Econ 103: Introduction to Econometrics

Lecture 10 — Hypothesis Testing

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §3.2–3.5; Stock & Watson (4th ed.), §5.1.

Where we are

Last lecture, the ***t*-statistic** gave us a *range* of plausible values:

$$\frac{b_k - \beta_k}{\mathrm{se}(b_k)} \sim t_{(N-2)} \implies \beta_2 \in [5.97,\ 14.45].$$

Today we point the *same* engine at specific conjectures. Decision-makers ask yes/no questions:

- Is there *any* relationship — is $\beta_2 = 0$?
- Will households spend more than \$5.50 of each extra \$100 — is $\beta_2 > 5.5$?

We will set up null and alternative hypotheses, build rejection regions, compute p-values, and — crucially — separate statistical from economic significance.

Today's plan

1. The logic of a hypothesis test
2. Rejection regions for the three alternatives
3. Worked tests on the food data
4. The p-value approach
5. Statistical vs. economic significance

Part 1

The logic of a hypothesis test

Five ingredients of every test

Components of a hypothesis test

1. a **null hypothesis** H_0,
2. an **alternative hypothesis** H_1,
3. a **test statistic**,
4. a **rejection region**,
5. a **conclusion** (stated in economic context).

Null $H_0 : \beta_k = c$

The maintained belief — held until the data convince us otherwise. c is a value that matters in context (often 0).

Alternative H_1

What we accept if we reject H_0. Three flavors: $\beta_k > c$, $\beta_k < c$, or $\beta_k \neq c$.

The test statistic and its logic

Recall $t = (b_k - \beta_k)/\mathrm{se}(b_k) \sim t_{(N-2)}$. *If $H_0 : \beta_k = c$ is true*, substitute c for β_k:

$$t = \frac{b_k - c}{\mathrm{se}(b_k)} \sim t_{(N-2)} \quad \text{under } H_0$$

A test statistic has a known distribution when H_0 is true, and some other distribution when it is false.

The chain of logic:

- If H_0 holds, t should land in the middle of the t-curve, where most probability sits.
- A t way out in a tail is *unlikely* under H_0.
- Observing such a value is evidence that H_0 is false — so we reject it.

How unlikely is "unlikely"? The level of significance

We draw the line with the **level of significance** α — the probability of landing in the rejection region *when* H_0 *is true*.

- Rejecting H_0 when it is actually true is a **Type I error**. $\mathbb{P}(\text{Type I error}) = \alpha$. We *choose* α (usually $0.01, 0.05, 0.10$).
- Failing to reject a *false* H_0 is a **Type II error**. Its probability depends on the unknown β_k — we can't set it directly.

Choosing α is choosing your risk

A costly false rejection $\Rightarrow$ small α (e.g. 0.01). The "$\alpha = 0.05$" default is just convention — Fisher's old "$t > 2$ is significant" rule of thumb, not a law of nature.

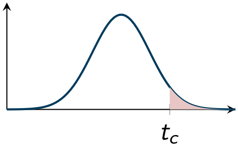
Part 2

Rejection regions for the three alternatives

One-tail and two-tail rejection regions

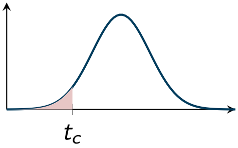
Where the rejection region sits depends on H_1. **Memory trick:** the rejection region is in the direction of the arrow in the alternative.

$H_1 : \beta_k > c$ (right tail)



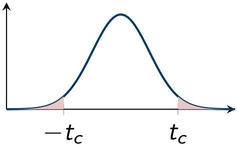
reject if $t \geq t_{(1-\alpha, N-2)}$

$H_1 : \beta_k < c$ (left tail)



reject if $t \leq t_{(\alpha, N-2)}$

$H_1 : \beta_k \neq c$ (two tails)



reject if $|t| \geq t_{(1-\alpha/2, N-2)}$

For a two-tail test the area α is split, $\alpha/2$ in each tail, so the critical value t_c is *larger* than for a one-tail test at the same α.

The five-step procedure

We run *every* test by the same checklist:

1. **Hypotheses.** State H_0 and H_1.
2. **Test statistic.** $t = \dfrac{b_k - c}{\mathrm{se}(b_k)}$, which is $t_{(N-2)}$ if H_0 is true.
3. **Rejection region.** Pick α; find the critical value(s) for the relevant tail(s).
4. **Compute.** Plug in b_k, c, $\mathrm{se}(b_k)$.
5. **Conclude.** Reject or do not reject — and say what it means for the economics.

A subtle but important habit: we say "reject H_0" or "*fail to reject* H_0" — never "accept H_0." Not rejecting only means the data are *compatible* with H_0, not that H_0 is true.

Part 3

Worked tests on the food data

Test of significance: is there any relationship?

Economic theory says food is a normal good, so $\beta_2 > 0$. Put the conjecture in H_1, so that rejecting H_0 *establishes* it.

1. $H_0 : \beta_2 = 0 \quad$ vs. $\quad H_1 : \beta_2 > 0$.
2. $t = b_2/\mathrm{se}(b_2) \sim t_{(38)}$ under H_0 (here $c = 0$).
3. $\alpha = 0.05$; right-tail critical value $t_{(0.95,\,38)} = 1.686$. Reject if $t \geq 1.686$.
4. $t = \dfrac{10.21}{2.09} = 4.88$.
5. $4.88 > 1.686$, so reject H_0. There is a statistically significant *positive* relationship between income and food expenditure.

Had we *failed* to reject, we would *not* conclude "theory is wrong" — only that this sample lacks the evidence.

Two-tail test of significance (what software reports)

If we have no prior sign in mind, test against $\neq$:

1. $H_0 : \beta_2 = 0 \quad$ vs. $\quad H_1 : \beta_2 \neq 0$.
2. $t = b_2/\mathrm{se}(b_2) \sim t_{(38)}$ under H_0.
3. $\alpha = 0.05$; critical values $\pm t_{(0.975, 38)} = \pm 2.024$. Reject if $|t| \geq 2.024$.
4. $t = 10.21/2.09 = 4.88$.
5. $|4.88| > 2.024$: reject H_0 — β_2 is significantly different from zero.

This is automatic

Every regression printout reports, for each coefficient, $t = b_k/\mathrm{se}(b_k)$ for $H_0 : \beta_k = 0$. In the food output, INCOME has $t = 4.88$ (slope significant); the intercept has $t = 1.92$ (not significant at 5%).

Testing an economic value, and why α matters

A new supermarket is profitable only if households spend more than \$5.50 of each extra \$100 on food. Put the make-or-break claim in H_1:

1. $H_0 : \beta_2 \leq 5.5 \quad \text{vs.} \quad H_1 : \beta_2 > 5.5.$
2. $t = \dfrac{b_2 - 5.5}{\mathrm{se}(b_2)} \sim t_{(38)}$ under H_0.
3. Choose a conservative $\alpha = 0.01$ (a wrong "build" is costly); critical value $t_{(0.99,\,38)} = 2.429$.
4. $t = \dfrac{10.21 - 5.5}{2.09} = 2.25.$
5. $2.25 < 2.429$: do not reject H_0. Not enough evidence of profitability — do not build (yet).

At $\alpha = 0.05$ (critical value 1.686) we *would* have rejected. The decision flips with α — choose it *before* seeing the data, by weighing the cost of a Type I error.

Part 4

The p-value approach

The p-value

Rather than memorize critical values, report the ***p*-value**: the probability, *if* H_0 *is true*, of a test statistic at least as extreme as the one observed.

The p-value rule

$$p \leq \alpha \;\Rightarrow\; \text{reject } H_0, \qquad p > \alpha \;\Rightarrow\; \text{do not reject } H_0.$$

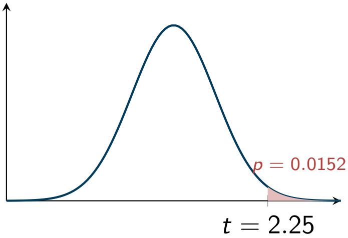
How it is computed depends on H_1 (same memory trick — look where the arrow points):

- $H_1 : \beta_k > c$: $p = \mathbb{P}(t_{(N-2)} \geq t)$ (right tail);
- $H_1 : \beta_k < c$: $p = \mathbb{P}(t_{(N-2)} \leq t)$ (left tail);
- $H_1 : \beta_k \neq c$: $p = 2\,\mathbb{P}(t_{(N-2)} \geq |t|)$ (both tails).

p-values on the food data

- **Supermarket test** ($H_1 : \beta_2 > 5.5$, $t = 2.25$): $p = \mathbb{P}(t_{(38)} \geq 2.25) = 0.0152$. Since $0.0152 > 0.01$, do not reject at $\alpha = 0.01$ — but *would* reject at 0.05. Same answer as the critical-value approach.
- **Two-tail significance** ($H_1 : \beta_2 \neq 0$, $t = 4.88$): $p \approx 0.0000$. Reject at any usual α.

Software's "Prob." column is exactly this two-tail p-value for $H_0 : \beta_k = 0$.



Right-tail p-value = shaded area beyond the observed t.

Why report p instead of just "reject"? It lets each reader apply their own α — a much more informative summary.

Part 5

Statistical vs. economic significance

Significant $\neq$ important

A statistically significant coefficient can be economically trivial.

Example (A large sample, a tiny effect)

Suppose $b_2 = 0.0001$ *with* $\mathrm{se}(b_2) = 0.00001$, *so*

$$t = \frac{0.0001}{0.00001} = 10.0.$$

We resoundingly reject $H_0 : \beta_2 = 0$ *—* b_2 *is statistically different from zero. But* 0.0001 *may be far too small to matter for any real decision. Statistically significant, economically negligible.*

The reverse also happens: an economically large effect can be statistically insignificant if the sample is small or noisy (a wide confidence interval).

Reading significance responsibly

- Sample size inflates significance. As N grows, $\mathrm{se}(b_k) \to 0$, so almost any $\beta_k \neq 0$ eventually becomes "significant." Significance alone says little about importance.
- Look at the magnitude. Is the estimated effect big enough to change a decision? Report the coefficient and its *units*, not just stars.
- Use the confidence interval. It shows both significance (does it exclude the null?) *and* the range of economically relevant values.
- State the economic conclusion. "We reject H_0" is not an answer — "income significantly raises food spending, by about \$10 per \$100, give or take \$4" is.

Statistical procedures are means, not ends. Always translate back to the economic question that motivated the test.

Recap

The machinery

- $H_0 : \beta_k = c$ vs. H_1 $(>, <, \neq)$
- under H_0: $t = \dfrac{b_k - c}{\mathrm{se}(b_k)} \sim t_{(N-2)}$
- reject if t falls in the tail(s); arrow points to the rejection region
- $\alpha = \mathbb{P}(\text{Type I})$; chosen by cost of false rejection

p-value

- prob. of a t this extreme under H_0
- reject iff $p \leq \alpha$

Food data

- $H_0 : \beta_2 = 0$ vs. $>$: $t = 4.88$, reject — significant positive effect
- $H_0 : \beta_2 \leq 5.5$: $t = 2.25$, $p = 0.015$ — reject at 5%, not at 1%

Significance

- statistical $\neq$ economic
- big $N \Rightarrow$ everything "significant"
- judge by magnitude + CI, in context

Where this leaves us (and the midterm)

We can now estimate, quantify precision, predict, and test in the simple regression model. After the midterm: goodness-of-fit (R^2), functional forms, and *multiple* regression.

Questions?