ECON 0150 | Economic Data Analysis

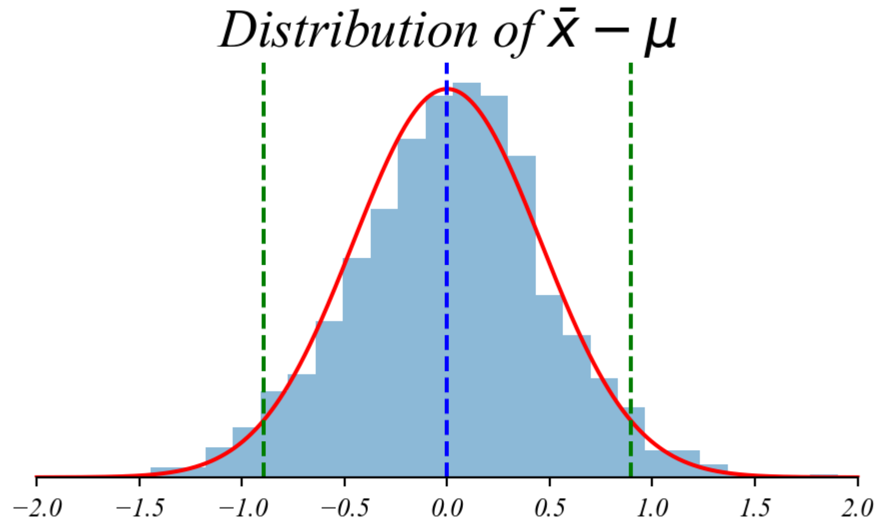
The economist's data analysis pipeline.

Part 3.4 | Testing Hypotheses

Confidence Intervals Recap

We used the distribution of sample means to systematize the probability of "closeness" of $\bar{x}$ and μ.

- *The difference between $\bar{x}$ and μ follows a t distribution with $SE = \frac{s}{\sqrt{n}}$*
- *95% of samples will have $\bar{x}$ no further than 1.96 standard errors from μ*



> in the wait time example, we asked "where is the true mean wait time?"

> but what if we want to test a specific claim about the mean?

Flipping The Question

What if we want to test a specific claim about the mean?

> *“my boss claims the mean wait time is 10 minutes”*

> *is our data consistent with that specific claim?*

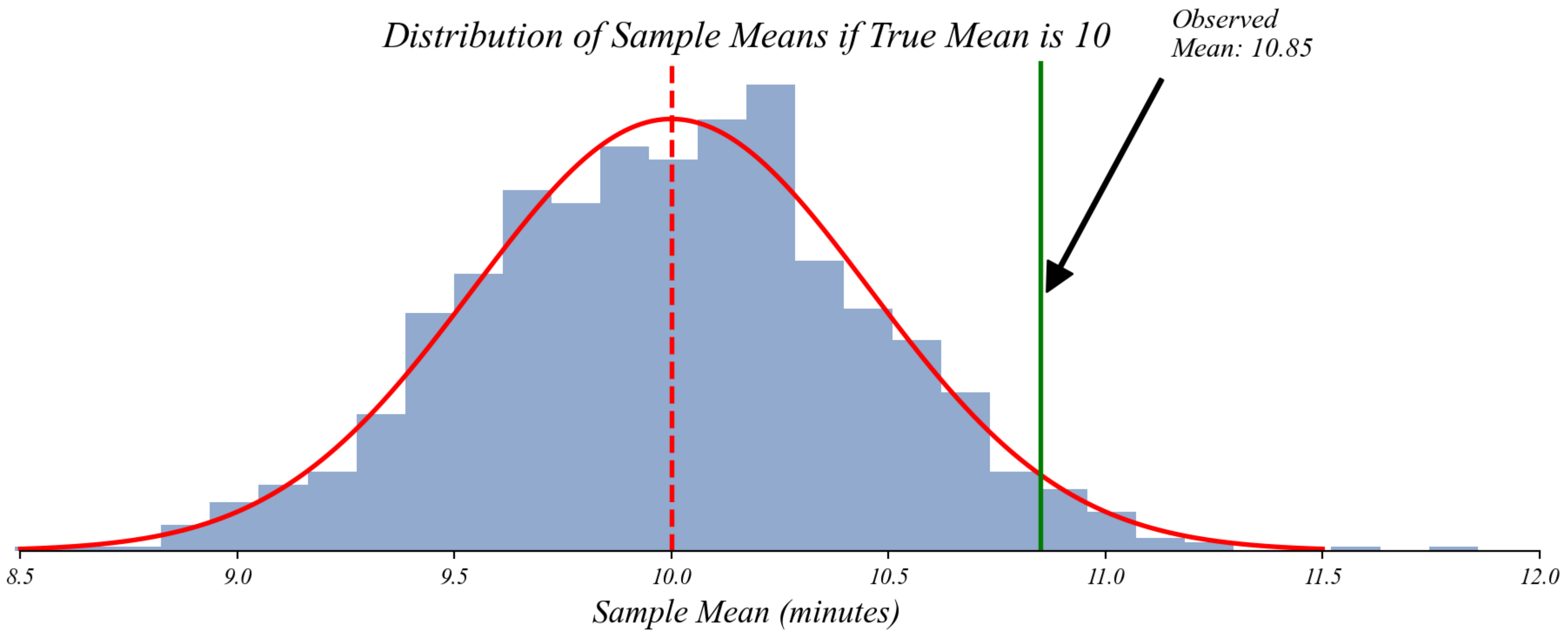
> *same math as last time, but a different question…*

> *instead of finding where some μ might be, we’re testing a specific value of μ*

Example: Wait Times

If $\bar{x} = 10.85$, is that consistent with $\mu_0 = 10$?

> *let's simulate data where $\mu = 10$ and see what sample means we'd get*



> *how "surprising" would our observed $\bar{x}$ be if μ actually was 10?*

> *notice we've centered the distribution on our hypothesis: μ_0*

Example: Wait Times

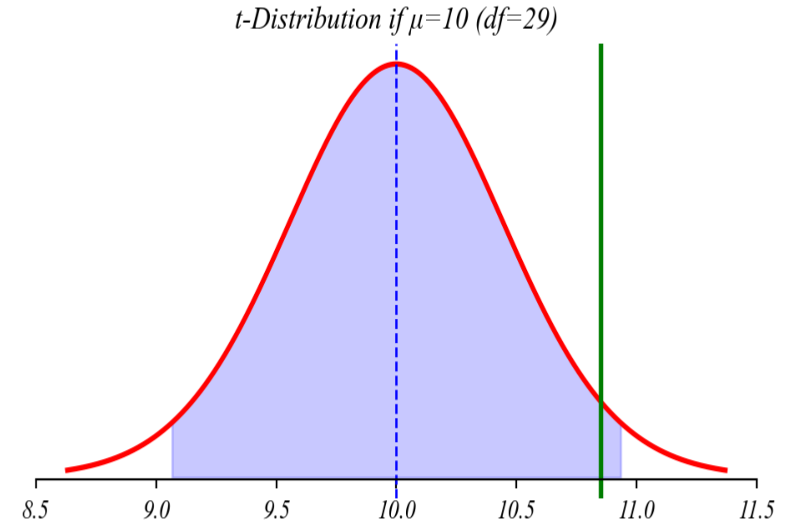
The math to answer this question is identical to confidence intervals.

If sample standard deviation is $s = 2.5$:

$$SE = \frac{s}{\sqrt{n}}$$

$$SE = \frac{2.5}{\sqrt{30}}$$

$$SE = 0.456$$




```
s = 2.5
n = 30
se = s / np.sqrt(30)
```

Example: Wait Times

The math to answer this question is identical to confidence intervals.

If sample standard deviation is $s = 2.5$:

$$SE = 0.456$$

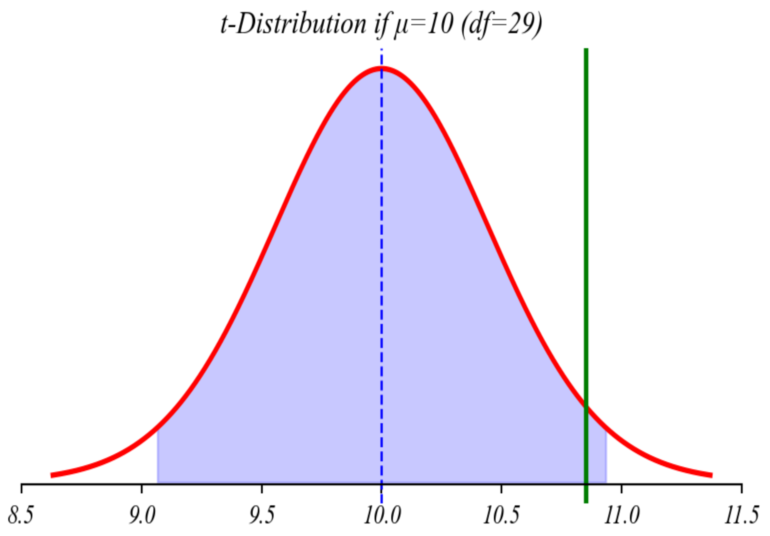
If true mean is $\mu_0 = 10$:

$$\bar{x} \sim t_{29}(10, 0.456)$$

So the critical value for 95%:

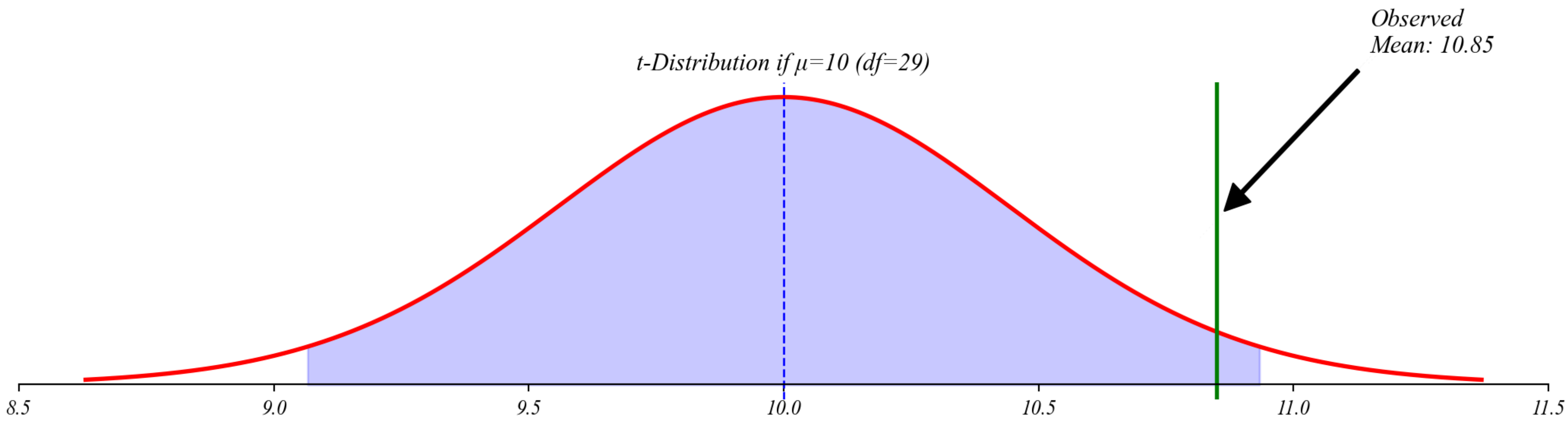
$$t_{crit} = 2.045$$

```
stats.t.interval(0.95, df=30)
```



Example: Wait Times

The math to answer this question is identical to confidence intervals.



A 95% confidence interval around μ_0 would be: $[9.07, 10.93]$

> *our observed mean ($\bar{x} = 10.85$) is within this interval — not surprising if μ=10*

> *but if we observed $\bar{x} = 11.5$, that would be outside the interval — surprising!*

The Null Hypothesis

We formalize this approach by setting up a "null hypothesis"

Null Hypothesis (H_0): *The specific value or claim we're testing*

- $H_0 : \mu = 10$ *(wait time is 10 minutes)*

Alternative Hypothesis (H_1 or H_a): *What we accept if we reject the null*

- $H_1 : \mu \neq 10$ *(wait time is not 10 minutes)*

Testing Approach:

- *Calculate how "surprising" our data would be if* H_0 *were true*
- *If sufficiently surprising, we reject* H_0

Quantifying Surprise: p-values

The p-value measures how compatible our data is with the null hypothesis.

p-value: *The probability of observing a test statistic at least as extreme as ours, if the null hypothesis were true*

For our example:

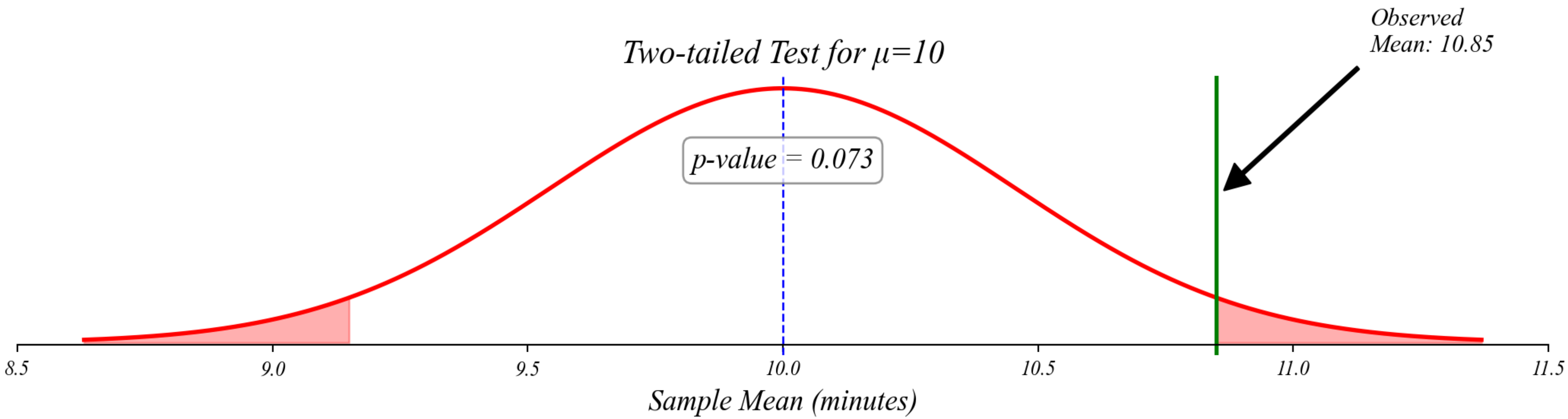
- *Null:* $\mu = 10$
- *Observed:* $\bar{x} = 10.85$

> How likely is it to get $\bar{x}$ this far or farther from 10, if the true mean is 10?

Quantifying Surprise: p-values

Example cont.: What is the probability of an error as large as the observed mean?

> *how likely is it to get $\bar{x}$ this far or farther from 10, if the true mean is 10?*




```
stats.t.cdf((mu_0-xbar)/se, df=n-1)) * 2
```

> *interpretation: if μ=10, we'd see $\bar{x}$ this far from 10 about 7.2% of the time*

> *often, we reject H_0 if p-value < 0.05 (5%)*

> *here, p-value > 0.05, so we don't reject the claim that μ=10*

Test Statistic: The t-statistic

We can standardize our result for easier interpretation

The **t-statistic** measures how many standard errors our sample mean is from the null value:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Where:

- $\bar{x}$ *is our sample mean (10.85)*
- μ_0 *is our null value (10)*
- s *is our sample standard deviation (2.5)*
- n *is our sample size (30)*

Test Statistic: The t-statistic

We can standardize our result for easier interpretation

The **t-statistic** measures how many standard errors our sample mean is from the null value:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{10.85 - 10}{2.5/\sqrt{30}} = \frac{0.85}{0.456} = 1.86$$

Where:

- *$\bar{x}$ is our sample mean (10.85)*
- *μ_0 is our null value (10)*
- *s is our sample standard deviation (2.5)*
- *n is our sample size (30)*

The t-test

This example has become a formal hypothesis test.

One-sample t-test:

- $H_0 : \mu = 10$
- $H_1 : \mu \neq 10$
- *Test statistic:* $t = 1.86$
- *Degrees of freedom: 29*
- *p-value: 0.072*

```python
# Imports
import numpy as np
from scipy import stats
```

```python
# Sample Data
sample_mu = 10.85
pop_mu = 10    # null hypothesis
std_dev = 2.5
n = 30
```

Decision rule:

- *If p-value < 0.05, reject* H_0
- *Otherwise, fail to reject* H_0

```python
# Calculate t-statistic
t_stat = (sample_mu - pop_mu) / (std_dev / np.sqrt(n))
```

```python
# Calculate p-value
p_value = 2 * (1 - stats.t.cdf(abs(t_stat), df=n-1))
```

> *t-tests are extremely common, especially in regression (coming soon!)*

Statistical vs. Practical Significance

A caution about hypothesis testing

Statistical significance:

- *Formal rejection of the null hypothesis ($p < 0.05$)*
- *Only tells us if the effect is unlikely due to chance*

Practical significance:

- *Whether the effect size matters in the real world*
- *A statistically significant result can still be tiny*

> *with large samples, even tiny differences can be statistically significant*

> *always consider the magnitude of the effect, not just the p-value*

Common Misinterpretations

What a p-value is NOT

❌ **Not:** The probability that H_0 is true

- *The p-value doesn't tell us if the null hypothesis is correct. It assumes the null is true and then calculates how surprising our result would be under that assumption.*
- *Example: A p-value of 0.04 doesn't mean there's a 4% chance the null hypothesis is true.*

Common Misinterpretations

What a p-value is NOT

❌ **Not:** The probability that the results occurred by chance

- *All results reflect some combination of real effects and random variation. The p-value doesn't separate these components.*
- *Example: A p-value of 0.04 doesn't mean there's a 4% chance our results are due to chance and 96% chance they're real.*

Common Misinterpretations

What a p-value is NOT

❌ **Not:** The probability that H_1 is true

- *The p-value doesn't directly address the alternative hypothesis or its likelihood.*
- *Example: A p-value of 0.04 doesn't mean there's a 96% chance the alternative hypothesis is true.*

Common Misinterpretations

What a p-value is NOT

✅ **Correct:** The probability of observing a test statistic at least as extreme as ours, if H_0 were true

- *It measures the compatibility between our data and the null hypothesis.*
- *Example: A p-value of 0.04 means: "If the null hypothesis were true, we'd see results this extreme or more extreme only about 4% of the time."*

> think of it like this: The p-value answers "How surprising is this data if the null hypothesis is true?" not "Is the null hypothesis true?"

Looking Forward

The t-test framework extends to many scenarios

Next time:

- *Comparing means between two groups*

Coming soon:

- *This same framework underlies regression analysis*
- *Regression coefficients are tested using t-tests*
- *ANOVA uses the same fundamental approach*

> *the hypothesis testing framework is foundational for modern science*