
ZADANIA

Zadanie 1. Pokaż, że $\log Y = \alpha + \beta \log X + \epsilon$ daje ten sam estymator najmniejszych kwadratów dla β niezależnie czy logarytmiczna we wzorze jest logarytmem przy podstawie 10 czy też logarytmem naturalnym. Czy jest to prawda dla estymatora najmniejszych kwadratów dla α ?

Zadanie 2. Znajdź transformacje, które zlinearyzują każdą z tych funkcji:

(a)

$$y = \frac{x}{\alpha x - \beta},$$

(b)

$$y = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}.$$

Zadanie 3. Załóżmy, że masz dwa niezależne nieobciążone estymatory, powiedzmy T_1 i T_2 tego samego parametru β , z różnymi wariancjami, odpowiednio σ_1 i σ_2 . Jaka kombinacja liniowa $T = c_1 T_1 + c_2 T_2$ jest nieobciążonym estymatorem o minimalnej wariancji parametru β ?

Zadanie 4. Chcemy dopasować model $y = \alpha + \beta x + \epsilon$, ale nie mamy do dyspozycji pełnego zestawu danych $[y \ X] = Z$. Zamiast tego mamy tylko $Z^T Z$:

$$Z^T Z = \begin{pmatrix} 100 & 10 & 25 \\ 10 & 20 & 0 \\ 25 & 0 & 75 \end{pmatrix}$$

Oblicz β , s^2 i R^2 .

Zadanie 5. Załóżmy, że modelem regresji liniowej modelujemy zależność x od y . Chcemy zminimalizować sumę kwadratów reszt (RSS), ale reszty są teraz odległościami w poziomie (zamiast odległościami mierzonymi w pionie), gdy na wykresie x odkładane jest na osi poziomej. To daje inną sumę reszt RSS, która będzie odpowiadała resztom w regresji x na y . Możemy więc dokonać regresji y na x lub x na y , ale tylko jedną wartość R^2 obliczamy dla zbioru danych. Niech b_{yx} i b_{xy} będą odpowiednio estymatorami współczynników prostych regresji y na x i x na y .

(a) Czy $b_{yx} = \frac{1}{b_{xy}}$?

(b) Czy zawsze b_{yx} i b_{xy} będą miały zawsze ten sam znak?

(c) Pokaż, że $r^2 = b_{yx} b_{xy}$.

- (d) Z (c) wynika, że $b_{yx} \leq \frac{1}{b_{xy}}$ pod warunkiem, że $b_{xy} > 0$. To znaczy, że na wykresie z X w osi poziomej regresja x na y ma większe nachylenie. Sprawdź część (c) i nierówność $b_{yx} \leq \frac{1}{b_{xy}}$, przy założeniu, że $b_{xy} > 0$, dla próby 200 obserwacji, dla których:

$$\sum x = 11,34 \quad \sum y = 20,72 \quad \sum x^2 = 12,16 \quad \sum y^2 = 84,96 \quad \sum xy = 22,13.$$

Zadanie 6. Rozważmy następujący model dla $n = 20$ obserwacji:

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

gdzie $i = 1, 2, \dots, n$.

Wiadomo, że $\sum_{i=1}^{20} y_i = 50$, $\sum_{i=1}^{20} y_i^2 = 225$, $\sum_{i=1}^{20} x_i = 30$, $\sum_{i=1}^{20} x_i^2 = 100$, $\sum_{i=1}^{20} x_i(y_i - \bar{y}) = 60$.

- (a) Oszacuj parametry strukturalne modelu $\hat{\beta}_0, \hat{\beta}_1$ metodą MNK.

- (b) Oblicz kowariancję:

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

- (c) Oblicz wariancję próbkową jako estymator nieobciążony:

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

oraz

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

- (d) Oblicz współczynnik korelacji:

$$r_{X,Y} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_X s_Y}.$$

- (e) Udowodnij, że $\sum_{i=1}^n e_i = 0$ oraz oblicz sumę kwadratów reszt (zmiennosć zmiennej objaśnianej niewyjaśnioną przez model):

$$SSE := \sum_{i=1}^n e_i^2$$

- (f) Oblicz zmiennosć zmiennej objaśnianej wyjaśnioną przez model:

$$SSR := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2;$$

(g) Oblicz całkowitą zmienność zmiennej objaśnianej:

$$SST := \sum_{i=1}^n (y_i - \bar{y})^2;$$

(h) Oblicz współczynnik determinacji:

$$R^2 := \frac{SSR}{SST}$$

oraz podaj jego interpretację.

Zadanie 7. Oceny parametrów strukturalnych modelu uzyskane metodą MNK są następujące:

$$\hat{Y}_i = 1 + 0,8x_i + \epsilon_i$$

dla $i = 1, 2, \dots, 51$. Wiadomo również, że $\sum_{i=1}^{51} x_i = 60$ i odchylenia standardowe z próby s_X, s_Y (jako estymatory nieobciążone) wyniosły 5. Znajdź oceny parametrów strukturalnych następującego modelu metodą MNK:

$$X_i = \alpha_0 + \alpha_1 y_i + \epsilon_i^*, \quad i = 1, 2, \dots, 51.$$

Zadanie 8. Udowodnij, że współczynnik determinacji R^2 jest równy kwadratowi współczynnika korelacji próbkowej r^2 .

Zadanie 9. Oszacowano model regresji liniowej metodą MNK:

$$\hat{Y}_i = 0,81x_i + 2,57 + \epsilon_i.$$

Wiadomo, że model regresji liniowej metodą MNK dla $\hat{X}_i$ przedstawia się następująco:

$$\hat{X}_i = \hat{\alpha}_1 y_i - 2 + \epsilon_i^*$$

oraz $\bar{x} = 3$.

(a) Oblicz wartość $\hat{\alpha}_1$ na podstawie danych.

(b) Oblicz współczynnik korelacji $r_{X,Y}$.

Zadanie 10. Wyznacz parametry strukturalne $\hat{\alpha}_0, \hat{\alpha}_1$ metodą MNK w poniższym modelu

$$X_i = \alpha_0 + \alpha_1 y_i + \epsilon_i^*, \quad i = 1, 2, \dots, n,$$

jeśli wiadomo, że:

$$\hat{Y}_i = 0,5 - 0,5x_i + \epsilon_i, \quad i = 1, 2, \dots, n,$$

oraz $\bar{y} = -0,25, |r_{X,Y}| = 0,6$.

Zadanie 11. Dane są modele $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ oraz $X_i = \alpha_0 + \alpha_1 y_i + \epsilon_i^*$ dla $i = 1, 2, \dots, n$. Dokonujemy estymacji parametrów strukturalnych MNK.

- (a) Na podstawie rozumowań przeprowadzonych w poprzednich zadaniach podaj jaka jest zależność między estymatorami $\hat{\alpha}_k, \hat{\beta}_k$ ($k = 0, 1$).
- (b) Oblicz współczynniki determinacji w obu modelach i porównaj je.

Zadanie 12. Rozważmy następujący model:

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

gdzie $i = 1, 2, \dots, n$. Wiadomo, że

$$\sum_{i=1}^n (\hat{Y}_i - Y_i)^2 = 10000, \quad \bar{Y} = 1400, \quad \sum_{i=1}^{20} x_i = 104, \quad \sum_{i=1}^n x_i^2 = 1520$$
$$n \sum_{i=1}^n x_i^2 - n^2 \bar{x}^2 = 1344, \quad \sum_{i=1}^n x_i y_i = 162400.$$

- (a) Oszacuj parametry strukturalne modelu $\hat{\beta}_0, \hat{\beta}_1$ metodą MNK.
- (b) Oblicz współczynnik determinacji i podaj jego interpretację.
- (c) Oblicz współczynnik korelacji $r_{X,Y}$.