PRZEGLĄD STATYSTYCZNY
R. LXI – ZESZYT 4 – 2014

MAREK WALESIAK

PRZEGLĄD FORMUŁ NORMALIZACJI WARTOŚCI ZMIENNYCH ORAZ ICH WŁASNOŚCI W STATYSTYCZNEJ ANALIZIE WIELOWYMIAROWEJ

1. WSTĘP

Punktem wyjścia zastosowania metod statystycznej analizie wielowymiarowej jest macierz danych $[x_{ij}]$, w której dowolny element x_{ij} ($i = 1,...,n$; $j = 1,...,m$) oznacza obserwację j-tej zmiennej dla i-tego obiektu. Normalizację przeprowadza się, gdy zmienne opisujące obiekty badania mierzone są na skali przedziałowej lub ilorazowej[1]. W odniesieniu do słabych skal pomiaru (nominalna, porządkowa) nie zachodzi potrzeba normalizacji, na ich wartościach bowiem nie wyznacza się ani relacji równości różnic i przedziałów, ani stosunków.

Celem normalizacji wartości zmiennych jest doprowadzenie zmiennych do porównywalności. Uzyskuje się to poprzez pozbawienie mian wyników pomiaru oraz ujednolicenie ich rzędów wielkości. Pierwszy cel normalizacji jest jednoznaczny. Stanowi on warunek *sine qua non* normalizacji. Cel drugi nie jest jednoznaczny, a zatem dopuszcza w tym zakresie różne rozwiązania. Ujednolicenie rzędów wielkości dla zmiennych uzyskuje się np. poprzez ujednolicenie wartości wszystkich zmiennych pod względem zmienności mierzonej odchyleniem standardowym (medianowym odchyleniem bezwzględnym dla miar pozycyjnych) lub przez zapewnienie stałości rozstępu dla znormalizowanych wartości zmiennych. Ogólnie rzecz biorąc ujednolicenie rzędów wielkości uzyskuje się przez wprowadzenie jednolicie określonej wartości zerowej dla wszystkich zmiennych (parametr A_j we wzorze (1)), a następnie przeskalowanie wartości zmiennych (parametr B_j we wzorze (1)).

W artykule zaprezentowano przegląd formuł normalizacyjnych, zaproponowano dwie nowe formuły normalizacyjne, pokazano związki między formułami normalizacyjnymi oraz wskazano przypadki nieprawidłowych formuł normalizacyjnych.

[1] Charakterystykę skal pomiaru zawarto m.in. w pracach (Stevens, 1946; Walesiak, 2011, s. 13–16).

2. FORMUŁY NORMALIZACJI WARTOŚCI ZMIENNYCH

Ze względu na to, że jedynymi dopuszczalnymi przekształceniami na skali przedziałowej i ilorazowej są przekształcenia liniowe, formuły normalizacyjne można wyrazić ogólnym wzorem (Walesiak, 1988; Walesiak, 1990):

$$z_{ij} = b_j x_{ij} + a_j = \frac{x_{ij} - A_j}{B_j} = \frac{1}{B_j} x_{ij} - \frac{A_j}{B_j} \quad (b_j > 0)\,, \tag{1}$$

gdzie: x_{ij} – wartość j-tej zmiennej dla i-tego obiektu,
z_{ij} – znormalizowana wartość j-tej zmiennej dla i-tego obiektu,
A_j – parametr przesunięcia do umownego zera dla j-tej zmiennej,
B_j – parametr skali dla j-tej zmiennej,
$a_j = -A_j/B_j$, $b_j = 1/B_j$ – parametry dla j-tej zmiennej określone w tab. 1.

Szczególnymi przypadkami wzoru (1) są formuły ujęte w tab. 1 (por. np. Abrahamowicz, 1985; Borys, 1978; Grabiński, 1992, s. 35–38; Jajuga, 1981; Jajuga, Walesiak, 2000; Milligan, Cooper, 1988; Młodak, 2006; Nowak, 1990, s. 38–39; Walesiak, 1988; Walesiak, 1993, s. 40; Walesiak, 1996, s. 38–40; Walesiak, 2002, s. 19).

Tabela 1.

Formuły normalizacyjne

Typ	Nazwa formuły	Parametr		Skala pomiaru zmiennych	
		b_j	a_j	przed normalizacją	po normalizacji
n0	Bez normalizacji	–	–	ilorazowa i (lub) przedziałowa	–
n1	Standaryzacja	$1/s_j$	$-\bar{x}_j/s_j$	ilorazowa i (lub) przedziałowa	przedziałowa
n2	Standaryzacja pozycyjna[2]	$1/mad_j$	$-med_j/mad_j$	ilorazowa i (lub) przedziałowa	przedziałowa
n3	Unitaryzacja	$1/r_j$	$-\bar{x}_j/r_j$	ilorazowa i (lub) przedziałowa	przedziałowa
n3a	Unitaryzacja pozycyjna	$1/r_j$	$-med_j/r_j$	ilorazowa i (lub) przedziałowa	przedziałowa

[2] Autorzy pracy Lira, Wagner, Wysocki (2002, s. 91) proponują przemnożenie mianownika przez stałą 1,4826. Uzasadnienie wprowadzenia stałej zawarto w pracy Młodak (2009, s. 18).

n4	Unitaryzacja zerowana	$1/r_j$	$-\min_i\{x_{ij}\}/r_j$	ilorazowa i (lub) przedziałowa	przedziałowa
n5	Normalizacja[3] w przedziale [–1; 1]	$\frac{1}{\max_i \lvert x_{ij}-\bar{x}_j \rvert}$	$\frac{-\bar{x}_j}{\max_i \lvert x_{ij}-\bar{x}_j \rvert}$	ilorazowa i (lub) przedziałowa	przedziałowa
n5a	Normalizacja pozycyjna w przedziale [–1; 1]	$\frac{1}{\max_i \lvert x_{ij}-med_j \rvert}$	$\frac{-med_j}{\max_i \lvert x_{ij}-med_j \rvert}$	ilorazowa i (lub) przedziałowa	przedziałowa
n6	Przekształcenia ilorazowe	$1/s_j$	0	ilorazowa	ilorazowa
n6a		$1/mad_j$	0	ilorazowa	ilorazowa
n7		$1/r_j$	0	ilorazowa	ilorazowa
n8		$1/\max_i\{x_{ij}\}$	0	ilorazowa	ilorazowa
n9		$1/\bar{x}_j$	0	ilorazowa	ilorazowa
n9a		$1/med_j$	0	ilorazowa	ilorazowa
n10		$1/\sum_{i=1}^{n} x_{ij}$	0	ilorazowa	ilorazowa
n11		$1/\sqrt{\sum_{i=1}^{n} x_{ij}^2}$	0	ilorazowa	ilorazowa
n12	Normalizacja	$\frac{1}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}}$	$\frac{-\bar{x}_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}}$	ilorazowa i (lub) przedziałowa	przedziałowa
n12a	Normalizacja pozycyjna	$\frac{1}{\sqrt{\sum_{i=1}^{n}(x_{ij}-med_j)^2}}$	$\frac{-med_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-med_j)^2}}$	ilorazowa i (lub) przedziałowa	przedziałowa
n13	Normalizacja z zerem usytuowanym centralnie[4]	$\frac{1}{r_j/2}$	$-\frac{m_j}{r_j/2}$	ilorazowa i (lub) przedziałowa	przedziałowa

z_{ij} – wartość j-tej zmiennej dla i-tego obiektu, z_{ij} – znormalizowana wartość j-tej zmiennej dla i-tego obiektu, $\bar{x}_j$ – średnia dla j-tej zmiennej, s_j – odchylenie standardowe dla j-tej zmiennej, r_j – rozstęp dla j-tej zmiennej, $m_j = \frac{\max_i\{x_{ij}\}+\min_i\{x_{ij}\}}{2}$ – środek rozstępu (*mid-range*), $med_j = \underset{i}{med}(x_{ij})$ – mediana dla j-tej zmiennej, $mad_j = \underset{i}{mad}(x_{ij})$ – medianowe odchylenie bezwzględne dla j-tej zmiennej.

Źródło: opracowanie własne.

---

3 Zob. Rybaczuk (2002, s. 147).

4 http://www.benetzkorn.com/2011/11/data-normalization-and-standardization/ (dostęp 1.06.2014).

W tab. 1 oprócz znanych formuł normalizacyjnych przedstawiono dwie nowe propozycje określone jako `n12` oraz `n12a`. Punktem wyjścia konstrukcji formuł normalizacyjnych `n12` i `n12a` jest formuła normalizacyjna `n11`. Od wartości x_{ij} odejmuje się w liczniku i mianowniku wartość średnią $\bar{x}_j$ (formuła `n12`) lub medianę $med_j = \underset{i}{med}(x_{ij})$ (formuła `n12a`). Dla formuły normalizacyjnej `n12` odchylenie standardowe maleje wraz ze wzrostem liczebności obserwacji (obiektów) w macierzy danych. Nie stanowi to wady tej formuły normalizacyjnej w statystycznej analizie wielowymiarowej, ponieważ normalizację przeprowadza się dla każdej zmiennej ze zbioru zmiennych dla ustalonej (jednakowej) liczby obserwacji (obiektów).

Normalizację wartości zmiennych należy odróżnić od różnych formuł przekształcających dane, które nie muszą być wyrażone w postaci funkcji liniowej określonej wzorem (1). Np. w porządkowaniu liniowym przy konstrukcji syntetycznego miernika rozwoju zachodzi niekiedy potrzeba ujednolicenia charakteru zmiennych w celu zapewnienia jednolitej preferencji zmiennych. Zmienne destymulanty oraz nominanty przekształca się w stymulanty z wykorzystaniem funkcji liniowych i nieliniowych (zob. np. Walesiak, 2011, s. 10).

Normalizację wartości zmiennych przeprowadza się w pakiecie `clusterSim` (zob. Walesiak, Dudek, 2014) programu R (R Development Core Team, 2014) z wykorzystaniem funkcji:

```
data.Normalization(x,type=”n0”,normalization=”column”)
```

gdzie: `x` – macierz danych,

`type` – typ formuły normalizacyjnej z tab. 1,

`normalization` – rodzaj normalizacji: `”column”` – normalizacja według zmiennych (kolumny w macierzy danych), `”row”` – normalizacja według obiektów (wiersze w macierzy danych).

W tab. 1 przedstawiono wzory na normalizację według zmiennych. Analogiczne wzory można przedstawić dla normalizacji według obiektów. Normalizacja według obiektów ma sens w przypadku, gdy wszystkie zmienne wyrażone są w tej samej jednostce miary. Taki przypadek ma miejsce np. w badaniach strukturalnych. Dalsze rozważania dotyczyć będą normalizacji według zmiennych, choć analogiczne spostrzeżenia odnoszą się do normalizacji według obiektów.

Ujednolicenie rzędów wielkości jest możliwe tylko w razie jednolitego określenia wartości zerowej dla wszystkich zmiennych (zob. Walesiak, 1988). Przekształcenia ilorazowe można stosować tylko wtedy, gdy zmienne są mierzone na skali ilorazowej (istnieje dla niej absolutny punkt zerowy). Gdy zbiór zawiera zmienne mierzone na skali przedziałowej lub przedziałowej i ilorazowej, wówczas do normalizacji można stosować pozostałe formuły normalizacyjne, wprowadzające jednolicie określoną wartość zerową (umowną) dla wszystkich zmiennych. Standaryzacja klasyczna (standaryzacja pozycyjna), normalizacja (normalizacja pozycyjna), unitaryzacja (unitaryzacja pozycyjna), normalizacja w przedziale [–1; 1] (normalizacja pozycyjna w przedziale [–1; 1]) określają umowną wartość zerową na poziomie średniej wartości zmiennej (mediany dla formuł pozycyjnych), unitaryzacja zerowana – na poziomie wartości minimalnej, a normalizacja z zerem usytuowanym centralnie – na poziomie środka rozstępu. Zastosowanie tych formuł normalizacyjnych do zmiennych mierzonych na skali ilorazowej, aczkolwiek formalnie poprawne, spowoduje stratę informacji wskutek „przejścia" wszystkich zmiennych na skalę przedziałową. Strata informacji przejawia się m.in. ograniczeniem zastosowania różnych technik statystycznych i ekonometrycznych.

3. WŁASNOŚCI FORMUŁ NORMALIZACJI WARTOŚCI ZMIENNYCH

Przy wyborze formuły normalizacyjnej należy brać pod uwagę nie tylko skale pomiaru zmiennych, ale również takie charakterystyki rozkładu zmiennych, jak: średnia arytmetyczna (mediana), odchylenie standardowe (medianowe odchylenie bezwzględne) i rozstęp wyznaczony dla znormalizowanych wartości zmiennych (por. tab. 2).

Tabela 2.

Charakterystyki rozkładu wartości zmiennych po normalizacji

Typ	Formuła	Średnia arytmetyczna / mediana*	Odchylenie standardowe / medianowe odchylenie bezwzględne*	Rozstęp
n1	$(x_{ij}-\bar{x}_j)/s_j$	0	1	r_j / s_j
n2	$(x_{ij}-med_j)/mad_j$	0	1	r_j / mad_j
n3	$(x_{ij}-\bar{x}_j)/r_j$	0	s_j / r_j	1
n3a	$(x_{ij}-med_j)/r_j$	0	mad_j / r_j	1
n4	$\left[x_{ij}-\min\limits_i\{x_{ij}\}\right]/r_j$	$\left[\bar{x}_j-\min\limits_i\{x_{ij}\}\right]/r_j$	s_j / r_j	1
n5	$(x_{ij}-\bar{x}_j)/\max\limits_i\lvert x_{ij}-\bar{x}_j\rvert$	0	$s_j/\max\limits_i\lvert x_{ij}-\bar{x}_j\rvert$	$r_j/\max\limits_i\lvert x_{ij}-\bar{x}_j\rvert$
n5a	$(x_{ij}-med_j)/\max\limits_i\lvert x_{ij}-med_j\rvert$	0	$mad_j/\max\limits_i\lvert x_{ij}-med_j\rvert$	$r_j/\max\limits_i\lvert x_{ij}-med_j\rvert$
n6	x_{ij}/s_j	$\bar{x}_j/s_j$	1	r_j / s_j
n6a	x_{ij}/mad_j	$\bar{x}_j/mad_j$	1	r_j / mad_j
n7	x_{ij}/r_j	$\bar{x}_j/r_j$	s_j / r_j	1
n8	$x_{ij}/\max\limits_i\{x_{ij}\}$	$\bar{x}_j/\max\limits_i\{x_{ij}\}$	$s_j/\max\limits_i\{x_{ij}\}$	$r_j/\max\limits_i\{x_{ij}\}$
n9	$x_{ij}/\bar{x}_j$	1	$s_j/\bar{x}_j$	$r_j/\bar{x}_j$
n9a	x_{ij}/med_j	1	mad_j / med_j	r_j / med_j

n10	$x_{ij}\Big/\sum_{i=1}^{n} x_{ij}$	$1/n$	$s_j\Big/\sum_{i=1}^{n} x_{ij}$	$r_j\Big/\sum_{i=1}^{n} x_{ij}$
n11	$x_{ij}\Big/\sqrt{\sum_{i=1}^{n} x_{ij}^2}$	$\bar{x}_j\Big/\sqrt{\sum_{i=1}^{n} x_{ij}^2}$	$s_j\Big/\sqrt{\sum_{i=1}^{n} x_{ij}^2}$	$r_j\Big/\sqrt{\sum_{i=1}^{n} x_{ij}^2}$
n12	$\frac{x_{ij}-\bar{x}_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}}$	0	$\sqrt{\frac{1}{n-1}}$	$\frac{r_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}}$
n12a	$\frac{x_{ij}-med_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-med_j)^2}}$	0	$\frac{mad_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-med_j)^2}}$	$\frac{r_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-med_j)^2}}$
n13	$\frac{x_{ij}-m_j}{r_j/2}$	$\frac{\bar{x}_j-m_j}{r_j/2}$	$\frac{s_j}{r_j/2}$	2

* mediana i medianowe odchylenie bezwzględne dla n2, n3a, n5a, n6a, n9a, n12a.
Źródło: opracowanie własne z wykorzystaniem prac: Jajuga (1981, s. 33), Walesiak (1996, s. 39), Walesiak (2011, s. 20), Jajuga, Walesiak (2000, s. 109), Lira, Wagner, Wysocki (2002, s. 91), Młodak (2006, s. 39–40).

Analiza tab. 2 pozwala sformułować następujące wnioski[5]:

a) formuły normalizacyjne (unitaryzacja, unitaryzacja pozycyjna, unitaryzacja zerowana, przekształcenie ilorazowe z podstawą normalizacji równą rozstępowi, normalizacja z zerem usytuowanym centralnie) są cenne, ponieważ zapewniają znormalizowanym wartościom zmiennych zróżnicowaną zmienność (mierzoną odchyleniem standardowym a dla normalizacji pozycyjnych medianowym odchyleniem bezwzględnym) i jednocześnie stały rozstęp dla wszystkich zmiennych;
b) standaryzacja klasyczna, standaryzacja pozycyjna, normalizacja oraz przekształcenie ilorazowe z podstawą normalizacji równą odchyleniu standardowemu i medianowemu odchyleniu bezwzględnemu powodują ujednolicenie wartości wszystkich zmiennych pod względem zmienności mierzonej odchyleniem standardowym (medianowym odchyleniem bezwzględnym dla miar pozycyjnych); oznacza to wyeliminowanie zmienności jako podstawy różnicowania obiektów;
c) przekształcenia ilorazowe z podstawą normalizacji równą maksimum oraz pierwiastkowi z sumy kwadratów obserwacji zapewniają znormalizowanym wartościom zmiennych zróżnicowaną zmienność, średnią arytmetyczną i rozstęp;
d) przekształcenia ilorazowe z podstawą normalizacji równą sumie, średniej arytmetycznej i medianie, normalizacja pozycyjna, normalizacja w przedziale [–1; 1] oraz normalizacja pozycyjna w przedziale [–1; 1] zapewniają znormalizowanym wartościom zmiennych zróżnicowaną zmienność i rozstęp oraz stałą dla wszystkich zmiennych średnią arytmetyczną (medianę dla miar pozycyjnych); pierwsza formuła stanowi podstawę normalizacji w badaniach strukturalnych (stosuje się tutaj normalizację według obiektów);

[5] Opracowanie własne z wykorzystaniem prac: Jajuga, Walesiak (2000, s. 110–111), Walesiak (2002, s. 20).

e) wszystkie formuły normalizacyjne, będące przekształceniami liniowymi obserwacji na każdej zmiennej, zachowują skośność i kurtozę rozkładu zmiennych[6];
f) dla każdej pary zmiennych wszystkie formuły normalizacyjne nie zmieniają wartości współczynnika korelacji liniowej Pearsona.

4. NORMALIZACJA WARTOŚCI ZMIENNYCH – ZWIĄZKI MIĘDZY FORMUŁAMI NORMALIZACYJNYMI I INNE SPOSTRZEŻENIA

W wyniku zastosowania wybranych formuł normalizacyjnych w dwóch następujących po sobie krokach otrzymuje się wyniki tożsame z zastosowaniem jednej z formuł normalizacyjnych (zob. tab. 3).

Tabela 3.

Formuły normalizacyjne odpowiadające normalizacji dwukrokowej

Zastosowana formuła normalizacyjna		Implikacja	Formuła normalizacyjna
Krok 1	Krok 2		
`n1`	`n7`	$\Rightarrow$	`n3`
`n2`	`n7`	$\Rightarrow$	`n3a`
`n5`	`n7`	$\Rightarrow$	`n3`
`n5a`	`n7`	$\Rightarrow$	`n3a`
`n3`	`n6`	$\Rightarrow$	`n1`
`n3a`	`n6a`	$\Rightarrow$	`n2`

Źródło: opracowanie własne.

W literaturze (por. np. Zeliaś, 2002, s. 794; Młodak, 2006, s. 40) proponowane są następujące formuły normalizacyjne:

$$z_{ij} = x_{ij} \Big/ \sum_{i=1}^{n} x_{ij}^2 , \tag{2}$$

$$z_{ij} = x_{ij} \Big/ \underset{i}{med}(x_{ij}^2) . \tag{3}$$

Formuły te są błędne, ponieważ jednym z celów normalizacji jest pozbawienie mian wyników pomiaru. W tym przypadku nie nastąpi pozbawienie mian wyników pomiaru.

[6] Obliczenia sprawdzające wykonano w pakiecie `e1071` (Meyer i in., 2014) programu R wykorzystując trzy wzory na skośność i kurtozę zaprezentowane w pracy Joanes, Gill (1998).

W literaturze (zob. Grabiński, 1988, s. 245; Grabiński, 1992, s. 35; Pawełek, 2008, s. 57) dyskutowana jest ogólna formuła normalizacyjna o postaci:

$$z_{ij} = \left(\frac{x_{ij} - A_j}{B_j} \right)^{p_j}, \tag{4}$$

gdzie: A_j – parametr przesunięcia do umownego zera dla *j*-tej zmiennej,
B_j – parametr skali dla *j*-tej zmiennej,
P_j – dodatnia liczba na ogół równa 1/2, 1, 2, ...

Tylko $\underset{j}{\forall} p_j = 1$ formuła ta jest identyczna z normalizacyjnym przekształceniem liniowym o postaci (1). Zastosowanie innych wartości w potędze spowoduje, że otrzyma się znormalizowane wartości zmiennych, które nie zachowają dwóch podstawowych własności formuł normalizacyjnych:

a) skośność i kurtoza rozkładu zmiennych przed i po normalizacji będzie inna,
b) współczynniki korelacji liniowej Pearsona dla każdej pary zmiennych przed i po normalizacji będą miały inne wartości.

5. PODSUMOWANIE

W artykule zaprezentowano przegląd formuł normalizacyjnych wartości zmiennych wyrażonych ogólną formułą liniową o postaci (1). Szczególne przypadki tej formuły ujęto w tab. 1.

Własności zaprezentowanych formuł normalizacyjnych przedstawiono w tab. 2. Przy wyborze formuły normalizacyjnej należy brać pod uwagę nie tylko skale pomiaru zmiennych, ale również takie charakterystyki rozkładu zmiennych, jak: średnia arytmetyczna (mediana dla formuł pozycyjnych), odchylenie standardowe (medianowe odchylenie bezwzględne dla formuł pozycyjnych) i rozstęp wyznaczony dla znormalizowanych wartości zmiennych.

Ponadto zaproponowano dwie nowe formuły normalizacyjne (`n12` i `n12a`), pokazano związki między formułami normalizacyjnymi oraz wskazano przypadki nieprawidłowych formuł normalizacyjnych.

Uniwersytet Ekonomiczny we Wrocławiu

LITERATURA

Abrahamowicz M., (1985), Konstrukcja syntetycznych mierników rozwoju w świetle twierdzenia Arrowa, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, nr 311, 5–25.

Borys T., (1978), Metody normowania cech w statystycznych badaniach porównawczych, *Przegląd Statystyczny*, 25 (2), 227–239.

Grabiński T., (1988), Metody statystycznej analizy porównawczej, w: Zeliaś A., (red.), *Metody statystyki międzynarodowej*, PWE, Warszawa, 235–260.

Grabiński T., (1992), *Metody taksonometrii*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

Jajuga K., (1981), *Metody analizy wielowymiarowej w ilościowych badaniach przestrzennych*, Akademia Ekonomiczna we Wrocławiu, Wrocław (praca doktorska).

Jajuga K., Walesiak M., (2000), Standardisation of data set under different measurement scales, w: Decker R., Gaul W., (red.), *Classification and information processing at the turn of the millennium*, Springer-Verlag, Berlin, Heidelberg, 105–112.

Joanes D. N., Gill C. A., (1998), Comparing Measures of Sample Skewness and Kurtosis, *The Statistician*, 47, 183–189.

Lira J., Wagner W., Wysocki F., (2002), Mediana w zagadnieniach porządkowania liniowego obiektów wielocechowych, w: Paradysz J. (red.), *Statystyka regionalna w służbie samorządu lokalnego i biznesu*, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej, Akademia Ekonomiczna w Poznaniu, Poznań, 87–99.

Meyer D., Dimitriadou E., Hornik K., Weingessel A., Leisch F., Chang C., Lin C., (2014), *e1071 pakkage*, URL http://www.R-project.org.

Milligan G. W., Cooper M. C., (1988), A Study of Standardization of Variables in Cluster Analysis, *Journal of Classification*, 5 (2), 181–204.

Młodak A., (2006), *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa.

Młodak A., (2009), Historia problemu Webera, *Matematyka Stosowana*, 37 (1), tom 10/51, 3–21.

Nowak E., (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa.

Pawełek B., (2008), *Metody normalizacji zmiennych w badaniach porównawczych złożonych zjawisk ekonomicznych*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.

R Development Core Team, (2014), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, URL http://www.R-project.org.

Rybaczuk M., (2002), Graficzna prezentacja struktury danych wielowymiarowych, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, nr 942, 146–153.

Stevens S.S., (1946), On the Theory of Scales of Measurement, *Science*, 103 (2684), 677–680.

Walesiak M., (1988), Skale pomiaru cech (w ujęciu zwężonym) a zagadnienie wyboru postaci analitycznej syntetycznych mierników rozwoju, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, nr 447, 63–71.

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

Walesiak M., (1993), *Statystyczna analiza wielowymiarowa w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 654, Seria: Monografie i Opracowania nr 101.

Walesiak M., (1996), *Metody analizy danych marketingowych*, PWN, Warszawa.

Walesiak M., (2002), *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.

Walesiak M., (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.

Walesiak M., Dudek A., (2014), c*lusterSim Package*, URL http://www.R-project.org.
Zeliaś A., (2002), Some Notes on the Selection of Normalisation of Diagnostic Variables, *Statistics in Transition*, 5 (5), 787–802.

PRZEGLĄD FORMUŁ NORMALIZACJI WARTOŚCI ZMIENNYCH ORAZ ICH WŁASNOŚCI W STATYSTYCZNEJ ANALIZIE WIELOWYMIAROWEJ

S t r e s z c z e n i e

Celem normalizacji wartości zmiennych jest doprowadzenie zmiennych do porównywalności poprzez pozbawienie mian wyników pomiaru oraz ujednolicenie ich rzędów wielkości. W artykule zaprezentowano przegląd formuł normalizacyjnych wartości zmiennych oraz ich własności. Zaproponowano dwie nowe formuły normalizacyjne, pokazano związki między formułami normalizacyjnymi oraz wskazano nieprawidłowe formuły normalizacyjne.

Słowa kluczowe: normalizacja, standaryzacja, unitaryzacja, przekształcenia ilorazowe, własności formuł normalizacyjnych

DATA NORMALIZATION IN MULTIVARIATE DATA ANALYSIS. AN OVERVIEW AND PROPERTIES

A b s t r a c t

The purpose of normalization is to adjust the size (magnitude) and the relative weighting of the input variables. The article presents an overview of the normalization formulas and their properties. Moreover a new formulas of normalization of the values of variables are proposed. The article discusses connection among normalization formulas and indicates incorrect normalization formulas.

Keywords: normalization, standardization, unitarization, quotient transformation, normalization formulas properties