# WEB-BASED INSTRUCTION FOR STATISTICAL POWER ANALYSIS

Ann A. O'Connell
University of Connecticut, United States
Rosemarie L. Ataya
University of South Florida, United States
Jiarong Zhao
University of Connecticut, United States
ann.oconnell@uconn.edu

*Use of the internet to support instruction in general and the use of statistical software in particular provides instructors and students with an opportunity to improve learning while maintaining effective use of limited classroom time. We have developed a Web site (http://power.education.uconn.edu/) that encompasses instruction in power analysis issues and teaches students and others how to use the* nQuery Advisor© *software to establish sample size for research designs ranging from the simple to the complex. The evaluation results of our Power Project Web site and materials are promising, and the purpose of this paper is to share our approach and materials with other instructors of statistics and research design.*

## INTRODUCTION

Determination of an appropriate sample size is one of the most common design concerns for persons planning a research study (Allison, Gorman and Primavera, 1993; Kraemer, 1985). As teachers, mentors and consultants to faculty and students pursuing applied research, we have found that many individuals asking sample size questions are relatively unfamiliar with the important relationships between power, effect size, and variability. People want a simple answer to "how many participants do I need?"! Often, the individuals asking sample size questions have recently completed courses in research design and statistical analysis at the graduate level, or already hold a Ph.D. While it is clear that on some level at least the right questions are being asked, it is unfortunate that training or coursework in applied quantitative methods for many educational researchers continues to contain only a small fraction of time devoted to issues of power and determination of sample size (Curtis and Harwell, 1998). This lack of attention to power issues has several adverse consequences, including "planned" studies using insufficient sample sizes, low power to detect meaningful effects, incorrect interpretation of results, and/or a persistent belief that statistical and practical significance are synonymous (Nickerson, 2000; Kirk, 2001). In short, we have found as have others (Aberson, Berger, Healy and Romero, 2002), that many individuals have misconceptions regarding principles and proper use of procedures in power analysis.

Given the time constraints of an already extensive introductory statistics sequence at the graduate level, we felt that instruction in power analysis issues could be enhanced through the use of technology. We developed a web-based support for power, effect size and sample size questions that is suitable for use as part of an introductory (or advanced) statistics course for graduate level educational researchers (http://power.education.uconn.edu/). Our web-site contains several components that are appropriate to the needs of students in educational psychology and research: (1) an introductory power-point tutorial on power analysis and sample size estimation, including factors affecting power; (2) presentation of a series of research-based lessons on power analysis to help students understand concepts relating to statistical power; (3) instruction in the use of power analysis software (*nQuery Advisor 5.0*) with examples from the research-based lessons for different statistical techniques; and, (4) a glossary and links for additional web-based support for power analysis.

## RESEARCH GOALS

The web-based tutorial is designed to be an adjunct to classroom lectures on effect size and power. We felt that our materials would be most meaningful if they could be accessed and revisited any number of times outside of the formal classroom. Our primary goals for the pedagogical use of our freely available website are to promote research planning skills among our

graduate students, and to strengthen understanding of the relationships among power, effect size, variability, and sample size in a manner that is authentic to the activities and experiences of investigators in educational research. We hope to improve transfer for later productivity in student-initiated research activities, particularly in research planning, beyond enhancement of short-term student learning in this area.

## THEORETICAL FRAMEWORK

According to Ferrall (1995), an effective computer-assisted tutorial should "teach the material and how to use an underlying statistics package simultaneously" (p. 1). The lessons we have developed correspond directly to the software we have selected to support power analysis in our School of Education. To enhance learning in statistics courses, we believe that technology can be most effectively used through the creation and use of situated learning activities appropriate to the level of the material and the topic being studied. The "cognitive apprenticeship model" (Collins, Brown and Newman, 1989) provides the theoretical framework that underlies the development and use of our power analysis tutorial. Through the tutorial and the supporting website, we hope that students will begin to make stronger and lasting connections among concepts pertinent to efficient research design, including effect size, sample size estimation, power, and practical versus statistical significance.

## DEVELOPMENT OF THE POWER PROJECT

Prior to making our software decision, we reviewed published critiques of power analysis software (Goldstein, 1988; Thomas ad Krebs, 1997) and tested some of the software and freeware available. We decided to purchase multiple copies of *nQuery Advisor 5.0* (Elashoff, 2002; http://www.statsol.ie/) for our courses, due to its flexibility, ease of use, and emphasis on determination of a meaningful effect size. Students have access to the *nQuery Advisor* software in our computer labs in our Neag School of Education. The software provides a seamless connection to the power analysis lessons, which we use to familiarize students with the relationships between power, effect size, sample size, variability, statistical significance, and type of statistical test. The use of the tutorial removes much of the didactic aspects of traditional approaches to teaching power analysis and sample size estimation, and supports familiar goals of statistics education, namely, to improve statistical literacy and competency through the active engagement of students in the process of statistics and statistical analysis (see Garfield, 1995; Holcomb and Ruffer, 2000).

After our software choice was made, we began with the development of research-based lessons taken from the published literature. We focused on research in educational psychology that used statistical techniques commonly studied during introductory quantitative methods courses, including the paired samples *t*-test, independent samples *t*-test, multiple regression with at least five independent variables of interest, and one-way analysis of variance. These techniques are among the most common analyses encountered in psychological research (Cohen, 1992), however, the *nQuery Advisor* is capable of much more complex analyses as well. Once the basics of power analysis are presented using *nQuery Advisor* for these simpler techniques, it is fairly straightforward to understand the requirements for power analysis given more complex applications.

The four research studies used for the lessons are described below.

1. Paired Samples *t*-test.

   Tiggemann, M. and Lynch, J. E. (2001). Body image across the life span in adult women: The role of self-objectification. *Developmental Psychology*, 37(2), 243-253.

2. Independent Samples *t*-test.

   Shankweiler, D., Lundquist, E., Dreyer, L. G., and Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing: An Interdisciplinary Journal*, 8, 267-294.

3. Multiple Regression.

   Stanovich, K. E., Cunningham, A. E., and Feeman, D. J. (1984). Intelligence, cognitive

skills, and early reading progress. *Reading Research Quarterly*, 19(3), 278-301.

4. One-way Analysis of Variance.

   Ball, E. W. and Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, 26(1), 49-66.

Our online materials include useful information about power analysis, design of experiments, freeware available from the Internet, and additional web-based resources. The website contains the research-based power analysis lessons, provides an introduction to power analysis using PowerPoint, contains a glossary of statistical terms, includes links to additional resources and websites on power analysis, and demonstrates for the user how to utilize the *nQuery* software.

For each of the power-analysis lessons, material on the website instructs the student in how to use the *nQuery* software to calculate effect sizes from the data for the particular analysis of interest, how to determine an appropriate variance estimate, and how to calculate sample size. The lessons are designed to illustrate how power calculations, effect size, sample size and variability are related, and how manipulation of one quantity affects the others. Although we recognize the limitations involved in post-hoc power analysis (Hoenig and Heisey, 2001), our emphasis is on using the lessons to guide learning about the *nQuery* software and power concepts, and as a method for approaching research design concerns for which sample size determination plays a critical role.

## EVALUATION METHODS

Students enrolled in introductory and advanced statistics courses at our Neag School of Education were asked to participate in a pilot study to assess the usefulness and quality of the power analysis lessons. Ten advanced and introductory students with diverse backgrounds volunteered to participate in the evaluation. We used three assessment strategies in the formative evaluation of our pedagogical approach. First, we included a brief pre- and post-test on power analysis concepts and terminology. Second, we asked for extensive evaluation feedback about the participants' likes and dislikes of our web-based approach and the design of the Power Project. Finally, we developed an additional post-test in the form of an authentic research design scenario which puts the student in the role of research designer. After completing the tutorial we asked the participants to determine the sample size and power for one of two proposed research scenarios. All information was collected using paper and pencil, although we are currently working on incorporating all assessments onto the website for on-line data collection. For our formative purposes, we randomly asked five of the participants to complete the two-sample *t*-test lesson, and the other five to complete the multiple regression lesson. The design-scenario post-test corresponded with the assigned lesson.

The pilot students were given the following outline for how to proceed with the navigation through the website.

a. Read the *Welcome* information
b. Complete the *Introduction to Power Analysis* PowerPoint presentation
c. Review the *Introduction to nQuery* pages.
d. Open the *nQuery* software (the tutorial works best if *nQuery* is running as you work through the examples).
e. Choose the desired lesson from the Power Project website.
f. Complete the Power Project evaluation forms when you have completed the lesson.

## RESULTS

### *Sample*

Most students in the pilot study were female (70%), pursuing a Ph.D. (90%), and all had taken (or were currently taking) at least two courses in statistics. All but two students were enrolled in a degree program in Educational Psychology; the other two were from Curriculum and Instruction and Family Studies.

*Pre- and Post-test Results*

The six items on the pre-test asked about definitions of power; the Greek symbol representing power; factors affecting power; and how high versus low power affects the findings from a research study. Out of the ten individuals we asked to review the Power Project web-based instruction, we found that 37 items (total) were answered incorrectly on the six-item pretest, while only ten items (total) were answered incorrectly on the post-test. Most of the remaining misconceptions on the post-test involved the impact of too high or too low power on a research study's findings. This pattern of errors on the pre- and post-test have suggested some areas for improvements to the tutorial.

On the design scenarios, all but one of the participants arrived at the correct answer for the sample size requested for the design using the *nQuery* software; this student provided the effect size rather than the sample size.

*Tutorial Evaluation Form*

We asked ten evaluation items at post-test regarding how the students felt about the information contained in the tutorial and the perceived usefulness of the Power Project web-based approach. In general, participants felt that an appropriate amount of knowledge was covered in the tutorial, and that the approach was helpful to understanding concepts and principles of statistical power and the corresponding relationship power has with effect size and sample size. One area of concern observed from these evaluations was that students were not consistent in their ratings of choice as to having power analysis taught through standard classroom lectures versus taught through a format such as the Power Project. Our goal, however, is to have the Power Project materials serve as an adjunct to classroom discussions covering power and research design, rather than replacing classroom discussion outright.

Three open-ended questions were used to identify student likes, dislikes, and suggestions for improvement of the Power Project site and approach. Overall, directions were found to be easy to follow, and students liked the power-point presentation and links to related resources and sites. The research scenarios were identified as a promising approach for teaching concepts in power analysis. However, navigation from the tutorial to the *nQuery* program needs to be more clear, and some window transitions improved. Students suggested adding more complex tutorial scenarios, which is part of our ongoing modifications to the Power Project site. Also suggested was a section link to information on when to use particular statistical tests or methods.

## DISCUSSION

Overall, our results suggest that the Power Project website is perceived by students to be helpful and instructive, and although our sample size limits our conclusions we did find positive effects on student knowledge of power as well as in application via (rather simple) authentic design situations. Our approach to teaching power analysis through research based power analysis lessons and hands-on access to power analysis software offers instructors a creative and effective option for instruction in these topics. Other instructional approaches, however, have proved useful as well. Horgan (1999) provides an excellent discussion and resource links on the use of spreadsheets for demonstrating concepts of power and variability in the classroom, although he acknowledges that spreadsheets may be limiting in terms of complexity of the desired statistical design. West and Ogden (1998) present a simple interactive demonstration on the World Wide Web which allows students to visualize how power changes as a function of sample size, standard deviation, and true mean change for a one-sample *Z*-test. We found their work to be a useful introduction to the definition of power of a statistical test. Kraemer (1985) discusses a strategy for using just one power table for teaching concepts of power, rather than the many different tables provided in Cohen's (1988) comprehensive volume. However, Gatti and Harwell (1998) argue that widely available power analysis software should be used for instructional purposes rather than often difficult-to-read power tables and charts (admittedly, much of this software was not available during Kraemer's work).

Although there are many websites that discuss power and effect size, and some that provide freeware for power analysis, we found none that actually offered a tutorial based on concepts of research design and planning, or that emphasized effect size estimation for the

statistical designs we were interested in including. We believe our approach may be a very efficient and productive way to learn about sample size calculations and related issues of power, variability and effect size. While we clearly prefer the *nQuery Advisor* to the alternative software and freeware available, it should be possible for instructors to adapt our approach to the software of their choice. Our materials and website are freely accessible and we welcome all constructive feedback on the Power Project.

REFERENCES

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

Allison, D.B., Gorman, B.S., and Primavera, L.H. (1993). Some of the most common questions asked of statistical consultants: Our favorite responses and recommended readings. *Genetic, Social and General Psychology Monographs,* 119(2), 155-185.

Ball, E. W. and Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly,* 26(1), 49-66.

Cohen, J. (1992). A power primer. *Psychological Bulletin,* 112(1), 155-159.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Collins, A., Brown, J.S. and Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser*, (pp. 453-494). Hillsdale, NJ: LEA.

Curtis, D. A. and Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education, 6*(1), http://www.amstat.org/publications/jse/v6n1/curtis.html.

Elashoff, J. D. (2002). *NQuery Advisor Version 5.0 User's Guide*. Los Angeles, CA: Statistical Solutions Limited.

Ferrall, C. (1995). Interactive statistics tutorials in *Stata*. *Journal of Statistics Education, 3*(3), http://www.amstat.org/publications/jse/v3n3/ferrall.html.

Garfield, J .B. (1995). How students learn statistics. *International Statistical Review*, 63, 25-34.

Gatti, G. G. and Harwell, M. (1998). Advantages of computer programs over power charts for the estimation of power. *Journal of Statistics Education,* 6(3), http://www.amstat.org/publications/jse/v6n3/gatti.html.

Goldstein, R. (1989). Power and sample size via MS/PC-DOS computers. *The American Statistician,* 43, 253-260.

Hoenig, J. M. and Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician,* 55(1), 19-24.

Holcomb, J. P., Jr. and Ruffer, R. L. (2000). Using a term-long project sequence in introductory statistics. *The American Statistician,* 54, 49-53.

Horgan, G. W. (1999). Use of spreadsheets for demonstrating experimental power and variability. *Journal of Statistics Education,* 7(1), http://www.amstat.org/publications/jse/v7n1/horgan.html.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement,* 56(5), 746-759.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement,* 61(2), 213-218.

Kraemer, H. C. (1985). A strategy to teach the concept and application of power of statistical tests. *Journal of Educational Statistics,* 10(3), 173-195.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods,* 5, 241-301.

Shankweiler, D., Lundquist, E., Dreyer, L. G., and Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing: An Interdisciplinary Journal,* 8, 267-294.

Stanovich, K. E., Cunningham, A. E., and Feeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly*, 19(3), 278-301.

Thomas, L. and Krebs, C.J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America*, 78(2), 126-139.

Tiggemann, M. and Lynch, J. E. (2001). Body image across the life span in adult women: The role of self-objectification. *Developmental Psychology*, 37(2), 243-253.

West, R.W. and Ogden, R.T. (1998). Interactive demonstrations for statistics education on the world wide web. *Journal of Statistics Education*, 6(3), http://www.amstat.org/publications/jse/v6n3/west.html.