Samples & Populations

Investigation 1

Standards:

Essential for 7.SP.A.1 Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences. *Problems 1, 3, and 4*

Essential for 7.SP.A.2 Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions *Problems 3 and 4*

Essential for 7.SP.B.3 Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *Problems 2 and 3*

7.SP.B.4Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *Problems 1, 2, 3, and 4*

7.NS.A.1 Apply and extend previous understandings of addition and subtraction to add and subtract rational numbers; represent addition and subtraction on a horizontal or vertical number line diagram. *Problem 2*

7.NS.A.1b Understand $p+q$ as the number located a distance $|q|$ from p, in the positive or negative direction depending on whether q is positive or negative. . . Interpret sums of rational numbers by describing real-world contexts. *Problem 2*

Objectives:

Investigation 1 focuses on comparing two or more sets of data (having the same number or different numbers of data values) using a variety of strategies first introduced in *Data About Us*. It also guides students toward using data from samples to make predictions about populations.

Three of the Problems (1.1, 1.2, and 1.4) involve numerical data; Problem 1.3 uses categorical data. The Problems focus attention on the use of measures of center and variability (or spread). Measures of center (mode with respect to categorical data, mean, and median) describe what is typical of the center of a distribution of data. Measures of variability (range, mean absolute deviation, and interquartile range) describe how data are spread about the center of a distribution. Students rely on measures of variability, in conjunction with measures of center, to make comparisons.

The intent is to build a deeper foundation—connecting measures of center and spread to compare data sets. In subsequent Investigations, students will use their knowledge to compare samples with samples and samples with populations.

Investigation 2

Standards:

7.SP.A.1 Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences. *Problems 1, 2, 3, and 4*

7.SP.A.2 Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions *Problems 1, 3, and 4*

7.SP.C.7 Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy. *Problems 3 and 4*

7.SP.C.7a Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events. *Problems 3 and 4*

Objectives:

In Investigation 2, students consider the statistical distinctions between samples and populations. They also use results of data analyses from samples to draw conclusions about characteristics or behaviors of a population. First, students consider the implications of making estimates about the entire U.S. population based on an Internet survey involving a few thousand people. The survey raises the issue of projecting the results collected from a sample to an entire population.

Next, students consider the differences among different types of sampling methods: convenience, voluntary-response, systematic, and random. They explore techniques for choosing samples randomly from a population—such as using spinners, number cubes, and random-number generators on calculators—and consider why random samples are often preferable to other types of sampling methods. Then, they investigate the relationship between sample size and accuracy of population estimates. Students study sampling distributions of means and medians, noticing that the means or medians from larger samples (e.g., 30 values) taken from the same population are less variable than those from smaller samples (e.g., 5 or 10 values). This helps students realize that measures of center from larger samples are more similar; it is possible to use the mean from a large sample to draw reliable conclusions about the given population.

Investigation 3

Standards:

7.NS.A.1 Apply and extend previous understandings of addition and subtraction to add and subtract rational numbers; represent addition and subtraction on a horizontal or vertical number line diagram. *Problem 2*

7.NS.A.1b Understand $p+q$ as the number located a distance $|q|$ from p, in the positive or negative direction depending on whether q is positive or negative. . . Interpret sums of rational numbers by describing real-world contexts. *Problem 2*

7.RP.A.2 Recognize and represent proportional relationships between quantities. *Problem 4*

7.SP.A.1 Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences. *Problems 1, 2, 3, and 4*

7.SP.A.2 Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. *Problems 3 and 4*

7.SP.B.3 Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *Problem 2*

7.SP.B.4 Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *Problems 1 and 2*

7.SP.C.5 Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event. *Problem 3*

7.SP.C.7 Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy. *Problem 3*

7.SP.C.7a Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events. *Problem 3*

Objective:

Students apply what they have learned about samples and populations to engaging real-world problems in Investigation 3.

First, they analyze measurements of Native American arrowheads found at six different archeological sites. Scientists know the approximate time periods during which four of the sites were settled. The time periods for two newer sites are unknown. Students explore how data from the known sites may be used to draw conclusions about the newer sites.

In Problem 3.2, students determine whether the differences found between two samples of basketball player heights (one sample from a male professional basketball player population and the other from a female professional basketball player population) are most likely due to naturally occurring variability or meaningful differences between the distributions of heights in

the two populations. They base this analysis on the relationship of the samples' means and MADs.

Next, students use 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. They use concepts that they have learned about sampling techniques and probability to solve this real-world quality control problem.

Last, in Problem 3.4, students simulate the capture–tag–recapture method that scientists and policymakers often use to sample and draw conclusions about the sizes of wildlife populations. In this Problem, students use containers of beans to represent deer populations.