Section 1

Getting Started

Introduction

The objective of the *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*) *Report* is to provide a conceptual framework for K–12 statistics education. The GAISE Framework outlines three statistical maturity levels—A, B, and C—that are based on experience, not on age or grade level. The framework stresses hands-on active learning and that statistical analysis is an investigative process that turns loosely formed ideas into scientific studies by doing the following:

1. Formulating a question that can be answered with data
2. Designing a plan to collect appropriate data
3. Analyzing the collected data by graphical and numerical methods
4. Interpreting the results to reflect light on the original question

The four-step statistical problem-solving process is also part of the Common Core State Standards, as found in the Grade 6 Statistics and Probability Content Standards, 6.SP.

Develop understanding of statistical variability.
6.SP.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *Ex. "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question, because one anticipates variability in students' ages.*

6.SP.2 Understand that a set of data collected to answer a statistical question has a distribution, which can be described by its center, spread, and overall shape.

6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Summarize and describe distributions.
6.SP.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

6.SP.5 Summarize numerical data sets in relation to their context, such as by the following:

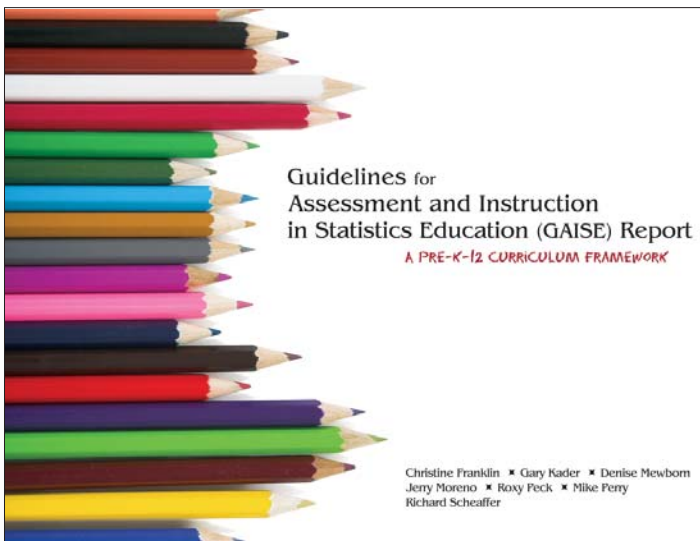
a. Reporting the number of observations

b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement

c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and striking deviations from the overall pattern with reference to the context in which the data were gathered

d. Relating the choice of measures of center and variability to the shape of the data distribution and context in which the data were gathered.

The purpose of the following investigation is to help your students learn how to formulate a statistical question—a question that can be answered with data.



Investigation 1.1
Formulating a Statistical Question

Overview

The GAISE report emphasizes the *process* of doing a statistical study. The first step in that process requires the investigator to formulate a question that will be the focus of the study. This investigation provides a framework for teachers to use to help their students construct questions that can be addressed through the collection and analysis of data. These types of questions are called **statistical questions**. In posing the questions, students will be encouraged to think about the **population** (subjects) to be studied, the **variable** (characteristic) to be measured, and the **variation** that may occur in the **measurement** of that characteristic.

As stated in the Common Core State Standards, a statistical question is one that "anticipates variability in the data and accounts for this variability in the analysis." The objective of this investigation is to assist students in generating statistical questions about their schools, neighborhoods, and interesting phenomena in the world and describing the type of data that would need to be collected to answer those questions.

GAISE Components

This investigation stresses the first component (formulate a question) of statistical problem solving put forth in the GAISE report. It can be used with GAISE Level A and Level B students.

Learning Goals

Students will be able to do the following after finishing this investigation:

- Distinguish statistical questions from nonstatistical questions
- Identify the population (subjects) to be studied
- Identify the data (values of a variable) to be collected
- Develop an intuitive understanding of the expected variation in the data

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
4. Model with mathematics.

Common Core State Standards Grade Level Content

6.SP1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

A well-written statistical question refers to a **population** of interest, a **measurement** of interest, and **anticipates answers that vary**.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 6–8 Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.

Materials

Statistical Questions Worksheet Level A and Level B

Estimated Time

One day

Instructional Plan

1. Begin the investigation by asking your students what they would be interested in finding out about their school, neighborhood, or families and friends. Many questions students generate will be interesting, but may not be statistical questions. Discuss with your students that a statistical question is one that can be answered with data and that **variability** in the data is expected. A well-written statistical question refers to a **population** of interest, a **measurement** of interest, and **anticipates answers that vary**.

   While the question, "How old is my math teacher?" might be of interest, it is not a statistical question because there is only a single subject, and hence no variability. "How old are the teachers in our school?" is a statistical question because "teachers in our school" is the population, "age" is the measurement variable, and we expect several ages.

Help your students understand that nonstatistical questions may be too broad or specific. The question, "Do people like pizza?" is too broad. It is unclear exactly what the population is. A better version would be, "Of all the 4th-graders in our school, who likes pepperoni pizza?" The population is "the 4th-graders in our school," the measurement is "like or don't like pepperoni pizza," and we would expect some people to like this type of pizza and some not to like it. The question, "How many words are in this sentence?" is narrow and has no variability in its answer. However, "How many words are in the sentences in this book?" is a statistical question. The population is "all the sentences in this book," the measurement is the "length of the sentences," and we would expect the sentences to be different lengths.

2. Pose the following question to your students and have them decide whether it is a statistical question. "How old is my pet dog?" This is not a statistical question because there is no variability—there is a single subject or unit, and hence no variability. Discuss with your students how this question could be rewritten into a statistical question about the class. One suggestion: "How old are the pets of the students in our class?" The population of interest is the "students' pets," the measurement is the "pets' ages," and we would expect the pets would be different ages.

3. Pose the question, "What is my favorite topping on a pizza?" Ask your students why this is not a statistical question, and ask them to rewrite it into a statistical question. Possible answer: "What do the students in this class prefer as their favorite topping on a pizza?" The population is the "students in class," the measurement is their "favorite pizza topping," and we would expect different answers such as cheese, sausage, or pepperoni.

4. Place your students into groups of four. Give each group the appropriate level A or B list of questions in Table 1.1.1 or Table 1.1.2. For each question, the groups should indicate whether the question is a statistical question and give reasons for their answer. If they answer that the question is a statistical one, they should specify the population, measurement taken, and expected variation. If it is not a statistical question, they should explain why it is not a statistical question and rewrite it so it is a statistical question.

Table 1.1.1: Level A Questions

Question	Statistical Question (Y or N)	Explain Your Answer	Question	Statistical Question (Y or N)	Explain Your Answer
What colors are the shoes worn by the teachers in our school?			How many languages does my friend speak?		
What are the shapes of all the buttons on the clothes worn by the students in this class?			How far can I jump?		
How many times does the word "bridge" appear in the rhyme "London Bridge Is Falling Down"?			Does my best friend like McDonald's Happy Meals?		
How many pockets do I have?			Is my last name the longest name in class?		
What is my fifth-grade sister's favorite animal at the zoo?			What is the favorite lunch of third-graders in our school?		

Table 1.1.2: Level B Questions

Question	Statistical Question (Y or No)	Explain Your Answer	Question	Statistical Question (Y or No)	Explain Your Answer
Can I roll my tongue?			Who was the oldest U.S. president when inaugurated?		
How do the lengths of the first names of students in class compare to the lengths of their last names?			Are students in our class who are 4′5″ or taller able to jump higher than students who are shorter than 4′5″?		
Am I going to win a prize at the school carnival?			Which brand of pizza has the most pepperoni?		
What is the longest-lasting brand of AA batteries?			Do plants grow better under colored lights?		
A teacher asks her class, "What is your shoe size?"			Is it easier to remember a set of objects or a list of words?		
Which brand of bubble gum holds its flavor the longest?					

5. Discuss with your students their answers to each question. Following are suggested answers.

Table 1.1.3: Suggested Answers to Level A Questions

Question	Statistical Question (Y or N)	Explain Your Answer
What colors are the shoes worn by the teachers in our school?	Y	Population is teachers in school; measurement is shoe color, data are various colors
What are the shapes of all the buttons on the clothes worn by the students in this class?	Y	Population is all the buttons worn by students in class; measurement is button shape; data are various shapes
How many times does the word "bridge" appear in the rhyme "London Bridge Is Falling Down"? What is the frequency of the words that appear in the nursery rhyme "London Bridge is Falling Down"?	N	There is one word in the population, hence no variability—only single frequency for an answer. Population is all words in the rhyme; measurement is number of times each appears; data are words with their frequency of occurrence
How many pockets do I have? How many pockets do the students in class have on the clothes they are wearing today?	N	There is one person in the population, hence no variability in number of pockets I have. Population is all students in class; measurement is number of pockets; data are 0, 1, 2, …
How many languages does my friend speak? How many languages do the students in my school speak?	N	There is one person in the population, hence no variability—only one number for an answer. Population is all students at my school; measurement is number of languages each student speaks; data are 1, 2, 3, …
How far can I jump? How far can the students in this class jump?	N	There is one person in the population, hence no variability—only one distance. Population is all students in class; measurement is distance one can jump; data are real numbers
Does my best friend like McDonald's Happy Meals? Of the fifth-graders in our school, who likes McDonald's Happy Meals?	N	There is one person in the population—my friend—so there is no variability only one answer—yes or no. Population is all fifth-graders in our school; measurement is yes/no liking of Happy Meals; data are a listing of students with yes or no response for each
What is my sister's favorite animal at the zoo? Which animal in the local zoo would the fifth-graders pick as their favorite?	N	There is one person in the population, hence no variability—only one animal name for an answer. Population is fifth-graders; measurement is favorite zoo animal; data are various animals
Is my last name the longest name in class? How long are the last names of students in this class?	N	There is one person in the population, hence there is no variability—only one answer. Population is class students; measurement is length of last name; data are various last names
What is the favorite lunch of third-graders in our school?	Y	Population is third-graders; measurement is the name of favorite lunch; variability—we would expect students to give answers such as pizza, sandwiches, or macaroni and cheese

Table 1.1.4: Suggested Answers to Level B Questions

Question	Statistical Question (Y or N)	Explain Your Answer
Can I roll my tongue? How do boys and girls compare regarding the ability to roll their tongues?	N	There is one person in the population, hence there is no variability—only one answer. Population is all boys and girls in class; measurement is whether a student can roll his/her tongue; data are yes/no for each student
How do the lengths of the first names of students in class compare to the lengths of their last names?	Y	Population is students in class; measurement is difference in length of first and last names; we would expect the differences to vary from student to student
If everyone in class plays "spin-the-wheel" at the school carnival, what are their chances of winning?	Y	Population is the students in class; measurement is win or not win; we would expect some students to win and others to not
What is the longest-lasting brand of AA batteries?	Y	Population is different brands of batteries; measurement is battery length of life; we would expect the brands to last varying lengths of time
A teacher asks her class, "What is your shoe size?"	Y	Population is implied to be all students in class; measurement is shoe size; we would expect students to have different shoe sizes
Which brand of bubble gum holds its flavor the longest?	Y	Population is brands of bubble gum; measurement is how long flavor lasts; we would expect different brands would vary on how long their flavor lasts
Who was the oldest U.S. president when inaugurated? How old were the U.S. presidents when they were inaugurated?	N	There is only one person in the population, the name of the oldest president when inaugurated, hence no variability. Population is all U.S. presidents; measurement is age when inaugurated; data are various ages
Are students in our class who are 4′5″ or taller able to jump higher than students who are shorter than 4′5″?	Y	Population is "tall" students and "short" students in our class; measurement is how high one can jump; we would expect students to jump different heights
Which brand of pizza has the most pepperoni?	Y	Population is different brands of pizza; measurement is the count of number of pieces of pepperoni on a pizza; we would expect different brands to have different amounts of pepperoni
Do plants grow better under colored lights? Do tomato plants grow taller under red light, blue light, or daylight?	N	More specific population should be listed. Define what it means to grow better, and need to specify the colors of light. Population is tomato plants; measurement is height under red light, blue light, or daylight; we expect heights to vary
Is it easier to remember a set of objects or a list of words? Are the seventh-graders able to memorize the names of a set of objects better than a list of words?	N	There is no population mentioned. Population is all seventh-graders; measurement is number of objects recalled and number of words recalled; we expect the number of correct responses to vary

Assessment with Answers

Level A

A third-grader's favorite sport was soccer. She asked all the students in her room, "Who likes to watch a soccer game?" Explain why this is a statistical question. The population is "all the students in her room." The measurement is "whether a student likes to watch soccer." Variation is expected with some students answering "yes" and some answering "no."

Level B

A group of seventh-grade students asked the question, "What's the fastest animal in the world?"

1. Explain why this is not a statistical question. There is no variability—there is just one fastest animal.

2. Rewrite the question so it is a statistical question. How many miles per hour can various animals run?

Extension

Ask your students to choose one of the questions they have decided is a statistical question or one they have rewritten into a statistical question. Have students discuss how they might collect data to help answer the question and describe the variability in the data they might expect.

Note: Some of the questions in tables 1.1.1 and 1.1.2 will be addressed in the investigations in this book.

References

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics, *www.corestandards.org*.