



Theses and Dissertations

2008-01-16

Using Interactive Diagrams to Teach Graduate Students About Statistical Power

Eric D. Hunter
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Educational Psychology Commons](#)

BYU ScholarsArchive Citation

Hunter, Eric D., "Using Interactive Diagrams to Teach Graduate Students About Statistical Power" (2008). *Theses and Dissertations*. 1287.
<https://scholarsarchive.byu.edu/etd/1287>

This Selected Project is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

USING INTERACTIVE DIAGRAMS IN A WEB-BASED TUTORIAL TO TEACH
GRADUATE STUDENTS ABOUT STATISTICAL POWER

by

Eric D. Hunter

A master's project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Instructional Psychology and Technology

Brigham Young University

December 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Eric D. Hunter

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Richard Sudweeks, Chair

Date

Paul Merrill

Date

Stephen Yanchar

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Eric D. Hunter in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable to fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Richard Sudweeks
Chair, Graduate Committee

Accepted for the Department

Andrew S. Gibbons
Department Chair

Accepted for the College

Richard Young
Dean, David O. McKay School of Education

ABSTRACT

USING INTERACTIVE DIAGRAMS IN A WEB-BASED TUTORIAL TO TEACH GRADUATE STUDENTS ABOUT STATISTICAL POWER

Eric D. Hunter

Department of Instructional Psychology and Technology

Master of Science

This report describes the design, development, implementation, and evaluation of a web-based tutorial designed to teach graduate students about the statistical concept of power. It contains a literature review of techniques used to teach statistics, similar computer-based programs for representing the concept of power, and instructional theories that pertain to web-based tutorials. It describes the process of designing and developing this tutorial in detail. The results section contains a description of the product implementation with three different groups and discusses the qualitative and quantitative findings from each of these implementations. Finally, there is a discussion of the tutorial's strengths and weaknesses.

ACKNOWLEDGEMENTS

I would like to express my gratitude to several people who helped me through the long process of this project. First, I want to thank my wife who took care of two energetic boys alone many nights while I read articles, programmed the tutorial, and wrote my project up. There is no way I could have finished this project without her help and support.

Second, I would like to thank Dr. Richard Sudweeks. Not only did he inspire me as a student in his classes through his excellence in teaching, but he went out of his way to help me complete this project in time to graduate. I hope to emulate his kindness, attention to detail, and passion in teaching.

Finally, I would like to thank Dr. Paul Merrill and Dr. Stephen Yanchar. Without Dr. Merrill's help, I would never have been able to program this tutorial. One of my regrets coming out of this program is that I did not take more classes from him. Dr. Yanchar also went out of his way to help me finish this project at the last minute at a time when he was very busy. I have benefited from his feedback and advice on many occasions.

Table of Contents

Chapter 1: Introduction.....	1
Statement of Purpose	2
Description of Target Audience.....	2
Chapter 2: Literature Review.....	4
Instructional Content.....	4
Instructional Materials	10
Power-Sim.....	10
Power of a Hypothesis Test Applet	11
The WISE Power Applet	11
Statistical Java.....	11
Instructional Theory Review.....	12
Chapter 3: Method	16
Analysis.....	16
Subject Matter.....	16
Audience	16
Delivery Needs.....	18
Design	19
Instructional Strategies.....	20
Scope and Sequence.....	20
Interface Design	21
Development.....	33
Phases.....	33

Instruments Developed	37
Implementation	39
Procedure with First Small Group	40
Procedure with Second Small Group	41
Chapter 4: Results	43
Results of First Small-Group Implementation	43
Quantitative Results	43
Results of Interviews	43
Results of Second Small-Group Implementation	44
Quantitative Results	44
Results of Interviews	44
Results from Both Classes	45
Qualitative Results	45
Quantitative Results	47
Chapter 5: Discussion and Recommendations	50
Product Strengths	50
Product Weaknesses	51
Reflections	52
Schedule	55
Budget	56
References	59
Appendix A	62
Appendix B	63

Appendix C	70
Appendix D	71
Appendix E	74
Appendix F	75
Appendix G	77
Appendix H	81

List of Tables

Table 1. Content outline of power tutorial.....	36
Table 2. Pretest and posttest results by class	49
Table 3. Estimated and actual schedule for power tutorial.....	57
Table 4. Estimated and actual costs for the power tutorial.....	59
Table 5. Summary of estimated and actual cost totals for the power tutorial.....	59

List of Figures

Figure 1. Screen shot of a Java applet in a web browser.	9
Figure 2. Screen shot of tutorial interface.	23
Figure 3. Screen shot of the table of contents.	23
Figure 4. Screen shot of "Things to Remember" section of interface.	25
Figure 5. Screen shot of "Things to Remember" pop-up.	26
Figure 6. Screen shot of the presentation section of the interface.	27
Figure 7. Screen shot of text file providing content for a question in the tutorial.	27
Figure 8. Screen shot of presentation template A.	30
Figure 9. Screen shot of presentation template B.	30
Figure 10. Screen shot of presentation template C.	31
Figure 11. Screen shot of text-only multiple-choice question template.	31
Figure 12. Screen shot of multiple-choice question template with graphics as options. ...	32
Figure 13. Screen shot of short-answer question template.	32
Figure 14. Screen shot of multiple-choice question template with graphic.	33
Figure 15. Screen shot of select-all-that-apply question template.	33
Figure 16. Screen shot of possible new interface for tutorial.	54

Chapter 1: Introduction

Taking at least one statistics class is a normal part of a graduate education in many disciplines. This class is meant to help students prepare to plan, conduct, and interpret research that is relevant to their degree of study. Having a basic understanding of statistics is important to the success of graduate students. At the same time, statistics can be a very discouraging course for many students; they come into these classes with varying competency levels in math and even more varying needs for knowing statistics. In recent years, statistics teachers have realized the importance of using a variety of methods to meet the varied needs of their students (Fernandez & Liu, 1999). Techniques have included more interaction with computers, group work, semester-long projects, and using meaningful data for analysis. Using these techniques has yielded positive results both in terms of student understanding and student enjoyment of the subject matter (Aberson, Berger, Healy, & Romero, 2002, Fernandez & Liu, 1999).

Statistics teachers are also focusing more on conceptual knowledge than computational ability (Howell, 2002). With the increased ease and power of using computer programs like SPSS and SAS to crunch numbers, students can focus more on a clear understanding of what a particular test statistic is used for and how it is helpful in analyzing data.

The Brigham Young University department of Instructional Psychology and Technology (IP&T) requires students to take at least one statistics course, IP&T 550. This course serves as an introduction to statistical research and methods. Students of the program can also take IP&T 651, which is a more in-depth statistics course. Rowe (2004) developed an interactive Flash-based tutorial for teaching simple linear regression to

students who take IP&T 550. His results indicated that students would like to use more of these tutorials to learn statistical concepts.

Statement of Purpose

To meet the needs of students enrolled in IP&T 550 and 651, the designer created an interactive Web-based tutorial on the subject of statistical power. This tutorial was designed to help students understand the concept of power in enough depth to both plan studies with adequate power and conduct post hoc analysis on studies that yielded non-significant results. This tutorial only teaches basic computation of power, and refers students to computer programs that can do more specific computations for them. Specifically, this tutorial is intended to teach students what power is, the factors that affect the power of a statistical test, how to plan a test with adequate power, and how to conduct a post hoc power analysis. The specific learning outcomes that this project is designed to facilitate are listed in Appendix A.

In addition to the tutorial, the designer also developed an assessment to determine how well students understand the concept of power. This assessment served as the pre- and posttest for the small-group evaluation. The designer also created a follow-up questionnaire to collect students' opinions about the effectiveness of the tutorial.

Description of Target Audience

This tutorial was developed for graduate-level students in the social sciences. These students are preparing to conduct their own research studies, so statistics is a vital subject for them to learn. Although they are generally motivated to learn the subject, these graduate students come from varied backgrounds; some feel that they have a strong

foundation in math and others list mathematics courses as their least favorite. The students are usually comfortable working with computers, and have used computer tutorials of one form or another before. Most say that they would be interested in using a computer tutorial for learning statistics. When asked what would make statistics easier for them to learn, the responses included making the subject more meaningful, reducing the amount of reading, and including more graphics and graphs to explain concepts.

Since professors will also utilize the materials, they are a second audience for this project. They are highly knowledgeable in the subject of statistics, but are pressed for time during lectures to fit everything in. They have a sincere desire to help students learn the subject matter and are always looking for ways to enhance their courses.

Chapter 2: Literature Review

Instructional Content

Despite the fact that it is now commonly taught in statistics classes, power continues to be one of the most ignored concepts in actual research studies, particularly in the behavioral sciences. In 1962, Cohen began publishing articles and books emphasizing the importance of power in experiments. Twenty-seven years later, Sedlmeier and Gigerenzer (1989) examined 64 studies to determine whether Cohen's work had had any effect. They found that only 2 of the 64 published experiments they examined even mentioned power. Additionally, they found that researchers retained the null hypothesis frequently despite the fact that the power of their experiments was very low (the median power for these experiments was .25). Behavioral scientists, however, are not the only researchers who neglect power. Thomas and Krebs (1997) point out that zoologists and ecologists are frequently unaware of the power of their tests. In general, researchers seem to disregard the risk of a Type II error (the probability of concluding that the null hypothesis is true when it should be rejected) because of their failure to consider the power of their experiments.

The effect size of treatments in the behavioral sciences tends to be small ($d = 0.20$) (Lipsey & Wilson, 1993). For example, consider a study performed by Williams (1990) in which he tested the effects of technology-based, non-technology-based, and combined teaching strategies on mathematically disadvantaged children. The effect size of the study was 0.14, and the sample size was 127 students. Based on this effect size and the desire for an 80% chance of being able to see a result if one had been present, this study should have had a sample size of at least 400 students. In the field of instructional psychology,

we should commonly expect to have small effect sizes, so for adequate power, we need large sample sizes. Despite this reality, most studies do not approach the needed number of participants. As a result, many researchers are not able to reject the null hypothesis when it is false simply because their studies lack adequate power.

There are many explanations as to why power seems to be ignored. Some suggest a perception among researchers that there is a lack of software able to compute power (Thomas & Krebs, 1997). Others propose a historical explanation, pointing out that the Fisherian ideas of null-hypothesis testing have been so predominant as to exclude analysis of power (Sedlmeier & Gigerenzer, 1989). It seems that researchers in the field of instructional psychology do not seem to understand the importance of power. They seem overly concerned with the likelihood of a Type I error, and, as a result, the chance of a Type II error is ignored. This may simply be due to tradition.

In addition, statistics teachers note that power can be very difficult to teach (Aberson et al., 2002) as it requires an integration of many concepts. A misunderstanding of statistical power could be the reason it is not commonly a part of research designs. Statistics teachers typically use diagrams to help learners visualize the principles associated with power, but these diagrams are difficult to produce on the fly and manipulate in front of a class of confused students. As Aberson et al. (2002) point out, the diagrams on the chalkboard must be accompanied by “a great deal of hand waving” (p. 2). Besides the difficulty of graphically representing the abstract concepts of statistics, other issues make teaching this subject difficult, including:

1. Since the abstract concepts are difficult to represent graphically, they are not intuitive and students must be taught to understand them.

2. Students come from a variety of backgrounds because most fields of study use statistics (e.g., some students may not have a strong background in math).
3. Students lack motivation to learn statistics, thinking that it is just a hurdle to jump and be done with (Fenandez & Liu, 1999). For example, it is possible that students in the class may never actually intend to use statistics to conduct their own research. This presents a challenge to the teacher to make the material meaningful to the students.
4. Class periods are short, and there is so much information to cover in the lecture that information is often condensed (Fenandez & Liu, 1999).
5. Class periods are so short that it is difficult to squeeze in a variety of learning activities like group work, etc. (Fenandez & Liu, 1999).
6. Statistical formulas use the Greek alphabet, which is unfamiliar to most learners. The combination of learning elements of a new language, doing math, and dealing with abstract concepts all at the same time can be quite cognitively taxing.
7. Statistical concepts require students to know and understand many details. Not all statistics students use the information they learn right away, and so it can be easily forgotten.
8. As with many subjects, statistics requires learners to know some foundational concepts. If the struggling students do not master the foundational ideas, they will struggle with statistics. In a typical statistics class there is so much information to cover that it is not possible to wait for each student to master the content before moving on.

There are, no doubt, many other factors that make statistics a difficult subject to teach and learn. Any product that seeks to teach statistics must address some of these points in order to really help students of statistics.

Statistics teachers are constantly trying to improve the teaching of their subject. Entire journals have been dedicated to improve teaching techniques. Ideas range from using jingles to having students work on semester-long projects gathering real data. Some of these techniques are for use in the classroom, and others are computer applications. One popular form of computer application in the nineties was the Java applet, which was specifically designed for Internet use (West & Ogden, 1998). An *applet* is a computer application written in the Java programming language which is specifically designed to be used in a web browser (see Figure 1). The particular applet in Figure 1 allows students to type in numbers and then click a button to update the drawing of the curves. This medium has been used to teach the concept of power (Aberson et. al, 2002; Anderson-Cook & Dorai-Raj, 2003; West & Ogden, 1998; Yu & Behrens, n.d.) in an attempt to address some of the difficulties in teaching statistics. For example, these applets are excellent for demonstrations. As a result, professors have used them in lectures to minimize time spent drawing diagrams on the board. This frees up time in the lecture, and has permitted some teachers to include other activities, such as group work (Fenandez & Liu, 1999). Instructors have also assigned them as a follow-up to the lecture. Students completed these follow-up assignments by accessing the applets online and, in some cases, completing study questions as they interacted with the applets (Aberson et. al, 2002; Anderson-Cook, Dorai-Raj, 2003). The research of Fernandez and Liu (1999), Aberson et. al (2002), and Anderson-Cook and Dorai-Raj (2003) suggests that these

applets have helped boost test scores, and surveys reveal that students seem to enjoy using them to learn.

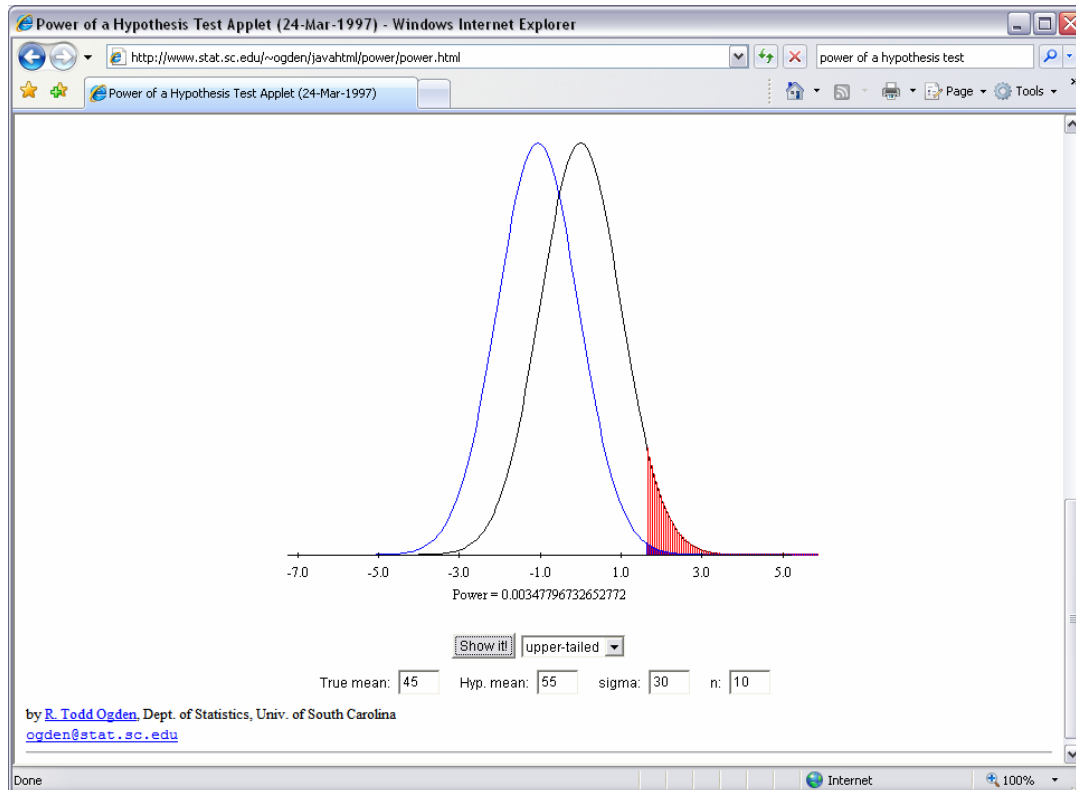


Figure 1. Screen shot of a Java applet in a web browser.

Applets are also powerful because they allow students a hands-on experience that they can not get from a textbook. It is easy to try and verbalize all of the information that statistics students need to know without letting them explore and learn through experimentation. Bruner (1964) suggests that there is a hierarchy to types of representations, and that it is easiest for novice learners to learn through hands-on activities (enactive representation), then through visual representations (iconic representation), and it is most difficult for them to learn through symbols (symbolic

representation). Unfortunately, it is easiest to teach through symbols. If we want to explain a complex idea, we often just start talking. According to Bruner (1964), speech and written language are forms of symbols. Combine this with the Greek alphabet used in statistics and it is easy to overwhelm learners. The applets can be used as an enactive representation, which should help novice learners greatly.

While they save time in class, help to better represent the concepts, and are possibly more aesthetically pleasing than some chalkboard drawings, Java applets were not designed to address all of problems facing teachers and learners of statistics. They are simply data-driven graphics that change based on the user adjusting some value or parameter. This quality is ideal for a lecture demonstration, but it has weakness as well. When used as a self-instructional unit in one instructional setting, students came away with many misconceptions (Yu & Behrens, n.d.). One instructor combated this problem by assigning his students a series of multiple-choice questions to answer while interacting with the applet. The distractors were common misconceptions, and students were encouraged to go back and interact with the applet any time they got a question wrong (Aberson et. al, 2002). These ideas provide the foundation for enhancing the helpfulness of an applet, but still leave instructional gaps.

Java applets could be more powerful tools if they—

1. Gave students guidance to use them effectively
2. Included some sort of knowledge scaffolding for students
3. Reduced the cognitive strain they place on students
4. Were placed in a meaningful context
5. Helped students remember and apply what they learn

6. Emphasized the importance of the subject they model

Each of these points adds a new dimension to the applet that would make it more useful.

Instructional Materials

In a desire to build on the work that had previously been done in the area of interactive statistics tutorials, the designer conducted a thorough search for these tutorials. Internet searches yielded large numbers of results, and several of these were developed specifically to teach statistical power. Statistics teachers have conducted studies in their classes using these products, and have seen positive results. The paragraphs that follow give a brief summary of some of the better products that the designer found. They are listed in the order in which they were developed.

Power-Sim

Yu and Behrens (n.d.) developed an interactive power diagram for use on the Macintosh platform called Power-Sim. This diagram allowed users to manipulate all three factors that effect power: effect size, sample size, and alpha. Further, users could easily control which information that they wanted to display. This makes focusing the learner's attention on a particular point much easier. Another positive aspect of this program is that it does not overwhelm students with numbers; the focus is much more conceptual with an emphasis on the three factors that influence power. Unfortunately, this product is not Web-based, and so is not a viable option for many statistics teachers to use. It is not a part of a tutorial, and so could only be used in a classroom demonstration.

Power of a Hypothesis Test Applet

West and Ogden (1998) developed the Power of a Test Hypothesis applet that creates a graph based on user input. Users are able to enter values for the true mean, the hypothesized mean, the standard deviation, and the sample size. Once these values have been entered, the applet will draw a graphic showing how the two distributions relate to each other. This applet produces excellent diagrams, but does not allow the user to change the value of alpha, and does not allow users to drag the curves to manipulate them.

The WISE Power Applet

Aberson et al. (2002) developed a tutorial that uses a paper-and-pencil worksheet to be completed while interacting with the WISE Power applet. This applet allows students to drag and manipulate the positions of the two curves (e.g., students can change d). As they do this, various values such as beta, power, and d are recalculated and displayed. Unfortunately, this applet only allows for one-tailed tests. Further, it is difficult to understand what exactly can be manipulated by dragging, and the connection of sample size, alpha, and effect size to power is not always clear. There is so much information available to the student, that it is difficult to focus on a particular aspect of power. This applet could be used very effectively in a classroom during a lecture as the teacher could help the students focus on one element at a time. This tutorial was not intended to be stand-alone instruction.

Statistical Java

Anderson-Cook and Dorai-Raj (2003) developed a series of java applets to teach power that are much more complex than their predecessors. These applets build on one

another and eventually give students full control over what to manipulate. They offer a much wider variety of tests (including one-tailed and two-tailed tests, etc.) and show concepts using different graphical representations than have previously been used. Unfortunately, students are once again bombarded with information. At least nine different values are calculated as students manipulate the graphics. Although helpful for advanced statistics students, these applets can be overwhelming to novices.

Instructional Theory Review

As in any type of training, multiple strategies are required to provide learners with an optimal environment for learning to take place. This section is a brief review of the foundational instructional strategies of this project.

According to Bruner (Driscoll, 1994), there are three means of presenting information to learners: enactively, iconically, and symbolically. An enactive representation is one where some sort of motor response is required. The second mode of representation is iconic, meaning information is presented using pictures. The third mode is symbolic representation. Language, spoken or written, is the primary means of symbolic representation. These modes are ordered from easiest to difficult, meaning that enactive is the simplest form of representation for learners to understand who are new to a subject and symbolic representation is the most difficult for novice learners to understand.

According to this theory, novice learners need more hands-on interaction. Interactive, data-driven graphics can be a powerful tool for them to grasp the concepts that they need. This served as a key strategy for this project: increase the amount of interaction in the tutorial. Not only does this have the potential for increasing learning outcomes, but several studies have documented the appeal and increased enjoyment of hands-on

interaction in statistics education (Fernandez & Lui, 1999; Anderson-Cook & Dorai-Raj, 2003).

Clark and Mayer (2003) present a variety of principles to make eLearning more effective and increase interactivity. The principles that related directly to this tutorial are (a) the multimedia principle, (b) the contiguity principle, (c) the personalization principle, and (d) the practice principle.

The multimedia principle states that graphics and words should be used together, rather than words alone (Clark & Mayer, 2003). Presented correctly, graphics can help learners form an accurate mental representation of the material being presented. Merrill and Bunderson (1981) point out that graphics can hinder learning if not used properly. They recommend that graphics should initially be simple and not contain "irrelevant cues" (p. 3). An overly complex graphic can place too much cognitive load on the learner and cause confusion.

Building on the multimedia principle, the contiguity principle states that corresponding words and graphics should be placed near each other (Clark & Mayer, 2003). Doing so can reduce the amount of cognitive strain imposed on learners as they struggle to understand a graphic or diagram. Placing words near text allows learners to more easily understand relationships by holding information in short-term or working memory long enough to process the information. This principle is further called into play with feedback. Feedback for practice exercises should be placed on the same screen. This allows learners to refer back to the question and their response as they read and process the feedback.

The personalization principle indicated that using a conversational, less-formal style of writing increases learning outcomes (Clark & Mayer, 2003). It is very common to use a very formal, academic style of writing in statistics textbooks. This formal style can actually hinder learning. Using a more conversational style, including the use of second-person pronouns, helps learners feel more engaged in the learning process instead of "simply receiving information" (p. 136).

Finally, the practice principle suggests that training should include a great deal of practice "interspersed throughout the lesson rather than all placed at the end" (Clark & Mayer, 2003, p. 159). This allows learners to encode information into long-term memory by spacing out review and practice of material. It allows them to test the mental models that they have created of concepts to make sure that they are encoding information correctly. Clark and Mayer also demonstrate that more practice decreases the amount of time that it takes learners to complete a task. Simply stated, the more learners practice, the more they will learn. Clark and Mayer also state that it is important to apply the previously stated principles to practice exercises.

Driscoll (1994) points out that in order to increase the likelihood of information being added to long-term memory (LTM), it should be represented to learners in a variety of different contexts. In other words, concepts should be presented in a range of forms to help learners discriminate between similar concepts. She recommends the use of a rational set generator, or a table showing how similar concepts differ from each other in a variety of scenarios, as a way of producing examples for explanation and testing that vary enough to help learners clearly see the boundaries between concepts in a variety of situations. A common blunder in instruction is to limit the range of examples for learners

in order to reduce cognitive load. While using the same examples repeatedly does reduce cognitive load, Clark, Nguyen, and Sweller (2006) have shown that the additional mental effort that varied examples and scenarios place upon the learner actually "end up with a much broader repertoire of . . . skills applicable to many . . . situations" (p. 11). While cognitive load can have a negative impact on encoding to LTM, not all cognitive load is the same, and the right kind of load, referred to as "germane load" by Clark, Nguyen, and Sweller (2006), is desirable and helpful for learners in the long run.

In summary, instruction should be carefully planned to include hands-on interaction. Carefully designed graphics should be used to reinforce mental models. Practice should be scattered throughout the tutorial to allow learners to test themselves as they progress. Material should be presented in a less-formal, conversational style. Examples and practice exercises should be presented in a variety of situations and scenarios to encourage a more broadly developed set of skills long term.

Chapter 3: Method

In order to develop an instructionally sound tutorial, the designer followed the instructional design process outlined by the ADDIE model (Analyze, Design, Develop, Implement, and Evaluate) (Lohr, 1998). The designer chose this model because it is commonly used in industry today and it met the needs of this project.

Analysis

The analysis phase consisted of three steps: an analysis of the subject matter, an audience analysis, and a delivery-needs analysis (Lohr, 1998).

Subject Matter

To analyze the subject matter, the designer reviewed both introductory and advanced textbooks (see Hinkle, Wiersma, & Jurs, 1998; Howell, 2002; Sprinthall, 2000), books specifically written about power (see Cohen, 1988; Murphy & Myors, 1998), as well as journal articles (see Cohen, 1970; Cohen, 1992; Lipsey & Wilson, 1993; Prentice & Miller, 1992; Sedlmeier & Gigerenzer, 1998). The designer also talked with Dr. Richard Sudweeks who served as the subject matter expert for this project.

Audience

Primary audience. The primary audience consisted of graduate students in the Instructional Psychology and Technology (IP&T) program. Since Rowe (2004) had recently completed a similar tutorial, many of the students the designer spoke with were familiar with the concept and were enthusiastic to share their thoughts on statistic tutorials. The designer conducted several focus groups with students in the IP&T lab, asking them how they felt about statistics, how they felt about other statistics tutorials,

and what they thought needed to be included in a new tutorial. The designer also interviewed Rowe (2004) on several occasions, as he had recently developed a similar tutorial. He was able to offer development suggestions based on the strengths and weaknesses of the tutorial he developed.

As development started, the designer conducted a low-fidelity implementation of the tutorial in Dr. Sudweeks' IP&T 651 class during winter semester 2005. At the conclusion of the class, students filled out the *Statistics Background Questionnaire* developed by the designer (see Appendix F). The questionnaire gathered basic demographic data, as well as information about students' experience with statistics, their familiarity with computers, access to the Internet, reasons for taking a statistics class, things that would motivate them to learn statistics, etc. Most of the students in the primary audience were seeking a PhD, and already had a master's degree. A majority of these learners said that they were "very interested" in learning statistics, and were likely to take another statistics class. At the same time, most reported that they were taking statistics because it was required for their program of study. Most of them reported that they were "very motivated" to learn statistics. All of the learners reported being familiar with Microsoft Excel, at least one type of statistical analysis software (e.g., SPSS, SAS, MiniTab), Microsoft Power Point, and web browsers. When asked whether they were slow, average, or fast learners in general, most reported being average, and no one reported being a fast learner. Finally, most of the students were very interested in a computer-based tutorial to help them study statistics, and most indicated that they would prefer to learn statistics using a computer-based tutorial over a textbook.

Secondary audience. A secondary audience for this project was professors who teach statistics in the classroom. Several journal articles indicated that this audience was very much in favor of dynamic graphics that could be easily manipulated to illustrate a principle (Aberson et. al, 2002, Anderson-Cook & Dorai-Raj, 2003, Gatti & Harwell, 1998). In an interview with Dr. Sudweeks, the designer learned that in dealing with power, he would like three interactive diagrams to illustrate the three factors that influence power: sample size (variance), effect size, and significance levels (the value of α). Based on these sources, the designer compiled a list of needs for this audience that included quick access to the interactive diagrams (i.e., they could not be embedded deep in a tutorial) and a delivery medium that was easily accessible for use in classes.

Delivery Needs

The system of delivery was greatly influenced by the needs of the audience and subject matter. Since one of the main objectives of the tutorial was to provide instructors and students with interactive diagrams, a computer-based tutorial was necessary. This medium also supports a self-study mode of learning which is consistent with the needs of an audience of adult learners. Making this an Internet-based tutorial as opposed to a client-side application provided several advantages: the audience did not need to worry about installation CDs, professors could easily access the tutorial for lectures or other teaching moments with no more than an Internet connection, and students could study the tutorial outside of class on any computer with an Internet connection. It was very obvious during the focus groups that the audience had easy access to the Internet both at home and at school.

Since the tutorial was to be delivered via the Internet, it was important to choose the right medium for getting it on the Internet. To this point, most applications of this nature have been programmed in Java and are accessible online as a Java Applet. The advantages of using Java Applets are that Java does not have as many cross-platform compatibility issues as other programming languages (i.e., it can run on a computer with Windows, Macintosh, or Linux operating systems), and it is powerful enough as a programming language that creating any type of data-driven animation is not a problem. Since the designer did not know Java though, it was not a viable option for this tutorial. Another strong option for creating the tutorial was to use Macromedia Flash. Flash is a very popular program for creating web-based applications. Its ActionScript is robust enough to create the types of animations needed for this tutorial. Further, it has a deep market penetration—98.7% of computers that can access the Internet today have a working version of the Flash Player installed on them, as opposed to 87.5% of computers with Java installed on them (Adobe, 2007). Finally, Flash includes an interface which allows developers a great deal of control over the aesthetics of the program. In the end, Flash appeared to be the best option for developing the tutorial.

Design

After completing the analysis, the designer began to design the tutorial. Lohr (1998) suggests that during the design phase of the ADDIE model, the designer should use "information from the analysis stage to create instructional strategies which address the instructional goals and objectives" (p. 5). The designer also created the scope and sequence, interface design, and tutorial templates during this phase.

Instructional Strategies

The analysis of instructional strategies began with a list of the objectives for this tutorial. The designer then determined what students should be expected to remember at various intervals (i.e., after one week, after a month, after six months). Next, the designer listed the current weaknesses of comparable products and what should be done to solve them. This process served as the basis for a review of instructional strategies. With these goals in mind, the designer reviewed several instructional design texts for instructional strategies. An attempt was made to find a specific strategy for dealing with each issue or objective. After compiling a list, the designer met with Dr. Stephanie Allen to review the list of strategies. After a review of the audience and the tutorial objectives, Dr. Allen and the designer evaluated each strategy. Several were identified as fundamental to the outcome of the tutorial, and others were eliminated based on the scope of the project. For example, the designer had wanted to make the tutorial adaptive to the learners. But after a discussion of the technical requirements of this feature (i.e., creating a database, creating an account for each user), it was determined that adaptability was not a critical need for a tutorial of this length.

Scope and Sequence

The scope and sequence was a detailed plan that showed when and how instructional objectives would be met. It included a list of the objectives each section would address, a list of the key terms and equations for the section, and an outline of the content that was included in the tutorial and the flow of instruction, including practice exercises. The designer developed the scope and sequence by reviewing several textbooks (see Hinkle, Wiersma, & Jurs, 1998; Howell, 2002; Sprinthall, 2000), articles (Cohen, 1970; Cohen,

1992; Lipsey & Wilson, 1993; Prentice & Miller, 1992; Sedlmeier & Gigerenzer, 1998), and books (Cohen, 1988; Murphy & Myors, 1998) written about power to determine how power was usually presented to students. The designer carefully outlined how these experts presented the concept of power, compared them to each other, and developed a combined expert outline that included all of the major points of each expert. In some cases slight adjustments in the order had to be made due to differences between the experts. Once the combined expert outline was complete, the designer analyzed it using the instructional objectives and instructional strategies. After making several modifications, the designer presented the revised scope and sequence to Dr. Richard Sudweeks who served as the SME and sponsor of the project. Based on his feedback, the designer made additional changes and clarifications. The final scope and sequence is shown in Appendix B.

Interface Design

The area of the interface was divided into three main sections: a table of contents, a "Things to Remember" section, and the presentation section where the instructional content appeared (see Figure 2). These elements are a mixture of web technologies. The table of contents and "Things to Remember" sections were both programmed with JavaScript. The presentation section contained Flash Shockwave files (SWF) which contained both presentation and practice content. Each section was a separate html file, and the files were presented together using html frames to separate them.

Table of contents. The table of contents (see Figure 3) served as a global navigation element, allowing learners to quickly jump to any section of the tutorial. It also allowed

learners to see the breadth of the content in the tutorial so they could anticipate what would be covered.

The screenshot shows a tutorial interface with a dark header bar containing 'Contents' and 'Section 1.1: Definition of Power'. On the left, a table of contents lists sections 1 through 6 and an appendix. The main content area is titled 'Slide 1 of 6' and contains three paragraphs of text. The first paragraph explains a scientist's task of checking water for a parasite. The second paragraph lists three tools: the naked eye, a magnifying glass, and a microscope. The third paragraph states that a microscope is the most powerful tool. To the right of the text are three icons: an eye, a magnifying glass, and a microscope. At the bottom right, there are 'previous' and 'next' navigation buttons.

Figure 2. Screen shot of tutorial interface.

Contents
Introduction
Section 1: What is Power?
1.1 Definition of Power
1.2 Review of a Normal Curve
1.3 Type I and II Errors
1.4 How Power is Measured
1.5 Factors that Influence Power
1.6 Section 1 Practice
Section 2: Sampling Error
Section 3: Effect Size
Section 4: Standards for Rejecting H_0
Section 5: δ (delta)
Section 6: Increasing the Power of a Study
Appendix: Interactive Animations

Figure 3. Screen shot of the table of contents.

The table of contents was organized into two levels: sections and subsections. Initially only the section names were visible. Clicking on the name of a section expanded it and the subsections then became visible. Once users selected a subsection, the link color changed to orange as a visual reminder to the users as to which section they were in.

The structure of this navigation system seemed to lend itself to the needs of adult learners. Users could easily jump to any section or subsection of the tutorial at any time. Seeing the table of contents could serve as an indication of progress through the tutorial. This was also the element of the interface that addressed the issue of learner versus program control.

Clark and Mayer (2003) point out three "arenas" of control, which are *content sequencing, pacing, and access to learning support* (p. 227). They further point out that although learners like to have more control over navigation, they learn better when the program controls the navigation until the learners have a strong understanding of the subject matter. In order to balance the needs of the audience, the table of contents allows learners to jump to a subsection, but within the subsection, they follow a program-imposed path through the content. Clark and Mayer (2003) found that learners with less familiarity with the subject matter appreciated a "course map" but rarely used it for navigation (p. 241).

The table of contents also allowed the secondary audience of professors quick and easy access to the interactive diagrams in the tutorial through the appendix links. Clicking on the "Appendix: Interactive Animations" link revealed links to the three interactive diagrams in the tutorial. These could be used during lectures in front of the class or

during one-on-one help sessions between the professor and students that needed extra help.

"Things to Remember" section. The "Things to Remember" section (see Figure 4) was also always visible to learners as they progressed through the tutorial. Since none of the learning objectives for this lesson included the memorization of formulas, key formulas used to compute power were accessible in this section.

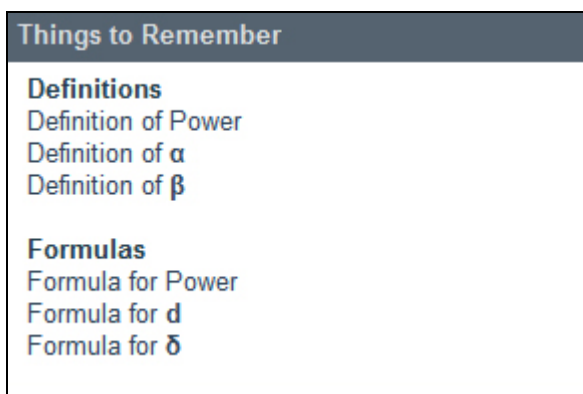


Figure 4. Screen shot of "Things to Remember" section of interface.

Clark and Mayer (2003) suggest that memory aid elements like those found in this section open a pop-up or tool tip so that learners can see the information they need without losing their place in the tutorial. Accordingly, clicking on any of the links in the "Things to Remember" section will open a small pop up containing the needed information. Figure 5 shows an example of a pop-up that appears when you click on the "Formula for d " link from this section. These pop-ups also serve as a way of eliminating what Clark, Nguyen, and Sweller (2006) refer to as "extraneous cognitive load" by "display[ing] content in external memory resources" (p. 139). This allowed learners to focus on the conceptual understanding called for by the bulk of the learning objectives.

The image shows a small window titled "Formula for d - Windo...". Inside the window, the text "Formula for d" is displayed above the mathematical formula $d = \left(\frac{\mu_A - \mu_O}{\sigma} \right)$.

Figure 5. Screen shot of "Things to Remember" pop-up.

Presentation section. The presentation section (see Figure 6) was the largest portion of the interface. This is where the instructional content was displayed through a Flash Shockwave File (SWF). Each subsection constituted a separate SWF, so in all, 40 separate SWFs were created for this tutorial. Because of the large number of screens and files, and to allow for greater flexibility, these SWFs were designed to be populated with content from separate text files. Figure 7 shows an example of a text file that populates a question in Section 1.1 of the tutorial. There were several distinct advantages to this design: (a) it allows for a greater number of content developers to contribute to the project without requiring special programs or tools, (b) it allows developers to change content without modifying the presentation templates, (c) it decreases the size of the files the end user will see, thus increasing the speed that they are loaded in a browser, and (d) it allows for content to be reused in other activities, lessons, etc. Each of these points will be discussed in greater detail.

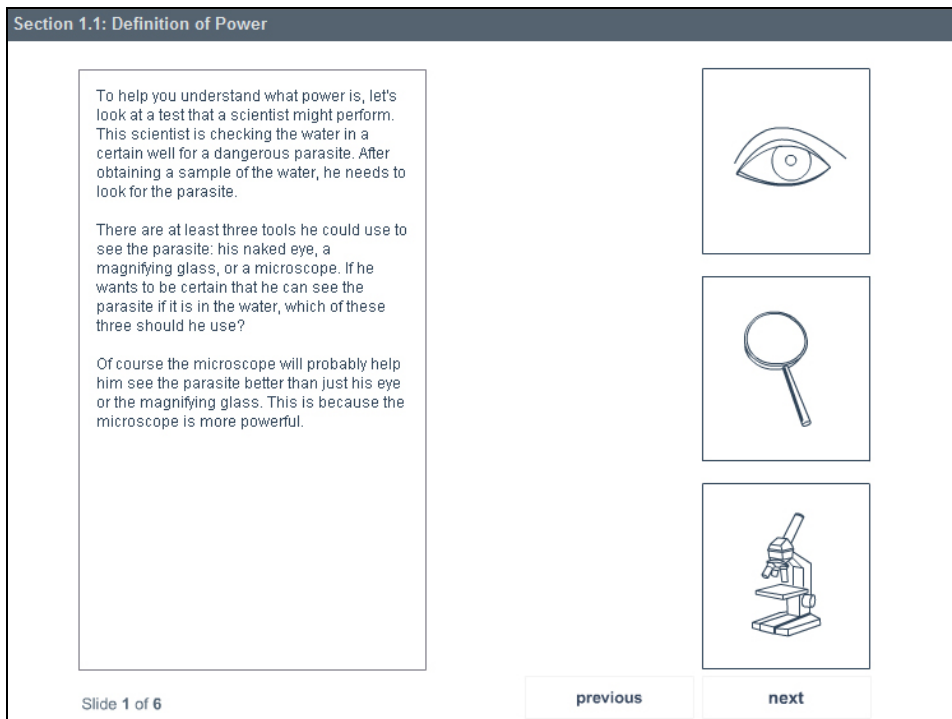


Figure 6. Screen shot of the presentation section of the interface.

```

numOptions=3
&question=which of these is a correct definition of power?(click on the
letter next to the question)
&option1=Power is the possibility that a test will <b>not</b> yield
statistically significant results.
&feedback1=<b>Incorrect</b><br><br>Power is all about getting significant
results. Is there another option that puts this better?
&option2=Power is the probability of rejecting a false null hypothesis.
&feedback2=<b>Correct</b><br><br>Since rejecting a false null hypothesis
means you got significant results, the probability that you will is power.
&option3=Power is the likelihood of rejecting a false alternative
hypothesis.
&feedback3=<b>Incorrect</b><br><br>The key word in this option is
<b>alternative hypothesis</b>. If you reject an alternative hypothesis, you
do not have enough power.
&option4=Power is the degree to which we are confident that we can reject a
true null hypothesis.
&feedback4=<b>Incorrect</b><br><br>we only want to reject a null hypothesis
if it is not true.

```

Figure 7. Screen shot of text file providing content for a question in the tutorial.

First, separating content from the delivery medium allows for a greater number of developers to contribute to content development without needing special programs or tools. If content were embedded within the SWF files, each person contributing to content development would have to use Flash. By pulling content from text files, anyone helping develop content can simply use a basic text editor like Notepad, which is included as a free accessory of Microsoft Windows. Not only does this help in the initial push to create the content, but it also allows for editors and proofreaders to easily access and modify content without requiring them to know how to use Flash or have access to Flash. This was, perhaps the most significant advantage in this project for using external content files. It allows SMEs to alter and update the content of this lesson after the developer has completed the project and can no longer maintain it.

Second, the separation of content from the delivery medium allows for quick changes to content without modifying delivery templates. It is faster and easier to correct a misspelled word in a text file and save it than to open Flash, find the misspelled word, and correct it. This was a great advantage during this project. The developer was able to make small editorial changes to the materials very quickly using this method.

Third, the file size of the SWFs is reduced when content is separated, allowing them to load more quickly over the Internet. The difference in size will vary depending on what is being loaded externally; it can range from a few kilobytes (KB) for external text files to several megabytes (MB) for external graphics. This impact can be very significant for users with a dial-up Internet connection. This was not a major advantage for this project since most users accessed this tutorial on BYU campus using a high-speed Internet connection. Based on the speed of their connection, they would probably never be able to tell a difference.

Finally, separating content allows it to be reused. For example, this is the second of two statistics-based tutorials that have been developed for BYU's Instructional Psychology and Technology department (see Rowe, 2004). If the department decided to create an online statistics course and wanted to use these two existing tutorials, they would probably want to create a unified look or template. Since the content for this tutorial is external to the template, it can easily be moved into a new template with no reworking of content. Since content development investments can be high, reusing content can be a way to maximize an investment. This is an important topic in training today, which has led to the *SCORM* movement to help make reusing content more cost effective through standardization. This factor had a minimal effect on this project since there are not currently any plans to integrate this tutorial into a larger statistics course.

In order to maximize the efficiency of developing in Flash, the designer created a limited number of templates that were flexible enough to be used with a variety of content scenarios. Three main templates (see Figures 8–10) served as a basic structure for presenting all of the content in the tutorial. Presentation template A (Figure 8) allowed for text with a quote, graphic, or formula on the right half of the screen. Presentation template B (Figure 9) was used when the graphic was large enough to take up the entire width of the screen, and presentation template C (Figure 10) was used with animations. Each of these templates pulled text, graphics, and animations dynamically into the SWF from external files. These templates were also specifically designed to allow students to select and copy the text from them into another program (e.g., Microsoft Word) for taking notes. During the implementations several students took advantage of this feature repeatedly and commented that they really appreciated it.

Templates were also used for practice questions. Although there were more of these templates, they also served to increase the efficiency of development. Figures 11–15 show the five main templates for practice items.

Section 1.1: Definition of Power

How does this relate to statistics? Well, in statistics we define power as **"the probability that a statistical test will yield statistically significant results"** (Cohen, 1988, p. 1). In other words, **power is the ability to see a result if there is one.**

If your test does not have enough power, you will not be able to reject the null hypothesis. Since you will probably spend a lot of time planning and conducting research, it is important that you make sure your tests have enough power.

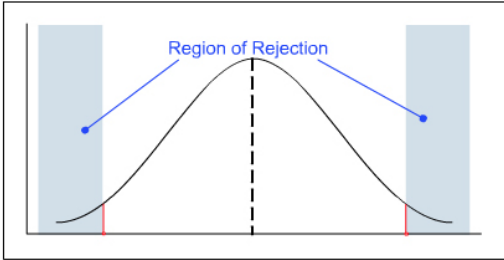
power is "the probability that a statistical test will yield statistically significant results"

Slide 2 of 6

previous next

Figure 8. Screen shot of presentation template A.

Section 1.2: Review of a Normal Curve



There are some other areas of the curve that you will need to be aware of.

Region of rejection

This is the area of the control groups curve to the right and left the two alpha levels. It is called the region of rejection because if the test groups mean falls in the region of rejection we are able to reject the null hypothesis.

Slide 2 of 2

previous next

Figure 9. Screen shot of presentation template B.

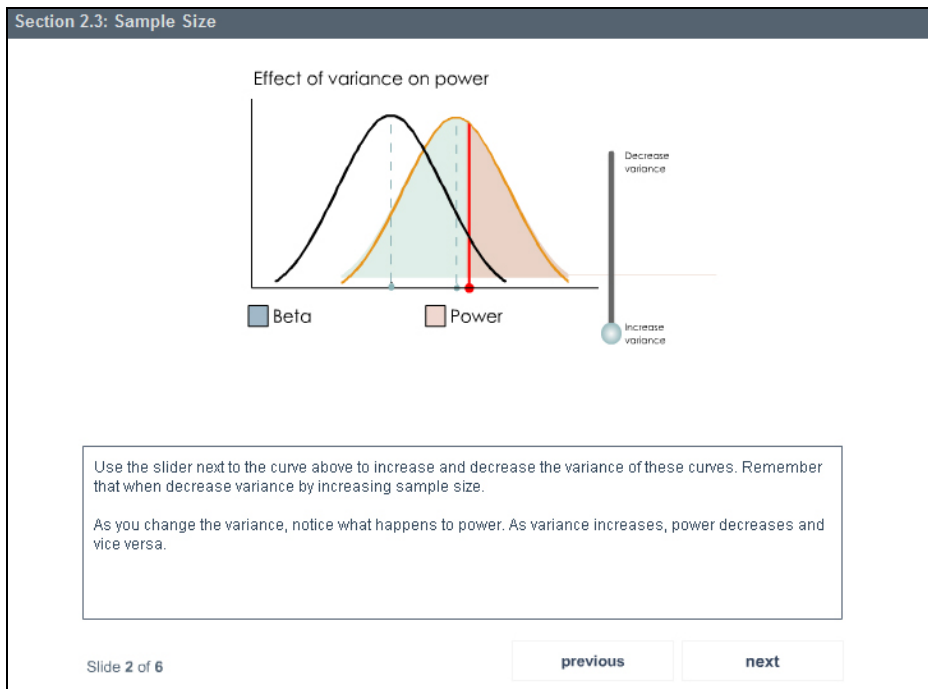


Figure 10. Screen shot of presentation template C.

Section 1.1: Definition of Power

Which of these is a correct definition of power?(click on the letter next to the question)

a Power is the possibility that a test will not yield statistically significant results.

b Power is the probability of rejecting a false null hypothesis.

c Power is the likelihood of rejecting a false alternative hypothesis.

d Power is the degree to which we are confident that we can reject a true null hypothesis..

Correct

Since rejecting a false null hypothesis means you got significant results, the probability that you will be power.

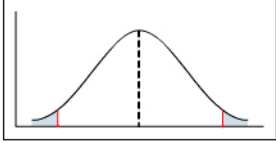
Slide 3 of 6

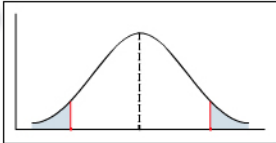
previous next

Figure 11. Screen shot of text-only multiple-choice question template.

Section 1.3: Type I and II Errors

Which study is more likely to make a Type I error?

a 

b 

Correct

Since alpha is bigger in this study, you are more likely to make a Type I error. the bigger alpha, the greater the chance of a Type I error.

Slide 13 of 14

previous next

Figure 12. Screen shot of multiple-choice question template with graphics as options.

Section 1.4: How Power is Measured

How much power do you have if Beta equals .4?

Correct

Since Power=1-Beta,
1-.4=.6.

Slide 6 of 8

previous next

Figure 13. Screen shot of short-answer question template.

Section 1.4: How Power is Measured

What do you estimate the power of this test to be?

a .2

b .8

Correct

This diagram looks like it has a lot of power. Since beta takes up about 20 percent of the area of the test group's curve, we can guess that power is around .8.

Slide 7 of 8

previous next

Figure 14. Screen shot of multiple-choice question template with graphic in the stem of the question.

Section 1.5: Factors that Influence Power

Which of the following things can influence the power of a test?

- the difference between the two group means
- the number of people in the study
- the number of people who will read the study
- the probability of a Type I error
- previous studies on the same topic

Check Answer

Click on the Check Answer button.

Slide 5 of 5

previous next

Figure 15. Screen shot of select-all-that-apply question template.

These templates allowed for multiple choice questions with text and graphics, short answer questions, multiple choice questions with graphics in the stem, and select-all-that-apply questions. Further, the templates were designed in such a way that each possible response would render unique and immediate feedback. Like the presentation templates, these practice templates were dynamically populated with content from external text and graphic files.

During the design of the interface, including the presentation and practice templates, the designer met regularly with Dr. Stephanie Allen to get feedback on instructional design techniques and strategies. The resulting templates proved to be flexible enough for most of the content in the tutorial.

Development

Phases

The development of this project was divided into two phases. The first phase consisted of writing the script for the tutorial, and the second phase consisted of developing the SWFs and checking them to make sure the content was populating them correctly. The content was developed in this order to assure that the templates would be developed based on the needs of the content and not the other way around. It is frequently a problem for instructional designers that the limitations of the templates will cause a loss of important instructional features (Evans, Kennedy, & Ure, 2004). This problem was greatly overcome by the fact that the roles of designer, content developer, and programmer were all filled by the same person. This concentration of roles had the advantage of being able to design in important features from the initial stages, but had the

downside of limiting the features to the programming knowledge of the designer, which was somewhat limited at the time of development.

Phase 1. When working with dynamic templates, it is essential to create standards to ensure that content will populate the templates correctly. Accordingly, some of the first steps in the development process included creating documents specifying a naming scheme for the external content files (including text, graphic, and animation files), html folder structure, and text length specifications. The text length specifications were very important so that scrollbars were unnecessary in the tutorial. Although all of the content was developed by the designer alone, these standards would have helped even more had a team of writers developed the content.

Based on the foundation of the scope and sequence document that was developed during the design phase (see Appendix B), writing the content was an efficient step. After writing several sections, the designer checked the content in some of the basic templates to ensure that the content sections were the right length, that the html formatting was being applied correctly, and that data structure of the external text files was correct. At this point several adjustments were made to get the content and templates working together correctly, and then the content development progressed quickly. The content outline is shown in Table 1.

Content was written in Microsoft Word and then copied and pasted into the Notepad text templates to minimize spelling and grammar errors. Once the text had been pasted into Notepad, the designer added html tags to format the text in a basic way, such as bolding headings. After the text was entered, curly quotation marks and apostrophes that Microsoft Word automatically adds had to be searched for and replaced with the straight version. Flash displays curly quotation marks and apostrophes as a box (□).

Table 1

Content outline of power tutorial

Section	Content
1	What is power? 1.1 Definition of power 1.2 Review of a normal curve 1.3 Type I and II errors 1.4 How power is measured 1.5 Factors that influence power
2	Sampling error 2.1 Definition of sampling error 2.2 Reliability and validity 2.3 Sample size
3	Effect size 3.1 Definition of effect size 3.2 d as a measure of effect size 3.3 Estimating effect size
4	Standards for rejecting H_0 4.1 Setting standards to make h_0 easier to reject
5	δ (delta) 5.1 Definition of δ (delta)
6	Increasing the power of a study 6.1 How to add power to your study 6.2 Ad hoc and post hoc power

Creating the graphics of the normal curves which appear in the animations in the tutorial, proved a slight challenge. For some reason, pasting the images directly from Microsoft Visio into Flash did not work, and Illustrator had to serve as a mediator between the two programs. Other graphics were created directly in Flash.

Phase 2. All templates were created directly in Macromedia Flash. These templates were developed using the Flash Slide Presentation format. This format appears as a hybrid between standard Flash movies and PowerPoint presentations. Each screen or page has its own timeline in Flash. Developers can create master and nested slides to cascade elements downward. Although this format is nearly equivalent to the features in standard Flash movies and it has the added benefit of the slide presentation schema, ActionScript functions differently in slide presentations. For example, the property "_root", which is a very common part of object instance names in standard Flash movies, is not required for most objects in the presentation. It became somewhat challenging to write some of the code for activities because of the unique needs of the script in the presentation mode. Documentation on the unique script needs for presentations is very limited.

Templates were developed to be very flexible for content, allowing for the possibility that content may change and so additional room was allocated to screens so that scroll bars would not be needed. After the templates were created and had successfully demonstrated their ability to pull content from the external files, a master slide file was created that each of the SWFs could be based on. Each time the developer created a new subsection, the master slide file was copied, renamed, and unneeded slides were deleted. This proved to be a very efficient way to access the templates.

Instruments Developed

The designer developed two instruments to evaluate the effectiveness of the tutorial: the *Statistical Power Test* and the *Power Tutorial Post Questionnaire*.

Statistical Power Test. In order to measure the effectiveness of the tutorial, the *Statistical Power Test* was developed. Sudweeks (2003) emphasized the need for creating a "detailed and explicit " plan for any test that includes information about the "primary purpose," "kinds of tasks to be tested," "content areas to be sampled and number of items allocated to each," "forms of test items to be used," and the difficulty of items (p. 3.5). Development began with a formal test plan (see Appendix D) that included a description of the purpose of the test, a statement of the decisions that would be made based on the test scores, an outline of the content of the test, a description of item formats, a list of the formulas that would be used for computation, and several sample items. This plan was quickly followed by a table of specifications for the test (see Appendix E). The table of specifications showed the number of items in the test and how they correlated with the learning objectives for the tutorial. With the tools of a test plan and a table of specifications in place, test items were developed to assess how well the learning objectives had been met.

The nature of the evaluation suggested that a pre/posttest design would best measure the effectiveness of the tutorial. Although a control group would have helped better control for variance in the study, the limited number of subjects available did not allow for a control group and treatment group. For this study, only a treatment group was used. Accordingly, two equivalent forms of a criterion-referenced test were created. After the test was created, Dr. Sudweeks reviewed it and made several suggestions for improving it.

The *Statistical Power Test* was developed as a criterion-referenced test. Linn and Gronlund (2000) describe a criterion-referenced test as one where the results are interpreted "in terms of a clearly defined and delimited domain of learning tasks" (p. 42). Because of this, items on the test were not eliminated based on failure to discriminate; items remained on the test if they were "directly relevant to the learning outcomes" (p.43). The test was also designed as an equivalent forms test in order to create pre/posttests. The designer worked with Dr. Sudweeks to align the test to the stated learning objectives. Each form of the test consisted of 12 items, and was worth 18 points. Each form contained five multiple-choice questions, four short-answer questions, two essay questions, and one binary response item. Most of the short-answer and essay questions were grouped together into context-dependent item sets.

During the first implementation of the tutorial, none of the students were able to complete the test in the time given. The designer worked with Dr. Richard Sudweeks to revise the test based on this problem. As a result, the test was revised and several items were removed or altered before the second implementation of the tutorial. These changes included separating one essay question into two so students would have a better idea of what information they were being asked to demonstrate, and removing several of the context-dependent items sets in order to shorten the test. Each form of the second draft of the test was made up of thirteen items and was worth eighteen points. Both forms included seven short answer questions, five multiple choice questions, and one binary response item. This version of the test contained one context-dependent item set.

Power Tutorial Post Questionnaire. The designer also developed the *Power Tutorial Post Questionnaire* to gather qualitative data about the tutorial. This questionnaire was

designed to gather students' perceptions of the tutorial and collect their suggestions for improving it. The final version consisted of eighteen questions, four of which gathered basic demographic data (see Appendix H).

Implementation

The power tutorial was implemented with three different groups: once in a section of IP&T 550 (Introduction to Quantitative Reasoning) and in two different sections of IP&T 651 (Quantitative Reasoning). The first implementation was a low-fidelity implementation designed to test whether the animations in the tutorial would be helpful. It was conducted in Dr. Sudweeks' IP&T 651 class during winter semester 2004 with 12 students. At that point, the tutorial was simply a Flash presentation meant to be used in a lecture. The main purpose of the implementation was to verify that the animations were helpful and to learn about the audience of the tutorial. The designer presented the lecture and afterwards the students filled out the *Statistics Background Questionnaire*, which was developed by the designer (see Appendix F). The information gathered in this questionnaire served as an important component of the audience analysis for the development of the final tutorial.

The second two implementations were both high-fidelity implementations of the entire tutorial and took place during fall semester 2004 and winter semester 2005 in Dr. Stephen Yanchar's IP&T 550 class and Dr. Richard Sudweeks' IP&T 651 class. Students were briefly introduced to the project and informed that participation was voluntary and would not have any affect on their grades in the courses.

Procedure with First Small Group

Seven students participated in the first implementation of the full power tutorial in IP&T 550. These students had all been admitted to the IP&T program of study and were seeking either a masters or doctoral degree. All students were taking the course for credit and were in their first semester of study in the program. There was one non-native English speaking student in this class.

During this implementation, students in IP&T 550 participated in the tutorial during their normal class time. Although not required, all students remained during the entire class time. This class already met in a room with computers for each student, so no special arrangements had to be made for them to access the Internet. The normal duration of IP&T 550 was one hour and forty minutes, and the entire time was taken up by the tutorial and evaluations.

After a brief introduction, each student completed a form of the pretest. The designer randomly distributed the two forms of the pretest to each class. After the students completed the test, they all started the tutorial. The designer walked around the lab taking observational notes.

With 15 minutes left in the class period, all students began the posttest. When they had completed the posttest, each participant completed a post questionnaire. Students were not allowed to access the tutorial during the posttest, but were allowed to access it during the questionnaire. The designer also conducted several one-on-one interviews with students who used the tutorial within a few days of their using it.

Procedure with Second Small Group

Nine students participated in the second implementation of the full power tutorial in IP&T 651. Seven of these students were pursuing degrees in the IP&T program and two were pursuing doctoral degrees in other programs in the college of education. All of the students had had at least one statistics class previous to this one, and four students reported having several statistics classes previous to IP&T 651. There were several non-native English speaking students in this class.

Since it did not normally meet in a room with a computer for each student, IP&T 651 met in the Macintosh Lab in the McKay building the day of the implementation to access the tutorial. This class normally met for its normal period of 50 minutes, which was entirely devoted to the tutorial and evaluation.

After a brief introduction, each student completed a form of the pretest. The designer randomly distributed one of the two forms of the pretest to each student. After the students completed the test, they all started the tutorial. The designer walked around the lab taking observational notes. Because the tutorial had not been previously tested on a Macintosh, there were a few problems with the table of contents working properly. As a result, the designer had to offer assistance to several people who had a challenge with the navigation.

With 15 minutes left in the class period, all students began the posttest. When they had completed the posttest, each participant completed a post questionnaire. Students were not allowed to access the tutorial during the posttest, but were allowed to access it during the questionnaire. The designer also conducted several one-on-one interviews with students who used the tutorial within a few days of their using it.

There were several reasons for multiple implementations of this tutorial. First, each class had less than ten students in it, and a somewhat larger group was desired. The second reason was that there were several features in the tutorial that were not working during the first implementation, and the designer wanted to try these features out on students.

Chapter 4: Results

Results of First Small-Group Implementation

Quantitative Results

Quantitative data were collected during this implementation through the pretests and posttests. All participants completed these tests as part of the attempt to determine whether learning objectives were met.

In the first implementation, all seven students completed the pretest and posttest. The mean score for the pretest of this group was 10.6 out of 18 possible points, and the standard deviation was 2.2. The mean score of the posttest was 13.4 with a standard deviation of 1.4. The average gain from pretest to posttest was 2.8 points. A paired-samples t test showed that the means of the pretest and the posttest were significantly different $t(6) = 3.447, p = .014$.

Results of Interviews

After the first implementation of the tutorial, the designer interviewed several of the students in a one-on-one setting to find out what they thought. The most common responses focused on a need for a "Table of Contents." All of the students interviewed in this setting expressed a desire for the ability to jump back to previous sections and noted a sense of annoyance because they did not know how long the sections were and how long the tutorial would keep going on. They felt like they had no concept of the structure of the tutorial. Since the table of contents had already been designed, the designer showed them the feature. All agreed that they thought it would adequately solve the problem they experienced.

Results of Second Small-Group Implementation

Quantitative Results

Quantitative data were also collected during this implementation through the pretests and posttests. All participants completed these tests as part of the attempt to determine whether learning objectives were met.

In this implementation, nine students completed the pretest and posttest. The mean score of the pretest was 10.9 out of 18 possible points ($SD = 3.1$). The mean score of the posttest was 12.1 ($SD = 4.8$). The average gain from pretest to posttest with this group was 1.2 points. A paired-samples t test suggested that the means of these two groups were not significantly different $t(8) = 1.537, p = .163$.

Results of Interviews

After the second implementation, the designer interviewed several students in the class in a one-on-one setting. The most common complaint was the lack of time. One of the students mentioned that she had not been able to access all of the sections of the tutorial, and felt like she missed one of the questions on the posttest as a result. The biggest recommendation to come from these students was to make the tutorial available to students to study outside of class and not try to squeeze it into a class period. All agreed that they would have enjoyed taking their time to go through the tutorial instead of having to rush through it.

Results from Both Classes

Qualitative Results

The designer collected qualitative data about the tutorial through the *Power Tutorial Post Questionnaire*, through observational notes, and through several one-on-one interviews with users.

Students in both implementations felt like information in the tutorial was presented in a clear way. They felt like they better understood power and how to plan a statistical test with adequate power after completing the tutorial. Students strongly felt that the tutorial had helped them understand the importance of power. They also liked all of the practice exercises, and several students mentioned that they would have liked even more practice for certain sections. Students felt like the interactive diagrams had helped them understand the concepts better, and that the "Things to Remember" section was helpful.

Students in both groups agreed that they would enjoy using this type of tutorial more often as they study statistics. While the majority of students in both groups further agreed that they would prefer this type of tutorial over a textbook to learn statistics, there were still several students in each group that would prefer to learn from a textbook.

Students in the first implementation found it easy to navigate through the tutorial. However, during that implementation the table of contents was not yet in place, and many of the students commented that they really wanted some indication of where they were at in the overall tutorial. As mentioned in the "Procedure" section, there were several problems with the table of contents on Macintosh computers during the second implementation. As a result, most students found navigation through the tutorial to be very challenging. Some of the students seemed to have problems navigating outside of

the known bug; it seemed that the design of the table of contents was confusing to them. Clicking on the name of the section expanded it to reveal the subsections, which could be clicked on to jump to the subsection of the tutorial. Several students did not expect this hierarchal navigation, and expected content to appear when they clicked on the section titles. These students repeatedly made this mistake. It became apparent that the two different kinds of links needed to look differently in the table of contents.

Unfortunately several students indicated that they did not see the interactive diagrams or know about the "Things to Remember" section. Through comments written on the questionnaires, it appears that the students who did not see the diagrams did not realize that they were interactive. This was very disappointing, as this was one of major instructional features of this tutorial. It was apparent that more care should have been taken to clearly distinguish between the many static diagrams and the few interactive diagrams. It is understandable that students would overlook them as they are not a common feature in statistics classes and so students were not expecting to see them. Other students, however, listed these diagrams as their favorite feature in the tutorial.

When asked about what they liked most about the tutorial, students most frequently responded that they liked the practice exercises. They justified this response by saying that it was because of the immediate chance to practice what was being learned and because the feedback was useful. This was very gratifying to hear, as considerable time was expended writing feedback for each option in practice questions. The next most frequent comments about favorite features included clear and concise content, real-world examples, and diagrams (both static and interactive). Other comments included

appreciation of the organization of the tutorial, the division of content into manageable chunks, self-pacing, thoroughness, and the look of the interface.

Students reported that their least favorite parts of the tutorial included unclear content, navigation problems, and lack of time to complete the tutorial. The students who reported that the content was unclear were all non-native English speakers. In truth, no special efforts were made to make the content more comprehensible for them. Profiling the vocabulary in the script and revising it to use the 2,000 most common words in English 95% of the time would have probably taken care of this problem. These same students all asked for the link to the tutorial to be emailed to them so they could study it more later. It appeared that they had to go through the tutorial more quickly than they would have liked. The second and third most common complaints came mostly from the IP&T 651 students. They mentioned that the navigation problem previously mentioned made the tutorial difficult for them, and several students regretted not having enough time to go through the entire tutorial or having to rush through it.

Quantitative Results

The fact that the first group experienced significant gains while the other did not is an interesting one. Although the second group had more experience with statistics, both groups had similar mean scores on the pretest. The standard deviation of the pretest scores for the two groups were slightly different, the second group having a higher standard deviation. On the posttest, the first group performed better on average than the second group. Interestingly, the standard deviation of the first group decreased on the posttest, and increased for the second group. These results are shown in Table 2.

Table 2

Pretest and posttest results by class

Class	<i>n</i>	Pretest		Posttest		Mean gain	<i>t</i>	<i>p</i>
		Mean	<i>SD</i>	Mean	<i>SD</i>			
IP&T 550	7	10.6	2.2	13.4	1.4	2.8	3.447	.014
IP&T 651	9	10.9	3.1	12.1	4.8	1.2	1.537	.163

The tutorial seemed to have each student in the first group increase their knowledge of power in about the same way. In the second group several of the students did very well on the posttest (several received a perfect score), but several students did not change their scores at all between the two tests. It seems that the students in the second group that came in with a good understanding of power saw the greatest growth in their knowledge, and students with the least amount of knowledge did not get as much out of the tutorial. One reason for this may have been the lack of time in completing the tutorial. Students in the second group had less time to complete the tutorial, and many of them did not complete it. Several students in this group seemed worried about time as they went through the tutorial and may have rushed. This group also included several non-native English speaking students who were not able to get through the tutorial as quickly as the native English speakers.

In summary, the tutorial does seem to be helpful to students and they enjoyed using it. They would like to use more tutorials of this type to learn statistics. It appears that self-pacing and adequate time were important factors for students. Many students in the second group complained that they did not have long enough to go through the tutorial and felt rushed. The difference between the two groups also came out in the post questionnaire. Students in the first group all said that they understood power better after

the tutorial, and answers to that question were mixed in the second group. Although the results concerning the effectiveness of the tutorial are mixed, it is difficult to make conclusions about the effectiveness of the tutorial based on the numbers alone. The groups in both implementations were small and statistical power was low. More research would need to be done to determine the true value of the tutorial. Future studies with this tutorial could include looking at the effect of self-pacing on learning outcomes as that seemed to be a confounding factor. A study comparing whether introducing the interactive diagrams earlier in the instruction or later in the instruction would also be very interesting.

Chapter 5: Discussion and Recommendations

Product Strengths

This tutorial had several strengths which included interactive diagrams, dynamic content, and the general instructional techniques used.

Students really enjoyed using the interactive diagrams. Even when they did not realize that the diagrams were interactive, they suggested that making them interactive would improve the tutorial. Students mentioned that they appreciated having so many graphics in general in the tutorial. This appears to support Bruner's theory that novice learners appreciate interaction and iconic representations (pictures) early on in learning as opposed to a simple text or audio description of a concept (Driscoll, 1994).

From a development point of view, the dynamic setup of the tutorial was a great strength. It facilitated content development and allowed for quick changes to the content. It will also be maintainable by someone who does not understand or have access to Macromedia Flash in the future. If the designer could redo the tutorial, he would use XML instead of simple text files as it is becoming a strong standard and would have increased the reusability of the content.

Students also appeared to appreciate the general instructional strategies that were used. The simple language of the tutorial was rated highly by students. Students commented that the practice exercises were their favorite part of the tutorial. Great care was taken to create a good balance of explanation and practice, with practice exercises coming up every few screens. Students requested even more practice in some areas, specifically when several skills were needed to solve the problems.

Product Weaknesses

The greatest weaknesses of the product were navigation issues, the amount of text to get through, and the placement of the interactive graphics.

In both implementations there were navigation issues. In the first, the table of contents was not present, and in the second implementation, it was not working correctly all of the time. Even when it did work, it confused several students as links to expand a section so subsection links were visible and the subsection links that actually opened lesson content looked the same. A design that included a small graphic in front of the section links that suggested expansion (i.e., ► vs. ▼ or + vs. – are commonly used for this purpose) or bringing up a section menu when the section headings were clicked would probably increase the usability of the product.

The limited time available to complete the tutorial seemed to have a negative impact on this study. Several students mentioned that the amount of text they had to get through was their least favorite part of the tutorial. Clark and Mayer (2003) suggested using audio narration as opposed to on-screen text worked more effectively in eLearning. This was a part of the original tutorial design, but in the end was left out due to development time required for the feature and the benefits of dynamic text. Clark, Nguyen, and Sweller, (2006) suggest that removing any unnecessary content from eLearning modules will improve learning outcomes. It would be interesting to analyze the content of the tutorial to see if there is anything that could be removed or simply put into an optional extra-mile section for students who want to know more about a particular topic.

Students seemed to really enjoy using the interactive graphics. These graphics were supposed to be the foundation of the tutorial, but somehow got pushed to the end. In an

attempt to provide adequate support for understanding them, the designer almost designed them out of the tutorial. For some students this was the case as they did not even realize the interactive diagrams were present. In retrospect, the diagrams should have been introduced very early in the tutorial and referred to more often.

Perhaps the greatest weakness, though, was in the design and development process itself. The designer failed to conduct one-on-one evaluations before the group implementations. Conducting these evaluations could have potentially eliminated all of the major design flaws from the tutorial before the tutorial was implemented on a larger scale. Rowe (2004) credited the success of his tutorial largely to these one-on-one evaluations. The designer believes that this formative evaluation could have completely eliminated the navigation problems and may have shown him that the interactive diagrams were not as self-explanatory as he had assumed.

Reflections

The designer completed the tutorial and implementations in 2005 and has worked as a full-time instructional designer and consultant since that time. Based on the experience of two years of learning about instructional design, the designer would make several changes in the tutorial. These changes would include a more intuitive navigation system, feedback to learners about time, contextualizing the learning more, and the technical structure of the tutorial.

There were problems with the navigation in both implementations of the tutorial. The main problem seems to have been the "Table of Contents," which was not present in one implementation and was only partially functional in the second. Getting this feature to work properly would have made the tutorial easier to navigate, but there are additional

things that could have been done to make the tutorial easier to navigate. For example, buttons could be added to the screen that allow learners to jump to any screen in the section. This could be helpful for several reasons. First, it gives learners a better sense of the length of each section. Second, if tool tips are added, learners can mouse over these buttons to get a sense of what the structure of the section will be like. Third, it allows learners to quickly jump back to screens that contain helpful information as they are answering practice questions. Figure 16 shows this possible enhancement to the tutorial.

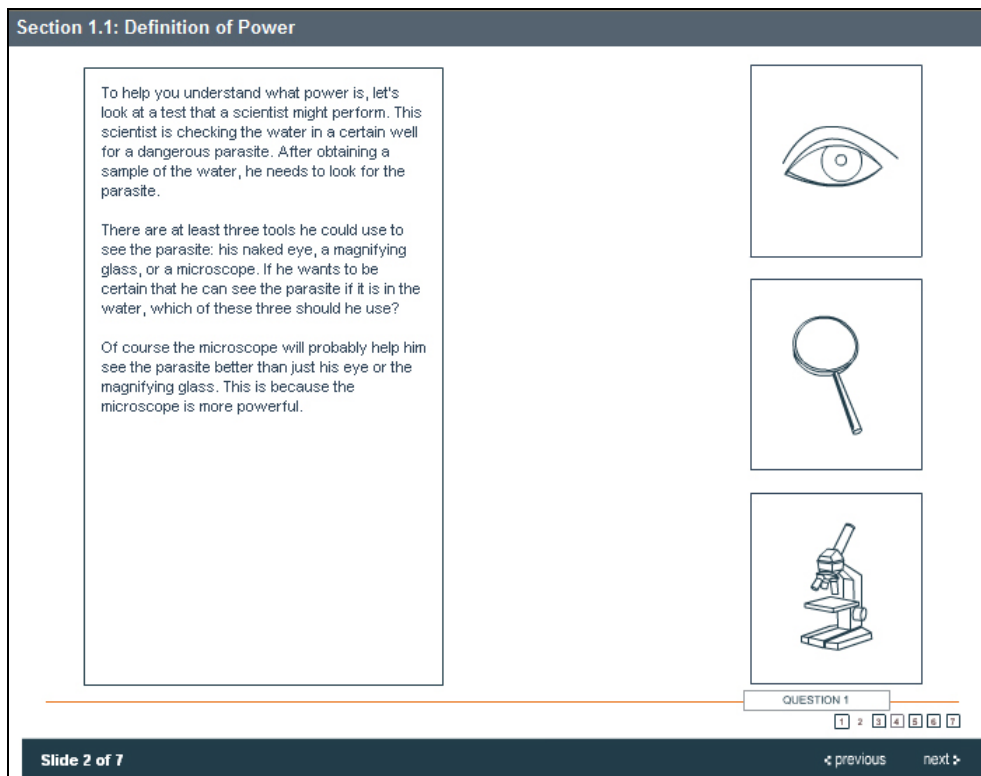


Figure 16. Screen shot of possible new interface for tutorial.

One of the most unfortunate aspects of the second implementation was the limited time that the learners had to complete the tutorial. The most basic way to solve this problem would have been to give learners more time to complete the tutorial. But including a note about the time that the entire tutorial takes to complete and notes about how long each section takes to complete are becoming a common addition to training modules. Learners like to have a sense about how long training will take. Simply adding a note to the title page of each section mentioning how long the section takes an average learner would be a simple way to add helpful feedback to the learner.

Another change that would have helped this tutorial would have been to contextualize it more. Adding a context to learning can make the overall experience more enjoyable and interesting, as well as adding cognitive benefits. It would be interesting to explore the idea of presenting the learners with a scenario in the first few screens of the tutorial where a researcher was dealing with a problem that required a knowledge of power. This scenario could have been structured in such a way that learners felt it was real and applicable to them. As they progressed through the tutorial, they would be relating what they were learning back to the scenario. One of the lesson objectives was to be able to plan a test with adequate power. Unfortunately, none of the learners met this objective. Making it the focus of a scenario from the beginning could potentially have made a big difference.

Finally, the designer would change many things about the technical set up of the tutorial. For example, each of the sub sections were separate files in this version of the tutorial. This made a large number of files to create and keep track of. It would probably have been better to reduce the number of files by collapsing each section into one file. It

could still have had the same number of divisions, and would look the same to the students, but it there would have been six SWF files in the end instead of nineteen. Further, the designer would have used XML structure for the external text files. XML is currently a very common format for external data. It is very flexible and can be used to populate software or to create documents. If content has the potential to be reused, then XML is an ideal choice as well, since it requires files to be set up in a way that makes them easier to reuse.

Schedule

Table 3 summarizes the projected and actual delivery dates and hours it took to complete the major deliverables for this project. A detailed project plan was created and managed in Microsoft Project. The software created weekly to-do lists for the project as well as creating and maintaining the project Gantt chart. Because the actual product was not scheduled to be implemented until fall semester 2004, the original dates began to slip once the spring semester ended and the designer started working full time for the summer. The estimates for hours to complete the project were very close for the most part, with the exception of the time it took to complete the tutorial. This discrepancy was due to the decision not to use audio narration throughout the tutorial. Using only text in the tutorial saved approximately 51 hours of development time.

Table 3

Estimated and actual schedule for power tutorial

Task	Projected delivery	Actual delivery	Estimated hours	Actual hours
Audience analysis	3/23/2004	3/23/2004	8	6
Instructional design	3/30/2004	3/30/2004	20	20
Scope and sequence	4/5/2004	3/24/2004	20	22
Script	4/12/2004	6/30/2004	20	21
Interface design	4/19/2004	6/25/2004	29	31
Complete tutorial	5/15/2004	7/15/2004	84	33
Test, review, and revise	5/26/2004	7/26/2007	28	28
Create assessment	6/2/2004	8/10/2004	17	14
Totals			226	175

Budget

The major expenses associated with this project were labor and materials. Of the two, the cost of labor represented the major portion of the budget. The planned cost of labor included the designer's production time, time spent by SME and instructional design experts reviewing the product and offering feedback, voice talent for audio recording, and an audio technician.

Materials for the project included three pieces of software: Microsoft Visio, Adobe Illustrator, and Macromedia Flash Professional. The designer was able to purchase each of these products in the BYU Bookstore for the education price, which greatly reduced the cost of materials. Copies were also a cost in the project. These included copies of the Post Tutorial Questionnaire, Statistical Power Test, Statistics Background Questionnaire, and job aides.

The cost of labor consisted in the work of the designer and the experts who offered assistance during the development process of the product, namely Dr. Richard Sudweeks, Dr. Paul Merrill, and Dr. Stephanie Allen. The original design of the product called for audio narration in the tutorial. Due to time constraints, this feature did not make it into the final product, and so the estimated hours of work from the voice talent and audio technician were not used. The lack of audio also reduced the number of hours it took the designer to produce the product.

Table 4 contains an itemized description of the estimated and actual costs for the development of this tutorial. The rate of the expert review was estimated based on information from the BYU Office of Research and Creative Activities (ORCA) since employee salary information is confidential. Table 5 summarizes the totals of estimated and actual costs for labor and materials for the project.

Table 4

Estimated and actual costs for the power tutorial

Expense	Rate	Estimated units	Actual units	Estimated cost	Actual cost
Labor					
Production	\$15/hr	226 hours	175 hours	\$3,390.00	\$2,625.00
Expert reviews	\$25/hr	32 hours	32 hours	\$800.00	\$800.00
Voice talent	\$15/hr	2 hours	0 hours	\$30.00	\$0.00
Audio technician	\$15/hr	4 hours	0 hours	\$60.00	\$0.00
Materials					
Microsoft Visio	\$75.00	1	1	\$75.00	\$75.00
Adobe Illustrator	\$100.00	1	1	\$100.00	\$100.00
Macromedia Flash	\$150.00	1	1	\$150.00	\$150.00
Copies and printouts	\$.05 ea.	300	350	\$15.00	\$17.50
Totals				\$4,620.00	\$3,767.50

Table 5

Summary of estimated and actual cost totals for the power tutorial

Expense	Estimated units	Actual units	Estimated cost	Actual cost
Labor	282 hours	332 hours	\$4,280.00	\$3,425.00
Materials			\$340.00	\$342.50

References

- Aberson, C. L., Berger, D. E., Healy, M. R., & Romero, V. L. (2002). An interactive tutorial for teaching statistical power. *Journal of Statistics Education, 10*(3).
- Adobe Systems Incorporated (2007). Retrieved July 16, 2007, from http://www.adobe.com/products/player_census/flashplayer/
- Anderson-Cook, C. M., & Dorai-Raj, S. (2003). Making the concepts of power and sample size relevant and accessible to students in introductory statistics courses using applets. *Journal of Statistics Education, 11*(3).
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. A., Pintrich, P. R., Raths, J., & Wittrock, M. C. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Clark, R. C. & Mayer, R. E. (2003). *E-learning and the science of instruction*. San Francisco: Jossey-Bass/Pfeiffer.
- Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence based guidelines to manage cognitive load*. San Francisco: Pfeiffer.
- Cohen, J. (1970). Approximate power and sample size determination for common one-sample and two-sample hypothesis tests. *Education and Psychological Measurement, 30*, 811–831.
- Cohen, J. (1988). *Power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
- Driscoll, M. P. (1994). *Psychology of learning for instruction*. Boston: Allyn and Bacon.

- Evans, S. D., Kennedy, M. M., & Ure, D. M. (2004). *A team-based approach to the design and development of a customer service training program for Desert Mutual Benefit administrators*. Unpublished master's project, Brigham Young University, Provo, Utah.
- Fernandez, G. C. & Liu, L. (1999). A technology-based teaching model that stimulates statistics learning. *Computers in the Schools*, 16(1), 173–191.
- Gatti, G. G. & Harwell, M. (1998). Advantages of computer programs over power charts for the estimation of power [Electronic version]. *Journal of Statistics Education*, 6(3).
- Howell, D. (2002). *Statistical Methods for Psychology* (5th ed.). Pacific Grove, California: Duxbury-Thomson Learning.
- Hinkle, D.E., Wiersma, W., & Jurs, S. (1998). *Applied statistics for the behavioral sciences* (4th ed.). Boston: Houghton Mifflin Company.
- Lipsey, M. W. & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: confirmation through meta-analysis. *American Psychologist*, 48, 1181–1209.
- Lohr, L. (1998). Using ADDIE to design a web-based training interface. In *SITE 98: Society for Information Technology & Teacher Education International Conference*. Washington DC. (ERIC Document Reproduction Service No. ED 421139)
- Merrill, P. F. & Bunderson, C. V. (1981). Preliminary guidelines for employing graphics in instruction. *Journal of Instructional Development*.4(4), 2–8.

- Murphy, K. R. & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. New Jersey: Lawrence Erlbaum Associates.
- Prentice, D. A. & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, 112(1), 160–164.
- Rowe, D. (2004). *Using graphics, animations, and data-driven animations to teach the principles of simple linear regression to graduate students*. Unpublished master's project, Brigham Young University, Provo, Utah.
- Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105(2), 309–316.
- Song, S. & Keller, J. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational Technology Research and Development*, 49(2), 5–22.
- Sprinthall, R. C. (2000). *Basic statistical analysis* (6th Ed.). Massachusetts: Allyn & Bacon.
- Sudweeks, R. (2003). *IP&T 652 course packet*. Unpublished course packet, Brigham Young University, Provo.
- Thomas, L. & Krebs, C. J. (1997). A review of statistical power analysis software. *Bulletin of Ecological Society of America*, 78(2), 126–139.
- West, R. W. & Ogden, R. T. (1998). Interactive demonstrations for statistics education on the World Wide Web [Electronic version]. *Journal of Statistics Education*, 6(3).
- Yu, C. H., & Behrens, J. T. (n.d.). *Identification of misconceptions in learning statistical power with dynamic graphics as a remedial tool*. Retrieved September 23, 2003 from <http://seamonkey.ed.asu.edu/~alex/pub/power/power.html>.

Appendix A

General Instructional Objectives (GIO) and Specific Learning Outcomes (SLO)

1. Understand what power is
 - a. Explain how power affects a statistical test
 - b. Identify the factors that influence the power of a test
 - c. Explain how each of these factors influence power
 - d. Classify examples of errors as a Type I or II
2. Understand why power is important to a statistical test
 - a. Given specific examples of experiments, determine which type of error will likely be more serious
 - b. Explain why experiments in the social and behavioral sciences often lack power
 - c. Explain the consequences of inadequate power for proposed studies or completed studies
3. Determine how much power a proposed or completed study has by using the appropriate statistics or resources (e.g., d or δ) to calculate the power of a test
4. Understand how to increase the power of a test
 - a. Analyze a test and determine why it has low power
 - b. Decide which of the factors that influence power should be manipulated to increase the power of a specific test
 - c. Design a test with adequate power

Appendix B

Scope and Sequence

Section 1: What is Power

Objectives

1. Understand what power is
 - a. Explain how power affects a statistical test
 - b. Infer the results of a test given its level of power or a factor(s) that influences power
 - c. Classify an error as a Type I or II
 - d. List the factors that influence the power of a test
2. Understand why power is important to a statistical test
 - a. Given specific examples of experiments, determine which type of error will be more serious
 - b. Explain why experiments in the social and behavioral sciences often lack power
 - c. Explain the consequences of inadequate power

Overview

Key terms: power, Type I error, Type II error, alpha, beta

Equations: $(1-\beta)$

Outline

- 1) Present
 - a) What is power
 - i) Definition of power
 - (1) Analogy of testing water for bacteria
 - (2) “The power of a statistical test is the probability that it will yield statistically significant results” (Cohen, 1988, p. 1)
 - (3) “[T]he probability of rejecting a false H_0 ” (Howell, 2002, p. 224)
- 2) Practice
 - a) Which of these is a definition of power? (give definitions that are just worded differently)
 - b) Which of these situations shows a concern about power?
 - i) Rational set generator (Driscoll, 1994, p.100–102)
 - c) What are the consequences of inadequate power?
- 3) Present & practice
 - a) Review normal curve
 - i) Mean

- ii) Alpha
 - iii) Curve itself (what does the high place mean)
 - iv) Standard deviation
- 4) Present
- a) Type I and II errors
 - i) Type I: Seeing an effect when there is not really one there.
 - (1) “When the null hypothesis is erroneously rejected” (Sprinthall, 2000, p. 242).
 - (2) “Rejecting H_0 when in fact it is true” (Howell, 2002, p. 105).
 - (3) Alpha is the probability of a Type I error
 - ii) Type II: Concluding that there is no effect when there is one there.
 - (1) “When the null hypothesis is accepted when it really should be rejected” (Sprinthall, 2000, p. 242).
 - (2) “Failing to reject H_0 when it is in fact false” (Howell, 2002, p. 105).
 - (3) Beta is the probability of a Type II error.
- 5) Practice
- a) Classify an error as type I or II in a given scenario
 - b) Decide which type of error would be more serious given a specific study
- 6) Present
- a) How power is measured
 - i) 1-beta
 - ii) Scale: 0.2 is low, 0.5 is medium, and 0.8+ is adequate
 - b) Problem in psychology and education
 - i) “On average, the Type-I error rate in psychology education is estimated to be at 50% or more” (Aberson & Berger, 2002, p. 2)
- 7) Practice
- a) What is the formula for power?
 - b) What is the power of this test?
 - c) Which test has more power?
- 8) Present
- a) There are three factors that influence the power of a study
 - i) Introduce
 - (1) Sampling error
 - (a) How accurate was your test? Did you really measure what you think you were measuring?
 - (b) How big was your sample size?
 - (i) What was the standard deviation?
 - (2) Effect size
 - (a) How much of a difference is there between the two means?
 - (3) The standards set to reject the null hypothesis
 - (a) How high is alpha? The bigger alpha is, the more power you have.
 - ii) Memory aid: jingle/chant
 - (1) Sampling error, effect size, and standards set to reject the null: they all control the power.
- 9) Practice
- a) What are the three things that influence power?

- b) Which of the three influences is this?

Section 2: How sampling error effects power

Objectives

1. Understand what power is
 - a. Explain why/how sampling error influences power

Overview

Key terms: sampling error, standard deviation, reliable (reliability), valid (validity)

Outline

- 1) Present
 - a) Sampling Error
 - i) Definition
 - ii) What contributes to sampling error?
 - (1) Sample size
 - (2) Instruments that are not adequate to measure the trait in question.
- 2) Practice
 - a) What is sampling error?
 - b) What causes sampling error?
- 3) Present
 - i) Reliable and valid instruments
 - (1) The problem of measurement error
 - (a) Example from Dr. Sudweeks test packet
 - (i) Virginia knows the answers to 46 questions on a test consisting of 60 multiple-choice items. She takes the test on two occasions. The first time she mistakenly marked the wrong answer, even though she knew the right answer. The second time she guessed on 11 items she wasn't sure of and got 4 right. We can never be sure what a person's true score is, so the obtained score is an estimate of the true score. How accurate is the estimate?
 - (2) Reliability is the degree of consistency in two or more measures of the same trait.
 - (a) Ways to increase the reliability of scores from a test:
 - (i) Increase the number of items in a test
 - (ii) Increase the variability of the scores (spread the scores out)
 - (iii) Make the test more difficult
 - (iv) Make sure the test is scored objectively
 - (3) Validity is the degree to which differences in the scores reflect differences in the degree to which subjects possess the trait being tested.

- (a) Ways to increase the validity of scores obtained from a test
 - (i) Explicitly define the trait that you are measuring
 - (ii) Carefully construct items that fit the definition of the trait
 - (iii) Only including items in the test that measure the trait
 - (iv) Taking steps to minimize the influence of extraneous factors on the scores (i.e., test wiseness, administration factors, etc.)
 - (v) Making sure there is a high correlation between each item and the total score of the test.
- 4) Practice
 - a) How do reliability and validity affect sampling error?
 - b) What is the difference between reliability and validity?
 - i) Rational set generator (Driscoll, 1994, p.100–102)
- 5) Present
 - i) How does sample size influence power?
 - (1) Review definition of standard deviation.
 - (2) When you have a large sample (n), you are less likely to choose individuals in a clump, or who lie in extremes.
 - (3) As sample size increases, the distribution becomes more normal and the variance decreases.
 - (a) Be prepared to show how a greater sample of the population will decrease variance.
 - ii) Show animation of variance decreasing.
- 6) Practice
 - a) How does sample size affect power?
 - b) Which of these tests has greater power (based on sample size)

Section 3: Influence of effect size on power

Objectives

- 2. Understand what power is
 - a. Explain why/how sampling error influences power

Overview

Key terms: effect size

Equations: formula for d ($d = \left(\frac{\mu_A - \mu_O}{\sigma} \right)$)

Outline

- 1) Present
 - a) Effect Size

- i) What is effect size?
 (1) Basically it is a measure of how much impact the treatment had on the trait being studied.
- 2) Practice
 (1) Which treatment had the strongest effect?

	Treatment 1	Treatment 2	Treatment 3
Control group:	Mean=70	Mean=70	Mean=70
Test group:	Mean=90	Mean=72	Mean=40

- 3) Present
 (1) d is a standardized measure of effect size.
 (a) Formula for d
 (i)
$$d = \left(\frac{\mu_A - \mu_O}{\sigma} \right)$$
 - This is the difference between the group means in terms of the standard deviation.
 - Example of a group with a large standard deviation and a small effect size. A d of .20 means that on average, each person's score is increased by 1/5 of a standard deviation when they receive the treatment.
- 4) Practice
 a) What is d?
 b) What does the formula for d mean?
 c) How can you increase effect size? (decrease standard deviation or increase difference between the two means)
- 5) Present
 a) How can you tell what size of effect you will have before you conduct the test?
 i) Small, medium, and large effect sizes
 (i) Effect size is of special concern to us as social scientists because our treatments typically have small effect sizes (which could explain why power is so low in social science).
 1. List examples from Lipsey & Wilson (1993)
- 6) Practice
 a) Classify an effect as small, medium, or large.
 b) Guess what the effect size of an experiment was
 c) What effect size can we usually expect?
 d) How can we estimate the effect size of an experiment?
- 7) Present
 a) Why is effect size so important?

- i) When we focus on effect size, we tend to look more at the practical effect of the treatment than whether or not we got significant results.

Section 4: Influence of standards for rejecting H_0

Objectives

1. Understand what power is
 - a. Explain why/how sampling error influences power

Overview

Key terms: alpha (review)

Outline

- 1) Setting standards to make H_0 easier to reject.
 - a) Change the alpha level.
 - i) If you increase alpha, you increase power, and if you decrease alpha, you decrease power.
 - ii) With an alpha of .05, you will get a type I error 5% of the time.
 - iii) “It is important to note that Type I errors can only occur when there is no true treatment effect (i.e., when the null hypothesis is true), which is rarely the case. As a result, Type I errors are probably quite rare, and the efforts to control these errors at the expenses of making more type II errors might be ill-advised” (Murphy, 1990).
 - iv) “If the null hypothesis is wrong, researchers cannot make a type I error” (Murphy & Myers, 1998, p. 80)

Section 5: Delta

Objectives

3. Determine how much power a test has
 - a. Use the appropriate statistics (e.g., d or delta) or resources (e.g., Excel, G*Power, etc.) to calculate the power of a test

Overview

Key terms: delta

Equations: delta

Outline

- 2) Present
 - a) The power statistic: δ
 - i) Calculating δ
 - (1) When we calculate δ , we should see the influence of all three factors in the formula.
 - ii) Estimating the effect size needed using delta
- 3) Practice
 - a) Calculate δ
 - b) Calculate sample size needed using delta

Section 6: Increasing the power of a study

Objectives

4. Understand how to increase the power of a test
 - a. Analyze a test and determine why it has low power
 - b. Decide which of the factors that influence power should be manipulated to increase the power of a specific test
 - c. Design a test with adequate power

Overview

Key terms: a priori power, post hoc power

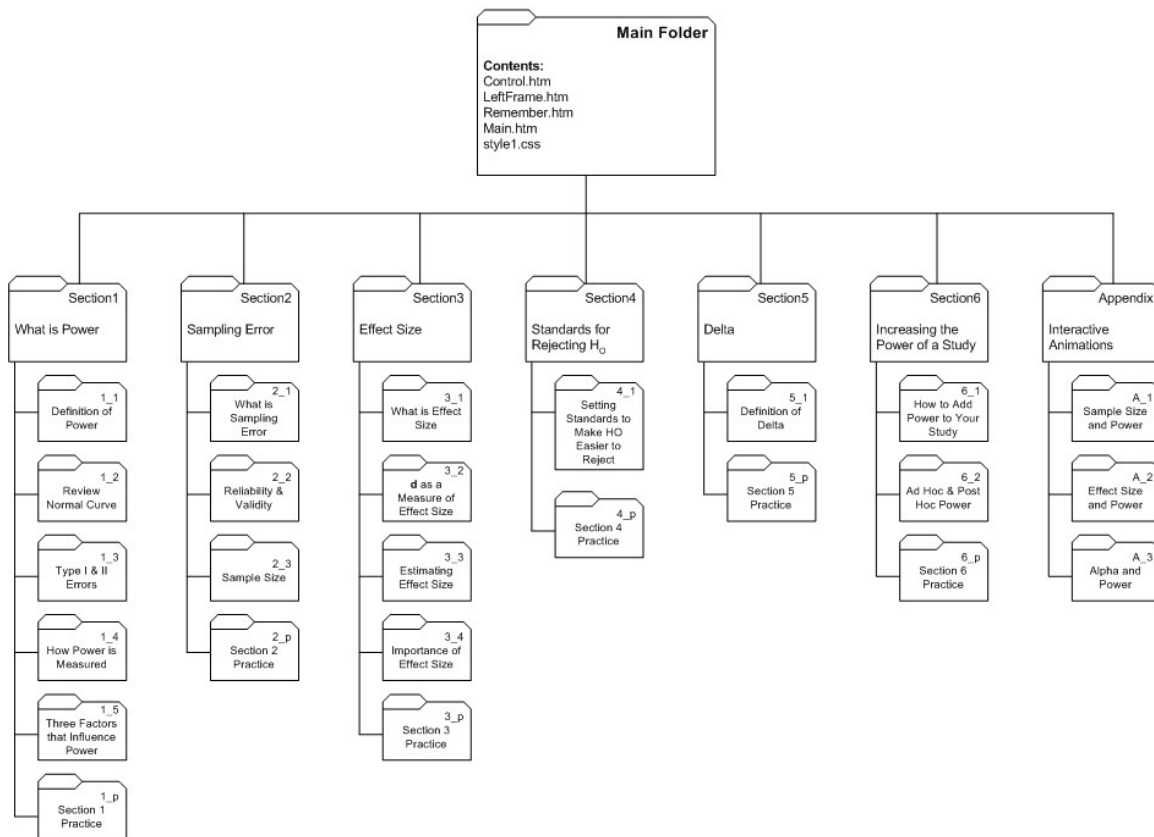
Outline

- 1) Present
 - a) What can you do to add power to your study?
 - i) Changing any of the three factors can increase the power of your study. However, the effect size is usually set, so the best way is to decrease the sampling error or by increasing alpha.
- 2) Practice
 - a) Given certain specifications, plan a study with adequate power.
- 3) Present
 - a) There are two time when you should calculate power
 - i) When planning a study
 - ii) When you are interpreting results of a completed study that failed to detect significant results
 - iii) Emily Dickinson poem
- 4) Practice
 - a) Given a study, can you determine whether the researcher did not get significant results because the H_0 was true, or because the study did not have adequate power?
 - b) What are the two times that you should analyze power

Appendix C

HTML File Structure

Power Lesson Folder Organization



Appendix D

Plan for Power Test

Purpose of the test

The purpose of the Power Assessment is to determine how effective the instruction was at teaching the concepts and principles of power. It will not, at least preliminarily, be used to assign grades.

Decisions that will be based on the scores

The scores and success or lack thereof on individual items will help me to revise the instruction. If students do poorly on a good item, I will revise the instruction of the objective that I based the item on.

Content of the test

All of the items in the test will be based on one or more of the instructional objectives below.

1. Understand what power is
 - a. Explain how power affects a statistical test
 - b. Identify the factors that influence the power of a test
 - c. Explain how each of these factors influence power
 - d. Classify examples of errors as a Type I or II
2. Understand why power is important to a statistical test
 - a. Given specific examples of experiments, determine which type of error will likely be more serious
 - b. Explain why experiments in the social and behavioral sciences often lack power
 - c. Explain the consequences of inadequate power for proposed studies or completed studies
3. Determine how much power a proposed or completed study has by using the appropriate statistics or resources (e.g., d or δ) to calculate the power of a test
4. Understand how to increase the power of a test
 - a. Analyze a test and determine why it has low power
 - b. Decide which of the factors that influence power should be manipulated to increase the power of a specific test
 - c. Design a test with adequate power

Item Formats

Remembering

1. Short answer
2. Multiple choice

Understanding

1. Multiple choice (using new information)
2. Binary response (type I/type II)
3. Essay questions

Applies

1. Short answer (solving problems and then giving the correct answer)

Analyze

1. Multiple choice

Evaluate

1. Essay

Create

1. Essay

The following formulas are used for computation

Power = 1-Beta

$$D = \left(\frac{\mu_A - \mu_O}{\sigma} \right)$$

$$\Delta = d\sqrt{(n/2)}$$

Sample Items

1. What is the formula for power? _____
2. When you solve the equation for power, what does it tell you?

3. Classify the following as examples of Type I or Type II errors.

<u>Type I</u>	<u>Type II</u>	
<input type="checkbox"/>	<input type="checkbox"/>	A doctor prescribes an antibiotic for a baby who does not really have an ear infection.
<input type="checkbox"/>	<input type="checkbox"/>	A man thinks that he has lost 10 pounds because he is standing on a broken scale.
<input type="checkbox"/>	<input type="checkbox"/>	A teacher thinks that her new way of teaching spelling helped increase test scores by 10%, but the scores really increased because the words were easier that week.
<input type="checkbox"/>	<input type="checkbox"/>	A man decides that his baby is crying because he is hungry when the baby really has a diaper rash.
<input type="checkbox"/>	<input type="checkbox"/>	A doctor decides that a woman that she has diabetes when she really doesn't.
<input type="checkbox"/>	<input type="checkbox"/>	A doctor decides that a new treatment for cancer does not work when it really does.
<input type="checkbox"/>	<input type="checkbox"/>	A missionary decides that his mission was a waste of time because none of his investigators were baptized. After he leaves, many of them join the Church.
<input type="checkbox"/>	<input type="checkbox"/>	A health clinic decides that smoking cessation classes do not help people stop smoking when, in fact, they really do help.
<input type="checkbox"/>	<input type="checkbox"/>	A French teacher decides that her new way of teaching vocabulary is not effective when it really is.

Appendix E

Table of Specifications for Power Test

Table 6

Distribution of number of questions by content area and cognitive process dimension of Bloom's revised taxonomy of educational objectives

Instructional Objectives							
Content Area	Remembers	Understands	Applies	Analyzes	Evaluates	Creates	Percentage
Effect of power		5					0.17
Type I Error		1			1		0.07
Type II Error		1			1		0.07
Sensitivity of the study	1	1		1			0.10
Effect Size	1	1		1			0.10
Standards for rejecting Ho	1	1		1			0.10
Power of a test			5	4		3	0.40
<i>Total</i>	3	10	5	7	2	3	1.00
<i>Percentage</i>	10.0	33.3	16.7	23.3	6.7	10.0	100.0
						Predicted #:	30
						Difference:	0

Table 7

Distribution of tutorial objectives by content area and cognitive process dimension of Bloom's revised taxonomy of educational objectives

Instructional Objectives						
Content Area	Remembers	Understands	Applies	Analyzes	Evaluates	Creates
Effect of power		1a,1b,1c,2b,2c				
Type I Error		1d			2a	
Type II Error		1d			2a	
Sensitivity of the study	1b	1c		4a,4b		
Effect Size	1b	1c		4a,4b		
Standards for rejecting Ho	1b	1c		4a,4b		
Power of a test			3	4a,4b		4c

Appendix F

Statistics Background Questionnaire

Age: _____

1. Gender:

- a. Male
- b. Female

2. Program of study:

3. What subjects have been your favorites in school?

4. What subjects have been your least favorites in school?

5. How many statistics classes have you taken in the past 5 years (including any classes you are currently enrolled in)?

6. If you were to learn about something, which way would you prefer to learn?

- a. Listen to a lecture or presentation.
- b. Read a booklet or pamphlet.
- c. Use the Internet.
- d. Use a computer program.

7. What is the highest level of education you have completed?

- a. High school.
- b. Trade school.
- c. Some college.
- d. Associate's degree.
- e. Bachelor's degree.
- f. Master's degree.
- g. Doctorate.

8. How often do you use a computer?

- 5 – Everyday
- 4
- 3
- 2

1 – Never

9. How interested are you in statistics?

- 6 – Very interested
- 5
- 4
- 3
- 2
- 1 – Not even slightly interested

10. How likely is it that you will take another statistics class?

- 6 – Very likely
- 5
- 4
- 3
- 2
- 1 – Not likely at all

11. Why have you taken statistics classes

- a. They were required for my program
- b. I want to understand statistics better in general
- c. I want to conduct my own research
- d. I need to learn statistics for a school-related project

12. Do you have access to the Internet?

- a. Yes
- b. No
- c. Unsure

13. What type of Internet connection do you use most frequently when doing schoolwork?

- a. Dial-up
- b. High-speed

14. Which of the following programs are you familiar with? (check all that apply)

- Excel
- SPSS
- MiniTab
- SAS
- PowerPoint
- Flash
- Web browsers (e.g., Internet Explorer, Netscape Navigator, Safari, etc.)

15. Does the computer that you use most frequently for schoolwork have a Flash player installed on it?
- Yes
 - No
 - Not sure
16. If you had to install a Flash player on your computer, would you know what to do?
- Yes
 - No
 - Not sure
17. If you had to classify yourself as a slow, average, or fast learner in statistics, which would you be in general?
- Slow
 - Average
 - Fast
18. Would you rather use a computer tutorial to learn statistics than read a textbook?
- Yes
 - No
 - Not sure
19. How motivated do you feel to learn statistics?
- 6 – Very motivated
5
4
3
2
1 – Not motivated at all
20. What is the most challenging thing about learning statistics?
- _____
- _____
- _____
- _____
21. What would make statistics easier for you to learn?
- _____
- _____
- _____
- _____
22. What would make you feel more motivated to learn statistics?
- _____
- _____
- _____
- _____
23. How interested are you in learning statistics using a computer tutorial?
- 6 – Very interested
5
4
3
2
1 – Not interested at all
24. Which of the following studies would be most interesting to you?
- Vocabulary development of third graders in their native language
 - A new methodology for teaching grammar to second-language learners
 - The effects of preparation on GRE scores
 - A new type of chemotherapy used on cancer

Appendix G

Statistical Power Test

1. What are the three things that affect the power of a statistical test? Briefly explain how each affects power. (6 points)

2. Which of the following statements about power is true? (1 point)
 - a. Power is most important for tests where the null hypothesis is true
 - b. The less power a test has, the more likely it is that the researcher will be able to reject the null hypothesis
 - c. The more power a test has, the more likely it is that the researcher will be able to reject the null hypothesis

3. Research has shown that experiments in the social and behavioral sciences often lack power. Why is this? (1 point)
 - a. Most researchers in the field design experiments that have too much variance
 - b. Most researchers in the field have hypothesis that are not true
 - c. Most of the treatments do not change the scores very much

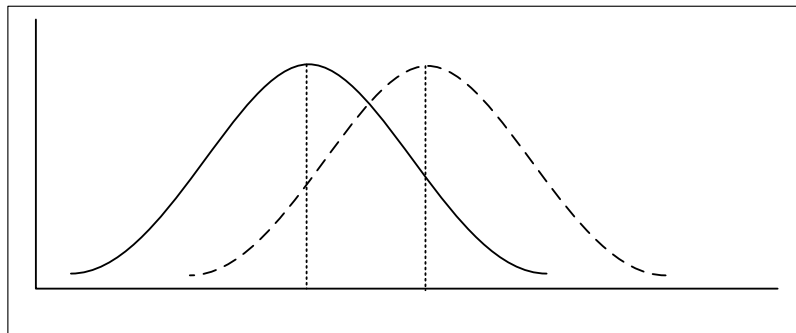
4. If a study does not have enough power, what will the most direct result be? (1 point)
 - a. The researcher will not be able to reject the null hypothesis if it is false
 - b. The researcher will not be able to generalize the results of the study
 - c. The results of the test that the researcher used will be invalid

5. A certain test has a p value of 0.01. If a researcher wanted to increase the power of her test, what should she change this value to? (1 point)

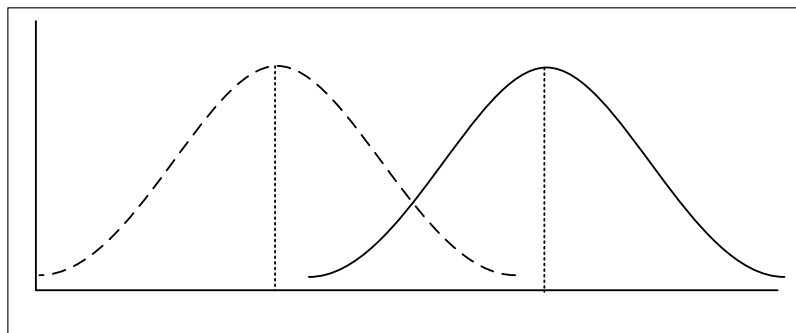
6. Dr. Sprocter is conducting a study to determine how well students need to understand a subject before they are effective peer tutors. He hypothesizes that they do not need to understand the subject much better than the person they are tutoring. For this study, the students will tutor each other in subjects that usually do not show that peer tutoring helps. Which of the following things will help him reach a correct conclusion about his hypothesis? (1 point)
 - a. Choosing a subject where peer tutoring has been shown to be very useful
 - b. Decreasing the value of p from 0.05 to 0.01
 - c. Decreasing the number of participants so that he won't have as much variance

7. Look at the following sets of curves. These represent different experiments in which the test group was given some kind of treatment and then tested to see whether this treatment had an effect. The curve with the solid line represents the control group's distribution of scores, and the curve with the dashed line represents the treatment group's distribution of scores. Which experiment had the most power? (1 point)

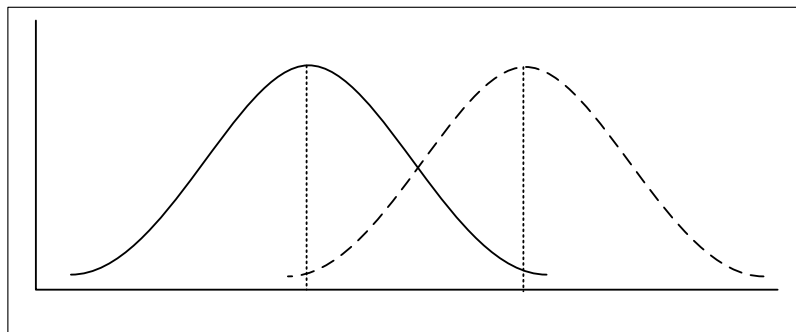
a.



b.



c.



8. Which of the following experiments has the most power? (1 point)

a. $n=20$

b. $n=40$

9. Classify each of the following as a Type I error, a Type II error, or no error by placing a check mark in the appropriate column. (2 points)

Type I	Type II	No error	
			A spelling teacher uses a new technique for teaching 5 th graders how to spell the days of the week. This new technique helps the students spell better, and they do well on her test. She concludes that the new way is better.
			A Spanish instructor teaches his students a mnemonic device for remembering the conjugations of several irregular verbs. This mnemonic does not really help, so the students study harder for the test. After testing them, the teacher concludes that his mnemonic device works.
			An English professor creates a packet for her students that leads them through prewriting steps. These handouts help the students organize their thoughts, and as a result the students write better. They all receive higher scores on their term papers than they would have. The professor decides that her new handouts are effective.

For questions 10-12, use the information in the box below.

In 1977, Hartley conducted a study on the effectiveness of individualized instruction in math for elementary and secondary students. Instead of teaching a large class of 20-30 students, teachers taught smaller classes, and focused on spending at least 20 minutes every two days with each student. Their scores in math were compared to the scores of students in traditional math classes after a month of using this technique. Hartley found that the effect size was 0.50, and 153 children participated in the study.

10. Assuming that Hartley used the standard 0.05, how much power did this study have?
11. How much of a chance did Hartley have of making a Type II error?
12. Assuming that the effect size and p-value do not change, how many participants would Hartley have needed to get a power of 0.80 or above?

Appendix H

Power Tutorial Post Questionnaire

Please rate each of the following statements on a scale of how much you agree or disagree with them. For all items the following key will be used

SA: Strongly Agree
a: agree
d: disagree
SD: Strongly Disagree

1. The information in this tutorial was presented clearly.	SD	d	a	SA
2. It was easy for me to navigate through the tutorial.	SD	d	a	SA
3. There were too many practice questions in the tutorial.	SD	d	a	SA
4. The “Things to Remember” section was helpful to me.	SD	d	a	SA
5. The interactive diagrams helped me understand the concepts better.	SD	d	a	SA
6. I feel like I understand power much better now.	SD	d	a	SA
7. This tutorial helped me realize the importance of power.	SD	d	a	SA
8. I understand how to plan a statistical test with adequate power now.	SD	d	a	SA
9. I would enjoy using this type of tutorial more often as I study statistics.	SD	d	a	SA
10. I would rather read the book than use this type of tutorial.	SD	d	a	SA
11. The person who wrote the content for this tutorial understood the subject well.	SD	d	a	SA

What did you like most about this tutorial?

What did you like least about this tutorial?

What would you change about this tutorial?