Econ 103: Introduction to Econometrics
Week 4: Confidence Interval and Hypothesis Testing

Francisco Díaz-Valdés*

Fall 2025

1 What is the distribution of our estimates?

Recall that our model is

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

where we assume that all the assumptions hold, including $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.[1] Then,

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2 \sum_i x_i}{N \sum_i (x_i - \overline{x}_i)^2}\right)$$

$$\hat{\beta}_2 \sim \mathcal{N}\left(\beta_2, \frac{\sigma^2}{\sum_i (x_i - \overline{x}_i)^2}\right)$$

The line-by-line derivation can be found in your textbook.

We do not know the value of σ, but we can estimate it

$$\hat{\sigma}^2 = \frac{\sum_i^N \hat{\varepsilon}_i^2}{N-2}$$

where $\hat{\varepsilon}_i = y_i - (\hat{\beta}_1 - \hat{\beta}_2 x_i)$, and we subtract 2 out of N because we have estimated two parameters $(\hat{\beta}_1, \hat{\beta}_2)$.

This estimator of the variance is unbiased, that is,

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2$$

2 Confidence Intervals

You know that if a random variable W follows $W \sim \mathcal{N}(\mu, \sigma^2)$, we can standardize it to obtain:

$$Z \equiv \frac{W - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Then, we can ask: What is the probability that our random variable Z lies between -1.96 and 1.96?

$$\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$$

*Many thanks to all previous TAs for providing the notes. All mistakes are my own. Please get in touch with me at fdiazvaldes@g.ucla.edu if you spot any typos or mistakes.

[1] In my notation $\mathcal{N}(0, \sigma^2)$, where σ is understood to be the standard deviation of the normal distribution. In some textbooks, you can find it written as $\mathcal{N}(0, \sigma)$, where σ is understood to be the standard deviation of the normal distribution. It is the same, just a notational convention.

An analogous formulation of the previous question is: If you take an infinite sample (large sample) of Z, what are the values of $Z_{\text{lower}} < 0 < Z_{\text{upper}}$ such that the probability of Z lying in the interval $(Z_{\text{lower}}, Z_{\text{upper}})$ is 95%? In mathematical terms,

$$\mathbb{P}(Z_{\text{lower}} \leq Z \leq Z_{\text{upper}}) = 0.95$$

Observation: We got two unknowns $(Z_{\text{lower}}, Z_{\text{upper}})$ and one equation. However, since the normal distribution is symmetrical, we can find those two values. More on this in a bit.

We can generalize the previous equation for any percentage we want. Let's say we want the $(1-\alpha)\%$, then we need to solve for $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$, such that,

$$\mathbb{P}(Z_{\alpha/2} \leq Z \leq Z_{1-\alpha/2}) = (1-\alpha)\%$$

where $Z_{\alpha/2}$ is a real-valued constant such that

$$\mathbb{P}(Z_{\alpha/2} \leq Z) = \frac{\alpha}{2}\%$$

and $Z_{1-\alpha/2}$ is a real-valued constant such that

$$\mathbb{P}(Z_{1-\alpha/2} \leq Z) = \frac{\alpha}{2}\%$$

What does all of this have to do with the distribution of our estimates and the confidence interval?
The answer: Everything. For instance, take the distribution of our $\hat{\beta}_2$

$$\hat{\beta}_2 \sim \mathcal{N}\left(\beta_2, \frac{\sigma^2}{\sum_i (x_i - \overline{x}_i)^2}\right)$$

To make the notation easier for us, define

$$\text{se}(\hat{\beta}_2) \equiv \sqrt{\frac{\sigma^2}{\sum_i (x_i - \overline{x}_i)^2}}$$

Then, we can write,

$$Z \equiv \frac{\hat{\beta}_2 - \beta}{\text{se}(\hat{\beta}_2)}$$

Then,

$$\begin{aligned}
& \mathbb{P}(Z_{\alpha/2} \leq Z \leq Z_{1-\alpha/2}) = (1-\alpha)\% \\
\Leftrightarrow \quad & \mathbb{P}\left(Z_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta}{\text{se}(\hat{\beta}_2)} \leq Z_{1-\alpha/2}\right) = (1-\alpha)\% \\
\Leftrightarrow \quad & \mathbb{P}\left(\hat{\beta}_2 - Z_{\alpha/2} \cdot \text{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + Z_{1-\alpha/2} \cdot \text{se}(\hat{\beta}_2)\right) = (1-\alpha)\%
\end{aligned}$$

And, congratulations, you have just found the confidence interval for β_2 when the errors in our model follow a normal distribution.

However, there is a little issue, we do not know the value of σ. We can estimate it and it yields $\hat{\sigma}^2$. Let's rewrite our notation for the standard error of our estimate

$$\hat{\text{se}}(\hat{\beta}_2) \equiv \sqrt{\frac{\hat{\sigma}^2}{\sum_i (x_i - \overline{x}_i)^2}}$$

If we use $\hat{\text{se}}(\hat{\beta}_2)$ to standardize our estimate, we will no longer get a normal distribution; instead, we will obtain a $t-$student distribution with $N-2$ degrees of freedom, which we denote by df. Note that $\text{df} = N-2$. Thus,

$$t = \frac{\hat{\beta}_2 - \beta_2}{\hat{\text{se}}(\hat{\beta}_2)} \sim t_{\text{df}}$$

The distribution for the $t-$student is symmetrical if $\text{df} > 3$ and it has fatter tails than a normal distribution. However, as df grows larger, the $t-$student distribution converges to the normal distribution.

Our confidence interval is now given by

$$\mathbb{P}\left(\hat{\beta}_2 - t_{\text{df},\alpha/2} \cdot \text{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\text{df},1-\alpha/2} \cdot \text{se}(\hat{\beta}_2)\right) = (1-\alpha)\%$$

Hence, the Confidence Interval (CI) at $1-\alpha\%$ is given by $[\beta_{2,\text{lower}}\ ,\ \beta_{2,\text{upper}}]$:

$$\beta_{2,\text{lower}} = \hat{\beta}_2 - t_{\text{df},\alpha/2} \cdot \text{se}(\hat{\beta}_2)$$
$$\beta_{2,\text{upper}} = \hat{\beta}_2 + t_{\text{df},1-\alpha/2} \cdot \text{se}(\hat{\beta}_2)$$

Question: Supposing everything equal, what Confidence Interval would be smaller at $1-\alpha\%$, the one constructed using σ^2 or the one constructed using $\hat{\sigma}^2$? Recall that, for a fixed α, the smaller the CI, the better. Try to provide an intuitive answer and explanation, with no math.

You can also try to construct a one-sided CI at $1-\alpha\%$ by recognizing that

$$\mathbb{P}(Z_\alpha \leq Z) = 1-\alpha \qquad \text{or} \qquad \mathbb{P}(Z \leq Z_{1-\alpha}) = 1-\alpha$$

3 Hypothesis Testing

For hypothesis testing, we need to specify

- The null hypothesis: H_0. Example: $H_0 : \beta_2 = c$
- The alternative hypothesis: H_A, which could be
  - One-sided: "Less than". Example: $H_A : \beta_2 < c$
  - One-sided: "Greater than". Example: $H_A : \beta_2 > c$
  - Two-sided. Example: $H_A : \beta_2 \neq c$
- The test statistic. Given H_0, we can compute our test statistic.
  For instance, if $H_0 : \beta_2 = c$, then our test statistic is

  $$t = \frac{\hat{\beta}_2 - c}{\hat{\text{se}}(\hat{\beta}_2)}$$

- The rejection region. This depends on our test statistic, the alternative hypothesis H_A, and the significance level we choose for our test, α. A lower value of α makes H_0 harder to reject.
  For example, using our t-statistic from before, the rejection rules are:
  - Less than (Left). Reject if $t \leq t_{\text{df},\alpha}$.
  - Greater than (Right). Reject if $t \geq t_{\text{df},1-\alpha}$.
  - Two-sided. Reject if $|t| \geq t_{\text{df},1-\alpha/2}$
- Conclusion: if our test statistic lies in the rejection region, we reject H_0; if not, we do not reject H_0. See your main textbook for illustrations.

The philosophy of hypothesis testing is that we hold the null hypothesis to be true and reject it if and only if the data overwhelmingly indicate, statistically speaking, that H_0 cannot hold true.

3.1 One-tailed test. Greater than (Right)

Suppose that

- $H_0 : \beta_k = c$
- $H_A : \beta_k > c$
- Level of significance is α.

Calculate the t statistic:

$$t_k = \frac{\hat{\beta}_k - c}{\hat{\text{se}}(\hat{\beta}_k)}$$

Then, reject H_0 if $t_k > t_{\text{df},1-\alpha}$. Otherwise, do not reject H_0. We can also calculate our $p-$value and see whether to reject or not H_0. The $p-$value is the accumulated probability that the $t-$distribution is greater than our t_k statistic. Hence,

$$p - \text{value} \equiv \mathbb{P}(t \geq t_k) = 1 - \mathbb{P}(t \leq t_k)$$

Note that if the $p-$value $< \alpha$, then the null is rejected.

3.2 One-tailed test. Less than (Left)

Suppose that

- $H_0 : \beta_k = c$
- $H_A : \beta_k < c$
- Level of significance is α.

Calculate the t statistic:

$$t_k = \frac{\hat{\beta}_k - c}{\hat{\text{se}}(\hat{\beta}_k)}$$

Then, reject H_0 if $t_k < t_{\text{df},\alpha}$. Otherwise, do not reject H_0. We can also calculate our $p-$value and see whether to reject or not H_0. The $p-$value is the accumulated probability that the $t-$distribution is less than our t_k statistic. Hence,

$$p - \text{value} \equiv \mathbb{P}(t \leq t_k)$$

Note that if the $p-$value $< \alpha$, then the null is rejected.

3.3 Two-sided test

Suppose that

- $H_0 : \beta_k = c$
- $H_A : \beta_k \neq c$
- Level of significance is α.

Calculate the t statistic:

$$t_k = \frac{\hat{\beta}_k - c}{\hat{\text{se}}(\hat{\beta}_k)}$$

Then, reject H_0 if $t_k > t_{\text{df},1-\alpha}$ or $t_k < t_{\text{df},\alpha}$. Otherwise, do not reject H_0. We can also calculate our $p-$value and see whether to reject or not H_0. The $p-$ value is the accumulated probability that the $t-$distribution is greater than our t_k statistic or that the $t-$distribution is less than our t_k. Hence,

$$\begin{aligned} p - \text{value} \equiv \mathbb{P}(|t| \geq |t_k|) &= \mathbb{P}\left(\left\{t \geq |t_k|\right\} \vee \left\{t \leq -|t_k|\right\}\right) \\ &= \mathbb{P}(t \geq |t_k|) + \mathbb{P}(t \leq -|t_k|) \\ &= 2\mathbb{P}\big(t > |t_k|\big) \\ &= 2\Big[1 - \mathbb{P}\big(|t_k| < t\big)\Big] \end{aligned}$$

We used that the $t-$distribution is symmetrical if it has more than 3 degrees of freedom. Note that if the $p-$ value $< \alpha$, then the null is rejected.

Again, for graphical illustrations, please see your textbook, chapter 3.