Still important ideas
Contrast the measurement of observable actions (and/or characteristics) with the theoretical constructs associated with those measurements
- Be able to identify examples of measurements versus constructs
- Describe what is meant by operational definitions
- Describe what is meant by validity versus reliability in measuring a construct
  - Provide an example of a measure that has high reliability and low validity?
  - Provide an example of a measure that has low reliability and high validity?
  - Describe the “target” example of reliability and validity
Contrast a dependent versus independent variable? - be able to identify them in an example
A study with a single independent variable would be associated with a univariate data set
- while a study with two variables (as in a correlation) would be associated with a bivariate data set
- a study with more than two variables would be associated with a multivariate data set
What are control (placebo) versus experimental (treatment) variables? - be able to identify them in an example.
What is a within versus between participant design? - also known as within and between subject designs?
What is random assignment - what's it good for?
What is random sampling - what's it good for?
Compare and contrast random sampling and random assignment
Populations versus samples - be able to define and identify examples of each
  - Define and contrast a statistic with a parameter
  - Sample data versus Population data
  - Sample mean (\( \bar{x} \)) versus Population mean (\( \mu \))

Describe what is meant by a placebo
Contrast the double-blind procedure with the single-blind procedure
Review the structure for organizing a memo

Contrast inferential statistics with descriptive statistics
  - Be able to identify each from a description of a study
Describe how quasi-experimental designs are especially vulnerable to
  - selection bias (when participants are not randomly assigned to groups)
  - especially from subject variables like gender, or political party
  - selection attrition (when some participants drop out of the study and create a bias in the sample)
Describe why we can talk about causation in ‘true’ experiments, but only relationships in correlational studies

Define continuous versus discrete variables
Define categorical versus numerical data
- These are also known as qualitative versus quantitative data
- Be able to identify each from examples
- Contrast verbal and coded labels of categorical data
  - Remember, that just because someone might code male = 1 and female = 2, this does not make “gender” a quantitative variable
- Describe what is meant by a binary variable

Levels of measurement: be able to identify each from an example
- Nominal, Ordinal, Interval, Ratio
5 Principles of questionnaire construction
Describe time series versus cross-sectional comparisons
Describe the “Likert Scales”

Describe surveys and what the goal of a survey would be
Contrast a census with a sample and contrast a population with a sample
- Review situations where sample or census might be preferred
- Describe how a sample can be representative versus biased
- Define random selection and random assignment relative to sampling for administering survey
Describe the five reasons to sample as described by the
Define a sampling frame
- How might a “sampling frame” be different from a “target population”
Describe why larger samples can be preferable to smaller samples
Describe why a higher response rate percentage is better than a smaller response rate
- Describe why response rate is so important for a sample to remain unbiased

Define and contrast probability sampling with non-probability sampling
Describe these sampling techniques
- simple random sampling (this is a probability sampling technique)
  - what is a random numbers table good for?
  - how would you find a random number using excel
- systematic random sampling (this is a probability sampling technique)
- stratified random sampling (this is a probability sampling technique)
  - proportional and disproportional stratified random sampling
- cluster sampling (this is a probability sampling technique)
- convenience sampling (this is a non-probability sampling technique)
- snowball sampling (this is a non-probability sampling technique)
- judgment sampling (this is a non-probability sampling technique)
Describe what is meant by sampling error

Describing Data Visually

Describe a frequency distribution
Describe the guidelines for constructing frequency distributions described in lecture
- mutually exclusive classes
- collectively exhaustive classes
- equal size intervals
Describe how raw data can be organized and presented in a dot plot
Describe what is meant by cumulative frequencies (be able to calculate them)
Describe what is meant by relative frequencies and cumulative relative frequencies (be able to calculate)
Describe a frequency histogram and a cumulative frequency histogram
Describe a frequency polygon and a cumulative frequency ogive

Describe the different shapes of frequency histograms and be able to interpret examples
- skewed left, skewed right (negatively or positively skewed)
- bimodal skewed left, bimodal skewed right
- skewed left with outliers, skewed right with outliers
- symmetric, multimodal symmetric, symmetric with outliers
- how does the skew of the distribution affect the order of the mean, median and mode
Describe a Bar chart, Pareto chart, stacked bar chart, pie chart, simple line chart and be able to interpret one
and when it should be used
Define a correlation: Describe and be able to interpret scatterplots
- strong positive pattern, strong negative, weak positive, weak negative, zero pattern, curvilinear pattern
Describe the relationship between the strength and direction of the correlation coefficient
Direction of correlation (positive, negative)
- Be able to identify a positive correlation from a correlation coefficient, a verbal description of the data, a scatterplot, or raw data
- Be able to generate examples of both negative and positive correlations
Strength of correlation (0 - +1.0 or 0 - -1.0)
- Be able to identify (or estimate) the strength of a correlation from a correlation coefficient, a scatterplot, or raw data
Correlation vs causation (when does a correlation imply, or provide evidence for causation?)
Describe linear vs curvilinear relationship -
- Be able to identify a linear or curvilinear relationship from raw data, a verbal description of data, or a scatterplot

Connecting intentions of the researchers (when designing the studies) with experimental methodologies, and appropriate statistical analyses and graphs. Be able to identify each of these 7 methodologies from a description of a study
1. Confidence intervals
   Using distributions to estimate means
2. t-test (2-means)
   Please note: t-tests always compares two means, there is one IV and one DV – typically is a bar graph
3. One-way ANOVA (note: ANOVA stands for “analysis of variance”)
   Please note: one-way ANOVA usually compares more than two means, there is one IV and one DV
   You may see either a bar graph representation or a line graph
4. Two-way ANOVA
   Please note: two-way ANOVA usually compares more than two means, there are two IVs and one DV
   You may see either a bar graph representation or more likely, a line graph
5. Correlational methodologies
   (Note: correlation uses two quantitative variables that must be interval/ratio numeric scale
   - uses scatterplot)
6. Simple and multiple regression
   Uses correlations to predict values on one variable based on values for the other variable
   - uses scatterplot
7. Chi Square
   Allows hypothesis testing for nominal data (just counting how many in each category)
   - also define and contrast these methodologies in terms of type of DV and IV
   - contrast quasi with true experiments (has to do with random assignment)
   - identify the DV and IV and whether it is a between or within participant design
   - also identify the most appropriate type of graph for each

Contrast inferential statistics with descriptive statistics
   Be able to identify each from a description of a study
- Describe why we can talk about causation in ‘true’ experiments, but only “relationships” in quasi or correlational studies. True experiments use random assignment that allows us to consider causal relationships, unlike quasi experiments
- Define categorical versus numerical data
  - These are also known as qualitative versus quantitative data
  - Be able to identify each from examples
- Levels of measurement: be able to identify each from an example - Nominal, Ordinal, Interval, Ratio
**Distributions**

Describe the dot plot display and frequency distributions and how they represent central tendency and dispersion and shape.

Define and contrast these three characteristics of distributions: central tendency (measure of location), dispersion (measure of variability), and shape.

Describe what it means to say that the "Mean is a measure of 'position', (it lives on one location of the curve) and than the standard deviation is a distance score of the spread of the distribution

- Describe the definitional formula for the standard deviation?
- Describe the relationship between sample size and standard error
  - as n goes up, df goes up, and variability goes down
  - describe why larger samples can be preferable to smaller samples
- How does this fit with the “Law of Large Numbers” (Gilbert Reading)

Describe the empirical rule for the normal curve

Normal distribution

Describe what is meant by a z-score and how area under the curve relates to a particular z-score

What is the probability that a score will fall above a z of 0 (50%)?

What is the probability that a score will fall between -1 and +1 standard deviation of the mean? - 68%

What is the probability that a score will fall between -1 or +1 standard deviation of the mean? - 34%
- notice, z = 1

What is the probability that a score will fall between -2 and +2 standard deviation of the mean? - 95%

What is the probability that a score will fall between -2 or +2 standard deviation of the mean? - 47.5%
- notice, z = 2

What is the probability that a score will fall between -3 and +3 standard deviation of the mean? - 99.7%

What is the probability that a score will fall between -3 or +3 standard deviation of the mean? - 49.85%
- notice, z = 3

Convert z scores to x scores

raw score = mean + (z score)(standard deviation)

Convert x scores to z scores (consider how the formula vary between samples and population)

z score = raw score – mean / Standard deviation

Finding z scores from raw scores
Finding z scores from probabilities (or area, percent, percentiles, or proportion of curve)
Finding raw scores from z scores
Finding raw scores from probabilities (or area, percent, percentiles or proportion of curve)
Finding probabilities from z scores
Finding probabilities from raw scores

How are the following related: Area under the curve, percent, probability, and proportion of the curve

Convert a raw score (x) to a percentile rank using the normal curve
Convert a percentile rank to a raw score (x) using the normal curve
Convert an x to a percentile ranking using the normal curve

Be able to use the z-table in our text.
Given an example of a distribution with a specific standard deviation, be able to estimate
- the score for a $z = 0$, $z = 1$, or $z = 2$ (from worksheet)
- the standard deviation (from worksheet)

Contrast positively skewed and negatively skewed distributions and how that affects the order of the mean, median and mode. *Remember, regardless of the skew, the median is always the middle score.*

Describe bimodal distribution

Given an example of a distribution with a mean and specific standard deviation

Define the characteristics of the normal curve
- Measured on continuous scale
  *Note: when range is large we often treat a discrete variable as continuous (exam scores for example)*

Measure of central tendency or “location”

Shape
- Possess clear central tendency
- Have only one peak (unimodal)
- Exhibit tapering tails
- Be symmetric around the mean (equal tails)
- Be bell-shaped

Domain
- theoretical and practical range in terms of standard deviations (or $z$ scores)

Measures of Central Tendency
- What are the three measures of “location” or measures of “central tendency”?
- Mode: most commonly occurring score – also tallest point on normal distribution
- Median: middle score also the midpoint score: (remember also the 50th percentile)
- Describe procedure for finding quartiles
- Describe how box plots use medians, quartiles, maximum and minimum scores
  *Note: just plots the minimum score, maximum score and first three quartiles*
- Mean: average score also balance point of distribution
  - define what is meant by a “trimmed mean”
  - define what is meant by a weighted mean

In a normal distribution mean = median = mode
In a positively skewed distribution mean > median > mode
In a negatively skewed distribution mean < median < mode
*Note: the mean is most influenced by extremely large or extremely small scores*
- which measure of central tendency is most affected by outliers

Probability

Describe what is meant by a “probability”

Describe and provide examples for the three approaches to probability
- Empirical
- Classical
- Subjective

Describe the “Law of Large Numbers”

Each measure of central tendency is useful for which type of data (nominal, ordinal, interval and ratio)

Measures of variability (dispersion)
- range: smallest score subtracted from the largest score
  note range makes no reference to scores between the largest and smallest scores
- variance: standard deviation squared
- standard deviation: typical amount observations deviate on either side of their mean

- $\Sigma(x - \bar{x}) = 0$ also $\Sigma(x - \mu) = 0$
- Describe the definitional formula for the population standard deviation (and variance)
- Describe the definitional formula for the sample standard deviation (and variance)
- Describe a “deviation score”
- memorize the standard deviation and variance (definitional) formula for samples and populations
- how is definitional formula different from the calculation formula for the standard deviation
  (be able to calculate standard deviation using calculation formula)
- how are the formula for standard deviations for samples different than for population
- how are standard deviation formula different from variance formulas
- describe what is meant by a deviation score
- what would happen if we took the average of the deviation scores (without taking the square or absolute values of each deviation)?
- describe the “mean deviation” - it uses the absolute values of the deviation scores
- what does it mean to say that the standard deviation is calculated relative to the mean
- be able to calculate standard deviation and variance from a set of scores
- be able to estimate it from a normal distribution when given an example deviation
- Note: standard deviation can also be estimated by range / 6

Describe what it means to say that the "Mean is a measure of 'position', (it lives on one location of the curve) and than the standard deviation is a distance score of the spread of the distribution

Provide examples of distributions that have the same mean but different variability and provide examples of distributions that have the same variability but different means (from worksheets in class)

Describe what is meant by
- “not unusual scores” (z of less than 2), “unusual scores” (z between 2 up to 3),
  “outliers” (z between 3 up to 4) and “extreme outliers” (z of 4 and above)

Contrast inferential statistics with descriptive statistics - Be able to identify each from a description of a study

Describe why we can talk about causation in ‘true’ experiments, but only “relationships” in correlational studies
True experiments use random assignment that allows us to consider causal relationships, unlike quasi experiments

**Describe the Central Limit Theorem**

Describe the three propositions derived from the central limit theorem
- Proposition 1: If sample size (n) is large enough, the mean of the sampling distribution will approach the mean of the population
- Proposition 2: If sample size (n) is large enough, the sampling distribution of means will be approximately normal, regardless of the shape of the population
- Proposition 3: The standard deviation of the sampling distribution equals the standard deviation of the population divided by the square root of the sample size. As n increases standard error of the mean (SEM) decreases.

Describe the sampling distribution of sample means
- what does this mean?
- how is this distribution formed? What is each point of data really mean in this distribution?
- how is this different from a population distribution of raw scores?
- how is the standard deviation different from the standard error of the mean (how are they similar?)
- what is the formula for standard error of the mean?
- why is the standard error of the mean nearly always smaller than the standard deviation of the population?

Describe what is meant by a point estimate, and an interval estimate – what are they for?

Contrast point estimates with confidence intervals
- Point estimates are a more specific estimation, but less likely to be exactly right
- Confidence intervals provide a range of scores, but is more likely to include the population parameter
- Describe the factors will affect the size of a confidence interval
  (especially variability, sample size and confidence level)
Plous text – Chapters 10, 11, 12, & 14

The Representativeness Heuristic (Chapter 10)

Describe what is meant by a heuristic (Plous, page 109)
Describe the pros and cons for using heuristics
Describe the “Linda the bank teller” example, what does this say about people’s perceptions of conjunction probability
Relate this situation with the nuclear war scenario described on page 112
Describe the quote by Tversky and Kahneman (1982) provided on page 111
Describe the “law of small numbers”
Describe what is meant by a “hot hand” or a “shooting streak”
Describe what is meant by “neglecting the base”
Describe what is meant by “regression to the mean”
Contrast clinical versus actuarial predictions

The Availability Heuristic (Chapter 11)

Describe what is meant by the availability heuristic (Plous, page 109)
Describe point made by comparing the probability of being killed by a shark vs by falling airplane parts
- and probability of being killed by violent versus disease-based process
- what does this say about the availability heuristic
Describe the point of the “imaginative study”, denial and vividness
Describe the point of the “legal significance of guacamole”

Probability and Risk (Chapter 12)

Describe the “game show problem” and Table 12.1 (Plous, page 132)
Describe the notion of “confusion of the inverse” (Plous, page 132)
Contrast simple probability with conditional probability (Plous, page 133)
Contrast conjunctive and disjunctive events (Plous, page 135)
Describe what is meant by conservative and anticonservative in decision making (Plous, page 138)
Describe the three recommendations to reduce biases in probability and risk estimation (Plous, page 143)

Perception of randomness (Chapter 14)

Describe the term “synchronicity” how does it relate to the main points of the chapter? (Plous, page 155)

How would you answer these two questions: (Plous, page 156)
1) Do people tend to see meaningful patterns in random arrangements of stimuli?
2) Can people behave randomly?

When is “lucky” an empirical fact, and when is it superstition? (Plous, page 156-157)
Describe Fischhoff and Slovic (1980) study that looked at peoples estimations about their own ability to predict the stock market

Note: Lecture notes can be found at http://courses.eller.arizona.edu/mgmt/delaney/