MODULE 9

INDEPENDENT-SAMPLES T-TEST

OUTLINE

1. Between-group designs
2. Quick review
3. Significance testing steps
4. Statistical assumptions
5. Study questions

OUTLINE

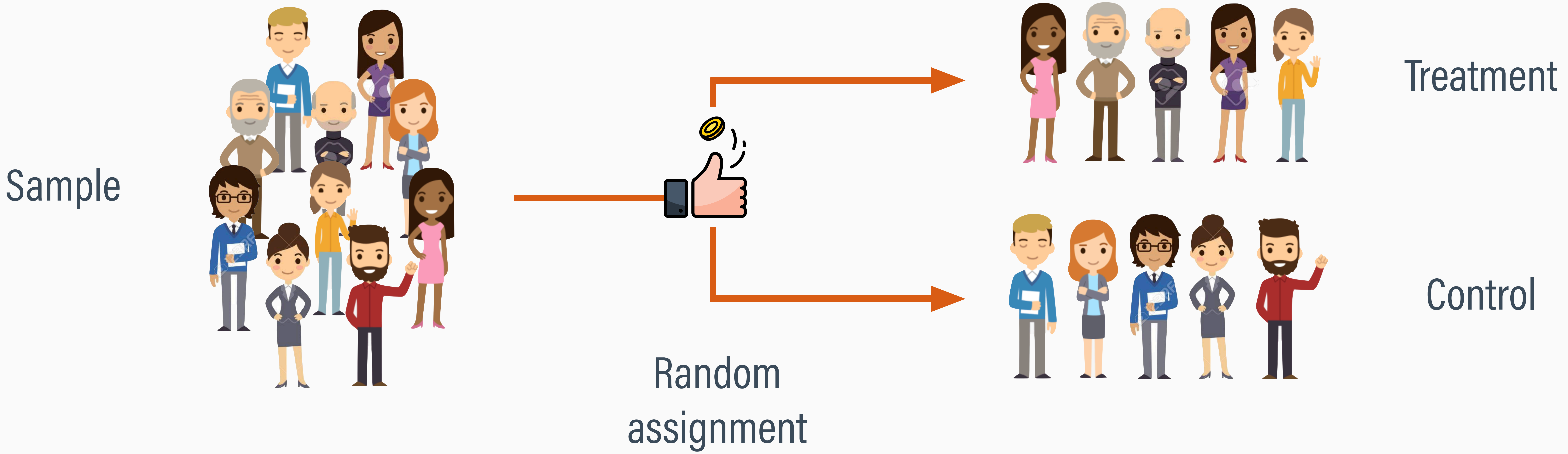
1. Between-group designs
2. Quick review
3. Significance testing steps
4. Statistical assumptions
5. Study questions

BETWEEN-GROUP RESEARCH DESIGNS

- A between-group research design seeks to compare two or more groups of participants
- Unlike the within-group design, each condition is comprised of different participants
- The classic example is a randomized experiment with a treatment and control group, but groups can reflect any qualitative characteristic (e.g., sociodemographic)

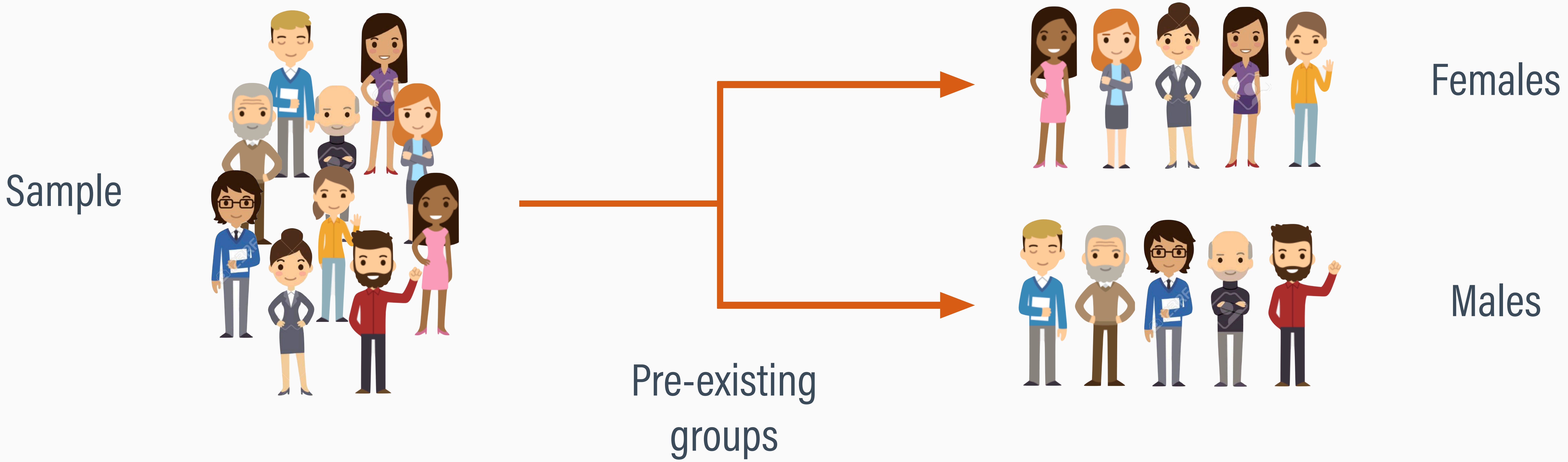
EXPERIMENTAL APPLICATION

- Participants are randomly assigned to either a treatment or a control condition



NON-EXPERIMENTAL APPLICATION

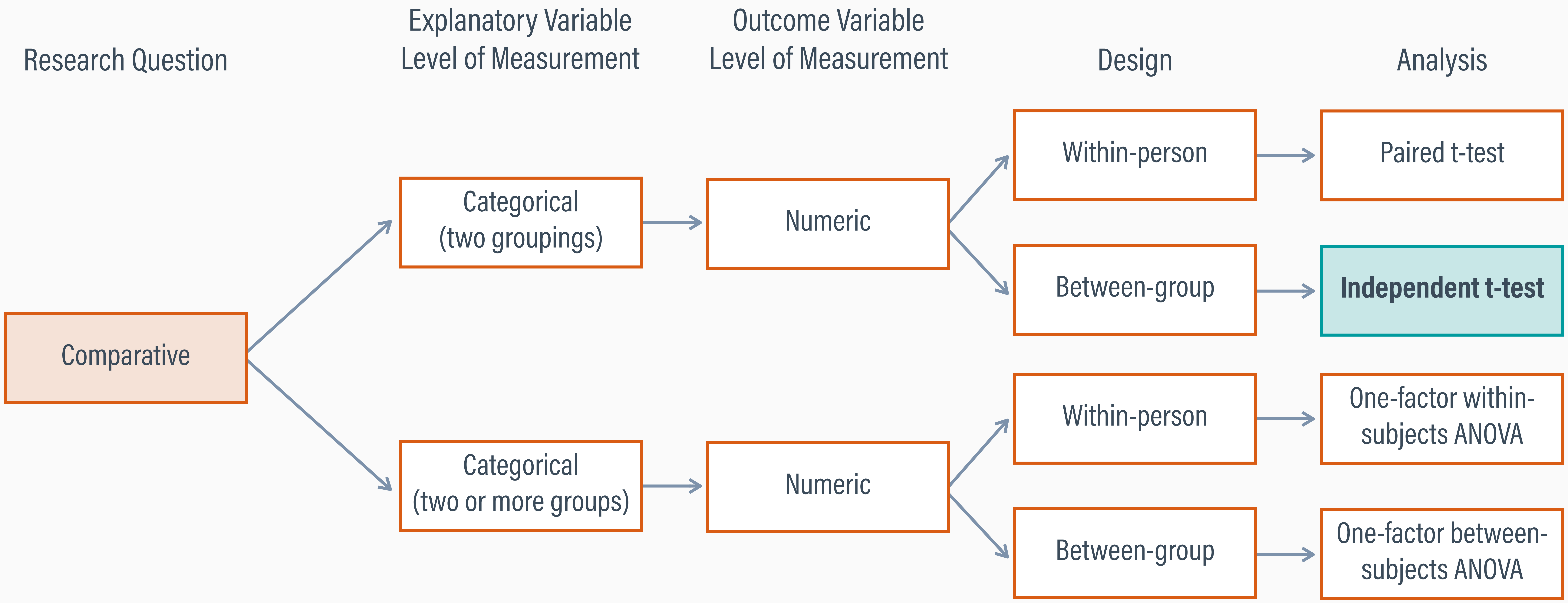
- Participants divide into subgroups based on a shared qualitative characteristic



INDEPENDENT-SAMPLES t TEST

- The independent-samples t-test is appropriate for between-group designs with two groups
- Applicable to comparative research questions and hypotheses involving the difference between two means obtained from the different individuals

STATISTICAL ORG CHART



OUTLINE

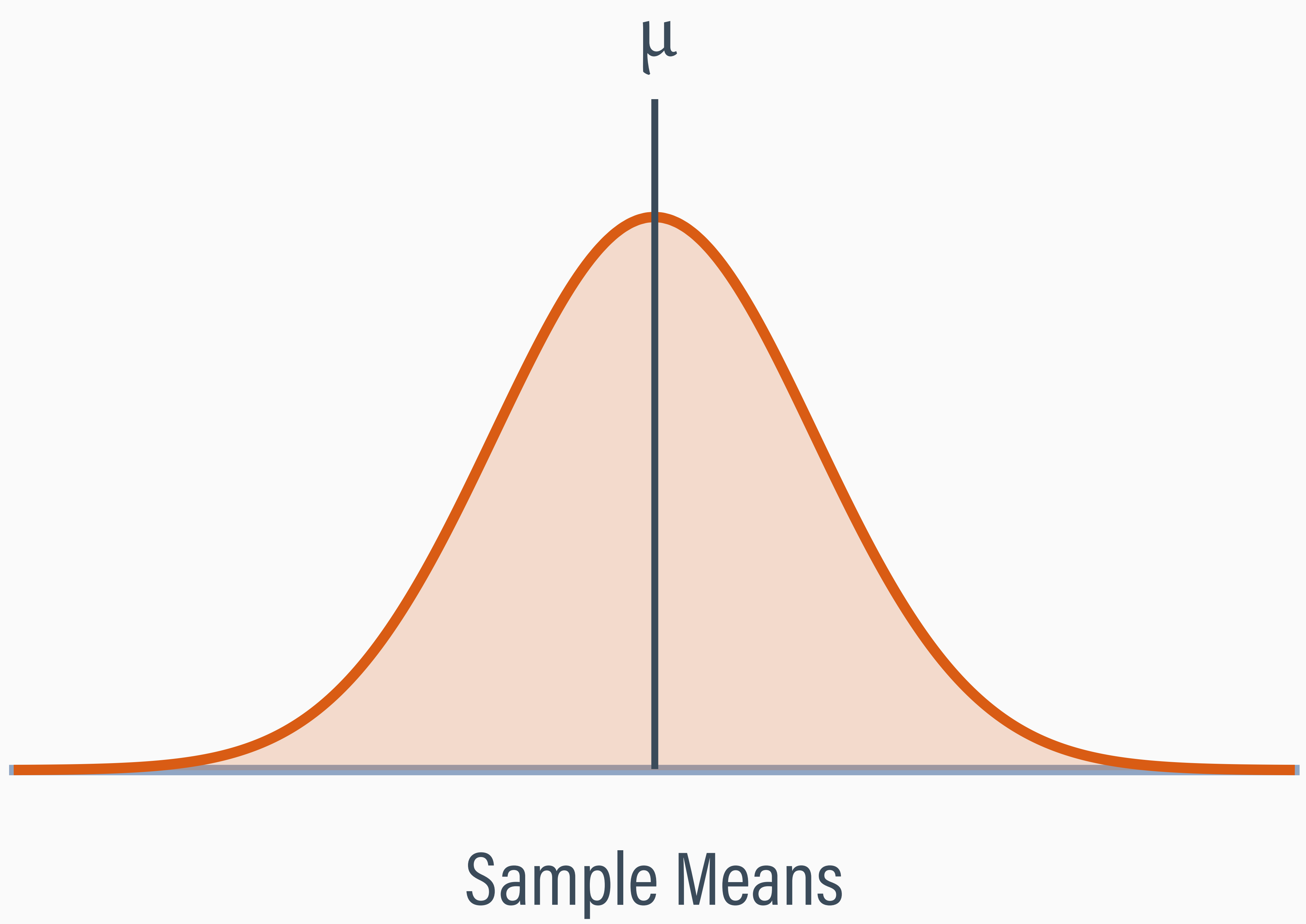
1. Between-group designs
2. Quick review
3. Significance testing steps
4. Statistical assumptions
5. Study questions

QUICK REVIEW: SAMPLING ERROR

- The frequentist paradigm imagines a single population that spawns many hypothetical random samples of data (one parameter, many hypothetical estimates)
- The amount by which an estimate differs from the true population statistic is called sampling error
- Every hypothetical sample has a different amount of error

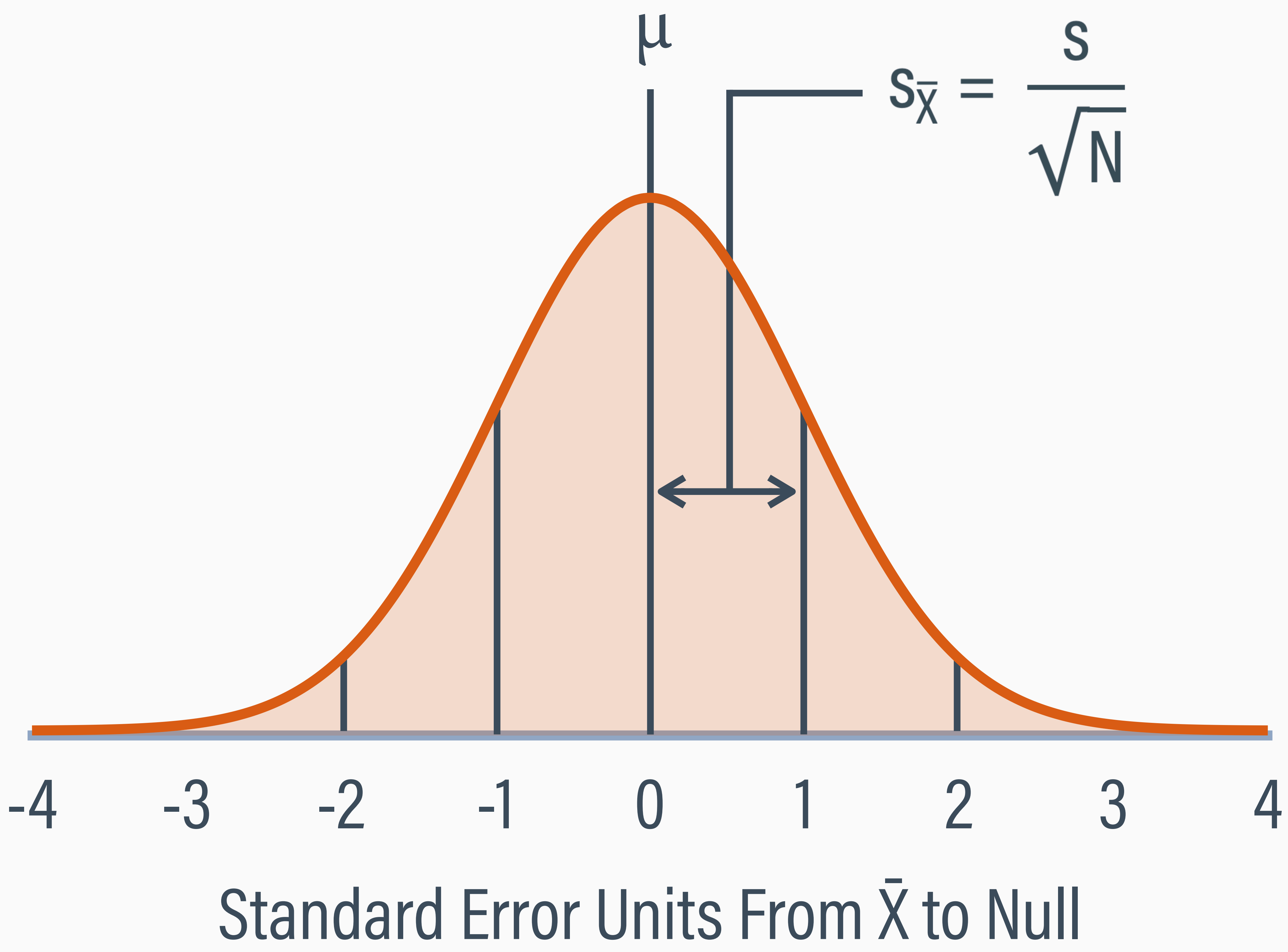
QUICK REVIEW: SAMPLING DISTRIBUTION

- The distribution of the estimates from many hypothetical samples is a sampling distribution
- With a large enough N, sample means follow a normal curve centered at the true mean
- Most estimates have small sampling errors, but a few have larger errors



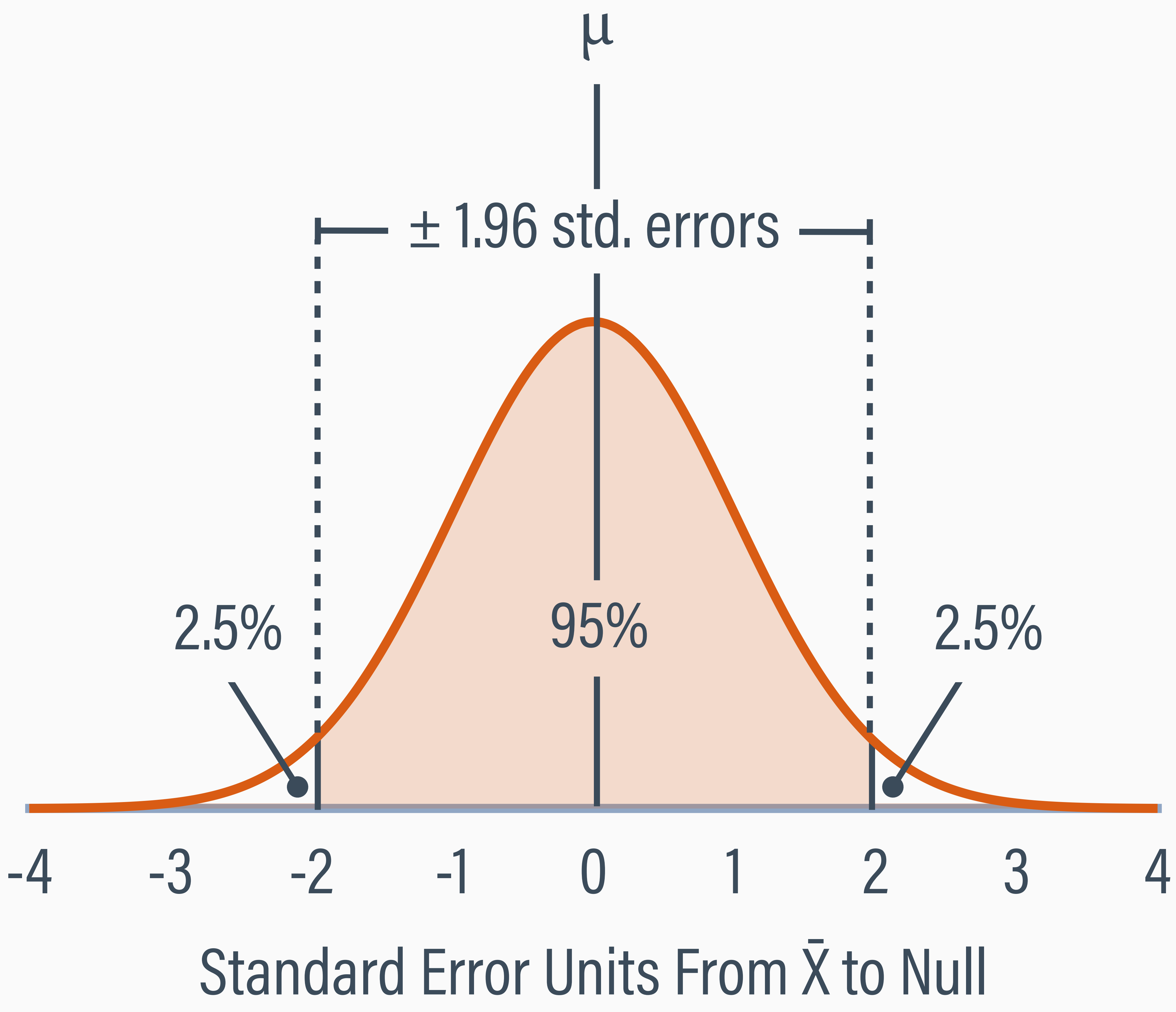
QUICK REVIEW: STANDARD ERROR

- The standard error is the average distance from a sample mean and the true mean
- $s_{\bar{X}}$ = standard deviation of the sample means
- The standard error is the average or expected amount of sampling error across many hypothetical samples



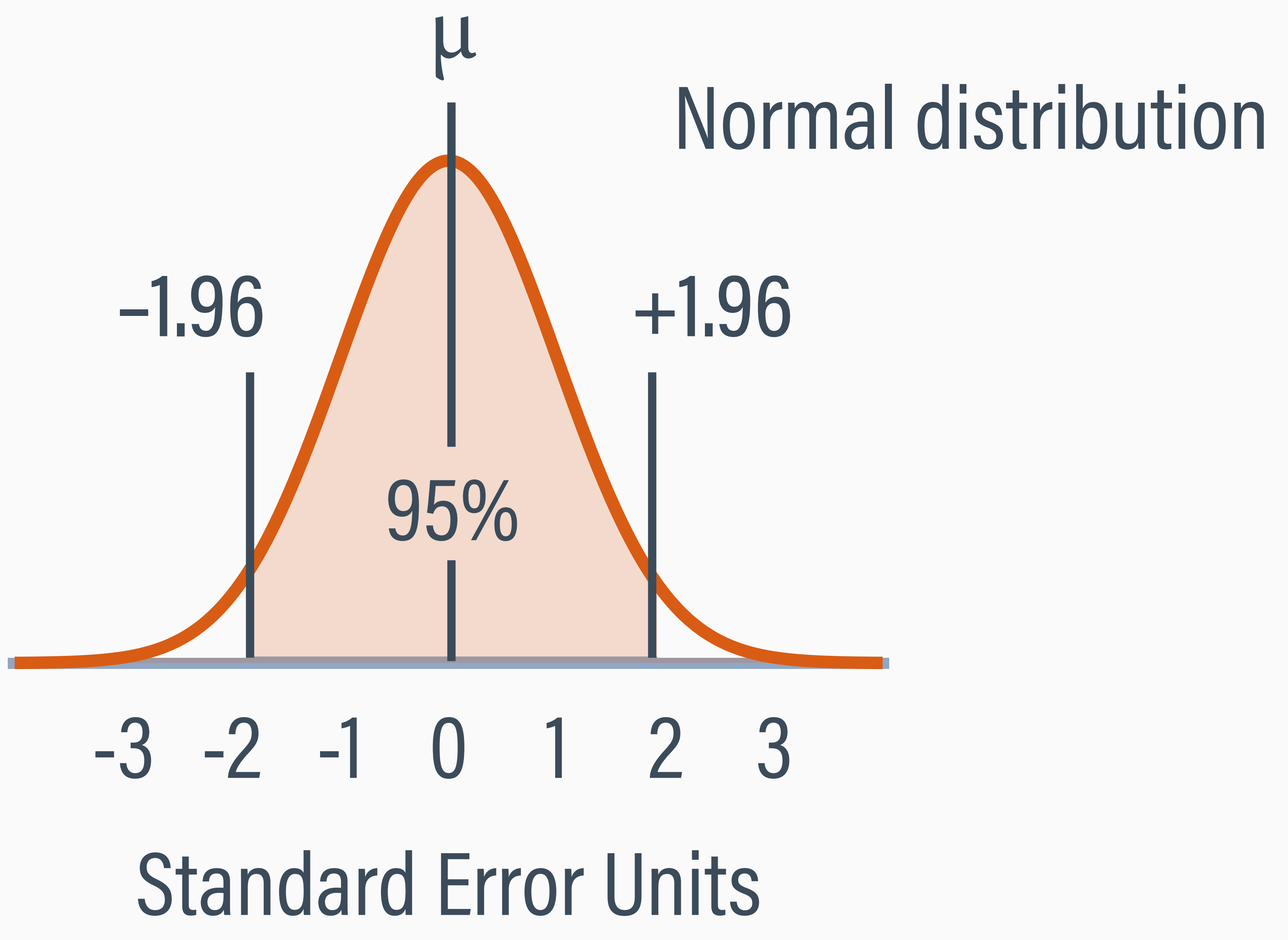
QUICK REVIEW: NORMAL CURVE RULE

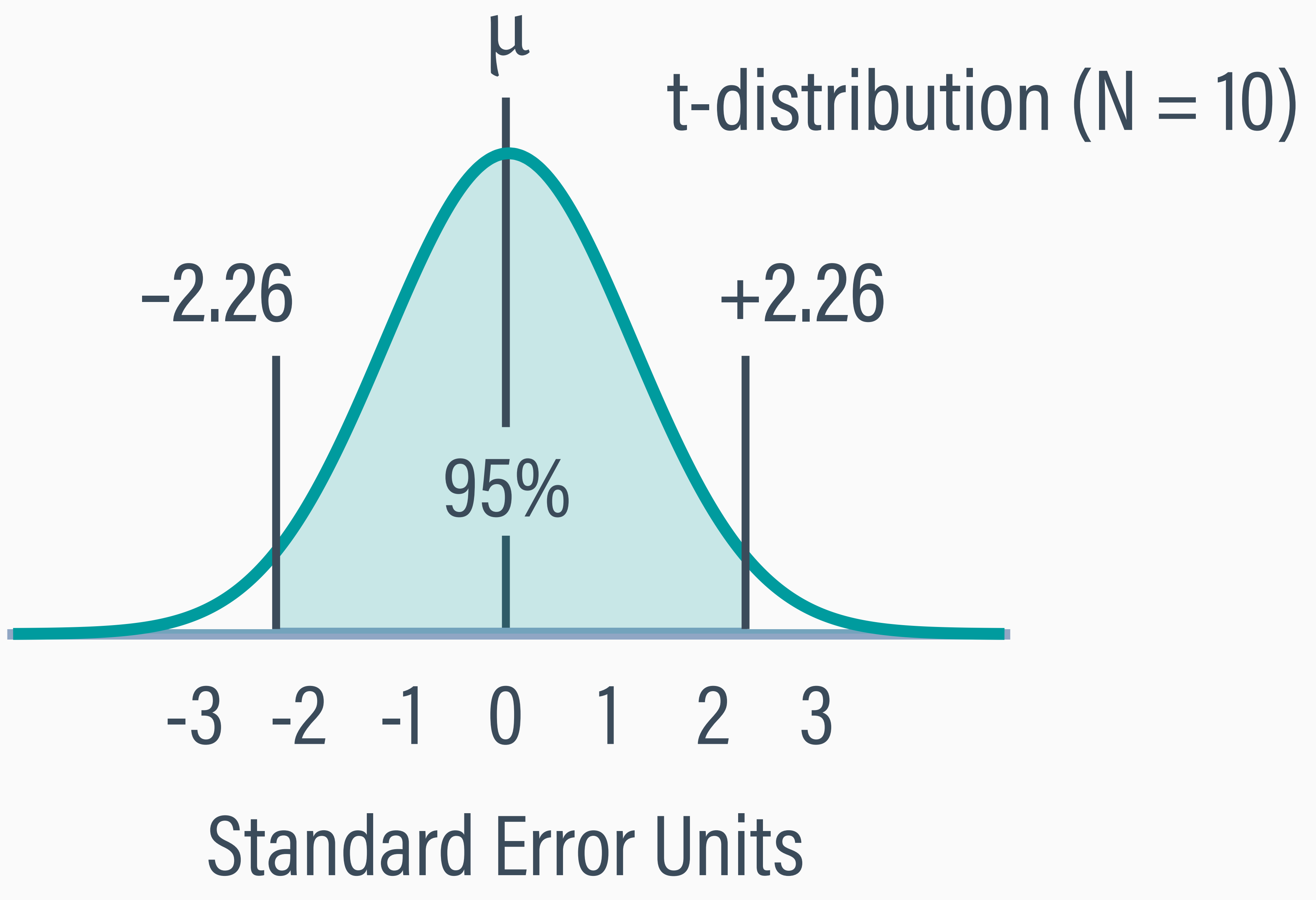
- The standard error is the standard deviation of many hypothetical sample means
- We can apply normal curve rules of thumb
- 95% of the means from large samples are within ± 1.96 standard errors of the true mean



QUICK REVIEW: T-DISTRIBUTION

- When using small samples, the normal curve is an inaccurate description of sampling error
- The t-distribution is a series of bell-shaped curves that stretch out (become more variable) as the N decreases
- Small samples are more likely to produce outlier estimates, and “stretching” the curve honors that

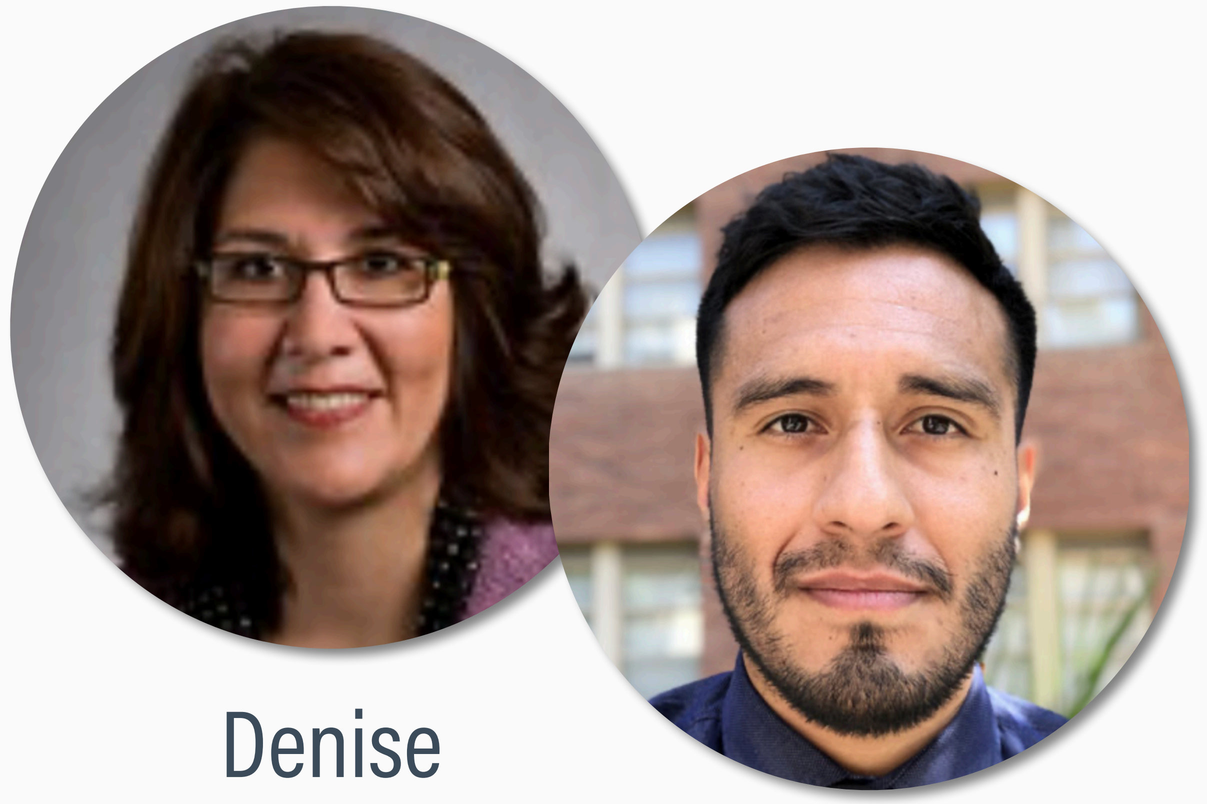




OUTLINE

DISCRIMINATION AND INTERNALIZING BEHAVIOR

There is a well-documented relationship between perceived discrimination and internalizing symptoms among Latinx adolescents. However, few studies have examined how this psychosocial stressor relates to multiple domains of functioning in rural Latinx adolescents simultaneously. This study tested a spillover model of internalizing symptom development, where the negative effects of perceived discrimination are experienced through peer and family relationships.



Denise Chavira

Giovanni Ramos

Ramos, G., Delgadillo, D., Fossum, J., Montoya, A.K., Thamrin, H., Rapp, A., Escovar, E., Chavira, D.A. (2021). Discrimination and internalizing symptoms in rural Latinx adolescents: An ecological model of etiology. *Children and Youth Services Review*, *130*, 1-10.

KEY VARIABLES

Perceived Discrimination

Perceived discrimination refers to individuals' perception of negative attitude, judgment, or unfair treatment due to their specific characteristics such as gender, race, ethnicity, and social status.

Gender

Respondents self-reported their biological gender.

RESEARCH QUESTION

- Question: Do male and female Latinx adolescents experience different levels of perceived discrimination?
- The explanatory (independent) variable, gender, consists of two groups: males and females
- The outcome (dependent) variable, perceived discrimination, is a numeric measure derived from questionnaire items

SIGNIFICANCE TESTING STEPS

1. Specify hypotheses
2. Define standard of evidence
3. Design study and collect data
4. Compare data to null hypothesis
5. Evaluate hypotheses and draw conclusion

MEAN DIFFERENCE STATISTIC

- Both groups have population means, μ_F and μ_M
- Hypotheses use a mean difference statistic that contrasts the two population means

$$\mu_{diff} = \mu_F - \mu_M$$

- The mean difference quantifies treatment effect in this case

NULL HYPOTHESIS

- In the population, there is no difference between males and females

$$H_0: \mu_{diff} = 0$$

- The null that $\mu_{diff} = 0$ is counter to expectations because researchers anticipate that gender could differentially impact perceived discrimination

TWO POSSIBLE ALTERNATIVE HYPOTHESES

- One-tailed alternate: Only one of the gender groups could conceivably experience higher perceived discrimination

$$H_A: \mu_{diff} < 0 \ (\mu_F < \mu_M) \quad \text{or} \quad H_A: \mu_{diff} > 0 \ (\mu_F > \mu_M)$$

- Two-tailed alternate: Perceived discrimination could be higher in either group

$$H_A: \mu_{diff} \neq 0 \ (\mu_F < \mu_M \text{ or } \mu_M > \mu_F)$$

SIGNIFICANCE TESTING STEPS

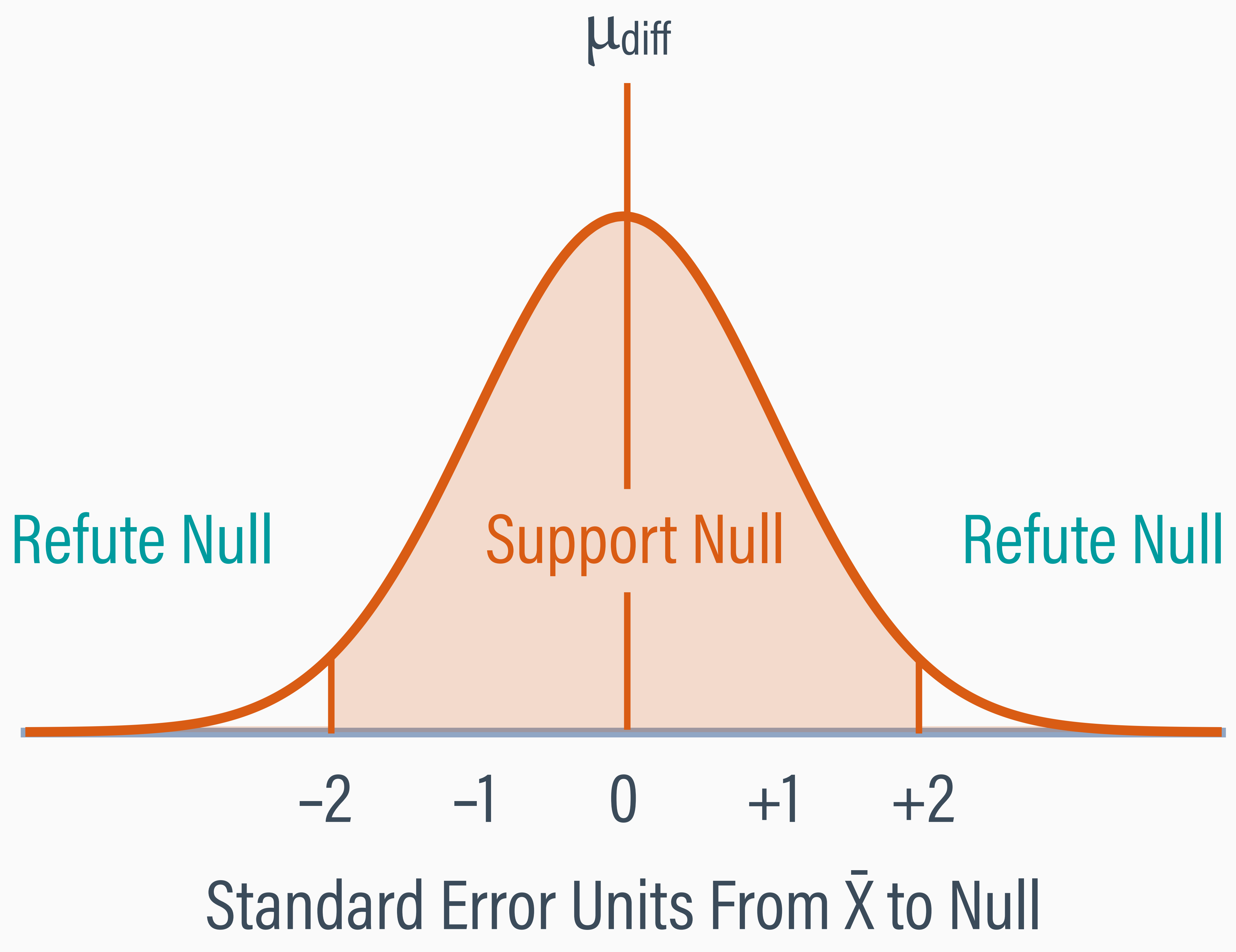
1. Specify hypotheses
2. Define standard of evidence
3. Design study and collect data
4. Compare data to null hypothesis
5. Evaluate hypotheses and draw conclusion

STANDARD OF EVIDENCE

- The data are the evidence that we use to conclude whether the null is plausible (“innocent”) or implausible (“guilty”)
- If the sample mean from our data is very different from the null mean, then we conclude that the null hypothesis is implausible
- How big a difference do we need to observe to refute the null?

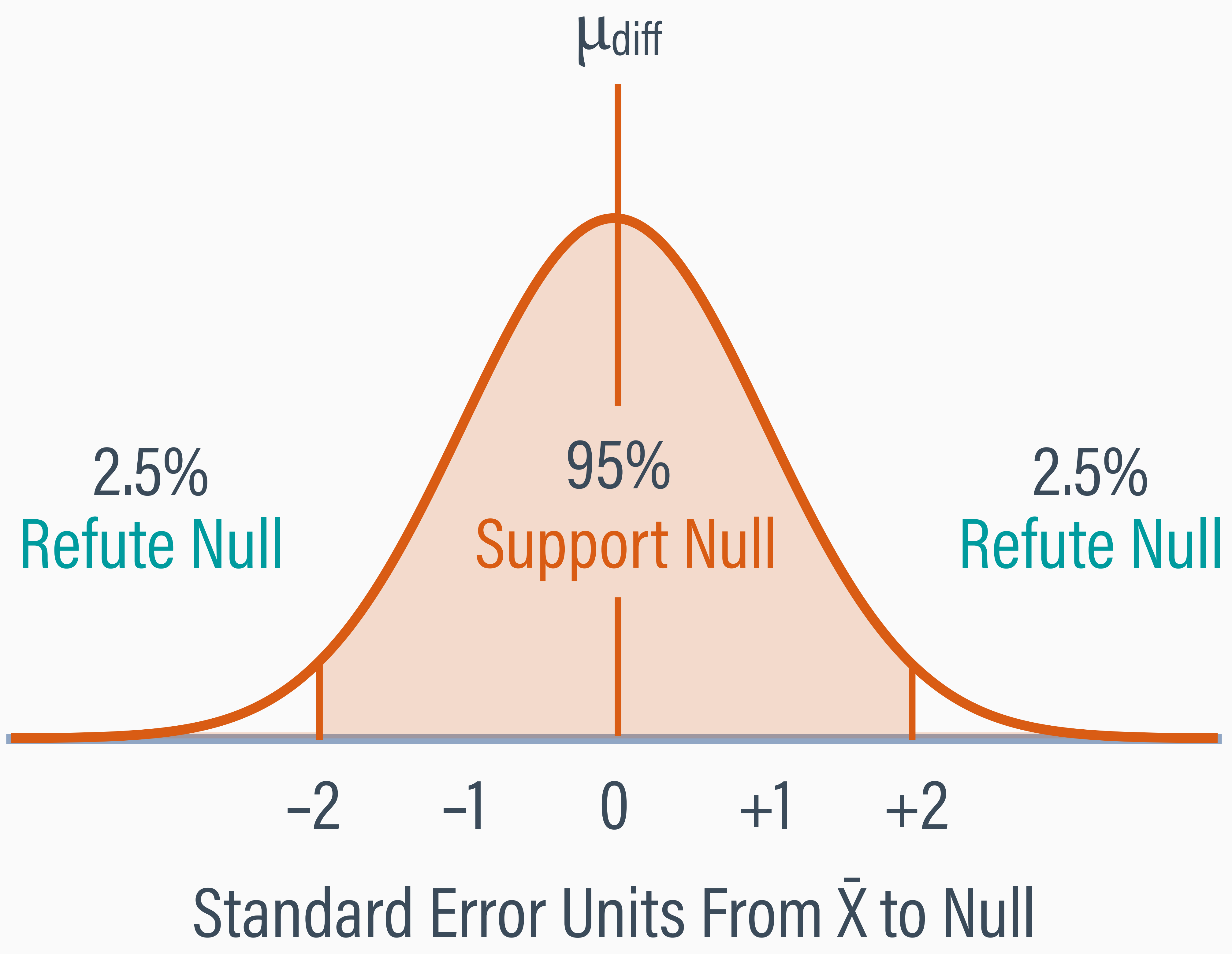
EVALUATING THE NULL

- Any $\bar{X}_{diff}$ near the middle of the sampling distribution ($\mu_{diff} = 0$) lends support to the null
- Such as sample has a high probability of originating from the null population
- We refute the null if the sample $\bar{X}_{diff}$ falls far from μ_{diff}
- Such as sample has a low probability of originating from the null population



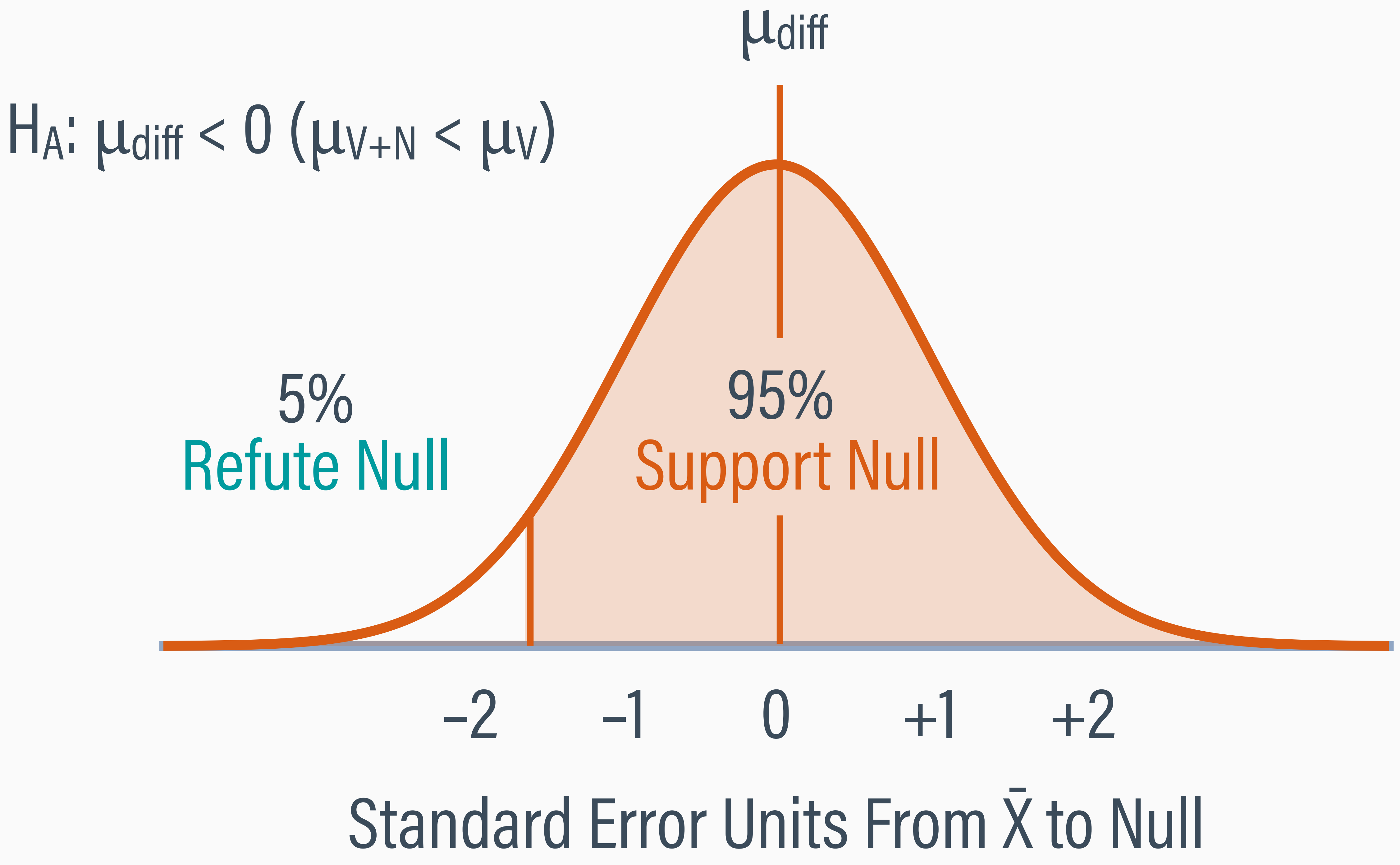
TWO-TAILED ALTERNATE HYPOTHESES

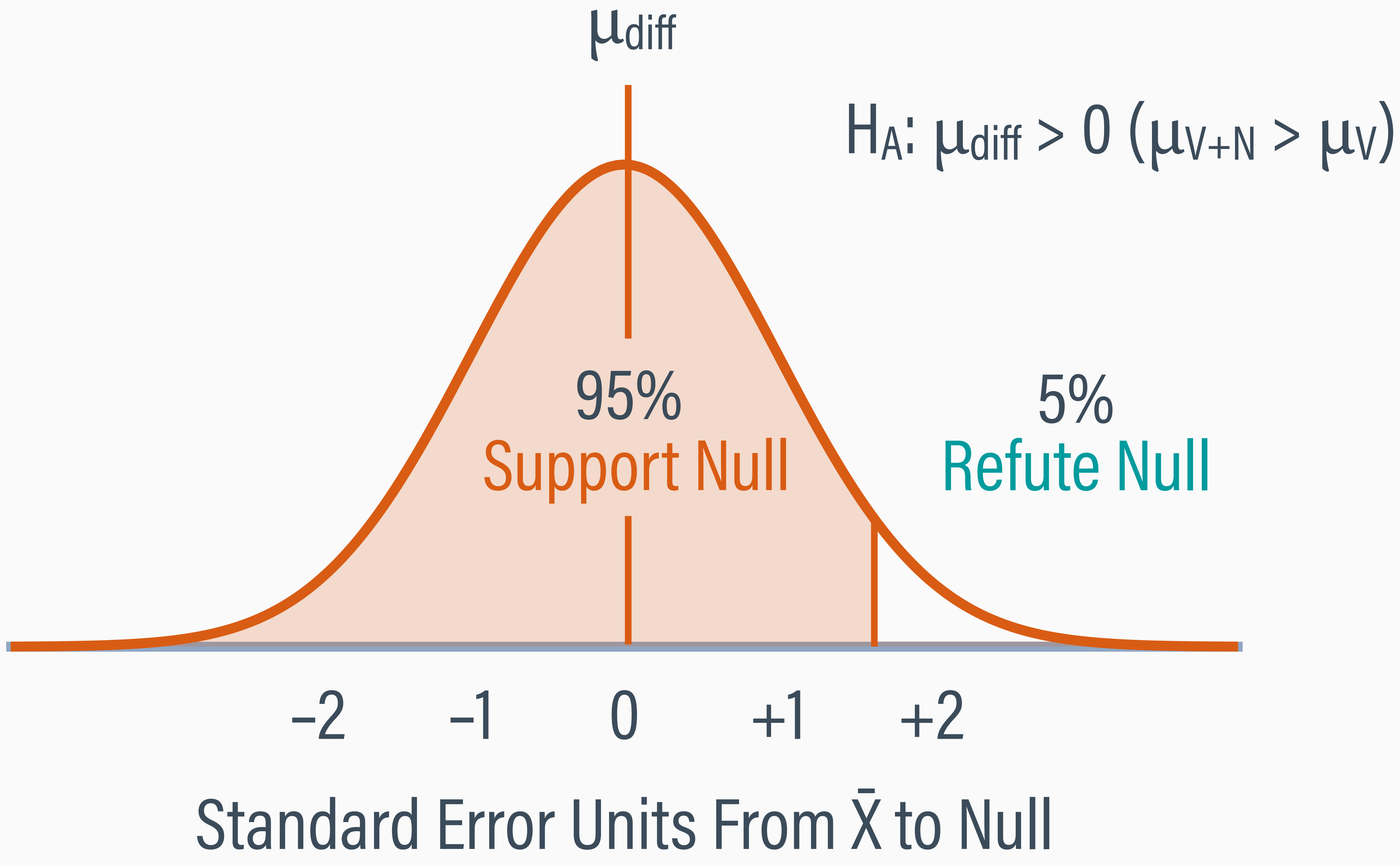
- By convention, we refute the null if the sample $\bar{X}_{diff}$ falls outside the middle 95% of the sampling distribution
- Such as sample has less than a 5% chance of originating from the null population ($p < .05$)
- The 5% rejection region (alpha level) is split in half to allow for the possibility that either an increase or a decrease provides evidence against H_0



ONE-TAILED ALTERNATE HYPOTHESES

- The 5% rejection region (alpha level) is placed in one tail, since only a positive (or negative) $\bar{X}_{diff}$ counts as evidence against H_0





SIGNIFICANCE TESTING STEPS

1. Specify hypotheses
2. Define standard of evidence
3. Design study and collect data
4. Compare data to null hypothesis
5. Evaluate hypotheses and draw conclusion

ANALYSIS SUMMARY

- N = 639 participants across two groups
- The sample mean difference is $\bar{X}_F - \bar{X}_M = +0.86$ discrimination points (13.97 – 13.11 = +0.86)
- Could a mean difference of 0.86 originate from a null population where both groups are equal?

Gender	$\bar{X}$	s	n
Female	13.97	4.57	339
Male	13.11	4.32	300

R OUTPUT

```
Gender: Female
               vars   n  mean   sd median trimmed  mad min max range skew kurtosis   se
Discrimination    1 339 13.97 4.57     13    13.6 4.45   5  34    29 0.94     1.62 0.25
-----------------------------------------------------------------------------------------
Gender: Male
               vars   n  mean   sd median trimmed  mad min max range skew kurtosis   se
Discrimination    1 300 13.11 4.32     12   12.77 3.71   4  31    27  0.9      1.2 0.25
```

EFFECT SIZE

- The effect size expresses the mean difference ($\bar{X}_{diff} = 0.86$) on a standardized metric
- The mean difference of 0.86 on the discrimination scale equates to 0.19 standard deviation units
- The two means differ by 0.19 z-score units

Gender	$\bar{X}$	s	n
Female	13.97	4.57	339
Male	13.11	4.32	300

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} =$$

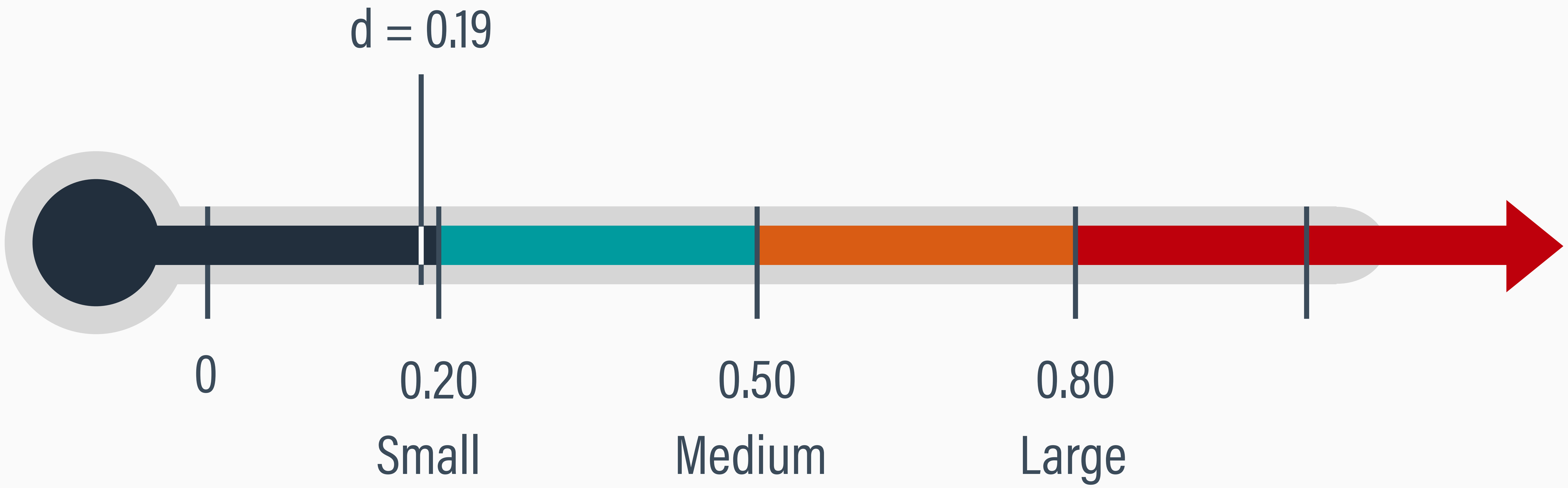
$$\frac{|13.97 - 13.11|}{\sqrt{\frac{(339 - 1)4.57^2 + (300 - 1)4.32^2}{339 + 300 - 2}}} = 0.19$$

R OUTPUT

```
Cohen d statistic of difference between two means
               lower effect upper
Discrimination -0.35  -0.19 -0.04

Multivariate (Mahalanobis) distance between groups
[1] 0.19

r equivalent of difference between two means
Discrimination
          -0.1
```

COHEN'S BENCHMARKS

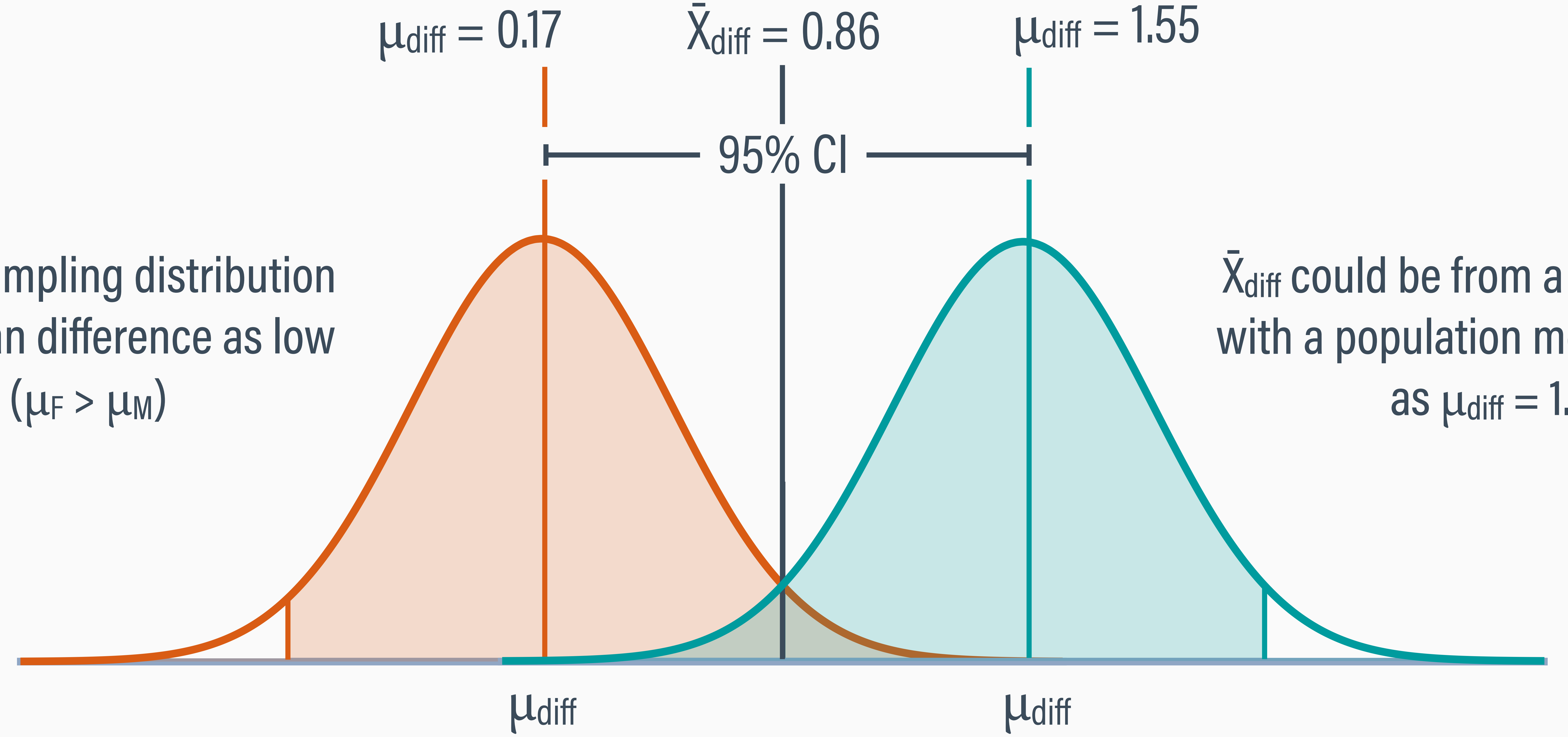
- A 0.19 standard deviation difference is very close to the small effect size cutoff



95% CONFIDENCE INTERVAL

- The 95% confidence interval gives the two most extreme values of the population mean that could have reasonably produced these data

$\bar{X}_{diff}$ could be from a sampling distribution with a population mean difference as low as $\mu_{diff} = 0.17$ $(\mu_F > \mu_M)$



$\bar{X}_{diff}$ could be from a sampling distribution with a population mean difference as high as $\mu_{diff} = 1.55$ $(\mu_F > \mu_M)$

R OUTPUT

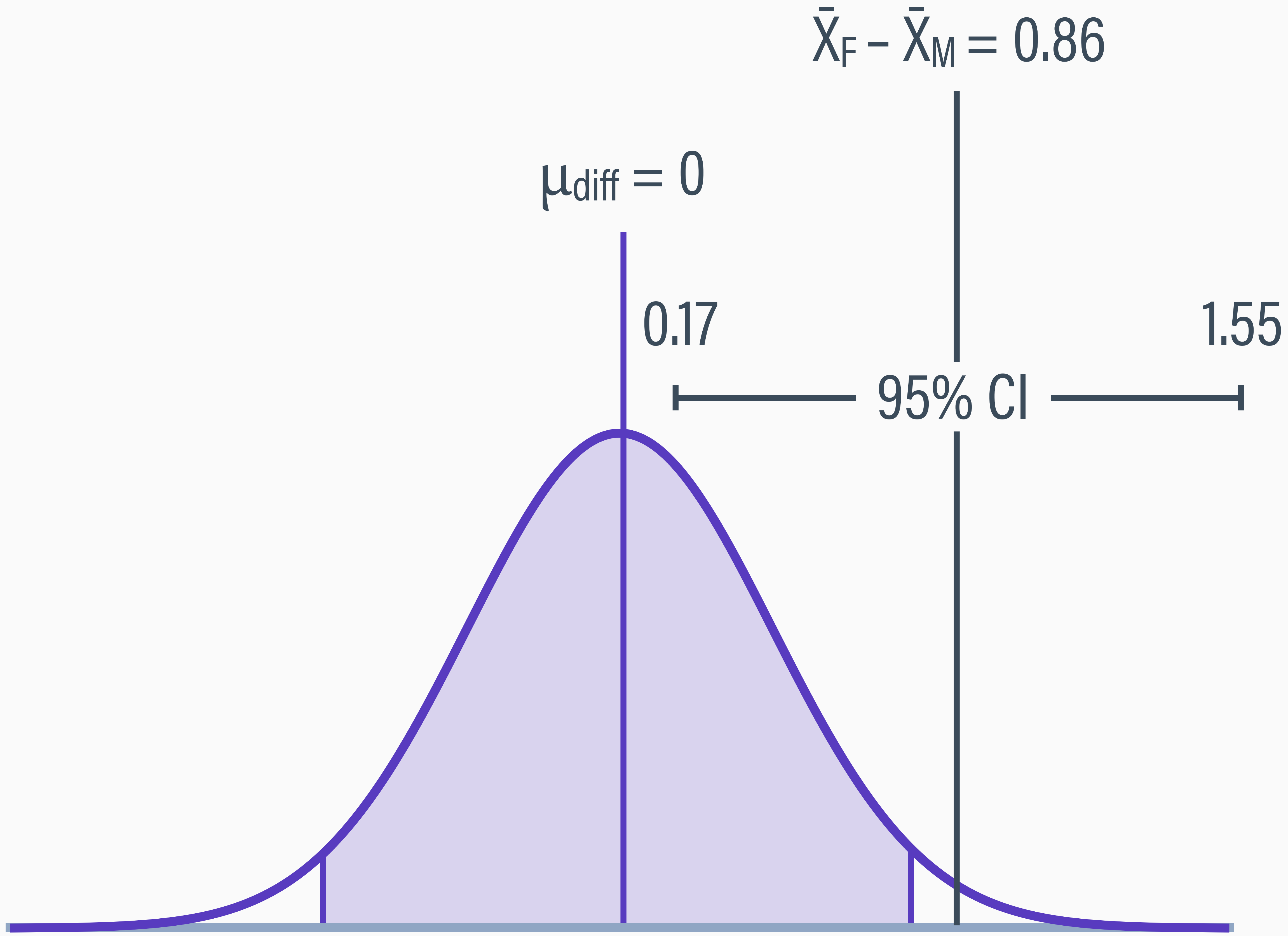
```
    Welch Two Sample t-test

data:  Discrimination by Gender
t = 2.4377, df = 634.16, p-value = 0.01506
alternative hypothesis: true difference in means between
group Female and group Male is not equal to 0

95 percent confidence interval:
 0.1667374 1.5483659

sample estimates:
mean in group Female   mean in group Male
            13.96755             13.11000
```

The study produced a mean difference and 95% confidence interval of $\bar{X}_{diff} = 0.86$ and $CI_{95\%} = [0.17, 1.55]$. In small groups of two or three, discuss whether this sample of $N = 639$ participants could have reasonably originated from a population where there is truly no discrimination difference between males and females ($\mu_{diff} = 0$).

SIGNIFICANCE TESTING WITH 95% INTERVALS

- A population with a mean difference of 0 is unlikely to have produced this sample because the null mean is outside the 95% interval

- The 95% confidence interval provides the same conclusion as a two-tailed significance test with a .05 significance criterion!



SIGNIFICANCE TESTING STEPS

1. Specify hypotheses
2. Define standard of evidence
3. Design study and collect data
4. Compare data to null hypothesis
5. Evaluate hypotheses and draw conclusion

COMPARING DATA TO THE NULL

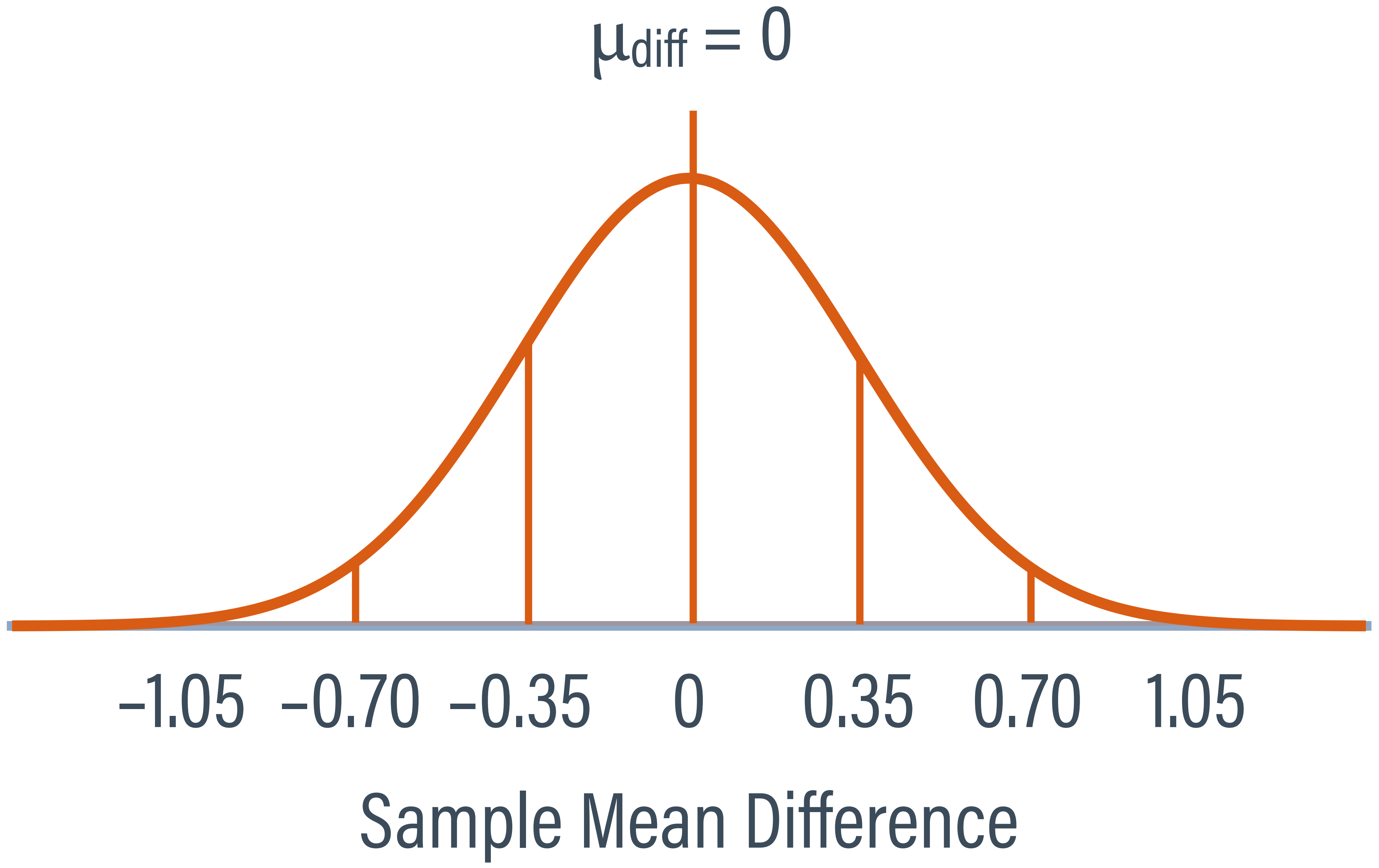
- Two ways to determine whether the sample $\bar{X}_{diff}$ is consistent (or inconsistent) with the null population mean
- The t-statistic gives a standardized distance between the sample mean and the null hypothesis mean (like a z-score)
- A p-value tells us how likely it is that hypothetical samples like our data would originate from the null population

STANDARD ERROR OF A MEAN DIFFERENCE

- The standard error gives the expected (average) sampling error in a mean difference statistic across many hypothetical samples

$$s_{\bar{X}_{diff}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{4.57^2}{339} + \frac{4.32^2}{300}} = 0.35$$

- Across many hypothetical samples from a null population, we would expect the gender mean difference to be ±0.35 perceived discrimination points from $\mu_{diff} = 0$



Consider the sampling distribution of sample means from a null population with $\mu_{diff} = 0$. The sample mean difference of $\bar{X}_{diff} = 0.86$ ($s_{\bar{X}_{diff}} = 0.35$). In small groups of two or three, discuss whether the data provide evidence for or against the null hypothesis.

t-STATISTIC

- The t-statistic quantifies the number of standard error units that separate the sample mean and null hypothesis population mean

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_{diff}}} = \frac{\text{distance from the null}}{\text{standard error (std. dev. of } \bar{X}_{diff})}$$

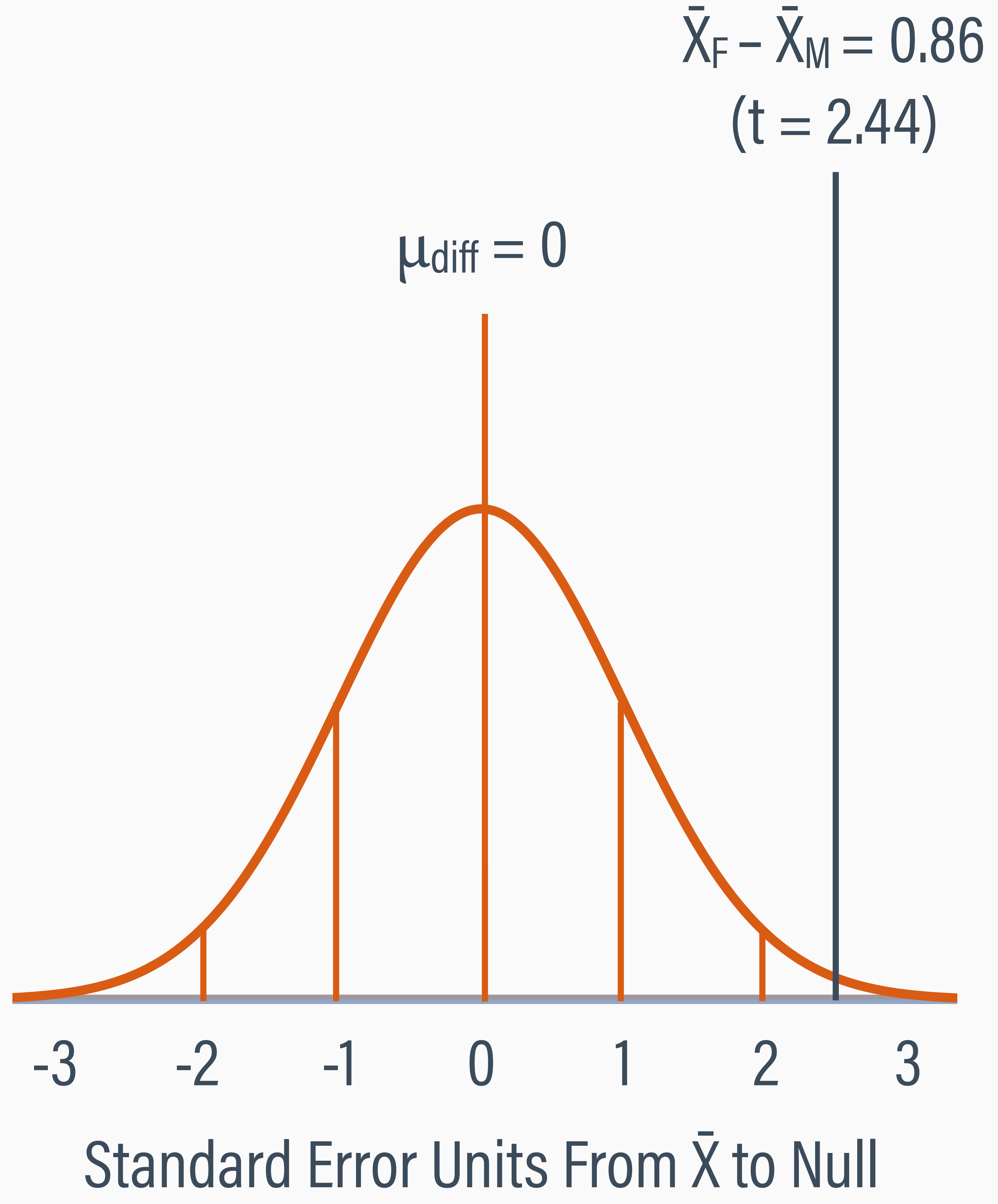
- The t-statistic is the same as a z-score (a standardized metric where distance is expressed in standard deviation units)

t-STATISTIC EXAMPLE

- The t-statistic indicates that 2.44 standard error units separate the sample mean and null

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_{diff}}} = \frac{(13.97 - 13.11) - 0}{0.35} = 2.44$$

- The positive sign is a result of subtracting the lower mean from the higher mean

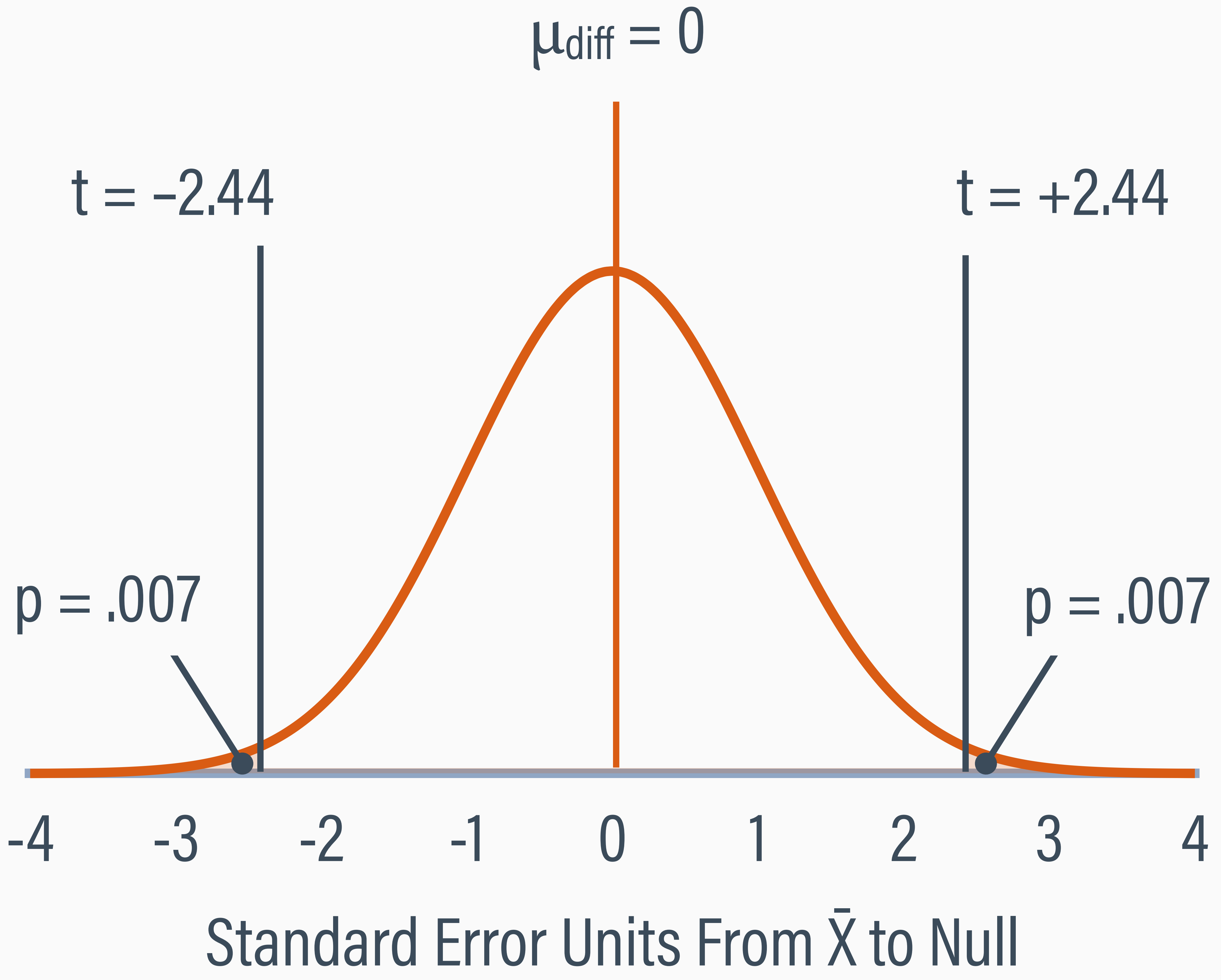


PROBABILITY VALUES (P-VALUES)

- A p-value is defined as proportion of hypothetical samples that have a t-statistic at least as large as the sample data
- Assuming the null is true, how likely is it to draw a sample with an effect at least as large as the one from our data?
- Visually, probability is an area under the curve, obtained by applying calculus integrals to the t-distribution function

TWO-TAILED P-VALUE

- The p-value tells how likely it is to draw a sample mean difference at least as extreme as ours from a null population with $\mu_{diff} = 0$
- The probability of drawing a sample from the null population with a t-statistic of at least ±2.44 is $p = .015$
- 1.5% of all hypothetical samples from a null population would have t-statistics this large



Suppose the researchers had instead specified a one-tailed test where the predicted that females would experience higher levels of perceived discrimination (i.e., they correctly predicted the direction). In small groups of two or three, discuss how the p-value and the conclusions would change with a one-tailed alternate hypothesis.

R OUTPUT

```
   Welch Two Sample t-test

data:  Discrimination by Gender
t = 2.4377, df = 634.16, p-value = 0.01506
alternative hypothesis: true difference in means between
group Female and group Male is not equal to 0

95 percent confidence interval:
 0.1667374 1.5483659

sample estimates:
mean in group Female   mean in group Male
            13.96755             13.11000

standard error of mean difference: 0.35179
```

SIGNIFICANCE TESTING STEPS

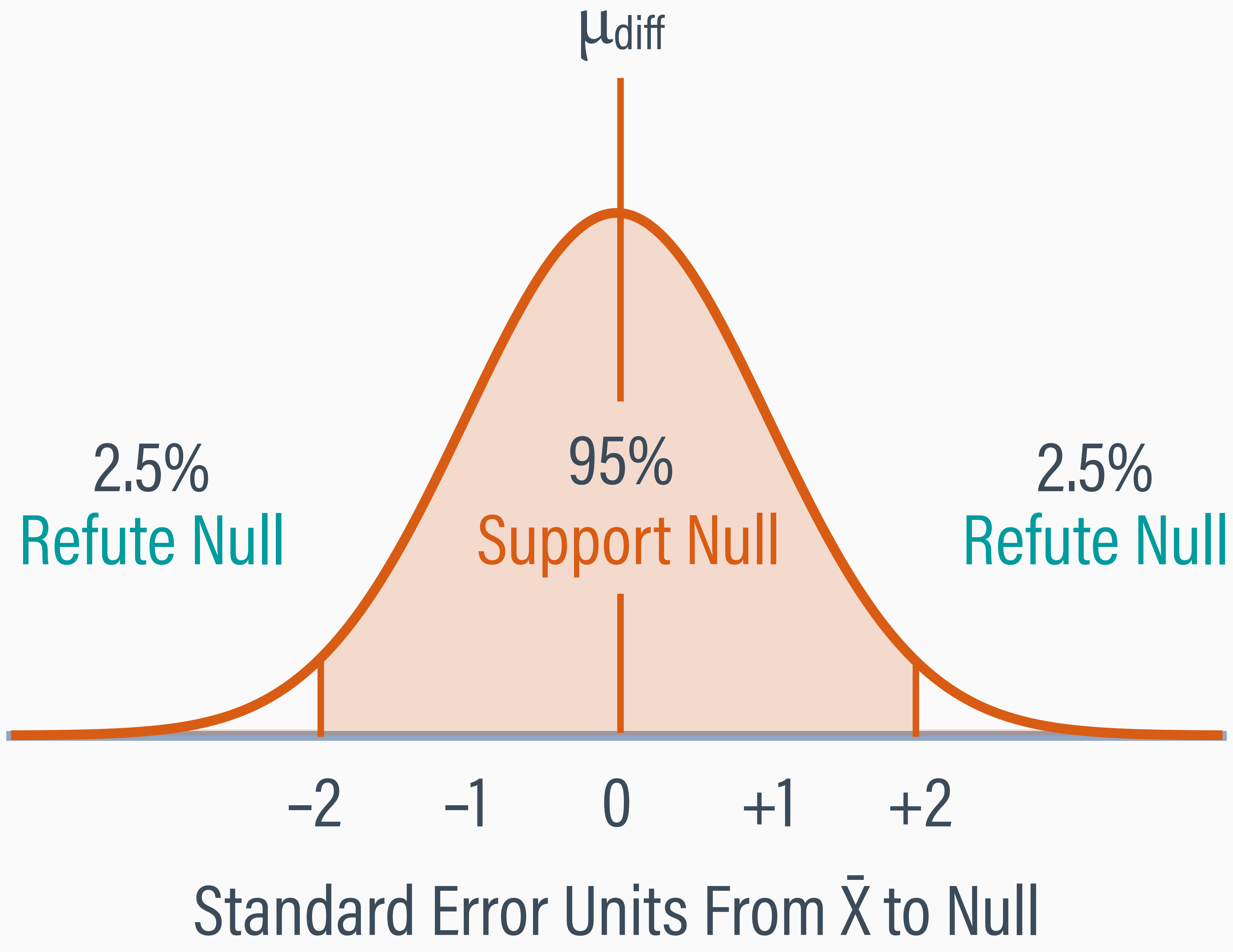
1. Specify hypotheses
2. Define standard of evidence
3. Design study and collect data
4. Compare data to null hypothesis
5. Evaluate hypotheses and draw conclusion

RESEARCH QUESTION REVISITED

- Studies typically attempt to answer a handful of research questions involving associations between key variables
- Question: Do male and female Latinx adolescents experience different levels of perceived discrimination?
- The null (no effect) hypothesis states that the discrimination means are identical (the population mean difference is zero)

5% SIGNIFICANCE CRITERION REVISITED

- By convention, we refute the null if the sample $\bar{X}_{diff}$ falls outside the middle 95% of the sampling distribution ($p < .05$)
- Such as sample has less than a 5% chance of originating from the null population
- We deem the null implausible because our data are unlikely to originate from that population



DECISION TREE

t-Statistic
and probability

$p < .05$

Significant

The data differ from the null

The null is proven "guilty"

There is an effect

Probability of the sample data is $< .05$

$p > .05$

Nonsignificant

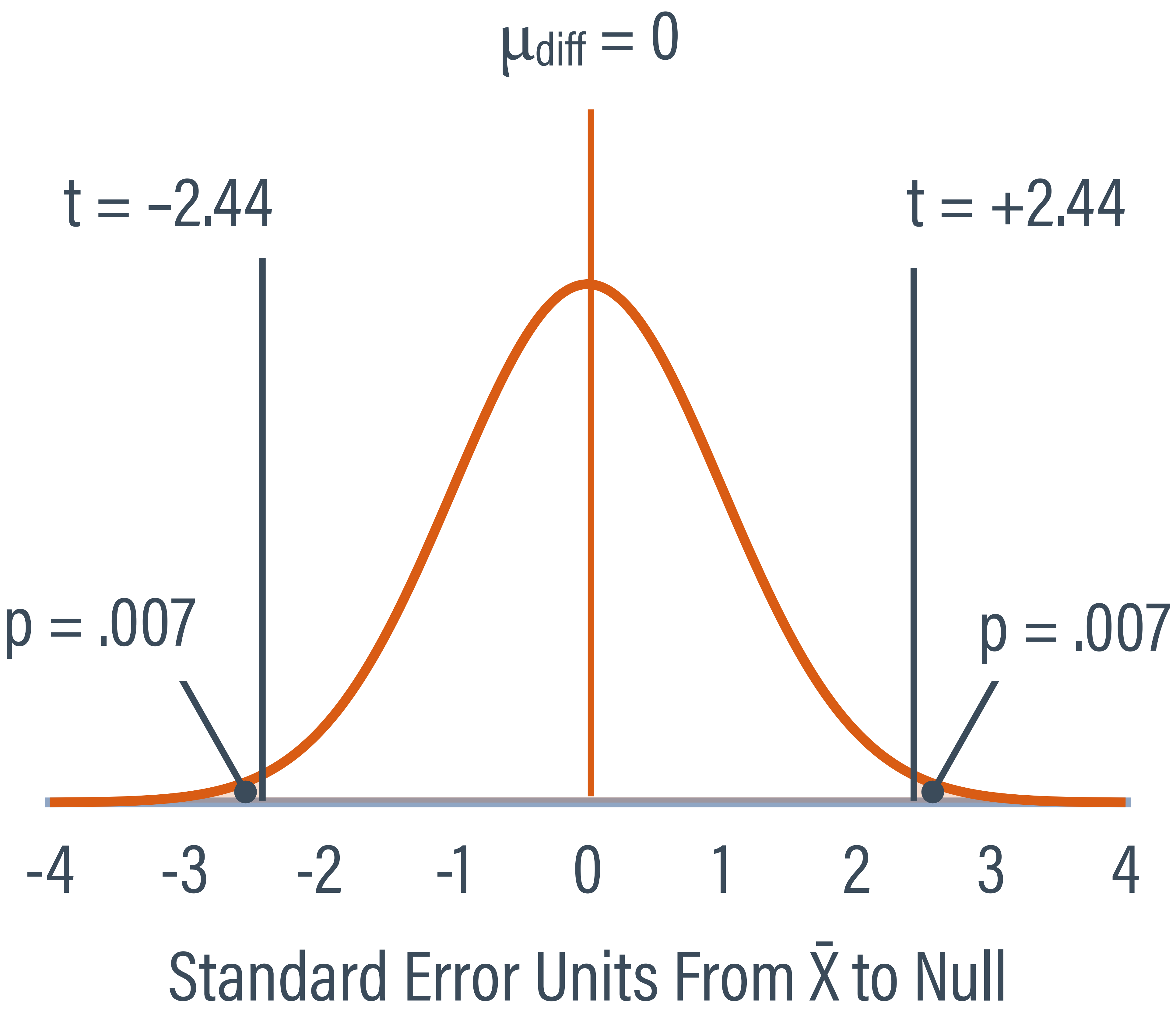
The data are similar to the null

The null is proven "innocent"

There is no effect

Probability of the sample data is $> .05$

The two-tailed probability for the study is p = .015. In small groups of two or three, discuss your decision about the the null hypothesis. Translate your decision into a tangible statement about gender differences in perceived discrimination.



CONCLUSION: TWO-TAILED ALTERNATE

- The p-value of .015 would lead us to refute (disconfirm) the null
- A mean difference as large as $\bar{X}_{diff} = \pm 0.85$ perceived discrimination points (or equivalently, a t-statistic at least ± 2.44) is unlikely to have originated from a null population with $\mu_{diff} = 0$
- The sample mean difference provides evidence that males and females are different in the population

FALSE POSITIVES (TYPE I ERRORS)

- The 5% rejection region is an area of the distribution that contains outlier samples that are unlikely *but not impossible*
- When $\bar{X}_{diff}$ falls in the rejection region (evidence against the null), there is still a 5% chance it came from the null population
- We conclude there is a change, while acknowledging that there is a 5% chance of a false positive—incorrectly rejecting the null when it is actually true (a Type I error)

APA-STYLE ANALYSIS SUMMARY

We used a independent-samples *t*-test to examine whether males and females reported different levels of perceived discrimination. Table 1 gives the descriptive statistics. The mean difference was approximately 0.86 points (females higher), with a 95% confidence interval for the mean difference that ranged from 0.17 (females higher) to 1.55 (females). An independent t-test revealed a significant difference between the two groups, $t(637) = 2.44$ $p = .015$. Finally, the standardized mean difference was just below Cohen's small effect size benchmark ($d = 0.19$), indicating a subtle mean difference.

OUTLINE

1. Between-group designs
2. Quick review
3. Significance testing steps
4. **Statistical assumptions**
5. Study questions

STATISTICAL ASSUMPTIONS

- The accuracy of t-tests (and other statistics) depends on certain conditions in the data being true (e.g., normality)
- Violations of assumptions can bias estimates, inflate or deflate standard errors, and distort significance tests
- Always check reasonableness of assumptions before drawing conclusions

INDEPENDENT T-TEST ASSUMPTIONS

- Numeric (approximately continuous) dependent variable
- Scores are approximately normal in each population
- Independence of observations (no participant's score influences any other participant's score)
- The two populations have equal variances (homogeneity of variance)

WITHIN-GROUP NORMALITY

- In small samples, normality violations can artificially inflate or deflate standard errors, thus distorting significance tests
- Normality is less of a concern if the sample size is large enough (e.g., Ns > 40 to 50)
- Normalizing transformations (e.g., the natural log of scores) are common in some domains

HOMOGENEITY OF VARIANCE

- The sample variances (squared standard deviations) differ by a factor of only 1.1 ($4.57^2 \div 4.32^2 = 1.12$), which is very small
- The classic Student's t-test assumes equal variation
- The Welch t-test we use relaxes this assumption, adjusting both the standard error and degrees of freedom
- Methodological literature shows Welch's test is quite robust to unequal variances and sample sizes

Gender	$\bar{X}$	s	n
Female	13.97	4.57	339
Male	13.11	4.32	300

OUTLINE

1. Between-group designs
2. Quick review
3. Significance testing steps
4. Statistical assumptions
5. Study questions

STUDY QUESTIONS

A researcher wants to determine whether male and female Latinx youth experience different levels of discrimination. To do so, he recruits a sample of n = 339 females and n = 300 males. The sample mean difference is $\bar{X}_{diff} = \bar{X}_{female} - \bar{X}_{male} = 13.97 - 13.11 = +0.86$. To evaluate whether gender differences exist, you will perform significance testing steps assuming a population with a true mean difference of $\mu_{diff} = 0$.

STUDY QUESTIONS (1)

1. State the null hypothesis, both as a sentence and using statistical symbols.

2. State the two-tailed alternate hypothesis, both as a sentence and using statistical symbols.

3. Explain why a independent-samples t-test is the appropriate statistical analysis for this scenario.

STUDY QUESTIONS (2)

4. The sampling distribution under the null hypothesis plays a vital role in hypothesis testing with the independent t-test. Explain how the 5% significance criterion is applied to this distribution, and how it is used to decide whether to reject the null hypothesis.

5. The standard error of the mean difference is $s_{\bar{X}diff} = 0.35$. Explain what the standard error measures. How does it help you gauge whether a mean difference of +0.86 is similar or different from the null?

STUDY QUESTIONS (3)

6. The t-statistic is $t = 2.43$. Explain what the t-statistic measures. What do the sign and the magnitude of the t-statistic indicate about the plausibility of the null hypothesis?

7. Researchers report the results as "statistically significant." What is your decision about the the null hypothesis. Translate your decision into a tangible statement about the difference in male and female discrimination experiences.

STUDY QUESTIONS (4)

8. The two-tailed p-value was .02. Provide an interpretation of the probability value (I am not asking whether the test is significant).

9. The sample mean difference $\bar{X}_{diff} = \bar{X}_{female} - \bar{X}_{male} = 13.97 - 13.11 = +0.86$. The 95% confidence interval limits are $CI_{95\%} = [0.17, 1.55]$. Provide an interpretation of the confidence interval (I am not asking about its statistical properties). Discuss whether the confidence interval supports or refutes the hypothesis that males and females have equal discrimination experiences.