Middle School Mathematics
A Guide to the Connected Mathematics™ Series

Samples and Populations

Prepared by members of
the Readington Middle
School math department
2004 Issue 1

Table of Contents

1 Introduction

Samples and Populations is part of the **Data Analysis and Probability** strand. This unit offers a review of statistics concepts introduced in earlier units. Students explore what samples are and how they are related to populations, ways to select samples, and use random samples.

Statistics is the science that relies on data to answer questions. A statistical investigation typically encompasses four interrelated components:

- Posting the question- Formulating key questions to explore and identify what data to collect in order to address the questions.
- Collecting the data – Deciding how to collect the data as well as collecting the data.
- Analyzing the data – Organizing, representing, summarizing, and describing the data and looking for patterns in the variation of the data.
- Interpreting the results – Predicting, comparing, and identifying relationships and using the results from the analyses to make decisions about the original questions.

Recognition and use of the process of statistical investigation is important in working with statistics.

2 Goals/Objectives

This unit will help students:

- Employ the process of statistical investigation to explore problems
- Analyze data using tables, stem-and-leaf plots, histograms, and box-and-whisker plots
- Compare data using measures of center (mean, median), measures of spread (range, percentiles), and data displays (stem-and-leaf plots, histograms, box-and-whisker plots)
- Explore relationships among data using scatter plots
- Distinguish between samples and populations, compare samples, and use information drawn from samples to make conclusions about populations
- Apply selected concepts from probability to understand the concept of randomness and to select random samples
- Explore concepts of “representativeness” and sample size as they relate to using random and nonrandom samples to draw conclusions about the characteristics of populations
- Design a survey, focusing on how questions are asked

3 Vocabulary

The following words and concepts are used in this unit. The concepts in the left column are those essential for student understanding in this and future units. The Descriptive Glossary (page 72) in the student text gives definitions for many of these words.

Essential Terms	Terms developed in previous units	Nonessential terms
biased sample	Data	Database
box-and-whiskers plot (box plot)	Equally likely	Quartile
convenience sample	Mean	Simulate
distribution	Median	
five-number summary	Probability	
histogram	Random	
population	Range	
random sample	Statistics	
representative sample	Survey	
sample		
scatter plot		
stem-and-leaf plot (stem plot)		
systematic sample		
voluntary-response sample		

4 Summary of Investigations

4.1 Investigation 1 – Comparing Data Sets (pp 5-23)

- Analyze data from a study on the price, quality and sodium content of a variety of peanut butter brands, which are classified by four attributes: natural or regular, creamy or chunky, salted or unsalted, and name brand or store brand.
- Review the use of the stem-and-left plots and measures of center and the box-and-whiskers plots as tools for comparing data.
- Explore relationships among data using scatter plots.

4.2 Investigation 2 – Conducting Surveys (pp 24-36)

- Consider the distinction between samples and populations and use results from analyses of data from samples to make estimates about population characteristics or behaviors.

- Consider the implications of making estimates about the entire US population based on a telephone survey.
- Consider a variety of sampling strategies, analyzing their advantages and disadvantages of each and determining which would produce the sample most representative of the population.
- Consider issues that are important in developing surveys as they design a survey and write survey questions.

4.3 Investigation 3 – Random Samples (pp 37-48)

- Consider the differences among convenience samples, voluntary-response samples, and random samples. Explore techniques for choosing samples randomly from a population – such as spinners, number cubes, and random-number generators on calculators.
- Investigate the idea that sample size affects the accuracy of population estimates.
- Through sampling and determining statistics, learn that by taking larger samples one can often reduce the variability of in-sample distributions and make more accurate predictions.

4.4 Investigation 4 – Solving Real-World Problems (pp 49-62)

- Analyze measurements of Native American arrowheads found in six different archaeological sites.
- Explore how the data from the known sites may be used to make predictions about the newer sites.
- Employ a sampling procedure to investigate how many chocolate chips must be added to a batch of cookie dough to ensure that each cookie in the batch will contain at least five chips.

5 Sample Problems and Solutions

This section provides solutions for selected ACE questions for each investigation.

5.1 Investigation 1

ACE Question 5 page 18

ANSWER

5a. You could use graph A, which shows the highest price of about $4.00 in Tokyo and the lowest price of about $0.90 in Caracas.

5b. You could use graph A to find the mean (about $1.80) or the median (about $1.70). Or you could use graph B, which shows the typical price is between $1.50 and $2.00.

5c. Possible answer: Both graphs could be titled “Prices of 2-liter Bottles of Soft Drink.” On graph A, the horizontal axis could be labeled “City” and the vertical

axis could be labeled “Price.” On graph B, the horizontal axis could be labeled “Price,” and the vertical axis could be labeled “Number of cities” or “Frequency.”
5d. Yes; from graph A, you could count how many cities have a price in each range and create graph B.
5e. No; the specific data values cannot be identified from graph B; we know only the *number* of prices in specific intervals.

5.2 Investigation 2

ACE question 1 page 31.

ANSWER
1. Systematic selection from class lists would probably give a representative sample, provided the lists include the names of all students in the school.

5.3 Investigation 3

ACE Question 7 page 59

ANSWER
7. The goal of the study was to determine the distance from a television set at which a typical person might use a remote control. It appears that the population being studied was owners of television sets; the sampling procedure is unclear. The report indicates only an average viewer distance, but it seems important to consider variability as well. The remote seems to have been designed for a fairly operational range. A box plot or stem plot would help in determining the range over which the remote control should be functional. The single data piece, average distance, doesn’t seem to give much confidence in the conclusion.

5.4 Investigation 4

ACE Question 2 page 55

ANSWER
2a. You could randomly choose some number of cookies (say 5), find the average number of raisins per cookie, and multiply by 48 to estimate the number of raisins in a box.
2b. 1000 / 48 = 21 raisins