

## Comparing Several Means (Part 1)

Ray Block, Jr.  
PS #585  
Research Methods

### Today's Blueprint

#### Last Class

- Comparing the Means of 2 *Dependent* Populations
- Comparing the Means of 2 *Independent* Populations
  - Difference Tests
  - Confidence Intervals (T-Interval)

#### Today's Class

- Comparing 2 or more means
  - **AN**alysis **O**f **V**ariance (ANOVA)
    - The Big Idea

### The Big Idea

### The Big Idea

#### Data Analysis:

- The goal of data analysis:
  - Infer population characteristics from sample data
    - You evaluate samples by how well they represent the population of interest
  - Statistical Models provides methods for doing this
    - You evaluate statistical models by how well they fit the sample data
- In the first data analysis section, we made inferences about populations by analyzing 1 variable at a time (uni-variate data analysis)
  - Measures of Central Tendency
  - Measures of Variability
- In the previous data analysis section, we made inferences about populations by analyzing 2 variables at a time (bi-variate data analysis)
  - Correlations
  - Mean Comparisons
- In this data analysis section, we will make inferences about populations by analyzing more than 2 variables at a time (multi-variate data analysis)
  - Multiple [Mean] Comparisons
  - **AN**alysis **O**f **V**ariance (ANOVA)

### The Big Idea

- Multiple Comparisons:
- T-Tests:
  - Tests of significance
    - Determines whether 2 sample means are *significantly different* from one another
    - Determines the probability that means were drawn from the same population
- Analysis of Variance (ANOVA):
  - Test of significance for more than 2 groups
  - Like T-Tests, ANOVA deals with differences between sample means
  - Unlike T-Tests, ANOVA imposes no restriction on the number of means
- Why not just use T-Tests?
  - Imagine that you have 3 groups (3 sample means)

- T-Tests would simply compare combinations of these groups one at a time (independent groups T-Tests)
- There are 2 potential problems with this approach
- Problem #1: Too many comparisons!!!
  - The number of T-Tests increases as the number of groups increases
  - Example: We would have to do 3 different T-Tests to compare 3 groups
    - Group 1 *versus* Group 2
    - Group 1 *versus* Group 3
    - Group 2 *versus* Group 3
- But what if we did more than 3 groups. For example, if we wanted to compare 6 groups we would need to do 15 comparisons

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Group 1	X	X	X	X	X	X
Group 2		X	X	X	X	X
Group 3			X	X	X	X
Group 4				X	X	X
Group 5					X	X
Group 6						X

- Problem #2: Inflated Probability of Type I Error
  - The probability of making at least one Type I Error increases as the number of T-Tests increases
  - Alpha at .05 means that we are willing to risk being wrong 5% of the time when we reject the  $H_0$
  - With multiple T-Tests, we are no longer just sampling one T-Value
  - Therefore, the alpha for the experiment is no longer equal to .05
- Think of the F-Test as an alternative to the T-Test
  - F-Test = *Omnibus* Test
  - Controls for Type 1 Errors

The F-Test is *Omnibus*:

- Omnibus: (adj.) Covers many things at once
  - F-Test = One overall comparison
    - Does not tell you which means differ
    - Simply indicates that not all population mean are equal
- The F-Test tests the null against all possible alternatives
  - $H_0: \mu_{\text{group1}} = \mu_{\text{group2}} = \mu_{\text{group3}}$
  - $H_1: \mu_{\text{group1}} \neq \mu_{\text{group2}} \neq \mu_{\text{group3}}$
- What are we testing?
  - Rejecting  $H_0$  tells us that at least two of the means are different
    - All of the means can be different from one another
    - Or, some means might differ while other are the same

The F-Test Controls for Type I Error:

- In addition to allowing you to tell whether or not means differ, multiple comparison techniques *control* or *reduce* the probability of making a Type I Error
- Different Kinds of Type I Errors:
  - Error Rate Per Comparison (PC)
  - Familywise Error Rate (FW)

#### Error Rate Per Comparison (PC):

- Every time we run a comparison or statistical test there is a chance that we have made a Type I error on that particular comparison
- The Probability of making a Type I Error on any given comparison
- Probability of PC is typically equivalent to our alpha (.05 or .01)

#### Familywise Error Rate (FW):

- If we run a set of comparisons, this series of comparisons are called a family
- The FW error rate is the probability that when we make a whole series of comparisons, that we will make at least one Type I Error
- If we are making a number of comparisons, we want to keep our type I Error rate low for the group of comparisons
  - Multiple comparisons procedures are directed towards controlling the Familywise Type I Error rates

#### Analysis of Variance (ANOVA):

- ANOVA = Type of multiple comparison
  - Means Difference comparisons use T-Tests
  - ANOVA uses the F test
- Assumptions of ANOVA:
  - *Homogeneity of variance*: Each of our populations has the same variance
  - *Normality*: The scores for each experimental condition are normally distributed around the population mean
  - *Independence of observations*: Observations within an experimental treatment (of the independent variables) are independent of one another
- What Happens if We Violate These Assumptions?
  - ANOVA is a robust test
  - It is minimally affected by violations of population normality and homogeneity so long as the samples are of equal size
  - The big problem is when you violate these assumptions when the sample sizes are different

#### Calculating an ANOVA:

- Same logic as before:
  - Assume that  $H_0$  is true
  - State an expected result based on this assumption
  - compute a sample statistic (T-Test, F-Test, etc)
  - Treat the statistic as a score in some known sampling distribution
  - If the test statistic falls within the rejection region, reject  $H_0$ , otherwise, retain  $H_0$
- What's new with ANOVA?
  - The Basic logic does not change between mean difference tests and ANOVA
  - The difference is that they use different statistical tests
  - Rather than using means to evaluate the  $H_0$ , the F-Test is based on the ratio of the *variances*