FYS-STK Week 37

Harald Tverdal

August 2024

1 Week 37

1.1 Exercise 1

1.1.1 1

$$y_i = X_{i,*}\beta + \epsilon_i \tag{1}$$

Taking the expectation value of y_i gives:

$$\mathbb{E}[y_i] = \mathbb{E}[X_{i,*}\beta + \epsilon_i] = \mathbb{E}[X_{i,*}\beta] + \mathbb{E}[\epsilon_i] \tag{2}$$

We know $\mathbb{E}[\epsilon_i] = 0$, which gives:

$$\mathbb{E}[y_i] = \mathbb{E}[X_{i,*}\beta] = X_{i,*}\beta \tag{3}$$

This is because $X_{i,*}\beta$'s expectation value is equal to itself, as it is independent of the probability distribution of ϵ.

1.1.2 2

$$Var(y_i) = \mathbb{E}\Big[[y_i - \mathbb{E}[y_i]]^2\Big] \tag{4}$$

We know

$$\mathbb{E}\Big[[y_i - \mathbb{E}[y_i]]^2\Big] = \mathbb{E}[y_i^2] - \mathbb{E}[y_i]^2 \tag{5}$$

Using $\mathbb{E}[y_i] = X_{i,*}\beta$ and $y_i = X_{i,*}\beta + \epsilon_i$ for the expression, we get:

$$\begin{aligned} &\mathbb{E}[(X_{i,*}\beta + \epsilon_i)(X_{i,*}\beta + \epsilon_i)] - (X_{i,*}\beta)^2 \\ &= \mathbb{E}[(X_{i,*}\beta)^2 + (2\epsilon_i X_{i,*}\beta) + \epsilon_i^2] - (X_{i,*}\beta)^2 \\ &= \mathbb{E}[(X_{i,*}\beta)^2] + \mathbb{E}[(2\epsilon_i X_{i,*}\beta] + \mathbb{E}[\epsilon_i]^2 - (X_{i,*}\beta)^2 \end{aligned} \tag{6}$$

Where $\mathbb{E}[(X_{i,*}\beta)^2] - (X_{i,*}\beta)^2 = 0$ and $\mathbb{E}[\epsilon_i] = 0$, which means we end up with:

$$\mathbb{E}[\epsilon_i]^2 = \sigma^2 \tag{7}$$

1.1.3 3

Using the expression $\hat{\beta} = (X^T X)^{-1} X^T y$, we get:

$$\begin{aligned} \mathbb{E}[\hat{\beta}] = \mathbb{E}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \mathbb{E}[y] \\ (X^T X)^{-1} X^T X \beta = \beta \end{aligned} \tag{8}$$

1.1.4 4

$$Var(\hat{\beta}) = \mathbb{E}\Big[(\hat{\beta} - \mathbb{E}[\hat{\beta}])(\hat{\beta} - \mathbb{E}[\hat{\beta}])^T\Big] \tag{9}$$

We know $\hat{\beta} = (X^T X)^{-1} X^T y$ and $E[\hat{\beta}] = \beta$:

$$\begin{aligned} Var(\hat{\beta}) = \mathbb{E}\Big[((X^T X)^{-1} X^T y - \beta)((X^T X)^{-1} X^T y - \beta)^T\Big] \\ = \mathbb{E}\Big[((X^T X)^{-1} X^T y - \beta)(y^T X (X^T X)^{-1} - \beta^T)\Big] \\ = \mathbb{E}\Big[(X^T X)^{-1} X^T y y^T X (X^T X)^{-1} - \beta\beta^T - \beta\beta^T + \beta\beta^T\Big] \\ = [(X^T X)^{-1} X^T \mathbb{E}\Big[y y^T\Big] X (X^T X)^{-1} - \beta\beta^T \end{aligned} \tag{10}$$

From previously we know $E[y^2] = E[y]^2 + \sigma^2$:

$$\begin{aligned} (X^T X)^{-1} X^T \mathbb{E}\Big[y y^T\Big] X (X^T X)^{-1} - \beta\beta^T \\ = (X^T X)^{-1} X^T (X \beta\beta^T X^T + \sigma^2) X (X^T X)^{-1} - \beta\beta^T \\ = (X^T X)^{-1} X^T (X \beta\beta^T X^T + \sigma^2) X (X^T X)^{-1} - \beta\beta^T \\ = [(X^T X)^{-1} X^T X \beta\beta^T X^T + (X^T X)^{-1} X^T \sigma^2] X (X^T X)^{-1} - \beta\beta^T \\ = [\beta\beta^T X^T + (X^T X)^{-1} X^T \sigma^2] X (X^T X)^{-1} - \beta\beta^T \\ = \beta\beta^T X^T X (X^T X)^{-1} + (X^T X)^{-1} X^T X \sigma^2 (X^T X)^{-1} - \beta\beta^T \\ = [\beta\beta^T + \sigma^2((X^T X)^{-1}] - \beta\beta^T \\ = \sigma^2((X^T X)^{-1} \end{aligned} \tag{11}$$

1.2 Exercise 2

1.2.1 1

$$\begin{aligned} \mathbb{E}[\hat{\beta}_{Ridge}] = \mathbb{E}[(X^T X + \lambda I)^{-1} X^T y] \\ = (X^T X + \lambda I)^{-1} X^T \mathbb{E}[y] \\ = (X^T X + \lambda I)^{-1} X^T X \beta \end{aligned} \tag{12}$$

1.2.2 2

$$Var[\hat{\beta}_{Ridge}] = \mathbb{E}[\hat{\beta}^2_{Ridge}] - \mathbb{E}[\hat{\beta}_{Ridge}]^2 \tag{13}$$

We already know $\mathbb{E}[\hat{\beta}_{Ridge}] = (X^TX + \lambda I)^{-1}X^TX\beta$.

To make calculations easier, $Q = (X^TX + \lambda I)^{-1}$.

$$\mathbb{E}[\hat{\beta}_{Ridge}]^2 = (Q^{-1}X^TX\beta)(Q^{-1}X^TX\beta)^T = Q^{-1}X^TX\beta\beta^TX^TX(Q^{-1})^T \tag{14}$$

$$\begin{aligned}\mathbb{E}[\hat{\beta}^2_{Ridge}] &= \mathbb{E}[(Q^{-1}X^Ty)(Q^{-1}X^Ty)^T] \\ &= \mathbb{E}[Q^{-1}X^Tyy^TX(Q^{-1})^T] \\ &= Q^{-1}X^T\mathbb{E}[yy^T]X(Q^{-1})^T\end{aligned} \tag{15}$$

We know $\mathbb{E}[yy^T] = X\beta\beta^TX^T + \sigma^2$

$$\begin{aligned}Q^{-1}X^T\mathbb{E}[yy^T]XQ^{-1} &= Q^{-1}X^T(X\beta\beta^TX^T + \sigma^2)X(Q^{-1})^T \\ &= (Q^{-1}X^TX\beta\beta^TX^T + Q^{-1}X^T\sigma^2)X(Q^{-1})^T \\ &= Q^{-1}X^TX\beta\beta^TX^TX(Q^{-1})^T + Q^{-1}X^T\sigma^2X(Q^{-1})^T\end{aligned} \tag{16}$$

Which means:

$$\begin{aligned}&Var[\hat{\beta}_{Ridge}] \\ &= Q^{-1}X^TX\beta\beta^TX^TX(Q^{-1})^T + Q^{-1}X^T\sigma^2X(Q^{-1})^T - \mathbb{E}[\hat{\beta}_{Ridge}]^2 \\ &= Q^{-1}X^TX\beta\beta^TX^TX(Q^{-1})^T + Q^{-1}X^T\sigma^2X(Q^{-1})^T \\ &\quad - Q^{-1}X^TX\beta\beta^TX^TX(Q^{-1})^T \\ &= Q^{-1}X^T\sigma^2X(Q^{-1})^T \\ &= \sigma^2(X^TX + \lambda I)^{-1}X^TX((X^TX + \lambda I)^{-1})^T\end{aligned} \tag{17}$$