**Psy 230 Stats/Methods I**

**Intro to Statistics**

**Chapter 1 (G&W)**

**Statistics** means “statistical procedures*”*

* Uses of statistics
  + Organize and summarize information
  + Determine exactly what general conclusions are justified based on the specific results that were obtained
* Goals of statistical procedures
  + Accurate and meaningful interpretation
  + Standardized evaluation procedures

**Variables**

A **variable** is a characteristic or condition that can change or take on different values.

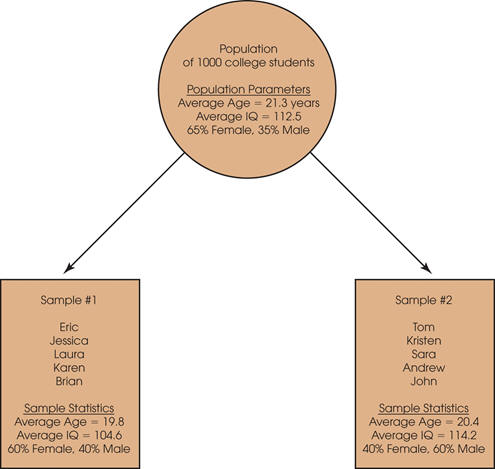
Most research begins with a general question about the relationship between two variables for a specific group of individuals.

**Populations and Samples**

The entire group of individuals is called the **population**.

For example, a researcher may be interested in the relation between class size (variable 1) and academic performance (variable 2) for the population of third-grade children.

Usually populations are so large that a researcher cannot examine the entire group. Therefore, a **sample** is selected to represent the population in a research study. The goal is to use the results obtained from the sample to help answer questions about the population.



**The Relationship between a Population and a Sample**

**A cyclic diagram shows four stages of the relationship between a population and a sample.
- The population: All of the individuals of interest
- The sample: is selected from the population.
- The sample: The individuals selected to participate in the research study.
- The results from the sample are generalized to the population.**

**Parameters and Statistics**

**Descriptive statistics** are methods for organizing and summarizing data.

When describing data with descriptive statistics, it is necessary to distinguish whether the data come from a population or a sample.

Typically, every population parameter has a corresponding sample statistic.

- Parameter—a descriptive value that describes a population

- Statistic—a descriptive value that describes a sample

**Descriptive Statistics vs. Inferential Statistics**

Descriptive Stats

techniques used to summarize, organize, and simplify data  
can't look at it all - get a quick, good impression

Inferential Stats

techniques used to study samples and then make generalizations about the populations from which they were selected. Use sample statistics to make inferences about the corresponding population parameters.

A drawback?

**Sampling Error**  
The discrepancy between a sample statistic and its population parameter is called **sampling error**.

Defining and measuring sampling error is a large part of inferential statistics.

**Observations, Measurement, and Variables**

* Science is **empirical**—it is based on observation
* The scores that make up the data from a research study are obtained by observing and measuring variables
* The process of measurement consists of applying carefully defined measurement procedures for each variable

**Constructs & Operational Definitions**

* **Constructs** 
  + Internal attributes or characteristics that cannot be directly observed
  + Useful for describing and explaining behavior
* **Operational definition** 
  + Identifies the set of operations for measuring an external (observable) behavior
  + Uses the resulting measurements as both a definition ***and*** a measurement of a hypothetical construct

**Variables  
discrete** - separate categories. No values can exist between two neighboring categories (e.g., dice)

**continuous** - infinite fineness. There are an infinite number of possible values that fall between any two observed values

* For example, time can be measured to the nearest minute, second, half-second, etc.

- each score corresponds to an interval of the scale

- the boundaries that separate these intervals are called real limits

**Measuring Variables**  
To establish relationships between variables, researchers must observe the variables and record their observations. This requires that the variables be **measured**.

**Scales of Measurement**The process of measuring a variable requires a set of categories called a scale of measurement and a process that classifies each individual into one category.

|  |  |  |
| --- | --- | --- |
| **Scale** | **Characteristics** | **Examples** |
| **Nominal** | * Label and categorize * No quantitative distinctions | * Gender * Diagnosis * Experimental or Control |
| **Ordinal** | * Categorizes observations * Categories organized by size or magnitude | * Rank in class * Clothing sizes (S, M, L, XL) * Olympic medals |
| **Interval** | * Ordered categories * Interval between categories  of equal size * Arbitrary or absent zero point | * Temperature * IQ * Golf scores (above/below par) |
| **Ratio** | * Ordered categories * Equal interval between categories * Absolute zero point | * Number of correct answers * Time to complete task * Gain in height and/or weight since last year |

**Three Data Structures  
Data structure 1: descriptive research (individual variables)**

* + One (or more) variables measured per individual
  + *Statistics* describe the observed variable
  + May use category and/or numerical variables
  + Not concerned with relationships between variables

Relationships between Variables

* + Two (or more) variables are observed and measured in order to determine a relationship
  + The resulting measurements can be classified into two distinct data structures that are used to determine what type of relationship exists

**Data structure 2: the correlational method**

* + One group of participants
  + Measurement of two variables for each participant
  + The goal is to describe type and magnitude of the relationship

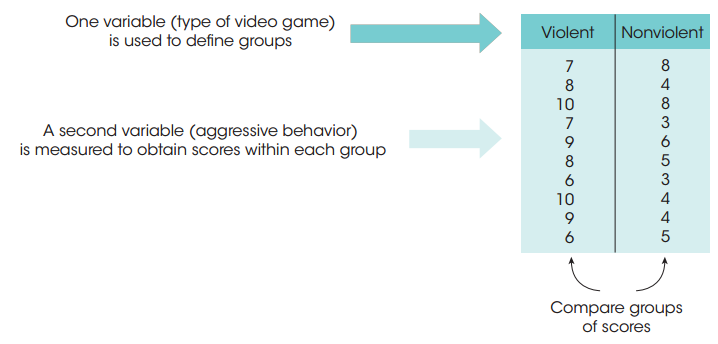
A table and a scatterplot show two data structures for studies evaluating the relationship between variables.
The table includes eight rows and three columns. The column headers are as follow: student, Facebook time, and academic performance. The row entries are as follows: A, 4, and 2.4; B, 2, and 3.6; C, 2, and 3.2; D, 5, and 2.2; E, 0, and 3.8; F, 3, and 2.2; G, 3, and 3.0; H, 1, and 3.0.
In the scatterplot, the horizontal axis represents Facebook time (0 equals least, 5 equals most) ranging from 0 to 5, in increments of 1. The vertical axis represents academic performance ranging from 2.0 to 3.8, in increments of 0.2. The data points on the graph reflect the data in the table.

What are the Limitations of the Correlational Method?

**Data structure 3: experimental and nonexperimental methods**

* + Comparing two (or more) groups of scores
  + One variable defines the groups
  + Second variable is the score, the measurement
  + Both experimental and nonexperimental studies use this structure

Structure 3: Comparing Two (or More) Groups of Scores. Experimental and Nonexperimental Methods



**The Experimental Method**

The goal of an **experiment** is to demonstrate a cause-and-effect relationship between two variables; that is, to show that changing the value of one variable causes changes to occur in a second variable.

In an **experiment**, one variable is manipulated to create treatment conditions. A second variable is observed and measured to obtain scores for a group of individuals in each of the treatment conditions.

The measurements are then compared to see if there are differences between treatment conditions. All other variables are controlled to prevent them from influencing the results.

The manipulated variable is called the independent variable and the observed variable is the dependent variable.

4 characteristics of true experiments:

1 – MANIPULATION

2 – MEASUREMENT

3 –COMPARISON

4 – CONTROL

**Independent variable:** the variable that is manipulated by the researcher

* + Independent because no other variable in the study influences its value; is manipulated *prior* to observing the dependent variable

**Dependent variable:** the one that is observed to assess the effect of treatment

* + Dependent because its value is thought to depend on the value of the independent variable

**Control Conditions in an Experiment**

* **Methods of control**
  + Random assignment of subjects
  + Matching of subjects
  + Holding the level of some potentially influential variables constant
* **Control condition** 
  + Individuals do not receive the experimental treatment
  + They either receive no treatment or they receive a neutral, placebo treatment
  + Purpose: to provide a baseline for comparison with the experimental condition
* **Experimental condition** 
  + Individuals do receive the experimental treatment

**Nonexperimental Methods: Nonequivalent Groups and Pre-Post Studies**

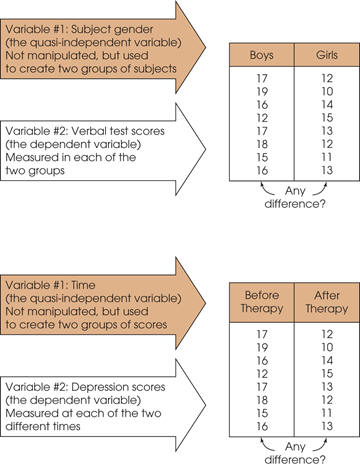
* Nonequivalent groups
  + Researcher compares groups of scores
  + Researcher cannot control who goes into which group
* Pretest/posttest
  + Individuals measured at two points in time
  + Researcher cannot control the influence of the passage of time
* Independent variable is **quasi-independent**

**Now try the Experimental Design Exercise**

**Difference between an experiment and quasi-experiment?**

Quasi-experiments aim to establish a *tentative* cause and effect relationship between two variables but cannot satisfy all of the strict requirements needed for a true experiment (often cannot not meet all of the above requirements in a natural settings). Introduce some treatment or manipulation.

Uses *some* of the rigor and control used in true experiments. But in some way, lack the control found in true experiments (usually lack random assignment of participants to conditions), so ability to draw a causal inference is impaired



**See G&W Appendix A for a math review--**I'm assuming you possess these basic skills...If not, you need to develop them before taking this class. **You must know and be VERY comfortable with the order of operations!!**

P E M D A S--**Order of Operations**

1. All calculations within parentheses are done first.

2. Squaring or raising to other exponents is done second.

3. Multiplying, and dividing are done third, and should be completed in order from left to right.

4. Summation with the Σ notation is done next.

5. Any additional adding and subtracting is done last and should be completed in order from left to right.  
  
**Statistical Notation**

•The individual measurements or scores obtained for a research participant will be identified by the letter X (or X and Y if there are multiple scores for each individual).

•The number of scores in a data set will be identified by N for a population or n for a sample.

•Summing a set of values is a common operation in statistics and has its own notation.  The Greek letter sigma, Σ, will be used to stand for "the sum of."  For example, ΣX identifies the sum of the scores.