STAT 245 HW5 Solution.

1. (a) $(I-P)^2 = I - 2P + P\cdot P = I - P$

$(I-P)P = P - P\cdot P = 0.$

(b) eigenvalues of P can only be 0's or 1's.

if P is a projection matrix in $\mathbb{R}^{n\times n}$.

$\because \lambda^2 v = P^2 v = Pv = \lambda v$ and $\because v \neq 0$. $\therefore \lambda^2 = \lambda$

$\therefore \lambda = 0$ or $\lambda = 1$.

$\because$ rank$(P) = r$. If we do eigendecomposition $P = U^T \Lambda U$

$\Lambda = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ & & & & 0 \end{bmatrix} \}r$ $\therefore$ there is $\lambda = 1$ with multiplicity r. and 0 with multiplicity $(n-r)$.

2. (a). $\left(X(X^TX)^{-1}X^T\right)^T = X(X^TX)^{-1}X^T$ (This is showing orthogonal not necessary).

$[X(X^TX)^{-1}X^T]\cdot[X(X^TX)^{-1}X^T] = X(X^TX)^{-1}X^T$. $(P=P^2)$

By definition, it is a projection matrix.

(b). $\left(n^{-1}\cdot 1\cdot 1^T\right)^T = n^{-1}\cdot 1\cdot 1^T$

$n^{-1}\, 1\cdot 1^T\cdot n^{-1}\cdot 1\cdot 1^T = n^{-2}\cdot 1\cdot(1^T 1)\cdot 1^T = n^{-2} n^{1}\, 1 1^T = n^{-1} 1 1^T$

By definition it is a projection matrix.

2 (c). If P_1, P_2 are symmetric, then obviously (P_1+P_2) is also symmetric

And $(P_1+P_2)(P_1+P_2) = P_1^2 + P_2P_1 + P_1P_2 + P_2^2$

$= P_1 + (P_1^T \cdot P_2^T)^T + P_1 \cdot P_2 + P_2$

$= P_1 + (P_1 \cdot P_2)^T + P_1 \cdot P_2 + P_2 = P_1 + 0 + 0 + P_2 = P_1 + P_2$.

3. $\because P \in \mathbb{R}^{n \times n}$ is a real ~~orthonormal~~ projection matrix, $\text{rank}(P) = r$

then $\exists$ Q is an orthonormal matrix and $\Lambda = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \ 0 \end{bmatrix}\}r$

s.t. $P = Q \Lambda Q^T$.

$z^T P z = z^T Q \Lambda Q^T z$. let $z^T Q = y^T$ then $y \sim N(0, Q^T I_n Q)$

$\therefore y \sim N(0, I_n)$

then $y^T \Lambda y = y^T \begin{bmatrix} I_r & \\ & 0 \end{bmatrix} y = \sum_{i=1}^{r} y_i^2 \sim \chi_r^2$.

4. $E(\hat{\theta} - \theta)^2 = E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2$

$= E[(\hat{\theta} - E(\hat{\theta}))^2 + 2(\hat{\theta} - E\hat{\theta})(E\hat{\theta} - \theta) + (E\hat{\theta} - \theta)^2]$

$= E(\hat{\theta} - E\hat{\theta})^2 + 2 \cdot E[(E\hat{\theta} - \theta) \cdot (\hat{\theta} - E\hat{\theta})] + (E\hat{\theta} - \theta)^2$

$= E(\hat{\theta} - E\hat{\theta})^2 + 2 (E\hat{\theta} - \theta) \cdot (E\hat{\theta} - E\hat{\theta}) + (E\hat{\theta} - \theta)^2$

$= \text{Var}\,\hat{\theta} + 0 + E(\hat{\theta} - \theta)^2$.

5. (a) $E\hat{\sigma}^2 = E\left(\frac{1}{n-p}\|\hat{e}\|^2\right) = \frac{1}{n-p}E(\|\hat{e}\|^2) = \frac{n-p}{n-p}\cdot\sigma^2 = \sigma^2$.

(b) $E(\hat{\sigma}^2-\sigma^2)^2$ is the MSE, from question 4:

$$E(\hat{\sigma}^2-\sigma^2)^2 = Var\,\hat{\sigma}^2 + (E\hat{\sigma}^2-\sigma^2)^2$$

$$= \frac{\sigma^4\cdot 2(n-p)}{(n-p)^2} + 0 = \frac{2\sigma^4}{n-p}$$

(c). $$E(\tilde{\sigma}^2-\sigma^2)^2 = Var\,\tilde{\sigma}^2 + (E\tilde{\sigma}^2-\sigma^2)^2$$

$$= \frac{2\sigma^4(n-p)}{n^2} + \left(\frac{(n-p)\sigma^2}{n} - \sigma^2\right)^2$$

$$= \frac{2(n-p)\sigma^4}{n^2} + \frac{p^2}{n^2}\sigma^4$$

$$= \frac{(2n-2p+p^2)\sigma^4}{n^2}$$

(d). $\sigma_c^2 = c\|\hat{e}\|^2$

$$E(\sigma_c^2-\sigma^2)^2 = Var\,\sigma_c^2 + (E\sigma_c^2-\sigma^2)^2$$

$$= 2c^2\cdot(n-p)\sigma^4 + [c(n-p)\sigma^2-\sigma^2]^2$$

Let $$\frac{\partial E(\sigma_c^2-\sigma^2)^2}{\partial c} = 4c\sigma^4(n-p) + 2\sigma^4\cdot(c(n-p)-1)(n-p) = 0$$

$$\Rightarrow \quad c = \frac{1}{n-p+2}$$

(e) choose (d) since it gives minimized MSE; (a) is unbiased, all these are assymptotically equivalent. Give partial credits for reasonable answers.

6.

(a) $f(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-X\beta)' \cdot (\sigma^2 I_n)^{-1} \cdot (y-X\beta)\right)$

$(\Sigma = \sigma^2 I_n)$

$$\ell(y) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2} \cdot \|y - X\beta\|_2^2$$

$$\frac{\partial \ell}{\partial \beta} = 0 \quad \Rightarrow \quad \hat{\beta} = (X^T X)^{-1} X^T y$$

plug in $\hat{\beta}$ and take $\frac{\partial \ell}{\partial \sigma^2} = 0$

$$\hat{\sigma}^2 = \frac{1}{n} y^T (I - X(X^T X)^{-1} X^T) y$$

(b) $\hat{\sigma}^2 = \frac{\|\hat{e}\|^2}{n}$ where $\|\hat{e}\|^2/\sigma^2 \sim \chi^2_{n-p}$.

$\therefore \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$ is a pivotal.

To find 95% CI:

$$\chi^2_{n-p,\,0.025} \le \frac{n\hat{\sigma}^2}{\sigma^2} \le \chi^2_{n-p,\,0.975}$$

$$95\%\ \text{CI}: \left[\frac{n\hat{\sigma}^2}{\chi^2_{n-p,0.975}},\ \frac{n\hat{\sigma}^2}{\chi^2_{n-p,0.025}} \right]$$

7.

(a). A paragraph of description about the data set.
What variables are their, how are they distributed?
Do they have correlations? Are there missing data
and very obvious outliers? It is nice to provide the
summary table of the data set, and some relavent figures
but you need to have a written paragraph for this question.

(b) ① State which subsets of rows selected.
② Why you selected these rows, corresponds to what question?

(c) ① If you used any variable selection methods, state
them. (Forward / Backward selection, etc.)

② Write out the linear model.

(d) ① Summary of the linear model. ~~How~~ evaluate it.
The R^2 is important to see how much of the variance
is explained and the p-values shows how ~~are the~~
significant are the variables selected. Also the coefficients
tells the relationships between the covariates and the
response.

② ~~Use~~ cooks distance to find the influential points,

carefully remove outliers with some reasoning.

refit the model after removing outliers.

③ There should be a summary on the final model after removing all the outliers.

(e) Conclusions for the final model:

① to have good prediction, R^2 need to be small.

② explain what features are significant in the model.

Does it make sense in the real world.