

Marek Walesiak

**UOGÓLNIANA MIARA ODLEGŁOŚCI GDM
W STATYSTYCZNEJ ANALIZIE
WIELOWYMIAROWEJ
Z WYKORZYSTANIEM PROGRAMU R**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu

Marek Walesiak

**UOGÓLNIONA MIARA ODLEGŁOŚCI GDM
W STATYSTYCZNEJ ANALIZIE
WIELOWYMIAROWEJ
Z WYKORZYSTANIEM PROGRAMU R**



**Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2011**

Senacka Komisja Wydawnicza

*Zdzisław Pisz (przewodniczący),
Andrzej Bąk, Krzysztof Jajuga, Andrzej Matysiak, Waldemar Podgórski,
Mieczysław Przybyła, Aniela Styś, Stanisław Urban*

Recenzent

Andrzej Sokołowski

Redakcja wydawnicza

Dorota Pitulec

Redakcja techniczna

Barbara Łopusiewicz

Korekta

Barbara Cibis

Skład i łamanie

Beata Mazur

Projekt okładki

Beata Dębska

Na okładce wykorzystano zdjęcie z zasobów 123 Royalty Free

Tytuł dofinansowano ze środków na działalność statutową Katedry Ekonometrii i Informatyki
Uniwersytetu Ekonomicznego we Wrocławiu

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISBN 978-83-7695-132-4

Druk: Drukarnia TOTEM

SPIS TREŚCI

WSTĘP	7
1. PODSTAWOWE ZAGADNIENIA STATYSTYCZNEJ ANALIZY WIELOWYMIAROWEJ	11
1.1. Zagadnienia wstępne	11
1.2. Typy skal pomiarowych i ich charakterystyka	13
1.3. Transformacja normalizacyjna i ujednocianie zmiennych	16
1.4. Pomiar podobieństwa obiektów w świetle skal pomiaru i wag zmiennych	22
1.5. Strategie postępowania w pomiarze odległości dla danych porządkowych	31
2. UOGÓLNIONA MIARA ODLEGŁOŚCI GDM	36
2.1. Wprowadzenie	36
2.2. Uogólniony współczynnik korelacji	36
2.3. Charakterystyka uogólnionej miary odległości	38
2.4. Silne i słabe strony uogólnionej miary odległości	44
2.5. Postać uogólnionej miary odległości dla zmiennych z różnych skal pomiaru	46
2.6. Postać uogólnionej miary odległości dla zróżnicowanych wag zmiennych	47
2.7. Kwadrat odległości euklidesowej a współczynnik korelacji liniowej Pearsona i cosinus kąta między wektorami	48
2.8. GDM a współczynnik korelacji liniowej Pearsona i cosinus kąta między wektorami	50
3. OBSZARY ZASTOSOWAŃ UOGÓLnionej MIARY ODLEGŁOŚCI GDM W STATYSTYCZNEJ ANALIZIE WIELOWYMIAROWEJ	54
3.1. Wyznaczanie macierzy odległości w procesie klasyfikacji obiektów ...	54
3.2. Ocena podobieństwa wyników klasyfikacji zbioru obiektów w czasie	68
3.3. Uogólniona miara odległości GDM jako syntetyczny miernik rozwoju w metodach porządkowania liniowego	73
3.4. Ocena podobieństwa wyników porządkowania liniowego zbioru obiektów w czasie	78
4. UOGÓLNIONA MIARA ODLEGŁOŚCI GDM W ŚWIETLE WYBRANYCH EKSPERYMENTÓW SYMULACYJNYCH	82
4.1. Losowe generowanie danych o znanej strukturze klas w pakiecie clusterSim	82

4.2. Analiza porównawcza metod klasyfikacji dla danych o znanej strukturze klas	87
4.3. Ocena wybranych procedur analizy skupień dla danych porządkowych	92
5. WYBRANE ZASTOSOWANIA UOGÓLNIONEJ MIARY ODLEGŁOŚCI GDM Z WYKORZYSTANIEM PROGRAMU R	98
5.1. Porządkowanie liniowe zbioru obiektów na podstawie danych porządkowych z rynku nieruchomości	98
5.2. Porządkowanie liniowe zbioru obiektów na podstawie danych metrycznych dotyczących warunków zamieszkiwania ludności w miastach ...	102
5.3. Ocena podobieństwa wyników porządkowania liniowego zbioru obiektów w czasie na podstawie danych metrycznych dotyczących warunków zamieszkiwania ludności w miastach	105
5.4. Analiza skupień zbioru obiektów opisanych danymi porządkowymi z rynku nieruchomości	108
5.5. Analiza skupień zbioru obiektów opisanych danymi metrycznymi dotyczącymi zanieczyszczenia powietrza	112
LITERATURA	118
SKOROWIDZ	125
SPIS RYSUNKÓW	129
SPIS TABEL	130

WSTĘP

Prezentowana książka stanowi podsumowanie rozważań autora zawartych w wielu opracowaniach dotyczących miary odległości, która została w pierwotnej wersji zaproponowana dla zmiennych porządkowych [Walesiak 1993a, s. 44-45], a następnie dla danych metrycznych [Walesiak 2002a] i nominalnych [Walesiak 2003c]. Podstawowe części książki zostały opublikowane m.in. w „Argumenta Oeconomica”, „Przeglądzie Statystycznym”, „Badaniach Operacyjnych i Decyzjach”, Pracach Naukowych Akademii Ekonomicznej we Wrocławiu (obecnie Uniwersytetu Ekonomicznego we Wrocławiu) oraz były referowane na konferencjach naukowych, w tym na konferencji Sekcji Klasyfikacji i Analizy Danych PTS (zob. [Walesiak, Bąk, Jajuga 2002; Walesiak 2003b; 2004b; 2011b; Walesiak, Dudek 2009a; 2010b]), konferencji Światowej Federacji Towarzystw Klasyfikacyjnych IFCS (zob. [Walesiak, Dziechciarz, Bąk 1998; Walesiak, Dudek 2010a]) oraz Niemieckiego Towarzystwa Klasyfikacyjnego (zob. [Jajuga, Walesiak, Bąk 2003]).

Dotychczas uogólniona miara odległości zaprezentowana została w zwartej postaci w dwóch wydaniach książkowych Wydawnictwa AE we Wrocławiu (zob. [Walesiak 2002b; 2006]). Obecna monografia zawiera istotne zmiany i uzupełnienia wynikające w znacznej mierze z oprogramowania miary GDM w pakiecie `clusterSim` programu **R**. Całkowicie nowe są podrozdziały 1.5, 3.1 i 3.3 oraz rozdziały czwarty i piąty.

Praca składa się z pięciu rozdziałów.

W rozdziale pierwszym przedstawiono podstawowe zagadnienia statystycznej analizy wielowymiarowej. Wyjaśniono w nim takie podstawowe pojęcia, jak obiekt, zmienna, macierz i kostka danych. Scharakteryzowano typy skal pomiarowych oraz zagadnienie transformacji normalizacyjnej i ujednociania zmiennych z punktu widzenia skal pomiarowych. Ponadto zaprezentowano szeroką klasyfikację miar podobieństwa obiektów z uwzględnieniem problematyki ważenia zmiennych oraz skal ich pomiaru. Rozdział kończą rozważania dotyczące strategii postępowania w pomiarze odległości dla danych porządkowych.

W rozdziale drugim przedstawiono szczegółową charakterystykę uogólnionej miary odległości GDM (*Generalised Distance Measure*). W konstrukcji miary odległości GDM wykorzystano ideę uogólnionego współczynnika korelacji, który obejmuje współczynnik korelacji liniowej Pearsona i współczynnik korelacji zmiennych porządkowych tau Kendalla. W związku z tym w części pierwszej tego rozdziału zaprezentowano uogólniony współczynnik korelacji. W dalszej części scharakteryzowano uogólnioną miarę odległości GDM dla jednakowych i zróżnicowanych wag zmiennych. Następnie wskazano silne i słabe strony uogólnionej miary odległości.

Rozważania teoretyczne zilustrowano licznymi przykładami poglądowymi. Zaprezentowano postać uogólnionej miary odległości GDM uwzględniającą zmienne mierzone na skali metrycznej, porządkowej, nominalnej oraz zmienne z różnych skal pomiaru. Ponadto przedstawiono analizę związków między kwadratem odległości euklidesowej a współczynnikiem korelacji liniowej Pearsona i cosinusem kąta między wektorami oraz między uogólnioną miarą odległości GDM a współczynnikiem korelacji liniowej Pearsona i cosinusem kąta między wektorami.

W rozdziale trzecim zaprezentowano obszary zastosowań uogólnionej miary odległości w statystycznej analizie wielowymiarowej. Podstawowymi obszarami zastosowań tej miary są wyznaczanie macierzy odległości w procesie klasyfikacji zbioru obiektów oraz zastosowanie miary GDM jako syntetycznego miernika rozwoju w metodach porządkowania liniowego. Ponadto w rozdziale tym zaprezentowano metody oceny podobieństwa wyników klasyfikacji zbioru obiektów oraz oceny podobieństwa wyników porządkowania liniowego zbioru obiektów w czasie.

Rozdział czwarty zawiera rezultaty wybranych eksperymentów symulacyjnych pozwalających ocenić zachowanie się uogólnionej miary odległości GDM przy różnych strukturach danych. W pierwszym podrozdziale scharakteryzowano zagadnienie losowego generowania danych o znanej strukturze klas w pakiecie `clusterSim`. W drugim podrozdziale przedstawiono analizę porównawczą metod klasyfikacji dla danych o znanej strukturze klas dla trzech typów danych. W dwóch pierwszych eksperymentach wykorzystano dane metryczne oraz porządkowe o znanej strukturze klas obiektów wygenerowane z wykorzystaniem z funkcji `cluster.Gen` pakietu `clusterSim`. W eksperymencie trzecim zbiory danych utworzono z wykorzystaniem funkcji pakietu `mlbench` (`spirals`, `smiley`, `cassini`) oraz zbiorów własnych (`worms`, `w3`, `skad`). W podrozdziale trzecim, na podstawie porządkowych danych symulacyjnych wygenerowanych z wykorzystaniem z funkcji `cluster.Gen` pakietu `clusterSim`, przeprowadzono ocenę przydatności wybranych procedur analizy skupień obejmujących miarę odległości GDM, dziewięć metod klasyfikacji oraz osiem indeksów służących ustaleniu liczby klas.

W rozdziale piątym zaprezentowano wybrane zastosowania uogólnionej miary odległości GDM1 i GDM2 w statystycznej analizie wielowymiarowej z wykorzystaniem programu **R**. Znaczna część skryptów wykorzystuje pakiet `clusterSim`. Zastosowania dotyczyły porządkowania liniowego i analizy skupień zbioru obiektów na podstawie danych porządkowych z rynku nieruchomości oraz porządkowania liniowego na podstawie danych metrycznych dotyczących warunków zamieszkiwania ludności w miastach i analizy skupień obiektów opisanych danymi metrycznymi dotyczącymi zanieczyszczenia powietrza. Ponadto dokonano oceny podobieństwa wyników porządkowania liniowego zbioru obiektów w czasie na podstawie danych metrycznych dotyczących warunków zamieszkiwania ludności w miastach.

Pracę zamyka zestawienie wykorzystywanej literatury, spis rysunków i tabel oraz skorowidz rzeczowy.

Wersję instalacyjną programu **R** oraz dodatkowe pakiety (w tym pakiet `clusterSim` autorstwa Marka Walesiaka i Andrzeja Dudka) można pobrać ze strony: <http://www.r-project.org/>. Wszystkie skrypty zawarte w książce przetestowano, używając wersji 2.14.1 programu **R**.

Na stronie internetowej <http://keii.ue.wroc.pl> znajdują się pliki zawierające wykorzystywane dane oraz skrypty realizujące zastosowania zamieszczone w książce.

Książka jest przeznaczona dla pracowników naukowych zajmujących się zastosowaniem metod statystycznej analizy wielowymiarowej w każdej dziedzinie wiedzy, w tym w badaniach ekonomicznych. Ponadto odbiorcami książki mogą być słuchacze wyższych uczelni studiujący zagadnienia statystycznej analizy wielowymiarowej i jej zastosowań.

LITERATURA

- Abrahamowicz M. (1985), *Konstrukcja syntetycznych mierników rozwoju w świetle twierdzenia Arrowa*, Prace Naukowe AE we Wrocławiu nr 311, AE, Wrocław, 5-25.
- Abrahamowicz M., Zajac K. (1986), *Metoda ważenia zmiennych w taksonomii numerycznej i procedurach porządkowania liniowego*, Prace Naukowe AE we Wrocławiu nr 328, AE, Wrocław, 5-17.
- Ackoff R.L. (1969), *Decyzje optymalne w badaniach stosowanych*, PWN, Warszawa.
- Adams E.W., Fagot R.F., Robinson R.E. (1965), *A theory of appropriate statistics*, „Psychometrika”, (30), 99-127.
- Ajvazjan S.A., Mchitarjan W.S. (1998), *Prikladnaja statistika i osnovy ekonometriki*, Juniti, Moskwa.
- Aldenderfer M.S., Blashfield R.K. (1984), *Cluster analysis*, Sage, Beverly Hills.
- Anderberg M.R. (1973), *Cluster analysis for applications*, Academic Press, New York, San Francisco, London.
- Atlas R.S., Overall J.E. (1994), *Comparative evaluation of two superior stopping rules for hierarchical cluster analysis*, „Psychometrika”, 59, 581-591.
- Bąk A. (1999), *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*, AE, Wrocław.
- Bock H.H., Diday E. (ed.) (2000), *Analysis of symbolic data*, Springer-Verlag, Berlin, Heidelberg.
- Borys T. (1978), *Metody normowania cech statystycznych w badaniach porównawczych*. „Przegląd Statystyczny”, z. 2, 227-239.
- Borys T. (1984), *Kategoria jakości w statystycznej analizie porównawczej*, Prace Naukowe AE we Wrocławiu nr 284, Seria: Monografie i Opracowania nr 23, AE, Wrocław.
- Breckenridge J.N. (2000), *Validating cluster analysis: consistent replication and symmetry*, „Multivariate Behavioral Research”, 35 (2), 261-285.
- Carmone F.J., Kara A., Maxwell S. (1999), *HINoV: a new method to improve market segment definition by identifying noisy variables*, „Journal of Marketing Research”, November, vol. 36, 501-509.
- Cegiełka K., Stachowski E., Szymański K. (red.) (2000), *Matematyka. Encyklopedia dla wszystkich*, WNT, Warszawa.
- Choyński M. (1971), *Pomiar w psychologii*, [w:] J. Koziński (red.), *Problemy psychologii matematycznej*, PWN, Warszawa, 15-41.
- Cormack R.M. (1971), *A review of classification (with discussion)*, „Journal of the Royal Statistical Society”, A, part 3, 321-367.
- Cox T.F., Cox M.A.A. (2000), *A general weighted two-way dissimilarity coefficient*, „Journal of Classification”, vol. 17, 101-121.
- Cramer H. (1958), *Metody matematyczne w statystyce*, PWN, Warszawa.
- Czerwińska D., Gemborowski H. (1975), *O współczynniku Renkonena podobieństwa zbiorów*, „Listy Biometryczne”, nr 49-50, 19-24.
- Dąbrowski M., Laus-Maczyńska K. (1978), *Metody wyszukiwania i klasyfikacji informacji*, WNT, Warszawa.
- Domański C. (1979), *Statystyczne testy nieparametryczne*, PWE, Warszawa.
- Dudoit S., Fridlyand J. (2002), *A prediction-based resampling method for estimating the number of clusters in a dataset*, „Genome Biology”, 3(7).
- Dziechciarz J., Strahl D., Walesiak M. (2001), *Data set normalisation for banks performance assessment*, Prace Naukowe AE we Wrocławiu nr 915, AE, Wrocław, 21-32.
- Everitt B.S. (1979), *The analysis of contingency tables*, Chapman and Hall, London.
- Everitt B.S., Landau S., Leese M. (2001), *Cluster analysis*, Edward Arnold, London.

- Everitt B.S., Landau S., Leese M., Stahl D. (2011), *Cluster analysis*, Wiley, Chichester.
- Fischer I., Poland J. (2004), *New methods for spectral clustering*, Technical Report No. IDSIA-12-04, Dalle Molle Institute for Artificial Intelligence, Manno-Lugano, Switzerland.
- Fowlkes E.B., Mallows C.L. (1983), *A method for comparing two hierarchical clusterings*, „Journal of the American Statistical Association”, no. 383, 553-569.
- Gatnar E. (1998), *Symboliczne metody klasyfikacji danych*, PWN, Warszawa.
- Girolami M. (2002), *Mercer kernel-based clustering in feature space*, IEEE Transactions on Neural Networks, vol. 13, no. 3, 780-784.
- Gnanadesikan R., Kettenring J.R., Tsao S.L. (1995), *Weighting and selection of variables for cluster analysis*, „Journal of Classification”, vol. 12, 113-136.
- Goodman L.A., Kruskal W.H. (1979), *Measures of association for cross classifications*, Springer-Verlag, New York, Heidelberg.
- Gordon A.D. (1981), *Classification*, Chapman and Hall, London.
- Gordon A.D. (1987), *A review of hierarchical classification*, „Journal of the Royal Statistical Society”, A, 119-137.
- Gordon A.D. (1996), *Hierarchical classification*, [w:] P. Arabie, L.J. Hubert, G. de Soete (eds.), *Clustering and classification*, World Scientific, Singapore, 65-121.
- Gordon A.D. (1999), *Classification*, 2nd Edition, Chapman and Hall/CRC, London.
- Gower J.C. (1971), *A general coefficient of similarity and some of its properties*, „Biometrics”, vol. 27, no. 4, 857-871.
- Grabiński T. (1980), *Metody analizy zbieżności wyników dyskryminacji zbiorów*, Zeszyty Naukowe AE w Krakowie nr 127, AE, Kraków, 5-36.
- Grabiński T. (1984), *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Zeszyty Naukowe AE w Krakowie, Seria specjalna: Monografie nr 61, AE, Kraków.
- Grabiński T. (1992), *Metody taksonometrii*, AE, Kraków.
- Grabiński T., Wydymus S., Zeliaś A. (1989), *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa.
- Grabisch M. (2001), *On preference representation on an ordinal scale*, [w:] S. Benferhat, P. Besnard (eds.), *Symbolic and quantitative approaches to reasoning with uncertainty*, Springer-Verlag, Berlin, Heidelberg, New York, 18-28.
- Hand D.J. (1996), *Statistics and the theory of measurement*, „Journal of the Royal Statistical Society”, A, (3), 445-492.
- Hastie T., Tibshirani R., Friedman J. (2001), *The elements of statistical learning. Data mining, inference, and prediction*, Springer-Verlag, New York, Berlin, Heidelberg.
- Hellwig Z. (1968), *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju i strukturę wykwalifikowanych kadr*, „Przegląd Statystyczny”, z. 4, 307-327.
- Hellwig Z. (1975), *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa.
- Hellwig Z. (1981), *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych*, [w:] W. Welfe (red.), *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, PWE, Warszawa, 46-68.
- Hubert L., Arabie P. (1985), *Comparing partitions*, „Journal of Classification”, no. 1, 193-218.
- Jajuga K. (1981), *Metody analizy wielowymiarowej w ilościowych badaniach przestrzennych*, Praca doktorska, AE, Wrocław
- Jajuga K. (1987), *Statystyka ekonomicznych zjawisk złożonych – wykrywanie i analiza niejednorodnych rozkładów wielowymiarowych*, Prace Naukowe AE we Wrocławiu nr 371, Seria: Monografie i Opracowania nr 39, AE, Wrocław.
- Jajuga K. (1989), *Podstawowe metody analizy wielowymiarowej w przypadku występowania zmiennych mierzonych na różnych skalach*, Praca wykonana w ramach CPBP 10.09, AE, Wrocław.

- Jajuga K. (1990), *Statystyczna teoria rozpoznawania obrazów*, PWN, Warszawa.
- Jajuga K. (1993), *Statystyczna analiza wielowymiarowa*, PWN, Warszawa.
- Jajuga K., Walesiak M. (2000), *Standardisation of data set under different measurement scales*, [w:] R. Decker, W. Gaul (eds.), *Classification and information processing at the turn of the millennium*, Springer-Verlag, Berlin, Heidelberg, 105-112.
- Jajuga K., Walesiak M. (2004), *Remarks on the dependence measures and the distance measures*, Prace Naukowe AE we Wrocławiu nr 1022, AE, Wrocław, 348-354.
- Jajuga K., Walesiak M., Bąk A. (2003), *On the general distance measure*, [w:] M. Schwaiger, O. Opitz (eds.), *Exploratory data analysis in empirical research*, Springer-Verlag, Berlin, Heidelberg, 104-109.
- Karatzoglou A. (2006), *Kernel methods. Software, algorithms and applications*, Rozprawa doktorska, Uniwersytet Techniczny w Wiedniu.
- Kaufman L., Rousseeuw P.J. (1990), *Finding groups in data: an introduction to cluster analysis*, Wiley, New York.
- Kendall M.G. (1955), *Rank correlation methods*, Griffin, London.
- Kendall M.G. (1966), *Discrimination and classification*, [w:] P.R. Krishnaiah (ed.), *Multivariate analysis I*, Academic Press, New York, London, 165-185.
- Kendall M.G., Buckland W.R. (1986), *Słownik terminów statystycznych*, PWE, Warszawa.
- Knapp T.R. (1990), *Treating ordinal scales as interval scales: an attempt to resolve the controversy*, „Nursing Research”, vol. 39, no. 2, 121-123.
- Kolonko J. (1979), *O wykorzystaniu w badaniach taksonomicznych danych pierwotnych mierzonych na skalach różnego typu*, Materiały konferencyjne nt. *Metody taksonomiczne i ich zastosowanie w badaniach ekonomicznych*, Szklarska Poręba, 25 października 1979 r. (materiał powielony).
- Kolonko J. (1980), *Analiza dyskryminacyjna i jej zastosowania w ekonomii*, PWN, Warszawa.
- Kolupa M. (1976), *Elementarny wykład algebry liniowej dla ekonomistów*, PWN, Warszawa.
- Kukuła K. (2000), *Metoda unitaryzacji zerowanej*, PWN, Warszawa.
- Legendre P., Legendre L. (2003), *Numerical ecology*, Elsevier, Amsterdam.
- Lerman J.C. (1988), *Comparing partitions (mathematical and statistical aspects)*, [w:] H.H. Bock (ed.), *Classification and related methods of data analysis*, North-Holland, Amsterdam, 121-131.
- Lira J., Wagner W., Wysocki F. (2002), *Mediana w zagadnieniach porządkowania liniowego obiektów wielocechowych*, [w:] J. Paradysz (red.), *Statystyka regionalna w służbie samorządu lokalnego i biznesu*, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej, AE, Poznań, 87-99.
- Macnaughton-Smith P., Williams W.T., Dale M.B., Mockett L.G. (1964), *Dissimilarity analysis: a new technique of hierarchical sub-division*, „Nature”, 202, 1034-1035.
- Makarenkov V., Legendre P. (2001), *Optimal variable weighting for ultrametric and additive trees and K-means partitioning methods and software*, „Journal of Classification”, vol. 18, 245-271.
- Milligan G.W. (1985), *An algorithm for generating artificial test clusters*, „Psychometrika”, no. 1, 123-127.
- Milligan G.W. (1989), *A validation study of a variable weighting algorithm for cluster analysis*, „Journal of Classification”, no. 1, 53-71.
- Milligan G.W. (1994), *Issues in applied classification: selection of variables to cluster*, Classification Society of North America Newsletter, November, Issue 37.
- Milligan G.W. (1996), *Clustering validation: results and implications for applied analyses*, [w:] P. Arabie, L.J. Hubert, G. de Soete (eds.), *Clustering and classification*, World Scientific, Singapore, 341-375.
- Milligan G.W., Cooper M.C. (1988), *A study of standardization of variables in cluster analysis*, „Journal of Classification”, no. 2, 181-204.

- Ng A., Jordan M., Weiss Y. (2002), *On spectral clustering: analysis and an algorithm*, W: T. Dietterich, S. Becker, Z. Ghahramani (eds.), *Advances in Neural Information Processing Systems 14*, MIT Press, 849-856.
- Nowak E. (1985), *Wskaźnik podobieństwa wyników podziałów*, „Przegląd Statystyczny”, z. 1, 41-48.
- Nowak E. (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa.
- Ostasiewicz W. (1975), *Możliwość uogólnienia miar zależności stochastycznej*, „Przegląd Statystyczny”, z. 2, 229-240.
- Ostasiewicz W. (red.) (1998), *Statystyczne metody analizy danych*, AE, Wrocław.
- Pawlukowicz R. (2006), *Klasyfikacja w wyborze nieruchomości podobnych dla potrzeb wyceny rynkowej nieruchomości*, *Ekonometria 16*, Prace Naukowe AE we Wrocławiu nr 1100, AE, Wrocław, 232-240.
- Pawlukowicz R. (2010), *Wykorzystanie metodyki porządkowania liniowego do określania wartości rynkowej nieruchomości*, *Taksonomia 17*, Prace Naukowe UE we Wrocławiu nr 107, UE, Wrocław, 377-385.
- Pawłowski T. (1969), *Metodologiczne zagadnienia humanistyki*, PWN, Warszawa.
- Pawłowski Z. (1973), *Prognozy ekonometryczne*, PWN, Warszawa.
- Perona P., Freeman W.T. (1998), *A factorization approach to grouping*, *Lecture Notes In Computer Science*; vol. 1406, *Proceedings of the 5th European Conference on Computer Vision*, vol. I, 655-670.
- Pociecha J. (1986), *Statystyczne metody segmentacji rynku*, *Zeszyty Naukowe AE w Krakowie*, Seria specjalna: Monografie nr 71, AE, Kraków.
- Podani J. (1999), *Extending Gowers general coefficient of similarity to ordinal characters*, „Taxon”, 48, 331-340.
- Poland J., Zeugmann T. (2006), *Clustering the Google distance with eigenvectors and semidefinite programming*, *Knowledge Media Technologies, First International Core-to-Core Workshop, Dagstuhl*, July 23-27, 2006, Germany (Klaus P. Jantke & Gunther Kreuzberger, Eds., *Diskussionsbeiträge*, Institut für Medien und Kommunikationswissenschaft, Technische Universität Ilmenau, No. 21, pp. 61-69, July 2006).
- Prymon M. (1991), *Marketing w kierowaniu firmą*, AG Master, Wrocław.
- Qiu W., Joe H. (2006), *Generation of random clusters with specified degree of separation*, „*Journal of Classification*”, vol. 23, 315-334.
- R Development Core Team (2011), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, URL <http://www.R-project.org>.
- Rand W.M. (1971), *Objective criteria for the evaluation of clustering methods*, „*Journal of the American Statistical Association*”, no. 336, 846-850.
- Rybaczuk M. (2002), *Graficzna prezentacja struktury danych wielowymiarowych*, Prace Naukowe AE we Wrocławiu nr 942, AE, Wrocław, 146-153.
- Sneath P.H.A., Sokal R.R. (1973), *Numerical taxonomy*, W.H. Freeman and Co., San Francisco.
- Soffritti G. (2003), *Identifying multiple cluster structures in a data matrix*, „*Communications in Statistics. Simulation and Computation*”, vol. 32, no. 4, 1151-1177.
- Sokołowski A. (1976), *Metoda porównywania wyników podziału zbioru skończonego*, XII Konferencja Naukowa Statystyków, Ekonometryków i Matematyków Akademii Ekonomicznych Polski Południowej, Karpacz, 2-4.04.1976 r. (materiał powielony).
- Sokołowski A. (1992), *Empiryczne testy istotności w taksonomii*, *Zeszyty Naukowe AE w Krakowie*, Seria specjalna: Monografie nr 108, AE, Kraków.
- Steczkowski J., Zeliaś A. (1997), *Metody statystyczne w badaniach cech jakościowych*, AE, Kraków.
- Steczkowski J., Zeliaś A. (1981), *Statystyczne metody analizy cech jakościowych*, PWE, Warszawa.
- Steinley D., Henson R. (2005), *OCLUS: an analytic method for generating clusters with known overlap*, „*Journal of Classification*”, vol. 22, 221-250.

- Stevens S.S. (1946), *On the theory of scales of measurement*, „Science”, vol. 103, no. 2684, 677-680.
- Stevens S.S. (1959), *Measurement, psychophysics and utility*, [w:] C.W. Churchman, P. Ratoosh (eds.), *Measurement; definitions and theories*, Wiley, New York, 18-61.
- Strahl D. (1978), *Propozycja konstrukcji miary syntetycznej*, „Przegląd Statystyczny”, z. 2, 205-215.
- Strahl D., Walesiak M. (1997), *Normalizacja zmiennych w skali przedziałowej i ilorazowej w referencyjnym systemie granicznym*, „Przegląd Statystyczny”, z. 1, 69-77.
- Szmigiel C. (1976), *Wskaźnik zgodności kryteriów podziału*, „Przegląd Statystyczny”, z. 4, 491-498.
- Tibshirani R., Walther G. (2005), *Cluster validation by predicting strength*, „Journal of Computational and Graphical Statistics”, vol. 14, no. 3, 511-528.
- Tibshirani R., Walther G., Hastie T. (2001), *Estimating the number of clusters in a data set via the gap statistic*, „Journal of the Royal Statistical Society”, B, vol. 63, part 2, 411-423.
- von Luxburg U. (2006), *A tutorial on spectral clustering*, Max Planck Institute for Biological Cybernetics, Technical Report TR-149.
- Walenta K. (1971), *Podstawowe pojęcia teorii pomiaru*, [w:] J. Koziński (red.), *Problemy psychologii matematycznej*, PWN, Warszawa, 43-63.
- Walesiak M. (1985a), *Metody klasyfikacji w badaniach strukturalnych*, Rozprawa doktorska, AE, Wrocław.
- Walesiak M. (1985b), *Sposoby rejestracji zmian w czasie w wynikach klasyfikacji*, Prace Naukowe AE we Wrocławiu nr 311, AE, Wrocław, 191-203.
- Walesiak M. (1988), *Skale pomiaru cech (w ujęciu zwężonym) a zagadnienie wyboru postaci analitycznej syntetycznych mierników rozwoju*, Prace Naukowe AE we Wrocławiu nr 447, AE, Wrocław, 63-71.
- Walesiak M. (1990a), *Porównawcza analiza miar zgodności wyników klasyfikacji w świetle postulowanych własności*, [w:] J. Pocięcha (red.), *Taksonomia – teoria i jej zastosowania*, AE, Kraków, 13-19.
- Walesiak M. (1990b), *Syntetyczne badania porównawcze w świetle teorii pomiaru*, „Przegląd Statystyczny”, z. 1-2, 37-46.
- Walesiak M. (1991), *O stosowalności miar korelacji w analizie wyników pomiaru porządkowego*, Prace Naukowe AE we Wrocławiu nr 600, AE, Wrocław, 13-19.
- Walesiak M. (1993a), *Statystyczna analiza wielowymiarowa w badaniach marketingowych*, Prace Naukowe AE we Wrocławiu nr 654, Seria: Monografie i Opracowania nr 101, AE, Wrocław.
- Walesiak M. (1993b), *Strategie postępowania w badaniach statystycznych w przypadku zbioru zmiennych mierzonych na skalach różnego typu*, „Badania Operacyjne i Decyzje”, nr 1, 71-77.
- Walesiak M. (1993c), *Zagadnienie oceny podobieństwa zbioru obiektów w czasie w syntetycznych badaniach porównawczych*, „Przegląd Statystyczny”, z. 1, 95-102.
- Walesiak M. (1995), *The analysis of factors influencing the choice of the methods in the statistical analysis of marketing data*, „Statistics in Transition”, June, vol. 2, no. 2, 185-194.
- Walesiak M. (1996), *Metody analizy danych marketingowych*, PWN, Warszawa.
- Walesiak M. (1999), *Distance measure for ordinal data*, „Argumenta Oeconomica”, no 2 (8), 167-173.
- Walesiak M. (2002a), *Propozycja uogólnionej miary odległości w statystycznej analizie wielowymiarowej*, [w:] J. Paradysz (red.), *Statystyka regionalna w służbie samorządu lokalnego i biznesu*, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej, AE, Poznań, 115-121.
- Walesiak M. (2002b), *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, AE, Wrocław.
- Walesiak M. (2002c), *Pomiar podobieństwa obiektów w świetle skal pomiaru i wag zmiennych*, Prace Naukowe AE we Wrocławiu nr 950, AE, Wrocław, 11-25.
- Walesiak M. (2003a), *Obszary zastosowań uogólnionej miary odległości GDM w statystycznej analizie wielowymiarowej*, Prace Naukowe AE we Wrocławiu nr 981, AE, Wrocław, 37-49.
- Walesiak M. (2003b), *Uogólniona miara odległości GDM jako syntetyczny miernik rozwoju w metodach porządkowania liniowego*, Taksonomia 10, Prace Naukowe AE we Wrocławiu nr 988, AE, Wrocław, 134-144.

- Walesiak M. (2003c), *Miara odległości obiektów opisanych zmiennymi mierzonymi na różnych skalach pomiaru*, Prace Naukowe AE we Wrocławiu nr 1006, AE, Wrocław, 261-267.
- Walesiak M. (2004a), *Problemy decyzyjne w procesie klasyfikacji zbioru obiektów*, Prace Naukowe AE we Wrocławiu nr 1010, AE, Wrocław, 52-71.
- Walesiak M. (2004b), *Uogólniona miara odległości dla obiektów opisanych zmiennymi z różnych skal pomiaru – oprogramowanie komputerowe*, Taksonomia 11, Prace Naukowe AE we Wrocławiu nr 1022, AE, Wrocław, 250-256.
- Walesiak M. (2004c), *Metody klasyfikacji*, [w:] E. Gatnar, M. Walesiak (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław, 316-350.
- Walesiak M. (2005a), *Rekomendacje w zakresie strategii postępowania w procesie klasyfikacji zbioru obiektów*, [w:] A. Zeliś (red.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, AE, Kraków, 185-203.
- Walesiak M. (2005b), *Uogólniona miara odległości GDM a współczynnik korelacji liniowej Pearsona i cosinus kąta między wektorami*, Prace Naukowe AE we Wrocławiu nr 1096, AE, Wrocław, 18-26.
- Walesiak M. (2005c), *Problemy selekcji i ważenia zmiennych w zagadnieniu klasyfikacji*, Taksonomia 12, Prace Naukowe AE we Wrocławiu nr 1076, AE, Wrocław, 106-118.
- Walesiak M. (2006), *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wydanie drugie rozszerzone, AE, Wrocław.
- Walesiak M. (2008), *Ocena stabilności wyników klasyfikacji z wykorzystaniem analizy replikacji*, [w:] Pocięcha J. (red.), *Modelowanie i prognozowanie zjawisk społeczno-gospodarczych*, UE, Kraków, 67-72.
- Walesiak M. (2009a), *Losowe generowanie danych o znanej strukturze klas w pakiecie clusterSim*, [w:] Pocięcha J. (red.), *Współczesne problemy modelowania i prognozowania zjawisk społeczno-gospodarczych*, Studia i Prace nr 2 Uniwersytetu Ekonomicznego w Krakowie, UE, Kraków, 391-399.
- Walesiak M. (2009b), *Analiza skupień*, [w:] M. Walesiak, E. Gatnar (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, PWN, Warszawa, 407-433.
- Walesiak M. (2011a), *Porządkowanie liniowe z wykorzystaniem uogólnionej miary odległości GDM2 dla danych porządkowych i programu R*, Prace Naukowe UE we Wrocławiu nr 163, UE, Wrocław, 9-18.
- Walesiak M. (2011b), *Odległość GDM2 w analizie skupień dla danych porządkowych z wykorzystaniem programu R*, Taksonomia 18, Prace Naukowe UE we Wrocławiu nr 176, UE, Wrocław, 40-52.
- Walesiak M. (2011c), *Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania*, Taksonomia 19, Prace Naukowe UE we Wrocławiu, UE, Wrocław (w druku).
- Walesiak M. (2011d), *Macierz danych i tablica danych*, [w:] E. Gatnar, M. Walesiak (red.), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, Wydawnictwo C.H. Beck, Warszawa.
- Walesiak M., Bąk A. (2000), *Conjoint analysis w badaniach marketingowych*, AE, Wrocław.
- Walesiak M., Bąk A., Jajuga K. (2002), *Uogólniona miara odległości – badania symulacyjne*, Taksonomia 9, Prace Naukowe AE we Wrocławiu nr 942, AE, Wrocław, 116-127.
- Walesiak M., Dudek A. (2008), *Identification of noisy variables for nonmetric and symbolic data in cluster analysis*, [w:] C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (eds.), *Data analysis, machine learning and applications*, Springer-Verlag, Berlin, Heidelberg, 85-92.
- Walesiak M., Dudek A. (2009a), *Ocena wybranych procedur analizy skupień dla danych porządkowych*, Taksonomia 16, Prace Naukowe UE we Wrocławiu nr 47, UE, Wrocław, 41-49.
- Walesiak M., Dudek A. (2009b), *Odległość GDM dla danych porządkowych a klasyfikacja spektralna*, Prace Naukowe UE we Wrocławiu nr 84, UE, Wrocław, 9-19.

- Walesiak M., Dudek A. (2010a), *Finding groups in ordinal data – an examination of some clustering procedures*, [w:] H. Locarek-Junge, C. Weihs (eds.), *Classification as a tool for research*, Heidelberg-Berlin, 185-192.
- Walesiak M., Dudek A. (2010b), *Klasyfikacja spektralna z wykorzystaniem odległości GDM*, *Taksonomia* 17, Prace Naukowe UE we Wrocławiu nr 107, 161-171.
- Walesiak M., Dudek A. (2011), *clusterSim package*, URL <http://www.R-project.org>.
- Walesiak M., Dziechciarz J., Bąk A. (1998), *Ordinal variables in the segmentation of advertisement receivers*, [w:] A. Rizzi, N. Vichi, H.H. Bock, *Advances in data science and classification*, Proceedings 6th Conference International Federation of Classification Societies in Rome, Springer, Heidelberg, 655-662.
- Wallace D.L. (1983), *Comment*, „Journal of the American Statistical Association”, vol. 78, no. 383, 569-576.
- Wedel M., Kamakura W.A. (1998), *Market segmentation. Conceptual and methodological foundations*, Kluwer Academic Publishers, Boston-Dordrecht-London.
- Wiśniewski J.W. (1986), *Korelacja i regresja w badaniach zjawisk jakościowych na tle teorii pomiaru*, „Przegląd Statystyczny”, z. 3, 239-248.
- Wiśniewski J.W. (1987), *Teoria pomiaru a teoria błędów w badaniach statystycznych*, „Wiadomości Statystyczne”, nr 11, 18-20.
- Zaborski A. (2001), *Skalowanie wielowymiarowe w badaniach marketingowych*, AE, Wrocław.
- Zakrzewska (1987), *O miarach podobieństwa obiektów i cech przydatnych w psychologicznych zastosowaniach analizy skupień*, [w:] J. Brzeziński (red.), *Wielozmienne modele statystyczne w badaniach psychologicznych*, rozdz. 7, PWN, Warszawa, Poznań.
- Zeliaś A. (1984), *Teoria prognozy*, PWE, Warszawa.
- Zeliaś A. (red.) (2000), *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, AE, Kraków.
- Zelnik-Manor L., Perona P. (2004), *Self-tuning spectral clustering*, [w:] Proceedings of the 18th Annual Conference on Neural Information Processing Systems (NIPS'04), <http://books.nips.cc/nips17.html>.
- Žukowska V.M., Mučnik J.G. (1976), *Faktornyj analiz v socialno-ekonomičeskich issledovanijach*, Statistika, Moskva.