EXERCISE 1

a) $\mathbb{E}(y) = \mathcal{X}\beta$ (show this)

we are working with the (OLS)

$\varepsilon \sim N(0, \sigma^2)$

OLS : $y_i = \mathcal{X}\beta + \varepsilon_i$ since the expectation of ε_i is 0 $\longrightarrow$ $\mathbb{E}(y_i) = x_i\beta + \mathbb{E}(\varepsilon_i)$

$\mathbb{E}(y_i) = x_i\beta + 0$

$\boxed{\mathbb{E}(y_i) = x_i\beta}$

b) $Var(y_i) = \sigma^2$ (show this)

OLS : $y_i = \underbrace{\mathcal{X}\beta}_{\text{this is a fixed and deterministic term, so its variance is 0}} + \varepsilon_i$ $\longrightarrow$ $Var(y_i) = Var(x_i\beta + \varepsilon_i) = \boxed{Var(\varepsilon_i)} = \sigma^2$

c) $y_i \sim N(\mathcal{X}_i\beta, \sigma^2)$

follow a normal distribution, we also know that $\varepsilon_i \sim N(0, \sigma^2)$, so any linear combination of normally distributed variables is also normally distributed

d) $\mathbb{E}(\hat{\beta}) = \beta$

true parameter β

expectation of the OLS estimator

$$\hat{\beta} = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T y$$

$y = \mathcal{X}\beta + \varepsilon$

$$\hat{\beta} = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot (\mathcal{X}\beta + \varepsilon)$$

$$\hat{\beta} = \cancel{(\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot \mathcal{X}}\beta + (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot \varepsilon$$

$$\hat{\beta} = \beta + (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot \varepsilon$$

now we take the Expectation

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}(\beta + (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot \varepsilon)$$

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}(\beta) + \mathbb{E}((\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\varepsilon)$$

$$\mathbb{E}(\hat{\beta}) = \beta + \mathbb{E}((\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\varepsilon)$$

$$\mathbb{E}(\hat{\beta}) = \beta + (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot \underbrace{\mathbb{E}(\varepsilon)}_{0} \longrightarrow \mathbb{E}(\hat{\beta}) = \beta$$

e) $Ver(\hat{\beta}) = \sigma^2(\mathcal{X}^T\mathcal{X})^{-1}$ (show this)

$\hat{\beta} = \beta + (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \cdot \varepsilon$ (from the dem: $\mathbb{E}(\hat{\beta}) = \beta$)

is costunt so we remove it

we tuse the Verience

$Ver(\hat{\beta}) = Ver((\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\varepsilon)$

now we use the standard property of verience if A matrix and v vector:

$Ver(Az) = A\,ver(z)A^T$

$Ver(\hat{\beta}) = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\,Ver(\varepsilon)(\mathcal{X}^T(\mathcal{X}^T\mathcal{X})^{-1})^T$

$= (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\boxed{Ver(\varepsilon)} \cdot \mathcal{X}(\mathcal{X}^T\mathcal{X})^{-1}$

$\varepsilon \sim N(0, \sigma^2 I)$

$= (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\sigma^2\cancel{I}\mathcal{X}(\mathcal{X}^T\mathcal{X})^{-1}$

$= (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\sigma^2\mathcal{X}(\mathcal{X}^T\mathcal{X})^{-1}$

$= \sigma^2(\mathcal{X}^T\mathcal{X})^{-1}\cancel{\mathcal{X}^T\mathcal{X}}\cancel{(\mathcal{X}^T\mathcal{X})^{-1}}$

$Ver(\hat{\beta}) = \sigma^2(\mathcal{X}^T\mathcal{X})^{-1}$

EXERCISE 2

a) $\mathbb{E}(\hat{\beta}^{ridge}) = (\mathcal{X}^T\mathcal{X} + \lambda I_{pp})^{-1}(\mathcal{X}^T\mathcal{X})\beta$ (show that)

$\downarrow$

$$\hat{\beta}_{ridge} = (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T y$$

$$\mathbb{E}(\hat{\beta}_{ridge}) = \mathbb{E}((\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T y)$$

$$y = \mathcal{X}\beta + \varepsilon$$

$$\mathbb{E}(\hat{\beta}_{ridge}) = \mathbb{E}((\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\mathcal{X}\beta + \varepsilon) \quad \varepsilon \sim N(0, \sigma^2)$$

$$\mathbb{E}(\hat{\beta}_{ridge}) = \mathbb{E}((\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T(\mathcal{X}\beta))$$

$$\mathbb{E}(\hat{\beta}_{ridge}) = (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\mathcal{X}\beta \qquad \mathbb{E}(\hat{\beta}_{OLS}) = \beta$$

$\downarrow$ if $\lambda = 0$ (like in OLS)

$$\mathbb{E}(\hat{\beta}_{ridge}) = \cancel{(\mathcal{X}^T\mathcal{X})^{-1}}\,\cancel{\mathcal{X}^T\mathcal{X}}\beta$$

if $\lambda > 0$ we have: $\mathbb{E}(\hat{\beta}_{ridge}) \neq \mathbb{E}(\hat{\beta}_{OLS})$

$$\mathbb{E}(\hat{\beta}_{ridge}) = \mathbb{E}(\hat{\beta}_{OLS}) = \beta$$

b) $\hat{\beta}_{ridge} = (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T y$

$y = \mathcal{X}\beta + \varepsilon$

$$Var(\hat{\beta}_{ridge}) = \mathbb{E}((\hat{\beta}_{ridge} - \mathbb{E}(\hat{\beta}_{ridge}))\cdot(\hat{\beta}_{ridge} - \mathbb{E}(\hat{\beta}_{ridge}))^T) \quad \text{(general expression of variance)}$$

$$\mathbb{E}(\hat{\beta}_{ridge}) = (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\mathcal{X}\beta$$

$$\hat{\beta}_{ridge} - \mathbb{E}(\hat{\beta}_{ridge}) = (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T(\mathcal{X}\beta + \varepsilon) - (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\mathcal{X}\beta$$

$$= (\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\varepsilon$$

$$Var(\hat{\beta}_{ridge}) = \mathbb{E}[(\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\varepsilon\varepsilon^T\mathcal{X}(\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}]$$

$$\mathbb{E}(\varepsilon\varepsilon^T) = \sigma^2 I$$

$$Var(\hat{\beta}_{ridge}) = \sigma^2(\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}\mathcal{X}^T\mathcal{X}(\mathcal{X}^T\mathcal{X} + \lambda I)^{-1}$$