Lecture Summary: Jan. 23, 2023

• p-value (continued)

For example, consider $H_0 : \beta_1 = 3$ vs $\mathrm{H}_1 : \beta_1 > 3$. The value of the t statistic is

$$t = \frac{\hat{\beta}_1 - \beta_{10}}{\text{s.e.}(\hat{\beta}_1)} = \frac{3.57 - 3}{0.35} = 1.63.$$

Under H_0, the probability that $t \geq 1.63$ is 0.058. Therefore, the p-value is 0.058. If the alternative is changed to $H_1 : \beta_1 \neq 3$, the corresponding p-value is $2 * 0.058 = 0.117$ (verify!).

• An alternative way of making a decision: Reject H_0 if the p-value is $\leq \alpha$.

• Power of the test

It is the probability of rejecting H_0 when the alternative holds at a given value.

<u>Power calculation</u>:

Using Table B.5. Example: If $\beta_{10} = 0, \beta_1 = 1$, $\text{s.d.}(\hat{\beta}_1) = 0.5$, we have $\delta = 1/0.5 = 2.0$. Let $\alpha = 0.05$. From Table B.5 we find the power is about 0.48.

<u>Derivation (not required for exams)</u>: TA will cover in discussion sessions.

• Confidence interval for β_1:

Assuming normality, a $100(1-\alpha)\%$ confidence interval (c.i.) for β_1 is

$$\hat{\beta}_1 \pm t_{n-2}\left(1 - \frac{\alpha}{2}\right) \text{s.e.}(\hat{\beta}_1),$$

where

$$\text{s.e.}(\hat{\beta}_1) = \sqrt{\frac{\text{MSE}}{\sum_i (x_i - \bar{x})^2}}.$$

Some concepts regarding the c.i.: Consider, for example, a 95% c.i.

1. For a numerical c.i. computed from one set of data, such as the Toluca company data, which results a 95% c.i. for β_1, (2.85, 4.29), it is wrong to

say the interval has a 95% chance of covering β_1 [because the interval (2.85, 4.29) either covers β_1, or it does not cover β_1].

2. The 95% confidence level is in the sense that the interval

$$\left[\hat{\beta}_1 - t_{n-2}\left(1-\frac{\alpha}{2}\right)\text{s.e.}(\hat{\beta}_1),\quad \hat{\beta}_1 + t_{n-2}\left(1-\frac{\alpha}{2}\right)\text{s.e.}(\hat{\beta}_1)\right],$$

with the lower and upper ends considered as random variables depending on the data collection, has 95% chance of covering β_1.

3. A alternative way of understanding: Suppose there are 100 people, each collecting a different set of data and hence producing a different 95% c.i. for β_1. Out of the 100 c.i.s, approximately 95 will cover β_1; however, we do not know which ones cover, and which ones don't.

• Confidence interval for β_0:

A $100(1-\alpha)\%$ c.v. for β_0 is

$$\hat{\beta}_0 \pm t_{n-2}\left(1-\frac{\alpha}{2}\right)\text{s.e.}(\hat{\beta}_0),$$

where

$$\text{s.e.}(\hat{\beta}_0) = \sqrt{\text{MSE}\left\{\frac{1}{n} + \frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}\right\}}.$$