Stat 331 Applied Linear Models – Assignment 1

You need to use the cover sheet provided in Learn. Due on Oct 1st (Wednesday) 12pm to the drop boxes located across the hall from MC 4065/4066.

1. Under the simple linear regression model, let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ denote the fitted value, and $r_i = y_i - \hat{y}_i$ denote the corresponding residual. We want to show that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n r_i^2}{n-2}$$

is an unbiased estimator of σ^2. Prove the following

   (a) Show that $cov(y_i, \hat{y}_i) = \sigma^2[\frac{1}{n} + \frac{(x_i-\bar{x})^2}{S_{xx}}]$ where $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$.

   (b) Show that $E(r_i) = 0$ and $V(r_i) = \sigma^2[1 - \frac{1}{n} - \frac{(x_i-\bar{x})^2}{S_{xx}}]$.

   (c) Show that $E(\sum_{i=1}^n r_i^2) = (n-2)\sigma^2$.

2. In some cases, it may be more realistic to consider a simple linear regression model through the origin:

$$y_i = \beta x_i + \epsilon_i, i = 1, 2, \ldots, n$$

where $\epsilon_i \sim N(0, \sigma^2)$.

   (a) Derive $\hat{\beta}$, the least squares estimator of β.

   (b) Show that $\hat{\beta}$ is an unbiased estimator of β and find an expression for $V(\hat{\beta})$.

   (c) Let $r_i = y_i - \hat{y}_i$ denote the residual. (i) Show that $\sum_{i=1}^n r_i x_i = 0$; (ii) Is it true that $\sum_{i=1}^n r_i = 0$? why?

3. Use the “Copier" dataset (available on Learn). The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, $\mathbf{x}$ is the number of copiers serviced and $\mathbf{y}$ is the total number of minutes spent by the service person.

   (a) Plot the data. Does a simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, 45$$

   seem appropriate? Explain.

   (b) What is the fitted regression line?

   (c) What is the estimated change in the mean service time when the number of copiers serviced increases by one?

   (d) Construct a 95% confidence interval for the slope and interpret.

   (e) Construct a t-test to determine whether or not there is a linear association between $\mathbf{x}$ and $\mathbf{y}$ here. Use 0.05 as the significance level. State the hypotheses, test statistic, p-value and conclusion.

   (f) Are your results in part (d) and part (e) consistent? Explain.

   (g) For a new call with 5 copiers serviced, what is the predicted total number of minutes spent by the service person? Construct a 95% prediction interval.