In many situations, data from samples is required to estimate characteristics (parameters) of large populations. The reason for using a sample may be that the population is too large to permit data to be gathered from every subject. For example, a national opinion poll about a political issue would require data from samples. Another reason is that sometimes the act of collecting the data ruins the item being studied, as when crash-testing automobiles.

Understandably, to get reliable results, one must minimize the sources of bias. Random selection of subjects is used to reduce bias in statistical studies. Problem 4-3 explains types of bias found within survey questions.

For additional information, see the Math Notes box in Lesson 4.1.1.

Examples

Determine whether each situation describes a sample or census. If the situation describes a sample, discuss the sampling technique and potential sources of bias.

a. The Chief of Police calls in his five newest officers to get their opinion on new requirements for department-wide promotions.
   Answer: The Chief is using a sample, which is not likely representative of the population affected by the new requirements. The newest officers would likely want to please the Chief and would echo his support or dislike of the new requirements.

b. A manager compares the annual sales totals for 12 store locations to rank their performance.
   Answer: This is a census. The manager has information from all 12 of the stores he is interested in comparing.

c. Shoppers are invited to fill out a questionnaire about their shopping experience at the cash register. The store manager uses the results in a report to corporate headquarters to demonstrate the level of customer satisfaction at her store.
   Answer: The manager is using a voluntary response sample (or “convenience sample”), which would not represent all customers very well. A sample like this tends to over-represent those with the strongest opinions.

d. The student council is trying to determine how much space they will need for the prom. Carmine, the Junior Class President, walks around during lunchtime at her school with a clipboard, asking, “Are you planning on going to the best prom ever?”
   Answer: Carmine is not taking a random sample. She is more likely to talk to students she knows, and they, as a group, are likely to be biased about the prom. There may also be a strong desire to please the interviewer in this situation. The wording of the question itself is biased, using “best prom ever” to describe the dance.
Problems

Determine whether each situation describes a sample or census. If it is a sample, discuss the sampling technique and potential sources of bias.

1. A math teacher wants to determine whether playing classical music during testing benefits high school math students. He plays classical music to half of his classes while they test, and compares their scores to the other half of his classes, who tested without the music.

2. A sports news reporter wants to know the win-loss record of the local high school girls lacrosse team, so he looks at the league website and sees that they have 8 wins and 5 losses.

3. A battery manufacturer wants to monitor the durability and peak voltage of its products. A machine is programmed to select every 1000th battery from the production line and subject it to a series of tests.

4. Scott is doing a report on the risk factors for cancer. He asks all of the members of his PE class: “Do you have a family history of cancer?” “Do you eat five servings of fruits and vegetables each day?”

Answers

1. Because the math teacher is using his data to make a statement about all high school math students (the population), his classes represent a sample of those students. His students would have many things in common, like where they live and how they were taught, which other high school math students do not have. They would be a poor representation of the population. The teacher may also be biased in grading the exams to obtain the results he favors.

2. This is a census. The reporter has access to every game outcome for the team he is researching and is not using the information to draw conclusions about other teams or the sport of lacrosse in general.

3. The batteries tested represent a sample of those being made. They might be a reasonable representation of the population of batteries. However, if the 1000th battery is usually made at the same time every day, or corresponds to some other variable like periodic machine maintenance, it might be subject to bias.

4. Scott’s PE class represents a sample of people or students. However this is a convenience sample and cannot be considered a reliable representation of the population of people or students. The question order may cause bias in the results as people being reminded of a disease like cancer would likely want to appear concerned enough to be eating a healthy diet with lots of fruits and vegetables. Members of a PE class may be more inclined to be thinking about health issues such as diet.