Lesson 12, part 1
Inferences for Categorical-Numerical Association

Outline of the Lesson[1]

As you know, "statistical inference" refers to the process of drawing conclusions about an entire population, by using data obtained from a sample drawn from that population. At this point, we are familiar with several types of statistical inference.

- In Lessons 6-8, we learned methods of statistical inference for a single sample, where the variable of interest was categorical. Specifically, we learned about these methods:
  - one-sample proportion hypothesis tests.
  - one-sample proportion confidence intervals.

- This same idea was addressed for numerical variables in Lesson 11, where we learned about the following methods:
  - one-sample mean hypothesis tests.
  - one-sample mean confidence intervals.

[1] This "part 1" document contains sections 12.1 through 12.6. The "part 2" document contains the remaining Section 12.7. Some, but not all, courses which use these lessons will study part 2 as well as part 1 of the lesson.

- In Lesson 10, we focused on categorical data for several different groups, developing two additional methods:
    - chi-squared hypothesis tests
    - confidence intervals for the difference between two population proportions (two-sample proportion confidence intervals) [2].

This lesson is similar to Lesson 10, but for numerical data. That is, we will be analyzing multiple populations with respect to a numerical variable. As in Lesson 10, we will be asking these questions:

- Is there a difference between the different populations?
- If so, how big is that difference?

For example, is there a difference between different brands of golf balls? Is it true that some golf balls go further than others when a golfer hits their drive? One way to think about this question is to realize that we are analyzing the connection (association, correlation) between two variables: the *brand* of the golf ball, and the *distance* it goes when struck by a driver. These are ideas we learned about back in Lesson 3. We begin with a quick review of the pertinent parts of that lesson. To prepare for this review, you should work the following exercise and check your answers.

Exercise 1[3]: The following table presents data for four brands of golf balls. The researcher randomly selected 10 balls of each brand. Each ball was struck by a mechanical driver, and the distance the ball traveled was measured.

Brand A	251	244	248	251	261	251	253	246	254	249
Brand B	263	263	266	254	264	258	263	264	261	256
Brand C	269	263	278	267	271	267	271	272	276	266
Brand D	252	249	249	242	247	251	262	249	247	246

a. Calculate the mean for each of the four samples. Based on the means, does it appear that, for the golf balls in the sample, the brand of ball made a difference in the distance the ball went when struck by the mechanical driver? Circle all statements you agree with.
    - No, it made no difference.
    - It appears that brands A & D went a shorter distance than the other two brands.
    - Brand C seems better than the other three.

b. Create side by side boxplots for the data. Does this graph seem to support your conclusions?

[2] You may not have learned this method – some courses that use these lessons treat this topic as optional.
[3] Reminder: Many examples in these lessons, including the current example, involve studies of the type which professional statisticians typically carry out, but with data created by the author of the lessons to illustrate the statistical concepts involved.

12.1 – Review: Association Between a Categorical Variable and a Numerical Variable

In Lesson 3, we studied the association between two variables. For example, we might ask these questions.

- Is there an association between whether one smokes and whether or not he or she has cancer?
- Is there an association between high school GPAs (grade point averages) and college GPAs?
- Is there an association between gender and SAT scores?

An **association** (or **relationship**, or **connection**, or **correlation**) exists between two variables if a particular value for one variable is more likely to occur with certain values of the other variable. If no association exists, we say the variables are **independent**.

In this lesson, one variable will be a categorical variable and the other a numerical variable. We will treat the categorical variable as the **explanatory** variable, and we will think of that variable as establishing the groups we wish to study. The numerical variable will be the **response** variable, the outcome from our survey or experiment. For example, to see if there is a difference between brands of golf balls, we can treat the brand as the explanatory variable and the distance the ball goes when hit by a driver as the response variable.

In general, there are a number of ways to ask the primary question we are interested in; all these English-language questions mean the same thing.

- Is there an association (correlation, relationship, connection) between the two variables?
- Are the groups different?
- Does the response variable depend on the explanatory variable?

The answer to the question could be either "yes" or "no." This table summarizes the different ways we can express these two possible answers, applied specifically to our example involving golf ball brand and distance.

yes	**no**
There *is* an association (connection, relationship, correlation) between the brand of golf ball and the distance it goes when struck by a driver.	There is *no* association (connection, relationship, correlation) between the brand of golf ball and the distance it goes when struck by a driver.
There *is* a difference between the different brands of golf ball, with respect to the distance it goes when struck by a driver.	There is *no* difference between the different brands of golf ball, with respect to the distance it goes when struck by a driver.
The distance the ball goes when struck by a driver *depends on* the brand of ball.	The distance the ball goes when struck by a driver *is independent of* the brand of ball.

What the table presents is not mathematics; it is English-language usage. An important note is this:

- To say that there *is an association* between the explanatory variable and the response variable says that the groups identified by the explanatory variable are *different* from one another, for the response variable you are studying. The value for the response variable can be expected to be different, depending on the value for the explanatory variable.

Exercise 2: In Exercise 1, we concluded that, for the balls in the sample, it appears that brands A and D go shorter distances than the others, and brand C goes the furthest. Use this conclusion to answer the following questions, for the balls in the sample.

a. Is there an association between the two variables (brand, distance)?
b. Are the two variables independent?
c. Does the distance depend on the brand?
d. Is there a difference between the brands, relative to the distance the ball goes when struck by a driver?

12.2 – The Logic of Hypothesis Testing for Categorical-Numerical Association

In this section we describe the logic of hypothesis testing for multiple groups, for the situation where we have one categorical variable that establishes the groups, and one numerical variable. The overall logic can be expressed in a short paragraph identical to that we used to describe hypothesis testing for the χ^2 test in Lesson 10.

> The null hypothesis states that there is no difference between the groups identified by the explanatory variable. The alternative hypothesis simply states that the null hypothesis is incorrect. To judge whether one should or should not reject the null hypothesis, the researcher takes a sample from the population. If the sample has data in which the groups are "close to" identical, this would seem to support the null hypothesis. On the other hand, if the sample's data is "far away from" what the null hypothesis claims, this is evidence against the null hypothesis. If the evidence is strong enough, the researcher rejects the null hypothesis.

Just as we have done for all the hypothesis test methods studied so far, we will judge whether the sample data is "far away from" what the null hypothesis claims by calculating a test statistic and a corresponding *p*-value. If the *p*-value is less than our chosen significance level α, we will reject the null hypothesis. The details of the corresponding calculations are outlined in the next two sections. For now, we will continue with our example involving golf balls to help us understand the overall structure of the hypothesis test involved.

Null and alternative hypotheses

Once again, we are using as an example the situation described in Exercises 1 and 2: you are studying four different brands of golf balls to see how far a mechanical driver hits them. You can think of this two ways:

1. There are four groups for which you are studying the numerical variable *distance*.
2. There are two variables we are studying, *brand* and *distance.*

The first way of thinking (there are four groups for which we are studying the *distance* variable) leads to one way of expressing the null hypothesis. We let μ_1 be the mean distance for all balls of the first brand, μ_2 the mean distance for all balls of the second brand, and similarly for μ_3 and μ_4. Then the null hypothesis says all these means are the same. The alternative hypothesis says they are not all the same (at least two are different from each other).

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_a: at least two of the population means are unequal

Caution: The alternative hypothesis *does not* state that all four means are different from each other. Rather, it states that the four means are not all the same. This is a subtle but important distinction. For the null hypothesis to be false, it is not necessary that all four numbers are different. For example, if it turned out that these were the four population means:

220	200	200	200

this would show the null hypothesis to be false, as would each of the following scenarios:

220	220	240	240
200	210	220	220

The alternative hypothesis simply states that there is some difference among the four numbers, and perhaps the simplest way to state this is that at least two of the means are different from each other.

The second way of thinking (there are two variables, *brand* and *distance*) leads to another possibility, which is exactly analogous to what we did for the chi-square test in Lesson 10. In this case, we can think in terms of association (relationship/connection/correlation) or in terms of independence and dependence. This leads to the following:

H_0 : brand of golf ball and driving distance are independent
H_a : driving distance depends on the brand of ball

or

H_0 : there is *no* association between brand of golf ball and driving distance
H_a : there *is* an association between brand of golf ball and driving distance

Example – looking at the data

As we have seen, one way to write the null hypothesis is to state that *in the entire populations*, the means for all four brands are equal:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

So, to decide whether or not to reject this hypothesis, it makes sense to examine the means *in the samples*. We did this in Exercise 1(a). *In the samples*, the means for the four brands were obviously different: 250.8 for Brand A, 261.2 for Brand B, 270.0 for Brand C, and 249.4 for Brand D. However, as always, we must acknowledge that there is a certain randomness introduced by the process of taking samples. We must try to decide between two possibilities:

- Reject the null hypothesis, concluding that the differences we see in the samples are due to differences in the entire populations.
- Fail to reject the null hypothesis, acknowledging that the means in the entire populations could be equal. In this case, we conclude that these four samples could be this different simply by chance, because of the randomness inherent in sampling.

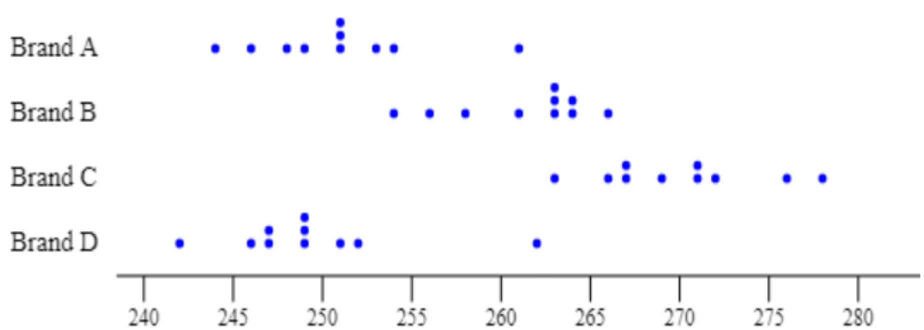
In Exercise 1(a) we tried to answer this question by looking at the four sample means, and making a judgment about how spread out they are. However, in general it is not enough simply to look at the variability of the sample means. To help us think about this, we will examine two different sets of data. The first set of data is the data presented in Exercise 1(a):

Brand A	251	244	248	251	261	251	253	246	254	249
Brand B	263	263	266	254	264	258	263	264	261	256
Brand C	269	263	278	267	271	267	271	272	276	266
Brand D	252	249	249	242	247	251	262	249	247	246

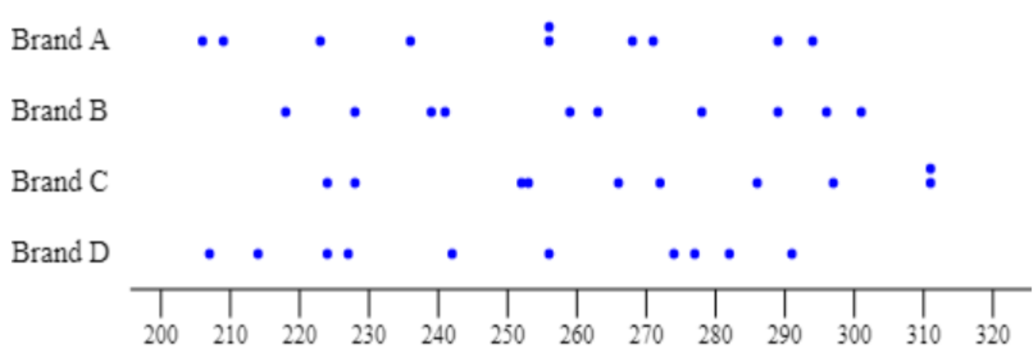
Here is another set of data, with this important property: ***The means for the four samples are identical to those for the first set of data.***

Brand A	206	209	223	236	256	256	268	271	289	294
Brand B	218	228	241	239	259	263	278	289	296	301
Brand C	224	228	253	252	266	272	286	297	311	311
Brand D	207	214	224	227	242	256	277	274	282	291

However, although the means are the same, the two sets are not at all similar. One easy way to see this is to create side-by-side dot plots for each set of data. For the first set this is the graph:



For the second set, here is the graph:



For the data depicted in the first graph, the graph seems to suggest that there is indeed a difference among the brands – the Brand B and Brand C balls seem to travel consistently further than the Brand A and Brand D balls, and the Brand C balls seem to travel consistently the furthest. Intuitively, one would suspect that for the four populations consisting of **all** balls of the four different brands, there is a difference in the average distance. That is, intuitively one would suspect that we should reject the null hypothesis.

On the other hand, for the data depicted in the second graph, the graph does not really suggest a strong difference. Intuitively, one would acknowledge that for the four populations consisting of **all** balls of the four different brands, it is conceivable that there is no difference in the average distance. That is, intuitively one would suspect that we should **not** reject the null hypothesis.

Comment. The graphs give a pictorial representation of the differences between the two sets of data. We can also see an important difference by examining the mean and standard deviation for each set. First, the first:

	Brand A	Brand B	Brand C	Brand D
Mean	250.8	261.2	270	249.4
Std. dev.	4.7093	3.9101	4.5947	5.2324

and for the second:

	Brand A	Brand B	Brand C	Brand D
Mean	250.8	261.2	270	249.4
Std. dev.	31.322	29.3023	31.3759	30.6529

The standard deviations in the second are much larger than in the first. Graphically, we saw this as the fact that each individual dot plot was more "spread out."

The hypothesis test we are learning about is called an ANOVA test, standing for **ANalysis Of VAriance**. The details of the calculations are outlined in Section 12.4. For now it is enough to say that it bases its calculations on the notion of *variance*. It examines the *variance* (think "variability") among the four sample means. However, as we observed, the four means for the data presented in the second graph are identical to the four means for the data in the first graph. It is obviously not enough, then, to examine how spread out the means are. One must somehow take into account the difference between these two graphs. One would certainly hope that the ANOVA test would be more likely to reject the null hypothesis for the data presented in the first dot plots, and less likely to do so for the data presented in the second dot plots.

The ANOVA test accomplishes this by also examining the variability within each of the samples. The sample data for each brand of ball depicted in the second graph is much more variable – much more spread out – than that in the first graph. Numerically, we can see this in the differences for the standard deviations from the comment above. By including calculations involving the variability within each sample, the ANOVA test is able to distinguish between the two data sets pictured in these two graphs. We will have more to say about this in the following sections.

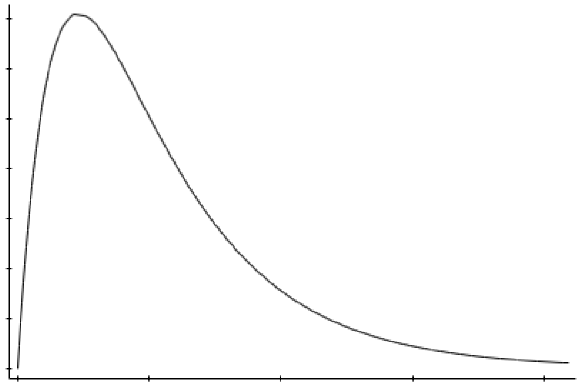
Should we reject the null hypothesis?

The question, "Should we reject the null hypothesis?" boils down to this related question: "How far is the sample data from what it should have been based on the null hypothesis?" To answer this question, we do two calculations.

1. First, we calculate a test statistic which measures the difference between the sample and the null hypothesis claim. This score is analogous to the χ^2-score of Lesson 10, but the calculations are of course different. As we indicated in the previous discussion, this test

statistic involves calculations involving *variance* – both variance among the means of the different samples, and variance within each sample.

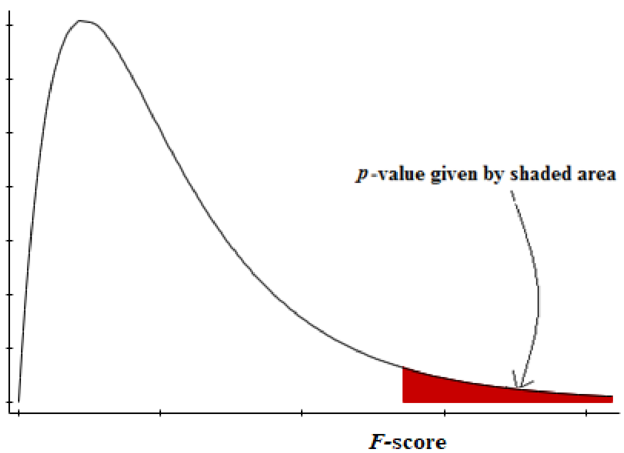
Statisticians refer to this test statistic as an *F*-score, in honor of the person, R. A. Fisher, who first developed the method. For the situation where the null hypothesis is true, the possible *F*-scores form a sampling distribution quite similar in shape to the χ^2 (chi-square) distribution we learned about in Lesson 10. See the figure.



The results are always positive. Referring to the two data sets whose dot plots we considered earlier, the first set of data would give a score out in the right tail of the distribution; the second set's *F*-score would be closer to zero. As always, our strategy is designed to control the occurrence of Type 1 error – that is, the likelihood of rejecting a true null hypothesis. Accordingly, if we get a large *F*-score (out in the tail of the F distribution), we will reject the null hypothesis.

2. To measure how far a particular F-score lies out in the tail of the F distribution, we calculate the corresponding *p*-value. See the figure below. This is quite similar to the χ^2 calculation of the *p*-value. As always, a small *p*-value indicates that we should reject the null hypothesis. If the *p*-value is less than our chosen significance level α, we reject the null hypothesis.

   (Note that the calculators provided with these lessons will calculate both the F-score and the *p*-value for us, based on the information about our sample.)



Stating conclusions

Just as in the hypothesis test methods we have already learned about, we ask the question: "Was there enough evidence to support the alternative hypothesis?"

- If the *p*-value is small, we reject the null hypothesis, and we write "There *was* evidence to support the alternative hypothesis."
- Otherwise, we fail to reject the null hypothesis, and we write, "There *was not* enough evidence to support the alternative hypothesis."

For the example with the golf balls, here are several ways to write the conclusion if it turns out that we reject the null hypothesis.

- "There was evidence to indicate that driving distance depends on the brand of ball."
- "There was evidence to indicate an association between brand of ball and driving distance."
- "There was evidence to indicate that the mean driving distances are not the same for all four brands. At least two of the brands differ from each other."

If we do not reject the null hypothesis, we will simply change "there *was* evidence" to "there was *not* enough evidence":

- "There was not enough evidence to indicate that driving distance depends on the brand of ball."
- "There was not enough evidence to indicate an association between brand of ball and driving distance."
- "There was not enough evidence to indicate that the mean driving distances are not the same for all four brands; the four population means *could be* equal."

As we learned in Lesson 10, another common way to express the conclusion uses the word **significant** or, more precisely, **statistically significant**. Use of this word indicates that the researchers rejected the null hypothesis. So, for a small *p*-value we might write:

> Researchers found a significant difference in the driving distance for the four different brands of golf ball.

On the other hand, if the *p*-value is not small, the words "found a significant difference" become "found *no* significant difference."

Comment: As always, these conclusions are conclusions about the *entire population*. Professional statisticians frequently just assume that fact, and write sentences such as, "There was evidence to indicate that driving distance depends on the brand of ball," without explicitly emphasizing that the statement is about populations, not merely about the balls that were sampled. If you want to be more precise, you might enhance the statement to something like this: "There was evidence to indicate that, for the populations consisting of *all* the golf balls of the four different brands, driving distance depends on the brand of ball."

Example[4]. How much do computer programmers earn? In answering this question, does it matter what region of the country the programmers are employed in? To study this, a researcher randomly sampled 146 computer programmers in states in the Southeast, Midwest, and Mid-Atlantic regions of the U.S. The data is summarized in the following table:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$73,267	$7,588
Southeast	63	$72,714	$6,602
Midwest	53	$68,623	$4,369

Describe a suitable hypothesis test to answer the question posed. Report the results using α = 0.05; using α = 0.01.

Solution. First of all, an ANOVA test is appropriate. There are two ways to think of this.

- We are asking about the association between a categorical variable (region of the country) and a numerical variable (salary), and ANOVA is designed for exactly this type of association. The explanatory variable is region of the country, with salary as the response variable.
- Alternatively, we can realize that ANOVA is indicated because we are examining the mean salary for three different populations – the mean salary for the Mid-Atlantic, the mean salary for the Southeast, and the mean salary for the Midwest.

Based on the data provided, we see that, ***in the sample***, the means are different. The Midwest mean is over $4000 less than either of the other two – with the other two seeming fairly similar. But it requires an ANOVA test to see if the difference observed in the sample is large enough to suggest a difference in the entire population.

Hypotheses. As usual, the null hypothesis states that the groups identified by the explanatory variable are all the same. Using the word *association*, we write:

H_0: there is *no* association between region of the country and salary for computer programmers

The alternative hypothesis states that the null hypothesis is false:

H_a: there *is* an association between region of the country and salary for computer programmers

If we wish, we can also write the null hypothesis in terms of the population means for the three regions. Using μ_{MA} for the population mean salary for computer programmers in the Mid-Atlantic, μ_{SE} for the Southeast population mean, and μ_{MW} for the Midwest population mean, we write:

$$H_0: \mu_{MA} = \mu_{SE} = \mu_{MW}$$

The alternative hypothesis states that the population means are **not** all equal – at least two are different from each other.

[4] Reminder: Many examples in these lessons, including the current example, involve studies of the type which professional statisticians typically carry out, but with data created by the author of the lessons to illustrate the statistical concepts involved. That is true for this example, although the data shown was adapted from real data.

Calculations. Using one of the techniques described in the following sections, we calculate a test statistic and a *p*-value. Here are the results we obtain:

$F = 8.2852$
p-value = 0.0004

Conclusions. The *p*-value 0.0004 is less than 0.05 and also less than 0.01, so either significance level leads to the same conclusion. We reject the null hypothesis. We have enough evidence to support the alternative hypothesis. In the context of this problem, here are some ways to phrase the conclusion:

- *We found evidence, at $\alpha = 0.01$, to suggest that there is an association between region and salary for computer programmers.*
- *We found strong evidence ($p = 0.0004$) that salary for computer programmers depends on the region of the country.*
- *Researchers reported a significant connection between region and salary for computer programmers.*
- *At significance level $\alpha = 0.01$, we conclude that the mean salary for computer programmers is not the same in all three regions. It appears that the mean is lower in the Midwest than in the other two regions.*

Exercise 3: The data in the previous example was provided in terms of salary in dollars. However, it is well-known that cost of living also varies by region. It is possible to adjust each of the reported salary figures to factor in cost of living. The researchers did this, with the results shown here:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$71,733	$4,362
Southeast	63	$74,881	$7,389
Midwest	53	$72,283	$5,712

a. Comment on the mean salaries in the sample.
b. Using the methods you will learn in later sections, we obtain a test statistic of $F = 3.6286$, with a corresponding *p*-value of 0.0290. At significance level $\alpha = 0.01$, does the evidence establish a connection between region and salary when adjusted for cost of living?
c. At $\alpha = 0.05$, did the researchers find a significant difference in the adjusted salaries for the three different regions?

Exercise 4: Researchers are studying the connection between systolic blood pressure and blood type. In a randomly chosen sample, they obtain data summarized as follows:

Blood type:	**O**	**A**	**B**	**AB**
Sample size:	8	6	4	6
Mean:	140.3750	160.6667	129.5000	113.5000
Standard deviation:	27.0235	17.0372	26.8887	13.6931

a. Which of the following are correct ways to write the null hypothesis?
   - There is a connection between systolic blood pressure and blood type.
   - The population mean systolic blood pressure is the same for all four types of blood.
   - Blood type and systolic blood pressure are independent.
   - Systolic blood pressure depends on blood type.
   - There is no correlation between systolic blood pressure and blood type.

b. Which of the following are correct ways to write the alternative hypothesis?
   - There is a connection between systolic blood pressure and blood type.
   - The population mean systolic blood pressure is different for all four types of blood.
   - At least two of the blood types have different population mean systolic blood pressure.
   - Blood type and systolic blood pressure are independent.
   - Systolic blood pressure depends on blood type.
   - There is no correlation between systolic blood pressure and blood type.

c. Using the methods you will learn in later sections, we obtain a test statistic of F = 4.8193, with a corresponding p-value of 0.0110. At significance level $\alpha = 0.01$, does the evidence establish a connection between systolic blood pressure and blood type?

d. Write the conclusion, at $\alpha = 0.05$, using the word *significant*.

The applet at the following link provides additional practice in formulating hypotheses and drawing conclusions for categorical – numerical association.

ANOVA - hypotheses and conclusions

When do we use ANOVA? A word of caution.

At this point in your study, you have learned about a variety of methods for confidence intervals and hypothesis tests. When confronted with a problem description, your first job is to determine which of these methods apply to your situation.

There are two ways to recognize that an ANOVA test is indicated. First of all, you may realize that the problem involves studying the association between a categorical variable and a numerical variable. Alternatively, the problem may present itself as studying a single numerical variable in a number of different populations.

Example: The golf ball example can be viewed in either of these two ways:

1. Study the association between brand of golf ball (categorical) and driving distance (numerical).
2. Compare the mean driving distance (numerical) for the four populations: all Brand A golf balls, all Brand B golf balls, all Brand C golf balls, and all Brand D golf balls.

Both ways of thinking about the problem suggest using an ANOVA test.

Comment: Perhaps it goes without saying that we cannot use the ANOVA test to study *different* numerical variables in the various populations. For example, it makes no sense to use the test to compare the average age for females to the average height for males.

A perhaps more common mistake among beginning students is to try to use the test to establish an association between two different numerical variables for a single population – for example, female heights and females ages. This too is not a correct use of the test. **The ANOVA test is NOT used to study different numerical variables for a single population – it is used to study a single numerical variable for multiple populations.** (Refer to Lesson 13 for methods to be used to establish association between multiple numerical variables for a single population.)

12.3 – Calculations for ANOVA Tests: Using Technology

In this section we will show you how to use technology – specifically, the online calculator[5] provided by the author of these lessons – to do the calculations involved in an ANOVA hypothesis test. Here again is a link to that calculator:

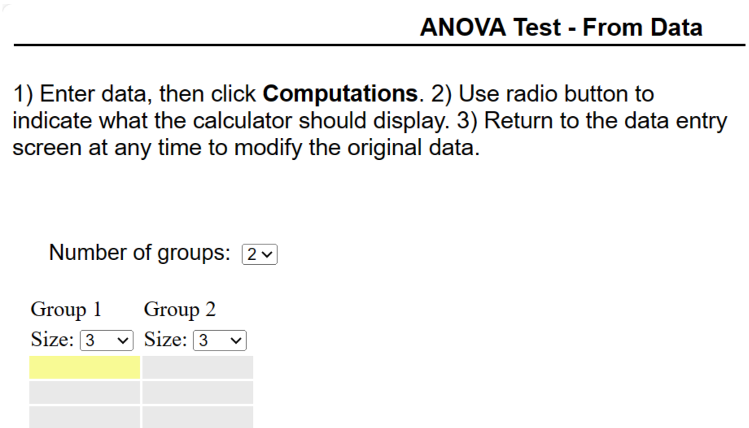
Statistical calculator

We will use the data from Exercise 1, reproduced below, to illustrate the process.

Brand A	251	244	248	251	261	251	253	246	254	249
Brand B	263	263	266	254	264	258	263	264	261	256
Brand C	269	263	278	267	271	267	271	272	276	266
Brand D	252	249	249	242	247	251	262	249	247	246

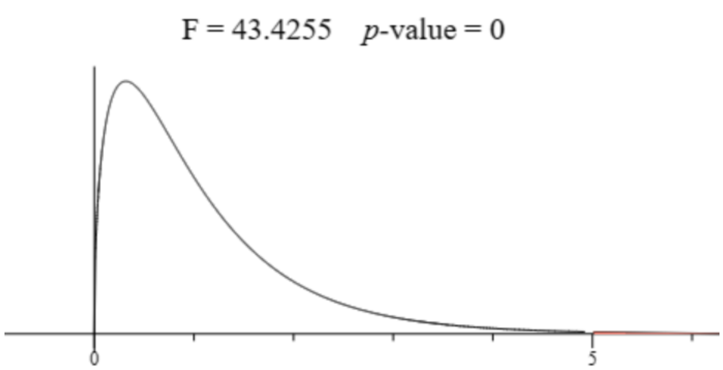
[5] In this section we describe the calculator designed for solving problems such as those typically encountered in an introductory statistics class. In Section 10.6 we will use the datafile-based calculator to analyze categorical variables in a data file.

Choose menu option *Tests* (this is a hypothesis test), then submenu option *ANOVA (from data)*, to obtain this screen:



The data entry process is identical to what we have learned for calculating statistics for several groups of data or for creating side by side box plots. In fact, if you did exercise 1 you may have saved the data to a file, in which case you can simply load that file rather then having to enter the data again.

When you press the *Computations* button you obtain the results:



The F test statistic is 43.4255, and the *p*-value is very small – so small that when the calculator rounded to four places the result was 0. We will reject the null hypothesis.

Notes:

1. These are the results obtained using the original set of data from Exercise 1. Recall that, based on the dot plot of this data, we intuitively felt that we should reject the null hypothesis. The extremely small *p*-value indicates that our intuition was correct in this case.

2. Recall also that we considered a completely different set of data which yielded exactly the same four sample means. But for this second set of data, the dot plot suggested that we might not reject the null hypothesis. It turns out that this set of data yields an *F*-score of 0.9917, with *p*-value 0.4077, we do not reject the null hypothesis Again, our intuition was correct for this second set of data.

3. The calculator provides an alternative method for the situation where you know summary information (size, mean, and standard deviation) about the samples but do not have the actual

data. Of course, it can also be used if you have the data and have already calculated the necessary statistics. For example, for the golf ball data we had shown the following earlier in this lesson:

	Brand A	Brand B	Brand C	Brand D
Mean	250.8	261.2	270	249.4
Std. dev.	4.7093	3.9101	4.5947	5.2324

We also know each sample consisted of 10 golf balls. Using menu option *Tests*, submenu choice *ANOVA (from sample statistics)*, we enter this information as shown:

Number of groups: 4

	Size	Mean	Standard deviation
Group 1	10	250.8	4.7093
Group 2	10	261.2	3.9101
Group 3	10	270	4.5947
Group 4	10	249.4	5.2324

When we press computations, we obtain F = 43.4255, p-value = 0, which of course agrees with the results we obtained when we entered the actual lists of data.

Exercise 5: Refer to Note 2 above.

a. Use technology to carry out an ANOVA test for the second set of golf ball data referred to in that note.
b. Do you reject the null hypothesis at $\alpha = 0.05$? (YES / NO)
c. Does the driving distance depend on the brand at $\alpha = 0.01$? (YES / NO)
d. Is there a significant difference between the four brands of golf balls? (YES/NO)

Exercise 6: Use technology to carry out an ANOVA test for the following data.

Group 1	Group 2	Group 3
4	7	8
12	8	9
13	9	7
7	4	3
13	7	6
9	10	2
9		7
11		

Do you reject the null hypothesis at $\alpha = 0.05$? (YES / NO)
Do you reject the null hypothesis at $\alpha = 0.01$? (YES / NO)
Is there a significant difference between the three groups? (YES/NO)

Exercise 7: Refer to Note 3 above.

a. We have this information for the second set of golf ball data, where again each sample has 10 balls.

	Brand A	Brand B	Brand C	Brand D
Mean	250.8	261.2	270	249.4
Std. dev.	31.322	29.3023	31.3759	30.6529

Use this to carry out an ANOVA test.

b. Carry out an ANOVA test for the summary information presented in Exercise 3, reproduced here:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$71,733	$4,362
Southeast	63	$74,881	$7,389
Midwest	53	$72,283	$5,712

The applet at the following link provides additional practice in formulating hypotheses, doing calculations, and drawing conclusions for ANOVA tests.

ANOVA - calculations

In addition, you may use this applet for more practice in drawing conclusions about the association between a categorical variable and a numerical variable.

ANOVA - interpretation

12.4 – Calculations for ANOVA Tests: "by hand" (partial description)

The ANOVA table presented by technology

When you use the calculator to run an ANOVA test, the output includes the following set of radio buttons:

Display:

- ◉ Results and graph
- ○ ANOVA table
- ○ Group statistics

By default the calculator displays the *F* statistic, the *p*-value, and a graph illustrating the *p*-value. In this section we discuss the second option, the ANOVA table. In general, any technology that does ANOVA calculations will also include this table, either as part of the output or (as in this calculator) as an optional output. For the golf ball example, if we choose this option, we obtain the following table:

	df	SS	MS	F	*p*-value
Factor	3	2799.5	933.1667	43.4255	0
Error	36	773.6	21.4889		
Total	39	3573.1			

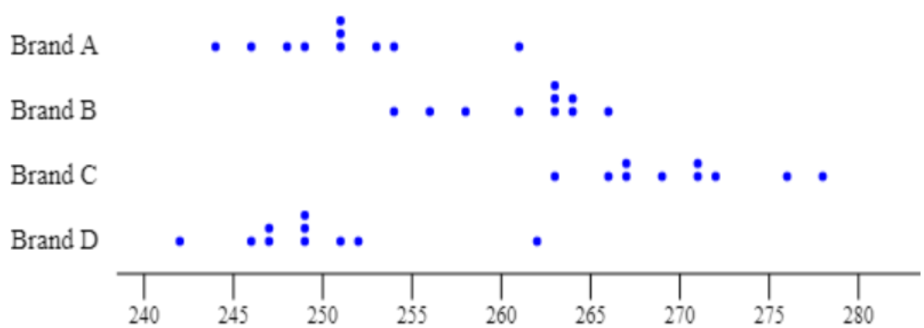
This table presents the following information:

- The *F* test statistic and the *p*-value: 43.4255 and 0.
- Three pieces of information (df, SS, and MS) about what the calculator refers to as “Factor”: 3, 2799.5, and 933.1667.
- The same three pieces of information (df, SS, and MS) about what both the calculator labels as “Error”: 36, 773.6, and 21.4889.
- A “Total” for the df and SS figures.

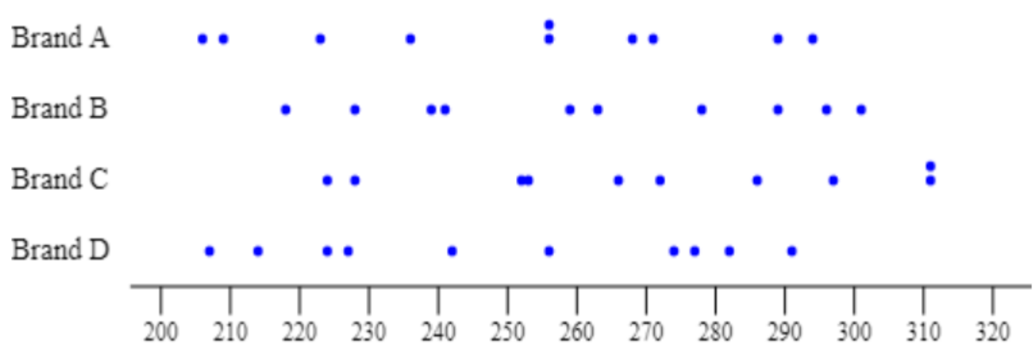
The df, SS, and MS pieces of information are intermediate steps in the calculation of the *F*-statistic. The following subsection gives some details – but not the entire story – about these intermediate steps.

The F test statistic

We begin by reproducing a part of the discussion from Section 12.2. In that section, we examined two different sets of data about four brands of golf balls. The first set of data produces the following side-by-side dot plots:



For the second set, here is the graph:



For the data depicted in the first graph, the graph seems to suggest that there is indeed a difference among the brands – the Brand B and Brand C balls seem to travel consistently further than the Brand A and Brand D balls. Intuitively, one would suspect that for the four populations consisting of **all** balls of the four different brands, there is a difference in the average distance. That is, intuitively one would suspect that we should reject the null hypothesis.

On the other hand, for the data depicted in the second graph, the graph does not really suggest a strong difference. Intuitively, one would acknowledge that for the four populations consisting of **all** balls of the four different brands, it is conceivable that there is no difference in the average distance. That is, intuitively one would suspect that we should **not** reject the null hypothesis.

The hypothesis test we are learning about is called an ANOVA test, standing for **ANalysis Of VAriance**. It bases its calculations on the notion of *variance*. It examines the *variance* (think "variability") among the four sample means. However, as we observed, the four means for the data presented in the second graph are identical to the four means for the data in the first graph. It is obviously not enough, then, to examine how spread out the means are. One must somehow take into account the difference between these two graphs. One would certainly hope that the ANOVA test would be more likely to reject the null hypothesis for the data presented in the first dot plots, and less likely to do so for the data presented in the second dot plots.

The ANOVA test accomplishes this by also examining the variability within each of the samples. The sample data for each brand of ball depicted in the second graph is much more variable – much more spread out – than that in the first graph. By including calculations involving the variability within each sample, the ANOVA test is able to distinguish between the two data sets pictured in these two graphs.

Summarizing, we want a larger *F*-score, and thus a smaller *p*-value, when either of the following situations occur:

- The variability **between** pairs of sample means is larger. The more spread out the sample means are, the more we might suspect that the population means are different.
- The variability **within** each sample is smaller. As we noted, the first graph, where each individual sample exhibits less variation, seems to suggest much more strongly that the populations means are different.

This is the reasoning which led statisticians to the definition of the *F* test statistic. Speaking generally, it is calculated as

$$F = \frac{between-groups-variability}{within-groups-variability}$$

Both the numerator and the denominator are calculated using formulas which are essentially calculations of *variance* (that is, calculations of standard deviation but without the final "square root" step). In Lesson 2 we saw the following formula for calculating the standard deviation:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

You perhaps never actually calculated the standard deviation using this formula, relying instead on technology for the calculations. For our purposes, the important thing to observe is that, prior to the square root step, we can summarize the calculations as follows:

1. Calculate the sum of a collection of terms, each of which is the square of an expression. In brief, calculate a "sum of squares."
2. Calculate the mean of the sum of squares, by dividing by the "degrees of freedom" ($n - 1$).

If we use SS to denote "sum of squares," MS to denote "mean of the sum of squares," and df to denote "degrees of freedom," we can summarize a variance calculation as:

$$MS = \frac{SS}{df}$$

Now back to our calculation of the *F* test statistic. The formula is:

$$F = \frac{between\!-\!groups\!-\!variability}{within\!-\!groups\!-\!variability}$$

We use the words "between-groups" and "within-groups" to remind ourselves that the numerator is concerned about the variability among the means of the groups, while the denominator is concerned with the variability within each group. Statisticians have come up with shorter descriptions of these two concepts. For the numerator, the calculator uses "Factor" but other technology (and textbooks) might use different terms such as "Columns" or "Group." We will use the term "Factor." For the denominator, standard usage is to use the term "Error."

Both the numerator and the denominator are essentially variance calculations, that is a "sum of squares" divided by a "degrees of freedom."

$$between - groups - variability = \frac{sum\ of\ squares\ for\ between\!-\!groups}{degrees\ of\ freedom\ for\ between\!-\!groups}$$

$$within - groups - variability = \frac{sum\ of\ squares\ for\ within\!-\!groups}{degrees\ of\ freedom\ for\ within\!-\!groups}$$

Using the notations we have just discussed, we can rewrite these formulas more succinctly, if less meaningfully, as:

$$Factor\ MS = \frac{Factor\ SS}{Factor\ df} \qquad \text{and} \qquad Error\ MS = \frac{Error\ SS}{Error\ df}$$

and finally:

$$F = \frac{Factor\ MS}{Error\ MS}$$

Example: Demonstrate these calculations for the golf ball example in Exercise 1(a).

Solution: As we pointed out earlier in this section, the calculator provides the following intermediate calculation results:

- Three pieces of information (df, SS, and MS) about what the calculator refers to as "Factor": 3, 2799.5, and 933.16667.
- The same three pieces of information (df, SS, and MS) about what both the calculator labels as "Error": 36, 773.6, and 21.488889.

We will not describe in detail the calculations involved in determining the SS figures. However, we do note that the Factor df (degrees of freedom) is one less than the number of groups, and the Error df is calculated as the total size of the sample minus the number of groups. In the golf ball example, there were four brands of balls, with 10 balls for each brand, for a total of 40 balls. So:

$$Factor\ df = \#groups - 1 = 4 - 1 = 3$$
$$Error\ df = sample\ size - \#groups = 40 - 4 = 36$$

In addition to these numbers, technology gives the Factor SS as 2799.5 and the Error SS as 773.6. Using these numbers, we can calculate:

$$Factor\ MS = \frac{Factor\ SS}{Factor\ df} = \frac{2799.5}{3} = 933.1667$$
$$Error\ MS = \frac{Error\ SS}{Error\ df} = \frac{773.6}{36} = 21.4889$$
$$F = \frac{Factor\ MS}{Error\ MS} = \frac{933.16667}{21.48889} = 43.4255$$

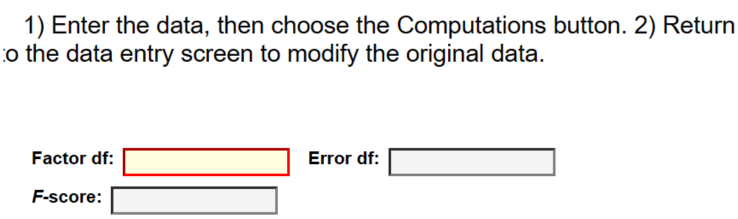
Exercise 8: Circle the correct answers.

a. F is large when (between-groups-variability / within-groups-variability) is much greater than (between-groups-variability / within-groups-variability). In this case, the p-value is (small / large) so we (reject / do not reject) the null hypothesis.

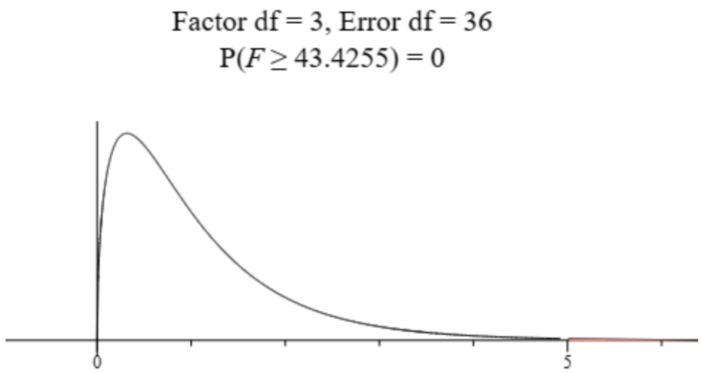
b. F is small when (between-groups-variability / within-groups-variability) is much smaller than (between-groups-variability / within-groups-variability). In this case, the p-value is (small / large) so we (reject / do not reject) the null hypothesis.

The *p*-value

Once you have the *F*-score, you can use tables to get a range of associated *p*-values. (This is the approach taken by many textbooks.) Preferably, you can use technology to do the calculation. Simply choose the menu option *Distributions*, with submenu option *F distribution p-value* to obtain this screen:



Continuing with our example, we have calculated *F* as 43.4255, *Factor df* as 3, and *Error df* as 36. We fill in those value and press *Computations* to obtain this result:



So F is 43.4255, and the p-value is extremely small – so small that when rounded to four places its value is 0.

Exercise 9: In an earlier example, we considered data on computer programmers' salary in three different regions of the country, with this summary information:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$73,267	$7,588
Southeast	63	$72,714	$6,602
Midwest	53	$68,623	$4,369

Answer the following for carrying out an ANOVA test using this data:

a. Factor df = _________
b. Error df = __________
c. Given that Factor SS = 621.7001 and Error SS = 5365.1766, calculate the F test statistic. Round to 4 decimal places in each step.
d. Calculate the p-value.

Exercise 10: Is test performance affected by the color of paper the test is printed on? To examine this, a professor gives the same test to everyone in her class, but with the color paper randomly determined. Five students had pink paper, four had yellow, eight had blue, and five had green. Here is a partial ANOVA table showing the results generated by technology. Complete the missing entries in the table.

	df	**SS**	**MS**	**F**	**P**
Factor		1509.5886			
Error		2054.2750			

12.5 Assumptions and Robustness

Just as we learned in Lesson 11 for the *t* test for a population mean, the ANOVA test is mathematically based upon certain assumptions.

1. We must be studying the same numerical variable in multiple populations.
2. The results must be based on random samples drawn from the populations being studied.
3. The variable must have a normal distribution for each of the populations.
4. All the populations must have the same standard deviation.

The first of these conditions can be thought of in another way: we are studying two variables – one categorical, one numerical – for a single population. The connection between these two ways of thinking can be established by thinking of the values of the categorical variable as dividing the single population into multiple groups.

Caution: The ANOVA test is used to study a single numerical variable for multiple populations. It is NOT used to study different numerical variables for a single population. (Lesson 13 covers linear regression, which can be used for this latter situation.)

The second condition is one we have discussed earlier, beginning in Lesson 4. Throughout these notes, we are assuming that the researcher has obtained a simple random sample from the population or populations being studied. In practice, it may be very difficult to achieve a simple random sample when carrying out a study. When you read about statistical studies, keep in mind that the results are trustworthy only to the extent that randomness has been achieved in carrying out the study, and to the extent that the sample has been drawn from the same populations the researcher is making claims about.

On the surface, the third and fourth conditions appear to pose a significant problem. In practice, it is difficult (impossible?) to guarantee that the populations being studied are normal, and even more difficult to guarantee that they all have the same standard distribution. Fortunately, just like the *t* test, the ANOVA test is ***robust***. Statisticians have established that the ANOVA test gives good results, especially for larger samples, provided at least one of these conditions are met:

- The samples from the various populations are all approximately the same size.
- The largest sample standard deviation in any of the samples is less than twice as big as the smallest sample standard deviation.

By "good results" we mean that the advertised significance level is reasonably accurate. If we do an ANOVA test using $\alpha = 0.05$, for example, the method does indeed restrict the likelihood of a Type 1 error to about 5%.

Example. In Section 12.2, we discussed an ANOVA test for this situation. A researcher randomly sampled 146 computer programmers in states in the Southeast, Midwest, and Mid-Atlantic regions of the U.S. The data is summarized in the following table:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	\$73,267	\$7,588
Southeast	63	\$72,714	\$6,602
Midwest	53	\$68,623	\$4,369

Justify the use of the ANOVA test.

Solution. First of all, it does make sense to consider an ANOVA test. As we have already noted in the earlier discussion of this example, there are two ways to think of this.

- We are asking about the association between a categorical variable (region of the country) and a numerical variable (salary).
- Alternatively, we can view the study as examining the population mean salary in three populations – the mean salary for the Mid-Atlantic, the mean salary for the Southeast, and the mean salary for the Midwest.

It is reasonable to conclude that the necessary assumptions for using ANOVA are met in this situation. The participants were randomly selected, and in general salary distributions tend to be mound-shaped, so the assumption that all three population distributions are normal is not unrealistic. Recall, also, that the ANOVA test is robust. It gives good results provided either: 1) the samples are approximately the same size, or 2) the largest sample standard deviation is less than twice the size of the smallest. In this example, the samples are not approximately the same size, but we do have that $7,588 is less than twice $4,369: that is, 7588 < 2(4369).

Exercise 11: Justify the use of ANOVA for the studies described in these exercises:
  a. Exercise 3
  b. Exercise 4

A note on technology

The examples and exercises above are based on problem descriptions which describe the summary statistics, *including the standard deviation*, for the various groups. This made it easy to compare the standard deviations, checking that the largest is less than twice the smallest. But what should you do if you are working a problem in which you are given only the raw data? The answer for the calculator supplied with these lessons is quite simple. When you press *Computations*, the calculator supplies the following set of radio buttons to choose what output you wish to see:

Display:
- ◉ Results and graph
- ○ ANOVA table
- ○ Group statistics

Simply by choosing the third option you can see the sizes, means, and standard deviations for each of the groups. For example, for our original golf ball data as presented in Exercise 1, choosing the third option yields this output:

	Brand A	Brand B	Brand C	Brand D
Size	10	10	10	10
Mean	250.8	261.2	270	249.4
Std. dev.	4.7093	3.9101	4.5947	5.2324

We can easily check that the largest standard deviation (for Brand D) is less than twice the smallest standard deviation (for Brand B): $5.2324 < 2 \cdot 3.9101$.

Note: By default the groups will be labeled Group 1, Group 2, etc. To get more the more meaningful labels Brand A, etc., the author has saved the data to a text file, and modified the labels in that text file.

Comment: If you are using technology other than the supplied calculator, you will likely have to calculate the standard deviations for the groups as a separate step.

12.6 –ANOVA Tests for Data Files

In Lessons 2, 3, 9, 10, and 11 we have examined the use of the second calculator provided with these lessons to analyze the data in a data file. In this lesson we will see that this same calculator can be used to carry out ANOVA hypothesis test for the data in a data file. The data file we use is the same used in those earlier lessons. You should have it saved to your own computing device, but in any case here again is a link to that file:

First day survey

As a reminder, the file contains student responses to a first day survey containing these questions:

1. What is your gender? (M) Male (F) Female
2. What is your class year? (FR) Freshman (SO) Sophomore (JR) Junior (SR) Senior
3. How many states have you visited?
4. Do you currently smoke? (Y) Yes (N) No
5. How tall are you (in inches)?
6. How many days per week do you read a newspaper?

To begin, open the data file calculator using the following link, then use the *Load vertical file* button to load the data file containing the student responses.

Data file calculator

Example. A researcher suspects that there is a connection between the class year of the students and how many states they have visited, for the population consisting of all students who take that class. Do an appropriate test.

Solution. First of all, an ANOVA test is appropriate because one variable (number of states visited) is numerical and the other (class year) is categorical. There are several ways we can write the hypotheses. First, using the concept of "connection" or "association":

H_0: there is **no** association between class year and number of states visited
H_a: there **is** an association between class year and number of states visited

Similarly, we can express association in terms of independence:

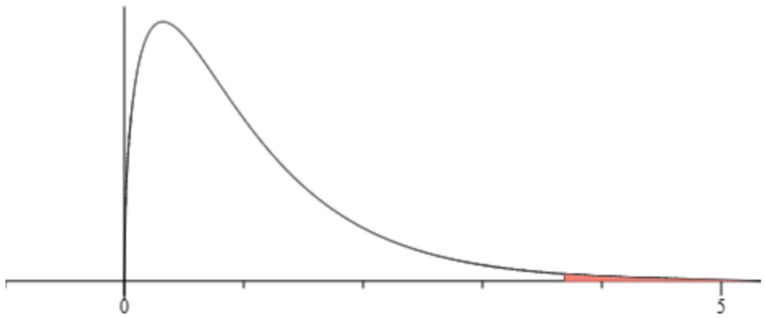
H_0: number of states visited does **not** depend on class year
H_0: number of states visited **does** depend on class year

Alternatively, we can express the hypotheses in terms of the mean number of states visited for the four populations (all freshmen, all sophomores, etc.). If μ_{FR} is the mean number of states visited for all the freshmen in the class, and so on, we write:

$H_0: \mu_{FR} = \mu_{SO} = \mu_{JR} = \mu_{SR}$
H_a: the population means are NOT identical – at least two are different from each other

Running the test is not difficult – choose menu option *Tests*, submenu option *ANOVA*, choose *States_Visited* as the numerical variable, *Class_Year* as the categorical variable, and press *Computations*. Here is the result.

F = 3.6667 *p*-value = 0.0172



Based on these results, we have evidence at α = 0.05 that the number of states visited does depend on the class year. However, there is a *caveat*: to be confident of our results, we should have at least one of these conditions met:

- the sample are all approximately the same size, or
- the largest standard deviation is less than twice the smallest.

Since neither of these conditions holds (see the output obtained by selecting the third radio button, below), we might want to take a larger sample and run the test again.

Statistics for the *States_Visited* variable, with groups determined by the values of the *Class_Year* variable

	FR	JR	SO	SR
Size	7	23	27	6
Mean	12	16.7826	14.3333	23.6667
Std. dev.	3.3166	7.0386	6.3063	12.4365

Exercise 12:
a. Test the claim that the height depends on the class year.
b. Is there an association between gender and states visited?

Exercise 13: In Exercise 16 of Lesson 2 you created a data file containing the data presented originally in Lesson 1. This data was collected in a statistics course at a public university. Use that data file to see if the number of times per week reading a newspaper depends on the political party.

Solutions to Exercises

1: The following table presents data for four brands of golf balls. The researcher randomly selected 10 balls of each brand. Each ball was struck by a mechanical driver, and the distance the ball traveled was measured.

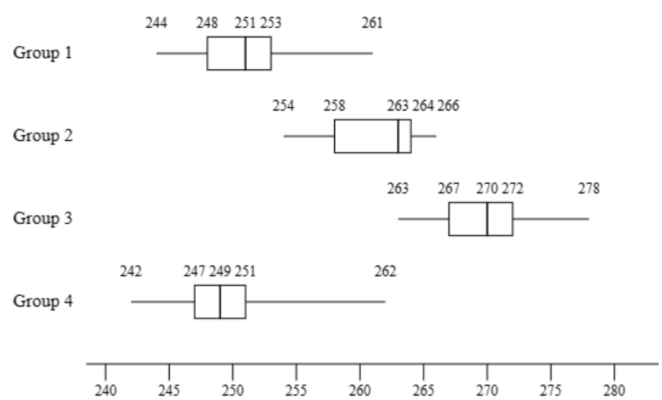
Brand A	251	244	248	251	261	251	253	246	254	249
Brand B	263	263	266	254	264	258	263	264	261	256
Brand C	269	263	278	267	271	267	271	272	276	266
Brand D	252	249	249	242	247	251	262	249	247	246

a. Calculate the mean for each of the four samples. Based on the means, does it appear that, for the golf balls in the sample, the brand of ball made a difference in the distance the ball went when struck by the mechanical driver? Circle all statements you agree with.
- No, it made no difference.
- It appears that brands A & D went a shorter distance than the other two brands.
- Brand C seems better than the other three.

The means for the fours samples are: Brand A 250.8, Brand B 261.2, Brand C 270.0, and Brand D 249.4. The second and third statement seem correct.

b. Create side by side boxplots for the data. Does this graph seem to support your conclusions?

The following plot definitely supports the conclusions.



2: In Exercise 1, we concluded that, for the balls in the sample, it appears that brands A and D go shorter distances than the others, and brand C goes the furthest. Use this conclusion to answer the following questions, for the balls in the sample.

a. Is there an association between the two variables (brand, distance)? YES
b. Are the two variables independent? NO
c. Does the distance depend on the brand? YES
d. Is there a difference between the brands, relative to the distance the ball goes when struck by a driver? YES

3: The data in the previous example was provided in terms of salary in dollars. However, it is well-known that cost of living also varies by region. It is possible to adjust each of the reported salary figures to factor in cost of living. The researchers did this, with the results shown here:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$71,733	$4,362
Southeast	63	$74,881	$7,389
Midwest	53	$72,283	$5,712

a. Comment on the mean salaries in the sample.
Mid-Atlantic and Midwest seem fairly close, with Southeast larger than those two.

b. Using the methods you will learn in later sections, we obtain a test statistic of $F = 3.6286$, with a corresponding p-value of 0.0290. At significance level $\alpha = 0.01$, does the evidence establish a connection between region and salary when adjusted for cost of living? NO, because 0.0290 is not less than 0.01

c. At $\alpha = 0.05$, did the researchers find a significant difference in the adjusted salaries for the three different regions? YES, because 0.0290 is less than 0.05

4: Researchers are studying the connection between systolic blood pressure and blood type. In a randomly chosen sample, they obtain data summarized as follows:

Blood type:	**O**	**A**	**B**	**AB**
Sample size:	8	6	4	6
Mean:	140.3750	160.6667	129.5000	113.5000
Standard deviation:	27.0235	17.0372	26.8887	13.6931

a. Which of the following are correct ways to write the null hypothesis? Correct answers highlighted.
- There is a connection between systolic blood pressure and blood type.
- The population mean systolic blood pressure is the same for all four types of blood.
- Blood type and systolic blood pressure are independent.
- Systolic blood pressure depends on blood type.
- There is no correlation between systolic blood pressure and blood type.

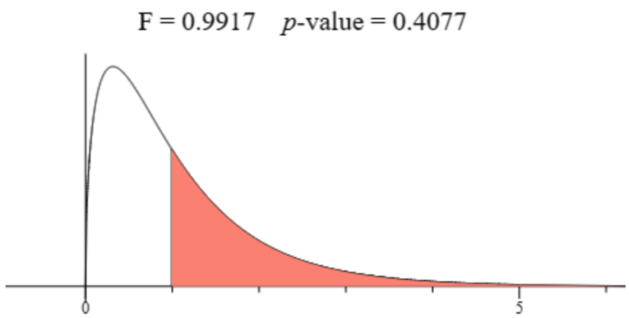
b. Which of the following are correct ways to write the alternative hypothesis?
- There is a connection between systolic blood pressure and blood type.
- The population mean systolic blood pressure is different for all four types of blood.
- At least two of the blood types have different population mean systolic blood pressure.
- Blood type and systolic blood pressure are independent.
- Systolic blood pressure depends on blood type.
- There is no correlation between systolic blood pressure and blood type.

c. Using the methods you will learn in later sections, we obtain a test statistic of $F = 4.8193$, with a corresponding p-value of 0.0110. At significance level $\alpha = 0.01$, does the evidence establish a connection between systolic blood pressure and blood type? NO, because 0.0110 is not less than 0.01

d. Write the conclusion, at α = 0.05, using the word *significant*. At α = 0.05, we found a significant connection between systolic blood pressure and blood type.

5: Refer to Note 3 above.

a. Use technology to carry out an ANOVA test for the second set of golf ball data referred to in that note.



$F = 0.9917, p = 0.4077$

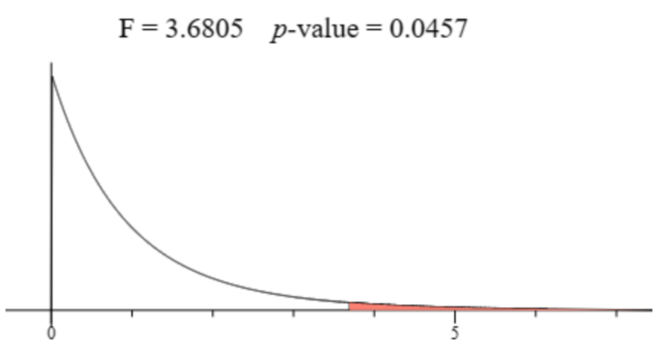
b. Do you reject the null hypothesis at α = 0.05? (YES / NO) NO

c. Does the driving distance depend on the brand at α = 0.01? (YES / NO) NO

d. Is there a significant difference between the four brands of golf balls? (YES/NO) NO

6: Use technology to carry out an ANOVA test for the following data.

Group 1	Group 2	Group 3
4	7	8
12	8	9
13	9	7
7	4	3
13	7	6
9	10	2
9		7
11		



F = 3.6805, p-value = 0.0457.

Do you reject the null hypothesis at α = 0.05? (YES / NO) YES

Do you reject the null hypothesis at α = 0.01? (YES / NO) NO

Is there a significant difference between the three groups? (YES/NO) YES at significance level 0.05 (but not at significance level 0.01)

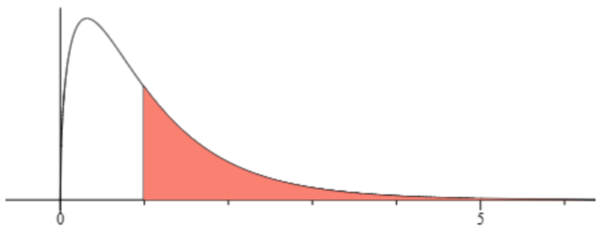
7: Refer to Note 3 above.

a. We have this information for the second set of golf ball data, where again each sample has 10 balls.

	Brand A	Brand B	Brand C	Brand D
Mean	250.8	261.2	270	249.4
Std. dev.	31.322	29.3023	31.3759	30.6529

Use this to carry out an ANOVA test.

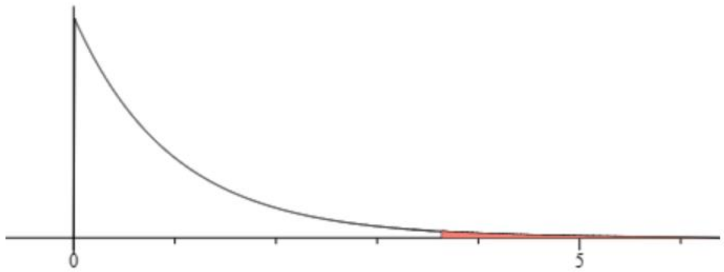
F = 0.9917 p-value = 0.4077



b. Carry out an ANOVA test for the summary information presented in Exercise 3, reproduced here:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$71,733	$4,362
Southeast	63	$74,881	$7,389
Midwest	53	$72,283	$5,712

F = 3.6292 $0.0289 \le p\text{-value} \le 0.0295$



In Exercise 3 we reported these values: F = 3.6286, p-value = 0.0290. These were obtained using commercially available software using the original data. Notice that our F-value is slightly different, due to the rounding that has occurred in calculating the mean and standard deviation for the three groups. In addition, when the groups are large, the calculator supplied with these lessons may sometimes give a range of p-values rather than a single value.

8: Circle the correct answers. (Correct answers shown highlighted)

a. F is large when (between-groups-variability / within-groups-variability) is much greater than (between-groups-variability / within-groups-variability). In this case, the p-value is (small / large) so we (reject / do not reject) the null hypothesis.

b. F is small when (between-groups-variability / within-groups-variability) is much smaller than (between-groups-variability / within-groups-variability). In this case, the p-value is (small / large) so we (reject / do not reject) the null hypothesis.

9: In an earlier example, we considered data on computer programmers' salary in three different regions of the country, with this summary information:

Region	***n***	**Mean**	**Standard deviation**
Mid-Atlantic	30	$73,267	$7,588
Southeast	63	$72,714	$6,602
Midwest	53	$68,623	$4,369

Answer the following for carrying out an ANOVA test using this data:

a. Factor df = 2 (calculated as: number groups – 1 = 3 – 1 = 2)

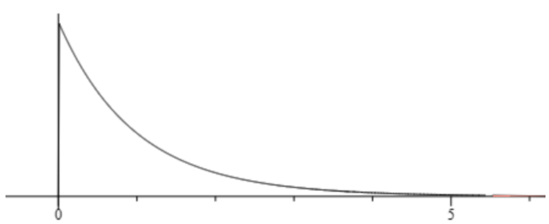
b. Error df = 143 (calculated as: total size of sample – number groups = (30+63+53) – 3 = 143)

c. Given that Factor SS = 621.7001 and Error SS = 5365.1766, calculate the *F* test statistic. Round to 4 decimal places in each step.

$$Factor\ MS = \frac{Factor\ SS}{Factor\ df} = \frac{621.7001}{2} = 310.8501$$
$$Error\ MS = \frac{Error\ SS}{Error\ df} = \frac{5365.1766}{143} = 37.5187$$
$$F = \frac{Factor\ MS}{Error\ MS} = \frac{310.8501}{37.5187} = 8.2852$$

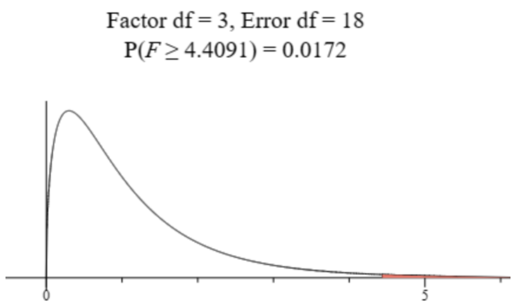
d. Calculate the *p*-value. 0.0004

Factor df = 2, Error df = 143
P($F \geq 8.2852$) = 0.0004



10: Is test performance affected by the color of paper the test is printed on? To examine this, a professor gives the same test to everyone in her class, but with the color paper randomly determined. Five students had pink paper, four had yellow, eight had blue, and five had green. Here is a partial ANOVA table showing the results generated by technology. Complete the missing entries in the table.

	df	**SS**	**MS**	**F**	**P**
Factor	3	1509.5886	503.1962	4.4091	0.0172
Error	18	2054.2750	114.1264		



NOTE: Other technology methods yield 0.0171as the *p*-value. As noted consistently throughout the calculator, it is generally accurate within 0.0001 but is NOT intended for professional use, only for classroom instruction!

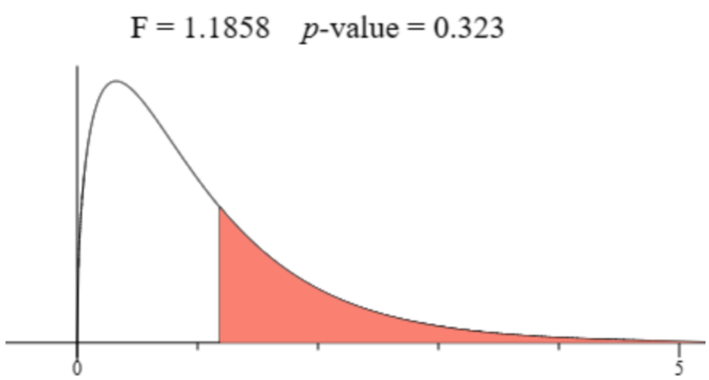
11: Justify the use of ANOVA for the studies described in these exercises:

a. Exercise 3 Although the sample sizes are not approximately equal, we do have that the largest sample standard deviation is less than twice the smallest: $7389 < 2(4362)$, that is $7369 < 8724$

b. Exercise 4 One could make the statement that the samples are approximately the same size, although all are pretty small. Also, 27.0235 < 2(13.6931), that is $27.0235 < 27.3862$, so the largest sample standard deviation is less than twice the smallest.

12:

a. Test the claim that the height depends on the class year.



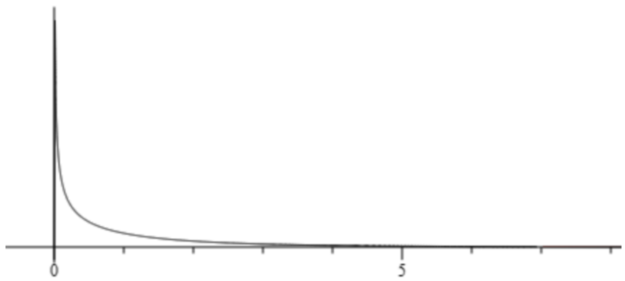
We do not have evidence to conclude that the height depends on the class year. NOTE: The use of ANOVA is justified; as shown below, the largest standard deviation is less than twice the smallest: $4.6413 < 2 \cdot 3.4847$.

Statistics for the *Height(in)* variable, with groups determined by the values of the *Class_Year* variable

	FR	JR	SO	SR
Size	7	23	27	6
Mean	70.8571	69.1304	67.8148	68.3333
Std. dev.	3.4847	3.5201	4.6413	3.7238

b. Is there an association between gender and height?

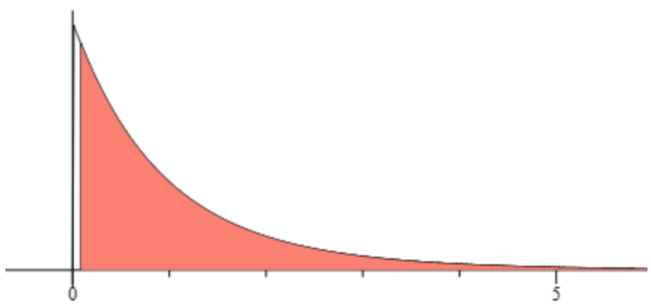
F = 55.1226 p-value = 0



Not surprisingly, the answer is yes. NOTE: The use of ANOVA is justified; as can be seen by choosing the third radio button, the largest standard deviation is less than twice the smallest: $3.221 < 2 \cdot 2.8301$.

13: In Exercise 16 of Lesson 2 you created a data file containing the data presented originally in Lesson 1. This data was collected in a statistics course at a public university. Use that data file to see if the number of times per week reading a newspaper depends on the political party.

F = 0.0851 p-value = 0.9187



The answer is no. NOTE: The use of ANOVA is justified; as can be seen by choosing the third radio button, the largest standard deviation is less than twice the smallest: $1.3119 < 2 \cdot 1$.