METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XVII/3, 2016, s. 95 – 107

OCENA ZMIAN STOPNIA ZANIECZYSZCZANIA ŚRODOWISKA W POLSCE W LATACH 2004-2014 PRZY WYKORZYSTANIU PODSTAWOWYCH NARZĘDZI ANALITYCZNYCH

Koszela Grzegorz
Katedra Ekonometrii i Statystyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: grzegorz_koszela@sggw.pl
Szczesny Wiesław
Katedra Informatyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: wieslaw_szczesny@sggw.pl

Streszczenie: W artykule podjęto próbę oceny zmian stopnia zanieczyszczenie środowiska na poziomie województw w latach 2004-2014. Ocenę tą przeprowadzono przy pomocy budowy rankingów województw. Rankingi te utworzono na podstawie zmiennych syntetycznych powstałych w wyniku normalizacji zmiennych metodą unitaryzacji zerowanej oraz przekształcenia ilorazowego. Zwrócono również uwagę na problem obserwacji odstających. Okazuje się, że w zależności od podejścia do tego problemu, można uzyskać znacząco różniące się wyniki dotyczące grupowania województw w klasy.

Słowa kluczowe: ranking, zmienna syntetyczna, unitaryzacja zerowana, przekształcenie ilorazowe, obserwacje odstające, gradacyjna analiza danych, ochrona środowiska, stopień zanieczyszczenia

WSTĘP

Od wielu lat w Polsce prowadzone są dyskusje nad poprawą jakości środowiska naturalnego. Jednakże postęp w zakresie technologii, chęć podniesienia poziomu życia oraz wysokie tempo życia powodują iż do natury trafia w Polsce dużo odpadów/zanieczyszczeń. Wejście Polski do UE według powszechnych oczekiwań, powinno skutkować sukcesywnym zmniejszaniem się negatywnego oddziaływania na środowisko. Celem artykułu jest kompleksowe –

z narzędziowego punktu widzenia - spojrzenie na wybrane problemy związane oceną stopnia negatywnego oddziaływania na środowisko, zarówno w ujęciu przestrzennym jak i na przestrzeni ostatniego 10-ciolecia. Zazwyczaj przedstawienie tego zagadnienia od strony analitycznej prowadzi do budowy rankingu (~ów) jednostek terytorialnych z zastosowaniem wielokryterialnych ocen i ewentualnie do porównania tego rankingu do innych rankingów tych jednostek terytorialnych (według charakterystyk opisujących ich zamożność i poziom inwestycji związanych z ochroną środowiska). Jednakże przy budowie wielokryterialnych ocen oraz publikacji rankingów opartych o te oceny, nie powinien być stosowany wyłącznie automatyzm uzasadnioniony faktem, że są to techniki ugruntowane. Bardzo interesującą pracą metodyczną ilustrującą sposób oceny negatywnego oddziaływania na środowisko w 2012 roku jest praca Profesora Karola Kukuły [Kukuła 2014]. Natomiast dyskusję na temat oceny stabilności rankingu przed jego upublicznieniem na przykładzie oceny stopnia zanieczyszczania środowiska w Polsce w ujęciu regionalnym, można znaleźć w pracy [Koszela, Szczesny 2015].

Metody i narzędzia badawcze zastosowane w niniejszym artykule to techniki z szerokiego zbioru pod nazwą „wielowymiarowa analiza porównawcza" [por np. Kukuła 2000]. Naturalną alternatywą dla technik uzanawanych za klasyczne są narzędzia z instrumentarium tzw. Gradacyjnej Analizy Danych (GAD). Z braku miejsca nie będą tu jednak przedstawione wyniki badań uzyskane przy ich użyciu [por. Szczesny 2002, Kowalczyk i in. 2004, Koszela 2016].

Praca ta powstała, aby zwrócić uwagę na stabilność ocen zawartych w tworzonych raportach, czyli na sprawdzeniu, czy w zależności od zastosowanej techniki, uzyskane wyniki nie prowadzą do różnych wniosków.

WYKORZYSTANE DANE

W celu analizy zmian dotyczących stopnia zanieczyszczenia środowiska w poszczególnych województwach w okresie 10 lat, w tym badaniu wytypowano pewne zmienne diagnostyczne, które tworzą macierz:

$$\mathbf{X} = \begin{bmatrix} x_{1,1,1} & x_{1,1,2} & \dots & x_{1,1,k} \\ x_{1,2,1} & x_{1,2,2} & \dots & x_{1,2,k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,n,1} & x_{1,n,2} & \dots & x_{1,n,k} \\ x_{2,1,1} & x_{2,1,2} & \dots & x_{2,1,k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{r,n,1} & x_{r,n,2} & \dots & x_{r,n,2} \end{bmatrix} = \left[x_{s,i,j}\right] \quad \begin{pmatrix} \mathrm{s} = 1,\dots,\mathrm{r} \\ \mathrm{i} = 1,\dots,\mathrm{n} \\ \mathrm{j} = 1,\dots,\mathrm{k} \end{pmatrix}, \tag{1}$$

gdzie: n - liczba obiektów, k - liczba zmiennych diagnostycznych, r - liczba lat w analizowanym okresie, $x_{s,i,j}$ - oznacza wartość zmiennej na obiekcie O_i w roku o numerze s.

Jako zmienne, które posłużyły do oceny zmian stopnia zanieczyszczania środowiska, wybrano subiektywnie 6 głównych aspektów dotyczących negatywnego wpływu na środowisko naturalne w podziale na województwa za okres 2004-2014, czyli od roku, w którym Polska przystąpiła do UE [GUS 2005-2015]. W celu pewnego ujednolicenia, dane te zostały przeliczone na 100 km^2:

X_1 - ścieki przemysłowe i komunalne wymagające oczyszczania odprowadzone do wód lub do ziemi (odprowadzone jako nieoczyszczone w hektometrach sześciennych),

X_2 - emisja zanieczyszczeń pyłowych z zakładów szczególnie uciążliwych (w tysiącach ton),

X_3 - emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych (w tysiącach ton),

X_4 - grunty wymagające rekultywacji (w hektarach)

X_5 - emisja metali ciężkich z zakładów szczególnie uciążliwych (w kg)

X_6 - odpady wytworzone według województw (magazynowane czasowo w tys. ton)

Wszystkie zmienne są stymulantami wspomagającymi ocenę stopnia zanieczyszczania środowiska.

METODA BADAWCZA

Budowa wskaźnika syntetycznego, który będzie wykorzystany do porównywania (budowy rankingu) województw ze względu na wielkość corocznego zanieczyszczania środowiska, wymaga wyboru zmiennych diagnostycznych oceniających różne aspekty tego zjawiska. Należy podkreślić, że ważnym czynnikiem określającym ilość informacji, jaką dostarczą wybrane do analizy zmienne (poza dokładnością pomiaru), jest typ wykorzystanej skali pomiarowej. W tym przypadku, czyli pomiaru wielkości poszczególnych zanieczyszczeń, jest to zwykle skala ilorazowa (czyli „mocniejsza" niż przedziałowa) [Luce i in. 1990, Hand i in. 2001, Walesiak 1990, Holder 1901]. Zmienne ilorazowe są podobne do zmiennych przedziałowych, lecz oprócz wszystkich cech skali przedziałowej, charakteryzuje je istnienie punktu absolutnego zera na skali. Dlatego w odniesieniu do zmiennych ilorazowych, prawomocne jest stwierdzenie typu: X_1 jest dwa razy większe niż X_2. Jednakże przy budowie wskaźników syntetycznych, jeśli chcemy, aby jego wartości pozostały na skali ilorazowej, mamy ograniczenie tylko do jednego rodzaju normowania, a mianowicie do przekształcenia ilorazowego. Należy jednak podkreślić, że w większości procedur statystycznych zaimplementowanych w pakietach komercyjnych, nie dokonuje się rozróżnienia pomiędzy skalami ilorazową i przedziałową.

Jednym z ważnych zagadnień podczas budowy wskaźnika syntetycznego, jest statystyczna analiza potencjalnych zmiennych diagnostycznych pod kątem ich ewentualnej eliminacji. W literaturze jest wiele podpowiedzi dla początkujących

analityków. Jako przesłanki do wyodrębnienia takich zmiennych podawane są najczęściej kryteria wykorzystujące różne miary nierówności/rozproszenia. Celem jest wyeliminowanie każdej takiej zmiennej, która różni się tylko nieznacznie od zmiennej stałej. Często taką zmienną nazywa się potocznie zmienną kwasi stałą. Jak zwykle problem tkwi w samej definicji pojęcia zmienna quasi stała. W wielu podręcznikach nadal jako quasi stałą rozumie się zmienną, której współczynnik zmienności V ma niską wartość (np. $V < 0,1$). Wynika to z potrzeby wskazywania zmiennych, które bez przekształcenia, nie mogły być wzięte do obliczeń współczynników regresji liniowej wykonywanych tradycyjnym algorytmem [Borkowski i in. 2007]. Według tego kryterium zmienna $X \sim N(2000;5)$ spełnia warunek, aby zostać nazwana quasi stałą. Inna bardziej naturalna definicja zmiennej quasi stałej mówi:

$$P(X=a) = 1-\alpha,\ P(X\neq a) = \alpha, \text{ gdzie } \alpha \text{ jest małe (np. } < 0,1).$$

Aby zilustrować rozterki badacza przy tworzeniu rankingu, rozważmy przykład (patrz tabela 1), dotyczący 10 obiektów opisanych za pomocą 4 zmiennych (przyjmijmy, że są to stymulanty). Przyjmijmy też, że mamy do czynienia ze zmiennymi ilorazowymi oraz przyjmijmy dla ustalenia uwagi założenie, że wszystkie zmienne są jednakowo ważne (czyli załóżmy, że możemy przyjąć jednakowe wagi). Zmienna ilorazowa jest także zmienną przedziałową, a zatem uprawnione jest zastosowanie normowania w postaci unitaryzacji zerowanej w celu zbudowania rankingu. Wartości wskaźników syntetycznych W_1, W_2 utworzono jako średnie z wartości unormowanych zmiennych X_1-X_4 przy użyciu odpowiednio unitaryzacji zerowanej (wskaźnik W_1) i przekształcenia ilorazowego (W_2, gdzie wykorzystano dzielenie przez średnią). W kolumnach R_1 i R_2 zamieszczono rankingi oparte o wartości wskaźników W_1 i W_2. Z tabeli 1 wynika, że rankingi różnią się znacznie. Miejsce obiektu O_{08} zmienia się aż o 7 pozycji, a obiektu O_{09} o 4 pozycje. Ten prosty zabieg, wykorzystujący tylko dwie różne normalizacje, wyraźnie sugeruje, że przed publikacją rankingu trzeba zasięgnąć opinii analityka. Jedną z przyczyn tego stanu rzeczy może być występowanie zmiennych quasi stałych lub elementów odstających, które mają znacząco inny wpływ w przypadku różnych technik normowania danych. Jeśli prześledzimy wartości współczynnika zmienności (wiersz V), to zauważymy, że w kolumnach X_2 i X_4 występują małe liczby, natomiast np. w kolumnie X_1 wartość jest znacząco wyższa, mimo iż zmienna przyjmuje (poza jednym przypadkiem) wartość równą 15. Po unormowaniu za pomocą unitaryzacji zerowanej – czyli przejściu na intuicyjny zakres wartości z przedziału [0;1] - zanika problem małych wartości współczynnika zmienności. Jednakże patrząc na wartości innych wskaźników po unormowaniu, np. na unormowaną wartość wskaźnika Gini (po unitaryzacji – kolumny UX_1 i UX_3- wynosi on odpowiednio 1,0 oraz 0,9) można dojść do wniosku, że rozkładom wartości zmiennych X_1 oraz X_3 należy przyjrzeć się dokładniej. Podobne ostrzeżenie pokazuje także w tym przypadku rozstęp międzykwartylowy IQR (wartości 0 oraz 0,063 dla danych po unitaryzacji

zerowanej). Ponieważ dane są małoliczne, więc wszystko (łącznie z występowaniem elementów odstających) jest wyraźnie widoczne. Nawet bez obliczania wskaźników widać, że zmienne X_1 oraz X_3 wymagają specjalnej uwagi. W przypadkach liczniejszych zbiorów danych, sytuacja może już nie być taka klarowna. Łatwo sprawdzić na tym przykładzie, że jeśli ograniczymy się tylko do zmiennych X_2 i X_4, to w przypadku zastosowania omawianych dwóch typów normalizacji rankingi będą ze sobą identyczne, a rozkłady wartości wskaźników syntetycznych prawie symetryczne, w przeciwieństwie do pokazanych w tabeli 1.
W tabeli 1 przyjęto następujące oznaczenia:
μ - średnia, σ - odchylenie standardowe, V-współczynnik zmienności, Q_i - i-ty kwartyl, IQR - rozstęp międzykwartylowy, GINI* - unormowany wskaźnik Gini, W_1 i W_2 – wskaźniki syntetyczne otrzymane jako średnia z unormowanych wartości zmiennych X_1-X_4 przy użyciu unitaryzacji zerowanej oraz przekształcenia ilorazowego wykorzystującego średnią, UX_1-UX_4 wartości zmiennych X_1-X_4 po unitaryzacji, R_i - rankingi według wartości wskaźnika syntetycznego W_i.

Tabela 1. Przykładowy zestaw danych

	X_1	X_2	X_3	X_4	UX_1	UX_2	UX_3	UX_4	W_1	W_2	R_1	R_2
O_{01}	15	12,70	16	13,70	0	0,765	0	0,765	0,382	0,	2	4
O_{02}	15	12,65	16	13,65	0	0,706	0	0,706	0,353	0,912	3	5
O_{03}	15	12,60	16	13,60	0	0,647	0	0,647	0,324	0,910	4	6
O_{04}	15	12,55	16	13,55	0	0,588	0	0,588	0,294	0,908	5	7
O_{05}	15	12,25	16	13,25	0	0,235	0	0,235	0,118	0,896	7	8
O_{06}	15	12,20	16	13,20	0	0,176	0	0,176	0,088	0,894	8	9
O_{07}	15	12,15	16	13,15	0	0,118	0	0,118	0,059	0,892	9	10
O_{08}	15	12,10	18	13,10	0	0,059	0,083	0,059	0,050	0,915	10	3
O_{09}	15	12,05	30	13,05	0	0,000	0,583	0,000	0,146	1,064	6	2
O_{10}	50	12,90	40	13,9	1	1	1	1	1,000	1,694	1	1
μ	18,5	12,415	20	13,415	0,1	0,429	0,167	0,429	0,281	1,000		
s	10,5	0,283	7,849	0,283	0,3	0,333	0,327	0,333	0,268	0,236		
V	0,568	0,023	0,392	0,021	3	0,775	1,962	0,775	0,953	0,236		
min	15	12,05	16	13,05	0	0	0	0	0,05	0,892		
max	50	12,90	40	13,9	1	1	1	1	1	1,694		
GINI**	0,189	0,014	0,180	0,013	1	0,482	0,900	0,482	0	0,000		
Q_1	15	12,16	16	13,16	0	0,132	0	0,132				
Q_3	15	12,64	17,5	13,64	0	0,691	0,063	0,691				
IQR	0	0,475	1,5	0,475	0	0,559	0,063	0,559				

Źródło: opracowanie własne

Z tego dość przerysowanego przykładu wynika, że w przypadku budowy rankingów i ewentualnego podziału obiektów na grupy w oparciu o wartości wskaźnika syntetycznego, należy odpowiednio zabezpieczyć się przed upublicznieniem nierzetelnego raportu. Zabezpieczenie takie polega na wykorzystaniu szerszego spektrum narzędzi analitycznych, dzięki któremu będzie możliwość wystawiania ostrzeżeń (reguł stopu), dających podstawę do zasięgnięcia

opinii u bardziej doświadczonego analityka. Jest to szczególnie ważne w przypadku, gdy wartości wszystkich zmiennych są na skali ilorazowej, bo wtedy aby wartości wskaźnika syntetycznego (będącego liniową kombinacją tych zmiennych) były także na skali ilorazowej, dopuszczalne jest (w procesie normalizacji zmiennych) użycie tylko przekształcenia ilorazowego. Przekształcenie ilorazowe nie zmienia takich parametrów rozkładów jak współczynnik zmienności, czy też wartość koncentracji Giniego, a jego wartości, niejednokrotnie są wskazywane w literaturze jako podstawa ustalania wartości wagowych dla poszczególnych zmiennych. Proponuje się aby te wagi były proporcjonalne do wskaźnika zmienności [por. np. Betti, Verma 1999, Sawiłow 2011]. Szerzej o technologii tworzenia wag można też przeczytać np. w pracach [Abrahamowicz, Zając 1986, Bąk 1999]. Z tego powodu jako jeden z sygnałów ostrzegawczych, przed publikacją wyników dotyczacych stopnia zanieczyszczania środowiska (który zazwyczaj będzie oparty na zmiennych o wartościach na skali ilorazowej) można uznać pomiędzy rankingami, przy tworzeniu których stosowano normalizację za pomocą przekształcenia ilorazowego i unitaryzacji zerowanej. Łatwo sprawdzić, że w przykładzie z tabeli 1 współczynnik korelacji $\rho(R_1; R_2) = 0{,}49$, gdzie R_1 i R_2 to dwa rankingi oparte o te dwie normalizacje.

W rozważanym przypadku badania stopnia poziomu corocznego zanieczyszczania środowiska, dla j=1,…,k wartości μ_j, a_j, b_j zostały policzone w następujący sposób:

$$\begin{aligned} \mu_j &= \sum_{s=1}^{r}\sum_{i=1}^{n} x_{s,i} \\ a_j &= min_{\{s,i\}}\{ x_{s,i,j} : s=1,...r,\ i=1,...n,\ j=1,...,k\} \\ b_j &= max_{\{s,i\}}\{ x_{s,i,j} : s=1,...r,\ i=1,...n,\ j=1,...,k\}. \end{aligned} \tag{2}$$

Wykorzystując te wartości dokonano normalizacji zmiennych X_1–X_5 według następujących wzorów:

$$U_j = \frac{X_j}{\mu_j}, \quad \mathrm{V_j} = \frac{X_j - a_j}{b_j - a_j}, \quad \mathrm{j} = 1,...,\mathrm{k} \tag{3}$$

Jako wskaźniki syntetyczne oceniające stopień zanieczyszczania środowiska naturalnego przyjęto:

$$W_U = \frac{1}{6}\sum_{j=1}^{6} U_j, \qquad W_V = \frac{1}{6}\sum_{j=1}^{6} V_j \tag{4}$$

Natomiast jako sygnalizację dotyczącą ostrożności w zakresie publikacji rankingu województw w badanym okresie, przyjęto podobieństwa wektorów wartości współczynników W_U i W_V w poszczególnych latach oraz podobieństwa rankingów zbudowanych w oparciu o te wartości. Jako miarę podobieństwa obydwu uporządkowań przyjęto współczynnik rho-Pearsona.

WYNIKI BADAŃ

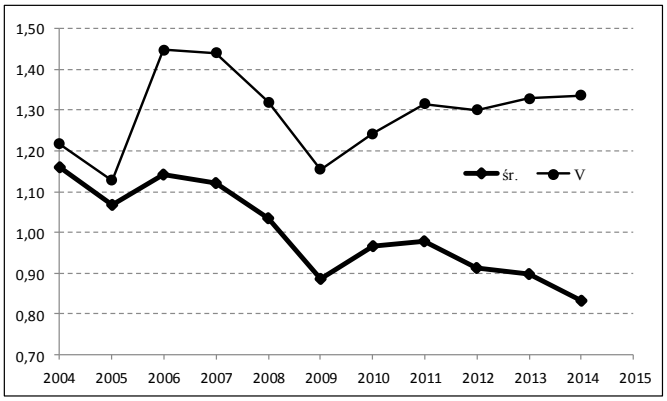
W tabelach 2 i 3 zaprezentowano odpowiednio wartości wskaźników syntetycznych W_V oraz W_U określonych wzorem (4). Natomiast na rysunku 1 zilustrowano zmiany średniej wartości oraz współczynnika zmienności V wskaźnika syntetycznego W_V. W oparciu o rysunek 1 oraz wartości w ostatnich trzech wierszach tabel 2 i 3, wyraźnie widać, ze zmniejsza się poziom corocznego zanieczyszczania środowiska. Jednakże wyraźny wzrost współczynnika zmienności w okresie 2009-2014 wskazuje, że zróżnicowanie pomiędzy województwami wzrasta. Co więcej w przypadku województw podlaskiego i świętokrzyskiego poziom corocznego zanieczyszczania środowiska w latach 2013-2014 wzrósł w porównaniu do poziomu z 2004 roku, podczas gdy w pozostałych województwach notowany jest jego wyraźny spadek. Ponadto warto zwrócić uwagę na bardzo duże zróżnicowanie województw pod względem poziomu corocznego zanieczyszczania środowiska (odchylenie standardowe niezależnie od sposobu normalizacji jest większe od średniej – por. tabele 2 i 3).

Tabela 2. Wartości wskaźnika W_V w okresie 2004-2014 oceniającego poziom corocznego zanieczyszczania środowiska naturalnego względem województw

Województwo	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Dolnośląskie	1,70	1,42	1,38	1,38	1,47	1,37	1,47	1,60	1,42	1,31	1,11
Kujawsko-Pomorskie	0,81	0,94	0,58	0,65	0,50	0,52	0,77	0,65	0,62	0,53	0,52
Lubelskie	0,37	0,36	0,32	0,33	0,28	0,54	0,24	0,34	0,24	0,22	0,21
Lubuskie	0,33	0,30	0,30	0,29	0,27	0,27	0,25	0,19	0,18	0,18	0,17
Łódzkie	1,11	1,14	1,04	1,02	1,05	0,88	0,84	0,92	0,92	0,94	0,90
Małopolskie	1,88	1,84	1,71	1,12	1,27	0,68	0,72	0,75	0,84	0,78	0,77
Mazowieckie	1,18	1,22	1,08	0,97	0,75	0,61	0,96	0,87	0,67	0,55	0,54
Opolskie	1,01	0,98	1,04	1,16	0,91	1,23	1,45	1,33	1,15	0,91	0,81
Podkarpackie	0,43	0,43	0,43	0,42	0,33	0,28	0,33	0,30	0,30	0,30	0,24
Podlaskie	0,24	0,24	0,21	0,22	0,21	0,20	0,18	0,19	0,19	0,37	0,38
Pomorskie	0,57	0,55	0,55	0,60	0,60	0,49	0,39	0,37	0,41	0,36	0,36
Śląskie	0,97	0,72	0,74	1,08	1,16	1,13	1,36	1,29	1,33	1,56	1,20
Świętokrzyskie	6,32	5,41	7,33	7,22	6,12	4,62	5,32	5,68	5,25	5,27	4,97
Warmińsko-Mazurskie	0,26	0,26	0,26	0,25	0,29	0,28	0,24	0,25	0,22	0,24	0,23
Wielkopolskie	0,70	0,68	0,68	0,63	0,64	0,61	0,53	0,54	0,55	0,54	0,58
Zachodniopomorskie	0,68	0,61	0,60	0,60	0,71	0,44	0,42	0,37	0,32	0,31	0,35
μ	1,16	1,07	1,14	1,12	1,04	0,89	0,97	0,98	0,91	0,90	0,83
s	1,41	1,21	1,65	1,61	1,37	1,02	1,20	1,29	1,19	1,19	1,11
V	1,22	1,13	1,45	1,44	1,32	1,15	1,24	1,32	1,30	1,33	1,34

Źródło: opracowanie własne

Rysunek 1. Zmiana średniego poziomu corocznego zanieczyszczania środowiska (W_V)



Źródło: opracowanie własne

W analizowanych danych wszystkie wartości zmiennych były na skali ilorazowej, a zatem naturalnym jest rozpatrywanie normowania wyłącznie w postaci przekształcenia ilorazowego. Jednakże bardziej popularnym jest przekształcenie w postaci unitaryzacji zerowanej, bo po unormowaniu otrzymujemy wartości z przedziału [0;1]. Dlatego w tabeli 3 przedstawiono wartości wskaźnika syntetycznego wykorzystującego tego typu normowanie. Należy podkreślić, że unitaryzacja zerowana jest bardziej wrażliwa na elementy odstające niż przekształcenie ilorazowe. W tym badaniu województwo śląskie znacząco odbiega poziomem zanieczyszczania środowiska od pozostałych województw, dlatego jeśli porównamy rankingi w 2004 roku stopnia zanieczyszczania środowiska przy wykorzystaniu wskaźników W_V i W_U, to zauważalne są spore różnice między nimi. Szczegóły dotyczace rankingu dla roku 2004 zamieszczono w tabeli 4. Miejsca w rankingach dla województwa mazowieckiego róznią się aż o 5 pozycji, a dla opolskiego o 3. Miara podobieństwa tych dwu rankingów w postaci współczynnika korelacji wynosi zaledwie 0,906. Zatem powstaje pytanie, który z tych rankingów upublicznić? Oczywiście w automatycznym systemie raportowania powinna zadziałać reguła stopu, aby o upublicznieniu zdecydował analityk danych. Jeśli do obliczenia parametrów wyznaczających oba normowania, tzn. przekształcenie ilorazowe oraz unitaryzację zerowaną (por wzór (3)), wykorzystamy tylko dane dla 15 województw pomijając województwo śląskie, to rankingi będą bardzo podobne do siebie w całym okresie 2004-2014 (będą oczywiście pojedyncze przesunięcia o jedną pozycję), a podobieństwo mierzone za pomocą rho Pearsona dla rozważanego roku 2004 wyniesie 0,994.

Tabela 3. Wartości wskaźnika W_U w okresie 2004-2014 oceniającego poziom corocznego zanieczyszczania środowiska naturalnego według województw

Województwo	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Dolnośląskie	0,30	0,25	0,26	0,25	0,27	0,24	0,23	0,29	0,27	0,26	0,23
Kujawsko-Pomorskie	0,15	0,16	0,13	0,13	0,11	0,12	0,14	0,13	0,13	0,12	0,12
Lubelskie	0,06	0,06	0,05	0,05	0,04	0,06	0,03	0,04	0,04	0,03	0,03
Lubuskie	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,02	0,03	0,02	0,02
Łódzkie	0,22	0,23	0,22	0,22	0,22	0,19	0,19	0,20	0,21	0,22	0,21
Małopolskie	0,24	0,22	0,21	0,16	0,16	0,08	0,09	0,10	0,10	0,10	0,10
Mazowieckie	0,15	0,15	0,14	0,12	0,10	0,08	0,12	0,11	0,09	0,08	0,07
Opolskie	0,24	0,24	0,25	0,26	0,23	0,27	0,29	0,26	0,23	0,19	0,17
Podkarpackie	0,07	0,07	0,06	0,05	0,04	0,03	0,03	0,03	0,03	0,03	0,02
Podlaskie	0,04	0,04	0,04	0,04	0,04	0,04	0,03	0,03	0,03	0,05	0,05
Pomorskie	0,08	0,08	0,08	0,09	0,09	0,07	0,07	0,07	0,07	0,07	0,07
Śląskie	0,19	0,15	0,16	0,19	0,20	0,20	0,24	0,23	0,24	0,26	0,22
Świętokrzyskie	0,79	0,71	0,86	0,85	0,75	0,61	0,68	0,73	0,68	0,68	0,66
Warmińsko-Mazurskie	0,07	0,07	0,07	0,07	0,07	0,06	0,06	0,06	0,06	0,06	0,06
Wielkopolskie	0,19	0,19	0,18	0,17	0,18	0,17	0,16	0,16	0,16	0,15	0,17
Zachodniopomorskie	0,09	0,09	0,09	0,08	0,09	0,06	0,06	0,05	0,05	0,03	0,05
μ	0,18	0,17	0,18	0,17	0,16	0,15	0,15	0,16	0,15	0,15	0,14
s	0,18	0,16	0,19	0,19	0,17	0,14	0,16	0,17	0,16	0,16	0,15
V	0,97	0,92	1,09	1,10	1,04	0,97	1,03	1,08	1,05	1,10	1,08

Źródło: opracowanie własne

Tabela 4. Ranking województw według wartości wskaźników W_V i W_U w roku 2004

Województwo	**DŚ**	**KP**	**LB**	**LS**	**ŁD**	**MP**	**MZ**	**OP**	**PK**	**PL**	**PM**	**ŚK**	**ŚL**	**WM**	**WP**	**ZP**
Rank(W_V)	3	8	13	14	5	2	4	6	12	16	11	7	1	15	9	10
Rank(W_U)	2	8	14	16	5	4	9	3	13	15	11	7	1	12	6	10
delta	**1**	**0**	**-1**	**-2**	**0**	**-2**	**-5**	**3**	**-1**	**1**	**0**	**0**	**0**	**3**	**3**	**0**

Źródło: opracowanie własne

Wykorzystując wartości z tabeli 2 dokonano podziału województw na cztery grupy pod względem poziomu corocznego zanieczyszczania środowiska. Wskaźnik W_V ma wartości na skali ilorazowej więc wobec dużego zróżnicowania województw ze względu na poziom zanieczyszczania przyjęto trzy progi: 0,5; 1; 2. Grupa o wartościach wskaźnika W_V poniżej 0,5 to grupa oznaczona numerem 1 (grupa najlepsza, bo o najniższym poziomie corocznego zanieczyszczania środowiska), a numerem 4 oznaczono grupę najgorszych, w której poziom zanieczyszczania środowiska mierzony wartością W_V przekracza wartość 2. Szczegółowe wyniki zawiera tabela 5. Wynika z niej, że do grupy województw znacząco zanieczyszczających środowisko dołączyło świętokrzyskie, które w stosunku do 2004 zwiększyło stopień corocznego zanieczyszczania środowiska, a województwo podlaskie, także mimo zwiększenia stopnia zanieczyszczania nadal pozostało w grupie 1. Województwo łódzkie natomiast znalazło się w grupie województw umiarkowanie zanieczyszczających corocznie środowisko.

Tabela 5. Podział województw na grupy pod względem poziomu corocznego zanieczyszczania środowiska przy wykorzystaniu przy wykorzystaniu wartości wskaźnika W_V z tabeli 2 oraz progów 0,5; 1; 2

Województwo	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Dolnośląskie	3	3	3	3	3	3	3	3	3	3	3
Kujawsko-Pomorskie	2	2	2	2	1	2	2	2	2	2	2
Lubelskie	1	1	1	1	1	2	1	1	1	1	1
Lubuskie	1	1	1	1	1	1	1	1	1	1	1
Łódzkie	3	3	3	3	3	2	2	2	2	2	2
Małopolskie	3	3	3	3	3	2	2	2	2	2	2
Mazowieckie	3	3	3	2	2	2	2	2	2	2	2
Opolskie	3	2	3	3	2	3	3	3	3	2	2
Podkarpackie	1	1	1	1	1	1	1	1	1	1	1
Podlaskie	1	1	1	1	1	1	1	1	1	1	1
Pomorskie	2	2	2	2	2	1	1	1	1	1	1
Śląskie	2	2	2	3	3	3	3	3	3	3	3
Świętokrzyskie	4	4	4	4	4	4	4	4	4	4	4
Warmińsko-Mazurskie	1	1	1	1	1	1	1	1	1	1	1
Wielkopolskie	2	2	2	2	2	2	2	2	2	2	2
Zachodniopomorskie	2	2	2	2	2	1	1	1	1	1	1

Źródło: opracowanie własne

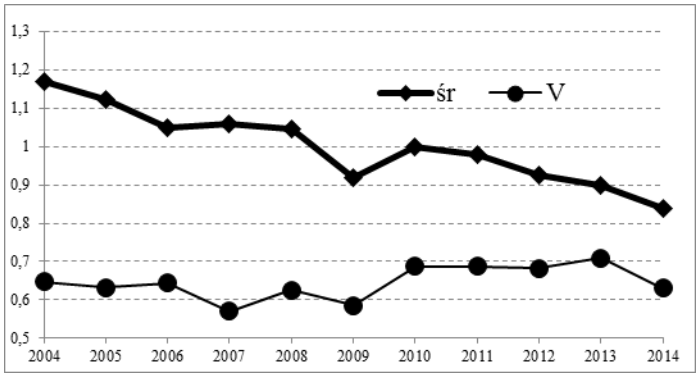
Jeśli mamy do czynienia z obiektem nietypowym (odstającym), to w przypadku podziału na grupy według standardowej wartości wskaźnika syntetycznego, należałoby potraktować go w sposób szczególny. W naszym badaniu podział na grupy według progów 0,5; 1, 2 został dokonany w oparciu o normalizację w postaci przekształcenia ilorazowego. Przekształcenie to wykorzystywało parametry μ_j (j=1,…6). Jeżeli przy wyznaczaniu parametrów μ_j, pominiemy dane dotyczące województwa śląskiego, otrzymamy nowe wyniki podziału województw na grupy wg. stopnia zanieczyszczenia, które prezentuje tabela 6. Takie podejście pozwoliło wskazać wyraźniej zachodzące zmiany w przynależności poszczególnych województw do określonych grup. Tym samym można np.: zaobserwować zmianę grupy dla województw podlaskiego i świętokrzyskiego, które zwiększyły poziom zanieczyszczania środowiska w stosunku do roku 2004. Okazało się również, że w 2004 do grupy 4-tej należało zaliczyć także województwa dolnośląskie i małopolskie, które poprzez zmniejszanie poziomu zanieczyszczania w latach następnych opuściły grupę największych trucicieli środowiska. Zmiany średniego poziomu W_V dla 15-tu województw (bez śląskiego) prezentuje rysunek 2.

Tabela 6. Podział województw na grupy pod względem poziomu corocznego zanieczyszczania środowiska przy wykorzystaniu wartości wskaźnika W_V (gdzie μ_j dla j = 1,…,6 nie uwzględnia danych dla woj. śląskiego) oraz progów 0,5; 1; 2

Województwo	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Dolnośląskie	4	4	4	4	4	3	4	4	4	3	3
Kujawsko-Pomorskie	3	3	2	2	2	2	3	2	2	2	2
Lubelskie	1	1	1	1	1	2	1	1	1	1	1
Lubuskie	1	1	1	1	1	1	1	1	1	1	1
Łódzkie	3	3	3	3	3	3	3	3	3	3	3
Małopolskie	4	4	4	3	4	3	3	3	3	3	3
Mazowieckie	3	3	3	3	3	2	3	3	3	2	2
Opolskie	3	3	3	3	3	3	4	3	3	3	3
Podkarpackie	2	2	2	2	2	1	2	2	2	2	1
Podlaskie	1	1	1	1	1	1	1	1	1	2	2
Pomorskie	2	2	2	2	2	2	2	1	2	1	2
Śląskie	3	2	3	3	4	3	4	4	4	4	4
Świętokrzyskie	4	4	4	4	4	4	4	4	4	4	4
Warmińsko-Mazurskie	1	1	1	1	1	1	1	1	1	1	1
Wielkopolskie	2	2	2	2	2	2	2	2	2	2	2
Zachodniopomorskie	3	2	2	2	2	2	2	1	1	1	1

Źródło: opracowanie własne

Rysunek 2. Zmiana średniego poziomu corocznego zanieczyszczania środowiska (W_V) dla 15 województw (z wyłączeniem województwa śląskiego)



Źródło: opracowanie własne

PODSUMOWANIE

Przy ocenie zmian poziomu zanieczyszczania środowiska w ciągu roku opieramy się często na wartościach zmiennych, których wartości są na skali ilorazowej (tzn. z zerem bezwzględnym). Aby nie stracić informacji możemy wykorzystać w procesie normalizacji w zasadzie tylko przekształcenie ilorazowe.

Zmienne o wartościach na skali ilorazowej mogą być poddane także metodom analizy wymagającym tylko skali przedziałowej. Dlatego w takich przypadkach warto oprócz przekształcenia ilorazowego wykorzystać intuicyjną normalizację w postaci unitaryzacji zerowanej, bo prowadzi ona do wartości na przedziale [0;1], które są dla wielu analityków zakresem najbardziej przyjaznym. Wykorzystanie tej innej normalizacji i porównanie wyników otrzymanych przy jej pomocy jest dobrą regułą stopu dla udostępnienia/publikacji raportu zawierającego ranking lub podział na grupy w oparciu o wskaźnik syntetyczny. W dobie praktycznie bez kosztowych obliczeń powinny być przeprowadzane procesy obliczeniowe co najmniej dwoma metodami i w przypadku każdej większej rozbieżności w uporządkowaniu obiektów lub podziale na grupy system raportowy powinien zatrzymać raport i skierować go do decyzji analityka danych.
W przypadku naszego badania należy podkreślić iż użycie dwu normalizacji pokazało znaczące rozbieżności w uporządkowaniu obiektów. Spowodowane to jest znaczącym odstępstwem stopnia zanieczyszczania środowiska w województwie Śląskim w stosunku do pozostałych województw. Natomiast w zbiorze 15 województw tzn. po wyeliminowaniu z analizy województwa śląskiego takiej rozbieżności w rankingach wykorzystujących te dwie normalizacje już nie ma. Dodatkowo podział na grupy według stopnia zanieczyszczenia środowiska przy wykorzystaniu tylko 15 województw do konstrukcji parametrów przekształceń normalizacyjnych jest znacząco lepszy niż ten oparty o standardową procedurę wykorzystującą dane z 16 województw.

BIBLIOGRAFIA

Abrahamowicz M., Zając K. (1986) Metoda ważenia zmiennych w taksonomii numerycznej i procedurach porządkowania liniowego. [w:] Prace Naukowe AE we Wrocławiu, 328, 5-17.

Bąk A. (1999) Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++. Wyd. AE, Wrocław.

Betti G., Verma V., (1999) Measuring the degree of poverty in a dynamic and comparative context: a multidimensional approach using fuzzy set theory. Proceedings of the ICCS-VI, Lahore, Pakistan, 11, 289–301.

Borkowski B., Dudek H., Szczesny W. (2007) Ekonometria. Wybrane zagadnienia. PWN, Warszawa.

Hand D., Maninila H., Smyth P. (2001) Principles of Data Mining. Cambridge: MIT Press.

Holder O. (1901) Die Axiome der Quantität und die Lehre vom Mass. Ber.Verh. Kgl. Sächsis. Ges. Wiss. Leipzig, Math.-Phys. Classe, 53, 1-64.

Koszela G. (2016) Wykorzystanie narzędzi gradacyjnej analizy danych do klasyfikacji podregionów pod względem struktury agrarnej. Wiadomości Statystyczne, 6, 10-30.

Koszela G., Szczesny W. (2015) Wykorzystanie narzędzi WAP do oceny poziomu zanieczyszczania środowiska w ujęciu przestrzennym. Metody Ilościowe w Badaniach Ekonomicznych, XVI/3, 183 – 193.

Kowalczyk T., Pleszczyńska E., Ruland F. (eds.) (2004) Grade Models and Methods of Data Analysis. With applications for the Analysis of Data Population. Studies in Fuzziness and Soft Computing, 151, Springer Verlag, Berlin - Heidelberg - New York.

Kukuła K. (2014) Wybrane problemy ochrony środowiska w Polsce w świetle wielowymiarowej analizy porównawczej. Metody Ilościowe w Badaniach Ekonomicznych, XV/3, 169 – 188.

Kukuła K. (2000) Metoda unitaryzacji zerowanej, PWN, Warszawa.

Luce R. D., Krantz D. H., Suppes P. C. (1990) Foundations of Measurement Volume III: Representation, Axiomatization, and Invariance. Academic Press, New York and London.

Sawiłow E. (2011) Ocena algorytmów wyceny nieruchomości w podejściu porównawczym. Studia i Materiały Towarzystwa Naukowego Nieruchomości, 19 (3), 20-32.

Szczesny W. (2002) Grade correspondence analysis applied to contingency tables and questionnaire data. Intelligent Data Analysis, 6 (1), IOS Press, Amsterdam.

Walesiak M. (1990) Syntetyczne badania porównawcze w świetle teorii pomiaru. Przegląd Statystyczny, 1-2, 37-46.

EVALUATION OF CHANGES OF ENVIRONMENTAL POLLUTION DEGREE IN POLAND 2004-2014 USING THE BASIC ANALYTICAL TOOLS

Abstract: The aim of the paper was to attempt to evaluate changes in the degree of pollution at the level of Voivodeships in the years 2004-2014. Assessment was carried out by construction of Voivodeship rankings. These rankings were created on the basis of synthetic variables resulting from the normalization of variables by unitarisation zeroed method and the quotient mapping. It was also paid attention to the problem of outliers. It was proved that depending on the approach to this problem, it can be obtained significantly different results for clustering Voivodeships into classes.

Keywords: ranking, synthetic variable, unitarisation zeroed, quotient mapping, outliers, grade data analysis, environmental protection, pollution degree