

1 Odhadovanie parametrov, vlastnosti odhadov.

Metóda maximálnej vierohodnosti.

1. Uvažujme náhodnú premennú, ktorej hustota $3\theta^3 \frac{1}{x^4}$ pre $x > \theta$, inak nulová. Dokážte, že odhad

$$\hat{\theta} = \frac{2}{3} \frac{1}{n} \sum_{i=1}^n X_i$$

je nevychýlený a konzistentný odhad parametra θ .

2. Ak náhodná premenná X má normálne rozdelenie $N(\mu, \sigma^2)$, tak rozdelenie náhodnej premennej $Y = e^X$ sa nazýva lognormálne (s parametrami μ, σ^2).

(a) Nájdite hustotu premennej Y a odhad parametrov lognormálneho rozdelenia z dát y_1, \dots, y_n .

(b) Vypočítajte medián tohto rozdelenia, čiže takú hodnotu m , pre ktorú platí

$$P(Y \leq m) = 1/2, P(Y \geq m) = 1/2.$$

(c) Zostrojte likelihood ratio test na testovanie hypotézy, že medián sa rovná e .

2 Lineárny regresný model, metóda najmenších štvorcov

Pokiaľ nie je povedané inak, uvažujeme model $y = X\beta + \varepsilon$, matica X , pričom má plnú hodnotu a ε má normálne rozdelenie $N(0, \sigma^2, I)$. Parameter $\beta = (\beta_0, \beta_1, \dots, \beta_{k-1})^T$ je k -rozmerný a model odhadujeme z n dát.

1. MNŠ odhad a jeho vlastnosti

Pre MNŠ odhad odvoďte:

- (a) vzorec $(X^T X)^{-1} X^T y$,
- (b) nevychýlenosť
- (c) kovariančnú maticu,
- (d) rozdelenie.

Ako odhadujeme kovariančnú maticu z časti (c)?

2. Testovanie signifikancie 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

$$Y = C(1) + C(2)x_1 + C(3)x_2 + C(4)x_3 + \varepsilon,$$

dostaneme výstup na obrázku 1. Ktoré parametre sú signifikantné?

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

Obr. 1: Príklady 2(d), 3(d), 4(c)

3. Testovanie signifikancie regresie

- Sformulujte nulovú hypotézu, ktorou testujeme signifikantnosť parametrov. Kedy je regresia signifikantná - ak H_0 zamietame alebo ak H_0 nezamietame?
- Akou štatistikou túto hypotézu testujeme - ako sa vypočíta a aké má rozdelenie, ak platí H_0 ?
- Aká je kritická oblasť testu, t.j. kedy H_0 zamietame?
- Znovu uvažujme model z predchádzajúceho príklad a výstup na obrázku 1. Je regresia signifikantná?

4. Testovanie lineráných hypotézy o parametroch

- Ako sa testuje hypotéza tvaru $a^T\beta = b$? Popíšte výpočet štatistiky, jej rozdelenie za platnosti nulovej hypotézy a obor zamietnutia.
- Ako sa testuje hypotéza tvaru $R\beta = r$? Popíšte výpočet štatistiky, jej rozdelenie za platnosti nulovej hypotézy a obor zamietnutia.
- V regresii z predchádzajúceho príkladu testujte hypotézu $C(2) = 20$.
- Vypočítajte odhad parametrov β a σ^2 metódou maximálnej vierohodnosti. Odvoďte likelihood ratio test na testovanie lineárnej hypotézy $R\beta = r$.

5. Doplňte vynechané hodnoty vo výstupe na obrázku 2.

6. Doplňte vynechané hodnoty vo výstupe na obrázku 3.

7. Prepodkladajme, že $E[\varepsilon_i] = 1$, ostatné predpoklady zostávajú v platnosti. Ktoré z odhadov $\beta_0, \beta_1, \dots, \beta_{k-1}$ budú vychýlené? Svoje tvrdenie dokážte.
8. Dokážte, že regresná priamka v modeli $y = \alpha + \beta x + \varepsilon$ prechádza bodom (\bar{x}, \bar{y}) .
9. Dokážte, že aj je výberová korelácia medzi dátami (x_i, y_i) kladná, tak MNŠ odhad parametra β v modeli $y = \alpha + \beta x + \varepsilon$ je kladný
10. Z dát (x_i, y_i) odhadneme modely

$$y = \alpha + \beta x + \varepsilon_1, \quad x = \gamma + \delta y + \varepsilon_2$$

Zakreslíme do jedného grafu získané regresné priamky $y = \hat{\alpha} + \hat{\beta}x$ a $x = \hat{\gamma} + \hat{\delta}y$. Budú tieto priamky zhodné? Svoje tvrdenie dokážte.

11. Z dát pre 22 firiem odhadujeme model

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

kde $y = \ln(\text{output})$, $x = \ln(\text{labor_hours_input})$ máme nasledovné informácie:

$$\bar{y} = 20, \quad \sum (y_i - \bar{y})^2 = 100$$

$$\bar{x} = 10, \quad \sum (y_i - \bar{y})^2 = 60$$

$$\sum (x_i - \bar{x})(x_i - \bar{x}) = 30$$

- (a) Vypočítajte MNŠ odhady parametrov α, β .
- (b) Testujte hypotézu $\beta = 1$.

Dependent Variable: Y23
 Method: Least Squares
 Date: 04/07/06 Time: 13:04
 Sample: 1 75
 Included observations: 75
 $Y23=C(1)+C(2)*X+C(3)*X^2$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	164.9689	16.56590		0.0000
C(2)	31.91071		3.407309	0.0011
C(3)		2.889207	-15.36713	0.0000

Obr. 2: Příklad 5

Dependent Variable: Y
 Method: Least Squares
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.45746		6.436957	0.0000
X	2.223056			0.0000
R-squared	0.758930	Mean dependent var		21.29375
Adjusted R-squared	0.745538	S.D. dependent var		6.676421
S.E. of regression	3.367872	Akaike info criterion		5.361079
Sum squared resid	204.1661	Schwarz criterion		5.460652
Log likelihood	-51.61079	F-statistic		56.66721
Durbin-Watson stat	2.121118	Prob(F-statistic)		0.000001

Obr. 3: Příklad 6