Konfidenčná oblasť parametra strednej hodnoty v zmiešanom lineárnom modeli

Volaufová, J., Witkovský, V., Bognárová, M.
Ústav merania a meracej techniky CEFV SAV
Dúbravská 9, 842 19 Bratislava

Zmiešaný lineárny model je daný následovne /pozri [2] /:

(1) $Y = X\beta + \varepsilon$, $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \sum_{i=1}^{p} \vartheta_i V_i$, kde Y je n - rozmerný náhodný vektor so strednou hodnotou $X\beta$, kde matice X typu $n \times k$ a V_i $i=1,\dots,p$ typu $n \times n$ sú známe. $\beta = (\beta_1, \dots, \beta_k)'$ a $\vartheta = (\vartheta_1, \dots, \vartheta_p)'$ sú neznáme vektorové parametre, $\beta \in R^k$, $\vartheta \in \Theta \subset R^p$, kde Θ obsahuje otvorenú množinu.

V modeli (1) sa vyšetrujú bodové odhady lineárnych funkcionálov parametra ϑ typu $g(\vartheta) = f'\vartheta$ a lineárne funkcionály $h(\beta) = p'\beta$ vektorového parametra β. Odhady $g(\vartheta)$ sa študujú obvykle v triede kvadratických odhadov $Q = \{Y'AY$, A symetrická $\}$, pričom sa požadujú vlastnosti ako je nevychýlenosť, lokálne, resp. globálne minimálna disperzia a invariantnosť vzhľadom na posun v strednej hodnote. Optimalitu odhadu $\widetilde{g(\vartheta)} \in Q$ je treba študovať prirodzene v závislosti od typu rozdelenia vektora Y, resp. ε .

Všeobecne známy je kvadratický odhad $\widehat{g(\vartheta)}$ typu MINQUE, ktorý v roku 1971 zaviedol C.R.Rao / pozri [4] /, pričom $\widehat{g(\vartheta)}$ je invariantný, nevychýlený a za predpokladu, že Y má n - rozmerné normálne rozdelenie má $\widehat{g(\vartheta)}$ vlastnosť lokálnej optimality vzhľadom na disperziu.

Vychádzajúc z odhadu $f'\vartheta$, resp. ϑ_i $i = 1,\dots,p$, lineárne odhady $h(\beta) = p'\beta$ majú vlastnosť optimality vzhľadom na disperziu vždy závisiacu od odhadu $\hat{\vartheta}_i$ $i = 1,\dots,p$.

Prirodzená otázka je, ako vyzerajú intervalové odhady funkcionálu $p'\beta$, resp. konfidenčná oblasť parametra $\beta \in R^k$. Optimálny lineárny bodový odhad $h(\beta)$ je daný vzťahom

(2) $\widehat{h(\beta)} = p'(X'V(\vartheta)^{-1}X)^{-1}X'V(\vartheta)^{-1}Y$.

Využívame predpoklad, že matica X má plnú hodnosť, t.j. $r(X) = k$ a $V(\vartheta) = \sum_{i=1}^{p} \vartheta_i V_i$ je regulárna pre každé $\vartheta \in \Theta$. Označme odhad parametra β založený na odhade parametra ϑ

(3) $\hat{\hat{\beta}} = (X'V(\hat{\vartheta})^{-1}X)^{-1}X'V(\hat{\vartheta})^{-1}Y$, kde $V(\hat{\vartheta}) = \sum_{i=1}^{p} \hat{\vartheta}_i V_i$.

Štatistické vlastnosti odhadu $\hat{\hat{\beta}}$ nie sú dostatočne preskúmané. Niektoré asymptotické vlastnosti sú uvedené v práci [3] / pozri [3] časť 1.5.3./.

Iný prístup je uvedený v práci Kubáček [1]. Úvaha je založená na predpoklade, že máme k dispozícii m nezávislých opakovaní vektora Y. Kvadratický odhad $\hat{\vartheta}$ využíva $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ a $S = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})(Y_i - \bar{Y})'$, pričom odhad $\tilde{\beta} = (X'V(\hat{\vartheta})^{-1}X)^{-1}X'V(\hat{\vartheta})^{-1}\bar{Y}$. V práci [1] je odvodený asymptotický konfidenčný interval pre lineárny funkcionál $G\beta$.

V ďalšom uvažujme $\hat{\vartheta}$ - MINQUE odhad parametra ϑ. Použijeme nasledujúce označenia : $M = I - X(X'X)^{-1}X' = I - XX^+$

$V_0 = V(\vartheta_0) = \sum_{i=1}^{p} \vartheta_{i_0} V_i$, kde $\vartheta_0 \in \Theta$ je pevne zvolený parameter

$S_{(MV_0M)^+}$ je matica, ktorej i, j - ty element je

$$\{S_{(MV_0M)^+}\}_{i,j} = \mathrm{tr}(MV_0M)^+V_i(MV_0M)^+V_j$$

Vychádzajúc z tohto značenia dá sa ukázať, že kovariančná matica MINQUE odhadu $\hat{\vartheta}$ v bode ϑ_0 je

$$(4) \quad \mathrm{var}\,\hat{\vartheta} = 2\,S^{-1}_{(MV_0M)^+}$$

V odhade $\hat{\hat{\beta}}$ zo vzťahu (3) použijeme $\hat{\vartheta}$ - MINQUE odhad.

<u>Lema 1.</u> Nech $L \in R^n$. Nech $Y \sim N_n(X\beta, V(\vartheta))$ a $Y'A_0Y = \hat{\vartheta}$, kde $A_0X = 0$. Potom $\mathrm{cov}(L'Y, Y'A_0Y) = 0$ pre každý vektor $L \in R^n$.

Dôkaz: K dôkazu si stačí uvedomiť skutočnosť: $\mathrm{cov}(L'Y, Y'A_0Y) = 2L'V(\vartheta)A_0X\beta$.

Aproximujme odhad $\hat{\hat{\beta}} = \hat{\hat{\beta}}(Y,\hat{\vartheta})$ lineárnym funkcionálom, t.j.

$$\hat{\hat{\beta}}(Y,\hat{\vartheta}) \approx \hat{\hat{\beta}}(X\beta,\vartheta) + \left[\frac{\partial\hat{\hat{\beta}}(Y,\hat{\vartheta})}{\partial Y'}\right]_{X\beta}(Y - X\beta) + \left[\frac{\partial\hat{\hat{\beta}}(Y,\hat{\vartheta})}{\partial \hat{\vartheta}'}\right]_{\vartheta}(\hat{\vartheta} - \vartheta)$$

, čo po vyjadrení parciálnych derivácií prejde do tvaru

$$(5) \quad \hat{\hat{\beta}}(Y,\hat{\vartheta}) \approx \beta + (X'V(\vartheta)^{-1}X)^{-1}X'V(\vartheta)^{-1}(Y - X\beta) +$$
$$+ (X'V(\vartheta)^{-1}X)^{-1}X'V(\vartheta)^{-1}(-V_1V(\vartheta)^{-1}v, -V_2V(\vartheta)^{-1}v, \ldots, -V_pV(\vartheta)^{-1}v).(\hat{\vartheta} - \vartheta)$$

Použili sme označenie $v = [I - X(X'V(\vartheta)^{-1}X)^{-1}X'V(\vartheta)^{-1}]Y$.

<u>Lema 2.</u> Kovariančná matica vektora $\hat{\hat{\beta}}(Y,\hat{\vartheta})$ v bode ϑ sa približne rovná

$$(6) \quad \mathrm{cov}_{\vartheta}\hat{\hat{\beta}}(Y,\hat{\vartheta}) \approx (X'V(\vartheta)^{-1}X)^{-1} + (X'V(\vartheta)^{-1}X)^{-1}X'V(\vartheta)^{-1}(V_1V(\vartheta)^{-1}v, \ldots, V_pV(\vartheta)^{-1}v).$$
$$.2\,S^{-1}_{(MV_0M)^+}\,.\begin{pmatrix} v'V(\vartheta)^{-1}V_1 \\ \vdots \\ v'V(\vartheta)^{-1}V_p \end{pmatrix}.\,V(\vartheta)^{-1}X(X'V(\vartheta)^{-1}X)^{-1}$$

Dôkaz: Vyplýva zo vzťahu (5) a z lemy 1.

Vzťah (5) udáva len veľmi hrubé priblíženie funkcie $\hat{\hat{\beta}}$ jej linearizáciou, preto aj kovariančná matica v leme 2 je hrubým priblížením ku skutočnej kovariančnej matici vektora $\hat{\hat{\beta}}(Y,\hat{\vartheta})$.

Naším cieľom bolo simuláciou porovnať empirický konfidenčný interval pre parameter β a intervalom, ktorý dostaneme využitím vzťahu (6), tzv. teoretickým intervalovým odhadom parametra β.

Model z ktorého sme vychádzali bol daný maticou $X = (1, 2, 3, 4, 5)'$, $\beta = 3{,}28$; $\vartheta_1 = 1$; $\vartheta_2 = 4$ a maticami $V_1 = \begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \end{pmatrix}$ a $V_2 = \begin{pmatrix} 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}$.

Simulovali sme 200 5 – rozmerných realizácií vektora Y z normálneho rozdelenia, rovnomerného rozdelenia a U – rozdelenia. Okrem uvedených symetrických rozdelení sme použili Burrovo rozdelenie, ktoré má parametre $\vartheta_1 = (0{,}059999)^2$ a $\vartheta_2 = (0{,}365999)^2$ / pozri [5] /.

Empirické 95 % – né intervalové odhady sú zhrnuté následovne:

normálne rozdelenie	$\langle 2{,}743120 \ , \ 3{,}748333 \rangle$
rovnomerné rozdelenie	$\langle 2{,}723232 \ , \ 3{,}804054 \rangle$
U – rozdelenie	$\langle 2{,}784301 \ , \ 3{,}765919 \rangle$
Burrovo rozdelenie	$\langle 3{,}246234 \ , \ 3{,}311091 \rangle$

U teoretických intervalových odhadov sa ukázala veľmi veľká závislosť od odhadov $\hat{\vartheta}$, čo sa prejavilo v tom, že až 30 – 50 % z 200 odhadov neprekrývalo skutočnú strednú hodnotu $\beta = 3{,}28$.

Vzťah (6) nemožno pokladať za dobré priblíženie kovariančnej matice $\hat{\beta}(Y, \hat{\vartheta})$ a v ďalšom je treba využiť jej vyjadrenie pomocou rozvoja do vyšších rádov.

Literatúra

[1] Kubáček, L.: Asymptotical confidence region in a replicated mixed linear model with an estimated covariance matrix. / zaslané do Aplikace matematiky /

[2] Kleffe, J.: Simultaneous estimation of expectation and covariance matrix in linear models. Math. Operationsforsch. Statist., Ser. Statistics, Vol. 9 / 1978 / No.3, 443 – 478.

[3] Humak, K.M.S.: Statistische Methoden der Modellbildung III. Akademie – Verlag Berlin, 1984

[4] Rao, C.R.: Estimation of variance and covariance components – MINQUE theory. J. Mult. Analysis 1, /1971 /, 445 – 456.

[5] Burr, I.W.: Parameters for a general system of distributions to match a grid of α^3 and α^4. Communications in Statistics 2, / 1973 /, 1 – 21 .