EPSY 5261 : Introductory Statistical Methods

Day 11

Hypothesis Testing for Comparing Two Means

Learning Goals

- At the end of this lesson, you should be able to…
  - List the steps of a hypothesis test
  - Describe the purpose of a hypothesis test for comparing groups
  - Describe a parametric approach to hypothesis testing for comparing two means
  - List the assumptions for using the t-distribution to test for a difference in means

Recall: Variable Types

- We have been working with quantitative data
  - The population mean (μ) has been our parameter of interest
- Sometimes we have two groups that we want to compare (this could be an additional categorical variable)
  - The parameter of interest is now $\mu_{group1} - \mu_{group2}$

Hypothesis Testing

- Purpose: to test a claim about a population parameter
- One Group
  - *RQ: Did the average movie length increase in 2022?*
- *Two Groups*
  - *RQ: Is there a difference in average movie length between dramas and comedies?*

Steps of Hypothesis Testing

1. Formulate a **research question**
2. Write your **hypotheses**
3. Find **Distribution** of the Null Hypothesis
4. **Compare** Sample to the Distribution of Null Hypothesis
5. Get a **p-value**
6. Make a **decision** to reject or fail to reject the p-value
7. Communicate your **conclusion** in context

Assumptions

- n $\geq$ 30 OR sample distribution looks reasonably normal
  - For single mean: only need to check this for one distribution
  - For difference in means: conditions must be met for both sample distributions
- If these are not met, better to use a randomization test

Use R Studio

- Use the t-distribution to help us get our estimate for the variability
- Use functions in R Studio to also give us our p-value
- We will explore the entire hypothesis test process in today's activity!

Z-test Hypothesis Testing Activity

Summary

- Hypothesis tests help us test a claim while taking into account sampling variability
- They provide one form of evidence to help answer a research question
- We can use a z-distribution to help us conduct our test when we have categorical data