Name _____ Algebra 2E

10C Samples and Populations

Population: the collection of all data, such as responses, measurements, or counts that you want information about.

Sample: subset of a population

A *census* consists of data from an entire population. But, unless a population is small, it is usually impractical to obtain all the population data. In most studies, information must be obtained from a *random sample.*

It is important for a sample to be representative of a population so that sample data can be used to draw conclusions about the population. When the sample is not representative of the population, the conclusions may not be valid. Drawing conclusions about populations is an important use of statistics.

Example 1: Identify the population and the sample in each of the following. Describe the sample.

a. In the United States, a survey of 2184 adults ages 18 and over found that 1328 of them own at least one pet.

b. To estimate the gasoline mileage of new cars sold in the United States, a consumer advocacy group tests 845 new cars and finds they have an average of 25.1 miles per gallon.

Parameter: a numerical description of a population characteristic.

Statistic: a numerical description of a sample characteristic.

Because some populations are too large to measure, a statistic, such as the sample mean, is used to estimate the parameter, such as the population mean.

Example 2: Distinguishing between parameters and statistics.

a. For all students taking the SAT in a recent year, the mean mathematics score was 514. Is the mean score a parameter or a statistic? Explain.

b. A survey of 1060 women, ages 20-29 in the United States, found that the standard deviation of their heights is about 2.6 inches. Is the standard deviation of the heights a parameter or a statistic? Explain.

Hypothesis: a claim about a characteristic of a population. Here are some examples:

1. A drug company claims that patients using its weight-loss drug lose an average of 24 pounds in the first 3 months.

2. A medical researcher claims that the proportion of U.S. adults living with one or more chronic conditions, such as high blood pressure, is 0.45 or 45%.

To analyze a hypothesis, you need to distinguish between results that can easily occur by change and results that are highly unlikely to occur by chance. One way to analyze a hypothesis is to perform a *simulation*. When the results are highly unlikely to occur, the hypothesis is probably false.

Example 3: You roll a six-sided die 5 times and do not get an even number. The probability of this happening is $\left(\frac{1}{2}\right)^5 = 0.03125$, so you suspect this die favors odd numbers. The maker claims that the die does not favor odd numbers. What should you conclude when you roll the actual die 50 times and get a) 26 odd numbers and b) 35 odd numbers?

Name _____ Algebra 2E

Samples and Populations Homework

1. Determine whether the data are collected from a population or a sample. Explain your reasoning.

   a. The number of high school students in the United States.

   b. The color of every third car that passes your house.

2. Identify the population and the sample. Describe the sample.

   a. In the United States, a survey of 1152 adults ages 18 and over found that 403 of them pretended to use their smartphones to avoid talking to someone.

   b. In a school district, a survey of 1777 adults ages 18 and over found that 1279 of them do some kinds of spring cleaning every year.

   c. In the United States, a survey of 2000 households with at least one child found that 1280 of them eat dinner together every night.

3. Determine whether the numerical value is a parameter or a statistic. Explain your reasoning.

   a. The average annual salary of some physical therapists in a state is $76,210.

b. In a recent year, 53% of the senators in the U.S. Senate were Democrats.

c. Seventy-three percent of all the students in a school would prefer to have school dances on Saturday.

4. **Error Analysis**: A survey of 1270 high school students found that 965 students felt added stress because of their workload. Describe and correct the error in identifying the population and the sample.

"The population consists of all the students in the high school. The sample consists of the 965 students who felt added stress."

5. You flip a coin 4 times and do not get a tails. You suspect this coin favors heads. The coin maker claims that the coin does not favor heads. You simulate flipping the coin 50 times by repeatedly drawing 200 random samples of size 50. What should you conclude when you flip the actual coin 50 times and get:

a. 27 heads

b. 33 heads

c. 17 heads

Name _____ 10C Algebra 2E

10.4 Collecting Data

Random Sample: each member of a population has an equal chance of being selected.

Self-Selected Sample: members of a population can volunteer to be in the sample.

Systematic Sample: a rule is used to select members of a population. For instance, selecting every other person.

Stratified Sample: a population is divided into smaller groups that share a similar characteristic. A sample is then randomly selected from each group.

Cluster Sample: a population is divided into groups, called clusters. All of the members in one of more of the clusters are selected.

Convenience sample: only members of a population who are easy to reach are selected.

Example 1: You want to determine whether students in your school like the new design of the school's website. Identify the type of sample described.

a. You list all of the students alphabetically and choose every sixth student.

b. You mail questionnaires and use only the questionnaires that are returned.

c. You ask all of the students in your algebra class.

d. You randomly select two students from each classroom.

Bias: an error that results in a misrepresentation of a population.

In order to obtain reliable information and draw accurate conclusions about a population, it is important to select an *unbiased sample.*

Unbiased Sample: is representative of the population that you want information about.

Biased Sample: a sample that overrepresents or under-represents part of the population is a biased sample.

When a sample is biased, the data are invalid. A random sample can help reduce the possibility of a biased sample.

Example 2: Identify the type of sample and explain why the sample is biased.

a. A news organization asks its viewers to participate in an online poll about bullying.

b. A computer science teacher wants to know how students at a school most often access the Internet. The teacher asks students in one of the computer science classes.

Example 3: You are a member of your school's yearbook committee. You want to poll members of the senior class to find out what the theme of the yearbook should be. There are 246 students in the senior class. Describe a method for selecting a random sample of 50 seniors to poll.

Experiment: imposes a treatment on individuals in order to collect data on their response to the treatment. The treatment may be a medical treatment, or it can be any action that might affect a variable in the experiment, such as adding methanol to gasoline and then measuring its effect on fuel efficiency.

Observational Study: observes individuals and measures variables without controlling the individuals or their environment. This type of student is used when it is difficult to control or isolate the variable being studied, or when it may be unethical to subject people to a certain treatment or to withhold it from them.

Survey: an investigation of one of more characteristics of a population. In a survey every member of a sample is asked one of more questions.

Simulation: uses a model to reproduce the conditions of a situation or process so that the simulated outcomes closely match the real-world outcomes. Simulations allow you to study situations that are impractical or dangerous to create in your life.

Example 4: Identify the method of data collection each situation describes.

a. A researcher records whether people at a gas station use hand sanitizer.

b. A landscaper fertilizes 20 lawns with a regular fertilizer mix and 20 lawns with a new organic fertilizer. The landscaper then compares the lawns after 10 weeks and determines which fertilizer is better.

When designing a survey, it is important to word survey questions so they do not lead to biased results. Answers to poorly worded questions may not accurately reflect the opinions or actions of those being surveyed. Questions that are flawed in a way that leads to inaccurate results are called **biased questions.**

Avoid questions that:

1. Encourage a particular response.

2. Are too sensitive to answer truthfully.

3. Do not provide enough information to give an accurate opinion.

4. Address more than one issue.

Example 5: A dentist surveys his patients by asking, "Do you brush your teeth at least twice per day and floss every day?" Explain why the question may be biased or otherwise introduce bias into the survey. Then describe a way to correct the flaw.

Name _____ Algebra 2E

Collecting Data Homework

1. Identify the type of sample described.

a. The owners of a chain of 260 retail stores want to assess employee job satisfaction. Employees from 12 stores near the headquarters are surveyed.

b. Each employee in a company writes their name on a card and places it in a hat. The employees whose names are on the first two cards drawn each win a gift card.

2. Identify the type of sample and explain why the sample is biased.

a. A town council wants to know whether residents support having an off-leash area for dogs in the town park. Eighty dog owners are surveyed at the park.

b. A sports writer wants to determine whether baseball couches think wooden bats should be mandatory in collegiate baseball. The sportswriter mails surveys to all collegiate coaches and uses the surveys that are returned.

3. Determine whether the sample is biased. Explain your reasoning.

a. Every third person who enters an athletic event is asked whether he or she supports the use of instant replay in officiating the event.

b. A governor wants to know whether voters in the state support building a highway that will pass through a state forest. Business owners in a town near the proposed highway are randomly surveyed.

4. The staff of a student newsletter wants to conduct a survey of the students' favorite television shows. There are 1225 students in the school. Describe a method for selecting a random sample of 250 students to survey.

5. Identify the method of data collection the situation describes.

a. A researcher uses technology to estimate the damage that will be done if a volcano erupts.

b. The owner of a restaurant asks 20 customers whether they are satisfied with the quality of their meals.

6. Explain why the survey question may be biased or otherwise introduce bias into the survey. Then describe and correct the flaw.

a. "Do you agree that the budget of our city should be cut?"

b. "Would you rather watch the latest award-winning movie or just read some book?"

7. A researcher studies the effect of fiber supplements on heart disease. The researcher identified 175 people who take fiber supplements and 175 people who do not take fiber supplements. The study found that those who took the supplements had 19.6% fewer heart attacks. The researcher concludes that taking fiber supplements reduces the chance of heart attacks.

a. Explain why the researcher's conclusion may not be valid.

b. Describe how the researcher could have conducted the study differently to produce valid results.