Chapter 13: Introduction to Analysis of Variance
Introduction

- Analysis of variance (ANOVA) is a hypothesis-testing procedure that is used to evaluate mean differences between two or more treatments (or populations).
- As with all inferential procedures, ANOVA uses sample data as the basis for drawing general conclusions about populations.
- The major advantage of ANOVA is that it can be used to compare two or more treatments.
- Thus, ANOVA provides researchers with much greater flexibility in designing experiments and interpreting results.
Population 1 (Treatment 1)

$\mu_1 = ?$

Sample 1

$n = 15$

$M = 23.1$

$SS = 114$

Population 2 (Treatment 2)

$\mu_2 = ?$

Sample 2

$n = 15$

$M = 28.5$

$SS = 130$

Population 3 (Treatment 3)

$\mu_3 = ?$

Sample 3

$n = 15$

$M = 20.8$

$SS = 101$
Terminology in Analysis of Variance

- When a researcher manipulates a variable to create the treatment conditions in an experiment, the variable is called an independent variable.
- On the other hand, when a researcher uses a non-manipulated variable to designate groups, the variable is called a quasi-independent variable.
- In the context of analysis of variance, an independent variable or a quasi-independent variable is called a factor.
- The individual conditions or values that make up a factor are called the levels of the factor.
Terminology in Analysis of Variance cont.

- Although ANOVA can be used in a wide variety of research situations, this chapter introduces ANOVA in its simplest form.
- Specifically, we consider only single-factor designs.
  - That is, we examine studies that have only one independent variable (or only one quasi-independent variable).
  - Second, we consider only independent-measures designs; that is, studies that use a separate sample for each treatment condition.
Statistical Hypotheses for ANOVA

- Null Hypothesis would be stated as follows:

\[ H_0: \mu_1 = \mu_2 = \mu_3 \]

- The Research Hypothesis would be stated as follows:

\[ H_1: \mu_1 \neq \mu_2 \neq \mu_3 \quad (\text{All three means are different.}) \]
The Test Statistic

- The test statistic for ANOVA is very similar to the t statistics used in earlier chapters.

\[ t = \frac{\text{obtained difference between two sample means}}{\text{standard error (the difference expected with no treatment effect)}} \]

- For ANOVA, the test statistic is called an F-ratio and has the following structure:

\[ F = \frac{\text{variance (differences) between sample means}}{\text{variance (differences) expected with no treatment effect}} \]

- As you can see from the formula above, the variance in the numerator of the F-ratio provides a single number that describes how big the differences are among all of the sample means.
The Test Statistic cont.

- In much the same way, the variance in the denominator of the F-ratio and the standard error in the denominator of the t statistic are both measuring the mean differences that would be expected if there is no treatment effect.
- Remember that two samples are not expected to be identical even if there is no treatment effect whatsoever.
- In the independent-measures t statistic we computed an estimated standard error to measure how much difference is reasonable to expect between two sample means.
- In ANOVA, we will compute a variance to measure how big the mean differences should be if there is no treatment effect.
The Test Statistic cont.

- Finally, you should realize that the t statistic and the F-ratio provide the same basic information.
- In each case, the numerator of the ratio measures the actual difference obtained from the sample data, and the denominator measures the difference that would be expected if there were no treatment effect.
- With either the F-ratio or the t statistic, a large value provides evidence that the sample mean difference is more than would be expected by chance alone (Box 13.1).
The Logic of Analysis of Variance

• Between-Treatments Variance
  – Remember that calculating variance is simply a method for measuring how big the differences are for a set of numbers.
  – When you see the term variance, you can automatically translate it into the term differences.
  – Thus, the between-treatments variance simply measures how much difference exists between the treatment conditions.
  – In addition to measuring the differences between treatments, the overall goal of ANOVA is to interpret the differences between treatments.
The Logic of Analysis of Variance cont.

- Specifically, the purpose for the analysis is to distinguish between two alternative explanations:
  - the differences are the result of sampling error
  - the differences have been caused by the treatment effects.
- Thus, there are always two possible explanations for the difference (or variance) that exists between treatments:
  - 1. Systematic Differences Caused by the Treatments
  - 2. Random, Unsystematic Differences
    - Two primary sources are usually identified for these unpredictable differences.
The Logic of Analysis of Variance cont.

» Individual differences
» Experimental error

– Thus, when we compute the between-treatments variance, we are measuring differences that could be caused by a systematic treatment effect or could simply be random and unsystematic mean differences caused by sampling error.

– To demonstrate that there really is a treatment effect, we must establish that the differences between treatments are bigger than would be expected by sampling error alone.
The Logic of Analysis of Variance cont.

- To accomplish this goal, we will determine how big the differences are when there is no systematic treatment effect; that is, we will measure how much difference (or variance) can be explained by random and unsystematic factors.

- **To measure these differences, we compute the variance within treatments.**

  - **Within-Treatments Variance**
    - Inside each treatment condition, we have a set of individuals who all receive exactly the same treatment; that is, the researcher does not do anything that would cause these individuals to have different scores.
Thus, the within-treatments variance provides a measure of how much difference is reasonable to expect from random and unsystematic factors.

In particular, the within-treatments variance measures the naturally occurring differences that exist when there is no treatment effect; that is, how big the differences are when $H_0$ is true.

Figure 13.4 shows the overall ANOVA and identifies the sources of variability that are measured by each of the two basic components.
The F-Ratio: The Test Statistic for ANOVA

- Once we have analyzed the total variability into two basic components (between treatments and within treatments), we simply compare them.
- The comparison is made by computing a statistic called an *F-ratio*.
- *For the independent-measures ANOVA, the F-ratio has the following structure:*

$$ F = \frac{\text{variance between treatments}}{\text{variance within treatments}} = \frac{\text{differences including any treatment effects}}{\text{differences with no treatment effects}} $$ \hspace{1cm} (13.1)

- When we express each component of variability in terms of its sources (see Figure 13.4), the structure of the *F-ratio is:*

$$ F = \frac{\text{systematic treatment effects + random, unsystematic differences}}{\text{random, unsystematic differences}} $$ \hspace{1cm} (13.2)
The Logic of Analysis of Variance cont.

- The value obtained for the F-ratio helps determine whether any treatment effects exist.
- Consider the following two possibilities:
  - 1. When there are no systematic treatment effects, the differences between treatments (numerator) are entirely caused by random, unsystematic factors.
    - In this case, the numerator and the denominator of the F-ratio are both measuring random differences and should be roughly the same size.
    - With the numerator and denominator roughly equal, the F-ratio should have a value around 1.00.
The Logic of Analysis of Variance cont.

- In terms of the formula, when the treatment effect is zero, we obtain

\[ F = \frac{0 \text{ + random, unsystematic differences}}{\text{random, unsystematic differences}} \]

- Thus, an F-ratio near 1.00 indicates that the differences between treatments (numerator) are random and unsystematic, just like the differences in the denominator.

- With an F-ratio near 1.00, we conclude that there is no evidence to suggest that the treatment has any effect.
The Logic of Analysis of Variance cont.

- 2. When the treatment does have an effect, causing systematic differences between samples, then the combination of systematic and random differences in the numerator should be larger than the random differences alone in the denominator.

  - In this case, the numerator of the F-ratio should be noticeably larger than the denominator, and we should obtain an F-ratio noticeably larger than 1.00.

  - Thus, a large F-ratio is evidence for the existence of systematic treatment effects; that is, there are significant differences between treatments.
The Logic of Analysis of Variance cont.

- In more general terms, the denominator of the F-ratio measures only random and unsystematic variability.
- For this reason, the denominator of the F-ratio is called the error term.
- Definition: For ANOVA, the denominator of the F-ratio is called the error term.
- The error term provides a measure of the variance due to random, unsystematic differences.
- When the treatment effect is zero *(Ho is true)*, *the error term* measures the same sources of variance as the numerator of the F-ratio, so the value of the F-ratio is expected to be nearly equal to 1.00.
ANOVA Notation and Formulas

- Because ANOVA most often is used to examine data from more than two treatment conditions (and more than two samples), we need a notational system to help keep track of all the individual scores and totals.
  
  1. The letter k is used to identify the number of treatment conditions—that is, the number of levels of the factor.
     - For an independent-measures study, k also specifies the number of separate samples.
     - For the data in Table 13,2, there are three treatments, so k = 3.
  
  2. The number of scores in each treatment is identified by a lowercase letter n. For the example in Table 13,2, n = 5 for all the treatments.
ANOVA Notation and Formulas cont.

- If the samples are of different sizes, you can identify a specific sample by using a subscript.
  - 3. The total number of scores in the entire study is specified by a capital letter \( N \).
    - When all the samples are the same size (\( n \) is constant), \( N = kn \).
    - For the data in Table 13,2, there are \( n = 5 \) scores in each of the \( k = 3 \) treatments, so we have a total of \( N = 3(5) = 15 \) scores in the entire study.
  - 4. The sum of the scores (\( \Sigma X \)) for each treatment condition is identified by the capital letter \( T \) (for treatment total).
ANOVA Notation and Formulas

- The total for a specific treatment can be identified by adding a numerical subscript to the T.
- For example, the total for the second treatment in Table 13.2 is $T_2 = 20$.

- 5. The sum of all the scores in the research study (the grand total) is identified by G.
  - You can compute G by adding up all of the treatment totals: $G = \sum T$.

- 6. Although there is no new notation involved, we also have computed SS and M for each sample, and we have calculated $\sum X^2$ for the entire set of $N = 15$ scores in the study.
ANOVA Formulas

• The entire process of ANOVA will require nine calculations:
  – Three values for SS, three values for df, two variances (between and within), and a final F-ratio.
• However, these nine calculations are all logically related and are all directed toward finding the final F-ratio.
  – Figure 13.5 shows the logical structure of ANOVA calculations.
The final goal for the ANOVA is an $F$-ratio

$$F = \frac{\text{Variance between treatments}}{\text{Variance within treatments}}$$

Each variance in the $F$-ratio is computed as $SS/df$

<table>
<thead>
<tr>
<th>Variance between treatments</th>
<th>Variance within treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS$ between treatments</td>
<td>$SS$ within treatments</td>
</tr>
<tr>
<td>$df$ between</td>
<td>$df$ within</td>
</tr>
</tbody>
</table>

To obtain each of the $SS$ and $df$ values, the total variability is analyzed into the two components

- **SS total**: $SS$ between, $SS$ within
- **df total**: $df$ between, $df$ within

Fig. 13-5, p. 404
Analysis of Sum of Squares

1. Total Sum of Squares, $SS_{total}$
   - As the name implies, $SS_{total}$ is the sum of squares for the entire set of $N$ scores.
   - To make this formula consistent with the ANOVA notation, we substitute the letter $G$ in place of $\sum X$ and obtain:

$$SS_{total} = \Sigma X^2 - \frac{G^2}{N}$$
Analysis of Sum of Squares cont.

- 2. Within-Treatments Sum of Squares, $SS_{within \ treatments}$
  - Now we are looking at the variability inside each of the treatment conditions.

$$SS_{within \ treatments} = \sum SS_{inside \ each \ treatment}$$
Analysis of Sum of Squares cont.

3. Between-Treatments Sum of Squares, $SS_{\text{between treatments}}$:
   - In the formula, each treatment total $(T)$ is squared and then divided by the number of scores in the treatment.
   - These values are added to produce the first term in the formula.
   - Next, the grand total $(G)$ is squared and divided by the total number of scores in the entire study to produce the second term in the formula.
   - Finally, the second term is subtracted from the first.

$$SS_{\text{between}} = \sum \frac{T^2}{n} - \frac{G^2}{N}$$
The Analysis of Degrees of Freedom

• The analysis of degrees of freedom (df) follows the same pattern as the analysis of SS.
  – First, we find df for the total set of N scores, and then we partition this value into two components:
    • Degrees of freedom between treatments
    • Degrees of freedom within treatments.
The Analysis of Degrees of Freedom

- 1. Total Degrees of Freedom, $df_{\text{total}}$
  \[ df_{\text{total}} = N - 1 \]

- 2. Within-Treatments Degrees of Freedom, $df_{\text{within}}$
  \[ df_{\text{within}} = N - k \]

- 3. Between-Treatments Degrees of Freedom, $df_{\text{between}}$
  \[ df_{\text{between}} = k - 1 \]
Calculation or Variances (MS) and the F-Ratio

- The next step in the analysis of variance procedure is to compute the variance between treatments and the variance within treatments to calculate the F-ratio (see Figure 13.S).
- In ANOVA, it is customary to use the term mean square or simply MS, in place of the term variance.
- Recall (from Chapter 4) that variance is defined as the mean of the squared deviations.
- In the same way that we use SS to stand for the sum of the squared deviations, we now will use MS to stand for the mean of the squared deviations.
Calculation or Variances (MS) and the F-Ratio cont.

- For the final F-ratio we will need an MS (variance) between treatments for the numerator and an MS (variance) within treatments for the denominator.

\[
MS_{\text{between}} = s^2_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}}
\]

\[
MS_{\text{within}} = s^2_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}}
\]
Calculation or Variances (MS) and the F-Ratio cont.

- The F-ratio simply compares these two variances:

\[ F = \frac{S^2_{\text{between}}}{S^2_{\text{within}}} = \frac{MS_{\text{between}}}{MS_{\text{within}}} \]
The Distribution of F-Ratios

• In analysis of variance, the F-ratio is constructed so that the numerator and denominator of the ratio are measuring exactly the same variance when the null hypothesis is true (see Equation 13.2).
• In this situation, we expect the value of F to be around 1.00.
• The problem now is to define precisely what we mean by "around 1.00."
• What values are considered to be close to 1.00, and what values are far away?
  – To answer this question, we need to look at all the possible F values—that is, the distribution of F-ratios.
The Distribution of F-Ratios cont.

- Before we examine this distribution in detail, you should note two obvious characteristics:
  - 1. Because F-ratios are computed from two variances (the numerator and denominator of the ratio), F values are always positive numbers.
    - Remember that variance is always positive.
  - 2. When $H_0$ is true, the numerator and denominator of the F-ratio are measuring the same variance.
    - In this case, the two sample variances should be about the same size, so the ratio should be near 1.
    - In other words, the distribution of F-ratios should pile up around 1.00.
The F Distribution Table

- For ANOVA, we expect F near 1.00 if $H_0$ is true, and we expect a large value for F if $H_0$ is not true.
As we noted previously, a significant mean difference simply indicates that the difference observed in the sample data is very unlikely to have occurred just by chance.

Thus, the term significant does not necessarily mean large, it simply means larger than expected by chance.

To provide an indication of how large the effect actually is, it is recommended that researchers report a measure of effect size in addition to the measure of significance.

For ANOVA, the simplest and most direct way to measure effect size is to compute $r^2$, the percentage of variance accounted for.
Measuring Effect Size for ANOVA cont.

• In simpler terms, $r^2$ measures how much of the differences between scores is accounted for by the differences between treatments.

\[ r^2 = \frac{SS_{between}}{SS_{total}} \]

• In published reports of ANOVA results, the percentage of variance accounted for by the treatment effect is usually called $\eta^2$ (the Greek letter eta squared) instead of using $r^2$.

• Thus, for the study in Example 13.1, $\eta^2 = 0.61$. 
Unequal Sample Sizes

• In the previous examples, all the samples were exactly the same size (equal ns).
• However, the formulas for ANOVA can be used when the sample size varies within an experiment.
• With unequal sample sizes, you must take care to be sure that each value of n is matched with the proper T value in the equations.
• You also should note that the general ANOVA procedure is most accurate when used to examine experimental data with equal sample sizes.
• Therefore, researchers generally try to plan experiments with equal ns.
Post Hoc Tests

• As noted earlier, the primary advantage of ANOVA (compared to t tests) is it allows researchers to test for significant mean differences when there are more than two treatment conditions.

• ANOVA accomplishes this feat by comparing all the individual mean differences simultaneously within a single test.

• Unfortunately, the process of combining several mean differences into a single test statistic creates some difficulty when it is time to interpret the outcome of the test.

• Specifically, when you obtain a significant F-ratio (reject $H_0$), it simply indicates that somewhere among the entire set of mean differences there is at least one that is statistically significant.
Post Hoc Tests cont.

• In other words, the overall F-ratio only tells you that a significant difference exists; it does not tell exactly which means are significantly different and which are not.

• Definition: Post hoc tests (or posttests) are additional hypothesis tests that are done after an ANOVA (assuming you reject $H_0$, and there are three or more treatments) to determine exactly which mean differences are significant and which are not.
  – Thus, Rejecting $H_0$ indicates that at least one difference exists among the treatments.
  – With $k = 3$ or more, the problem is to find where the differences are.
Post-Tests and Type I Errors

- In general, a post hoc test enables you to go back through the data and compare the individual treatments two at a time.
- In statistical terms, this is called making pairwise comparisons.
  - For example, with $k = 3$, we would compare $\mu_1$ versus $\mu_2$, then $\mu_2$ versus $\mu_3$, and then $\mu_1$ versus $\mu_3$.
- In each case, we are looking for a significant mean difference.
- The process of conducting pairwise comparisons involves performing a series of separate hypothesis tests, and each of these tests includes the risk of a Type I error.
Post-Tests and Type I Errors cont.

- As you do more and more separate tests, the risk of a Type I error accumulates and is called the experiment-wise alpha level (see Box 13.1).
- **Definition**: The experiment-wise alpha level is the overall probability of a Type I error that accumulates over a series of separate hypothesis tests.
- Typically, the experiment-wise alpha level is substantially greater than the value of alpha used for anyone of the individual tests.
Planned versus Unplanned Comparisons

• Statisticians often distinguish between what are called planned and unplanned comparisons.
• As the name implies, planned comparisons refer to specific mean differences that are relevant to specific hypotheses the researcher had in mind before the study was conducted.
Post Hoc Tests to Control the Experiment-Wise Alpha Level

• Tukey’s Test
  – Tukey's test allows you to compute a single value that determines the minimum difference between treatment means that is necessary for significance.
  – This value, called the honestly significant difference, or HSD, is then used to compare any two treatment conditions.
  – If the mean difference exceeds Tukey's HSD, you conclude that there is a significant difference between the treatments.
  – Otherwise, you cannot conclude that the treatments are significantly different.
Post Hoc Tests to Control the Experiment-Wise Alpha Level

- The formula for Tukey's HSD is:

\[ HSD = q \sqrt{\frac{MS_{within}}{n}} \]

- \( q \) = look-up on table B5 (pg 734)
  - Left column (df\(_{within}\)) or \( \sum(n - 1) \)
  - Top column (k)
  - Alpha at .05 or .01
- Tukey's test requires that the sample size, \( n \), be the same for all treatments.
- Then subtract each mean from one another and look at the differences.
- If the differences between any two means is more than the HSD value, then they can be considered to be statistically significant.
Post Hoc Tests to Control the Experiment-Wise Alpha Level

• The Scheffe Test
  – Because it uses an extremely cautious method for reducing the risk of a Type I error, the Scheffe test has the distinction of being one of the safest of all possible post hoc tests (smallest risk of a Type I error).
  – We will not be calculating this test statistic by hand in class.