Honors Applied Statistical Methods Fall 2011 4 credits Nancy Pfenning

Seminar Format

In addition to three one-hour periods a week devoted to lecture interspersed with discussion, a single one-hour period per week features three students, each taking a turn as presenter/discussion leader. Each of these three has chosen a topic from the previous week's lecture—for example, inference for relationships between two categorical variables. The student seeks out a journal article or a report from the media, preferably pertaining to his or her major area of study—for example, a pre-pharmacy major may locate a report on side effects suffered by patients given a drug versus a placebo. The article is made available to instructor and classmates to be read in advance. During the actual "fifteen minutes of fame", the student summarizes the article's content, stressing its statistical aspects. Then he or she facilitates discussion of pertinent issues—for example, how were the subjects recruited for the study? Can we be fairly certain that the subjects were blind to who received drug and who received placebo? Are there any potential flaws in the study's design? How were the side-effects assessed—for example, if "insomnia" is listed, how did researchers decide whether or not a particular subject should be classified as having suffered from insomnia? Finally, is there good reason for takers of the drug to fear those particular side-effects?

If the next student is to make a presentation on inference for a single categorical variable using a chi-square goodness-of-fit procedure, there may be a report on whether certain racial/ethnic groups are under-represented by characters in prime-time television entertainment series. On one hand, this presents an entirely different context for the use of statistical tools. On the other hand, the same basic processes must be addressed: How were the data produced? How were they summarized? Is what we see in the data indicative of some general trend?

During the part of the course that deals with probability, students can meet in the Statistics Computing Lab to perform simulations of sampling distributions and then display and summarize their findings, to illustrate the workings of the Central Limit Theorem. Other seminars for this part of the course require a student to carry out a random sampling exercise in advance, then summarize and explain the outcomes.

With a class-size limited to 20 students, during a fifteen-week course, each student will be able to make two fifteen-minute presentations. Students sign up for topics in lecture, and can get help from the instructor or course assistant, if necessary, to track down articles or reports that feature the tools under current discussion. Depending on class size, some of the Alternative Activities can be used as seminar topics towards the end of the semester so that all students complete the requirement of two presentations. If time allows, special presentations will be made by statistics faculty, to be followed by questions and discussion. There will be weekly homework assignments and a weekly practice quiz for self-assessment.

Grading

Each seminar is worth 25 points. The following aspects will be taken into account:

- well-prepared?

- interesting?
- well-presented?
- statistical correctness
- relevance to material covered

Seminar Topics

The instructions below should serve as guidelines; presenters must focus on the assigned lecture topic, but can approach it in other ways in addition to what is suggested. Care should be taken to plan a presentation that lasts approximately 15 minutes. The success of presentations about statistical studies will depend to a great extent on the particular variables or relationships considered. Therefore, you should make an effort to discover reports of particular interest to you or potential interest to the class, in order to stimulate discussion.

The articles or reports discussed must be made available to the class no later than Monday of that week, either by posting the url in Blackboard or by handing out hard copies in Monday's lecture. All students are responsible for reading these materials before the Thursday seminar.

Thursday, September 8 (provide url by Monday, Sept. 5 or hardcopies to me by Tuesday, Sept. 6)

1(a). Find (and distribute) examples in newspapers or journals or on the internet of studies involving the five variable situations C, Q, $C \to Q$, $C \to C$, and $Q \to Q$ and describe the variables in each situation. Discuss if, and how, the variables could have been handled as a different type from the one actually used.

1(b). Find (and distribute) several articles or reports about studies for which a sample has been taken. Tell what you know, or can reasonably guess, about what kind of sampling plan was used in each. Be sure to mention whether the sampling plan included randomness, and whether it seems susceptible to some kind of bias. Tell whether the sample size seems fairly large and, if not, whether it would be feasible to obtain a larger sample.

2. Find (and distribute) several sample surveys posted on the internet. For each, tell if each variable of interest is quantitative or categorical. Tell what is the suggested population of interest. Then tell how the individuals were selected and whether or not you believe they adequately represent the population of interest. Discuss whether there are any clear sources of bias in the wording of the question, or in the options offered if the questions are closed. If there is a sensitive topic involved, discuss if people can be expected to respond honestly.

Thursday, Sept. 15 (provide url or hardcopies by Monday, Sept. 12)

3. Find (and distribute) an article or report about an observational study. Tell what the variables of interest are, whether they are quantitative or categorical, which is explanatory and response (if there are two variables). Are there any potential confounding variables that should have been controlled for?

4. Find (and distribute) an article or report about an experiment. Tell what the variables of interest are, whether they are quantitative or categorical, and which is explanatory and response. Describe the subjects, treatments, whether or not the subjects were blind, whether or not the researchers were blind, and whether there are any obvious problems with the experiment's design.

5. Find (and distribute) a journal or newspaper article or internet report about a study that involves just a single categorical variable. Tell how the data were produced (experiment, observational study, or sample survey), and if the variable is summarized with counts, percents, or both. Tell if results are being reported for a sample or for an entire population, and report the number of individuals studied, if known. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values.

Thursday, Sept. 22 (provide url or hardcopies by Monday, Sept. 19)

6. Find (and distribute) a newspaper or journal article or internet report about a study that involves a single quantitative variable whose values have been summarized, and displayed with a stemplot, histogram, and/or boxplot. Tell how the data were produced—was it an experiment or an observational study? Tell if results are being reported for a sample or for an entire population. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values. Discuss the center, spread, and shape of the distribution as evidenced in the summaries and display, and comment on whether other summaries/displays would have been equally effective.

7. Find (and distribute) a newspaper or journal article or internet report about a study that involves a single quantitative variable *whose values are specified*. Tell how the data were produced—was it an experiment or an observational study? Tell if the variable is summarized with mean, median, or neither. Use software to find the mean and standard deviation of the data set. Display the data and discuss the shape of the distribution. Report how well or badly the data set conforms to the 68-95-99.7 Rule. Tell if results are being reported for a sample or for an entire population. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values.

8. In this student-led activity, students calculate their individual z scores for certain variables, then consider the pattern of z scores for the entire class.

(a) Find the mean and standard deviation of the shoe sizes of class members. Each student in the class uses a calculator to find his or her z score for shoe size and reports it to the class. Together the class constructs a histogram to display all the z scores and comments on its shape, noting whether there are both large negative and large positive z scores. Determine what percentage of z scores are less than 1, 2, and 3 in absolute value to see how well students' shoe sizes conform to the 68-95-99.7 Rule. Discuss whether results would differ for a larger class.

(b) Students report data on the number of minutes spent watching TV the day before (or doing school work, or on the computer, or on the phone), and the leader uses a calculator to find the mean and standard deviation. Students calculate and report their individual z scores, then the class constructs a histogram for the data, comments on its shape, and notes whether there are both large negative and large positive z scores. Determine what percentage of z scores are less than 1, 2, and 3 in absolute value to check conformity with the 68-95-99.7 Rule. Discuss whether reults would differ for a larger class.

(c) Depending on how much time is available, students examine their z scores for other variables such as number of hours slept the night before (to the nearest half hour), age in years (to the nearest tenth), number of siblings, credits taken, etc.

Thursday, Sept. 29 (provide url or hardcopies by Monday, Sept. 26)

9. Find (and distribute) a journal or newspaper article or internet report about a study *that specifies the values* of a quantitative variable being compared for two or more categorical groups. If the design was two-sample, discuss whether or not a paired study would have been feasible. Tell how the data were produced—was it an experiment or an observational study? Display and summarize the data and tell which group apparently has the highest values for the quantitative variable and which has the lowest. Tell if results are being reported for a sample or for an entire population. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values.

10. Find (and distribute) a journal or newspaper article or internet report about a study that involves two categorical variables. Tell how the data were produced—was it an experiment or an observational study? Which variable is explanatory and which is response? Report which group apparently has the highest percentage in the response of interest and which has the lowest. Tell if results are being reported for a sample or for an entire population, and the number of individuals studied, if known. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values. If there is enough information provided, construct a two-way table of counts and display the information with a bar graph.

11. Find (and distribute) a journal or newspaper article or internet report about a study that involves two quantitative variables. Tell how the data were produced—was it an experiment or an observational study? Which variable is explanatory and which is response? Is there a positive or negative relationship between the two variables?

Does the report suggest that the relationship is strong or weak? If the data values are specified, use software to report the correlation, and comment on its sign and size. Tell if results are being reported for a sample or for an entire population, and the number of individuals studied, if known. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values.

Thursday, Oct. 6 (provide url or hardcopies by Monday, Oct. 3)

12. Find (and distribute) a journal or newspaper article or internet report about a study that *specifies the values* for two quantitative variables. Tell how the data were produced—was it an experiment or an observational study? Which variable is explanatory and which is response? Is there a positive or negative relationship between the two variables? Does the report suggest that the relationship is strong or weak? Report the equation of the regression line and interpret the slope, intercept, and value of s in this context. Tell if results are being reported for a sample or for an entire population. Tell if there is reason to suspect bias that arises if the sample is not representative, or bias in the design for assessing the variable's values.

13. This student-led activity gives students practice using the Non-Overlapping "Or" Rule, the Independent "And" Rule, and the "Not" Rule. Students should begin by thinking about this question: What is the chance that at least two people in a group of 60 share the same birthday? At their seats they write down a guess; then the presenter sketches a quick histogram of their guesses, with ranges of 0 to 0.1, 0.1 to 0.2, all the way up to 0.9 to 1.0. Their guesses will vary considerably, showing that intuition is not very helpful in this case. Using notes provided by the instructor, the presenter walks the class through the solution process. He or she can finish the activity by going through the months one at a time and asking students to raise their hand if they themselves or their mother or father was born in that month, and if so, tell the date, in order to discover if there are at least two out of 60 with the same birthday.

14. This activity gives students practice solving for probabilities with the aid of two-way tables or tree diagrams.

    Materials: the student leader should be prepared with 3 envelopes, preferably of three different colors, and labeled A, B, C. One of the envelopes hides a prize, for example a dollar bill. The leader plays the role of game show host, and chooses a volunteer student to play the role of contestant. The host explains that a prize is hidden in one of the three envelopes A, B, or C. After the student picks one of the three, the host offers the student a "golden opportunity": host reveals one of the remaining two envelopes, showing no prize. Host gives the student a chance to switch to the envelope he or she did not select originally. After the student makes the decision to "keep" or "switch", ask the rest of the class if this was a wise choice: Is the probability of winning the prize higher with the "keep" strategy, or the "switch" strategy, or perhaps the same for both? Students will almost surely differ in their opinions, confirming that this is a probability problem for which our ordinary intuition tends not to be adequate. At the board the host can show the class how to contruct probability trees to calculate the probability of winning using each strategy.

Friday, Oct. 7 (provide url or hardcopies by Monday, Oct. 3)

14(b). **Activity: Can Humans Mimic Random Behavior?** Make a 15-minute presentation on this topic, including a discussion of the game Rock, Paper, Scissors.

Thursday, Oct. 20 (provide url or hardcopies by Monday, Oct. 17)

15. This activity helps students see rules for means and variances in action. The sampling and calculations should be done in advance, and results presented to the class with a handout.

Access survey data in

`http://www.pitt.edu/~nancyp/bigpicture/surveydata.txt`

that accompanies this book.

Add to the existing columns Height, Shoesize, and Age by creating additional columns for height with a 1-inch heel “HeightPlus1”, height in centimeters “2.5TimesHeight”, sum of height and shoesize “HeightPlus Shoe”, and sum of height and age “HeightPlusAge”.

Now find means and standard deviations of Height, Shoesize, Age, HeightPlus1, 2.5TimesHeight, HeightPlusShoesize, HeightPlusAge. Verify the following, and explain:

(a) The mean of the column “HeightPlusOne” **equals** one plus the mean of the column “Height” .

(b) The standard deviation of the column “HeightPlusOne” **equals** the standard deviation of the column “Height”.

(c) The standard deviation of the column “2.5TimesHeight” **equals** 2.5 times the standard deviation of the column “Height”.

(d) The mean of the column “HeightPlusShoesize” **equals** the mean of the column “Height” plus the mean of the column “Shoesize”.

(e) The mean of the column “HeightPlusAge” **equals** the mean of the column “Height” plus the mean of the column “Age”.

Calculate variances of Height, Shoesize, Age, HeightPlusShoesize, and HeightPlusAge by squaring standard deviations.

Discuss whether height and shoe size should be dependent or independent for college students; discuss whether height and age should be dependent or independent for college students. Create plots of heights versus shoe sizes and heights versus ages and discuss whether or not each pair of variables appears to be independent.

Explain why these would *not* be equal, and verify that they are quite different: The variance of the column “HeightPlusShoesize” **does not equal** the variance of the column “Height” plus the variance of the column “Shoesize”.

Explain why these should be approximately equal, and verify that they are close: The variance of the column "HeightPlusAge" **approximately equals** the variance of the column "Height" plus the variance of the column "Age".

In Chapter 11 we will establish a formula for standard deviation of sample mean $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$:

$$Variance[(X_1+X_2+\cdots+X_n)/n] = (Variance[X_1]+Variance[X_2]+\cdots+Variance[X_n])/n$$

Discuss why this cannot be claimed if the sampled values X_1 to X_n are *dependent*. Discuss why inference methods that rely on this formula require the population to be at least 10 times the sample size when sampling without replacement.

16. This activity helps students explore the behavior of sample proportion in the binomial setting.

    Each student needs a coin to flip; alternatively, each student can roll a die and count 1, 2, 3 as tails and 4, 5, 6 as heads.

    Discuss and predict center (mean), spread (standard deviation) and shape of the distribution of proportion of heads in (a) 1 coinflip (b) 4 coinflips (c) 20 coinflips. [Use values of n and p, and formulas for mean and standard deviation of sample proportion $\hat{p}$.]

    (a) Each student flips a coin 1 time and reports aloud the proportion of heads as being 0 or 1. The presenter constructs a histogram of the class's sample proportions so shape can be discussed; mean and standard deviation of the class's sample proportions are found (either by everyone in the class or by the presenter) with a calculator, and these are compared to the exact mean and standard deviation discussed in (a) above.

    (b) Each student flips a coin 4 times and reports aloud the proportion of heads as being 0, 0.25, 0.5, .75, or 1. The presenter constructs a histogram of the class's sample proportions so shape can be discussed; mean and standard deviation of the class's sample proportions are found, and these are compared to the exact mean and standard deviation discussed in (b) above.

    (c) (If time permits.) Each student flips a coin 20 times and reports aloud the proportion of heads as being 0, 0.05, 0.10, etc. up to 1. The presenter constructs a histogram of the class's sample proportions so shape can be discussed; mean and standard deviation of the class's sample proportions are found, and these are compared to the exact mean and standard deviation discussed in (c) above.

    Which of the three histograms has a shape closest to normal? Would the activity's results be any different if the class size were much larger? Discuss. Students may also discuss to what extent their histograms would differ if they recorded counts instead of proportions.

17. (With my permission, you may substitute alternate Activity #36.) This computer activity helps to reinforce the 68-95-99.7 Rule as well as more specific information about tails of the normal curve [which we can call the "90-95-98-99 Rule"].

Access data from the book's existing data file about several hundred students surveyed previously; specifically, work with the column of 391 Verbal SAT scores, whose distribution is quite close to normal, with mean 591.84 and standard deviation 73.24. To standardize these values, create a column VerbalZ from the original column Verbal by subtracting the mean 591.84 and dividing by the standard deviation 73.24. Sort this column from lowest to highest. Produce histograms of the original and standardized scores to verify that centers, spreads, and shapes appear as they should. Then report how many of the 391 z scores are less than -3, less than -2, less than -1, greater than $+1$, greater than $+2$, and greater than $+3$. Using this information, calculate the proportion of standardized values between -1 and $+1$; between -2 and $+2$; between -3 and $+3$. Looking at these proportions, discuss how well the z scores conform to the 68-95-99.7 Rule.

Now discuss the fact that for standard normal values z, 1% exceed 2.576 in absolute value, 2% exceed 2.326 in absolute value, and 10% exceed 1.645 in absolute value. Report how many of the 391 z scores are less than -2.576, less than -2.326, less than -1.645, greater than $+1.645$, greater than $+2.326$, and greater than $+2.576$. Using this information, calculate the proportion of standardized values exceeding 2.576, 2.326, and 1.645 in absolute value, and discuss how well the z scores conform to a perfectly normal distribution.

Thursday, Oct. 27 (provide url or hardcopies by Monday, Oct. 24)

18. Lead an exploration of normal probability plots: The probability that a standard normal variable takes a value below -3 is .0015, the probability of taking a value below -2 is 0.025, below -1 is 0.16, below 0 is 0.5, below $+1$ is 0.84, below $+2$ is 0.975, and below $+3$ is 0.9985. If we plot the probabilities 0.0015, 0.025, 0.16, 0.5, 0.84, 0.975, 0.9985 against the values $-3, -2, -1, 0, 1, 2, 3$, then the result is a curve that starts flat, rises steeper, then flattens again. It is possible to re-scale the vertical axis to make the plot linear. The z-values and probabilities mentioned above are just the "tip of the iceberg": if we plot any set of standard normal values along the x-axis, and re-scaled probabilities on the y-axis, the result should be a straight line. For that matter, they need not be standard normal: if they are unstandardized, there's just a different scale on the horizontal axis.

    If, on the other hand, the data do not conform well to a normal distribution, the graph of re-scaled normal probabilities versus data values will deviate noticeably from a straight line. Consider the impact on our plot if the data are heavy-tailed, or light-tailed, or left-skewed, or right-skewed...

    Statistical software packages include a test for normality like the one described above, and we start with a visual test: a plot that appears linear suggests that the data set is normal.

    There are also a variety of test statistics and accompanying P-values. Consider the role of sample size in the size of the P-value...

    The activity leader first discusses with the class the variables in the survey data set, such as Math and Verbal SAT scores, height, age, hours slept, time spent on the telephone, etc. Together they hypothesize which variables should be approximately normal, which will be left-skewed, and which would be right-skewed, listing each variable in one of three columns on the blackboard to keep track. Then all students access the data set and perform a normality test for one variable at a time, commenting on how well or badly the normal plot shows the data to conform to a normal distribution, and whether or not their original guesses were correct.

19. **Distribution of Sample Proportion** This activity helps students understand the sampling distribution of sample proportion, including the roles played by sample size and underlying population proportion. The task is to consider the center (mean), spread (standard deviation), and shape of the distribution of sample proportion for random samples of

    (a) size $n = 10$ from a population of students with proportion $p = .5$ living on campus

    (b) size $n = 40$ from a population of students with proportion $p = .5$ living on campus

    (c) size $n = 10$ from a population with proportion of ambidextrous people $p = .03$

    (d) size $n = 40$ from a population with proportion of ambidextrous people $p = .03$

    Complete activity based on Lecture 19 detailed in

    `http://www.pitt.edu/~nancyp/stat-1000/seminars19-20.html`

    and report your findings to the class, showing relevant summaries and graphs.

20. **Distribution of Sample Mean** This activity helps students understand the sampling distribution of sample mean, including the roles played by sample size and center, spread, and shape of the underlying population. The task is to consider the center (mean), spread (standard deviation), and shape of the distribution of sample mean for random samples of

    (a) size $n = 10$ from a fairly normal population with mean 610, standard deviation 72

    (b) size $n = 40$ from a fairly normal population with mean 610, standard deviation 72

    (c) size $n = 10$ from a skewed population with mean 3.776, standard deviation 6.503

    (d) size $n = 40$ from a skewed population with mean 3.776, standard deviation 6.503

    Complete activity based on Lecture 20 detailed in

    `http://www.pitt.edu/~nancyp/stat-1000/seminars19-20.html`

    and report your findings to the class, showing relevant summaries and graphs.

Thursday, Nov. 3 (provide url or hardcopies by Monday, Oct. 31)

21. Find (and distribute) an article or report that includes mention of sample size and summarizes values of a categorical variable with a count, proportion, or percentage. Based on that information, set up a 95% confidence interval for population proportion in the category of interest. If the article reports a margin of error, tell whether or not it is consistent with the one you calculated. Discuss sampling and study design. Be prepared with several such reports, to be discussed if time permits.

22. This activity helps students understand the meaning of confidence levels and of significance levels in setting up intervals for, and tests about, unknown population proportion.

    Each student randomly samples a teaspoon of M&Ms and uses the sample proportion of blues to construct a 95% confidence interval for "unknown" population proportion that are blue. This is done several times by each student so that altogether the class has constructed one hundred 95% confidence intervals. (For example, if there are 20 students in the class, each should construct 5 confidence intervals.) Then the presenter checks to see what count, then percentage, of 95% confidence intervals contain the actual population proportion that are blue, 0.17. In the context of this activity, students should discuss the meaning of a 95% confidence interval.

    Now each student uses his or her sample proportions to carry out several tests of $H_0 : p = 0.17$ against the two-sided alternative $H_a : p \neq 0.17$. Altogether, the class as a group should carry out 100 tests. Then the presenter checks to see what count, then percentage, of the students' tests have a P-value less than 0.05. In the context of this activity, students should discuss the meaning of rejecting a null hypothesis with a P-value less than 0.05 (or other level).

    Discuss the relationship between confidence intervals and hypothesis test results.

    Finally, the presenter should explain how this activity differs from the activity carried out in class for Lecture 19.

23. Find (and distribute) an article or report that includes mention of a confidence interval or P-value concerning a population proportion when there is a single categorical variable of interest. If sample proportion and sample size are mentioned, construct the confidence interval or carry out the test yourself to verify the reported results. Discuss sampling and study design. Discuss the implications of the study's conclusions.

Thursday, Nov. 10 (provide url or hardcopies by Monday, Nov. 7)

24. [Alternative: create a computer activity...] This activity helps students understand the meaning of confidence levels in setting up intervals for unknown population mean.

    Materials needed: many ordinary (6-sided) dice: either 1 per student or 8 for each row of students; students' own calculators for finding sample means.

    Each student rolls a die 8 times (or rolls 8 dice and passes them on to the next student). Each averages the 8 numbers with the aid of a calculator, records this sample mean $\bar{x}$, and uses it to construct 68%, 95%, and 99.7% confidence intervals for "unknown" population mean μ. This should be done using known standard deviation $\sigma = 1.7$ (as well as sample size $n = 8$).

Meanwhile, the presenter reviews the following: An approximate

- 68% confidence interval for μ is $\bar{x} \pm 1(\frac{\sigma}{\sqrt{n}})$
- 95% confidence interval for μ is $\bar{x} \pm 2(\frac{\sigma}{\sqrt{n}})$
- 99.7% confidence interval for μ is $\bar{x} \pm 3(\frac{\sigma}{\sqrt{n}})$

For example, suppose 8 dice are rolled. Since σ for the population of rolls 1,2,3,4,5,6 is 1.7, an approximate

- 68% confidence interval for μ is $\bar{x} \pm 1(\frac{1.7}{\sqrt{8}}) = \bar{x} \pm 0.6$
- 95% confidence interval for μ is $\bar{x} \pm 2(\frac{1.7}{\sqrt{8}}) = \bar{x} \pm 1.2$
- 99.7% confidence interval for μ is $\bar{x} \pm 3(\frac{1.7}{\sqrt{8}}) = \bar{x} \pm 1.8$

Suppose a student rolls 8 dice and gets a sum of 23, so $\bar{x} = \frac{23}{8} = 2.875$. Then a

- 68% confidence interval for μ is $2.875 \pm 0.6 = (2.275, 3.475)$
- 95% confidence interval for μ is $2.875 \pm 1.2 = (1.675, 4.075)$
- 99.7% confidence interval for μ is $2.875 \pm 1.8 = (1.075, 4.675)$

The 68% confidence interval either captures the population mean μ or it does not. We happen to know that μ in this case is 3.5, so the interval (2.275, 3.475) did *not* capture it. Somebody else's 68% confidence interval might well contain 3.5. In the long run, 68% of such intervals should succeed in capturing μ. If 20 people each roll 8 dice and set up a 68% confidence interval for μ based on each sample mean $\bar{x}$, then about 14 of them should succeed in capturing μ—that is, roughly 14 of those intervals will contain 3.5. Similarly, about 95%—roughly 19—of their 95% confidence intervals should contain 3.5, and it's practically guaranteed that all of their 99.7% confidence intervals will contain 3.5.

Now the presenter checks to see what count, then percentage, of 68%, 95%, and 99.7% confidence intervals contain the true population mean, which we happen to know is 3.5.

Finally, based on the 95% confidence interval results, predict how many tests, based on sample mean rolls, would reject $H_0 : \mu = 3.5$ at the 5% level.

25. Find (and distribute) a journal article or internet report about a study that involves just a single quantitative variable for a large sample size, where a z confidence interval or hypothesis test result has been reported. Explain all the steps in constructing the interval or carrying out the test. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the implications of the study's conclusions.

26. Find (and distribute) a journal article or internet report about a study that involves a single quantitative variable *whose values are specified*, where a t confidence interval

or hypothesis test result has been carried out. Explain all the steps in constructing the interval or carrying out the test. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the implications of the study's conclusions.

Thursday, Nov. 17 (provide url or hardcopies by Monday, Nov. 14)

27. This activity helps students understand the meaning of confidence levels and of significance levels in setting up intervals for, and tests about, unknown population mean.

    Each student accesses data from the book's existing data file about several hundred students surveyed previously; specifically, they will work with the column "Sleep" of hours slept the night before students were surveyed, whose mean is 7.1 (hours). Each student randomly samples 40 values from that column, puts them in a column SleepSample and constructs a 95% confidence interval for "unknown" population mean hours slept based on the sampled values. This is done several times by each student so that altogether the class has constructed one hundred 95% confidence intervals. Then the presenter checks to see what count, then percentage, of 95% confidence intervals contain the population mean 7.1. If there is time, students also construct several 90% and 99% confidence intervals to see what percentage contain the mean (7.1) in each case. Next, each student uses the columns of sampled sleep times to carry out several tests of $H_0 : \mu = 7.1$ against the two-sided alternative $H_a : \mu \neq 7.1$. Altogether, the class as a group should carry out 100 tests. Then the presenter checks to see what count, then percentage, of the students' tests have a P-value less than 0.05. If there is time, the presenter can also check to see the count and percentage of tests with a P-value less than 0.10 and less than 0.01.

    In the context of this activity, students should discuss the meaning of a 95% (or other) confidence interval and the meaning of rejecting a null hypothesis with a P-value less than 0.05 (or other level).

28. (With my permission, you may substitute alternate Activity #38.) Find (and distribute) a journal or newspaper article or internet report about a study that uses a paired t procedure to draw conclusions about the population mean of differences, based on information from sampled differences. Explain all the steps in carrying out the procedure. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the consequences of committing Type I or II Errors in this context.

    Alternatively, find a journal or newspaper article or internet report about a study that reports pairs of measurements for each individual, such as a before-and-after study. Use software to carry out a test about the population mean of differences. Report a confidence interval for population mean of differences and discuss the extent to which the categorical explanatory variable impacts the quantitative response.

29. Find (and distribute) a journal or newspaper article or internet report about a study that uses a two-sample t procedure to compare mean values of some quantitative variable for two populations, based on information from a sample (or samples). Explain

all the steps in carrying out the procedure. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the consequences of committing Type I or II Errors in this context.

Monday, Dec. 5 (provide url or hardcopies by Monday, Nov. 28)

30. Find (and distribute) a journal or newspaper article or internet report about a study that uses an ANOVA (F) test to compare mean values of some quantitative variable for more than two populations, based on information from a sample (or samples). Explain all the steps in carrying out the procedure. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the consequences of committing Type I or II Errors in this context.

31. (With my permission, you may substitute alternate Activity #37.) Find (and distribute) a journal or newspaper article or internet report about a study that uses an ANOVA (F) test to compare mean values of some quantitative variable for more than two populations, based on information from a sample (or samples). Explain all the steps in carrying out the procedure. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the consequences of committing Type I or II Errors in this context.

32. Find (and distribute) a journal or newspaper article or internet report about a study that uses a chi-square test to explore the relationship between two categorical variables, based on information from a sample (or samples). Explain all the steps in carrying out the procedure. Explain whether there is reason to suspect bias because the sample is not representative of the larger population of interest. Discuss the implications of the study's conclusions, and the possible consequences of committing Type I or II Errors in this context.

Thursday, Dec. 8 (provide url or hardcopies by Monday, Dec. 5)

33. Find (and distribute) a journal or newspaper article or internet report about a study of a drug's effectiveness, where a Type I Error is believed to have been committed. Find another about a study of a drug's side effects, where a Type II Error is believed to have been committed. Discuss the consequences of these errors. If details about the studies are provided, such as sample sizes, sample counts or proportions, etc., discuss these.

34. Find (and distribute) a journal article or internet report about a study that uses inference to draw conclusions about the relationship between two quantitative variables for a larger population, based on the relationship observed in a sample. Explain whether there is reason to suspect bias because of samples not being representative of the larger population of interest. Report as many details as possible: correlation, slope and/or regression line equation, test statistic, degrees of freedom, and P-value. Discuss the implications of the study's results.

35. Find (and distribute) a journal article or internet report about a study concerning the relationship between two quantitative variables, where the data values themselves are reported. Explain whether there is reason to suspect bias because of samples not being representative of the larger population of interest. Use software to set up prediction and confidence intervals for at least two values of the explanatory variable. Discuss the difference between prediction and confidence interval widths, and how interval centers and widths are affected by which value the explanatory variable takes. Calculate rough estimates for the prediction and confidence intervals by hand ($\hat{y} \pm 2s$ and $\hat{y} \pm 2s/\sqrt{n}$), and tell how the sample size plays a role in how good or bad these estimates are. (With my permission, you may report on the Halberda/Mazzocco article in Nature Oct 2008 about number acuity.)

36. **Alternate Activity: Sampling Without Replacement** This activity explores the standard deviation of the distribution of sample mean when sampling without replacement, first from a population that satisfies our Rule of Thumb (size at least 10n) and then from a population that is not much larger than the sample. For this activity, the class needs to access three normally distributed populations, each with approximate mean 50 and standard deviation 5. Their sizes are 1000, 150, and 100, respectively. They are posted as a text file on the course website, after the student survey data link. Once everyone has them downloaded into MINITAB, they should use descriptive stats to verify the means and standard deviations as stated above.

    Now, each student uses MINITAB to randomly sample 100 rows from normalpop1000 and store them in a new column, named samplefrom1000. (Sample WITHOUT replacement: do NOT check the sample with replacement option box.) Then everyone finds the mean of their sample and tells the activity leader. The leader enters those 20 means in the computer at the front, and finds the mean and standard deviation of those sample means obtained from taking samples of size 100 from a population of size 1000 (ten times as large).

    Next, each student samples 100 rows from normalpop150 and stores them in a column named samplefrom150 (again without replacement). Again they report their means, the leader enters them in the computer and finds the mean and standard deviation of those sample means. Note that now the population isn't much larger than the sample size.

    Finally, each student samples 100 rows from normalpop100, and so on. Can they anticipate what will happen with the mean and standard deviation this time?

37. **Alternate Activity: Role of I and N in F test results** This activity explores how the number of groups and the total sample size affect the distribution of F and also how they affect the size of the F statistic.

    First we consider how I and N impact the distribution of F.

    Suppose the sample size is $N = 20$. These individuals might be divided into 2 groups of 10, or 4 groups of 5, or 5 groups of 4, or 10 groups of 2. Use Calc>Probability Distributions>F to find the inverse cumulative probability for the correct degrees of freedom in the numerator (DFG) and in the denominator (DFG) in each of the four abovementioned situations (I=2, 4, 5, or 10), inputting 0.95 as the constant. This identifies the F value that yields 0.05 as the tail probability, and tells us how large our F statistic must be to let us reject at the 0.05 level.

In general, as far as the spread of the distribution of F is concerned, is it easier to reject the null hypothesis when our sample is divided into few groups or many groups?

Now suppose the sample size is $N = 40$. These individuals might be divided into 2 groups of 20, or 4 groups of 10, or 10 groups of 4, or 20 groups of 2. Use Calc>Probability Distributions>F to find the inverse cumulative probability for the correct degrees of freedom in the numerator (DFG) and in the denominator (DFG) in each of these situations, again inputting 0.95 as the constant.

As far as the spread of the distribution of F is concerned, is it easier to reject the null hypothesis when our sample is divided into few groups or many groups? (Do you see the same trend for $N = 40$ as you saw for $N = 20$?)

As far as the distribution of F is concerned, is it easier to reject the null hypothesis for a smaller total sample size N or a larger N?

Next we consider how I and N impact the size of the F statistic.

Our formula for F can be rearranged to show that SSG/SSE is being multiplied by $\frac{N-I}{I-1}$. For $N = 20$, does this quantity increase or decrease as I increases from 2 to 4 to 5 to 10? As far as the size of F is concerned, is it easier to reject the null hypothesis when our sample is divided into few groups or many groups?

For $N = 40$, does the quantity increase or decrease as I increases? (Do you see the same trend for $N = 40$ as you saw for $N = 20$?)

As N increases from 20 to 40, does $\frac{N-I}{I-1}$ increase or decrease? As far as the size of F is concerned, is it easier to reject the null hypothesis for a smaller total sample size N or a larger N?

In summary, can we say that our P-value tends to be smaller when the data are divided into few or many groups I? Can we say that our P-value tends to be smaller for smaller or larger total sample size N?

38. **Alternate Activity: Two-Sample vs. Paired Procedures** This activity demonstrates the importance of distinguishing between paired and two-sample studies.

    Access the survey data for Fall 2003 and unstack heights, mom's heights, and dad's heights by the variable sex of student. Is there a significant difference between mean height of female students and mean height of the mothers? Decide whether the test should be paired or two-sample, then carry it out and state conclusions. Next, carry out the wrong test and see if the same conclusions are reached. Likewise, test for a difference between mean heights of male students and fathers, then compare results to those for the incorrect procedure.

    If there is a subtle difference, are we better equipped to detect it with a paired or with a two-sample test?

39. **Alternate Activity: Lead Lab 5** Make a 5-minute presentation on the relationship between smoking and exercise, summarizing statistical evidence to date. Lead the class in completing Lab 5 on the relationship between these variables based on data from our surveyed Pitt students.

40. **Alternate Activity: Lead Lab 6** Make a 5-minute presentation on the relationship between age and hours slept, summarizing statistical evidence to date. Lead the class in completing Lab 6 on the relationship between year and sleep time based on data from our surveyed Pitt students.

41. **Alternate Activity: Lead Lab 7** Make a 5-minute presentation on the relationship between weight and time spent watching television, summarizing statistical evidence to date. Lead the class in completing Lab 7 on the relationship between these variables based on data from our surveyed Pitt students.

42. **Alternate Activity: Lead Lab 8** Make a 5-minute presentation on the relationship between weight and hours slept, summarizing statistical evidence to date. Lead the class in completing Lab 8 on the relationship between these variables based on data from our surveyed Pitt students.

43. **Alternate Activity: Power: a test's ability to detect (i.e. reject) a false null hypothesis**

Consider our several hundred surveyed students to be a **population** whose Math SAT scores have $\mu = 610.44$, $\sigma = 72.14$.

(a) Sample 100 with replacement and test $H_0 : \mu = 590$ vs. $H_a : \mu > 590$ at the $\alpha = 0.05$ level. Notice that the alternative is in fact correct. What is the probability of correctly rejecting H_0?

In general, we reject H_0 in favor of $H_a : \mu > \mu_0$ at the 0.05 level when our standardized statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is greater than 1.645. In this case, we reject when

$$\frac{\bar{x} - 590}{72.14/\sqrt{100}} > 1.645 \rightarrow \bar{x} > 590 + \frac{1.645(72.14)}{\sqrt{100}} = 601.87$$

What is the probability of getting a sample mean greater than 601.87, given $\mu = 610.44$ and $\sigma = 72.14$? Find

$$P(\bar{X} > 601.87 \text{ given } \mu = 610.44) = P(Z > \frac{601.87 - 610.44}{72.14/\sqrt{100}}) = P(Z > -1.19) = 0.883$$

There is an 88% chance, in this case, that our hypothesis test will correctly favor the claim that $\mu > 590$: this is the power of the test.

Activity: have each student randomly sample 100 *with* replacement from all the Math SAT scores, store them in column "sample1", and carry out a z test as outlined above. Determine what percentage get a P-value small enough to reject—that is, less than 0.05.

(b) What if $n = 10$ instead of $n = 100$? Should we have more or less power with a smaller sample size? Discuss, calculate, carry out the activity based on samples of size 10, stored in column "sample2".

(c) What if we test $H_0 : \mu = 605$ vs. $H_a : \mu > 605$ instead of $H_0 : \mu = 590$ vs. $H_a : \mu > 590$? Should we have more or less power now? Discuss, calculate, carry out the activity using "sample1" but specifying the new alternative.

(d) What if we increase α from 0.05 to 0.10? Should we have more or less power now? In this case, the critical z changes from 1.645 to 1.28. Discuss, calculate, review the P-values from the first test to see how many are less than 0.10 instead of less than 0.05.

Summarize: power increases with smaller or larger sample sizes? with alternatives that are further from or closer to the true μ? with smaller or larger cut-off level α?