

WYKORZYSTANIE NARZĘDZI WAP DO OCENY POZIOMU ZANIECZYSZCZANIA ŚRODOWISKA W UJĘCIU PRZESTRZENNYM

Grzegorz Koszela

Katedra Ekonometrii i Statystyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: grzegorz_koszela@sggw.pl

Wiesław Szczesny

Katedra Informatyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: wieslaw_szczesny@sggw.pl

Streszczenie: W pracy została omówiona problematyka oceny stabilności rankingu przed jego upublicznieniem na przykładzie oceny zanieczyszczenia środowiska w Polsce w ujęciu regionalnym. Została zaproponowana procedura oceny stabilności przy wykorzystaniu miary zróżnicowania zbioru wektorów. Praca ta jest pewnym uzupełnieniem wyników i dyskusji uzyskanych przez [Kukuła 2014].

Słowa kluczowe: ranking, wskaźnik zróżnicowania, wskaźnik syntetyczny, gradacyjna analiza danych, program GradeStat

WSTĘP

Wpływ działalności człowieka na stan środowiska aktualnie jest coraz częściej przedmiotem wielu ocen i porównań. Rejestrowane i udostępniane publicznie dane pozwalają badaczom na porównania pomiędzy różnymi jednostkami terytorialnymi zarówno w kraju jak i na świecie. Do porównań wykorzystywane są zazwyczaj standardowe narzędzia Wielowymiarowej Analizy Porównawczej (WAP). Narzędzia te są prezentowane szeroko także w polskiej literaturze (por. np. [Kukuła 2000], [Zeliaś 2000], [Malina 2004], [Młodak 2006], [Gatnar, Walesiak 2009]). Przy użyciu tych narzędzi budowane są rankingi, obszarów, które po udostępnieniu szerszemu odbiorcy czasami wywołują dyskusje dotyczące przedstawianego uporządkowania/rankingu. Zazwyczaj powodowane są

one faktem, iż twórcy rankingu nie sprawdzają jego stabilności. A przecież przy budowie porównań jest wiele miejsca na subiektywizm. Co więcej powszechność automatyzacji systemów raportowych powoduje, że wiele rankingów powstaje bez udziału doświadczonego analityka.

Celem niniejszej pracy jest przedstawienie jak przy wykorzystaniu podstawowych narzędzi WAP można ustrzec się przed zaprezentowaniem rankingu, który przy ustalonym zestawie przyjętych zmiennych może być uznany za dyskusyjny. Zatem można przyjąć, że zawarte w tej pracy rozważania są w pewnym sensie kontynuacją pracy profesora Karola Kukuły [Kukuła 2014]. Należy podkreślić iż zagadnienie to jest szczególnie istotne, gdyż obecnie występuje często automatyczne raportowanie w ramach różnych systemów informacyjnych, które w przypadku mało lub średnio krytycznych odbiorców może nawet wprowadzać w błąd. Aby lepiej przedstawić problemy, które stoją przed analitykiem budującym ranking, rozpatrzmy pewien „przerysowany” sztuczny przykład, który zostanie zaprezentowany w następnej sekcji.

PRZYKŁAD WPROWADZAJĄCY

W Tabeli 1 zamieszczono przykład obejmujący 5 zmiennych i 8 obiektów. Wartości zmiennych X_1, \dots, X_5 zostały tak dobrane, aby zilustrować kilka problemów pojawiających się przed analitykiem jednocześnie. Zmienne posiadają różne zakresy wartości, wartość odstającą (nietypową) oraz zróżnicowanie wartości niektórych z nich mierzone współczynnikiem zmienności V (iloraz odchylenia standardowego do wartości średniej) jest bardzo małe.

Przystępując do konstrukcji rankingu analityk dysponujący wybranym zestawem stymulant, jeśli to ma miejsce w zautomatyzowanym systemie raportowo-informacyjnym, kieruje się zazwyczaj następującymi przesłankami przy budowie wskaźnika syntetycznego:

- należy wyeliminować zmienne quasi-stałe,
- wartości poszczególnych zmiennych należy sprowadzić do porównywalności poprzez proces normowania,
- ustalić wagi poszczególnych zmiennych zgodnie ze wskazaniem ekspertów lub w oparciu o wskazania ogólne (np. [Młodak 2006], [Betti, Soldi, Talev 2015]).

W literaturze wskazuje się, że jeśli nie ma bezpośrednich merytorycznych wskazań dla wag, to te najprostsze wskazówki ograniczają się do następujących:

- wszystkie zmienne mają jednakowe wagi,
- wagi są proporcjonalne do wartości współczynników zmienności dla zmiennych oryginalnych,
- wagi są proporcjonalne do wartości wskaźników zróżnicowania/zmienności dla zmiennych przekształconych za pomocą przekształcenia normalizacyjnego,

- wagi proporcjonalne do odwrotności sum wartości bezwzględnych współczynników w wierszu macierzy korelacji pomiędzy wybranymi do badania stymulantami,
- wagi proporcjonalne do iloczynu wartości wag ustalonych dwoma powyższymi metodami.

Ta ostatnia procedura ustalania wag w wersji uszczegółowionej została szerzej opisana m. in. w pracach ([Betti, Verma 1999], [Panek 2011], [Betti, Soldi, Talev 2015]). Została ona oficjalnie przyjęta przez Eurostat [2002] w pomiarze ubóstwa oraz wykluczenia społecznego. Zaletą tej procedury jest to, iż uwzględnia ona problem redundancji (tzn. nie powiela tych samych informacji zawartych w różnych cząstkowych wskaźnikach w danym wymiarze) oraz przypisuje mniejszą wagę tym wskaźnikom cząstkowym, które charakteryzują się niewielką zmiennością,

Tabela 1. Sztuczny przykład prezentujący 5 stymulant dla 8-miu obiektów w ujęciu nominalnym (zmiennie X1, ..., X5) oraz po unitaryzacji zerowanej (zmiennie Z1, ..., Z5)

Obiekty	X1	X2	X3	X4	X5	Z1	Z2	Z3	Z4	Z5
O01	15,13	149,17	751,48	31,00	2009,00	0,077	0,933	0,475	1,000	0,571
O02	14,00	154,15	751,10	26,00	2013,00	0,061	1,000	0,000	0,500	0,857
O03	13,60	145,61	751,90	25,00	2011,00	0,055	0,885	1,000	0,400	0,714
O04	10,60	124,46	751,70	28,00	2007,00	0,013	0,600	0,750	0,700	0,429
O05	11,50	115,61	751,65	30,00	2015,00	0,026	0,480	0,687	0,900	1,000
O06	9,70	106,23	751,60	24,00	2003,00	0,000	0,354	0,625	0,300	0,143
O07	12,40	97,73	751,75	27,00	2005,00	0,038	0,239	0,813	0,600	0,286
O08	80,00	80,00	751,20	21,00	2001,00	1,000	0,000	0,125	0,000	0,000
maximum	80,00	154,15	751,90	31,00	2015,00	1,000	1,000	1,000	1,000	1,000
minimum	9,70	80,00	751,10	21,00	2001,00	0,000	0,000	0,000	0,000	0,000
średnia	20,87	121,62	751,55	26,50	2008,00	0,159	0,561	0,559	0,550	0,500
odch. stand.	22,41	24,94	0,26	3,04	4,58	0,319	0,336	0,321	0,304	0,327
V	1,074	0,205	0,000	0,115	0,002	2,007	0,599	0,574	0,553	0,655
max-min	70,30	74,15	0,80	10,00	14,00	1,000	1,000	1,000	1,000	1,000
max/min	8,247	1,927	1,001	1,476	1,007	x	x	x	x	x
kwartyl 1	10,83	99,86	751,27	24,25	2003,50	0,016	0,268	0,213	0,325	0,179
kwartyl 3	14,85	148,28	751,74	29,50	2012,50	0,073	0,921	0,797	0,850	0,821
IQR=Kw3 -	4,02	48,43	0,47	5,25	9,00	0,057	0,653	0,584	0,525	0,643

Źródło: opracowanie własne

Zauważmy iż w przypadku gdy użyjemy wag proporcjonalnych do współczynników zmienności dla danych oryginalnych, to mamy do czynienia z sytuacją marginalizacji tych zmiennych, które mają małe współczynniki zmienności. W celu ilustracji wpływu poszczególnych działań na wartości wskaźnika syntetycznego oraz na ranking obiektów, w Tabeli 2 zamieszczono wartości 5-ciu wskaźników otrzymanych w oparciu o wartości zmiennych Z1, ..., Z5 z Tabeli 1, odpowiednio: dla wag jednakowych (W1), dla wag proporcjonalnych do współczynników

zmienności dla X_1, \dots, X_5 (W2), dla wag proporcjonalnych do współczynników zmienności dla Z_1, \dots, Z_5 (W3), dla wag wyznaczonych w oparciu o macierz korelacji (W4) oraz W5 wykorzystujący wagi proporcjonalne do iloczynów wag użytych do stworzenia W3 i W4. Natomiast w Tabeli 3 umieszczono analogiczne wartości wskaźników dla danych różniących się od zawartych w Tabeli 1 jedną wartością. Jest to wartość zmiennej X_1 dla obiektu O08, której wartość zmieniono na wartość progu θ - po przekroczeniu którego wartości są oznaczane jako nietypowe. Wartość θ wyliczono według wzoru:

$$\theta = q_3 + 1,5 \cdot (q_3 - q_1) \quad (1)$$

gdzie q_1 i q_3 oznaczają odpowiednio kwartył pierwszy i trzeci rozkładu wartości analizowanej zmiennej. Otrzymana wartość 20,88 jest nadal najwyższa, ale w procesie unitaryzacji nie powoduje tak dużego jak wartość oryginalna zniżenia wartości Z_1 dla pozostałych obiektów (poza O08), co znacząco wpływa na wartość wskaźników syntetycznych. Szczególnie jest to widoczne przy analizie wartości wskaźnika W2, który praktycznie nie wykorzystuje (z uwagi na wielkość wag) informacji o wartościach zmiennych X_3 i X_5 (współczynniki zmienności nieznacznie większe od zera) i mocno przeważa informację o wartościach X_1 (duża wartość współczynnika zmienności dla X_1 z uwagi na wartość odstającą).

Tabela 2. Wartości wskaźników syntetycznych i ranking obiektów dla danych z Tabeli 1

Obiekty	W1	W2	W3	W4	W5	RW1	RW2	RW3	RW4	RW5
O01	0,611	0,280	0,436	0,608	0,456	2	2	2	3	3
O02	0,484	0,236	0,355	0,449	0,348	5	3	5	5	6
O03	0,611	0,207	0,434	0,664	0,494	3	4	3	1	1
O04	0,498	0,156	0,338	0,532	0,382	4	6	6	4	5
O05	0,619	0,166	0,430	0,633	0,464	1	5	4	2	2
O06	0,284	0,077	0,189	0,324	0,229	7	8	8	7	8
O07	0,395	0,115	0,275	0,440	0,321	6	7	7	6	7
O08	0,225	0,769	0,474	0,200	0,424	8	1	1	8	4

Źródło: opracowanie własne

Tabela 3. Wartości wskaźników syntetycznych i ranking obiektów dla danych z Tabeli 1 po zmianie wartości zmiennej X_1 dla O08 z 80 na 20,88

Obiekty	W1	W2	W3	W4	W5	RW1	RW2	RW3	RW4	RW5
O01	0,693	0,753	0,671	0,677	0,658	1	1	1	2	2
O02	0,548	0,633	0,545	0,518	0,518	4	2	4	5	4
O03	0,670	0,556	0,645	0,695	0,669	2	3	2	1	1
O04	0,512	0,397	0,470	0,523	0,482	5	6	5	4	5
O05	0,646	0,431	0,608	0,637	0,601	3	4	3	3	3
O06	0,284	0,190	0,256	0,306	0,276	7	8	8	7	8
O07	0,436	0,314	0,412	0,454	0,429	6	7	6	6	6
O08	0,225	0,429	0,289	0,225	0,287	8	5	7	8	7

Źródło: opracowanie własne

Ten przykład pokazuje, iż w raportowaniu powinno się pokazywać oprócz rankingu i wartości wskaźnika syntetycznego, także informację o stabilności dokonanej oceny. Jako miary stabilności oceny rankingu można użyć jednej z wielu dostępnych w literaturze miar zróżnicowania zbioru wektorów (zbiorów wielowymiarowych). Najlepiej, aby wartości tej miary były z przedziału $[0;1]$. W tym przypadku pomiar ten zilustrowano przy wykorzystaniu wskaźnika τ_{abs} , który wykorzystuje wartości wskaźnika zróżnicowania ar_{max} . Wskaźniki te należą do instrumentarium tzw. gradacyjnej analizy danych (por. [Szczesny 2002], [Kowalczyk i inni 2004], [Borkowski i inni 2008], [Binderman i inni 2014]).

Aby przybliżyć stosowane pojęcia wprowadzimy następujące oznaczenia, niech: $\mathbf{x} = (x_1, \dots, x_n)$ i $\mathbf{y} = (y_1, \dots, y_n)$ będą dwoma strukturami tzn. $x_i, y_i \geq 0$ oraz

$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = 1$. Przy tych oznaczeniach krzywą łączącą punkty:

$$(0;0), (x_1, y_1), (x_1 + x_2; y_1 + y_2), \dots, (x_1 + \dots + x_n; y_1 + \dots + y_n) = (1;1)$$

nazywa się zwykle krzywą zróżnicowania struktury \mathbf{y} względem struktury \mathbf{x} i oznacza symbolem $C_{[\mathbf{y}:\mathbf{x}]}$. Krzywa ta jest wykresem pewnej funkcji ciągłej $C_{[\mathbf{y}:\mathbf{x}]} : [0;1] \rightarrow [0;1]$. Wskaźnik zróżnicowania struktury \mathbf{y} względem struktury \mathbf{x} można określić jako:

$$ar(\mathbf{y} : \mathbf{x}) = ar(C_{[\mathbf{y}:\mathbf{x}]}) = 1 - 2 \int_0^1 C_{[\mathbf{y}:\mathbf{x}]}(t) dt \quad (2)$$

Łatwo zauważyć, iż wartość tego wskaźnika zależy od uporządkowania współrzędnych. W wielu przypadkach praktycznych nie ma zadanej ustalonej kolejności współrzędnych i w związku z tym zamiast wskaźnika ar używa się wskaźnika ar_{max} , który przyjmuje największą wartość ar dla wszystkich możliwych ustawień współrzędnych pary struktur \mathbf{y} i \mathbf{x} . Największa wartość jest przyjmowana przy ustawieniu wspomnianych par według wartości ilorazów y_i/x_i (por. [Kowalczyk i inni 2004]). Natomiast do oceny zróżnicowania zbioru wektorów $\{\mathbf{W}_1, \dots, \mathbf{W}_k\}$, z których każdy ma współrzędne równe wartościom odpowiedniego wskaźnika syntetycznego W_i ($i=1, \dots, k$) określonego na obiektach O_1, \dots, O_n użyjemy formuły:

$$\tau_{abs}(\{\mathbf{W}_1, \dots, \mathbf{W}_k\}) = \frac{1}{k(k-1)} \sum_{i=1}^k \sum_{j=1}^k [ar_{max}(\mathbf{x}_i : \mathbf{x}_j)], \quad (3)$$

gdzie: $\mathbf{W}_i = (w_{i1}, w_{i2}, \dots, w_{in}) = (W_i(O_1), \dots, W_i(O_n))$, $|\mathbf{W}_i| = \sum_{s=1}^n w_{is}$,

$$\mathbf{x}_i = \left(\frac{W_i(O_1)}{|\mathbf{W}_i|}, \dots, \frac{W_i(O_n)}{|\mathbf{W}_i|} \right), \quad \text{dla } i = 1, \dots, k$$

W przypadku danych zamieszczonych w Tabelach 2 i 3, wartości wskaźnika określonego wzorem (3) wynoszą odpowiednio 0,196 oraz 0,077. Obliczenia wartości można wykonać przy pomocy programu GradeStat (<http://gradestat.ipipan.waw.pl/>). Można posłużyć się również inną miarą niepodobieństwa wektorów, a mianowicie wstawić w miejsce wskaźnika ar_{max} we wzorze (3) inną miarę niepodobieństwa/odległości. Na przykład gdybyśmy posłużyli się zamiast ar_{max} miarą:

$$\mu := 0,5 \cdot (1 - \rho(W_i, W_j))\mu, \quad (4)$$

gdzie: $\rho(W_i, W_j)$ oznacza współczynnik korelacji Pearsona pomiędzy i-tym i j-tym wskaźnikiem syntetycznym, to otrzymalibyśmy odpowiednio wartości 0,308 i 0,62. W obu przypadkach widoczny jest wyraźny spadek zróżnicowania po zastosowaniu zmiany sposobu normowania zmiennych, polegającego na specjalnym potraktowaniu elementu odstającego. Oznacza to, iż po tym zabiegu uporządkowania mniej się różnią między sobą. W Tabeli 4 zamieszczono zróżnicowanie wartości utworzonych 5-ciu wskaźników syntetycznych między sobą. Łatwo zauważyć iż wskaźnik W_2 wyraźnie różni się od pozostałych. Ma to miejsce zarówno w przypadku gdy nie dokonaliśmy modyfikacji normowania jak i po modyfikacji. Przyczyną tego stanu rzeczy jest przyjęcie do budowy wag wskaźnika zmienności V dla zmiennych oryginalnych, co w praktyce oznacza wyeliminowanie w dalszej analizie zmiennych X_3 i X_5 (wagi są bardzo małe).

Tabela 4. Macierze wskaźników zróżnicowania ar_{max} dla wektorów wartości wskaźników syntetycznych odpowiednio z Tabel 2 i 3

	x_1	x_2	x_3	x_4	x_5
x_1	0	0,40	0,11	0,04	0,09
x_2	0,40	0	0,28	0,42	0,32
x_3	0,11	0,28	0	0,15	0,05
x_4	0,04	0,42	0,15	0	0,10
x_5	0,09	0,32	0,05	0,10	0

	x_1	x_2	x_3	x_4	x_5
x_1	0	0,16	0,03	0,02	0,03
x_2	0,16	0	0,13	0,17	0,14
x_3	0,03	0,13	0	0,05	0,02
x_4	0,02	0,17	0,05	0	0,03
x_5	0,03	0,14	0,02	0,03	0

Źródło: opracowanie własne

Zatem naturalnym postępowaniem wydaje się eliminacja wskaźnika W_2 . Po wyeliminowaniu W_2 notujemy duży procentowy spadek wartości wskaźnika τ_{abs} gdy dokonujemy modyfikacji normowania, kiedy przyjmuje on odpowiednio wartości 0,090 i 0,028. Reasumując, mała wartość wskaźnika zróżnicowania, a dokładniej jego wrażliwość na problem neutralizacji ("usuwania") elementów odstających oraz ewentualnych nietypowych wskaźników syntetycznych, wydaje się być dobrym ostrzeżeniem przed publikowaniem dyskusyjnego rankingu. Szczególnie, gdy jego konsekwencje w sytuacjach praktycznych mogą być nieprzyjemne, jak na przykład w systemie podziału funduszu premiowego pomiędzy jednostki biznesowe w dużej korporacji, bądź przy podziale środków budżetowych pomiędzy jednostki samorządowe itp. Zilustrujemy użyteczność zastosowania ta-

kiej miary w przypadku danych rzeczywistych zawartych w artykule prof. Kukuły [Kukuła 2014].

WYNIKI BADAŃ

W pracy [Kukuła 2014] do budowy rankingu województw ze względu na stopień zanieczyszczenia środowiska wykorzystano następujące 7 cech diagnostycznych: X_1 - nieoczyszczone ścieki przemysłowe i komunalne odprowadzone do wód lub do ziemi w $m^3/100 km^2$, X_2 - emisja zanieczyszczeń pyłowych z zakładów szczególnie uciążliwych dla środowiska w $t/100 km^2$, X_3 - emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych dla środowiska w $t/100 km^2$, X_4 - udział powierzchni gruntów zdewastowanych i zdegradowanych (wymagających rekultywacji) w ogólnej powierzchni województwa w %, X_5 - emisja ołowiu z zakładów szczególnie uciążliwych dla środowiska w $g/100 km^2$, X_6 - emisja rtęci z zakładów szczególnie uciążliwych dla środowiska w $g/100 km^2$, X_7 - wielkość odpadów komunalnych wytworzonych i niezabezpieczonych w $kg/osobę$.

Można podjąć próbę pokazania wrażliwości uporządkowania województw ze względu na nietypowe wartości badanych zmiennych (patrz Tabela 5). Wartości zmiennych dla wygody powtórzono za pracą [Kukuła 2014]. W ostatnich 3-ech wierszach zamieszczono wartości współczynników zmienności oraz dolną i górną granicę dzielące wartości typowe od nietypowych. Górną granicę wyznaczono według wzoru (1), a wartości nietypowe wyróżniono w treści tabeli:

Tabela 5. Wartości zmiennych X_1 - X_7

Województwo	X_1	X_2	X_3	X_4	X_5	X_6	X_7
dolnośląskie	31,58	19,48	80,41	4,05	22490	582	54
kujawsko-pomorskie	15,58	11,67	47,05	2,51	1246	384	59
lubelskie	0,80	8,35	21,20	1,25	2464	299	73
lubuskie	3,57	8,31	14,69	1,16	93	43	48
łódzkie	12,08	18,94	224,18	2,63	285	1603	105
małopolskie	30,96	25,88	69,69	2,59	2720	356	88
mazowieckie	71,99	12,98	78,30	1,15	2351	1108	102
opolskie	89,25	22,71	133,80	3,07	2072	2773	49
podkarpackie	5,04	9,51	18,83	0,95	824	235	28
podlaskie	0,25	4,55	7,33	1,39	163	1511	89
pomorskie	2,73	15,17	37,69	1,67	486	421	48
śląskie	532,71	85,66	331,19	3,91	289198	3116	32
świętokrzyskie	204,94	23,15	108,39	2,92	4329	632	38
warmińsko-mazurskie	2,48	4,90	6,26	1,97	4	4	86
wielkopolskie	2,68	15,56	56,15	3,29	660	3758	44
zachodniopomorskie	3,93	11,35	40,35	1,32	175	502	39
granica dolna	-55,73	-7,38	-79,59	-1,19	-3148,25	-1446,63	-22,88
granica górna	100,13	36,89	187,60	5,45	5933,75	3322,38	152,13
wsp. zmienności V	2,09	0,99	1,06	0,44	3,38	1,04	0,40

Źródło: opracowanie własne na podstawie [Kukuła 2014]

Tak jak opisano w poprzedniej rozdziale po normalizacji dokonanej techniką unitaryzacji zerowanej tworzymy cztery wskaźniki syntetyczne W_1 - W_4 wykorzystując odpowiednio: wagi jednakowe (1/7), proporcjonalne do wartości współczynników zmienności zmiennych po unitaryzacji zerowanej, proporcjonalne do odwrotności sumy współczynników korelacji z macierzy korelacji oraz do iloczynu tych dwu ostatnich wag. Wartości tych współczynników zamieszczono w Tabeli 6. Dodatkowo aby sprawdzić stabilność została dokonana normalizacja w ten sposób, iż dla największej wartości odstającej przyjęto wartość równą granicy dla największej wartości oraz o 5 jednostek mniejszą wartość dla drugiej mniejszej co do wielkości wartości odstającej, aby zachować uporządkowanie wartości. Otrzymane wskaźniki syntetyczne dla tak zmodyfikowanych danych oznaczono symbolami WN_1 - WN_4 i również podano w Tabeli 6. Zróżnicowanie wartości wskaźników W_1 - W_4 i WN_1 - WN_4 mierzone według wzoru (3) wynosi odpowiednio 0,120 oraz 0,064, co wskazuje iż mogą zajść spore zmiany w uporządkowaniu województw. Uzyskane rangi zawiera w Tabeli 7, która pokazuje duży wpływ wprowadzonej modyfikacji normowania na miejsce w rankingu w przypadku województw wielkopolskiego i świętokrzyskiego. W przypadku pozostałych województw różnice są niewielkie. Mała wartość wskaźnika 0,064 wskazuje na stabilność sporządzonych rankingów RN1-RN4. Gdy nie ma wskazań ekspertów co do wag poszczególnych wskaźników cząstkowych, to należałoby – naszym zdaniem – wybrać ranking RN4, bo wybrane wagi uwzględniają zarówno korelacje pomiędzy wskaźnikami cząstkowymi, jak i ich zmienność.

Tabela 6. Wartości zmiennych X1-X7

Województwo	W1	W2	W3	W4	WN1	WN2	WN3	WN4
dolnośląskie	0,291	0,197	0,311	0,217	0,528	0,517	0,495	0,490
kujawsko-pomorskie	0,179	0,107	0,213	0,132	0,261	0,244	0,276	0,257
lubelskie	0,123	0,070	0,177	0,104	0,198	0,185	0,250	0,228
lubuskie	0,060	0,034	0,082	0,049	0,079	0,068	0,100	0,087
łódzkie	0,406	0,253	0,485	0,313	0,515	0,472	0,577	0,527
małopolskie	0,275	0,172	0,336	0,214	0,456	0,427	0,487	0,453
mazowieckie	0,255	0,168	0,339	0,223	0,447	0,457	0,516	0,511
opolskie	0,355	0,250	0,369	0,272	0,614	0,635	0,573	0,600
podkarpackie	0,025	0,021	0,022	0,020	0,068	0,070	0,059	0,062
podlaskie	0,191	0,107	0,272	0,161	0,203	0,183	0,292	0,260
pomorskie	0,119	0,075	0,139	0,090	0,175	0,157	0,182	0,165
śląskie	0,834	0,913	0,741	0,854	0,849	0,877	0,742	0,787
świętokrzyskie	0,268	0,217	0,257	0,216	0,539	0,567	0,474	0,508
warmińsko-mazurskie	0,156	0,078	0,228	0,124	0,159	0,130	0,236	0,195
wielkopolskie	0,322	0,205	0,346	0,236	0,387	0,371	0,378	0,370
zachodniopomorskie	0,084	0,055	0,094	0,064	0,125	0,116	0,126	0,119

Źródło: opracowanie własne

Tabela 7. Rankingi Wartości województw według wartości wskaźników z Tabeli 6

Województwo	R1	R2	R3	R4	RN1	RN2	RN3	RN4
dolnośląskie	5	6	7	6	4	4	5	6
kujawsko-pomorskie	10	9	11	10	9	9	10	10
lubelskie	12	13	12	12	11	10	11	11
lubuskie	15	15	15	15	15	16	15	15
łódzkie	2	2	2	2	5	5	2	3
małopolskie	6	7	6	8	6	7	6	7
mazowieckie	8	8	5	5	7	6	4	4
opolskie	3	3	3	3	2	2	3	2
podkarpackie	16	16	16	16	16	15	16	16
podlaskie	9	10	8	9	10	11	9	9
pomorskie	13	12	13	13	12	12	13	13
śląskie	1	1	1	1	1	1	1	1
świętokrzyskie	7	4	9	7	3	3	7	5
warmińsko-mazurskie	11	11	10	11	13	13	12	12
wielkopolskie	4	5	4	4	8	8	8	8
zachodniopomorskie	14	14	14	14	14	14	14	14

Źródło: opracowanie własne

W badanym przypadku opisanym przez Tabelę 5 wynika, iż województwo śląskie jest elementem mocno odstającym w badanym zbiorze obiektów (bardzo duże wartości w przypadku 4-ech zmiennych, wartości te znacząco przekraczają granicę oddzielającą elementy statycznie uznawane jako odstające od pozostałych). Zatem można go usunąć z obliczeń i dokonać uporządkowania 15-tu województw nadając pozycje 2-16 i sprawdzić z każdym poprzednich uporządkowań. Jeśli przyjmiemy jako miarę podobieństwa współczynnik korelacji, to jego wartości dla tych 4-ech nowych rankingów i kolumn R1-R5 z Tabeli 7 wynoszą odpowiednio: 0,959; 0,944; 0,918; 0,909 oraz 0,968; 0,968; 0,971; 0,953 dla kolumn RN1-RN4. Zatem podobieństwo tych nowych rankingów z rankingami RN1-RN4 jest większe niż z rankingami R1-R5. Jest to dodatkowy argument, że opisane postępowanie w procesach normowania z elementami o wartościach odstających jest właściwe. Pozostaje zatem ustalenie, który z tych czterech rankingów (RN1-RN4) ma zostać upubliczniony. Wydaje się, iż przydatnym narzędziem powinna być macierz podobieństw (lub odległości) pomiędzy wektorami WN1-WN4 lub RN1-RN4. Łatwo można sprawdzić, iż jeśli oprzemy się na macierzach korelacji, to w obu przypadkach najlepszym wyborem będzie ranking RN4 (kryterium maksymalna wartość sumy w wierszu macierzy korelacji). W przypadku użycia macierzy niepodobieństw wektorów WN1-WN4 przy wykorzystaniu wskaźnika ar_{max} , też uzyskamy identyczną odpowiedź.

PODSUMOWANIE

Problem stworzenia stabilnego i nie budzącego zastrzeżeń rankingu obiektów jest w dobie licznych systemów raportowych ważnym zagadnieniem. Dodatkowo w wielu zastosowaniach praktycznych (np. w przypadku jednostek terytorialnych, oddziałach dużych korporacji) po upublicznieniu takie rankingi są dokładnie oglądane i komentowane. Dlatego przed wypuszczeniem konkretnego rankingu warto sprawdzić jego stabilność. Zazwyczaj publikuje się nie tylko samo miejsce w rankingu danego obiektu, ale także wartości wskaźnika syntetycznego, które są podstawą publikowanego rankingu. Jedną z najczęstszych stosowanych metod normalizacyjnych przy sprowadzaniu do porównywalności jest metoda unitaryzacji zerowanej. W badaniu wybraliśmy tę metodę normalizacji z uwagi na fakt, iż sprowadza ona wartości fizyczne zmiennych do wartości bezwymiarowych z przedziału $[0;1]$, co nadaje im najbardziej intuicyjny charakter z praktycznego punktu widzenia. Najczęściej jako syntetyczną ocenę wybiera się średnią lub średnią ważoną z wartości wskaźników częściowych wziętych do budowy rankingu, gdyż wtedy w naturalny sposób wartości wskaźnika syntetycznego należą do przedziału $[0;1]$. Wybór wag to ważny proces, ale w tej pracy z uwagi na ograniczoność miejsca nie był szerzej dyskutowany. Wykorzystaliśmy tylko najbardziej popularne techniki, które w sumie posłużyły do oceny stabilności uporządkowań obiektów za pomocą wartości 4-ech wskaźników syntetycznych. Jako ocenę stabilności rankingu przyjęliśmy wybraną przykładową miarę zróżnicowania wektorów. Po zastosowaniu modyfikacji w procesie normalizacji dotyczącej elementów uznanych za odstające, miara ta wskazuje, że różnice te przy różnych technikach tworzenia wag są niewielkie (wartość przyjętego wskaźnika zróżnicowania poniżej 0,1) czyli można uznać iż rankingi zachowują się stabilnie. Na zakończenie należy podkreślić, że w badaniu stabilności rankingu powinny być użyte także inne techniki normowania wskaźników częściowych. Z uwagi na ograniczoność miejsca, ten ważny problem został w pracy pominięty.

BIBLIOGRAFIA

- 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*, Vol. 11, pp. 289–301.
- Betti G., Soldi R., Talev I. (2015) Fuzzy Multidimensional indicators of quality of life: the empirical case of Macedonia, *Social Indicators Research*, doi:10.1007/s11205-015-0965-y.
- Borkowski B., Dudek H., Szczesny W. (2008) Spatial differentiation of food production structure and consumption profile in enlarged European Union, *Journal of Economic Asymmetries*, Vol. 5, No. 2, 145-156.

- Binderman Z., Koszela G., Szczesny W. (2014) Zmiany w strukturze gospodarstw rolnych w krajach Unii Europejskiej w latach 2003-2010 (aspekty metodyczne), Problemy rolnictwa światowego, T14 z. 3, Wydawnictwo SGGW, Warszawa.
- Gatnar E., Walesiak M. (2009) Statystyczna analiza danych z wykorzystaniem programu R, PWN, Warszawa.
- Kowalczyk T., Pleszczyńska E., Ruland F. (2004) Grade Models and Methods of Data Analysis. With applications for the Analysis of Data Population, Studies in Fuzziness and Soft Computing, Vol. 151, Springer Verlag, Berlin - Heidelberg - New York.
- Kukuła K. (2000) Metoda unitaryzacji zerowanej, PWN, Warszawa.
- Kukuła K. (2014) Wybrane problemy ochrony środowiska w Polsce w świetle wielowymiarowej analizy porównawczej, Metody Ilościowe w Badaniach Ekonomicznych, Tom XV/3, str. 169 – 188.
- Malina A. (2004) Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw, AE, Seria Monografie nr 162, Kraków.
- Młodak A. (2006) Analiza taksonomiczna w statystyce regionalnej, DIFIN, Warszawa.
- Panek T. (2011) Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, Warszawa.
- Szczesny W. (2002) Grade correspondence analysis applied to contingency tables and questionnaire data, Intelligent Data Analysis, Vol. 6(1), IOS Press, Amsterdam.
- Zeliaś A. (2000) Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym, AE Kraków.

THE USE OF MULTIDIMENSIONAL COMPARATIVE ANALYSIS TOOLS TO ASSESS THE LEVEL OF ENVIRONMENTAL POLLUTION IN SPATIAL TERMS USE OF TOOLS

Abstract: The paper discusses the problem of assessing the ranking stability before publicizing on the example of the environmental pollution evaluation in Polish regions. It was proposed the stability assessment process with the use of measures of set of vectors diversity. The paper is a complement to the results and discussion obtained by [Kukuła 2014].

Keywords: ranking, diversity index, a synthetic index, grade data analysis, GradeStat program