
ZADANIA

Zadanie 1. Udowodnij następującą równość dotyczącą estymatora najmniejszych kwadratów współczynnika kierunkowego prostej regresji

$$\text{Cov}(\bar{y}, \hat{\beta}_1) = 0.$$

Zadanie 2. Oblicz współczynnik korelacji r_{XY} między zmiennymi X i Y , jeśli modele regresji liniowej MNK przedstawiają się następująco:

$$\hat{Y} = 0,9x + 0,42, \quad \hat{X} = ay + 0,9$$

oraz $\bar{y} = 1,5$.

Zadanie 3. Model ekonometryczny

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

oszacowano 3 różnymi metodami, w tym metodą MNK. Wektory reszt dla kolejnych metod są następujące:

$$\mathbf{a} = \begin{bmatrix} 1 \\ -3 \\ 2 \\ 4 \\ -1 \\ -3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 6 \\ 4 \\ -5 \\ -6 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 \\ -1 \\ 7 \\ -7 \\ 2 \\ -2 \end{bmatrix}$$

Który wektor odpowiada metodzie MNK? Odpowiedź uzasadnij.

Zadanie 4. Rozważmy następujący model:

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

gdzie $i = 1, 2, \dots, 20$. Wiadomo, że

$$\sum_{i=1}^{20} (\hat{Y}_i - Y_i)^2 = 4,75, \quad \sum_{i=1}^{20} Y_i = 108, \quad \sum_{i=1}^{20} x_i^2 = 40$$
$$20 \sum_{i=1}^{20} x_i^2 - \left(\sum_{i=1}^{20} x_i \right)^2 = 544, \quad \sum_{i=1}^{20} x_i Y_i = 100.$$

- (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 $r_{X,Y}$.

Zadanie 5. Badamy zależność wydatków konsumpcyjnych od dochodów. W tym celu przebadaliśmy wydatki i dochody 10 rodzin otrzymując następujące dane:

lp	Wydatki	Dochody
1	110	200
2	200	320
3	130	150
4	150	310
5	120	240
6	180	210
7	220	270
8	130	180
9	250	360
10	320	490

W tabeli podano miesięczny dochód oraz wydatek konsumpcyjny w przeliczeniu na osobę. Oszacuj model regresji liniowej metodą MNK wpływu dochodów na wydatki konsumpcyjne.

- Narysuj wykres rozproszenia danych. Czy na podstawie wykresu można stwierdzić, że między wydatkami i dochodami istnieje zależność?
- Przedstaw model w zapisie macierzowym.
- Oszacuj parametry strukturalne modelu $\hat{\beta}_0, \hat{\beta}_1$ metodą MNK.
- Oblicz wariancję próbkową jako estymator nieobciążony wariancji dochodów oraz wydatków:

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

oraz

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

- Oblicz współczynnik korelacji:

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

- Oblicz zmienność zmiennej objaśnianej wyjaśnioną przez model:

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

- 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 6. Udowodnij korzystając z założeń Gaussa-Markowa, że estymatory $\hat{\beta}_0, \hat{\beta}_1$ uzyskane metodą MNK są estymatorami zgodnymi parametrów strukturalnych β_0, β_1 .

Zadanie 7. Rozważmy następujący model:

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

oszacowany w zadaniu 5. Przy założeniach Gaussa-Markowa podaj nieobciążony estymator wariancji składnika losowego ϵ_i .

Zadanie 8. Pokaż, że $SST = SSreg + RSS$, czyli całkowita wariancja ($SST = \sum_{i=1}^n (y_i - \bar{y})^2$) jest równa sumie regresyjnej sumy kwadratów ($SSreg = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$) i sumy kwadratów reszt ($RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$). W tym celu pokaż, że

(a) $\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$,

(b) $(y_i - \hat{y}_i) = (y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})$,

(c) $\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$.

(d) Korzystając z faktu, że $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ pokaż, że $(\hat{y}_i - \bar{y}) = \hat{\beta}_1(x_i - \bar{x})$.

Zadanie 9. Rozpatrzy prosty model regresji liniowej $y = \beta_0 + \beta_1 x + \epsilon$, gdzie $E(\epsilon) = 0$, i $\text{Var}(\epsilon) = \sigma^2$ oraz błędy są nieskorelowane. Pokaż, że

(a)

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}\sigma^2}{S_{xx}},$$

(b)

$$\text{Cov}(\bar{y}, \hat{\beta}_1) = 0.$$

Zadanie 10. Załóżmy, że dopasowaliśmy do danych prostą regresji $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$, ale w rzeczywistości zmienna odpowiedzi y zależy również od zmiennej x_2 , więc właściwy model regresji powinien mieć postać

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

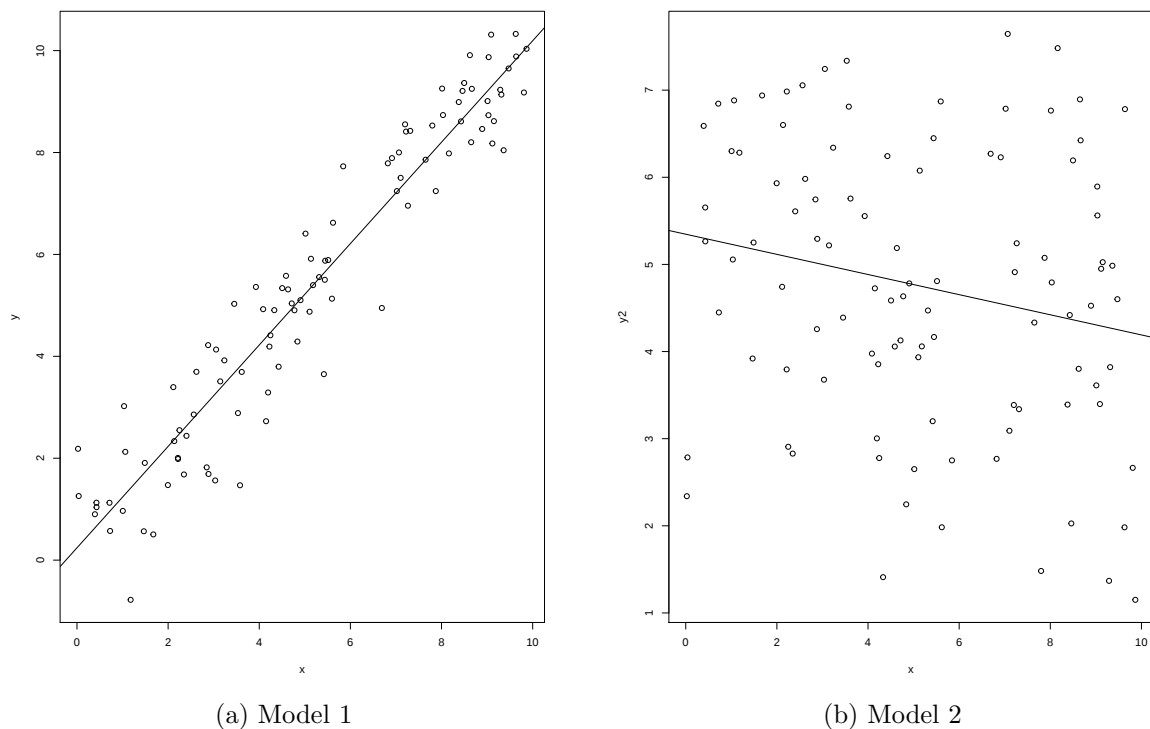
(a) Czy w tej sytuacji estymator najmniejszych kwadratów $\hat{\beta}_1$ współczynnika kierunkowego prostej regresji jest nieobciążony?

(b) Wyznacz obciążenie estymatora $\hat{\beta}_1$.

Zadanie 11. Dla y zaproponowano dwa alternatywne modele prostej regresji liniowej. W pierwszym modelu y jest funkcją liniową x_1 , podczas gdy w drugim modelu y jest funkcją liniową x_2 . Wykres w pierwszej kolumnie na rysunku 1 przedstawia zależność Y od x_1 , natomiast wykres w drugiej kolumnie to wykres y w zależności od x_2 . Na wykresach pokazane są również proste regresji wyznaczone metodą najmniejszych kwadratów. W poniższych stwierdzeniach RSS oznacza sumę kwadratów reszt, a $SSreg$ oznacza residualną sumę kwadratów ($SSreg = \sum(\hat{y}_i - \bar{y})^2$). Które z poniższych stwierdzeń jest prawdziwe?

- (a) RSS dla modelu 1 jest większy niż RSS dla modelu 2, natomiast $SSreg$ dla modelu 1 jest większy niż $SSreg$ dla modelu 2.
- (b) RSS dla modelu 1 jest mniejszy niż RSS dla modelu 2, podczas gdy $SSreg$ dla modelu 1 jest mniejszy niż $SSreg$ dla modelu 2.
- (c) RSS dla modelu 1 jest większy niż RSS dla modelu 2, a $SSreg$ dla modelu 1 jest mniejszy niż $SSreg$ dla modelu 2.
- (d) RSS dla modelu 1 jest mniejszy niż RSS dla modelu 2, podczas gdy $SSreg$ dla modelu 1 jest większy niż $SSreg$ dla modelu 2.

Podaj argumenty dla poparcia swojego wyboru.



Rysunek 1: Wykresy punktowe oraz proste regresji dla modelu 1 i dla modelu 2.

Uwaga. Odpowiedzi na pytania z zadania 11 można sprawdzić korzystając z zamieszczonych na stronie do danych.

Barbara Jasiulis-Góldyn
Krzysztof Topolski