notes w3

Sam Reeves

A Modern Approach to Regression Ch 2.2

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

$$E(\hat{\beta}_1 | X) = \beta_1$$

$$\text{Var}(\hat{\beta}_1 | X) = \frac{\sigma^2}{SXX}$$

sooooooo. . . .

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{SXX}} N(0, 1)$$

If sigma is not known, as is usually the case, we can replace sigma with S:

$$T = \frac{\hat{\beta}_1 - \beta_1}{S / \sqrt{SXX}} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$$

For hypothesis testing, $H_0 : \beta_1 = \beta_1^0$

The test statistic is:

$$T = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \text{with } t_{n-2}$$

Confidence Intervales for the Population Regression Line

Here we are looking for a population regression line at a specific x value deonted x^*:

$$E(Y | X = x^*) = \beta_0 + \beta_1 x^*$$

an appropriate estimator:

$$\hat{y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

under our basic assumptions:

$$E(\hat{y}^*) = E(\hat{y} | X = x^*) = \beta_0 + \beta_1 x^*$$

The Z score (sigma is known):

$$Z = \frac{\hat{y}^* - (\beta_0 + \beta_1 x^*)}{\sigma\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{SXX}}} N(0,1)$$

The T score (sigma not known):

$$T = \frac{\hat{y}^* - (\beta_0 + \beta_1 x^*)}{S\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{SXX}}} t_{n-2}$$

The confidence interval is then:

$$\hat{\beta_0} + \hat{\beta_1} x^* \pm t(\alpha/2, n-2) S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{SXX}}$$

Prediction Intervals

The prediction interval for a given vlaue x^* is:

$$\hat{\beta_0} + \hat{\beta_1} x^* \pm t(\alpha/2, n-2) S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{SXX}}$$

Analysis of Variance

To test if there is a linear association between Y and X we have to test $H_0 : \beta_1 = 0$ against $H_A : \beta_1 \neq 0$.

We can use:

$$T = \frac{\hat{\beta_1} - 0}{\text{se}(\hat{\beta_1})} t_{n-2}$$

2.3

The manager of the purchasing department of a large company would like to develop a regression model to predict the average amount of time it takes to process a given number of invoices. Over a 30-day period, data are collected on the number of invoices processed and the total time taken (in hours). The data are available on the book web site in the file invoices.txt. The following model was fit to the data: Y = b 0 + b 1 x + e where Y is the processing time and x is the number of invoices. A plot of the data and the fitted model can be found in Figure 2.7 . Utilizing the output from the fit of this model provided below, complete the following tasks.

(a) Find a 95% confidence interval for the start-up time, i.e., β_0 .

(b) Suppose that a best practice benchmark for the average processing time for an additional invoice is 0.01 hours (or 0.6 minutes). Test the null hypothesis $H_0 : \beta_1 1 = 0.01$ against a two-sided alternative. Interpret your result.

(c) Find a point estimate and a 95% prediction interval for the time taken to process 130 invoices.
