B. PLANNING AND CONDUCTING SURVEYS

The purpose of a sample survey is to determine one or more characteristics about a population from the results of the sample. The characteristic of the population is a **population parameter**. The result of the sample survey used to estimate the parameter is a **sample statistic**. There are many methods to obtain a sample from a population.

Sampling Methods

A **voluntary response sample** is composed of individuals who choose to respond to a survey because of interest in the subject, particularly those with strong opinions or attraction.

A **convenience sample** is composed of individuals who are easily accessed or contacted.

Example: A journalism class prints a survey in their school newspaper. Readers are asked to clip the survey from the paper, complete it, and return it to a drop box in the school cafeteria. What type of sample is this?

Answer: This is a voluntary response sample. It is likely that only students interested in the survey's subject matter will take the time to respond.

Example: A journalism class stations pollsters in front of the stadium during a football game. They ask each student who enters his or her opinion of the quality of the school's athletics program. What type of sample is this?

Answer: This is a convenience sample. The site was likely chosen because of the ease of gathering data there.

A **random sample (or probability sample)** is composed of individuals selected by chance. There are several types of random samples.

A **simple random sample (SRS)** is a sample where n individuals are selected from a population in a way that every possible combination of n individuals is equally likely. In general, one may think of an SRS as having all individuals of the population in a hat and selecting n individuals, without replacement, from the hat.

Note: Not only is it required of a simple random sample that each *individual* has an equally likely chance of being chosen, but also each possible group must be equally likely. That is, if an SRS is to have a size of $n = 5$, no possible group of five individuals can be any more or less likely to be selected than any other group.

A **stratified random sample** is a sample in which simple random samples are selected from each of several homogeneous subgroups of the population, known as **strata** (singular, **stratum**).

Note: In a stratified random sample, every individual in each subgroup has an equally likely chance of being chosen, but every individual in the population may not. Additionally, each possible group of individuals from the population will not have the same chance of being chosen.

A **cluster sample** is a sample in which a simple random sample of heterogeneous subgroups of the population is selected. The subgroups are known as **clusters**. The selected clusters may also themselves be subject to random sampling, or a census done for each.

A **multistage sample** is a sample resulting from multiple applications of cluster, stratified, and/or simple random sampling.

A **systematic random sample** is a sample where every kth individual is selected from a list or queue. The first selection is randomly chosen from the first k individuals.

Example: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The voters' ID numbers are entered into a computer and 500 are chosen at random, without replacement, to comprise the sample. What type of sampling design is this?

Answer: This is a simple random sample. Individuals are chosen at random, without replacement, from a list of the entire population. This is a "select n from a hat"-type of sampling.

Note: In this scenario, every possible group of 500 individuals has an equally likely chance of being selected, as does every individual.

Example: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The voters' ID numbers and party affiliations (Democrat, Republican, Independent) are entered into a computer. The firm randomly selects, without replacement,

200 Democrats, 200 Republicans, and 100 Independents to comprise the sample. What type of sampling design is this?

Answer: This is a stratified random sample. Individuals were first placed in strata, in this case, party affiliation. Then, simple random samples from each stratum were selected to make up the final sample.

Note: In this scenario, all individuals within their respective strata have an equally likely chance of being chosen, but all individuals in the population may not. If there are more Democrats than Republicans, then it is more likely that a particular Republican voter will be in the sample than a particular Democrat voter. Additionally, every possible group of 500 individuals from the population does not have an equally likely chance of being chosen, since samples of 500 Democrats, or 250 Democrats and 250 Republicans, for example, are not possible.

Example: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The voters' ID numbers and voting precinct numbers (01–75) are entered into a computer. The firm first randomly selects 20 of the voting precincts. Then, from each of the chosen precincts, 25 voters are chosen with simple random samples. What type of sampling design is this?

Answer: This is a cluster sample. Individuals were first grouped into clusters by voting precinct, of which several were selected. Then, simple random samples from each cluster were selected to make up the final sample.

Note: An additional step could have been added to the previous example. If voters in the selected precincts were then grouped by party affiliation, and an SRS taken from each group, the result would have been a multistage sample.

Example: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The list is sorted by voter ID number. The firm randomly selects one person from the first 100 on the list; then it selects every 100th person after that. The sample size is 150. What type of sampling design is this?

Answer: This is a systematic random sample. One of the first k individuals on the list is randomly selected—in this case $k = 100$—and every kth individual on the list is selected after that.

EXAMPLE:

A large shipment of apples is received at a warehouse. Five different samples are selected to determine the average weight of an apple in the shipment. The average weight of the apples in each sample is determined. Although the sample averages are close in value, they are not the same. Does this mean that the sampling techniques were in error? Does each sample produce a good estimate of the actual average weight of all the apples in the shipment?

The average weight of all the apples in the shipment is not known. Each sample produces an estimate of the average weight. Each sample estimate could be valid, even though they are different. If good sampling techniques were used to make each selection, each sample is valid, and some variation in the results is expected. It is critical that each sample be representative of the entire population of apples.

Sources of Bias in Surveys

Bias occurs in surveys when one outcome is favored over another. Bias can be intentional or unintentional. Most bias occurs because the sampling process did not produce representative samples.

Undercoverage Bias

This type of bias occurs when part of the population is excluded from the sampling process.

EXAMPLE:
A survey of college students to determine their feelings about a building expansion is taken at lunchtime in the cafeteria. This survey shows undercoverage bias since it excludes evening students. A survey is taken of motorists at a gas station asking them whether they favor rent control. This survey shows undercoverage bias since it does not include people who do not have cars and use public transportation. These excluded people might be mostly renters and have a very different opinion than home owners.

Voluntary Response Bias

This type of bias occurs when the sample is self-selected. People with strong opinions are more likely to want to participate. When people offer to be included in a survey, their opinions are not usually representative of the population.

EXAMPLE:
The city council wants to determine the opinion of residents about installing speed bumps on a particular street. A questionnaire is sent out to all residents in the neighborhood. Only a small percentage of the questionnaires are returned, and most of those do not want the speed bumps. This is an example of voluntary response bias. The small response probably included those people most interested in the speed bumps, either because they use the street where the bumps are to be installed, or they live on the street. Their strong opinions will produce biased results.

Nonresponse Bias

This type of bias occurs when people chosen for a survey refuse to respond or are unable to respond.

EXAMPLE:
A several page questionnaire is sent out by a car dealership seeking information about your satisfaction with the new car you recently purchased. Although you like the car, you don't want to take the time to fill out the lengthy questionnaire. This represents a nonresponse bias. Someone who is dissatisfied with their new car might be more likely to fill out the long questionnaire. This could also lead to voluntary response bias.

Wording Bias

The way a question is worded can influence the results. If the question is stated in a way to emphasize the positive aspects of the situation, it is more likely to result in a favorable response. If the question is stated with a negative emphasis, an unfavorable response might be more likely.

EXAMPLE:

Opinions about a local tax to help schools build new gymnasiums is the subject of a survey. A positive response is more likely if the question focuses on "helping young people stay in school by encouraging participation in after school athletics" than if the question focuses on "the tax will cost property owners $320 in additional taxes each year."

EXAMPLE:

A prescription drug manufacturer is more likely to get a favorable response from a survey if it emphasizes the benefits of taking a new medication for diabetes than if it emphasizes the side effects. (Notice that in commercials on television about prescription drugs, most of the time is spent talking about the positive aspects, and the last few seconds relate the negative side effects.) The wording of a question can produce an intentional or unintentional wording bias.

Response Bias

Response bias occurs in several situations. If an oral interview is conducted or if the questionnaire is not anonymous, the respondent may want to please the interviewer rather than give their true opinion. If questions are of a personal nature, respondents might not tell the truth. If a question is phrased in a way that makes it difficult to understand, the respondent might be embarrassed that they do not understand and give a false answer.

EXAMPLE:

A teacher wants to determine whether he presented the material in a chapter in a way that students enjoyed. He asks the students in his class to choose a number from 1 (bad) to 5 (good) that represent their opinions of his teaching style. He asks them to write the number at the top of their chapter test. Would this technique produce biased results?

Students usually want to please their teacher, especially while he is grading their test. Students are more likely to give favorable replies if their responses are not anonymous.

Selection Bias

If a sample is chosen simply because it is easy to obtain, it is known as a **convenience sample.** If you want to choose five students in your class and you choose the five that sit closest to you, you have obtained a convenience sample. Convenience samples might or might not produce biased results. If the convenience aspects of the sample are related in any way to the survey, then bias has occurred.

If a survey about child care is taken on a popular website, several types of bias can occur. First, only people who visit the website, and, therefore have or use computers, are included in the survey. Not all visitors to the website might choose to participate. Only those with a particular interest in the question being surveyed would have a tendency to answer. The type of website could have a lot to do with the responses. The survey on a financial website would probably produce different results than the survey on a marriage and counseling website.

Unintentional Bias

It is difficult for a person to choose survey subjects randomly. If you ask someone to choose five random digits chosen from the digits 0 through 9, they are very likely to choose five different digits. Although they think they are choosing randomly, they are not. After choosing the first random digit, they mentally eliminate that digit when choosing subsequent digits. If the random choice is left to the interviewer, this unintentional bias can occur. If you have ever taken a multiple choice test and for a particular problem you have no clue as to which answer is correct, you can make a random selection. If you think back to see how many of your previous choices were "a" and how many were "b," then you are introducing unintentional bias in your random selection of an answer. The best way to avoid unintentional bias is to use acceptable selection techniques as illustrated in the following sections.