Data Analysis 4 Hint:

On part 2, question a, you are asked to find the 10th and 90th percentiles of a Standard Normal Distribution. The Standard Normal is a special case of the normal distribution with a fixed mean and standard deviation. We discussed this distribution initially on Tuesday of week 4 in the Normal Distribution notes.

You will need your answer from part a to answer part b.

DA 4 is due tonight!

Week 6

Hypothesis Testing

ST 314
Introduction to Statistics for Engineers

Oregon State University

Estimation vs. Hypothesis Testing

Want to know more about a population characteristic. → Goal is to estimate the population characteristic (unknown parameter). → **Estimation** Point estimates and confidence intervals

Have a claim or guess about a population characteristic. → Goal is to use sampled data to test the validity of the claim. → **Hypothesis Testing** Hypotheses, test statistics, and p-values

Examples of when to use a hypothesis test:

- Test the average train ride time against the advertised time
- Test the proportion of defective parts in a manufacturing process against a specified standard
- Compare whether the average tensile strength of rubber seals is different between machines

Hypothesis Test Example

A high-speed train between Florence and Rome, Italy is advertised to take 94 minutes.

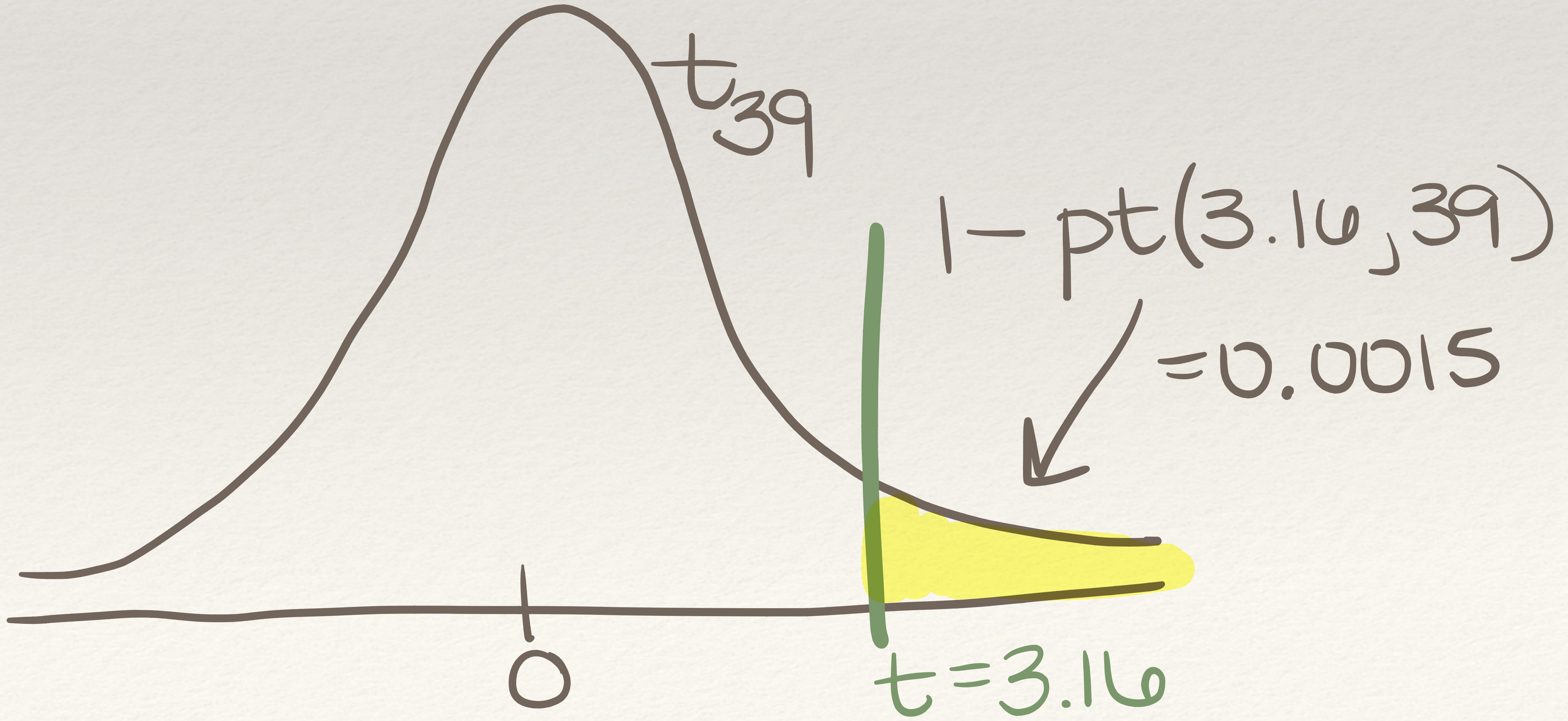
A frequent rider is consistently late and is convinced the average ride time is longer. She decided to record the time of 40 random rides. Her sampled average is 99 minutes. The sample has a standard deviation of 10 minutes.

$\bar{X} = 99 \quad n = 40 \quad df = 39$

$s = 10 \quad \mu = 94$

If the advertised time of 94 minutes is the actual average, what is the chance the mean from a sample of 40 rides would exceed 99 minutes?

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{99 - 94}{10/\sqrt{40}} = 3.16 \sim t_{39}$$



Steps in Performing a Hypothesis Test

1. State the question of interest
2. Identify the parameter (μ or p)
3. State the null and alternative hypotheses
4. Check the conditions for the test
5. Determine the null distribution
6. Calculate the test statistic & p-value
7. Make a conclusion

The Hypotheses

The Null Hypothesis

- Establishes a claim
- Assume the null hypothesis is true when performing the test.
- Always statement of equality =
- Denote H_0

The Alternative Hypothesis

- reflects the question of interest
- contradict the null
- either one- or two-sided
- Denoted H_A (or H_1)

$H_0: \mu = 94$

$H_A: \mu > 94$

Always in terms of the pop. parameter!

Hypothesis Test Example

A high-speed train between Florence and Rome, Italy is advertised to take 94 minutes.

A frequent rider is consistently late and is convinced the average ride time is longer. She decided to record the time of 40 random rides. Her sampled average is 99 minutes. The sample has a standard deviation of 10 minutes.

Determine the null and alternative hypotheses for this test based on the train site's claim that rides take 94 minutes.

$H_0: \mu = 94$

$H_A: \mu > 94$

where μ = average ride time

One- and Two-sided Alternatives

μ_0 = claimed mean

p_0 = claimed proportion

Lower one-sided alternative

Question of interest: is the parameter **less than** the claimed value?

$H_A: \mu < \mu_0$ | $H_A: p < p_0$

Upper one-sided alternative

Question of interest: is the parameter **greater than** the claimed value?

$H_A: \mu > \mu_0$ | $H_A: p > p_0$

Two-sided alternative

Question of interest: does the parameter **differ** from the claimed value?

$H_A: \mu \neq \mu_0$ | $H_A: p \neq p_0$

One- or Two-sided Practice

Consider the following alternative hypothesis examples. Which are one-sided? Which are two-sided?

1. The average time to finish a maze is not equal to 16 minutes. two-sided
2. The average running pace is less than 7 mph. lower one-sided
3. The average IQ score of OSU students is greater than 119. upper one-sided
4. The average exam score is different than 75%. two-sided
5. The average lunch break is less than 30 minutes. lower one-sided

Test Statistics ~~& p-values~~

Similar to "standardizing", a **test statistic** compares a sample statistic to a hypothesized value while accounting for the variability of the statistic.

The general form of a test statistic is:

$$\frac{\text{estimate} - \text{claim}}{\text{standard error}}$$

The Null Distribution

- $H_0 : \mu = \mu_0$

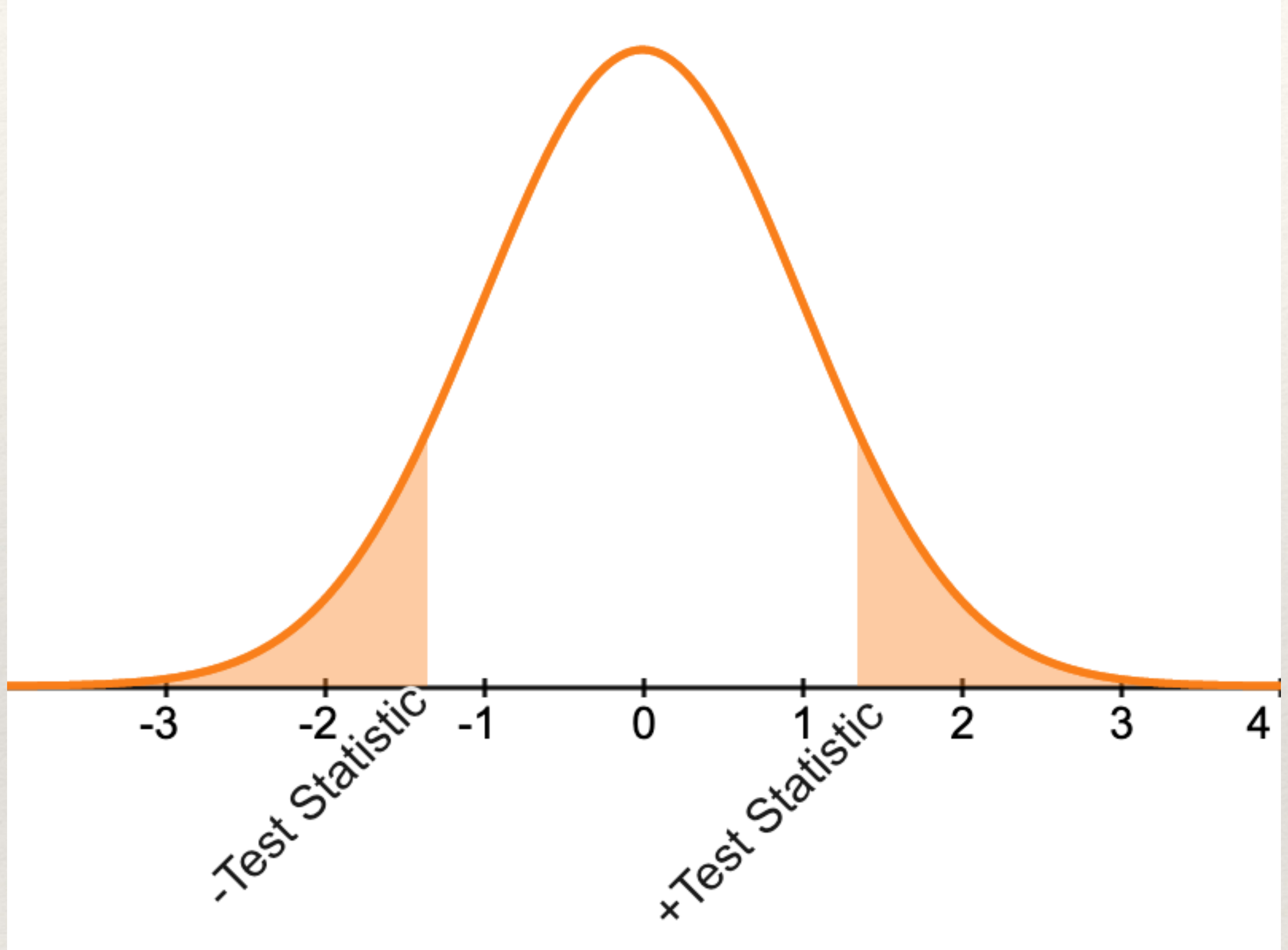
$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

- $H_0 : p = p_0$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$$

p-values

Two-sided

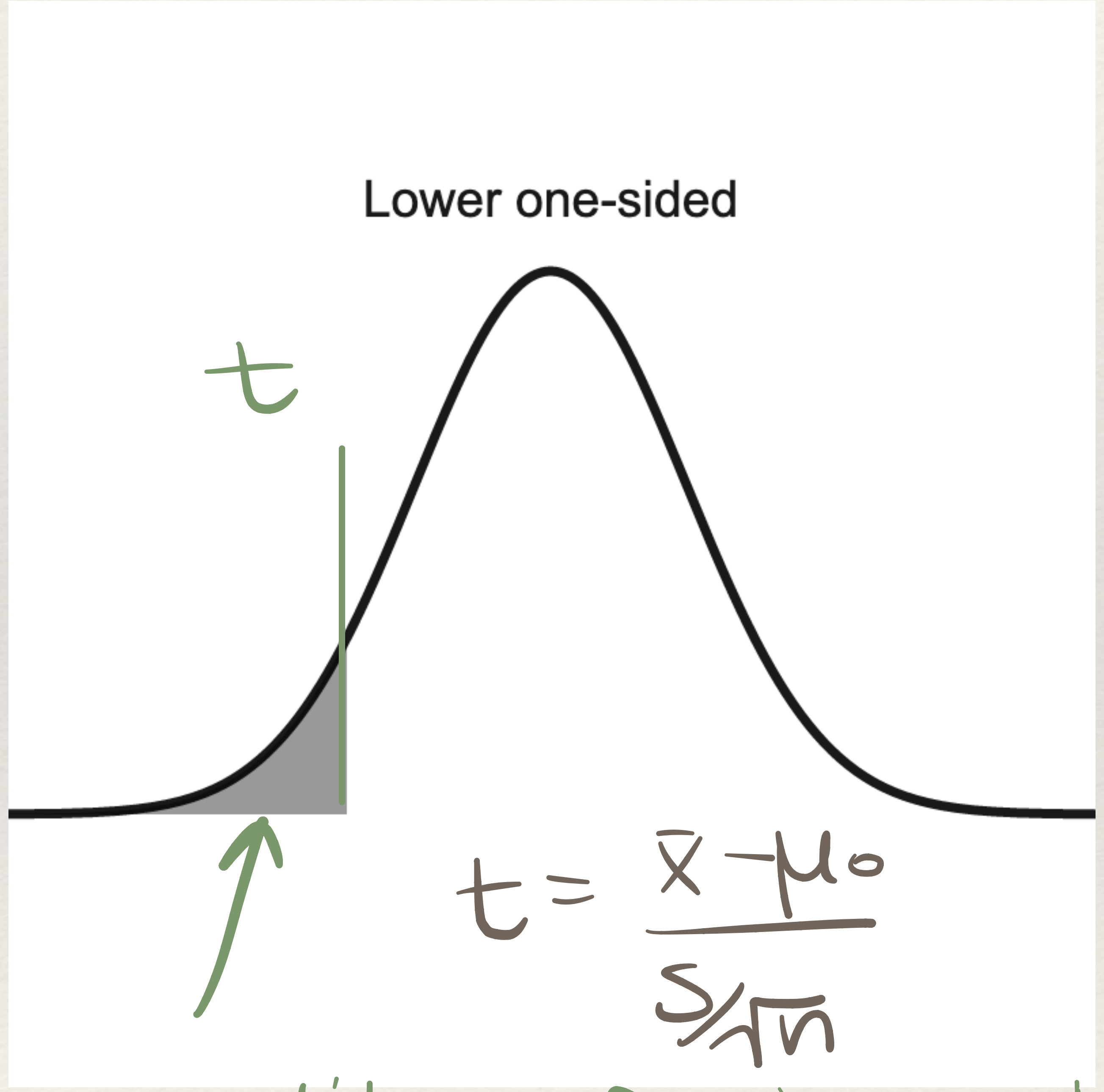


The p-value is a probability. It is the area under the tail(s) of the null distribution curve.

Assuming the hypothesized parameter value is true, the probability the sample statistic would take on a value as or more extreme than the one we actually observed is the p-value.

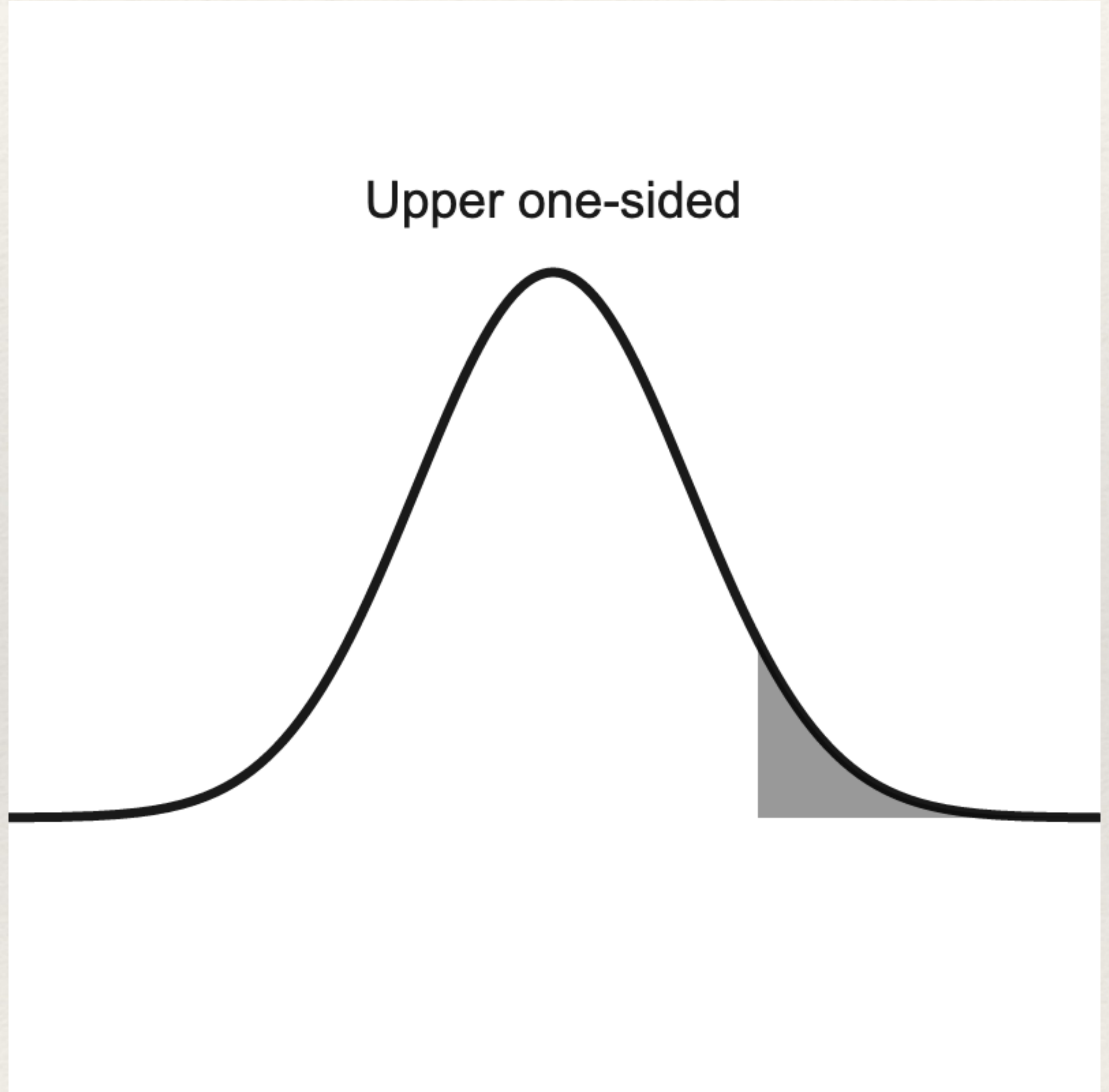
One- and Two-sided p-values

$H_A: \mu < \mu_0$

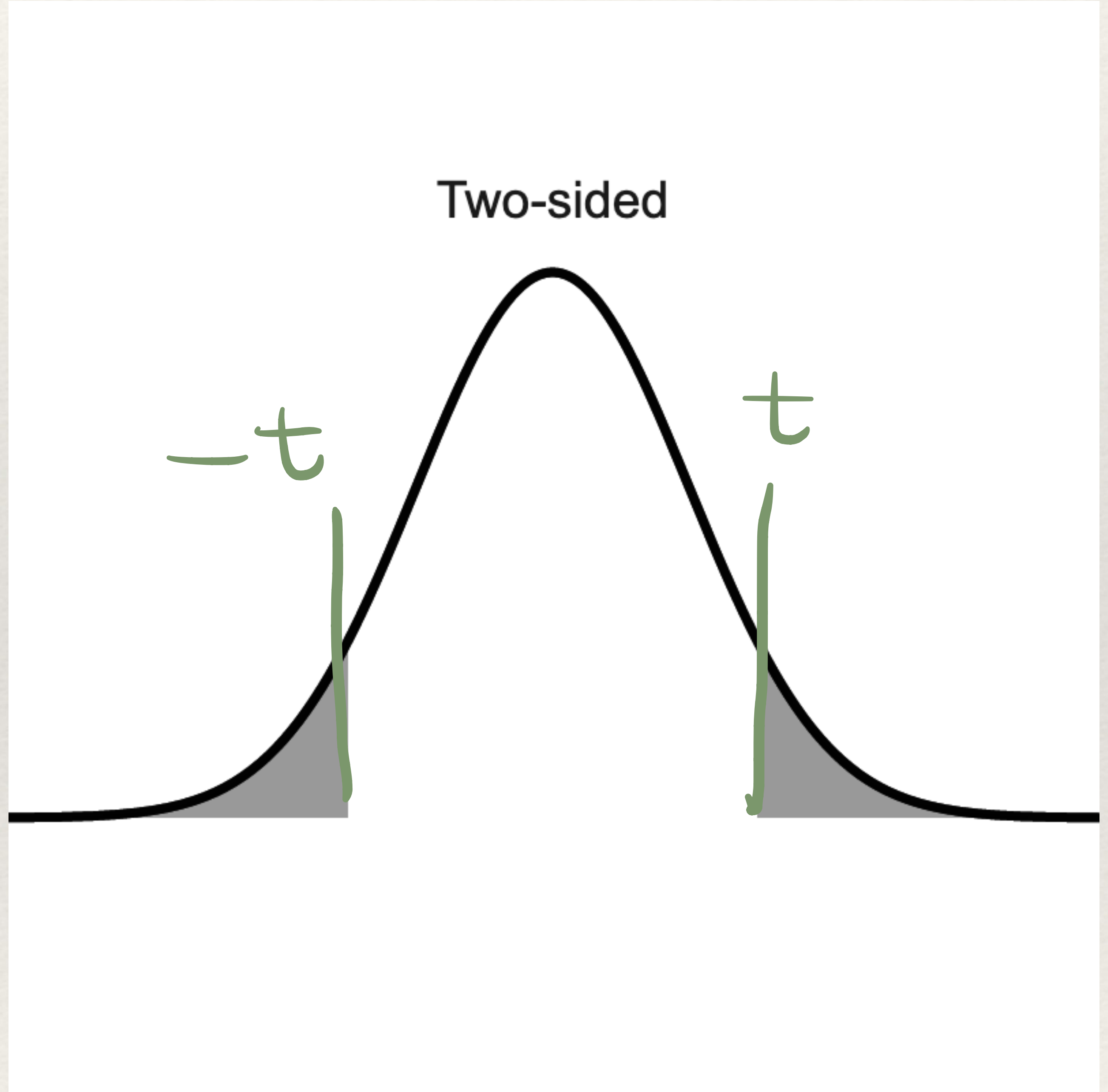


probability of observing a sample mean less than or equal to $\bar{x}$ when the true mean is μ_0

$H_A: \mu > \mu_0$



$H_A: \mu \neq \mu_0$



t-Test for a Mean

- When to use: test average value, μ, for single population
- Conditions required for inference:
  - representative sample (random)
  - sufficiently large sample (Central Limit Theorem Notes)
- Null & Alternative hypotheses: $H_0: \mu = \mu_0$

$H_A: \mu < \mu_0$ OR $H_A: \mu > \mu_0$ OR $H_A: \mu \neq \mu_0$

- Test statistic: $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$

t-Test for a Mean Example

A high-speed train between Florence and Rome, Italy is advertised to take 94 minutes.

A frequent rider is consistently late and is convinced the average ride time is longer. She decided to record the time of 40 random rides. Her sampled average is 99 minutes. The sample has a standard deviation of 10 minutes.

$n = 40$ ✓

We want to perform a hypothesis test to test whether the average ride time is greater than 94 minutes. To do so, first check the conditions. If they are met, proceed to calculating the test statistic.

$\bar{x} = 99 \quad n = 40 \quad s = 10 \quad \mu_0 = 94$

$$t = \frac{99 - 94}{10/\sqrt{40}} = 3.16$$

t-Test for a Mean Example

$H_0: \mu = 94$

$H_A: \mu > 94$

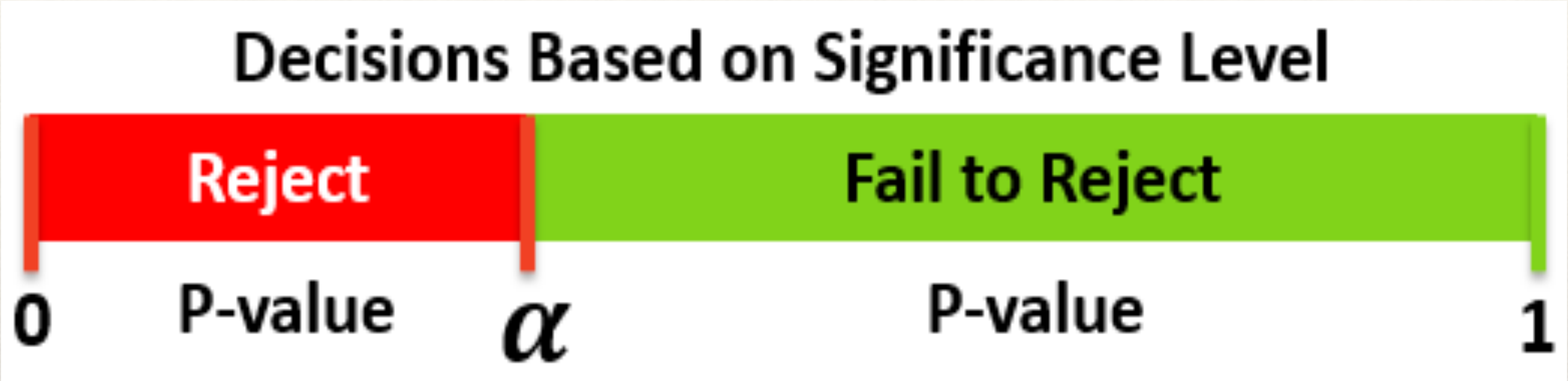
A high-speed train between Florence and Rome, Italy is advertised to take 94 minutes.

A frequent rider is consistently late and is convinced the average ride time is longer. She decided to record the time of 40 random rides. Her sampled average is 99 minutes. The sample has a standard deviation of 10 minutes.

Now that we've calculated the test statistic, we know how many standard deviations away from the hypothesized mean our sample mean is. We can use this value to calculate the probability we would have observed a sample mean as or more extreme than 99 minutes if we assume the true population mean to be 94 minutes (this calculation is the p-value). Calculate the p-value.

Decisions Based on p-values

The **significance level**, denoted by α, is a predetermined cut off to which we will reject the null hypothesis. Common values for α are: $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.1$

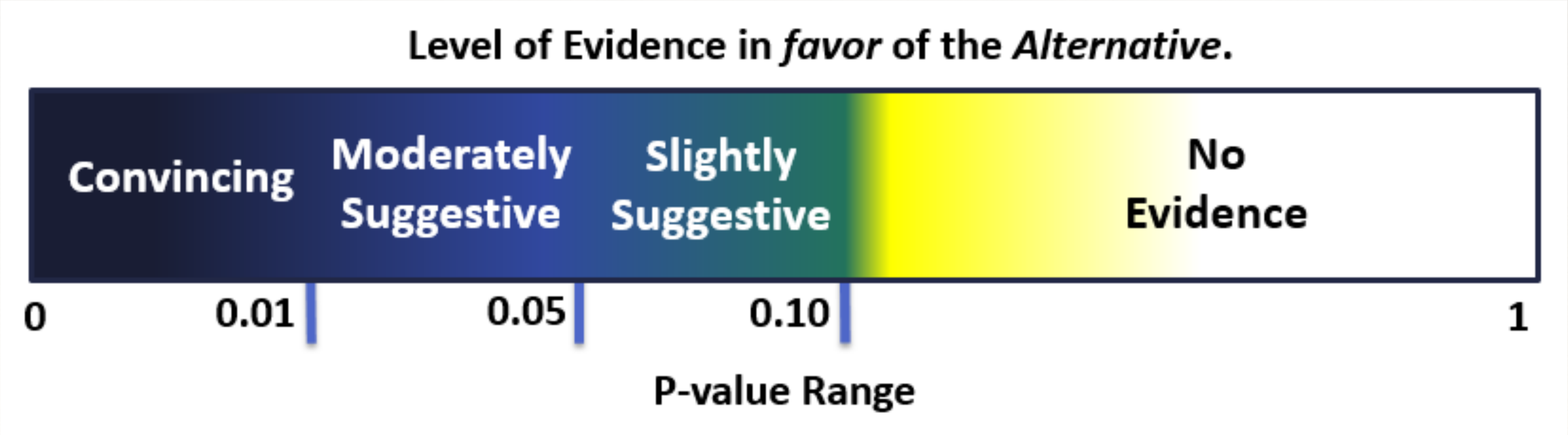


A smaller α requires a more extreme test statistic to reject the null. Decreasing α will reject less often, but will decrease the chance of error.

Never say the null is "false". Instead, we say "the evidence is very strong to reject the null"

Statements in terms of the alternative

- Using only terms like "reject" and "fail to reject" in conclusions may confuse novice readers.
- We'll provide a more complete conclusion by providing a statement of evidence in terms of the alternative hypothesis that reflects the question of interest.



Four-part Conclusion

A good conclusion must have context and should include the following four important pieces.

1. Statement for the strength of evidence for the alternative hypothesis.
2. Whether to reject or fail to reject the null hypothesis based on α.
3. The point estimate for the parameter of interest.
4. A $(1 - \alpha)100\,\%$ confidence interval estimate for the parameter of interest.

p-value = 0.0015 $\alpha = 0.05$

Hypothesis Test Conclusion Example

A high-speed train between Florence and Rome, Italy is advertised to take 94 minutes.

A frequent rider is consistently late and is convinced the average ride time is longer. She decided to record the time of 40 random rides. Her sampled average is 99 minutes. The sample has a standard deviation of 10 minutes.

Write a four-part conclusion for this hypothesis test.

There is convincing evidence to suggest that average ride time between Florence and Rome is greater than 94 minutes. We reject the null hypothesis at the 0.05 significance level that the average ride time is 94 minutes. We are 95% confident that the average ride time is between 95.8 minutes and 102.2 minutes with a point estimate of 99 minutes.

Errors in Hypothesis Testing

	H_0 True	H_0 False
Reject H_0	Type I Error	Correct Conclusion
Fail to reject H_0	Correct Conclusion	Type II Error

- When a test performs as it should, the value α is the probability of making a Type 1 error.
- When the consequences are high, reduce the chance of making a Type I error by choosing a smaller significance level.

Type I and Type II are related. Decreasing the chance of one error will increase the chance of the other.

z-Test for a Proportion

- When to use: want to test the population proportion from a single population
- Conditions required for inference:

sample is representative of population (random)
Independent observations

$np_0 \geq 10$ $n(1-p_0) \geq 10$ p_0 = hypothesized proportion

- Null & Alternative hypotheses: $H_0: p = p_0$

$H_A: p < p_0$ OR $H_A: p > p_0$ OR $H_A: p \neq p_0$

- Test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

z-Test Example

Pew Research asked a random sample of 1000 American adults whether they supported the increased usage of coal to produce energy. Set up hypotheses to evaluate whether a majority of American adults support or oppose the increased usage of coal.

z-Test Example

Confidence Intervals vs. Hypothesis Tests

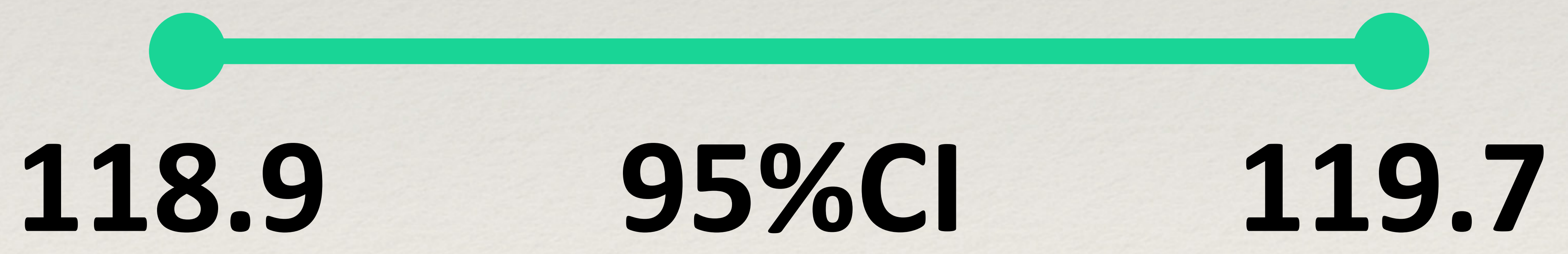
If a $(1-\alpha)100\,\%$ confidence interval, does not contain μ_0, the hypothesized value, then μ_0 it is not a plausible value for our ~~estimate~~ parameter. The two-sided hypothesis test will reject the null at the α significance level.

Example

Suppose a manufacturer of boards for construction would like to test the length of a batch of boards with the following null and alternative hypotheses:

$H_0 : \mu = 120$ inches
$H_A : \mu \neq 120$ inches

$$\mu_0 = 120$$



Reject at 5% Significance!

1. Do a majority of US adults believe raising the minimum wage will help the economy, or is there a majority who do not believe this? A Rasmussen Reports survey of 1,000 US adults found that 42% believe it will help the economy. Construct a 95% confidence interval and conduct a hypothesis test at the 0.05 significance level to help answer the research question.

p = proportion of US adults that believe increasing min wage will help economy

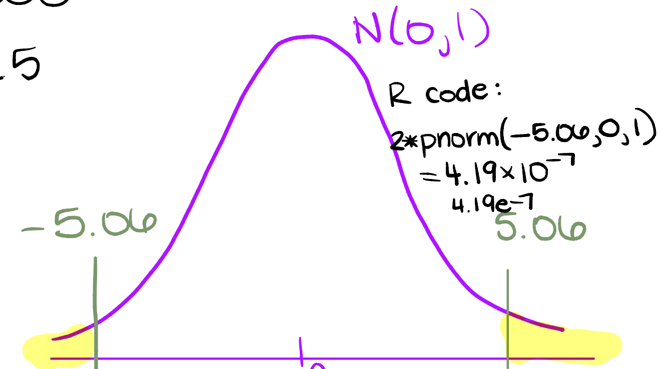
Confidence Interval: $n\hat{p} \geq 10$ $\quad 1000(0.42) \geq 10$ $\qquad n(1-\hat{p}) \geq 10$ $\quad 1000(1-0.42) \geq 10$

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.42 \pm 1.96\sqrt{\frac{0.42(1-0.42)}{1000}} = (0.389, 0.451)$$

Hypothesis test: $H_0: p = 0.5 \qquad H_A: p \neq 0.5$

$np_0 \geq 10$ $\quad 1000(0.5) > 10$ $\qquad n(1-p_0) \geq 10$ $\quad 1000(0.5) > 10$

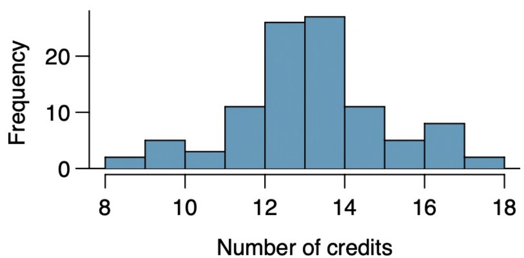
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.42 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{1000}}} = -5.06$$



P-value = 4.19 x10^-7

We have convincing evidence that there is a majority that either believes or does not believe that raising minimum will help the economy. We reject the null that the proportion of US adults that believe raising minimum wage will help the economy is equal to the proportion that do not believe this to be true. We are 95% confident that the true prop of adults that believe raising min wage will help the economy is between 0.389 and 0.451 with a point est of 0.42.

2. A college counselor is interested in estimating how many credits a student typically enrolls in each semester. The counselor decides to randomly sample 100 students by using the registrar's database of students. The histogram below shows the distribution of the number of credits taken by these students. Sample statistics for this distribution are also provided.



Min	8
Q1	13
Median	14
Mean	13.65
SD	1.91
Q3	15
Max	18

a. Construct a 90% confidence interval for the average number of credits taken per semester at this college. Interpret the interval

b. The college expects that students enroll in 13 credit hours per semester on average. Perform a hypothesis test to test if this college's average credit load is more than 13 per semester.

3. Georgianna claims that in a small city renowned for its music school, the average child takes less than 5 years of piano lessons. We have a random sample of 20 children from the city, with a mean of 4.6 years of piano lessons and a standard deviation of 2.2 years.

   a. Evaluate Georgianna's claim (or that the opposite might be true) using a hypothesis test.
   b. Construct a 95% confidence interval for the number of years students in this city take piano lessons, and interpret it in context of the data.
   c. Assess the conditions for inference required to rely on the results of the confidence interval and hypothesis test.