Disclaimer

© 2017 Your Stat Class. Latrica L. Williams

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means, graphic, electronic, or mechanical including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, information storage and retrieval systems, without written permission of the company or author. Use in violation of these terms and conditions is prohibited.

REPRODUCING THIS WORK BY ANY MEANS WILL RESULT IN LEGAL ACTION. VIOLATORS WILL BE PROSECUTED TO THE FULL EXTENT OF THE LAW.
Chapter 2

RESEARCH DESIGN
CASE STUDY:

For Americans, has the cost of monthly groceries increased since 2012?

To answer this question, researchers sampled 2500 Americans and asked two questions, which were the following:

**Question 1**: How much money did you spend in a typical month on groceries in 2012?

**Question 2**: How much money do you spend in a typical month on groceries nowadays?

Responses were documented. After analyzing the data, it was found that Americans are spending more money today on groceries than in 2012. So, how were the researchers able to answer the question and draw a conclusion about all Americans?

Well, they had to address the following concepts before analyzing the data:

1. Identify a research question.
2. Define the population.
3. Determine a sample.
4. Determine the variable of interest.
5. Determine how the data will be collected.
6. Determine what type of study will be performed.

In the following chapter, you will learn about these concepts and their applications.
Section 2.1:

Research Question

Student Learning Outcomes

By the end of the section,

1. You will understand the importance of stating a research question.
2. You will define population and sample.
In this book, the focus will be on applied statistics. In applied statistics, researchers generally begin by stating a research question.

**Definition**

Research Question: a clearly-defined question researchers would like to answer.

Stating a research question is the core of many studies. Research questions can be broad or specific. It depends on the interest of the researcher.

For example, researchers can be interested in determining the average age of community college students in the United States.

In contrast, they can be interested in determining the average age of community college students in the state of Arizona.

Research questions define the population.

**Definition**

Population: all subjects of interest.

Suppose the research question was to determine the average age of community college students in the United States. The population would be ALL community college students in the United States.

On the contrary, suppose the research question was to determine the average age of community college students in the state of Arizona. The population would be ALL community college students in the state of Arizona.
Ideally, researchers would like for every subject of interest to be a part of the study. However, due to time, money and resources, researchers are not always able to obtain every subject of interest. Therefore, they gather a sample.

**DEFINITION**

**Sample:** a subset of the population.

For example, suppose the research question was indeed to determine the average age of community college students in the United States. The population would be ALL community college students in the United States. Theoretically, researchers would like for every community college student in the United States to be a part of the study. However, it would be time consuming to gather information from every community college student in the United States. Moreover, if a small research company was interested in addressing this question, it may not have the funds or resources to gather the required information from all subjects. Therefore, a sample is gathered.

The size of the sample depends on the preference of the researcher. The sample size can be 30 or 30,000. (Later in the book, formulas will be used to determine an appropriate sample size.)

**STOP!!** Although the sample size can vary, ensure the sample size is not too small. Moreover, extremely large samples are not necessarily required.

Samples are important in statistics because researchers rarely analyze entire populations. In a sense, statistics is the “study of samples”. Thus, researchers use sample information to make inferences about the population.
Section 2.2:

Sampling Techniques and Bias

Student Learning Outcomes

By the end of the section,

1. You will define random and representative in the context of statistics.
2. You will identify different types of sampling techniques.
3. You will recognize types of sampling bias.
When obtaining a sample, ensure the subjects in the sample are **BOTH** random and representative.

### Definition

**Random:** each subject has an equal chance of being selected.

**Representative:** each subject is representative of the population.

Selecting subjects **randomly** helps to ensure reliability and validity of the study. Moreover, the subjects should also be **representative** of the population.

For example, if a researcher is interested in analyzing the average annual income of residents in Georgia, residents of Maine should not be included in the study. Residents of Maine would not be representative of the population.

There are several sampling techniques that can be used to gather a sample. The first technique is a simple random sample.

### Definition

**Simple Random Sample:** a sample in which each subject of the population has an equal chance of being selected.

Selecting a sample using this technique is similar to placing names in a hat, then randomly choosing subjects from the hat to be a part of the sample.

Technology such as a random number generator can be used to randomly select subjects as well.
The second technique is a stratified sample.

**Definition**

**Stratified Sample:** a sample in which subjects of the population are first divided into strata, then subjects are randomly selected from each stratum.

Suppose a researcher was interested in comparing people from different socioeconomic levels. To make the comparison, the researcher would need to have a sample of subjects that represent each socioeconomic level. Using this sampling technique, the researcher would begin by dividing the population into lower-class, middle-class and upper-class groups. Next, the researcher would randomly select subjects from each group to be a part of the sample.

This sampling technique is advantageous over the simple random sample because it allows the researcher to ensure that each group is represented.

If a simple random sample was used, the researcher could randomly select all subjects from the middle-class group due to chance.

The third technique is a cluster sample.

**Definition**

**Cluster Sample:** a sample in which clusters of subjects are randomly selected from the population.

Suppose a researcher was interested in the average annual income of South Dakota residents. Using this sampling technique, the researcher would divide the population into clusters such as counties.
Next, the researcher would select ALL subjects from randomly selected counties (clusters) to be a part of the sample.

The fourth technique is a systematic sample.

**Definition**

Systematic Sample: a sample in which every $x^{th}$ subject of the population is randomly selected.

Selecting a sample using this technique begins by choosing an arbitrary starting point (it doesn’t have to always be the first subject). From there, every $x^{th}$ subject is selected.

*Note:* $x$ can represent any number. For example, every $4^{th}$, $15^{th}$, or $100^{th}$ subject is randomly selected from the population.

The last technique is a convenience sample.

**Definition**

Convenience Sample: a sample in which the subjects of the population are selected conveniently.

Subjects selected using this technique are easily obtainable. Examples include selecting the first 50 people leaving a movie theater or selecting the first 100 customers entering a mall.

In conclusion, some sampling techniques are more advantageous than others. Choosing which technique to use depends on the researcher and the research question.
When selecting a sample, one must be aware of potential bias. There are three common types of bias. The first type of bias is nonresponse bias.

**DEFINITION**

Nonresponse Bias: bias that occurs when subjects do not respond to a questionnaire (survey) or cannot be contacted.

For example, a questionnaire is mailed to 300 subjects but only 25 respond. This is an example of nonresponse bias.

The second type of bias is response bias.

**DEFINITION**

Response Bias: bias that occurs when a question is poorly worded or when an interviewer can influence the subjects’ response.

For example, a teacher asks a student, “Have you ever cheated on a test?” This is an example of response bias because most students would respond a particular way due to the person asking the question.

The last type of bias is undercoverage bias.

**DEFINITION**

Undercoverage Bias: bias that occurs when representative subjects are not included in the sample selection process.

For example, a researcher is interested in the average age of Alabama residents. The researcher decides to sample subjects from the entire state except for the Southeast region of the state.
This is an example of undercoverage bias because the entire state was not represented in the sample.

All in all, when selecting a sample, ensure the subjects are randomly selected and representative of the population.

Also, refrain from using bias when selecting a sample because it can distort the results of the study.
Section 2.3:

Data

Student Learning Outcomes

By the end of the section,

1. You will define variable of interest.
2. You will identify the two types of data.
3. You will describe the types of data collection techniques.
Once a sample is obtained, information is gathered from the subjects. The information gathered depends on the variable of interest, which is often identified in the research question.

**DEFINITION**

**Variable of Interest:** a characteristic of the population.

There can be one or more variables that are of interest to the researcher.

Let’s look at an example. Suppose researchers are interested in the average weight of babies born at a local hospital.

The variable of interest is weight because weight is the information to be gathered from the babies.

Researchers gather information from subjects in every research study. The information gathered from the subjects creates a set of **data**.

There are two types of data. Data can be qualitative or quantitative.

**DEFINITION**

**Qualitative Data:** data that can be categorized.

Examples of qualitative data include gender, marital status, makes of cars, majors and hair color.

Suppose a researcher is interested in the hair color of models at a popular modeling agency. The variable of interest is hair color. The information gathered could include brown, red, grey, blond and black.
Examples of quantitative data include weight, height, time, number of students and miles.

Quantitative data can be further classified into two categories. Quantitative data can be either discrete or continuous.

Examples of discrete data include number of bikes, number of computers, number of customers and number of books.

Examples of continuous data include length of marlin, temperature and age of breast cancer patients.

Data can be collected using one or more of the following data collection techniques:

- a personal interview
- a mailed questionnaire
- direct observation
- a survey of records

The data collection technique used in a study is based on the researcher’s preference and accessibility.
For example, suppose researchers are interested in determining the average weight of a 5-year old child. Researchers might choose to survey records rather than interview the children.
Section 2.4:

Types of Studies

Student Learning Outcomes

By the end of the section,

1. You will define and recognize an observational study.
2. You will define and recognize an experiment.
The type of study that a researcher performs depends on the research question and the focus of the study.

If a researcher is merely observing the subjects and documenting the data, an observational study is performed. For example, suppose a researcher wanted to know if drivers of red colored cars were more likely to run a red light. The researcher could simply be positioned at an intersection and observe the colors of cars running red lights.

In this situation, the researcher has not applied any treatment to the subjects. The researcher has simply “observed” the results.

On the contrary, a researcher can perform an experiment. With an experiment, treatment is applied to the subjects and the conditions are manipulated before responses are recorded. The type of treatment can vary. Again, it depends on the researcher and research question. For example, the treatment can be a new drug, a new teaching technique or a newly designed product.

In an experiment, there are typically two groups: a treatment group and a control group.

The treatment group receives the treatment; and the control group receives no treatment or a placebo. A placebo is an unmedicated drug designed to resemble the real drug.

Experiments can also be blind or double-blind.

DEFINITION

Blinding: a technique where the subjects do not know whether they received the treatment or placebo.

Double-Blinding: a technique where neither the subjects nor the experimenter(s) know which subjects received the treatment or placebo.
Blinding, more importantly double-blinding, a study is preferred by many researchers. It helps to keep the study and results unbiased. There are times when subjects may respond favorably to a drug if they are aware they are receiving the drug. This is known as the **Hawthorne effect**. As a result, the results can be skewed. To eliminate bias by subjects and manipulations by experimenters, researchers most often double-blind an experiment.

As stated earlier, deciding whether a study is observational or experimental depends on the research question. There are times when an observational study can only be performed. For example, does the use of cell phones cause brain cancer? It would be unethical to provide cell phones to subjects and require the subjects to use them to determine if they will get brain cancer.

For the most part, researchers prefer performing an experimental study over an observational study. With observational studies, researchers merely observe the results and draw conclusions based on the findings.

With experiments, researchers can apply treatment and control the testing situation. With all things being equal other than the treatment, researchers can determine if the treatment is effective or not. This also allows them to make **causal statements**. Causal statements can be made with experimental studies because the situation is controlled and all aspects are equal other than the treatment. Causal statements cannot be made with observational studies because the testing situation is not controlled. Thus, only observations are made.