

## Lesson 1.1.1

**The Statistical Analysis Process**

---

STUDENT NAME \_\_\_\_\_ DATE \_\_\_\_\_

**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.

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

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?

## Lesson 1.1.1

**The Statistical Analysis Process**

---

**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?

**Signs of the Zodiac 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*.

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.

## Lesson 1.1.1

**The Statistical Analysis Process**

---

**TRY THESE**

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?

**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?

**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?

**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.

**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.

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?

- 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?

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.

- 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.

## Lesson 1.1.1

## The Statistical Analysis Process

## NEXT STEPS

## Using probability to discover what may happen by chance

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.

- 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.  $0.714 = 71.4\%$ .

- 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  $1/3$ ?
- 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?

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.

Lesson 1.1.1

The Statistical Analysis Process

---

- 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.

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    |                     |

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

## Lesson 1.1.1

**The Statistical Analysis Process**

---

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

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?

- 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 \_\_\_\_\_.

- 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.

- 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.

## Lesson 1.1.1

### The Statistical Analysis Process

---

#### Draw a Conclusion and Communicate the Results.

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?

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:

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.

Lesson 1.1.1

The Statistical Analysis Process

NEXT STEPS

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.

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. |                                 |
| 2. Decide what to measure and then collect data.           |                                 |
| 3. Summarize and analyze the data.                         |                                 |
| 4. Draw a conclusion and communicate the results.          |                                 |

Lesson 1.1.1

The Statistical Analysis Process

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  $229/561 \approx 41\%$ . This would be very unlikely if the true percentage was not less than 50%.*

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

**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.

| Steps in Statistical Analysis                              | Study 1 |
|--|---------|
| 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.          |         |

Lesson 1.1.1

The Statistical Analysis Process

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.

| Steps in Statistical Analysis                              | Study 2 |
|--|---------|
| 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.          |         |

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?

## Lesson 1.1.1

**The Statistical Analysis Process**

---

DATE \_\_\_\_\_

STUDENT NAME \_\_\_\_\_

**TAKE IT HOME**

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.)

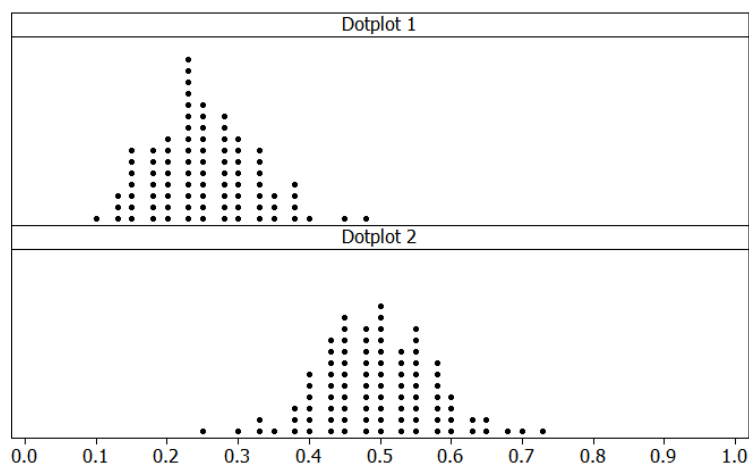
## Lesson 1.1.1

**The Statistical Analysis Process**

- 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?



- 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.)

Lesson 1.1.1

**The Statistical Analysis Process**

---

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.

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. |                     |
| 2. Decide what to measure and then collect data.           |                     |
| 3. Summarize and analyze the data.                         |                     |
| 4. Draw a conclusion and communicate the results.          |                     |

## Lesson 1.1.1

**The Statistical Analysis Process**

---

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

+++++

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](http://carnegiefoundation.org).

+++++

Statway™ is a trademark of the Carnegie Foundation for the Advancement of Teaching. It may be retained on any identical copies of this Work to indicate its origin. If you make any changes in the Work, as permitted under the license [CC BY NC], you must remove the service mark, while retaining the acknowledgment of origin and authorship. Any use of Carnegie's trademarks or service marks other than on identical copies of this Work requires the prior written consent of the Carnegie Foundation.

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/). (CC BY-NC)

