Homework 1 Derivations

Data Analytics in Business

Toyaj Singh

1 Question 1

1.1 Choose the correct statement regarding the sum of residuals calculated using Ordinary Least Squares (OLS)

Let us consider a general regression model given by Eq 1:

$$\hat{y}_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{1}$$

Note $\hat{y}$ is representative of the models estimate for any instance of response variable. This should be distinguished from the actual values y.

Now recall the objective of Ordinary Least Squares (OLS) is given in its name: to minimize to sum of squares of the residuals between the values from our model and the actual. The sum of squared errors are given in Eq 2 and in the expanded form in Eq 3.

$$SSE = \sum_{i=1}^{n}(e_i^2) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

$$= \sum_{i=1}^{n}(y_i - \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)^2 \tag{3}$$

Note here we have simply mate the substitution for $\hat{y}$.

Now to minimize we will, take the first derivative and set it to 0, as per standard procedure for minimization. For refresher, please refer to any elementary calculus notes.

Naturally we must take the derivative with respect to each one of the βs, however to explain this question, consider only the derivative with respect to β_0. To take the derivative easily, we can use the chain rule with the following substitution:

$$u = y_i - \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{4}$$

The derivative then simplifies to:

$$\frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^{n} 2(y_i - \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)(-1) \tag{5}$$

Recall the definition of the residual e_i:

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{i=1}^{n} e_i \tag{6}$$

Now as per minimization constraint, this derivative is equal to 0:

$$-2 \sum_{i=1}^{n} e_i = 0 \tag{7}$$

$$\Rightarrow \sum_{i=1}^{n} e_i = 0 \tag{8}$$

Hence the sum of residuals must be equal to 0.