STAT 110: Introduction to Descriptive Statistics

- ***Statistics:*** **The science of data.**
- **We gain insight and make conclusions based on data.**

Principles We Will Study:

1. **What is a good design of an experiment?**
2. **How should samples be gathered?**
3. **What are the best ways to summarize data graphically?**
4. **What are the best ways to summarize data using numbers?**
5. **What are common statistical errors to recognize and to avoid?**
6. **What conclusions can we make about a situation using data from a sample?**

Why Do We Need Data?

- **Suppose we want to know the most popular TV shows?**
- **What shows do you watch? What shows does your circle of friends watch?**
- **What shows do people always seem to be talking about around town?**
- **Will answering these questions give us correct answers of the viewership of TV shows?**
- **Advertisers need *accurate* and *precise* answers to make business decisions**

Data Versus Anecdotes

- **Question: Does living near power lines cause cancer?**
- ***TV report*: Televised interview with local mother of a child who has leukemia**
- **Family lives 150 feet from power lines!**
- ***Scientific Study*: 5-year, $5 million study by NCI**
- **Found no link between leukemia and power lines**
- **Which conclusion is more likely to be accurate?**
- **Which report is more likely to be make an impact in people's minds?**

Data Versus Anecdotes, Example 2

- **Question: Is Columbia a safe place to live?**
- ***TV news*: What stories lead off the news most nights?**
- ***Scientific Study*: Comparative crime rates for Columbia and other cities of similar size**
- **Which conclusion is more likely to be accurate?**
- **Which report is more likely to be make an impact in people's minds?**

Where Do Data Come From?

- **Example 1: National unemployment rate**
- **Example 2: Internet poll about social networking usage**
- **Example 3: Political opinion polls**

Individuals and Variables

- ***Individuals:*** **Objects (could be people or things) described by a data set**
- ***Variable:*** **A characteristic of an individual that could be measured or observed.**
- **When we measure one or more variables for a set of individuals, we have a data set.**

"Practice" Clicker Quiz 1

The Census Bureau gathers information about U.S. residents every ten years. What are the individuals in the resulting data set?

A. Census bureau workers

B. U.S. households

C. U.S. residents

D. Years

"Practice" Clicker Quiz 2

The Census Bureau gathers information about U.S. residents every ten years. What is most likely a variable in the resulting data set?

A. Hispanic

B. Race

C. Caucasian

D. Year

Types of Variables

- **Data sets typically contain lots of variables.**
- **Some variables are *numerical*, others *categorical*.**

Table 1: Selected golfers at the 2015 Masters tournament.

Name	Country	Round 1 score	Round 2 score	Round 3 score	Round 4 score	Total score	Earnings
Jordan Spieth	United States	64	66	70	70	270	$1,800,000
Phil Mickelson	United States	70	68	67	69	274	$880,000
Justin Rose	England	67	70	67	70	274	$880,000
Rory McIlroy	Northern Ireland	71	71	68	66	276	$480,000
Hideki Matsuyama	Japan	71	70	70	66	277	$400,000

- ***Numerical variables:*** **Round 1 score, Round 2 score, Round 3 score, Round 4 score, Total score, Earnings**
- ***Categorical variable:*** **Country**
- **Which type of variable can we perform mathematical calculations (sums, averages) on?**
- **Example analysis: Could calculate average earnings, separately for each country.**
- ***Response variable:*** **A variable that measures the major outcome of interest in a study.**

Types of Studies

- ***Observational study:*** **Researcher measures a response variable *without* attempting to influence or control the response.**
- ***Experiment:*** **Researcher intentionally applies some *treatments* to *subjects* and then measures a response variable. (Question: How does the treatment affect the response?)**

"Practice" Clicker Quiz 3

In 1954, Dr. Jonas Salk gave some children a trial vaccine and gave other children a *placebo*. After a period of time, he then recorded whether the children in the study contracted polio.

A. Observational Study

B. Experiment

C. Both

"Practice" Clicker Quiz 4

Researchers measure birth weights of infants and record whether the mothers consumed alcohol during the pregnancy, to study whether drinking alcohol affects infant birth weight.

A. Observational Study

B. Experiment

C. Both

Populations and Samples

- **In statistics, the *population* is the entire set of *individuals* about which we want to make conclusions.**
- **Usually a VERY LARGE number of individuals – we can't observe all of them!**
- **Instead, we select a *sample* – a part of the population for which we actually collect data.**
- ***If the sample is well-chosen*, the data from the sample can help us make conclusions about the population.**

More Types of Studies

- ***Sample Survey*: Researchers carefully select members of a population of interest and measure their responses to various questions.**
- **The survey questions are the *variables* in the sample survey.**
- ***Census*: Similar to the sample survey, but with a census the researchers (attempt to) collect data on the entire population.**
- **A census is much more difficult and expensive to carry out (relatively rare).**

Example of a Sample Survey

- ***General Social Survey*: Done every 2 years by National Opinion Research Center (Univ. of Chicago)**
- **Variables are background characteristics (race, family history, personal habits) and opinion questions about social issues.**
- ***Population of interest*: U.S. adults living in households. (What people are left out?)**
- ***Sample*: 3000 adults personally interviewed. (Is this sample size large enough?)**