Cvičenia z ekonometrie 2006/2007

Písomka 2 - riešenie

Pokyny:

- *Podpíšte sa aj na papier so zadaním a odovzdajte ho spolu s riešením.*
- *Môžete používať kalkulačku a ťahák A4.*
- *Neodpisujte...*

---

1. *(4 body, v každej otázke 0.25 b. za odpoveď, 0.75 b. za zdôvodnenie) Rozhodnite, ktoré z nasledujúcich tvrdení sú pravdivé a ktoré nepravdivé. Svoje tvrdenie zdôvodnite.*

   (a) *Ak v modeli $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \varepsilon$ testujeme signifikantnosť regresie, testujeme hypotézu $\alpha_0 = 0$, $\alpha_1 = 0$, $\alpha_2 = 0$.*

   (b) *Ak je submodel lepší ako pôvodný model, má väčší koeficient determinácie.*

   (c) *Rozklad sumy štvorcov TSS = ESS + RSS v modeli $y = \beta x + \varepsilon$ neplatí.*

   (d) *Ak je v modeli heteroskedasticita, MNŠ odhad nie je nevychýlený.*

   Riešenie:

   (a) Nepravdivé, testujeme hypotézu $\alpha_1 = 0$, $\alpha_2 = 0$.

   (b) Nepravdivé, submodel nikdy nemá väčší koeficient determinácie ako pôvodný model.

   (c) Pravdivé, na dôkaz rozkladu sumy štvrocov potrebujeme vlastnosť $\sum e_i = 0$, ktorú tento model nemá:

   $$\begin{aligned} TSS &= \sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \\ &= \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \\ &= RSS + ESS + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \end{aligned}$$

   Posledný sčítanec je

   $$2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 2\sum e_i(\hat{\beta} x_i - \bar{y}) = 2\left[\hat{\beta}\sum e_i x_i - \bar{y}\sum e_i\right] \neq 0,$$

   keďže $\sum x_i e_i = 0$ (nutná podmienka optima), $\sum e_i \neq 0$ (videli sme v prvej písomke).

   (d) Nepravdivé, k nevychýlenosti potrebujeme predpoklad $E[\varepsilon] = 0$, homoskedasticita, resp. heteroskedasticita na to nemá vplyv.

   Poznámky:

   - Model $y = \beta x + \varepsilon$, ani $y = \beta X + \varepsilon$, nemôže byť maticový zápis modelu $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1}$. V maticovom zápise je β k-rozmerný vektor, t.j. jeho rozmer je $k \times 1$, X je matica $n \times k$ (kde n je počet dát). Súčin βX nie je v tomto prípade definovaný.

2. *(3 body) Vyberte si jednu z úloh (a), (b), (c), každá je za 3 body. Ak ich odovzdáte viac, zarátajú sa vám body za tú, za ktorú dostanete viac bodov.*

   (a) *Dokážte, že v modeli $y = \alpha + \beta x + \varepsilon$ sa koeficient determinácie R^2 rovná druhej mocnine výberového korelačného koeficientu $r(x, y)$.*

   (b) *Uvažujme model $Y = X\beta + \varepsilon$, kde $\varepsilon \sim N(0, \sigma^2 I)$ a hypotézu $R\beta = r$. Vyjadrite štatistiku likelihood testu na testovanie tejto hypotézy pomocou RSS a RSS_*.*

(c) *Uvažujme model $y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2\Omega)$, kde Ω je známa diagonálna matica $diag(\sigma_1^2, \dots, \sigma_n^2)$. Dokážte, že odhad zovšeobecnenou metódou najmenších štvorcov je riešením optimalizačnej úlohy tvaru*

$$\sum_{i=1}^{n} w_i \left(y_i - \beta_0 - \beta_1 x_{1i} - \ldots - \beta_k x_{ki}\right)^2 \to \min$$

a určte váhy w_i

Riešenie:

(a) V modeli $y = \alpha + \beta x + \varepsilon$ pre MNŠ odhady platí:

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \ \bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}.$$

Preto

$$\begin{aligned} R^2 &= \frac{ESS}{TSS} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\sum\left[(\hat{\alpha} + \hat{\beta}x_i) - (\hat{\alpha} + \hat{\beta}\bar{x}\right]^2}{\sum(y_i - \bar{y})^2} = \frac{\hat{\beta}^2\sum(x_i - \bar{x})^2}{\sum(y_i - \bar{y})^2} \\ &= \frac{\left[\sum(x_i - \bar{x})(y_i - \bar{y})\right]^2\sum(x_i - \bar{x})^2}{\left[\sum(x_i - \bar{x})^2\right]^2\sum(y_i - \bar{y})^2} = \frac{\left[\sum(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum(x_i - \bar{x})^2\sum(y_i - \bar{y})^2} = [r(x,y)]^2. \end{aligned}$$

(b)
- Model bez reštrikcie: $\hat{beta} = \hat{\beta}_{mns}$, $\hat{\sigma}^2 = \frac{RSS}{n}$.
- Model s reštrikciou: Log-likelihood funkcia je

$$\ln L(\beta, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta),$$

  maximalizujeme ju za podmienky $R\beta = r$. Dostaneme $\hat{\beta}_*$, $\hat{\sigma}^2 = \frac{RSS_*}{n}$
- Likelihood ratio štatistika:

$$\begin{aligned} LR &= -2\left[\ln L(\hat{\beta}_*, \hat{\sigma}_*^2) - lnL(\hat{\beta}, \hat{\sigma}^2)\right] = \\ &= -2\left[\left(-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\frac{RSS_*}{n} - \frac{n}{2}\right) - \left(-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\frac{RSS}{n} - \frac{n}{2}\right)\right] = \\ &= -2\left[-\frac{n}{2}\left(\ln\frac{RSS_*}{n} - \ln\frac{RSS}{n}\right)\right] = n\ln\frac{RSS_*}{RSS}. \end{aligned}$$

(c) Optimalizačnú úlohu môžeme zapísať v maticovom tvare:

$$(y - X\beta)^T W(y - X\beta) \to \min,$$

kde $W = diag(w_1, \ldots, w_n)$. Označme

$$F(\beta) = (y - X\beta)^T W(y - X\beta) = y^T W y - 2y^T W X\beta + \beta^T X^T W X\beta,$$

v bode optima β^* je derivácia podľa β rovná nule:

$$\left.\frac{\partial F}{\partial\beta}\right|_{\beta^*} = -2X^T W y + 2X^T W X\beta^* = 0$$

a teda riešenie tejto optimallizácie je

$$\beta^* = (X^T W X)^{-1}(X^T W y).$$

To sa zhoduje s odhadom zovšeobecnenou metódou najmenších štvorcov

$$\hat{\beta}_{gls} = (X^T\Omega^{-1}X)^{-1}(X^T\Omega^{-1}y)$$

pre $W = \Omega^{-1}$, čiže $w_i = \frac{1}{\sigma_i^2}$.

Poznámky:

- Príklady (a), (b) boli v príkladoch na precvičnie ako príklady 5(a) a 9.
- Príklady (c) je podobný príkladu 2(b) vo vzorovej písomke.

3. *(3 body) Odhadnutím modelu dostaneme*

$$\hat{y} = -6.38 + 10,41x$$

$$(1.93)\ (3.65)$$

kde v zátvorkách sú odhady štandardných odchýlok. Určte koeficient determinácie, ak model bol odhadnutý zo 150 dát.

Riešenie:

- t štatistika na testovanie hypotézy $\beta = 0$ v modeli $y = \alpha + \beta x + \varepsilon$ je

$$t = \frac{\hat{\beta}}{s.\hat{d}.(\hat{\beta})} = \frac{10.41}{3.65}.$$

- F štatistika na testovanie signifikacie tejto regresie je

$$F = \left(\frac{10.41}{3.65}\right)^2.$$

- R^2 vyjadríme pomocou F štatistiky na testovanie signifikanci regresie:

$$\begin{aligned} F &= \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{ESS}{RSS/(n-2)} = (n-2)\frac{ESS}{RSS} \\ R^2 &= \frac{ESS}{TSS} = \frac{ESS}{ESS+RSS} = \frac{\frac{ESS}{RSS}}{\frac{ESS}{RSS}+1} = \frac{\frac{1}{n-2}F}{\frac{1}{n-2}F+1} \end{aligned}$$

a dosadíme:

$$R^2 = \frac{\frac{1}{148}\left(\frac{10.41}{3.65}\right)^2}{\frac{1}{148}\left(\frac{10.41}{3.65}\right)^2 + 1} = 0.052.$$

Poznámky:

- Ide o rovnaký typ príklad, ako bol voliteľný príklad 2(c) v prvej písomke.

4. *(10 bodov) Z 22 dát odhadujeme model $y = \alpha + \beta x + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$. Platí:*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 10, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 20,$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 60 \quad \sum_{i=1}^{n}(y_i - \bar{y})^2 = 100,$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 30.$$

(n je počet dát). Vypočítajte:

(a) odhad parametrov $\hat{\alpha}$, $\hat{\beta}$

(b) koeficient determinácie

(c) hodnotu štatistiky na testovanie hypotézy $H_0 : \beta = 1$ a určte jej rozdelenie za platnosti H_0

(d) rezidálnu sumu štvorcov

Riešenie:

(a) Odhady parametrov sú

$$\begin{aligned}\hat{\beta} &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{30}{60} = frac12, \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} = 20 - \frac{1}{2}10 = 15.\end{aligned}$$

(b) Podľa príkladu 2(a) v tejto písomke (resp. podľa príkadu 5(a) z príkladov na precvičenie) máme

$$R^2 = [r(x,y)]^2 = \left[\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}\right]^2 = \left[\frac{30}{\sqrt{60 \times 100}}\right]^2 = \frac{3}{20}.$$

(c) Testovacia štatistika je

$$t = \frac{\hat{\beta} - 1}{s.\hat{d}.(\hat{\beta})}$$

a za platnosti H_0 má Studentovo $t_{n-k} = t_{22-2} = t_{20}$ rozdelenie. Potrebujeme teda $s.\hat{d}.(\hat{\beta})$:

- V príklade 3 sme pre model $y = \alpha + \beta x + \varepsilon$ odvodili, že $R^2 = \frac{\frac{1}{n-2}F}{\frac{1}{n-2}F+1}$, kde F je štatistika na testovanie signifikancie regresie.
- Z toho: $F = (n-2)\frac{R^2}{1-R^2}$.
- Keďže $F = \left[\frac{\hat{\beta}}{s.\hat{d}.(\hat{\beta})}\right]^2$, dostaneme

$$s.\hat{d}.(\hat{\beta}) = \sqrt{\frac{\hat{\beta}^2}{(n-2)\frac{R^2}{1-R^2}}}.$$

- Dosadíme:

$$t = \frac{\frac{1}{2} - 1}{\sqrt{\frac{\left(\frac{1}{2}\right)^2}{(22-2)\frac{\frac{3}{20}}{1-\frac{3}{20}}}}} = -1.8787.$$

(d) $RSS = TSS(1 - R^2) = 100\left(1 - \frac{3}{20}\right) = 100 - 15 = 85.$