UNIVERSITAS INDONESIA
Veritas, Probitas, Iustitia
FAKULTAS ILMU KOMPUTER

PENGANTAR STATISTIKA

Introduction to Hypothesis Testing

Pengajar:

- Prof. Achmad Nizar Hidayanto
- Ave Adriana Pinem
- Bayu Distiawan
- Ika Chandra Hapsari
- Fathia Prinastiti Sunarso
- Muhammad Misbah
- Nabila Clydea
- Pramitha Dwi Larasati

LEARNING OBJECTIVES

1. Understand logic of hypothesis testing
2. State hypotheses and locate critical region(s)
3. Conduct *z*-test and make decision
4. Define and differentiate Type I and Type II errors
5. Understand effect size and compute Cohen's *d*
6. Make directional hypotheses and conduct one-tailed test

Tools You Will Need

- *z*-Scores
- Distribution of sample means
  - Expected value
  - Standard error
  - Probability and sample means

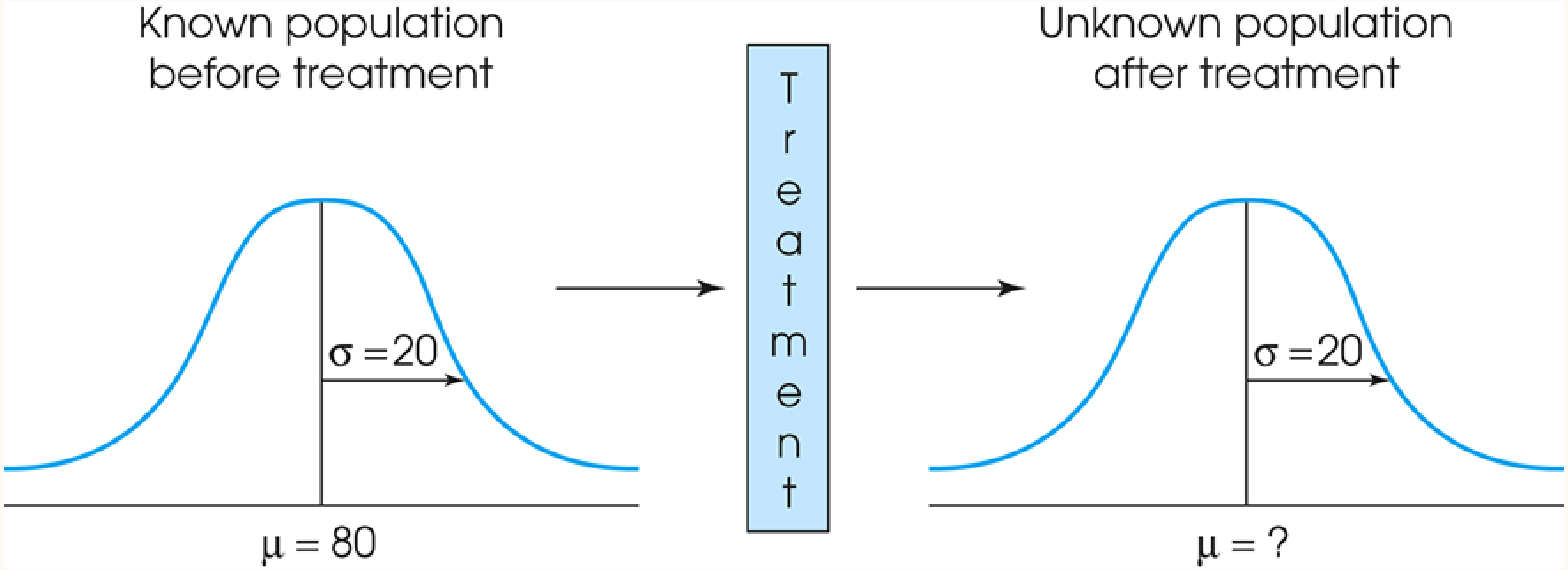
8.1. Hypothesis Testing Logic

- Hypothesis testing is one of the most commonly used inferential procedures
- Definition: a statistical method that uses sample data to evaluate the validity of a hypothesis about a population parameter

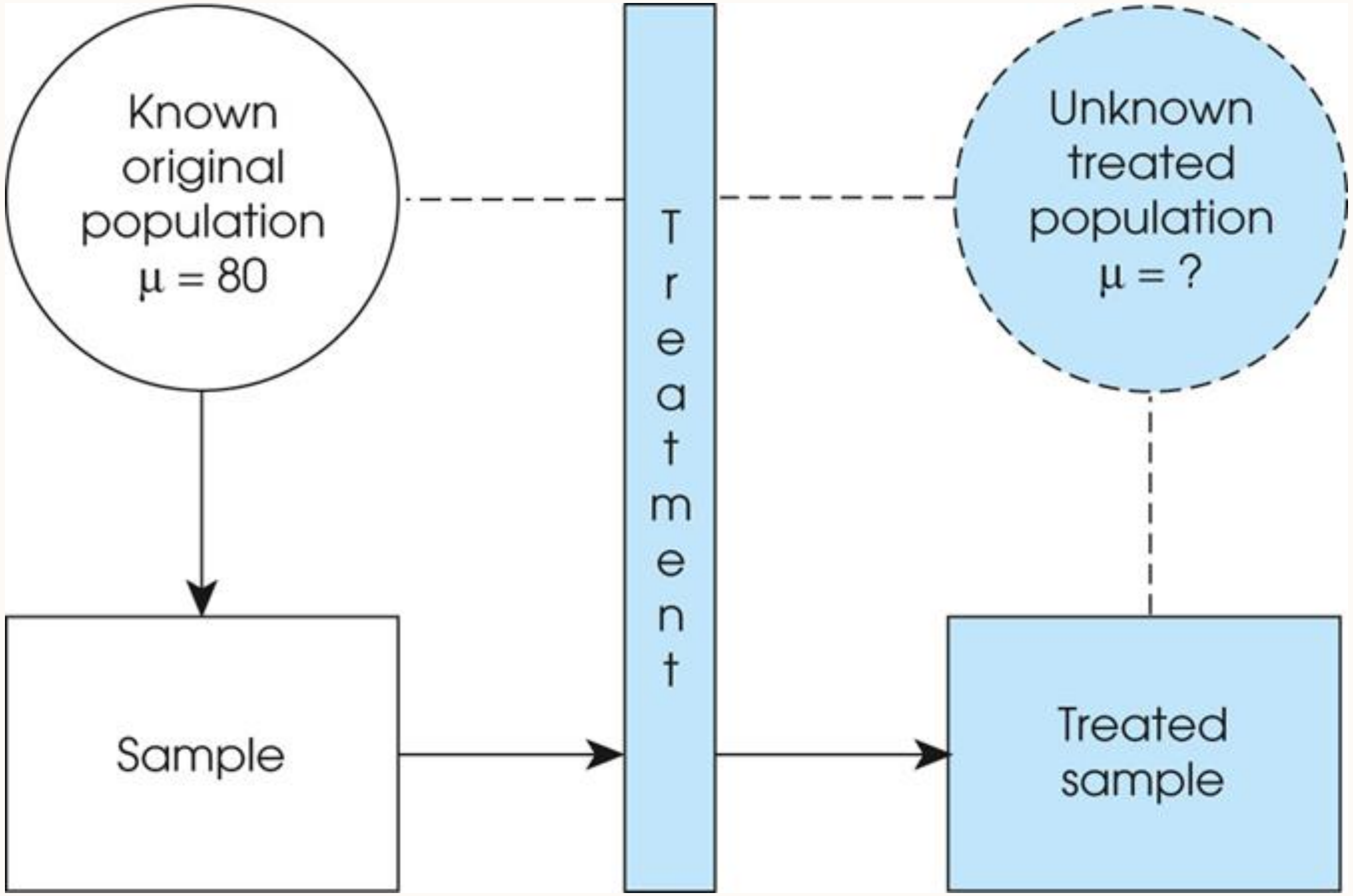
Logic of Hypothesis Test

- State hypothesis about a population
- Predict the *expected characteristics* of the sample based on the hypothesis
- Obtain a random sample from the population
- Compare the obtained sample data with the prediction made from the hypothesis
  - If consistent, hypothesis is reasonable
  - If discrepant, hypothesis is rejected

Basic Experimental



Population in Basic



Four Steps in Hypothesis Testing

Step 1: State the hypotheses

Step 2: Set the criteria for a decision

Step 3: Collect data; compute sample statistics

Step 4: Make a decision

Step 1: State Hypotheses

- Null hypothesis (H_0) states that, in the general population, there is no change, no difference, or is no relationship
- Alternative hypothesis (H_1) states that there is a change, a difference, or there is a relationship in the general population

Step 2: Set the Decision Criterion

- Distribution of sample outcomes is divided
  - Those likely if H_0 is true
  - Those “very unlikely” if H_0 is true
- Alpha level, or *significance level*, is a probability value used to define “very unlikely” outcomes
- Critical region(s) consist of the extreme sample outcomes that are “very unlikely”
- Boundaries of critical region(s) are determined by the probability set by the alpha level

Means

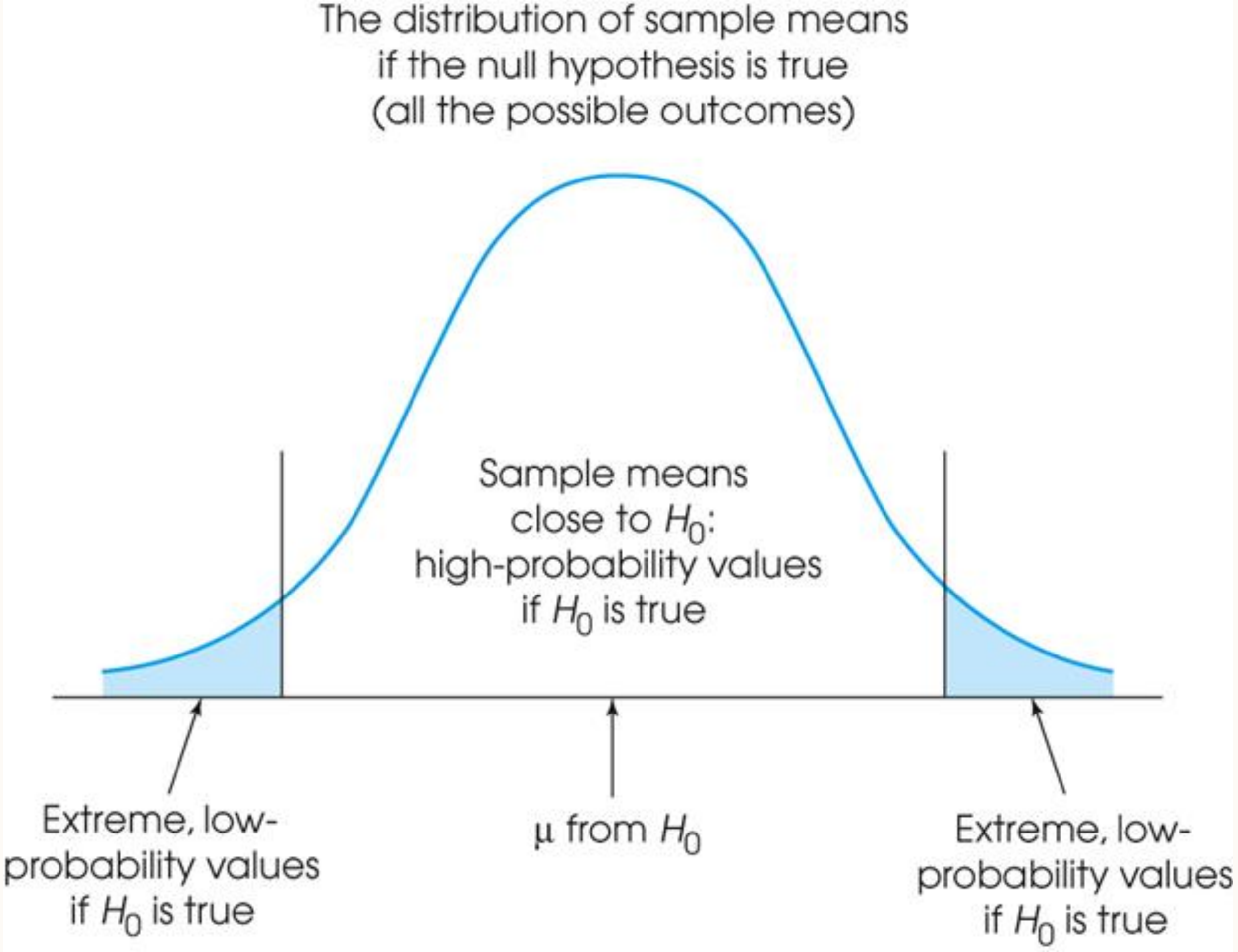


Figure 8.3. "Unlikely" Parts of Distribution of Sample

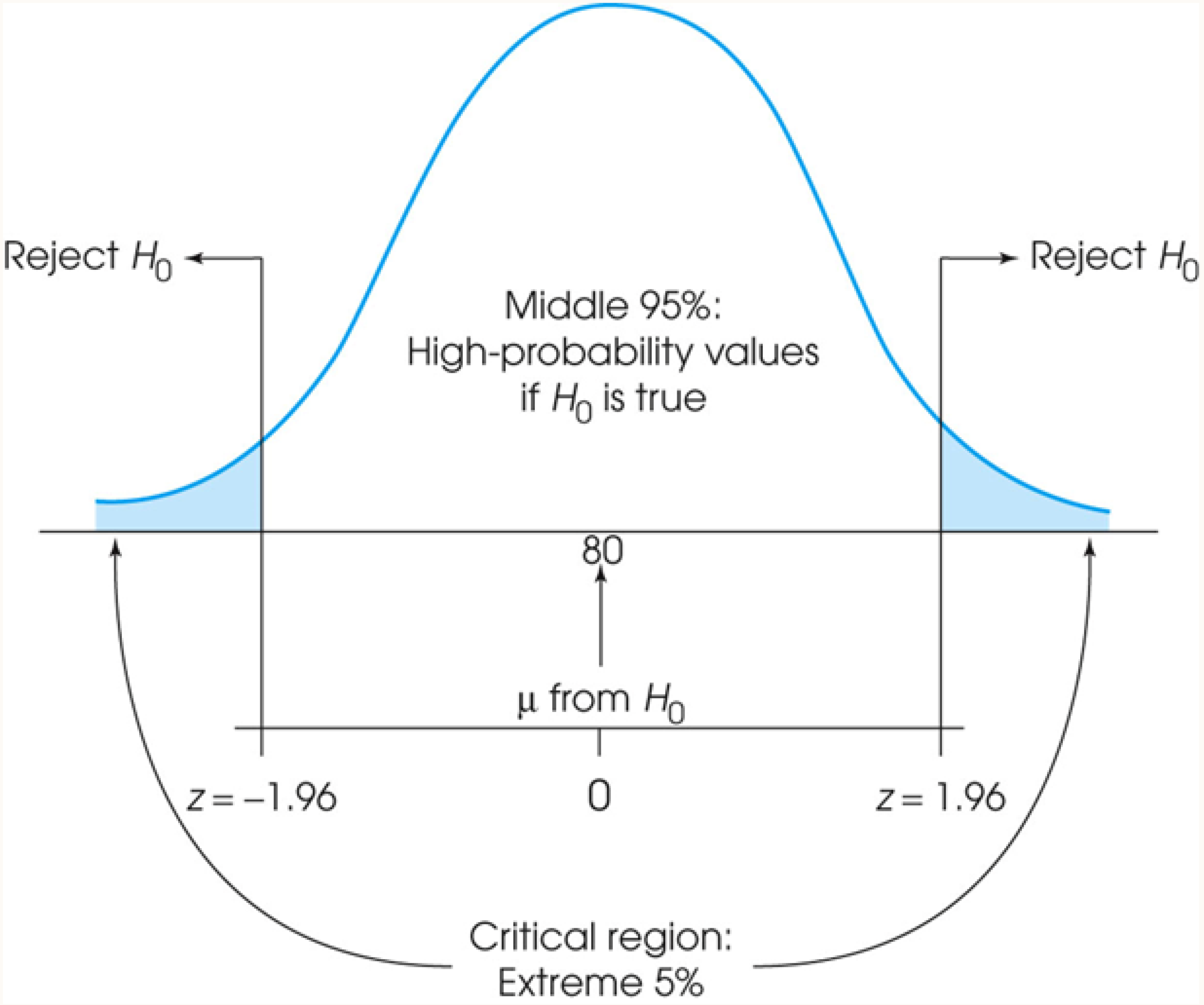


Figure 8.4 Critical region(s) for α = .05

Learning Check

- A sports coach is investigating the impact of a new training method. In words, what would the null hypothesis say?

A
- The new training program produces different results from the existing one

B
- The new training program produces results about like the existing one

C
- The new training program produces better results than the existing one

D
- There is no way to predict the results of the new training program

Learning Check - Answer

- A sports coach is investigating the impact of a new training method. In words, what would the null hypothesis say?

A
- The new training program produces different results from the existing one

B
- The new training program produces results about like the existing one

C
- The new training program produces better results than the existing one

D
- There is no way to predict the results of the new training program

Learning Check

- Decide if each of the following statements is True or False.

T/F

- If the alpha level is decreased, the size of the critical region decreases

T/F

- The critical region defines unlikely values if the null hypothesis is true

Learning Check - Answers

True

- Alpha is the proportion of the area in the critical region(s)

True

- This is the definition of “unlikely”

Step 3: Collect Data (and...)

- Data *always* collected after hypotheses stated
- Data *always* collected after establishing decision criteria
- This sequence assures objectivity

Step 3: (continued)…
Compute Sample Statistics

- Compute a sample statistic (z-score) to show the exact position of the sample

$$z = \frac{M - \mu}{\sigma_M}$$

- In words, z is the difference between the observed sample mean and the hypothesized population mean divided by the standard error of the mean

Step 4: Make a decision

- If sample statistic (z) is located in the critical region, the null hypothesis is rejected

- If the sample statistic (z) is not located in the critical region, the researcher fails to reject the null hypothesis

Jury Trial:
Hypothesis Testing Analogy

- Trial begins with the null hypothesis "not guilty" (defendant's innocent plea)
- Police and prosecutor gather evidence (data) relevant to the validity of the innocent plea
- With *sufficient* evidence against, jury rejects null hypothesis innocence claim to conclude "guilty"
- With *insufficient* evidence against, jury fails to convict, i.e., fails to reject the "not guilty" claim (but does not conclude defendant is innocent)

Learning Check

- Decide if each of the following statements is True or False.

T/F

- When the z-score is quite extreme, it shows the null hypothesis is true

T/F

- A decision to retain the null hypothesis means you proved that the treatment has no effect

Learning Check - Answer

False

- An extreme *z*-score is in the critical region—very unlikely if H_0 is true

False

- Failing to reject H_0 does not prove it true; there is just not enough evidence to reject it

Uncertainty and Errors in Hypothesis Testing

- Hypothesis testing is an inferential process
  - Uses limited information from a sample to make a statistical decision, and then from it a general conclusion
  - Sample data used to make the statistical decision allows us to make an inference and draw a conclusion about a population
- Errors are possible

Type I Errors

- Researcher rejects a null hypothesis that is actually *true*
- Researcher concludes that a treatment has an effect *when it has none*
- Alpha level is the probability that a test will lead to a Type I error

Type II Errors

- Researcher fails to reject a null hypothesis that is really false
- Researcher has failed to detect a real treatment effect
- Type II error probability is not easily identified

Table 8.1

		Actual Situation	
		No Effect = H_0 True	***Effect Exists = H_0 False***
Researcher's Decision	***Reject H_0***	Type I error (α)	Decision correct
	Fail to reject H_0	Decision correct	Type II error (β)

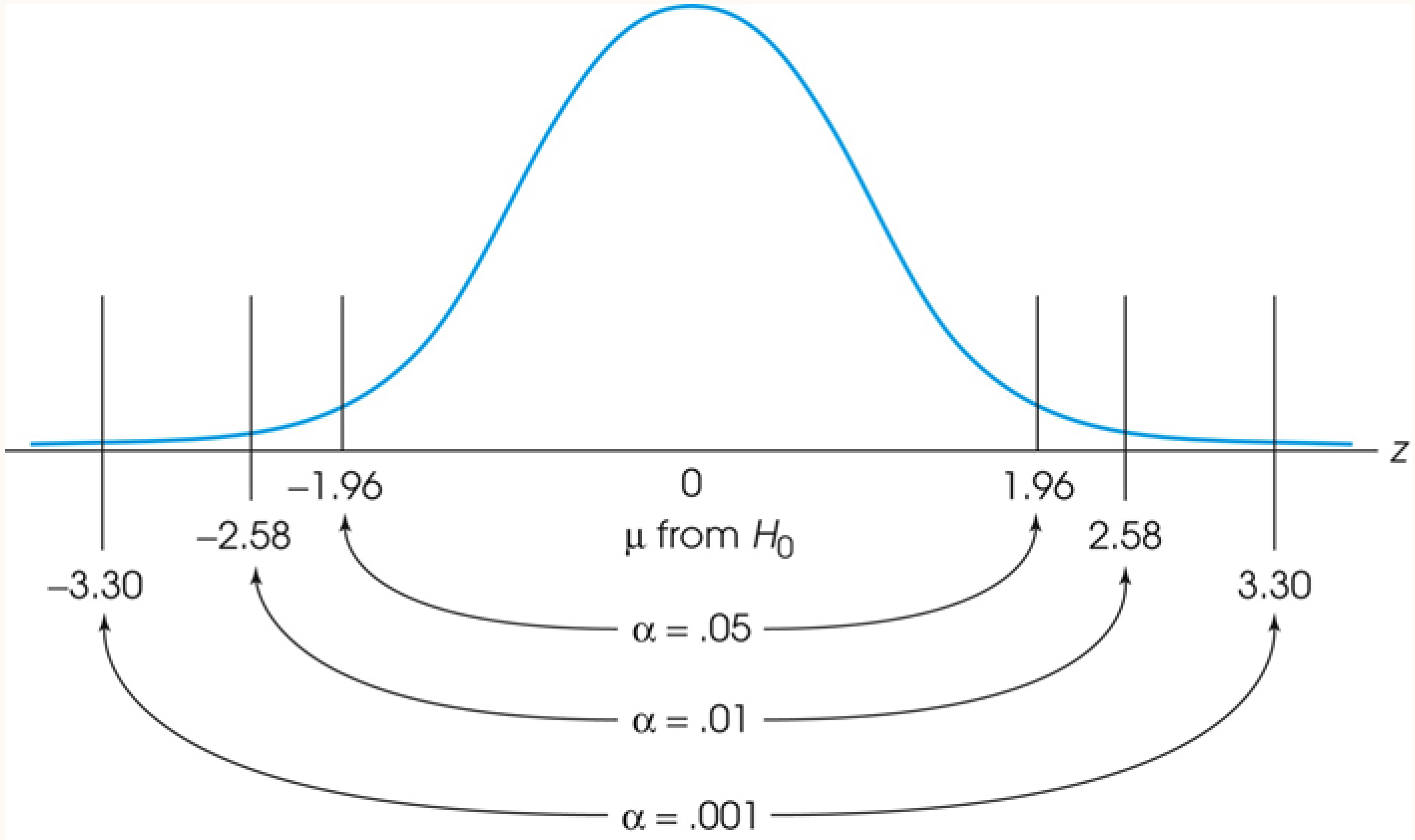


Figure 8.5 Location of Critical Region Boundaries

Learning Check

- Decide if each of the following statements is True or False.

T/F

- A Type I error is like convicting an innocent person in a jury trial

T/F

- A Type II error is like convicting a guilty person in a jury trial

Learning Check - Answer

True

- Innocence is the “null hypothesis” for a jury trial; conviction is like rejecting that hypothesis

False

- Convicting a guilty person is not an error; but acquitting a guilty person would be like Type II error

8.3 Hypothesis Testing Summary

- Step 1: State hypotheses and select alpha level
- Step 2: Locate the critical region
- Step 3: Collect data; compute the test statistic
- Step 4: Make a probability-based decision about H_0: Reject H_0 if the test statistic is unlikely when H_0 is true—called a "significant" or "statistically significant" result

In the Literature

- A result is *significant* or *statistically significant* if it is very *unlikely* to occur when the null hypothesis is true; conclusion: reject H_0
- In APA format
    - Report that you found a significant effect
    - Report value of test statistic
    - Report the *p*-value of your test statistic

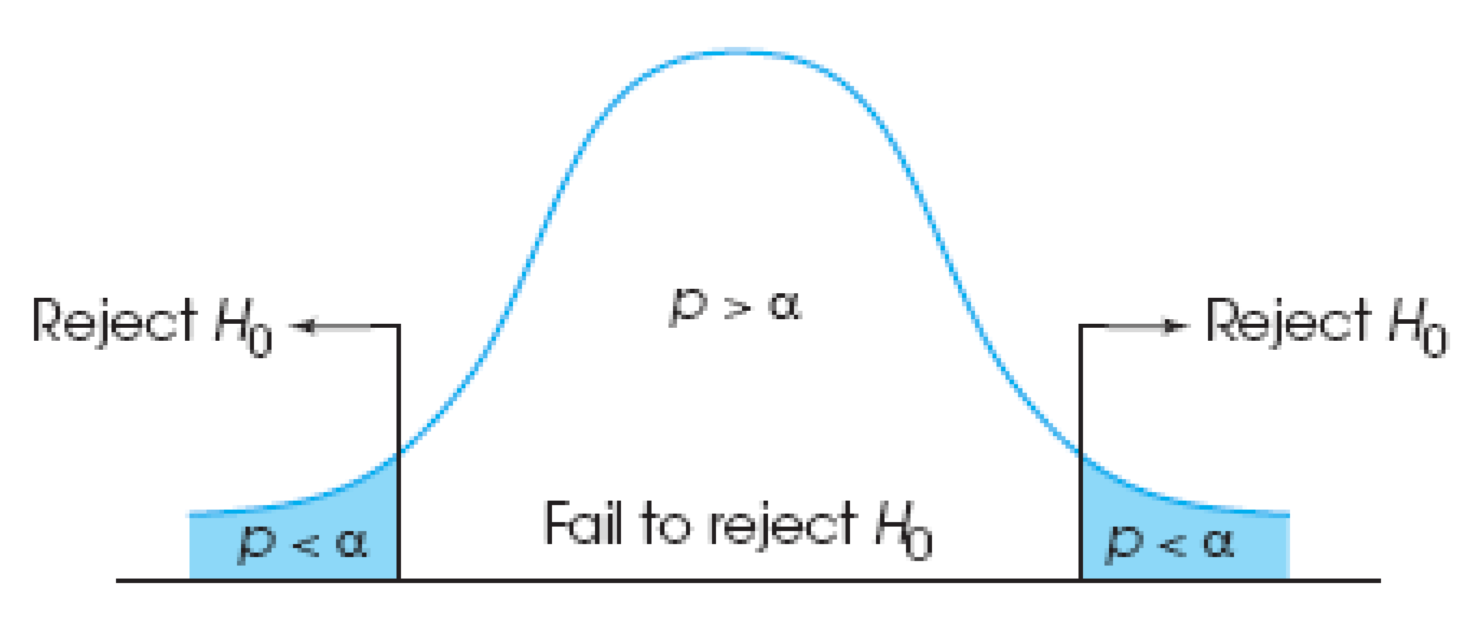


Figure 8.6

Critical Region for Standard Test

8.3 Assumptions for Hypothesis Tests with z-Scores

- Random sampling
- Independent Observation
- Value of σ is not changed by the treatment
- Normally distributed sampling distribution

Factors that Influence the Outcome of a Hypothesis Test

- Size of difference between sample mean and original population mean
  - Larger discrepancies □ larger z-scores
- Variability of the scores
  - More variability □ larger standard error
- Number of scores in the sample
  - Larger n □ smaller standard error

Learning Check

- A researcher uses a hypothesis test to evaluate H_0: $\mu = 80$. Which combination of factors is most likely to result in rejecting the null hypothesis?

A
- $\sigma = 5$ and $n = 25$

B
- $\sigma = 5$ and $n = 50$

C
- $\sigma = 10$ and $n = 25$

D
- $\sigma = 10$ and $n = 50$

Learning Check - Answer

- A researcher uses a hypothesis test to evaluate H_0: $\mu = 80$. Which combination of factors is most likely to result in rejecting the null hypo

A
- $\sigma = 5$ and $n = 25$

B
- $\sigma = 5$ and $n = 50$

C
- $\sigma = 10$ and $n = 25$

D
- $\sigma = 10$ and $n = 50$

Learning Check

- Decide if each of the following statements is True or False.

T/F

- An effect that exists is more likely to be detected if n is large

T/F

- An effect that exists is less likely to be detected if σ is large

Learning Check - Answers

True

- A larger sample produces a smaller standard error and larger *z*

True

- A larger standard deviation increases the standard error and produces a smaller *z*

8.4 Directional Hypothesis Tests

- The standard hypothesis testing procedure is called a two-tailed (non-directional) test because the critical region involves both tails to determine if the treatment increases or decreases the target behavior
- However, sometimes the researcher has a *specific prediction about the direction of the treatment*

8.4 Directional Hypothesis Tests (Continued)

- When a *specific direction* of the treatment effect can be predicted, it can be incorporated into the hypotheses
- In a directional (one-tailed) hypothesis test, the researcher specifies *either* an increase *or* a decrease in the population mean as a consequence of the treatment

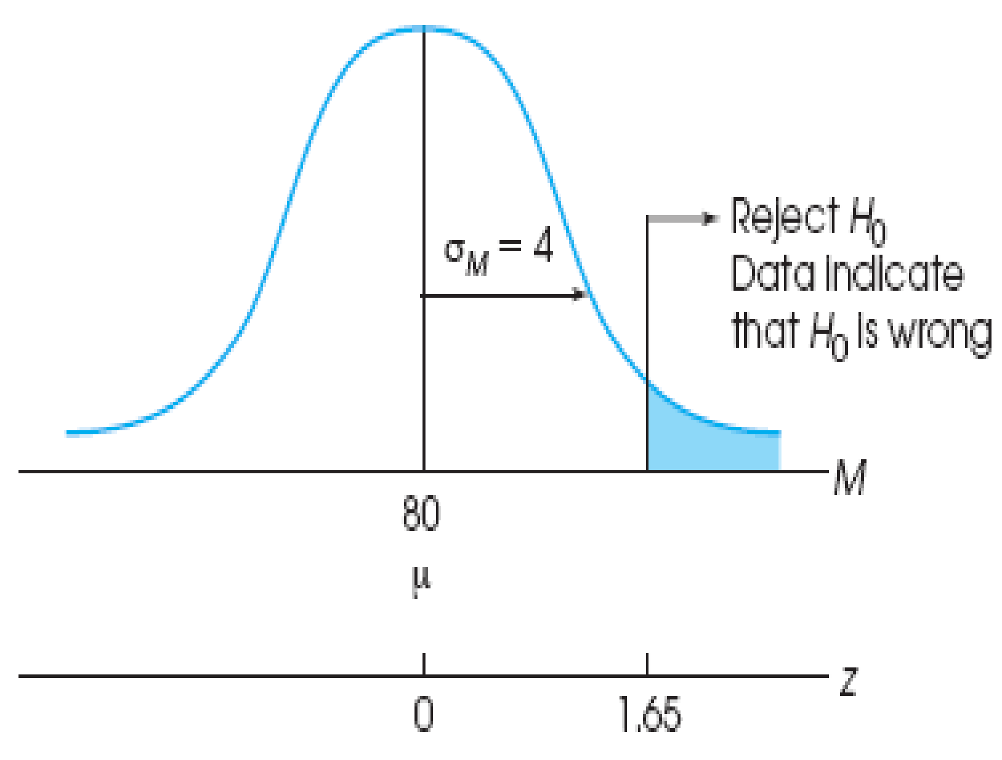


Figure 8.7 Critical Region

One-tailed and Two-tailed Tests Compared

- One-tailed test allows rejecting H_0 with relatively small difference **provided** the difference is in the predicted direction
- Two-tailed test requires relatively large difference regardless of the direction of the difference
- In general two-tailed tests should be used unless there is a strong justification for a directional prediction

Learning Check

- A researcher is predicting that a treatment will decrease scores. If this treatment is evaluated using a directional hypothesis test, then the critical region for the test.

A
- would be entirely in the right-hand tail of the distribution

B
- would be entirely in the left-hand tail of the distribution

C
- would be divided equally between the two tails of the distribution

D
- cannot answer without knowing the value of the alpha level

Learning Check - Answer

- A researcher is predicting that a treatment will decrease scores. If this treatment is evaluated using a directional hypothesis test, then the critical region for the test.

A
- would be entirely in the right-hand tail of the distribution

B
- would be entirely in the left-hand tail of the distribution

C
- would be divided equally between the two tails of the distribution

D
- cannot answer without knowing the value of the alpha level

8.5 Hypothesis Testing Concerns: Measuring Effect Size

- Although commonly used, some researchers are concerned about hypothesis testing
  - Focus of test is data, not hypothesis
  - Significant effects are not always substantial
- Effect size measures the absolute magnitude of a treatment effect, independent of sample size
- Cohen's *d* measures effect size simply and directly in a standardized way

Cohen's *d*: Measure of Effect Size

$$\text{Cohen's } d = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{\mu_{treatment} - \mu_{no\ treatment}}{\sigma}$$

Magnitude of *d*	Evaluation of Effect Size
$d = 0.2$	Small effect
$d = 0.5$	Medium effect
$d = 0.8$	Large effect

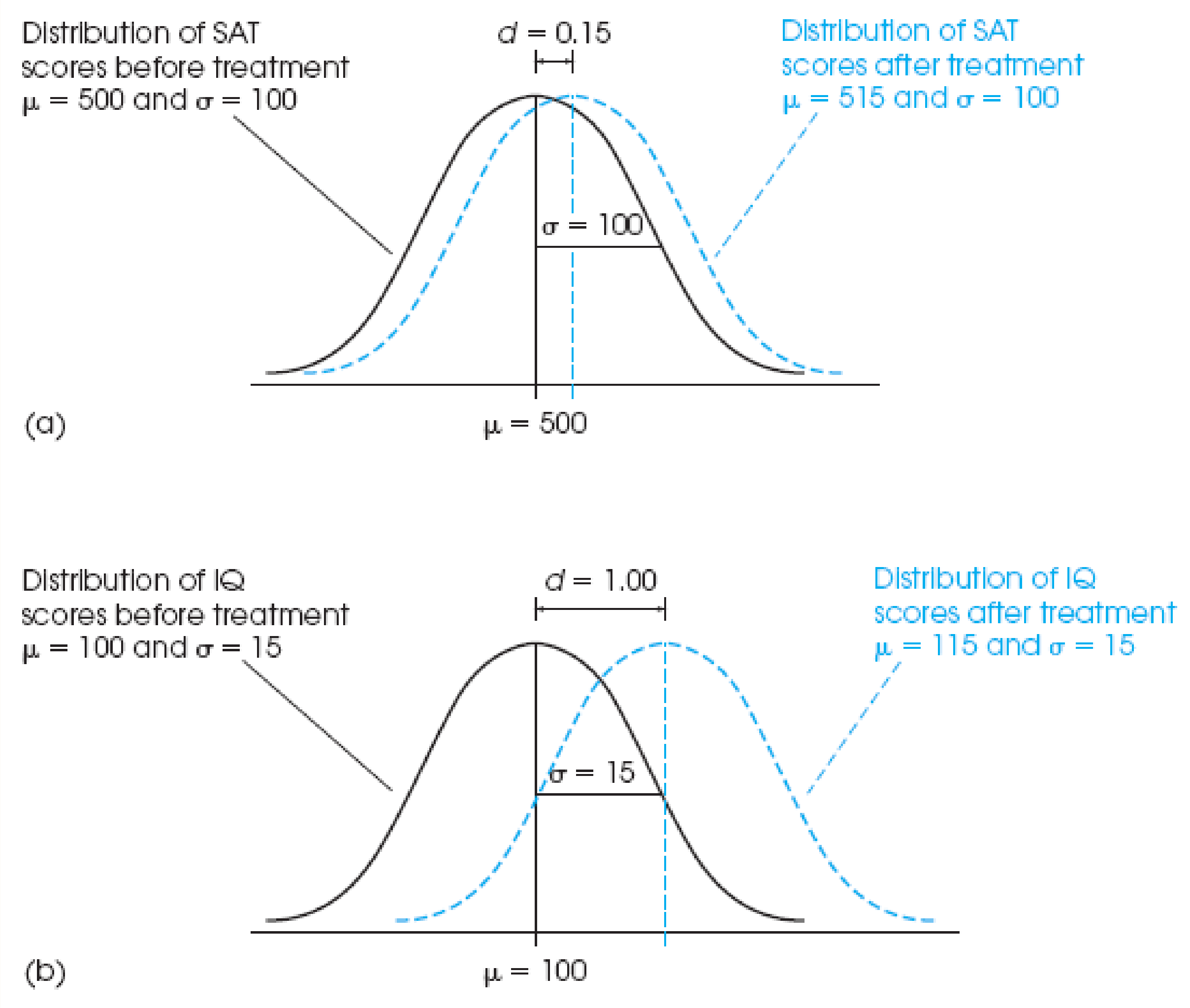


Figure 8.8 When is a 15-point Difference a "Large" Effect?

Learning Check

- Decide if each of the following statements is True or False.

T/F	• Increasing the sample size will also increase the effect size
T/F	• Larger differences between the sample and population mean increase effect size

Learning Check -Answers

False

- Sample size does not affect Cohen's *d*

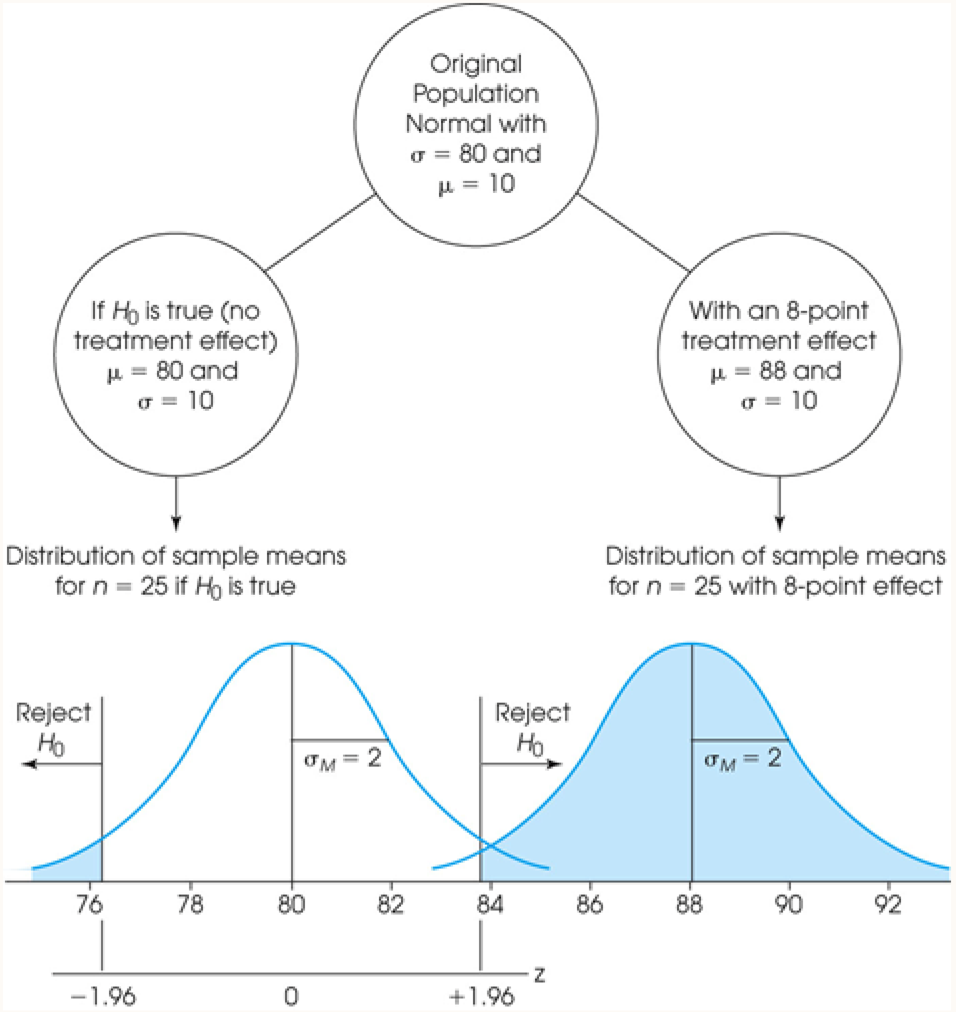
True

- The mean difference is in the numerator of Cohen's *d*

8.6 Statistical Power

- The power of a test is the probability that the test will correctly reject a false null hypothesis
  - It will detect a treatment effect if one exists
  - Power = $1 - \beta$ [where β = probability of a Type II error]
- Power usually estimated before starting study
  - Requires several assumptions about factors that influence power

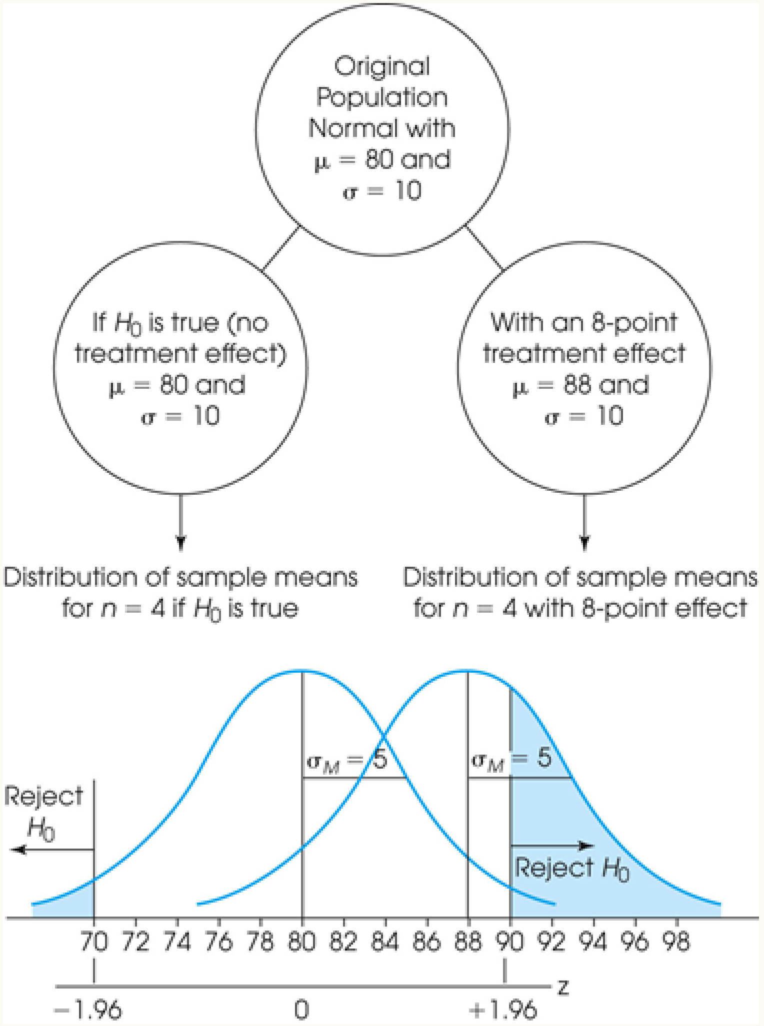
Measuring Statistical



Influences on Power

- Increased Power
  - As effect size increases, power also increases
  - Larger sample sizes produce greater power
  - Using a one-tailed (directional) test increases power (relative to a two-tailed test)
- Decreased Power
  - Reducing the alpha level (making the test more stringent) reduces power
  - Using two-tailed (non-directional) test decreases power (relative to a one-tailed test)

Sample Size Affects



Learning Check

- The power of a statistical test is the probability of _____

A
- rejecting a true null hypothesis

B
- supporting true null hypothesis

C
- rejecting a false null hypothesis

D
- supporting a false null hypothesis

Learning Check - Answer

- The power of a statistical test is the probability of _____

A
- rejecting a true null hypothesis

B
- supporting true null hypothesis

C
- rejecting a false null hypothesis

D
- supporting a false null hypothesis

Learning Check

- Decide if each of the following statements is True or False.

T/F

- Cohen's *d* is used because alone, a hypothesis test does not measure the size of the treatment effect

T/F

- Lowering the alpha level from .05 to .01 will increase the power of a statistical test

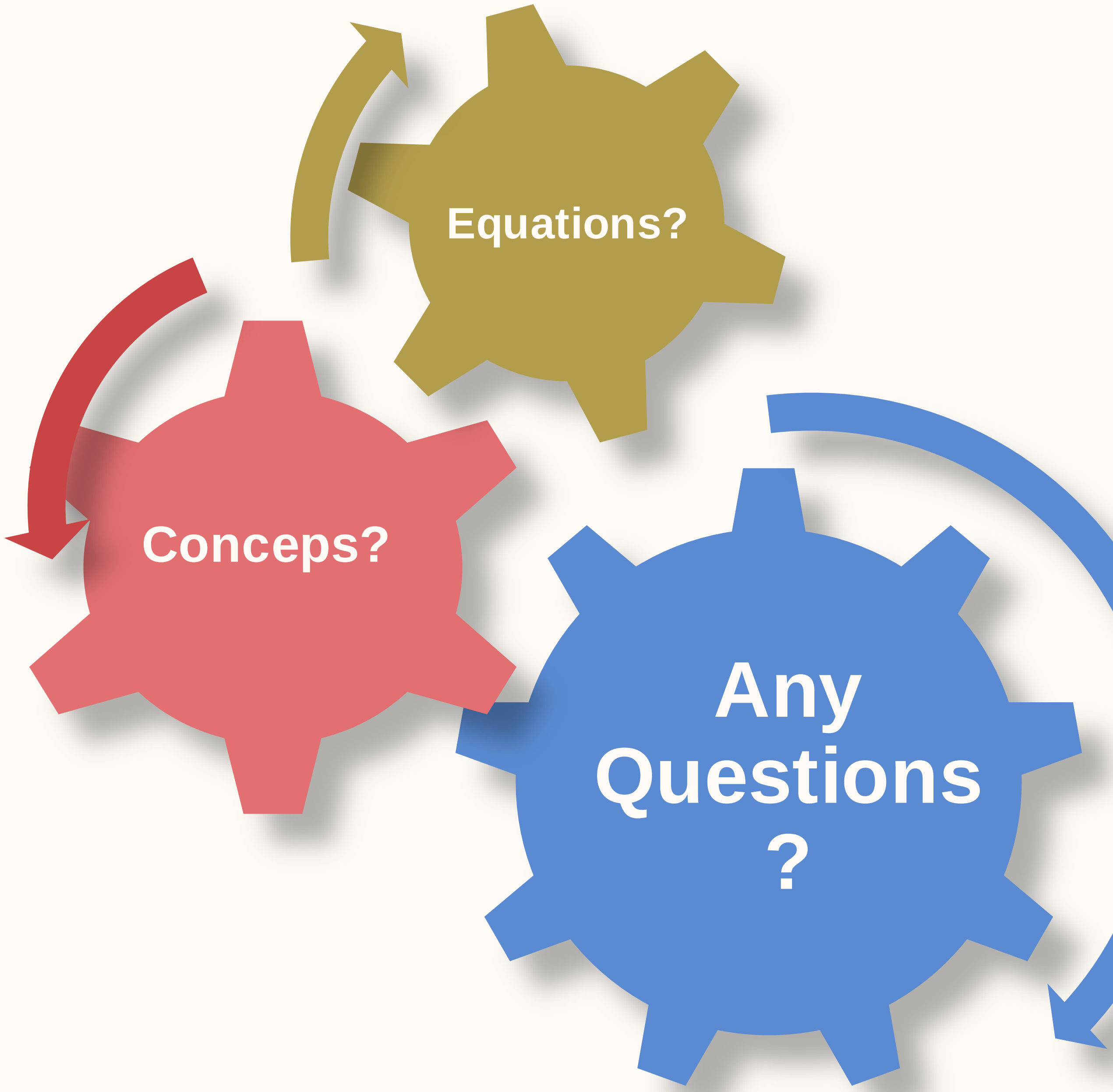
Answer

True

- Differences might be significant but not of substantial size

False

- It is less likely that H_0 will be rejected with a small alpha


Equations?
Conceps?
Any
Questions
?

THANK YOU