Module 2 Lecture - Review of Inferential Statistics, Power, and Assumptions

Analysis of Variance

Quinton Quagliano, M.S., C.S.P

Department of Educational Psychology

Table of Contents

1 Overview and Introduction

1.1 Learning Objectives

- This module will help us introduce some of the problems and strategies that lead us to consider alternative analysis methods, like non-parametric tests, one of the foci of this semester
- Students should be able to:
  - Understand and apply the vocabulary of errors and decision from hypothesis testing
  - Describe how statistical power changes in response to different circumstances
  - Describe common assumptions to check for, and how to check for those

1.2 Introduction

- Last module, we left off talking about ______________ testing, central to how we draw conclusions using ______________ statistics.
- With this module, we'll review ______________ errors, and how we have to navigate possible issues that may inflate those types of errors, i.e., assumption violations
- We'll also cover some common strategies for looking for assumption violation, prior to next module, when we'll work through ______________ and addressing assumption violations

2 Decisions, Type I and Type II Errors, and Power

2.1 Quick Review on Hypothesis Testing

- *This section is review of module 1, please review that section if this terminology is unfamiliar*
- At the end of an inferential test, we will have a ______________ and compare that with a pre-set ______________ value
  - A statistically ______________ result:

    * $p < \alpha$
    * Test statistic is more extreme (i.e., further from 0) than critical value
  - A non-statistically ________ result:
    * $p \geq \alpha$
    * Test statistic is equal to or less extreme than critical value

? Review: A statistically significant result indicates that we should [BLANK] the null hypothesis

- (•) A) Retain
- ( ) B) Reject
- ( ) C) Question
- ( ) D) Review

Explanation:

2.2 Decision-making

- When we gather and ________ the data we gather from a sample, we use those results to make a decision
  - The decision comes down to whether we can ________ the null hypothesis or not, called 'retaining' the null hypothesis
  - Ideally we avoid making a type I or type II ________ with our decision

Reality (Truth)	Decision: Reject H_0	Decision: Fail to Reject H_0
H_0 is True	**Type I Error (α)** False positive	**Correct Decision (1 - β; power)** (True negative)
H_1 is True	**Correct Decision (1 - α)** (True positive)	**Type II Error (β)** False negative

2.3 Types of Errors

- A type I error is when we ________ to reject the null hypothesis, when it is actually true

- The ________ of a Type I error occurring (i.e., $P(TypeI)$) is given as α
- Remember, α is under the assumption that the null hypothesis is ________, and that the ________ hypothesis is false

? What alpha value is most commonly set?

- (•) A) 0.10
- ( ) B) 0.50
- ( ) C) 0.20
- ( ) D) 0.05

Explanation:

- A type II error is when we decide to retain the null hypothesis when it is actually ________
  - The ________ of a Type II error occurring (i.e., $P(TypeII)$) is given as β
  - Opposite to α, β and by extension, power, are based upon the assumption that the null hypothesis is ________ (i.e., ________ hypothesis is true)

! Important

When I was first trained on statistics, I was taught to view the Type I error as being more 'egregious' - 'the number one error to avoid!' - but realistically, they both result in flawed conclusions which could be dangerous

? In a hypothetical experiment when comparing students before and after a new instructional style, I decide I can reject the null hypothesis that there was no change. This is accurate, as there actually was a change. What error occurred here?

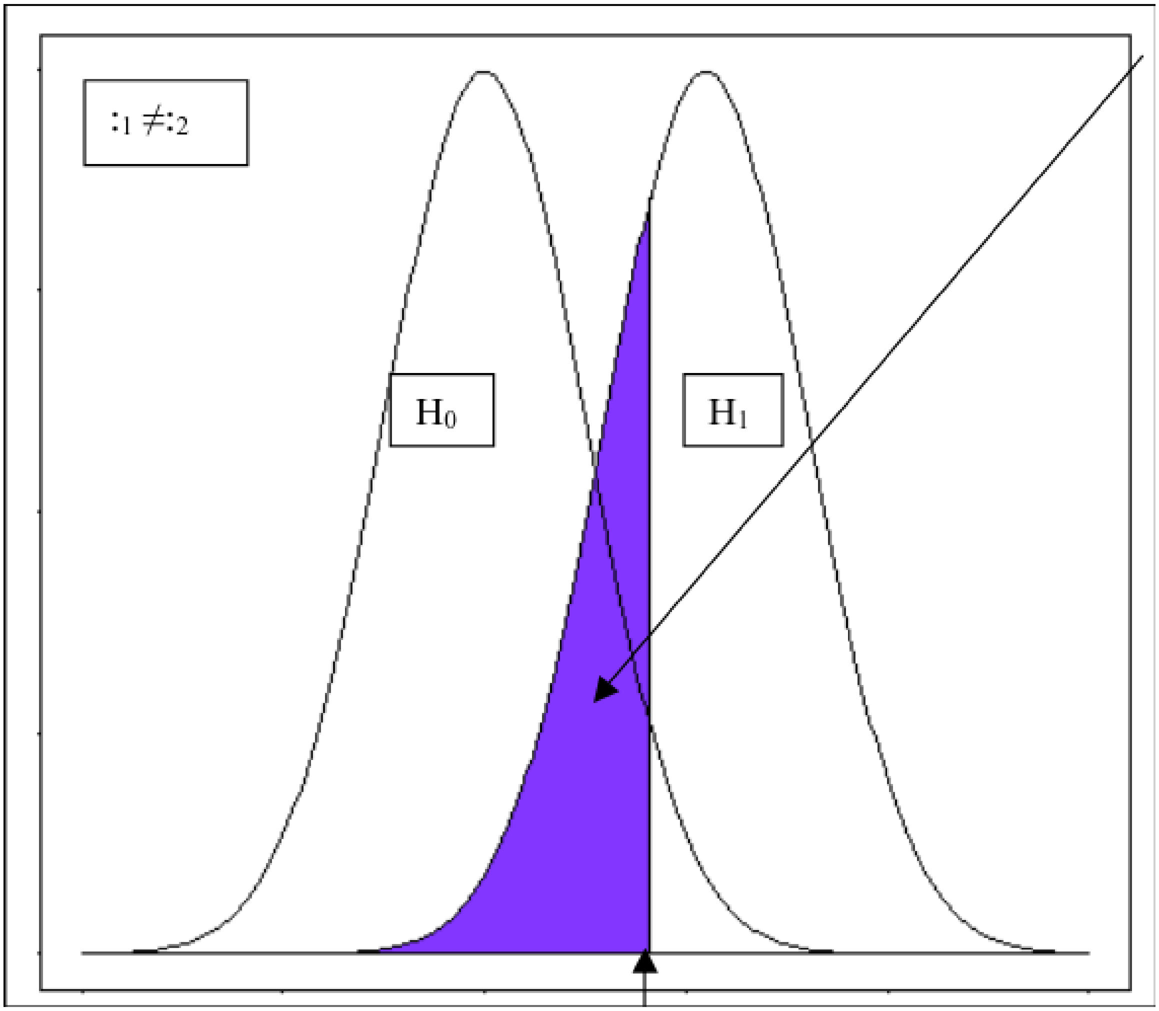
- (•) A) No error occurred
- ( ) B) Type I
- ( ) C) Type II
- ( ) D) Not enough information

Explanation:

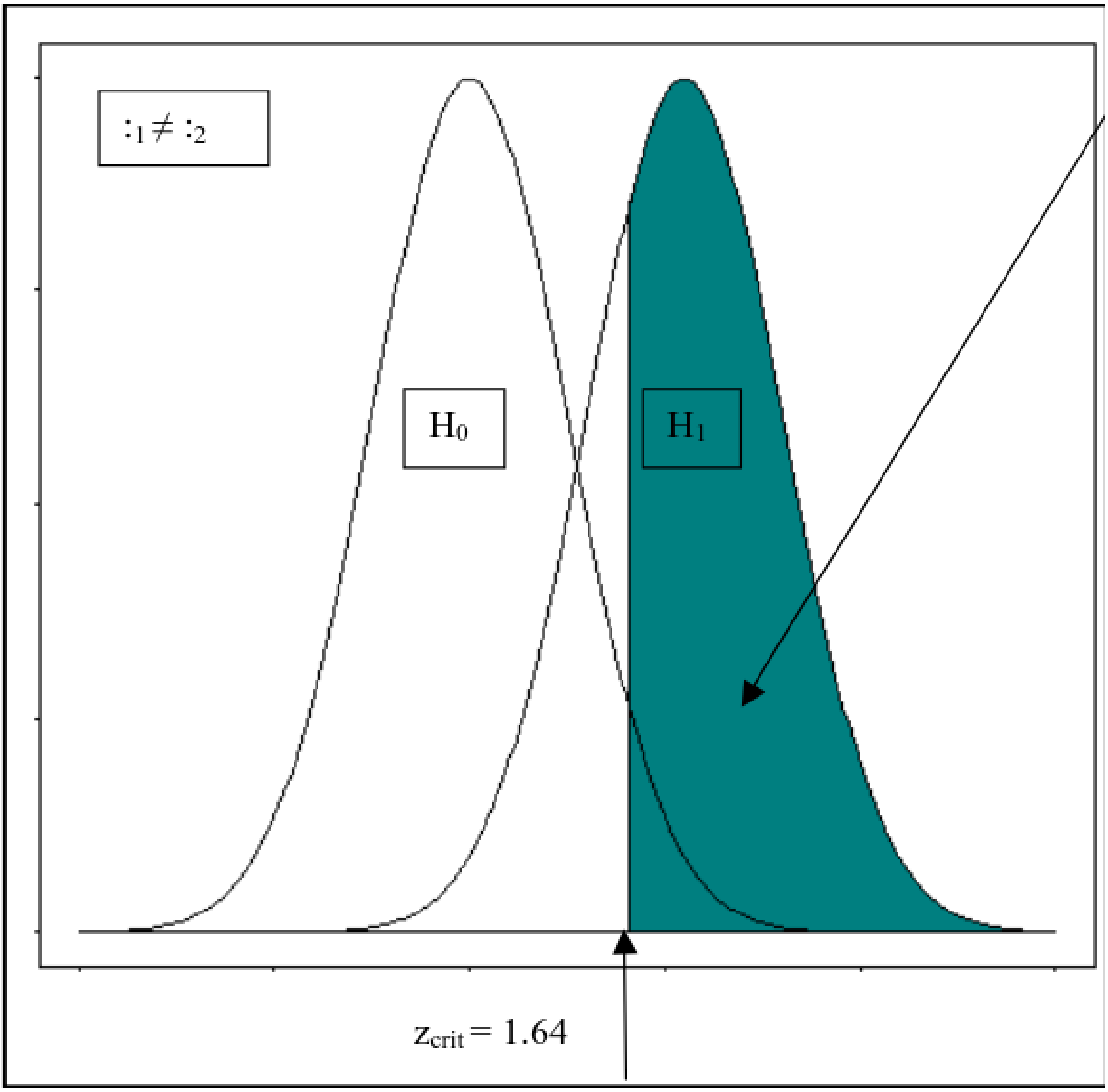
2.4 Power

- Ideally, both α and β should be as ______ as possible, as to try avoid making an error of any sort!
  - An extension of this is that power, the ______ we correctly reject the null when it is false
  - $Power = 1 - \beta$
  - Power is *very* ______ on several characteristics of our analysis:
    * ______ level
    * Sample ______
    * ______ of measures
    * Study ______, i.e., designing stronger studies when expected effect size is small
  - We'll focus on the first two

Discuss: Review: Describe the relationship between confidence level and alpha

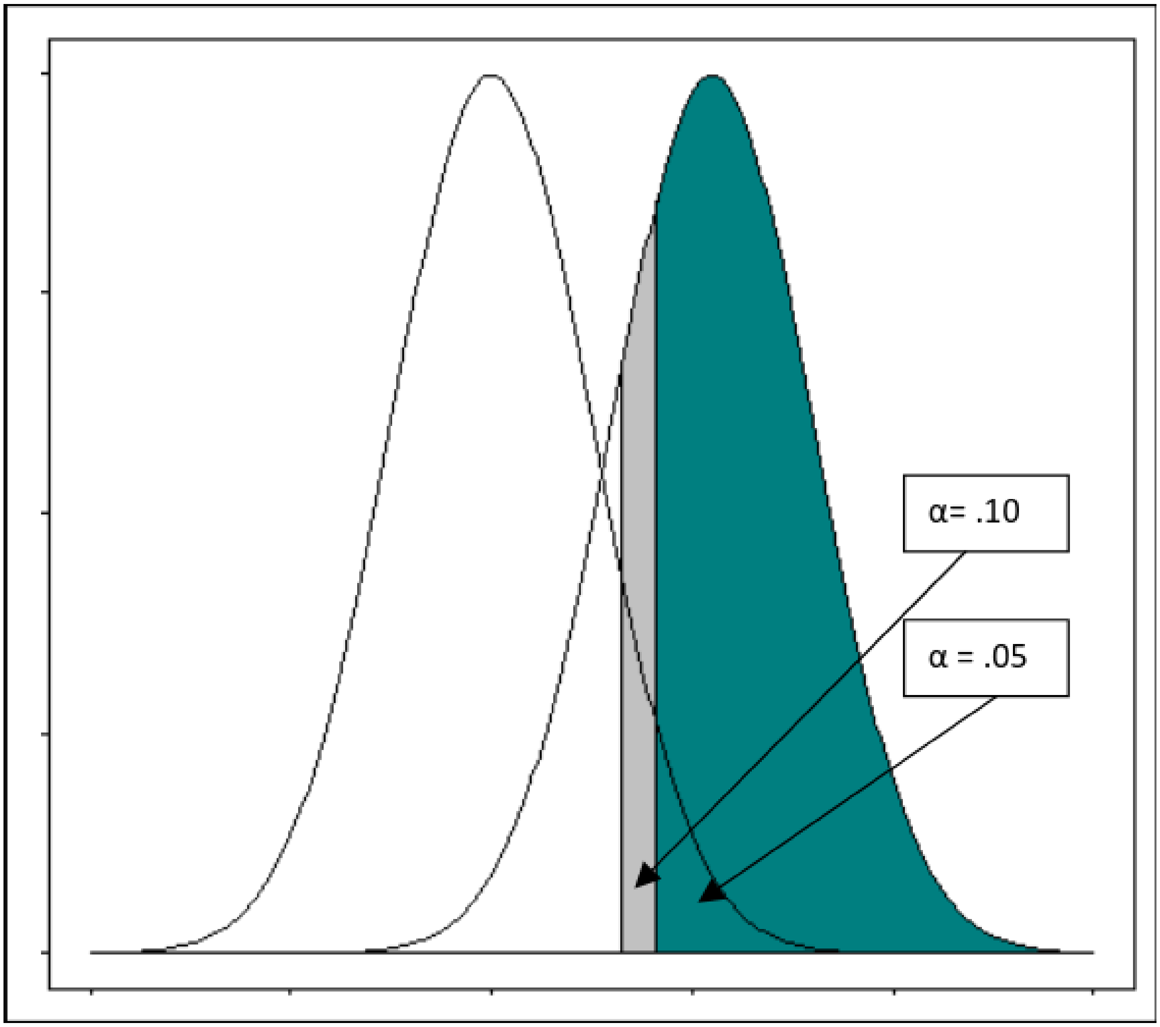


- The purpure area here is ________________ or our probability of failing to reject the null hypothesis when it is false
  - Thus, it refers to the region of the alternative hypothesis curve that is ________________ beyond our critical value.



- The teal area here is ____________, or our probability of rejecting the null hypothesis, when the null hypothesis is false
  - Thus, it refers to the region of the alternative hypothesis curve ____________ the critical value

2.4.1 Alpha level



- ______ α level/confidence level increases the area beyond α, and thus, also increases ______.
  - Recall: We, the researcher, choose our α level, but a higher α also increase Type ______ error rate
  - We must be mindful of possible ______ arising from such a possible Type I error

Discuss: Try to describe a type of research scenario in which a Type I error would be particularly devastating or dangerous

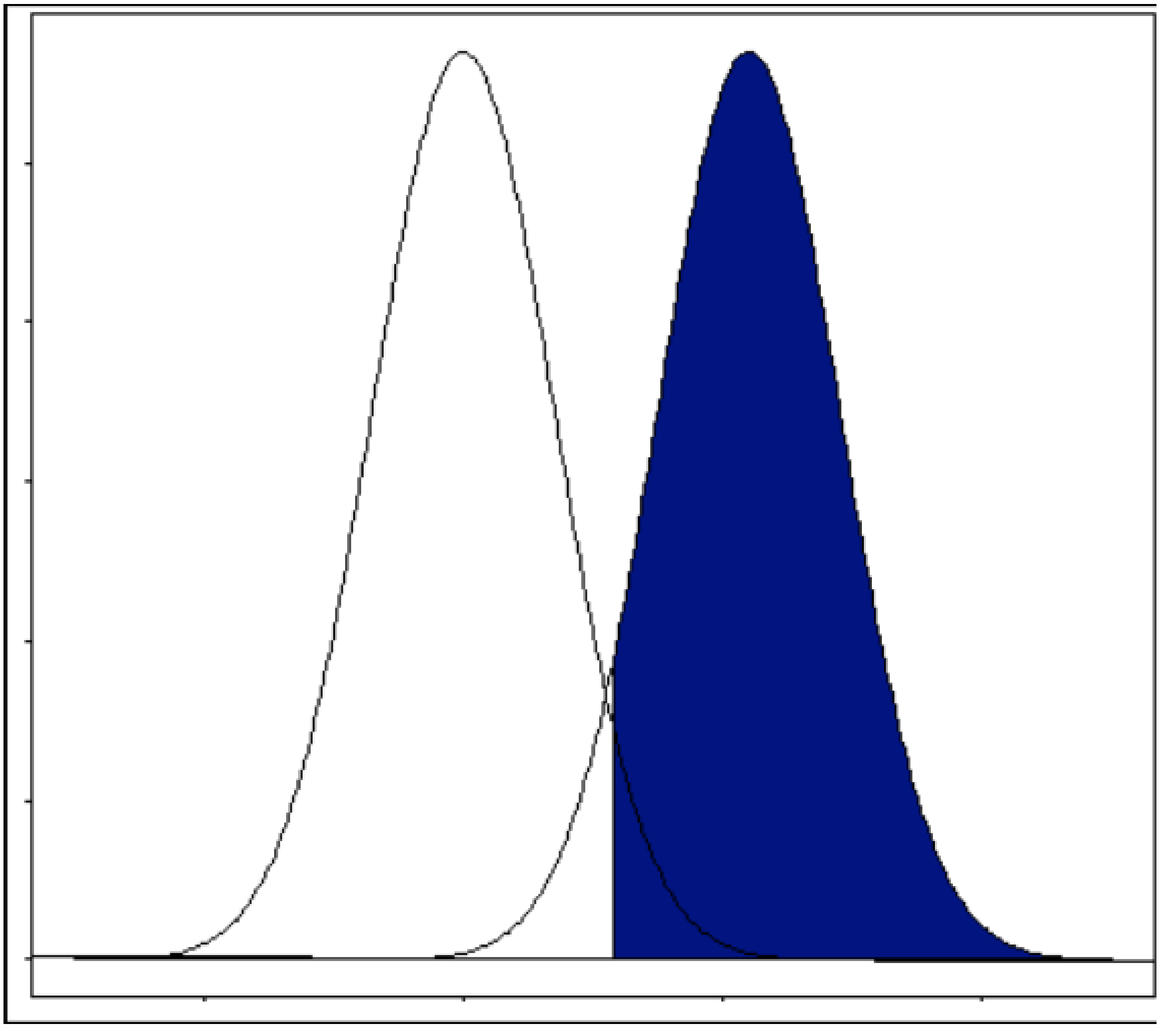
Important

Quick trivia, $\alpha = 0.05$ was originally mentioned by Sir Ronald Fisher (a statistics pioneer) in a letter, and has since become convention. However, there are more modern papers that suggest changing the 'default'

2.4.2 Sample Size

Important

Sample size is, by far, the most common way researchers improve power.

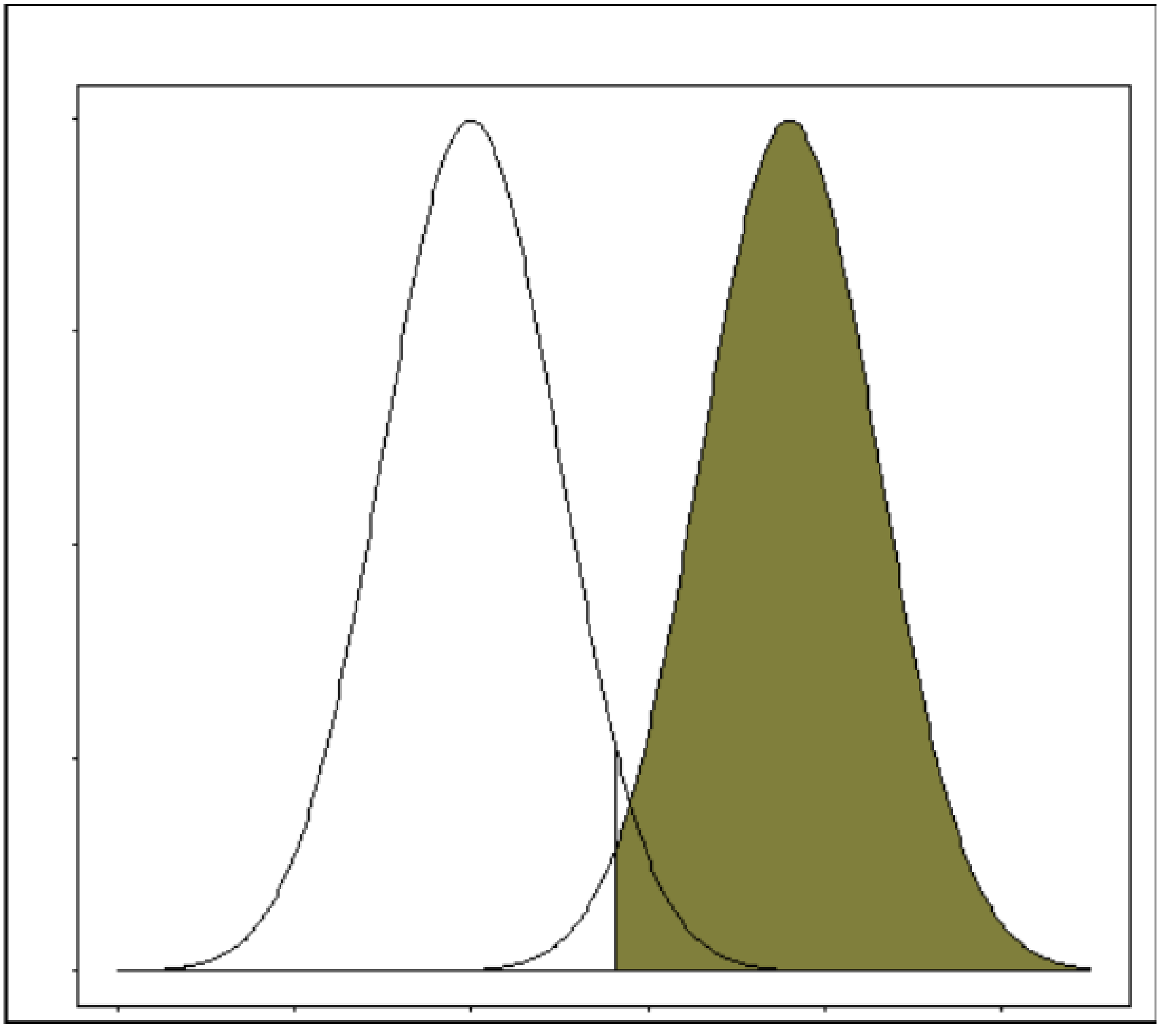


- As sample size (n) ________, power ________, this is rooted in the law of large numbers and central limit theorem, and how greater sample size contributes to less ________
- With standard Error of mean, it is given as: $\frac{\sigma}{\sqrt{N}}$, where:
  - σ: standard deviation
  - N: sample size
- Given the formula for the SE, an increase in N (in the ________) decreases standard error for the mean, which results in less variability, and thus, more power

2.5 Effect Size

- Effect size, is a measure of the ________ of an effect, e.g.:
  - ________ between two populations

  - __________ explained in one variable, by another
- Sometimes, we may have some __________ of an "expected" effect size, like substantial prior literature on the topic
  - However, sometimes it can be __________ to know the exact effect size to expect, which is why we need to __________ studies well, gather large samples, and choose sensitive measures
- An increase in effect size in the __________ will result in higher power, i.e., if a large effect does, in reality, exist $\longrightarrow$ power will be __________



Important

Effect size is defined by different formulas and meaning based upon what inferential test is being used.

? Review: which of the following is treated as an 'effect size' of correlation?

- (x) A) r
- ( ) B) r-squared
- ( ) C) d
- ( ) D) t

Explanation:

3 Identifying Assumption Violations

3.1 Introduction

- Every ________ test we use has some amount of assumptions, that is some ________ pre-requisites to using a certain model or test, an implicit ________ of the data for the test to work as expected
  - The main ________ of assumption violations, is that they reduce power and raise Type ________ error rate
  - Thus, we have to be mindful of possible problems that may arise, relevant to our chosen test

Important

Remember, different tests have different assumptions, and we need to check those relevant to the current test. But, they do share some similarities

- Truthfully, assumption-checking can be very subjective at times, and even oft used tests and checks have been ________ for being insufficient
  - Most importantly, focus on ________ showing the process and provide *enough* information

3.2 Common Assumptions to Check

- For the ________ t-test, there are several assumptions to meet, which are also relevant to other tests. These ________ assumptions are:

- ________
- ________ of Variances, i.e., are the variable distributed similarly
- ________

3.3 Normality Assumption

- The normality assumption is the assumption that is each group's measure is normally distributed in the ________
  - Critically, it's not just the ________! We have to have some notion as to whether we believe our sample is fully ________ of the population
- Histogram strategy
  - One common ________ method is to plot a frequency histogram of a variable for each level of the ________ variable
  - Sometimes one may overlay a theoretical ________ curve to assess it as well

> Discuss: Try to explain what exactly a frequency histogram is, in your own words

> Important
>
> Contrary to what we are used to, in many of the following tests we actually hope for non-significant results, as that points to normality and/or homogeneity of variances

- The Kolmogorov-Smirnov Test (K-S)
  - An assumption test often used to test for normality, but can technically be used to compare against ________ distribution
  - ________ result suggests normality

- The Shapiro-Wilk Test (S-W)
  - An assumption test __________ used to test for normality
  - Like K-S, __________ result suggests normality

3.4 Homogeneity of Variances

- The homogeneity of variances assumption is the assumption that is each group's measure is distributed roughly the __________ as one another, in the population
- Histogram strategy
  - Much like with normality, we may use frequency histograms to exam for __________, kurtosis, or multi-modality in either level of the categorical variable
- Levene's Test
  - An assumption test used to test against the __________ hypothesis that group variances are equal to one another
  - Like the other tests before, __________ result suggests normality

3.5 Independence Assumption

- The independence assumption is an assumption that two variables or events are unrelated, and __________ independently from the two populations from which they belong
- There is not a __________ test or realistic method in analysis to detect whether this is an issue or not $\rightarrow$, more of an issue of research __________

4 Conclusion

4.1 Recap

- Our decision making in choosing to reject or retain the null hypothesis can be confounded by trying to avoid Type I and Type II errors, while still maintain adequate power to detect effects.

- There are several ways to increase power in our study, with the most common method being increasing sample size
- However, assumption violations can detract from our power, when using parametric tests, so we need to be wary of detecting those violations.
- We introduced several useful methods by which to consider finding problems in normality or homogeneity of variances, while also briefly discussing independence.
- Next week will involve a discussion and demonstration on how to address and possibly solve some assumption violation issues

Key Terms

A

assumptions Statistical pre-requisites to using a certain model or test, an implicit requirement of the data for the test to work as expected 12

D

decision In statistics, this is our determination on whether we can reject the null hypothesis, based upon the gathered data and results of analyses upon it 3

E

Effect size A measure of the magnitude of an effect, e.g. a difference between two populations 10

H

homogeneity of variances assumption Assumption that is each group's measure is distributed roughly the same as one another, in the population 14

I

independence assumption An assumption that two variables or events are unrelated, and sampled independently from the two populations from which they belong 14

K

Kolmogorov-Smirnov Test (K-S) An assumption test often used to test for normality, but can technically be used to compare against any distribution; sometimes considered weaker than Shapiro-Wilk; non-significant result suggests normality 13

L

Levene's Test An assumption test used to test whether variances between two groups are equal to one another, tests against the null hypothesis that they are; non-significant result suggests equal variances 14

N

normality assumption Assumption that is each group's measure is normally distributed in the population 13

P

power In statistics, power is the likelihood that we correctly reject the null hypothesis when it is false 5

S

sample size The number of individuals or observations measured, from a subset of a population, usually represented as n or N 10

Shapiro-Wilk Test (S-W) An assumption test only used to test for normality; sometimes considered stronger than Kolmogorov-Smirnov; non-significant result suggests normality 14

standard Error The standard deviation of sample means in a normal distribution of infinitely many possible sample means (i.e., the sampling distribution of the mean) 10

T

type I error When we make the decision to reject the null hypothesis, when it is actually true 3

type II error When we make the decision to retain or fail to reject the null hypothesis 4

The instructor-provided glossary may not include all terms worth memorizing, make sure you consider using the vocabulary list in you book and your own judgment to make sure you have all relevant terms