The 7-step SOP for Hypothesis Tests: A More Inclusive Way to Teach Statistics to Social Science Students

Lei Zhang

Department of Sociology

Jan. 9th, 2024 @ TLC

Outline

- **Challenges**
  - Diversified Background of Students
  - Quantitative Readiness
- **Goal: the balance between**
  - Being scientifically rigorous
  - Being practically realistic
- **A Possible Solution: the 7-step SOP for Hypothesis Tests**
- **Examples**
  - Example 1: Teaching T-test of Population Mean
  - Example 2: T-test of Population Mean in Exam

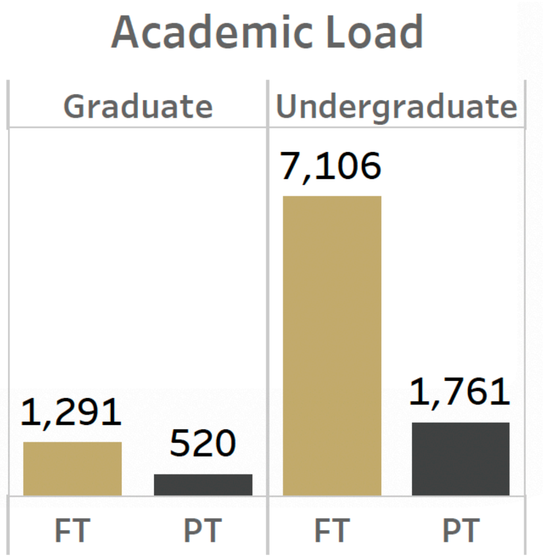
Challenges

- **Diversified Background**
  - UCCS facts of diversity, 2023
  - Before-class survey, SOC 3170/5020 Social Statistics, 2023 Fall
- **Quantitative Readiness**
  - Before-class survey, SOC 3170/5020 Social Statistics, 2023 Fall

UCCS facts of diversity, 2023

Challenge 1: Diversified Background

Diversified Background

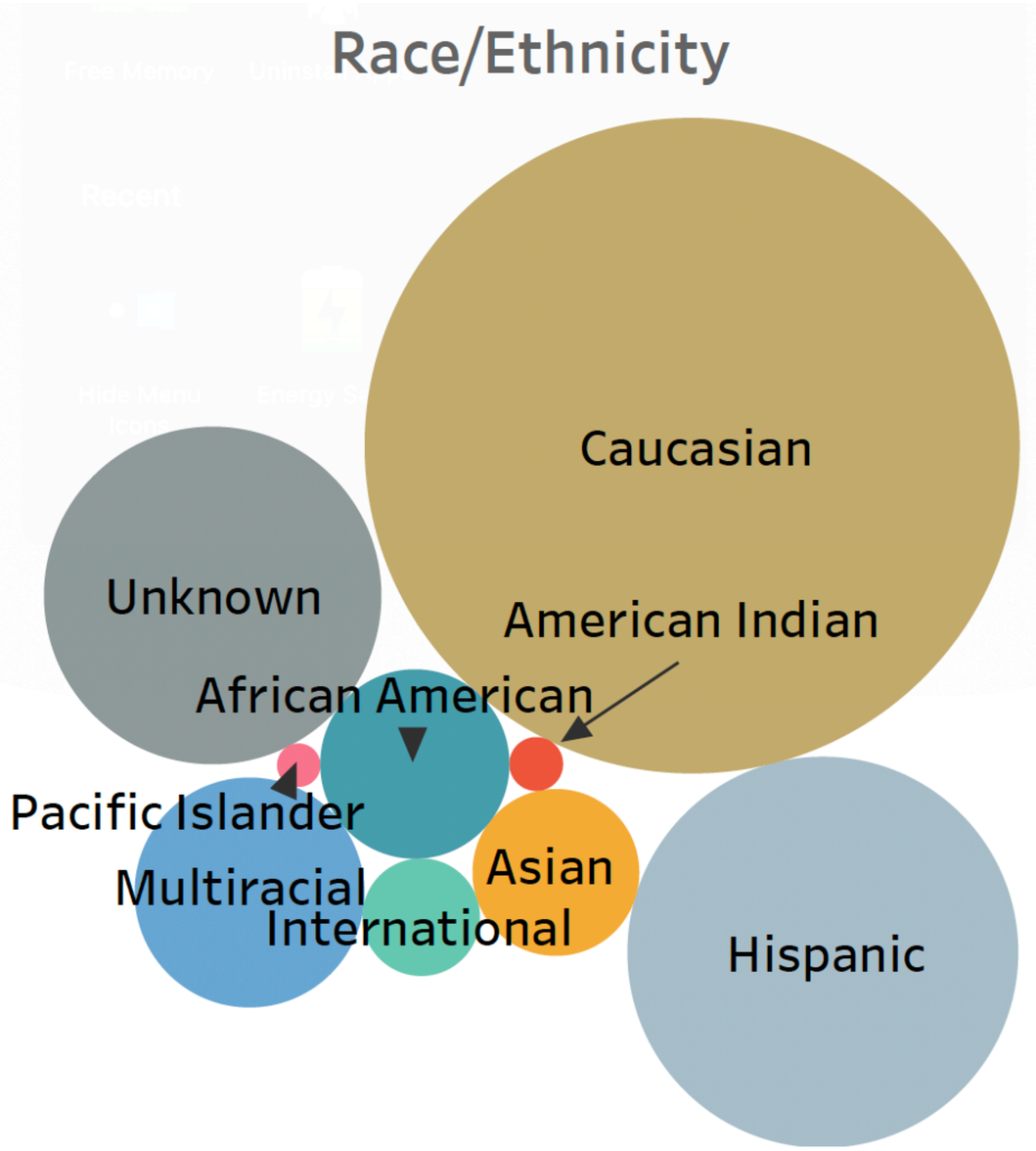


* FT = Full-time, PT = Part-time

* Unknown gender = 2.6%

Average Age

Graduate	34
Undergraduate	23

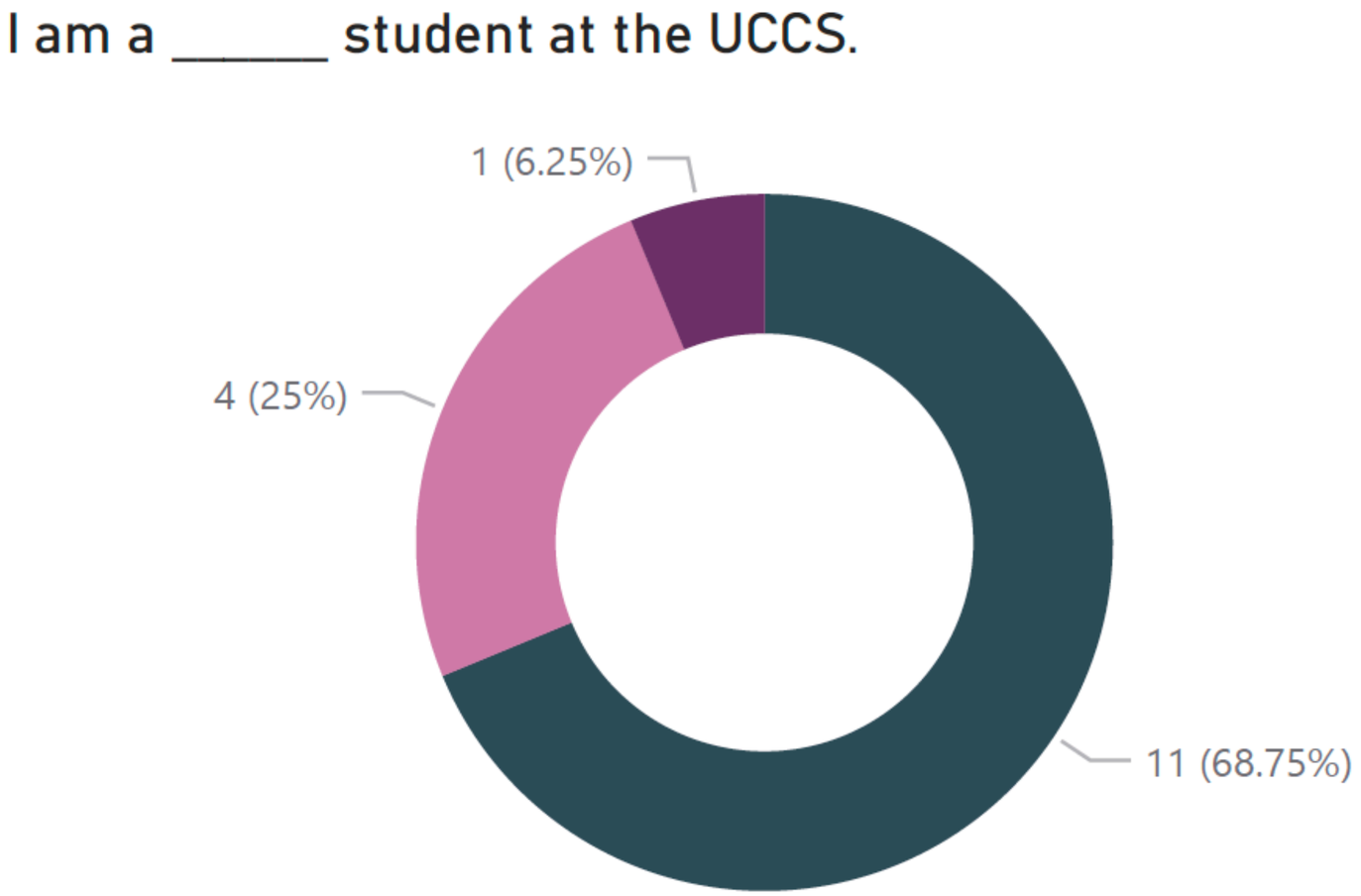


Source: https://ir.uccs.edu/quick-facts

Before-class survey, SOC 3170/5020 Social Statistics, 2023 Fall

Challenge 1: Diversified Background

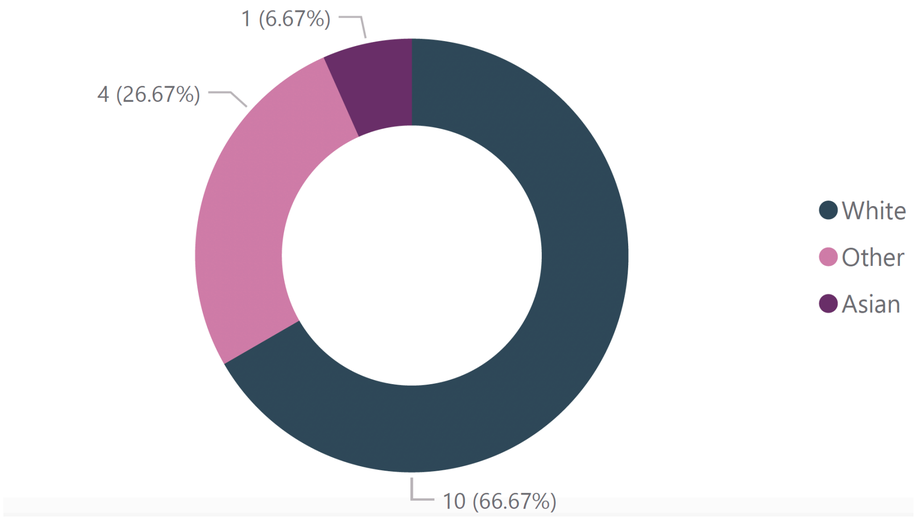
Diversified Background



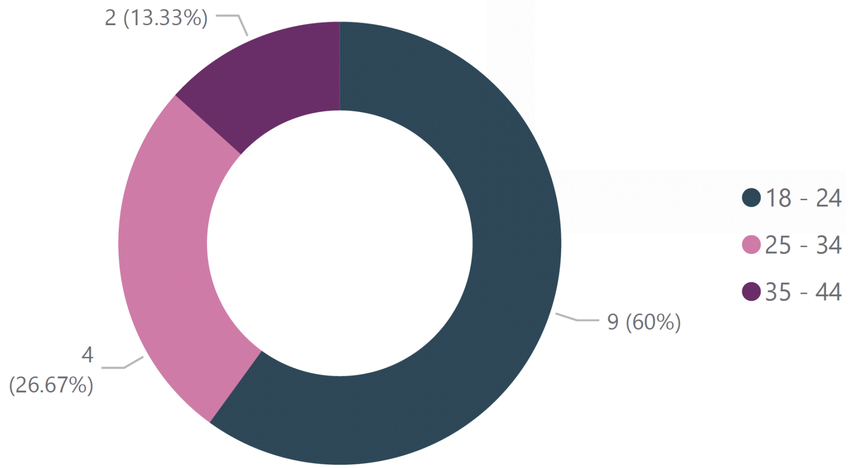
Source: Before-class survey of 16 students in my SOC 3170/5020 Basic Social Statistics, Fall 2023

Diversified Background

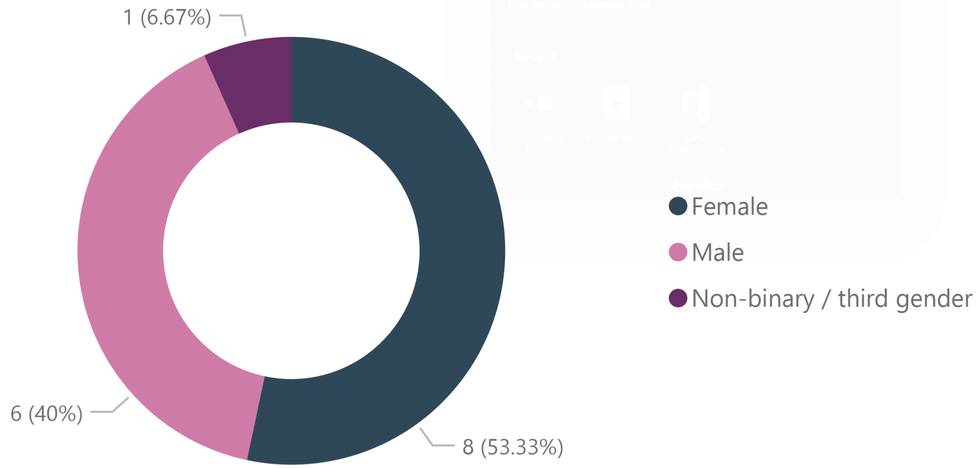
Race of students



Age of students



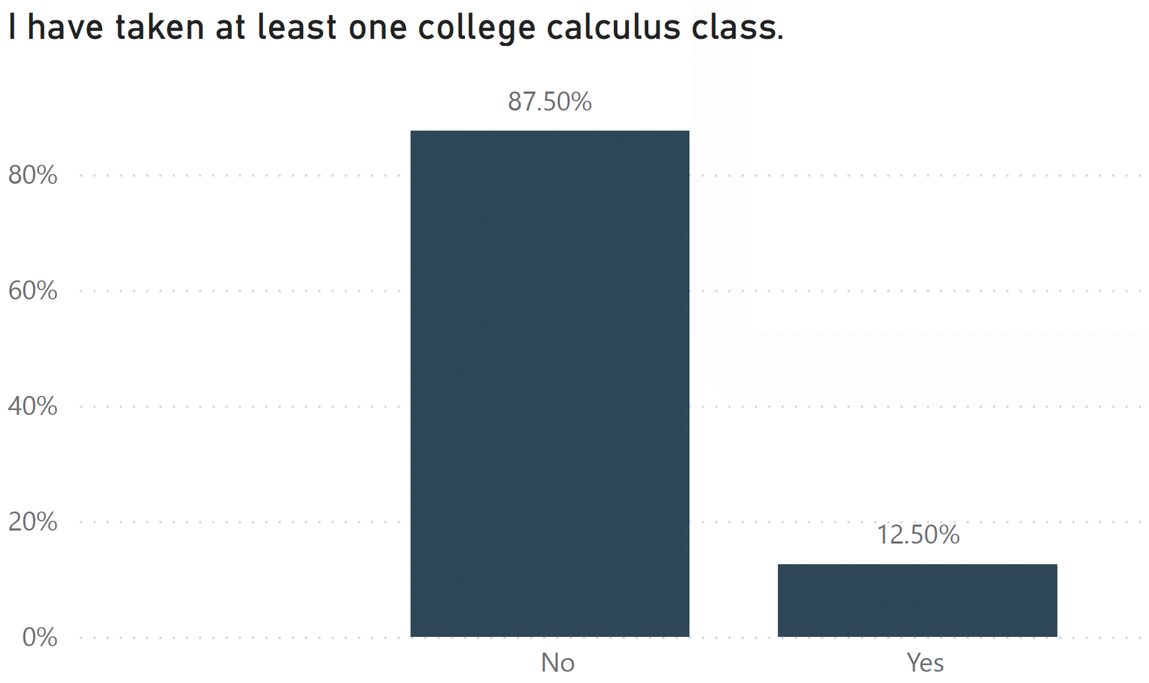
Gender identities of students

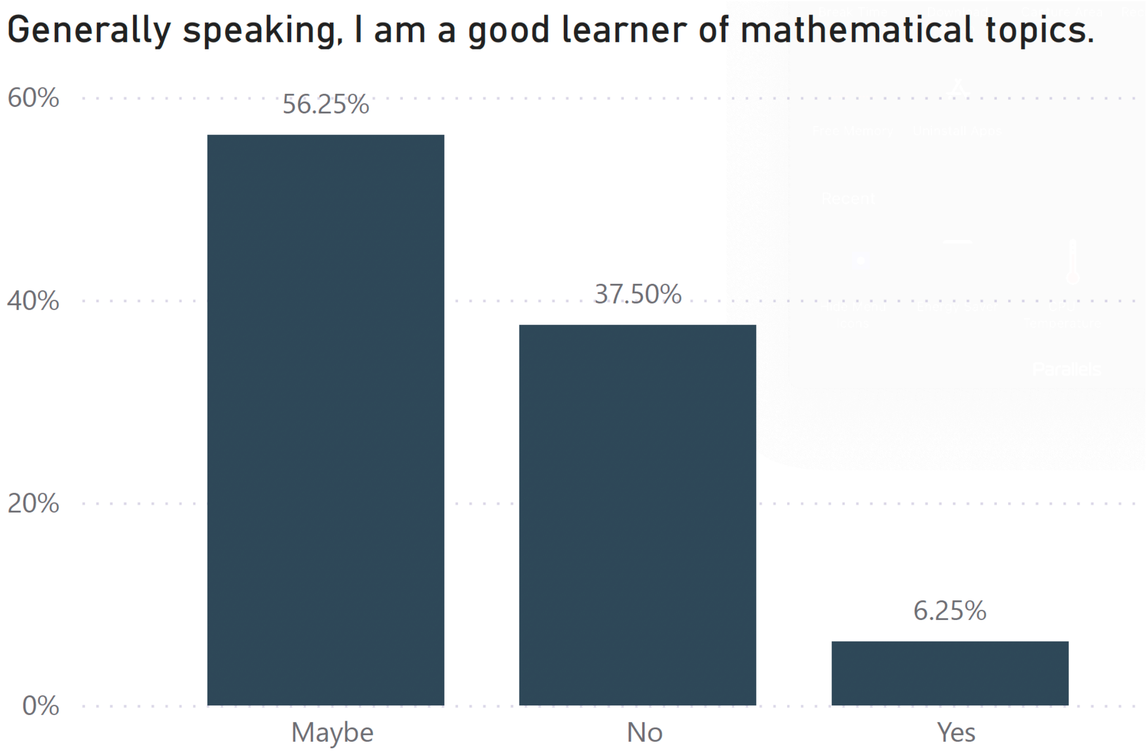


Before-class survey, SOC 3170/5020 Social Statistics, 2023 Fall

Challenge 2: Quantitative Readiness

Diversified Background

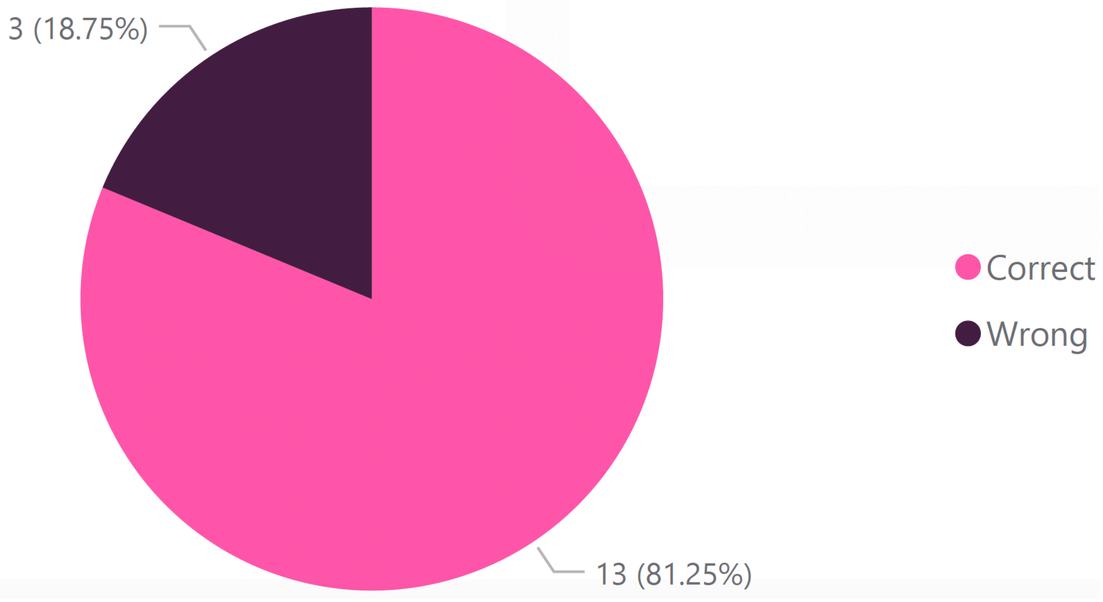




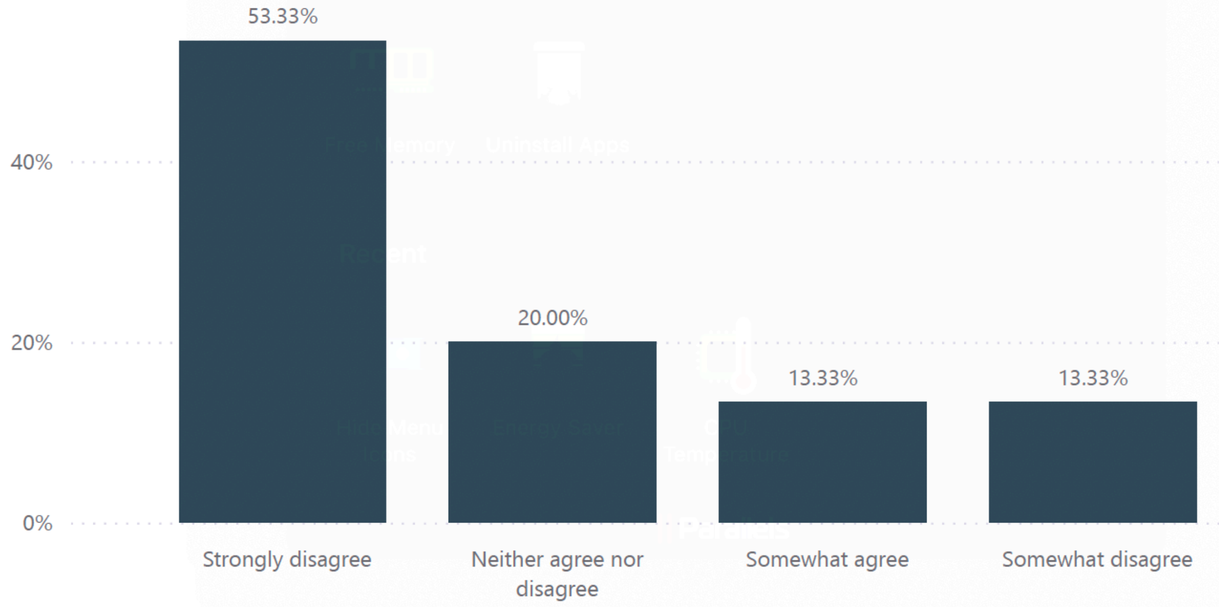
Source: Before-class survey of 16 students in my SOC 3170/5020 Basic Social Statistics, Fall 2023

Quantitative Readiness

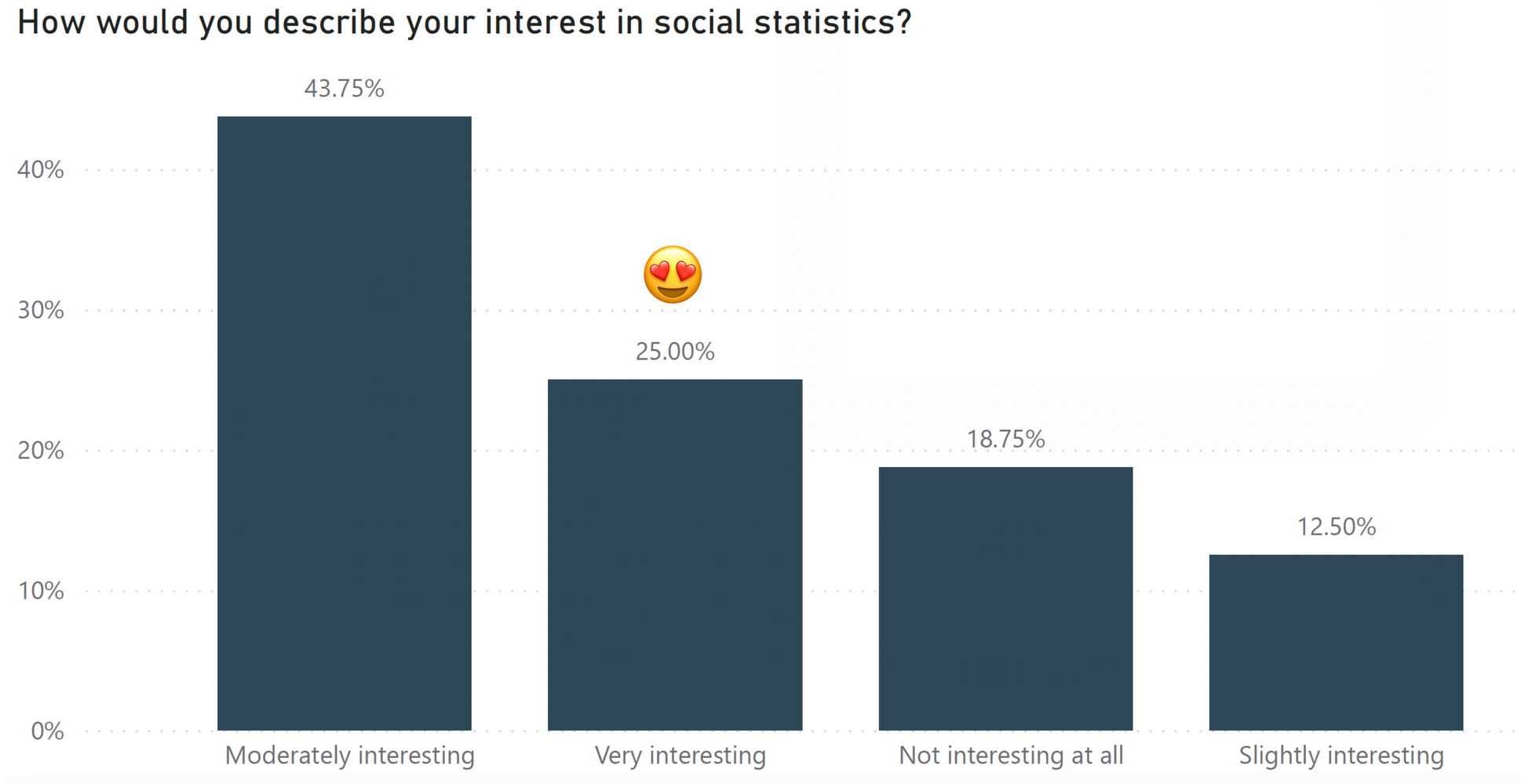
Without using a calculator, my answer to 9+10 x 3 is _____.



I am comfortable reading math notations.



As a result, I face the following pattern



Goal

- **Our main goal is to reach a balance between**
  - Being scientifically rigorous when teaching statistics: we give students the training they need to qualify for quantitative job positions.
  - Being practically realistic: we allow highly diversified social science undergraduate students with relatively low quantitative readiness to learn the essence of quantitative topics.

Being scientifically rigorous

A Possible Solution: The 7-step SOP

The 7-step SOP for Hypothesis Tests

Step 1. Write down null (H_0) and research (H_1) hypotheses.

Step 2. Define the alpha level as the maximum probability of making the Type I Error.

Step 3. Define and calculate a test statistic.

Step 4. Obtain the critical value(s).

Step 5. Compare the calculated test statistic in Step 3 to the critical value(s) in Step 4.

Step 6. If the test statistic exceeds a critical value, reject H_0; otherwise, do not reject H_0.

Step 7. Write down the test conclusion based on the decision in Step 6.

The 7-step SOP is a universal framework

- It applies to any hypothesis test questions, such as undergraduate-level topics of one-sample t-test, independent-samples t-test, chi-squared test of independence, t-and-z-test of correlation coefficient, F-test of ANOVA, and t-tests of OLS regression coefficients.
- It also applies to intermediate topics, such as z-tests for proportions, tests of generalized linear models, tests for log-linear models, and other commonly used nonparametric tests.

Step 1 reflects the logic of scientific research

- A statement about a **population parameter** is called a "hypothesis."
- Our existing knowledge of a population parameter is the null hypothesis (denoted as H_0). And we make scientific contributions by rejecting the null hypothesis.
- The logical negation of H_0 is called the research hypothesis (denoted as H_1).
- For most entry-level hypothesis tests, we want to show that H_0 is highly possible to be false and H_1 is more likely to be true.

Step 2 reflects the uncertainty of scientific research

Decisions	True Status of Hypothesis	
	H_0 is true	H_0 is false
Reject H_0	**Type I error**	Correct decision
Don't reject H_0	Correct decision	Type II error

The α level is the highest tolerable probability of making the Type I error when rejecting the null hypothesis. It is usually a tiny probability, such as 0.10, 0.05, 0.01, or 0.001 for social science studies.

Step 3-7 follows the logic of proof by contradiction

- Q: How do we show H_0 is highly possible to be false?
- A: We borrow the idea of proof by contradiction.
  1. We assume H_0 is true, then we follow certain deductive logic to reach "a new conclusion" that is almost impossible to be true.
  2. Why the new conclusion is almost impossible to be true? Since the deductive logic is perfect, we can only conclude that the H_0 is highly likely to be false.
  3. Since H_0 is highly possible to be false, H_1 is highly possible to be true, given the current sample, alpha level, and the testing method.

Step 3-7 follows the logic of proof by contradiction

- Q: By assuming H_0 is true, what is that "new conclusion" derived from a perfect deductive logic?
- A: It is a calculated "test statistic" using sample statistics and the null hypothesis. This test statistic follows a known theoretical distribution, like SND and T distribution.
- Q: What does "almost impossible" mean?
- A: We use the "small probability event" assumption to define "almost impossible."

Step 3-7 follows the logic of proof by contradiction

- Some events rarely happen, such as winning a lottery, massive earthquakes (magnitude >7.0), and fatal commercial aviation accidents.

✓**Good and bad luck is equally rare.** (Thinking in this way is good for your mental and physical health 😈)

➡**Assumption: a small probability event will be almost impossible in just one trial. You will eventually observe a small probability event only when you repeat trying a large number of times.**

- Just like an old saying in Chinese moral education: frequently walk along the riverside, shoes eventually get wet (常在河边走，哪有不湿鞋).

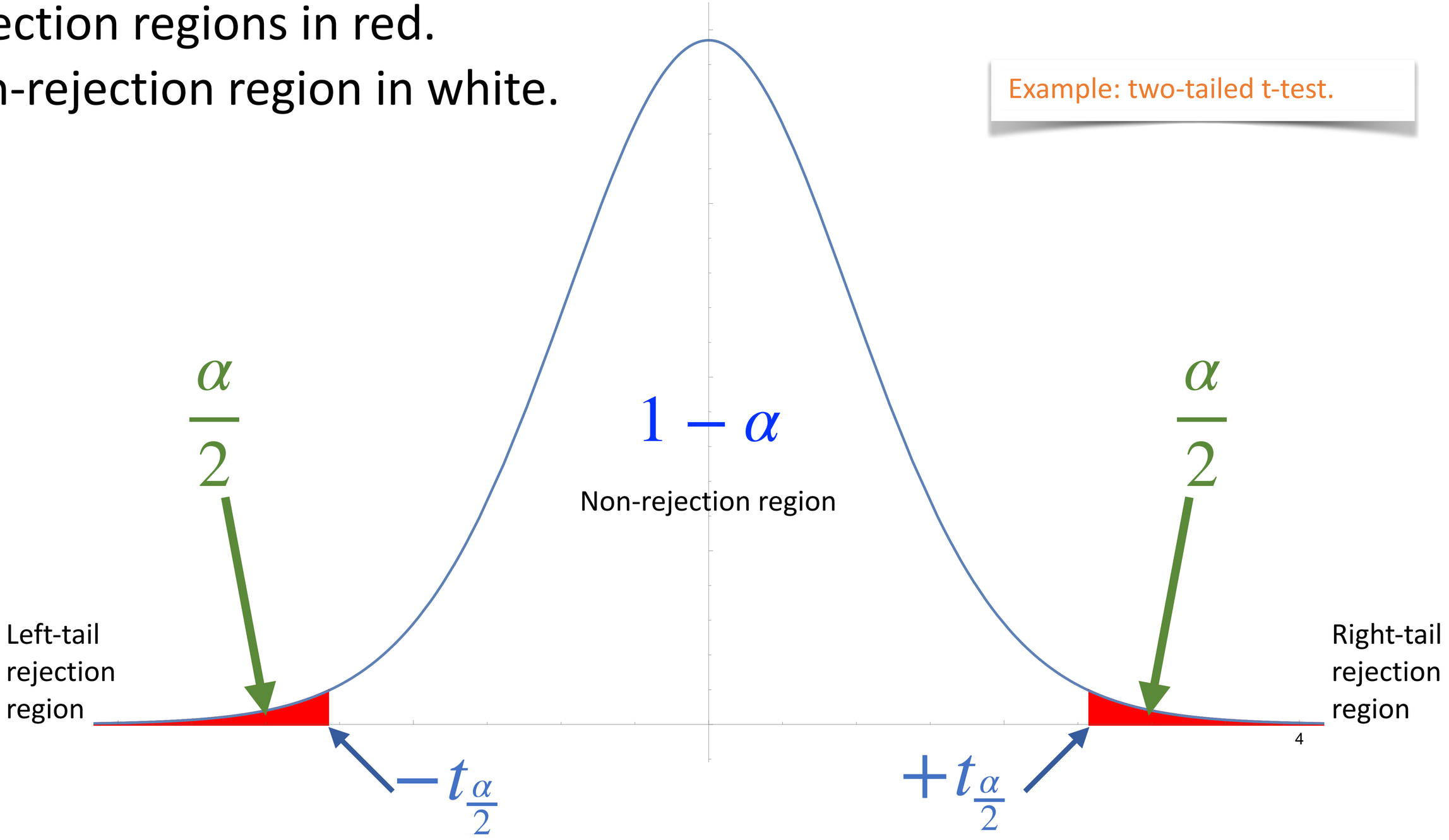
Step 3-7 follows the logic of proof by contradiction

- Q: How do we link "test statistic" to "small probability event"?
- A: Use the alpha level in Step 2 to define "small probability." Usually, we use 0.10, 0.05, 0.01, and 0.001 in social science studies.
- Then, given the alpha level, find the critical value(s) from a known probability distribution of the test statistic.
- Finally, compare the value of the test statistic t* with critical value(s). if t* exceeds the critical value, we face something highly improbable in just one trial. It violates the small probability assumption, so we must reject H_0. This conclusion could be wrong, but the probability of rejecting a true H_0 [making the Type I Error] is smaller than the predefined alpha level in Step 2.

Rejection regions in red.

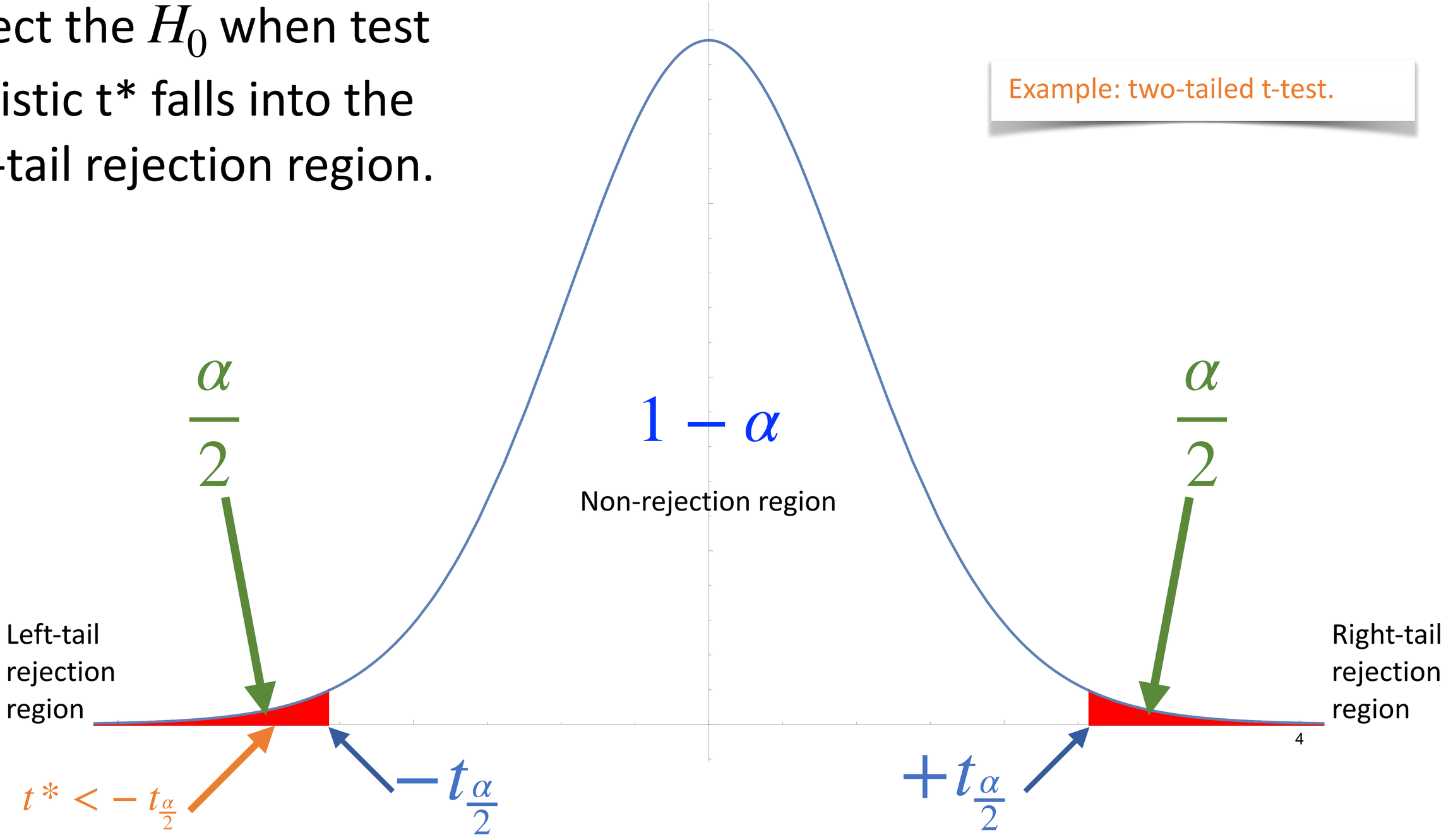
Non-rejection region in white.

Example: two-tailed t-test.



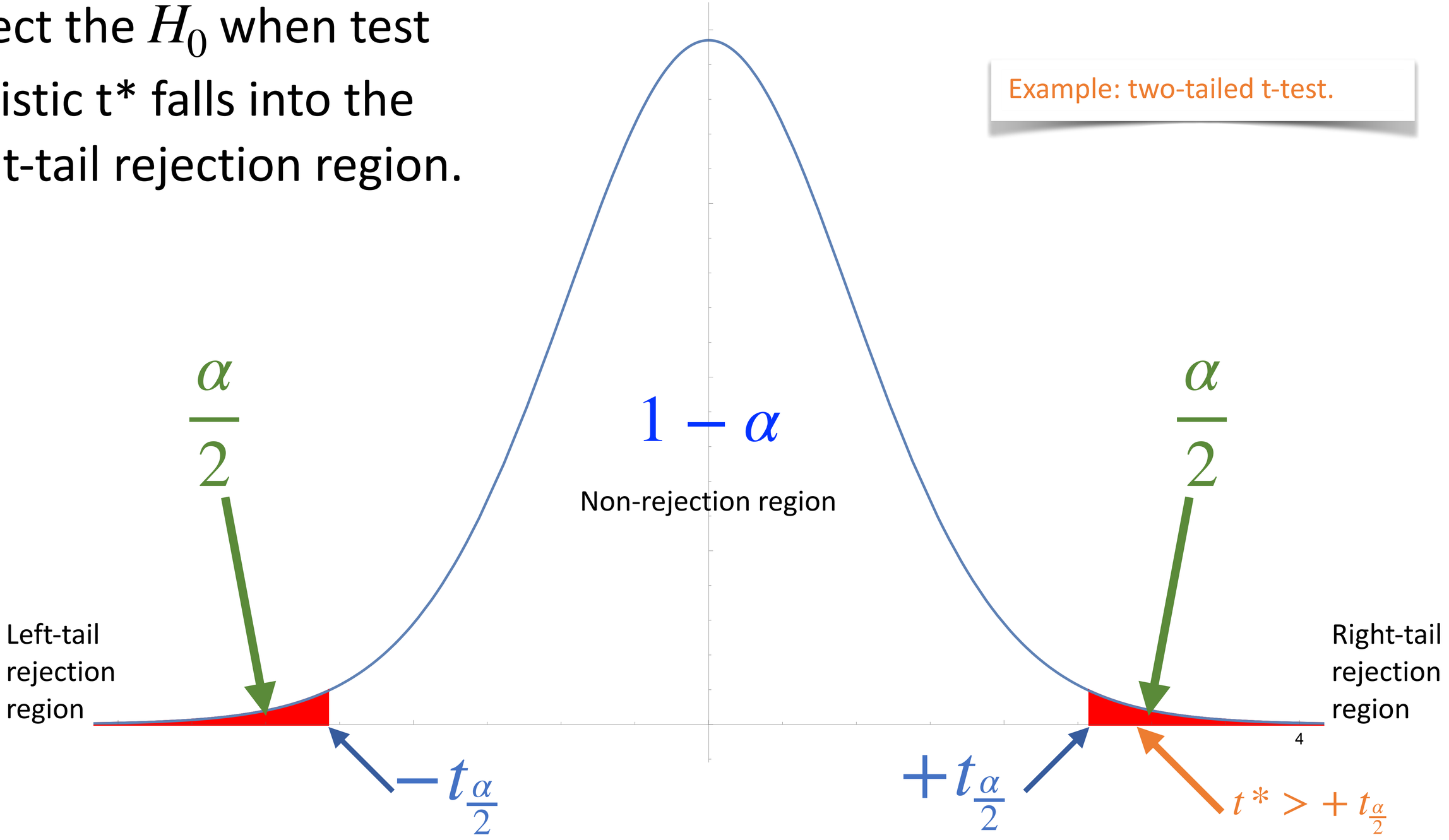
Reject the H_0 when test statistic t* falls into the left-tail rejection region.

Example: two-tailed t-test.



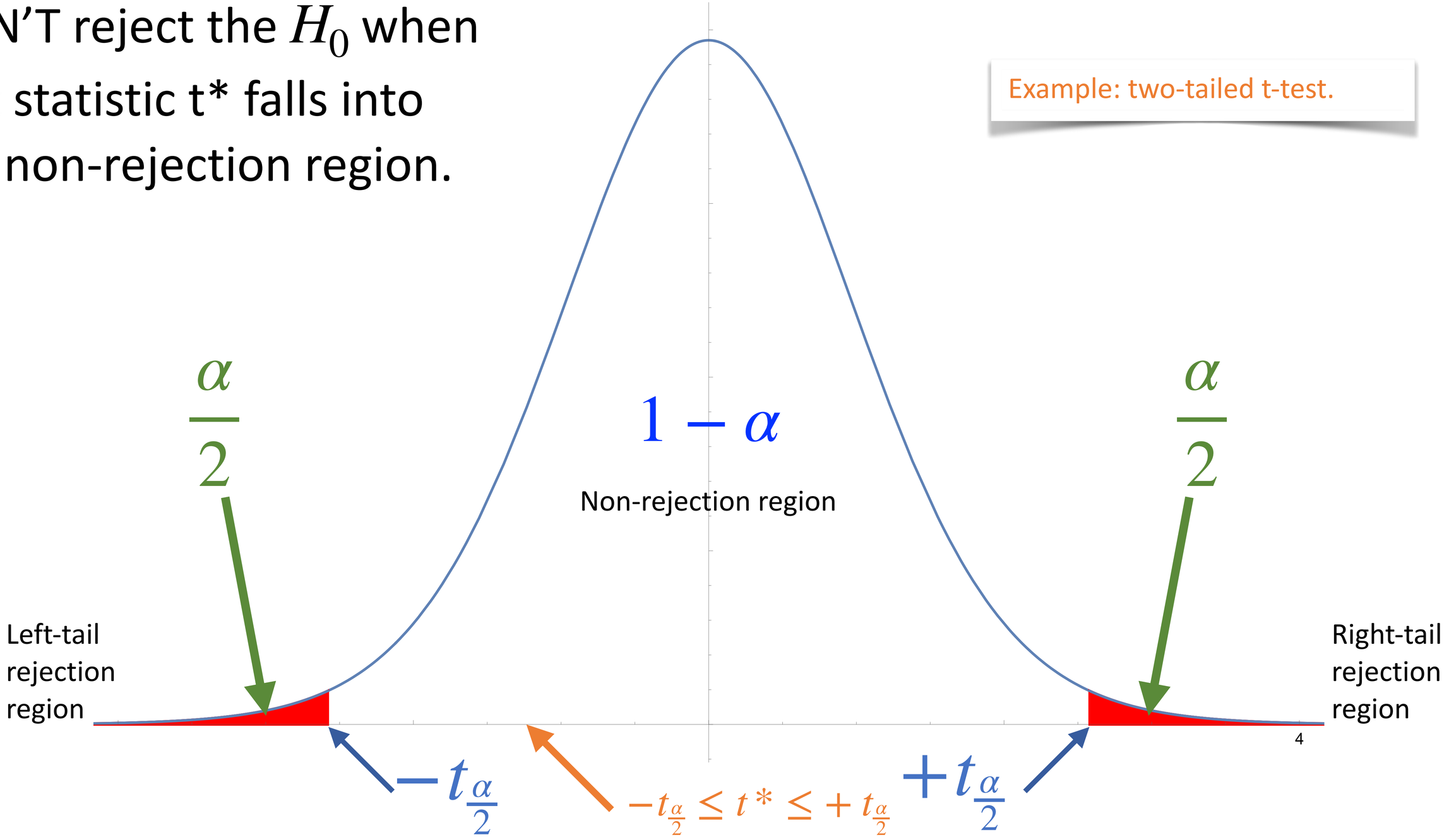
Reject the H_0 when test statistic t* falls into the right-tail rejection region.

Example: two-tailed t-test.



DON'T reject the H_0 when test statistic t* falls into the non-rejection region.

Example: two-tailed t-test.



Being practically realistic

A Possible Solution: The 7-step SOP

The 7-step SOP in teaching

1. The instructor writes the null hypothesis; students complete the research one.
2. Explain the alpha level as the maximum probability of making the Type I Error.
3. Define and calculate a test statistic. Emphasize the correct order of operations.
4. Show how to get critical value(s) given the alpha level in Step 2 by using the textbook appendix and Excel smartsheet.
5. Lead students to compare the test statistic in Step 3 to the critical value(s) in Step 4.
6. Let students decide whether to reject or not reject H_0.
7. Write an example conclusion based on the decision in Step 6.

Example 1: Teaching T-test of Population Mean

- H_0: On average, the US adult population wants two children.
- GSS 2018 data shows that the average ideal number of children = 3.157 with a standard deviation of 1.925 based on 1,381 valid cases. Confined to this sample, how can we test the null hypothesis?
- Let's follow the ***7-step SOP***:

1. Write down null and research hypotheses

   $H_0 : \mu = 2$

   $H_1 : \mu \neq 2$ Hint: this step is a "fill-in-blanks" question in exams.

2. Define the α level. Hint: this step will be completed for you in exams.

Let's use $\alpha = 0.05$ as the maximum probability of making the Type I error.

Example 1: Teaching T-test of Population Mean

3. Define and calculate a ***t*** statistic, which follows a ***T distribution***

$$t = \frac{\bar{Y} - \mu}{\frac{S_Y}{\sqrt{N}}} = \frac{3.157 - 2}{\frac{1.925}{\sqrt{1381}}} = \frac{1.157}{\frac{1.925}{\sqrt{1381}}} \approx \frac{1.157}{\frac{1.925}{37.162}} \approx \frac{1.157}{0.052} \approx 22.25$$

   Hint: the formula will be given in exams. And partially correct calculations give you partial credit(s). The degree-of-freedom (df) is not a required topic.

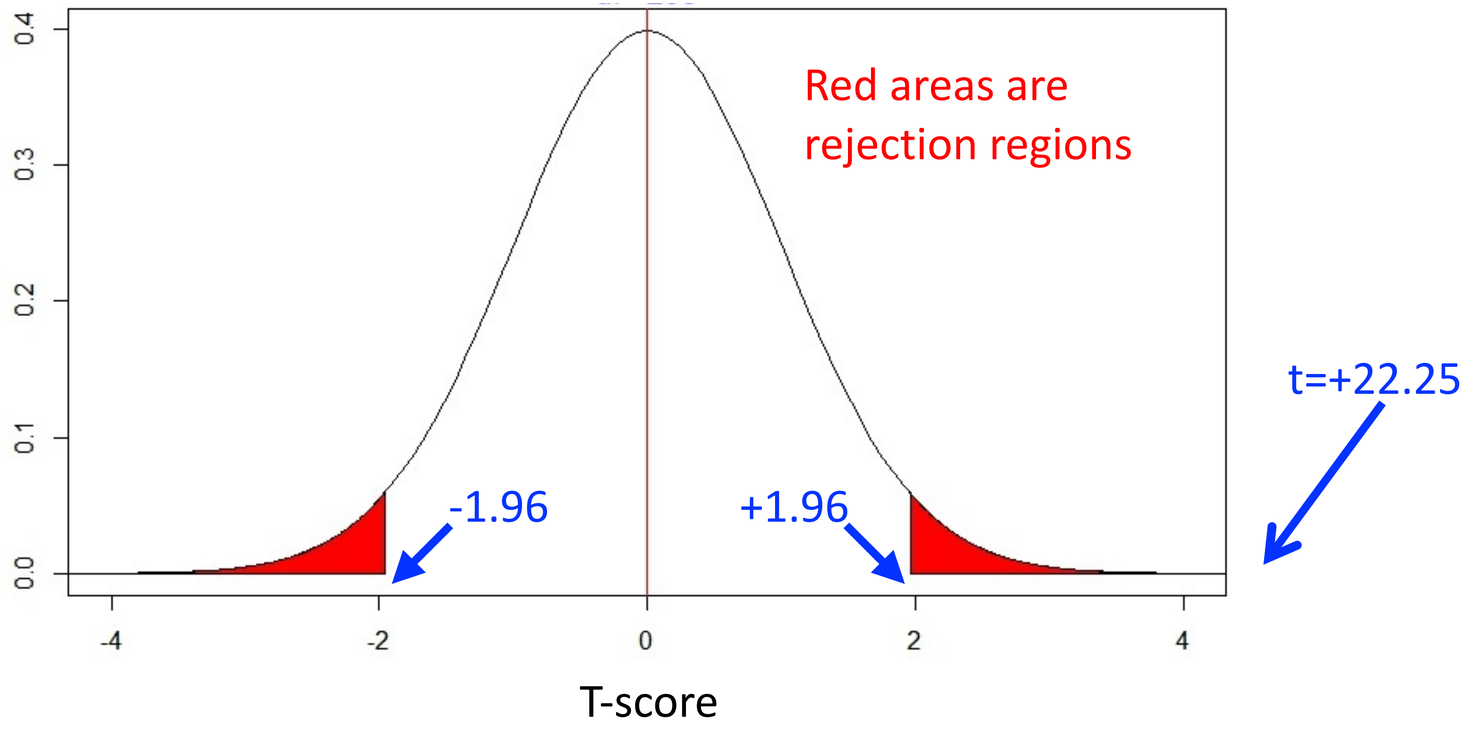
4. Critical values from a T distribution with df=1380 given the alpha level 0.05 in Step 2 are ± 1.96. Hint: the critical values are given in exams.

Example 1: Teaching T-test of Population Mean

5. Compare the test statistic to critical values. Hint: this step will be a "Label-the-graph" question in exams.

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$



Example 1: Teaching T-test of Population Mean

6. The test statistic exceeds the critical value [falls into a rejection region]. We have enough evidence to reject H_0. Hint: this step will be a "Please circle one" question in exams, i.e., "We have/have no (please circle one) enough evidence to reject H_0."
7. We have enough evidence to say that the US adult population does not want to have two children on average. This conclusion could be wrong, but the probability of making the Type I Error is no more than 0.05. Hint: this step will be a "fill-in-blanks" question in exams.

The 7-step SOP in exams

1. One of the null and alternative hypotheses will be a blank to fill in.
2. The alpha level will be given in the question.
3. The formula of test statistic will be given in the question. Partially correct calculations lead to partial credits.
4. The critical value(s) will be given in the question.
5. Compare the calculated test statistic in Step 3 to the critical value(s) in Step 4 by labeling rejection region(s) and test statistic under a distribution curve.
6. The decision of rejecting or not rejecting H_0 will be a "Please circle one" question.
7. The test conclusion based on the decision in Step 6 will be a "Fill-in-blanks" question.

Example 2: T-test of Population Mean in Exam (answer keys in red)

ID: **Name:**

9. In 1980s, the average attitude of US adults on gun control in a 20-point scale (1=strongly support gun control;…; 10=neutral;…; 20=strongly against gun control) was 10. Thirty-five years later, is this average attitude still highly possible to be true? To re-evaluate it, 500 U.S. adults are randomly selected to answer the same 20-point scale on gun control and their mean attitude is 8 with a sample standard deviation of 5. Use a hypothesis testing procedure to complete this re-evaluation (set $\alpha = 0.05$ and use ± 1.960 as critical values). **[20 points]**

Fill in the blank for the alternative hypotheses in math notations or in English: [2 points]

$$H_0: \mu = 10$$
$$H_1: \mu \neq 10$$

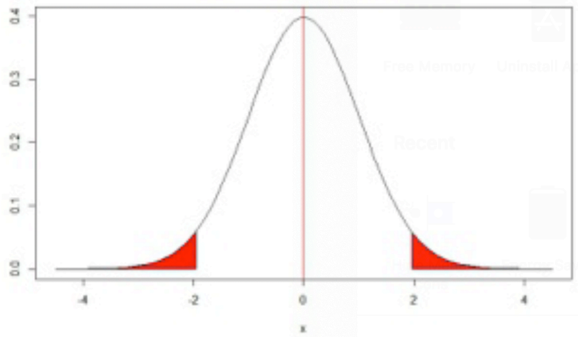
Which test statistic should you use? (Please circle one) [3 point]

A. $t = \frac{\bar{Y}-\mu_0}{S_{\bar{Y}}} = \frac{\bar{Y}-\mu_0}{S_Y/\sqrt{N}}$; B. $F = \frac{MSB}{MSW} = \frac{SSB/df_B}{SSW/df_W}$; C. $z = \frac{p-\pi}{\sqrt{\frac{pq}{N}}}$

Calculate the test statistic [8 points]

$$t = \frac{\bar{Y}-\mu_0}{S_{\bar{Y}}} = \frac{\bar{Y}-\mu_0}{S_Y/\sqrt{N}} = \frac{8-10}{5/\sqrt{500}} \approx -8.944$$

Label critical values and your t-statistic on the following graph [2 point]



Can you reject your null hypothesis? Yes or No? (Please circle one) [2 point]

How to explain your conclusion? [3 points] Remember to cite numbers in your answer!
This conclusion means that after 35 years, the average attitude of gun control for the entire US adult population [1 point] is highly possible [1 point] to be different from 10 [1 point] (Or, US adults' mean attitude tends to be more supportive to gun control [1 point]).

Thank you!

Q&A