Listy Biometryczne - Biometrical Letters
Vol. 32 (1995), No. 2, 101-129

Metody analizy serii doświadczeń wielokrotnych i wieloletnich zakładanych w układzie o blokach niekompletnych

Stanisław Czajka

Katedra Metod Matematycznych i Statystycznych, Akademia Rolnicza
Wojska Polskiego 28, 60-637 Poznań

Streszczenie

W pracy podany jest model analizy wyników serii doświadczeń zakładanych z tymi samymi genotypami w jednakowym układzie blokowym w wielu miejscowościach przez okres kilku lat. Dla tego modelu przedstawiona jest analiza wariancji oraz testy dla hipotez ogólnych dotyczących interesujących źródeł zmienności. Rozpatrywane są także procedury testów jednoczesnych dla testowania różnych układów hipotez szczegółowych. Ponadto zastosowano metodę regresji dla wyjaśnienia zależności interakcji genotypowo-środowiskowej od średniej środowiskowej. Praktyczne zastosowanie prezentowanych w części teoretycznej metod pokazano na konkretnym przykładzie serii doświadczeń odmianowych.

1. Wstęp

Analiza statystyczna pojedynczego doświadczenia, którego obiektami są odmiany, rody, linie rodzicielskie, mieszańce (ogólnie genotypy) dostarcza jedynie informacji o zachowaniu się porównywanych genotypów w warunkach glebowo-klimatycznych istniejących w środowisku, w którym doświadczenie to było założone. Ponieważ z reguły żądamy informacji o zachowaniu się genotypów w rozmaitych warunkach, tak glebowych jak i pogodowych, winno się zakładać doświadczenia wieloletnie i to nie w jednej lecz w wielu miejscowościach. Tak prowadzone doświadczenia nazywane są doświadczeniami wielokrotnymi i wieloletnimi lub seriami doświadczeń powtarzanych w wielu miejscowościach przez

Słowa kluczowe: analiza serii doświadczeń, wielozmienna analiza wariancji, testowanie hipotez, analiza regresji.

okres kilku lat (por. Neyman, 1932; Elandt, 1956; Caliński, 1967; Caliński, Czajka i Kaczmarek, 1987a,b).

Metody opracowywania wyników serii doświadczeń wielokrotnych i wieloletnich są przedmiotem zainteresowania rolników i biometryków już od przeszło pięćdziesięciu lat. Bogata literatura na ten temat sprowadza się w głównej mierze do analizy doświadczeń prowadzonych w najprostszych układach blokowych, przeważnie kompletnych.

Z punktu widzenia oceny wartości adaptacyjnej badanych w doświadczeniach genotypów, najbardziej interesującą jest interakcja genotypowo-środowiskowa, oznaczana symbolem GE (patrz na przykład Comstock i Moll, 1963; Hinkelmann, 1974; Freeman, 1973, 1975; Hill, 1975; Wood, 1976; Shorter i in., 1977; Freeman i Crisp, 1979; Lin, 1982; Cox, 1984; a także Freeman, 1985). Nazwą tą określamy zjawisko polegające na tym, że reakcja fenotypowa roślin (organizmów) na zmianę środowiska nie jest jednakowa dla wszystkich genotypów. W konsekwencji różnice obserwowane między porównywanymi genotypami zależą od środowiska, w którym to porównanie jest przeprowadzone.

W analizie serii doświadczeń wykorzystywane były najrozmaitsze metody statystyczne, z których znaczna część została zaproponowana już w latach trzydziestych naszego wieku przez Neymana (1932) oraz Yatesa i Cochrana (1938).

Analiza wariancji, jako podstawowa metoda przy opracowywaniu serii doświadczeń, została w pierwszym rzędzie zaadoptowana przez Neymana (1932), Yatesa i Cochrana (1938), a następnie przez wielu autorów takich jak Federer (1955), Cochran i Cox (1957), Finney (1964), a także Steel i Torrie (1980).

Również podejście regresyjne zainicjowane zostało przez Neymana (1932) oraz Yatesa i Cochrana (1938). Było one później rozwijane przede wszystkim przez Finlaya i Wilkinsona (1963), Eberharta i Russella (1966), Tai (1971), Freemana (1973) oraz Wrighta (1976). W podejściu tym zakładano, że interakcja genotypowo-środowiskowa może być całkowicie (lub niemal całkowicie) wyjaśniona za pomocą współczynników regresji. Oczywiście w sytuacji, gdy tylko nieznaczna część interakcji mogła być wyjaśniona w ten sposób, metoda regresyjna okazywała się mało efektywna dla charakteryzowania genotypów pod względem ich stabilności lub adaptacyjności do różnych warunków środowiskowych.

Możliwości wykorzystania komponentów wariancyjnych do oszacowania udziału każdego z badanych genotypów w interakcji ze środowiskami poprzez podział sumy kwadratów dla interakcji zaproponowali Caliński (1960) i niezależnie Wricke (1962, 1964, 1965) oraz nieco później Shukla (1972).

Podejście wielozmienne, wykorzystujące dekompozycję składników interakcji za pomocą zmiennych kanonicznych bądź składowych głównych, zostało zaadoptowane do analizowania serii doświadczeń przez takich autorów jak Okuno i in. (1971), Freeman i Dowker (1973), Freeman (1973) i Kempton (1984).

We wszystkich tych podejściach punktem startu był zawsze model klasycznej analizy wariancji (ANOVA), często modyfikowany i rozszerzany w celu zastosowania pożądanych metod z tego bądź innego podejścia. Okazało się jednak później, że każda z proponowanych do analizy metod może logicznie wynikać z jednego wspólnego modelu, a mianowicie modelu wielozmiennej analizy wariancji (MANOVA). Model taki został zaproponowany i szczegółowo omówiony przez Calińskiego i in. (1987a). Dotyczył on serii doświadczeń zakładanych w klasycznym układzie o blokach kompletnych. Interpretacja wyników przykładowej analizy opartej na wspomnianym modelu została przedstawiona w pracy Calińskiego i in. (1987b). Model oraz szczegółową analizę doświadczeń wielokrotnych zakładanych w blokach niekompletnych przedstawił w swojej pracy Kaczmarek(1986).

W pracy niniejszej dokonano uogólnienia modelu z pracy Calińskiego i in. (1987a) na wypadek doświadczeń wielokrotnych i wieloletnich zakładanych w dowolnym układzie blokowym, zarówno kompletnym jak i niekompletnym.

Celem pracy jest przedstawienie metod statystycznej analizy serii doświadczeń prowadzonych z tymi samymi genotypami w takim samym (spójnym) układzie blokowym w wielu miejscowościach przez okres kilku lat. Analiza ta odnosi się zarówno do tak zwanej serii ortogonalnej doświadczeń jak i nieortogonalnej. Przez serię ortogonalną rozumiemy taką, w której dwukierunkowa klasyfikacja doświadczeń według miejscowości i lat jest pełna, to znaczy doświadczenia prowadzone w danej miejscowości powtarzane są we wszystkich latach i wszystkie one wchodzą do analizy. Natomiast seria nieortogonalna charakteryzuje się tym, że doświadczenia nie wypełniają całkowicie wspomnianej klasyfikacji według lat i miejscowości.

2. Matematyczny model obserwacji

Przyjmijmy, że plon i-tego ($i = 1,2,...,I$) genotypu w j-tej ($j = 1,2,...,J$) miejscowości i k-tym ($k = 1,2,...,K$) roku z poletka w l-tym ($l = 1,2,...,L$) bloku doświadczenia założonego w układzie o blokach niekompletnych ma postać

$$y_{ijkl} = m_i(j,k,l) + e_{ijkl} ,$$

gdzie $m_i(j,k,l)$ oznacza średni "prawdziwy" plon, a e_{ijkl} oznacza "błąd" (odchylenie losowe) danej obserwacji y_{ijkl}. Przyjmujemy założenie, że błędy losowe e_{ijkl} pojedynczych obserwacji są zmiennymi losowymi o rozkładach niezależnych i identycznych, ze średnią zero i wspólną wariancją σ_e^2. Zakładamy ponadto, że rozkłady te są niezależne od średnich prawdziwych $m_i(j,k,l)$.

Z analizy pojedynczego doświadczenia otrzymujemy tak zwane "średnie plony poprawione" (ze względu na stosowany układ blokowy) dla wszystkich I genoty-

pów. Plon taki dla genotypu i z doświadczenia w miejscowości j oraz roku k można przedstawić w postaci modelu

$$y_{ijk} = m_i(j,k) + e_{ijk} \quad , \tag{2.1}$$

gdzie $m_i(j,k)$ oznacza średni plon "prawdziwy" i-tego genotypu z doświadczenia w j-tej miejscowości i k-tym roku, a e_{ijk} oznacza odpowiedni "błąd średni ważony" dla genotypu i z doświadczenia w miejscowości j oraz roku k, będący funkcją liniową błędów e_{ijkl} określoną przez stosowany układ blokowy.

Wyjaśnienia wymaga tutaj użycie pojęcia średniego plonu prawdziwego. Prawdziwość tę odnosi się do warunków danego doświadczenia określonych nie tylko przez daną miejscowość i dany rok, ale również przez środowisko konkretnego pola wybranego pod doświadczenie. Te warunki środowiskowe mogą się zmieniać w obrębie danej miejscowości i danego roku w sposób losowy, niezależny od eksperymentatora. Wobec tego wielkości $m_1(j,k)$, $m_2(j,k),\ldots,m_I(j,k)$ traktujemy jako zmienne losowe, o odpowiednich wartościach oczekiwanych $\mu_1(j,k)$, $\mu_2(j,k),\ldots,\mu_I(j,k)$, oraz wariancjach i kowariancjach $\sigma_{ii'}(j,k)$, $i,i' = 1,2,\ldots,I$.

Parametry te określają plonowanie genotypów w j-tej miejscowości i k-tym roku w sensie ogólnym, w przekroju wszystkich warunków środowiskowych danej miejscowości i danego roku.

Przy przyjętych powyżej założeniach wektory błędów średnich

$$\mathbf{e}_{jk} = [e_{1jk}, e_{2jk},\ldots,e_{Ijk}]' \tag{2.2}$$

mają dla każdego j oraz k zerowy wektor wartości oczekiwanych, $\mathrm{E}(\mathbf{e}_{jk}) = \mathbf{0}$, i macierz dyspersji $\mathrm{D}(\mathbf{e}_{jk}) = \sigma_e^2\,\mathbf{\Omega}_e$, gdzie $\mathbf{\Omega}_e$ jest macierzą Tochera (1952) zależną od zastosowanego układu blokowego. Ponadto rozkład wektora $\mathbf{e}_{jk}$ dla każdego j oraz k jest niezależny od wektora średnich prawdziwych dla genotypów

$$\mathbf{m}(j,k) = [m_1(j,k),\ m_2(j,k),\ldots,m_I(j,k)]' \quad , \tag{2.3}$$

którego wartością oczekiwaną jest

$$\mathrm{E}\{\mathbf{m}(j,k)\} = [\mu_1(j,k),\ \mu_2(j,k),\ldots,\mu_I(j,k)]' \quad ,$$

oraz macierzą dyspersji jest

$$\mathrm{D}\{\mathbf{m}(j,k)\} = \mathbf{\Sigma}_{m(j,k)} = [\sigma_{ii'}(j,k)] \quad .$$

Ponieważ wyniki doświadczeń z różnych miejscowości i różnych lat można uważać za niezależne, przyjmujemy, że rozkłady powyższych wektorów losowych, tak (2.2) jak i (2.3), są dla różnych par (j,k) niezależne.

Uzupełniając powyższe założenia zauważmy jeszcze, że aby dla rozpatrywanego modelu można było zastosować technikę analizy wariancji konieczne jest przyjęcie założenia, iż wszystkie macierze dyspersji są równe, czyli

$$\mathbf{\Sigma}_{m(j,k)} = \mathbf{\Sigma}_m = [\sigma_{ii'}] \quad \text{dla każdego } j \text{ oraz } k\,, \tag{2.4}$$

to znaczy założenie o równości macierzy dyspersji dla wszystkich miejscowości i lat.

Aby zapisać rozważany model (2.1) w postaci powszechnie stosowanej, uwzględniającej różne źródła zmienności, wprowadzimy następujące parametry:

$\mu_i = \mu_i(\cdot\,,\cdot)$ na oznaczenie przeciętnej plenności i-tego genotypu poprzez wszystkie miejscowości i lata,

$\alpha_i^P(j) = \mu_i(j,\cdot) - \mu_i(\cdot\,,\cdot)$ na oznaczenie przeciętnego odchylenia plenności i-tego genotypu w j-tej miejscowości,

$\alpha_i^T(k) = \mu_i(\cdot\,,k) - \mu_i(\cdot\,,\cdot)$ na oznaczenie przeciętnego odchylenia plenności i-tego genotypu pod wpływem k-tego roku,

$\alpha_i^{PT}(j,k) = \mu_i(j,k) - \mu_i(j,\cdot) - \mu_i(\cdot\,,k) + \mu_i(\cdot\,,\cdot)$ na oznaczenie przeciętnego odchylenia plenności i-tego genotypu pod wpływem interakcji j-tej miejscowości z k-tym rokiem, gdzie odpowiednie średnie brzegowe są zdefiniowane następująco:

$$\mu_i(\cdot\,,\cdot) = \sum_{j=1}^{J}\sum_{k=1}^{K} v_j w_k \mu_i(j,k) \Big/ \sum_{j=1}^{J}\sum_{k=1}^{K} v_j w_k\,,$$

$$\mu_i(j,\cdot) = \sum_{k=1}^{K} w_k\,\mu_i(j,k) \Big/ \sum_{k=1}^{K} w_k\,, \qquad \mu_i(\cdot\,,k) = \sum_{j=1}^{J} v_j\,\mu_i(j,k) \Big/ \sum_{j=1}^{J} v_j\,,$$

przy czym $\mathbf{v} = [v_1, v_2, \ldots, v_J]'$ jest wektorem (dodatnich) wag przypisanych poszczególnym miejscowościom, a $\mathbf{w} = [w_1, w_2, \ldots, w_K]'$ jest wektorem (dodatnich) wag przypisanych poszczególnym latom.

Ponadto wprowadźmy zmienną losową $\alpha_i^E(j,k) = m_i(j,k) - \mu_i(j,k)$ na oznaczenie odchylenia średniego plonu prawdziwego i-tego genotypu od jego wartości oczekiwanej, pod wpływem warunków środowiska, w którym przeprowadzono doświadczenie w j-tej miejscowości i k-tym roku (warunków nie zdeterminowanych jednoznacznie przez daną miejscowość w danym roku). Ponieważ w każdym z N środowisk (miejscowość, rok) przeprowadza się zwykle jedno doświadczenie, więc nie jest możliwe dokonanie rozróżnienia między składnikami modelu $\alpha_i^{PT}(j,k)$ i $\alpha_i^E(j,k)$. Dlatego konieczne jest upraszczające założenie, że $\alpha_i^{PT}(j,k) = 0$ dla każdego i, j oraz k, co jest na ogół do przyjęcia, gdy doświadczenia prowadzone są w ramach względnie jednorodnego rejonu.

Przy tak zdefiniowanych wielkościach, stosując notację macierzową, model (2.1) można zapisać w postaci

$$\mathbf{y}_{jk} = \boldsymbol{\mu} + \boldsymbol{\alpha}^{P}(j) + \boldsymbol{\alpha}^{T}(k) + \mathbf{a}^{E}(j,k) + \mathbf{e}_{jk} \ , \tag{2.5}$$

gdzie

$$\mathbf{y}_{jk} = [y_{1jk}, y_{2jk},\ldots,y_{Ijk}]' \ , \quad \boldsymbol{\mu} = [\mu_1, \mu_2,\ldots, \mu_I]',$$

$$\boldsymbol{\alpha}^{P}(j) = [\alpha_1^{P}(j), \alpha_2^{P}(j),\ldots, \alpha_I^{P}(j)]' \ , \quad \boldsymbol{\alpha}^{T}(k) = [\alpha_1^{T}(k), \alpha_2^{T}(k),\ldots, \alpha_I^{T}(k)]' \ ,$$

$$\mathbf{a}^{E}(j,k) = [a_1^{E}(j,k), \ a_2^{E}(j,k),\ldots, a_I^{E}(j,k)]' \ ,$$

natomiast wektor $\mathbf{e}_{jk}$ jest taki, jak został określony w (2.2).

Zgodnie z przyjętymi definicjami wektory parametrów $\boldsymbol{\mu}$, $\boldsymbol{\alpha}^{P}(j)$, $\boldsymbol{\alpha}^{T}(k)$ są efektami stałymi modelu, a składniki $\mathbf{a}^{E}(j,k)$, $\mathbf{e}_{ijk}$ są zmiennymi losowymi.

Z przyjętych wcześniej założeń wynika, że

$$\mathrm{E}\{\mathbf{a}^{E}(j,k)\} = \mathrm{E}\{\mathbf{e}_{jk}\} = \mathbf{0} \ , \quad \mathrm{D}\{\mathbf{a}^{E}(j,k)\} = \boldsymbol{\Sigma}_{m} \ , \quad \mathrm{D}\{\mathbf{e}_{jk}\} = \sigma_e^2 \boldsymbol{\Omega}_e \ ,$$

oraz że $\mathrm{E}\{\mathbf{a}^{E}(j,k)\,\mathbf{e}'_{jk}\} = \mathbf{0}$, a także, że wszystkie inne kowariancje między wektorami $\mathbf{a}^{E}(j,k)$ oraz $\mathbf{e}_{jk}$, $j = 1,2,\ldots,J$, $k = 1,2,\ldots,K$, są równe zero.

Dla skompletowania rozważanych założeń umożliwiających zastosowanie wielozmiennej analizy wariancji, przyjmijmy jeszcze, że wchodzące w grę rozkłady wielowymiarowe są normalne. Wobec tego wszystkie N wektory losowe $\mathbf{y}_{jk}$ mają niezależne i identyczne I-wymiarowe rozkłady normalne z wartością oczekiwaną $\boldsymbol{\mu} + \boldsymbol{\alpha}^{P}(j) + \boldsymbol{\alpha}^{T}(k)$ i macierzą kowariancji

$$\boldsymbol{\Sigma}_{y} = \boldsymbol{\Sigma}_{m} + \sigma_e^2 \boldsymbol{\Omega}_e \ , \tag{2.6}$$

co zapisujemy krótko

$$\mathbf{y}_{jk} \sim N_I\{\boldsymbol{\mu} + \boldsymbol{\alpha}^{P}(j) + \boldsymbol{\alpha}^{T}(k) \ , \ \boldsymbol{\Sigma}_{y}\} \ .$$

Dla przeprowadzenia w analizie opartej na modelu (2.5) porównań między genotypami, niezbędne jest jeszcze przekształcenie wektora $\mathbf{y}_{jk}$ do postaci

$$\mathbf{z}_{jk} = \mathbf{G}\mathbf{y}_{jk} \quad \text{dla każdego } j \text{ i } k, \tag{2.7}$$

gdzie

$$\mathbf{G} = \mathbf{I}_I - \mathbf{1}_I\mathbf{1}_I'/I \ , \tag{2.8}$$

przy czym $\mathbf{I}_I$ jest I-wymiarową macierzą jednostkową, a $\mathbf{1}_I$ oznacza I-wymiarowy wektor jedynek.

Po tym przekształceniu mamy model

$$\mathbf{z}_{jk} = \boldsymbol{\alpha}^{G} + \boldsymbol{\alpha}^{GP}(j) + \boldsymbol{\alpha}^{GT}(k) + \mathbf{a}^{GE}(j,k) + \mathbf{f}_{jk} \ , \tag{2.9}$$

o składnikach określonych następująco:

$$\boldsymbol{\alpha}^G = \mathbf{G}\boldsymbol{\mu} = [\alpha_1^G, \alpha_2^G, \ldots, \alpha_I^G]' \quad , \tag{2.10}$$

gdzie $\alpha_i^G = \mu_i - \mu.$ jest efektem głównym i-tego genotypu,

$$\boldsymbol{\alpha}^{GP}(j) = \mathbf{G}\,\boldsymbol{\alpha}^P(j) = [\alpha_1^{GP}(j), \alpha_2^{GP}(j), \ldots, \alpha_I^{GP}(j)]' \quad , \tag{2.11}$$

gdzie $\alpha_i^{GP}(j) = \alpha_i^P(j) - \alpha_.^P(j)$ jest interakcją i-tego genotypu z j-tą miejscowością,

$$\boldsymbol{\alpha}^{GT}(k) = \mathbf{G}\,\boldsymbol{\alpha}^T(k) = [\alpha_1^{GT}(k), \alpha_2^{GT}(k), \ldots, \alpha_I^{GT}(k)]' \quad , \tag{2.12}$$

gdzie $\alpha_i^{GT}(k) = \alpha_i^T(k) - \alpha_.^T(k)$ jest interakcją i-tego genotypu z k-tym rokiem,

$$\mathbf{a}^{GE}(j,k) = \mathbf{G}\mathbf{a}^E(j,k) = [a_1^{GE}(j,k), \ldots, a_I^{GE}(j,k)]' \quad , \tag{2.13}$$

gdzie $a_i^{GE}(j,k) = a_i^E(j,k) - a_.^E(j,k)$ jest interakcją i-tego genotypu z warunkami środowiskowymi doświadczenia przeprowadzonego w j-tej miejscowości i k-tym roku,

$$\mathbf{f}_{jk} = \mathbf{G}\mathbf{e}_{jk} = [f_{1jk}, f_{2jk}, \ldots, f_{Ijk}]' \quad , \tag{2.14}$$

gdzie $f_{ijk} = e_{ijk} - e_{.jk}$ jest odchyleniem błędu i-tego genotypu w (j,k)-tym doświadczeniu.

Z przyjętych założeń dla modelu (2.5) wynika, że wszystkie wektory $\mathbf{a}^{GE}(j,k)$ mają niezależne wielowymiarowe rozkłady normalne $N_I\{\mathbf{0}, \mathbf{G}\boldsymbol{\Sigma}_m\mathbf{G}\}$, gdzie $\mathbf{G}\boldsymbol{\Sigma}_m\mathbf{G} = [(\sigma_{ii'} - \sigma_{i.} - \sigma_{.i'} + \sigma_{..})]$, podczas gdy wszystkie wektory $\mathbf{f}_{jk}$ mają niezależne wielowymiarowe rozkłady normalne $N_I\{0, \sigma_e^2\mathbf{G}\,\boldsymbol{\Omega}_e\mathbf{G}\}$, niezależne od rozkładów wektorów $\mathbf{a}_{GE}(j,k)$. W związku z tym losowe wektory $\mathbf{z}_{jk}$ określone w (2.7) mają identyczne i niezależne I wymiarowe rozkłady normalne

$$\mathbf{z}_{jk} \sim N_I\{\boldsymbol{\alpha}^G + \boldsymbol{\alpha}^{GP}(j) + \boldsymbol{\alpha}^{GT}(k)\,,\ \boldsymbol{\Sigma}_z\} \quad , \tag{2.15}$$

gdzie

$$\boldsymbol{\Sigma}_z = \mathbf{G}\boldsymbol{\Sigma}_m\mathbf{G} + \sigma_e^2\mathbf{G}\boldsymbol{\Sigma}_e\mathbf{G} \tag{2.16}$$

dla wszystkich (j,k) środowisk, ściślej dla wszystkich podklas klasyfikacji według J miejscowości i K lat.

3. Sumy kwadratów i iloczynów w analizie wariancji i ich rozkłady

Aby zastosować wielozmienną analizę wariancji i uzyskać macierze sum kwadratów i iloczynów potrzebne do przeprowadzenia tej analizy wygodnie jest zapisać I-wymiarowe obserwacje $\mathbf{y}_{jk}$ jako wiersze $N\times I$ wymiarowej macierzy $\mathbf{Y}$, o liczbie wierszy N odpowiadającej liczbie doświadczeń w analizowanej serii.

Przy tak uformowanej macierzy obserwacji $\mathbf{Y}$, możemy model wielozmiennej analizy wariancji zapisać równością

$$\mathbf{Y} = \mathbf{1}_N\boldsymbol{\mu}' + \boldsymbol{\Delta}_P'\boldsymbol{\alpha}^P + \boldsymbol{\Delta}_T'\boldsymbol{\alpha}^T + \mathbf{a}^E + \mathbf{e} \ , \tag{3.1}$$

gdzie $\mathbf{1}_N$ jest wektorem N jedynek, $\boldsymbol{\Delta}_P'$ jest $N\times J$ wymiarową macierzą układu dla miejscowości, $\boldsymbol{\Delta}_T'$ jest $N\times K$ wymiarową macierzą układu dla lat, oraz gdzie macierze parametrów mają postać

$$\boldsymbol{\alpha}^P = [\boldsymbol{\alpha}^P(1),\ \boldsymbol{\alpha}^P(2),\ldots,\boldsymbol{\alpha}^P(J)]' \ , \qquad \boldsymbol{\alpha}^T = [\boldsymbol{\alpha}^T(1),\ \boldsymbol{\alpha}^T(2),\ldots,\boldsymbol{\alpha}^T(K)]' \ .$$

Aby można było oprzeć analizę na modelu przedstawionym w postaci (3.1) należy uzupełnić go relacjami wynikającymi ze sposobu definiowania występujących w nim parametrów. Zauważmy mianowicie, że parametry te podlegają następującym ograniczeniom związanym z odpowiednimi wagami wprowadzonymi w rozdziale 2:

$$\mathbf{v}'\boldsymbol{\alpha}^P = \mathbf{0}' = \mathbf{w}'\boldsymbol{\alpha}^T \ . \tag{3.2}$$

Wówczas, dla tak przyjętego modelu określonego w (3.1) wraz z relacjami określonymi w (3.2), interesujące nas macierze sum kwadratów i iloczynów są następujące:

(i) dla miejscowości

$$\mathbf{S}_P = \mathbf{Y}_P'\mathbf{Q}_P\mathbf{Y} \ , \tag{3.3}$$

gdzie macierz $\mathbf{Q}_P = \mathbf{F}_T\,\boldsymbol{\Delta}_P'\,\boldsymbol{\Omega}_P\boldsymbol{\Delta}_P\mathbf{F}_T$ jest rzędu J–1, $\boldsymbol{\Omega}_P^{-1} = \boldsymbol{\Delta}_P\mathbf{F}_T\boldsymbol{\Delta}_P' + \mathbf{v}\mathbf{v}'/\mathbf{1}_J'\mathbf{v}$, $\mathbf{F}_T = \mathbf{I}_N - \boldsymbol{\Delta}_T'\,\mathbf{J}^{-\delta}\boldsymbol{\Delta}_T$ i $\mathbf{J}^{-\delta} = \operatorname{diag}[J_1^{-1}, J_2^{-1},\ldots,J_K^{-1}]$.

(ii) dla lat

$$\mathbf{S}_T = \mathbf{Y}'\mathbf{Q}_T\mathbf{Y} \ , \tag{3.4}$$

gdzie macierz $\mathbf{Q}_T = \mathbf{F}_P\boldsymbol{\Delta}_T'\boldsymbol{\Omega}_T\boldsymbol{\Delta}_T\mathbf{F}_P$ jest rzędu K–1, $\boldsymbol{\Omega}_T^{-1} = \boldsymbol{\Delta}_T\,\mathbf{F}_P\boldsymbol{\Delta}_T' + \mathbf{w}\mathbf{w}'/\mathbf{1}_K'\mathbf{w}$, $\mathbf{F}_P = \mathbf{I}_N - \boldsymbol{\Delta}_P'\mathbf{K}^{-\delta}\boldsymbol{\Delta}_P$ i $\mathbf{K}^{-\delta} = \operatorname{diag}[K_1^{-1}, K_2^{-1},\ldots,K_J^{-1}]$.

(iii) dla środowisk

$$\mathbf{S}_E = \mathbf{Y}'\,\mathbf{Q}_E\mathbf{Y} \ , \tag{3.5}$$

gdzie macierz $\mathbf{Q}_E = \mathbf{F}_T(\mathbf{I}_N - \boldsymbol{\Delta}'_P\boldsymbol{\Omega}_P\boldsymbol{\Delta}_P)\mathbf{F}_T = \mathbf{F}_P(\mathbf{I}_N - \boldsymbol{\Delta}'_T\boldsymbol{\Omega}_T\boldsymbol{\Delta}_T)\,\mathbf{F}_P$ jest rzędu $N-J-K+1$.

(iv) dla genotypów

$$\mathbf{S}_G = \mathbf{GSG} \ , \tag{3.6}$$

gdzie $\mathbf{S} = \mathbf{Y}'\mathbf{QY}$, macierz $\mathbf{Q} = \mathbf{tt}'/\mathbf{t}'\mathbf{t} = \mathbf{uu}'/\mathbf{u}'\mathbf{u}$ jest rzędu 1, $\mathbf{t} = (\mathbf{I}_N - \mathbf{F}_P\boldsymbol{\Delta}'_T\boldsymbol{\Omega}_T\boldsymbol{\Delta}_T)\boldsymbol{\Delta}'_P\mathbf{K}^{-\delta}\mathbf{v}$ i $\mathbf{u} = (\mathbf{I}_N - \mathbf{F}_T\boldsymbol{\Delta}'_P\boldsymbol{\Omega}_P\boldsymbol{\Delta}_P)\boldsymbol{\Delta}'_T\mathbf{J}^{-\delta}\mathbf{w}$.

(v) dla interakcji genotypów z miejscowościami

$$\mathbf{S}_{GP} = \mathbf{GS}_P\mathbf{G} \ , \tag{3.7}$$

(vi) dla interakcji genotypów z latami

$$\mathbf{S}_{GT} = \mathbf{GS}_T\mathbf{G} \ , \tag{3.8}$$

(vii) dla interakcji genotypów ze środowiskami

$$\mathbf{S}_{GE} = \mathbf{GS}_E\mathbf{G} \ . \tag{3.9}$$

Macierze $\mathbf{S}_G$, $\mathbf{S}_{GP}$, $\mathbf{S}_{GT}$ mają rozkłady Wisharta z, odpowiednio, stopniami swobody 1, $J-1$, $K-1$, macierzą dyspersji $\boldsymbol{\Sigma}_z$, i macierzami parametrów niecentralności oznaczanymi dalej $\boldsymbol{\tau}_G$, $\boldsymbol{\tau}_{GP}$ i $\boldsymbol{\tau}_{GT}$. Ponadto z analiz poszczególnych doświadczeń otrzymujemy

(viii) połączoną sumę kwadratów dla błędu

$$SS_{\mathrm{e}} = \sum_{(j,k)} SS_{\mathrm{e}}(j,k) \ , \tag{3.10}$$

gdzie $SS_{\mathrm{e}}(j,k)$ jest sumą kwadratów dla błędu z analizy doświadczenia w j-tej miejscowości i k-tym roku.

Zauważmy, że między macierzami $\mathbf{Q}_P$, $\mathbf{Q}_T$, $\mathbf{Q}_E$ i $\mathbf{Q}$ zachodzą poniższe związki

$$\mathbf{Q}_P\mathbf{Q}_T \neq 0 \ , \quad \mathbf{Q}\,\mathbf{Q}_P \neq 0 \ , \quad \mathbf{Q}\,\mathbf{Q}_T \neq 0 \ , \quad \mathbf{Q}_P\mathbf{Q}_E = \mathbf{Q}_T\mathbf{Q}_E = \mathbf{Q}\,\mathbf{Q}_E = 0 \ . \tag{3.11}$$

Z macierzy (3.3)-(3.5) otrzymujemy następujące sumy kwadratów odchyleń:

– dla miejscowości

$$SS_P = \mathbf{1}'_I\mathbf{S}_P\mathbf{1}_I/I \ , \tag{3.12}$$

– dla lat

$$SS_T = \mathbf{1}'_I\mathbf{S}_T\mathbf{1}_I/I \ , \tag{3.13}$$

– dla środowisk

$$SS_E = \mathbf{1}'_I\mathbf{S}_E\mathbf{1}_I/I \ . \tag{3.14}$$

Suma kwadratów SS_P ma taki rozkład jak zmienna $I(\sigma_E^2 + \sigma_e^2\omega_{..})\,\chi^2(J-1, \tau_P^2)$, gdzie zmienna $\chi^2(J-1, \tau_p^2)$ ma niecentralny rozkład chi-kwadrat z $J-1$ stopniami swobody i parametrem niecentralności $\tau_p^2 = (\boldsymbol{\alpha}_{\cdot}^{P})'\,\boldsymbol{\Omega}_P^{-1}\boldsymbol{\alpha}_{\cdot}^{P} / (\sigma_E^2 + \sigma_e^2\omega_{..})$, przy czym $\omega_{..} = \mathbf{1}'\boldsymbol{\Omega}_e\mathbf{1}/I^2$, $\boldsymbol{\alpha}_{\cdot}^{P} = I^{-1}\boldsymbol{\alpha}^{P}\mathbf{1} = [\alpha_{\cdot}^{P}(1), \alpha_{\cdot}^{P}(2),\ldots, \alpha_{\cdot}^{P}(J)]'$, $\sigma_E^2 = \mathbf{1}'\,\boldsymbol{\Sigma}_m\,\mathbf{1}/I^2 = \sigma_{..}$. Wobec tego $\mathrm{E}\{SS_P\} = \mathbf{I}(\sigma_E^2 + \sigma_e^2\omega_{..})(J-1+\tau_P^2)$.

Suma kwadratów SS_T ma taki rozkład jak zmienna $I(\sigma_E^2 + \sigma_e^2\omega_{..})\,\chi^2(K-1, \tau_T^2)$, gdzie zmienna $\chi^2(K-1,\tau_T^2)$ ma niecentralny rozkład chi-kwadrat z $K-1$ stopniami swobody i parametrem niecentralności $\tau_T^2 = (\boldsymbol{\alpha}_{\cdot}^{T})'\;\boldsymbol{\Omega}_T^{-1}\boldsymbol{\alpha}_{\cdot}^{T} / (\sigma_E^2 + \sigma_e^2\omega_{..})$, przy czym $\boldsymbol{\alpha}_{\cdot}^{T} = I^{-1}\boldsymbol{\alpha}^{T}\mathbf{1} = [\alpha_{\cdot}^{T}(1), \alpha_{\cdot}^{T}(2),\ldots, \alpha_{\cdot}^{T}(K)]'$. Stąd $\mathrm{E}\{SS_T\} = I(\sigma_E^2 + \sigma_e^2\omega_{..})(K-1+\tau_T^2)$.

Suma kwadratów SS_E ma taki rozkład jak zmienna $I(\sigma_E^2+\sigma_e^2\omega_{..})\,\chi^2(N-J-K+1)$, gdzie zmienna $\chi^2(N-J-K+1)$ ma centralny rozkład chi-kwadrat z $N-J-K+1$ stopniami swobody. Zatem $\mathrm{E}\{SS_E\} = (N-J-K+1)I(\sigma_E^2 + \sigma_e^2\omega_{..})$.

Z macierzy (3.6) – (3.9) otrzymujemy dalsze sumy kwadratów odchyleń

– dla genotypów

$$SS_G = \mathrm{tr}(\mathbf{S}_G)\ , \tag{3.15}$$

– dla interakcji genotypów z miejscowościami

$$SS_{GP} = \mathrm{tr}(\mathbf{S}_{GP})\ , \tag{3.16}$$

– dla interakcji genotypów z latami

$$SS_{GT} = \mathrm{tr}(\mathbf{S}_{GT})\ , \tag{3.17}$$

– dla interakcji genotypów ze środowiskami

$$SS_{GE} = \mathrm{tr}(\mathbf{S}_{GE})\ . \tag{3.18}$$

Suma kwadratów SS_G ma taki rozkład jak funkcja $\sum_{r=1}^{I-1}\xi_r\,\chi_r^2(1, \tau_{G(r)}^2)$, gdzie $I-1$ zmiennych $\chi_r^2(1, \tau_{G(r)}^2)$ ma niezależne niecentralne rozkłady chi-kwadrat, każdy z 1 stopniem swobody i parametrem niecentralności $\tau_{G(r)}^2 = \mathbf{q}_r'\boldsymbol{\Sigma}_z\,\boldsymbol{\tau}_G\,\mathbf{q}_r/\xi_r$, przy czym $\mathbf{q}_1, \mathbf{q}_2,\ldots,\mathbf{q}_{I-1}$ są wektorami własnymi macierzy $\boldsymbol{\Sigma}_z$ odpowiadającymi niezerowym wartościom własnym $\xi_1, \xi_2,\ldots, \xi_{I-1}$.

Suma kwadratów SS_{GP} ma taki rozkład jak funkcja $\sum_{r=1}^{I-1}\xi_r\,\chi_r^2(J-1, \tau_{GP(r)}^2)$, gdzie $I-1$ zmiennych $\chi_r^2(J-1, \tau_{GP(r)}^2)$ ma niezależne niecentralne rozkłady chi-kwadrat, każdy z $J-1$ stopniami swobody i z parametrem niecentralności $\tau_{GP(r)}^2 = \mathbf{q}_r'\boldsymbol{\Sigma}_z\,\boldsymbol{\tau}_{GP}\,\mathbf{q}_r/\xi_r$.

Suma kwadratów SS_{GT} ma taki rozkład jak funkcja $\sum_{r=1}^{I-1}\xi_r\,\chi_r^2(K-1, \tau_{GT(r)}^2)$, gdzie I-1 zmiennych $\chi_r^2(K-1, \tau_{GT(r)}^2)$ ma niezależne niecentralne rozkłady chi-kwadrat, każdy z $K-1$ stopniami swobody i z parametrem niecentralności $\tau_{GT(r)}^2 = \mathbf{q}_r'\boldsymbol{\Sigma}_z\,\boldsymbol{\tau}_{GT}\,\mathbf{q}_r/\xi_r$.

Suma kwadratów SS_{GE} ma taki rozkład jak funkcja $\sum_{r=1}^{I-1} \xi_r \chi_r^2(N-J-K+1)$, gdzie $I-1$ zmiennych $\chi_r^2(N-J-K+1)$ ma niezależne centralne rozkłady chi-kwadrat, z $N-J-K+1$ stopniami swobody.

Sumy kwadratów $SS_e(j,k)$ mają taki rozkład jak zmienna $\sigma_e^2 \chi^2(\nu_{e,jk})$. Wobec tego $E\{SS_e\} = \sum_{(j,k)} E\{SS_e(j,k)\} = \sum_{(j,k)} \sigma_e^2 \nu_{e,jk}$. Gdy każdy układ jest spójny i jednakowy w całej serii (co zakładamy), to SS_e ma rozkład $\sigma_e^2 \chi^2(N\nu_e)$ i $E\{SS_e\} = N\nu_e \sigma_e^2$, przy czym $\nu_e = \nu_{e,jk}$ dla wszystkich j oraz k.

Zauważmy jeszcze, że dzięki równościom (3.11), zarówno rozkład sumy kwadratów SS_P jak i SS_T jest niezależny od rozkładu sumy kwadratów SS_E. Ponadto wszystkie te trzy rozkłady są niezależne od rozkładu sumy kwadratów SS_e, oraz rozkłady sum kwadratów SS_G, SS_{GP} i SS_{GT} są niezależne od rozkładu sumy kwadratów SS_{GE}.

Powyższe rozkłady determinują następujące wartości oczekiwane:

$$E\{SS_G\} = \mathrm{tr}(\mathbf{\Sigma}_z \boldsymbol{\tau}_G) + (I-1)\sigma_{GE}^2 + \sigma_e^2(\mathrm{tr}\ \boldsymbol{\Omega}_e - I\omega_{..}) ,$$

$$E\{SS_{GP}\} = \mathrm{tr}(\mathbf{\Sigma}_z \boldsymbol{\tau}_{GP}) + (I-1)(J-1)\sigma_{GE}^2 + (J-1)\sigma_e^2(\mathrm{tr}\ \boldsymbol{\Omega}_e - I\omega_{..}) ,$$

$$E\{SS_{GT}\} = \mathrm{tr}(\mathbf{\Sigma}_z \boldsymbol{\tau}_{GT}) + (I-1)(K-1)\sigma_{GE}^2 + (K-1)\sigma_e^2(\mathrm{tr}\ \boldsymbol{\Omega}_e - I\omega_{..}) ,$$

$$E\{SS_{GE}\} = (I-1)(N-J-K+1)\sigma_{GE}^2 + (N-J-K+1)\sigma_e^2(\mathrm{tr}\boldsymbol{\Omega}_e - I\omega_{..}) ,$$

gdzie

$$\sigma_{GE}^2 = (I-1)^{-1}\sum_i \mathrm{var}\{a_i^{GE}(j,k)\} = (I-1)^{-1}\mathrm{tr}(\mathbf{G}\mathbf{\Sigma}_m\mathbf{G}) = (I-1)^{-1}\sum_i(\sigma_{ii} - \sigma_{..}) \quad (3.19)$$

Rezultaty te pozwalają na przedstawienie analizy wariancji rozważanej serii doświadczeń w formie takiej, jak pokazuje Tabela 1.

Wykorzystując przedstawione w tej tabeli wartości oczekiwane średnich kwadratów można podać nieobciążone estymatory komponentów wariancyjnych σ_E^2 i σ_{GE}^2, które są odpowiednio równe

$$\hat{\sigma}_E^2 = I^{-1}(MS_E - I\omega_{..}MS_e)$$

i

$$\hat{\sigma}_{GE}^2 = MS_{GE} - (I-1)^{-1}(\mathrm{tr}\boldsymbol{\Omega}_e - I\omega_{..})MS_e ,$$

gdzie $MS_E = SS_E/(N-J-K+1)$, $MS_{GE} = SS_{GE}/((I-1)(N-J-K+1))$ i $MS_e = SS_e/(N\nu_e)$. Natomiast estymatory parametrów stałych modelu (3.1) otrzymane metodą najmniejszych kwadratów przy uwzględnieniu warunków ograniczających (3.2) są postaci

$$\hat{\boldsymbol{\mu}} = \mathbf{t}'\mathbf{Y}/\mathbf{1}_J'\mathbf{v} = \mathbf{u}'\mathbf{Y}/\mathbf{1}_K'\mathbf{w} ,$$

Tabela 1
Analiza wariancji

Źródło zmienności	Liczba stopni swobody	Suma kwadratów	Wartość oczekiwana średniego kwadratu
P	$J-1$	SS_P	$I(\sigma_E^2+\sigma_e^2\omega_{..})\tau_P^2/(J-1)+I(\sigma_E^2+\sigma_e^2\omega_{..})$
T	$K-1$	SS_T	$I(\sigma_E^2+\sigma_e^2\omega_{..})\tau_T^2/(K-1)+I(\sigma_E^2+\sigma_e^2\omega_{..})$
E	$N-J-K+1$	SS_E	$I(\sigma_E^2+\sigma_e^2\omega_{..})$
G	$I-1$	SS_G	$\mathrm{tr}(\boldsymbol{\Sigma}_z\boldsymbol{\tau}_G)/(I-1)+\sigma_{GE}^2+\sigma_e^2(\mathrm{tr}\ \boldsymbol{\Omega}_e-I\omega_{..})/(I-1)$
GP	$(I-1)(J-1)$	SS_{GP}	$\mathrm{tr}(\boldsymbol{\Sigma}_z\boldsymbol{\tau}_{GP})/\{(I-1)(J-1)+\sigma_{GE}^2+\sigma_e^2(\mathrm{tr}\ \boldsymbol{\Omega}_e-I\omega_{..})/(I-1)$
GT	$(I-1)(K-1)$	SS_{GT}	$\mathrm{tr}(\boldsymbol{\Sigma}_z\boldsymbol{\tau}_{GT})/\{(I-1)(K-1)+\sigma_{GE}^2+\sigma_e^2(\mathrm{tr}\ \boldsymbol{\Omega}_e-I\omega_{..})/(I-1)$
GE	$(I-1)(N-J-K+1)$	SS_{GE}	$\sigma_{GE}^2+\sigma_e^2(\mathrm{tr}\boldsymbol{\Omega}_e-I\omega_{..})\ /\ (I-1)$
Błąd	$N\nu_e$	SS_e	σ_e^2

$$\hat{\boldsymbol{\alpha}}^P = \boldsymbol{\Omega}_P\boldsymbol{\Delta}_P\mathbf{F}_T\mathbf{Y}\ ,$$

$$\hat{\boldsymbol{\alpha}}^T = \boldsymbol{\Omega}_T\boldsymbol{\Delta}_T\mathbf{F}_P\mathbf{Y}\ ,$$

skąd

$$\hat{\boldsymbol{\alpha}}^G = \mathbf{G}\hat{\boldsymbol{\mu}}\ ,$$

$$(\hat{\boldsymbol{\alpha}}^{GP})' = \mathbf{G}(\hat{\boldsymbol{\alpha}}^P)'$$

$$(\hat{\boldsymbol{\alpha}}^{GT})' = \mathbf{G}(\hat{\boldsymbol{\alpha}}^T)'$$

W końcu jako nieobciążone estymatory macierzy dyspersji $\boldsymbol{\Sigma}_y$ i $\boldsymbol{\Sigma}_z$ otrzymujemy odpowiednio

$$\hat{\boldsymbol{\Sigma}}_y = (N-J-K+1)^{-1}\mathbf{S}_E$$

i

$$\hat{\boldsymbol{\Sigma}}_z = \mathbf{G}\hat{\boldsymbol{\Sigma}}_y\mathbf{G} = (N-J-K+1)^{-1}\mathbf{S}_{GE}\ .$$

4. Kryteria testowe i ich rozkłady

Analiza Tabeli 1 oraz przedstawione rozkłady sum kwadratów przedstawione w poprzednim rozdziale wskazują jak tworzyć statystyki testowe dla interesujących hipotez.

Hipoteza o równości średnich efektów głównych wszystkich miejscowości, czyli hipoteza H_P : wszystkie $\alpha_{\cdot}^{P}(j) = 0$, może być testowana za pomocą ilorazu

$$F_P = MS_P/MS_E \ . \tag{4.1}$$

Przy prawdziwości hipotezy H_P iloraz (4.1) ma centralny rozkład F z $\nu_P = J-1$ i $\nu_E = N-J-K+1$ stopniami swobody. Hipotezę H_P odrzucimy więc na poziomie istotności α wtedy i tylko wtedy, gdy $F_P > F_{\alpha,\nu_P,\nu_E}$, gdzie F_{α,ν_P,ν_E} jest 100α-procentową wartością krytyczną z rozkładu F dla ν_P i ν_E stopni swobody.

Podobnie hipoteza o równości średnich efektów głownych wszystkich lat, czyli hipoteza H_T : wszystkie $\alpha_{\cdot}^{T}(k) = 0$, może być testowana za pomocą ilorazu

$$F_T = MS_T/MS_E \ . \tag{4.2}$$

Przy prawdziwości hipotezy H_T iloraz (4.2) ma centralny rozkład F z $\nu_T = K-1$ i $\nu_E = N-J-K+1$ stopniami swobody. Hipotezę H_T odrzucimy więc na poziomie istotności α wtedy i tylko wtedy, gdy $F_T > F_{\alpha,\nu_T,\nu_E}$.

Hipoteza o braku efektów głównych środowisk wewnątrz miejscowości i lat, a ściślej hipoteza H_E : $\sigma_E^2 = 0$, może być testowana za pomocą ilorazu

$$F_E = MS_E/(I\omega_{\cdot\cdot}MS_e) \ . \tag{4.3}$$

Przy prawdziwości hipotezy H_E iloraz (4.3) ma centralny rozkład F z ν_E i $N\nu_e$ stopniami swobody. Hipotezę H_E odrzucimy na poziomie istotności α wtedy i tylko wtedy, gdy $F_E > F_{\alpha,\nu_E,N\nu_e}$.

Hipoteza o równości efektów głównych wszystkich genotypów, a ściślej hipoteza H_G : wszystkie $\alpha_i^G = 0$, może być testowana za pomocą statystyki opartej na dwóch macierzach, $\mathbf{S}_G$ i $\mathbf{S}_{GE}$, które wygodnie jest transformować odpowiednio do $\mathbf{C}_0'\mathbf{S}_G\mathbf{C}_0$ i $\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0$, gdzie $\mathbf{C}_0$ jest dowolną macierzą o I wierszach i $(I-1)$ kolumnach taką, że rząd$(\mathbf{G}\,\mathbf{C}_0) = I-1$. Statystyką tą jest statystyka Hotellinga postaci

$$T_G^2 = \nu_E \operatorname{tr}\{(\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0)^{-1}(\mathbf{C}_0'\mathbf{S}_G\mathbf{C}_0)\} = \nu_E\tilde{N}(\mathbf{C}_0'\hat{\boldsymbol{\alpha}}^G)'\,(\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0)^{-1}\mathbf{C}_0'\hat{\boldsymbol{\alpha}}^G, \tag{4.4}$$

gdzie $\tilde{N} = (\mathbf{1}_J'\mathbf{v})^2/\mathbf{t}'\mathbf{t} = (\mathbf{1}_K'\mathbf{w})^2/\mathbf{u}'\mathbf{u}$.

Statystyka (4.4) pomnożona przez $(\nu_E - I+2)/\{(I-1)\nu_E\}$ ma (por. Rao, 1973; Morrison, 1976) niecentralny rozkład F z $I-1$ i ν_E-I+2 stopniami swobody oraz parametrem niecentralności $\operatorname{tr}(\boldsymbol{\tau}_{C_0'G}) = \tilde{N}(\mathbf{C}_0'\boldsymbol{\alpha}^G)'(\mathbf{C}_0'\boldsymbol{\Sigma}_z\mathbf{C}_0)^{-1}\mathbf{C}_0'\boldsymbol{\alpha}^G$. Przy prawdzi-

wości hipotezy H_G ma ona centralny rozkład F z tymi samymi stopniami swobody. Zatem hipoteza H_G może być testowana przez porównanie wartości statystyki

$$F_G = \{(I-1)\nu_E\}^{-1}(\nu_E - I + 2)\, T_G^2 \tag{4.5}$$

z wartością krytyczną $F_{\alpha, I-1, \nu_E - I+2}$, przy założeniu, że $\nu_E > I-1$.

Z kolei rozważmy testowanie hipotezy o braku interakcji genotypów z miejscowościami, czyli hipotezy H_{GP} : wszystkie $\alpha_i^{GP}(j) = 0$, którą można zapisać równoważnie jako H_{GP} : $\mathbf{C}_0'\boldsymbol{\alpha}^{GP}(j) = \mathbf{0}$ dla każdego j, z macierzą $\mathbf{C}_0$ określoną powyżej. Transformując macierze (3.7) i (3.9) odpowiednio do postaci $\mathbf{C}_0'\mathbf{S}_{GP}\mathbf{C}_0$ i $\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0$ zauważamy, że do sprawdzenia hipotezy H_{GP} można wykorzystać jedną ze statystyk testowych stosowanych w wielozmiennej analizie wariancji (por. Anderson, 1958; Gabriel, 1968; Morrison, 1976). Zastosujemy tu statystykę opartą na śladzie Hotellinga-Lawleya

$$T_0^2 = \nu_E \operatorname{tr}\{(\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0)^{-1}\mathbf{C}_0'\mathbf{S}_{GP}\mathbf{C}_0\} = \nu_E \sum_{h=1}^{s} \lambda_h^{GP} \;, \tag{4.6}$$

gdzie $\lambda_1^{GP} > \lambda_2^{GP} > \ldots > \lambda_s^{GP} > 0$ są niezerowymi wartościami własnymi macierzy $(\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0)^{-1}\mathbf{C}_0'\mathbf{S}_{GP}\mathbf{C}_0$, przy czym $s = \min(I-1, J-1)$. Ponieważ tablice wartości krytycznych dla statystyki T_0^2 (por. Pillai, 1960; Roy i in. 1971; Pearson i Hartley, 1972) są bardzo ograniczone, wobec tego stosowanie powyższej statystyki testowej w praktycznych zastosowaniach w postaci dokładnej jest bardzo utrudnione. W związku z tym proponuje się zastąpienie wartości krytycznych dla powyższej statystyki testowej aproksymowaną wartością krytyczną dla statystyki F. Zastosujemy tu aproksymację McKeona (1974), która ma postać

$$F_{GP} = \{f_2(\nu_E - I)/f_1(f_2-2)\}T_0^2/\nu_E \;, \tag{4.7}$$

gdzie

$$f_1 = (I-1)(J-1), \quad f_2 = 4+(f_1+2)/(B-1), \quad B = (\nu_E+J-I-1)(\nu_E-1)/\{(\nu_E-I-2)(\nu_E-I+1)\} \;.$$

Statystyka (4.7) ma przy prawdziwości hipotezy H_{GP}, w przybliżeniu rozkład F z f_1 i f_2 stopniami swobody. Zatem hipotezę H_{GP} odrzucimy na poziomie istotności α wtedy i tylko wtedy, gdy $F_{GP} > F_{\alpha, f_1, f_2}$.

To samo kryterium testowe może być zastosowane do sprawdzenia hipotezy o braku interakcji genotypów z latami, czyli hipotezy H_{GT}: wszystkie $\alpha_i^{GT}(k) = 0$.

Dla zweryfikowania hipotezy o braku interakcji genotypów ze środowiskami, tzn. hipotezy H_{GE} : $\sigma_{GE}^2 = 0$, zastosujemy statystykę

$$F_{GE} = \operatorname{tr}\{(\mathbf{C}_0'\mathbf{G}\boldsymbol{\Omega}_e\mathbf{G}\mathbf{C}_0)^{-1}\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0\}/\{(I-1)\nu_E MS_e\} \;, \tag{4.8}$$

która ma przy prawdziwości hipotezy H_{GE} rozkład centralny F z $(I-1)\nu_E$ i $N\nu_e$ stopniami swobody. Hipotezę H_{GE} odrzucimy więc na poziomie istotności α wtedy i tylko wtedy, gdy $F_{GE} > F_{\alpha,(I-1)\nu_E, N\nu_e}$.

5. Procedury testowania jednoczesnego hipotez szczegółowych

Każda z hipotez ogólnych rozważanych w rozdziale 4 implikuje wiele różnych hipotez szczegółowych, które mogą być testowane jednocześnie. Odpowiednie procedury testów jednoczesnych opracowane są zarówno dla hipotez w jednozmiennej analizie wariancji (ANOVA) jak i dla hipotez w wielozmiennej analizie wariancji (MANOVA) (patrz Gabriel, 1968; Caliński i in., 1979). Dla nas interesujące może być zastosowanie ich do hipotez dotyczących efektów głównych genotypów oraz ich interakcji z miejscowościami, latami i środowiskami. Rozważać będziemy tutaj w szczególności skończone rodziny hipotez szczegółowych implikowanych przez hipotezy ogólne, które w naszym wypadku dotyczyć będą pojedynczych genotypów lub porównań między nimi.

Rozważmy skończoną rodzinę hipotez $H_{\mathbf{c}'_{i^o}G}: \mathbf{c}'_{i^o}\boldsymbol{\alpha}^G = 0$, implikowanych przez hipotezę H_G gdzie $\mathbf{c}_{i^o}$, $i^o = 1,2,\ldots,I^o$, są pewnymi wektorami takimi, że $\mathbf{G}\,\mathbf{c}_{i^o} \neq \mathbf{0}$ dla każdego i^o. Wówczas konsekwentną i harmonijną procedurą testów jednoczesnych jest procedura polegająca na odrzuceniu na poziomie α hipotezy ogólnej H_G wtedy i tylko wtedy, gdy $\max\limits_{1\le i^o\le I^o} F_{\mathbf{c}'_{i^o}G} > F_{\max,\alpha}$, gdzie

$$F_{\mathbf{c}'_{i^o}G} = T^2_{\mathbf{c}'_{i^o}G} = \nu_E \tilde{N}(\mathbf{c}'_{i^o}\boldsymbol{\alpha}^G)^2 / (\mathbf{c}'_{i^o}\mathbf{S}_{GE}\,\mathbf{c}_{i^o}) \tag{5.1}$$

i gdzie wartość krytyczna $F_{\max,\alpha}$ jest tak dobrana, że

$$P\{\max_{1\le i^o\le I^o} F_{\mathbf{c}'_{i^o}G} > F_{\max,\alpha} \mid H_G\} = \alpha\ , \tag{5.2}$$

oraz na odrzuceniu dowolnej implikowanej przez H_G hipotezy $H_{\mathbf{c}'_{i^o}G}$ wtedy i tylko wtedy, gdy odpowiadająca jej statystyka (5.1) przekracza wartość $F_{\max,\alpha}$. Taką rodziną hipotez może być w szczególności zbiór $I^o=I$ hipotez wyznaczonych przez wektory $\mathbf{c}'_1 = [1,0,\ldots,0]$, $\mathbf{c}'_2 = [0,1,0,\ldots,0],\ldots,$ $\mathbf{c}'_I = [0,0,\ldots,0,1]$, przy czym i-tą hipotezą z tej rodziny jest wówczas hipoteza $H_{G(i)}: \alpha_i^G = 0$ dla $i = 1,2,\ldots,I$. Inną rodzinę mogą tworzyć hipotezy $H_{\mathbf{c}'_{i^o}G}$ dotyczące zbioru I^o, wybranych z góry, niezależnie od analizowanych wyników doświadczalnych, porównań (kontrastów) między genotypami. Wówczas i^o-ta hipoteza z tej rodziny jest postaci $H_{\mathbf{c}'_{i^o}G}: \mathbf{c}'_{i^o}\boldsymbol{\alpha}^G = 0$, gdzie $\mathbf{c}'_{i^o} = [c_{i^o1}, c_{i^o2},\ldots,c_{i^oI}]$, z $\sum_{i=1}^{I} c_{i^o i} = 0$, jest wektorem współczynników wyznaczających i^o-ty wybrany kontrast między genotypami ($i^o = 1,2,\ldots,I^o$).

Istnieją jednak pewne ograniczenia w stosowaniu powyższego kryterium testowego $F_{\max}$, wynikające stąd, że wartości krytyczne $F_{\max,\alpha}$ zdefiniowane w (5.2)

są trudno dostępne (z wyjątkiem niektórych szczególnych przypadków). W praktycznych zastosowaniach trudność tę można ominąć aproksymując $F_{\max,\alpha}$ nierównością Bonferroniego. Oznacza to posługiwanie się wartością

$$\tilde{F}_{\max,\alpha} = F_{\alpha/I^\circ;1,\nu_E} \ , \tag{5.3}$$

gdzie $F_{\alpha/I^\circ;1,\nu_E}$ jest $100\alpha/I^\circ$-procentową wartością krytyczną w tablicy rozkładu F dla 1 i ν_E stopni swobody.

Podobne podejście może być zastosowane dla skończonych rodzin hipotez implikowanych przez hipotezy H_{GP} lub H_{GT}. Przedstawimy je dla hipotezy H_{GP}. Skończoną rodzinę hipotez $H_{\mathbf{c}'_{i^\circ}GP}: \mathbf{c}'_{i^\circ}(\boldsymbol{\alpha}^{GP})' = \mathbf{0}'$ implikowanych przez H_{GP}, weryfikujemy procedurą testów jednoczesnych $F_{\max}$ opartą na statystyce

$$F_{\mathbf{c}'_{i^\circ}GP} = T_{0,\,\mathbf{c}'_{i^\circ}GP} / (J-1) = \{\nu_E/(J-1)\}\,\mathbf{c}'_{i^\circ}\mathbf{S}_{GP}\,\mathbf{c}_{i^\circ}(\mathbf{c}'_{i^\circ}\mathbf{S}_{GE}\,\mathbf{c}_{i^\circ}) \ , \tag{5.4}$$

którą porównywać będziemy z wartością krytyczną $\tilde{F}_{\max,\alpha}$ aproksymowaną nierównością Bonferroniego

$$\tilde{F}_{\max,\alpha} = F_{\alpha/I^\circ;J-1,\nu_E} \ . \tag{5.5}$$

Przechodząc do testowania skończonej rodziny hipotez implikowanych przez hipotezę $H_{GE}: \sigma^2_{GE} = 0$ zauważmy, że hipotezę tą możemy zapisać równoważnie w postaci $H_{GE}: \mathbf{G}\boldsymbol{\Sigma}_m\mathbf{G} = \mathbf{0}$. Hipoteza ta implikuje każdą hipotezę postaci $H_{\mathbf{c}'_{i^\circ}GE}: \mathbf{c}'_{i^\circ}\mathbf{G}\,\boldsymbol{\Sigma}_m\mathbf{G}\,\mathbf{c}_{i^\circ} = 0$. Skończoną rodzinę takich hipotez będziemy weryfikować procedurą testów jednoczesnych $F_{\max}$ opartą na statystyce

$$F_{\mathbf{c}'_{i^\circ}GE} = MS_{\mathbf{c}'_{i^\circ}GE}/MS_e \ , \tag{5.6}$$

gdzie

$$MS_{\mathbf{c}'_{i^\circ}GE} = (\mathbf{c}'_{i^\circ}\mathbf{G}\boldsymbol{\Omega}_e\mathbf{G}\,\mathbf{c}_{i^\circ})^{-1}\mathbf{c}'_{i^\circ}\mathbf{S}_{GE}\,\mathbf{c}_{i^\circ}/\nu_E \ . \tag{5.7}$$

Statystykę (5.6) porównywać będziemy z jedną wspólną wartością krytyczną $\tilde{F}_{\max,\alpha}$ aproksymowaną nierównością Bonferroniego

$$\tilde{F}_{\max,\alpha} = F_{\alpha/I^\circ;\nu_E,N\nu_e} \ . \tag{5.8}$$

Inne (zwykle skończone) rodziny hipotez implikowanych przez hipotezę H_{GP} mogą dotyczyć liniowych kombinacji między miejscowościami. Każda taka hipoteza z rodziny, implikowana przez hipotezę H_{GP}, może być przedstawiona w postaci $H_{GP\mathbf{d}_{j^\#}}: \mathbf{d}'_{j^\#}\boldsymbol{\alpha}^{GP} = \mathbf{0}$, przy czym $\mathbf{d}_{j^\#}$ jest odpowiednim wektorem o J składowych ($j^\# = 1,2,\ldots,J^\#$).

Konsekwentna i harmonijna procedura testów jednoczesnych zastosowana na poziomie α polega tu na odrzuceniu hipotezy H_{GP} wtedy i tylko wtedy, gdy

$$\max_{1 \le j^{\#} \le J^{\#}} T_{GP\mathbf{d}_j^{\#}} > T^2_{\max,\alpha} \ , \tag{5.9}$$

przy czym

$$T_{GP\mathbf{d}_{j^{\#}}} = \nu_E \{ \mathbf{d}'_{j^{\#}} (\mathbf{\Omega}_P - \mathbf{1}_J \mathbf{1}'_J / \mathbf{1}'_J \mathbf{v}) \mathbf{d}_{j^{\#}} \}^{-1} \mathbf{d}'_{j^{\#}} \hat{\boldsymbol{\alpha}}^{GP} \mathbf{C}_0 (\mathbf{C}'_0 \mathbf{S}_{GE} \mathbf{C}_0)^{-1} \mathbf{C}'_0 (\hat{\boldsymbol{\alpha}}^{GP})' \mathbf{d}_{j^{\#}} \ , \tag{5.10}$$

gdzie wartość krytyczna $T^2_{\max,\alpha}$ spełnia warunek

$$P \{ \max_{1 \le j^{\#} \le J^{\#}} T_{GP\mathbf{d}_{j^{\#}}} > T^2_{\max,\alpha} \mid H_{GP} \} = \alpha \ ,$$

oraz na odrzuceniu jakiejkolwiek implikowanej przez H_{GP} hipotezy $H_{GP\mathbf{d}_{j^{\#}}}$ wtedy i tylko wtedy, gdy odpowiadająca jej statystyka określona w (5.10) przewyższa wartość $T^2_{\max,\alpha}$. Ze względu na brak odpowiednich tablic wartości krytycznych $T^2_{\max,\alpha}$ (z wyjątkiem niektórych przypadków) dokonamy tu także aproksymacji $T^2_{\max,\alpha}$ nierównością Bonferroniego. Oznacza to posługiwanie się wartością

$$\tilde{T}^2_{\max,\alpha} = \nu_E \{ (I-1) / (\nu_E - I + 2) \} F_{\alpha/J^{\#}} \ , \tag{5.11}$$

gdzie $F_{\alpha/J^{\#}}$ jest $100\alpha/J^{\#}$-procentową wartością krytyczną w tablicy rozkładu F dla $I-1$ i $\nu_E - I + 2$ stopni swobody.

W szczególności taką rodziną hipotez jest zbiór $J^{\#} = J$ hipotez określonych przez wektory $\mathbf{d}_1 = [1,0,\ldots,0]'$, $\mathbf{d}_2 = [0,1,\ldots,0]', \ldots, \mathbf{d}_J = [0,0,\ldots,1]'$, gdzie j-ty wektor określa hipotezę $H_{GP\mathbf{d}_j} \equiv H_{GP(j)}$: $\boldsymbol{\alpha}^{GP}(j) = \mathbf{0}$ $(j = 1,2,\ldots,J)$.

6. Zastosowanie regresji w badaniu interakcji genotypowo-środowiskowej

Z omówionego w rozdziale 2 modelu matematycznego obserwacji widoczne jest, że losowe składniki przedstawiające interakcje poszczególnych genotypów ze środowiskami, mogą być skorelowane dla danego (j, k). Skorelowane mogą być także ich funkcje liniowe. Toteż przy badaniu interakcji genotypowo-środowiskowej interesująca jest cała macierz dyspersji wektorów $\mathbf{a}^{GE}(j,k)$ określonych w (2.13). W związku z tym badanie interakcji *GE* związane jest z określeniem wpływu na nie pewnej zmiennej środowiskowej, przeważnie rozumianej jako funkcja liniowa zmiennych obserwowanych w poszczególnych środowiskach. Niech funkcją tą będzie $\mathbf{x}_{jk} = \mathbf{g}'\mathbf{y}_{jk}$, gdzie wektor $\mathbf{g} \neq \mathbf{0}$.

Z rozkładu wektora losowego $\mathbf{z}_{jk}$ przedstawionego w (2.15) można określić jego rozkład warunkowy, gdy zmienną ustaloną jest funkcja $\mathbf{x}_{jk}$. Wówczas

$$\mathbf{z}_{jk} \mid \mathbf{x}_{jk} \sim N_I\{\mathrm{E}(\mathbf{z}_{jk}) + \boldsymbol{\beta}(\mathbf{g})[\mathbf{x}_{jk} - \mathrm{E}(\mathbf{x}_{jk})],\ \boldsymbol{\Sigma}_z - \boldsymbol{\beta}(\mathbf{g})\mathbf{g}'\boldsymbol{\Sigma}_y\mathbf{G}\}\ , \qquad (6.1)$$

gdzie

$$\boldsymbol{\beta}(\mathbf{g}) = \mathbf{G}\boldsymbol{\Sigma}_y\mathbf{g}/\mathbf{g}'\boldsymbol{\Sigma}_y\mathbf{g}$$

jest wektorem współczynników regresji wektora losowego $\mathbf{z}_{jk}$ względem zmiennej losowej $\mathbf{x}_{jk}$ (por. Anderson, 1958), którego estymator nieobciążony o wariancji minimalnej ma postać

$$\hat{\boldsymbol{\beta}}(\mathbf{g}) = \mathbf{G}\hat{\boldsymbol{\Sigma}}_y\mathbf{g}/\mathbf{g}'\hat{\boldsymbol{\Sigma}}_y\mathbf{g} = \mathbf{G}\mathbf{S}_E\mathbf{g}/\mathbf{g}'\mathbf{S}_E\mathbf{g}\ . \qquad (6.2)$$

Hipoteza o braku regresji efektów interakcji *GE* względem efektu środowiska, $\mathrm{H}_{\beta(\mathbf{g})}$: $\boldsymbol{\beta}(\mathbf{g}) = \mathbf{0}$, którą można zapisać równoważnie jako $\mathrm{H}_{\mathbf{C}_0'\beta(\mathbf{g})}$: $\mathbf{C}_0'\boldsymbol{\beta}(\mathbf{g}) = \mathbf{0}$, ($\mathbf{C}_0'\boldsymbol{\beta}(\mathbf{g}) = \mathbf{0}$ wtedy i tylko wtedy, gdy $\boldsymbol{\beta}(\mathbf{g}) = \mathbf{0}$), może być weryfikowana za pomocą statystyki

$$T^2_{\mathbf{C}_0'\beta(\mathbf{g})} = T^2_{\beta(\mathbf{g})} = (\nu_E - 1)\frac{\mathbf{g}'\mathbf{S}_E\mathbf{G}\mathbf{C}_0(\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0)^{-1}\mathbf{C}_0'\mathbf{G}\mathbf{S}_E\mathbf{g}}{\mathbf{g}'\mathbf{S}_E\mathbf{g} - \mathbf{g}'\mathbf{S}_E\mathbf{G}\mathbf{C}_o(\mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0)^{-1}\mathbf{C}'_0\mathbf{G}\mathbf{S}_E\mathbf{g}}\ . \qquad (6.3)$$

Statystyka (6.3) jest statystyką T^2 Hotellinga, która pomnożona przez $(\nu_E - I + 1)/\{(\nu_E - 1)(I-1)\}$ ma przy prawdziwości hipotezy $\mathrm{H}_{\beta(\mathbf{g})}$ centralny rozkład F z $I-1$ oraz $\nu_E - I + 1$ stopniami swobody.

Przejdźmy teraz do zagadnienia znalezienia odpowiedniej procedury testowania jednoczesnego skończonej rodziny hipotez $\mathrm{H}_{\mathbf{c}_{i^o}'\beta}$: $\mathbf{c}_{i^o}'\boldsymbol{\beta} = 0$, ($i^o = 1,2,\ldots,I^o$), implikowanych przez hipotezę $\mathrm{H}_{\beta(\mathbf{g})}$. Procedura konsekwentna i harmonijna polegać tutaj będzie na odrzuceniu hipotezy $\mathrm{H}_{\mathbf{c}_{i^o}'\beta}$ wtedy i tylko wtedy, gdy

$$T_{\mathbf{c}_{i^o}'\beta(\mathbf{g})} = (\nu_E - 1)\frac{(\mathbf{c}_{i^o}'\mathbf{G}\mathbf{S}_E\mathbf{g})^2}{\mathbf{g}'\mathbf{S}_E\mathbf{g}\,\mathbf{c}_{i^o}'\mathbf{S}_{GE}\mathbf{c}_{i^o} - (\mathbf{c}_{i^o}'\mathbf{G}\mathbf{S}_E\mathbf{g})^2} > T^2_\alpha\ , \qquad (6.4)$$

gdzie wartość krytyczna T^2_α jest dobrana tak, aby zachodziła równość

$$P\{\max_{1 \le i^o \le I^o} T^2_{\mathbf{c}_{i^o}'\beta(\mathbf{g})} > T^2_\alpha \mid \bigcap_{i^o=1}^{I^o} \mathrm{H}_{\mathbf{c}_{i^o}'\beta(\mathbf{g})}\} = \alpha\ .$$

Ponieważ występują trudności ze znalezieniem wartości T^2_α, również tutaj skorzystamy z nierówności Bonferroniego, zastępując nieznaną wartość T^2_α wartością nieco od niej większą, mianowicie wartością $F_{\alpha/I^o;1,\nu_E-1}$.

Odrzucenie hipotezy $\mathrm{H}_{\beta(\mathbf{g})}$, a tym samym stwierdzenie istotności regresji efektów interakcji względem $\mathbf{x}_{jk} = \mathbf{g}'\mathbf{y}_{jk}$, pociąga za sobą pytanie, czy rozważana

interakcja jest wyjaśniona tą regresją całkowicie. Tak więc postawione powyżej pytanie jest w rzeczywistości pytaniem, czy prawdziwa jest hipoteza

$$\mathrm{H}_{GE|E}: \ \boldsymbol{\Sigma}_{GE|E} = \mathbf{0} \ , \tag{6.5}$$

gdzie

$$\boldsymbol{\Sigma}_{GE|E} = \boldsymbol{\Sigma}_{GE} - \boldsymbol{\beta}(\mathbf{g})\mathbf{g}'\boldsymbol{\Sigma}_y\mathbf{G} = \boldsymbol{\Sigma}_{GE} - \mathbf{g}'\boldsymbol{\Sigma}_y\mathbf{g}\ \boldsymbol{\beta}(\mathbf{g})\boldsymbol{\beta}'(\mathbf{g})$$

jest macierzą kowariancji w rozkładzie warunkowym wektora $\mathbf{a}^{GE}(j,k)$ przy ustaleniu zmiennej $\mathbf{x}_{jk}$. Prawdziwość tej hipotezy oznacza, że jeśli interakcja genotypowo-środowiskowa istnieje, to jest ona całkowicie uwarunkowana regresją względem zmiennej $\mathbf{x}_{jk}$.

Statystykę do weryfikacji hipotezy (6.5) lub równoważnej z nią hipotezy $\mathrm{H}_{GE|E}: \ \sigma^2_{GE|E} = 0$, gdzie $\sigma^2_{GE|E} = (I-1)^{-1}\ \mathrm{tr}\ \{\boldsymbol{\Sigma}_{GE|E}\}$, jest iloraz

$$F_{GE|E} = \mathrm{tr}\ \{(\mathbf{C}_0'\mathbf{G}\boldsymbol{\Omega}_e\mathbf{G}\,\mathbf{C}_0)^{-1}\mathbf{C}_0'\mathbf{S}_{GE|E}\,\mathbf{C}_0\}\,/\,\{(I-1)(\nu_E-1)MS_e\} \ , \tag{6.6}$$

przy czym

$$\mathbf{C}_0'\mathbf{S}_{GE|E}\mathbf{C}_0 = \mathbf{C}_0'\mathbf{S}_{GE}\mathbf{C}_0 - \mathbf{C}_0'\mathbf{G}\,\mathbf{S}_E\,\mathbf{g}\,(\mathbf{g}'\mathbf{S}_E\,\mathbf{g})^{-1}\mathbf{g}'\mathbf{S}_E\,\mathbf{G}\mathbf{C}_0 \ . \tag{6.7}$$

Ma ona przy prawdziwości hipotezy $\mathrm{H}_{GE|E}$ rozkład centralny F z $(I-1)(\nu_E-1)$ i $N\nu_e$ stopniami swobody. Hipotezę $\mathrm{H}_{GE|E}$ odrzucimy więc na poziomie istotności α wtedy i tylko wtedy, gdy $F_{GE|E} > F_{\alpha,(I-1)(\nu_E-1),N\nu_e}$.

Aby znaleźć odpowiednią procedurę dla testowania jednoczesnego rodziny hipotez złożonej z hipotezy ogólnej $F_{GE|E}$, oraz implikowanych przez nią hipotez szczegółowych postaci

$$\mathrm{H}_{\mathbf{c}_{i^o}'GE|E}: \ \mathbf{c}_{i^o}'\mathbf{G}\boldsymbol{\Sigma}_{GE|E}\,\mathbf{G}\,\mathbf{c}_{i^o} = 0 \ , \tag{6.8}$$

należy postąpić analogicznie jak to uczyniono w rozdziale 5. Skończoną rodzinę I^o hipotez implikowanych przez $\mathrm{H}_{GE|E}$ będziemy weryfikować procedurą testów jednoczesnych $F_{\max}$ opartą na statystyce

$$F_{\mathbf{c}_{i^o}'GE|E} = MS_{\mathbf{c}_{i^o}'GE|E}\,/\,MS_e \ , \tag{6.9}$$

gdzie

$$MS_{\mathbf{c}_{i^o}'GE|E} = (\mathbf{c}_{i^o}'\mathbf{G}\boldsymbol{\Omega}_e\mathbf{G}\,\mathbf{c}_{i^o})^{-1}\mathbf{c}_{i^o}'\mathbf{S}_{GE|E}\,\mathbf{c}_{i^o}\,/\,(\nu_E-1) \ . \tag{6.10}$$

Statystykę (6.9) porównywać będziemy z jedną wspólną wartością krytyczną $\tilde{F}_{\max,\alpha}$ aproksymowaną nierównością Bonferroniego

$$\tilde{F}_{\max,\alpha} = F_{\alpha/I^o;\nu_E-1,N\nu_e} \ . \tag{6.11}$$

Powyższa teoria dotycząca regresji w badaniu interakcji genotypowo środowiskowej jest przedstawiona w sposób ogólny poprzez zastosowanie wektora **g**. W praktyce najczęściej przyjmuje się, że $\mathbf{g} = \mathbf{1}/I$. Implikuje to, że $\mathbf{x}_{jk} = \mathbf{y}_{.jk}$, co znaczy, że mamy do czynienia z regresją interakcji genotypowo środowiskowej względem średniej środowiskowej.

7. Przykład zastosowania

Dla zilustrowania przebiegu analizy statystycznej serii doświadczeń odmianoych posłużymy się wynikami N=26 doświadczeń z I=10 odmianami (genotypami) pszenicy ozimej (Jana(1), Modra(2), Begra(3), Beta(4), Liwilla(5), Salwa(6), Asta(7), Emika(8), Jawa(9), Weneda(10)), przeprowadzonych w K=4 latach, 1982-85 w J=7 stacjach oceny odmian (Głogowa(G), Sulejów(S), Kawęczyn(K), Seroczyn(E), Marianowo(M), Krzyżewo(R), Cicibór(C)), położonych w środkowo-wschodniej części Polski (w rejonie IV według klasyfikacji COBORU). Dane te stanowią wycinek dużej serii doświadczeń prowadzonych przez COBORU z wieloma odmianami w ośmiu rejonach Polski. Przyjmujemy, że każde z rozpatrywanych doświadczeń założone było w jednakowym układzie o pięciu blokach niekompletnych.

W wyniku przeprowadzonej dla każdego doświadczenia analizy wariancji uzyskano tak zwane średnie poprawione dla odmian, y_{ijk} (Tabela 2), oraz sumy kwadratów dla błędu, pozwalające otrzymać łączną sumę kwadratów dla błędu doświadczeń tej serii (SS_e = 6167.42, przy ν_e = 676).

Średnie poprawione dla odmian z poszczególnych miejscowości i lat, zestawione w Tabeli 2, stanowią podstawę do wykonania analizy wariancji (Tabela 3). W analizie tej wyznaczone są sumy kwadratów odchyleń dla miejscowości(P), dla lat(T), dla środowisk(E), dla odmian(G), dla interakcji odmiany × miejscowości(GP) i odmiany × lata(GT) oraz interakcji odmiany × środowiska(GE). Obliczone są wartości funkcji testowych wyrażonych odpowiednimi statystykami F oraz podane są dla nich właściwe wartości krytyczne umożliwiające weryfikację hipotezy H_P o równości średnich efektów głównych miejscowości, hipotezy H_T o równości średnich efektów głównych lat, hipotezy H_E o braku efektów głównych środowisk, testowanych za pomocą ilorazów odpowiednich średnich kwadratów, oraz hipotezy H_G, o równości efektów głównych odmian, H_{GP} o braku interakcji odmian z miejscowościami i H_{GT} o braku interakcji odmian z latami, testowanych za pomocą testów wielozmiennych T^2 (dla H_G) i T_0^2 (dla H_{GP} oraz H_{GT}) transformowanych następnie do testu F. Hipoteza H_{GE} o braku interakcji odmian ze środowiskami wewnątrz miejscowości i lat może być weryfikowana za pomocą zwyczajnego testu F.

Tabela 2

Wartości średnie z doświadczeń dla poszczególnych miejscowości, lat i odmian

Miejscowość	Rok	Odmiana (Genotyp) 1	2	3	4	5	6	7	8	9	10
Głogowa (G)	1982	60.7	63.3	60.6	65.0	70.1	70.0	68.8	67.0	63.1	63.4
	1983	65.0	76.7	65.8	70.2	75.8	74.8	65.4	72.4	79.0	71.8
	1984	60.4	64.8	61.1	66.6	67.8	66.1	58.5	63.6	76.8	72.1
	1985	49.6	59.5	50.2	54.6	68.6	62.7	61.1	59.4	61.8	63.6
Sulejów (S)	1982	31.6	21.9	29.1	34.3	35.4	40.6	27.5	36.4	27.6	35.4
	1983	45.5	54.9	48.5	51.2	48.4	55.4	43.4	49.8	59.3	55.8
	1984	40.2	40.3	34.2	39.8	52.4	44.3	42.5	42.0	55.2	47.9
	1985	31.2	46.7	47.6	53.2	47.1	42.4	37.4	44.2	60.6	49.1
Kawęczyn (K)	1982	–	–	–	–	–	–	–	–	–	–
	1983	–	–	–	–	–	–	–	–	–	–
	1984	57.3	64.0	66.7	68.5	72.0	66.6	59.4	71.2	79.6	70.0
	1985	41.2	51.4	49.8	55.4	54.3	53.0	48.3	53.4	50.0	56.4
Seroczyn (E)	1982	48.0	34.1	45.6	47.0	50.5	53.5	49.1	49.6	47.9	53.1
	1983	54.8	64.5	58.6	63.2	58.9	67.0	53.1	55.4	69.0	64.4
	1984	55.7	52.2	45.2	57.0	66.0	62.5	51.1	63.3	62.1	56.6
	1985	61.9	60.8	48.0	60.2	45.6	60.4	55.1	61.9	56.4	57.1
Marianowo (M)	1982	39.8	41.9	41.5	42.7	47.4	53.2	47.9	49.3	38.9	41.8
	1983	69.8	69.4	69.6	78.2	74.2	74.8	62.8	71.4	77.8	75.4
	1984	52.0	33.4	35.9	42.4	48.0	48.9	49.8	49.8	34.7	40.5
	1985	49.6	44.8	47.0	51.1	52.5	51.2	52.2	56.8	44.7	49.3
Krzyżewo (R)	1982	56.4	63.2	56.5	53.1	62.0	61.6	55.1	59.3	65.9	62.2
	1983	69.0	70.2	67.3	74.8	66.5	75.2	68.3	67.0	81.7	63.0
	1984	56.8	51.5	58.2	58.7	60.0	61.9	53.8	58.0	64.7	61.8
	1985	53.7	55.7	52.1	58.7	53.8	57.8	50.8	61.8	56.1	55.1
Cicibór (C)	1982	31.4	22.2	39.9	36.9	42.5	48.1	33.6	44.8	32.7	38.9
	1983	57.2	56.1	60.7	61.2	59.8	61.0	48.4	53.9	67.7	66.0
	1984	43.2	46.9	42.9	46.1	52.0	48.2	41.8	49.1	51.7	49.6
	1985	48.6	43.9	46.1	51.9	44.0	49.5	41.9	51.4	43.6	44.7

W rozważanym przykładzie niemal wszystkie weryfikowane hipotezy odrzucone zostały co najmniej na poziomie istotności $\alpha = 0.05$. Spośród uzyskanych wyników testowań najbardziej interesujący jest fakt wysoce istotnego zróżnicowania odmian pod względem ich efektów głównych oraz wystąpienie istotnej, losowej interakcji odmianowo-środowiskowej. Ze stałych interakcji odmian z miejscowościami i odmian z latami tylko ta ostatnia okazała się znacząca. Po-

Tabela 3
Analiza wariancji

Źródło zmienności	Liczba stopni swobody	Suma kwadratów	Średni kwadrat	Statystyka F	Wartość krytyczna	
					5%	1%
Miejscowości(P)	6	14797.49	2466.25	7.84	2.74	4.20
Lata(T)	3	9338.80	3112.93	9.90	3.24	5.29
Środowiska(E)	16	5031.46	314.47	137.92	1.66	2.03
Regresja względem interakcji GE	9	4236.17	470.69	4.14	3.68	6.72
Odchylenia od regresji	7	795.29	113.61			
Odmiany(G)	9	2062.09	229.12	24.43	3.39	5.91
Odmiany× miejscowości (GP)	54	1234.56	22.86	0.93	2.23	3.20
Odmiany×Lata (GT)	27	1266.20	96.90	3.20	2.48	3.74
Odmiany×Środowiska (GE)	144	2151.82	14.94	6.35	1.23	1.34
Regresja względem środowiska	9	372.65	41.41			
Odchylenia od regresji	135	1779.17	13.18	5.61	1.23	1.35
Błąd doświadczeń	676	6167.42	9.12			

wstaje pytanie, w jakim stopniu tę niejednakową reakcję odmian na zmiany warunków środowiska można wytłumaczyć ich regresją liniową względem efektów środowiska. Dokonany został w tym celu podział sumy kwadratów dla interakcji GE na dwa składniki; sumę kwadratów dla regresji efektów interakcji GE względem efektu środowiska i sumę kwadratów dla odchyleń od tej regresji. Równolegle dokonano rozbicia sumy kwadratów dla środowisk na sumę kwadratów dla regresji wielokrotnej efektów środowiska względem efektów interakcyjnych dla poszczególnych odmian i sumę kwadratów dla odchyleń od tej regresji. W rozważanym przykładzie hipoteza $\boldsymbol{\beta} = \mathbf{0}$ o braku regresji efektów interakcji GE względem efektu środowiska została odrzucona na poziomie istotności $\alpha = 0.05$ ($F_{\beta} = 4.14$ przy $F_{0.05} = 3.68$). Oznacza to, że istotną część interakcji GE można wyjaśnić za pomocą wspomnianej regresji. O tym czy pozostała część interakcji

odmian ze środowiskami jest nieistotna można dowiedzieć się porównując wartość statystyki F dla odchyleń od regresji z odpowiednimi wartościami krytycznymi. W naszym przykładzie wartość tej statystyki, wynosząca 3.61, znacznie przewyższa obie wartości krytyczne $F_{0.05} = 1.23$ i $F_{0.01} = 1.35$, czyli jedynie część interakcji *GE* wynika z jej regresji liniowej względem efektu środowiska.

Testowane w analizie wariancji hipotezy dotyczyły informacji ogólnych o badanym zestawie odmian w sensie równości ich efektów głównych oraz wielkości interakcji *GP*, *GT* i *GE*. Aby uzyskać informacje dotyczące poszczególnych odmian niezbędne jest przeprowadzenie testowania odpowiednich hipotez szczegółowych implikowanych przez hipotezy ogólne. Analiza poszczególnych odmian i ich interakcji z miejscowościami, latami i środowiskami przedstawiona jest w Tabeli 4. W tabeli tej każda z odmian oceniona jest pod względem przeciętnej wysokości plonowania, wyrażonej efektem głównym badanej odmiany i odpowiadającą jej wartością statystyki F, a także pod względem wielkości interakcji danej odmiany

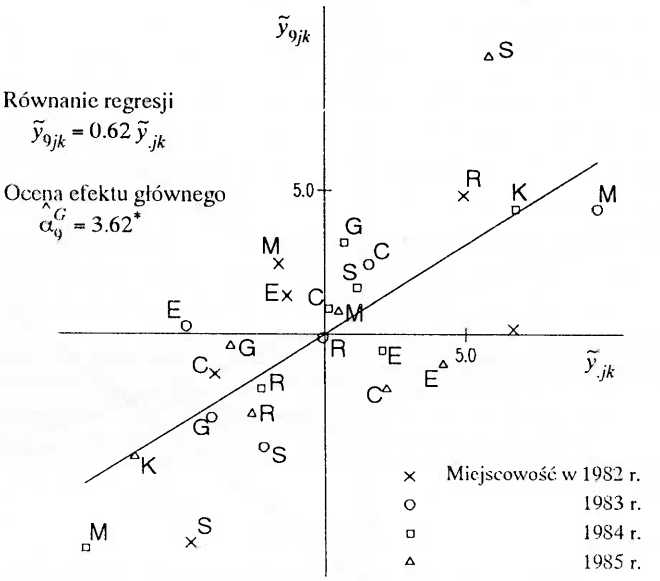
Tabela 4

Oceny i wyniki testowania hipotez dotyczących odmian i ich interakcji *GP*, *GT* oraz *GE*

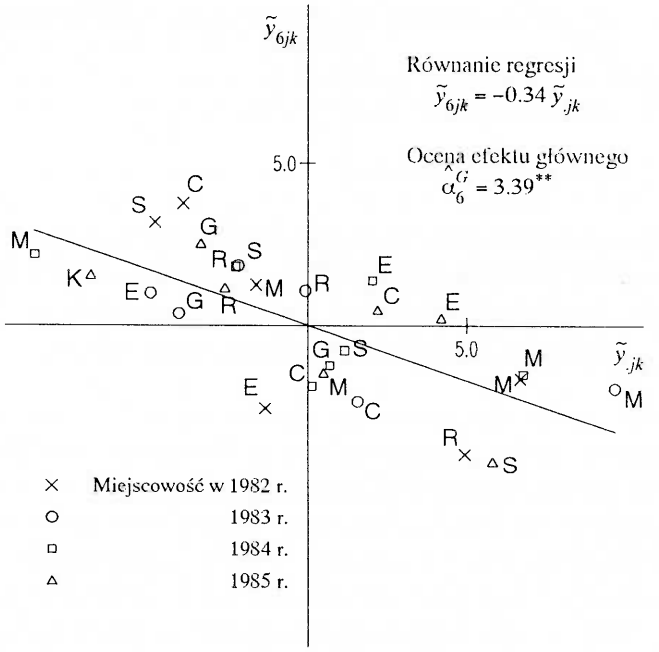
Odmiana	Ocena efektu głównego	Statystyka F dla				Współczynnik		Statystyka F dla	
		efektu głównego	*GP*	*GT*	*GE*	regresji	determinacji	regresji	odchyleń
1	–4.10	21.43	2.37	0.47	8.89	–0.16	4.2	0.67	9.08
2	–2.59	7.33	0.77	3.49	10.41	0.75	8.7	1.43	10.14
3	–3.15	27.13	2.62	2.42	4.16	0.05	1.0	0.16	4.39
4	0.84	1.84	0.57	2.82	4.37	0.14	6.4	1.03	4.36
5	2.12	6.61	0.95	2.50	7.74	–0.12	2.7	0.41	8.03
6	3.39	38.72	0.71	3.87	3.37	–0.34	50.1	15.07	1.79
7	–3.90	27.64	1.69	3.84	6.25	–0.32	24.1	4.77	5.06
8	1.45	5.15	1.02	5.14	4.66	–0.14	6.1	0.97	4.67
9	3.62	14.14	2.42	4.90	10.51	0.62	54.4	17.93	5.11
10	2.32	18.53	1.33	0.84	3.28	0.01	0.1	0.01	3.50
Wartości krytyczne dla testowania jednoczesnego na poziomie									
a = 0.05		10.58	4.91	6.30	2.17			10.80	2.22
a = 0.01		16.12	6.80	9.01	2.50			16.59	2.56
(przybliżone nierównością Bonferroniego)									

z miejscowościami, latami i środowiskami wyrażonych wartościami odpowiednich statystyk F.

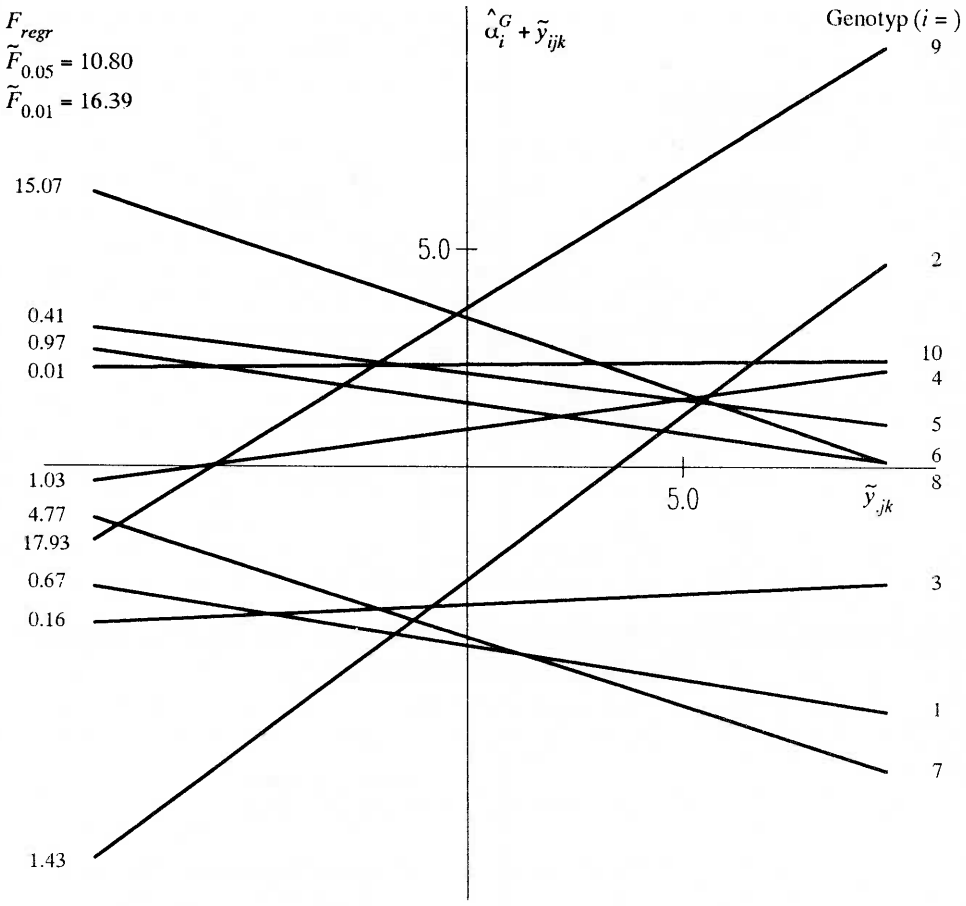
Badane jest także występowanie regresji w interakcji GE, to znaczy wyliczone są współczynniki regresji i determinacji oraz testowane są hipotezy o braku regresji efektów interakcyjnych odmian względem efektu środowiska i hipotezy o braku odchyleń od regresji. Spośród badanych odmian najbardziej interesującymi okazały się odmiany Salwa(6), Weneda(10) i Jawa(9), a w pewnym stopniu także odmiana Liwilla(5). Zwracają uwagę, zamieszczone w Tabeli 4, wysokie oceny efektów głównych tych odmian, dla których, wyłączywszy odmianę Liwilla(5), wartości statystyki F przekraczają wartość krytyczną $\widetilde{F}_{0.05} = 10.58$. Wszystkie te odmiany odznaczają się ponadto bardzo niskimi, nieistotnymi na poziomie $\alpha = 0.05$, interakcjami z miejscowościami i latami, a dwie pierwsze także najniższą spośród wszystkich badanych odmian interakcją ze środowiskami. Z kolei odmiana Jawa(9) należy do odmian najbardziej wrażliwych na zmieniające się warunki środowiska, o czym świadczy bardzo wysoka wartość statystyki F (=10.51) dla interakcji GE, przewyższająca zdecydowanie wartość krytyczną $\widetilde{F}_{0.01} = 2.50$. Okazuje się jednak, że interakcję tej odmiany, a także interakcję GE odmiany Salwa(6), można w znacznej mierze wyjaśnić regresją efektów interakcyjnych tych odmian względem efektów środowiska ($F_{\beta_9} = 17.93$, $F_{\beta_6} = 15.07$ przy $\widetilde{F}_{0.05} = 10.80$ i $\widetilde{F}_{0.01} = 16.59$). Współczynniki regresji i determinacji dla odmiany Jawa(9) i Salwa(6) wynoszą odpowiednio $\hat{\beta}_9 = 0.62$ (54.4%) i $\hat{\beta}_6 = -0.34$ (50.1%), przy czym dla tej ostatniej odmiany nieistotną okazała się również statystyka dla odchyleń od regresji, F=1.79 wobec $\widetilde{F}_{0.05} = 2.22$. Biorąc pod uwagę powyższe rezultaty obliczeń oraz średnie plony odmian Jawa(9) i Salwa(6), najwyższe ze wszystkich, można stwierdzić, że odmiana Jawa(9) jest wybitnie intensywna, to znaczy plonująca szczególnie dobrze w tych środowiskach, w których plonowanie jest przeciętnie wyższe, natomiast odmiana Salwa(6) jest odmianą ekstensywną, czyli plonującą dobrze w środowiskach, generalnie nie odznaczających się wysokim przeciętnym plonowaniem. Pełniejszą interpretację interakcji obu odmian można uzyskać analizując dodatkowo Rysunki 1 i 2, przedstawiające dla każdej z tych odmian rozmieszczenie środowisk w układzie współrzędnych prostokątnych o osi odciętych reprezentującej zmienną $\tilde{y}_{jk}$ o wartościach będących elementami wektora $\mathbf{Q}_E\mathbf{Y}\,\mathbf{1}\,/I$, określającą obserwowane efekty główne środowiska, i osi rzędnej reprezentującej zmienną $\tilde{y}_{9jk}$ ($\tilde{y}_{6jk}$ dla odmiany Salwa) o wartościach stanowiących 9-tą (6-tą) kolumnę macierzy $\mathbf{A}^{GE} = \mathbf{Q}_E\mathbf{Y}\,\mathbf{G}$, określającej efekty interakcyjne GE analizowanej odmiany. Na rysunkach tych punktami zaznaczone są środowiska, przy czym ich znaczny rozrzut wokół osi odciętych jak to ma miejsce na Rysunku 1 wykreślonym dla odmiany Jawa(9) świadczy o małej stabilności tej odmiany, to znaczy o jej dużej wrażliwości na zmiany warunków środowiska, natomiast stosunkowo niewielki rozrzut punktów reprezentujących



Rys. 1. Regresja efektów interakcyjnych *GE* odmiany Jawa(9), $\tilde{y}_{9jk}$, względem efektów głównych środowiska $\tilde{y}_{.jk}$



Rys. 2. Regresja efektów interakcyjnych *GE* odmiany Salva(6), $\tilde{y}_{6jk}$, względem efektów głównych środowiska $\tilde{y}_{.jk}$



Rys. 3. Proste regresji dla poszczególnych odmian wyznaczone za pomocą analizy regresji efektów interakcyjnych *GE*, $\tilde{y}_{ijk}$, względem efektów głównych środowiska, $\tilde{y}_{.jk}$. Punkty przecięcia tych prostych z osią rzędnych wyznaczają oceny efektów głównych odpowiednich odmian.

środowiska jaki można obserwować na Rysunku 2, wykreślonym dla odmiany Salwa(6), wskazuje na jej większą stabilność. Wykreślone na Rysunkach 1 i 2 proste regresji opisane równaniami odpowiednio $\tilde{y}_{9jk} = 0.62\,\tilde{y}_{.jk}$ i $\tilde{y}_{6jk} = -0.34\,\tilde{y}_{.jk}$ informują dodatkowo o różnych kierunkach wpływu warunków środowiskowych na obie odmiany, dodatnim dla odmiany Jawa(9) i ujemnym dla odmiany Salwa(6). Niezbyt duży rozrzut punktów wokół prostej regresji, widoczny na rysunku 2, świadczy o tym, że interakcja *GE* odmiany Salwa jest w znacznym stopniu wytłumaczona regresją.

Przy analizowaniu regresji dla poszczególnych odmian pomocny okazuje się Rysunek 3, który przedstawia indywidualne proste regresji dla odmian wyznaczoe w układzie współrzędnych $\tilde{y}_{jk}$ jako odciętych oraz $\hat{a}_i^G + \tilde{y}_{ijk}$ jako rzędnych. Punkt przecięcia się prostej z osią rzędnych pokazuje ocenę efektu głównego odpowiedniej odmiany. Przy każdej z wykreślonych prostych regresji podana jest wartość statystyki F dla regresji, porównywalna z wartościami krytycznymi na poziomie $\alpha = 0.05$ i $\alpha = 0.01$. Rysunek 3 potwierdza też graficznie wyniki uzyskane z analizowania Tabeli 4 dotyczące odmian najsłabszych w sensie plonowania w poszczególnych środowiskach. Odmiany Jana(1), Asta(7), Begra(3) to odmiany, których średnie plony są istotnie niższe od średniego plonu wszystkich 10-ciu analizowanych odmian. Co więcej, wysokiej interakcji tych odmian ze środowiskami nie można wytłumaczyć regresją, co tym samym uniemożliwia określenie dla tych odmian optymalnych warunków środowiskowych.

LITERATURA

Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Caliński, T. (1960). On a Certain Statistical Method of Investigating Interaction in Serial Experiments with Plant Varieties. *Bulletin de l'Academie Polonaise des Sciences* II, **8**, 565–568.

Caliński, T. (1967). Model analizy wariancji dla doświadczeń wielokrotnych. *Roczniki Nauk Rolniczych*, Ser. A, **93**, 3, 549–579.

Caliński, T., Czajka, S. and Kaczmarek, Z. (1987a). A model for the analysis of a series of experiments repeated at several places over a period of years. I. Theory. *Biuletyn Oceny Odmian* **10**, 7–33.

Caliński, T., Czajka, S. and Kaczmarek, Z. (1987b). A model for the analysis of a series of experiments repeated at several places over a period of years. II. Example. *Biuletyn Oceny Odmian* **10**, 35–71.

Caliński, T., Dyczkowski, A. i Sitek, M. (1979). Procedury testów jednoczesnych w wielozmiennej analizie wariancji. *Matematyka Stosowana* **14**, 5–31.

Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*. 2nd ed. New York: Wiley.

Comstock, R.E. and Moll, R.H. (1963). Genotype-environment interactions. Statistical Genetics and Plant Breeding. *Nat. Acad. Sc. Natl. Res. Coun. Publ.* **982**, 164–196.

Cox, D.R. (1984). Interaction. *Int. Statist. Rev.* **52**, 1–31.

Eberhart, S.A. and Russell, W.A. (1966). Stability parameters for comparing varietes. *Crop Science* **6**, 26–40.

Elandt, R. (1956). O pewnych testach interakcji w doświadczeniach wieloletnich i wielokrotnych. Zagadnienie rejonizacji. *Zastosowania Matematyki* **3**, 8–45.

Federer, W.T. (1955). *Experimental Designs*. New York: Wilay.

Finlay, K.W. and Wilkinson, G.N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research* **14**, 742–754.

Finney, D.J. (1964). The replication of variety trials. *Biometrics* **20**, 1–15.

Freeman, G.H. (1973). Statistical Methods for the analysis of genotype-environment interactions. *Heredity* **31**, 339–354.

Freeman, G.H. (1975). Analysis of interactions in incomplete two-way tables. *Applied Statistics* **24**, 46–55.

Freeman, G.H. (1985). The analysis and interpretation of interactions. *Journal of Applied Statistics* **12**, 1, 3–10.

Freeman, G.H. and Crisp, P. (1979). The use of related variables in explaning genotype-environment interactions. *Heredity* **42**, 1–11.

Freeman, G.H. and Dowker, B.D. (1973). The analysis of variation between and within genotypes and environments. *Heredity* **30**, 97–109.

Gabriel, K.R. (1968). Simultaneous test procedures in multivariate analysis of variance. *Biometrika* **55**, 489–504.

Hill, J. (1975). Genotype environment interactions – a challenge for plant breeding. *Journal of Agricultural Science* **85**, 477–493.

Hinkelmann, K. (1974). Genotype environment interaction: Aspect of statistical designs, analysis and interpretation. *8th International Biometric Conference*, Constanta, 1–33.

Kaczmarek, Z. (1986). Analiza doświadczeń wielokrotnych zakładanych w blokach niekompletnych. Roczniki Akademii Rolniczej w Poznaniu, *Rozprawy Naukowe*, Zeszyt 155.

Kempton, R.A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science* **103**, 123–135.

Lin, C.S. (1982). Grouping genotypes by a cluster method directly related to genotype environment interaction mean square. *Theoretical and Applied Genetics* **62**, 277–280.

McKeon, J.J. (1974). *F* aproximations to the distribution of Hotteling's T_0^2. *Biometrika* **61**, 381–383.

Morrison, D.F. (1976). *Multivariate Statistical Methods*, 2nd ed. New York: Mc Graw-Hill.

Neyman, J. (1932). O metodach opracowywania doświadczeń wielokrotnych. *Roczniki Nauk Rolniczych i Leśnych* **28**, 154–210.

Okuno, T., Kikuchi, F., Kumagai, K., Okuno, C., Shiyomi, M. and Tabuchi, H. (1971). Evaluation of varietal performance in several environments. *The Bulletin of the National Institute of Agricultural Sciences*, Series A **18**, 93–147.

Pearson, E.S. and Hartley, H.O. (1972). *Biometrika Tables for Statisticians*. Vol. 2. Cambridge University Press.

Pillai, K.C.S. (1960). *Statistical Tables for Tests of Multivariate Hypotheses*. Manila: The Statistical Center, University of the Philippines.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.

Roy, S.N., Gnanadesikan, R. and Srivastava, J.N. (1971). *Analysis and Design of Certain Quantitative Multiresponse Experiments*. Oxford: Pergamon Press.

Shorter, R., Byth, D.E. and Mungomery, V.E. (1977). Genotype × environment interactions and environmental adaptation. 2. Assessment of environmental contributions. *Australian Journal of Agricultural Research* **28**, 223–235.

Shukla, G.K. (1972). Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity* **29**, 237–245.

Steel, R.G.D. and Torrie, J.H. (1980). *Principles and Procedures of Statistics.* 2nd ed. New York: Mc Graw-Hill.

Tai George, C.C. (1971). Genotypic stability analysis and its application to potato regional trials. *Crop Science* **14**, 587–590.

Tocher, K.D. (1952). The design and analysis of block experiments. *Journal of the Royal Statistical Society*, Series B **14**, 49-91.

Wood, J.T. (1976). The use of environmental variables in the interpretation of genotype-environment interaction. *Heredity* **37**, 1–7.

Wricke, G. (1962). Uber eine Methode zur Erfassung der okologschen Streubreite in Feldversuchen. *Z. Pflanzenzuchtung* **47**, 92–96.

Wricke, G. (1964). Zur Berechnung der Okovalenz bei Sommerweizen und Hafer. *Z. Pflanzenzuchtung* **52**, 127–138.

Wricke, G. (1965). Die Erfassung der Wechselwirkung zwischen Genotyp und Unvelt beiquantitativan Eigenchaften. *Z. Pflanzenzuchtung* **53**, 266–343.

Wright, A.J. (1976). Bias in the estimation of regression coefficients in the analysis of genotype-environmental interaction. *Heredity* **37**, 299–303.

Yates, F. and Cochran, W.G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science* **28**, 556–580.

Praca wpłynęła 26.09.1995

Statistical methods for the analysis of a series of multiple experiments carried out in incomplete block design

Summary

A model for analysing of results of a series of experiments repeated with the same set of genotypes in block design at many places over a period of years is presented. For this model the analysis of variance and tests of general hypotheses concerning interesting sources of variation are given. Simultaneous test procedures for various families of particular hypotheses are proposed. Furthermore some linear regression methods for studying relations between genotype-environment interactions and the environmental mean are suggested. The application of the theory presented in the paper is illustrated by a practical example of variety testing experiments.

Key words: analysis of a series of experiments, multivariate analysis of variance, hypothesis testing, regression analysis