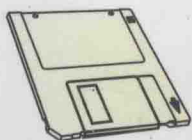


STATISTICS IN CRIMINAL JUSTICE

ANALYSIS AND

INTERPRETATION

Jeffery T. Walker



disk included

AN ASPEN PUBLICATION



STATISTICS IN CRIMINAL JUSTICE

Analysis and Interpretation

Jeffery T. Walker, PhD

Professor

Department of Criminal Justice
University of Arkansas, Little Rock
Little Rock, Arkansas



AN ASPEN PUBLICATION®

Aspen Publishers, Inc.
Gaithersburg, Maryland
1999

Library of Congress Cataloging-in-Publication Data

Walker, Jeffery T.
Statistics in criminal justice: analysis and interpretation /
Jeffery T. Walker.

p. cm

Includes bibliographical references and index.

ISBN 0-8342-1086-X (alk. paper)

1. Criminal justice, Administration of—Statistical methods.

2. Criminal statistics—Research. I. Title.

HV7415.W33 1999

364'.01'5195—dc21

99-29104

CIP

Copyright © 1999 by Aspen Publishers, Inc.

A Wolters Kluwer Company

www.aspenpublishers.com

All rights reserved.

Aspen Publishers, Inc., grants permission for photocopying for limited personal or internal use.
This consent does not extend to other kinds of copying, such as copying for general distribution,
for advertising or promotional purposes, for creating new collective works, or for resale.

For information, address Aspen Publishers, Inc., Permissions Department,
200 Orchard Ridge Drive, Suite 200, Gaithersburg, Maryland 20878.

Orders: (800) 638-8437

Customer Service: (800) 234-1660

About Aspen Publishers • For more than 35 years, Aspen has been a leading professional publisher in a variety of disciplines. Aspen's vast information resources are available in both print and electronic formats. We are committed to providing the highest quality information available in the most appropriate format for our customers. Visit Aspen's Internet site for more information resources, directories, articles, and a searchable version of Aspen's full catalog, including the most recent publications: <http://www.aspenpublishers.com>

Aspen Publishers, Inc. • The hallmark of quality in publishing
Member of the worldwide Wolters Kluwer group.

Editorial Services: Nora Fitzpatrick

Library of Congress Catalog Card Number: 99-29104

ISBN: 0-8342-1086-X

Printed in the United States of America

1 2 3 4 5

This book is dedicated to
Orval and Ann Walker.

Table of Contents

PART I—FUNDAMENTAL CONCEPTS OF STATISTICAL ANALYSIS

Chapter 1—The Logic of Comparisons and Analysis	1
Introduction: Why Analyze Data?	1
Some Statistical History	1
Uses of Statistics	2
Theory Construction at a Glance	2
What Is Theory?	3
Theory and Research	3
Observation and Inquisitiveness	5
Primary Questions	5
Research Questions	5
Research: Movement from Theory to Data and Back	6
Formulating Hypotheses	6
Constructing the Research Design	7
Developing Concepts	7
Operationalizing	8
Gathering the Data	9
Statistical Analysis: The Art of Making Sound Comparisons	10
Foundations of Valid Comparisons	11
Comparing Appropriate Phenomena	11
Using Comparable Measures	11
Choosing Analysis Methods That Best Summarize the Data	12
Drawing Conclusions	12
Communicating the Results	12
Data and Purposes of This Book	13

Chapter 2—Variables and Measurement19

 The Variable Defined19

 Transforming Characteristics into Data: The Process of Measurement19

 How Variables Can Differ21

 Levels of Measurement21

 Nominal22

 Ordinal23

 Interval24

 Ratio27

 The Process for Determining Level of Measurement28

 Changing Levels29

 Scale Continuity30

 Use in the Research Process31

 Dependent Variable31

 Independent Variable31

 Confounding Variable32

 Conclusion34

PART II—UNIVARIATE STATISTICS

**Chapter 3—Not All Statistical Analyses Involve Numbers: Summarizing Data
through Graphical Representation43**

 Introduction43

 Frequency Distributions: A Chart of a Different Color44

 Conventions for Building Distributions45

 Frequency Distributions46

 Frequency Distributions for Grouped Data47

 Percentage Distributions48

 Combination Distributions50

 Graphical Representation of Frequencies50

 Pie Charts51

 Histograms and Bar Charts52

 Polygons and Area Charts54

 Analyzing Univariate Statistics57

 Analyzing Change58

 Line Charts58

 Ogives59

 Analyzing Bivariate and Multivariate Data60

 Scatter Plots60

 Normal Probability or P-P Plots61

 Path Diagrams63

 Analyzing Geographic Distributions64

 Pin, Spot, or Point Maps64

 Choropleth Maps64

 Constructing Effective Choropleth Maps66

Problems With Choropleth Maps	66
Conclusion	69
Chapter 4—Measures of Central Tendency	73
Measures of Central Tendency	73
Mode	74
Median	78
Median for Ungrouped Data	79
Median for Grouped Data	79
Mean	83
Selecting the Most Appropriate Measure of Central Tendency	85
Conclusion	86
Chapter 5—Measures of Dispersion	93
Introduction	93
Deviation and Dispersion	93
Measures of Dispersion	94
Range	95
Index of Dispersion	96
Mean Absolute Deviation	99
Variance	100
Standard Deviation	103
What Use Do I Have for the Variance and Standard Deviation?	104
Selecting the Most Appropriate Measure of Dispersion	104
Conclusion	105
Chapter 6—The Form of a Distribution	111
Introduction	111
Moments of a Distribution	111
Number of Modes	111
Skewness	112
Analysis of Skew	113
Kurtosis	115
Analysis of Kurtosis	115
The Importance of Skewness and Kurtosis	115
Design of the Normal Curve	116
Points to Remember About the Normal Curve	121
Conclusion	121
PART III—BIVARIATE STATISTICS	
Chapter 7—Introduction To Bivariate Descriptive Statistics	127
Introduction	127
Bivariate Tables and Analysis	128
Statistical Tables versus Presentation Tables	128
Constructing Bivariate Tables	130
Ordinal Level Table Construction	131

Statistical Tables	133
Nominal Level Table Construction	134
Analysis of Bivariate Tables	134
Conclusion	135
 Chapter 8—Measures of Existence/Statistical Significance of a Relationship	139
Nominal Level Measures of Existence	139
Tables, Percentages, and Differences	139
Chi-square	143
Requirements for Using Chi-square	149
Limitations of Chi-square	151
A Final Note on Chi-square	152
Tests of Existence for Ordinal and Interval Level Data	152
Calculation and Interpretation for Ordinal Data	152
Spearman's Rho and Pearson's <i>r</i>	153
An Issue Of Significance	156
Conclusion	156
 Chapter 9—Measures of the Strength of a Relationship	163
What Is Association?	164
Nominal Level Data	164
Calculation	166
Interpretation	167
Ordinal Level Data	169
Tau	170
Calculation	170
Interpretation	174
Gamma	177
Calculation	177
Interpretation	178
Somers' d	179
Calculation	179
Interpretation	180
Spearman's Rho	181
Calculation	181
Interpretation	182
Limitations	183
Interval Level Data	183
Pearson's <i>r</i>	185
Calculation	187
Correlation Matrixes	188
Interpretation	188
Coefficient of Determination	189
Correlation and Causation	189
Limitations	191
Conclusion: Selecting the Most Appropriate Measure of Strength	191

Chapter 10—Measures of Direction and Nature of a Relationship	197
Direction of the Association	197
Establishing Direction for Ordinal Level Data	197
Establishing Direction for Interval/Ratio Level Data	200
Nature of the Association	201
Establishing the Nature of the Distribution for Nominal and Ordinal Level Data	201
Establishing the Nature of the Distribution for Interval/Ratio Level Data	202
Conclusion	204

PART IV—MULTIVARIATE STATISTICS: BRIDGING THE GAP BETWEEN DESCRIPTIVE AND INFERENTIAL STATISTICS

Chapter 11—Introduction to Multivariate Statistics	213
Introduction: When Two Variables Just Aren't Enough	213
Interaction Among Variables	213
Causation	216
Association	216
Temporal Ordering	216
Elimination of Rival Factors and Relationships	218
Other Issues for Multivariate Analysis	219
Robustness	219
Error	220
Parsimony	221
Conclusion	221
Chapter 12—Multivariate Measures of Association	225
Introduction	225
Regression	225
Assumptions	226
Analysis and Interpretation	228
Limitations	233
Factor Analysis	234
Assumptions	235
Analysis and Interpretation	236
Step 1—Univariate Analysis	236
Step 2—Preliminary Analyses	236
Step 3—Factor Extraction	237
Step 4—Factor Rotation	241
Step 5—Use of Factors in Other Analyses	246
ANOVA	246
Assumptions	247
Analysis and Interpretation	248
Conclusion	251

PART V—INFERENCEAL STATISTICS

Chapter 13—Introduction to Inferential Analysis	255
Descriptive and Inferential Analyses	255
Terminology and Assumptions	256
The Normal Curve	257
Probability	258
Sampling	260
Probability Sampling	261
Simple Random Sampling	261
Systematic Sampling	262
Stratified Sampling	262
Cluster or Multistage Sampling	262
Nonprobability Sampling	263
Purposive Sampling	263
Quota Sampling	263
Snowball Sampling	264
Accidental/Convenience Sampling	264
Sampling Distributions	264
Central Limit Theorem	266
Confidence Intervals	267
Calculating Confidence Intervals	267
Interpreting Confidence Intervals	269
Conclusion	270
Chapter 14—Hypothesis Testing	273
Introduction	273
Null and Research Hypotheses	273
Steps in Hypothesis Testing	275
Step 1—Developing the Research Hypothesis	275
Step 2—Developing the Null Hypothesis	275
Step 3—Drawing Samples	276
Step 4—Selecting the Test	276
One-Tailed and Two-Tailed Tests	276
Step 5—The Obtained Value	277
Step 6—Significance and Critical Regions	277
Step 7—Making a Decision	278
Type I and Type II Errors	279
Which Is Better, a Type I or Type II Error?	281
Power of Tests	281
Conclusion	283
Chapter 15—Hypothesis Tests	287
Introduction: Tests of Hypothesis	287
Z-Test	287
Calculation and Example	288
Interpretation and Application: Known Probability of Error	290

One-Sample versus Two-Sample Z-Tests	293
t -Test	294
Assumptions of a t -Test	295
Calculation and Example	296
SPSS Analysis for Z-Tests and t -Tests	297
One-Sample t -Test	298
Two-Sample t -Test	299
F -Test	301
Calculation and Example	301
SPSS Output for an F -Test	303
Chi-square Test for Independence	304
Conclusion	305
Chapter 16—Putting It All Together	309
Introduction	309
The Relationship Between Statistics, Methodology, and Theory	309
Describe It or Make Inferences	310
Abuses of Statistics	312
When You Are on Your Own	312
Conclusion	313
Appendix A—Math Review and Practice Test	315
Addition, Subtraction, and Negative Numbers	315
Multiplication, Division, and Negative Numbers	316
Statistical Addition: Summation	316
Exponents and Roots	317
Order of Operations	317
Putting It All Together: Simple Math for Statistical Analysis	317
Practice Test	318
Tables	320
Table A-1	320
Table A-2	322
Table A-3	323
Table A-4	324
Appendix B—List of Variables	327
Index	351
Disk Instructions	363

CHAPTER 1

The Logic of Comparisons and Analysis

*Statistical thinking will one day be as necessary for
efficient citizenship as the ability to read and write.*

—H. G. Wells

INTRODUCTION: WHY ANALYZE DATA?

Discovery and innovation may be the distinguishing characteristics between modern human activity and that of our ancestors. The Renaissance brought forth an emphasis on learning and advancing our way of doing things that has prevailed to the present. Scientists, inventors, and others involved in the process of scientific inquiry have often been held in awe for their works. Galileo, Einstein, Madam Curie, and others are singled out in grade school books for their works and discoveries. As you will learn, so too should be people like Pearson, Kendall, and Yule.

Statistical analysis is all about discovery. The process of *scientific inquiry* provides a method of examining things that interest us in a systematic manner. This process generally requires evidence to support an argument. One of the clearest methods of gathering evidence is by examining numbers associated with the objects being studied. That examination takes place through statistical analysis. As such, statistical analysis is the linchpin of discovery, and mastery of it draws us closer to Einstein and Galileo.

SOME STATISTICAL HISTORY

The earliest form of what is now considered statistical analysis was developed by Pythagoras in the sixth century B.C. This was the forerunner of descriptive statistics (what would eventually be known as the mean, or is commonly known as an average). The other type of statistical analysis (inferential statistics) is thought to have first developed in the Orient around 200 B.C. (Dudycha and Dudycha, 1972). This was a form of probability analysis used in assessing whether an expected child was likely to be male or female. Probability theory, as it would come to be known, continued in the form of gambling mathematics in the works of Blaise Pascal (1623–1662) and Lord Christianus Huygens (1629–1695) (David, 1962). Many of the other descriptive statistics were developed in the late 1800s and early 1900s by mathematicians and scientists such as Galton (1883) and Pearson (1895).

Statistics moved beyond gambling and purely mathematical concepts through what was called *political arithmetic*, a term coined because of its close association with those studying political topics, including economic. (This probably began the close association between political lying and statistical lying.) The first known use of this political arithmetic was by John Graunt, who used what is now called descriptive statistics to study London's death rates (1662). Although there is fierce debate concerning the original use of the term *statistics* (Yule, 1905), the greatest support argues that it was coined by Eberhard August Wilhelm von Zimmerman in the preface of *A Political Survey of the Present State of Europe* (1787). The modern use of the term *statistics* (as opposed to mathematics) is often attributed to R. A. Fisher and his work

Statistical Methods for Research Workers (1925), wherein he stated that “a statistic is a value calculated from an observed sample with a view to characterizing the population from which it is drawn.”

Since that time, statisticians have added to the techniques available to analyze data, many adding their names to the procedures; and the addition of statistical techniques continues today. Analysis procedures have been added to the statistical repertoire in the past few years that have greatly increased the ability of criminology, criminal justice, and other researchers to more accurately examine the relationship between variables.

A single death is a tragedy, a million deaths is a statistic.

—Joseph Stalin

USES OF STATISTICS

The term *statistics* is often misunderstood because there are actually two practical applications of it. The first, reflecting the history of the term, is a collection of data—often expressed in summary form—that is collected and preserved. The best example of these are census statistics or mortality statistics, which depict the characteristics of the living or the causes of death, respectively. The second application is the subject of this text: a method of analyzing data. Statistics as you will come to know them are methods used to examine the data collected in the process of scientific inquiry. These methods allow researchers to think logically about the data and to do one of two things: to come to some succinct and meaningful conclusions about the data (*descriptive statistics*), or to determine, or infer, characteristics of large groups based on the data collected on small parts (samples) of the group (*inferential statistics*). For example, data could be gathered on all correctional officers in Nebraska for a research project to determine the sex and race breakdown of the officers. This would be a descriptive analysis that could be used to examine the employment patterns for the Nebraska Department of Correction. Alternatively, a sample of correctional officers from each state could be collected and the data from the sample used to make statements about all correctional officers in the nation. This would be drawing conclusions (inferences) about a large group based on information about a sample of the group.

Statistical analysis is the workhorse of discovery and advancement. The scientific process, using research to test theory, requires that empirical evidence data drawn from the research subjects be examined systematically. The use of mathematics in general and statistical analysis in particular allows researchers to make these comparisons and to discover new information that will provide a better understanding of their subject.

In the scientific process, the purpose is usually to discover something that was previously unknown or to prove something true or false that was previously thought to be true but was never supported by hard evidence. The way to obtain that hard evidence is by gathering information (data) and subjecting it to statistical analysis.

THEORY CONSTRUCTION AT A GLANCE

Three elements in social science research, or any research for that matter, are essential to sound investigation. These are theory, research methods, and statistical analysis. Although these elements are intimately linked, there is debate—even among those most supportive of the research process—on their ordering, importance, and what should be included from each element in any particular textbook. It is not possible to adequately cover all of these elements in one course or in one textbook, so it becomes an issue of how much of each element should be included in a discussion of the other. In this text, theory will be

primarily covered in this chapter, research will span this chapter and several that follow, and statistical analysis will prevail thereafter.

What Is Theory?

At the most basic level, theory consists of statements concerning the relationship or association among *social phenomena* such as events and characteristics of people or things. For example, in criminology, there are theories that address how people learn to be criminal. In these theories, statements are constructed that deal with the role of peers in a person's learning criminal behavior, how the rewards from a crime can influence behavior, and what effect punishment can have on the decision to commit a crime.

The goal of these statements is to develop explanations of why things are as they appear and to try to explain their meaning. From an early age, humans have ideas about the causes of events and why things work the way that they do. The problem with these explanations, however, is that they are often too simplistic to be of any real value. Theory attempts to provide a stronger foundation for these ideas by asking questions about them, such as:

- What is the point of all of this?
- What does it mean?
- Why are things this way?

Without theory, there is often only conjecture and war stories. With theory, we may begin to develop statements or ideas that are based on sound observation and thought.

Theory and Research

Theory may be developed in several ways. Researchers may look at the world around them, find the social phenomena that pique their interest and begin to develop statements concerning why these phenomena work the way that they do. This is called *induction*. An example might be a researcher who follows crime trends in a city for a number of years. She may begin to see that the crimes follow a definite pattern of movement in the city—moving from east to west across the city. From this, she might set out to determine what the cause of this movement could be, ultimately developing a theory of crime movement in urban areas. This is a process of moving from data to theory and attempting to make sense of the data with the theory. It should be noted that Sherlock Holmes was not exactly correct in his understanding of the difference between induction and deduction. When Holmes made his famous statement, “brilliant deduction Watson!” he should have actually been commending Watson on his inductive reasoning. Watson was actually drawing conclusions based on what he had observed, not testing previously developed conclusions, as will be discussed below.

Alternatively, researchers may become curious about something and set out to develop statements and then to test them. This is called *deduction*. The process of deduction begins with an idea and attempt to test the idea with data and analysis. For example, a researcher might believe that increased supervision of probationers would prevent them from becoming involved in subsequent crimes. This researcher might create an experiment where a random sample of probationers are put under intensive supervision while another random sample receives a normal amount of supervision. The results of this experiment could either support or refute the researcher's initial beliefs. This is a process of moving from theory to data, where the data tests the theory.

Finally, and probably most often the case, a researcher may start with either induction or deduction, but, by the time a project is finished, the researcher has used both induction and deduction. This is called *retro-duction*. With this process the researcher investigating supervision of probationers might conduct the in-

tensive supervision experiment as a deductive process. After examining the data, however, it might be obvious that the experiment could be done better or that there was something in the data that needed further explanation for example, those probationers who received the most supervision were successful; while those who received intensive, but less than the most intensive, supervision were not successful. The researcher, then, might rethink part of the theory and set out to retest it. This process might continue until the theory was supported or disproven. This is a process of moving from theory to data to theory and so on; or data to theory to data and so on. The key here is that it is an alternating process between induction and deduction.

The process of scientific inquiry (using a deductive method) is shown in Figure 1-1. As shown in this diagram, theory is at the starting point of the process. Theory is driven by observations, and leads re-

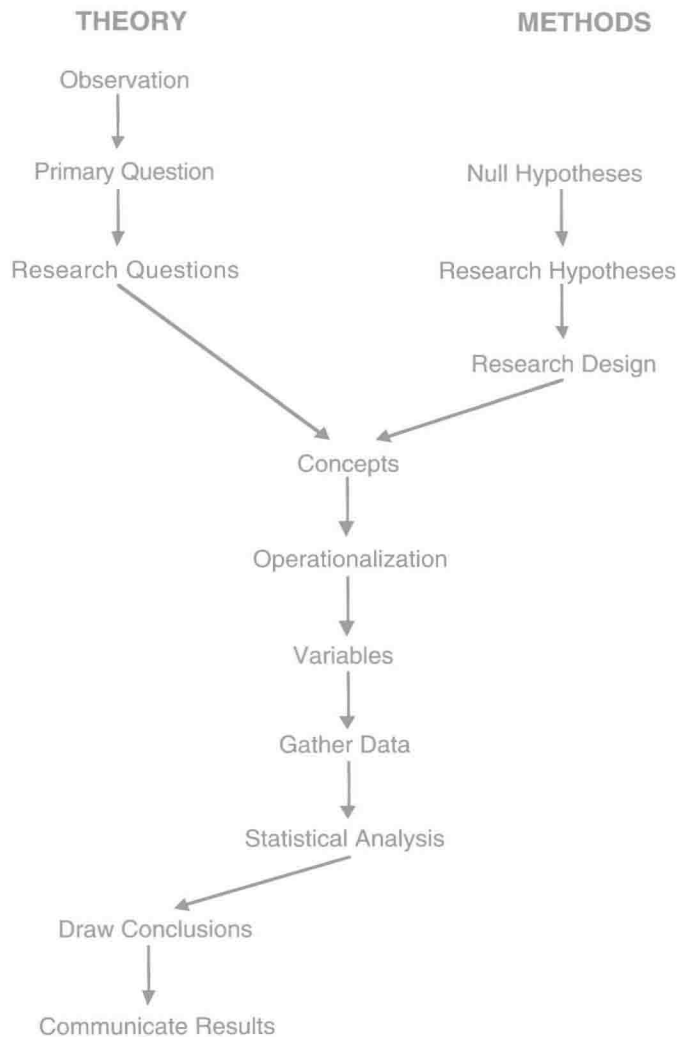


Figure 1-1 Process of scientific inquiry: Theory, research, and statistical analysis.

searchers to initiate the research process through primary questions and research questions. It is from this process of theory building that researchers follow the process from developing a null hypothesis to communicating results. The process of scientific inquiry and its individual parts are further discussed in the remainder of this chapter.

Observation and Inquisitiveness

The first step in the process of scientific inquiry, and one of the more important, is often overlooked: observation and inquisitiveness. Many research projects are never begun because the researcher was not aware of his or her surroundings or did not recognize something as a topic worthy of research.

It is often theory that stimulates observation and scientific inquiry. As you go through school and read research and material that you find interesting, sometimes you will think you have a better way to do something or what you read may stimulate you in other areas. By using a structured, scientific process to evaluate your observations and formulate statements of why these phenomena are behaving the way that they are, you are developing theory.

An example of inductive theory development can be shown in Robert Burgess' Zonal hypothesis. Students at the University of Chicago were making maps of Chicago showing different characteristics of neighborhoods, such as welfare, infant mortality, and housing. Burgess observed that these maps followed very similar patterns throughout the city. His observations led him to develop a theory about how cities grow and change. The theory he developed from these maps proposed that cities grow in rings similar to when a rock is thrown into the water. In this configuration, the rings closest to the center of the city will be the most run down, have the highest level of infant mortality, and be characterized by other social ills that are not present in the outer rings. All of this was developed simply by examining the students' maps developed and by using inductive theory building.

Primary Questions

A primary question is the one, driving thought behind a research project. It should represent the whole reason for the study. Primary questions are important because how well a researcher meets the goals of the primary question will often be the criteria by which the research will be evaluated. The primary question should be a carefully worded, specific phrase that states exactly the focus of the study.

For example, in research on police use of deadly force, a possible primary question might be, What factors most influence police use of deadly force?¹ This question is very broad, and somewhat vague, but it can easily represent the goal of a research project.

Research Questions

Often the primary question will be theoretical, vague, and quite possibly not directly addressable through research. Research questions break down the primary question into subproblems that are more manageable, and make the primary question testable through research. If the primary question establishes the goal of the research, the research questions suggest ways of achieving that goal.

In the example concerning police use of deadly force used above, some possible research questions might include the following:

- What is the relationship between an officer's shift and the likelihood of the officer using deadly force?
- What is the relationship between the violent crime rate of an area and police use of deadly force?

¹ This example will be used throughout this discussion to show the research process using a single example in each stage.

- What is the relationship between an officer's level of education and the likelihood of the officer using deadly force?

These research questions break down the primary question into smaller parts that can be more easily examined. The answers to these questions are derived from the research process and statistical analysis and allow the researcher to answer the primary question.

RESEARCH: MOVEMENT FROM THEORY TO DATA AND BACK

Theory cannot stand alone. Nor can research or statistics. Theory without research and statistical analyses to back it up is little more than fable. Research without theory is like building a house without plans; and research without statistical analysis is like building a house without nails: it is possible and it is done, but it would be more effective with them. Statistical analysis without theory and research methodology to guide the research is like having your own intercontinental ballistic missile—nice to have and may impress your neighbors, but it is not really useful.

Whatever the technique of developing theory, research is the method used to test and validate it. Research, in its purest form is a scientific, systematic study to discover new information or to test the validity of theories previously developed. The prime goal of research is discovery. Depending on whether an inductive or deductive process is used, research is a systematic way of turning statistical observation and analysis into theory (induction) or testing theory through statistical analysis, through deduction.

Although there are no exact, cookbook steps that must be followed in conducting a research project, there are some general guidelines that should be followed, either consciously or unconsciously, to ensure that nothing is left out of the study. These steps are included in this chapter and provide a kind of map of where research and statistics fits into the overall process of scientific inquiry.

Formulating Hypotheses

Once the research questions have been developed, you must decide what the research is attempting to determine. Hypotheses are questions or statements whose answers will support or refute the theoretical propositions of the research. Hypotheses are generally broken down into *research hypotheses* and *null hypotheses*. Although these will be covered in more detail in chapter 13, a brief definition is provided here.

A research hypothesis is generally a statement, similar to a research question, that states the expected outcome of a part of a research project. If a research question in a project asked "What is the relationship between an officer's shift and the likelihood of using deadly force?" the research hypothesis might be that there is a statistically significant correlation between an officer's shift and the likelihood of using deadly force. In using research hypotheses, the researcher turns the relatively abstract wording of theory development into a more concrete, testable form appropriate for statistical analysis.

One of the often difficult to understand but vital elements of statistical analysis and hypothesis testing is that research alone cannot prove anything. Even when researchers find a great deal of support for an association between two variables, it may be that these results are occurring because information is missing or the model is somehow flawed. Other researchers may very well be able to disprove these findings by conducting additional research. If research cannot prove anything, then what can it do? It can be used to disprove something or eliminate alternatives. For example, even though research cannot prove that officers on night shifts use deadly force more than their daytime counterparts, it may disprove that there is no relationship between shift and use of deadly force. This is accomplished with a null hypothesis. A null hypothesis generally states one of the following:

- There is no statistically significant difference between the groups being compared.