JAN FEB MAR APR MAY JUN
JUL AUG SEP OCT NOV DEC

Stat 408 MIDTERM

Rachel Gordon

1a. X_i is the i^{th} observation of the predictor or independent variable. It is known and comes from the dataset. observation X is a random variable

Y_i is the i^{th} observation of the response variable or the dependent variable. Y_i in this model is unknown because it is what's being predicted. Y is a random variable.

β_0 is the intercept and β_1 is the coefficient of X_i. They are the model parameters and they are unknown and fixed while $\hat{\beta}_0$ and $\hat{\beta}_1$ are random variables.

ε is the error and it is unknown and a random variable

1b.
- model assumption, follows linearity structure, $E(y) = X\beta$
- error assumption, normality $\varepsilon \sim N(0, \sigma^2 I)$
- no unusual observations or outliers
- constant variance

2a. Based on the model summary output, none of the predictors appear to be significant at the 5% level as the p-values are greater than 0.05. However, based on the F-test, the p-value is 0.01902, suggesting that these 4 predictors collectively have a relationship to the response. However, these 2 conclusions conflict with one another thus other interactions or transformations of these predictors should be explored

2b. This code conducts an F-test, resulting in a p-value of 0.468. Based on this result, the smaller model without both R Str and L Str and instead just the sum of those two would be a better choice because the p-value is less than a 5% significance level.

$$H_0: \beta_{RStr} = \beta_{LStr}$$

$$H_a: \beta_{RStr} \neq \beta_{LStr}$$

Based on this we fail to reject the null hypothesis, concluding that there is insufficient evidence that LStr and RStr do not have the same effect on distance

2c. $df = n - p - 1$

$8 = n - 4 - 1$

$8 = n - 5$

$n = 13$

No, we cannot compare two models with two different response variables because they are measuring two completely different things.

3a. The love coefficient shows that the happiness score is expected to increase by approximately 1.919 for every one unit increase in love. Therefore, a person with deep belonging and caring (3) is expected to have a happiness score that is 3.838 greater than someone who is lonely (love = 1)

3b. The clove variable changes so that a love value less than 3 (1 or 2) is coded as a 0 while a love value of 3 is coded as a 1. Therefore, someone who ~~is~~ has deep belonging is expected to have a happiness score about 2.296 greater than someone who doesn't. This slightly changes the interpretation because it removes the distinction between lonely and ~~securing~~ secure relationships and simply refers to it all as deep belonging or not.

4. 1 - outlier and influential point because it is far from the data and the overall fit line the graph would make

2 - outlier but not influential because it is far from the data but not the overall fit of the graph

3 - neither because it is not far from the data points or the overall fit

5. $RSS(\beta_0, \beta_1) = \sum_{i=1}^{n} e_i^2 = e_1^2 + e_2^2 + \ldots + e_n^2 = (y_1 - \beta_0 - \beta_1 x_1)^2 + \ldots + (y_n - \beta_0 - \beta_1 x_1)^2$

$= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$ in $Y_i = \beta_0 + \varepsilon_i$, $\beta_1 = 0$

$= \sum_{i=1}^{n} (y_i - \beta_0)^2$

$$\frac{\partial}{\partial \beta_0}\left(\sum_{i=1}^{n} (y_i - \beta_0)^2\right) = 0$$

$$-\cancel{2} \sum_{i=1}^{n} (y_i - \beta_0) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \beta_0 = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \beta_0$$

$$\beta_0 = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$$

$RSS(\beta) = \sum_{i=1}^{n} e_i^2 = e^T e$ $\qquad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 - \beta_0 \\ \vdots \\ y_n - \beta_0 \end{pmatrix} = y - X\beta$

$$e^T e = \begin{pmatrix} y_1 - \beta_0 & y_2 - \beta_0 & \cdots & y_n - \beta_0 \end{pmatrix} \begin{pmatrix} y_1 - \beta_0 \\ y_2 - \beta_0 \\ \vdots \\ y_n - \beta_0 \end{pmatrix} = (y_1 - \beta_0)^2 + (y_2 - \beta_0)^2 + \ldots + (y_n - \beta_0)^2$$

$$= \sum_{i=1}^{n} (y_i - \beta_0)^2$$

$$\frac{\partial e^T e}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\left(\sum_{i=1}^{n} (y_i - \beta_0)^2\right) = 0$$

$$-2 \sum_{i=1}^{n} (y_i - \beta_0) = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \beta_0$$

$$\beta_0 = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$$