostromięcki, Zajac, Mantaj 2015	identyfikacja i ocena uwarunkowań oraz stopnia dofinansowania gospodarstwa rolnego z dochodów uzyskiwanych przez rolników z pozarolniczej działalności gospodarczej	indywidualne gospodarstwa rolne Polski południowo-wschodniej prowadzące działalność pozarolniczą	logit
Świadek, Szopik-Depczyńska 2015	powiązania z dostawcami w kształtowaniu aktywności innowacyjnej przemysłu spożywczego w Polsce	przedsiębiorstwa	probit
Szopik-Depczyńska, Świadek 2015	powiązania z odbiorcami na polu aktywności innowacyjnej w polskim przemyśle spożywczym	przedsiębiorstwa	probit
Zielińska-Chmielewska 2015	ocena sytuacji finansowej z punktu widzenia zagrożenia upadłością przedsiębiorstw przetwórstwa mięsnego	spółki akcyjne	modele dyskryminacyjne
Kołodziejczak, Wysocki 2016	ocena szans i prawdopodobieństwa zmiany stanu aktywności ekonomicznej ludności wiejskiej w Polsce w zależności od wybranych cech demograficzno-społecznych	osoby aktywne ekonomicznie	logit
Sączewska-Piotrowska 2016	ocena ryzyka wejścia i wyjścia ze sfery ubóstwa w miejskich i wiejskich gospodarstwach domowych	gospodarstwa domowe	logit

Sompolska-Rzechuła, Śwityk 2016	określenie czynników wpływających na prawdopodobieństwo poprawy przychodów gospodarstw rolnych, specjalizujących się w produkcji mleka	gospodarstwa rolne	logit
Zielińska-Sitkiewicz 2016	ocena 3 polskich modeli dyskryminacji w zakresie prognozowania zagrożenia upadłością firm sektora spożywczego	spółki sektora spożywczego	modele dyskryminacyjne
Kmieć 2017	określenie czynników wpływających na aktywność edukacyjną ludności wiejskiej w wieku produkcyjnym	mieszkańcy wsi w wieku produkcyjnym	logit
Zavrotniy, Bilyk 2017	ocena przemian w produkcji oleju słonecznikowego na Ukrainie	przedsiębiorstwa	modele dyskryminacyjne
Janik 2018	zastosowanie polskich modeli dyskryminacyjnych do oceny ryzyka upadłości spółki Alma Market S.A.	spółka Alma Market S.A.	modele dyskryminacyjne
Maponya i in. 2018	identyfikacja społeczno-ekonomicznych czynników determinujących udział małych przedsiębiorstw ogrodniczych w oficjalnych rynkach zbytu w rejonie Przylądka Wschodniego w Afryce Południowej	małe gospodarstwa ogrodnicze	logit
Strzelecka, Zawadzka, Kurdyś-Kujawska 2018	określenie czynników wewnętrznych wpływających na prawdopodobieństwo dzierżawy gruntów przez gospodarstwa rolne w Polsce	indywidualne gospodarstwa rolne	logit
Tomczak 2018	weryfikacja efektywności modeli dyskryminacyjnych na przykładzie sektora przetwórstwa spożywczego	przedsiębiorstwa	modele dyskryminacyjne
Kurdyś-Kujawska, Sompolska-Rzechuła 2019	identyfikacja i ocena czynników determinujących zakup ubezpieczenia na życie przez osoby prowadzące działalność rolniczą	właściciele lub zarządcy gospodarstwa rolnego	logit
Adams, Osei-Amponsah, Jumpah 2020	analiza determinant, ograniczeń i szans drobnych farmerów w północnej Ghanie w dostępie do czynników produkcji	drobne gospodarstwa rolne	probit
Kubwimana 2020	badanie percepcji ryzyka w działalności biznesowej sprzedawców warzyw w Prowincji Północnej Rwandy	hurtownicy i pośrednicy w handlu warzywami	logit
Mdoda 2020	ustalenie czynników wpływających na świadomość zmian klimatycznych i wybór strategii adaptacyjnych na przykładzie Przylądka Wschodniego w Afryce Południowej	drobne gospodarstwa rolne	logit
Sechube, Belete, Hlongwane 2020	analiza wpływu dostępu do maszyn rolniczych na efektywność produkcji kukurydzy na przykładzie drobnych rolników z prowincji Mpumalanga w Afryce Południowej	drobne gospodarstwa rolne	logit
Sekele 2020	określenie społeczno-ekonomicznych czynników wpływających na udział drobnych rolników w lokalnym rynku drobiu w prowincji Limpopo	drobne gospodarstwa rolne	logit
Strzelecka, Zawadzka 2020	identyfikacja czynników wpływających na skłonność do oszczędzania wiejskich gospodarstw domowych	wiejskie gospodarstwa domowe Pomorza Środkowego	logit
Zaród 2020	klasyfikacja polskich miast na podstawie warunków środowiskowych i komunikacyjnych	miasta na prawach powiatu	analiza dyskryminacyjna*
Zawadzka, Strzelecka 2020	identyfikacja cech społeczno-ekonomicznych wiejskich gospodarstw domowych, wpływających na prawdopodobieństwo uzyskania wyższego przeciętnego miesięcznego dochodu netto <i>per capita</i>	wiejskie gospodarstwa domowe Pomorza Środkowego	logit

*budowa i estymacja własnych funkcji dyskryminacyjnych

Źródło: opracowanie własne

Literatura

- Adams A., Osei-Amponsah Ch., Jumpah E.T. (2020) Analysing the determinants, constraints and opportunities of smallholder farmers' access to input markets: evidence from Northern Ghana. *Journal of Agribusiness and Rural Development*, 2(56), 133-143.
- Altman E.I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, September, 589-609.
- Balina R., Pochopień J. (2012) Skuteczność modeli do prognozowania bankructwa przedsiębiorstw handlu hurtowego żywnością, napojami i wyrobami tytoniowymi. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 96, 255-263.
- Bieniasz A., Gołaś Z., Łuczak A. (2014) Wielowymiarowa analiza kondycji finansowej gmin wiejskich w Polsce w latach 2007-2011. *Więś i Rolnictwo*, 2(163), 101-121.
- Drejerska N. (2010) Uwarunkowania aktywności zawodowej ludności wiejskiej z perspektywy teorii „swoich” i „obcych”. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 10(25), 1, 51-58.
- Dudek M. (2009) Sukcesja w gospodarstwach rolników w wieku przedemerytalnym. *Zagadnienia Ekonomiki Rolnej*, 3, 112-123.
- Dudek M. (2010) Czynniki wpływające na proces likwidacji gospodarstw indywidualnych. *Journal of Agribusiness and Rural Development*, 1(15), 45-54.
- Dziechciarz J. (red.) (2002) *Ekonometria. Metody, przykłady, zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Dzikowski P. (2015) Wpływ powiązań z dostawcami i odbiorcami w łańcuchu dostaw na aktywność innowacyjną przemysłu spożywczego w zachodniej Polsce. *Journal of Agribusiness and Rural Development*, 2(36), 189-196.
- Firlej K., Bargieł A., Szymański M. (2014) Zagrożenie upadłością przedsiębiorstw przemysłu spożywczego w Polsce na przykładzie spółek z indeksu WIG-Spożywczy. *Folia Pomeranae Universitatis Technologiae Stettinensis*, 74(1), 63-72.
- Gatnar E., Walesiak M. (red.) (2004) *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Gierałtowska U. (2004) Wykorzystanie liniowej funkcji dyskryminacyjnej do klasyfikacji spółek giełdowych. *Taksonomia*, 11, Klasyfikacja i analiza danych. Teoria i zastosowania, 15-22.
- Gierałtowska U. (2008) Możliwości wykorzystania funkcji dyskryminacyjnej na polskim rynku kapitałowym. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego*, 9, 537-551.
- Gołaś Z. (2014) Dochodowość pracy w rolnictwie krajów Unii Europejskiej. *Więś i Rolnictwo*, 3(164), 7-23.
- Gołaś Z., Kurzawa I. (2014) Zastosowanie uporządkowanego modelu logitowego w analizie rentowności branż przemysłu spożywczego. *Zagadnienia Ekonomiki Rolnej*, 1, 78-96.
- Góralski P., Pietrzak S., Jędralski P. (2012) Ocena kondycji finansowej oraz zagrożenia upadłością spółek ANR. *Zeszyty Naukowe Uniwersytetu Szczecińskiego, Finanse, Rynki Finansowe, Ubezpieczenia*, 737(56), 57-68.
- Gruszczyński M. (2002) *Modele i prognozy zmiennych jakościowych w finansach i bankowości*. Oficyna Wydawnicza SGH, Warszawa.
- Gruszczyński M., Kluza A., Winek D. (2003) *Ekonometria*. Dom Wydawniczy Elipsa, Warszawa.
- Grzegorzewska E. (2012) Model wczesnego ostrzegania jako narzędzie oceny kondycji finansowej spółek hodowli roślin i zwierząt. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 96, 231-249.
- Hadasik D. (1998) Upadłość przedsiębiorstw w Polsce. Metody jej prognozowania. *Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu, Seria 2, Prace habilitacyjne*, 153.
- Hamrol M., Chodakowski I. (2008) Prognozowanie zagrożenia finansowego przedsiębiorstw, wartość predykcyjna polskich modeli analizy dyskryminacyjnej. *Badania Operacyjne i Decyzje*, 3. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.

- Jagiełło R. (2013) Analiza dyskryminacyjna i regresja logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw. Materiały i Studia NBP, 286.
- Janik K. (2018) Bankructwo czy stabilność rynkowa – analiza wybranych polskich modeli dyskryminacyjnych na przykładzie spółki Alma Market SA. Zeszyty Naukowe Politechniki Częstochowskiej, Zarządzanie, 29, 235-248.
- Juszczak S. (2010) Prognozowanie upadłości przedsiębiorstw. Ekonomista, 5, 701-728.
- Kata R., Walenia A. (2015) Financial exclusion of farmers and rural entrepreneurs. Journal of Agribusiness and Rural Development, 2(36), 225-235.
- Kisielińska J. (2004) Zastosowanie analizy dyskryminacyjnej do prognozowania sytuacji finansowej gospodarstwa rolniczego. Wieś i Rolnictwo, 1, 100-111.
- Kisielińska J. (2006) Wykorzystanie analizy dyskryminacyjnej oraz sieci neuronowych do prognozowania sytuacji finansowej gospodarstw rolniczych z weryfikacją roli czasu. Acta Scientiarum Polonorum, Oeconomia, 5(2), 37-46.
- Kisielińska J., Skórnik-Pokarowska U. (2006) Wykorzystanie liniowej funkcji dyskryminacyjnej oraz metody głównych składowych w procesie doboru spółek giełdowych do portfela inwestycyjnego. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej, 60, 159-167.
- Kisielińska J. (2008) Modele klasyfikacyjne prognozowania sytuacji finansowej gospodarstw rolniczych. Rozprawy i Monografie. Wydawnictwo SGGW, Warszawa.
- Kisielińska J., Stańko S. (2009) Wielowymiarowa analiza danych w ekonomice rolnictwa. Roczniki Nauk Rolniczych, Seria G – Ekonomika Rolnictwa, 96(2), 63-76.
- Kisielińska J., Waszkowski A. (2015) Zagregowana ocena kondycji finansowej przedsiębiorstw z wykorzystaniem polskich modeli upadłościowych. Ekonomista, 5, 679-692.
- Kmieć D. (2017) Uwarunkowania aktywności edukacyjnej ludności wiejskiej w wieku produkcyjnym. Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu, 18(2), 149-153.
- Kołodziejczak W., Wysocki F. (2016) Wielomianowa analiza logitowa w badaniach aktywności ekonomicznej ludności wiejskiej. Wieś i Rolnictwo 2(171), 11-39.
- Kubwimana J.J. (2020) Risk analysis of vegetables marketing in Rwanda. A case of carrots and cabbages produced in Northern Province, Rubavu District and supplied across the country. Journal of Agribusiness and Rural Development, 2(56), 183-200.
- Kufel T. (2007) Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL. Wydawnictwo Naukowe PWN, Warszawa.
- Kurdyś-Kujawska A., Sompolska-Rzechuła A. (2019) Determinants of demand for life insurance: the example of farmers from North-West Poland. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 63(7), 71-81.
- Lach B. (2013) Selekcja akcji z wykorzystaniem modeli dyskryminacyjnych oraz optymalizacji podziału. Zeszyty Naukowe Akademii Ekonomicznej w Katowicach, Studia Ekonomiczne, 163, Modelowanie Preferencji a Ryzyko '13, 45-58.
- Letkowski D. (2014) Weryfikacja skuteczności wykorzystania funkcji dyskryminacyjnej do przewidywania cen kursów akcji. Zeszyty Naukowe Uniwersytetu Szczecińskiego, 804, Finanse, Rynki Finansowe, Ubezpieczenia, 67, 315-325.
- Lewandowska M.S. (2014) Innovation barriers and international competitiveness of enterprises from Polish food processing industry. Research results. Acta Scientiarum Polonorum, Oeconomia, 13(4), 103-113.
- Łuniewska M. (2003) Wykorzystanie metod ilościowych do tworzenia portfela papierów wartościowych. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
- Łuniewska M., Tarczyński W. (2006) Metody wielowymiarowej analizy porównawczej na rynku kapitałowym. Wydawnictwo Naukowe PWN, Warszawa.
- McFadden D. (1974) Conditional logit analysis of qualitative behavior. Frontiers in Econometrics. Academic Press, New York-San Francisco-London, 105-142.
- Maciąg A., Boboła A. (2015) Zastosowanie finansowych modeli wielowymiarowych do oceny możliwości kontynuowania działalności na przykładzie wybranych spółek giełdowych z branży gastronomicznej i hotelarsko-turystycznej. Przedsiębiorczość i Zarządzanie, 16(4), cz. II, 133-146.

- Maddala G.S. (2006) *Ekonometria*. Wydawnictwo Naukowe PWN, Warszawa.
- Maponya P., Kekana V., Senyolo G.M., Venter S.L. (2018) Socioeconomic factors influencing market participation of horticultural smallholder farmers in the Alfred NZO District, Eastern Cape, South Africa. *Journal of Agribusiness and Rural Development*, 4(50), 421-427.
- Mączyńska E., Zawadzki M. (2006) Dyskryminacyjne modele predykcji upadłości przedsiębiorstw. *Ekonomista*, 2, 1-24.
- Mdoda L. (2020) Factors influencing farmers' awareness and choice of adaptation strategies to climate change by smallholder crop farmers. *Journal of Agribusiness and Rural Development*, 4(58), 401-413.
- Meyer P.A., Pifer W.H. (1970) Prediction of bank failures. *Journal of Finance*, 25, 853-868.
- Noga T., Adamowicz K., Jakubowski J. (2014) Metody dyskryminacyjne w ocenie sytuacji finansowej przedsiębiorstw sektora leśno-drzewnego. *Acta Scientiarum Polonorum, Silvarum Colendarum Ratio Industria Lignaria*, 13(1), 25-35.
- Ostasiewicz W. (red.) (1999) *Statystyczne metody analizy danych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Ostromęcki A., Zając D., Mantaj A. (2015) The importance of non-agricultural economic activity of farmers in the modernization process of farms. *Acta Scientiarum Polonorum, Oeconomia*, 14(4), 93-92.
- Prusak B. (2005) *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*. Difin, Warszawa.
- Roszkowska-Mądra B., Mańkowski D.R. (2010) determinanty decyzji rolników o korzystaniu z funduszy Unii Europejskiej i kredytów na działalność rolniczą: przykład dla rolnictwa z rozwiniętym systemem produkcji mlecznej w województwie podlaskim. *Roczniki Nauk Rolniczych, Seria G*, 97(1), 14-27.
- Rószkiewicz M. (2002) *Metody ilościowe w badaniach marketingowych*. Wydawnictwo Naukowe PWN, Warszawa.
- Ryś-Jurek R. (2005) Zastosowanie analizy logitowej do badania struktur podmiotowych w rolnictwie. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 7(5), 74-79.
- Ryś-Jurek R. (2008a) Using the discriminant analysis to estimate the profitability of production types according to the economic size in the European Union (with the use of FADN data). *Journal of Agribusiness and Rural Development*, 4(10), 109-122.
- Ryś-Jurek R. (2008b) Ocena sytuacji ekonomicznej indywidualnych gospodarstw rolnych z wykorzystaniem wybranych metod ilościowych. *Zastosowania metod statystycznych w badaniach naukowych. III. StatSoft Polska*, 173-180.
- Sączewska-Piotrowska A. (2016) Dynamika ubóstwa w miejskich i wiejskich gospodarstwach domowych. *Wiadomości Statystyczne*, 7, 39-59.
- Sechube M.P., Belete A., Hlongwane J.J. (2020) Analysing the effectiveness of access to tractor service on the technical efficiency of small-scale maize farmers in Mpumalanga Province: a case study of the Masibuyele Emasimini Programme. *Journal of Agribusiness and Rural Development*, 3(57), 317-326.
- Sekele L.S., Mokhaukhau J.P., Cholo M.S., Mayekiso A. (2020) A binary logistic analysis on factors affecting the participation of smallholder farmers in the market of indigenous chickens. *Journal of Agribusiness and Rural Development*, 4(58), 415-422.
- Sołoma A., Plesiewicz J. (2011) Wykorzystanie wielowymiarowych modeli analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw przemysłu mięsnego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 90, 155-169.
- Sompolska-Rzechuła A., Świtłyk M. (2016) Factors affecting probability of income increase in agricultural holdings specialized in milk production. *Zagadnienia Ekonomiki Rolnej*, 4(349), 104-117.
- Strzelecka A., Zawadzka D. (2020) Factors determining the tendency of rural households in Central Pomerania to save – pilot study results. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 23(72), 180-190.
- Strzelecka A., Zawadzka D., Kurdyś-Kujawska A. (2018) Wpływ czynników endogenicznych na prawdopodobieństwo dzierżawy gruntów przez gospodarstwa rolne w Polsce. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 105(2), 38-49.
- Szopik-Decpzyńska K., Świadek A. (2015) Odbiorcy a aktywność innowacyjna w przemyśle spożywczym w Polsce. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 382, 401-409.

- Świadek A. (2012) Determinanty aktywności innowacyjnej przemysłu spożywczego w Polsce. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego, 12(27), 2, 123-131.
- Świadek A., Szopik-Depczyńska K. (2015) Relationships with suppliers and innovation activity in food industry in Poland. *Journal of Agribusiness and Rural Development*, 2(36), 335-341.
- Tarczyński W. (1996) Analiza dyskryminacyjna na giełdzie papierów wartościowych. *Przegląd Statystyczny*, 43(1-2), 49-66.
- Tomczak S.K. (2018) Verification of the effectiveness of re-estimated discriminant models on the example of the food manufacturing sector. *Zarządzanie i Finanse/Journal of Management and Finance*, 16(4), 193-205.
- Wardzińska K. (2012) Przykład zastosowania analizy dyskryminacyjnej do oceny sytuacji finansowej przedsiębiorstw. *Ekonomia i Zarządzanie*, 3, 197-208.
- Wasilewski M., Domańska T. (2012) Wykorzystanie modeli dyskryminacyjnych do oceny kondycji finansowej Zakładów Tłuszczowych Kruszwica S.A. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 96, 303-321.
- Waszkowski A. (2013) Zastosowanie wielowymiarowej analizy porównawczej w doborze spółek do portfela inwestycyjnego. *Zarządzanie i Finanse*, 11(2), cz. 4, 465-476.
- Wierzba D. (2000) Wczesne wykrywanie przedsiębiorstw zagrożonych upadłością na podstawie analizy wskaźników finansowych – teoria i badania empiryczne. *Zeszyty Naukowe Wyższej Szkoły Ekonomiczno-Informatycznej w Warszawie*, 9, 79-105.
- Wysocki F., Kozera A. (2012) Wykorzystanie analizy dyskryminacyjnej w ocenie ryzyka upadłości przedsiębiorstw przemysłu mięsnego. *Journal of Agribusiness and Rural Development*, 4(46), 167-182.
- Zaród J. (2009) Wykorzystanie analizy dyskryminacyjnej do podziału województwa zachodniopomorskiego na rejony przydatności rolniczej. *Journal of Agribusiness and Rural Development*, 3(13), 245-254.
- Zaród J. (2020) Classification of Polish cities based on environmental and communication conditions. *Econometrics/Ekonometria*, 24(1), 12-22.
- Zavrotniy R., Bilyk O. (2017) Transformation of sunflower oil production in Ukraine due to acute economic crisis. *Journal of International Studies*, 10(1), 225-236.
- Zawadzka D., Strzelecka A. (2020) Socio-economic features of rural households in Central Pomerania and their profitability – pilot study results. *Acta Scientiarum Polonorum, Oeconomia*, 19(1), 101-109.
- Zielińska-Chmielewska A. (2015) Use of chosen discrimination models in the assessment of bankruptcy risk in meat processing enterprises. *Journal of Agribusiness and Rural Development*, 2(36), 363-370.
- Zielińska-Sitkiewicz M. (2016) Zastosowanie metod wielowymiarowej analizy dyskryminacyjnej do prognozowania upadłości wybranych spółek sektora spożywczego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 113, 117-129.
- Zuba M. (2011) Wielowymiarowa analiza dyskryminacyjna jako skuteczna metoda przewidywania kryzysu finansowego przedsiębiorstw na przykładzie spółdzielni mleczarskich w Polsce. *Problemy Zarządzania*, 9(1), 72-89.
- Zygmunt A., Szewczyk M. (2013) Zastosowanie metod dyskryminacyjnych jako narzędzia umożliwiającego wspomaganie procesu dywersyfikacji ryzyka inwestycyjnego w akcje. *Zeszyty Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Nauki o Finansach*, 1(14), 115-127.
- Żelazowska-Przewłoka A. (2012a) Prognozowanie zagrożenia finansowego przedsiębiorstw agrobiznesu w województwie świętokrzyskim w latach 2005-2010. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 96, 221-230.
- Żelazowska-Przewłoka A. (2012b) Prognozowanie zagrożenia sytuacji ekonomicznej wybranych przedsiębiorstw agrobiznesu w województwie świętokrzyskim w latach 2004-2010. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 14(4), 141-146.

5. Bezwzorcowe metody klasyfikacji (grupowania) obiektów

Klasyfikacja bezwzorcowa jest stosowana w sytuacji, kiedy strukturę grup należy dopiero odkryć. Analiza skupień (tzw. taksonomia numeryczna, grupowanie obiektowe, ang. *cluster analysis*) stanowi najbogatszą i najpopularniejszą klasę metod rozpoznawania bez nauczyciela, stosowaną w badaniach ekonomicznych. Analizę skupień wykorzystuje się bardzo często w badaniach porównujących czy grupujących obiekty (takie jak kraje, regiony, województwa, powiaty) ze względu na wiele cech, na przykład: poziom rozwoju gospodarczego, poziom i jakość życia ludności, rozwój demograficzny, jakość środowiska naturalnego itp.

Za jedną z pierwszych prac z zakresu analizy skupień uznaje się artykuł R.L. Thorndike'a „Who belongs in the family?” z 1953 roku, w którym po raz pierwszy zastosowano tego rodzaju metody analizy [Domański (red.) 2011]. Metody grupowania oraz związane z nimi problemy są opisane w wielu pracach naukowców zagranicznych i polskich [Pociecha i in. 1988; Sokołowski 1992; Gordon 1999; Gatnar, Walesiak (red.) 2004; Gan, Ma, Wu 2007; Domański (red.) 2011; Everitt i in. 2011; Panek, Zwierzchowski 2013].

Celem analizy skupień jest utworzenie jak najbardziej jednorodnych grup obiektów (skupień) ze względu na podobieństwa, w zakresie wewnętrznej struktury [Kisielińska 2009; Panek, Zwierzchowski 2013]. Elementy w tej samej grupie powinny być do siebie podobne (powinna istnieć homogeniczność w obrębie skupienia), a jednocześnie jak najbardziej odmienne od elementów z pozostałych grup (powinna istnieć heterogeniczność między skupieniami).

5.1. Opis metod

Informacje o obiektach są umieszczane w macierzy, zwanej macierzą obserwacji, o liczbie wierszy równej liczbie obiektów i liczbie kolumn równej liczbie cech. Każdy z n obiektów może być traktowany jako punkt w k -wymiarowej przestrzeni cech. Tak zdefiniowaną zbiorowość należy podzielić na podgrupy, które spełniają warunek zupełności, rozłączności i niepustości [Gatnar, Walesiak (red.) 2004]. Warunek zupełności oznacza, że każdy obiekt przypisany jest do jakiejś grupy. Warunek rozłączności mówi, że każdy obiekt musi należeć do jednej klasy, zaś warunek niepustości – że każde skupienie musi zawierać przynajmniej jeden obiekt.

Z analizą skupień wiążą się cztery podstawowe problemy decydujące o stopniu trudności danego problemu klasyfikacji [Domański (red.) 2011]. Podstawowy problem związany jest z liczbą analizowanych obiektów. W przypadku podziału n analizowanych obiektów na s skupień liczba możliwych podziałów obiektów jest duża. Można ją określić z następującego wzoru:

$$K(n, s) = \frac{1}{s!} \sum_{q=1}^s (-1)^{s-q} \binom{s}{q} q^n. \quad (5.1)$$

Kolejny problem jest związany z liczbą analizowanych zmiennych opisujących badane obiekty. Wraz ze wzrostem liczby zmiennych konieczne jest zastosowanie odpowiednich algorytmów metod klasyfikacji. Następny problem dotyczy rozmieszczenia analizowanych obiektów w wielowymiarowej przestrzeni oraz braku powszechnie akceptowanej definicji grupy (klasy). Ostatni dotyczy braku szeroko akceptowanej i ujednocionej teorii klasyfikacji.

W metodach taksonomicznych stopień podobieństwa obiektów mierzony jest za pomocą miar odległości. Odległości dla wszystkich par analizowanych obiektów przedstawia się w postaci macierzy odległości. Istotne znaczenie dla wyników analizy porównawczej ma wybór miary odległości. Miary te wyliczone są dla znormalizowanych zmiennych. Metody normalizacji omówiono w siódmym rozdziale.

Najbardziej ogólną miarą odległości pomiędzy obiektami, scharakteryzowanymi przez zmienne mierzone w skalach przedziałowej lub ilorazowej, jest metryka Minkowskiego ($p \geq 1$):

$$d_{il} = \left[\sum_{j=1}^k |z_{ij} - z_{lj}|^p \right]^{\frac{1}{p}}, \quad (5.2)$$

gdzie d_{il} to miara odległości między obiektami i oraz l , p parametr w postaci liczby naturalnej, z_{ij} znormalizowana zmienna x_j dla obiektu i , k liczba zmiennych.

W szczególności dla $p = 2$ jest to odległość Euklidesa:

$$d_{il} = \left[\sum_{j=1}^k (z_{ij} - z_{lj})^2 \right]^{\frac{1}{2}} \quad (5.3)$$

dla $p = 1$ – odległość miejska:

$$d_{il} = \sum_{j=1}^k |z_{ij} - z_{lj}| \quad (5.4)$$

dla $p \rightarrow \infty$ – odległość Czebyszewa:

$$d_{il} = \max_j |z_{ij} - z_{lj}|. \quad (5.5)$$

Inną miarą odległości jest odległość potęgowa:

$$d_{il} = \left[\sum_{j=1}^k |z_{ij} - z_{lj}|^a \right]^{\frac{1}{b}}, \quad (5.6)$$

gdzie a, b są parametrami sterującymi wagami zmiennych.

Odległość Mahalanobisa o postaci:

$$d_{il} = \left[\sum_{j=1}^k \sum_{m=1}^k s_{jm} (x_{ij} - z_{lj}) (x_{im} - x_{lm}) \right]^{\frac{1}{2}} \quad (5.7)$$

jest miarą odległości opartą na oryginalnych wartościach zmiennych, gdzie s_{jm} jest to jm -ty element macierzy odwrotnej do macierzy kowariancji zbioru obserwacji.

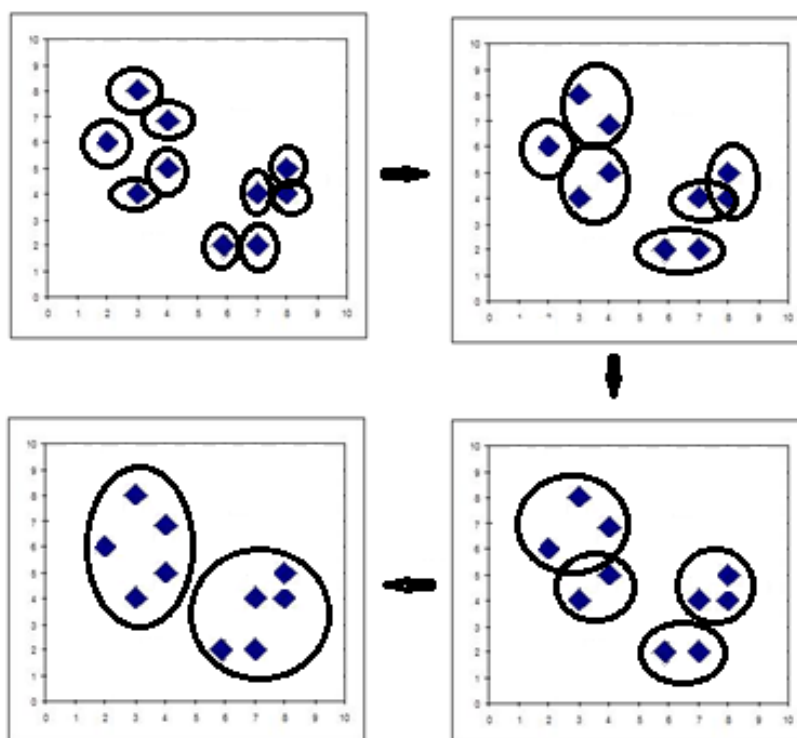
Oprócz wymienionych miar odległości można ponadto wyróżnić miary odległości Canberra, Braya-Curtisa czy tzw. uogólnioną miarę odległości [Domański (red.) 2011].

Podstawowy podział metod analizy skupień obejmuje metody następujące:

- 1) hierarchiczne metody aglomeracyjne i deglomeracyjne,
- 2) metody obszarowe,
- 3) metody optymalizujące wstępny podział obiektów.

5.1.1. Hierarchiczne metody aglomeracyjne i deglomeracyjne

W hierarchicznych metodach grupowania zakłada się, że dana jest macierz odległości obserwacji $\mathbf{D} = [d_{il}]$, $i, l = 1, 2, \dots, n$. Wstępnie przyjmuje się liczbę klas równą liczbie obiektów (tzn. każda obserwacja sama tworzy klasę), a następnie łączy się klasy najbardziej do siebie podobne, redukując w każdym kroku liczbę klas o 1. Po r -tym kroku grupowania liczba klas, które pozostały, wynosi $n - r$. Następnie określa się odległość klasy na nowo powstałej od pozostałych klas. Tworzona jest nowa macierz odległości między klasami o wymiarze $(n - r) \times (n - r)$. Procedurę tę przeprowadza się $n - 1$ razy, aż do uzyskania jednej klasy obejmującej wszystkie obiekty. Na rysunku 5.1 przedstawiono graficzną interpretację metody aglomeracyjnej.

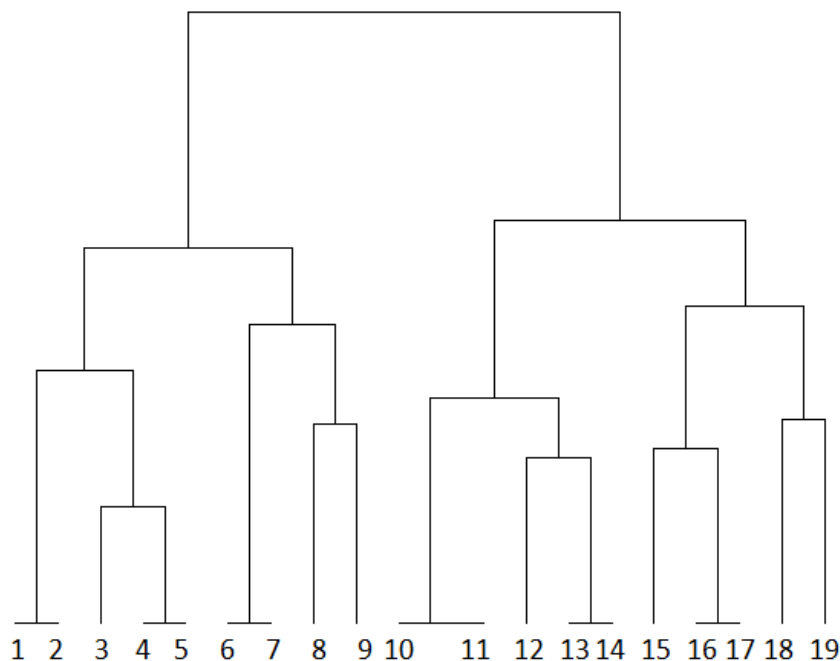


Rysunek 5.1. Interpretacja metody aglomeracyjnej

Źródło: opracowanie własne

Hierarchiczne metody aglomeracyjne mają najlepiej opracowaną metodologię [Gatnar, Walesiak (red.) 2004]. Zalety tej metody klasyfikacji to, że:

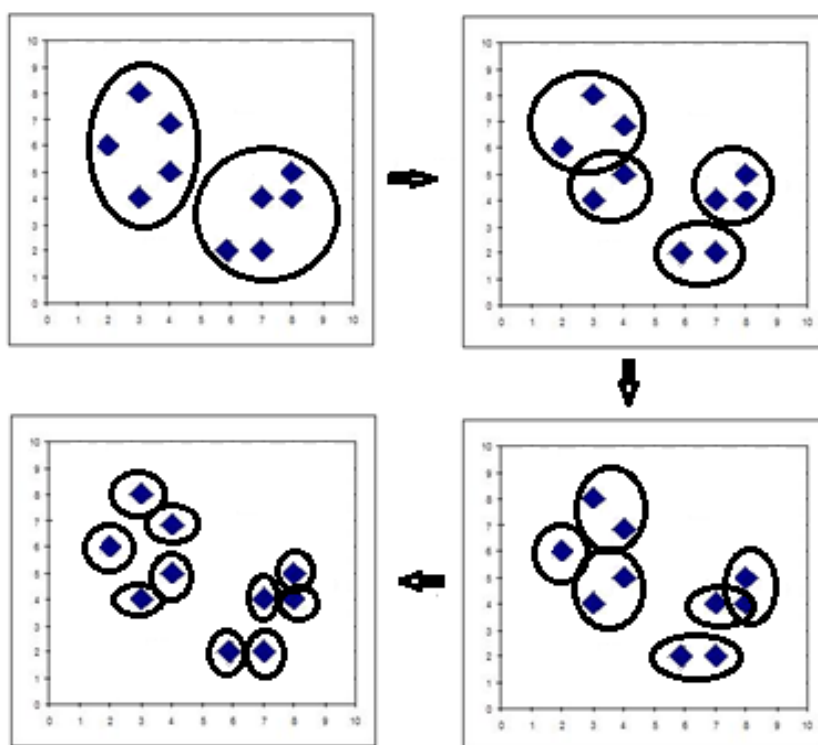
- 1) Działają zgodnie z jedną procedurą.
- 2) Efekty klasyfikacji są przedstawiane w postaci ciągu klasyfikacji.
- 3) Wyniki klasyfikacji można przedstawić w formie dendrogramu (tj. drzewa połączeń), który wskazuje na kolejność połączeń między grupami (rys. 5.2). Jest to (binarne) drzewo, którego węzły reprezentują skupienia, a liście – pojedyncze obiekty. Liście umieszczone są na poziomie zerowym, pozostałe węzły drzewa umieszczone są na wysokości odpowiadającej mierze niepodobieństwa pomiędzy skupieniami reprezentowanymi przez węzły – potomki. Przedstawiona na dendrogramach hierarchia umożliwia precyzyjne określenie, jak są rozmieszczone poszczególne grupy oraz same obiekty w nich zawarte.



Rysunek 5.2. Przykładowy dendrogram

Źródło: opracowanie własne

W hierarchicznych metodach deglomeracyjnych (rozdzielających, dedukcyjnych, zstępujących) procedura jest odwrotna. Na początku zakłada się istnienie jednej klasy. W każdym kolejnym kroku liczbę klas zwiększa się o jeden, aż do uzyskania liczby klas równej liczbie obiektów. Metody rozdzielające są bardziej pracochłonne i są na ogół stosowane w mniejszym zakresie niż metody aglomeracyjne. Do grupy tych metod można zaliczyć metody Huberta oraz metody dendrytowe (taksonomii wrocławskiej oraz metoda najkrótszej sieci połączeń Prima [Gatnar, Walesiak (red.) 2004]. Na rysunku 5.3 przedstawiono graficzną interpretację metody deglomeracyjnej.



Rysunek 5.3. Interpretacja metody deglomeracyjnej

Źródło: opracowanie własne

Przy grupowaniu aglomeracyjnym czy deglomeracyjnym miarą podobieństwa obiektów $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ i klas są odległości pomiędzy nimi. Ze względu na istnienie wielu sposobów określania tych odległości, skutkuje to istnieniem wielu hierarchicznych metod grupowania. Popularne metody to [Jajuga 1990]:

- Metoda pojedynczego wiązania (zwana też metodą najbliższego sąsiedztwa). W metodzie tej odległość pomiędzy dwoma skupieniami C_f i C_h jest określona jako minimalna odległość pomiędzy dwoma najbliższymi obiektami należącymi do tych skupień:

$$d(C_f, C_h) = \min d_{il}(x_i \in C_f, x_l \in C_h). \quad (5.8)$$

- Metoda pełnego wiązania (najdalszego sąsiedztwa). W metodzie tej odległość pomiędzy dwoma skupieniami C_f i C_h jest określona jako największa z odległości pomiędzy dwoma dowolnymi obiektami należącymi do tych skupień:

$$d(C_f, C_h) = \max d_{il}(x_i \in C_f, x_l \in C_h). \quad (5.9)$$

- Metoda średniej międzyklasowej. W metodzie tej odległość pomiędzy dwoma skupieniami C_f i C_h jest określona jako średnia arytmetyczna odległości pomiędzy obserwacjami należącymi do tych klas:

$$d(C_f, C_h) = \frac{1}{n_f n_h} \sum_{x_i \in C_f, x_l \in C_h} d_{il}, \quad (5.10)$$

gdzie n_f i n_h oznaczają liczbę obiektów należących do klasy C_f i C_h .

- Metoda mediany. W metodzie tej odległość pomiędzy dwoma skupieniami C_f i C_h określa się jako medianę odległości pomiędzy obserwacjami należącymi do tych klas:

$$d(C_f, C_h) = \text{med}(d_{il}) \quad (x_i \in C_f, x_l \in C_h). \quad (5.11)$$

- Metoda środka ciężkości. W metodzie tej odległość pomiędzy dwoma skupieniami C_f i C_h określa się jako odległość pomiędzy środkami ciężkości tych klas:

$$d(C_f, C_h) = d(g(C_f), g(C_h)), \quad (5.12)$$

gdzie $g(C_f) = \frac{1}{n_f} \sum_{i \in C_f} \mathbf{x}_i$.

- Metoda Warda. W metodzie tej odległością pomiędzy skupieniami C_f i C_h jest wartość, o jaką zwiększy się wariancja wewnątrzgrupowa po połączeniu grup. Metoda ta zmierza do minimalizacji sumy kwadratów odchyień dowolnych dwóch skupień, które mogą zostać uformowane na każdym etapie. Metoda ta jest traktowana jako bardzo efektywna, z tym, że zmierza do tworzenia skupień o małej wielkości. Miarą zróżnicowania skupienia względem wartości średnich jest ESS (ang. *error sum of squares*), zwane również błędem sumy kwadratów.

5.1.2. Metody obszarowe

W metodach obszarowych przestrzeń wielowymiarową dzieli się na rozłączne podzbiory (podobszary) [Grabiński, Wydymus, Zeliaś 1989; Kisielińska 2009; Panek, Zwierzchowski 2013]. Obiekty znajdujące się w poszczególnych podobszarach zalicza się do jednej grupy obiektów. Podobszary mogą być różnego rodzaju i stanowić je mogą wielowymiarowe kule (hiperkule) czy prostopadłości (hiperkostki).

Grupowanie obiektów metodami obszarowymi zależy od wyjściowego grupowania. Do metod obszarowych należą: taksonomia hiperkul, taksonomia stochastyczna, metoda Thorndike'a, metoda Hartigana, metoda katowicka.

W taksonomii hiperkul znana jest między innymi metoda wrocławska, zaproponowana przez statystyków z Wrocławia [Bukietyński i in. 1969]. W metodzie tej wykorzystuje się elementy macierzy odległości pomiędzy obiektami – promienie hiperkul.

Promienie hiperkul można wyznaczyć, wykorzystując następujące wzory [Panek, Zwierzchowski 2013]:

$$d_0 = \bar{d} + c \cdot S(d) \quad (5.13)$$

lub

$$d_0 = \max_i \min_l \{d_{il}\}, \quad i, l = 1, 2, \dots, n, \quad (5.14)$$

przy czym:

$$\bar{d} = \frac{1}{n} \sum d_i, \quad (5.15)$$

$$S(d) = \left[\frac{1}{n} \sum (d_i - \bar{d})^2 \right]^{\frac{1}{2}}, \quad (5.16)$$

$$d_i = \min_l \{d_{il}\}, i, l = 1, 2, \dots, n, \quad (5.17)$$

gdzie d_{il} jest odległością pomiędzy i -tym i l -tym obiektem należącymi do tej samej grupy obiektów, $S(d)$ odchyleniem standardowym, a c liczbą rzeczywistą, wybraną dowolnie, określaną w sposób empiryczny (może przyjmować wartości 2, -1, -1/2, 0, 1/2, 1 itp.).

Dla ustalonego promienia d_0 tworzone są hiperkule z różną liczbą obiektów. Obiekty należące do najliczniejszej hiperkuli tworzą grupę obiektów i wyłączane są z dalszej procedury grupowania. W przypadku wystąpienia więcej niż jednej hiperkuli o maksymalnej liczbie obiektów, to pierwszą wyodrębnioną grupę obiektów tworzą obiekty należące do hiperkuli, której środek leży najbliżej początku układu współrzędnych. Środkami hiperkul są poszczególne grupowane obiekty. Algorytm ten powtarza się w stosunku do obiektów spoza wyodrębnionych grup obiektów, do momentu uzyskania rozłącznych grup obiektów [Panek, Zwierzchowski 2013].

Kolejną metodą należącą do metod obszarowych jest metoda katowicka. Jest to metoda hiperkostek, zaproponowana przez statystyków pracujących w ówczesnej Wyższej Szkole Ekonomicznej w Katowicach [Kolonko, Stolarska, Zadra 1970]. W tej metodzie, w odróżnieniu do metody wrocławskiej, nie wyznacza się elementów macierzy odległości pomiędzy obiektami.

Algorytm grupowania obiektów jest następujący. W pierwszym kroku zakres zmienności każdej ze zmiennych dzieli się na k z góry ustalonych, równych części (klas), czyli dokonuje się podziału na części zakresu zmienności wartości zmiennych na każdej z pól współrzędnych, wyznaczających przestrzeń klasyfikacji obiektów. W wyniku czego każdy z obiektów otrzymuje swój m -elementowy numer identyfikacyjny, w którym j -ty element to numer klasy j -tej zmiennej (j -tej osi współrzędnych przestrzeni klasyfikacji). Poszczególne elementy numeru identyfikacyjnego obiektów przyjmują wartości z przedziału $\langle 1; k \rangle$. W każdej z hiperkostek może znajdować się zero, jeden lub więcej niż jeden obiekt. Każdy obiekt otrzymuje numer identyfikacyjny hiperkostki, do której należy.

W kolejnym kroku grupowania poszukuje się pojedynczych albo połączonych ze sobą hiperkostek otoczonych hiperkostkami pustymi. Porównuje się ze sobą elementy składowe numerów identyfikacyjnych obiektów. Im więcej występuje różnic między numerami identyfikacyjnymi pary obiektów, tym podobieństwo tych obiektów jest mniejsze. Przyjmuje się, że dwa obiekty należą do tych samych lub sąsiednich kostek, gdy ich numery identyfikacyjne są identyczne lub różnią się co najwyżej jednym elementem.

Finalny efekt grupowania obiektów zależy od liczby klas, na które dzielone są zakresy zmiennych oraz liczby jednostkowych różnic numerów identyfikacyjnych obiektów. Liczba ta określa progowy stopień sąsiedztwa hiperkostek.

Przedstawiony algorytm może skutkować tym, że w grupach obiektów mogą znaleźć się hiperkostki znacznie od siebie oddalone w przestrzeni wielowymiarowej. Panek i Zwierzchowski [2013] zaproponowali taką modyfikację tej metody, aby w poszczególnej grupie obiektów znalazły się tylko obiekty należące do tych samych kostek lub będące w sąsiadujących ze sobą kostkach w danej przestrzeni wielowymiarowej.

5.1.3. Metody optymalizujące wstępny podział obiektów

W metodach optymalizujących wstępny podział obiektów pierwszym krokiem jest określenie pożądanej liczby grup obiektów (klas), które chcemy utworzyć. Ustalany jest również wstępny skład poszczególnych klas [Grabiński, Wydymus, Zeliaś 1989; Panek, Zwierzchowski 2013]:

- 1) w sposób losowy,
- 2) poprzez korzystanie ocen ekspertów,
- 3) poprzez wykorzystanie arbitralnie wybranej zmiennej,
- 4) poprzez grupowanie dokonane za pomocą dowolnej metody taksonomicznej,
- 5) poprzez porządkowanie obiektów według ich odległości od środka ciężkości poszczególnych grup obiektów.

Zadaniem metod optymalizujących jest poprawa wstępnego podziału obiektów, polegająca na przenoszeniu obiektów między klasami.

Najczęściej stosowaną metodą, optymalizującą wstępny podział obiektów, jest metoda *k*-średnich (ang. *k-means*). Klasyczny algorytm *k*-średnich został spopularyzowany przez Hartigana [Hartigan 1975; Hartigan, Wong 1979]. Podstawa algorytmu jest stosunkowo prosta. Jest to procedura iteracyjna.

Najczęściej wykorzystywany wariant metody *k*-średnich przebiega w czterech etapach [Witkowska 2002; Łuniewska, Tarczyński 2006; Panek, Zwierzchowski 2013; Zalewska 2017].

Etap 1. Ustala się arbitralnie maksymalną liczbę iteracji poprawiania dobroci grupowania oraz liczbę grup *k*, na jakie ma być podzielony analizowany zbiór obiektów, przy czym $k \in [2, N - 1]$, gdzie *N* oznacza liczbę obiektów. Podziału tego dokonuje się *a priori* lub w sposób losowy.

Etap 2. Dla każdego skupienia (grupy) obliczany jest środek ciężkości (zwany centroidem). Wyjściową macierz środków ciężkości zapisujemy jako:

$$\mathbf{G} = [g_{mj}] \quad (5.18)$$

dla $m = 1, 2, \dots, k, j = 1, 2, \dots, k$, gdzie *k* oznacza liczbę zmiennych.

Następnie obiekty przyporządkowuje się do poszczególnych grup. Dany obiekt przyporządkowujemy do tej grupy, dla której odległość tego obiektu od środka ciężkości dla danej grupy jest najmniejsza.

Etap 3. Wyznacza się wartość wyjściowego błędu podziału obiektów między s grup:

$$\varepsilon_m = \sum_{i=1}^n d_{im}^2, \quad (5.19)$$

gdzie d_{im} jest odległością euklidesową pomiędzy i -tym obiektem a najbliższym centroidem.

Etap 4. W sposób iteracyjny, przenosząc obiekty pomiędzy grupami, optymalizujemy wstępny podział obiektów w taki sposób, aby zminimalizować błędy grupowań. W tym celu dla pierwszego obiektu określa się zmianę błędu grupowania, wynikającą z przyporządkowania go kolejno do wszystkich aktualnie występujących grup:

$$\Delta\varepsilon_m^1 = \frac{n_m d_{1m}^2}{n_m+1} - \frac{n_{m_1} d_{1m_1}^2}{n_{m_1}+1}, \quad (5.20)$$

gdzie n_m jest liczbą obiektów w m -tej grupie, d_{1m} odległością pierwszego obiektu od środka ciężkości m -tej grupy, n_{m_1} liczebnością grupy zawierającej pierwszy obiekt, d_{1m_1} odległością pierwszego obiektu od najbliższego środka ciężkości. Jeżeli dla wszystkich $m \neq m_1$ minimalna wartość wyrażenia (5.20) jest ujemna, to obiekt 1 jest klasyfikowany do grupy, dla której $\Delta\varepsilon_m^1 = \min$. W kolejnym kroku, dla aktualnej grupy, przelicza się środki ciężkości grup oraz wyznaczana jest aktualna wartość błędu pomiaru ε_s . Jeśli dla bieżącej transformacji obiektu minimalna wartość wyrażenia $\Delta\varepsilon_m^1$ jest dodatnia lub równa 0, to nie następuje żadna zmiana.

Etapy 1-4 powtarza się dla każdego kolejnego obiektu. Kończy to pierwszą iterację. Jeżeli w danej iteracji nie następuje zmiana klas dla obiektów lub gdy osiągnięto założoną liczbę iteracji, to postępowanie jest zakończone. Jeśli natomiast nastąpi przesunięcie obiektów między grupami, to rozpoczyna się kolejną iterację.

Inna wersja metody k -średnich opiera się na środkach ciężkości klas. W metodzie tej optymalne grupowanie obiektów uzyskuje się poprzez minimalizację ich odległości od środków ciężkości klas. W początkowym etapie tej metody, dla każdej grupy, obliczany jest środek ciężkości. W kolejnym kroku każdy obiekt przypisywany jest do tej klasy, dla której odległość obiektu od środka ciężkości jest najmniejsza. Dla tak powstałych grup jest liczony nowy środek ciężkości. Jeśli podział na grupy nie uległ zmianie, to uznaje się go za optymalny. W przeciwnym przypadku kontynuowana jest procedura, aż nie nastąpi zmiana składu klas lub osiągnięta zostanie założona liczba iteracji [Panek, Zwierzchowski 2013].

Kolejna wersja metody k -średnich opiera się na maksymalizacji stosunku zmienności międzygrupowej do zmienności wewnątrzgrupowej, zamiast maksymalizacji spadku wartości błędu klasyfikacji [Panek, Zwierzchowski 2013].

Uogólniając w grupowaniu metodą k-średnich, staramy się tak przenosić obiekty między grupami, aby osiągnąć jak najmniejszą wariancję wewnątrz każdej z nich i jak największą wariancję między nimi, czyli aby uzyskać największy poziom istotności analizy dla prowadzonej analizy wariancji.

5.1.4. Weryfikacja uzyskanej klasyfikacji

Istotnymi problemami w analizie skupień jest ocena wyniku grupowania (ang. *cluster validity*), jak również weryfikacja wyników grupowania uzyskanych różnymi metodami analizy. Istnieje bardzo wiele metod grupowania i w konsekwencji jest również wiele metod oceny uzyskanej struktury grupowej.

Do oceny jakości podziału opracowano wiele wskaźników [Domański (red.) 2011; Migdał-Najman 2011], spośród których wyróżnić można indeks sylwetkowy (ang. *silhouette value*), zaproponowany przez Rousseeuwa [1987]. Indeks ten bada niepodobieństwo – odległość każdego analizowanego i -tego obiektu względem skupienia, do którego należy. Wskaźnik ten określony jest wzorem:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad (5.21)$$

gdzie $a(i)$ to średnia odległość obiektu i od pozostałych obiektów z tej samej grupy, a $b(i)$ średnia odległość obiektu i od obiektów z najbliższej sąsiedniej grupy.

Indeks $s(i)$ jest obliczany dla wszystkich elementów skupienia, co kwantyfikuje trafność przypisania każdego obiektu do danej grupy. Zakresem wartości, jakie może przyjmować wskaźnik $s(i)$, jest przedział $(-1; +1)$. Przy czym wartości bliskie $+1$ świadczą o poprawnym przypisaniu (silniejsza przynależność obiektu do klasy do której obiekt przypisano), natomiast bliskie -1 – o ewidentnie złym przypisaniu obiektu do danej klasy. Elementy mające wartość $s(i)$ bliskie 0 znajdują się na pograniczu grup.

Dla każdego skupienia C_q , otrzymane dla wszystkich n_q obiektów z tego skupienia, indeksy $s(i)$ są uśredniane:

$$S(q) = \frac{1}{n_q} \sum_{i=1, i \in C_q}^{n_q} s(i), \quad (5.22)$$

a następnie dla wszystkich analizowanych s skupień wyznacza się ogólną wartość indeksu sylwetkowego $SI(s)$:

$$SI(s) = \frac{1}{s} \sum_{q=1}^s S(q). \quad (5.23)$$

Wartość $SI(s) \geq 0,71$ świadczy o silnej strukturze uzyskanych grup, $SI(s) \in \langle 0,51, 0,70 \rangle$ potwierdza sensowną strukturę grupowania, $SI(s) \in \langle 0,26, 0,50 \rangle$ wskazuje, że wyodrębniona struktura jest słaba, a wynik przeprowadzonej analizy skupień może być sztuczny, natomiast $SI(s) \leq 0,25$

nakazuje odrzucić uzyskane rozwiązanie, wskazując na brak skupień w zbiorze [Kaufman, Rousseeuw 1990; Domański (red.) 2011].

W tej procedurze odległość pomiędzy obiektami może być obliczona dowolną metodą, ale należy zastosować tę samą miarę odległości, co w analizie, która jest oceniana [Migdał-Najman, Najman 2006; Domański (red.) 2011].

Do walidacji wyników analizy skupień można zastosować również tzw. replikację (tj. wielokrotne powtórzenie klasyfikacji). Replikacja dotyczy przeprowadzenia procesu klasyfikacji zbioru obiektów na podstawie dwóch prób wylosowanych z danego zbioru danych, a następnie na ocenie zgodności otrzymanych rezultatów [Walesiak 2008a, 2008b; Domański (red.) 2011].

5.2. Zastosowania metod klasyfikacji bezwzorcowej

Zastosowania analizy skupień w badaniach z zakresu ekonomiki rolnictwa są bardzo różnorodne, a ich przydatność wydaje się być duża. Przykłady użycia omówionych metod przedstawiono w tabeli 5.1.

Analizując pozycje literatury zestawione w tabeli 5.1, można zauważyć, że najczęściej wykorzystywaną metodą klasyfikacji była metoda Warda, a następnie metoda k-średnich (tylko w trzech pracach została użyta metoda pojedynczego wiązania i w jednej pracy metoda najdalszego sąsiada). Stosowano je zarówno w analizach regionalnych, ograniczonych do terytorium danego kraju, jak i w badaniach o szerszym zakresie, odnoszących się do obszaru Unii Europejskiej (UE) czy krajów OECD.

Niewątpliwie znaczącą grupę stanowią prace analizujące ekonomiczno-społeczne aspekty działalności gospodarstw rolnych w krajach UE. W ten nurt wpisują się prace Błażejczyk-Majki i Kali [2004, 2005], w których badane były zasoby siły roboczej rolnictwa polskiego i krajów UE czy zasoby ziemi charakteryzujące użytki rolne. Pietrzykowski i Kobus [2008] analizowali wyniki ekonomiczno-produkcyjne gospodarstw rolnych w wybranych państwach UE, a Pocza, Średzińska i Pawlak [2008] oraz Pocza, Średzińska i Standar [2008] oceniali sytuację finansową gospodarstw rolnych krajów UE według grup typologicznych gospodarstw, stworzonych na podstawie ich potencjału produkcyjnego oraz ich wyników produkcyjno-ekonomicznych. W pracy Matuszczak [2010] dokonano oceny taksonomicznej zróżnicowania zasobów czynników produkcji zaangażowanych w produkcję rolną oraz relacji zachodzących między nimi, a także efektywności ich wykorzystania w regionach UE-25. Ryś-Jurek [2012] analizowała miejsce polskiego rolnictwa w rolnictwie krajów UE, biorąc pod uwagę notowania gospodarstw rolnych uczestniczących w FADN. Majchrzak [2014] badaniom taksonomicznym poddał determinanty kształtowania struktur agrarnych w krajach UE. Ossowska i Janiszewska [2014] analizowały regionalne zróżnicowanie rolnictwa Polski, Czech, Słowacji i wschodnich Niemiec na podstawie cech potencjału produkcyjnego. Autorki te oceniały również zróżnicowanie zasobów ziemi czy funkcji lasów, a także sytuację na rynku pracy na obszarach wiejskich oraz relację

zatrudnienia i intensywność rolnictwa w krajach UE [Ossowska, Janiszewska 2015, 2016, 2017, 2018]. Wyrzykowski [2014] analizował popyt na żywność w krajach UE w okresie spowolnienia gospodarczego. Jankowska [2016] dokonała typologii rolnictwa krajów kandydujących do UE ze względu na wybrane cechy sektora rolnego. Smędzik-Ambroży [2016] analizie poddała uwarunkowania zasobowe i dochody rolnicze gospodarstw w krajach UE. Średzińska [2016] badała zróżnicowanie sytuacji gospodarstw rolnych krajów Unii UE (na podstawie FADN). Kryszak i Staniszewski [2017] analizowali zróżnicowanie krajów UE o różnej strukturze kosztów w rolnictwie w kontekście elastyczności dochodu w rolnictwie. Kryszak [2018] weryfikował efektywność dochodową gospodarstw rolnych w UE w ramach regionów FADN. Strojny i Musiał [2019] badali stan, zakres zróżnicowania i znaczenie obszarów wiejskich w poszczególnych krajach UE. Janiszewska i Ossowska [2015, 2017, 2018] analizowały zróżnicowanie krajów UE z punktu widzenia uwarunkowań rolnictwa dla produkcji energii odnawialnej z biomasy rolniczej, aspektów oddziaływania rolnictwa na środowisko oraz produkcji energii odnawialnej z rolnictwa i leśnictwa. Badania Guth [2015, 2016, 2017] i Śmigli [2014] koncentrowały się zaś na determinantach produkcji mleka w regionach UE i jej zróżnicowaniu.

Analizę skupień wykorzystano również do badania usług w krajach UE. Kołodziejczak [2015, 2016, 2020] dokonała oceny krajów ze względu na znaczenie i poziom korzystania z usług weterynaryjnych w rolnictwie oraz oceny korzystania z usług w gospodarstwach rolnych. Zalewski [2018] analizie poddał popyt na rolnicze usługi mechanizacyjne w krajach UE. Kraje UE Zalewski [2020] oceniał również ze względu na zmiany wartości zużytych nawozów i środków ochrony roślin.

W nurt badań o charakterze regionalnym wpisuje się Majewski [2005], który w swojej pracy poddał analizie zróżnicowanie skupu mleka w Polsce. Osowska [2006] oceniała typologię funkcjonalną obszarów wiejskich Pomorza Środkowego. Bezat [2008] koncentrowała się na kwestii spełniania przez polskich rolników wymogów dotyczących standardów jakości wytwarzanych produktów odnośnie ich identyfikowalności, podczas gdy Kowalska [2015] oceniała stan rolnictwa ekologicznego w Polsce w kontekście rozwoju zrównoważonej konsumpcji. Walory naturalnego środowiska jako determinant rozwoju turystyki wiejskiej analizowały Kałowska i Poczta-Wajda [2008]. Pietrzykowski i Wicki [2010] odnieśli się w swojej pracy do zmian dysproporcji regionalnych rolnictwa mierzonej poziomem nawożenia. Wąs i in. [2011] oceniali efektywność inwestowania w systemy automatyzacji udoju w regionie podlaskim, a Grużewska, Rymuza i Woch [2012] porównywali gospodarstwa rolne specjalizujące się w produkcji mleka w gminie Kąkolewnica. Kamińska i Nowak [2014] oraz Chrobościńska i Łukiewska [2015] oceniały zróżnicowanie regionalne potencjału produkcyjnego rolnictwa w Polsce. Karmowska i Marciniak [2015] badały zróżnicowanie przestrzenne rozwoju społeczno-gospodarczego obszarów wiejskich duoregionu Pomorze, tworzonego przez województwa zachodniopomorskie i pomorskie. Leszczyńska [2015] analizowała regionalne zróżnicowanie produkcji jęczmienia w Polsce, zaś Sułek i Jaśkiewicz [2015] w swoich badaniach określili wpływ czynników decydujących o regionalnym zróżnicowaniu produkcji pszenicy w Polsce. Guth i Smędzik-Ambroży [2017] podjęły się odpowiedzi na pytanie, czy na obszarach różniących się uwarunkowaniami

zasobowymi po integracji Polski z Unią Europejską występowały różnice w zakresie zrównoważenia ekonomicznego i środowiskowego. Z kolei Ciupa i Suligowski [2018] analizowali użytkowanie ziemi w kontekście stabilności ekologicznej obszarów wiejskich.

Poczta [2005] taksonomię numeryczną zastosował do oceny poziomu i struktury wsparcia finansowego rolnictwa w krajach OECD. Kiryluk-Dryjska [2008, 2009] dokonała oceny alokacji środków planu rozwoju obszarów wiejskich (PROW) oraz w ujęciu regionalnym skuteczności działań PROW 2004-2006 w opiniach rolników i ekspertów. Czyżewski i Kułyk [2009] w swoich badaniach podjęli problem konwergencji i dywergencji mechanizmów wsparcia sektora rolnego w 27 krajach świata. Grzelak i Kiełbasa [2014] oceniali wykorzystanie środków z UE do wsparcia inwestycji w gospodarstwach rolnych w regionalnej perspektywie, a Sieczko i Sieczko [2014] dokonali wielowymiarowej analizy pozyskiwania środków finansowych w ramach działania 132 PROW 2007-2013 przez kraje UE.

Kisielińska [2003] badała zróżnicowanie gospodarstw rolniczych, a Adamowicz i Nowak [2006] analizowali typy wiejskich gospodarstw domowych. Kwasek [2010] wyznaczała wzorce konsumpcji żywności w grupach społeczno-ekonomicznych gospodarstw domowych. Smędzik-Ambroży [2015] porównała uwarunkowania zasobowe i dochody indywidualnych gospodarstw rolnych w Polsce. Kozera [2018] oceniała własny potencjał dochodowy gmin wiejskich w Polsce. Szafrąńska i Żmija [2017] zajęli się zagadnieniem typologią strategii oszczędnościowych wiejskich gospodarstw domowych, zaś Maciejewski [2020] oceniał gospodarstwa domowe ze względu na ich zachowania na rynku żywności.

Analiza skupień wykorzystywana była również w analizie sektora przemysłowego. Wysocki [1999] oceniał strategie marketingowe w polskim przemyśle mleczarskim. Szczepaniak i Wigier [2003] analizowali innowacyjność małych i bardzo małych firm przemysłu spożywczego. Pawlak i Poczta [2008] badali pozycję konkurencyjności polskiego sektora rolno-spożywczego na jednolitym rynku europejskim. Prevužňáková i Zemková [2011] analizowały zadłużenie przedsiębiorstw produkcji rolnej na Słowacji, podczas gdy Kontsevaya [2017] oceniała efektywność wykorzystania rządowych subwencji inwestycyjnych w rolnictwie w Federacji Rosyjskiej. Urban i Kowalska [2015] poddali badaniu inwestycje oraz podstawowe środki trwałe w rolnictwie.

Angowski, Domańska i Kijek [2017] dokonali segmentacji nabywców na rynku produktów spożywczych, a Czyżewski i Czakowski [2018] analizowali podobieństwo rynków poszczególnych produktów rolnych w Polsce po wstąpieniu do UE.

Przedstawiony przegląd literatury ukazuje szeroki wachlarz wykorzystania poszczególnych metod analizy skupień zarówno w ekonomice rolnictwa, jak i gospodarce żywnościowej.

Tabela 5.1. Przykłady zastosowania klasyfikacji bezwzorcowej w problemach z zakresu ekonomiki rolnictwa

Autor	Problem	Obiekty	Cechy	Metoda
Wysocki 1999	strategie marketingowe w polskim przemyśle mleczarskim	143 polskich mleczarni	34 cechy	metoda Warda
Kisielińska 2003	zróźnicowanie gospodarstw rolniczych	998 gospodarstw	16 wskaźników finansowych	sieć neuronowa Kohonena
Szczepaniak, Wigier 2003	innowacyjność małych i bardzo małych firm przemysłu spożywczego	36 przedsiębiorstw	około 500 cech pochodzących z ankiet	metoda k-średnich
Błażejczyk-Majka, Kala 2004	zasoby siły roboczej rolnictwa polskiego i krajów UE	15 państw członków UE w 2000 roku i Polska	3 cechy	metoda najdalszego sąsiada
Błażejczyk-Majka, Kala 2005	charakterystyka użytków rolnych wybranych państw UE	15 państw członków UE w 2000 roku i Polska	5 cech opisujących zasoby ziemi	metoda pojedynczego i pełnego wiązania
Majewski 2005	regionalne zróźnicowanie skupu mleka w Polsce	16 województw	6 cech	metoda Warda
Poczta 2005	poziom i struktura wsparcia finansowego rolnictwa w krajach OECD	13 państw	10 cech	metoda Warda
Adamowicz, Nowak 2006	typy wiejskich gospodarstw domowych	611 wiejskich gospodarstw domowych z województwa lubelskiego	4 cechy	metoda k-średnich
Osovska 2006	typologia funkcjonalna obszarów wiejskich Pomorza Środkowego	65 gmin wiejskich i wiejsko-miejskich Pomorza Zachodniego	7 cech	metoda Warda
Bezat 2008	ocena spełniania przez polskich rolników wymogów dotyczących standardów jakości wytwarzanych produktów w odnośnie ich identyfikowalności	gospodarstwa rolne	4 cechy	metoda Warda
Kiryłuk-Dryjska 2008	ocena alokacji środków planu rozwoju obszarów wiejskich (PROW) przy wykorzystaniu metod optymalizacyjnych	800 ankiet	9 cech	metoda Warda
Kałowska, Poczta-Wajda 2008	walory naturalnego środowiska jako determinant rozwoju turystyki wiejskiej w Wielkopolsce	31 powiatów województwa wielkopolskiego	5 cech	metoda Warda
Pawlak, Poczta 2008	pozycja konkurencyjna polskiego sektora rolno-spożywczego na jednolitym rynku europejskim	12 grup towarów	2 cechy	metoda Warda
Pietrzykowski, Kobus 2008	analiza wyników ekonomiczno-produkcyjnych gospodarstw rolnych wybranych państw UE	25 państw	4 cechy	metoda pojedynczego wiązania, metoda Warda, metoda k-średnich

Poczta, Średzińska, Standar 2008	ocena sytuacji finansowej gospodarstw rolnych krajów UE według ich potencjału produkcyjnego	gospodarstwa rolne krajów UE	5 cech	metoda Warda
Poczta, Średzińska, Pawlak 2008	analiza sytuacji finansowej gospodarstw rolnych państw UE według ich wyników produkcyjno-ekonomicznych	gospodarstwa rolne 23 państw UE	14 cech	metoda Warda
Czyżewski, Kułyk 2009	problem konwergencji i dywergencji mechanizmów wsparcia sektora rolnego	27 państwa	3 cechy	metoda Warda
Kiryłuk-Dryjska 2009	ocena skuteczności działań PROW 2004-2006 w opinii rolników i ekspertów (ujęcie regionalne)	792 ankiety	9 cech	metoda Warda
Kisielińska 2009	prezentacja bezwzorcowej klasyfikacji obiektów w ekonomice rolnictwa	dane hipotetyczne	dane hipotetyczne	prezentacja metod
Kwasek 2010	wyznaczanie wzorców konsumpcji żywności metodą Warda	7 grup społeczno-ekonomicznych gospodarstw domowych	30 cech	metoda Warda
Matuszczak 2010	ocena taksonomiczna zróżnicowania zasobów czynników produkcji zaangażowanych w produkcję rolną oraz relacji zachodzących między nimi, a także efektywności ich wykorzystania w regionach UE-25	gospodarstwa rolne ze 122 regionów UE	cechy opisujące zaangażowanie czynnika ziemi, pracy i kapitału oraz ich efektywność	metoda Warda
Pietrzykowski, Wicki 2010	ocena dynamiki zmian dysproporcji regionalnych rolnictwa mierzona poziomem nawożenia	16 województw	2 cechy	metoda k-średnich
Prevužňáková, Zemková 2011	analiza zadłużenia przedsiębiorstw rolnych na Słowacji	przedsiębiorstwa rolne	4 cechy	metoda Warda
Wąs i in. 2011	ocena efektywności inwestowania w systemy automatyzacji udoju w regionie podlaskim	gospodarstwa mleczne	10 cech	metoda Warda, metoda k-średnich
Grużewska i in. 2012	analiza porównawcza gospodarstw rolnych w gminie Kąkolewnica	gospodarstwa mleczne	14 cech	metoda pojedynczego wiązania
Ryś-Jurek 2012	miejsce polskiego rolnictwa w rolnictwie UE-27 według notowań gospodarstw rolnych uczestniczących w FADN w 2009 roku	kraje UE	30 cech	metoda Warda
Grzelak, Kiełbasa 2014	analiza wykorzystania środków z UE do wsparcia inwestycji w gospodarstwach rolnych w regionalnej perspektywie	województwa	3 cechy	metoda Warda
Kamińska, Nowak 2014	ocena zróżnicowania regionalnego potencjału produkcyjnego rolnictwa w Polsce	16 województw	11 cech	metoda Warda
Majchrzak 2014	determinanty kształtowania struktur agrarnych	kraje UE	6 cech	metoda Warda
Ossowska, Janiszewska 2014	regionalne zróżnicowanie rolnictwa Polski, Czech, Słowacji i wschodnich Niemiec na podstawie wybranych cech potencjału produkcyjnego	31 regionów	4 cechy	metoda Warda
Sieczko, Sieczko 2014	wielowymiarowa analiza pozyskiwania środków finansowych w ramach działania 132 PROW 2007-2013 przez kraje członkowskie UE	kraje UE	11 cech	metoda Warda

Tabela 5.1 (cd.)

Autor	Problem	Obiekty	Cechy	Metoda
Śmigła 2014	badanie zróżnicowania produkcji mleka w makroregionach UE	regiony UE	4 cechy	metoda Warda
Wyrzykowski 2014	analiza popytu na żywność w krajach UE w warunkach spowolnienia gospodarczego	kraje UE	4 cechy	metoda Warda
Chrobocińska, Łukiewska 2015	przestrzenne zróżnicowanie potencjału produkcyjnego rolnictwa w Polsce	16 województw	6 cech	metoda Warda
Guth 2015	badanie zróżnicowania gospodarstw mlecznych FADN w regionach UE	gospodarstwa mleczne regionów EU	3 cechy	metoda Warda
Janiszewska, Ossowska 2015	zróżnicowanie uwarunkowań rolnictwa dla produkcji energii odnawialnej z biomasy rolniczej	kraje UE	6 cech	metoda Warda
Jaśkiewicz 2015	regionalne zróżnicowanie produkcji gryki w Polsce	16 województw	19 cech	metoda Warda
Karmowska, Marciniak 2015	ocena zróżnicowania przestrzennego rozwoju społeczno-gospodarczego obszarów wiejskich duoregionu Pomorze	powiaty	5 cech	metoda k-średnich
Kołodziejczak 2015	ocena krajów UE ze względu na znaczenie i poziom korzystania z usług weterynaryjnych w rolnictwie	kraje UE	7 cech	metoda Warda
Kowalska 2015	ocena stanu rolnictwa ekologicznego w Polsce w kontekście rozwoju zrównoważonej konsumpcji	województwa	3 cechy	metoda Warda
Kozera, Wysocki 2015	typ funkcjonalny a samodzielność finansowa gmin wiejskich	117 gmin wiejskich województwa wielkopolskiego	9 cech	metoda Warda, metoda k-średnich
Kułyk 2015	relacje strukturalne w finansowym wsparciu rolnictwa w grupach państw o różnym poziomie rozwoju gospodarczego	48 państw	3 cechy	metoda Warda
Leszczyńska 2015	regionalne zróżnicowanie produkcji jęczmienia w Polsce	16 województw	18 cech	metoda Warda
Ossowska, Janiszewska 2015	zróżnicowanie zasobów ziemi	kraje UE	5 cech	metoda Warda
Smędzik-Ambroży 2015	uwarunkowania zasobowe a dochody indywidualnych gospodarstw rolnych w Polsce w latach 2004-2012 (próba oceny zbieżności na podstawie doświadczeń gospodarstw FADN)	województwa (regiony) w Polsce (gospodarstwa FADN)	9 cech	metoda Warda
Sulek, Jaśkiewicz 2015	określenie wpływu czynników decydujących o regionalnym zróżnicowaniu produkcji pszenicy w Polsce	16 województw	20 cech	metoda k-średnich
Urban, Kowalska 2015	ocena inwestycji oraz podstawowych środków trwałych w rolnictwie	16 województw	3 cechy	metoda Warda

Guth 2016	ekonomiczne determinanty produkcji mleka	gospodarstwa mleczne w makroregionach UE (gospodarstwa FADN)	3 cechy	metoda Warda
Jankowska 2016	typologia rolnictwa krajów kandydujących do UE ze względu na wybrane cechy sektora rolnego	8 krajów kandydujących oraz potencjalne kraje kandydujące do UE	4 cechy (początkowo 7 cech)	metoda Warda
Kołodziejczak 2016	ocena korzystania z usług w gospodarstwach rolnych	kraje UE	6 cech	metoda Warda
Ossowska, Janiszewska 2016	zróżnicowanie funkcji lasów	kraje UE	6 cech	metoda Warda
Smędzik-Ambroży 2016	uwarunkowania zasobowe a dochody rolnicze	kraje UE	3 cechy	metoda Warda
Średzińska 2016	badanie zróżnicowania sytuacji gospodarstw rolnych krajów UE	gospodarstwa rolne krajów UE	25 cech	metoda Warda
Angowski, Domańska, Kijek 2017	segmentacja nabywców na rynku produktów spożywczych	358 studentów	18 cech	metoda Warda, metoda k-średnich
Czyżewski, Kryszak 2017	wpływ typów rolnictwa na emisję gazów cieplarnianych	40 kraje świata	3 cechy	metoda Warda
Guth 2017	określenie determinant produkcji mleka w skupieniach gospodarstw FADN o podobnych cechach	gospodarstwa mleczne FADN	3 cechy	metoda Warda
Guth, Smędzik-Ambroży 2017	zasoby a zrównoważony rozwój rolnictwa w Polsce po integracji z UE	gospodarstwa rolne w Polsce	4 cechy	metoda Warda
Janiszewska, Ossowska 2017	wybrane aspekty oddziaływania rolnictwa na środowisko	kraje UE	6 cech	metoda Warda
Kontsevaya 2017	ocena efektywności wykorzystania rządowych subwencji inwestycyjnych w rolnictwie w Federacji Rosyjskiej	regiony Federacji Rosyjskiej	7 cech	metoda k-średnich
Kryszak, Staniszewski 2017	analiza zróżnicowania krajów UE o różnej strukturze kosztów w kontekście elastyczności dochodu w rolnictwie	kraje UE	21 cech	metoda Warda
Ossowska, Janiszewska 2017	sytuacja na rynku pracy na obszarach wiejskich	kraje UE	4 cechy	metoda Warda
Szafrańska, Żmija 2017	typologia strategii oszczędnościowych wiejskich gospodarstw domowych	384 respondentów województwa małopolskiego	7 cech	metoda k-średnich
Ciupa, Suligowski 2018	użytkowanie ziemi a stabilność ekologiczna obszarów wiejskich	69 gmin wiejskich oraz 28 miejsko-wiejskich województwa świętokrzyskiego	4 cechy	metoda Warda

Tabela 5.1 (cd.)

Autor	Problem	Obiekty	Cechy	Metoda
Czyżewski, Czakowski 2018	ocena podobieństwa rynków produktów rolnych w Polsce po akcesji do UE	rynki produktów rolnych	5 cech	metoda Warda, metoda k-średnich
Janiszewska, Ossowska 2018	analiza zróżnicowania krajów UE z punktu widzenia produkcji energii odnawialnej z rolnictwa i leśnictwa	kraje UE	6 cech	metoda Warda
Kozera 2018	analiza kształtowania się własnego potencjału dochodowego gmin wiejskich w Polsce	gminy wiejskie	17 cech	metoda Warda, metoda k-średnich
Kryszak 2018	weryfikacja efektywności dochodowej gospodarstw rolnych UE w ramach regionów FADN	regiony FADN	3 cechy	metoda Warda
Ossowska, Janiszewska 2018	zatrudnienie a intensywność rolnictwa	kraje UE	6 cech	metoda Warda
Zalewski 2018	popyt na rolnicze usługi mechanizacyjne	kraje UE	2 cechy	metoda Warda
Strojny, Musiał 2019	ocena stanu, zakres zróżnicowania i znaczenia obszarów wiejskich w krajach UE	kraje UE	4 cechy	metoda Warda
Kołodziejczak 2020	wykorzystanie usług rolniczych w regionach UE	regiony UE	6 cech	metoda Warda
Maciejewski 2020	typy gospodarstw domowych ze względu na ich zachowania na rynku żywności	900 gospodarstw domowych	11 cech	metoda Warda, metoda k-średnich
Zalewski 2020	zmiany wartości zużytych nawozów i środków ochrony roślin	kraje UE	2 cechy	metoda Warda

Źródło: opracowanie własne

Literatura

- Adamowicz M., Nowak A. (2006) Charakterystyczne typy wiejskich gospodarstw domowych na przykładzie województwa lubelskiego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 8(4), 14-18.
- Angowski M., Domańska K., Kijek T. (2017) Zastosowanie metody k-średnich w segmentacji nabywców na rynku produktów spożywczych. *Metody Ilościowe w Badaniach Ekonomicznych*, 18(4), 521-530.
- Bezat A. (2008) Polish farms in the light of quality requirements. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 3(18), 7-13.
- Błażejczyk-Majka L., Kala R. (2004) Porównanie zasobów siły roboczej rolnictwa polskiego i krajów UE w latach 1990-1999. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 6(5), 10-15.
- Błażejczyk-Majka L., Kala R. (2005) Metody analizy skupień do charakterystyki użytków rolnych wybranych państw Unii Europejskiej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 7(5), 5-10.
- Bukietyński W., Hellwig Z., Królik U., Smoluk A. (1969) Uwagi o dyskryminacji zbiorów skończonych. *Prace Naukowe WSE we Wrocławiu*, 21, 111-122.
- Chrobocińska K., Łukiewska K. (2015) Przestrzenne zróżnicowanie potencjału produkcyjnego rolnictwa w Polsce. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 102(3), 56-65.
- Ciupa T., Suligowski R. (2018) Użytkowanie ziemi a stabilność ekologiczna obszarów wiejskich województwa świętokrzyskiego. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 105(1), 18-31.
- Czyżewski A., Czakowski D. (2018) Similarities and differences in the core markets for agricultural products in Poland after the accession to the European Union. *Journal of Agribusiness and Rural Development*, 1(47), 3-19.
- Czyżewski A., Kułyk P. (2009) Konwergencja czy dywergencja mechanizmów wsparcia sektora rolnego? *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 8(23), 41-51.
- Czyżewski B., Kryszak Ł. (2017) Wpływ typów rolnictwa na emisję gazów cieplarnianych. *Wieś i Rolnictwo*, 1(174), 99-122. DOI: 10.7366/wir012017/05
- Domański Cz. (red.) (2011) *Nieklasyczne metody oceny efektywności i ryzyka. Otwarte fundusze emerytalne.* PWE, Warszawa.
- Everitt B.S., Landau S., Leese M., Stahl D. (2011) *Cluster Analysis* (5th edition). Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.
- Gan G., Ma C., Wu J. (2007) *Data clustering: theory, algorithms, and applications.* SIAM, Philadelphia.
- Gatnar E., Walesiak M. (red.) (2004) *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych.* Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Gordon A.D. (1999) *Classification.* Chapman & Hall/CRC, London.
- Grabiński T., Wydymus S., Zeliaś A. (1989) *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych.* Wydawnictwo Naukowe PWN, Warszawa.
- Grużewska A., Rymuza K., Woch S. (2012) An application of cluster analysis to compare selected agricultural holdings in the Kąkolewnica commune. *Acta Scientiarum Polonorum, Oeconomia*, 11(1), 47-59.
- Grzelak A., Kiełbasa B. (2014) Assessment of the use of the European Union funds to support investments on Polish farms in the regional perspective. *Acta Scientiarum Polonorum, Oeconomia*, 13(2), 49-60.
- Guth M. (2015) Diversity of FADN milk farms in the regions of the European Union in 2011. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(3), 119-124.
- Guth M. (2016) Ekonomiczne determinanty produkcji mleka w gospodarstwach mlecznych FADN w makroregionach Unii Europejskiej z przewagą ekstensywnej produkcji w 2011 roku. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(2), 114-123.
- Guth M. (2017) Determinants of milk production diversity in the macroregions of the European Union. *Acta Scientiarum Polonorum, Oeconomia*, 16(1), 33-42.

- Guth M., Smędzik-Ambroży K. (2017) Zasoby a zrównoważony rozwój rolnictwa w Polsce po integracji z UE. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(32)(3), 101-110. DOI: 10.22630/PRS.2017.17.3.57
- Hartigan J.A. (1975) *Clustering Algorithms (Probability & Mathematical Statistics)*. John Wiley & Sons, Chicester.
- Hartigan J.A., Wong M.A. (1979) Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, 28(1), 100-108.
- Jajuga K. (1990) *Statystyczna teoria rozpoznawania obrazów*. Wydawnictwo Naukowe PWN, Warszawa.
- Janiszewska D.A., Ossowska L. (2015) Zróżnicowanie uwarunkowań rolnictwa dla produkcji energii odnawialnej z biomasy rolniczej w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 15(30)(2), 75-84.
- Janiszewska D., Ossowska L. (2017) Wybrane aspekty oddziaływania rolnictwa na środowisko w państwach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(32)(1), 73-83. DOI: 10.22630/PRS.2017.17.1.7
- Janiszewska D., Ossowska L. (2018) Diversification of European Union member states due to the production of renewable energy from agriculture and forestry. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(33)(2), 95-104.
- Jankowska A. (2016) Typologia rolnictwa krajów kandydujących do Unii Europejskiej ze względu na wybrane cechy sektora rolnego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(2), 145-153.
- Jaśkiewicz B. (2015) Czynniki decydujące o regionalnym zróżnicowaniu produkcji gryki w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(5), 77-83.
- Kałowska J., Poczta-Wajda A. (2008) Walory naturalnego środowiska jako determinant rozwoju turystyki wiejskiej w Wielkopolsce. *Więś i Rolnictwo*, 2 (139), 115-128.
- Kamińska A., Nowak A. (2014) Zastosowanie analizy skupień do badania zróżnicowania regionalnego potencjału produkcyjnego rolnictwa w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 16(3), 126-130.
- Karmowska G., Marciniak M. (2015) Analiza zróżnicowania przestrzennego rozwoju społeczno-gospodarczego duoregionu Pomorze z wykorzystaniem metody k-średnich. *Polityki Europejskie, Finanse i Marketing*, 13(62), 63-72.
- Kaufman L., Rousseeuw P.J. (1990) *Finding groups in data: an introduction to cluster analysis*. Wiley, New York.
- Kiryłuk-Dryjska E. (2008) Ocena alokacji Środków Planu Rozwoju Obszarów Wiejskich przy wykorzystaniu metod optymalizacyjnych. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 4(19), 246-258.
- Kiryłuk-Dryjska E. (2009) Ocena skuteczności działań PROW 2004–2006 w opinii rolników i ekspertów (ujęcie regionalne). *Więś i Rolnictwo*, 1 (142), 132-145.
- Kisielińska J. (2003) Klasyfikacja gospodarstw rolniczych siecią neuronową Kohonena w oparciu o wybrane wskaźniki finansowe. *Zagadnienia Ekonomiki Rolnej*, 2, 49-64.
- Kisielińska J. (2009) Bezwzorcowo klasyfikacja obiektów w ekonomice rolnictwa. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 8(23), 104-115.
- Kolonko J., Stolarska E., Zadra K. (1970) Prosta metoda dyskryminacji zbiorów skończonych. *Przegląd Statystyczny*, 2, 173-186.
- Kołodziejczak M. (2015) Typologia krajów Unii Europejskiej ze względu na znaczenie i poziom korzystania z usług weterynaryjnych w rolnictwie. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 15(30)(1), 39-48.
- Kołodziejczak M. (2016) Ocena korzystania z usług w gospodarstwach rolnych krajów Unii Europejskiej – analiza typologiczna. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(3), 192-198.
- Kołodziejczak M. (2020) Wykorzystanie usług rolniczych w regionach Unii Europejskiej – analiza typologiczna. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 22(1), 159-169.
- Kontsevaya S. (2017) Analysis and control of government subsidies for investments in agriculture of the Russian Federation. *Acta Scientiarum Polonorum, Oeconomia*, 16(3), 25-33.

- Kowalska A. (2015) Rolnictwo ekologiczne jako czynnik rozwoju zrównoważonej konsumpcji. *Journal of Agribusiness and Rural Development*, 3(37), 467-476.
- Kozera A. (2018) Własny potencjał dochodowy gmin wiejskich w Polsce. *Polityki Europejskie, Finanse i Marketing*, 19(68), 95-106.
- Kozera A., Wysocki F. (2015) Typ funkcjonalny a samodzielność finansowa gmin wiejskich województwa wielkopolskiego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(6), 133-139.
- Kryszak Ł. (2018) Profit efficiency in EU FADN farms under different types of agriculture. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(33)(3), 196-207.
- Kryszak Ł., Staniszewski J. (2017) The elasticity of agricultural income in the EU member states under different cost structures. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(32)(4), 182-192.
- Kułyk P. (2015) Relacje strukturalne w finansowym wsparciu rolnictwa w grupach państw o różnym poziomie rozwoju gospodarczego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(6), 140-145.
- Kwasek M. (2010) Wyznaczanie wzorców konsumpcji żywności metodą Warda. *Wiadomości Statystyczne*, 11(594), 31-46.
- Leszczyńska D. (2015) Regionalne zróżnicowanie produkcji jęczmienia w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(6), 151-156.
- Łuniewska M., Tarczyński W. (2006) *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*. Wydawnictwo Naukowe PWN, Warszawa.
- Maciejewski G. (2020) Typy gospodarstw domowych ze względu na ich zachowania na rynku żywności. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 20(1), 52-66. DOI: 10.22630/PRS.2020.20.1.5
- Majchrzak A. (2014) Determinanty kształtowania struktur agrarnych w państwach członkowskich Unii Europejskiej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 16(4), 176-183.
- Majewski J. (2005) Regionalne zróżnicowanie skupu mleka w Polsce oraz czynniki je determinujące. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 7(5), 56-60.
- Matuszczak A. (2010) Alokacja czynników wytwórczych a wyniki działalności rolniczej w regionach rolnych UE-25. Ocena taksonomiczna. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 10(25)(2), 71-79.
- Migdał-Najman K. (2011) Ocena jakości wyników grupowania – przegląd bibliografii. *Przegląd Statystyczny*, 57(3-4), 281-299.
- Migdał-Najman K., Najman K. (2006) Wykorzystanie indeksu silhouette do ustalania optymalnej liczby skupień. *Wiadomości Statystyczne*, 6, 1-10.
- Ossowska L. (2006) Typologia funkcjonalna obszarów wiejskich Pomorza Środkowego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 7(4), 254-257.
- Ossowska L., Janiszewska D.A. (2014) Regionalne zróżnicowanie rolnictwa Polski, Czech, Słowacji i wschodnich Niemiec na podstawie wybranych cech potencjału produkcyjnego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 14(29)(2), 161-169.
- Ossowska L., Janiszewska D.A. (2015) Zróżnicowanie zasobów ziemi w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 15(30)(3), 102-111.
- Ossowska L., Janiszewska D.A. (2016) Zróżnicowanie funkcji lasów w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(3), 292-300.
- Ossowska L., Janiszewska D. (2017) Sytuacja na rynku pracy na obszarach wiejskich państw Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(32)(3), 176-185. DOI: 10.22630/PRS.2017.17.2.37
- Ossowska L., Janiszewska D. (2018) Zatrudnienie a intensywność rolnictwa państw Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(3), 238-247. DOI: 10.22630/PRS.2018.18.3.82

- Panek T., Zwierzchowski J. (2013) Statystyczne metody wielowymiarowej analizy porównawczej. Teoria i zastosowanie. Oficyna Wydawnicza SGH, Warszawa.
- Pawlak K., Poczta W. (2008) Pozycja konkurencyjna polskiego sektora rolno-spożywczego na jednolitym rynku europejskim. *Więś i Rolnictwo*, 4(141), 81-102.
- Pietrzykowski R., Kobus P. (2008) Wielowymiarowe metody statystyczne w analizie wyników ekonomiczno-produkcyjnych gospodarstw rolnych wybranych państw Unii Europejskiej. *Problemy Rolnictwa Światowego*, 4(19), 371-388.
- Pietrzykowski R., Wicki L. (2010) Dynamika zmian dysproporcji regionalnych rolnictwa mierzona poziomem nawożenia. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 12(3), 317-323.
- Pociecha J., Podolec B., Sokołowski A., Zajac K. (1988) Metody taksonomiczne w badaniach społeczno-ekonomicznych. Wydawnictwo Naukowe PWN, Warszawa.
- Poczta A. (2005) Poziom i struktura wsparcia finansowego rolnictwa w krajach OECD po powstaniu WTO. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 7(7A), 221-227.
- Poczta W., Średzińska J., Pawlak K. (2008) Sytuacja finansowa gospodarstw rolnych krajów UE sklasyfikowanych według ich wyników produkcyjno-ekonomicznych. *Problemy Rolnictwa Światowego*, 4(19), 379-387.
- Poczta W., Średzińska J., Standar A. (2008) Sytuacja finansowa gospodarstw rolnych krajów UE według potencjału produkcyjnego. *Journal of Agribusiness and Rural Development*, 4(10), 83-94.
- Prevužňáková J., Zemková K. (2011) The analysis of the indebtedness of Slovak agricultural enterprises in years 2003 and 2009 and multi-criteria analysis of Slovak agricultural enterprises in years 2003 and 2009. *Polityki Europejskie, Finanse i Marketing*, 7(56), 92-101.
- Rousseeuw P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(1), 53-65.
- Ryś-Jurek R. (2012) Miejsce polskiego rolnictwa w rolnictwie UE-27 według notowań gospodarstw rolnych uczestniczących w FADN w 2009 roku. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 12(27)(4), 124-134.
- Sieczko A., Sieczko L. (2014) Wielowymiarowa analiza pozyskiwania środków finansowych w ramach działania 132 PROW 2007-2013 przez kraje członkowskie Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 14(29)(2), 214-225.
- Smędzik-Ambroży K. (2015) Uwarunkowania zasobowe a dochody indywidualnych gospodarstw rolnych w Polsce w latach 2004-2012 (próba oceny zbieżności na podstawie doświadczeń gospodarstw FADN). *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(4), 281-287.
- Smędzik-Ambroży K. (2016) Uwarunkowania zasobowe a dochody rolnicze w krajach UE. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(2), 314-323.
- Sokołowski A. (1992) Empiryczne testy jednorodności w taksonomii. *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, Monografie*, 108.
- Strojny J., Musiał W. (2019) Obszary wiejskie w systemie ekonomiczno-społecznym państw Unii Europejskiej. *Polityki Europejskie, Finanse i Marketing*, 22(71), 193-209.
- Sułek A., Jaśkiewicz B. (2015) Regionalne zróżnicowanie produkcji pszenicy w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(4), 308-313.
- Szafrańska M., Żmija J. (2017) Typologia strategii oszczędnościowych wiejskich gospodarstw domowych (na przykładzie województwa małopolskiego). *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 118, 19-30.
- Szczepaniak I., Wigier M. (2003) Identyfikacja czynników wpływających na innowacyjność małych i bardzo małych firm przemysłu spożywczego. *Zagadnienia Ekonomiki Rolnej*, 4, 47-61.
- Śmigła M. (2014) Determinanty produkcji mleka w regionach Unii Europejskiej o bardzo dużych gospodarstwach mlecznych po 2004 roku. *Journal of Agribusiness and Rural Development*, 1(31), 143-150.
- Średzińska J. (2016) The diversification of the economic situation of the EU countries' farms (based on FADN). *Journal of Agribusiness and Rural Development*, 4(42), 669-677.
- Thorndike R.L. (1953) Who belongs in the family? *Psychometrika*, 18(4), 267-276.

- Urban S., Kowalska A. (2015) Inwestycje oraz podstawowe środki trwałe w rolnictwie. *Wiadomości Statystyczne*, 9(652), 66-76.
- Walesiak M. (2008a) Ocena stabilności wyników klasyfikacji z wykorzystaniem analizy replikacji. Modelowanie i prognozowanie zjawisk społeczno-gospodarczych. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków, 67-72.
- Walesiak M. (2008b) Procedura analizy skupień z wykorzystaniem programu komputerowego ClusterSim i środowiska R. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia*, 15, 7(1207), 44-56.
- Wąs A., Majewski E., Cygański Ł., Bartolini F., Floridi M., Viaggi D. (2011) Assessment of economic effects of innovations in automatic milking systems in Podlaskie region (Poland) with the use of real option approach. *Acta Scientiarum Polonorum, Oeconomia*, 10(2), 107-120.
- Witkowska D. (2002) *Sztuczne sieci neuronowe i metody statystyczne*. Wydawnictwo C.H. Beck, Warszawa.
- Wyrzykowski P. (2014) Popyt na żywność w krajach UE w warunkach spowolnienia gospodarczego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 14(29)(1), 130-140.
- Wysocki F. (1999) Metody statystyczne w badaniu strategii marketingowych w polskim przemyśle mleczarskim. *Roczniki Nauk Rolniczych, Seria G*, 88(1), 61-76.
- Zalewska E. (2017) Zastosowanie analizy skupień i metody porządkowania liniowego w ocenie polskiego szkolnictwa wyższego. *Prace Naukowe Uniwersytetu Ekonomicznego We Wrocławiu, Taksonomia* 29, *Klasyfikacja i analiza danych – teoria i zastosowania*, 469, 234-242.
- Zalewski A. (2018) Popyt na rolnicze usługi mechanizacyjne w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(4), 50-59. DOI: 10.22630/PRS.2018.18.4.96
- Zalewski A. (2020) Zmiany wartości zużytych nawozów i środków ochrony roślin w krajach Unii Europejskiej w latach 2010-2018. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 20(1), 78-87. DOI: 10.22630/PRS.2020.20.1.7

6. Przestrzenna analiza taksonomiczna

Określenie analiza taksonomiczna wywodzi się z nauk przyrodniczych, a ujęcie taksonomii w sposób ilościowy przypisuje się niemieckiemu przyrodnikowi F. Heinckemu, zaś w Polsce antropologowi J. Czekanowskiemu [1913]. Początki taksonomii związane są z naukami przyrodniczymi [Sokal, Sneath 1963], antropologicznymi [Czekanowski 1930] i psychologicznymi [Choynowski 1971; Brzeziński 1997]. Dopiero w późniejszym okresie wprowadzono ich zastosowanie do nauk technicznych [Deptuła 2017], ekonomicznych [Fierich 1957; Pocięcha i in. 1988; Walesiak 2004], geograficznych [Chojnicki, Czyż 1973; Chojnicki i in. 1978] i innych. Popularność i rozwój metod klasyfikacyjnych przypadł na lata 50. i 60. XX wieku. W tym okresie powstała również metoda taksonomii wrocławskiej [Florek i in. 1951]. W wyniku rozwoju procedur taksonomicznych w różnych dziedzinach nastąpiły zmiany w nazewnictwie. W literaturze społeczno-ekonomicznej zamiast określenia taksonomia wprowadzono pojęcia: grupowania, klastrowania, typologii, delimitacji, rejonizacji i regionalizacji [Fajferek 1966; Kosiedowski, Kufel, Popławski 1989; Frankowski 1991; Balicki 2009]. Na gruncie nauk ekonomicznych zamiast pojęcia taksonomii operuje się pojęciem klasyfikacji lub analizy skupień. W konsekwencji wielu autorów dla określenia metod służących konstrukcji miar syntetycznych oraz porządkowania liniowego dla obiektów wielocechowych postuluje się pojęciem wielowymiarowej analizy porównawczej (WAP) [Grabiński 1988; Panek 2009]. Podsumowując, pojęcie taksonomii może zatem określać zbiór metod klasyfikacji lub analizy skupień. Z tego powodu w przypadku analiz przestrzennych możemy mówić o przestrzennej analizie taksonomicznej (PAT) jako metodzie klasyfikacji (grupowania) czy analizy skupień mającej zastosowanie do obiektów przestrzennych.

W badaniach przestrzennych celem prowadzonych analiz jest pogrupowanie obiektów ze względu na ich położenie lub określenie wzajemnych związków zachodzących między obiektami. Lokalizację obiektów możemy przedstawić w przestrzeni geograficznej jako dwa punkty (x, y) służące geolokalizacji z przypisanymi im wartościami cech. Grupowanie takich danych prowadzi się w odniesieniu do koordynat geograficznych (x, y) , dla których uzyskujemy klastry przestrzenne sąsiadujących obiektów. W ramach uzyskanych podziałów badamy cechy charakteryzujące wyodrębnione obiekty. W tego typu analizach wykorzystuje się algorytmy oparte na macierzy odległości (k -średnich, PAM, CLARA) lub gęstość lokalizacji (statystyka DBSCAN). Możliwe jest również grupowanie danych na podstawie hierarchicznego klastrowania z ograniczeniem przestrzennym. W dalszej części rozdziału przedstawione zostaną te algorytmy oraz propozycja grupowania przestrzennego, które uwzględniają wzajemne oddziaływanie obiektów w ich lokalizacjach.

6.1. Opis metod

Celem w przestrzennej analizie taksonomicznej (PAT) jest uzyskanie podziału badanych obiektów zlokalizowanych w przestrzeni na p homogenicznych skupień. Skupienie można uznać za homogeniczne, jeżeli obiekty badane ze względu na rozpatrywane cechy należą tylko do tego skupienia i nie wchodzi w skład innych [Seber 2004]. Zatem w skupieniach homogenicznych zlokalizowane są obiekty najbardziej do siebie podobne, natomiast w skupieniach różniących się od siebie znajdują się obiekty jak najmniej do siebie podobne. Podstawowe metody taksonomiczne możemy podzielić na następujące typy: metody hierarchiczne, optymalizacyjne, obszarowe i inne [Cormack 1971; Everitt i in. 2011]. Dla klasycznej analizy skupień Gordon [1999] wyróżnia następującą procedurę działania. Po pierwsze określenie obiektów oraz zmiennych, które je charakteryzują. Następnie wybór odpowiednich miar odległości i metod analizy skupień. Ważne jest również określenie ostatecznej liczby skupień oraz ocena jakości doboru obiektów do poszczególnych grup (klastrow). Takie postępowanie jest jak najbardziej prawidłowe dla obiektów przestrzennych.

Słabą stroną klasycznej analizy skupień jest jej mała odporność na skorelowanie zmiennych. Występowanie zmiennych silnie skorelowanych będzie powodowało zaburzenie struktury skupień. Jak podają Hair i in. [1995], zmienne silnie skorelowane będą silnie wpływały na wynik grupowania. Jednym ze sposobów na rozwiązanie tego problemu może być tzw. procedura dualna, która polega na wstępnym podziale badanych zmiennych na grupy silnie ze sobą powiązane, a następnie wybór reprezentantów, z których tworzy się listę zmiennych diagnostycznych [Nowak 1984]. Niezależnie jednak od przyjętego ostatecznie podejścia i metody punktem wyjścia w dalszej analizie jest macierz korelacji lub utworzona na jej podstawie macierz odległości [Walesiak 2002].

W analizach przestrzennych obiekty są zlokalizowane w przestrzeni, natomiast cechy je opisujące mogą mieć charakter rolniczy, społeczny lub ekonomiczny. Miary odległości wykorzystywane w analizach przestrzennych mogą być standardowymi miarami wykorzystywanymi w analizie skupień lub oddawać miarę odległości między badanymi obiektami. Miary odległości wykorzystywane w analizie skupień zostały przedstawiono w formułach (6.1)–(6.6) (odległości są tu liczone dla nieu-normowanych zmiennych). Odległość Euklidesa zapisana formułą (6.1) jest wykorzystywana do analiz wielowymiarowych jako najbardziej „naturalna” miara odległości.

$$d_{il} = \left[\sum_{j=1}^k (x_{ij} - x_{lj})^2 \right]^{1/2}. \quad (6.1)$$

Średnia z odległości euklidesowej (6.2) określana jest często jako średnia odległość taksonomiczna i jest miarą najczęściej używaną w taksonomii (naukach biologicznych).

$$d_{il} = \left[\frac{1}{k} \sum_{j=1}^k (x_{ij} - x_{lj})^2 \right]^{1/2}. \quad (6.2)$$

Dla wartości standaryzowanych, przy założeniu, że każda z k cech jest niezależna i ma rozkład $N(0, 1)$, jej wartość oczekiwana zbliża się do $\sqrt{2}$, a wariancja w przybliżeniu równa się $1/m$ [Sneath, Sokal 1973]. Odległość euklidesową określa się jako klasyfikację sferyczną, a odległość Mahalanobisa jest stosowana w klasyfikacji eliptycznej [Zeliaś (red.) i in. 1991]. Modyfikacją podstawowej odległości euklidesowej jest jej kwadrat (6.3):

$$d_{il} = \sum_{j=1}^k (x_{ij} - x_{lj})^2. \quad (6.3)$$

Odległość miejska (Hamminga) charakteryzuje klasyfikację kubiczną, czyli grupy mają postać hiperkostek – formuła (6.4):

$$d_{il} = \sum_{j=1}^k (x_{ij} - x_{lj})^2. \quad (6.4)$$

Odległość Czebyszewa:

$$d_{il} = \max_j |x_{ij} - x_{lj}|. \quad (6.5)$$

Odległość Sneatha (6.6) jest wykorzystywana w przypadku cech jakościowych [Suchecka (red.) 2014].

$$d_{il} = \frac{1}{k} \sum_{j=1}^k I(x_{ij} \neq x_{lj}). \quad (6.6)$$

Innymi miarami odległości dla cech jakościowych są miary Rogersa i Tanimoto [1960], współczynnik Hamanna [1961], odległość Russela i Rao [Rohlf 1994], odległość Kulczyńskiego [Rohlf 1994] i miara Gowera [Sneath, Sokal 1973]. Te miary są jednak rzadko stosowane.

Poza tym miernikami dla cech ilościowych stosowana jest również odległość Canberra:

$$d_{il} = \sum_{j=1}^k \left[\frac{|x_{ij} - x_{lj}|}{(x_{ij} + x_{lj})} \right], \quad (6.7)$$

gdzie $i, l = 1, 2, \dots, n$ jest liczbą obiektów, $j = 1, 2, \dots, k$ jest liczbą zmiennych, x_{ij}, x_{lj} to znormalizowane wartości j -tej cechy obiektów i oraz l . Należy tu również wymienić rzadko stosowane odległości Jeffreysa-Matusita, Braya-Curtisa, Clarka i Lance-Williamsa [Zeliaś (red.) i in. 1991]. Po obliczeniu odległości między każdą parą obiektów otrzymuje się symetryczną macierz odległości \mathbf{D} , której postać zapisano formułą (6.8):

$$\mathbf{D} = \begin{bmatrix} 0 & d_{12} & \dots & d_{1j} & \dots & d_{1k} \\ d_{21} & 0 & \dots & d_{2j} & \dots & d_{2k} \\ \vdots & \vdots & 0 & \vdots & \vdots & \vdots \\ d_{i1} & d_{i2} & \dots & d_{ij} & \dots & d_{ik} \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nj} & \dots & 0 \end{bmatrix}. \quad (6.8)$$

Macierz \mathbf{D} stanowi podstawę do tworzenia skupień. Technika stosowana do tworzenia skupień w grupowaniu obiektów przestrzennych to metoda najbliższego sąsiada. W tej technice grupowania, określanej często jako technika pojedynczego wiązania, odległość pomiędzy dwoma skupieniami jest określana przez odległość pomiędzy dwoma najbliższymi obiektami należącymi do różnych skupień. Przeciwnością tej techniki jest technika najdalszego sąsiada, określana również jako technika

pełnego wiązania. Kolejną techniką jest technika średnich wiązań, w której odległości pomiędzy dwoma skupieniami oblicza się jako średnią odległość pomiędzy wszystkimi parami obiektów należących do dwóch różnych skupień. Ta technika określana jest również jako UPGMA (ang. *unweighted pair-group method using arithmetic averages*). Kolejną techniką jest technika środków ciężkości. W tej technice odległość pomiędzy skupieniami jest wyznaczana jako różnica między środkami ciężkości. Jako środek ciężkości przyjmuje się średni punkt w przestrzeni wielowymiarowej zdefiniowanej przez te wymiary. Sneath i Sokal [1973] określają skrót dla tej techniki jako UPGMC (ang. *unweighted pair-group method using centroid average*). Kolejną techniką wykorzystywaną w analizie skupień jest technika mediany. Jest ona niemal identyczna jak technika środków ciężkości. Jedyną różnicą polega na zmianie środka ciężkości, którym jest mediana. W tej technice do obliczeń wprowadza się ważenie, aby uwzględnić różnice między liczbą obiektów występujących w danym skupieniu. Sneath i Sokal [1973] określają ją jako WPGMC (ang. *weighted pair-group method using centroid average*). Często wykorzystywaną techniką w analizie skupień jest technika Warda. W tej technice szacuje się odległości pomiędzy skupieniami, wykorzystując analizę wariancji, czyli dwie grupy będą należały do skupienia, wtedy kiedy suma kwadratów odchyłeń od wszystkich obiektów tych dwóch grup od środka ciężkości nowej grupy stanowi minimum [Suchecka (red.) 2014].

W analizie przestrzennej jednostką przestrzenną może być kraj, województwo, powiat lub gmina, ale również miasto, dzielnica czy gospodarstwo. Ważne jest umiejscowienie badanej jednostki w przestrzeni i zaimplantowanie tego położenia do analizy statystycznej. W badaniach przestrzennych należy również uwzględnić interakcje między jednostkami przestrzennymi, ponieważ wiadomo, że badane jednostki przestrzenne nie funkcjonują w izolowanej przestrzeni. Klasyczna analiza skupień nie uwzględnia powyższych założeń. Należy zatem zdefiniować interakcje i powiązania przestrzenne na bazie definicji sąsiedztwa.

Najprostszą definicję sąsiedztwa można zapisać według następującego schematu:

$$\begin{cases} w_{ij} = 1, & \text{gdy obiekt } i\text{-ty jest sąsiadem obiektu } j\text{-tego} \\ w_{ij} = 0, & \text{gdy obiekt } i\text{-ty nie jest sąsiadem obiektu } j\text{-tego,} \\ w_{ij} = 0, & \text{gdy } i = j \end{cases} \quad (6.9)$$

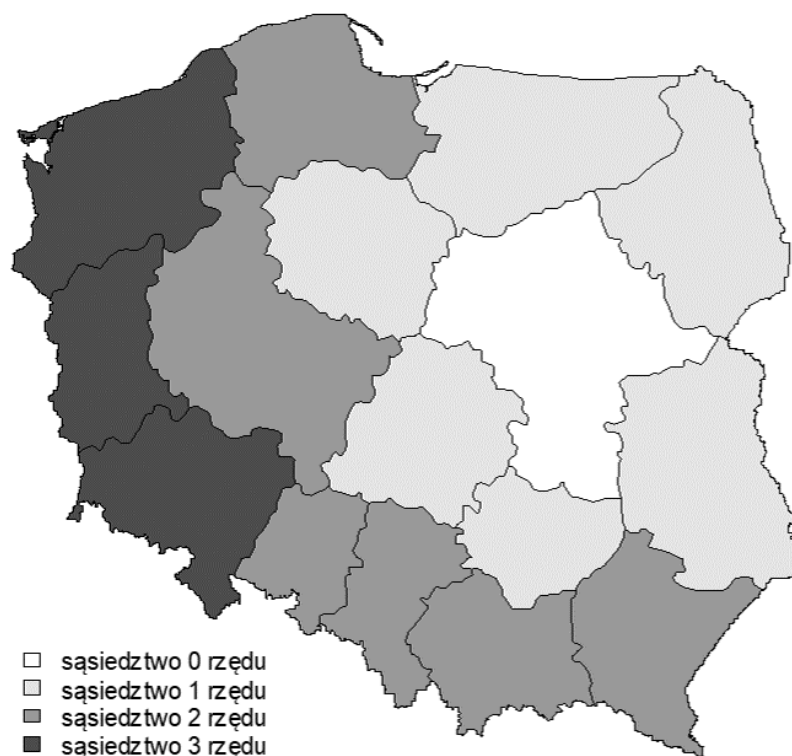
gdzie w_{ij} jest elementem macierzy sąsiedztwa.

Taki sposób określenia sąsiedztwa występuje wtedy, kiedy dwie jednostki przestrzenne mają wspólną granicę. W przypadku kiedy dwa obiekty (jednostki przestrzenne) sąsiadują ze sobą w macierzy sąsiedztwa, zaznaczamy to jako 1, a w przypadku, kiedy obiekty nie są sąsiadami, przyznajemy wartość 0. Na diagonalu macierzy sąsiedztwa wstawiamy zera, ponieważ dany obiekt nie może być swoim sąsiadem. Na rysunku 6.1 przedstawiono schemat powiązań według kryterium wspólnej granicy, a na rysunku 6.2 – sąsiedztwo pierwszego, drugiego i trzeciego rzędu dla województwa mazowieckiego.



Rysunek 6.1. Schemat powiązań według kryterium wspólnej granicy dla województw na mapie referencyjnej Polski

Źródło: obliczenia własne



Rysunek 6.2. Schemat powiązań według kryterium wspólnej granicy uwzględniający rząd sąsiedztwa dla województwa mazowieckiego na mapie referencyjnej Polski.

Źródło: obliczenia własne

Występowanie sąsiedztwa pierwszego rzędu obserwujemy dla obiektów sąsiadujących bezpośrednio ze sobą. Zatem schemat powiązań przedstawiony na rysunku 6.1 jest przykładem sąsiedztwa pierwszego rzędu. Uwzględnienie kolejnych sąsiadów powoduje wydzielenie dalszych rzędów, tzn. sąsiedztwa drugiego i trzeciego rzędu. Na rysunku 6.2 zaznaczono kolejne rzędy sąsiedztwa w gradiencie szarości, na przykład lubuskie, zachodniopomorskie i dolnośląskie są sąsiadami trzeciego rzędu dla województwa mazowieckiego [Pietrzykowski 2011a].

Innym sposobem definicji sąsiedztwa jest określenie odległości pomiędzy sąsiadami (np. w kilometrach) na zasadzie:

$$\begin{cases} w_{ij} = 1, & \text{gdy obiekt } i\text{-ty jest oddalony od obiektu } j\text{-tego o } d \text{ km lub mniej} \\ w_{ij} = 0, & \text{gdy obiekt } i\text{-ty jest oddalony od obiektu } j\text{-tego o więcej niż } d \text{ km,} \\ w_{ij} = 0, & \text{gdy } i = j \end{cases} \quad (6.10)$$

gdzie oznaczenia jak w formule (6.9).

Nawiązując do analizy skupień i określania odległości pomiędzy obiektami, interakcje przestrzenne można zdefiniować przez wprowadzenie odległości społecznych [Doreian 1980] lub ekonomicznych [Conley 1999], które dotyczą wzajemnych stosunków handlowych, przepływu kapitału i migracji [Groot, Abreu, Florax 2005]. Jeszcze innym podejściem do analizy sąsiedztwa jest propozycja Dacey [1968], uwzględniającą długość granicy pomiędzy sąsiadującymi obiektami według następującego wzoru:

$$w_{ij} = b_{ij}\alpha_i\beta_{ij}, \quad (6.11)$$

gdzie w_{ij} jest elementem macierzy sąsiedztwa, b_{ij} binarnym współczynnikiem sąsiedztwa, α_i udziałem powierzchni obiektu i -tego w badanej powierzchni całkowitej, β_{ij} miarą granicy obiektu w badanej powierzchni całkowitej.

Tak zdefiniowane wzajemne interakcje i powiązania zestawia się w macierzy \mathbf{W} , którą określa się jako macierz wag lub macierz sąsiedztwa. W formule (6.12) przedstawiono najprostszą macierz wag uzyskaną zgodnie z zasadą zapisaną formułą (6.9).

$$16 \times 16 \quad \mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (6.12)$$

Składa się ona z zer i jedynek, z tego powodu określa się ją jako binarną macierz wag. Macierz wag opisana formułą (6.12) została zbudowana dla Polski, gdzie obiektami są województwa, które sąsiadują ze sobą (czyli mamy sąsiadów pierwszego rzędu). Każda z jedynek określa sąsiada dla poszczególnych województw. W pierwszej kolumnie tej macierzy mamy województwo wielkopolskie, które ma siedmiu sąsiadów (kujawsko-pomorskie, łódzkie, dolnośląskie, lubuskie, opolskie, pomorskie, zachodniopomorskie). W tabeli 6.1 zamieszczono nazwy województw i liczby sąsiadów pierwszego rzędu.

Tabela 6.1. Liczba sąsiadów pierwszego rzędu dla poszczególnych województw Polski

Województwa	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13	w14	w15	w16
Liczba sąsiadów	3	5	6	4	3	3	6	4	3	3	4	4	6	4	7	3

Kody dla poszczególnych województw: dolnośląskie - (w1), kujawsko-pomorskie - (w2), łódzkie - (w3), lubelskie - (w4), lubuskie - (w5), małopolskie - (w6), mazowieckie - (w7), opolskie - (w8), podkarpackie - (w9), podlaskie - (w10), pomorskie - (w11), śląskie - (w12), świętokrzyskie - (w13), warmińsko-mazurskie - (w14), wielkopolskie - (w15), zachodniopomorskie - (w16).

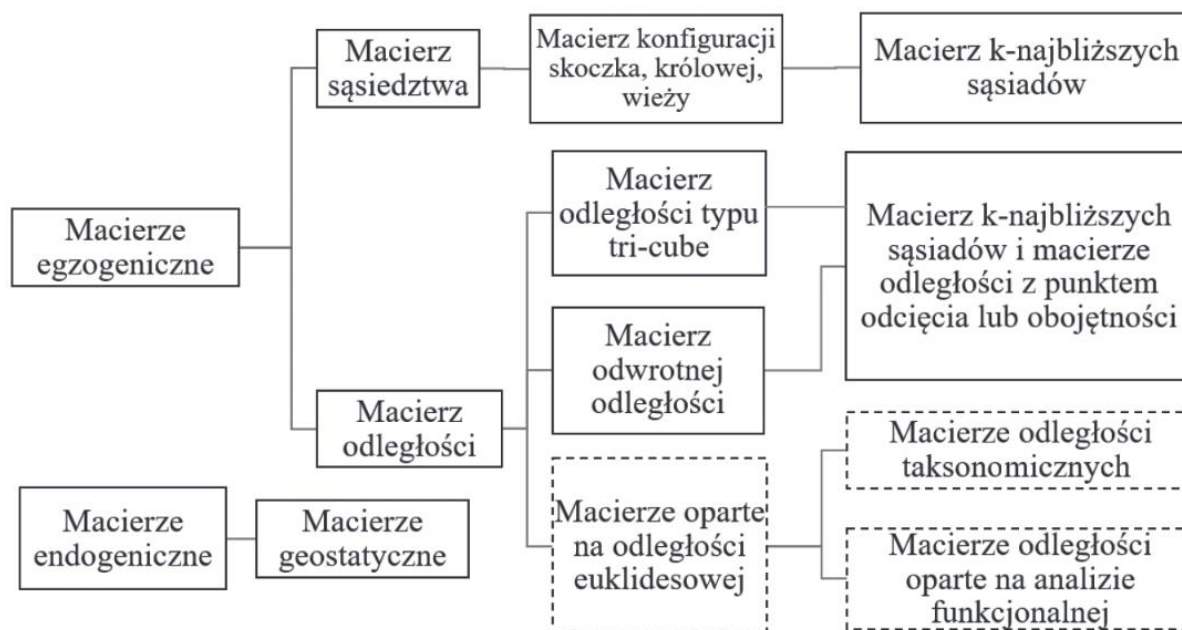
Źródło: obliczenia własne

Macierz wag \mathbf{W} jest macierzą kwadratową i symetryczną, czyli mamy tu analogię do macierzy odległości (6.8). Zatem zasadę oddziaływania symetrii $w_{ij} = w_{ji}$ można powiązać z definicją metryki, zgodnie z którą zachodzi $\forall s_i, s_j \in D: |s_i - s_j| = |s_j - s_i|$. Tym samym zależności przestrzenne podlegają operacjonalizacji i mogą być kwantyfikowane funkcjami odległości bądź sąsiedztwa w symetrycznej macierzy wag \mathbf{W} . Macierz taka będzie spełniała następujący warunek:

$$(\mathbf{I} - \rho\mathbf{W})^{-1} = (\mathbf{I} - \rho\mathbf{W}^T)^{-1}, \quad (6.13)$$

gdzie ρ jest miarą korelacji przestrzennej.

Na rysunku 6.3 przedstawiono klasyfikację rodzajów macierzy wag przestrzennych ze względu na sposób ich budowy na podstawie prac badawczych Getisa i Alstadta [2004]. Przedstawiono również propozycję rozbudowy klasyfikacji rodzajów macierzy wag, które w swojej strukturze zawierają odległości ekonomiczne. Poza tym w takiej postaci macierzy, oprócz interakcji między obiektami, zawarty jest również element czasowy. Przykład dotyczący macierzy wag opartej na odległościach taksonomicznych można znaleźć u Pietrzykowskiego [2014]. Na rysunku 6.3 przerywaną linią zaznaczono propozycje poszerzenia dotychczasowych badań dotyczących macierzy wag, przede wszystkim o analizę funkcjonalną [Pietrzykowski 2018].



Rysunek 6.3. Klasyfikacja rodzajów macierzy wag przestrzennych ze względu na rodzaj ich budowy

Źródło: opracowanie własne na podstawie [Suchecka (red.) 2014]

Estymatorem korelacji przedstawionej w formule (6.13) jest współczynnik korelacji Morana [1950] określany jako globalny współczynnik korelacji, który jest miarą autokorelacji przestrzennej. Globalny współczynnik korelacji Morana może być traktowany jako miernik grupowania badanych jednostek w przestrzeni. Konstrukcja tego miernika wykorzystuje koncepcję współczynnika korelacji Pearsona [Sang-Il 2001] oraz statystykę gamma, która łączy dwa rodzaje informacji dotyczące podobieństw obserwacji w przestrzeni. Informacje te zawarte są w macierzy wag \mathbf{W} oraz w macierzy korelacji \mathbf{A} , które dotyczą zależności badanej zmiennej w poszczególnych lokalizacjach. Ogólną postać statystyki gamma możemy zapisać:

$$\Gamma = \mathbf{WA} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} a_{ij}, \quad (6.14)$$

gdzie w_{ij} jest elementem macierzy wag \mathbf{W} , a_{ij} elementem macierzy \mathbf{A} .

Statystykę Morana można zapisać jako:

$$\Gamma = \sum_i \sum_j w_{ij} (z_i \cdot z_j), \quad (6.15)$$

gdzie z_i i z_j są standaryzowanymi wartościami zmiennej losowej X .

Biorąc pod uwagę dowód Cliffa i Orda [1981], globalna statystyka korelacji przestrzennej Morana ma rozkład asymptotycznie normalny, a jej zapis macierzowy przedstawiono w formule (6.16):

$$I_g = \frac{n}{S_0} \frac{\mathbf{z}^T \mathbf{W} \mathbf{z}}{\mathbf{z}^T \mathbf{z}}, \quad (6.16)$$

$$Z(I_g) = \frac{I_g - E(I_g)}{\text{var}(I_g)^{1/2}} \sim N(0,1), \quad (6.17)$$

gdzie \mathbf{z} jest wektorem o elementach $z_i = x_i - \bar{x}$, S_0 sumą wszystkich elementów macierzy wag.

Jak wcześniej wspomniano, globalna autokorelacja przestrzenna bada wzajemne powiązania między obiektami, a zatem można ją sprowadzić do weryfikacji następującej hipotezy:

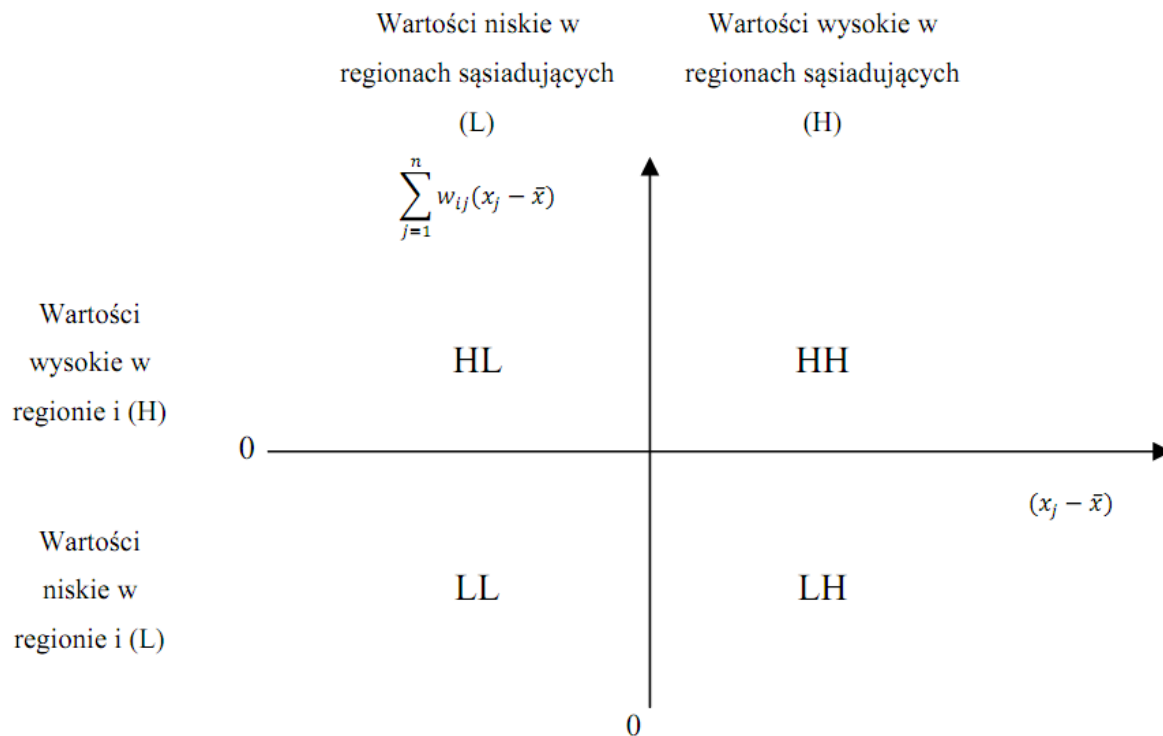
$$H_0: \rho = 0, \quad H_1: \rho \neq 0, \quad (6.18)$$

gdzie ρ oznacza współczynnik korelacji przestrzennej.

Nieodrzućenie hipotezy zerowej oznacza brak autokorelacji przestrzennej, a zatem obserwowane wartości cechy są rozmieszczone w sposób losowy pomiędzy poszczególnymi lokalizacjami. Odrzućenie hipotezy zerowej oznacza istnienie autokorelacji przestrzennej i implikuje wniosek, że wartości obserwowanej zmiennej nie są rozmieszczone w sposób losowy pomiędzy poszczególnymi lokalizacjami.

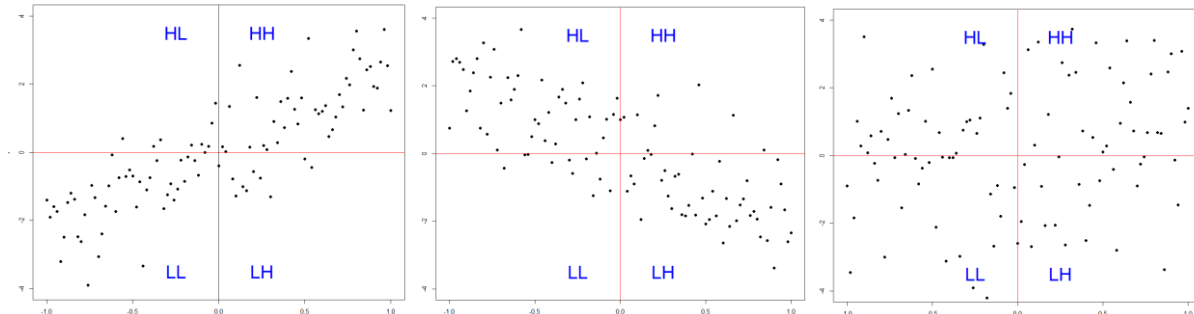
Graficzną prezentacją autokorelacji przestrzennej jest wykres rozrzutu Morana, który dotyczy wartości współczynnika globalnego Morana (I_g) (rys. 6.4). Wykres ten jest wykorzystywany do wizualizacji związków przestrzennych i określenia kierunku autokorelacji przestrzennej. Wykres podzielony jest na cztery części względem wartości zerowych.

Rozłożenie punktów na wykresie Morana świadczy o obserwowanej autokorelacji i tak: punkty w kwadracie HL i LH wskazują na ujemną autokorelację, a w kwadratach LL i HH na dodatnią autokorelację. Równomierne rozłożenie punktów w czterech kwadratach świadczy o braku autokorelacji przestrzennej (rys. 6.5). Na wykresie Morana umieszcza się również prostą regresji, której współczynnik kierunkowy jest tożsamy z współczynnikiem globalnej autokorelacji Morana, czyli w przypadku ujemnej autokorelacji przechodzi on przez punkt (0, 0) i kwadraty HL i LH. Rozmieszczenie punktów w kwadracie LL i HH wskazuje, że obiekty w tym regionie przyjmują małe lub duże wartości ze względu na badaną cechę. Należy zauważyć dużą złożoność metody klasyfikacji na bazie współczynnika autokorelacji Morana i macierzy sąsiedztwa, która może zawierać dodatkowe informacje o badanych zmiennych.



Rysunek 6.4. Diagram Morana określający zależności między regionem a sąsiadującymi obiektami

Źródło: opracowanie własne



Rysunek 6.5. Przykłady autokorelacji przestrzennej na diagramie Morana

Źródło: opracowanie własne

Obiekty są klasyfikowane na podstawie ich przydziału do czterech ćwiartek wykresu Morana. Pierwsze skupienie znajduje się w kwadracie HL, drugie w HH, trzecie i czwarte odpowiednio w kwadratach LL i LH. Pewnym mankamentem tej metody jest zawsze podział na cztery grupy skupień. Dodatkowo należy zauważyć problem dotyczący macierzy wag. Konstrukcja macierzy w bardzo dużym stopniu będzie wpływała na podział badanych obiektów. Jak można zauważyć, określenie macierzy wag w analizach przestrzennych jest bardzo ważne. Dyskusję na temat określania macierzy wag można znaleźć u Cliffa i Orda [1981], Anselina [1988] oraz Uptona i Fingletona [1985]. Jak już wcześniej wspomniano, macierz wag określa się *a priori*, a jakość dalszych analiz zależy od jej

specyfikacji, dobór odpowiedniej macierzy wag stanowi więc poważny problem metodologiczny. Najczęściej stosowana macierz wag to macierz oparta na kryterium wspólnej granicy, pierwszego rzędu, standaryzowana wierszowo (tzn. dla każdego wiersza suma jego elementów musi wynosić 1).

Pietrzykowski [2011a] zaproponował modyfikację macierzy wag \mathbf{W} , która uwzględniałaby interakcje związane z położeniem badanej jednostki przestrzennej, ale również miała wagi związane z wybranymi cechami ekonomicznymi:

$$\mathbf{W}_p = \mathbf{W} + \mathbf{D}, \quad (6.19)$$

gdzie \mathbf{W}_p jest zmodyfikowaną macierz wag, \mathbf{W} macierzą wag pierwszego rzędu, standaryzowana wierszowo, \mathbf{D} macierzą odległości.

Macierz wag \mathbf{W} nie jest macierzą symetryczną. Dlatego macierz \mathbf{W}_p poddaje się transformacji w celu zapewnienia prawidłowości wnioskowania na podstawie aproksymowanych momentów z próby [Anselin 2001] zgodnie z formułą:

$$\mathbf{W}_m = \frac{\mathbf{W}_p + \mathbf{W}_p^T}{2}, \quad (6.20)$$

gdzie \mathbf{W}_m jest macierzą wag uzyskaną według formuły (6.19).

Oprócz globalnego współczynnika autokorelacji przestrzennej, którym możemy posłużyć się do klasyfikacji obiektów w przestrzeni, mamy również lokalne współczynniki autokorelacji. Lokalne współczynniki korelacji pozwalają na bardziej szczegółowy wgląd w strukturę rozmieszczenia przestrzennego badanej zmiennej na danym obszarze, rozpoznanie wzorców heterogeniczności, identyfikację obszarów niestacjonarności, a także obserwacji nietypowych (*hot spots* i *cold spots*) oraz jednorodnych podobszarów (reżimów przestrzennych). Lokalne współczynniki korelacji są dekompozycją globalnego współczynnika korelacji na części, które dotyczą poszczególnych lokalizacji [Suchecki 2010]. Lokalny współczynnik Morana jest zaliczany do grupy współczynników określonych przez Anselina [1995] jako LISA (ang. *local indicators of spatial association*) o postaci:

$$I_i = z_i \sum_j w_{ij} z_j. \quad (6.21)$$

Współczynniki lokalne (6.21) spełniają kryterium proporcjonalności do odpowiedniego współczynnika korelacji globalnej, to znaczy suma wszystkich współczynników lokalnych w poszczególnych lokalizacjach przyjmuje wartość współczynnika globalnego autokorelacji [Suchecki 2010], co można przedstawić:

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j \quad (6.22)$$

i wtedy statystykę Morana zapisujemy jako:

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j \quad \text{lub} \quad I = \frac{\sum_i I_i}{S_0 \frac{\sum_i z_i^2}{n}}, \quad (6.23)$$

gdzie $S_0 = \sum_i \sum_j w_{ij}$ a pozostałe symbole jak we wzorze (6.16).

6.2. Zastosowania metod przestrzennej analizy taksonomicznej

Zastosowania metody klasyfikacji przestrzennej lub, jak określono w tytule rozdziału, przestrzennej analizy taksonomicznej (PAT) można podzielić na wykorzystanie klasycznych metod analizy skupień oraz klasyfikacji na bazie lokalnego i globalnego współczynnika Morana. W tabeli 6.2 zestawiono przykłady wykorzystania metod PAT w ekonomice rolnictwa oraz zagadnieniach związanych z ekonomią gospodarczą. W tabeli 6.2 zestawiono prace z ostatnich dziesięciu lat, chcąc pokazać, że zagadnienie przestrzeni w klasyfikacji obiektów jest cały czas aktualne. Zamieszczone przykłady w większości przypadków odnoszą się do Polski, ale są również prace dotyczące Unii Europejskiej (UE), Słowacji i Chin. W prowadzonych badaniach zakres analiz przestrzennych dotyczy: państw [Tłuczak 2013; Zalewski 2020], województw [Pietrzykowski i Wicki 2011; Pietrzykowski 2011b; Tłuczak 2014; Łukiewska, Chrobocińska 2015; Kobylińska 2016; Pietrzykowski 2018; Wicki, Pietrzykowski 2018], powiatów [Malinowski 2017], gmin [Foryś i Putek-Szeląg 2013; Gospodarowicz 2018], prowincji [Han i in. 2020], regionów [Xiao-Ni i in. 2011; Głębocki, Kacprzak i Kossowski 2019], podregionów [Pietrzak 2014] miast [Herbst i Wójcik 2013], gospodarstw domowych [Maciejewski 2020] i poletek [Koutsos, Menexes, Mamolos 2021]. Tak szerokie spektrum analiz wykazuje potrzebę metod klasyfikacyjnych zlokalizowanych w przestrzeni gospodarczej. Z klasycznych metod analizy skupień najczęściej w badaniach wykorzystuje się technikę Warda i odległość euklidesową [Foryś, Putek-Szeląg 2013; Łukiewska, Chrobocińska 2015; Kobylińska 2016; Zalewski 2020] lub jej kwadrat [Maciejewski 2020], a także najbardziej ogólną jej postać, tzn. odległość Minkowskiego [Pietrzykowski, Wicki 2011]. W pracy Pietrzykowskiego [2011b] przedstawiono zastosowanie globalnego współczynnika korelacji Morana oraz diagramu Morana do klasyfikacji obiektów z modyfikacją macierzy wag. Rozwinięcie tej metody zostało opisane również dla macierzy wag zbudowanej na bazie analizy funkcjonalnej [Pietrzykowski 2018]. W pozostałych pracach wykorzystywano do określenia reżimów przestrzennych lokalnego (LISA) lub globalnego współczynnika Morana. W tabeli 6.2 wymieniono problematykę, jaką zajmowali się autorzy, badane obiekty oraz zmienne wykorzystane w analizach przestrzennych (PAT). Podsumowując, w większości prac analiza dotyczy zagadnień regionalnych i jest często stosowana w badaniach związanych z ekonomią rolnictwa.

Tabela 6.2. Przykłady użycia metod przestrzennych oraz klasycznej analizy skupień, jak i jej modyfikacji w ekonomice rolnictwa i naukach społeczno-ekonomicznych

Autor	Problem	Obiekty	Metoda
Xiao-Ni i in. 2011	W pracy badano przestrzenne rozmieszczenie metali ciężkich w gruntach rolniczych w Chinach (dookoła Pekinu). Przestrzenna zależność 4 metali była zależna od skali odległość progowej 13 km, 32 km, 50 km i 29 km odpowiednio dla Cr, Ni, Zn i Hg. Analiza przestrzenna pozwoliła na wykrycie klastrow przestrzennych i wartości odstających (<i>hot-spots</i>).	regiony	PAT na bazie globalnego i lokalnego (LISA) współczynnika Morana, diagram Morana, wykorzystano 8 różnych macierzy wag: macierz pierwszego rzędu (formacja wieży i królowej), od 5 do 8 najbliższych sąsiadów, dystans 4 km.
Pietrzykowski i Wicki 2011	Badania dotyczyły regionalnego zróżnicowania natężenia wykorzystania środków w ramach działań wspólnej polityki rolnej WPR. Badanymi cechami były: liczba umów o dofinansowanie inwestycji na 1000 gospodarstw rolnych, które złożyły w danym roku wnioski o płatności bezpośrednie, wartość wypłaconych środków w przeliczeniu na 1 gospodarstwo składające wnioski o płatności bezpośrednie, wartość wypłaconych dotacji w przeliczeniu na 1 ha użytków rolnych.	województwa	PAT klasyczny na bazie techniki Warda, odległość Minkowskiego z $p = 4$, klasyfikacje przedstawiono na mapie Polski.
Pietrzykowski 2011b	Badanie ekonomiczne struktury regionalnej Polski. Macierz wag zawierała informacje o interakcjach między województwami ze względu na wybrane cechy ekonomiczne, jakimi były: jednolite dopłaty bezpośrednie, uzupełniające dopłaty bezpośrednie, dopłaty do ONW, nawożenie mineralne (w kg/ha), powierzchnia użytków rolnych według klas bonitacyjnych w hektarach dla klas V i VI. Analizę prowadzono ze względu na ceny gruntów rolniczych.	województwa	PAT na podstawie statystyki Morana i macierzy wag uzyskanej na bazie formuły (6.20), macierz odległości euklidesowych.
Foryś i Putek-Szeląg 2013	Przedmiotem badań była przestrzenna klasyfikacja gmin ze względu na rodzaj zbywanych użytków gruntowych w województwie zachodniopomorskim przez ANR w latach 1995-2011. Wykorzystano następujące cechy: powierzchnia każdego użytku zbywanej nieruchomości (w ha), powierzchnia poszczególnych klas gruntów ornych (ha), lokalizacja w gminie województwa zachodniopomorskiego, data zawarcia umowy, cena transakcyjna użytku gruntowego (w zł), cena transakcyjna 1 ha użytku gruntowego (w zł/ha).	gminy województwa zachodniopomorskiego	PAT, klasyczna analiza skupień techniką Warda, wykorzystano miarę zróżnicowania struktur [Kukuła 2010], macierz odległości euklidesowych.
Tłuczak 2013	Prowadzone badania miały za zdanie określić zależności i interakcje przestrzenne na poziomie państw UE. Cechami badanymi były: wskaźniki produkcji roślinnej w cenach producenta z lat 2010 i 2011 oraz wskaźniki produkcji zwierzęcej w cenach producenta z lat 2010 i 2011.	27 państw UE	PAT na podstawie statystyki Morana, wykorzystano diagram Morana do klasyfikacji obiektów, klasyfikację przedstawiono na mapie Europy.
Herbst i Wójcik 2013	W pracy przedstawiono problem zwiększania spójności gospodarczej w ujęciu przestrzennym, a także dyfuzji wielkich miast o cechach metropolitalnych. Dla celów analizy wykorzystamy miarę poziomu rozwoju odnosząc się do dochodów budżetów gmin: wielkość dochodów własnych w przeliczeniu na 1 mieszkańca. Rozważano wszystkie miasta wojewódzkie jako potencjalnych kandydatów o charakterze metropolitalnym.	województwa, gminy, 18 miast wojewódzkich	PAT na podstawie globalnej i lokalnej (LISA) statystyki Morana, macierz odległości dotyczyła gminy i danego miasta w województwie.

Tabela 6.2 (cd.)

Autor	Problem	Obiekty	Metoda
Tłuczak 2014	Analiza dotyczyła badań regionalnych Polski. W pracy badano zróżnicowanie i zależności przestrzenne: cen bydła i trzody w skupie (w zł/1 kg) w latach 2005-2012.	województwa	PAT na podstawie statystyki lokalnej Morana, statystyka LISA, klasyfikacje przedstawiono na mapie Polski.
Tłuczak 2014	Analiza dotyczyła badań regionalnych Polski. W pracy badano zróżnicowanie i zależności przestrzenne: cen bydła i trzody w skupie (w zł/1 kg) w latach 2005-2012.	województwa	PAT na podstawie statystyki lokalnej Morana, statystyka LISA, klasyfikacje przedstawiono na mapie Polski.
Pietrzak 2014	Analiza dotyczyła badań regionalnych i pozwoliła na ocenę sytuacji gospodarczej w Polsce oraz tendencji w rozwoju gospodarczym podregionów. Cechami badanymi były: PKB <i>per capita</i> , podmioty gospodarki narodowej wpisane do rejestru REGON na 10 000 ludności, nakłady inwestycyjne w przedsiębiorstwach <i>per capita</i> , przeciętne miesięczne wynagrodzenie brutto oraz stopa bezrobocia rejestrowanego.	66 wybranych podregionów (NUTS3)	pTMR modyfikacja taksonomicznego miernika rozwoju (TMR) uwzględniającego potencjalną siłę interakcji między regionami. Oprócz statystyki Morana wykorzystano model SAR.
Łukiewska i Chrobocińska 2015	W pracy wykonano przestrzenną analizę zróżnicowania potencjału produkcyjnego rolnictwa. Cechami badanymi były zasoby produkcyjne, czyli ziemia, kapitał i praca ujęta jako: powierzchnia użytków rolnych, wartość brutto środków trwałych oraz liczba pracujących w rolnictwie w poszczególnych województwach Polski w latach 2011-2013. Dokonano podziału na 4 klasy typologiczne.	województwa	PAT klasyczny na bazie techniki Warda, odległość euklidesowa.
Kobylińska 2016	Celem pracy było określenie przestrzennego zróżnicowania cen rynku nieruchomości w 2014 roku. Badanymi cechami były: liczba transakcji kupna/sprzedaży lokali mieszkalnych na 10 000 mieszkańców, średnia cena lokali mieszkalnych (w zł/m ²), liczba transakcji kupna/sprzedaży nieruchomości zabudowanych budynkami mieszkalnymi na 10 000 mieszkańców, średnia cena nieruchomości zabudowanych budynkami mieszkalnymi (w zł/m ²).	województwa	PAT klasyczny na bazie techniki Warda, odległość euklidesowa.
Rzeszutko 2016	Celem pracy była ocena zmian w strukturze ekonomicznej gospodarstw rolnych w Polsce w latach 2005-2013 w warunkach WPR. Analizie poddano zarówno liczbę gospodarstw rolnych według ich wielkości ekonomicznej, jak i wartość wytwarzanej SO.	województwa	PAT klasyczna analiza skupień techniką Warda, wykorzystano miary zróżnicowania struktury zaproponowane przez Kukułę [2010].
Malinowski 2017	W pracy badano przestrzenny charakter poziomu życia mieszkańców w 101 powiatach w województwach lubelskim, podkarpackim, podlaskim, świętokrzyskim i warmińsko-mazurskim. W pracy wykorzystano 23 zmienne diagnostyczne w 7 grupach tematycznych: rynek pracy, ochrona zdrowia, środowisko, transport, warunki mieszkaniowe, kultura i edukacja.	powiaty	PAT na bazie metody podziałowej PAM (algorytm grupowania wokół medoidów), analiza skupień techniką Warda, współczynnik globalny Morana i diagram Morana.

Gospodarowicz 2018	Celem prowadzonych analiz była delimitacja zjawiska spójności terytorialnej pod względem rozwoju infrastruktury technicznej oraz identyfikacja klastrów na poziomie gmin w Polsce. Badaniem objęto lata 2005, 2010 i 2015. W pracy na bazie metody Hellwiga określono syntetyczny wskaźnik agregujący miary dotyczące gęstości sieci wodociągowej, kanalizacyjnej i gazowej (na 100 km ²) oraz odsetka ludności korzystającej z poszczególnych mediów. Syntetyczny wskaźnik wykorzystano w klasyfikacji do oceny poziomu rozwoju demograficznego oraz społeczno-ekonomicznego.	gminy	PAT na bazie globalnej i lokalnej (LISA) statystyki Morana, diagram Morana.
Pietrzykowski 2018	Prowadzone badania miały za zadanie klasyfikację badanych obiektów, uwzględniając interakcje czasowo-przestrzenne. Zmienną badaną był dochód z gospodarstwa rolnego w latach 2004-2012. W pracy porównano klasyfikację, wykorzystując klasyczną analizę skupień (technika Warda) z klasyfikacją uzyskaną z wykorzystaniem analiz czasowo-przestrzennych.	województwa	PAT na bazie statystyki Morana i diagramu Morana, macierz wag uzyskano w wyniku analizy funkcjonalnej i macierzy binarnej.
Wicki i Pietrzykowski 2018	Celem pracy było określenie zmian przestrzennego zróżnicowania wykorzystania środków na modernizację gospodarstw rolnych w Polsce w ramach PROW na lata 2007-2013 i 2014-2020. W pracy wykorzystano następujące zmienne: liczba gospodarstw rolnych, liczba gospodarstw według wielkości ekonomicznej określonej wartością produkcji standardowej i określonej według wielkości standardowej nadwyżki bezpośredniej, liczba zatrudnionych w rolnictwie, powierzchnia użytków rolnych, liczba gospodarstw korzystająca z programu „Modernizacja gospodarstw rolnych”, uzyskana kwota wsparcia w ramach tego programu.	województwa	PAT na bazie statystyki Morana i diagramu Morana, macierz wag uzyskano w wyniku odległości euklidesowych i macierzy binarnej.
Głębocki i in. 2019	W pracy podjęto próbę syntetycznego opisu przestrzennej niejednorodności rozwoju polskiego rolnictwa. Zbiór danych składał się z 69 zmiennych oraz 3069 obiektów (LAU2, co korespondowało z NUTS5). Badane zmienne podzielono na pięć grup: system własności ziemi i organizacja przestrzeni produkcyjnej, zasoby i jakość pracy, środki produkcji, zasoby naturalne oraz struktura upraw i chów zwierząt, skutki produkcji rolnej.	gminy	PAT na bazie statystyki lokalnej (LISA) i globalnej Morana, technika k-średnich.
Han i in. 2020	W pracy analizowano przestrzenno-czasową ewolucję przemysłu hodowli zwierząt roślinożernych w gospodarce Chin. Poza tym starano się określić czynniki wpływające na tak prowadzoną hodowlę. Jako cechy diagnostyczne wybrano produkcję wołowiny, jagnięciny i mleka. Poza tym, aby uwzględnić wpływ miasta, przyjęto również PKB <i>per capita</i> , wskaźnik urbanizacji i współczynnik Engla. Badano również produktywność w rolnictwie, poziom mechanizacji i stopień nawadniania.	provincje	PAT z wykorzystaniem statystyki lokalnej (LISA) i globalnej Morana, diagramu Morana, przestrzennego modelu Durbina, elipsę odchylenia standardowego, macierz wag w formacji królowej.

Tabela 6.2 (cd.)

Autor	Problem	Obiekty	Metoda
Maciejewski 2020	Celem pracy było zaprezentowanie wyników typologii gospodarstw domowych przeprowadzonych w Polsce i na Słowacji. Próba liczyła 900 gospodarstw domowych. W analizach wykorzystano 11 zmiennych diagnostycznych: zakupu żywności dokonujemy na podstawie wcześniej sporządzonej listy produktów. Przy zakupie żywności zwracamy uwagę na jej składniki. Przy zakupie żywności zwracamy uwagę na termin przydatności do spożycia. Przy zakupie żywności zwracamy uwagę na jej walory smakowo-zapachowe. Przy zakupie żywności zwracamy uwagę na jej cenę. Żywność kupujemy w sklepach, do których mamy zaufanie. Kupujemy żywność przez Internet. Przy zakupie żywności zwracamy uwagę na rodzaj opakowania. Przy zakupie żywności zwracamy uwagę na jej pochodzenie. Informujemy się o nowych ofertach żywnościowych. Przed zakupem zastanawiamy się, czego potrzebujemy. Odpowiedzi były rejestrowane w 7-stopniowej skali. Analizy prowadzono dla województwa mazowieckiego, województwa śląskiego i Słowacji.	gospodarstwa domowe	PAT klasyczna analiza skupień techniką Warda i k-średnich, macierz kwadratowych odległości euklidesowych.
Zalewski 2020	W pracy podjęto próbę przedstawienia i porównania zmiany wartości zużytych nawozów i środków ochrony roślin w krajach UE w latach 2010-2018. Jako cechy diagnostyczne przyjęto: wartości zużytych nawozów i środków ochrony roślin w przeliczeniu na 1 ha UR oraz udziału środków produkcji w zużyciu pośrednim.	państwa UE	PAT klasyczna analiza skupień techniką Warda, macierz odległości euklidesowych.
Koutsos, Menexes, Mamolos 2021	Celem pracy było przedstawienie korzyści wynikających z przestrzennej analizy upraw na bazie zrównoważonego rozwoju. W tym przypadku przestrzenna analiza danych dotyczących plonów upraw w postaci analizy autokorelacji przestrzennej została wykorzystana jako praktyczne, zrównoważone podejście do lokalizacji statystycznie istotnych obszarów o niskiej produkcji. W analizie wykorzystano dane z farmy w Argentynie. Jako zmienne wykorzystano plon kukurydzy, nawożenie azotowe w latach 1991 i 2001. Dane zebrano z 6 odrębnych doświadczeń (różny poziom nawożenia) na 3 powtórzonych blokach zawierających 18 poletek (pasków) w poprzek pola.	poletka/paski	PAT na bazie lokalnego (LISA) i globalnego współczynnika Morana, zmodyfikowany diagram Morana, macierz wag określono jako odwrotność odległości euklidesowej.

Źródło: opracowanie własne

Literatura

- Anselin L. (1988) Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20, 1-17.
- Anselin L. (1995) Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27(2), 93-115.
- Anselin L. (2001) *Spatial econometrics*. Oxford, Basil Blackwell.
- Balicki A. (2009) *Statystyczna analiza wielowymiarowa i jej zastosowania społecznoekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
- Brzeziński J. (1997) *Metodologia badań psychologicznych*. PWN, Warszawa
- Chojnicki Z., Czyż T. (1973) *Metody taksonomii numerycznej w regionalizacji geograficznej*. PWN, Warszawa.
- Chojnicki Z., Czyż T., Parysek J., Ratajczak W. (1978) *Badanie przestrzennej struktury społeczno-ekonomicznej Polski metodami czynnikowymi*, PWN, Warszawa-Poznań.
- Choynowski M. (1971) *Pomiar w psychologii*, [w:] J. Koziński (red.): *Problemy psychologii matematycznej*, Państwowe Wydawnictwo Naukowe, Warszawa, 15-41.
- Cliff A.D., Ord K. (1981) *Spatial Process: Models and Applications*, Pion, London.
- Conley T.G. (1999) GMM estimation with cross selection dependence. *Journal of Econometrics*, 92(1), 1-45
- Cormack R.M. (1971) A Review of Classification. *Journal of Royal Statistical Society, seria A*, 134(3), 321-367.
- Czekanowski J. (1913) *Zarys metod statystycznych w zastosowaniu do antropologii*. Towarzystwo Naukowe Warszawskie, Warszawa.
- Czekanowski J. (1930) *Zarys antropologii Polski*. Jakubowski i Sp., Lwów.
- Dacey M. (1968) A review of measures of contiguity for two and k-color maps, [w:] B. Berry, D. Marble (red.) *Spatial analysis: A Reader in Statistical Geography*. Prentice Hall, Englewood Cliffs, 479-495.
- Deptuła A.M. (2017) Analysis of criteria used in the risk assessment of technical innovations. *Procedia Engineering*, 182, 135-142.
- Doreian P. (1980) Linear models with spatial distributed data. Spatial disturbances or spatial effects. *Sociological Methods and Research*, 9, 29-60.
- Everitt B.S., Landau S., Leese M., Stahl D. (2011) *Cluster Analysis (5th Edition)*. Series in Probability and Statistics. John Wiley & Sons, New York.
- Fajferek A. (1966) *Region ekonomiczny i metody analizy regionalnej*. PWE, Warszawa.
- Fierich J. (1957) *Próba zastosowania metod taksonomicznych do rejonizacji systemów rolniczych w województwie krakowskim*. *Myśl Gospodarcza*, 1, 73-100.
- Florek K., Łukasiewicz J., Perkal J., Steinhaus H., Zubrzycki S. (1951) *Taksonomia wrocławska*. *Przegląd Antropologiczny*, 17, 193-210.
- Foryś I., Putek-Szeląg E. (2013) *Przestrzenna klasyfikacja gmin ze względu na sprzedaż użytków gruntowych zbywanych przez ANR w województwie zachodniopomorskim*. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 21, 279, 67-76.
- Frankowski Z. (1991) *Zastosowanie metod taksonomicznych w badaniach przestrzennych*. Agencja Wydawnicza Instytutu Gospodarki Przestrzennej i Komunalnej, Warszawa.
- Getis A., Alstadt J. (2004) Constructing the Spatial Weights Matrices Using a Local Statistics. *Geographical Analysis*, 36(2), 147-163.
- Głębocki B., Kacprzak E., Kossowski T. (2019) Multicriterion Typology of Agriculture: A Spatial Dependence Approach. *Quaestiones Geographicae*, 38(2), 29-49.
- Gordon A.D. (1999) *Classification Methods for the Exploratory Analysis of Multivariate Data (2nd edition)*. Chapman & Hall, London.
- Gospodarowicz M. (2018) Spójność terytorialna gmin w Polsce pod względem rozwoju infrastruktury technicznej w latach 2005-2015 w ujęciu miar autokorelacji przestrzennej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 20(5), 62-69.
- Grabiński T. (1988) *Metody statystycznej analizy porównawczej*, [w:] A. Zeliaś (red.) *Metody statystyki międzynarodowej*. PWE, Warszawa, 235-259
- Groot H., Abreu M., Florax R.J.G.M. (2005) *Space and Growth: A Survey of Empirical Evidence and Methods*. *Région et Développement*, 21, 13-44.

- Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1995) *Multivariate Data Analysis with Readings*. Prentice Hall International, London.
- Han C., Wang G., Zhang Y., Song L., Zhu L. (2020) Analysis of the temporal and spatial evolution characteristics and influencing factors of China's herbivorous animal husbandry industry. *PLoS ONE*, 15(8), 023782. DOI: 10.1371/journal.pone.0237827
- Hamann U. (1961) Merkmalbestand und Verwandtschaftsbeziehungen der Farinosae, Ein Beitrag zum System der Monokotyledonen. *Willdenowia*, 2, 639-768.
- Herbst M., Wójcik P. (2013) Delimitacja dyfuzji rozwoju z miast metropolitalnych z wykorzystaniem korelacji przestrzennej. *Studia Regionalne i Lokalne*, 4(54), 5-22.
- Kobylińska M. (2016) Metoda aglomeracyjna w ocenie przestrzennego zróżnicowania obrotu lokalami mieszkalnymi oraz nieruchomościami zabudowanymi budynkami mieszkalnymi. *Metody Ilościowe w Badaniach Ekonomicznych*, 17(3), 73-83.
- Kosiedowski W., Kufel T., Popławski W. (1989) Regionalizacja województwa wrocławskiego za pomocą metod taksonomicznych. *Wiadomości Statystyczne*, 3, 33-40.
- Koutsos T.M., Menexes G.C., Mamolos A.P. (2021) The Use of Crop Yield Autocorrelation Data as a Sustainable Approach to Adjust Agronomic Inputs. 2021. *Sustainability*, 13(4), 2362. DOI: 10.3390/su13042362
- Kukuła K. (2010) *Statystyczne studium struktury agrarnej w Polsce*. Wydawnictwo Naukowe PWN, Warszawa.
- Łukiewska K., Chrobocińska K. (2015) Przestrzenne zróżnicowanie potencjału produkcyjnego rolnictwa w Polsce. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 102 (3), 56-65.
- Maciejewski G. (2020) Typy gospodarstw domowych ze względu na ich zachowania na rynku żywności. *Problemy Rolnictwa Światowego*, 20 (35)(1), 56-66.
- Malinowski M. (2017) Przestrzenne zróżnicowanie poziomu życia ludności w ujęciu powiatów. *Wiadomości Statystyczne*, 2(669), 52-71.
- Moran P.A.P. (1950) Notes on Continuous Stochastic Phenomena. *Biometrika*, 37, 17-33.
- Nowak E. (1984) *Problemy doboru zmiennych do modelu ekonometrycznego*. PWN, Warszawa.
- Panek T. (2009) *Statystyczne metody wielowymiarowej analizy porównawczej*. Oficyna Wydawnicza SGH, Warszawa.
- Pietrzak M. (2014) Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych. *Przebieg Statystyczny*, 61(2), 181-201.
- Pietrzykowski R. (2011a) Wykorzystanie metod statystycznej analizy przestrzennej w badaniach ekonomicznych. *Roczniki Ekonomiczne Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy*, 4, 97-112.
- Pietrzykowski R. (2011b) Koncepcja i zastosowanie modyfikacji macierzy wag w przestrzennych badaniach ekonomicznych. *Metody Ilościowe w Badaniach Ekonomicznych*, 12(2), 270-278.
- Pietrzykowski R. (2014) Application of spatial techniques for panel data analysis of agricultural real estate market in the years 2004-2012. *Quantitative Methods in Economics*, 15(2), 188-197.
- Pietrzykowski R. (2018) Klasyfikacja obiektów w ujęciu czasowo-przestrzennym z zastosowaniem analizy funkcjonalnej. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia* 31, 508, 170-179.
- Pietrzykowski R., Wicki L. (2011) Regionalne zróżnicowanie wykorzystania środków z programów wspólnej polityki rolnej na modernizację rolnictwa. *Roczniki Nauk Rolniczych, Seria G*, 98(4), 7-22.
- Pociecha J., Podolec B., Sokołowski A., Zajac K. (1988) *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. PWN, Warszawa.
- Rogers D.J., Tanimoto T.T. (1960) A computer program for classifying plants. *Science*, 132, 1115-1118.
- Rohlf F.J. (1994) *NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System. Version 1.80*. Exeter Software, Seatuket, New York.
- Rzeszutko A. (2016) Zmiany struktury ekonomicznej gospodarstw rolnych w Polsce w ujęciu regionalnej analizy strukturalnej. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(4), 34-47.
- Sang-II L. (2001) Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I . *Journal of Geographical System*, 3, 369-385.
- Seber G.A.F. (2004) *Multivariate Observation*, Wiley Series in Probability and Statistics. John Wiley & Sons, New York.
- Sneath P.H.A., Sokal R.R. (1973) *Numerical taxonomy. The principles and practice of numerical classification*. W.H. Freeman, San Francisco.

- Sokal R.R., Sneath P.H.A. (1963) Principles of numerical taxonomy. W.H. Freeman, San Francisco.
- Sucecki B. (2010) Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych. Wydawnictwo C.H. Beck, Warszawa.
- Sucecka J. (red.) (2014) Statystyka przestrzenna. Metody analizy struktur przestrzennych. Wydawnictwo C.H. Beck, Warszawa.
- Tłuczak A. (2013) The analysis of the phenomenon of spatial autocorrelation of indices of agricultural output. *Metody Ilościowe w Badaniach Ekonomicznych*, 14(2), 261-271.
- Tłuczak A. (2014) Regionalne zróżnicowanie cen żywności w skupie w Polsce w latach 2005-2012. *JARD*, 1(31), 151-158.
- Upton G., Fingleton B. (1985) Spatial Data Analysis by Example. Wiley, New York.
- Walesiak M. (2002) Uogólniona miara odległości w statystycznej analizie wielowymiarowej. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Walesiak M. (2004) Metody klasyfikacji, [w:] E. Gatnar, M. Walesiak, A. Bąk (red.) Metody statystycznej analizy wielowymiarowej w badaniach marketingowych. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław, 316-350.
- Wicki L., Pietrzykowski R. (2018) Zróżnicowanie przestrzenne wykorzystania środków na modernizację gospodarstw rolnych z Programu Rozwoju Obszarów Wiejskich. *Zeszyty Naukowe Szkoły Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 124, 93-108.
- Xiao-Ni H., Wei-Wei Z., Dan-Feng S., Hong L., Lian-Di Z., Bao-Guo L. (2011) Spatial Pattern Analysis of Heavy Metals in Beijing Agricultural Soils Based on Spatial Autocorrelation Statistics. *Journal of Environmental Research and Public Health*, 8, 2074-2089.
- Zalewski A. (2020) Zmiany wartości zużytych nawozów i środków ochrony roślin w krajach Unii Europejskiej w latach 2010-2018. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 20(35)(1), 78-87.
- Zeliaś A. (red.), Grabiński T., Ludwiczak B., Malina A. (1991) Ekonometria przestrzenna. PWE, Warszawa.

7. Metody porządkowania liniowego

Porządkowanie liniowe jest metodą przydatną do porównywania obiektów, których charakterystyka, z punktu widzenia pewnego zjawiska złożonego, wynika z wielu własności czy cech. Zakładamy, że własności czy cechy – nazywane zmiennymi diagnostycznymi – dane są jako zmienne numeryczne (mierzone w skali przedziałowej lub ilorazowej). Na ich podstawie tworzona jest zmienna, zwana syntetyczną, która pozwala uporządkować obiekty od najlepszego do najgorszego ze względu na badane kryterium. Utworzenie zmiennej syntetycznej odbywa się w kilku etapach, z których każdy może być przeprowadzony na wiele różnych sposobów.

Bąk [2013] zwraca uwagę, że koncepcję uporządkowania obiektów opisanych wieloma cechami przy pomocy miary syntetycznej, z zastosowaniem obiektu wzorcowego, jako pierwszy przedstawił Hellwig [1968]. Wielu autorów rozwijało następnie zaproponowaną metodę, wielu stosowało ją praktycznie.

Metody porządkowania liniowego opisane są w wielu publikacjach. Jako przykłady podać można prace następujące: Hellwig [1968], Bartosiewicz [1976], Pluta [1977], Borys [1978], Strahl [1978], Nowak [1989], Grabiński [1992], Walesiak [1993a], Kukuła [2000], Zeliaś [2002], Gatnar i Walesiak (red.) [2004], Młodak [2006], Pawełek [2008], Panek [2009], Wysocki [2010], Kukuła i Luty [2015a, 2015b].

7.1. Opis metod

Zastosowanie metod porządkowania liniowego przebiegać może w następujących etapach:

1. Dobór i selekcja zmiennych diagnostycznych.
2. Normalizacja zmiennych.
3. Wyznaczenie zmiennej lub zmiennych syntetycznych stanowiących agregaty zmiennych diagnostycznych. Sporządzenie rankingu lub rankingów (jeśli stosowanych jest równocześnie wiele metod).
4. Wybór rankingu ostatecznego.
5. Podział zbiorowości na klasy.

W pewnych przypadkach niektóre etapy mogą być pominięte, na przykład jeśli celem jest sporządzenie jedynie rankingu i nie ma potrzeby dzielenia zbiorowości na klasy, ostatni etap nie jest potrzebny. Etap wyboru rankingu ostatecznego też należy pominąć, jeśli zastosowano jedynie jedną metodę.

7.1.1. Dobór i selekcja zmiennych diagnostycznych

Wybór zmiennych diagnostycznych wynika ściśle z charakteru zjawiska złożonego, jakim zainteresowany jest badacz. Różne zestawy zmiennych mogą skutkować różną oceną jego poziomu. Wstępny dobór prowadzony jest przede wszystkim według kryteriów merytorycznych, przy czym konieczne jest uwzględnienie dostępności danych. W następnym kroku prowadzona jest selekcja na podstawie analizy statystycznej. Do jej przeprowadzenia najczęściej wykorzystuje się wartości współczynników korelacji Pearsona pomiędzy zmiennymi diagnostycznymi. Współczynnik ten jest miarą liniowego związku między zmiennymi. Jeśli występuje podejrzenie o zależność nieliniową, można przeprowadzić odpowiednie przekształcenia zmiennych. Współczynnik korelacji obliczony dla zmiennych po transformacji, może być wówczas traktowany jako miara zależności nieliniowej. Jeśli korelacja między zmiennymi diagnostycznymi jest zbyt silna, zasoby informacji w nich zawarte są dublowane, zwiększając ich wpływ na zmienną syntetyczną. Konieczne jest założenie pewnej granicznej wartości współczynnika korelacji, powyżej którego eliminowana będzie jedna ze zmiennych. Pamiętać należy, że metody porządkowania liniowego są stosowane dla całych populacji (nie mają bowiem sensu w przypadku prób losowych). Wobec tego nie jest możliwe zastosowanie testu statystycznego na istotność współczynnika korelacji (błędy tego typu często pojawiają się w literaturze). Wybierając graniczną wartość współczynnika korelacji, należy wziąć pod uwagę, że poziom zbyt wysoki prowadzi do wspomnianej możliwości dublowania informacji, ale równocześnie nie może być zbyt niski, ponieważ może nastąpić znacząca jej utrata.

Zagadnienie doboru zmiennych są szerzej omawiane między innymi w pracach: Grabiński, Wydymus i Zeliaś [1982], Nowak [1984], Kukuła [2000], Panek [2009], Tarka [2010] i wielu innych.

7.1.2. Normalizacja zmiennych

Normalizacja zmiennych ma na celu sprowadzenie ich do podobnego rzędu wielkości. Można ją przeprowadzić wieloma sposobami, które porównywał Walesiak [2014]. Formuły stosowane do normalizacji zmiennych zależą od rodzaju ich związku ze zjawiskiem złożonym. Wyróżniamy stymulanty, destymulanty i nominanty (rozważania zostaną ograniczone do nominanty jednodomodalnej, zdefiniowanej przez Borysa [1984, patrz: Walesiak, Gatnar (red.) 2004]). Stymulanta (destymulanta) to zmienna, której duże (małe) wartości są pożądane z punktu widzenia badanego zjawiska złożonego. Im większa (mniejsza) wartość stymulanta (destymulanta), tym lepiej. W przypadku nominanty pożądany jest określony jej poziom. Im wartość nominanta jest bardziej oddalona od niego (wszystko jedno w którą stronę), tym gorzej oceniana jest jednostka ze względu na tę cechę. Destymulanty i nominanty można przekształcić na stymulanty przed normalizacją zmiennych.

Można też prowadzić normalizację, wykorzystując różne formuły dla stymulant, destymulant i nominant. W niniejszej pracy przedstawione zostaną formuły umożliwiające zamianę destymulant i nominant w stymulanty, a następnie formuły normalizacyjne jedynie dla stymulant

Zamianę destymulant w stymulanty można wykonać według wzorów podanych przez Walesiaka i Gatnara [(red.) 2004], stosując przekształcenie ilorazowe lub różnicowe. Przekształcenie ilorazowe jest następujące:

$$x_{ij} = b \cdot (x_{ij}^D)^{-1}, \quad (7.1)$$

gdzie $x_{ij}^D \neq 0$ jest wyjściową wartością j -tej cechy (destymulanty) dla i -tego obiektu, a $b > 0$ to stała przyjmowana arbitralnie (np. $b = 1$, czy $b = \min_i \{x_{ij}^D\}$). Przekształcenie różnicowe określa formuła:

$$x_{ij} = a - b \cdot x_{ij}^D, \quad (7.2)$$

gdzie a i $b > 0$ są stałymi przyjmowanymi arbitralnie (np. $b = 1$, $a = 0$ czy $a = \max_i \{x_{ij}^D\}$).

Obydwa przekształcenia można stosować dla zmiennych diagnostycznych mierzonych w skali ilorazowej. Jeśli zmienna diagnostyczna mierzona jest w skali przedziałowej, jedynym możliwym przekształceniem jest przekształcenie różnicowe.

Zamianę nominant jednomodalnych w stymulanty można również przeprowadzić, stosując przekształcenie ilorazowe [Walesiak, Gatnar (red.) 2004]:

$$x_{ij} = \frac{\min\{nom_j; x_{ij}^N\}}{\max\{nom_j; x_{ij}^N\}} \quad (7.3)$$

lub różnicowe:

$$x_{ij} = -|x_{ij}^N - nom_j|. \quad (7.4)$$

We wzorach (7.3) i (7.4) x_{ij}^N jest wyjściową wartością j -tej cechy (nominanty) dla i -tego obiektu, a nom_j jest pożądanym poziomem tej cechy. Ograniczenie dla stosowania przekształcenia ilorazowego jest takie samo jak dla destymulant (tylko dla zmiennych mierzonych w skali ilorazowej).

W tabeli 7.1 przedstawiono wybrane formuły normalizacyjne w wersji dla stymulant. Więcej wariantów normalizacji można znaleźć na przykład w pracach: Kukuła [2000], Zeliaś [2002], Walesiak i Gatnar (red.) [2004], Walesiak [2014]. Standaryzację pozycyjną przedstawiono zgodnie z publikacjami: Lira, Wagner i Wysocki [2002], Łuczak i Wysocki [2013].

Tabela 7.1. Wybrane formuły normalizujące stymulanty

Nazwa metody	Przekształcenie dla stymulant
Standaryzacja	$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$
Standaryzacja pozycyjna	$z_{ij} = \frac{x_{ij} - m\tilde{e}d_j}{1,4826 \cdot m\tilde{a}d_j}$
Unitaryzacja zerowana	$z_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}$
Metoda Nowaka	$z_{ij} = \frac{x_{ij}}{\bar{x}_j}$
Metoda Strahl	$z_{ij} = \frac{x_{ij}}{\max_i x_{ij}}$
Normalizacja na przedziale [-1,1]	$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\max_i x_{ij} - \bar{x}_j }$

Oznaczenia zastosowane w tabeli:

z_{ij} unormowana i -ta obserwacja zmiennej j -tej, dla $i=1, \dots, n$, oraz $j=1, \dots, k$, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$, $s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$, $m\tilde{a}d_j = med_i |x_{ij} - m\tilde{e}d_j|$, gdzie wektor $\hat{\theta} = [m\tilde{e}d_1, m\tilde{e}d_2, \dots, m\tilde{e}d_k]^T$, jest wektorem medianowym Webera, będącym rozwiązaniem zadania optymalizacyjnego: $\min_{\theta_j} \sum_{i=1}^n \sqrt{\sum_{j=1}^k (x_{ij} - \theta_j)^2}$. Symbol med_i oznacza medianę w zbiorze obiektów.

7.1.3. Wyznaczenie zmiennej lub zmiennych syntetycznych

Zmienne syntetyczne stanowią agregaty zmiennych diagnostycznych, na podstawie których sporządzane są rankingi. Na podstawie unormowanych zmiennych wyznaczone są zmienne syntetyczne, które stanowią podstawę opracowania rankingów. Są dwie grupy metod ich tworzenia: metody bezwzorcowe i wzorcowe. W metodach bezwzorcowych przeprowadza się agregację zmiennych syntetycznych. Najprostszą metodą agregacji jest obliczenie średniej arytmetycznej lub sumy unormowanych zmiennych diagnostycznych. Jeżeli są przesłanki do uznania różnego wkładu poszczególnych zmiennych na badane zjawisko złożone, można obliczyć średnią ważoną. Jest wiele metod doboru wag dla poszczególnych zmiennych syntetycznych, które zostaną dalej omówione. W przypadku metod wzorcowych wykorzystuje się wartości zmiennych pewnych obiektów wybranych jako wzorcowe (mogą być to obiekty fikcyjne). Wzorcem może być tzw. dolny lub górny biegun rozwoju. Dolny biegun rozwoju (dla stymulant oraz destymulant i nominant przekształconych na stymulanty) tworzy fikcyjny obiekt o najmniejszych wartościach zmiennych diagnostycznych, górny zaś o największych.

W tabeli 7.2 przedstawiono różne metody tworzenia zmiennych syntetycznych – zarówno wzorcowe, jak i bezwzorcowe. W niektórych z nich założono jednakowe wagi, z jakimi zmienne wprowadzane są do zmiennych syntetycznych, w innych uwzględniono wagi zróżnicowane.

Wiele metod agregacji przedstawiono na przykład w pracy Kukuły [2000]. Metodę pozycyjną Hellwiga przedstawiono zgodnie z publikacją Liry, Wagnera i Wysockiego [2002] oraz Łuczak i Wysockiego [2013].

Metodą bezwzorcową, w której omija się problem doboru wag dla zmiennych syntetycznych, jest metoda PCA (ang. *principal component analysis*), czyli metoda pierwszej głównej składowej. Procedurę konstruowania zmiennej syntetycznej przy jej użyciu przedstawia Bąk [2018].

Tabela 7.2. Wybrane metody tworzenia zmiennych syntetycznych

Nazwa metody	Zmienna syntetyczna q_i dla i -tego obiektu
Metody bezwzorcowe	
Wagi jednakowe	$q_i = \frac{1}{k} \sum_{j=1}^k z_{ij}$ lub $q_i = \sum_{j=1}^k z_{ij}$
Wagi zróżnicowane	$q_i = \sum_{j=1}^k w_j \cdot z_{ij}$, gdzie $\sum_{j=1}^k w_j = 1$
Średnia geometryczna	$q_i = \sqrt[k]{\prod_{j=1}^k z_{ij}}$
Metody wzorcowe	
Hellwiga	$q_j = 1 - \frac{d_i^+}{d_0}, \text{ gdzie: } d_i^+ = \sqrt{\sum_{j=1}^k (z_{ij} - z_j^+)^2}$ $d_0 = \bar{d} + 2 \cdot s_d, \quad \bar{d} = \frac{1}{n} \sum_{i=1}^n d_i^+, \quad s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i^+ - \bar{d})^2}$
Pozycyjna Hellwiga	$q_i = 1 - \frac{d_i^+}{d_0}, \text{ gdzie: } d_i^+ = \text{med}_j z_{ij} - z_j^+ $ $d_0 = \text{med}_i(d_i^+) + 2,5 \cdot \text{mad}(d_i^+),$ $\text{mad}(d_i^+) = \text{med}_i d_i^+ - \text{med}_i(d_i^+) $
TOPSIS	$q_j = \frac{d_i^-}{d_i^- + d_i^+}, \text{ gdzie: } d_i^- = \sqrt{\sum_{j=1}^k (z_{ij} - z_j^-)^2}, d_i^+ \text{ jak wyżej}$
Uogólniona miara odległości (GDM) ¹	$q_{iW} = \frac{1}{2} - \frac{L}{2 \cdot M}$ $L = - \sum_{j=1}^k w_j \cdot (z_{ij} - z_{Wj})^2 + \sum_{j=1}^k \sum_{l=1, l \neq i, W}^n w_j \cdot (z_{ij} - z_{lj}) \cdot (z_{Wj} - z_{lj})$ $M = \sqrt{\sum_{j=1}^k \sum_{l=1}^n w_j \cdot (z_{ij} - z_{lj})^2 \cdot \sum_{j=1}^k \sum_{l=1}^n w_j \cdot (z_{Wj} - z_{lj})^2} \quad i = 1, \dots, n \text{ są}$ <p style="text-align: center;">numerami obiektów</p>

Oznaczenia zastosowane w tabeli: w_j – waga dla zmiennej diagnostycznej x_j . Fikcyjny obiekt o współrzędnych $z_j^- = \min_i z_{ij}$ oraz $z_j^+ = \max_i z_{ij}$, dla $j = 1, \dots, k$, nazywane są odpowiednio dolnym i górnym biegunem rozwoju.

Uwaga: ¹ 1. W tabeli podano GDM jedynie dla wariantu zmiennych diagnostycznych mierzonych w skalach przedziałowej i ilorazowej. Formuły dla przypadku zmiennych wyrażonych w skali jedynie porządkowej znaleźć można w pracach Walesiaka [2003] czy Gatnara i Walesiaka [(red.) 2004]. 2. Jako wzorzec rozwoju W przyjmuje się albo dolny, albo górny biegun rozwoju w zależności od sposobu uporządkowania wartości zmiennych syntetycznych na potrzeby tworzonego rankingu. Przyjmuje się, że im odległość obiektu od dolnego bieguna rozwoju jest większa, tym wyższa jego pozycja w rankingu, natomiast im odległość obiektu od górnego bieguna rozwoju jest większa, tym pozycja jest niższa.

Jeśli istnieje podejrzenie, że poszczególne zmienne diagnostyczne mają różne udziały w ocenie badanego zjawiska złożonego, należy w formułach agregujących zastosować dla nich różne wagi. Wagi te można dobierać metodą ekspercką [Kukuła 2000] bądź „naturalną” wynikającą z udziałów części składowych w badanym zjawisku złożonym [Kukuła 2000, za Nowak 1977]. Można także przeprowadzić analizę statystyczną zmiennych diagnostycznych.

Jako przykłady metod stosujących statystyczne własności cech można wymienić następujące:

- metodę opartą na współczynnikach zmienności [Grabiński, Wydymus, Zeliaś 1989],
- metodę opartą na współczynnikach korelacji [Grabiński, Wydymus, Zeliaś 1989],
- metodę CCSD (ang. *correlation coefficient and standard deviation*) [Wang, Luo 2010; Łuczak, Wysocki 2014],
- metodę Bettiego i Vermy [Betti, Verma 2008; Panek 2009],
- metodę CRITIC (ang. *criteria importance through intercriteria correlation*) [Diakoulaki, Mavrotas, Papayannakis 1995].

Dwie pierwsze metody wymagają podania jedynie prostych formuł pozwalających obliczyć wagi, z dwoma pozostałymi wiążą się bardziej złożone procedury ich wyznaczania. Postuluje się, aby wagi spełniały dwa warunki: $0 \leq w_j \leq 1$, dla $j = 1, \dots, k$, oraz $\sum_{j=1}^k w_j = 1$.

W metodzie opartej na współczynnikach zmienności preferowane są zmienne diagnostyczne (będą miały większe wagi), które cechuje większa zmienność mierzona współczynnikiem klasycznym. Wagi te obliczane są zgodnie ze wzorem:

$$w_j = \frac{v_j}{\sum_{m=1}^k v_m}, \quad (7.5)$$

gdzie v_j jest klasycznym współczynnikiem zmienności znormalizowanej j -tej zmiennej.

W metodzie stosującej współczynniki korelacji większe wagi uzyskują te cechy, których poziom skorelowania ze zmiennymi pozostałymi jest wysoki. Wagi te obliczane są jako:

$$w_j = \frac{\sum_{m=1}^k |r_{jm}|}{\sum_{j=1}^k \sum_{m=1}^k |r_{jm}|}. \quad (7.6)$$

Metodę CCSD zaproponowali Wang i Luo [2010], w polskiej literaturze przedstawiają ją Łuczak i Wysocki [2014]. W metodzie tej wagi nadawane poszczególnym zmiennym diagnostycznym są rozwiązaniem następującego zadania optymalizacji z ograniczeniami:

$$\begin{aligned} \min_{w_j} \sum_{j=1}^k \left(w_j - \frac{\sigma_j \sqrt{1-r_j}}{\sum_{m=1}^k \sigma_m \sqrt{1-r_m}} \right)^2, \\ \sum_{j=1}^k w_j = 1, \\ w_j \geq 0, j = 1, \dots, k, \end{aligned} \quad (7.7)$$

gdzie $\sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2}$ jest odchyleniem standardowym znormalizowanej j -tej zmiennej, a r_j współczynnikiem korelacji między znormalizowaną j -tą zmienną i zmienną syntetyczną wyznaczoną z cech pozostałych. Współczynnik r_j obliczany jest według wzoru:

$$r_j = \frac{\sum_{i=1}^n (z_{ij} - \bar{z}_j) \cdot (q_{ij} - \bar{q}_j)}{\sqrt{\sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 \cdot \sum_{i=1}^n (q_{ij} - \bar{q}_j)^2}}, \quad (7.8)$$

gdzie $\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij}$, $\bar{q}_j = \frac{1}{n} \sum_{i=1}^n q_{ij}$, $q_{ij} = \sum_{m=1, m \neq j}^k w_m \cdot z_{im}$.

Betti i Verma [2008] proponują, aby waga była wprost proporcjonalna do zmienności cechy diagnostycznej i odwrotnie proporcjonalna do jej skorelowania z cechami pozostałymi. Wagi mogą być skalowane do wartości dowolnej⁹, choć jak zauważają, zwykle skalowane są do 1. Betti i Verma zaproponowali, aby waga była iloczynem dwóch wag cząstkowych:

$$w_j = w_j^a \cdot w_j^b. \quad (7.9)$$

Człon pierwszy, wynikający ze współczynnika zmienności cechy, może być obliczony na przykład zgodnie ze wzorem (7.5) [Panek 2009]. Człon drugi, wynikający ze skorelowania cechy, jest następujący [Panek 2009]:

$$w_j^b = \left(\frac{1}{1 + \sum_{m=1, r_{jm} < r^*}^k |r_{jm}|} \right) \cdot \left(\frac{1}{\sum_{m=1, r_{jm} \geq r^*}^k |r_{jm}|} \right), \quad (7.10)$$

gdzie r^* jest progową wartością współczynnika korelacji. Betti i Verma [2008] zaproponowali, aby tę wartość określić na podstawie największej luki występującej pomiędzy uporządkowanymi niemalejąco współczynnikami korelacji (opis tej procedury znajduje się w pracy [Dudek, Szczesny, 2015]). Propozycję opartą na formule minimaksowej podaje Panek [2009]:

$$r^* = \min_j \max_m |r_{jm}|, \text{ dla } j \neq m, j, m = 1, \dots, k. \quad (7.11)$$

W oryginalnym wzorze określającym składową w_j^b [Betti, Verma, 2008] nie ma wartości bezwzględnych dla współczynników korelacji. Dudek i Szczesny [2015] tę rozbieżność wyjaśniają tym, że na etapie obliczania zmiennej syntetycznej wszystkie zmienne po normalizacji są stymulantami, wobec czego współczynniki korelacji między nimi są dodatnie.

W metodzie CRITIC [Diakoulaki, Mavrotas, Papayannakis 1995] wagi obliczane są jako:

$$w_j = \frac{c_j}{\sum_{m=1}^k c_m}, c_j = \sigma_j \cdot \sum_{m=1}^k (1 - r_{jm}), \quad (7.12)$$

gdzie σ_j jest odchyleniem standardowym znormalizowanej j -tej zmiennej diagnostycznej, a r_{jm} współczynnikiem korelacji znormalizowanych zmiennych j -tej i m -tej.

⁹ Chodzi o to, aby suma wag miała określoną wartość. W przypadku zmiennej syntetycznej nie ma bowiem znaczenia, czy mnożymy ją, czy dzielimy dla wszystkich obiektów przez dowolną liczbę.

7.1.4. Wybór rankingu ostatecznego

Każdy sposób normalizacji i tworzenia zmiennych syntetycznych prowadzi do różnych wartości zmiennych syntetycznych, a tym samym kolejność obiektów w uzyskanych rankingach może być różna. Jeśli do sporządzenia rankingu wykorzystano wiele metod, pojawia się potrzeba wyboru spośród nich rankingu ostatecznego, ustalającego pozycję wszystkich obiektów. Kukuła [1986, 2012, 2014; Kukuła, Luty 2015a] zaproponował następującą miarę podobieństwa rankingów:

$$\bar{u}_p = \frac{1}{\nu-1} \sum_{l=1, l \neq p}^{\nu} \left(1 - \frac{2 \cdot \sum_{i=1}^n |c_{ip} - c_{il}|}{n^2 - e} \right), \quad (7.13)$$

gdzie ν jest liczbą sporządzonych rankingów, c_{ip} pozycją i -tego obiektu w p -tym rankingu, a $e = \begin{cases} 0, & n \bmod 2 = 0 \\ 1, & n \bmod 2 = 1 \end{cases}$. Najlepszym rankingiem jest ranking najbardziej podobny do pozostałych, czyli taki, dla którego \bar{u}_p jest największe.

Walesiak [1993a, 1993b] zaproponował dwa mierniki podobieństwa zbioru obiektów w czasie oparte na wartościach zmiennych syntetycznych. Mierniki te można wykorzystać do porównania dwóch różnych rankingów. Pierwszy z nich może być stosowany dla zmiennych mierzonych w skali ilorazowej i przedziałowej, drugi zaś dla skali od porządkowej. Zastosowanie miernika drugiego (współczynnika τ Kendalla) dla zmiennych mierzonych w skali silniejszej wiąże się z utratą informacji (podobnie jak w przypadku miary podobieństwa rankingów).

Dalej przedstawiony zostanie jedynie miernik pierwszy. Drugi miernik można znaleźć w pracach Walesiaka [1993a, 1993b]. Do konstrukcji pierwszego miernika, autor wykorzystał koncepcję Theila na miernik rzędu dokładności prognozy typu *ex post* [Walesiak 1993b; Gatnar, Walesiak (red.) 2004].

Miernik podobieństwa oparty na wartościach p -tej i l -tej zmiennej obliczany jest jako:

$$\begin{aligned} P_{pl} &= \frac{1}{n} \sqrt{P_1^2 + P_2^2 + P_3^2}, \\ P_1^2 &= \bar{q}_p - \bar{q}_l, \\ P_2^2 &= s_p - s_l, \\ P_3^2 &= 2 \cdot s_p \cdot s_l \cdot (1 - r_{pl}), \end{aligned} \quad (7.14)$$

gdzie $\bar{q}_p = \frac{1}{n} \sum_i^n q_{ip}$, a q_{ip} to p -ta zmienna syntetyczna dla i -tego obiektu, s_p jest odchyleniem standardowym p -tej zmiennej syntetycznej, r_{pl} jest współczynnikiem korelacji między p -tą i l -tą zmienną syntetyczną.

Jeżeli nie ma różnic między p -tą i l -tą zmienną syntetyczną, to miernik ten jest równy 0. Wartość miernika P_{pl} informuje, jaki jest przeciętny rząd odchyłeń wartości porównywanych cech syntetycznych. Mierniki takie można policzyć między p -tą zmienną syntetyczną i wszystkimi pozostałymi. Na ich podstawie można obliczyć łączną miarę niepodobieństwa p -tej zmiennej syntetycznej z pozostałymi jako:

$$\bar{y}_p = \frac{1}{\nu-1} \sum_{l=1, l \neq p}^{\nu} P_{pl}. \quad (7.15)$$

Dla rankingu najbardziej podobnego do pozostałych \bar{y}_p będzie najmniejszy (odwrotnie niż w przypadku \bar{u}_p).

Znacznie prostszą koncepcję porównania rankingów przedstawiła Kisieleńska [2016]. Propozycja polega na obliczeniu zagregowanej pozycji i -tego obiektu w ν rankingach:

$$wz_i = \frac{1}{\nu} \sum_{l=1}^{\nu} c_{il}, \quad (7.16)$$

gdzie c_{il} to pozycja i -tego obiektu w l -tym rankingu. Na podstawie wartości zagregowanej pozycji i -tego obiektu można wyznaczyć jego pozycję w rankingu ostatecznym¹⁰.

Zastosowanie zagregowanej pozycji i -tego obiektu w ν rankingach również wiąże się z utratą informacji. Można więc pokusić się o wyznaczenie zagregowanej wartości zmiennej syntetycznej dla i -tego obiektu, czyli:

$$qz_i = \frac{1}{\nu} \sum_{l=1}^{\nu} q_{il}, \quad (7.17)$$

gdzie q_{il} to l -ta zmienna syntetyczna dla i -tego obiektu. Wówczas o pozycji i -tego obiektu w rankingu ostatecznym decydować będzie zagregowana wartość zmiennej syntetycznej qz_i .

7.1.5. Podział zbiorowości na klasy

Zmienna syntetyczna (także zagregowana) umożliwia podział badanej zbiorowości na klasy. W literaturze spotyka się podziały głównie na trzy do pięciu klas, różne są też reguły ich przeprowadzania. Oczywiście jest, że liczba klas powinna być dostosowana do liczebności badanej zbiorowości. Zbyt mała ich liczba spowoduje zaliczenie do jednej klasy obiektów znacznie różniących się od siebie, z kolei zbyt duża utrudnia analizę. Liczbę klas można określić zgodnie z zaleceniami sformułowanymi dla przypadku tworzenia szeregów rozdzielczych [Wasilewska 2011].

Najbardziej popularne sposoby podziału oparte są na wartości średniej \bar{q} i odchyleniu standardowym s_q zmiennej syntetycznej. Najczęściej wyróżniane są cztery klasy w sposób następujący ([Wysocki i Kozera 2012] i wielu innych):

- klasa I – poziom wysoki: $q_i \geq \bar{q} + s_q$,
- klasa II – poziom średni-wyższy: $\bar{q} \leq q_i < \bar{q} + s_q$,
- klasa III – poziom średni-niższy: $\bar{q} - s_q \leq q_i < \bar{q}$,
- klasa IV – poziom niski: $q_i < \bar{q} - s_q$.

Podział taki ma dwie słabe strony. Pierwsza z nich wynika z tego, że typowy obszar zmienności cechy¹¹ dzielony jest na dwie części, a druga z możliwych różnic w liczebnościach klas. Jeśli założone

¹⁰ Problem pojawi się, gdy różne obiekty mają taką samą pozycję zagregowaną. W takim przypadku może rozstrzygnąć wartość miary kolejnej, podanej we wzorze (7.17).

zostanie, że zmienna syntetyczna ma rozkład normalny, to przy tym podziale w klasach I i IV znajdzie się po 15,9% obiektów, odpowiednio o jej największych i najmniejszych wartościach. Klasy II i III zawierać będą po 34,1% obiektów, dla których wartość zmiennej syntetycznej różni się od jej średniego poziomu o mniej niż odchylenie standardowe. Kisielińska zaproponowała bardziej równomierny podział, wyróżniając pięć klas obejmujących po 20% obiektów [Kisielińska 2017]. Wartościami granicznymi są wówczas kwintyle rozkładu normalnego, a taki podział jest następujący:

- klasa I (bardzo wysoki poziom badanego zjawiska): $\bar{q} + 0,8416 \cdot s_q \leq q_i$,
- klasa II (wysoki poziom): $\bar{q} + 0,2533 \cdot s_q \leq q_i < \bar{q} + 0,8416 \cdot s_q$,
- klasa III (średni poziom): $\bar{q} - 0,2533 \cdot s_q \leq q_i < \bar{q} + 0,2533 \cdot s_q$,
- klasa IV (niski poziom): $\bar{q} - 0,8416 \cdot s_q \leq q_i < \bar{q} - 0,2533 \cdot s_q$,
- klasa V (bardzo niski poziom): $q_i < \bar{q} - 0,8416 \cdot s_q$.

Jeżeli zmienna syntetyczna ma rozkład różny od normalnego, udziały obiektów w poszczególnych klasach mogą być oczywiście różne od 20%.

Na podstawie średniej i odchylenia standardowego zmiennej syntetycznej zbiorowość jest dzielona w sposób następujący:

- klasa I (wysoki poziom badanego zjawiska): $\bar{q} + P \cdot s_q \leq q_i$,
- klasa II (średni poziom): $\bar{q} - P \cdot s_q \leq q_i < \bar{q} + P \cdot s_q$,
- klasa III (niski poziom): $q_i < \bar{q} - P \cdot s_q$.

Jeśli przyjęte zostanie $P = 1$, mogą być zaburzone proporcje w liczebności poszczególnych klas (przy założeniu normalności rozkładu w klasach I i III będzie zaledwie po 16% obiektów, podczas gdy w II 68% – czyli większość). Lepszym rozwiązaniem jest przyjęcie mniejszego mnożnika dla odchylenia standardowego (np. $W = 0.5$ [Ossowska 2015]).

Zakres zmienności zmiennej syntetycznej może być dzielony na równe pod względem szerokości części poprzez podział rozstępu na K równych części, gdzie K to założona liczba klas (np. [Kukuła 2014]). Można także dokonać podziału, który zapewni równą lub niemal równą liczebność poszczególnych klas. W takim przypadku należy wyznaczyć kwantyle odpowiedniego rzędu badanej populacji. I tak mediana dzieli zbiorowość na dwie równoliczne części, tercyle, kwartyle na cztery, kwintyle na pięć, decyle na dziesięć a percentyle na sto.

Często stosowanym w praktyce rozwiązaniem jest arbitralne przyjęcie granic dla poszczególnych klas.

¹¹ Klasyczny typowy obszar zmienności dla cechy y stanowi przedział $[\bar{y} - s_y, \bar{y} + s_y]$, gdzie \bar{y} jest średnią, a s_y odchyleniem standardowym cechy [Wasilewska 2011].

7.2. Zastosowania metod porządkowania liniowego

Metoda porządkowania liniowego jest jedną z najczęściej stosowanych metod WAD w ekonomice rolnictwa. Wynika to prawdopodobnie z tego, że przedmiotem badania jest cała populacja, a badany zbiór obiektów nie musi spełniać żadnych warunków wstępnych. Nie ma też praktycznie możliwości porażki – każdy sporządzony ranking może być podstawą analiz merytorycznych. Pozwala na porównywanie obiektów ze względu na pozycję w nim zajętą, wynikającą z wartości zmiennej syntetycznej reprezentującej badane zjawisko złożone. Jeśli na bazie zmiennej syntetycznej dokonywany jest podział zbiorowości na klasy (grupy) obiektów o podobnym jej poziomie, można dodatkowo porównywać wyróżnione klasy.

W tabeli 7.3 przedstawiono przykłady zastosowań porządkowania liniowego w ekonomice rolnictwa (przegląd taki pochodzący z lat 2002-2007 przedstawili Kisielińska i Stańko [2009]). Badanymi obiektami najczęściej były jednostki terytorialne – województwa (szesnaście pozycji) i gminy, głównie wiejskie (czternaście pozycji) i oraz państwa UE (dziewięć pozycji). Znacznie rzadziej badano regiony (trzy pozycje), powiaty, przedsiębiorstwa czy typy produkcyjne gospodarstw (zaledwie po jednej pozycji).

Badania prowadzone na poziomie gmin obejmowały przede wszystkim gminy wiejskie i najczęściej dotyczyły sytuacji społeczno-gospodarczej na ich terenie. Jeśli chodzi o problemy społeczne, to badano warunki życia ludności, warunki mieszkaniowe i sytuację demograficzną. Analizy gospodarcze dotyczyły poziomu przedsiębiorczości, atrakcyjności dla prowadzenia działalności gospodarczej na obszarach wiejskich oraz możliwości ich inteligentnego rozwoju. Wiele prac poświęcono również sytuacji finansowej gmin. Rzadziej zajmowano się ściśle rolnictwem czy aspektem środowiskowym obszarów wiejskich.

W przypadku województw badane zjawiska złożone odnosiły się natomiast głównie do potencjału rolnictwa, jego efektywności i wyposażenia technicznego (osiem pozycji). Badano także poziom kapitału ludzkiego i intelektualnego na wsi (cztery pozycje). Problematyka badawcza objęła ponadto gospodarkę odpadami, ekonomiczne zrównoważenie rolnictwa oraz potencjał do produkcji żywności ekologicznej (po jednym artykule).

Większość badań na poziomie państw prowadzono na podstawie danych pochodzących z gospodarstw rolniczych. Odnosiły się one do ich sytuacji finansowej i produkcyjnej, pozycji konkurencyjnej, możliwości rozwojowych czy efektywności wytwarzania. Badano ponadto poziom rolnictwa ekologicznego, konkurencyjność przemysłu spożywczego, kwestie tworzenia i wykorzystania wiedzy na potrzeby biogospodarki. Jedną publikację poświęcono wykluczeniu społecznemu na wsi.

Na poziomie regionów UE badano jakość życia na obszarach wiejskich, na poziomie makroregionów UE potencjał produkcyjny gospodarstw mlecznych, a na poziomie polskich podregionów dezagryzując produkcję w gospodarstwach rolnych.

W jednej publikacji sporządzono ranking typów produkcyjnych gospodarstw ze względu na ich sytuację finansowo-majątkową, w jednej badano rozwój przedsiębiorstw mleczarskich oraz przetwórstwa owocowo-warzywnego. Także w jednej badano stan środowiska w powiatach.

Tematyka badań jest różnorodna, co świadczy o dużych możliwościach aplikacyjnych metod porządkowania liniowego w badaniach społeczno-ekonomicznych.

Wykorzystując metody porządkowania liniowego, badacz podejmuje decyzje co do wyboru stosowanych rozwiązań. Na etapie normalizacji należy kierować się własnościami różnych metod, biorąc pod uwagę cel prowadzonych badań [Zeliaś 2002]. Agregacja prowadząca do uzyskania zmienionych syntetycznych także może być zrealizowana w rozmaity sposób, różnie można także obliczać odległości między obiektami. Warto więc zastosować kilka metod i porównać uzyskane wyniki, ponieważ dobór metody ma bezpośredni wpływ na kolejność obiektów w tworzonych rankingach. Spośród przedstawionych w tabeli 7.3 przykładów jedynie w sześciu pracach stosowano więcej niż jedną metodę i spośród uzyskanych rankingów wybierano najlepszy. Pięciokrotnie zastosowano miarę podobieństwa rankingów Kukuły (wzór 7.13), a jeden raz dodatkowo zagregowaną pozycję obiektu (wzór 7.16) [Kisielińska 2016]. Bąk zastosował współczynnik korelacji rang Spearmana oraz zagregowany miernik jakości procedur porządkowania liniowego [Bąk 2018].

Najczęściej stosowaną metodą normalizacji była metoda unitaryzacji zerowanej (w trzydziestu jeden pracach), a następnie standaryzacja (w dziesięciu pracach) oraz standaryzacja pozycyjna (w sześciu pracach). Zarówno metodę Strahl, jak i metodę Nowaka zastosowano trzykrotnie, a normalizację na przedziale $[-1,1]$ jedynie raz. W trzech publikacjach nie podano metody normalizacji.

Najpopularniejszą metodą agregacji była agregacja bezwzorcowa (w dwudziestu siedmiu pracach). Metodę TOPSIS zastosowano w czternastu artykułach, a Hellwiga w osiemnastu (przy czym w sześciu w wersji pozycyjnej). Uogólnioną miarę odległości użyto raz [Kisielińska 2017], również raz wykorzystano pierwszą główną składową [Bąk 2018].

Różne wagi podczas agregacji zastosowano jedynie w trzech publikacjach. W jednym przypadku zastosowano metodę CRITIC przy agregacji metodą TOPSIS [Bieniasz, Gołaś 2015], w jednym wagi określano arbitralnie [Ossowska 2015] i w jednej pracy zastosowano kilka metod doboru wag (na podstawie współczynników zmienności, współczynników korelacji, metodą CCSD oraz metodą Bettiego i Vermy [Kisielińska 2018]).

Zmienne syntetyczne były wykorzystywane do tworzenia rankingu obiektów, ale również do ich grupowania. Klas nie wyróżniono jedynie w siedmiu pracach. Najczęściej wyróżniano cztery klasy, następnie trzy oraz pięć. W jednym przypadku nie założono z góry liczby klas i zastosowano metodę klasyfikacji Warda [Kijek, Chojnacki 2016].

Najczęściej autorzy wyróżniali cztery klasy obiektów na podstawie średniej i odchylenia standardowego zmiennej syntetycznej (w trzynastu pracach), dwukrotnie na podstawie odchylenia w ujęciu pozycyjnym [Müller-Frączek, Muszyńska 2015; Muszyńska, Müller-Frączek 2015]. Dwukrotnie do podziału wykorzystano kwartyle ([Bąk 2018; Prus i in. 2021]), a raz rozstęp podzielono na cztery równe części ([Kukuła, Luty 2018]). Autorzy pięciu publikacji nie wyjaśnili, jak podział przeprowadzili.

Trzy klasy obiektów wyróżniono w dziesięciu pracach, przy czym w czterech przypadkach podziału dokonano na podstawie średniej i odchylenia standardowego zmiennej syntetycznej, w dwóch zaś podzielonego na trzy równe części rozstępu ([Kukuła 2014; Zwolińska-Ligaj i in. 2018]) i w dwóch granice przyjęto arbitralnie ([Wojewodzic 2014; Kutkowska i in. 2019]). Pozostałe podziały zastosowano jednokrotnie, a dokonano ich na podstawie tercylów [Molenda-Grysa 2016] i percentyli [Pietrzak 2018].

Podziału na pięć klas dokonano w pięciu artykułach: dwukrotnie granice przyjęto arbitralnie [Hadyński 2014; Janiszewska 2017], dwukrotnie według kwintyli rozkładu normalnego [Kisielińska 2017, 2018] i jeden raz na podstawie średniej i odchylenia standardowego zmiennej syntetycznej [Pomianek, Kapaj 2018].

Tabela 7.3. Przykłady zastosowań porządkowania liniowego w ekonomice rolnictwa

Autor	Problem	Obiekty	Metoda	Klasyfikacja
Poczta, Bartkowiak 2012	potencjał produkcyjny rolnictwa	województwa	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Wysocki, Kozera 2012	potencjał produkcyjny rolnictwa i efektywność wykorzystania czynników produkcji	województwa	unitaryzacja zerowana, TOPSIS	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Baer-Nawrocka, Markiewicz 2013	efektywność wytwarzania w rolnictwie	państwa	unitaryzacja zerowana, TOPSIS	wyróżniono 4 klasy – nie wyjaśniono, jak dokonano podziału
Adamska, Golinowska 2014	wymiar środowiskowy obszarów wiejskich	gminy	unitaryzacja zerowana, agregacja bezwzorcowa	nie wyróżniono klas
Hadyński 2014	jakość życia na obszarach wiejskich	regiony UE	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 5 klas na podstawie wartości zmiennej syntetycznej, granice określone arbitralnie
Juchniewicz 2014	konkurencyjność przemysłu spożywczego	państwa	standaryzacja, agregacja bezwzorcowa	nie wyróżniono klas
Kowalewska 2014	kapitał ludzki na obszarach wiejskich	województwa	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 3 klasy, rozstęp zmiennej syntetycznej podzielono na 3 równe części
Kukuła 2014	wyposażenie techniczne rolnictwa	województwa	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 3 klasy, rozstęp zmiennej syntetycznej podzielono na 3 równe części
Nowak, Kamińska, Różańska-Baczuła 2014	potencjał produkcyjny rolnictwa	województwa	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Rzeszutko 2014	efektywność produkcyjna rolnictwa	województwa	unitaryzacja zerowana, agregacja metodą Hellwiga	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Wojewodziec 2014	dezagraryzacja produkcyjno- -ekonomiczna gospodarstw	podregiony	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 3 klasy na podstawie wartości zmiennej syntetycznej, granice określone arbitralnie
Bąk, Szczecińska 2015	warunki życia w gminach wiejskich	gminy	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej

Bieniasz, Gołaś 2015	kondycja finansowa gmin	gminy	metoda TOPSIS z wagami wyznaczonymi metodą CRITIC	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Kukuła, Luty 2015b	poziom rolnictwa ekologicznego	państwa	(1) unitaryzacja zerowana, agregacja metodą Hellwiga, (2) unitaryzacja zerowana, agregacja metodą TOPSIS, (3) standaryzacja pozycyjna, agregacja metodą pozycyjną Hellwiga, (4) standaryzacja, agregacja bezwzorcowa, (5) unitaryzacja zerowana, agregacja bezwzorcowa, (6) normalizacja metodą Strahl, agregacja bezwzorcowa	nie wyróżniono klas
Lisek 2015	rozwój przedsiębiorstw mleczarskich i przetwórstwa owocowo-warzywnego	przedsiębiorstwa	unitaryzacja zerowana, agregacja bezwzorcowa (suma)	nie wyróżniono klas
Müller-Frączek, Muszyńska 2015	kapitał intelektualny mieszkańców wsi, poziom życia na wsi	województwa	standaryzacja pozycyjna, agregacja metodą pozycyjną Hellwiga	wyróżniono 4 klasy metodą odchyłań w ujęciu pozycyjnym na podstawie mediany
Muszyńska, Müller-Frączek 2015	kapitał intelektualny mieszkańców wsi	województwa	standaryzacja pozycyjna, agregacja metodą pozycyjną Hellwiga	wyróżniono 4 klasy metodą odchyłań w ujęciu pozycyjnym na podstawie mediany
Ossowska 2015	poziom przedsiębiorczości na obszarach wiejskich	gminy	normalizacja metodą , agregacja bezwzorcowa ze zróżnicowanymi wagami przyjętymi arbitralnie	wyróżniono 3 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Głowicka-Wołoszyn, Wysocki 2016	kondycja finansowa gmin wiejskich	gminy	unitaryzacja zerowana, agregacja metodą TOPSIS	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Janiszewska, Ossowska 2016	możliwości rozwojowe funkcji rolniczej	państwa	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 3 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Kijek, Chojnacki 2016	zdolności krajów do tworzenia i wykorzystania wiedzy na potrzeby biogospodarki	państwa	standaryzacja, agregacja metodą Hellwiga	metodą Warda

Tabela 7.3 (cd.)

Kisielińska 2016	potencjalne możliwości zaspokojenia zapotrzebowania na produkty rolnicze	państwa	(1) standaryzacja, agregacja metodą Hellwiga, (2) standaryzacja, agregacja metodą TOPSIS, (3) standaryzacja pozycyjna, agregacja metodą pozycyjną Hellwiga, (4) standaryzacja, agregacja bezwzorcowa, (5) unitaryzacja zerowana, agregacja bezwzorcowa, (6) normalizacja metodą Strahl, agregacja bezwzorcowa, (7) normalizacja metodą Nowaka, agregacja bezwzorcowa	nie wyróżniono klas
Molenda-Grysa 2016	potencjał do produkcji żywności ekologicznej	województwa	normalizacja na przedziale [-1, 1], agregacja bezwzorcowa	wyróżniono 3 klasy, podziału dokonano na podstawie tercylu
Pomianek, Gralak 2016	kondycja finansowa obszarów wiejskich	gminy	agregacja metodą Hellwiga	wyróżniono 3 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Tłuczak 2016	ekonomiczne zrównoważenie rolnictwa	województwa	standaryzacja, agregacja metodą Hellwiga	wyróżniono 4 klasy – nie wyjaśniono, jak dokonano podziału
Janiszewska 2017	rolnicze nakłady produkcyjne	gminy	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 5 klas na podstawie wartości zmiennej syntetycznej, granice określone arbitralnie
Kisielińska 2017	potencjał rolnictwa	województwa	(1) standaryzacja, agregacja bezwzorcowa, (2) unitaryzacja zerowana, agregacja bezwzorcowa, (3) normalizacja metodą Nowaka, agregacja bezwzorcowa, (4) normalizacja metodą Strahl, agregacja bezwzorcowa, (5) standaryzacja, agregacja metodą Hellwiga, (6) standaryzacja, agregacja metodą TOPSIS, (7) standaryzacja pozycyjna, agregacja metodą pozycyjną Hellwiga, (8) standaryzacja, uogólniona miara odległości	wyróżniono 5 klas na podstawie kwintyli rozkładu normalnego
Misztal 2017	sytuacja finansowo-majątkowa gospodarstw rolnych	typy produkcyjne	metoda Strahl, agregacja bezwzorcowa	nie wyróżniono klas
Rosa, Jakubowska 2017	wykluczenie społeczne na obszarach wiejskich	państwa	zmodyfikowana unitaryzacja zerowana, agregacja bezwzorcowa średnią geometryczną	wyróżniono 4 klasy – nie wyjaśniono, jak dokonano podziału
Rzeszutko 2017	poziom rozwoju rolnictwa	województwa	unitaryzacja zerowana, agregacja metodą Hellwiga	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej

Bąk 2018	stan środowiska	powiaty	(1) standaryzacja, agregacja metodą Hellwiga, (2) standaryzacja, agregacja metodą TOPSIS, (3) standaryzacja, metoda bezwzorcową wg pierwszej głównej składowej	wyróżniono 4 klasy na podstawie kwartyli
Kisielińska 2018	sytuacja towarowych gospodarstw rolnych	państwa	unitaryzacja zerowana, agregacja bezwzorcową oraz wagi dobierane: na podstawie współczynników zmienności, współczynników korelacji, metodą CCSD oraz metodą Bettiego i Vermy	wyróżniono 5 klas na podstawie kwintyli rozkładu normalnego
Kukuła, Luty 2018	gospodarka odpadami w Polsce	województwa	(1) standaryzacja, agregacja metodą Hellwiga, (2) standaryzacja, agregacja metodą TOPSIS, (3) standaryzacja pozycyjna, agregacja metodą pozycyjną Hellwiga, (4) standaryzacja, agregacja bezwzorcową, (5) unitaryzacja zerowana, agregacja bezwzorcową, (6) normalizacja metodą Strahl, agregacja bezwzorcową, (7) normalizacja metodą Nowaka, agregacja bezwzorcową	wyróżniono 4 klasy, rozstęp zmiennej syntetycznej podzielono na 4 równe części
Pietrzak 2018	poziomu kapitału ludzkiego na obszarach wiejskich	województwa	unitaryzacja zerowana, agregacja bezwzorcową	wyróżniono 3 klasy, podziału dokonano wg percentyli
Poczta, Średzińska, Mikołajczyk 2018	sytuacja finansowa gospodarstw ogrodniczych	państwa	unitaryzacja zerowana, agregacja metodą TOPSIS	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Pomianek, Kapaj 2018	sytuacja demograficzna obszarów wiejskich i półmiejskich	gminy	agregacja metodą Hellwiga	wyróżniono 5 klas na podstawie średniej arytmetycznej i odchylenia standardowego zmiennej syntetycznej
Standar 2018	kondycja finansowa gmin	gminy	standaryzacja, agregacja metodą TOPSIS	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennych syntetycznych
Zwolińska-Ligaj, Guzal-Dec, Adamowicz 2018	inteligentny rozwój jednostek terytorialnych na obszarach wiejskich	gminy	unitaryzacja zerowana, agregacja bezwzorcową	wyróżniono 3 klasy, rozstęp zmiennej syntetycznej podzielono na 3 równe części

Tabela 7.3 (cd.)

Głowicka-Wołoszyn, Stanisławska, Wołoszyn 2019	warunki mieszkaniowe ludności na obszarach wiejskich i miejskich województwa wielkopolskiego	gminy	unitaryzacja zerowana, TOPSIS	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennych syntetycznych
Kutkowska i in. 2019	poziom rozwoju społeczno-gospodarczego obszarów wiejskich	gminy	standaryzacja, agregacja bezwzorcowa	wyróżniono 3 klasy na podstawie wartości zmiennej syntetycznej, granice określone arbitralnie
Paszkowski, Sarniak 2019	różnice w potencjale rozwojowym obszarów wiejskich	województwa	unitaryzacja zerowana, agregacja metodą TOPSIS	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennych syntetycznych
Prus, Dziekański 2019	poziomu rozwoju a sytuacja finansowa gmin wiejskich województwa świętokrzyskiego	gminy	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 4 klasy – nie wyjaśniono, jak dokonano podziału
Smędzik-Ambroży, Rutkowska, Kirbas 2019	zdolność konkurencyjna polskiego sektora rolnego	kraje UE	unitaryzacja zerowana, agregacja bezwzorcowa	nie wyróżniono klas
Kozera-Kowalska, Uglis 2020	atrakcyjności gmin wiejskich dla rozwoju działalności gospodarczej w Polsce	gminy wiejskie	unitaryzacja zerowana, agregacja bezwzorcowa	wyróżniono 4 klasy na podstawie średniej arytmetycznej i odchylenia standardowego zmiennych syntetycznych
Baer-Nawrocka, Bartczak 2020	potencjał produkcyjny gospodarstw	makroregiony UE	agregacja metodą Hellwiga	wyróżniono 4 klasy – nie wyjaśniono, jak dokonano podziału
Prus i in. 2021	potencjał rolnictwa	województwa	unitaryzacja zerowana, TOPSIS	wyróżniono 4 klasy na podstawie kwartyli

Źródło: opracowanie własne

Literatura

- Adamska H., Golinowska M. (2014) Identyfikacja i ocena wymiaru środowiskowego obszarów wiejskich województwa dolnośląskiego w aspekcie zrównoważonego rozwoju. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 16(2), 9-14.
- Baer-Nawrocka A., Bartczak W. (2020) Production potential of holdings specializing in cow milk production in macroregions of the European Union – a typological analysis. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 22(1), 11-19.
- Baer-Nawrocka A., Markiewicz N. (2013) Relacje między czynnikami produkcji a efektywność wytwarzania w rolnictwie Unii Europejskiej. *Journal of Agribusiness and Rural Development*, 3(29), 5-16.
- Bartosiewicz S. (1976) Propozycja metody tworzenia zmiennych syntetycznych. *Zeszyty Naukowe Akademii Ekonomicznej we Wrocławiu*, 84.
- Bąk A. (2013) Metody porządkowania liniowego w polskiej taksonomii – pakiet pllord. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 20, 54-62.
- Bąk A. (2018) Zastosowanie metod wielowymiarowej analizy porównawczej do oceny stanu środowiska w województwie dolnośląskim. *Wiadomości Statystyczne*, 1, 7-20.
- Bąk I., Szczecińska B. (2015) Zastosowanie metod taksonomicznych w badaniu warunków życia w gminach wiejskich województwa zachodniopomorskiego. *Journal of Agribusiness and Rural Development*, 1(35), 7-15.
- Betti G., Verma V. (2008) Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective. *Statistical Methods and Applications*, 17, 225-250.
- Bieniasz A., Gołaś Z. (2015) Kondycja finansowa gmin wiejskich Wielkopolski w latach 2007-2011. *Journal of Agribusiness and Rural Development*, 1(35), 27-37.
- Borys T. (1978) Metody normowania cech w statystycznych badaniach porównawczych. *Przegląd Statystyczny*, 25(2), 227-239.
- Borys T. (1984) Kategoria jakości w statystycznej analizie porównawczej. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 284, Monografie i Opracowania, 23.
- Diakoulaki D., Mavrotas G., Papayannakis L. (1995). Determining objective weights in multiple criteria problems: The critic method. *Computers & Operations Research*, 22(7), 763-770.
- Dudek H., Szczesny W. (2015) Zastosowanie funkcji przynależności w analizie subiektywnego postrzegania jakości życia. *Ekonometria*, 4(50), 62-78.
- Gatnar E., Walesiak M. (red.) (2004) *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Głowicka-Wołoszyn R., Wysocki F. (2016) Kondycja finansowa gmin wiejskich a źródła ich dochodów w województwie wielkopolskim. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 18(1), 50-58.
- Głowicka-Wołoszyn R., Stanisławska J., Wołoszyn A. (2019) Wielowymiarowa ocena warunków mieszkaniowych ludności na obszarach wiejskich i miejskich województwa wielkopolskiego. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 21(2), 79-87.
- Grabiński T. (1992) *Metody taksonometrii*. Wydawnictwo AE, Kraków.
- Grabiński T., Wydymus S., Zeliaś A. (1982) *Metody doboru zmiennych w modelach ekonometrycznych*. PWN, Warszawa.
- Grabiński T., Wydymus S., Zeliaś A. (1989) *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*. Państwowe Wydawnictwo Naukowe, Warszawa.
- Hadyński J. (2014) Zróżnicowanie jakości życia na obszarach wiejskich Unii Europejskiej. *Journal of Agribusiness and Rural Development*, 2(32), 69-77.
- Hellwig Z. (1968) Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, 15(4), 307-327.

- Janiszewska D. (2017) Spatial differentiation of agricultural productive inputs in zachodniopomorskie voivodeship. *Polityki Europejskie, Finanse i Marketing*, 22(71), 763-770.
- Janiszewska D., Ossowska L. (2016) Zróżnicowanie rolnictwa krajów Unii Europejskiej na podstawie wybranych cech. *Problemy Rolnictwa Światowego*, 16(2), 134-144.
- Juchniewicz M. (2014) Tendencje zmian międzynarodowej konkurencyjności przemysłu spożywczego państw Unii Europejskiej. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 101(1), 31-40.
- Kijek T., Chojnacki P. (2016) Ocena zdolności krajów Unii Europejskiej do tworzenia i wykorzystania wiedzy na potrzeby biogospodarki. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 18(5), 72-77.
- Kisielińska J. (2016) Ranking państw UE ze względu na potencjalne możliwości zaspokojenia zapotrzebowania na produkty rolnicze z wykorzystaniem metod porządkowania liniowego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(3), 142-152.
- Kisielińska J. (2017) Ranking województw ze względu na potencjał rolnictwa. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 104(1), 56-71.
- Kisielińska J. (2018) Ocena sytuacji towarowych gospodarstw rolnych państw UE z wykorzystaniem metod porządkowania liniowego. *Problemy Rolnictwa Światowego*, 18(1), 66-79.
- Kisielińska J., Stańko S. (2009) Wielowymiarowa analiza danych w ekonomice rolnictwa. *Roczniki Nauk Rolniczych, Seria G*, 96(2), 63-76.
- Kowalewska M. (2014) Terytorialne zróżnicowanie poziomu kapitału ludzkiego na obszarach wiejskich. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 16(1), 105-110.
- Kozera-Kowalska M., Uglis J. (2020) Przestrzenne zróżnicowanie atrakcyjności gmin wiejskich dla rozwoju działalności gospodarczej w Polsce. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 22(2), 88-98.
- Kukuła K. (1986) Propozycja miary zgodności układów porządkowych. *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, 22.
- Kukuła K. (2000) Metoda unitaryzacji zerowanej. *Wydawnictwo Naukowe PWN, Warszawa*.
- Kukuła K. (2012) Propozycja budowy rankingu obiektów z wykorzystaniem cech ilościowych oraz jakościowych. *Metody Ilościowe w Badaniach Ekonomicznych*, 13(1), 5-16.
- Kukuła K. (2014) Budowa rankingu województw ze względu na wyposażenie techniczne rolnictwa w Polsce. *Wiadomości Statystyczne*, 7, 62-76.
- Kukuła K., Luty L. (2015a) Propozycja procedury wspomagającej wybór metody porządkowania liniowego. *Przegląd Statystyczny*, 62(2), 219-231.
- Kukuła K., Luty L. (2015b) Ranking państw UE ze względu na wybrane wskaźniki charakteryzujące rolnictwo ekologiczne. *Metody Ilościowe w Badaniach Ekonomicznych*, 16(3), 225-236.
- Kukuła K., Luty L. (2018) O wyborze metody porządkowania liniowego do oceny gospodarki odpadami w Polsce. *Problemy Rolnictwa Światowego*, 18(2), 183-192.
- Kutkowska B., Pilawka T., Rybchak V., Rybchak O. (2019) Zróżnicowanie poziomu rozwoju społeczno-gospodarczego obszarów wiejskich województwa dolnośląskiego w latach 2002 i 2010. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 21(2), 170-187.
- Lira J., Wagner W., Wysocki F. (2002) Mediana w zagadnieniach porządkowania obiektów wielocechowych, [w:] J. Paradyś (red.) *Statystyka regionalna w służbie samorządu lokalnego i biznesu*. Internetowa Oficyna Wydawnicza Centrum Statystyki Regionalnej, AE w Poznaniu, Poznań, 87-99.
- Lisek S. (2015) Rozwój wybranych przedsiębiorstw mleczarskich i przetwórstwa owocowo-warzywnego w latach 2010-2012. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(2), 139-143.
- Łuczak A., Wysocki F. (2013) Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 20, 63-73.
- Łuczak A., Wysocki F. (2014) Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 22, 49-59

- Miształ A. (2017) Ewaluacja sytuacji finansowo-majątkowej polskich gospodarstw rolnych w latach 2004-2014. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 117, 47-59.
- Młodak A. (2006) Analiza taksonomiczna w statystyce regionalnej. Difin, Warszawa.
- Molenda-Grysa I. (2016) Przesłanki do produkcji żywności ekologicznej w Polsce – typologia potencjałów województw. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(3), 66-74.
- Müller-Frączek I., Muszyńska J. (2015) Kapitał intelektualny polskiej wsi a poziom życia jej mieszkańców. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(6), 198-203.
- Muszyńska J., Müller-Frączek I. (2015) Ocena stopnia zróżnicowania kapitału intelektualnego polskiej wsi. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 17(4), 179-184.
- Nowak A., Kamińska A., Różańska-Baczuła M. (2014) Przestrzenne zróżnicowanie potencjału produkcyjnego rolnictwa w Polsce. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 30, 363-372.
- Nowak E. (1977) Syntetyczne mierniki plonów w krajach europejskich. *Wiadomości Statystyczne*, 10, 19-22.
- Nowak E. (1984) Problemy doboru zmiennych do modelu ekonometrycznego. PWN, Warszawa.
- Nowak E. (1989) Metody taksonomiczne w klasyfikacji obiektów społeczno-ekonomicznych. PWE, Warszawa.
- Ossowska L. (2015) Zróżnicowanie poziomu przedsiębiorczości obszarów wiejskich województwa zachodniopomorskiego w zależności od położenia wobec miast. *Journal of Agribusiness and Rural Development*, 4(38), 811-820.
- Panek T. (2009) Statystyczne metody wielowymiarowej analizy porównawczej. Oficyna Wydawnicza SGH, Warszawa.
- Paszkowski S., Sarniak Ł. (2019) Differences in rural development potential across Polish regions. *Journal of Agribusiness and Rural Development*, 1(51), 59-68.
- Pawełek B. (2008) Metody normalizacji zmiennych w badaniach porównawczych złożonych zjawisk ekonomicznych. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, Seria specjalna: Monografie*, 187, Kraków.
- Pietrzak P. (2018) Klasyfikacja obszarów wiejskich w Polsce pod względem poziomu kapitału ludzkiego. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 20(5), 154-159.
- Pluta W. (1977) Wielowymiarowa analiza porównawcza w badaniach ekonomicznych. PWN, Warszawa.
- Poczta W., Bartkowiak N. (2012) Regionalne zróżnicowanie rolnictwa w Polsce. *Journal of Agribusiness and Rural Development*, 1(23), 95-109.
- Poczta W., Średzińska J., Mikołajczyk P. (2018) Sytuacja finansowa gospodarstw ogrodniczych w Unii Europejskiej – ujęcie syntetyczne. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 20(5), 167-172.
- Pomianek I., Gralak A. (2016) Ocena kondycji finansowej obszarów wiejskich w województwie mazowieckim w latach 2004-2015 według wybranych koncepcji podziału regionu. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 18(6), 169-175.
- Pomianek I., Kapaj A. (2018) Demographic changes in rural and semi-urban areas in Poland (2003-2016). *Annals of Marketing Management & Economics*, 4(12), 89-101.
- Prus P., Dziekański P. (2019) Zróżnicowanie przestrzenne poziomu rozwoju a sytuacja finansowa gmin wiejskich województwa świętokrzyskiego. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 21(1), 74-82.
- Prus P., Dziekański P., Bogusz M., Szczepanek M. (2021) Spatial Differentiation of Agricultural Potential and the Level of Development of Voivodeships in Poland in 2008-2018. *Agriculture*, 11(3), 1-29.
- Rosa A., Jakubowska A. (2017) Zagrożenie wykluczeniem społecznym na obszarach wiejskich krajów Unii Europejskiej. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 19(3), 244-249.
- Rzeszutko A. (2014) Efektywność produkcyjna rolnictwa w Polsce a relacje między czynnikami produkcji. *Journal of Agribusiness and Rural Development*, 2(32), 141-154.
- Rzeszutko A. (2017) Zmiany struktury ekonomicznej gospodarstw rolnych w Polsce – ocena opóźnień w rozwoju strukturalnym. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 104(4), 48-61.

- Smędzik-Ambroży K., Rutkowska M., Kirbas H. (2019) Zdolność konkurencyjna polskiego sektora rolnego na tle krajów Unii Europejskiej w latach 2004-2017 na podstawie gospodarstw rolnych FADN. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 21(3), 422-431.
- Standar A. (2018) Realizacja wydatków inwestycyjnych a kondycja finansowa JST. Przykład gmin wiejskich województwa wielkopolskiego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 19(68), 248-261.
- Strahl D. (1978) Propozycja konstrukcji miary syntetycznej. *Przegląd Statystyczny*, 25(2), 205-215.
- Tarka D. (2010) Własności cech diagnostycznych w badaniach typu taksonomicznego. *Ekonomia i Zarządzanie*, 4, 194-205.
- Tłuczak A. (2016) Typologia województw pod względem ekonomicznego zrównoważenia rolnictwa. *Annals of the Polish Association of Agricultural and Agribusiness Economists*, 18(4), 219-223.
- Walesiak M. (1990) Syntetyczne badania porównawcze w świetle teorii pomiaru. *Przegląd Statystyczny*, 37(2), 37-46.
- Walesiak M. (1993a) Statystyczna analiza wielowymiarowa w badaniach marketingowych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 654, Monografie i Opracowania, 101.
- Walesiak M. (1993b) Zagadnienie oceny podobieństwa zbioru obiektów w czasie w syntetycznych badaniach porównawczych. *Przegląd Statystyczny*, 1, 95-102.
- Walesiak M. (2003) Uogólniona miara odległości GDM jako syntetyczny miernik rozwoju w metodach porządkowania liniowego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 10, 134-144.
- Walesiak M. (2014) Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej. *Przegląd Statystyczny*, 61(4), 363-372.
- Walesiak M., Gatnar E. (red.) (2004) *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Wang Y.M., Luo Y. (2010) Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making. *Mathematical and Computer Modelling*, 51, 1-12.
- Wasilewska E. (2011) *Statystyka opisowa od podstaw*. Wydawnictwo SGGW, Warszawa.
- Wojewodziec T. (2014) Dezagraryzacja produkcyjno-ekonomiczna gospodarstw rolnych w Polsce – próba pomiaru zjawiska. *Journal of Agribusiness and Rural Development*, 4(34), 213-223.
- Wysocki F. (2010) *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*. Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań.
- Wysocki F., Kozera A. (2012) Potencjał produkcyjny rolnictwa i efektywność wykorzystania czynników produkcji. *Wiadomości Statystyczne*, 4, 49-64
- Zeliaś A. (2002) Uwagi na temat wyboru metody normowania zmiennych diagnostycznych, [w:] T. Kufel, M. Piątowska (red.) *Analiza szeregów czasowych na początku XXI wieku*. Uniwersytet M. Kopernika w Toruniu, Toruń, 25-41.
- Zwolińska-Ligaj M., Guzal-Dec D., Adamowicz M. (2018) Koncepcja inteligentnego rozwoju lokalnych jednostek terytorialnych na obszarach wiejskich regionu peryferyjnego na przykładzie województwa lubelskiego. *Więś i Rolnictwo*, 2(179), 247-280.

8. Metoda głównych składowych oraz analiza czynnikowa

W analizie wielowymiarowych danych statystycznych często celem jest wskazanie istotnych zależności, jakie zachodzą między zmiennymi opisującymi obiekty wielowymiarowe [Ostasiewicz (red.) 1999]. Metody koncentrujące się na badaniu wewnętrznych powiązań w zbiorze analizowanych zmiennych noszą nazwę metod analizy powiązań. Należy do nich między innymi metoda głównych składowych PCA (ang. *principal components method*) oraz analiza czynnikowa FA (ang. *factor analysis*).

Celem analizy głównych składowych jest dekompozycja zmienności zbioru obserwacji na zbiór składowych w taki sposób, że pierwsza składowa wyjaśnia największą część zmienności, a druga największą część pozostałej zmienności itd. [Aczel 2005]. Analiza czynnikowa prezentuje inne podejście. W metodzie tej należy odnaleźć czynniki mające znaczenie w określeniu istoty badanego problemu. Zadaniem obu metod jest odpowiednia transformacja i ewentualna redukcja zestawu zmiennych [Gatnar, Walesiak (red.) 2004]. Transformacja pozwala na zastąpienie szeregu skorelowanych ze sobą zmiennych obserwowalnych zmiennymi nieobserwowalnymi, które skorelowane już nie będą. Redukcja wymiaru przestrzeni cech do dwóch daje na przykład możliwość przedstawienia obserwacji wielowymiarowych na płaszczyźnie. Ponadto przekształcone za pomocą obu metod zmienne mogą być wykorzystane w innych metodach (np. w regresji wielorakiej) [Kisielińska, Stańko 2009].

8.1. Opis metod

8.1.1. Metoda głównych składowych

Metodę głównych składowych zaproponował K. Pearson w 1901 roku, a potem Hotelling ją rozwinął i zastosował do analizy struktury zależności [Hotelling 1933].

Metoda ta polega na ortogonalnym przekształceniu p -wymiarowego zestawu zmiennych opisujących obserwacje wielowymiarowe w nowy zestaw zmiennych nieskorelowanych, tzw. głównych składowych. Transformacja taka wyznacza nowy zbiór zmiennych opisanych w nowym układzie współrzędnych (jeśli nowy układ współrzędnych ma ten sam wymiar, to przekształcenie takie jest po prostu obrotem w przestrzeni p -wymiarowej).

Związki między zmiennymi obserwowalnymi a zmiennymi nieobserwowalnymi można opisać odpowiednim równaniem macierzowym. Jeżeli $\mathbf{X} = [X_1, \dots, X_p]^T$ jest wektorem zmiennych obserwowalnych, to wektor nowych nieskorelowanych zmiennych $\mathbf{PC} = [PC_1, \dots, PC_p]^T$ – głównych składowych – jest następującą liniową transformacją wektora \mathbf{X} :

$$\mathbf{PC} = \mathbf{A}^T \mathbf{X}, \quad (8.1)$$

gdzie $\mathbf{A} = [A_1, \dots, A_p]$ jest tzw. macierzą przekształcenia ortogonalnego. Wobec tego j -tą główną składową $PC_j, j = 1, \dots, p$, definiuje się następująco:

$$PC_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p. \quad (8.2)$$

Główne składowe są więc kombinacjami liniowymi zmiennych obserwowalnych, ponadto są ze sobą nieskorelowane (ortogonalne) i unormowane (co oznacza, że $\sum_{k=1}^p a_{jk}^2 = 1, j = 1, \dots, p$).

Elementy macierzy \mathbf{A} określają udział każdej zmiennej obserwowalnej w tworzeniu poszczególnych składowych. Macierz \mathbf{A} znajduje się w sposób iteracyjny [Pluta 1986]:

- najpierw wyznacza się taki wektor A_1 , dla którego wariancja zmiennej PC_1 jest maksymalna,
- w kolejnym kroku należy dobrać wektor A_2 tak, aby wariancja PC_2 była maksymalna przy założeniu, że PC_1 i PC_2 są nieskorelowane,
- analogiczne czynności wykonuje się dla pozostałych zmiennych PC_j .

Kryterium maksymalizacji wariancji nowych zmiennych wynika stąd, że zakres zmienności głównych składowych powinien w jak największym stopniu odzwierciedlać zakres zmienności zmiennych obserwowalnych.

Formalnie dla zbioru p zmiennych obserwowalnych można wyznaczyć maksymalnie p głównych składowych. Za pomocą omawianej metody można również zastępować liczny zbiór zmiennych obserwowalnych mniejszą liczbą zmiennych. Wtedy główne składowe obejmują jedynie część całkowitej zmienności zmiennych wyjściowych. Często wyznaczenie już dwóch lub trzech składowych wystarczy, aby w wysokim stopniu odzwierciedlić zmienność zestawu wyjściowego, na przykład w 75%.

Transformacja wektora \mathbf{X} w \mathbf{PC} (czyli rozwiązanie układu równań $\mathbf{PC} = \mathbf{A}^T \mathbf{X}$) wymaga znalezienia macierzy \mathbf{A} przekształcenia ortogonalnego. Wyznacza się ją na podstawie macierzy kowariancji \mathbf{C} dla zmiennych obserwowalnych \mathbf{X} , wyrażanej jako

$$\mathbf{C} = \mathbf{A}\mathbf{A}^T, \quad (8.3)$$

gdzie $\mathbf{A} = [\lambda_l]$ to diagonalna macierz zawierająca wartości własne macierzy \mathbf{C} . Wartości własne macierzy \mathbf{C} są pierwiastkami następującego wielomianu charakterystycznego:

$$\det(\mathbf{C} - \lambda_l \mathbf{I}) = 0. \quad (8.4)$$

Wartości własne ułożone w kolejności nierosnącej $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ są wytycznymi budowy zestawów reprezentacji wyników w wybranym wymiarze. Przykładowo największa wartość własna jest związana z pierwszą osią i reprezentuje wariancję wyjaśnianą przez tę oś.

Każdej wartości własnej λ_l jest przypisany wektor własny V_l . Wektory własne są wyznaczone z równości $(\mathbf{C} - \lambda_l \mathbf{I})V_l = 0$. Ostatecznie elementy macierzy \mathbf{A} są obliczane jako:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}^{1/2}, \quad (8.5)$$

gdzie \mathbf{V} jest macierzą wektorów własnych macierzy kowariancji \mathbf{C} , a $\mathbf{\Lambda}$ to macierz diagonalna, której elementami diagonalnymi są uporządkowane malejąco wartości własne macierzy \mathbf{C} .

8.1.2. Analiza czynnikowa

Podobnie jak metoda głównych składowych analiza czynnikowa pozwala na zbadanie struktury obserwacji wielowymiarowych i pomaga w zrozumieniu wewnętrznych zależności, jakie zachodzą w zbiorze zmiennych opisujących te obserwacje [Ostasiewicz (red.) 1999]. Zakłada się, że zależności te są wywoływane przez pewien czynnik lub małą liczbę czynników, które nie są bezpośrednio obserwowalne (czynniki latentne). W metodzie buduje się model, który wyjaśnia zależności występujące w macierzy korelacji zmiennych opisujących obiekty wielowymiarowe.

Analiza czynnikowa została skonceptualizowana i rozwinięta w psychologii anglosaskiej. Po raz pierwszy metoda ta została opisana przez Ch. Spearmana, aczkolwiek pełny opis jej podstaw teoretycznych oraz praktycznych zastosowań podał dopiero L.L. Thurstone w latach 30. XX wieku [Spearman 1904; Thurstone 1935].

W modelu analizy czynnikowej wektor zmiennych rzeczywistych $\mathbf{X} = [X_1, \dots, X_p]^T$ jest iloczynem macierzy ładunków czynnikowych $\mathbf{A} = [a_{ij}]^T$ i wektora czynników wspólnych $\mathbf{F} = [F_1, \dots, F_k]^T$ (zmiennych nieobserwowalnych) powiększonym o wektor czynników specyficznych $\mathbf{u} = [u_1, \dots, u_p]^T$, czyli:

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \mathbf{u}. \quad (8.6)$$

Przyjmuje się więc, że każda z obserwowalnych zmiennych $X_i, i = 1, \dots, p$, jest liniową funkcją zmiennych latentnych $F_j, j = 1, \dots, k, (k < p)$ oraz czynnika specyficznego u_i :

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{ik}F_k + u_i. \quad (8.7)$$

Ponadto wymaga się, aby zmienne X_i były zestandaryzowane ($E(X_i) = 0, D^2(X_i) = 1$).

Celem metody jest ukazanie, iż związki między obserwowalnymi cechami statystycznymi są rezultatem oddziaływania zespołu pewnych nieobserwowalnych (ukrytych) czynników. Zmienne ukryte kształtują jednak zmienne rzeczywiste tylko z pewną dokładnością, określoną składnikiem losowym.

Zakłada się, że:

- czynniki wspólne F_j nie są ze sobą skorelowane (są ortogonalne),
- czynniki specyficzne u_i nie są ze sobą skorelowane,

- czynniki wspólne nie są skorelowane z czynnikami specyficznymi,
- czynniki wspólne są zestandaryzowane, tzn. $E(F_i) = 0, D^2(F_i) = 1$,
- czynniki specyficzne są niezależnymi zmiennymi losowymi o rozkładach normalnych, przy czym $E(u_i) = 0, D^2(u_i) = \sigma_{u_i}^2$.

Jeżeli spełnione są powyższe założenia, to wariancję zmiennej obserwowalnej X_i można wyrazić jako:

$$\begin{aligned} D^2(X_i) &= E(a_{i1}F_1 + a_{i2}F_2 + \dots + a_{ik}F_k + u_i)^2 = \\ &= a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2 + \sigma_{u_i}^2 = h_i^2 + \sigma_{u_i}^2 = 1. \end{aligned} \quad (8.8)$$

Wariancja zmiennej X_i jest więc sumą dwóch składników: zasobu zmienności wspólnej (wariancja wyjaśniona przez czynniki wspólne F_j) oraz zasobu zmienności specyficznej (wariancja uwzględniająca wpływ zmiennej specyficznej u_i). Ogólnie macierz kowariancji wektora \mathbf{X} można przedstawić w postaci:

$$\mathbf{C} = \mathbf{A}\mathbf{A}^T + \mathbf{\Psi}, \quad (8.9)$$

gdzie \mathbf{A} jest macierzą ładunków czynnikowych, a $\mathbf{\Psi}$ to diagonalna macierz, w której elementami głównej przekątnej są kolejne wariancje specyficzne. Dalsze obliczenia w analizie czynnikowej oparte są na macierzy współczynników korelacji opisujących zależności między wystandaryzowanymi zmiennymi X_i :

$$\mathbf{R}_{p \times p} = [r_{ij}] = \frac{1}{n} \mathbf{X}^T \mathbf{X}. \quad (8.10)$$

Elementy na przekątnej macierzy \mathbf{R} interpretowane są jako wariancja całkowita każdej zmiennej (korelacja zmiennej z nią samą) i wynoszą 1. Elementy poza główną przekątną, dla $i \neq j$, wynoszą:

$$r_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{D^2(X_i)D^2(X_j)}} = \sum_{m=1}^k a_{im}a_{jm}. \quad (8.11)$$

Przed przystąpieniem do wykonania procedury analizy czynnikowej należy jednak sprawdzić, czy zmienne X_i pozostają wobec siebie w określonych relacjach. Niewielkie korelacje między zmiennymi pozwoliłyby wątpić w możliwość utworzenia czynników wspólnych F_j . Na podstawie macierzy korelacji \mathbf{R} można wyznaczyć wartość wskaźnika Kaisera-Mayera-Olkinia [Keiser 1974]:

$$KMO = \frac{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2}{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2 + \sum_{i \neq j} \sum_{j \neq i} \hat{r}_{ij}^2} \in [0, 1], \quad (8.12)$$

gdzie r_{ij} to współczynnik korelacji między i -tą i j -tą zmienną, a \hat{r}_{ij} jest współczynnikiem korelacji cząstkowej między nimi [Zeliaś 2000]. Im większa wartość KMO , tym bardziej zasadne stosowanie analizy czynnikowej. Pożądane są wartości KMO większe od 0,5. Z kolei test sferyczności Bartletta służy do oceny istotności macierzy korelacji. Formuluje się w nim następujące hipotezy: $H_0: \mathbf{R} = \mathbf{I}$

oraz $H_1: \mathbf{R} \neq \mathbf{I}$. Prawdziwość H_0 oznacza, że zbiór współczynników jest nieistotny. Sprawdzianem jest statystyka testowa $\chi^2 = -\left(n - 1 - \frac{2p+5}{6}\right) \log(\det \mathbf{R})$, w której n wyraża liczebność próby, p liczbę zmiennych obserwowalnych, a \mathbf{I} to macierz jednostkowa. Statystyka ta ma rozkład χ^2 o liczbie stopni swobody $p(p-1)/2$. Jeśli $\chi^2 > \chi^2_*$, to H_0 odrzucamy i zbiór współczynników korelacji uznaje się za istotny.

Celem wyeliminowania czynników specyficznych (nieistotnych dla analizy) tworzy się zredukowaną macierz korelacji \mathbf{R}^1 :

$$\mathbf{R}_{p \times p}^1 = \mathbf{A}\mathbf{A}^T = [r_{ij}^1] = \begin{cases} h_i^2 & \text{dla } i = j \\ r_{ij} & \text{dla } i \neq j \end{cases}. \quad (8.13)$$

Wyznaczenie zasobów zmienności wspólnej może przebiegać:

- metodą największej korelacji (wtedy $h_i^2 = \max_k \{r_{ki}\}$ dla $i \neq k$),
- metodą triad (wtedy $h_i^2 = \frac{r_{ki} - r_{ji}}{r_{jk}}$ dla $i, j, k = 1, \dots, p, i \neq k$, a zmienne j oraz k są najsilniej skorelowane ze zmienną i),
- metodą korelacji przeciętnej (wtedy $h_i^2 = \frac{1}{p-1} \sum_{k=1}^p r_{ki}$ dla $i \neq k$).

Wprowadzenie macierzy zredukowanej pozwala napisać:

$$\mathbf{X} = \mathbf{A}\mathbf{F}. \quad (8.14)$$

Przy założeniu $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ wartości czynników wspólnych wyznacza się następująco:

$$\mathbf{F} = \mathbf{A}^{-1}\mathbf{X}. \quad (8.15)$$

Rozwiązanie tego układu równań uzyskuje się, wykorzystując różne techniki, na przykład omówioną wcześniej analizę głównych składowych, metody osi głównych czy metodę największej wiarygodności [Stanimir 2006].

Przykładowo stosując metodę głównych składowych, wyznacza się elementy macierzy \mathbf{A} na podstawie macierzy \mathbf{R}^1 :

$$\mathbf{R}^1 = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T, \quad (8.16)$$

gdzie $\mathbf{\Lambda} = [\lambda_l]$ jest diagonalną macierzą zawierającą wartości własne macierzy \mathbf{R}^1 . Wartości własne macierzy \mathbf{R}^1 są pierwiastkami wielomianu charakterystycznego $\det(\mathbf{R}^1 - \lambda_l \mathbf{I}) = 0$ i ustawione w kolejności nierosnącej $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ są wytycznymi budowy zestawów reprezentacji wyników analizy czynnikowej w wybranym wymiarze. Wartościom własnym przypisane są wektory własne wyznaczane z równości $(\mathbf{R}^1 - \lambda_l \mathbf{I})\mathbf{V}_l = 0$. Elementy macierzy ładunków modelu analizy czynnikowej \mathbf{A} są obliczane jako:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}^{1/2}, \quad (8.17)$$

gdzie \mathbf{V} to macierz wektorów własnych zredukowanej macierzy korelacji \mathbf{R}^1 , $\mathbf{\Lambda}$ jest macierzą diagonalną, której elementami diagonalnymi są uporządkowane malejąco wartości własne zredukowanej macierzy \mathbf{R}^1 .

Ładunki czynników wyrażają siłę i kierunek skorelowania zmiennej obserwowalnej ze stojącym przy ładunku czynnikiem wspólnym. Im większa wartość ładunku (bliższa 1 lub -1), tym większe jest powiązanie zmiennej rzeczywistej i zmiennej latentnej. Wartości ładunków selekcionują więc zmienne obserwowalne, co daje możliwość nadania interpretacji czynnikom wspólnym F_j .

Jak już wspomniano, celem analizy czynnikowej, a także metody głównych składowych może być redukcja rzeczywistego wymiaru przestrzeni zależności zmiennych. Postępując dalej, wyznacza się więc wariancje dla każdego czynnika wspólnego z osobna, a ich suma określa łączne zróżnicowanie wszystkich zmiennych objaśniających. Jeśli zatem dla pierwszych dwóch lub trzech czynników suma ich wariancji stanowi znaczną część łącznej sumy wariancji zmiennych obserwowalnych (co najmniej 75%), to na takiej liczbie czynników się zwykle poprzestaje.

Liczbę czynników wspólnych można wyznaczać, korzystając z następujących kryteriów [Stanimir 2006; Sztemberg-Lewandowska 2008]:

- kryterium „łokcia” wykorzystujące wykres osypiska (wykres tworzony z wartości własnych ułożonych nierosnąco): należy zachować tyle czynników, ile tworzy „zbocze”, natomiast zignorować te, które tworzą „osypisko”; numer wartości własnej, po przejściu której wykres ma przebieg łagodny, wskazuje na najlepszą liczbę czynników [Kim, Mueller 1978; Gatnar 1998; Hair i in. 1998; Gatnar, Walesiak (red.) 2004];
- kryterium Keisera: pod uwagę brane są tylko $\lambda_l > 1$ [Kim, Mueller 1978; Gatnar 1998; Hair i in. 1998];
- kryterium wyjaśnionej wariancji: należy pozostawić taką liczbę czynników, aby wyjaśniony został założony procent wariancji (np. 80% lub 90%), czyli liczba czynników wyznaczona jest przez nierówność $\sum_{l=1}^k \frac{\lambda_l}{\sum_{i=1}^p \lambda_i} > w_0$, gdzie arbitralnie określono $w_0 < 1$ [Gatnar 1998; Hair i in. 1998].

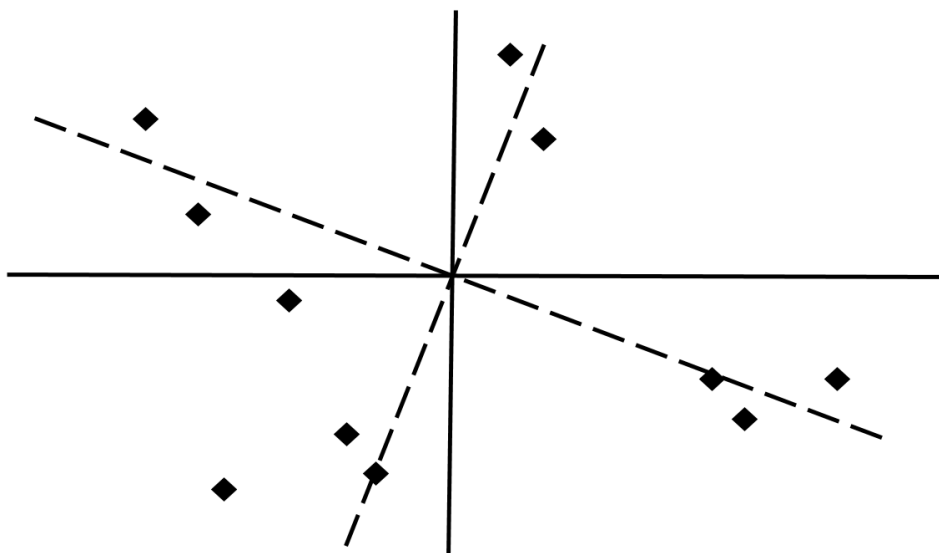
Rozwiązanie układu równań określających model analizy czynnikowej uzyskane jako pierwsze charakteryzuje się na ogół bardzo dużymi wartościami ładunków stojących przy pierwszym czynniku wspólnym dla większości zmiennych. Rodzi to trudności związane z interpretacją czynników latentnych. Wspomniany układ równań ma więcej niż jedno rozwiązanie (co dowodzi niejednoznaczności analizy czynnikowej), dlatego też poprzez tzw. rotację osi układu współrzędnych można uzyskać alternatywne rozwiązania. W trakcie rotacji znajdowana jest macierz ortogonalna \mathbf{B} , której elementy określają wielkości kątów, o jakie należy obrócić układ osi przestrzeni czynników wspólnych z położenia zadanego macierzą \mathbf{A} do położenia \mathbf{A}' (czyli wykonywane jest działanie $\mathbf{A}' = \mathbf{B}\mathbf{A}^T$). Przeprowadzana rotacja sprowadza się do wyzerowania lub zminimalizowania

niektórych ładunków, tak aby zmienna obserwowalna była kształtowana tylko przez jeden czynnik [Rószkiewicz 2002a].

Wyróżnia się rotacje ortogonalne i ukośne. Te pierwsze zachowują ortogonalność czynników, a drugie dodatkowo wprowadzają korelacje między czynnikami (bo rotują każdą oś oddzielnie), co jednak w konsekwencji utrudnia interpretację wyników.

Wartości ładunków określają współrzędne położenia poszczególnych zmiennych rzeczywistych w układzie osi czynnikowych (por. rys. 8.1). Rotacje z kolei wiążą się z różnymi sposobami obrotu osi czynnikowych. Obrót układu osi czynnikowych do położenia wskazanego na rysunku przerywanymi liniami powoduje, że w konsekwencji niektóre punkty zbliżają się do obróconych osi. Pozwoli to na lepsze pogrupowanie zmiennych obserwowalnych według ich siły powiązania z czynnikami wspólnymi.

Rotacja ortogonalna zachowuje kąt prosty między osiami współrzędnych. Wśród rotacji tego typu wyróżniają się techniki Varimax, Quatrimax, Equimax. Varimax skupia się na uproszczeniu kolumn macierzy ładunków czynnikowych, Quatrimax na uproszczeniu jej wierszy, a Equimax na transformacji kolumn i wierszy tej macierzy jednocześnie [Gatnar, Walesiak (red.) 2012; Tarka 2013]. Nowy układ osi ma być taki, aby odpowiadająca mu macierz ładunków czynnikowych pozwalała na interpretację czynników wspólnych. Macierz taka powinna więc spełniać kryteria prostej struktury [Mulaik 1972]. Niejednokrotnie ten ostatni etap analizy czynnikowej stanowi poważny problem, a interpretacja i nazwanie czynnika nie jest możliwe.



Rysunek 8.1. Rotacja ortogonalna osi czynnikowych zmieniająca współrzędne zmiennych rzeczywistych

Źródło: opracowanie własne

8.1.3. Porównanie analizy czynnikowej z metodą głównych składowych

W obu omówionych metodach wektor zmiennych obserwowanych \mathbf{X} jest określony przez wektor zmiennych nieobserwowalnych. W metodzie głównych składowych jest nim wektor składowych \mathbf{PC} , a w analizie czynnikowej wektor czynników wspólnych \mathbf{F} (patrz tabela 8.1). W obydwu metodach zadanie polega na wyznaczeniu macierzy przekształcenia tak, aby nowe zmienne nie były ze sobą skorelowane. Różnica między metodami polega na tym, że w metodzie głównych składowych zmienne nieobserwowalne w pełni objaśniają zmienne obserwowalne, w analizie czynnikowej uwzględnione są dodatkowo czynniki specyficzne [Kisielińska, Stańko 2009].

Tabela 8.1. Porównanie metody głównych składowych z analizą czynnikową

Metoda głównych składowych	Analiza czynnikowa
równanie podstawowe: $\mathbf{PC} = \mathbf{A}^T \mathbf{X}$	równanie podstawowe: $\mathbf{X} = \mathbf{A} \cdot \mathbf{F} + \mathbf{u}$
zmienne obserwowalne są liniową funkcją nieobserwowalnych głównych składowych	zmienne obserwowalne są liniową funkcją nieobserwowalnych czynników wspólnych oraz zmiennej specyficznej
cel podstawowy: transformacja zmiennych	cel podstawowy: redukcja zmiennych
uwzględniona jest tylko wariancja badanych zmiennych	uwzględniona jest wariancja badanych zmiennych oraz wariancja nieobjaśniona
punktem wyjścia jest całkowita macierz kowariancji	punktem wyjścia jest zredukowana macierz korelacji
zmienne nie muszą być losowe	zmienne muszą być losowe
zmienne nie muszą mieć rozkładu normalnego	zmienne powinny mieć rozkład normalny
zmienne nie muszą być wystandaryzowane	zmienne powinny być wystandaryzowane

Źródło: opracowanie własne na podstawie [Gatnar 2004]

8.2. Zastosowania metod głównych składowych i analizy czynnikowej

Zastosowanie metody głównych składowych lub analizy czynnikowej umożliwia redukcję zbioru zmiennych obserwowalnych. Może się to okazać użyteczne na przykład przy budowie modelu regresji wielorakiej, szczególnie gdy zbiór zmiennych objaśniających przekracza liczbę obserwacji. Redukcja zmiennych do dwóch lub trzech składowych nie powoduje dużego ubytku informacji, gwarantuje natomiast niezależność nowych zmiennych (dzięki ortogonalizacji uzyskiwane zmienne będą nieskorelowane), co jest bardzo ważne z punktu widzenia analizy regresji [Walesiak, Bąk 1997]. Należy nadmienić, że przeprowadzana transformacja skutkuje nowymi zmiennymi, którym często brak interpretacji.

Analiza czynnikowa pozwala na podzielenie zbioru zmiennych obserwowalnych na podzbiory, które są silniej determinowane przez pewne czynniki i słabiej determinowane przez inne. Zmiennie latentne wykorzystuje się potem w modelowaniu ekonometrycznym jako nowy rodzaj zmiennych wyrażających zjawiska nieobserwowalne. Zmiennie te stają się nawet podstawą grupowania obiektów [Rószkiewicz 1998].

Obie metody po raz pierwszy zastosowano w psychologii [Spearman 1904; Hotelling 1933; Thurstone 1935]. Analiza czynnikowa w psychologii jest najczęściej kojarzona z badaniami inteligencji, choć jest również wykorzystywana w badaniach osobowości, postaw, przekonań, w psychometrii oraz w badaniach międzykulturowych. W innych dziedzinach nauki wyróżniają się zastosowania w naukach fizycznych takich jak geochemia, hydrochemia, astrofizyka i kosmologia, a także w naukach biologicznych, na przykład w ekologii, biologii molekularnej, neuronauce i biochemii (przykładowo analiza czynnikowa może być użyta do podsumowania danych z mikromacierzy oligonukleotydowych DNA przy badaniu ekspresji genów).

Najczęstsze ekonomiczne zastosowania analizy czynnikowej dotyczą badań marketingowych. Metoda ta służy wtedy do przeprowadzenia segmentacji rynku, określenia preferencji konsumentów, badania produktów lub cen, analizy efektywności działań reklamowych czy konstrukcji tzw. map percepcji [Walesiak, Bąk 1997; Rószkiewicz 2002a, 2002b; Gatnar, Walesiak (red.) 2004; Stanimir 2006; Sztemberg-Lewandowska 2008]. Metody redukcji danych są wykorzystywane również w badaniach rynku pracy [Landmesser 2010].

Zastosowania metody głównych składowych oraz analizy czynnikowej w badaniach z zakresu ekonomiki rolnictwa są różnorodne, a ich przydatność wydaje się być duża.

Na podstawie przedstawionego w tabeli 8.2 przeglądu problemów, w których rozwiązywaniu pomocne były zaprezentowane metody, można stwierdzić, że jednym z obszarów ich wykorzystania jest badanie zachowań konsumentów na rynkach produktów żywnościowych. Z pierwotnie dużego zbioru zmiennych, w wyniku przeprowadzenia analizy czynnikowej, otrzymuje się kilka czynników wspólnych, które decydują na przykład wyborze danego produktu przez konsumenta. W literaturze można znaleźć przykłady na identyfikację czynników determinujących wybory konsumentów na rynku produktów mleczarskich [Angowski, Bujanowicz-Haraś 2019] lub wśród różnych rodzajów kaw [Walesiak, Gatnar (red.) 2012]. Istnieją też badania atrybutów zaufania dotyczących jakości produktów żywnościowych [Żakowska-Biemans, Gutkowska 2018]. Były analizowane determinanty zakupu żywności mieszkańców dużych miast [Sojkin, Bartkowiak 2017] oraz aktywność prosumpcyjna wśród mieszkańców miast i wsi [Iwanicka 2017].

Inny obszar dotyczy badania produkcji. Wykorzystując PCA, określano ekonomiczne determinanty produkcji mleka w gospodarstwach FADN dla regionów UE [Śmigła 2013, 2014; Guth, Stępień 2016; Guth 2017a, 2017b]; makroczynniki determinujące uzyskiwane efekty w uprawie różnych

odmian pszenicy bądź żyta [Pietrzykowski, Rakoczy-Trojanowska, Zieliński 1997; Sieczko i in. 2004; Ukalska i in. 2008; Laudański, Mańkowski, Flaszka 2012]; analizowano zróżnicowania systemów produkcji w gospodarstwach zajmujących się uprawą tytoniu [Hryniewski i in. 2011]; identyfikowano i klasyfikowano miody produkowane w Polsce [Jadamus-Hacura, Hacura 2009]. Wśród innych badań należy wymienić również te dotyczące konkurencyjności przedsiębiorstw. Oceniano na przykład wpływ barier innowacji na sprawność innowacyjną, a także międzynarodową konkurencyjność przedsiębiorstw polskiego przemysłu spożywczego [Lewandowska 2014] lub konsumencką percepcję nowych technologii stosowanych w produkcji żywności [Jeżewska-Zychowicz, Królak 2015].

Prowadzono również badania podstawowych aspektów produkcji rolniczej w poszczególnych województwach z wykorzystaniem PCA oraz FA [Ostasiewicz (red.) 1999] oraz badano czynniki decydujące o regionalnym zróżnicowaniu wykorzystania rolniczej przestrzeni produkcyjnej w Polsce [Krasowicz, Filipiak 1999]. W ostatniej z wspomnianych prac stwierdzono między innymi, że uwarunkowania przyrodniczo-agrotechniczne i organizacyjno-ekonomiczne objaśniają około 70% regionalnego zróżnicowania produkcji roślinnej. Inne badania z zakresu wyodrębniania regionów rolniczych wskazują na większą rolę dla zróżnicowania rolnictwa czynników ekonomiczno-organizacyjnych (struktura agrarna, rozproszenie produkcji, udział gospodarstw towarowych, poziom wiedzy fachowej) niż przyrodniczych. Badano też poziom rolniczej produkcji roślinnej w krajach UE [Strojny 2006].

W literaturze występują liczne przykłady zastosowań PCA i FA w badaniach regionalnych z zakresu rozwoju ekonomicznego obszarów wiejskich. Analizowano na przykład rozwój ekonomiczny obszarów wiejskich w krajach UE w latach 2000-2014 [Siudek, Vashchik 2014]. Najwyższy poziom rozwoju ekonomicznego obszarów wiejskich odnotowano w takich krajach, jak: Luksemburg, Holandia, Słowenia, a najniższy na Litwie, w Rumunii, Słowacji, Polsce i Czechach. Czyż metodami analizy czynnikowej badała ekonomiczną strukturę regionalną Polski [Czyż 1971]. Podobnie Malina wykorzystła FA do klasyfikacji i typologii regionów Polski ze względu na poziom rozwoju społeczno-gospodarczego [Malina 2006]. Badano również regionalną konkurencyjność rolnictwa w Polsce [Kołodziejczak, Kossowski 2014]. Określenia głównych determinant rozwoju obszarów wiejskich w podregionie pilskim przed akcesją i po niej Polski do UE podjęli się Czyżewski i Strońska-Ziemann [2016, 2017]. Identyfikacją barier rozwoju wsi i rolnictwa na poziomie regionalnym (region północno-zachodni Polski) zajęli się Beba i Kiryluk-Dryjska [2016]. Z kolei Bąk [2018] wykorzystał ideę głównych składowych do porządkowania liniowego powiatów ze względu na stan środowiska (na podstawie wartości pierwszej głównej składowej).

Ostatnią wyróżniającą się grupą badań z zastosowaniem PCA i FA w zakresie ekonomiki rolnictwa są badania dotyczące zagadnień finansowych. Można znaleźć przykłady prac, w których identyfikowano determinanty dochodów z działalności rolniczej [Kołoszko-Chomentowska 2003,

2004; Średzińska, Standar 2017]. W tym wypadku przeprowadzenie analizy czynnikowej wykazało, że najważniejszymi determinantami dochodów gospodarstw rolnych są zasoby czynników wytwórczych, dotacje do działalności operacyjnej, wielkość produkcji i intensywność wytwarzania. Zabolotnyy, Felczak i Wasilewski badali za pomocą PCA zależność między efektywnością wykorzystania zasobów wytwórczych a sytuacją finansową przedsiębiorstw rolniczych [Zabolotnyy, Felczak, Wasilewski 2018]. Siudek z kolei oceniał sytuację ekonomiczno-finansową banków spółdzielczych w Polsce (stosując FA spośród trzydziestu czterech wskaźników finansowych wyodrębnił siedem czynników głównych) [Siudek 2006]. Kobus i Pietrzykowski wykorzystali PCA do analizy wyników ekonomiczno-produkcyjnych gospodarstw rolnych w wybranych państwach UE [Kobus, Pietrzykowski 2008]. Inne interesujące zastosowanie PCA stanowi praca Zydronia i Szoszkiewicza, w której pod rozwagę poddano gotowość społeczeństwa do zapłacenia za to dobro, jakim jest środowisko naturalne [Zydroń, Szoszkiewicz 2013].

Podsumowując, należy stwierdzić, że metoda głównych składowych oraz analiza czynnikowa są coraz częściej stosowane w badaniach z zakresu ekonomiki rolnictwa. Wynika to z tego, że problemy spotykane w rzeczywistości gospodarczej rzadko są proste, a ich opis wymaga wnikliwego rozpoznania zależności między analizowanymi zmiennymi.

Tabela 8.2. Przykłady zastosowań analizy czynnikowej i metody składowych głównych w ekonomice rolnictwa

Autor	Problem	Obiekty	Metoda
Czyż 1971	badanie ekonomicznej struktury regionalnej Polski	powiaty	analiza czynnikowa
Pietrzykowski, Rakoczy-Trojanowska, Zieliński 1997	ocena zmienności somaklonalnej u żyta ozimego	somaklony	metoda głównych składowych
Krasowicz, Filipiak 1999	czynniki decydujące o regionalnym zróżnicowaniu wykorzystania rolniczej przestrzeni produkcyjnej w Polsce	regiony	analiza czynnikowa
Ostasiewicz (red.) 1999	cechy charakteryzujące produkcję rolniczą w województwach	województwa	metoda głównych składowych, analiza
Kołoszko-Chomentowska 2003, 2004	czynniki determinujące dochody z działalności rolniczej	gospodarstwa	metoda głównych składowych
Sieczko i in. 2004	charakterystyka zmienności genetycznej w kolekcji zasobów genowych pszenicy twardej	genotypy	metoda głównych składowych
Malina 2006	poziom rozwoju społeczno-gospodarczego regionów w Polsce	województwa	analiza czynnikowa
Siudek 2006	sytuacja ekonomiczno-finansowa banków spółdzielczych w Polsce	banki spółdzielcze	analiza czynnikowa
Stanimir 2006	ocena cech jogurtu owocowego	konsumenci	analiza czynnikowa z metodą głównych składowych
Strojny 2006	poziom rolniczej produkcji roślinnej krajów UE	państwa	metoda głównych składowych
Kobus, Pietrzykowski 2008	analiza wyników ekonomiczno-produkcyjnych gospodarstw rolnych w państwach UE	państwa	metoda głównych składowych
Ukalska i in. 2008	badanie zmienności i współzależności cech użytkowych w kolekcji roboczej pszenicy ozimej		metoda głównych składowych
Jadamus-Hacura, Hacura 2009	identyfikacja i klasyfikacja polskich miodów	miody	metoda głównych składowych, analiza
Hryniewski i in. 2011	ocena zróżnicowania systemów produkcji w gospodarstwach zajmujących się uprawą tytoniu	gospodarstwa	metoda głównych składowych
Krasowska, Banaszuk 2011	procesy kształtujące skład chemiczny wód cieku w krajobrazie rolniczym	próbki wody rzecznej	analiza czynnikowa
Laudański, Mańkowski, Flaszka 2012	czynniki produkcji determinujące uzyskiwane efekty w uprawie pszenicy ozimej	gospodarstwa	eksploracyjna analiza czynnikowa
Walesiak, Gatnar (red.) 2012	czynniki mające wpływ na wybór kawy przez konsumenta	rodzaje kawy	analiza czynnikowa
Zydroń, Szoszkiewicz 2013	wartość środowiska a gotowość społeczeństwa do zapłacenia za to dobro	respondenci ankiety	metoda głównych składowych

Kołodziejczak, Kossowski 2014	regionalna konkurencyjność rolnictwa	regiony	analiza czynnikowa z metodą głównych składowych
Lewandowska 2014	ocena wpływu barier innowacji na sprawność innowacyjną i międzynarodową konkurencyjność przedsiębiorstw przemysłu spożywczego	przedsiębiorstwa	analiza czynnikowa
Siudek, Vashchik 2014	określenie rozwoju ekonomicznego obszarów wiejskich w krajach UE	państwa	analiza czynnikowa, wskaźnik syntetyczny
Śmigła 2013, 2014	determinanty produkcji mleka w regionach UE	gospodarstwa	analiza czynnikowa
Jeżewska-Zychowicz, Królak 2015	postawy konsumentów wobec nowych technologii w produkcji żywności	konsumenci	analiza czynnikowa
Beba, Kiryluk-Dryjska 2016	identyfikacja barier rozwoju wsi i rolnictwa na poziomie regionalnym	gminy	analiza czynnikowa
Czyżewski, Strońska-Ziemann 2016, 2017	określenie determinant rozwoju obszarów wiejskich w podregionie pilskim	gminy	analiza czynnikowa
Guth, Stępień 2016	czynniki wpływające na produkcję mleka w gospodarstwach w makroregionach UE	gospodarstwa	analiza czynnikowa
Sroka 2016	ocena potencjału produkcyjnego rolnictwa w miastach i obszarach podmiejskich	gminy	analiza czynnikowa
Guth 2017a, 2017b	determinanty produkcji mleka w gospodarstwach	gospodarstwa	analiza czynnikowa
Iwanicka 2017	specyfika prosumpcji jako zmiany w zachowaniach rynkowych nabywców finalnych ze względu na miejsce ich zamieszkania (wieś/miasto)	nabywcy finalni	analiza czynnikowa
Sojkin, Bartkowiak 2017	analiza determinant zakupu żywności przez mieszkańców dużych miast	konsumenci	eksploracyjna analiza czynnikowa
Średzińska, Standar 2017	badanie determinant dochodów gospodarstw rolnych w krajach Europy Środkowo-Wschodniej	gospodarstwa rolne	analiza czynnikowa
Bąk 2018	porządkowanie liniowe powiatów ze względu na stan środowiska	powiaty	metoda głównych składowych
Feizabadi, Gorji 2018	analiza czynników wpływających na gospodarkę wodną w rolnictwie	eksperci	analiza czynnikowa
Zabolotnyy, Felczak, Wasilewski 2018	zależność między czynnikami efektywności a sytuacją finansową przedsiębiorstw rolniczych	przedsiębiorstwa rolnicze	metoda głównych składowych
Żakowska-Biemans, Gutkowska 2018	czynniki determinujące wybory konsumentów związane z żywnością	konsumenci	analiza czynnikowa z metodą głównych składowych
Angowski, Bujanowicz-Haraś 2019	identyfikacja czynników determinujące wybór ekologicznych produktów spożywczych na rynku mleczarskim	nabywcy żywności	analiza czynnikowa
Domagalska-Grędyś 2020	identyfikacja postaw wśród rolników utrzymujących zwierzęta ras zachowawczych	rolnicy	metoda głównych składowych

Źródło: opracowanie własne

Literatura

- Aczel A.D. (2005) Statystyka w zarządzaniu. Wydawnictwo Naukowe PWN, Warszawa.
- Angowski M., Bujanowicz-Haraś B. (2019) Consumers on Organic Food Market – Factors Determining the Choice of Dairy Products. *Acta Scientiarum Polonorum, Oeconomia*, 18(2), 5-12.
- Bąk A. (2018) Zastosowanie metod wielowymiarowej analizy porównawczej do oceny stanu środowiska w województwie dolnośląskim. *Wiadomości Statystyczne*, 1(680), 7-20.
- Beba P., Kiryluk-Dryjska E. (2016) Identyfikacja barier rozwoju wsi i rolnictwa w Polsce na przykładzie regionu północno-zachodniego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 433, 29-39.
- Czyż T. (1971) Zastosowanie metody analizy czynnikowej do badania ekonomicznej struktury regionalnej Polski. Instytut Geografii Polskiej Akademii Nauk, *Prace Geograficzne*, 92. Zakład Narodowy im. Ossolińskich – Wydawnictwo PAN, Warszawa.
- Czyżewski A., Strońska-Ziemann J. (2016) Obszary wiejskie w podregionie piłskim przed i po akcesji polski do UE. Wydawnictwo Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy, Bydgoszcz.
- Czyżewski A., Strońska-Ziemann J. (2017) Factors of Changes in Farming in Piła Subregion in the View of Agricultural Censuses in 1996, 2002 and 2010. *Management*, 21(1), 259-270.
- Domagalska-Grędyś M. (2020) Evaluation of Dimensions of Farmer Attitudes in Principal Component Analysis (PCA). *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 22(1), 66-76.
- Feizabadi Y., Gorji E.M. (2018) Analysis of Effective Factors on Agricultural Water Management in Iran. *Journal of Water and Land Development*, 38, 35–41.
- Gatnar E. (1998) Analiza czynnikowa. Teoria i zastosowanie. Akademia Ekonomiczna w Katowicach, Katowice [maszynopis powielony].
- Gatnar E., Walesiak M. (red.) (2004) Metody statystycznej analizy wielowymiarowej w badaniach marketingowych. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Guth M. (2017a) Determinants of Milk Production Diversity in the Macroregions of the European Union. *Acta Scientiarum Polonorum, Oeconomia*, 16(1), 33-42.
- Guth M. (2017b) Diversity of Milk Production Determinants in EU Macro-Regions with a Predominance of Intensive and Extensive Production in 2011. *Journal of Agribusiness and Rural Development*, 44(2), 329-336.
- Guth M., Stępień S. (2016) Zróżnicowanie gospodarstw mlecznych FADN w makroregionach Unii Europejskiej względem determinant produkcji mleka. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 18(4), 86-91.
- Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1998) *Multivariate Data Analysis*. Prentice Hall, Englewood Cliffs.
- Hotelling H. (1933) Analysis of a Complex of Statistical Variables into Principal Components. Warwick & York, Baltimore.
- Hryniewski R., Gozdowski D., Mądry W., Roszkowska-Mądra B. (2011) Typology of Tobacco-Based Farming Systems at the Farm Level in South-Eastern Poland. *Acta Scientiarum Polonorum, Oeconomia*, 10(1), 21-33.
- Iwanicka A. (2017) Specyfika prosumpcji jako zmiany w zachowaniach rynkowych nabywców finalnych ze względu na miejsce ich zamieszkania. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 117, 107-121.
- Jadamus-Hacura M., Hacura A. (2009) Wykorzystanie metod wielowymiarowej analizy danych spektralnych do identyfikacji i klasyfikacji polskich miódów. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia*, 16(47), 469-476.
- Jeżewska-Zychowicz M., Królak M. (2015) Konsumentcka percepcja nowych technologii stosowanych w produkcji żywności. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 111, 71-80.
- Keiser H.F. (1974) An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kim J.O., Mueller C.W. (1978) *Factor Analysis. Statistical Methods and Practical Issues*. Sage Publishing, Beverly Hills.

- Kisielińska J., Stańko S. (2009) Wielowymiarowa analiza danych w ekonomice rolnictwa. *Roczniki Nauk Rolniczych, Seria G*, 96(2), 63-76.
- Kobus P., Pietrzykowski R. (2008) Wielowymiarowe metody statystyczne w analizie wyników ekonomiczno-produkcyjnych gospodarstw rolnych wybranych państw Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 4(19), 371-378.
- Kołodziejczak A., Kossowski T. (2014) Regional Competitiveness of Agriculture in Poland. *Więś i Rolnictwo*, 3(164), 57-70.
- Kołoszko-Chomentowska Z. (2003) Zastosowanie składowych głównych w badaniach nad czynnikami determinującymi dochody z działalności rolniczej. *Roczniki Nauk Rolniczych, Seria G*, 90(2), 34-44.
- Kołoszko-Chomentowska Z. (2004) Wykorzystanie analizy składowych głównych w badaniach ekonomiczno-rolniczych. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu, Seria G*, 6(5), 55-59.
- Krasowicz S., Filipiak K. (1999) Czynniki decydujące o regionalnym zróżnicowaniu wykorzystania rolniczej przestrzeni produkcyjnej w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 1(1), 153-158.
- Krasowska M., Banaszuk P. (2011) Zastosowanie analizy czynnikowej w badaniach hydrochemicznych w zlewni niewielkiego cieku w krajobrazie rolniczym. *Inżynieria Ekologiczna*, 26, 147-155.
- Landmesser J. (2010) A Dynamic Approach to the Study of Unemployment Duration. *Metody Ilościowe w Badaniach Ekonomicznych*, 11(1), 212-222.
- Laudański Z., Mańkowski D.R., Flaszka M. (2012) Eksploracyjna analiza czynnikowa w badaniach struktury zespołu zmiennych obserwowanych. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 263, 75-89.
- Lewandowska M. (2014) Innovation Barriers and International Competitiveness of Enterprises from Polish Food Processing Industry: Research Results. *Acta Scientiarum Polonorum, Oeconomia*, 13(4), 103-113.
- Malina A. (2006) Analiza czynnikowa jako metoda klasyfikacji regionów w Polsce. *Przegląd Statystyczny*, 53(1), 33-48.
- Mulaik S.A. (1972) *The Foundations of Factor Analysis*. McGraw Hill, New York.
- Ostasiewicz W. (red.) (1999) *Statystyczne metody analizy danych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Pietrzykowski R., Rakoczy-Trojanowska M., Zieliński W. (1997) Wykorzystanie wielowymiarowych metod statystycznych do oceny zmienności somaklonalnej u żyta ozimego *Secale cereale* L., [w:] *Hodowla Roślin. Materiały I Krajowej Konferencji*, Poznań, 19-20 listopada, 1, 171-174.
- Pluta W. (1986) *Wielorównaniowa analiza porównawcza w modelowaniu ekonometrycznym*. PWN, Warszawa.
- Rószkiewicz M. (1998) *Zarys metod Statystyki wielowymiarowej z wykorzystaniem programów komputerowych Statgraphics wersja 6 oraz SPSS wersja 5*. Oficyna Wydawnicza SGH, Warszawa.
- Rószkiewicz M. (2002a) *Metody ilościowe w badaniach marketingowych*. Wydawnictwo Naukowe PWN, Warszawa.
- Rószkiewicz M. (2002b) *Narzędzia statystyczne w analizach marketingowych*. Wydawnictwo C.H. Beck, Warszawa.
- Sieczko L., Mądry W., Zieliński A., Paderewski J., Urbaś-Szwed K. (2004) Zastosowanie analizy składowych głównych w badaniach nad wielo cechową charakterystyką zmienności genetycznej w kolekcji zasobów genowych pszenicy twardej (*Triticum durum* L.). *Colloquium Biometryczne*, 34, 223-239.
- Siudek T. (2006) Ocena sytuacji ekonomiczno-finansowej banków spółdzielczych w Polsce przy wykorzystaniu wskaźnika syntetycznego. *Roczniki Nauk Rolniczych, Seria G*, 92(2), 7-15.
- Siudek T., Vashchik M. (2014) Economic Development of Rural Areas in European Union Member States in 2000-2012. *Acta Scientiarum Polonorum, Oeconomia*, 13(3), 95-104.
- Sojkin B., Bartkowiak P. (2017) Determinanty zakupu żywności mieszkańców dużych miast w Polsce. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 119, 155-164.
- Spearman Ch. (1904) General Intelligence, Objectively Determined and Measured. *American Journal of Psychology*, 15, 201-293.
- Sroka W. (2016) Potencjał produkcyjny rolnictwa w miastach i obszarach podmiejskich. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 18(2), 249-255.

- Stanimir A. (2006) Analiza danych marketingowych. Problemy, metody, przykłady. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Strojny J. (2006) Poziom rolniczej produkcji roślinnej krajów UE. *Więś i Rolnictwo*, 4(133), 103-125.
- Sztemberg-Lewandowska M. (2008) Analiza czynnikowa w badaniach marketingowych. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Śmigła M. (2013) Determinanty rozwoju gospodarstw mleczarskich w regionach Unii Europejskiej po 2003 roku w świetle analizy czynnikowej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 15(3), 344-349.
- Śmigła M. (2014) Determinanty produkcji mleka w regionach Unii Europejskiej o bardzo dużych gospodarstwach mlecznych po 2004 roku. *Journal of Agribusiness and Rural Development*, 1(31), 143-150.
- Średzińska J., Standar A. (2017) Wykorzystanie analizy czynnikowej do badania determinant dochodów gospodarstw rolnych (na przykładzie krajów Europy Środkowo-Wschodniej). *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 118, 5-17.
- Tarka P. (2013) Geometrical Perspective on Rotation and Data Structure Diagnosis in Factor Analysis. *Econometrics*, 1(39), 198-209.
- Thurstone L.L. (1935) *The Vectors of Mind*. University of Chicago Press, Chicago.
- Ukalska J., Ukalski K., Śmiałowski T., Mądry W. (2008) Badanie zmienności i współzależności cech użytkowych w kolekcji roboczej pszenicy ozimej (*Triticum aestivum* L.) za pomocą metod wielowymiarowych. Cz. II. Analiza składowych głównych na podstawie macierzy korelacji fenotypowych i genotypowych. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 249, 45-57.
- Walesiak M., Bąk A. (1997) Wykorzystanie analizy czynnikowej w badaniach marketingowych. *Badania Operacyjne i Decyzje*, 1, 75-87.
- Walesiak M., Gatnar E. (red.) (2012) *Statystyczna analiza danych z wykorzystaniem program R*. Wydawnictwo Naukowe PWN, Warszawa.
- Zabolotny S., Felczak T., Wasilewski M. (2018) Relations between input efficiency and financial situation of agricultural companies. *Zagadnienia Ekonomiki Rolnej*, 2(355), 41-60.
- Zeliaś A. (2000) *Metody statystyczne*. Wydawnictwo Naukowe PWN, Warszawa.
- Zydroń A., Szoszkiewicz K. (2013) Wartość środowiska a gotowość społeczeństwa do zapłacenia za to dobro. *Annual Set The Environment Protection. Rocznik Ochrona Środowiska*, 15, 2874-2886.
- Żakowska-Biemans S., Gutkowska K. (2018) Atrybuty zaufania dotyczące jakości jako czynnik wpływający na decyzje nabywcze konsumentów żywności. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 121, 115-130.

9. Analiza struktury

Problematyka badań struktury jest od dawna obecna w wielu dziedzinach nauki. Słowo struktura należy do najczęściej używanych terminów w wielu badaniach, opracowaniach, podręcznikach, artykułach i monografiach [Bogocz i in. 2010]. Pojęcie to wprowadził do teorii ekonomii w pierwszej połowie XX wieku E. Wagemann, dając początek badaniom strukturalnym na większą skalę, których rozwój przypadł na lata 50. i 60. XX wieku [Walesiak 1984].

Badania strukturalne, w tym porównania struktur oraz określenie stopnia zmian struktur w czasie, odgrywają istotną rolę w naukach ekonomicznych, w tym w ekonomice rolnictwa. Mimo iż termin struktura ma charakter interdyscyplinarny, w odniesieniu do sfery ekonomicznej używa się raczej pojęć struktury gospodarczej bądź struktury ekonomicznej. Terminy te często traktowane są czysto opisowo, ale należy wskazać istnienie ścisłych, matematycznych ich definicji [Bogocz i in. 2010]. Struktura może być traktowana jako specyficzny obiekt złożony, charakteryzowany nie zestawami wartościami cech, lecz ciągiem pewnych wskaźników, zwanych wskaźnikami struktury (lub udziałami). Można więc – w pewnym sensie – wskaźniki struktury traktować jako cechy, a liczbę składników struktury jako liczbę wymiarów przestrzeni klasyfikacji cech [Strahl 1998; Malina 2004].

W dosłownym rozumieniu struktura oznacza układ i wzajemne relacje elementów stanowiących całość, czyli budowę tejże całości. Porównania struktur można dokonywać w aspekcie zarówno przestrzennym, porównując struktury dwóch lub więcej obiektów, jak i czasowym, w którym bada się strukturę zjawiska w dwóch lub więcej okresach. Pojęcie struktury nie jest jednoznacznie rozumiane – można wyróżnić szerokie oraz wąskie jego rozumienie. Według Urbańczyka i Jurka struktura w ujęciu szerokim oznacza dynamiczny system celowych przemian w czasie i przestrzeni, a w ujęciu wąskim badanie struktury sprowadza się do badania rozkładów zmiennych losowych [Urbańczyk, Jurek 1998]. Jedną z propozycji określenia struktury ekonomicznej jest definicja Marciniaka, według której struktura ekonomiczna to całokształt wzajemnie powiązanych stosunków między poszczególnymi elementami danej gospodarki, przy czym struktura wyraża stosunek poszczególnych części do całości oraz stosunek tych części wzajemnie do siebie [Marciniak 1970]. Poza takim podejściem pojęcie struktury używane jest również do określania pewnych pojęć ekonomicznych, w tym pojęć ekonomiki rolnictwa, takich jak: struktura kapitału, struktura majątku, struktura spożycia, struktura produkcji, struktura zasiewów itd.

9.1. Opis metod

Struktura może być traktowana jako wektor w przestrzeni wielowymiarowej [Kukuła 1975, 1986; Stawicki 2004]. Wyróżnia się przy tym dwa podejścia. W pierwszym podejściu strukturę liczebności pewnej zbiorowości traktuje się jako wektor, którego składowe reprezentują liczbę jednostek tejże zbiorowości, należących do poszczególnych klas (stanów) utworzonych w wyniku jej podziału według określonego kryterium. W tym podejściu suma składowych rozważanego wektora daje ogólną liczebność zbiorowości. W drugim podejściu autor rozpatruje strukturę udziałów w pewnej zbiorowości, rozumianą jako wektor, którego składowe przedstawiają udział liczebności poszczególnych klas (stanów), utworzonych w wyniku podziału rozważanej zbiorowości według określonego kryterium, w ogólnej liczebności tejże zbiorowości. W tym przypadku suma składowych wektora struktury równa jest jeden. Oznacza to, iż pierwszemu ujęciu pojęcia struktury przypisuje się wektory liczebności w danej zbiorowości, natomiast drugiemu podejściu odpowiada tradycyjne rozumienie wektorów struktury, w którym poszczególne wskaźniki (składniki) struktury w danej zbiorowości tworzą odpowiedni wektor kolumnowy. Drugie podejście może stanowić punkt wyjścia we wszelkich analizach strukturalnych. Przy tym szczególnego znaczenia nabierają porównania struktur oraz określenie stopnia zmian struktur w czasie.

Porównania struktur mogą być dokonywane w aspekcie:

- 1) przestrzennym, w którym porównuje się struktury dwóch lub więcej zbiorowości (obiektów, np.: krajów, województw, gmin, przedsiębiorstw, gospodarstw),
- 2) czasowym, w którym bada się strukturę zjawiska w dwóch lub więcej okresach.

Strukturę traktuje się jako kolumnowy wektor udziałów, którego składowe oznaczają udziały (wskaźniki struktury) badanego zjawiska w danej zbiorowości. Przy tym zbiorowość może być badana w okresie bądź momencie czasowym – mowa jest wówczas o dynamice struktur. W proponowanym podejściu m składników struktury rozpatrywanych dla k zbiorowości tworzy wektor:

$$U_k^T = [w_{1k}, w_{2k}, \dots, w_{mk}], \quad (9.1)$$

gdzie w_{ik} jest udziałem i -tego składnika struktury w strukturze ogółem, rozpatrywanej dla zbiorowości k . Zachodzą przy tym relacje: $0 \leq w_{ik} \leq 1$ oraz $\sum_{i=1}^m w_{ik} = 1$ oznaczające, iż wszystkie wskaźniki struktury są liczbami z przedziału $[0,1]$ oraz suma wskaźników struktury dla danej zbiorowości k równa jest 1.

Rozpatrując n zbiorowości, gdzie w każdej z nich wyodrębniono m składników struktury, kompletne dane o składnikach struktury tworzą macierz $\mathbf{W} = [w_{ik}]$ o wymiarach $m \times n$, gdzie $i = 1, \dots, m$, a $k = 1, \dots, n$:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix}. \quad (9.2)$$

Porównania dwóch struktur można dokonać poprzez obliczenie wartości odpowiedniej miary zgodności charakteryzującej ich stopień podobieństwa lub zróżnicowania. Szeroki wachlarz stosowanych miar przedstawiony jest w pracach Kukuły [1996] oraz Bogocz i in. [2010]. W niniejszej pracy przedstawione zostały wybrane z nich.

Jedną z najczęściej stosowanych miar podobieństwa dwóch struktur jest miara oparta na wskaźniku podobieństwa zbiorów Renkonena. Jeśli rozpatrujemy dwie zbiorowości (obiektów, okresów, momentów) k i q , przy czym w każdej z nich wyodrębniono m wskaźników struktury, tę miarę można przedstawić następująco:

$$p_{kq} = \sum_{i=1}^m \min(w_{ik}, w_{iq}), \quad (9.3)$$

gdzie p_{kq} jest miarą podobieństwa struktur dwóch zbiorowości k oraz q , przy czym: $k, q = 1, 2, \dots, n$, w_{ik} jest udziałem i -tego składnika w strukturze zbiorowości k , w_{iq} jest udziałem i -tego składnika w strukturze zbiorowości q , gdzie $i = 1, 2, \dots, m$. Zachodzi ponadto: $p_{kq} = p_{qk}$ oraz $p_{kk} = 1$.

Miara p_{kq} podobieństwa struktur dwóch zbiorowości (obiektów, okresów, momentów czasowych) charakteryzuje się następującymi własnościami:

- Jest miarą unormowaną, czyli przyjmuje wartości z przedziału $[0, 1]$.
- Przyjmuje wartość 0 w przypadku całkowitego braku podobieństwa porównywanych struktur.
- Przyjmuje wartość 1 w przypadku całkowitej zgodności badanych struktur.
- Przyjmuje wartości tym większe, im bardziej podobne są do siebie porównywane struktury.
- Przyjmuje wartości tym mniejsze, im bardziej różnią się od siebie porównywane struktury.

Na podstawie miary (9.3) podobieństwa struktur można zaproponować miarę zróżnicowania struktur następującej postaci [Chomątowski, Sokołowski 1978]:

$$q_{kq} = 1 - p_{kq} = 1 - \sum_{i=1}^m \min(w_{ik}, w_{iq}). \quad (9.4)$$

Wartości tej miary, podobnie jak poprzedniej, mieszczą się w przedziale $[0; 1]$. Jednak w tym przypadku, zgodnie z postulatem, jaki stawia się miarom zróżnicowania, wartość równa 0 wskazuje na brak zróżnicowania porównywanych struktur (struktury są identyczne), natomiast wraz ze zbliżaniem się wartości miary do 1, zróżnicowanie to wzrasta.

Miara podobieństwa struktur dwóch zbiorowości oparta na odległości miejskiej jest następująca [Bogocz i in. 2010]:

$$V_{k/q} = \sum_{i=1}^m |w_{ik} - w_{iq}|. \quad (9.5)$$

Miara ta przyjmuje wartości z przedziału $[0; 2]$. Zero wskazuje, że porównywane struktury są identyczne. Im wartość miary jest bliższa liczbie 2, tym zróżnicowanie struktur jest większe.

Jako miarę przeciętnego zróżnicowania wskaźników struktury dwóch zbiorowości k i q można przyjąć miarę postaci [Kukuła 1975, 1986]:

$$d_{k/q} = \frac{\sum_{i=1}^m |w_{ik} - w_{iq}|}{m}. \quad (9.6)$$

Wartości miary $d_{k/q}$ mieszczą się w przedziale $[0; 2/m]$. Gdy $d_{k/q}$ przyjmuje wartości bliskie 0, stwierdza się, że przeciętne zróżnicowanie wskaźników struktury zbiorowości k i q jest niewielkie, natomiast im większe (bliższe $2/m$) są wartości $d_{k/q}$, tym większe jest przeciętne zróżnicowanie porównywanych struktur.

Według miary (9.6) można skonstruować miarę o wartościach z przedziału $[0, 1]$. Uzyskuje się ją, dzieląc miarę $d_{k/q}$ przez liczbę $2/m$ [Kukuła 1975, 1986]:

$$d_{k/q}^* = \frac{\sum_{i=1}^m |w_{ik} - w_{iq}|}{2}. \quad (9.7)$$

Ciekawą koncepcyjnie jest miara bazująca na kącie rozwarcia między wektorami struktury. Zróżnicowanie struktur dwóch zbiorowości (obiektów, okresów lub momentów czasowych) jest tym większe, im bardziej dwa wektory reprezentujące te struktury różnią się od siebie. Można więc powiedzieć, iż dwa wektory są tym bardziej podobne, im bardziej zbliżone są składowe obu wektorów [Rutkowski 1981; Żwirbła 2006]. Tak więc im większe są różnice w składowych wektorów, tym bardziej różnią się porównywane struktury.

Jeśli wektory \mathbf{U}_k oraz \mathbf{U}_q charakteryzujące struktury w dwóch zbiorowościach k i q (objektach, okresach lub momentach czasowych) tworzą pewien kąt α , można w prosty sposób określić wartość funkcji sinus bądź cosinus tego kąta. Za miarę natężenia zmian struktury można przyjąć wartość funkcji sinus. Wybór taki ma przejrzystą interpretację, gdyż duży kąt rozwarcia α między wektorami \mathbf{U}_k oraz \mathbf{U}_q (wartość funkcji sinus jest bliższa 1) świadczy o znacznych różnicach w strukturze zbiorowości k w porównaniu ze strukturą zbiorowości q . I na odwrót – im mniejszy kąt rozwarcia α między tymi wektorami (wartość funkcji sinus jest bliższa 0), tym mniejsze są różnice w strukturze porównywanych zbiorowości. W szczególnym przypadku, gdy porównywane struktury są identyczne, kąt między wektorami struktury wynosi 0° i wówczas $\sin \alpha = 0$. Maksymalny kąt rozwarcia α , jaki mogą tworzyć dwa wektory struktury, wynosi 90° (ponieważ wszystkie współrzędne obydwu wektorów są dodatnie). Tak więc wartość funkcji $\sin \alpha$ jest liczbą z przedziału $[0; 1]$,

co odpowiada relacji spełnianej przez składniki struktury, czyli $0 \leq w_{ik} \leq 1$. Ponadto funkcja sinus jest w przedziale $[0^\circ; 90^\circ]$ rosnąca, stąd też uzasadniony jest wybór tej właśnie funkcji trygonometrycznej jako miary natężenia zróżnicowania struktur porównywanych zbiorowości – wraz ze wzrostem wartości funkcji $\sin \alpha$ wzrasta zróżnicowanie porównywanych zbiorowości.

Miarę natężenia zmian struktury jako wartość funkcji sinus kąta rozwarcia między dwoma wektorami struktury można więc zapisać następująco [Kukuła 1975, 1986]:

$$s_{k/q} = \sin \alpha = \sqrt{1 - \frac{(\sum_{i=1}^m w_{ik} w_{iq})^2}{\sum_{i=1}^m w_{ik}^2 \cdot \sum_{i=1}^m w_{iq}^2}}. \quad (9.8)$$

Należy zauważyć, że chcąc ustalić umowne przedziały, które określałyby bardzo małe, małe, umiarkowane, duże i bardzo duże zróżnicowanie struktur, to można za końce takich przedziałów przyjąć wartości funkcji sinus odpowiadające podziałowi kąta 90° na pięć równych części. Gdy $s_{k/q} \in [0; 0,3]$, można powiedzieć o bardzo małym zróżnicowaniu porównywanych struktur (tj. obie struktury są bardzo zbliżone), gdy $s_{k/q} \in [0,3; 0,6)$ o małym zróżnicowaniu, gdy $s_{k/q} \in [0,6; 0,8)$ o umiarkowanym, gdy $s_{k/q} \in [0,8; 0,95)$ o dużym, a gdy $s_{k/q} \in [0,95; 1]$ o bardzo dużym zróżnicowaniu porównywanych struktur (tj. podobieństwo obu struktur jest nieznaczące) [Wasilewska 2007]. Należy jednak przy tym zaznaczyć, iż z uwagi na to, że funkcja sinus nie jest liniowa względem kąta, należy ostrożnie interpretować wielkości zmian struktury w ujęciu procentowym. Posługiwanie się pojęciem zmian procentowych traktuje się jako informację przybliżoną, przy czym należy mieć na uwadze, że taka interpretacja powiększa faktyczny stopień zmian [Bogocz i in. 2010].

Jeśli rozpatrywany jest uporządkowany ciąg n zbiorowości, wówczas stosując wcześniej przedstawioną formułę (9.8), można wprowadzić jednopodstawowe i łańcuchowe miary zmian struktury. Strukturę dowolnej k -tej zbiorowości można porównywać ze strukturą zbiorowości bazowej, oznaczonej jako 0. Otrzymuje się wówczas jednopodstawowe miary zmian struktury. Jeśli k -tej zbiorowości porównuje się ze strukturą zbiorowości $k - 1$, otrzymuje się miary łańcuchowe. Miary odpowiednio jednopodstawowe i łańcuchowe wyznacza się z następujących formuł:

$$s_{k/0} = \sqrt{1 - \frac{(\sum_{i=1}^m w_{ik} w_{i0})^2}{\sum_{i=1}^m w_{ik}^2 \cdot \sum_{i=1}^m w_{i0}^2}}, \quad (9.9)$$

$$s_{k/k-1} = \sqrt{1 - \frac{(\sum_{i=1}^m w_{ik} w_{i,k-1})^2}{\sum_{i=1}^m w_{ik}^2 \cdot \sum_{i=1}^m w_{i,k-1}^2}}. \quad (9.10)$$

Miara jednopodstawowa $s_{k/0}$ pokazuje natężenie zmian struktury badanego zjawiska k -tej zbiorowości w stosunku do zbiorowości bazowej, a miara łańcuchowa $s_{k/k-1}$ obrazuje natężenie tych zmian w zbiorowości k w stosunku do zbiorowości $k - 1$. Interpretacja tych wielkości jest analogicz-

na jak $s_{k/q}$ (wzór 9.8). Wartości bliskie 0 świadczą o nieznacznych różnicach struktury w porównywanych zbiorowościach.

Miary jednopodstawowe i łańcuchowe przeciętnego zróżnicowania wskaźników struktury w porównywanych zbiorowościach można przedstawić następująco:

$$d_{k/0} = \frac{\sum_{i=1}^m |w_{ik} - w_{i0}|}{m}, \quad (9.11)$$

$$d_{k/k-1} = \frac{\sum_{i=1}^m |w_{ik} - w_{i,k-1}|}{m}. \quad (9.12)$$

Miara jednopodstawowa $d_{k/0}$ wyraża przeciętne zmiany wskaźników struktury w zbiorowości k w porównaniu ze zbiorowością bazową, a łańcuchowy $d_{k/k-1}$ daje obraz tych zmian w zbiorowościach k i $k - 1$. Wartości tych miar bliskie 0 świadczą o znikomym przeciętnym zróżnicowaniu wskaźników struktury porównywanych zbiorowości.

W celu kompleksowego opisu zmian struktury dla wszystkich zbiorowości łącznie można znaleźć średnią wartość łańcuchowych miar $s_{k/k-1}$ oraz $d_{k/k-1}$. Otrzymuje się w ten sposób miary średnie o następującej postaci:

$$\bar{s} = \frac{\sum_{k=2}^n s_{k/k-1}}{n-1}, \quad (9.13)$$

$$\bar{d} = \frac{\sum_{k=2}^n d_{k/k-1}}{n-1}. \quad (9.14)$$

9.1.1. Analiza koncentracji

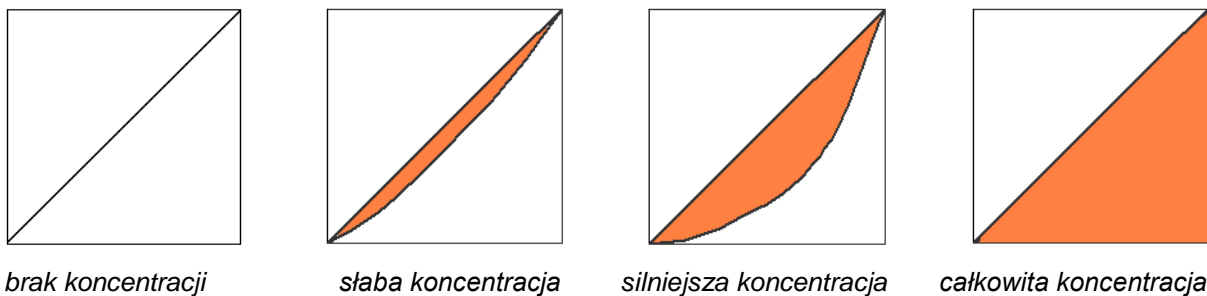
Jednym z częściej pojawiających się problemów badawczych w analizach strukturalnych jest ocena koncentracji. Wykorzystując dane strukturalne, można oceniać poziom koncentracji „dobra” w posiadaniu jednostek (np. gospodarstw), a także równomierność rozmieszczenia podmiotów partycypujących w procesie gospodarczym realizowanym na danym terytorium podzielonym na mniejsze jednostki przestrzenne [Bogocz i in. 2010].

Rozróżnia się dwa rodzaje koncentracji: koncentrację rozumianą jako nierównomierny podział zjawiska w zbiorowości oraz koncentrację zbiorowości wokół średniej (koncentrację wartości cechy wokół średniej), tzw. kurtozę. Koncentracja w pierwszym znaczeniu zachodzi wówczas, gdy występuje nierównomierny rozdział ogólnej sumy wartości cechy (łącznego funduszu cechy), na przykład dochodu, produkcji czy majątku, między poszczególne jednostki zbiorowości, na przykład przedsiębiorstwa, gospodarstwa domowe lub indywidualne osoby [Sobczyk 2007]. W praktyce pojęcie koncentracji ograniczone jest dwoma skrajnymi przypadkami – koncentracją całkowitą (zupelną) oraz brakiem koncentracji. Koncentracja całkowita występuje wówczas, gdy jedna jednostka zbiorowości skupia łączny fundusz cechy (tzw. rozkład elitarny), na przykład jedno gospodarstwo rolne dysponuje całością arealu użytków rolnych w regionie. Z brakiem koncentracji mamy do czynienia, gdy na każdą jednostkę zbiorowości przypada taka sama część łącznego funduszu cechy

(tzw. rozkład egalitarny) – na jedno gospodarstwo przypada taka sama część łącznego arealu użytków rolnych.

Pomiar koncentracji może być dokonany graficznie – za pomocą wieloboku koncentracji Lorenza (krzywej koncentracji Lorenza) lub analitycznie. Metoda analityczna sprowadza się do wyznaczenia odpowiedniego współczynnika, z których najbardziej popularnym jest współczynnik Giniego, nazywany również indeksem Giniego lub współczynnikiem Lorenza. Współczynnik ten został opracowany w XX wieku przez włoskiego statystyka C. Giniego do pomiaru nierówności rozkładu wysokości dochodów w społeczeństwie.

W celu sporządzenia krzywej koncentracji Lorenza w układzie współrzędnych na osi odciętych odmierza się skumulowane częstości względne liczebności (%), a na osi rzędnych – skumulowane częstości względne łącznej sumy wartości cechy (%). Łącząc otrzymane punkty, uzyskuje się krzywą koncentracji. W przypadku rozkładu egalitarnego wszystkie punkty byłyby położone na przekątnej kwadratu, stąd przekątna kwadratu nazywa jest linią równomiernego podziału. Powierzchnia zawarta między przekątną kwadratu a krzywą Lorenza to tzw. powierzchnia Lorenza. Im większa jest powierzchnia Lorenza, tym silniejsza jest koncentracja (rys. 9.1).



brak koncentracji

słaba koncentracja

silniejsza koncentracja

całkowita koncentracja

Rysunek 9.1. Ilustracja różnej siły koncentracji

Źródło: Wasilewska 2011

Współczynnik koncentracji Giniego jest stosunkiem pola powierzchni koncentracji a i pola powierzchni trójkąta będącego obrazem całkowitej koncentracji (por. rys. 9.1):

$$G = \frac{a}{0,5} \text{ lub } G = \frac{a}{5000}, \quad (9.15)$$

gdzie a to pole powierzchni koncentracji. Drugi z tych wzorów stosowany jest wówczas, gdy częstości względne wyrażone są w procentach. Pole powierzchni koncentracji można obliczyć jako różnicę pola trójkąta (będącego obrazem całkowitej koncentracji) i pola pod krzywą Lorenza, a to pole z kolei jako sumę pól trapezów i trójkąta, na jakie można podzielić pole pod krzywą. Pole to można wówczas zapisać jako:

$$G = 1 - \sum_{i=1}^k (z_i^{sk} + z_{i-1}^{sk}) w_i \text{ lub } G = 1 - \frac{\sum_{i=1}^k (z_i^{sk} \cdot 100 + z_{i-1}^{sk} \cdot 100) w_i \cdot 100}{10\ 000}, \quad (9.16)$$

gdzie w_i jest wskaźnikiem struktury liczebności dla i -tego przedziału, z_i^{sk} skumulowanym wskaźnikiem struktury ogólnej sumy wartości cechy dla i -tego przedziału, a k liczbą przedziałów (klas) w szeregu rozdzielczym.

Współczynnik Giniego przyjmuje wartości z przedziału $[0, 1]$. Im jego wartość jest bliższa 1, tym koncentracja jest większa. W skrajnych przypadkach przyjmuje wartość 0 – przy braku koncentracji, natomiast gdy koncentracja jest całkowita – wartość 1. Można przyjąć umowne przedziały pozwalające ocenić siłę koncentracji: 0-0,3 – słaba koncentracja; 0,3-0,6 – umiarkowana koncentracja; 0,6-1 – silna koncentracja. Do oceny rozwarstwienia dochodowego społeczeństwa zazwyczaj przyjmuje się przedziały: 0-0,3 (niskie rozwarstwienie dochodowe); 0,3-0,45 (średnie dysproporcje dochodów); 0,45-1 (wysokie rozwarstwienie dochodowe) [Panek 2011].

W praktyce współczynnik Giniego może być obliczany z wielu formuł [Panek 2011], przykładowo:

$$G = \frac{1}{2n^2\bar{x}} \sum_{i>j} \sum_j |x_i - x_j|, \quad (9.17)$$

gdzie x_i to wartość cechy (np. dochód) dla i -tej jednostki, x_j wartość cechy (np. dochód) dla j -tej jednostki, n liczba obserwacji, a \bar{x} średnia wartość cechy.

Wzór (9.17) wskazuje, że współczynnik Giniego przyjmuje wartość równą stosunkowi połowy średniej absolutnej różnicy wartości cechy wszystkich jednostek do średniej wartości cechy.

Alternatywnie współczynnik Giniego może być wyznaczany następująco:

$$G = \frac{\sum_{i=1}^n (2i-n-1)x_i}{n^2\bar{x}}. \quad (9.18)$$

Badanie koncentracji może być przeprowadzone przy wykorzystaniu wielu innych miar, między innymi: współczynników koncentracji przestrzennej według koncepcji Kukuły [Bogocz i in. 2010], miary wykorzystującej pojęcie entropii w teorii informacji [Panek 2011] czy tzw. miary ar będącej uogólnieniem współczynnika Giniego, opracowanej przez zespół autorski [Binderman, Koszela, Szczesny 2014; Binderman, Szczesny 2016].

9.2. Zastosowania metod

W tabeli 9.1 przedstawiono zastosowania metod analizy struktur w badaniach z zakresu ekonomiki rolnictwa. Szeroka problematyka poruszana w pracach cytowanych autorów wskazuje na znaczenie tego typu badań. Binderman i Szczesny wskazują na konieczność śledzenia zmian nie tylko w strukturze agrarnej, ale też zmian strukturalnych zachodzących w rozmaitych obszarach (np.: płace, koszty działania, portfele produktów i klientów), podkreślając, że badanie struktur różnego rodzaju zjawisk społecznych i gospodarczych jest ważnym elementem służącym ocenie prawidłowości rozwoju społeczno-ekonomicznego [Binderman, Szczesny 2016].

Dominująca liczba prac skupia się wokół problematyki badania koncentracji zjawisk (trzydzieści pięć prac wśród sześćdziesięciu). Głównym obszarem zastosowań analizy koncentracji są badania w zakresie struktury obszarowej gospodarstw [Binderman, Koszela, Szczesny 2014; Binderman, Szczesny 2016; Majchrzak 2018]. Analizowano też wpływ koncentracji ziemi w rolnictwie na jego wyniki ekonomiczne i sferę środowiskową [Staniszewski 2016] oraz wpływ zmian w strukturze rolnictwa na podstawowe wskaźniki stanu rozwoju gospodarczego [Wąs, Małażewska 2012] oraz zmiany zachodzące w finansowym wsparciu rolnictwa [Jankowska 2015; Kułyk 2015].

Wykorzystując miary koncentracji, podjęto próby regionalizacji rolnictwa na świecie [Rokicki, Golonko, Perkowska 2017; Golonko, Wysokiński, Gromada 2021] oraz regionalizacji produkcji zwierzęcej [Rokicki, Golonko, Perkowska 2018a; Rokicki, Ringdorfer 2020; Rokicki, Perkowska, Ziółkowska 2020]. Badano również stopień występowania systemów agroleśnych w krajach UE [Rokicki, Golonko, Perkowska 2018b].

Analizowano procesy koncentracji czynników produkcji i wsparcia oraz ich związek z produktywnością w gospodarstwach rolniczych [Czyżewski, Staniszewski 2017; Kisielińska 2019]. Prowadzono również badania koncentracji rolnictwa ekologicznego [Makowska, Gotkiewicz, Pawlewicz 2015].

Wiele uwagi poświęcono nierównościom dochodowym w rolnictwie i ubóstwu ekonomicznemu na wsi [Murawska 2014; Raczkowska 2015; Kobus 2016; Kryszak 2016; Murawska 2017; Raczkowska, Wrześcińska-Kowal 2018; Średzińska 2018; Czyżewski, Kata, Matuszczak 2019]. Wskazano na związki nierówności dochodowych z bezpieczeństwem żywnościowym [Grzelak 2016].

Częste zastosowania analizy strukturalnej w ekonomice rolnictwa, w tym prostej analizy wskaźników struktury, dotyczą zagadnień związanych z produkcją rolniczą. Prowadzono badania koncentracji produkcji mleka [Komorowska 2006; Dzwonkowski 2015], koncentrację i siłę rynkową w przemyśle spożywczym [Kufel, Hamulczuk 2015], pozycji polskiego przetwórstwa spożywczego [Podlińska 2017], przepływy produktów grup producentów owoców i warzyw [Bieniek-Majka 2016]. Badano strukturę produkcji indywidualnych gospodarstw rolnych w Polsce na tle krajów UE, analizowano strukturę obszarową gospodarstw rolnych, strukturę użytków rolnych, klas bonitacyjnych, zasiewów oraz wpływ regionalnego zróżnicowania struktury zasiewów na środowisko przyrodnicze [Kukuła 2007; Wasilewska 2008; Ryś-Jurek 2010; Matyka 2017]. Analizowano także kierunki i przyczyny przemian struktur podmiotowych przemysłu spożywczego [Chechelski 2018], związek nakładów na działalność innowacyjną w przedsiębiorstwach przemysłu mięsnego z wykształceniem pracowników [Wasilewska, Wasilewska 2016] oraz wpływ struktury użytków rolnych na wyniki ekonomiczne gospodarstw [Rokicki 2007]. Analiza strukturalna wykorzystana została także do określenia zmian zasobów pracy w rolnictwie [Szuba-Barańska, Mrówczyńska-Kamińska, Poczta 2019] oraz zmian struktury ekonomicznej gospodarstw rolnych [Rzeszutko 2016].

Inna grupa badań, w których zastosowano analizy strukturalne, dotyczyła zagadnień finansowych. Analizowano modele finansowania w przedsiębiorstwach rolniczych [Wasilewski, Oliynyk, Adamenko 2015], badano związek koncentracji kapitału własnego na jego efektywność w spółdzielniach mleczarskich [Zuba-Ciszewska 2015]. Rozważano również związki między zobowiązaniami kredytowymi rolników a koniunkturą i zmianami strukturalnymi w rolnictwie [Kata 2020].

Wśród cytowanych prac można znaleźć również takie, w których wykorzystano analizę struktur do badania zachowań konsumentów na rynku produktów żywnościowych. Badaniom poddano zmiany poziomu i struktury wydatków na żywność oraz struktury konsumpcji żywności [Świątlik 2017, 2020].

Badania struktur dotyczyły również zagadnień związanych z problematyką demograficzną. Rozważano zmiany demograficzne na obszarach wiejskich [Rakowska 2011], analizowano problem starzenia się polskiej wsi na tle miast [Wasilewska 2017], badano związek zmian płodności na obszarach wiejskich i miejskich z bezrobociem [Podogrodzka 2012].

Podsumowując, można stwierdzić, że możliwości zastosowania analiz strukturalnych w ekonomice rolnictwa są znaczne, a mnogość problemów badawczych rozważanych w cytowanych pracach wskazuje na dużą przydatność tego typu analiz. Należy przy tym podkreślić, że często analizy struktury zjawisk stanowią tylko wstęp do dalszych, pogłębionych badań.

Tabela 9.1. Przykłady zastosowań analizy struktury w ekonomice rolnictwa

Autor	Zagadnienie	Zbiorowość	Metody
Komorowska 2006	zmiany w koncentracji produkcji mleka	gospodarstwa rolne	analiza wskaźników struktury
Wasilewska 2006	zmiany struktury aktywności ekonomicznej ludności wiejskiej	mieszkańcy wsi	analiza wskaźników struktury
Kukuła 2007	struktura obszarowa, użytków rolnych, klas bonitacyjnych, zasiewów	województwa	analiza strukturalna
Rokicki 2007	wpływ struktury użytków rolnych na wyniki ekonomiczne gospodarstw	gospodarstwa owczarskie	analiza wskaźników struktury
Wasilewska 2007	zmiany struktury aktywności ekonomicznej ludności wiejskiej	mieszkańcy wsi	analiza strukturalna
Wasilewska 2008	zmiany struktury zasiewów	gospodarstwa rolne	analiza wskaźników struktury
Pawlak, Poczta 2010	konkurencyjność potencjalna (zasobowo-nakładowa) polskiego rolnictwa	gospodarstwa rolne	analiza wskaźników struktury, analiza koncentracji (krzywa koncentracji Lorenza)
Ryś-Jurek 2010	struktura produkcji indywidualnych gospodarstw rolnych w Polsce na tle UE-27	gospodarstwa rolne	analiza wskaźników struktury
Gąsiorowska-Mącznik 2011	rozwój przedsiębiorczości pozarolniczej na obszarach wiejskich	przedsiębiorcy	analiza wskaźników struktury
Rakowska 2011	zmiany demograficzne na obszarach wiejskich	mieszkańcy miast i wsi	analiza wskaźników struktury
Podogrodzka 2012	bezrobocie a zmiany płodności na obszarach wiejskich i miejskich	województwa	analiza wskaźników struktury
Wąs, Małażewska 2012	zmiany w rolnictwa a podstawowe wskaźniki stanu rozwoju gospodarczego	gospodarstwa rolne	analiza koncentracji (współczynnik Giniego)
Binderman, Koszela, Szczesny 2014	zmiany w strukturze gospodarstw rolnych	gospodarstwa rolne	analiza koncentracji (współczynnik Giniego, krzywa Lorenza, współczynnik ar)
Murawska 2014	sytuacja finansowa wiejskich gospodarstw domowych	gospodarstwa domowe	analiza koncentracji (współczynnik Giniego)
Sroka 2014	identyfikacja i ocena znaczenia rolnictwa miejskiego	gospodarstwa indywidualne na terenie miast	analiza wskaźników struktury
Wrzaszcz, Zegar 2014	ocena sprawności ekonomicznej wybranych form gospodarstw przyjaznych dla środowiska przyrodniczego	gospodarstwa rolne	analiza wskaźników struktury

Tabela 9.1 (cd.)

Autor	Zagadnienie	Zbiorowość	Metody
Dzwonkowski 2015	procesy koncentracji produkcji mleka a popyt na pasze przemysłowe w żywieniu krów mlecznych	gospodarstwa rolne	analiza wskaźników struktury
Jankowska 2015	koncentracja dopłat bezpośrednich	kraje Europy	analiza wskaźników struktury, analiza koncentracji (współczynnik Giniego)
Kufel, Hamulczuk 2015	koncentracja a wywieranie siły rynkowej w przemyśle spożywczym	przedsiębiorstwa przemysłu spożywczego	analiza wskaźników struktury
Kułyk 2015	zmiany zachodzące w finansowym wsparciu rolnictwa oraz związki pomiędzy transferami	państwa	analiza koncentracji (współczynnik Giniego)
Makowska, Gotkiewicz, Pawlewicz 2015	rozwój rolnictwa ekologicznego w Polsce	gospodarstwa rolne	analiza koncentracji (krzywa koncentracji Lorenza)
Popa, Dinu 2015	zarządzanie gospodarstwem w warunkach rozdrobnionego rolnictwa	gospodarstwa rolne	analiza koncentracji (współczynnik Giniego)
Raczkowska 2015	nierówności dochodowych w układzie miast-wieś	mieszkańcy miast i wsi	analiza koncentracji (współczynnik Giniego)
Wasilewski, Oliynyk, Adamenko 2015	modele finansowania w przedsiębiorstwach rolniczych	przedsiębiorstwa rolnicze	analiza koncentracji (krzywa Lorenza)
Zaród 2015	kierunek i tempo przemian w rolnictwie	gospodarstwa rolne	analiza strukturalna
Zuba-Ciszewska 2015	koncentracja kapitału własnego a jego efektywność w spółdzielniach mleczarskich	spółdzielnie mleczarskie	analiza wskaźników struktury
Bieniek-Majka 2016	przepływy produktów grup producentów owoców i warzyw	grupy producentów	analiza wskaźników struktury
Binderman, Szczesny 2016	zmiany struktury obszarowej gospodarstw	gospodarstwa rolne	analiza koncentracji
Grzelak 2016	bezpieczeństwo żywnościowe a nierówności dochodowe	kraje	analiza koncentracji (współczynnik Giniego)
Klepacki, Gołasa, Wysokiński 2016	emisja gazów cieplarnianych w rolnictwie	kraje UE	analiza koncentracji (krzywa koncentracji Lorenza, współczynnik Giniego)
Kobus 2016	poziom nierówności w rolnictwie	kraje UE	analiza koncentracji (współczynnik Giniego)

Kryszak 2016	nierówności dochodowe w rolnictwie	gospodarstwa rolne	analiza koncentracji (współczynnik Giniego)
Rzeszutko 2016	zmiany struktury ekonomicznej gospodarstw rolnych	gospodarstwa rolne	regionalna analiza strukturalna
Staniszewski 2016	wpływ koncentracji ziemi w rolnictwie na jego wyniki ekonomiczne i sferę środowiskową	kraje UE	analiza koncentracji (współczynnik Giniego)
Wasilewska, Wasilewska 2016	nakłady na działalność innowacyjną w przedsiębiorstwach przemysłu mięsnego a wykształcenie pracowników	przedsiębiorstwa	analiza strukturalna
Chechelski 2017	przemiany struktur podmiotowych przemysłu spożywczego i branż przetwórstwa produktów pochodzenia roślinnego	przedsiębiorstwa	analiza wskaźników struktury
Czyżewski, Staniszewski 2017	struktura czynników produkcji w rolnictwie	kraje Europy	analiza strukturalna, analiza koncentracji (krzywa Lorenza)
Góral, Rembisz 2017	wydajność pracy w rolnictwie	kraje UE	analiza koncentracji (współczynnik Giniego, krzywa Lorenza)
Lisek 2017	struktura wielkościowa przedsiębiorstw w Polsce	przedsiębiorstwa	analiza strukturalna, metoda eliminacji wektorów
Matyka 2017	regionalne zróżnicowanie struktury zasiewów a środowisko przyrodnicze	gospodarstwa rolne	analiza strukturalna, analiza skupień
Murawska 2017	nierówności społeczne na obszarach wiejskich	gospodarstwa domowe	analiza koncentracji (współczynnik Giniego)
Podlińska 2017	pozycja polskiego przetwórstwa spożywczego	przedsiębiorstwa	analiza wskaźników struktury
Rokicki, Golonko, Perkowska 2017	regionalizacja rolnictwa na świecie	kontynenty	analiza koncentracji (współczynnik Giniego, krzywa Lorenza)
Świetlik 2017	poziom i struktura wydatków na żywność oraz spożycia żywności	gospodarstwa domowe	analiza wskaźników struktury, analiza dynamiki
Wasilewska 2017	starzenie się wsi na tle miast	mieszkańcy miast i wsi	analiza wskaźników struktury
Chechelski 2018	kierunki i przyczyny przemian struktur podmiotowych przemysłu spożywczego	przedsiębiorstwa	analiza wskaźników struktury
Majchrzak 2018	koncentracja gruntów rolnych w państwach UE	kraje UE	analiza koncentracji (krzywa Lorenza, współczynnik Giniego)

Tabela 9.1 (cd.)

Autor	Zagadnienie	Zbiorowość	Metody
Raczkowska, Wrzesińska-Kowal 2018	nierówności i ubóstwo ekonomiczne	mieszkańcy miast i wsi	analiza koncentracji (współczynnik Giniego)
Rokicki, Golonko, Perkowska 2018a	regionalizacja produkcji zwierzęcej na świecie	kontynenty	analiza koncentracji (krzywa Lorenza, współczynnik Giniego)
Rokicki, Golonko, Perkowska 2018b	stopień występowania systemów agroleśnych w krajach UE	kraje UE	analiza koncentracji (współczynnik Giniego, krzywa Lorenza)
Średzińska 2018	zróżnicowanie dochodów uzyskiwanych przez rolników	rolnicy	analiza koncentracji (współczynnik Giniego)
Czyżewski, Kata, Matuszczak 2019	wydatki budżetowe na rolnictwo a dochody rolników	gospodarstwa domowe rolników	analiza koncentracji (współczynnik Giniego)
Kisielińska 2019	koncentracja czynników produkcji i wsparcia a produktywność w gospodarstwach rolniczych	gospodarstwa rolne	analiza koncentracji (współczynnik Giniego)
Szuba-Barańska, Mrówczyńska- -Kamińska, Poczta 2019	zmiany zasobów pracy w rolnictwie	kraje EŚW	analiza strukturalna
Kata 2020	zobowiązania kredytowe rolników a koniunktura i zmiany strukturalne w rolnictwie	gospodarstwa rolne	analiza wskaźników struktury
Mikołajczyk 2020	porównanie zmienności struktury wydatków publicznych	kraje EŚW	analiza wskaźników struktury
Rokicki, Perkowska, Ziółkowska 2020	koncentracja produkcji zwierzęcej	województwa	analiza koncentracji (krzywa Lorenza, współczynnik Giniego)
Rokicki, Ringdorfer 2020	produkcja mięsa baraniego	gospodarstwa rolne	analiza koncentracji (krzywa Lorenza, współczynnik Giniego)
Świetlik 2020	zmiany poziomu i struktury konsumpcji żywności	gospodarstwa domowe	analiza wskaźników struktury
Golonko, Wysokiński, Gromada 2021	koncentracja i regionalizacja rolnictwa na świecie	kontynenty	analiza koncentracji (krzywa Lorenza, współczynnik Giniego)

Źródło: opracowanie własne

Literatura

- Bieniek-Majka M. (2016) Kierunki przepływów input/output produktów grup producentów owoców i warzyw. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 28(1), 23-27.
- Binderman Z., Koszela G., Szczesny W. (2014) Zmiany w strukturze gospodarstw rolnych w krajach Unii Europejskiej w latach 2003-2010 (aspekty metodyczne). *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 14(19)(3), 15-26.
- Binderman Z., Szczesny W. (2016) Ocena zmian struktury obszarowej w Polsce na tle nowych państw UE (aspekty metodyczne). *Optimum Studia Ekonomiczne*, 2(80), 81-96.
- Bogocz D., Bożek J., Kukuła K. (red.), Strojny J. (2010) *Statystyczne studium struktury agrarnej w Polsce*. Wydawnictwo Naukowe PWN, Warszawa.
- Chechelski P. (2017) procesy koncentracji struktur podmiotów gospodarczych w branżach przetwórstwa produktów pochodzenia roślinnego w Polsce. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 19(1), 14-19.
- Chechelski P. (2018) Przemiany struktury podmiotowej przemysłu spożywczego w Polsce w XXI wieku. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 20(5), 29-35.
- Chomętowski S., Sokołowski A. (1978) Taksonomia struktur. *Przegląd Statystyczny*, 25(2), 217-226.
- Czyżewski A., Kata R., Matuszczak A. (2019) The redistribution function in Poland's agricultural budgets in the long term. *Acta Scientiarum Polonorum, Oeconomia*, 18(2), 25-35.
- Czyżewski A., Staniszewski J. (2017) Agricultural production factors structures in selected EU-15 countries and Poland. Similarities and differences. *Journal of Agribusiness and Rural Development*, 2(44), 295-304.
- Dzwonkowski W. (2015) Ocena wpływu procesów koncentracji produkcji mleka na popyt na pasze przemysłowe w żywieniu krów mlecznych. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(4), 75-80.
- Gąsiorowska-Mącznik E. (2011) Stan i możliwości rozwoju pozarolniczej przedsiębiorczości na obszarach wiejskich województwa świętokrzyskiego w świetle badań ankietowych. *Więś i Rolnictwo*, 4(153), 157-177.
- Golonko M., Wysokiński M., Gromada A. (2021) Concentration and regionalization of agriculture in the world. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 23(1), 36-47.
- Góral J., Rembisz W. (2017) Wynagrodzenia i wydajność pracy w rolnictwie polskim na tle pozostałych krajów Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 119, 119-139.
- Grzelak A. (2016) Związki między bezpieczeństwem żywnościowym a nierównościami dochodowymi. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 18(3), 95-100.
- Jankowska A. (2015) Koncentracja dopłat bezpośrednich w krajach środkowej i wschodniej Europy w 2012 roku. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(4), 98-103.
- Kata R. (2020) Zadłużenie rolników w Polsce w aspekcie przemian strukturalnych i koniunktury w rolnictwie. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 23(72), 33-44.
- Kisielińska J. (2019) Concentration of production factors and support and their productivity in EU farm. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 21(4), 204-214.
- Klepacki B., Gołasa P., Wysokiński M. (2016) Efektywność emisji gazów cieplarnianych w rolnictwie Unii Europejskiej. *Więś i Rolnictwo*, 3(172), 129-144. DOI: 10.7366/wir032016/06
- Kobus P. (2016) Inequalities in Agricultural Subsidies in European Union. *Zeszyty Naukowe Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 16(31)(4), 179-187.
- Komorowska D. (2006) Koncentracja produkcji mleka w Polsce. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 61, 159-167.
- Kryszak Ł. (2016) Nierówności dochodowe w rolnictwie krajów Unii Europejskiej w kontekście koncepcji zrównoważonego rozwoju. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(2), 166-171.

- Kufel J., Hamulczuk M. (2015) Koncentracja a wywieranie siły rynkowej w polskim przemyśle spożywczym. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(5), 158-163.
- Kukuła K. (1975) Propozycja w zakresie pewnych miar dynamiki struktur. *Przegląd Statystyczny*, 3(22), 453-461.
- Kukuła K. (1986) Przegląd wybranych miar zgodności struktur. *Przegląd Statystyczny*, 4(33), 385-401.
- Kukuła K. (1996) *Statystyczne metody analizy struktur ekonomicznych*. Wydawnictwo Edukacyjne, Kraków.
- Kukuła K. (2007) Z problematyki nad strukturą agrarną w Polsce w ujęciu przestrzennym. *Acta Scientiarum Polonorum, Oeconomia*, 6(4), 19-27.
- Kułyk P. (2015) Relacje strukturalne w finansowym wsparciu rolnictwa w grupach państw o różnym poziomie rozwoju gospodarczego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(6), 140-145.
- Lisek S. (2017) Struktura wielkościowa przedsiębiorstw w Polsce. *Metody Ilościowe w Badaniach Ekonomicznych*, 18(4), 635-642.
- Majchrzak A. (2018) Koncentracja gruntów rolnych w państwach członkowskich Unii Europejskiej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 20(5), 99-104.
- Makowska M., Gotkiewicz W., Pawlewicz A. (2015) Rolnictwo ekologiczne w Polsce w ujęciu przestrzennym i środowiskowym. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 17(4), 160-165.
- Malina A. (2004) Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Marciniak S. (1970) *Struktura produkcji a dynamika wzrostu gospodarczego*. PWN, Warszawa.
- Matyka M. (2017) Ocena regionalnego zróżnicowania struktury zasiewów w kontekście oddziaływania na środowisko przyrodnicze. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 19(3), 188-192.
- Mikołajczyk K. (2020) Zmienność struktury wydatków publicznych: analiza porównawcza dla krajów Europy środkowo-wschodniej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 23(72), 134-135.
- Murawska A. (2014) Sytuacja finansowa wiejskich gospodarstw domowych w aspekcie zrównoważonego rozwoju. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 16(1), 147-152.
- Murawska A. (2017) Ocena nierówności społecznych na obszarach wiejskich w Polsce w świetle realizacji celów polityki spójności. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 19(3), 206-211.
- Panek T. (2011) *Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru*. Oficyna Wydawnicza SGH, Warszawa.
- Pawlak K., Poczta W. (2010) Potencjał polskiego rolnictwa pięć lat po akcesji Polski do UE jako przesłanka jego konkurencyjności. *Więś i Rolnictwo*, 1(146), 21-47.
- Podlińska O. (2017) Przetwórstwo spożywcze w Polsce na tle krajów Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 18(67), 137-145.
- Podogrodzka M. (2012) Bezrobocie determinantą przestrzennego zróżnicowania płodności według miejsca zamieszkania w Polsce w latach 1999–2009. *Więś i Rolnictwo*, 1(154), 119-131.
- Popa D.G., Dinu T.A. (2015) Correlations between fragmentation of farms in the Republic of Moldova and its impact on farm incomes compared to Poland and Romania. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 15(30)(4), 166-179.
- Raczkowska M. (2015) Nierówności dochodowe w Polsce w relacji miasto–wieś [Income discrepancies in Poland in urban-rural relations]. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 112, 111-124.
- Raczkowska M., Wrzesińska-Kowal J. (2018) Nierówności i ubóstwo ekonomiczne na obszarach wiejskich w Polsce. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 105(1), 7-17.
- Rakowska J. (2011) Zmiany demograficzne na obszarach wiejskich Polski. *Więś Jutra*, 11(12), 7-10.
- Rokicki T. (2007) Wpływ struktury użytków rolnych na wyniki ekonomiczne gospodarstw zajmujących się hodowlą owiec. *Roczniki Nauk Rolniczych, Seria G*, 94(1), 46-50.
- Rokicki T., Golonko M., Perkowska A. (2017) Regionalizacja rolnictwa na świecie. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 17(32)(3), 239-248.

- Rokicki T., Golonko M., Perkowska A. (2018a) Regionalizacja produkcji zwierzęcej na świecie. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(33)(3), 272-281
- Rokicki T., Golonko M., Perkowska A. (2018b) Zastosowanie systemów agroleśnych w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Problemy Rolnictwa Światowego*, 18(33)(2), 249-258.
- Rokicki T., Perkowska A., Ziółkowska P. (2020) Changes in the concentration of animal production in Poland. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 22(3), 175-186.
- Rokicki T., Ringdorfer F. (2020) Sheep meat production in Austria and Poland similarities and differences. *Acta Scientiarum Polonorum, Oeconomia*, 19(3), 97-106.
- Rutkowski J. (1981) Podobieństwo struktur i zmiany strukturalne – zagadnienia kwantyfikacji. *Wiadomości Statystyczne*, 8, 20-23.
- Ryś-Jurek R. (2010) Struktura produkcji indywidualnego gospodarstwa rolnego w Polsce i w krajach UE-27. *Analiza podobieństw w 2007 roku. Wieś i Rolnictwo*, 3(148), 73-87.
- Rzeszutko A. (2016) Zmiany struktury ekonomicznej gospodarstw rolnych w Polsce w ujęciu regionalnej analizy strukturalnej. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(4), 32-47.
- Sobczyk M. (2007) *Statystyka*. Wydawnictwo Naukowe PWN, Warszawa.
- Sroka W. (2014) Rolnictwo w obrębie miast – wybrane aspekty zmian strukturalnych przed i po przystąpieniu Polski do Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 108, 33-43.
- Staniszewski J. (2016) Zrównoważona intensyfikacja rolnictwa w państwach Unii Europejskiej o zróżnicowanej strukturze agrarnej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 18(1), 238-243.
- Stawicki J. (2004) *Wykorzystanie łańcuchów Markowa w analizie rynku kapitałowego*. Wydawnictwo UMK, Toruń.
- Strahl D. (1998) *Taksonomia struktur w badaniach regionalnych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Szuba-Barańska E., Mrówczyńska-Kamińska A., Poczta W. (2019) Labour resources in agribusiness in Central-Eastern Europe. *Journal of Agribusiness and Rural Development*, 2(52), 179-187.
- Średzińska J. (2018) Zróżnicowanie poziomu dochodów rolników w gospodarstwach z różnych klas wielkości ekonomicznej w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 20(69), 215-223.
- Świetlik K. (2017) Preferencje konsumentów na rynku żywności w Polsce w warunkach deflacji w latach 2014-2015. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 117, 75-92.
- Świetlik K. (2020) Trends in behaviour of modern food consumers in Poland. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Polityki Europejskie, Finanse i Marketing*, 23(72), 191-207.
- Urbańczyk E., Jurek M. (1998) *Wielokierunkowa analiza wyniku finansowego przedsiębiorstwa*. Fundacja na Rzecz Uniwersytetu Szczecińskiego, Szczecin.
- Walesiak M. (1984) Pojęcie, klasyfikacja i wskaźniki podobieństwa struktur gospodarczych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 285, 57-78.
- Wasilewska E. (2006) Tendencje zmian w strukturze aktywności ekonomicznej ludności wiejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 61, 131-150.
- Wasilewska E. (2007) Zastosowanie syntetycznych mierników dynamiki struktury w analizie zmian aktywności ekonomicznej ludności wiejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 62, 67-80.
- Wasilewska E. (2008) Zmiany w strukturze zasiewów w Polsce w latach 1996-2007. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomia i Organizacja Gospodarki Żywnościowej*, 71, 123-134.
- Wasilewska E. (2011) *Statystyka opisowa od podstaw. Podręcznik z zadaniami*. Wydawnictwo SGGW, Warszawa.

- Wasilewska E. (2017) Starość demograficzna obszarów wiejskich i jej zróżnicowanie. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 104(3), 75-85.
- Wasilewska A., Wasilewska E. (2016) Syntetyczne miary zróżnicowania struktur w ocenie związku struktury wykształcenia z poziomem innowacyjności w przedsiębiorstwach przemysłu mięsnego. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 103(3), 75-83.
- Wasilewski M., Oliynyk O., Adamenko V. (2015) Equity and debt financing models of agricultural enterprises in Ukraine: challenges and perspectives. *Zarządzanie Finansami i Rachunkowość*, 3(3), 93-106.
- Wąs A., Małażewska S. (2012) Przemiany strukturalne w rolnictwie w wybranych krajach europejskich. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 99(4), 75-88.
- Wrzaszcz W., Zegar J.S. (2014) Sprawność ekonomiczna wybranych form rolnictwa przyjaznego dla środowiska przyrodniczego. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 16(1), 226-232.
- Zaród J. (2015) Dynamika i kierunki przemian strukturalnych w rolnictwie woj. zachodniopomorskiego. *Wiadomości Statystyczne*, 7, 90-101.
- Zuba-Ciszewska M. (2015) Wpływ koncentracji kapitału własnego na jego efektywność w wybranych spółdzielniach mleczarskich w Polsce w latach 2003-2012. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Ekonomika i Organizacja Gospodarki Żywnościowej*, 109, 107-121.
- Żwirbla A. (2006) Próba konstrukcji mierników struktury oraz zmian strukturalnych. *Wiadomości Statystyczne*, 10, 10-22.