

Jadwiga ZARÓD

PODZIAŁ WOJEWÓDZTWA ZACHODNIOPOMORSKIEGO NA REJONY PRZYDATNOŚCI ROLNICZEJ METODĄ K -ŚREDNICH¹

DISTRIBUTION OF ZACHODNIOPOMORSKIE PROVINCE ON AREAS OF AGRICULTURAL USABILITY BY K -MEANS ALGORITHM

Katedra Zastosowań Matematyki w Ekonomii, Zachodniopomorski Uniwersytet Technologiczny
w Szczecinie, 71–270 Szczecin, ul. K. Janickiego 36

Abstract. Natural conditions for rural production in Zachodniopomorskie Province show a wide diversity. Every community was then described by means of variables concerning the quality of soil, water conditions, agroclimate and relief. These data were subject to standardization. Afterwards, by the usage of k -means algorithm, the Province was divided into 10 areas with similar agricultural usability. The purpose of this article is to reduce the number of objects to several basic classes that can be a subject of further analysis. The comparison of production capacity of areas instead of communities would decrease work input and the length of analysis.

Słowa kluczowe: analiza wariancji, metoda k -średnich, rejony województwa zachodniopomorskiego, standaryzacja danych.

Key words: analysis of variance, areas of Zachodniopomorskie Province, data standardization, k -means algorithm,

WSTĘP

Coraz częściej do grupowania obiektów rolniczych wykorzystywane są metody badań ilościowych. Metoda k -średnich należy do metod podziałowych analizy skupień. Pozwala ona na uzyskanie jednorodnych grup o typowych cechach, zredukowanie dużej liczby obiektów do kilku zasadniczych kategorii, podlegających dalszej analizie, oraz na porównanie obiektów wielocechowych.

W województwie zachodniopomorskim warunki przyrodnicze produkcji rolnej wykazują duże zróżnicowanie. Celem tej pracy jest pogrupowanie gmin województwa na kilka rejonów o zbliżonych warunkach glebowych, wodnych, podobnym agroklimacie oraz ukształtowaniu terenu. Porównywanie możliwości produkcyjnych gmin wymagałoby znacznych nakładów pracy i czasu. Metoda k -średnich pozwoli zredukować te nakłady, umożliwiając wyodrębnienie jednorodnych rejonów o najbardziej typowych cechach, przy stosunkowo niewielkich stratach informacji o poszczególnych gminach.

¹ Praca naukowa finansowana ze środków na naukę w latach 2008–2010 jako projekt badawczy.

MATERIAŁ I METODY

Przedmiotem badań były wszystkie gminy województwa zachodniopomorskiego (106), bez podziału na gminy miejskie i wiejskie. Stosując metodę bonitacji punktowej opracowaną przez IUNG (Stuczyński 2000), każdą gminę opisano za pomocą takich czynników środowiskowych, jak: gleba, agroklimat, rzeźba terenu i warunki wodne. Czynniki te zsumowane tworzą ogólny wskaźnik jakości rolniczej przestrzeni produkcyjnej. Następnie dane te poddano standaryzacji zgodnie z wzorem (Młoda 2006):

$$z_{ij} = \frac{x_{ij} - \bar{x}}{s} \quad (1)$$

gdzie:

x_{ij} – wartość i -tego obiektu (gminy) dla j -tej zmiennej,

\bar{x} – wartość średnia danej zmiennej,

s – odchylenie standardowe.

W wyniku standaryzacji przekształcone wartości mają wariancję równą 1 i średnią arytmetyczną równą 0. Ta transformacja umożliwia porównywanie rozkładu wartości dla wielu zmiennych i grup.

Do grupowania gmin w rejony wykorzystano metodę k -średnich. Istota tej metody polega na minimalizowaniu wielkości wewnątrzgrupowej wariancji lub maksymalizowaniu wartości międzygrupowej wariancji. Algorytm postępowania (Hartigan 1979):

1. Ustalić maksymalną liczbę literacji oraz liczbę grup k , na jakie ma być podzielony analizowany zbiór obiektów, przy czym $k \in (2, n - 1)$, gdzie n – liczba obiektów.

2. Ustalić wyjściową macierz środków ciężkości grup

$$B(b_{sj}) \quad s = 1, 2, \dots, k \quad j = 1, 2, \dots, l$$

gdzie:

l – liczba zmiennych oraz przyporządkowanie obiektów do takich grup, dla których ich odległość od środka ciężkości danej grupy jest najmniejsza.

3. Wyznaczyć wartość wyjściową błędu podziału obiektów między k -grupami

$$\varepsilon = \sum_{i=1}^n d_{is}^2 \quad (2)$$

gdzie:

d_{is}^2 – odległość Euklidesa między i -tym obiektem a najbliższym s -tym środkiem ciężkości:

$$d_{is}^2 = \sum_{i=1}^n (x_{ij} - b)^2 \quad i = 1, 2, \dots, n \quad (3)$$

4. Dla pierwszego obiektu określić zmiany błędu podziału wynikające z przyporządkowania go kolejno do wszystkich aktualnie występujących grup

$$\Delta \varepsilon_s^1 = \frac{n_s d_{1s}^2}{n_s + 1} - \frac{n_{s_1} d_{1s_1}^2}{n_{s_1} + 1} \quad (4)$$

gdzie:

n_s – liczebność s -tej grupy,

d_{1s}^2 – odległość pierwszego obiektu od środka ciężkości s -tej grupy,

n_{s_1} – liczebność grupy zawierającej pierwszy obiekt,

$d_{1s_1}^2$ – odległość pierwszego obiektu od najbliższego środka ciężkości.

Jeżeli minimalna wartość wyrażenia (4) dla wszystkich $s \neq s_1$ jest ujemna, to pierwszy obiekt przypisuje się do grupy, dla której $\Delta \varepsilon_s^1 = \min$. Następnie należy obliczyć środki ciężkości grupy B, uwzględniając dokonaną transformację obiektu oraz wyznaczyć aktualną wartość błędu podziału (2). Jeżeli minimalna wartość wyrażenia (4) jest dodatnia lub równa zero, to nie dokonuje się żadnych zmian.

5. Operacje z punktu 4 powtarza się dla każdego obiektu, co kończy pierwszą literację procedury.

6. Jeżeli w danej literacji nie występują żadne przesunięcia obiektów z grupy do grupy, to postępowanie się kończy. W przeciwnym wypadku należy rozpocząć następną literację aż do momentu, w którym liczba literacji nie przekroczy ustalonej wartości.

WYNIKI I DISKUSJA

Metoda k -średnich, przenosząc obiekty (gminy) pomiędzy skupieniami (rejonami), dąży do minimalizacji zmienności wewnątrz skupień i maksymalizacji zmienności między skupieniami (Kisielińska 2004). Tabela 1 podaje analizę wariancji zmiennych opisujących przydatność rolniczą wszystkich gmin.

Tabela 1. Analiza wariancji

Zmienne	Zmienność między skupieniami	Zmienność wewnątrz skupień	Statystyka F
Gleba	90,536	14,464	66,768
Agroklimat	83,883	21,117	42,370
Rzeźba terenu	77,793	27,207	30,499
Warunki wodne	88,436	16,564	56,948
Ogólny wskaźnik jakości rolniczej przestrzeni produkcyjnej	91,213	13,787	70,568

Wysoka zmienność między rejonami i niska wewnątrz rejonów świadczy o prawidłowym pogrupowaniu gmin ze względu na analizowane zmienne. Wielkość statystyki F wskazuje, w jakim stopniu dany wymiar dyskryminuje skupienie. Przydatność rolnicza rejonów w najmniejszym stopniu zależała od rzeźby terenu oraz agroklimatu.

Gminy województwa zachodniopomorskiego zostały podzielone na 10 rejonów. Liczbę skupień ustalono na podstawie wzorów: $k = \sqrt{n}$ i $k \approx 1 + 3,322 \log n$, gdzie $n = 106$ (liczba gmin).

Przypisanie gmin do rejonów oraz odległość od środka skupienia przedstawia tab. 2.

Tabela 2. Podział województwa zachodniopomorskiego na rejon przydatności rolniczej

Rejony	Gminy	Odległość od środka skupienia
I	Bierzwnik	0,258
	Brzeźno	0,517
	Chociwel	0,268
	Dobra	0,417
	Dobra Szczecińska	0,509
	Dobrzany	0,427
	Drawno	0,246
	Dziwnów	0,638
	Łobez	0,313

cd. tab. 2

	Maszewo	0,214
	Nowogard	0,262
	Nowogródek	0,321
	Płoty	0,477
	Radowo Małe	0,349
	Recz	0,220
	Resko	0,257
	Szczecin	0,527
	Świerzno	0,404
	Wałcz	0,656
	Wolin	0,452
II	Białogard	0,314
	Kołobrzeg	0,505
	Manowo	0,521
	Sianów	0,288
	Siemyśl	0,414
	Świeszyno	0,277
	Tychowo	0,429
III	Barwice	0,238
	Bobolice	0,638
	Borne Sulinowo	0,606
	Czaplinek	0,400
	Drawsko Pomorskie	0,486
	Grzmiąca	0,303
	Malechowo	0,752
	Ostrowice	0,311
	Polanów	0,667
	Połczyn	0,727
	Rąbino	0,324
	Rymań	0,384
	Świdwin	0,471
	Złocieniec	0,347
IV	Barlinek	0,093
	Boleszkowice	0,555
	Cedynia	0,595
	Chojna	0,477
	Choszczno	0,247
	Dębno	0,309
	Dolice	0,519
	Gryfino	0,309
	Krzęcin	0,177
	Lipiany	0,477
	Mieszkowice	0,295
	Moryń	0,544
	Myślibórz	0,438
	Pelczyce	0,411
	Stara Dąbrowa	0,390
	Stare Czarnowo	0,309
	Stargard	0,443
	Suchań	0,254
	Trzcińsko Zdrój	0,327
	Węgorzyno	0,641
	Widuchowa	0,429
V	Golczewo	0,373
	Gościno	0,475
	Gryfice	0,125
	Kamień Pomorski	0,354
	Karnice	0,263
	Rewal	0,288
	Trzebiatów	0,249
VI	Będzino	0,280
	Biesiekierz	0,268

cd. tab. 2

	Brojce	0,243
	Dygowo	0,186
	Karlino	0,211
	Koszalin	0,523
	Ustronie	0,125
VII	Goleniów	0,143
	Kobylanka	0,351
	Marianowo	0,439
	Międzyzdroje	0,331
	Nowe Warpno	0,589
	Osina	0,336
	Police	0,319
	Przybiernów	0,377
	Stepnica	0,304
	Świnoujście	0,465
VIII	Banie	0,442
	Bielice	0,411
	Kołbaskowo	0,278
	Kozielice	0,155
	Przelewice	0,293
	Pyrzyce	0,205
	Warnice	0,925
IX	Darłowo	0,112
	Mielno	0,350
	Postomino	0,210
	Sławno	0,365
X	Biały Dwór	1,046
	Człopa	0,636
	Ińsko	0,312
	Kalisz Pomorski	0,401
	Mirosławiec	0,334
	Sławoborze	0,361
	Szczecinek	0,412
	Tuczno	0,418
	Wierzchowo	0,316

Gminy przydzielone do tych samych rejonów sąsiadują ze sobą i charakteryzują się zbliżonymi warunkami naturalnymi (jakość gleby, klimat, nawodnienie, rzeźba terenu) produkcji rolniczej.

WNIOSKI

1. Metoda *k*-średnich pozwala zredukować dużą liczbę gmin do kilku podstawowych rejonów, które mogą być traktowane jako przedmiot dalszej analizy.

2. Porównanie możliwości produkcyjnych rejonów, a nie gmin zmniejszy nakłady pracy i skróci czas analiz.

3. Zastosowana metoda daje możliwość porównywania obiektów wielocechowych, bowiem każda gmina została opisana za pomocą zmiennych dotyczących jakości gleb, agroklimatu, rzeźby terenu, warunków wodnych i ogólnego wskaźnika jakości rolniczej przestrzeni produkcyjnej.

4. Standaryzacja danych wejściowych o gminach czyni wyniki analiz statystycznych całkowicie niezależnymi od jednostek pomiaru poszczególnych zmiennych.

5. Metodę *k*-średnich można wykorzystać jako narzędzie wspomagające podział województwa zachodniopomorskiego na rejony o różnych możliwościach produkcji rolnej.

PIŚMIENNICTWO

Hartigan J.A., Wong M.A. 1979. A K-Means Clustering Algorithm. *App Stat.* 28 (1), 100–108.

Kisielińska J. 2004. Porównanie bezwzorcowej klasyfikacji siecią neuronową Kohonena i metodą k-średnich 1001 gospodarstw rolniczych w oparciu o wybrane wskaźniki finansowe [w: *Metody ilościowe w badaniach ekonomicznych*]. Warszawa, Wydaw. SGGW, 37–46.

Młodak A. 2006. Analiza taksonomiczna w statystyce regionalnej. Warszawa, Centrum Doradztwa i Informacji Difin, 118–127.

Stuczyński T. 2000. Waloryzacja rolniczej przestrzeni produkcyjnej Polski. *Biul. Infor. IUNG Puławy* 12.