Contributors:
* Annika Bätz: annika.baetz@uni-potsdam.de
* Max Serra: maximilia-manuel.serra.lasierra@uni-potsdam.de
* Adrián de Miguel: adrian.de.miguel.palacio@uni-potsdam.de
* Ignacio Llorca: ignacio.llorca.rodriguez@uni-potsdam.de

SDA - Sheet 06 - Exercise 2

Linear Regression Problem.

$$y = X\beta + \epsilon \quad \text{with } \epsilon_i \sim N(0, \sigma^2) \qquad \text{(I)}$$

LS-estimator: $$\hat{\beta} = (X^T X)^{-1} X^T y \qquad \text{(II)}$$

Some useful equations / rules from our lecture:

$$\hat{y} = X\hat{\beta} = Hy \qquad \text{(III)}$$

$$\hat{\epsilon} = y - \hat{y} = y - X\hat{\beta} \qquad \text{(IV)}$$

$$H = X(X^T X)^{-1} X^T \qquad \text{(V)}$$

$$y \sim (X\beta, \sigma^2 I_n) \qquad \text{(VI)}$$

$$\hat{\epsilon} \overset{\text{(IV)}}{=} y - \hat{y} \overset{\text{(III)}}{=} y - Hy = (I_n - H)\, y \qquad \text{(VII)}$$

1. Task

Show that $\mathbb{E}[\hat{\epsilon}] = 0$.

$$\begin{aligned}
\mathbb{E}[\hat{\epsilon}] &\overset{\text{(IV)}}{=} \mathbb{E}[y - \hat{y}] \\
&\overset{(1)}{=} \mathbb{E}[y] - \mathbb{E}[\hat{y}] \\
&\overset{\text{(VI)},\text{(III)}}{=} X\beta - \mathbb{E}[X\hat{\beta}] \\
&\overset{(1)}{=} X\beta - X\,\mathbb{E}[\hat{\beta}] \\
&\overset{(2)}{=} X\beta - X\beta \\
&= 0 \quad \square
\end{aligned}$$

(1) linearity of expectation

(2) In sheet 05 we have proven: $\mathbb{E}[\hat{\beta}] = \beta$

The result $\mathbb{E}[\hat{\boldsymbol{\epsilon}}] = 0$ does fit our finding in the lecture that the residuals are zero on average:

$$\bar{\hat{\epsilon}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_i = 0 \qquad (\text{VIII})$$

$\mathbb{E}[\hat{\boldsymbol{\epsilon}}] = 0$ alone does not imply (VIII), but it proves a necessary condition for (VIII) to hold:

$$\mathbb{E}[\hat{\boldsymbol{\epsilon}}] = 0 \implies \mathbb{E}[\bar{\hat{\epsilon}}] = 0$$

Because $\mathbb{E}[\boldsymbol{\epsilon}] = 0$, it holds: $\underset{\text{error / noise}}{\mathbb{E}[\boldsymbol{\epsilon}]} = \underset{\text{residual}}{\mathbb{E}[\hat{\boldsymbol{\epsilon}}]}$, i.e. $\mathbb{E}[\epsilon_i] = \mathbb{E}[\hat{\epsilon}_i]$

Thus, we could see $\hat{\epsilon}_i$ as an estimator for the error ϵ_i.

2. Task

Determine $Cov(\hat{\boldsymbol{\epsilon}})$.

$$\begin{aligned}
Cov(\hat{\boldsymbol{\epsilon}}) &\overset{(\text{IV})}{=} Cov(y - \hat{y}) \\
&\overset{(\text{III})}{=} Cov(y - Hy) \\
&= Cov((I_n - H)\, y) \\
&\overset{(3)}{=} (I_n - H)\, Cov(y)\, (I_n - H)^T \\
&\overset{(\text{VI})}{=} (I_n - H)\, \sigma^2 I_n\, (I_n - H)^T \\
&= \sigma^2 (I_n - H)(I_n - H)^T \\
&\overset{(4)}{=} \sigma^2 (I_n - H)
\end{aligned} \qquad (\text{VII})$$

(3) It generally holds:
Let $\mathbf{X} \in \mathbb{R}^n$ be a multivariate RV and $\mathbf{A} \in \mathbb{R}^{m \times n}$ a matrix of constants. Then, it holds: $Cov(\mathbf{A}\mathbf{X}) = \mathbf{A}\, Cov(\mathbf{X})\, \mathbf{A}^T$

Here: $\mathbf{A}$ is given by $(I_n - H)$ with $H := X(X^T X)^{-1} X^T$
$\mathbf{X}$ is given by y

(4) We have seen in the lecture that $(I_n - H)$ is symmetric and idempotent.
$\implies (I_n - H)(I_n - H)^T = (I_n - H)(I_n - H) = (I_n - H)$

(Prove of this can be found on the next page)

$\mathrm{Cov}(\hat{\epsilon}) = \sigma^2 (I_n - H)$

Interpretation / Comments:

- In contrast to the error term ϵ for which the covariance matrix is a diagonal matrix: σI_n, i.e. $\mathrm{cov}(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$, i.e. each error term ϵ_i is independent, here we have dependencies:

  The matrix $\mathrm{Cov}(\hat{\epsilon})$ is not a diagonal matrix and $\mathrm{cov}(\epsilon_i, \epsilon_j) \neq 0$ in general (not necessarily though).

- We know that $H = X(X^T X)^{-1} X^T$, this means $\mathrm{Cov}(\hat{\epsilon})$ depends on σ^2 and X, <u>not</u> on y.

- For the variance of each residual $\hat{\epsilon}_i$ it holds:

  $$\mathrm{Var}(\hat{\epsilon}_i) = \sigma^2 (1 - h_{ii})$$

  entry in H in column i and row i

  It can be shown that $\frac{1}{n} \leq h_{ii} \leq 1$

  https://stats.stackexchange.com/questions/61924/diagonal-elements-of-the-projection-matrix

  This implies:

  $$\mathrm{Var}(\hat{\epsilon}_i) \leq \sigma^2 = \mathrm{Var}(\epsilon_i)$$

  The variance of each residual term $\hat{\epsilon}_i$ is smaller our equal to the variance of our noise / error terms ϵ_i.

Extra: Prove that $(I_n - H)$ is symmetric and idempotent.

$$H := X(X^T X)^{-1} X^T$$

$$H^T = \left(X(X^T X)^{-1} X^T\right)^T = X\left((X^T X)^{-1}\right)^T X^T = X(X^T X)^{-1} X^T = H$$

$\rightarrow$ H is symmetric

$$HH = X\underbrace{(X^T X)^{-1} X^T X}_{I_p}(X^T X)^{-1} X^T = X I_p (X^T X)^{-1} X^T$$

$$= X(X^T X)^{-1} X^T = H$$

$\rightarrow$ H is idempotent

$$(I_n - H)^T = I_n^T - H^T = I_n - H$$

$\rightarrow$ $(I_n - H)$ is symmetric

$$(I_n - H)(I_n - H) = I_n I_n - I_n H - H I_n + HH = I_n - 2H + H$$

$$= I_n - H$$

$\rightarrow$ $I_n - H$ is idempotent