INSTRUCTOR NOTES FOR M358K FOR PART V
(FROM THE DATA AT HAND TO THE WORLD AT LARGE: CHAPTERS 18-22) OF DEVEAUX, VELLEMAN AND BOCK, *STATS: DATA AND MODELS*, 3RD ED.

Chapter 18: Sampling Distribution Models

- Note that this textbook discusses sampling distributions for proportions before sampling distributions before means (and then inference for proportions before inferences for means). Traditionally, inference for means has appeared in textbooks before inference for proportions, but many instructors have come to believe that students learn better if these topics are presented in the opposite order. I think the crux of the matter may be that students don't have to deal with the t-distribution so soon.
- Don't rush this chapter -- it lays the foundation for inference and requires students to look at things in a new way.
- Students may have seen a different form of the Central Limit Theorem in M 362K, so the section title on p. 431 may cause some confusion and/or warrant some explanation. Indeed, there are many theorems misleadingly referred to as "The" Central Limit Theorem.
- Draw attention to the footnote on p. 432; this points out an area of confusion for many students.
- Students may find the ActivStats applet Learn About the Sampling Distribution of a Proportion helpful in understanding the concept of sampling distribution.
- The web demo Sampling Distributions at http://www.prenhall.com/agresti/applet_files/samplingdist.html can show sampling distributions for several types of population distribution and a couple of relevant statistics (including counts for binary distributions). See instructions on website.
- pp. 432 – 436 seem a little out of optimal order: On pp. 433 (informally) and 434 (more formally, in the math box), the sample proportion $\hat{p}$ is regarded as a random variable, but it is not explicitly called such until p. 436. So you may want to lecture on this, but in a slightly different order. I suggest:
  1. Giving a discussion such as pp. 431 – 432, but pointing out that we are now considering the sample proportion $\hat{p}$ as a random variable. Using the characterization of a random variable as a variable whose value depends on a random process is helpful here: The random process for $\hat{p}$ (considered as a random variable) is "pick a random sample."
  2. Go through the calculations for the mean and standard deviation of $\hat{p}$ (done fuzzily on p. 433 and more rigorously in the math box on p. 434), pointing out that talking about the mean and standard deviation of $\hat{p}$ makes sense when we consider $\hat{p}$ as a random variable.
  3. You might at this point introduce the concepts of *estimator* and *unbiased* estimator: (Note: These are also introduced in the supplement "Why Does the Sample Variance Have n-1 in the Denominator?")
     - An *estimator* is a random variable whose underlying random process is choosing a sample, and whose value is a statistic (as

defined on p. 285), based on that sample, that is used to estimate a population parameter. So $\hat{p}$ (considered as a random variable) is an estimator of p, the population proportion.

- Note that the concepts of *estimate* and *estimator* are related but not the same: a particular value (calculated from a particular sample) of the estimator is an estimate. The distinction is similar to the distinction between a function and the value of the function when evaluated at a particular number.
- An *unbiased estimator* is one with the property that E(estimator) = the parameter being estimated. Point out that intuitively, this seems like a desirable property for an estimator*.
- The calculation E($\hat{p}$) = p on p. 434 shows that $\hat{p}$ is an unbiased estimator of p.

4. Then continue from the bottom of p. 434 (How Good is the Normal Model) through p. 336, taking note of the following:
   - The blue box on p. 435 points out a common source of confusion.
   - See the supplement *Where Does the 10% Condition Come From?* for more detail on the 10% conditions mentioned on pp. 407, 435, and 441.**
   - You might also consider introducing the intuitive idea of a consistent estimator (one, such as the sample mean, that gets better and better as the sample size increases)
   - You might also consider introducing Robert Kass' "big picture" of statistical inference (or some variant thereof), from p. 3 of http://www.stat.cmu.edu/~kass/papers/bigpic.pdf

- pp. 434, 442, and 443 As pointed out in the notes on Chapter 16, the authors depart here from the practice they articulated in Chapter 16 of using capital letters for a random variable and small letters for the values of random variables. I recommend adopting the policy stated in Chapter 16, since I find it helps students understand the sampling distribution better.
- p. 437 A nice example. You might consider assigning this as reading and going through a similar example in class. The Just Checking is also good.
- pp. 348 – 439 The Step-by-Step Example provides a good example of statistical thinking.
- pp. 439 – 440 This section seems well done. The disk activity Learn About Sampling Distributions of Means may be helpful for some students.
- p. 440-441
  - *Caution*: Students may have seen a different form of the Central Limit Theorem in M 362K, so be sure to mention that there are several (related) theorems that are called the Central Limit Theorem.
  - One more general form of the CLT that is worth mentioning is that linear combinations of independent random variables (and note that the mean is such, if we have sampling with replacement or from an infinite population) are (if the distributions aren't too weird) approximately normal. This helps explain why so many distributions that occur in nature are approximately normal. For example, height is the sum of heights of

lots of smaller body parts, which might be close enough to independent to be expected to add up to an approximately normal distribution.***

- Some students find a non-computer simulation helpful here. One that is easy is the Birthday Simulation: Have a transparency (or sheet of paper to show on the do-cam) prepared on which every student marks the day of the month on which they and each of their parents were born (they can substitute siblings, friends not in the class, etc. if they don't know their parents' birthdays), and then another transparency on which each student marks the average of the three days they marked on the first transparency (or do the histograms on the board). Then discuss the different shapes, and why the histogram of averages reasonably should be more mound shaped: To get an average toward the beginning of the month, all three birthdays would have to fall at the beginning of the month, which would be very unusual. On the other hand, there are lots of ways to get an average lying in the middle of the month, so those would be more common as averages.
- The Rice Virtual Lab in Statistics Sampling Distribution Demo is a convenient way to show sampling distributions for a variety of distributions, including ones students might suggest. See the External Links section of http://www.ma.utexas.edu/users/mks/M358KInstr/M358KInstructorMaterials.html for cases to try where the CLT appears to be iffy.
- See also the link to Mary Parker's CLT simulations.
- See the supplement *An Example where the Central Limit Theorem Fails* if you or any of your students are interested in footnote 9 on p. 440.

- pp. 442 -443
  - The text just states that the sampling distribution of the mean " is centered at the population mean." For this class, prove (or assign as an exercise to prove) that the expected value of the sample means is the population mean – and thus the sample mean is an unbiased estimator of the population mean.
  - The notation SD($\bar{y}$) or $\sigma(\bar{y})$ is helpful: Many people use "standard error" to refer to *both* the standard deviation of the sampling distribution and the approximation to it obtained by substituting the sample standard deviation s for σ, which causes confusion that often leads to misunderstanding.
  - Note that the authors have abandoned the convention they introduced in Chapter 16 of using capital letters for random variables and small letters for the values of these random variables. I recommend that you instead follow the convention introduced earlier – it helps some students see the distinction between a statistic and an estimator.
- pp. 444-445 These are nicely chosen and well-explained examples. Some students may have seen similar examples in M 362K, but most likely without careful checking of model assumptions, so it is worth assigning some exercises along similar lines.
- 445 – 446 This section (including the box on p. 446) points out some important practical considerations.
- p. 447 The "Just Checking" is good, and the last section provides a good summary.

- p. 448 The “What Can Go Wrong” and Connections are important.
- This might be a good time to explain why the sample standard deviation has n-1 rather than n in the denominator; see the supplement “Why Does the Sample Variance Have n-1 in the Denominator?”.
- Suggestions for Exercises (pp. 450 – 456): #1, 3, 9, 15, 27, 29, 31, 35 as self-check; #18, 20, 36, 40 to hand in. You might also want to assign the Exercise in the supplement “Why Does the Sample Variance Have n-1 in the Denominator?”

Footnotes for Chapter 18 notes:

*However, as you may or may not know, there are times when a biased estimator with small variance is actually preferable to an unbiased estimator.

** *Of possible interest to you or some of your students*: One current active area of research in statistics is developing good methods for analyzing large-scale data sets, such as microarray data from gene studies. In these data sets, cases are typically correlated rather than independent. Like sampling without replacement, this creates difficulties in calculating variances for quantities of interest. One method that has been developed for dealing with this for one such quantity is an approximate formula for the variance that has one term similar to the variance for binomial probabilities, and a second term (called a *correlation penalty*) that involves the root mean square (i.e., the square root of the mean of the squared correlations) of the correlations between the cases. For more information, see Chapters 7 (Estimation Accuracy) and 8 (Correlation questions) of B. Efron, Large Scale Inference, Cambridge, 2010, or Stats 329: Large Scale Inference

***You might also mention that lots of things in nature arise as products rather than sums, but if we take logs we get sums, so lognormal distributions can be expected to be common. (See, e.g., http://stat.ethz.ch/~stahel/lognormal/websitefor examples.)

Chapter 19: Confidence Intervals for Proportions

- pp. 458 – 461 give a particularly good introduction to confidence intervals.
  - But you might want to supplement with a little algebra and more pictures at the top of p. 459
  - Items 1-5 on p. 459 seem particularly well done.
  - Still, the ActivStats applet Learn How to Estimate a Population Parameter might help some students understand (or at least accept) the point of p. 459: why the concept of confidence interval isn’t as simple as we’d like it to be.
  - Many instructors find they can make this point by bringing in enough small packets of multi-colored candies such as M and M’s, Reeces’ Pieces, etc. for each student to get one packet as their “sample.” (Or else bring a couple of large packages and dole out a pile to each student) Have them count out some number, say 15 (without making choices of color!) Specify a color and have each student calculate a confidence interval, from their sample, for that color.
    - →Before counting and calculations, ask someone to speculate what they think the proportion of that color is.

    Then compare CI’s to get a feel for the variability of CI’s depending on the sample.
    - Also make a histogram of the individual sample proportions to simulate the sampling distribution of the sample proportions.
    - You might then want to pool the samples, to show how a larger sample gives a smaller CI.

        - Ask how many got a CI containing the conjectured proportion.
        - Ask them to write down/remember their proportion, to use for a hypothesis test in Chapter 20 or 21.
    - You may want to point out that the "personal probability" perspective does *not* fit well with confidence intervals (although it does fit with the Bayesian concept of "credible interval")
    - The "Just Checking" on p. 460 is good.
    - The ActivStats applet Compute a Confidence Interval for a Proportion initially looked like it might be helpful, but it had some glitches when I tried it.
    - The ActivStats activity Simulate the Randomness of Confidence Intervals seems OK, but he web demo Confidence Intervals for a Proportion, at http://www.prenhall.com/agresti/applet_files/propci.html, may be better for helping students understand the picture on p. 461.
- pp. 461 – 462 The ActivStats applet Learn About Balancing Precision and Certainty has some merit in explaining the tension between precision and confidence, but there are some problems with it. First, it refers to the sample proportion as p, not p-hat. Second, the marks on the axis are not very good for estimating the ends of the confidence interval.
- pp. 463 – 464 This gives an important recap of things too often neglected in practice. (Comment re the last sentence under "10% Condition" on p. 464: I've encountered students who don't know what "negligible" means.)
- p. 466 Be sure to point out how the formula tells us that a larger sample size gives a smaller confidence interval, with the same level of confidence. (If you did the candy activity, remind them of what happened with the CI when you pooled samples to get a larger sample.)
    - Be sure to point out the information in the blue box – and explain why the assertion in the box is true.
- p. 467 This gives some good examples of the types of thinking that need to go into good statistical practice – how making good decisions means taking the context into account.
- pp. 468 – 470 Emphasize the "What Can Go Wrong" section.
    - I find that in order to really understand what a confidence interval is and is not, most students need more than being told several times. (In part, the human urge to make things simple needs to be countered.) In addition to a web demo (as mentioned above), I often use a class discussion activity such as "What Is a Confidence Interval?" linked under Supplementary Materials. I give it out for students to read and think about before class, then give students a few minutes to discuss what they think with their neighbors, then go over each item in class.
    - As usual, the Violations of Assumptions section (p. 470) is important.
- Suggestions for exercises (pp 473 – 476): Self-chaeck: #3, 5, 7, 9, 11, 27. Written: #8, 12, 18, 24, 28. Also, the handout "What is a Confidence Interval?" for class discussion.

Chapter 20: Testing Hypotheses about Proportions

- *Please note*: This chapter is *not* intended to teach all about hypothesis tests for proportions -- the next chapter continues the task. I recommend that you look over both chapters before starting this one, to see what is covered in which. In brief: this chapter focuses on the general outline of hypothesis testing and setting up and performing hypothesis tests, whereas Chapter 21 goes more into addressing misunderstandings about P-values (as well as Type I and II errors and power).
- pp. 477 - 479
    - Some students might find the combination of the ActivStats activities Testing a Claim, Learn the Reasoning of Hypothesis Testing, Understand Unlikely Outcomes, and Learn the Terminology of Significance Testing helpful.
    - Point out that the top blue box on p. 478 is broad – it includes both the math definition and the statistics/science definition of hypotheses. In the statistical definition, the *null* hypothesis will be used in the sense of a math hypothesis, to deduce other things that will be true when the null hypothesis is true. (The alternate hypothesis will usually be used that way *only* when discussing Type II error and power.)
    - Point out that the "always" in the Notation Alert (p. 478) just applies to this textbook – some people use other notation, such as H_{null}, H_a, H_{alt} or a subscript labeling a value of a parameter.
    - The box "Why is this a standard deviation …" (p. 478) addresses a common source of confusion.
- p. 480
    - Unfortunately, not everyone uses the notation as described in the "Notation Alert," so notation does need to be interpreted in context.
    - The paragraph starting "When the P-value is high …" is important! The lower blue box and the section "What to Do with an "Innocent" Defendant" help reinforce the message.
    - The phrasing of the last sentence of the paragraph starting, "When the P-value is low …" seems rather unclear to me. I usually state the point of this paragraph by saying that if we obtain a low enough P-value, we are faced with three alternatives: Either the model is not good enough, or the null hypothesis is false, or we happen to have one of the rare samples that give unusually low P-values. If we are confident of the model, then the low P-value casts a reasonable doubt on the null hypothesis, so we could then consider rejecting the null hypothesis in favor of the alternate hypothesis to be a rational decision, based on the evidence provided by the data.
- pp. 481 – 484 (The Reasoning of Hypothesis Testing) In general, this section seems well done.
    - Including the terminology "null model" (p. 481) is good, since this terminology seems to be increasingly used in applications.
    - The top blue box on p. 482 is important.

- One important caution: The statement "The conditions for the one-proportion z-test are the same as for the one-proportion z-interval," (in the box in the middle of the p 482) requires an important qualification: That the *hypothesized* proportion p_0 (*not* the sample proportion $\hat{p}$) is used for checking the success-failure condition. (Note that the For Example on p. 482 and the Step-by-Step Example on pp. 488 - 490 do correctly use p_0 for checking the success-failure condition for the hypothesis test.
- The lower blue box on p. 482 is important, but could use some elaboration, namely that there are really three conditions in the conditional probability describing the P-value:
  1. The form of the model fits.
  2. The null hypothesis is true.
  3. Only samples of the same size are considered.
- The last paragraph on p. 483 and the box on the top of p. 484 are important.

- pp. 484 – 486 (Alternate Alternatives)
  - Some students might find the ActivStats activity Learn About the Alternative Hypothesis helpful.
  - The example on p. 484 is good for illustrating the point that the appropriate alternative hypothesis depends on the question being asked – and that different researchers may ask different questions about the same situation, hence use different alternate hypotheses. One consequence: When reading results of a study, check to see whether or not the researchers are asking the same question as you are.
  - The sentence starting "For the same data …" and the following sentences at the bottom of p. 484 may require a little explanation (including what "conservative" means in this context) for the students.
- pp. 486 – 487 (P-Values and Decisions: …) This section is important and generally well done.
  - First paragraph of p. 486: I'd add, "Unfortunately, some people are uncomfortable with the uncertainty of a confidence interval, and prefer hypothesis tests precisely because they have an air of certainty. Don't fall into that trap – when we are using statistical inference, we always have some degree of uncertainty. Using confidence intervals helps us remember that, as uncomfortable as it might be for some people. But also don't fall into the trap of thinking that the confidence interval expresses all the uncertainty: There may still be uncertainty from other sources, such as data collection or a not-good-enough model." (Some math students in particular may have trouble with the uncertainty, if they were attracted to math because they see it as having "right" answers. I recall a student who once asked, "But if we can't be certain, what's the point of doing the hypothesis test?" I don't recall exactly how I responded, but later thought that "Half a loaf is better than none" would have been a good reply.)
  - First paragraph of p. 487: I'd add to the disease-screening example: What if the treatment has serious risks of harm? (For discussion of

risks vs benefits for breast cancer and prostate cancer screening, see the *Significance* articles Breast Cancer Screening: Cons as Well as Pros and A Closer Look at Prostate Cancer Screening)

- pp. 488 – 490
    - The "Just Checking" is good, especially question 6.
    - The Step-by-Step Example provides a good model for students to emulate. Note especially checking the success/failure condition (pp. 488 and 490) and using the appropriate standard error for the confidence interval (p. 290). (You might want to point out that 0.517 is not in the confidence interval and ask if this is surprising or expected.)
- p. 491 Both the "What Can Go Wrong?" and the blue box give guidance for good (and ethical) statistical practice. Unfortunately, these precepts are often not followed (often through ignorance), resulting in lots of literature in many fields that just doesn't stand up under replication. (See, for example, the recent *Nature* article Bad Copy and references therein.)

- Suggestions for exercises (pp. 495 – 498): #1, 3, 5, 7, 9 for self check; #10, 12, and maybe 20 or 30 to hand in.

Chapter 21: More About Tests and Intervals

- pp. 499- 504 These pages emphasize points that are often misunderstood
    - p. 503 Some students might find the ActivStats activity "Perform a Significance Test on the Therapeutic Touch Data" helpful if they still don't understand two-sided p-values. (*Note*: This activity appears to be labeled "Testing Therapeutic Touch" in the textbook, and mislabeled "Hypothesis Tests are Random" in the ActivStats notes. You can find it listed in the ActivStats Table of Contents under Chapter 21, Section 1: More About Inference.)
- pp. 505 – 506 This section seems generally well done.
    - Be sure to note the box in the middle of p. 505.
    - Struggling students might find the ActivStats activity Learn About Rejecting Null Hypotheses helpful.
    - The box at the top of p. 506 and the subsequent paragraph are important.
    - If you did the class activity with candies, and people remember the proportions they got, have each student do a hypothesis test for the conjectured proportion, using the proportion in their sample.Ask how many rejected the null at a .05 level, then ask if that sounds reasonable.
- pp. 507 – 509
    - If you did the candies class activity, you may want to use the confidence intervals they calculated to check out the plausibility of a conjectured proportion of the color in question.
    - p. 508 You might want to mention that there is such a thing as one-sided confidence intervals, which are useful in some situations but beyond the scope of this course.
    - Be sure to have them read (and understand) the Math Box on p. 508
    - Just Checking #3 on p. 509 is particularly good.

- pp. 510 – 511 In case you or one of your students would like to learn more about the various alternative methods for confidence intervals for proportions: The Wikipedia page Binomial Proportion Confidence Interval seems unusually good. (The "exact" method is called the Clopper-Pearson method there.)
- pp. 511 – 513 Making Errors
  - p. 511 Struggling students may find the ActivStats activity Learn About Type I and Type II Errors helpful.
  - If you had students do hypothesis tests with M and M's or some similar candy, talk about the results of their hypothesis tests in terms of Type I error rate.
  - Paragraph starting on p. 512 and finished on p. 513: One may argue that in a jury situation, a Type I error may be worse than a Type II error, since it not only means convicting an innocent person, but may be letting a guilty person go free as well.
  - The top blue box on p. 512 is worth note.
  - I agree with the bottom blue box on p. 512 that calculating sample size (beyond what was done in Chapter 19, and perhaps an analogous calculation for a z-test for a mean) to achieve a desired Type II error rate is beyond the scope of this class. There are online calculators (as well as calculators in most statistical software packages) for power/sample size for many of the common tests. However, for anything sophisticated, simulations are increasingly being used – but have to be done carefully. Some companies have a statistician who spends most of their time doing power and sample size calculations. (For a little discussion and some starting references for those who might be interested, see the bottom section of http://www.ma.utexas.edu/users/mks/statmistakes/powersamplesize.html. Also, http://www.ma.utexas.edu/blogs/mks/2012/06/24/misguided-choice-of-sample-size/ discusses some misguided methods of choosing sample size). But if you want to get or give a little of the flavor of what might be involved, you might try (or have students try) the power curve simulation in Significance Tests for a Proportion, at http://www.prenhall.com/agresti/applet_files/propht.html
  - The "For Example" on p. 513 is nice.
  - The web demo Significance Tests for a Proportion, at http://www.prenhall.com/agresti/applet_files/propht.html, can be helpful here:
    - For example, try n = 30, Null p = 0.5, and True p = 06. Do 1000 simulations (ten clicks, since each click gives 100 samples).
      - Check to see how close the proportion of Type I errors is to the significance level.
      - Also use the displayed results to estimate the Type II error rate, beta.
    - Repeat with a different value of True p (say, 0.51) and compare.

  - It's worth reiterating here that the inevitableness of Type I errors is why replicating studies is important. (The *Nature* article Bad Copy mentioned above is relevant here.)
- pp. 513 – 517 Power, Effect Size, A Picture …, and Reducing Both …
  - These four sections are really all parts of a bigger whole, so definitely should be read and discussed as a whole. Suggestions follow.
  - p. 513 The definition of power given here is somewhat vague. In practice, we can only talk about the power of a test against a specific parameter value, as was discussed for Type II error on p. 512. Often this is called "power against a specific alternative hypothesis". For example, if the null hypothesis is p = .5, then the power of the test against the specific alternate hypothesis H_1: p = .53 would be different from the power of the test against the specific alternate hypothesis H_2: p = .6; a specific alternate hypothesis just corresponds to a specific value of a parameter within the range of the "general" alternate hypothesis H_A (e.g. "p > .5" would be called a "general" alternate hypothesis, for which H_1 and H_2 are examples of specific alternate hypotheses.) http://www.ma.utexas.edu/users/mks/statmistakes/power.html has a picture showing power for two specific alternate hypotheses for a test of means; you can easily draw a similar one for a test of proportions – just make the values on the horizontal axis suitable for proportions. The idea of power against a specific alternative is somewhat implicit in the discussion of Effect Size on p. 514. To make this more explicit: We can't calculate power against a general alternate hypothesis, but we can (at least theoretically) calculate power to detect a specific effect size. Similarly, if we want a test to have a certain power, we need to specify the effect size we want to detect in order to (at least theoretically) calculate a sample size needed to detect that effect with the desired power.
  - p. 513 The ActivStats activity Simulate Hypothesis Tests to See the Effects of Randomness (mentioned on p. 513 of the text as "Hypothesis Tests are Random", and found in the ActivStats Table of Contents in Section 2 under Chapter 21) is pretty good (*except* that it uses a test for means, which hasn't been covered yet). Try the web demo Significance Tests for a Proportion (mentioned above) instead.
  - p. 514 The discussion of Effect Size is generally good.
    - The ActivStats activity Learn About the Power of a Test might be helpful for students – but it would be better if it included a picture such at that on the bottom of p. 515.
    - Revisiting the web demo Significance Tests for a Proportion may help her also.
    - One thing missing from the discussion is that in calculating power or sample size, thought needs to go into choice of what effect size (difference) to detect. For example, it is pointless to try to detect a difference that is smaller than can be measured by the measurement method being used. Similarly, it is pointless to try to

detect a difference that is not of practical importance. For more, see Common Mistakes Involving Power.

- The Just Checking on p. 515 is good.
- The "A Picture …" section (pp. 515 – 516) is good, but note that the caption of Figure 21.3 is not a complete explanation of the picture; p. 516 gives a fuller explanation. Because of this, and just because of the nature of the ideas, many students will need an explanation in class before understanding the picture.
  - Also see comments above for p. 513 and those below.
- pp. 516 – 517 This is pretty well done, but many students will need something more "interactive" to understand. A good oral explanation, constructing the diagram step-by-step with explanation of each step, then drawing attention to which part is being referred to where, might help. The following might also:
  - The ActivStats activity Power and Sample Size could be useful in helping students understand power and the relationship with sample size – but be sure to try it out before hand, to see how it works. It could be better if it actually changed the size of the sampling distribution curve rather than just changing the scale. Also, it deals with a test for means, which hasn't been covered yet.
  - The WISE Statistical Power Applet can also be used to demonstrate the interaction between power, sample size, and specific alternative – but it is for means, not proportions.

- pp. 517 – 518 The What Can Go Wrong and Connections sections are both good.
  - This would be a good time to do a class discussion exercise such as What Is a P-Value? (analogous to the "What Is a Confidence Interval?" exercise for Chapter 19). I would assign it for the students to think about before class, give them a few minutes to discuss with their neighbor in class, then go through it item by item, asking students what they think and augmenting/clarifying/correcting/reinforcing as needed.
- pp. 519 – 520 Good comments on computer output.
- Suggested Exercises (pp. 520 – 524): Self-check: #1,3,5,7,9,13,15,17,19,23. Written: #6, 12, 14, 18, 20, 24. Also the "What Is a P-Value?" exercise (discussed above), for class discussion.

Chapter 22: Comparing Two Proportions (pp. 525 – 549)

- p. 525 "Where are we going?" makes some good points.
- p. 526 Emphasize how the independence assumption is important for getting the formula for variances here. Without independence (or at least, no correlation), we have no reason to believe the formula works.
- pp.527-528 As always, the Assumptions and Conditions are important.
  - Point out the parenthetical note at the end of the For Example on p. 528.
- pp. 528 – 529 The Sampling Distribution
  - The reason given in the blue box is a little vague. For this class, it would be better to refer to the version of the Central Limit Theorem that says that a linear combination of independent normal random variables is

approximately normal. The linear combination in this case is the difference of the two sample proportions, and each sample proportion is an average of the “indicator” random variables for that group. (For example, for the first group, let X_i have value 0 or 1, accordingly as the ith “trial” is a failure or success. Then the sample proportion for that first group, considered as a random variable, is $(X_1 + \ldots + X_n)/n$, where n is the number of trials.)

- p. 529 The Activ/Stats activity Compare Two Proportions might be helpful for struggling students.
- pp. 529 – 531 The Step-By-Step Example is nicely done.
    - The blue box just below the Step-By-Step Example, and the Just Checking are also good.
- p. 532 Everyone Into the Pool. This is nicely explained.
- p 533 The Activ/Stats activity Test for a Difference between Two Proportions might be helpful for some students.
- p. 535 The Just Checking questions are good.
- p. 536 – 537 As usual, the What Can Go Wrong section is important.
    - Note especially the examples in the first and third cautions.
    - The Connections box is also good – it helps prepare for Chapter 26 (Comparing Proportions)
- Suggested Exercises (pp 540 – 544): Self-check: #1, 3, 5, 7, 9, 11, 13, 15, 17. Written: #10, 12, 18

Suggestions for Part V Review Exercises (pp. 545 – 549):

Self-check: #1, 3, 5, 11, 17, 19, 21, 23, 25, 33, 35, 37

Written: #8, 12, 16, 20, 26, 30