Задача 39*[теорема Гаусса–Маркова]. Для регрессионной модели

$$Y_i = \beta x_i + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i] = 0, \quad \mathrm{D}(\varepsilon_i) = \sigma^2, \quad \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j, \qquad (1^*)$$

докажите частный случай теоремы Гаусса–Маркова. А именно, покажите, что

(a) МНК-оценка $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ является линейной по вектору $Y = (Y_1, \ldots, Y_n)^T$,

(b) МНК-оценка $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ является несмещенной оценкой параметра β,

(c) МНК-оценка $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ имеет наименьшую дисперсию в классе всех линейных по $Y = (Y_1, \ldots, Y_n)^T$, несмещенных оценок параметра β.

Решение. (Способ 1: доказательство в скалярной форме)

(a) Пусть $Y = (Y_1, \ldots, Y_n)^T$, $Z = (Z_1, \ldots, Z_n)^T$ и $\lambda \in \mathbb{R}$. Функция $\hat{\beta}(X, Y) = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ является линейной по переменной Y, поскольку

$$\hat{\beta}(X, Y+Z) = \frac{\sum_{i=1}^n x_i (Y_i + Z_i)}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} + \frac{\sum_{i=1}^n x_i Z_i}{\sum_{i=1}^n x_i^2} = \hat{\beta}(X, Y) + \hat{\beta}(X, Z)$$

и

$$\hat{\beta}(X, \lambda Y) = \frac{\sum_{i=1}^n x_i (\lambda Y_i)}{\sum_{i=1}^n x_i^2} = \lambda \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} = \lambda \hat{\beta}(X, Y).$$

(b) Оценка $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ является несмещенной, поскольку

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}\left[\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}\right] = \frac{1}{\sum_{i=1}^n x_i^2} \mathbb{E}\left[\sum_{i=1}^n x_i Y_i\right] = \frac{1}{\sum_{i=1}^n x_i^2} \sum_{i=1}^n x_i \mathbb{E}\left[Y_i\right] = \frac{1}{\sum_{i=1}^n x_i^2} \sum_{i=1}^n x_i \beta x_i = \beta.$$

(c) Прежде всего, заметим, что всякая линейная по вектору $Y = (Y_1, \ldots, Y_n)^T$ оценка $\tilde{\beta}$ имеет вид:

$$\tilde{\beta} = \sum_{i=1}^n c_i Y_i.$$

Представим МНК-оценку $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ в виде

$$\hat{\beta} = \sum_{i=1}^n a_i Y_i, \qquad \text{где } a_i := \frac{x_i}{\sum_{j=1}^n x_j^2}, \ i = 1, \ldots, n.$$

Тогда оценку $\tilde{\beta}$ можно переписать в виде

$$\tilde{\beta} = \sum_{i=1}^n (a_i + b_i) Y_i,$$

где $b_i := c_i - a_i$, $i = 1, \ldots, n$.

Далее, учитывая, что МНК-оценка $\hat\beta$ является несмещенной, т.е. $\mathbb{E}[\hat\beta]=\beta$ для любого $\beta\in\mathbb{R}$, получаем соотношение

$$\mathbb{E}[\hat\beta]=\mathbb{E}\left[\sum_{i=1}^n a_iY_i\right]=\sum_{i=1}^n a_i\mathbb{E}[Y_i]=\sum_{i=1}^n a_i\beta x_i=\beta\sum_{i=1}^n a_ix_i=\beta,$$

верное при любом $\beta\in\mathbb{R}$. Следовательно, верное и при $\beta=1$. Значит,

$$\sum_{i=1}^n a_ix_i=1. \qquad (2*)$$

В свою очередь, оценка $\tilde\beta$ также является несмещенной, т.е. $\mathbb{E}[\tilde\beta]=\beta$ для любого $\beta\in\mathbb{R}$. Следовательно, имеет место соотношение

$$\mathbb{E}[\tilde\beta]=\mathbb{E}\left[\sum_{i=1}^n (a_i+b_i)Y_i\right]=\sum_{i=1}^n (a_i+b_i)\mathbb{E}[Y_i]=\sum_{i=1}^n (a_i+b_i)\beta x_i=\beta\sum_{i=1}^n (a_i+b_i)x_i=\beta,$$

верное при любом $\beta\in\mathbb{R}$. Стало быть, верное и при $\beta=1$. Поэтому

$$\sum_{i=1}^n (a_i+b_i)x_i=1. \qquad (3*)$$

Из соотношений $(2*)$ и $(3*)$ получаем, что

$$\sum_{i=1}^n b_ix_i=0. \qquad (4*)$$

В свою очередь, отсюда следует, что

$$\sum_{i=1}^n a_ib_i=\sum_{i=1}^n \frac{x_i}{\sum_{j=1}^n x_j^2}b_i=\frac{1}{\sum_{j=1}^n x_j^2}\underbrace{\sum_{i=1}^n x_ib_i}_{=0\ (\text{по }(4*))}=0. \qquad (5*)$$

Теперь сравним дисперсии $\mathrm{D}(\hat\beta)$ и $\mathrm{D}(\tilde\beta)$. Имеем:

$$\mathrm{D}(\hat\beta)=\mathrm{D}(\sum_{i=1}^n a_iY_i)=\sum_{i=1}^n \mathrm{D}(a_iY_i)=\sum_{i=1}^n a_i^2\,\mathrm{D}(Y_i)=$$

$$=\sum_{i=1}^n a_i^2\,\mathrm{D}(\beta x_i+\varepsilon_i)=\sum_{i=1}^n a_i^2\,\mathrm{D}(\varepsilon_i)=\sum_{i=1}^n a_i^2\sigma^2=\sigma^2\sum_{i=1}^n a_i^2;$$

$$\mathrm{D}(\tilde\beta)=\mathrm{D}(\sum_{i=1}^n (a_i+b_i)Y_i)=\sum_{i=1}^n \mathrm{D}((a_i+b_i)Y_i)=\sum_{i=1}^n (a_i+b_i)^2\,\mathrm{D}(Y_i)=$$

$$=\sum_{i=1}^n (a_i+b_i)^2\,\mathrm{D}(\beta x_i+\varepsilon_i)=\sum_{i=1}^n (a_i+b_i)^2\sigma^2=\sigma^2\sum_{i=1}^n (a_i+b_i)^2=$$

$$=\sigma^2\sum_{i=1}^n a_i^2+2\sigma^2\underbrace{\sum_{i=1}^n a_ib_i}_{=0\ (\text{по }(5*))}+\sigma^2\underbrace{\sum_{i=1}^n b_i^2}_{\ge 0}\ge\sigma^2\sum_{i=1}^n a_i^2=\mathrm{D}(\hat\beta),$$

т.е. дисперсия МНК-оценки $\hat\beta$ нестрого меньше, чем дисперсия произвольной линейной по вектору $Y=(Y_1,\ldots,Y_n)^T$, несмещенной оценки $\tilde\beta$.

(Способ 2: доказательство в матричной форме)

Перепишем модель $(1*)$ в матричной форме:

$$
\left(Y_i = \beta x_i + \varepsilon_i,\ i = 1, \ldots, n\right) \ \Leftrightarrow\ \begin{cases} Y_1 = \beta x_1 + \varepsilon_1, \\ Y_2 = \beta x_2 + \varepsilon_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ Y_n = \beta x_n + \varepsilon_n, \end{cases} \ \Leftrightarrow\ \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{=:Y} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{=:X} \cdot \beta + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{=:\varepsilon}.
$$

Таким образом, модель $(1*)$ может быть представлена в виде:

$$
Y = X\beta + \varepsilon, \quad \text{где } \mathbb{E}[\varepsilon] = \mathbf{0} \text{ и } \mathrm{V}(\varepsilon) = \sigma^2 I. \tag{6*}
$$

В матричных обозначениях МНК-оценка $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$ имеет вид: $\hat{\beta} = (X^T X)^{-1} X^T Y$.

(a) Пусть $Y = (Y_1, \ldots, Y_n)^T$, $Z = (Z_1, \ldots, Z_n)^T$ и $\lambda \in \mathbb{R}$. Функция $\hat{\beta}(X, Y) = (X^T X)^{-1} X^T Y$ является линейной по переменной Y, поскольку

$$
\hat{\beta}(X, Y+Z) = (X^T X)^{-1} X^T (Y+Z) = (X^T X)^{-1} X^T Y + (X^T X)^{-1} X^T Z = \hat{\beta}(X, Y) + \hat{\beta}(X, Z)
$$

и

$$
\hat{\beta}(X, \lambda Y) = (X^T X)^{-1} X^T (\lambda Y) = \lambda (X^T X)^{-1} X^T Y = \lambda \hat{\beta}(X, Y).
$$

(b) Оценка $\hat{\beta} = (X^T X)^{-1} X^T Y$ является несмещенной, поскольку

$$
\mathbb{E}[\hat{\beta}] = \mathbb{E}\left[(X^T X)^{-1} X^T Y\right] = (X^T X)^{-1} X^T \mathbb{E}\left[Y\right] = (X^T X)^{-1} X^T X\beta = \beta.
$$

(c) Прежде всего, заметим, что всякая линейная по вектору $Y = (Y_1, \ldots, Y_n)^T$ оценка $\tilde{\beta}$ имеет вид:

$$
\tilde{\beta} = c^T Y, \quad \text{где } c \in \mathbb{R}^{n \times 1}.
$$

Представим МНК-оценку $\hat{\beta} = (X^T X)^{-1} X^T Y$ в виде

$$
\hat{\beta} = a^T Y, \quad \text{где } a^T = (X^T X)^{-1} X^T.
$$

Тогда оценку $\tilde{\beta}$ можно переписать в виде

$$
\tilde{\beta} = (a+b)^T Y,
$$

где $b := c - a$.

Далее, учитывая, что МНК-оценка $\hat{\beta}$ является несмещенной, т.е. $\mathbb{E}[\hat{\beta}] = \beta$ для любого $\beta \in \mathbb{R}$, получаем соотношение

$$
\mathbb{E}[\hat{\beta}] = \mathbb{E}\left[a^T Y\right] = a^T \mathbb{E}\left[Y\right] = a^T X\beta = \beta,
$$

верное при любом $\beta \in \mathbb{R}$. Следовательно, верное и при $\beta = 1$. Значит,

$$
a^T X = 1. \tag{7*}
$$

В свою очередь, оценка $\tilde{\beta}$ также является несмещенной, т.е. $\mathbb{E}[\tilde{\beta}] = \beta$ для любого $\beta \in \mathbb{R}$. Следовательно, имеет место соотношение

$$\mathbb{E}[\tilde{\beta}] = \mathbb{E}\Big[(a+b)^T Y\Big] = (a+b)^T \mathbb{E}\big[Y\big] = (a+b)^T X\beta = \beta,$$

верное при любом $\beta \in \mathbb{R}$. Стало быть, верное и при $\beta = 1$. Поэтому

$$(a+b)^T X = 1. \tag{8*}$$

Из соотношений $(7*)$ и $(8*)$ получаем, что

$$b^T X = 0. \tag{9*}$$

В свою очередь, отсюда следует, что

$$a^T b = (X^T X)^{-1} \underbrace{X^T b}_{=0 \text{ (по (9*))}} = 0. \tag{10*}$$

Теперь сравним дисперсии $\mathrm{D}(\hat{\beta})$ и $\mathrm{D}(\tilde{\beta})$. Имеем:

$$\mathrm{D}(\hat{\beta}) = \mathrm{V}(\hat{\beta}) = \mathrm{V}(a^T Y) = a^T\,\mathrm{V}(Y)a = a^T\,\mathrm{V}(X\beta+\varepsilon)a = a^T\,\mathrm{V}(\varepsilon)a = a^T\sigma^2 I a = a^T a;$$

$$\mathrm{D}(\tilde{\beta}) = \mathrm{V}(\tilde{\beta}) = \mathrm{V}((a+b)^T Y) = (a+b)^T\,\mathrm{V}(Y)(a+b) =$$

$$= (a+b)^T\,\mathrm{V}(X\beta+\varepsilon)(a+b) = (a+b)^T\,\mathrm{V}(\varepsilon)(a+b) =$$

$$= (a+b)^T\sigma^2 I(a+b) = \sigma^2(a+b)^T(a+b) = \sigma^2 a^T a + 2\sigma^2 \underbrace{a^T b}_{=0 \text{ (по (10*))}} + \underbrace{b^T b}_{\ge 0} \ge \sigma^2 a^T a = \mathrm{D}(\hat{\beta}),$$

т.е. дисперсия МНК-оценки $\hat{\beta}$ нестрого меньше, чем дисперсия произвольной линейной по вектору $Y = (Y_1, \ldots, Y_n)^T$, несмещенной оценки $\tilde{\beta}$. □