Estimation and Testing for Slope

In the model where

$$E[Y] = \beta_0 + \beta_1 x$$

it is of interest to test the hypothesis

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \end{aligned}$$

i.e. H_0 implies that there is no systematic contribution of x to the variation of y.

To test H_0 vs H_a we us the test statistic

$$t = \frac{\widehat{\beta}_1}{\text{e.s.e}(\widehat{\beta}_1)} = \frac{\widehat{\beta}_1}{s_{\widehat{\beta}_1}}$$

where e.s.e$(\widehat{\beta}_1)$ is the *Estimated Standard Error* of $\widehat{\beta}_1$, computed as

$$\text{e.s.e}(\widehat{\beta}_1) = \frac{s}{\sqrt{SS_{xx}}}$$

where s is the estimate of σ defined previously.

If H_0 is true, and $\beta_1 = 0$, then

$$t = \frac{\widehat{\beta}_1}{s/\sqrt{SS_{xx}}} \sim \text{Student}(n-2)$$

so we can carry out a significance test at level α in the usual way (use a p-value, or construct the rejection region).

Note: we might also consider a one-sided test, where $H_a : \beta_1 > 0$, say.

- If $H_a : \beta_1 \neq 0$, we use the *two-sided* rejection region, with critical values
$$C_R = \pm St_{\alpha/2}(n-2)$$
- If $H_a : \beta_1 > 0$, we use the *one-sided* rejection region, with critical value
$$C_R = +St_{\alpha}(n-2)$$
- If $H_a : \beta_1 < 0$, we use the *one-sided* rejection region, with critical value
$$C_R = -St_{\alpha}(n-2)$$

Note: To test

$$H_0 \;:\; \beta_1 = b$$
$$H_a \;:\; \beta_1 \neq b$$

for any b, the test statistic is

$$t = \frac{\widehat{\beta}_1 - b}{s/\sqrt{SS_{xx}}}$$

(for example, $b = 1$ may be of interest. If H_0 is true

$$t \sim \text{Student}(n-2)$$

Confidence Interval

A $100(1-\alpha)\%$ confidence interval for β_1 is

$$\widehat{\beta}_1 \pm St_{\alpha/2}(n-2) \times s_{\widehat{\beta}_1}$$

where

$$\begin{aligned} St_{\alpha/2}(n-2) &: \quad \alpha/2 \text{ prob. point of Student}(n-2) \text{ distn.} \\ s_{\widehat{\beta}_1} &: \quad \text{Estimated standard error of } \widehat{\beta}_1 \end{aligned}$$

Note: we could perform a similar analysis for β_0, but this is generally of less interest.

The only quantity that needs attention is the estimated standard error of $\widehat{\beta}_0$. It can be shown that

$$\text{e.s.e.}(\widehat{\beta}_0) = s_{\widehat{\beta}_0} = \sqrt{\frac{1}{n}\left(1 + \frac{n\bar{x}^2}{SS_{xx}}\right)}$$

2.1.5 The Coefficient of Correlation

To measure the *strength of association* between the two variables x and y we can use the

Pearson Product Moment Coefficient Of Correlation

or *correlation coefficient* which measures the strength of the **linear** relationship between x and y.

The coefficient, r, is defined by

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

where

$$SS_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2 \quad SS_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

Note: $-1 \leq r \leq 1$.

- If r is close to 1, there is a strong linear relationship between x and y where y **increases** with x.
- If r is close to -1, there is a strong linear relationship between x and y where y **decreases** with x.

Note: In the model

$$y = \beta_0 + \beta_1 x$$

$\beta_1 = 0 \implies r \approx 0$, so tests for $\beta_1 = 0$ can also be used to deduce a lack of correlation between the variables.

Notes

1. A strong linear relationship is not necessarily a **causal** relationship, that is, just because $r \approx 1$ does not mean that x **causes** changes in y (we may have a *spurious* correlation).

2. Just because $r \approx 0$ does not mean that that x and y are unrelated, merely that they are **uncorrelated**. That is, it is possible to construct examples where x and y have a strong functional relationship, but where $r = 0$.

Examples where $r \approx 0$.