Objective 1

- Define statistics and statistical thinking

**Statistics** is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. In addition, statistics provides a measure of confidence in any conclusions.

The information referred to in the definition is **data**. **Data** are “facts or propositions used to draw a conclusion or make a decision.” Data describe characteristics of an individual.

A key aspect of data is that they vary.

Is everyone in your class the same height? No!

Is the number of beers one can drink within an hour and be legal to drive the same for everyone? No!

One goal of statistics is to describe and understand sources of variability.
Objective 2

• Explain the Process of Statistics

The entire group of individuals to be studied is called the population. An individual is a person or object that is a member of the population being studied. A sample is a subset of the population that is being studied.

A parameter is a numerical summary of a population.
Examples: \( \mu, \sigma, p \) (population mean, population standard deviation, population proportion)

A statistic is a numerical summary based on a sample.
Examples: \( \bar{x}, s, \hat{p} \) (sample mean, sample standard deviation, sample proportion)

EXAMPLE  Parameter versus Statistic

Suppose the percentage of all students on your campus who have a job is 84.9%.  This value represents a parameter or a statistic?

Suppose a sample of 250 students is obtained, and from this sample we find that 86.3% have a job.  This value represents a statistic or a parameter?

Descriptive statistics consist of organizing and summarizing data. Descriptive statistics describe data through numerical summaries, tables, and graphs.

Inferential statistics uses methods that take results from a sample, extends them to the population, and measures the reliability of the result.
Objective 3

- Distinguish between Qualitative and Quantitative Variables

**Qualitative or Categorical variables** allow for classification of individuals based on some attribute or characteristic.

**Quantitative variables** provide numerical measures of individuals. Arithmetic operations such as addition and subtraction can be performed on the values of the quantitative variable and provide meaningful results.

EXAMPLE: **Distinguishing between Qualitative and Quantitative Variables**

Researcher Elisabeth Kvaavik and others studied factors that affect the eating habits of adults in their mid-thirties. (Source: Kvaavik E, et. al. Psychological explanations of eating habits among adults in their mid-30’s (2005) International Journal of Behavioral Nutrition and Physical Activity (2):9.) Classify each of the following variables considered in the study as qualitative or quantitative.

- a. Nationality: Qualitative
- b. Number of children: Quantitative
- c. Household income in the previous year: Quantitative
- d. Level of education: Qualitative
- e. Daily intake of whole grains (measured in grams per day): Quantitative

Objective 4

Distinguish between Discrete and Continuous Variables

A **discrete variable** is a quantitative variable that either has a finite number of possible values or a countable number of possible values. The term “countable” means the values result from counting such as 0, 1, 2, 3, and so on.

A **continuous variable** is a quantitative variable that has an infinite number of possible values it can take on and can be measured to any desired level of accuracy.

EXAMPLE: **Distinguishing between Qualitative and Quantitative Variables**

Researcher Elisabeth Kvaavik and others studied factors that affect the eating habits of adults in their mid-thirties. Classify each of the following quantitative variables considered in the study as discrete or continuous.

- a. Number of children: Discrete
- b. Household income in the previous year: Continuous
- c. Daily intake of whole grains (measured in grams per day): Continuous
Objective 5

- Determine the Level of Measurement of a Variable

There are FOUR of them
TWO Qualitative
TWO Quantitative

Qualitative
A variable is at the nominal level of measurement if its values name, label, or categorize so that the naming scheme does not allow for ranking.
- Color, Ethnicity, Gender, Birthplace

A variable is at the ordinal level of measurement if it has the properties of the nominal level of measurement and the naming scheme allows ranking.
- Months of the year, Military Rank, Socio-Economic Status

Quantitative
A variable is at the interval level of measurement if it is numerical, a difference in values has meaning, zero does not mean the absence of, and division of values has no meaning.
- Temperature in Celsius, Birth year

A variable is at the ratio level of measurement if it is numerical, a ratio of values has meaning, and zero means the absence of.
- Income, Number of pets, Weight

EXAMPLE  Determining the Level of Measurement of a Variable

Classify each of the following variables and determine whether the quantitative variables are discrete or continuous.

a. The distance between two cars  
   Ratio, continuous

b. The flavor of a beverage  
   Nominal

c. Class rank (Freshman, Sophomore, Junior, Senior)  
   Ordinal

d. A student’s IQ score  
   Interval, discrete

Section 1.2
Observational Studies Versus Designed Experiments
Objectives

1. Distinguish between an observational study and an experiment
2. Explain the various types of observational studies

Objective 1

Distinguish between Observational Study and Experiment

An observational study measures the value of the response variable without trying to influence the value of the outcome.

Example: Observing how many guys leave with a date from a singles bar.

If a researcher assigns subjects in a study to a certain group, intentionally changes the value of the explanatory variable, and then records the value of the response variable for each group, the researcher is conducting a designed experiment.

Example: (1) Some guys get complete makeover  (2) Some guys given best pick-up lines ever to use (3) Some guys given inflated status with the help of the bartender

Example 1

Cellular Phones and Brain Tumors

Researchers Joachim Schirr and associates wanted "to investigate cancer risk among Danish cellular telephone users who were followed for up to 21 years." To do so, they kept track of 420,095 people whose first cellular telephone prescription was between 1982 and 1995. In 2002, they recorded the number of people out of the 420,095 who had a brain tumor and compared the rate of brain tumors in this group to the rate of brain tumors in the general population. They found no significant difference in the rate of brain tumors between the two groups. The researchers concluded "cellular telephone use was not associated with increased risk for brain tumors." (Source: Joachim Schirr et al. "Cellular Telephone Use and Cancer Risk: Update of a Nationwide Danish Cohort," Journal of the National Cancer Institute 99(2):1707-1713, 2006)
In both studies, the goal of the research was to determine if radio frequencies from cell phones increase the risk of contracting brain tumors. Whether or not brain cancer was contracted is the response variable. The level of cell phone usage is the explanatory variable.

In research, we wish to determine how varying the amount of an explanatory variable affects the value of a response variable.

Confounding/lurking: Erroneous conclusions from a study may occur when the effects of two or more explanatory variables are not separated. A lurking variable is an explanatory variable that was not considered in a study, but that affect the value of the response variable in the study. In addition, lurking variables are typically related to any explanatory variables considered in the study.

Examples: (1) Higher consumption of ice creams results in a higher crime rate. (2) Cleaner cars results in better fuel efficiency.

Objective 2

- Explain the Various Types of Observational Studies

Cross-sectional Studies Observational studies that collect information about individuals at a specific point in time, or over a very short period of time.

Case-control Studies These studies are retrospective, meaning that they require individuals to look back in time or require the researcher to look at existing records. In case-control studies, individuals that have certain characteristics are matched with those that do not.

Cohort Studies A cohort study first identifies a group of individuals to participate in the study (cohort). The cohort is then observed over a period of time. Over this time period, characteristics about the individuals are recorded. Because the data is collected over time, cohort studies are prospective.
A census is a list of all individuals in a population along with certain characteristics of each individual.

Objective

1. Obtain a Simple Random Sample

Random sampling is the process of using chance to select individuals from a population to be included in the sample.
A sample of size \( n \) from a population of size \( N \) is obtained through **simple random sampling** if every possible sample of size \( n \) has an equally likely chance of occurring. The sample is then called a **simple random sample**.

**EXAMPLE Illustrating Simple Random Sampling**

Suppose a study group consists of 5 students: Bob, Patricia, Mike, Jan, and Maria. 2 of the students must go to the board to demonstrate a homework problem. List all possible samples of size 2 (without replacement).

- Bob, Patricia
- Bob, Mike
- Bob, Jan
- Bob, Maria
- Patricia, Mike
- Patricia, Jan
- Patricia, Maria
- Mike, Jan
- Mike, Maria
- Jan, Maria

**Steps for Obtaining a Simple Random Sample**

1) Obtain a frame that lists all the individuals in the population of interest.

2) Number the individuals in the frame 1 - \( N \).

3) Use a random number table, graphing calculator, or statistical software to randomly generate \( n \) numbers where \( n \) is the desired sample size.
Section 1.4

Other Effective Sampling Methods

Objectives

1. Obtain a Stratified Sample
2. Obtain a Systematic Sample
3. Obtain a Cluster Sample

A stratified sample is one obtained by separating the population into homogeneous, non-overlapping groups called strata, and then obtaining a simple random sample from each stratum.

EXAMPLE: Obtaining a Stratified Sample

In 2008, the United States Senate had 49 Republicans, 49 Democrats, and 2 Independents. The president wants to have a luncheon with 4 Republicans, 4 Democrats and 1 Other. Obtain a stratified sample in order to select members who will attend the luncheon.
Objective 2

- Obtain a Systematic Sample

A systematic sample is obtained by selecting every \( k \)th individual from the population. The first individual selected is a random number between 1 and \( k \).

**EXAMPLE Obtaining a Systematic Sample**

A quality control engineer wants to obtain a systematic sample of 25 bottles coming off a filling machine to verify the machine is working properly. Design a sampling technique that can be used to obtain a sample of 25 bottles coming from a shift that fills 1189 bottles.

**STEPS IN SYSTEMATIC SAMPLING, POPULATION SIZE KNOWN**

1. **Step 1:** Determine the population size, \( N \). (\( N=1189 \))
2. **Step 2:** Determine the sample size desired, \( n \). (\( n=25 \))
3. **Step 3:** Compute \( N/n \) and round down to the nearest integer. This value is \( k \). (\( 1189/25=47.56 \); \( k=47 \))
4. **Step 4:** Randomly select a number between 1 and \( k \). Call this number \( p \). (\( p = \text{randint}(1,47) = 23 \))
5. **Step 5:** The \( i \)th sample selected is \( p + (i-1)k \)

\[ i=1: 23 \hspace{1cm} i=2: 70 \hspace{1cm} i=3: 117 \hspace{1cm} i=4: 164 \ldots \]

\[ i=n=25: 1151 \]
Objective 3

• Obtain a Cluster Sample

A cluster sample is obtained by selecting all individuals within a randomly selected collection or group of individuals.

EXAMPLE Obtaining a Cluster Sample

A school administrator wants to obtain a sample of students in order to conduct a survey. She randomly selects 10 classes and administers the survey to all the students in the class.
A **convenience sample** is one in which the individuals in the sample are easily obtained.

Any studies that use this type of sampling generally have results that are suspect. Results should be looked upon with extreme skepticism.

**Objective**

1. Explain the Sources of Bias in Sampling

**Sampling bias** means that the technique used to obtain the individuals to be in the sample tend to favor one part of the population over another.

**Undercoverage** is a type of sampling bias. **Undercoverage** occurs when the proportion of one segment of the population is lower in a sample than it is in the population.
An experiment is a controlled study conducted to determine the effect of varying one or more explanatory variables or factors has on a response variable. Any combination of the values of the factors is called a treatment.

The experimental unit (or subject) is a person, object or some other well-defined item upon which a treatment is applied.

A control group serves as a baseline treatment that can be used to compare to other treatments.

A placebo is an innocuous medication, such as a sugar tablet, that looks, tastes, and smells like the experimental medication.

Blinding refers to nondisclosure of the treatment an experimental unit is receiving.

A single-blind experiment is one in which the experimental unit (or subject) does not know which treatment he or she is receiving.

A double-blind experiment is one in which neither the experimental unit nor the researcher in contact with the experimental unit knows which treatment the experimental unit is receiving.

The English Department of a community college is considering adopting an online version of the freshman English course. To compare the new online course to the traditional course, an English Department faculty member randomly splits a section of her course. Half of the students receive the traditional course and the other half is given an online version. At the end of the semester, both groups will be given a test to determine which performed better.

EXAMPLE The Characteristics of an Experiment

(a) Who are the experimental units? The students in the class
(b) What is the population for which this study applies? All students who enroll in the class
(c) What are the treatments? Traditional vs. online instruction
(d) What is the response variable? Exam score
(e) Why can’t this experiment be conducted with blinding? Both the students and instructor know which treatment they are receiving
EXAMPLE  A Matched-Pairs Design

Xylitol has proven effective in preventing dental caries (cavities) when included in food or gum. A total of 75 Peruvian children were given milk with and without Xylitol and were asked to evaluate the taste of each. The researchers measured the children’s’ ratings of the two types of milk. *(Source: Castillo JL, et al (2005) Children’s acceptance of milk with Xylitol or Sorbitol for dental caries prevention. BMC Oral Health (5)6.)*

(a) What is the response variable in this experiment? Rating
(b) Think of some of the factors in the study. Which are controlled? Which factor is manipulated?
   Age and gender of the children; Milk with and without Xylitol is the factor that was manipulated
(c) What are the treatments? How many treatments are there?
   Milk with Xylitol and milk without xylitol; 2
(d) Identify the experimental units. 75 Peruvian children
(e) Why would it be a good idea to randomly assign whether the child drinks the milk with Xylitol first or second?
   Remove any effect due to order in which milk is drunk.
(f) Do you think it would be a good idea to double-blind this experiment? Yes!