Chapter 7

Linear regression

Exercise 1 The y variable is weight and the x variable is height. The OLS estimator of β_1 and β_0 are:

$$\hat{\beta}_1 = \frac{\text{COV}[x, y]}{\text{VAR}[x]} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Notice that:

$$\begin{aligned}
\bar{x} &= \frac{\sum_{i=1} n x_i}{n} = \frac{892}{5} = 178.4 \\
\text{VAR}[x] &= \frac{\sum_{i=1} n (x_i - \bar{x})^2}{n-1} = \frac{411.2}{4} = 102.8 \\
\bar{y} &= \frac{\sum_{i=1} n y_i}{n} = \frac{373}{5} = 74.6 \\
\text{COV}[x, y] &= \frac{\sum_{i=1} n (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{300.8}{4} = 75.2
\end{aligned}$$

Therefore:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\text{COV}[x, y]}{\text{VAR}[x]} = \frac{75.2}{102.8} = 0.732 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 74.6 - 0.732(178.4) = -55.99
\end{aligned}$$

Exercise 2 The OLS estimates of β_1 and β_0 are equal to:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{25.2}{44.5} = 0.566 \\
\hat{\beta}_0 &= 1980 - 0.566(1280) = 1255.52
\end{aligned}$$

Exercise 3 Recall that the coefficient of determination R^2 is equal to coefficient of correlation squared, i.e.

$$R^2 = (r_{xy})^2 = \left(\frac{\text{COV}[x, y]}{s_x s_y} \right)^2$$

The coefficient of correlation is equal to the covariance between x and y divided by the product of standard deviations, i.e.

$$r_{xy} = \frac{\text{COV}[x, y]}{s_x s_y}$$

Therefore, we have

$$r_{xy} = \frac{25.2}{\sqrt{44.5}\sqrt{29.7}} = 0.693$$

and

$$R^2 = 0.693^2 = 0.480$$

i.e. 48% of the variation in the data is explained by the explanatory variable.

Exercise 4

- The hypothesis tested is:

$$\text{H}_0 : \beta_1 = 0 \quad \text{versus} \quad \text{H}_0 : \beta_1 \neq 0$$

  Since $n = 30$, and the parameters to estimate are two, the relevant critical value at 5% significance level for this test is $\text{cv}_\alpha = t_{\frac{\alpha}{2},n-2} = t_{0.025,28} = 2.048$. The test statistic is

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\text{VAR}}[\beta_1]}} = \frac{0.2 - 0}{0.08} = 2.5$$

  We reject H_0 if $|T| > \text{cv}_\alpha$. Since $2.5 > 2.048$ we reject the null hypothesis at a 5% level of significance.

- The hypothesis tested is:

$$\text{H}_0 : \beta_1 = 0.3 \quad \text{versus} \quad \text{H}_0 : \beta_1 < 0.3$$

  Since $n = 30$, and the parameters to estimate are two, the relevant critical value at 5% significance level for this test is $\text{cv}_\alpha = t_{\alpha,n-2} = t_{0.05,28} = 1.701$. The test statistic is

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\text{VAR}}[\beta_1]}} = \frac{0.2 - 0.3}{0.08} = -1.25$$

  We reject H_0 if $T < -\text{cv}_\alpha$. Since $-1.25 < -1.701$ we fail to reject the null hypothesis at a 5% level of significance.