

Between-Groups ANOVA

NOTE TO INSTRUCTORS

This essential chapter introduces the concept of between-groups ANOVA, one of the most frequently used statistical analyses in psychological research. Although conceptually understanding ANOVAs should not be much more difficult than understanding t tests, your students may have difficulty with the calculation component of working with ANOVAs. Therefore, two hands-on activities have been included for students to practice working with ANOVAs. It is important to give students plenty of examples to work with to help reduce any anxiety they may experience. Also, students may have difficulty understanding that their work is not completely done once they have calculated an ANOVA. If there is statistical significance, students will also need to conduct post-hoc tests. Emphasize to students that post-hoc tests are really no more complicated than t tests. The only difference is that they adjust for an increasing alpha error. Once students realize that post-hoc tests are similar to a familiar statistical concept, they should have greater confidence working with them.

OUTLINE OF RESOURCES

- I. **Using the F Distribution with Three or More Samples**
 - Discussion Question 12-1 (p. 106)
 - Discussion Question 12-2 (p. 107)
 - Discussion Question 12-3 (p. 107)
 - Discussion Question 12-4 (p. 108)
 - Discussion Question 12-5 (p. 108)
- II. **One-Way Between-Groups ANOVA**
 - Discussion Question 12-6 (p. 109)
 - Discussion Question 12-7 (p. 109)
 - Discussion Question 12-8 (p. 110)
 - Classroom Activity 12-1: Snack Food/Drink Product Testers (p. 111)
 - Classroom Activity 12-2: Yoga and Stress (p. 111)

III. Beyond Hypothesis Testing

- Discussion Question 12-9 (p. 112)
- Classroom Activity 12-3: One-Way Analysis of Variance: Singers' Height (p. 113)
- Classroom Activity 12-4: One-Way ANOVAs in Context (p. 113)

IV. Next Steps: The Bonferroni Test

- Discussion Question 12-10 (p. 114)
- Additional Readings (p. 114)
- Online Resources (p. 114)

V. Handouts and Transparency Masters

- Handout 12-1: Snack Food/Drink Product Testers (p. 115)
- Handout 12-2: Yoga and Stress (p. 116)
- Handout 12-3: One-Way ANOVA in Context (p. 117)
- Transparency Master 12-1: The Six Steps of Hypothesis Testing for One-Way Analysis of Variance: Singers' Height (p. 118)

CHAPTER GUIDE

I. Using the F Distribution with Three or More Samples

1. Tests based on the z distribution, the t distributions, and the F distributions are all variations on the normal, bell-shaped curve.
2. When calculating the statistic for these tests, we are essentially dividing a numerator that represents the difference between groups by a denominator that represents the variability within groups.
3. When we want to compare three or more groups, we use an F distribution and conduct an analysis of variance (**ANOVA**).
4. It is a hypothesis test typically used with one or more nominal independent variables (with at least three groups overall) and a scale dependent variable.

> Discussion Question 12-1

What is an ANOVA, and how is different from a z test or t test?

Your students' answers should include:

- An analysis of variance (ANOVA) is a hypothesis test that is used to compare group mean differences categorized by one independent variable (with three or more levels) on an interval dependent variable.
 - It is different from a z test or t test because it allows you to compare more than two groups.
5. The **F statistic** is a ratio of two measures of variance—between-groups variance and within-groups variance. **Between-groups variance** is an

estimate of the population variance based on the differences among the means. In contrast, **within-groups variance** is an estimate of the population variance based on the differences within each of the three (or more) sample distributions.

6. If the between-groups variance is much larger than the within-groups variance, we can infer that the sample means are different from one another.

> Discussion Question 12-2

What is the difference between within-groups variance and between-groups variance?

Your students' answers should include:

- The difference between within-groups variance and between-groups variance is that within-groups variance refers to the variability among scores for subjects in the same group, compared to between-groups variance, which refers to the variability among scores for subjects in different groups.
7. The F table is basically an expansion of the t table. Just as there are many t distributions based on sample size, there are many F distributions in the F table. Similar to the t table, the F table includes several extreme probabilities and the full range of sample sizes, represented by the degrees of freedom.
 8. A **one-way ANOVA** is a hypothesis test that includes one nominal independent variable with more than two levels and an interval dependent variable.
 9. A one-way ANOVA can have one of two research designs. In a **within-groups ANOVA** (or repeated-measures ANOVA), there are more than two samples and each sample comprises the same participants. The second type, a **between-groups ANOVA**, is when there are more than two samples and each sample is composed of different participants.

> Discussion Question 12-3

What is the difference between a within-groups and a between-groups ANOVA? What would be an example of each type of design?

Your students' answers should include:

- The difference between a within-groups and between-groups ANOVA is that, in within-groups ANOVAs, all participants are in all levels of the independent variable, and in between-groups ANOVAs, participants can only be in one level of the independent variable.
- An example of a within-groups design would be a study comparing the effects of stress on memory performance (dependent variable). There is one independent variable (stress) with three levels (heat stress, cold stress,

and no stress) and participants are in all three levels of the independent variable.

- An example of a between-groups design is a study comparing the effectiveness of three different fertilizers on rate of plant growth (dependent variable). There is one independent variable (fertilizer) with three levels (fertilizer A, fertilizer B, and fertilizer C), and participants are in only one level of the independent variable.

10. Regardless of the type used, they share the same assumptions. The first assumption is that the sample is randomly selected from all members of the population. Second, the population distributions for the samples should be normal. Lastly, the population from which the samples are drawn should have the same spread, usually as measured by variance. As with z and t tests, the ANOVA is robust against these assumptions.

> Discussion Question 12-4

What three assumptions do you make when conducting an ANOVA?

Your students' answers should include:

- The three assumptions you make when conducting an ANOVA are that:
 - the sample is randomly selected from the population of interest,
 - the underlying population is normally distributed, and
 - the population from which the samples are drawn have equal variance.

11. The third assumption for an ANOVA is called homoscedasticity (also called homogeneity of variance). **Homoscedastic** populations have the same variance. In contrast, **heteroscedastic** populations have different variances.

> Discussion Question 12-5

What are homoscedastic populations? What are heteroscedastic populations?

Your students' answers should include:

- Homoscedastic populations are populations that have equal variances.
- Heteroscedastic populations are populations with different variances.

II. One-Way Between-Groups ANOVA

1. As with our other tests, there are six steps of hypothesis testing in the context of a one-way between-groups ANOVA.
2. We first identify the populations, distributions, and assumptions.
3. Second, we state the null and research hypotheses. However, in the case of an ANOVA, the null hypothesis is slightly different because we can reject the null hypothesis in an ANOVA, even if only one group, on average, is different from the others.

> Discussion Question 12-6

Why is the null hypothesis slightly different for an ANOVA compared to a t test or a z test?

Your students' answers should include:

- The null hypothesis is slightly different for an ANOVA compared to a t test or a z test because, for an ANOVA, you can reject the null hypothesis on the basis that at least one group mean is different from the others.
4. Third, we determine the characteristics of the comparison distribution. There are two sets of degrees of freedom for an ANOVA. To calculate the degrees of freedom for the between-groups variance estimate ($df_{between}$), we subtract one from the number of samples. We determine the degrees of freedom for the within-groups variance (df_{within}) by subtracting one from the number of individuals in each sample, and then sum the number of all of the samples.

> Discussion Question 12-7

How do we calculate the $df_{between}$ and the df_{within} ?

Your students' answers should include:

- Calculate the $df_{between}$ by subtracting the number of groups from 1 ($N_{groups} - 1$). The df_{within} is calculated by summing the degrees of freedom for each group. The degree of freedom for each group is calculated by subtracting 1 from the number of people in that group.
5. The fourth step is to determine the critical values, or cutoffs. To find the critical value, we will need to use the between-groups and within-groups degrees of freedom. When using the F table, first find the appropriate within-groups degrees of freedom along the left side of the page. Then find the appropriate between-groups degrees of freedom along the top.
 6. The fifth step is to calculate the test statistic. The F statistic is simply an estimate of between-groups variability divided by an estimate of within-groups variability.
 7. Our last step is to make a decision to reject or fail to reject the null hypothesis.
 8. The value of the F statistic will be large when the means of the groups are far apart (between-groups variability is large) and the scores within each group are not very spread out (within-groups variability is small). If both of our estimates of variance are the same, our F statistic will be 1.0, which would mean that there is no difference between our groups. This is different from z or t tests, where no difference is represented by a value of zero.
 9. When conducting an ANOVA, we use a **source table** to present the important calculations and final results of an ANOVA in a consistent and easy-to-read format.
 10. The first column is labeled *source* and lists the sources or origins of the

two estimates of population variance. One source comes from the spread between means and the other comes from the spread among scores within each sample. There is also a *total* row.

11. In our fifth column, we label F , which is determined by dividing the estimate of between-groups variance by the estimate of the within-groups variance.
12. The fourth column, MS , describes how we arrived at those estimates. MS stands for “mean square” because variance is the arithmetic mean of the squared deviations. $MS_{between}$ and MS_{within} refer to between-groups variance and within-groups variance, respectively, and we divide $MS_{between}$ by MS_{within} to get F .
13. The third column is labeled df and shows the degrees of freedom. The degrees of freedom between and degrees of freedom within are each written in the appropriate row, and the degrees of freedom total is obtained by adding $df_{between}$ and df_{within} .
14. Column two is labeled SS and includes the three sums of squares— $SS_{between}$, SS_{within} , and SS_{total} . As with the df column, SS_{total} is obtained by adding $SS_{between}$ and SS_{within} . The variance estimates are obtained by dividing the SS with its matching df .
15. The **grand mean** (GM) is the mean of every score in a study, regardless of which sample it came from, and is symbolized as: The formula for SS_{total} is: $SS_{total} = \Sigma(X - GM)^2$.
16. For the within-groups sum of squares, we create deviations for each set of sample scores, but the deviations are around the mean of that particular group. Once we have all of our deviations, we square them and sum them. The formula for the within-groups sum of squares is: $SS_{within} = \Sigma(X - M)^2$.
17. For the between-groups sum of squares, the grand mean is subtracted from the mean of each group. The individual scores are never involved—only the sample means and the grand mean are involved. The formula for the $SS_{between}$ is: $SS_{between} = \Sigma(M - GM)^2$.

> Discussion Question 12-8

How do you calculate the $SS_{between}$ and SS_{within} ? How do you use this information to determine the test statistic?

Your students' answers should include:

- Calculate the $SS_{between}$ by subtracting the grand mean from the appropriate sample mean of each group, and calculate the SS_{within} by subtracting the sample mean, rather than the grand mean, from each score.
- To use this information to determine the test statistic: Once you add your within-groups sum of squares to your between-groups sum of squares to see if they equal the total sum of squares, you can insert these numbers into your source table to calculate your F statistic.

18. A statistically significant ANOVA tells us that something important is going on somewhere in the study, but it does not specify which pairs of means are responsible for the statistically significant difference between groups.
19. Once we have a statistically significant ANOVA, we use a **post-hoc test**, which is a statistical procedure frequently carried out after we reject the null hypothesis in an analysis of variance.

Classroom Activity 12-1

Snack Food/Drink Product Testers

This activity will give students hands-on practice calculating and interpreting one-way ANOVAs.

- Divide the class into three groups. Each group should have one particular type or brand of food or drink.
- Have students rate their item based on a variable that can be decided by the instructor or the class (e.g., tastiness, thirst quenching, satisfying, etc.). For example, each student could be given one of three types of soft drinks and would rate the drink on degree of refreshment.
- Depending on the size of the class, either students can be given descriptive statistics about each of the three groups or they can calculate the descriptives themselves in order to calculate a one-way ANOVA.

Use Handout 12-1, found at the end of this chapter, to aid in this process.

Classroom Activity 12-2

Yoga and Stress

In recent years, yoga has become very popular because of its supposed stress-reducing properties. In this activity, the class will determine whether yoga does indeed relieve stress (at least within your class).

- Divide students into three groups:
 - Group One will do no physical activity for the following week.
 - Group Two will do 30 minutes of yoga each day for the following week.
 - Group Three will do 30 minutes of some other physical activity (e.g., walking) each day for the next week.
- At the end of the week, have students rate their stress levels on a scale from 1 (not at all stressed) to 7 (extremely stressed).
- Depending on the size of the class, either students can be given descriptive statistics about each of the three groups' ratings or they can calculate the descriptives themselves so that they can then calculate a one-way ANOVA.

Use Handout 12-2, found at the end of this chapter, for this activity.

III. Beyond Hypothesis Testing

1. When using an ANOVA to look at differences between three or more means, we use R^2 to calculate effect size. R^2 is the proportion of variance in the dependent variable that is accounted for by the independent variable. It is often symbolized as η^2 , pronounced "eta squared."

2. R^2 is calculated using the formula $R^2 = SS_{between}/SS_{total}$.
3. There are also conventions for R^2 to let us know whether the effect is small, medium, or large. An R^2 of .01 is considered small, an R^2 of .06 is considered medium, and an R^2 of .14 is considered large.
4. When we are faced with multiple comparisons guided by an existing theory or a previous finding, we can use *a priori* comparisons or **planned comparisons**, a test that is conducted when there are multiple groups of scores, but specific comparisons have been specified prior to data collection.
5. With planned comparisons, the researcher states the exact comparisons that he or she will be making before collecting the data and can choose to conduct one or more independent-samples *t* tests with a *p* level of 0.05 or one with a more conservative *p* value such as a Bonferroni test.
6. When we do not have a smaller set of comparisons established before collecting the data, we must use the more conservative **post-hoc test**.
7. A **post-hoc test** is a statistical procedure frequently carried out after we reject the null hypothesis in an analysis of variance. It allows us to make multiple comparisons among several means.
8. One of the most common post-hoc tests is the **Tukey HSD test**, which determines the differences between means in terms of standard error. The *HSD* is compared to a critical value.
9. The Tukey *HSD* test involves the calculation of differences between each pair of means and the division of each difference by the standard error.
10. We compare the *HSD* for each pair of means to a critical value, *q*, using a table. *HSD* is calculated using the formula:

$HSD = \frac{(M_1 - M_2)}{s_M}$ for any two sample means. The formula for the

standard error is $s_M = \sqrt{\frac{(MS_{within})}{N}}$, where *N* is the sample size within each group.

11. If we have different sample sizes, we need to calculate a weighted sample size known as the harmonic mean which is calculated by $N' = \frac{N_{groups}}{\Sigma\left(\frac{1}{N}\right)}$. If the sample sizes are not equal, the formula for s_M uses

N' instead of *N*.

> Discussion Question 12-9

What are post-hoc tests? Why are they necessary? What is a Tukey HSD test?

Your students' answers should include:

- Post-hoc tests are statistical procedures frequently conducted after rejecting the null hypothesis in an ANOVA, which allows you to make multiple comparisons among several means and helps you avoid increases in Type I errors.

- The **Tukey *HSD* test** is a widely used post-hoc test that determines the differences between means in terms of standard error. It is compared to a critical value.

Classroom Activity 12-3

One-Way Analysis of Variance: Singers' Height

Here you will find a dataset of which your class can do several types of analyses, including a one-way analysis of variance at the following Web site: <http://lib.stat.cmu.edu/DASL/Datafiles/Singers.html>.

- Cut and paste the data since they are not in SPSS or a downloadable format. The independent measure is singing voice (with four levels), while the dependent measure is height in inches (self-reported). The Web site also provides examples of box plots.
- Complete Tukey's *HSD* on the data and note the differences in results.

Complete a series of independent sample *t* tests. Note that the results of ANOVA and post-hoc tests and *t* tests are the same when each comparison is completed. Use Transparency Master 12-1, found at the end of this chapter. (You may display the transparency on an overhead projector by photocopying it onto acetate, or you may use PowerPoint by scanning the transparency master into your computer.)

- Also, this is a good time to review the spiraling alpha error problem.
- Save this data file—with a minor update you can use it to demonstrate a two-way analysis of variance for Chapter 14.

The main DASL link is <http://lib.stat.cmu.edu/DASL/DataArchive.html>, where you will find many useful data files with brief scenario and background information.

Classroom Activity 12-4

One-Way ANOVAs in Context

Have students read the article: Storbeck, J., and Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory. *Psychological Science*, 16(10), 785–791. (To view or purchase this article, go to your local library or visit Blackwell Publishing online at <http://www.blackwellpublishing.com>.)

- First, have students try to identify the one-way ANOVA (found in experiment 1).
- Next, have students interpret the findings from the ANOVA.
- Alternatively, you could present students with a different article that also contains at least one straightforward one-way ANOVA and have students interpret that article instead.

Use Handout 12-3, found at the end of this chapter, to aid your students in understanding the analysis in context.

IV. Next Steps: The Bonferroni Test

1. The **Bonferroni test** is a post-hoc test that provides a more strict critical value for every comparison of means.
2. In the Bonferroni test, we divide the p level by the number of comparisons. In many cases when using a Bonferroni test, the difference between the two means must be quite extreme and we may fail to detect real differences that are not quite extreme enough, thus committing a Type II error.

> Discussion Question 12-10

What is the difference between a Tukey HSD test and the Bonferroni test?

Your students' answers should include:

- Both the Tukey *HSD* and Bonferroni tests are post-hoc tests, but the Bonferroni test provides a more conservative test, meaning that we may fail to detect real differences that are not quite extreme enough.
3. When conducting hypothesis tests, we are always striking a balance between reducing Type I and Type II errors.

Additional Readings

Fisher, Ronald A. (1925). *Statistical Methods for Research Workers*. Originally published in London by Oliver and Boyd. Oxford University Press, 1990.

Fisher is generally credited with developing the analysis of variance statistical technique, and this book contains the first textbook presentation of it. This is the original reference for the work.

Tukey, John W. (1977). *Exploratory Data Analysis*. New York: Addison-Wesley.

This is a classic book about data analysis. Although it was written in the precomputer age and therefore does not include computer computations, it still clearly expresses the ideas behind the computations and is well worth the read.

Online Resources

This site provides a visual interactive demonstration to help explain how variability between and within groups contributes to the F statistic: <http://www.psych.utah.edu/stat/introstats/anovaflash.html>. This can be useful in a class demonstration. The demonstration requires FLASH software.

TRANSPARENCY MASTER 12-1
(for use with Classroom Activity 12-3)

The Six Steps of Hypothesis Testing for One-Way Analysis of Variance: Singers' Height

Step 1. Populations, distributions, and assumptions.

The populations are sopranos, altos, tenors, and basses, all members in the NY Choral Society in 1979.

The comparison is the distribution and will be the F distribution.

The hypothesis test is the one-way between-groups ANOVA.

The data are not from random samples, so we must be cautious about generalizations.

Step 2. State the null and research hypotheses.

H_0 There will be no difference in singers' height.

H_1 There will be a difference in singers' height.

Step 3. Determine the characteristics of the comparison distributions.

We are using the F distribution: $df_{between}$ is 3; df_{within} is 125.

Step 4. Determine the critical values, or cutoffs:

$p < 0.05$; the degrees of freedom are 3125.

Step 5. Calculate the test statistics.

$F = \text{between groups variability} / \text{within groups variability}$.

Step 6. Make a decision.