W. OKTABA (Lublin)

*ESTYMACJA KOMPONENTÓW WARIANCYJNYCH
W NIEORTOGONALNYCH MODELACH LOSOWYCH
OPARTYCH NA KOMBINACJI KLASYFIKACJI KRZYŻOWEJ
Z HIERARCHICZNĄ*

1. Przegląd piśmiennictwa i problem. Interesujemy się takim modelem losowym, w którym efekty są nieskorelowane i mają średnie równe zeru oraz pewne nieznane wariancje $\sigma_A^2, \sigma_B^2, \dots, \sigma_e^2$, nazywane również komponentami wariancyjnymi (por. [3], [4], [20]). Korzystamy z terminologii przyjętej w pracy [15].

W przypadku danych nieortogonalnych dokładne metody oceny wymienionych komponentów, a mianowicie: bezpośrednia i pośrednia (por. [8]) polegają m. in. na obliczeniu wartości oczekiwanych średnich kwadratów według zasady najmniejszych kwadratów dla poszczególnych źródeł zmienności. Z uwagi na konieczność odwracania macierzy dla znalezienia współczynników przy komponentach w oczekiwanych wartościach, metody te mogą być stosowane wtedy, gdy dane liczbowe są zgrupowane w niezbyt wielkiej liczbie podklas. W przypadku okazałej ich liczby trudności rachunkowe mogą być usunięte z pomocą maszyny elektronicznej lub przy użyciu tzw. pierwszej przybliżonej metody Hendersona (por. [10] i [19]).

Metoda ta, z której częściowo będziemy korzystali, polega na obliczaniu sum kwadratów, jak w standardowej analizie wariancji, odpowiednich danych ortogonalnych oraz na przyrównaniu tych sum do ich wartości oczekiwanych. Rozwiązanie otrzymanego układu równań ze względu na niewiadome wariancje, dostarcza nieobciążonych ocen wariancji poszczególnych efektów (por. [18]).

Poniżej, w punktach A, B i C, wspominamy o problematyce ocen komponentów wariancyjnych odpowiednio dla modeli: 1) ortogonalnego; 2) nieortogonalnego klasyfikacji hierarchicznej i 3) nieortogonalnego klasyfikacji krzyżowej. Problem pracy referujemy w punkcie D dla nieortogonalnego modelu opartego na dowolnej kombinacji obu klasyfikacji

krzyżowej i hierarchicznej (por. [14]) i na niekompletnej klasyfikacji krzyżowej (por. § 2).

A. *Dowolny model ortogonalny*. Przez dane ortogonalne rozumiemy dane liczbowe, zawierające w każdej z podklas równą liczbę obserwacji. Zagadnienie estymacji komponentów wariancyjnych w ortogonalnych modelach nie tylko losowych, lecz również stałych, mieszanych i w tak zwanych modelach skończonej populacji najogólniejszego typu, a więc opartych na dowolnej kombinacji klasyfikacji krzyżowej z hierarchiczną, zostało rozwiązane przez Bennetta i Franklina ([2]). Reguła Crumpa (por. [5]) jest szczególnym przypadkiem reguły Bennetta-Franklina, gdy ograniczymy się do modelu losowego.

Jak wiadomo, reguła Crumpa głosi, że wartość oczekiwana średniego kwadratu dowolnego efektu jest kombinacją liniową wariancji losowej σ_e^2 i tylko tych wariancji, których wskaźniki zawierają wszystkie litery efektu; współczynniki przy tych wariancjach są iloczynami liczb poziomów tych czynników, których jest brak we wskaźnikach wariancji. Tak np. $E(V_{BC})$, czyli wartość oczekiwana średniego kwadratu dla interakcji BC w przypadku trójkierunkowej klasyfikacji krzyżowej $A \times B \times C$ z $n = abc$ obserwacjami i a, b, c — liczbami poziomów klasyfikacji A, B i C , obejmuje wariancję losową σ_e^2 ze współczynnikiem 1 i wariancje σ_{ABC}^2 i σ_{BC}^2 ze współczynnikami, które są iloczynami liczby powtórzeń r i liczb poziomów tych czynników, które nie występują przy składnikach wariancyjnych, tj.

$$E(V_{BC}) = \sigma_e^2 + r\sigma_{ABC}^2 + r\sigma_{BC}^2.$$

Iloczyn liczby powtórzeń i liczby poziomów czynników będących wskaźnikami przy wariancji jest (z wyjątkiem σ_e^2) równy n (np. $r \cdot abc = = ra \cdot bc = n$). Korzystając z tej reguły można natychmiast wypisać wartości oczekiwane średnich kwadratów dla poszczególnych źródeł zmienności modelu losowego. Rozwiązanie układu równań powstałych z przyrównania wymienionych wartości oczekiwanych do odpowiadających im średnich kwadratów daje nieobciążone oceny komponentów wariancyjnych.

B. *Dowolny model nieortogonalny klasyfikacji hierarchicznej*. Prosty problem oceny komponentów wariancyjnych dla przypadku ortogonalnego staje się nieco bardziej złożony dla danych nieortogonalnych, tj. danych, w których liczby obserwacji zmieniają się nieregularnie od podklasy do podklasy. Dla modeli nieortogonalnych i losowych opartych na klasyfikacji hierarchicznej zagadnienie oceny komponentów wariancyjnych zostało rozwiązane przez statystyka hinduskiego Ganguli ([6]). Odpowiednie uwagi można znaleźć w innych pracach (por. [1]); przypadek ogólny klasyfikacji hierarchicznej dowolnej krotności, tj. z dowolną

liczbą czynników wraz z przykładem liczbowym dla pięciu czynników występuje w pracy autora (por. [16]).

Podobnie jak w przypadku ortogonalnym, estymację komponentów wariancyjnych realizujemy przez rozwiązanie układu równań powstałych z przyrównania wartości oczekiwanych średnich kwadratów poszczególnych czynników do ich średnich kwadratów. Jednakże współczynniki, które są funkcjami liczebności podklas przy komponentach wariancyjnych w wymienionych równaniach, jakkolwiek tworzą się w sposób regularny, mają postać raczej złożoną.

C. *Dowolny model nieortogonalny klasyfikacji krzyżowej.* Zagadnienie wyznaczania ocen komponentów wariancyjnych dla nieortogonalnego modelu losowego opartego na klasyfikacji krzyżowej z dowolną liczbą czynników rozwiązał Le Roy ([12]). Przedstawił on metodę, nazwijmy ją R, pozwalającą wypisać wartości oczekiwane składników sum ważonych kwadratów odchyłeń poszczególnych czynników jako kombinacje liniowe komponentów wariancyjnych. Współczynniki przy tych komponentach wyznacza się bezpośrednio według prostej reguły jako funkcje liczebności podklas bez potrzeby obliczania wartości oczekiwanych średnich kwadratów. Rozwiązanie układu równań powstałych z przyrównania wartości oczekiwanych wymienionych składników sum do wartości liczbowych tych składników daje oceny komponentów.

D. *Losowy model nieortogonalny, oparty na dowolnej kombinacji klasyfikacji krzyżowej z hierarchiczną,* występuje często w zastosowaniach, np. w pewnych problemach genetycznych. Z uwagi na to, że jest on uogólnieniem wyżej wspomnianych modeli, interesującą rzeczą jest rozwiązanie zagadnienia oceny komponentów wariancyjnych w tym przypadku. Było to celem niniejszej pracy.

Dostrzeżliśmy, że dla uzyskania ocen komponentów wariancyjnych można zastosować metodę R dla dowolnego nieortogonalnego modelu losowego bez względu na to, czy opiera się on na klasyfikacji krzyżowej, hierarchicznej czy też na dowolnej kombinacji obu klasyfikacji. Stosuje się ona również do danych liczbowych z pewną liczbą podklas pustych (por. § 5). Uzasadnienie teoretyczne wraz z odpowiednim twierdzeniem i dowodem znajduje się w innej pracy autora ([17]). Z twierdzenia tego wynika reguła postępowania, która pozwala w prosty sposób wyznaczyć współczynniki przy komponentach w układzie równań ze względu na niewiadome wariancje bez uciekania się do każdorazowego obliczania wartości oczekiwanych średnich kwadratów. Szczególnym przypadkiem tej reguły, nazwijmy ją RO, jest reguła R. Metoda RO oceny komponentów wariancyjnych nieortogonalnych modeli losowych opiera się na metodzie R w tym sensie, że przenosi jej reguły postępowania z przypadku klasyfikacji krzyżowej na ogólną, tj. na dowolną kombinację

klasyfikacji krzyżowej z hierarchiczną (k. k. k. h) i na modele oparte na danych z pewnymi podklasami bez obserwacji. Metoda RO uogólnia symbolikę metody R przez wprowadzenie rygorystycznie określonej notacji dla modeli statystycznych (oznaczenie źródeł zmienności w analizie wariancji, efektów głównych i interakcyjnych w modelu, wariancji w populacji oraz ich ocen dla dowolnych modeli opartych na k. k. k. h.) ze wskazaniem na wykorzystanie metody opisanej w [13] do modeli nieortogonalnych. Prosta symbolika dla klasyfikacji krzyżowej nieco się komplikuje w k. k. k. h. W związku z tym metoda R zostaje odpowiednio zmodyfikowana i przybiera postać metody RO. Wprowadzenie rygorystycznej symboliki pozwala na wyraźne uwypuklenie różnicy w interpretacji wariancji i ich ocen dla czynników rozważanych klasyfikacji hierarchicznej i krzyżowej (oznaczono np. σ_A^2 , σ_C^2 , σ_{AC}^2 dla klasyfikacji krzyżowej, σ_A^2 i $\sigma_{C(A)}^2$ dla klasyfikacji hierarchicznej oraz $\sigma_{BC(A)}^2$, $\sigma_{D(C(A))}^2$, $\sigma_{BD(C(A))}^2$ dla k. k. k. h., por. §§ 2, 3, 4). Można łatwo sprawdzić, że dzięki regule RO otrzymamy w szczególności wzory podane przez Ganguli dla klasyfikacji hierarchicznej ([6]).

W następnych paragrafach przedstawimy regułę RO i zilustrujemy odpowiedni tok postępowania dla wyznaczenia komponentów wariancyjnych na przykładzie liczbowym opartym na kombinacji klasyfikacji krzyżowej i hierarchicznej.

Szereg przykładów i zagadnień dotyczących ogólnego problemu oceny komponentów wariancyjnych nie tylko w modelach losowych, lecz również mieszanych znajdzie czytelnik w wielu pracach (por. np. [7], [9], [11]).

2. Klasyfikacja krzyżowa i hierarchiczna. Ponieważ w niniejszej pracy jest mowa o danych liczbowych podlegających dowolnym klasyfikacjom, zdefiniujemy dwie klasyfikacje podstawowe: krzyżową i hierarchiczną ([18]).

DEFINICJA 1. Jeżeli każdy z I poziomów (klas), np. $A_1, A_2, \dots, A_I$, czynnika (klasyfikacji) A występuje z każdym z J poziomów, np. $B_1, B_2, \dots, B_J$, czynnika B , to czynniki A i B tworzą *dwukierunkową* (podwójną) *klasyfikację krzyżową* $A \times B$ z IJ podklasami.

Przyjmując $I = 3$ i $J = 9$ otrzymujemy schemat dwukierunkowej klasyfikacji krzyżowej, wyobrażony na rysunku 1 w formie zakreślonego prostokąta, złożonego z $IJ = 27$ prostokątów częściowych.

DEFINICJA 2. Jeżeli każdy z K poziomów czynnika C , np. $C_1, C_2, \dots, C_K$ występuje z każdą z IJ kombinacji poziomów czynników A i B , tj. z $A_1B_1, A_1B_2, \dots, A_IB_J$, to czynniki A, B i C tworzą *trójkierunkową* (potrójną) *klasyfikację krzyżową* $A \times B \times C$ z IJK podklasami.

W przestrzeni trójwymiarowej tej sytuacji odpowiada prostopadłościan złożony z IJK prostopadłościanów częściowych. Uogólnienie defi-

nieji na przypadek klasyfikacji krzyżowej $A \times B \times C \times \dots$, z dowolną liczbą czynników $A, B, C, \dots$, jest bezpośrednio widoczne.

Jeżeli w każdym z częściowych „prostokątów”, tj. w każdej z podklas, występuje taka sama liczba obserwacji $r \geq 1$, mówimy, że dane liczbowe podlegają *ortogonalnej klasyfikacji krzyżowej*, a jeżeli liczba

		<i>B</i>								
		1	2	3	4	5	6	7	8	9
<i>A</i>	1									
	2									
	3									

ZM-467

Rys. 1. Dwukierunkowa klasyfikacja krzyżowa $A \times B$

obserwacji jest różna, to dane liczbowe stanowią *nieortogonalną klasyfikację krzyżową*.

DEFINICJA 3. Jeżeli każdy poziom czynnika B występuje tylko z jednym poziomem czynnika A w obserwacjach, to mówimy, że poziomy czynnika B są *uplasowane hierarchicznie* wewnątrz czynnika A ; symbolicznie notujemy $B(A)$ i czytamy „ B w A ”.

Na rysunku 2 podajemy przykład klasyfikacji hierarchicznej $B(A)$; pola zakreskowane wskazują podklasy (kombinacje poziomów czynni-

		<i>B</i>								
		1	2	3	4	5	6	7	8	9
<i>A</i>	1									
	2									
	3									

lub

		<i>B</i>	
<i>A</i>	1	1	2
	2	3	4
	3	5	6
		7	8
		9	

ZM-468

Rys. 2. Nieortogonalna klasyfikacja hierarchiczna

ków A i B), w których występują obserwacje. W prostokątach niezakreskowanych nie ma obserwacji.

Z rysunku 2 widzimy, że dwa poziomy czynnika B , tj. B_1 i B_2 , występują z pierwszym poziomem czynnika A , inne dwa poziomy czynnika B ,

a mianowicie B_3 i B_4 , z drugim poziomem czynnika A i wreszcie pozostałych pięć poziomów czynnika B , tj. B_5, B_6, B_7, B_8 i B_9 , występuje z trzecim poziomem czynnika A .

Taką klasyfikację hierarchiczną nazywamy (niezrównoważoną) nieortogonalną; wyjaśnia to następująca definicja.

DEFINICJA 4. Jeżeli (1) te same liczby poziomów czynnika B występują hierarchicznie wewnątrz czynnika A i jeżeli (2) w każdej pod-

		B								
		1	2	3	4	5	6	7	8	9
A	1	■	■	■						
	2				■	■	■			
	3							■	■	■

ZM-469

Rys. 3. Ortogonalna klasyfikacja hierarchiczna $B(A)$
z $r = 5$ obserwacjami w każdej z 9 podklas

klasie znajduje się taka sama liczba obserwacji, to mamy *ortogonalną* (zrównoważoną) klasyfikację hierarchiczną. Gdy jeden z tych dwu warunków jest naruszony, klasyfikacja jest *nieortogonalna*.

Zauważmy, że warunkiem koniecznym na to, by klasyfikacja hierarchiczna $C(A)$ była ortogonalna jest to, aby liczba poziomów czynnika C była krotnością liczby poziomów czynnika A .

		B								
		1	2	3	4	5	6	7	8	9
A	1	■	■	■						
	2	■	■							
	3				■	■	■	■	■	■

ZM-470

Rys. 4. Niekompletna klasyfikacja krzyżowa

DEFINICJA 5. Jeżeli czynnik B nie jest uplasowany hierarchicznie wewnątrz czynnika A i niektóre podklasy są puste, mówimy o *niekompletnej klasyfikacji krzyżowej*. Ilustruje to rysunek 4.

DEFINICJA 6. Jeżeli te same liczby poziomów czynnika C występują hierarchicznie wewnątrz czynnika B , który tworzy ortogonalną

klasyfikację hierarchiczną z czynnikiem A postaci $B(A)$, to czynniki A , B i C tworzą podwójną klasyfikację hierarchiczną $C(B(A))$; symbol ten czytamy „ C w B w A ”. Gdy ponadto w każdej podklasie występuje taka sama liczba obserwacji, mówimy o *zrównoważonej podwójnej klasyfikacji hierarchicznej*. Uogólnienie na dowolną wielokrotną klasyfikację hierarchiczną jest widoczne.

DEFINICJA 7. Jeżeli każdą kombinację poziomów takich czynników A i C , że C w A tworzy pojedynczą klasyfikację hierarchiczną, zestawimy z każdym poziomem czynnika B , to otrzymamy *kombinację klasyfikacji krzyżowej z hierarchiczną*; przypisujemy jej symbol $C(A) \times B$ i czytamy: „ C w A razy B ”.

Nawiasy okrągłe wskazują klasyfikację hierarchiczną; znak mnożenia „ $\times$ ” — klasyfikację krzyżową.

W podobny sposób można tworzyć bardziej złożone kombinacje obu klasyfikacji dla dowolnej liczby czynników. Za przykład może posłużyć choćby kombinacja opisana symbolem $D(C(A)) \times B$, o której jest mowa w paragrafie następnym.

Uwaga. Dla zaznaczenia, że w podklasach występuje więcej obserwacji niż jedna (powtórzenia) wprowadzamy znak E , który dopisywać będziemy z lewej strony symbolu rozważanej kombinacji. Tak więc dane liczbowe opisane kombinacjami czynników A , B , C i D w formie $D(C(A)) \times B$ odnotowujemy w postaci $E[D(C(A)) \times B]$ i czytamy: „ E w podklasach kombinacji D w C w A razy B ”.

3. Nieortogonalny losowy model typu $E[D(C(A)) \times B]$. Metoda RO. Przedstawimy metodę oceny komponentów wariancyjnych na przykładzie nieortogonalnego losowego modelu typu $E[D(C(A)) \times B]$. Ostatni symbol czytamy: „ E w podklasach kombinacji D w C w A razy B ”; oznacza on, że indywidualne obserwacje (E) występują w podklasach kombinacji poziomów czynnika A , C i D skrzyżowanych z B , przy czym poziomy czynnika D występują hierarchicznie w poziomach czynnika C a te z kolei występują hierarchicznie w poziomach czynnika A . Ilustruje to schemat przedstawiony w tablicy 1.

Występujące w tej tablicy wielkości n i Y (z odpowiednimi indeksami) zdefiniowane są na str. 446 i 447, pod tablicą 2. W tablicy 1 numerujemy poziomy czynnika C niezależnie w zakresie każdego poziomu nadrzędnego czynnika A . W ten sposób poziom C_1 w A_1 jest czymś innym niż poziom C_1 w A_2 . Zgodnie z rysunkiem 3 należałoby tu mówić o pięciu poziomach czynnika C : C_1, C_2, C_3, C_4, C_5 , którym odpowiadają C_1 w A_1 , C_2 w A_1 , C_3 w A_1 , C_1 w A_2 , C_2 w A_2 . Dla wygody dalszych rozważań w zakresie każdego poziomu czynnika nadrzędnego numerujemy poziomy podrzędny czynnika zaczynając od jedynki. Te same uwagi odnoszą się do czynnika D , który jest podrzędnym względem czynnika C .

TABLICA 1
Schemat danych liczbowych nieortogonalnego losowego modelu typu (1)

A	C w A	D w C	B ₁	B ₂	B ₃
1	2	3	4	5	6
A ₁ Y _{1...} n _{1...}	C ₁ Y _{1.1.} n _{1.1.}	D ₁ Y _{1.11.} n _{1.11}	$\left. \begin{matrix} y_{11111} \\ y_{11112} \end{matrix} \right\} \begin{matrix} Y_{1111.} \\ n_{1111} \end{matrix}$	$\left. \begin{matrix} y_{12111} \\ y_{12112} \\ y_{12113} \end{matrix} \right\} \begin{matrix} Y_{1211.} \\ n_{1211} \end{matrix}$	$\left. \begin{matrix} y_{13111} \\ y_{13112} \end{matrix} \right\} \begin{matrix} Y_{1311.} \\ n_{1311} \end{matrix}$
		D ₂ Y _{1.12.} n _{1.12}	$\left. \begin{matrix} y_{11121} \\ y_{11122} \\ y_{11123} \end{matrix} \right\} \begin{matrix} Y_{1112.} \\ n_{1112} \end{matrix}$	$\left. \begin{matrix} y_{12121} \\ y_{12122} \\ y_{12123} \end{matrix} \right\} \begin{matrix} Y_{1212.} \\ n_{1212} \end{matrix}$	$\left. \begin{matrix} y_{13121} \\ y_{13122} \end{matrix} \right\} \begin{matrix} Y_{1312.} \\ n_{1312} \end{matrix}$
		D ₁ Y _{1.21.} n _{1.21}	$\left. \begin{matrix} y_{11211} \\ y_{11212} \\ y_{11213} \end{matrix} \right\} \begin{matrix} Y_{1121.} \\ n_{1121} \end{matrix}$	$\left. \begin{matrix} y_{12211} \\ y_{12212} \\ y_{12213} \end{matrix} \right\} \begin{matrix} Y_{1221.} \\ n_{1221} \end{matrix}$	$\left. \begin{matrix} y_{13211} \\ y_{13212} \end{matrix} \right\} \begin{matrix} Y_{1321.} \\ n_{1321} \end{matrix}$
	C ₂ Y _{1.2.} n _{1.2.}	D ₂ Y _{1.22.} n _{1.22}	$\left. \begin{matrix} y_{11221} \\ y_{11222} \\ y_{11223} \end{matrix} \right\} \begin{matrix} Y_{1122.} \\ n_{1122} \end{matrix}$	$\left. \begin{matrix} y_{12221} \\ y_{12222} \\ y_{12223} \end{matrix} \right\} \begin{matrix} Y_{1222.} \\ n_{1222} \end{matrix}$	$\left. \begin{matrix} y_{13221} \\ y_{13222} \end{matrix} \right\} \begin{matrix} Y_{1322.} \\ n_{1322} \end{matrix}$
		D ₃ Y _{1.23.} n _{1.23}	$\left. \begin{matrix} y_{11231} \\ y_{11232} \\ y_{11233} \end{matrix} \right\} \begin{matrix} Y_{1123.} \\ n_{1123} \end{matrix}$	$\left. \begin{matrix} y_{12231} \\ y_{12232} \\ y_{12233} \end{matrix} \right\} \begin{matrix} Y_{1223.} \\ n_{1223} \end{matrix}$	$\left. \begin{matrix} y_{13231} \\ y_{13232} \end{matrix} \right\} \begin{matrix} Y_{1323.} \\ n_{1323} \end{matrix}$
		D ₁ Y _{1.31.} n _{1.31}	$\left. \begin{matrix} y_{11311} \\ y_{11312} \\ y_{11313} \end{matrix} \right\} \begin{matrix} Y_{1131.} \\ n_{1131} \end{matrix}$	$\left. \begin{matrix} y_{12311} \\ y_{12312} \\ y_{12313} \end{matrix} \right\} \begin{matrix} Y_{1231.} \\ n_{1231} \end{matrix}$	$\left. \begin{matrix} y_{13311} \\ y_{13312} \end{matrix} \right\} \begin{matrix} Y_{1331.} \\ n_{1331} \end{matrix}$
	C ₃ Y _{1.3.} n _{1.3.}	D ₂ Y _{1.32.} n _{1.32}	$\left. \begin{matrix} y_{11321} \\ y_{11322} \\ y_{11323} \end{matrix} \right\} \begin{matrix} Y_{1132.} \\ n_{1132} \end{matrix}$	$\left. \begin{matrix} y_{12321} \\ y_{12322} \\ y_{12323} \end{matrix} \right\} \begin{matrix} Y_{1232.} \\ n_{1232} \end{matrix}$	$\left. \begin{matrix} y_{13321} \\ y_{13322} \end{matrix} \right\} \begin{matrix} Y_{1332.} \\ n_{1332} \end{matrix}$
		D ₃ Y _{1.33.} n _{1.33}	$\left. \begin{matrix} y_{11331} \\ y_{11332} \\ y_{11333} \end{matrix} \right\} \begin{matrix} Y_{1133.} \\ n_{1133} \end{matrix}$	$\left. \begin{matrix} y_{12331} \\ y_{12332} \\ y_{12333} \end{matrix} \right\} \begin{matrix} Y_{1233.} \\ n_{1233} \end{matrix}$	$\left. \begin{matrix} y_{13331} \\ y_{13332} \end{matrix} \right\} \begin{matrix} Y_{1333.} \\ n_{1333} \end{matrix}$
		D ₁ Y _{1.41.} n _{1.41}	$\left. \begin{matrix} y_{11411} \\ y_{11412} \\ y_{11413} \end{matrix} \right\} \begin{matrix} Y_{1141.} \\ n_{1141} \end{matrix}$	$\left. \begin{matrix} y_{12411} \\ y_{12412} \\ y_{12413} \end{matrix} \right\} \begin{matrix} Y_{1241.} \\ n_{1241} \end{matrix}$	$\left. \begin{matrix} y_{13411} \\ y_{13412} \end{matrix} \right\} \begin{matrix} Y_{1341.} \\ n_{1341} \end{matrix}$
	C ₄ Y _{1.4.} n _{1.4.}	D ₂ Y _{1.42.} n _{1.42}	$\left. \begin{matrix} y_{11421} \\ y_{11422} \\ y_{11423} \end{matrix} \right\} \begin{matrix} Y_{1142.} \\ n_{1142} \end{matrix}$	$\left. \begin{matrix} y_{12421} \\ y_{12422} \\ y_{12423} \end{matrix} \right\} \begin{matrix} Y_{1242.} \\ n_{1242} \end{matrix}$	$\left. \begin{matrix} y_{13421} \\ y_{13422} \end{matrix} \right\} \begin{matrix} Y_{1342.} \\ n_{1342} \end{matrix}$
		D ₃ Y _{1.43.} n _{1.43}	$\left. \begin{matrix} y_{11431} \\ y_{11432} \\ y_{11433} \end{matrix} \right\} \begin{matrix} Y_{1143.} \\ n_{1143} \end{matrix}$	$\left. \begin{matrix} y_{12431} \\ y_{12432} \\ y_{12433} \end{matrix} \right\} \begin{matrix} Y_{1243.} \\ n_{1243} \end{matrix}$	$\left. \begin{matrix} y_{13431} \\ y_{13432} \end{matrix} \right\} \begin{matrix} Y_{1343.} \\ n_{1343} \end{matrix}$
		D ₁ Y _{1.51.} n _{1.51}	$\left. \begin{matrix} y_{11511} \\ y_{11512} \\ y_{11513} \end{matrix} \right\} \begin{matrix} Y_{1151.} \\ n_{1151} \end{matrix}$	$\left. \begin{matrix} y_{12511} \\ y_{12512} \\ y_{12513} \end{matrix} \right\} \begin{matrix} Y_{1251.} \\ n_{1251} \end{matrix}$	$\left. \begin{matrix} y_{13511} \\ y_{13512} \end{matrix} \right\} \begin{matrix} Y_{1351.} \\ n_{1351} \end{matrix}$
Suma			Y _{11...} n _{11.}	Y _{12...} n _{12.}	Y _{13...} n _{13.}

A_2 $Y_{2...}$ $n_{2...}$	C_1 $Y_{2.1..}$ $n_{2.1.}$	D_1 $Y_{2.11.}$ $n_{2.11}$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$
	C_2 $Y_{2.2..}$ $n_{2.2.}$	D_2 $Y_{2.21.}$ $n_{2.21}$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$
		D_3 $Y_{2.23.}$ $n_{2.23}$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$
Suma			$Y_{21...}$ $Y_{22...}$ $Y_{23...}$	$n_{21..}$ $n_{22..}$ $n_{23..}$	$Y_{2321.}$ $Y_{2322.}$ $Y_{2323.}$	$n_{2321.}$ $n_{2322.}$ $n_{2323.}$
Suma sum			$Y_{21...}$ $Y_{22...}$ $Y_{23...}$	$n_{21..}$ $n_{22..}$ $n_{23..}$	$Y_{2321.}$ $Y_{2322.}$ $Y_{2323.}$	$n_{2321.}$ $n_{2322.}$ $n_{2323.}$

Schemat ten reprezentuje kombinację podwójnej klasyfikacji krzyżowej $A \times B$ z potrójną klasyfikacją hierarchiczną $E(D(C(A)))$. Pomińcie bowiem czynnika B daje potrójną klasyfikację hierarchiczną (por. [16]) z następującymi źródłami zmienności: 1° A , 2° C w A , tj. $C(A)$, 3° D w C w A , czyli $D(C(A))$ i 4° między indywidualnymi wynikami E , tj. E w D w C w A czyli $E(D(C(A)))$.

Dane te są nieortogonalne, ponieważ w podklasach występują różne liczby obserwacji.

Matematyczny model dla rozpatrywanego schematu ma postać

$$(1) \quad y_{ijkmg} = \mu + a_i + b_j + (ab)_{ij} + c(a)_{k(i)} + bc(a)_{jk(i)} + \\ + d(c(a))_{m(k(i))} + bd(c(a))_{jm(k(i))} + e_{ijkmg},$$

gdzie wskaźniki i, j, k, m, g odpowiadają czynnikom $A, B, C(A), D(C(A))$ i powtórzeniom $E[D(C(A)) \times B]$ przyjmując następujące zakresy:

$$\begin{array}{ll} A & i = 1, \dots, I, \\ B & j = 1, \dots, J, \\ C(A) & k = 1, \dots, c_i, \\ D(C(A)) & m = 1, \dots, d_{ik}, \\ E[D(C(A)) \times B] & g = 1, \dots, n_{ijkm}. \end{array}$$

Symbol a_i oznacza efekt i -tej klasy klasyfikacji A , b_j — efekt j -tej klasy klasyfikacji B , $(ab)_{ij}$ — efekt interakcyjny podklasy (ij) klasyfikacji AB , $c(a)_{k(i)}$ — efekt k -tej klasy klasyfikacji C wewnątrz i -tej klasy klasyfikacji A , $bc(a)_{jk(i)}$ — efekt interakcyjny podklasy (jk) klasyfikacji BC w i -tej klasie klasyfikacji A , itd. (por. [18]).

W modelu (1) wszystkie efekty poza μ są losowe i nieskorelowane. Losowość efektów jest uwarunkowana losowym wyborem klas klasyfikacji A , klas klasyfikacji B , klas C wewnątrz A i klas D wewnątrz C .

Nadto zakładamy, że ich średnie w populacji są równe zeru tzn., że

$$\mathcal{E}(a_i) = \mathcal{E}(b_j) = \mathcal{E}(ab)_{ij} = \dots = \mathcal{E}e_{ijkmg} = 0,$$

gdzie symbol $\mathcal{E}$ oznacza wartość oczekiwaną. Wariancje efektów losowych $a_i, b_j, \dots, e_{ijkmg}$ definiujemy jako

$$\begin{aligned} \mathcal{E}a_i^2 &= \sigma_A^2, & \mathcal{E}b_j^2 &= \sigma_B^2, & \mathcal{E}(ab)_{ij}^2 &= \sigma_{AB}^2, & \mathcal{E}[c(a)_{k(i)}]^2 &= \sigma_{C(A)}^2, \\ \mathcal{E}[bc(a)_{jk(i)}]^2 &= \sigma_{BC(A)}^2, & \mathcal{E}[d(c(a))_{m(k(i))}]^2 &= \sigma_{D(C(A))}^2, \\ \mathcal{E}[bd(c(a))_{jm(k(i))}]^2 &= \sigma_{BD(C(A))}^2, & \mathcal{E}e_{ijkmg}^2 &= \sigma_e^2. \end{aligned}$$

Dla wszystkich ocen komponentów wariancyjnych $\sigma_A^2, \sigma_B^2, \sigma_{AB}^2, \sigma_{C(A)}^2, \sigma_{BC(A)}^2, \sigma_{D(C(A))}^2, \sigma_{BD(C(A))}^2$ i σ_e^2 przedstawiamy następującą metodę,

nazwijmy ją RO; jest ona uogólnieniem metody R (por. § 1) na przypadek dowolnej kombinacji klasyfikacji krzyżowej z hierarchiczną ([17]).

Metoda RO obejmuje następujące 4 punkty: I, II, III i IV, w tym regułę RO(III), opisaną dalej w 1°, 2° i 3° (por. str. 448):

I. Ustalenie modelu statystycznego dla danych liczbowych i przyjęcie odpowiedniej symboliki (por. §§ 2 i 3).

II. Wyznaczenie p składników sum kwadratów z odpowiadającymi im stopniami swobody na podstawie tablicy analizy wariancji (w oparciu o pierwszą metodę Hendersona; por. tablicę 2) lub wprost na podstawie zespołów wskaźników stojących przy składnikach modelu.

III. Reguła RO, czyli metoda wyznaczania współczynników W przy komponentach wariancyjnych (por. [17]).

IV. Rozwiązanie układu p równań ze względu na p niewiadomych: kwadrat średniej w populacji, μ^2 , i $p-1$ komponentów wariancyjnych: $\sigma_e^2, \sigma_A^2, \dots$

Metoda RO opiera się na tych samych punktach co metoda R, z tym, że uzupełnia ją nową symboliką a to z uwagi na jej stosowalność wobec dowolnych modeli podlegających kombinacji klasyfikacji krzyżowej z hierarchiczną. Nadto wobec uogólnienia symboliki modyfikuje regułę wyznaczania współczynników W . Dzięki wprowadzonej symbolice ustala się bez trudu odpowiedniość wzajemnie jednojednoznaczna między wskaźnikami i efektami modelu oraz wariancjami tych efektów bez względu na złożony charakter danych liczbowych.

Metodę RO zilustrujemy korzystając z modelu (1). Najpierw postępujemy podobnie jak przy analizie wariancji modelu ortogonalnego. Wypisujemy źródła zmienności, stopnie swobody dla modelu ortogonalnego i ważone sumy kwadratów odchyłeń (por. [13]) zgodnie z pierwszą metodą Hendersona (por. [10]). Odpowiednie pozycje wraz z kolumną 2 zatytułowaną „Stopnie swobody dla modelu nieortogonalnego” zestawiono w tablicy 2.

Dla modelu ortogonalnego mamy: $c_i = K = \text{const}$, $d_{ik} = D = \text{const}$, $n_{ijkm} = G = \text{const}$, wtedy $p = IK$ i $r = IKD$, $n = IJKDG$. Liczba podklas wynosi $IJKD$.

Zauważmy, że w kolumnie 2 dla stopni swobody modelu nieortogonalnego $\sum_i c_i = p$ oznacza łączną liczbę poziomów klasyfikacji C we wszystkich klasach A , tj. w $A_1, A_2, \dots, A_I$, a $\sum_i \sum_k^I c_i d_{ik} = r$ oznacza łączną liczbę poziomów klasyfikacji D we wszystkich podklasach AC , tj. w $A_1 C_1, A_1 C_2, \dots, A_1 C_{c_1}, \dots, A_I C_{c_I}$.

Z ważonych sum kwadratów odchyłeń (kol. 4 tablicy 2) wybieramy wszystkie możliwe różne składniki. Liczba ich jest dokładnie równa

TABLICA 2

Analiza wariancji dla oceny komponentów

Źródło zmienności	Stopnie swobody dla modelu nieortogonalnego	Stopnie swobody dla modelu ortogonalnego ($c_i = K$, $d_{ik} = D$, $n_{ijkm} = G$)
1	2	3
1. A	$I - 1$	$I - 1$
2. B	$J - 1$	$J - 1$
3. $A \times B$	$(I - 1)(J - 1)$	$(I - 1)(J - 1) = IJ - I - J + 1$
4. C w A lub $C(A)$	$\sum_i^I (c_i - 1) = p - I$	$I(K - 1) = IK - I$
5. $B \times C$ w A lub $BC(A)$	$\sum_i^I (c_i - 1)(J - 1) =$ $= (J - 1)(p - I)$	$I(J - 1)(K - 1) =$ $= IJK - IJ - IK + I$
6. D w C w A lub $D(C(A))$	$\sum_i^I \sum_k^{c_i} (d_{ik} - 1) = r - p$	$IK(D - 1) = IKD - IK$
7. $B \times D$ w C w A lub $BD(C(A))$	$\sum_i^I \sum_k^{c_i} (J - 1)(d_{ik} - 1) =$ $= (J - 1)(r - p)$	$IK(J - 1)(D - 1) =$ $= IJKD - IJK - IKD + IK$
8. Indywidualne w $ABCD$	$\sum_i \sum_j \sum_k \sum_m (n_{ijkm} - 1) = n - Jr$	$IJKD(G - 1) = IJKDG - IJKD$

Symbol n_{ijkm} oznacza liczbę obserwacji w podklasie $(ijkm)$, czyli w podklasie $ABCD$. Definicje symboli n użytych dla oznaczenia liczebności poszczególnych podklas są następujące:

$$n_{ijk.} = \sum_m n_{ijkm}$$

$$n_{j..} = \sum_i n_{ij..}$$

$$n_{i..} = \sum_j n_{ij..}$$

$$n_{i...} = \sum_j n_{ij..}$$

$$n_{i.k.} = \sum_m n_{i.k.m}$$

$$n = \sum_i n_{i...} \quad - \text{liczba wszystkich wyników}$$

$$n_{ij..} = \sum_k n_{ijk.}$$

wariancyjnych nieortogonalnego modelu losowego (1)

Ważone sumy kwadratów odchyień

4

$$\sum_i \frac{Y_{i....}^2}{n_{i...}} - \frac{Y^{2....}}{n} = [2] - [1]$$

$$\sum_j \frac{Y_{.j...}^2}{n_{j..}} - \frac{Y^{2....}}{n} = [3] - [1]$$

$$\sum_i \sum_j \frac{Y_{ij...}^2}{n_{ij..}} - \sum_i \frac{Y_{i....}^2}{n_{i...}} - \sum_j \frac{Y_{.j...}^2}{n_{j..}} + \frac{Y^{2....}}{n} = [4] - [2] - [3] + [1]$$

$$\sum_i \sum_k \frac{Y_{i.k..}^2}{n_{i.k.}} - \sum_i \frac{Y_{i....}^2}{n_{i...}} = [5] - [2]$$

$$\sum_i \sum_j \sum_k \frac{Y_{ijk..}^2}{n_{ijk.}} - \sum_i \sum_j \frac{Y_{ij...}^2}{n_{ij..}} - \sum_i \sum_k \frac{Y_{i.k..}^2}{n_{i.k.}} + \sum_i \frac{Y_{i....}^2}{n_{i...}} = [7] - [4] - [5] + [2]$$

$$\sum_i \sum_k \sum_m \frac{Y_{i.km.}^2}{n_{i.km}} - \sum_i \sum_k \frac{Y_{i.k..}^2}{n_{i.k.}} = [6] - [5]$$

$$\begin{aligned} \sum_i \sum_j \sum_k \sum_m \frac{Y_{ijkm.}^2}{n_{ijkm}} - \sum_i \sum_j \sum_k \frac{Y_{ijk..}^2}{n_{ijk.}} - \sum_i \sum_k \sum_m \frac{Y_{i.km.}^2}{n_{i.km}} + \sum_i \sum_k \frac{Y_{i.k..}^2}{n_{i.k.}} = \\ = [8] - [7] - [6] + [5] \end{aligned}$$

$$\sum_i \sum_j \sum_k \sum_m \sum_g \frac{Y_{ijkmg}^2}{n_{ijkmg}} - \sum_i \sum_j \sum_k \sum_m \frac{Y_{ijkm.}^2}{n_{ijkm}} = [9] - [8]$$

Definicje sum wyników poszczególnych podklas są następujące:

$$Y_{ijkm.} = \sum_g y_{ijkmg}$$

$$Y_{ij...} = \sum_k Y_{ijk..}$$

$$Y_{ijk..} = \sum_m Y_{ijkm.}$$

$$Y_{i....} = \sum_j Y_{ij...}$$

$$Y_{i.km.} = \sum_j Y_{ijkm.}$$

$$Y_{.j...} = \sum_i Y_{ij...}$$

$$Y_{i.k..} = \sum_m Y_{i.km.}$$

$$Y^{2....} = \sum_i Y_{i....}$$

liczbie nieznanych komponentów wariancyjnych; w naszym przykładzie zestawiono je w kolumnie 2 tablicy 3; jest ich dziewięć. Ponieważ tablica 3 stanowi podstawę do wyznaczania ocen komponentów wariancyjnych, więc opiszemy metodę jej konstrukcji.

W kolumnie 3 wypisujemy wskaźniki, które wystąpiły pod sumami składników kolumny 2, tj. $ijkmg, ijkm, \dots, i$; dziewiętemu składnikowi nie zawierającemu znaku sumy przypisujemy symbol 1. W kolumnie 4 wypisujemy stopnie swobody odpowiadające kolejnym dziewięciu składnikom; $n, Jr, \dots, I, 1$; obliczamy je sumując jedynek względem zaznaczonych wskaźników. W nagłówkach kolumn 5, 6, 7, 8, 9, 10, 11, 12 i 13 wypisujemy μ^2 i osiem wariancji, które mamy oszacować: $\sigma_A^2, \sigma_B^2, \dots, \sigma_e^2$. Nad tymi symbolami odnotowujemy wskaźniki występujące przy odpowiednich efektach modelu (1): $i, j, ij, k(i), jk(i), m(k(i)), jm(k(i)), ijkmg$, pisząc przy tym symbol 0 nad μ^2 . Zauważmy odpowiedniość między wskaźnikami przy efektach modelu i symbolami przy wariancjach.

W kolumnach oznaczonych numerami 5, 6, ..., 13 znajdują się liczby, które nazywamy ogólnie współczynnikami W i które występują przy komponentach wariancyjnych wartości oczekiwanych dziewięciu składników.

Metoda wyznaczania współczynników W przy μ^2 i wariancjach $\sigma_A^2, \sigma_B^2, \dots, \sigma_e^2$ opiera się na regule RO, którą wyrażamy w następujących punktach (por. twierdzenie w [17]):

1° Przy μ^2 występuje zawsze współczynnik n a przy σ_e^2 — liczba stopni swobody, odpowiadająca danemu składnikowi sumy kwadratów.

2° Przy wariancji opatrzonej znaczkami $A, B, AB, \dots$ wypisujemy współczynnik W z pierwszą grupą wskaźników takich samych jak pod znakami Σ w składniku (w odpowiednim wierszu) i z drugą grupą (pomiędzy nawiasy okrągłe) w prostokątnych nawiasach tych wskaźników, które odpowiadają symbolom wariancji a zostały umieszczone w kolumnach nad wariancjami. W grupie wskaźników w nawiasach prostokątnych pomijamy te, które wystąpiły przed nawiasami.

3° Jeżeli wszystkie wskaźniki w nawiasach zostaną pominięte, to na ich miejsce piszemy zero a zamiast współczynnika W liczbę wszystkich obserwacji czyli n ; tak więc

$$W_{j[i]} = W_{i[i]} = W_{ik[i]} = \dots = n.$$

Przyjmujemy również umowę: $W_{i[i]} = W_i$; $W_{i[j]} = W_j, \dots, W_{i[ijkm]} = W_{ijkm}$.

Przedstawioną metodę wyznaczania współczynników zilustrujemy na kilku przykładach. W tym celu obserwujemy tablicę 3.

TABLICA 3
Wartości oczekiwane składników sum ważonych kwadratów odchylen dla modelu (1)

Lp.	Składniki sum ważonych kwadratów odchyżeń S	Wskaźniki przy sumach	Stopnie swobody ν	Współczynniki przy komponentach wariancyjnych dla wartości oczekiwanych składników sum ważonych kwadratów odchyżeń E(S)									
				0	i	j	ij	k(i)	jk(i)	m(k(i))	jm(k(i))	ijkmg	
				μ^2	σ_A^2	σ_B^2	σ_{AB}^2	$\sigma_C(A)$	$\sigma_{BC(A)}^2$	$\sigma_D(C(A))$	$\sigma_{BD(C(A))}^2$	σ_e^2	
1		2	4	5	6	7	8	9	10	11	12	13	
1	$\sum_i \sum_j \sum_k \sum_m \sum_g \frac{Y_{ijkmg}^2}{n}$	$ijkmg$	$\sum_i \sum_j \sum_k \sum_m \sum_g 1 = n$	n	n	n	n	n	n	n	n	n	
2	$\sum_i \sum_j \sum_k \sum_m \frac{Y_{ijkm}^2}{n_{ijkm}}$	$ijkm$	$\sum_i \sum_j \sum_k \sum_m 1 = Jr$	n	n	n	n	n	n	n	n	Jr	
3	$\sum_i \sum_k \sum_m \frac{Y_{ikm}^2}{n_{ikm}}$	ikm	$\sum_i \sum_k \sum_m 1 = r$	n	n	$W_{ikm[j]}$	$W_{ikm[j]}$	n	$W_{ikm[j]}$	n	$W_{ikm[j]}$	r	
4	$\sum_i \sum_j \sum_k \frac{Y_{ijk}^2}{n_{ijk}}$	ijk	$\sum_i \sum_j \sum_k 1 = Jp$	n	n	n	n	n	n	$W_{ijk[m]}$	$W_{ijk[m]}$	Jp	
5	$\sum_i \sum_k \frac{Y_{ik}^2}{n_{ik}}$	ik	$\sum_i \sum_k 1 = p$	n	n	$W_{ik[j]}$	$W_{ik[j]}$	n	$W_{ik[j]}$	$W_{ik[m]}$	$W_{ik[m]}$	p	
6	$\sum_i \sum_j \frac{Y_{ij}^2}{n_{ij}}$	ij	$\sum_i \sum_j 1 = IJ$	n	n	n	n	$W_{ij[k]}$	$W_{ij[k]}$	$W_{ij[km]}$	$W_{ij[km]}$	IJ	
7	$\sum_j \frac{Y_j^2}{n_j}$	j	$\sum_j 1 = J$	n	$W_{j[i]}$	n	$W_{j[i]}$	$W_{j[ik]}$	$W_{j[ik]}$	$W_{j[km]}$	$W_{j[km]}$	J	
8	$\sum_i \frac{Y_i^2}{n_i}$	i	$\sum_i 1 = I$	n	n	$W_{i[j]}$	$W_{i[j]}$	$W_{i[k]}$	$W_{i[k]}$	$W_{i[km]}$	$W_{i[km]}$	I	
9	$\frac{Y^2}{n}$	1	1	n	W_i	W_j	W_{ij}	W_{ik}	W_{ijk}	W_{ikm}	W_{ijkm}	1	

Wzory na współczynniki W zestawione w tablicy 3 łatwo zapamiętać z uwagi na prostą prawidłowość, którą dostrzegamy między dwiema grupami wskaźników przy W i dwiema grupami sum: 1) przed ułamkiem i 2) w liczniku ułamka (por. [17]):

1. $W_{ijk[m]} = \sum_i \sum_j \sum_k \frac{\sum_m^2 n_{ijk}^2}{n_{ijk}}$
2. $W_{ikm[j]} = \sum_i \sum_k \sum_m \frac{\sum_j^2 n_{ijk}^2}{n_{i,km}}$
3. $W_{ik[j]} = \sum_i \sum_k \frac{\sum_j^2 n_{ijk}^2}{n_{i,k}}$
4. $W_{ik[m]} = \sum_i \sum_k \frac{\sum_m^2 n_{i,km}^2}{n_{i,k}}$
5. $W_{ik[mj]} = \sum_i \sum_k \frac{\sum_j^2 \sum_m n_{ijk}^2}{n_{i,k}}$
6. $W_{ij[k]} = \sum_i \sum_j \frac{\sum_k^2 n_{ijk}^2}{n_{ij..}}$
7. $W_{ij[km]} = \sum_i \sum_j \frac{\sum_k^2 \sum_m n_{ijk}^2}{n_{ij..}}$
8. $W_{j[i]} = \sum_j \frac{\sum_i^2 n_{ij..}^2}{n_{j..}}$
9. $W_{j[ik]} = \sum_j \frac{\sum_i^2 \sum_k n_{ijk}^2}{n_{j..}}$
10. $W_{j[ikm]} = \sum_j \frac{\sum_i^2 \sum_k \sum_m n_{ijk}^2}{n_{j..}}$
11. $W_{i[j]} = \sum_i \frac{\sum_j^2 n_{ij..}^2}{n_{i..}}$
12. $W_{i[k]} = \sum_i \frac{\sum_k^2 n_{i,k}^2}{n_{i..}}$
13. $W_{i[jk]} = \sum_i \frac{\sum_j^2 \sum_k n_{ijk}^2}{n_{i..}}$
14. $W_{i[jkm]} = \sum_i \frac{\sum_j^2 \sum_k \sum_m n_{ijk}^2}{n_{i..}}$
15. $W_i = \frac{\sum_i^2 n_{i..}^2}{n}$
16. $W_j = \frac{\sum_j^2 n_{j..}^2}{n}$
17. $W_{ij} = \frac{\sum_i^2 \sum_j^2 n_{ij..}^2}{n}$
18. $W_{ik} = \frac{\sum_i^2 \sum_k^2 n_{i,k}^2}{n}$
19. $W_{ijk} = \frac{\sum_i^2 \sum_j^2 \sum_k^2 n_{ijk}^2}{n}$
20. $W_{ikm} = \frac{\sum_i^2 \sum_k^2 \sum_m^2 n_{i,km}^2}{n}$
21. $W_{ijkm} = \frac{\sum_i^2 \sum_j^2 \sum_k^2 \sum_m^2 n_{ijkm}^2}{n}$

PRZYKŁAD 1. Współczynnik w drugim wierszu przy $\sigma_{BD(C(A))}^2$ ma postać $W_{ijk[m(k(i))]}$, a po opuszczeniu okrągłych nawiasów $W_{ijk[mki]}$. Zgodnie z punktem 2° i 3° otrzymujemy następnie $W_{ijk[m0]} = n$.

W początkowej postaci współczynnika przed nawiasem kwadratowym wypisaliśmy grupę wskaźników $ijkm$, gdyż są to wskaźniki odpowiadające składnikowi wiersza drugiego $\sum_i \sum_j \sum_k \sum_m \frac{Y_{ijkm}^2}{n_{ijkm}}$; w nawiasie wypisaliśmy $jm(k(i))$, gdyż jest to zespół wskaźników stojących nad $\sigma_{BD(C(A))}^2$.

PRZYKŁAD 2. Współczynnik stojący w wierszu składnika $\sum_i \sum_j \sum_k \frac{Y_{ijk..}^2}{n_{ijk..}}$ przy wariancji $\sigma_{D(C(A))}^2$ jest postaci $W_{ijk[m(k(i))]} = W_{ijk[mki]} = W_{ijk[m]}$, gdyż w składniku występują wskaźniki ijk a nad wariancją $\sigma_{D(C(A))}^2$ wskaźniki $m(k(i))$. Pominęliśmy w nawiasie prostokątnym wskaźniki ik , gdyż występowały już one przed tym nawiasem.

PRZYKŁAD 3. Współczynnik stojący na wysokości składnika $Y^2 \dots / n$ przy wariancji $\sigma_{BC(A)}^2$ ma kształt $W_{1[jk(i)]} = W_{1[jki]} = W_{ijk}$.

Obliczywszy współczynniki W zgodnie z podanymi wzorami, możemy, korzystając z tablicy 3, wypisać 9 równań z 9 niewiadomymi: μ^2 , σ_A^2 , σ_B^2 , σ_{AB}^2 , $\sigma_{C(A)}^2$, $\sigma_{BC(A)}^2$, $\sigma_{D(C(A))}^2$, $\sigma_{BD(C(A))}^2$ i σ_e^2 .

Po lewej stronie równań występują składniki sum ważonych kwadratów odchyłeń a po prawej wymienione niewiadome z zaznaczonymi współczynnikami. Tak np. równanie 5 jest formy

$$\sum_i \sum_k \frac{Y_{i.k..}^2}{n_{i.k..}} = n\mu^2 + n\sigma_A^2 + W_{ik[j]} \sigma_B^2 + W_{ik[j]} \sigma_{AB}^2 + \\ + n\sigma_{C(A)}^2 + W_{ik[j]} \sigma_{BC(A)}^2 + W_{ik[m]} \sigma_{D(C(A))}^2 + W_{ik[m]} \sigma_{BD(C(A))}^2 + p\sigma_e^2.$$

Rozwiązanie układu daje oceny ośmiu komponentów wariancyjnych i ocenę kwadratu średniej, μ^2 .

Uogólnienie przedstawionej metody estymacji komponentów wariancyjnych na dowolną kombinację klasyfikacji krzyżowej z hierarchiczną jest widoczne. Liczba równań w układzie jest równa liczbie składników po prawej stronie modelu matematycznego o postaci analogiczne do postaci modelu (1).

4. Przykład liczbowy modelu $E[C(A \times B)]$. Dla ilustracji stosowalności metody RO w celu wyznaczenia ocen komponentów wariancyjnych rozważmy fragment danych liczbowych uzyskanych z Katedry Ogólnej Hodowli Zwierząt Wyższej Szkoły Rolniczej w Lublinie. Odpowiednie dane zestawione w tablicy 4, obejmują nieśność $n = 2099$ kurocerek (klasyfikacja E) w teście 500-dniowym pochodzącym od $r = 43$

TABLICA 4
Nieśność kur-córek w teście 500-dniowym

Lata A	F e r m y							
	B ₁		B ₂		B ₃		B ₄	
	koguty	kury-córki	koguty	kury-córki	koguty	kury-córki	koguty	kury-córki
A ₁ Y _{1...} = 105866 n _{1..} = 683	C ₁₁₁ Y _{111.} = 4440 n ₁₁₁ = 30	y ₁₁₁₁ = 107 y ₁₁₁₂ = 169 y ₁₁₁₃ = 92	C ₁₂₁ Y _{121.} = 8353 n ₁₂₁ = 44	.	C ₁₃₁ Y _{131.} = 4693 n ₁₃₁ = 30	.	C ₁₄₁ Y _{141.} = 2977 n ₁₄₁ = 21	.
	C ₁₁₂ Y _{112.} = 9115 n ₁₁₂ = 63	.	C ₁₂₂ Y _{122.} = 5354 n ₁₂₂ = 32	.	C ₁₃₂ Y _{132.} = 3722 n ₁₃₂ = 24	.	C ₁₄₂ Y _{142.} = 1496 n ₁₄₂ = 10	.
	C ₁₁₃ Y _{113.} = 9934 n ₁₁₃ = 72	.	C ₁₂₃ Y _{123.} = 1751 n ₁₂₃ = 10	.	C ₁₃₃ Y _{133.} = 748 n ₁₃₃ = 6	.	C ₁₄₃ Y _{143.} = 1246 n ₁₄₃ = 9	.
	C ₁₁₄ Y _{114.} = 12864 n ₁₁₄ = 93	.	C ₁₂₄ Y _{124.} = 7929 n ₁₂₄ = 45	.	C ₁₃₄ Y _{134.} = 7652 n ₁₃₄ = 47	.	C ₁₄₄ Y _{144.} = 791 n ₁₄₄ = 6	.
	—	—	C ₁₂₅ Y _{125.} = 6029 n ₁₂₅ = 35	.	—	—	C ₁₄₅ Y _{145.} = 6781 n ₁₄₅ = 47	.
	—	—	C ₁₂₆ Y _{126.} = 9475 n ₁₂₆ = 55	.	—	—	C ₁₄₆ Y _{146.} = 516 n ₁₄₆ = 4	.
	Y _{11..} = 33553 n _{11.} = 258	.	Y _{12..} = 38891 n _{12.} = 221	.	Y _{13..} = 16815 n _{13.} = 107	.	Y _{14..} = 13807 n _{14.} = 97	.
Suma		.		.		.		.

A_2 $Y_{2...} = 204457$ $n_{2..} = 1416$	C_{211} $Y_{211.} = 5947$ $n_{211} = 47$	• • •	C_{221} $Y_{221.} = 6855$ $n_{221} = 39$	• • •	C_{231} $Y_{231.} = 10220$ $n_{231} = 66$	• • •	C_{241} $Y_{241.} = 13450$ $n_{241} = 94$	• • •
	C_{212} $Y_{212.} = 5102$ $n_{212} = 44$	• • •	C_{222} $Y_{222.} = 8431$ $n_{222} = 50$	• • •	C_{232} $Y_{232.} = 4647$ $n_{232} = 31$	• • •	C_{242} $Y_{242.} = 12756$ $n_{242} = 82$	• • •
	C_{213} $Y_{213.} = 5859$ $n_{213} = 53$	• • •	C_{223} $Y_{223.} = 6660$ $n_{223} = 39$	• • •	C_{233} $Y_{233.} = 7612$ $n_{233} = 54$	• • •	C_{243} $Y_{243.} = 17290$ $n_{243} = 114$	• • •
	C_{214} $Y_{214.} = 5691$ $n_{214} = 49$	• • •	C_{224} $Y_{224.} = 5729$ $n_{224} = 34$	• • •	C_{234} $Y_{234.} = 5383$ $n_{234} = 34$	• • •	C_{244} $Y_{244.} = 11687$ $n_{244} = 74$	• • •
	C_{215} $Y_{215.} = 7494$ $n_{215} = 68$	• • •	C_{225} $Y_{225.} = 10941$ $n_{225} = 61$	• • •	C_{235} $Y_{235.} = 8255$ $n_{235} = 58$	• • •	C_{245} $Y_{245.} = 13574$ $n_{245} = 95$	• • •
	C_{216} $Y_{216.} = 8001$ $n_{216} = 69$	• • •	C_{226} $Y_{226.} = 11672$ $n_{226} = 77$	• • •	—	—	C_{246} $Y_{246.} = 12101$ $n_{246} = 84$	• • •
	$Y_{21..} = 38094$ $n_{21.} = 330$	—	$Y_{22..} = 49388$ $n_{22.} = 300$	—	$Y_{23..} = 36117$ $n_{23.} = 243$	—	$Y_{24..} = 80858$ $n_{24.} = 543$	—
	$Y_{21..} = 74447$ $n_{21.} = 588$	—	$Y_{22..} = 88279$ $n_{22.} = 521$	—	$Y_{23..} = 52932$ $n_{23.} = 350$	—	$Y_{24..} = 94665$ $n_{24.} = 640$	—
	Suma							
	Suma sum							

kogutów (klasyfikacja C) z $J = 4$ ferm (klasyfikacja B) w $I = 2$ kolejnych latach: 1958/59 i w 1959/60 (klasyfikacja A). Wyniki zebrano z ferm: w Kobylance, pow. Gorlice; Jodłowniku, pow. Tylmanowa; Chorzeliowie, pow. Mielec i w Borowinie, pow. Puławy. Jest widoczne, że dane są nieortogonalne i stanowią kombinację klasyfikacji krzyżowej lata $\times$ fermy z klasyfikacją hierarchiczną kury-córki (E) w obrębie kogutów (C) wewnątrz fermo-lat ($A \times B$). Stąd układowi takiemu przypisujemy symbol $E[C(A \times B)]$ i model matematyczny postaci

$$(2) \quad y_{ikg} = \mu + a_i + b_j + (ab)_{ij} + c(ab)_{k(ij)} + e_{ikg}.$$

Wskaźniki i, j, k, g odpowiadają czynnikom A (lata), B (fermy), C (koguty) i indywidualnym obserwacjom (E), jak następuje:

A : $i = 1, \dots, I$; $I = 2$ lata;

B : $j = 1, \dots, J$; $J = 4$ fermy;

$C(AB)$: $k = 1, \dots, d_{ij}$; d_{ij} oznaczają liczbę kogutów w i -tym roku na j -tej fermie (gdyby dane były ortogonalne, mielibyśmy $d_{ij} = D = \text{const}$ dla każdego roku w każdej fermie, tzn. mielibyśmy tę samą liczbę kogutów w każdej fermie w każdym roku);

$E[C(A \times B)]$: $g = 1, \dots, n_{ijk}$; n_{ijk} oznaczają liczbę kur-córek po k -tym kogucie w i -tym roku na j -tej fermie (gdyby dane były ortogonalne to mielibyśmy $n_{ijk} = \text{const}$, a więc mielibyśmy tę samą liczbę kur-córek po każdym kogucie).

Symbol a_i określa efekt i -tego sezonu, b_j — efekt j -tej fermy, $(ab)_{ij}$ — efekt interakcyjny odpowiadający i -temu sezonowi i j -tej fermie, $c(ab)_{k(ij)}$ — efekt k -tego koguta w i -tym sezonie j -tej fermy, e_{ikg} — efekt losowy odpowiadający wynikowi y_{ikg} , który określa liczbę jaj zniesionych w teście 500-dniowym przez g -tą kure, córkę k -tego koguta z j -tej fermy w i -tym roku; tak więc $y_{1113} = 92$ (por. tablica 4) oznacza, że kura będąca trzecią córką pierwszego koguta z pierwszej fermy w pierwszym roku zniosła 92 jaja. W tablicy 4 w podklasie (111) z braku miejsca podano tylko 3 początkowe wartości zamiast 30 na liczbę zniesionych jaj y a mianowicie: 107, 169, 92; pozostałe 27 wartości były następujące: 143, 145, 190, 169, 121, 214, 147, 113, 143, 155, 160, 174, 161, 194, 176, 173, 158, 138, 137, 122, 141, 155, 127, 138, 118, 104 i 156. Suma tych $n_{111} = 30$ wartości wynosi $Y_{111} = 4440$ i oznacza liczbę jaj zniesionych przez wszystkie kury-córki po 1-szym kogucie 1-szej fermy w 1-szym roku. W podobny sposób wyznaczono pozostałe pozycje tablicy 4.

W modelu (2) wszystkie efekty poza μ są losowe, gdyż traktujemy dane dwa lata za wylosowane z populacji lat, uważamy również, że cztery fermy zostały wylosowane z populacji ferm, następnie, że koguty były wylosowane w poszczególnych fermach. Omawiany model nazwalibyśmy

stałym, gdyby wszystkie jego efekty poza błędem e_{ijk} potraktować jako stałe, tzn. gdyby rozważania ograniczyć tylko do wybranych kogutów z danych ferm w dwu rozpatrywanych latach. W przypadku modelu stałego porównujemy ze sobą średnie podklas lub klas (np. średnie dla ferm, lat, kogutów w fermo-latach, ...), podczas gdy w modelu losowym interesują nas jedynie wariancje poszczególnych efektów (np. wariancje ferm, lat, kogutów w fermo-latach, ...).

Wróćmy do modelu losowego (2). Zakładamy również, że efekty są nieskorelowane i mają średnie równe zeru, tzn. że

$$\mathcal{E}(a_i) = \mathcal{E}(b_j) = \mathcal{E}(ab)_{ij} = \mathcal{E}c(ab)_{k(ij)} = \mathcal{E}e_{ijk} = 0.$$

Problem polega na wyznaczeniu ocen pięciu wariancji: σ_A^2 , σ_B^2 , σ_{AB}^2 , $\sigma_{C(AB)}^2$ i σ_e^2 metodą RO.

Korzystając z uwag poprzedniego paragrafu konstruujemy tablicę 5, analogiczną do tablicy 3. Tak więc pod kolumną wartości oczekiwanych tworzymy 6 kolumn o numerach 5, 6, 7, 8, 9 i 10, tj. dokładnie tyle, ile jest efektów po prawej stronie równości w modelu (2): jako tytuły tych kolumn wypisujemy odpowiednie wskaźniki efektów, przyporządkowując zero średniej μ : 0, i , j , ij , $k(ij)$, ijk . Poniżej zamieszczamy odpowiednio μ^2 i wariancje efektów: σ_A^2 , σ_B^2 , σ_{AB}^2 , $\sigma_{C(AB)}^2$ i σ_e^2 . W kolumnie 3 notujemy wskaźniki wypisane wyżej z pominięciem nawiasów i z zastąpieniem zera jedyneką: ijk , ijk , ij , j , i , 1.

Na podstawie wskaźników kolumny 3 tworzymy składniki sum ważonych kwadratów odchyłeń (por. kol. 2), pisząc sumy względem wszystkich wskaźników zaznaczonych i umieszczając za znakami sum Y^2/n z odpowiednimi identycznymi wskaźnikami przy Y^2 i n .

W kolumnie 4 wypisujemy liczby stopni swobody, które wyznacza się sumując jedynekę względem zaznaczonych wskaźników; tak więc mamy $\sum_i \sum_j \sum_k \sum_g 1 = n = 2099$ kur-córek, $\sum_i \sum_j \sum_k 1 = 43$ koguty, $\sum_i \sum_j 1 = IJ = 8$ fermo-lat, $\sum_j 1 = J = 4$ fermy, $\sum_i 1 = I = 2$ lata.

Współczynniki pod kolumnami 5, 6, 7, 8, 9 i 10 wyznaczamy zgodnie z metodą RO przedstawioną w paragrafie poprzednim (por. [17]). Zanim wypiszemy równania ze względu na niewiadome μ^2 i wariancje, wykonajmy rachunki wstępne, polegające na obliczeniu wartości składników kolumny 2 i wartości współczynników z kolumn 5-10.

Otóż w składnikach kolumny 2 występują sumy Y postaci $Y_{ijk..}$, $Y_{ij..}$, $Y_{j..}$, $Y_{i..}$, których wartości podano w tablicy 4; np. $Y_{213..} = 5859$, $Y_{13..} = 16815$, $Y_{2..} = 88279$, $Y_{2...} = 204457$. Kropki stoją na miejscach wskaźników, względem których sumowano; stąd definicje sum Y są następujące:

TABLICA 5

Wartości oczekiwane składników sum ważonych kwadratów odchyżeń dla modelu (2)

Lp.	Składniki sum ważonych kwadratów odchyżeń S	Wskaźniki przy sumach	Stopnie swobody ν	Współczynniki przy komponentach wariancyjnych dla wartości oczekiwanych składników sum ważonych kwadratów odchyżeń $E(S)$						
				0	i	j	ij	k(ij)	ijk	σ_e^2
				μ^2	σ_A^2	σ_B^2	σ_{AB}^2	$\sigma_{C(AB)}^2$		
1	2	3	4	5	6	7	8	9	10	
1	$\sum_i \sum_j \sum_k \sum_l y_{ijkl}^2$	ijkl	$n = 2099$ kur-córek	n	n	n	n	n	n	n
2	$\sum_i \sum_j \sum_k \frac{Y_{ijk}^2}{n_{ijk}}$	ijk	$r = 43$ koguty	n	n	n	n	n	n	r
3	$\sum_i \sum_j \frac{Y_{ij..}^2}{n_{ij.}}$	ij	$IJ = 8$ fermo-lat	n	n	n	n	$W_{ij[k]}$	IJ	
4	$\sum_j \frac{Y_{j..}^2}{n_{j.}}$	j	$J = 4$ fermy	n	$W_{j[i]}$	n	$W_{j[i]}$	$W_{j[ik]}$	J	
5	$\sum_i \frac{Y_{i...}^2}{n_{i..}}$	i	$I = 2$ lata	n	n	$W_{i[j]}$	$W_{i[j]}$	$W_{i[jk]}$	I	
6	$\frac{Y^{2....}}{n}$	1	1 poprawka	n	W_i	W_j	W_{ij}	W_{ijk}	1	

$Y_{ijk} = \sum_g y_{ijk_g}$ — suma wyników k -tego koguta w i -tym roku na j -tej fermie,

$Y_{ij..} = \sum_k Y_{ijk}$ — suma wyników w i -tym roku na j -tej fermie,

$Y_{.j..} = \sum_i Y_{ij..}$ — suma wyników j -tej fermy,

$Y_{i...} = \sum_j Y_{ij..}$ — suma wyników i -tego roku,

$Y_{....} = \sum_i Y_{i...}$ — suma wszystkich obserwacji.

Definicje symboli n użytych dla oznaczenia liczebności poszczególnych podklas są następujące:

$n_{ij} = \sum_k n_{ijk}$ — liczba kur w i -tym roku na j -tej fermie,

$n_{.j} = \sum_i n_{ij}$ — liczba kur na j -tej fermie,

$n_{i..} = \sum_j n_{ij}$ — liczba kur w i -tym roku,

$n = \sum_i n_{i..}$ — liczba wszystkich kur.

Otóż suma kwadratów $n = 2099$ wyników wynosi $\sum_i \sum_j \sum_k \sum_g y_{ijk_g}^2 = 48196293$.

Obliczamy dalsze składniki, korzystając z danych tabelicy 4:

$$\begin{aligned} \sum_i \sum_j \sum_k \frac{Y_{ijk}^2}{n_{ijk}} &= \frac{Y_{111}^2}{n_{111}} + \frac{Y_{112}^2}{n_{112}} + \dots + \frac{Y_{246}^2}{n_{246}} = \\ &= \frac{4440^2}{30} + \frac{9115^2}{63} + \dots + \frac{12101^2}{84} = 46597689,638 \end{aligned}$$

jest sumą 43 składników, wobec 43 kogutów,

$$\begin{aligned} \sum_i \sum_j \frac{Y_{ij..}^2}{n_{ij.}} &= \frac{Y_{11..}^2}{n_{11.}} + \frac{Y_{12..}^2}{n_{12.}} + \dots + \frac{Y_{24..}^2}{n_{24.}} = \\ &= \frac{36353^2}{258} + \frac{38891^2}{221} + \dots + \frac{80858^2}{543} = 46510566,097 \end{aligned}$$

jest sumą 8 składników (8 fermo-lat),

$$\begin{aligned} \sum_j \frac{Y_{.j..}^2}{n_{.j.}} &= \frac{Y_{.1..}^2}{n_{.1.}} + \frac{Y_{.2..}^2}{n_{.2.}} + \frac{Y_{.3..}^2}{n_{.3.}} + \frac{Y_{.4..}^2}{n_{.4.}} = \\ &= \frac{74447^2}{588} + \frac{88279^2}{521} + \frac{52932^2}{350} + \frac{94665^2}{640} = 46391310,547 \end{aligned}$$

jest sumą 4 składników (4 fermy),

$$\sum_i \frac{Y_{i...}^2}{n_{i..}} = \frac{Y_{1...}^2}{n_{1..}} + \frac{Y_{2...}^2}{n_{2..}} = \frac{105866^2}{683} + \frac{204457^2}{1416} = 45931040,96$$

jest sumą 2 składników (2 lata), a

$$\frac{Y_{...}^2}{n} = \frac{310323^2}{2099} = 45879163,57.$$

Wyznamy wartości dziewięciu współczynników W z tablicy 5 na podstawie liczebności n tablicy 4; wzory na nie są analogiczne do wzorów podanych w paragrafie poprzednim:

$$\begin{aligned} W_{ij[k]} &= \sum_i \sum_j \frac{\sum_k n_{ijk}^2}{n_{ij.}} = \frac{n_{111}^2 + n_{112}^2 + \dots + n_{114}^2}{n_{11.}} + \\ &+ \frac{n_{121}^2 + n_{122}^2 + \dots + n_{126}^2}{n_{12.}} + \dots + \frac{n_{241}^2 + n_{242}^2 + \dots + n_{246}^2}{n_{24.}} = \\ &= \frac{30^2 + 63^2 + 72^2 + 93^2}{258} + \frac{44^2 + 32^2 + \dots + 55^2}{221} + \\ &+ \dots + \frac{94^2 + 82^2 + \dots + 84^2}{543} = 435,279 \end{aligned}$$

(występuje tu 8 ułamków wobec 8 fermo-lat),

$$\begin{aligned} W_{\{i\}} &= \sum_j \frac{\sum_i n_{ij.}^2}{n_{.j.}} = \\ &= \frac{n_{11.}^2 + n_{21.}^2}{n_{.1.}} + \frac{n_{12.}^2 + n_{22.}^2}{n_{.2.}} + \dots + \frac{n_{14.}^2 + n_{24.}^2}{n_{.4.}} = \\ &= \frac{258^2 + 330^2}{588} + \frac{221^2 + 300^2}{521} + \dots + \frac{97^2 + 543^2}{640} = 1231,723, \end{aligned}$$

$$\begin{aligned} W_{\{ik\}} &= \sum_j \frac{\sum_i \sum_k n_{ijk}^2}{n_{.j.}} = \frac{n_{111}^2 + n_{112}^2 + \dots + n_{126}^2}{n_{.1.}} + \\ &+ \frac{n_{121}^2 + n_{122}^2 + \dots + n_{226}^2}{n_{.2.}} + \dots + \frac{n_{141}^2 + n_{142}^2 + \dots + n_{246}^2}{n_{.4.}} = \\ &= \frac{30^2 + 63^2 + \dots + 69^2}{588} + \frac{44^2 + 32^2 + \dots + 77^2}{521} + \\ &+ \frac{21^2 + 10^2 + \dots + 84^2}{640} = 242,845, \end{aligned}$$

$$\begin{aligned}
 W_{i[j]} &= \sum_i \frac{\sum_j n_{ij}^2}{n_{i..}} = \\
 &= \frac{n_{11.}^2 + n_{12.}^2 + \dots + n_{14.}^2}{n_{1..}} + \frac{n_{21.}^2 + n_{22.}^2 + \dots + n_{24.}^2}{n_{2..}} = \\
 &= \frac{258^2 + 221^2 + \dots + 97^2}{683} + \frac{330^2 + 300^2 + \dots + 543^2}{1416} = 589,91,
 \end{aligned}$$

$$\begin{aligned}
 W_{i[jk]} &= \sum_i \frac{\sum_j \sum_k n_{ijk}^2}{n_{i..}} = \\
 &= \frac{n_{111}^2 + n_{112}^2 + \dots + n_{146}^2}{n_{1..}} + \frac{n_{211}^2 + n_{212}^2 + \dots + n_{246}^2}{n_{2..}} = \\
 &= \frac{30^2 + 63^2 + \dots + 4^2}{683} + \frac{47^2 + 44^2 + \dots + 84^2}{1416} = 119,89,
 \end{aligned}$$

$$W_i = \frac{\sum_i n_{i..}^2}{n} = \frac{n_{1..}^2 + n_{2..}^2}{n} = \frac{683^2 + 1416^2}{2099} = 1177,48,$$

$$W_j = \frac{\sum_j n_{.j}^2}{n} = \frac{n_{.1}^2 + n_{.2}^2 + \dots + n_{.4}^2}{n} = \frac{588^2 + 521^2 + \dots + 640^2}{2099} = 547,55,$$

$$W_{ij} = \frac{\sum_i \sum_j n_{ij}^2}{n} = \frac{n_{11.}^2 + n_{12.}^2 + \dots + n_{24.}^2}{n} = \frac{258^2 + 221^2 + \dots + 543^2}{2099} = 328,28,$$

$$W_{ijk} = \frac{\sum_i \sum_j \sum_k n_{ijk}^2}{n} = \frac{n_{111}^2 + n_{112}^2 + \dots + n_{246}^2}{n} = \frac{30^2 + 63^2 + \dots + 84^2}{2099} = 63,17.$$

Uwzględniając obliczone wartości składników sum ważonych kwadratów odchyłeń i wartości współczynników W oraz korzystając z tablicy 5 otrzymujemy następujący układ sześciu równań z sześcioma niewiadomymi $\hat{\mu}^2$, $\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$, $\hat{\sigma}_{AB}^2$, $\hat{\sigma}_{C(AB)}^2$ i $\hat{\sigma}_e^2$:

$$\begin{aligned}
 48196293 &= 2099\hat{\mu}^2 + 2099\hat{\sigma}_A^2 + 2099\hat{\sigma}_B^2 + 2099\hat{\sigma}_{AB}^2 + 2099\hat{\sigma}_{C(AB)}^2 + 2099\hat{\sigma}_e^2, \\
 46597689,638 &= 2099\hat{\mu}^2 + 2099\hat{\sigma}_A^2 + 2099\hat{\sigma}_B^2 + 2099\hat{\sigma}_{AB}^2 + 2099\hat{\sigma}_{C(AB)}^2 + 43\hat{\sigma}_e^2, \\
 46510566,097 &= 2099\hat{\mu}^2 + 2099\hat{\sigma}_A^2 + 2099\hat{\sigma}_B^2 + 2099\hat{\sigma}_{AB}^2 + 435,279\hat{\sigma}_{C(AB)}^2 + 8\hat{\sigma}_e^2, \\
 46391310,547 &= 2099\hat{\mu}^2 + 1231,723\hat{\sigma}_A^2 + 2099\hat{\sigma}_B^2 + 1231,723\hat{\sigma}_{AB}^2 + \\
 &\quad + 242,845\hat{\sigma}_{C(AB)}^2 + 4\hat{\sigma}_e^2, \\
 45931040,960 &= 2099\hat{\mu}^2 + 2099\hat{\sigma}_A^2 + 589,91\hat{\sigma}_B^2 + 589,91\hat{\sigma}_{AB}^2 + \\
 &\quad + 119,89\hat{\sigma}_{C(AB)}^2 + 2\hat{\sigma}_e^2, \\
 45879163,570 &= 2099\hat{\mu}^2 + 1177,48\hat{\sigma}_A^2 + 547,55\hat{\sigma}_B^2 + 328,28\hat{\sigma}_{AB}^2 + \\
 &\quad + 63,17\hat{\sigma}_{C(AB)}^2 + \hat{\sigma}_e^2.
 \end{aligned}$$

W przypadku uzyskania ujemnych ocen wariancji sugeruje się, aby wyeliminować z modelu statystycznego odpowiednie efekty i znaleźć oceny wariancji dla nowego modelu według metody RO.

Ponieważ rozwiązanie tego układu daje ujemną ocenę na $\hat{\sigma}_A^2$, więc przyjmujemy $\hat{\sigma}_A^2 = 0$. Przy tym założeniu rozwiązujemy układ równań, pomijając równanie z członem

$$\sum_i \frac{Y_{i...}^2}{n_{i..}} = 45931040,960.$$

W ten sposób uzyskano następujące oceny: $\hat{\sigma}_A^2 = \hat{\sigma}_{\text{lata}}^2 = 0$, $\hat{\sigma}_B^2 = \hat{\sigma}_{\text{fermy}}^2 = 251,10$, $\hat{\sigma}_{AB}^2 = \sigma_{\text{lata} \times \text{fermy}}^2 = 125,93$, $\hat{\sigma}_{C(AB)}^2 = \hat{\sigma}_{\text{koguty w fermo-latach}}^2 = 36,01$, $\hat{\sigma}_e^2 = 777,53$, $\hat{\mu}^2 = 21770,98$.

5. Nieortogonalny model losowy typu $D(C(A)) \times B$ wyrażony w formie

$$(3) \quad y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \{c(a)\}_{k(i)} + \{bc(a)\}_{jk(i)} + e_{ijk}$$

z czynnikami A , B i C oraz odpowiadającymi im wskaźnikami

$A: i = 1, \dots, I; \quad C(A): k = 1, \dots, m_i;$

$B: j = 1, \dots, J; \quad \text{Indywidualna w } ABC: g = 1, \dots, n_{ijk};$

można uzyskać z modelu (1) przez usunięcie czynnika D i odpowiadającego mu wskaźnika m . W ten sposób także definicje liczebności podklas $n_{ij.}$, $n_{i.k}$ itd. oraz definicje sum podklas formy $Y_{ijk.}$, $Y_{ij..}$ itd. otrzymuje się z odpowiednich definicji modelu (1) usuwając D i wskaźnik m .

Dane liczbowe dla modelu (3), gdy A występuje w dwóch poziomach, B w trzech a C w trzech w zakresie A_1 i w dwóch w zakresie A_2 , można zestawzić według następującego schematu:

		B_1	B_2	B_3
A_1	C_1	.	.	.
	C_2	.	.	.
	C_3	.	.	.
A_2	C_1	.	.	.
	C_2	.	.	.

Jeżeli z tablicy 2 usuniemy źródła zmienności z D oznaczone numerami 6 i 7, to otrzymamy analizę wariancji dla modelu (3).

Jeżeli teraz z tablicy 3 usuniemy wskaźnik m i odpowiadający mu czynnik D , to otrzymamy układ równań, którego rozwiązanie da nam oceny dla μ^2 i 6 komponentów wariancyjnych: σ_A^2 , σ_B^2 , σ_{AB}^2 , $\sigma_{C(A)}^2$, $\sigma_{BC(A)}^2$ i σ_e^2 .

W przypadku danych ortogonalnych przyjmujemy: $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$ ($m_i = K$); $g = 1, \dots, G$ ($n_{ijk} = G$) i wtedy $n_{ij.} = KG$, $n_{i.k} = JG$, $n_{i..} = JKG$, $n_{.j.} = IKG$, $n = IJKG$, $p = \sum_I m_i = IK$.

6. Niepełne i nieortogonalne modele losowe. Należy zauważyć, że metoda RO stosuje się również wówczas gdy dane liczbowe są niepełne (niekompletna klasyfikacja krzyżowa), tzn. gdy zostaje naruszona symetria klasyfikacji lub gdy pewne podklasy są puste. Za przykład posłużyć nam mogą dane tablicy 4, uzupełnione piątą fermą, która występuje tylko w pierwszym roku. Takie właśnie były oryginalne wyniki liczbowe, których część reprezentuje tablica 4. Innym przykładem danych niepełnych są te, które obejmują nieśność w teście 500-dniowym u kur pochodzących z lęgów w kilku fermach i latach. Ponieważ lęgi w jednym roku nie zawsze występują w tych samych tygodniach co lęgi w drugim roku, więc w tabeli zestawiającej dane liczbowe znajdują się podklasy puste, bez obserwacji.

Przy stosowaniu metody RO dla danych niezupełnych wzory na składniki sum kwadratów odchyień i na współczynniki W są analogiczne jak przy danych pełnych z tym, że pomijamy podklasy puste. Przy rachunkach należy zwrócić szczególną uwagę na obliczanie stopni swobody, będących współczynnikami przy wariancji σ_e w układzie równań ze względu na nieznane komponenty wariancyjne.

Jestem głęboko zobowiązany dr W. Gluchowskiemu z Katedry Ogólnej Hodowli Zwierząt Wyższej Szkoły Rolniczej w Lublinie za udostępnienie mi danych liczbowych zestawionych w tablicy 4.

Prace cytowane

- [1] R. L. Anderson i T. A. Bancroft, *Statistical Theory in Research*, New York 1952.
- [2] C. A. Bennett i N. L. Franklin, *Statistical Analysis in Chemistry and the Chemical Industry*, New York 1954, str. 413-415, 423-425.
- [3] K. A. Brewnlee, *Statistical Theory and Methodology in Science and Engineering*, New York 1960.
- [4] — *Industrial Experimentation*, 3rd American ed., New York 1949.
- [5] S. L. Crump, *The estimation of variance components in analysis of variance*, *Biometrics* 2 (1946), str. 7-11.
- [6] M. Ganguli, *A note on nested sampling*, *Sankhya*, 5 (1941), str. 449-452.
- [7] C. E. Gould i W. M. Hampton, *Statistical methods applied to the manufacture of spectacle glasses*, *J. Royal Stat. Soc. Supp.* 3 (1936), str. 137.
- [8] W. R. Harvey, *Least squares analysis of data with unequal subclass numbers*, Agricultural Research Service, U. S. Department of Agriculture, ARS-20-8, July 1960, str. 36-37, 46-47.
- [9] L. N. Hazel, *The covariance of multiple classification tables with unequal subclass numbers*, *Biomet. Bull.* 2 (1946), str. 21-25.

- [10] C. R. Henderson, *Estimation of variance and covariance components*, Biometrics 9 (1953), str. 226-252.
- [11] S. C. King i C. R. Henderson, *Variance component analysis in heritability studies*, Poultry Sci. 33 (1954), str. 155-169.
- [12] H. L. Le Roy i W. Gluckowski, *Die bestimmung der Varianzkomponenten im a.b.c-Faktorenversuch mit ungleichen Klassenfrequenzen*, Biometrische Zeitschrift 3 (1961), str. 73-91.
- [13] W. Oktaba, *O wzajemnej i jednoznacznej odpowiedniości między sumami kwadratów i stopniami swobody w analizie wariancji*, Annales UMCS, Lublin, Sec. E., 13 (1958), str. 327-344.
- [14] — *Elementy statystyki matematycznej i metodyka doświadczalnictwa*, Łódź, Warszawa 1962.
- [15] — *Słownik polsko-rosyjsko-angielski statystyki matematycznej i teorii doświadczania*, Lublin 1963.
- [16] — *Nieortogonalne modele losowe klasyfikacji hierarchicznej*, Roczniki Nauk Rolniczych, 1963, 82 B-3, str. 417-435.
- [17] — *O twierdzeniu dotyczącym metody wyznaczania ocen komponentów wariancyjnych w nieortogonalnych modelach losowych*, w przygotowaniu dla Zastosow Mat.
- [18] H. Scheffé, *The Analysis of Variance*, New York 1959, str. 188.
- [19] S. R. Searle, *Sampling variances of estimates of components of variance*, AMS 29 (1958), str. 167-178.
- [20] R. G. D. Steel i J. H. Torrie, *Principles and Procedures of Statistics*, New York 1960.

Praca wpłynęła 5. 4. 1963

В. ОКТАБА (Люблин)

**ОЦЕНКА КОМПОНЕНТОВ ДИСПЕРСИИ
В НЕОРТОГОНАЛЬНЫХ СЛУЧАЙНЫХ МОДЕЛЯХ
ОСНОВАННЫХ НА КОМБИНАЦИИ ПЕРЕКРЕСТНОЙ КЛАССИФИКАЦИИ
С ИЕРАРХИЧЕСКОЙ**

РЕЗЮМЕ

Рассматривается случайная модель с p некоррелированными эффектами, $a, b, \dots, e$ средние которых являются нулями а дисперсии которых равны соответственно $\sigma_A^2, \sigma_B^2, \dots, \sigma_e^2$. Численные данные, соответствующие этой модели представляют собой произвольную комбинацию перекрестной классификации с иерархической и могут содержать диспропорциональные числа наблюдений в подгруппах, и в частности некоторые подгруппы без наблюдения.

При этих предположениях представлено здесь метод РО оценки компонентов дисперсии (ср. § 3 и § 4, а также [17]), используя метод R, который Леруа применил к неортогональной случайной модели, основанной на перекрестной классификации (ср. [12]), а источником которого являются результаты Гендерсона (ср. [10], [19]). Метод РО при использовании соответствующих обозначений, позволяет выписать систему $p+1$ уравнений с $p+1$ неизвестными $\mu^2, \sigma_A^2, \sigma_B^2, \dots, \sigma_e^2$ оперируя исключительно индексами эффектов случайной модели и не прибегая

к вычислению $p+1$ величин ожидаемых слагаемых взвешенных сумм квадратов. Соответствующие коэффициенты при неизвестных, определяемые указанным методом, являются простыми функциями количеств наблюдений в подгруппах (ср. § 3 и § 4, а также [17]).

Формулы полученные Ганголи (ср. [6]) для неортогональной случайной модели в случае иерархической классификации, получаются этим методом как частный случай. Правило Крампа (ср. § 1) для ортогональных моделей также можно вывести из правила RO.

Метод RO и вычислительную технику иллюстрируется на примере неортогональных генетических данных с области культуры домашних птиц, которые являются комбинацией двусторонней перекрестной классификации (годы $\times$ фермы) с двусторонней иерархической классификацией (курицы-дочеры в сфере петухов внутри фермо-лет).

W. OKTAŁA (Lublin)

*ESTIMATION OF VARIANCE COMPONENTS IN NON-ORTHOGONAL
RANDOM MODELS BASED ON THE COMBINATION
OF CROSS- AND NESTED CLASSIFICATION*

SUMMARY

The author considers a random model with p uncorrelated effects $a, b, \dots, e$ with means zero and variances $\sigma_a^2, \dots, \sigma_e^2$. The numerical data corresponding to this model form an arbitrary combination of cross and nested classification, and they may contain disproportional numbers of observations in subclasses or, in particular, in some subclasses there may be no observations.

Under these assumptions the so-called RO method of estimating the variance components is presented (see § 3, § 4 and [17]). Use is made of the so-called R method applied by Le Roy to the non-orthogonal random model based on cross classification (see [12]); this in turn is based on Henderson's result (see [10], [19]). The RO method, by the use of suitable symbols, allows us to write down a system of $p+1$ equations with $p+1$ unknowns $\mu^2, \sigma_a^2, \sigma_b^2, \dots, \sigma_e^2$. The advantage is that we deal with indices of effects of the random model without computing $p+1$ expected values of the components of weighted sums of squares. The corresponding coefficients standing by the unknowns and determined by the above mentioned method turn out to be simple functions of the frequencies of subclasses (see § 3, § 4 and [17]).

The formulas derived by Gangula [6] for non-orthogonal random models in the case of nested classification are obtained here as a special case. Crump's rule (see § 1) for orthogonal models may also be obtained from the RO rule.

The RO method and computing technique are illustrated with an example of non-orthogonal genetic data from chicken farms: they form a combination of a two-way cross classification (years $\times$ farms) with a two-way nested classification (chickens-daughters belonging to one rooster in given farm and year).