Simultánne obojstranné tolerančné intervaly v lineárnom regresnom modeli

Martina Chvosteková

Ústav merania,
Slovenská akadémia vied

5.2. 2010 Králíky

Obsah

Viacrozmerná regresia

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma\boldsymbol{Z} = (\mathbf{1}, \boldsymbol{X}_1)\boldsymbol{\beta} + \sigma\boldsymbol{Z}$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{q-1} x_{i,q-1} + \sigma Z_i, \quad i = 1, \ldots, n$$

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ - $n \times 1$ vektor nezávislých pozorovaní
- $\boldsymbol{X}$ - $n \times q$ - známa matica plánu, $\mathrm{h}(\boldsymbol{X}) = q, \; n > q$
- $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{q-1})^T$, $\sigma > 0$ - neznáme parametre
- $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$, predpokladáme $\boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{I}_n)$

$(\gamma, 1-\alpha)$ nesimultánny tolerančný interval v LRM

- úroveň spoľahlivosti $1-\alpha$, pokrytie γ (coverage alebo content)
- pozorovanie pre pevné $\boldsymbol{x}$ (vysvetľujúca premenná)
  $Y(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta} + \sigma Z = \beta_0 + \beta_1 x_1 + \ldots + \beta_{q-1}x_q + \sigma Z, \quad Z \sim N(0,1)$
- používaný tvar obojstranného tolerančného intervalu pre $Y(\boldsymbol{x})$
  [1] $\langle \boldsymbol{x}^T\hat{\boldsymbol{\beta}} - \lambda(\boldsymbol{x})S, \boldsymbol{x}^T\hat{\boldsymbol{\beta}} + \lambda(\boldsymbol{x})S \rangle$
- pokrytie
  $C(\boldsymbol{x}; \hat{\boldsymbol{\beta}}, S) = P_{Y(\boldsymbol{x})}(\boldsymbol{x}^T\hat{\boldsymbol{\beta}} - \lambda(\boldsymbol{x})S \leq Y(\boldsymbol{x}) \leq \boldsymbol{x}^T\hat{\boldsymbol{\beta}} + \lambda(\boldsymbol{x})S | \hat{\boldsymbol{\beta}}, S)$

$$P_{\hat{\boldsymbol{\beta}},S}(C(\boldsymbol{x}; \hat{\boldsymbol{\beta}}, S) \geq \gamma) = 1-\alpha$$

[1] $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}, \quad S^2 = \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\beta})^T(\boldsymbol{Y}-\boldsymbol{X}\hat{\beta})}{n-q}$

Simultánny obojstranný tolerančný interval v LRM

$$P_{\hat{\boldsymbol{\beta}},S}(C(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) \geq \gamma \;\; \forall \boldsymbol{x}) = 1-\alpha$$

$$P_{\hat{\boldsymbol{\beta}},S}(\min_{\boldsymbol{x}} C(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) \geq \gamma) = 1-\alpha$$

Klasický postup

$$Y(\boldsymbol{x}) \sim N(\boldsymbol{x}^T\beta,\sigma^2), \quad \frac{Y(\boldsymbol{x})-\boldsymbol{x}^T\boldsymbol{\beta}}{\sigma} \sim N(0,1)$$

Pokrytie

$$C(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) = P_{Y(\boldsymbol{x})}(\boldsymbol{x}^T\hat{\boldsymbol{\beta}} - \lambda(\boldsymbol{x})S \leq Y(\boldsymbol{x}) \leq \boldsymbol{x}^T\hat{\boldsymbol{\beta}} + \lambda(\boldsymbol{x})S|\hat{\boldsymbol{\beta}},S)$$
$$= \Phi(\boldsymbol{x}^T\boldsymbol{b} + \lambda(\boldsymbol{x})u) - \Phi(\boldsymbol{x}^T\boldsymbol{b} - \lambda(\boldsymbol{x})u) = C_1(\boldsymbol{x}^T\boldsymbol{b},\lambda(\boldsymbol{x})u)$$

Pivotné premenné

$$\boldsymbol{b} = \frac{\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}}{\sigma} \sim N(0,(\boldsymbol{X}^T\boldsymbol{X})^{-1}), \quad u = \frac{S}{\sigma}, \quad (n-q)u^2 \sim \chi^2_{n-q}$$

Princíp confidence-set prístupu

Simultánne tolerančné intervaly

$$P_{\boldsymbol{b},u}(\min_{\boldsymbol{x}} C_1(\boldsymbol{x}^T\boldsymbol{b}, \lambda(\boldsymbol{x})u) \geq \gamma) = 1 - \alpha$$

G - $(1-\alpha)$-pivotná oblasť

Konfidenčná oblasť pre parametre modelu

$$\{(\boldsymbol{\beta}, \sigma) : (\hat{\boldsymbol{\beta}} - \boldsymbol{b}S/u, S/u) \quad \forall(\boldsymbol{b}, u) \in G\}$$

Tolerančný faktor pre $\boldsymbol{x}$

$$\lambda(\boldsymbol{x}) = \min\{\lambda : C_1(\boldsymbol{x}^T\boldsymbol{b}, \lambda(\boldsymbol{x})u) \geq \gamma \quad \forall(\boldsymbol{b}, u) \in G\}$$

Wilsonova metóda

Wilsonova pivotná oblasť

$$G_W = \{(\boldsymbol{b}, u) : \boldsymbol{b}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} + 2v(u-k)^2 \leq c\}$$

$c = \chi^2_{q+1}(1-\alpha)$, $k = \sqrt{\frac{2v-1}{2v}}$, $v = n - q$

R. A. Fisher (1928), *Statistical Methods for Research Workers*, 2nd Edition, pp. 96-97.

$$(2\chi^2_{n-q})^{1/2} \approx N([2(n-q)-1]^{1/2}, 1)$$

$$u \approx N(\sqrt{(2v-1)/(2v)}, 1/(2v))$$

$(n-q)u^2 \sim \chi^2_{n-q}$, $\boldsymbol{b} \sim N(\boldsymbol{0}, (\boldsymbol{X}^T\boldsymbol{X})^{-1})$

Limamova-Thomasova metóda

Pivotná oblasť

$$G_B = \{(\boldsymbol{b}, u) : \boldsymbol{b}(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} \le u^2 k_1^2 \quad \text{a} \quad u \ge k_2\}.$$

- $(1-\alpha/2)$-konfidenčná oblasť pre β

$$E_1 = \{\beta : (\hat{\beta}-\beta)^T(\boldsymbol{X}^T\boldsymbol{X})(\hat{\beta}-\beta) \le S^2 k_1^2\}, \;\; k_1^2 = qF_{q,n-q}(1-\alpha/2)$$
$$E_1^* = \{(\boldsymbol{b}, u) : \boldsymbol{b}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} \le u^2 k_1^2\}$$

- zhora ohraničenú $(1-\alpha/2)$ konfidenčnú oblasť pre σ

$$E_2 = \{\sigma : 0 < \sigma \le S/k_2\}, \;\; k_2^2 = \chi^2_{n-q}(\alpha/2)/(n-q)$$
$$E_2^* = \{u : k_2 \le u\}$$

Modifikovaná Wilsonova metóda

Wilsonova pivotná oblasť

$$G_W = \{(\boldsymbol{b}, u) : \boldsymbol{b}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} + 2v(u-k)^2 \leq \chi^2_{q+1}(1-\alpha) = c\},\ k = \sqrt{(2v-1)/(2v)}$$

$$|\boldsymbol{x}^T\boldsymbol{b}| \leq \sqrt{c - 2v(u-k)^2}\sqrt{\boldsymbol{x}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}} = A_{\boldsymbol{x}}(u) \quad \text{pre každé} \quad \boldsymbol{x}$$

$$C_1(\boldsymbol{x}^T\boldsymbol{b}, \lambda u) \geq C_1(A_{\boldsymbol{x}}(u), \lambda u) \quad \text{pre } (\boldsymbol{b}, u) \in G_W$$

$$G_{MW} = G_{MW1} \cup G_{MW2}$$

$$G_{MW1} = \{(\boldsymbol{b}, u) : \boldsymbol{b}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} + 2v(u-k)^2 \leq c_m \quad k - \sqrt{c_m/(2v)} \leq u \leq k\}$$

$$G_{MW2} = \{(\boldsymbol{b}, u) : \boldsymbol{b}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} \leq u^2 c_m/k^2 \quad u \geq k\}$$

Konštanta c_m je dopočítaná, aby platilo $P(G_{MW}) = 1 - \alpha$

Testovacia štatistika

Hypotéza

$H_0 : (\boldsymbol{\beta}, \sigma) = (\boldsymbol{\beta}_0, \sigma_0)$ versus $H_1 : (\boldsymbol{\beta}, \sigma) \neq (\boldsymbol{\beta}_0, \sigma_0)$

Testovacia štatistika pre napozorované $\boldsymbol{y}$

$$\mathcal{D}(\boldsymbol{y} \mid \boldsymbol{X}) = \frac{1}{\sigma_0^2}(\boldsymbol{y} - \boldsymbol{X}\beta_0)^T(\boldsymbol{y} - \boldsymbol{X}\beta_0) - n \log\left(\frac{\frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})}{\sigma_0^2}\right) - n,$$

Rozdelenie testovacej štatistiky

$\mathcal{D}(\boldsymbol{Y}) \sim Q_q + Q_{n-q} - n\log(Q_{n-q}) + n(\log(n) - 1)$

$Q_q \sim \chi_q^2$ and $Q_{n-q} \sim \chi_{n-q}^2$

Distribučná funkcia LRT

$$\begin{aligned}\mathcal{F}_{LR}(x) &= P(\mathcal{D}(\boldsymbol{Y}\,|\,n,q) \leq x) \\ &= P(W_q \leq x - W_{n-q} + n\log(W_{n-q}) - n\,(\log(n) - 1)) \\ &= \int_0^\infty \mathcal{F}_{\chi^2_q}\,(x - w_{n-q} + n\log(w_{n-q}) - n(\log(n) - 1)) \\ &\qquad \times f_{\chi^2_{n-q}}(w_{n-q}) \;\; w_{n-q},\end{aligned}$$

$$\mathcal{F}_{LR}(\mathcal{D}_{1-\alpha}) = 1 - \alpha$$

H_0 zamieta na hladine významnosti $\alpha \in (0,1)$ $\quad \mathcal{D}(y) > \lambda_{1-\alpha}$

Kritické hodnoty

Tabuľka: $\alpha = 0.05$

n / q	1	2	3	4	5
10	6.6103	9.2228	12.0691	15.3518	19.3464
20	6.2825	8.4408	10.5656	12.7292	14.9765
30	6.1817	8.2174	10.1695	12.1014	14.0460
40	6.1328	8.1115	9.9864	11.8187	13.6384
50	6.1038	8.0497	9.8809	11.6577	13.4094
60	6.0847	8.0092	9.8122	11.5537	13.2627
70	6.0712	7.9806	9.7640	11.4811	13.1606
80	6.0611	7.9594	9.7282	11.4274	13.0855
90	6.0533	7.9429	9.7006	11.3861	13.0279
100	6.0470	7.9299	9.6788	11.3534	12.9823
∞	5.9915	7.8147	9.4877	11.0705	12.5916

Pomocou presného testu pomerom vierohodnosti pre testovanie nulovej hypotézy

$H_0 : (\boldsymbol{\beta}, \sigma) = (\boldsymbol{\beta}_0, \sigma_0)$ proti alternatíve $H_1 : (\boldsymbol{\beta}, \sigma) \neq (\boldsymbol{\beta}_0, \sigma_0)$

môžeme skonštruovať konfidenčnú oblasť pre všetky parametre lineárneho regresného modelu simultánne. Presná $(1 - \alpha)$-oblasť spoľahlivosti pre parametre β a σ je daná

$$\mathcal{C}_{1-\alpha}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \{(\boldsymbol{\beta}, \sigma) : \mathcal{D}(\boldsymbol{Y} \,|\, \boldsymbol{X}) \leq \mathcal{D}_{1-\alpha}\} .$$

Chvosteková, M., Witkovský, V. (2009), "Exact Likelihood Ratio Test for the Parameters of the Linear Regression Model with Normal Errors." *Measurement Science Review*, Vol. 9, No. 1, pp. 1-8.

Ďakujem za pozornosť!