

Unit 1

1.1.1: Textbook Reading

- **Two Types of Data:**
 - **Qualitative Variables:** Measures a *quality/characteristic*; also called *categorical data* because it can be split into categories
 - **Quantitative variables:** Measures a numerical *quantity*. There are two types of Quantitative variables: **Discrete** and **Continuous**.
 - * **Discrete Variables:** Finite/countable value or number of values
 - * **Continuous Variables:** Infinite number of values (ex. on a line)

1.1.2: Errors in Ads (and Other Claims)

Examples of Errors in Ads

1. “Our new car sales are up 25% over the last year.” How much were the car sales last year? Does it just look like this on paper?
2. “The death rate is lowest in cold-weather states, such as Connecticut and Minnesota.” What other factors are affecting the death rate? (Are there more old people?)
3. “MoreHair is a revolutionary new product that stops hair loss in two-thirds of men with certain types of hair loss.” “Certain types of hair loss?” What if that “certain type” is just some super rare type of hair loss that a lot of other hair products can stop? Also, how are they measuring the population of men? Is the sample size balanced, or do they look at more older men than younger men?
4. “Our local forecast shows the chance of rain at 30 percent.” What do you mean, “30 percent”? 30 percent of the area? Where was this forecast made?

1.1.3 Study: Welcome to Statistics

Descriptive vs Inferential Statistics

- Descriptive statistics is used to describe/summarize things known to you
 - Example: statistics about SAT scores at a high school you go to
- Inferential statistics does more than descriptive statistics; you compare groups of **descriptive** statistics and make inferences from that – in other words, make a hypothesis.
 - Example: Randomly pick and drive 50 of a sample size of cars to find out if a claim that says the mileage is 30 mpg is true

The Phases of a Statistical Study

- The three phases of a statistical study are:

- Data Gathering
- Data organization and analysis
- Probability-based inference
- Note that the first two are descriptive statistics and the third one is inferential.
- **Data gathering** is any process that gets you data; questionnaires, counting, computer programs to reorganize old data; this is the most important part to get right.
- **Data organization** is the process of making tables/graphs/lists from the data. You also calculate statistics in this phase like average and standard deviation and quartiles.
 - **Data Analysis** is where you look for patterns in the statistics.
 - * **Exploratory Data Analysis** is when a study doesn't try to answer a question
- **Probability-based Inference** is when you make a prediction based on probability (WOW!)

Data

- **Data** is defined as a bunch of bits of information (a bit is a datum); however, we can define a data set as singular, a bigger unit of datum(s?)
- **Numeric Data** can be sorted and can be manipulated mathematically
 - **Continuous Data** is data that can be measured; for example, how much you weigh, height, width, etc.
 - **Discrete Data** is data that can be counted; for example, the number of brothers in a family. There can be no value in between the different pieces of data
- **Categorical Data** is non-numeric data that can't be manipulated mathematically

1.2.1: Textbook Reading/Quiz (?)

- **Variable:** Characteristic that changes/varies over time for different stuff; examples include body temperature, religion, nationality, height, age
- **Experimental Unit:** Object on which the variable is actually measured
- **Measurement:** The measure of the actual variable
- **Population:** Set of all important measurements
- **Sample:** Subset of the population

1.2.4: Introduction to Frequency Data and Their Graphs

- **Frequency Tables** show how frequently a value occurs; value is one column, frequency is other.
 - You can also do a meta-frequency table; values can be ranges of values as well. These ranges of values are called *classes*. You have to find the “Goldilocks” class width; not too small (too much data) and not

to big (too little data)

- A **Histogram** graphs these meta-frequency tables. **HOW TO DRAW A HISTOGRAM**
 1. Draw a horizontal axis with intervals equally spaced
 2. Draw a vertical axis that shows frequencies at least up to 15
 3. Draw the bars and add labels to the axes
- Relative frequencies can be calculated by dividing the frequency of one value by the total
-