Homework 1

Aiden Kenny
STAT GR5205: Linear Regression Models
Columbia Univeristy
September 21, 2020

Question 1

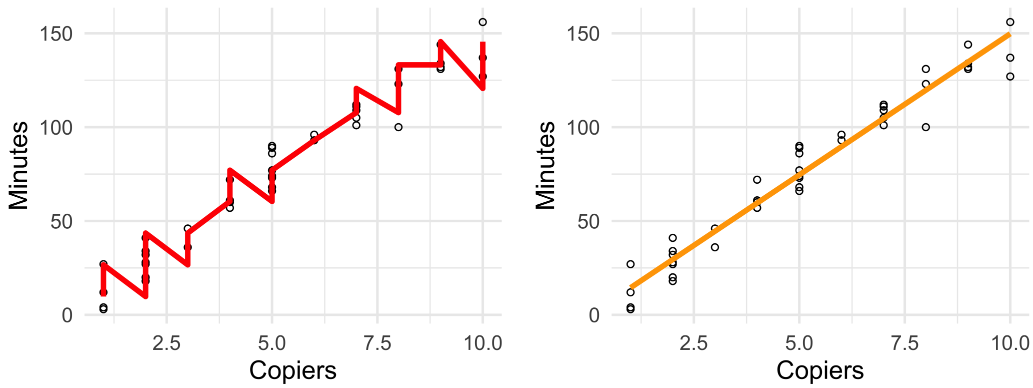

Figure 1: Left: applying a LOWESS smoother to a scatterplot of the data. Right: plotting the linear regression model to the scatterplot.

(a) Not sure what she wants here, ask more about LOWESS smoothers.

(b) Using `R`, our estimated coefficients are given by $\hat{\beta}_0 = -0.5801567$ and $\hat{\beta}_1 = 15.0352480$, and so our estimated linear regression function is given by

$$\hat{Y} = -0.5801567 + 15.0352480 \cdot X.$$

The estimated linear regression model has been overlayed on a scatterplot of the data in the right plot in Figure 1, and the estimated function seems to fit the data well. The general trend, where an increase in number of copiers results in an increased number of minutes on call, is captured by the model.

(c) $\hat{\beta}_1$ can be interpreted as follows. If the number of copiers serviced during a call increased by one, the total number of minutes of the call is expected to *increase* by 15.0352480 minutes.

(d) $\hat{\beta}_0$ can be interpreted as follows. If there are zero copiers serviced during a call, then we can expect the call to last for, on average, -0.5801567 minutes. This does *not* provide any useful or relevant information; a call cannot ever have negative time, and a customer would never call if they did not have any copiers to service (where $X = 0$).

(e)

(f)

(g) Using `R`, we can see that the residuals sum to 0; this is easy to do since the residuals are are included in the fitted model. We can think of Q as a *function* of β_0 and β_1, and we want to find the values of β_0 and β_1 that minimize Q. The observed residuals e_i, when plugged into Q, give the smallest value of Q that can possibly be obtained. Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ be the random vector containing the n residuals, and let $\mathbf{e} = (e_1, \ldots, e_n)^T$ be the n realized residuals from $\hat{\beta}_0$ and $\hat{\beta}_1$. Using this notation, we have $Q = \|\boldsymbol{\varepsilon}\|_2^2$, and

$$\|\mathbf{e}\|_2^2 = \min \|\boldsymbol{\varepsilon}\|_2^2 = \min Q.$$

Question 2 fafad

Question 3 The coefficients for the estimated linear regression function $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X$ are given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$