DOI: 10.37317/biul-2011-0055

ZYGMUNT KACZMAREK [1]
DARIUSZ R. MAŃKOWSKI [2]
[1] Instytut Genetyki Roślin, Polskiej Akademii Nauk w Poznaniu
Pracownia Biometrii
[2] Instytut Hodowli i Aklimatyzacji Roślin — Państwowy Instytut Badawczy w Radzikowie
Pracownia Ekonomiki Nasiennictwa i Hodowli Roślin

Wprowadzenie do statystycznych analiz wielozmiennych∗ Część II. Przykład zastosowania

An introduction to multivariate statistical analyses Part II. The application

Prowadząc doświadczenia oraz analizując dane pochodzące z doświadczeń często obserwuje się i analizuje wiele cech charakteryzujących pewne obiekty doświadczalne. Często każda z tych cech analizowana jest osobno. Jednak, by mieć pełen obraz wyłaniający się z prowadzonych badań, należy sięgnąć do analizy wielocechowej pozwalającej na całościowe podejście do badanego problemu. W pracy przedstawiono praktyczny przykład obliczeniowy opisanych w części pierwszej metod wielocechowych analiz statystycznych. W szczególności przedstawiono wielozmienną analizę wariancji oraz powiązane z nią: analizę kontrastów oraz analizę zmiennych kanonicznych.

Słowa kluczowe: analizy wielocechowe, grupowanie wielocechowe obiektów, jęczmień jary, macierz korelacji, macierz kowariancji, MANOVA, wielocechowe mary podobieństwa

While conducting research and analysing experimental data, one often observes and analyses multiple characteristics of certain experimental objects. Often each of these characteristics is analyzed separately. However, to have a complete picture emerging from the research you should use multivariate analysis which allowing a holistic approach to the investigation of the problem. The paper presents a practical example of the calculation methods of multivariate statistical analysis formally described in the first part. In particular, this paper presents the MANOVA analysis and an analysis of contrasts and canonical variables analysis associated with it.

Key words: multivariate analysis, multivariate grouping, spring barley, correlation matrix, covariance matrix, MANOVA, multivariate similarity measures

∗ Praca była prezentowana w ramach I Warsztatów Biometrycznych, które odbyły się w IHAR-PIB w Radzikowie w dniach 14-15 września 2010 r.

WSTĘP

W pierwszej części zaprezentowano teoretyczne podstawy analizy eksperymentalnych danych wielocechowych za pomocą wielozmiennej analizy wariancji (MANOVA). W części drugiej omówiony zostanie przykład przeprowadzenia takich analiz i zasady wnioskowania na podstawie uzyskanych wyników.

Opisywany przykład rozpocznie charakterystyka doświadczenia i danych, następnie przeprowadzona zostanie wielozmienne analiza wariancji, analiza zmiennych kanonicznych oraz analiza podobieństwa pomiędzy badanymi obiektami.

W przygotowaniu niniejszego opracowania skorzystano z następujących publikacji: Caliński (1970), Anderberg (1973), Caliński i Kaczmarek (1973), Kaczmarek (1975), Morrison (1976), Seber (1984), Krzysko (2000), Timm (2002), Kaczmarek i in. (2008).

Omawiane analizy wykonano za pomocą programów: ABS-35 (Ceranka i in., 1975), ABS-36 (Caliński i in., 1975) i ABS-45 (Caliński i in., 1976) oraz w Systemie SAS® w wersji 9.2 (SAS Institute Inc., 2009).

1. OPIS DOŚWIADCZENIA WIELOCECHOWEGO

Analizowane dane wielocechowe pochodzą z doświadczenia polowego z jęczmieniem jarym wykonanego w Instytucie Genetyki Roślin PAN w Poznaniu. W doświadczeniu porównywano 26 linii jęczmienia jarego (tab. 1), spośród których 12 było liniami nieoplewionymi (nagimi) [N1–N12], 12 było liniami oplewionymi [P1–P12], a dwie formy [P] i [N] były formami rodzicielskimi, pierwsza oplewiona a druga nieoplewiona. Dla każdego obiektu wykonano trzy powtórzenia. Doświadczenie było założone w układzie losowanych bloków.

W doświadczeniu prowadzono obserwacje czterech cech: średnicy źdźbła, współczynnika elastyczności, długości kłosa i masy tysiąca ziaren (MTZ). Wartości średnie dla tych cech przedstawiono w tabeli 2.

W celu scharakteryzowania zmienności i relacji pomiędzy analizowanymi zmiennymi wyznaczono macierz kowariancji (tab. 3) oraz macierz korelacji (tab. 4).

W macierzy kowariancji, na przekątnej znajdują się wariancje dla poszczególnych zmiennych, a poza przekątną, na przecięciu dwóch różnych zmiennych — kowariancje dla tych zmiennych. W wyniku standaryzacji macierzy kowariancji uzyskano macierz korelacji, w której na głównej przekątnej znajdują się same jedynki, natomiast na przecięciu dwóch zmiennych współczynniki korelacji liniowej Pearsona pomiędzy tymi zmiennymi.

Analiza współczynników korelacji liniowej (tab. 5) wskazuje na występowanie silnej, wysoce istotnej i wprost proporcjonalnej współzależności pomiędzy średnicą źdźbła i współczynnikiem elastyczności (r = 0,73). Ponadto stwierdzono występowanie nieco słabszych i odwrotnie proporcjonalnych współzależności pomiędzy długością kłosa oraz średnicą źdźbła (r = -0,44) oraz długością kłosa i współczynnikiem elastyczności (r = -0,35).

Tabela 1

Wartości obserwowanych cech (zmiennych) u form oplewionych i nieoplewionych jęczmienia jarego badanych w doświadczeniu polowum założonym w układzie losowanych bloków
Values of observed traits (variables) from the experiment carried out in randomized block design with the husked and naked forms of spring barley

Obiekt Object	Średnica źdźbła Stalk diameter			Współczynnik elastyczności Coefficient of elasticity			Długość kłosa Spike length			MTZ Mass of 1000 grains		
	I	II	III	I	II	III	I	II	III	I	II	III
N1	2,52	2,84	2,45	22,03	24,81	22,91	7,7	7,1	7,7	54,55	50,51	47,06
N2	2,39	2,5	2,4	25,65	26,71	24,06	8,3	7,6	7,7	43,56	50,23	52,17
N3	2,45	2,59	2,5	25,48	24,72	26,01	8,7	9,9	8,7	46,61	58,11	48,64
N4	2,41	2,56	2,4	23,03	22,01	24,08	7,6	8,3	8,5	47,28	48,58	49,52
N5	2,45	2,61	2,89	25,01	26,12	26,01	9,8	8,8	10	52,34	44,3	47,53
N6	2,4	2,68	2,41	26,01	26,22	22,32	9	7,3	9	53,61	48,04	55,16
N7	2,31	2,48	2,42	22,41	23,46	21,01	9,5	8,6	7,8	52,5	52,97	48,92
N8	2,65	2,71	2,6	25,01	25,74	25,94	8,5	9,2	8,6	53,93	50,79	54,55
N9	2,55	2,71	2,4	25,8	25,89	23,04	7,6	9,5	9,7	52,71	52,34	54,01
N10	2,34	2,46	2,67	26,89	27,01	27,01	8,1	9,5	8,3	52,61	55,56	51,89
N11	2,51	2,6	2,48	24,81	25	23,01	8,2	7,4	8,1	64,13	57,31	59,91
N12	2,45	2,4	2,81	23,2	24,73	25,06	6,6	8,2	7	50,53	54,03	52,17
P1	2,89	3,01	3,05	31,73	32,03	28,15	8	6,9	7,2	51,06	50	55,5
P2	2,75	3,01	2,74	28,01	27,21	27,4	6,6	5,8	6,1	53,93	54,9	51,74
P3	3,01	3,2	3,05	28,76	30,11	27,91	7,2	7,7	7,3	54,27	55,85	58,64
P4	2,81	3,01	2,9	29,03	29,01	27,98	7,3	7,2	6,7	51,46	53,9	49,46
P5	2,85	2,69	2,81	25,91	26,17	24,46	8,9	8,1	7,6	57,28	57	55,4
P6	3,01	3,1	2,81	32,01	29,88	27,82	7	7	7,1	51,58	56,7	55,26
P7	2,85	3,11	3,06	30,11	30,24	31,42	8,7	6,1	7,5	57,27	50	53,85
P8	2,73	3,11	2,9	25,81	26,89	27,01	7	6,9	6,7	55,62	52,12	56,32
P9	2,8	3,01	2,81	32,01	31,26	28,32	7,1	6,5	6,6	56,25	55,41	50,62
P10	2,91	3,02	2,82	32,01	31,21	27,01	8,9	7,6	8,6	58,72	53,65	60,38
P11	2,6	2,7	2,92	29,71	28,41	29,16	6,4	7	6,9	54,94	58,19	60,22
P12	2,8	3,01	2,99	28,97	32,49	27,81	6,7	9,4	6,2	53,3	49,4	49,46
P	2,8	2,62	3,11	31,23	31	30,96	6,9	6,8	8,3	45,5	49,47	48,4
N	2,69	2,45	2,7	26,76	26,03	26,05	7,8	8,5	8,3	45,11	53,65	50,6

Tabela 2

Wartości średnie analizowanych cech dla form jęczmienia jarego badanych w wielocechowym doświadczeniu polowym
Mean values of analyzed traits for the forms of spring barley tested in the multivariate experiment

Obiekty Objects	Średnica źdźbła Stalk diameter	Współczynnik elastyczności Coefficient of elasticity	Długość kłosa Spike length	MTZ Mass of 1000 grains
1	2	3	4	5
Bloki — Blocks				
I	2,65	27,21	7,85	52,72
II	2,77	27,47	7,80	52,81
III	2,73	26,23	7,77	52,97
Formy nieoplewione — Naked forms				
N1	2,60	23,25	7,50	50,70
N2	2,43	25,47	7,87	48,65
N3	2,51	25,40	9,10	51,12
N4	2,45	23,04	8,13	48,46
N5	2,65	25,71	9,53	48,05
N6	2,49	24,85	8,43	52,27
N7	2,40	22,29	8,63	51,46

c. d. Tabela 2

1	2	3	4	5
N8	2,65	25,56	8,76	53,09
N9	2,55	24,91	8,93	53,02
N10	2,49	26,97	8,63	53,35
N11	2,53	24,27	7,90	60,45
N12	2,55	24,33	7,27	52,24
N	2,61	26,28	8,20	49,79
Formy oplewione — Husked forms				
P1	2,98	30,64	7,37	52,19
P2	2,83	27,54	6,17	53,52
P3	3,08	28,93	7,40	56,25
P4	2,91	28,67	7,07	51,61
P5	2,78	25,51	8,20	56,56
P6	2,97	29,90	7,03	54,51
P7	3,01	30,59	7,43	53,71
P8	2,91	26,57	6,87	54,69
P9	2,87	30,53	6,73	54,09
P10	2,92	30,08	8,37	57,58
P11	2,74	29,09	6,77	57,78
P12	2,93	29,76	7,43	50,72
P	2,84	31,06	7,33	47,79

Tabela 3

Macierz kowariancji dla analizowanych zmiennych uzyskanych z doświadczenia polowego z formami nagimi i oplewionymi jęczmienia jarego
Covariance matrix for the analyzed variables obtained from the field experiment with the naked and husked forms of spring barley

	Średnica źdźbła Stalk diameter	Współczynnik elastyczności Coefficient of elasticity	Długość kłosa Spike length	MTZ Mass of 1000 grains
Średnica źdźbła Stalk diameter	0,055466	0,489400	-0,105200	0,146238
Współczynnik elastyczności Coefficient of elasticity	0,489400	8,084765	-1,012796	0,780378
Długość kłosa Spike length	-0,105200	-1,012796	1,015536	-0,027680
MTZ Mass of 1000 grains	0,146238	0,780378	-0,027680	15,692180

Tabela 4

Macierz korelacji dla analizowanych zmiennych uzyskanych z doświadczenia polowego z formami nagimi i oplewionymi jęczmienia jarego
Correlation matrix for the analyzed variables obtained from the field experiment with the naked and husked forms of spring barley

	Średnica źdźbła Stalk diameter	Współczynnik elastyczności Coefficient of elasticity	Długość kłosa Spike length	MTZ Mass of 1000 grains
Średnica źdźbła Stalk diameter	1,0000	0,7308**	-0,4433**	0,1568 ^{NS}
Współczynnik elastyczności Coefficient of elasticity	0,7308**	1,0000	-0,3535**	0,0693 ^{NS}
Długość kłosa Spike length	-0,4433**	-0,3535**	1,0000	-0,0069 ^{NS}
MTZ Mass of 1000 grains	0,1568 ^{NS}	0,0693^{NS}	-0,0069 ^{NS}	1,0000

** — istotne przy α=0,01; NS — nie istotne statystycznie; ** — significant at α=0.01; NS — not significant

Tabela 5

Macierze sum kwadratów i iloczynów dla badanych cech uzyskanych z doświadczenia polowego z formami nagimi i oplewionymi jęczmienia jarego
Sum of squares and products matrices for the analyzed variables obtained from the field experiment with the naked and husked forms of spring barley

	Średnica źdźbła Stalk diameter	Współczynnik elastyczności Coefficient of elasticity	Długość kłosa Spike length	MTZ Mass of 1000 grains
Dla bloków — For blocks				
Średnica źdźbła Stalk diameter	0,21185384615E+00			
Współczynnik elastyczności Coefficient of elasticity	0,36050000000E–01	0,22431479487E+02		
Długość kłosa Spike length	-0,93230769231E–01	0,80432051282E+00	0,71025641024E–01	
MTZ Mass of 1000 grains	0,22480000000E+00	-0,38041884615E+01	-0,23946153847E+00	0,89691538463E+00
Dla obiektów — For objects				
Średnica źdźbła Stalk diameter	0,32631538462E+01			
Współczynnik elastyczności Coefficient of elasticity	0,35122858974E+02	0,53612089872E+03		
Długość kłosa Spike length	-0,75936153846E+01	-0,84242769231E+02	0,52638461538E+02	
MTZ Mass of 1000 grains	0,15075302564E+02	0,11334309487E+03	-0,48464076923E+02	0,75191951282E+03
Dla błędu — For error				
Średnica źdźbła Stalk diameter	0,85134615385E+00			
Współczynnik elastyczności Coefficient of elasticity	0,30142833333E+01	0,72059320513E+02		
Długość kłosa Spike length	-0,51876923077E+00	0,44403461538E+01	0,26502307692E+02	
MTZ Mass of 1000 grains	-0,38935333333E+01	-0,48669444872E+02	0,46544461538E+02	0,47117361795E+03

Zapis $x\mathrm{E}\pm n$ należy rozumieć jako $x \cdot 10^{\pm n}$; form $x\mathrm{E}\pm n$ should be understood as $x \cdot 10^{\pm n}$

2. WIELOZMIENNA ANALIZA WARIANCJI — MANOVA

Model liniowy wielozmiennej jednoczynnikowej analizy wariancji w układzie bloków losowanych dla $y_{ij}^{(r)}$, czyli obserwacji *i*-tego poziomu czynnika A ($i = 1,2,...,a$) oraz *j*-tego bloku ($j = 1,2,...,b$) dotyczącej *r*-tej cechy ($r = 1,2,...,p$), można zapisać analogicznie jak w (3.1) z pierwszej części pracy (Kaczmarek i Mańkowski, 2011):

$$y_{ij}^{(r)} = \mu^{(r)} + \gamma_j^{(r)} + \alpha_i^{(r)} + \varepsilon_{ij}^{(r)}$$

gdzie $\mu^{(r)}$jest średnią ogólną dla *r*-tej cechy,$\alpha_i^{(r)}$jest efektem *i*-tego poziomu czynnika A dla *r*-tej cechy,$\gamma_j^{(r)}$ jest efektem *j*-tego bloku dla *r*-tej cechy,$\varepsilon_{ij}^{(r)}$ są błędami eksperymentalnymi dla *r*-tej cechy.

Pierwszym etapem jednocechowej analizy wariancji jest wyznaczenie sum kwadratów odchyleń dla wszystkich źródeł zmienności analizowanych w doświadczeniu i występujących w modelu liniowym. W przypadku wielozmiennej analizy wariancji wyznacza się macierze sum kwadratów i iloczynów dla wszystkich elementów modelu analizy wariancji, czyli dla bloków, obiektów i błędu losowego. Macierze te przedstawiono w postaci półlogarytmicznej w tabeli 5.

Na przekątnej macierzy sum kwadratów i iloczynów (tab. 6) znajdują się sumy kwadratów odchyleń dla poszczególnych cech. Wartości poza głównymi przekątnymi tych macierzy to wartości kowariancji pomiędzy tymi cechami. Biorąc z macierzy sum kwadratów i iloczynów do analizy tylko wartości sum kwadratów odchyleń dla poszczególnych cech, uzyskuje się klasyczny model jednocechowej analizy wariancji.

Tabela 6

Porównania szczegółowe — kontrasty
Contrasts — detailed comparison

N.k.	Porównanie Comparison	Statystyka F dla kontrastu wielocechowego F-statistic for multivariate contrast	Średnica źdźbła Stalk diameter		Współczynnik elastyczności Coefficient of elasticity		Długość kłosa Spike length		MTZ Mass of 1000 grains	
			ocena kontrastu contrast estimate	statystyka F F-statistic	ocena kontrastu contrast estimate	statystyka F F-statistic	ocena kontrastu contrast estimate	statystyka F F-statistic	ocena kontrastu contrast estimate	statystyka F F-statistic
1	2	3	4	5	6	7	8	9	10	11
Porównanie linii nagich z formą rodzicielską N — Comparison of the naked lines with a parental form N										
1	N1-N	2,69*	-0,01	0,01 ^{NS}	-3,03	9,56**	-0,70	1,39 ^{NS}	0,92	0,13 ^{NS}
2	N2-N	0,91^{NS}	-0,18	2,96 ^{NS}	-0,81	0,68 ^{NS}	-0,33	0,31 ^{NS}	-1,13	0,20 ^{NS}
3	N3-N	0,90^{NS}	-0,10	0,88 ^{NS}	-0,88	0,80 ^{NS}	0,90	2,29 ^{NS}	1,33	0,28 ^{NS}
4	N4-N	3,32*	-0,16	2,16 ^{NS}	-3,24	10,93**	-0,07	0,01 ^{NS}	-1,33	0,28 ^{NS}
5	N5-N	2,67*	0,037	0,12 ^{NS}	-0,57	0,33 ^{NS}	1,33	5,03*	-1,73	0,48 ^{NS}
6	N6-N	0,67^{NS}	-0,12	1,20 ^{NS}	-1,43	2,13 ^{NS}	0,23	0,15 ^{NS}	2,48	0,98 ^{NS}
7	N7-N	4,49**	-0,21	3,89 ^{NS}	-3,99	16,54**	0,43	0,53 ^{NS}	1,67	0,45 ^{NS}
8	N8-N	0,72^{NS}	0,04	0,14 ^{NS}	-0,72	0,53 ^{NS}	0,57	0,91 ^{NS}	3,30	1,74 ^{NS}
9	N9-N	0,95^{NS}	-0,06	0,32 ^{NS}	-1,37	1,95 ^{NS}	0,73	1,52 ^{NS}	3,23	1,66 ^{NS}
10	N10-N	1,24^{NS}	-0,12	1,34 ^{NS}	0,69	0,50 ^{NS}	0,43	0,53 ^{NS}	3,57	2,02 ^{NS}
11	N11-N	5,76**	-0,08	0,61 ^{NS}	-2,01	4,19*	-0,30	0,25 ^{NS}	10,66	18,10**
12	N12-N	1,17^{NS}	-0,06	0,32 ^{NS}	-1,95	3,96 ^{NS}	-0,93	2,47 ^{NS}	2,46	0,96 ^{NS}
Porównanie linii oplewionych z formą rodzicielską P — Comparison of the husked lines with a parental form P										
13	P1-P	1,55 ^{NS}	0,14	1,73 ^{NS}	-0,43	0,19 ^{NS}	0,03	<0,01 ^{NS}	4,40	3,08 ^{NS}
14	P2-P	5,36**	-0,01	0,01 ^{NS}	-3,52	12,92**	-1,17	3,85 ^{NS}	5,73	5,23*
15	P3-P	6,43**	0,24	5,22*	-2,14	4,75*	0,07	0,01 ^{NS}	8,46	11,40**
16	P4-P	2,38 ^{NS}	0,06	0,35 ^{NS}	-2,39	5,95*	-0,27	0,20 ^{NS}	3,82	2,32 ^{NS}
17	P5-P	9,99**	-0,06	0,32 ^{NS}	-5,55	32,06**	0,86	2,13 ^{NS}	8,77	12,24**
18	P6-P	3,27*	0,13	1,49 ^{NS}	-1,16	1,40 ^{NS}	-0,30	0,25 ^{NS}	6,72	7,20**
19	P7-P	2,48 ^{NS}	0,16	2,35 ^{NS}	-0,47	0,23 ^{NS}	0,10	0,03 ^{NS}	5,92	5,57*
20	P8-P	7,55**	0,07	0,43 ^{NS}	-4,49	21,01**	-0,47	0,62 ^{NS}	6,90	7,57**
21	P9-P	2,86*	0,03	0,08 ^{NS}	-0,53	0,30 ^{NS}	-0,60	1,02 ^{NS}	6,30	6,32*
22	P10-P	4,19**	0,07	0,47 ^{NS}	-0,99	1,01 ^{NS}	1,03	3,02 ^{NS}	9,79	15,27**
23	P11-P	5,71**	-0,10	0,94 ^{NS}	-1,97	4,04*	-0,57	0,91 ^{NS}	9,99	15,90**
24	P12-P	1,17 ^{NS}	0,09	0,71 ^{NS}	-1,31	1,78 ^{NS}	0,10	0,03 ^{NS}	2,93	1,37 ^{NS}
Błąd standardowy Standard error			0,1065		0,9802		0,5944		2,5065	

c. d. Tabela 6

1	2	3	4	5	6	7	8	9	10	11
Porównanie formy rodzicielskiej N ze średnim wynikiem dla linii nagich Comparison of the parental form N with mean value for the naked lines										
25	Nsr-N	1,30 ^{NS}	-0,08556	1,19 ^{NS}	-1,60750	4,97*	0,19167	0,19 ^{NS}	2,12056	1,32 ^{NS}
Błąd standardowy Standard error			0,0784		0,7214		0,4375		1,8447	
Porównanie formy rodzicielskiej P ze średnim wynikiem dla linii oplewionych Comparison of parental form P with mean value for the husked lines										
26	Psr-P	5,74**	0,06917	0,78 ^{NS}	-2,07917	8,31**	-0,09722	0,05 ^{NS}	6,64472	12,97**
Błąd standardowy Standard error			0,0784		0,7214		0,4375		1,8447	
Porównanie średniego wyniku dla linii nagich ze średnim wynikiem dla linii oplewionych Comparison of mean value for the naked lines with mean value for the husked lines										
27	Psr-Nsr	120,91**	0,38472	156,47**	4,31167	232,19**	-1,15556	45,35**	2,52750	12,20**
Błąd standardowy Standard error			0,0308		0,2830		0,1716		0,7236	
Porównanie formy rodzicielskiej P z formą rodzicielską N Comparison of the parental form P with the parental form P										
28	P-N	7,10**	0,23000	4,66*	4,78333	23,81**	-0,86667	2,13 ^{NS}	1,99667	0,63 ^{NS}
Błąd standardowy Standard error			0,1065		0,9802		0,5944		2,5065	

Psr — wartość średnia dla wszystkich linii oplewionych; Nsr — wartość średnia dla wszystkich linii nagich; ** — istotne przy $\alpha = 0{,}01$;
* — istotne przy $\alpha = 0{,}05$; ^{NS} — nie istotne.
Psr — mean value for all husked lines; Nsr — mean value for all naked lines; ** — significant at $\alpha = 0.01$; * — significant at $\alpha = 0.05$;
^{NS} — not significant.
N.k. — numer kontrastu; Contrast number

Postawiono hipotezę ogólną w postaci: $H_{0,0}$: nie ma istotnych różnic pomiędzy porównywanymi obiektami pod względem wszystkich analizowanych cech. Wartość statystyki testowej wynosiła $F = 7{,}73$ i była istotna statystycznie przy $\alpha = 0{,}05$. Pozwoliło to na odrzucenie hipotezy ogólnej i stwierdzenie, że pomiędzy badanymi obiektami występowały istotne różnice w ramach obserwowanych cech.

Odrzucenie hipotezy ogólnej pozwoliło dokonać weryfikacji hipotez szczegółowych dotyczących testowania istotności różnic pomiędzy badanymi obiektami w ramach każdej z analizowanych cech osobno (jednocechowa analiza wariancji). W procesie testowania hipotez szczegółowych H_{0i} ($i = 1, 2, 3, 4$) uzyskano następujące wyniki:

— $H_{0,1}$: badane obiekty nie różniły się istotnie ze względu na średnicę źdźbła. Statystyka testowa $F = 7{,}67$ była istotna statystycznie. Pozwoliło to na odrzucenie hipotezy zerowej i stwierdzenie, że obiekty różniły się między sobą ze względu na średnicę źdźbła.

— $H_{0,2}$: badane obiekty nie różniły się istotnie ze względu na współczynnik elastyczności. Statystyka testowa $F = 14{,}88$ była istotna statystycznie. Pozwoliło to na odrzucenie hipotezy zerowej i stwierdzenie, że obiekty różniły się między sobą ze względu na współczynnik elastyczności.

— $H_{0,3}$: badane obiekty nie różniły się istotnie ze względu na długość kłosa. Statystyka testowa $F = 3{,}97$ była istotna statystycznie. Pozwoliło to na odrzucenie hipotezy zerowej i stwierdzenie, że obiekty różniły się między sobą ze względu na długość kłosa.

— $H_{0,4}$: badane obiekty nie różniły się istotnie ze względu na MTZ. Statystyka testowa F = 3,19 była istotna statystycznie. Pozwoliło to na odrzucenie hipotezy zerowej i stwierdzenie, że obiekty różniły się między sobą ze względu na MTZ.

Wyznaczono kontrasty służące do porównania badanych obiektów oraz błędy standardowe porównań i statystki testowe służące do testowania hipotez zerowych mówiących, że poszczególne kontrasty są równe 0, czyli mówiącej, że nie ma różnic pomiędzy obiektami budującymi dany kontrast ze względu na wszystkie cechy jednocześnie oraz ze względu na każdą z analizowanych cech z osobna (tab. 6).

Wyznaczenie kontrastów jest jednym ze sposobów statystycznego porównania badanych obiektów lub ich grup w wielozmiennej analizie wariancji. W uproszczeniu polega na wykonaniu oddzielnych analiz tylko dla porównywanych obiektów lub ich grup. Pierwszym krokiem analizy jest określenie, które obiekty, bądź grupy obiektów, należy porównać. Następnie wskazane kontrasty zapisuje się w postaci macierzy **C** o k kolumnach i g wierszach, gdzie k jest liczbą analizowanych cech, a g jest liczbą testowanych kontrastów (Timm, 2002). Po określeniu tej macierzy można przeprowadzić testowanie globalnej hipotezy w postaci (Kaczmarek i Mańkowski, 2011):

$$H_{00}: \mathbf{CTM} = 0.$$

Testowanie poszczególnych kontrastów może przebiegać na dwóch płaszczyznach. Istnieje możliwość testowania pojedynczych kontrastów wielocechowych uwzględniających wszystkie analizowane cechy jednocześnie, wówczas hipoteza zerowa jest postaci (Kaczmarek i Mańkowski, 2011):

$$H_{f0}: \mathbf{c}_f'\mathbf{TM} = \mathbf{0}, \qquad f = 1,2,\dots,g^*$$

Drugą możliwością jest szczegółowe testowanie kontrastów jednocechowych dla wszystkich badanych cech, wówczas hipotezę zerową zapisuje się w postaci (Kaczmarek i Mańkowski, 2011):

$$H_{fl}: \mathbf{c}_f'\mathbf{Tm}_l = \mathbf{0},$$

Istotnie różne pod względem analizowanych cech od formy rodzicielskiej N okazały się linie N1, N4, N5, N7 oraz N11. Istotnie różne pod względem analizowanych cech od formy rodzicielskiej P były linie P2, P3, P5, P6, P8, P9, P10 oraz P11. Stwierdzono również występowanie istotnych różnic pomiędzy liniami oplewionymi oraz liniami nieoplewionymi, zarówno w ramach wszystkich analizowanych cech łącznie, jak i dla każdej cechy osobno (tab. 6).

3. ANALIZA ZMIENNYCH KANONICZNYCH

Dane wielocechowe można przedstawić jako rozmieszczenie badanych obiektów w przestrzeni n-wymiarowej, gdzie n jest liczbą analizowanych cech. W omawianym doświadczeniu obserwowano 4 cechy, w związku z tym przestrzeń danych jest przestrzenią czterowymiarową. Ponieważ człowiek jest w stanie zaprezentować graficznie najwyżej przestrzeń trójwymiarową, poszukuje się sposobu na redukcję wymiarów danych wielocechowych. Jedną z takich możliwości daje analiza zmiennych kanonicznych.

Zmienne kanoniczne są liniową kombinacją analizowanych zmiennych. Zawierają w sobie jednak znacznie więcej informacji niż pojedyncze zmienne w zbiorze danych. Najpierw wyznaczono statystyki F dla analizowanych zmiennych (tab. 7). Istotną wartość statystyki F miała pierwsza zmienna kanoniczna i objaśniała ona 76% obserwowanej zmienności. Druga zmienna kanoniczna objaśniała 13% zmienności. Ograniczając się tylko do dwóch pierwszych zmiennych kanonicznych 'utracono' jedynie 11% obserwowanej zmienności. Czyli strata wynikająca ze zmniejszenia liczby wymiarów dla analizowanych obiektów z czterech do dwóch wynosiła jedynie 11% obserwowanej zmienności. Natomiast 89% zmienności było zawarte w przestrzeni dwuwymiarowej. Wektory własne, opisujące strukturę zmiennych kanonicznych, przedstawiono w tabeli 8. Wektory własne charakteryzują zależności pomiędzy poszczególnymi analizowanymi cechami a wyznaczonymi zmiennymi kanonicznymi.

Tabela 7

Wartości statystyki F dla analizowanych zmiennych oraz stopień wyjaśnianej zmienności
F-statistic values and percent of explained variation for analyzed variables

Zmienne kanoniczne Canonical variables	Lamba Wiksa Wilks lambda	Statystyka F F-statistic	Procent zmienności Percent of variance
V_1	12,9019	5,87	75,94%
V_2	2,1273	0,97	12,52%
V_3	1,1061	0,50	6,51%
V_4	0,8539	0,39	5,03%

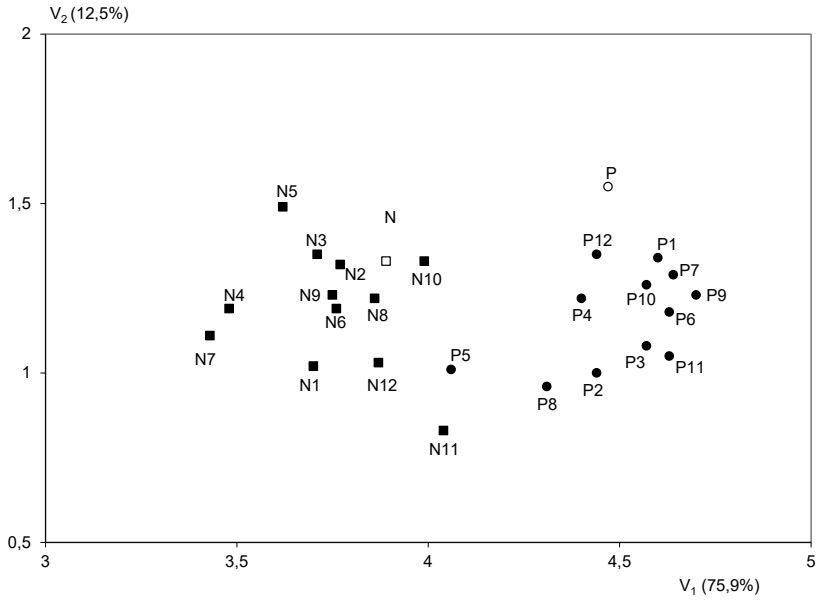
Tabela 8

Wektory własne
Eigenvectors

Zmienne Variables	V_1	V_2	V_3	V_4
Średnica źdźbła Stalk diameter	0,235207	0,102064	–0,12246	0,032005
Współczynnik elastyczności Coefficient of elasticity	0,248945	–0,06484	–0,11209	0,03308
Długość kłosa Spike length	1,07355	–0,05099	0,1167	–0,00268
MTZ Mass of 1000 grains	–0,38415	0,033945	0,090056	0,027901

Wyznaczono współrzędne kanoniczne dla wszystkich 26 obiektów i przedstawiono ich rozmieszczenie w układzie dwóch pierwszych zmiennych kanonicznych (rys. 1). Na rysunku widać wyraźnie rozdzielające się grupy linii nagich i oplewionych.

W celu określenia dokładnych odległości między poszczególnymi liniami wyznaczono odległości Mahalanobisa ze względu na cztery analizowane cechy. Przedstawiono je w postaci 3 tabel. Najpierw wyznaczono odległości pomiędzy liniami nagimi (tab. 9), następnie odległości Mahalanobisa między liniami oplewionymi i nagimi (tab. 10) oraz odległości pomiędzy liniami oplewionymi (tab. 11).



Rys. 1. Rozmieszczenie badanych linii na płaszczyźnie w układzie dwóch pierwszych zmiennych kanonicznych V_1 i V_2
Fig. 1. Location of the studied lines in the two dimensional space at the level of the first two canonical variables V_1 and V_2

Tabela 9

Odległości D^2 Mahalanobisa pomiędzy liniami nieoplewionymi
D^2 Mahalanobis distances between the naked lines

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12
N2	2,92											
N3	2,98	1,98										
N4	1,96	2,60	2,50									
N5	3,74	3,39	1,87	2,98								
N6	2,20	1,57	1,24	2,36	2,91							
N7	2,29	3,19	2,66	1,13	3,29	2,51						
N8	2,57	2,65	1,58	3,17	2,64	1,38	3,27					
N9	2,43	2,31	0,97	2,54	2,39	0,89	2,51	0,91				
N10	4,12	2,30	2,39	4,36	3,98	2,13	4,55	2,34	2,44			
N11	3,93	4,46	4,48	5,23	6,02	3,46	4,99	3,46	3,69	3,61		
N12	1,68	2,33	2,98	2,99	4,34	1,81	3,28	2,36	2,42	3,01	2,80	
N	2,76	1,61	1,59	3,07	2,75	1,38	3,57	1,42	1,64	1,88	4,04	2,22

Wartości krytyczne: $D^2_{\alpha=0,05} = 2,7$, $D^2_{\alpha=0,01} = 3,26$
Critical values: $D^2_{\alpha=0.05} = 2.7$, $D^2_{\alpha=0.01} = 3.26$

Tabela 10

Odległości D^2 Mahalanobisa pomiędzy liniami nieoplewionymi oraz liniami oplewionymi
D^2 Mahalanobis distances between the naked lines and the husked lines

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N
P1	6,87	6,17	6,37	7,98	7,08	6,05	8,42	5,36	6,15	4,76	5,68	5,74	5,05
P2	5,36	5,36	6,04	6,88	7,14	5,18	7,26	4,88	5,55	4,58	3,95	4,05	4,62
P3	6,51	6,78	6,67	8,01	7,35	6,18	8,25	5,27	6,13	5,45	4,95	5,59	5,49
P4	5,20	4,87	5,19	6,47	6,00	4,67	6,92	4,06	4,85	3,85	4,43	4,11	3,75
P5	3,29	4,21	3,70	4,75	4,64	3,04	4,74	2,15	2,87	3,40	2,41	2,69	3,02
P6	6,86	6,46	6,73	8,18	7,60	6,21	8,53	5,56	6,35	5,05	5,19	5,65	5,40
P7	7,08	6,53	6,65	8,27	7,39	6,29	8,65	5,56	6,36	5,03	5,63	5,95	5,38
P8	4,43	5,18	5,34	6,13	6,23	4,56	6,40	3,92	4,68	4,47	3,38	3,52	4,11
P9	7,49	6,66	7,19	8,67	8,23	6,65	9,06	6,23	6,92	5,22	5,65	6,10	5,85
P10	7,03	6,45	6,27	8,14	7,14	5,96	8,32	5,16	5,92	4,56	5,01	5,88	5,27
P11	7,20	6,53	7,09	8,46	8,41	6,37	8,70	6,10	6,70	5,03	4,65	5,68	5,89
P12	5,84	5,14	5,34	6,89	6,02	5,04	7,37	4,37	5,15	3,94	5,15	4,80	3,97
P	6,76	5,30	5,81	7,44	6,49	5,66	8,04	5,29	5,90	4,19	6,28	5,64	4,49

Wartości krytyczne: $D^2_{\alpha=0,05} = 2,7$, $D^2_{\alpha=0,01} = 3,26$
Critical values: $D^2_{\alpha=0.05} = 2.7$, $D^2_{\alpha=0.01} = 3.26$

Tabela 11

Odległości D^2 Mahalanobisa pomiędzy liniami oplewionymi
D^2 Mahalanobis distances between the husked lines

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P2	2,99											
P3	2,33	2,81										
P4	1,76	1,85	2,29									
P5	4,59	3,62	3,69	3,23								
P6	1,17	2,32	1,72	1,83	4,39							
P7	0,58	3,01	1,96	1,99	4,59	0,86						
P8	3,50	1,81	2,36	2,05	2,24	2,97	3,45					
P9	1,75	2,71	3,05	2,60	5,29	1,35	1,69	3,93				
P10	1,90	3,61	2,26	2,70	4,17	1,86	1,58	3,69	2,49			
P11	2,99	2,50	3,39	3,15	4,94	2,19	2,81	3,75	1,63	2,86		
P12	1,16	2,76	2,60	1,03	3,86	1,88	1,60	3,01	2,52	2,41	3,45	
P	2,10	3,90	4,27	2,60	5,32	3,04	2,65	4,63	2,85	3,45	4,02	1,82

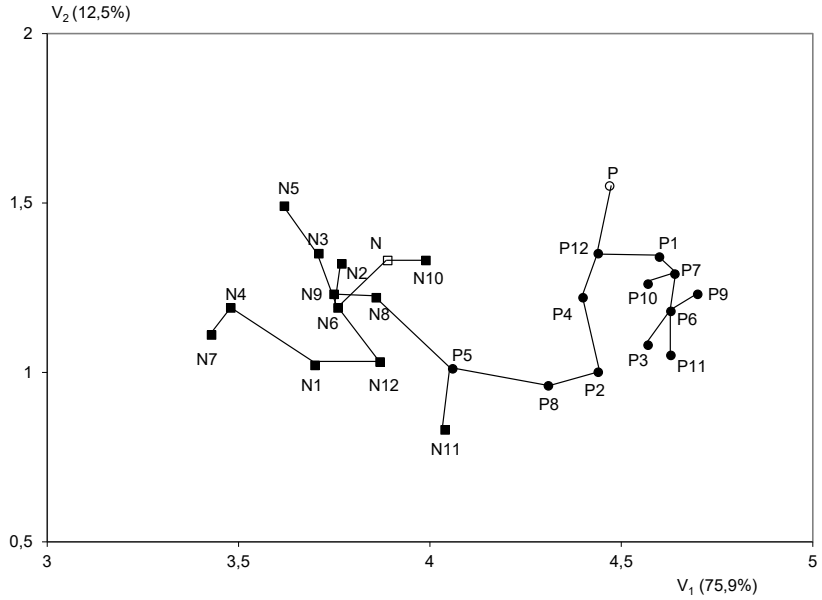
Wartości krytyczne: $D^2_{\alpha=0,05} = 2,7$, $D^2_{\alpha=0,01} = 3,26$
Critical values: $D^2_{\alpha=0.05} = 2.7$, $D^2_{\alpha=0.01} = 3.26$

Obliczone odległości D^2 Mahalanobisa mogą być testowane statystycznie. Stawia się hipotezę zerową, mówiącą że odległość ta jest równa zero ($H_0: D^2 = 0$). Następnie porównuje się wyznaczoną odległość z odległością krytyczną przy założonym poziomie istotności. Jeżeli odległość wyliczona jest większa od odległości krytycznej, to hipotezę zerową odrzucamy (stwierdzamy, że odległość pomiędzy dwoma porównywanymi obiektami jest istotna). W przeciwnym przypadku — nie mamy podstaw do odrzucenia hipotezy zerowej orzekającej, że badane obiekty nie różnią się.

Istnieje ścisły związek pomiędzy wynikami testowania odległości Mahalanobisa pomiędzy dwoma obiektami, a wynikiem testowania kontrastu pomiędzy tymi obiektami w wielocechowej analizie wariancji. Jeżeli wartość kontrastu była istotna statystycznie

(istniała istotna wielocechowa różnica pomiędzy obiektami) to odległość Mahalanobisa między tymi obiektami również będzie istotna statystycznie (będzie istotnie różna od zera).

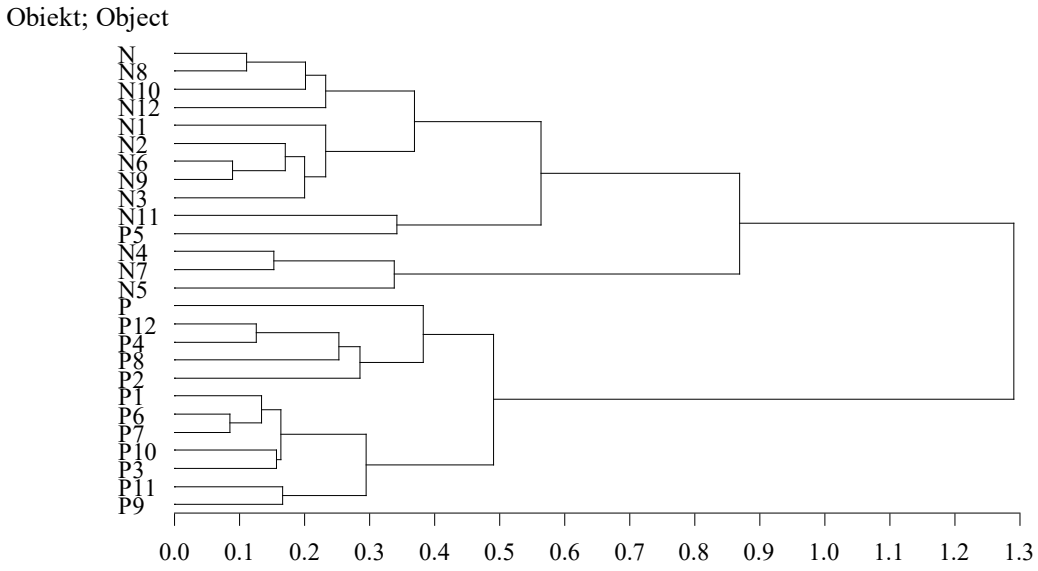
Na podstawie wyznaczonych odległości Mahalanobisa można utworzyć dendrogram (zgodnie z metodyką analizy skupień), bądź też dendryt najkrótszych połączeń. Dendryt najkrótszych połączeń polega na łączeniu ze sobą tych obiektów, które ze względu na wszystkie analizowane cechy są sobie najbliższe, czyli odległość między nimi jest najmniejsza, a co za tym idzie, połączenie tych obiektów obarczone będzie najmniejszym błędem (rys. 2). Dendrogram natomiast powstaje na skutek hierarchicznego łączenia ze sobą obiektów sobie najbliższych. W wyniku takiego podejścia uzyskuje się wykres przypominający drzewo. Z dendrogramu można odczytać kolejność tworzenia się skupień. Im połączenie pomiędzy obiektami lub skupieniami jest bliższe początku wykresu tym obiekty tworzące dane skupienie są mniej zróżnicowane między sobą. Na rysunku 3 przedstawiono dendrogram uzyskany dla badanych w wielozmiennym doświadczeniu form jęczmienia jarego z wykorzystaniem metody aglomeracji średniego wiązania (UPGMA).



Rys. 2. Rozmieszczenie badanych linii na płaszczyźnie w układzie dwóch pierwszych zmiennych kanonicznych V_1 i V_2 z dendrytem najkrótszych połączeń rozpiętym na punktach opisujących linie
Fig. 2. Location of the studied lines in the two dimensional space at the level of the first two canonical variables V_1 and $V_2$with the shortest dendrite connections stretched to the point describing the lines

Porównując uzyskany dendryt najkrótszy (rys. 2) oraz dendrogram hierarchicznej analizy skupień (rys. 3) stwierdzono wyraźne różnice w sposobie tworzenia skupień i ich strukturze. Różnice te wynikały z różnego podejścia prezentowanych metod do sposobu i kryteriów tworzenia skupień. Do przeprowadzenia hierarchicznej analizy skupień metodą UPGMA wykorzystano macierz odległości D^2 Mahalanobisa. Połączenia obiektów w

skupienia przebiegało zgodnie z kryterium najmniejszej średniej odległości pomiędzy łączonymi obiektami lub skupieniami. Natomiast dendryt najkrótszy powstaje na skutek połączenia ze sobą obiektów, dla których błąd (wyrażony wartością statystyki F) powodowany przez ich połączenie jest najmniejszy.



Rys. 3. Dendrogram uzyskany metodą UPGMA przedstawiający bliskość badanych w wielozmiennym doświadczeniu polowym nagich i oplewionych form jęczmienia jarego
Fig. 3. Tree-diagram obtained by UPGMA method, showing the proximity of the naked and husked spring barley forms tested in the multivariate field experiment

4. PODSUMOWANIE

Omówiony przykład pokazuje, jak wykonywać analizy wielocechowe: wielocechową analizę wariancji (MANOVA), analizę zmiennych kanonicznych oraz wielocechową ocenę podobieństwa badanych obiektów.

Wymienione analizy pozwoliły wykazać występowanie różnic pomiędzy badanymi obiektami, zarówno pojedynczo, jak i pomiędzy grupami obiektów (MANOVA, analiza kontrastów), pozwoliły również na zredukowanie wielowymiarowej przestrzeni, w której były opisane badane obiekty, do przestrzeni dwuwymiarowej z niewielką stratą informacji (analiza zmiennych kanoniczych), pozwoliły wreszcie na pogrupowanie badanych obiektów i scharakteryzowanie zróżnicowania między nimi (odległość Mahalanobisa).

Wielozmienna analiza wariancji często znajduje zastosowanie w badaniach wielocechowych. W literaturze można odnaleźć wiele prac, w których zastosowano tę metodę analizy danych doświadczalnych. Spośród prac opublikowanych w ostatnich latach można wymienić między innymi badania Crossa i Franco (2004). Pisali oni o znaczeniu

MANOVA w statystycznej klasyfikacji genotypów badanych w wielozmiennych doświadczeniach hodowlanych. Możliwość wykorzystania wielozmiennej analizy wariancji w połączeniu z analizą dyskryminacyjną do wyboru form rodzicielskich w hodowli kiwi przedstawili Daoyu i Lawes (2000). Do podobnego celu zastosowali wielozmienną analizę wariancji Nyassé i in. (2002), którzy poszukiwali rodziców do hodowli odpornościowej roślin kakao. MANOVA znalazła zastosowanie w ocenie zróżnicowania w kolekcjach zasobów genowych i kolekcjach roboczych wykorzystywanych w hodowli. Ukalska i in. (2007) prowadzili badania nad zmiennością fenotypową w kolekcji zasobów genowych truskawki. Ukalska i in. (2008) badali zmienność i współzależności cech użytkowych w kolekcji roboczej pszenicy ozimej.

Należy wskazać również możliwość wykorzystania tej metody w analizie kowariancji. Można tu wymienić badania Kenga i in. (2006), którzy wykorzystali wielozmienną analizę wariancji do oszacowania genotypowych i fenotypowych relacji pomiędzy składowymi plonu sorga. Takie zastosowanie MANOVA opisali również Mądry i in. (2010).

LITERATURA

Anderberg M.R. 1973. Cluster analysis for applications. Academic Press, New York.

Caliński T. 1970. Wielozmienna analiza wariancji i pokrewne metody wielowymiarowe. PAN, Warszawa.

Caliński T., Czajka S., Kaczmarek Z. 1975. Analiza składowych głównych i jej zastosowania. Algorytmy biometryczne i statystyczne (ABS-36). AR Poznań.

Caliński T., Dyczkowski A., Kaczmarek Z. 1976. Testowanie hipotez w wielozmiennej analizie wariancji i kowariancji. Algorytmy biometryczne i statystyczne (ABS-45). AR Poznań.

Caliński T., Kaczmarek Z. 1973. Metody kompleksowej analizy doświadczenia wielocechowego. Trzecie Colloquium Metodologiczne z Agro-Biometrii, PAN i PTB Warszawa: 257 — 320.

Ceranka B., Chudzik H., Kaczmarek Z., Krzyśko M. 1975. Wielozmienna analiza wyników doświadczeń w układach blokowych. Algorytmy biometryczne i statystyczne (ABS-35). AR Poznań.

Crossa J., Franco J. 2004. Statistical methods for classifying genotypes. Euphytica, 137: 19 — 37.

Daoyu Z., Lawes G. S. 2000. Manova and discriminant analyses of phenotypic data as a guide for parent selection in kiwifruit (*Actinidia deliciosa*) breeding. Euphytica, 114: 151 — 157.

Kaczmarek Z. 1975. Wielozmienna analiza kowariancji i jej niektóre zastosowania. Matematyka Stosowana 5: 139 — 156.

Kaczmarek Z., Czajka S., Adamska E. 2008. Propozycja metody grupowania obiektów jedno- i wielocechowych z zastosowaniem odległości Mahalanobisa i analizy skupień. Biul.IHAR 249: 9 — 18.

Kaczmarek Z., Mańkowski D. R. 2011. Wprowadzenie do statystycznych analiz wielozmiennych. Część I. Podstawy teoretyczne. Biuletyn IHAR, *w druku*.

Kenga R., Tenkouano A., Gupta S. C., Alabi S. O. 2006. Genetic and phenotypic association between yield components in hybrid sorgum (*Sorgum bicolor* (L.) Moench) populations. Euphytica, 150: 319 — 326.

Krzyśko M. 2000. Wielowymiarowa analiza statystyczna. Uniwersytet im. A. Mickiewicza w Poznaniu. Poznań.

Mądry W., Mańkowski D. R., Kaczmarek Z., Krajewski P., Studnicki M. 2010. Metody statystyczne oparte na modelach liniowych w zastosowaniach do doświadczalnictwa, genetyki i hodowli roślin. Monografie i Rozprawy Naukowe IHAR Radzików, nr 34.

Morrison D. F. 1976. Multivariate statistical methods. McGraw-Hill. New York.

Nyassé S., Despréaux D., Cilas C. Validity of a leaf inoculation test to assess the resistance to *Phytophthora megakarya* in a cocoa (*Theobroma cacao* L.) diallel mating design. Euphytica, 123: 395 — 399.

SAS Institute Inc. 2009. SAS/STAT 9.2 User's Guide, Second Edition. Cary, NC, USA: SAS Publishing, SAS Institute Inc.

Seber G.A.F. 1984. Multivariate observations. Wiley. New York.

Timm N. H. 2002. Applied multivariate analysis. New York, USA: Springer-Verlag Inc.

Ukalska J., Mądry W., Ukalski K., Masny A. 2007. Wielowymiarowa ocena różnorodności fenotypowej w kolekcji zasobów genowych truskawki. Cz. II. Grupowanie genotypów. Zesz. Probl. Post. Nauk Rol. 517: 759 — 766.

Ukalska J., Ukalski K., Śmiałowski T., Mądry W. 2008. Badanie zmienności i współzależności cech użytkowych w kolekcji roboczej pszenicy ozimej (*Triticum aestivum* L.) za pomocą metod wielowymiarowych. Cz. II. Analiza składowych głównych na podstawie macierzy korelacji fenotypowych i genotypowych. Biul. IHAR 249: 45 — 57.