Listy Biometryczne
Nr 34-36 (1972)

Bronisław Ceranka i Zygmunt Kaczmarek (Poznań)

ZASTOSOWANIE WIELOZMIENNEJ ANALIZY WARIANCJI DLA NIEORTOGONALNYCH UKŁADÓW BLOKOWYCH

1. Wstęp

W niniejszej pracy przedstawiono wielowymiarowe uogólnienie modelu Tochera [8] i Reesa [7]. Powyższy model pozwala na przeprowadzenie analizy wielu zmiennych dla dowolnego układu doświadczalnego o dwukierunkowej klasyfikacji z eliminacją niejednorodności w jednym kierunku. Przeprowadza się testowanie hipotezy dotyczącej różnicy wektorów poprawionych średnich obiektowych wraz z obliczaniem odległości Mahalanobisa, między poszczególnymi parami obiektów.

Zastosowanie tego modelu pokazano na przykładzie doświadczenia założonego w układzie częściowo zrównoważonych bloków niekompletnych. W każdym bloku występują dodatkowo dwa wzorce.

2. Wielozmienny model analizy wariancji dla doświadczeń blokowych

Model analizy wariancji jednej zmiennej dla doświadczeń blokowych rozważany był przez Tochera [8]. Dalsze uogólnienia i uzupełnienia zostały wprowadzone przez Jonesa [4], Reesa [7] i Pearce'a [6]. W pracy tej rozważać będziemy wielozmienny model analizy wariancji dla doświadczeń blokowych.

Ogólny model liniowy analizy wariancji jednej zmiennej

$$\underline{y} = \underline{A}\,\underline{\theta} + \underline{e} \tag{1}$$

w zastosowaniu do doświadczeń blokowych można zapisać w postaci

$$\underline{y} = \begin{bmatrix} \underline{1} & \underline{A}_1 & \underline{A}_2 \end{bmatrix} \begin{bmatrix} \mu \\ \underline{\beta} \\ \underline{\tau} \end{bmatrix} + \underline{e},$$

gdzie $\underline{y}$ jest wektorem obserwacji typu $N \times 1$, $\underline{A} = \begin{bmatrix} \underline{1} & \underline{A}_1 & \underline{A}_2 \end{bmatrix}$ jest macierzą układu typu $N \times (1+b+t)$ złożoną z wektora jedynkowego $\underline{1}$ typu $N \times 1$ oraz macierzy $\underline{A}_1$ typu $N \times b$ dla pierwszej klasyfikacji (bloków) i macierzy $\underline{A}_2$ typu $N \times t$ dla drugiej klasyfikacji (obiektów); $\underline{\theta}' = [\mu \ \underline{\beta}' \ \underline{\tau}']$ jest wektorem parametrów typu $(1+b+t)\times 1$ złożonym z parametru wspólnego μ (skalara), wektora $\underline{\beta}$ typu $b \times 1$ parametrów blokowych oraz wektora $\underline{\tau}$ typu $t \times 1$ parametrów obiektowych i gdzie $\underline{e}$ typu $N \times 1$ jest wektorem błędów losowych takich, że $E(\underline{e}) = \underline{0}$ oraz $\underline{E}(\underline{e}\,\underline{e}') = \sigma^2\,\underline{I}$, przy czym $\underline{I}$ jest macierzą jednostkową typu $N \times N$.

Rozważany model ograniczymy do przypadku, gdy macierz $\underline{A}$ jest rzędu $b+t-1$. Równania normalne w przypadku modelu (1) mają postać

$$\underline{A}'\underline{A}\,\underline{\theta} = \underline{A}'\underline{y},$$

a stąd oszacowanie parametrów $\underline{\theta}$ według zasady najmniejszych kwadratów ma postać

$$\hat{\underline{\theta}} = (\underline{A}'\underline{A})^{-}\,\underline{A}'\underline{y}$$

gdzie $(\underline{A}'\underline{A})^{-}$ jest uogólnioną macierzą odwrotną macierzy $(\underline{A}'\underline{A})$, tzn. macierzą spełniającą równanie $(\underline{A}'\underline{A})\,(\underline{A}'\underline{A})^{-}(\underline{A}'\underline{A}) = (\underline{A}'\underline{A})$. Ponieważ macierz $\underline{A}$ jest rzędu $b+t-1$, zatem istnieje macierz $\underline{D}$ typu $(1+b+t)\times 2$ rzędu 2 taka, że $\underline{A}\,\underline{D} = \underline{0}$. Istnieje również macierz $\underline{C}$ typu $2\times(1+b+t)$ rzędu 2 taka, że $|\,\underline{C}\ \ \underline{D}\,| \neq 0$ oraz $\underline{C}\,\underline{\theta} = \underline{0}$. Uogólnioną macierzą odwrotną macierzy $(\underline{A}'\underline{A})$ jest macierz $(\underline{A}'\underline{A}+\underline{C}'\underline{C})^{-1}$; tzn. spełniona jest równość $(\underline{A}'\underline{A})\,(\underline{A}'\underline{A}+\underline{C}'\underline{C})^{-1}\,(\underline{A}'\underline{A}) = (\underline{A}'\underline{A})$. Zatem

$$\hat{\underline{\theta}} = (\underline{A}'\,\underline{A} + \underline{C}'\,\underline{C})^{-1}\,\underline{A}'\,\underline{y}$$

jest oszacowaniem wektora parametrów $\underline{\theta}$ według zasady najmniejszych kwadratów. Dalej mamy

$$\underline{A}'\underline{y} = \begin{bmatrix} G \\ \underline{T}_1 \\ \underline{T}_2 \end{bmatrix},$$

gdzie G jest sumą ogólną, $\underline{T}_1$ jest wektorem typu $b \times 1$ sum blokowych, $\underline{T}_2$ jest wektorem typu $t \times 1$ sum obiektowych. Wprowadźmy następujące oznaczenia:

$$\underline{Q}_1 = \underline{T}_1 - \frac{1}{N}\underline{k}\,G$$

gdzie N jest liczbą obserwacji, $\underline{k}$ jest wektorem typu $b \times 1$, którego elementy określają wielkość bloków, oraz

$$\underline{Q}_2 = \underline{T}_2 - \underline{n}\,\underline{k}^{-d}\,\underline{T}_1$$

gdzie $\underline{n}$ jest macierzą incydencji typu $t \times b$, $\underline{k}^{-d}$ jest macierzą odwrotną do macierzy diagonalnej $\underline{k}^d$ typu $b \times b$ powstałej z wektora $\underline{k}$, oraz

$$\Omega^{-1} = \underline{r}^d - \underline{n}\,\underline{k}^{-d}\,\underline{n}' + \underline{r}\,\underline{r}'/N$$

gdzie $\underline{r}$ jest wektorem typu $t \times 1$ powtórzeń obiektów, $\underline{r}^d$ jest macierzą diagonalną typu $t \times t$ powstałą z wektora $\underline{r}$. Tocher [8] pokazał, że $\sigma^2\,\Omega$ jest macierzą kowariancji poprawionych średnich obiektowych, tzn. macierzą kowariancji wektora $\hat{\mu}\,\underline{1} + \hat{\underline{\tau}}$.

Po rozwiązaniu układu równań normalnych oszacowanie wektora parametrów $\underline{\theta}$ według zasady najmniejszych kwadratów dane jest w postaci

$$\hat{\underline{\theta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\underline{\beta}} \\ \hat{\underline{\tau}} \end{bmatrix} = \begin{bmatrix} G/N \\ \underline{k}^{-d}(\underline{Q}_1 - \underline{n}'\Omega\,\underline{Q}_2 + \underline{k}\,\underline{r}'\Omega\,\underline{Q}_2/N) \\ \Omega\,\underline{Q}_2 \end{bmatrix}.$$

Stąd wektor poprawionych średnich dla obiektów jest równy

$$\frac{1}{N}\underline{1}\,G + \Omega\,\underline{Q}_2\,.$$

Suma kwadratów dla błędu określona jest wzorem

$$E = (\underline{y} - \underline{A}\,\hat{\underline{\theta}})'\,(\underline{y} - \underline{A}\,\hat{\underline{\theta}}) = (\underline{y}'\underline{y} - \frac{1}{N}G^2) - (\underline{T}_1'\underline{k}^{-d}\underline{T}_1 - \frac{1}{N}G^2) - \underline{Q}_2'\Omega\,\underline{Q}_2 =$$

$$= \underline{y}'\left\{\underline{I} - \underline{A}_1\,\underline{k}^{-d}\,\underline{A}_1' - (\underline{A}_2 - \underline{A}_1\,\underline{k}^{-d}\,\underline{n}')\,\Omega\,(\underline{A}_2' - \underline{n}\,\underline{k}^{-d}\,\underline{A}_1')\right\}\underline{y} = \underline{y}'\underline{M}_E\,\underline{y}\,.$$

Poprawiona suma kwadratów dla obiektów wynosi

$$H = \underline{y}'\left\{(\underline{A}_2 - \underline{A}_1\,\underline{k}^{-d}\,\underline{n}')\,\Omega\,(\underline{A}_2' - \underline{n}\,\underline{k}^{-d}\,\underline{A}_1')\right\}\underline{y} = \underline{y}'\underline{M}_H\,\underline{y}\,.$$

Statystyką do testowania hipotezy $H_0 : \underline{\tau} = \underline{0}$ jest

$$F = \frac{H/(t-1)}{E/(N-t-b+1)}\,. \tag{2}$$

Jeśli hipoteza H_0 jest prawdziwa, to iloraz ten ma rozkład F z t-1 i N-t-b+1 stopniami swobody. Tym samym, jeśli chcemy zweryfikować H_0 na poziomie istotności α, to przyjmujemy tę hipotezę, gdy $F \leqslant F_{\alpha;\ t-1,\ N-t-b+1}$, a odrzucimy ją w przeciwnym razie.

Przy uogólnieniu na wielozmienny model analizy wariancji w modelu (1) wektor $\underline{y}$ zastępuje się macierzą

$$\underline{Y} = \left[\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_p\right],$$

a wektor parametrów $\underline{\theta}$ zastępuje się macierzą parametrów

$$\underline{\theta} = \left[\underline{\theta}_1, \underline{\theta}_2, \ldots, \underline{\theta}_p\right].$$

Macierz układu $\underline{A}$ pozostaje bez zmiany. Wektor $\underline{e}$ zastępuje się macierzą

$$\underline{e} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ e_{N4} & e_{N2} & \cdots & e_{Np} \end{bmatrix},$$

o której zakładamy, że $E(\underline{e}) = \underline{0}$ oraz, że

$$E(e_{ij}\, e_{i'j'}) = \begin{cases} 0 & \text{gdy} \quad i \neq i' \\ \sigma_{jj'} & \text{gdy} \quad i = i'. \end{cases}$$

Przy omawianiu testów i przedziałów ufności będziemy nadto zakładać, że każdy wiersz macierzy $\underline{e}$ ma ten sam wielowymiarowy rozkład normalny $N(\underline{0}, \underline{\Sigma})$, gdzie $\underline{\Sigma} = (\sigma_{jj'})$. Wówczas wiersze macierzy $\underline{e}$ będą statystycznie niezależne.

Wiadomo, że każde zagadnienie testowania hipotezy jednozmiennej za pomocą statystyki (2) można uogólnić na przypadek wielu zmiennych zastępując po prostu sumy kwadratów występujące w (2) przez ich uogólnienie macierzowe $\underline{H}$ i $\underline{E}$ i następnie stosując w miejsce testu F odpowiedni test wielozmienny (patrz Caliński [1]). Elementy H_{rr} macierzy $\underline{H}$ są sumami kwadratów odchyleń, a $H_{rr'}$ $(r \neq r')$ są sumami iloczynów odchyleń (r, r' = 1, 2, ..., p). Analogicznie jest dla elementów E_{rr} i $E_{rr'}$ macierzy $\underline{E}$.

3. Test F w wielozmiennej analizie wariancji

Macierze $\underline{H}$ i $\underline{E}$ służą do testowania hipotezy $H_0: \underline{\tau} = \underline{0}$ przy pomocy statystyki

$$F = \frac{ms-2n}{p(t-1)} \cdot \frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}$$

gdzie $m = N-t-b+1 + \frac{t-p-2}{2}$; $n = \frac{p(t-1)-2}{4}$;

$$s = \sqrt{p^2(t-1)^2-4} \Big/ \sqrt{p^2+(t-1)^2-5},$$

natomiast Λ zwana dalej Λ Wilksa jest ilorazem wyznacznikowym

$$\Lambda = |\underline{E}| \,/\, |\underline{H}+\underline{E}|.$$

Tak zdefiniowana statystyka przy prawdziwości hipotezy H_0 ma w przybliżeniu rozkład F z $p(t-1)$ i $ms-2n$ stopniami swobody. Hipotezę zerową odrzucamy gdy

$$F > F_{\alpha;\, p(t-1),\, ms-2n}.$$

W analogiczny sposób możemy przeprowadzić testowanie hipotezy o braku różnic między wektorami średnich dla poszczególnych bloków. Celem doświadczenia jest jednak przede wszystkim badanie istotności różnic między poszczególnymi obiektami i temu zagadnieniu poświęcimy nieco miejsca.

4. Porównanie średnich obiektowych. Odległość Mahalanobisa

W analizie wielu zmiennych tj. gdy rozpatrujemy jednocześnie wiele cech, problem porównania między sobą obiektów polega na zastosowaniu odpowiedniej miary odległości między dwoma obiektami, która uwzględniałaby wszystkie p cech. Najwłaściwszą miarą tej odległości jest tzw. odległość Mahalanobisa. Oceny odległości Mahalanobisa między dwoma obiektami wyliczamy z następującej formy kwadratowej

(3) $$D^2(i,i') = (N-t-b+1)(\overline{\underline{y}}^{(i)}-\overline{\underline{y}}^{(i')})' \, \underline{E}^{-1}(\overline{\underline{y}}^{(i)}-\overline{\underline{y}}^{(i')}).$$

Odległość krytyczną znajdujemy przy pomocy statystyki T^2 Hotellinga, która jest ściśle powiązana z rozkładem F. Ponieważ istnieje bezpośredni związek między D^2 i T^2, więc tym samym można podać "najmniejszą istotną odległość" między dwoma obiektami.

Statystyka

$$T^2 = \frac{1}{z_{ii'}} D^2(i, i') \tag{4}$$

gdzie $z_{ii'} = w_{ii} - 2w_{ii'} + w_{i'i'}$, przy czym $w_{ii'}$ są elementami macierzy kowariancji poprawionych średnich obiektowych Ω, ma przy prawdziwości hipotezy H_0 rozkład T^2 (Hotellinga) z N-t-b+1 stopniami swobody, natomiast statystyka

$$F = \frac{N-b-t-p+2}{(N-b-t+1)p} T^2$$

przy prawdziwości hipotezy H_0 ma rozkład F z p i (N-b-t-p+2) stopniami swobody. Zatem

$$D^2(i, i') = z_{ii'} \frac{(N-t-b+1)p}{N-t-b-p+2} F_{\alpha;\, p,\, N-t-b-p+2} \tag{5}$$

będzie wartością krytyczną na poziomie istotności α dla funkcji (3).

W praktyce wygodniej jest posługiwać się odległością wielowymiarową będącą pierwiastkiem z (5). Tak zdefiniowana odległość jest równoznaczna z odległością geometryczną w uogólnionej przestrzeni euklidesowej.

5. Udział poszczególnych zmiennych w zróżnicowaniu obiektów

Doświadczalnika mającego do swej dyspozycji większą liczbę cech może zainteresować ustalenie ich kolejności ze względu na udział poszczególnych cech w zmienności między obiektami oraz ewentualne wyeliminowanie cech nieistotnych. Statystycznie powyższe badania można przeprowadzić przy pomocy techniki analizy kowariancji dla 1 zmiennych niezależnych i (p-l) zmiennych zależnych. Testem stwierdzającym istotność udziału poszczególnych zmiennych przy ustalonych pozostałych jest kryterium ilorazowe Λ Wilksa. Dla ustalenia udziału l-tej zmiennej kryterium ilorazowe Wilksa przyjmie postać

$$\Lambda_{l/12\ldots(l-1)(l+1)\ldots p} = \Lambda_{12\ldots p} / \Lambda_{12\ldots(l-1)(l+1)\ldots p}\,. \tag{6}$$

Kryterium to stanowi pewną względną miarę "ważności" zmiennych. Można je przedstawić za pomocą statystyki F

$$F_l = \frac{N-b-t-p+2}{t-1} \cdot \frac{1-\Lambda_{1/12\ldots(l-1)(l+1)\ldots p}}{\Lambda_{1/12\ldots(l-1)(l+1)\ldots p}} . \tag{7}$$

Wartość krytyczną F odczytujemy z tablic rozkładu F dla t - 1 i (N-t-b-p+2) stopni swobody. Korzystając z (7) możemy wyliczyć wartość F dla każdej zmiennej i uporządkować zmienne w kolejności od najmniejszej do największej wartości F. Porównując obliczone wartości F z wartościami krytycznymi można przeprowadzić eliminację cech nieistotnych. Odbywa się to w ten sposób, że najpierw odrzucamy cechę mającą najmniejszą nieistotną wartość F, a po jej odrzuceniu ponownie badamy udział pozostałych, tym razem już przy mniejszej (o jeden) liczbie cech ustalonych. Odrzucamy drugą z kolei cechę nieistotną (oczywiście gdy taka istnieje). Ponownie badamy udział pozostałych. Powyższe obliczenia prowadzimy do momentu, gdy wszystkie cechy będą miały wartość $F_{obl.}$ większą od wartości krytycznej F. Po wyeliminowaniu cech nieistotnych przeprowadzamy ponownie analizę wariancji wraz z obliczeniem odległości Mahalanobisa między wszystkimi parami obiektów.

6. Przykład

Jako przykład wielozmiennej analizy wariancji dla układu bloków nieortogonalnych niech posłuży analiza danych pochodzących z badań hodowlanych nad słonecznikiem, przeprowadzonych w SHB Borowo. Doświadczenie założono metodą bloków niekompletnych częściowo zrównoważonych z t = 25 obiektami w r = 4 powtórzeniach w b = =20 blokach, każdy o stałej wielkości k = 5. Dodatkowo w każdym bloku znajdowały się dwa wzorce. Na dane składają się obserwacje 11 następujących cech:

1) wysokość roślin w cm
2) średnica tarczy w cm
3) plon nasion w g/roślinę
4) plon nasion w q/ha
5) ciężar 1000 nasion w g
6) procent łuski
7) liczba roślin do sprzętu
8) procent roślin z odrostami
9) procent tłuszczu w całych niełupkach

Tabela 1. Średnie poprawione

Rody \ Cechy	Wysokość roślin w cm	Średnica tarczy w cm	Plon nasion w g/roślinę	Plon nasion w q/ha	Ciężar 1000 nasion w g
1	72,23	15,86	52,95	24,60	57,30
2	92,24	17,64	54,26	23,97	55,82
3	68,22	14,70	45,85	23,07	58,47
4	74,88	15,99	52,03	23,61	61,35
5	70,43	15,05	47,33	23,67	58,94
6	67,17	15,00	45,07	21,94	63,65
7	59,40	13,03	34,93	14,65	62,12
8	77,13	15,70	50,40	24,44	59,31
9	72,30	14,67	46,17	23,57	70,13
10	72,73	14,95	46,41	24,13	62,51
11	73,81	15,20	44,55	25,98	64,18
12	68,80	16,40	48,29	19,91	63,00
13	75,56	15,35	47,60	25,25	53,83
14	76,23	15,02	47,06	21,02	69,92
15	62,96	13,05	39,62	24,11	60,29
16	63,86	14,65	41,86	24,01	54,17
17	69,41	15,47	53,18	28,85	58,34
18	67,28	15,16	47,60	21,34	51,83
19	70,36	15,45	45,99	22,79	56,39
20	67,08	13,66	42,32	21,20	54,49
21	60,22	13,77	46,47	20,97	59,63
22	70,23	15,58	44,98	21,37	52,18
23	59,87	13,09	39,81	18,15	70,85
24	60,51	13,46	39,35	18,17	62,50
25	74,45	15,18	49,17	28,21	62,67
wzorzec 1	67,12	14,13	42,10	23,84	62,25
wzorzec 2	73,20	15,13	47,70	22,98	61,90

jedenastu cech rodów słonecznika

Procent łuski	Liczba roślin do sprzętu	Procent roślin z odrostami	Procent tłuszczu w całych niełupkach	Plon oleju w q/ha	Procent tłuszczu w niełupkach odłuszczonych
25,97	12,91	27,80	49,14	12,01	62,63
31,47	12,94	5,43	42,94	10,19	59,24
27,59	14,65	8,76	48,84	11,26	63,49
29,98	13,25	36,34	46,43	10,73	61,33
26,89	14,67	28,32	47,63	11,25	63,92
27,58	14,07	19,34	47,07	10,35	61,12
27,05	12,34	21,22	47,31	6,89	61,08
29,27	14,12	12,03	47,16	11,55	62,18
25,80	14,48	29,72	47,11	11,09	61,23
27,12	14,98	15,74	49,98	12,51	64,05
28,65	16,44	3,97	46,23	11,88	61,60
26,02	11,41	18,83	48,15	9,52	62,33
28,45	15,20	23,68	49,42	12,28	63,81
29,99	12,85	16,65	46,10	9,55	62,54
26,57	17,16	3,69	48,20	11,57	62,52
26,22	16,23	6,88	49,30	11,76	63,90
27,88	15,64	14,67	48,53	13,89	63,45
29,20	12,91	23,43	48,59	10,38	62,93
27,79	13,93	33,45	48,07	11,02	63,08
28,02	14,14	25,91	49,95	10,65	63,87
27,10	13,28	42,53	48,63	10,08	62,65
27,59	13,34	29,41	48,19	10,29	63,50
28,34	13,15	23,40	45,95	8,36	61,24
29,19	13,18	29,64	47,62	8,44	61,83
26,50	16,25	4,05	47,59	13,29	62,59
27,15	16,20	34,53	47,97	11,40	62,27
28,45	13,90	20,99	47,23	10,81	62,83

10) plon w oleju w q/ha

11) procent tłuszczu w niełupkach odłuszczonych.

Średnie poprawione uzyskane z obserwacji dla poszczególnych cech i rodów zestawione zostały w tabeli 1. Macierz poprawionych sum kwadratów i iloczynów dla obiektów została zestawiona w tabeli 2, a dla błędu w tabeli 3.

Wyznaczniki potrzebne do obliczenia statystyki Wilksa przyjmują tu następujące wartości:

$$|\underline{E}| = 0,371567988 \cdot 10^{27}$$

$$|\underline{H} + \underline{E}| = 0,258960884 \cdot 10^{30}.$$

Stąd

$$\Lambda = |\underline{E}| / |\underline{H} + \underline{E}| = 0,001434842.$$

Po przekształceniu tej statystyki w statystykę F otrzymujemy $F_{obl.} = 2,797$. Wartość tę porównujemy z górnymi wartościami krytycznymi odczytanymi dla p(t-1) = 286 i ms-2n = 884 stopni swobody. Ponieważ $F_{obl.}$ jest znacznie większe od $F_{0,01;286;884}$, więc odrzucamy hipotezę o równości rodów pod względem średnich wartości badanych cech.

Dokonujemy obecnie badania udziału poszczególnych cech, tzn. chcemy odpowiedzieć na pytanie w jakim stopniu każda z badanych cech wpływa na zmienność między rodami. Przy pomocy kryterium Λ Wilksa i testu F będziemy prowadzić eliminację cech, których udział jest nieistotny.

Dla poszczególnych cech uzyskano następujące wartości F:

$$F_1 = 3,354$$
$$F_2 = 1,252$$
$$F_3 = 1,236$$
$$F_4 = 1,605$$
$$F_5 = 3,433$$
$$F_6 = 2,104$$
$$F_7 = 1,526$$
$$F_8 = 3,142$$
$$F_9 = 1,747$$
$$F_{10} = 1,286$$
$$F_{11} = 2,340$$

Wartość krytyczna wynosi F = 1,65. Najniższą nieistotną wartość F ma cecha nr 3, czyli plon nasion w gramach na roślinę. Eliminujemy ją i ponownie badamy udział pozostałych cech. Eliminacja cech

T a b e l a 2. Macierz poprawionych sum kwadratów i iloczynów dla obiektów

4641,3	624,6	2482,3	975,7	-397,0	488,8	-114,4	-272,3	-484,2	342,0	-161,0
	118,0	454,4	135,9	-160,9	52,5	-52,6	-337,5	-48,4	51,5	-10,8
		2294,8	793,2	-547,7	203,3	-191,6	-1016,3	-123,7	339,7	22,3
			843,5	-249,4	-34,6	289,0	-819,6	37,8	407,0	70,5
				2353,9	-50,0	12,74	154,7	-283,9	-180,7	-228,0
					184,9	-71,4	-275,1	-118,9	-48,4	-53,3
						242,9	-171,7	50,0	149,8	31,6
							13577,4	405,7	-297,6	95,1
								208,7	72,5	131,9
									211,9	68,5
										126,0

T a b e l a 3. Macierz sum kwadratów i iloczynów dla błędu

1846,9	114,2	904,4	399,2	-243,9	3,9	-24,6	87,9	2,6	188,2	85,1
	118,3	437,3	139,1	160,9	4,8	-56,4	76,3	-2,1	63,6	9,2
		4754,6	1335,9	1358,4	27,7	-617,2	1567,2	-132,0	607,9	41,4
			996,8	362,3	31,4	192,4	-11,6	-53,8	465,6	11,1
				2898,1	94,8	-222,8	282,5	-235,9	94,7	-110,6
					171,0	10,0	36,5	-14,6	11,5	7,0
						309,9	-510,1	12,8	95,0	-0,6
							13414,3	-14,0	-32,3	-19,4
								97,4	-3,4	56,8
									227,4	17,8
										89,0

T a b e l a 4. Badanie udziału cech

Numer cechy	Wartości F: Dla wszystkich 11-tu cech	Dla 10-ciu cech po wyeliminowaniu cechy 3	Dla 9-ciu cech po wyeliminowaniu cechy 3 i 2	Dla 8-miu cech po wyeliminowaniu cech 3, 2 i 10
(1)	3,354 *	3,547 *	4,457 *	4,447 *
(2)	1,252	<u>1,183</u>		
(3)	<u>1,236</u>			
(4)	1,605	1,415	1,487	1,658 *
(5)	3,433 *	3,488 *	3,487 *	3,064 *
(6)	2,104 *	2,158 *	2,186 *	2,221 *
(7)	1,526	1,362	2,201 *	2,135 *
(8)	3,142 *	3,162 *	3,246 *	3,242 *
(9)	1,747 *	1,690 *	1,578	2,684 *
(10)	1,286	1,244	<u>1,279</u>	
(11)	2,340 *	2,357 *	2,402 *	2,428 *
Wartość krytyczna $F_{0,05}$	1,65	1,64	1,64	1,63

* oznacza istotność udziału danej cechy na poziomie $\alpha = 0,05$

—— oznacza eliminację podkreślonej cechy.

nieistotnych trwała dopóty, dopóki wszystkie pozostałe cechy nie okazały się istotnymi. Są to cechy:

1) wysokość roślin w cm	F_1	= 4,447
4) plon nasion w q/ha	F_2	= 1,658
5) ciężar 1000 nasion w g	F_5	= 3,064
6) procent łuski	F_6	= 2,221
7) liczba roślin do sprzętu	F_7	= 2,135
8) procent roślin z odrostami	F_8	= 3,242
9) procent tłuszczu w całych niełupkach	F_9	= 2,648
11) procent tłuszczu w niełupkach odłuszczonych	F_{11}	= 2,428.

W tabeli 4 pokazane są kolejne etapy eliminacji poszczególnych cech. Dla cech istotnych ponownie przeprowadzamy analizę wariancji. Wartości wyznaczników macierzy $|\underline{E}|$ i $|\underline{H} + \underline{E}|$ wynoszą tym razem odpowiednio

$$|\underline{E}| = 0{,}811710125 \cdot 10^{22}$$
$$|\underline{H} + \underline{E}| = 0{,}216668331 \cdot 10^{25}.$$

Kryterium Λ Wilksa jest równe $\Lambda = 0{,}003746326$, a $F_{obl.} = 2{,}266$.

Ponieważ $F_{obl.}$ przewyższa wartość krytyczną $F_{0,05;208;683}$, więc odrzucamy hipotezę o równości rodów pod względem ośmiu cech, których udział był istotny. Biorąc pod uwagę te cechy, wyliczamy dla wszystkich par rodów odległości Mahalanobisa, których jednak z uwagi na ich liczbę tutaj nie zamieszczamy.

Obliczenia wykonane zostały według programów własnych na maszynie elektronicznej Mińsk 22.

7. Zakończenie

Autorzy pragną podziękować Doc. dr T. Calińskiemu za cenne wskazówki i pomoc oraz dr Z. Kłoczowskiemu za udostępnienie materiału, który stanowił przykład obliczeń.

Literatura

[1] Caliński, T., Teoria układów eksperymentalnych, część II. Wielozmienna analiza wariancji i pokrewne metody wielowymiarowe, Warszawa 1970.

[2] Ceranka, B., Caliński, T., Harabasz, J. S., ABS7. Analiza wariancji dla doświadczeń nieortogonalnych w klasyfikacji dwukierunkowej bez interakcji, Algorytmy biometryczne i statystyczne, z. 1, Roczn. WSR Poznań 61 (1972), str. 77-92.

[3] Harabasz, J. S., Ceranka, B,, Kaczmarek, Z., ABS 1. Odwracanie macierzy symetrycznej, Algorytmy biometryczne i statystyczne, z. 1, Roczn. WSR Poznań 61 (1972), str. 5-16.

[4] Jones, R. M,, On a property of incomplete blocks, J. R. Statist. Soc., Series B, 21 (1959), str. 172-179.

[5] Oktaba, W., Teoria układów eksperymentalnych, część I. Modele stałe, Warszawa 1970.

[6] Pearce, S. C., The efficiency of block designs in general, Biometrika 57 (1970), str. 339-346.

[7] Rees, D. H., The analysis of variance of designs with many non-orthogonal classifications, J. R. Statist. Soc., Series B, 28 (1966), str. 110-117.

[8] Tocher, K. D., The design and analysis of block experiments, J. R. Statist. Soc., Series B, 14 (1952), str. 45-91.