Chapter 2 – Designing Observational Studies and Experiments
Section 1 – Simple Random Sampling

Materials Needed: TI Calculators, StatCrunch

Objectives
1. Identify the individuals, the variables, and the observations of a study.
2. Describe the five steps of statistics.
3. Identify the sample and the population of a study.
4. Identify descriptive statistics and inferential statistics.
5. Select a simple random sample.
6. Identify the sampling bias, the nonresponse bias, and the response bias of a study.
7. Identify sampling and non-sampling error.

Vocabulary
1. individuals
2. variable
3. observations
4. population
5. census
6. sample
7. descriptive statistics
8. inferential statistics/inferences
9. simple random sample/sampling
10. with/without replacement
11. frame
12. seed
13. sampling/non-sampling error
14 bias

Lesson/Activity
OBJECTIVE 1 – Identify the individuals, the variables, and the observations of a study.

Individuals are the people or objects we want to learn about.

Definition: Variable
In statistics, a **variable** is a characteristic of the individuals to be measured or observed.

Observations are data that we observe for a variable.

1. The number of listens of certain songs during the week ending Sunday, February 2015 on Last.fm are shown in the following table.

Song	Band	Genre	Number of Listens
"Breezeblocks"	Alt-J	Alternative	1436
"10 Bands"	Drake	Hip-Hop	5817
"Backseat Freestyle"	Kendrick Lamar	Hip-Hop	1466
"Chandelier"	Sia	Pop	3256
"Blue Moon"	Beck	Alternative	5360

Source: Last.fm

a. Identify the individuals.   b. Identify the variables.   c. Identify the observations for each variable.

d. Of the songs listed in the table, which one was listened to the most?

OBJECTIVE 2 – Describe the five steps of statistics.

Definition: Statistics
Statistics is the practice of the following five steps:
1. **Raise a precise question about one or more variables.**
2. **Create a plan to answer the question.**
3. **Collect the data.**
4. **Analyze the data.**
5. **Draw a conclusion about the question.**

2. *Motivational enhancement therapy* (MET) is a counseling process that aims to motivate drug abusers to stop using drugs. Some researchers wanted to test whether MET works. Out of 70 drug abusers who participated in the study, 35 received the standard drug treatment provided at a hospital. The other 35 individuals received MET in addition to the standard drug treatment. After receiving treatment for 12 weeks, all 70 individuals took a questionnaire which measures a drug abuser's motivation to stop using drugs. The researchers concluded that MET effectively motivates drug abusers to stop using drugs. (Source: Motivational Enhancement Therapy for Substance Abusers: A Quasi Experimental Study, Seema Rani etal.). Describe the five steps of the study. If no details are given about a step, say so.

OBJECTIVE 3 – Identify the sample and the population of a study.

Definition: Population
A **population** is the entire group of individuals about which we want to learn.

A **census** is a study in which data are collected about all members of the population.

Definition: Sample
A **sample** is the part of a population from which data are collected.

Collecting data from a sample, rather than the entire population, saves time, money, and effort.

3. In a poll of 1000 American adults, 13% of respondents said that concerns about global warming are unwarranted. The study then made a conclusion about all American adults. (Source: NBC News/Wall Street Journal)

   a. Define a variable for the study.   b. Identify the sample.   c. Identify the population.

OBJECTIVE 4 – Identify descriptive statistics and inferential statistics.

   Descriptive statistics is the practice of using tables, graphs, and calculations about a sample to draw conclusions about only the sample.

   Inferential statistics is the practice of using information from a sample to draw conclusions about the entire population. When we perform inferential statistics, we call the conclusions **inferences**.

4. Researchers wanted to find out how a restricted diet would affect a Labrador Retriever's lifespan. Of 48 Labrador Retrievers being observed, 24 were fed a normal diet and 24 were fed a restricted diet. The dogs' lifetimes were recorded. The study concluded that a typical Labrador Retriever on a restricted diet tends to live longer than a typical Labrador Retriever on a normal diet (Source: Effects of Diet Restriction on Life Span and Age-Related changes in Dogs, Richard D. Kealy, PhD, et al.).

   a. What question were the researchers trying to answer?   b. Identify the population.

   c. Identify the sample.   d. Identify the conclusion. Is it part of descriptive or inferential statistics?

OBJECTIVE 5 – Select a simple random sample.

When a study makes use of a sample, it is important that the sample represents the population well.

Definition: Simple random sample
A process of selecting a sample of size *n* is **simple random sampling**
if every sample of size n has the same chance of being chosen.
A sample selected by such a process is called a **simple random sample**.

- If we allow an individual to be selected more than once, then we are **sampling with replacement**.
- If we do not allow an individual to be selected more than once, then we are **sampling without replacement**.
- A **frame** is a numbered list of all the individuals in the population.

5. a. Create a frame of all the students in your class.

   b. Use technology to randomly select 6 students without replacement.

6. In 2013, 1523 out of 4270 colleges were 2-year colleges (Source: U.S. Department of Education).

   a. What proportion of the colleges were 2-year colleges?

   b. By using the numbers 1 through 1523 to represent the 2-year colleges and the numbers 1524 through 4270 to represent the other colleges, the author used technology to randomly select 500 colleges without replacement. There were 173 2-year colleges in the sample. What proportion of the sample was 2-year colleges?

   c. If you did not know the proportion of the population that are 2-year colleges and used the result you found in Part (b) to estimate it, would that be part of descriptive statistics or inferential statistics? Explain.

Definition: Sampling error
Sampling error is the error involved in using a sample to estimate
information about a population due to randomness in the sample.

7. Find the sampling error for the estimation you made in Problem 6.

If random sampling is used to select a large enough sample, the sampling error will not be too large.

OBJECTIVE 6 – Identify the sampling bias, the nonresponse bias, and the response bias of a study.

Definition: Bias
A sampling method that consistently underemphasizes or overemphasizes
some characteristic(s) of the population is said to be biased.

If a sampling method is biased, then any inferences made will be misleading.

There are three types of bias:

1. **Sampling bias** occurs if the sampling technique favors one group of individuals over another.
2. **Nonresponse bias** happens if individuals refuse to be part of the study or if the researcher cannot track down individuals identified to be in the sample.
3. **Response bias** occurs if surveyed people's answers do not match with what they think.

Guidelines for Constructing Survey Questions

When constructing a survey question:

- Do not include judgmental words.
- Avoid asking a yes/no question.
- If the question includes two or more choices,
  switch the order of the choices for different respondents.
- Address just one issue.

For each study, identify possible forms of bias. Also, discuss sampling error.

8. A liberal radio station conducts a call-in survey, asking callers, "Should Congress do the right thing and increase the budget for education?" Of the 1140 callers, 969 answered yes. The station concludes that 85% of all Americans think that the budget for education should be increased.

9. A group of students surveys other students who enter and exit the gym and finds that 67 students out of 100 students exercise every day. The group concludes that 67% of all students at the college exercise every day.

OBJECTIVE 7 – Identify sampling and non-sampling error.

Definition: Nonsampling error
Nonsampling error is error from using biased sampling,
recording data incorrectly, and analyzing data incorrectly.

10. A hotel randomly selects some customers who stayed at the hotel in the past month, asking whether they had taken clothes hangers and/or towels from their room. The proportion of the sampled customers who said they had taken clothes hangers and/or towels was 0.01. But the population proportion is 0.12. Identify whether sampling error, nonsampling error, or both have likely occurred.

Homework/Assessment
1, 3, 7, 11, 13, 19, 21, 23, 27, 33, 35, 41, 43, 58, 59