

IWONA MÜLLER-FRĄCZEK¹

PROPOZYCJA MIARY SYNTETYCZNEJ

1. WSTĘP

Analiza wielu zagadnień życia społeczno-gospodarczego wymaga uporządkowania liniowego obiektów wielocechowych. Jedną z możliwości rozwiązania tego problemu jest wykorzystanie zmiennych syntetycznych, które są jednowymiarowym obrazem niemierzalnego zjawiska złożonego, reprezentowanego przez wiele zmiennych diagnostycznych. Sztuką jest skonstruowanie „dobrej” zmiennej syntetycznej, a nauka polska ma ogromne osiągnięcia w tej dziedzinie.

Po pionierskiej propozycji Hellwiga (1968) powstało wiele jej rozwinięć, jak również nowatorskich pomysłów. Wśród nich warto wymienić prace (chronologicznie): Cieślak (1974), Dittmann, Pisz (1975), Bartosiewicz (1976), Pluta (1976), Nowak (1977), Strahl (1978), Panek (1984), Jajuga (1986), Lira i inni (2002), Strahl (2006), Strahl, Markowska (2008), Młodak (2009), Walesiak (2011), Wydymus (2013), Pietrzak (2014).

Początkowo miary syntetyczne stosowano do badania poziomu rozwoju gospodarczego obiektów, stąd ich nazwa mierniki rozwoju. Z czasem zaczęto je wykorzystywać do charakteryzowania innych zjawisk złożonych.

Celem artykułu było zaproponowanie nowej metody konstrukcji zmiennej syntetycznej. Aby ocenić poziom rozwoju obiektu, analizowano jego odległość od wzorca (antywzorca) względem sumy odległości od wzorca (antywzorca) innych obiektów. W ten sposób scharakteryzowano obiekt na tle wszystkich pozostałych². Powstała w ten sposób miara jest ograniczona. Zmienne diagnostyczne nie wymagają normalizacji ani zamiany na stymulanty. Podstawową jej zaletą jest porównywalność analiz prowadzonych dla różnych lat.

Zaproponowaną metodę zilustrowano badaniem empirycznym dotyczącym stanu ochrony środowiska. Wartości miar porównano z wynikami otrzymanymi na podstawie

¹ Uniwersytet Mikołaja Kopernika w Toruniu, Wydział Nauk Ekonomicznych i Zarządzania, Katedra Ekonometrii i Statystyki, ul. Gagarina 13a, 87-100 Toruń, Polska, e-mail: muller@econ.umk.pl.

² W pracy Wydymusa (2013) przedstawiono inną metodę odzwierciedlenia w zmiennej syntetycznej wpływu wszystkich obiektów na sytuację obiektu badanego. Dla każdej cechy analizowano ilorazy jej wartości dla różnych obiektów. Powstałe w ten sposób macierze zrelatywizowanych wskaźników były podstawą konstrukcji miernika.

najbardziej rozpowszechnionej metody wzorcowej – Hellwiga. Były one podobne, ale nie tożsame.

W kolejnym kroku porównano obie metody pod względem poprawności. Co ciekawe w różnych analizowanych aspektach wyraźną przewagę uzyskiwała zawsze jedna z metod.

2. KONSTRUKCJA MIERNIKA

Załóżmy, że dany jest zestaw zmiennych diagnostycznych charakteryzujących badane zjawisko, który spełnia zarówno kryteria merytoryczne jak i statystyczne. W zestawie tym mogą występować stymulanty, destymulanty oraz nominanty, dla których znana jest wartość nominalna. Proponowana w artykule miara syntetyczna składa się z dwóch części. Jedna z nich charakteryzuje obiekt względem wzorca, druga – antywzorca.

Oznaczmy przez x_{ip} wartość p -tej zmiennej ($p = 1, \dots, r$) dla i -tego obiektu ($i = 1, \dots, n$). Oznaczmy przez S , D oraz N zbiory indeksów odpowiednio dla stymulant, destymulant oraz nominant.

Przyjmijmy jednakowy dla wszystkich obiektów wzorec $(x_1^+, x_2^+, \dots, x_r^+)$ postaci:

$$x_p^+ = \begin{cases} \max_i x_{ip} & \text{dla } p \in S, \\ \min_i x_{ip} & \text{dla } p \in D, \\ x_p^{nom} & \text{dla } p \in N. \end{cases} \quad (1)$$

Określmy również antywzorec $(x_1^-, x_2^-, \dots, x_r^-)$. W przypadku, gdy w zestawie cech nie ma nominant, będzie on miał jednoznaczną postać dla wszystkich obiektów:

$$x_p^- = \begin{cases} \min_i x_{ip} & \text{dla } p \in S, \\ \max_i x_{ip} & \text{dla } p \in D. \end{cases}$$

Natomiast, gdy w zestawie cech występuje nominanta, dzielimy obiekty na dwie grupy względem wartości nominalnej i określimy antywzorec inaczej dla każdej grupy:

$$x_p^- = \begin{cases} \min_i x_{ip} & \text{dla } p \in N \wedge x_{ip} < x_p^{nom}, \\ \max_i x_{ip} & \text{dla } p \in N \wedge x_{ip} > x_p^{nom}. \end{cases}$$

W przypadku obiektu, dla którego cecha osiąga dokładnie wartość nominalną, za antywzorcową przyjmujemy wartość, która jest od niego najbardziej odległa.

Popularne metody wzorcowe różnią się zarówno sposobem wyboru wzorca, jak i metodą agregacji zmiennych. Jednak najczęściej wykorzystuje się różne wer-

sje pomiaru odległości obiektu od wzorca w wielowymiarowej przestrzeni cech. Proponowana metoda konstrukcji oparta jest na analizie jednowymiarowej.

W przestrzeni p -tej zmiennej diagnostycznej rozważmy udział odległości od wzorca i -tego obiektu w łącznej odległości od wzorca wszystkich obiektów:

$$m_{ip}^+ = \frac{|x_{ip} - x_p^+|}{\sum_{j=1}^n |x_{jp} - x_p^+|} \quad \text{dla } p \in S \cup D \cup N.$$

Podobnie określamy:

$$m_{ip}^- = \begin{cases} \frac{|x_{ip} - x_p^-|}{\sum_{j=1}^n |x_{jp} - x_p^-|} & \text{dla } p \in S \cup D, \\ \frac{n_{ip}}{n} \cdot \frac{|x_{ip} - x_p^-|}{\sum_j |x_{jp} - x_p^-|} & \text{dla } p \in N. \end{cases}$$

Występująca w wersji dla nominant wielkość n_{ip} oznacza liczbę obiektów, które mają tę co obiekt i wartość antywzorcową. Tylko te obiekty bierzemy pod uwagę przy sumowaniu w mianowniku tego wzoru.

Zauważmy, że sytuacja obiektu jest tym lepsza, im niższa jest wielkość m_{ip}^+ oraz im wyższe jest m_{ip}^- . Oznaczmy ich średnie wartości:

$$m_i^+ = -\frac{1}{r} \sum_{p=1}^r m_{ip}^+,$$

$$m_i^- = \frac{1}{r} \sum_{p=1}^r m_{ip}^-.$$

Należy podkreślić, że miary m_i^+ oraz m_i^- charakteryzują „położenie” obiektu nie tylko indywidualnie względem wzorca czy antywzorca, ale odzwierciedlają jego sytuację na tle wszystkich analizowanych obiektów. W przypadku większości tego typu konstrukcji wpływ sytuacji innych obiektów na wartość miary syntetycznej jest niewielki, uwzględniony wyłącznie w postaci parametrów rozkładu wykorzystanych do normalizacji cech.

Ostatecznie miernik rozwoju i -tego obiektu definiujemy w postaci:

$$m_i = \frac{1 + m_i^- + m_i^+}{2}. \quad (2)$$

3. WŁASNOŚCI MIARY

Miara syntetyczna opisana formułą (2) spełnia wymagany dla tego typu konstrukcji postulat – poziom rozwoju obiektu jest tym wyższy, im większa jest wartość miernika. Miara ta jest niemianowana i ograniczona do przedziału $[0,1]$. Jej średnia wartość nie zależy od konkretnych danych, zawsze wynosi 0,5. Porównywanie mierników z tą wielkością posiada intuicyjną interpretację. Jeśli wartość miernika jest większa od 0,5 oznacza to, że obiektowi jest relatywnie (tzn. na tle innych obiektów) bliżej do wzorca niż antywzorca, natomiast wartości miernika mniejsze od 0,5 świadczą o tym, że obiekt relatywnie leży bliżej antywzorca niż wzorca.

Główną zaletą opisanej konstrukcji jest porównywalność wyników w czasie. Wzrost wartości miernika świadczy o tym, że na tle wszystkich obiektów sytuacja badanej jednostki uległa poprawie, natomiast spadek informuje o pogorszeniu jej sytuacji.

Inne własności:

1. $m_i^- \in [0,1]$.
2. $m_i^+ \in [-1,0]$.
3. $\sum_{i=1}^n m_i^- = 1$.
4. $\sum_{i=1}^n m_i^+ = -1$.
5. $\overline{m^-} = \frac{1}{n} \sum_{i=1}^n m_i^- = \frac{1}{n}$.
6. $\overline{m^+} = \frac{1}{n} \sum_{i=1}^n m_i^+ = -\frac{1}{n}$.
7. $\sum_{i=1}^n m_i = \frac{n}{2}$.

Przypadki szczególne:

1. Miara m_i^+ przyjmuje największą wartość równą 0, gdy dla wszystkich zmiennych diagnostycznych obiekt i był wzorcem.
2. Miara m_i^- przyjmuje najmniejszą wartość równą 0, gdy dla wszystkich zmiennych diagnostycznych obiekt i był antywzorcem.
3. Miary m_i^+ oraz m_i^- nie mogą jednocześnie być zerami.
4. $m_i^+ = -1$, gdy dla wszystkich zmiennych tylko obiekt i nie jest wzorcem.
5. $m_i^- = 1$, gdy dla wszystkich zmiennych tylko obiekt i nie jest antywzorcem.

Należy podkreślić, że proponowana konstrukcja nie opiera się bezpośrednio na zmiennych diagnostycznych, lecz na wielkościach m_{ip}^+ oraz m_{ip}^- . Ponieważ są one porównywalne (niemianowane i ograniczone do tego samego przedziału), nie zachodzi potrzeba normalizacji zmiennych diagnostycznych³. Ponadto konstrukcja ma dwie inne zalety techniczne:

- nie ma potrzeby przekształcania destymulant na stymulanty,
- cecha może przyjmować wartość 0.

³ Normalizacja zmiennych jest szczególnie dyskusyjna w przypadku analiz prowadzonych w różnych latach. Sprowadzenie zmiennych do porównywalności może powodować na przykład utratę unormowania miary syntetycznej. Niektóre problemy związane z normalizacją rozwiązuje wykorzystanie tzw. mediany Webera (por. Lira i inni, 2002). Więcej formuł normalizacyjnych można znaleźć np. w pracy Walesiaka (2014).

4. SYNTETYCZNA ANALIZA STANU OCHRONY ŚRODOWISKA W POLSCE

W celu ilustracji przydatności zaproponowanej konstrukcji miary syntetycznej analizowano zróżnicowanie województw pod względem stanu ochrony środowiska w latach 2006–2015. Badanie przeprowadzono w oparciu o dane dostępne w statystyce publicznej (Bank Danych Lokalnych). Wstępny zestaw zmiennych diagnostycznych charakteryzujących stan ochrony środowiska dobrano na podstawie pracy Młodaka (2006). Były to:

1. ścieki odprowadzone bezpośrednio do wód lub do ziemi w dm^3 na 1 km^2 (D_1),
2. udział ścieków oczyszczonych w ściekach wymagających oczyszczenia (S_2),
3. emisja pyłowych zanieczyszczeń powietrza na 1 km^2 w tonach na rok (D_3),
4. emisja gazowych zanieczyszczeń powietrza na 1 km^2 w tonach na rok,
5. pyłowe zanieczyszczenia zatrzymane lub zneutralizowane w urządzeniach do redukcji zanieczyszczeń w % zanieczyszczeń wytworzonych,
6. gazowe zanieczyszczenia zatrzymane lub zneutralizowane w urządzeniach do redukcji zanieczyszczeń w % zanieczyszczeń wytworzonych (S_4),
7. odpady wytworzone w ciągu roku na jednego mieszkańca w tonach (D_5),
8. odpady wykorzystane i unieszkodliwione w % wytworzonych,
9. wskaźnik lesistości w % (S_6),
10. powierzchnia obszarów prawnie chronionych w % powierzchni ogółem (S_7),
11. nakłady inwestycyjne na ochronę środowiska i gospodarkę wodną w % powierzchni ogółem (S_8),
12. ludność obsługiwana przez oczyszczalnie ścieków w % ludności ogółem (S_9).

Po przeprowadzeniu analizy zmiennościowo-korelacyjnej wstępny zestaw zmiennych zredukowano do dziewięciu, wśród których znalazło się sześć stymulant (oznaczono je w spisie literą S) oraz trzy destymulanty (oznaczone literą D).

Wyselekcjonowane zmienne były podstawą wyznaczenia miar opisanych wzorem (2). Wartości obliczonych mierników dla poszczególnych lat oraz odpowiadające im pozycje w rankingach przedstawiono w tabeli 1. Pozwoliły one dokonać syntetycznej oceny stanu ochrony środowiska w województwach Polski, jak również ocenić zmiany jakie zaszły w tym zakresie na przestrzeni dziesięciu lat.

Tabela 1.

Mierniki obliczone zgodnie z formułą (2) oraz odpowiadające im pozycje w rankingach

Województwo	Miernik	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
DŚL	Wartość	0,498	0,494	0,492	0,490	0,484	0,490	0,489	0,491	0,499	0,497
	Pozycja	12	12	13	13	14	13	13	13	10	11
K-P	Wartość	0,501	0,502	0,504	0,504	0,503	0,501	0,501	0,496	0,498	0,502
	Pozycja	10	8	6	6	7	9	8	10	11	10

Tabela 1. (cd.)

Województwo	Miernik	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
LBL	Wartość	0,502	0,500	0,498	0,497	0,497	0,499	0,499	0,496	0,498	0,493
	Pozycja	9	10	10	12	12	12	10	11	12	12
LBU	Wartość	0,523	0,523	0,524	0,522	0,521	0,522	0,521	0,519	0,523	0,530
	Pozycja	2	2	1	2	2	2	2	2	2	1
ŁDZ	Wartość	0,502	0,501	0,501	0,502	0,501	0,509	0,504	0,499	0,491	0,486
	Pozycja	8	9	9	9	9	5	7	9	13	14
MLP	Wartość	0,500	0,503	0,502	0,502	0,501	0,503	0,501	0,505	0,507	0,507
	Pozycja	11	7	8	8	8	7	9	6	5	4
MAZ	Wartość	0,481	0,484	0,491	0,489	0,487	0,484	0,487	0,487	0,488	0,491
	Pozycja	14	14	14	14	13	14	14	14	14	13
OPO	Wartość	0,510	0,517	0,509	0,497	0,499	0,499	0,508	0,516	0,515	0,516
	Pozycja	3	3	4	10	10	11	5	3	3	3
PKR	Wartość	0,508	0,513	0,515	0,513	0,516	0,512	0,515	0,515	0,509	0,507
	Pozycja	6	4	3	4	3	4	3	4	4	6
PDL	Wartość	0,510	0,508	0,508	0,504	0,507	0,505	0,506	0,504	0,505	0,504
	Pozycja	4	6	5	7	6	6	6	7	6	9
POM	Wartość	0,523	0,525	0,523	0,526	0,528	0,526	0,526	0,527	0,526	0,526
	Pozycja	1	1	2	1	1	1	1	1	1	2
ŚL	Wartość	0,463	0,465	0,469	0,472	0,461	0,456	0,462	0,465	0,461	0,460
	Pozycja	16	16	16	15	16	16	16	16	16	16
ŚW	Wartość	0,472	0,468	0,471	0,465	0,471	0,479	0,477	0,475	0,474	0,462
	Pozycja	15	15	15	16	15	15	15	15	15	15
W-M	Wartość	0,509	0,511	0,504	0,514	0,514	0,512	0,509	0,507	0,502	0,506
	Pozycja	5	5	7	3	4	3	4	5	7	7
WLP	Wartość	0,495	0,493	0,496	0,497	0,498	0,499	0,497	0,495	0,502	0,507
	Pozycja	13	13	11	11	11	10	12	12	8	5
ZPM	Wartość	0,503	0,494	0,495	0,506	0,513	0,503	0,497	0,502	0,501	0,506
	Pozycja	7	11	12	5	5	8	11	8	9	8

Źródło: opracowanie własne na podstawie danych GUS.

Analiza wyników zawartych w tabeli 1. nasuwa spostrzeżenie, że w badanym okresie w większości województw odnotowano wartość miary syntetycznej charak-

teryzującej stan ochrony środowiska powyżej neutralnego progu 0,5. Oznacza to, że więcej było regionów, którym relatywnie bliżej było do wzorca niż antywzorca. Wyjątek stanowił rok 2013, w którym tyle samo było województw, dla których wartość miernika była powyżej 0,5, jak tych dla których była poniżej tego progu. W pięciu województwach sytuacja była korzystna przez cały okres badania. Byli to liderzy klasyfikacji: pomorskie i lubuskie oraz województwa: podkarpackie, podlaskie i warmińsko-mazurskie. Dla czterech województw przez cały okres badania miara była niższa niż wartość neutralna. Były to województwa: śląskie, świętokrzyskie i mazowieckie (najgorsze w klasyfikacji) oraz dolnośląskie.

Rozstęp wartości mierników w badanym okresie ulegał zmianom, jednak w każdym roku rozstęp dolny (wartości mniejsze niż 0,5) wyraźnie przeważał nad rozstępem górnym (wartości większe niż 0,5). Różnice między liderami klasyfikacji były niewielkie, natomiast różnica między województwem ostatnim a przedostatnim była spora, w 2011 roku wynosiła nawet połowę dolnego rozstępu.

Największą poprawę na przestrzeni 10 lat zarówno pod względem wartości miernika jak i awansu w klasyfikacji odnotowano w województwie wielkopolskim oraz małopolskim. Natomiast w województwie mazowieckim poprawa pozycji była minimalna mimo, że wartość miernika wzrosła równie silnie jak w wielkopolskim. Najbardziej znacząca obniżka wartości miernika odnotowano w łódzkim, co przełożyło się na znaczny spadek pozycji tego województwa w rankingu.

W analizie rok do roku dużą zmienność sytuacji można zaobserwować w województwach opolskim i zachodniopomorskim, osiągały zarówno największe w danym roku spadki jak i wzrosty wartości mierników. Skutkowało to silnymi zmianami pozycji w klasyfikacji.

5. OCENA JAKOŚCI METODY

Jednym ze sposobów oceny jakości proponowanej miary syntetycznej jest porównanie wyników badania empirycznego z wynikami uzyskanymi w oparciu o inną miarę. W artykule wykorzystano w tym celu miernik zaproponowany przez Hellwiga (1968).

W pierwszym kroku metody Hellwiga zmienne poddaje się standaryzacji, po czym analizuje się odległość obiektów od wzorca postaci (1). Miernik rozwoju określony jest wówczas wzorem:

$$m_i = 1 - \frac{d_i^+}{\bar{d} + 2s_d}, \quad (3)$$

w którym

d_i^+ jest r -wymiarową odległością euklidesową i -tego obiektu od wzorca, \bar{d} , s_d są odpowiednio średnią i odchyleniem standardowym tej odległości.

Wyniki analizy stanu ochrony środowiska z wykorzystaniem mierników Hellwiga przedstawiono w tabeli 2.

Tabela 2.

Mierniki obliczone zgodnie z formułą (3) oraz odpowiadające im pozycje w rankingach

Województwo	Miernik	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
DŚL	Wartość	0,22	0,21	0,20	0,18	0,14	0,16	0,13	0,16	0,20	0,24
	Pozycja	11	10	12	11	12	12	12	13	11	12
K-P	Wartość	0,24	0,24	0,27	0,28	0,24	0,23	0,21	0,19	0,23	0,29
	Pozycja	9	8	6	8	8	10	11	11	10	8
LBL	Wartość	0,12	0,12	0,08	0,08	0,11	0,15	0,13	0,09	0,12	0,12
	Pozycja	13	13	14	15	14	13	13	14	14	14
LBU	Wartość	0,40	0,39	0,42	0,39	0,33	0,39	0,35	0,35	0,42	0,57
	Pozycja	1	2	2	2	4	2	3	4	2	1
ŁDZ	Wartość	0,19	0,19	0,20	0,22	0,20	0,26	0,24	0,21	0,15	0,16
	Pozycja	12	12	11	10	11	9	7	8	13	13
MLP	Wartość	0,30	0,33	0,31	0,31	0,27	0,33	0,29	0,33	0,37	0,42
	Pozycja	5	5	5	5	7	4	6	5	3	3
MAZ	Wartość	0,07	0,10	0,14	0,15	0,13	0,09	0,09	0,18	0,20	0,26
	Pozycja	14	14	13	13	13	15	14	12	12	11
OPO	Wartość	0,34	0,37	0,33	0,28	0,28	0,29	0,34	0,37	0,36	0,40
	Pozycja	3	3	4	7	6	6	4	3	4	4
PKR	Wartość	0,29	0,34	0,40	0,36	0,39	0,38	0,38	0,43	0,34	0,39
	Pozycja	6	4	3	3	3	3	2	2	5	6
PDL	Wartość	0,25	0,23	0,23	0,17	0,22	0,19	0,22	0,21	0,23	0,27
	Pozycja	8	9	10	12	10	11	10	9	8	10
POM	Wartość	0,40	0,44	0,50	0,51	0,50	0,49	0,46	0,49	0,48	0,55
	Pozycja	2	1	1	1	1	1	1	1	1	2
ŚL	Wartość	0,03	0,08	0,06	0,08	0,03	0,00	0,02	0,05	-0,01	0,05
	Pozycja	15	15	16	14	16	16	16	16	16	15
ŚW	Wartość	0,02	0,00	0,06	0,04	0,09	0,10	0,08	0,07	0,08	0,05
	Pozycja	16	16	15	16	15	14	15	15	15	16
W-M	Wartość	0,25	0,28	0,23	0,30	0,30	0,28	0,29	0,29	0,23	0,29
	Pozycja	7	7	9	6	5	7	5	7	9	9
WLP	Wartość	0,23	0,20	0,25	0,24	0,23	0,28	0,23	0,21	0,30	0,37
	Pozycja	10	11	8	9	9	8	9	10	7	7
ZPM	Wartość	0,32	0,29	0,27	0,36	0,40	0,33	0,24	0,30	0,31	0,39
	Pozycja	4	6	7	4	2	5	8	6	6	5

Źródło: opracowanie własne na podstawie danych GUS.

W celu porównania zgodności wyników syntetycznego badania stanu ochrony środowiska, uzyskanych na podstawie obu metod, wykorzystano współczynnik korelacji rang Spearmana oraz (opisany dalej) miernik, który zaproponował Nowak (1982).

Przez d_{ij}^k oznaczymy dystans między i -tym a j -tym obiektem w jednowymiarowej przestrzeni k -tej zmiennej syntetycznej ($k = 1, 2$). Ponieważ dla różnych metod wielkości te są nieporównywalne, dokonujemy normalizacji:

$$c_{ij}^k = \frac{d_{ij}^k}{d_{k0}}, \quad k = 1, 2$$

przy czym

$$d_{k0} = \max_{i,j} d_{ij}^k.$$

W kolejnym kroku dla wszystkich obiektów obliczamy różnice unormowanych odległości:

$$u_{ij} = |c_{ij}^1 - c_{ij}^2|.$$

Miernik zgodności dwóch metod tworzenia taksonomicznych mierników rozwoju ma wówczas postać:

$$M0 = 1 - \frac{2 \sum_{i=1}^{n-1} \sum_{j>i}^n u_{ij}}{n^2 - n} \in [0,1]. \quad (4)$$

Jest on tym większy, im bardziej zgodne wyniki dają porównywane metody.

Obliczone wartości współczynników korelacji rang Spearmana oraz mierników $M0$ dla kolejnych lat badania przedstawiono w tabeli 3.

Tabela 3.

Zgodność miar określonych równaniami (2) oraz (3) scharakteryzowana współczynnikiem korelacji rang Spearmana r oraz miernikiem $M0$ określonym wzorem (4)

Rok	2007	2009	2011	2012	2013	2014	2015
r	0,835	0,897	0,859	0,885	0,906	0,838	0,903
$M0$	0,867	0,875	0,878	0,884	0,885	0,879	0,878

Źródło: opracowanie własne na podstawie danych zawartych w tabeli 1 oraz tabeli 2.

Analizując wyniki zawarte w tabeli można zauważyć, że w kolejnych latach wartości obu mierników zgodności zmieniały się, jednak w żadnym roku metody nie były tożsame. Oznacza to, że proponowana konstrukcja nie tylko różni się formalnie od miernika Hellwiga, ale również inaczej niż on odzwierciedla rzeczywistość.

W celu analizy poprawności zaproponowanej konstrukcji wykorzystano mierniki opisane w pracach Grabiński (1984) oraz Grabiński i inni (1989). Im mniejsze są wartości tych mierników, tym bardziej poprawna jest zmienna syntetyczna. Charakteryzują one następujące aspekty:

1. Zgodność odwzorowania:

$$M1 = \frac{\sum_{i=1}^{n-1} \sum_{j>i}^n (d_{ij} - \delta_{ij})^2}{\sum_{i=1}^{n-1} \sum_{j>i}^n \delta_{ij}^2},$$

$$M2 = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^n \left(\frac{d_{ij} - \delta_{ij}}{\delta_{ij}} \right)^2,$$

$$M3 = \frac{\sum_{i=1}^{n-1} \sum_{j>i}^n \frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}}}{\sum_{i=1}^{n-1} \sum_{j>i}^n \delta_{ij}},$$

gdzie:

d_{ij} – dystans między obiektami w jednowymiarowej przestrzeni zmiennej syntetycznej,

δ_{ij} – dystans między obiektami w r -wymiarowej przestrzeni zmiennych diagnostycznych,

n – liczba obiektów.

2. Korelację liniową zmiennej syntetycznej ze zmiennymi diagnostycznymi:

$$M4 = 1 - \frac{1}{r} \sum_{p=1}^r r_p,$$

$$M5 = \frac{1}{r} \sum_{p=1}^r l(r_p),$$

gdzie:

r_p – współczynnik korelacji między p -tą zmienną diagnostyczną a zmienną syntetyczną,

r – liczba zmiennych,

$$l(r_p) = \begin{cases} 0 & \text{dla } r_p \in [0,5; 1], \\ 1 & \text{dla } r_p \in [0; 0,5), \\ 2 & \text{dla } r_p \in [-0,5; 0), \\ 3 & \text{dla } r_p \in [-1; -0,5). \end{cases}$$

3. Korelację rangową zmiennej syntetycznej ze zmiennymi diagnostycznymi:

$$M6 = 1 - \frac{1}{r} \sum_{p=1}^r \rho_p,$$

$$M7 = \frac{1}{r} \sum_{p=1}^r l(\rho_p),$$

$$M8 = \frac{2}{rl} \sum_{i=1}^n \sum_{p=1}^r |\dot{x}_{ip} - \dot{q}_i|,$$

gdzie:

ρ_p – współczynnik korelacji rangowej między p -tą cechą a zmienną syntetyczną,

$$l(\rho_p) = \begin{cases} 0 & \text{dla } \rho_p \in [0,5; 1], \\ 1 & \text{dla } \rho_p \in [0; 0,5), \\ 2 & \text{dla } \rho_p \in [-0,5; 0), \\ 3 & \text{dla } \rho_p \in [-1; -0,5), \end{cases}$$

\dot{x}_{ip} – ranga i -tego obiektu ze względu na p -tą zmienną diagnostyczną,

\dot{q}_i – ranga i -tego obiektu ze względu na zmienną syntetyczną,

$$l = \begin{cases} n^2 & \text{dla } n \text{ parzystych,} \\ n^2 - 1 & \text{dla } n \text{ nieparzystych.} \end{cases}$$

4. Zmienność i koncentrację zmiennej syntetycznej:

$$M9 = -\frac{s_m}{\bar{m}},$$

$$M10 = \frac{s_\Delta}{\bar{\Delta}},$$

gdzie:

\bar{m} , s_m – średnia i odchylenie standardowe zmiennej syntetycznej,

$\bar{\Delta}$, s_Δ – średnia i odchylenie standardowe z wielkości $\Delta_k = \tilde{m}_k - \tilde{m}_{k-1}$, przy czym \tilde{m}_k to uporządkowane niemalejąco realizacje zmiennej syntetycznej, $k = 1, 2, \dots, n$.

5. Odległość taksonomiczną zmiennej syntetycznej od zmiennych pierwotnych:

$$M11 = \frac{1}{nr} \sum_{i=1}^n \sum_{p=1}^r |x'_{ip} - q'_i|,$$

$$M12 = \left[\frac{1}{nr} \sum_{i=1}^n \sum_{p=1}^r (x'_{ip} - q'_i)^2 \right]^{1/2},$$

gdzie:

x'_{ip} – standaryzowana wartość p -tej zmiennej pierwotnej dla i -tego obiektu,

q'_i – standaryzowana wartość zmiennej syntetycznej dla i -tego obiektu.

W tabeli 4 przedstawiono wartości mierników poprawności metody (pierwsza kolumna) obliczone dla proponowanej konstrukcji oraz dla miary Hellwiga.

Tabela 4.

Porównanie poprawności miary $m^{(1)}$ opisanej wzorem (2) oraz miary Hellwiga $m^{(2)}$ (wzór (3))

M	Miara syntet.	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
M1	$m^{(1)}$,9998	,9998	,9998	,9999	,9999	,9999	,9998	,9999	,9999	,9999
	$m^{(2)}$,9976	,9981	,9978	,9985	,9986	,9987	,9984	,9986	,9986	,9984
M2	$m^{(1)}$,9997	,9997	,9997	,9998	,9996	,9997	,9996	,9997	,9998	,9998
	$m^{(2)}$,9967	,9970	,9966	,9973	,9967	,9967	,9967	,9974	,9973	,9971
M3	$m^{(1)}$,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998
	$m^{(2)}$,9972	,9976	,9973	,9981	,9979	,9980	,9977	,9981	,9981	,9979
M4	$m^{(1)}$	0,669	0,663	0,706	0,701	0,659	0,671	0,694	0,688	0,673	0,608
	$m^{(2)}$	0,682	0,676	0,715	0,709	0,666	0,686	0,703	0,700	0,683	0,615
M5	$m^{(1)}$	0,889	0,889	1,000	1,000	0,778	1,000	0,778	0,778	0,778	0,778
	$m^{(2)}$	0,889	0,889	1,111	0,889	0,889	0,889	1,111	1,000	0,778	0,667
M6	$m^{(1)}$	0,696	0,670	0,728	0,690	0,629	0,678	0,686	0,658	0,665	0,633
	$m^{(2)}$	0,728	0,700	0,765	0,753	0,676	0,727	0,723	0,690	0,697	0,650
M7	$m^{(1)}$	1,000	1,000	0,889	0,778	0,778	0,889	0,889	0,889	0,889	0,778
	$m^{(2)}$	1,000	1,000	1,111	1,000	0,889	1,000	1,111	1,000	0,889	0,778
M8	$m^{(1)}$	0,547	0,543	0,563	0,527	0,505	0,522	0,529	0,519	0,526	0,516
	$m^{(2)}$	0,557	0,549	0,580	0,565	0,530	0,553	0,544	0,538	0,538	0,514
M9	$m^{(1)}$	-0,03	-0,03	-0,03	-0,03	-0,03	-0,03	-0,03	-0,03	-0,03	-0,04
	$m^{(2)}$	-0,50	-0,50	-0,50	-0,50	-0,50	-0,50	-0,50	-0,50	-0,50	-0,50
M10	$m^{(1)}$	1,168	1,026	1,291	1,063	0,869	1,217	0,938	0,844	0,971	1,319
	$m^{(2)}$	0,880	0,777	0,874	0,930	0,816	0,839	0,871	0,829	0,793	1,021

Tabela 4. (cd.)

<i>M</i>	Miara syntet.	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
<i>M11</i>	$m^{(1)}$	1,103	1,110	1,122	1,083	1,053	1,061	1,099	1,088	1,065	1,028
	$m^{(2)}$	1,039	1,022	1,063	1,049	1,013	1,026	1,081	1,063	1,045	0,998
<i>M12</i>	$m^{(1)}$	1,440	1,441	1,475	1,459	1,439	1,447	1,478	1,465	1,446	1,401
	$m^{(2)}$	1,365	1,356	1,393	1,380	1,365	1,385	1,429	1,408	1,402	1,346

Źródło: opracowanie własne na podstawie danych zawartych w tabeli 1 oraz tabeli 2.

Analiza uzyskanych rezultatów nasuwa spostrzeżenie, że w każdym analizowanym aspekcie jedna z metod wykazywała wyraźną przewagę. Miara Hellwiga miała lepszą zgodność odwzorowania (mierniki *M1–M3*), przy czym było to widoczne dopiero na trzecim miejscu po przecinku. Przeważała również pod względem zmienności i koncentracji (*M9, M10*) oraz odległości taksonomicznej zmiennej syntetycznej od zmiennych pierwotnych (*M11, M12*). Natomiast proponowana w artykule zmienna syntetyczna była silniej skorelowana ze zmiennymi diagnostycznymi zarówno liniowo (*M4, M5*) jak i rangowo (*M6, M7*).

6. PODSUMOWANIE

Zaproponowana w artykule konstrukcja miary syntetycznej, mimo że w zasadzie przebiega w przestrzeni jednowymiarowej, wydaje się być użytecznym narzędziem wielowymiarowej analizy porównawczej. Posiada ona własność postulowaną dla tego typu miar: jej wartości rosną wraz ze wzrostem poziomu rozwoju obiektu. Ponadto miara jest niemianowana, ograniczona oraz określona również w przypadku, gdy cecha przyjmuje wartość 0. Główną jej zaletą jest porównywalność wartości w latach, przez co możliwa staje się analiza dynamiki zjawiska złożonego.

W opisanej metodzie, aby syntetycznie ocenić sytuację obiektu, analizuje się jego odległości od wzorca (antywzorca) względem sumy odległości od wzorca (antywzorca) wszystkich obiektów. Przy czym dla nominant antywzorzec określono dwojako w zależności od sytuacji danego obiektu. Inaczej niż dla większości konstrukcji tego typu, na wartość miernika dla danego obiektu silny wpływ ma sytuacja pozostałych obiektów, tzn. miara charakteryzuje poziom rozwoju obiektu na tle wszystkich pozostałych.

W badaniu empirycznym proponowana miara dała wyniki podobne – ale nie tożsame – z wynikami uzyskanymi z wykorzystaniem bardzo popularnej miary Hellwiga. Porównanie obu mierników pod względem poprawności nie ukazało wyraźnej przewagi żadnej z metod. Miara Hellwiga lepiej wypadła pod względem odległości miary syntetycznej od zmiennych pierwotnych oraz zmienności i koncentracji miernika, natomiast proponowana konstrukcja przeważała pod względem korelacji miary syntetycznej ze zmiennymi diagnostycznymi.

Prosta konstrukcja proponowanej miary pozwala na jej różnorodne modyfikacje, które będą przedmiotem dalszych badań. Na przykład wprowadzając inny rodzaj odległości można zwiększyć zmienność miary. Ciekawe warianty można również otrzymać wykorzystując koncepcje opisane w pracach: Jajuga (1986), Walesiak (2011), Pietrzak (2014).

LITERATURA

- Bartosiewicz S., (1976), Propozycja metody tworzenia zmiennych syntetycznych, *Prace Naukowe AE we Wrocławiu*, 84, 5–7.
- Cieślak M., (1974), Taksonomiczna procedura programowania rozwoju gospodarczego i określenia zapotrzebowania na kadry kwalifikowane, *Przegląd Statystyczny*, 21 (1), 29–39.
- Dittmann P., Pisz Z., (1975), Metoda dynamicznego badania zróżnicowania przestrzennego zjawisk społeczno-ekonomicznych, *Wiadomości Statystyczne*, 11.
- Grabiński T., (1984), *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Wydawnictwo AE, Kraków.
- Grabiński T., Wydymus S., Zeliś A., (1989), *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWE, Warszawa.
- Hellwig Z., (1968), Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę kwalifikowanych kadr, *Przegląd Statystyczny*, 15 (4), 307–327.
- Jajuga K., (1986), *Metoda tworzenia zmiennych syntetycznych na podstawie klasyfikacji rozmytej*, Prace Naukowe AE, 360, Wrocław.
- Lira J., Wagner W., Wysocki F., (2002), Mediana w zagadnieniach porządkowania obiektów wielocechowych, w: Paradysz J., (red.), *Statystyka regionalna w służbie samorządu terytorialnego i biznesu*, Akademia Ekonomiczna w Poznaniu, 87–99.
- Młodak A., (2006), *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa.
- Młodak A., (2009), Hierarchiczność a porządkowanie obiektów w statystyce regionalnej, *Wiadomości Statystyczne*, 4, 11–25.
- Nowak E., (1977), Syntetyczne mierniki plonów w niektórych krajach europejskich, *Wiadomości Statystyczne*, 10, 19–22.
- Nowak E., (1982), Badanie zgodności metod konstruowania taksonomicznych mierników rozwoju, *Przegląd Statystyczny*, 29 (3/4), 455–463.
- Panek T., (1984), Taksonomiczna modyfikacja dystansowej metody pomiaru poziomu życia ludności, *Prace Naukowe AE we Wrocławiu*, 262.
- Pietrzak M. B., (2014), Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych, *Przegląd Statystyczny*, 59 (2), 181–201.
- Pluta W., (1976), Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego, *Przegląd Statystyczny*, 23 (4), 511–518.
- Strahl D., (1978), Propozycja konstrukcji miary syntetycznej, *Przegląd Statystyczny*, 25 (2), 205–215.
- Strahl D., (2006), Strukturalna miara rozwoju obiektów hierarchicznych, *Prace Naukowe AE we Wrocławiu, Ekonometria*, 16, 11–20.
- Strahl D., Markowska M., (2008), Klasyfikacja dynamiczno-przestrzenna obiektów hierarchicznych z wykorzystaniem statystyk porządkowych, w: Dziechciarz J., (red.), *Zastosowania metod ilościowych*, Prace Naukowe AE we Wrocławiu, Ekonometria, 20, Wrocław, 16–27.
- Walesiak M., (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*, Wydawnictwo UE, Wrocław.

Walesiak M., (2014), Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej, *Przegląd Statystyczny*, 61 (4), 363–372.

Wydymus S., (2013), Rozwój gospodarczy a poziom wynagrodzeń w krajach Unii Europejskiej – analiza taksonomiczna, *Finanse, Rynki Finansowe, Ubezpieczenia*, 57, 631–645.

DODATEK

Dowody własności:

1. Oczywiste.
2. Oczywiste.
3. Bez straty ogólności założymy, że w zbiorze zmiennych diagnostycznych występuje tylko jedna nominanta, nadajmy jej numer 1. Założymy, że jej wartość nominalna dzieli uporządkowane niemalejąco wartości dla obiektów w proporcji $n_1:n-n_1$. Odpowiednie wartości antyzwzorcowe oznaczmy x_1^- oraz x_1^+ . Wówczas:

$$\begin{aligned} \sum_{i=1}^n m_i^- &= \sum_{i=1}^n \left(\frac{1}{r} \sum_{p=1}^r m_{ip}^- \right) = \frac{1}{r} \sum_{p=1}^r \sum_{i=1}^n m_{ip}^- = \frac{1}{r} \left(\sum_{i=1}^n m_{i1}^- + \sum_{p=2}^r \sum_{i=1}^n m_{ip}^- \right) = \\ &= \frac{1}{r} \left(\sum_{i=1}^{n_1} m_{i1}^- + \sum_{i=n_1+1}^n m_{i1}^- + \sum_{p=2}^r \sum_{i=1}^n m_{ip}^- \right) = \\ &= \frac{1}{r} \left(\sum_{i=1}^{n_1} \frac{n_1}{n} \cdot \frac{|x_{i1}^- - x_1^-|}{\sum_{j=1}^{n_1} |x_{ip}^- - x_p^-|} + \sum_{i=n_1+1}^n \frac{n-n_1}{n} \cdot \frac{|x_{i1}^- - x_1^+|}{\sum_{j=n_1+1}^n |x_{ip}^- - x_p^-|} + \sum_{p=2}^r \sum_{i=1}^n \frac{|x_{ip}^- - x_p^-|}{\sum_{i=1}^n |x_{ip}^- - x_p^-|} \right) = \\ &= \frac{1}{r} \left(\frac{n_1}{n} + \frac{n-n_1}{n} + r - 1 \right) = 1. \end{aligned}$$

4. Podobnie jak 3.
5. Oczywiste na podstawie 3.
6. Oczywiste na podstawie 4.
7. $\sum_{i=1}^n m_i = \sum_{i=1}^n \left(\frac{1+m_i^-+m_i^+}{2} \right) = \frac{n}{2} + \sum_{i=1}^n m_i^- + \sum_{i=1}^n m_i^+ = \frac{n}{2} - 1 + 1 = \frac{n}{2}$.

Przypadki szczególne:

1. $m_i^+ = 0 \Leftrightarrow \frac{1}{r} \sum_{p=1}^r m_{ip}^+ = 0 \Leftrightarrow \bigwedge_p m_{ip}^+ = 0 \Leftrightarrow \bigwedge_p |x_{ip} - x_p^+| = 0 \Leftrightarrow \bigwedge_p x_{ip} = x_p^+$.
2. Analogicznie jak 1.
3. $m_i^+ = 0 \wedge m_i^- = 0 \Leftrightarrow \bigwedge_p (|x_{ip} - x_p^+| = 0 \wedge |x_{ip} - x_p^-| = 0) \Leftrightarrow \bigwedge_p x_p^+ = x_p^-$
co jest sprzeczne z założeniem, że zmienne spełniają kryterium zmiennościowe.
4. $m_i^+ = -1 \Leftrightarrow \frac{1}{r} \sum_{p=1}^r m_{ip}^+ = 1 \Leftrightarrow \bigwedge_p m_{ip}^+ = 1 \Leftrightarrow \bigwedge_p \frac{|x_{ip} - x_p^+|}{\sum_{j=1}^n |x_{jp} - x_p^+|} =$
 $= 1 \Leftrightarrow \bigwedge_p |x_{ip} - x_p^+| = \sum_{j=1}^n |x_{jp} - x_p^+| \Leftrightarrow \bigwedge_p \sum_{j \neq i} |x_{jp} - x_p^+| =$
 $= 0 \Leftrightarrow \bigwedge_p \bigwedge_{j \neq i} x_{jp} = x_p^+.$
5. Podobnie jak 4.

PROPOZYCJA MIARY SYNTETYCZNEJ

Streszczenie

Celem artykułu było zaproponowanie nowej metody konstrukcji zmiennej syntetycznej. Aby ocenić poziom rozwoju obiektu, analizowano jego odległość od wzorca (antywzorca) względem sumy odległości od wzorca (antywzorca) wszystkich obiektów. W ten sposób scharakteryzowano region na tle wszystkich pozostałych. Opisana miara jest ograniczona. Zmienne diagnostyczne nie wymagają normalizacji ani zamiany na stymulanty. Podstawową zaletą sugerowanego podejścia jest porównywalność analiz prowadzonych dla różnych lat. Proponowaną metodę zilustrowano przykładem empirycznym. Wartości miar porównano z wynikami otrzymanymi na podstawie metody wzorcowej Hellwiga. Były podobne, ale nie tożsame. W kolejnym kroku badania porównano obie metody pod względem poprawności. Żadna z nich nie wykazała znaczącej przewagi.

Słowa kluczowe: zmienna syntetyczna, miernik rozwoju

PROPOSITION OF SYNTHETIC MEASURE

Abstract

The aim of the study was to propose a new method for construction of synthetic variable. We analyzed distance between object and pattern (antipattern) in relation to sum of all distances between the pattern (antipattern) and other objects. This allowed us to characterize position of a region among others. The described measure is limited. Diagnostic variables do not require normalization or conversion to stimulants. Its main advantage is comparability of analyzes carried out for various years. The proposed method was illustrated by an empirical example. The values of measurement were compared with those obtained by Hellwig's method. They were similar but not identical. Then two methods were compared in terms of correctness. None of them had a significant advantage.

Keywords: synthetic variable, measure of development