Statistics and Sociology

Computation and Application

This book is designed to provide you with a set of tools that can and should be readily *applied* to solve certain kinds of problems. Although there are several ways in which the subject of social statistics might be introduced, we have selected the approach that is the most *practical*. Obviously, this is not because it is the only way or even because it is somehow the "best" way. Rather, it is because this approach will help you to see clearly and immediately why and how you can benefit from learning the concepts, principles, and techniques that are presented.

Even while you are taking your first course in statistics, you will be able to use what you are learning here in other classes in which you are enrolled this semester. These classes might be in sociology, in another one of the social science disciplines, or in other fields. Knowledge of basic statistics will serve you well in more advanced courses in social science and in graduate studies if you choose to go on after your bachelor's degree. Your research reports and term papers will reflect a more mature grasp of the strengths and the limitations of the scientific method, and of what one can (or cannot) "prove" with factual observations. On a more general level, your reasoning powers will become more orderly and more disciplined than they might otherwise have been. Ultimately, the lessons learned from a practical introduction to statistics such as this will serve you well in getting a job in any of several professions and in doing good, effective work after you are hired.

The "applying" part of the title of this book is reflected in several small ways and in two major themes that run throughout the chapters that follow. One of these focuses on how we *use* social statistics with the personal computer (PC). The second focuses on how applied sociologists and other practicing social scientists *apply* statistics in their assessment research, evaluation studies, community organizing, and various types of consulting for governments, nonprofit organizations, and businesses. The next few pages introduce these two themes and discuss how they are integrated with the basic principles and techniques of social statistics.

Statistics for Sociologists[1]

The Origins of Sociology and Modern Statistics

One could argue almost endlessly about who the founder of sociology was. A major reason why this question is difficult to resolve is that sociology did not have one inventor. Rather, several scholars in various parts of the world developed the elements of the discipline over the course of centuries. However, if we had to select the person most often identified as the first "true" sociologist, it would be Auguste Comte (born in Montpellier, France, 1798, and died in Paris, 1857).

Comte is accorded this place in intellectual history partly because he invented the term (*sociologie* in French). But his fame is also the result of the set of principles he summarized, which have endured as the foundation of the field to this day. These are:

1. The scientific study of human relationships and the application of the results of this study to improve the human condition are necessary components of contemporary society. As in any other scientific field, the truths of sociology must ultimately be based on careful observation of relevant phenomena. What we observe (see, hear, etc.), not what we deduce or wish was the case, forms the raw material of sociological thinking. Following his teacher, Henri Saint-Simon, Comte referred to this doctrine of "seeing is believing" as *positivism*—another of his concepts still in use.

2. In contrast to psychology and biographical research, the focus of sociology is not on individuals as such but on aggregates: relationships (e.g., friendships, buyers and sellers), families and other groups, organizations, institutions, and societies. This is not to deny the importance of individuals, for they obviously are the participants in relationships, members of families, etc. However, specific friendships come and go, family members die and new ones are born, organizations experience turnover, yet the aggregate forms endure. In an important sense, there are no lone individuals in this world: all of us are defined by and act within the framework of aggregates. Thus, to understand humanity in an authentic scientific manner, we must understand the aggregate level of behavior.

3. Human behavior and experience at the aggregate level are exceedingly complex compared to the objects of study of other scientific fields. Numerous, sometimes enormously many, variables affect how we relate to one another, from two-person interactions (*dyads*) to international relations. Also, unlike the phenomena studied by other sciences, humans in the aggregate set goals, act with intentions, and make and execute plans—granted, often without much success and with frequent unintended outcomes. This means that the social world is shaped by a combination of physical, biological, psychological, and environmental forces,

but also by behavior motivated by the goal of affecting the shape of the social world. This fact underlies Herbert Spencer's invention of the term *superorganic*. This concept indicates that the principles that govern sociocultural phenomena include but also extend beyond the principles that govern other biological phenomena.

4. As in other sciences, those who study superorganic matters need a common language. This language must be universal in scope, it must avoid ambiguity, it must be precise, and it must facilitate inductive as well as deductive thinking. In brief, this common language must be mathematical.

If we ask what mathematical language deals most effectively (1) with truths based on observation, (2) at the aggregate level, and (3) with highly complex phenomena, the answer clearly is *statistics*. As Comte was living during an era of significant breakthroughs in mathematics, especially in probability theory and statistics, he viewed these fields as the most appropriate tools for sociological discourse. In recognition of this, he first called his new "positive science" *social physics*. However, he settled on *sociology* because "social physics" had been popularized by the Belgian statistician and sociologist Adolphe Quetelet (1796–1874; figure 1.1). As the *Collier Encyclopedia* notes, Quetelet

> developed many of the rules governing modern census taking and stimulated statistical activity in other countries. Applying statistics to social phenomena, he developed the concept of the 'average man' and established the theoretical foundations for the use of statistics in social physics or, as it is now known, sociology. Thus, he is considered by many to be the founder of modern quantitative social science.

Comte and Quetelet are two of the best-known founders to appreciate the fact that scientific sociology and statistics are closely related. However, in their age as in ours, this observation has been made and demonstrated by many researchers (in both fields). This book presents another illustration of the strong ties between statistics and sociology. We are not merely applying statistics to sociological examples, although there is much of this as well. Rather, we demonstrate that many of the key sociological issues of today—public sociology, civil society, sexualities, social inequality—are also, at root, statistical issues. That is, they are capable of formulation in statistical terms and they can be better understood through the use of statistical techniques.

Using Social Statistics with the Personal Computer

When the digital computer began to be used for business and educational applications in the years following World War II, it became clear to statisticians that a new era had dawned on their field. As you will soon see, many of the formulas employed to calculate statistical measures are quite simple in appearance, and they are also easy to understand once they are explained. However, when the number of units[2] to which these formulas are applied is large (in some cases even 10 or more is large), computation even with a good desktop calculator could be long, tedious, and prone to error. The mainframe computer introduced in the postwar era was ideal for such number crunching. The increased capacity to perform many repetitive calculations relatively quickly allowed social statisticians to work with larger and larger samples. In addition, it allowed them to test established theories and principles that had been developed prior to the computer era in experiments that used only a small and (what was then) a manageable number of cases.

FIGURE 1.1 Adolphe Quetelet (1796–1874), founder of statistical sociology. Photo by Milton Micallef, used with permission.

The Origins of Statistical Computing

The development of computer-oriented techniques of data analysis occurred in conjunction with the evolution of computer hardware. In the late 1950s, still during the mainframe-only era, a few programming languages were created that were in some respects the precursors of the codes used today. One of special interest to statisticians is FORTRAN (for "FORmula TRANslator"). This language, invented in 1959, is well suited to handling large databases and calculating all of the statistical measures discussed in this book: sums, averages, correlations, frequency distributions, crosstabs, and the like.[3] Several of FORTRAN's features led programmers to experiment with what we now term "packages." These are program-like languages that are compiled on a FORTRAN base (or a similar language) but that allow the user to write commands and to input data directly. By far the most widely used of these packages among sociologists and other social scientists is IBM–SPSS, which was known until 2009 as SPSS® (originally Statistical Package for the Social Sciences), produced by SPSS, Inc., based in Chicago and having a web site at http://www.spss.com. PASW began as mainframe software that quickly caught on among computer-oriented social researchers because of its ease and its repertoire of commands that fit well with the ongoing interests of users. Since the 1960s, it has served to teach

uncounted hundreds of thousands of students (and many of their teachers) how to employ the computer in quantitative research. (As of October 2009, SPSS, Inc., was acquired by IBM, motivating the name change to PASW and a new advertising emphasis on the use of the software to support business decision making.) By the late 1980s, statistical computer applications, and SPSS in particular, had made the transition to the PC. SPSS-PC, now known as IBM–SPSS and available for Microsoft Windows, Macintosh, and Linux, is a powerful, easy-to-use package that has turned a task that once required a thick manual and some ingenuity to decipher it into a point-and-click operation. The computer applications presented in this book are all geared to recent SPSS versions.

The Fundamentals of Statistical Computing

Most of the chapters of _Applying Social Statistics_ include an IBM–SPSS application in the _Study Guide_ that accompanies this book, for a total of 11 in all. Each application uses SPSS-PC, version 14.0 or later, and one or more of the data files included with the software. Topics and the corresponding chapters are:

Chapter 2: Opening a Data File, Sampling, and Comparing Two Data Files

Chapter 3: Creating Your Own Data File

Chapter 4: FREQUENCY and CROSSTABS Commands

Chapter 6: Central Tendency and Variability

Chapter 7: Sampling and Sampling Distributions

Chapter 8: Finding and Using Confidence Intervals

Chapter 10: Two-Sample t-Tests

Chapter 11: ONE WAY Analysis of Variance

Chapter 12: Chi-square, Confidence Intervals and Differences of Proportions, and Spearman's Rank-Order Correlation

Chapter 13: Correlation

Chapter 14: Regression

Social Statistics in Application

Until the early 1980s, few people thought of majoring in sociology or related social science disciplines as preparation for a nonacademic occupation. Rather, students and instructors alike viewed a B.A. in sociology, political science, anthropology, etc., as essentially a basic studies degree with no particular career focus. It was expected that a

small minority of B.A. holders in these fields would continue with graduate studies and eventually earn a Ph.D. This, in turn, would lead to a career of conducting research and teaching at the college or university level. Most undergraduate majors would, however, end up doing something that had no direct ties to sociology; the rest would find themselves teaching college-level sociology and perpetuating the pattern.

Since that time, much has changed. Increasingly many opportunities now exist for graduates with a bachelor's or master's degree in sociology and related disciplines to apply outside of academe the skills they have acquired. Professional organizations dedicated to promoting application have been created and are actively involved in changing the teaching philosophies and the public images of social science. Their goal is to make it clear that one can "do something" with a degree in one of these fields other than teach it. Two of the most important of these organizations are the Association for Applied and Clinical Sociology (AACS) and the Sociological Practice Section of the American Sociological Association (SP-ASA).[4] To emphasize how social statistics is used in solving real-world problems, this book discusses the work of these and similar organizations and their members at several points. In particular, each chapter includes one or two case studies based on the publications of AACS that illustrate the chapter's main concepts or techniques *in practice.*

These studies and other aspects of this book that are application-focused will make it clear that one of the most valuable skills that an undergraduate degree in sociology or a related social science offers is the ability to understand statistical reports and to perform basic statistical operations, especially with SPSS. Unlike many of the other social science courses required for a degree, in fact unlike most other courses regardless of the discipline, basic social statistics provides skills that can immediately be added to one's résumé and pay off on the job market. Although it remains difficult for a graduate to find a job as a "sociologist," or a "political scientist," employers are very interested in someone who can fit the requirements of a "statistical analyst," a "statistical report writer," or an "SPSS statistical programmer." Of course, neither this book nor any introductory statistics course can make one an expert in such fields. But they can give you a considerable head start in becoming an expert on the job and/or with continuing technical training.

By way of illustrating how social statistics can be used in practical contexts, each chapter includes "Solution-Centered Applications" that accompany each of the main principles and techniques introduced in the text. These are drawn from actual projects on which applied sociologists have worked during the past few years. They were provided by some of the leading practitioners in the field: Robert A. Dentler, William J. Hauser, AnnMarie Scarisbrick Hauser, Stephen F. Steele, Roger Straus, William DuBois, R. Dean Wright, and Richard E. Stephens. The authors and titles of these

books are listed at the end of this chapter, and full citations are included in the reference section following Chapter 15.[5] You are expected to give some careful thought to each of these projects, to carry each out with the statistical tools that are provided in the text, and to write a brief report on each. As you work toward achieving these goals, it will become obvious that social statistics is meant to be *applied*, and you will better appreciate how it can provide a very *useful* set of skills for solving real-world problems.

The Limits of Statistical Sociology

Statistics clearly plays a central role in sociology, especially in applied sociology. Yet it cannot substitute for well-developed theory, careful data collection, and old-fashioned hard work. A debate has been going on since the middle of the twentieth century between sociologists who are strongly committed to quantitative methods and computer application, on one side, and others who believe that such a commitment has led the field astray. The latter faction has argued that statistics is in danger of becoming a *substitute* for good sociology.

As with most other debates of this type, a middle ground does exist. It is not always appropriate to conduct research with quantitative variables that are susceptible to statistical analysis using SPSS or similar software. Whether such techniques are applicable depends on the problem at hand, the specific groups and individuals who are the subjects of the research, the availability of previously collected data, and numerous other factors. At the same time, quantitative research in sociology does have many advantages, including precision of findings and economy of presentation. Thus, statistical applications have a place, and an important place. In some respects they are a necessary part of the enterprise. But statistics is only *one part* of the enterprise. It may often be necessary, but it is not a sufficient basis for doing good sociology.

In Box 1.1 is an excerpt from an essay by William Foote Whyte about the limits of statistical, computer-assisted sociology. Whyte was a pioneer in participant observation research, made famous by his study of a Boston neighborhood, *Street Corner Society* (Whyte, 1955).

Looking Ahead

Now that we have outlined the main ways in which this is a first course in applied statistics, the next step is to get down to the business of examining the kinds of things you can expect to encounter as you learn. In this spirit, Chapter 2 turns to an overview of the field that includes basic definitions, a discussion of the relationship between statistics and the scientific method, and a close look at the processes of observation

BOX 1.1   # Statistics for Sociologists

Participant Observation and Statistical Sociology

WILLIAM FOOTE WHYTE

Until recently, it appeared to me that [participant-observer] research had gone out of style, overwhelmed by questionnaires processed with all the sophisticated methods made possible by modern technology. If I were to write a history of sociology from World War II into the 1970s, I would be inclined to give it the title of "Captured by Computers." It seems to me that the computerization of sociology began during World War II with the large-scale and impressive surveys leading to *The American Soldier*. Since that time, computer technology has advanced with such extraordinary rapidity as to make possible statistical and mathematical analysis hardly dreamed of in earlier periods.

Many sociologists have become so impressed with the powers of the computer as to lead them into planning research in terms of what the computer can do rather than in terms of what needs to be learned to advance knowledge regarding human society and social change. Let me hasten to add that I am not advocating abandonment of the computer. In collaboration with others more competent in surveys and statistical analysis, I have benefited greatly from the use of the computer. I am simply arguing that surveys—the most commonly used method of data gathering for computer processing—are not the only research method to be practiced by sociologists and learned by students.

I see encouraging signs that increasing numbers of sociologists now share this view. On the other hand, we should recognize that this is still distinctly a minority view. I was depressed recently to hear that in the graduate program of one of our leading departments of sociology, the standard way for a student to do a doctoral thesis did not require any field work at all. The student just got hold of a tape from the Institute of Social Research, the National Opinion Research Center, or some other research center, figured out a set of hypotheses different from those involved in the original survey, re-ran the tape through the local computer, reported the figures that came out, and then went on to try to find some plausible explanations for the apparent support or non-support of the hypotheses. To be sure, at this university students are not told that it is against the rules to go out and get their own data, but those who insist upon "doing it the hard way" are looked upon as rather peculiar types by their fellow students.

Apparently this is not a unique case. The chairman of a department at a major university described to me his recent experience in recruiting a new assistant professor. The department invited nine candidates from a number of departments to visit the

campus and give a talk on their doctoral theses. The professor reported that eight of the nine presented theses that conformed to the pattern I have described. When, in the discussion following the presentation, professors asked questions about what was going on in the field from which the numbers were produced, the candidates were unable to answer. Furthermore, their reactions suggested that they considered such questions illegitimate. They seemed to be saying, "If we have command of the numbers, what else can you ask of us?"

If this is to become the only way that sociological research is done, in the future we can visualize a complete division of labor between those who gather the data and those who design the research instrument, analyze the data, and publish the results. In this research style, the professional skills of research design and analysis are concentrated among people who have little if any contact with the field they study, and the field work is done generally by people without advanced training in sociology or any other social science. In fact, as is well known, in sociology or any other social science. In fact, as is well known, large survey organizations prefer to hire as data gatherers housewives, school teachers, and others without advanced social science training.

This style of research also deprives the sociologist of the experience necessary for effective performance in applied sociology or action research. It is a rare case indeed in which a sociologist simply reports how the numbers came out and then finds practitioners able to make practical use of the findings. Those who learn to interview and observe in the field are also acquiring skills of social interaction that are transferable in the action side of action research. The sociologist who can only interact with the computer and with his academic colleagues will not be able to work effectively with people in that "real world" beyond the Ivory Tower.

I expect increasing numbers of students to be attracted to field work, including participant observation, but many will still be troubled by the question of the legitimacy of such field methods. They will ask: "Are these methods scientific or artistic?" In my view, both science and art are involved. The student must not only grasp intellectually the nature of the skills required; he or she must also learn to perform those skills—which is where art comes in, for we must assume that some people will become more skillful than others.

At the same time, we are selling participant observation and related field methods short if we view them simply as art. These methods rest upon a scientific foundation, . . .

Those who fail to recognize this scientific foundation may be misled by a common error in describing the interviewing we do, along with participant observation, as nondirective or unstructured. If we are rigidly nondirective in the sense that we simply encourage the respondent to talk and let the respondent completely determine

the content of the talk, it is impossible to gather comparable data from a number of informants.

It seems to me that "flexibly structured" is a more accurate characterization of what we do with interviewing in the field. We do not pin the respondent down with a rigidly pre-structured series of questions, but neither do we simply invite the respondent to talk about whatever topic comes to mind in whatever way seems most comfortable to the respondent.

With [this] research style . . . the structure emerges in the mind of the field worker as he or she gains increasing familiarity with the people and problems in the field of study. We begin in a highly unstructured fashion, recognizing that we won't know really what are the important questions to ask until we have learned the lay of the land socially. As we see patterns emerging out of our data and out of our personal experience, we gradually build more structure into our field work. Increasingly, we focus questions designed [to] support or deny the pattern we think we see emerging. Even as we close in on this structure, we try to keep our minds open to the possibility that there is some important element that we have overlooked, and therefore we encourage the respondent to go beyond simple answers to our questions in ways that may point out to us relationships among topics that we had not previously perceived.

---

and measurement. At times, we may find ourselves deeply involved in technical details, which is to be expected in a course such as this. If this does occur and you find yourself asking, "What is this all for?" it will help to remember the preceding overviews of IBM–SPSS and applied social science. This might also be a good time to explore the IBM–SPSS exercises and the solution-centered projects provided at the end of most of the chapters. Although these aspects of the book may not always help to make every detail perfectly clear, they will allow you to see the larger picture and to make better sense of the place of the smaller details in it.