**Chapter 1: Statistics, Data, and Statistical Thinking**

**This Chapter is filled with useful vocabulary words, which makes this chapter VERY helpful but not THAT interesting.**

**Bear with me statistics will be more than vocabulary words ☺**

The Science of Statistics

* Statistics is a SCIENCE (not really a Math).
* Your book defines Statistics as the science of data (collecting, classifying, summarizing, organizing, analyzing, and interpreting numerical information)

Types of Statistical Applications

* **Descriptive Statistics** – numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data set, and to present the information in a convenient form
	+ Batting averages
	+ Daily rainfall
	+ Graph of sample survey (like on Facebook)
* **Inferential Statistics** – using sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data
	+ Inferring political winner from sample polling
	+ Estimating average SAT score from sample

Fundamental Elements of Statistics

* **Experimental Unit** – an object upon which we collect data
* **Variable** – characteristic or property of an individual experimental unit
* **Population** – a set of units that we are interested in studying
* **Sample** – a subset of the units of a population
* **Census** – measure a variable for every experimental unit in the population (not usually done)

Population versus Sample

* This process can be repeated multiple times to get multiple samples

Population

* + - Parameter
			* Usually greek letters
			(α, β, ρ, μ)
			* Usually unknown

Sample

* + - Statistic
			* Usually alphabetic letters
			(x, y, z, p)
			* Can be found from data then use to infer about population

**Example – Identify the experimental units, population, variable, and sample of each of the following:**

* **A politician who is running for the office of mayor of a city with 25,000 registered voters commissions a survey. In the survey, 48% of the 200 registered voters interviewed say they plan to vote for her.**

**Experimental units –**

**Population –**

**Variable –**

**Sample –**

* **A manufacturer of computer chips claims that less than 10% of his products are defective. When 1,000 chips were drawn from a large production run, 7.5% were found to be defective.**

**Experimental units –**

**Population –**

**Variable –**

**Sample –**

Four Elements of Descriptive Statistical Problems

1. Identify the population or sample of interest
2. Identify one or more variables that are to be investigated
3. Use Tables, graphs, or numerical summary tools (we will learn these in Chapter 2)
4. Identify patterns in data

Five Elements of Inferential Statistical Problems

1. Identify the population of interest
2. Identify one or more variables that are to be investigated
3. Sample population units
4. Make Inference about the population based on information contained in the sample (we will learn in Chapters 5, 6, and 10)
5. Give a measure of reliability (a statement [usually quantified] about the degree of uncertainty associated with a statistical inference) for the inference

Types of Data

* **Quantitative** – measurements that are recorded on a naturally occurring numerical scale
* **Qualitative** – measurements that cannot be measured on a natural numerical scale; they can only be classified into one of a group of categories

**Example – Identify the following as either qualitative or quantitative data**

* **ACT scores**
* **Number of hunting accidents in a state during deer season**
* **Religious Affiliation**
* **Time spent studying for first statistics exam**
* **Lifetimes of electric light bulbs**
* **Number of cars ticketed for illegal parking on campus today**
* **Languages of United States**

**Example - Recently the manager of the call center for a large Internet bank asked his staff to collect data on a random sample of the bank’s customers. Data on the following variables were collected and placed in a file…**

****

* **Is Account Number Qualitative or Quantitative?**
* **Is Caller Gender Qualitative or Quantitative?**
* **Is Account Holder Gender Qualitative or Quantitative?**
* **Is Past Due Amount Qualitative or Quantitative?**
* **Is Current Amount Due Qualitative or Quantitative?**
* **Is “Was This a Billing Question” Qualitative or Quantitative?**

Collecting Data

* Published Sources
	+ Census of US
	+ Online Collections of Data (GSS, Sports information)
	+ Newspapers
* Designed Experiments
	+ Medical study on a particular drug
	+ Which is the best fertilizer
	+ Taste tests (Pepsi vs. Coke)
* Survey
	+ Polling
	+ Online surveys (see next slide)
	+ Nielson surveys
* Observational Study
	+ Closing stock prices
	+ Profit trends

\*\*Caution: When collecting data it is important that you make sure the data was collected from a **representative sample**, that is a sample that is typical of the population of interest. In particular online data or polling can be unreliable data because it is limited to those with access to a computer and usually with volunteer responses you only get people participating that feel strongly about the topic at hand.

Representative Sample

* One way to obtain this representative sample is to take a **random sample**

**Random Sample** – is a sample selected from the population in such a way that every different sample of size n has an equal chance of selection (we will learn how to do this in Chapter 3)

**\*\*Note:** This is equivalent to putting all names in a hat and drawing n number of names out to be in the sample.

Types of Bias and Errors – The Role of Statistics in Managerial Decision Making

* When a collection of data is not a representative sample it is often the result of **selection bias**

**Selection Bias**- when a subset of the experimental units in the population is excluded so that these units have no chance of being selected for the sample (Ex. Online polls , is a selection bias against those people without computers)

* When conducting a survey it is often times the case that people will not respond. If a large number of people (relative to the number of people asked) do not respond to the survey, this is known as **nonresponse bias.**

**Nonresponse bias**- when the researchers conducting a survey or study are unable to obtain data on all experimental units selected for the sample

* It is important when preparing a survey that the questions not be leading (encouraging an answer in particular direction) or misleading (confusing), this would result in **measurement error.**

**Measurement error**- inaccuracies in the values of the data recorded.

**Example – A bank with branches in a large metropolitan area is considering opening its offices on Saturday, but it is uncertain whether customers will prefer (1) having walk-in hours on Saturday or (2) having extended branch hours during the week. Listed below are some of the ideas proposed for gathering data. For each, indicate what (if any)**

* **Put a big ad in the newspaper asking people to log their opinions on the bank’s Web site.**
* **Randomly select one of the branches and contact every customer at that bank by phone.**
* **Send a survey to every customer’s home, and ask the customer to fill it out and return it.**
* **Randomly select 20 customers from each branch. Send each a survey, and follow up with a phone call if he or she does not return the survey within a week.**