Lesson 1.1.1

The Statistical Analysis Process

INSTRUCTOR SPECIFIC MATERIAL IS INDENTED AND APPEARS IN GREY

ESTIMATED TIME

1 hour 40 minutes

MATERIALS REQUIRED

- Chart of personality traits by birthday (Lesson 1.1.1 Supplement – Astrology Investigation).
- Three index cards per student.
- Lesson 1.1.1 Match Applet, an Excel file.
- Lesson 1.1.1 Working in Groups, a Word file.

BRIEF SUMMARY

This lesson introduces students to the four-step process in Statistics. Students learn about this process by collecting data from the class to learn if someone's astrological sign is associated with their personality. Students summarize the four steps at the end of the activity, and then proceed to summarize the four steps for two other statistical studies.

LEARNING GOALS

Students should understand that:

- Statistical investigations are a cycle of forming a research question, designing a study and collecting relevant data, data analysis, statistical inference, and drawing appropriate conclusions.
- Chance variability plays a role in the statistical decision-making process.
- Unusual results can provide evidence for a claim about a population proportion.

Lesson 1.1.1
The Statistical Analysis Process

Students should be able to:

- Identify the steps in the statistical analysis process given a description of a statistical study.
- Interpret data displayed in a dotplot.
- Evaluate the strength of evidence against a claim about a population proportion.

INTRODUCTION

The four-step process is used throughout the course both in the lessons and in MyStatway. We open with a brief summary of the reason for studying statistics and then introduce the four steps. Following this the students examine how we can use the tools of statistics to determine whether astrology has any basis in reality.

This lesson is designed to be 100 minutes long. If your class is shorter, you will have to find a good place to stop. Make sure to leave the students with a preview of how the lesson will conclude.

 STUDENT MATERIAL IS NOT INDENTED AND APPEARS IN BLACK

INTRODUCTION

What is statistics? Why do we study statistics?

Statistics is about using **data** to answer questions. Data is information that we collect from our world. Data involves facts and observations that we make. Before scientists created statistics and before people used data, they would use opinions and hunches to explain how the world worked. A lot of times these explanations were wrong.

For example, people once believed that the earth was the center of the universe. When people started making observations and using measures they discovered that this was incorrect. The earth was not the center of the universe. Data helps us make better conclusions. In this example data helped us see that the earth goes around the sun, not the other way around.

Data can help us answers many types of questions.

- Students can use data to help pick a college that is best for them.
- Teachers use data to find the best ways to educate their students.
- Medical professionals use data to learn if new treatments actually work.
- Voters need data about their society and planet to create a better democracy.
- Politicians use data to better represent the people who elect them.

Lesson 1.1.1
The Statistical Analysis Process

In statistics, we gather, summarize, and analyze data to search for answers to our questions.

> **Instructor Note:** Question #1 is the first time students actually work through a rich task in groups. This is an open-ended question that can be answered well in many ways. Allow students to use their imaginations, and then have them share their ideas with the rest of the class. Tell students that there are no correct answers to this question. We just want them to come up with some possible ideas.

Let's begin with an idea that will help us think about how statistics can help us answer a question that we may have.

1 Imagine that you want to learn whether someone's birth date can influence her or his personality. You know that data involves facts, observations, and measures about a particular topic or idea. If you wanted to gather information about birth dates and personality, what kind of information would you look for? How could you use data to answer this question?

> **Answer:** One possible answer is that students may ask people for their birth date and then ask about their personality. They might suggest comparing answers to questions about personality for people with similar birthdates. There are many possibilities.
>
> Instructor Note: The point is to get students thinking. Do not correct students here, although helping students to elaborate on their ideas is a good thing. Try to draw out student reasoning.

Statistical analysis is the process of looking at data to learn about something bigger. Looking at data allows us to make **generalizations** about **populations** that are large and difficult to understand. We can think of the statistical analysis process in 4 steps.

Steps in a Statistical Analysis

1. **Ask a question that can be answered by collecting data.**
2. **Decide what to measure and then collect data.**
3. **Summarize and analyze the data.**
4. **Draw a conclusion and communicate the results.**

We will now do an activity that will help us learn about the statistical analysis process. In this activity we use the statistical analysis process to investigate a question about whether we can use astrology to predict personality traits. We are answering the same question we saw above: Can someone's birth data influence his or her personality?

Lesson 1.1.1
The Statistical Analysis Process

Signs of the Zodiac and Personality Traits

You can read the section below to your students as an introduction to the activity:

For this task, you need to know the difference between astrology and astronomy. In both astrology and astronomy, people study the stars, sun, moon, and planets in outer space. But, they are very different.

Astronomy is a science that studies the movement of stars, moons, planets and other celestial bodies in outer space. Scientists who study astronomy are called astronomers. Most people do not debate if astronomy is real because it is a science.

In astrology, people believe that the positions of the stars, sun, moon, and planets influence human actions. People that study astrology are called astrologers. In astrology, you have a zodiac sign based on your birthday. On your birthday, the sun passes through one constellation and this constellation is your zodiac sign. There are 12 zodiac signs. Many people believe that these 12 zodiac signs influence the personality traits people have.

Note: You could mention that the actual dates of when the sun passes though each constellation do not coincide with the actual signs anymore. Astronomers (not astrologers) recalculated the zodiac calendar based on the actual dates. The links below are about this discovery.

Reference:
http://www.post-gazette.com/pg/11014/1117842-153.stm and
http://www.csmonitor.com/Science/2011/0114/New-zodiac-signs-2011-Can-one-guy-just-change-the-zodiac-like-that).

(Here pass out the sheets with birthdates and personality traits)

Can a person's birthday determine his or her personality traits, like being kind or jealous? Your instructor will give you a list of personality traits. These traits are grouped by birthdays. Each birthday group has 3 sets of personality traits. Only 1 set actually goes with that birthday group (according to the *zodiac* calendar). The other 2 sets are not associated with the zodiac **prediction** for the birthday group.

Look for your birthday on the handout. Read the 3 sets of personality traits for your birthday group. Choose the set of personality traits that you think matches your own personality. None of the choices will match your personality perfectly. Pick the choice that describes your personality the *best*.

Students may not know some of the personality traits. If they have questions, tell them what the traits mean or tell them to use the traits that they know from each group to make their choice. Also emphasize that students are supposed to pick based on which set matches their personality. They are not trying to guess which goes with their zodiac sign.

Lesson 1.1.1
The Statistical Analysis Process

Write the number for your choice below.

Choice #_____.

Let's take a closer look at the **statistical analysis** process. We want to investigate if the zodiac's personality predictions are likely to be correct. Before going into great detail, let's think about those steps.

TRY THESE

Before students engage in an activity requiring Productive Struggle we want to help them frame the activity so that they don't become so frustrated that they give up. We also want to forestall the conclusion that more struggle means they can't cut it - that they lack the potential to succeed. Here is script that you can use to help set the scene for students to Productively Struggle with the Statway™ materials.

One important part of the Statway™ is to do problems that are going to stretch you beyond what you know. Although this may be uncomfortable sometimes, we do this on purpose so you can learn. It's important to know that when you feel stretched, it's not a sign that you can't be successful. Usually it's a sign that you're learning.

In the past, when some people get a problem wrong, they might have thought that they just don't have the ability to study math--that they're not math people. But when you talk to professional mathematicians, the people who are best at math, it turns out that they work a long time on the same problem--and they only spend their time on problems that they struggle with the most. And even though you might think they make up answers on their own, almost always mathematicians have to ask people for help.

Like a professional mathematician, the goal of the Statway is to help you learn things that you can actually use in life--not so you can memorize it for a test and then forget it. And so in the Statway™, sometimes we will work on problems that either don't have a correct answer or that, for most, the answer isn't obvious right away. At those times, it's important to realize that the way that you thought about math in the past--like getting right means you are smart and getting in wrong means that your aren't--that type of thinking just isn't true in the Statway™. Instead we will work on problems that take three things: effort, a good strategy, and help from others. If you get it right, it's because of one of those three things, and if you get it wrong, then you know what to do: change your effort, change your strategy, or ask for more help.

Lesson 1.1.1
The Statistical Analysis Process

> Have students think about the following questions alone for about 3-4 minutes and then have them discuss it in groups for about 5 minutes.
>
> To help you facilitate group work, refer to included Word file: Working in Groups
>
> Emphasize what we expect students to do during this time – i.e. TRY to answer the questions. There are no wrong answers here. They will pull this all together during the wrap-up. The goal is for students to become comfortable TRYING to answer. The largest goal in this activity is setting the classroom norm for participation and safety, NOT correctness. There will be more of this in the Mindset Activity in Lesson 1.1.2, but you can briefly talk to them that thought exercises like these are a hugely important part of the learning process, regardless of the answer they come up with.

2 Answer each of the following questions about the four steps of the statistical analysis process.

Step 1: Ask a question that can be answered with data.

A What question are we trying to answer in our investigation of the zodiac birthday groups?

> **Answer:** Can we use the zodiac birthday groups to predict personality traits?

Step 2: Decide what to measure and then collect data.

B What information can we get from each student? Are the data related to the question we are trying to answer?

> **Answer:** We asked each student which set of characteristics best matched their personality traits. If the zodiac calendar can predict personality traits, most people should pick the set of traits that matches their zodiac sign.

Step 3: Summarize and Analyze Data

C Each student chose a set of personality traits. Some choices will match the zodiac predictions, and others won't. How could we *summarize* all of the individual student responses?

> **Answer:** We could count how many students picked the set of traits that match their zodiac sign or we could compute the percentage of students who picked the set of traits that match their zodiac sign. We might look to see if more than half of the class picked the matching traits.

Lesson 1.1.1
The Statistical Analysis Process

Step 4: Draw a conclusion and communicate the results.

D Once the data are summarized, how can we use the summary to answer the research question? What would we expect to learn from your data? Can we generalize, or **infer**, the results to something bigger than our class?

Language Tip
To *infer* means to use statistical evidence to make a conclusion that applies to a larger group than the sample.

Answer: If most students pick the set of traits that matches their zodiac sign, this might mean that astrology can be used to predict personality characteristics. If only a few students pick the matching set of traits, this might mean that astrology doesn't work.

WRAP-UP

Have a discussion about the best way to determine whether the zodiac is good at making predictions about personality traits. If computing percentages or proportions of correct predictions has not been mentioned, steer the conversation toward this. Have students talk about the value of the proportion of correct predictions if the zodiac does not work, and also if it does work.

Lead a discussion with students on how we might be able to rule out chance as a possibility for deviations from the expected results. Move the discussion toward the idea of a simulation that would allow us to see what type of results we would expect to see by chance alone.

Note: The overall goal is to give students a sense of what a statistical analysis is all about. You can talk about the data collection plan and whether it appears sound (e.g., Are the responses by this class representative of a larger population? The larger focus here is actually on the *inferential process*. Do the results in this experiment generalize to something bigger? What sorts of outcomes would we expect in an experiment through pure chance? If you can rule out chance as an explanation for a class result, you have evidence to support the zodiac. Make sure that students see the big picture (the four step process and ruling out chance). If students do not see the big picture, you can return to the coin example; ask students whether/why they would be surprised for you to toss 10 heads in a row. They should be able to respond "because that would not happen by chance."

NEXT STEPS

As we proceed, we will address the question of *whether a person's birthday group (their zodiac sign) plays a part in determining his or her personality traits*. We have actually completed the data collecting process by having everyone record the personality traits that match them best.

Lesson 1.1.1
The Statistical Analysis Process

We now prepare to *summarize and analyze* our data. To do this, we need to think about how the data can be used answer our question. We will use *probability* to help in this. Talk about the following questions with your group to begin this process.

3 In the first part of this lesson you chose one set of traits that best matched your personality. The zodiac defines personality traits based on birth date. Suppose that personality traits *are not* related to birth dates. Will the choices students make match the zodiac prediction from time to time? Explain why you think this.

> **Answer:** Yes, some students will pick the matching traits just by chance.

Suppose that personality traits *are not* related to birth dates. Out of all students in your class, about what **fraction** of choices should match the zodiac personality predictions? Why do you think this?

> **Answer:** About 1/3. One third of the choices match the zodiac.

4 Suppose that personality traits *are* related to birth dates. Would you expect the **fraction** of student choices that match the zodiac prediction to be *greater* or *less than* the fraction that you chose above?

> **Answer:** Greater.

Think about the entire class. What fraction of the class would need to choose traits matching the zodiac for you *to be convinced* that someone's zodiac sign can predict their personality traits? Tell why you think this.

> **Answer:** Answers vary. Many may say more than ½. The number should be enough over 1/3 that we can be convinced.

5 Imagine that half the students in your class select traits matching the zodiac. Does this *guarantee* that the theory could be true? If not, give another explanation for why so many students picked the correct set of traits.

> **Answer:** It is not a guarantee. Maybe more than ½ of students in our class would pick the matching traits just by chance.

Lesson 1.1.1
The Statistical Analysis Process

NEXT STEPS

Using probability to discover what may happen by chance

> **Note:** The term *match* is used to convey students' picking traits that match zodiac predictions. Later, students write match/no match on index cards. Make it clear to students that the cards represent the correct and incorrect traits for each zodiac sign. It is also important to make clear what the term "sufficiently greater" means in this context. How will students gauge whether something is "sufficiently greater"?

If the zodiac sign has nothing to do with personality traits, then we expect the fraction of students in the class who pick the personality traits that corresponds to their zodiac sign to be around 1/3. But how far above 1/3 would convince us that the opposite is true? We will answer this by learning about the values that are most probable. Probable values are those that are most likely to occur by **chance variation**.

Language Tip

***Chance variation* describes the type of differences we would naturally expect to see between different samples.**

> **Discussion Question A**: Why do you expect the fraction to be around 1/3? That is, why is the fraction 1/3 used?
>
> **Note:** Students were given three descriptions to choose from. If they are choosing at random, you expect them to choose the correct description about one out of three times. You might consider drawing this on the board so that students have a visual representation of your explanation.
>
> **Discussion Question B:** Why is it "around 1/3" and not "exactly 1/3"?
> **Note:** You do not expect a class of students to hit the 1/3 mark exactly. You may get a little more or a little less than 1/3 by chance. The 1/3 applies over thousands and thousands of attempts.
>
> Earlier in this lesson, we asked if you think your birth date or zodiac sign *determines* your personality traits. In statistics, we need to use more careful language, because it is very difficult to say for sure that something determines or causes something else. Instead, we should have said: "is there a *relationship* between personality characteristics and zodiac sign?" All we can tell from the data we collected is if there is a relationship between your zodiac sign and your personality characteristics.
>
> To be *convinced* or persuaded that there is a relationship between personality characteristics and zodiac sign, the following needs to happen:
>
> More than 1/3 of the class needs to pick the matching set. This is the set of personality traits that matches their zodiac sign. If students in class picked 1 of the 3 sets at random, it would not be surprising that a little more than 1/3 picked the matching set by **chance**.

Lesson 1.1.1
The Statistical Analysis Process

If we found out that more than 1/3 of students picked the matching set "by chance," then this means that it was just random or by accident. This would not show or prove that there is a relationship between personality characteristics and zodiac sign.

How do you figure out that our results are not "by chance"? To be convinced, you need a fraction that is **sufficiently greater** than 1/3 that could not be explained just by chance. To be sure that "chance" does not explain our results, you need to *rule out* chance as a possible explanation. **To rule out chance as a possible explanation, you need to have an idea of what values to expect for the fraction of the class who pick the matching set just by chance.** This means that you need to have an idea of how large the fraction of students would be who picked the matching set by chance

Note: For this section, you will give students 3 index cards, one with the word "Match" and the other 2 with "No Match" written on them. Introduce this activity by explicitly telling students that these cards represent the three possible choices on their supplementary handout. This is why there is one card labeled "Match" and two cards labeled "No Match." They are going to choose 1 of the 3 cards randomly and then use this data to calculate the class proportion. They will repeat this several times to try and see a pattern. The purpose of this activity should be made explicit to students so that they can make a link between task 1 and task 2.)

6 Your instructor has given you three cards, one with "Match" and two with "No Match" written on them. Mix the cards and choose one of them. Complete the following information.

- My card says:
- Number of students in the class:
- Number of students who randomly selected the card that said "Match":
- Fraction of the students who randomly selected "Match":
- **Decimal proportion** of the students who randomly selected "Match":

In statistics, a **proportion** is a number between 0 and 1. It represents a portion out of the total. We usually give proportions as decimals or percents. We can calculate a decimal proportion by dividing the **numerator** of a fraction by the **denominator**. For example, if the fraction of students who picked *match* is 5/7, then you divide 5 by 7. The proportion would be 0.714. To change to a percent we multiply by 100 or move the decimal 2 places to the right. 0.714 = 71.4%.

Note: Of course results vary from class to class, but make sure students understand the distinctions between *count*, *fraction*, and *proportion*. You can tell them that proportions can be expressed as fractions, decimals, or percents, but we rarely use fractions since they are harder to compare. This might be a good place to define 'match proportion' as the proportion of students whose card says "Match". Be very consistent in your

Lesson 1.1.1
The Statistical Analysis Process

language, always referring to the proportion of something. Students will see proportion throughout the course.

You can have students do 6 and 7 in their groups or do these as a faculty facilitated discussion. Questions 8 and 9 are more complicated and will probably work best with faculty facilitated discussion.

7 Is the proportion of students in the class who picked the *match* card equal to 1/3? If not, is it greater than or less than 1/3?

Answer: Answers will vary.

8 If the class repeats this process a second time, would we get exactly the same *match* proportion for the class? Why or why not?

Answer: When we choose cards a second time we may not get the same answer since the picks are random.

Instructor Note: Have a class wide discussion at this point to help students with the following ideas.

Discussion Question: Each of your classmates mixed the cards and picked one at random. We then computed the proportion of students in the class who picked the *match* card. Is knowing this one proportion enough to determine the values of the "Match" proportion we expect to observe just by chance?

Answer: This just tells us *one value,* but what we really need to see is a **pattern** of values. This could have been an unusual outcome. We want to look at many outcomes to see how what values come up most often.

Discussion Question: Knowing this proportion is not enough. What additional information would help us get a better understanding of the proportion of matches that are likely to occur by *chance*?

Answer: Make sure that the discussion ends up with "we need to try this more times."

We will repeat the process of observing *chance* outcomes a large number of times. This will help us understand what kinds of match proportions are consistent with picking cards at random.

9 First, write the *class proportion* of matches in the following table under **Observed Proportion** next to Trial 1. Second, work with your classmates to repeat the process of picking cards until you have proportions for 10 trials.

Lesson 1.1.1
The Statistical Analysis Process

Fill in the table below with the proportions of students who picked *match*. (*Observed Proportion* means the proportion we see from the data we collect in each trial.)

Trial	Observed Proportion
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Note: You may want to explain to students that a *trial* is one round of an experiment and that the "observed proportion" is the proportion obtained in each trial. Have students shuffle the cards, pick one card, compute the class "Match" proportion, and enter the value into table. Repeat until 10 proportions have been calculated.

Introduce a way of graphically displaying the data [the dotplot] using the following material:

> A convenient way to summarize all the observed proportions in the table is to construct a graphical display. You will learn about graphical displays in Module 2, but for now we will construct a simple graphical display called a *dotplot*. In a dotplot, each numerical value is plotted as a point along a number line. "

Plot the 10 observed class proportions on the board for students to see. You can get a head start by drawing the number line while the students are calculating their proportions.

10 Answer these questions after your professor has constructed a **dotplot** which shows the observed "Match" proportions for the 10 trials where the students picked one of three choices at random.

A What was the smallest *match* proportion observed?

B What was the largest *match* proportion observed?

C Did the match proportion differ much from trial to trial?

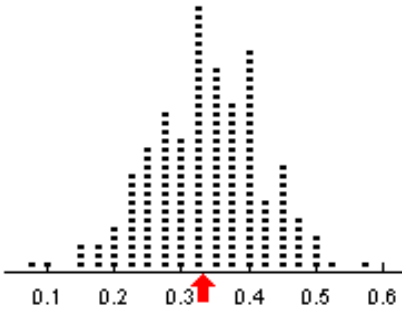
Lesson 1.1.1
The Statistical Analysis Process

Note: You will use the Lesson 1.1.1 Match Applet to carry out 100 trials and build a dotplot that shows the sampling variability in the match proportion. To use the applet, enter the theoretical match proportion (0.333) and the class size (varies). Hit "Enter" after typing in the class size and the applet will generate a dotplot of 100 sample proportions. Tell students that each dot represents the proportion of matches in one class trial with the index cards.

Introduce the applet by saying something along these lines:

> "To really understand how the observed proportion varies from one trial to another, we need to look at many more trials. This process can take a long time, so we will turn to technology to perform more trials."

Results from the applet will vary by class size but may look something like this, centered around 1/3.



Confirm with students that they understand that each dot represents the proportion of matches in one set of *n* picks [*n* = class size] where each pick was made completely at random.

Note: Once the applet has been used to create the dotplot, this next part can be done as a whole-class discussion, as a mini-lecture, or in small groups. Be sure that the following questions are addressed.

11 Where is the dotplot centered? Why do you think this is?

Answer: This dotplot should be centered around 0.33 or 1/3. We expect most proportions to be fairly close to 1/3.

Lesson 1.1.1
The Statistical Analysis Process

12 Summarize what your dotplot tells you. Why did you create this graph?

> **Answer:** This dotplot tells us the various proportions we might see if we repeated the match card activity many times.
>
> **Note:** Remind students that they wanted to know about the chance variation in the proportion of matches under the condition that everyone was picking at random. Chance variation is an important term that you should explain to students. Also students may need an explanation about the kinds of answers that you expect in a "summary".

A **distribution** of data (like match proportion) shows what the possible values are and how often each of the values occurs. We can compare the distribution of *match* proportions with the activity on zodiac sign and personality traits. In both cases you picked one item out of three. First you chose one set of personality traits out of three. Next you picked one card out of three. If the zodiac does not predict traits well, each activity should result in similar proportions – the zodiac should be right about one time in three, and you would have picked a "match" card about one time in three.

13 Look at the dotplot. How can this graph help you decide whether your class's proportion of correct personality traits would support the zodiac sign theory?

> **Answer:** This graph shows the types of proportions we might expect to see if there is nothing more to astrology than chance. If the proportion of matches in our class is higher than what we see in the dotplot, that would be a surprising result and might mean that someone's zodiac sign can be used to predict personality traits.

14 Use the dotplot to answer the following question.

If students in the class were picking at random, it would be *unusual* to see a match proportion as large as _____.

> **Answer:** The most likely answers are around 0.5
>
> **Note:** Class results vary, but student answers should be a value in the "tail" of the distribution generated in the applet, like above 0.45 or 0.5. Students may benefit from a brief discussion of the structure of distributions e.g., what is the "tail"? You could put labels on the dotplot (like center and tail).

15 Suppose that the proportion of students who picked the set of personality traits that matched their zodiac sign was 0.40. Would this convince you that personality traits and zodiac signs are related? Explain your answer using the dotplot.

Lesson 1.1.1
The Statistical Analysis Process

> **Answer:** Although 0.4 is above 0.333, it is not very far out in the tail of the dotplot. This result does not appear that unusual and therefore probably would not convince us that personality traits and zodiac signs are related.
>
> **Note**: From the dotplot, students should get a sense that 0.4 is closer to the middle of the distribution rather than way out in the tail.

16 How large does the proportion of students who picked matching sets have to be to convince you there is a relationship between personality traits and zodiac sign? Look back at your answer to questions 14 and 15 to help you answer this question.

> **Answer:** If the proportion is around 0.5 or higher, this is a very unusual outcome if personality characteristics are not related to sign according to the dotplot. So, if the class results are that large, this could mean that students had more matches than expected just by chance and there could be something to the sign theory.

Draw a Conclusion and Communicate the Results.

> **Instructor Note:** We are about to show students the correct choices for their zodiac sign. They will enjoy this part, seeing whether they got it "right" or "wrong". Remind them we are most interested in whether they matched the choice or not.

Now it is time to take a look at the *actual* class data gathered on personality traits. You will decide if the results provide evidence that there is a relationship between personality traits and zodiac sign. Fill in your answers:

Choice that best matched your personality:

Set that is the correct match for your zodiac sign:

Did you match the correct set of traits for your sign?

Lesson 1.1.1
The Statistical Analysis Process

Number of students in the class:

Number of students who picked the correct match for their zodiac sign:

To provide a *summary* of your class's data – compute the proportion of correct zodiac matches:

> **Answer:** Correct matches are

Aries (3/21 to 4/20)	Choice 1		Libra (9/24 to 10/23)	Choice 1
Taurus (5/21 to 5/21)	Choice 2		Scorpio (10/24 to 11/22)	Choice 2
Gemini (5/22 to 6/21)	Choice 3		Sagittarius (11/23 to 12/21)	Choice 3
Cancer (6/22 to 7/22)	Choice 1		Capricorn (12/22 to 1/20)	Choice 1
Leo (7/23 to 8/22)	Choice 2		Aquarius (1/21 to 2/19)	Choice 2
Virgo (9/23 to 9/23)	Choice 3		Pisces (2/20 to 3/20)	Choice 3

Decision time! It is now time to make an inference. An inference is a decision based on the evidence we have gathered.

17 Does the class proportion of correct zodiac matches provide convincing evidence that personality characteristics are related to zodiac sign? Why or why not? Explain your answer using question 18 and the dotplot in your reasoning.

> **Instructor Note:** Students need to compare the class results to the dotplot generated by the applet and see whether the class results are in the tail of the distribution. If the class proportion is in the tail, this provides evidence that personality characteristics might be related to sign or at least convincing evidence that random chance alone is not a sufficient counter argument.
>
> **WRAP-UP**
>
> Discuss the overall process students have experienced here—state a theory, collect data to test the theory, and consider what kind of evidence provides support for this theory. Students then evaluated two possible explanations—the theory is correct *or* the theory is not correct and the observed result is due to chance variability. To conclude that there is evidence in support of the theory, they need to rule out chance as a plausible explanation for what was observed. Students do this by generating results assuming the theory is false and what they are seeing is only due to chance variability; they then compare the observed results to see whether they can be considered consistent with chance.

Lesson 1.1.1
The Statistical Analysis Process

NEXT STEPS

> **Estimated time**: 30–40 minutes

The task you have just completed with the astrology example illustrates the **statistical analysis process**, which we have described in four steps. These are given again below.

Steps in a Statistical Investigation

1. Ask a question that can be answered by collecting data.
2. Decide what to measure and then collect data.
3. Summarize and analyze the data.
4. Draw a conclusion and communicate the results.

A statistical investigation is an ongoing process. Often, researchers analyze results of one study and this leads them to think of other research questions. Then they conduct more research. Also, researchers start to think more carefully about how they collect data and this may also lead them to think of ways to improve the data collection process.

> **Note:** You may use whole-group discussion or group work to fill in the following table for the astrology example. Answers are italicized and do not appear in the Student Handout.

18 Identify each step of the Statistical Analysis Process for the astrology investigation below.

Steps in Statistical Analysis	For the Astrology Investigation
1. Ask a question that can be answered by collecting data.	*Is a person's astrological sign related to their personality traits?*
2. Decide what to measure and then collect data.	*We had each student choose one from among 3 sets of traits and checked whether it matched those predicted by their zodiac sign.*

Lesson 1.1.1
The Statistical Analysis Process

3. Summarize and analyze the data.	*We calculated the proportion of matches for the students in class and compared it to what might happen by chance.*
4. Draw a conclusion and communicate the results.	*If our results were higher than we would expect by chance, then we saw that as evidence for a relationship between zodiac sign and personality.*

19 Look at the study descriptions on the next two pages. For each study, identify the four steps of the statistical investigation process to complete the tables.

Study 1 – A Study about a Population

A group of researchers studied women who visit a fertility clinic. The researchers wondered if less than half of women who visit the clinic would want to choose the gender of their future child.

They mailed a survey to women who had visited the clinic. The survey asked women if they would choose the gender of their future child, if they were able to do so. Five hundred sixty one women responded to the survey. Of these 561 women, 229 said that they wanted to choose the gender of their future child.

The researchers did a statistical analysis of the data. Based on their statistical analysis of these data, the researchers concluded that there is **convincing evidence** that *less than half* of women who visit the clinic would choose the sex of a future child. The researchers based this conclusion on this fact:

> *In a sample of 561 women, it would be very unusual to observe a percentage as low as 229/561 ≈ 41%. This would be very unlikely if the true percentage was not less than 50%.*

Language Tip

***Convincing evidence* is information that provides very strong support for a conclusion. The evidence is beyond what would be expected due to chance variation.**

Lesson 1.1.1
The Statistical Analysis Process

Now that you know the details from Study 1, complete the four steps of the statistical analysis process in the table below.

Steps in Statistical Analysis	Study 1
1. Ask a question that can be answered by collecting data.	*Do less than half of women visiting fertility clinics want to choose the sex of their child?*
2. Decide what to measure and then collect data.	*Surveys were sent to women who visited the clinic to ask whether they would want to choose the gender of their child.*
3. Summarize and analyze the data.	*41% of 561 respondents said they would choose a girl.*
4. Draw a conclusion and communicate the results.	*There is evidence that less than half of women would want to choose the gender of their child.*

Study 2 – A Study about an Experimental Treatment

Researchers wanted to know if people think a task will be hard to accomplish when the instructions are difficult to read. To answer this question, researchers randomly divided twenty student volunteers into two groups of 10 students each. Researchers gave instructions to each group of students using different fonts (see below). Instructions for one group were written in a large upright font. The other group was given the *same* instructions but in a font that used *hard-to-read italics*. Researchers asked students to read the directions and say how many minutes they thought the task would take. Researchers did this in order to figure out if the fonts used for the instructions made a difference.

This is the easy-to-read upright font that was used in the study.

This is the hard-to-read italic font that was used in the study.

The first group of students, those that read the instructions printed in the easy font, had an average time estimate of 8.23 minutes. The other group, the group that read the instructions in the *hard-to-read italic* font, had an average time estimate of 15.1 minutes.

Researchers concluded that such a large difference between the averages was not likely to have occurred by chance. There was evidence that people think a task will be harder when instructions are difficult to read.

Lesson 1.1.1
The Statistical Analysis Process

Steps in Statistical Analysis	Study 2
1. Ask a question that can be answered by collecting data.	*Do people think a task will be harder if the instructions for the task are harder to read?*
2. Decide what to measure and then collect data.	*Students were asked to estimate how long a task would take using different fonts, where one was harder to read.*
3. Summarize and analyze the data.	*The average time estimate for the task with the hard-to-read font was 15.1 minutes compared to 8.23 minutes with the easier-to-read font.*
4. Draw a conclusion and communicate the results.	*There is evidence that people think a task will be harder when the instructions are harder to read.*

20 Both Study 1 and Study 2 follow the same general process but they are different in some ways. What are two ways that these studies are different?

> **Note:** Focus on the type of question that is asked [one asks about characteristics of a population, and one asks about the effect of some treatment] and the manner the data were collected [sampling versus experimentation]. This leads in to the next lessons.
>
> **WRAP-UP**
>
> The main point here is for students to get an overview of the *types of studies* they will see and will be able to analyze. Emphasize that the reasoning process to judge whether a result is significant follows very similar lines to the astrology study. Also emphasize that the other steps (deciding what to measure, obtaining valid measurements, reporting conclusions) are also critical to this process and should always be given careful attention as well.

Lesson 1.1.1
The Statistical Analysis Process

STUDENT NAME _____ **DATE** _____

TAKE IT HOME

> The following is an introduction to the Homework section that you can choose to deliver to the class:

In the lesson, your class looked at two different sets of data:

- Hypothetical Data: These were the data collected from match card activity and the applet, where you were given three cards (2 cards that had "no match" and 1 that had "match"). These were *hypothetical* data because we used a computer to create much of the data. The dotplot the class looked at showed the data on a graph. The hypothetical data allowed us to see what the proportion and dotplot would look like when there was *no relationship* between personality traits and zodiac sign. We saw that if there is no relationship between zodiac sign and personality traits than about 1/3 (or 0.33) of the class will picking the matching traits.

- Actual Data: These were the data collected from when you had to read three sets of personality traits that were under your birthday.

You compared the *hypothetical* results to the *actual* result for the class. You did this to see whether the proportion in the class was consistent with **chance variation**. If the class result was unlikely to occur by chance alone, this gave some evidence to the astrological theory.

1 Imagine the investigation had given each student a choice of *four* sets of personality characteristics for each sign instead of three. If there is *no relationship* between personality characteristics and sign, about what fraction of the students in the class do you expect to pick the description that astrologists say matched their sign? Why do you think this? (Remember, the "matching set" is the set of traits that astrologists say matches a person's zodiac sign.)

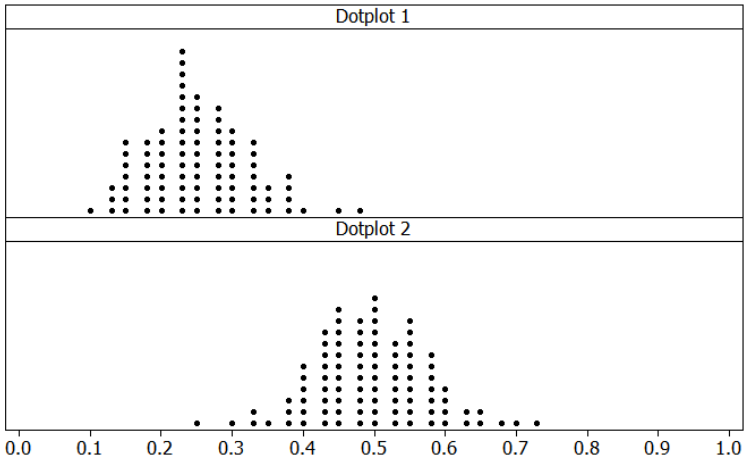
> **Answer:** Now there is one correct choice of four matches, so a correct match occurs by chance about one time in four, for a match proportion [in the long-run] of about 0.25.

2 There are two dotplots below. One of the dotplots was made by asking 40 students to pick one of four cards at random. Each card had a set of traits written on it. Researchers computed the proportion of students that chose the card with the set of traits that matched their zodiac signs by chance. This process was repeated a large number of times to generate the data used to construct this one dotplot.

Out of the two dotplots below, which dotplot do you think is the one that was constructed this way?

Why did you pick this dotplot?

Lesson 1.1.1
The Statistical Analysis Process



> **Answer:** These data represent what happens when people are just guessing, so the proportions should center around 0.25. The correct choice is Dotplot 1.

3 Imagine that each of these 40 students then made their choices from the list of four personality types. What proportion of the 40 students needs to match correctly to provide convincing evidence that there is a connection between sign and personality type? Explain your reasoning. (Helpful hint: Use your answer from Question 2.)

> **Answer:** Results vary a bit, but a proportion of 0.35 or larger seems reasonable as this puts the observed proportion in the tail of the distribution [Dotplot 1]. With Dotplot 2, they might say anything above 0.6.

4 Read the following study description:

The United States Government recommends that to stay physically fit, middle-aged adults (ages 40 to 60) need to use 150 to 400 calories per day doing exercise. Researchers at Minnesota State University wanted to learn whether middle-aged adults who used the Wii Fit video games exercised enough to meet the government's fitness recommendations. The Wii Fit is a video game that includes exercises.

The researchers taught 20 middle-aged adult volunteers how to use the Wii Fit video games. On the day after they were trained, the adults exercised for 20 minutes with the Wii Fit. Researchers measured the total amount of energy each of the adults in the study used in calories. They found that the average energy used was 116 calories for the 20 minute session.

Based on the results of the study, the researchers concluded the Wii Fit video games could be a helpful form of exercise for middle aged adults. But, for exercise with Wii Fit to meet the government's recommendation, the researchers stated that the length of the exercise session should be increased from 20 minutes to 30 minutes.

Lesson 1.1.1
The Statistical Analysis Process

A Complete the four steps of the statistical investigation process for the study in the table.

Steps in a Statistical Investigation	Wii Fit Video Study
1. Ask a question that can be answered by collecting data.	*Does the Wii Fit video game burn enough calories to be considered suitable exercise?*
2. Decide what to measure and then collect data.	*The total number of calories burned was gathered for a 20-minute session.*
3. Summarize and analyze the data.	*The average energy expenditure was 116 calories. This is below the recommended 150 to 400 calories.*
4. Draw a conclusion and communicate the results.	*The Wii Fit video game does not appear to burn enough calories in a 20-minute session, but a 30-minute session would possibly be enough.*

B Based on the results of this study, what next steps do you recommend to the researchers investigating this issue?

> **Answer:** Researchers could next measure the 30-minute sessions to see whether the calories burned are now in line with the recommendations. They could also try this with a larger and/or more diverse group of subjects [advantages to this are discussed in the next lesson]. Students may also suggest expanding beyond middle-age adults.

Lesson 1.1.1
The Statistical Analysis Process

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. The original version of this work, version 1.0, was created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is a research and development community that seeks to harvest the wisdom of its diverse participants through systematic and disciplined inquiry to improve developmental mathematics instruction. For more information on the Statway™ Networked Improvement Community, please visit **carnegiefoundation.org.**

+++++