Cvičenia z ekonometrie 2005/2006

Písomka 1

Podpíšte sa aj na tento paper so zadaniami a odovzdajte ho spolu s riešeniami.

1. **Príklad 1:** *[0,5 b. za každú z otázok (a)-(j), spolu 5 b.]*
   V úlohách (d)-(j) uvažujeme regresiu z obrázku 1. V úlohe (j) použite výstup z obrázku 2.

Dependent Variable: Y9
Method: Least Squares
Date: 04/13/06 Time: 22:12
Sample: 1 65
Included observations: 65

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1	16.13952	6.211975	2.598130	0.0117
X	0.033667	5.027141	0.006697	0.9947
X^2	12.05221	3.854306	3.126948	0.0027
X^3	-0.287895	0.682737	-0.421678	0.6747

R-squared	0.928670	Mean dependent var	104.7032
Adjusted R-squared	0.925162	S.D. dependent var	104.2532
S.E. of regression	28.52008	Akaike info criterion	9.598657
Sum squared resid	49617.10	Schwarz criterion	9.732466
Log likelihood	-307.9564	F-statistic	264.7265
Durbin-Watson stat	2.552180	Prob(F-statistic)	0.000000

Obr. 1: Regresia - príklad 1.

Null Hypothesis:	C(1)+C(2)+C(3)+C(4)=25 C(3)-C(4)=15		
F-statistic	0.753607	Probability	0.474997

Obr. 2: Testovanie hypotézy - k príkladu 1.

(a) Uvažujme model $Y = X\beta + \varepsilon$ a odhad parametra β metódou najmenších štvorcov $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$. Predpokladajme, že vektor ε má nulovú strednú hodnotu a kovariančnú maticu $\sigma^2 I$. Aká je kovariančná matica odhadu? Ako odhadujeme túto kovariančnú maticu?

(b) Aké je pravdepodobnostné rozdelenie odhadu, ak o vektore ε predpokladáme, že má normálne rozdelenie $N(0, \sigma^2 I)$? Kde potrebujeme tento predpoklad?

(c) Akú hypotézu testujeme pri testovaní signifikancie parametra? Kedy je parameter signifikantný: ak túto hypotézu zamietame alebo ak ju nezamietame?

(d) V regresii na obrázku 1 odhadujeme model $Y = C(1) + C(2)X + C(3)X^2 + C(4)X^3 + \varepsilon$. Ktoré parametre sú signifikantné na hladine významnosti $0,05$?

(e) Nájdite 95% interval spoľahlivosti pre parameter $C(2)$.

(f) Testujte hypotézu $C(2) = 10$ na hladine významnosti 0,05.

(g) Budeme testovať hypotézu

$$C(1) + C(2) + C(3) + C(4) = 25,\ C(3) - C(4) = 15.$$

Zapíšte ju v maticovom tvare $R\beta = r$, kde β je vektor parametrov $(C(1), C(2), C(3), C(4))^T$. Ako vypočítame štatistiku, ktorou túto hypotézu testujeme?

(h) Aké je rozdelenie tejto štatistiky, ak platí nulová hypotéza?

(i) Aká je kritická oblasť testu, t.j. pre aké hodnoty štatistiky nulovú hypotézu zamietame?

(j) Na obrázku 2 je výstup testovania tejto hypotézy zo softvéru - hodnota F štatistiky a P hodnota. Rozhodnite, či hypotézu zamietame alebo nezamietame

- na hladine významnosti $\alpha = 0,05$,
- na hladine významnosti $\alpha = 0,01$.

2. **Príklad 2:** *[1 b. za každú hodnotu, spolu 4 b.]*
Doplňte vynechané hodnoty v regresii na obrázku 3.

Dependent Variable: Y9
Method: Least Squares
Date: 04/13/06 Time: 15:32
Included observations: [ ]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1	2.174009	9.791746	0.222025	0.8251
X	-53.43271	[ ]	[ ]	0.0000

R-squared	0.787117	Mean dependent var	111.7111
Adjusted R-squared	[ ]	S.D. dependent var	105.1810
S.E. of regression	48.94626	Akaike info criterion	10.65209
Sum squared resid	138952.7	Schwarz criterion	10.72190
Log likelihood	-317.5626	F-statistic	214.4506
Durbin-Watson stat	0.824331	Prob(F-statistic)	0.000000

Obr. 3: Príklad 2.

3. **Príklad 3:** *[3 b.]*
Uvažujme model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

kde

$$x_i = -1 + (i-1)\,0,1 \text{ pre } i = 1, \ldots, 31.$$

Vektor ε má normálne rozdelenie $N(0, \sigma^2 I)$. Zistite, či odhady $\hat{\beta}_0$ a $\hat{\beta}_1$ sú nekorelované a svoje tvrdenie dokážte.

4. **Príklad 4:** *[3 b.]*
V knihe *Atkinson a kol.: Psychologie. Portál, 2003* sa píše:

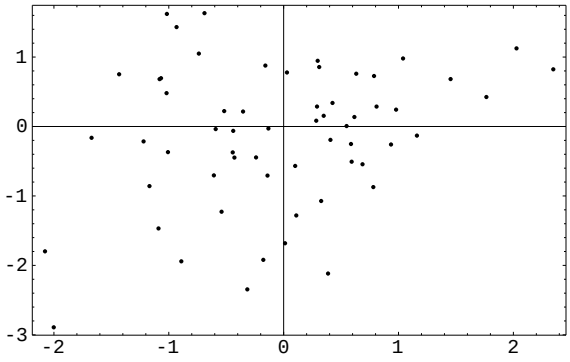
> V psychologickém výzkumu se hodnota korelačního koeficientu 0,60 a vyšší považuje za dosti vysokou. Korelace mezi 0,20 a 0,60 mají praktickou a teoretickou hodnotu a jsou užitečné ve vytváření předpokladů. Korelace mezi 0 a 0,20 musí být posuzovány velmi opatrně a při vytváření předpovědí jsou užitečné jen minimálně.

Hodnota korelácie 0,2 sa môže zdať príliš malá na to, aby sme z nej vyvodzovali nejakú závislosť. Na obrázku 4 je 60 bodov (x_i, y_i), pričom medzi vektormi $x = (x_1, \dots x_{60})$ a $y = (y_1, \dots y_{60})$ je korelácia 0,225.

Predpokladajme teraz, že z dát z obrázku 4 odhadneme regresný model

$$y_i = a + bx_i + \varepsilon_i.$$

Bude táto regresia signifikantná na hladine významnosti 0,05? Svoje tvrdenie dokážte.



Obr. 4: Príklad 4.

Cvičenia z ekonometrie 2005/2006

Písomka 1 - riešenie

1. (a) Kovariančná matica odhadu je $\sigma^2(X^TX)^{-1}$, odhadujeme ju maticou $s^2(X^TX)^{-1}$, kde $s^2 = \frac{RSS}{n-k}$ je odhad parametra σ^2.

   (b) Rozdelenie odhadu je $N\left(\beta, \sigma^2(X^TX)^{-1}\right)$. Predpoklad o normalite potrebujeme pri testovaní hypotéz pomocou t a F štatistík. Bez tohto predpokladu nemajú tieto štatistiky Studentovo, resp. Fisherovo rozdelenie.

   (c) Testujeme hypotézu $H_0 : \beta_i = 0$. Parameter je signifikantný, ak túto hypotézu zamietame.

   (d) Signifikantné sú parametre $C(1)$ a $C(3)$ (tie, pre ktoré je P hodnota dosiahnutá pri testovaní hypotézy $C(i) = 0$ menšia ako hladina významnosti 0,05).

   (e) Hľadaný interval spoľahlivosti je

   $$IS = \left(\hat{C}(2) - t_{krit}\ \hat{sd}(\hat{C}(2)),\ \hat{C}(2) + t_{krit}\ \hat{sd}(\hat{C}(2))\right),$$

   kde

   - $\hat{C}(2)$ je odhad parametra $C(2)$, v našom prípade 0,033667,
   - $\hat{sd}(\hat{C}(2))$ je odhad štandardnej odchýlky $\hat{C}(2)$, v našom prípade 5,027141,
   - t_{krit} je kritická hodnota Studentovho rozdelenia s $n-k$ stupňami voľnosti, v našom prípade $n - k = 65 - 4 = 61$ a príslušná kritická hodnota je (z tabuliek) 1,99962.

   Dosadením dostaneme

   $$IS = (-10,019;\ 10.086)\,.$$

   (f) Testovacia štatistika

   $$t = \frac{\hat{C}(2) - 5}{\hat{sd}(\hat{C}(2))}$$

   má za platnosti nulovej hypotézy Studentovo rozdelenie s $n - k$, t.j. v našom prípade 61 stupňami voľnosti. Dosadením dostaneme hodnotu štatistiky, rovná sa -0,9879. Hypotézu zamietame, ak je t-štatistika v absolútnej hodnote väčšia ako kritická hodnota. Pretože $|-0,9879| < 1.99962$, hypotézu nezamietame.

   Iné riešenie: Hodnota 5 patrí do intervalu spoľahlivosti pre parameter $C(2)$ vypočítaného v predchádzajúcej úlohe, preto hypotézu $C(2) = 5$ nezamietame

   (g) Matica R a vektor r v zápise $R\beta = r$ sú:

   $$R = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix},\ \ r = \begin{pmatrix} 25 \\ 15 \end{pmatrix},$$

   Hypotézu testujeme štatistikou

   $$F = \frac{(R\beta - r)^T \left(R(X^TX)^{-1}R^T\right)^{-1} (R\beta - r)/q}{e^Te/(n-k)},$$

   kde n je počet dát (65), k je počet parametrov v modeli (4) a q je počet reštrikcií v hypotéze (2).

   (h) Fisherovo rozdelenie s q a $n - k$ stupňami voľnosti, t.j. $F(2, 61)$.

   (i) Hypotézu zamietame, ak je hodnota štatistiky väčšia ako kritická hodnota $F(q, n-k)$ rozdelenia.

(j) P hodnota sa rovná 0,475. To je viac ako 0,05 aj ako 0,01, preto hypotézu nezamietame ani na jednej z týchto hladín významnosti.

2. • Keďže v modeli je jedna vysvetľujúca premenná, F-štatistika z testu signifikancie regresie je druhou mocninou t-štatistiky z testu signifikancie parametra pri X, ktorú máme vypočítať. To znamená, že $t^2 = 214,4506$. Ďalej vieme, že $t = \frac{coeff.}{std.error}$. Pretože odhad parametra je záporný ($coeff. = -53.43271$) a odhad štandardnej odchýlky musí byť kladný, t-štatistika je záporná. Teda

$$t = -\sqrt{F} = -\sqrt{214.4506} = -14.6441.$$

• Teraz môžeme vypočítať štandardnú odchýlku odhadu:

$$t = \frac{coeff.}{std.error} \Rightarrow std.error = \frac{coeff.}{t} = \frac{-53,4321}{-14,6441} = 3,64875$$

• Vzťah medzi F-štatistikou z testu signifikancie regresie a koeficientom determinácie je

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{R^2}{1-R^2}\frac{n-k}{k-1}.$$

Z tejto rovnosti poznáme všetky hodnoty okrem n, ktoré teda môžeme vyjadriť:

$$n = F(k-1)\frac{1-R^2}{R^2} + k$$

a dosadiť:

$$n = 214,4506\,(2-1)\,\frac{1-0,787117}{0,787117} + 2 = 60,0001.$$

Počet pozorovaní, z ktorých bol model odhadnutý je teda 60.

• Máme už počet dát n, takže z R^2 môžeme vypočítať upravené R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1-R^2) = 1 - \frac{60-1}{60-2}\,(1-0,787117) = 0,783447.$$

3. V modeli $Y = \beta_0 + \beta_1 x + \varepsilon$ sú odhady nekorelované práve vtedy, keď $\sum x_i = 0$.[1]

Dôkaz tejto podmienky: Kovariančná matica odhadu je $\sigma^2(X^TX)^{-1}$. Z toho, že v tomto modeli sa matica X rovná

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},$$

dostaneme

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

To znamená, že

$$cov\left(\hat{\beta}_0, \hat{\beta}_1\right) = \sigma^2 \frac{1}{det(X^TX)}\left(-\sum x_i\right).$$

Náhodné premenné sú nekorelované práve vtedy, keď je ich kovariancia nulová, čo je v tomto prípade ekvivalentné s rovnosťou $\sum_{i=1}^{n} x_i = 0$.

Riešenie príkladu: Podmienka $\sum x_i = 0$ je ekvivalentná s tým, že priemer hodnôt x_i je nulový. Máme

$$x_i = -1 + (i-1)\,0,1 \text{ pre } i = 1, \ldots, 31,$$

t.j. body x_i sú rovnomerne rozložené na intervale $[-1, 2]$. Ich priemer sa teda rovná stredu tohto intervalu, čo je 0,5. To je nenulové číslo, a teda odhady $\hat{\beta}_0$ a $\hat{\beta}_1$ sú korelované.

[1] Odvodiť nutnú a postačujúcu podmienku nekorelovanosti odhadov v tomto modeli bol tretí príklad v druhej časti príkladov na precvičenie.

4. Vieme, že v modeli $y_i = a + bx_i + \varepsilon_i$. sa koeficient determinácie R^2 rovná druhej mocnine výberového koeficientu korelácie medzi $x = (x_1, \ldots, x_n)$ a $y = (y_1, \ldots, y_n)$, t.j. $R^2 = r^2$. Pomocou koeficientu determinácie vypočítame F štatistiku na testovanie signifikancie regresie: $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$.

   Dosadíme $n = 60$, $k = 2$, $R^2 = r^2 = 0,225^2$, dostaneme

$$F = \frac{0,225^2/(2-1)}{(1-0,225^2)/(60-2)} = 3,09282.$$

   Pri testovaní signifikancie regresie porovnávame hodnotu F-štatistiky s kritickou hodnotou rozdelenia $F(k-1, n-k)$. Ak je väčšia ako kritická hodnota, regresia je signifikantná (lebo zamietame nulovú hypotézu, že všetky koeficienty okrem absoútneho člena sú nulové). V tomto prípade potrbujeme kritickú hodnotu rozdelenia $F(1, 58)$, čo je $4,00686$. Hodnota F-štatistiky je menšia, preto regresia nie je signifikantná.