Test for Regression Slope

This document provides a derivation of the distribution of the test statistic for the slope in a simple linear regression. The result can be used to compute the power of the test when conditions are specified.

The simple linear regression model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\ \epsilon_i &\sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

Let $\hat{\beta}_1$ be the ordinary least squares (OLS) estimate of β_1. This is a linear combination of the y_i and also has a Normal distribution,

$$\begin{aligned} \hat{\beta}_1 &\sim \mathrm{N}\left(\beta_1, \mathrm{Var}(\hat{\beta}_1)\right) \\ \mathrm{Var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Typically the error variance is estimated by

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \\ &= \frac{SSE}{n-2} \end{aligned}$$

Another useful result is

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$$

Combining the normality of $\hat{\beta}$ and the distribution for the quadratic form involving SSE, the statistic

$$\begin{aligned} T &= \frac{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\beta_1)}}}{\sqrt{\frac{SSE}{\sigma^2(n-2)}}} \\ &\sim t_{n-2} \end{aligned}$$

has a t distribution with $n-2$ degrees of freedom.

A common test for simple linear regression has the form

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

The test statistic T from above can be modified to obtain the typical test statistic for this hypothesis, denoted as T_0

$$\begin{aligned} T_0 &= \frac{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\beta_1)}} + \frac{\beta_1}{\sqrt{Var(\beta_1)}}}{\sqrt{\frac{SSE}{\sigma^2(n-2)}}} \\ &\sim t_{n-2}(\delta) \\ \delta &= \frac{\beta_1}{\sqrt{Var(\beta_1)}} \\ &= \beta_1 \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}} \end{aligned}$$

Note that this statistic adds an additional term in the numerator. The addition of this term gives T_0 a non-central t distribution with $n-2$ degrees of freedom and non-centrality parameter δ. **Thus, the usual test statistic for the regression slope follows a non-central t distribution when the true slope is not zero.**

Just to be sure that T_0 is the usual test statistic, the simplification can be completed.

$$\begin{aligned}
T_0 &= \frac{\frac{\hat{\beta}_1-\beta_1}{\sqrt{Var(\beta_1)}} + \frac{\beta_1}{\sqrt{Var(\beta_1)}}}{\sqrt{\frac{SSE}{\sigma^2(n-2)}}} \\
&= \frac{\frac{\hat{\beta}_1}{\sqrt{Var(\beta_1)}}}{\sqrt{\frac{SSE}{\sigma^2(n-2)}}} \\
&= \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}}
\end{aligned}$$

Inspection of the non-centrality parameter δ reveals that its magnitude depends on the true slope, the error variance, and the distribution of x values.

The power of the test is then the probability that the statistic T_0 will be more extreme than the appropriate quantiles of the central t distribution with $n-2$ degrees of freedom,

$$\text{Power} = 1 - \Pr[t_{\alpha/2;n-2} \leq T_0 \leq t_{1-\alpha/2;n-2}]$$

This computation will require computing probabilities for the non-central t distribution; this can be done with the `pt` function in R using the `ncp` option.