Problem 2

October 2020

1 (a)

Answers: The equality of error variance means that as the number of predictors increases, the scatter of the residuals doesn't change.

2 (b)

Answers: We are going to find a regression line such that the line minimize $e_1^2 + e_2^2 + e_3^2 + ... + e_n^2 = \sum_{i=1}^{N} e_i^2$ where e_i's are the residuals.

3 (c)

Answers: Conceptually, $\hat{Y}$ is a linear function of X, when we talk about a relationship between residuals and $\hat{Y}$, we are actually talking about the relationship between residuals and X in different scale and position. Mathematically, suppose $r = b_1\hat{Y} + b_0$ and $\hat{Y} = a_1X + a_0$, then $r = b_1a_1X + b_1a_0 + b_0$.

4 (d)

Answers: Sum of square of total $= \sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \sum_{i=1}^{N}(Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2$
$= \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{N} 2(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$
$= SSE + SSR + \sum_{i=1}^{N} 2e_i(\hat{Y}_i - \bar{Y})$
Since $\sum_{i=1}^{N} e_i = 0$, therefore Sum of square of total = SSE + SSR

5 (e)

Answers: SSE $= \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{N}(Y_i - b_1 - b_0X_i)^2$. We want to find b_1 and b_0 such that SSE is minimized. Let's take the derivative with respect to b_1 and b_0 for SSE.
Then we get $\frac{\partial SSE}{\partial b_1} = \sum_{i=1}^{N} -2(Y_i - b_1 - b_0X_i)$ and $\frac{\partial SSE}{\partial b_0} = \sum_{i=1}^{N} -2(Y_i - b_1 -$

$b_0X_i)X_i$.
In order to minimize SSE, we set the first derivatives of SSE to 0. $\sum_{i=1}^{N} -2(Y_i - b_1 - b_0X_i) = 0$ and $\sum_{i=1}^{N} -2(Y_i - b_1 - b_0X_i)X_i = 0$.
For b_1 we got $\sum_{i=1}^{N} b_1 = \sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} b_0X_i$
$Nb_1 = N\bar{Y} - N\bar{X}b_0$
$b_1 = \bar{Y} - \bar{X}b_0$ then we plug this b_1 into $\frac{\partial SSE}{\partial b_0}$, we got $\sum_{i=1}^{N} -2(Y_i - \bar{Y} + \bar{X}b_0 - b_0X_i)X_i = 0$
$\sum_{i=1}^{N}(Y_i - \bar{Y} + \bar{X}b_0 - b_0X_i) = 0$
$b_0 \sum_{i=1}^{N}(\bar{X} - X_i) = \sum_{i=1}^{N}(\bar{Y} - Y_i)$
$b_0 = \frac{\sum_{i=1}^{N}(\bar{Y} - Y_i)}{\sum_{i=1}^{N}(\bar{X} - X_i)}$
$= \frac{\frac{\sum_{i=1}^{N}(X_i - \bar{X}) * (Y_i - \bar{Y})}{N-1}}{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}}$
$= \frac{S_{XY}}{S_X^2}$