Exercises week 37

Exercise 1

i) Show that $\mathbb{E}(y_i) = \sum_j x_{ij} \beta_j$

From the standard linear model we have that

$$y_i = \underline{X}_{i,*} \underline{\beta} + \varepsilon_i$$

The expectation is has the linerity property. Thus

$$\mathbb{E}[y_i] = \mathbb{E}[\underline{X}_{i,*} \underline{\beta}] + \mathbb{E}[\varepsilon_i]$$

assuming $\varepsilon_i \sim N(0, \sigma^2)$ we get

$$\mathbb{E}[\varepsilon_i] \equiv 0$$

Thus we have reduced it to

$$\mathbb{E}[y_i] = \mathbb{E}[\underline{X}_{i,*} \underline{\beta}]$$

$$= \mathbb{E}\left[\sum_j x_{ij} \beta_j\right]$$

This sum is a constant and thus

$$\mathbb{E}[y_i] = \sum_j x_{ij} \beta_j$$

ii) Show that $\mathrm{var}(y_i) = \sigma^2$

The variance is just an expectation value itself defined as

$$\mathrm{var}(X) = E\left[\left(X - E(X) \right)^2 \right]$$

which by applying the property of linearity can be written as

$$\mathrm{Var}(X) = E(X^2) - E(x)^2$$

Thus

$$\mathrm{var}(y_i) = \mathbb{E}(y_i^2) - \mathbb{E}(y_i)^2$$

$$= E\left[\left(\sum_j x_{ij} \beta_j + \varepsilon_i \right)^2 \right] - E\left[\sum_j x_{ij} \beta_j \right]^2$$

$$= E\left[\left(\sum_j x_{ij} \beta_j \right)^2 + 2 \varepsilon_i \sum_j x_{ij} \beta_j + \varepsilon_i^2 \right]$$

$$-E\left[\sum_j x_{ij}\beta_j\right]^2$$

Using the linearity we get

$$=\left(\sum_j x_{ij}\beta_j\right)^2 + 0 + \mathbb{E}\left[\varepsilon_i^2\right] - \left(\sum_j x_{ij}\beta_j\right)^2$$

Since $\mathrm{Var}(\varepsilon_i) = E(\varepsilon_i^2) - \underbrace{E(\varepsilon_i)^2}_{\to 0}$
we get

$$\mathrm{Var}(y_i) = \mathrm{Var}(\varepsilon_i) = \sigma^2$$

since by defenition $\mathrm{var}(\varepsilon_1) = \sigma^2$

iii) Show that $E(\hat{\beta}) = \beta$

From earlier we know $\hat{\beta} = (X^TX)^{-1}X^T\underline{y}$

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

Further $\underline{y} = X\underline{\beta}$ and thus we are taking the expection of only constants. Therefore we can pull them out and

$$\mathbb{E}(\underline{\hat{\beta}}) = \underbrace{(X^TX)^{-1}X^TX}_{1}\underline{\beta} \Rightarrow E(\hat{\beta}) = \underline{\beta}$$

iv) Show that $Var(\hat{\underline{\beta}}) = \sigma^2(X^TX)^{-1}$

$$Var(\hat{\underline{\beta}}) = \mathbb{E}\left[\left(\hat{\underline{\beta}} - \mathbb{E}(\hat{\underline{\beta}})\right)^2\right],$$

Bruker $^{(1)}\underline{a}^2 = \underline{a}\,\underline{a}^T$, $^{(2)}\mathbb{E}(\hat{\underline{\beta}}) = \underline{\beta}$, $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$ for videre utregninger

$$Var(\hat{\underline{\beta}}) = \mathbb{E}(\hat{\underline{\beta}}^2) - (\mathbb{E}(\hat{\underline{\beta}}))^2$$

$$= \mathbb{E}\left[(X^TX)^{-1}X^T\underline{y}\left((X^TX)^{-1}X^T\underline{y}\right)^T\right] - \underline{\beta}\beta^T$$

We have now used (1) and (2). We now pull out the constans from the expection and are left with

$$= (X^TX)^{-1}X^T\,\mathbb{E}(\underline{y}\,\underline{y}^T)\,X(X^TX)^{-1} - \underline{\beta}\,\underline{\beta}^T$$

Using $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ we get

$$\underline{y}\,\underline{y}^T = (X\underline{\beta} + \underline{\varepsilon})(X\underline{\beta} + \underline{\varepsilon})^T$$

$$= X\underline{\beta}X^T\underline{\beta}^T + X\underline{\beta}\,\underline{\varepsilon}^T + \underline{\varepsilon}X^T\underline{\beta}^T + \varepsilon\varepsilon^T$$

Taking the expectation we aknowledge that $\varepsilon_i \sim N(0, \sigma^2)$

thus term two and three has expectation 0 further we use $\mathbb{E}(\underline{\varepsilon}\,\underline{\varepsilon}^T) = \sigma^2 I$ yielding

$$\mathbb{E}(\underline{y}\,\underline{y}^T) = X\underline{\beta}\,\underline{\beta}^T X^T + \sigma^2 I$$

and

$$= (X^TX)^{-1}X^T\left(X\underline{\beta}\,\underline{\beta}^T X^T + \sigma^2 I\right)X\,(X^TX)^{-1} - \underline{\beta}\,\underline{\beta}^T$$

$$= \cancel{(X^TX)^{-1}}\,\cancel{X^TX}\,\underline{\beta}\,\underline{\beta}^T\,\cancel{X^TX}\,\cancel{(X^TX)^{-1}}$$

$$+ (\cancel{X^T}X)^{-1}\cancel{X^T}\sigma^2 I\,\cancel{X}(X^T\cancel{X})^{-1} - \underline{\beta}\,\underline{\beta}^T$$

$$= \cancel{\underline{\beta}\,\underline{\beta}^T} + \sigma^2(X^TX)^{-1} - \cancel{\underline{\beta}\,\underline{\beta}^T}$$

$$= \sigma^2(X^TX)^{-1} \qquad \blacksquare$$

Exercise 2

i) show that $E(\hat{\beta}^{ridge}) = (X^TX - \lambda I)^{-1}(X^TX)\beta^{ols}$

$$E(\hat{\beta}^{ridge}) = E\left[(X^TX + \lambda I)^{-1}X^T\underline{y}\right]$$

As earlier $\underline{y} = X\beta^{OLS}$ thus

$$= E\left[(X^TX + \lambda I)^{-1}(X^TX)\,\beta^{OLS}\right]$$

All of these are const. therfore:

$$E(\hat{\beta}^{ridge}) = (X^TX - \lambda I)^{-1}X^TX\,\beta^{ols}$$

(For derivations in ii) I define:

$$W_\lambda = (X^TX + \lambda I)^{-1}X^TX \qquad (*)$$

giving us $\mathbb{E}(\hat{\beta}^{ridge}) = W_\lambda \underline{\beta}^{OLS}$)

ii) Show that

$$\mathrm{Var}(\hat{\beta}^{ridge}) = \sigma^2 (X^TX+\lambda I)^{-1} X^TX \left[(X^TX+\lambda I)^{-1}\right]^T$$

Using W_λ defined in i) we have

$$\mathrm{Var}(\hat{\underline{\beta}}^{ridge}) = \mathrm{Var}(W_\lambda \hat{\underline{\beta}}^{OLS})$$

using $\mathrm{Var}(A\underline{x}) = A\,\mathrm{var}(\underline{x})A^T$ we get:

$$\mathrm{Var}(\hat{\beta}^{ridge}) = W_\lambda \mathrm{Var}(\hat{\underline{\beta}}^{OLS}) W_\lambda^T$$

From 1. iv) we have $\mathrm{Var}(\hat{\underline{\beta}}^{OLS}) = \sigma^2(X^TX)^{-1}$ yielding

$$= W_\lambda \sigma^2 (X^TX)^{-1} W_\lambda^T$$

since σ^2 is a cons we pull it out of the matrix product. Further subbing (*) for W_λ we get

$$= \sigma^2 (X^TX-\lambda I)^{-1} \cancel{X^TX(X^TX)^{-1}} \left((X^TX-\lambda I)^{-1}X^TX\right)^T$$

$$= \sigma^2 (X^TX-\lambda I)^{-1} X^TX \left((X^TX-\lambda I)^{-1}\right)^T \quad \square$$