Econ 103: Introduction to Econometrics

Lecture 9 — Confidence Intervals

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §3.1, 3.6; Stock & Watson (4th ed.), §5.2.

Where we are

We have everything we need to start **statistical inference**:

$$b_2 = 10.21, \qquad \mathrm{se}(b_2) = 2.09, \qquad b_2 \mid x \sim N\left(\beta_2,\ \frac{\sigma^2}{\sum(x_i - \bar{x})^2}\right).$$

A point estimate alone says nothing about its reliability. "$\beta_2 = 10.21$" — but how sure are we? Could it plausibly be 6? 14?

Today we report a range of plausible values — an **interval estimate** (confidence interval). We will

- turn the normal b_2 into a usable **t-statistic**,
- build the interval $b_2 \pm t_c\,\mathrm{se}(b_2)$ and interpret it *carefully*,
- and extend it to linear combinations like $\mathbb{E}(y \mid x_0) = \beta_1 + x_0\beta_2$.

Today's plan

1. From the normal to the t-distribution
2. Confidence intervals for a coefficient
3. Linear combinations of parameters

Part 1

From the normal to the t-distribution

Standardizing b_2

Under SR1–SR6, b_2 is conditionally normal. Subtract its mean and divide by its standard deviation to standardize:

$$Z = \frac{b_2 - \beta_2}{\sqrt{\sigma^2 / \sum (x_i - \bar{x})^2}} \sim N(0, 1).$$

Z is **pivotal**: its $N(0,1)$ distribution involves no unknown parameters — so we can read probabilities straight off the normal table, e.g.

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

One snag

Z still contains the unknown σ^2 in the denominator. We cannot actually compute it. We must replace σ^2 with its estimate $\hat{\sigma}^2$ — and that changes the distribution.

Swapping σ^2 for $\hat{\sigma}^2$: the t-distribution

Replace σ^2 by $\hat{\sigma}^2 = \mathrm{SSE}/(N-2)$. The denominator becomes the *standard error*, and the statistic is no longer normal — it follows **Student's t-distribution**:

$$t = \frac{b_2 - \beta_2}{\sqrt{\hat{\sigma}^2 / \sum (x_i - \bar{x})^2}} = \frac{b_2 - \beta_2}{\mathrm{se}(b_2)} \sim t_{(N-2)}$$

and the same holds for b_1. In general, for $k = 1, 2$,

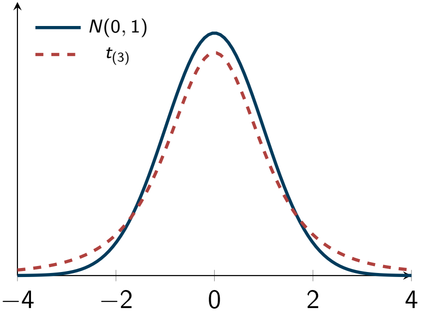
$$t = \frac{b_k - \beta_k}{\mathrm{se}(b_k)} \sim t_{(N-2)}.$$

This single equation is the engine of *both* confidence intervals (today) and hypothesis tests (next lecture). It too is pivotal — no unknown parameters, no dependence on x.

What the t-distribution looks like

- Bell-shaped, symmetric, centered at 0 — like the standard normal.
- But fatter tails and more spread: estimating σ^2 injects extra uncertainty.
- Governed by one number, the **degrees of freedom** $\mathrm{df} = N - 2$.
- As $\mathrm{df} \to \infty$, the t converges to the normal — for large N, $t_c \approx 1.96$.

This is why S&W, who lean on large samples, just use the normal and 1.96. With $N = 40$ we use the exact t.



The t (red) sits lower in the middle, fatter in the tails.

Part 2

Confidence intervals for a coefficient

Building the interval

From Statistical Table 2, pick the **critical value** $t_c = t_{(1-\alpha/2,\, N-2)}$ that puts $\alpha/2$ in each tail, so

$$\mathbb{P}(-t_c \le t \le t_c) = 1 - \alpha.$$

Substitute $t = (b_k - \beta_k)/\mathrm{se}(b_k)$ and rearrange to isolate β_k:

$$\mathbb{P}\big[\, b_k - t_c\,\mathrm{se}(b_k) \le \beta_k \le b_k + t_c\,\mathrm{se}(b_k) \,\big] = 1 - \alpha.$$

The $100(1-\alpha)\%$ interval estimator

$$b_k \pm t_c\,\mathrm{se}(b_k)$$

Three ingredients: the estimate b_k, its standard error, and a critical value that bakes in both the confidence level and the sample size (via df).

Food data: a 95% interval for β_2

$N = 40$, so $\mathrm{df} = 38$ and for $\alpha = 0.05$, $t_c = t_{(0.975,\,38)} = 2.024$. With $b_2 = 10.21$ and $\mathrm{se}(b_2) = 2.09$:

$$b_2 \pm t_c\,\mathrm{se}(b_2) = 10.21 \pm 2.024(2.09) = [\,5.97,\ 14.45\,].$$

Reading it in context

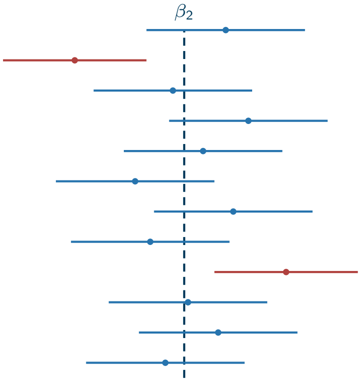
We estimate *with 95% confidence* that from an extra \$100 of weekly income, households spend between \$5.97 and \$14.45 more on food. The range is wide — a single regressor leaves a lot of uncertainty about β_2.

A supermarket CEO planning store capacity would stress-test decisions across this whole range, not just the point estimate 10.21.

What "95% confidence" really means

The confidence is in the procedure, not in the one interval:

- Across *all possible samples*, 95% of the intervals built this way contain β_2.
- *Our* interval $[5.97, 14.45]$ either contains β_2 or it does not — and we will never know which.
- It is *wrong* to say "β_2 has a 95% probability of being in $[5.97, 14.45]$." β_2 is fixed; the interval is what is random.



Most intervals cover β_2; the red ones miss.

The width of the interval is information

$$\text{half-width} = t_c \, \mathrm{se}(b_k) = t_c \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}.$$

- Narrow interval ⇔ small se ⇔ the data pin down β_k sharply — we have *learned a lot*.
- Wide interval ⇔ large se ⇔ little information about β_k.
- Everything that shrank $\mathrm{se}(b_2)$ in Lecture 7 — smaller σ^2, more spread-out x, larger N — also narrows the interval.
- Higher confidence (99% vs. 95%) means a *bigger* t_c, hence a *wider* interval: more coverage costs precision.

The handy large-sample shortcut

When $\mathrm{df} = N - 2 > 30$, $t_c \approx 2$, so a quick 95% interval is $b_k \pm 2\,\mathrm{se}(b_k)$. This is the rule of thumb behind "two standard errors."

Part 3

Linear combinations of parameters

Beyond one coefficient at a time

Often the quantity of interest mixes *both* parameters — a **linear combination**

$$\lambda = c_1\beta_1 + c_2\beta_2,$$

with c_1, c_2 constants we choose. The headline case is the conditional mean at a specific x_0:

$$\mathbb{E}(y \mid x = x_0) = \beta_1 + x_0\,\beta_2 \qquad (c_1 = 1,\ c_2 = x_0).$$

We estimate it the natural way, and — because b_1, b_2 are BLUE — the estimator $\hat{\lambda} = c_1 b_1 + c_2 b_2$ is the best linear unbiased estimator of λ:

$$\mathbb{E}(\hat{\lambda} \mid x) = c_1\mathbb{E}(b_1 \mid x) + c_2\mathbb{E}(b_2 \mid x) = c_1\beta_1 + c_2\beta_2 = \lambda.$$

The standard error of a linear combination

The point estimate is easy; the standard error needs the variance-of-a-sum rule from Lecture 3 — and the covariance term matters:

$$\mathrm{Var}(\hat{\lambda} \mid x) = c_1^2 \,\mathrm{Var}(b_1 \mid x) + c_2^2 \,\mathrm{Var}(b_2 \mid x) + 2c_1 c_2 \,\mathrm{Cov}(b_1, b_2 \mid x).$$

Plug in the *estimated* variances/covariance (Lecture 8) and take the root:

$$\mathrm{se}(\hat{\lambda}) = \sqrt{\widehat{\mathrm{Var}}(\hat{\lambda} \mid x)}.$$

Don't forget the covariance

A common mistake is to add only $c_1^2\widehat{\mathrm{Var}}(b_1) + c_2^2\widehat{\mathrm{Var}}(b_2)$. Because b_1 and b_2 are correlated (recall $\mathrm{Cov}(b_1, b_2) < 0$ when $\bar{x} > 0$), the cross term is essential.

Food data: a CI for expected food expenditure

"Estimate average weekly food spending for households with \$2,000 income" means $\mathbb{E}(y \mid x_0 = 20) = \beta_1 + 20\beta_2$. Point estimate:

$$\hat{\lambda} = b_1 + 20\, b_2 = 83.42 + 20(10.21) = 287.61.$$

Using $\widehat{\mathrm{Var}}(b_1) = 1884.44$, $\widehat{\mathrm{Var}}(b_2) = 4.3818$, $\widehat{\mathrm{Cov}}(b_1, b_2) = -85.90$:

$$\widehat{\mathrm{Var}}(\hat{\lambda}) = 1884.44 + 20^2(4.3818) + 2(20)(-85.90) = 201.02,$$

$$\mathrm{se}(\hat{\lambda}) = \sqrt{201.02} = 14.18.$$

A 95% interval ($t_c = 2.024$):

$$287.61 \pm 2.024(14.18) = [\,258.91,\ 316.31\,].$$

With 95% confidence, the *average* such household spends between \$258.91 and \$316.31 on food.

Mean vs. outcome: two different bands

Note how *tight* the interval for the mean is ($\pm$\$28.70) compared with predicting a single household's spending.

CI for the mean $\mathbb{E}(y \mid x_0)$

[258.91, 316.31]. Only sources of error: estimating b_1, b_2. This is today's linear-combination interval.

Prediction interval for y_0

Much wider: it must also absorb the new household's own shock e_0. Built from the forecast error of Lecture 8 — the full mechanics come in Lecture 11.

Same center (287.61), very different widths. Always ask: am I estimating an average, or forecasting an individual outcome?

Recap

The t-statistic

- standardize b_k, swap $\sigma^2 \to \hat{\sigma}^2$: $\frac{b_k - \beta_k}{\mathrm{se}(b_k)} \sim t_{(N-2)}$
- t: bell, fat tails, df $= N-2$; $\to$ normal as N grows

Confidence interval

- $b_k \pm t_c\,\mathrm{se}(b_k)$, $t_c = t_{(1-\alpha/2, N-2)}$
- food: $\beta_2 \in [5.97, 14.45]$
- confidence is in the procedure; 95% of intervals cover

Linear combinations

- $\hat{\lambda} = c_1 b_1 + c_2 b_2$ (BLUE of λ)
- $\mathrm{Var}(\hat{\lambda}) = c_1^2 \mathrm{Var}\, b_1 + c_2^2 \mathrm{Var}\, b_2 + 2c_1 c_2 \mathrm{Cov}(b_1, b_2)$
- $\mathbb{E}(y \mid x_0{=}20)$: 287.61, CI $[258.91, 316.31]$
- mean band $\ll$ outcome (prediction) band

Next time (Lecture 10): hypothesis testing

Same $t = (b - c)/\mathrm{se}(b)$ engine, pointed at a conjecture: is $\beta_2 = 0$? > 5.5? We will reject or fail to reject, with p-values and a word on *statistical vs. economic* significance.

Questions?