Stat 415 Assignment 3.

1. (a). True. $E[XY] = \iint xy P_{x,y}(x,y)\,dx\,dy = \int x P_x(x)dx \int y P_y(y)dy = E[X]E[Y]$

(b). False $E[\varepsilon|x] = 0$. $E[E[\varepsilon|x]] = E[\varepsilon] = E[0] = 0$, we can just prove that.

$\mathrm{cov}(\varepsilon, x) = E(\varepsilon x) - E(\varepsilon)E(x) = E(x)E(\varepsilon|x) - E(\varepsilon)E(x) = 0$ However, even though u and x are uncorrelated, they can still be dependent. E.g $x \sim U(-1,1)$, $\varepsilon = -x$

Then $E[\varepsilon|x] = E(\varepsilon) = \int \varepsilon P(\varepsilon)\,d\varepsilon = 0$

(c) True Already Proved in (b)

(d) True $P_x = x(x^T x)^{-1} x^T$, It is obvious that $x^T x$ is symmetric, i.e $[(x^T x)^{-1}]^T = (x^T x)^{-1}$

$P_x^T = [x(x^T x)^{-1} x^T]^T = (x^T)^T (x^T x)^{-1} x^T = x(x^T x)^{-1} x^T = P_x$; Hence. Q.E.D

(e) True. $(I - P_x)^2 = (I - P_x)(I - P_x) = I^2 - P_x I - I P_x + P_x^2 = I - P_x - P_x + P_x^2$

$P_x^2 = x(x^T x)^{-1} x^T x (x^T x)^{-1} x^T = x\underbrace{[(x^T x)^{-1}[x^T x]]}_{=I}(x^T x)^{-1} x^T = x(x^T x^{-1}) x^T = P_x$

Hence $(I - P_x)^2 = I - P_x$ we can easily conclude that $(I - P_x)^{100} = I - P_x$

(f). False. If $y_i = x^T\beta + \varepsilon_i$ where ε_i is i.i.d; then the fitted values under OLS is

$\hat{y} = P_x y = x(x^T x)^{-1} x^T y$; Then the residuals are $y - \hat{y} = (I - P_x)y$

The Covariance matrix of the residuals is.

$\mathrm{Cov}((I - P_x)y) = (I - P_x)\mathrm{Cov}(y)(I - P_x)^T = (I - P_x)(\sigma^2 I)(I - P_x)^T = \sigma^2(I - P_x)$

Which is not diagonal, so the residuals might not be identically distributed

(g) True $\hat{\varepsilon} = y - \hat{y} = (I - P_x)Y$

Then $\hat{\varepsilon}^T x = [(I - P_x)Y]^T x = Y^T(I - P_x)^T x = Y^T(I - P_x)X = Y^T(I - X(x^T x)^{-1} x^T)X$

$= Y^T[X - X(x^T x)^{-1}(x^T x)] = Y^T[X - X] = 0$

2 (a) We know that $\hat{\beta} = (x^T x)^{-1} x^T y = \beta + (x^T x^{-1}) x^T \varepsilon$

Then $\mathrm{cov}(\hat{\beta}) = \mathrm{var}(\hat{\beta}|x) = \mathrm{Var}(\beta|x + (x^T x^{-1}) x^T \varepsilon|x) = \mathrm{Var}(\beta|x) + \mathrm{Var}((x^T x^{-1}) x^T \varepsilon|x)$

$= (x^T x^{-1}) X^T \mathrm{var}(\varepsilon|x)[(x^T x)^{-1} x^T]^T$

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

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

(b). Consider the vector $Z = (Z_j, Z_{oj})$. Z_{nj} for the jth feature, Its response is calculated by.

Hence. $\hat{Y}_j = Z^T(x^T x)^{-1} x^T Y$, the prediction is unbiased

Since $E(\hat{Y}_j) = E(Z^T)E((x^T x)^{-1} x^T Y) = \underbrace{E(Z^T)}_{= Z^T} \underbrace{E(\hat{\beta})}_{= \beta} = Z^T\beta = E(Y_j)$

(c) $\mathrm{Var}(\hat{Y}_j) = \sigma^2 Z^T (x^T x)^{-1} Z = Z^T \sigma^2(\hat{\beta}) Z$

$s^2(\hat{Y}_j) = MSE(Z^T(x^T x)^{-1} x) = Z^T s^2(\hat{\beta}) Z$

The $1-\alpha$ confidence interval of $E(\hat{Y}_j)$ is given by. $\hat{Y}_j \pm t(1 - \alpha/2; n-p)\, s(\hat{Y}_j)$