Časť I

1. **Odhad metódou najmenších štvorcov**
   Odhadujeme model
   $$Y = X\beta + \varepsilon,$$
   pričom matica X má rozmer $n \times k$ a má lineárne nezávislé stĺpce, vektor ε má nulovú strednú hodnotu a kovariančnú maticu $\sigma^2 I$.

   (a) Ako vypočítame odhad parametra β metódou najmenších štvorcov?

   (b) Dokážte jeho nevychýlenosť.

   (c) Aká je jeho kovariančná matica? Ako odhadujeme túto kovariančnú maticu?

   (d) Aké je pravdepodobnostné rozdelenie odhadu, ak o vektore ε predpokladáme, že má normálne rozdelenie? Kde sa využíva tento predpoklad?

2. **Signifikancia parametrov**

   (a) Sformulujte nulovú hypotézu, ktorou testujeme signifikantnosť parametrov. Kedy je parameter signifikantný - ak H_0 zamietame alebo ak H_0 nezamietame?

   (b) Akou štatistikou túto hypotézu testujeme - ako sa vypočíta a aké má rozdelenie, ak platí H_0?

   (c) Aká je kritická oblasť testu, t.j. kedy H_0 zamietame?

   (d) Odhadujeme model
   $$Y = C(1) + C(2)x + C(3)x^2 + C(4)x^3 + \varepsilon,$$
   pričom ε má $N(0, \sigma^2 I)$ rozdelenie. Dostaneme:

   Dependent Variable: Y10
   Method: Least Squares
   Date: 03/18/06 Time: 16:08
   Sample: 1 101
   Included observations: 101
   Y10=C(1)+C(2)*X+C(3)*X^2+C(4)*X^3

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	144.8358	15.22934	9.510310	0.0000
C(2)	28.00946	8.708314	3.216405	0.0018
C(3)	-44.53205	1.335532	-33.34405	0.0000
C(4)	0.102766	0.521756	0.196961	0.8443

R-squared	0.924280	Mean dependent var	-233.6867
Adjusted R-squared	0.921938	S.D. dependent var	365.1702
S.E. of regression	102.0270	Akaike info criterion	12.12715
Sum squared resid	1009722.	Schwarz criterion	12.23072
Log likelihood	-608.4211	F-statistic	394.6779
Durbin-Watson stat	2.152210	Prob(F-statistic)	0.000000

   Ktoré parametre sú signifikantné na hladine významnosti $\alpha = 0,05$?

3. **Signifikancia regresie**

   (a) Sformulujte nulovú hypotézu, ktorou testujeme signifikantnosť regresie. Kedy je regresia signifikantná - ak H_0 zamietame alebo ak H_0 nezamietame?

   (b) Akou štatistikou túto hypotézu testujeme - ako sa vypočíta a aké má rozdelenie, ak platí H_0?

   (c) Aká je kritická oblasť testu, t.j. kedy H_0 zamietame?

   (d) Je regresia z príkladu 2(d) signifikantná?

4. **Intervaly spoľahlivosti (intervalové odhady)**
   Vypočítajte 95% interval spoľahlivosti pre každý parameter z regresie z príkladu 2(d).

5. **Testovanie jednej lineárnej hypotézy o parametroch**
   Predpokladajme, že v modeli $Y = X\beta + \varepsilon$ testujeme hypotézu $c^T\beta = r$.

   (a) Akou štatistikou túto hypotézu testujeme - ako sa vypočíta a aké má rozdelenie, ak platí H_0?

   (b) Aká je kritická oblasť testu, t.j. kedy H_0 zamietame?

   (c) V modeli z príkladu 2(d) testujte hypotézu, že $C(2) = 20$.

6. **Testovanie lineárnych hypotéz o parametroch**
   Predpokladajme, že v modeli $Y = X\beta + \varepsilon$ testujeme hypotézu $R\beta = r$.

   (a) Akou štatistikou túto hypotézu testujeme - ako sa vypočíta a aké má rozdelenie, ak platí H_0?

   (b) Aká je kritická oblasť testu, t.j. kedy H_0 zamietame?

   (c) Uveďte konkrétny príklad modelu a hypotézy o jeho parametroch tohto typu, spolu s ekonomickou interpretácioiu hypotézy.

7. **Odhad metódou maximálnej vierohodnosti**
   Odhadujeme model

   $$Y = X\beta + \varepsilon,$$

   pričom matica X má rozmer $n \times k$ a má lineárne nezávislé stĺpce, vektor ε má normálne rozdelenie, nulovú strednú hodnotu a kovariančnú maticu $\sigma^2 I$.

   (a) Ako vypočítame odhad parametra β metódou maximálnej vierohodnosti?

   (b) Ako vypočítame odhad parametra σ^2 metódou maximálnej vierohodnosti?

   (c) Ktoré z odhadov z častí (a) a (b) sú nevychýlené? Svoju odpoveď zdôvodnite.

Časť II

1. Doplňte vynechané hodnoty vo výstupe z regresie na obrázku 1.[1]

2. Doplňte vynechané hodnoty vo výstupe z regresie na obrázku 2.

3. Rozhodnite, či sú nasledujúce tvrdenia pravdivé. Svoje tvrdenie zdôvodnite.

   (a) Ak do regresného modelu prodáme ďalšiu vysvetľujúcu premennú, koeficient determinácie R^2 sa nemôže zmenšiť.

   (b) Ak do regresného modelu prodáme ďalšiu vysvetľujúcu premennú, upravený koeficient determinácie $\bar{R}^2$ sa nemôže zmenšiť.

[1] V príkladoch tohto typu môžu byť výpočty, v ktorých treba poznať význam hodnôt *Included observations, Coefficient, Std. Error, t-statistic, R squared, Adjusted R squared, F statistics* a vzťahy medzi nimi

Dependent Variable: Y43
Method: Least Squares
Date: 04/07/06 Time: 12:59
Sample: 1 50
Included observations: 50
Y43=C(1)+C(2)*X

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	330.4089		9.076391	0.0000
C(2)		12.42420	19.64899	0.0000

R-squared		Mean dependent var	-292.1046
Adjusted R-squared		S.D. dependent var	377.3411
S.E. of regression	126.7786	Akaike info criterion	12.56194
Sum squared resid	771494.9	Schwarz criterion	12.63842
Log likelihood	-312.0485	F-statistic	
Durbin-Watson stat	1.354564	Prob(F-statistic)	0.000000

Obr. 1: Doplňte vynechané hodnoty

(c) Parameter je signifikantný na hladine významnosti α práve vtedy, keď $(1-\alpha)\%$ interval spoľahlivosti pre tento parameter neobsahuje nulu.

(d) V modeli $Y = C(1) + C(2)x + \varepsilon$ je signifikancia regresie je ekvivalentná so signifikantnosťou parametra $C(2)$.

4. Dokážte, že v modeli

$$Y = \beta_0 + \beta_1 x_1 + \ldots \beta_{k-1} x_{k-1} + \varepsilon,$$

ktorý odhadneme MNŠ, je súčet rezíduí nulový.

5. (a) Dokážte, že v modeli

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

sa koeficient determinácie R^2 rovná druhej mocnine výberového korelačného koeficienta medzi x_i a Y_i.

(b) Dokážte, že v modeli

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$

sa koeficient determinácie R^2 rovná druhej mocnine výberového korelačného koeficienta medzi skutočnými hodnotami Y_i a odhadnutými hodnotami $\hat{Y}_i$.

6. Z tých istých dát (je ich 60) sme odhadli dva modely:

$$Y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \varepsilon, \quad (1)$$
$$Y = b_0 + b_1 x^2 + \eta. \quad (2)$$

Koeficient determinácie v prvej regresii je $R^2 = 0.9412$. V druhej regresii je $R^2 = 0.9394$. Testujte hypotézu, že v modeli (1) je $a_1 = 0$, $a_3 = 0$.

7. Z tých istých dát (je ich 110) sme odhadli dva modely:

$$Y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \varepsilon, \quad (3)$$
$$Y = b_0 + b_1 x^2 + \eta. \quad (4)$$

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. 2: Doplňte vynechané hodnoty

Reziduálna suma štvorcov v prvej regresii je $RSS = 204.54$, v druhej regresii je $RSS = 224.95$. Testujte hypotézu, že v modeli (3) je $a_1 = 0$, $a_3 = 0$.

8. Uvažujme dva regresné modely pre tie isté dáta $y_1, \ldots, y_n$. Vieme, že odhady variancie náhodných chýb ε_{1i} (v prvom modeli) a ε_{2i} (v druhom modeli) sú

$$\hat{\sigma}_1^2 = \frac{RSS_1}{n-k_1}, \quad \hat{\sigma}_2^2 = \frac{RSS_2}{n-k_2}.$$

Ukážte, že ak $\hat{\sigma}_1^2 < \hat{\sigma}_2^2$, tak prvý model má väčší upravený koeficient determinácie.

9. Odhadujeme metódou maximálnej vierohodnosti model $Y = X\beta + \varepsilon$, pričom matica X má rozmer $n \times k$ a má lineárne nezávislé stĺpce, vektor ε má normálne rozdelenie, nulovú strednú hodnotu a kovariančnú maticu $\sigma^2 I$. Odvoďte likelihood ratio test na testovanie lineárnej hypotézy o parametroch $R\beta = r$.

Časť III

V súbore `domy.txt` nájdete dáta pre 26 predaných domov[2], spolu s ich popisom. Vytvorte regresný model na vysvetlenie ceny domu od ostatných premenných (môžete ich aj transformovať), ktorý obsahuje aspoň dve vysvetľujúce premenné.

- Ktoré parametre sú vo vašom modeli signifikantné? Je regresia signifikantná?
- Aké sú očakávané znamienka parametrov? Zhodujú sa odhady s týmto očakávaním?
- Vytvorte submodel a testujte ho pomocou F testu.
- Uveďte iné spôsoby, ako porovnať dva modely. Ukážte ich použitie na vašom modeli a submodeli z predchádzajúcej časti.

[2] Narula, Wellington, 1977