ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
SERIA III: MATEMATYKA STOSOWANA IV (1975)

BRONISŁAW CERANKA i ZYGMUNT KACZMAREK (Poznań)

Wielozmienna analiza kowariancji dla układów blokowych*

1. Wstęp. W niniejszej pracy przedstawiono metodę wielozmiennej analizy kowariancji dla układów blokowych. Metoda ta pozwala na przeprowadzenie analizy kowariancji dla dowolnego układu doświadczalnego o dwukierunkowej klasyfikacji (przy braku interakcji) oraz na przeprowadzenie testowania następujących hipotez: o braku regresji zmiennych głównych od zmiennych towarzyszących i o równości średnich obiektowych z uwzględnieniem regresji.

W rozdziale 2 podano ogólne założenia wielozmiennego modelu analizy kowariancji, natomiast w rozdziale 3 szczegółowo omówiono analizę kowariancji dla układów blokowych przy jednej zmiennej głównej i wielu towarzyszących. W rozdziale 4 dokonano uogólnienia na przypadek wielu zmiennych głównych. W zakończeniu przedstawiono przykład analizy danych doświadczalnych pochodzących z badań hodowlanych nad słonecznikiem.

2. Ogólny wielozmienny model liniowy analizy kowariancji. Ogólny model liniowy analizy kowariancji wielu zmiennych rozważany był przez wielu autorów (patrz na przykład [6], [5]).

Model ten można zapisać następująco:

$$Y = AT + XB + E, \tag{2.1}$$

gdzie

$$Y = [y_1, y_2, \ldots, y_p] \tag{2.2}$$

jest macierzą n p-wymiarowych obserwacji zmiennych głównych (zależnych), natomiast macierz

$$T = [\tau_1, \tau_2, \ldots, \tau_p], \tag{2.3}$$

jest macierzą $q \times p$ nieznanych parametrów, przy czym q oznacza jednakową liczbę parametrów dla i-tej zmiennej głównej ($i = 1, 2, \ldots, p$).

Macierz A typu $n \times q$ jest macierzą układu. Ponadto macierz

$$X = [x_1, x_2, \ldots, x_t] \tag{2.4}$$

jest macierzą n t-wymiarowych zaobserwowanych wartości zmiennych towarzyszących (niezależnych), natomiast macierz

*Praca wykonana w ramach problemu węzłowego 06.1.1. koordynowanego przez Instytut Matematyczny PAN.

(2.5) $$B = [\beta_1, \beta_2, \ldots, \beta_p],$$

jest macierzą $t \times p$ współczynników regresji zmiennych głównych względem zmiennych towarzyszących, przy czym β_i jest wektorem współczynników regresji i-tej zmiennej głównej względem t zmiennych towarzyszących ($i = 1, 2, \ldots, p$).

Wreszcie macierz

(2.6) $$E = [e_1, e_2, \ldots, e_p] = [e_{1.}, e_{2.}, \ldots, e_{n.}]'$$

jest macierzą odchyleń losowych (błędów).

Dla modelu (2.1) przyjmujemy następujące założenia:

(2.7) $$E(e_i) = \mathbf{0}, \qquad E(e_i e_i') = \sigma_{ii}^2 \mathbf{I}, \qquad i = 1, 2, \ldots, p,$$

gdzie $\mathbf{I}$ jest macierzą jednostkową typu $n \times n$.

Obok założeń (2.7) odnoszących się do każdej kolumny macierzy $\mathbf{E}$, przyjmujemy dla każdego wiersza tej macierzy założenie, że

(2.8) $$E(e_{h.} e_{h.}') = \Sigma = (\sigma_{ii'}), \qquad i, i' = 1, 2, \ldots, p; \ h = 1, 2, \ldots, n,$$

to znaczy założenie, że macierz kowariancji błędów doświadczalnych dla każdej jednostki doświadczalnej jest jednakowa, równa Σ.

Przy omawianiu testów i przedziałów ufności będziemy nadto zakładać, że każdy wiersz macierzy $\mathbf{E}$ ma ten sam wielowymiarowy rozkład normalny $N(\mathbf{0}, \Sigma)$.

Wobec (2.7), wiersze macierzy $\mathbf{E}$ będą wówczas statystycznie niezależne.

3. Jednozmienna analiza kowariancji dla układów blokowych. Model liniowy (2.1) dla jednej zmiennej głównej ($p = 1$) i t zmiennych towarzyszących przyjmuje postać

(3.1) $$y = A\tau + X\beta + e = [A, X][\tau', \beta']' + e.$$

Układ równań normalnych dla modelu (3.1) jest postaci

(3.2) $$\begin{bmatrix} A'A & A'X \\ X'A & X'X \end{bmatrix} \begin{bmatrix} \tau \\ \beta \end{bmatrix} = \begin{bmatrix} A' \\ X' \end{bmatrix} y.$$

W zastosowaniu do układów blokowych model (3.1) wygodnie jest zapisać następująco:

(3.3) $$y = [\mathbf{1}, A_1, A_2][\mu, \xi', \gamma']' + [X - \mathbf{1}\mathbf{1}' X/n]\beta + e =$$
$$= [\mathbf{1}, A_1, A_2, X - \mathbf{1}\mathbf{1}' X/n][\mu, \xi', \gamma', \beta']' + e,$$

gdzie y jest wektorem obserwacji zmiennej zależnej typu $n \times 1$, $A = [\mathbf{1}, A_1, A_2]$ jest macierzą układu typu $n \times (1 + b + a)$ złożoną z wektora $\mathbf{1}$ typu $n \times 1$ oraz macierzy układu A_1, typu $n \times b$, dla pierwszej klasyfikacji (bloków) i macierzy układu A_2, typu $n \times a$, dla drugiej klasyfikacji (obiektów), $\tau = [\mu, \xi', \gamma']'$ jest wektorem parametrów typu $(1 + b + a) \times 1$ złożonym z parametru wspólnego μ (skalara), wektora parametrów blokowych ξ typu $b \times 1$, oraz wektora parametrów obiektowych γ typu $a \times 1$. Ponadto wektor $\beta = [\beta_1, \beta_2, \ldots, \beta_t]'$ o wymiarach $t \times 1$ jest wektorem współczynników regresji zmiennej głównej względem t zmiennych towarzyszących, natomiast macierz X została określona w (2.4). Wprowadzenie do modelu (3.3) macierzy $[X - \mathbf{1}\,\mathbf{1}' X/n]$ zamiast macierzy X uzasadnio-

ne jest tym, że zwykle w interpretacji wyników doświadczenia pożądana jest eliminacja wpływu odchyleń wartości zmiennych towarzyszących od ich wartości średnich, a nie eliminacja wpływu samych wartości tych zmiennych.

Wektor $\boldsymbol{e} = [e_1, e_2, \ldots, e_N]'$ jest wektorem błędów losowych. Założenia odnośnie modelu (3.3) są takie same jak dla modelu (2.1).

W dalszych rozważaniach przydatny będzie następujący zapis wielkości charakteryzujących układ blokowy. Liczby replikacji poszczególnych obiektów zapiszemy w postaci wektora $\boldsymbol{r} = [r_1, r_2, \ldots, r_a]'$ gdzie r_i jest liczbą replikacji i-tego obiektu a wielkości bloków przedstawimy wektorem $\boldsymbol{k} = [k_1, k_2, \ldots, k_b]'$, gdzie k_j jest liczbą jednostek doświadczalnych w j-tym bloku ($i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$). Oznaczmy macierz diagonalną o elementach diagonalnych równych $r_1, r_2, \ldots, r_a$ przez $\boldsymbol{R}^\delta$ oraz macierz diagonalną o elementach diagonalnych równych $k_1, k_2, \ldots, k_b$ przez $\boldsymbol{K}^\delta$. Przez $\boldsymbol{R}^{-\delta}$ rozumieć będziemy macierz odwrotną do $\boldsymbol{R}^\delta$ a przez $\boldsymbol{K}^{-\delta}$ macierz odwrotną do $\boldsymbol{K}^\delta$.

Układ blokowy możemy opisać jednoznacznie za pomocą macierzy incydencji $\boldsymbol{N}$, której wiersze odpowiadają poszczególnym obiektom a kolumny poszczególnym blokom. Zatem macierz ta jest typu $a \times b$. Element n_{ij} tej macierzy oznacza liczbę jednostek należących równocześnie do i-tego obiektu oraz j-tego bloku.

Dodatkowo oznaczmy wektor sum obiektowych dla zmiennej głównej przez $\boldsymbol{v}_y = [v_{1y}, v_{2y}, \ldots, v_{ay}]'$, gdzie v_{iy} jest sumą obserwacji zmiennej głównej dotyczących i-tego obiektu, oraz wektor sum blokowych dla zmiennej głównej $\boldsymbol{z}_y = [z_{1y}, z_{2y}, \ldots, z_{by}]'$, gdzie z_{jy} jest sumą obserwacji zmiennej głównej dla j-tego bloku. Sumę obserwacji z wszystkich n jednostek dla zmiennej głównej oznaczmy przez w_y. Ponadto wprowadźmy jeszcze wektor sum $\boldsymbol{u}_y = [w_y, \boldsymbol{v}_y, \boldsymbol{z}_y]'$ dla zmiennej głównej.

Analogicznie oznaczmy sumy obserwacji dla zmiennych towarzyszących. Niech $\boldsymbol{V}_X = [\boldsymbol{v}_{1X}, \boldsymbol{v}_{2X}, \ldots, \boldsymbol{v}_{tX}]$ będzie macierzą $a \times t$ sum obiektowych dla zmiennych towarzyszących, $\boldsymbol{Z}_X = [\boldsymbol{Z}_{1X}, \boldsymbol{Z}_{2X}, \ldots, \boldsymbol{Z}_{tX}]$ macierzą $b \times t$ sum blokowych dla zmiennych towarzyszących, oraz $\boldsymbol{w}_X = [w_{1X}, w_{2X}, \ldots, w_{tX}]'$ wektorem sum wszystkich obserwacji dla poszczególnych zmiennych towarzyszących. Przez $\boldsymbol{U}_X = [\boldsymbol{w}'_X, \boldsymbol{V}'_X, \boldsymbol{Z}'_X]'$ oznaczmy macierz sum dla zmiennych towarzyszących.

Między wprowadzonymi wielkościami można zauważyć następujące związki:

$$\boldsymbol{N}\boldsymbol{1} = \boldsymbol{r} = \boldsymbol{A}'_2\boldsymbol{1}, \qquad \boldsymbol{N}'\boldsymbol{1} = \boldsymbol{k} = \boldsymbol{A}'_1\boldsymbol{1},$$

$$\boldsymbol{r}'\boldsymbol{1} = n = \boldsymbol{k}'\boldsymbol{1}, \qquad \boldsymbol{N} = \boldsymbol{A}'_2\boldsymbol{A}_1, \qquad \boldsymbol{K}^\delta = \boldsymbol{A}'_1\boldsymbol{A}_1, \qquad \boldsymbol{R}^\delta = \boldsymbol{A}'_2\boldsymbol{A}_2,$$

$$\boldsymbol{z}_y = \boldsymbol{A}'_1\boldsymbol{y}, \qquad \boldsymbol{Z}_X = \boldsymbol{A}'_1\boldsymbol{X}, \qquad \boldsymbol{v}_y = \boldsymbol{A}'_2\boldsymbol{y}, \qquad \boldsymbol{V}_X = \boldsymbol{A}'_2\boldsymbol{X},$$

$$\boldsymbol{1}'\boldsymbol{z}_y = w_y = \boldsymbol{1}'\boldsymbol{v}_y, \qquad \boldsymbol{1}'\boldsymbol{Z}_X = \boldsymbol{w}_X = \boldsymbol{1}'\boldsymbol{V}_X, \qquad \boldsymbol{u}_y = \boldsymbol{A}'\boldsymbol{y}, \qquad \boldsymbol{U}_X = \boldsymbol{A}'\boldsymbol{X},$$

gdzie $\boldsymbol{1}$ oznacza zawsze odpowiedni wektor złożony z samych jedynek.

Wprowadźmy jeszcze następujące macierze i wektory:

$$
\begin{aligned}
d_y &= v_y - N K^{-\delta} z_y, \\
D_X &= V_X - N K^{-\delta} Z_X, \\
c_y &= z_y - (w_y/n) k, \\
C_X &= Z_X - k (w'_X/n), \\
N_0 &= N - r k'/n, \\
\Omega^{-1} &= R^{\delta} - N K^{-\delta} N' + r r'/n.
\end{aligned} \tag{3.4}
$$

Zauważmy, że zachodzą następujące związki: $1'd_y = 1'c_y = 0$, $1'D_X = 1'C_X = 0'$, $\Omega^{-1} 1 = r$, $\Omega r = 1$, $N_0 1 = 0$.

Przy powyższych oznaczeniach równania normalne (3.2) w zastosowaniu do układów blokowych przyjmują postać

$$
\begin{bmatrix} n & k' & r' & 0' \\ k & K^{\delta} & N' & C_X \\ r & N & R^{\delta} & V_X - r(w'_X/n) \\ 0 & C'_X & V'_X - (w_X/n) r' & X'X - w_X w'_X/n \end{bmatrix} \begin{bmatrix} \mu \\ \xi \\ \gamma \\ \beta \end{bmatrix} = \begin{bmatrix} w_y \\ z_y \\ v_y \\ X'y - w_X w_y/n \end{bmatrix} \tag{3.5}
$$

Rozważany model (3.3) ograniczymy do przypadku, gdy macierz układu A jest rzędu $b + a - 1$. Przy tym założeniu oraz przy warunkach ubocznych $k' \xi = r' \gamma$, parametry μ, ξ i γ są estymowalne ([7]).

Zatem, po rozwiązaniu układu równań normalnych (3.5) otrzymujemy estymatory najmniejszych kwadratów dla układów blokowych w postaci wektora

$$
\begin{bmatrix} \hat{\mu} \\ \hat{\xi} \\ \hat{\gamma} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} w_y/n \\ K^{-\delta} (c_y - C_X \hat{\beta} - N'_0 \gamma) \\ \Omega(d_y - D_X \hat{\beta}) \\ S^{-1}_{e;X,X} S_{e;X,y} \end{bmatrix}, \tag{3.6}
$$

gdzie

$$
S_{e;X,X} = X'X - w_X w'_X/n - C'_X K^{-\delta} C_X - D'_X \Omega D_X \tag{3.7}
$$

jest macierzą sum kwadratów i iloczynów dla błędu zmiennych towarzyszących, natomiast

$$
S_{e;X,y} = X'y - w_X w_y/n - C'_X K^{-\delta} c_y - D'_X \Omega d_y \tag{3.8}
$$

jest wektorem sum iloczynów dla błędu zmiennych towarzyszących i zmiennej głównej. Do wyniku (3.6) można dojść na tej samej drodze jak w przypadku jednej zmiennej głównej i jednej zmiennej towarzyszącej (patrz [3]).

Stąd wektor tak zwanych poprawionych średnich obiektowych dla zmiennej głównej z uwzględnieniem układu doświadczalnego i regresji przyjmuje postać

(3.9) $$\bar{y}_{\text{popr}} = \hat{\mu}\mathbf{1} + \hat{\gamma} = (w_y/n)\,\mathbf{1} + \boldsymbol{\Omega}(\boldsymbol{d}_y - \boldsymbol{D}_X\, \boldsymbol{S}^{-1}_{e;X,X}\, \boldsymbol{S}_{e;X,y}).$$

Macierzą kowariancji tego wektora jest

(3.10) $$\Sigma\bar{y}_{\text{popr}} = (n-a-b-t+1)^{-1}\, SS_e\, \boldsymbol{\Phi},$$

gdzie

(3.11) $$SS_e = SS_{e;y} - \boldsymbol{S}'_{e;X,y}\, \boldsymbol{S}^{-1}_{e;X,X}\, \boldsymbol{S}_{e;X,y},$$

jest sumą kwadratów dla błędu z uwzględnieniem regresji, przy czym $SS_{e;y}$ oznacza sumę kwadratów dla błędu zmiennej głównej, a macierz $\boldsymbol{\Phi}$ ze wzoru (3.10) jest zdefiniowana jako

(3.12) $$\boldsymbol{\Phi} = \boldsymbol{\Omega} + \boldsymbol{\Omega}\,\boldsymbol{D}_X\, \boldsymbol{S}^{-1}_{e;X,X}\, \boldsymbol{D}'_X\, \boldsymbol{\Omega}.$$

Jeżeli interesuje nas kontrast $\psi = \boldsymbol{s}'\,\boldsymbol{\gamma}$, gdzie $\boldsymbol{s}$ jest wektorem odpowiednich współczynników, wtedy błąd standardowy tego kontrastu jest postaci

(3.13) $$s_\psi = \sqrt{(n-a-b-t+1)^{-1}\, SS_e\, \boldsymbol{s}'\,\boldsymbol{\Omega}\,\boldsymbol{s}}.$$

T e s t o w a n i e h i p o t e z y H_β : $\boldsymbol{\beta} = \mathbf{0}$. Uwzględnienie wpływu zmiennych towarzyszących jest zbędne, jeżeli $\boldsymbol{\beta} = \mathbf{0}$. Chcąc sprawdzić tę hipotezę, musimy znaleźć odpowiednią sumę kwadratów do jej testowania. Sumę tę uzyskujemy jako minimum przy warunku $\boldsymbol{\beta} = \mathbf{0}$ odpowiedniej formy kwadratowej, to znaczy jako

(3.14) $$\min_{\boldsymbol{\beta}=\mathbf{0}}\, [\boldsymbol{y} - \boldsymbol{A}\,\boldsymbol{\tau} - \boldsymbol{X}\boldsymbol{\beta}]'[\boldsymbol{y} - \boldsymbol{A}\,\boldsymbol{\tau} - \boldsymbol{X}\boldsymbol{\beta}] = SS_{e;y}.$$

Stąd funkcja testowa dla testowania hipotezy $\boldsymbol{\beta} = \mathbf{0}$ jest postaci ([6]),

(3.15) $$F = \frac{n-a-b-t+1}{t} \cdot \frac{SS_\beta}{SS_e},$$

gdzie

(3.16) $$SS_\beta = \boldsymbol{S}'_{e;X,y}\, \boldsymbol{S}^{-1}_{e;X,X}\, \boldsymbol{S}_{e;X,y},$$

natomiast suma kwadratów SS_e jest taka jak zdefiniowano w (3.11). Funkcja (3.15) ma przy założeniu prawdziwości hipotezy H_β rozkład F z t i $n-a-b-t+1$ stopniami swobody.

T e s t o w a n i e h i p o t e z y H_γ : $\gamma = 0$. Dla sprawdzenia hipotezy $H_\gamma : \boldsymbol{\gamma} = \mathbf{0}$ należy zminimalizować formę kwadratową ze wzoru (3.14) przy warunku $\boldsymbol{\gamma} = \mathbf{0}$. Otrzymana wówczas suma kwadratów dla testowania hipotezy $\boldsymbol{\gamma} = \mathbf{0}$ jest postaci ([3]):

(3.17) $$SS_\gamma = \boldsymbol{d}'_y\, \boldsymbol{\Omega}\,\boldsymbol{d}_y + SS_{e;y} - \\ - (\boldsymbol{d}'_y\, \boldsymbol{\Omega}\boldsymbol{D}_X + \boldsymbol{S}'_{e;X,y})\,(\boldsymbol{D}'_X\, \boldsymbol{\Omega}\boldsymbol{D}_X + \boldsymbol{S}_{e;X,X})^{-1}\, (\boldsymbol{d}'_y\boldsymbol{\Omega}\,\boldsymbol{D}_X + \boldsymbol{S}'_{e;X,y})'.$$

Stąd odpowiednią funkcją testową jest

$$F = \frac{n - a - b - t + 1}{a - 1} \cdot \frac{SS_\gamma - SS_e}{SS_e}. \tag{3.18}$$

Funkcja (3.18) ma przy założeniu prawdziwości hipotezy H_γ rozkład F z $(a - 1)$ i $(n - a - b - t + 1)$ stopniami swobody.

4. Wielozmienna analiza kowariancji dla układów blokowych. Wielozmienny model analizy kowariancji (2.1) w zastosowaniu do układów blokowych przyjmuje postać

$$\mathbf{Y} = [\mathbf{1}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{X} - \mathbf{1}\mathbf{1}'\,\mathbf{X}/n]\,[\boldsymbol{\mu}, \boldsymbol{\Xi}', \boldsymbol{\Gamma}', \mathbf{B}']' + \mathbf{E}. \tag{4.1}$$

W porównaniu z modelem (3.3) wektor $\mathbf{y}$ został zastąpiony macierzą $\mathbf{Y}$ (wzór (2.2)), wektor $\boldsymbol{\tau} = [\mu, \boldsymbol{\xi}', \boldsymbol{\gamma}']'$ został zastąpiony macierzą $\mathbf{T} = [\boldsymbol{\mu}, \boldsymbol{\Xi}', \boldsymbol{\Gamma}']'$ typu $(1 + b + a) \times p$, gdzie $\boldsymbol{\mu}, \boldsymbol{\Xi}$ i $\boldsymbol{\Gamma}$ są uogólnieniami parametrów μ, $\boldsymbol{\xi}$ i $\boldsymbol{\gamma}$ z modelu (3.3) na p zmiennych. Również wektor współczynników regresji $\boldsymbol{\beta}$ został zastąpiony macierzą $\mathbf{B}$ podaną w (2.5), natomiast wektor $\mathbf{e}$ zastąpiono macierzą $\mathbf{E}$ (wzór (2.6)), z założeniami (2.7) i (2.8). Macierz układu $\mathbf{A}$ oraz macierz zmiennych towarzyszących $\mathbf{X}$ pozostają bez zmian. Wiadomo, że każde zagadnienie testowania hipotezy jednozmiennej można uogólnić na przypadek wielu zmiennych, zastępując sumy kwadratów dla błędu i hipotezy przez ich uogólnienia macierzowe i następnie stosując odpowiedni test wielozmienny (patrz [1]).

W związku z powyższym wypiszmy dla modelu (4.1) interesujące nas macierze sum kwadratów i iloczynów. Macierz sum kwadratów i iloczynów dla błędu zmiennych towarzyszących $\mathbf{S}_{e;X,X}$ pozostaje jak w (3.7).

Macierz sum iloczynów dla błędu zmiennych towarzyszących i zmiennych głównych jest postaci

$$\mathbf{S}_{e;X,Y} = \mathbf{X}'\mathbf{Y} - \mathbf{w}_X\mathbf{w}_Y'/n - \mathbf{C}_X'\mathbf{K}^{-\delta}\,\mathbf{C}_Y - \mathbf{D}_X'\boldsymbol{\Omega}\mathbf{D}_Y, \tag{4.2}$$

gdzie $\mathbf{w}_Y$, $\mathbf{C}_Y$ i $\mathbf{D}_Y$ są odpowiednimi uogólnieniami na p zmiennych $\mathbf{w}_y$, $\mathbf{c}_y$ i $\mathbf{d}_y$ (patrz (3.4)).

Macierz sum kwadratów i iloczynów dla błędu z uwzględnieniem regresji, będąca uogólnieniem odpowiedniej sumy kwadratów dla błędu (3.11) jest równa

$$S_e = \mathbf{S}_{e;Y,Y} - \mathbf{S}_{e;X,Y}'\,\mathbf{S}_{e;X,X}^{-1}\,\mathbf{S}_{e;X,Y}, \tag{4.3}$$

gdzie $\mathbf{S}_{e;Y,Y}$ oraz $\mathbf{S}_{e;X,Y}$ są uogólnieniami na p zmiennych odpowiednio sumy kwadratów $SS_{e;y}$ i wektora $\mathbf{S}_{e;X,y}$. Macierz sum kwadratów i iloczynów dla testowania hipotezy H_B: $\mathbf{B} = \mathbf{0}$, będąca uogólnieniem sumy kwadratów (3.16), jest postaci:

$$\mathbf{S}_B = \mathbf{S}_{e;X,Y}'\,\mathbf{S}_{e;X,X}^{-1}\,\mathbf{S}_{e;X,Y}, \tag{4.4}$$

natomiast macierz sum kwadratów i iloczynów w regresji dla hipotezy $H_\Gamma : \boldsymbol{\Gamma} = \mathbf{0}$, będąca uogólnieniem sumy kwadratów dla hipotezy w regresji (3.17), jest postaci

$$\mathbf{S}_\Gamma = \mathbf{D}_Y'\boldsymbol{\Omega}\mathbf{D}_Y + \mathbf{S}_{e;Y,Y} - (\mathbf{D}_Y'\boldsymbol{\Omega}\mathbf{D}_X + \mathbf{S}_{e;X,Y}')(\mathbf{D}_X'\boldsymbol{\Omega}\mathbf{D}_X + \mathbf{S}_{e;X,X})^{-1}(\mathbf{D}_Y'\boldsymbol{\Omega}\mathbf{D}_X + \mathbf{S}_{e;X,Y}')'. \tag{4.5}$$

Przy pomocy określonych w (4.3), (4.4) i (4.5) macierzy sum kwadratów i iloczynów w regresji, można dokonać testowania hipotez H_B i H_Γ.

Hipotezę H_B: $\boldsymbol{B} = \boldsymbol{0}$ testujemy przy pomocy statystyki

$$F = \frac{mf - 2l}{pt} \cdot \frac{1 - \Lambda_B^{1/f}}{\Lambda_B^{1/f}}, \tag{4.6}$$

gdzie

$$\Lambda_B = \frac{|\boldsymbol{S}_e|}{|\boldsymbol{S}_e + \boldsymbol{S}_B|} = \frac{|\boldsymbol{S}_e|}{|\boldsymbol{S}_{e;Y,Y}|},$$

natomiast

$$m = n - a - b - t + 1 - \frac{p - t + 1}{2}; \quad f = \sqrt{\frac{p^2 t^2 - 4}{p^2 + t^2 - 5}} \text{ gdy } p^2 + t^2 \neq 5;$$

$$l = \frac{pt - 2}{4}.$$

W przypadku, gdy co najmniej jedna z liczb p, t jest równa 1 lub 2, to f przyjmuje postać $f = \min(p, t)$ i statystyka F określona w (4.6) ma pod hipotezą H_B dokładny rozkład F. W pozostałych przypadkach funkcja F ma przy prawdziwości hipotezy H_B w przybliżeniu rozkład F z pt i $mf - 2l$ stopniami swobody. Oznacza to, że hipotezę tę odrzucamy wtedy i tylko wtedy, gdy $F > F_{\alpha;pt,mf-2l}$.

Hipotezę H_Γ: $\boldsymbol{\Gamma} = \boldsymbol{0}$ testujemy przy pomocy statystyki

$$F = \frac{mf - 2l}{p(a-1)} \cdot \frac{1 - \Lambda_\Gamma^{1/f}}{\Lambda_\Gamma^{1/f}}, \tag{4.7}$$

gdzie

$$\Lambda_\Gamma = \frac{|\boldsymbol{S}_e|}{|\boldsymbol{S}_e + \boldsymbol{S}_\Gamma|},$$

natomiast

$$m = n - a - b - t + 1 - \frac{p - a + 2}{2}; \quad f = \sqrt{\frac{p^2 (a-1)^2 - 4}{p^2 + (a-1)^2 - 5}} \text{ gdy } p^2 + (a-1)^2 \neq 5;$$

$$l = \frac{p(a-1) - 2}{4}.$$

Statystyka (4.7) ma przy prawdziwości hipotezy H_Γ w przybliżeniu rozkład F z $p(a-1)$ i $mf - 2l$ stopniami swobody.

Oszacowań macierzy $\boldsymbol{B}$ nieznanych współczynników regresji zmiennych głównych względem zmiennych towarzyszących oraz macierzy parametrów $\boldsymbol{\Gamma}$ dokonuje się przez zastąpienie we wzorze (3.6) wektora $\boldsymbol{y}$ macierzą $\boldsymbol{Y}$. Są one odpowiednio postaci:

(4.8) $$\hat{B} = S^{-1}_{e;X,X} S_{e;X,Y},$$

(4.9) $$\hat{\Gamma} = \Omega(D_Y - D_X \hat{B}).$$

Stąd macierz poprawionych średnich obiektowych względem układu doświadczalnego i regresji, będąca uogólnieniem wzoru (3.9), jest równa

(4.10) $$\overline{Y}_{\text{popr}} = (w_Y/n)\,1 + \Omega(D_Y - D_X S^{-1}_{e;X,X} S_{e;X,Y}).$$

Macierz kowariancji każdej kolumny macierzy (4.10) jest określona wzorem (3.10).

Porównanie między sobą obiektów przeprowadza się podobnie jak w wielozmiennej analizie wariancji ([4]).

5. Przykład. Opisaną wyżej metodę zilustrujemy przykładem analizy doświadczenia ze słonecznikiem przeprowadzonego w IHAR Borowo.

Doświadczenie to, założone w rozszerzonym afinicznie rozkładalnym układzie blokowym, zostało opracowane za pomocą wielozmiennej analizy wariancji w pracy [4]. W wyniku zastosowania metody wyboru zmiennych przedstawionej w pracy [2], z rozpatrywanych pierwotnie 11 cech, do analizy kowariancji jako zmienne główne zakwalifikowano:

1. wysokość roślin w cm,
2. plon nasion w q/ha,
3. ciężar 1000 nasion w g,
4. procent łuski,
5. procent tłuszczu w całych niełupkach,
6. plon oleju w q/ha,
7. procent tłuszczu w niełupkach odłuszczonych,

a jako zmienne towarzyszące

1. średnicę tarczy w cm,
2. liczbę roślin do sprzętu.

Z dalszej analizy odrzucone zostały zatem cechy: plon nasion w g/roślinę oraz procent roślin z odrostami.

Stosując dla tak wybranych cech wielozmienną analizę kowariancji, uzyskujemy możliwość porównania rodów po wyeliminowaniu wpływu cech towarzyszących, to jest zróżnicowanych wartości średnicy tarczy i liczby roślin do sprzętu. Oczywiście wybór właśnie tych cech jako cech towarzyszących może być dyskusyjny i w naszym przykładzie został dokonany przede wszystkim dla pokazania metody obliczeń. Inne względy odgrywają tutaj rolę podrzędną. Jeśli badacza interesuje porównanie rodów bez eliminacji wpływu zmiennych towarzyszących, wystarczy przeprowadzić wielozmienną analizę wariancji (patrz przykład w pracy [4]).

Przy pomocy wzorów (4.3), (4.4) i (4.5) obliczone zostały macierze sum kwadratów i iloczynów dla błędu w regresji oraz macierze do testowania hipotez H_B i H_Γ. Macierze te zostały zestawione w tabelach 1, 2 i 3.

Aby przeprowadzić testowanie istotności współczynników regresji obu zmiennych towarzyszących względem zmiennych głównych, to znaczy testowanie hipotezy H_B : $B = 0$, korzystamy ze statystyki (4.6). Wartość tej statystyki jest równa $F = 6.144$. Ponieważ jest ona większa od wartości krytycznej $F_{0.01;\,14,\,172} = 2.19$, więc badane zmienne towarzyszące istotnie wpływają na zmienne główne.

TABELA 1

Macierz sum kwadratów i iloczynów dla błędu w regresji S_{ε}

1733.45	237.59	−383.79	−2.06	3.39	113.55	75.78
237.59	596.70	306.69	14.47	−62.18	276.20	−3.19
−383.79	306.69	2603.96	94.61	−226.98	72.90	−121.20
−2.06	14.47	94.61	170.23	−14.99	3.47	6.46
3.39	−62.18	−226.98	−14.99	96.82	−7.58	56.85
113.55	276.20	72.90	3.47	−7.58	137.78	11.22
75.78	−3.19	−121.20	6.46	56.85	11.22	88.24

TABELA 2

Macierz sum kwadratów i iloczynów dla testowania hipotezy H_B

113.41	161.64	139.85	5.96	−27.97	74.66	9.27
161.64	400.13	55.60	16.91	254.56	189.37	14.24
139.85	55.60	294.11	0.21	4.17	21.83	10.56
5.96	16.91	0.21	0.73	24.99	8.04	0.54
−27.97	254.56	4.17	24.99	213.11	102.54	−57.48
74.66	189.37	21.83	8.04	102.54	89.69	6.61
9.27	14.24	10.56	0.54	−57.48	6.61	0.77

TABELA 3

Macierz sum kwadratów i iloczynów dla testowania hipotezy H_{Γ}

2363.51	33.32	−1238.40	329.60	−343.01	−62,90	−158.82
33.32	187.58	−493.66	−42.05	57.03	101.89	50.03
−1238.40	−493.66	2202.32	−132.08	−219.30	−279.12	−211.55
329.60	−42.05	−132.08	169.30	−103.45	−47.65	−50.29
−343.01	57.03	−219.30	−103.45	195.35	77.76	129.15
−62.90	101.89	−279.12	−47.65	77.76	69.00	57.90
−158.82	50.03	−211.55	−50.29	129.15	57.90	124.92

TABELA 4

Macierz współczynników regresji zmiennych towarzyszących względem zmiennych głównych

1.0158	1.6119	1.1134	0.0615	0.0025	0.7490	0.0840
0.1058	0.9144	−0.5160	0.0435	0.0419	0.4429	0.0133

Macierz oszacowanych na podstawie (4.8) współczynników regresji przedstawia tabela 4. Wobec istotności regresji, celowe jest testowanie hipotezy H_{Γ}: $\boldsymbol{\Gamma} = \mathbf{0}$, dotyczącej równości średnich obiektowych z uwzględnieniem regresji. Korzystając ze statystyki (4.7) uzyskujemy wartość $F = 14.735$. Przewyższa ona znaczenie wartość krytyczną $F_{0.01;\,182,\,596} = 1.32$; tym samym hipotezę H_{Γ} odrzucamy.

Macierz poprawionych średnich obiektowych wyznaczona na podstawie (4.10) jest zestawiona w tabeli 5.

TABELA 5

Średnie poprawione siedmiu cech rodów słonecznika

Rody \ Cechy	Wysokość roślin w cm	Plon nasion w q/roślinę	Ciężar 1000 nasion w g	Procent łuski	Procent tłuszczu w całych niełupkach	Plon oleju w q/ha	Procent tłuszczu w niełupkach odłuszczonych
1	71.39	24.23	55.56	25.97	49.19	11.86	62.56
2	89.59	20.72	52.12	31.36	42.98	8.70	59.03
3	68.38	22.99	58.91	27.58	48.82	11.21	63.50
4	73.88	22.74	59.64	29.95	46.46	10.34	61.25
5	70.23	23.01	59.02	26.86	47.61	10.94	63.90
6	67.09	21.91	63.48	27.58	47.07	10.34	61.11
7	61.51	19.38	63.24	27.25	47.39	9.12	61.27
8	76.33	23.23	58.37	29.23	47.16	10.99	62.11
9	72.51	23.72	70.52	25.80	47.10	11.15	61.25
10	72.60	23.34	62.85	27.08	49.95	12.13	64.04
11	73.26	23.45	64.99	28.54	46.13	10.67	61.55
12	67.57	20.05	59.89	26.05	48.26	9.64	62.24
13	75.01	23.58	53.85	28.38	49.38	11.51	63.76
14	76.26	22.07	69.10	30.04	46.16	10.06	62.55
15	64.53	24.39	63.87	26.55	48.08	11.65	62.63
16	63.91	22.56	55.50	26.15	49.22	11.05	63.89
17	68.68	26.63	58.44	27.78	48.47	12.83	63.38
18	67.16	22.12	50.87	29.24	48.64	10.77	62.93
19	69.83	22.16	55.63	27.77	48.08	10.73	63.04
20	68.33	23.27	55.83	28.10	49.95	11.61	63.97
21	61.46	23.65	60.42	27.21	48.67	11.34	62.75
22	69.62	21.06	50.98	27.59	48.22	10.16	63.45
23	61.82	22.04	72.32	28.50	46.00	10.19	61.40
24	62.08	21.43	63.57	29.32	47.67	9.97	61.96
25	73.95	25.89	63.41	26.40	47.50	12.17	62.54
26	67.69	23.27	64.14	27.11	47.88	11.10	62.31
27	73.00	22.90	61.49	28.45	47.24	10.78	62.82

(Ze względu na objętość, nie zamieszczamy w pracy tablicy odległości między wszystkimi rodami).

Porównanie wyników uzyskanych z analizy kowariancji z wynikami uzyskanymi z analizy wariancji (patrz [4]), daje możliwość zorientowania się, czy różnice między rodami spowodowane są tylko niejednakową średnicą tarczy słonecznika i różną liczbą roślin, czy też innymi względami. Weźmy dla przykładu pod uwagę dwa rody: 15 i 25. Przeprowadzona w cytowanej pracy [4] wielozmienna analiza wariancji wykazała wysoce istotne zróżnicowanie obu rodów (odległość Mahalanobisa między nimi wynosiła 4.165).

Porównując teraz poprawione w wyniku analizy kowariancji średnie wymienionych rodów (ród 15 cechuje nieco niższa od przeciętnej średnica tarczy (= 13.65) i bardzo duża liczba roślin do sprzętu (= 17.16); natomiast ród 25 – duża średnica tarczy (= 15.18) i także duża liczba roślin do sprzętu (= 16.25)), można zauważyć stosunkowo niewielkie odchylenia od pierwotnych wartości cech dla pierwszego rodu i znaczne zmniejszenie wartości poszczególnych cech drugiego rodu. Między poprawionymi średnimi większości cech

głównych obu rodów różnice są obecnie niewielkie. Świadczy choćby o tym odległość Mahalanobisa między rodem 15 i 25, która w wyniku zastosowania analizy kowariancji jest nieistotna i wynosi 2.527.

Jak z powyższego wynika, te same rody, które różniły się w wyniku przeprowadzonej analizy wariancji, nie różnią się istotnie w wyniku zastosowania analizy kowariancji. Łatwo zauważyć, że na tę sytuację wpływa przede wszystkim taki czy inny wybór cech towarzyszących.

Kwalifikowanie cech do grupy cech towarzyszących należy zatem przeprowadzać bardzo ostrożnie. Badacz, traktujący pewne cechy jako towarzyszące powinien zdawać sobie sprawę z konsekwencji takiego postępowania i szczególnie uważać przy interpretacji wyników.

Bibliografia

[1] T. C a l i ń s k i, Z. K a c z m a r e k, *Metody kompleksowej analizy doświadczenia wielocechowego,* III Colloqium Agrobiometryczne, Warszawa 1973.

[2] T. C a l i ń s k i, A. D y c z k o w s k i, Z. K a c z m a r e k, *Wybór zmiennych do analizy doświadczenia wielocechowego,* Roczniki A.R. Poznań, Algorytmy biometryczne i statystyczne 3 (1974), str. 117–128.

[3] B. C e r a n k a, *Analiza kowariancji dla doświadczalnych układów blokowych,* Wydział V Nauk Rolniczych i Leśnych, PAN, Warszawa 1972.

[4] B. C e r a n k a, Z. K a c z m a r e k, *Zastosowanie wielozmiennej analizy wariancji dla nieortogonalnych układów blokowych,* Listy Biometryczne 34–36 (1972).

[5] D. F. M o r r i s o n, *Multivariate Statistical Methods,* New York 1967.

[6] C. R. R a o, *Linear Statistical Inference and Its Applications,* New York 1965.

[7] H. S c h e f f é, *The Analysis of Variance,* New York 1959.

ZAKŁAD METOD MATEMATYCZNYCH I STATYSTYCZNYCH AR W POZNANIU
ZAKŁAD GENETYKI ROŚLIN PAN W POZNANIU