David Weisburd
Chester Britt
David B. Wilson
Alese Wooditch

Basic Statistics in Criminology and Criminal Justice

Fifth Edition

Springer

Basic Statistics in Criminology and Criminal Justice

Basic Statistics in Criminology and Criminal Justice

Fifth Edition

David Weisburd

Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA and Institute of Criminology, Faculty of Law, Hebrew University of Jerusalem, Jerusalem, Israel

Chester Britt

Iowa State University, Ames, IA, USA

David B. Wilson

Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

and

Alese Wooditch

Department of Criminal Justice, Temple University, Philadelphia, PA, USA

Springer

David Weisburd
Department of Criminology, Law and Society
George Mason University
Fairfax, VA, USA

Institute of Criminology
Faculty of Law
Hebrew University of Jerusalem
Jerusalem, Israel

David B. Wilson
Department of Criminology, Law and Society
George Mason University
Fairfax, VA, USA

Chester Britt (deceased)
Iowa State University
Ames, IA, USA

Alese Wooditch
Department of Criminal Justice
Temple University
Philadelphia, PA, USA

Additional material can be downloaded from www.springer.com

ISBN 978-3-030-47966-4 ISBN 978-3-030-47967-1 (eBook)
https://doi.org/10.1007/978-3-030-47967-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Oliver Wendell Holmes, the distinguished Associate Justice of the Supreme Court, was noted for his forgetfulness. On a train leaving Washington, D.C., he is said to have been approached by a conductor who requested his ticket. Holmes, searching through his case and his pockets, could not locate his pass. After a few awkward moments, the conductor recognized the distinctive-looking, well-known jurist and suggested that he just send the rail company the ticket when he found it. Justice Holmes is said to have looked sternly at the conductor and responded, "Young man, the problem is not where is my ticket; the problem is where am I going."

For the student of statistics, a textbook is like a train ticket. Not only does it provide a pass the student can use for entering a new and useful area of study, it also defines the route that will be taken and the goals that are important to achieve. Different textbooks take different approaches and emphasize different types of material. *Basic Statistics in Criminology and Criminal Justice* emphasizes the uses of statistics in research in crime and justice. This text is meant for students and professionals who want to gain a basic understanding of statistics in this field. In the first chapter, the main themes of the text are outlined and discussed. This preface describes how the text is organized.

The text takes a building-block approach. This means that each chapter helps prepare you for the chapters that follow. It also means that the level of sophistication of the text increases as the text progresses. Basic concepts discussed in early chapters provide a foundation for the introduction of more complex statistical issues later. One advantage to this approach is that it is easy to see, as time goes on, how much you have learned about statistics. Concepts that would have seemed impossible to understand, had they been introduced at the outset, are surprisingly simple when you encounter them later on. If you turn to the final chapters of the book now, you will see equations that are quite forbidding. However, when you come to these equations after covering the material in earlier chapters, you will be surprised at how easy they are to understand.

Throughout the text, there is an emphasis on *comprehension* and not *computation*. This approach is meant to provide readers with an accessible but sophisticated understanding of statistics that can be used to examine real-life criminal justice problems. In the opening chapters of the book, basic themes and materials are presented. Chapter 1 provides an introduction to how we use statistics in criminal justice and the problems we face in applying statistics to real-life research problems. Chapters 2 through 5 introduce basic concepts of measurement and basic methods for

graphically representing data and using statistics to describe data. Many of the statistics provided here will be familiar to you; however, remember that the more advanced statistics presented in later chapters build on the themes covered in these early chapters.

One of the fundamental problems researchers face is that they seek to make statements about large populations (such as all U.S. citizens) but are generally able to collect information or data on only a sample, or smaller group, drawn from such populations. In Chaps. 6 through 12, the focus is on how researchers use statistics to overcome this problem. What is the logic that underlies the statistics we use for making statements about populations based on samples? What are the different types of statistical procedures or tests that can be used? What special problems are encountered in criminal justice research, and how should the researcher approach them? Some texts skip over the basics, moving students from test to test before they understand the logic behind the tests. The approach here is to focus in greater detail on relatively simple statistical decisions before moving on to more complex ones.

Having examined how we can make statements about populations from information gained from samples, we turn to how we describe the strength of association between variables. In the social sciences, it is often essential not only to determine whether factors are related but also to define the strength and character of those relationships. Accordingly, in Chaps. 13 and 14, we look at measures of association, and in Chap. 15, we examine bivariate regression. These are likely to be new topics for you, though they are statistics commonly used in criminology and criminal justice.

While it is always difficult in statistics to decide where an introductory text should stop, with an understanding of these techniques you will have the basic tools to comprehend and conduct statistical analysis in criminology and criminal justice. Of course, these tools constitute a building block for more advanced methods. In a second volume, *Advanced Statistics for Criminology and Criminal Justice*, we will go beyond these basic statistical issues to address many of the more complex questions that researchers ask.

Each chapter starts with a statement of the basic concepts and problems addressed and ends with a full chapter summary. There is also a list of equations, when relevant, at the end of the chapter. These materials should help you to review what you have learned, and to identify the basic knowledge you need to move on to subsequent chapters.

All of the chapters contain a list of key terms with short definitions. The key terms appear in boldface the first time they are mentioned in the chapter. Sometimes a term may have been briefly explained in an earlier chapter, but is designated as a key term in the chapter where the concept is more central. A general glossary of key terms appears at the end of the book. The chapters also have a set of questions at the end. The questions are designed to make you think about the subjects covered in the chapter.

Sometimes they are straightforward, following directly from the text. Sometimes they ask you to develop ideas in slightly different ways than in the text. In constructing the questions, we sought to make working on statistical issues as much fun as possible. In statistics, it is crucial to go over material more than once. The questions are meant to reinforce the knowledge you have gained.

A working knowledge of computers is not required to understand the statistical concepts or procedures presented in the text. However, computers have become a very important part of research in statistics, and thus we provide computer exercises for relevant chapters and a website where you can access the data needed for those exercises (see the computer exercises at the end of Chap. 2 for details). We provide examples for computer questions in SPSS, Stata, and R, since these are the most widely used statistical computer packages in criminology.

Statistics in Criminology and Criminal Justice will allow you to approach statistics in a familiar context. It emphasizes the statistics and the problems that are commonly encountered in criminology and criminal justice research. It focuses on understanding rather than computation. However, it takes a serious approach to statistics, which is relevant to the real world of research in crime and justice. The text is meant not only as an introduction for students but also as a reference for researchers. The approach taken will help both students who want an introduction to statistics and professionals who seek a straightforward explanation for statistics that have become a routine tool in contemporary criminal justice systems. SPSS and Stata datasets and syntax for this volume are available at: https://github.com/AleseWooditch/Weisburd_et_al_Basic_Stats.

Jerusalem, Israel — David Weisburd
Fairfax, VA, USA — David B. Wilson
Philadelphia, PA, USA — Alese Wooditch

Acknowledgments

This volume draws substantial materials from *Statistics in Criminal Justice*, authored by David Weisburd and Chester Britt. We want to acknowledge our reliance on core parts of that book for our volume, and also our debt to Chester Britt, who passed away before we developed this book. Chester was a good friend and colleague, and was very much missed in the updating of the statistical materials and reorganization of topics for this book. He remains a key author for this volume because his intellectual contributions are found throughout these materials. On a personal note, we want to recognize how much we and criminologists more generally were saddened by his being taken from us so suddenly. We hope that through this book, his contributions will continue to be recognized by a new generation of scholars.

Contents

Chapter six

Chapter seven

Chapter eight

Chapter nine

Chapter ten

Chapter eleven

Chapter thirteen

Chapter fourteen

David Weisburd is a leading researcher and scholar in criminology and criminal justice. He is Distinguished Professor of Criminology, Law and Society at George Mason University in Virginia and Walter E. Meyer Professor of Law and Criminal Justice at the Hebrew University of Jerusalem. Professor Weisburd has received many awards and prizes for his contributions to criminology and criminal justice including the Stockholm Prize in Criminology and the Sutherland and Vollmer Awards from the American Society of Criminology.

Chester Britt was a leading researcher and scholar in the field of criminology. During his career, he taught at a number of universities and led departments at Northeastern University, Arizona State University, and the University of Iowa. His research addressed theories of criminal behavior and victimization, demography of crime and criminal careers, criminal justice decision-making, and quantitative research methods.

David B. Wilson is a Professor in the Criminology, Law and Society Department at George Mason University in Virginia. He is a social psychologist and leading applied statistician in the field of criminology, and was the recipient of the Mosteller Award from the Campbell Collaboration for his contributions to the science of systematic review and meta-analysis.

Alese Wooditch is an Assistant Professor of Criminal Justice at Temple University. She received her PhD from George Mason University. Professor Wooditch is interested in innovative spatial statistical analyses in the area of criminology and criminal justice experimental and computational criminology, and quantitative methodological issues.

Chapter one

Introduction: Statistics as a Research Tool

Initial hurdles

Do statisticians have to be experts in mathematics?

Are computers making statisticians redundant?

Key principles

What is our aim in choosing a statistic?

What basic principles apply to different types of statistics?

What are the different uses of statistics in research?

D. Weisburd et al., *Basic Statistics in Criminology and Criminal Justice*,
https://doi.org/10.1007/978-3-030-47967-1_1

The purpose of statistical analysis is to clarify and not confuse. It is a tool for answering questions. It allows us to take large bodies of information and summarize them with a few simple statements. It lets us come to solid conclusions even when the realities of the research world make it difficult to isolate the problems we seek to study. Without statistics, conducting research about crime and justice would be virtually impossible. Yet, there is perhaps no other subject in their university studies that criminology and criminal justice students find so difficult to approach.

A good part of the difficulty lies in the links students make between statistics and math. A course in statistics is often thought to mean long hours spent solving equations. In developing your understanding of statistics in criminology and criminal justice, you will come to better understand the formulas that underlie statistical methods, but the focus will be on concepts and not on computations. There is just no way to develop a good understanding of statistics without doing some work by hand. But in the age of computers, the main purpose of doing computations is to gain a deeper understanding of how statistics work.

Researchers no longer spend long hours calculating statistics. In the 1950s, social scientists would work for months developing results that can now be generated on a desktop computer in a few seconds. Today, you do not need to be a whiz kid in math to carry out a complex statistical analysis. Such analyses can be done with user-friendly computer programs. Why then do you need a course in statistics? Why not just leave it to the computer to provide answers? Why do you still need to learn the basics?

The computer is a powerful tool and has made statistics accessible to a much larger group of students and researchers. However, the best researchers still spend many hours on statistical analysis. Now that the computer has freed us from long and tedious calculations, what is left is the most challenging and important part of statistical analysis: identifying the statistical analysis methods that will best serve a researcher in answering their research question and interpreting the results for others.

The goal of this text is to provide you with the basic skills you will need to choose a statistical method for a research question and how to interpret the results. It is meant for students of criminology and criminal justice. As in other fields, there are specific techniques that are commonly used and specific approaches that have been developed over time by researchers who specialize in this area of study. These are the focus of this text. Not only do we draw our examples from crime and justice issues; we also pay particular attention to the choices that criminology and criminal justice researchers make when approaching statistical problems.

Before we begin our study of statistics in criminology and criminal justice, it is useful to state some basic principles that underlie the approach taken in this text. They revolve around four basic questions. First, what should our purpose be in choosing a statistic? Second, why do we use statistics to answer research questions? Third, what basic principles apply across very different types of statistics? And finally, what are the different uses for statistics in research?

The Purpose of Statistics Is to Clarify

It sometimes seems as if researchers use statistics as a kind of secret language. In this sense, statistics provide a way for the initiated to share ideas and concepts without including the rest of us. Of course, it is necessary to use a common language to report research results. This is one reason why it is important for you to take a course in statistics. But the reason we use statistics is to make research results easier—not more difficult—to understand. For example, if you wanted to provide a description of three offenders you had studied, you would not need to search for statistics to summarize your results. The simplest way to describe your sample would be just to tell us about your subjects. You could describe each offender and his or her criminal history without creating any real confusion. But what if you wanted to report on 20 offenders? It would take quite a long time to tell us about each one in some detail, and it is likely that we would have difficulty remembering who was who. It would be even more difficult to describe 100 offenders. With thousands of offenders, it would be just about impossible to take this approach.

This is one example of how statistics can help to simplify and clarify the research process. Statistics allow you to use a few summary statements to provide a comprehensive portrait of a large group of offenders. For example, instead of providing the name of each offender and telling us how many crimes he or she committed, you could present a single statistic that described the average number of crimes committed by the people you studied. You might say that, on average, the people you studied committed

two to three crimes in the last year. Thus, although it might be impossible to describe each person you studied, you could, by using a statistic, give your audience an overall picture. Statistics make it possible to summarize information about a large number of subjects with a few simple statements.

Given that statistics should simplify the description of research results, it follows that the researcher should utilize the simplest statistics appropriate for answering the research questions that he or she raises. Nonetheless, it sometimes seems as if researchers go out of their way to identify statistics that few people recognize and even fewer understand. This approach does not help the researcher or his or her audience. There is no benefit in using statistics that are not understood by those interested in your research findings. Using a more complex statistic when a simpler one is appropriate serves no purpose beyond reducing the number of people who will be influenced by your work.

The best presentation of research findings will communicate results in a clear and understandable way. When using complex statistics, the researcher should present them in as straightforward a manner as possible. The mark of good statisticians is not that they can mystify their audiences, but rather that they can communicate even complex results in a way that most people can understand.

Statistics Are Used to Solve Problems

The large variety of statistical methods that exist developed because of a need to deal with many different types of question or problem. In the example above, you were faced with the dilemma that you could not describe each person in a very large study without creating a good deal of confusion. We suggested that an average might provide a way of using one simple statistic to summarize a characteristic of all the people studied. The average is a statistical method or solution. It is a tool for solving the problem of how to describe many subjects with a short and simple statement.

As you will see in later chapters, statistics have been developed to deal with many different types of problems that researchers face. Some of these may seem difficult to understand at the outset, and indeed it is natural to be put off by the complexities of some statistics. However, the solutions that statisticians develop are usually based on simple common sense. Contrary to what many people believe, statistics follow a logic that you will find quite easy to follow. Once you learn to trust your common sense, learning statistics will turn out to be surprisingly simple. Indeed, our experience is that students who have good common sense, even if they have very little formal background in this area, tend to become the best statisticians. But in

order to be able to use common sense, it is important to approach statistics with as little fear as possible. Fear of statistics is a greater barrier to learning than any of the computations or formulas that we will use. It is difficult to learn anything when you approach it with great foreboding. Statistics is a lot easier than you think. The job of this text is to take you step by step through the principles and ideas that underlie basic statistics for criminology and criminal justice researchers. At the beginning, we will spend a good deal of time examining the logic behind statistics and illustrating how and why statisticians choose a particular solution to a particular statistical problem. What you must do at the outset is take a deep breath and give statistics a chance. Once you do, you will find that the solutions statisticians have developed make good sense.

Basic Principles Apply Across Statistical Techniques

A few basic principles underlie much of the statistical reasoning you will encounter in this text. Stating them at the outset will help you to see how statistical procedures in later chapters are linked one to another. To understand these principles, you do not need to develop any computations or formulas; rather, you need to think generally about what we are trying to achieve when we develop statistics.

The first is simply that *in developing statistics, we seek to reduce the amount of error as much as possible*. The purpose of research is to provide answers to research questions. In developing these answers, we want to be as accurate as we can. Clearly, we want to make as few mistakes as possible. The best statistic is one that provides the most accurate statement about your study. Accordingly, a major criterion in choosing which statistic to use—or indeed in defining how a statistic is developed—is the amount of error that a statistic incorporates. In statistics, we try to minimize error whenever possible.

Unfortunately, it is virtually impossible to develop any description without some degree of error. This fact is part of everyday reality. For example, we do not expect that our watches will tell perfect time or that our thermostats will be exactly correct. At the same time, we all know that there are better watches and thermostats and that one of the factors that leads us to define them as "better" is that they provide information with less error. Similarly, although we do not expect a stockbroker to be correct all of the time, we are likely to choose the broker who we believe will make the fewest poor investments.

In choosing a statistic, we also use a second principle to which we will return again and again in this text: *Statistics based on more information are generally preferred over those based on less information*. This principle is

common to all forms of intelligence gathering and not just those that we use in research. Good decision-making is based on information. The more information available to the decision-maker, the better he or she can weigh the different options that are presented. The same goes for statistics. A statistic that is based on more information, all else being equal, will be preferred over one that utilizes less information. There are exceptions to this rule, often resulting from the quality or form of the information or data collected. We will discuss these in detail in the text. But as a rule, the best statistic utilizes the maximum amount of information.

Our third principle relates to a danger that confronts us in using statistics as a tool for describing information. In many studies, there are cases that are very different from all of the others. Indeed, they are so different that they might be termed deviant cases or, as statisticians sometimes call them *outliers*. For example, in a study of criminal careers, there may be one or two offenders who have committed thousands of crimes, whereas the next most active criminal in the sample has committed only a few hundred crimes. Although such cases form a natural part of the research process, they often have very significant implications for your choice of statistics and your presentation of results.

In almost every statistic we will study, outliers present a distinct and troublesome problem. A deviant case can make it look as if your offenders are younger or older than they really are—or less or more criminally active than they really are. Importantly, deviant cases often have the most dramatic effects on more complex statistical procedures. And it is precisely here, where the researcher is often preoccupied with other statistical issues, that deviant cases go unnoticed. But whatever statistic is used, the principle remains the same: *Outliers present a significant problem in choosing and interpreting statistics.*

The final principle is one that is often unstated in statistics, because it is assumed at the outset: *Whatever the method of research, the researcher must strive to systematize the procedures used in data collection and analysis.* As Albert J. Reiss, Jr., a pioneer in criminology and criminal justice methods, has noted, that *systematic* means in part "that observation and recording are done according to explicit procedures which permit replication and that rules are followed which permit the use of scientific inference" (Reiss, 1971).

While Reiss' comment will become clearer as statistical concepts are defined in coming chapters, his point is simply that you must follow clearly stated procedures and rules in developing and presenting statistical findings. It is important to approach statistics in a systematic way. You cannot be sloppy or haphazard, at least if the statistic is to provide a good answer to the research question you raise. The choice of a statistic should follow a consistent logic from start to finish. You should not jump from one statistical test to another statistical test merely because the latter produced an

outcome more favorable to your thesis. In learning about statistics, it is also important to go step by step—and to be well organized and prepared. An important principle in statistical analysis is the replicability of analysis; can you repeat the steps in your statistical analysis (even years later) and reproduce your original results? You cannot learn statistics by cramming in the last week of classes. The key to learning statistics is to adopt a systematic process and follow it each week.

Statistical procedures are built on all of the research steps that precede them. If these steps are faulty, then the statistics themselves will be faulty. In later chapters, we often talk about this process in terms of the assumptions of the statistics that we use. We assume that all of the rules of good research have been followed up to the point where we decide on a statistic and calculate it. Statistics cannot be disentangled from the larger research process that comes before them. The numbers that we use are only as good as the data collection techniques that we have employed. Very complex statistics cannot hide bad research methods. A systematic approach is crucial not only to the statistical procedures that you will learn about in this text but to the whole research process.

The Uses of Statistics

In the chapters that follow, we will examine two broad types of statistics or two ways in which statistics are used in criminal justice. The first is called **descriptive statistics**, because it helps in the summary and description of research findings. The second, **inferential or inductive statistics**, allows us to make inferences or statements about large groups from studies of smaller groups, or samples, drawn from them.

Descriptive Statistics

We are all familiar in some way with descriptive statistics. We use them often in our daily lives, and they appear routinely in newspapers and on television. Indeed, we use them so often that we sometimes don't think of them as statistics at all. During an election year, everyone is concerned about the percentage support that each candidate gains in the primaries. Students at the beginning of the semester want to know what proportion of their grades will be based on weekly exercises. In deciding whether our salaries are fair, we want to know what the average salary is for other people in similar positions. These are all descriptive statistics. They summarize in one simple numeric statement the characteristics of many people or other units of interest. As discussed above in the example concerning criminal histories, descriptive statistics make it possible for us to summarize or describe large amounts of data.

In the chapters that follow, we will be concerned with two types of descriptive statistics: **measures of central tendency** and **measures of dispersion**. Measures of central tendency are measures of typicality. They tell us in one statement what the average case is like. If we could take only one person as the best example for all of the subjects we studied, who would it be? If we could choose only one level of criminal activity to typify the frequency of offending of all subjects, what level would provide the best snapshot? If we wanted to give our audience a general sense of how much, on average, a group of offenders stole in a year, what amount would provide the best portrait? Percentages, proportions, medians, and means are all examples of measures of central tendency that we commonly use. In the coming chapters, you will learn more about these statistics, as well as others with which you may not be familiar.

Having a statistic that describes the average case is very helpful in describing research results. However, we might also want to know how typical this average case is of the subjects in our study. The answer to this question is provided by measures of dispersion. They tell us to what extent the other subjects we studied are similar to the case or statistic we have chosen to represent them. Although we don't commonly use measures of dispersion in our daily lives, we do often ask similar questions without the use of such statistics.

For example, in deciding whether our income is fair, we might want to know not only the average income of others in similar positions but also the range of incomes that such people have. If the range was very small, we would probably decide that the average provides a fairly good portrait of what we should be making. If the range was very large, we might want to investigate more carefully why some people make so much more or less than the average. The range is a measure of dispersion. It tells us about the spread of scores around our statistic. In the chapters that follow, we will look at other measures of dispersion—for example, the standard deviation and variance, which may be less familiar to you. Without these measures, our presentation of research findings would be incomplete. It is not enough simply to describe the typical case; we must also describe to what degree other cases in our study are different from or similar to it.

Inferential Statistics

Inferential statistics allow us to make statements about a population, or the larger group of people we seek to study, on the basis of a sample drawn from that population. Without this very important and powerful tool, it would be very difficult to conduct research in criminal justice. The reason is simple. When we conduct research, we do so to answer questions about populations. But in reality, we seldom are able to collect information on the whole population, so we draw a sample from it. Statistical inference makes it possible for us to infer characteristics from that sample to the population.

Why is it that we draw samples if we are really interested in making statements about populations? In good part it is because gaining information on most populations is impractical and/or too expensive. For example, if we seek to examine the attitudes of U.S. citizens toward criminal justice processing, we are interested in how all citizens feel. However, studying all citizens would be a task of gigantic proportion and would cost billions of dollars. Such surveys are done every few years and are called censuses. The last census in the United States took many years to prepare and implement and cost over $5 billion to complete. If every research study of the American population demanded a census, then we would have very few research projects indeed. Even when we are interested in much smaller populations in the criminal justice system, examination of the entire population is often beyond the resources of the criminal justice researcher. For example, to study all U.S. prisoners, we would have to study over two million people.

Even if we wanted to look at the 200,000 or so women prisoners, it would likely cost millions of dollars to complete a simple study of their attitudes. This is because the most inexpensive data collection can still cost tens of dollars for each subject studied. When you consider that the National Institute of Justice, the primary funder of criminology and criminal justice research in the United States, provides a total of about $100 million a year for all social science research and development (National Institute of Justice, 2016), it is clear that criminology and criminal justice research cannot rely on studies of whole populations.

It is easy to understand, then, why we want to draw a sample or subset of the larger population to study, but it is not obvious why we should believe that what we learn from that sample applies to the population from which it is drawn. How do we know, for example, that the attitudes toward criminal justice expressed by a sample of U.S. citizens are similar to the attitudes of all citizens? The sample is a group of people drawn from the population; it is not the population itself. How much can we rely on such estimates? And to what extent can we trust such statistics? You have probably raised such issues already, in regard to either the surveys that now form so much a part of public life or the studies that you read about in your other college or university classes. When a news organization conducts a survey of 1,000 people to tell us how all voters will vote in the next election, it is using a sample to make statements about a population.

The criminology and criminal justice studies you read about also base their conclusions about populations—whether of offenders, criminal justice agents, crime-prone places, or criminal justice events—on samples. Statistical inference provides a method for deciding to what extent you can have faith in such results. It allows you to decide when the outcome observed in a sample can be generalized to the population from which it was drawn. Another way to think about this is that inferential statistics provide information about the precision of our descriptive statistics or the likelihood that our

findings are a statistical fluke or are credible. Statistical inference is a very important part of statistics and one we will spend a good deal of time discussing in this text.

Chapter Summary

Statistics seem intimidating because they are associated with complex mathematical formulas and computations. Although some knowledge of math is required, an understanding of the concepts is much more important than an in-depth understanding of the computations. Today's computers, which can perform complex calculations in a matter of seconds or fractions of seconds, have drastically cut the workload of the researcher. They cannot, however, replace the key role a researcher plays in choosing the most appropriate statistical tool for each research problem.

The researcher's aim in using statistics is to communicate findings in a clear and simple form. As a result, the researcher should always choose the simplest statistic appropriate for answering the research question.

Statistics offer commonsense solutions to research problems. The following principles apply to all types of statistics: (1) in developing statistics, we seek to reduce the level of error as much as possible; (2) statistics based on more information are generally preferred over those based on less information; (3) outliers present a significant problem in choosing and interpreting statistics; and (4) the researcher must strive to systematize the procedures used in data collection and analysis.

There are two principal uses of statistics discussed in this book. In **descriptive statistics**, the researcher summarizes large amounts of information in an efficient manner. Two types of descriptive statistics that go hand in hand are **measures of central tendency**, which describe the characteristics of the average case, and **measures of dispersion**, which tell us just how typical this average case is. We use **inferential statistics** to make statements about a population on the basis of a sample drawn from that population.

Key Terms

Descriptive statistics A broad area of statistics that is concerned with summarizing large amounts of information in an efficient manner. Descriptive statistics are used to describe or represent in summary form the characteristics of a sample or population.

Inferential, or inductive, statistics A broad area of statistics that provides the researcher with tools for making statements

about populations on the basis of knowledge about samples. Inferential statistics allow the researcher to make inferences regarding populations from information gained in samples.

Measures of central tendency Descriptive statistics that allow us to identify the typical case in a sample or population. Measures of central tendency are measures of typicality.

Measures of dispersion Descriptive statistics that tell us how tightly clustered or dispersed the cases in a sample or population are. They answer the question, "how typical is the typical case?"

References

Reiss, A. J., Jr. (1971). Systematic Observation of Natural Social Phenomena. *Sociological Methodology, 3,* 3–33. doi:10.2307/270816

National Institute of Justice (2016). *National Institute of Justice Annual Report: 2016.* - Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Chapter two

Measurement: The Basic Building Block of Research

Criminal justice research as a science

How do criminal justice researchers develop knowledge?

The four levels of measurement

What are they?

What are their characteristics?

How do the different levels interconnect?

Defining a good measure

Which is the most appropriate level of measurement?

What is meant by the validity of a measure?

What is meant by the reliability of a measure?

Electronic supplementary material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47967-1_2) contains supplementary material, which is available to authorized users. The datasets and SPSS/Stata syntax files can be found https://github.com/AleseWooditch/Weisburd_et_al_Basic_Stats.

D. Weisburd et al., *Basic Statistics in Criminology and Criminal Justice*,
https://doi.org/10.1007/978-3-030-47967-1_2

Measurement lies at the heart of statistics. Indeed, no statistic would be possible without the concept of measurement. Measurement is also an integral part of our everyday lives. We routinely classify and assign values to people and objects without giving much thought to the processes that underlie our decisions and evaluations. In statistics, such classification and ordering of values must be done in a systematic way. There are clear rules for developing different types of measures and defined criteria for deciding which are most appropriate for answering a specific research question.

Although it is natural to focus on the end products of research, it is important for the researcher to remember that measurement forms the first building block of every statistic. Even the most complex statistics, with numbers that are defined to many decimal places, are only as accurate as the measures upon which they are built. Accordingly, the relatively simple rules we discuss in this chapter are crucial for developing solid research findings. A researcher can build a very complex structure of analysis, but if the measures that form the foundation of the research are not appropriate for the analyses that are conducted, the findings cannot be relied upon.

We begin Chap. 2 by examining the basic idea of measurement in science. We then turn to a description of the main types of measures in statistics and the criteria used to distinguish among them. We are particularly concerned with how statisticians rank or classify measurement based on the amount of information that a measure includes. This concept, known as levels of measurement, is very important in choosing which statistical procedures are appropriate for the data. Finally, we discuss some basic criteria for defining a good measure.

Science and Measurement: Classification as a First Step in Research

Criminology and criminal justice research is a scientific enterprise that seeks to develop knowledge about the nature of crimes, criminals, victims, crime places, and the criminal justice system. The development of knowledge can, of course, be carried out in a number of different ways. Researchers may, for example, observe the actions of criminal justice agents or speak to offenders. They may examine routine information collected by government or criminal justice agencies or develop new information through analyses of the content of records in the criminal justice system. Knowledge may be developed through historical review or even through examination of archaeological records of legal systems or sanctions of ancient civilizations.

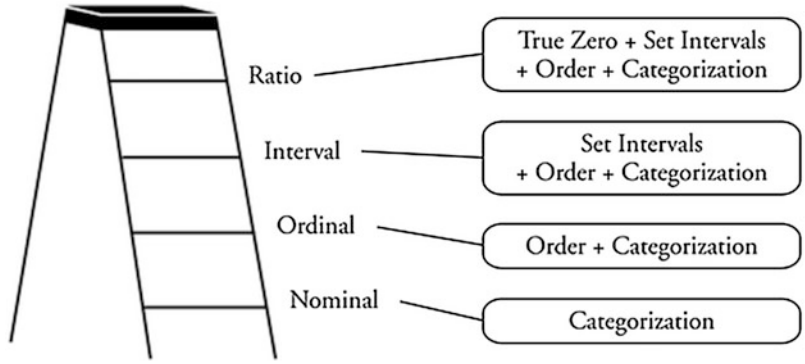
The methods that criminology and criminal justice researchers use vary. What they have in common is an underlying philosophy about how knowledge may be gained and what scientific research can tell us. This philosophy, which is predominant in scientific study in the modern world, is usually called positivism (Black, 1973). At its core is the idea that science is based on facts and not values or at a minimum strives to be value free. Science cannot make decisions about the way the world should be (although scientific observation may inform such decisions). Rather, it allows us to examine and investigate the realities of the world as we know it. The major tool for defining this reality in science is **measurement**.

Measurement in science begins with the activity of distinguishing groups or phenomena from one another. This process, which is generally termed **classification**, implies that we can place units of scientific study—such as victims, offenders, crimes, or crime places—in clearly defined categories or along some continuum. Each piece of information we classify is a datum, and we use a collection of multiple datum, referred to as **data**, to answer research questions. The classification of these data leads to the creation of **variables**. A variable is a trait, characteristic, or attribute that can be measured. What differentiates measurement in science from measurement in our everyday lives is that there must be systematic criteria for determining both what each category of a variable represents and the boundaries between categories. We now turn to a discussion of these criteria as they relate to different **levels of measurement**.

Levels of Measurement

Classification forms the first step in measurement. There are a number of different ways we can classify the people, places, or phenomena we wish to study. We may be content to simply distinguish one category from another.

Figure 2.1 *Levels of measurement*



But, we may also be interested in how those categories relate to one another. Do some represent more serious crime or less serious crime? Can we rank how serious various crimes are in a clear and defined order? Is it possible to define exactly how serious one crime is relative to another?

As these types of questions suggest, measurement can be a lot more complex than simply distinguishing one group from another. Recognizing this complexity, statisticians have defined four basic groups of measures, or levels of measurement, based on the amount of information that each takes advantage of. The four are generally seen as occupying different positions, or levels, on a ladder of measurement (see Fig. 2.1): nominal, ordinal, interval, and ratio. Nominal and ordinal variables are referred to as **qualitative** (data with qualities or characteristics). Interval and ratio data are measured on a **quantitative** scale (numeric data that represent quantities). Quantitative data can be further classified into **discrete** or **continuous** data. Discrete variables have values that are obtained through counting and thus are whole numbers (e.g., the number of heads flipped). Continuous variables have values that are obtained through measurement and can be any value within a range of values, including decimals or fractions.

Following a principle stated in Chap. 1—that statistics based on more information are generally preferred—measures that include more information rank higher on the level of measurement.

Nominal Scales

At the bottom of the ladder of measurement are **nominal scales**. Nominal-scale variables simply distinguish one phenomenon from another. Nominal scales are considered **categorical** variables because they divide cases into groups or categories, where the order of the categories does not matter.

Suppose, for example, that you want to measure crime types. In your study, you are most interested in distinguishing between violent crime and other types of crime. To fulfill the requirements of a nominal scale, and thus the minimum requirements of measurement, you need to be able to take all of the crime events in your study and place them in one of two categories: either violent crime or other crime. There can be no overlap. In practice, you might come across many individual events that seem difficult to classify. For example, what would you do with a crime event in which the offender first stole from his victim and then assaulted him? This event includes elements of both violent and property crime. What about the case where the offender did not assault the victim but merely threatened her? Would you decide to include this in the category of violent crime or other crime?

In measurement, you must make systematic choices that can be applied across events. You cannot decide one way for one event and another way for another. In the situation described above, you might conclude that the major issue in your study was the presence of violence. Thus, all cases with any violent events would be placed in the violent category. Similarly, you might conclude that violence had to include physical victimization. Whatever your choice, to meet the requirements of measurement you must define clearly where all events in your study are to be placed.

In criminology and criminal justice, we often make use of nominal-scale variables. Many of these reflect simple dichotomies, like the distinction between violent and other crime. For example, criminologists often seek to examine differences between men and women in their involvement in criminality or treatment in the criminal justice system. It is common as well to distinguish between those who are sentenced to prison and those who are not or those who commit more than one crime ("recidivists") and those who are only one-time offenders.

It is often necessary to distinguish among multiple categories of a nominal-level variable. For example, if you wanted to describe legal representation in court cases, you would provide a very simplistic picture if you simply distinguished between those who had some type of legal representation and those who did not. Some of the offenders would be likely to have private attorneys and others court-appointed legal representation. Still others might gain help from a legal aid organization or a public defender. In order to provide a full portrait of legal representation, you would likely want to create a nominal-scale variable with five distinct categories: no attorney, legal aid, court appointed, public defender, and private attorney. Table 2.1 presents a number of examples of nominal-level scales commonly used in research in criminology and criminal justice.

Nominal-scale measures can include any number of different categories. The Uniform Crime Reporting system, which keeps track of arrests in the United States, includes some 29 categories of crime. These range from

Table 2.1 Nominal-scale variables commonly found in criminology and criminal justice

VARIABLE	COMMON CATEGORIES
Gender	Male, female, non-binary
Race-ethnicity	Non-Hispanic Black, non-Hispanic White, Hispanic (any race), other/multiple
Marital status	Single, married, separated, divorced, widowed
Pretrial release status	Detained, released
Type of case disposition	Dismissed, acquitted, diverted, convicted
Method of conviction	Negotiated guilty plea, non-negotiated guilty plea, bench trial, jury trial
Type of punishment	Incarceration, non-incarceration

violent crimes, such as murder or robbery, to vagrancy and vandalism. Although there is no statistical difficulty with defining many categories, the more categories you include, the more confusing the description of the results is likely to be. If you are trying to provide a sense of the distribution of crime in your study, it is very difficult to practically describe 20 or 30 different crime categories. Keeping in mind that the purpose of statistics is to clarify and simplify, you should try to use the smallest number of categories that will accurately describe the research problem you are examining.

At the same time, do not confuse collection of data with presentation of your findings. You do not lose anything by collecting information in the most detailed way that you can. If you collect information with a large number of categories, you can always collapse a group of categories into one. For example, if you collect information on arrest events utilizing the very detailed categories of the criminal law, you can always combine them later into more general categories. But, if you collect information in more general categories (e.g., just violent crime and property crime), you cannot identify specific crimes such as robbery or car theft without returning to the original source of your information.

Though nominal-scale variables are commonly used in criminology and criminal justice, they provide us with very limited knowledge about the phenomenon we are studying. As you will see in later chapters, they also limit the types of statistical analyses that the researcher can employ. In the hierarchy of measurement, nominal-scale variables form the lowest step in the ladder. One step above are **ordinal scales**.

Ordinal Scales

What distinguishes an ordinal from a nominal scale is the fact that we assign a clear order to the categories included. Now, not only can we distinguish between one category and another, we also can place these categories on a continuum. This is a very important new piece of information; it allows us to rank events and not just categorize them. In the case of crime, we might decide to rank in order of seriousness. In measuring crime in this way, we

would not only distinguish among categories, such as violent, property, and victimless crimes, we might also argue that violent crimes are more serious than property crimes and that victimless crimes are less serious than both violent and property crimes. We need not make such decisions arbitrarily. We could rank crimes by the amount of damage done or the ways in which the general population rates or evaluates different types of crime.

Ordinal-scale variables are also commonly used in criminal justice and criminology. Indeed, many important criminal justice concepts are measured in this way. For example, in a well-known London survey of victimization, fear of crime was measured using a simple four-level Likert-type scale. Researchers asked respondents: "Are you personally concerned about crime in London as a whole? Would you say you are (1) very concerned, (2) quite concerned, (3) a little concerned, or (4) not concerned at all?" (Sparks et al., 1977).

Ranking crime by seriousness and measuring people's fear of crime are only two examples of the use of ordinal scales in criminal justice research. We could also draw examples regarding severity of court sentences, damage to victims, complexity of crime, or seriousness of prior records of offenders, as illustrated in Table 2.2. What all of these variables have in common is that they classify events and order them along a continuum. What is missing is a precise statement about how various categories differ one from another.

Interval and Ratio Scales

Nominal and ordinal scales are qualitative measures (descriptive or categorical meaning), whereas interval and ratio scales are considered quantitative (numeric meaning). **Interval scales** not only classify and order people or events, they also define the exact differences between them. An interval scale requires that the intervals measured be equal for all of the categories of the scale examined. Interval scales also have an arbitrary zero,

Table 2.2 Ordinal-scale variables commonly found in criminology and criminal justice

VARIABLE	COMMON CATEGORIES
Level of education	Less than high school, some high school, high school graduation, some college or trade school, college graduate, graduate/professional school
Severity of injury in an assault	None; minor, no medical attention; minor, medical attention required; major, medical attention required with no hospitalization; major, medical attention required with hospitalization
Attitude and opinion survey questions	Strongly disagree, disagree, no opinion, agree, strongly agree, very high, high, moderate, low, very low
Bail-release decision	Released on own recognizance, released on bail, detained—unable to post bail, denied release
Type of punishment	Probation/community service, jail incarceration, prison incarceration, death sentence

and it is not meaningful to perform multiplication or division with variables measured at this level. Internal-level variables are uncommon in the field of criminal justice. The classic examples of interval-level measures are the two common scales for measuring temperature: Fahrenheit and Celsius. A change in 1 degree Fahrenheit from 50 to 51 degrees is the same amount of change as from 80 to 81 degrees Fahrenheit. Similarly, on the Celsius scale, each change of one degree equals the same amount of change in temperature as any other one-degree change—the intervals are equal; the zero point on each of these scales is arbitrary and does not represent the absence of temperature, as zero on the Kelvin scale represents. Because of this, ratios of temperatures measured on these two interval scales are not meaningful. For example, 80 degrees Fahrenheit is not twice as hot as 40 degrees Fahrenheit. On the Celsius scale, these temperatures would be 26.667 and 4.444 degrees, a sixfold difference. Other common examples of interval scales are credit score, GRE score, time (e.g., 4 PM, 8 PM), and pH.

Quantitative variables where the values have equal intervals but there is a true zero are measured at the **ratio scale**. This means simply that zero represents the absence of the trait under study. For instance, number of arrests is measured at the ratio scale; it does not make sense to have a negative number of arrests (ratio scales can have negative numbers, indicating a negative quantity of what is being measured, such as a negative bank balance). Criminology and criminal justice researchers use ratio scales to present findings about criminal justice agency resources, criminal sentences, and a whole host of other issues related to crimes and criminals. For example, we can measure the amount spent by criminal justice agencies to pay the salaries of police officers or to pay for the healthcare costs of prison inmates. We can measure the financial costs of different types of crime by measuring the amount stolen by offenders or the amount of time lost from work by violent crime victims. We can measure the number of years served in prison/sentenced or the age at which offenders were first arrested. Also, unlike interval scales, it may be meaningful to perform multiplication and division on ratio-level variables. Table 2.3 provides examples of criminal justice variables that are measured at the ratio scale.

It is also important to realize that some criminology and criminal justice researchers commonly treat measures that are technically ordinal as if they were interval measures, such as with Likert-type scales. Another example that would be meaningful to most students is letter grades. At least within the United States, final course grades are assigned as the letters A, B, C, D, and F, with A being the top grade and F reflecting course failure. These may also be modified with pluses and minus, creating additional ordinal categories. It is also common to compute a student's grade point average or GPA. This is done by assigning each of the letters a number value, starting at 4 for the letter A and descending to a 0 for the letter F. In doing so, we are assuming that the difference between an A and a B is the same as the

Table 2.3 Variables commonly found in criminology and criminal justice research that are measured on a ratio scale

VARIABLE	COMMON CATEGORIES
Age	Years
Education	Years
Income or salary	Dollars
Number of crimes in a hot spot/community/city/county state nation	Count
Crime rates for a city/county/state/nation	Count of crimes, adjusted for the size of the population
Self-reported delinquent acts	Count

difference between a B and a C, etc. Most any student, however, would argue that the differences between these letter grades are often not equal. A consequence of this is that GPA is only an *approximate* ranking of student performance across a collection of classes because the computation of each student's GPA has assumed equal intervals even though that assumption may not fully hold. This aspect differentiates ordinal scales because the difference between values must be equal with ratio and interval scales.

Once aspect that influences a researcher's decision to treat an ordinal scale as interval or ratio is the number of ordinal categories which also influences the options a researcher has in selecting a statistical method. With modern computers, complex statistical methods for ordinal data are feasible and easy to implement; however, these methods generally only work for ordinal scales with a limited number of categories. For example, while it is quite feasible to use methods for ordinal data to analyze a single item on a 5-point Likert scale (e.g., 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree), it is generally not feasible to apply these methods to a composite of 20 such items. A composite (either sum or average) of 20 5-point Likert (pronounced *lick urt*) scales will have 96 possible values (a summed score can range from 5 to 100). While such a composite is technically still an ordinal measure (since the difference between Likert-values is not always equivalent), it is common practice in the social sciences to treat such a composite as an interval-level scale. In so doing, however, we are assuming that these data are behaving in a fashion consistent with interval-level data, and computing the mean of such values would be treating the measure as if it were an interval scale. These are generally safe practices so long as the shape of the distribution is appropriate for the statistical method used.

Now that we have defined each step in the ladder of measurement, we can summarize. As is illustrated in Table 2.4, as you move up the ladder of measurement, the amount of information that is gained increases. At the lowest level, you have only categorization. At the next level, you add knowledge about the order of the categories included. With interval scales, you not only classify and order your measure but also define how much

Table 2.4 Summary of the information required for each level of measurement

LEVEL OF MEASUREMENT	CATEGORIZATION	ORDER + CATEGORIZATION	SET INTERVALS ORDER + CATEGORIZATION	TRUE ZERO + SET INTERVALS + ORDER + CATEGORIZATION
Ratio	X	X	X	X
Interval	X	X	X	
Ordinal	X	X		
Nominal	X			

categories differ one from another (but have an arbitrary zero). A ratio scale requires all of these characteristics as well as a non-arbitrary, or true, zero value.

Relating Interval, Ordinal, and Nominal Scales: The Importance of Collecting Data at the Highest Level Possible

One important lesson we can draw from the ladder of measurement is that you should measure variables in a study at the highest level of measurement your data allow. This is because each higher level of measurement requires additional information. And if you fail to collect that information at the outset, you may not be able to add it at the end of your study. In general, variables measured lower on the ladder of measurement cannot be transformed easily into measures higher on the ladder. Conversely, variables measured higher on the ladder of measurement can be transformed easily into measures lower on the ladder. Take, for example, the measurement of victimization. If you decided to simply compare the types of victimization involved in a crime event, you would measure victimization using a nominal scale. You might choose the following categories: events involving loss of money or property, events including physical harm, a combination of such events, and all other events. But, let us assume for a moment that at some time after you collected your data, a colleague suggests that it is important to distinguish not only the type of event but also the seriousness of crimes within each type. In this case, you would want to distinguish not only whether a crime included monetary loss or violence but also the seriousness of each loss. However, because your variable is measured on a nominal scale, it does not include information on the seriousness of loss. Accordingly, from the information available to you, you cannot create an ordinal-level measure of how much money was stolen or how serious the physical harm was.

Similarly, if you had begun with information only on the order of crime seriousness, you could not transform that variable into one that defined the exact differences between categories you examined. Let's say, for example, that you received data from the police that ranked monetary victimization for each crime into four ordinally scaled categories: no monetary harm, minor monetary harm (up to $500), moderate monetary harm ($501–10,000), and serious monetary harm ($10,001 and above). If you decide that it is important to know not just the general order of monetary harm but also the exact differences in harm between crimes, these data are insufficient. Such information would be available only if you had received data about harm at the interval or ratio level of measurement. In this case, the police would provide information not on which of the four categories of harm a crime belonged to, but rather on the exact amount of harm in dollars caused by each crime.

While you cannot move up the ladder of measurement, you can move down it. Thus, for example, if you have information collected at a ratio level, you can easily transform that information into an ordinal-scale measure. In the case of victimization, if you have information on the exact amount of harm caused by a crime in dollars, you could at any point decide to group crimes into levels of seriousness. You would simply define the levels and then place each crime in the appropriate level. For example, if you defined crimes involving harm between $501 and $10,000 as being of moderate victimization, you would take all of the crimes that included this degree of victimization and redefine them as falling in this moderate category. Similarly, you could transform this measure into a nominal scale just by distinguishing between those crimes that included monetary harm and those that did not.

Beyond illustrating the connections among different levels of measurement, our discussion here emphasizes a very important rule of thumb for research. You should always collect information at the highest level of measurement possible. You can always decide later to collapse such measures into lower-level scales. However, if you begin by collecting information lower on the ladder of measurement, you will not be able to decide later to use scales at a higher level.

What Is a Good Measure?

In analysis and reporting of research results, measures that are of a higher scale are usually preferred over measures that are of a lower scale. Higher-level measures are considered better measures, based on the principle that they take into account more information. Nonetheless, this is not the only criterion we use in deciding what is a good variable in research. The

researcher must raise two additional concerns. First, does the variable reflect the phenomenon to be described? Second, will the variable yield results that can be trusted?

The first question involves what those who study research methods call **validity**. Validity addresses the question of whether the variable used actually measures the concept or construct it is intended to reflect. Is the variable measuring what it intends to measure? Few measures have perfect validity; the goal is to use or create measures that are sufficiently valid for the research purpose at hand. Thus, for example, collecting information on age in a sample is not a valid way of measuring criminal history. Age, although related to criminal history, is not a measure of criminal history. Similarly, work history may be related to criminality, but it does not make a valid measure of criminality. But, even if we restrict ourselves to variables that directly reflect criminal history, there are often problems of validity to address.

Let's say that you wanted to describe the number of crimes that offenders committed over a 1-year period. This seems like a straightforward thing to measure, but as we will discuss, finding a valid measure of this construct is difficult. One option you might have is to examine their criminal history as it is recorded on the Federal Bureau of Investigation's (FBI) criminal history record or rap sheet. The rap sheet includes information on arrests, convictions, and incarcerations. Although each of these variables tells us something about a person's criminal history, they are not all equally valid in terms of measuring the construct of interest.

The most valid measure of frequency of offending is the one that most accurately reflects how many crimes an individual has committed over a 1-year period, independent of whether that crime was detected by law enforcement. Clearly, arrests, convictions, and incarcerations fall short of this goal. Furthermore, each of these three variables is associated with additional validity concerns. Incarceration, for example, is more a measure of the seriousness of crime and the strength of the evidence against an individual than simply frequency of offending. This is because judges may impose a number of different types of sanctions, given sufficient evidence of guilt, and they are more likely to impose a prison sentence for more serious crimes. Many crimes that result in a conviction lead not to incarceration but rather to probation, fines, or community service. Thus, if we use incarceration to measure frequency of offending, we are likely to miss many crime events in an offender's criminal record. Accordingly, incarceration provides a biased picture of the number of offenses committed by an offender. It is not a highly valid measure of this concept. Similarly, convictions are also affected by the strength of evidence and a prosecutor's decision to press charges. Not all arrests result in convictions, even in the case of an arrestee that is in fact guilty.

Using this logic, criminologists have generally assumed that arrest is the more valid measure of frequency of offending among these three variables (i.e., that can be gained from official data sources, such as the FBI rap sheet). Arrests are much closer in occurrence to the actual behavior we seek to study and are not filtered by the negotiations found at later stages of the legal process. However, even arrests are not perfectly valid as not all arrests reflect actual wrong doing and not all wrong doing results in an arrest. While criminologists have assumed that arrests reflect criminal behavior more accurately than convictions or incarceration, some legal scholars contend that arrests are a less valid measure of criminality precisely because they come before the court reaches a conclusion regarding the innocence or guilt of a defendant. They contend that someone has not committed a crime until the legal system defines an act as such.

In contrast to these official measures of criminal behavior, many measures used in criminology and criminal justice research rely on self-reports of offending, often coming from data collected in surveys. Individuals are generally the best source of information regarding their behaviors, attitudes, opinions, beliefs, or other characteristics. Valid survey items and measures built from those items (e.g., measures may be based on a single item or composites of items) must still reflect the construct measured. This is often assessed through face validity: does the item or composite score appear on its face to measure the construct of interest? For example, a measure of procedural justice based on a series of Likert items would be judged to be face valid if the items seem to reflect the four principles of procedural justice: fairness of the process, transparency of the process, providing the opportunity for voice for those involved, and an impartial decision-maker. If the items don't reflect these principles or address other issues, such as satisfaction with the outcome, then the measure would be judged as invalid. Additionally, the items may look valid, but if people are unwilling to answer them honestly or if they have difficulty understanding the meaning of the question, then validity is compromised. There are other more complex ways of assessing a measure's validity, though these go beyond our discussions of measurement in this volume.

Reliability addresses the question of whether a measure gains information in a consistent manner. Will you get the same result for each person or other unit of interest on repeated measurement? If different people have similar characteristics, will your measure reflect that similarity? Returning to the above example of criminal history, we would not ask whether the measure reflects the concept of frequency of offending but whether measurement of the concept is consistent in the number it assigns to each person.

Returning to official measures of the number of crimes that a person has committed are likely to be reliable in that criminal justice agencies are careful to ensure that their official records are accurate and truly reflect

whether an individual has been arrested, convicted, or incarcerated. Thus, if two members of your research team were to determine the number of arrests for each person in your sample during a one-year period, the two sets of assessments are likely to agree with few if any exceptions. However, as discussed above, this doesn't mean that this reliable indicator of criminal behavior has high validity.

For self-report measures, we often think about reliability in terms of the consistency across two administrations of the same items or test-retest reliability. This assumes that what you are measuring hasn't changed during the intervening time period and that the individuals don't simply remember how they answered the questions the first time. For example, a measure of a teenager's risk-taking tendency would be judged as reliable if scores measured 3 months apart on a sample of youth were highly correlated. Risk-taking tendencies are likely to be fairly stable over time. For measures of constructs that may vary within an individual over time, such as mood, reliability is often assessed by examining the internal consistently of a collection of items that are all indicators of this construct. For example, a measure of procedural justice that is based on a composite of numerous individual questions would be judged reliable if those individual questions were internally consistent with one another, that is, if they tended to rank people in the same way.

It is important to realize that a measure can be valid but not reliable and reliable but not valid. It should be obvious that we want to use measures that are high on both dimensions. In practice, this is often not the case, and researchers must assess whether the measures are adequate for their research purpose. You should keep in mind that no variable is perfect. Some threat to validity is likely to be encountered, no matter how careful you are. Some degree of unreliability is almost always present in measurement. Your task is to develop or choose the best measure you can. The best measure is the one that most closely reflects the concept you wish to study and assesses it in a consistent and reliable way across subjects or events.

Chapter Summary

In science, we use **measurement** to make accurate observations. All measurement must begin with a **classification** process—a process that in science is carried out according to systematic criteria. This process implies that we can place units of scientific study in clearly defined categories. The end result of classification is the development of **variables**.

There are four **levels of measurement**: nominal, ordinal, interval, and ratio. Nominal and ordinal variables are **qualitative** or **categorical** in nature, whereas interval and ratio variables are **quantitative**. Quantitative

data can be **discrete** (take on only whole values) or **continuous** (take on any value including decimals in a range). With a **nominal scale**, information is organized by simple classification. The aim is merely to distinguish between different phenomena. There can be no overlap between categories nor can there be cases that do not fit any one category. There is no theoretical limit to the number of nominal categories possible. With an **ordinal scale**, not only is information categorized, but these categories are then placed in order of magnitude. An **interval scale** is one that, in addition to permitting the processes of categorization and ordering, also defines the exact difference between objects, characteristics, and events (without a true zero). A **ratio scale** is an interval scale for which a non-arbitrary or true zero value can be identified.

Data collected at a higher level of measurement may subsequently be reduced to a lower level, but data collected at a lower level may not be transformed to a higher one. For this reason, it is always advisable to collect data at the highest level of measurement possible.

There are three separate factors that affect the quality of a measure. The researcher should strive for a measure that has (1) a high scale of measurement (one that uses the most information); (2) a high level of **validity** (one that provides an accurate reflection of the concept being studied); and (3) a high level of **reliability** (one that provides consistent results).

Key Terms

Categorical Data that are grouped into categories that cannot vary in degree or rank.

Classification The process whereby data are organized into categories or groups.

Continuous Data that can be measured and take on any value within a specified range (including decimals or fractions).

Data Multiple pieces of information (datum) that are used to answer a research question.

Discrete Data that can be counted and take on whole numbers (integer values).

Interval scale A scale of measurement that uses a common and standard unit and enables the researcher to calculate exact differences between scores (but with an arbitrary zero), in addition to categorizing and ordering data.

Levels of measurement Types of measurement that make use of progressively larger amounts of information.

Measurement The assignment of numerical values to objects, characteristics, or events in a systematic manner.

Nominal scale A scale of measurement that assigns each piece of information to an appropriate category without suggesting any order for the categories created.

Ordinal scale A scale of measurement that categorizes information and assigns it an order of magnitude without using a standard scale of equal intervals.

Qualitative Data with qualities or characteristics.

Quantitative Numeric data that represent quantities.

Ratio scale A scale of measurement identical to an interval scale in every respect except that, in addition, a value of zero on the scale represents the absence of the phenomenon.

Reliability The extent to which a measure provides consistent results.

Validity The extent to which a variable accurately reflects the concept being measured.

Variable A trait, characteristic, or attribute of a person/object/event that can be measured at least at the nominal-scale level.

Exercises

2.1 For each of the following examples of criminal justice studies, state whether the level of measurement used is nominal, ordinal, or ratio. Explain your choice.

(a) In a door-to-door survey, residents of a neighborhood are asked how many times over the past year they (or anyone in their household) have been the victim of any type of crime.

(b) Parole board members rate inmate behavior on a scale with values ranging from 1 to 10; a score of 1 represents exemplary behavior.

(c) One hundred college students are asked whether they have ever been arrested.

(d) A researcher checks prison records to determine the racial background of prisoners assigned to a particular cell block.

(e) In a telephone survey, members of the public are asked which of the following phrases best matches how they feel about the performance of their local police force: *totally dissatisfied*, *dissatisfied*, *indifferent*, *satisfied*, or *very satisfied*.

(f) A criminologist measures the diameters (in centimeters) of the skulls of inmates who have died in prison, in an attempt to develop a biological theory of the causes of criminality.

(g) Legal assistants at a top legal firm are asked the following question: "Over the past year, have you been the victim of sexual harassment—and if so, how many times?" Answers are categorized as follows: *never*, *once*, *two or three times*, *more than three times*, or *refused to answer*.

2.2 You have been given access to a group of 12 jurors with a mandate from your senior researcher to "go and find out about their prior jury experience." Under each of the following three sets of restrictions, devise a

question to ask the jurors about the number of experiences they have had with previous juries.

(a) The information may be recorded *only* on a nominal scale of measurement.

(b) The information may be recorded on an ordinal scale but not on any higher scale of measurement.

(c) The information may be recorded on a ratio scale.

Your senior researcher subsequently informs you that she wishes to know the answers to the following five questions:

—How many of the jurors have served on a jury before?

—Who is the juror with the most prior experience?

—What is the sum total of previous jury experience?

—Is there anyone on the jury who has served more than three times?

—What is the average amount of prior jury experience for this group?

(d) If you had collected data at the nominal-level, which (if any) of the above questions would you be in a position to answer?

(e) If you had collected data at the ordinal level, which (if any) of the above questions would you be in a position to answer?

(f) If you had collected data at the ratio level, which (if any) of the above questions would you be in a position to answer?

2.3 You have been asked to measure the public's level of support for the death penalty. Devise questions to gauge each of the following:

(a) Overall support for the death penalty.

(b) Support for the death penalty if there are other punishment options.

(c) Support for using the death penalty if the chances of an innocent person being executed are:

(i) 1 in 1,000

(ii) 1 in 100

(iii) 1 in 10

2.4 You are investigating the effects of a defendant's prior record on various punishment decisions made by the court. One variable that you have access to in local court records is the total number of prior felony arrests for each defendant.

(a) What kinds of questions would you be able to answer with prior record measured in this way?

(b) Explain how you would recode this information on a nominal scale of measurement. What kinds of questions would you be able to answer with prior record measured in this way?

(c) Explain how you would recode this information on an ordinal scale of measurement. What kinds of questions would you be able to answer with prior record measured in this way?

2.5 Because the Ministry of Transport (MOT) is concerned about the number of road accidents caused by motorists driving too close together, it has painted *chevrons* (lane markings) every few meters in each lane on an experimental 2-km stretch of road. By the roadside, it has erected a sign that reads, "KEEP YOUR DISTANCE: STAY AT LEAST 3 CHEVRONS FROM THE CAR IN FRONT!" The MOT has asked you to measure the extent to which this instruction is being followed. There are a number of possible measures at your disposal. Assess the reliability and validity of each approach suggested below. Which is the best measure?

(a) Stand on a bridge over the experimental stretch of road and count how many of the cars passing below do not keep the required distance.

(b) Compare police figures on how many accidents were recorded on that stretch of road over the periods before and after it was painted.

(c) Study the film from a police camera situated 5 km farther down the same stretch of road (after the end of the experimental stretch) and count how many cars do not keep a safe distance.

2.6 The police are planning to introduce a pilot community relations strategy in a particular neighborhood and want you to evaluate whether it has an effect on the willingness of citizens to report crimes to the police. There are a number of possible measures at your disposal. Assess the reliability and validity of each approach suggested below. Which is the best measure?

(a) Telephone every household and ask respondents to measure, on a Likert-type scale of 1 to 10, how willing they are to report particular types of crime to the police. Repeat the experiment after the scheme has been in operation 6 months.

(b) Compare a list of offenses reported by members of the neighborhood in the 6 months before introduction of the scheme with a similar list for the 6 months after introduction of the scheme. (It is standard procedure for the police to record the details of the complainant every time a crime is reported to them.)

2.7 You are comparing the psychological condition of three inmates serving out long terms in different high-security prisons, and you are interested in the amount of contact each one has with the outside world. You wish to determine how many letters each one has sent over the past 12 months.

No official records of this exist. There are a number of possible measures at your disposal. Assess the reliability and validity of each approach suggested below. Which is the best measure?

(a) Ask each prisoner how many letters he or she sent over the past year.

(b) Check the rules in each of the prisons to see how many letters high security prisoners are allowed to send each year.

(c) Check the records of the prison postal offices to see how many times each prisoner bought a stamp over the past year.

2.8 The government is interested in the link between employment and criminal behavior for persons released from prison. In a study designed to test for an effect of employment, a group of people released from prison are randomly assigned to a job training program where they will receive counseling, training, and assistance with job placement. The other offenders released from prison will not receive any special assistance. There are a number of possible measures at your disposal. Assess the reliability and validity of each approach suggested below. Which is the best measure?

(a) Eighteen months after their release from prison, interview all the offenders participating in the study and ask about their criminal activity to determine how many have committed criminal acts.

(b) Look at prison records to determine how many offenders were returned to prison within 18 months of release.

(c) Look at arrest records to determine how many offenders were arrested for a new crime within 18 months of release.

2.9 In a recent issue of a criminology or criminal justice journal, locate a research article on a topic of interest to you. In this article, there should be a section that describes the data. A well-written article will describe how the variables were measured.

(a) Make a list of the variables included in the article and how each was measured.

(b) What is the level of measurement for each variable—nominal, ordinal, interval, or ratio level? Explain why.

(c) Consider the main variable of interest in the article. Assess its reliability and validity.

Computer Exercises

There are a number of statistical software packages available for data analysis. Most spreadsheet programs will also perform the basic statistical analyses of the kind described in this text through Chap. 15. The computer exercises included in this text focus on the use of three different software programs: SPSS, Stata, and R. Most

universities make at least one of these three statistical programs available in student computer labs. There are also student versions of SPSS and Stata that can be purchased separately, sometimes through a university bookstore or other offices on campus—see each company's website for details (www.spss.com or www.stata.com). R is a freely available software that can be downloaded at www.r-project.org. For the R exercises, it is expected that the user already has a working knowledge of the software. There are many excellent reference books on these programs for statistical data analysis—our intent here is not to repeat what is said in those books but rather our goal with the computer exercises is to illustrate some of the power available to you in three widely used packages. In real-world situations where you are performing some type of statistical analysis, you rarely work through a problem by hand, especially if the number of observations is large.[1]

Several SPSS and Stata files are available at http://extras.springer.com. These data files can be imported into R to complete the exercises. The data file we will use first represents a subset of the data from the National Youth Survey, Wave 1. The sample of 1,725 youth is representative of persons aged 11–17 years in the United States in 1976, when the first wave of data was collected. While these data may seem old, researchers continue to publish reports based on new findings and interpretations of these data. One of the apparent strengths of this study was its design; the youth were interviewed annually for 5 years from 1976 to 1980 and then were interviewed again in 1983 and 1987. The data file available in this book's supplemental resources was constructed from the full data source available at the Inter-University Consortium for Political and Social Research, which is a national data archive. Data from studies funded by the National Institute of Justice (NIJ) are freely available to anyone at www.icpsr.umich.edu/NACJD. All seven waves of data from the National Youth Survey are available, for example. The datasets and SPSS/Stata syntax files can be found https://github.com/AleseWooditch/Weisburd_et_al_Basic_Stats.

SPSS

To begin our exploration of SPSS, we will focus here on some of the data management features available to users. After starting the SPSS program on your computer, you will need to open the National Youth Survey data file from the website (nys_1.sav). For those readers working with the Student Version of SPSS, you are limited to data files with no more than 50 variables and 1,500 cases. We have also included a smaller version of the NYS data (nys_1_student.sav) that contains a random sample of 1,000 cases (of the original 1,725).

After you start SPSS and open the data file, the raw data should appear in a window that looks much like a spreadsheet. Each column represents a different

[1] If you would like an indepth tutorial of R, please see A Beginner's Guide to Statistics for Criminology and Criminal Justice using R by Wooditch, Johnson, Solymosi, Medina, and Langton (2021).

variable, while each row represents a different observation (individual, here). If you scroll down to the end of the data file, you should see that there are 1,725 lines of data (or 1,000 if using the student version of the data file).

There are three direct ways to learn about the variables included in this data file. First, notice the buttons in the lower center of the spreadsheet. One button (which should appear as the darker shade of the two buttons) is labeled "Data View," and the other is labeled "Variable View." The Data View button presents us with the spreadsheet of values for each observation and variable. If you click on the button labeled "Variable View," you should now see another spreadsheet, in which variable names are listed in the first column and the other columns contain additional information about each variable. For example, the first column provides the name of the variable, another column provides a label for the variable (allowing us to add a more informative description of our variable), and an additional column provides value labels. It is from this column that we will be able to learn more about each variable. For example, click on the cell in this column for the sex variable, and you should see a small gray box appear in the cell. Now click on this small gray box and you will be presented with a new window that lists possible values for the sex variable and the corresponding labels. Here, we see that males have been coded as "1" and females as "2." If you click on "OK" or "Cancel," the window disappears. You can then perform this same operation for every other variable.

A second way of obtaining information about the variables in an SPSS data file involves using the "Variables" command. To execute this command, click on "Utilities" on the menu bar; then, click on "Variables." What you should see is a list of variables on the left and another window on the right that presents information about the highlighted variable. If you click on the sex variable, you should see information on its coding and values in the window on the right. This command is particularly useful if you are working with an SPSS data file and simply need a reminder of how the variables are coded and what categories or values are included.

A third way of obtaining information about the variables in an SPSS data file involves the "Display Data File Information" command. To run this command, click on "File" on the menu bar; then, click on "Display Data File Information" and select the option for "Working File" (the data you have already opened up in SPSS). This command generates text for the output window in SPSS. This output contains all the information SPSS has on every variable in a data file. Executing this command is equivalent to executing the "Variables" command for every variable in the dataset and saving that information in another file. Be aware that using this command on a data file with many variables will produce a very large output file. This command is most useful when you are first working with an SPSS dataset that someone else has conveniently set up for you where you need to verify the contents of the dataset and the nature of the variables included in the dataset (similar to what you are now doing with the NYS data file).

In subsequent chapters, the computer exercises will make use of syntax commands and files. These are the command lines that each program uses to run a particular procedure. To begin, open a syntax window in SPSS by clicking on "File,"