

1. **Statistics:** The science of collecting, analyzing, and interpreting data.
2. **Data:** The numbers or information collected in a study or experiment (singular: datum).
3. **Variable:** Any characteristics of a person or thing that can be expressed as a number.
  
4. **Quantitative data** (aka **numerical**): Numerical values for which arithmetic operations make sense (you **can average**).
5. **Qualitative data** (aka **categorical**): **COUNT** data. Recording which **category** a person or thing falls into.
  
6. **Questions to ask about data**
  - a. \_\_\_\_\_
  - b. \_\_\_\_\_
  - c. \_\_\_\_\_
  - d. \_\_\_\_\_
  
7. **Descriptive statistics:** Methods of organizing, displaying, and describing data using tables, graphs, and summary measures.
8. **Inferential statistics:** Using collected data to make generalizations about an entire group. Produces answers to specific questions and a statement of how confident we can be that the answer is correct.
  - a. **\*NOTE\***: We cannot predict with absolute certainty! We can't make predictions about individual cases.
9. **Census:** An attempt to include the entire population.
  
10. **Population:** The entire group to be studied (Population of interest)
11. **Parameter:** A numerical measurement describing some characteristics of a **Population**.
  
12. **Sample:** The part of a population that is actually examined.
13. **Statistic:** A numerical measurement describing some characteristic of a **Sample**.
  
14. **Discrete Data:** Values are only isolated points on a number line (Thing integer values ex. {1,2,3,4,...}).
15. **Continuous Data:** Values are from an entire interval on the number line, measurements of time or lengths (Think of heights between/that range from 60 to 100 inches... a continuous value can be any measurement in the range).
  
16. **Bias:** When a study systematically favors certain outcomes
  - a. Response
  - b. Nonresponse
  - c. Voluntary Response
  - d. Undercoverage
  - e. Selection
  - f. Measurement
  - g. Wording of the Questions
  - h. Convenience Sampling

17. **\*\*Randomization\*\***: The use of chance to assign subjects to different treatments **TO REDUCE BIAS** (or randomly select subjects for samples) Here's your mantra: randomization reduces bias, randomization reduces bias, **randomization reduces bias... Memorize this!!!**

### **RANDOM SAMPLING VS RANDOM ASSIGNMENT (UNDERSTAND THE DIFFERENCE!):**

#### **RANDOM SAMPLE**

(of subjects)

Allows us to GENERALIZE our results to the larger population from which we sampled

(very rare in an experiment)

vs

#### **RANDOM ASSIGNMENT**

(of subjects to treatments)

Allows us to draw CAUSAL conclusions (cause and effect) between the explanatory and response variables

### **18. Types of Sampling Methods: Goal is to get a sample that is REPRESENTATIVE of the population.**

- a. **Simple Random Sample (SRS)**: Every unit/thing/person/subject and **every possible sample of size  $n$  are possible**
- b. **Stratified Random Sample**: The population is divided into groups (strata) with similar characteristics, then an SRS is selected from each group (Conducted to ensure each group is represented in the sample... **TO ENSURE A REPRESENTATIVE SAMPLE**)
- c. **Cluster Sample**: Population is generally already separated into parts (Think each classroom at WHS is its own premade/naturally occurring cluster), randomly select clusters. Randomly selected clusters combine (we take a census of each randomly selected cluster) to make the sample.
- d. **Multistage Sample**: Any method that incorporates randomness at MORE THAN ONE stage.  
*Common example: Form clusters, taken a random selection of clusters... and THEN take an SRS from within each of the randomly selected clusters (instead of taking a census of each cluster) to form the sample.*
- e. **Voluntary Response Sample**: only those who choose to respond are evaluated; biased
- f. **Convenience Sample**: Sample that includes the individuals or experimental units that are readily available, not a great plan (because convenience samples are usually biased), but much easier than SRS.

19. **Observational Study: No treatment is imposed**; researchers merely observe a characteristic(s)... LOTS of observational studies are used to collect and analyze data!

20. **Experimental Study: Treatment is imposed**. (or Treatments are imposed if more than one) Uses a deliberate treatment to observe a response and measure the effect of the treatment(s) and/or control.

**Memorize!!!**

21. **Treatment**: The specific experimental process applied to each case/unit/subject

22. **Experimental Units**: Objects on which the experiment is performed (received treatments and/or control)

23. **Subjects**: Human experimental units

24. **Confounding (variables)**: When one or many variables confuse the ability to determine the cause of a characteristic. To overcome confounding it is best to conduct a well-designed, well controlled experiment.

- a. Randomization helps overcome confounding because subjects/experimental units are randomly assigned to treatment(s) and/or control groups. An approximately equal number of

subjects/experimental units are assigned to each of the groups... As a result, the confounding variable(s) is no longer a concern with respect to measuring the effect of the treatment(s).

25. **Lurking (extraneous) variable:** A variable that has an important effect on the response, but is not included among the explanatory variables studied. Most statisticians refer to this as a confounding variable.
26. **EXplanatory Variable:** Used to explain the observed outcome (most often the **X** variable).
27. **Response Variable:** A variable used to measure the outcome of a study (most often the **Y** variable in bivariate data).
28. **Treatment Group(s):** The group of subjects/experimental units that receive the treatment(s).
29. **Control/Placebo Group:** The group of subject/experimental units that receive no treatment (control group) or a fake treatment (placebo group) which actually has no effect.
30. **Placebo:** A dummy/fake/sham treatment such as a odorless, tasteless, sugar pill (which has little to no impact physiologically) when testing, for example, a new weight loss pill.
31. **\*\*Experimental Design\*\*: Control, Randomization, Replication... CONTROL, RANDOMIZATION, REPLICATION**

**Memorize these!!!!**

- a. Control: \_\_\_\_\_
- b. Randomization: \_\_\_\_\_
- c. Replication: \_\_\_\_\_

**32. Types of Experimental Designs**

- a. **Completely Randomized Experiment:** Subjects are randomly assigned to treatment(s) and control (if there is a control group), treatments are imposed, response variable is measured for each group, response variable is compared for each group.
- b. **Randomized Block Design:** A Block (**\*\*\*which is not a stratified random sample!!!!\*\*\*)** is a group that is known to be similar before the experiment in some way that is **expected to affect the response variable** to the treatment. **Blocking reduces variation** within the treatments so we can compare the variation between the treatments. Blocking separates/sorts the units to reduce variation (and is conducted intentionally by the researcher).
  - i. **Matched Pairs Design or just Paired Design** is a special type of block design where subjects/experimental units are paired based on a similar characteristic OR each subject acts as her own block (with a before and after test OR all treatments are randomly assigned to one individual). Used if we need to control variation due to individual differences.
33. **Blinding (may be used in any type of experiment if possible...which we can usually figure out a way to blind)**
  - a. **Blind Experiment:** The subject does not know which treatment s/he is receiving.
  - b. **Double Blind Experiment:** The subject **nor the evaluator** (person recording the response variable) know which treatment the subject received.
34. **Statistically Significant:** A difference too large to be attributed to chance (an outcome that occurs has an extremely small probability of occurring under the assumed conditions... An outcome with a probability of .50 would occur half the time **by chance**. An outcome with a probability of .00124 would only occur about 1 out of

1000 times and we could not attribute its occurrence to chance... Something else is at work if that outcome shows up!)

### 35. **Experimental Design Example:**

The 6 steps of Data Analysis

1. Understand the nature of the problem
2. Decide what to measure and how to measure it
3. Collect the data
4. Data summarization and preliminary analysis
5. Formal data analysis
6. Interpretation and generalization (if possible) of results

### **Vocab from textbook**

Retrospective – an observational study in which subjects are selected and their previous conditions or behaviors are studied

Prospective – an observational study in which subjects are followed (usually a longitudinal study.. The Minnesota Twin Family Study:MTFS)

Sampling Frame – a list of individuals from which the sample is drawn (basically defining who or what you're drawing your sample from)

Additional space for notes and or examples: