

Andrzej Woźniak, Jakub Sikora

**WYKORZYSTANIE SKALOWANIA
WIELOWYMIAROWEGO W ANALIZIE
POTENCJAŁU INFRASTRUKTURALNEGO
GOSPODARSTW ROLNYCH W WYBRANYCH
GMINACH WOJEWÓDZTWA MAŁOPOLSKIEGO**

Streszczenie

Wyznaczenie potencjału infrastrukturalnego gminy możliwe jest przez zmierzenie wielu cech diagnostycznych opisujących ten potencjał w gminie – liczba tych zmiennych zależy głównie od celu analizy. W tym opracowaniu do zobrazowania zróżnicowania infrastruktury rolniczej gospodarstw rolnych na poziomie gmin przyjęto 13 zmiennych uznanych za diagnostyczne. Zestaw tych zmiennych wyznacza w przestrzeni wielowymiarowej punkt charakterystyczny dla każdego badanego obiektu i różnicujący obiekty między sobą. Oceny potencjału dokonano, opierając się na syntetycznej mierze rozwoju będącej agregatem cech diagnostycznych. Obliczony wskaźnik syntetyczny redukuje wielowymiarowe dane do jednej liczby i zastępuje punkt w przestrzeni wielowymiarowej punktem na osi w przestrzeni jednowymiarowej.

Celem artykułu jest wykazanie czy obraz zbiorowości gmin w ich wielu wymiarach i oryginalnym kształcie widziany przez pryzmat współczynnika syntetycznego, jest podobny do oryginału. Czy otrzymana odległość jednowymiarowa dobrze odzwierciedla odległości w pierwotnej przestrzeni badanego potencjału infrastruktury?

W pracy do analizy jakości informacyjnej, jaką niesie współczynnik syntetyczny obliczony dla badanych gmin wykorzystano skalowanie wielowymiarowe. Metoda obliczeniowa skalowania wymiarowego polega na minimalizacji funkcji zwanej funkcją stresu (w polskiej terminologii używa się też nazwy naprężenie) lub funkcji nieco zmodyfikowanej stres standardowy, tzw. współczynnik alienacji, gdzie d_{ij} są utworzonymi od-

ległościami przy danej liczbie współrzędnych w przestrzeni skalowania. W metodzie skalowania nie nakłada się żadnych ograniczeń na postać wyrażenia definiującego odległość. Aby ułatwić interpretację i ograniczyć objętość artykułu przyjęto dwa wymiary przestrzeni skalowania

Do oceny jakości skalowania wykorzystano wykres Sheparda.

Interpretacji współrzędnych dokonano na podstawie przeprowadzonej analizy korelacji współrzędnych przestrzeni skalowania ze zmiennymi wejściowymi.

Badaniami objęto 178 gmin województwa małopolskiego, dla których można było pozyskać dane statystyczne i oddzielając infrastrukturę rolniczą od infrastruktury komunalnej badanej gminy.

Słowa kluczowe: infrastruktura rolnicza gospodarstw rolnych, syntetyczna miara rozwoju, skalowanie wielowymiarowe, województwo małopolskie

WPROWADZENIE

Rola i znaczenie klasyfikacji jako czynności badawczych są ściśle związane z systematyką, biologiczną, na której gruncie i regułach metod statystycznych powstała w ostatnich latach odrębna dyscyplina – taksonomia numeryczna, zajmująca się teoretycznymi zasadami, procedurami i regułami klasyfikowania obiektów w ujęciu numerycznym [Woźniak 2001]. Taksonomia jest dziedziną statystycznej analizy wielowymiarowej, która zajmuje się teoretycznymi zasadami i regułami klasyfikacji obiektów wielowymiarowych. Taksonometria jest to dziedzina wiedzy zajmująca się pomiarem, ilościowym opisem i analizą związków strukturalnych występujących wśród elementów zbiorów wielocechowych obiektów. Punktem wyjścia w badaniach taksonomicznych jest ustalenie przedmiotu klasyfikacji i przestrzeni klasyfikacji. Przedmiot klasyfikacji definiuje się jako policzalny zbiór elementów w wielowymiarowej przestrzeni klasyfikacji dowolnej natury. Przestrzeń klasyfikacji określa się jako zbiór własności charakteryzujących elementy zbioru wielowymiarowej przestrzeni klasyfikacji.

Jeżeli zbadamy wiele gmin to zbiór punktów w wielowymiarowej przestrzeni zastępujemy zbiorem punktów w przestrzeni o jednym wymiarze. Mamy tu zatem do czynienia z odwzorowaniem homomorficznym. Pojawia się pytanie dotyczące własności tego odwzorowania i jego jakości. Jest to zagadnienie teorii pomiaru i teorii modelowania. Statystyki pozwalają na sposoby analizy wymiarowości przestrzeni konfiguracji i redukcji jej wymiarów. Sposoby te opierają się albo

na tradycyjnych technikach analizy składowych macierzy kowariancji i analizie czynnikowej, albo na skalowaniu wielowymiarowym, gdzie redukcja liczby wymiarów odbywa się na podstawie macierzy odległości między punktami przestrzeni konfiguracyjnej, a nie macierzy kowariancji.

Istnieją metody związane z wyszukiwaniem schematów zachowań danych w przestrzeni wielowymiarowej i poszukiwanie ukrytych wymiarów (analiza czynnikowa), określane ogólnie mianem teorii równań strukturalnych. Metody te pozwalają na tworzenie i badanie modeli z wykorzystaniem ukrytych zmiennych i oceną istotności ich wpływu na opis rzeczywistości za pomocą modelu [Dąbkowski 1998].

W artykule zastosowano metodę skalowania wielowymiarowego, która pozwoli na ocenę struktury przestrzeni konfiguracyjnej i możliwości redukcji liczby wymiarów tej przestrzeni. Jest to metoda opisu populacji za pomocą mniejszej liczby wymiarów, która pierwotnie została zmierzona.

TEORETYCZNE PODSTAWY METODY

Skalowanie wielowymiarowe redukuje liczbę wymiarów na zasadzie dopasowania konfiguracji odległości punktów pomiarowych w przestrzeni wyników do przestrzeni zredukowanej. Podejście to prowadzi do zmniejszenia liczby zmiennych opisujących badaną populację. Analizy oparte na macierzach kowariancji lub korelacji pozwalają na przeprowadzenie klasycznych testów istotności, ale za to wymagają założenia o normalności rozkładu wielowymiarowego. Skalowanie wielowymiarowe nie wymaga żadnych założeń o rozkładach, pozwala na stosowanie dowolnej miary odległości i daje mniejszą liczbę wymiarów [Dąbkowski 1998].

Konieczne obliczenia przeprowadzono za pomocą pakietu STATISTICA 6.0. Wykonywanie w tym pakiecie procedury skalowania wielowymiarowego cechuje się ograniczoną liczbą punktu w przestrzeni skalowania. Możliwa największa liczba punktów wynosi 90. W tym celu zbiór obiektów został podzielony na cztery grupy pod względem współczynnika syntetycznego, analiza przeprowadzana była oddzielnie dla poszczególnych grup.

Metodą redukcji liczby wymiarów jest przybliżenie wyznaczonej doświadczalnie macierzy odległości między obiektami za pomocą mniejszej liczby wymiarów. W opracowaniu dysponujemy 13 zmiennymi obrazującymi potencjał infrastruktury rolniczej gospodarstw

wiejskich, a liczebność grupy pierwszej (obszary o niskich wartościach miary rozwoju) wynosi 67, grupy drugiej (obszary o średnich wartościach miary rozwoju) wynosi 77, trzeciej (obszary o wysokich wartościach miary rozwoju) 19, zaś w grupie czwartej (obszary o bardzo wysokich wartościach miary rozwoju) wynosi 13. Z tych zbiorów danych tworzymy macierze odległości przypadków, które mają wymiary macierzy kwadratowych, i tak kolejno do grupy pierwszej jest to macierz o wymiarach 67 x 67, grupy drugiej 77 x 77, grupy trzeciej 19 x 19 i grupy czwartej 13 x 13. Do przeprowadzenia analizy została wybrana zwykła odległość euklidesowa.

Wyliczenie odległości za pomocą metryki Minkowskiego. W taksonometrii przyjmuje się, że mierniki odległości powinny być określone przez funkcje metryczne [Woźniak 2001].

$$d_{ik} = \left[\sum_{j=1}^m |x_{ij} - x_{kj}|^p \right]^{\frac{1}{p}} \quad (i, k = 1, \dots, n), \quad (1)$$

gdzie:

- x_{ij}, x_{kj} – realizacja j -tej cechy dla i -tego oraz k -tego obiektu,
- n – liczba obiektów,
- m – liczba cech,
- p – liczba naturalna:

w praktyce wykorzystuje się odległość miejską ($p = 1$), odległość Euklidesa ($p = 2$), odległość Czybyszewa ($p = \infty$).

W metodzie skalowania nie nakłada się żadnych ograniczeń na postać wyrażeń definiujących odległości z ograniczeniem, aby spełniały one warunek bycia odległością, tzn. $d_{ik} = d_{kj} \geq 0$, nierówności trójkąta. W szczególności można stosować miarę związaną z macierzą korelacji $d_{ik} = 1 - r_{jk}$ [Dąbkowski 1998].

Metoda obliczeniowa skalowania wielowymiarowego polega na minimalizacji funkcji:

$$S_{(d_{ij})} = \sum (d_{ij} - f(\delta_{ij}))^2 \quad (2)$$

zwanej funkcją stresu.

Gdzie:

d_{ij} – otworzone odległości przy danej liczbie współrzędnych w przestrzeni skalowania,

$f(\delta_{ij})$ – monotonicznymi funkcjami odległości wejściowych.

Zmienne, które analizujemy, obrazują potencjał infrastruktury gospodarstw wiejskich, przedstawiają bardzo różnorodne wielkości, wyrażające się w różnych jednostkach i są zróżnicowane pod względem wartości. Aby pozbyć się zjawiska dominacji w odległościach zmiennych o dużych wartościach zastosowano standaryzację zmiennych, aby pozbyć się mian i wielkości.

Mnogość metod standaryzacji oraz istniejących formuł normalizacji cech diagnostycznych w wielu wypadkach może nastęrczać problemów z wyborem najwłaściwszej. Niektórzy autorzy, np. Borys [1978] i Grabiński [1984], sugerują, aby wybór formuły normalizacji połączyć z wyborem formuły agregacji. Najlepsze właściwości formalne ma metoda unitaryzacji, w której przez rozstęp dzielona jest odległość danej wartości od zaobserwowanej wartości najgorszej (minimalnej). Metoda ta spełnia warunek ograniczenia zakresu cechy do domkniętego przedziału $\langle 0, 1 \rangle$, jak również postulowany warunek nieujemności cechy znormalizowanej [Woźniak 2001].

Normalizacja cech diagnostycznych przebiega więc według wzoru:

$$Z_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}} \quad (3)$$

gdzie:

- i – numer obiektu,
- j – numer cechy diagnostycznej,
- x_{ij} – realizacja cechy j w obiekcie i .

Powyższa formuła dotyczy zmiennych o charakterze stymulanty, tzn. takich, których „duże” wartości mają duże znaczenie dla badanych procesów rozwoju. Z formalnego punktu widzenia każdą zmienną o charakterze stymulanty można przekształcić w destymulantę i odwrotnie, pod warunkiem zachowania sensu merytorycznego zmiennej dokładnie zostało to opisane w artykule [Woźniak, Sikora 2005].

SYNTETYCZNA MIARA ROZWOJU JAKO KRYTERIUM PORZĄDKOWANIA LINIOWEGO. POMIAR METRYKI ODLEGŁOŚCI

Kluczowym zagadnieniem dla rozwiązania postawionego problemu badawczego jest dobór cech diagnostycznych i sposób ich agregacji. W pracach poświęconych miarom syntetycznym można

wielokrotnie napotkać postulat niezależności zmiennych używanych ostatecznie do tworzenia wskaźnika syntetycznego. Z drugiej jednak strony zasady sprawozdawczości statystycznej podają ograniczony zestaw danych statystycznych w skali gmin. Z uwagi więc na koszty badań należy poszukiwać rozwiązań pośrednich, to znaczy poszukiwać kompromisu pomiędzy podejściem merytorycznym i statystycznym. Sam przedmiot badań – infrastruktura obszarów wiejskich – ogranicza kryterium formalne na rzecz merytorycznego doboru cech diagnostycznych do zestawu prezentowanych w statystyce publicznej wskaźników charakteryzujących infrastrukturę techniczną gmin.

Ostatecznie, po wstępnej weryfikacji wartości informacyjnej cech, z dostępnego zestawu cech diagnostycznych przyjęto do analizy 13 wskaźniki wyposażenia gospodarstw w infrastrukturę techniczną (tab. 1).

Cechy zostały wybrane z zestawu tak, by reprezentująca je lista zmiennych odznaczała się następującymi własnościami:

- zmienne były słabo skorelowane między sobą,
- były silnie korelowane ze zmiennymi odrzuconym.

Tabela 1. Wybrane spośród infrastruktury technicznej gospodarstw zmienne diagnostyczne

X_j	Cecha diagnostyczna	Wymiar cechy
X_1	liczba samochodów ciężarowych na 1 ha UR	[szt./ha]
X_2	liczba ciągników 1 ha UR	[szt./ha]
X_3	liczba kombajnów zbożowych 100 ha UR	[szt./100 ha]
X_4	liczba kombajnów ziemniaczanych na 100 ha UR	[szt./100 ha]
X_5	liczba kombajnów buraczanych na 100 ha UR	[szt./100 ha]
X_6	powierzchnia obór ogółem na 1 ha UR	[m ² /ha]
X_7	powierzchnia innych pomieszczeń na 1 ha UR	[m ² /ha]
X_8	powierzchnia chlewni ogółem na 1 ha UR	[m ² /ha]
X_9	powierzchnia kurników ogółem na 1 ha UR	[m ² /ha]
X_{10}	powierzchnia stodół ogółem na 1 ha UR	[m ² /ha]
X_{11}	powierzchnia wiat ogółem na 1 ha UR	[m ² /ha]
X_{12}	powierzchnia garaży ogółem na 1 ha UR	[m ² /ha]
X_{13}	powierzchnia budynków wielofunkcyjnych na 1 ha UR	[m ² /ha]

Przyjęty zestaw zmiennych zweryfikowany został pod względem dostatecznej zmienności przestrzennej jako potencjalnej wartości informacyjnej każdej z cech.

Za podstawę obliczeń przyjęto współczynnik zmienności:

$$V_j = \frac{S_j}{\bar{x}_j} \quad (j = 1, 2, \dots, k) \quad (4)$$

gdzie:

\bar{x} – średnia arytmetyczna badanej cechy,
 S_j – odchylenie standardowe.

Cechy, dla których jest spełniona nierówność: $V_j \leq \varepsilon$ zostały wyeliminowane, gdzie ε jest arbitralnie przyjętą wartością progową współczynnika zmienności ustaloną na poziomie $\varepsilon = 0,1$.

Spośród wstępnie wybranych 198 gmin województwa małopolskiego o różnych typach: wiejskich, miejsko-wiejskich i miejskich, wytypowano gminy, dla których statystyki publiczne zawierały dane w formie umożliwiającej oddzielenie infrastruktury technicznej gospodarstw wiejskich od urządzeń infrastrukturalnych miast i części miejskich gmin miejsko-wiejskich.

Dalszej analizie poddano 176 gmin, reprezentujących wyłącznie obszary wiejskie. Uporządkowany liniowo, w zakresie $\langle 0, 1 \rangle$ zbiór miar syntetycznych dla wybranych obiektów przyjmowały wartość z przedziału od 0,27 dla gmin najlepszych do 0 dla najgorszej dokładnie zostało to opisane w artykule [Woźniak, Sikora 2005].

Tabela 2. Charakterystyka klasyfikacji obiektów z uwagi na wartość syntetycznej miary rozwoju

Grupa	Charakterystyka grupy	Zakres zmienności grupy
I	obszary o niskich wartościach miary rozwoju	$0 \leq q_i \leq \min \{q_i\} + 0,2R$
II	obszary o średnich wartościach miary rozwoju	$\min \{q_i\} + 0,2R \leq q_i \leq \min \{q_i\} + 0,4R$
III	obszary o wysokich wartościach miary rozwoju	$\min \{q_i\} + 0,4R \leq q_i \leq \min \{q_i\} + 0,6R$
IV	obszary o bardzo wysokich wartościach miary rozwoju	$\min \{q_i\} + 0,6R \leq q_i \leq \max$

q_i – wartość miary syntetycznej ustalona dla i -tego obiektu

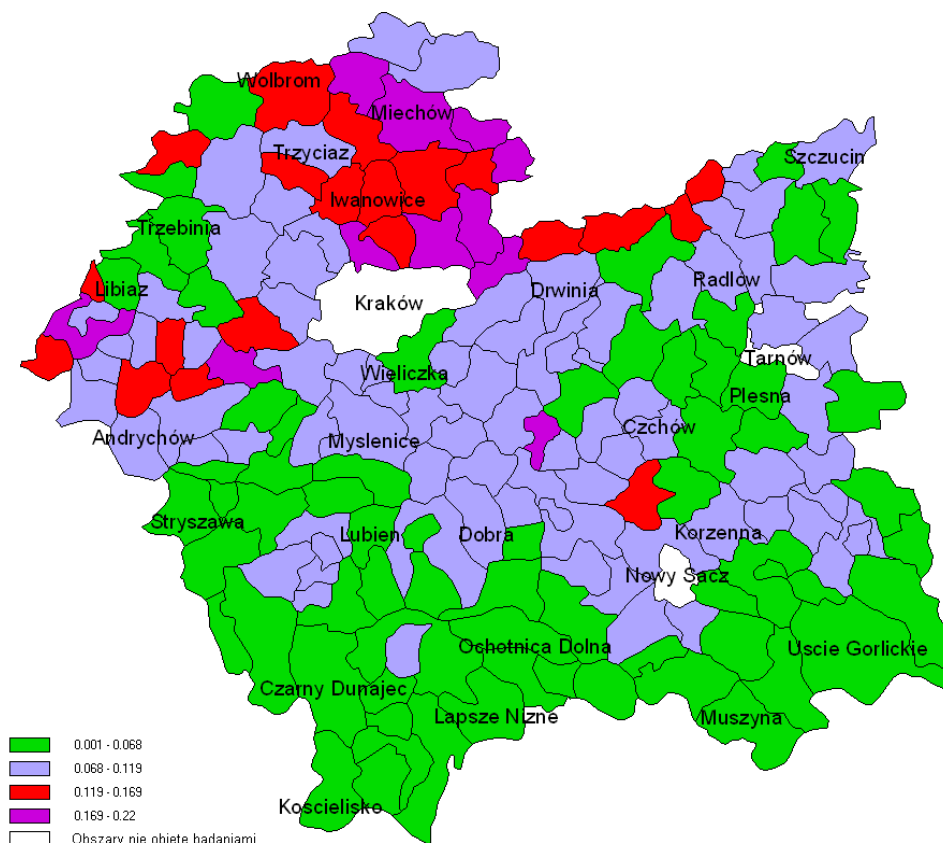
R – rozstęp wartości syntetycznej miary rozwoju

WYNIKI BADAŃ I WNIOSKI

Źródłem materiałów statystycznych do wyznaczania wskaźnika nasycenia gmin infrastrukturą techniczną gospodarstw stanowiły wyniki powszechnego spisu rolnego 2002 dla województwa małopolskiego.

We wstępnej fazie badań analizą objęto wszystkie gminy województwa małopolskiego, po weryfikacji danych przyjęto 176 gmin województwa małopolskiego. Z uzyskanego zbioru wyznaczono syntetyczną miarę rozwoju według cytowanej metodyki [Woźniak, Sikora 2005].

Obiekty badań (gminy) zostały podzielone na grupy zgodnie z klasyfikacją przedstawioną w tabeli 2. W grupie o niskich wartościach miary rozwoju z przedziału od 0,01 do 0,068 znalazło się 67 obiektów, które zajmują obszary południowe województwa małopolskiego, powierzchnia tych gmin wynosi 662 974 ha i stanowi 43,65% całości powierzchni województwa. Grupa o średnich wartościach miary rozwoju zawiera się w przedziale od 0,068 do 0,119 współczynnika syntetycznego, jest to najliczniejsza grupa reprezentowana przez 77 gmin, które zajmują 573 019 ha i stanowią 37,72% całego obszaru województwa. Następne grupy są mniej liczne i ich obiekty pochodzą z obszarów północnych i zachodnich województwa. Grupa o wysokich wartościach miary rozwoju z przedziału od 0,119 do 0,169, w tej grupie wystąpiło 19 gmin o łącznej powierzchni 131 300 ha, co stanowiło 8,64% całości powierzchni województwa. Grupa czwarta jest najmniej liczna o bardzo wysokich wartościach miary rozwoju z przedziału od 0,169 do 0,22, w tej grupie wystąpiło 13 gmin o łącznej powierzchni 98 121 ha, co stanowi 6,4% powierzchni województwa. Przestrzenne zróżnicowanie i rozmieszczenia gmin podzielonych na grupy obrazuje rysunek 1, z przedstawionej analizy uwidacznia się, gdzie położone gminy o najmniejszych wartościach współczynnika syntetycznego pokrywają się z obszarem o największym rozproszeniu rolnictwa w województwie małopolskim. Obiekty o bardzo wysokich wartościach miary rozwoju są rozmieszczone blisko dużych aglomeracji, tj. Kraków, Oświęcim i Tarnów. Gminy o bardzo wysokim potencjale infrastruktury rolniczej są z terenów uznawanych za bardzo dobrze rozwijające się rolniczo: są to gminy z powiatów miechowskiego, proszowickiego, bocheńskiego, oświęcimskiego i części krakowskiego.



Rysunek 1. Przestrzenne rozmieszczenie wydzielonych grup według wskaźnika syntetycznego infrastruktury rolniczej gospodarstw gmin województwa małopolskiego

Obliczenia skalowania wielowymiarowego polega na minimalizacji funkcji zwanej funkcją stresu w polskiej terminologii używa się też nazwy naprężenia (2). W funkcji stresu d_{ij} stanowią odtworzone odległości przy danej liczbie współrzędnych w przestrzeni skalowania, natomiast $f(d_{ij})$ monotonicznymi funkcjami odległości wejściowych. Program STATISTICA oblicza dwie pomocnicze macierze odległości, które pomagają w ocenie jakości (dobroci) dopasowania (tzw. D^* i D^\wedge).

Dobór ilości wymiarów w artykule został przyjęty przez autorów, najlepiej interpretuje się małą liczbę wymiarów od dwóch do trzech. W artykule została ona przyjęta jako przestrzeń dwuwymiarowa.

Wtedy otrzymane współrzędne w przestrzeni skalowania są stosunkowo łatwe do interpretacji. Graficzne ilustracje jakości dopasowania stanowią wykresy:

– wykres Sheparda wykres odległości dopasowanych względem rzeczywistych,

– wykres D^{\wedge} w zależności od odległości rzeczywistych δ_{ij} .

Jeżeli dopasowanie jest dobre, punkty rozrzutu na wykresie układają się blisko linii prostej $y = x$. Zwłaszcza wykres Sheparda uważa się za miarodajną ocenę jakości skalowania. Interpretacja współrzędnych w skalowaniu wielowymiarowym jest zadaniem trudnym i nie można podać jednoznacznej procedury jej przeprowadzania. Wybrana ilość współrzędnych (X ; Y) w przestrzeni skalowania, jakimi dysponujemy, zastępują nam współrzędne w przestrzeni skalowania, jakimi dysponujemy, zastępują cały potencjał infrastruktury rolniczej badanych obiektów. Mamy tu jednak do czynienia z dowolnością, kierunek osi układu współrzędnych wybrany jest w pewnym sensie dowolnie w zależności od algorytmu obliczania. Stosowany w pakiecie STATISTICA algorytm obliczania pozwala na wybór konfiguracji początkowej. Program automatycznie przyjmuje konfigurację Guttmana-Lingoesa, która opiera się na analizie głównych składowych i oś pierwszej współrzędnej przyjmuje jako kierunek największej wariancji. To znaczy, że największa wariancja będzie odpowiadać kierunkowi największej odległości pomiędzy obiektami, ponieważ analiza przeprowadzona w tej pracy została przeprowadzona w układzie przypadków zbioru danych, a nie zmiennych zbioru danych.

Analiza danych zawartych w tabeli 3 nie pozwala dopatrzeć się wyraźnego schematu zależności. W każdej grupie inne zmienne są słabo lub silnie skorelowane ze współrzędnymi w przestrzeni skalowania. Można zauważyć, że w grupach I i II mocno skorelowana ze współrzędną X jest zmienna X_{13} , czyli powierzchnia budynków wielofunkcyjnych. Ze współrzędną X w pierwszej grupie są mocno skorelowane zmienne X_2 (liczba ciągników), X_6 (powierzchnia obór ogółem) i X_{10} (powierzchnia stodół ogółem). Na podstawie tej analizy można stwierdzić, że współrzędna X w pierwszej grupie może być najbardziej stymulowana zmiennymi X_2 i X_{13} natomiast współrzędna Y zmiennymi X_3 i X_{10} . Co daje nam obraz w tym rejonie rolnictwa nieukierunkowanego. W drugiej grupie najsilniej skorelowana ze współrzędną X jest tylko jedna zmienna X_{10} , która najbardziej będzie determinować współrzędną X . Najbardziej skorelowana ze współrzędną Y w drugiej grupie jest zmienna X_{10} . W grupie trzeciej ze współrzędną X najsilniej korelowały zmienne: X_1 , X_3 i X_4 a współrzędną Y zmienne X_6 i X_{13} . Współrzędna X w grupie czwartej koreluje ze zmiennymi: X_3 , X_4 , X_7 ,

X_8 , i X_{11} , a współrzędna Y ze zmiennymi X_1 , X_2 , X_5 , i X_{12} . Gminy z grup trzeciej i czwartej pochodzą z części województwa najbardziej rozwiniętego rolniczo.

Tabela 3. Korelacja zmiennych ze współrzędnymi w przestrzeni skalowania (X, Y)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	
Grupa I	X	0,17	0,85*	0,25*	-0,06	-	-0,42*	-0,34*	-0,23	-0,26*	-0,46*	-0,07	-0,26*	0,94*
	Y	0,15	0,38*	0,64*	0,12	-	-0,06	0,15	0,20	0,00	0,86*	0,41*	0,41*	-0,01
Grupa II	X	0,07	0,33*	0,26*	0,04	0,00	-0,48*	-0,14	-0,04	-0,31*	-0,64*	-0,22	-0,05	0,97*
	Y	0,07	-0,18	-0,18	-0,08	-0,04	-0,08	0,88*	-0,06	-0,18	-0,04	-0,15	0,07	0,01
Grupa III	X	0,58*	-0,03	0,76*	0,93*	-0,30	-0,15	-0,47*	0,39	-0,28	0,32	0,05	0,33	-0,35
	Y	-0,39	0,10	-0,11	-0,03	-0,04	0,81*	-0,16	0,52*	-0,08	-0,08	0,14	-0,32	-0,70*
Grupa IV	X	-0,14	-0,36	-0,86*	-0,83*	-0,18	-0,24	0,64*	-0,84*	0,47	-0,37	-0,54	-0,37	0,56*
	Y	0,81*	0,72*	-0,29	0,02	-0,58*	0,08	0,17	0,05	-0,51	0,10	-0,35	0,56*	0,44

*wartość istotna na poziomie 0,05

Usytuowane są blisko dużych aglomeracji gdzie gospodarstwa są nastawione na produkcję warzywniczą, w gospodarstwach tych istotny udział w wyposażeniu parku maszynowego mają samochody dostawcze. Ich liczny udział determinuje współczynnik syntetyczny, stąd silna korelacja samochodów ciężarowych w tych grupach ze współrzędnymi skalowania.

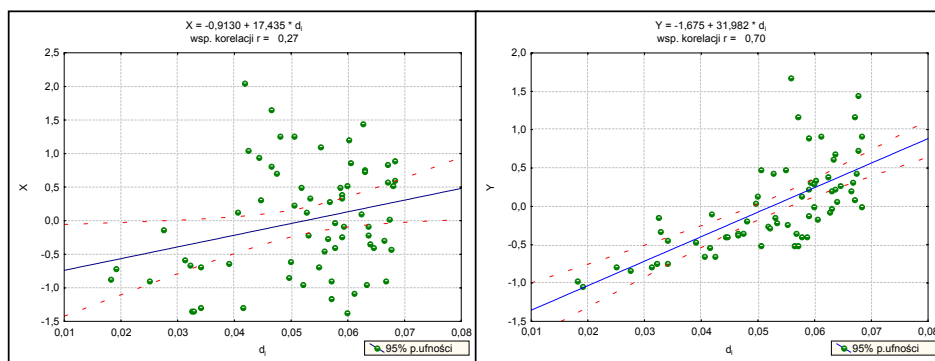
Wyniki obliczeń współczynników korelacji pomiędzy wymiarami a wskaźnikiem syntetycznym przedstawiono w tabeli 4, zaś ilustrację graficzną przedstawiono na rysunkach 2–13. Redukcja liczby wymiarów przestrzeni konfiguracyjnej metodą skalowania wielowymiarowego i już przy dwóch wymiarach w przestrzeni skalowania można otrzymać obraz obiektów. Opis dwuwymiarowy dobrze odzwierciedla strukturę grup obiektów, pozwala na łatwą identyfikację obiektów odstających oraz może stanowić dobre uzupełnienie narzędzi eksploatacji danych wejściowych obrazujących wyposażenie w infrastrukturę rolniczą badanych obiektów.

Tabela 4. Korelacja w grupach pomiędzy współrzędnymi przestrzeni skalowania a wskaźnikiem syntetycznym

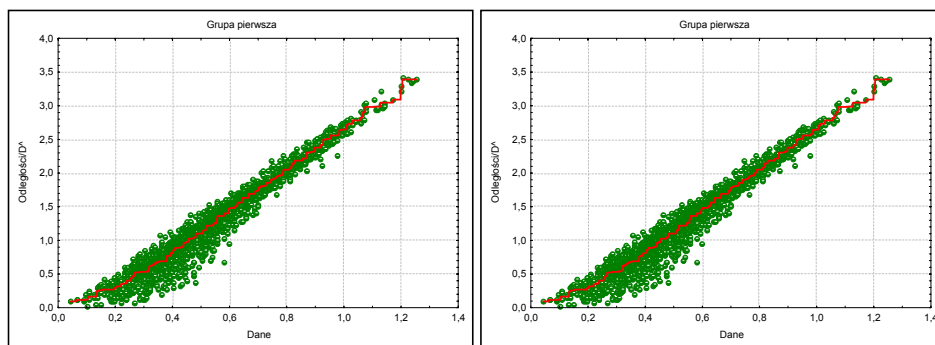
	Grupa I		Grupa II		Grupa III		Grupa IV	
	X	Y	X	Y	X	Y	X	Y
d_i	0,27*	0,7*	-0,09	0,24*	0,51*	-0,04	-0,50	-0,49

* wartość istotna na poziomie 0,05

Grupa pierwsza



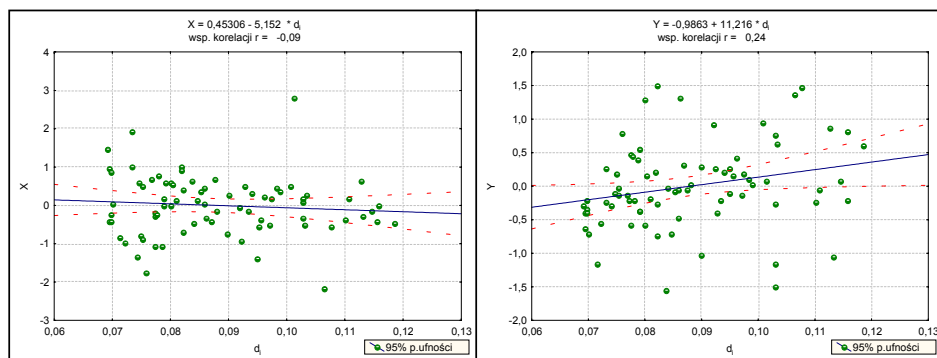
Rysunek 2. Korelacja w grupie pierwszej między współrzędnymi przestrzeni skalowania a wskaźnikiem syntetycznym



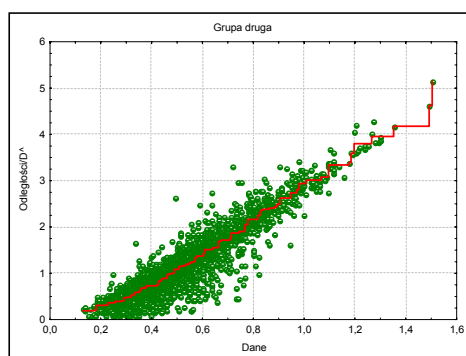
Rysunek 3. Wykres Sheparda dla dwóch wymiarów grupy pierwszej

Rysunek 4. Wykres D^4 dla dwóch wymiarów grupy pierwszej

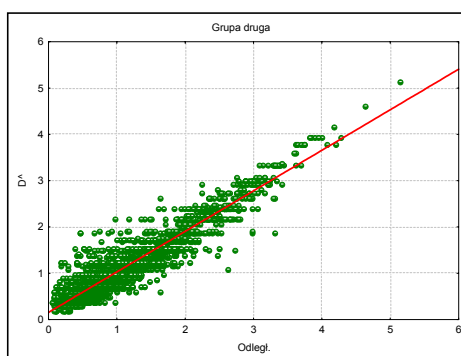
Grupa druga



Rysunek 5. Korelacja w grupie drugiej między współrzędnymi przestrzeni skalowania a wskaźnikiem syntetycznym

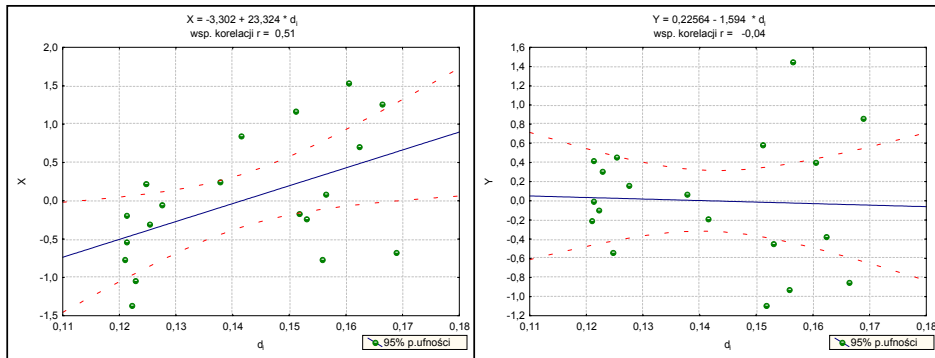


Rysunek 6. Wykres Sheparda dla dwóch wymiarów grupy pierwszej

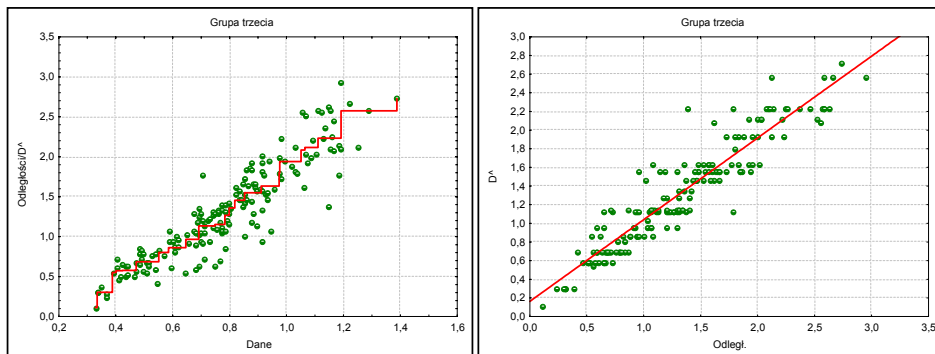


Rysunek 7. Wykres D² dla dwóch wymiarów grupy pierwszej

Grupa trzecia



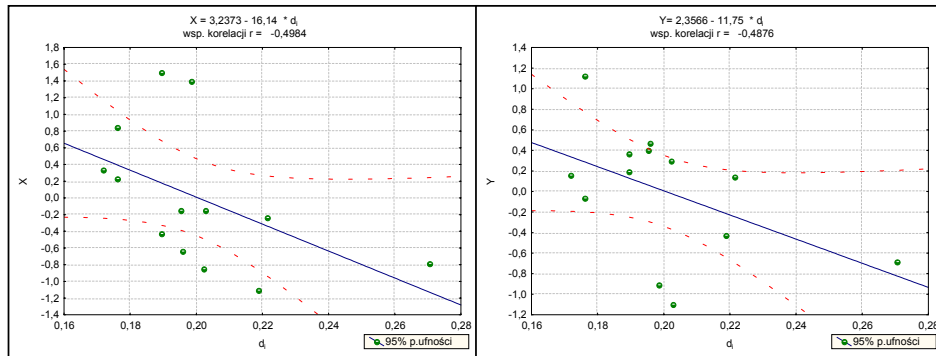
Rysunek 8. Korelacja w grupie trzeciej między współrzędnymi przestrzeni skalowania a wskaźnikiem syntetycznym



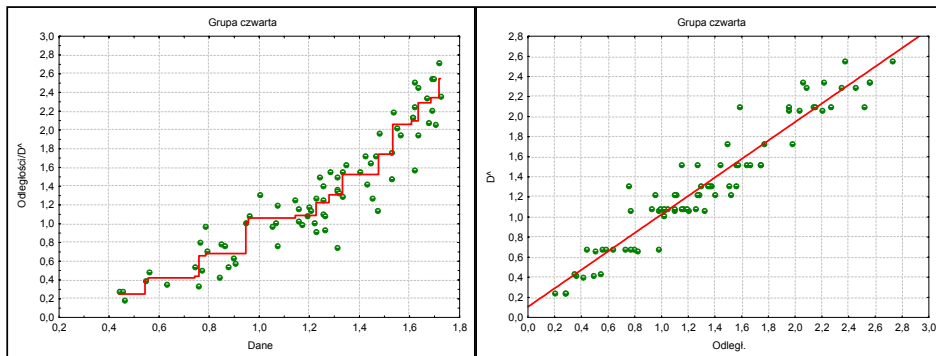
Rysunek 9. Wykres Sheparda dla dwóch wymiarów grupy pierwszej

Rysunek 10. Wykres D^a dla dwóch wymiarów grupy pierwszej

Grupa czwarta



Rysunek 11. Korelacja w grupie czwartej między współzrędnymi przestrzeni skalowania a wskaźnikiem syntetycznym



Rysunek 12. Wykres Sheparda dla dwóch wymiarów grupy pierwszej

Rysunek 13. Wykres D^{\wedge} dla dwóch wymiarów grupy pierwszej

BIBLIOGRAFIA

- Borys T. *Propozycja agregatowej miary rozwoju obiektów*. Przegląd Statystyczny, XXV, 1978, 3.
- Dąbkowski J. *Metoda oceny postępu technicznego w rolnictwie z zastosowaniem analizy wielowymiarowej*. Zesz. Nauk. AR Kraków, rozprawa nr 242, 1998.
- Grabiński T. *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*. Akademia Ekonomiczna w Krakowie, monografia, 61, 1984.
- Woźniak A. *Relacje przestrzenne i zróżnicowanie infrastruktury obszarów wiejskich województwa małopolskiego*. Zesz. Nauk. AR Kraków nr 365. Sesja Naukowa z. 72, 2000.
- Woźniak A. *Relacje przestrzenne w infrastrukturze i technicznym wyposażeniu rolnictwa w województwie małopolskim*. Inżynieria Rolnicza nr 5(25) Seria - Rozprawy habilitacyjne nr 7. Kraków 2001.
- Woźniak A. Sikora J. *Zróżnicowanie obszarów wiejskich woj. małopolskiego pod względem wyposażenia w środki techniczne gospodarstw rolnych*. Infrastruktura i Ekologia Terenów Wiejskich. nr 3 Kraków 2005.
- Zeliaś A. *Ekonometria przestrzenna*. Państwowe Wydawnictwo Ekonomiczne. Warszawa 1991.

Dr Andrzej Woźniak,
mgr inż. Jakub Sikora
Katedra Technicznej Infrastruktury Wsi
ul. Balicka 116B 30-149 Kraków

Recenzent: *Prof. dr hab. Zdzisław Wójcicki*

Andrzej Woźniak, Jakub Sikora

THE USE OF MULTIDIMENSIONAL SCALING IN THE ANALYSIS OF INFRASTRUCTURAL POTENTIAL OF AGRICULTURAL HOLDINGS IN SELECTED MUNICIPALITIES OF THE MALOPOLSKIE PROVINCE

SUMMARY

Determining infrastructural potential of a municipality is possible by measuring a number of diagnostic features describing this potential in the municipality – the number of these variables depends mainly on the purpose of the analysis. 13 variables considered as diagnostic were used in this paper to illustrate diversification of

agricultural infrastructure of farms on the municipality level. The set of these variables marks a point in the multidimensional space, which is characteristic for each investigated object but differentiating objects among themselves. The potential was estimated on the basis of synthetic measure of development which is an aggregate of diagnostic features. The computed synthetic index reduces multidimensional data to a single number and replaces the point in the multidirectional space by a point on an axis in one-dimensional space.

$$d_{ik} = \left[\sum_{j=k}^m |x_{ij} - x_{kj}|^p \right]^{\frac{1}{p}}$$

Where:

- x_{ij} – realisation of the j -th feature for the i -th and k -th object
- n – number of objects
- m – number of features
- p – natural number

The goal of the article is to show if the picture of a population of municipalities in their many dimensions and original shape seen through synthetic coefficient is similar to the original and whether the obtained one-dimensional distance properly reflects distances in the initial space of the studied infrastructure potential.

In the presented work the information quality provided by the synthetic coefficient computed for the studied districts has been analysed using multidimensional scaling. Computational method of multidimensional scaling involves minimisation of the function called stress function or slightly modified standard stress function, so called coefficient of alienation, where d_{ij} are reproduced distances at a given number of coordinates in the space of scaling. The stress function has been written as the following formula:

$$S_{(d_{ij})} = \sum (d_{ij} - f(\delta_{ij}))^2$$

Where:

- d_{ij} – reproduced distances at a given number of coordinates in the space of scaling
- $f(\delta_{ij})$ – monotonic functions of initial distances

In the scaling method no limitation are imposed on the form of expressions defining the distance. In order to facilitate the interpretation and limit the scope of the article, only two dimensions of the space of scaling were assumed.

Shepard diagram was used for scaling quality assessment.

A method of reduction of dimensions number is approximation of experimentally determined matrix of distance between objects using a lower number of dimensions. In this work we use 13 variables illustrating the potential of agricultural infrastructure of farms and the numerical force of the 1st group (areas with low value of development measure) is 67, the 2nd group (areas with average values of development measure) is 77, 3rd group (areas with high values of development measure) is 19 and the 4th group (areas of very high value of development) is 13. These data sets were used to formulate distance matrices, which have dimensions of following square

matrices: the 67x 67 matrix for the first group, the 77 x 77 matrix for the second group, 19 x 19 for the third group and 13 x 13 for the fourth group. The usual Euclidean distance has been used for the analysis conducted.

The coordinates were interpreted on the basis of conducted correlation analysis of scaling space coordinates with initial variables.

The studies comprised 178 municipalities of the malopolskie province for which it was possible to obtain statistical data and distinguish between agricultural infrastructure and municipal infrastructure of the investigated municipality.

Key words: agricultural infrastructure of farms, synthetic measure of development, multidimensional scaling, malopolskie province