

## WIELOWYMIAROWA ANALIZA DANYCH W EKONOMICE ROLNICTWA

*Joanna Kisielińska, Stanisław Stańko*

Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych Szkoły  
Głównej Gospodarstwa Wiejskiego w Warszawie  
Kierownik: prof. dr hab. Henryk Manteuffel

Słowa kluczowe: analiza wielowymiarowa, regresja wieloraka, klasyfikacje obiektów,  
analiza dyskryminacyjna, czynnikiowa

*Key words: multivariate analysis, multiple regression, discriminant analysis, factor  
analysis*

S y n o p s i s. Przedstawiono metody wielowymiarowej analizy danych wraz z przykła-  
dami ich stosowania w ekonomice rolnictwa. Obejmują one regresję wieloraką, wzor-  
cową i bezwzorcową klasyfikację obiektów, liniowe porządkowanie, analizę czynnikiową  
i metodę głównych składowych.

### WSTĘP

Wraz z rozwojem społeczno-gospodarczym zwiększają się powiązania między przedsię-  
biorstwami i krajami, regionami, nasilają się procesy integracji i globalizacji. Wszystko to  
powoduje, że na efekty działalności, dynamikę zjawisk i procesów gospodarczych oraz  
społecznych oddziałuje jednocześnie wiele różnych czynników. Skutkuje to również dużym  
różnicowaniem funkcjonujących w tych warunkach jednostek gospodarczych. Dlatego w  
ekonomii często mamy do czynienia z danymi statystycznymi pochodzącymi z wielu jedno-  
stek, które dodatkowo charakteryzują się różnorodnymi właściwościami. Tak zgromadzony  
materiał powinien być usystematyzowany i zestawiony stosownie do celów, w jakich został  
zgromadzony, na przykład do identyfikacji grup podobnych obiektów, odkrywania niezna-  
nej struktury analizowanych danych, do porównywania obiektów, czy też poszukiwania  
zależności między wieloma cechami je opisującymi. Jest to problem o kapitalnym znaczeniu,  
polega on na mniej lub bardziej różnicowanym podziale niejednorodnej zbiorowości na  
możliwie jednorodne grupy według obranych kryteriów. W tej sytuacji coraz większego  
znaczenia w badaniach ekonomicznych nabierają metody wielowymiarowe pozwalające na  
przeprowadzenie różnego rodzaju analiz badających więcej niż jedną zmienną.

Celem opracowania jest przedstawienie wybranych technik (metod) wielowymiarowej  
analizy danych wraz z przykładami ich stosowania w ekonomice rolnictwa.

## ISTOTA ANALIZY WIELOWYMIAROWEJ

Wielowymiarowa analiza danych zajmuje się badaniem zbiorów obiektów, opisanych wieloma cechami. Informacje o obiektach umieszcza się w macierzy zwanej macierzą obserwacji, której wiersze odpowiadają obiektom, zaś kolumny zebranych cechom.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

gdzie:  $n$  – liczba obiektów,  $k$  – liczba cech.

W niektórych metodach z macierzy obserwacji wyodrębnia się zmienną zależną  $y$ , np. w regresji i analizie dyskryminacyjnej.

Kluczowym momentem każdego badania przy zastosowaniu metody wielowymiarowej analizy jest wybór cech diagnostycznych. Spośród wielu zmiennych należy wybrać te, które mają największy wpływ na badane zjawisko oraz dostarczają najistotniejszych informacji. Jest to podstawowy problem natury merytorycznej, jak i statystycznej.

Większość współczesnych autorów jest zdania, że dobór cech nie powinien być subiektywny, a oparty na określonej procedurze [np. Grabiński i in. 1982, Grabiński 1992, Gatnar 1999, Witkowska 2005]. Według tej procedury postępowanie jest następujące:

- ustalanie wstępnego zestawu cech (zmiennych) na podstawie kryteriów merytorycznych,
- poddawanie tego zestawu analizie formalno-statystycznej,
- na tej podstawie wybór optymalnego zestawu zmiennych objaśniających.

Kryteria merytoryczne doboru cech można ująć następująco:

1. Należy preferować te cechy, które pozostają w pewnym związku merytorycznym z modelowanym zjawiskiem. Pominięcie ich może spowodować zubożenie prowadzonej analizy lub powodować trudności przy estymacji.
2. Zmienne powinny być dobrymi reprezentantami różnych aspektów badanego odcinka rzeczywistości gospodarczej, co można nazwać uniwersalnością.
3. Zmienne powinny cechować logiczność wzajemnych powiązań.
4. Cechy powinny mieć określone tradycje badawcze. Chodzi tu o ustaloną interpretację i kontrolę merytoryczną.
5. Powinny być dostępne i wiarygodne dane statystyczne, dotyczące wyróżnionych cech.
6. Cechy powinny mieć charakter mierzalny (bezpośredni lub pośredni). Cechy, które nie są bezpośrednio kwantyfikowalne należy przekształcić tak, aby otrzymać cechy mierzalne.

W wyniku wskazań merytorycznych otrzymuje się wstępną listę cech, które są zakwalifikowane do dalszych badań. Cechy w takim zbiorze są potencjalnymi cechami diagnostycznymi, która mogą, ale nie muszą, być uwzględnione w dalszym postępowaniu. Cechy diagnostyczne powinny także spełniać określone kryteria formalno-statystyczne. Na podstawie tych kryteriów ze wstępnej listy cech eliminuje się te, które ich nie spełniają. Najczęściej kryteria te są następujące:

- cechy powinny charakteryzować się określoną zmiennością; zmienność tę najczęściej mierzy się współczynnikiem zmienności; ze zbioru eliminuje się te zmienne, które nie różnicują badanych obiektów (są *quasi* stałe),
- cechy nie powinny być skorelowane i współliniowe; chodzi tu o wyeliminowanie zjawiska powtarzania tych samych informacji niesionych przez różne cechy; współliniowość wpływa również na zwiększenie wariancji estymatorów oraz może wpływać na

zmianę znaków algebraicznych estymatorów i parametrów, nawet gdy z merytorycznej analizy wynika, że znaki te powinny być dodatnie; model dostarcza w takiej sytuacji fałszywych wskazań,

- wymagana jest zwykle kompletność danych dla wszystkich obiektów,
- zestawy zmiennych powinny posiadać określony potencjał informacyjny.

Zastosowanie kryteriów formalno-statystycznych pozwala wyeliminować te zmienne, które nie spełniają wymaganych kryteriów. Nie znaczy to jednak, że pozostałe muszą być uwzględnione w badaniach. Z otrzymanego zestawu należy wybrać zbiór optymalny, spełniający określone kryteria. Metody takiego doboru są bardzo zróżnicowane. Można tutaj wymienić: analizę współczynników korelacji, regułę „stop”, metodę wskaźników pojemności informacyjnej Hellwiga. Poza selekcją *a priori* zastosować można również eliminację *a posteriori*, stosując metody krokowe.

Największe uznanie wśród ekonometryków prowadzących badania empiryczne w Polsce zdobyła metoda integralnych informacji Hellwiga [1969, Grabiński i in. 1982]. Wadą jej jest lawinowy wzrost liczby badanych kombinacji wraz ze wzrostem wymiarowości problemu.

Znanych jest wiele różnorodnych metod wielowymiarowej analizy danych. W ekonomice rolnictwa najczęściej stosowano:

- regresję wieloraką,
- wzorcową klasyfikację obiektów,
- bezwzorcową klasyfikację obiektów,
- liniowe porządkowanie obiektów,
- analizę czynnikową i metodę składowych głównych.

## REGRESJA WIELORAKA

Najbardziej ogólnym modelem regresji jest związek funkcyjny między zmienną zależną  $y$  i wektorem zmiennych niezależnych  $x$ . Ponieważ przewidywanie wartości zmiennej  $y$  na podstawie znajomości  $x$  obarczone jest zwykle pewnym błędem, do równania wprowadza się element zwany składnikiem losowym. Ogólny model regresji zapisuje się jako:

$$y = g(x) + \varepsilon$$

Problem sprowadza się do dobrania funkcji  $g$ .

Zwykle poszukiwania właściwego modelu ogranicza się do konkretnej postaci funkcji, która może być w szczególności funkcją nieliniową. Mowa jest wówczas o regresji nieliniowej:

$$y = f(x, \beta) + \varepsilon$$

Zadanie polega na wyznaczeniu wektora parametrów  $\beta$ .

Jeżeli funkcja  $f$  jest liniowa, to mówimy wówczas o regresji liniowej, którą można zapisać jako:

$$y = \beta^T \cdot x + \varepsilon$$

Model regresyjny budowany jest na podstawie zaobserwowanych wartości zmiennej zależnej  $y$  i odpowiadających im wartości zmiennych niezależnych  $x$ , tak aby uzyskać jak najmniejsze błędy. Przez błędy rozumie się różnicę między wartością rzeczywistą zmiennej zależnej, a jej wartością przewidywaną. Miarą błędu modelu jest najczęściej suma kwadratów tych różnic. Dopasowanie modelu tak, aby otrzymać najmniejszą sumę kwadratów błędów nazywa się metodą najmniejszych kwadratów. Sposób wyznaczenia modelu regresyjnego metodą najmniejszych kwadratów zależy od rodzaju modelu.

W regresji liniowej wektor  $b$  stanowiący oszacowanie wektora parametrów  $\beta$  w równaniu regresyjnym określa następująca formuła macierzowa:

$$b = (X^T \cdot X)^{-1} \cdot X^T \cdot y$$

Nieliniowe zadania regresyjne dzielone są na dwie grupy: zadania linearyzowalne oraz ściśle nieliniowe. Modele linearyzowane sprowadza się do modeli liniowych przez odpowiednie transformacje zmiennych. Natomiast modeli ściśle nieliniowych poszukuje się metodami optymalizacji nieliniowej, w których funkcje zmiennych decyzyjnych pełnią wektory parametrów  $\beta$ .

Jeżeli chodzi o ogólny model regresji należy podkreślić, że nie ma metody wyznaczania nieznannej funkcji  $g$  metodą najmniejszych kwadratów. Problem taki można niejako obejść stosując sieć neuronową. Udowodniono bowiem, że sieć perceptronowa czy o radialnych funkcjach bazowych stanowi aproksymator uniwersalny. Oznacza to, że budując odpowiednią sieć można odwzorować dowolną zależność między jej wejściem, a wyjściem. Przykłady zastosowań tej metody w ostatnich latach przedstawiono w tabeli 1. Rodzaje zmiennych zależnych występujące w modelach wskazują na różnorodność zjawisk, które badacze chcą wyjaśniać metodami regresyjnymi. Najwięcej przykładów odnosi się do dochodów i kosztów oraz wskaźników finansowych.

Tabela 1. Przykłady zastosowań regresji wielorakiej w ekonomice rolnictwa (od 2000 r.)

Autor	Zmienna zależna	Metoda (liczba zmiennych niezależnych)
Gołębiewska, Klepacki [2000]	dochód rolniczy	regresja liniowa (8,5,6)
Szuk [2001]	nakłady pracy ludzkiej i mechanicznej	regresja liniowa (5)
Kowalczyk [2002]	wartość użytków rolnych	funkcja potęgowa (7)
Kisielińska [2003a]	dochód rolniczy	regresja liniowa i sieć neuronowa (12)
Pietrzak [2004]	wartość dodatkowa ze sprzedaży na litr skupu mleka	regresja liniowa (3)
Płonka, Sobczyński [2004]	ryzyko utraty zdolności płatniczej	funkcja kwadratowa (13)
Majewski [2005]	wielkość skupu mleka na 1 hektar UR	regresja liniowa (5)
Pawłowska, Piereliğin [2005]	zysowność	regresja liniowa (4)
Czekaj [2006]	dochód rolniczy	
Stępień [2006]	koszty jednostkowe bezpośrednie, pośrednie i całkowite na 1 kg żywca wieprzowego	regresja liniowa i funkcja potęgowa (3)
Wysocki, Kurzawa [2006]	wydatki na spożycie	regresja liniowa (4)
Mańko, Sass, Sobczyński [2007]	liczba inwestorów w gminie	funkcja potęgowa-wykładnicza (6)
Pietrzak [2007]	wielkość produkcji	regresja liniowa i wielomianowa (11)
Bieniasz, Czerwińska-Kayzer, Gołaś [2007]	płynność bieżąca, szybkość i natychmiastowa	funkcja potęgowa (2) regresja liniowa (5)

Źródło: zestawienie własne.

Najczęściej stosowana była regresja liniowa, ale znaleźć można też wiele przykładów wykorzystania regresji nieliniowej w postaci funkcji potęgowych, wykładniczych czy wielomianowych. Sieci neuronowe używane są rzadko, chociaż są lepszym narzędziem przynajmniej do wstępnego szacowania modelu. Pozwalają bowiem nie tylko dobrać zmienne niezależne, ale również przez możliwość generowania wykresów odpowiedzi, zobrazować zależności między zmiennymi. Doбору zmiennych objaśniających dokonywali autorzy różnymi metodami.

## WZORCOWA KLASYFIKACJA OBIEKTÓW

Chcąc przeprowadzić wzorcową klasyfikację obiektów należy założyć, że populację można podzielić na części, zwane klasami. Najczęściej rozważane są przypadki dwóch klas, np. firmy upadłe i firmy w dobrej kondycji, klienci banku spłacający kredyty i ci, którzy nie spłacają.

Model klasyfikacyjny jest opracowywany na podstawie zbioru obiektów, o których wiadomo, do jakiej klasy należą (stąd nazwa klasyfikacja wzorcowa). Następnie może on zostać użyty do klasyfikacji obiektów o nieznanym przynależności.

Klasyfikacja wzorcowa, zwana też rozpoznawaniem z nauczycielem, jest w istocie szczególnym przypadkiem regresji. Jest to model regresyjny, w którym zmienna zależna jest cechą jakościową. Najczęściej stosowane metody klasyfikacji wzorcowej to:

- liniowa analiza dyskryminacyjna,
- modele prawdopodobieństwa,
- sieci neuronowe.

W analizie dyskryminacyjnej dążymy do utworzenia kombinacji liniowej zmiennych niezależnych, która najlepiej dyskryminuje (tzn. „rozdziela”) dwie lub więcej grupy określone *a priori*. Model klasyfikacyjny jest liniową kombinacją cech:

$$LFD = \lambda^T \cdot x$$

Opracowanie jego polega na wyznaczeniu wektora parametrów  $\lambda$  i granicznej wartości funkcji dyskryminacyjnej  $f_g$ . Jeżeli LFD ma wartość mniejszą od granicznej, stwierdzamy, że obiekt należy do klasy powiedzmy 0, natomiast w przypadku przeciwnym – do klasy 1. W literaturze [np. Maddala 1994] znaleźć można wzór na wektor współczynników  $\lambda$ , który maksymalizuje relację wariancji międzygrupowej do wariancji wewnątrzgrupowej funkcji dyskryminacyjnej. Wyznaczenie wartości granicznej natomiast wymaga założenia normalności rozkładów cech i równości macierzy wariancji-kowariancji w klasach. Poszukiwana  $f_g$  jest wówczas równa logarytmowi naturalnemu z ilorazu liczebności klas w próbie.

W modelach prawdopodobieństwa szacowane jest prawdopodobieństwo przynależności obiektu do jednej z klas. W zależności od przyjętej postaci funkcji mówimy o modelu liniowym, logitowym i probitowym, których formuły są następujące:

$$P = \beta^T \cdot x, \quad P = \frac{1}{1 + \exp(-\beta^T \cdot x)}, \quad P = \int_{-\infty}^{\beta^T \cdot x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

W modelu liniowym zakładany jest związek liniowy między cechami a szacowanym prawdopodobieństwem, w modelu logitowym jest to funkcja logistyczna, w probitowym zaś dystrybuanta standaryzowanego rozkładu normalnego.

We wszystkich modelach prawdopodobieństwa problem sprowadza się do wyznaczenia wektora parametrów  $\beta$ . Można wykazać, że liniowy model prawdopodobieństwa jest równoważny liniowej funkcji dyskryminacyjnej. Pozostałe modele prawdopodobieństwa są ściśle nieliniowe, co oznacza, że do szacowania ich parametrów należy użyć metod optymalizacyjnych.

Sieci neuronowe można zastosować do zadania klasyfikacji wzorcowej tak jak dla zadania regresyjnego zakładając jedynie dwie wartości zmiennej zależnej 0 i 1. Podejście takie jest analogiczne do liniowej analizy dyskryminacyjnej. Zbudowanie sieciowego modelu prawdopodobieństwa wymaga ograniczenia zakresu zmian jej wyjścia do przedziału  $\langle 0, 1 \rangle$ , co można zrealizować wprowadzając neurony sigmoidalne na wyjściu sieci. Przedstawione w tabeli 2 przykłady wykorzystania wzorcowej klasyfikacji obiektów wskazują na szerokie zastosowa-

Tabela 2. Przykłady zastosowań wzorcowej klasyfikacji obiektów w ekonomice rolnictwa (od 2000 r.)

Autor	Problem	Metoda
Czerwińska-Kajzer [2001]	czynniki wpływające na decyzje inwestycyjne rolników indywidualnych	logitowy model prawdopodobieństwa
Boratyńska [2004]	efekty zmian dokonanych w przedsiębiorstwach piwowarskich przez inwestorów zagranicznych	liniowa analiza dyskryminacyjna
Kisielińska [2004]	prognozowanie kondycji finansowej gospodarstw rolniczych	liniowa analiza dyskryminacyjna i sieci neuronowe
Lemanowicz [2004]	czynniki sukcesu rolników działających w grupach producentów rolnych	logitowy model prawdopodobieństwa
Ryś-Jurek [2004,2005]	ocena sytuacji ekonomicznej indywidualnych gospodarstw rolnych	liniowa analiza dyskryminacyjna i logitowy model prawdopodobieństwa
Daniłowska [2005]	mikroekonomiczne determinanty zaciągania kredytów przez gospodarstw indywidualne	liniowy model prawdopodobieństwa
Siudek [2005]	prognozowanie upadłości banków spółdzielczych w Polsce	liniowa analiza dyskryminacyjna
Domagalska-Grędyś [2006]	prognozowanie upadłości banków spółdzielczych w Polsce	logitowy model prawdopodobieństwa
Kisielińska [2008]	ocena prawdopodobieństwa obniżki kosztów w grupach producenckich	liniowa analiza dyskryminacyjna, sieci neuronowe, probitowy model
Grzegorzewska [2008]	prognozowanie sytuacji finansowej gospodarstw rolniczych	prawdopodobieństwa
Grzegorzewska [2008]	zagrożenie upadłością przedsiębiorstw rolniczych	liniowa analiza dyskryminacyjna i logitowe modele prawdopodobieństwa

Źródło: jak w tab. 1.

nie tych metod klasyfikacji. Obejmują one zagadnienia oceny sytuacji finansowej i upadłości przedsiębiorstw, ekonomiki produkcji, oceny efektów wprowadzonych zmian, wyodrębnienia determinant zaciągania kredytów. Często przy wykorzystaniu tej metody autorzy stosują jedynie liniową analizę dyskryminacyjną. Takie rozwiązanie nie zawsze pozwala właściwie dokonać poprawnej klasyfikacji. Wskazane jest stosowanie więcej niż jednej funkcji, co pozwala lepiej poznać występujące zależności i cechy różnicujące badane obiekty.

## BEWZORCOWA KLASYFIKACJA OBIEKTÓW

Celem klasyfikacji bezwzorcowej, zwanej też analizą skupień, jest połączenie obiektów w klasy (grupy). Podział zbiorowości powinien zostać przeprowadzony w taki sposób, aby podobieństwo obiektów w obrębie klas oraz różnice między obiektami z różnych klas były jak największe. W klasyfikacji bezwzorcowej, w przeciwieństwie do wzorcowej, nie jest znana przynależność obiektów do klas. Klasyfikacja opiera się na odległościach między obiektami.

Metody klasyfikacji wzorcowej podzielić można na następujące grupy:

- hierarchiczne aglomeracyjne,
- hierarchiczne deglomeracyjne,
- obszarowe,
- optymalizujące wstępny podział obiektów,
- sieci neuronowe.

W hierarchicznych metodach aglomeracyjnych wstępnie przyjmuje się liczbę klas równą liczbie obiektów, a następnie łączy się klasy najbardziej do siebie podobne, redukując w każdym kroku liczbę klas o 1, aż do uzyskania jednej klasy obejmującej wszystkie obiekty. Miarą podobieństwa obiektów i klas są odległości między nimi. Różne odmiany metod

aglomeracyjnych różnią się przede wszystkim sposobem wyznaczania odległości między klasami. Popularne metody z tej grupy to metoda:

- pojedynczego wiązania (najbliższego sąsiedztwa) – odległość między dwoma skupieniami jest określona przez odległość między dwoma najbliższymi obiektami należącymi do różnych skupień,
- pełnego wiązania (najdalszego sąsiedztwa) – odległością między skupieniami jest największa z odległości między dwoma dowolnymi obiektami należącymi do różnych skupień,
- Warda – odległością pomiędzy skupieniami jest wartość o jaką zwiększy się wariancja wewnątrzgrupowa po połączeniu grup.

W hierarchicznych metodach deaglomeracyjnych najpierw zakłada się istnienie jednej klasy. W każdym kolejnym kroku liczbę klas zwiększa się o jeden, aż do uzyskania liczby klas równej liczbie obiektów. Różne odmiany metod różnią się sposobem wyboru klasy dzielonej. W klasie tej określa się dwa obiekty leżące najdalej od siebie i na podstawie odległości od nich dokonuje się podziału obiektów pozostałych (przypisując je do najbliższego obiektu wyróżnionego).

W metodach obszarowych wielowymiarowa przestrzeń dzielona jest na rozłączne podobszary. Obiekty znajdujące się w tych obszarach zalicza się do jednej klasy. Stosuje się różne rodzaje podobszarów – mogą to być wielowymiarowe kule czy prostopadłościany.

Metody optymalizujące wstępny podział obiektów startują od pewnego początkowego podziału zbiorowości na klasy. Liczba skupień jest z góry określona, a przydziału do nich dokonać można w sposób losowy (lub przeprowadzić inną metodą). Zadaniem metod optymalizacyjnych jest poprawa przyjętego wstępnego podziału zbiorowości.

Popularną metodą z tej grupy jest metoda k-średnich. W metodzie tej dla każdego skupienia obliczany jest środek ciężkości (zwany centroidem<sup>1</sup>). Następnie obiekty przenosi się do klas o najbliższych środkach ciężkości. Powstają w ten sposób nowe klasy, dla których ponownie oblicza się środki ciężkości. Procedurę kończy się, gdy nie następuje zmiana klas dla obiektów.

Tabela 3. Przykłady zastosowań bezwzorcowej klasyfikacji obiektów w ekonomice rolnictwa (od 2000 r.)

Autor	Problem	Metoda
Borkowski, Szczesny [2002]	przestrzenne zróżnicowanie rolnictwa	metoda pojedynczego wiązania i metoda k-średnich
Kisielińska [2003b] Szczepaniak, Wigier [2003]	zróżnicowanie gospodarstw rolniczych innowacyjność małych i bardzo małych firm przemysłu spożywczego	sieć neuronowa Kohonena metoda k-średnich
Błażejczyk-Majka, Kala [2004]	zasoby siły roboczej rolnictwa polskiego i krajów UE	metoda najdalszego sąsiada
Błażejczyk-Majka, Kala [2005]	charakterystyka użytkowników rolnych wybranych państw Unii Europejskiej	metoda pojedynczego i pełnego wiązania
Majewski [2005] Pocza [2005]	regionalne zróżnicowanie skupu mleka w Polsce poziom i struktura wsparcia finansowego rolnictwa w krajach OECD	metoda Ward'a metoda Ward'a
Adamowicz, Nowak [2006]	typy wiejskich gospodarstw domowych	metoda k-średnich
Osowska [2006]	typologia funkcjonalna obszarów wiejskich Pomorza Środkowego	metoda Ward'a

Źródło: jak w tab. 1.

<sup>1</sup> Centroidem nazywa się wielowymiarową wartość średnią (oczekiwaną) populacji lub rozkładu.

Do wykonania klasyfikacji bezwzorcowej można użyć również sieci neuronowych [Tadeusiewicz 1993, Osowski 2000] – są to tzw. sieci samoorganizujące, których odmianą jest sieć Kohonena. Uczenie jej polega na takim doborze wag neuronów na wyjściu, aby rozpoznawały określony typ wzorców (w naszym przypadku określony typ obiektów).

W przypadku sieci Kohonena, podobnie jak w metodach optymalizujących wstępny podział obiektów, należy założyć z góry liczbę neuronów na wyjściu równą liczbie spodziewanych klas. Zasadnicza różnica funkcjonalna między siecią Kohonena a metodą k-średnich polega jednak na tym, że w sieci niektóre klasy mogą pozostać puste (nie reprezentują żadnego wzorca), natomiast w metodzie k-średnich – liczba klas jest z góry określona. Przykłady zastosowań tej metody w ekonomice rolnictwa przedstawiono w tabeli 3.

Przykłady zastosowań bezwzorcowej klasyfikacji obiektów wskazują, że stosowano ją najczęściej do typologii przestrzennego zróżnicowania rolnictwa i wiejskich gospodarstw domowych. Najczęściej w klasyfikacji stosowano metody aglomeracyjne, w tym głównie metodę Warda oraz metodę k-średnich. Pewne niebezpieczeństwo związane ze stosowaniem analizy skupień wynika z faktu, że w przypadkach wielowymiarowych, gdy brak jest możliwości zobrazowania danych na wykresie, stwierdzenie czy w zbiorowości są skupienia czy też zbiór jest jednorodny jest bardzo utrudnione. Może się więc zdarzyć, że poszukiwane są podgrupy, których w istocie nie ma. Warto więc przeprowadzić test jednorodności dla analizowanego zbioru danych.

## LINIOWE PORZĄDKOWANIE OBIEKTÓW

Celem liniowego porządkowania obiektów jest utworzenie zmiennych syntetycznych, które pozwolą uszeregować obiekty od najlepszego do najgorszego według kryterium reprezentującego pewne zjawisko złożone. Zmienne syntetyczne (ZS) tworzone są jako sumy lub iloczyny przekształconych cech. Transformację należy przeprowadzić tak, aby cechy w jednakowym kierunku wpływały na zmienną syntetyczną, miały ten sam rząd wielkości i nie posiadały miana.

Wyróżniane są trzy rodzaje zmiennych ze względu na kierunek ich oddziaływanie na badane zjawisko złożone:

- stymulanty – zmienne, których wysokie wartości wpływają pozytywnie na badane zjawisko złożone,
- destymulanty – zmienne, których wysokie wartości na badane zjawisko wpływają negatywnie,
- nominanty – zmienne, których określone wartości lub przedziały wartości są korzystne z punktu widzenia badanego zjawiska złożonego.

Warunkiem prawidłowej konstrukcji zmiennej syntetycznej jest przekształcenie wszystkich cech na stymulanty (bądź destymulanty). Można wyróżnić następujące etapy tworzenia zmiennych syntetycznych:

- przekształcenia destymulant i nominant na stymulanty,
- normalizacja cech,
- agregacja cech.

Przekształcenie cech prowadzi się przy użyciu formuł ilorazowych lub różnicowych, o następujących postaciach:



formuły ilorazowe:  $x'_{ij} = \frac{a}{x_{ij}}$  (dla destymulant),  $x'_{ij} = \frac{\min(nom_j; x_{ij})}{\max(nom_j; x_{ij})}$  (dla nominant),

formuły różnicowe:  $x'_{ij} = b - a \cdot x_{ij}$  (dla destymulant),  $x'_{ij} = -|x_{ij} - nom_j|$  (dla nominant).

gdzie:  $a$  i  $b$  – parametry,  $nom_j$  – nominalny poziom  $j$ -tej zmiennej.

Warunkiem poprawnego zastosowania tej metody jest przekształcenie cech w taki sposób, aby rozpatrywane łącznie spełniały warunek porównywalności. Transformację taką nazywa się normalizacją (lub normowaniem zmiennych). Można wyodrębnić kilka grup metod normalizacji [Kukuła 2000, Zeliaś 2002, Gatnar, Walesiak 2004]:

standaryzację:  $x'_{ij} = \frac{x_{ij} - a}{s_j}$ , gdzie  $a = \bar{x}_j$  lub 0,

unitaryzację<sup>2</sup>:  $x'_{ij} = \frac{x_{ij} - a}{\max_i x_{ij} - \min_i x_{ij}}$ , gdzie  $a = \bar{x}_j$ , 0 lub  $\min_i x_{ij}$ ,

przekształcenie ilorazowe:  $x'_{ij} = \frac{x_{ij}}{a}$ , gdzie  $a = \bar{x}_j$ ,  $\min_i x_{ij}$  lub  $\max_i x_{ij}$ ,

użycie metod rangowych:  $x'_{ij} = \frac{l-1}{r-1}$ , gdzie:  $l$  – ranga cechy,  $r$  – ranga obiektu najlepszego.

Cechy znormalizowane mogą być agregowane. Do agregacji stosuje się formuły addytywne i multiplikatywne z wagami lub bez, a także odległości między obiektami, przy czym można użyć różnych miar odległości. Może być to odległość euklidesowa, miejska itp.

Agregacja cech pozwala uzyskać zmienną syntetyczną, która szereguje obiekty od najlepszego do najgorszego. Do porządkowania obiektów można zastosować także metodę Hellwiga [1981], która wymaga uprzedniego przekształcenia destymulant i nominant na stymulanty oraz ich normalizację. Następnie wyróżnia się obiekt wzorcowy i oblicza odległości wszystkich obiektów od niego. Obiektem wzorcowym jest pewien obiekt fikcyjny, którego wszystkie współrzędne przyjmują wartości maksymalne w badanej zbiorowości ( $x_{0j} = \max x_{ij}$ ). Wyznaczone odległości  $d$  (mogą być liczone różne miary odległości) pozwalają obliczyć względne wskaźniki rozwoju obiektu jako:

$$z_i = 1 - \frac{d_i}{d_0}, \text{ gdzie } i \text{ numer obiektu, } d_0 = \bar{d} + 2 \cdot s_d, \quad \bar{d} = \frac{1}{n} \sqrt{\sum_{i=1}^n d_i}, \quad s_d = \frac{1}{n} \sqrt{\sum_{i=1}^n (d_i - \bar{d})^2}$$

Niektóre przykłady zastosowań porządkowania liniowego obiektów w ekonomice rolnictwa przedstawiono w tabeli 4.

Dotychczas porządkowanie liniowe najczęściej było stosowane do tworzenia rankingu różnych jednostek administracyjnych (województwa, powiaty, gminy). Coraz częściej stosuje się ją także do podmiotów gospodarczych. Natomiast spośród metod normalizacji najczęściej wykorzystywano unitaryzację zerowaną, standaryzację, a także przekształcenie ilorazowe.

<sup>2</sup> w metodzie unitaryzacji zerowanej przyjmuje się  $a = \min_i x_{ij}$

Tabela 4. Wybrane przykłady zastosowań porządkowania liniowego w ekonomice rolnictwa (od 2002 r.)

Autor	Problem	Obiekty	Metoda
Borkowski, Szczesny [2002]	przestrzenne zróżnicowanie rolnictwa	województwa	3 rodzaje normalizacji zmiennych – standaryzacja, przekształcenie ilorazowe i unitaryzacja zerowana, ZS jako odległość euklidesowa od wzorca
Kasztelan [2002]	ocena efektywności ekonomicznej przedsiębiorstw	przedsiębiorstwa rolnicze	metoda unitaryzacji zerowanej, ZS jako odległość euklidesowa od wzorca pozytywnego,
Bożek [2002]	przestrzenne zróżnicowanie infrastruktury technicznej gospodarstw rolnych	województwa	metoda unitaryzacji zerowanej i standaryzacja, ZS jako suma zmiennych znormalizowanych lub odległość od antywzorca
Jarka [2003]	ocena głębokości restrukturyzacji przedsiębiorstw przejętych z zasobów AWRSP	przedsiębiorstwa rolnicze	metoda unitaryzacji zerowanej, ZS jest średnią geometryczną obliczoną ze wskaźników cząstkowych równych średniej arytmetycznej z dwóch cech
Puchała [2004]	kierunki i bariery rozwoju gmin w rejonie Małopolski	gminy	metoda unitaryzacji zerowanej
Binderman [2005]	przestrzenne zróżnicowanie potencjału rolnictwa w Polsce	województwa	normalizacja zmiennych – standaryzacja, przekształcenie ilorazowe i unitaryzacja zerowana, ZS jako odległości Minkowskiego od hipotetycznego wzorca
Ossowska [2005]	poziom infrastruktury obszarów wiejskich Pomorza Środkowego	powiaty	metoda Hellwiga
Majchrzak, Wysocki [2007]	potencjał produkcyjny rolnictwa w województwie wielkopolskim	gminy z Wielkopolski	metoda unitaryzacji zerowanej, ZS jako odległość euklidesowa od wzorca najlepszego i najgorszego
Krawiec, Landmesser [2007]	aktywność ekonomiczna ludności na obszarach wiejskich	województwa	normalizacja zmiennych – standaryzacja, przekształcenie ilorazowe i unitaryzacja zerowana, ZS jako średnia arytmetyczna oraz syntetyczny miernik rozwoju Hellwiga obliczony dla odległości miejskiej i euklidesowej
Strojny [2007]	analiza komparatywna produkcji zwierzęcej w krajach UE	państwa	metoda unitaryzacji zerowanej, ZS jako średnia arytmetyczna
Zioło, Jaworska [2007]	zróżnicowanie banków spółdzielczych województwa lubelskiego pod względem wskaźników efektywności	banki działające na terenie Lubelszczyzny	metoda unitaryzacji zerowanej, ZS jako ważona średnia arytmetyczna

Źródło: jak w tab. 1.

Stosując liniowe porządkowanie obiektów należy zwrócić uwagę na wybór metody normalizacji, bowiem ma ona bezpośrednio wpływ na uzyskane rankingi. Należy kierować się własnościami różnych metod oraz celem prowadzonych badań [Zeliaś 2002]. Również agregacja prowadząca do uzyskania zmiennych syntetycznych może być zrealizowana w rozmaity sposób, różnie można także obliczać odległości między obiektami. Warto więc zastosować kilka rozwiązań i porównać uzyskane wyniki.

## ANALIZA CZYNNIKOWA I METODA SKŁADOWYCH GŁÓWNYCH

Analiza czynnikowa i metoda składowych głównych pełnią rolę niejako usługową wobec innych metod wielowymiarowej analizy danych. Celem analizy głównych składowych jest dekompozycja zmienności wielowymiarowego zbioru obserwacji na zbiór składowych (komponentów) w taki sposób, że pierwsza składowa wyjaśnia największą część zmienności, druga z kolei największą część pozostałej zmienności itd. [Aczel 2005].

W analizie czynnikowej stosuje się odmienne podejście. W metodzie tej należy odnaleźć czynniki mające istotne znaczenie w określeniu istoty badanego problemu [Aczel 2005]. Zadaniem obu metod [Gatnar, Walesiak 2004] jest odpowiednia transformacja i ewentualnie redukcja zestawu zmiennych niezależnych. Przekształcone zmienne mogą być następnie wykorzystane w innych metodach. Transformacja ma za zadanie zastąpienie wielu skorelowanych ze sobą niezależnych zmiennych obserwowalnych zmiennymi nieobserwowalnymi, które skorelowane nie będą.

Związki między zmiennymi obserwowalnymi, a zmiennymi nieobserwowalnymi w obydwu metodach określają odpowiednie równania macierzowe.

W metodzie składowych głównych wektor zmiennych obserwowanych  $X$  przedstawiany jest jako iloczyn macierzy współczynników głównych składowych  $B$  i wektora głównych składowych  $S$ , będących poszukiwanymi zmiennymi nieobserwowalnymi:

$$X = B \cdot S$$

W analizie czynnikowej natomiast wektor  $X$  jest iloczynem macierzy ładunków czynnikowych  $A$  i wektora czynników wspólnych  $F$  (zmiennych nieobserwowalnych) powiększony o wektor czynników specyficznych  $U$ .

$$X = A \cdot F + U$$

Wektor zmiennych obserwowalnych  $X$  jest określony przez wektory zmiennych nieobserwowalnych. W metodzie składowych głównych stanowi go wektor głównych składowych  $S$ , w analizie czynnikowej wektor czynników wspólnych  $F$ . W obydwu metodach zadanie polega na wyznaczeniu macierzy przekształcenia, aby nowe zmienne nie były ze sobą skorelowane.

Różnica między metodami polega na tym, że w metodzie składowych głównych zmienne nieobserwowalne w pełni objaśniają zmienne obserwowalne, w analizie czynnikowej natomiast konieczne jest dodatkowo uwzględnienie czynników specyficznych.

Przykłady użycia obydwu omawianych metod przedstawiono w tabeli 5. Zastosowanie metod analizy czynnikowej i głównych składowych w ekonomice rolnictwa jest stosunkowo rzadkie, chociaż ich przydatność jest duża.

Tabela 5. Przykłady zastosowań analizy czynnikowej i metody składowych głównych w ekonomice rolnictwa

Autor	Problem	Obiekty	Metoda
Kołoszko-Chomentowska [2003]	czynniki determinujące dochody z działalności rolniczej	gospodarstwa	składowe główne
Malina [2006]	klasyfikacja regionów w Polsce	województwa	analiza czynnikowa
Siudek [2006]	sytuacja ekonomiczno-finansowa banków spółdzielczych w Polsce	banki spółdzielcze	analiza czynnikowa
Strojny [2006]	poziom rolniczej produkcji roślinnej krajów UE	państwa	składowe główne

Źródło: jak w tab. 1.

## PODSUMOWANIE

Metody wielowymiarowej analizy danych są coraz częściej stosowane w badaniach ekonomicznych. Wynika to z faktu, że problemy spotykane w rzeczywistości gospodarczej rzadko są proste – zwykle do ich opisu konieczna jest więcej niż jedna zmienna. Kluczowym momentem w analizie zjawiska jest dobór cech diagnostycznych. Spośród wielu potencjalnych należy wybrać te, które mają największy wpływ na badane zjawisko oraz dostarczają najistotniejszej informacji. Dobór cech nie powinien być subiektywny, a oparty o kryteria merytoryczne i formalno-statystyczne. Do badania zjawisk mogą być ponadto zastosowane różne metody. Wykorzystując różne metody nie należy poprzestawać na pobieżnej analizie wyników, ale szukać uzasadnienia merytorycznego dla otrzymanych zależności. Ostateczny wynik zależy od badacza, który może posługiwać się różnymi narzędziami, w tym także wielowymiarową analizą porównawczą.

## LITERATURA

- Aczel A.D. 2005: Statystyka w zarządzaniu. Wyd. Naukowe PWN, Warszawa.
- Adamowicz M., Nowak A. 2006: Charakterystyczne typy wiejskich gospodarstw domowych na przykładzie województwa lubelskiego. *Roczniki Naukowe SERiA*, t. VII, z. 4.
- Bieniasz A., Czerwińska-Kajzer D., Gołaś Z. 2007: Czynniki kształtujące płynność finansową przedsiębiorstw branży spożywczej. *Zagadnienia Ekonomiki Rolnej*, nr 4.
- Binderman A. 2005: Klasyfikacja polskich województw według poziomu rozwoju rolnictwa. *Roczniki Nauk Rolniczych*, seria G, t. 92, z. 1.
- Błażejczyk-Majka L., Kala R. 2004: Porównanie zasobów siły roboczej rolnictwa polskiego i krajów UE w latach 1990-1999. *Roczniki Naukowe SERiA*, t. VI, z. 5.
- Błażejczyk-Majka L., Kala R. 2005: Metody analizy skupień do charakterystyki użytków rolnych wybranych państw Unii Europejskiej. *Roczniki Naukowe SERiA*, t. VII, z. 5.
- Boratyńska K. 2004: Efekty zmian dokonanych w przedsiębiorstwach piwowarskich przez inwestorów zagranicznych na przykładzie grupy Żywiec S.A. *Roczniki Naukowe SERiA*, t. VI, z. 2.
- Borkowski B., Szczesny W. 2002: Metody taksonomiczne w badaniach przestrzennego zróżnicowania rolnictwa. *Roczniki Nauk Rolniczych*, seria G, t. 89, z. 2.
- Bożek J. 2002: O niektórych metodach porządkowanie liniowego. *Wiadomości Statystyczne*, nr 9.
- Czekaj T. 2006: Obserwacje odstające i wpływy w analizie regresji – analiza dochodowości materialnych czynników produkcji w gospodarstwach rolnych. *Roczniki Naukowe SERiA*, t. VII, z. 5.
- Czerwińska-Kajzer D. 2001: Czynniki wpływające na decyzje inwestycyjne rolników indywidualnych. *Roczniki Naukowe SERiA*, t. III, z. 5.
- Daniłowska A. 2005: Mikroekonomiczne determinanty zaciągania kredytów przez gospodarstw indywidualne. *Roczniki Naukowe SERiA*, t. VII, z. 4.
- Domagalska-Grędyś M. 2006: Zastosowanie funkcji regresji logistycznej do oceny prawdopodobieństwa obniżki kosztów w grupach producenckich. *Roczniki Naukowe SERiA*, t. VII, z. 5.
- Gatnar E. 1999: Metody wyboru cech w nieparametrycznej analizie dyskryminacyjnej. *Taksonomia 6. Klasyfikacja i analiza danych. Teoria i zastosowania*.
- Gatnar E., Walesiak M. (red.) 2004: Metody statystycznej analizy wielowymiarowej w badaniach marketingowych. Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Gołębiowska B., Klepacki B. 2000: Czynniki kształtujące dochód rolniczy w gospodarstwach o zróżnicowanym poziomie towarowości. *Roczniki Nauk Rolniczych*, seria G, t. 88, z. 2.
- Grabiński T. 1984: Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych. *Zeszyty Naukowe, Monografie*, nr 61. AE w Krakowie, Kraków.
- Grabiński T. 1992: Metody taksonometrii. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Grabiński T., Wydymus S., Zeliaś A. 1982: Metody doboru zmiennych w modelach ekonometrycznych. PWN, Warszawa.

- Grzegorzewska E. 2008: Ocena zagrożenia upadłością przedsiębiorstw w sektorze rolniczym. *Zeszyty Naukowe SGGW. Ekonomika i Organizacja Gospodarki Żywnościowej*, nr 70.
- Hellwig Z. 1981: Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych. PWE, Warszawa.
- Jarka S. 2003: Wykorzystanie metody taksonomicznej do oceny głębokości restrukturyzacji. *Roczniki Nauk Rolniczych*, seria G, t. 90, z. 2.
- Kasztelan P. 2002: Ocena efektywności ekonomicznej przedsiębiorstw z wykorzystaniem metod ilościowych. *Roczniki Nauk Rolniczych*, seria G, t. 89, z. 2.
- Kisielińska J. 2003a: Ocena dochodu rolniczego na podstawie analizy finansowej, przy użyciu sieci neuronowej i analizy regresji. *Roczniki Nauk Rolniczych*, seria G, t. 90, z. 1.
- Kisielińska J. 2003b: Klasyfikacja gospodarstw rolniczych siecią neuronową Kohonena w oparciu o wybrane wskaźniki finansowe. *Zagadnienia Ekonomiki Rolnej*, nr 2.
- Kisielińska J. 2004: Zastosowanie analizy dyskryminacyjnej do prognozowania sytuacji finansowej gospodarstw rolniczych. *Przegląd Statystyczny*, nr 2.
- Kisielińska J. 2008: Modele klasyfikacyjne prognozowania sytuacji finansowej gospodarstw rolniczych. *Rozprawy i Monografie*. Wyd. SGGW, Warszawa.
- Kołoszko-Chomentowska Z. 2003: Zastosowanie składowych głównych w badaniach nad czynnikami determinującymi dochody z działalności rolniczej. *Roczniki Nauk Rolniczych*, seria G, t. 90, z. 2.
- Kowalczyk S. 2002: Model wartości rynkowej gruntów rolniczych. *Roczniki Nauk Rolniczych*, seria G, t. 89, z. 1.
- Krawiec M., Landmesser J. 2007: Analiza taksonomiczna aktywności ekonomicznej ludności na obszarach wiejskich w Polsce. *Roczniki Naukowe SERiA*, t. IX, z. 2.
- Kukuła K. 2000: Metoda unitaryzacji zerowanej. PWN, Warszawa.
- Lemanowicz M. 2004: Czynniki sukcesu rolników działających w grupach producentów rolnych. *Roczniki Naukowe SERiA*, t. VI, z. 1.
- Maddala G.S. 1994: Limited – dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge.
- Majewski J. 2005: Regionalne zróżnicowanie skupu mleka w Polsce oraz czynniki je determinujące. *Roczniki Naukowe SERiA*, t. VII, z. 5.
- Majchrzak A., Wysocki F. 2007: Potencjał produkcyjny rolnictwa w województwie wielkopolskim. *Roczniki Naukowe SERiA*, t. IX, z. 2.
- Malina A. 2006: Analiza czynnikowa jako metoda klasyfikacji regionów w Polsce. *Przegląd Statystyczny*, 1.
- Mańko S., Sass R., Sobczyński T. 2007: Czynniki kształtujące aktywność inwestycyjną rolników w podregionie bydgoskim. *Roczniki Naukowe SERiA*, t. IX, z. 2.
- Osowska L. 2006: Typologia funkcjonalna obszarów wiejskich Pomorza Środkowego. *Roczniki Naukowe SERiA*, t. VII, z. 4.
- Osowski S. 2000: Sieci neuronowe do przetwarzania informacji. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
- Ossowska J. 2005: Poziom infrastruktury obszarów wiejskich Pomorza Środkowego. *Roczniki Naukowe SERiA*, t. VII, z. 4.
- Pawłowska L., Piereligin M. 2005: Czynniki ekonomiczne analizy dochodów przedsiębiorstw przemysłu piekarniczego. *Roczniki Naukowe SERiA*, t. VII, z. 7B.
- Pietrzak M. 2004: Efektywność ekonomiczna spółdzielni mleczarskich – koncepcja pomiaru oraz czynniki wzrostu w świetle badań empirycznych. *Zagadnienia Ekonomiki Rolnej*, nr 3.
- Pietrzak M. 2007: Efektywność finansowa spółdzielni mleczarskich – koncepcja oceny. *Rozprawy i Monografie*, Wyd. SGGW, Warszawa.
- Plonka R., Sobczyński T. 2004: Przydatność wybranych wskaźników finansowych w ocenie zdolności kredytowej gospodarstw rolniczych. *Roczniki Naukowe SERiA*, t. VI, z. 5.
- Pocza A. 2005: Poziom i struktura wsparcia finansowego rolnictwa w krajach OECD po powstaniu WTO. *Roczniki Naukowe SERiA*, t. VII, z. 7A.
- Puchała J. 2004: Kierunki i bariery rozwoju gmin w rejonie Małopolski. *Rocz. Nauk. SERiA*, t. 6, z. 4.
- Rys-Jurek R. 2004: Zastosowanie analizy dyskryminacyjnej do oceny ekonomicznej sytuacji indywidualnych gospodarstw rolnych. *Roczniki Naukowe SERiA*, t. VI, z. 5.

- Siudek T. 2005: Prognozowanie upadłości banków spółdzielczych w Polsce przy użyciu analizy dyskryminacyjnej. *Roczniki Naukowe SERiA*, t. VII, z. 5.
- Siudek T. 2006: Ocena sytuacji ekonomiczno-finansowej banków spółdzielczych w Polsce przy wykorzystaniu wskaźnika syntetycznego. *Roczniki Nauk Rolniczych*, seria G, t. 92, z. 2.
- Stępień S. 2006: Koszty jednostkowe w gospodarstwach o zróżnicowanej strukturze organizacyjnej. *Roczniki Naukowe SERiA*, t. VIII, z. 1.
- Strojny J. 2006: Poziom rolniczej produkcji roślinnej krajów UE. *Więś i Rolnictwo*, nr 4.
- Strojny J. 2007: Analiza komparatywna produkcji zwierzęcej w krajach UE. *Roczniki Naukowe SERiA*, t. IX, z. 1.
- Szczepaniak I., Wigier M. 2003: Identyfikacja czynników wpływających na innowacyjność małych i bardzo małych firm przemysłu spożywczego. *Zagadnienia Ekonomiki Rolnej*, nr 4.
- Szuk T. 2001: Czynniki warunkujące nakłady pracy przy produkcji pszenicy ozimej. *Roczniki Naukowe SERiA*, t. III, z. 5.
- Tadeusiewicz R. 1993: Sieci neuronowe. Akademia Oficyna Wydawnicza RM, Warszawa.
- Witkowska D. 2005: Podstawy ekonometrii i teorii prognozowania. Oficyna Ekonomiczna, Kraków.
- Wysocki F., Kurzawa I. 2006: Kształtowanie się preferencji konsumpcyjnych artykułów żywnościowych w relacji miasto-wieś. *Zagadnienia Ekonomiki Rolnej*, nr 2.
- Zeliaś A. 2002: Uwagi na temat wyboru metody normowania zmiennych diagnostycznych. [W:] Analiza szeregów czasowych na początku XXI wieku, Kufel T., Piłatowska M. (red). Uniwersytet M. Kopernika w Toruniu, Toruń.
- Zioło M., Jaworska M. 2007: Zróżnicowanie banków spółdzielczych województwa lubelskiego pod względem wskaźników efektywności. *Roczniki Naukowe SERiA*, t. IX, z. 3.

*Joanna Kisielińska, Stanisław Stańko*

#### MULTIDIMENSIONAL DATA ANALYSIS IN AGRICULTURAL ECONOMICS

##### *Summary*

The paper presents methods for multidimensional data analysis with examples of their use in agricultural economics. These include multiple regression, the discriminant and cluster analysis, linear organization, factor analysis and principal component method.

Adres do korespondencji:  
dr hab. Joanna Kisielińska  
dr hab. Stanisław Stańko, prof. SGGW  
Szkoła Główna Gospodarstwa Wiejskiego  
Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych  
ul. Nowoursynowska 166  
02-787 Warszawa  
e-mail: joanna\_kisielinska@sggw.pl