

ROMAN ZMYŚLONY (Wrocław)

Kwadratowo dopuszczalne estymatory w modelach losowych

(Praca przyjęta do druku 20.12.1974)

1. Wstęp. W niniejszej pracy rozważamy następujący model:

$$(1) \quad \mathbf{y} = \sum_{i=1}^k \mathbf{X}_i \mathbf{b}_i,$$

gdzie \mathbf{X}_i są znanymi n -wierszowymi macierzami, a \mathbf{b}_i , z wyjątkiem \mathbf{b}_1 , wektorami losowymi. Dodatkowo zakładamy, że $\mathbf{X}_1 = (1, \dots, 1)' = \mathbf{1}$ oraz $\mathbf{X}_k = \mathbf{I}$ jest macierzą jednostkową. Dla $i = 2, \dots, k$ wektory losowe \mathbf{b}_i mają wielowymiarowe rozkłady normalne $N(0, \sigma_i^2 \mathbf{V}_i)$, gdzie \mathbf{V}_i są znanymi macierzami, natomiast $\sigma_i^2 \geq 0$ dla $i = 2, \dots, k-1$ oraz $\sigma_k^2 > 0$. Ponadto zakładamy, że $\mathbf{V}_k = \mathbf{I}$ oraz że wektory \mathbf{b}_i są niezależne, tzn. $E(\mathbf{b}_i \mathbf{b}_j') = 0$ dla $i \neq j$. Dla wygody rzeczywisty parametr \mathbf{b}_1 zastępujemy przez σ_1 . Modele o wyżej opisanej strukturze noszą nazwę *modeli losowych*, a parametry $\sigma_2^2, \dots, \sigma_k^2$ nazywają się *komponentami wariancyjnymi*.

Dla estymacji komponentów wariancyjnych powszechnie stosowana jest *metoda Hendersona* [3], zwana inaczej *metodą analizy wariancji*. Otrzymane estymatory są nieobciążone i są formami kwadratowymi wektora \mathbf{y} . Przy założeniu, że macierze \mathbf{V}_i są macierzami jednostkowymi, Graybill i Hultquist [1] podali warunki dostateczne na to, aby estymatory otrzymane za pomocą tej metody miały minimalną wariancję w klasie wszystkich estymatorów nieobciążonych.

W pracy [4] podana jest inna metoda, która prowadzi do otrzymania estymatorów nieobciążonych dla komponentów wariancyjnych czy też dowolnych liniowych funkcji parametrów $\sigma_1^2, \dots, \sigma_k^2$. Estymatorami dla σ_i^2 , $i = 1, \dots, k$, są funkcje $\hat{\sigma}_i^2 = \varphi_i(\mathbf{y})$, które minimalizują normę macierzy $\mathbf{y}\mathbf{y}' - E(\mathbf{y}\mathbf{y}')$, gdzie norma w przestrzeni \mathcal{A} macierzy symetrycznych $n \times n$ jest indukowana przez iloczyn skalarny $\langle \mathbf{A}, \mathbf{B} \rangle$ określony jako ślad iloczynu macierzy \mathbf{A} i \mathbf{B} . Estymatorem funkcji postaci

$$h = \sum_{i=1}^k c_i \sigma_i^2 \text{ jest}$$

$$\hat{h} = \sum_{i=1}^k c_i \hat{\sigma}_i^2$$

zwany *estymatorem otrzymanym metodą najmniejszych kwadratów* lub krótko *estymatorem NK*. W przypadku, gdy przestrzeń \mathcal{A}_1 generowana przez macierze $\mathbf{A}_1 =$

$= \mathbf{1}'$, $\mathbf{A}_2 = \mathbf{X}_2 \mathbf{V}_2 \mathbf{X}_2'$, ..., $\mathbf{A}_k = \mathbf{X}_k \mathbf{V}_k \mathbf{X}_k'$ ma wymiar k , to estymatory NK są estymatorami nieobciążonymi dla dowolnej liniowej funkcji parametrów $\sigma_1^2, \dots, \sigma_k^2$. Jeśli ponadto \mathcal{A}_1 jest podprzestrzenią kwadratową przestrzeni \mathcal{A} , to znaczy, jeśli $\mathbf{A} \in \mathcal{A}_1$ implikuje $\mathbf{A}^2 \in \mathcal{A}_1$, to estymatory NK mają minimalną wariancję w klasie kwadratowych i nieobciążonych estymatorów.

W niniejszej pracy rozważamy takie modele losowe, dla których \mathcal{A}_1 nie jest podprzestrzenią kwadratową przestrzeni \mathcal{A} . Przykładem takiego modelu jest model losowy dla jednokierunkowej klasyfikacji z niejednakową liczbą obserwacji w podklasach. Dla tego modelu Harville [2] podał ogólną postać każdego kwadratowo dopuszczalnego estymatora. W niniejszej pracy rozszerzony zostanie wynik Harville'a na dowolne modele losowe. Ponadto udowodnimy, że estymatory NK są kwadratowo dopuszczalne.

2. Estymatory kwadratowo dopuszczalne. W niniejszej pracy będziemy zakładali, że podprzestrzeń \mathcal{A}_1 przestrzeni \mathcal{A} nie jest podprzestrzenią kwadratową. W tym przypadku estymatory otrzymane metodą najmniejszych kwadratów nie mają minimalnej wariancji w klasie estymatorów kwadratowych i nieobciążonych (patrz [4], twierdzenie 2). Za kryterium optymalności estymatora przyjmujemy pochodzącą od Harville'a następującą definicję estymatora kwadratowo dopuszczalnego. Niech Ω oznacza podzbiór R^k zdefiniowany w następujący sposób:

$$\Omega = \{ \sigma : \sigma = (\sigma_1^2, \dots, \sigma_k^2)', \sigma_i^2 \geq 0, i = 1, \dots, k-1, \sigma_k^2 > 0 \}.$$

DEFINICJA. Funkcję $\mathbf{y}'\mathbf{D}\mathbf{y}$, gdzie $\mathbf{D} \in \mathcal{A}$, nazywamy *kwadratowo dopuszczalnym estymatorem wartości oczekiwanej* $E(\mathbf{y}'\mathbf{D}\mathbf{y})$, jeśli nie istnieje macierz $\mathbf{G} \in \mathcal{A}$ taka, że

$$(2) \quad \mathbf{E}\mathbf{y}'\mathbf{G}\mathbf{y} = \mathbf{E}\mathbf{y}'\mathbf{D}\mathbf{y} \quad \text{dla każdego } \sigma \in \Omega,$$

oraz

$$(3) \quad \text{Var}(\mathbf{y}'\mathbf{G}\mathbf{y}) \leq \text{Var}(\mathbf{y}'\mathbf{D}\mathbf{y}) \quad \text{dla każdego } \sigma \in \Omega$$

z silną nierównością dla co najmniej jednego $\sigma \in \Omega$.

Niech $\mathcal{A}_2 \subset \mathcal{A}$ będzie najmniejszą podprzestrzenią kwadratową zawierającą \mathcal{A}_1 . Podprzestrzeń taka zawsze istnieje, ponieważ \mathcal{A} jest niewłaściwą podprzestrzenią kwadratową i iloczyn mnogościowy podprzestrzeni kwadratowych jest podprzestrzenią kwadratową.

Poniższe twierdzenie podaje warunek konieczny na to, by estymator był kwadratowo dopuszczalny.

TWIERDZENIE 1. *Jeśli estymator $\mathbf{y}'\mathbf{D}\mathbf{y}$ jest kwadratowo dopuszczalny, to $\mathbf{D} \in \mathcal{A}_2$.*

D o w ó d. Załóżmy, że \mathcal{A}_2 jest właściwą podprzestrzenią kwadratową przestrzeni \mathcal{A} . Wystarczy pokazać, że jeśli $\mathbf{D} \notin \mathcal{A}_2$, to istnieje macierz \mathbf{G} spełniająca warunki (2) i (3). Niech $\mathbf{D} = \mathbf{G} + \mathbf{G}_1$, gdzie $\mathbf{G} \in \mathcal{A}_2$, a $\mathbf{G}_1 \perp \mathbf{G}$, tzn. $\langle \mathbf{G}, \mathbf{G}_1 \rangle = 0$. Ponieważ $\mathcal{A}_1 \subset \mathcal{A}_2$, przeto $\mathbf{G}_1 \in \mathcal{A}_1^\perp$, czyli jest elementem przestrzeni ortogonalnej do \mathcal{A}_1 . Stąd, na mocy lematu 2 z pracy [4], otrzymujemy, że $E(\mathbf{y}'\mathbf{G}\mathbf{y}) = E(\mathbf{y}'\mathbf{D}\mathbf{y})$. Aby wykazać (3), wykażemy najpierw, że $\text{Cov}(\mathbf{y}'\mathbf{G}\mathbf{y}, \mathbf{y}'\mathbf{G}_1\mathbf{y}) = 0$ dla każdego $\sigma \in \Omega$.

Z uwagi na to, że wektor \mathbf{y} ma rozkład normalny z macierzą kowariancji $\mathbf{W}\sigma = \sum_{i=2}^k \sigma_i^2 \mathbf{A}_i$, prawdziwy jest wzór

$$\text{Cov}(\mathbf{y}'\mathbf{G}\mathbf{y}, \mathbf{y}'\mathbf{G}_1\mathbf{y}) = 2\langle \mathbf{W}\sigma\mathbf{G}\mathbf{W}\sigma + \sigma_1^2(\mathbf{A}_1\mathbf{G}\mathbf{W}\sigma + \mathbf{W}\sigma\mathbf{G}\mathbf{A}_1), \mathbf{G}_1 \rangle.$$

Ponieważ $\mathbf{W}\sigma$, \mathbf{G} i \mathbf{A}_1 są elementami podprzestrzeni kwadratowej \mathcal{A}_2 , więc na mocy lematu 4 z pracy [4] otrzymujemy, że $\mathbf{W}\sigma\mathbf{G}\mathbf{W}\sigma + \sigma_1^2(\mathbf{A}_1\mathbf{G}\mathbf{W}\sigma + \mathbf{W}\sigma\mathbf{G}\mathbf{A}_1)$ również należy do \mathcal{A}_2 , co wraz z warunkiem $\mathbf{G}_1 \in \mathcal{A}_2^\perp$ implikuje, że $\text{Cov}(\mathbf{y}'\mathbf{G}\mathbf{y}, \mathbf{y}'\mathbf{G}_1\mathbf{y}) = 0$. Zatem

$$\text{Var}(\mathbf{y}'\mathbf{D}\mathbf{y}) = \text{Var}(\mathbf{y}'\mathbf{G}\mathbf{y}) + \text{Var}(\mathbf{y}'\mathbf{G}_1\mathbf{y}).$$

Ponieważ $\mathbf{G}_1 \neq \mathbf{0}$ i ponieważ rozkład wektora \mathbf{y} jest nieosobliwy w każdym punkcie $\sigma \in \Omega$, więc stąd wynika, że $\text{Var}(\mathbf{y}'\mathbf{G}\mathbf{y}) < \text{Var}(\mathbf{y}'\mathbf{D}\mathbf{y})$ dla każdego $\sigma \in \Omega$.

W przypadku, gdy $\mathcal{A}_2 = \mathcal{A}$, teza twierdzenia jest w sposób oczywisty spełniona.

Z powyższego twierdzenia wynika natychmiast

WNIOSEK. Niech $\mathbf{B}_1, \dots, \mathbf{B}_r$ będzie bazą przestrzeni \mathcal{A}_2 . Wtedy każdy kwadratowo dopuszczalny estymator można przedstawić w postaci

$$\sum_{i=1}^r \mathbf{b}_i \mathbf{y}' \mathbf{B}_i \mathbf{y},$$

gdzie $\mathbf{b} = (b_1, \dots, b_r)'$ jest pewnym wektorem z \mathbb{R}^r .

PRZYKŁAD. Pokażemy teraz, jak z twierdzenia powyższego można otrzymać rezultat Harville'a.

Rozważmy model losowy dla jednokierunkowej klasyfikacji

$$(4) \quad y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, n_i.$$

Tutaj μ jest parametrem stałym, natomiast α_i , e_{ij} , $i = 1, \dots, a$, $j = 1, \dots, n_i$ niezależnymi zmiennymi losowymi o rozkładach normalnych. Ponadto zakładamy, że $\alpha_i \sim N(0, \sigma_\alpha^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, przy czym $\sigma_\alpha^2 \geq 0$, $\sigma_e^2 > 0$ oraz $a \geq 2$ i $n_1 + \dots + n_a > a$.

Rozważać będziemy tylko ten przypadek, kiedy nie wszystkie n_i są jednakowe. Oznaczmy przez $\eta_1, \dots, \eta_{a^*}$ liczby, które powtarzają się co najmniej dwa razy w ciągu n_1, \dots, n_a , a przez $\eta_{a^*+1}, \dots, \eta_{a'}$ liczby, które w tym ciągu występują dokładnie jeden raz. Niech $S_i = \{j: n_j = \eta_i\}$ dla $i = 1, \dots, a'$ oraz niech ζ_i oznacza liczbę elementów zbioru S_i . Oczywiście, $\zeta_i > 1$ dla $i = 1, \dots, a^*$ oraz $\zeta_i = 1$ dla $i = a^*+1, \dots, a'$.

Harville [2] udowodnił, że przy powyższych założeniach każdy kwadratowo dopuszczalny estymator można przedstawić w postaci

$$(5) \quad c \sum_{i,j} y_{ij}^2 + \sum_{u=1}^{a^*} b_u \sum_{i \in S_u} \bar{y}_i^2 + \sum_{u=1}^{a'} \sum_{r \geq u} e_{ur} \bar{y}_u^* \bar{y}_r^*,$$

gdzie

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad y_i^* = \frac{1}{\zeta_i} \sum_{j \in S_i} \bar{y}_j,$$

podczas gdy c, b_u, e_{ur} są pewnymi liczbami rzeczywistymi.

Przedstawimy teraz dowód tego faktu, korzystając z udowodnionych poprzednio twierdzenia i wniosku. W tym celu wprowadzimy wektor y określony w następujący sposób. Niech $y_i = (y_{i1}, \dots, y_{in_i})$, $i = 1, \dots, a$, oraz niech $z_k = (y_{i_1}, \dots, y_{i_{\zeta_k}})$, $k = 1, \dots, a'$, gdzie $i_j \in S_k$ dla $j = 1, \dots, \zeta_k$. Kolejność elementów w ciągu $(y_{i_1}, \dots, y_{i_{\zeta_k}})$ może być dowolna. W końcu, niech

$$(6) \quad y = (z_1, \dots, z_{a^*}, z_{a^*+1}, \dots, z_{a'})'.$$

Zauważmy, że wektor y zawiera wszystkie y_{ij} , $i = 1, \dots, a$, $j = 1, \dots, n_i$, oraz że każdy y_{ij} występuje w wektorze y dokładnie jeden raz.

Niech $N_i = \eta_i \zeta_i$ oraz niech J_{ij} oznacza macierz jedynek o wymiarze $N_i \times N_j$, $i, j = 1, \dots, a'$, natomiast J_i macierz jedynek o wymiarze $\eta_i \times \eta_i$ dla $i = 1, \dots, a^*$. Następnie niech $D_i = \text{diag}_{\zeta_i}(J_i, \dots, J_i)$, dla $i = 1, \dots, a^*$, gdzie symbol $\text{diag}_p(B_1, \dots, B_p)$ oznacza macierz

$$\begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & B_p \end{bmatrix}.$$

Łatwo zauważyć, że przestrzeń \mathcal{A}_1 dla modelu (4), przy kolejności y_{ij} określonej przez (6), jest generowana przez macierze

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{1}\mathbf{1}', \\ \mathbf{A}_2 &= \text{diag}_{a'}(D_1, \dots, D_{a^*}, J_{a^*+1, a^*+1}, \dots, J_{a'a'}), \\ \mathbf{A}_3 &= I. \end{aligned}$$

Jeżeli nie wszystkie n_i są jednakowe, to \mathcal{A}_1 nie jest podprzestrzenią kwadratową przestrzeni \mathcal{A} , gdyż $A_1^2 \notin \mathcal{A}_1$.

Znajdziemy teraz bazę dla najmniejszej podprzestrzeni kwadratowej \mathcal{A}_2 zawierającej podprzestrzeń \mathcal{A}_1 . Rozbijmy w tym celu macierz $\mathbf{1}\mathbf{1}'$ na a'^2 podmacierzy

$$\begin{bmatrix} J_{11} & \dots & J_{1a'} \\ J_{21} & \dots & J_{2a'} \\ \dots & \dots & \dots \\ J_{a'1} & \dots & J_{a'a'} \end{bmatrix}.$$

Niech C_{ij} , dla $i \geq j$, będzie macierzą kwadratową $n \times n$ otrzymaną z macierzy $\mathbf{1}\mathbf{1}'$ przez zastąpienie wszystkich podmacierzy J_{kl} macierzami zerowymi z wyjątkiem J_{ij} oraz J_{ji} . Następnie, niech C_i dla $i = 1, \dots, a^*$ oznacza macierz otrzymaną z \mathbf{A}_2 przez zastąpienie wszystkich podmacierzy J_{ii} oraz \mathbf{D}_k , z wyjątkiem \mathbf{D}_i , macierzami zerowymi. Macierze C_{ij} , C_i oraz I są liniowo niezależne, gdyż żadna z nich nie da

się przedstawić jako liniowa kombinacja pozostałych. Ponadto, korzystając z własności podprzestrzeni kwadratowych, łatwo można udowodnić, że przestrzeń generowana przez te macierze jest kwadratową podprzestrzenią przestrzeni \mathcal{A} i zawiera \mathcal{A}_1 . Aby udowodnić, że przestrzeń ta pokrywa się z przestrzenią \mathcal{A}_2 , wystarczy sprawdzić, że C_{ij} oraz σ_i należą do \mathcal{A}_2 . Wykażemy najpierw, że $C_{ii} \in \mathcal{A}_2$ dla $i = a^* + 1, \dots, a'$ oraz że $C_i \in \mathcal{A}_2$ dla $i = 1, \dots, a^*$. Ponieważ $A_2 \in \mathcal{A}_2$, przeto dla $i = 1, \dots, a'$ macierze $B_i = \eta_i A_2 - A_2^2$ należą do \mathcal{A}_2 . Ponadto można łatwo udowodnić, że macierze B_i, B_j dla $i, j = 1, \dots, a'$ komutują, a stąd i na mocy lematu 4 z pracy [4] dowolny skończony iloczyn tych macierzy należy do \mathcal{A}_2 . Co więcej, $i \neq j$ implikuje, że $\eta_i \neq \eta_j$, a w takim razie macierze

$$G_i = \left(\prod_{j \neq i} B_j \right) \left(\eta_i^{a'-1} \prod_{j \neq i} (\eta_i - \eta_j) \right)^{-1}, \quad i = 1, \dots, a',$$

są dobrze określone i należą do \mathcal{A}_2 .

Nietrudno zauważyć, że

$$G_i = \begin{cases} C_i & \text{dla } i = 1, \dots, a^*, \\ C_{ii} & \text{dla } i = a^* + 1, \dots, a'. \end{cases}$$

Udowodnimy teraz, że $C_{ii} \in \mathcal{A}_2$ dla $i = 1, \dots, a^*$ oraz że $C_{ij} \in \mathcal{A}_2$ dla $i > j$, $j = 1, \dots, a'$.

Istotnie, ponieważ $A_1 \in \mathcal{A}_2$ oraz dla $i = 1, \dots, a^*$

$$C_{ii} = \frac{1}{\eta_i^2} C_i A_1 C_i,$$

przeto na mocy wyżej cytowanego lematu macierze te należą do \mathcal{A}_2 . Podobnie łatwo można sprawdzić, że dla $i > j$, $i, j = 1, \dots, a'$,

$$C_{ij} = \frac{1}{2N_i N_j} (C_{ii} A_1 C_{jj} + C_{jj} A_1 C_{ii}),$$

więc macierze te również należą do \mathcal{A}_2 .

Z wniosku otrzymujemy, że każdy kwadratowo dopuszczalny estymator da się przedstawić w postaci

$$(7) \quad c\mathbf{y}'\mathbf{I}\mathbf{y} + \sum_{u=1}^{a^*} b'_u \mathbf{y}' C_u \mathbf{y} + \sum_{u=1}^{a'} \sum_{r \geq u} e_{ru} \mathbf{y}' C_{ru} \mathbf{y},$$

gdzie c, b_u, e_{ru} są ustalonymi liczbami rzeczywistymi.

Teraz z (7) otrzymujemy wynik Harville'a z uwagi na oczywiste równości

$$\begin{aligned} \mathbf{y}'\mathbf{I}\mathbf{y} &= \sum_{i,j} \mathbf{y}_{ij}^2, & \mathbf{y}'C_u \mathbf{y} &= \eta_u^2 \sum_{i \in S_u} \bar{y}_i^2, \\ \mathbf{y}'C_{uu} \mathbf{y} &= N_u^2 (\bar{y}_i^*)^2, & \mathbf{y}'C_{ur} \mathbf{y} &= 2N_u N_r \bar{y}_u^* \bar{y}_r^* \end{aligned}$$

dla $u > r$.

Udowodnimy, że estymatory NK są kwadratowo dopuszczalne.

TWIERDZENIE 2. *Każdy estymator NK jest kwadratowo dopuszczalny.*

D o w ó d. Niech $y'Fy$ będzie estymatorem NK dla swojej wartości oczekiwanej $E(y'Fy)$. Z lematu 2 z pracy [4] wynika, że każdy inny nieobciążony estymator funkcji $E(y'Fy)$ jest postaci $y'(F+G)y$, gdzie $G \in \mathcal{A}_1^\perp$. Stąd dla $\sigma = (0, \dots, 0, 1)'$ otrzymujemy $\text{Var}(y'(F+G)y) = 2\langle F, F \rangle + 4\langle F, G \rangle + 2\langle G, G \rangle = 2\langle F, F \rangle + 2\langle G, G \rangle = \text{Var}(y'Fy) + \text{Var}(y'Gy) > \text{Var}(y'Fy)$, co oznacza, że $y'Fy$ jest estymatorem kwadratowo dopuszczalnym.

Prace cytowane

- [1] F. A. Graybill and R. A. Hultquist, *Theorem concerning Eisenhart's Model II* Ann. Math. Statist. 32 (1961), str. 261–269.
 - [2] Harville, *Quadratic unbiased estimation of variance components for the one-way classification*, Biometrika 56 (1969), str. 313–326.
 - [3] C. R. Henderson, *Estimation of variance components*, Biometrics 9 (1951), str. 226–252
 - [4] R. Zmyślony, *Estimation of variance components in random models*, Applicationes Mathematicae 13.4 (1973), str. 521–527.
-