1

INTRODUCTION

Many students and researchers are intimidated by statistical procedures. This may in part be due to a fear of math, problematic math teachers in earlier education, or the lack of exposure to a "discovery" method for understanding difficult procedures. Readers of this book should realize that they have the ability to succeed in understanding statistical processes.

APPROACH OF THE BOOK

This is an introduction to statistics using EXCEL® and SPSS® to make it more understandable. Ordinarily, the first course leads the student through the worlds of descriptive and inferential statistics by highlighting the formulas and sequential procedures that lead to statistical decision making. We will do all this in this book, but I place a good deal more attention on conceptual understanding. Thus, rather than memorizing a specific formula and using it in a specific way to solve a problem, I want to make sure the student first understands the nature of the problem, why a specific formula is needed, and how it will result in the appropriate information for decision making.

By using statistical software, we can place more attention on understanding how to *interpret findings*. Statistics courses taught in mathematics departments, and in some social science departments, often place primary emphases on the formulas/processes themselves. In the extreme, this can limit the usefulness of the analyses to the practitioner. My approach encourages students to focus more on how to understand and make applications of the results of statistical analyses. EXCEL®

Understanding Educational Statistics Using Microsoft Excel® and SPSS®. By Martin Lee Abbott.

and other statistical programs are much more efficient at performing the analyses; the key issue in my approach is how to interpret the results in the context of the research question.

Beginning with my first undergraduate course through teaching statistics with conventional textbooks, I have spent countless hours demonstrating how to conduct statistical tests by hand and teaching students to do likewise. This is not always a bad strategy; performing the analysis by hand can lead the student to understand how formulas treat data and yield valuable information. However, it is often the case that the student gravitates to memorizing the formula or the steps in an analysis. Again, there is nothing wrong with this approach as long as the student does not stop there. The outcome of the analysis is more important than memorizing the steps to the outcome. Examining the appropriate output derived from statistical software shifts the attention from the nuances of a formula to the wealth of information obtained by using it.

It is important to understand that I do indeed teach the student the nuances of formulas, understanding why, when, how, and under what conditions they are used. But in my experience, forcing the student to scrutinize statistical output files accomplishes this and teaches them the appropriate use and limitations of the information derived.

Students in my classes are always surprised (ecstatic) to realize they can use their textbooks, notes, and so on, on my exams. But they quickly find that, unless they really understand the principles and how they are applied and interpreted, an open book is not going to help them. Over time, they come to realize that the analyses and the outcomes of statistical procedures are simply the ingredients for what comes next: building solutions to research problems. Therefore, their role is more detective and constructor than number juggler.

This approach mirrors the recent national and international debate about math pedagogy. In my recent book, *Winning the Math Wars* (Abbott et al., 2010), my colleagues and I addressed these issues in great detail, suggesting that, while traditional ways of teaching math are useful and important, the emphases of reform approaches are not to be dismissed. Understanding and memorizing detail are crucial, but problem solving requires a different approach to learning.

PROJECT LABS

Labs are a very important part of this course since they allow students to take charge of their learning. This is the "discovery learning" element I mentioned above. Understanding a statistical procedure in the confines of a classroom is necessary and helpful. However, learning that lasts is best accomplished by students directly engaging the processes with actual data and observing what patterns emerge in the findings that can be applied to real research problems.

In this course, we will have several occasions to complete Project Labs that pose research problems on actual data. Students take what they learn from the book material and conduct a statistical investigation using EXCEL® and SPSS®. Then, they have the opportunity to examine the results, write research summaries, and compare findings with the solutions presented at the end of the book.

These are labs not using data created for classroom use but instead using real-world data from actual research databases. Not only does this engage students in the learning process with specific statistical processes, but it presents real-world information in all its “grittiness.” Researchers know that they will discover knotty problems and unusual, sometimes idiosyncratic, information in their data. If students are not exposed to this real-world aspect of research, it will be confusing when they engage in actual research beyond the confines of the classroom.

The project labs also introduce students to two software approaches for solving statistical problems. These are quite different in many regards, as we will see in the following chapters. EXCEL® is widely accessible and provides a wealth of information to researchers about many statistical processes they encounter in actual research. SPSS® provides additional, advanced procedures that educational researchers utilize for more complex and extensive research questions. The project labs provide solutions in both formats so the student can learn the capabilities and approaches of each.

REAL-WORLD DATA

As I mentioned, I focus on using real-world data for many reasons. One reason is that students need to be grounded in approaches they can use with “gritty” data. I want to make sure that students leave the classroom prepared for encountering the little nuances that characterize every research project.

Another reason I use real-world data is to familiarize students with contemporary research questions in education. Classroom data often are contrived to make a certain point or show a specific procedure, which are both helpful. But I believe that it is important to draw the focus away from the procedure per se and understand how the procedure will help the researcher resolve a research question. The research questions are important. Policy reflects the available information on a research topic, to some extent, so it is important for students to be able to generate that information as well as to understand it. This is an “active” rather than “passive” learning approach to understanding statistics.

RESEARCH DESIGN

People who write statistics books have a dilemma with respect to research design. Typically, statistics and research design are taught separately in order for students to understand each in greater depth. The difficulty with this approach is that the student is left on their own to synthesize the information; this is often not done successfully.

Colleges and universities attempt to manage this problem differently. Some require statistics as a prerequisite for a research design course, or vice versa. Others attempt to synthesize the information into one course, which is difficult to do given the eventual complexity of both sets of information. Adding somewhat to the problem is the approach of multiple courses in both domains.

I do not offer a perfect solution to this dilemma. My approach focuses on an in-depth understanding of statistical procedures for actual research problems. What this means is that I cannot devote a great deal of attention in this book to research design apart from the statistical procedures that are an integral part of it. However, I try to address the problem in two ways.

First, wherever possible, I connect statistics with specific research designs. This provides an additional context in which students can focus on using statistics to answer research questions. The research question drives the decision about which statistical procedures to use; it also calls for discussion of appropriate design in which to use the statistical procedures. We will cover essential information about research design in order to show how these might be used.

Second, I am making available an online course in research design as part of this book. In addition to databases and other research resources, you can follow the web address in the Preface to gain access to the online course that you can take in tandem with reading this book or separately.

"PRACTICAL SIGNIFICANCE"—IMPLICATIONS OF FINDINGS

I emphasize "practical significance" (effect size) in this book as well as statistical significance. In many ways, this is a more comprehensive approach to uncertainty, since effect size is a measure of "impact" in the research evaluation. It is important to measure the likelihood of chance findings (statistical significance), but the extent of influence represented in the analyses affords the researcher another vantage point to determine the relationship among the research variables.

I call attention to problem solving as the important part of statistical analysis. It is tempting for students to focus so much on using statistical procedures to create meaningful results (a critical matter!) that they do not take the next steps in research. They stop after they use a formula and decide whether or not a finding is statistically significant. I strongly encourage students to think about the findings in the context and words of the research question. This is not an easy thing to do because the meaning of the results is not always cut and dried. It requires students to think beyond the formula.

Statisticians and practitioners have devised rules to help researchers with this dilemma by creating criteria for decision making. For example, squaring a correlation yields the "coefficient of determination," which represents the amount of variance in one variable that is accounted for by the other variable. But the next question is, How much of the "accounted for variance" is meaningful?

Statisticians have suggested different ways of helping with this question. One such set of criteria determines that 0.01 (or 1% of the variance accounted for) is considered "small" while 0.05 (5% of variance) is "medium," and so forth. (And, much to the dismay of many students, there are more than one set of these criteria.)

But the material point is that these criteria do not apply equally to every research question.

If a research question is, "Does class size affect math achievement," for example, and the results suggest that class size accounts for 1% of the variance in math achievement, many researchers might agree it is a small and perhaps even inconsequential impact. However, if a research question is, "Does drug X account for 1% of the variance in AIDS survival rates," researchers might consider this to be much more consequential than "small"!

This is not to say that math achievement is any less important than AIDS survival rates (although that is another of those debatable questions researchers face), but the researcher must consider a range of factors in determining meaningfulness: the intractability of the research problem, the discovery of new dimensions of the research focus, whether or not the findings represent life and death, and so on.

I have found that students have the most difficult time with these matters. Using a formula to create numerical results is often much preferable to understanding what the results mean in the context of the research question. Students have been conditioned to stop after they get the right numerical answer. They typically do not get to the difficult work of what the right answer *means* because it isn't always apparent.

COVERAGE OF STATISTICAL PROCEDURES

The statistical applications we will discuss in this book are "workhorses." This is an introductory treatment, so we need to spend time discussing the nature of statistics and basic procedures that allow you to use more sophisticated procedures. We will not be able to examine advanced procedures in much detail. I will provide some references for students who wish to continue their learning in these areas. It is hoped that, as you learn the capability of EXCEL® and SPSS®, you can explore more advanced procedures on your own, beyond the end of our discussions.

Some readers may have taken statistics coursework previously. If so, my hope is that they are able to enrich what they previously learned and develop a more nuanced understanding of how to address problems in educational research through the use of EXCEL® and SPSS®. But whether readers are new to the study or experienced practitioners, my hope is that statistics becomes meaningful as a way of examining problems and debunking prevailing assumptions in the field of education.

Often, well-intentioned people can, through ignorance of appropriate processes promote ideas in education that may not be true. Furthermore, policies might be offered that would have a negative impact even though the policy was not based on sound statistical analyses. Statistics are tools that can be misused and influenced by the value perspective of the wielder. However, policies are often generated in the absence of compelling research. Students need to become "research literate" in order to recognize when statistical processes should be used and when they are being used incorrectly.