Estimation, Hypothesis Tests, and Confidence Intervals

02 Sep 2026 | Chapter 3

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Recall: Sampling variation

From Lecture 3

The commuting example treated daily times as independent draws from a population with the following mean and SD.

$$\mu_Y = 20, \quad \sigma_Y = 10, \quad n = 100$$

Calculate $SD(\bar{Y})$. Does it describe variation in individual trips or variation across repeated 100-day study averages?

What the sampling example already gives us

Answer

$$SD(\bar{Y}) = 10/\sqrt{100} = 1 \text{ minute}$$

This is the spread of averages across repeated 100-day studies, not the spread of individual trip times. The population mean remains fixed.

In practice we observe the sample but do not know the population mean. Can we reverse the reasoning?

An estimator is a rule; an estimate is its realized value

Before observing the sample

$\bar{Y}$ is an estimator: a random variable with a sampling distribution.

After observing this sample

22.64 is the estimate: the numerical value produced by the rule.

$$E(Y_1) = E(\bar{Y}) = \mu_Y, \qquad \text{Var}(\bar{Y}) = \text{Var}(Y_1)/200$$

Both rules are centered correctly, but $\bar{Y}$ uses all 200 observations and varies less across samples.

First estimate variation among the individual wages

The population SD is unknown

For wages $Y_1, \ldots, Y_n$, measure each wage relative to the sample mean.

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

s_Y describes differences among workers. Dividing it by $\sqrt{n}$ will describe uncertainty about their population mean.

Convert wage dispersion into uncertainty about the mean

Across individual graduates

$$s_Y = 18.14 \text{ dollars}$$

Across repeated sample means

$$\widehat{SE}(\bar{Y}) = \frac{s_Y}{\sqrt{n}} = \frac{18.14}{\sqrt{200}} = 1.283$$

The standard error is the estimated SD of the estimator, not the SD of individual wages.

A hypothesis test begins with a population claim

$$H_0 : \mu_Y = 20 \qquad \text{versus} \qquad H_1 : \mu_Y \neq 20$$

Null world

Pretend for the calculation that the population mean really is 20.

Evidence question

Would random sampling often produce a sample mean at least this far from 20?

The test concerns the fixed population mean, not whether the observed mean equals 20.

The test statistic compares signal with sampling noise

$$t = \frac{\bar{Y} - \mu_0}{\widehat{SE}(\bar{Y})} = \frac{22.64 - 20}{1.28} = 2.06$$

Numerator: signal

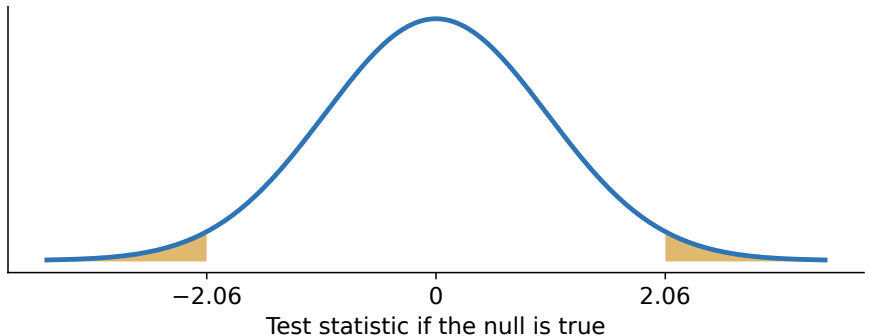
The estimate is 2.64 dollars above the null value.

Denominator: noise scale

A comparable sample mean typically varies by about 1.28 dollars.

The estimate is 2.06 estimated standard errors above the null benchmark.

The p-value asks how unusual this result is under the null



For a two-sided test, count both tails: $p \approx 2[1 - \Phi(|2.06|)] \approx 0.039$.

This is $P(\text{data at least this extreme} \mid H_0)$, not $P(H_0 \mid \text{data})$.

Choose the error rate before looking at the result

$$\alpha = .05, \qquad p \approx .039 < .05 \quad \Longrightarrow \quad \text{reject } H_0$$

If the null is true

The procedure rejects with probability about α: a Type I error.

If the null is false

Power is the probability that the procedure detects the departure.

Rejection is evidence against $\mu_Y = 20$ under the sampling assumptions; it does not reveal an economic mechanism.

A second survey gives a different decision

A separate illustrative survey estimates mean earnings at 24 dollars with an estimated standard error of 2.5 dollars. Test $H_0 : \mu_Y = 20$ against a two-sided alternative at 5%, using the normal approximation.

Solution: A second survey gives a different decision

$$t = (24 - 20)/2.5 = 1.6, \qquad p \approx 2[1 - \Phi(1.6)] \approx 0.110$$

Do not reject at 5%. The estimated difference is larger than before, but the survey is also less precise.

A confidence interval collects the values that pass the test

A candidate value μ_0 remains in the 95% confidence set when its two-sided 5% test does not reject:

$$\left| \frac{\bar{Y} - \mu_0}{\widehat{SE}(\bar{Y})} \right| \leq 1.96$$

$$\bar{Y} - 1.96\widehat{SE}(\bar{Y}) \leq \mu_0 \leq \bar{Y} + 1.96\widehat{SE}(\bar{Y})$$

The interval and the test are two views of the same sampling calculation.

Apply the interval to the wage sample

$$22.64 \pm 1.96(1.28) = [20.13, 25.15] \text{ dollars per hour}$$

Compatibility

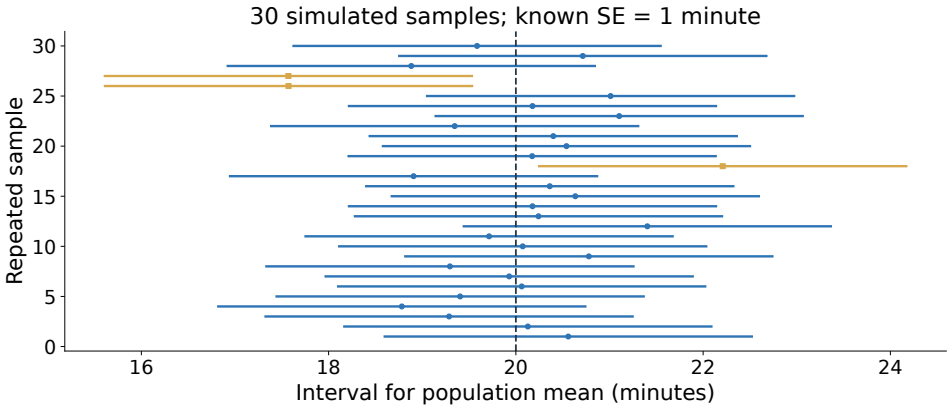
The value 20 lies just outside the interval, matching the earlier 5% rejection.

Precision

The margin of error is $1.96(1.28) = 2.51$ dollars.

The interval concerns the population mean; it is not a range containing 95% of individual wages.

The confidence level belongs to the procedure



Stata: verify the calculation and see the exact reference distribution

With summary statistics, the command order is sample size, sample mean, sample SD, and null mean.

```
ttesti 200 22.64 18.14 20
```

Match the hand calculation

Locate the estimated SE, t statistic, p-value, and 95% interval.

Explain the small difference

Stata uses a Student t_{199} reference; the slides use the large-sample normal approximation.

Extension: compare two population means with the same logic

For independent random samples from populations A and B, the target is $\Delta = \mu_A - \mu_B$ and the estimator is $\hat{\Delta} = \bar{Y}_A - \bar{Y}_B$.

$$\widehat{SE}(\hat{\Delta}) = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$t = \frac{\hat{\Delta} - \Delta_0}{\widehat{SE}(\hat{\Delta})}, \qquad CI_{95\%} = \hat{\Delta} \pm 1.96\widehat{SE}(\hat{\Delta})$$

Only the target and standard error change; the test-and-interval logic does not.

Worked extension: compare two independent groups

Group A: $\bar{Y}_A = 25$, $s_A = 10$, $n_A = 100$. Group B: $\bar{Y}_B = 22$, $s_B = 8$, $n_B = 100$.

$$\widehat{\Delta} = 3, \qquad \widehat{SE}(\widehat{\Delta}) = \sqrt{10^2/100 + 8^2/100} = 1.28$$

$$CI_{95\%} \approx 3 \pm 1.96(1.28) = [0.49, 5.51]$$

The data estimate a positive population mean difference; the comparison alone does not explain why the groups differ.

Research design determines whether a difference is causal

If workers were randomly offered training, compare the offer and no-offer groups:

$$\widehat{\tau} = \bar{Y}_{\text{offer}} - \bar{Y}_{\text{no offer}}$$

Random assignment

Makes treatment offers comparable in expectation and supports a causal interpretation of the offer contrast.

Random sampling

Supports generalizing the estimated contrast to a target population.

Sampling uncertainty quantifies noise; research design determines what the comparison means.

One inferential spine connects the entire lecture

Step	Question
Target	Which population quantity?
Estimate	What did the sample produce?
Standard error	How much would the estimate vary?
Test	Is a benchmark unusually far away?
Interval	Which benchmarks remain compatible?
Design	What substantive claim is justified?

A larger estimated difference can be less convincing when its standard error is larger.