# Q2.

(a) $E(\varepsilon) = 0$.

$E(\varepsilon) = E(Y) - E(\beta_0) - E(\beta^T X) = E(Y) - (\beta_0 + \beta^T E(X))$

$\beta_0 = E(Y) - \beta^T E(X)$ . So. $E(\varepsilon) = E(Y) - [E(Y) - \beta^T E(X)] - \beta^T E(X)$

(b). $\operatorname{cov}(\varepsilon, X) = 0$ $\qquad \varepsilon = Y - \beta_0 - \beta^T X$

$= \operatorname{cov}(Y, X) - \operatorname{cov}(\beta_0, X) - \beta^T \operatorname{cov}(X, X)$ $\qquad \beta_0$ constant.

$= \operatorname{cov}(Y, X) - \beta^T \operatorname{var}(X)$ $\qquad \beta = \operatorname{Var}(X)^{-1} \operatorname{cov}(X, Y)$

$= \operatorname{cov}(Y, X) - [\operatorname{var}(X)^{-1} \operatorname{cov}(X, Y)]^T \operatorname{var}(X)$

$= \operatorname{cov}(Y, X) - \operatorname{cov}(X, Y)^T = 0$.

(c) No. $\operatorname{cov}(\varepsilon, X) = 0$ does not mean they are independent.

$\operatorname{var}(\varepsilon | X) = \operatorname{var}(Y - \beta_0 - \beta^T X | X)$

$= \operatorname{Var}(Y) - 0 - 0 = \operatorname{var}(Y)$

Since we don't have $\varepsilon$, $X$ independent, so that $\operatorname{var}(\varepsilon | X)$ is not constant with $X$.

(d). $E(\hat{\beta} | X^n)$

$= (X^T X)^{-1} X^T E(\vec{Y} | X^n)$ $\qquad \vec{Y} = \vec{\varepsilon} + X\beta$

$= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T E(\vec{\varepsilon} | X^n)$

$= \beta + (X^T X)^{-1} X^T E(\vec{\varepsilon} | X^n)$

So we can see $E(\hat{\beta} | X^n) = \beta$ only when $E(\vec{\varepsilon} | X^n) = 0$.

(e) $\operatorname{var}(\hat{\beta} | X^n)$ $\qquad \operatorname{var}(\vec{\varepsilon} | X^n) = \Sigma$

$= \cancel{(X^T X)^{-1} X^T X X^T Y}\ \operatorname{var}((X^T X)^{-1} X^T \vec{Y} | X^n)$

$= (X^T X)^{-1} X^T \operatorname{var}(\vec{Y} | X^n) X (X^T X)^{-1}$

$= (X^T X)^{-1} X^T \cdot (\operatorname{var}(\vec{\varepsilon} | X^n) + \underbrace{\cancel{\operatorname{var}(X\beta | X^n)}}_{0} + \underbrace{\cancel{2 \cdot \operatorname{cov}(\vec{\varepsilon}, X\beta | X^n)}}_{0}) X (X^T X)^{-1}$

$= (X^T X)^{-1} X^T \cdot \Sigma X (X^T X)^{-1}$

So when $\Sigma = \sigma^2 I$, $\operatorname{var}(\hat{\beta} | X^n) = (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1}$

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