Module 10 Lecture - Mixed Effects Designs

Analysis of Variance

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

Department of Educational Psychology

Table of Contents

1 Overview and Introduction

1.1 Textbook Learning Objectives

- Classify hypothesis tests by type.
- Conduct and interpret hypothesis tests for two population means, population standard deviations unknown.
- Conduct and interpret hypothesis tests for matched or paired samples.

1.2 Instructor Learning Objectives

- Understand when it is appropriate to test with "two-sample" tests vs "one-sample" tests, as described in Module 9
- Appreciate the different conditions and assumptions that must be met for each of the two-sample tests
- Be able to correctly choose an appropriate test, when reading a narrative/anecdotal example of "real variables"
- Understand the difference between "paired" and "independent samples"

1.3 Introduction

- Last module we focused on ____________ testing, and how one can build a structure for ____________ testing by creating paired null and alternative hypotheses
  - E.g., Those in a new math curriculum will perform better on an exam than those in the traditional curriculum
  - ____________ hypothesis: $H_A : Exam_{New} > Exam_{Old}$
    * Remember that the alternative hypothesis is usually something along the lines of our ____________ hypothesis
  - ____________ hypothesis: $H_0 : Exam_{New} \leq Exam_{Old}$
  - This example is technically for a b58fc729-690b-4000-b19f-365a4093b2ff-7B7B3C20626C616E6B2074776F203E7D7D–sample case, where we are comparing two or ____________ groups

Discuss: Review from Module 9: Is saying 'Those in a new math curriculum will perform better than or *equal to* on an exam than those in the traditional curriculum' a valid alternative hypothesis? Why or why not?

- Compare this to a one-sample hypothesis, which was more a focus of last module: E.g., I hypothesize that my class average on the midterm exam will be less than 60.
  - ______ hypothesis: $H_A : Exam < 60$ lines of our ______ hypothesis
  - ______ hypothesis: $H_0 : Exam \geq 60$

Discuss: Review from Module 9: Will the prior one-sample example be more accurately classified as a one-tailed or two-tailed test?

- This module, we'll review b58fc729-690b-4000-b19f-365a4093b2ff-7B7B3C20626C616E6B206F6E65203E7D cases from module 9, scenarios, and computations, before we move into discussing how to conduct two-sample tests, like the type you'd use for the prior example
  - This will also help us demonstrate how the ______ of our alternative and null hypothesis is ______ depending on the situation

Important

The two-sample scenario is very common and more often used than one-sample cases in real research - many educational intervention research is likely to look familiar to this!

2 Review of One-Sample Cases (Module 9) and Computation

2.1 Introduction

- Regardless of the scenario, our hypothesis testing is in service of determining whether we have ______ that our results are such a rare event under the null hypothesis that there is compelling evidence that it is incorrect
  - Refresher example: If a person said that there are 100 $200 bills in a bag, and 1 $10 bill, I'd call BS if I somehow grabbed the $10
  - If we do have ______ evidence, by virtue of our p-value being less than α, then we can reject the null hypothesis

? Making a poor decision on rejecting or retaining a null hypothesis can lead to errors! What is it called if I reject the null hypothesis, when it is actually true?

- (●) A) Type I
- ( ) B) Type II
- ( ) C) Type III
- ( ) D) Not an error

Explanation:

- The one-sample case is when we compare a ______ against some pre-set standard or ______
  - E.g. The prior exam with the midterm graded tested against 60

2.2 Prerequisites to Using the One-Sample Tests

! Important

Remember that choosing the 'right' test for the scenario is perhaps the most valuable skill a good statistician can have!

- Realistically, this is used when we want to understand if a ______ we have differs from an arbitrary expectation on results, and there are several tests we can use for it

  - One-sample b58fc729-690b-4000-b19f-365a4093b2ff-7B7B3C20626C616E6B207A203E7D7D–test, using the ______ distribution
  - One-sample b58fc729-690b-4000-b19f-365a4093b2ff-7B7B3C20626C616E6B2074203E7D7D–test, using the t-distribution
  - One-sample test of ______, e.g., like a percentage
- Both the z-test and t-test are appropriate to variables that are ______ distributed in the sample and the population

Discuss: If a variable is 'normally distributed', then must it be continuous, discrete, or categorical? Explain why

- The z-test and t-test differ in what ______ information we must have about the population
  - The z-test uses the base normal distribution, and thus requires that we know the population ______ deviation parameter (σ)
  - The t-test allows us to use the ______ standard deviation statistic instead, by virtue of using the t-distribution

2.3 Steps in a One-sample Hypothesis Test

1. Set up an appropriate alternative hypothesis comparing some variable against some set value, and then write the corresponding ______ hypothesis.
2. Set α at an appropriate level to minimize risk and balance Type I and Type II error chances

? What is the value that alpha is most often set to, out of tradition?

- (●) A) 0.01
- ( ) B) 0.03
- ( ) C) 0.05
- ( ) D) 0.10

Explanation:

1. Ensure variable is normally distributed (and also, ____________)

2. Determine whether you have the population standard deviation or not
   - If you ____________ $\rightarrow$ one-sample z-test
   - If you ____________ $\rightarrow$ one-sample t-test

1. Compute the p-value and ____________ against α
   - If p-value is ____________ than $\alpha \rightarrow$ reject null
   - If p-value is ____________ or equal to than $\alpha \rightarrow$ retain null
   - This is now done with calculators and computers (see practical assignment walkthrough this week with SPSS), but was historically done by determining whether test statistics (z or t) was greater than a critical value corresponding to the sample size and degrees of ____________

1. Make ____________ and conclusion based upon results

Discuss: What might we say about our 'confidence level' (CL) given our chosen alpha value? How are they related?

2.4 Example of an Applied One-Sample Scenario

- (Set hypotheses) I predict that professors will spend more than 40 hours this week working on teaching classes
  - $H_A : Hours > 40$
  - $H_0 : Hours \leq 40$
- (Set alpha) I decide to set $\alpha = 0.10$, saying that I am not particularly concerned about Type I errors
- (Ensure normality and continuous nature) I'll assume the hours worked by professor is normally-distributed, and this is a continuous variable because hours worked is continuous

- (Do I have population standard deviation?) No, I don't know what the population standard deviation, thus I know I need to rely on the t-distribution
- (Compute p-value to make decision against alpha) I find a p-value of 0.09, which is less than $\alpha = 0.10$, thus I reject the null hypothesis
- (Make decision) I have evidence that professors spend more 40 hours this week working on teaching classes

Discuss: Why do I say 'have evidence' rather than prove?

3 Comparing One- and Two-Sample Cases

3.1 Introduction

- While one-sample tests cases do exists, we often want to test groups against __________ when we believe they __________ in some meaningful way
  - E.g., students who got a new intervention or tool, patients who got a new medication vs. an already established medication, the same group of children before and after a new after-school program (called a "paired" sample, we'll come back to this later!)
- The focus of the later part of this unit will be on these two-sample tests

3.2 Differences in Notation and Hypotheses

- One-sample cases are written to compare against a certain __________:
  - E.g., $H_A : Hours > 40$, $H_A : Exam < 60$
- Two-sample cases are written to compare against two groups or __________ :
  - E.g., $H_A : Exam_{New} > Exam_{Old}$, $Hours_{NonTenure} < Hours_{Tenure}$

Discuss: Try writing one example each of a one-sample and a two-sample hypothesis

3.3 Differences in Tests

- One-sample:
  - One-sample z-test
  - One-sample t-test
- Two-samples (__________ of these tests to be discussed soon):
  - Independent-samples t-test
  - Dependent-samples/paired t-test
  - z-test (but very __________ - won't be something we cover)

Discuss: Try to explain, why do you think the two-sample z-test would be so rare in practice?

Important

Make sure you feel comfortable in the differences in these scenarios before moving on!

4 Two Population Means with Unknown Standard Deviations

4.1 Introduction

> **Important**
>
> In the following sections, we consider the scenario that we have two separate groups with no shared members - we'll talk about comparing people to themselves in a future section

- The most common two-sample problem to run into for real research will be when you ________ know the ________ population standard deviations of our two fully separate groups
  - Just like the one-sample, we love our t-distribution!
- To compare our two groups, we have to consider both their ________ and ________
  - We have to have a measure of ________ here, because otherwise large variance alone could ________ for a perceived difference

4.2 Formula for (Welch's) Independent-Samples t-test

> **Important**
>
> A common misconception is that we 'test' using the two groups, but we actually calculate the p-value using the *difference* between them!

- Thus, we need to get the ________ error (SE) of the difference between our means with the following equation:

$$SE_{diff} = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

- Then, our full t-statistic formula for Welch's independent-samples t-test with ________ variances is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{SE_{diff}}$$

- Associated degrees of freedom:

$$df = \frac{(\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2})^2}{(\frac{1}{n_1-1})(\frac{(s_1)^2}{n_1})^2 + (\frac{1}{n_2-1})(\frac{(s_2)^2}{n_2})^2}$$

- Where:
  - s_1 and s_2, the ________ standard deviations, are estimates of σ_1 and σ_2, respectively.
  - σ_1 and σ_2 are the ________ population standard deviations.
  - $\bar{x}_1$ and $\bar{x}_2$ are the ________ means.
  - Notice that all of these values are ________ that we can calculate from a data sample!
- Calculation of p-value is done with calculator or computer
- Wait - who's Welch?
  - Student originally developed a version of this two-sample test that assumes that the ________ between the two group are ________, and "pools" the variances in the formula (see pooling)
  - The Welch version does not make this same assumption, and thus is a bit more ________

Important

Your book recommends Welch-version at baseline and a lot of people forget to report which one they used! Be careful of looking at which specific test they used and know which one you used!

4.3 Worked Example of Independent-Samples t-test

4.3.1 Formulas

$$SE_{diff} = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{SE_{diff}}$$

$$df = \frac{(\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2})^2}{(\frac{1}{n_1-1})(\frac{(s_1)^2}{n_1})^2 + (\frac{1}{n_2-1})(\frac{(s_2)^2}{n_2})^2}$$

4.3.2 Scenario

- Given two hypothetical groups measured on IQ:
  - I hypothesize that group 2 will have a greater IQ then group 1
    - $H_A : IQ_1 < IQ_2$
    - $H_0 : IQ_1 \geq IQ_2$
  - Group 1
    - $\bar{x}_1 = 100$
    - $s_1 = 15$
    - $n_1 = 50$
  - Group 2
    - $\bar{x}_2 = 120$
    - $s_2 = 10$
    - $n_2 = 50$

4.3.3 Substituted Formulas

$$SE_{diff} = \sqrt{\frac{(15)^2}{50} + \frac{(10)^2}{50}} = 2.55$$

$$t = \frac{(100 - 120)}{2.55} = -7.84$$

$$df = \frac{(\frac{(15)^2}{50} + \frac{(10)^2}{50})^2}{(\frac{1}{50-1})(\frac{(15)^2}{50})^2 + (\frac{1}{50-1})(\frac{(10)^2}{50})^2} = 84.9$$

- From computer: $p < 0.0001$, thus I reject the null hypothesis and conclude that group 2 is significantly higher than group 1.

Discuss: Now you try the same process to calculate SE, t, and df with these two groups. Group 1) mean = 50, sd = 12, n = 40, group 2) mean = 45, sd = 30, n = 40

5 Matched or Paired Samples

5.1 Introduction

- Paired samples are (usually) those which have two sample measurements are drawn from the ____________ people at two different ____________ points
  - E.g., before and after an intervention; maybe I want to see if a lesson changes student's knowledge on biopsychology, quiz before and after and test if there is a change

Discuss: Come up with another example of a time you may see a paired sample

Important

There are a lot of good methodological reasons to use paired samples in research designs, so they show up often!

5.2 Formulas

Important

In the following equations, n is the number of pairs/people

- Unlike with the previous formulas, we have to first calculate the mean and standard deviation of the ________________ between each pair of scores with:

$$\bar{d} = \frac{d_1 + d_2 + ... + d_n}{n}$$

$$s_d = \frac{\sqrt{(d_1 - \bar{d})^2 + (d_2 - \bar{d})^2 + ... + (d_n - \bar{d})^2}}{n-1}$$

- We then use those values to find t-statistic and degrees of freedom with:

$$SE_d = \frac{s_d}{\sqrt{n}}$$

$$t = \frac{\bar{d}}{SE}$$

$$df = n - 1$$

5.3 Worked Example of Paired-Samples t-test

5.3.1 Formulas

$$\bar{d} = \frac{d_1 + d_2 + ... + d_n}{n}$$

$$s_d = \frac{\sqrt{(d_1 - \bar{d})^2 + (d_2 - \bar{d})^2 + ... + (d_n - \bar{d})^2}}{n-1}$$

$$SE_d = \frac{s_d}{\sqrt{n}}$$

$$t = \frac{\bar{d}}{SE}$$

$$df = n - 1$$

5.3.2 Scenario

- Given one group of students given a 10pt pre-test and post-test on content from a biology lecture
  - $X_{pre} = \{0, 1, 2, 3, 2, 1\}$
  - $X_{post} = \{5, 6, 8, 8, 3, 7\}$
  - $d = \{5, 5, 6, 5, 1, 6\}$

5.3.3 Substituted Formulas

- Statistics:
  - $\bar{d} = 4.67$
  - $s_d = 1.86$

$$SE_d = \frac{1.86}{\sqrt{6}} = 0.760$$

$$t = \frac{4.67}{0.760} = 6.14$$

$$df = 6 - 1 = 5$$

Discuss: Now you try with groups of paired data: (1, 1, 2, 2, 3) and (5, 9, 8, 7, 6) - calculate df, SE, and t

6 Conclusion

6.1 Recap

- Hypothesis testing with two-samples follows the same process as with one-samples, but introduces comparisons between groups, rather than just testing against some value
- When testing two samples, they may be independent or paired, depending on the design of the study, and that changes what computation process we use
- The independent-samples test may either assume equal variance's (Student's) or not (Welch's)
- While there are two-sample tests for proportions and when the population standard deviation is known, it is far more common to use the t-distribution

Key Terms

D

Dependent-samples/paired t-test An inferential statistical test for comparing two sets of paired measurements, like those that come from measuring the same people twice 8

I

Independent-samples t-test An inferential statistical test for comparing two entirely separate groups of measurements or people to determine if they are significantly different 8

O

One-sample t-test An inferential statistical test for comparing a single sample against an arbitrary value using the normal distribution; appropriate when the population standard deviation is unknown 8

one-sample tests Tests and hypotheses that focus on the comparison of a single sample group to an arbitrary standard or value 7

One-sample z-test An inferential statistical test for comparing a single sample against an arbitrary value using the normal distribution; appropriate when the population standard deviation is known 8

P

Paired samples When two samples of measurements are connected, usually by virtue of having come from the same person at two time points 12
pooling In t-tests, it a specific procedure that treats variances and standard deviations of the groups as equivalent - not used in Welch's, but its used in Student's 10

T

two-sample tests Tests and hypotheses that focus on the comparison between two samples or groups of measurements 7

W

Welch's independent-samples t-test A specific formulation of the independent samples t-test that does not pool variances and thus does not assume equal variances between groups, unlike Student's version 9

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