Homework 7 Solutions

1-(a). We have learned that

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} = \frac{\|y - X\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{n-p}.$$

By the CLT, we have

$$\frac{\chi^2_{n-p} - (n-p)}{\sqrt{2(n-p)}} \rightsquigarrow N(0,1).$$

Hence,

$$\sqrt{n-p}(\hat{\sigma}^2 - \sigma^2) \sim \frac{\chi^2_{n-p} - (n-p)}{\sqrt{n-p}}\sigma^2 \rightsquigarrow N(0, 2\sigma^4).$$

1-(b). By the delta method (variance-stabilizing transformation), we have

$$\sqrt{n-p}(g(\hat{\sigma}^2) - g(\sigma^2)) \rightsquigarrow N(0, 2\sigma^4|g'(\sigma^2)|^2).$$

Hence, we want $g'(u) = \frac{1}{u\sqrt{2}}$. Therefore, we let $g(u) = \frac{\log(u)}{\sqrt{2}}$.

1-(c). As we have

$$\sqrt{\frac{n-p}{2}}(\log(\hat{\sigma}^2) - \log(\sigma^2)) \rightsquigarrow N(0,1),$$

we have for large n

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{\frac{n-p}{2}}(\log(\hat{\sigma}^2) - \log(\sigma^2)) \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

Accordingly,

$$\mathbb{P}\left(\hat{\sigma}^2 \exp\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{2}{n-p}}\right) \leq \sigma^2 \leq \hat{\sigma}^2 \exp\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{2}{n-p}}\right)\right) \approx 1 - \alpha.$$

Hence, an approximate $(1-\alpha)$-confidence interval for σ^2 is

$$\left[\hat{\sigma}^2 \exp\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{2}{n-p}}\right), \hat{\sigma}^2 \exp\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{2}{n-p}}\right)\right].$$

2-(a). We have $\hat{\beta}_0 = \bar{y} = \sum_{i=1}^n y_i/n$. As $\mathbb{E}(\bar{y}) = \beta_0$, we have

$$\mathbb{E}(\hat{\beta}_0 - \beta_0)^2 = \mathbb{E}(\bar{y} - \beta_0)^2 = \mathrm{var}(\bar{y}) = \frac{\sigma^2}{n}.$$

2-(b). We still have the usual estimators

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

In this case, we can treat $y_i \sim N(\beta_0, \sigma^2)$ as $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ with $\beta_1 = 0$. From this, one can notice that we still have the usual results, namely,

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad \text{var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Hence,

$$\mathbb{E}(\hat{\beta}_0 - \beta_0)^2 = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

2-(c). While both estimators are unbiased, the MSE of (a) is smaller than that of (b) if $\bar{x} \neq 0$, meaning that a simpler model is better in terms of estimation accuracy.

3-(a). We have $\hat{\beta}_0 = \bar{y} = \sum_{i=1}^n y_i / n$. Hence,

$$\mathbb{E}(\hat{\beta}_0 - \beta_0)^2 = \mathbb{E}(\bar{y} - \beta_0)^2 = \text{var}(\bar{y}) + (\mathbb{E}(\bar{y}) - \beta_0)^2 = \frac{\sigma^2}{n} + \beta_1^2 \bar{x}^2,$$

where we use

$$\mathbb{E}(\bar{y}) = \beta_0 + \beta_1 \frac{\sum_{i=1}^n x_i}{n} = \beta_0 + \beta_1 \bar{x}.$$

3-(b). We have learned that

$$\mathbb{E}(\hat{\beta}_0 - \beta_0)^2 = \text{var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

3-(c). While both estimators are unbiased, the MSE of (b) is smaller than that of (a) if

$$\bar{x} \neq 0 \quad \text{and} \quad \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} < \beta_1^2.$$

Hence, in this case, the more complicated model is better in terms of estimation accuracy.

4-(a). Under H_0, we have $y \sim N(X_{S^c}\beta_{S^c}, \sigma^2 I_n)$. Hence, under H_0, we have

$$\hat{\beta}_{H_0} \sim N(\beta_{S^c}, \sigma^2 (X_{S^c}^T X_{S^c})^{-1}).$$

4-(b). It suffices to show that

$$(y-\hat{y})^T(\hat{y}-\hat{y}_{H_0}) = y^\top(I_n - X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T - X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T)y = 0.$$

We have

$$(I_n - X(X^TX)^{-1}X^T)X(X^TX)^{-1}X^T = 0.$$

Meanwhile,

$$X(X^TX)^{-1}X^TX_{S^c} = X_{S^c},$$

which follows from the fact that $X(X^TX)^{-1}X^T$ is a projection matrix to the column space of X which obviously contains the columns of X_{S^c}. Hence,

$$X(X^TX)^{-1}X^TX_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T = X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T.$$

Therefore,

$$(I_n - X(X^TX)^{-1}X^T)X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T = X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T - X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T = 0.$$

4-(c). From the above calculation, we have derived that $y-\hat{y} = Ay$ and $\hat{y}-\hat{y}_{H_0} = By$ with $AB = 0$ for symmetric matrices

$$\begin{aligned} A &= I_n - X(X^TX)^{-1}X^T, \\ B &= X(X^TX)^{-1}X^T - X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T. \end{aligned}$$

Then, from $\mathrm{Cov}(y-\hat{y}, \hat{y}-\hat{y}_{H_0}) = \mathrm{Cov}(Ay, By) = \sigma^2 AB = 0$, we conclude that $y-\hat{y}$ and $\hat{y}-\hat{y}_{H_0}$ are independent.

4-(d). Under H_0, we have $y \sim N(X_{S^c}\beta_{S^c}, \sigma^2 I_n)$, which allows us to reuse the standard results with $X_{S^c} \in \mathbb{R}^{n\times(p-s)}$ in the place of $X \in \mathbb{R}^{n-p}$. From this, we can deduce that

$$\frac{\|y-\hat{y}_{H_0}\|^2}{\sigma^2} \sim \chi^2_{n-p+s}.$$

Meanwhile, from (b), we have $X(X^TX)^{-1}X^TX_{S^c} = X_{S^c}$, which implies

$$B^2 = (X(X^TX)^{-1}X^T - X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T)^2 = X(X^TX)^{-1}X^T - X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T = B.$$

Hence, under H_0, we have

$$\hat{y}-\hat{y}_{H_0} = By \sim N(0, \sigma^2 B^2) = N(0, \sigma^2 B).$$

As B is symmetric and idempotent,

$$\mathrm{rank}(B) = \mathrm{Tr}(B) = \mathrm{Tr}(X(X^TX)^{-1}X^T) - \mathrm{Tr}(X_{S^c}(X_{S^c}^TX_{S^c})^{-1}X_{S^c}^T) = p-(p-s) = s,$$

which implies, under H_0, that

$$\frac{\|\hat{y}-\hat{y}_{H_0}\|^2}{\sigma^2} = \frac{\|By\|^2}{\sigma^2} \sim \chi^2_{\mathrm{rank}(B)} = \chi^2_s.$$

4-(e). From the above calculations, under H_0, we have

$$\frac{\|\hat{y} - \hat{y}_{H_0}\|^2/(\sigma^2 s)}{\|y - \hat{y}\|^2/(\sigma^2(n-p))} \sim F_{s,n-p}.$$

Hence, we reject the null H_0 if

$$\frac{\|\hat{y} - \hat{y}_{H_0}\|^2/s}{\|y - \hat{y}\|^2/(n-p)} > F_{s,n-p,1-\alpha}.$$

4-(f). If $S = \{1, \ldots, p-1\}$ and thus $S^c = \{0\}$, we have $\hat{y}_{H_0} = \bar{y}\mathbb{1}_n$. Then, (b) becomes

$$\|y - \bar{y}\mathbb{1}_n\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - \bar{y}\mathbb{1}_n\|^2.$$

Also, (c) follows as $y - \hat{y}$ and $\hat{y} - \bar{y}\mathbb{1}_n$ are independent, and (d) becomes

$$\frac{\|y - \bar{y}\mathbb{1}_n\|^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{and} \quad \frac{\|\hat{y} - \bar{y}\mathbb{1}_n\|^2}{\sigma^2} \sim \chi^2_{p-1}.$$

Lastly, (e) becomes the F-test based on

$$\frac{\|\hat{y} - \bar{y}\mathbb{1}_n\|^2/(p-1)}{\|y - \hat{y}\|^2/(n-p)} \sim F_{p-1,n-p}.$$