

Tadeusz C a l i ń s k i, Stanisław C z a j k a i  
Zygmunt K a c z m a r e k (Poznań)

WIELOZMIENNA ANALIZA WARIANCJI  
W ZASTOSOWANIU DO WYNIKÓW SERII DOŚWIADCZEŃ <sup>1)</sup>

1. Wstęp i streszczenie

Większość badawczych zagadnień rolniczych rozwiązywana jest przez prowadzenie doświadczeń powtarzanych w czasie i przestrzeni. Jest to w szczególności typowe dla doświadczeń z porównywaniem odmian roślin, czyli dla tzw. doświadczeń odmianowych. W zależności od tego jak traktuje się wybór lat i miejscowości można budować model analizy wariancji stały lub mieszany. Na ogół w zastosowaniach przyjmuje się model mieszany, który pozwala na bardziej ogólne wnioskowanie.

W pracy tej opracowano model mieszany analizy wariancji dla przypadku ortogonalnego. Model ten jest skonstruowany dla zmiennej wektorowej o dwóch kierunkach klasyfikacji krzyżowej A i B, przy czym czynnik A (miejscowości), czynnik B (lata) oraz interakcja  $A \times B$  (miejscowości  $\times$  lata) są traktowane jako losowe, natomiast składowe wektorów (odmiany) jako stałe.

Do opracowania tego modelu wykorzystano teorię wielozmiennej analizy wariancji.

W rozdziale drugim sformułowano wektorowy model mieszany analizy wariancji dla klasyfikacji dwukierunkowej, gdzie kierunkami klasyfikacji są miejscowości i lata, natomiast zmiennymi są odmiany.

---

<sup>1)</sup> Praca wykonana w ramach problemu węzłowego 06.1.1 koordynowanego przez Instytut Matematyczny PAN.

W rozdziale trzecim podano macierze sum kwadratów i iloczynów oraz ich rozkłady.

W następnym rozdziale omówiono testowanie hipotez, przy czym zastosowano tutaj testy: F i  $\Lambda$  Wilksa. Do testowania różnic między odmianami zbudowano przybliżoną statystykę  $T^2$  Hotellinga oraz podano sposób sprawdzenia tej hipotezy również dla poszczególnego przypadku nieistotności interakcji odmian z miejscowościami, bądź odmian z latami.

W rozdziale piątym badany jest przypadek, gdy różnice między odmianami są istotne. Ponieważ interesują nas te różnice między odmianami, które miały główny wpływ na odrzucenie hipotezy o braku różnic, wobec tego zbudowano wektor maksymalnego kontrastu między odmianami i podano jak można go wykorzystać do wartościowania poszczególnych odmian. Skonstruowano również test S i przedział ufności Scheffé'go dla wszystkich porównań.

Rozdział szósty poświęcony jest zastosowaniu analizy składowych głównych do badania interakcji kontrastów odmian z miejscowościami względnie latami. Dzięki tej metodzie można z minimalną stratą informacji przedstawić na płaszczyźnie rozmieszczenie miejscowości lub lat pod względem zachowania się kontrastów wykazujących największą interakcję. Wyznaczone składowe główne są właśnie takimi kontrastami ortogonalnymi.

W rozdziale siódmym pokazano na przykładach praktyczne zastosowanie wyżej wymienionych metod.

Przykłady zostały obliczone na emc ODRA 1204 i MINSK 22 według programów własnych, które zostaną opublikowane w serii algorytmów biometrycznych i statystycznych Akademii Rolniczej w Poznaniu.

## 2. Sformułowanie modelu

Wprowadźmy następujące oznaczenia:

Niech  $\underline{y}_{ij} = (y_{ij}^{(1)}, y_{ij}^{(2)}, \dots, y_{ij}^{(p)})'$  oznacza wektor obserwowanych średnich plonów p odmian w i-tej miejscowości ( $i = 1, 2, \dots, \dots, m$ ) i j-tym roku ( $j = 1, 2, \dots, l$ ). Przez  $\underline{\mu} = (\mu^{(1)}, \mu^{(2)}, \dots, \dots, \mu^{(p)})'$  oznaczymy wektor prawdziwych średnich ogólnych. Efekty główne odmian w i-tej miejscowości oznaczymy za pomocą wektora  $\underline{a}_i = (a_i^{(1)}, a_i^{(2)}, \dots, a_i^{(p)})'$ , a efekty główne odmian w j-tym roku za pomocą wektora  $\underline{b}_j = (b_j^{(1)}, b_j^{(2)}, \dots, b_j^{(p)})'$ . Interakcję i-tej miej-

scowości z  $j$ -tym rokiem dla poszczególnych odmian oznaczymy za pomocą wektora  $\underline{c}_{ij} = (c_{ij}^{(1)}, c_{ij}^{(2)}, \dots, c_{ij}^{(p)})'$ . Natomiast przez  $\underline{e}_{ij} = (e_{ij}^{(1)}, e_{ij}^{(2)}, \dots, e_{ij}^{(p)})'$  oznaczymy wektor średnich odchyłek losowych (błędów). Przy tych oznaczeniach model liniowy średnich obserwowanych plonów  $p$  odmian w  $i$ -tej miejscowości i  $j$ -tym roku możemy zapisać w postaci

$$y_{ij} = \mu + a_i + b_j + c_{ij} + e_{ij}.$$

$$(i = 1, \dots, m; j = 1, \dots, l; m > p, l > p)$$

przy czym wektory  $\{a_i\}$ ,  $\{b_j\}$ ,  $\{c_{ij}\}$ ,  $\{e_{ij}\}$  są tu niezależnymi zmiennymi losowymi o rozkładach normalnych z wartościami oczekiwanymi

$$E(a_i) = E(b_j) = E(c_{ij}) = E(e_{ij}) = 0$$

i macierzami kowariancji odpowiednio

$$\sum_A^{p \times p} = [\sigma_{rr}^a], \quad \sum_B^{p \times p} = [\sigma_{rr}^b], \quad \sum_C^{p \times p} = [\sigma_{rr}^c], \quad \sum_E^{p \times p} = [\sigma_{rr}^e],$$

gdzie  $\sigma_{rr}^e = \delta_{rr} \cdot \sigma_e^2$ , z symbolem  $\delta_{rr}$  oznaczającym deltę Kroneckera ( $r, r' = 1, 2, \dots, p$ ).

### 3. Macierzowe składniki analizy wariancji i ich rozkłady

Wprowadzając pomocnicze oznaczenia:

$$P_{ir} = \sum_{j=1}^l y_{ij}^{(r)}, \quad Q_{jr} = \sum_{i=1}^m y_{ij}^{(r)},$$

$$G_r = P_{1r} + P_{2r} + \dots + P_{mr} = Q_{1r} + Q_{2r} + \dots + Q_{lr}$$

otrzymujemy następujące elementy macierzy sum kwadratów i iloczynów:

dla miejscowości -

$$A_{rr'} = \frac{1}{l} \sum_{i=1}^m P_{ir} P_{ir'} - \frac{1}{ml} G_r G_{r'},$$

dla lat -

$$B_{rr'} = \frac{1}{m} \sum_{j=1}^l Q_{jr} Q_{jr'} - \frac{1}{ml} G_r G_{r'},$$

dla interakcji miejscowości  $\times$  lata -

$$G_{rr'} = \sum_{i=1}^m \sum_{j=1}^1 (y_{ij}^{(r)} \cdot y_{ij}^{(r')}) - \frac{1}{1} \sum_{i=1}^m P_{ir} P_{ir'} - \frac{1}{m} \sum_{j=1}^1 Q_{jr} Q_{jr'} + \frac{1}{m1} G_r G_{r'}$$

dla błędu (wspólnego dla odmian, miejscowości i lat) -

$$E_{rr'} = \frac{1}{k} \delta_{rr'} \sum_{i=1}^m \sum_{j=1}^1 E_{ij}^{\#}$$

gdzie  $E_{ij}^{\#}$  jest sumą kwadratów dla błędu z doświadczenia w  $i$ -tej miejscowości i  $j$ -tym roku (dzielenie przez  $k$ , tzn. stałą liczbę replikacji, wynika stąd, że obliczenia dotyczące analizy całej serii doświadczeń przeprowadza się na średnich arytmetycznych z poszczególnych doświadczeń, podczas gdy sumy kwadratów  $E_{ij}^{\#}$  obliczone są na podstawie wyników poletkowych, to znaczy z indywidualnych jednostek),

dla zmienności całkowitej -

$$T_{rr'} = A_{rr'} + B_{rr'} + C_{rr'} + E_{rr'}$$

Zbadajmy teraz rozkłady macierzy sum kwadratów i iloczynów. Macierz sum kwadratów i iloczynów dla miejscowości możemy zapisać w postaci

$$A^{p \times p} = 1 \sum_{i=1}^m (\underline{g}_i - \underline{g}_.) (\underline{g}_i - \underline{g}_.)' = 1 \sum_{i=1}^m \underline{g}_i \underline{g}_i' - m 1 \underline{g}_. \underline{g}_.'$$

gdzie  $\underline{g}_i^{p \times 1} = \underline{a}_i + \underline{c}_i + \underline{e}_i$ .

Wektory  $\{\underline{g}_i\}$  są niezależne i mają jednakowe rozkłady normalne o wartości oczekiwanej

$$E(\underline{g}_i) = E(\underline{a}_i + \underline{c}_i + \underline{e}_i) = \underline{0}$$

i macierzy kowariancji

$$E(\underline{g}_i - E\underline{g}_i) (\underline{g}_i - E\underline{g}_i)' = E(\underline{g}_i \underline{g}_i') =$$

$$= E \begin{bmatrix} g_i^{(1)} g_i^{(1)}, g_i^{(1)} g_i^{(2)}, \dots, g_i^{(1)} g_i^{(p)} \\ g_i^{(2)} g_i^{(1)}, g_i^{(2)} g_i^{(2)}, \dots, g_i^{(2)} g_i^{(p)} \\ \dots \\ g_i^{(p)} g_i^{(1)}, g_i^{(p)} g_i^{(2)}, \dots, g_i^{(p)} g_i^{(p)} \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma_{11}^g, & \sigma_{12}^g, & \dots, & \sigma_{1p}^g \\ \sigma_{21}^g, & \sigma_{22}^g, & \dots, & \sigma_{2p}^g \\ \dots & \dots & \dots & \dots \\ \sigma_{p1}^g, & \sigma_{p2}^g, & \dots, & \sigma_{pp}^g \end{bmatrix} = \underline{\Sigma}_g,$$

gdzie  $\sigma_{rr}^g = \sigma_{rr}^a + \frac{1}{l} \sigma_{rr}^c + \frac{1}{l} \sigma_{rr}^e$ .

Stąd  $\underline{\Sigma}_g = \underline{\Sigma}_A + \frac{1}{l} \underline{\Sigma}_C + \frac{1}{l} \underline{\Sigma}_E$ .

Wartość oczekiwana macierzy  $\underline{A}$  jest równa

$$E(\underline{A}) = E\left(1 \sum_{i=1}^m \mathbf{g}_i \mathbf{g}_i' - m \mathbf{1} \mathbf{g}' \mathbf{g}'\right) = l(m-1) \underline{\Sigma}_g.$$

Zatem

$$E\left(\frac{1}{m-1} \underline{A}\right) = l \underline{\Sigma}_g = l \underline{\Sigma}_A + \underline{\Sigma}_C + \underline{\Sigma}_E.$$

Macierz  $\frac{1}{m-1} \underline{A}$  ma więc rozkład Wisharta  $W\left(\frac{1}{m-1} \underline{\Sigma}_g, m-1\right)$  (patrz [1], [13]).

Macierz sum kwadratów i iloczynów dla lat zapiszemy jako:

$$\underline{B}^{p \times p} = m \sum_{j=1}^l (\underline{d}_j - \underline{d}_\cdot)(\underline{d}_j - \underline{d}_\cdot)' = m \sum_{j=1}^l \underline{d}_j \underline{d}_j' - m \mathbf{1} \underline{d}_\cdot \underline{d}_\cdot',$$

gdzie

$$\underline{d}_j^{p \times 1} = \underline{b}_j + \underline{c}_j + \underline{e}_j.$$

Wektory  $\{\underline{d}_j\}$  są niezależne i mają jednakowe rozkłady  $N(\underline{0}, \underline{\Sigma}_d)$ .

Macierz kowariancji  $\underline{\Sigma}_d^{p \times p} = \left[\sigma_{rr}^d\right]$ , gdzie  $\sigma_{rr}^d = \sigma_{rr}^b + \frac{1}{m} \sigma_{rr}^c + \frac{1}{m} \sigma_{rr}^e$ .

Wobec tego

$$\underline{\Sigma}_d = \underline{\Sigma}_B + \frac{1}{m} \underline{\Sigma}_C + \frac{1}{m} \underline{\Sigma}_E.$$

Zatem

$$E(\underline{B}) = E\left(m \sum_{j=1}^l \underline{d}_j \underline{d}_j' - m \mathbf{1} \underline{d}_\cdot \underline{d}_\cdot'\right) = m(l-1) \underline{\Sigma}_d.$$

Stąd

$$E\left(\frac{1}{l-1} \underline{B}\right) = m \underline{\Sigma}_d = m \underline{\Sigma}_B + \underline{\Sigma}_C + \underline{\Sigma}_E.$$

Macierz  $\frac{1}{l-1} \underline{B}$  ma więc rozkład Wisharta  $W\left(\frac{m}{l-1} \underline{\Sigma}_d, l-1\right)$ .

Macierz sum kwadratów i iloczynów dla interakcji miejscowości  $\times$  lata można przedstawić następująco:

$$\underline{C}^{p \times p} = \sum_{i=1}^m \sum_{j=1}^1 (\underline{h}_{ij} - \underline{h}_{i.} - \underline{h}_{.j} + \underline{h}_{..}) (\underline{h}_{ij} - \underline{h}_{i.} - \underline{h}_{.j} + \underline{h}_{..})',$$

gdzie

$$\underline{h}_{ij}^{p \times 1} = \underline{c}_{ij} + \underline{e}_{ij}.$$

Wektory  $\{\underline{h}_{ij}\}$  są niezależne i mają jednakowe rozkłady  $N(\underline{0}, \underline{\Sigma}_h)$ , gdzie

$$\underline{\Sigma}_h^{p \times p} = \begin{bmatrix} \sigma_{rr}^h \end{bmatrix}, \quad \sigma_{rr}^h = \sigma_{rr}^e + \sigma_{rr}^c.$$

Stąd otrzymamy

$$\underline{\Sigma}_h = \underline{\Sigma}_C + \underline{\Sigma}_E.$$

Natomiast wartość oczekiwana macierzy  $\underline{C}$  jest równa

$$E(\underline{C}) = (m-1)(1-1) \underline{\Sigma}_h.$$

Wobec tego

$$E\left(\frac{1}{(m-1)(1-1)} \underline{C}\right) = \underline{\Sigma}_h = \underline{\Sigma}_C + \underline{\Sigma}_E.$$

Macierz  $\frac{1}{(m-1)(1-1)} \underline{C}$  ma więc rozkład Wisharta

$$W\left(\frac{1}{(m-1)(1-1)} \underline{\Sigma}_h, (m-1)(1-1)\right).$$

Powyższe wyniki wygodnie jest przedstawić tak jak w tabeli 1.

Tabela 1

Analiza	Zmiennosc	Liczba stopni swobody	Macierze sum kwadratów i iloczynów	Wartości oczekiwane macierzy średnich kwadratów i iloczynów
Wielozmienna	Miejscowości	$m-1$	$\underline{A}$	$\underline{\Sigma}_E + \underline{\Sigma}_C + 1 \underline{\Sigma}_A$
	Lata	$1-1$	$\underline{B}$	$\underline{\Sigma}_E + \underline{\Sigma}_C + m \underline{\Sigma}_B$
	Miejscowości $\times$ lata	$(m-1)(1-1)$	$\underline{C}$	$\underline{\Sigma}_E + \underline{\Sigma}_C$
Jednozmenna	Błąd	$m1(p-1)(k-1)$	$\underline{E}$	$\underline{\Sigma}_E$

Należy zauważyć, że macierze  $\underline{A}$ ,  $\underline{B}$  i  $\underline{C}$  wyliczane są przy traktowaniu wyników jako obserwacji wielowymiarowych, podczas gdy macierz  $\underline{E}$  obliczana jest z analiz przeprowadzonych dla poszczególnych doświadczeń przy traktowaniu wyników jako obserwacji jednowymiarowych (patrz także uwaga na początku rozdziału). Z tych względów liczba stopni swobody dla błędu nie jest porównywalna z liczbami stopni swobody podanymi wyżej w tabeli 1.

#### 4. Testowanie hipotez

Przed przystąpieniem do testowania hipotez wprowadźmy następujące oznaczenia:

$$\underline{m}_0^{p \times 1} = [1, 1, \dots, 1]'$$

$$\underline{M}_1^{p \times (p-1)} = [\underline{m}_1, \underline{m}_2, \dots, \underline{m}_{p-1}],$$

gdzie  $\underline{m}_k$  ( $k = 1, 2, \dots, p-1$ ) jest wektorem współczynników  $k$ -tego z  $(p-1)$  interesujących nas kontrastów między odmianami, co oznacza, że zachodzi równość  $\underline{m}_0' \underline{M}_1 = \underline{0}$

Z tabeli 1 jest widoczny sposób budowy odpowiednich testów do sprawdzania kolejnych hipotez.

1. Hipotezę o braku różnic między miejscowościami pod względem sumarycznych plonów odmian można zapisać w postaci

$$H_{0,A} : \underline{m}_0' \sum_A \underline{m}_0 = 0.$$

Do zbudowania odpowiedniej funkcji testowej potrzebne są następujące wielkości

$$SS_A = \underline{m}_0' \underline{A} \underline{m}_0 \qquad MS_A = \frac{SS_A}{m-1}$$

$$SS_C = \underline{m}_0' \underline{C} \underline{m}_0 \qquad MS_C = \frac{SS_C}{(m-1)(l-1)}.$$

Funkcja testowa  $F = \frac{MS_A}{MS_C}$  ma przy prawdziwości hipotezy  $H_{0,A}$  rozkład  $F$  z  $(m-1)$  i  $(m-1)(l-1)$  stopniami swobody. Hipotezę  $H_{0,A}$  odrzucimy na poziomie istotności  $\alpha$  gdy  $F > F_{\alpha; m-1, (m-1)(l-1)}$ .

2. Hipotezę o braku różnic między latami pod względem sumarycznych plonów odmian zapiszemy w postaci

$$H_{0,B} : \underline{m}_0' \sum_B \underline{m}_0 = 0,$$

Hipotezę tę weryfikuje się analogicznie jak w punkcie 1 z tym, że należy zamiast macierzy  $\underline{A}$  dla miejscowości wstawić macierz  $\underline{B}$  dla lat, zaś liczbę poziomów dla miejscowości zamienić z liczbą poziomów dla lat czyli  $m$  zamienić z  $l$ .

3. Hipotezę o braku interakcji miejscowości z latami pod względem sumarycznych plonów odmian zapiszemy w postaci

$$H_{0,C} : \underline{m}_0' \underline{C} \underline{m}_0 = 0.$$

Jeśli przyjmiemy następujące oznaczenia

$$SS_C = \underline{m}_0' \underline{C} \underline{m}_0 \qquad MS_C = \frac{SS_C}{(m-1)(l-1)}$$

$$SS_E = \underline{m}_0' \underline{E} \underline{m}_0 \qquad MS_E = \frac{SS_E}{ml(k-1)(p-1)}$$

to funkcja testowa  $F = \frac{MS_C}{MS_E}$  ma przy prawdziwości hipotezy  $H_{0,C}$  rozkład  $F$  z  $(m-1)(l-1)$  i  $ml(p-1)(k-1)$  stopniami swobody. Hipotezę  $H_{0,C}$  odrzucimy na poziomie  $\alpha$  jeśli  $F > F_{\alpha, (m-1)(l-1), ml(k-1)(p-1)}$ .

4. Hipotezę o braku interakcji miejscowości z odmianami zapiszemy w postaci

$$H_{0,AV} : \underline{m}_1' \underline{A} \underline{m}_1 = 0.$$

Jako kryterium do sprawdzania tej hipotezy służy statystyka  $\Lambda$  Wilksa (por. [12]).

Wprowadzając macierze

$$\underline{S}_A = \underline{m}_1' \underline{A} \underline{m}_1, \qquad \underline{S}_C = \underline{m}_1' \underline{C} \underline{m}_1,$$

typu  $(p-1) \times (p-1)$  można zapisać statystykę  $\Lambda$  jako iloraz

$$\Lambda = \frac{|\underline{S}_C|}{|\underline{S}_A + \underline{S}_C|}.$$

Dla odczytania wartości krytycznych z ogólnie dostępnych tabeli statystycznych konieczne jest odpowiednie przekształcenie statystyki  $\Lambda$ . Można tu wykorzystać przybliżony test  $F$  zaproponowany przez Rao [9], dla którego funkcją testową jest

$$(4.1) \quad F = \frac{m's' - 2\lambda'}{gu} \cdot \frac{1 - \Lambda^{1/s'}}{\Lambda^{1/s'}}$$



gdzie

$$m' = (m-1)(l-1) - \frac{p-m+1}{2}, \quad s' = \sqrt{\frac{(m-1)^2(p-1)^2 - 4}{(m-1)^2 + (p-1)^2 - 5}},$$

$$\lambda' = \frac{(m-1)(p-1) - 2}{4}, \quad g = m - 1, \quad u = p - 1.$$

Statystyka  $F$  ma przy prawdziwości hipotezy  $H_{0,AV}$  w przybliżeniu rozkład  $F$  z  $gu$  i  $(m's' - 2\lambda')$  stopniami swobody. Hipotezę  $H_{0,AV}$  odrzucimy na poziomie istotności  $\alpha$  jeśli  $F_{Obl} > F_{\alpha, gu, m's' - 2\lambda'}$ .

5. Hipoteza o braku interakcji lat z odmianami jest następująca:

$$H_{0,BV} : \underline{M}'_1 \underline{\Sigma}'_B \underline{M}_1 = 0.$$

Testowanie tej hipotezy jest analogiczne jak hipotezy  $H_{0,AV}$  z tym, że przy tworzeniu odpowiednich macierzy do testu zastępujemy macierz  $\underline{A}$  macierzą  $\underline{B}$  oraz zamieniamy  $m$  z  $l$ .

6. Hipotezę o braku interakcji miejscowości z latami i odmianami można zapisać w postaci

$$H_{0,CV} : \underline{M}'_1 \underline{\Sigma}'_C \underline{M}_1 = 0.$$

Jeśli przyjmujemy następujące oznaczenia

$$\underline{S}_C = \underline{M}'_1 \underline{C} \underline{M}_1, \quad \underline{S}_E = \underline{M}'_1 \underline{E} \underline{M}_1,$$

to

$$\Lambda = \frac{|\underline{S}_E|}{|\underline{S}_C + \underline{S}_E|}.$$

Wówczas w funkcji testowej (4.1)

$$m' = ml(k-1)(p-1) - \frac{p - (m-1)(l-1)}{2},$$

$$s' = \sqrt{\frac{[(m-1)(l-1)]^2 (p-1)^2 - 4}{[(m-1)(l-1)]^2 + (p-1)^2 - 5}}, \quad u = p - 1$$

$$\lambda' = \frac{(m-1)(l-1)(p-1) - 2}{4}, \quad g = (m-1)(l-1).$$

7. Hipoteza o braku zróżnicowania odmian jest następująca

$$H_{0,V} : \underline{M}'_1 \underline{\mu} = 0.$$

Właściwym testem do sprawdzania tej hipotezy jest test  $T^2$  Hotelinga (patrz [6]). Aby zbudować statystykę  $T^2$  musimy wykonać sze-

reg pomocniczych obliczeń. Niech

$$\underline{y}^{p \times 1} = \frac{1}{m-1} \sum_{i=1}^m \sum_{j=1}^1 \underline{y}_{ij}.$$

Wektor  $\underline{y} \dots$  ma rozkład  $N(\underline{\mu}, \frac{1}{m-1} \underline{\Sigma})$ , gdzie

$$\underline{\Sigma}^{p \times p} = 1 \underline{\Sigma}_A + m \underline{\Sigma}_B + \underline{\Sigma}_C + \underline{\Sigma}_E.$$

Dokonyjemy teraz przekształcenia wektora  $\underline{y} \dots$ . Wektor  $\underline{y}^{\pi(p-1) \times 1} = \underline{M}'_1 \underline{y} \dots$  ma rozkład  $N(\underline{\mu}^{\pi}, \frac{1}{m-1} \underline{\Sigma}'_1)$ , gdzie

$$\begin{aligned} \underline{\mu}^{\pi(p-1) \times 1} &= \underline{M}'_1 \underline{\mu}, \\ \underline{\Sigma}'_1^{(p-1) \times (p-1)} &= \underline{M}'_1 \underline{\Sigma} \underline{M}_1. \end{aligned}$$

Macierz  $\underline{\Sigma}'_1$  można zapisać w innej postaci:

$$\begin{aligned} \underline{\Sigma}'_1 &= \underline{M}'_1 \underline{\Sigma} \underline{M}_1 = \underline{M}'_1 (1 \underline{\Sigma}_A + m \underline{\Sigma}_B + \underline{\Sigma}_C + \underline{\Sigma}_E) \underline{M}_1 = \\ &= \underline{M}'_1 (1 \underline{\Sigma}'_g + m \underline{\Sigma}'_d - \underline{\Sigma}'_h) \underline{M}_1 = 1 \underline{M}'_1 \underline{\Sigma}'_g \underline{M}_1 + m \underline{M}'_1 \underline{\Sigma}'_d \underline{M}_1 - \underline{M}'_1 \underline{\Sigma}'_h \underline{M}_1 = \\ &= E\left(\frac{1}{m-1} \underline{M}'_1 \underline{A} \underline{M}_1\right) + E\left(\frac{1}{1-1} \underline{M}'_1 \underline{B} \underline{M}_1\right) - E\left(\frac{1}{(m-1)(1-1)} \underline{M}'_1 \underline{C} \underline{M}_1\right) = \\ &= E\left(\frac{1}{m-1} \underline{S}_A\right) + E\left(\frac{1}{1-1} \underline{S}_B\right) - E\left(\frac{1}{(m-1)(1-1)} \underline{S}_C\right). \end{aligned}$$

Nieobciążoną oceną macierzy  $\underline{\Sigma}'_1$  jest

$$\hat{\underline{\Sigma}}'_1 = \frac{1}{m-1} \underline{S}_A + \frac{1}{1-1} \underline{S}_B - \frac{1}{(m-1)(1-1)} \underline{S}_C.$$

Z powyższego wzoru widać, że nie można zbudować dokładnej statystyki  $T^2$ . Zatem do sprawdzania hipotezy  $H_{0,V}$  skonstruujemy przybliżoną statystykę  $T^2$  Hotellinga.

Macierze  $\underline{S}_A$ ,  $\underline{S}_B$ ,  $\underline{S}_C$  są niezależne i mają rozkład Wisharta

$$\frac{1}{m-1} \underline{S}_A \sim W\left(\frac{1}{m-1} \underline{M}'_1 \underline{\Sigma}'_g \underline{M}_1, m-1\right)$$

$$\frac{1}{1-1} \underline{S}_B \sim W\left(\frac{m}{1-1} \underline{M}'_1 \underline{\Sigma}'_d \underline{M}_1, 1-1\right)$$

$$\frac{1}{(m-1)(1-1)} \underline{S}_C \sim W\left(\frac{1}{(m-1)(1-1)} \underline{M}'_1 \underline{\Sigma}'_h \underline{M}_1, (m-1)(1-1)\right).$$

Przybliżony rozkład macierzy  $\hat{\underline{\Sigma}}'_1$  rozkładem Wisharta  $W\left(\frac{1}{v} \underline{\Sigma}'_1, v\right)$ , gdzie  $v$  określimy z warunku, że  $\hat{\underline{\Sigma}}'_1$  i  $W\left(\frac{1}{v} \underline{\Sigma}'_1, v\right)$  będą miały jednakowe wariancje i kowariancje (one mają już tę samą wartość oczekiwaną). Przy obliczaniu liczby stopni swobody korzystać będziemy z następującego twierdzenia (patrz [10]):

Niech  $\underline{Z} \sim W(\underline{Z}, k)$  i niech  $\underline{L}$  będzie ustalonym wektorem. Wtedy  $\underline{L}'\underline{Z}\underline{L} \sim \sigma_{L}^2 \chi^2(k)$ , gdzie  $\sigma_{L}^2 = \underline{L}'\underline{Z}\underline{L}$ .

Stąd dla ustalonego wektora  $\underline{L}^{(p-1) \times 1}$

$$\frac{1}{m-1} \underline{L}'\underline{S}_A\underline{L} \sim \frac{1}{m-1} \sigma_{L,g}^2 \chi^2(m-1),$$

gdzie  $\sigma_{L,g}^2 = \underline{L}'\underline{M}_1'\underline{Z}_g\underline{M}_1\underline{L}$ ;

$$\frac{1}{1-1} \underline{L}'\underline{S}_B\underline{L} \sim \frac{m}{1-1} \sigma_{L,d}^2 \chi^2(1-1),$$

gdzie  $\sigma_{L,d}^2 = \underline{L}'\underline{M}_1'\underline{Z}_d\underline{M}_1\underline{L}$ ;

$$\frac{1}{(m-1)(1-1)} \underline{L}'\underline{S}_C\underline{L} \sim \frac{1}{(m-1)(1-1)} \sigma_{L,h}^2 \chi^2((m-1)(1-1)),$$

gdzie  $\sigma_{L,h}^2 = \underline{L}'\underline{M}_1'\underline{Z}_h\underline{M}_1\underline{L}$ ;

$$\underline{L}'\underline{Z}_1\underline{L} \sim \frac{1}{v} \sigma_{L,q}^2 \chi^2(v),$$

gdzie  $\sigma_{L,q}^2 = \underline{L}'\underline{M}_1'\underline{Z}_1\underline{M}_1\underline{L}$ .

Wówczas

$$\text{Var} \left( \frac{1}{m-1} \sigma_{L,g}^2 \chi^2(m-1) \right) = \frac{2(1 \sigma_{L,g}^2)^2}{m-1},$$

$$\text{Var} \left( \frac{m}{1-1} \sigma_{L,d}^2 \chi^2(1-1) \right) = \frac{2(m \sigma_{L,d}^2)^2}{1-1},$$

$$\text{Var} \left( \frac{1}{(m-1)(1-1)} \sigma_{L,h}^2 \chi^2((m-1)(1-1)) \right) = \frac{2(\sigma_{L,h}^2)^2}{(m-1)(1-1)},$$

$$\text{Var} \left( \frac{1}{v} \sigma_{L,q}^2 \chi^2(v) \right) = \frac{2(\sigma_{L,q}^2)^2}{v}.$$

Porównując powyższe wariancje otrzymamy

$$\frac{2(\sigma_{L,q}^2)^2}{v} = \frac{2(1 \sigma_{L,g}^2)^2}{m-1} + \frac{2(m \sigma_{L,d}^2)^2}{1-1} + \frac{2(\sigma_{L,h}^2)^2}{(m-1)(1-1)}$$

skąd

$$v = \frac{(\sigma_{L,q}^2)^2}{\frac{(1 \sigma_{L,g}^2)^2}{m-1} + \frac{(m \sigma_{L,d}^2)^2}{1-1} + \frac{(\sigma_{L,h}^2)^2}{(m-1)(1-1)}}.$$

Zatem ocena  $\hat{\nu}$  jest równa

$$(4.2) \quad \hat{\nu} = \frac{(\hat{\sigma}_{L,q}^2)^2}{\frac{(1\hat{\sigma}_{L,g}^2)^2}{m-1} + \frac{(m\hat{\sigma}_{L,d}^2)^2}{1-1} + \frac{(\hat{\sigma}_{L,h}^2)^2}{(m-1)(1-1)}} ,$$

gdzie

$$\hat{\sigma}_{L,q}^2 = L' \left( \frac{1}{m-1} S_A + \frac{1}{1-1} S_B - \frac{1}{(m-1)(1-1)} S_C \right) L ,$$

$$1\hat{\sigma}_{L,g}^2 = L' \left( \frac{1}{m-1} S_A \right) L ,$$

(4.3)

$$m\hat{\sigma}_{L,d}^2 = L' \left( \frac{1}{1-1} S_B \right) L ,$$

$$\hat{\sigma}_{L,h}^2 = L' \left( \frac{1}{(m-1)(1-1)} S_C \right) L .$$

Na podstawie powyższych rozważań przybliżoną statystykę  $T^2$  możemy zapisać następująco:

$$T^2 = m l (\underline{y}_{\dots})' (\hat{\Sigma}_1)^{-1} (\underline{y}_{\dots}) ,$$

gdzie

$$\hat{\Sigma}_1 = \frac{1}{m-1} S_A + \frac{1}{1-1} S_B - \frac{1}{(m-1)(1-1)} S_C .$$

Wówczas  $\frac{\hat{\nu}-p+2}{p-1} \frac{T^2}{\hat{\nu}}$  ma w przybliżeniu rozkład  $F$  z  $(p-1)$  i  $(\hat{\nu}-p+2)$  stopniami swobody i parametrem niecentralności

$$\delta = m l \underline{\mu}' \underline{M}_1 (\hat{\Sigma}_1)^{-1} \underline{M}_1' \underline{\mu} .$$

Hipoteza  $\underline{M}_1' \underline{\mu} = \underline{0}$  jest odrzucana wtedy i tylko wtedy, gdy

$$\frac{\hat{\nu}-p+2}{p-1} \cdot \frac{T^2}{\hat{\nu}} > F_{\alpha, p-1, \hat{\nu}-p+2} .$$

Należy jeszcze rozpatrzyć przypadek szczególny, gdy hipotezy interakcyjne  $H_{0,AV}$  lub  $H_{0,BV}$  są prawdziwe, gdyż w tym przypadku można skonstruować dokładną statystykę  $T^2$  Hotellinga. Gdy hipoteza  $H_{0,AV}$  jest prawdziwa, to wówczas otrzymamy następujące interesujące nas elementy do konstrukcji statystyki  $T^2$ :

$$\hat{\Sigma}_1 = \frac{1}{1-1} S_B, \quad \hat{\nu} = 1 - 1 .$$

Wówczas  $\frac{1-p+1}{p-1} \cdot \frac{T^2}{1-1}$  ma rozkład  $F$  z  $(p-1)$  i  $(1-p+1)$  stopniami swobody i parametrem niecentralności  $\delta = 1 \underline{\mu}' \underline{M}_1 (\underline{M}_1' \hat{\Sigma}_1 \underline{M}_1)^{-1} \underline{M}_1' \underline{\mu}$ . Analogicznie przebiega testowanie, gdy hipoteza  $H_{0,BV}$  jest praw-

dziwa. W tym przypadku

$$\hat{\Sigma}_1 = \frac{1}{m-1} S_A, \quad \hat{\nu} = m - 1.$$

Natomiast, gdy prawdziwe są hipotezy  $H_{0,AV}$  i  $H_{0,BV}$  to

$$\hat{\Sigma}_1 = \frac{1}{(m-1)(l-1)} S_C, \quad \hat{\nu} = (m-1)(l-1).$$

### 5. Kontrast maksymalny

Przy testowaniu hipotezy  $\underline{M}'_1/\underline{\mu} = \underline{0}$  istnieje problem określenia wektora  $\underline{l}$  przy obliczaniu przybliżonej liczby stopni swobody  $\nu$ . Proponujemy tutaj następujące rozwiązanie tego zagadnienia.

Odrzucenie hipotezy  $\underline{M}'_1/\underline{\mu} = \underline{0}$  następuje (także) wtedy i tylko wtedy, gdy dla conajmniej jednego kontrastu  $\psi$  zachodzi nierówność

$$\frac{\hat{\nu} - p + 2}{(p-1)\hat{\nu}} \cdot \frac{\hat{\Psi}^2}{\hat{\sigma}_{\psi}^2} > F_{\alpha, p-1, \hat{\nu} - p + 2},$$

gdzie

$$\hat{\Psi} = \underline{h}' \underline{M}'_1 \underline{y} \dots \sim N(\psi, \sigma_{\psi}^2),$$

przy czym

$$\psi = \underline{h}' \underline{M}'_1 \underline{\mu}, \quad \underline{h} = (h^{(1)}, h^{(2)}, \dots, h^{(p-1)})',$$

$$\sigma_{\psi}^2 = \frac{1}{m-1} \underline{h}' \hat{\Sigma}_1 \underline{h}.$$

Nieobciążoną oceną  $\sigma_{\psi}^2$  jest

$$\hat{\sigma}_{\psi}^2 = \frac{1}{m-1} \underline{h}' \hat{\Sigma}_1 \underline{h}.$$

Zatem zamiast testu  $T^2$  Hotellinga można stosować test S Scheffe'go ([11])

$$\frac{\hat{\nu} - p + 2}{(p-1)\hat{\nu}} \max \frac{\hat{\Psi}^2}{\hat{\sigma}_{\psi}^2} > F_{\alpha, p-1, \hat{\nu} - p + 2}.$$

Wobec tego należy znaleźć taki wektor  $\underline{h}$ , który maksymalizuje wyrażenie (patrz [8])

$$v = \frac{\hat{\Psi}^2}{\hat{\sigma}_{\psi}^2} = \frac{(\underline{h}' \underline{M}'_1 \underline{y} \dots)^2}{\frac{1}{m-1} \underline{h}' \hat{\Sigma}_1 \underline{h}}.$$

Wektor o tej własności oznaczać będziemy w dalszym ciągu przez  $\underline{h}_{\max}$ . Natomiast  $\psi_{\max} = \underline{h}'_{\max} \underline{M}'_1 \underline{\mu}$  nazywać będziemy kontrastem

maksymalnym. Wektor  $\underline{h}_{\max}$  wyliczamy z układu równań

$$\hat{\Sigma}_1 \underline{h} = \underline{M}'_1 \underline{y} \dots$$

Wówczas

$$(5.1) \quad \underline{h}_{\max} = (\hat{\Sigma}_1)^{-1} \underline{M}'_1 \underline{y} \dots$$

$$\text{Stąd } V_{\max} = m_1 \underline{y} \dots \underline{M}_1 (\hat{\Sigma}_1)^{-1} \underline{M}'_1 \underline{y} \dots$$

Na podstawie powyższych obliczeń test S Scheffe'go przyjmie postać

$$\frac{\hat{v}_{-p+2}}{(p-1)\hat{v}} m_1 \underline{y} \dots \underline{M}_1 (\hat{\Sigma}_1)^{-1} \underline{M}'_1 \underline{y} \dots > F_{\alpha, p-1, \hat{v}_{-p+2}}$$

gdzie  $\hat{v}$  jest określone wzorem (4.2), przy czym za wektor  $\underline{L}$  przyjmujemy wektor  $\underline{h}_{\max}$ . Wektor  $\underline{M}_1 \underline{h}_{\max}$ , tzn. wektor współczynników maksymalnego kontrastu przyjmuje te same wartości dla każdej dowolnej macierzy  $\underline{M}_1$  tego samego rzędu (w tym przypadku rząd ten wynosi  $p-1$ ).

Jeśli dla niektórych kontrastów interesują nas przedziały ufności, to wtedy możemy określić przedział ufności Scheffe'go

$$\hat{\Psi} - s \hat{\sigma} \hat{\psi} \leq \Psi \leq \hat{\Psi} + s \hat{\sigma} \hat{\psi}$$

$$\text{gdzie } s^2 = \frac{\hat{v}(p-1)}{\hat{v}_{-p+2}} F_{\alpha, p-1, \hat{v}_{-p+2}}$$

przy czym przy obliczaniu  $\hat{v}$  we wzorach (4.3) za wektor  $\underline{M}_1 \underline{L}$  należy przyjąć wektor współczynników danego kontrastu. Analogicznie przebiegają rozważania w przypadku szczególnym, gdy hipotezy  $H_{0,AV}$  lub  $H_{0,BV}$  są prawdziwe, z tym, że za  $\hat{\Sigma}_1$  i  $\hat{v}$  przyjmujemy odpowiednie elementy jak przy testowaniu szczególnego przypadku hipotezy  $\underline{M}'_1 \underline{\mu} = \underline{0}$ .

Znalezienie wektora maksymalnego kontrastu ( $\underline{M}_1 \underline{h}_{\max}$ ) okazuje się również bardzo przydatne do oceny wartości poszczególnych odmian. Zauważmy bowiem, że maksymalny kontrast przypisuje poszczególnym odmianom takie współczynniki, które dają największe liniowe skontrastowanie tych odmian. Przy tym na tworzenie tych współczynników wpływają nie tylko średnie, ale także wariancje i kowariancje. W ten sposób wykorzystane są informacje dotyczące nie tylko wysokości plonowania, ale także dotyczące jego stabilności oraz współzależności (podobieństwa) między odmianami. Wartości liczbowe współczynników maksymalnego kontrastu można traktować jako pewne "wagi jakościowe" poszczególnych odmian, natomiast znak (+ lub -) pozwala zaliczyć każdą z odmian do grupy dobrych lub

złych. Wagi te można wykorzystać przy formułowaniu zaleceń dotyczących doboru odmian do uprawy.

Jeśli chcemy w odpowiednich proporcjach zalecać wszystkie odmiany z wyjątkiem najgorszej, to należy przesunąć współczynniki maksymalnego kontrastu tak, aby odmiana najgorsza otrzymała współczynnik o wartości zero. Uzyskane w ten sposób nowe współczynniki stanowiąc będą wspomniane "wagi jakościowe". Można je wyrazić w procentach sumy wartości wszystkich "wag". Wówczas w podejmowaniu decyzji uprawowych procenty te można traktować jako wskazania dotyczące udziału poszczególnych odmian na określonym obszarze uprawowym.

Natomiast w przypadku, gdy chcemy zalecać tylko odmiany, które w kontraście maksymalnym znalazły się w grupie odmian dobrych, to aby określić propozycje uprawowe należy zbudować nowy układ kontrastów liniowo niezależnych (macierz  $\underline{M}_1$ ), który przypisuje wszystkim odmianom zaliczonym do grupy odmian złych tę samą wartość ujemną.

Dalsze postępowanie w tym przypadku jest analogiczne do przedstawionego powyżej (szczegóły będą podane w przykładzie). W ten sposób przedstawione propozycje można interpretować jako pomocnicze wytyczne dla decyzji uprawowych pozwalających oczekiwać najwyższego sumarycznego plonu z najmniejszą jego wariancją.

## 6. Badanie interakcji za pomocą składowych głównych

W rozdziale czwartym testowane są interakcje miejscowości  $\times$  odmiany i lata  $\times$  odmiany. Jeżeli którakolwiek z tych interakcji jest istotna, to można dokonać jej szczegółowej analizy za pomocą składowych głównych (patrz [4], [7], [10]). Można bowiem przedstawić wyniki z poszczególnych miejscowości (lub lat) w postaci kontrastów między odmianami i następnie te zmienne przekształcić w składowe główne. Często pierwsze dwie składowe główne będące nowymi kontrastami o maksymalnej interakcji z miejscowościami (lub latami) stanowiąc będą wystarczającą informację o różnicach między odmianami w poszczególnych miejscowościach (lub latach). Użycie tych składowych jako współrzędnych umożliwi nadto przedstawienie wyników dla miejscowości (bądź lat) na wykresie dwuwymiarowym, co ułatwi interpretację występującej interakcji.

Niech

$$(6.1) \quad \psi_k = \sum_{r=1}^p l_k^{(r)} \mu^{(r)} = \underline{l}'_k \underline{\mu},$$

gdzie  $r = 1, 2, \dots, p$ ;  $k = 1, 2, \dots, p-1$ ;

$$\sum_{r=1}^p l_k^{(r)} = 0; \quad (l_k^{(1)}, l_k^{(2)}, \dots, l_k^{(p)}) = \underline{l}'_k$$

będzie dowolnym kontrastem liniowo niezależnym między efektami głównymi odmian. Jego oceną nieobciążoną będzie kontrast obserwowany

$$(6.2) \quad \hat{\psi}_k = \sum_{r=1}^p l_k^{(r)} y_{..}^{(r)} = \underline{l}'_k \underline{y} \dots$$

Układ  $p-1$  kontrastów obserwowanych można zapisać w postaci

$$(6.3) \quad \hat{\underline{\psi}} = \begin{bmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \\ \vdots \\ \hat{\psi}_{p-1} \end{bmatrix} = \begin{bmatrix} \underline{l}'_1 \underline{y} \dots \\ \underline{l}'_2 \underline{y} \dots \\ \vdots \\ \underline{l}'_{p-1} \underline{y} \dots \end{bmatrix} = \begin{bmatrix} \underline{l}'_1 \\ \underline{l}'_2 \\ \vdots \\ \underline{l}'_{p-1} \end{bmatrix} \underline{y} \dots = \\ = \underline{L}^{(p-1) \times p} \underline{y} \dots \underline{p} \times 1.$$

Podobnie można zapisać układ takich kontrastów dla każdej z  $m$  miejscowości. Dla  $i$ -tej miejscowości będziemy mieli

$$(6.4) \quad \hat{\underline{\psi}}_i = \begin{bmatrix} \hat{\psi}_{1,i} \\ \vdots \\ \hat{\psi}_{p-1,i} \end{bmatrix} = \underline{L} \underline{y}_{i..},$$

gdzie

$$\hat{\psi}_{ki} = \underline{l}'_k \underline{y}_{i..}$$

Macierz sum kwadratów i iloczynów odchyłek kontrastów  $\hat{\underline{\psi}}_i$  od  $\hat{\underline{\psi}}$ , czyli macierz interakcji kontrastów z miejscowościami, można przedstawić następująco:

$$\underline{S}_A \hat{\underline{\psi}} = 1 \sum_{i=1}^m (\hat{\underline{\psi}}_{i..} - \hat{\underline{\psi}}) (\hat{\underline{\psi}}_{i..} - \hat{\underline{\psi}})'$$



Jej wymiar jest  $(p-1) \times (p-1)$ . Zauważmy, że

$$(6.5) \quad \underline{S}_A \hat{\Psi} = 1 \sum_{i=1}^m (\underline{L} y_{1..} - \underline{L} y_{...}) (\underline{L} y_{1..} - \underline{L} y_{...})' = \underline{L} \underline{A} \underline{L}'.$$

W podobny sposób można oczywiście przedstawić kontrasty między odmiannami dla każdego z  $l$  lat, a także dla każdego z  $m$  doświadczeń. Oznaczając  $k$ -ty kontrast obserwowany w  $j$ -tym roku przez  $\hat{\Psi}_{k,j}$ , a  $k$ -ty kontrast obserwowany w doświadczeniu z  $i$ -tej miejscowości oraz  $j$ -tego roku przez  $\hat{\Psi}_{kij}$ , można zapisać odpowiednie układy tych kontrastów jako wektory

$$(6.6) \quad \hat{\Psi}_{1j} = \begin{bmatrix} \hat{\Psi}_{1,1j} \\ \vdots \\ \hat{\Psi}_{p-1,1j} \end{bmatrix} = \underline{L} y_{1j}, \quad \hat{\Psi}_{.j} = \begin{bmatrix} \hat{\Psi}_{1,.j} \\ \vdots \\ \hat{\Psi}_{p-1,.j} \end{bmatrix} = \underline{L} y_{.j}.$$

Wprowadźmy macierz

$$\underline{S}_C \hat{\Psi} = \sum_{i=1}^m \sum_{j=1}^l (\hat{\Psi}_{1j} - \hat{\Psi}_{1.} - \hat{\Psi}_{.j} + \hat{\Psi}) (\hat{\Psi}_{1j} - \hat{\Psi}_{1.} - \hat{\Psi}_{.j} + \hat{\Psi}),$$

która będzie macierzą sum kwadratów i iloczynów dla interakcji rozważanych kontrastów z miejscowościami i latami. Zauważmy, że można ją zapisać także w postaci

$$(6.7) \quad \underline{S}_C \hat{\Psi} = \sum_{i=1}^m \sum_{j=1}^l (\underline{L} y_{1j} - \underline{L} y_{1..} - \underline{L} y_{.j} + \underline{L} y_{...}) (\underline{L} y_{1j} - \underline{L} y_{1..} - \underline{L} y_{.j} + \underline{L} y_{...})' = \underline{L} \underline{C} \underline{L}'.$$

Można sobie postawić zadanie znalezienia takich kontrastów, które wykazują największą interakcję z miejscowościami. Zagadnienie to rozwiążemy stosując analizę składowych głównych. Weźmy pod uwagę macierze  $\underline{S}_A \hat{\Psi}$  i  $\underline{S}_C \hat{\Psi}$  i znajdziemy pierwiastki charakterystyczne równania wyznacznikowego

$$(6.8) \quad |\underline{S}_A \hat{\Psi} - \lambda \underline{S}_C \hat{\Psi}| = 0.$$

Aby rozwiązać równanie (6.8) wygodnie jest przekształcić je najpierw do postaci

$$(6.9) \quad |\underline{R}^{-1} \underline{S}_A \hat{\Psi} \underline{R}^{-1} - \lambda \underline{I}| = 0$$

przy czym  $\underline{S}_C \hat{\Psi} = \underline{R}' \underline{R}$ . Pierwiastki równania (6.9) ustawiamy w porządku nierosnącym:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p-1}.$$

Dla każdego pierwiastka znajdujemy odpowiadający mu wektor charakterystyczny  $\underline{q}_k^{\pi}$ , tzn. wektor spełniający równanie

$$(6.10) \quad (\underline{R}^{-1} \underline{S}_A \hat{\Psi} \underline{R}^{-1} - \lambda_k \underline{I}) \underline{q}_k^{\pi} = 0.$$

Otrzymane wektory charakterystyczne normalizujemy. Wówczas

$$(6.11) \quad \underline{q}_k^{\pi} \underline{q}_{k'}^{\pi} = \begin{cases} 1 & \text{gdy } k = k' \\ 0 & \text{gdy } k \neq k' \end{cases} \text{ oraz}$$

$$\underline{q}_k^{\pi} \underline{R}^{-1} \underline{S}_A \hat{\Psi} \underline{R}^{-1} \underline{q}_{k'}^{\pi} = \begin{cases} \lambda_k & \text{gdy } k = k' \\ 0 & \text{gdy } k \neq k'. \end{cases}$$

Przedstawiając równanie (6.10) w postaci

$$(6.12) \quad (\underline{S}_A \hat{\Psi} - \lambda \underline{S}_G \hat{\Psi}) \underline{R}^{-1} \underline{q}_k^{\pi} = 0$$

zauważmy, że wektorem charakterystycznym spełniającym to równanie jest  $\underline{q}_k = \underline{R}^{-1} \underline{q}_k^{\pi}$ . Łatwo pokazać, że są także spełnione warunki

$$(6.13) \quad \underline{q}_k \underline{S}_G \hat{\Psi} \underline{q}_{k'} = \begin{cases} 1 & \text{gdy } k = k' \\ 0 & \text{gdy } k \neq k' \end{cases} \text{ i}$$

$$\underline{q}_k \underline{S}_A \hat{\Psi} \underline{q}_{k'} = \begin{cases} \lambda_k & \text{gdy } k = k' \\ 0 & \text{gdy } k \neq k'. \end{cases}$$

Oznaczmy przez  $Q = (\underline{q}_1, \underline{q}_2, \dots, \underline{q}_{p-1})$  macierz wszystkich wektorów charakterystycznych  $\underline{q}_k$  i dokonajmy następującego przekształcenia

$$(6.14) \quad Z = Q \hat{\Psi} = Q' \underline{L} \underline{Y} \dots = \begin{bmatrix} \underline{q}'_1 \underline{L} \\ \underline{q}'_2 \underline{L} \\ \vdots \\ \underline{q}'_{p-1} \underline{L} \end{bmatrix} \underline{Y} \dots = \begin{bmatrix} \underline{l}'_1 \\ \underline{l}'_2 \\ \vdots \\ \underline{l}'_{p-1} \end{bmatrix} \underline{Y} \dots = \underline{l}'^{\pi} \underline{Y} \dots,$$

przyjmując oznaczenie  $\underline{q}'_k \underline{L} = \underline{l}'_k$ . Zatem wektor

$$\underline{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p-1} \end{bmatrix} = \underline{l}'^{\pi} \underline{Y} \dots,$$

jest wektorem nowych zmiennych zwanych składowymi głównymi.

Interesujące są składowe główne uzyskane z wektorów charakterystycznych odpowiadających największym pierwiastkom charakterystycznym  $\lambda_1, \lambda_2, \dots, \lambda_{p-1}$ , bowiem wobec równości (6.13) można powiedzieć, że każdy pierwiastek pochłania część zmienności między obiektami. Wartość każdego pierwiastka charakterystycznego można określić procentowo w następujący sposób:

$$\frac{\lambda_k}{\lambda_1 + \dots + \lambda_{p-1}} \cdot 100\% \quad (k = 1, \dots, p-1).$$

Tym samym można stwierdzić w jakim procencie w stosunku do wszystkich kontrastów każdy nowy kontrast wykazuje największą interakcję z miejscowościami bądź z latami. Biorąc pod uwagę dwa pierwsze kontrasty (składowe główne)  $Z_1$  i  $Z_2$  oraz stosując przekształcenie (6.14) można przedstawić rozmieszczenia poszczególnych miejscowości (lub lat) na płaszczyźnie w układzie współrzędnych  $Z_1$  i  $Z_2$ . Ze stratą informacji równą  $\left(1 - \frac{\lambda_1 + \lambda_2}{\lambda_1 + \dots + \lambda_{p-1}}\right) 100\%$  w stosunku do przestrzeni  $p-1$  wymiarowej można wówczas określić wzajemne położenie miejscowości (lub lat) w przestrzeni dwuwymiarowej.

## 7. Przykłady

**Przykład 1.** Do analizy statystycznej wzięto wyniki średnich plonów nasion w q/ha pięciu odmian łąbinu żółtego: 1. Bielański Pastewny, 2. As, 3. Popularny, 4. Mazowiecki, 5. Pomorski Pastewny z siedmiu miejscowości i siedmiu kolejnych lat (patrz [5] oraz materiały dotychczas nieopublikowane).

Średnie przedstawione są w tabeli 2 i 3.

Tabela 2

Odmiany Miejscowości	1	2	3	4	5	Średnie
Dąbrówka	18,70	18,39	17,79	17,69	17,77	18,07
Wysoka Włkp.	13,61	12,31	11,33	12,16	11,50	12,18
Głodowo	14,37	13,77	13,03	13,14	12,27	13,32
Bratoszewice	18,46	17,39	16,84	16,74	16,09	17,10
Marianowo	16,41	16,17	15,97	15,51	15,76	15,96
Cicibór	15,01	14,34	14,09	14,34	13,79	14,31
Łaziska	13,14	14,67	14,30	13,64	13,87	13,92
Średnie	15,67	15,29	14,76	14,75	14,43	14,98

Tabela 3

Odmiany Lata	1	2	3	4	5	Średnie
	1961	16,41	15,14	13,47	15,04	
1962	16,99	15,31	14,50	14,80	14,84	15,29
1963	11,71	15,21	13,44	14,34	13,00	13,54
1964	10,27	12,54	12,41	12,59	11,37	11,84
1965	19,90	16,33	16,53	17,01	15,28	17,01
1966	16,93	16,49	16,20	13,80	16,13	15,91
1967	17,50	16,01	16,79	15,64	16,10	16,41
Średnie	15,67	15,29	14,76	14,75	14,43	14,98

## Testowanie hipotez

## 1. Hipoteza o braku wpływu miejscowości:

$$SS_A = 4750,12 \text{ z } 6 \text{ stopniami swobody, } MS_A = 791,69$$

$$SS_C = 15399,45 \text{ z } 26 \text{ stopniami swobody, } MS_C = 427,76$$

$$F = 1,851; F_{0,05} = 2,36; F_{0,01} = 3,35.$$

Nie ma podstaw do odrzucenia hipotezy.

## 2. Hipoteza o braku wpływu lat:

$$SS_B = 3338,73 \text{ z } 6 \text{ stopniami swobody, } MS_B = 556,45$$

$$SS_C = 15399,45 \text{ z } 36 \text{ stopniami swobody, } MS_C = 427,76$$

$$F = 1,301; F_{0,05} = 2,36; F_{0,01} = 3,35.$$

Nie ma podstaw do odrzucenia hipotezy.

3. Hipoteza o braku interakcji miejscowości  $\times$  lata

$$SS_C = 15399,45 \text{ z } 36 \text{ stopniami swobody, } MS_C = 427,76$$

$$SS_E = 3160,50 \text{ z } 980 \text{ stopniami swobody, } MS_E = 3,23$$

$$F = 132,640; F_{0,05} = 1,47; F_{0,01} = 1,71.$$

Hipotezę odrzucamy.

4. Hipoteza o braku interakcji miejscowości  $\times$  odmiany

$$|\underline{S}_C| = 243 \ 550 \ 890 \quad |\underline{S}_A + \underline{S}_C| = 337 \ 636 \ 212$$

$$\wedge = 0,721 \ 341.$$

$$F = 0,476 \text{ z } 24 \text{ i } 166 \text{ stopniami swobody, } F_{0,05} = 1,63; F_{0,01} = 1,98.$$

Nie ma podstaw do odrzucenia hipotezy.

5. Hipoteza o braku interakcji lata  $\times$  odmiany

$$|\underline{S}_C| = 243\ 550\ 890 \quad |\underline{S}_B + \underline{S}_C| = 1\ 158\ 637\ 551$$

$$\wedge = 0,210205.$$

$F = 2,733$  z 24 i 166 stopniami swobody,  $F_{0,05} = 1,63$ ;  $F_{0,01} = 1,98$ .  
Hipotezę odrzucamy.

6. Hipoteza o braku interakcji miejscowości  $\times$  lata  $\times$  odmiany:

$$|\underline{S}_E| = 798\ 202\ 650\ 650 \quad |\underline{S}_C + \underline{S}_E| = 1\ 390\ 166\ 408\ 800$$

$$\wedge = 0,574178$$

$F = 4,042$  z 144 i 3894 stopniami swobody,  $F_{0,05} = 1,24$ ;  $F_{0,01} = 1,36$ .  
Hipotezę odrzucamy.

7. Hipoteza o braku różnic między odmianami:

Ponieważ nie ma podstaw do odrzucenia hipotezy o braku interakcji miejscowości  $\times$  odmiany, natomiast odrzucona została hipoteza o braku interakcji lata  $\times$  odmiany, więc jako macierz  $\hat{\Sigma}_1$  do statystyki  $T^2$  przyjmujemy macierz  $\frac{1}{I-1} \underline{S}_B$ .

Wówczas

$$T^2 = 41,104$$

$F = 5,138$  z 4 i 3 stopniami swobody,  $F_{0,05} = 9,12$ ;  $F_{0,01} = 28,71$ .  
Nie ma podstaw do odrzucenia hipotezy.

K o n t r a s t m a k s y m a l n y

Aby znaleźć maksymalny kontrast tworzymy najpierw układ kontrastów między poszczególnymi odmianami a odmianą trzecią (Popularny). Wektor kontrastów oraz macierz kowariancji wynoszą dla nich odpowiednio

$$\underline{M}'_1 \underline{y} \dots = \begin{bmatrix} 0,910 \\ 0,529 \\ -0,016 \\ -0,328 \end{bmatrix},$$

$$\hat{\Sigma}_1 = \begin{bmatrix} 33,86 & -0,91 & 3,06 & 4,31 \\ -0,91 & 6,27 & 5,42 & 3,20 \\ 3,06 & 5,42 & 12,48 & 1,36 \\ 4,31 & 3,20 & 1,36 & 3,94 \end{bmatrix}.$$

Wówczas na podstawie wzoru (5.1)

$$\underline{h}_{\max} = \begin{bmatrix} 0,18 \\ 0,77 \\ -0,29 \\ -0,80 \end{bmatrix}, \text{ skąd } \underline{M}_1 \underline{h}_{\max} = \begin{bmatrix} 0,18 \\ 0,77 \\ 0,14 \\ -0,29 \\ -0,80 \end{bmatrix}.$$

Maksymalny kontrast ( $\underline{M}_1 \underline{h}_{\max}$ ) rozgraniczył nam wszystkie odmiany na dwie grupy. Do grupy odmian dobrych zostały zaliczone odmiany: Bielański Pastewny, As, Popularny, natomiast do grupy odmian złych: Mazowiecki i Pomorski Pastewny.

Jeśli chcemy w odpowiednich proporcjach zalecać wszystkie odmiany z wyjątkiem najgorszej (Pomorski Pastewny), to wystarczy dokonać przesunięcia współczynników maksymalnego kontrastu tak, aby współczynnik odpowiadający tej odmianie otrzymał wartość zero oraz wyrazić wartość każdego tak przekształconego współczynnika odpowiadającego danej odmianie w procentach sumy wartości wszystkich współczynników. Otrzymujemy wówczas:

$$\underline{\tilde{\lambda}} = \begin{bmatrix} 0,98 \\ 1,57 \\ 0,94 \\ 0,51 \\ 0,00 \end{bmatrix} \quad \underline{\tilde{\lambda}}^* = \begin{bmatrix} 24,5\% \\ 39,2\% \\ 23,5\% \\ 12,8\% \\ 0,0\% \end{bmatrix} \begin{array}{l} \text{- Bielański Pastewny} \\ \text{- As} \\ \text{- Popularny} \\ \text{- Mazowiecki} \\ \text{- Pomorski Pastewny.} \end{array}$$

W przypadku, gdy chcemy zalecać tylko odmiany, które znalazły się w grupie odmian dobrych (Bielański Pastewny, As, Popularny) budujemy układ nowych kontrastów (Macierz  $\underline{M}_1$ ) między każdą z odmian z grupy dobrych, a wszystkimi odmianami z grupy złych. W tym przypadku układ ten będzie miał postać:

$$\underline{M}_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

Wektor kontrastów oraz macierz kowariancji są dla nich odpowiednio:

$$\underline{M}_1 \underline{y} \dots = \begin{bmatrix} 2,164 \\ 1,402 \\ 0,344 \end{bmatrix},$$

$$\hat{\Sigma}_1 = \begin{bmatrix} 125,11 & -16,48 & 4,39 \\ -16,48 & 9,73 & 1,90 \\ 4,39 & 1,90 & 19,14 \end{bmatrix} .$$

Wówczas na podstawie wzoru (5.1)

$$\underline{h}_{\max} = \begin{bmatrix} 0,048 \\ 0,228 \\ -0,016 \end{bmatrix} , \text{ skąd } \underline{M}_1 \underline{h}_{\max} = \begin{bmatrix} 0,096 \\ 0,457 \\ -0,031 \\ -0,261 \\ -0,261 \end{bmatrix} .$$

Dokonyjemy teraz przesunięcia współczynników maksymalnego kontrastu tak, aby współczynniki odpowiadające odmianom zaliczonym do grupy złych otrzymały wartość zero. Ponadto nowo uzyskane wartości współczynników wyrażamy w procentach. Otrzymujemy wówczas:

$$\underline{\lambda} = \begin{bmatrix} 0,357 \\ 0,718 \\ 0,229 \\ 0,000 \\ 0,000 \end{bmatrix} , \quad \underline{\lambda}^{\%} = \begin{bmatrix} 27,35\% \\ 55,05\% \\ 17,60\% \\ 0,00\% \\ 0,00\% \end{bmatrix} \begin{array}{l} \text{- Bielański Pastewny} \\ \text{- As} \\ \text{- Popularny} \\ \text{- Mazowiecki} \\ \text{- Pomorski Pastewny} \end{array}$$

Składowe wektora  $\underline{\lambda}^{\%}$  mogą stanowić procentowo określone propozycje co do udziału każdej z odmian w areale uprawnym na terenie objętym badaniami.

### B a d a n i a i n t e r a k c j i

Z dwóch testowanych hipotez dotyczących braku interakcji odmian z miejscowościami bądź latami, odrzucona została na poziomie istotności  $\alpha = 0,05$  tylko hipoteza dotycząca braku interakcji odmian z latami. Stosując w tym wypadku analizę składowych głównych możemy znaleźć nowe kontrasty o maksymalnej interakcji z latami. Można także uzyskać graficzne przedstawienie średnich z poszczególnych lat na płaszczyźnie.

Rozwiązując równanie wyznacznikowe  $|\underline{S}_B \hat{\psi} - \lambda \underline{S}_C \hat{\psi}| = 0$ , gdzie  $\underline{S}_B \hat{\psi}$  i  $\underline{S}_C \hat{\psi}$  są odpowiednio macierzami sum kwadratów i iloczynów dla interakcji kontrastów z latami oraz dla ich interakcji z miejscowościami i latami, otrzymujemy następujące pierwiastki charakterystyczne (w nawiasach podajemy procentową wartość pierwiastka w stosunku do sumy ich wartości):

$$\lambda_1 = 1,140533 \quad (52,9\%),$$

$$\lambda_2 = 0,768493 \quad (35,7\%),$$

$$\lambda_3 = 0,191151 \quad (8,9\%),$$

$$\lambda_4 = 0,054834 \quad (2,5\%),$$

---


$$\text{Razem} \quad 2,155011 \quad (100,0\%).$$

Ponieważ  $\lambda_1 + \lambda_2 = 88,6\%$ , więc strata informacji odnośnie zmienności między średnimi dla lat przy przejściu z pierwotnych kontrastów ( $\underline{L}$ ) do pierwszych dwóch składowych głównych (nowych kontrastów  $\underline{1}_1^{\pi}$  i  $\underline{1}_2^{\pi}$ ) wyniesie  $100 - 88,6 = 11,4\%$ . Jest to strata stosunkowo niewielka, pozwalająca wykorzystać nowe zmienne do graficznego przedstawienia wyników na płaszczyźnie.

Dwa pierwsze wektory charakterystyczne, tzn. wektory odpowiadające największym pierwiastkom charakterystycznym  $\lambda_1$  i  $\lambda_2$  przyjmują następujące wartości:

$$\underline{q}_1' = (0,077553 \quad 0,021552 \quad 0,062726 \quad -0,034247),$$

$$\underline{q}_2' = (0,028676 \quad -0,043679 \quad -0,087536 \quad 0,083006).$$

Mnożąc wektory  $\underline{q}_1'$  i  $\underline{q}_2'$  przez macierz współczynników kontrastów  $\underline{L}$  uzyskamy współczynniki dwóch nowych kontrastów odmian o największej interakcji z latami

$$\underline{1}_1^{\pi} = (0,069365 \quad 0,001323 \quad 0,027653 \quad -0,073387 \quad -0,024955)'$$

$$\underline{1}_2^{\pi} = (0,025648 \quad -0,044239 \quad -0,065275 \quad 0,100628 \quad -0,016761)'$$

Dla ułatwienia interpretacji wyrażmy powyższe współczynniki kontrastów w procentach w stosunku do odpowiedniej sumy współczynników o jednakowym znaku.

$$\tilde{1}_1^{\pi} : (70,6\% \quad 1,3\% \quad 28,1\% \quad -74,6\% \quad -25,4\%)$$

$$\tilde{1}_2^{\pi} : (20,3\% \quad -35,0\% \quad -51,7\% \quad 79,7\% \quad -13,3\%).$$

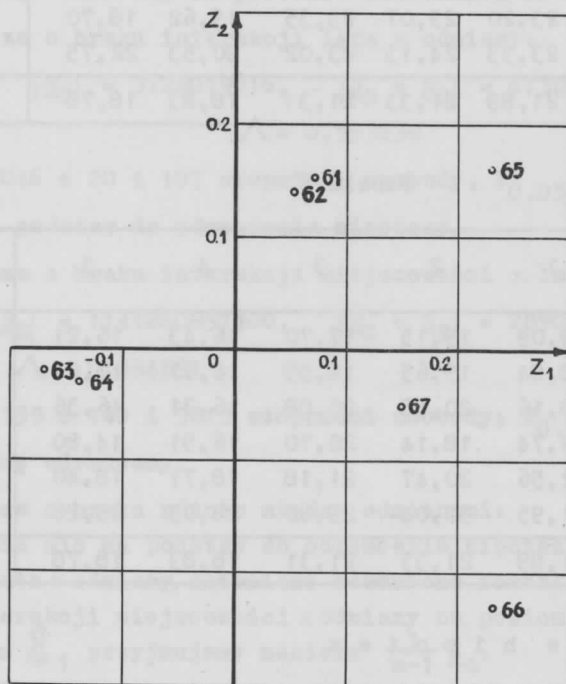
Współczynniki te wskazują, w jakim stopniu każda z odmian uwzględniana jest w danym kontraście.

Mnożąc składowe wektorów  $\underline{1}_1^{\pi}$  i  $\underline{1}_2^{\pi}$  przez średnie dla odmian z poszczególnych lat otrzymujemy dla każdego roku następujące wartości nowych kontrastów o największej interakcji, które oznaczamy jako  $Z_1$  i  $Z_2$ :



Rok	$Z_1$	$Z_2$
1961	0,070	0,146
1962	0,143	0,052
1963	-0,173	-0,025
1964	-0,135	-0,026
1965	0,229	0,165
1966	0,229	-0,234
1967	0,150	-0,051

Mając współrzędne ( $Z_1, Z_2$ ) dla poszczególnych lat możemy każdy rok przedstawić jako punkt na najlepiej dopasowanej płaszczyźnie. Otrzymuje się ostatecznie obraz jak na rysunku 1. Konfrontacja tego rysunku z warunkami meteorologicznymi występującymi w latach objętych badaniami umożliwia właściwą interpretację reagowania poszczególnych odmian na zmienność tych warunków.



Rys. 1. Pierwsza i druga składowa główna

Przykład 2. Do analizy statystycznej wzięto wyniki średnich plonów nasion w q/ha pięciu odmian grochu: 1. Ceser, 2. Wiktorja Łagiewnicki, 3. Kujawski Wczesny, 4. Wiktorja PZHR, 5. Czwartkowy z ośmiu miejscowości i sześciu kolejnych lat (patrz [2], [3]). Średnie przedstawione są w tabelach 4 i 5.

Tabela 4

Odmiany Miejscowości	1	2	3	4	5	Średnie
Banaski	24,95	24,90	20,08	20,73	21,90	22,51
Radostowo	22,40	19,80	20,12	18,57	18,08	19,79
Fałęcin	16,97	16,72	19,02	14,00	15,58	16,46
Sobótka	20,22	16,97	20,18	17,67	15,28	18,06
Zabornia	19,07	19,25	19,75	15,03	15,68	17,76
Czesławice	25,10	25,83	24,93	24,52	22,25	24,53
Wojciechowice	23,20	23,07	23,35	19,62	18,70	21,59
Majków	23,33	24,13	23,02	20,53	22,75	22,73
Średnie	21,89	21,33	21,31	18,83	18,78	20,43

Tabela 5

Odmiany Lata	1	2	3	4	5	Średnie
1957	19,09	19,15	17,70	16,43	16,21	17,72
1958	18,84	19,65	18,59	16,61	17,64	18,27
1959	19,16	20,55	20,08	16,31	16,36	18,49
1960	19,74	18,14	20,70	16,91	14,80	18,06
1961	22,56	20,47	21,18	18,71	18,40	20,26
1962	31,95	30,04	29,60	28,03	29,26	29,78
Średnie	21,89	21,33	21,31	18,83	18,78	20,43

### Testowanie hipotez

#### 1. Hipoteza o braku wpływu miejscowości:

$$SS_A = 8503,70 \text{ z } 7 \text{ stopniami swobody, } MS_A = 1214,81$$

$$SS_C = 39500,62 \text{ z } 35 \text{ stopniami swobody, } MS_C = 1128,59$$

$$F = 1,076; F_{0,05} = 2,30; F_{0,01} = 3,21.$$

Nie ma podstaw do odrzucenia hipotezy.

## 2. Hipoteza o braku wpływu lat:

$$SS_B = 21759,47 \text{ z } 5 \text{ stopniami swobody, } MS_B = 4351,89$$

$$SS_C = 39500,62 \text{ z } 35 \text{ stopniami swobody, } MS_C = 1128,59$$

$$F = 3,856; F_{0,05} = 2,49; F_{0,01} = 3,61.$$

Hipotezę odrzucamy.

3. Hipoteza o braku interakcji miejscowości  $\times$  lata:

$$SS_C = 39500,62 \text{ z } 35 \text{ stopniami swobody, } MS_C = 1128,59$$

$$SS_E = 3456,00 \text{ z } 960 \text{ stopniami swobody, } MS_E = 3,60$$

$$F = 313,497; F_{0,05} = 1,47; F_{0,01} = 1,71.$$

Hipotezę odrzucamy.

4. Hipoteza o braku interakcji miejscowości  $\times$  odmiany:

$$|\underline{S}_C| = 3728019016 \quad |\underline{S}_A + \underline{S}_C| = 13872008403$$

$$\wedge = 0,268744$$

$$F = 1,834 \text{ z } 28 \text{ i } 117 \text{ stopniami swobody; } F_{0,05} = 1,57; F_{0,01} = 1,89.$$

Hipotezę odrzucamy na poziomie istotności  $\alpha = 0,05$ .

5. Hipoteza o braku interakcji lata  $\times$  odmiany:

$$|\underline{S}_C| = 3728019016, \quad |\underline{S}_B + \underline{S}_C| = 6738546665,$$

$$\wedge = 0,553238$$

$$F = 1,046 \text{ z } 20 \text{ i } 107 \text{ stopniami swobody, } F_{0,05} = 1,68; F_{0,01} = 2,06.$$

Nie ma podstaw do odrzucenia hipotezy.

6. Hipoteza o braku interakcji miejscowości  $\times$  lata  $\times$  odmiany:

$$|\underline{S}_E| = 1141260857400, \quad |\underline{S}_C + \underline{S}_E| = 2890067533200$$

$$\wedge = 0,394891$$

$$F = 7,159 \text{ z } 140 \text{ i } 3813 \text{ stopniami swobody; } F_{0,05} = 1,24; F_{0,01} = 1,36.$$

Hipotezę odrzucamy.

## 7. Hipoteza o braku różnic między odmianami:

Ponieważ nie ma podstaw do odrzucenia hipotezy o braku interakcji lata  $\times$  odmiany, natomiast odrzucona została hipoteza o braku interakcji miejscowości  $\times$  odmiany na poziomie  $\alpha = 0,05$ , to za macierz  $\sum_1$  przyjmujemy macierz  $\frac{1}{m-1} \underline{S}_A$ .

Wówczas

$$T^2 = 99,778$$

$$F = 14,254 \text{ z } 4 \text{ i } 4 \text{ stopniami swobody, } F_{0,05} = 6,39; F_{0,01} = 15,98.$$

Hipotezę odrzucamy na poziomie  $\alpha = 0,05$ .

## K o n t r a s t   m a k s y m a l n y

Aby znaleźć maksymalny kontrast tworzymy najpierw układ kontrastów między poszczególnymi odmianami a odmianą o najniższym średnim plonie (Czwartkowy). Wektor kontrastów oraz macierz kowariancji są dla nich odpowiednio

$$\underline{M}_1' \underline{Y} \dots = \begin{bmatrix} 3,111 \\ 2,554 \\ 2,527 \\ 0,054 \end{bmatrix} ,$$

$$\hat{\Sigma}_1 = \begin{bmatrix} 14,42 & 4,67 & 10,06 & 12,09 \\ 4,67 & 8,93 & 2,86 & 4,44 \\ 10,06 & 2,86 & 32,21 & 13,17 \\ 12,09 & 4,44 & 13,17 & 18,06 \end{bmatrix} .$$

Wówczas na podstawie wzoru (5.1)

$$\underline{h}_{\max} = \begin{bmatrix} 0,40 \\ 0,24 \\ 0,09 \\ -0,39 \end{bmatrix} , \quad \text{skąd} \quad \underline{M}_1 \underline{h}_{\max} = \begin{bmatrix} 0,40 \\ 0,24 \\ 0,09 \\ -0,39 \\ -0,34 \end{bmatrix} .$$

Maksymalny kontrast ( $\underline{M}_1 \underline{h}_{\max}$ ) rozgraniczył nam wszystkie odmiany na dwie grupy. Do grupy odmian dobrych zostały zaliczone odmiany: Ceser, Wiktorja Łagiewnicki, Kujawski Wczesny, natomiast do grupy odmian złych: Wiktorja PZHR i Czwartkowy.

Jeśli tylko eliminujemy odmianę najgorszą (Wiktorja PZHR), to wystarczy dokonać przesunięcia współczynników maksymalnego kontrastu tak, aby współczynnik odpowiadający tej odmianie otrzymał wartość zero oraz wyrazić wartość każdego tak przekształconego współczynnika odpowiadającego danej odmianie w procentach sumy wartości wszystkich współczynników. Otrzymujemy wówczas

$$\underline{\tilde{\lambda}} = \begin{bmatrix} 0,79 \\ 0,63 \\ 0,48 \\ 0,00 \\ 0,05 \end{bmatrix} , \quad \underline{\tilde{\lambda}}^{\#} = \begin{bmatrix} 40,5\% \\ 32,3\% \\ 24,6\% \\ 0,0\% \\ 2,6\% \end{bmatrix} \begin{array}{l} \text{- Ceser} \\ \text{- Wiktorja Łagiewnicki} \\ \text{- Kujawski Wczesny} \\ \text{- Wiktorja PZHR} \\ \text{- Czwartkowy.} \end{array}$$

W przypadku, gdy eliminujemy wszystkie odmiany zaliczone do grupy odmian złych (Wiktorja PZHR, Czwartkowy) budujemy układ no-

wych kontrastów (macierz  $\underline{M}_1$ ) między każdą z odmian z grupy dobrych, a wszystkimi odmianami z grupy złych. W tym przypadku układ ten będzie miał postać:

$$\underline{M}_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix} .$$

Wektor kontrastów oraz macierz kowariancji są dla nich odpowiednio:

$$\underline{M}_1 \underline{y} \dots = \begin{bmatrix} 6,168 \\ 5,054 \\ 5,000 \end{bmatrix} ,$$

$$\underline{\Sigma}_1 = \begin{bmatrix} 27,41 & 3,70 & 7,79 \\ 3,70 & 36,02 & -5,71 \\ 7,79 & -5,71 & 94,23 \end{bmatrix} .$$

Wówczas na podstawie wzoru (5.1)

$$\underline{h}_{\max} = \begin{bmatrix} 0,195 \\ 0,127 \\ 0,045 \end{bmatrix} , \quad \text{skąd} \quad \underline{M}_1 \underline{h}_{\max} = \begin{bmatrix} 0,390 \\ 0,254 \\ 0,090 \\ -0,367 \\ -0,367 \end{bmatrix} .$$

Dokonujemy teraz przesunięcia współczynników maksymalnego kontrastu tak, aby współczynniki odpowiadające odmianom zaliczonym do grupy złych otrzymały wartość zero. Ponadto nowo uzyskane wartości współczynników wyrażemy w procentach. Otrzymujemy wówczas:

$$\underline{\tilde{\lambda}} = \begin{bmatrix} 0,70 \\ 0,62 \\ 0,46 \\ 0,00 \\ 0,00 \end{bmatrix} \quad \underline{\tilde{\lambda}}^{\%} = \begin{bmatrix} 41,2\% \\ 33,9\% \\ 24,9\% \\ 0,0\% \\ 0,0\% \end{bmatrix} \quad \begin{array}{l} - \text{Ceser} \\ - \text{Wiktoria Łąglewnicki} \\ - \text{Kujawski Wczesny} \\ - \text{Wiktoria PZHR} \\ - \text{Czwartkowy.} \end{array}$$

Składowe wektora  $\underline{\tilde{\lambda}}^{\%}$  mogą stanowić procentowo określone propozycje co do udziału każdej z odmian w areale uprawnym na terenie objętym badaniami.

Przedział ufności Scheffe'go dla maksymalnego kontrastu jest określony nierównością

$$0,68 \leq \Psi \leq 3,48 .$$

### B a d a n i e i n t e r a k c j i

Ponieważ istotna okazała się interakcja odmian z miejscowościami, zatem stosując analizę składowych głównych warto znaleźć nowe kontrasty między odmianami o możliwie wysokiej interakcji z miejscowościami. Pierwiastki równania  $|\underline{S}_A \hat{\Psi} - \lambda \underline{S}_C \hat{\Psi}| = 0$  przedstawiają się następująco:

$$\begin{aligned} \lambda_1 &= 0,736025 \quad (44,9\%), \\ \lambda_2 &= 0,555053 \quad (33,9\%), \\ \lambda_3 &= 0,200527 \quad (12,2\%), \\ \lambda_4 &= 0,148109 \quad (9,0\%), \end{aligned}$$

---


$$\text{Razem} \quad 1,639714 \quad (100\%) .$$

Widzimy, że pierwsze dwa pierwiastki pochłaniają 78,7% całkowitej zmienności między miejscowościami. Przyjmujemy, że to wystarczy, aby korzystać z pierwszych dwóch składowych głównych.

Pierwiastkom charakterystycznym  $\lambda_1$  i  $\lambda_2$  odpowiadają następujące wektory charakterystyczne:

$$\begin{aligned} \underline{q}_1' &= (0,048043 \quad 0,048841 \quad -0,021540 \quad -0,054887) \\ \underline{q}_2' &= (0,021476 \quad -0,006855 \quad -0,048833 \quad 0,070176) . \end{aligned}$$

Stąd nowe kontrasty między odmianami o największej interakcji z miejscowościami mają postać:

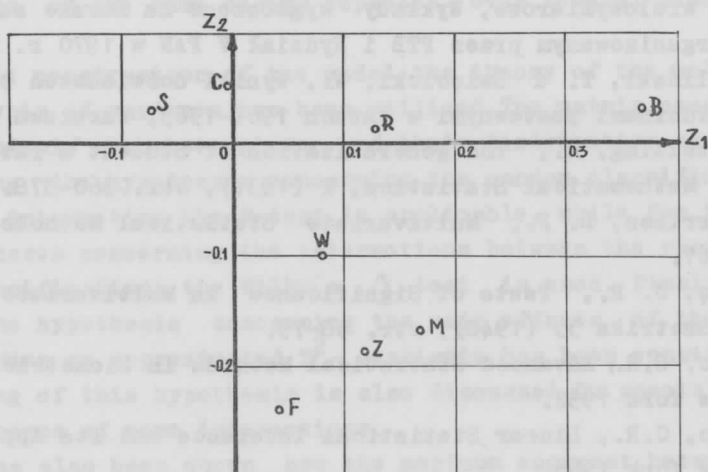
$$\begin{aligned} \underline{1}_1^{\#} &= (0,042971 \quad 0,031555 \quad -0,042429 \quad -0,054859 \quad 0,022763)' \\ \underline{1}_2^{\#} &= (0,019209 \quad -0,010765 \quad -0,042686 \quad 0,066743 \quad -0,032501)' . \end{aligned}$$

Poniżej zapisane są wartości procentowe tych współczynników

$$\begin{aligned} \tilde{1}_1^{\#} &: (44,2\% \quad 32,4\% \quad -43,6\% \quad -56,4\% \quad 23,4\%) \\ \tilde{1}_2^{\#} &: (22,4\% \quad -12,5\% \quad -49,7\% \quad 77,6\% \quad -37,8\%) . \end{aligned}$$

Mnożąc składowe wektorów  $\underline{1}_1^{\#}$  i  $\underline{1}_2^{\#}$  przez średnie dla odmian z poszczególnych miejscowości otrzymujemy dla każdej miejscowości następujące wartości nowych kontrastów o największej interakcji, które oznaczamy jako  $Z_1$  i  $Z_2$ :

Miejscowości	$Z_1$	$Z_2$
Banaski	0,366	0,026
Radostowo	0,127	0,010
Falęcin	0,036	-0,238
Sobótka	-0,074	0,027
Zabornia	0,121	-0,190
Czesławice	-0,003	0,053
Wojciechowice	0,084	-0,098
Majków	0,174	-0,165



Rys. 2. Pierwsza i druga składowa główna

Na rysunku 2 pokazane jest rozmieszczenie wszystkich w/w miejscowości w odpowiednio dopasowanej płaszczyźnie. Współrzędnymi otrzymanego układu są pierwsze dwie składowe główne, czyli kontrasty  $Z_1$  i  $Z_2$ . Przy interpretacji merytorycznej uzyskanych wyników dotyczących nowych kontrastów między odmianami i związanego z nimi obrazu miejscowości na rysunku 2, należy wziąć pod uwagę czynniki charakteryzujące te miejscowości (np. gleby w miejscowościach, warunki klimatyczne dla tych miejscowości). Tutaj ogra-

niczono się jedynie do pokazania metod analizy statystycznej i nie wchodzi się w zasadniczą interpretację merytoryczną.

#### Literatura cytowana

- [1] Anderson, T. W., An Introduction to Multivariate Statistical Analysis, New York 1958.
- [2] Barbacki, S. i Caliński, T., Wyniki doświadczeń odmianowych z grochem, peluszką i bobikiem wykonanych w latach 1957-1960, Warszawa 1963.
- [3] Byszewski, W., Kalinowski-Zdun, M., Błaszczyk-Ostrowska, D., Wyniki doświadczeń z odmianami grochu, peluszką i bobiku przeprowadzonych w latach 1961-1964, Warszawa 1968.
- [4] Caliński, T., Wielozmienna analiza wariancji i pokrewne metody wielowymiarowe, Wykłady wygłoszone na Kursie szkoleniowym zorganizowanym przez PTB i Wydział V PAN w 1970 r. w Lublinie.
- [5] Caliński, T. i Świącicki, W., Wyniki doświadczeń odmianowych z łubinami pastewnymi w latach 1961-1965, Warszawa 1970
- [6] Hotelling, H., The generalization of Student's ratio, *Annals of Mathematical Statistics*, 2 (1931), str. 360-378.
- [7] Morrison, D. F., *Multivariate Statistical Methods*, New York 1967.
- [8] Rao, C. R., Tests of Significance in Multivariate Analysis, *Biometrika* 35 (1948), str. 58-79.
- [9] Rao, C.R., *Advanced Statistical Methods in Biometric Research*, New York 1952.
- [10] Rao, C.R., *Linear Statistical Inference and Its Applications*, New York 1965.
- [11] Scheffe, H., *The Analysis of Variance*, New York 1959.
- [12] Wilks, S. S., Certain generalizations in the analysis of variance, *Biometrika* 24 (1932), str. 471-494.
- [13] Wishart, J., The generalized product moment distribution in samples from a normal multivariate population, *Biometrika* 20 (1928), str. 32-52.



T. Caliński, S. Czajka, Z. Kaczmarek (Poznań)

MULTIVARIATE ANALYSIS OF VARIANCE APPLIED TO THE RESULTS  
OF A SERIES OF ANALOGICAL EXPERIMENTS

Summary

A mixed model for the analysis of variance in the case of there-way orthogonal classifications, one fixed and two random, has been presented. The model is applicable in the analysis of experiments of the same design repeated over several places and years.

In the construction of the model the theory of the multivariate analysis of variance has been utilized. The matrix components of the analysis has been given and their distribution examined. For testing the hypotheses concerning the random classifications and their interaction the F-test is applicable, while for testing the hypotheses concerning the interactions between the random and fixed classifications the Wilks's  $\Lambda$  test is used. Finally, for testing the hypothesis concerning the main effects of the fixed classification an approximated  $T^2$ -statistic has been constructed. The testing of this hypothesis is also discussed for special cases of the absence of some interactions.

It has also been shown how the maximum contrast between the main fixed effects can be found. Its particular application for series of variety trials is indicated and its use as a mean of evaluating the varieties is discussed. Scheffé's simultaneous test and confidence interval for any contrast has also been given.

It has also been shown how principal components may be used for the analysis of the interaction between variety contrast and places or years.