DOI: 10.37317/biul-2008-0028

ZYGMUNT KACZMAREK [1]
STANISŁAW CZAJKA [2]
ELŻBIETA ADAMSKA [1]
[1] Instytut Genetyki Roślin PAN, Poznań
[2] Katedra Metod Matematycznych i Statystycznych, Uniwersytetu Przyrodniczego w Poznaniu

Propozycja metody grupowania obiektów jedno- i wielocechowych z zastosowaniem odległości Mahalanobisa i analizy skupień

A proposed method for grouping uni- and multivariate treatments using Mahalanobis distance and cluster analysis

Praca zawiera propozycję metody grupowania wielocechowych obiektów o rozkładach normalnych ze wspólną macierzą kowariancji. Metoda ta jest stosowana w analizie skupień i wykorzystuje odległości Mahalanobisa jako miarę podobieństwa dwóch obiektów. Grupowanie polega na łączeniu obiektów najbardziej do siebie podobnych, czyli takich między którymi odległość Mahalamnobisa jest najmniejsza. Jako kryterium zastopowania procesu grupowania, zarówno w przypadku obiektów wielocechowych jak i jednocechowych, przyjęto „najmniejszą istotną odległość" wyznaczoną każdorazowo dla porównywanych obiektów. Prezentowana metoda jest zilustrowana przykładem numerycznym grupowania linii DH rzepaku ozimego.

Słowa kluczowe: ANOVA, MANOVA, analiza skupień, grupowanie obiektów, najmniejsza istotna odległość, odległość Mahalanobisa

A method for grouping the multivariate normal treatments with common covariance matrix is proposed. The method is applicable in cluster analysis and utilizes the Mahalanobis distance as a of measure similarity of two treatments. It is based on the successive merging of two most similar samples taken from the treatments in the sense of minimal distance between them. The "least significant distance" is used as the criterion to determine a stopping rule for the grouping uni- and multivariate treatments. A numerical example from the winter oilseed rape data is included to illustrate the proposed method.

Key words: ANOVA, MANOVA, cluster analysis, grouping of treatments, least significant distance, Mahalanobis distance

WSTĘP

Analiza wariancji wyników doświadczenia, którego obiekty obserwowane są pod względem określonych zmiennych (cech) kończy się z reguły testowaniem hipotezy o równości jedno - lub wielocechowych średnich obiektowych. Odrzucenie tej hipotezy nie daje jeszcze żadnej informacji o tym, które z badanych obiektów różnią się między sobą istotnie, a między którymi różnice te nie są istotne. Testowanie hipotez ogólnych i implikowanych przez nie hipotez szczegółowych omawiane jest w pracach wielu autorów, między innymi przez Gabriela (1968), Calińskiego i Kaczmarka (1973), Calińskiego i wsp. (1976, 1979), Morrisona (1976), Mardię i wsp. (1979) oraz Sebera (1984). Jednymi z najczęściej stosowanych w praktyce metod porównań wielokrotnych w jednozmiennej analizie wariancji (ANOVA) są: test Duncana, Dunnetta, Newmana-Keulsa, Scheffego, test F, test t oraz test Tukeya. W wielozmiennej analizie wariancji (MANOVA) stosowane są różne testy do weryfikacji hipotez: pierwiastek maksymalny Roya, ślad Lawleya-Hottelinga, iloraz wiarogodności Wilksa oraz ślad Pillaia. Testy te wykorzystywane są w wielu programach i pakietach statystycznych. Znacznie rzadziej przedstawiane są i stosowane w programach obliczeniowych metody grupowania obiektów. O ile jeszcze wyróżnianie grup jednorodnych w przypadku jednej cechy można niekiedy znaleźć w niektórych programach komputerowych (Gabriel 1964; Harabasz i Wiśniewski 1984), to już procedury obliczeniowe i pakiety statystyczne dotyczące analizy doświadczeń wielocechowych rzadko uwzględniają potrzebę grupowania obiektów. Zawarte w nich metody wielozmienne takie jak analiza składowych głównych czy też analiza zmiennych kanonicznych (Caliński i in., 1975) umożliwiają co prawda znalezienie graficznych obrazów rozmieszczenia obiektów na płaszczyźnie, pozwalają nawet wyznaczyć pewne ich skupienia, jednakże nie dokonują formalnego i obiektywnego podziału obiektów na grupy w maksymalnym stopniu wewnętrznie jednorodne pod względem badanego zespołu cech. Również odległości Mahalanobisa a także odległości Euklidesa stanowią co najwyżej podstawę wykreślenia dendrogramu lub dendrytu najkrótszych połączeń mimo że uznawane są często za miarę wielocechowego podobieństwa obiektów (Caliński, Kaczmarek, 1969, 1973; Mądry, Kubicka, 1988).

W latach siedemdziesiątych i osiemdziesiątych zaproponowano co prawda wiele metod statystycznych i procedur obliczeniowych grupowania obiektów wielocechowych (Mc Queen, 1966; Lance i Williams, 1967; Caliński, 1969; Karoński, 1971; Caliński i Harabasz, 1974; Harabasz i Karoński, 1977; Chudzik i Karoński, 1979; Karoński i Caliński, 1979) lecz nie znalazły one częstego wykorzystania w praktyce.

W niniejszej pracy przedstawimy propozycję metody grupowania obiektów opartej na analizie skupień. W metodzie tej, grupującej obiekty o wielowymiarowych rozkładach normalnych ze wspólną macierzą kowariancji, wykorzystana będzie odległość Mahalanobisa jako miara podobieństwa między dwoma obiektami. Metoda może być stosowana zarówno dla obiektów jednocechowych, jak i wielocechowych. Dla obu sytuacji obowiązywać będzie także to samo kryterium podziału na rozłączne grupy jednorodne.

2. METODA

Weźmy pod uwagę doświadczenie, w którym każdy z k obiektów jest replikowany n_i razy ($i = 1,2,...,k$) i obserwowany pod względem p cech ($l = 1,2,...,p$). Zapiszmy wektory średnich z powtórzeń dla i-tego oraz j-tego obiektu ($i \neq j = 1,2,...,k$) w postaci:

$$\bar{y}_i = \begin{bmatrix} \bar{y}_{i1} \\ \bar{y}_{i2} \\ \vdots \\ \bar{y}_{ip} \end{bmatrix}, \bar{y}_j = \begin{bmatrix} \bar{y}_{j1} \\ \bar{y}_{j2} \\ \vdots \\ \bar{y}_{jp} \end{bmatrix}, \tag{1}$$

gdzie $\bar{y}_{il}$ ($\bar{y}_{jl}$) są średnimi i-tego (j-tego) obiektu dla l-tej cechy (l = 1,2,...,p).

2.1. Grupowanie obiektów wielocechowych (p > 1)

Jeśli jako miarę wielocechowego podobieństwa obiektu i-tego z j-tym przyjmiemy uogólnioną odległość Mahalanobisa (Caliński, Kaczmarek, 1969), to jej kwadrat można zapisać w postaci następującej formy kwadratowej

$$D_{ij}^2 = (\bar{y}_i - \bar{y}_j)'\mathbf{S}_\mathrm{e}^{-1}(\bar{y}_i - \bar{y}_j), \tag{2}$$

gdzie $\mathbf{S}_\mathrm{e}^{-1}$ jest odwrotnością macierzy średnich kwadratów i iloczynów dla błędu w wielozmiennej analizie wariancji a wektory $\bar{y}_i$ i $\bar{y}_j$ zostały zdefiniowane w (1).

Dla weryfikacji hipotezy orzekającej, że oba obiekty maja takie same p-cechowe wektory średnich, czyli, że odległość między nimi wynosi zero, można zastosować następującą „odległość krytyczną" (Morrison, 1976; Camussi i in., 1985)

$$D_\alpha^2 = \left(\frac{1}{n_i} + \frac{1}{n_j}\right)\frac{p\upsilon_e}{\upsilon_e - p - 1}F_{\alpha;p,\upsilon_e - p+1}, \tag{3}$$

gdzie n_i (n_j) są liczbami replikacji dla i-tego (j-tego) obiektu, p oznacza liczbę cech, υ_e jest liczbą stopni swobody dla macierzy $\mathbf{S}_\mathrm{e}$, a $F_{\alpha;p,\upsilon_e - p+1}$ jest wartością krytyczną odczytaną z tablic rozkładu F Fishera-Snedecora na poziomie istotności α dla p i ν-p+1 stopni swobody. Testowanie to można przeprowadzić przy założeniu istnienia wspólnej macierzy kowariancji oraz faktu, że łączny rozkład obserwowanych p zmiennych jest normalny. Proponowana metoda grupowania polegać będzie na łączeniu obiektów najbardziej podobnych „krok za krokiem" i każdorazowym porównywaniu kwadratu odległości Mahalanobisa, najkrótszej w aktualnym zbiorze obiektów, z odległością krytyczną D^2_α.

2.2. Grupowanie obiektów jednocechowych (p = 1)

W przypadku jednej l-tej cechy odległość Mahalanobisa zdefiniowaną w (2) można dla pary średnich obiektowych $\bar{y}_i$ oraz $\bar{y}_j$ zapisać następująco:

$$D_{ij}^2 = (\bar{y}_{il} - \bar{y}_{jl})(\mathrm{MS_e})^{-1}(\bar{y}_{il} - \bar{y}_{jl}) = \frac{(\bar{y}_{il} - \bar{y}_{jl})^2}{\mathrm{MS_e}}, \tag{4}$$

gdzie MS_e jest średnim kwadratem dla błędu wyznaczonym dla *l*-tej cechy w ANOVA. Odległość krytyczna ze wzoru (3) przyjmie teraz postać

$$D_{\alpha}^{2} = \left(\frac{1}{n_i} + \frac{1}{n_j}\right) F_{\alpha;1,\upsilon_e}. \qquad (5)$$

Jeżeli $D^2_{ij} > D^2_{\alpha}$ to odrzucamy na poziomie istotności α hipotezę, że odległość między *i*-tym oraz *j*-tym obiektem jest równa zero, czyli hipotezę mówiącą, że obiekty *i*-ty oraz *j*-ty nie różnią się pod względem średniej *l*-tej cechy. Zauważmy, że istnieje w tym przypadku pełna zgodność nierówności między D^2_{ij} i D^2_{α}, czyli nierówności

$$\frac{(\bar{y}_{il} - \bar{y}_{jl})^2}{\mathrm{MS_e}} > \left(\frac{1}{n_i} + \frac{1}{n_j}\right) F_{\alpha;1,\upsilon_e} \qquad (6)$$

z testem *F* dla weryfikacji hipotezy dotyczącej kontrastu między *i*-tym oraz *j*-tym obiektem pod względem *l*-tej cechy, którego ocena jest różnicą między średnimi $\bar{y}_{il}$ oraz $\bar{y}_{jl}$. Kontrast ten jest istotny na poziomie α jeśli

$$F = \frac{(\hat{k})^2}{(\frac{1}{n_i} + \frac{1}{n_j}) MS_e} > F_{\alpha;1,\upsilon_e}.$$

2.3. Przebieg procesu grupowania obiektów

Proponowana metoda analizy skupień *k* obiektów wielocechowych lub jednocechowych przyjmuje jako punkt wyjścia tablicę odległości Mahalanobisa wyznaczonych dla wszystkich możliwych par tych obiektów. Jej przebieg jest następujący:

— 1. Znajdujemy najmniejszą odległość Mahalanobisa, min D^2_{ij}, między dwoma obiektami i porównujemy ja z odległością krytyczną D^2_{α}.

— 2. Jeżeli:

$$\min D^2_{ij}, \le D^2_{\alpha}., \qquad (7)$$

dokonujemy połączenia tych obiektów w jeden obiekt łączny zawierający obserwacje obu połączonych obiektów i oraz *j*.

Realizacja metody w punktach 1–2 prowadzi do zmniejszenia liczby obiektów z *k* do *k*-1, czyli podziału *k* obiektów na *k*-1 grup.

— 3. Uaktualniamy tablicę odległości Mahalanobisa poprzez rezygnację z odległości obiektów i oraz *j* z pozostałymi i wprowadzenie wyliczonych odległości między nowym obiektem łącznym o liczbie replikacji $n_i + n_j$ a pozostałymi obiektami.

— 4. Wracamy do punktu (1) znajdując dla *k*-1 obiektów najmniejszą odległość Mahalanobisa i porównując ją z odległością krytyczną D^2_{α}.

— 5. Proces tworzenia grup jednorodnych trwa do momentu, w którym min $D^2 > D^2_{\alpha}$, czyli do momentu, w którym nierówność (7) nie zachodzi. Wystąpienie tej nierówności oznacza zakończenie procesu grupowania i przyjęcie oprócz dotychczas utworzonych grup obiektów, jednoelementowych grup zawierających obiekty do tego momentu niepogrupowane.

3. PRZYKŁAD I DYSKUSJA

W celu zobrazowania przebiegu grupowania według proponowanej metody wzięto fragment wyników doświadczenia linia × tester z rzepakiem ozimym z 7 liniami DH, 4 testerami i 28 mieszańcami F_1 przeprowadzonego w układzie losowanych bloków w 3 powtórzeniach na Polu Doświadczalnym w Cerekwicy k/Poznania. Wybrany fragment danych dotyczy 7 linii DH obserwowanych pod względem 4 cech: długości łuszczyny, liczby nasion w łuszczynie, masy tysiąca nasion i zawartości tłuszczu w oleju nasion. Średnie wartości tych cech dla poszczególnych linii zawiera tabela 1.

Tabela 1

Średnie wartości linii DH rzepaku ozimego dla poszczególnych cech
Mean values of DH lines of winter rape for individual traits

Linia DH DH line (i =)	Długość łuszczyny Length of seeds silique (cm)	Liczba nasion w łuszczynie Number of per silique	MTN TSW (g)	Zawartość tłuszczu Oil content (%)
1	9,2	25,1	4,7	44,3
2	7,7	19,1	5,1	44,6
3	7,7	25,6	4,6	44,0
4	7,2	17,3	5,3	44,0
5	7,6	19,7	4,6	45,5
6	9,8	27,0	4,7	43,9
7	9,0	22,9	4,6	43,9

3.1 Grupowanie obiektów wielocechowych

W wyniku przeprowadzonej wielozmiennej analizy wariancji odrzucona została hipoteza ogólna o braku różnic między wektorami średnich poszczególnych linii DH. Obliczone, ze względu na wyżej wymienione 4 cechy, kwadraty odległości Mahalanobisa przedstawia tabela 2. Odległości te stanowią pierwszy etap obliczeń w procesie grupowania obiektów.

Tabela 2

Odległości Mahalanobisa między liniami DH
Mahalanobis distances between DH lines

	Odległości Mahalanobisa — Mahalanobis distances (i =)					
(j =)	1	2	3	4	5	6
2	15,3					
3	33,2	36,2				
4	22,4	**1,3**	38,2			
5	16,7	3,1	27,5	6,4		
6	1,4	25,0	40,8	33,4	27,7	
7	1,4	19,0	41,2	25,9	21,6	2,1

Ponieważ min D^2_{ij} = 1,3 (dla pary obiektów 2 i 4) < $D^2_{0.05}$ = 11,39 zatem obiekty 2 i 4 tworzą jeden obiekt łączny (o numerze 2) o średnich wartościach cech: 7,5, 18,2, 5,2, 44,3 i liczbie replikacji r = 6. W etapie pierwszym wyróżniono grupę obejmującą obiekty 2 i 4. Obiekty niepogrupowane: 1, 3, 5, 6. 7

Etap 2. Obliczamy ponownie odległości Mahalanobisa między obiektami uwzględniając fakt dołączenia obiektu nr 4 do obiektu 2.

Odległości Mahalanobisa — Mahalanobis distances						
			(i =)			
(j =)	1	2	3	4	5	6
2	15,3					
3	33,2	36,8				
4	-		-			
5	16,7	4,4	27,5	-		
6	1,4	28,9	40,8	-	27,7	
7	**1,4**	22,2	41,2	-	21,6	2,1

Ponieważ min D^2_{ij} = 1.4 (dla par obiektów 1, 7) $<D^2_{0,05}$ = 11,39 zatem obiekty 1 i 7 tworzą jeden obiekt łączny (o numerze 1) o średnich wartościach cech: 9,1, 24,0 4,6, 44,1 i liczbie replikacji r = 6.Oznacza to, że obliczenia w etapie drugim pozwoliły wyróżnić grupy:

— grupa 1: obiekty 2, 4
— grupa 2: obiekty 1, 7

Obiekty niepogrupowane: 3, 5, 6.

Etap 3. Obliczamy odległości Mahalanobisa między dwoma obiektami łącznymi (grupami 1 i 2) oraz obiektami 3, 5 i 6.

Odległości Mahalanobisa — Mahalanobis distances						
			(i =)			
(j =)	1	2	3	4	5	6
2	20,0					
3	36,9	36,8				
4	-	-	-	-	-	
5	18,8	4,4	27,5	-		
6	**1,4**	28,9	40,8	-	27,7	
7	-	-	-	-	-	-

Ponieważ min D^2_{ij} = 1,4 (dla par obiektów 1 i 6) $<D^2_{0,05}$ = 8,54 zatem obiekty 1, 7 i 6 tworzą jeden obiekt łączny (grupę o numerze 1) o średnich wartościach cech: 9,3, 25,0, 4,6, 44,0 i liczbie replikacji r = 9. Obliczenia w etapie 3 pozwoliły wyróżnić grupy:

— grupa 1: obiekty 2, 4
— grupa 2: obiekty 1, 7, 6

Obiekty niepogrupowane: 3, 5.

Etap 4. Obliczamy odległości Mahalanobisa między dwoma obiektami łącznymi (grupami 1 i 2) oraz obiektami 3 i 5.

Odległości Mahalanobisa — Mahalanobis distances						
			(i =)			
(j =)	1	2	3	4	5	6
2	22,6					
3	37,9	36,8				
4	-	-	-			
5	21,4	**4,4**	27,5-			
6	-	-	-	-	-	
7	-	-	-			-

Ponieważ min D^2_{ij} = 4,4 (dla par obiektów 2 i 5) <$D^2_{0,05}$ = 8,54 zatem obiekt nr 5 został dołączony do grupy 2 utworzonej z obiektów 2 i 4, której aktualne średnie wartości cech są następujące: 7,5, 18,7, 5,0, 44,7 przy liczbie replikacji r = 9. W wyniku obliczeń wykonanych w etapie 4 utworzono następujące grupy obiektów:

— grupa 1: obiekty 2, 4, 5
— grupa 2: obiekty 1, 7, 6

Obiekt niepogrupowany: obiekt 3

Etap 5. Obliczamy odległości Mahalanobisa między dwoma obiektami łącznymi (grupami 1 i 2) oraz obiektem nr 3.

Odległości Mahalanobisa — Mahalanobis distances						
			(i =)			
(j =)	1	2	3	4	5	6
2	22,6					
3	37,9	32,7				
4	-	-	-			
5	-	-	-	-		
6	-	-	-	-	-	
7	-	-	-	-	-	-

Ponieważ min D^2_{ij} = 22,6 (dla par obiektów 1 i 2) >$D^2_{0,05}$ = 3,80 zatem obiekt 3 tworzy rozłączną grupę i proces grupowania zostaje zakończony. W rezultacie obowiązujący jest wynik grupowania polegający na utworzeniu 3 grup jednorodnych o następującym składzie obiektów:

— grupa 1: obiekty 2, 4, 5
— grupa 2: obiekty 1, 7 i 6,
— grupa 3:obiekt 3.

3.2 Grupowanie obiektów dla poszczególnych cech

<u>Długość łuszczyny</u>

Test F w analizie wariancji: F = 8.635> $F_{0.05}$ = 2.85

Hipoteza H_0 została odrzucona

Skład grup jednorodnych	Średnia
1, 7. 6	9,3
2, 3, 5, 4	7,5

Liczba nasion

Test F w analizie wariancji: F = 5,172 > $F_{0.05}$ = 2,85

Hipoteza H_0 została odrzucona

Skład grup jednorodnych	Średnia
1, 3. 6,7	25,2
2, 5, 4	18,7

Masa tysiąca nasion

Test F w analizie wariancji:

F = 2.276 < $F_{0.05}$ = 2.85

Hipoteza H_0 o braku różnic między średnimi nie została odrzucona

Zawartość tłuszczu

Test F w analizie wariancji:

F = 1.708 < $F_{0.05}$ = 2.85

Hipoteza H_0 o braku różnic między średnimi nie została odrzucona

Identyczne podziały obiektów na grupy jednorodne uzyskano dla każdej cechy stosując metodę Gabriela (1964) wykorzystującą procedurę testów jednoczesnych.

WNIOSKI

1. Wyróżnienie grup jednorodnych za pomocą opisanej w pracy metody grupowania może stać się dodatkową informacją i uzupełnieniem wyników testowania istotności określonych kontrastów dotyczących par obiektów i grup obiektów.
2. Szczególnym przypadkiem proponowanej metody grupowania jest możliwość wykorzystania jako miary wielocechowego podobieństwa obiektów -odległość Euklidesa zamiast odległości Mahalanobisa. Proces grupowania przebiega w sposób analogiczny do opisanego wyżej, jednakże bez możliwości testowania odległości D_{ij}^2.

   Jest to równoznaczne z brakiem kryterium, które pozwoliłoby zatrzymać proces łączenia obiektów w grupy.
3. Przy łączeniu obiektów w proponowanej metodzie można stosować różne metody analizy skupień, takie jak metoda najbliższego sąsiedztwa, metoda mediany, metoda średniej grupowej, metoda środka ciężkości, czy tez metoda Warda (Lance, Williams, 1967; Wishart, 1969; Karoński, Caliński, 1973).
4. Dla oszacowanych odległości Mahalanobisa mogą być wykorzystane wartości krytyczne przy stosowaniu jednoczesnej procedury testowej oparte na teście Rao lub teście Roya.
5. Zastosowanie jako kryterium grupowania tej samej funkcji testowej dla obiektów wielocechowych (p > 1) i jednocechowych (p = 1) daje możliwość bardziej

obiektywnego porównania wyników uzyskanych dla wyżej wymienionych przypadków.

LITERATURA

Caliński T. 1964. On the application of cluster analysis to experimental results W: proceedings of the 37th Session of the International Statistical Institute ISI, London.
Caliński T., Czajka S., Kaczmarek Z. 1975. Analiza składowych głównych i jej zastosowania. Roczniki AR w Poznaniu, Algorytmy biometryczne i statystyczne, 36: 159 — 185.
Caliński T., Dyczkowski A., Kaczmarek Z. 1976. Testowanie hipotez w wielozmiennej analizie wariancji i kowariancji. Roczniki AR w Poznaniu, Algorytmy biometryczne i statystyczne 45: 77 — 94.
Caliński T., Dyczkowski A I Sitek M.(1979). Procedury testów jednoczesnych w wielozmiennej analizie wariancji. Matematyka stosowana XIV: 5 — 31.
Caliński T., Harabasz J. S. 1974. A dendrite method for cluster analysis. Communications in Statistics 3.
Caliński T., Kaczmarek Z.1969. A note the calculation and use of the generalized distance between multivariate samples. Zeszyty Naukowe UAM, Geografia 8: 7 —17.
Caliński T., Kaczmarek Z. 1973. Metody kompleksowej analizy doświadczenia wielocechowego W: Trzecie Colloquim Metodologiczne z Agrobiometrii, PAN, PTB Warszawa, 258 —320.
Caliński T., Karoński M. 1977. Grupowanie populacji o wielorozkładowych rozkładach normalnych za pomocą procedury testów jednoczesnych. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne 56: 123 — 134.
Caliński T., Wagner W. 1974. Grupowanie średnich obiektowych w jednozmiennej analizie wariancji. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne 24: 61 — 73.
Camussi A., Ottaviano E., Caliński T., Kaczmarek Z. 1985. Genetic distances based on quantitative traits. Genetics 111: 945 —962.
Chudzik H., Karoński M.1979. Skupianie obserwacji metodą k-średnich. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne 78: 133 – 152.
Gabriel K. R. 1964. A procedure for testing the homogeneity of all sets of means in analysis of variance. Biometrics, vol. 20, 3: 459 —477.
Gabriel K. R. 1968. Simultaneous test procedures in multivariate analysis of variance. Biometrika 55: 489 — 504.
Górczyński J., Mądry W. 1988. A study of genetic divergence of plants by multivariate methods. Genetica Polonica, vol. 29, No 3–4: 341 —352.
Harabasz J. S., Karoński M. 1977. Dendrytowa metoda analizy skupień. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne, 57: 135 — 148.
Harabasz J. S., Wiśniewski P. 1984. Grupowanie obiektów jednocechowych za pomocą programowania dynamicznego. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne 109: 147 — 154.
Karoński M. 1971. Algorytm grupowania populacji w rozkładach metodą krok po kroku.. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne 4: 30 — 33.
Karoński M., Caliński T. 1973. Grupowanie populacji o rozkładach normalnych na podstawie odległości Mahalanobisa. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne, 16, 107 — 115.
Karoński M., Caliński T. 1973. Grupowanie obiektów wielocechowych na podstawie odległości euklidesowych. Roczniki AR w Poznaniu. Algorytmy biometryczne i statystyczne 17: 117 — 129.
Lance G. M., Williams W. T. 1967. A general theory of classificatory sorting strategies. Hierarchical systems, Computer J. 9: 373 —380.
Mac Queen J. B. 1967. Some methods for classification and analysis of multivariate observations. Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability Theory. Berkeley University of California Press. vol. 1, 281 — 287.
Mardia K. V., Kent J. T., Bibby J. M. 1979. Multivariate Analysis. Academic Press, London.
Morrison D. F. 1976. Multivariate statistical methods. McGraw-Hill, New York.

Mądry W., Kubicka H. 1988. Wielocechowa ocena zróżnicowania linii wsobnych wyselekcjonowanych z odmian uprawnych żyta ozimego (*S. cereale* L.). Hod. Rośl. Aklim. i Nasien. 32, 3/4: 16 — 26.
Rohlf F. J. 1970. Adaptive hierarchical clustering schemes. Syst. Zool. 19: 58 — 82.
Rajfura A., Mądry W. 2000. Wydzielenie grup jednocechowo podrzędnych w środowiskach w stosunku do genotypów wykazujących interakcję jakościową. Biul. IHAR 216: 27 — 37.
Seber G. A. F. 1984. Multivariate Observations. Wiley, New York.
Wishart D. 1969. An algorithm for hierarchical classifications. Biometrics 25, 1: 165 — 170.