Final 2019

Problem 1

- Left: the spread of residual increases as fitted value increases. Violation of constant variance assumption. WLS
- Middle: quadratic pattern. Add a quadratic term.
- Right: skewed and an outlier. Remove the outlier and may consider different distribution of ε.

Problem 2

(a) $R^2 = 1 - \frac{RSS}{TSS}$, R-Squared is the proportion of the variation explained by the model.

(b) Since

$$F = \frac{ESS/p}{RSS/(n-p-1)} = n - 2\left(\frac{1}{\frac{1}{R^2} - 1}\right) = (n-2)\frac{R^2}{1-R^2} \sim F_{1,n-2}$$

A 0.95 CI for F would be

$$0 \leq F \leq F_{0.05,1,n-2}$$

Since,

$$F = (n-2)\frac{R^2}{1-R^2} \leq F_{0.05,1,n-2}$$

Solve this, we get a 0.95 CI for R^2 is

$$0 \leq R^2 \leq \frac{F_{0.05,1,n-2}}{n-2+F_{0.05,1,n-2}}$$

Another way to approach this problem is to use bootstrap:

(1) Sample $(y_1^*, x_1^*), \ldots, (y_n^*, x_n^*)$ with replacement from $(y_1, x_1), \ldots, (y_n, x_n)$.

(2) Fit the OLS use the bootstrap sample and use $\hat{\beta}_i^*$ and $(y_1^*, x_1^*), \ldots, (y_n^*, x_n^*)$ to compute $(\hat{R}^2)^*$.

(3) repeat 1-2 B times to get bootstrap distribution of $(\hat{R}^2)^*$.

The bootstrap 0.95 CI is:

$$((\hat{R}^2)^*_{[0.025B]}, (\hat{R}^2)^*_{[0.975B]})$$

(c) pls see practice midterm.

Problem 3

(a) Let $D_i,\ i = 1, 2$ to be

$$D_i = \begin{cases} 1 & \text{if level } i \\ 0 & \text{if not level } i \end{cases}$$

The mean function for model 1 is

$$\mathrm{E}[Y|X] = \beta_0 + \beta_{1,1} D_1 + \beta_{1,2} D_2 + \beta_3 X_1 + \beta_4 X_2 = \begin{cases} \beta_0 + \beta_3 X_1 + \beta_4 X_2 & \text{base level} \\ \beta_0 + \beta_{1,1} + \beta_3 X_1 + \beta_4 X_2 & \text{if level 1} \\ \beta_0 + \beta_{1,2} + \beta_3 X_1 + \beta_4 X_2 & \text{if level 2} \end{cases}$$

The mean function for model 2 is

$$\begin{aligned} \mathrm{E}[Y|X] &= \beta_0 + \beta_{1,1} D_1 + \beta_{1,2} D_2 + \beta_2 X_1 + \beta_3 X_2 + \beta_{4,1} D_1 X_1 + \beta_{4,2} D_2 X_1 + \beta_5 X_1 X_2 \\ &= \begin{cases} \beta_0 + \beta_2 X_1 + \beta_3 X_2 + \beta_5 X_1 X_2 & \text{base level} \\ \beta_0 + \beta_{1,1} + \beta_2 X_1 + \beta_3 X_2 + \beta_{4,1} X_1 + \beta_5 X_1 X_2 & \text{if level 1} \\ \beta_0 + \beta_{1,2} + \beta_2 X_1 + \beta_3 X_2 + \beta_{4,2} X_1 + \beta_5 X_1 X_2 & \text{if level 2} \end{cases} \end{aligned}$$

(b)

$$\beta_0 : \text{the intercept for the base level } i$$

$$\beta_{1,i}, i = 1, 2 : \text{the intercept difference between the base level and level } i$$

$$\beta_i, i = 3, 4 : \text{increase 1 unit then increase by } \beta_i$$

(c)

$$AIC = 100\log(\exp(1.1)) + 2 \times 5 = 120$$
$$AIC = 100\log(\exp(1)) + 2 \times 8 = 116$$
$$BIC = 100\log(\exp(1.1)) + \log(100) \times 5 = 133$$
$$BIC = 100\log(\exp(1)) + \log(100) \times 8 = 136.8$$

Based on AIC, Model 2 is better. Based on BIC, Model 1 is better, BIC tend to select smaller model.

(d) $df_{H_0} = 100-(1+2+1+1) = 95$, $df_{H_A} = 100-(1+2+1+1+2\times1+1\times1) = 92$, so $F_{3,92}$.

Problem 4

(a) Consider

$$\min_\beta \sum_i w_i(Y_i - \beta X_i)^2$$

$$\begin{aligned}\frac{\partial}{\partial\beta}\sum_i w_i(Y_i - \beta X_i)^2 &= \frac{\partial}{\partial\beta}(Y - X\beta)^T W(Y - X\beta)\\ &= -2X^TWY + 2X^TWX\beta\\ &= 0\\ &\Rightarrow \hat\beta = (X^TWX)^{-1}X^TWY\end{aligned}$$

(b) $Y_i \sim N(X_i\beta, \sigma^2/w_i)$.

$$P(Y_1,\ldots,Y_n) = \prod_{i=1}^n \frac{w_i^{1/2}}{(2\pi)^{1/2}\sigma}\exp\{-\frac{(y_i - X_i\beta)^2}{2\sigma^2/w_i}\} = \frac{\prod_{i=1}^n w_i^{1/2}}{(2\pi)^{n/2}\sigma^n}\exp\{-\frac{\sum_{i=1}^n w_i(y_i - X_i\beta)^2}{2\sigma^2}\}$$

(c) MLE of β is WLS estimate.

$$\frac{\partial}{\partial\sigma^2} - \frac{n}{2}\log((2\pi\sigma^2)) - \frac{\sum_{i=1}^n w_i(y_i - X_i\beta)^2}{2\sigma^2} = \frac{\sum_{i=1}^n w_i(y_i - X_i\beta)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0$$
$$\Rightarrow \hat\sigma^2 = \frac{\sum_{i=1}^n w_i(y_i - X_i\beta)^2}{n}$$

(d) Ridge:

$$\min_{\beta} \sum_i w_i (Y_i - \beta X_i)^2 + \lambda \beta^T \beta$$

$$\begin{aligned}
\frac{\partial}{\partial \beta} \sum_i w_i (Y_i - \beta X_i)^2 + \lambda \beta^T \beta &= \frac{\partial}{\partial \beta} (Y - X\beta)^T W (Y - X\beta) + 2\lambda\beta \\
&= -2X^T W Y + 2(X^T W X + \lambda I)\beta \\
&= 0 \\
&\Rightarrow \hat{\beta} = (X^T W X + \lambda I)^{-1} X^T W Y
\end{aligned}$$