



Lesson 1: Posing Statistical Questions

Student Outcomes

- Students distinguish between statistical questions and those that are not statistical.
- Students formulate a statistical question and explain what data could be collected to answer the question.
- Students distinguish between categorical data and numerical data.

Classwork

Statistics is about using data to answer questions. In this module, the following four steps will summarize your work with data:

Step 1: Pose a question that can be answered by data.

Step 2: Determine a plan to collect the data.

Step 3: Summarize the data with graphs and numerical summaries.

Step 4: Answer the question posed in Step 1 using the data and summaries.

You will be guided through this process as you study these lessons. This first lesson is about the first step – what is a statistical question, and what does it mean that a question can be answered by data?

Example 1 (15 minutes): What is a Statistical Question?

Example 1: What is a Statistical Question?

Jerome, a 6th grader at Roosevelt Middle School, is a huge baseball fan. He loves to collect baseball cards. He has cards of current players and of players from past baseball seasons. With his teacher's permission, Jerome brought his baseball card collection to school. Each card has a picture of a current or past major league baseball player, along with information about the player. When he placed his cards out for the other students to see, they asked Jerome all sorts of questions about his cards. Some asked:

- How many cards does Jerome have altogether?
- What is the typical cost of a card in Jerome's collection?
- Where did Jerome get the cards?

Introduce the situation described in Example 1 to the class. You may want to show the students an example of a baseball card so that they can see the varied amount of information on the card.

Then, consider the questions that follow the description and ask the students:

- Which of these questions do you think might be statistical questions?
- What do you think I mean when I say “a statistical question”?

Students do not have a definition or understanding of what a statistical question is at this point. Allow them to discuss and make conjectures about what that might mean before guiding them to the following:

A statistical question is one that can be answered with data and for which it is anticipated that the data (information) collected to answer the question will vary.

The second and third questions are statistical questions because the answer for each card in the collection could vary. The 1st question, “How many cards do you have in your collection?” is not a statistical question because we do not anticipate any variability in the data collected to answer this question. There is only one data value and no variability.

Convey the main idea that a question is statistical if it can be answered with data that varies. Point out to the students the concept of variability in the data means that not all data values have the same value.

The question, “How old am I?” is not a statistical question because it is not answered by collecting data that vary. The question, “How old are the students in my school?” is a statistical question because when you collect data on the ages of students at the school, the ages will vary – not all students are the same age.

Ask students if the following questions would be answered by collecting data that vary:

- How tall is your 6th grade math teacher?
- What is your hand span (measured from tip of the thumb to the tip of the small finger)?

Ask students which of these data sets would have the most variability.

- Number of minutes students in your class spend getting ready for school.
- Number of pockets on the clothes of students in your class.

After arriving at this understanding as a class, post the informal definition of *statistical question* on the board for students to refer to for the remainder of the class.

Exercises 1–5 (10 minutes)

These question sets are designed to reinforce the definition of a statistical question. The main focus is on whether there is variability in the data that would be used to answer the question. You may want to have students share their answers to Exercise 3 with a partner and have the partner decide whether or not the question is a statistical question.

Exercises 1–5

1. For each of the following, determine whether or not the question is a statistical question. Give a reason for your answer.
 - a. Who is my favorite movie star?
No, not answered by collecting data that vary.
 - b. What are the favorite colors of 6th graders in my school?
Yes, colors will vary.
 - c. How many years have students in my school's band or orchestra played an instrument?
Yes, number of years will vary.
 - d. What is the favorite subject of 6th graders at my school?
Yes, subjects will vary.
 - e. How many brothers and sisters does my best friend have?
No, not answered by collecting data that vary.
2. Explain why each of the following questions is not a statistical question.
 - a. How old am I?
Not answered by data that vary.
 - b. What's my favorite color?
Not answered by data that vary – I just have one favorite color.
 - c. How old is the principal at our school?
The principal has just one age at the time I ask the principal's age. Answered by data that does not vary.
3. Ronnie, a 6th grader, wanted to find out if he lived the farthest from school. Write a statistical question that would help Ronnie find the answer.
What is a typical distance from home to school (in miles) for students at my school?
4. Write a statistical question that can be answered by collecting data from students in your class.
What is the typical number of pets owned by students in my class?
How many hours each day does a typical student in my class play video games?
5. Change the following question to make it a statistical question: "How old is my math teacher?"
What is the typical age of teachers in my school?

Example 2 (10 minutes): Types of Data

To answer statistical questions, we collect data. In the context of baseball cards, we might record the cost of a card for each of 25 baseball cards. This would result in a data set with 25 values. We might also record the age of a card or the team of the player featured on the card.

Example 2: Types of Data

We use two types of data to answer statistical questions: numerical data and categorical data. If we recorded the age of 25 baseball cards, we would have numerical data. Each value in a numerical data set is a number. If we recorded the team of the featured player for 25 baseball cards, you would have categorical data. Although you still have 25 data values, the data values are not numbers. They would be team names, which you can think of as categories.

- What are other examples of categorical data? Eye color, the month in which you were born, and the number that may be used to identify your classroom are examples of categorical data.
- What are other examples of numerical data? Height, number of pets, and minutes to get to school are all examples of numerical data.

To help students distinguish between the two data types, encourage them to think of possible data values. If the possible data values include words or categories, then the variable is categorical.

Suppose that you collected data on the following. What are some of the possible values that you might get?

- Eye color
- Favorite TV show
- Amount of rain that fell during storms
- High temperatures for each of 12 days

Exercises 6–7 (5 minutes)

Have students complete the exercise to reinforce students' understanding of the two types of data.

Exercises 6–7

6. Identify each of the following data sets as categorical (C) or numerical (N).

- Heights of 20 6th graders N
- Favorite flavor of ice cream for each of 10 6th graders C
- Hours of sleep on a school night for 30 6th graders N
- Type of beverage drank at lunch for each of 15 6th graders C
- Eye color for each of 30 6th graders C
- Number of pencils in each desk of 15 6th graders N

7. For each of the following statistical questions, students asked Jerome to identify whether the data are numerical or categorical. Explain your answer, and list four possible data values.

- a. How old are the cards in the collection?

Numerical data, as I anticipate data will be a number.

Possible data values: 2 years, $2\frac{1}{2}$ years, 4 years, 20 years

- b. How much did the cards in the collection cost?

Numerical data, as I anticipate data will be a number.

\$ 0.20, \$ 1.50, \$ 10.00, \$ 35.00

- c. Where did you get the cards?

Categorical, as I anticipate the data represents the name of a place.

(e.g., a store, a garage sale, from my brother, from a friend)

Lesson Summary

A statistical question is one that can be answered by collecting data that vary (i.e., not all of the data values are the same).

There are two types of data: numerical and categorical. In a numerical data set, every value in the set is a number. Categorical data sets can take on non-numerical values, such as names of colors, labels, etc. (e.g., “large,” “medium,” or “small”).

Statistics is about using data to answer questions. In this module, the following 4 steps will summarize your work with data:

Step 1: Pose a question that can be answered by data.

Step 2: Determine a plan to collect the data.

Step 3: Summarize the data with graphs and numerical summaries.

Step 4: Answer the question posed in Step 1 using the data and the summaries.

Exit Ticket (10 minutes)

Name _____

Date _____

Lesson 1: Posing Statistical Questions

Exit Ticket

1. Indicate whether each of the following two questions is a statistical question. Explain why or why not.
 - a. How much does Susan's dog weigh?

 - b. How much do the dogs belonging to students at our school weigh?

2. If you collected data on the weights of dogs, would the data be numerical or categorical? Explain how you know it is numerical or categorical.

Exit Ticket Sample Solutions

1. Indicate whether each of the following two questions is a statistical question. Explain why or why not.
 - a. How much does Susan's dog weigh?
This is not a statistical question. This question is not answered by collecting data that vary.
 - b. How much do the dogs belonging to students at our school weigh?
This is a statistical question. This question would be answered by collecting data on weights of dogs. There is variability in these weights.
2. If you collected data on the weights of dogs, would the data be numerical or categorical?
Numerical

Problem Set Sample Solutions

1. For each of the following, determine whether the question is a statistical question. Give a reason for your answer.
 - a. How many letters are in my last name?
No, this question is not answered by collecting data that vary.
 - b. How many letters are in the last names of the students in my 6th grade class?
Yes, there is variability in the lengths of the last names.
 - c. What are the colors of the shoes worn by the students in my school?
Yes, we expect variability in the colors.
 - d. What is the maximum number of feet that roller coasters drop during a ride?
Yes, we expect variability in the feet to drop for different roller coasters; they are not all the same.
 - e. What are the heart rates of the students in a 6th grade class?
Yes, we expect variability – not all 6th graders have exactly the same heart rate.
 - f. How many hours of sleep per night do 6th graders usually get when they have school the next day?
Yes, we do not expect that all 6th graders sleep the same number of hours.
 - g. How many miles per gallon do compact cars get?
Yes, we expect variability in the miles per gallon – not all compact cars get the same miles per gallon.

2. Identify each of the following data sets as categorical (C) or numerical (N). Explain your answer.
- a. Arm spans of 12 6th graders
N; the arm span can be measured as number of inches for example, so the data is numerical.
 - b. Number of languages spoken by each of 20 adults
N; number of languages is clearly numerical.
 - c. Favorite sport of each person in a group of 20 adults
C; a sport falls into a category, such as "soccer" or "hockey" and cannot be measured numerically.
 - d. Number of pets for each of 40 3rd graders
N; number of pets is clearly numerical.
 - e. Number of hours a week spent reading a book for a group of middle school students
N; number of hours is clearly numerical.
3. Rewrite each of the following questions as a statistical question.
- Answers will vary*
- a. How many pets does your teacher have?
How many pets do students in our school have?
 - b. How many points did the high school soccer team score in its last game?
How many points did the high school soccer team score in soccer games this season?
 - c. How many pages are in our math book?
How many pages are in the books in the school library?
 - d. Can I do a handstand?
Can most 6th graders do a handstand?
4. Write a statistical question that would be answered by collecting data from the 6th graders in your classroom.
Answers will vary. Check if the question would be answered by collecting data that vary.
5. Are the data you would collect to answer that question categorical or numerical? Explain your answer.
Answers will vary.