

A Data-Oriented, Active Learning, Post-Calculus Introduction to Statistical Concepts, Methods, and Theory

Allan J. Rossman
California Polytechnic State University
USA

Beth L. Chance
California Polytechnic State University
USA

Abstract

Our paper describes our project to develop curricular materials for a course that introduces students at the post-calculus level to statistical concepts, methods, and theory. This course provides a more balanced introduction to the discipline of statistics than the standard sequence in probability and mathematical statistics. The materials incorporate many features of successful statistics education projects that target less mathematically prepared students. The student audiences targeted by this project are particularly important because they have been overlooked by previous curricular reform projects. Most importantly, the proposed audience includes prospective teachers of statistics, introducing them to content and pedagogy that prepare them for implementing NCTM Standards with regard to statistics and probability and for teaching the Advanced Placement course in Statistics.

Background

The past decade has seen the development of a reform movement in statistics education, emphasizing features such as statistical thinking, active learning, conceptual understanding, genuine data, technology use, collaborative learning, and communication skills.¹ A wide variety of materials have been developed to support this type of instruction.² These include:

- Textbooks with more emphasis on statistical thinking, conceptual understanding, and genuine data are now widely available.
- Activity books and lab manuals provide investigations to foster students' active learning.
- Depositories of genuine datasets have been compiled in books and on the web.
- JAVA applets and new software allow for more interactive, visual explorations of statistical concepts.
- Assessment tools, such as projects, focusing more on students' conceptual understanding and ability to think statistically.

As these materials become more readily available, noticeable changes are occurring in introductory courses, especially in the areas of teaching methods, course content, and use of technology (see Garfield, 2000).

The Problem

The vast majority of these educational reform efforts have been directed at what we will call "Stat 101," an introductory, algebra-based, service course for non-majors. Relatively little attention has been paid to introductory statistics courses for mathematically inclined students majoring in fields such as mathematics, economics, the sciences, engineering, and even statistics.

Mathematics majors and other students with strong mathematical backgrounds typically choose between two options for introductory study in statistics: 1) take the Stat 101 course, or 2) take a standard two-semester sequence in probability and mathematical statistics. The first option is far from ideal, because the Stat 101 course is aimed at a different student audience and is not at a challenging mathematical level. Due to its lack of substantial mathematical content, this course often does no count

towards the student's major, providing a strong disincentive from taking the course. Unfortunately, the second and more common option is also fraught with problems.

Concerns about the nature of this sequence are not new. For example, the 1991 report of the MAA's Committee on the Undergraduate Program in Mathematics (CUPM) stated: "The traditional undergraduate course in statistical theory has little contact with statistics as it is practiced and is not a suitable introduction to the subject." This "math stat" sequence often presents a full semester of probability before proceeding to statistics, and then the statistics covered is often abstract in nature. As a result, students do not emerge from the sequence with a modern and balanced view of the applied as well as the theoretical aspects of the discipline of statistics. In fact, students often leave this course with less intuition and conceptual understanding than students who have taken a lower level course (e.g., data collection issues, statistical vs. practical significance, association vs. causation, robustness, diagnostics). An unfortunate consequence of this may be that the courses fail to attract some good students who would be excited by statistical applications.

Importance for Prospective Teachers

Especially unfortunate is that reform efforts in statistics education have largely failed to reach prospective teachers of mathematics and statistics, most of whom experience statistics, if at all, through this "math stat" sequence. In addition to the problems described above, the "math stat" sequence also does not typically adopt the pedagogical reform features (e.g., active learning, conceptual focus, group work, written reports) that have been demonstrated to enhance student learning (Garfield, 1995). Thus, future teachers emerging from a traditional "math stat" sequence generally do not experience a model of data-oriented, activity-based teaching practices that they will be expected to adopt in their own teaching.

In particular, the Curriculum Standards of the National Council of Teachers of Mathematics (2000) and the College Board's description of the Advanced Placement course in Statistics (2002) both emphasize the need for teachers who understand the fundamental concepts of statistics and can teach the subject using activities focused on data. Fortunately, awareness is growing in the United States that this calls for changes in the mathematical preparation of teachers. A recently released report on this issue from the Conference Board of the Mathematical Sciences (2001) recognizes the importance of better training in statistics for prospective teachers of mathematics.

Previous Efforts

There have been some efforts to incorporate more data and applications into the "math stat" sequence. Moore (1992) provides several examples for how he infuses the second semester course with more data and concrete applications, and Witmer (1992) offers a supplementary book towards these goals. Texts such as Rice (1994) now include more genuine data and applied topics such as two-way ANOVA and normal probability plots. More recently, a new text by Terrell (1999) aims to present a "unified introduction" to statistics by using statistical motivations for probability theory; its first two chapters are devoted to structural models for data and to least squares methods, before the introduction of probability models in chapter 3. Additionally, a new supplement by Nolan and Speed (2000) provides lab activities that integrate real scientific applications into statistical investigations in order to motivate the theory presented.

These changes are directed toward the second course in the two-course sequence, presumably leaving the first course to cover probability theory. This approach is especially a disservice to students who only take the first course. These students (e.g., engineering majors, mathematics education majors) often just do not have room in their curriculum for a second course. Other students, failing to see the relevance to their own discipline, may simply choose not to continue to the second course. As a consequence, Berk (1998) advocates that we should maximize the amount of statistics in the first semester."

Thus, while there have been efforts, they have not yet achieved widespread integration throughout the entire sequence as has been hoped. As David Moore wrote in support of our grant proposal in 1998: “The question of what to do about the standard two-course upperclass sequence in probability and statistics for mathematics majors is the most important unresolved issue in undergraduate statistics education.” We propose a rethinking of the entire two-course sequence so that the first course also addresses the call of Cobb and Moore (1997) to “design a better one-semester statistics course for mathematics majors.”

Course Materials

In response to this challenge, we are initially developing curricular materials for an introductory course at the post-calculus level, introducing mathematically inclined students to statistical concepts, methods, and theory through a data-oriented, active learning pedagogical approach. We consider it essential that this course provide a self-contained introduction to statistics, focusing on concepts and methods but also introducing some of their mathematical underpinnings. The materials provide a mixture of activities and exposition, with the activities leading students to explore statistical ideas and construct their own conceptual understanding.

The principles guiding our development of these course materials are:

- Motivate with real data, problems.
- Foster active explorations by students.
- Make use of mathematical competence to investigate underpinnings.
- Use variety of computational tools.
- Develop assortment of problem-solving skills.
- Use simulations (tactile, technology) throughout.
- Focus on the process of statistical investigation in each setting.
- Introduce probability “just in time.”

While several of these principles are equally relevant to the Stat 101 course, the focus on mathematical underpinnings sets this course apart. Students also develop several strategies for addressing problems; for example, the use of simulation as an analysis tool and not just as a learning device is emphasized throughout. With regard to use of technology tools, students use spreadsheet programs as well as statistical analysis packages. The focus is on a modern approach to these problems. Students will still learn basic rules and properties of probability, but in the context of statistical issues. Students will be motivated by a recent case study or statistical application and when necessary will “detour” to a lesson in the appropriate probabilistic technique. In each scenario, students will follow the problem from the origin of the data to the final conclusion.

The pedagogical approach is a combination of investigative activities and exposition. Some of the activities will be quite prescriptive, leading students clearly to a specific learning outcome, while others will be very open-ended. Examples of the former include guiding students to discover that the power of a test increases as the sample size does (other factors being equal), while examples of the latter include asking students to investigate the performance of alternative confidence interval procedures.

The sequencing of topics emphasizes the distinction between different types of studies and scope of conclusions by repeatedly modeling the process of statistical inquiry through data collection and statistical inference. Students first study comparisons between groups through experiments and observational studies with categorical then quantitative data, then they learn about randomly selecting samples from larger population first for one sample than two. They see in the two sample case that the mathematical computations are identical to the comparison of groups in an experiment but the interpretations differ. The final chapters focus on analyzing relationships among variables. A preliminary outline appears below:

Chapter 1: Comparisons and Conclusions for Categorical Data – descriptive analyses of 2×2 tables (segmented bar graphs, conditional proportions, relative risk, odds ratio), types of variables, observational studies vs. controlled experiments, confounding variables, causation, simulation, randomization, hypergeometric probabilities, Fisher’s Exact test, Simpson’s paradox

Chapter 2: Comparisons and Conclusions for Quantitative Data – descriptive analyses of quantitative data (dotplots, mean, standard deviation, five number summary, boxplots, stemplots, histograms) resistance, empirical rule, simulation of randomization test, effects of variability and sample size on significance.

Chapter 3: Variation and Random Sampling – probability sampling methods, bias, effect of sample size on sampling distribution, bootstrapping, Bernoulli process, Binomial tests and intervals, types of errors, binomial approximation to hypergeometric, sign test

Chapter 4: Models – normal distribution and other probability models, normal probability plots and normal probability calculations, Central Limit Theorem for sample counts and sample means, large sample z procedures for one proportion, t procedures for one mean, meaning of confidence, alternative confidence interval procedures, prediction intervals.

Chapter 5: Comparing Two Populations – Comparison of two population proportions, large sample z procedures, odds-ratio inference procedures, effect of sample size, types of error, comparison of two population means, standard errors, t procedures, effect of sample size and standard deviation, bootstrapping, pairing, t approximation to randomization test.

Chapter 6: Association and Prediction – simple linear regression (descriptive and inferential), logistic regression, one-way ANOVA, chi-square tests of independence, homogeneity of proportions.

Sample Activities

Below we present descriptions of four sample activities in order to provide a better sense for the materials being developed. We have chosen these both to illustrate the course principles described above and also to highlight differences between activities for a Stat 101 course and for the more mathematically inclined audience that we are addressing.

Sample Activity 1: Randomization Test

This activity concerns a psychology experiment to study whether having an observer with a vested interest in a subject’s performance on a cognitive task detracts from that performance (Butler & Baumeister, 1998). Twenty-three subjects played a video game ten times to establish their skill level. They were then told that they would win a prize in the next game if they surpassed a threshold value chosen for each individual so that he or she had beaten it three times in ten practice games. Subjects were randomly assigned to one of two groups. One group (A) was told that their observer would also win a prize if the threshold was surpassed; the other (B) was told nothing about the observer winning a prize. It turned out that 3 of 12 subjects in group A achieved the threshold score, compared to 8 of 11 in group B.

Students are asked to use cards (11 black cards for “winners” who surpass the threshold and 12 red cards for “losers”) to simulate random assignment of these subjects to treatment groups, under the assumption that group membership has no effect on performance. They pool their results in class to obtain an approximate sampling distribution of the number of “winners” randomly assigned to group A. By determining the proportion of cases in which that number is three or less, they approximate the p -value of the randomization test. Students thus begin to develop an intuitive understanding of the concept of statistical significance and an appreciation that statistical inference asks the fundamental question,

“How often would such sample results occur by chance?” Following their tactile simulation, we direct student to a java applet ([www.rossmanchance.com/applets/Friendly/ Friendly.html](http://www.rossmanchance.com/applets/Friendly/Friendly.html)) through which they simulate the process thousands of times to improve their estimate of the empirical p-value (Figure 1).

Friendly Observer Simulation

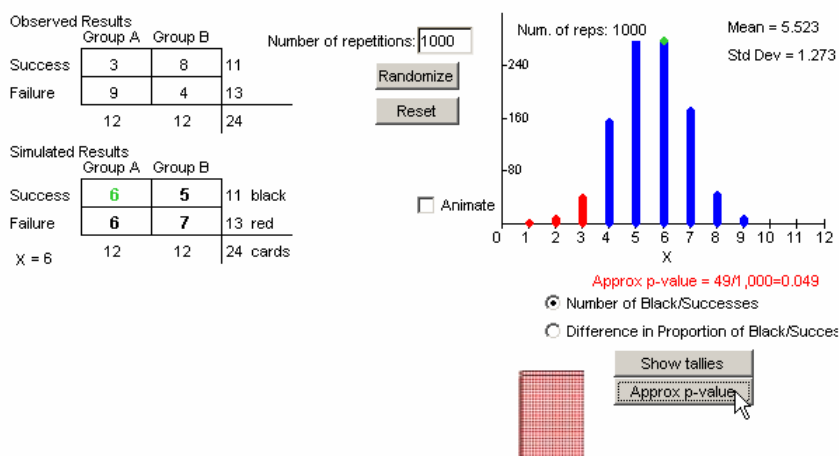


Figure 1: Simulation of “Friendly Observers” study

To this point the activity is very similar to ones appropriate for Stat 101 students, for example as found in *Activity-Based Statistics* (Scheaffer, et. al., 1996) and *Workshop Statistics* (Rossman and Chance, 2001). With this audience of mathematically inclined students, however, it is appropriate to ask them to take the next step and to calculate the exact p-value using hypergeometric probabilities. Thus, we take this opportunity to develop the hypergeometric distribution by studying counting rules and combinations and the equal likeliness assumption, motivated by their preliminary investigations. This probability “detour” comes “just in time” for students to explore with more exactness the statistical concept of significance in the context of real data from a scientific study.

Sample Activity 2: Bootstrapping

This activity has students analyze a data set obtained by Cal Poly students to compare the prices at two different local grocery stores. A random sample of 35 items was selected from an inventory list and the exact item was found at both stores. If the exact item was not available at both stores, the item was adjusted slightly (size or brand name) so that the identical item could be priced at both stores. Students then describe the distribution of the differences and discuss issues in “matched pairs” designs. They are then asked to carry out a bootstrap resampling to investigate the amount of sampling variability. They do this by programming a small Minitab macro. Based on the bootstrap estimate of the standard deviation of the sample mean, they decide whether the observed price difference between the two stores is statistically significant.

The bootstrap procedure is becoming increasingly popular with practicing statisticians but is seldom discussed in introductory courses. It also provides an intuitive understanding of the concepts of sampling variability and standard error. With these students the simulation is easy for them to implement and instead of always giving them applets where the simulation is transparent to run, we expect these students to program the details of the simulation procedure themselves.

Sample Activity 3: Type I and Type II Errors

We ask students to consider a baseball player who has been a .250 career hitter but genuinely improves to the point of becoming a .333 hitter, a very practically significant difference in this context. The question then is how likely the player is to establish that he has improved in a sample of 30 at-bats. Again we ask students to investigate this issue first through simulation. They use technology to simulate the results of 1000 samples of 30 at-bats for a .250 hitter and also for a .333 hitter, note the substantial overlap between the two distributions, and estimate the (low) power of the test from the simulation results. They then increase the sample size and note the resulting decrease in probability of type II error. At that point we ask students to study the binomial distribution and to apply it, with the benefit of technology, to calculating the theoretical probability of type II error for these tests. They then set up their statistical or spreadsheet package to perform these calculations efficiently for a variety of user-supplied sample sizes, significance levels, and values of the alternative probability. Finally, students conclude the activity by sketching graphs of power vs. these factors and writing a report (to either the player or manager) explaining the effects of these factors on power.

While a side benefit of these activities is that students become familiar with the binomial distribution and with calculations involving it, the emphasis is on helping them to understand the concepts of type I and type II errors. Technology again serves as an indispensable tool for minimizing computational burdens, enabling students to explore these ideas and develop their own understandings of them.

Sample Activity 4: Confidence

In this activity we begin by collecting some interesting data from the students. They are asked whether they would prefer to hear good news first or to hear bad news first. Typically the sample proportion preferring bad news first is quite large (e.g., .90) and our class sizes are around 30-50 students. As with Stat 101 students, students in this course begin to study the concept of confidence as they study sampling distributions through physical and then technology simulations. We then ask these students to use their mathematical abilities to investigate the meaning of “confidence” more formally through a follow-up to the *Activity-Based Statistics* activity that leads students to view a confidence interval as the set of plausible values of a population parameter based on the observed sample. Students use their knowledge of the binomial distribution along with technology to create exact binomial confidence intervals for a population proportion. They do this by considering all parameter values from .001 to .999 and calculating the probability of obtaining a sample proportion as extreme or more extreme as the actual with each proposed parameter value in either direction. Parameter values for which this probability exceeds α are considered plausible and are therefore included in the $100(1-\alpha)\%$ confidence interval. Through this activity students also discover the duality between confidence intervals and two-sided tests of significance.

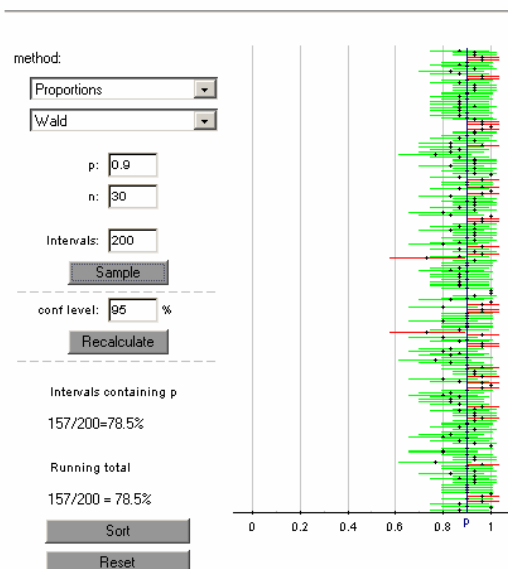
We then present students with two formulas for constructing approximate, large-sample 95% confidence intervals for a population proportion: $\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ and $p^* \pm 1.96 \sqrt{\frac{p^*(1-p^*)}{n+4}}$, where

$\hat{p} = \frac{x}{n}$ is the ordinary sample proportion of “successes” and $p^* = \frac{x+2}{n+4}$ is an estimator that “shrinks” the sample proportion toward one-half.

The former, of course, is the standard technique based on the normal approximation to the binomial distribution, and the second was proposed by Agresti and Coull (1998). We ask students to compare the performance of these two interval estimators (e.g., either through computer simulations and/or java applets as in Figure 2). This question naturally leads to a discussion of coverage probabilities and of interval lengths, and we proceed to ask students to perform simulations to determine coverage rates of both types of intervals for various values of the sample size n and the parameter value p . Students find

that for values of n and p with $np < 10$, the standard procedure produces nominal 95% confidence intervals that actually contain the population parameter less often than claimed, whereas the second procedure achieves a coverage rate much closer to the nominal confidence level.

Simulating Confidence Intervals



Simulating Confidence Intervals

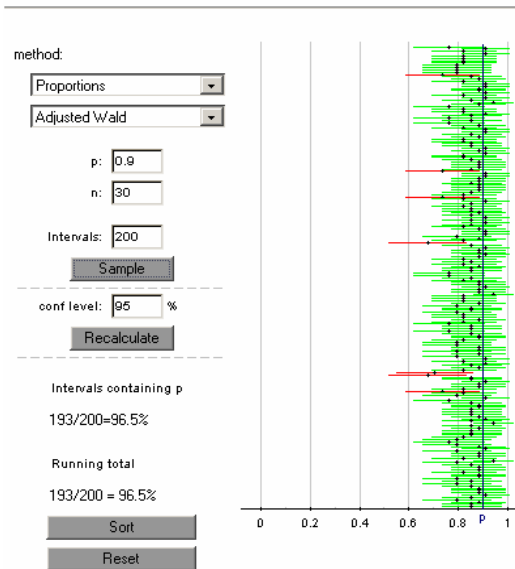


Figure 2: Comparison of confidence interval procedures with large p and small n

These exercises achieve several goals. First, they help students to deepen their understanding of confidence level as a coverage probability under repeated sampling. Second, they afford students the opportunity to investigate and evaluate recently published statistical methods, demonstrating the dynamic nature of the discipline. Finally, they provide students with still more experience of the utility of simulation as a powerful problem-solving tool.

Sample Activity 5: Measures of Center, Minimization Criteria

In this activity, students begin with data collected by Cal Poly students on the price of a house in San Luis Obispo County. They randomly selected prices for 8 houses from realestate.com (Figure 3).

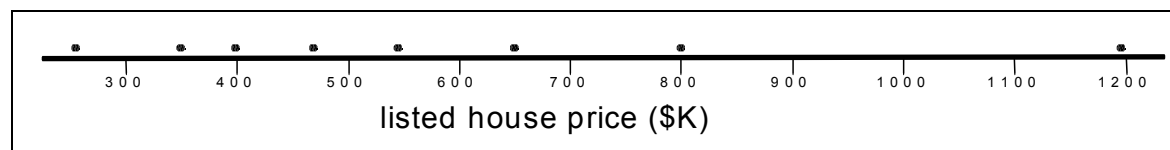


Figure 3: Housing Prices in San Luis Obispo County, Nov. 20, 2003

We ask students to propose criteria for comparing point estimates of the center of this distribution. Common criteria suggested include the sum of absolute deviations between the data values and the estimate, and the sum of squared deviations:

$$SAD(m) = |255 - m| + |349 - m| + |399 - m| + |469 - m| + |545 - m| + |649 - m| + |799 - m| + |1195 - m|.$$

$$SSD(m) = \sum_{i=1}^n (x_i - m)^2$$

Students investigate the behavior of these functions, analytically and graphically, and discover that the SAD function has an interesting piecewise linear appearance, minimized not at a unique value but for any value between 469 and 1545 (the two middle data values), inclusive. Naturally, the SSD function is parabolic, minimized at the sample mean. Students then use their calculus skills to prove that the sum of squared deviations is always minimized at the sample mean. In addition to examining these mathematical properties, students also use this activity to follow up on previously discussed properties such as resistance of the median but not the mean to outliers. They also investigate other criteria, finding for example that the midrange minimizes the maximum absolute (or squared) deviation. This activity also provides an introduction to the fundamental notion of “residual.”

Pilot Studies

The development began through a collection of activities that could be utilized to enhance an existing course. These activities were developed and used by the authors at their respective institutions. After securing an NSF grant, development continued to further field-test these activities, to further develop accompanying java applets, and to organize these activities and exposition into a stand-alone text. Sample activities were tested at over a dozen institutions and students were given the opportunity to react to individual activities through webforms and instructors provided summative feedback as well. This feedback further informed the development and revisions of the materials. Issues raised through this feedback included:

- Amount of structure and discussion provided for the students, including definitions of key terms
- Length of activities and more flexibility in selecting different portions of activities for use with students
- Structure of student time outside of class
- Detailed changes of individual activities, for example, slightly modifying the numbers in a data table to be less confusing.

Our external evaluator to the grant, Joan Garfield, compiled the feedback from the students and instructors and noted:

“Some consistent themes in the student comments were that they enjoyed hands-on activities, problems with real world contexts or applications, working in pairs or small groups instead of listening to lectures, and learning to use software to simulate or calculate results. Many commented that they liked using the computer to help them visualize concepts.”

The class testers and their institutions were rather diverse, illustrating wide interest in replacing the existing courses for these students. Two of the testers were teaching mathematical statistics courses, one was teaching an introductory statistics course that had a calculus prerequisite, one taught a probability course, one taught a probability and statistics course for high school teachers, and one taught a high school statistics course. All courses were small, varying from 4 to 25 students in the class. The textbooks used varied from an introductory text (e.g., Moore’s *Basic Practice of Statistics*) to mathematical statistics texts (three testers used the text by Wackerly, Mendenhall, and Schaeffer). All class testers combined a variety of instructional methods in their classes, including lecture, discussion, use of technology, and activities.

As individual chapters for the stand-alone text were developed, they were reviewed by three statistics faculty and two former students. Their comments also led to further refinement before the full course was to be taught. Specific feedback included more highlighting of the key ideas for easier reference, isolating definitions instead of having them completely embedded in the context of an activity, and estimating the class time of individual sections.

Current Evaluation Efforts

The course will be taught in its entirety several times this spring. In particular, Beth Chance will teach the course to science majors at St. Olaf University (a liberal arts college in Minnesota) and Allan Rossman will teach the course to mathematics majors, especially those planning to teach, and computer science majors at Cal Poly. This latter course will substitute for a comparable course with a more traditional curriculum. Individual class activities will be evaluated by student comments, observations of student engagement, and be external review of selected student writings.

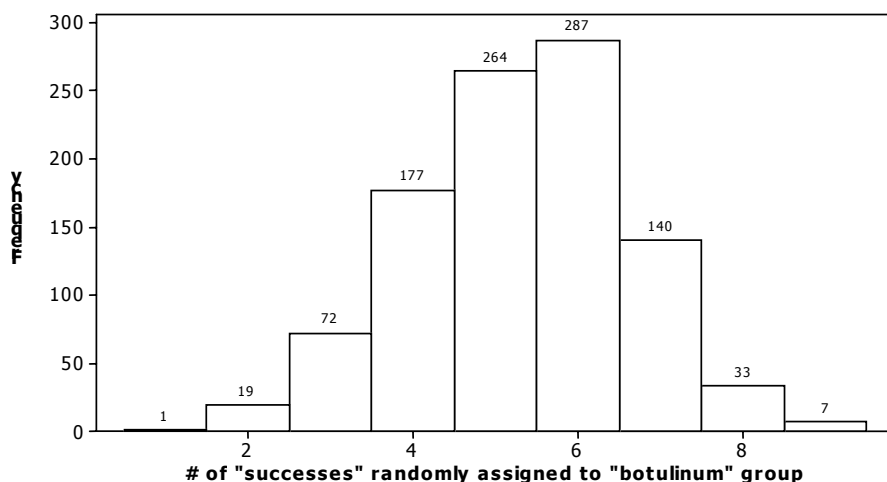
These courses will be evaluated by student feedback and by comparing performance on common final exam questions aimed at the above principles that have been given in the Cal Poly course by the same instructors in recent quarters. For example, questions will focus on students' conceptual understanding of key concepts such as significance and confidence (see example below) and students will be asked to carry out a small simulation to explore a topic not discussed in the course (see example below). We also have responses to "what is a p-value" by graduating statistics majors to compare to. Student performance on more standard calculation questions will also be compared. Student attitudes toward the discipline of statistics will also be assessed.

The authors are also overseeing a senior project that is exploring the impact of technology, especially java applets, on students' understanding, engagement, and retention. In this project, led by Katie Pesicka, she is observing students in the class while also developing assessments of specific learning gains from individual activities (see example below). We also plan to incorporate some specific assessments of the mathematics majors planning to teach in regards to their ideas about their own future teaching of statistics. Results of all these evaluations will be updated in March and in June and will be presented at the roundtable.

Example Assessment Items

1. A study published in the journal *Neurology* examined whether the drug botulinum toxin A is helpful for reducing pain among patients who suffer from chronic low back pain. Thirty-one subjects participated in the study. They were randomly assigned to one of two treatment groups: 15 received the drug itself, and the other 16 received a placebo of normal saline. The subjects' pain levels were evaluated at the beginning of the study and again after three weeks and after eight weeks. Naturally, the researchers were looking for evidence that the drug was more effective than the placebo for reducing back pain. The results of the study after eight weeks were that 9 of the 15 subjects who received the drug experienced a substantial reduction in pain, compared to 2 of the 16 subjects in the placebo group.

The following histogram displays the results of simulating a randomization test to assess whether the difference between the two groups is statistically significant. The variable displayed is the number of "successes" (those who experienced a reduction in pain) randomly assigned to the drug group.



(a) What conclusion would you draw from this simulation analysis about whether the difference between the two groups (botulinum and placebo) is statistically significant? Explain clearly how your conclusion follows from these simulation results. Include an approximation of the p-value of the test.

(b) Would you conclude that botulinum causes more reduction in back pain than the placebo? Explain. Be sure to refer to the type of study that was conducted here in your response.

2. Suppose that a new machine is designed to fill boxes of cereal by weight and you want to estimate μ , the mean weight of boxes produced by this machine. In class we discussed the interval $\bar{x} \pm t_{n-1,\alpha/2} s/\sqrt{n}$ as an approximate 95% confidence interval for μ . However, often managers want their employees to be able to calculate these intervals immediately on the factory floor. It would be much simpler if they could use the sample range R as the measure of variability instead of the sample standard deviation s . It turns out that with a sample of size 5, an unbiased estimator of the population standard deviation σ is $R/2.325$.

(a) Perform a simulation analysis to determine whether the procedure $\bar{x} \pm 2.571 R/(2.325\sqrt{5})$ produces 95 percent confidence intervals for μ . Use a sample size of $n=5$, and assume that the population of box weights follows a normal distribution. Clearly explain what you did in your simulation, the results of the simulation, and your conclusion. Also be clear about how your decision follows from your simulation analysis.

(b) Now calculate t -intervals for your simulated samples, and compare the t -intervals to those based on the sample range R . Compare the performance of these two procedures on both coverage probability and interval width.

Sample Student Feedback Form (K. Pesicka)

1. Have you learned anything more about confidence intervals after using this applet?

Yes

No

1a. If you answered "yes": What have you learned/ what concepts are more clear about confidence intervals now that you have worked with this applet?

1b. If you answered "no": Of the concepts you already understand about confidence intervals, which ones were presented in this applet?

2. A common response students give when asked to define what 95% confidence means, is... “In the long run, if we were to continue drawing samples of size n from the same population, 95% of the time *this specific* interval we calculated for μ [ex; (5.34, 17.9)] will contain the true population value of μ ”. After running this applet, explain why this is incorrect.

Summer Immersion Workshop

Up to thirty faculty members will be invited to a faculty development workshop on implementing these curricular materials at Cal Poly in June of 2004. The goals of the workshop will be to provide faculty with instruction on how to successfully select, adapt, and use activities in their introductory, post-calculus classes, how to balance in-class and out-of-class time spent on activities and follow-up work, how to prepare students to use software as part of activities, how to manage class time, and how to provide summary discussions and wrap-ups that clarify the big ideas and connect the concepts in an activity to the broader scope of statistical content.

We are currently accepting applications from individuals who are interested in learning more about the course and who plan to make some changes in their existing courses. Participants will meet in a computer lab and will work through some of the activities directly while discussing issues related to the first introduction to statistics for post-calculus students. A summary of participant reactions and their feedback will be discussed at the roundtable and will inform the next revision of the materials. An email discussion board will also be established to share feedback and suggestions during the following year.

Conclusion

We have argued that while the statistics education reform movement has made great strides and produced important materials for revamping “Stat 101” courses, a pressing need to reform introductory statistics courses for mathematically inclined students persists. We propose to address this need by developing materials to support a data-centered, active learning pedagogical style at the post-calculus level. Some of the key features of these materials are illustrated in common elements of the sample activities presented above, including:

- Students conduct investigations of statistical concepts and properties.
- Probability models are introduced in the context of statistical ideas, applied to real data.
- Mathematical skills of students that are utilized include familiarity with functions, graphical and analytical, as well as counting techniques and calculus optimization methods.
- Technology is used as a tool for such techniques as simulation and to assist with graphical displays and investigating effects of parameter changes.
- Data from scientific studies, popular media, or student-collected motivate the student explorations.

Our hope is that this re-designed course sequence will provide a more balanced introduction to statistical concepts and methods as well as theory, will increase interest in statistics as a potential career or side interest among mathematically inclined students, and will better prepare future teachers to employ student-centered pedagogy in their future classes.

References

- Agresti, A. & Coull, B. (1998). Approximate is better than ‘exact’ for interval estimation of binomial proportions. *The American Statistician*, 52:119-126.
- Berk, K. (1998). Revamping the mathematical statistics course. *1998 Proceedings of the Section on Statistical Education*. Alexandria, VA: American Statistical Association.

- Butler, J. & Baumeister, R. (1998). The trouble with friendly faces: Skilled performance with a supportive audience. *Journal of Personality and Social Psychology*, 75: 1213-1230.
- Cobb, G. (1992). Teaching Statistics. In L. Steen (ed.), *Heeding the Call for Change: Suggestions for Curricular Action*. Washington, DC: Mathematical Association of America, Notes #22: 3-43.
- Cobb, G. (1993). Statistics Education: A National Science Foundation Conference. *Journal of Statistics Education*, 1(1).
- Cobb, G. & Moore, D. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104: 801-824.
- College Board. (2002). *AP Statistics Course Description*. New York: The College Board. www.collegeboard.com/ap/statistics/.
- Committee on the Undergraduate Program in Mathematics (CUPM). (1991). The undergraduate major in the mathematical sciences. In L. Steen (ed.), *Heeding the Call for Change: Suggestions for Curricular Action*. Washington: Mathematical Association of America, Notes #22, 225-247.
- Conference Board of the Mathematical Sciences. (2001). *The Mathematical Education of Teachers*. Washington, DC: CBMS. Available online at: www.cbmsweb.org/MET_Document/.
- Garfield, J. (1995). How Students Learn Statistics. *International Statistical Review*, 63(1): 25-34.
- Garfield, J. (2000). An evaluation of the impact of statistics reform: Year 1 report. National Science Foundation, REC-9732404.
- Moore, D. (1997). New pedagogy and new content: The case of statistics (with discussion). *International Statistical Review*, 65:123-165.
- Moore, D. (2004). *The Basic Practice of Statistics* (3rd edition). New York: W.H. Freeman and Company.
- Moore, T. (1992). Getting more data into theoretical statistics courses. *Primus: Problems, Resources, and Issues in Undergraduate Mathematics Studies*, 2: 348-356.
- Moore, T. (ed). (2000). *Teaching Resources for Undergraduate Statistics*. Washington: Mathematical Association of America, Notes #51.
- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, Virginia: NCTM. www.nctm.org/standards/.
- Nolan, D. & Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications*, New York: Springer-Verlag.
- Rice, J. (1994). *Mathematical Statistics and Data Analysis* (2nd edition). Pacific Grove, California: Wadsworth Publishing.
- Rossman, A. & Chance, B. (2001). *Workshop Statistics: Discovery with Data* (2nd edition). Emeryville, California: Key College Publishing.
- Scheaffer, R., Gnanadesikan, M., Watkins, A., & Witmer, J. (1996). *Activity-Based Statistics*. New York: Springer-Verlag.
- Terrell, G. (1999). *Mathematical Statistics: A Unified Introduction*. New York: Springer-Verlag.
- Wackerly, D., Mendenhall, W. & Scheaffer, R. (2002). *Mathematical Statistics with Applications* (6th edition). Belmont, CA: Duxbury Press.
- Witmer, J. (1992). *Data Analysis: An Introduction*. Englewood Cliffs, New Jersey: Prentice-Hall.

Endnotes

¹ For overviews of this reform movement see Cobb (1992), Cobb (1993) and Moore (1997).

² For descriptions of such teaching resources see Moore (2000).