Module 3 Lecture - Transformations and Non-parametric Comparisons for Two Groups

Analysis of Variance

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

Department of Educational Psychology

Table of Contents

1 Overview and Introduction

1.1 Objectives

- In module 2, we discussed the intricacies of avoiding _____
  I and II errors in hypothesis testing, the connection to statistical power, and
  how assumption violations can create problems in Type II errors

- In this module, students should be able to:
  - Understand the different possible options to pursue in light of assumption
    violations
  - Appreciate the use of variable transformations in addressing problems in
    skew, and using trimming/winsorizing to address kurtosis, i.e., solving
    problems in normality
  - Understand non-parametric tests as viable alternative to our usual para-
    metric tests, with certain nuances to be aware of in application.

1.2 Introduction

- Previously, we talked about _____ problems in 3 types of com-
  mon assumptions:
  - _____
  - _____ of variances
  - _____ of sampling variables
- Now we will go through the Solutions for Assumption Violations, and how they
  may (or may not) solve our problems
  - Our data is _____ ideal! We might need to readily con-
    sider these options

> 📢 Discuss: Of the top of your head, try to remember the different methods of
> identifying the above assumption violations

2 Solutions for Assumption Violations

2.1 3 General Strategies

> ❗ Important
>
> There is considerable controversy and conflicting ideas in this area, as to 'what should we do' - take non-parametric class for more fun!

- We need a way to _____ assumption violations when they occur (well mostly - see Do Nothing!)
  - As mentioned before, the main _____ of assumption violations, is that they reduce power and raise Type _____ error rate
  - Since we are coming into most studies with a hypothesis of differences _____, we want to plan an analysis that has the sufficient power to detect that hypothesized difference
- We have 3 _____ for approaching assumption violations:
  - Do Nothing!
  - Correct for Violations in Data
  - Use Non-parametric Tests
- Before we go through those, lets talk about some reasonable advice for working through these:

2.2 Advice in 'Fixing' Assumption Violations

- First, always _____ checks/tests for assumptions of the test you are using
  - E.g., _____ tests like Kolmogorov-Smirnov and Shapiro-Wilk; _____ to examine skew and kurtosis; Levene's test for homogeneity of _____, etc.
  - Review notes from module 2 for more details on each of those
- Second, you should report assumption checks both _____ AND _____ you attempt a fix
  - I.e., If you _____/trim a variable, you should check assumption both before AND after the transformation/trim
  - Re-check _____ assumptions, not just the ones you attempted to fix with the transformation or _____

- Third, be _____ and transparent when it seems like a solution does not work or when the downsides outweigh the results
  - We'll talk about different _____ when using each strategy
  - It's good practice to always be straightforward about _____ of our analyses

> ❗ **Important**
>
> The emphasis is always on transparently reporting results!

2.3 Do Nothing!

- Without going into too much detail, we are concerned with how robust a test is, or how _____ a test is to assumption violations, and how well it works under less-than-ideal circumstances

- Some researcher's hold that many commonly used tests, i.e., t-tests are reasonably _____ at baseline to assumption violations
  - This becomes even more true at large n size (See discussion in Sample Size section in module 2 notes)
  - However, when samples are small or the groups being compared are _____ - this route cannot be recommended

> 📢 Discuss: How often do you think this option is taken in 'real' research?

3 Correct for Violations in Data

3.1 Transformations / Dealing with Skewness

- There are several options for selectively _____ and or _____ our data to correct for certain patterns of skewness, kurtosis, or _____

  – However, be aware these are not panaceas - they come with their own issues

- Making mathematical $\boxed{\text{variable transformations}}$ is largely used to address the Normality Assumption, but maybe indirectly solve other issues as well

- The exact transformation is dependent on the type of problem, specifically the _____ :

  – _____ /right skew: Logarithmic (Severe) or Square Root (Moderate)
    * Logarithmic method: $\log_{10}(X)$
    * Square root method: $\sqrt{X}$
    * X is the variable of interest
  – _____ /left skew: Reflect and Logarithmic (Severe) or Reflect and Square Root (Moderate)
    * Logarithmic method: $\log_{10}(K - X)$
    * Square root method: $\sqrt{K - X}$
    * K is a constant from which each score is subtracted, so that the smallest score is equal to 1

- Advantages:
  – Makes use of _____ available data
  – Allows for use of a _____ well-known technique.

- Disadvantages:
  – _____ can be questionable
  – Fixing one assumption violation can _____ others

? I have a variable of numeric test scores, in which a histogram shows that there is severe bunching of scores to the left, with a tail off to the right. Which of the above transformation might be the best to try first?

    A) Log
    B) Square Root
    C) Reflect and Log
    D) Reflect and Square Root

Explanation:

3.2   Trimming and Winsorizing / Dealing with Kurtosis

- Another option, particularly useful for _____ kurtotic (platykur-
  tic) distributions (relatively flat distributions with an unusual number of obser-
  vations in the tails) is to use variable trimming .

- A trimmed sample is a sample where a _____ percentage of
  extreme values is removed from each tail.
    - Of course, if you are comparing groups, you would want to trim the same
      percentage from the tails of _____ groups to be fair.
    - _____ percent or 0.10 from each tail is the most common
      amount to trim
    - Example Mean: $((6.0 + 8.1 + 8.3 + 9.1 + 9.9)/5) = 8.28$
    - Example 20% Trimmed Mean: $(8.1 + 8.3 + 9.1)/3 = 8.50$

- Another related option is using winsorizing
    - A Winsorized sample replaces the trimmed values by the most _____
      value remaining in each tail.
    - Example Dataset: 2, 4, 7, 8, 11, 14, 18, 23, 23, 27, 35, 40, 49, 50, 55,
      60, 61, 61, 62, 75
    - Example 20% Winsorized Dataset: 7, 7, 7, 8, 11, 14, 18, 23, 23, 27, 35,
      40, 49, 50, 55, 60, 61, 61, 61, 61
    - Note: _____ of freedom (df) for a test on a trimmed or
      Winsorized sample must be adjusted for the data trimming. For both,
      you would subtract your total N by the number of trimmed cases to get a
      new value for N. (We do this even for Winsorized samples b/c the added
      values are really pseudovalues.)

- Advantages:
    - Allows for use of a _____ well-known technique.
    - Interpretability of variable remains _____

- Disadvantages:
    - Loss of information - what if those _____ were an impor-
      tant part of the phenomenon

> 📣 Discuss: Using the variable data 1, 2, 3, 4, 5, try 20 percent trimming and winsorizing this data

4 Use Non-parametric Tests

4.1 Introduction

- Non-parametric test are _____ -free statistical tests that are not based on _____ (i.e., means or standard deviations) or assumptions about the normality underlying data distribution.

- They do still have *some* _____ , just not about normality! More on that in Assumptions of Non-parametric Tests

- Instead, non-parametric tests are based on amounts such as percentages (Chi-square) or ranks, i.e., _____ or _____ - transformed data

- Non-parametric tests are just as _____ as traditional tests, and under situations of violated assumptions can be much more _____

- We'll discuss the Wilcox Rank Sum Test and the Mann Whitney U Test as they are non- parametric tests comparing the _____ orderings of two independent samples
  - These can be thought of as the non-parametric versions of the _____ samples t-test

- We'll also cover the Wilcox Matched Pairs Signed Rank Test, sort of an analog to the dependent-samples t-test

- Because these test use ranks, they can actually use _____ data, unlike the t-test!

> 📣 Discuss: Think of a hypothesis/scenario in which you'd use the independent samples t-test

4.2 Wilcox Rank Sum Test

- The Wilcox Rank Sum Test is based on the logic that if there truly is a significant difference between two groups, the _____ from one group should generally be lower than the _____ from the other group.

- Following from that, if the groups are different, the _____ of one group should be lower than the _____ of the ranks from the other group.
    - If the sum of the one group is too small relative to the other sum we will reject the null hypothesis.

4.2.1 Calculating Test Statistic

- First, combine both groups and _____ their _____ values from smallest (starting at 1) to largest
    - In the scenario of _____, you can assign them all the mean rank or assign adjacent ranks at random

Group A Scores ($n_1 = 4$)	Group B Scores ($n_2 = 5$)
85	70
92	82
88	75
95	78
	80

Score	Group	Rank
70	B	1

Score	Group	Rank
75	B	2
78	B	3
80	B	4
82	B	5
85	A	6
88	A	7
92	A	8
95	A	9

- Second, calculate the _____ of the ranks of the smaller group, also called W_s.
    - The "smaller" group is the one with a smaller n

Group A Ranks	Group B Ranks
6	1
7	2
8	3
9	4
	5
Sum (W_A) = 30	**Sum (W_B) = 15**

- Third, compare W_s against _____ value table, which is derived from sample _____ from the smaller and larger groups
    - Realistically, we'll use SPSS for this

📢 Discuss: Try calculating the rank sums for groups of data: A: 23, 34, 89; B: 16, 12, 40, 50; which of these is W_s?

❗ Important

Contrary to what we are used to, we want our test statistic to be *less* than the critical value

- This is a _____ -tailed test that tests to see if the ranks of the smaller group are sufficiently smaller than the larger group.

4.2.2   When Smaller Group Has Larger Ranks

- A problem we may run into in the above process is if the smaller group (by n) actually has the _____ ranks
    - To solve for this, we can backwards rank the data, i.e., start with the _____ value as 1
- We can check for that possibility by calculating W'_s or the compliment of W_s
    - $W'_s = 2\bar{W} - W_s$ where
- Where:
    - $2\bar{W} = n_1(n_1 + n_2 + 1)$
    - n_1: sample size of smaller sample
    - n_2: sample size of the larger sample

> **❗ Important**
>
> So effectively, we have to consider the statistic both ways to account for both directions. This gets covered by our next test.

4.3   Mann Whitney U Test

- The ┃ Mann Whitney U Test ┃ is a test completely _____ to the Wilcoxon Rank Sum, however, it is slightly more _____ in order to eliminate the need to compute W'_s.

> **❗ Important**
>
> For this reason, it tends to be a bit more immediately useful, because it doesn't require going through the equation twice.

- The Mann Whitney U, like the Wilcoxon Rank Sum, is based on ranked data.

4.3.1   Calculating Test Statistic

- We start with the exact same steps as the Wilcox Rank Sum Test, calculating W_s
- The we calculate the U statistic:

$$U = \frac{n_1(n_1 + 2n_2 + 1)}{2} - W_s$$

- Where:

      – n_1: sample size of smaller sample
      – n_2: sample size of the larger sample

> **❗ Important**
>
> SPSS will only calculate Mann Whitney U, but because these tests are equivalent, that is fine!

4.4 Wilcox Matched Pairs Signed Rank Test

- The Wilcox Matched Pairs Signed Rank Test is a non- parametric test testing the null hypothesis that two _____ samples were drawn from identical populations with the same mean.
  - This is a non-parametric version of the _____ /paired samples t-test.
- The logic of the Signed Ranks Test rests on measuring the direction and _____ of change.

4.4.1 Calculating Test Statistic

- Compute difference scores (d) between time 1 t_1 and time 2 (t_2)

Participant	Before	After	Difference (d_i)
1	120	115	-5
2	135	130	-5
3	110	120	+10
4	145	145	0
5	130	110	-20
6	125	122	-3

- Next, rank *regardless of sign*

$|d_i|$	Absolute Rank	Sign
3	1	Negative
5	2.5	Negative
5	2.5	Negative
10	4	Positive
20	5	Negative

- There are two kinds of _____ possible this time:

- Like before, you can have rank ties which you would handle by either random ranks or tied ranks.
    - Now, though, you have the possibility for paired values to be equal (difference = 0). When this happens do not rank that case, and _____ it from your sample.
- Next, Sum positive and negative ranks

Absolute Rank	Positive Ranks (R^+)	Negative Ranks (R^-)
1		1
2.5		2.5
2.5		2.5
4	4	
5		5
Sum (W)	$W^+ = 4$	$W^- = 11$

- Our test statistic is T, which is the smaller of the two _____ sums

4.5 Assumptions of Non-parametric Tests

- These mentioned non-parametric tests allow us to ignore _____ problems in our data
    - However, we still need to test for and address problems in homogeneity of _____
    - See prior module lecture on examining _____ and _____ test to help identify these issues

4.6 Effect Size Under Non-parametric Tests

- We can use this quick equation and rule-of-thumb for effect _____ under these tests

$$r = \frac{z}{\sqrt{n}}$$

- Where:
    - z is the _____ test value
    - n is the sample size
- Interpretation
    - _____ : 0.1
    - _____ : 0.2
    - _____ : 0.3

"I don't mind not knowing. It doesn't scare me." — Richard P. Feynman

> ❗ **Important**
>
> Like with many statistics, there are *many* other ways of calculating effect size, some with much more complex equations - but for the scope of this class, we'll use this

4.7 Advantages and Disadvantages of Non-parametric Tests

- Advantages of Non-Parametrics
    - _____ of variables remains intact
    - Makes use of all available data
    - We still have good options for statistical _____ testing, _____ size, and other normal comforts of parametric tests
- Disadvantages
    - Non-parametric tests are relatively _____. Many researchers would not know how to _____ them, and thus, may prefer just to see the parametric version, regardless of issues

> 📢 Discuss: Do you think this is disadvantage is a sufficient reason to avoid these tests? What issues (if any) do you have with this view?

5 Conclusion

5.1 Recap

- We have now discussed various methods we may pursue when dealing with oddities in our data that result in assumption violations

- Some of these strategies involve carefully modifying our data with transforming, trimming, or winsorizing; some strategies involve alternative selection of tests;

and then finally, we may do nothing (mindfully)

- Each of our options result in some new things to be careful about, like transformation changing the scale of our data, trimming removing data, or non-parametric tests changing interpretation of results

- Our discussion of the non-parametric analogs for the t-test will help lead us into a similar discussion on alternative tests to the one-way ANOVA (and others) later in the semester

- Paramount to all of these things, we must be mindful of transparency when working through assumption problems

Key Terms

M

Mann Whitney U Test A non-parametric test similar to the Wilcox Rank Sum Test, but used for the purpose of doing the test 'both ways'; preferred over Wilcox Rank Sum, and equivalent to it 10

N

Non-parametric test Tests without normality assumptions and without mean or standard deviation parameters 7

R

robust A description of how resilient a test is to assumption violations, and how well it works under less-than-ideal circumstances 4

V

variable transformations Mathematical transformations usually used to counteract skew, to try and meet the normality assumption 5
variable trimming Removing a certain amount of variables at the high and low end of a variable, usually to solve for platykurtic-ness 6

W

Wilcox Matched Pairs Signed Rank Test A non-parametric test used to compare dependent-samples, i.e., those that are the same individuals taken at two times points 11
Wilcox Rank Sum Test A non-parametric test for testing whether the ranks of one group are significantly smaller than another; akin to a independent-samples t-test 8
winsorizing A process of replacing the most extreme values of a variable with the last remaining one 6

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