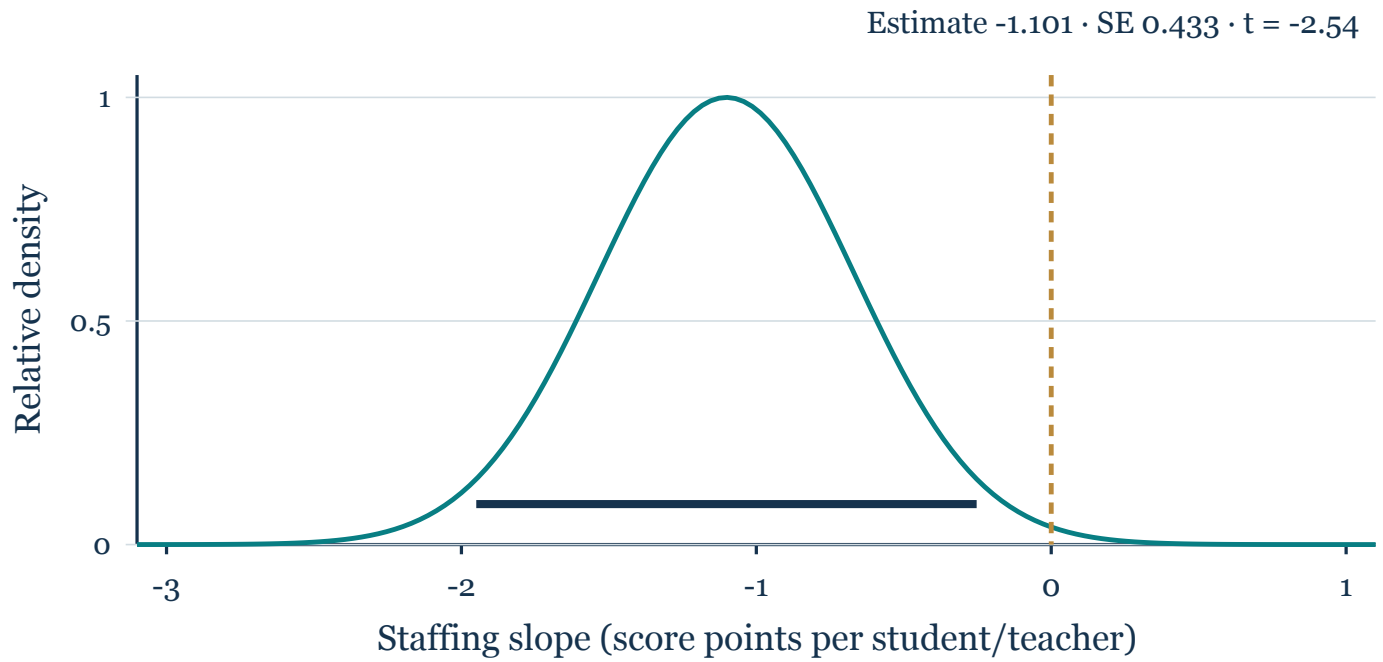
Is the staffing relationship large enough to matter?

A decision needs magnitude and uncertainty.



Actual staffing estimate and robust SE. The reference curve is a large-sample normal approximation, not the distribution of district outcomes.

> Question: how accurately does the sample measure the relevant magnitude?

> Inference adds an uncertainty argument to the fitted estimate.

Recall which coefficient the model estimates

Keep its units and adjustment fixed.

> The adjusted staffing slope is -1.101 points per additional student per teacher.

> Before calculating a test, state the exact target.

State a null before calculating a p-value

The evidence is evaluated under a specific claim.

> $H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$

> $t = (\widehat{\beta}_1 - \beta_{1,0}) / SE(\widehat{\beta}_1)$

> p-value: a tail probability under the null and reference assumptions.

Normal-error exact inference and robust approximations differ

Know which argument supports the reference distribution.

> Exact conventional t/F: normal, independent, equal-variance errors under the linear model.

> Robust: large-sample approximation allowing unequal variance; dependence needs separate treatment.

State the reference distribution before choosing a critical value

An exact classical t test uses its residual degrees of freedom.

> Under the classical normal-error model and $H_0 : \beta_j = b_j^0$, $T = (\widehat{\beta}_j - b_j^0)/SE_j \sim t_{n-p}$.

> Choose c from $P(|t_{n-p}| > c) = 0.05$.

> Interval: $\widehat{\beta}_j \pm cSE_j$.

Build a confidence interval step by step

Combine the estimate with a stated reference rule.

> $\widehat{\beta} \pm 1.96\,\mathrm{SE}(\widehat{\beta})$

One-sided alternatives must express a prior scientific question

The direction of the alternative changes the tail probability.

> Two-sided: $H_1 : \beta_j \neq b_j^0$; upper-sided: $H_1 : \beta_j > b_j^0$.

> Upper-sided p-value: $P(T \geq t_{obs} \mid H_0)$.

> Choose the direction before inspecting the estimate.

Test a substantively meaningful value

A zero null is not the only relevant question.

```
. test str=-1

 ( 1)  str = -1

       F(  1,   417) =    0.05
            Prob > F =    0.8151
```

Compute uncertainty for the policy contrast

Scale the SE along with the estimate.

```
. lincom -2*str

------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   2.202592   .8656943     2.54   0.011     .5009232     3.90426
------------------------------------------------------------------------------
```

$$\mathrm{SE}(-2\widehat{\beta}_1) = 2\,\mathrm{SE}(\widehat{\beta}_1)$$

A linear contrast uses cross-coefficient covariance

Combine coefficients in the units of the intended fitted comparison.

> Contrast $\hat{\theta} = ab_1 + cb_2$. Here $a = -2$ and $c = 5$.

> $\operatorname{Var}(\hat{\theta}) = a^2 \operatorname{Var}(b_1) + c^2 \operatorname{Var}(b_2) + 2ac \operatorname{Cov}(b_1, b_2)$.

> Test $\theta = \theta_0$ with $(\hat{\theta} - \theta_0)/SE(\hat{\theta})$ and the justified reference.

Write an uncertainty statement for a policy reader

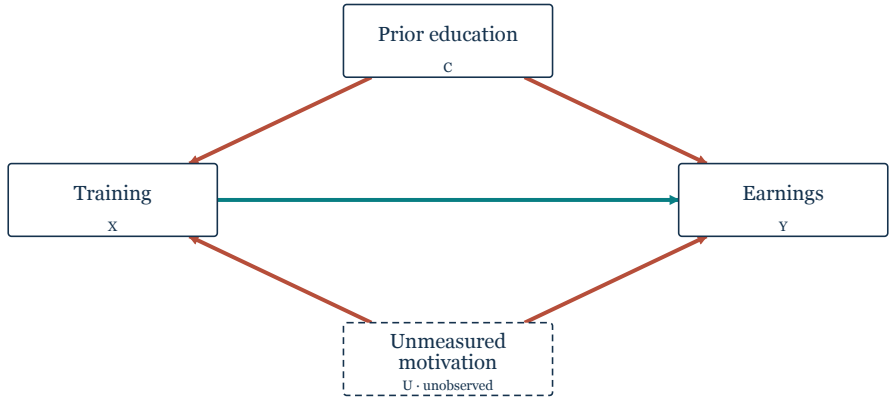
Include the comparison, units and conditions.

> A concise result sentence names the adjusted comparison, gives a feasible policy contrast and reports its uncertainty method.

> For example, describe a fitted two-student ratio difference at fixed language composition and give its robust interval in score points.

DAG practice: uncertainty is not adjustment

Trace the routes before choosing a control.



1 open directed route · 2 open alternatives · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

Practice

Select prior education as a control in the training diagram. Which alternative route remains? If the analyst changes conventional SEs to heteroskedasticity-robust SEs, does that close the route or change the training coefficient on an identical specification and sample?

DAG practice: uncertainty is not adjustment: explanation

Explain what conditioning changes in this stated graph.

> The route through unmeasured motivation remains open after conditioning on education. The common-cause arrows are properties of the proposed causal story, not of the covariance formula. Switching the SE option does not block them. On an unchanged model and sample, the OLS coefficient remains the same while the estimated covariance, intervals and tests can change. Robust inference addresses uncertainty under appropriate conditions. A credible causal comparison needs an additional argument about motivation or a design that addresses its role.

A joint hypothesis asks whether several restrictions hold together

One coefficient test cannot replace the combined question.

```
. test 1.high 1.high#c.str

 ( 1)  1.high = 0
 ( 2)  1.high#c.str = 0

       F(  2,   415) =   98.88
            Prob > F =    0.0000
```

- Example joint null: $H_0 : \beta_{el} = 0,\ \beta_{income} = 0$.
- Estimate the restricted and unrestricted models on the same observations.
- Use the reference distribution justified by the covariance assumptions.

Classical fit-loss F tests require matched models

Restricted and unrestricted fits must answer one nested comparison.

> $$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n - p_U)}$$

> Example: $F = \dfrac{(110 - 100)/2}{100/(100 - 5)} = 4.75$

School staffing: reproduce and explain

Reproduce the evidence, then explain its economic meaning.

Practice

Add district income to the model already controlling for language share. Recalculate the fitted gain for a two-student reduction, with its robust SE. Keep your commands, log, output and interpretation.

School staffing: reproduce and explain: explanation

A reproducible result needs both a calculation and an interpretation.

> Original two-student fitted gains: +4.56 and +2.20 points.

> With income: gain 0.137551, robust SE 0.580319.

> The smaller adjusted association does not establish a causal intervention effect.

From finite-sample inference to large-sample reasoning

> From finite-sample inference to large-sample reasoning

Confidence interval

Try the comparison before opening the explanation.

Practice

An estimate is 0.30 with robust SE 0.12. Use the large-sample 1.96 rule to form a 95% interval and test a zero effect.

Confidence interval: explanation

Follow the calculation and interpret its units.

> $0.30 \pm 1.96 \times 0.12$ gives [0.0648,0.5352]. The t ratio is 2.5, so reject zero at approximately 5% under the relevant large-sample assumptions. This says nothing by itself about the economic importance or causal validity of the estimate.

A rejection is not an effect-size statement

Statistical significance and practical value answer different questions.

> A statistically significant result passes a chosen testing threshold.

> For a staffing decision, translate the slope into a feasible ratio change and compare the implied learning difference with costs and relevant alternatives.

A confidence interval can contain an important effect

Failing to reject is not proof of no effect.

> If an interval includes zero and also includes effects large enough to matter, the evidence may be too imprecise for the decision.

> Report the point estimate and interval, explain which practical values remain compatible with the model and discuss what data would improve precision.

Economic versus statistical size

Try the comparison before opening the explanation.

Practice

A precisely estimated policy effect is 0.001 outcome units. Is a small p-value enough to recommend the policy?

Economic versus statistical size: explanation

Follow the calculation and interpret its units.

> No. Report the units, realistic policy contrast, interval, policy cost and relevant outcome scale. Statistical detectability alone neither establishes causal validity nor a worthwhile benefit.

More observations help only under the relevant conditions

The information gain depends on what the observations add.

> Illustrative rate: SE $\propto n^{-1/2}$

Avoid repeated specification searches for a preferred p-value

The search changes what an isolated test can mean.

> If an analyst tries many outcomes, control sets and sample restrictions and reports only the most significant result, the reported p-value does not describe the whole search.

> Keep a clear account of the main question and intended specification.

Check the reference distribution and scale

An interval is meaningful only for a defined quantity.

> Before interpreting an interval, identify whether it concerns a coefficient, a fitted contrast, an average outcome or an individual prediction.

> Also check whether the displayed interval uses a normal approximation or a t reference and whether the covariance option matches the observations.

Use the evidence to identify the next useful step

Precision, fit and design can require different improvements.

> An imprecise but credible comparison may benefit from more informative observations.

> These are different diagnoses.

Report magnitude and uncertainty together

Reproduce the evidence, then explain its economic meaning.

Practice

Use the fitted two-student gain to assess a null gain of one score point, rather than zero. Calculate the corresponding test ratio and explain its economic meaning. Keep your commands, log, output and interpretation.

Report magnitude and uncertainty together: explanation

A reproducible result needs both a calculation and an interpretation.

> Two-student gain: 2.202592, robust SE 0.865694.

> Null gain of one: $(2.202592 - 1)/0.865694 = 1.389164$.

> Separate a policy-relevant magnitude from a test against zero.