Readings:
Lind 1 – 11 (with emphasis on chapters 5, 6, 7, 8, 9 10 & 11)
Online readings: Appendix D, E & F
Plous – Chapters 10, 11, 12 and 14

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 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?
Populations versus samples - be able to define and identify examples of each
- Define and contrast a statistic with a parameter Lind, page 59, 60
- Sample data - Population data
- Sample mean (x̄) - Population mean (µ)

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 Lind, page 6
  - 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 Lind, page 9
Define categorical versus numerical data
- These are also known as qualitative versus quantitative data Lind, page 8
- 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 Lind, page 9 - 13
- Nominal, Ordinal, Interval, Ratio
Define what is meant by naturalistic observation and field observation
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 *Lind, page 7*
- 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 *Lind Text Lind, page 225*
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) *Lind, page 226*
  - 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) *Lind, page 229*
- stratified random sampling (this is a probability sampling technique) *Lind, page 229*
  - proportional and disproportional stratified random sampling
- cluster sampling (this is a probability sampling technique) *Lind, page 230*
- 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 *Lind, page 233*

**Describing Data Visually**

Describe a frequency distribution *Lind, page 23, 29*
Describe how raw data can be organized and presented in a dot plot *Lind, page 96 - 98*
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 a contingency table *Lind, page 113, 143*
Describe a tree diagrams *Lind, page 145*

Describe the different shapes of frequency histograms and be able to interpret examples
- skewed left, skewed right (negatively or positively skewed) *Lind, page 69, 106, 107*
- 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 *Lind, page 107*
Describe a Bar chart *Lind, page 24*
Describe a 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
- See also Lind 111, 112
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

Distributions
Describe the dot plot display and frequency distributions and how they represents 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 Lind, page 85 and 206

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 -1 or +1 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 -1 or +1 standard deviation of the mean? – 49.85%
  - notice, z = 3
Convert z scores to x scores *Lind, page 204 and 208*
\[ \text{raw score} = \text{mean} + (z \text{ score})(\text{standard deviation}) \]
Convert x scores to z scores (consider how the formula vary between samples and population)
\[ z \text{ score} = \frac{\text{raw score} - \text{mean}}{\text{Standard deviation}} \]

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 on page 529 of 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 *Lind, page 202*
- 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)

Describe the five major objectives (or goals) of using research in business
1. exploration
2. description
3. explanation
4. prediction
5. Influence

Measures of Central Tendency
- What are the three measures of “location” or measures of “central tendency”? *Lind, page 58*
- Mode: most commonly occurring score – also tallest point on normal distribution *Lind, page 66*
- Median: middle score also the midpoint score: (remember also the 50th percentile) *Lind, page 63*
- Describe procedure for finding quartiles *Lind, page 99, 100*
- Describe how box plots use medians, quartiles, maximum and minimum scores *(Lind, page 103)*
  *Note: just plots the minimum score, maximum score and first three quartiles*
- Mean: average score also balance point of distribution *Lind, page 59*
  - define what is meant by a “trimmed mean”
  - define what is meant by a weighted mean *Lind, page 63*
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” *Lind, page 127*
Describe and provide examples for the three approaches to probability *Lind, page 129 - 132*
- Empirical
- Classical
- Subjective
Describe the “Law of Large Numbers” *Lind, page 130*

Describe
- Probability of an event must lie within an interval of 0 and 1 *Lind, page 127*
- Probabilities of all simple events must sum to 1
- Complement Rule *Lind, page 135*
- Union of Two Events
- Intersection of Two Events (Joint Probability) *Lind, page 138*
- General Law of Addition *Lind, page 135*
- Mutually Exclusive Events *Lind, page 129*
- Special Law of Addition *Lind, page 134*
- Collectively Exhaustive *Lind, page 130*
- Conditional Probabilities *Lind, page 142*
- Dichotomous and binary variables
- Multiplication Law for Independent Events *Lind, page 140, 141*

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

**Measures of variability (dispersion)** *Lind, page 73, 74*
- range: smallest score subtracted from the largest score *Lind, page 75*
  - note range makes no reference to scores between the largest and smallest scores
- variance: standard deviation squared *Lind, page 78, 79*
- standard deviation: typical amount observations deviate on either side of their mean *Lind, page 78, 79*
  - $\Sigma(x - \bar{x}) = 0$ also $\Sigma(x - \mu) = 0$
- Describe the definitional formula for the standard deviation?
- Describe a “deviation score”
- memorize the standard deviation and variance (definitional) formula for samples and populations
  *Lind, page 79 - 82*
- 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 *Lind, page 75, 76*
- 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 and *Lind, page 203*)
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** *Lind, page 238*
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 *Lind, page 234 and 245*
- 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? *Lind, page 244*
- *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? *Lind, page 259*
Contrast point estimates with confidence intervals *Lind, page 259*
- 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)

**Standard error of the mean** *Lind, page 261*
Confidence Intervals with z scores *Lind, page 259 – 263 and 267-271*
Why do we care about the middle 95% of the curve?
- confidence intervals
- usual versus unusual scores
- hypothesis testing

Be able to calculate a confidence interval *Lind, page 259*
- for a z score distribution
- for different confidence intervals (80%, 90%, 95%, 99%)
- review different ways of interpreting confidence intervals (from lecture)
- how can we make our confidence interval smaller?
  - decrease level of confidence
  - decrease variability

*remember, variability can be decreased through 1) an increase in sample size and 2) improvement of reliability of assessment
tools and careful data collection techniques – our worksheet discussed this too*
Hypothesis tests
Describe what is meant by hypothesis testing
Describe generally what is meant by a null hypothesis, and contrast it with the alternative hypothesis
Describe what is meant by "we assume weird (or rare) things just don't happen"
Describe what is meant by a "critical z", "critical t", "critical value" or "critical statistic"
Describe the relationship between the observed z and critical z - which has to be bigger for significance?
- which has to be bigger to reach significance, or to reject the null, or to claim \( p < \alpha \) (for example \( p < 0.05 \))
Describe how "not rejecting the null hypothesis" is different from "accepting the null hypothesis"

If the observed z falls beyond the critical z in the distribution (curve):
- then it is so rare, we conclude it must be from some other distribution
- decision considers effect size and variability
- then we reject the null hypothesis – we have a significant result
- then we reject the null hypothesis
- \( p < \alpha \) (for example \( p < 0.05 \))
- then we have support for our alternative hypothesis

If the observed z falls within the critical z in the distribution (curve):
- then we know it is a common score and is likely to be part of this null distribution, we conclude it must be from this distribution (either because the effect size is too small or because the variability is too big)
- then we do not reject the null hypothesis
- \( p \) is not less than \( \alpha \) (for example \( p > 0.05 \)) \( p \) is n.s.
- then we do not have support for our alternative hypothesis

The z-test - Remember, the numerator of our test statistics (z) estimate the difference between groups (between means) while the denominators estimate the within group variability.

5 steps for hypothesis testing
1. identify research problem, describe null and alternative hypotheses,
2. identify decision rule, (alpha, d.f., critical value how do we find critical values?)
3. Calculations
   - be able to calculate standard deviations, standard error of the mean and z-scores
   (as we did in homework and worksheet)
4. make decision,
   - when do you reject the null hypothesis?
   - what does that mean relative to the alternative hypothesis?
   - what is our assumption about rare and common outcomes?
5. conclusion (what are the four parts?)
   1. Restatement of the hypothesis (or prediction)
   2. Restatement of the general statistical finding
   3. Statement tying them together - was our prediction right?
   4. what does this tell us about the big question

Generally the numerator of our test statistics (z, t, and F) estimates the difference between groups (between means) while the denominators estimate the within group variability.

Describe generally what is meant by a null hypothesis, and contrast it with the alternative hypothesis
Describe what is meant by a “critical z”, “critical t”, “critical r”, “critical F” or “critical statistic”
Describe the relationship between the observed statistic and the critical statistic z
- which has to be bigger to reach significance, or to reject the null, or to claim \( p < \alpha \) (for example \( p < 0.05 \))
- Be able to find the critical z for different levels of alpha and 1 vs 2 tailed tests Lind, page 298, 299, 303
  - What happens to our critical statistic when alpha gets larger?
  - What happens to our critical statistic when we move from 2-tailed to 1-tailed test?
- What is meant by conventional levels of significance? (- the common ones .05, .01 and .001)
- Statistical significance versus practical importance
Please contrast "significance" meaning important or relevant with "significance" meaning statistically different (or statistically reliable)

- Describe how the variance of the distribution effects the likelihood of rejecting the null hypothesis
  - larger variance makes it harder to reject (effect size has to be bigger) than a smaller variance
  - describe the relationship between effect size, variance, and sample size
- Region of rejection - what is it? – what does that mean?
- Be able to draw where it is on a distribution
- If we want to reject the null we want our z (or t) to be as big as possible and our p to be as little as possible

Note: These topics will not appear on Exam 2 but will be covered in the next section

1 vs 2 tailed tests: how do we decide when to use each

- One vs two tailed tests - what are the pros and cons of each
- how does it affect our critical values for an alpha of 0.05?
- how do we construct the hypothesis for a one versus two tailed test

Type I vs Type II errors Lind. page 295

- Be able to define them
- Be able to identify each type of error from a situation
- Which is false alarm, which is miss?
- What determines which is worse?
  - Consider criminal trial versus drug testing versus fire in the house
- How is alpha related to Type I error?
- How is beta related to Type II error
- Please note: power of a test is sometimes called "sensitivity"

Analogy between hypothesis tests and confidence intervals

"The two-tailed hypothesis test at the 5 percent level of significance (α = .05) is exactly equivalent to asking whether the 95% confidence interval for the mean includes the hypothesized mean..."

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
   (Please 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
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/