Lecture 8

Estimation of σ^2, t-statistic, and $(1-\alpha)\%$ confidence interval for β_1

Let $$\hat{\sigma}^2 = \frac{SSR}{N-2} = \frac{\sum_1^N \hat{u}_i^2}{N-2} = \frac{\sum_1^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{N-2}.$$

Then

<u>Property 8</u>: $E(\hat{\sigma}^2) = \sigma^2$ (unbiased)

<u>Property 9</u>: Under the assumed truth that

$$H_0: \beta_1 = \beta_1^0$$ ($\beta_1^0 = 0$ is a case often of interest.)

The statistic,

$$t = \frac{\hat{\beta}_1 - \beta_1^0}{\sqrt{\frac{\hat{\sigma}^2}{\sum_1^N (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_1 - \beta_1^0}{se(\hat{\beta}_1)}$$

<u>in repeated samples</u>, has a t-distribution with $v = N-2$ degrees of freedom.

Property 10: under the assumed truth that

$$H_0: \beta_0 = \beta_0^\circ$$

the statistic

$$t = \frac{\hat{\beta}_0 - \beta_0^\circ}{\sqrt{\dfrac{\sum_1^N x_i^2}{N\sum_1^N (x_i - \bar{x})^2} \cdot \hat{\sigma}^2}} = \frac{\hat{\beta}_0 - \beta_0^\circ}{se(\hat{\beta}_0)}$$

follows a t-distribution with $v = N-2$ degrees of freedom in repeated sampling.

See Section 4.2 in Wooldridge for a discussion of two-tailed and one-tailed tests of hypotheses concerning β_1 (one-sided alternative versus two-sided alternative). Also see section 4.2 for a discussion of how you calculate the probability value (p-value) of a t-statistic for a one-sided alternative and a two-side alternative (Figure 4.6).

$(1-\alpha)\%$ Confidence Interval for $\hat{\beta}_1$

Inversion of the t-statistic into a confidence interval.

$$Pr\left(-t_{v,\alpha/2} < \frac{\hat{\beta}_1 - \beta_1^\circ}{\sqrt{\frac{\hat{\sigma}^2}{\sum_1^N (X_i - \bar{X})^2}}} < t_{v,\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \quad Pr\left(-t_{v,\alpha/2} \cdot se(\hat{\beta}_1) < \hat{\beta}_1 - \beta_1^\circ < t_{v,\alpha/2} \cdot se(\hat{\beta}_1)\right) = 1-\alpha$$

where $se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_1^N (X_i - \bar{X})^2}}$

$$\Rightarrow \quad Pr\left(\hat{\beta}_1 - t_{v,\alpha/2} \cdot se(\hat{\beta}_1) < \beta_1^\circ < \hat{\beta}_1 + t_{v,\alpha/2} \cdot se(\hat{\beta}_1)\right) = 1-\alpha.$$

The $(1-\alpha)\%$ confidence interval

$$\left[\hat{\beta}_1 - t_{v,\alpha/2} \cdot se(\hat{\beta}_1),\ \hat{\beta}_1 + t_{v,\alpha/2} \cdot se(\hat{\beta}_1)\right]$$

is a random interval in the sense

that the endpoints of the confidence interval (and hence the confidence interval itself) are random in repeated samples. It is then proper to say that $(1-\alpha)\%$ of the above so-constructed confidence intervals will encompass the unknown population value β_1^0. It should not be said that there is a $(1-\alpha)\%$ chance that the given confidence interval includes the unknown population value β_1^0. For any given confidence interval, either β_1^0 is in it or it is not in the confidence interval.