PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 384

Taksonomia 24
Klasyfikacja i analiza danych – teoria i zastosowania

Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 384 • 2015
Taksonomia 24 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Małgorzata Misztal
Uniwersytet Łódzki
e-mail: mmisztal@uni.lodz.pl

O ZASTOSOWANIU KANONICZNEJ ANALIZY KORESPONDENCJI W BADANIACH EKONOMICZNYCH

Streszczenie: Celem analizy korespondencji (*Correspondence Analysis*, CA) jest graficzna prezentacja zależności między zmiennymi jakościowymi w przestrzeni o mniejszej liczbie wymiarów, przy zachowaniu jak największej liczby pierwotnych informacji. Dodatkowa informacja, ułatwiająca interpretację uzyskanych wyników, może być uwzględniona na mapie percepcji w postaci tzw. punktów dodatkowych, o zerowej masie i zerowym wkładzie w całkowitą inercję. W kanonicznej analizie korespondencji (*Canonical Correspondence Analysis*, CCA) dodatkowa informacja, opisana za pomocą tzw. zmiennych środowiskowych (objaśniających), jest uwzględniania w sposób aktywny w tworzeniu osi ordynacyjnych poprzez przyjęcie założenia, że kolejne osie są kombinacjami liniowymi dostępnych zmiennych objaśniających. Kanoniczna analiza korespondencji zyskała popularność przede wszystkim w badaniach ekologicznych, trudno natomiast znaleźć jej zastosowania w badaniach ekonomiczno-społecznych. Celem pracy jest prezentacja możliwości aplikacyjnych CCA do danych o charakterze ekonomicznym.

Słowa kluczowe: analiza korespondencji, kanoniczna analiza korespondencji, badania ekonomiczne.

DOI: 10.15611/pn.2015.384.21

1. Wstęp

Inspirację prezentowanej pracy stanowią badania ekologiczne, w których dysponujemy zwykle informacjami o składzie gatunkowym (*species*) zaobserwowanym w kolejnych badanych miejscach (*sites/samples*). Występowanie gatunków w próbach może być mierzone bezpośrednio w sposób ilościowy – wielkość biomasy, liczebność, udział procentowy w próbie [Piernik 2008]. Dodatkowo można także dla każdej badanej próby uzyskać informacje dotyczące tzw. zmiennych środowiskowych, opisujących warunki panujące w analizowanych miejscach (np. wilgotność, zawartość azotu, odczyn gleby itp.). Analiza tego rodzaju danych wykorzy-

stuje najczęściej metody redukcji wymiarowości zwane w ekologii metodami ordynacyjnymi (*ordination*) lub gradientowymi (*gradient analysis*).

Jak podaje Piernik [2008, s. 63], „termin *ordynacja* oznacza w ekologii uporządkowanie prób wzdłuż gradientu reprezentowanego przez oś diagramu ordynacyjnego, na podstawie danych o składzie gatunkowym", a celem analiz gradientowych jest „takie uporządkowanie prób, aby obiekty o podobnym składzie gatunkowym i udziale gatunków były położone blisko siebie, a oddalone od nich były obiekty odmienne".

Wyróżnia się dwie grupy metod ordynacyjnych [Jongman, ter Braak, van Tongeren 1987] – techniki ordynacji pośredniej (*indirect* / *unconstrained ordination*) oraz techniki ordynacji bezpośredniej (*direct* / *constrained ordination*). Do technik ordynacji pośredniej należy m.in. analiza składowych głównych – PCA oraz analiza korespondencji – CA. Do technik ordynacji bezpośredniej z kolei należy m.in. analiza redundancji – RDA oraz kanoniczna analiza korespondencji – CCA.

Analiza korespondencji (CA) jest szeroko stosowana w psychologii, socjologii, ekologii, marketingu czy badaniach ekonomicznych. Kanoniczna analiza korespondencji (CCA) zyskała natomiast popularność przede wszystkim w badaniach ekologicznych. Jak podaje ter Braak [2011], w ciągu 25 lat od chwili pojawienia się jego pracy o CCA [ter Braak 1986] miała ona według Web of Science ponad 3000 cytowań, z czego tylko około 6% dotyczyło zastosowań innych niż ekologiczne.

Szczególnie trudno znaleźć zastosowania kanonicznej analizy korespondencji w badaniach ekonomiczno-społecznych. Stąd też celem pracy jest prezentacja możliwości aplikacyjnych kanonicznej analizy korespondencji (CCA) do danych o charakterze ekonomicznym.

2. Analiza korespondencji i kanoniczna analiza korespondencji

Analiza korespondencji (*Correspondence Analysis*, CA) jest eksploracyjną techniką analizy tablic kontyngencji, która zmierza do odtworzenia odległości między punktami reprezentującymi wiersze i/lub kolumny w przestrzeni o mniejszej liczbie wymiarów przy jednoczesnym zachowaniu jak największej liczby pierwotnych informacji [Gatnar i Walesiak (red.) 2004, s. 284]. Do analizy można wprowadzić także punkty dodatkowe (pasywne), ułatwiające interpretację uzyskanych wyników, przy czym punkty te nie są uwzględnianie podczas wyznaczania wartości własnych, a zatem nie wpływają na wartość inercji.

W roku 1986 ter Braak zaproponował algorytm kanonicznej analizy korespondencji (*Canonical Correspondence Analysis*, CCA), w którym dodatkowa informacja, opisana za pomocą zmiennych środowiskowych (czyli objaśniających), jest uwzględniania w sposób aktywny w tworzeniu osi ordynacyjnych. Kolejne osie ordynacyjne są kombinacjami liniowymi dostępnych zmiennych środowiskowych.

Różnice między CA i CCA w skrócony sposób przedstawia Greenacre [2010].

Niech $\boldsymbol{N}$ oznacza $(I \times J)$-wymiarową macierz danych o elementach nieujemnych, a $\boldsymbol{P} = \left(\frac{1}{n}\right)\boldsymbol{N}$ – macierz korespondencji. Niech $\boldsymbol{r}$ i $\boldsymbol{c}$ oznaczają kolejno masy wiersza i kolumny oraz niech $\boldsymbol{D_r}$ i $\boldsymbol{D_c}$ będą macierzami diagonalnymi, które na głównej przekątnej posiadają odpowiednio elementy $\boldsymbol{r}$ i $\boldsymbol{c}$.

Poszukiwanie optymalnej podprzestrzeni w analizie korespondencji opiera się na rozkładzie według wartości osobliwych (SVD) macierzy

$$\boldsymbol{S} = \boldsymbol{D_r}^{-\frac{1}{2}}\left(\boldsymbol{P} - \boldsymbol{r}\boldsymbol{c}^{\boldsymbol{T}}\right)\boldsymbol{D_c}^{-\frac{1}{2}} = \boldsymbol{U}\boldsymbol{D}_{\sigma}\boldsymbol{V}^{\boldsymbol{T}}, \text{ gdzie } \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}.$$

Współrzędne główne wierszy i kolumn wynoszą odpowiednio: $\boldsymbol{F} = \boldsymbol{D_r}^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{D}_{\sigma}$ oraz $\boldsymbol{G} = \boldsymbol{D_c}^{-\frac{1}{2}}\boldsymbol{V}\boldsymbol{D}_{\sigma}$, czyli zostają przeskalowane w taki sposób, że $\boldsymbol{F}^T\boldsymbol{D}_r\boldsymbol{F} = \boldsymbol{G}^T\boldsymbol{D}_C\boldsymbol{G} = \boldsymbol{D}_{\sigma}^2$. Macierz $\boldsymbol{D}_{\sigma}^2$ jest macierzą diagonalną, utworzoną z niezerowych wartości osobliwych σ_k uporządkowanych malejąco. Suma kwadratów wszystkich wartości osobliwych nazywana jest inercją całkowitą: $\lambda = \sum \sigma_k^2$. Współrzędne standardowe profili wierszowych i kolumnowych są równe odpowiednio:

$$\boldsymbol{F}_s = \boldsymbol{D_r}^{-\frac{1}{2}}\boldsymbol{U} \text{ oraz } \boldsymbol{G}_s = \boldsymbol{D_c}^{-\frac{1}{2}}\boldsymbol{V}.$$

W przypadku kanonicznej analizy korespondencji (CCA) wymiary są zdefiniowane jako kombinacja liniowa (regresja wieloraka) zmiennych środowiskowych. Maksymalna liczba wymiarów jest równa liczbie zmiennych środowiskowych. Schemat obliczeń jest analogiczny do CA, jednakże wymagane jest wstępne rzutowanie danych na przestrzeń wyznaczoną przez zmienne objaśniające.

Przyjmijmy, że $\boldsymbol{X}$ jest $(I \times K)$-wymiarową macierzą zmiennych objaśniających oraz że zmienne te zostały wystandaryzowane. Wówczas macierz rzutowania będzie postaci: $\boldsymbol{Q} = \boldsymbol{D_r}^{\frac{1}{2}}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{D}_r\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{D_r}^{\frac{1}{2}}$. Macierz $\boldsymbol{S}$ zdefiniowana wyżej przyjmie postać: $\boldsymbol{S}^* = \boldsymbol{Q}\boldsymbol{S}$.

W przeciwieństwie do CA kanoniczna analiza korespondencji poszukuje optymalnych współrzędnych osi głównych, biorąc pod uwagę przestrzeń kanoniczną, wyznaczoną przez zmienne środowiskowe. Jak wskazuje Greenacre [2007, s. 188], na tym polega zasadnicza różnica między CA i CCA. Całkowita inercja w tym przypadku podlega dwukrotnej dekompozycji – na część związaną z przestrzenią kanoniczną (*restricted* / *canonical space*) i pozostałą część w przestrzeni niekanonicznej (*non-canonical* / *unconstrained space*), a następnie w każdej z tych podprzestrzeni na odpowiednie części związane z wyznaczonymi osiami głównymi.

Szczegółowy opis obu metod przedstawia np. Greenacre [2007] oraz ter Braak i Prentice [1988].

Istotną zaletą obu metod jest możliwość prezentacji graficznej uzyskanych wyników. W przypadku CA punkty reprezentujące poszczególne profile kategorii zmiennych można przedstawić graficznie w układzie kartezjańskim. Ocenie podle-

ga: (1) odległość punktu reprezentującego profil danej kategorii zmiennej od początku układu (im bliżej, tym rozkład danej zmiennej jest bardziej zbliżony do przeciętnego), (2) wzajemne położenie punktów odpowiadających kategoriom tej samej zmiennej (im bliżej siebie są położone, tym mniejsze jest zróżnicowanie struktury tej zmiennej), (3) wzajemne położenie punktów odpowiadających kategoriom różnych zmiennych (blisko siebie położone punkty wskazują na ich współwystępowanie).

W przypadku kanonicznej analizy korespondencji prezentacją graficzną wyników jest triplot. Zmienne środowiskowe przedstawione są na takim wykresie w postaci wektorów. Kierunek wektora wskazuje kierunek największej zmienności danej zmiennej środowiskowej, a jego długość jest proporcjonalna do znaczenia tej zmiennej.

Analizując triplot, należy zwrócić uwagę na: (1) położenie wektorów środowiskowych względem osi (kąt nachylenia wektora w kierunku osi wskazuje na korelację zmiennej z osią ordynacyjną; im mniejszy kąt nachylenia, tym silniejsza korelacja), (2) położenie wektorów środowiskowych względem siebie (wektory prostopadłe informują o braku korelacji między zmiennymi środowiskowymi, im bliżej siebie położone, tym silniejsza dodatnia korelacja między zmiennymi, położone po przeciwnych stronach wskazują na korelację ujemną zmiennych środowiskowych) oraz (3) położenie punktów odpowiadających zmiennym kolumnowym i wierszowym względem wektorów (punkty położone blisko końców wektorów lub w przeciwnym kierunku na przedłużeniu wektorów świadczą o silnym dodatnim lub ujemnym skorelowaniu zmiennych wierszowych i kolumnowych z daną zmienną środowiskową).

3. Przykład zastosowania kanonicznej analizy korespondencji

Do zilustrowania możliwości aplikacyjnych CCA do analizy danych o charakterze ekonomicznym wykorzystano dane dotyczące struktury towarowej produkcji rolniczej[1] według województw (por. tab. 1).

Sposób prezentacji danych nasuwa skojarzenia z przedstawianiem danych ekologicznych – odpowiednikiem gatunków są wyszczególnione produkty rolnicze, a odpowiednikiem miejsc obserwacji (*samples*/ *sites*) – województwa.

W tabeli 2 przedstawiono dane dotyczące 5 zmiennych środowiskowych wybranych[2] spośród wstępnie rozpatrywanych 15 zmiennych mogących mieć wpływ na zróżnicowanie rozkładów struktury produkcji.

[1] Towarowa produkcja rolnicza, jak definiuje Rocznik Statystyczny Rolnictwa [2013, s. 51], stanowi sumę sprzedaży produktów rolnych do skupu i na targowiskach.

[2] Selekcji zmiennych środowiskowych dokonano, posługując się algorytmem opisanym w pracy Lepša i Šmilauera [2003] wykorzystującym metody Monte Carlo i testy permutacyjne.

Tabela 1. Struktura towarowej produkcji rolniczej w 2011 r. (ceny stałe z 2010 r.)

Województwo/ (skrócona nazwa)		Zboża	Uprawy przemysł.	Ziemniaki	Warzywa	Owoce	Żywiec wołowy	Żywiec wieprz.	Drób	Jaja	Mleko
Dolnośląskie	D	38,5	15,5	3,5	6,7	3,8	2,7	4,6	8,4	6,2	6,6
Kujawsko-pomorskie	C	13,1	9,8	2,9	14,2	1,7	4,6	20,5	9,5	2,5	19,8
Lubelskie	L	7,7	8,0	2,6	13,0	28,2	4,2	12,5	5,6	2,9	11,6
Lubuskie	F	10,9	4,0	1,8	10,5	2,4	3,8	14,8	26,1	7,7	9,3
Łódzkie	E	4,0	1,5	13,3	12,4	12,5	8,2	15,7	9,9	3,5	17,2
Małopolskie	K	6,3	1,6	4,8	25,2	10,2	9,4	9,6	5,1	6,4	12,0
Mazowieckie	W	6,8	1,4	6,1	12,2	19,3	4,4	8,6	11,9	3,4	20,1
Opolskie	O	32,0	16,5	0,9	4,1	0,8	2,4	16,2	9,6	2,2	13,3
Podkarpackie	R	9,2	5,6	1,5	13,7	8,0	3,3	12,7	12,1	7,6	20,5
Podlaskie	B	3,8	0,7	1,9	2,9	1,2	8,5	10,2	7,4	2,1	59,3
Pomorskie	G	18,7	6,9	8,5	4,4	1,3	4,5	27,4	10,0	4,1	11,4
Śląskie	S	9,0	3,1	4,4	8,0	3,3	6,4	13,3	18,0	15,9	11,3
Świętokrzyskie	T	5,6	2,7	3,4	17,6	20,6	7,6	8,6	8,9	2,9	18,3
Warmińsko-mazurskie	N	13,4	3,7	1,3	2,6	1,2	3,2	13,9	23,2	2,1	31,9
Wielkopolskie	P	8,2	5,9	1,9	5,7	0,7	7,7	23,7	11,9	13,0	16,0
Zachodniopomorskie	Z	27,9	8,7	5,1	4,3	2,6	2,2	13,4	20,4	4,0	8,5

Źródło: Rocznik Statystyczny Rolnictwa [2013] i Rocznik Statystyczny Województw [2013].

Tabela 2. Wybrane zmienne środowiskowe

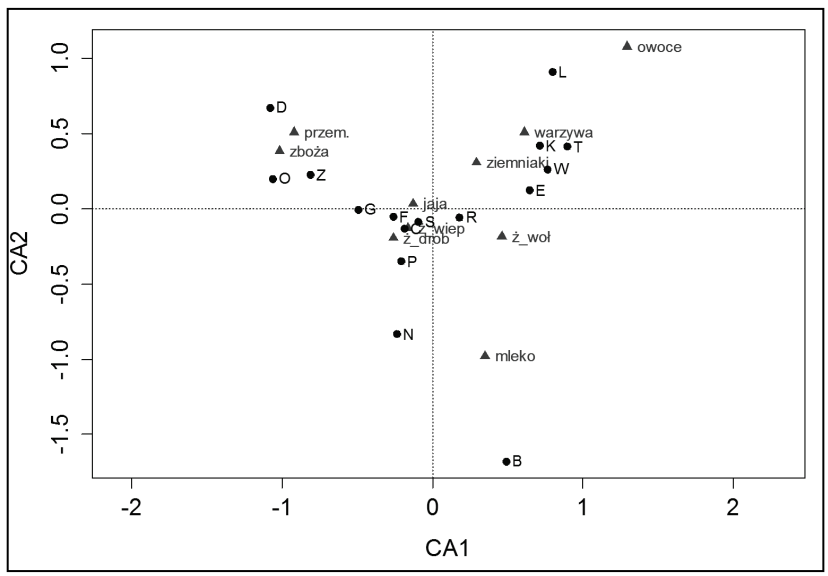
Województwo/ (skrócona nazwa)		X1	X2	X3	X4	X5
Dolnośląskie	D	62,6	10,24	61 099	8,8	7 123,3
Kujawsko-pomorskie	C	48,3	8,50	64 770	10,4	9 041,3
Lubelskie	L	26,9	9,41	177 525	21,7	11 334,9
Lubuskie	F	38,6	9,09	22 354	6,7	2 827,2
Łódzkie	E	20,2	9,51	130 241	18,0	9 809,3
Małopolskie	K	8,3	9,75	152 176	47,1	6 668,0
Mazowieckie	W	15,3	11,86	234 503	14,8	18 019,8
Opolskie	O	70,3	8,38	26 832	9,4	4 870,6
Podkarpackie	R	7,5	9,26	134 024	41,9	5 824,9
Podlaskie	B	12,6	8,51	80 873	11,5	8 405,1
Pomorskie	G	39,0	9,40	40 035	8,3	5 017,5
Śląskie	S	37,5	11,63	64 803	26,3	5 218,4
Świętokrzyskie	T	8,0	10,31	92 654	29,7	4 991,8
Warmińsko-mazurskie	N	36,4	7,66	43 995	6,3	7 217,3
Wielkopolskie	P	59,4	10,47	123 228	11,5	18 422,3
Zachodniopomorskie	Z	46,9	9,23	28 739	5,1	6 033,7

X1 – zużycie nawozów wapniowych na 1 ha użytków (naw_wap); X2 – odsetek osób z wyższym wykształceniem na wsi (w.wyż); X3 – gospodarstwa rolne ogółem (gosp); X4 – pracujący w rolnictwie na 100 ha użytków rolnych (prac.100ha); X5 – wartość brutto środków trwałych ogółem w mln zł (ST).

Źródło: Rocznik Statystyczny Rolnictwa [2013] i Rocznik Statystyczny Województw [2013].

Analizę korespondencji oraz kanoniczną analizę korespondencji przeprowadzono z wykorzystaniem pakietu `vegan` ze środowiska R.

Wyniki analizy korespondencji przedstawiono na rys. 1. Całkowita inercja wynosi 0,458. Dwa wymiary wyjaśniają łącznie 64% inercji (oś 1 – 38%, wartość własna: 0,175; oś 2 – 26%, wartość własna: 0,120).



Rys. 1. Wyniki analizy korespondencji (CA) – mapa percepcji

Źródło: opracowanie własne na podstawie danych z tab. 1.

Analizując rozmieszczenie punktów na rys. 1, można zauważyć, że blisko środka układu leżą punkty reprezentujące produkcję jaj, żywca wieprzowego i drobiu oraz województwa śląskie i kujawsko-pomorskie. Są to profile najbardziej zbliżone do profili przeciętnych. Oznacza to, że wielkość produkcji jaj, żywca wieprzowego i drobiu charakteryzuje się stosunkowo niewielkim zróżnicowaniem w badanych województwach w porównaniu do pozostałych produktów rolnych. Województwo śląskie i kujawsko-pomorskie określić można mianem typowych ze względu na strukturę towarowej produkcji rolniczej.

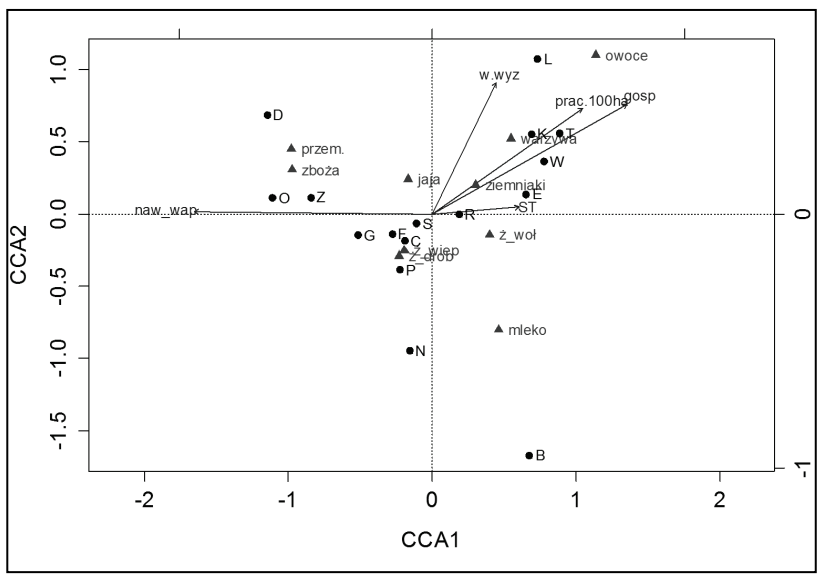
Rozkłady struktury produkcji najbardziej odbiegają od rozkładu przeciętnego dla województwa podlaskiego (B), lubelskiego (L), dolnośląskiego (D) i opolskiego (O).

Oś pierwsza wyraźnie oddziela zboża i uprawy przemysłowe od pozostałych typów produkcji. Ze względu na produkcję zbóż i upraw przemysłowych szczególnie wyróżniają się województwa dolnośląskie (D), opolskie (O) i zachodniopomorskie (Z). Województwa te, jak podaje Agencja Rynku Rolnego [2013a], należą do grupy województw o najwyższym skupie zbóż w kraju.

Największy udział produkcji mleka w produkcji całkowitej występuje dla woj. podlaskiego (B) i warmińsko-mazurskiego (N). Jak podaje Czernewski [2012], podmioty skupujące w województwie podlaskim skupiły w ciągu dwunastu miesięcy 2011 r największą ilość mleka spośród wszystkich województw – 2 632,6 mln litrów, co stanowiło 29,2 % skupu krajowego.

Największy udział produkcji owoców w produkcji całkowitej występuje dla woj. lubelskiego (L), które jest liderem w produkcji owoców z krzewów i jagodowych i zapewnia ponad połowę krajowych zbiorów tych owoców, por. [ARR, 2013b].

Wyniki kanonicznej analizy korespondencji przedstawiono na rys. 2. Zmienne środowiskowe są na diagramie przedstawione w postaci wektorów.



Rys. 2. Wyniki kanonicznej analizy korespondencji (CCA) – mapa percepcji (triplot)

Źródło: opracowanie własne na podstawie danych z tab. 1 i 2.

Inercja w układzie osi kanonicznych wynosi 0,2922, co stanowi 64% inercji całkowitej. Wartości własne osi kanonicznych są równe: 0,153 dla osi 1 oraz 0,087 dla osi 2. Pierwsza oś kanoniczna wyjaśnia 52% zmienności w układzie kanonicznym (33% całkowitej zmienności), a druga oś kanoniczna – 30% (19% zmienności całkowitej). 36% zmienności całkowitej nie jest wyjaśniane przez przyjęte zmienne środowiskowe, należy zatem uznać, że oprócz prezentowanych w analizie istnieją inne czynniki mające wpływ na zróżnicowanie rozkładów zmiennych (np. nakłady inwestycyjne itp.).

Największy wpływ na zmienność rozkładów ma zużycie nawozów wapniowych (najdłuższy wektor). Zmienna ta jest najsilniej skorelowana z pierwszą osią kano-

niczną. Z drugą osią kanoniczną najsilniej koreluje odsetek mieszkańców wsi z wyższym wykształceniem.

Analizując współczynniki korelacji[3] między zmiennymi kolumnowymi (rodzaj produkcji) a zmiennymi środowiskowymi, można zauważyć m.in., że wzrost zużycia nawozów wapniowych ma związek z wysokim udziałem produkcji zbóż i upraw przemysłowych w produkcji ogółem dla danego województwa. Z kolei w województwach z wysokim odsetkiem rolników z wyższym wykształceniem obserwuje się większy udział produkcji warzyw i owoców.

Spośród uwzględnionych zmiennych środowiskowych relatywnie mały wpływ na zróżnicowanie rozkładów zmiennych ma wartość brutto środków trwałych ogółem w mln zł (ST).

4. Zakończenie

Przedstawiony w pracy przykład zastosowania kanonicznej analizy korespondencji, jak i inne, niezaprezentowane w artykule badania własne pozwalają przypuszczać, że metoda ta może być stosowana do analizy danych o charakterze ekonomicznym. Dodatkowych badań wymaga natomiast wskazanie tego obszaru nauk ekonomicznych, w którym CCA byłaby szczególnie przydatna.

Prezentacja graficzna wyników CCA z wykorzystaniem triplotu ułatwić może analizę powiązań między zmiennością rozkładów badanych kategorii ekonomicznych i czynnikami mogącymi wpływać na tę zmienność.

Warto też zauważyć, że kanoniczna analiza korespondencji nie ma tak silnych założeń jak wiele innych metod wielowymiarowych, a zatem może być stosunkowo łatwo stosowana w analizach zmiennych, które zwykle nie spełniają założeń, np. dotyczących normalności rozkładu. Zmienne środowiskowe mogą być mierzone na różnych skalach pomiaru (przedziałowej, ilorazowej, nominalnej) bez konieczności spełnienia założeń dotyczących np. postaci rozkładu.

Literatura

Agencja Rynku Rolnego, 2013a, *Rynek zbóż w Polsce*, http://www.arr.gov.pl/data/00321/rynek_zboz_2013_pl.pdf (28.12.2014).

Agencja Rynku Rolnego, 2013b, *Oddział Terenowy Agencji Rynku Rolnego w Lublinie*, http://www.arr.gov.pl/data/00321/2013_ot_lublin.pdf (28.12.2014 r.).

Czernewski K., 2012, *Wielkość skupu i ceny mleka w 2011 r.*, http://www.mlekpol.com.pl/ (28.12.2014 r.).

Gatnar E., Walesiak M. (red.), 2004, *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.

Greenacre M., 2007, *Correspondence Analysis in Practice*, 2ed., Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton.

[3] Ze względu na ograniczoną zawartość pracy pominięto szczegółowe wyniki liczbowe.

Greenacre M., 2010, *Canonical Correspondence Analysis in Social Science Research*, [w:] Locarek-Junge H., Weihs C. (red.), *Classification as a Tool for Research. Studies in Classification, Data Analysis and Knowledge Organization*, Springer, Berlin, Heidelberg, s. 279-286.
GUS, 2013, *Rocznik Statystyczny Rolnictwa*, GUS, Warszawa.
GUS, 2013, *Rocznik Statystyczny Województw*, GUS, Warszawa.
Jongman R.H.G., ter Braak C.J.F., van Tongeren O.F.R. (red.), 1995, *Data Analysis in Community and Landscape Ecology*, Cambridge University Press, Cambridge.
Lepš J., Šmilauer P., 2003, *Multivariate Analysis of Ecological Data using CANOCO*, Cambridge University Press, Cambridge.
Piernik A., 2008, *Metody numeryczne w ekologii*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika w Toruniu, Toruń.
ter Braak C.J.F., 1986, *Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis*, Ecology, vol. 67(5), s. 1167-1179.
ter Braak C.J.F., 2011, *The history of canonical correspondence analysis*, Talk given on the occasion of the CARME 2011, www.youtube.com/watch?v=fO88UjIgk1s (15.10.2014).
ter Braak C.J.F., Prentice I.C., 1988, *A theory of gradient analysis*, Advances in Ecological Research, vol. 18, s. 271-317.

ON THE USE OF CANONICAL CORRESPONDENCE ANALYSIS IN ECONOMIC RESEARCH

Summary: The aim of correspondence analysis (CA) is graphical gpresentation of relations among qualitative variables in low-dimensional subspace with optimal explanation of inertia [Greenacre 2007, p. 185]. Additional information can be added to the map in the form of supplementary (passive) points with zero mass and zero inertia in order to interpret their positions relative to the active points. By contrast, in canonical correspondence analysis (CCA) the dimensions are assumed to be responses in a regression-like relationship with external variables i.e. dimensions are found with the same CA objective but with the restriction that the dimensions are linear combinations of a set of explanatory variables [Greenacre 2007, p. 192]. CCA is very popular in ecological research but almost completely unknown in socio-economic research. The objective of the paper is to present the possibilities of application of CCA to economic data.

Keywords: correspondence analysis, canonical correspondence analysis, economic study.