

Section 6.1 Sampling

Population – each element (or person) from the set of observations that can be made (entire group)

Sample – a subset of the population

Census – systematically getting information about an entire population

Sampling – studying a part (a sample) in order to gain information about an entire group

Sampling Frame – the list of individuals from which a sample is actually selected

Types of Sampling –

A **simple random sample (SRS)** consists of individuals from the population chosen in such a way that every set of individuals has an equal chance to be the sample actually selected.

A **probability sample** gives each member of the population a known chance to be selected.

A **stratified sample** divides the population into smaller groups, called strata, and chooses an SRS in each stratum and combines these to form the full sample.

For example, to obtain a stratified sample of university students, the researcher would first organize the population by college class and then select appropriate numbers of freshmen, sophomores, juniors, and seniors.

In **multistage sample** design samples are taken from various subsets of the population until a manageable number of samples to interview are arrived upon.

For example, you want to conduct a survey of salespeople for a nationwide retail chain with stores all over the country. You could randomly select states, randomly select counties in each state, randomly select stores in each county, and randomly select salespeople in those stores.

Convenience sampling is a non-probability type of sample where the sample is chosen based on their convenient accessibility and proximity (leads to under coverage bias)

Voluntary Response sample – people who choose themselves by responding to a general appeal (over represents people with strong opinions)

Convenience sampling and voluntary response sampling are examples of bad sample design.

Example 1: In each of the following, identify the population and the sample then describe the sampling method that was used.

a. Students at an elementary school were assigned numbers. Fifty numbers were chosen at random and the students that were assigned those numbers were asked to rate their Spring carnival.

Population: All students

Sample: 50 chosen students

Type of Sampling: Simple Random Sample (SRS)

b. A grocery store gave samples of some strawberries to their customers and asked them if they liked the product or not. Fifty customers responded; forty-one said yes.

Population: Store customers

Sample: 50 responders

Type of Sampling: Voluntary

c. A random sample of ten employees from each of five different departments was selected and each was asked to rate their department's performance.

Population: Employees

Sample: 10 employees from 5 departments

Type of Sampling: Stratified

d. A random sample of large urban school districts throughout Houston was selected and the selected districts were identified as target districts. Within each district, a SRS of its high schools was chosen and the principals of those high schools were interviewed. (Note: In this case, results of the interviews can be used to infer responses of the population of the high school principals in Houston.)

Population: All urban school districts in Houston

Sample: HS principals

Type of Sampling: Multistage

Random Sample

→ Computer Software
→ Calculator
→ Table of Random Digits

A **table of random digits** is a long string of the digits 0 – 9 where each entry in the table is equally likely to be any of the 10 digits and the entries are independent of each other. This is **Table B in the appendix of your book.**

Here's a glimpse at the table:

Table of Random Digits

Line								
101	98360	26534	47384	94612	88666	14170	10847	05567
102	55556	59863	86607	00094	77213	35711	52851	42108
103	31634	15399	73476	77412	06186	16636	54307	14947
104	13785	11509	54891	98375	68377	50572	08453	80376
105	80376	73842	95465	59746	38078	25727	78502	95324

Example 2: Using the table of random digits in the appendix, **starting at line 130**, to choose an SRS of 5 players from the football team to interview about last week's game. Assume there are 65 players on the team.

130 | 07831 | 65838 | 52005 | 63845 | 08148 | 35315 | 13588 | 06969

Step 1: Label the players 01 – 65.

Step 2: Go to line 130, select two digits at a time. If one falls outside 01 – 65, discard.

07, ~~83~~, 16, 58, 38, 52, ~~00~~, 56

Which **five players** will be interviewed?

07, 16, 58, 38, 52

Once a population is identified and a sample is chosen, we must determine how to get the information we need. This can be done by **conducting an experiment** or **performing an observational study**.

Experiment – actively impose some treatment in order to observe the response

Example: Children playing violent video games, does it make them more prone to violent behavior? Have a group of children come into a lab and play violent video games, then watch those children play with other children and record any violent behavior.

All experiments have **extraneous factors** (the time of day the experiment was run, the room temperature, etc).

Observational study – investigators observe subjects and measure variables of interest without assigning treatments to the subjects. The treatment that each subject receives in an observational study is beyond the control of the investigator.

Example: Observing eating habits in elderly people at a nursing home.

Example 3: Determine if the study is an experiment or an observational study.

a. A personnel director at a large company studied the eating habits of employees by watching the movements of a selected group of employees at lunchtime. The purpose of the study was to determine the proportion of employees who buy lunch in the cafeteria, bring their own lunches, or go out to lunch.

Experiment or Observational

b. A pharmacy student would like to know if there is a difference in results from a specific brand of drug and its equivalent generic prescription. She randomly selects 50 people who take the drug and has them complete a questionnaire regarding their symptoms and improvements after taking the specific brand or generic equivalent.

Experiment or Observational

The design of an experiment or study is **biased** if it methodically favors certain outcomes.

Under-coverage is a type of bias that occurs when some groups in the population are left out of the process of choosing the sample.

*Example: In Example 3, part a, if the director limits this study to only one department, then that would be an example of **convenience sampling**. Recall: Convenience sampling is a non-probability type of sample where the sample is chosen based on their convenient accessibility and proximity.*

Nonresponse is another type of bias that occurs when an individual chosen for the sample cannot be contacted or refuses to cooperate.

Example: Sending out a questionnaire to a selected number of subjects and not all return the questionnaire.

Two elements are confounded when their effects on a response variable cannot be distinguished from one another.

For example, a study of coffee drinking and lung cancer. If coffee drinkers were also more likely to be cigarette smokers, and the study measured coffee drinking but not smoking, the results may seem to show that coffee drinking increases the risk of lung cancer, which may not be true. However, if a confounding factor (in this example, smoking) is recognized, adjustments can be made in the study design or data analysis so that the factor does not confound (confuse) the study results.

Statistical inference provides ways to answer specific questions from data with some guarantee that the answers are good ones. In inference we must think about how to produce data as well as analyze data.

The **wording of questions** can also introduce strong bias when misleading or confusing questions are used.

Example 4: Here are two wordings for the same question:

1. Should laws be passed to eliminate all possibilities of special interests giving huge sums of money to candidates?

2. Should laws be passed to prohibit interest groups from contributing to campaigns, or do groups **have a right** to contribute to the candidates they support?

One of these questions drew 40% favoring banning contributions; the other drew 80% with this opinion.

a. Which question produced the 40% and which got 80%?


#2 #1

b. Can this question be reworded to eliminate all bias?

Yes. For example, "Should laws be passed to prohibit interest groups from contributing to the campaigns of political candidates?"