Comparing Several Means (Part 2)

Ray Block, Jr.
PS #585
Research Methods

Today's Blueprint

Last Class

- Comparing Several Means (Part 1)
  - **AN**alysis **O**f **VA**riance (ANOVA)
    - The Big Idea

Today's Class

- Comparing Several Means (Part 2)
  - Calculating an ANOVA
  - Some Examples

Calculating an ANOVA

Calculating an ANOVA

The Logic of ANOVA

- Similar 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 a ratio of the *variances*

- Essentially, ANOVA entails calculating two estimates of the population variance:
  - Variance among treatment means
  - Variance among multiple groups from the same underlying population

- Variance among treatment means:
  - For any given treatment, the variance of the scores that compose that treatment could be estimated and used as a measure of the underlying population variance

- Multiple groups from the same underlying population:
  - Assume homogeneity of variance
  - Thus, multiple groups represent multiple estimates of the common population variance

- If the two estimates agree, we have *no reason to reject the H_0*
- If they disagree, conclude that *the underlying treatment changed the second estimate*

The F-Test

- To calculate the F-Test, you must distinguish between two types of Variance
  - Between-group Variance
  - Within-group Variance

- Within-Group Variance:
  - Unexplained variance
  - Variance caused by individual differences, errors in measurement, etc.
  - Sometimes called *error variance*

- Between-Group Variance:
  - Same sources as within-group variance (individual differences, measurement error, etc.)
  - But also includes any variance due to group differences (i.e. the experimental manipulation)

- F = The ratio of 2 independent variance estimates:

$$F = \frac{\text{Variance between sample means}}{\text{Variance expected by chance (error)}}$$

- Basically, you divide the first variance estimate by the second variance estimate:

$$F = \frac{\text{Differences between treatment group means}}{\text{Difference among multiple variances treated alike}}$$

Calculating an ANOVA

- If H_0 is true:

$$F = \frac{\text{Differences between treatment group means}}{\text{Difference among multiple variances treated alike}} = 1$$

Calculating an ANOVA

- If H_0 is false:

$$F = \frac{\text{Sampling Error + Effect of IV + Measurement Error (For Estimate 1)}}{\text{Sampling Error + Measurement Error (For Estimate 2)}} > 1$$

Calculating an ANOVA

- To compute F, we need:
  - Mean Squared Error between Groups (MS_{Error})
  - Mean Squared Error for Treatment Groups ($MS_{Treatment}$)
- To obtain those, we need:
  - Total Sums of Squares (SS_{Total})
  - Error Sums of Squares (SS_{Error})
  - Group (between) Sums of Squares ($SS_{Between}$)
  - Total Degrees of Freedom (df_{Total})
  - Error Degrees of Freedom (df_{Error})

- Group Degrees of Freedom ($df_{Between}$)

Calculating an ANOVA

- Some ANOVA notation:
  - N = number of total data values.
  - n_i = number of data values in i^{th} group
  - t = number of groups (or "factor levels")
  - y_{ij} = j^{th} data value in i^{th} group
  - $Y_{i.}$ = sum of data values in i^{th} group
  - Y.. = sum of all data values

Calculating an ANOVA

- Computational Formula:

$$F = \frac{MS_{between}}{MS_{within}}$$

Review

- ANOVA = Division of the overall variability in data values in order to compare means.
- Overall (or "**total**") variability is divided into two components:
  - the variability "**between**" groups, and
  - the variability "**within**" groups
- Summarized in an "ANOVA" table:

General ANOVA Table

Source of variation (SOURCE)	Degrees of Freedom (DF)	Sum of Squares (SS)	Mean Square (MS)	F-Statistic (F)	P-value (F)
Between	T - 1	SS (Between)	MSB	MSB/MSE	Significance
Within	N - T	SS (Error)	MSE		
Total	N - 1	SS Total			