Chapter 1

1. Individuals of a Study
2. The Variable of a study
   - What is being **measured**
   - Qualitative vs. Quantitative
   - Emphasize the difference and point out it goes beyond "numbers" vs. words...
3. Parameter vs. Statistic
4. Levels of Measurement
   - Nominal
   - Ordinal
   - Interval
   - Ratio
   - A good thing to point out is that Ratio levels of measurement are usually are not negative valued because of the existence of this "absolute zero"
5. Descriptive vs. Inferential Statistics
6. Sampling Techniques
   - Simple Random – Each individual is equally likely to be selected in the sample
   - Cluster – Entire, randomly selected groups are included
   - Stratified – Random samples from every group is included
   - Systematic – Choosing names from a list in a systematic fashion
7. Experimental Design
   - Placebo
   - Treatment vs. control
   - Double blind

2.1 Frequency Distributions, Histograms, Etc.

1. Describe how frequency tables organize data into classes and then lists the numbers of data points in each class
2. Describe the procedure to finding class limits/widths for **integer** valued data
   (a) $class\ width = \dfrac{largest\ data\ value - smallest\ data\ value}{number of class limits}$

(b) Then **increase** this to the next integer, even if the above computation yields an integer. For example, $= 4.4 \to 5$ and $= 7 \to 8$.

3. How to compute limits,boundaries/midpoints of classes.
4. If the data is decimals, then multiply by an appropriate power of 10 to convert all data points to integers. Proceed as above. Then divide class limits, boundaries and midpoints by said power of 10.
5. Histograms
6. Relative frequency in tables and histograms
7. Shapes of distributions
    - Left-skewed data means the long-tail is on the **left side** of the median
    - Right-skewed data means the long-tail is on the **right side** of the median
8. Cumulative frequency tables
9. Ogive (pronounced oh-jive and rhymes with hive)

2.2 Types of Graphs

1. Bar Graphs
2. Circle Graphs (commonly known as Pie Charts)
    - Only used for parts of a whole
3. Time series data/graphs
4. Pareto graphs
    - A special type of bar graph/histogram
    - The rectangles are arranged from most frequent to least frequent
    - Not necessarily left-to-right; can be arranged vertically

2.3 Stem and Leaf Displays

1. Pay attention to the **key**.
2. Note the leaves need not be ordered
3. Can also be back-to-back to show data tables for two different groups.

3.1 Measures of Central Tendency: Mode, Median, Mode

1. Mode: The most frequent data value
2. Median: The "middle" data value
   - Explain how to compute this. Use the formula $\frac{n+1}{2}$ as a guide.
   - Be sure to explain the difference when there is an even or odd number of data points.
3. Mean: The arithmetic average.
   - For a **sample**, $\bar{x} = \frac{\sum x}{n}$ and a **population**, $\mu = \frac{\sum x}{N}$.
   - State our convention of using Greek letters to denote parameters and Latin letters (our usual letters) for statistics. Also, N is for a population size and n is for a sample size.
4. Weighted Average
   - Be sure to mention that how "weights" can be anything and it is not necessary for them to be percentages.
   - State how their grade for this class will be computed by a weighted average.

Participation	5%
Reading Assignments	10%
Chapter Reviews	10%
Quizzes	15%
Midterm 1	15%
Midterm 2	15%
Final Exam	20%

   - Also take this time to mention how the final exam score could replace a worse midterm score.

3.2 Measures of Variation

1. Give two data sets like

$$\{10, 9, 9, 8, 8, 8, 7, 7, 7, 7, 6, 6, 6, 5, 5, 4\}$$

$$\{10, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 4\}$$

   - Ask the students "Which data set is more varied?" Explain how standard deviation/variance measures this property.
   - Range: $highest - lowest$. Show both data sets have the same range.

2. Formula for Variance/Standard Deviation

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

   - As with means, be sure to point out the difference between a population variance and sample standard deviation. Students often get the two mixed up.
   - Always emphasize which one should be computing/using in formulas and discussions.
   - Explain that the second data set has a smaller standard deviation than the first.
   - Mention that we will not have to compute by hand, so no need to ponder the computational formulas.

3. Coefficient of Variation: Describes standard deviation as a percentage of the mean.

   - Perhaps note that for $\{100, 70, 70, 70, \ldots, 70, 40\}$, the standard deviation and mean is ten times than for the data set above; but are they really that different?

4. Chebyshev's Theorem: $1 - \frac{1}{k^2}$ of data falls within $\mu \pm k\sigma$ for $k \geq 1$.

   - Students will have difficulty understanding this lemma. A good idea state the following results:

$$\geq 75\% \text{ within} \mu \pm 2\sigma$$
$$\geq 88.9\% \text{ within} \mu \pm 3\sigma$$
$$\geq 93.8\% \text{ within} \mu \pm 4\sigma$$

   - Expect questions about this on Monday.
   - Note that for the mound-shape symmetric distributions, the result is much stronger. We will see that later.

3.3 Percentiles and Quartiles

1. Definition of a percentile
2. Box-and-whisker
   - The two ends and middle lines for the box represent quartile marks.
3. This concept is valuable in future sections.
4. Specifically mention quartiles as 25^{th}, 50^{th}, 75^{th}-percentiles (I wouldn't mention it now, but be prepared for next class to clarify that this textbook finds Q1 as the median of data BELOW Q2 (not including Q2). Not all textbooks agree and there are at least 9 different and sensible definitions for this values for discrete data.)
5. Visual display of LOWEST, Q1, Q2, Q3, HIGHEST (also know as the five-number summary)
6. Be sure to emphasize that **all** statistics in this chapter is handled by `1-VAR STATS` on the calculator.

4.1 What is Probability?

1. Summary box on page 137 pretty much sums up the key points.
2. A statistical experiment is any random activity that results in a definite outcome. Give examples like drawing cards, rolling dice, spinning the big wheel on "The Price is Right", etc.
   - Sample space, simple event, event. Give an example of these terms with dice rolling.
   - Probability will always be a value between 0 and 1 (inclusive). State when an event has probability 0 or 1.
   - I don't care if students use fractions, decimals, or percentages.
3. Notation
   - $P(A)$.
   - $P(A^c)$.
   - $P(A) + P(A^c) = 1$.
4. Computed by
   - Theoretical design and counting simple events. Be sure to emphasize formula of
   $$P(A) = \frac{\text{Number of outcomes in } A}{\text{Number of all possible outcomes.}}.$$
   - Empirical data and relative frequencies. Include the statement of the law of large numbers. Also emphasize that these methods will always **approximate** the probability of an event and may not ever get with total accuracy.
5. Why is probability relevant to statistics?
   - Probability is about assigning a likelihood to a particular outcome of an unknown experiment. The most relevant situation is assigning probabilities to sampling from a **known** population.
   - Statistics is about using the results of sampling to infer information about an **unknown** population.

4.2 Some Probability Rules—Compound Events

1. There are some formulas in this section, but often I try to de-emphasize them in trade for "careful counting".
2. The concept of conditional probability is always tricky for students. Carefully explain the concept with emphasis that for $P(A|B)$, the event B is **known** to have happened.
3. Note the two key terms
   - Mutually exclusive
   - Independent
4. Be sure to explain the difference between to conjunctions **or** and **and**. The use of a Venn diagram will be of help.
5. Describe a standard deck of 52 cards.
   - Show how to compute P(Ace), P(Heart),
     P(Ace or Heart), P(Ace and Heart), P(Ace|Heart).
6. Provide a contingency table and compute some probabilities with it.

	Political Affiliation			
Employee Type	Democrat (D)	Republican (R)	Independent (I)	Row Total
Executive (E)	5	34	9	48
Production Worker (PW)	63	21	8	92
Column Total	68	55	17	140

- $P(D)$ and $P(E)$.
- $P(D$ and $E)$.
- $P(D|E)$.

5.1 Discrete Probability Distributions

1. Refer to the "sum of two dice" exercise as an example of such a probability distribution.
2. Discuss **expected value** and **standard deviation** of a probability distribution.
   - Point how the formulas are like the formulas we already know for mean, variance and standard deviation but also what makes them different.

$$E(X) = \mu = \sum xP(x)$$

$$Var(x) = \sum (x-\mu)^2 p(x) = \sum x^2 P(x) - \mu^2 = E(X^2) - E(x)^2.$$

- These last two formulas for variance are not in the book, but are easier to use if someone were to compute them by hand.
- Note that 1-Var Stats can still get the job done for us, like a weighted average or frequency table. Use the probabilities in the "frequency table".

3. As far as the linear combination stuff, we don't do much with this topic. I plan to skip this, since the Excel discussion will be of more value. Mention it and tell them to read it carefully.

5.2 & 5.3 Binomial Distribution

1. Describe the features of a binomial experiment. Begin with the example of flipping a coin 10 times in a row and counting the number of heads (which we consider a success).

   - A **fixed number** n of trials.
   - Each trial is independent of all others.
   - Each trial has two outcomes: a success (with probability p) and failure (with probability $1-p = q$).
   - The goal is to count the number of successes r in n trials.

2. Present the formula, $P(X = r) = \binom{n}{r}p^r(1-p)^{n-r} = \binom{n}{r}p^r q^{n-r}$.

   Because we didn't cover Section 4.3 about counting, the reference to $C_{n,r} = \binom{n}{r}$ will be to state that it is the number of ways that the r successes could have fallen in those n trials.

3. Note that the TI calcs have binompdf and binomcdf. Refer them to the screencasts if they don't know how to use those functions already, but make a point to review them right before the worksheet next Wednesday.

   Specifically (on Wednesday) review the syntax and application of `binompdf(n,p,r)` which computes the probability of EXACTLY r successes out of n trials, $P(X = r)$, while `binomcdf(n,p,r)` computes the probability of at most r success out of n trials, $P(X \leq r)$. Discuss how we can compute the probability of at least r successes (say) when neither function is explicitly designed to do that by using compliments.

4. A quick note of the formulas for expected value and standard deviation for a binomially distributed random variable. They won't use 1-Var Stats, most likely, because the formulas are so much simpler.

$$E(X) = np$$

$$Var(x) = npq = np(1-p)$$

Excel Intro

1. Use the "Excel for Friday" file. There is a Highlights sheet that has some key ideas to cover as a point of reference for you to plan, a DATASHEET sheet with the some data set up to demonstrate to computations, and an Answer Key sheet. (If you aren't sure what to do, practice by referring to the "Answer Key".)

2. Also, tell them to download the Excel Sheet in D2L **before** class on Friday if they plan to bring their laptops. I found it is more beneficial for them if they can follow along on their own.

3. The idea is to give a student who has little or no experience with Excel a quick look at how to enter a basic formula or function to complete a computation on the DATASHEET page.

4. The dataset is small so after using the function to compute the result, a quick visual verification can be done. The dataset for the project is much larger and hand verification is unrealistic.

5. Also, another item to stress as you do the demo is the idea of robustness, which is accomplished through cell referencing. So, for example, when you compute % Heads, you will use the cell reference as `=A3/B1` rather than **hard coding** in as `=31/50`. That way, WHEN new data is entered to overwrite the original data, the % will update automatically.

    - Note that part of the grading process will be to actually paste a new set of data over their original data set to see if the objects they created update and adjust to the new data set. So, it is important that they understand this expectation.
    - As an example, after having coded in the robust functions, like =MIN(A3:A22) to see LOW value, go into the data and overwrite the low value with a LOWER one so that the students can see the outcome for the formula change.

I am happy to provide a brief tutorial on Excel, if you need it. Just let me know and we will figure out a time to meet.

The primary goal here is to give the students an outline of material to review for the Midterm on Wednesday. Since they have a quiz this day, I will not have enough time to fill in details. I plan to write this out and if someone has a specific question on something, I will then go into detail.

1. Remind them that Chapter Review (Chapter 4) on WebAssign is due tonight AND although Chapter Review (Chapter 5) is not due until 2/19, the midterm on 2/17 covers Chapter 5 material, so it will be to their advantage to complete it before the midterm.
2. As far as resources during the midterm:
   - Calculator (with NO internet access) When a calculator function is used, **work required is to write the function and input values used.** I do not supply calculators. If they show up without one on exam day, then they will suffer. Phones will not be allowed as a substitute.
   - Formula sheet: A copy of the "Frequently Used Formulas" for Chapter 1-5 (as shown in the back cover of the textbook; they can even see that in the online textbook).
3. Resources for more practice problems:
   - WebAssign problems with "Practice Another Version". Not all problems are programmed with this option in WebAssign.
   - Odd-numbered problems from the textbook where the correct answer can be checked in the back of the textbook (even the online version).
4. The worksheet on Monday will be a "mini-midterm" useful for review. Homework for Monday is to start studying for the midterm.

Outline of content by Chapter

1. CHAPTER 1
   (a) Vocabulary terms: Individual, population, quantitative variable, qualitative variable, statistic, parameter, descriptive statistics, inferential statistics
   (b) Levels of measurement: Nominal, ordinal, interval, ratio
   (c) Sampling techniques: Random, stratified, systematic, cluster, convenience, multi-stage
   (d) Basics of experimental design: Observational study vs experiment, Control group, placebo and placebo effect
2. CHAPTER 2
   (a) Displaying data: Frequency tables
   (b) Class limits, class boundaries, class width, midpoint, relative frequency, cumulative frequency
   (c) Histograms & ogives
   (d) Symmetry and skewness of histogram
   (e) Graphs: Bar graph, Pareto chart, Circle (pie) graph, Time-Series graph
   (f) Stem-and-Leaf displays

3. CHAPTER 3 : `1-Var Stats`

   (a) Central tendencies: Mean, median, mode, trimmed mean, weighted mean

   (b) Variation

   - Range, variance, standard deviation (statistic and parameter)
   - Coefficient of variation
   - Chebyshev's Theorem

   (c) Percentiles and Box-and-Whisker

   (d) Quartiles, IQR, 5-number summary

4. CHAPTER 4

   (a) Elementary probability theory

   - Sample space, notation $P(A)$, $P(A^c)$, equally likely outcomes, using relative frequency
   - $P(A$ or $B)$, $P(A$ and $B)$, $P(A|B)$, independence, mutual exclusivity

5. CHAPTER 5

   (a) Discrete probability distribution

   (b) Valid probability distribution, mean (expected value), standard deviation

   (c) Binomial probabilities

   - Criteria of a binomial experiment
   - Probability of exactly r successes, at most r successes, at least r successes, fewer than r successes, more than r successes, etc.

I hope to have my Midterms graded by Friday. So, I may start with reviewing any commonly missed problems, if there seems to be some obvious choices for that discussion. Otherwise, I do NOT intend to go over the midterm in class; students will have to seek out help in office hours on an individual basis.

6.1 Normal Probability Distribution

1. Ask the class about the difference between a discrete random variable and a continuous random variable.
2. Ask the class what the sum of the probabilities in a discrete probability distribution must equal.
3. Introduce the normal distribution
    - Unimodal, symmetric, approaching x-axis, area under curve is 1.
    - Draw a picture and label the mean. As note the distance between the maximum and inflection points is equal to the standard deviation, which is 1.
    - I see no need to write down the density function, as it will likely be met with vacant stares.
    - Point out that the students should pay close attention to the Empirical Rule in the reading and to note its connection to Chebyshev's Theorem.
4. Introduce the concept of a Control Chart.
    - Using a mean and standard deviation computed from historical data or industry standards, a control chart is a mechanism to determine if the variable is in statistical control. This chart is essentially following the expected distribution from the mean and standard deviation.
    - There are 3 "out-of-control" signals. Make note of them from the reading.
        - One reading beyond 3σ of average.
        - 9 **consecutive** on one side of average.
        - 2 of 3 points beyond 2σ of average.

Sections 6.2 & 6.3 Computing Area under Normal Curves

1. In these sections, they will see examples and applications of computing the area under the normal curve above, below, and between specified points.

   - `NORMALCDF` (under the `DISTR` menu) will be their friend here. Somewhat similar in concept to `BINOMCDF`, it is more flexible with the ability to accept both lower and upper bounds.
   - If they have no lower (or upper bound) for a probability question, have them use `1E99` for ∞ and `-1E99` for $-\infty$.
   - The book will show them methods using a table. Those methods are NOT necessary, if they have the calculator. If they desire, tables are to be made available for quizzes, exams and the final.

2. The z-value (or z-score) of a data value x gives the number of standard deviations between the data value and the mean.

3. When the variable z is used, the assumption is that mean = 0 and standard deviation = 1.

4. When the variable x (or any other variable for that matter) is used, we will need to be provided the information on the values of the mean and standard deviation.

5. EITHER z- or x-information can be entered into the `NORMALCDF` function.

6. Another function that will be useful is `INVNORM` (also under `DISTR`). This is another somewhat restricted function in that it can **only be applied to a LEFT-TAIL area**.

7. The information about "checking for normality" is not something that we will assess in this course. That is not to say that it is worthless, it is just not something that we will explore in this class.

1. The content of 6.4 and 6.5 are really the heart of what makes almost all the rest of the material in the class work. So, the big picture of what is happening is valuable.

2. We are focusing on the normal distribution. Review the shape, mean and standard deviation of this distribution. Even though the population we inquire about may not be normally distributed, this distribution is none-the-less important.

3. If we have a random sample from a population, we can certainly compute the mean of that sample, but how well can we expect that sample mean to represent the population mean?

6.4 & 6.5 Sampling Distributions and the Central Limit Theorem

1. If we were able to take every possible random sample of a specified size n from a population and compute the sample mean for each of those samples, what would that distribution of sample means look like?

2. The Central Limit Theorem tells us that we can expect that (with certain restrictions) the distribution is normal. This is a pretty profound and powerful result. No matter how scattered the population itself is, with some proper guidelines, we can expect that the sample means to be well behaved.

3. Because of this, we can determine the probability of finding a sample with a specific mean (assuming we know the population mean). This is very important as it is the cornerstone of why confidence intervals and hypothesis testing are sound reasoning.

4. The middle (mean) of that distribution of sample means is equal to the population mean itself. So, although there are some sample means that lie far off in the tails of the distribution, they aren't very likely.

5. Further, the standard deviation of that distribution of sample means is equal to the population standard deviation divided by the square root of the sample size. So this implies that the larger the sample size, the smaller the spread of the sample means, which should make sense. The larger the sample size, the more likely it represents the population well.

Confidence Intervals

When we can't collect a measurement for every member of a population, how can we determine a population mean?

1. Without a full set of population data, we can never be 100% certain that we know the population mean, but with certain restrictions applied, we can use the data from a random sample to estimate a population parameter.

2. Even with a large random sample, the value of the sample mean is usually not exactly equal to the population mean. But, according to the Central Limit Theorem, we can have some expectations on how likely it is that the sample mean falls within a certain interval around the population mean. A confidence interval is an interpretation of precisely this application.

3. What criterion is required to apply the Central Limit Theorem?

4. The idea is that we start with a sample statistic (called a point estimate). We then create a margin of error around that point estimate which yields an interval of values that is asserted as one that contains the population parameter (at least with some high level of, but not 100%, certainty)

   (a) So, this looks like (point – error, point + error) or a guess ±error.

   (b) The size of the error depends on what level of certainty we want to assert.

   (c) The most prevalent example of a confidence interval are during elections and predicting election results.

5. Keep in mind that although we will be asserting that we have an interval that contains the population parameter, there is no indication where within the interval we expect it to lie.

7.1 Estimating μ when σ is known

1. The first look at confidence intervals assumes that we know σ. (This might seem a bit contrived, because why would we know σ if we don't know μ.)

2. The TI `Z-INTERVAL` function will compute the confidence interval, either by entering the actual list of sample data or by entering the sample mean. For that reason some of the details of the formulation developed in the reading are not so critical, but understanding what the function does and how to interpret the result is critical.

3. When reading, pay special attention to

   - The formula for E in this case, as it should make sense why this is the correct formula.
   - The definition of z_c, as this should also be something that you already know how to compute.
   - The interpretation of a confidence interval, as it is easy to misinterpret what the interval found means.

7.2 Estimating μ when σ is unknown

1. In this case, a seemingly more common case, we don't know σ. So, we must estimate the value of σ too. This translates into a slightly larger margin of error to compensate for the potential error in our guess of σ.

2. Rather than the standard normal distribution, we use the Student's t-distributions. There is a slightly different distribution for each sample size, but they are all bell-shaped. As n gets larger, the t-distributions approach the standard normal distribution.

3. The formulation and interpretation of a confidence interval in this case is very similar to that when σ is known, just the distributions from which the critical values are determined has changed.

4. The `T-INTERVAL` function the TI can do the work for us.

5. When reading, pay special attention to

   - The formula for E (which should look VERY similar to that from Section 7.1).
   - The formula for degrees of freedom.

6. If your calculator does not already have an `InvT` function, there is a screencast available about how to create a program in your calculator that does the job.

Internet Visual for Confidence Intervals

1. Start with an overview of what this whole confidence interval and confidence level thing is all about. This applet could be effective in confirming what idea of the c-level, also as a way to compare how the size of the interval changes with changes to the variables.

2. The website is `http://www.rossmanchance.com/applets/ConfSim.html`

3. Start with Means, Normal, z with σ.

   - Show 1 sample
   - Show 10 samples

- Show 100 samples, maybe a few times to show how the number of "good" confidence intervals can vary.
- Change Conf level
- Change n

4. Reset and look at Means, Normal, t. Note how they can have different lengths. (this is sometimes subtle) Why?

Section 7.3 Confidence Interval for a Proportion

1. This time, rather than a mean, we are estimating a population proportion, like "What percentage of all college students change their major at least once in their first four years?" which is different than "What is the average number of times a college student changes their major within their first four years?"
2. The calculator function is `1-PropZInt`.
3. In the reading watch out for
   - The requirement on the sample size, it is more complicated than just $n \geq 30$.
   - The formula for E.
   - The formulas for finding sample size.
   - Interpreting poll results.

Section 7.4 Confidence Intervals for Differences

1. As our last look at confidence intervals, we look at differences (between two means or between two proportions), as a way to tell if two populations are different.
2. This section refers only to independent samples, but check out the reading for the definitions as it will matter later (when we do tests).
3. `2-SampZInt`, `2-SampTInt`, and `2-PropZInt` are the calculator functions.
4. In the reading watch out for
   - The degrees of freedom for Tint
   - The criteria on sample size (it is again different for p than for μ).
   - The interpretation of the confidence interval
     - When the interval contains only negative values
     - When the interval contains only positive values
     - When the interval contains both positive and negative values.

Hypothesis Testing

In Chapter 7, we estimated the value of population parameters (mean and proportion) using confidence intervals. Another method of statistical inference is to make decisions concerning the value of a population parameter, which we do in Chapter 8 with hypothesis testing.

1. Suppose that you roll a regular six-sided die 600 times. About how many times would you expect to see a 4 rolled within those 600?
   (a) If you saw 105 rolls that were 4, would this be surprising... enough to question the fairness of the die?
   (b) What if you saw 595 rolls that were 4, would this be surprising... enough to question the fairness of the die?
   (c) Where would you draw the line between "not so surprising" and "surprising"?
2. The basic idea in hypothesis testing is to start with an assumption of what "should" happen and to draw a line on what extreme outcomes would be "surprising". If the random sample indicates a "surprising" result, we have evidence to abandon our assumption... if the ransom sample indicates a "not so surprising" result, we do not have adequate evidence to abandon our assumption and we must stick with it.
3. Note, as with the case of the rolls of the die, the result of the random sample may be very "surprising" (595 of our 600 rolls were 4), but it will never serve as PROOF that our assumption is wrong (as it's possible that this 595/600 happens with a completely fair die).

8.1 Introduction to Statistical Tests

1. This section introduces the language and formalizes the concepts of "drawing a line" and interpreting the results of the test.
2. When reading, pays close attention to
   (a) The notation and definition/usage of the null hypothesis and the alternate hypothesis.
   (b) How we categorize the test as right-tailed, left-tailed, or two-tailed.
   (c) What the P-value measures and how it is used to draw the conclusion of the test.
   (d) The usage and meaning of the conventional language of "Reject H_0" and "Fail to Reject H_0".
3. I plan to save the discussion of types of errors until Monday when they should have a slightly better sense of what this is all about.

8.2 Testing the Mean

1. The calculator functions that will be useful are `Z-Test` and `T-Test`. Like `Z-Interval` and `T-Interval`, one is used with σ is known and the other when σ is unknown.
2. They may find the "critical regions" method helpful in solidifying the concepts of hypothesis testing, but they will be required to compute and interpret P-values as well. So, this "critical regions" method should be considered a secondary method.

8.3 Testing the Proportion

The calculator function that will be useful is `1-PropZTest`.

8.4 Paired Data

1. A discussion of dependent (paired) versus independent data will need to be had. We treat paired data quite differently than independent data.
2. A good guideline is the following: If you implement two distinct processes on **the same group** of individuals, then the data is paired.
3. When the data can be paired, the null hypothesis is always that the mean of the differences is 0 (there is no difference).
4. They will need to created a "new" data set that is the list of differences of the pairs. Then use T-Test to complete the test. By typing $L_1 - L_2$ in the label of the stats editor will generate a difference column.

8.5 Independent Populations

1. Here the null hypothesis is always that the difference of the means (or proportions) is 0 (there is no difference).
2. The `2-SampZTest`, `2-SampTTest`, and `2-PropZTest` are the key calculator functions.
3. Computations are easy with the calculator... reading and deciphering which function does the job for a specific problem is the tough part.

Project 2

1. I think that simply stepping through the R Tutorial sheet (or similar exercises) is a good demo.
2. Please announce that there are screencasts in D2L to help show them some of the key functions.
3. It is the workspace (.RData) file that they need to submit. Not their code (.r file), not a picture of their code or a text version of their code... we don't need to see their code... just the workspace containing the objects created from their code.
4. A computer will be grading their workspace objects. This means that all required objects must be named exactly as written in the project or the grader will not see it. Note that extra objects will simply be ignored, so they can have MORE than what is required with no penalty. But changes in case, transposition of letters, or extraneous symbols in a required object's name will appear to the grader as if the object simply does not exist. Partial credit may be possible on some objects, but NOT as a result of spelling/typing errors in the object name.
5. Items 2-6 on the worksheet refer to values of objects in their workspace. To earn full credit on these items, the values on the worksheet must match the values in the workspace. Further, these particular objects will be randomly generated when the code is executed. So, rerunning their code can/will change the values of the workspace objects. Therefore, they should not attempt to complete those worksheet items until their workspace is finalized!!

The primary goal here is to give the students an outline of material to review for the Midterm on Wednesday. Since they have a quiz this day, I will not have enough time to fill in details. I plan to write this out and if someone has a specific question on something, I will then go into detail.

1. Remind them that although Chapter Review (Chapter 8) is not due until 4/15, the midterm on 4/13 covers Chapter 8 material, so it will be to their advantage to complete it before the midterm.
2. As far as resources during the midterm:
   - Calculator (with NO internet access) When a calculator function is used, **work required is to write the function and input values used.** I do not supply calculators. If they show up without one on exam day, then they will suffer. Phones will not be allowed as a substitute.
   - Formula sheet: A copy of the "Frequently Used Formulas" for Chapter 1-8 (as shown in the back cover of the textbook; they can even see that in the online textbook).
   - The "Calculator Functions Syntax" sheet on D2L. This just states the necessary parameters for each function with no explanation of what they mean.
3. Resources for more practice problems:
   - WebAssign problems with "Practice Another Version". Not all problems are programmed with this option in WebAssign.
   - Odd-numbered problems from the textbook where the correct answer can be checked in the back of the textbook (even the online version).
4. The worksheet on Monday will be a "mini-midterm" useful for review. Homework for Monday is to start studying for the midterm.

Outline of content by Chapter

1. CHAPTER 6: Normal Curves and Sampling Distributions
   (a) Properties of Normal distribution
   (b) Empirical Rule: Approximation of how much area is between $\mu \pm k\sigma$ for $k = 1, 2, 3$.
   (c) Control Charts, we have 3 "warning signs".
      - Any point beyond $\pm 3\sigma$.
      - Nine consecutive points all above or all below μ.
      - Two of Three consecutive points beyond $\pm 2\sigma$.
   (d) Conversion to standard normal distribution: $z = \frac{x-\mu}{\sigma}$ or $x = \mu + z\sigma$. A z-score represents the number of standard deviations from normal an observation is.
   (e) Using `normalcdf` to compute probabilities. Use $\pm 1E99$ for bounds $\pm\infty$.
   (f) The Central Limit Theorem

2. CHAPTER 7: Estimation (or Confidence Intervals)

   (a) General Philosophy of an $x\%$ confidence interval: A **process** that produces an interval which contains the desired parameter $x\%$ of the time.

   (b) This **does not** mean given an $x\%$ confidence interval, the chance the interval contains the parameter is $x\%$. This is like saying a 99% accurate test (which means the text gets the **correct** diagnosis 99% of the time) that says you have a disease means the probability of you **actually having the disease** is 99%; which is a false, but common conclusion people make.

   (c) The above point is a very common misconception and is often propagated. I myself am guilty of this. The good news however is that the confidence intervals we produce in this class actually have this additional property, so it is probably worth mentioning the fallacy, but not spending much time on it.

   (d) CI types:

   - CI for μ, z or t-based.
   - CI for p a population proportion. Recall the requirements!
   - CI for $\mu_1 - \mu_2$, $p_1 - p_2$
   - Mention the relevant functions.

3. CHAPTER 8: Hypothesis Testing

   (a) Null Hypothesis: The assumption that a parameter is equal to some value.

   (b) Alternate Hypothesis: The belief being "tested", phrased that the parameter is different (in some fashion) from what is assumed.

   (c) p-value and α, the level of significance.

   (d) Error types

   (e) Test types – completely determined by H_1.

   (f) Reject H_0 when $p <= \alpha$. Phrasing a proper conclusion.

   (g) Hypothesis test types:

   - z and t tests
   - Paired vs. Unpaired data. Be sure to state the null hypothesis is **always** $H_0 : \mu = 0$ for such tests.

9.1 & 9.2 Linear Correlation

1. The key ideas here are, of course, the correlation coefficient and the least squares line of best fit. This is frequently a topic that students have some familiarity with ... and some intuition for.
2. If you are so compelled (inspired by) the actual formulas for the various pieces, then by all means, discuss. However, as for most things so far, the calculator can take care of it for them.
3. I do think a brief conceptual discussion about "least-squares" is based on is very much worthwhile. A discussion of how r measures the "goodness-of-fit" and how to interpret its value is.
4. The key calculator function is `LinReg(a+bx)`.
   - Note that there is also a `LinReg(ax+b)` that does precisely the same thing. HOWEVER, because we will be referring to the population slope with the letter β, the former is a better choice.
   - Be sure to mention that we will **only** be using `LinReg(a+bx)` in this class to compute the line of best fit.
   - In order to get the r and r^2 values to display, the `DiagnosticsON` must be set. To do this, go to `Catalog` (which is located by 2^{nd} + 0), select `DiagnosticsON` and push `Enter` twice.
5. A brief mention of the difference of interpolation and extrapolation would also be a good note.
6. Mention the coefficient of determination and what it measures.

9.3 Inference for Correlation

1. Note the notation for population parameters ρ, β, and y.
2. Most calculators should have `LinRegTTest` which simultaneously tests the sign of ρ and β. Note that like null hypothesis in both cases is "$= 0$".
3. Now, the TI-84 will likely have `LinRegTInterval` to determine a confidence interval for the value of β, but TI-83 will not. There is a screencast to help them program a function to compute the interval.
4. Further, neither TI-83 nor TI-84 will have a built-in function to compute a prediction interval (confidence interval) for y. (The TI-89 should.) The formula for E is not impossible to evaluate with the function, but it is certainly a bit of a pain. It may be good to present the formula so that they have a preview of what will be expected of them if they don't program the function.
5. There is a screencast on D2L and definitely mention this to them in class.
6. One conceptual point to highlight, the closer x is to the $\bar{x}$, the narrower the prediction interval is.

10.2 Chi-Square: Goodness of Fit

1. Start with an overview of what "goodness of fit" means. The null hypothesis is always that the population "fits". The more deviant the data is from the expected distribution the worse the fit.

2. Although this is a test that is supported by the TI, I like to spend some time analyzing the formula for chi-square and the distribution. It just so easy to see what it is measuring and why large test statistic values imply small P-value.

3. Note that we are using a different distribution (than standard normal or Student's t).

4. Note, we have yet another measure of degrees of freedom.

5. The key calculator function is χ^2`GOF-Test`.

10.5 ANOVA

1. My intent here is NOT to really get them to delve into the "why" and "how" the derivation of the sample F ratio, as I think that would require more time that we have. So, you can note that in the reading they will see a BUNCH of steps and intermediate computations, but they will not be responsible for understanding all those details for this class.

2. Talk about the null and alternate hypotheses of the test.

3. Talk about the sample F ratio is a measure of the variance BETWEEN populations versus the variance WITHIN each population. When there is significant variance BETWEEN, our F ratio will be big (we are removing the impact of the variance within each separate population and basically isolating the difference between populations).

4. Note that we are using yet another distribution.

5. Note that evidence supporting that there is a difference between the populations provides no direct evidence on which population might be the different one.

6. The key calculator function is `ANOVA`.