Interpreting Tests and Choosing Controls

14 Oct 2026 | Chapter 7

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

From Lecture 14: what did the joint test tell us?

For the school model with STR, English learners, and spending,

$$H_0 : \beta_{STR} = \beta_{Spend} = 0, \qquad F(2, 416) = 5.43, \quad p = 0.0047.$$

Discuss with a partner

Which statements follow from this result?

1. Both coefficients must be individually significant.
2. The two zero restrictions are rejected together.
3. Reducing class size will necessarily raise scores.

A joint rejection leaves two questions open

The statistical conclusion

The data provide evidence against the claim that both conditional coefficients are zero.

- They do not tell us that *each* coefficient is nonzero.
- They do not remove omitted-variable bias or policy selection.

A test answers its stated null. A policy recommendation needs a credible causal comparison.

Three questions a superintendent might ask

The outcome is the district test score. STR is students per teacher; spending is dollars per pupil; EL is the percentage of English learners.

$$Score_i = \beta_0 + \beta_{STR} STR_i + \beta_{EL} EL_i + \beta_{Spend} Spend_i + u_i.$$

1. Is there a conditional association between STR and scores?
2. What fitted change accompanies two fewer students per teacher *and* \$1, 000 more spending per pupil?
3. Do STR and spending contribute anything jointly, conditional on EL?

Group task: translate the question before choosing a test

Work in groups of three – 6 minutes

For each question on the previous page:

1. Write the coefficient or combination of coefficients of interest.
2. If a test is wanted, write its null and count the restrictions.
3. Identify what remains fixed in the comparison.

For the proposed policy, ask for an estimate and an interval, not merely whether two coefficients are separately significant.

Debrief: two coefficients can define one restriction

Question	Mathematical target	q
STR coefficient	$H_0 : \beta_{STR} = 0$	1
Combined change	$\Delta = -2\beta_{STR} + 1000\beta_{Spend}$; test $H_0 : \Delta = 0$ if required	1
Joint contribution	$H_0 : \beta_{STR} = 0$ and $\beta_{Spend} = 0$	2

The combined change can be zero even when both coefficients are nonzero: their contributions can cancel.

What makes a comparison useful for a causal question?

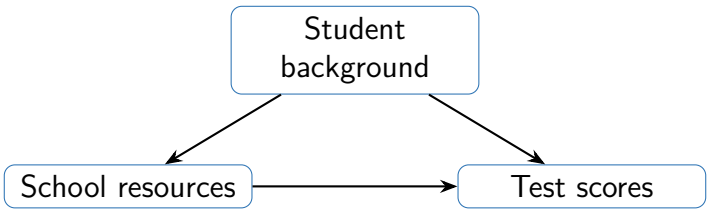
Suppose districts with more disadvantaged students choose different staffing and spending levels. Student background also predicts test scores.

The role of a control

We want to compare districts with similar relevant background, rather than attribute background differences to school resources.

- A small p-value does not establish that this comparison is credible.
- Robust SEs address heteroskedasticity, not confounding.

Student background is not randomly assigned



This is a possible causal story, not a result established by the regression.

The subsidized-lunch share is a proxy for economic background. Its coefficient is not automatically the effect of providing lunches.

Stata: begin with the same school specification

```
. regress score str el spend, vce(robust)
Linear regression                               Number of obs     =        420
                                                R-squared         =     0.4366
------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234002     .661203
          el |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       spend |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

Locate the STR estimate, its interval, and the EL coefficient. Keep this result for the next comparison.

Stata: add the background proxy on the same sample

```
. regress score str el spend meal_pct, vce(robust)
Linear regression                               Number of obs     =        420
                                                R-squared         =     0.7834
------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -.2353884   .3248148    -0.72   0.469    -.8738757    .4030989
          el |  -.1283415   .0324397    -3.96   0.000    -.1921082   -.0645748
       spend |    .003622   .0009447     3.83   0.000       .001765    .0054791
    meal_pct |  -.5463929   .0231685   -23.58   0.000    -.5919351   -.5008506
       _cons |   665.9882   10.37683    64.18   0.000     645.5905    686.3859
------------------------------------------------------------------------------
```

`meal_pct` is the subsidized-lunch share, in percentage points. Both regressions use the same 420 districts.

Group task: audit the specification comparison

Use the two output pages – 7 minutes

1. Did the STR estimate change much? Did its precision change?
2. Why might the EL coefficient change after controlling for economic background?
3. Does the higher R^2 prove that school resources now have causal effects?
4. Should we drop STR solely because its p-value is above 0.05?

Prepare a two-sentence recommendation that separates fit, precision, and identification.

Debrief: better fit is not the same as identification

- STR: estimate $-0.286 \rightarrow -0.235$; SE $0.482 \rightarrow 0.325$.
- EL: estimate $-0.656 \rightarrow -0.128$ after adding the background proxy.
- R^2: $0.437 \rightarrow 0.783$, on the same observations.

A defensible conclusion

The comparison shows sensitivity to the conditioning set and improved fit. It does not prove that remaining omitted factors are unrelated to school resources.

A second case: decide which comparison is credible

Hospital waiting times – 6 minutes

A researcher studies the effect of adding nurses on average patient waiting time. Candidate controls are:

- patient volume measured before staffing changes;
- the pre-existing severity mix of patients;
- service capacity created *by* the added nurses.

Which belong in a model for the *total* staffing effect? Give a causal reason for each decision.

Debrief and the next step

For the total staffing effect

Consider pre-existing volume and severity as potential confounders. Do not automatically control for capacity that is created by the staffing change.

1. State the empirical target.
2. Choose a comparison and controls that fit that target.
3. Use an appropriate test or interval, then qualify the causal claim.

On Sunday we will carry these targets into Stata and calculate the uncertainty of a combined change.