W37 + Part D)

Our assumptions:

$f(x) \Rightarrow$ continous function with error ϵ which is like $\epsilon \sim N(0, \sigma^2)$

These describe our data like

$y = f(x) + \epsilon$

Design matrix.

y is approximated like $\tilde{y} = X\beta$

Show expectation value for y for given element i:

We have our linear regression model:

$y = X\beta + \epsilon$

y = model-output
X = design matrix
β = Vector of regression coefficients
ϵ = error.

For a given observation i, we know that element y_i is:

$y_i = \sum_j x_{ij} \beta_j + \epsilon$

x_{ij} is the jth predictor for the i-th observation.

We now take the expectation of the same element y_i:

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

Using linearity to split:

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

We know from our assumption that ϵ has zero mean.

And $x_{ij}\beta_j$ is deterministic (no randomness)

This means:

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

Therefor

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

Using matrix notation since $x_{ij}\beta_j$ is the dot product of the i-th row of X with β.

$$x_{ij}\beta_j = X_{i,*}\beta \Rightarrow \underline{\mathbb{E}[y_i] = \sum_{ij}\beta_j = X_{i,*}\beta}.$$

Show: $Var(y_i) = \sigma^2$

Using the same assumptions

Have our regression model (linear).

$y = X\beta + \epsilon$

As illustrated earlier:

$y_i = \sum_j x_{ij} \beta_j + \epsilon_i,$

$\sum x_{ij} \beta_j$ is deterministic term, which result in 0 variance. Therefer the $var(y_i)$ is only influenced by ϵ. ϵ is independent of the predictor.

$Var(y_i) = var(\overbrace{\sum x_{ij} \beta_j}^{\to 0}) + Var(\epsilon_i)$

$var(y_i) = var(\epsilon_i)$

From assumption $\epsilon \sim N(0, \sigma^2)$

$Var(\epsilon) = \sigma^2$

$\underline{var(y_i) = \sigma^2}$

Show $\mathbb{E}[\hat{\beta}] = \beta$ for OLS.

For OLS, we have

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

Our linear model say:

$$y = X\beta + \epsilon \qquad (2)$$

Substitute y in expression (1)

$$\hat{\beta} = (X^T X)^{-1} X^T (X\beta + \epsilon)$$

$$= (X^T X)^{-1} X^T X \beta + X^T \epsilon (X^T X)^{-1}$$

We notice that

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

$$\hat{\beta} = I\beta + (X^T X)^{-1} X^T \epsilon$$

Taking the expection:

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[\beta + (X^T X)^{-1} X^T \epsilon]$$

Expected value of 0, so it makes the whole expression vanish.

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[\underbrace{\beta}_{\text{Deterministic}} + 0]$$

$$\mathbb{E}[\hat{\beta}] = \beta$$

Show $\text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$

As stated:

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

Substitute the linear regression model $y = X\beta + \epsilon$ in.

$$\hat{\beta} = (X^T X)^{-1} X^T (X\beta + \epsilon)$$

Expand:

$$\hat{\beta} = (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \epsilon$$

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

$$\Rightarrow \hat{\beta} = \beta + (X^T X)^{-1} X^T \epsilon$$

$$\text{Var}(\hat{\beta}) = \text{Var}(\overset{=0}{\beta} + (X^T X)^{-1} X^T \epsilon)$$

- $(X^T X)^{-1} X^T$ is a constant matrix and $\text{Var}(A\epsilon) = A\,\text{var}(\epsilon) A^T$ for any matrix A, so we factor out our expression.

$$\text{Var}(\hat{\beta}) = (X^T X)^{-1} X^T \,\text{Var}(\epsilon)\, X (X^T X)^{-1}$$

- $\text{Var}(\epsilon) = \sigma^2 I$

$$\text{Var}(\hat{\beta}) = \sigma^2 \underbrace{(X^T X)^{-1} X^T X}_{I} (X^T X)^{-1}$$

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