Cvičenia z ekonometrie 2004/2005

Písomka 1 - vzor
Riešenie

Príklad 1 (***2,5 b.***)
Rozhodnite, či sú nasledujúce tvrdenia pravdivé (A - áno, N - nie). Správna odpoveď ***0,5 b.****, nesprávna odpoveď* ***-0,5 b.****, žiadna odpoveď* ***0 b.*** *Odpovede napíšte do nasledujúcej tabuľky:*

tvrdenie	1.	2.	3.	4.	5.
odpoveď					

Uvažujte lineárny regresný model $Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1} + \varepsilon$. *Predpokladajte splnenie klasických predpokladov lineárnej regresie a normalitu náhodnej zložky* ε .

Tvrdenia:

1. *Ak sa P hodnota testu o parametroch rovná* 0,0750*, tak nulová hypotéza sa na hladine významnosti* $\alpha = 0{,}05$ *nezamieta.*
2. *Pri porovnávaní modelu s alternatívnym modelom pomocou Akaikeho informačného kritéria (AIC) je lepší ten, ktorý má väčšiu hodnotu AIC.*
3. *Vždy platí:* $\sum_{i=1}^{n} \varepsilon_i = 0$.
4. *Pridaním vysvetľujúcej premennej do modelu sa* R^2 *môže zmenšiť.*
5. *Odhad každého parametra je náhodná premenná, ktorej stredná hodnota sa rovná skutočnej hodnote parametra.*

Správne odpovede:

1. ÁNO: $\alpha < P$, takže nulová hypotéza sa nezamieta.
2. NIE: lepší je ten model, ktorý má hodnotu AIC menšiu
3. NIE: $\sum_{i=1}^{n} \varepsilon_i$ je náhodná premenná (s nulovou strednou hodnotou, lebo $E[\varepsilon_i] = 0$, ale nemusí sa rovnať nule)
4. NIE: dokazovali sme na cvičení
5. ÁNO: $E[\beta] = E[(X^T X)^{-1} X^T Y] = (X^T X)^{-1} X^T E[Y] = (X^T X)^{-1} X^T X \beta = \beta$, lebo $E[Y] = X\beta$ (keďže $Y = X\beta + \varepsilon$, pričom $E[\varepsilon] = 0$.)

Príklad 2 (***5 b.***)
Uvažujme dáta z domácej úlohy, v ktorej sa skúmal dopyt po ekonomických časopisoch. Použité premenné: CI - náklad, PP - index ceny, DS - rovná sa 1, ak časopis vydáva spoločnosť, inak sa rovná 0, HM - index kvality. Použitím 18 dostupných dát odhadneme dva modely:

$$CI = \alpha_0 + \alpha_1 PP + \alpha_2 DS + \alpha_3 HM + \varepsilon \quad (1)$$

$$CI = \beta_0 + \beta_1 PP + \varepsilon \quad (2)$$

Výstup pre model (1):

	estimate	st. error	t stat.	P value
1	-5417.29	1505.14	-3.5992	0.0029
PP	4874.47	457.137	10.6631	0.0000
DS	572.99	704.054	0.8138	0.4294
HM	414.14	127.409	3.2505	0.0058

ESS	RSS	TSS
4.81021 $.\,10^8$	2.9797 $.\,10^7$	5.10824 $.\,10^8$

1. *Ktoré koeficienty sú signifikantné na hladine významnosti $\alpha = 0{,}05$?*
   $\alpha_0,\ \alpha_1,\ \alpha_3$ (tie, pre ktoré sa hypotéza $\alpha_i = 0$ zamieta na danej hladine významnosti, t.j. P hodnota je menšia ako 0,05).

2. *Vypočítajte koeficient determinácie R^2.*
   $R^2 = \frac{ESS}{TSS} = \frac{4.81021\times 10^8}{5.10824\times 10^8} = 0{,}9417$

3. *Vypočítajte F štatistiku a rozhodnite o signifikancii regresie na hladine významnosti $\alpha = 0{,}05$.*
   $F = \frac{ESS/(k-1)}{RSS/(n-k)}$, pričom $n = 18$ (počet dát) a $k = 4$ (počet odhadovaných parametrov). Teda $F = \frac{4.81021\times 10^8/3}{2.9797\times 10^7/14} = 75.3353$. To je viac ako kritická hodnota $F(k-1, n-k) = F(3, 14)$ rozdelenia pre $\alpha = 0{,}05$, ktorá sa rovná 3.34. Hypotéza $\alpha_1 = \alpha_2 = \alpha_3 = 0$ sa teda zamieta a regresia je signifikantná.

4. *Uvažujme hypotézu, že časopis s rovnakým indexom ceny a rovnakým indexom kvality má rovnaký očakávaný náklad bez ohľadu na vydavateľa. Sformulujte túto hypotézu ako hypotézu o parametroch modelu a testujte ju.*
   Hypotéza: koeficient pri premennej DS je nulový, teda $\alpha_2 = 0$. Táto hypotéza sa testuje pomocou t štatistiky. V tabuľke máme P hodnotu 0.4294, preto hypotézu nezamietame.

5. *Testujte hypotézu, že v modeli (1) platí $\alpha_2 = 0,\ \alpha_3 = 0$.*
   Hypotézu testujeme F štatistikou $F = \frac{(RSS_* - RSS)/q}{RSS/(n-k)}$, kde RSS_* je reziduálna suma štvorcov z regresie s reštrikciou (model (2)), RSS je reziduálna suma štvorcov z regresie bez reštrikcie (model (1)), $n = 18$ je počet dát, $k = 4$ je počet parametrov, $q = 2$ je počet ohraničení daných reštrikciou. Teda $F = \frac{(5.36965\times 10^7 - 2.9797\times 10^7)/2}{2.9797\times 10^7/14} = 5.61454$, čo je viac ako kritická hodnota rozdelenia $F(q, n-k) = F(2, 14)$ pre $\alpha = 0{,}05$, ktorá sa rovná 3.74. Hypotézu preto zamietame.

Model (2):

	estimate	st. error	t stat.	P value
1	-688.976	650.464	-1.059	0.3052
PP	5681.180	486.782	11.6709	0.0000

ESS	RSS	TSS
4.57124 $.\,10^8$	5.36965 $.\,10^7$	5.10821 $.\,10^8$

PRÍKLAD 3 (***2,5 b.***)
Uvažujte lineárny regresný model

$$Y = \alpha + \beta x + \varepsilon.$$

Určte šírku 95% *intervalu spoľahlivosti pre predikciu* $E[Y]$*, ak je daná hodnota* x*. Zistite, pre akú hodnotu* x *je jeho šírka minimálna a svoje tvrdenie dokážte.*

Interval spoľahlivosti je

$$\left(\hat{y} - t_{0.025}(n-2)s\sqrt{c^T(X^TX)^{-1}c}, \hat{y} + t_{0.025}(n-2)s\sqrt{c^T(X^TX)^{-1}c}\right),$$

kde $\hat{y}$ je vyrovnaná hodnota, t.j. $\hat{\alpha} + \hat{\beta}x$, t je kritická hodnota, s je odhad parametra σ, matica $X = \begin{pmatrix} 1 & x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix}$ a vektor $c = (1, x)^T$. Šírka tohto intervalu teda je

$$2t_{0.025}(n-2)s\sqrt{c^T(X^TX)^{-1}c},$$

pričom od bodu x závisí len výraz $c^T(X^TX)^{-1}c$. Šírka intervalu je minimálna pre to x, pre ktoré je minimálny výraz $c^T(X^TX)^{-1}c$.

Platí

$$(X^TX)^{-1} = \frac{1}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix},$$

a preto

$$c^T(X^TX)^{-1}c = \frac{1}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}(1, x) \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} =$$

$$= \frac{1}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}\left(\sum x_i^2 - 2x\sum x_i + nx^2\right) =$$

$$= \frac{1}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}\left(n\left(x^2 - 2\frac{\sum x_i}{n}\right) + \sum x_i^2\right) =$$

$$= \frac{1}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}\left(n(x-\bar{x})^2 - n\frac{\left(\sum x_i\right)^2}{n^2} + \sum x_i^2\right) =$$

$$= \frac{n}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}(x-\bar{x})^2 + \frac{\frac{1}{n}\left(-\left(\sum x_i\right)^2 + n\sum x_i^2\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} =$$

$$= \frac{1}{\left(\sum x_i^2\right) - \frac{1}{n}\left(\sum x_i\right)^2}(x-\bar{x})^2 + \frac{1}{n}.$$

Minimum sa teda nadobúda v bode $x = \bar{x}$.