

Claremont Colleges Scholarship @ Claremont

CGU Faculty Publications and Research

CGU Faculty Scholarship

1-1-2002

An Interactive Tutorial for Teaching Statistical Power

Christopher L. Aberson '99
Claremont Graduate University

Dale E. Berger
Claremont Graduate University

Michael R. Healy '04
Claremont Graduate University

Victoria L. Romero '07
Claremont Graduate University

Recommended Citation

Aberson, C.L., Berger, D.E., Healy, M.R., & Romero, V.L. (2002). An interactive tutorial for teaching statistical power. *Journal of Statistics Education* [online], 10 (3). Retrieved from <http://www.amstat.org/publications/jse/v10n3/aberson.html>

This Article is brought to you for free and open access by the CGU Faculty Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in CGU Faculty Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

An Interactive Tutorial for Teaching Statistical Power

Christopher L. Aberson
Humboldt State University

Dale E. Berger, Michael R. Healy, and Victoria L. Romero
Claremont Graduate University

Journal of Statistics Education Volume 10, Number 3 (2002), www.amstat.org/publications/jse/v10n3/aberson.html

Copyright © 2002 by Christopher L. Aberson, Dale E. Berger, Michael R. Healy, and Victoria L. Romero, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Internet; Introductory statistics; Statistical inference; Statistical power; Tutorial.

Abstract

This paper describes an interactive Web-based tutorial that supplements instruction on statistical power. This freely available tutorial provides several interactive exercises that guide students as they draw multiple samples from various populations and compare results for populations with differing parameters (for example, small standard deviation versus large standard deviation). The tutorial assignment includes diagnostic multiple-choice questions with feedback addressing misconceptions, and follow-up questions suitable for grading. The sampling exercises utilize an interactive Java applet that graphically demonstrates relationships between statistical power and effect size, null and alternative populations and sampling distributions, and Type I and II error rates. The applet allows students to manipulate the mean and standard deviation of populations, sample sizes, and Type I error rate. Students ($n = 84$) enrolled in introductory and intermediate statistics courses overwhelmingly rated the tutorial as clear, useful, easy to use, and they reported increased comfort with the topic of statistical power after using the tutorial. Students who used the tutorial outperformed those who did not use the tutorial on a final exam question measuring knowledge of the factors influencing statistical power.

1. Introduction: Teaching statistical power

Statistical power considerations are important to adequate research design ([Cohen 1988](#)). Without sufficient statistical power, data-based conclusions may be useless. Students and researchers often misunderstand factors relating to statistical power. One common misunderstanding is the relationship between Type II errors and Type I errors. Given a design with a 5% Type I error rate, students and researchers often predict the rate of Type II errors also to be 5% ([Hunter 1997](#); [Nickerson 2000](#)). Of course, the probability of a Type II error is generally much greater than 5%, and in a given study, the probability of a Type II error is inversely related to the Type I error rate. Another misconception is the belief that failure to reject the null hypothesis is sufficient evidence that the null hypothesis is true (that is, failing to reject suggests that the null hypothesis is true; see [Nickerson 2000](#)). The prevalence of underpowered studies in many fields is striking evidence of a lack of comprehension of the relevance of statistical power to research design (for example, on average, the Type II error rate in psychology and education is estimated to be 50% or more; [Sedlmeier and Gigerenzer 1989](#); [Lipsey and Wilson 1993](#)).

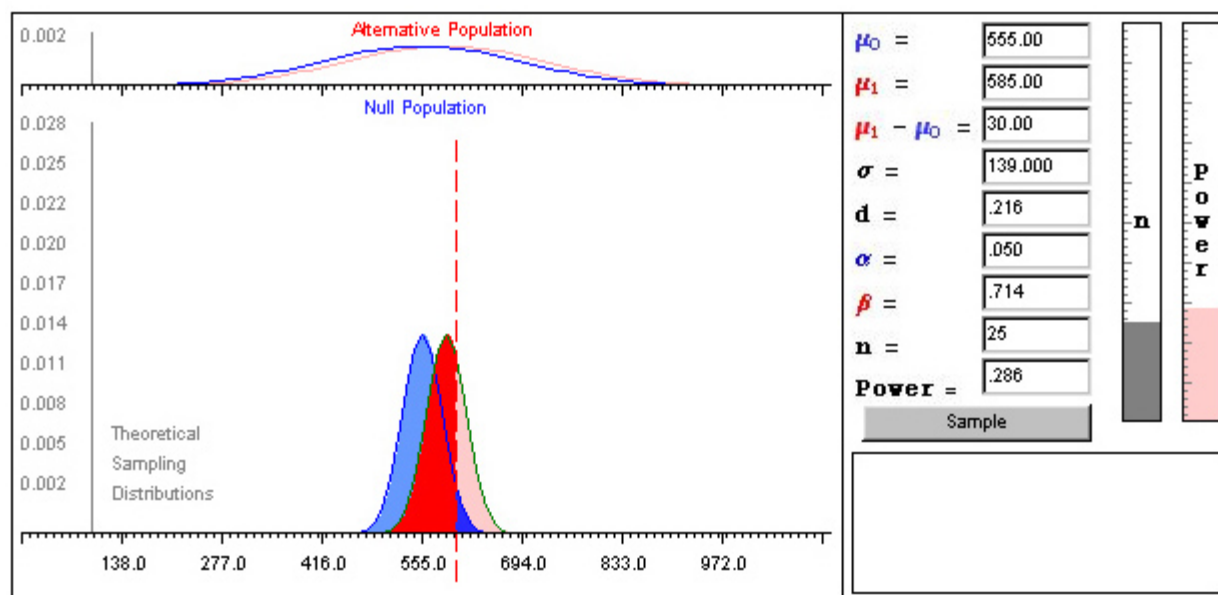
Unfortunately, statistical power remains among the most difficult topics to teach. A full understanding requires

integration of the relationships among null and alternative distributions, Type I and Type II error rates, sample size, and effect size. A typical in-class demonstration might involve sketching null and alternative sampling distributions, using different colored shading to illustrate the areas associated with Type I, Type II, and power. The instructor typically repeats the process to address changes in sample size, standard deviations, and the size of the difference between means, and accompanies the drawings with a great deal of hand waving. Students often find this topic confusing, because the relationships between power concepts are difficult to illustrate with traditional classroom media.

These problems and misconceptions suggest a need for improved instruction on statistical power. Greater understanding of statistical power may increase appreciation of the application and limitations of hypothesis testing, and potentially lead to improvements in student research design.

2. The WISE Power Applet

The WISE power applet graphically represents the populations, the samples, and sampling distributions for the training and no training groups. Using the applet, students manipulate null and alternative population means, standard deviations, sample sizes, Type I, Type II, power, effect size, and simulate drawing samples. The applet includes two additional functions that are not used in the tutorial. A "thermometer" graphic interface allows the student to raise or lower sample size and observe the corresponding changes in the sampling distributions and the effect on power. The applet also allows the student to click and drag the mean of the alternative population, and instantly see how differences between the null and alternative means influence power. We often use these functions for in-class demonstrations. A permanent working version of the applet represented in [Figure 1](#) can be accessed at www.amstat.org/publications/jse/v10n3/aberson/power_applet.html. The authors' version of the same applet can be accessed at the following site: wise.cgu.edu/power/power_applet.html.



[Figure 1](#)

Figure 1. Web Interface for Statistics Education Power Applet.

A paper and pencil assignment (described below) includes several exercises guiding student use of the applet. For each exercise, the student draws multiple samples from the specified alternative population. The applet presents a plot of the distribution of individual scores in the sample, an arrow representing the location of the sample mean, the calculated value for the sample mean, z , the probability of obtaining a value of z or greater than that obtained, and power. The student compares the obtained z for each sample to a one-tailed criterion (for example, $z > 1.645$) and decides whether to reject or retain the null hypothesis. The applet uses color-coding to distinguish between areas on both distributions with the Type I error in dark blue, power in pink, and Type II error in red, and provides a decision regarding the null

hypothesis, allowing the students to check their conclusions.

Following the simulation exercise with the applet, students answer several multiple-choice questions. Incorrect choices correspond to common misconceptions or errors. When students choose an incorrect answer, they receive feedback that discusses why that answer is wrong, and they receive guidance to finding the correct answer. To account for the possibility that students may choose a correct answer even though they do not understand why the answer is correct, the tutorial provides feedback explaining correct answers as well.

3. A Tutorial for Teaching Statistical Power

3.1 Description of the Tutorial

Our Web-based interactive tutorial consists of a paper and pencil assignment that guides use of the power applet, on-screen multiple-choice questions, and follow-up questions. The assignment assumes basic knowledge of normal distributions and hypothesis testing for one mean. The tutorial includes a short review of effect sizes (specifically Cohen's d) and provides a link to a tutorial on hypothesis testing. We designed the computer-based portion of the tutorial so that students could complete it in a single 50-minute laboratory session. The follow-up questions generally take the students additional time to complete.

The tutorial begins by describing a research scenario (wise.cgu.edu/power/index.html). The student reads that the task is to investigate the effectiveness of standardized test preparation courses. The tutorial presents the mean and standard deviation for a population of students who took a standardized test but did not take part in any training course and the means for populations of students who completed one of two training programs. One program is very effective and the other is slightly effective.

For the first set of exercises (beginning on wise.cgu.edu/power/power19.html), students examine test scores for people who have completed either of two training programs (that is, alternative populations) in comparison to those who received no training. The first training program is very effective and the second is only slightly effective, corresponding to a large and a small effect size, respectively ([Cohen 1988](#)). The student begins by using the interactive applet to draw 10 samples of test scores for 25 program graduates from the very effective program. For each sample, the student determines whether to reject or retain the null hypothesis. Next, the student repeats the procedure above for samples drawn from the slightly effective program and then suggests reasons for differences in the rejection rates.

To examine the impact of standard deviation, the student reduces the standard deviation for the slightly effective program and again draws 10 samples. The student compares these results to the sample results for the program with the same mean but larger standard deviation from the earlier exercise.

Following these exercises is a multiple-choice question on effect size. Through multiple-choice questions, we force the student to confront possible misconceptions and test their knowledge. Students examine three scenarios, one with a large magnitude of difference between the null and alternative means and a large standard deviation, one with a small magnitude of difference and a small standard deviation, and one with a moderate magnitude of difference and moderate standard deviation. The tutorial asks the student to choose the situation with the greatest statistical power. Incorrect answers correspond to common errors such as choosing the option with the largest magnitude of differences between means or the option with the smallest standard deviation instead of considering the relationship between the two parameters (that is, effect size). When students choose these answers, they receive feedback as to why the answer is wrong and guidance to finding the correct answer.

Next, the student examines the effects of the sample size by drawing 10 samples of $n = 4$, $n = 25$, and $n = 100$, respectively. The student explains the effect of sample size on power in his or her own words and completes a multiple choice question on the relationship between sample size and the shape of the sampling distribution.

The final exercise (wise.cgu.edu/power/power37.html) asks the student to consider how the results of the initial sampling exercises change if a Type I criterion of 1% is used in place of the 5% criterion that was used in the initial

exercises. Again, multiple-choice questions follow.

After completing the sampling exercises, students answer several follow-up questions that do not involve computer-based feedback (wise.cgu.edu/power/followup.html). These questions are appropriate for grading and/or for in-class discussion. One question asks students to evaluate a researcher's conclusions regarding another training program. For this program, the sample mean is nearly identical to the null mean. However, with a sample of 10,000 program graduates, the researcher correctly rejected the null hypothesis. The researcher concluded that this program does a good job of preparing individuals for the standardized test. The student provides an interpretation of the findings and evaluates the conclusion. This question forces the student to consider issues of practical versus statistical significance and provides a striking example reminding the student that statistical significance does not imply practical value (see [Sowey 2001](#)).

3.2. Tutorial Effectiveness

So far, we have described the tutorial assignment. We collected data on student reactions to the tutorial and tutorial effectiveness as well. Eighty-four students, in introductory psychological statistics and intermediate psychological statistics classes from three institutions, used the WISE power tutorial as part of their regular instruction. Students, approximately three quarters of whom were female, ranged from sophomores to graduate students. The institutions included a large urban state university, a small, private liberal arts college, and a medium-sized rural state university. Students from a single introductory statistics class (included in the 84 students above) used the power tutorial as an extra credit assignment. From this class, 18 students completed the assignment and seven chose not to use the assignment.

Students used the tutorial as either a laboratory or homework assignment. Before using the tutorial, students attended an introductory lecture on statistical power. After completion of the tutorial assignment, students rated ease of use, utility in teaching statistics, interest in using similar assignments in the future, comfort with the topic prior to using the tutorial, and comfort with the topic after tutorial completion. From the class including tutorial users and non-users, we compared scores on a single essay question taken from the final examination asking the student to identify and explain the factors that influence statistical power.

Most students rated the tutorial as either somewhat easy or very easy to use (89.2%), the explanation of statistical concepts as somewhat clear or very clear (92.6%), indicated that the tutorial was somewhat useful or very useful for teaching statistics (98.8%), and that they were somewhat interested or very interested in using similar assignments to learn about other statistical topics (96.2%). Student comfort with the topic of statistical power after tutorial use (0% not at all comfortable, 64.6% somewhat comfortable, and 35.4% very comfortable) improved compared to retrospective ratings of comfort before tutorial use (24.1% not at all, 73.4% somewhat, and 2.5% very; Wilcoxon $z = 6.15, p < .001$).

Data from final examinations provide some support for the effectiveness of the tutorial in teaching about factors related to statistical power. As self-selection into use and no tutorial use groups is a problem with this comparison, we used an ANCOVA with final points in the course (removing points earned on the power laboratory from the dependent variable) as a covariate. Those students who used the power tutorial ($M = 13.0$ out of 15 possible) scored significantly higher on an exam question related to statistical power than did students who chose not to use the power tutorial ($M = 8.4$), $F(1,22) = 10.7, p = 0.003$, eta-squared = 0.33.

4. Discussion

Nearly all students using the WISE power tutorial viewed the explanation of statistical concepts as clear and useful, found the tutorial easy to use, indicated greater comfort with the topic of statistical power after using the tutorial, and were interested in using similar tutorials to learn about other topics in the future. This combination of student interest, ease of use, and increased student comfort with topics indicates that students accept the tutorial, and that the tutorial can be a useful tool for teaching concepts about statistical power. Additionally, there is some, albeit limited, evidence that students using the power tutorial understand the factors related to statistical power better than those who did not

use the tutorial.

The WISE tutorials incorporate many important principles of good instruction ([Romero, Berger, Healy, and Aberson 2000](#)). In particular, the power tutorial promotes active engagement in elaborative processing (see [Hofer, Yu, and Pintrich 1998](#)). To answer the questions, students must interact with the applet, interpret and integrate findings, and explain and apply the concepts that they have learned.

Another issue addressed by the power tutorial is that of false confidence, whereby learners feel that they understand concepts better than they really do ([Bjork 1995](#)). The choices of answers in the multiple-choice questions correspond to common misconceptions. Thus, students are forced, on a limited basis, to confront their errors, and the tutorial provides instruction that address specific misunderstandings.

The power tutorial incorporates visual and numeric information to enhance learning. According to [Paivio's \(1971\)](#) dual-coding theory, information presented in multiple formats improves memory. Our interactive applet shows relationships between numeric values for the mean, z , and power and graphs that reflect Type I and Type II error rates, power, and sample means in relation to the null and true distributions.

Student reactions to this tutorial and performances following tutorial use demonstrate that interactive computer-based tutorials integrated into introductory statistics courses can be accepted by students as useful supplements, or even replacements for, traditional statistics assignments. This result agrees with earlier assessments of WISE tutorials indicating positive student reactions (see, for example, [Aberson, Berger, Healy, and Romero, in press](#)).

Acknowledgments

The Web Interface for Statistics Education project is supported by grants from the Mellon Foundation, Claremont Graduate University, and Humboldt State University.

Portions of this paper have been presented at Syllabus, San Jose, CA 2001 and the American Psychological Association Annual Convention, San Francisco, CA 2001.

References

- Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D., and Romero, V. L. (2000), "Evaluation of an Interactive Tutorial for Teaching the Central Limit Theorem," *Teaching of Psychology*, 27, 289-291.
- Aberson, C. L., Berger, D. E., Healy, M. R., and Romero, V. L. (in press), "Evaluation of an Interactive Tutorial for Teaching Hypothesis Testing Concepts," *Teaching of Psychology*.
- Bjork, R. A. (1995), "Memory and Metamemory Considerations in the Training of Human Beings," in *Metacognition: Knowing about Knowing*, ed. J. M. A. P. Shimamura, Cambridge, MA: The MIT Press, 185-205.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), New York: Academic Press.
- Hofer, B. K., Yu, S. L., and Pintrich, P. R. (1998), "Teaching College Students to be Self-Regulated Learners," in *Self-Regulated Learning: From Teaching to Self-Reflective Practice*, eds. D. H. Schunk, and B. J. Zimmerman, New York: Guilford Press, 57-85.
- Hunter, J. A. (1997), "Needed: A Ban on the Significance Test," *Psychological Science*, 8, 3-7.
- Lipsey, M. W., and Wilson, D. B. (1993), "The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation from Meta-analysis," *American Psychologist*, 48, 1181-1209.

Nickerson, R. S. (2000), "Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy," *Psychological Methods*, 5, 241-301.

Paivio, A. (1971), *Imagery and Cognitive Processes*, New York: Holt, Rinehart, and Winston.

Romero, V. L., Berger, D. E., Healy, M. R., and Aberson, C. L. (2000), "Using Cognitive Theory to Design Effective On-line Statistics Tutorials," *Behavior Research Methods, Instruments, and Computers*, 32, 246-249.

Sedlmeier, P., and Gigerenzer, G. (1989), "Do Studies of Statistical Power Have an Effect on the Power of Studies?" *Psychological Bulletin*, 105, 309-316.

Sowey, E. (2001), "Striking Demonstrations in Teaching Statistics," *Journal of Statistics Education* [Online], 9(1). (www.amstat.org/publications/jse/v9n1/sowey.html)

Christopher L. Aberson
Department of Psychology
Humboldt State University
Arcata, CA 95521
USA
cla18@humboldt.edu

Dale E. Berger, Michael R. Healy, and Victoria L. Romero
Department of Psychology
Claremont Graduate University
Claremont, CA 91711
USA
dale.berger@cgu.edu

[Volume 10 \(2002\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Information Service](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)