

# CK-12 Probability and Statistics - Advanced (Second Edition)

---

Ellen Lawsky  
Larry Ottman  
Raja Almukkahal  
Brenda Meery  
Danielle DeLancey

## Chapter 11

Analysis of Variance  
and the F-Distribution

Say Thanks to the Authors

Click <http://www.ck12.org/saythanks>

*(No sign in required)*



To access a customizable version of this book, as well as other interactive content, visit [www.ck12.org](http://www.ck12.org)

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the **FlexBook®**, CK-12 intends to pioneer the generation and distribution of high-quality educational content that will serve both as core text as well as provide an adaptive environment for learning, powered through the **FlexBook Platform®**.

Copyright © 2014 CK-12 Foundation, [www.ck12.org](http://www.ck12.org)

The names “CK-12” and “CK12” and associated logos and the terms “**FlexBook®**” and “**FlexBook Platform®**” (collectively “CK-12 Marks”) are trademarks and service marks of CK-12 Foundation and are protected by federal, state, and international laws.

Any form of reproduction of this book in any format or medium, in whole or in sections must include the referral attribution link <http://www.ck12.org/saythanks> (placed in a visible location) in addition to the following terms.

Except as otherwise noted, all CK-12 Content (including CK-12 Curriculum Material) is made available to Users in accordance with the Creative Commons Attribution-Non-Commercial 3.0 Unported (CC BY-NC 3.0) License (<http://creativecommons.org/licenses/by-nc/3.0/>), as amended and updated by Creative Commons from time to time (the “CC License”), which is incorporated herein by this reference.

Complete terms can be found at <http://www.ck12.org/terms>.

Printed: December 17, 2014

**flexbook**  
next generation textbooks



## AUTHORS

Ellen Lawsby  
Larry Ottman  
Raja Almukkahal  
Brenda Meery  
Danielle DeLancey

---

CHAPTER

**11**

# Analysis of Variance and the F-Distribution

## Chapter Outline

---

- 11.1 THE F-DISTRIBUTION AND TESTING TWO VARIANCES
  - 11.2 THE ONE-WAY ANOVA TEST
  - 11.3 THE TWO-WAY ANOVA TEST
-

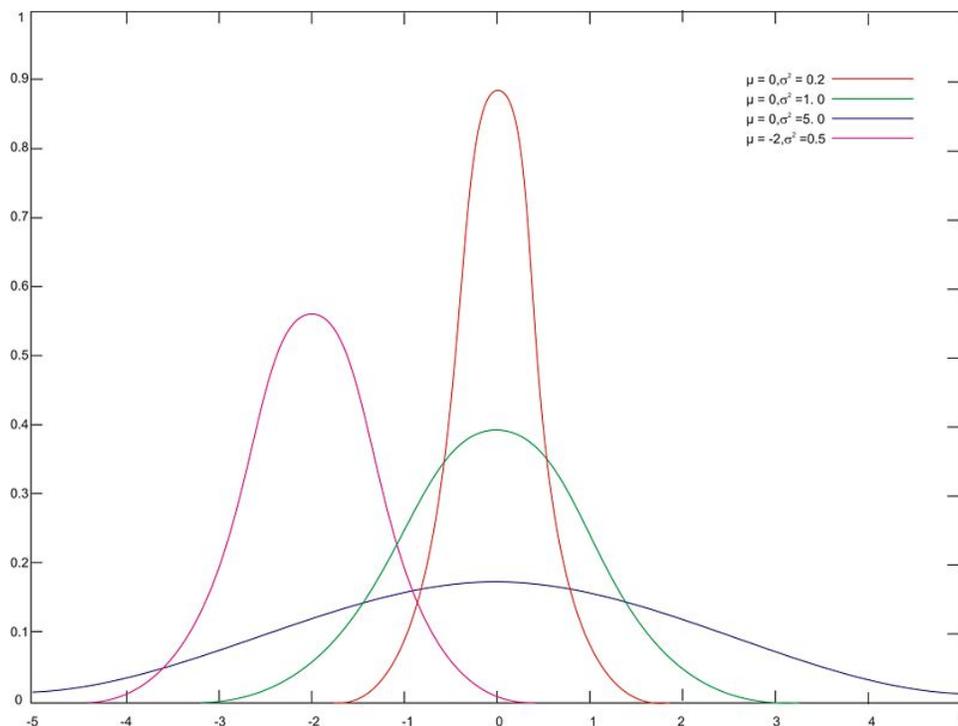
# 11.1 The F-Distribution and Testing Two Variances

## Learning Objectives

- Understand the differences between the  $F$ -distribution and Student's  $t$ -distribution.
- Calculate a test statistic as a ratio of values derived from sample variances.
- Use random samples to test hypotheses about multiple independent population variances.
- Understand the limits of inferences derived from these methods.

## Introduction

In previous lessons, we learned how to conduct hypothesis tests that examined the relationship between two variables. Most of these tests simply evaluated the relationship of the means of two variables. However, sometimes we also want to test the variance, or the degree to which observations are spread out within a distribution. In the figure below, we see three samples with identical means (the samples in red, green, and blue) but with very different variances:



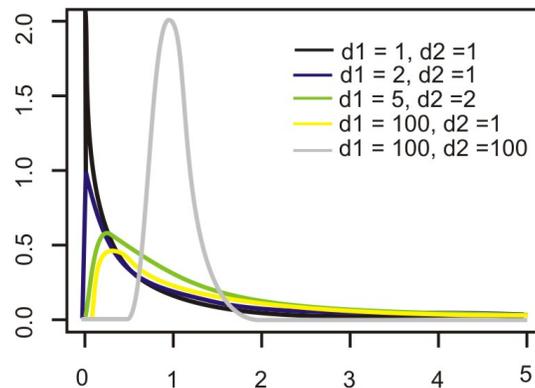
So why would we want to conduct a hypothesis test on variance? Let's consider an example. Suppose a teacher wants to examine the effectiveness of two reading programs. She randomly assigns her students into two groups, uses a different reading program with each group, and gives her students an achievement test. In deciding which reading program is more effective, it would be helpful to not only look at the mean scores of each of the groups, but

also the “spreading out” of the achievement scores. To test hypotheses about variance, we use a statistical tool called the  $F$ -distribution.

In this lesson, we will examine the difference between the  $F$ -distribution and Student’s  $t$ -distribution, calculate a test statistic with the  $F$ -distribution, and test hypotheses about multiple population variances. In addition, we will look a bit more closely at the limitations of this test.

## The

The  $F$ -distribution is actually a family of distributions. The specific  $F$ -distribution for testing two population variances,  $\sigma_1^2$  and  $\sigma_2^2$ , is based on two values for degrees of freedom (one for each of the populations). Unlike the normal distribution and the  $t$ -distribution,  $F$ -distributions are not symmetrical and span only non-negative numbers. (Normal distributions and  $t$ -distributions are symmetric and have both positive and negative values.) In addition, the shapes of  $F$ -distributions vary drastically, especially when the value for degrees of freedom is small. These characteristics make determining the critical values for  $F$ -distributions more complicated than for normal distributions and Student’s  $t$ -distributions.  $F$ -distributions for various degrees of freedom are shown below:



We use the  $F$ -ratio test statistic when testing the hypothesis that there is no difference between population variances. When calculating this ratio, we really just need the variance from each of the samples. It is recommended that the larger sample variance be placed in the numerator of the  $F$ -ratio and the smaller sample variance in the denominator. By doing this, the ratio will always be greater than 1.00 and will simplify the hypothesis test.

*Example:* Suppose a teacher administered two different reading programs to two groups of students and collected the following achievement score data:

Program 1

$$n_1 = 31$$

$$\bar{x}_1 = 43.6$$

$$s_1^2 = 105.96$$

Program 2

$$n_2 = 41$$

$$\bar{x}_2 = 43.8$$

$$s_2^2 = 36.42$$

What is the  $F$ -ratio for these data?

$$F = \frac{s_1^2}{s_2^2} = \frac{105.96}{36.42} \approx 2.909$$

When we test the hypothesis that two variances of populations from which random samples were selected are equal,  $H_0 : \sigma_1^2 = \sigma_2^2$  (or in other words, that the ratio of the variances  $\frac{\sigma_1^2}{\sigma_2^2} = 1$ ), we call this test the  $F$ -Max test. Since we have a null hypothesis of  $H_0 : \sigma_1^2 = \sigma_2^2$ , our alternative hypothesis would be  $H_a : \sigma_1^2 \neq \sigma_2^2$ .

Establishing the critical values in an  $F$ -test is a bit more complicated than when doing so in other hypothesis tests. Most tables contain multiple  $F$ -distributions, one for each of the following: 1 percent, 5 percent, 10 percent, and 25 percent of the area in the right-hand tail. (Please see the supplemental link for an example of this type of table.) We also need to use the degrees of freedom from each of the samples to determine the critical values.

### On the Web

<http://www.statsoft.com/textbook/sttable.html#f01>  $F$ -distribution tables.

*Example:* Suppose we are trying to determine the critical values for the scenario in the preceding section, and we set the level of significance to 0.02. Because we have a two-tailed test, we assign 0.01 to the area to the right of the positive critical value. Using the  $F$ -table for  $\alpha = 0.01$ , we find the critical value at 2.203, since the numerator has 30 degrees of freedom and the denominator has 40 degrees of freedom.

Once we find our critical values and calculate our test statistic, we perform the hypothesis test the same way we do with the hypothesis tests using the normal distribution and Student's  $t$ -distribution.

*Example:* Using our example from the preceding section, suppose a teacher administered two different reading programs to two different groups of students and was interested if one program produced a greater variance in scores. Perform a hypothesis test to answer her question.

For the example, we calculated an  $F$ -ratio of 2.909 and found a critical value of 2.203. Since the observed test statistic exceeds the critical value, we reject the null hypothesis. Therefore, we can conclude that the observed ratio of the variances from the independent samples would have occurred by chance if the population variances were equal less than 2% of the time. We can conclude that the variance of the student achievement scores for the second sample is less than the variance of the scores for the first sample. We can also see that the achievement test means are practically equal, so the difference in the variances of the student achievement scores may help the teacher in her selection of a program.

### The Limits of Using the

The test of the null hypothesis,  $H_0 : \sigma_1^2 = \sigma_2^2$ , using the  $F$ -distribution is only appropriate when it can safely be assumed that the population is normally distributed. If we are testing the equality of standard deviations between two samples, it is important to remember that the  $F$ -test is extremely sensitive. Therefore, if the data displays even small departures from the normal distribution, including non-linearity or outliers, the test is unreliable and should not be used. In the next lesson, we will introduce several tests that we can use when the data are not normally distributed.

---

## Lesson Summary

We use the  $F$ -Max test and the  $F$ -distribution when testing if two variances from independent samples are equal.

The  $F$ -distribution differs from the normal distribution and Student's  $t$ -distribution. Unlike the normal distribution and the  $t$ -distribution,  $F$ -distributions are not symmetrical and go from 0 to  $\infty$ , not from  $-\infty$  to  $\infty$  as the others do.

When testing the variances from independent samples, we calculate the  $F$ -ratio test statistic, which is the ratio of the variances of the independent samples.

When we reject the null hypothesis,  $H_0 : \sigma_1^2 = \sigma_2^2$ , we conclude that the variances of the two populations are not equal.

The test of the null hypothesis,  $H_0 : \sigma_1^2 = \sigma_2^2$ , using the  $F$ -distribution is only appropriate when it can be safely assumed that the population is normally distributed.

---

## Review Questions

1. We use the  $F$ -Max test to examine the differences in the \_\_\_ between two independent samples.
2. List two differences between the  $F$ -distribution and Student's  $t$ -distribution.
3. When we test the differences between the variances of two independent samples, we calculate the \_\_\_.
4. When calculating the  $F$ -ratio, it is recommended that the sample with the \_\_\_ sample variance be placed in the numerator, and the sample with the \_\_\_ sample variance be placed in the denominator.
5. Suppose a guidance counselor tested the mean of two student achievement samples from different SAT preparatory courses. She found that the two independent samples had similar means, but also wants to test the variance associated with the samples. She collected the following data:

SAT Prep Course #1

$$n = 31$$

$$s^2 = 42.30$$

SAT Prep Course #2

$$n = 21$$

$$s^2 = 18.80$$

- (a) What are the null and alternative hypotheses for this scenario?
  - (b) What is the critical value with  $\alpha = 0.10$ ?
  - (c) Calculate the  $F$ -ratio.
  - (d) Would you reject or fail to reject the null hypothesis? Explain your reasoning.
  - (e) Interpret the results and determine what the guidance counselor can conclude from this hypothesis test.
6. True or False: The test of the null hypothesis,  $H_0 : \sigma_1^2 = \sigma_2^2$ , using the  $F$ -distribution is only appropriate when it can be safely assumed that the population is normally distributed.

## 11.2 The One-Way ANOVA Test

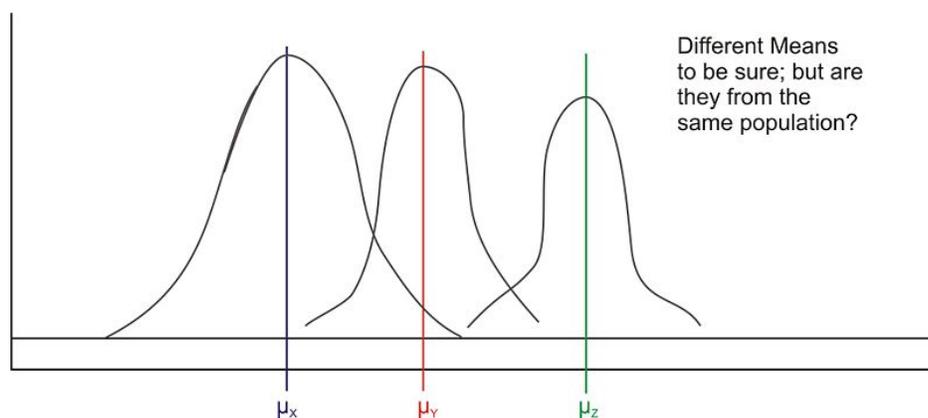
### Learning Objectives

- Understand the shortcomings of comparing multiple means as pairs of hypotheses.
- Understand the steps of the ANOVA method and the method's advantages.
- Compare the means of three or more populations using the ANOVA method.
- Calculate pooled standard deviations and confidence intervals as estimates of standard deviations of populations.

### Introduction

Previously, we have discussed analyses that allow us to test if the means and variances of two populations are equal. Suppose a teacher is testing multiple reading programs to determine the impact on student achievement. There are five different reading programs, and her 31 students are randomly assigned to one of the five programs. The mean achievement scores and variances for the groups are recorded, along with the means and the variances for all the subjects combined.

We could conduct a series of  $t$ -tests to determine if all of the sample means came from the same population. However, this would be tedious and has a major flaw, which we will discuss shortly. Instead, we use something called the Analysis of Variance (ANOVA), which allows us to test the hypothesis that multiple population means and variances of scores are equal. Theoretically, we could test hundreds of population means using this procedure.



### Shortcomings of Comparing Multiple Means Using Previously Explained Methods

As mentioned, to test whether pairs of sample means differ by more than we would expect due to chance, we could conduct a series of separate  $t$ -tests in order to compare all possible pairs of means. This would be tedious, but we could use a computer or a TI-83/84 calculator to compute these quickly and easily. However, there is a major flaw with this reasoning.

When more than one  $t$ -test is run, each at its own level of significance, the probability of making one or more type I errors multiplies exponentially. Recall that a type I error occurs when we reject the null hypothesis when we should not. The level of significance,  $\alpha$ , is the probability of a type I error in a single test. When testing more than one pair of samples, the probability of making at least one type I error is  $1 - (1 - \alpha)^c$ , where  $\alpha$  is the level of significance for each  $t$ -test and  $c$  is the number of independent  $t$ -tests. Using the example from the introduction, if our teacher conducted separate  $t$ -tests to examine the means of the populations, she would have to conduct 10 separate  $t$ -tests. If she performed these tests with  $\alpha = 0.05$ , the probability of committing a type I error is not 0.05 as one would initially expect. Instead, it would be 0.40, which is extremely high!

### The Steps of the ANOVA Method

With the *ANOVA method*, we are actually analyzing the total variation of the scores, including the variation of the scores within the groups and the variation between the group means. Since we are interested in two different types of variation, we first calculate each type of variation independently and then calculate the ratio between the two. We use the  $F$ -distribution as our sampling distribution and set our critical values and test our hypothesis accordingly.

When using the ANOVA method, we are testing the null hypothesis that the means and the variances of our samples are equal. When we conduct a hypothesis test, we are testing the probability of obtaining an extreme  $F$ -statistic by chance. If we reject the null hypothesis that the means and variances of the samples are equal, and then we are saying that the difference that we see could not have happened just by chance.

To test a hypothesis using the ANOVA method, there are several steps that we need to take. These include:

1. Calculating the *mean squares between groups*,  $MS_B$ . The  $MS_B$  is the difference between the means of the various samples. If we hypothesize that the group means are equal, then they must also equal the population mean. Under our null hypothesis, we state that the means of the different samples are all equal and come from the same population, but we understand that there may be fluctuations due to sampling error. When we calculate the  $MS_B$ , we must first determine the  $SS_B$ , which is the sum of the differences between the individual scores and the mean in each group. To calculate this sum, we use the following formula:

$$SS_B = \sum_{k=1}^m n_k (\bar{x}_k - \bar{x})^2$$

where:

$k$  is the group number.

$n_k$  is the sample size of group  $k$ .

$\bar{x}_k$  is the mean of group  $k$ .

$\bar{x}$  is the overall mean of all the observations.

$m$  is the total number of groups.

When simplified, the formula becomes:

$$SS_B = \sum_{k=1}^m \frac{T_k^2}{n_k} - \frac{T^2}{n}$$

where:

$T_k$  is the sum of the observations in group  $k$ .

$T$  is the sum of all the observations.

$n$  is the total number of observations.

Once we calculate this value, we divide by the number of degrees of freedom,  $m - 1$ , to arrive at the  $MS_B$ . That is,  $MS_B = \frac{SS_B}{m-1}$

2. Calculating the *mean squares within groups*,  $MS_W$ . The mean squares within groups calculation is also called the *pooled estimate of the population variance*. Remember that when we square the standard deviation of a sample, we are estimating population variance. Therefore, to calculate this figure, we sum the squared deviations within each group and then divide by the sum of the degrees of freedom for each group.

To calculate the  $MS_W$ , we first find the  $SS_W$ , which is calculated using the following formula:

$$\frac{\sum(x_{i1} - \bar{x}_1)^2 + \sum(x_{i2} - \bar{x}_2)^2 + \dots + \sum(x_{im} - \bar{x}_m)^2}{(n_1 - 1) + (n_2 - 1) + \dots + (n_m - 1)}$$

Simplified, this formula becomes:

$$SS_W = \sum_{k=1}^m \sum_{i=1}^{n_k} x_{ik}^2 - \sum_{k=1}^m \frac{T_k^2}{n_k}$$

where:

$T_k$  is the sum of the observations in group  $k$ .

Essentially, this formula sums the squares of each observation and then subtracts the total of the observations squared divided by the number of observations. Finally, we divide this value by the total number of degrees of freedom in the scenario,  $n - m$ .

$$MS_W = \frac{SS_W}{n - m}$$

3. Calculating the test statistic. The formula for the test statistic is as follows:

$$F = \frac{MS_B}{MS_W}$$

4. Finding the critical value of the  $F$ -distribution. As mentioned above,  $m - 1$  degrees of freedom are associated with  $MS_B$ , and  $n - m$  degrees of freedom are associated with  $MS_W$ . In a table, the degrees of freedom for  $MS_B$  are read across the columns, and the degrees of freedom for  $MS_W$  are read across the rows.

5. Interpreting the results of the hypothesis test. In ANOVA, the last step is to decide whether to reject the null hypothesis and then provide clarification about what that decision means.

The primary advantage of using the ANOVA method is that it takes all types of variations into account so that we have an accurate analysis. In addition, we can use technological tools, including computer programs, such as SAS, SPSS, and Microsoft Excel, as well as the TI-83/84 graphing calculator, to easily perform the calculations and test our hypothesis. We use these technological tools quite often when using the ANOVA method.

*Example:* Let's go back to the example in the introduction with the teacher who is testing multiple reading programs to determine the impact on student achievement. There are five different reading programs, and her 31 students are randomly assigned to one of the five programs. She collects the following data:

Method

1	2	3	4	5
1	8	7	9	10
4	6	6	10	12
3	7	4	8	9
2	4	9	6	11
5	3	8	5	8
1	5	5		
6		7		
		5		

Compare the means of these different groups by calculating the mean squares between groups, and use the standard deviations from our samples to calculate the mean squares within groups and the pooled estimate of the population variance.

To solve for  $SS_B$ , it is necessary to calculate several summary statistics from the data above:

Number ( $n_k$ )	7	6	8	5	5	31
Total ( $T_k$ )	22	33	51	38	50	= 194
Mean ( $\bar{x}$ )	3.14	5.50	6.38	7.60	10.00	= 6.26
Sum of Squared Obs. $\left(\sum_{i=1}^{n_k} x_{ik}^2\right)$	92	199	345	306	510	= 1,452
$\frac{\text{Sum of Obs. Squared}}{\text{Number of Obs}} \left(\frac{T_k^2}{n_k}\right)$	69.14	181.50	325.13	288.80	500.00	= 1,364.57

Using this information, we find that the sum of squares between groups is equal to the following:

$$SS_B = \sum_{k=1}^m \frac{T_k^2}{n_k} - \frac{T^2}{N}$$

$$\approx 1,364.57 - \frac{(194)^2}{31} \approx 150.5$$

Since there are four degrees of freedom for this calculation (the number of groups minus one), the mean squares between groups is as shown below:

$$MS_B = \frac{SS_B}{m-1} \approx \frac{150.5}{4} \approx 37.6$$

Next, we calculate the mean squares within groups,  $MS_W$ , which is also known as the pooled estimate of the population variance,  $\sigma^2$ .

To calculate the mean squares within groups, we first use the following formula to calculate  $SS_W$ :

$$SS_W = \sum_{k=1}^m \sum_{i=1}^{n_k} x_{ik}^2 - \sum_{k=1}^m \frac{T_k^2}{n_k}$$

Using our summary statistics from above, we can calculate  $SS_W$  as shown below:

$$\begin{aligned} SS_W &= \sum_{k=1}^m \sum_{i=1}^{n_k} x_{ik}^2 - \sum_{k=1}^m \frac{T_k^2}{n_k} \\ &\approx 1,452 - 1,364.57 \\ &\approx 87.43 \end{aligned}$$

This means that we have the following for  $MS_W$ :

$$MS_W = \frac{SS_W}{n - m} \approx \frac{87.43}{26} \approx 3.36$$

Therefore, our  $F$ -ratio is as shown below:

$$F = \frac{MS_B}{MS_W} \approx \frac{37.6}{3.36} \approx 11.19$$

We would then analyze this test statistic against our critical value. Using an F-distribution table for  $\alpha=0.01$  (equivalent to a two-tailed significance of 0.02), and also the numerator degrees of freedom of  $m-1=5-1=4$  and the denominator degrees of freedom of  $m-n=31-5=26$ , we find our critical value equal to 4.140. Since our test statistic of 11.19 exceeds the critical value of 4.140, we reject the null hypothesis. We can conclude, therefore, that not all of the population means of the five programs are equal and that obtaining an F-ratio this extreme by chance is highly improbable.

### On the Web

<http://www.statsoft.com/textbook/sttable.html#f01> F-distribution tables.

### Technology Note: Calculating a One-Way ANOVA with Excel

Here is the procedure for performing a one-way ANOVA in Excel using this set of data.

Enter the table above (i.e., with data on how the 31 students divided into five groups performed on reading comprehension) into an empty Excel worksheet.

Click the Data choice on the toolbar, then select 'Data Analysis,' and then choose 'Regression' from the list that appears (Note, if Data Analysis does not appear as a choice on your Data page need to follow the add-in instructions below).

Place the cursor in the 'Input Y range' field and select the third column.

Place the cursor in the 'Input X range' field and select the first and second columns.

Place the cursor in the 'Output Range' field and click somewhere in a blank cell below and to the left of the table.

Click 'Labels' so that the names of the predictor variables will be displayed in the table.

Click 'OK', and the results shown below will be displayed.

*Note: In Excel 2007, to add Data Analysis to your Data page, perform the following functions. Click the Microsoft Office Button in the upper left, then click on Excel Options. Click on Add-ins, then highlight the Analysis ToolPak, click Go, make sure the Analysis ToolPak box is checked off, and then click OK. The Data Analysis choice should now appear on your Excel Data page. Follow the remaining instructions above.*

Anova: Single Factor

**TABLE 11.1:** SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	7	22	3.142857	3.809524
Column 2	6	33	5.5	3.5
Column 3	8	51	6.375	2.839286
Column 4	5	38	7.6	4.3
Column 5	6	50	10	2.5

**TABLE 11.2:** ANOVA

Source of Variation	SS	df	MS	F	P- value	F crit
Between Groups	150.5033	4	37.62584	11.18893	2.05e-05	2.742594
Within Groups	87.43214	26	3.362775			
Total	237.9355	30				

**Technology Note: One-Way ANOVA on the TI-83/84 Calculator**

Enter raw data from population 1 into **L1**, population 2 into **L2**, population 3 into **L3**, population 4 into **L4**, and so on.

Now press [**STAT**], scroll right to **TESTS**, scroll down to 'ANOVA(', and press [**ENTER**]. Then enter the lists to produce a command such as 'ANOVA(L1, L2, L3, L4)' and press [**ENTER**].

**Lesson Summary**

When testing multiple independent samples to determine if they come from the same population, we could conduct a series of separate  $t$ -tests in order to compare all possible pairs of means. However, a more precise and accurate analysis is the Analysis of Variance (ANOVA).

In ANOVA, we analyze the total variation of the scores, including the variation of the scores within the groups, the variation between the group means, and the total mean of all the groups (also known as the *grand mean*).

In this analysis, we calculate the  $F$ -ratio, which is the total mean of squares between groups divided by the total mean of squares within groups.

The total mean of squares within groups is also known as the pooled estimate of the population variance. We find this value by analysis of the standard deviations in each of the samples.

**Review Questions**

1. What does the ANOVA acronym stand for?
2. If we are testing whether pairs of sample means differ by more than we would expect due to chance using multiple  $t$ -tests, the probability of making a type I error would \_\_\_\_.
3. In the ANOVA method, we use the \_\_\_\_ distribution.
  - a. Student's  $t$ -

- b. normal  
c.  $F$ -
4. In the ANOVA method, we complete a series of steps to evaluate our hypothesis. Put the following steps in chronological order.
- Calculate the mean squares between groups and the mean squares within groups.
  - Determine the critical values in the  $F$ -distribution.
  - Evaluate the hypothesis.
  - Calculate the test statistic.
  - State the null hypothesis.
5. A school psychologist is interested in whether or not teachers affect the anxiety scores among students taking the AP Statistics exam. The data below are the scores on a standardized anxiety test for students with three different teachers.

**TABLE 11.3:** Teacher's Name and Anxiety Scores

Ms. Jones	Mr. Smith	Mrs. White
8	23	21
6	11	21
4	17	22
12	16	18
16	6	14
17	14	21
12	15	9
10	19	11
11	10	
13		

- (a) State the null hypothesis.
- (b) Using the data above, fill out the missing values in the table below.

**TABLE 11.4:**

	Ms. Jones	Mr. Smith	Mrs. White	Totals
Number ( $n_k$ )			8	=
Total ( $T_k$ )		131		=
Mean ( $\bar{x}$ )		14.6		=
Sum of Squared Obs. ( $\sum_{i=1}^{n_k} x_{ik}^2$ )				=
Sum of Squared/Number of Obs. ( $\frac{T_k^2}{n_k}$ )				=

- (c) What is the value of the mean squares between groups,  $MS_B$ ?
- (d) What is the value of the mean squares within groups,  $MS_W$ ?
- (e) What is the  $F$ -ratio of these two values?
- (f) With  $\alpha = 0.05$ , use the  $F$ -distribution to set a critical value.
- (g) What decision would you make regarding the null hypothesis? Why?

## 11.3 The Two-Way ANOVA Test

### Learning Objectives

- Understand the differences in situations that allow for one-way or two-way ANOVA methods.
- Know the procedure of two-way ANOVA and its application through technological tools.
- Understand completely randomized and randomized block methods of experimental design and their relation to appropriate ANOVA methods.

### Introduction

In the previous section, we discussed the one-way ANOVA method, which is the procedure for testing the null hypothesis that the population means and variances of a single independent variable are equal. Sometimes, however, we are interested in testing the means and variances of more than one independent variable. Say, for example, that a researcher is interested in determining the effects of different dosages of a dietary supplement on the performance of both males and females on a physical endurance test. The three different dosages of the medicine are low, medium, and high, and the genders are male and female. Analyses of situations with two independent variables, like the one just described, are called two-way ANOVA tests.

**TABLE 11.5:** Mean Scores on a Physical Endurance Test for Varying Dosages and Genders

	Dietary Supplement Dosage	Dietary Supplement Dosage	Dietary Supplement Dosage	Average
	Low	Medium	High	
Female	35.6	49.4	71.8	52.3
Male	55.2	92.2	110.0	85.8
Average	45.4	70.8	90.9	

There are several questions that can be answered by a study like this, such as, "Does the medication improve physical endurance, as measured by the test?" and "Do males and females respond in the same way to the medication?"

While there are similar steps in performing one-way and two-way ANOVA tests, there are also some major differences. In the following sections, we will explore the differences in situations that allow for the one-way or two-way ANOVA methods, the procedure of two-way ANOVA, and the experimental designs associated with this method.

### The Differences in Situations that Allow for One-way or Two-Way ANOVA

As mentioned in the previous lesson, ANOVA allows us to examine the effect of a single independent variable on a dependent variable (i.e., the effectiveness of a reading program on student achievement). With *two-way ANOVA*, we are not only able to study the effect of two independent variables (i.e., the effect of dosages and gender on the results of a physical endurance test), but also the interaction between these variables. An example of interaction between the two variables gender and medication is a finding that men and women respond differently to the medication.

We could conduct two separate one-way ANOVA tests to study the effect of two independent variables, but there are several advantages to conducting a two-way ANOVA test.

*Efficiency.* With simultaneous analysis of two independent variables, the ANOVA test is really carrying out two separate research studies at once.

*Control.* When including an additional independent variable in the study, we are able to control for that variable. For example, say that we included IQ in the earlier example about the effects of a reading program on student achievement. By including this variable, we are able to determine the effects of various reading programs, the effects of IQ, and the possible interaction between the two.

*Interaction.* With a two-way ANOVA test, it is possible to investigate the interaction of two or more independent variables. In most real-life scenarios, variables do interact with one another. Therefore, the study of the interaction between independent variables may be just as important as studying the interaction between the independent and dependent variables.

When we perform two separate one-way ANOVA tests, we run the risk of losing these advantages.

### Two-Way ANOVA Procedures

There are two kinds of variables in all ANOVA procedures—dependent and independent variables. In one-way ANOVA, we were working with one independent variable and one dependent variable. In two-way ANOVA, there are two independent variables and a single dependent variable. Changes in the dependent variables are assumed to be the result of changes in the independent variables.

In one-way ANOVA, we calculated a ratio that measured the variation between the two variables (dependent and independent). In two-way ANOVA, we need to calculate a ratio that measures not only the variation between the dependent and independent variables, but also the interaction between the two independent variables.

Before, when we performed the one-way ANOVA, we calculated the total variation by determining the variation within groups and the variation between groups. Calculating the total variation in two-way ANOVA is similar, but since we have an additional variable, we need to calculate two more types of variation. Determining the total variation in two-way ANOVA includes calculating: variation within the group (within-cell variation), variation in the dependent variable attributed to one independent variable (variation among the row means), variation in the dependent variable attributed to the other independent variable (variation among the column means), and variation between the independent variables (the interaction effect).

The formulas that we use to calculate these types of variations are very similar to the ones that we used in the one-way ANOVA. For each type of variation, we want to calculate the total sum of squared deviations (also known as the sum of squares) around the grand mean. After we find this total sum of squares, we want to divide it by the number of degrees of freedom to arrive at the mean of squares, which allows us to calculate our final ratio. We could do these calculations by hand, but we have technological tools, such as computer programs like Microsoft Excel and graphing calculators, that can compute these figures much more quickly and accurately than we could manually. In order to perform a two-way ANOVA with a TI-83/84 calculator, you must download a calculator program at the following site: <http://www.wku.edu/~david.neal/statistics/> .

The process for determining and evaluating the null hypothesis for the two-way ANOVA is very similar to the same process for the one-way ANOVA. However, for the two-way ANOVA, we have additional hypotheses, due to the additional variables. For two-way ANOVA, we have three null hypotheses:

1. In the population, the means for the rows equal each other. In the example above, we would say that the mean for males equals the mean for females.
2. In the population, the means for the columns equal each other. In the example above, we would say that the means for the three dosages are equal.
3. In the population, the null hypothesis would be that there is no interaction between the two variables. In the example above, we would say that there is no interaction between gender and amount of dosage, or that all

effects equal 0.

Let’s take a look at an example of a data set and see how we can interpret the summary tables produced by technological tools to test our hypotheses.

*Example:* Say that a gym teacher is interested in the effects of the length of an exercise program on the flexibility of male and female students. The teacher randomly selected 48 students (24 males and 24 females) and assigned them to exercise programs of varying lengths (1, 2, or 3 weeks). At the end of the programs, she measured the students’ flexibility and recorded the following results. Each cell represents the score of a student:

**TABLE 11.6:**

Gender		Length of Program	Length of Program	Length of Program	
		1 Week	2 Weeks	3 Weeks	
Females		32	28	36	
		27	31	47	
		22	24	42	
		19	25	35	
		28	26	46	
		23	33	39	
		25	27	43	
		21	25	40	
	Males		18	27	24
			22	31	27
			20	27	33
			25	25	25
			16	25	26
			19	32	30
			24	26	32
			31	24	29

Do gender and the length of an exercise program have an effect on the flexibility of students?

Solution:

From these data, we can calculate the following summary statistics:

**TABLE 11.7:**

Gender			Length of Program	Length of Program	Length of Program	Total
			1 Week	2 Weeks	3 Weeks	
Females	<i>n</i>		8	8	8	24
	Mean		24.6	27.4	41.0	31.0
	St. Dev.		4.24	3.16	4.34	8.23
Males	<i>n</i>		8	8	8	24
	Mean		21.9	27.1	28.3	25.8
	St. Dev.		4.76	2.90	3.28	4.56
Totals	<i>n</i>		16	16	16	48
	Mean		23.3	27.3	34.6	28.4
	St. Dev.		4.58	2.93	7.56	7.10

As we can see from the tables above, it appears that females have more flexibility than males and that the longer

programs are associated with greater flexibility. Also, we can take a look at the standard deviation of each group to get an idea of the variance within groups. This information is helpful, but it is necessary to calculate the test statistic to more fully understand the effects of the independent variables and the interaction between these two variables.

**Technology Note: Calculating a Two-Way ANOVA with Excel**

Here is the procedure for performing a two-way ANOVA with Excel using this set of data.

1. Copy and paste the earlier table (with the flexibility data from the 48 students) into an empty Excel worksheet, without the labels 'Length of program' and 'Gender'.
2. Select 'Data Analysis' from the Tools menu and choose 'ANOVA: Two-Factor Without Replication' from the list that appears.
3. Place the cursor in the 'Input Range' field and select the entire table.
4. Place the cursor in the 'Output Range' field and click somewhere in a blank cell below the table.
5. Click 'Labels' only if you have also included the labels in the table. This will cause the names of the predictor variables to be displayed in the table.
6. Click 'OK', and the results shown below will be displayed.

Using technological tools, we can generate the following summary table:

**TABLE 11.8:**

Source	SS	df	MS	F	Critical Value of $F^*$
Rows (gender)	582.58	15	38.84	1.62	2.015
Columns (length)	1,065.5	2	532.75	22.22	3.32
Error	719.17	30	23.97		
Total	2,367.25	47			

\*Statistically significant at  $\alpha = 0.05$ .

From this summary table, we can see that all three  $F$ -ratios exceed their respective critical values.

This means that we can reject all three null hypotheses and conclude that:

In the population, the mean for males differs from the mean of females.

In the population, the means for the three exercise programs differ.

There is an interaction between the length of the exercise program and the student's gender.

### Experimental Design and its Relation to the ANOVA Methods

*Experimental design* is the process of taking the time and the effort to organize an experiment so that the data are readily available to answer the questions that are of most interest to the researcher. When conducting an experiment using the ANOVA method, there are several ways that we can design an experiment. The design that we choose depends on the nature of the questions that we are exploring.

In a totally randomized design, the subjects or objects are assigned to treatment groups completely at random. For example, a teacher might randomly assign students into one of three reading programs to examine the effects of the different reading programs on student achievement. Often, the person conducting the experiment will use a computer to randomly assign subjects.

In a randomized block design, subjects or objects are first divided into homogeneous categories before being randomly assigned to a treatment group. For example, if an athletic director was studying the effect of various

physical fitness programs on males and females, he would first categorize the randomly selected students into homogeneous categories (males and females) before randomly assigning them to one of the physical fitness programs that he was trying to study.

In ANOVA, we use both randomized design and randomized block design experiments. In one-way ANOVA, we typically use a completely randomized design. By using this design, we can assume that the observed changes are caused by changes in the independent variable. In two-way ANOVA, since we are evaluating the effect of two independent variables, we typically use a randomized block design. Since the subjects are assigned to one group and then another, we are able to evaluate the effects of both variables and the interaction between the two.

---

## Lesson Summary

With two-way ANOVA, we are not only able to study the effect of two independent variables, but also the interaction between these variables. There are several advantages to conducting a two-way ANOVA, including efficiency, control of variables, and the ability to study the interaction between variables. Determining the total variation in two-way ANOVA includes calculating the following:

Variation within the group (within-cell variation)

Variation in the dependent variable attributed to one independent variable (variation among the row means)

Variation in the dependent variable attributed to the other independent variable (variation among the column means)

Variation between the independent variables (the interaction effect)

It is easier and more accurate to use technological tools, such as computer programs like Microsoft Excel, to calculate the figures needed to evaluate our hypotheses tests.

---

## Review Questions

1. In two-way ANOVA, we study not only the effect of two independent variables on the dependent variable, but also the \_\_\_ between the two independent variables.
2. We could conduct multiple  $t$ -tests between pairs of hypotheses, but there are several advantages when we conduct a two-way ANOVA. These include:
  - a. Efficiency
  - b. Control over additional variables
  - c. The study of interaction between variables
  - d. All of the above
3. Calculating the total variation in two-way ANOVA includes calculating \_\_\_ types of variation.
  - a. 1
  - b. 2
  - c. 3
  - d. 4
4. A researcher is interested in determining the effects of different doses of a dietary supplement on the performance of both males and females on a physical endurance test. The three different doses of the medicine are low, medium, and high, and again, the genders are male and female. He assigns 48 people, 24 males and 24 females, to one of the three levels of the supplement dosage and gives a standardized physical endurance test. Using technological tools, he generates the following summary ANOVA table:

TABLE 11.9:

Source	SS	df	MS	F	Critical Value of F
Rows (gender)	14.832	1	14.832	14.94	4.07
Columns (dosage)	17.120	2	8.560	8.62	3.23
Interaction	2.588	2	1.294	1.30	3.23
Within-cell	41.685	42	992		
Total	76,226	47			

$$*\alpha = 0.05$$

- What are the three hypotheses associated with the two-way ANOVA method?
- What are the three null hypotheses for this study?
- What are the critical values for each of the three hypotheses? What do these tell us?
- Would you reject the null hypotheses? Why or why not?
- In your own words, describe what these results tell us about this experiment.

#### **On the Web**

[http://www.ruf.rice.edu/~lane/stat\\_sim/two\\_way/index.html](http://www.ruf.rice.edu/~lane/stat_sim/two_way/index.html) Two-way ANOVA applet that shows how the sums of square total is divided between factors  $A$  and  $B$ , the interaction of  $A$  and  $B$ , and the error.

<http://tinyurl.com/32qaufs> Shows partitioning of sums of squares in a one-way analysis of variance.

<http://tinyurl.com/djob5t> Understanding ANOVA visually. There are no numbers or formulas.

#### **Keywords**

ANOVA method

Experimental design

$F$ -distribution

$F$ -Max test

$F$ -ratio test statistic

Grand mean

Mean squares between groups

Mean squares within groups

Pooled estimate of the population variance

$SS_B$

$SS_W$

Two-way ANOVA