

# Investigation 9

## Survey Says?

### *Analyzing Categorical Data in a Statistical Study*



## Overview

This investigation is the first of several lessons that focus on the analysis of two categorical variables. This investigation shares an initiative carried out by students at an urban high school that involved collecting data from the student body. This initiative involved creating a survey, obtaining a sample of completed surveys, and analyzing the sample to answer statistical questions posed by the students. The students addressed whether the sample was representative of the student body of their school.

The four components of statistical problem solving as put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report* are addressed in this investigation. The four components are formulate a statistical question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This investigation is a GAISE Level B activity.

## Instructional Plan

### Brief Overview

- » Define and give examples of categorical data.
- » Read the scenario and study the survey questions.
- » Develop a statistical question based on the survey questions.

- » Discuss the four options to collect survey data.
- » Discuss the collection plan used by high-school students.
- » Summarize survey results.

### Introduction to Categorical Data

Ask your students to give examples of *numerical* data they have analyzed. Have them share what types of graphs and calculations they did in their analysis of numerical data.

*Possible answer: Examples from previous investigations include height and arm span, length of baseball games in hours, time to complete a memory game, and homework times. Students would have constructed box plots, dot plots, and histograms and found means, medians, IQRs, and standard deviations.*

Discuss another type of data called *categorical data*. Categorical variables take on values that are names or labels. Share some examples of categorical data such as an answer to a true or false question, an answer to a multiple-choice question (A, B, C, or D), the size of a T-shirt (small, medium, or large), or the breed of a dog (shepherd, terrier, lab).

Ask your students to share other examples of categorical data.

Hand out Student Worksheet 9.1 Scenario

Explain that the case study presented was conducted by high-school students from an urban high school who analyzed categorical data.

Direct students to read the scenario from Student Worksheet 9.1.

### Scenario

The administration at Rufus King High School, a United States urban high school of students in grades 9 to 12, was in the process of evaluating the school's academic and extracurricular programs. The high-school administration considered distributing and analyzing a survey addressing the school's programs that would be similar to the process businesses use to evaluate

their products and services. They asked the students enrolled in an 11th grade mathematics class if they would help with the design, distribution, and analysis of a survey project.

Statistical studies about a school's services might result in decisions that alter a school's daily schedule, curriculum, course offerings, extracurricular opportunities, etc. Rufus King students wanted to be part of a study that might alter their school's academic and extracurricular programs. Students designed

### Learning Goals

- » Investigate methods for obtaining a representative sample of responses to a survey from a large population.
- » Evaluate a random sample of students' responses to a survey.
- » Summarize a population using the results of a random sample.

### Mathematical Practices Through a Statistical Lens

*MP2. Reason abstractly and quantitatively.*

Statistically proficient students are able to summarize data to answer statistical questions. Students explain their summaries of data using proportions.

### Materials

Student worksheets are available at [www.statisticsteacher.org/statistics-teacher-publications/focus](http://www.statisticsteacher.org/statistics-teacher-publications/focus)

- » Student Worksheet 9.1. Scenario
- » Student Worksheet 9.2 Data Collection Methods
- » Student Worksheet 9.3 Survey Results
- » Exit Ticket

### Estimated Time

One 50-minute class period. This lesson introduces a case study that is expanded in investigations 10 and 11.

### Pre-Knowledge

Students should be able to convert a proportion to a percent.

**Question 1:** Indicate your gender:  
 Female (F)     Male (M)     Prefer not want to respond

**Question 2:** Indicate your grade level in high school:  
 9<sup>th</sup> grade     10<sup>th</sup> grade     11<sup>th</sup> grade     12<sup>th</sup> grade

**Question 3:** Do you consider yourself a dog person, a cat person, or neither?  
 A. I consider myself a dog person.  
 B. I consider myself a cat person.  
 C. I do not consider myself a dog or cat person.

**Question 4:** What is your main goal after completing high school?  
 A. To attend a college, university, or technical school.  
 B. To get a job.  
 C. To enlist in the military.  
 D. Other

**Question 5:** Do you participate in one or more of the athletic programs at your school (basketball, football, soccer, hockey, tennis, volleyball, etc.)  
 Yes (Y)     No (N)

**Question 6:** Do you exercise daily?  
 Yes (Y)  No (N)

**Question 7:** Do you spend at least 1 hour a week involved in an outdoor activity (walking, running, playing a game etc.)?  
 Yes (Y)  No (N)

**Question 8:** Are you involved in any community service activity?  
 Yes (Y)     No (N)

Figure 9.1: Survey questions developed by students at Rufus King High School

a survey they thought would address several important statistical questions related to the school's academic and extracurricular programs. A few of the survey questions are listed in Figure 9.1.

After students read the scenario, discuss the following questions:

1. Why do you think the students designing this survey wanted to know the grade level of a student completing the survey (Survey Question 2)?

*Possible answer:* Several of the other questions might be answered differently based on a student's grade level. For example, is it possible 9th graders might be more or less involved in extracurricular activities than 12th graders? Does

*a student's plan after completing high school change over time? Might 12th graders answer Question 4 differently than 9th graders?*

2. Why do you think students included Survey Question 3? What was a possible reason to consider this question important?

*Possible answer:* Survey Question 3 might be used to examine the growing interest in therapy pets to address stress or anxiety. If a program of this type were pursued, what type of pets would be selected? Do most students have a similar interest in the pets selected?

3. Why might it be important to know if students exercise daily? Will most students understand what this question is asking? Will most students answer this question?

*Possible answer: Exercise may have different meanings to students. Some students may think of this as organized school activity. Other students may think of this as an individual activity involving walking, running, stretching, etc. The question was considered adequate by the Rufus King students, but whether they collected accurate responses was not clear. This question was unclear to some students answering the question and is a good example of how questions of this type have different interpretations. After the survey was distributed, students discussed that the following rewording of this question might have clarified some of the confusion: “Do you participate in at least 10 minutes a day of physical exercise either alone or as part of a group?”*

4. Why might it be important to know if students are involved in community service? Do students agree on what is meant by community service? Will most students answer this question?

*Possible answer: This question also needed more clarity. Community service was viewed by some students as service activities by the school; other students interpreted community service as an activity organized by other organizations or groups. Here again is an opportunity to discuss the importance of whether or not a question provided the intended information needed in the statistical study.*

Take the students through the process the Rufus King students followed.

### Formulate a Statistical Question

The Rufus King students developed a series of statistical questions designed to provide a summary of the school’s student population. Several possible statistical questions emerged from this project. For example, are students typically going to attend a college or university after high school or pursue other options? Are students likely to participate in the school’s

athletic programs? Do students typically spend at least one hour per week outdoors?

### Discussion About Different Ways to Collect Appropriate Data

Explain that after the survey was designed and approved by the administration, a plan was needed to organize how students would complete the survey. It was possible, but not practical, to analyze completed surveys from more than 1,200 students enrolled in the high school. Students (under the direction of their teachers) discussed ways in which they might distribute surveys to obtain a sample that provided all students in the school the same opportunity to complete the survey.

Ask the students to read the four data collection options and answer the two questions for each option.

For each of the following four options, answer the two questions:

- » Do you think this option will provide an accurate summary of the responses from students in the school?
- » If this option is used, are there any groups of students who may not be represented? Explain your answer.

#### Option 1:

Consider placing computers at various locations around school (e.g., the cafeteria, library, computer lab) that are monitored by students from the mathematics class involved with this project. Students in the vicinity of the computers would be asked to complete the survey provided on the computer. After a student completed the survey, the students monitoring the computer would save the results and load a new survey for the next student to complete. At the end of the day, the

responses from the completed surveys would represent the representative sample for analyzing the questions.

**Option 2:**

There are 35 students in the mathematics class involved with this project. Each member of the class would be encouraged to anonymously complete the survey. The completed surveys would comprise the representative sample for analyzing the questions.

**Option 3:**

Students in the mathematics class involved with this project would post the survey online using a service provided by a private company. Each member of the class would encourage friends to complete the survey, both through word of mouth and also through their social media accounts. The online service would provide completed surveys that comprise the representative sample for analyzing the questions.

**Option 4:**

Students enrolled in the mathematics class would distribute surveys both before or after school at various locations in the school building. At the end of the day, the completed surveys would comprise the representative sample for analyzing the questions.

After the students have read the four options and answered the two questions for each, have the students share their answers for each option.

*Option:* Place students in groups. Have each group create a poster that lists the pros and cons of each method. Also, list their choice for a method.

**Discussion Points**

Discuss with students that each of the options would provide a sample, but there would be questions as to whether a representative sample

of the student population would have completed the surveys. In general, these options result in a *convenience sample*, or a sample that did not provide an opportunity for a cross section of the school's students to complete the survey. It is important to indicate that a statistical study based on a convenience sample, or any sample not representative of the population, may result in *bias* that raises questions regarding any conclusions of the study.

*Note:* Consider discussing with your students how they would collect a representative sample at their school. Would any of the previous options provide a representative sample? What other options might be considered?

**Design and Implementation of a Plan to Collect Data**

Hand out Student Worksheet 9.2 Data Collection Methods

Have the students read the Plan to Collect Data section on Student Worksheet 9.2.

**Plan to Collect Data**

The following is a summary of the plan implemented at Rufus King High School.

All students attending Rufus King are required to take an English course. Students involved in the survey project arranged providing the option of completing a survey during an English class with the school's English teachers. They estimated it would take fewer than five minutes to complete the survey. A specific day was identified to complete the survey. Students were also told by their English teachers that they did not have to complete the survey. Students involved in organizing this project provided an explanation of the project to the students several days before the survey was distributed by way of an all-school announcement. In addition, a flier was sent home to inform parents and guardians about the project.



The number of students who completed the survey was 1103.

Each survey was collected and given a specific identification number. Identification numbers from 1 to 1103 were assigned to the completed surveys. It was decided that 50 randomly selected surveys would form the sample for this study. Students generated 50 random numbers from 1 to 1103 using a graphing calculator. The 50 numbers generated by the calculator represented the 50 identification numbers and the 50 surveys selected to form the sample.

Ask your students to answer questions 1 to 5.

1. Do you think the above plan resulted in a sample that provided all students an equal chance to be selected in the sample? Explain your answer.

*Possible answer: Essentially, every student who was in attendance had an opportunity to complete the survey. This plan would result in a sample in which each completed survey had an equal chance of selection.*

2. Why do you think it was important to inform students about the project before they received the survey?

*Possible answer: It is important to emphasize that data of this type must convince students their time to complete the survey and their responses are important. If this study had been a research study conducted by professionals, a careful review of the survey questions would need to be conducted by a team of advisers. This team would also be responsible for evaluating details about the research project and how it would be communicated to students and their parents or guardians. Authentic statistical studies are held to high standards of communication and review.*

3. Why do you think it was important to inform parents and guardians about the project?

*Possible answer: Emphasize again the importance of maintaining communication in a statistical study. Also, students in high school are considered minors. Involving parents and guardians was an important requirement of the administration.*

4. Using the plan described, which students would not have completed the survey?

*Possible answer: Students absent from school or students who opted out of completing the survey would not have been included.*

5. Do you think the sample of 50 completed surveys represents a representative sample of all students?

*Possible answers: Encourage students to express their opinions to this question. It is anticipated students may comment that a sample of only 50 students would likely not be representative of the school's population. Students may also indicate a sample of only 50 surveys would not be large enough to obtain adequate summaries of the survey questions.*

### Analyze the Data

Ask your students what type of data was collected for each of the eight survey questions.

*Answer: Categorical data.*

Ask students how they would summarize the responses to Question 1. What measure would they use to communicate what they have collected?

*Possible discussion points: Students may initially focus on the counts of Male responses or Female responses. Although the count of each category is important, it is the proportion of the*

*number of males or the number of females to the sample size that will be the more important summary of the question in this statistical study. A similar proportion would be calculated for each of the other questions in the survey.*

Return to the discussion concerning whether the sample of 50 completed surveys represents a representative sample of all students. The main question is whether the selection of 50 surveys is large enough to estimate the proportions of the school population for each question. Will the proportion of females, proportion of students who exercise daily, or proportion of students who are involved in community service based on this sample of 50 students be the same as the school population?

The following example will help students understand that the random sample selected by students is likely to provide an adequate summary of the school population.

The sample of 50 students indicated 33 females and 17 males. A first step was to convert the number of females to a proportion— $33/50$  or 0.66 or 66% of the sample of 50 students was female and  $17/50$  or 0.34 or 34% of the students was male.

Share with your students the following summary of the school population posted on the school's website:

- » Total enrollment (September 5): 1204 students
- » Total number of females: 775
- » Total number of males: 429

Based on this information summarizing the school population, the proportion of females at Rufus King High School at the time this project was conducted was  $775/1204$ , or approximately 0.644 or 64%. The sample pro-

portion of 66% was similar to the proportion of females of the school's population.

### **Interpret the Results in the Context of the Original Statistical Questions**

Hand out Student Worksheet 9.3 Survey Results.

Direct students individually or in small groups to use the data presented on Student Worksheet 9.3 and answer Question 6.

*Answers:*

#### **Q1 (Survey Question 1)**

- » Proportion of females:  $33/50$  or 0.66
- » Proportion of males:  $17/50$  or 0.34

#### **Q2 (Survey Question 2)**

- » Proportion of students in 9th grade:  $15/50$  or 0.30
- » Proportion of students in 10th grade:  $14/50$  or 0.28
- » Proportion of students in 11th grade:  $16/50$  or 0.32
- » Proportion of students in 12th grade:  $5/50$  or 0.10

#### **Q3 (Survey Question 3)**

- » Proportion of students who indicate they are a “dog person”:  $23/50$  or 0.46
- » Proportion of students who indicate they are a “cat person”:  $24/50$  or 0.48
- » Proportion of students who indicate they are neither:  $3/50$  or 0.06

#### **Q4 (Survey Question 4)**

- » Proportion of students who plan to attend college after high school:  $30/50$  or 0.60

- » Proportion of students who plan to get a job after high school:  $9/50$  or  $0.18$
- » Proportion of students who plan to enlist in the military after high school:  $6/50$  or  $0.12$
- » Proportion of students who selected other:  $5/50$  or  $0.10$

#### Q5 (Survey Question 5)

- » Proportion of students who participate in the school's athletic program:  $30/50$  or  $0.60$

#### Q6 (Survey Question 6)

- » Proportion of students who exercise daily:  $30/50$  or  $0.60$

#### Q7 (Survey Question 7)

- » Proportion of students who spend at least one hour per week outdoors:  $10/50$  or  $0.20$

#### Q8 (Survey Question 8)

- » Proportion of students involved in community service:  $23/50$  or  $0.46$

Direct students to individually answer questions 7 to 12. After they have completed the questions, discuss these questions with the whole class.

7. Based on the above summaries, provide a brief description of the students attending this high school.

**Summary answer:** Students can focus on one or two summaries they find interesting about the school. For example, slightly more students considered themselves a cat person or only 20% of the students spend at least one hour outdoors. Remind students they are us-

ing a representative sample to describe the students in the school population.

8. What is your estimate of the *number of* students who participate in an athletic program from the total enrollment of 1204 students? Do you think your estimate is the exact number of students who participate in an athletic program?

**Answer:** Assume the proportion of the school population participating in an athletic program is the same as the proportion of the sample. Therefore, an estimate of  $0.60$ , or  $60\%$ , of the 1204 students is 722 students. This estimate is likely not the exact number of students who participate in an athletic program.

9. Why might it be important to know the number of students and the proportion of students who participate in a school athletic program?

**Possible answer:** Results could be used to evaluate the interest in an athletic program and whether the school's facility can effectively address the interest. Estimating the number of students involved in an athletic program might be used to determine whether the facilities (e.g., gyms or volleyball courts or bathrooms) are sufficient.

10. What is your estimate of the students who participate in community service? Do you think your estimate is the exact number of students who participate in community service?

**Answer:** Assume the proportion of the school population participating in community service is the same as the proportion of the sample. Therefore, an estimate of  $0.46$ , or  $46\%$ , of the 1204 students is approximately 553.84 or 554 students. This estimate is likely not the exact number of students who participate in community service.



11. Why might it be important to know the number and proportion of students who participate in community service?
12. Why might it be important to know the proportion of students who spend at least one hour involved in outdoor activities?

*Possible answer:* If the school is considering improving students' participation in community service, it is important to determine the current involvement.

*Possible answer:* Several research studies have linked outdoor activity to student achievement. This survey question might be connected to other items (e.g., exercise, gender, grade level) that examine whether there are noticeable differences in outdoor activity based on these other categories.



## Exit Ticket

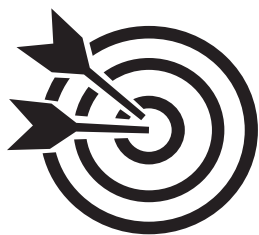
---

1. Based on the estimate of the relative frequencies from the surveys and the summary of the high-school enrollment of 775 females and 361 ninth graders, determine an estimate for each of the following two questions. For each question, indicate how you determined your estimate.
- a. How many females do you think are involved in an athletic program at King?

*Summary answer:* Assume the proportion females participating in an athletic program is the same as the proportion of students participating in an athletic program based on the sample. Therefore, 60% of the 775 female students, or 465 female students, is an estimate of the number of females who participate in an athletic program.

- b. How many 9th graders do you think are involved in at least 1 hour of outdoor activities per week?

*Answer:* Assume the proportion of 9th graders involved in at least one hour of outdoor activities is the same as the proportion of students involved in at least one hour of outdoor activities from the sample. Therefore, assume 20% of the 361 9th grade students, or approximately 72.2 or 73 students, is an estimate. This also assumes the proportion of 9th grade students involved in outdoor activities is the same as the proportion of other grade levels.



## Exit Ticket Cont.

---

2. Consider the following data collection option:

Students in the mathematics class involved with the project would number each table in the cafeteria. They would select 10 random tables at each lunch period and ask everyone sitting at the selected table to answer the survey.

Do you think this option will provide an accurate summary of the responses from students in the school? If this option is used, are there any groups of students who may not be represented? Explain your answer.

*Possible answer: Since students usually sit with their friends at the same lunch table, it is likely students at the table would answer the survey questions in a similar manner. Students who do not eat in the cafeteria or go out or home for lunch are not represented.*

### Further Explorations and Extension

---

Interest in completing a survey project at your school may also be considered a viable extension of this investigation. If the entire project could not be completed (due to time constraints or other challenges), designing a plan to carry out a project at your school that is similar to the one in this investigation might also be a valuable discussion and an exploration to consider.