

Przypomnijmy założenia modelu regresji liniowej.

Założenie 1 (Liniowość)

$$y_i = \sum_{j=1}^K \beta_j x_{ij} + \varepsilon_i \quad i=1, 2, \dots, n,$$

gdzie β są nieznanymi parametrami
a ε_i jest nieobserwowanym błędem

W zapisie macierzyowym założenie 1
możemy zapisać w następującej postaci

$$\begin{array}{ccccc} y & = & X & \beta & + & \varepsilon \\ (n \times 1) & & (n \times K) & (K \times 1) & & (n \times 1) \\ & & \underbrace{\hspace{1.5cm}} & & & \\ & & (n \times 1) & & & \end{array}$$

Założenie 2

$$E(\varepsilon_i | X) = 0 \quad i=1, 2, \dots, n$$

W bardziej precyzyjnej bezpośredniej
postaci założenie 2 możemy zapisać
nie odwołując się do macierzy X

$$E(\varepsilon_i | x_1, \dots, x_n) = 0 \quad i=1, 2, \dots, n$$

Założenie 3 (brak współliniowości)

Rząd macierzy danych X rozmiaru $n \times K$ jest z prawdopodobieństwem 1 równy K .

Założenie 4 (sferyczność wariancji błędów)

$$E(\varepsilon_i^2 | X) = \sigma^2 > 0, \quad i=1, 2, \dots, n$$

$$E(\varepsilon_i \varepsilon_j | X) = 0, \quad i, j=1, 2, \dots, n; i \neq j$$

Założenie 4 możemy zapisać w postaci

$$\text{Var}(\varepsilon | X) = \sigma^2 I_n$$

co tłumaczy nazwę sferyczność wariancji.

Oznaczmy przez b estymator najmniejszych kwadratów parametru β , wtedy

$$b = (X^T X)^{-1} X^T y$$

$$\text{Ponieważ } (X^T X)^{-1} X^T y = \left(\frac{X^T X}{n} \right)^{-1} X^T y / n$$

otrzymujemy równoważną postać

$$b = S_{xx}^{-1} S_{xy}$$

gdzie

$$S_{xx} = \frac{1}{n} X^T X = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$$

$$S_{xy} = \frac{1}{n} X^T y = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i$$

3

Własności estymatora najmniejszych kwadratów b parametru β w modelu regresji liniowej

Przy założeniu założeń 1-4

(a) (nieobciążoność)

$$E(b | X) = \beta$$

$$(b) \quad \text{Var}(b | X) = \sigma^2 (X^T X)^{-1}$$

$$(c) \quad \text{Cov}(b, e | X) = 0,$$

$$\text{gdzie } e = y - Xb$$

TWIERDZENIE (Gausa-Markowa)

Przy założeniu założeń 1-4 estymator najmniejszych kwadratów jest efektywny w klasie liniowych, nieobciążonych estymatorów.

To znaczy, że dla każdego nieobciążonego estymatora $\hat{\beta}$, który jest liniowy ze względu na y

$$\text{Var}(\hat{\beta} | X) \geq \text{Var}(b | X)$$

w sensie maciernowym.

UWAGA

Niech A i B będą macierzami kwadratowymi tego samego rozmiaru. Powiemy, że $A \geq B$ gdy macierz $A - B$ jest macierzą nieujemnie zdefiniowaną.

Macierz C rozmiaru $K \times K$ jest nieujemnie zdefiniowana gdy dla dowolnego K -wymiarowego wektora x zachodzi nierówność postaci

$$x^T C x \geq 0$$

Dowód twierdzenia Gaussa-Markowa

Ponieważ estymator $\hat{\beta}$ jest liniowy względem y możemy jego postać zapisać jako

$$\hat{\beta} = C y,$$

gdzie C jest pewną macierzą, która może być funkcją X .

Niech $D = C - A$ lub $C = D + A$

gdzie $A = (X^T X)^{-1} X^T$. Wtedy

(5)

$$\begin{aligned}\hat{\beta} &= (D+A)y \\ &= Dy + Ay \\ &= D(X\beta + \epsilon) + b\end{aligned}$$

Ostatnia równość wynika z faktu, że

$$y = X\beta + \epsilon$$

oraz $Ay = (X^T X)^{-1} X^T y = b$

$$= DX\beta + D\epsilon + b$$

Nakładając warunkową wartość oczekiwaną otrzymamy

$$E(\hat{\beta} | X) = DX\beta + E(D\epsilon | X) + E(b | X)$$

Ponieważ zarówno b jak i $\hat{\beta}$ są niezobciążone natomiast

$$E(D\epsilon | X) = D E(\epsilon | X) = 0$$

otrymujemy, że $DX\beta = 0$.

Ostatnia równość jest prawdziwa dla każdego β tylko wtedy gdy

$$DX = 0 \text{ co implikuje } \hat{\beta} = D\epsilon + b$$

$$\text{ i } \hat{\beta} - \beta = D\epsilon + (b - \beta)$$

UWAGA

Przypomnijmy, że błąd prognozy $b - \beta$ można wyrazić w terminach ϵ

$$\begin{aligned} b - \beta &= (X^T X)^{-1} X^T y - \beta \\ &= (X^T X)^{-1} X^T (X\beta + \epsilon) - \beta \end{aligned}$$

pierwsza równość wynika z postaci estymatora b natomiast druga jest konsekwencją równości

$$y = X\beta + \epsilon \quad (\text{zakożenie 1})$$

$$\begin{aligned} &= (X^T X)^{-1} (X^T X) \beta + (X^T X)^{-1} X^T \epsilon - \beta \\ &= \beta + (X^T X)^{-1} X^T \epsilon - \beta \\ &= (X^T X)^{-1} X^T \epsilon \end{aligned}$$

Komputując z tej uwagi równość

$$\hat{\beta} - \beta = D\epsilon + (b - \beta)$$

możemy zapisać w postaci

$$\hat{\beta} - \beta = (D + A)\epsilon$$

$$\text{gdzie } A = (X^T X)^{-1} X^T$$

Tak więc

$$\begin{aligned}\text{Var}(\hat{\beta} | X) &= \text{Var}(\hat{\beta} - \beta | X) \\ &= \text{Var}[(D+A)\epsilon | X] \\ &= (D+A)\text{Var}(\epsilon | X)(D^T + A^T)\end{aligned}$$

ponieważ zarówno A jak i D są funkcjami X

$$= \sigma^2 (D+A)(D^T + A^T)$$

ponieważ $\text{Var}(\epsilon | X) = \sigma^2 I_n$ (założenie 4)

$$= \sigma^2 (DD^T + AD^T + DA^T + AA^T)$$

Zauważmy, że $DA^T = DX(X^T X)^{-1} = 0$
ponieważ $DX = 0$.

Również $AA^T = (X^T X)^{-1}$ (*) Uwaga 2

Wsc

$$\begin{aligned}\text{Var}(\hat{\beta} | X) &= \sigma^2 [DD^T + (X^T X)^{-1}] \\ &\geq \sigma^2 (X^T X)^{-1}\end{aligned}$$

gdzie DD^T jest nieujemnie
zdefiniowana

$$= \text{Var}(b | X)$$

Ostatnia równość jest konsekwencją
własności (b) estymatora b

2)

UWAGA 2 $AA^T = (X^T X)^{-1}$

$$\underbrace{(X^T X)^{-1} X^T}_A \underbrace{[(X^T X)^{-1} X^T]^T}_{A^T} = (X^T X)^{-1}$$

UWAGA 3

Założenie 2 mówi, że dla $i=1, 2, \dots, n$

$$E(\varepsilon_i | X) = 0$$

jest istotne dla dowodu nieobciążoności estymatora b.

Stabsze założenie np. założenie

$$E(\varepsilon_i | x_i) = 0, \text{ dla wszystkich } i$$

lub założenie

$$E(x_i \cdot \varepsilon_i) = 0, \text{ dla wszystkich } i$$

są niewymagające.

W tym takich słabszych założeniach estymator najmniejszych kwadratów może być obciążony.

Co twierdzenie Gaussa-Markowa nie mówi.

Przy założeniach 1-4 o modelu regresji możemy wyznaczyć estymator liniowy niekoniecznie nieobciążony ale o mniejszej wariancji niż nieobciążony i liniowy estymator.

Możemy również wyznaczyć estymator nieliniowy o mniejszej wariancji.

Jeżeli jednak dodamy dodatkowe założenie

Założenie 5

Pod warunkiem X rozkład bieżący ϵ jest normalny.

UWAGA. Biorąc pod uwagę założenia 1 i 4 założenie 5 implikuje, że rozkład ϵ pod warunkiem X jest rozkładem wielowymiarowym normalnym $N(0, \sigma^2 I_n)$

to możemy udowodnić następujące twierdzenie

TWIERDZENIE. Przy założeniu zachodzenia założeń 1-5 estymator najmniejszych kwadratów b parametru β w modelu regresji liniowej jest najlepszym nieobciążonym estymatorem. To znaczy, że żaden nieobciążony estymator, który nie musi być liniowy nie ma mniejszej waronkowej wariancji.

UWAGA

Przy założeniach 1-4 estymator b jest BLUE (best linear unbiased estimator).

Przy założeniach 1-5 estymator b jest BUE (best unbiased estimator)

UWAGA

Podobnie jak w przypadku prostej regresji liniowej gdy założony normalność błędów estymator najmniejszych kwadratów pokrywa się z estymatorem największej wiarygodności.

Efektywność możemy badać korzystając z nierówności Cramera-Rao. Patrz przykład ze statystyki.