

## **INTRO**

- Sampling error:** discrepancy between the sample statistics and the population parameter is sampling error
- Data:** collection of numerical observations from a survey/experiment (single datum is called raw score)

### **Quantitative vs. Qualitative Data**

- Qualitative data:** (always discrete data) single observation, which represents a class or category
- Quantitative Data :** (not always continuous) A single observation is an amount or a count

### **Discrete vs. Continuous Data**

- Discrete:** countable number of possible values (ex. # cars in parking lot)
- Continuous:** infinite number of possible values on a scale in which there are no gaps or interruptions

### **Experimental Design**

\*\*\*Importance in Random Assignment

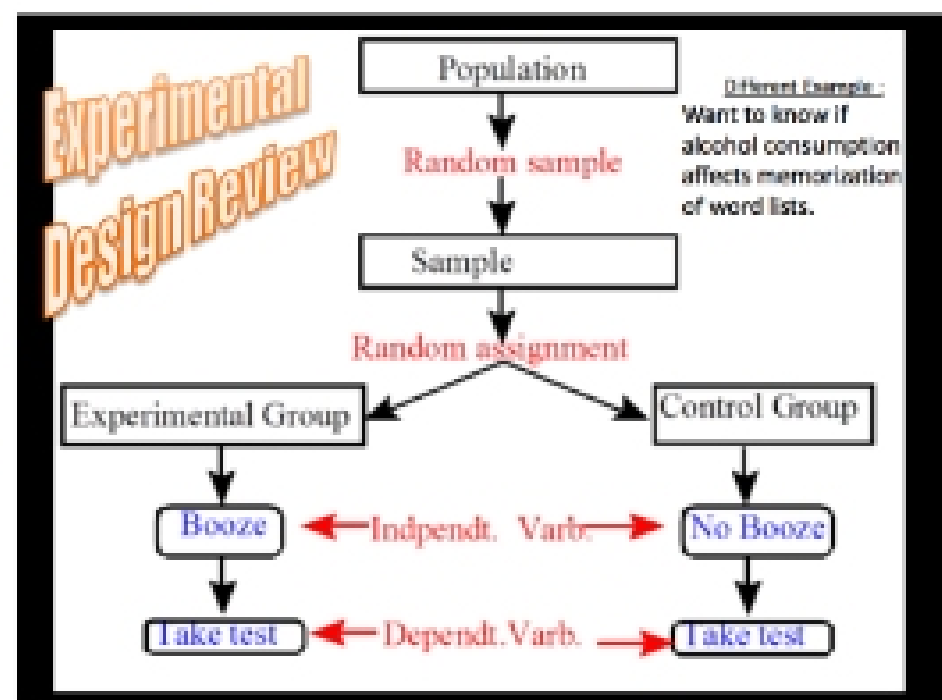
- Variable:** A characteristic or property of organisms, events, or objects that can take on different values
- Constant:** characteristic/property that doesn't change
- Independent Variable:** manipulated by the investigator
- Dependent variable:** variable measured by the investigator

### **Quasi-experimental (non experimental)**

- cases where researcher has no control of group assignment
- inability to rigorously control for extraneous variables

### **Correlational Study**

- investigator measures two DV's and looks for relationship
- REMEMBER:** correlation doesn't determine causation



### **Scales of Measurement**

1. **Nominal:** refers to data that consists of names, labels, categories
2. **Ordinal:** refers to data or scores that can be arranged in some order (give us rank order. ex. 1<sup>st</sup>, 20<sup>th</sup>, etc.)

3. Interval: refers to data that have meaningful differences between scores. Tells us identity and rank order

- Equidistant Scale: intervals of values distributed in equal units
- No true zero: doesn't mean "absence" (such as 0 degrees F isn't the absence of heat, just less)

4. Ratio: refers to data on a scale with a true zero point

- All properties of nominal, ordinal, and intervals scale (name + order + intervals + true 0)
- Is an interval scale with a true zero (complete absence of data being measured)

Summation sign ( $\Sigma$  sigma)

$\Sigma$  = "the sum of"

$\Sigma X$  = add all the scores for variable X

$\Sigma x^2$  = first square each value of X, then add all squared values

$\Sigma(X+1)$  = first add 1 to each value of x, then add all x values)

## **FREQUENCY DISTRIBUTIONS**

-Step one: Plot Data

Graph or table, tells characteristics of data, patterns emerge

-Frequency distribution: how often each observations occurs

Basic setup:

X = score type (ex. how many males), categorical r numerical

F = frequency of occurrence

Categorical frequency distribution table (used for nominal/ordinal data)

-Contingency table: look at 2 categorical variables simultaneously

Shows joint frequency of crossing two variables

Marginal Distribution: depicted in the "total" columns

-Calculating  $\Sigma X$  from Frequency Table:

1. Recreate the original set of scores and sum them up
2. Multiply each score (X) with its frequency (f) and then add up the resulting values (ex.  $15(.25) + 30(.50) + 32(.25) = \underline{\quad}$ )

-Group Frequency Table (when you have too wide of a range of numbers, scores get grouped into class intervals)

Suggestions:

1. There should be about 10 class intervals
2. The width of each interval should be a simple number (e.g. 2, 5, 10, or 20)
3. The bottom score in each class interval should be multiple of the width
4. All intervals should be the same width

5. Equation for determining number of intervals and their width (always round up the number of intervals!)

$$\# \text{ of intervals} = \frac{\text{highest score} - \text{lowest score} + 1}{\text{Interval width}}$$

#### Relative Frequency Distribution

- $p = f(X) / N$
- $n$  or  $N$  = (number of observations, total frequencies)
- $p = \text{percentage} * 100$

#### Outliers

-Extreme values in a frequency distribution; may be excluded, BUT make sure that you mention that the data has been excluded

#### Bar Graph

- Type of histogram used to graph qualitative data
- Each bar is separated from other bars

#### Histogram

- \*\*\*Different than bar graph b/c columns touch, histograms are continuous
- Graph of a frequency distribution in which a rectangular bar is drawn over each value on the x-axis
- Classes are plotted on the x axis – Interval or ratio scale, discrete
- Frequency is plotted on the y axis – x axis: horizontal; y axis: vertical

#### Frequency Polygon

- line graph of frequency distribution, classes are plotted in the x axis and frequencies plotted on y axis
- Interval or ratio, continuous
- line graph version of histogram

#### Shapes of Histograms and Frequency Polygons

1. Normal Distribution
  - a. Familiar bell shape curve
  - b. Symmetrical
  - c. Describes many naturally occurring phenomena
2. Bimodal Distribution
  - a. Two modes or “humps” (usually due to systematic influence)
  - b. Most observations fall in or around one of two classes
  - c. Ex). Ages of members of two generation families (parents and kids)
3. Skewed Distribution
  - a. With few extreme values
  - b. Positively skewed: extreme values are large
  - c. Negatively skewed: extreme values are small