Problem Sheet 05 - Exercise 2.

$y = X\beta + \epsilon$; y, ϵ ~~are~~ vector notation

$y \in \mathbb{R}^{n\times 1}$, $X \in \mathbb{R}^{n\times(p+1)}$, $\beta \in \mathbb{R}^{(p+1)\times 1}$, $\epsilon \in \mathbb{R}^{n\times 1}$

$\hat{\beta} = (X^TX)^{-1}X^Ty$, if X has full column rank $(p+1)$ (See Lecture 10, p.: 11)

$= (X^TX)^{-1}X^T(X\beta + \epsilon)$ { Since X is a full rank matrix X^TX positive semidefinite and invertible (nonsingular).

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

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

1) $\mathbb{E}[\hat{\beta}] = \mathbb{E}[\beta + (X^TX)^{-1}X^T\epsilon]$; (By using linearity of expectation which we already prove PS1Ex1)

$= \mathbb{E}[\beta] + \mathbb{E}[(X^TX)^{-1}X^T\epsilon]$; (Since in regression we condition on X $(X^TX)^{-1}X^T$ deterministic matrix

$= \beta + (X^TX)^{-1}X^T\mathbb{E}[\epsilon]$ and β vector as unknown param. are constants.)

$= \beta + 0$ ($\mathbb{E}[\epsilon] = 0$ from lecture notes as mentioned it's i.i.d)

$= \beta$

$Var(x_i) = \mathbb{E}[(x_i - \mathbb{E}[x_i])^2]$, $Var(\epsilon_i) = \sigma^2$

$Cov(x_i, x_j) = \mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$

$Cov(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$

$Var(\epsilon) = \mathbb{E}[\epsilon^2] - \mathbb{E}[\epsilon]^2 = \sigma^2 - 0^2 = \sigma^2$

2) $Cov(\hat{\beta}) = \mathbb{E}[(\hat{\beta} - \mathbb{E}[\hat{\beta}])(\hat{\beta} - \mathbb{E}[\hat{\beta}])^T]$; (From Ex1: $\mathbb{E}[\hat{\beta}] = \beta$

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

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

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

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

$$= (X^TX)^{-1}X^T\, \mathbb{E}[\epsilon\epsilon^T]\, X(X^TX)^{-1} \quad ; (\text{As } \epsilon x 1).$$

$\mathbb{E}[\epsilon\epsilon^T] = \sigma^2 I_n$ (also see: Gauss-Markov Theorem)

$$\mathbb{E}[\epsilon\epsilon^T] = \mathbb{E}\left[\begin{bmatrix}\epsilon_1\\ \vdots \\ \epsilon_n\end{bmatrix}\begin{bmatrix}\epsilon_1 \dots \epsilon_n\end{bmatrix}\right] = \mathbb{E}\begin{bmatrix}\epsilon_1^2 & \dots & \epsilon_1\epsilon_n \\ \vdots & \ddots & \vdots \\ \epsilon_n\epsilon_1 & \dots & \epsilon_n^2\end{bmatrix} = \begin{bmatrix}\mathbb{E}[\epsilon_1^2] & \dots & \mathbb{E}[\epsilon_1\epsilon_n] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\epsilon_n\epsilon_1] & \dots & \mathbb{E}[\epsilon_n^2]\end{bmatrix}$$

Since $Var(\epsilon_i, \epsilon_j) = 0$ (it's not autocorrelated, it's i.i.d) $\mathbb{E}[\epsilon_i\epsilon_j] = 0$

$$\Rightarrow \begin{bmatrix}\sigma^2 & \dots & 0 \\ \vdots & \sigma^2 & \\ & & \ddots \\ 0 & & \sigma^2\end{bmatrix} = \sigma^2 . I_n$$

$$Cov(\hat\beta) = \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1} \quad ; ((X^TX)^{-1}(X^TX) = I \text{ as shown } 1)$$

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

---

3) $$\mathbb{E}[\hat\sigma^2_{ad}] = \mathbb{E}\left[\frac{1}{n-p-1}\hat{e}^T\hat{e}\right] = \frac{1}{n-p-1}\mathbb{E}[\hat{e}^T\hat{e}]$$

$$= \frac{1}{n-p-1}\mathbb{E}[(y-X\hat\beta)^T(y-X\hat\beta)] \quad ; \hat\beta = (X^TX)^{-1}X^Ty$$

$$= \frac{1}{n-p-1}\mathbb{E}[(y-X(X^TX)^{-1}X^Ty)^T(y-X(X^TX)^{-1}X^Ty)] \quad ; (X(X^TX)^{-1}X^T = H)$$

$$= \frac{1}{n-p-1}\mathbb{E}[(y-Hy)^T(y-Hy)]$$

$$\boxed{\begin{aligned}&\mathbb{E}[x^TAx] = tr(A\Sigma) + \mu^TA\mu \\ &Cov(x) = \Sigma \quad \mathbb{E}[x] = \mu\end{aligned}}$$

$$= \frac{1}{n-p-1}\mathbb{E}[y^T(I_n - H)^T(I_n-H)y]$$

$$= \frac{1}{n-p-1}[tr((I_n-H)\sigma^2 I_n) + \beta^TX^T(I_n-H)X\beta$$

$$= \frac{1}{n-p-1}[(tr(I_n) - tr(H))\sigma^2 + \beta^TX^T(I_n - X(X^TX)^{-1}X^T)X\beta]$$

$$= \frac{1}{n-p-1}[(n - rk(H))\sigma^2 + \beta^TX^TX\beta - \beta^T\underbrace{X^TX(X^TX)^{-1}}_{I}X^TX\beta]$$

$$= \frac{1}{n-p-1}[(n-(p+1))\sigma^2 + \beta^TX^TX\beta - \beta^TX^TX\beta]$$

$$= \frac{1}{n-p-1}(n-p-1)\sigma^2 = \sigma^2$$

Since H is idempotent $tr(H) = rk(H) = p+1$ | Lecture 10, p.15