

CVIČENIA Z EKONOMETRIE  
LETNÝ SEMESTER 2008/2009

PÍ SOMKA 1 - RIEŠENIE

1. Uvažujme regresný model  $Y = X\beta + \varepsilon$ , pričom matica  $X$  má plnú hodnosť a obsahuje jednotkový stĺpec, model odhadujeme z  $n$  dát, vektor parametrov má  $k$  zložiek a vektor  $\varepsilon$  má normálne rozdelenie s nulovou strednou hodnotou a kovariančnou maticou  $\sigma^2 I$ .

(a) Ako sa vypočíta MNSĎ odhad?

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

(b) Aká je kovariančná matica MNSĎ odhadu?

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

(c) Aké je rozdelenie MNSĎ odhadu (typ rozdelenia, parametre)?

normálne rozdelenie  $N(\beta, \sigma^2 (X^T X)^{-1})$ .

(d) Napíšte nevychýlený odhad parametra  $\sigma^2$ .

$$RSS / (n - k)$$

2. Uvažujme regresiu z nasledujúceho výstupu:

```
Dependent Variable: Y
Method: Least Squares
Sample: 1 25
Included observations: 25
Y=C(1)+C(2)*X1+C(3)* X2
```

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	224.5178	501.3519	0.447825	0.6587
C(2)	4.068341	2.784569	1.461031	0.1581
C(3)	0.001687	0.003059	0.551662	0.5867

R-squared	0.817364	Mean dependent var	1809.625
Adjusted R-squared	0.800761	S.D. dependent var	1212.295
S.E. of regression	541.1224	Akaike info criterion	15.53733
Sum squared resid	6441896	Schwarz criterion	15.68360
Log likelihood	-191.2167	Durbin-Watson stat	1.930395

(a) Ktoré parametre sú signifikantné na hladine významnosti 5 percent?

Ani jeden.

(b) Ktoré parametre sú signifikantné na hladine významnosti 10 percent?

Ani jeden.

(c) Prepokladajme, že vynecháme premennú  $X_1$ . Čo sa stane s upraveným koeficientom determinácie?

Zníži sa.

(d) Testujte hypotézu, že koeficient pri premennej  $X_1$  sa rovná 2,5.

Testovacia štatistika:  $(4,068341 - 2,5)/2,784569 = 0,5632$

Kritická hodnota: 2,074 (22 stupňov voľnosti)

Záver: hypotézu nezamietane

3. Doplňte vynechané hodnoty:

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	338.1415	139.3245	2.427007	0.0259
X1	0.752529		8.636284	0.0000
R-squared		Mean dependent var		1366.245
Adjusted R-squared		S.D. dependent var		714.5984
S.E. of regression		Akaike info criterion		14.49227
Sum squared resid	1886286.	Schwarz criterion		14.59184
Log likelihood	-142.9227	F-statistic		
Durbin-Watson stat	1.135521	Prob(F-statistic)		0.000000

• Pri testovaní signifikácie parametrov platí pre t-štatistiku vzťah  $t - Statistic = \frac{Coefficient}{Std.Error}$ , preto  $Std.Error = \frac{Coefficient}{t-Statistic}$ . Vynechaná štandardná odchýlka preto je  $Std.Error = 0.752529/8.636284 = 0.087136$ .

• T-štatistika 8.636384 testuje rovnakú hypotézu ako vynechaná F-štatistika, platí  $F = t^2$ , takže  $F - statistic = 8.636384^2 = 74.58540$

• F štatistika sa dá vyjadriť pomocou koeficientu determinácie ako  $f = \frac{R^2}{1-R^2} \frac{n-k}{n-1}$ , kde  $n$  je počet dát a  $k$  je počet parametrov modelu. Z tohoto vzťahu sa dá vypočítať  $R^2 = \frac{1}{\frac{1}{F} \frac{n-k}{k-1} + 1}$ . V našom prípade  $n = 20$  (20 dát),  $k = 2$  (2 parametre). Dostaneme  $R - squared = 0.805585$ .

Inou možnosťou výpočtu koeficienta determinácie (bez využitia F-štatistiky) je použiť reziduálnu sumu štvorcov (*Sum squared resid*) a celkovú sumu štvorcov (dá sa určiť zo štandardnej odchýlky *S.D. dependent var*).

• Dosadíme  $R^2$  do vzťahu  $\bar{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2)$ , dostaneme *Adjusted R - squared* = 0.794784.

• *S.E. of regression* je odhad parametra  $\sigma$ , platí  $\hat{\sigma} = \sqrt{RSS/(n-k)}$ . Reziduálna suma štvorcov je v riadku *Sum squared resid*, takže *S.E. of regression* =  $\sqrt{1886286/18} = 323.7185$ .

4. Z 30 dát sme odhadli dva modely:  $Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \varepsilon_1$   $Y = \beta_0 + \beta_1 x_1 + \varepsilon_2$   
 V prvom modeli bol upravený koeficient determinácie 0.77, v druhom 0.76. Testujte hypotézu, že v prvom modeli je  $\alpha_2 = 0$ .

Druhý model je model s reštrikciou, F-štatistiku na testovanie tejto reštrikcie môžeme vyjadriť pomocou koeficientov determinácie  $R_1^2$  (z modelu bez reštrikcie) a  $R_2^2$  (z modelu z reštrikciou):

$$F = \frac{(R_1^2 - R_2^2)/q}{(1 - R_1^2)/(n - k)},$$

kde  $n$  je počet dát (v našom prípade 30),  $k$  je počet parametrov (v modeli, v ktorom testujeme hypotézu - v našom prípade teda 3),  $q$  je počet ohraničení v nulovej hypotéze (v našom prípade 1). My máme k dispozícii upravené koeficienty determinácie, preto z nich musíme koeficienty determinácie najskôr vyjadriť:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2) \Rightarrow R^2 = 1 - \frac{n-k}{n-1}(1 - \bar{R}^2)$$

a teda  $\bar{R}_1^2 = 0.77, n = 30, k = 3 \Rightarrow R_1^2 = 0.7859, \bar{R}_2^2 = 0.76, n = 30, k = 2 \Rightarrow R_2^2 = 0.7683$ . Dosadíme do F-štatistiky a dostaneme  $F = 2.2195$ .

K dispozícii máme kritické hodnoty t-rozdelenia. Tá istá hypotéza sa dá testovať t-štatistikou a platí  $F = t^2$ . Preto

$$|t| = \sqrt{F} = \sqrt{2.2195} < t_{crit., n-k} = t_{crit., 28} = 2.048,$$

a teda hypotézu nezamietame.

#### 5. Z 22 dát odhadujeme model

$$y = \alpha + \beta x + \varepsilon,$$

pričom máme k dispozícii nasledovné informácie:

$$\bar{y} = 20, \sum (y_i - \bar{y})^2 = 100, \bar{x} = 10, \sum (y_i - \bar{y})(x_i - \bar{x}) = 30, \sum (x_i - \bar{x})^2 = 60$$

- (a) *Testujte signifikanciu parametrov. Napíšte hypotézu, ktorú testujete, použitú štatistiku a jej hodnotu a záver (signifikantný/nesignifikantný).*

Odhady parametrov sú:

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

Odhad kovariančnej matice MNŠ odhadu je  $s^2(X^T X)^{-1}$ , kde  $s^2 = RSS/(n - k)$  a

$$(X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}.$$

Keďže  $\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$ , tak  $\sum x_i^2 = 2260$ , a teda

$$(X^T X)^{-1} = \frac{1}{22 \times 2260 - (22 \times 10)^2} \begin{pmatrix} 2260 & -22 \times 10 \\ -22 \times 10 & 22 \end{pmatrix}.$$

Potrebuje ešte RSS. Vieme, že

$$R^2 = r^2 = \frac{900}{100 \times 60} = 0.15$$

a preto

$$R^2 = 1 - \frac{RSS}{TSS} \Rightarrow RSS = (1 - R^2)TSS = (1 - R^2) \sum (y_i - \bar{y})^2 = 85.$$

Môžeme už testovať signifikanciu parametrov:

- Pri testovaní signifikancie parametra  $\alpha$  testujeme hypotézu  $H_0 : \alpha = 0$ . Testovacia štatistika:

$$t = \frac{\hat{\alpha}}{\sqrt{\widehat{Var}(\hat{\alpha})}} = \frac{\hat{\alpha}}{\sqrt{\sigma^2(X^T x)_{11}^{-1}}} = \frac{15}{\sqrt{\frac{85}{20} \frac{1}{22 \times 2260 - (22 \times 10)^2} 2260}} = 5.5607.$$

Za platnosti novej hypotézy má táto štatistika Studentovo rozdelenie  $t_{n-k} = t_{20}$ , a preto jej absolútnu hodnotu porovnávame s kritickou hodnotou tohto rozdelenia, čo je 2.086. Je väčšia ako táto kritická hodnota, preto nulovú hypotézu zamietame a parameter je signifikantný.

- Pri testovaní signifikancie parametra  $\beta$  testujeme hypotézu  $H_0 : \beta = 0$ . Testovacia štatistika:

$$t = \frac{\hat{\beta}}{\sqrt{\widehat{Var}(\hat{\beta})}} = \frac{\hat{\beta}}{\sqrt{\sigma^2(X^T x)_{22}^{-1}}} = \frac{1/2}{\sqrt{\frac{85}{20} \frac{1}{22 \times 2260 - (22 \times 10)^2} 22}} = 1.8787.$$

Za platnosti novej hypotézy má táto štatistika Studentovo rozdelenie  $t_{n-k} = t_{20}$ , a preto jej absolútnu hodnotu porovnávame s kritickou hodnotou tohto rozdelenia, čo je 2.086. Je menšia ako táto kritická hodnota, preto nulovú hypotézu nezamietame a parameter nie je signifikantný.

- (b) *Aká hypotéza sa testuje pri testovaní signifikancie regresie? Je táto regresia signifikantná? Zdôvodnite.*

Testuje sa hypotéza  $\beta = 0$ , čo je rovnaká hypotéza ako pri testovaní signifikancie parametra  $\beta$  v predchádzajúcej úlohe. Túto hypotézu sme nezamietli, takže hypotéza nie je signifikantná.

- (c) *Vypočítajte koeficient determinácie, upravený koeficient determinácie, Akaikeho a Schwarzovo informačné kritérium.*

Už sme vypočítali  $RSS = 85$ ,  $R^2 = 0.15$ . Z toho dostaneme ostanté požadované hodnoty:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2) = 1 - 0.85 \frac{21}{20} = 0.1075,$$

$$AIC = 1 + \ln(2\pi) + \ln\left(\frac{RSS}{n}\right) + \frac{2k}{n} = 1 + \ln(2\pi) + \ln\left(\frac{85}{22}\right) + \frac{4}{22} = 4.3713,$$

$$SIC = 1 + \ln(2\pi) + \ln\left(\frac{RSS}{n}\right) + \frac{k \ln n}{n} = 1 + \ln(2\pi) + \ln\left(\frac{85}{22}\right) + \frac{2 \ln(22)}{22} = 4.4705.$$