

Hypothesis Testing

Grinnell College

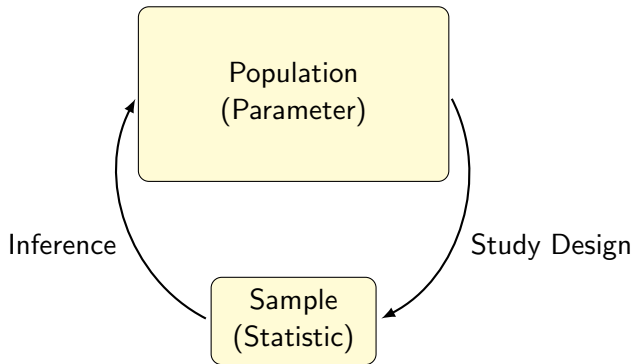
October 16, 2026

Warm-up

1. Why do we use a t distribution instead of the standard normal when constructing a confidence interval for a mean? What determines the degrees of freedom?
2. For the same sample and confidence level, would a t -based interval be wider or narrower than one using a normal critical value? Why?
3. We observe $\bar{x} = 35$ from a sample of $n = 15$, with an estimated standard error of 1. Someone claims $\mu = 36.5$. What is the t statistic, and what do its sign and magnitude tell us? Is 36.5 a plausible population mean using an 80% confidence interval?

Part II: Hypothesis Testing (not today)

The Statistical Framework



Hypothesis Testing

Hypothesis testing involves:

1. Formulating an *unambiguous* statement about a population parameter, called our **null hypothesis**
2. Collecting observational or experimental data
3. Determining if the data collected is consistent with our hypothesis
4. Either *rejecting* or *failing to reject* our hypothesis based on the *strength* of the evidence

Null Hypothesis

Our hypothesis about a parameter prior to seeing any data is called our **null hypothesis**, typically expressed in the form

$$H_0 : \mu = \mu_0$$

where μ_0 (“mew-naught”) represents a specific quantity.

This addresses the fact that, in actuality, *we do not know the value of μ* . Instead, our goal will be to posit a value and determine whether or not this hypothesis is consistent with the data observed.

Test statistics

We relate the data that we have observed (i.e., $\bar{x}$, s) with our null hypothesis with the use of **test statistics**

For example, if $H_0 : \mu = \mu_0$ is correct, then

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

will be centered at zero and will follow a t -distribution. If H_0 is *not* correct, the distribution of t statistics will be centered somewhere else instead

The **null distribution** describes the distribution that our test statistics will follow if the null hypothesis is true

Hypothesis Testing and Confidence

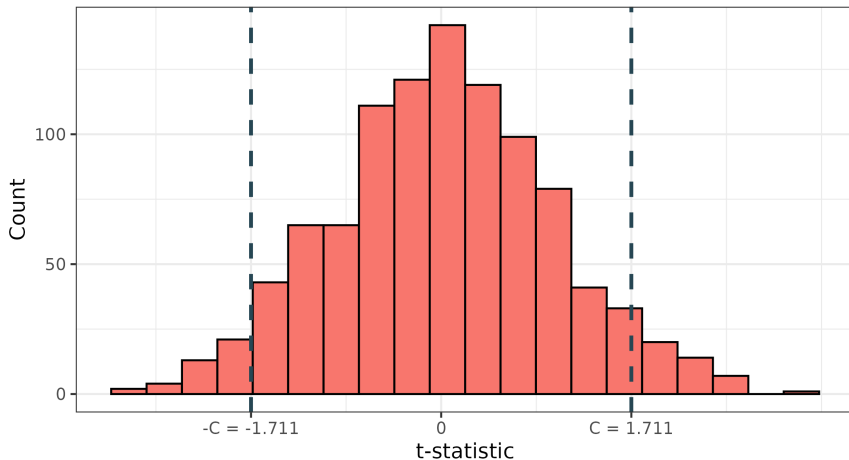
How do we go about using a null distribution?

Assuming the null hypothesis is true, we know what distribution our statistic should follow and, accordingly, the *critical values* associated with the bounds of our distribution

Test statistics that fall sufficiently outside of the bounds of where we expect our data to fall may be considered evidence *against* the null hypothesis

90% of t-statistics fall between -1.711 and 1.711

Under the null hypothesis; $df = 24$



The General Idea

The general idea is this:

1. We specify some null hypothesis $H_0 : \mu = \mu_0$
2. *Assuming that the null hypothesis is true*, the test statistic

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

will follow a t -distribution, centered at 0

3. From our data, we will compute a test statistic using both $\bar{x}$ and μ_0
4. We will then check our test statistic t against critical values C
5. If $C < |t|$, our test statistically is sufficiently far from what we expect and we reject our null hypothesis

This is what it looks like when we subtract true mean and this is what it looks like when we subtract WRONGO MEAN. The more wrong, the more different it is. This is a pregnant idea

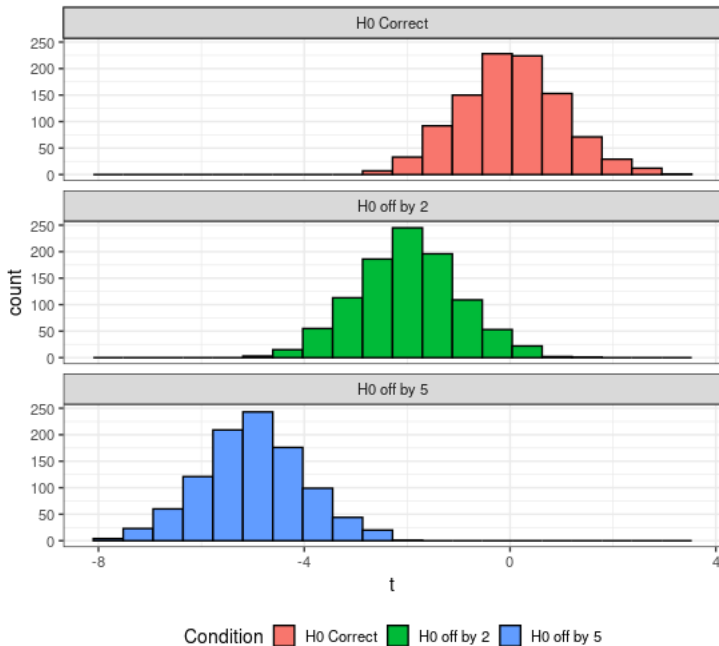
Null vs Sampling Distribution

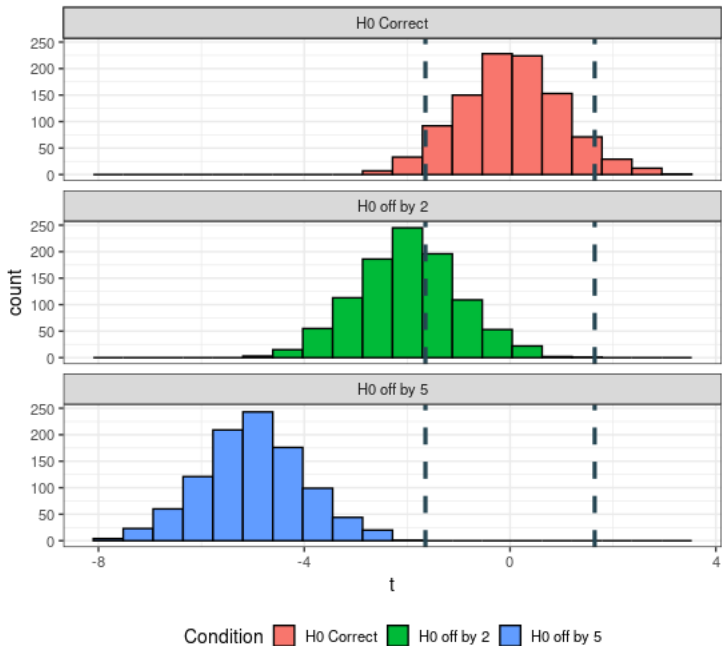
Suppose there are true population parameters $\mu = 10$ and $\sigma = 5$. Further, suppose we have three separate hypotheses for H_0 :

1. $H_0 : \mu = 10$
2. $H_0 : \mu = 12$
3. $H_0 : \mu = 15$

If we were to collect many samples of $\bar{X}$ with $n = 25$ and construct t statistics from each one, only one of the resulting distributions will be centered at 0. For both $H_0 : \mu = 12$ and $H_0 : \mu = 15$, will we be subtracting too much

This will impact how likely we are to reject our null hypothesis





Penguin Example

Suppose we are studying the length (mm) of male penguin flippers, beginning with two competing hypotheses:

1. $H_{0_1} : \mu = 203$
2. $H_{0_2} : \mu = 215$

We collect sample of size $n = 20$, finding sample statistics of:

1. $\bar{x} = 209.7$
2. $s/\sqrt{n} = 3.342$

How would we test these hypotheses at confidence levels of 80%, 90%, and 95%?

Penguin Example

We essentially have two ways that are mathematically equivalent:

1. Check if μ_0 is in our confidence interval $\bar{X} \pm C \times \left(\frac{\hat{\sigma}}{\sqrt{n}} \right)$
2. Check if our t statistic falls between critical values $|t| < C$

With $n = 20$ we have $df = 19$, and from our critical value sheet we find critical values for 80, 90, and 95 to be

$$C_{80} = 1.33, \quad C_{90} = 1.73, \quad \text{and} \quad C_{95} = 2.09$$

Confidence Interval Method (clunky)

We have $\bar{x} = 209.7$ and $\hat{\sigma}/\sqrt{n} = 3.34$, with critical values

$$C_{80} = 1.33, \quad C_{90} = 1.73, \quad \text{and} \quad C_{95} = 2.09$$

Using

$$\bar{X} \pm C \times \left(\frac{s}{\sqrt{n}} \right)$$

we find intervals:

- ▶ 80%: (205.26, 214.14)
- ▶ 90%: (203.92, 215.48)
- ▶ 95%: (202.72, 216.68)

We can simply check inclusion of $H_{0_1} : \mu = 203$ and $H_{0_2} : \mu = 215$

Critical value method (elegant)

We have $\bar{x} = 209.7$ and $\hat{\sigma}/\sqrt{n} = 3.34$, with critical values

$$C_{80} = 1.33, \quad C_{90} = 1.73, \quad \text{and} \quad C_{95} = 2.09.$$

We can then construct t -statistics from each of our hypotheses
 $H_{0_1} : \mu = 203$ and $H_{0_2} : \mu = 215$

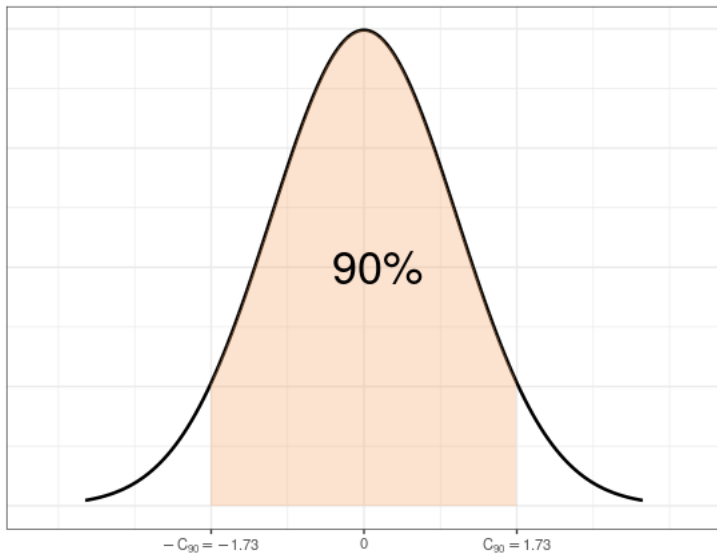
$$t_1 = \frac{209.7 - 203}{3.34} = 2.006, \quad t_2 = \frac{209.7 - 215}{3.34} = -1.587$$

By checking if $|t| < C$ we will come to the same conclusions:

1. Reject both with 80% confidence
2. Reject H_{0_1} with 90% confidence but not H_{0_2}
3. Reject neither with 95% confidence

$$t_1 = 2.006 \text{ and } t_2 = -1.587$$

Null Distribution



Review Steps

- ▶ Formulate a null hypothesis H_0
- ▶ Use this and your sample data to construct a test statistic (i.e., t -statistic)
- ▶ If the null hypothesis is true, the t -statistic will follow a t -distribution centered at zero. This is our *null distribution*
- ▶ Find the critical values of the null distribution, i.e., *if the null hypothesis were true*, what are the bounds of where we would expect to see our data?
- ▶ Compare your statistic to the critical values. If our statistic exceeds those bounds, we reject H_0