Karol Kukuła
Zakład Statystyki Matematycznej AR w Krakowie

Analiza własności metody unitaryzacji zerowanej

Wstęp

Istnieje poważna liczba zjawisk, które można określić mianem złożone. Wszelkie porównania przestrzenne, a niekiedy również czasowe w zakresie zjawisk złożonych, stwarzają konieczność sporządzania ocen pewnych obiektów, a w dalszej kolejności budowy ich rankingów. Zjawiska złożone z natury rzeczy są charakteryzowane wieloma cechami, które mają różne miana i wykazują różne rzędy wielkości. Aby ocena zjawiska na podstawie cech go opisujących była możliwa, należy w sposób w miarę prosty dokonać przekształcenia ich wartości oryginalnych. Przekształcone zmienne są pozbawione mian i przybierają wartości zbliżonego rzędu wielkości. Takie sposoby transformacji wartości oryginalnych cech diagnostycznych nazywane są metodami normowania. Unormowane wartości zmiennych diagnostycznych mogą być poddane procesowi agregacji, co prowadzi do uzyskania konkretnych ocen zmiennej agregatowej. Wartości zmiennej agregatowej stanowią oceny obiektów ze względu na stan rozwojowy badanego zjawiska złożonego. Dysponowanie ocenami poszczególnych obiektów pozwala zbudować ich ranking, tj. układ, w którym obiekty są uporządkowane w kolejności od najlepszego do najgorszego ze względu na wartość zmiennej agregatowej, zwanej również zmienną syntetyczną.

W długim łańcuchu czynności prowadzących do pozyskania wielokryterialnych ocen obiektów oraz budowy ich rankingu szczególną rolę do odegrania ma proces normowania cech diagnostycznych za pomocą określonych metod. Istnieje wiele metod normowania zmiennych diagnostycznych. Metody te dają częstokroć dość zróżnicowane wyniki unormowań, nawet gdy dotyczą tych samych zmiennych. Zatem należy przyjąć, iż wybór metody ma istotny wpływ na oceny obiektów.

Wyniki badań symulacyjnych dotyczących wyników normowania przy zastosowaniu dziesięciu formuł (zob. [5], ss. 111–151) upoważniają do zwrócenia szczególnej uwagi na metodę unitaryzacji zerowanej, w skrócie MUZ.

Kwintesencją badań zjawisk złożonych jest ich ujęcie porównawcze, co oznacza, iż poziom zjawiska rozpatruje się w obiektach. Obiekty w początkowej fazie charakteryzowane są za pomocą wielu cech. Przyjmijmy, że O oznaczać będzie zbiór obiektów:

$$O = \{ O_1, O_2, \ldots, O_r \}, \tag{1}$$

gdzie r jest liczbą badanych obiektów. Każdy z obiektów jest opisany przez zbiór zmiennych diagnostycznych X.

W podejściu badawczym określonym przymiotnikiem „statyczne" rozpatrywane zjawisko analizowane jest w jednym, wybranym okresie. Zwykle jest to ostatni z okresów, gwarantujący pozyskanie odpowiednich informacji statystycznych. W podejściu tym zrezygnowano z oznaczania zmiennych indeksem t. Niezbędne w badaniach wielowymiarowych dane tworzą macierz o postaci:

$$\mathbf{X} = [x_{ij}] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1s} \\ x_{21} & x_{22} & \cdots & x_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ x_{r1} & x_{r2} & \cdots & x_{rs} \end{bmatrix}, \quad \begin{pmatrix} i = 1, \ldots, r \\ j = 1, \ldots, s \end{pmatrix}, \tag{2}$$

gdzie x_{ij} oznacza realizację zmiennej X_j w obiekcie O_i. Zatem i-ty obiekt charakteryzuje następujący wektor zmiennych:

$$\mathbf{X}_i = \begin{bmatrix} x_{i1} & x_{i2} & \ldots & x_{is} \end{bmatrix}, \qquad (i = 1, \ldots, r). \tag{3}$$

Wektor $\mathbf{X}_i$ jest s-wymiarową obserwacją charakteryzującą obiekt O_i. Łatwo zauważyć, iż każdemu obiektowi odpowiada punkt w przestrzeni s-wymiarowej.

Podział metod normujących

Rozpatrzenie zalet, jak również ewentualnych wad metody unitaryzacji zerowanej (MUZ) rodzi konieczność jej prezentacji na tle innych metod normujących cechy diagnostyczne w możliwie szerokim kontekście. Dla realizacji tego celu niezbędne wydają się omówienie i wszechstronna analiza własności procedur normowania najczęściej stosowanych i mających już stałe miejsce w literaturze przedmiotu. Ze względu na znaczną ich liczbę, prezentację procedur normowania ograniczono do wybranych metod. Przedstawiono i omówiono te, które znajdują akceptację wśród badaczy wykorzystujących aparat wielowymiarowej analizy porównawczej w pracach empirycznych. Z drugiej zaś strony przy wyborze uwzględniono tradycyjny już podział procedur normowania (zob. T. Borys – [2] lub T. Grabiński – [4], s. 33–34), który obejmuje:

- metody standaryzacji,
- metody unitaryzacji,
- przekształcenia ilorazowe,
- metody rangowe.

Normowanie jest działaniem mającym na celu przysposobić zmienne diagnostyczne do roli kryteriów cząstkowych w procesie oceny zjawiska złożonego. Dodajmy, iż zwykle cechy diagnostyczne wyrażone są w różnych jednostkach miary oraz odpowiadają im zróżnicowane zakresy liczbowe. Uwzględniając potrzebę pozbycia się mian oraz ujednolicenia zakresów liczbowych zmiennych diagnostycznych, metody normujące służą transformacji bezwzględnych wartości na wartości względne. W tym sensie każda z metod normowania (oprócz metody rangowej) jest pewnym przekształceniem ilorazowym, które w końcowym wyniku daje zmienną diagnostyczną transformowaną. Zmienna ta jest pozbawiona miana i ujednolicona co do zakresu wartości, jakie może przyjmować. Powstaje zatem pewna wątpliwość, czy słuszne jest określenie „metody oparte na przekształceniu ilorazowym”, do których zalicza się m.in. metodę zaproponowaną przez D. Strahl [7], E. Nowaka [6], S. Bartosiewicz [1] oraz M. Cieślak [3], skoro wszystkie wymienione metody (z metodami standaryzacyjnymi i unitaryzacyjnymi włącznie) są skonstruowane w formie ilorazowej. Oznacza to dzielenie oryginalnej wartości cechy bądź różnicy między tą wartością a określonym parametrem (średnią, minimum wszystkich wartości itp.) przez odpowiednią wartość stałą wyrażoną tą samą jednostką co zmienna oryginalna. Dlatego też, biorąc pod uwagę zaistniały już podział metod normujących (zob. [2] lub [4]), proponuję przyjąć następującą ich specyfikację:

A. Metody oparte na formule przekształcenia ilorazowego.
B. Metody rangowe.

Metody z grupy A oparte na formule przekształcenia ilorazowego przyjmują różne punkty odniesienia, które można określić jako:

1) miary zróżnicowania cech, takie jak:
   a) odchylenie standardowe zmiennej (ten punkt odniesienia wykorzystują metody standaryzacyjne),

b) rozstęp zmiennej (ten punkt odniesienia wykorzystują metody unitaryzacyjne);

2) inne parametry stałe cechy, takie jak:
   a) średnia arytmetyczna zmiennej,
   b) maksymalna wartość zmiennej,
   c) minimalna wartość zmiennej,
   d) długość wektora realizacji zmiennej,
   e) suma realizacji zmiennej.

Przy omawianiu metod normowania cech diagnostycznych skoncentrowano uwagę na metodach z grupy A. Metody rangowe zastosowane do zmiennych mierzonych na skalach ilorazowej i przedziałowej wprawdzie są możliwe do zastosowania, co tłumaczy łatwość przejścia z wyższych skal do niższych, lecz takie postępowanie należy zawsze łączyć z poważną stratą informacji. A oto wybrane formuły normujące z grupy metod A:

(I) $$z_{ij} = \frac{x_{ij} - \overline{X}_j}{S(X_j)}, \qquad S(X_j) > 0,$$

(II) $$z_{ij} = \frac{x_{ij}}{S(X_j)}, \qquad S(X_j) > 0,$$

(III) $$z_{ij} = \frac{x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad \max_i x_{ij} > \min_i x_{ij},$$

(IV) $$z_{ij} = \frac{x_{ij} - \overline{X}_j}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad \max_i x_{ij} > \min_i x_{ij},$$

(V) $$z_{ij} = \frac{x_{ij} - \min\limits_i x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad \max_i x_{ij} > \min_i x_{ij},$$

(VI) $$z_{ij} = \frac{x_{ij}}{\max\limits_i x_{ij}}, \qquad \max_i x_{ij} \neq 0,$$

(VII) $$z_{ij} = \frac{x_{ij}}{\min\limits_i x_{ij}}, \qquad \min\limits_i x_{ij} \neq 0,$$

(VIII) $$z_{ij} = \frac{x_{ij}}{\overline{X}_j}, \qquad \overline{X}_j \neq 0,$$

(IX) $$z_{ij} = \frac{x_{ij}}{\sum\limits_{i=1}^{r} x_{ij}}, \qquad \sum\limits_{i=1}^{r} x_{ij} \neq 0,$$

(X) $$z_{ij} = \frac{x_{ij}}{\left[\sum\limits_{i=1}^{r} x_{ij}^2\right]^{0,5}}, \qquad \sum\limits_{i=1}^{r} x_{ij}^2 > 0.$$

Szczególne miejsce poświęcono przypadkowi normowania zmiennych diagnostycznych przyjmujących wartości ze zbioru liczb rzeczywistych R (chodzi o takie zmienne, które mogą być zarówno liczbami ujemnymi, jak i dodatnimi, a także mogą przyjmować wartość zero).

Formuły wartościujące w metodzie unitaryzacji zerowanej

Celem operacji normowania jest pozbawienie zmiennych mian oraz ujednolicenie ich przedziałów zmienności. Zmienne unormowane będziemy określać symbolem Z. Zbiór cech diagnostycznych podzielono na trzy podzbiory:

- S – stymulant,
- D – destymulant,
- N – nominant.

W MUZ najczęściej stosowane są następujące formuły normalizacyjne:

$$z_{ij} = \frac{x_{ij} - \min\limits_i x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, \qquad X_j \in S, \tag{4}$$

$$z_{ij} = \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}, \qquad X_j \in D, \tag{5}$$

$$z_{ij} = \begin{cases} \dfrac{x_{ij} - \min_i x_{ij}}{c_{0j} - \min_i x_{ij}}, & \text{gdy } x_{ij} < c_{0j} \\ 1, & \text{gdy } x_{ij} = c_{0j} \\ \dfrac{x_{ij} - \max_i x_{ij}}{c_{0j} - \max_i x_{ij}}, & \text{gdy } x_{ij} > c_{0j} \end{cases}, \qquad X_j \in N, \tag{6}$$

przy czym c_{0j} – nominalna wartość j-tej zmiennej,

oraz

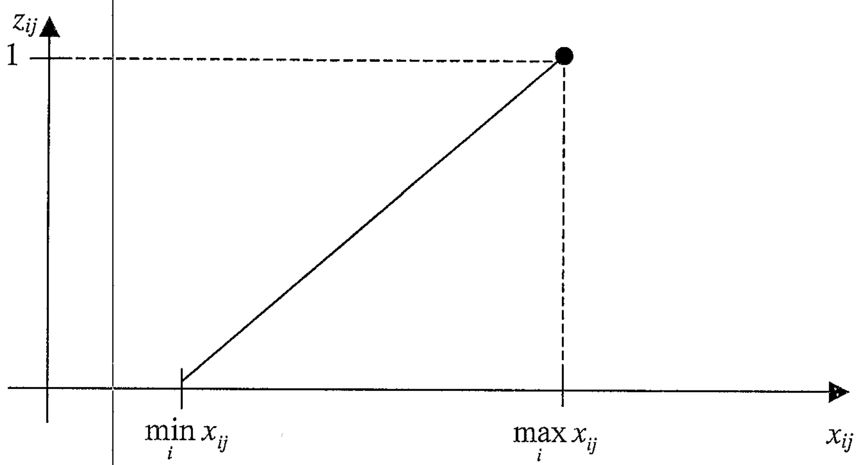
$$z_{ij} = \begin{cases} \dfrac{x_{ij} - \min_i x_{ij}}{c_{1j} - \min_i x_{ij}}, & \text{gdy } x_{ij} < c_{1j} \\ 1, & \text{gdy } c_{1j} \le x_{ij} \le c_{0j}, \\ \dfrac{x_{ij} - \max_i x_{ij}}{c_{2j} - \max_i x_{ij}}, & \text{gdy } x_{ij} > c_{2j} \end{cases} \qquad X_j \in N. \tag{7}$$

Formuła (7) jest przewidziana dla normowania nominant z określonym przedziałem wartości nominalnych $[c_{1j}, c_{2j}]$.

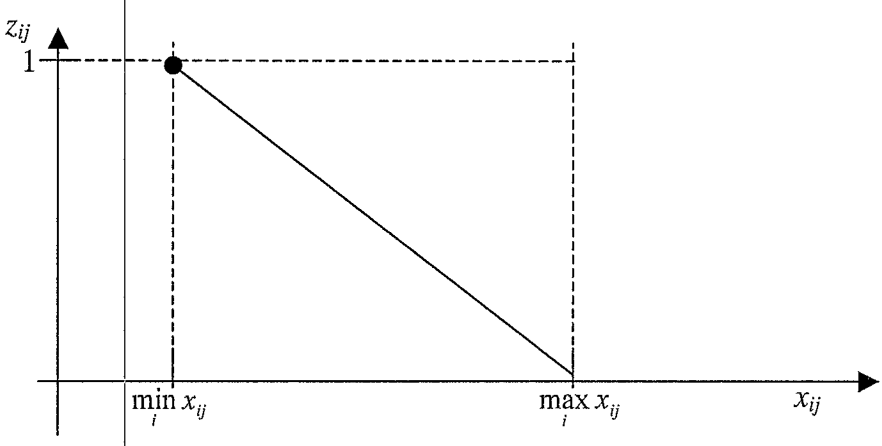
W każdej z przedstawionych formuł otrzymujemy:

$$z_{ij} \in [0\ ,\ 1] \tag{8}$$

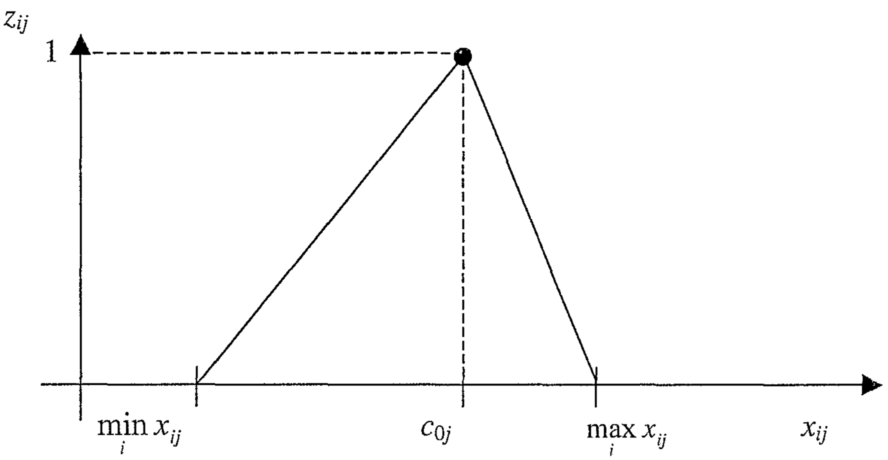
Działanie MUZ w swej klasycznej wersji opisane formułami (4)–(7) ilustrują rysunki 1–4.



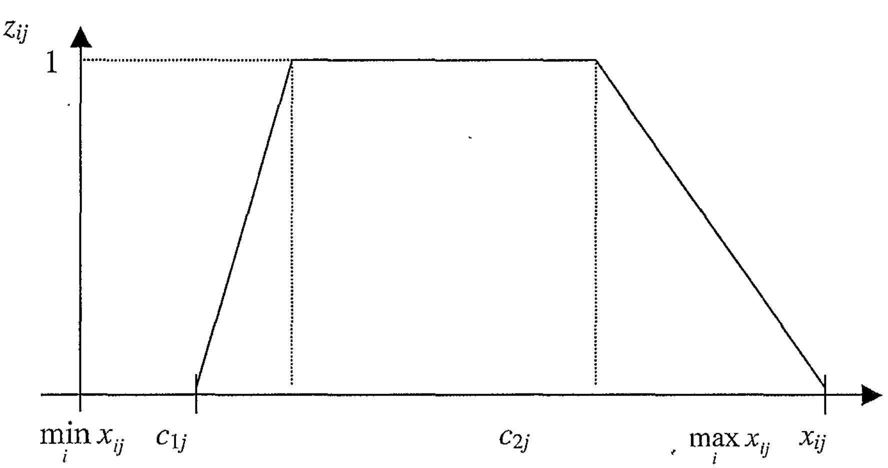
Rysunek 1
($X_j \in S$)



Rysunek 2
($X_j \in D$)



Rysunek 3
$(X_j \in N)$



Rysunek 4
$(X_j \in N)$

Postulaty stawiane metodom normującym cechy diagnostyczne

Normowanie cech diagnostycznych nie jest czynnością li tylko formalną, ma ono bowiem do spełnienia pewne cele natury praktycznej. Główne cele normowania według Borysa [2] można ująć następująco:
1) wprowadzenie addytywności w zbiorach wartości cech o różnych mianach;
2) określenie na zbiorze wartości cechy funkcji preferencyjnej; w zależności od rodzaju cechy funkcja preferencyjna jest transformacją stymulacyjną, destymulacyjną lub stymulacyjno-destymulacyjną (przypadek nominant).

Normowanie cech ma za zadanie umożliwienie realizacji szeroko zakrojonych badań porównawczych obiektów ze względu na poziom wielu zmiennych (cech) przyjętych jako kryteria oceny rozpatrywanego zjawiska złożonego. Aby zadanie to mogło być należycie wypełnione, nie należy traktować obojętnie własności charakteryzujących poszczególne formuły normalizacyjne. Jest rzeczą niezmiernie ważną, aby stosowane przez praktyków procedury normujące spełniały określone wymogi. Postulaty te można sformułować w kilku punktach:
1) pozbawienie mian (jednostek), w których są wyrażone cechy diagnostyczne;
2) sprowadzenie rzędu wielkości zmiennych diagnostycznych do stanu porównywalności, co oznacza wyrównanie zakresów zmienności cech, a w konsekwencji możliwość ich dodawania;
3) równość długości przedziałów zmienności wartości wszystkich cech unormowanych (stałość rozstępu z_{ij}) oraz równość dolnej i górnej granicy ich przedziału zmienności, w szczególności chodzi o przedział [0, 1];
4) możliwość normowania cech diagnostycznych przyjmujących wartości zarówno dodatnie, jak i ujemne lub tylko ujemne;
5) możliwość normowania cech przyjmujących wartość zero;
6) nieujemność wartości cech unormowanych;
7) istnienie prostych formuł – w ramach danej procedury normalizacyjnej – ujednolicających charakter zmiennych.

Warto podkreślić, że prawie wszystkie formuły normalizacyjne spełniają dwa pierwsze postulaty. Pozostałe postulaty uwzględniają tylko niektóre z formuł, i to nie zawsze wszystkie naraz. Można pokusić się o stwierdzenie, że metoda normalizacyjna, która spełnia wymienione postulaty, gwarantuje uniwersalne unormowanie wszystkich cech niezależnie od ich charakteru, rzędu wielkości czy też znaku.

Założono, iż w analizie porównawczej własności formuł normalizacyjnych brane są pod uwagę stymulanty jako zmienne najczęściej występujące w badaniach empirycznych. Ponadto, formuły dla destymulant i nominant są z reguły pokrewne konstrukcyjnie w stosunku do wzorów dla stymulant i mają te same bądź zbliżone do nich własności. W związku z tym wnioski płynące z analizy formuł przeznaczonych dla stymulant można uogólnić na wszystkie formuły.

Własności MUZ na tle pozostałych metod normowania

W tym podrozdziale starano się sprecyzować, a następnie poddać analizie własności dziesięciu wybranych formuł normujących cechy diagnostyczne (I–X) głównie pod kątem sformułowanych postulatów. Aby doprowadzić do zbiorczego ich porównania, zbudowano tabelę 1, w której w sposób dychotomiczny określono zgodność formuł normowania w stosunku do postulatów (1)–(7) oraz posiadania lub nieposiadania stałych wartości parametrów charakteryzujących zmienne unormowane. Z analizy zawartości tej tabeli wynika, iż żadna z formuł przekształceniowych nie ma samych plusów. Najwięcej jednak pozytywów wiąże się z formułą (V), właściwą metodzie unitaryzacji zerowanej. W dalszej kolejności plasują się formuły (I) i (IX). Gradacji tej nie należy kojarzyć jako jednoznacznego wskazania metod najlepszych, o wyborze bowiem metody w konkretnych zastosowaniach mogą decydować dodatkowe kryteria (tu nie uwzględnione) bądź też spełnienie lub niespełnienie jednego tylko z wymienionych postulatów.

Zauważmy na podstawie wyników zawartych w tabeli 1, że najtrudniejszym do spełnienia wymogiem w kontekście rozważanych formuł okazał się postulat (3). Postulat ten jest uszczegółowionym rozwinięciem postulatu (2), który tu traktujemy jako spełniony przez wszystkie formuły, bardziej w sensie intencjonalnym. Napotykamy dość wyraźne zróżnicowania wyników zarówno pod względem wartości rozstępu, jak i usytuowania dolnych i górnych granic przedziałów zmienności zmiennych unormowanych poszczególnymi metodami. Z wszystkich formuł branych pod uwagę tylko MUZ – formuła (V) – daje unormowania w stałym przedziale [0, 1]. Względnie ustabilizowane – wg przyjętych kryteriów – wyniki normowania uzyskano przy transformacjach opisanych wzorami (I), (IV), (IX) i (X).

Istnieje wiele trudności ze wskazaniem najlepszej metody odpowiadającej celom i zakresowi konkretnej analizy porównawczej. Na podstawie spostrzeżeń dokonanych w kontekście omawianych formuł transformacyjnych podejmujemy próbę sformułowania kilku wskazówek o charakterze aplikacyjnym.

Tabela 1

Zestawienie porównawcze formuł normujących cechy diagnostyczne

Formuła normująca		Postulaty stawiane formułom normującym						Stałość parametrów zmiennych unormowanych	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	$\overline{Z}_j$	$S^2(Z_j)$
(I)	+	+	–	+	+	–	–	+	+
(II)	+	+	–	+	+	–	–	–	+
(III)	+	+	–	+	+	–	+	–	–
(IV)	+	+	–	+	+	–	–	+	–
(V)	+	+	+	+	+	+	+	–	–
(VI)	+	+	–	–	–	+	+	–	–
(VII)	+	+	–	–	–	+	+	–	–
(VIII)	+	+	–	–	+	+	–	+	–
(IX)	+	+	–	–	+	+	+	+	–
(X)	+	+	–	–	+	+	+	–	–

Legenda:

(+) należy kojarzyć ze spełnieniem postulatu bądź ze stałością parametru charakteryzującego zmienną unormowaną,

(–) należy kojarzyć z niespełnieniem postulatu bądź z niestałością parametru charakteryzującego zmienną unormowaną.

Źródło: Opracowanie własne.

Duży wpływ na rezultaty porządkowania liniowego obiektów a tym samym na budowę ich rankingu ma wybór odpowiedniej formuły normującej. Przy czym w tym przypadku zaleca się wybór tych formuł, które dają stabilne bądź prawie stabilne przedziały zmienności zmiennych unormowanych, ze szczególnym wskazaniem MUZ.

Nieco innego wyboru metod można dokonać mając na celu modelowanie zjawisk złożonych. Tu często wykorzystuje się stałość parametrów charakteryzujących zmienne unormowane (por. E. Nowak [6], s. 80–81). W takich przypadkach można zalecać formuły (I), (VIII) lub (IX).

W przypadkach, w których sumowanie poszczególnych wartości realizacji zmiennych po obiektach przejawia sens, można zastosować formułę (IX). W wyniku tego zabiegu otrzymujemy struktury przestrzenne, które zawierają dodatkową informację o badanym zjawisku. Tego typu normowania można z powodzeniem stosować w przestrzennych analizach produkcji przemysłowej, rolniczej itp.

Wreszcie kilka uwag o normowaniu zmiennych przyjmujących wartości ujemne. Należy zauważyć, iż stosunkowo rzadko napotykamy konieczność transformowania zmiennych przyjmujących wartości zarówno dodatnie, jak i ujemne. Niemniej istnieje kilka przypadków, w których konieczność ta wystąpi. Są to zwykle badania porównawcze dotyczące kondycji finansowej firm, banków i innych instytucji. W badaniach tych nie sposób pominąć kategorii określonej mianem wynik finansowy, który może przybierać wartości zarówno dodatnie, jak i zero czy też wartości ujemne. Musimy zatem dobrać taką metodę normowania, która transformuje zmienne diagnostyczne o wszystkich możliwych wartościach ($X_j \in R$). W tej sytuacji mogą być wykorzystane formuły (I), (IV) i (V).

Literatura

1. BARTOSIEWICZ S., Propozycja metody tworzenia zmiennych syntetycznych, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1976, nr 84.
2. BORYS T., Metody normowania cech w statystycznych badaniach porównawczych, *Przegląd Statystyczny* 1978, z. 2.
3. CIEŚLAK M., *Taksonomiczna procedura programowania rozwoju gospodarczego i określania zapotrzebowania na kadry kwalifikowane*, PWN, Warszawa 1976.
4. GRABIŃSKI T., Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk gospodarczych, *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, seria specjalna: *Monografie*, nr 61, Kraków 1984.
5. KUKUŁA K., *Metoda unitaryzacji zerowanej*, Wydawnictwo Naukowe PWN Warszawa 2000.
6. NOWAK E., Propozycja prostej metody konstruowania miernika rozwoju i jego wykorzystania do badań regresyjnych, *Przegląd Statystyczny* 1979, z. 1–2.
7. STRAHL D., Propozycja konstrukcji miary syntetycznej, *Przegląd Statystyczny* 1978, z. 2.

Properties of Zero Unitarization Method

Abstract

The paper presents properties of zero unitarization method in relation with some other normalisation methods.

Formulas for normalising stimulants, destimulants and nominants are shown and compared. Special attention is given to the problem of normalisation of nominants.

In the concluding part the method discussed in the paper is proven to be universal and easy in many applications.