SECTION C

Should the Letters t and z be Eliminated From the Introductory Statistics Course?

Dr. Eugene L. Round

Embry-Riddle Aeronautical University Worldwide Campus

ABSTRACT

The first statistics course that students take in their undergraduate program, the introductory statistics course, is critical to their attitudes toward statistics, research, and their performance in later research efforts. This paper discusses current recommendations in regard to teaching undergraduate statistics, outlines some activities and techniques for improving the introductory statistics course, and explores some new techniques that are made possible by the use of the computer in the classroom. Offering a good introductory statistics course is essential to the success of the university's Quality Enhancement Plan.

Introduction

For those familiar with teaching statistics, the title might seem somewhat heretical since the z or standard normal distribution and its use in teaching such things as probability, confidence intervals, and hypothesis testing are foundations of the introductory statistics course, a term used to refer to the first statistics course that students take in their undergraduate programs. Likewise, use of the t-distribution in constructing confidence intervals and conducting hypothesis tests when the population standard deviation is unknown seems essential to the introductory course and indeed to much of what is done in statistics. Where did this odd-ball idea come from? Surely the proposers can't be serious! This was among ideas discussed by Drs. Alan Rossman and Beth Chance, prominent statistics educators and professors at California Polytechnic Institute, San Luis Obispo, at the 2011 United States Conference on Teaching Statistics. The theme of the conference was, "The Next Big Thing," and in their plenary address, they proposed several ideas of what might be the next big thing in teaching statistics. This was one of their thoughts. Were they serious? More on that later.

The topic of this year's Teaching and Learning Effectiveness Symposium is Research-Based Learning. When you hear the word "research," what other words come to mind? For most, probably somewhere not too far down the line would be the word "statistics." Some use of statistical techniques is essential to much of the research students will do. The importance of statistics is recognized in requiring most students to take at least one course in the subject. Why is that required? There are probably several reasons. First, a college-educated person in today's society should have a basic knowledge of statistical terms and techniques in order to function as an informed citizen. Second, in today's world people are inundated with data, and knowing how to draw informed conclusions from that data will be essential to students in their work environment. In a recent interview, Hal Varian, Chief Economist at Google, talked about statistics being the sexy job of the next decade. A larger quote is as follows:

> The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it is going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.... I think statisticians are part of it, but it's just a part. You also want to be able to visualize the data, communicate the data, and utilize it effectively. But I do think those skills - of being able to access, understand, and communicate the insights you get from data analysis - are going to be extremely important. Managers need to be able to access and understand the data themselves. (Manyika, 2008)

Such ideas regarding the importance of statistics are not new. H. G. Wells said, "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write" (PSI Psychology Tutor, 2007). Florence Nightingale called statistics,

> . . .the most important science in the whole world: for upon it depends the practical application of every other science and of every art; the one science essential to all political and social administration, all education, all organisation based upon experience, for it only gives the results of our experience. (PSI Psychology Tutor, 2007)

As indicated by Varian in the previously mentioned quote, one reason for the rising interest in statistics is the seemingly unending volumes of data now available through the internet and the need for companies and organizations to know how to use and interpret it.

So for good reasons most students are required to take statistics. What is the typical reaction of students to their first statistic course? It is probably accurate to say that most don't want to take statistics. Part of the problem may be that they see statistics as just another math class and consequently go into the course with a poor attitude and the feeling that it is just another course the university requires them to take covering material that they will never use. Most students complete the introductory statistics course with little understanding of the material, and few remember it more than a few months after the course ends—a few months is probably optimistic. Embry-Riddle Worldwide is not unique in this regard. Mathews and Clark (1997) asked eight randomly selected students who had just received a grade of A in an introductory statistics course to explain the mean, standard deviation, and Central Limit Theorem and found that although most could compute the mean of a set of numbers, many could not provide an adequate explanation of what the mean represents and none understood the standard deviation or Central Limit Theorem.

The following are comments from students who recently completed ERAU's online Math 222, Business Statistics course:

1. Theoretical application does not usually pertain to practical application. It is understood that the problem sets and methods taught during the course have value, however, without correlating them to actual value and repeatable uses throughout life, makes it meaningless. The idea behind taking a college course is to retain knowledge and use it throughout the course of your life
2. In this case, I will be performing an information purge immediately upon completion of the course. This is not a negative thing towards the instructor; however it is against the design of the course. Had there been real world applications and settings

> incorporated into the course that showed the value of the material and how it can be used on a regular basis, then perhaps it would have been a far better experience. However, I feel that my money was wasted on a text book and software that I will never use again, and that I received no real life benefit from this course. The only good thing from it is that I am three credit hours closer to my degree.

These are discouraging comments, but probably not atypical of the feelings many students have after finishing the introductory statistics course. Is the course worth teaching if most students don't learn or remember the most important concepts and many can't wait to do an information dump of the little they might remember? What can be done to:

- Change student attitudes towards statistics?
- Help students understand it?
- Help students remember it?
- Help students transfer what they learned?

Remembering, understanding, and transferring concepts from the basic statistics course will be very important if the Quality Enhancement Plan is to be successful.

Guidelines for Assessment and Instruction in Statistics Education

The focus of this paper will be some thoughts and recommendations on teaching the introductory statistics course. The Guidelines for Assessment and Instruction in Statistics Education (GAISE, 2005) funded by the American Statistical Association gave the following six recommendations for the introductory college statistics course:

1. Emphasize statistical literacy and develop statistical thinking,
2. Use real data,
3. Stress conceptual understanding, rather than mere knowledge of procedures,

4. Foster active learning in the classroom,

5. Use technology for developing conceptual understanding and analyzing data,

6. Use assessments to improve and evaluate student learning (p. 4).

Each of these will be briefly discussed in relation to Embry-Riddle Worldwide statistics courses.

Emphasize Statistical Literacy and Develop Statistical Thinking

Statistical literacy has been discussed and defined by several prominent statistics educators. Garfield and Ben-Zvi (2008) state, "Statistical literacy is a key ability expected of citizens in information-laden societies, and is often touted as an expected outcome of schooling and as a necessary component of adults' numeracy and literacy" (p. 34). They further state that statistical literacy should be a desired outcome of all introductory statistics courses. Many definitions of statistical literacy focus on the ability of people to see, understand, critically evaluate, and discuss the statistics that they are inundated with every day in the media. This includes the ability to understand and use basic statistical symbols, charts, graphs, and terms, and to interpret them in context (Gal, 2000; Garfield, 1999). Several authors have developed lists of learning outcomes appropriate to statistical literacy. Rumsey (2002) summarizes them:

> Each list seems to encompass two different types of learning outcomes for our students: being able to function as an educated member of society in this age of information and having a basic foundational understanding of statistical terms, ideas, and techniques. (p. 4).

Garfield and Ben-Zvi (2008) quoting from Wild and Pfannkuch (1999) state, "Statistical thinking is the way professional statisticians think" (p. 34). Cobb (2000) says, "Any introductory course should take as its main goal helping students to learn the basic elements of statistical thinking" (p. 4). He goes on to define the basic elements as appreciating and understanding the need for data, the importance of data production, the omnipresence of variability, and the quantification and explanation of variability.

A learning outcome from Embry-Riddle Aeronautical University's introductory statistics course reads as follows:

> Be a critical consumer of statistics presented by the media and other sources. Accurately interpret the statistics presented, identify ways in which they might be subject to misinterpretation either intentional or unintentional, and apply ethics to the interpretation and presentation of statistics. (Embry-Riddle, 2011, p.2)

The University's general education core competencies include Critical Thinking, Quantitative Reasoning, Information Literacy, Communication, and Scientific Literacy. The Worldwide Campus added Life Long Personal Growth to this list. In today's world, statistical literacy is critical to all of these competencies.

Use Real Data

Earlier in this paper, some comments from students who had just completed the online Business Statistics course were quoted. The assignments these students were given in the course they had just completed used textbook data sets drawn from real business situations and included problems with aviation-related data that had been recently downloaded from the Bureau of Transportation Statistics website. An example of a problem from the text is as follows:

> The manufacturer of Boston and Vermont asphalt shingles knows that product weight is a major factor in the customer's perception of quality. Moreover, the weight represents the amount of raw materials being used and is therefore very important to the company from a cost standpoint. The last stage of the assembly line packages the shingles before they are placed on wooden pallets. Once a pallet is full (a pallet for most brands holds 16

> squares of shingles), it is weighed, and the measurement is recorded. The data file Pallet contains the weight (in pounds) from a sample of 368 pallets of Boston shingles and 330 pallets of Vermont shingles. Completely analyze the differences in weights of the Boston and Vermont shingles, using $\alpha = 0.05$. (Berensen, Levine, & Krehbiel, 2009, p. 431)

An example from one of the problem sets written by the course developer reads:

> Do Americans favor the use of full-body scanners at U.S. Airports? According to a poll conducted by CBS News in January 2010, 74% of Americans approved of the use of full-body scanners. In a later poll taken by CBS news in November 2010, 81% approved of the use of full-body scanners. The sample size for the January poll was 1216 and the sample size for the November poll was 1137. At the .05 level of significance is there evidence of a difference in the proportion of Americans favoring the use of full body scanners between January 2010 and November 2010?

After working problems similar to these throughout the course, how could students come away seeing no applications for statistics or its importance in their work lives and everyday lives? How could they feel that they received no benefits from the course? Apparently, although the data were real, and the applications were there, there was a failure to make an essential connection between the concepts covered in the course and the life applications. Moore (2000) states, "Statistics is about data. Data are numbers, but they are not 'just numbers.' *Data are numbers with a context*" (p. xix). To see the context, students need have some appreciation for how the data were collected and the effort, time, expense, and perhaps pain that went into the data collection. Gould (2010) suggests that even though the data given to students are "real," they may not be real to students because of the way they are presented and students' lack of first-hand experience in the fields from which the data came. In this case, it seems that even though the context for the data was given, students didn't take time to consider the context, but viewed the data merely as a bunch of numbers to be plugged into a computer using a set of steps to get an answer and complete another homework problem.

Students need to have a stake in the data, but how can that be done? Perhaps by having them concentrate more on what the data given in the textbook problems represent and think about the procedures that were needed to collect the data that they are given in a neat Excel file. What went into collecting the data on Boston and Vermont shingles? How might the pallets in the sample have been selected? Actual experience with that approach has shown little success. Another approach would be to have students begin the course by identifying a problem of interest to them that can be addressed using statistics and then spend time throughout the course collecting the data and analyzing it to address the problem they identified. Most companies that employ ERAU Worldwide students use statistics so there must be an employee or group of employees somewhere in the company who collect and analyze the statistics. Another possibility could be to require students to find that group of employees in their company, find out what problems of interest to the company they are currently analyzing and what data sets they are using, and bring those problems and data sets to class.

Materials from the workshop Computationally Intensive Methods in Teaching Introductory Statistics conducted by Drs. Roger Woodard and Webster West contain the following recommendations for getting students involved in the data.

- Give the back story: How was the data collected and who collected it?
- What is the statistical question we need to ask? Why do we care?
- Discuss the data with the students: Brainstorm sources of variability, ask about personal involvement with the data, discuss the practical impact.

- Use visuals: Create graphics of the data, discuss individuals in the data, are there any outliers?
- Use photos of the data collection process and photos of the individuals involved.

Throughout their text, Sharpe, De Veaux, and Velleman (2012) emphasize having students think about the what, why, when, where, and how of the data.

Students could be required to actually go through the process of collecting the data themselves. Several statistics educators have stated that students need to work with the "messy" data that occur in the real world rather than the neat data sets seen in texts. (Garfield & Ben-Zvi, 2008). A problem with collecting real data is that it takes time and there is little of that in the introductory statistics course. By the time all the required topics are covered and the "really important stuff," like confidence intervals and hypothesis tests, is reached, there is little time left. An option would be to reduce the amount of material covered in the introductory course and strive for real understanding of the topics that are covered and equipping students to learn the other techniques on their own when they need them. That topic will be covered in more detail later. Eliminating some of the less important and time-consuming topics would allow introducing inference earlier in the course and perhaps allow time for some more meaningful activities related to tying all the course topics together and having students do some analysis of data they collected themselves and presenting their reports to the class. An excellent data collection activity involves constructing and "flying" paper helicopters, then collecting and analyzing the data, and presenting conclusions to the class (Box, 1991).

Stress Conceptual Understanding Rather Than Mere Knowledge of Procedures

The first recommendation would be: Don't teach statistics like a math course. Perhaps a starting point would be to stop naming statistics courses Math XXX. Most Worldwide students struggle with mathematics and calling statistics courses Math XXX may contribute to the negative attitude they have toward the subject before they sit in the first class session. Many authors have discussed the differences between mathematics and statistics. Citing Cobb and Moore (1997), Garfield and Ben-Zvi (2008) state, "And unlike mathematics, where the context obscures the underlying structure, in statistics, context provides meaning for the numbers and data cannot be meaningfully analyzed without paying careful consideration to their context: how they were collected and what they represent" (p. 8). In answer to the question posed in the title of his article, "Should Mathematicians Teach Statistics?"Moore (1988) states:

> No! Statistics is no more a branch of mathematics than is economics, and should no more be taught by mathematicians. It is a separate discipline that makes heavy and essential use of mathematical tools, but has origins, subject matter, foundational questions and standards that are distinct from those of mathematics. (p. 3)

He goes on to list four differences between statistics and mathematics:

- Statistics does not originate within mathematics,
- The aims and foundational controversies of statistics are unrelated to those of mathematics,
- The standards of excellence in statistics differ from those of mathematics,
- Statistics does not participate in the interrelationships among subfields that characterize contemporary mathematics. (p. 5)

Rossman, Chance, and Medina (2006) conclude that mathematics and statistics involve different types of reasoning and intellectual skills.

As mentioned previously, recommendations of prominent statistics educators include less emphasis on formulas and more emphasis on data. Cobb (2000) says, "Almost any course in introductory statistics can be improved by more emphasis on data and concepts at the expense of less theory and fewer recipes" (p. 3). Moore (1988) states, "..., the mathematical theory of statistics is of secondary importance in teaching. Almost any first course in statistics could be improved by less emphasis on both theory and recipes and more emphasis on what statisticians do and why" (p. 6). Some specific recommendations are:

- To the maximum possible extent, construction of graphics by hand should be eliminated. (Garfield & Ben-Zvi, 2008)
- Reduce emphasis on such things as requiring students to compute standard deviations, linear regression equations, and correlation coefficients. Going through the calculation once or twice by hand for a small data set may be useful, but beyond that the time required outweighs the value. Students learn little about statistical concepts from this time-consuming exercise.
- Reduce the amount of time spent teaching counting techniques and probability. Probability is a difficult topic for students and too much emphasis on probability causes students to become discouraged early in the course. Since probability forms the foundation of inferential statistics, some basic probability needs to be taught, but no more than what students need to understand the statistics that follows.
- The one-sample z-test for the mean may be useful in introducing hypothesis testing, but a lot of time shouldn't be spent on it. The two-sample z-test for population means should

be eliminated. Both of these assume that the population variance is known, an unrealistic assumption.

- Eliminate teaching the two-sample t-test that assumes equal population variances and the *F*-test for population variances. Regarding the t-test, Hayden (2000) states, "This was relevant in the days of hand calculations, but now most statisticians prefer not to make this assumption" (p. 96). Moore (1988) states, "…the *F*-test for equal variance is so sensitive to deviations from normality as to be almost useless in practice" (p. 6).

In the interest of having time to go more deeply into other important topics, some topics such as Analysis of Variances and perhaps Chi-Square statistics could also be dropped from the introductory course.

One item on a recent survey of ERAU Worldwide statistics instructors asked if they required students to use software to do statistics. It then asked those who didn't require software use to explain why they didn't require it. One of the most common answers was that there was not time to teach students how to use the software in addition to teaching them how to do statistics. That attitude should be changed to one of "learning to use the software is an essential part of learning how to do statistics."

Foster Active Learning in the Classroom and Use Technology for Developing Conceptual Understanding and Analyzing Data

GAISE (Aliaga et al., 2005) recommendations 4 and 5 as stated above are closely related in practice, and will be discussed together. There are numerous websites with applets related to statistics and almost all texts have applets that can be used. Many of the activities and associated materials are described in Garfield and Ben-Zvi (2008) and can be downloaded from the Adapting and Implementing Innovative Material in Statistics (AIMS) website at http://www.tc.umn.edu/~aims/aimstopics.htm. AIMS is an NSF funded project led by Dr. Joan Garfield of the University of Minnesota. The materials are grouped under nine topics that correspond to chapters in the Garfield and Ben-Zvi book, then into lessons and activities. In the following discussion, these will be identified by AIMS Chapter, Lesson, and Activity. More specific references can be found on the AIMS website.

A key to understanding statistics lies in understanding variation and where it comes from. A key question in statistics is whether the variation in the data is due to a real difference or just due to random variation. Students have a difficult time understanding variation. An activity related to this is "Measuring Heads" (AIMS Variability Chapter, Lesson 1, Activity 1). In this activity, all students in the class measure and record the circumference of their own heads, then all measure and record the circumference of the head of one selected class member. The measurements are then recorded on a dot plot in front of the class and the class discusses the variation in the two data sets. Why is there variation in individual head sizes? What is the cause of that variation? Why is there variation when they all measured the same person's head? What caused that variation? Would instructions on how to do the measurements help reduce the variation? Will all of the variation ever be eliminated? Some variation has a reason (different people have different size heads), and some variation is random with no identifiable reason (one person's head size does not vary in a short period of time).

In another type activity (AIMS Distribution Chapter, Lesson 2, Activity 3), students are given the description of two data sets, and asked to match them to given histograms. Figure 1 gives an example of this.

Below are two histograms (the scales have been left off intentionally). One corresponds to the age of people applying for marriage licenses, the other corresponds to the last digit of a sample of social security numbers. Label the horizontal axis of each graph with either AGE or LAST DIGIT, and write an explanation of why you made that choice.

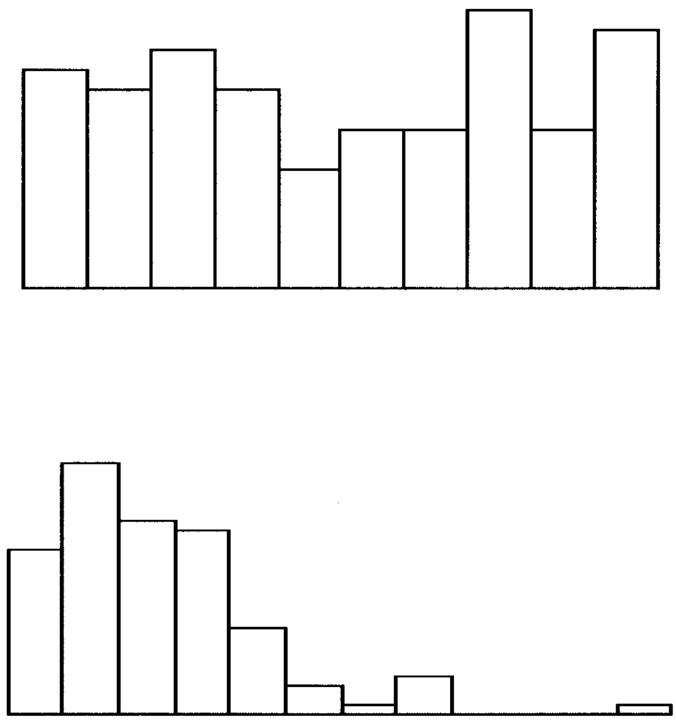
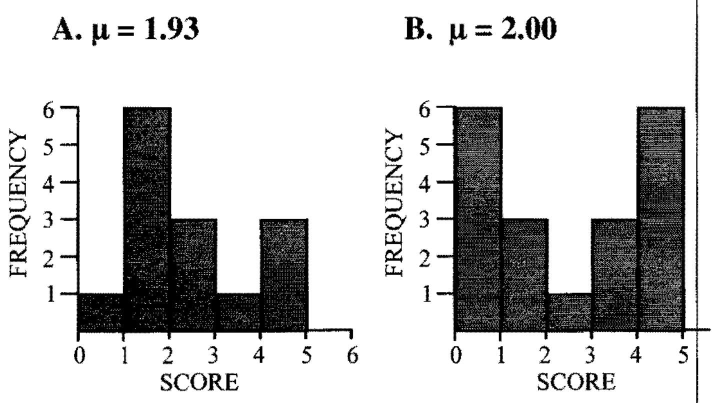


Figure 1. Histogram activity from AIMS Distribution Chapter, Lesson 2, Activity 3.

In a similar activity in the same lesson, students are given the description of a set of data such as "Students' guesses of their professor's age on the first day of class," and asked to draw a histogram of what the distribution would look like and justify their answers.

Understanding standard deviation is difficult for students. The AIMS materials (AIMS Variability Chapter, Lesson 2, Activity 2) include activities such as that shown in Figure 2, asking students to compare two histograms and identify which has the larger standard deviation and what makes the standard deviation larger. Another activity (AIMS Comparing Groups Chapter, Lesson 3, Activity 2) asks students to match histograms with boxplots. (See Figure 3.)



A has a larger standard deviation than B

B has a larger standard deviation than A

Both graphs have the same standard deviation

Figure 2. Standard distribution activity from AIMS Variability Chapter, Lesson 2, Activity 2.

Garfield and Ben-Zvi (2008) state, "To the maximum extent feasible, calculations and graphics should be automated" (p. 11). The following activities start with guesses, move to simple simulations not using computers, then to computer simulations using applets.

An AIMS activity supporting understanding of sampling distributions is Reese's Pieces (AIMS Samples/Sampling Distributions Chapter, Lesson 1, Activity 1). In this lesson, students are first told that Reese's Pieces candies come in three colors: orange, brown, and yellow and asked to guess the proportion of each color in a bag of Reese's Pieces. Students are then asked to pretend that each of 10 students in the class has a sample of 25 Reese's Pieces and instructed to write down the number of orange candies they think would be in each student's sample. The

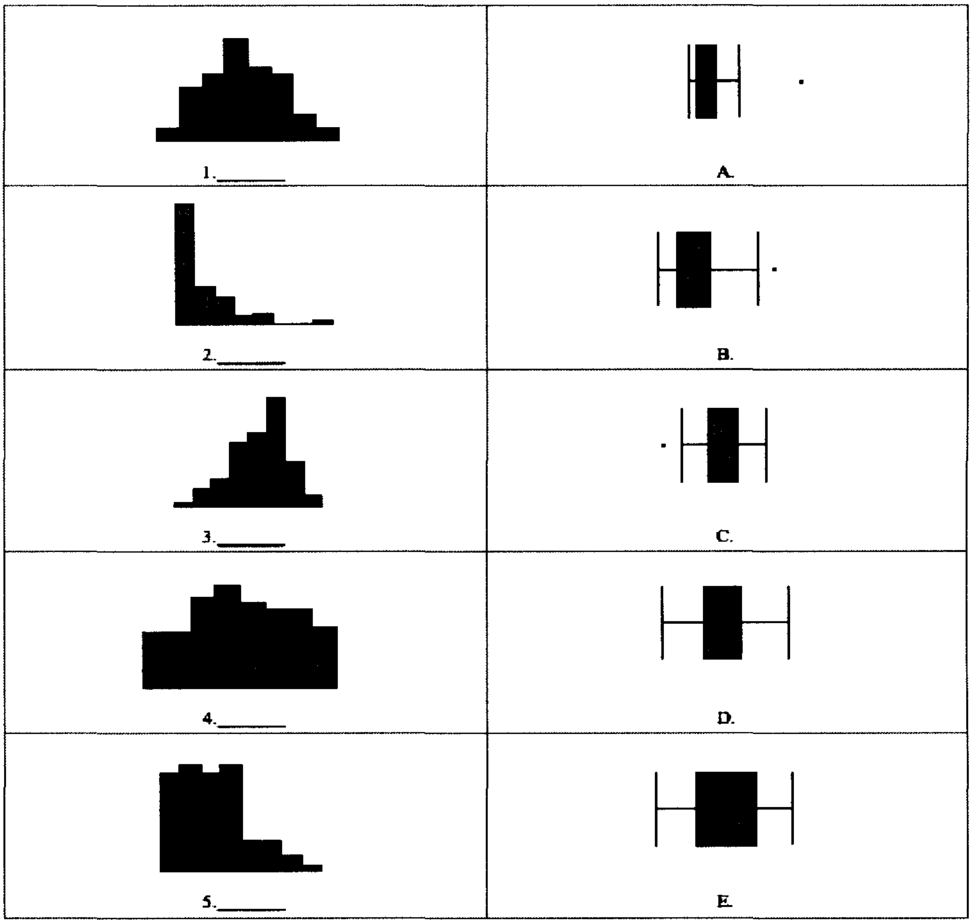


Figure 3. Histogram and boxplot activity from AIMS Comparing Groups, Lesson 3, Activity 2.

differences in the numbers students write represent the variability in each sample. Students are then given a sample of 25 Reese's Pieces candies and told to record the number and proportion of each color candy in their sample. A dot plot of the numbers and proportions of orange candies in each student's sample is constructed on the board so that all in the class can see it. Students are then asked to respond to questions such as:

- Did everyone in the class have the same number or proportion of orange candies?
- How do the actual sample values compare to the ones you estimated earlier?
- Describe the variability of the distribution of sample proportions of orange candies in terms of shape, center, and spread.

- Do you know the proportion of orange candies in the population?
- Which one (statistic or parameter) can always be calculated and which one has to be estimated?
- Does the value of the parameter change each time you take a sample?
- Does the value of the statistic change each time you take a sample?
- How does the sample proportion compare to the population parameter (the proportion of all orange Reese's Pieces candies produced by Hershey's Company that are orange)?

Next students are directed to an applet that enables them to simulate drawing many samples of 25 candies (http://statweb.calpoly.edu/chance/applets/Reeses/ReesesPieces.html). They are told that the population proportion of orange candies is 0.45 and asked how that compares to the proportion in their sample and to the center of the class distribution of sample proportions. Students then use the technology to draw one sample of size 25, then another, then 500 samples. They are then asked to describe the shape, center, and spread of the distribution of sample proportions and how that distribution compares to the one constructed by the class. In other activities, students are asked to vary the sample size and see what impact that may have on the distribution of sample proportions. They are asked, "As sample size increases, what happens to the distance the sample statistics are from the population parameter?"

An activity supporting understanding of the sampling distribution of the mean and the Central Limit Theorem is Sampling Pennies (AIMS Chapter Samples/Sampling Distributions, Lesson 2, Activity 3). This works best if students in the class all record the ages of coins in their possession, then construct a histogram of the ages. This will be a very right-skewed distribution. Another of the Rossman-Chance Applets is then used to generate samples of pennies that were collected in 1999. Students again start with small sample sizes and move to larger ones to see that even though the original population is highly skewed, the sampling distribution is approximately normal as the sample size increases.

The One Son Modeling Activity (AIMS Lesson Statistical Models and Modeling, Lesson 1, Activity 1) describes a policy adopted by China in the early 1990s in an attempt to reduce the country's birth rate. Each family was allowed to continue having children until they had a son. Students are asked to state a null and alternate hypothesis regarding what they think the average number of children per family would be under this policy. To simulate this, students are then given a cup with two slips of paper labeled B and G and asked to draw slips with replacement until they draw the B slip. They then record the number of "children" the family had. They do this for 10 simulated families, then compute the average number of children per family and compare that to their hypotheses.

Another worthwhile activity is the Let's Make a Deal Simulation based on the 1970s Monty Hall television program, Let's Make a Deal. In the program, a contestant was given a choice of three doors and told that a valuable prize was behind one of them. After the contestant picked a door, Monty would open one of the other doors so as not to reveal the big prize. The contestant was then given the choice of sticking with their original choice of doors or switching to the other remaining door. Students are asked to develop a null and alternate model about what would be the best decision, sticking or switching. Using an applet, they then simulate many contestant choices and the outcomes to test their hypotheses.

The recommendations are for more emphasis on use of software to do routine computations and more emphasis on interpreting the meaning of the output from the software. The technology reduces the necessity to use formulas and construct graphs by hand, but one set of rote procedures without understanding (plugging numbers into formulas) should not be traded for another (blindly following a list of steps for using the software). Using activities and applets such as those discussed above guide students through the process of doing statistics and help them see the concepts involved. Properly structured assessment activities can also help overcome this.

Efron and Tibshirani (1993) discuss statistical techniques called bootstrapping and resampling. These methods are potential replacements for traditional *z* and *t* tests that have been taught and used in statistics since their development in the early 1900s. The following is a brief illustration of these methods here using two examples and using StatCrunch software developed by Dr. Webster West of Texas A&M University. The following problem comes from Sharpe, De Veaux, and Velleman (2012).

> American League baseball teams play their games with the designated hitter rule, meaning that pitchers do not bat. The league believes that replacing the pitcher, traditionally a weak hitter, with another player in the batting order produces more runs and generates more interest among fans. The data provided in the file on the CD include the average numbers of runs scored per game by American League and National League teams for the 2009 season. With a 95% confidence interval, estimate the mean number of runs scored by American League teams. Do the data suggest that the American League's designated hitter rule may lead to more runs per game scored? (pp. 438-439)

A *t*-confidence interval would normally be used to answer the first question. The StatCrunch "Bootstrap a Statistic" applet can also be used to construct the confidence interval. After setting up the applet, the screen shown in Figure 4 is seen. It shows the results for American League Teams from 2009 and shows the mean number of runs scored per team to be 4.819. Bootstrapping basically involves resampling with replacement. The idea is to select another sample of size 14 from the 14 American League teams by selecting one team, noting its average runs scored, putting it back into the group, selecting another team, noting its average runs scored, putting it back into the group, and continuing this process until 14 teams are selected. Because the sampling is done with replacement, a team could be selected more than one time or not at all in the same sample. This is what yields the differences that will be seen in the sample means.

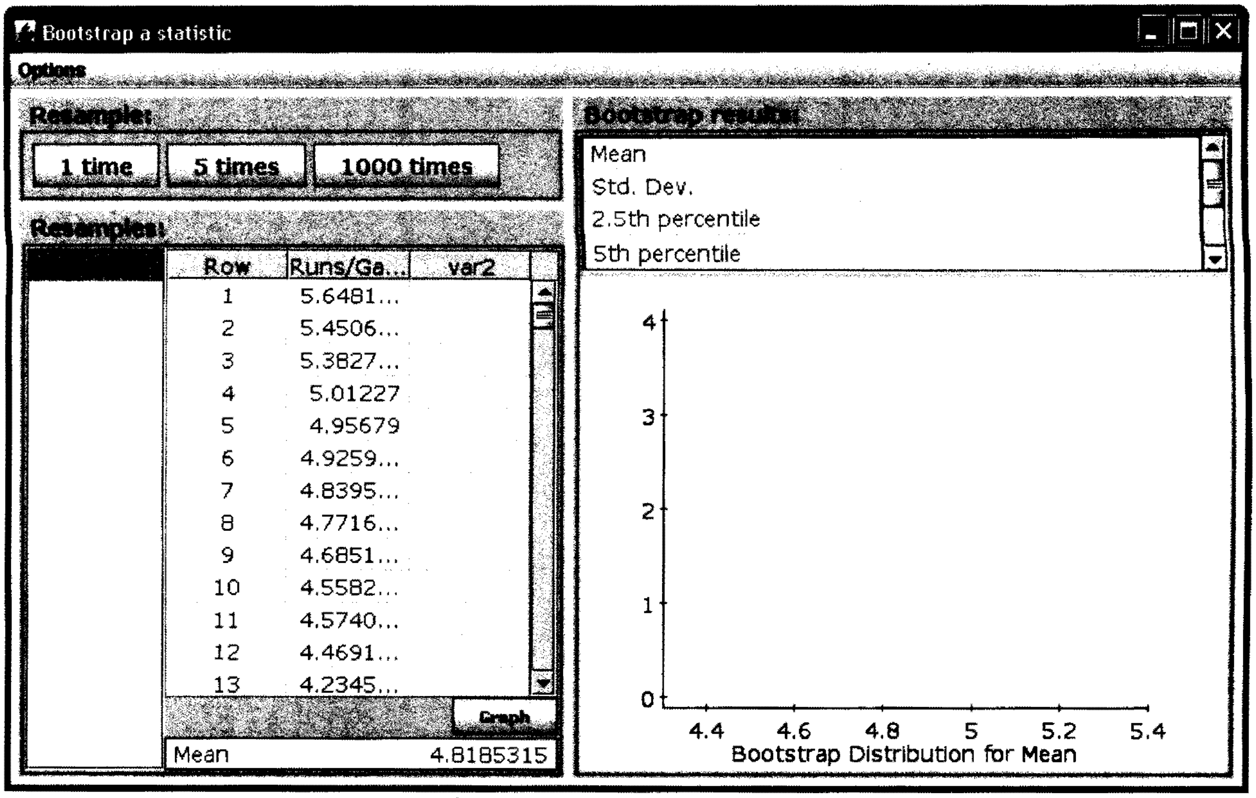


Figure 4. StatCrunch bootstrap applet showing the average number of runs per team in the American League in 2009.

When "1 time" at the upper left in Figure 4 is selected, the software draws one resample of the data as shown in Figure 5, then adds it to the display in Figure 6. The applet shows the mean for the new sample and plots a bar on the graph on the right side of the display.

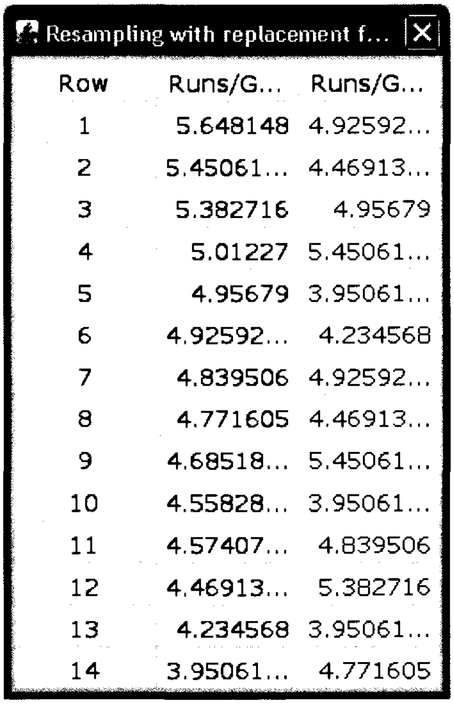


Figure 5. StatCrunch bootstrap applet showing the first resample of the American League data.

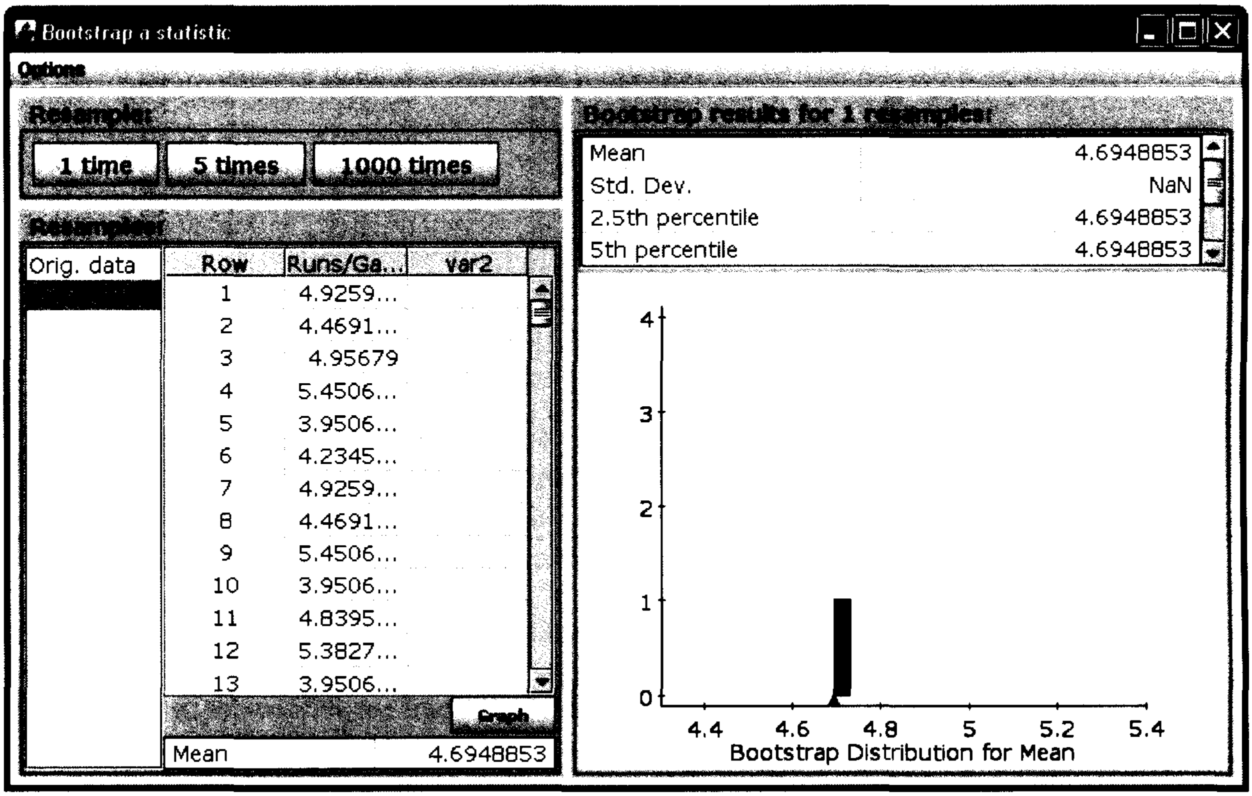


Figure 6. StatCrunch bootstrap applet showing the first resample for the baseball data.

The power of the computer is displayed when "1000 times" at the top of the display is selected. The applet takes 1000 resamples of the data and builds a plot of the distribution for the mean. Figure 7 shows the distribution of 3001 sample means. Each resample is shown on the left. Clicking on the sample number reveals the results for that sample. Above the plot of the distribution is a display that shows the 95% confidence interval. The 2.5th percentile is 4.578. Scrolling down the display shows that the 97.5th percentile is 5.062. So the 95% confidence interval is 4.578 to 5.062.

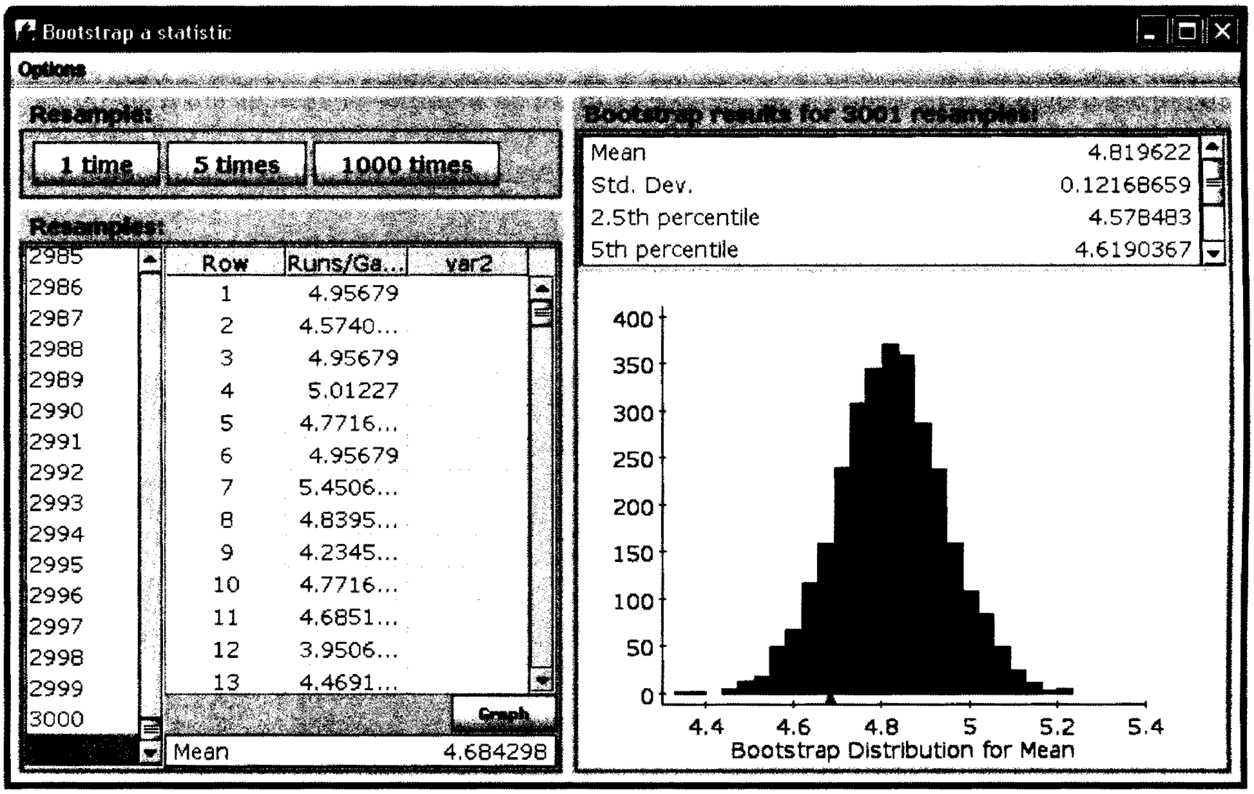


Figure 7. StatCrunch bootstrap applet showing 3001 resamples of the American League data.

The standard *t* confidence interval for these data, would be 4.549 to 5.088. In regard to the accuracy of the confidence interval constructed in this manner, Efron and Tibshirani (1993) say:

> …in large samples, the coverage of the bootstrap-*t* interval tends to be closer to the desired level than the coverage of the standard interval or the interval based on the *t* table.

It is interesting that like the *t* approximation, the gain in accuracy is at the price of generality. The standard normal table applies to all samples, and all sample sizes; the *t* table applies all samples of a fixed size *n*; the bootstrap-*t* applies *only to the given sample*. However with the availability of fast computers, it is not impractical to derive a "bootstrap table" for each new problem we encounter. (p. 161)

Next, a hypothesis test will be done to determine if there is a significant difference between the number of runs scored per game between the American and National Leagues. To do that, another StatCrunch applet called "Randomization Test for Two Means" will be used. After opening the applet and setting it up for this problem, the screen shown in Figure 8 will be seen. Notice that it shows the original data with the mean for the American League, the mean for the National League, and the difference of +0.3859, meaning that on average, the American League scored 0.3859 more runs per game than the National League.

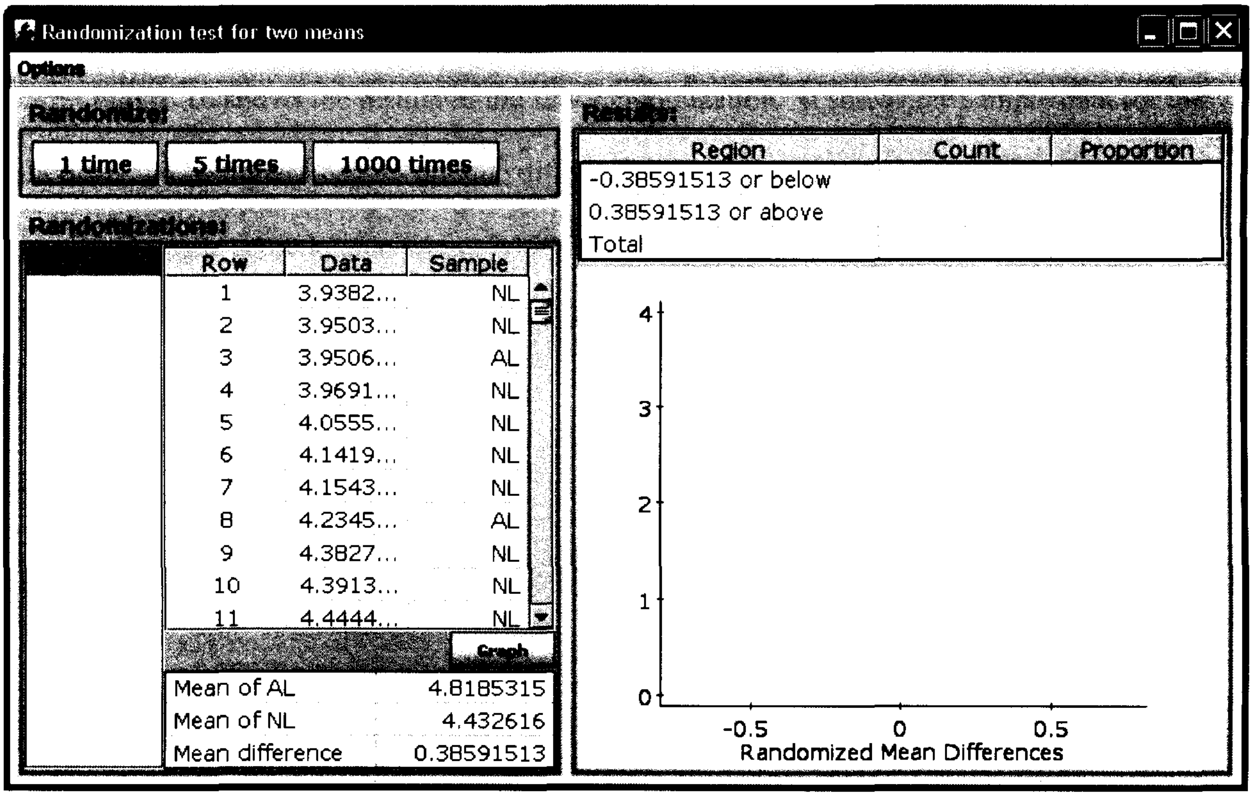


Figure 8. StatCrunch randomization applet showing the 2009 baseball data.

By clicking on "1 time" at the top of the screen, the applet will keep the data in the same order, but randomly shuffle the league assignment for each data point as shown in Figure 9. A description of this randomization adapted from Efron and Tibshirani (1993) so that it applies to this example is as follows:

> If the null hypothesis (of no difference in average runs per game between the leagues) is correct, any of the mean number of runs could come equally well from either league. So we combine all the 30 observations from both groups together, then take a sample of size 14 without replacement to represent the American League; the remaining 16 observations constitute the National League. We compute the difference between group means and then repeat this process a large number of times. If the original difference in sample means (.3859 in this example) falls outside the middle 95% of the distribution of differences, the two-sided permutation test rejects the null hypothesis at a 5% level. (p. 205 with adaptations)

Figure 10 shows the results of the first randomization. It shows the sample data for the American and National Leagues and the difference of -0.1796 meaning that in this sample, the American League scored, on average, 0.1796 fewer runs than the National League. As in the previous example, the power of the computer is shown when the randomization is run a large number of times.

Randomizing original data values

Row	Data	Sample
1	3.93827...	NL
2	3.95031...	AL
3	3.95061...	NL
4	3.96913...	AL
5	4.055556	NL
6	4.14197...	NL
7	4.154321	AL
8	4.234568	NL
9	4.382716	AL
10	4.39130...	NL
11	4.444444	AL
12	4.46913...	AL
13	4.50617...	AL
14	4.537037	AL
15	4.55828...	AL
16	4.57407...	NL
17	4.68518...	NL
18	4.765432	AL
19	4.771605	AL
20	4.81481...	NL
21	4.839506	AL
22	4.845679	NL
23	4.92592...	AL
24	4.95679	NL
25	4.962963	AL
26	5.01227	NL
27	5.06172...	NL
28	5.382716	NL

Figure 9. StatCrunch randomization applet showing the first "reshuffling" of the 2009 baseball data.

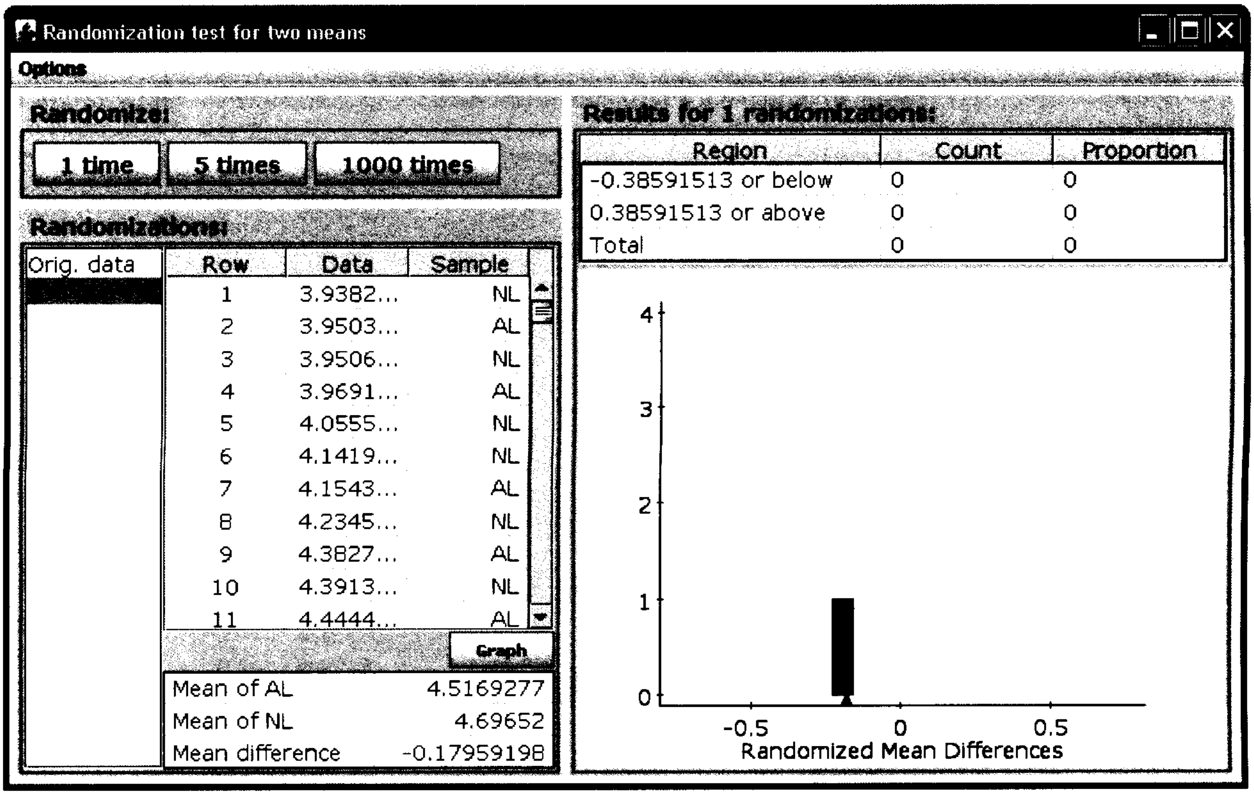


Figure 10. Statcrunch randomization applet showing the plot of the first "reshuffling" of the 2009 baseball data.

Figure 11 shows the results for 3001 randomizations. It shows that of the 3001 randomizations, a total of 22 or a proportion of 0.0073 were below -0.3859 and a total of 30 or a proportion of approximately 0.01 were above +0.3859 where 0.3859 is the mean difference between the leagues in the original data. The total of these proportions, 0.0173, would represent the *p*-value of this test—the probability of obtaining a result more extreme than the one in the original data if the null hypotheses of no difference in means is true. In StatCrunch, this area is shown in red on the chart at the right in Figure 11 as indicated by the circled areas in the tails of the distribution. The majority of students never comprehend the meaning of a *p*-value. This chart provides an excellent visual representation of its meaning. This *p*-value would lead to the rejection of the null hypothesis of no difference in the mean number of runs scored and lead to the conclusion that there was a significant difference in the mean number of runs scored between the two leagues. For comparison purposes, a t-test not assuming equal variances gives a p-value of .0208 while a t-test assuming equal variances gives a p-value of .0183.

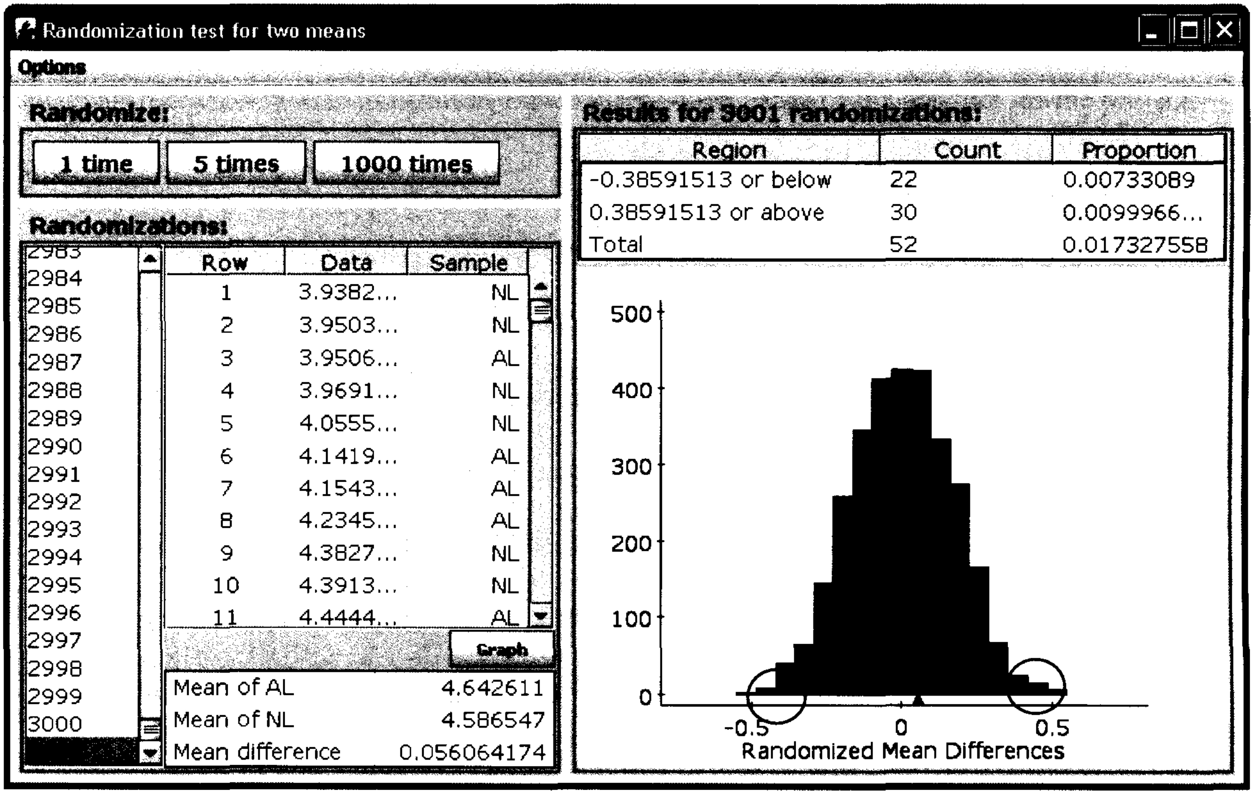


Figure 11. Statcrunch randomization applet showing the plot of 3001 "reshufflings" of the 2009 baseball data.

Use Assessments to Improve and Evaluate Student Learning

Authentic assessment supporting the course learning objectives should be incorporated into courses, rather than added on at the end. Garfield and Ben-Zvi (2008) give a good discussion of authentic statistics assessment in Chapter 4 of their book. The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) website (https://app.gen.umn.edu/artist/) has excellent assessment resources and a bank of great test or quiz questions. There is also a test builder function that allows instructors to view and select items they want to use in a test or quiz. This allows instructors to tailor the test to their class.

Conclusion

Should t and z be eliminated from the introductory statistics course? It could happen! Drs. Alan Rossman and Beth Chance are working on an NSF funded project titled Concepts of Statistical Inference: A Randomization-Based Curriculum. Part of the abstract for the project reads:

> This project is developing a fundamentally different curriculum for the introductory statistics course that emphasizes the entire process of statistical investigations, from design of data collection through statistical inference, throughout the course. The inference techniques are based on randomness introduced in data collection, specifically randomization and permutation tests, rather than on normal-based probability models. The goal is to lead students to develop a deeper understanding of fundamental concepts of statistical inference and of the process through which statisticians investigate research questions by collecting, analyzing, and drawing conclusions from data (Chance, B., & Rossman, A., 2009).

The W.H. Freeman Company has published an optional chapter, "Bootstrap Methods and Permutation Test," that accompanies one of their statistics texts (Hesterburg, Monaghan, Moore, Clipson, & Epstein, 2003). The chapter starts with the following:

> Statistics is changing. Modern computers and software make it possible to look at data graphically and numerically in ways previously inconceivable. They let us do more realistic, accurate, and informative analyses than can be done with pencil and paper. The bootstrap, permutation tests, and other resampling methods are part of this revolution. Resampling methods allow us to quantify uncertainty by

> calculating standard errors and confidence intervals and performing significance tests. They require fewer assumptions than traditional methods and generally give more accurate answers (sometimes very much more accurate). (p. 18-4)

The real underlying purpose of the Quality Enhancement Plan (QEP) that the university has selected is to better equip students to be successful in their college experience, in their work beyond college, and in other phases of life. For the undergraduate research course that is proposed as part of the QEP effort to be successful and therefore for the QEP itself to be successful, students need to come out of their first statistics course with a positive attitude toward statistics and research, an awareness of the statistics that they are inundated with everyday, and a solid understanding of basic statistics concepts and techniques and how they can be applied to investigate problems. This paper has suggested some things that might be done to improve the basic statistics course so that instead of getting end-of-course comments like those mentioned earlier in the paper, comments might read more like:

> Wow! This course really opened my eyes to all the statistics around me and to the ways statistics impact my life every day. I came into it with a negative attitude thinking it was just another math course and asking, "where will I ever use this stuff?' I won't ask that question again! The techniques used in the course enabled me to understand the concepts and gain confidence in my ability to do well in the undergraduate research course.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., . . .Witmer, J. (2005). Guidelines for assessment and instruction in statistics education: College report. Retrieved from http://www.amstat.org/education/gaise/GAISECollege_TOC.pdf

Berenson, M., Levine, D., & Krehbiel. (2009). *Basic business statistics: Concepts and applications.* Upper Saddle River, NJ: Pearson Education.

Box, G. (1991). Teaching engineers experimental design with a paper helicopter. *Quality Engineering*, *4*(3).

Chance, B., & Rossman, A. (2009). Concepts of statistical inference: a randomized-based curriculum. Retrieved from http://statweb.calpoly.edu/csi/

Cobb, G. (2000). Teaching statistics: More data, less lecturing. In T. Moore, (Ed.). *Resources for undergraduate instructors: Teaching Statistics* (pp.3-5). Washington, DC: Mathematical Association of America and American Statistical Association.

Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*, 801-823.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap.* Boca Raton, FL: Chapman & Hall/CRC.

Embry-Riddle Aeronautical University, Master Course Outline for MA 222. (2011). Retrieved from https://erau.blackboard.com/bbcswebdav/orgs/ERAU_EVP/MA%20222%20 Final.pdf

Gal, I. (Ed). (2000). *Adult numeracy development: Theory, research, practice.* Cresskill, NJ: Hampton Press.

Garfield, J. (1999). *Thinking about statistical reasoning, thinking, and literacy.* Paper presented at the first international research forum on statistical reasoning, thinking, and literacy (STRL-I), Kibbutz Be'eri, Israel.

Gould, R. (2010). Statistics and the Modern Student. Retrieved from http://preprints.stat.ucla.edu/590/Gould.pdf

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* New York: Springer.

Hayden, R. (2000). Advice to mathematics teachers on evaluating introductory statistics textbooks. In T. Moore, (Ed.). *Resources for undergraduate instructors: Teaching Statistics* (pp. 93-103). Washington, DC: Mathematical Association of America and American Statistical Association.

Hesterburg, T., Monaghan, S., Moore, D., Clipson,, A., & Epstein,R. (2003). Bootstrap methods and permutation tests: Companion chapter 18 to *The Practice of Business Statistics.* Retrieved from http://bcs.whfreeman.com/pbs/cat_160/PBS18.pdf.

Manyika, J. (Interviewer) & Valrian, Hal. (Interviewee). (2008. *Hal Varian on how the Web challenges managers* [Interview audio file]. Retrieved from McKinsey Quarterly website: http://mckinseyquarterly.com/

Mathews, D., & Clark, J. (1997). Successful students' conceptions of mean, standard deviation, and the Central Limit Theorem. Paper presented at the Midwest Conference on Teaching Statistics, Oshkosh, WI.

Moore, D. S. (1988). Should mathematicians teach statistics? *The college mathematics journal, (19)*, 3-7.

Moore, D. S. (2000). *Statistics: Concepts and controversies.* New York, NY: W.H. Freeman.

PSI Psychology Tutor. (2007, December 18). Why study statistics in psychology? [Web log message]. Retrieved from http://psychologytutor.wordpress.com/2007/12/18/why-study-statistics-in-psychology/

Rossman, A., Chance, B., & Medina, E. (2006). Some important comparisons between statistics and mathematics and why teachers should care. In G.F. Burrill (Ed.), *Thinking and reasoning about data and chance: Sixty-eighth NCTM Yearbook* (pp. 323-334). Reston, VA: National Council of Teachers of Mathematics.