STT 4660
Homework #2

Timothy Stubblefield

September 14, 2020

1.21

a) From the data, we can calculate $\bar{x}$ with the formula,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

1) Thus, $\bar{x} = 1$
2) Similarly, we can calculate $\bar{y}$ with the formula,

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

3) Hence, $\bar{y} = 14.2$
4) Now, utilize the regression formula

$$\hat{y} = b_0 + b_1 x$$

where, $b_1 = \frac{S_{xx}}{S_{xy}}$ and $b_0 = \hat{y} - b_1\bar{x}$

5) So, $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$
6) Evaluating (5), we get,

$$S_{xx} = 10$$

7) Now, $S_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$
8) Evaluating, (7), we get,

$$S_{yy} = 177.6$$

9) So, $S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$
10) Evaluating (9), we get,

$$S_{xy} = 182 - 142$$

$$S_{xy} = 40$$

11) From our information gathered in 5-10 we can apply (4) yielding,

$$b_1 = \frac{S_{xy}}{S_{xx}}$$

$$b_1 = \frac{40}{10}$$

$$b_1 = 4$$

12) Similarly, we can solve for b_0 like so,

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_0 = 14.2 - 4(1)$$

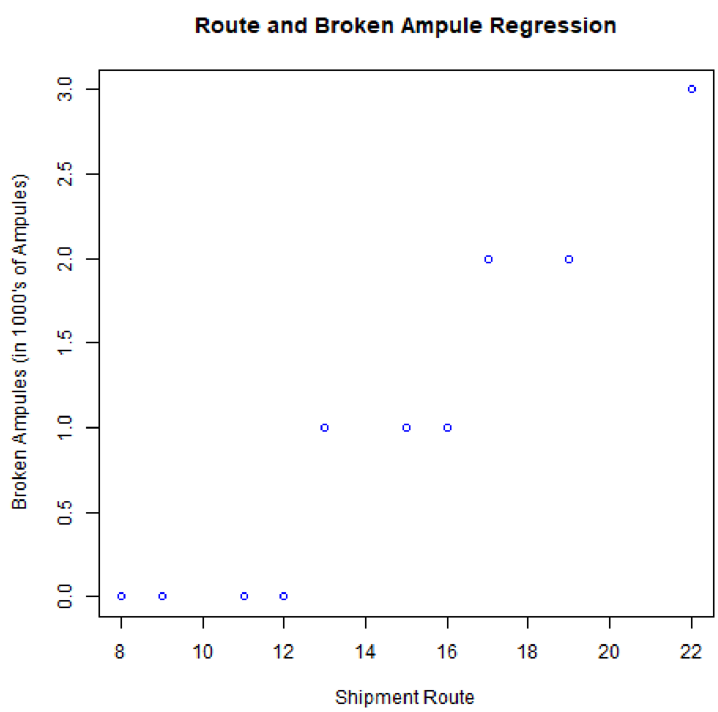
$$b_0 = 14.2 - 4$$

$$b_0 = 10.2$$

13) Thus, the linear regression function is,

$$\hat{y} = 10.2 + 4x$$

14) Below is the scatterplot of the points,



b) Now, from the previous part, the regression function is,

$$\hat{y} = 10.2 + 4x$$

1) We must find evaluate the function at $X = 1$.
2) So, we have the following,

$$\hat{y_1} = 10.2 + 4(1)$$

$$\hat{y_1} = 10.2 + 4$$

$$\hat{y_1} = 14.2$$

c) Now use the same formula for two transfers.

1) Now, we want to evaluate the function at $X = 2$.
2) Thus, we have the following,

$$\hat{y_2} = 10.2 + 4(2)$$

$$\hat{y_2} = 10.2 + 8$$

$$\hat{y_2} = 18.2$$

d) Now, I will show that the regression line goes through $(\bar{x}, \bar{y})$

1) Substitute in $\bar{x}$ and $\bar{y}$ into the regression function.
2) Thus, by inserting $\bar{x} = 1$ and $\bar{y} = 14.2$, we have

$$14.2 = 10.2 + 4(1)$$

$$14.2 = 10.2 + 4$$

$$14.2 = 14.2$$

■

1.25

a) I want to find the residual for the first case.

1) The formula for the residual is,

$$e_i = y_i - \hat{y}_i$$

2) Thus, for the first case, we have,

$$e_1 = y_1 - \hat{y}_1$$

3) Substitute in values for y_1 and $\hat{y_1}$, yielding,

$$e_1 = 16 - 14.2$$

$$e_1 = 1.8$$

b) Now, lets find $\sum e_i^2$ and MSE

1) After calculating each e_i, I determined that,

$$\sum e_i^2 = SSE = 17.6$$

2) This can also be calculated another way, as shown,

$$SSE = S_{yy} - b_1 S_{xy}$$

$$SSE = 177.6 - 4(40)$$

$$SSE = 177.6 - 160$$

$$SSE = 17.6$$

3) And, we know $MSE = S^2$

4) Now, we calculate the MSE to estimate the σ^2

5) And, we can use the following formula,

$$S^2 = MSE = \frac{SSE}{n-2}$$

6) Just evaluate the expression with the calculated SSE from (1) and n, yielding

$$S^2 = MSE = \frac{17.6}{10-2}$$

$$S^2 = MSE = \frac{17.6}{8}$$

$$S^2 = MSE = 2.2$$

1.33

We want to derive the Least Squares Estimator of β_0 for the regression model,

$$Y_i = \beta_0 + \epsilon_i$$

1) We know the formula for SSE of Least Squares is,

$$SSE = \sum (y_i - \beta_0 - \beta_1 x)^2$$

2) Since $\beta_1 = 0$, the formula becomes,

$$Q(\beta_0) = SSE = \sum (y_i - \beta_0)^2$$

3) To minimize SSE, take the derivative and set to 0, yielding

$$\frac{dQ}{d\beta_0} = \frac{d}{d\beta_0} \sum (y_i - \beta_0)^2 = 0$$

$$\frac{dQ}{d\beta_0} = 2 \sum (y_i - \beta_0)(-1) = 0$$

$$\frac{dQ}{d\beta_0} = -2 \sum (y_i - \beta_0) = 0$$

$$\frac{dQ}{d\beta_0} = \sum (y_i - \beta_0) = 0$$

4) Thus, the Normal Equation for β_0 estimator is,

$$\sum (y_i - b_0) = 0$$

5) Now, rearrange the equation from (4) to solve for b_0.

$$\sum y_i - nb_0 = 0$$

$$nb_0 = \sum y_i$$

$$b_0 = \frac{\sum y_i}{n}$$

6) Hence, the Least Squares Estimator for β_0 is,

$$b_0 = \frac{\sum y_i}{n}$$

■

1.34

We want to show that the Least Squares Estimator, b_0, of β_0, is unbiased.

1. Recall from Problem 1.33 that,

$$b_0 = \frac{\sum y_i}{n}$$

2. By applying definition of $\bar{y}$ we can see,

$$b_0 = \bar{y}$$

3. Now, evaluate the expected value of b_0, yielding

$$E(b_0) = E(\bar{y})$$

$$E(b_0) = E(\frac{1}{n}\sum y_i)$$

4. Remember,

$$E(\sum y_i) = \sum \beta_0 + \beta_1 x_i$$

5. Thus, applying (4) to (3), we have,

$$E(b_0) = \frac{1}{n}\sum(\beta_0 + \beta_1 x_i)$$

6. But, remember $\beta_1 = 0$, so

$$E(b_0) = \frac{1}{n}\sum \beta_0$$

$$E(b_0) = \frac{1}{n} n\beta_0$$

$$E(b_0) = \beta_0$$

7. Thus, b_0 is unbiased.

■