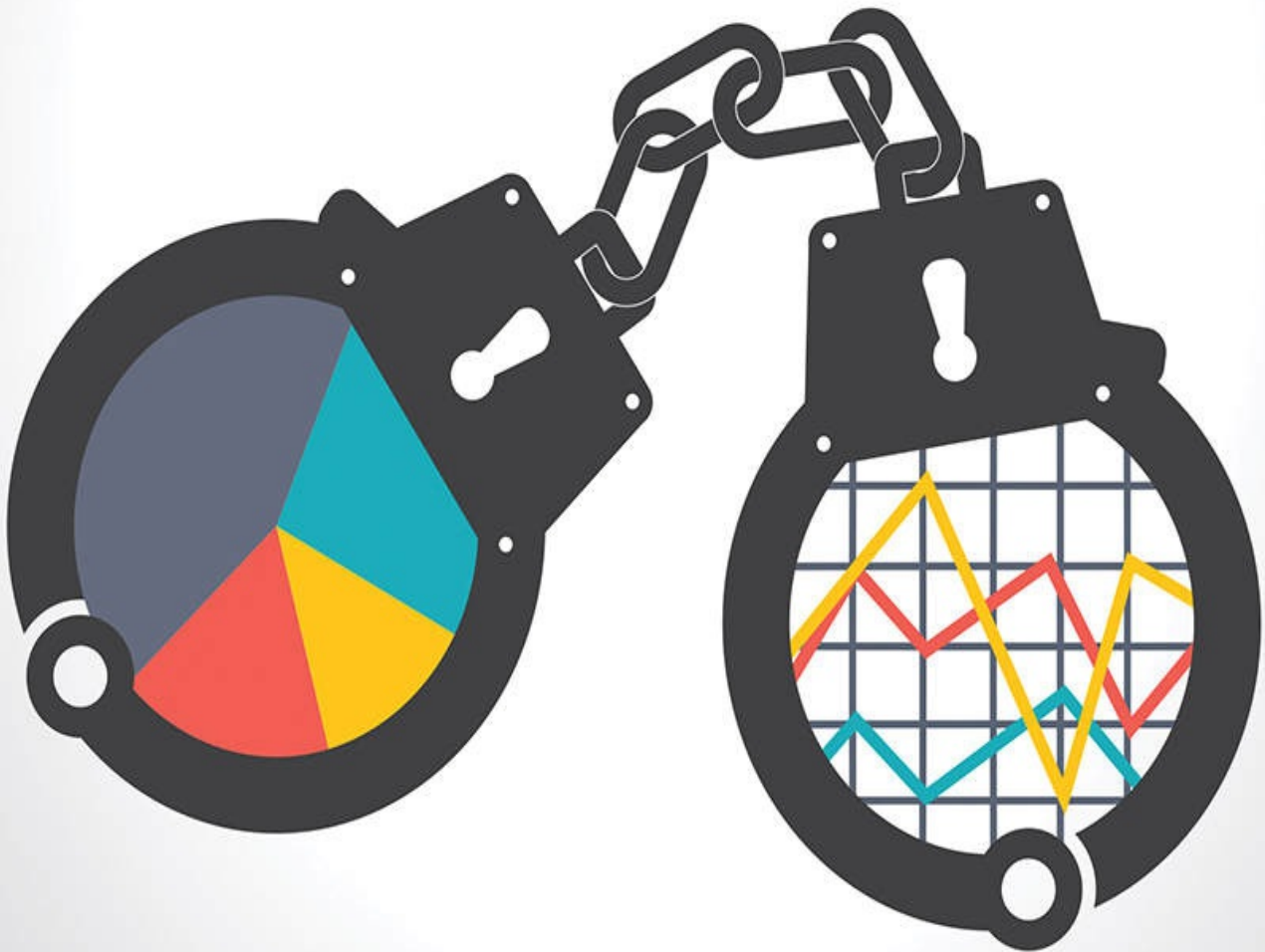# STATISTICS
FOR CRIMINOLOGY
AND CRIMINAL JUSTICE

Third Edition

JACINTA M. GAU

Statistics for Criminology and Criminal Justice

Third Edition

Statistics for Criminology and Criminal Justice

Third Edition

Jacinta M. Gau
University of Central Florida

Brief Contents

Detailed Contents

Preface to the Third Edition

In 2002, James Comey, the newly appointed U.S. attorney for the Southern District of New York who would later become the director of the Federal Bureau of Investigation, entered a room filled with high-powered criminal prosecutors. He asked the members of the group to raise their hands if they had never lost a case. Proud, eager prosecutors across the room threw their hands into the air, expecting a pat on the back. Comey's response befuddled them. Instead of praising them, he called them chickens (that is not quite the term he used, but close enough) and told them the only reason they had never lost is that the cases they selected to prosecute were too easy.[1] The group was startled at the rebuke, but they really should not have been. Numbers can take on various meanings and interpretations and are sometimes used in ways that conceal useful information rather than revealing it.

[1.] Eisinger, J. (2017). *The chickens**t club: Why the Justice Department fails to prosecute executives.* New York: Simon & Schuster.

This book enters its third edition at a time when the demand for an educated, knowledgeable workforce has never been greater. This is as true in criminal justice and criminology as in any other university major and occupational field. Education is the hallmark of a professional. Education is not just about knowing facts, though—it is about thinking critically and treating incoming information with a healthy dose of skepticism. All information must pass certain tests before being treated as true. Even if it passes those tests, the possibility remains that additional information exists that, if discovered, would alter our understanding of the world. People who critically examine the trustworthiness of information and are open to new knowledge that challenges their preexisting notions about what is true and false are actively *using* their education, rather than merely possessing it.

At first glance, statistics seems like a topic of dubious relevance to everyday life. Convincing criminology and criminal justice students that they should care about statistics is no small task. Most students approach the class with apprehension because math is daunting, but many also express frustration and impatience. The thought, "But I'm going to be a [police officer, lawyer, federal agent, etc.], so what do I need *this* class for?" is on many students' minds as they walk through the door or log in to the learning management system on the first day. The answer is surprisingly simple: Statistics form a fundamental part of what we know about the world. Practitioners in the criminal justice field rely on statistics. A police chief who alters a department's deployment plan so as to allocate resources to crime hot spots trusts that the researchers who analyzed the spatial distribution of crime did so correctly. A prison warden seeking to classify inmates according to the risk they pose to staff and other inmates needs assessment instruments that accurately predict each person's likelihood of engaging in behavior that threatens institutional security. A chief prosecutor must recognize that a high conviction rate might not be testament to assistant prosecutors' skill level but, rather, evidence that they only try simple cases and never take on challenges.

Statistics matter because what unites all practitioners in the criminology and criminal justice occupations and

professions is the need for valid, reliable data and the ability to critically examine numbers that are set before them. Students with aspirations for graduate school have to understand statistical concepts because they will be expected to produce knowledge using these techniques. Those planning to enter the workforce as practitioners must be equipped with the background necessary to appraise incoming information and evaluate its accuracy and usefulness. Statistics, therefore, is just as important to information consumers as it is to producers.

The third edition of *Statistics for Criminology and Criminal Justice*, like its two predecessors, balances quantity and complexity with user-friendliness. A book that skimps on information can be as confusing as one overloaded with it. The sacrificed details frequently pertain to the underlying theory and logic that drive statistical analyses. The pedagogical techniques employed in this text draw from the scholarship of teaching and learning, wherein researchers have demonstrated that students learn best when they understand logical connections within and across concepts, rather than merely memorizing key terms or steps to solving equations. In statistics, students are at an advantage if they first understand the overarching goal of the techniques they are learning before they begin working with formulas and numbers.

This book also emphasizes the application of new knowledge. Students can follow along in the step-by-step instructions that illustrate plugging numbers into formulas and solving them. Additional practice examples are embedded within the chapters, and chapter review problems allow students to test themselves (the answers to the odd-numbered problems are located in the back of the book), as well as offering instructors convenient homework templates using the even-numbered questions.

Real data and research also further the goal of encouraging students to apply concepts and showing them the relevance of statistics to practical problems in the criminal justice and criminology fields. Chapters contain *Data Sources* boxes that describe some common, publicly available data sets such as the Uniform Crime Reports, National Crime Victimization Survey, General Social Survey, and others. Most in-text examples and end-of-chapter review problems use data drawn from the sources highlighted in the book. The goal is to lend a practical, tangible bent to this often-abstract topic. Students get to work with the data their professors use. They get to see how elegant statistics can be at times and how messy they can be at others, how analyses can sometimes lead to clear conclusions and other times to ambiguity.

The *Research Example* boxes embedded throughout the chapters illustrate criminal justice and criminology research in action and are meant to stimulate students' interest. They highlight that even though the math might not be exciting, the act of scientific inquiry most definitely is, and the results have important implications for policy and practice. In the third edition, the examples have been expanded to include additional contemporary criminal justice and criminology studies. Most of the examples contained in the first and second editions were retained in order to enhance diversity and allow students to see firsthand the rich variety of research that has been taking place over time. The full texts of all articles are available on the SAGE companion site (http://www.sagepub.com/gau) and can be downloaded online by users with institutional access to the SAGE journals in which the articles appear.

This edition retains the *Learning Check* boxes. These are scattered throughout the text and function as mini-

quizzes that test students' comprehension of certain concepts. They are short so that students can complete them without disrupting their learning process. Students can use each *Learning Check* to make sure they are on the right track in their understanding of the material, and instructors can use them for in-class discussion. The answer key is in the back of the book.

Where relevant to the subject matter, chapters end with a section on IBM® SPSS® Statistics[2] and come with one or more shortened versions of a major data set in SPSS file format. Students can download these data sets to answer the review questions presented at the end of the chapter. The full data sets are all available from the Inter-University Consortium for Political and Social Research at www.icpsr.umich.edu/icpsrweb/ICPSR/ and other websites as reported in the text. If desired, instructors can download the original data sets to create supplementary examples and practice problems for hand calculations or SPSS analyses.

[2] SPSS is a registered trademark of International Business Machines Corporation.

The third edition features the debut of *Thinking Critically* sections. These two-question sections appear at the end of each chapter. The questions are open-ended and designed to inspire students to think about the nuances of science and statistics. Instructors can assign them as homework problems or use them to initiate class debates.

The book is presented in three parts. Part I covers descriptive statistics. It starts with the basics of levels of measurement and moves on to frequency distributions, graphs and charts, and proportions and percentages. Students learn how to select the correct type(s) of data display based on a variable's level of measurement and then construct that diagram or table. They then learn about measures of central tendency and measures of dispersion and variability. These chapters also introduce the normal curve.

Part II focuses on probability theory and sampling distributions. This part lays out the logic that forms the basis of hypothesis testing. It emphasizes the variability in sample statistics that precludes direct inference to population parameters. Part II ends with confidence intervals, which is students' first foray into inferential statistics.

Part III begins with an introduction to bivariate hypothesis testing. The intention is to ease students into inferential tests by explaining what these tests do and what they are for. This helps transition students from the theoretical concepts covered in Part II to the application of those logical principles. The remaining chapters include chi-square tests, *t* tests and tests for differences between proportions, analysis of variance (ANOVA), correlation, and ordinary least squares (OLS) regression. The sequence is designed such that some topics flow logically into others. Chi-square tests are presented first because they are the only nonparametric test type covered here. Two-population *t* tests then segue into ANOVA. Correlation, likewise, supplies the groundwork for regression. Bivariate regression advances from correlation and transitions into the multivariate framework. The book ends with the fundamentals of interpreting OLS regression models.

This book provides the foundation for a successful statistics course that combines theory, research, and practical application for a holistic, effective approach to teaching and learning. Students will exit the course ready to put their education into action as they prepare to enter their chosen occupation, be that in academia,

law, or the field. Learning statistics is not a painless process, but the hardest classes are the ones with the greatest potential to leave lasting impressions. Students will meet obstacles, struggle with them, and ultimately surmount them so that in the end, they will look back and say that the challenge was worth it.

Acknowledgments

About the Author

Jacinta M. Gau, Ph.D.,

is an associate professor in the Department of Criminal Justice at the University of Central Florida. She received her Ph.D. from Washington State University. Her primary areas of research are policing and criminal justice policy, and she has a strong quantitative background. Dr. Gau's work has appeared in journals such as *Justice Quarterly*, *British Journal of Criminology*, *Criminal Justice and Behavior, Crime & Delinquency, Criminology & Public Policy, Police Quarterly, Policing: An International Journal of Police Strategies & Management*, and the *Journal of Criminal Justice Education*. In addition to *Statistics for Criminology and Criminal Justice*, she is author of *Criminal Justice Policy: Origins and Effectiveness* (Oxford University Press) and coauthor of *Key Ideas in Criminology and Criminal Justice* (SAGE). Additionally, she coedits *Race and Justice: An International Journal*, published by SAGE.

Part I Descriptive Statistics

Chapter 1 Introduction to the Use of Statistics in Criminal Justice and Criminology

Learning Objectives

- Explain how data collected using scientific methods are different from anecdotes and other nonscientific information.
- List and describe the types of research in criminal justice and criminology.
- Explain the difference between the research methods and statistical analysis.
- Define samples and populations.
- Describe probability sampling.
- List and describe the three major statistics software packages.

You might be thinking, "What do statistics have to do with criminal justice or criminology?" It is reasonable for you to question the requirement that you spend an entire term poring over a book about statistics instead of one about policing, courts, corrections, or criminological theory. Many criminology and criminal justice undergraduates wonder, "Why am I here?" In this context, the question is not so much existential as it is practical. Luckily, the answer is equally practical.

You are "here" (in a statistics course) because the answer to the question of what statistics have to do with criminal justice and criminology is "Everything!" Statistical methods are the backbone of criminal justice and criminology as fields of scientific inquiry. Statistics enable the construction and expansion of knowledge about criminality and the criminal justice system. Research that tests theories or examines criminal justice phenomena and is published in academic journals and books is the basis for most of what we know about criminal offending and the system that has been designed to deal with it. The majority of this research would not be possible without statistics.

Statistics can be abstract, so this book uses two techniques to add a realistic, pragmatic dimension to the subject. The first technique is the use of examples of statistics in criminal justice and criminology research. These summaries are contained in the *Research Example* boxes embedded in each chapter. They are meant to give you a glimpse into the types of questions that are asked in this field of research and the ways in which specific statistical techniques are used to answer those questions. You will see firsthand how lively and diverse criminal justice and criminology research is. Research Example 1.1 summarizes seven studies. Take a moment now to read through them.

The second technique to add a realistic, pragmatic dimension to the subject of this book is the use of real data from reputable and widely used sources such as the Bureau of Justice Statistics (BJS). The BJS is housed within the U.S. Department of Justice and is responsible for gathering, maintaining, and analyzing data on various criminal justice topics at the county, state, and national levels. Visit http://bjs.ojp.usdoj.gov/ to familiarize yourself with the BJS. The purpose behind the use of real data is to give you the type of hands-on experience that you cannot get from fictional numbers. You will come away from this book having worked with some of the same data that criminal justice and criminology researchers use. Two sources of data that will be used in upcoming chapters are the Uniform Crime Reports (UCR) and the National Crime Victimization Survey (NCVS). See Data Sources 1.1 and 1.2 for information about these commonly used measures of criminal incidents and victimization, respectively. All the data sets used in this book are publicly available and

were downloaded from governmental websites and the archive maintained by the Inter-University Consortium for Political and Social Research at www.icpsr.umich.edu.

Research Example 1.1 What Do Criminal Justice and Criminology Researchers Study?

Researchers in the field of criminology and criminal justice examine a wide variety of issues pertaining to the criminal justice system and theories of offending. Included are topics such as prosecutorial charging decisions, racial and gender disparities in sentencing, police use of force, drug and domestic violence courts, and recidivism. The following are examples of studies that have been conducted and published. You can find the full text of each of these articles and of all those presented in the following chapters at (**www.sagepub.com/gau**).

1. *Can an anticrime strategy that has been effective at reducing certain types of violence also be used to combat open-air drug markets?* The "pulling levers" approach involves deterring repeat offenders from crime by targeting them for enhanced prosecution while also encouraging them to change their behavior by offering them access to social services. This strategy has been shown to hold promise with gang members and others at risk for committing violence. The Rockford (Illinois) Police Department (RPD) decided to find out if they could use a pulling levers approach to tackle open-air drug markets and the crime problems caused by these nuisance areas. After the RPD implemented the pulling levers intervention, Corsaro, Brunson, and McGarrell (2013) used official crime data from before and after the intervention to determine whether this approach had been effective. They found that although there was no reduction in violent crime, nonviolent crime (e.g., drug offenses, vandalism, and disorderly conduct) declined noticeably after the intervention. This indicated that the RPD's efforts had worked, because drug and disorder offenses were exactly what the police were trying to reduce.

2. *Are prisoners with low self-control at heightened risk of victimizing, or being victimized by, other inmates?* Research has consistently shown that low self-control is related to criminal offending. Some studies have also indicated that this trait is a risk factor for victimization, in that people with low self-control might place themselves in dangerous situations. One of the central tenets of this theory is that self-control is stable and acts in a uniform manner regardless of context. Kerley, Hochstetler, and Copes (2009) tested this theory by examining whether the link between self-control and both offending and victimization held true within the prison environment. Using data gathered from surveys of prison inmates, the researchers discovered that low self-control was only slightly related to in-prison offending and victimization. This result could challenge the assumption that low self-control operates uniformly in all contexts. To the contrary, something about prisoners themselves, the prison environment, or the interaction between the two might change the dynamics of low self-control.

3. *Does school racial composition affect how severely schools punish black and Latino students relative to white ones?* Debates about the so-called school-to-prison pipeline emphasize the long-term effects of school disciplinary actions such as suspension, expulsion, and arrest or court referral. Youth who experience these negative outcomes are at elevated risk for dropping out of school and getting involved in delinquency and, eventually, crime. Hughes, Warren, Stewart, Tomaskovic-Devey, and Mears (2017) set out to discover whether schools' and school boards' racial composition affects the treatment of black, Latino, and white students. The researchers drew from two theoretical perspectives: The racial threat perspective argues that minorities are at higher risk for punitive sanctions when minority populations are higher, because whites could perceive minority groups as a threat to their place in society. On the other hand, the intergroup contact perspective suggests that racial and ethnic diversity reduces the harshness of sanctions for minorities, because having contact with members of other racial and ethnic groups diminishes prejudice. Hughes and colleagues used data from the Florida Department of Education, the U.S. Census Bureau, and the Uniform Crime Reports. Statistical results provided support for both perspectives. Increases in the size of the black and Hispanic student populations led to higher rates of suspension for students of these groups. On the other hand, intergroup contact among school board members of different races reduced suspensions for all students. The researchers concluded that interracial contact among school board members equalized disciplinary practices and reduced discriminatory disciplinary practices.

4. *What factors influence police agencies' ability to identify and investigate human trafficking?* Human trafficking has been recognized as a transnational crisis. Frequently, local police are the first ones who encounter victims or notice signs suggesting the presence of trafficking. In the United States, however, many local police agencies do not devote systematic attention to methods that would enable them to detect and investigate suspected traffickers. Farrell (2014) sought to learn more about U.S. police agencies' antitrafficking efforts. Using data from two national surveys of medium-to-large municipal police departments, Farrell found that 40% of departments trained their personnel on human trafficking, 17% had written policies pertaining to this crime, and 13% dedicated personnel to it. Twenty-eight percent had investigated at least one

5. *How safe and effective are conducted energy devices as used by police officers?* Conducted energy devices (CEDs) have proliferated in recent years. Their widespread use and the occasional high-profile instances of misuse have generated controversy over whether these devices are safe for suspects and officers alike. Paoline, Terrill, and Ingram (2012) collected use-of-force data from six police agencies nationwide and attempted to determine whether officers who deployed CEDs against suspects were more or less likely to sustain injuries themselves. The authors' statistical analysis suggested a lower probability of officer injury when only CEDs were used. When CEDs were used in combination with other forms of force, however, the probability of officer injury increased. The results suggest that CEDs can enhance officer safety, but they are not a panacea that uniformly protects officers in all situations.

6. *How prevalent is victim precipitation in intimate partner violence?* A substantial number of violent crimes are initiated by the person who ultimately becomes the victim in an incident. Muftić, Bouffard, and Bouffard (2007) explored the role of victim precipitation in instances of intimate partner violence (IPV). They gleaned data from IPV arrest reports and found that victim precipitation was present in cases of both male and female arrestees but that it was slightly more common in instances where the woman was the one arrested. This suggests that some women (and, indeed, some men) arrested for IPV might be responding to violence initiated by their partners rather than themselves being the original aggressors. The researchers also discovered that victim precipitation was a large driving force behind dual arrests (cases in which both parties are arrested), because police could either see clearly that both parties were at fault or, alternatively, were unable to determine which party was the primary aggressor. Victim precipitation and the use of dual arrests, then, could be contributing factors behind the recent rise in the number of women arrested for IPV against male partners.

7. *What are the risk factors in a confrontational arrest that are most commonly associated with the death of the suspect?* There have been several high-profile instances of suspects dying during physical confrontations with police wherein the officers deployed CEDs against these suspects. White and colleagues (2013) collected data on arrest-related deaths (ARDs) that involved CEDs and gained media attention. The researchers triangulated the data using information from medical-examiner reports. They found that in ARDs, suspects were often intoxicated and extremely physically combative with police. Officers, for their part, had used several other types of force before or after trying to solve the situation using CEDs. Medical examiners most frequently attributed these deaths to drugs, heart problems, and excited delirium. These results suggest that police departments should craft policies to guide officers' use of CEDs against suspects who are physically and mentally incapacitated.

In this book, emphasis is placed on both the production and interpretation of statistics. Every statistical analysis has a producer (someone who runs the analysis) and a consumer (someone to whom an analysis is being presented). Regardless of which role you play in any given situation, it is vital for you to be sufficiently versed in quantitative methods that you can identify the proper statistical technique and correctly interpret the results. When you are in the consumer role, you must also be ready to question the methods used by the producer so that you can determine for yourself how trustworthy the results are. Critical thinking skills are an enormous component of statistics. You are not a blank slate standing idly by, waiting to be written on—you are an active agent in your acquisition of knowledge about criminal justice, criminology, and the world in general. Be critical, be skeptical, and never hesitate to ask for more information.

Data Sources 1.1 The Uniform Crime Reports

The Federal Bureau of Investigation (FBI) collects annual data on crimes reported to police agencies nationwide and maintains the UCR. Crimes are sorted into eight index offenses: homicide, rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson. An important aspect of this data set is that it includes only those crimes that come to the attention of police—crimes that are not reported or otherwise detected by police are not counted. The UCR also conforms to the hierarchy rule, which

mandates that in multiple-crime incidents only the most serious offense ends up in the UCR. If, for example, someone breaks into a residence with intent to commit a crime inside the dwelling and while there, kills the homeowner and then sets fire to the structure to hide the crime, he has committed burglary, murder, and arson. Because of the hierarchy rule, though, only the murder would be reported to the FBI—it would be as if the burglary and arson had never occurred. Because of underreporting by victims and the hierarchy rule, the UCR undercounts the amount of crime in the United States. It nonetheless offers valuable information and is widely used. You can explore this data source at www.fbi.gov/about-us/cjis/ucr/ucr.

Data Sources 1.2 The National Crime Victimization Survey

The U.S. Census Bureau conducts the periodic NCVS under the auspices of the BJS to estimate the number of criminal incidents that transpire each year and to collect information about crime victims. Multistage cluster sampling is used to select a random sample of households, and each member of that household who is 12 years or older is asked to participate in an interview. Those who agree to be interviewed are asked over the phone or in person about any and all criminal victimizations that transpired in the 6 months prior to the interview. The survey employs a rotating panel design, so respondents are called at 6-month intervals for a total of 3 years, and then new respondents are selected (BJS, 2006). The benefit of the NCVS over the UCR is that NCVS respondents might disclose victimizations to interviewers that they did not report to police, thus making the NCVS a better estimation of the total volume of crime in the United States. The NCVS, though, suffers from the weakness of being based entirely on victims' memory and honesty about the timing and circumstances surrounding criminal incidents. The NCVS also excludes children younger than 12 years, institutionalized populations (e.g., persons in prisons, nursing homes, and hospitals), and the homeless. Despite these problems, the NCVS is useful because it facilitates research into the characteristics of crime victims. The 2015 wave of the NCVS is the most recent version available as of this writing.

Science: Basic Terms and Concepts

There are a few terms and concepts that you must know before you get into the substance of the book. Statistics are a tool in the larger enterprise of scientific inquiry. **Science** is the process of systematically collecting reliable information and developing knowledge using techniques and procedures that are accepted by other scientists in a discipline. Science is grounded in **methods**—research results are trustworthy only when the procedures used to arrive at them are considered correct by experts in the scientific community. Nonscientific information is that which is collected informally or without regard for correct methods. Anecdotes are a form of nonscientific information. If you ask one person why he or she committed a crime, that person's response will be an anecdote; it cannot be assumed to be broadly true of other offenders. If you use scientific methods to gather a large group of offenders and you survey all of them about their motivations, you will have data that you can analyze using statistics and that can be used to draw general conclusions.

> **Science**: The process of gathering and analyzing data in a systematic and controlled way using procedures that are generally accepted by others in the discipline.
>
> **Methods**: The procedures used to gather and analyze scientific data.

In scientific research, **samples** are drawn from **populations** using scientific techniques designed to ensure that samples are representative of populations. For instance, if the population is 50% male, then the sample should also be approximately 50% male. A sample that is only 15% male is not representative of the population. Research-methods courses instruct students on the proper ways to gather representative samples. In a statistics course, the focus is on techniques used to analyze the data to look for patterns and test for relationships. Together, proper methods of gathering and analyzing data form the groundwork for scientific inquiry. If there is a flaw in either the gathering or the analyzing of data, then the results might not be trustworthy. Garbage in, garbage out (GIGO) is the mantra of statistics. Data gathered with the best of methods can be rendered worthless if the wrong statistical analysis is applied to them; likewise, the most sophisticated, cutting-edge statistical technique cannot salvage improperly collected data. When the data or the statistics are defective, the results are likewise deficient and cannot be trusted. Studies using unscientific data or flawed statistical analyses do not contribute to theory and research or to policy and practice because their findings are unreliable and could be erroneous.

> **Sample**: A subset pulled from a population with the goal of ultimately using the people, objects, or places in the sample as a way to generalize to the population.
>
> **Population**: The universe of people, objects, or locations that researchers wish to study. These groups are often very large.

Learning Check 1.1

Identify whether each of the following is a sample or a population.

1. A group of 100 police officers pulled from a department with 300 total officers
2. Fifty prisons selected at random from all prisons nationwide
3. All persons residing in the state of Wisconsin
4. A selection of 10% of the defendants processed through a local criminal court in 1 year

Everybody who conducts a study has an obligation to be clear and open about the methods they used. You should expect detailed reports on the procedures used so that you can evaluate whether they followed proper scientific methods. When the methods used to collect and analyze data are sound, it is not appropriate to question scientific results on the basis of a moral, emotional, or opinionated objection to them. On the other hand, it is entirely correct (and is necessary, in fact) to question results when methodological or statistical procedures are shoddy or inadequate. Remember GIGO!

A key aspect of science is the importance of **replication**. No single study ever proves something definitively; quite to the contrary, much testing must be done before firm conclusions can be drawn. Replication is important because there are times when a study is flawed and needs to be redone or when the original study is methodologically sound but needs to be tested on new populations and samples. For example, a correctional treatment program that reduces recidivism rates among adults might or might not have similar positive results with juveniles. Replicating the treatment and evaluation with a sample of juvenile offenders would provide information about whether the program is helpful to both adults and juveniles or is only appropriate for adults. The scientific method's requirement that all researchers divulge the steps they took to gather and analyze data allows other researchers and members of the public to examine those steps and, when warranted, to undertake replications.

Replication: The repetition of a particular study that is conducted for purposes of determining whether the original study's results hold when new samples or measures are employed.

Types of Scientific Research in Criminal Justice and Criminology

Criminal justice and criminology research is diverse in nature and purpose. Much of it involves theory testing. **Theories** are proposed explanations for certain events. **Hypotheses** are small "pieces" of theories that must be true in order for the entire theory to hold up. You can think of a theory as a chain and hypotheses as the links forming that chain. Research Example 1.1 discusses a test of the general theory of crime conducted by Kerley et al. (2009). The general theory holds that low self-control is a static predictor of offending and victimization, regardless of context. From this proposition, the researchers deduced the hypothesis that the relationship between low self-control and both offending and victimization must hold true in the prison environment. Their results showed an overall lack of support for the hypothesis that low self-control operates uniformly in all contexts, thus calling that aspect of the general theory of crime into question. This is an example of a study designed to test a theory.

> **Theory**: A set of proposed and testable explanations about reality that are bound together by logic and evidence.
>
> **Hypothesis**: A single proposition, deduced from a theory, that must hold true in order for the theory itself to be considered valid.

Evaluation research is also common in criminal justice and criminology. In Research Example 1.1, the article by Corsaro et al. (2013) is an example of evaluation research. This type of study is undertaken when a new policy, program, or intervention is put into place and researchers want to know whether the intervention accomplished its intended purpose. In this study, the RPD implemented a pulling levers approach to combat drug and nuisance offending. After the program had been put into place, the researchers analyzed crime data to find out whether the approach was effective.

> **Evaluation research**: Studies intended to assess the results of programs or interventions for purposes of discovering whether those programs or interventions appear to be effective.

Exploratory research occurs when there is limited knowledge about a certain phenomenon; researchers essentially embark into unfamiliar territory when they attempt to study this social event. The study by Muftić et al. (2007) in Research Example 1.1 was exploratory in nature because so little is known about victim precipitation, particularly in the realm of IPV. It is often dangerous to venture into new areas of study when the theoretical guidance is spotty; however, exploratory studies have the potential to open new areas of research that have been neglected but that provide rich information that expands the overall body of knowledge.

> **Exploratory research**: Studies that address issues that have not been examined much or at all in prior research and that therefore might lack firm theoretical and empirical grounding.

Finally, some research is **descriptive** in nature. White et al.'s (2013) analysis of CED-involved deaths illustrates a descriptive study. White and colleagues did not set out to test a theory or to explore a new area of research—they merely offered basic descriptive information about the suspects, officers, and situations involved in instances where CED use was associated with a suspect's death. In descriptive research, no

generalizations are made to larger groups; the conclusions drawn from these studies are specific to the objects, events, or people being analyzed. This type of research can be very informative when knowledge about a particular phenomenon is scant.

> **Descriptive research**: Studies done solely for the purpose of describing a particular phenomenon as it occurs in a sample.

With the exception of purely descriptive research, the ultimate goal in most statistical analyses is to generalize from a sample to a population. A population is the entire set of people, places, or objects that a researcher wishes to study. Populations, though, are usually very large. Consider, for instance, a researcher trying to estimate attitudes about capital punishment in the general U.S. population. That is a population of more than 300 million! It would be impossible to measure everyone directly. Researchers thus draw samples from populations and study the samples instead. **Probability sampling** helps ensure that a sample mirrors the population from which it was drawn (e.g., a sample of people should contain a breakdown of race, gender, and age similar to that found in the population). Samples are smaller than populations, and researchers are therefore able to measure and analyze them. The results found in the sample are then generalized to the population.

> **Probability sampling**: A sampling technique in which all people, objects, or areas in a population have a known chance of being selected into the sample.

Learning Check 1.2

For each of the following scenarios, identify the type of research being conducted.

1. A researcher wants to know more about female serial killers. He gathers news articles that report on female serial killers and records information about each killer's life history and the type of victim she preyed on.
2. A researcher wants to know whether a new in-prison treatment program is effective at reducing recidivism. She collects a sample of inmates that participated in the program and a sample that did not go through the program. She then gathers recidivism data for each group to see if those who participated had lower recidivism rates than those who did not.
3. The theory of collective efficacy predicts that social ties between neighbors, coupled with neighbors' willingness to intervene when a disorderly or criminal event occurs in the area, protect the area from violent crime. A researcher gathers a sample of neighborhoods and records the level of collective efficacy and violent crime in each one to determine whether those with higher collective efficacy have lower crime rates.
4. A researcher notes that relatively little research has been conducted on the possible effects of military service on later crime commission. She collects a sample of people who served in the military and a sample of people that did not and compares them to determine whether the military group differs from the nonmilitary group in terms of the numbers or types of crimes committed.

Software Packages for Statistical Analysis

Hand computations are the foundation of this book because seeing the numbers and working with the formulas facilitates an understanding of statistical analyses. In the real world, however, statistical analysis is generally conducted using a software program. Microsoft Excel contains some rudimentary statistical functions and is commonly used in situations requiring only basic descriptive analyses; however, this program's usefulness is exhausted quickly because researchers usually want far more than descriptives. Many statistical packages are available. The most common in criminal justice and criminology research are SPSS, Stata, and SAS. Each of these packages has strengths and weaknesses. Simplicity and ease of use makes SPSS a good place to start for people new to statistical analysis. Stata is a powerful program excellent for regression modeling. The SAS package is the best one for extremely large data sets.

This book incorporates SPSS into each chapter. This allows you to get a sense for what data look like when displayed in their raw format and permits you to run particular analyses and read and interpret program output. Where relevant, the chapters offer SPSS practice problems and accompanying data sets that are available for download from www.sagepub.com/gau. This offers a practical, hands-on lesson about the way that criminal justice and criminology researchers use statistics.

Organization of the Book

This book is divided into three parts. Part I (Chapters 1 through 5) covers descriptive statistics. Chapter 2 provides a basic overview of types of variables and levels of measurement. Some of this material will be review for students who have taken a methods course. Chapter 3 delves into charts and graphs as means of graphically displaying data. Measures of central tendency are the topic of Chapter 4. These are descriptive statistics that let you get a feel for where the data are clustered. Chapter 5 discusses measures of dispersion. Measures of dispersion complement measures of central tendency by offering information about whether the data tend to cluster tightly around the center or, conversely, whether they are very spread out.

Part II (Chapters 6 through 8) describes the theoretical basis for statistics in criminal justice and criminology: probability and probability distributions. Part I of the book can be thought of as the nuts-and-bolts of the mathematical concepts used in statistics, and Part II can be seen as the theory behind the math. Chapter 6 introduces probability theory. Binomial and continuous probability distributions are discussed. In Chapter 7, you will learn about population, sample, and sampling distributions. Chapter 8 provides the book's first introduction to inferential statistics with its coverage of point estimates and confidence intervals. The introduction of inferential statistics at this juncture is designed to help ease you into Part III.

Part III (Chapters 9 through 14) of the book merges the concepts learned in Parts I and II to form the discussion on inferential hypothesis testing. Chapter 9 offers a conceptual introduction to this framework, including a description of the five steps of hypothesis testing that will be used in every proceeding chapter. In Chapter 10, you will encounter your first bivariate statistical technique: chi-square. Chapter 11 describes two-population t tests and tests for differences between proportions. Chapter 12 covers analysis of variance, which is an extension of the two-population t test. In Chapter 13, you will learn about correlations. Finally, Chapter 14 wraps up the book with an introduction to bivariate and multiple regression.

The prerequisite that is indispensable to success in this course is a solid background in algebra. You absolutely must be comfortable with basic techniques such as adding, subtracting, multiplying, and dividing. You also need to understand the difference between positive and negative numbers. You will be required to plug numbers into equations and solve those equations. You should not have a problem with this as long as you remember the lessons you learned in your high school and college algebra courses. Appendix A offers an overview of the basic mathematical techniques you will need to know, so look those over and make sure that you are ready to take this course. If necessary, use them to brush up on your skills.

Statistics are cumulative in that many of the concepts you learn at the beginning form the building blocks for more-complex techniques that you will learn about as the course progresses. Means, proportions, and standard deviations, for instance, are concepts you will learn about in Part I, but they will remain relevant throughout the remainder of the book. You must, therefore, learn these fundamental calculations well and you must remember them.

Repetition is the key to learning statistics. Practice, practice, practice! There is no substitute for doing and

redoing the end-of-chapter review problems and any other problems your instructor might provide. You can also use the in-text examples as problems if you just copy down the numbers and do the calculations on your own without looking at the book. Remember, even the most advanced statisticians started off knowing nothing about statistics. Everyone has to go through the learning process. You will complete this process successfully as long as you have basic algebra skills and are willing to put in the time and effort it takes to succeed.

Thinking Critically

1. Media outlets and other agencies frequently conduct opinion polls to try to capture information about the public's thoughts on contemporary events, controversies, or political candidates. Poll data are faster and easier to collect than survey data are, because they do not require adherence to scientific sampling methods and questionnaire design. Agencies conducting polls often do not have the time or resources to engage in full-scale survey projects. Debate the merits of poll data from a policy standpoint. Is having low-quality information better than having none at all? Or is there no place in public discussions for data that fall short of scientific standards? Explain your answer.

2. Suppose you tell a friend that you are taking a statistics course, and your friend reacts with surprise that a criminology or criminal justice degree program would require students to take this class. Your friend argues that although it is necessary for people whose careers are dedicated to research to have a good understanding of statistics, this area of knowledge is not useful for people with practitioner jobs, such as police and corrections officers. Construct a response to this assertion. Identify ways in which people in practical settings benefit from possessing an understanding of statistical concepts and techniques.

Review Problems

1. Define science and explain the role of methods in the production of scientific knowledge.
2. What is a population? Why are researchers usually unable to study populations directly?
3. What is a sample? Why do researchers draw samples?
4. Explain the role of replication in science.
5. List and briefly describe the different types of research in criminal justice and criminology.
6. Identify three theories that you have encountered in your criminal justice or criminology classes. For each one, write one hypothesis for which you could collect data in order to test that hypothesis.
7. Think of three types of programs or policies you have heard about or read about in your criminal justice or criminology classes. For each one, suggest a possible way to evaluate that program's or policy's effectiveness.
8. If a researcher were conducting a study on a topic about which very little is known and the researcher does not have theory or prior evidence to make predictions about what she will find in her study, what kind of research would she be doing? Explain your answer.
9. If a researcher were solely interested in finding out more about a particular phenomenon and focused entirely on a sample without trying to make inference to a population, what kind of research would he be doing? Explain your answer.
10. What does GIGO stand for? What does this cautionary concept mean in the context of statistical analyses?

Key Terms

Chapter 2 Types of Variables and Levels of Measurement

Learning Objectives

- Define variables and constants.
- Define unit of analysis and be able to identify the unit of analysis in any given study.
- Define independent and dependent variables and be able to identify each in a study.
- Explain the difference between empirical associations and causation.
- List and describe the four levels of measurement, including similarities and differences between them, and be able to identify the level of measurement of different variables.

The first thing you must be familiar with in statistics is the concept of a **variable**. A variable is, quite simply, something that varies. It is a coding scheme used to measure a particular characteristic of interest. For instance, asking all of your statistics classmates, "How many classes are you taking this term?" would yield many different answers. This would be a variable. Variables sit in contrast to **constants**, which are characteristics that assume only one value in a sample. It would be pointless for you to ask all your classmates whether they are taking statistics this term because of course the answer they would all provide is "yes."

Variable: A characteristic that describes people, objects, or places and takes on multiple values in a sample or population.

Constant: A characteristic that describes people, objects, or places and takes on only one value in a sample or population.

Units of Analysis

It seems rather self-evident, but nonetheless bears explicit mention, that every scientific study contains *something* that the researcher conducting the study gathers and examines. These "somethings" can be objects or entities such as rocks, people, molecules, or prisons. This is called the **unit of analysis**, and it is, essentially, whatever the sample under study consists of. In criminal justice and criminology research, individual people are often the units of analysis. These individuals might be probationers, police officers, criminal defendants, or judges. Prisons, police departments, criminal incidents, or court records can also be units of analysis. Larger units are also popular; for example, many studies focus on census tracks, block groups, cities, states, or even countries. [Research Example 2.2](#) describes the methodological setup of a selection of criminal justice studies, each of which employed a different unit of analysis.

> **Unit of analysis**: The object or target of a research study.

Independent Variables and Dependent Variables

Researchers in criminal justice and criminology typically seek to examine relationships between two or more variables. Observed or **empirical** phenomena give rise to questions about the underlying forces driving them. Take homicide as an example. Homicide events and city-level rates are empirical phenomena. It is worthy of note that Washington, D.C., has a higher homicide rate than Portland, Oregon. Researchers usually want to do more than merely note empirical findings, however—they want to know *why* things are the way they are. They might, then, attempt to identify the criminogenic (crime-producing) factors that are present in Washington but absent in Portland or, conversely, the protective factors possessed by Portland and lacked by Washington.

Empirical: Having the qualities of being measurable, observable, or tangible. Empirical phenomena are detectable with senses such as sight, hearing, or touch.

Research Example 2.1 Choosing Variables for a Study on Police Use of Conducted Energy Devices

Conducted energy devices (CEDs) such as the Taser have garnered national—indeed, international—attention in the past few years. Police practitioners contend that CEDs are invaluable tools that minimize injuries to both officers and suspects during contentious confrontations, whereas critics argue that police sometimes use CEDs in situations where such a high level of force is not warranted. Do police seem to be using CEDs appropriately? Gau, Mosher, and Pratt (2010) addressed this question. They sought to determine whether suspects' race or ethnicity influenced the likelihood that police officers would deploy or threaten to deploy CEDs against those suspects. In an analysis of this sort, it is important to account for other variables that might be related to police use of CEDs or other types of force; therefore, the researchers included suspects' age, sex, and resistance level. They also measured officers' age, sex, and race. Finally, they included a variable indicating whether it was light or dark outside at the time of the encounter. The researchers found that police use of CEDs was driven primarily by the type and intensity of suspect resistance but that even controlling for resistance, Latino suspects faced an elevated probability of having CEDs either drawn or deployed against them.

Research Example 2.2 Units of Analysis

Each of the following studies used a different unit of analysis.

1. *Do prison inmates incarcerated in facilities far from their homes commit more misconduct than those housed in facilities closer to home?* Lindsey, Mears, Cochran, Bales, and Stults (2017) used data from the Florida Department of Corrections to find out whether distally placed inmates (i.e., those sent to facilities far from their homes) engaged in more in-prison misbehavior, and, if so, whether this effect was particularly pronounced for younger inmates. Individual prisoners were the units of analysis in this study. The findings revealed a curvilinear relationship between distance and misconduct: Prisoners' misconduct increased along with distance up to approximately 350 miles, but then the relationship inverted such that further increases in distance were associated with less misconduct. As predicted, this pattern was strongest among younger inmates. Visitation helped offset the negative impact of distance but did not eliminate it. The researchers concluded that family visitation might have mixed effects on inmates. Inmates might be less inclined to commit misconduct if they fear losing visitation privileges, but receiving visits might induce embarrassment and shame when their family sees them confined in the prison environment. This strain, in turn, could prompt them to act out. Those who do not see their families much or at all do not experience this unpleasant emotional reaction.

2. *Is the individual choice to keep a firearm in the home affected by local levels of crime and police strength?* Kleck and Kovandzic (2009), using individual-level data from the General Social Survey (GSS) and city-level data from the FBI, set out to determine whether city-level homicide rates and the number of police per 100,000 city residents affected GSS respondents' likelihood of owning a firearm. There were two units of analysis in this study: individuals and cities. The statistical models indicated that high homicide rates and low police levels both modestly increased the likelihood that a given person would own a handgun; however, the relationship between city homicide rate and individual gun ownership decreased markedly

when the authors controlled for whites' and other nonblacks' racist attitudes toward African Americans. It thus appeared that the homicide–gun ownership relationship was explained in part by the fact that those who harbored racist sentiments against blacks were more likely to own firearms regardless of the local homicide rate.

3. *How consistent are use-of-force policies across police agencies?* The U.S. Supreme Court case *Graham v. Connor* (1989) requires that police officers use only the amount of force necessary to subdue a resistant suspect; force exceeding that minimum is considered excessive. The Court left it up to police agencies to establish force policies to guide officers' use of physical coercion. Terrill and Paoline (2013) sought to determine what these policies look like and how consistent they are across agencies. The researchers mailed surveys to a sample of 1,083 municipal police departments and county sheriff offices nationwide, making the agency the unit of analysis. Results showed that 80% of agencies used a force continuum as part of their written use-of-force policies, suggesting some predictability in the way in which agencies organize their policies. However, there was substantial variation in policy restrictiveness and the placement of different techniques and weapons. Most agencies placed officer presence and verbal commands at the lowest end, and deadly force at the highest, but between those extremes there was variability in the placement of soft and hard hand tactics, chemical sprays, impact weapons, CEDs, and other methods commonly used to subdue noncompliant suspects. These findings show how localized force policies are, and how inconsistent they are across agencies.

4. *Does gentrification reduce gang homicide?* Gentrification is the process by which distressed inner-city areas are transformed by an influx of new businesses or higher-income residents. Gentrification advocates argue that the economic boost will revitalize the area, provide new opportunities, and reduce crime. Is this assertion true? Smith (2014) collected data from 1994 to 2005 on all 342 neighborhoods in Chicago with the intention of determining whether gentrification over time reduces gang-motivated homicide. Smith measured gentrification in three ways: Recent increases in neighborhood residents' socioeconomic statuses, increases in coffee shops, and demolition of public housing. The author predicted that the first two would suppress gang homicide and that the last one would increase it; even though public-housing demolition is supposed to reduce crime, it can also create turmoil, residential displacement, and conflict among former public-housing residents and residents of surrounding properties. Smith found support for all three hypotheses. Socioeconomic-status increases were strongly related to reductions in gang-motivated homicides, coffee-shop presence was weakly related to reductions, and public-housing demolition was robustly associated with increases. These results suggest that certain forms of gentrification might be beneficial to troubled inner-city neighborhoods but that demolishing public housing might cause more problems than it solves, at least in the short term.

Researchers undertaking quantitative studies must specify **dependent variables** (DVs) and **independent variables** (IVs). Dependent variables are the empirical events that a researcher is attempting to explain. Homicide rates, property crime rates, recidivism among recently released prisoners, and judicial sentencing decisions are examples of DVs. Researchers seek to identify variables that help predict or explain these events. Independent variables are factors a researcher believes might affect the DV. It might be predicted, for instance, that prisoners released into economically and socially distressed neighborhoods and given little support during the reentry process will recidivate more frequently than those who receive transitional housing and employment assistance. Different variables—crime rates, for instance—can be used as both IVs and DVs across different studies. The designation of a certain phenomenon as an IV or a DV depends on the nature of the research study.

Dependent variable: The phenomenon that a researcher wishes to study, explain, or predict.

Independent variable: A factor or characteristic that is used to try to explain or predict a dependent variable.

Relationships Between Variables: A Cautionary Note

It is vital to understand that *independent* and *dependent* are not synonymous with *cause* and *effect*, respectively. A particular IV might be related to a certain DV, but this is far from definitive proof that the former is the cause of the latter. To establish causality, researchers must demonstrate that their studies meet three criteria. First is **temporal ordering**, meaning that the IV must occur prior to the DV. It would be illogical, for instance, to predict that adolescents' participation in delinquency will impact their gender; conversely, it does make sense to predict that adolescents' gender affects the likelihood they will commit delinquent acts. The second causality requirement is that there be an **empirical relationship** between the IV and the DV. This is a basic necessity—it does not make sense to try to delve into the nuances of a nonexistent connection between two variables. For example, if a researcher predicts that people living in high-crime areas are more likely to own handguns for self-protection, but then finds no relationship between neighborhood-level crime rates and handgun ownership, the study cannot proceed.

> **Temporal ordering**: The causality requirement holding that an independent variable must precede a dependent variable.
>
> **Empirical relationship**: The causality requirement holding that the independent and dependent variables possess an observed relationship with one another.

The last requirement is that the relationship between the IV and the DV be **nonspurious**. This third criterion is frequently the hardest to overcome in criminology and criminal justice research (indeed, all social sciences) because human behavior is complicated, and each action a person engages in has multiple causes. Disentangling these causal factors can be difficult or impossible.

> **Nonspuriousness**: The causality requirement holding that the relationship between the independent variable and dependent variable not be the product of a third variable that has been erroneously omitted from the analysis.

The reason spuriousness is a problem is that there could be a third variable that explains the DV as well as, or even better than, the IV does. This third variable might partially or fully account for the relationship between the IV and DV. The inadvertent exclusion of one or more important variables can result in erroneous conclusions because the researcher might mistakenly believe that the IV strongly predicts the DV when, in fact, the relationship is actually partially or entirely due to intervening factors. Another term for this problem is **omitted variable bias**. When omitted variable bias (i.e., spuriousness) is present in an IV–DV relationship but erroneously goes unrecognized, people can reach the wrong conclusion about a phenomenon. Research Example 2.3 offers examples of the problem of omitted variables.

> **Omitted variable bias**: An error that occurs as a result of unrecognized spuriousness and a failure to include important third variables in an analysis, leading to incorrect conclusions about the relationship between the independent and dependent variables.

A final caution with respect to causality is that statistical analyses are examinations of aggregate trends. Uncovering an association between an IV and a DV means only that the presence of the IV has the *tendency* to be related to either an increase or a reduction in the DV in the sample as a whole—it is not an indication that

the IV–DV link inevitably holds true for every single person or object in the sample. For example, victims of early childhood trauma are more likely than nonvictims to develop substance abuse disorders later in life (see Dass-Brailsford & Myrick, 2010). Does this mean that every person who was victimized as a child has substance abuse problems as an adult? Certainly not! Many people who suffer childhood abuse do not become addicted to alcohol or other drugs. Early trauma is a risk factor that elevates the risk of substance abuse, but it is not a guarantee of this outcome. Associations present in a large group are not uniformly true of all members of that group.

Research Example 2.3 The Problem of Omitted Variables

In the 1980s and 1990s a media and political frenzy propelled the "crack baby" panic to the top of the national conversation. The allegations were that "crack mothers" were abusing the drug while pregnant and were doing irreparable damage to their unborn children. Stories of low-birth-weight, neurologically impaired newborns abounded. What often got overlooked, though, was the fact that women who use crack cocaine while pregnant are also likely to use drugs such as tobacco and alcohol, which are known to harm fetuses. These women also frequently have little or no access to prenatal nutrition and medical care. Finally, if a woman abuses crack —or any other drug—while pregnant, she could also be at risk for mistreating her child after its birth (see Logan, 1999, for a review). She might be socially isolated, as well, and have no support from her partner or family. There are many factors that affect fetal and neonatal development, some under mothers' control and some not; trying to tie children's outcomes definitively to a single drug consumed during mothers' pregnancies is inherently problematic.

In the 1980s, policymakers and the public became increasingly concerned about domestic violence. This type of violence had historically been treated as a private affair, and police tended to take a hands-off approach that left victims stranded and vulnerable. The widely publicized results of the Minneapolis Domestic Violence Experiment suggested that arrest effectively deterred abusers, leading to lower rates of recidivism. Even though the study's authors said that more research was needed, states scrambled to enact mandatory arrest laws requiring officers to make arrests in all substantiated cases of domestic violence. Subsequent experiments and more detailed analyses of the Minneapolis data, however, called the effectiveness of arrest into question. It turns out that arrest has no effect on some offenders and even increases recidivism among certain groups. Offenders' employment status, in particular, emerged as an important predictor of whether arrest deterred future offending. Additionally, the initial reduction in violence following arrest frequently wore off over time, putting victims back at risk. Pervasive problems collecting valid, reliable data also hampered researchers' ability to reach trustworthy conclusions about the true impact of arrest (see Schmidt & Sherman, 1993, for a review). The causes of domestic violence are numerous and varied, so it is unwise to assume that arrest will be uniformly advantageous.

In sum, you should always be cautious when interpreting IV–DV relationships. It is better to think of IVs as *predictors* and DVs as *outcomes* rather than to view them as causes and effects. As the adage goes, correlation does not mean causation. Variables of all kinds are related to each other, but it is important not to leap carelessly to causal conclusions on the basis of statistical associations.

Levels of Measurement

Every variable possesses a **level of measurement**. Levels of measurement are ways of classifying or describing variable types. There are two overarching classes of variables: **categorical** (also sometimes called *qualitative*) and **continuous** (also sometimes referred to as *quantitative*). Categorical variables comprise groups or classifications that are represented with labels, whereas continuous variables are made of numbers that measure how much of a particular characteristic a person or object possesses. Each of these variable types contains two subtypes. This two-tiered classification system is diagrammed in Figure 2.1 and discussed in the following sections.

Level of measurement: A variable's specific type or classification. There are four types: *nominal, ordinal, interval*, and *ratio*.

Categorical variable: A variable that classifies people or objects into groups. There are two types: *nominal* and *ordinal*.

Continuous variable: A variable that numerically measures the presence of a particular characteristic. There are two types: *interval* and *ratio*.

The Categorical Level of Measurement: Nominal and Ordinal Variables

Categorical variables are made up of categories. They represent ways of divvying up people and objects according to some characteristic. Categorical variables are subdivided into two types: *nominal* and *ordinal*. The **nominal** level of measurement is the most rudimentary of all the levels. It is the least descriptive and sometimes the least informative. Race is an example of a nominal-level variable. See Tables 2.1 and 2.2 for examples of nominal variables (see also Data Sources 2.1 for a description of the data set used in these tables). The variable in Table 2.1 comes from a question on the survey asking respondents whether or not they personally know a police officer assigned to their neighborhood. This variable is nominal because respondents said "yes" or "no" in response and so can be grouped accordingly. In Table 2.2, the variable representing the races of stopped drivers is nominal because races are groups into which people are placed. The labels offer descriptive information about the people or objects within each category. Data are from the Police–Public Contact Survey (PPCS).

> **Nominal variable**: A classification that places people or objects into different groups according to a particular characteristic that cannot be ranked in terms of quantity.

Figure 2.1 Levels of Measurement
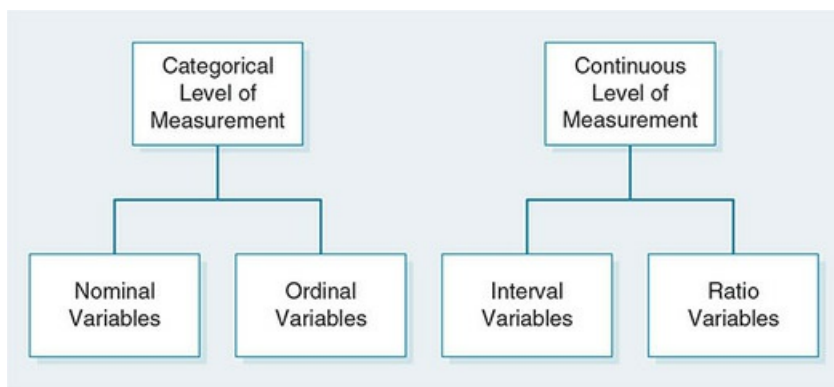
Table 2.1 Knowledge of Police in the Neighborhood (PPCS)	
Do you know any police officers that work in your neighborhood by name or by sight?	*Frequency*
Yes	9,246
No	32,263
	Total = 41,509

Table 2.2 Race of Stopped Drivers (PPCS)	
Driver Race	*Frequency*
White	416
Black	63
Asian	14
Other, including multiracial	26
	Total = 519

Gender is another example of a nominal variable. Table 2.3 displays the gender breakdown among people who reported that they had sought help from the police within the past year.

Table 2.3 Gender of People Who Sought Help From Police (PPCS)	
Gender	Frequency
Male	1,511
Female	1,904
	Total = 3,415

Much information is missing from the nominal variables in Tables 2.1 through 2.3. For instance, the question about knowing a local police officer does not tell us how often respondents talk to the officers they know or whether they provide these officers with information about the area. Similarly, the race variable provides fairly basic information. This is why the nominal level of measurement is lowest in terms of descriptiveness and utility. These classifications represent only *differences*; there is no way to arrange the categories in any meaningful rank or order. Nobody in one racial group can be said to have "more race" or "less race" than someone in another category—they are merely of different races. The same applies to gender. Most people identify as being either female or male, but members of one gender group do not have more or less gender relative to members of the other group.

One property that nominal variables possess (and share with other levels) is that the categories within any given variable are **mutually exclusive** and **exhaustive**. They are mutually exclusive because each unit in the data set (person, place, and so on) can fall into only one category. They are exhaustive because all units have a category that applies to them. For example, a variable measuring survey respondents' criminal histories that asks them if they have been arrested "0–1 time" or "1–2 times" would not be mutually exclusive because a respondent who has been arrested once could circle both answer options. This variable would also violate the principle of exhaustiveness because someone who has been arrested three or more times cannot circle any available option because neither is applicable. To correct these problems, the answer options could be changed to, for instance, "no arrests," "1–2 arrests," and "3 or more arrests." Everyone filling out the survey would have one, and only one, answer option that accurately reflected their experiences.

Mutually exclusive: A property of all levels of measurement whereby there is no overlap between the categories within a variable.

Exhaustive: A property of all levels of measurement whereby the categories or range within a variable capture all possible values.

Ordinal variables are one step up from nominal variables in terms of descriptiveness because they can be ranked according to the quantity of a characteristic possessed by each person or object in a sample. University students' class level is an ordinal variable because freshmen, sophomores, juniors, and seniors can be rank-ordered according to how many credits they have earned. Numbers can also be represented as ordinal classifications when the numbers have been grouped into ranges like those in Table 2.3, where the income categories of respondents to the General Social Survey (GSS; see Data Sources 2.2) are shown. Table 2.4 displays another variable from the PPCS. This survey question queried respondents about how often they drive. Respondents were offered categories and selected the one that most accurately described them.

> **Ordinal variable**: A classification that places people or objects into different groups according to a particular characteristic that can be ranked in terms of quantity.

Ordinal variables are useful because they allow people or objects to be ranked in a meaningful order. Ordinal variables are limited, though, by the fact that no algebraic techniques can be applied to them. This includes ordinal variables made from numbers such as those in Table 2.3. It is impossible, for instance, to subtract *<$1,000* from *$15,000–$19,999*. This eliminates the ability to determine exactly how far apart two respondents are in their income levels. The difference between someone in the *$20,000–$24,999* group and the ≥*$25,000* group might only be $1 if the former makes $24,999 and the latter makes $25,000. The difference could be enormous, however, if the person in the ≥*$25,000* group has an annual family income of $500,000 per year. There is no way to figure this out from an assortment of categories like those in Table 2.3. The same limitation applies to the variable in Table 2.4. A glance at the table reveals general information about the frequency of driving, but there is no way to add, subtract, multiply, or divide the categories to obtain a specific measurement of how more or less one person in the sample drives relative to another.

Learning Check 2.1

✓

In a study of people incarcerated in prison, the variable "offense type" captures the crime that each person was convicted of and imprisoned for. This variable could be coded as either a nominal or an ordinal variable. Explain why this is. Give an example of each type of measurement approach.

Table 2.4 Annual Family Income (GSS)

Family Income	Frequency
<$1,000	40
$1,000–$2,999	32
$3,000–$3,999	19
$4,000–$4,999	14
$5,000–$5,999	20
$6,000–$6,999	18
$7,000–$7,999	16
$8,000–$9,999	50
$10,000–$14,999	150
$15,000–$19,999	114
$20,000–$24,999	157
≥$25,000	1,684
	Total = 2,314

Table 2.5 Driving Habits (PPCS)

How often do you drive?	Frequency
Every day or almost every day	29,442
A few days a week	5,833
A few days a month	1,235
A few times a year	395
Never	4,623
	Total = 41,528

Data Sources 2.2 The General Social Survey

The National Opinion Research Center has conducted the GSS annually or every 2 years since 1972. Respondents are selected using a multistage clustering sample design. First, cities and counties are randomly selected. Second, block groups or districts are selected from those cities and counties. Trained researchers then canvass each block group or district on foot and interview people in person. Interviews are offered in both English and Spanish. The GSS contains a large number of variables. Some of these variables are asked in every wave of the survey, whereas others are asked only once. The variables include respondents' attitudes about religion, politics, abortion, the death penalty, gays and lesbians, persons of racial groups other than respondents' own, free speech, marijuana legalization, and a host of other topics (Davis & Smith, 2009). The most current wave of the GSS available at this time is the one conducted in 2014.

Table 2.6 contains another example of an ordinal variable. This comes from the GSS and measures female respondents' educational attainment. You can see that this variable is categorical and ranked. Someone who

has attended junior college has a higher educational attainment than someone who did not complete high school. As before, though, no math is possible. Someone whose highest education level is junior college might be in her first semester or quarter, putting her barely above someone in the high-school-only category, or she might be on the verge of completing her associate's degree, which places her nearly two years above.

Table 2.6 Female Respondents' Educational Levels (GSS)	
Highest Schooling Completed	Frequency
Less than High School	198
High School	673
Associate's Degree	122
Bachelor's Degree	257
Graduate Degree	147
	Total = 1,397

The Continuous Level of Measurement: Interval and Ratio Variables

Continuous variables differ from categorical ones in that the former are represented not by categories but rather by numbers. **Interval variables** are numerical scales in which there are equal distances between all adjacent points on those scales. Ambient temperature is a classic example of an interval variable. This scale is measured using numbers representing degrees, and every point on the scale is exactly one degree away from the nearest points on each side. Twenty degrees Fahrenheit, for instance, is exactly 1 degree cooler than 21 degrees and exactly 4 degrees warmer than 16 degrees.

> **Interval variable**: A quantitative variable that numerically measures the extent to which a particular characteristic is present or absent and does not have a true zero point.

An example of an interval-level variable is the GSS's scoring of respondents' occupational prestige. The GSS uses a ranking system to assign each person a number representing how prestigious his or her occupation is. Figure 2.2 displays the results. The scores range from 16 to 80 and so are presented as a chart rather than a table.

As can be seen in Figure 2.2, the prestige scores are numerical. This sets them apart from the categories seen in nominal and ordinal variables. The scores can be subtracted from one another; for instance, a respondent with a prestige score of 47 has 9 points fewer than someone whose score is 56. Importantly, however, the scores cannot be multiplied or divided. It does not make sense to say that someone with a score of 60 has twice the occupational prestige as someone with a 30. This is because interval data, such as this scale, do not have a true zero point. The numbers, although useful and informative, are ultimately arbitrary. An occupational prestige score of 30, for instance, could have been any other number if a different coding system had been employed.

Table 2.7 contains another interval variable. This one measures GSS respondents' political views. The scale ranges from 1 (extremely liberal) to 7 (extremely conservative). This variable in Table 2.7 is interval because it is a scale (as opposed to distinctly separate categories), but lacks a zero point and therefore is an arbitrary numbering system. The 1-to-7 numbering system could be replaced by any other 7-point consecutive sequence without affecting the meaning of the scale. These are the hallmark identifiers of interval-level data.

Figure 2.2 Respondents' Occupational Prestige Scores (GSS)
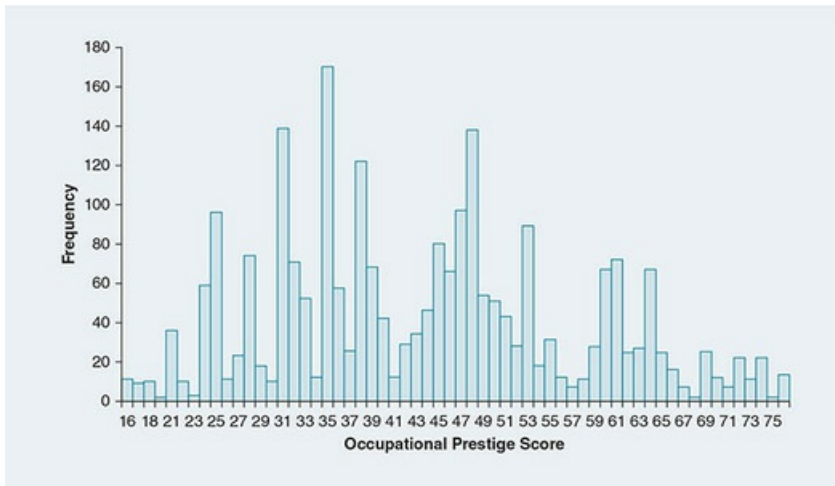
Table 2.7 Respondents' Political Views (GSS)	
Political Views	**Frequency**
1 (extremely liberal)	94
2	304
3	263
4	989
5	334
6	358
7 (extremely conservative)	107
	Total = 2,449

Another attitudinal variable measured at the interval level is shown in Table 2.8. This is a scale tapping into how punitive GSS respondents feel toward people convicted of crimes. The scale ranges from the lower range of punitiveness to the upper range. The idea behind attitudinal scores being interval (as opposed to categorical) is that attitudes are best viewed as a continuum. There is a lot of gray area that prevents attitudes from being accurately portrayed as ordinal; chunking attitudes up into categories introduces artificial separation between the people falling into adjacent positions on the scale.

Table 2.8 Punitiveness Scores (GSS)	
Political Views	**Frequency**
1 (least punitive)	191
2	358
3	643
4 (most punitive)	1,002
	Total = 2,194

Ratio variables are the other subtype within the continuous level of measurement. The ratio level resembles the interval level in that ratio, too, is numerical and has equal and known distance between adjacent points. The difference is that ratio-level scales, unlike interval ones, have meaningful zero points that represent the absence of a given characteristic. Temperature, for instance, is not ratio level because the zeros in the various temperature scales are just placeholders. Zero does not signify an *absence* of temperature. Likewise, the data

presented in [Figure 2.2](#) and [Tables 2.7](#) and [2.8](#), as discussed previously, do not have meaningful zero points and therefore cannot be multiplied or divided. You could not, for instance, say that someone who scores a 4 on the punitiveness scale is twice as punitive as a person with a 2 on this measure. Ratio-level data permit this higher level of detail.

> **Ratio variable**: A quantitative variable that numerically measures the extent to which a particular characteristic is present or absent and has a true zero point.

Criminal justice and criminology researchers deal with many ratio-level variables. Age is one example. Although it is strange to think of someone as having zero age, age can be traced close enough to zero to make analytically reasonable to think of this variable as ratio. A 40-year-old person is twice as old as someone who is 20.

[Table 2.9](#) displays the number of state prisoners who were executed in 2013. These data come from the BJS (Snell, 2014; see [Data Sources 2.3](#)). The left-hand column of the table displays the number of people executed per state, and the right-hand column shows the frequency with which each of those state-level numbers occurred (note that the frequency column sums to 50 to represent all of the states in the country). Can you explain why *number of persons executed* is a ratio-level variable?

In [Table 2.10](#), the number of children GSS respondents report having is shown. Since the zero is meaningful here, this qualifies as ratio level.

Table 2.9 Number of Persons Executed, 2013	
Number Executed	**Frequency**
0	41
1	3
2	2
3	1
6	1
7	1
16	1
	Total = 50

Another example of the ratio level of measurement is offered in <u>Figure 2.3</u>. These data come from a question on the PPCS asking respondents who had been the subject of a traffic or pedestrian stop in the past 12 months the number of minutes the stop lasted. Time is ratio level because, in theory, it can be reduced down to zero, even if nobody in the sample actually reported zero as their answer.

Figure 2.3 Length of Stop, in Minutes (PPCS)
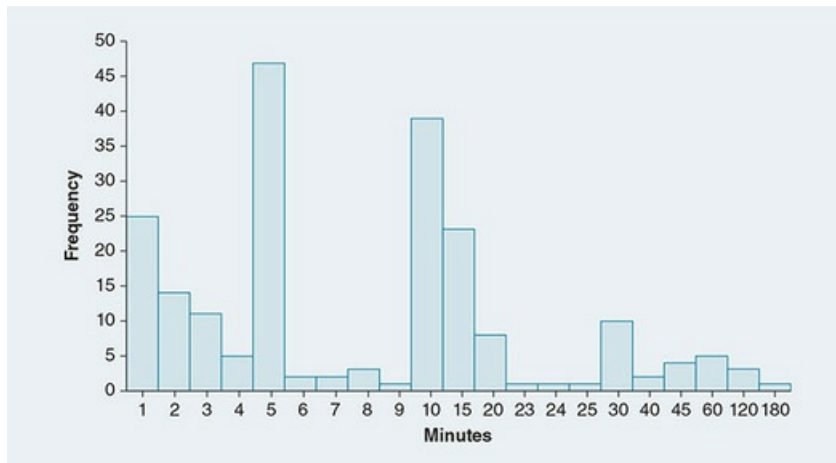
Table 2.10 Number of Children (GSS)

Children	Frequency
0	704
1	419
2	658
3	394
4	198
5	88
6	40
7	14
8	15
	Total = 2,530

Data Sources 2.3 The Bureau of Justice Statistics

The BJS is the U.S. Department of Justice's repository for statistical information on criminal justice–related topics. The BJS offers downloadable data and periodic reports on various topics that summarize the data and present them in a user-friendly format. Researchers, practitioners, and students all rely on the BJS for accurate, timely information about crime, victims, sentences, prisoners, and more. Visit http://bjs.ojp.usdoj.gov/ and explore this valuable information source.

In the real world of statistical analysis, the terms *interval* and *ratio variables* are often used interchangeably. It is the overarching categorical-versus-continuous distinction that usually matters most when it comes to statistical analyses. When a researcher is collecting data and has a choice about level of measurement, the best strategy is to always use the highest level possible. A continuous variable can always be made categorical later, but a categorical variable can never be made continuous.

Level of measurement is a very important concept. It might be difficult to grasp if this is the first time you have been exposed to this idea; however, it is imperative that you gain a firm understanding because level of measurement determines what analyses can and cannot be conducted. This fundamental point will form an underlying theme of this entire book, so be sure you understand it. Do not proceed with the book until you can readily identify a given variable's level of measurement. Table 2.11 summarizes the basic characteristics that define each level and distinguish it from the others.

Table 2.11 Characteristics of Each Level of Measurement

| Level of Measurement | Variable Characteristic | | | |
	Mutually exclusive and exhaustive	Rank-orderable	Equal intervals	True zero
Nominal	✓			
Ordinal	✓	✓		
Interval	✓	✓	✓	
Ratio	✓	✓	✓	✓

This chapter discussed the concept of a variable. You also read about units of analysis, independent variables, dependent variables, and about the importance of not drawing strict causal conclusions about statistical relationships. It is important to consider whether meaningful variables have been left out and to keep in mind that empirical associations do not imply that one variable causes the other.

This chapter also described the two overarching levels of measurement: categorical and continuous. Categorical variables are qualitative groupings or classifications into which people or objects are placed on the basis of some characteristic. The two subtypes of categorical variables are nominal and ordinal. These two kinds of variables are quite similar in appearance, with the distinguishing feature being that nominal variables cannot be rank-ordered, whereas ordinal variables can be. Continuous variables are quantitative measurements of the presence or absence of a certain characteristic in a group of people or objects. Interval and ratio variables are both continuous. The difference between them is that ratio-level variables possess true zero points and interval-level variables do not.

You must understand this concept and be able to identify the level of measurement of any given variable because, in statistics, the level at which a variable is measured is one of the most important determinants of the graphing and analytic techniques that can be employed. In other words, each type of graph or statistical analysis can be used with some levels of measurement and cannot be used with others. Using the wrong statistical procedure can produce wildly inaccurate results and conclusions. You must therefore possess an understanding of level of measurement before leaving this chapter.

Thinking Critically

1. Suppose you have been contacted by a reporter from the local newspaper who came across data showing that men tend to be sentenced more harshly than women (e.g., more likely to be sent to prison, given longer sentence lengths). The reporter believes this to be a clear case of discrimination and asks you for comment. What is your response? Do you agree that gender discrimination has been demonstrated here, or do you need more information? If the latter, what additional data would you need before you could arrive at a conclusion?
2. Many researchers have tried to determine whether capital punishment deters murder. Suppose a new study has been published analyzing how death-sentence rates in one year relate to murder rates the following year. The researchers who conducted this study included only the 32 states that authorize the death penalty, and excluded the remaining states. Do you think this is a justifiable approach to studying the possible deterrent effects of the death penalty? Would you trust the results of the analysis and the conclusions the researchers reach on the basis of those results? Explain your answer.

Review Problems

1. A researcher wishes to test the hypothesis that low education affects crime. She gathers a sample of people aged 25 and older.
   1. What is the independent variable?
   2. What is the dependent variable?
   3. What is the unit of analysis?
2. A researcher wishes to test the hypothesis that arrest deters recidivism. She gathers a sample of people who have been arrested.
   1. What is the independent variable?
   2. What is the dependent variable?
   3. What is the unit of analysis?
3. A researcher wishes to test the hypothesis that poverty affects violent crime. He gathers a sample of neighborhoods.
   1. What is the independent variable?
   2. What is the dependent variable?
   3. What is the unit of analysis?
4. A researcher wishes to test the hypothesis that prison architectural design affects the number of inmate-on-inmate assaults that take place inside a facility. He gathers a sample of prisons.
   1. What is the independent variable?
   2. What is the dependent variable?

3. What is the unit of analysis?

5. A researcher wishes to test the hypothesis that the amount of money a country spends on education, health, and welfare affects the level of violent crime in that country. She gathers a sample of countries.
    1. What is the independent variable?
    2. What is the dependent variable?
    3. What is the unit of analysis?

6. A researcher wishes to test the hypothesis that police officers' job satisfaction affects the length of time they stay in their jobs. He gathers a sample of police officers.
    1. What is the independent variable?
    2. What is the dependent variable?
    3. What is the unit of analysis?

7. A researcher wishes to test the hypothesis that the location of a police department in either a rural or an urban area affects starting pay for entry-level police officers. She gathers a sample of police departments.
    1. What is the independent variable?
    2. What is the dependent variable?
    3. What is the unit of analysis?

8. A researcher wishes to test the hypothesis that the level of urbanization in a city or town affects residents' social cohesion. She gathers a sample of municipal jurisdictions (cities and towns).
    1. What is the independent variable?
    2. What is the dependent variable?
    3. What is the unit of analysis?

9. Suppose that a researcher found a statistical relationship between ice cream sales and crime—during months when a lot of ice cream is purchased, crime rates are higher. The researcher concludes that ice cream causes crime. What has the researcher done wrong?

10. Suppose that in a random sample of adults a researcher found a statistical relationship between parental incarceration and a person's own involvement in crime. Does this mean that every person who had a parent in prison committed crime? Explain your answer.

11. Identify the level of measurement of each of the following variables:
    1. Suspects' race measured as *white, black, Latino*, and *other*
    2. The age at which an offender was arrested for the first time
    3. The sentences received by convicted defendants, measured as *jail, prison, probation, fine*, and *other*
    4. The total number of status offenses that adult offenders reported having committed as juveniles
    5. The amount of money, in dollars, that a police department collects annually from drug asset forfeitures
    6. Prison security level, measured as *minimum*, *medium*, and *maximum*
    7. Trial judges' gender

12. Identify the level of measurement of each of the following variables:
    1. The amount of resistance a suspect displays toward the police, measured as *not resistant, somewhat resistant*, or *very resistant*
    2. The number of times someone has shoplifted in her or his life
    3. The number of times someone has shoplifted, measured as *0–2, 3–5*, or *6 or more*
    4. The type of attorney a criminal defendant has at trial, measured as *privately retained* or *publicly funded*
    5. In a sample of juvenile delinquents, whether or not those juveniles have substance abuse disorders
    6. Prosecutors' charging decisions, measured as *filed charges* and *did not file charges*
    7. In a sample of offenders sentenced to prison, the number of days in their sentences

13. If a researcher is conducting a survey and wants to ask respondents about their self-reported involvement in shoplifting, there are a few different ways he could phrase this question.
    1. Identify the level of measurement that each type of phrasing shown below would produce.
    2. Explain which of the three possible phrasings would be the best one to choose and why this is.

        Possible phrasing 1: How many times have you taken small items from stores without paying for those items? Please write in: _____

        Possible phrasing 2: How many times have you taken small items from stores without paying for those items? Please circle one of the following:

Never 1–2 times 3–4 times 5+ times

Possible phrasing 3: Have you ever taken small items from stores without paying for those items? Please circle one of the following:

Yes   No

14. If a researcher is conducting a survey and wants to ask respondents about the number of times each of them has been arrested, there are a few different ways she could phrase this question. Write down the three possible phrasing methods.

15. The following table contains BJS data on the number of prisoners under sentence of death, by region. Use the table to do the following

Number of Prisoners Under Sentence of Death at Year-End 2012 (Snell, 2014)	
Region	*Prisoners*
Northeast	211
Midwest	221
South	1,556
West	989
	Total = 2,977

1. Identify the level of measurement of the variable *region*.
2. Identify the level of measurement of *number of prisoners under sentence of death*.

16. The following table contains data from the 2012 National Crime Victimization Survey showing the number of victimization incidents and whether or not those crimes were reported to the police. The data are broken down by victims' household income level. Use the table to do the following:

Number of NCVS Victimizations Reported to Police, by Household Income (U.S. Department of Justice, 2012)			
	Victimization Reported		
Income	*Yes*	*No*	*Total*
$12,499 or less	399	697	1,096
$12,500–$24,999	393	840	1,233
$25,000–$49,999	680	1,206	1,886
$50,000 or more	1,048	1,840	2,888
Total	2,520	4,583	$N = 7,103$

1. Identify the level of measurement of the *variable income*.
2. Identify the level of measurement of the variable *victimization reported*.

17. Haynes (2011) conducted an analysis to determine whether victim advocacy affects offender sentencing. She gathered a sample of courts and measured victim advocacy as a yes/no variable indicating whether or not there was a victim witness office located inside each courthouse. She measured sentencing as the number of months of incarceration imposed on convicted offenders in the courts.

1. Identify the independent variable in this study.
2. Identify the level of measurement of the independent variable.
3. Identify the dependent variable in this study.
4. Identify the level of measurement of the dependent variable.
5. Identify the unit of analysis.

18. Bouffard and Piquero (2010) wanted to know whether arrested suspects' perceptions of the way police treated them during the encounter affected the likelihood that those suspects would commit more crimes in the future. Their sample consisted of males who had been arrested at least once during their lives. They measured suspects' perceptions of police behavior as *fair* or

unfair. They measured recidivism as the number of times suspects came into contact with police after that initial arrest.

1. Identify the independent variable in this study.
2. Identify the level of measurement of the independent variable.
3. Identify the dependent variable in this study.
4. Identify the level of measurement of the dependent variable.
5. Identify the unit of analysis.

19. Kleck and Kovandzic (2009; see Research Example 2.2) examined whether the level of homicide in a particular city affected the likelihood that people in that city would own firearms. They measured homicide as the number of homicides that took place in the city in 1 year divided by the total city population. They measured handgun ownership as whether survey respondents said they did or did not own a gun.

1. Identify the independent variable used in this study.
2. Identify the level of measurement of the independent variable.
3. Identify the dependent variable in this study.
4. Identify the level of measurement of the dependent variable.
5. Identify the unit of analysis. (Hint: This study has two!)

20. Gau et al. (2010; see Research Example 2.1) examined whether suspects' race or ethnicity influenced the likelihood that police would brandish or deploy Tasers against them. They measured race as *white, Hispanic, black*, or *other*. They measured Taser usage as *Taser used* or *some other type of force used*.

1. Identify the independent variable used in this study.
2. Identify the level of measurement of the independent variable.
3. Identify the dependent variable in this study.
4. Identify the level of measurement of the dependent variable.
5. Identify the unit of analysis.

Key Terms

Chapter 3 Organizing, Displaying, and Presenting Data

Data are usually stored in electronic files for use with software programs designed to conduct statistical analyses. The program SPSS is one of the most common data software programs in criminal justice and criminology. The SPSS layout is a data spreadsheet. You might be familiar with Microsoft Excel; if so, then the SPSS data window should look familiar. Figure 3.1 depicts what a typical SPSS file might look like.

The data in Figure 3.1 are from the Bureau of Justice 2011 Police–Public Contact Survey (PPCS; see Data Sources 2.1). Each row (horizontal line) in the grid represents one respondent, and each column (vertical line) represents one variable. Where any given row and column meet is a **cell** containing a given person's response to a particular question.

Cell: The place in a table or spreadsheet where a row and a column meet.

Your thoughts as you gaze on the data screen in Figure 3.1 can probably be aptly summarized as "Huh?" That is a very appropriate response, because there is no way to make sense of the data when they are in this raw format. This brings us to the topic of this chapter: methods for organizing, displaying, and presenting data. You can see from the figure that something has to be done to the data set to get it into a useful format. This chapter will teach you how to do just that.

Chapter 2 introduced levels of measurement (nominal, ordinal, interval, and ratio). A variable's level of measurement determines which graphs or charts are and are not appropriate for that variable. Various data displays exist, and many of them can be used only with variables of particular types. As you read this chapter, take notes on two main concepts: (1) the proper construction of each type of data display and (2) the level of measurement for which each display type is applicable.

Data Distributions

Univariate Displays: Frequencies, Proportions, and Percentages

Perhaps the most straightforward type of pictorial display is the **univariate** (one variable) frequency distribution. A **frequency** is simply a raw count; it is the number of times a particular characteristic appears in a data set. A frequency distribution is a tabular display of frequencies. Table 3.1 shows the frequency distribution for the variable *respondent gender* in the 2011 PPCS. Let us pause and consider two new symbols. The first is the *f* that sits atop the right-hand column in Table 3.1. This stands for *frequency*. There are 25,078 males in this sample. An alternative way of phrasing this is that the characteristic *male* occurs 25,078 times. A more formal way to write this would be f_{male} = 25,078; for females, f_{female} = 27,451. The second new symbol is the *N* found in the bottom right-hand cell. This represents the total sample size; here, $f_{male} + f_{female} = N$. Numerically, 25,078 + 27,451 = 52,529.

> **Univariate**: Involving one variable.
>
> **Frequency**: A raw count of the number of times a particular characteristic appears in a data set.

Figure 3.1 SPSS Data File

Table 3.1 Gender of PPCS Respondents	
Gender	*f*
Male	25,078
Female	27,451
	N = 52,529

Raw frequencies are of limited use in graphical displays because it is often difficult to interpret them and they do not offer much information about the variable being examined. What is needed is a way to standardize the numbers to enhance interpretability. Proportions do this. **Proportions** are defined as the number of times a particular characteristic appears in a sample relative to the total sample size. Formulaically,

$$p = \frac{f}{N},$$

Formula 3(1)

where

p = proportion

f = raw frequency

N = total sample size

Proportion: A standardized form of a frequency that ranges from 0.00 to 1.00.

Proportions range from 0.00 to 1.00. A proportion of exactly 0.00 indicates a complete absence of a given characteristic. If there were no males in the PPCS, their proportion would be

$$p_{male} = \frac{0}{52,529} = 0.00$$

Conversely, a trait with a proportion of 1.00 would be the only characteristic present in the sample. If the PPCS contained only men, then the proportion of the sample that was male would be

$$p_{male} = \frac{52,529}{52,529} = 1.00$$

Another useful technique is to convert frequencies into percentages (abbreviated *pct*). **Percentages** are a variation on proportions and convey the same information, but percentages offer the advantage of being more readily interpretable by the public. Percentages are computed similarly to proportions, with the added step of multiplying by 100:

$$pct = \left(\frac{f}{N}\right)100 \qquad\qquad Formula\ 3(2)$$

Percentage: A standardized form of a frequency that ranges from 0.00 to 100.00.

Proportions and percentages can be used in conjunction with frequencies to form a fuller, more informative display like that in Table 3.2. Note that the two proportions (p_{males} and $p_{females}$) sum to 1.00 because the two categories contain all respondents in the sample; the percentage column sums to 100.00 for the same reason. The "Σ" symbol in the table is the Greek letter *sigma* and is a summation sign. It instructs you to add up everything that is to the right of the symbol. The number to the right of the equal sign in "Σ =" is the summed total. As long as all cases in a sample have been counted once and only once, proportions will sum to 1.00 and percentages to 100.00 or within rounding error of these totals. Significant deviation from 1.00 or 100.00 that is greater than what could be explained by rounding error suggests a counting or calculation error.

Another useful technique for frequency distributions is the computation of cumulative measures. Cumulative frequencies, cumulative proportions, and cumulative percentages can facilitate meaningful interpretation of distributions, especially when data are continuous. Consider Table 3.3, which contains data from identity theft victims who discovered that someone had filed a fraudulent tax return in their name. The variable in the table reflects the dollar amount victims personally lost from the crime. The data are part of the larger National Crime Victimization Survey (NCVS; see Data Sources 1.2) and come from a 2014 supplement tapping into

respondents' experiences with various forms of identity theft. The dollar amounts reported by victims and the frequency of each response are located in the two left-most columns, respectively. To their right is a column labeled *cf*, which stands for *cumulative frequency*. The *cp* and *cpct* columns contain cumulative proportions and percentages, respectively.

Sex	f	p	pct
Male	25,078	.48	47.74
Female	27,451	.52	52.26
	N = 52,529	Σ = 1.00	Σ = 100.00

Table 3.2 Sex of PPCS Respondents

Cumulative columns are constructed by summing the *f*, *p*, and *pct* columns successively from row to row. The arrows in the table are intended to help you visualize this process. In the second row of Table 3.3's *cf* column, for instance, 48 is the sum of 45 and 3. In the *cp* column, likewise, .87 is the sum of .82 + .05. **Cumulatives** allow assessments of whether the data are clustered at one end of the scale or spread fairly equally throughout. In Table 3.3, it can be readily concluded that the data cluster at the low end of the scale, because .82 (or 81.82%) of victims did not incur any personal financial costs, and .95 (or 94.55%) spent $70 or less. This suggests that whatever profits offenders enjoy from income-tax fraud, victims themselves suffer minimal out-of-pocket expense.

Cumulative: A frequency, proportion, or percentage obtained by adding a given number to all numbers below it.

What is the level of measurement of the variable *state*? What is the level of measurement of the variable *property crime rate*? Refer to Chapter 2 if needed.

Table 3.3 Personal Costs Incurred by Victims of Fraudulent Tax Returns

Dollar amount of cost	f	cf	p	cp	pct	cpct
0	45 → 45		.82	.82	81.82	81.82
2	3 → 48		.05	.87	5.45	87.27
30	1	49	.02	.89	1.82	89.09
50	2	51	.04	.93	3.64	92.73
70	1	52	.02	.95	1.82	94.55
200	1	53	.02	.97	1.82	96.37
400	1	54	.02	.99	1.82	98.19
4,400	1	55	.02	1.01	1.82	100.01
	$N = 55$		$\Sigma = 1.01$		$\Sigma = 100.01$	

Univariate Displays: Rates

Suppose someone informed you that 1,579,527 burglaries were reported to the police in 2015. What would you make of this number? Nothing, probably, because raw numbers of this sort are simply not very useful. They lack a vital component—a denominator. The question that would leap to your mind immediately is "1,579,527 *out of what?*" You would want to know if this number was derived from a single city, from a single state or from the United States as a whole. This is where rates come in. A *rate* is a method of standardization that involves dividing the number of events of interest (e.g., burglaries) by the total population:

$$rate = \frac{f}{population} \qquad \qquad Formula\ 3(3)$$

Table 3.4 contains data from the 2015 Uniform Crime Report (UCR; see Data Sources 1.1). The column titled *Rate* displays the rate per capita that is obtained by employing Formula 3(3).

Table 3.4 UCR Index Offenses and Offense Rates, 2015			
Crime	f	Rate	Rate per 10,000
Burglary	1,579,527	.005	49.14
Aggravated assault	764,449	.002	23.78
Motor vehicle theft	707,758	.002	22.02

Note: 2012 U.S. Population = 321,418,820
Note: 2012 U.S. Population = 321,418,820

Note how tiny the numbers in the rate column are. Rates per capita do not make sense in the context of low-frequency events like crime because they end up being so small. It is, therefore, customary to multiply rates by a certain factor. This factor is usually 1,000, 10,000, or 100,000. You should select the multiplier that makes the most sense with the data at hand. In Table 3.4, the 10,000 multiplier has been used to form the *Rate per 10,000* column. Multiplying in this fashion lends clarity to rates because now it is no longer the number of crimes *per person* but, rather, the number of crimes *per 10,000 people*. If you randomly selected a sample of 10,000 people from the population, you would expect 49.14 of them to have been the victim of burglary in the past year and 23.78 to have experienced aggravated assault. These numbers and their interpretation are more real and more tangible than those derived using Formula 3(3) without a multiplier.

Sometimes rates must be calculated using multiple denominators, in contrast to Table 3.4 where there was only one. Table 3.5 shows a sample of states and the number of property crimes they reported in 2015.

States have different populations, so rates have to be calculated using each state's unique denominator. Table 3.6 expands upon the counts in Table 3.5 by including the property-crime rates per 10,000.

Table 3.5   UCR Property Crimes in Four States

State	f	Population
Alabama	144,746	4,858,979
California	1,024,914	39,144,818
Delaware	25,455	945,934
Michigan	187,101	9,922,576

Table 3.6   UCR Property-Crime Rates in Four States

State	f	Population	Rate per 10,000
Alabama	144,746	4,858,979	297.89
California	1,024,914	39,144,818	261.83
Delaware	25,455	945,934	269.10
Michigan	187,101	9,922,576	188.56

Learning Check 3.3

Rates and percentages have similar computational steps but very different meanings. Explain the differences between them, including the additional information needed to calculate a rate that is not needed for a percentage, the reason rates do not sum to 100 the way percentages do, and the substantive meaning of each one (i.e., the information that is provided by each type of number).

Bivariate Displays: Contingency Tables

Researchers are often interested not just in the frequency distribution of a single variable (a univariate display) but, rather, in the overlap between two variables. The Census of Jails (COJ; see Data Sources 3.1) collects data from all jails in the United States that hold inmates past arraignment (i.e., excluding lockups like stationhouse holding cells where inmates are confined for only a brief period before being transferred elsewhere).

Data Sources 3.1 The Census of Jails

The BJS has conducted the COJ periodically since 1970. Every 5 years, the BJS sends surveys to all jails operated by federal, state, and local governments, as well as by private corporations. The surveys capture institution-level data such as the total inmate population, the number of correctional staff, the number of inmate assaults against staff, whether the facility offers services (vocational, educational, or mental-health), and so on. The 2013 wave of the COJ is the most recent version available, but the 2006 version of the survey contains information on jail facilities that is not found in the 2013 version, so the examples in this book draw from both data sets.

Researchers studying jails might be interested in knowing whether facilities of various sizes differ in terms of whether they provide inmates with the opportunity to take secondary education courses that advance them toward the acquisition of a GED certificate. One might predict, for example, that larger jails will be busier and more crowded and that institutional security will therefore take precedence over the provision of services like GED courses. A **contingency table** (also sometimes called *crosstabs*) allows us to see the overlap between these two variables. Table 3.7 contains this display. This is a **bivariate** analysis, meaning it contains two variables. You can see from Table 3.7 that 155 small jails, 315 medium-sized ones, and 518 large ones provide inmates the opportunity to take GED classes while incarcerated.

Contingency table: A table showing the overlap between two variables.

Bivariate: An analysis containing two variables. Usually, one is designated the independent variable and the other the dependent variable.

Raw frequencies such as those shown in Table 3.7 offer a basic picture of bivariate overlap, but are not as informative as they could be. It is not immediately apparent from Table 3.7 whether the provision of GED courses differs across small, medium, and large facilities, since each of these groups is a different size (i.e., there are 778 small jails, 786 medium ones, and 807 large ones). To organize the data into a more readily interpretable format, proportions—or, more commonly, percentages—can be computed and entered into the contingency table in place of frequencies.

There are two types of proportions and percentages that can be computed in a bivariate contingency table: row and column. **Row proportions and percentages** are computed using the row marginals in the denominator, whereas **column proportions and percentages** employ the column marginals. There is no rule about which variable to place in the rows and which one in the columns, or whether you should compute row or column marginal (or both). These decisions should always be made based on the variables at hand and the point the researcher is trying to make with the data. Here, we want to discover the percentage of each facility type that

offers GED courses. It is common (though, again, not required) to place the independent variable in the rows and the dependent variable in the columns. Since we are predicting that facility size influences GED offerings, size is the IV and GED course availability is the DV. Hence, we will place size in the rows and calculate row percentages. Table 3.8 shows the percentage distribution.

> **Row proportions and percentages**: In a contingency (crosstabs) table, the proportions and percentages that are calculated using row marginals as the denominators. Each row sums to 1.00 and 100.00, respectively.
>
> **Column proportions and percentages**: In a contingency (crosstabs) table, the proportions and percentages that are calculated using column marginals as the denominators. Each column sums to 1.00 and 100.00, respectively.

Table 3.7 Jail Size and GED Courses Offered, per Facility (Frequencies)			
	GED Courses Offered?		
Jail Size	**Yes**	**No**	**Row Total**
Small	155	623	778
Medium	315	471	786
Large	518	289	807
Column Total	988	1,383	N = 2,371

Interesting results emerge from the row percentages in Table 3.8. It appears that small jails are the least likely to offer GED classes (19.92%) and that large ones are the most likely (64.19%), with medium-sized facilities falling in the middle (40.08%). The prediction that larger jails would be too preoccupied with security matters to concern themselves with providing educational services is not empirically supported. Perhaps the existence of a GED program inside a jail is more dependent on available resources, and smaller jails might lack the money and personnel needed to offer this type of benefit to inmates.

Learning Check 3.4

✓

Table 3.8 displays row percentages (i.e., percentages computed on the basis of row marginal). This shows the percentage of each type of facility that offers GED classes. How would column percentages be interpreted in this instance? As practice, calculate the column percentages for Table 3.8 and explain their meaning.

Table 3.8   Jail Size and GED Courses Offered, per Facility (Row Percentages)

Jail Size	GED Courses Offered?		Row Total
	Yes	No	
Small	$\left(\dfrac{155}{778}\right)100 = 19.92$	$\left(\dfrac{623}{778}\right)100 = 80.08$	100.00
Medium	$\left(\dfrac{315}{786}\right)100 = 40.08$	$\left(\dfrac{471}{786}\right)100 = 59.92$	100.00
Large	$\left(\dfrac{518}{807}\right)100 = 64.19$	$\left(\dfrac{289}{807}\right)100 = 35.81$	100.00
	988	1,383	$N = 2,371$

Research Example 3.1 Does Sexual-Assault Victimization Differ Between Female and Male Jail Inmates?

Research has shown that criminal offenders experience victimization at higher rates than the general population. This victimization might have predated the beginning of their criminal involvement, or it might have occurred because of the risky lifestyles that many offenders lead. Female offenders, in particular, experience high levels of sexual abuse and assault. Lane and Fox (2013) gathered data on a sample of jail inmates and asked them about their victimization histories and their fear of future victimization. (The numbers in the table are averages, with higher scores indicating greater levels of fear.) The following table displays the results, broken down by gender.

More than half of female inmates (51%) reported having been sexually assaulted; this number was 7% for men. Women were also more worried about future victimization of all types, although their scores on the fear-of-victimization variables showed that their fear of sexual assault outweighed their fear of other crimes. It has been argued in the past that sexual assault takes a unique physical and psychological toll on women—even those who have never actually experienced it—because it is an ever-present threat. Lane and Fox's results confirm the poignant impact that sexual assault has on female offenders.

	Women	Men
Sexual-assault victimization	51%	7%
Fear of property crime	1.73	1.55
Fear of violent crime	1.87	1.69
Fear of sexual assault	2.12	1.59
Fear of gang crime	1.74	1.57

Do Victim Impact Statements Influence Jurors' Likelihood of Sentencing Murder Defendants to Death?

Victim impact statements (VIS) are controversial. On the one hand, victims' rights advocates claim that speaking out in court is cathartic for victims and ensures they play a role in the process of punishing defendants who have hurt them or their loved ones. Critics, however, fear that these statements bias juries by injecting emotional pleas into what should be a neutral, detached review of the evidence presented during the sentencing phase of trial. Nuñez, Myers, Wilkowski, and Schweitzer (2017) sought a greater understanding of the effects of VIS on juror decision making in capital sentencing trials. In particular, they wanted to find out whether VIS that were sad in tone differed from those couched in anger. Nuñez and colleagues gathered a sample of people eligible to serve on juries and randomly assigned them to view one of six videotaped mock trials. In two trials, the murder victim's wife (played by an actor) read impact statements infused with either anger or sadness. The third trial contained no VIS. Each VIS condition was repeated across trial scenarios in which the mitigating evidence favoring the defendant was either weak or strong. The researchers then asked the participants whether they would sentence the defendant to life or to death. The table shows the sentencing decisions made by mock jurors in each of the three VIS conditions, broken down by whether the mitigating evidence in the case was weak or strong.

	Life (%)	Death (%)
No VIS		
Weak Evidence	53	47
Strong Evidence	66	34
Sad VIS		
Weak Evidence	57	43
Strong Evidence	60	40
Angry VIS		
Weak Evidence	33	67
Strong Evidence	51	49

The table reveals several interesting findings, foremost of which is that angry VIS do indeed seem to appeal to jurors' emotions and lead them to hand down death sentences more often; this conclusion flows from the fact that the participants in the "angry VIS" condition voted for death more often than those in either other VIS condition. The impact of an angry VIS is especially pronounced when mitigating evidence is weak. The effect of sad VIS is less pronounced; the presence of a sad victim does not substantially alter sentencing outcomes compared to having no victim present during sentencing. This study suggests that jurors making capital sentencing decisions are emotionally swayed by angry victims and might rest their sentencing decisions partially upon this stirring up of anger.

Graphs and Charts

Frequency, proportion, and percentage distributions are helpful ways of summarizing data; however, they are rather dull to look at. It is sometimes desirable to arrange data in a more attractive format. If you were giving a presentation to a local police department or district attorney's office, for instance, you would not want to throw numbers at your audience for 45 minutes. The monotony would be boring. Additionally, discerning a table's meaning typically requires close examination, which taxes an audience's ability to follow along with ease. Presentations can be diversified by the introduction of charts and graphs, of which there are many different types. Charts and graphs inject variety, color, and interest value into written and oral presentations of quantitative results. This chapter concentrates on five of the most common: *pie charts, bar graphs, histograms, frequency polygons*, and *line graphs*. Some of these charts and graphs are limited to specific levels of measurement, while others are useful with multiple types of data.

Categorical Variables: Pie Charts

Pie charts can be used only with categorical data, and they are most appropriate for variables that have relatively few **classes** (i.e., categories or groups) because pie charts get messy fast. A good general rule is to use a pie chart only when a variable contains five or fewer classes. Pie charts are based on percentages: The entire circle represents 100%, and the slices are sized according to the size of their contribution to that total.

Classes: The categories or groups within a nominal or ordinal variable.

The variable that will be used here to illustrate a pie chart is the race and ethnicity of stopped drivers from Table 2.2 (see page 22). The first step is to transform the raw frequencies into percentages using Formula 3(2). Once percentages have been computed, the pie chart can be built by dividing 100% into its constituent parts. Figure 3.2 contains the pie chart. Flip back to Table 2.2 and compare this pie chart to the raw frequency distribution to note the dramatic difference between the two presentation methods.

Figure 3.2 Race of Stopped Drivers (Percentages)
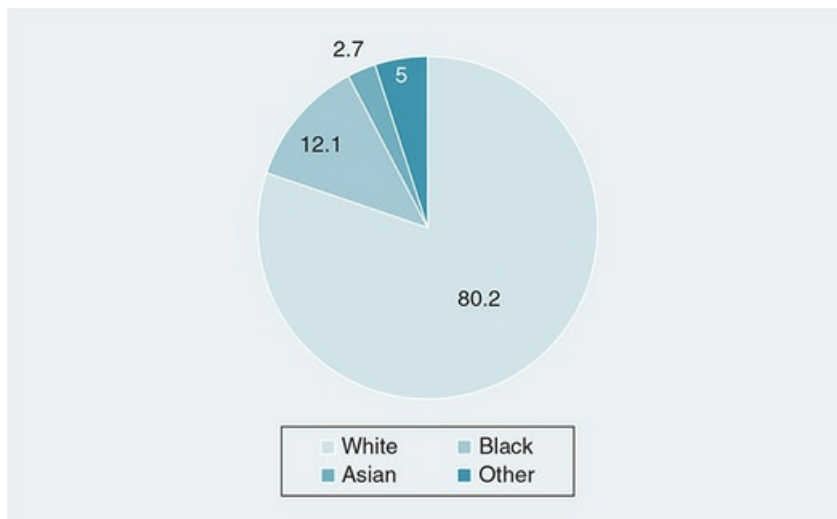
A pie chart can also be used to display data from the 2013 Law Enforcement Management and Administrative Statistics (LEMAS; see Data Sources 3.2) survey. This survey captures data on the types of police agencies nationwide (i.e., their primary jurisdiction). Figure 3.3 contains the percentages of agencies that are operated at the municipal (city or town police), county (sheriffs' offices), state, and tribal level. As you can see, the vast majority of police agencies are municipally operated.

Figure 3.3 Police Agency Type (Percentages)

0.3 1.2

19.6

78.9

Municipal ■ County ■ State ■ Tribal
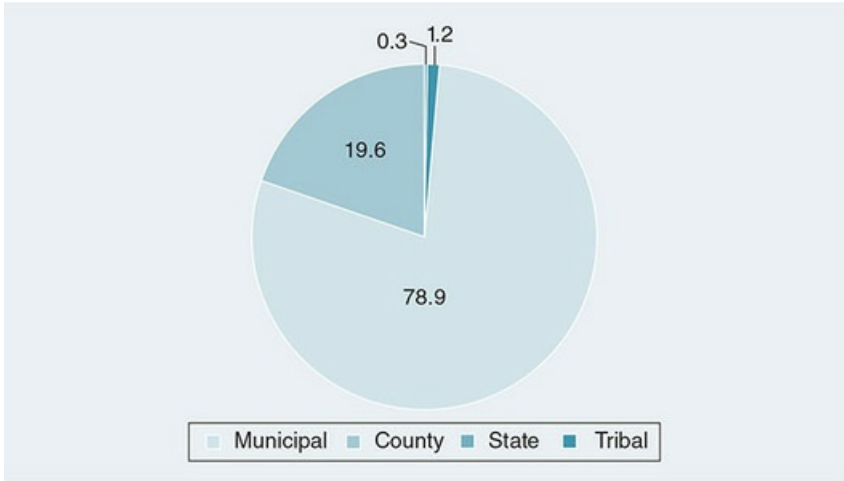
Learning Check 3.5

✓

Pie charts like the one in Figures 3.2 and 3.3 are made of percentages. Rates cannot be used for pie charts. Why is this true?

Data Sources 3.2 The Law Enforcement Management and Administrative Statistics Survey

The BJS conducts the Law Enforcement Management and Administrative Statistics (LEMAS) survey every 3 to 4 years. The sampling design involves two stages. First, all police agencies with 100 or more sworn personnel are included. Second, the BJS pulls a random sample of agencies with fewer than 100 officers. The agencies in the sample are sent surveys to fill out. The surveys capture agency-level data such as the number of sworn law-enforcement personnel an organization employs, the number of civilian personnel, whether the agency participates in community policing, whether the agency has specialized units, and so on. At this time, the 2013 LEMAS survey is the most recent wave available.

Categorical Variables: Bar Graphs

Like pie charts, bar graphs are meant to be used with categorical data; unlike pie charts, though, bar graphs can accommodate variables with many classes without damage to the charts' readability. Bar graphs are thus more flexible than pie charts. For variables with five or fewer classes, pie charts and bar graphs might be equally appropriate; when there are six or more classes, bar graphs should be used.

In Chapter 1, you learned that one of the reasons for the discrepancy between crime prevalence as reported by the UCR and the NCVS is that a substantial portion of crime victims do not report the incident to police. Truman and Morgan (2016) analyzed NCVS data and reported the percentage of people victimized by different crime types who reported their victimization to police. Figure 3.4 contains a bar graph illustrating the percentage of victims who contacted the police to report the crime. Bar graphs provide ease of visualization and interpretation. It is simple to see from Figure 3.4 that substantial portions of all types of victimizations are not reported to the police and that motor vehicle theft is the most reliably reported crime.

Figure 3.4 Percentage of Victimizations Reported to Police in 2015
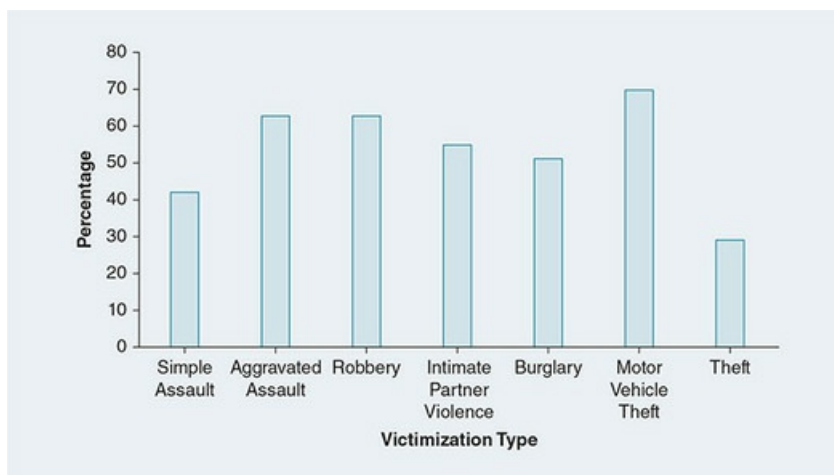
Rates can also be presented as bar graphs. Figure 3.5 is a bar graph of the rates in Table 3.6.

Figure 3.5 2015 Property Crime Rates (per 10,000) in Four States

A useful feature of bar graphs is that they can also be used to show the overlap between two variables. Figure 3.6 draws from LEMAS and the agency-type variable used in Figure 3.3. In the figure, the bars represent the number of agencies within each category that do and do not allow members of the public to report crimes through email or text message. Many agencies are adopting these electronic methods in hopes of encouraging people to report crimes by making it easier and more convenient for them to do so.

The frequencies in Figure 3.6 can also be turned into percentages and graphed like Figure 3.7. Now, instead of representing frequencies, the bars mark the percentage of agencies of each type that allow (or do not allow) people to report crimes through text or email.

Continuous Variables: Histograms

Histograms are for use with continuous data. Histograms resemble bar charts, with the exception that in histograms, the bars touch one another. In bar charts, the separation of the bars signals that each category is distinct from the others; in histograms, the absence of space symbolizes the underlying continuous nature of the data. Figure 3.8 contains a histogram showing the ages of Hispanic respondents to the 2011 PPCS (see Data Sources 2.1) who reported that they had called the police for help within the past 24 months.

Figure 3.6 Police Agencies' use of Email and Text for Crime Reporting, by Agency Type
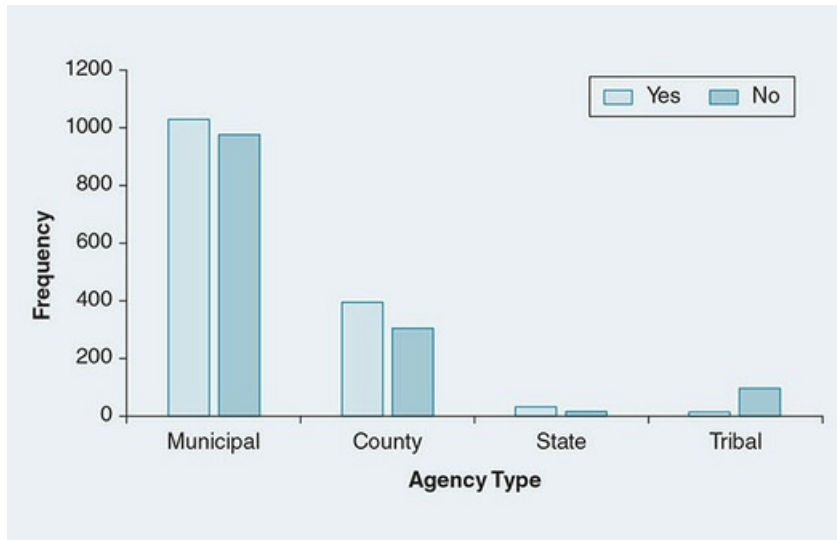
Figure 3.7 Police Agencies' use of Email and Text for Crime Reporting, by Agency Type

Figure 3.8 Age of Hispanic Respondents Who Called the Police in the Past 24 Months
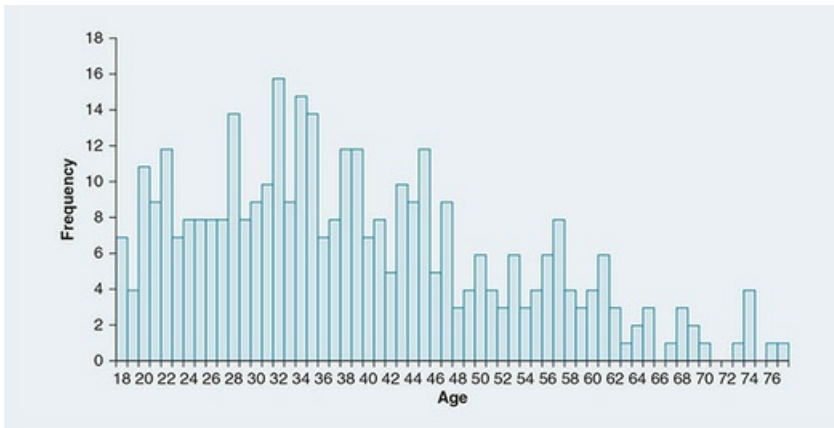
Research Example 3.2 Are Women's Violent-Crime Commission Rates Rising?

Women have historically experienced a very small arrest rate, much smaller than men's rate. Over the past few years, arrest rates for women have risen dramatically, even as crime itself has been declining or remaining stable. Some observers believe that women commit more crime today than they did in the past. Critics, however, argue that the change in the arrest rate is not caused by actual increases in women's criminality; instead, they say, it is because tough-on-crime policies have eroded the leniency that women used to receive from police. Who is right? Has women's violent-crime involvement truly risen? Or are women just more likely to be arrested today? Steffensmeier, Zhong, Ackerman, Schwartz, and Agha (2006) set out to answer these questions. The researchers used two data sources. First, they relied on the UCR to track the arrest rates for women and men across different types of violent crime from 1980 through 2003. Second, they used the NCVS to measure violent victimizations perpetrated by women and men during this same period. The authors knew that the UCR would show a rise in women's arrest rates, at least for certain kinds of offenses. The real question was whether the NCVS would also show a rise. The authors displayed their results in histograms.

The top histogram displays UCR arrest data and the bottom one shows NCVS victimization data. The shorter bars in each one represent the percentage of arrestees and offenders, respectively, who were female. Together, these graphs show that even though female arrests for assault have risen, their participation in criminal assaults has not; assaults perpetrated by women have remained constant. The authors concluded that women are not committing violent crimes at higher rates now than in the past; their increasing arrest rates are caused by criminal-justice policies that lead police to arrest women in situations where, a few decades ago, they would have shown them leniency.

Note: Includes aggravated and simple assaults.

Assault Index (NCVS)

Male Rates   Female Rates   △ Female Percentages
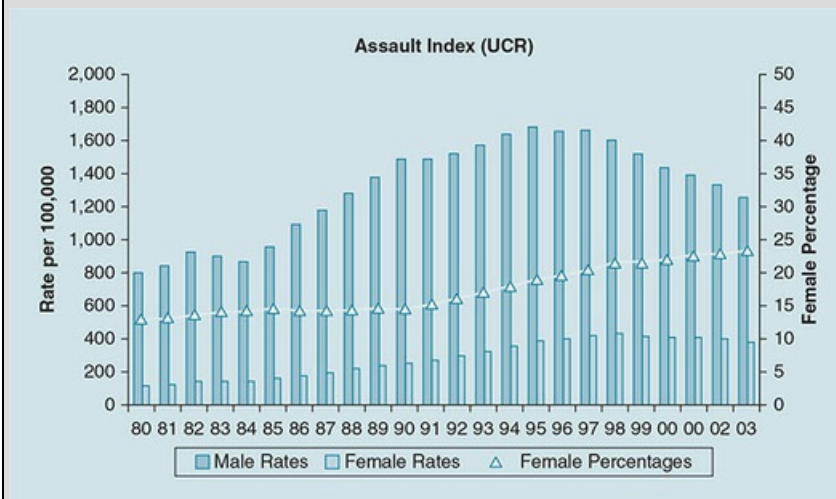
Note: Includes aggravated and simple assaults.
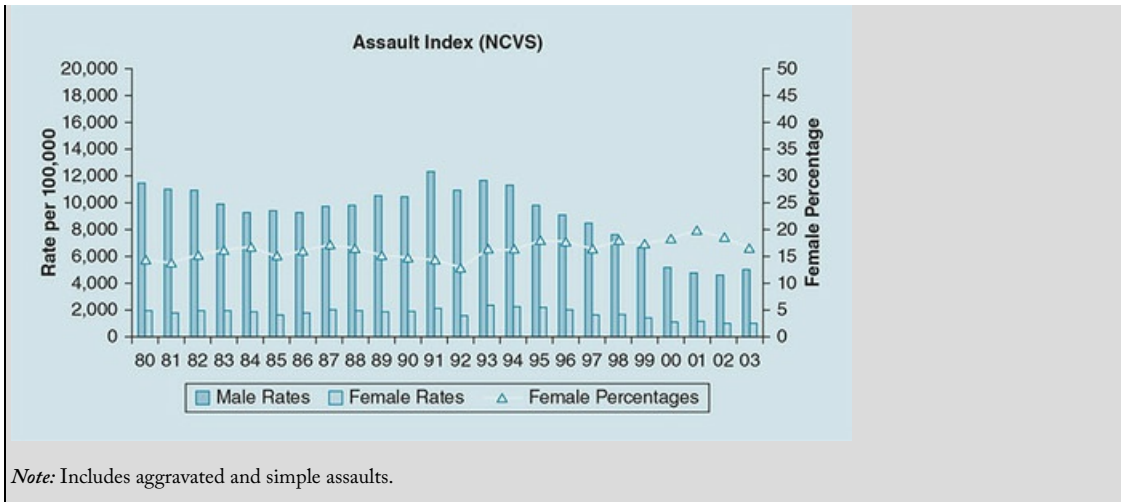
Histograms can also feature percentages instead of frequencies. For example, we can use a histogram to display the average daily population of small jails using data from the COJ facilities (Data Sources 3.1). Figure 3.9 shows the results. The horizontal axis lists the numbers of inmates each jail reported containing, and the vertical axis represents the percent of all facilities that housed each quantity. A glance at the histogram allows us to see that while small jails are fairly diverse in size, they are somewhat clustered toward the smaller end.

Figure 3.9 Number of Inmates in Small Jails
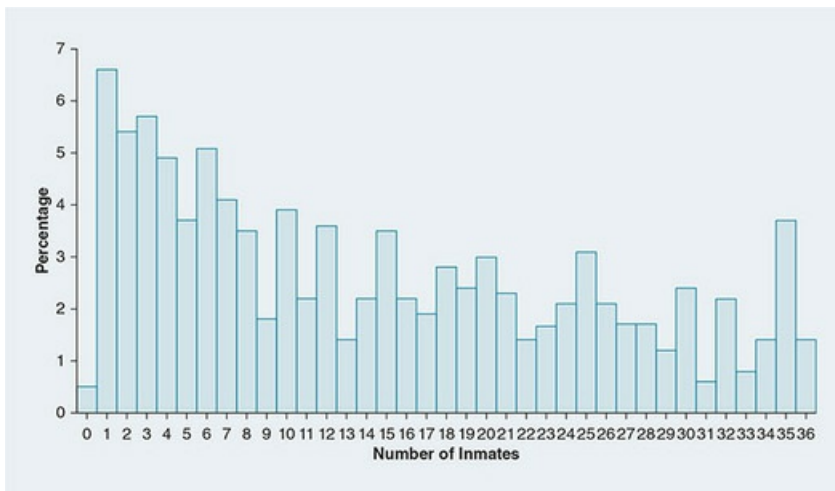
Continuous Variables: Frequency Polygons

Frequency polygons are an alternative to histograms. There is no right or wrong choice when it comes to deciding whether to use a histogram or a frequency polygon with a particular continuous variable; the best strategy is to mix it up a bit so that you are not using the same chart type repeatedly. Figure 3.10 contains the frequency polygon for data similar to that used in Figure 3.8 except this time it includes the ages of non-Hispanic respondents instead of Hispanic respondents. Frequency polygons are created by placing a dot in the places where the tops of the bars would be in a histogram and then connecting those dots with a line.

Another type of continuous variable that could be graphed using a frequency polygon is the number of female sergeants in police departments and sheriffs' offices. Nationwide, police agencies are seeking not merely to recruit and retain female officers but also to promote them to ranks with supervisory and management responsibilities. Figure 3.11 shows the number of female sergeants in sheriffs' offices, according to the LEMAS survey (Data Sources 3.2).

Figure 3.10 Age of Non-Hispanic Respondents Who Called the Police in the Past 24 Months
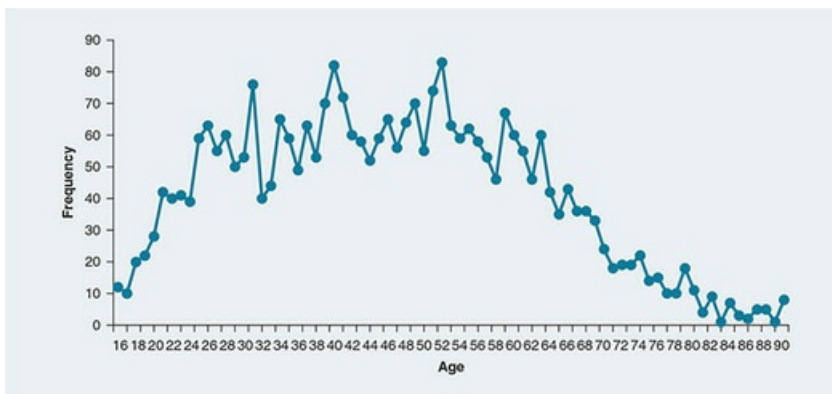
Figure 3.11 Number of Female Sergeants in Sheriffs' Offices
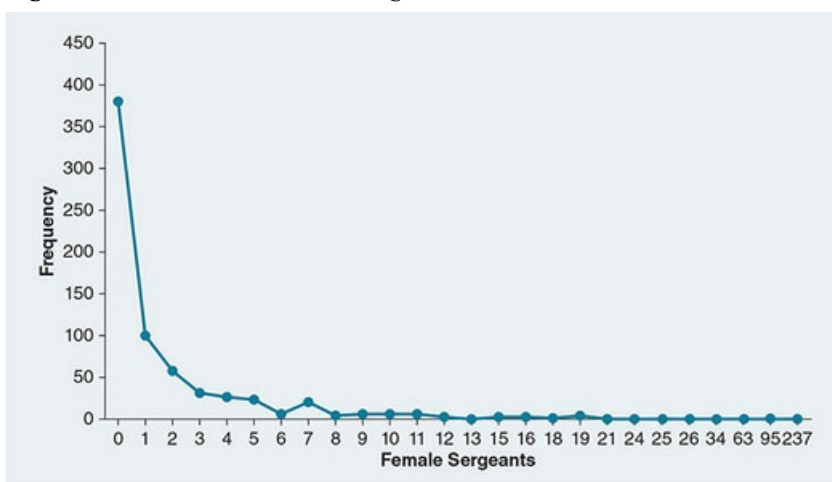
Longitudinal Variables: Line Charts

People who work with criminal justice and criminology data often encounter **longitudinal variables**. Longitudinal variables are measured repeatedly over time. Crime rates are often presented longitudinally as a means of determining **trends**. Line graphs can make it easy to discern trends.

> **Longitudinal variables**: Variables measured repeatedly over time.
>
> **Trends**: Patterns that indicate whether something is increasing, decreasing, or staying the same over time.

Figure 3.12 shows a line graph of data from the Uniform Crime Reports measuring the annual number of hate-crime incidents from 1996 to 2015. Figure 3.13 shows the annual percentage of all hate crimes that are motivated by sexual-orientation bias. Together, these two line charts show two trends. First, total hate-crime incidents have declined slightly in recent years. Second, despite the downward trend in total incidents, the percentage of incidents that are based on the victims' sexual orientation has risen steadily, albeit with a small drop from 2011 to 2012.

Figure 3.12 Annual Number of Hate-Crime Incidents, 1996–2015
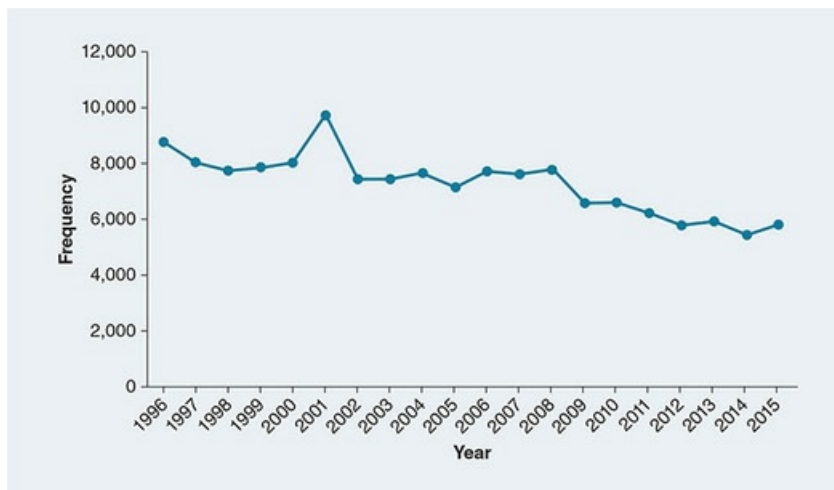
Line charts can employ percentages, as well. We might ask not merely the number of hate-crime incidents that occur annually, but the percentage of these incidents motivated by sexual orientation bias. Figure 3.13 shows this graph.

Together, Figures 3.12 and 3.13 reveal that although hate crimes have declined over the past 20 years, there has been a slight rise in the percentage of these crimes arising from prejudice against members of the gay and lesbian individuals.

Figure 3.13 Percentage of Hate-Crime Incidents Motivated by Sexual Orientation Bias, 1996–2015
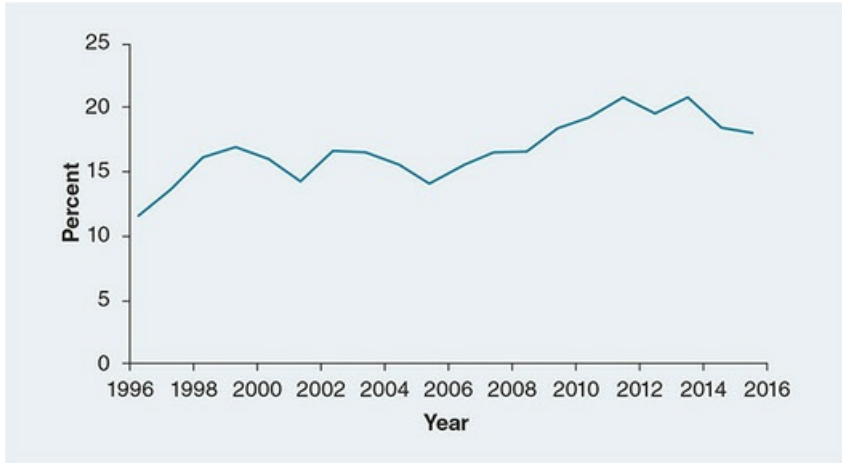
Grouped Data

The overarching purpose of a frequency distribution, chart, or graph is to display data in an accessible, readily understandable format. Sometimes, though, continuous variables do not lend themselves to tidy displays. Consider Table 3.9's frequency distribution for the amount of money, in dollars per person, that local governments in each state spent on criminal justice operations (Morgan, Morgan, & Boba, 2010; Data Sources 3.3). Figure 3.14 displays a histogram of the data. You can see that neither the frequency distribution nor the histogram is useful; there are too many values, and most of the values occur only once in the data set. There is no way to discern patterns or draw any meaningful conclusion from these data displays.

Data Sources 3.3 CQ Press's State Factfinder Series

The Factfinder Series' Crime Rankings are compilations of various crime and crime-related statistics from the state and local levels. These volumes are comprehensive reports containing data derived from the Federal Bureau of Investigation (FBI), BJS, U.S. Census Bureau, and U.S. Drug Enforcement Administration. The data used here come from Morgan et al. (2010).

Table 3.9	Per Capita Local Government Expenditures on Criminal Justice, Ungrouped				
Dollars per Capita	f	Dollars per Capita	f	Dollars per Capita	f
123	1	266	1	357	2
164	1	267	2	358	1
169	1	268	1	374	1
188	1	272	1	375	1
209	1	273	1	387	1
214	1	281	1	390	1
215	2	287	1	414	1
220	1	292	1	427	1
230	1	306	2	433	1
246	1	310	1	439	1
248	1	321	1	462	1
251	1	345	1	500	1
252	1	348	1	552	1
255	1	351	1	603	1
258	1	354	1	623	1
263	1				N = 50

Figure 3.14 Per Capita Local Government Expenditures on Criminal Justice, Ungrouped
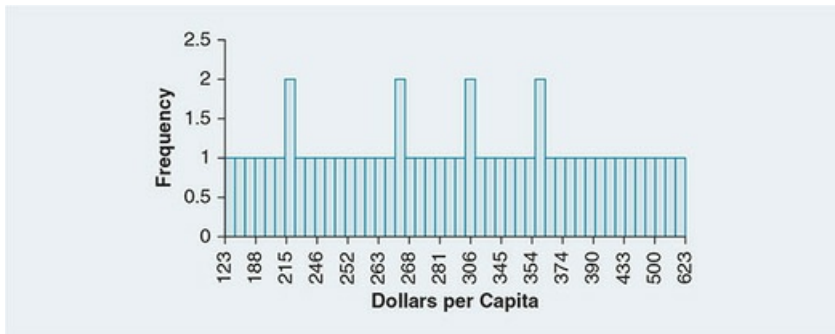
Grouping the data can provide a solution to this problem by transforming a continuous variable (either interval or ratio) into an ordinal one. There are several steps to grouping. First, find the range in the data by subtracting the smallest number from the highest. Second, select the number of intervals you want to use. This step is more art than science; it might take you a bit of trial and error to determine the number of intervals that is best for your data.

The ultimate goal is to find a middle ground between having too few and too many intervals—too few can leave your data display flat and uninformative, whereas too many will defeat the whole purpose of grouping.

Third, determine interval width by dividing the range by the number of intervals. This step is probably best illustrated in formulaic terms:

$$width = \frac{range}{intervals}$$

This will often produce a number with a decimal, so round up or down depending on your reasoned judgment as to the optimum interval width for your data. Fourth, construct the stated class limits by starting with the smallest number in the data set and creating intervals of the width determined in Step 3 until you run out of numbers. Finally, make a new frequency (*f*) column by counting the number of people or objects within each stated class interval.

Let us group the legal expenditure data in Table 3.9. First, we need the range:

Range = 623 − 123 = 500

Now we choose the number of intervals we want to use. With a range as large as 500, it is advisable to select relatively few intervals so that each interval will encompass enough raw values to make it meaningful. We will start with 10 intervals. The next step is to compute the interval width. Using the formula from above,

$$width = \frac{500}{10} = 50$$

Each interval will contain 50 raw scores. Now the stated class limits can be constructed. Take a look at the left-hand column in Table 3.10. There are three main points to keep in mind when building stated class limits. The stated limits must be *inclusive*—in this example, the first interval contains the number 123, the

number 172, and everything in between. They also must be *mutually exclusive* and *exhaustive*. Once the stated class limits have been determined, the frequency for each interval is calculated by summing the number of raw data points that fall into each stated class interval. The sum of the frequency column in a grouped distribution should equal the sum of the frequencies in the ungrouped distribution.

You can see that Table 3.10 is much neater and more concise than Table 3.9. It is more condensed and easier to read. Where you will really see the difference, though, is in the histogram. Compare Figure 3.15 to Figure 3.14. Quite an improvement! It has a real shape now. This demonstrates the utility of data grouping.

Table 3.10 Per Capita Local Government Expenditures on Criminal Justice, Grouped	
Stated Class Limits	*f*
123–172	3
173–222	6
223–272	13
273–322	8
323–372	7
373–422	5
423–472	4
473–522	1
523–572	1
573–622	1
623–672	1
	$N = 50$

Figure 3.15 Per Capita Local Government Expenditures on Criminal Justice, Grouped
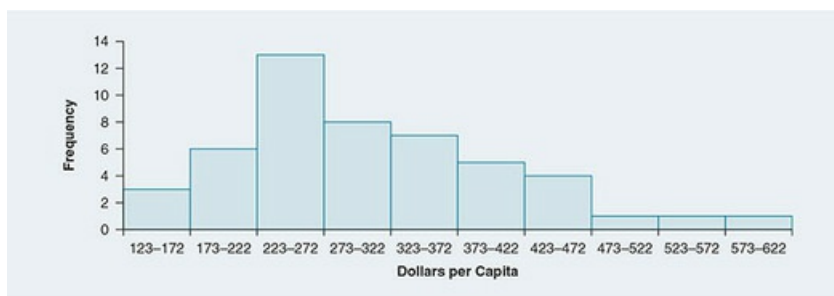
Another variable that could benefit from grouping is the daily population of small jails. Although the histogram in Figure 3.9 is not too bad to look at, it is rather busy. Grouping could clean the distribution up and give it a more streamlined appearance. This would also be helpful if you were writing a report and had space constraints.

Table 3.11	Number of Inmates Housed in Small Jails, Ungrouped		
Number	f	Number, cont'd	f
0	4	19	19
1	51	20	23
2	42	21	18
3	44	22	11
4	38	23	13
5	29	24	16
6	40	25	24
7	32	26	16
8	27	27	13
9	14	28	13
10	30	29	9
11	17	30	19
12	28	31	5
13	11	32	17
14	17	33	6
15	27	34	11
16	17	35	29
17	15	36	11
18	22		N = 778

We will group the data in Table 3.11 using the three-step process described earlier. First, the range is 36 – 0 = 36. Second, with a range of 36, let us try 8 intervals. As noted, there are no specific rules guiding the selection of intervals. Eight seems like a good number here because it will create enough intervals so that the distribution does not lose its shape (a danger when there are too few intervals in the grouped distribution) but will significantly reduce the complexity of the ungrouped data and make for a clean grouped histogram. Third,

the width is $\dfrac{36}{8} = 4.75$, which we will round to 5. Each interval will contain five numbers.

The stated class limits will start with zero and continue until the highest number (here, 36) has been included

in an interval. The left column in Table 3.12 shows the limits. To calculate the frequency for each group, we sum the frequencies for each of the numbers in each one. For example, the frequency for the "0 – 4" group is 4 + 51 + 42 + 44 + 38 = 179. The frequency for the "5 – 9" group is 29 + 40 + 32 + 27 + 14 = 142. We complete this process until all the ungrouped data have been shrunk down into the stated class limits.

Table 3.12 Number of Inmates Housed in Small Jails, Grouped	
Stated Class Limits	**f**
0–4	179
5–9	142
10–14	103
15–19	100
20–24	81
25–29	75
30–34	58
35–39	40
	$N = 778$

After finishing the frequency table (Table 3.12), we can display the grouped data in a histogram. Refer to Figure 3.16. A comparison between Figures 3.16 and 3.9 reveals the benefits of grouping. The bars in Figure 3.9 jump around a lot, but the bars in Figure 3.16 show a distinct and constant decline that quickly conveys information about inmate populations in small jails. This figure would allow an audience to easily see that small jails tend to hold only a few inmates.

Figure 3.16 Number of Inmates Housed in Small Jails, Grouped
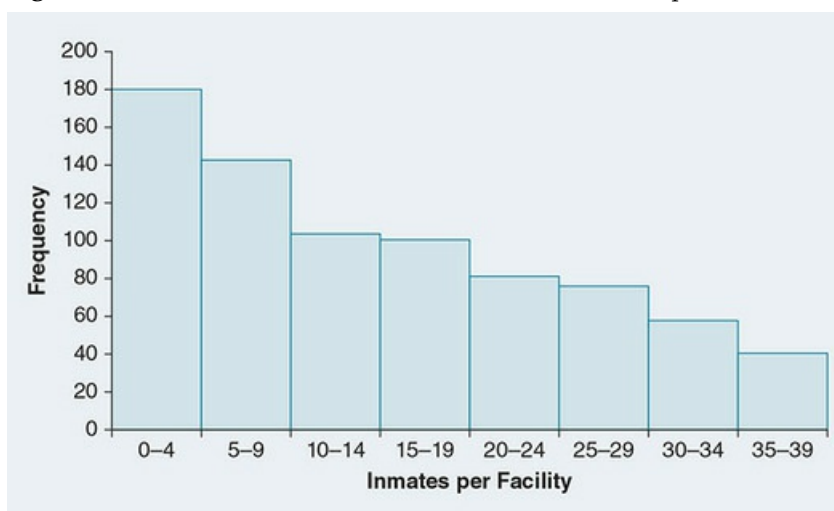
SPSS

This is our first encounter with SPSS. There are a few preliminary things you should know before you start working with data. First, GIGO alert! Recall that when garbage goes in, the output is also garbage. Using the wrong statistical technique will produce unreliable and potentially misleading results. Statistical software programs generally do not alert you to errors of this sort; they will give you output even if that output is wrong and useless. It is your responsibility to ensure that you are using the program correctly. Second, pay attention to the SPSS file extension. The *.sav* extension signifies an SPSS data file: Anytime you see this extension, you know the file contains data in SPSS format. SPSS is the only program that will open a file with the *.sav* extension, so make sure you are working on a computer equipped with SPSS.

To obtain a frequency distribution, click on *Analyze → Descriptive Statistics → Frequencies*, as shown in Figure 3.17. Select the variable you want from the list on the left side and either drag it to the right or click the arrow to move it over. For this illustration, we will use the LEMAS data displayed in the bar chart in Figure 3.3. This variable captures the number (and corresponding percentage) of each type of policing agency in the United States. Following the steps depicted in Figure 3.17 will produce the table shown in Figure 3.18.

The SPSS program can also be used to produce graphs and charts. The Chart Builder (accessible from the *Graphs* drop-down menu) allows you to select a chart type and then choose the variable you want to use. The SPSS Chart Builder requires that the level of measurement for each variable be set properly. SPSS will not permit certain charts to be used with some levels of measurement. Before constructing graphs or charts, visit the *Measure* column in the Variable View and make sure that continuous variables are marked as *Scale* and that nominal and ordinal variables are designated as such. To begin the chart or graph, click *Graphs → Chart Builder*. This will produce the dialog boxes shown in Figure 3.19, where a pie chart has been selected from the list on the bottom left. In the Element Properties box on the right, you can change the metric used in the chart. Counts are the default, but for reasons we covered in this chapter percentages are often preferable to raw counts.

Figure 3.17 Running Frequencies in SPSS

Figure 3.18 SPSS Frequency Output

Frequencies

Statistics

Agency Type

N	Valid	2826
	Missing	0

Agency Type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Municipal	2036	72.0	72.0	72.0
	County	717	25.4	25.4	97.4
	State	50	1.8	1.8	99.2
	Tribal	23	.8	.8	100.0
	Total	2826	100.0	100.0	
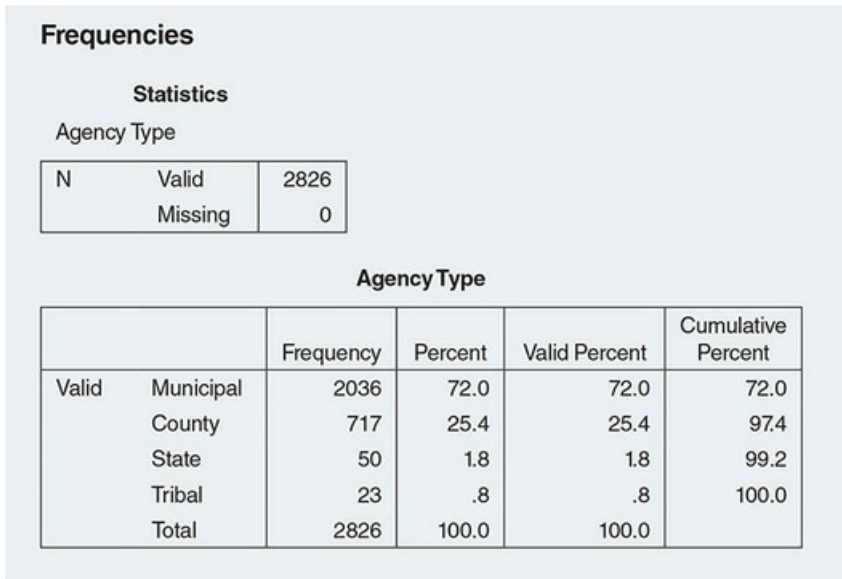
Figure 3.19 Using the SPSS Chart Builder to Create a Pie Chart
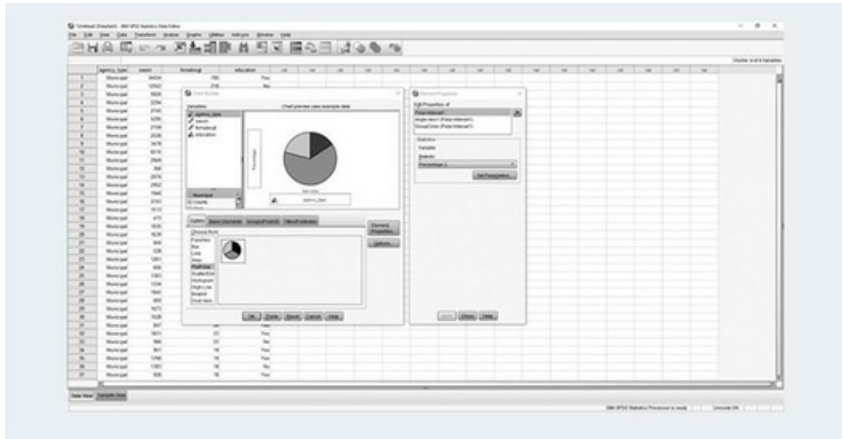
The Chart Builder can be used for bar graphs, too. Figure 3.20 displays the dialog box you will see if you select Bar from the list on the lower left. For this graph, we will stick with the count default. Figures 3.21 and 3.22 show the pie chart and bar graph output, respectively.

Figure 3.20 Using the SPSS Chart Builder to Create a Bar Graph
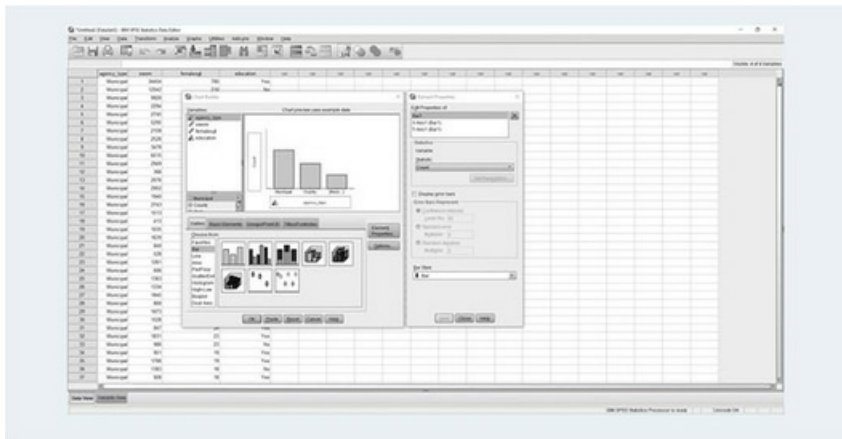
Figure 3.21 SPSS Pie Chart
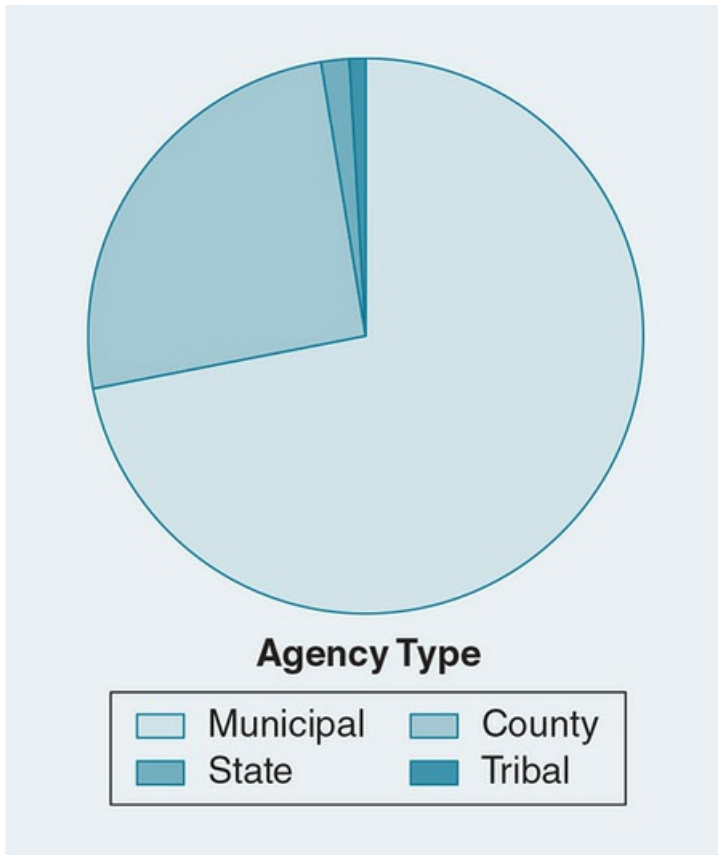
Figure 3.22 SPSS Bar Graph
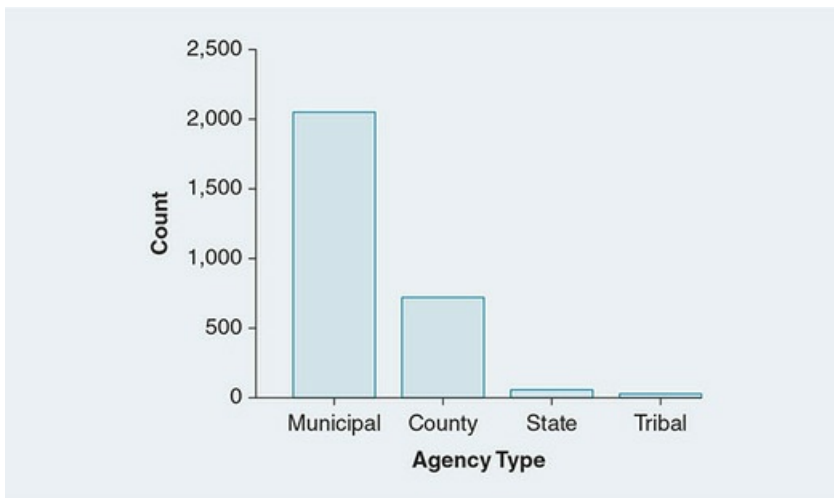
Bivariate contingency tables are available in SPSS, too. Locate this option through *Analyze → Descriptive Statistics → Crosstabs*. Select the variable that you would like to put in the rows of the table and the one you would like to place in the column. (Recall that there are no rules governing this choice, although it is common to put the variable you conceptualize as the independent variable or predictor in the rows and the one you think is a dependent variable or outcome in the columns.) For this example, we will again work with the LEMAS variable capturing police agency types. The second variable we will use comes from the same LEMAS survey and measures whether each agency offers incentive pay (i.e., raises for educational attainment intended to encourage officers to obtain college degrees). It makes more sense to think that agency type would

affect the offering of education incentives than to think the offering of educational incentives would affect an agency's jurisdiction (an unrealistic scenario). Therefore, the agency type variable will go into the rows and the incentive variable will form the columns. We obtain percentages by clicking the *Cells* button and selecting row or column percentages. For present purposes, we will opt for row percentages so SPSS will display the percentage of each type of agency that does (and does not) offer educational incentives. Figure 3.23 shows the output.

Finally, SPSS can be used to transform raw numbers into rates. To demonstrate this, we will keep working with LEMAS and compute the number of municipal police officers per 1,000 residents for each of the city police departments in the sample. Clicking on the *Transform* button at the very top of the SPSS data screen and then clicking *Compute* will produce the *Compute Variable* box pictured in Figure 3.24.

Figure 3.23 SPSS Contingency Table Output (With Row Percentages)

Agency Type *Agency offers incentives for educational achievement? Crosstabulation

| | | | Agency offers incentives for educational achievement? | | |
			Yes	No	Total
Agency Type	Municipal	Count	944	1068	2012
		% within Agency Type	46.9%	53.1%	100.0%
	County	Count	228	481	709
		% within Agency Type	32.2%	67.8%	100.0%
	State	Count	19	26	45
		% within Agency Type	42.2%	57.8%	100.0%
	Tribal	Count	8	15	23
		% within Agency Type	34.8%	65.2%	100.0%
Total		Count	1199	1590	2789
		% within Agency Type	43.0%	57.0%	100.0%

Figure 3.24 Creating Rates in SPSS
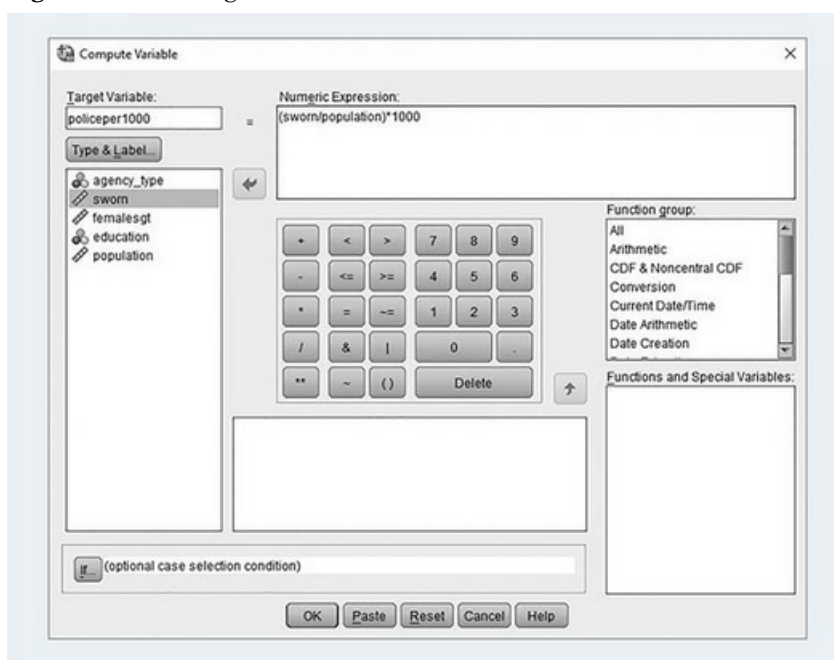
In the *Target Variable* box (see Figure 3.24), type the name you want to give your new variable; in the present

example, the variable will be called *policeper1000*. In the "Numeric Expression" area, type the equation you wish SPSS to follow to create the new variable. Here, the portion of the equation reading "(police/population)" tells the program to begin the computation by dividing the total number of police in a jurisdiction by that jurisdiction's population size, and the "*1000" instructs SPSS to then multiply the results by 1,000. Click *OK*, and a new variable will appear at the very end of the data set. This is your rate variable.

Learning Check 3.7

There is a good reason for using rates (such as the number of officers per 1,000 residents) to determine the size of a police agency rather than simply using the number of officers employed by that organization. Can you explain why rates are better than raw counts in this instance?

The chapter review questions contain directions for accessing data sets that you can use to practice constructing charts and transforming variables into rates in SPSS. Play around with the data! Familiarize yourself with SPSS; we will be visiting it regularly throughout the book, so the more comfortable you are with it, the better prepared you will be. You do not have to worry about ruining the data set—if you make a mistake, just click the "Undo" button or hit "Don't save" when you exit the program, and the file will be as good as new.

Chapter Summary

This chapter discussed some of the most common types of graphs and charts. Frequency distributions offer basic information about the number of times certain characteristics appear in a data set. Frequencies are informative and can convey valuable information; however, numbers are often difficult to interpret when they are in a raw format. Proportions and percentages offer a way to standardize frequencies and make it easy to determine which characteristics occur more often and which less often. Rates are another option for enhancing the interpretability of frequencies. Rates are generally multiplied by a number such as 1,000, 10,000, or 100,000.

Graphs and charts portray this same information—frequencies, proportions, percentages, and rates—using pictures rather than numbers. Pictorial representations are more engaging than their numerical counterparts and can capture audiences' interest more effectively. Pie charts can be used with categorical variables that have five or fewer classes and are in percentage format. Bar graphs are useful for categorical variables with any number of classes. They can be made from frequencies, proportions, percentages, or rates. For continuous data, histograms and frequency polygons can be used to graph frequencies, proportions, or percentages. Line graphs are useful for longitudinal data. Finally, some continuous variables that do not have a clear shape and are difficult to interpret in their raw form can be grouped. Grouping transforms continuous variables into ordinal ones. Histograms can be used to display data that have been grouped.

It is good to diversify a presentation by using a mix of pie charts, bar graphs, histograms, frequency polygons, and line charts. Simplicity and variety are the keys to a good presentation. Simplicity ensures that your audience can make sense of your data display quickly and easily. Variety helps keep your audience engaged and interested. Good data displays are key to summarizing data so that you and others can get a sense for what is going on in a data set.

Thinking Critically

1. Suppose that in the past year, there were 15 incidents of violence committed by inmates against correctional staff in a nearby prison. An administrator from the prison calls you to ask for an assessment of the severity of the problem of violence within the prison (i.e., how prevalent inmate-on-staff violence is within this facility). What additional data will you request from the prison administrator and what calculations will you perform? Think about rates, percentages, and longitudinal trends, as well as any additional analyses that would help you understand this issue.

2. Consider three settings in which a researcher would potentially give presentations that summarize the results of data analyses: academic conferences, presentations to practitioners (such as administrators from police agencies or jails), and presentations to the community (such as community-based agencies or city councils). Based on the different audiences in each of these three settings, explain what types of data displays you would use. In other words, how would you tailor your presentation to make sure it was appropriate for your audience? Identify the displays you would use and explain your choices.

Review Problems

1. The following table contains data from the BJS's State Court Processing Statistics, which includes data on felony defendants in large urban counties in 2009 (Reaves, 2013). The variable is *most serious arrest charge*, which captures the most severe of the offenses for which defendants were arrested.

Most Serious Arrest Charge	f
Violent Offense	13,938
Property Offense	16,241
Drug Offense	18,220
Public-Order Offense	7,504
	N = 55,903

   1. Construct columns for proportion, percentage, cumulative frequencies, cumulative proportions, and cumulative percentages.
   2. Identify the types of charts or graphs that could be used to display this variable.
   3. Based on your answer to (b), construct a graph or chart for this variable using percentages.

2. The following table contains data from the State Court Processing Statistics showing the number of urban felony defendants sentenced to prison time after conviction for different types of property offenses (Reaves, 2013).

Offense	f
Burglary	1,191
Larceny/Theft	901
Motor Vehicle Theft	329
Forgery	194
Fraud	256
Other Property Offense	365
	N = 3,236

   1. Construct columns for proportion, percentage, cumulative frequencies, cumulative proportions, and cumulative percentages.
   2. Identify the types of charts or graphs that could be used to display this variable.
   3. Based on your answer to (b), construct a graph or chart for this variable using percentages.

3. The following table contains data from the COJ facilities showing the number of inmates on work release in medium-sized facilities that offer this program for inmates.
   1. Choose the appropriate graph type for this variable and construct that graph using frequencies.
   2. Group this variable using 10 intervals.
   3. Choose an appropriate graph type for the grouped variable and construct that graph using frequencies.

Inmates on Work Release	f	Inmates on Work Release	f	Inmates on Work Release	f
0	50	11	4	24	3
1	39	12	12	25	3
2	33	13	2	27	2
3	28	14	9	28	1
4	34	15	6	29	1
5	15	17	3	30	4
6	20	18	2	31	1
7	6	20	6	40	1
8	11	21	1	42	1
9	5	22	3		$N = 324$
10	17	23	1		

4. The following table contains PPCS data on the ages of respondents who said that the police had requested consent to search their car during their most recent traffic stop.

Age	f	Age	f	Age	f
16	1	29	1	46	2
17	7	30	3	47	3
18	4	31	4	49	2
19	9	32	2	50	3
20	8	34	2	51	1
21	7	35	1	52	1
22	6	37	1	53	1
23	3	38	2	54	1
24	5	39	3	55	1
25	7	41	1	57	1
26	5	42	4	58	1
27	6	44	3	59	1
28	5	45	4		N = 122

1. Choose the appropriate graph type for this variable and construct that graph using frequencies.
2. Group this variable using six intervals.
3. Choose an appropriate graph type for the grouped variable and construct that graph using frequencies.

5. The General Social Survey (GSS) asks respondents whether they think that marijuana should be legalized. The following table contains the percentage of respondents who supported legalization in each wave of the GSS from 1990 to 2014. Construct a line graph of these data, and then interpret the longitudinal trend. Does support for marijuana legalization appear to be increasing, decreasing, or staying the same over time?

Year	pct	Year	pct
1990	17	2002	36
1991	19	2004	36
1993	23	2006	37
1994	24	2008	40
1996	27	2010	48
1998	29	2012	48
2000	34	2014	55

6. The following table displays data from the official website of the U.S. courts (www.uscourts.gov) on the number of wiretap authorizations issued by state and federal judges every year from 1997 to 2016. Construct a line graph of the data, and then interpret the longitudinal trend. Have wiretap authorizations been increasing, decreasing, or staying the same over time?

Year	f	Year	f
1997	1,186	2007	2,208
1998	1,329	2008	1,891
1999	1,350	2009	2,376
2000	1,190	2010	3,194
2001	1,491	2011	2,732
2002	1,358	2012	3,395
2003	1,442	2013	3,576
2004	1,710	2014	3,554
2005	1,773	2015	4,148
2006	1,839	2016	3,168

7. The following table contains data on the number of violent crimes that occurred in six cities during 2015. The table also displays each city's population.

City	Violent Crimes	Population
Birmingham, AL	3,707	212,291
Portland, ME	215	66,816
San Francisco, CA	6,710	863,782
Tampa, FL	2,298	364,383
Ann Arbor, MI	228	118,730
Atlanta, GA	5,203	464,710

1. Compute the rate of violent crime per 1,000 city residents in each city.
2. Select the appropriate graph type for this variable and construct that graph using rates.

8. The following table contains data on the number of property crimes that occurred in six cities during 2015. The table also displays each city's population.

City	Property Crimes	Population
Phoenix, AZ	54,456	1,559,744
Hartford, CT	5,468	124,553
Bellingham, WA	4,067	83,976
Cincinnati, OH	16,446	298,478
Pinehurst, NC	91	15,667
St. Louis, MO	20,028	317,095

1. Compute the rate of property crime per 1,000 city residents in each city.
2. Select the appropriate graph type for this variable and construct that graph using rates.

9. The website for this chapter (http://www.sagepub.com/gau) contains a data set called *Work Release for Chapter 3.sav*. These are the data from Review Problem 3. Use the SPSS Chart Builder to construct a frequency histogram showing the number of inmates on work release in facilities offering this service.

10. The file *City Police for Chapter 3.sav* contains data from the 2013 LEMAS, narrowed to municipal (town and city) departments. This data file contains the variable *forfeiture*, which indicates whether each agency includes asset forfeiture revenues as a funding source in its operating budget. Run a frequency analysis to find out the percentage of departments that do and do not include forfeiture in their operating budgets.

11. In the *City Police for Chapter 3.sav file*, there is a variable called *sworn* that displays each police department's number of full-time sworn officers and a variable called *population* that records the city population served by each department. Use these two variables and the compute function to calculate the number of officers per 1,000 population for each department.

12. The website (www.sagepub.com/gau) also features a data file called *Hate Crimes for Chapter 3.sav*. This is the same data set used in the in-text demonstration of line graphs. Use the SPSS Chart Builder to construct a line graph mirroring the one in the text.

13. The website (www.sagepub.com/gau) contains a data set called *Crime Attitudes for Chapter 3.sav*. This is a trimmed version of the 2014 GSS. One of the variables contained in the data set is *courts*, which measures respondents' opinions about how well courts punish people who commit crimes. Select an appropriate graph or chart and create it using SPSS.

14. The *Crime Attitudes for Chapter 3.sav* file also contains the variable *marijuana*, which captures respondents' beliefs about whether marijuana should be made legal. Use SPSS to construct a contingency table with *courts* in the rows and *marijuana* in the columns. Include row percentages in your table. Provide a brief written explanation of the percentages and apparent overlap between these two variables.

Key Terms

Cell 37

Univariate 37

Frequency 37

Proportion 38

Percentage 38

Cumulative 40

Contingency table 43

Bivariate 43

Classes 47

Longitudinal variables 56

Trends 37

Glossary of Symbols and Abbreviations Introduced in This Chapter

f	Frequency
p	Proportion
N	Sample size
pct	Percentage
cf	Cumulative frequency
cp	Cumulative proportion
$cpct$	Cumulative percentage

Chapter 4 Measures of Central Tendency

Learning Objectives

- Define the three types of data distribution shapes.
- Identify the shape of a distribution based on a comparison of the mean and median.
- Describe the mode, the median, and the mean.
- Explain which level(s) of measurement each measure of central tendency can be used with.
- Identify, locate, or calculate the mode, the median, and the mean for a variety of variables and variable types.
- Explain the difference between the two mean formulas and correctly use each one on the appropriate types of data.
- Explain deviation scores and their relationship to the mean.
- Use SPSS to produce the mode, the median, and the mean.

Criminal justice and criminology researchers and practitioners are often interested in averages. Averages offer information about the centers or middles of distribution. They indicate where data points tend to cluster. This is important to know. Consider the following questions that might be of interest to a researcher or practitioner:

1. What is the most common level of educational attainment among police officers?
2. How does the median income for people living in a socioeconomically disadvantaged area of a certain city compare to that for all people in the city?
3. What is the average violent crime rate across all cities and towns in a particular state?
4. Do prison inmates, in general, have a lower average reading level compared to the general population?

All of these questions make some reference to an average, a middle point, or, to use a more technical term, a **measure of central tendency**. Measures of central tendency offer information about where the bulk of the scores in a particular data set are located. A person who is computing a measure of central tendency is, in essence, asking, "Where is the middle?"

> **Measures of central tendency**: Descriptive statistics that offer information about where the scores in a particular data set tend to cluster. Examples include the mode, the median, and the mean.

Averages offer information about the normal or typical person, object, or place in a sample. A group of people with an average age of 22, for instance, probably looks different from a group averaging 70 years of age. Group averages help us predict the score for any individual within that group. Suppose in two samples of people, the only information you have is that one group's average weight is 145 pounds and that the other's is 200 pounds. If someone asked you, "How much does an individual person in the first group weigh?" your response would be, "About 145 pounds." If you were asked, "Who weighs more, a person randomly selected from the first group or from the second group?" you would respond that the person from the second group is probably the heavier of the two. Of course, you do not know for sure that you are correct; there might be people in the first group who are heavier than some people in the second group. The average, nonetheless, gives you predictive capability. It allows you to draw general conclusions and to form a basic level of understanding about a set of objects, places, or people.

Measures of central tendency speak to the matter of distribution shape. Data distributions come in many

different shapes and sizes. [Figure 4.1](#) contains data from the Police–Public Contact Survey (PPCS; see [Data Sources 2.1](#)) showing the ages of non-Hispanic respondents who reported having called the police for help within the past 24 months. This is the same variable used in the frequency polygon shown in [Figure 3.10](#) in the [previous chapter](#). The shape this variable assumes is called a **normal distribution**. The normal curve represents an even distribution of scores. The most frequently occurring values are in the middle of the curve, and frequencies drop off as one traces the number line to the left or right. Normal distributions are ideal in research because the average is truly the best predictor of the scores for each case in the sample, since the scores cluster around that value.

> **Normal distribution**: A set of scores that clusters in the center and tapers off to the left (negative) and right (positive) sides of the number line.

Standing in contrast to normal curves are skewed distributions. Skew can be either positive or negative. The distribution in [Figure 4.2](#) contains the data from the Census of Jails (COJ; see [Data Sources 3.1](#)) showing the number of inmate-on-staff assaults each jail reported experiencing in the past year. The distribution in [Figure 4.2](#) manifests what is called a **positive skew**. Positively skewed data cluster on the left-hand side of the distribution, with extreme values in the right-hand portion that pull the tail out toward the positive side of the number line. Positively skewed data are common in criminal justice and criminology research.

> **Positive skew**: A clustering of scores in the left-hand side of a distribution with some relatively large scores that pull the tail toward the positive side of the number line.

Figure 4.1 Ages of Non-Hispanic Respondents Who Called the Police in the Past 24 Months
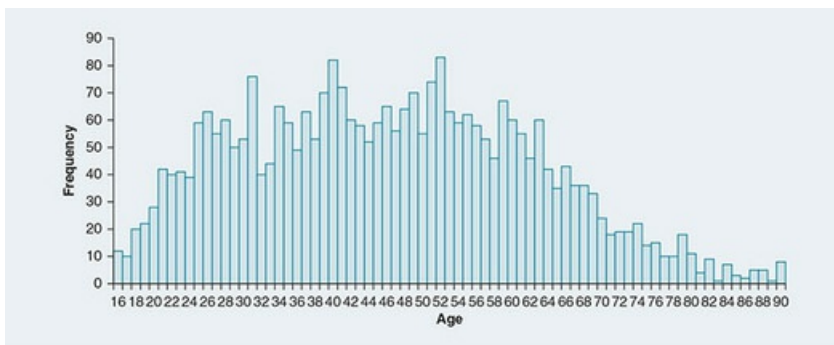
Figure 4.2 Number of Inmate Assaults on Jail Staff
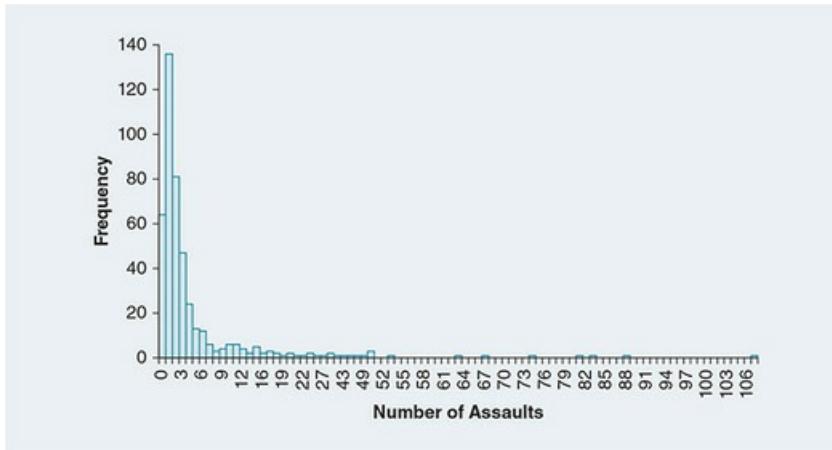
[Figure 4.3](#) shows 2014 General Social Survey (GSS; see [Data Sources 2.2](#)) respondents' annual family incomes. This distribution has a **negative skew**: Scores are sparse on the left-hand side, and they increase in frequency on the right side of the distribution.

Negative skew: A clustering of scores in the right-hand side of a distribution with some relatively small scores that pull the tail toward the negative side of the number line.

Knowing whether a given distribution of data points is normal or skewed is vital in criminal justice and criminology research. The average is an excellent predictor of individual scores when the curve is normal. When a distribution departs from normality, however, the average becomes less useful and, in extreme cases, can be misleading. For example, the mean number of inmate-on-staff assaults is 5.5 per jail, but you can see in [Figure 4.2](#) that the vast majority of jails had four or fewer assaults; in fact, a full two-thirds experienced just two, one, or even zero assaults. A statement such as "Jails had an average of 5.5 inmate-on-staff assaults in 2013" would be technically correct, but would be very misleading because the typical jail has fewer incidents, and many have a great deal more than that as well. Because this distribution is so skewed, the mean loses its usefulness as a description of the middle of the data. Distribution shape plays a key role in more-complicated statistical analyses (more on this in [Parts II](#) and [III](#)) and is important in a descriptive sense so that information conveyed to academic, practitioner, or lay audiences is fully accurate and transparent. You must always know where the middle of your data set is located; measures of central tendency give you that information.

Figure 4.3 GSS Respondents' Annual Family Incomes

Frequency vs Annual Income histogram. X-axis categories: <$1,000, $1,000–2,999, $3,000–3,999, $4,000–4,999, $5,000–5,999, $6,000–6,999, $7,000–7,999, $8,000–9,999, $10,000–14,999, $15,000–19,999, $20,000–24,999, $25,000+. Y-axis: Frequency (0 to 1,700).
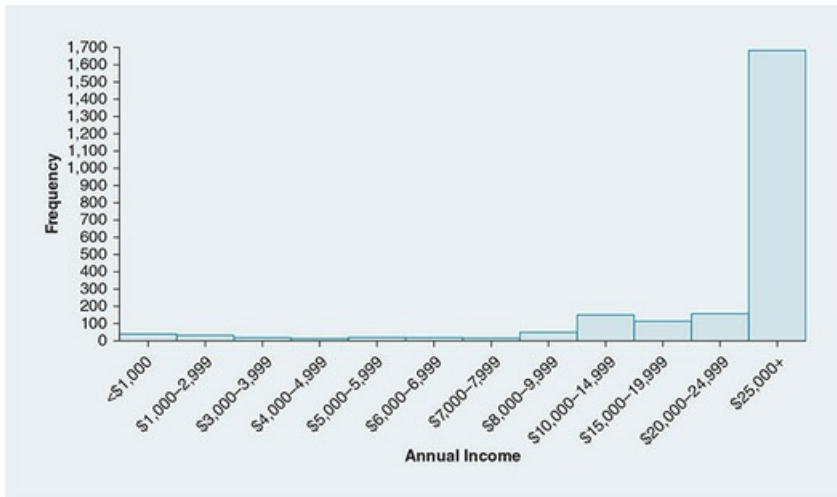
The Mode

The **mode** is the simplest of the three measures of central tendency covered in this chapter. It requires no mathematical computations and can be employed with any level of measurement. It is the only measure of central tendency available for use with nominal data. The mode is simply the most frequently occurring category or value. Table 4.1 contains data from the 2011 PPCS (Data Sources 2.1). Interviewers asked PPCS respondents whether they had been stopped by police while driving a vehicle. The people who answered yes were then asked to report the reason for that stop. This is a nominal-level variable. Table 4.1 presents the distribution of responses that participants gave for their stop. The mode is *speeding* because that is the stop reason that occurs most frequently (i.e., 2,040 people said that this is the violation for which they were pulled over by police).

> **Mode**: The most frequently occurring category or value in a set of scores.

A frequency bar graph of the same data is shown in Figure 4.4. The mode is easily identifiable as the category accompanied by the highest bar.

The mode can also be used with continuous variables. Instead of identifying the most frequently occurring category as with nominal or ordinal data, you will identify the most common value. Figure 4.5 shows a frequency histogram for the variable from the PPCS that asks respondents how many face-to-face contacts they had with the police in the past 12 months. The sample has been narrowed to include only female respondents who were 21 or younger at the time of the survey. Can you identify the modal number of contacts? If you answered "1," you are correct!

Table 4.1 Among Stopped Drivers, Reason for the Stop	
Stop Reason	*Frequency*
Speeding	2,040
Vehicle Defect	599
Record Check	381
Roadside Check	66
Seatbelt Violation	202
Illegal Turn or Lane Change	257
Stop Sign or Light Violation	275
Cellphone Usage	76
Other	283
Total	$N = 4,179$

Figure 4.4 Among Stopped Drivers, Reason for the Stop
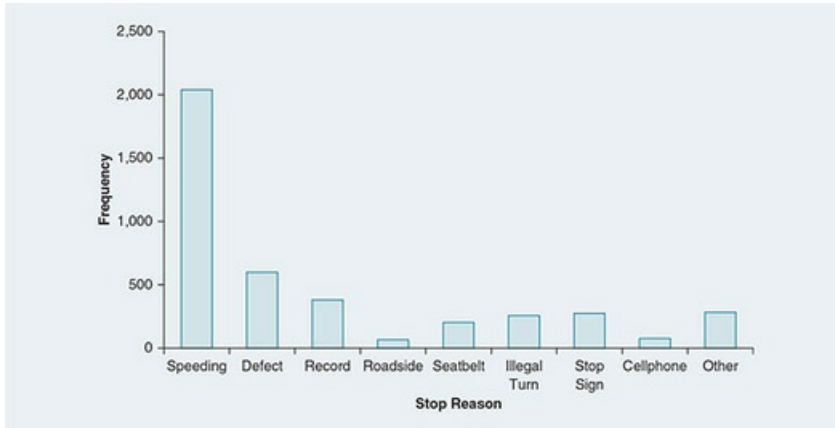
Figure 4.5 Number of Police Contacts in Past 12 Months Among Females Age 21 and Younger
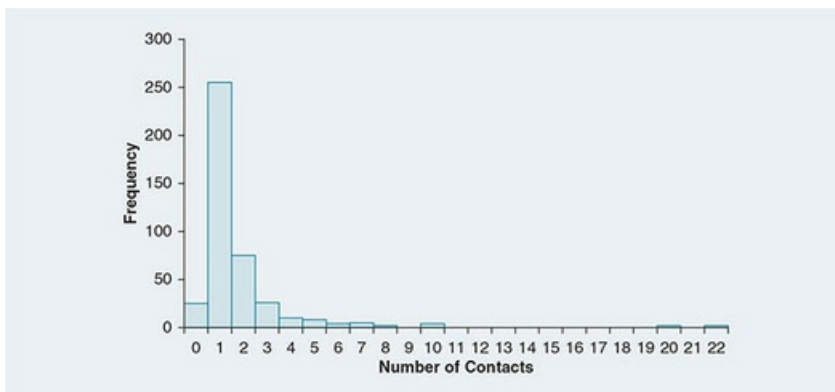
Research Example 4.1 Are People Convicted of Homicide More Violent in Prison Than People Convicted of Other Types of Offenses?

Sorensen and Cunningham (2010) analyzed the institutional conduct records of all inmates incarcerated in Florida state correctional facilities in 2003, along with the records of inmates who entered prison that same year. They divided inmates into three groups. The stock population consisted of all people incarcerated in Florida prisons during 2003, regardless of the year they were admitted into prison. The new persons admitted into prison during 2002 and serving all of 2003 composed the admissions cohort. The close custody group was a subset of the admissions cohort and was made of the inmates who were considered to be especially high threats to institutional security. The table below contains descriptive information about the three samples. For each group, can you identify the modal custody level and modal conviction offense type? (Hint: These are column percentages.) Visit **www.sagepub.com/gau** to view the full article and see the results of this study.

Demographics	Stock Population (N = 51,527)	Admissions Cohort (N = 14,088)	Close Custody Sample (N = 4,113)
Custody level (%)			
Community	6.0	10.4	0.0
Minimum	17.2	23.7	0.0
Medium	27.2	33.8	0.0
Close	48.9	32.0	100.0
Death row	0.6	0.1	0.0
Conviction offense type (%)			
Homicide	18.6	5.9	10.9
Other violent	39.6	33.8	46.2
Property	21.4	26.7	21.0
Drugs	14.7	24.0	14.7
Public order/weapons	5.7	9.6	7.2

Source: Adapted from Table 1 in Sorensen and Cunningham (2010).

Source: Adapted from Table 1 in Sorensen and Cunningham (2010).

Do Latino Drug Traffickers' National Origin and Immigration Status Affect the Sentences They Receive?

The United States is experiencing significant demographic shifts, one of the most notable being the steady increase in the percentage of the population that is Latino. Immigration (both legal and illegal) is a significant contributor to this trend. The influx has inspired many native-born U.S. citizens to react with a get-tough, law enforcement–oriented mind-set toward immigrants who commit crimes in this country. Mexican immigrants, in particular, are frequently seen as a threat, and this belief could translate into harsher treatment for those immigrants who commit crimes. Logue (2017) examined whether Latino drug traffickers' countries of origin and their immigration status impacted the severity of the sentences they receive in federal court. One of the variables she included in her analysis of sentencing outcomes was the type of drug that defendants were alleged to have trafficked. The table shows the breakdown of drug types across defendants of different origins and immigration statuses. Can you identify the modal drug type for each of the four defendant groups? (Hint: These are column percentages.) Visit www.sagepub.com/gau to view the full article and see the results of this study.

	Mexican, Legal	Mexican, Undocumented	Non-Mexican, Legal	Non-Mexican, Undocumented
Drug Type (%)				
Powder cocaine	27.5	21.5	44.1	53.6
Crack cocaine	0.4	0.7	3.9	4.3
Heroin	3.7	4.5	28.1	25.9
Marijuana	48.9	49.4	15.4	7.5
Methamphetamine	19.2	23.8	4.5	6.6
Other	0.2	0.1	4.0	2.1

Source: Adapted from Table 1 in Logue (2017).

Source: Adapted from Table 1 in Logue (2017).

Learning Check 4.3

✔

Remember that the mode is the category or value that *occurs* the most frequently—it is *not* the frequency itself. Check your ability to tell the difference between the mode and its frequency by looking back at . Identify the modal number of assaults and the approximate frequency for this number.

The previous two examples highlight the relative usefulness of the mode for categorical data as opposed to continuous data. This measure of central tendency can be informative for the former but is usually not all that interesting or useful for the latter. This is because there are other, more-sophisticated measures that can be calculated with continuous data.

The strengths of the mode include the fact that this measure is simple to identify and understand. It also, as mentioned previously, is the only measure of central tendency that can be used with nominal variables. The mode's major weakness is actually the flipside of its primary strength: Its simplicity means that it is usually too superficial to be of much use. It accounts for only one category or value in the data set and ignores the rest. It also cannot be used in more-complex computations, which greatly limits its usefulness in statistics. The mode, then, can be an informative measure for nominal variables (and sometimes ordinal, as well) and is useful for audiences who are not schooled in statistics, but its utility is restricted, especially with continuous variables.

The Median

The **median** (*Md*) is another measure of central tendency. The median can be used with continuous and ordinal data; it cannot be used with nominal data, however, because it requires that the variable under examination be rank orderable, which nominal variables are not.

> **Median**: The score that cuts a distribution in half such that 50% of the scores are above that value and 50% are below it.

The median is the value that splits a data set exactly in half such that 50% of the data points are below it and 50% are above it. For this reason, the median is also sometimes called the *50th percentile*. The median is a positional measure, which means that it is not so much calculated as it is located. Finding the median is a three-step process. First, the categories or scores need to be rank ordered. Second, the median position (*MP*) can be computed using the formula

$$MP = \frac{N+1}{2} \qquad\qquad Formula\ 4(1)$$

where

  N = total sample size

The median position tells you where the median is located within the ranked data set. The third step is to use the median position to identify the median. When N is odd, the median will be a value in the data set. When N is even, the median will have to be computed by averaging two values.

Let us figure out the median violent crime rate among the five Colorado cities listed in Table 4.2. The variable *violent crime rate* is continuous (specifically, ratio level), and the median is therefore an applicable measure of central tendency. The numbers in Table 4.2 are derived from the 2015 Uniform Crime Reports (UCR; see Data Sources 1.1).

Table 4.2 Violent Crime Rates in Five Colorado Cities	
City	*Violent Crimes per 10,000 Residents*
Aspen	8.77
Colorado Springs	43.83
Denver	67.39
Woodland Park	15.30
Steamboat Springs	22.74
$N = 5$	

The first step is to rank the rates in either ascending or descending order. Ranked in ascending order, they look like this:

8.77[1] 15.30[2] 22.74[3] 43.83[4] 67.39[5]

Superscripts have been inserted in the ranked list to help emphasize the median's nature as a positional measure that is dependent on the *location* of data points rather than these points' actual values. The superscripts represent each number's position in the data set now that the values have been rank ordered.

Second, the formula for the median position will tell you where to look to find the median. Here,

$$MP = \frac{5+1}{2} = \frac{6}{2} = 3$$

This means that the median is in position 3. Remember that the *MP* is *not* the median; rather, it is a "map" that tells you where to look to find the median.

Finally, use the *MP* to identify the median. Since the median is in position 3, we can determine that *Md* = 22.74. This group of five Colorado cities has a median violent crime rate of 22.74 per 10,000.

In this example, the sample had five cases. When there is an even number of cases, finding the median is slightly more complex and requires averaging two numbers together. To demonstrate this, we will use 2015 property crime rates in six North Dakota cities (Table 4.3).

First, rank the values: 38.50^1 130.37^2 149.67^3 239.09^4 297.24^5 325.03^6.

Second, find the median position:

$$MP = \frac{6+1}{2} = \frac{7}{2} = 3.5$$

Table 4.3 Property Crime Rates in Six North Dakota Cities	
City	*Property Crimes per 10,000 Residents*
Bismarck	130.37
Fargo	297.24
Burlington	38.50
Minot	325.03
Valley City	239.09
Lisbon	149.67
N = 6	

Notice that the *MP* has a decimal this time—this is what happens when the number of cases is even rather than odd. What this means is that the median is halfway between positions 3 and 4, so finding *Md* requires averaging these two numbers. The median is

$$Md = \frac{149.67 + 239.09}{2} = \frac{388.76}{2} = 194.38$$

This sample of six North Dakota cities has a median property crime rate of 194.38 per 10,000 residents. Of these cities, 50% have rates that are lower than this, and 50% have rates that are higher.

For another example of locating the median in an even-numbered sample size, we will use state-level

homicide rates (per 100,000 population) from the 2015 UCR. Table 4.4 shows the data. To increase the complexity a bit, we will use eight states.

Table 4.4 Homicide Rates in Eight States	
State	Homicides per 100,000 Residents
Alabama	7.16
Arkansas	6.08
Illinois	5.79
Hawai'i	1.33
Idaho	1.93
Indiana	5.63
Kansas	4.40
Massachusetts	1.88
N = 8	

Following the three steps, we first rank order the values:

1.33^1 1.88^2 1.93^3 4.40^4 5.63^5 5.79^6 6.08^7 7.16^8

Next, we use the *MP* formula to locate the position of the median in the data:

$$MP = \frac{8+1}{2} = \frac{9}{2} = 4.5$$

The *MP* of 4.5 tells us the median is halfway between the numbers located in positions 4 and 5, so we take these two numbers and find their average. The number in position 4 is 4.40 and the number in position 5 is 5.63. Their average is

$$Md = \frac{4.40+5.63}{2} = \frac{10.03}{2} = 5.02$$

The median homicide rate (per 100,000) in this sample of eight states is 5.02.

Medians can also be found in ordinal-level variables, though the median of an ordinal variable is less precise than that of a continuous variable because the former is a category rather than a number. To demonstrate, we will use an item from the PPCS that measures the driving frequency among respondents who said that they had been pulled over by police for a traffic stop within the past 6 months, either as a driver or as a passenger in the stopped vehicle. Table 4.5 displays the data.

The first two steps to finding the *Md* of ordinal data mirror those for continuous data. First, the median position must be calculated using Formula 4(1). For ordinal data, the total sample size (*N*) is found by summing the frequencies. In Table 4.5, N = 1,544 and so the *MP* is calculated as

$$MP = \frac{1,544 + 1}{2} = \frac{1,545}{2} = 772.5$$

The median position in this case is a person—the person who is in position 772.5 is the median.

The second step involves identifying the category in which the *MP* is located. Instead of ranking the categories according to frequencies as we did with continuous data, we are now going to arrange them in either ascending or descending order according to the categories themselves. In other words, the internal ranking system of the categories themselves is used to structure the sequence of the list. In Table 4.5, the categories are arranged from the most-frequent driving habits (*Every Day or Almost Every Day*) to the least-frequent ones (*Never*). As such, the categories are already in descending order and do not need to be rearranged.

Table 4.5 Driving Frequency of PPCS Respondents Who Experienced Traffic Stops	
How often do you drive?	**f**
Every Day or Almost Every Day	751
A Few Days a Week	217
A Few Days a Month	82
A Few Times a Year	37
Never	457
	N = 1,544

Next, add cumulative frequencies of the rank-ordered categories until the sum meets or exceeds the *MP*. Table 4.6 illustrates this process. Here, 751 + 217 = 968, so the median is located in the *A Few Days a Week* category. In other words, if you lined up all 1,544 people in this sample in the order in which they answered the question, labeling the people in each group accordingly, and then counted them until you reached 772.5,

the person in that position would be standing in the *A Few Days a Week* group. Therefore, we now know that half the people in this sample drive a few days a week or more, and half drive a few days a week or less.

Table 4.6 Driving Frequency of PPCS Respondents Who Experienced Traffic Stops		
How often do you drive?	f	cf
Every Day or Almost Every Day	751	751
A Few Days a Week	217	968
A Few Days a Month	82	1,050
A Few Times a Year	37	1,087
Never	457	1,544
	N = 1,544	

Note how much less informative the median is for ordinal variables as compared to continuous ones. For the crime rates in Tables 4.2 through 4.4, we were able to identify the specific, numerical median; for the driving-frequency variable in Table 4.5, we are able to say only that the median case is contained within the *A Few Days a Week* category. This is a rough estimate that paints a limited picture.

Learning Check 4.5

Rearrange the frequencies from Table 4.5 so that they are in descending order of driving frequency, rather than in ascending order as was the case in the demonstration. Complete the cumulative-frequencies exercise to locate the median. What is your conclusion?

The median has advantages and disadvantages. Its advantages are that it uses more information than the mode does, so it offers a more-descriptive, more-informative picture of the data. It can be used with ordinal variables, which is advantageous because, as we will see, the mean cannot be.

A key advantage of the median is that it is not sensitive to extreme values or outliers. To understand this concept, revisit Table 4.3 and replace Minot's rate of 325.03 with 600.00; then relocate the median. It did not change, despite a near doubling of this city's crime rate! That is because the median does not get pushed and pulled in various directions when there are extraordinarily high or low values in the data set. As we will see, this feature of the median gives it an edge over the mean, the latter of which *is* sensitive to extremely high or extremely low values and does shift accordingly.

The median has the disadvantage of not fully using all available data points. The median offers more information than the mode does, but it still does not account for the entire array of data. This shortfall of the median can be seen by going back to the previous example regarding Minot's property crime rate. The fact that the median did not change when the endpoint of the distribution was noticeably altered demonstrates how the median fails to offer a comprehensive picture of the entire data set. Another disadvantage of the median is that it usually cannot be employed in further statistical computations. There are limited exceptions to this rule, but, generally speaking, the median cannot be plugged into statistical formulas for purposes of performing more-complex analyses.

The Mean

This brings us to the third measure of central tendency that we will cover: the **mean**. The mean is the arithmetic average of a data set. Unlike locating the median, calculating the mean requires using every raw score in a data set. Each individual point exerts a separate and independent effect on the value of the mean. The mean can be calculated only with continuous (interval or ratio) data; it cannot be used to describe categorical variables.

> **Mean**: The arithmetic average of a set of data.

There are two formulas for the computation of the mean, each of which is for use with a particular type of data distribution. The first formula is one with which you are likely familiar from college or high school math classes. The formula is

$$\overline{x} = \frac{\sum x}{N},$$

<div align="right">*Formula 4(2)*</div>

where

$\overline{x}$ (x bar) = the sample mean,
Σ (*sigma*) = a summation sign directing you to sum all numbers or terms to the right of it,
x = values in a given data set, and
N = the sample size.

This formula tells you that to compute the mean, you must first add all the values in the data set together and then divide that sum by the total number of values. Division is required because, all else being equal, larger data sets will produce larger sums, so it is vital to account for sample size when attempting to construct a composite measure such as the mean.

For the example concerning computation of the mean, we can reuse the Colorado violent crime rate data from Table 4.2:

$$\overline{x} = \frac{8.77 + 43.83 + 67.39 + 15.30 + 22.74}{5} = \frac{158.03}{5} = 31.61$$

In 2015, these five cities had a mean violent crime rate of 31.61 per 10,000 residents. Let us try one more example using data from Table 4.4 (homicide rates in 8 states). The mean is calculated as

$$\overline{x} = \frac{7.16 + 6.08 + 5.79 + 1.33 + 1.93 + 5.63 + 4.40 + 1.88}{8} = \frac{34.20}{8} = 4.28$$

The second formula for the mean is used for large data sets that are organized in tabular format using both an *x* column that contains the raw scores and a frequency (*f*) column that conveys information about how many times each *x* value occurs in the data set. Table 4.7 shows data from the Bureau of Justice Statistics (BJS) on the number of death-sentenced prisoners received per state in 2013 (Snell, 2014). Note that the *f* column sums to 36 rather than 50 because in 2013, 14 states did not authorize the death penalty and are thus excluded from the analysis. (This number climbed to 15 that same year when Maryland abolished capital punishment and the governor commuted the sentences of the four people remaining on death row.) Table 4.7 contains the numbers that states reported receiving and the frequency of each number. For instance, 20 states that authorize the death penalty did not admit any new prisoners to death row, while 5 admitted one new inmate each.

To calculate the mean using frequencies, first add a new column to the table. This column—titled *fx*—is the product of the *x* and *f* columns. The results of these calculations for the death row data are located in the right-hand column of Table 4.8. The *fx* column saves time by using multiplication as a shortcut and thereby avoiding cumbersome addition. Using the conventional mean formula would require extensive addition because you would have to sum 36 numbers (i.e., 20 zeroes plus 5 ones plus 4 twos, and so on). This process is unwieldy and impractical, particularly with very large data sets. Instead, merely multiply each value by its frequency and then sum these products to find the total sum of all *x* values. You can see from Table 4.8 that, in 2013, states received 81 new death-sentenced inmates.

Table 4.7 Number of Death-Sentenced Prisoners Received by States, 2013	
Number Received (x)	f
0	20
1	5
2	2
3	2
4	3
5	1
9	1
15	1
25	1
	N = 36

Once the *fx* column is complete and has been summed, the mean can be calculated using a formula slightly different from the one presented in Formula 4(2). The mean formula for large data sets is

$$\overline{x} = \frac{\sum fx}{N},$$

Formula 4(3)

where

f = the frequency associated with each raw score x and

fx = the product of x and f.

Table 4.8 Number of Death-Sentenced Prisoners Received by States, 2013		
Number Received (x)	f	fx
0	20	0
1	5	5
2	2	4
3	2	6
4	3	12
5	1	5
9	1	9
15	1	15
25	1	25
	N = 36	Σ = 81

The process of computing this mean can be broken down into three steps: (1) Multiply each x by its f, (2) sum the resulting fx products, and (3) divide by the sample size N. Plugging the numbers from Table 4.8 into the formula, it can be seen that the mean is

$$\overline{x} = \frac{81}{36} = 2.25$$

In 2013, the 36 states that authorized use of the death penalty each received a mean of 2.25 new death-sentenced prisoners.

Learning Check 4.7

✓

Anytime you need to compute a mean, you will have to choose between Formulas 4(2) and 4(3). This is a simple enough choice if you just consider that in order to use the formula with an *f* in it, there must be an *f* column in the table. If there is no *f* column, use the formula that does not have an *f*. Refer back to Table 2.9 in Chapter 2. Which formula would be used to calculate the mean? Explain your answer. As practice, use the formula and compute the mean number of prisoners executed per state.

Note that Table 2.9 contains all 50 states, including those states that have abolished capital punishment, whereas Tables 4.7 and 4.8 contain only those 36 states that allowed the death penalty at the time the data were collected. What would happen to the mean calculated based on Table 4.8 if the 14 states that did not allow the death penalty (and thus had zero admissions) were added to the table and to the computation of the mean? Alternatively, what would happen to the mean you calculated on Table 2.9 if the 14 states without capital punishment (which therefore had no executions) were removed from the calculation of the mean? Think about these questions theoretically and make a prediction about whether the mean would increase or decrease. Then make the change to each set of data and redo the means. Were your predictions correct?

Research Example 4.2 How Do Offenders' Criminal Trajectories Impact the Effectiveness of Incarceration?

It is well known that some offenders commit a multitude of crimes over their life and others commit only a few, but the intersection of offense volume (offending rate) and time (the length of a criminal career) has received little attention from criminal justice/criminology researchers. Piquero, Sullivan, and Farrington (2010) used a longitudinal data set of males in South London who demonstrated delinquent behavior early in life and were thereafter tracked by a research team who interviewed them and looked up their official conviction records. The researchers were interested in finding out whether males who committed a lot of crimes in a short amount of time (the short-term, high-rate [STHR] offenders) differed significantly from those who committed crimes at a lower rate over a longer time (the long-term, low-rate [LTLR] offenders) on criminal justice outcomes. The researchers gathered the following descriptive statistics. The numbers not in parentheses are means. The numbers in parentheses are standard deviations, which we will learn about in the next chapter.

You can see from the table that the LTLR offenders differed from the STHR offenders on a number of dimensions. They were, overall, older at the time of their first arrest and had a longer mean career length. They committed many fewer crimes per year and were much less likely to have been sentenced to prison.

Piquero et al.'s (2010) analysis reveals a dilemma about what should be done about these groups of offenders with respect to sentencing; namely, it shows how complicated the question of imprisonment is. The STHR offenders might appear to be the best candidates for incarceration based on their volume of criminal activity, but these offenders' criminal careers are quite short. It makes no sense from a policy and budgetary perspective to imprison people who would not be committing crimes if they were free in society. The STHR offenders also tended to commit property offenses rather than violent ones. The LTLR offenders, by contrast, committed a disproportionate number of violent offenses despite the fact that their overall number of lifetime offenses was lower than that for the STHR group. Again, though, the question of the utility of imprisonment arises: Is it worth incarcerating someone who, though he might still have many years left in his criminal career, will commit very few crimes during that career? The dilemma of sentencing involves the balance between public safety and the need to be very careful in the allotting of scarce correctional resources.

Variable	Offender Type	
	LTLR (N = 44)	STHR (N = 21)
Overall career length (years)	14.5 (6.50)	10.8 (4.40)
Offenses committed per year	.42 (.31)	1.25 (.72)
Age at first conviction	17.8 (4.70)	13.5 (2.30)
Total convictions	4.9 (2.40)	11.5 (4.10)
Percentage ever incarcerated	9.1%	61.9%
Number of times incarcerated	1.46 (.50)	1.23 (.66)
Years incarcerated	1.23 (.76)	1.05 (.56)

Source: Adapted from Table 1 in Piquero et al. (2010).

Source: Adapted from Table 1 in Piquero et al. (2010).

Can Good Parenting Practices Reduce the Criminogenic Impact of Youths' Time Spent in Unstructured Activities?

Youths spending time with peers, away from parents and other adults, are at elevated risk for engaging in delinquency either individually or in a group setting. The chances of unstructured, unsupervised activities leading to antisocial acts increases in disadvantaged urban settings, where opportunities for deviance are higher and youths who have never been in trouble before are more likely to encounter delinquent peers. Janssen, Weerman, and Eichelsheim (2017) posited that parents can reduce the chances that their children's unstructured time will result in deviance or delinquency. The researchers hypothesized that strong bonds between parents and children mitigate the criminogenic impacts of time spent in urban environments with peer groups, as does the extent to which parents monitor their children and set boundaries. They used a longitudinal research design, wherein a random sample of adolescents was interviewed twice over a span of two years. Data were collected on how active parents were in their children's lives, the quality of each parent–child relationship, time spent in unstructured activities within disorderly environments, and the number of delinquent acts each child had committed. The table contains the means and the standard deviations for each wave of data collection.

The authors found that all three measures of positive parenting practices significantly reduced children's delinquency. There was no indication that positive parenting buffers children against the deleterious impacts of criminogenic environments; however, good parenting and strong parent–child bonds did negate the effects of increases in the amount of time adolescents spent in these environments. These results suggest that although parenting practices alone do not fully combat the bad influence of unstructured time spent in disorderly neighborhoods, they can offset the effects of an increase in time spent in this manner. Parents have an important role in preventing their children from slipping into deviance and delinquency.

Variable	Wave 1		Wave 2	
	Mean	SD	Mean	SD
Parental monitoring	17.35	2.66	16.36	4.56
Parental limit setting	16.86	2.66	16.61	2.25
Quality of relationship	23.05	3.30	22.76	3.42
Time spent in unstructured activities	0.73	1.93	1.22	2.43
Delinquency	5.50	9.00	4.31	7.15

Source: Adapted from Table 1 in Janssen et al. (2017).
Source: Adapted from Table 1 in Janssen et al. (2017).

Learning Check 4.8

✓

Explain why the mean cannot be calculated on ordinal data. Look back at the driving frequency variable in Table 4.6 and identify the information that is missing and prevents you from being able to calculate a mean.

The mean is sensitive to extreme values and outliers, which gives it both an advantage and a disadvantage relative to the median. The advantage is that the mean uses the entire data set and accounts for "weird" values that sometimes appear at the high or low ends of the distribution. The median, by contrast, ignores these values. The disadvantage is that the mean's sensitivity to extreme values makes this measure somewhat unstable; it is vulnerable to the disproportionate impact that a small number of extreme values can exert on the data set.

To illustrate this property of the mean, consider Table 4.9, which contains the 2015 homicide rates for six cities in California. Trace the changes in the mean homicide rates from left to right. Do you notice how the rate increases with the successive introductions of Los Angeles, Soledad, and San Bernardino? Los Angeles pulls the mean up from 2.17 to 3.41, and Soledad tugs it to 5.11. The most dramatic increase occurs when San Bernardino is added: The mean shoots up to 7.65. The successive addition of higher-rate cities caused, in total, more than a threefold increase from the original mean across the three low-crime cities.

Table 4.9 Homicide Rates in California Cities			
Urban Area	Homicides per 100,000	Homicides per 100,000	Homicides per 100,000	Homicides per 100,000
San Diego	2.64	2.64	2.64	2.64
Redlands	2.81	2.81	2.81	2.81
Santa Monica	1.07	1.07	1.07	1.07
Los Angeles	–	7.12	7.12	7.12
Soledad	–	–	11.92	11.92
San Bernardino	–	–	–	20.33
Sample size	$N = 3$	$N = 4$	$N = 5$	$N = 6$
Mean	$\bar{x} = \dfrac{6.52}{3} = 2.17$	$\bar{x} = \dfrac{13.64}{4} = 3.41$	$\bar{x} = \dfrac{25.56}{5} = 5.11$	$\bar{x} = \dfrac{45.89}{6} = 7.65$

This demonstrates how the inclusion of extreme values can cause the mean to move in the direction of those values. A score that is noticeably greater than the others in the sample can draw the mean upward, while a value that is markedly lower than the rest can drive the mean downward. There is a good reason why, for example, average income in the United States is reported as a median rather than a mean—a mean would lump extremely poor people who are barely scraping by in with multibillionaires. That would not be accurate at all! The apparent "average" income in the United States would be huge. Finding the point at which 50% of households sit below that particular annual income and 50% above it is more useful and accurate. Because of its sensitivity to extreme values, the mean is most informative when a distribution is normally distributed; the accuracy of the mean as a true measure of central tendency is reduced in distributions that are positively or negatively skewed, and the mean can actually become not merely inaccurate but downright misleading. For instance, a severely economically divided city could have a mean income in the hundreds of thousands of

dollars, even if a significant portion of the local population is impoverished.

Another implication of the mean's sensitivity to extreme values is that the mean and the median can be compared to determine the shape of a distribution, as described in the following section.

Table 4.10 Summary of the Measures of Central Tendency Available for Each Level of Measurement			
	Mode	Median	Mean
Nominal	✓		
Ordinal	✓	✓	
Interval	✓	✓	✓
Ratio	✓	✓	✓

Using the Mean and the Median to Determine Distribution Shape

Given that the median is invulnerable to extreme values but the mean is not, the best strategy is to report both of these measures when describing data distributions. The mean and the median can, in fact, be compared to form a judgment about whether the data are normally distributed, positively skewed, or negatively skewed. In normal distributions, the mean and median will be approximately equal. Positively skewed distributions will have means markedly greater than their medians. This is because extremely high values in positively skewed distributions pull the mean up but do not affect the location of the median. Negatively skewed distributions, on the other hand, will have medians that are noticeably larger than their means because extremely low numbers tug the mean downward but do not alter the median's value. Figure 4.6 illustrates this conceptually.

To give a full picture of a data distribution, then, it is best to make a habit of reporting both the mean and the median.

The mean—unlike the mode or the median—forms the basis for further computations; in fact, the mean is an analytical staple of many inferential hypothesis tests. The reason that the mean can be used in this manner is that the mean is the **midpoint of the magnitudes**. This point is important and merits its own section.

Midpoint of the magnitudes: The property of the mean that causes all deviation scores based on the mean to sum to zero.

Figure 4.6 The Mean and Median as Indicators of Distribution Shape

126
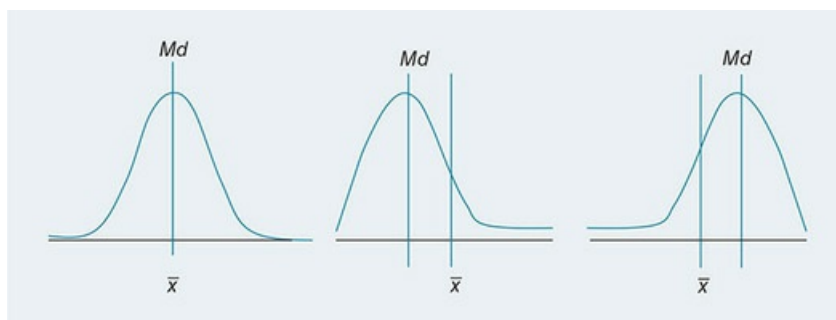
Deviation Scores and the Mean as the Midpoint of the Magnitudes

The mean possesses a vital property that enables its use in complex statistical formulas. To understand this, we must first discuss **deviation scores**. A deviation score is a given data point's distance from its group mean. The formula for a deviation score is based on simple subtraction:

$$d_i = x_i - \overline{x},$$

<div style="text-align: right">*Formula 4(4)*</div>

Deviation score: The distance between the mean of a data set and any given raw score in that set.

where

d_i = the deviation score for a given data point x_i, and

x_i = a given data point,

$\overline{X}$ = the sample mean.

Suppose, for instance, that a group's mean is $\overline{X}$ = 24. If a certain raw score x_i is 22, then $d_{x=22}$ = 22 − 24 = −2. A raw score of 25, by contrast, would have a deviation score of $d_{x=25}$ = 25 − 24 = 1.

A deviation score conveys two pieces of information. The first is the absolute value of the score or, in other words, how far from the mean a particular raw score is. Data points that are exactly equal to the mean will have deviation scores of 0; therefore, deviation scores with larger absolute values are farther away from the mean, while deviation scores with smaller absolute values are closer to it.

The second piece of information that a deviation score conveys is whether the raw score associated with that deviation score is greater than or less than the mean. A deviation score's sign (positive or negative) communicates this information. Positive deviation scores represent raw scores that are greater than the mean and negative deviation scores signify raw numbers that are less than the mean. You can thus discern two characteristics of the raw score x_i that a given deviation score d_i represents: (1) the distance between x_i and $\overline{X}$ and (2) whether x_i is above $\overline{X}$ or below it. Notice that you would not even need to know the actual value of x_i or $\overline{X}$ in order to effectively interpret a deviation score. Deviation scores convey information about the position of a given data point with respect to its group mean; that is, deviation scores offer information about raw scores' *relative*, rather than absolute, positions within their group. Figure 4.7 illustrates this.

What lends the mean its title as the midpoint of the magnitudes is the fact that deviation scores computed using the mean (as opposed to the mode or median) always sum to zero (or within rounding error of it). The mean is the value in the data set at which all values below it balance out with all values above it. For an example of this, try summing the deviation scores in Figure 4.7. What is the result?

To demonstrate this concept more concretely, the raw homicide counts that were used as rates in Table 4.9 are listed in Table 4.11. These cities had a mean of 87.25 homicides in 2015. The *d* column contains the deviation score for each homicide count.

Illustrative of the mean as the fulcrum of the data set, the positive and negative deviation scores ultimately cancel each other out, as can be seen by the sum of zero at the bottom of the deviation-score column. This represents the mean's property of being the midpoint of the magnitudes—it is the value that perfectly balances all of the raw scores. This characteristic is what makes the mean a central component in more-complex statistical analyses. You will see in later chapters that the mean features prominently in many calculations.

Figure 4.7 Deviation Scores in a Set of Data With a Mean of 24
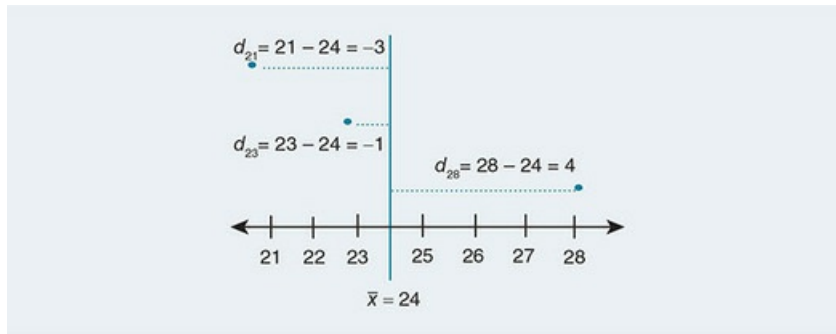
Table 4.11 Homicides in California Cities

City	Homicides	d
San Diego	37	$37 - 63.50 = -26.50$
Redlands	2	$2 - 63.50 = -61.50$
Santa Monica	13	$13 - 63.50 = -50.50$
Los Angeles	282	$282 - 63.50 = 218.50$
Soledad	3	$3 - 63.50 = -60.50$
San Bernardino	44	$44 - 63.50 = -19.50$
	$\bar{x} = \dfrac{381}{6} = 63.50$	$\Sigma = 0.00$

Learning Check 4.9

To test your comprehension of the concept of the mean as the midpoint of the magnitudes, go back to Table 4.3. You calculated the mean property-crime rate in Learning Check 4.6. Use that mean to compute each city's deviation score, and then sum the scores.

SPSS

Criminal justice and criminology researchers generally work with large data sets, so computing measures of central tendency by hand is not feasible; luckily, it is not necessary, either, because statistical programs such as SPSS can be used instead. There are two different ways to obtain central tendency output. Under the *Analyze → Descriptive Statistics* menu, SPSS offers the options *Descriptives* and *Frequencies*. Both of these functions will produce central tendency analyses, but the *Frequencies* option offers a broader array of descriptive statistics and even some charts and graphs. For this reason, we will use *Frequencies* rather than *Descriptives*. Once you have opened the *Frequencies* box, click on the *Statistics* button to open a menu of options for measures of central tendency. Select *Mean*, *Median*, and *Mode*, as shown in Figure 4.8. Then click *OK*, and the output displayed in Figure 4.9 will appear. The variable used in this example comes from Table 4.7, measuring the number of prisoners received under sentence of death in 2015.

You can see from Figure 4.9 that the mean is identical to the result we arrived at by hand earlier. The mode is zero, which you can verify by looking at Table 4.7. The median is zero, meaning half the states did not receive any new death-row inmates and half received one or more. We can also compare the mean and the median to determine the shape of the distribution. With a mean of 2.25 and a median of zero, do you think that this distribution is normally distributed, positively skewed, or negatively skewed? If you said positively skewed, you are correct!

Figure 4.8 Running Measures of Central Tendency in SPSS
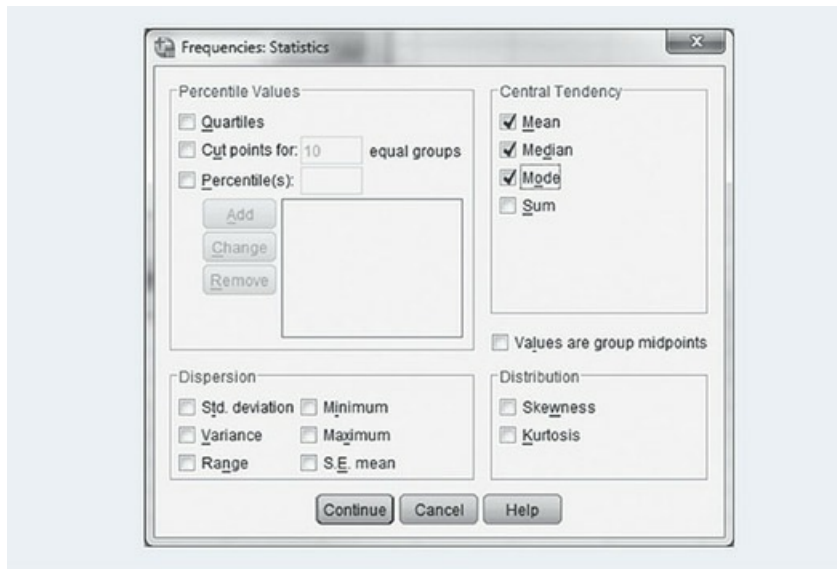
Figure 4.9 SPSS Output

Frequencies

Statistics

Prisoners received 2013

N	Valid	36
	Missing	0
Mean		2.2500
Median		.0000
Mode		.00

For another example, we will use 2015 homicide rates (per 100,000) in all California cities. Figure 4.10 shows the mean, the median, and the mode for this variable.

> Follow along with this example of using SPSS to obtain measures of central tendency by downloading the file *California Homicides for Chapter 4.sav* at www.sagepub.com/gau.

Figure 4.10 Homicide Rates in California Cities

Frequencies

Statistics

California homicide rates, 2015

N	Valid	460
	Missing	0
Mean		6.5782
Median		1.0839
Mode		.00

The mode is zero, but because homicide rates vary so widely across this large sample of cities ($N = 460$), the mode is not a useful or informative measure. More interesting are the mean and the median. The mean (6.58) is much larger than the median (1.08), indicating significant positive skew. This can be verified by using SPSS to produce a histogram of the data using the Chart Builder; refer back to Chapter 3 if you need a refresher on the use of the Chart Builder. Figure 4.11 shows the histogram.

Figure 4.11 Histogram of Homicide Rates in California Cities

Graph

Mean = 6.5782
Std. Dev. = 49.11422
N = 460

California homicide rates, 2015
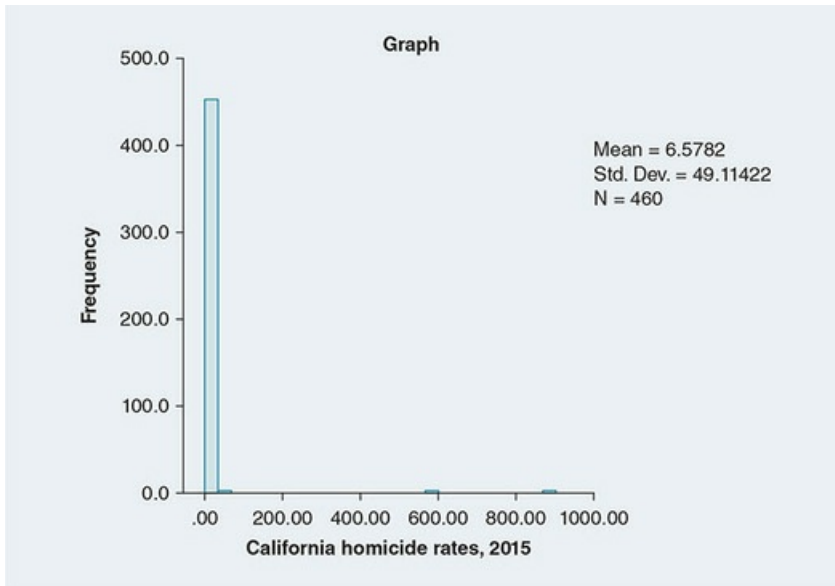
Figure 4.11 confirms the severe skew in this variable: The data cluster toward the lower end so much that the values out in the tail are barely visible in the histogram.

There is a GIGO alert relevant here. It is your responsibility to ensure that you use the correct measure(s) of central tendency given the level of measurement of the variable with which you are working. The SPSS program will not produce an error message if you make a mistake by, for instance, telling it to give you the mean of a nominal or ordinal variable. You will get a mean, just the same as you get when you correctly ask for the mean of a continuous variable. To illustrate this, Figure 4.12 contains output from the National Crime Victimization Survey (NCVS) showing respondents' marital status. Although this is a nominal variable—making the median and the mean inappropriate—SPSS went ahead with the calculations anyway and produced results. Of course, the mean and the median of a variable measuring whether someone is married, separated, divorced, and so on are nonsense, but SPSS does not know that. This statistical program is not a substitute for knowing which techniques are appropriate for which data types.

Figure 4.12 Respondent Marital Status

Statistics

Marital status

N	Valid	84229
	Missing	594
Mean		2.07
Median		1.00
Mode		1

Chapter Summary

This chapter introduced you to three measures of central tendency: mode, median, and mean. These statistics offer summary information about the middle or average score in a data set. The mode is the most frequently occurring category or value in a data set. The mode can be used with variables of any measurement type (nominal, ordinal, interval, or ratio) and is the only measure that can be used with nominal variables. Its main weakness is in its simplicity and superficiality—it is generally not all that useful.

The median is a better measure than the mode for data measured at the ordinal or continuous level. The median is the value that splits a data set exactly in half. Since it is a positional measure, the median's value is not affected by the presence of extreme values; this makes the median a better reflection of the center of a distribution the mean is when a distribution is highly skewed. The median, though, does not take into account all data points in a distribution, which makes it less informative than the mean.

The mean is the arithmetic average of the data and is used with continuous variables only. The mean accounts for all values in a data set, which is good because no data are omitted; the flipside, however, is that the mean is susceptible to being pushed and pulled by extreme values.

It is good to report both the mean and the median because they can be compared to determine the shape of a distribution. In a normal distribution, these two statistics will be approximately equal; in a positively skewed distribution, the mean will be markedly greater than the median; and in a negatively skewed distribution, the mean will be noticeably smaller than the median. Reporting both of them provides your audience with much more information than they would have if you just reported one or the other.

The mode, the median, and the mean can all be obtained in SPSS using the *Analyze → Descriptive Statistics → Frequencies* sequence. As always, GIGO! When you order SPSS to produce a measure of central tendency, it is your responsibility to ensure that the measure you choose is appropriate to the variable's level of measurement. If you err, SPSS will probably not alert you to the mistake—you will get output that looks fine but is actually garbage. Be careful!

Thinking Critically

1. According to the Law Enforcement Management and Administrative Statistics (LEMAS) survey, the mean number of officers per police agency (of all types) is 163.92. Do you trust that this mean is an accurate representation of the middle of the distribution of police agency size? Why or why not? If not, what additional information would you need in order to gain an accurate understanding of this distribution's shape and central tendency?
2. The COJ reports that the modal number of inmates per jail is 1. This value occurs more frequently than any other population value (51 times among 2,371 jails). Use this example to discuss the limitations and drawbacks of using the mode to describe the central tendency of a continuous variable. Then identify the measure(s) you would use instead of the mode, and explain why.

Review Problems

1. A survey item asks respondents, "How many times have you shoplifted?" and allows them to fill in the appropriate number.
   1. What level of measurement is this variable?
   2. What measure or measures of central tendency can be computed on this variable?
2. A survey item asks respondents, "How many times have you shoplifted?" and gives them the answer options: *0, 1–3, 4–6, 7 or more*.
   1. What level of measurement is this variable?
   2. What measure or measures of central tendency can be computed on this variable?
3. A survey item asks respondents, "Have you ever shoplifted?" and tells them to circle *yes* or *no*.
   1. What level of measurement is this variable?
   2. What measure or measures of central tendency can be computed on this variable?
4. Explain what an extreme value is. Include in your answer (1) the effect extreme values have on the median, if any, and (2) the effect extreme values have on the mean, if any.
5. Explain why the mean is the midpoint of the magnitudes. Include in your answer (1) what deviation scores are and how they are calculated and (2) what deviation scores always sum to.
6. In a negatively skewed distribution . . .
   1. the mean is less than the median.
   2. the mean is greater than the median.
   3. the mean and the median are approximately equal.
7. In a normal distribution . . .
   1. the mean is less than the median.
   2. the mean is greater than the median.
   3. the mean and the median are approximately equal.
8. In a positively skewed distribution . . .
   1. the mean is less than the median.
   2. the mean is greater than the median.
   3. the mean and the median are approximately equal.
9. In a positively skewed distribution, the tail extends toward _____ of the number line.
   1. the positive side
   2. both sides
   3. the negative side
10. In a negatively skewed distribution, the tail extends toward _____ of the number line.
    1. the positive side
    2. both sides
    3. the negative side

11. The following table contains 2015 UCR data on the relationship between murder victims and their killers, among those crimes for which the relationship status is known.

Victim–Offender Relationships Among Murder Victims	
Offender's Relationship to Victim	f
Family Member	1,099
Intimate Partner	1,270
Friend	365
Acquaintance	2,801
Other Nonstranger	113
Stranger	1,375
	N = 7,023

1. Identify this variable's level of measurement and, based on that, state the appropriate measure or measures of central tendency.
2. Determine or calculate the measure or measures of central tendency that you identified in part (a).

12. The frequency distribution in the following table shows rates of violent victimization, per victim racial group, in 2014 according to the NCVS (Truman & Langton, 2015). Use this table to do the tasks following the table.

Violent Victimization Rate per 1,000 Persons	
Race	Victimization Rate
Black/African American	22.5
White	20.3
Hispanic/Latino	16.2
Other	23.0
N = 4	

1. Identify the median victimization rate using all three steps.
2. Compute the mean victimization rate across all racial groups.

13. Morgan, Morgan, and Boba (2010) report state and local government expenditures, by state, for police protection in 2007. The data in the following table contain a random sample of states and the dollars spent per capita in each state for police services.

Dollars Spent per Capita on Police Protection, 2007	
State	Dollars Spent per Capita
Illinois	317
Arkansas	170
Alabama	211
Ohio	258
Washington State	219
Florida	345
Maine	176
Texas	220
N = 8	

1. Identify the median dollar amount using all three steps.
2. Calculate the mean dollar amount.

14. The following frequency distribution shows LEMAS data on the number of American Indian officers employed by state police agencies.

American Indian Officers in State Police Agencies

Number of Officers	f	Number of Officers	f
0	8	13	1
1	4	14	1
2	3	15	1
3	2	19	1
4	3	20	1
6	2	27	1
7	4	40	1
9	3	44	1
10	1	45	1
11	1	83	1
12	1		$N = 42$

1. Identify the modal number of American Indian officers in this sample of agencies.
2. Compute the mean number of American Indian officers in this sample.
3. The median number of American Indian officers is 6.00. Based on this median and the mean you calculated, would you say that this distribution is normally distributed, positively skewed, or negatively skewed? Explain your answer.

15. The following frequency distribution shows a variable from the PPCS measuring, among female respondents who had been stopped by police while walking or riding a bike, the number of minutes those stops lasted.

Number of Minutes Stop Lasted

Minutes	f	Number	f
1	15	20	1
2	4	24	1
3	2	30	3
4	2	45	2
5	15	60	2
7	1	120	2
8	1	180	1
10	11		$N = 70$
15	7		

1. Compute the mean number of minutes per stop.
2. The median number of minutes was 5.00. Based on this median and the mean you calculated earlier, would you say that this distribution is normally distributed, positively skewed, or negatively skewed? Explain your answer.

16. The following table shows UCR data on the number of aggravated assaults that occurred in five Wyoming cities and towns in 2015.

Aggravated Assaults in Wyoming Cities and Towns	
City or Town	**Agg. Assaults**
Pine Bluffs	6
Jackson	14
Laramie	32
Afton	1
Cheyenne	95
$N = 5$	

1. Identify the median number of assaults in this sample.
2. Calculate the mean number of assaults.
3. Calculate each city's deviation score, and sum the scores.

17. The following table displays the number of juveniles arrested for arson in select states in 2015, according to the UCR.

Juveniles Arrested for Arson	
State	**Arrests**
Alaska	31
Arkansas	26
Connecticut	31
Kansas	15
Montana	13
South Carolina	27
Vermont	4
$N = 7$	

1. Identify the median number of arson arrests per state.
2. Calculate the mean number of arrests in this sample.
3. Calculate each state's deviation score and sum the scores.

18. The data set *NCVS for Chapter 4.sav* (www.sagepub.com/gau) contains the ages of respondents to the 2015 NCVS. Run an SPSS analysis to determine the mode, the median, and the mean of this variable. Summarize the results in words.

Helpful Hint: When running measures of central tendency on large data sets in SPSS, deselect the "Display frequency tables" option in the "Frequencies" dialog box. This will not alter the analyses you are running but will make your output cleaner and simpler to examine.

19. The data set *NCVS for Chapter 4.sav* (www.sagepub.com/gau) also contains portions of the Identity Theft Supplement survey. The variable "purchases" asked respondents how many times they had made online purchases in the past year. Run an SPSS analysis to determine the mode, the median, and the mean of this variable. Summarize the results in words.

20. The data file *Staff Ratio for Chapter 4.sav* (www.sagepub.com/gau) contains a variable from the 2006 COJ showing the ratio of inmates to security staff per institution. Run an SPSS analysis to determine the mode, the median, and the mean of this variable. Summarize the results in words.

Key Terms

Measures of central tendency 76

Normal distribution 76

Positive skew 77

Negative skew 77

Mode 78

Median 83

Mean 88

Midpoint of the magnitudes 96

Deviation score 97

Glossary of Symbols and Abbreviations Introduced in This Chapter

MP	Median position
Md	Median
$\bar{X}$	Mean
d_i	Deviation score for a raw score x_i

Chapter 5 Measures of Dispersion

Consider this question: Do you think that there is more variability in the physical size of housecats or in that of dogs? In other words, if I gathered a random sample of housecats and a random sample of dogs, would I find a greater diversity of sizes and weights in the cat sample or in the dog sample? In the dog sample, of course! Dogs range from puny things you might lose in your sofa cushions all the way up to behemoths that might eat your sofa. Cats, on the other hand, are . . . well, pretty much just cat-sized. If I were to draw separate size distributions for dogs and housecats, they might look something like Figure 5.1.

The dog distribution would be wider and flatter than the cat distribution, while the cat distribution would appear somewhat tall and narrow by comparison. This is because dogs' sizes vary more than cats' sizes do.

Now, let us consider a situation in which two distributions have the same mean. (The dog and cat distributions would, of course, have different means.) Consider the hypothetical raw data for variables X_1 and X_2 in Table 5.1.

These distributions have the same mean, so if this were the only piece of information you had about them, you might be tempted to conclude that they are similar to one another. This conclusion would be quite wrong, though. Look at Figure 5.2, which displays a line chart for these two variables. Which series do you think represents Sample 1? Sample 2? If you said that the stars are Sample 1 and diamonds are Sample 2, you are correct. You can see that the raw scores of Sample 1 cluster quite tightly around the mean, while Sample 2's scores are scattered about in a much less cohesive manner.

The previous examples highlight the importance of the subject of this chapter: measures of dispersion. **Dispersion** (sometimes called variability) is the amount of "spread" present in a set of raw scores. Dogs have more dispersion in physical size than cats have, just as Sample 2 has more dispersion than Sample 1 has. Measures of dispersion are vital from an informational standpoint for the reason exemplified by the thought experiment just presented—measures of central tendency convey only so much information about a distribution and can actually be misleading, because it is possible for two very discrepant distributions to have similar means. It is also necessary to know the shape of a data distribution so that you can assess whether it appears fairly normal or whether it deviates from normality. We talked about skew in Chapter 3: Skew occurs when values cluster at one end of the distribution or the other. The dispersion analog of skew is **kurtosis**. There are two types of kurtosis: **Leptokurtosis** happens when values cluster together very tightly, and

platykurtosis is evident when values are markedly spread out. In Figure 5.1, the dog-size distribution is platykurtic (representing a spread-out smattering of values), and the cat-size curve is leptokurtic (indicating that values are highly clustered and have minimal variability).

> **Dispersion**: The amount of spread or variability among the scores in a distribution.
>
> **Kurtosis**: A measure of how much a distribution curve's width departs from normality.
>
> **Leptokurtosis**: A measure of how peaked or clustered a distribution is.
>
> **Platykurtosis**: A measure of how flat or spread out a distribution is.

Figure 5.1 Hypothetical Distributions of Dog and Housecat Sizes
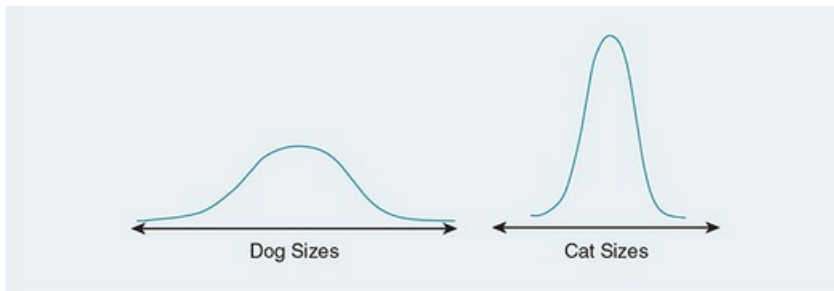
Table 5.1 Hypothetical Data in Two Samples

Sample 1	Sample 2
10	0
11	15
9	20
10	5
10	10
$\bar{x}_1 = 10$	$\bar{x}_2 = 10$

For all these reasons, measures of dispersion are necessary pieces of information about any distribution. They go hand in hand with measures of central tendency, and the two types of descriptive statistics are usually presented alongside one another. This chapter discusses four of the most common types of measures of dispersion: the variation ratio, the range, the variance, and the standard deviation. Like measures of central tendency, each measure of dispersion is suitable for variables of only certain levels of measurement.

Figure 5.2 Line Chart for Variables X_1 and X_2
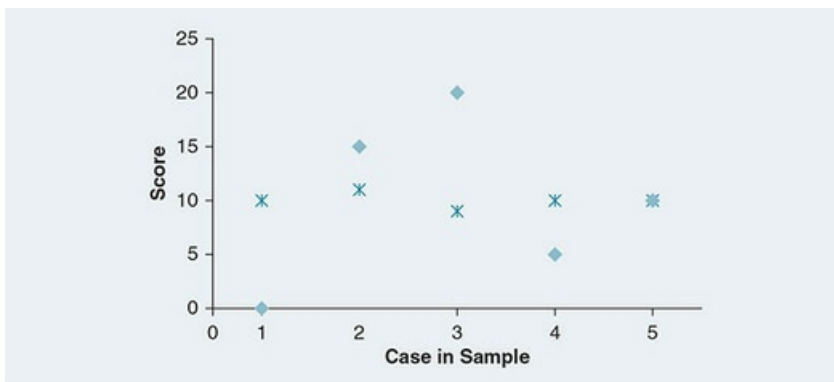
The Variation Ratio

The **variation ratio** (*VR*) is the only measure of dispersion discussed in this chapter that can be used with categorical (nominal and ordinal) data; the remainder of the measures that will be covered are restricted to continuous data (interval or ratio). The *VR* is based on the mode (see [Chapter 4](#)). It measures the proportion of cases that are *not* in the modal category. Recall from the discussion of proportions in [Chapter 3](#) that the proportion of cases that are in a certain category can be found using Formula 3(1): $p = \dfrac{f}{N}$ It is easy to take this formula a step farther in order to calculate the proportion that is outside any given category.

> **Variation ratio**: A measure of dispersion for variables of any level of measurement that is calculated as the proportion of cases located outside the modal category. Symbolized as *VR*.

For finding the proportion of cases outside a specific category, we rely on the **bounding rule**, which states that proportions always range from 0.00 to 1.00. The bounding rule leads to the **rule of the complement**. This rule states that the proportion of cases in a certain category (call this category *A*) and the proportion located in other categories (call this *Not A*) always sum to 1.00. The proportion of cases in *A* is written as "*p*(*A*)" and the proportion of cases outside category *A* is "*p*(*Not A*)." Formally,

$p(A) + p(Not\ A) = 1.00$

> **Bounding rule**: The rule stating that all proportions range from 0.00 to 1.00.
>
> **Rule of the complement**: Based on the bounding rule, the rule stating that the proportion of cases that are not in a certain category can be found by subtracting the proportion that are in that category from 1.00.

If *p*(*A*) is known, the formula can be reconfigured thusly in order to calculate *p*(*Not A*):

$$p(Not\ A) = 1.00 - p(A) \hspace{3cm} Formula\ 5(1)$$

The variation ratio is a spinoff of the rule of the complement:

$$VR = 1.00 - \frac{f_{mode}}{N}, \hspace{3cm} Formula\ 5(2)$$

where

f_{mode} = the number of cases in the modal category and
N = the sample size.

To illustrate use of the *VR*, consider [Table 5.2](#), which contains data from the 2007 Census of State Court Prosecutors (CSCP; see [Data Sources 5.1](#)). The variable used here is the size of the population served by full-time prosecutorial offices (as summarized by Perry & Banks, 2011).

Table 5.2 Population Served by Full-Time Prosecutors' Offices	
Population Served	**f**
1,000,000 or more	43
250,000 to 999,999	211
100,000 to 249,999	341
99,999 or fewer	1,389
	N = 1,984

To compute the *VR*, first identify the mode and its associated frequency. Here, the mode is *99,999 or fewer*, and its frequency is 1,389. Now, plug the numbers into Formula 5(2):

$$VR = 1.00 - \frac{1,389}{1,984} = 1.00 - .70 = .30$$

The variation ratio is .30, which means that .30 (30%) of the cases fall outside the modal category. This is a fairly limited amount of variation—70% of offices serve relatively small populations, and just 30% serve midsized or large jurisdictions.

Data Sources 5.1 Census of State Court Prosecutors

Every 4 to 5 years, the Bureau of Justice Statistics (BJS) sends surveys to chief prosecutors in state courts. Some of the waves use random samples of all offices nationwide, and some waves employ censuses (i.e., contain the total population of offices). The most recent available wave is that for 2007, which was a census. The survey is delivered by phone, Internet, or mail to chief prosecutors, who are asked to report on various aspects of their offices' organization and operation. The data set contains information such as the number of attorneys on staff, the number of felony cases closed in the prior year, and whether the office regularly makes use of DNA analysis.

Table 5.3 displays another variable from the CSCP. This survey item asked prosecutors whether, in the past year, their offices had pursued criminal actions against defendants suspected of elder abuse.

Table 5.3 Prosecution of Elder-Abuse Cases	
Prosecuted Elder-Abuse Case in Past Year?	**f**
Yes	1,092
No	887
	N = 1,979

The mode is *Yes*, and the variation ratio is

$$VR = 1.00 - \frac{1,092}{1,979} = 1.00 - .55 = .45$$

This *VR* of .45 is close to .50, indicating that the offices were split nearly in half with respect to whether or not they had prosecuted cases of elder abuse.

In sum, the variation ratio offers information about whether the data tend to cluster inside the modal category or whether a fair number of the cases are in other categories. This statistic can be used with nominal and ordinal variables, which makes it unique relative to the range, the variance, and the standard deviation. The problem with the *VR* is that it uses only a small portion of the data available in a distribution. It indicates the proportion of cases not in the modal category, but it does not actually show where those cases are located. It would be nice to know where the data are rather than merely where they are not.

Learning Check 5.1

✓

The bounding rule and the rule of the complement are going to resurface in later chapters, so it is a good idea to make sure that you fully understand them before moving on to the of this chapter. To check your comprehension of this idea, quickly fill in the following blanks.

55% of a sample is female, and _____% is male.

96% of defendants in a sample were convicted, and _____% were not.

A judge choosing from three punishment options sentenced 70% of defendants to jail, 20% to probation, and the remaining _____% to pay fines.

The Range

The **range** (R) is the simplest measure of dispersion for continuous-level variables. It measures the span of the data or, in other words, the distance between the smallest and largest values. The range is very easy to compute:

$$R = x_{maximum} - x_{minimum}$$

Formula 5(3)

> **Range:** A measure of dispersion for continuous variables that is calculated by subtracting the smallest score from the largest. Symbolized as R.

The first step in computing the range is identification of the maximum and minimum values. The minimum is then subtracted from the maximum, as shown in Formula 5(3). That is all there is to it!

For an example, we can use data from the Uniform Crime Reports (UCR; Data Sources 1.1) on the number of juveniles arrested on suspicion of homicide in 2015 in eight states. Table 5.4 shows the data.

To calculate the range, first identify the maximum and minimum values. The maximum is 20 and the minimum is 0. The range, then, is

$R = 19 - 0 = 19$

Table 5.4 Number of Juveniles Arrested for Homicide by State, 2015	
State	*Arrests*
Maine	2
New Hampshire	1
Vermont	0
Massachusetts	3
Rhode Island	0
Connecticut	1
New York	10
New Jersey	19
$N = 8$	

Another example of the calculation of the range can be found using another variable from the CSCP. This variable collects information about the salary of the chief prosecutor in each office. Table 5.5 shows these data for the eight districts in Maine.

The largest value in Table 5.5 is 98,000, and the smallest is 66,511.06, so the range is

$R = 98,000 - 66,511.06 = 31,488.94.$

Table 5.5 Chief Prosecutor's Salary in Maine Districts	
District	**Salary ($)**
District 1	87,573.00
District 2	91,628.60
District 3	66,511.06
District 4	97,500.00
District 5	98,000.00
District 6	96,860.33
District 7	97,572.00
District 8	97,000.00
$N = 8$	

There is a difference of nearly $32,000 between the lowest-paid and the highest-paid chief prosecutors in Maine.

The range has advantages and disadvantages. First, it is both simple and straightforward. It offers useful information and is easy to calculate and understand. This same feature, however, makes this measure of dispersion too simplistic to be of much use. The range is superficial and offers minimal information about a variable. It is silent as to the distribution of the data, such as whether they are normally distributed or whether there is kurtosis present. The range can be misleading, too: Table 5.5 shows clearly that salaries cluster at the upper end of the distribution (i.e., around the $90,000 mark) and that the salary of $66,511.06 is an unusually low value. The range of $31,488.94 is mathematically correct but, in a practical sense, overstates the amount of variation in salaries. Finally, because the range does not use all of the available data, it has no place in further computations.

Learning Check 5.2

If most prosecutors' salaries nationwide were between $90,000 and $100,000, with very few values that are higher or lower, would this distribution be normally distributed, leptokurtic, or platykurtic? Explain your answer.

The Variance

Like the range, the **variance** can be used only with continuous data; unlike the range, the variance uses every number in a data set. The variance and its offshoot, the standard deviation (discussed next), are the quintessential measures of dispersion. You will see them repeatedly throughout the remainder of the book. For this reason, it is crucial that you develop a comprehensive understanding of them both formulaically and conceptually.

Variance: A measure of dispersion calculated as the mean of the squared deviation scores. Notated as $s2$.

The formula for the variance can be rather intimidating, so let us work our way up to it piece by piece and gradually construct it. We will use the juvenile homicide arrest data from Table 5.4 to illustrate the computation as we go. First, recall the concept of deviation scores that was discussed in the last chapter. Go back and review if necessary. A deviation score (symbolized d_i) is the difference between a raw score in a distribution and that distribution's mean. Recall Formula 4(4):

$$d_i = x_i - \overline{x}_i,$$

where

x_i = a given data point and

$\overline{x}$ = the sample mean

The variance is constructed from mean-based deviation scores. The first step in computing the variance is to compute the mean, and the second is to find the deviation score for each raw value in a data set. The first piece of the variance formula, therefore, is $x_i - \overline{x}$ (i.e., a deviation score for each raw value in the data). Table 5.6a shows the original data from Table 5.4 along with a deviation score column to the right of the raw scores. The mean of the data set has been calculated ($\overline{x}$ = 4.50). In the deviation score column, the mean has been subtracted from each raw score to produce a variety of positive and negative deviation scores. Recall that the sigma symbol (Σ) is a summation sign.

Deviation scores are a good first step, but what we end up with is an array of numbers. A table full of deviation scores is no more informative than a table full of raw scores. What we need is a single number that combines and represents all of the individual deviation scores. The most obvious measure is the sum—sums are good ways of packaging multiple numbers into a single numerical term. The problem with this approach, though, should be obvious: Deviation scores sum to zero. Since adding the deviation-score column will always produce a zero (or within rounding error of it), the sum is useless as a measure of variance.

Table 5.6a	Number of Juveniles Arrested for Homicide in Eight States, 2015	
State	Arrests	$x_i - \bar{x}$
Maine	2	2 − 4.50 = −2.50
New Hampshire	1	1 − 4.50 = −3.50
Vermont	0	0 − 4.50 = −4.50
Massachusetts	3	3 − 4.50 = −1.50
Rhode Island	0	0 − 4.50 = −4.50
Connecticut	1	1 − 4.50 = −3.50
New York	10	10 − 4.50 = 5.50
New Jersey	19	19 − 4.50 = 14.50
$N = 8$	$\bar{x} = \dfrac{\Sigma x}{N} = \dfrac{36}{8} = 4.50$	$\Sigma(x_i - \bar{x}) = .00$

We need to find a way to get rid of those pesky negative signs. If all of the numbers were positive, it would be impossible for the sum to be zero. Squaring is used to accomplish this objective. Squaring each deviation score eliminates the negative signs (because negative numbers always become positive when squared), and, as long as the squaring is applied to all the scores, it is not a problematic transformation of the numbers. Table 5.6b contains a new right-hand column showing the squared version of each deviation score.

Now the numbers can be summed. We can write the sum of the right-hand column in Table 5.6b as

$(x_i - \overline{x})^2$, and we can compute the answer as such:

$(x_i - \overline{x})^2$ = 6.25 + 12.25 + 20.25 + 2.25 + 20.25 + 12.25 + 30.25 + 210.25 = 314.00

This sum represents the total squared deviations from the mean. We are not quite done yet, though, because sample size must be taken into consideration in order to control for the number of scores present in the variance computation. You would not be able to compare the sum of squared deviation scores across two data sets with different sample sizes. The sum of the squared deviation scores must be divided by the sample size.

Table 5.6b Variance Calculation Table for the Number of Juveniles Arrested for Homicide in Eight States, 2015

State	Arrests	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Maine	2	$2 - 4.50 = -2.50$	$(-2.50)^2 = 6.25$
New Hampshire	1	$1 - 4.50 = -3.50$	$(-3.50)^2 = 12.25$
Vermont	0	$0 - 4.50 = -4.50$	$(-4.50)^2 = 20.25$
Massachusetts	3	$3 - 4.50 = -1.50$	$(-1.50)^2 = 2.25$
Rhode Island	0	$0 - 4.50 = -4.50$	$(-4.50)^2 = 20.25$
Connecticut	1	$1 - 4.50 = -3.50$	$(-3.50)^2 = 12.25$
New York	10	$10 - 4.50 = 5.50$	$(5.50)^2 = 30.25$
New Jersey	19	$19 - 4.50 = 14.50$	$(14.50)^2 = 210.25$
$N = 8$	$\bar{x} = \frac{\Sigma x}{N} = \frac{36}{8} = 4.50$	$\Sigma(x_i - \bar{x}) = .00$	$\Sigma(x_i - \bar{x})^2 = 314.00$

There is a bit of a hiccup, however: The variance formula for samples tends to produce an estimate of the population variance that is downwardly biased (i.e., too small) because samples are littler than populations and therefore typically contain less variability. This problem is especially evident in very small samples, such as when $N < 50$.

The way to correct for this bias in sample-based estimates of population variances is to subtract 1.00 from the sample size in the formula for the variance. Reducing the sample size by 1.00 shrinks the denominator and thereby slightly increases the number you get when you perform the division. The sample variance is symbolized s^2, whereas the population variance is symbolized σ^2, which is a lowercase Greek sigma. The reason for the exponent is that, as described earlier, the variance is made up of *squared* deviation scores. The symbol s^2 communicates the squared nature of this statistic.

We can now assemble the entire s^2 formula and compute the variance of the juvenile homicide arrest data. The formula for the sample variance is

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{N - 1} \qquad \text{Formula 5(4)}$$

Plugging in the numbers and solving yields

$$s^2 = \frac{\Sigma\left(x_i - \bar{x}\right)^2}{N-1} = \frac{314.00}{8-1} = \frac{314.00}{7} = 44.86$$

The variance of the juvenile homicide arrest data is 44.86. This is the average squared deviation from the mean in this data set.

Let's try a second example of the variance. The CSCP captures the number of felony cases each office tried before a jury in the past year. The vast majority of criminal defendants nationwide plead guilty, with a small portion opting to exercise their right to trial by jury. Table 5.7a shows the number of felony jury trials handled in the past year by a random sample of seven offices serving populations of 50,000 or less. (The letters in the left column represent the offices.) Table 5.7b displays the deviation scores and squared deviation scores.

Table 5.7a	Number of Felony Jury Trials Handled by Prosecutors' Offices	
Office		**Trials**
A		8
B		15
C		12
D		10
E		1
F		40
G		16
$N = 7$		

Table 5.7b	Variance Calculation Table for the Number of Felony Jury Trials Handled by Prosecutors' Offices		
Office	**Trials**	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
A	8	−6.57	43.16
B	15	.43	.18
C	12	−2.57	6.60
D	10	−4.57	20.88
E	1	−13.57	184.14
F	40	25.43	646.68
G	16	1.43	2.04
$N = 7$	$\bar{x} = \frac{\Sigma x}{N} = \frac{102}{7} = 14.57$	$\Sigma(x_i - \bar{x}) = .01$	$\Sigma(x_i - \bar{x})^2 = 903.7$

Applying Formula 5(4),

$$s^2 = \frac{\Sigma\left(x_i - \bar{x}\right)^2}{N-1} = \frac{903.71}{7-1} = \frac{903.71}{6} = 150.62$$

These offices' jury-trial variance is 150.62. In other words, the average squared deviation from the mean is 150.62 units.

The variance is preferable to the variation ratio and range because it uses all of the raw scores in a data set and is therefore a more informative measure of dispersion. It offers information about how far the scores are from the mean. Every case in the sample is used when the variance is calculated. When the data are continuous, the variance is better than either the *VR* or the range.

Learning Check 5.5

Variances and squared deviation scores are always positive; it is impossible to do the calculations correctly and arrive at a negative answer. Why is this?

The Standard Deviation

Despite its usefulness, the variance has an unfortunate hitch. When we squared the deviations (refer to Table 5.7b), by definition we also squared the *units* in which those deviations were measured. The variance of the juvenile homicide arrest data, then, is 44.86 *arrests squared*, and the variance for prosecutors' offices is 150.62 *trials squared*. This obviously makes no sense. The variance produces oddities such as *crimes squared* and *years squared*. The variance is also difficult to interpret—the concept of the "average squared deviation score" is not intuitive and does not provide a readily understandable description of the data. Something needs to be done to correct this. Luckily, a solution is at hand. Since the problem was created by squaring, it can be solved by doing the opposite—taking the square root. The square root of the variance (symbolized *s*, or sometimes *sd*) is the **standard deviation**:

$$s = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{N-1}} = \sqrt{s^2} \qquad \textit{Formula 5(5)}$$

> **Standard deviation**: Computed as the square root of the variance, a measure of dispersion that is the mean of the deviation scores. Symbolized as *s* or *sd*.

The standard deviation of the juvenile homicide arrest data is

$$s = \sqrt{44.86} = 6.70$$

And *s* for the prosecutors' jury trials is

$$s = \sqrt{150.62} = 12.27$$

The square root transformation restores the original units; we are back to *arrests* and *trials* now and have solved the problem of impossible and nonsensical squares. Note that the standard deviation—just like the squared deviation scores and the variance—can never be negative.

Substantively interpreted, the standard deviation is the mean of the deviation scores. In other words, it is the average distance between the individual raw scores and the distribution mean. It indicates the general spread of the data by conveying information as to whether the raw values cluster close to the mean (thereby producing a relatively small standard deviation) or are more dispersed (producing a relatively large standard deviation). The standard deviation is generally presented in conjunction with the mean in the description of a continuous variable. We can say, then, that in 2015, the eight states considered here had a mean of 4.50 juvenile arrests for homicides, with a standard deviation of 6.70. Similarly, this sample of prosecutors' offices handled a mean of 14.57 felony jury trials, with a standard deviation of 12.27.

To put the entire process together (from calculating the deviation scores all the way through to computing the standard deviation), we will use data from the Census of Jails. Table 5.8a contains a random sample of 10 jails and shows the number of juvenile (i.e., younger than age 18) girls held in each one.

Table 5.8a Number of Female Juveniles	
Jail	*Female Juveniles*
A	7
B	0
C	11
D	0
E	1
F	60
G	3
H	0
I	2
J	1
$N = 10$	

Table 5.8b Variance Calculation Table for the Number of Female Juveniles in Jails			
Jail	*Female Juveniles*	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
A	7	−1.50	2.25
B	0	−8.50	72.25
C	11	2.50	6.25
D	0	−8.50	72.25
E	1	−7.50	56.25
F	60	51.50	2652.25
G	3	−5.50	30.25
H	0	−8.50	72.25
I	2	−6.50	42.25
J	1	−7.50	56.25
$N = 10$	$\bar{x} = \dfrac{\Sigma x}{N} = \dfrac{85.00}{10} = 8.50$	$\Sigma(x_i - \bar{x}) = .00$	$\Sigma(x_i - \bar{x})^2 = 3,062.50$

Applying Formula 5(4),

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{N-1} = \frac{3,062.50}{10-1} = \frac{3,062.50}{9} = 340.28$$

Now we know the variance is 340.28. To find the standard deviation, we take the square root:

$$s = \sqrt{340.28} = 18.45$$

The standard deviation is reported more often than the variance is, for the reasons explained earlier. It is a good idea to present the standard deviation alongside any mean that you report (also expect others to do the same!). We will be using the standard deviation a lot in later chapters; it is a fundamental descriptive statistic critical to many formulas. The standard deviation also has another useful property: It can be employed to determine the upper and lower boundaries of the "typical" range in a normal distribution. When you know the

mean and the standard deviation for a sample, you can find out some important information about that distribution, as discussed in the [next section](#).

The Standard Deviation and the Normal Curve

Recall that the standard deviation is the mean of the deviation scores; in other words, it is the mean deviation between each raw score and the distribution mean. Larger standard deviations represent greater variability, whereas smaller ones suggest less dispersion. [Figure 5.3](#) displays the relationship between the mean and the standard deviation for populations and samples.

Figure 5.3 Mean and Standard Deviation for a Normally Distributed, Continuous Variable
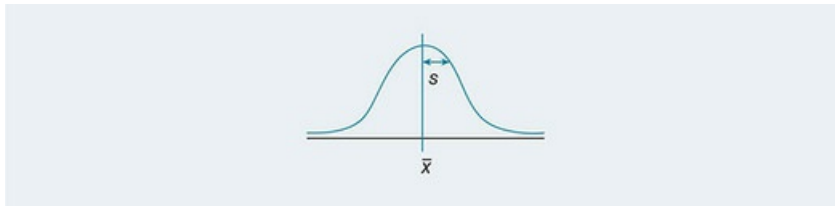
Learning Check 5.6

✓

Like the mean, the standard deviation plays a key role in many statistical analyses. We will be using both the mean and the standard deviation a lot in the following chapters; therefore, be sure that you have a solid grasp on the calculation and the interpretation of the standard deviation. As practice, revisit Table 5.5 and try calculating the standard deviation of chief prosecutors' salaries.

Research Example 5.1 Does the South Have a Culture of Honor That Increases Gun Violence?

Scholars have frequently noted that the South leads the nation in rates of violence and that gun violence is particularly prevalent in this region. This has led to the formation and proposal of multiple theories attempting to explain southern states' disproportionate involvement in gun violence. One of these theories is the "culture of honor" thesis that predicts that white male southerners are more likely than their counterparts in other regions of the country to react violently when they feel that they have been disrespected or that they, their family, or their property has been threatened. Copes, Kovandzic, Miller, and Williamson (2014) tested this theory using data from a large, nationally representative survey of adults' reported gun ownership and defensive gun use (the use of a gun to ward off a perceived attacker). A primary independent variable (IV) was whether respondents themselves currently lived in the South. The main dependent variable (DV) was the number of times respondents had used a firearm (either fired or merely brandished) to defend themselves or their property against a perceived human threat within the past 5 years. The researchers reported the descriptive statistics shown in the table below.

The authors ran a statistical analysis to determine if living in the South or in a state where the majority of the population was born in the South was related to defensive gun use. They found that it was not: Neither currently living in the South nor living in a state populated primarily with southerners increased the likelihood that respondents had used a gun defensively in the past 5 years. These findings refuted the southern culture of honor thesis by suggesting that southern white males are no more likely than white males in other areas of the country to resort to firearms to defend themselves.

Variable	Mean	Standard Deviation	Percentage
DV: Number of defensive gun uses in the past 5 years	1.04	.19	
Percentage of sample currently living in the South			34.0
Percentage of white state population born in the South	70.88	12.85	
White homicide rates in cities of respondents' residence	5.60	3.45	
Respondents' perceptions of crime in their neighborhoods (5-point scale)	2.50	1.12	

Source: Adapted from Table 1 in Copes et al. (2014).

Source: Adapted from Table 1 in Copes et al. (2014).

Do Neighborhoods With Higher Immigrant Concentrations Experience More Crime?

Many studies have examined the possible link between immigration and crime. Popular assumptions that immigrants commit crimes at high rates have been proven false by numerous studies uncovering no relationship between immigrant concentrations and crime rates, or even negative relationships wherein crime rates were lower in cities with larger immigrant populations than in those with fewer immigrants. Most of the research into immigration and crime, however, has been based in the United States. Sydes (2017) extended this empirical inquiry to Australia. She collected data on neighborhoods in two Australian cities. Means and standard deviations for the DV and several of the IVs are shown in the table. After running sophisticated regression models, Sydes found no relationship between immigrant concentrations and violent crime within neighborhoods in any given year, but also discovered that increases in immigrant populations from one year to the next were associated with statistically significant reductions in violence. These results are similar to most of the research conducted in the United States and lend support to the conclusion that immigration does not increase crime and might even reduce it.

Variable	Sydney		Brisbane	
	Mean	Standard Deviation	Mean	Standard Deviation
DV: Violent crime counts	58.91	87.21	21.47	25.42
Percent immigrant	30.75	12.98	23.81	8.75
Language diversity	47.20	22.27	29.04	15.74
Disadvantage	.11	1.06	.11	.82
Population density	29.15	21.76	14.27	10.61
Percent residential land use	66.62	18.99	66.12	18.17
Nearby violent crime	6.42	.71	5.40	1.03

Source: Adapted from Tables 1 and 2 in Sydes (2017).
Source: Adapted from Tables 1 and 2 in Sydes (2017).

A distinguishing characteristic of normal distributions is that approximately two-thirds of the scores in a distribution are located within one standard deviation below and one standard deviation above the mean. Figure 5.4 depicts this important property of normal distributions. This distance between one standard deviation below and one standard deviation above the mean constitutes the "normal" or "typical" range of values. Most of the scores fall within this range, with a smaller number being higher or lower.

To illustrate this, let us suppose we have a sample of housecats with a mean weight of 11 pounds and a standard deviation of 2 pounds (see Figure 5.5). We can find the range of weights that fall within the "normal zone" using subtraction and addition. Two-thirds of the cats would be between 11 − 2 = 9 pounds (i.e., one standard deviation below the mean) and 11 + 2 = 13 pounds (one *sd* above the mean). The remaining one-third would weigh less than 9 pounds (i.e., they would be more than one *sd* below the mean) or more than 13 pounds (greater than one *sd* above the mean). These extreme values would indeed occur but with relatively low frequency. People, objects, and places tend to cluster around their group means, with extreme values being infrequent and improbable. Remember this point; we will come back to it in later chapters.

Whenever you know the mean and the standard deviation of a normally distributed, continuous variable, you

can find the two values between which two-thirds of the cases lie (i.e., the outer boundaries of the "typical" area). In a distribution with a mean of 100 and a standard deviation of 15, two-thirds of cases will be between 85 and 115.

Learning Check 5.7

The concept regarding two-thirds of cases lying between −1 *sd* and +1 *sd* will form a fundamental aspect of later discussions. You will need to understand this concept both mathematically and conceptually. Practice computing the upper and lower limits using the means and the standard deviations listed in the first table in Research Example 5.1.

Figure 5.4 In a Normal Distribution, Approximately Two-Thirds of the Scores Lie Within One Standard Deviation of the Mean
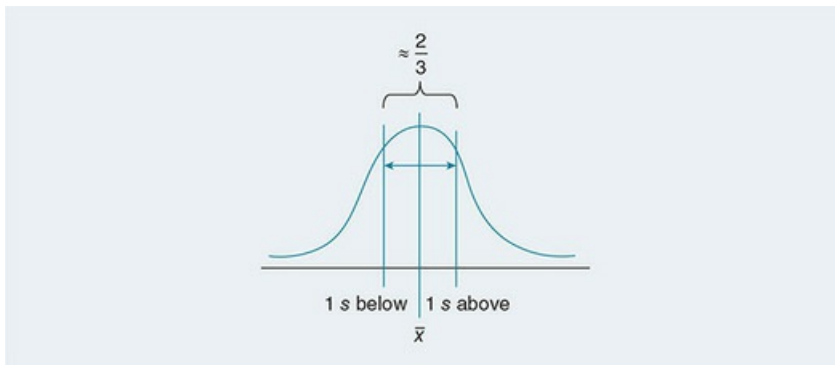
Figure 5.5 Hypothetical Distribution of Housecats' Weights
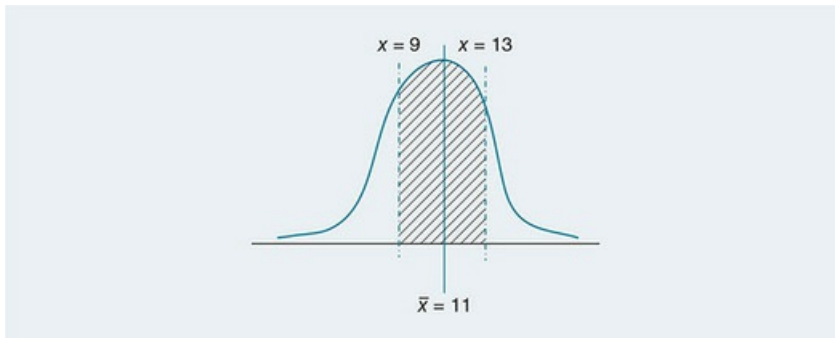
Research Example 5.2 Why Does Punishment Often Increase—Rather Than Reduce—Criminal Offending?

Deterrence is perhaps the most pervasive and ingrained punishment philosophy in the United States and, indeed, in much of the world. It is commonly assumed that punishing someone for a criminal transgression will lessen the likelihood of that person reoffending in the future. Several studies have noted, however, that the chance of offending actually *increases* after someone has been punished. Offenders who have been caught, moreover, tend to believe that their likelihood of being caught again are very low because they think that it is improbable that the same event would happen to them twice. This flawed probabilistic reasoning is called the "gambler's bias" or "gambler's fallacy."

Pogarsky and Piquero (2003) attempted to learn more about the way that the gambler's bias could distort people's perceptions of the likelihood of getting caught for criminal wrongdoing. They distributed a survey with a drunk driving scenario to a sample of university students and asked the respondents about two DVs: (a) On a scale of 0 to 100, how likely they would be to drive under the influence in this hypothetical situation, and (b) on a scale of 0 to 100, how likely it is that they would be caught by police if they did drive while intoxicated. The researchers also gathered data on several IVs, such as respondents' criminal histories (which were used to create a risk index scale), levels of impulsivity in decision making, and ability to correctly identify the probability of a flipped coin landing on tails after having landed heads side up four times in a row. The coin-flip question tapped into respondents' ability to use probabilistic reasoning correctly; those who said that the coin is more likely to land tails up were coded as engaging in the type of logical fallacy embodied by the gambler's bias. The researchers obtained the means and standard deviations shown in the table on

page 126.

Pogarsky and Piquero (2003) divided the sample into those at high risk of offending and those at low risk and then analyzed the relationship between the gambler's fallacy and perceived certainty of punishment within each group. They found that the high-risk respondents' perceptions of certainty were not affected by the gambler's bias. Even though 26% of people in this group engaged in flawed assessments of probabilities, this fallacious reasoning did not impact their perceptions of the certainty of punishment.

Among low-risk respondents, however, there was a tendency for those who engaged in flawed probabilistic reasoning to believe that they stood a very low chance of detection relative to those low-risk respondents who accurately assessed the probability of a coin flip landing tails side up. The researchers concluded that people who are at high risk of offending might not even stop to ponder their likelihood of being caught and will proceed with a criminal act when they feel so inclined. Those at low risk, on the other hand, attempt to employ probabilities to predict their chances of being apprehended and punished. The gambler's bias, therefore, might operate only among relatively naive offenders who attempt to use probabilistic reasoning when making a decision about whether or not to commit a criminal offense.

Variable	Mean	Standard Deviation
DV: Offending likelihood	38.28	36.14
DV: Perceived certainty of punishment	28.48	36.13
Number times pulled over while driving drunk	.24	.63
Number of days in past month consumed 3+ drinks	4.69	5.00
Number of times shoplifted	3.69	11.74
Number of times vandalized property	3.78	4.46
Number of 10 closest friends who have driven drunk at least once	4.50	3.34
Total risk index score	4.84	2.67

Source: Adapted from the appendix in Pogarsky and Piquero (2003).
Source: Adapted from the appendix in Pogarsky and Piquero (2003).

SPSS

SPSS offers ranges, variances, and standard deviations. It will not provide you with variation ratios, but those are easy to calculate by hand. The process for obtaining measures of dispersion in SPSS is very similar to that for measures of central tendency. The juvenile homicide data set from Table 5.4 will be used to illustrate SPSS. First, use the *Analyze → Descriptive Statistics → Frequencies* sequence to produce the main dialog box. Then click the *Statistics* button in the upper right to produce the box displayed in Figure 5.6. Select *Std. deviation*, *Variance*, and *Range* to obtain these three statistics.

Figure 5.6 Using SPSS to Obtain Measures of Dispersion
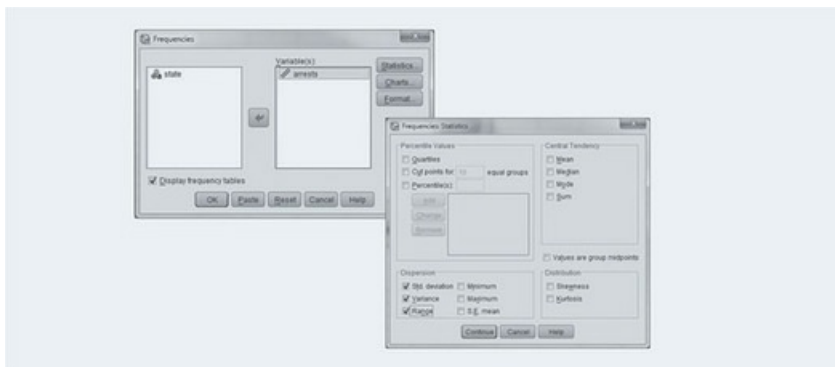
After you click *Continue* and *OK*, the output in Figure 5.7 will appear.

You can see in Figure 5.7 that all the numbers generated by SPSS match those that we obtained by hand.

Follow along with this example of using SPSS to obtain measures of central tendency by downloading the file *Juvenile Arrests for Chapter 5.sav* at **www.sagepub.com/gau**.

Figure 5.7 SPSS Output for Measures of Dispersion

Chapter Summary

This chapter introduced four measures of dispersion: variation ratio, range, variance, and standard deviation. The *VR* tells you the proportion of cases not located in the modal category. *VR* can be computed on data of any level of measurement; however, it is the only measure of dispersion covered in this chapter that is available for use with categorical data. The range, the variance, and the standard deviation, conversely, can be used only with continuous variables.

The range is the distance between the lowest and highest value on a variable. The range provides useful information and so should be reported in order to give audiences comprehensive information about a variable. This measure's usefulness is severely limited by

the fact that it accounts for only the two most extreme numbers in a distribution. It ignores everything that is going on between those two endpoints.

The variance improves on the range by using all of the raw scores on a variable. The variance is based on deviation scores and is an informative measure of dispersion. This measure, though, has a conceptual problem: Because computation of the variance requires all of the deviation scores to be squared, the units in which the raw scores are measured also end up getting squared. The variance thus suffers from a lack of interpretability.

The solution to the conceptual problem with the variance is to take the square root of the variance. The square root of the variance is the standard deviation. The standard deviation is the mean of the deviation scores. Raw scores have a mean, and so do deviation scores. The former is the distribution mean, and the latter is the standard deviation.

The mean and the standard deviation together are a staple set of descriptive statistics for continuous variables. The standard deviation will be key to many of the concepts that we will be discussing in the next few chapters. Approximately two-thirds of the cases in any normal distribution are located between one standard deviation below and one standard deviation above the mean. These scores are within the normal or typical range; scores that are more than one standard deviation below or above the mean are relatively uncommon.

Thinking Critically

1. It is common practice to present the mean and standard deviation together (as seen in the research examples in this chapter). Why is this? In other words, explain (1) the reasons why a measure of central tendency and a measure of dispersion would be reported together to describe the same set of data, and (2) why the mean and standard deviation are the measures typically selected for this. Refer back to Chapter 4 if needed.
2. Suppose you are reading a news article discussing bank-fraud victimization. The report features the story of one victim who was 20 years old when she discovered her bank account had been compromised and her money had been stolen. Because you are familiar with the National Crime Victimization Survey's Identity Theft Supplement, you know that the mean age of bank-fraud victims is 45.52 ($s = 15.26$). Based on this, is the victim in the report typical of bank-fraud victims in terms of age? How do you know? Do you trust that her experiences are representative of other bank-fraud victims? Explain your answer.

Review Problems

1. Explain the reason why measures of dispersion are necessary in addition to measures of central tendency. That is, what information is given by measures of dispersion that is not provided by measures of central tendency?
2. A data distribution that was very narrow and peaked would be considered
   1. normal.
   2. platykurtic.
   3. leptokurtic.
3. In a normal curve, approximately _____ of values are between one standard deviation above and one standard deviation below the mean.
4. A data distribution that was very flat and spread out would be considered
   1. normal.
   2. platykurtic.
   3. leptokurtic.
5. The following table contains data from the Census of Jails. Compute the variation ratio.

Facility Size	f
Small	778
Medium	786
Large	807
	N = 2,371

6. The following table contains data from the PPCS. Compute the variation ratio.

Household Income	f
Less than $20,000	1,276
$20,000 – $49,999	1,057
$50,000 or more	1,948
	N = 4,281

7. The following table contains data from the Census of Jails. Compute the variation ratio.

Race of Security Staff in Small Jails	f
White	6,036
Black	607
Hispanic	309
Other	4
	N = 6,956

8. The table below contains UCR data showing the 2015 rate of officer assaults (per 100 officers) in each of the four regions of the country. Use this table to do the following:

Region	Assaults per 100 Officers
Northeast	7.70
Midwest	8.90
South	9.60
West	12.60
N = 4	

   1. Calculate the range.
   2. Calculate the mean assault rate.
   3. Calculate the variance.
   4. Calculate the standard deviation.

9. The following table shows data from the BJS on the number of death-row inmates housed by states in the Midwest in 2013. Use this table to do the following:

State	Inmates
Louisiana	84
Mississippi	50
Nevada	81
Arizona	122
Pennsylvania	190
Oklahoma	48
N = 6	

   1. Calculate the range.
   2. Calculate the mean number of inmates per state.
   3. Calculate the variance.
   4. Calculate the standard deviation.

10. The following table contains UCR data on the number of hate crimes in 2015. Use this table to do the following:

State	Hate Crimes
Alaska	19
Delaware	11
Maine	45
Idaho	46
Florida	83
New Hampshire	15
Alabama	12
Illinois	101
N = 8	

1. Calculate the range.
2. Calculate the mean number of firearm-perpetrated murders in these states.
3. Calculate the variance.
4. Calculate the standard deviation.

11. The following table shows UCR data on the percentage of burglaries cleared by arrest in 2015, as broken down by region. Use this table to do the following:

Region	Clearance Rate
New England	13.70
Middle Atlantic	18.70
East North Central	9.90
West North Central	12.40
South Atlantic	16.20
East South Central	14.00
West South Central	10.50
Mountain	10.40
Pacific	11.00
N = 9	

1. Calculate the range.
2. Calculate the mean clearance rate per region.
3. Calculate the variance.
4. Calculate the standard deviation.

12. For each of the following means and standard deviations, calculate the upper and lower limits of the middle two-thirds of the distribution.

1. $\overline{X}$ = 6.00, sd = 1.50
2. $\overline{X}$ = 14.00, sd = 3.40

3. $\overline{X}$ = 109.32, *sd* = 14.98

13. For each of the following means and standard deviations, calculate the upper and lower limits of the middle two-thirds of the distribution.

    1. $\overline{X}$ = 63.10, *sd* = 18.97
    2. $\overline{X}$ = 1.75, *sd* = .35
    3. $\overline{X}$ = 450.62, *sd* = 36.48

14. Explain the conceptual problem with the variance that is the reason why the standard deviation is generally used instead.

15. Explain the concept behind the standard deviation; that is, what does the standard deviation represent substantively?

16. The companion website (www.sagepub.com/gau) has an SPSS file called *Firearm Murders for Chapter 5.sav* that contains 2015 UCR data showing the percentage of murders, per state, that were committed with firearms. (There are 49 states in this file because Florida did not submit UCR data. In addition, murders for which no weapon data were submitted are excluded from these counts.) Use SPSS to obtain the range, the variance, and the standard deviation for the variable *percent*.

17. There is an SPSS file called *Census of State Prosecutors for Chapter 5.sav* on the website (www.sagepub.com/gau) that contains data from the 2007 CSCP. This data set contains three variables. For the variable *yearsinoffice*, which measures the number of years chief prosecutors have held their positions, use SPSS to obtain the range, the standard deviation, and the variance.

18. Using *Census of State Prosecutors for Chapter 5.sav*, use SPSS to obtain the range, standard deviation, and variance of the variable *assistants*, which shows the number of full-time assistant prosecutors employed by each office.

19. Using *Census of State Prosecutors for Chapter 5.sav* and the variable *felclosed*, which captures the number of felony cases each office closed in 2007, use SPSS to obtain the range, the standard deviation, and the variance.

Key Terms

Glossary of Symbols and Abbreviations Introduced in This Chapter

VR	Variation ratio
R	Range
s^2	Variance
s	Standard deviation
sd	Alternative notation for the standard deviation

Part II Probability and Distributions

Part I of this book introduced you to descriptive statistics. You learned the mathematical and conceptual underpinnings of proportions, means, standard deviations, and other statistics that describe various aspects of data distributions. Many times, though, researchers want to do more than merely describe a sample—they want to run a statistical test to analyze relationships between two or more variables. The problem is there is a gap between a sample and the population from which it was drawn. Sampling procedures produce a subset of the population, and it is not correct to assume that this subset's descriptive statistics (such as its mean) are equivalent to those of the population. Going back to our housecat example in the previous chapter, suppose you go door to door in your neighborhood recording the weights of all the cats, and you find a mean weight of 9.50 pounds. Would it be safe to conclude that if you weighed all housecats in the world (i.e., the entire population), the mean would be exactly 9.50? Definitely not! In the process of pulling your sample, you might have picked up a few cats that are atypically small or large. We saw in the previous chapter how extreme values can pull the sample mean up or down; if you got an extremely large or extremely small cat in your sample, the mean would be thrown off as a result. There is always a possibility that any given sample contains certain values that cause the mean, the proportion, or other descriptive statistic to be higher or lower than that for the entire population from which the sample was derived. For this reason, sample statistics cannot be automatically generalized to populations. We need a way of bridging the gap.

Inferential statistics (also called hypothesis testing, the subject of Part III of this book) provide this bridge between a descriptive statistic and the overarching population. The purpose of inferential statistics is to permit a descriptive statistic to be used in a manner such that the researcher can draw an inference about the larger population. This procedure is grounded in **probability theory**. Probability forms the theoretical foundation for statistical tests and is therefore the subject of Part II of this book. You can think of Part I as having established the foundational mathematical and formulaic concepts necessary for inferential tests and of Part II as laying out the theory behind the strategic use of those descriptive statistics. Part III is where these two areas of knowledge converge.

Inferential statistics: The field of statistics in which a descriptive statistic derived from a sample is employed probabilistically to make a generalization or inference about the population from which the sample was drawn.

Probability theory: Logical premises that form a set of predictions about the likelihood of certain events or the empirical results that one would expect to see in an infinite set of trials.

Part II is heavily grounded in theory. You will not see SPSS sections or much use of research examples. This is because probability is largely concealed from view in criminal justice and criminology research; probability is the "man behind the curtain" who is pulling the levers and making the machine run but who usually remains

hidden. Although there will be formulas and calculations that you will need to understand, your primary task in [Part II](#) is to form conceptual comprehension of the topics presented. When you understand the logic behind inferential statistics, you will be ready for [Part III](#).

Chapter 6 Probability

A probability is the likelihood that a particular event will occur. We all use probabilistic reasoning every day. If I buy a lottery ticket, what are my chances of winning? What is the likelihood that I will get a promotion if I put in extra effort at work? What is the probability that I will get pulled over if I drive 5 miles per hour over the speed limit? These musings involve predictions about the likelihood that a certain event will (or will not) occur. We use past experiences and what we know (or what we assume to be true) about the world to inform our judgments about the chance of occurrence.

Probability: The likelihood that a certain event will occur.

Probabilities are linked intricately with proportions; in fact, the probability formula is a spinoff of Formula 3(1). Flip back to this formula right now for a refresher. Let us call a particular event of interest A. Events are phenomena of interest that are being studied. The probability that event A will occur can be symbolized as p (A) (pronounced "p of A") and written formulaically as

$$p(A) = \frac{\text{The number of times event } A \text{ can occur}}{\text{The total number of possible outcomes}} \qquad \text{Formula 6(1)}$$

Coin flips are a classic example of probabilities. Any two-sided, fair coin will land on either heads or tails when flipped, so the denominator in the probability formula is 2 (i.e., there are two possible outcomes of the coin flip). There is one tails side on a coin, so the numerator is 1 (i.e., there is only one option for the coin to land tails-side up). Probabilities, like proportions, are expressed as decimals. The probability of the coin landing tails side up, then, is

$$p(A) = \frac{1}{2} = .50$$

Any time you flip a fair coin, there is a .50 probability that the coin will land on tails. Of course, the probability that it will land heads side up is also .50. Note that the two probabilities together sum to 1.00:

$p(tails) + p(heads) = .50 + .50 = 1.00$

The probabilities sum to 1.00 because heads and tails are the only two possible results and thus constitute an exhaustive list of outcomes. The coin, moreover, must land (it will not hover in midair or fly around the

room), so the probability sums to 1.00 because the landing of the coin is inevitable.

Think about rolling one fair die. A die has six sides, so any given side has

$$p(\text{any side}) = \frac{1}{6} = .17$$

of being the one that lands faceup. If you were asked, "What is the probability of obtaining a 3 on a single roll of a fair die?" your answer would be ".17." There are 52 cards in a standard deck and only one of each number and suit, so during any random draw from the deck, each card has

$$p(\text{any card}) = \frac{1}{52} = .02$$

probability of being the card that is selected. If someone asked you, "What is the probability of selecting the two of hearts from a full deck?" you would respond, ".02." Note, too, that in the instance of a die and a deck of cards—just as with the flipped coin—the sum of all events' individual probabilities is 1.00 (or within rounding error of it). This is a reflection of the bounding rule and the rule of the complement that we discussed in Chapter 5. A die has six sides, so there are six total possible events: The die can land on 1, 2, 3, 4, 5, or 6. The probability that it will land on a specific predicted value (say, 5, e.g.) is .17, whereas the probability that it will land on *any* of its six sides is .17 + .17 + .17 + .17 + .17 + .17 = 1.02 (technically, this sum is zero, but the numbers used in the example contain rounding error). Likewise, the probability of pulling a predetermined playing card (for instance, the four of clubs) is .02, but the chance that you will, in fact, retrieve *some* card from the deck when you pull one out is .02(52) = 1.04 (exceeding 1.00 due to rounding error).

Probabilities can be used to make specific predictions about how likely it is that an event will occur. Maybe you want to predict that you will draw any card from the diamonds suit. Decks contain four suits, each with 13 cards, meaning there are 13 diamond cards. The probability of drawing any one of them is

$$p(diamond) = \frac{13}{52} = .25.$$

This is different from the previous example using the two of hearts because here we do not care what number the card is as long as it belongs to the diamond suit. Likewise, there are four cards of each number (one of each suit), so the probability that your random draw would produce a

$$p(9) = \frac{4}{52} = .08.$$

nine of any suit would be Again, this is different from .02 because the prediction that a draw will yield a nine of any suit opens up more options (four, to be exact) than when the suit is restricted (resulting in just one option).

Table 6.1 Gender of Police–Public Contact Survey (PPCS) Respondents		
Gender	f	p
Male	25,078	.48
Female	27,451	.52
	$N = 52,529$	$\Sigma = 1.00$

The major difference between proportions and probabilities is that proportions are purely descriptive, whereas probabilities represent predictions. Consider Table 6.1, which shows the proportion of the respondents to the Police–Public Contact Survey (PPCS; see Data Sources 2.1) that is male and the proportion that is female. If you threw all 52,529 people into a gigantic hat and randomly drew one, what is the probability that the person you selected would be female? It is .52! The probability that the person would be male? The answer is .48. This is the relationship between proportions and probabilities.

Let us try another example. Table 6.2 shows data from the Uniform Crime Reports (UCR; see Data Sources 1.1) showing the percentage of crimes cleared by arrest in 2015.

Table 6.2 Percentage of Crimes Cleared by Arrest	
Crime Type	Percentage Cleared
Homicide	61.50
Rape	37.80
Robbery	29.30
Aggravated assault	54.00
Burglary	12.90
Larceny-theft	21.90
Motor vehicle theft	13.10

We can use the percentages in Table 6.2 just as if they were proportions. (Flip back to Formulas 3[1] and 3[2] if you need a refresher on this concept.) What is the probability that a robbery reported to police will result in an arrest of the perpetrator? You can see that 29.30% of these crimes are cleared by arrest, which means that

any given robbery has a .29 chance of resulting in an arrest. Try this exercise with the other crime types listed in the table.

Probability theory is grounded in assumptions about infinite trials; in other words, probabilities concern what is expected over the long run. Think back to the coin flip example. Since p (*tails*) on any given flip is .50, then over the course of a long-term flipping session, half of the flips will yield tails and the other half will land on heads. In other words, if you spent several hours flipping one coin and recording each outcome, eventually you would end up with a 50/50 split.

There is a distinct difference, though, between **theoretical predictions** and **empirical outcomes.** (Empirical outcomes can also be called *observed outcomes* ; we will use the terms interchangeably.) Quite simply, and as you have undoubtedly already learned in life, you do not always get what you expect. You can see this in practice. First, imagine that you flipped a coin six times. How many of those flips would you expect to land tails side up? Three, of course, because every flip has p (*tails*) = .50 and so you would expect half of all flips to result in tails. Now, grab a real coin, flip it six times, and record each outcome. How many tails did you get? Try 20 flips. Now how many tails? Transfer the following chart to a piece of paper and fill it in with your numbers.

6 flips proportion tails = $\dfrac{?}{6}$ =

20 flips proportion tails = $\dfrac{?}{20}$ =

Theoretical prediction: A prediction, grounded in logic, about whether or not a certain event will occur.

Empirical outcome: A numerical result from a sample, such as a mean or frequency. Also called observed outcomes.

You might have found in your six-flip exercise that the number of tails departed noticeably from three; you might have gotten one, five, or even six or zero tails. On the other hand, the number of tails yielded in the 20-flip exercise should be approximately 10 (with a little bit of error; it might have been 9 or 11). Why is this? It is because you increased the number of trials and thereby allowed the underlying probability to emerge in the empirical outcomes. The knowledge that half of coin flips will produce tails *over time* is a theoretical

prediction based on the concept of infinite coin flipping (nobody has actually spent infinity flipping a coin), whereas the flip experiment you just conducted is an empirical test and outcome. Theory guides us in outlining what we expect to see (e.g., if you are asked how many times you would expect to see tails in a series of six flips, your best guess would be three), but sometimes empirical results do not match expectations (e.g., you might see one tail, or possibly all six will be tails).

The relationship between theoretical predictions and empirical findings is at the heart of statistical analysis. Researchers constantly compare observations to expectations to determine if empirical outcomes conform to theory-based predictions about those outcomes. The extent of the match (or mismatch) between what we expect (theory) and what we see (reality) is what leads us to make certain conclusions about both theory and reality.

Learning Check 6.3

✓

If you are finding the material in this chapter to be about as clear as mud, then you are experiencing the perfectly normal learning process for this topic. These ideas take time to sink in. A highly recommended study technique is to spend time simply thinking about these concepts. Work them over in your mind. Take breaks as you read so that you can digest concepts before continuing. Practice additional probability calculations or try concrete examples like drawing cards, flipping a coin, or rolling a die.

Now, let us add another layer to the discussion. We have thus far covered four examples of situations in which probability can be used to make predictions: coin tosses, die rolls, card selections, and clearance rates. You might have noticed that clearance rates stand apart from the other three examples—unlike coins, dice, and cards, the probability of clearance is not equal across the different crime types. Whereas a rolled die offers equal probability of landing on two versus six (each one is .17), a crime will vary in its clearance probability based on the type of crime that it is. Motor vehicle theft has a .13 clearance probability, and homicide has a .62 chance of being cleared. Unlike each of the six sides of a die, any given crime does not have an equal probability of clearance; instead, a crime's probability of being solved depends in large part on the type of crime it is.

This leads us to an important point: In most cases, the different possible outcomes have *unequal* probabilities of occurring. The existence of outcomes that have greater or lesser theoretical probabilities of being the one that actually occurs forms the basis for everything else we are going to discuss in this chapter. Put simply, some outcomes are more likely than others.

Learning Check 6.4

✓

The general population is roughly 50% female, meaning that approximately 50% of newborns are female. Suppose there are two hospitals, Hospital A and Hospital B. In Hospital A, four babies are born each day. In Hospital B, 20 babies are born per day. In which hospital would you expect to see roughly half of the babies born each day be female? That is, in which hospital would the sex of the babies born each day be representative of the entire population? Explain your answer. What about over the course of a year? If you tracked the sex of babies born for a year, would there be a difference between the two hospitals' female-baby percentages? Why or why not?

A **probability distribution** is a table or graph showing the full array of theoretical probabilities for any given variable. These probabilities represent not what we *actually* see (i.e., tangible empirical outcomes) but, rather, the gamut of potential empirical outcomes and each outcome's probability of being the one that actually happens. Probability distributions are theoretical. A probability distribution is constructed on the basis of an underlying parameter or statistic (such as a proportion or a mean) and represents the probability associated with each possible outcome. Two types of probability distributions are discussed in this chapter: binomial and continuous.

> **Probability distribution**: A table or graph showing the entire set of probabilities associated with every possible empirical outcome.

Discrete Probability: The Binomial Probability Distribution

A **trial** is a particular act with multiple different possible outcomes (e.g., rolling a die, where the die will land on any one of its six sides). **Binomials** are trials that have exactly two possible outcomes; this type of trial is also called *dichotomous* or *binary*. Coin flips are binomials because coins have two sides. describes two types of binomials that criminal justice and criminology researchers have examined. Binomials are used to construct **binomial probability distributions**. The binomial probability distribution is a list of expected probabilities; it contains all possible results over a set of trials and lists the probability of each result.

> **Trial**: An act that has several different possible outcomes.
>
> **Binomial**: A trial with two possible outcomes. Also called a dichotomous or binary variable empirical outcome.
>
> **Binomial probability distribution**: A numerical or graphical display showing the probability associated with each possible outcome of a trial.

So, how do we go about building the binomial probability distribution? The distribution is constructed using the **binomial coefficient**. The formula for this coefficient is a bit intimidating, but each component of the coefficient will be discussed individually in the following pages so that when you are done reading, you will understand the coefficient and how to use it. The binomial coefficient is given by the formula

$$p(r) = \binom{N}{r} p^r q^{N-r}, \qquad \text{Formula 6(2)}$$

where

$p(r)$ = the probability of r,

r = the number of successes,

N = the number of trials or sample size,

p = the probability that a given event will occur, and

q = the probability that a given event will not occur.

> **Binomial coefficient**: The formula used to calculate the probability for each possible outcome of a trial and to create the binomial probability distribution.

> Research Example 6.1 Are Police Officers Less Likely to Arrest an Assault Suspect When the Suspect and the Alleged Victim Are Intimate Partners?
>
> Critics of the police response to intimate partner violence have accused police of being "soft" on offenders who abuse intimates. Klinger (1995) used a variable measuring whether or not police made an arrest when responding to an assault of any type. The variable was coded as *arrest/no arrest*. He then examined whether the probability of arrest was lower when the perpetrator and victim were intimates as compared to assaults between strangers or non-intimate acquaintances. The results indicated that police were unlikely to make arrests in *all* types of assault, regardless of the victim–offender relationship and that they were not less likely to arrest offenders who victimized intimate partners relative to those who victimized strangers or acquaintances.

Are ex-prisoners returning to society less likely to recidivate if they are given reentry services?

Researchers, corrections agencies, and community groups constantly seek programming that reduces recidivism among recently released prisoners by helping them transition back into society. The reentry process is difficult—newly released ex-prisoners struggle to find housing, transportation, and employment. Many have ongoing problems with substance dependence. Ray, Grommon, Buchanan, Brown, and Watson (2017) evaluated recidivism outcomes for clients of a federally funded program implemented across five organizations in Indiana that provided reentry services to adult ex-prisoners with histories of drug or alcohol abuse. Recidivism was coded as *yes/no* to indicate whether each person was returned to prison after completing the program. Approximately one-third recidivated. There were significant differences in recidivism rates across the five agencies, suggesting that better-funded agencies combining multiple different therapeutic and social-support services produce lower recidivism rates compared to those with fewer resources that offer a narrow range of services.

The ultimate goal of binomials is to find the value of $p(r)$ for every possible value of r. The resulting list of $p(r)$ values is the binomial probability distribution.

Before we get into the math, let's first consider a conceptual example using the clearance data. Table 6.2 shows that 61.50% of homicides result in arrest, meaning that any given homicide has a .62 probability of clearance. Suppose you gathered a random sample of 10 homicide cases. Out of these 10, there are 11 separate possible (i.e., theoretical) outcomes: Anywhere from none of them to all 10 could have been cleared by arrest. Each of these individual outcomes has a certain probability of being the one that occurs in reality. You might have already guessed that the most likely outcome is that 6 of them would be cleared (since 61.50% of 10 is 6.25), but what are the other possible outcomes' probabilities of occurrence? This is what we use the binomial probability distribution for.

Successes and Sample Size: N and r

The binomial coefficient formula contains the variables N and r. The N represents sample size or the total number of trials. As an example, we will use a hypothetical study of jail inmates who have been booked on arrest and are awaiting word as to whether they will be released on bail or whether they will remain confined while they await criminal-court proceedings. The Bureau of Justice Statistics (BJS) reports that 62% of felony defendants are released from jail prior to the final disposition of their case (Reaves, 2013), so $p = .62$. Let's say we draw a random sample of five recently arrested jail inmates. This means that for the binomial coefficient formula, $N = 5$.

In the binomial coefficient, r represents the number of **successes** or, in other words, the number of times the outcome of interest happens over N trials. A researcher decides what "success" will mean in a given study. In the current example, we will define success as a defendant obtaining pretrial release. There are multiple values that r takes on in any given study, since there are multiple possible successes. If we wanted to find the probability that three of the five defendants would be released, we would input $r = 3$ into the binomial coefficient formula; if we wanted the probability that all five would be released, then $r = 5$.

Success: The outcome of interest in a trial.

The Number of Ways r Can Occur, Given N: The Combination

In our sample of five defendants, there are a lot of possibilities for any given value of r. For instance, if one defendant was released (i.e., $r = 1$), then it could be that Defendant 1 was released and the remaining four were detained. Alternatively, Defendant 4 might have been the one who made bail. The point is, a given value of r can occur in multiple different ways. If a release is symbolized r, then, for an illustration, let us call a detention d. (Remember that release and detention are binary; each defendant receives one of these two possible outcomes.) Consider the following sets of possible arrangements of outcomes for one release and four detentions:

$$\{r, d, d, d, d\}\{d, r, d, d, d\}\{d, d, r, d, d\}\{d, d, d, r, d\}\{d, d, d, d, r\}$$

What these sets tell you is that there are five different ways for $r = 1$ (i.e., one success) to occur over $N = 5$ trials. The same holds true for any number of successes—there are many different ways that $r = 2$, $r = 3$, and so on can occur in terms of which defendants are the successes and which are the failures.

The total number of ways that r can occur in a sample of size N is called a **combination** and is calculated as

$$\frac{N!}{r!(N-r)!},$$
Formula 6(3)

where

    N = the total number of trials or total sample size,

    r = the number of successes, and

    ! = factorial.

Combination: The total number of ways that a success r can occur over N trials.

Learning Check 6.5

✓

In this and subsequent chapters, you will see various notations representing multiplication. The most popular symbol, $\times$, will not be used; in statistics, x represents raw data values, and so using this symbol to also represent multiplication would be confusing. Instead, we will rely on three other indicators of multiplication. First are parentheses; numbers separated by parentheses should be multiplied together. This could appear as 2(3) = 6, or (2)(3) = 6. Second, sometimes no operand is used, and it is merely the placement of two numbers or symbols right next to each other that signals multiplication. An example of this is xy, where whatever numbers x and y symbolized, respectively, would be multiplied together. If $x = 2$ and $y = 3$, then $xy = 2(3) = 6$. Third is the centered dot that connects the numbers being multiplied, such as $2 \cdot 3 = 6$.

Learning Check 6.6

✔

Most calculators meant for use in math classes (which excludes the calculator that came free with that new wallet you bought) will compute factorials and combinations for you. The factorial function is labeled as *!*. The combination formula is usually represented by *nCr* or sometimes just *C*. Depending on the type of calculator you have, these functions are probably either located on the face of the calculator and accessed using a *2nd* or *alpha* key, or can be found in a menu accessed using the *math, stat* , or *prob* buttons, or a certain combination of these functions. Take a few minutes right now to find these functions on your calculator.

Also, note that 1 factorial and 0 factorial both equal 1. Try this out on your calculator to see for yourself. Since 0! = 1, you do not need to worry if you end up with a zero in the denominator of your fractions.

1! = 1

0! = 1

The **factorial** symbol *!* tells you to multiply together the series of descending whole numbers starting with the number to the left of the symbol all the way down to 1.00. If $r = 3$, then $r! = 3 \cdot 2 \cdot 1 = 6$. If $N = 5$, then $N! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

Factorial: Symbolized *!* , the mathematical function whereby the first number in a sequence is multiplied successively by all numbers below it down to 1.00.

There is also shorthand notation for the combination formula that saves us from having to write the whole thing out. The shorthand notation is $\binom{N}{r}$, which is pronounced "N choose r." Be very careful—this is not a fraction! Do not mistake it for "N divided by r ."

The combination formula can be used to replicate the previously presented longhand demonstration wherein we were interested in the number of ways that one release and four detentions can occur. Plugging the numbers into the combination formula yields

$$\binom{5}{1} = \frac{5!}{1!(5-1)!} = \frac{120}{1(4)!} = \frac{120}{1(24)} = 5$$

When we wrote out the possible ways for $r = 1$ to occur, we concluded that there were five options; we have now confirmed this mathematically using the combination formula. There are five different ways for one person to be released and the remaining four to be detained.

We will do one more example calculation before moving on. Suppose that three of the five defendants were released. How many combinations of three are there in a sample of five? Plug the numbers into the combination formula to find out:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{120}{3!(2)!} = \frac{120}{6(2)} = \frac{120}{12} = 10$$

There are 10 combinations of three in a sample of five. In the context of the present example, there are 10 different ways for three defendants to be released and two to be detained. This has to be accounted for in the computation of the probability of each possible result, which is why you see the combination formula included in the binomial coefficient.

The Probability of Success and the Probability of Failure: *p* and *q*

The probability of success (symbolized *p*) is at the heart of the binomial probability distribution. This number is obtained on the basis of prior knowledge or theory. In the present example pertaining to pretrial release, we know that 62% of felony defendants obtain pretrial release. This means that each defendant's probability of release is *p* = .62.

We know that release is not the only possible outcome, though—defendants can be detained, too. The opposite of a success is a **failure**. Because we are dealing with events that have two potential outcomes, we need to know the probability of failure in addition to that of success. The probability of failure is represented by the letter *q* ; to compute the value of *q* , the bound rule and the rule of the complement must be invoked. In , you learned that

p (*A*) + *p* (*Not A*) = 1.00; therefore,

p (*Not A*) = 1.00 − *p* (*A*) .

Failure: Any outcome other than success or the event of interest.

What we are doing now is exactly the same thing, with the small change that *p* (*A*) is being changed to simply *p* , and *p* (*Not A*) will now be represented by the letter *q*. So,

$$q = 1.00 - p \qquad\qquad Formula\ 6(4)$$

With *q* = .62, the probability that a felony defendant will be detained prior to trial (i.e., will *not* obtain pretrial release) is

q = 1.00 − .62 = .38

Putting It All Together: Using the Binomial Coefficient to Construct the Binomial Probability Distribution

Using p and q, the probability of various combinations of successes and failures can be computed. When there are r successes over N trials, then there are $N - r$ failures over that same set of trials. There is a formula called the **restricted multiplication rule for independent events** that guarantees that the probability of r successes and $N - r$ failures is the product of p and q. In the present example, there are three successes (each with probability $p = .62$) and two failures (each with probability $q = .38$). You might also recall from prior math classes that exponents can be used as shorthand for multiplication when a particular number is multiplied by itself many times. Finally, we also have to account for the combination of N and r. The probability of three successes is thus

$$p(3) = \binom{5}{3} p(success) \ p(success) \ p(success) \ p(failure) \ p(failure)$$

$$= \binom{5}{3} p \ p \ p \ q \ q$$

$$= \binom{5}{3} p^3 q^2$$

$$= \binom{5}{3}(.62^3)(.38^2)$$

$$= 10(.24)(.14)$$

$$= .34$$

> **Restricted multiplication rule for independent events**: A rule of multiplication that allows the probability that two events will both occur to be calculated as the product of each event's probability of occurrence: That is, $p(A \text{ and } B) = p(A) \cdot p(B)$.

This result means that given a population probability of .62, there is a .34 probability that if we pulled a sample of five felony defendants, three would obtain pretrial release.

To create the binomial probability distribution, repeat this procedure for all possible values of r. The most straightforward way to do this is to construct a table like Table 6.3. Every row in the table uses a different r value, whereas the values of N, p, and q are fixed. The rightmost column, $p(r)$, is obtained by multiplying the

$$\binom{N}{r}$$

, p^r, and q^{N-r} terms.

The values in the $p(r)$ column tell you the probability of each possible outcome being the one that actually occurs in any given random sample of five defendants. The probability that two of the five defendants will be released pending trial is .19, and the probability that all five will be released is .09. Probabilities can also be added together. The probability of four *or more* defendants being released is $p\,(4) + p\,(5) = .29 + .09 = .38$. The chance that two *or fewer* would be released is $p\,(2) + p\,(1) + p\,(0) = .19 + .06 + .01 = .26$.

Our original question was, "With a population probability of .62, what is the probability that three of five defendants would be released?" You can see from the table that the answer is .34. This, as it happens, is the highest probability in the table, meaning that it is the outcome that we would most expect to see in any given random sample. Another way to think of it is to imagine that someone who was conducting this study asked you to predict—based on a population probability of .62 and $N = 5$ defendants—how many defendants would be released. Your "best guess" answer would be three, because this is the outcome with the highest likelihood of occurring.

Table 6.3	Binomial Probability Distribution for Pretrial Release Among Five Felony Defendants			
r	$\binom{N}{r}$	p^r	q^{N-r}	p(r)
0	$\binom{5}{0} = 1$	$.62^0 = 1.00$	$.38^{5-0=5} = .01$	.01
1	$\binom{5}{1} = 5$	$.62^1 = .62$	$.38^{5-1=4} = .02$	.06
2	$\binom{5}{2} = 10$	$.62^2 = .38$	$.38^{5-2=3} = .05$	.19
3	$\binom{5}{3} = 10$	$.62^3 = .24$	$.38^{5-3=2} = .14$	.34
4	$\binom{5}{4} = 5$	$.62^4 = .15$	$.38^{5-4=1} = .38$	.29
5	$\binom{5}{5} = 1$	$.62^5 = .09$	$.38^{5-5=0} = 1.00$	.09

Contrast this to the conclusion we would draw if none of the five defendants in the sample had been released. This is an extremely improbable event with only $p\,(0) = .01$ likelihood of occurring. It would be rather surprising to find this empirical result, and it might lead us to wonder if there was something unusual about our sample. We might investigate the possibility that the county we drew the sample from had particularly strict policies regulating pretrial release or that we happened to draw a sample of defendants charged with especially serious crimes.

There are two neat and important characteristics of the binomial probability distribution. The first is that it can be graphed using a bar chart (Figure 6.1) so as to form a visual display of the numbers in Table 6.3. The horizontal axis contains the r values and the bar height (vertical axis) is determined by $p(r)$. The bar chart makes it easy to determine at a glance which outcomes are most and least likely to occur.

The second important thing about the binomial probability distribution is that the $p(r)$ column sums to 1.00. This is because the binomial distribution is exhaustive; that is, all possible values of r are included in it. Any time an exhaustive list of probabilities is summed, the result will be 1.00. In Table 6.3, we covered all of the values that r can assume (i.e., zero to five); therefore, all possible probabilities are included, and the sum is 1.00. Memorize this point! It is applicable in the context of continuous probability distributions, too, and will be revisited shortly.

Figure 6.1 The Binominal Probability Distribution for Pretrial Release in a Sample of Five Defendants, With $p = .62$
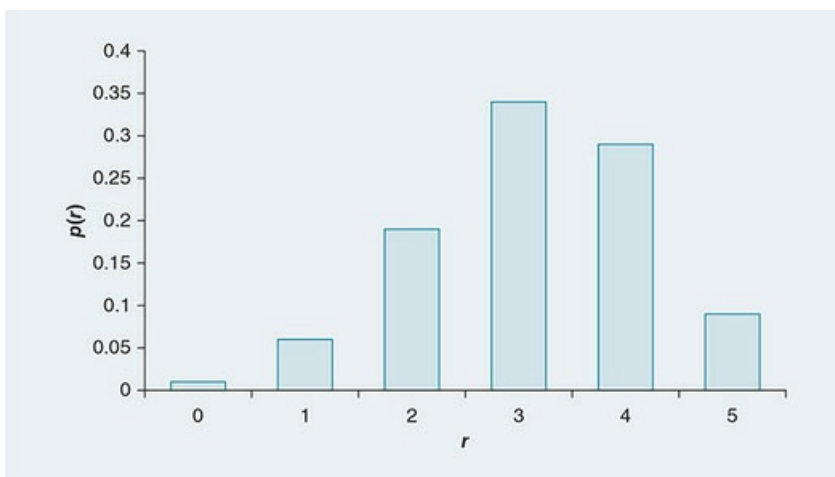
Learning Check 6.7

✓

Recall from Table 6.2 and the discussion that 61.50% of homicides are cleared by arrest. Use this clearance rate to construct the binomial probability distribution for a hypothetical sample of six homicide cases.

Continuous Probability: The Standard Normal Curve

The binomial probability distribution is applicable for trials with two potential outcomes; in other words, it is used for dichotomous categorical variables. Criminal justice and criminology researchers, however, often use continuous variables. The binomial probability distribution has no applicability in this context. Continuous variables are represented by a theoretical distribution called the **normal curve**. Figure 6.2 shows this curve's familiar bell shape.

> **Normal curve**: A distribution of raw scores from a sample or population that is symmetric and unimodal and has an area of 1.00. Normal curves are expressed in raw units and differ from one another in metrics, means, and standard deviations.

Figure 6.2 The Normal Curve for Continuous Variables
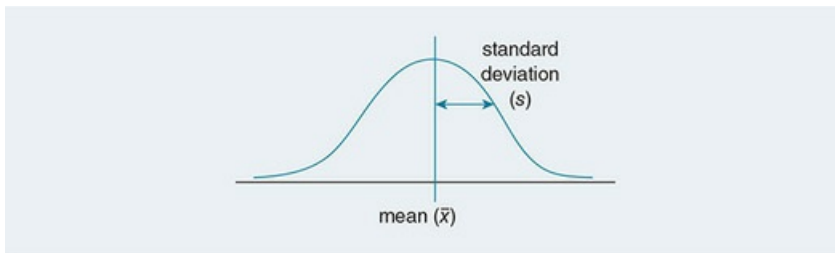
Learning Check 6.8

✓

We have already visited the concept of a normal distribution and have encountered four ways that a curve can deviate from normality. Can you remember those four deviation shapes? Revisit Chapters 4 and 5, if needed.

The normal curve is a *unimodal*, *symmetric* curve with an area of 1.00. It is unimodal because it peaks once and only once. In other words, it has one modal value. It is symmetric because the two halves (split by the mean) are identical to one another. Additionally, it has an area of 1.00 because it encompasses all possible values of the variable in question. Just as the binomial probability distribution's *p* (r) column always sums to 1.00 because all values that *r* could possibly take on are contained within the table, so too the normal curve's tails stretch out to negative and positive infinity. This might sound impossible, but remember that this is a *theoretical* distribution. This curve is built on probabilities, not on actual data.

Research Example 6.2 What Predicts Correctional Officers' Job Stress and Job Satisfaction?

Correctional officers work in high-stress environments, and their job performance has implications for the quality and safety of the correctional institution. High stress levels can reduce these workers' performance levels and can increase their likelihood of physical injury; likewise, low job satisfaction can lead to constant staff turnover and to burnout, which also dampens the overall effectiveness of the correctional institution.

Paoline, Lambert, and Hogan (2006) sought to uncover the predictors of correctional officers' attitudes toward their jobs. They gathered a sample of jail security staff and administered surveys that asked these respondents several questions about the levels of stress they experience at work and the amount of satisfaction they derive from their job. There were six stress variables and five satisfaction variables. Each set of variables was summed to form a single, overarching score on each index for each respondent. The indexes were continuous. Jail staff members with higher scores on the stress index experienced greater work anxiety and tension; likewise, those with lower scores on the satisfaction index had relatively poor feelings about their job.

Paoline et al. (2006) found that the most consistent predictors of both stress and satisfaction were organizational factors specific to the jail itself; in particular, officers who felt that that the jail policies were clear and fair and who had positive views toward their coworkers experienced significantly less stress and greater satisfaction as compared to those officers who were not happy about the policies and their coworkers. These findings suggest that jail managers who seek to foster a positive work environment for their employees should ensure clear, fair policies and should promote harmony and teamwork among jail staff.

The characteristics that determine a normal curve's location on the number line and its shape are its mean and standard deviation, respectively. When expressed in raw units, normal curves are scattered about the number line and take on a variety of shapes. This is a product of variation in metrics, means, and standard deviations. Figure 6.3 depicts this variation.

This inconsistency in locations and dimensions of normal curves can pose a problem in statistics. It is impossible to determine the probability of a certain empirical result when every curve differs from the rest. What is needed is a way to standardiz e the normal curve so that all variables can be represented by a single curve. This widely applicable single curve is constructed by converting all of a distribution's raw scores to **z scores.** The *z* -score transformation is straightforward:

$$z_x = \frac{x - \bar{x}}{s},$$

where

z_x = the z score for a given raw score x,

x = a given raw score,

$\bar{x}$ = the distribution mean, and

s = the distribution standard deviation.

> **z score**: A standardized version of a raw score that offers two pieces of information about the raw score: (1) how close it is to the distribution mean and (2) whether it is greater than or less than the mean.

Figure 6.3 Variation in Normal Curves: Different Means and Standard Deviations
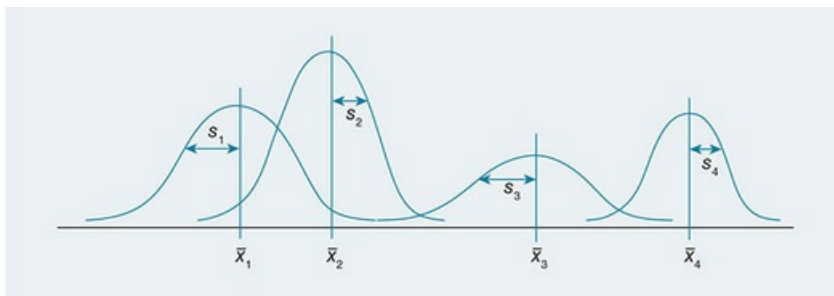
A z score conveys two pieces of information about the raw score on which the z score is based. First, the absolute value of the z score reveals the location of the raw score in relation to the distribution mean. Z scores are expressed in standard deviation units. A z score of .25, for example, tells you that the underlying raw score is one-fourth of one standard deviation away from the mean. A z score of –1.50, likewise, signifies a raw score that is one-and-one-half standard deviations away from the mean.

The second piece of information is given by the sign of the z score. Although standard deviations are always positive, z scores can be negative. A z score's sign indicates whether the raw score that the z score represents is greater than the mean (producing a positive z score) or is less than the mean (producing a negative z score). A z score of .25 is above the mean, while a score of –1.50 is below it. Figure 6.4 shows the relationship between raw scores and their z-score counterparts. You can see that every raw score has a corresponding z score. The z score tells you the distance between the mean and an individual raw score, as well as whether that raw score is greater than or less than the mean.

Figure 6.4 Raw Scores and z Scores
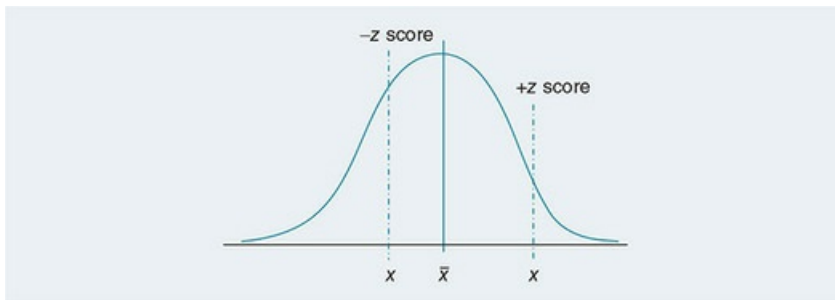
When all of the raw scores in a distribution have been transformed to z scores and plotted, the result is the **standard normal curve**. The z-score transformation dispenses with the original, raw values of a variable and replaces them with numbers representing their position *relative to* the distribution mean. A normal curve, then, is a curve constructed of raw scores, while the standard normal curve is composed of z scores.

Standard normal curve: A distribution of z scores. The curve is symmetric and unimodal and has a mean of zero, a standard deviation of 1.00, and an area of 1.00.

Like ordinary normal curves, the standard normal curve is symmetric, unimodal, and has an area of 1.00. Unlike regular normal curves, though, the standard normal curve's mean and standard deviation are fixed at 0 and 1, respectively. They remain constant regardless of the units in which the variable is measured or the original distribution's mean and standard deviation. This allows probabilities to be computed.

To understand the process of using the standard normal curve to find probabilities, it is necessary to comprehend that, in this curve, area is the same as proportion and probability. A given area of the curve (such as the area between two raw scores) represents the proportion of scores that are between those two raw values. Figures 6.5 and 6.6 display the relationship between z scores and areas.

In Chapter 5, you learned that approximately two-thirds of the scores in any normal distribution lie between

one standard deviation below and one standard deviation above the mean for that set of scores. (Refer back to Figure 5.4.) In Figure 6.5, you can see that .3413 (or 34.13%) of the scores are located between the mean and one standard deviation. If you add the proportion of cases (i.e., area under the curve) that is one standard deviation below the mean to the proportion or area that is one standard deviation above, you get .3413 + .3413 = .6826, or 68.26%. This is just over two-thirds! The bulk of scores in a normal distribution cluster fairly close to the center, and those scores that are within one standard deviation of the mean (i.e., z scores that have absolute values of 1.00 or less) are considered the typical or *normal* scores. This confirms what we saw in Chapter 5 when we talked about the "normal" range being between 1 *sd* above and 1 *sd* below the mean.

Z scores that are greater than 1.00 or less than −1.00 are relatively rare, and they get increasingly rare as you trace the number line away from zero in either direction. These very large z scores do happen, but they are improbable, and some of them are incredibly unlikely. In Figure 6.6, you can see that a full 95% of scores (or .9544, to be exact) are within two standard deviations above and below the mean. In other words, only about 5% of scores in a normal distribution will be either greater than 2 *sd* above the mean (i.e., have a z value greater than 2.00) or more than 2 *sd* below the mean (a z value less than −2.00).

This is all getting very abstract, so let us get an example going. According to the Census of Jails (COJ; see Data Sources 3.1), medium-sized jails have a mean of 23.54 full-time correctional officers. The standard deviation is 17.66.

Suppose that a facility has 19 officers. To find this jail's z score, plug the numbers into Formula 6(5):

$$z_{19} = \frac{19 - 23.54}{17.66} = \frac{-4.54}{17.66} = -.26$$

This facility is approximately one-fourth of one standard deviation below the group mean on staffing. Substantively speaking, this facility is inside the typical zone of ±1 *sd* .

Figure 6.5 Standard Normal Curve: Area Between the Mean and One Standard Deviation
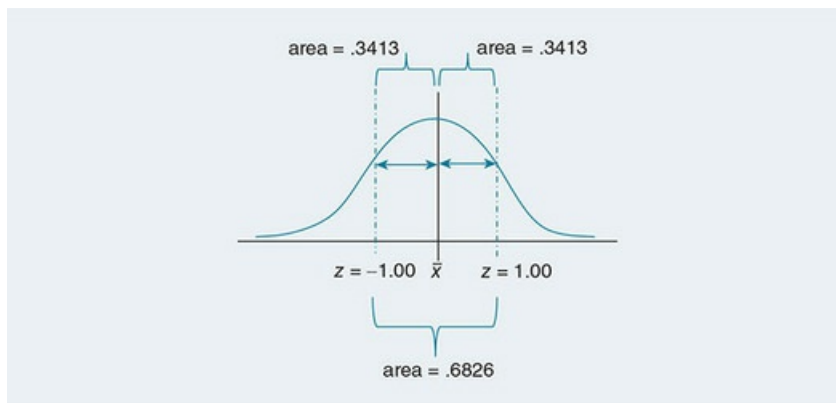
Figure 6.6 Standard Normal Curve: Area Between the Mean and Two Standard Deviations
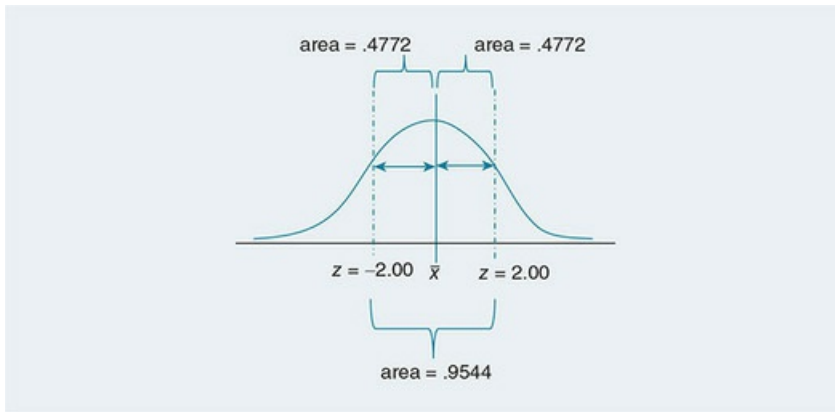
Now for another one. One institution employed 55 correctional officers, so its z score is

$$z_{55} = \frac{55 - 23.54}{17.66} = \frac{31.46}{17.66} = 1.78$$

This institution is roughly one-and-three-fourths standard deviations above the mean, well outside the typical zone.

One more example. Suppose a facility has 80 officers. Its z score is thus

$$z_{80} = \frac{80 - 23.54}{17.66} = \frac{56.46}{17.66} = 3.20$$

This jail's raw score is three-and-one-fifths standard deviations greater than the sample mean, which is very far out in the right-hand tail of the distribution.

Let's draw a hypothetical plot that offers a visualization of the three raw and z scores used in these examples and each one's relationship to the mean. You can see in Figure 6.7 each pair of x's and z's and how close to or far from the mean they are, as well as whether they are above or below it. The larger the z score is, the farther away from the mean that score is.

The z Table and Area Under the Standard Normal Curve

It is also possible to use z scores to find the area (i.e., proportion of values) between the mean and a particular score, or even the area that is *beyond* that score (i.e., the area that is in the tail of the distribution).

Figure 6.7 Raw and z Scores in Relation to the Sample Mean
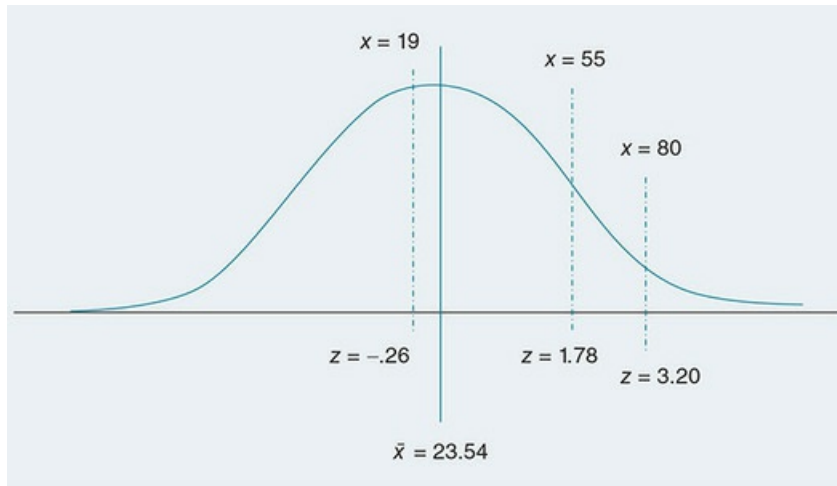
To do this, the **z table** is used. The z table is a chart containing the area of the curve that is between the mean and a given z score. Appendix B contains the z table. The area associated with a particular z score is found by decomposing the score into an *x.x* and *.0x* format such that the first half of the decomposed score contains the digit and the number in the 10ths position, and the second half contains a zero in the 10ths place and the number that is in the 100ths position. In the three previous examples, we calculated z scores of −.26, 1.78, and 3.20. The first z score would be broken down as −.2 + .06. Note that there are no negative numbers in the z table, but this does not matter because the standard normal curve is symmetric, and, therefore, the z table is used the same way irrespective of an individual score's sign. Go to the z table and locate the .2 row and .06 column; then trace them to their intersection. The area is .1026. (Area cannot be negative, so the area between the mean and any given z score is positive even if the score itself is negative.) This means that approximately .10 or 10% of the scores in the distribution sit between the mean and a raw score of 19.

> **z table**: A table containing a list of z scores and the area of the curve that is between the distribution mean and each individual z score.

The z score of 1.78 would decompose as 1.7 + .08. Locating the 1.7 row and .08 column, we can see that the area is .4625. This tells us that roughly .46 or 46% of scores lie between the mean and a raw score of 55. Lastly, to find the area for a z score of 3.20, we look at the 3.2 row and .00 column. Here we hit a bump because the z table does not have a row for 3.2; it has only 3.0 and 3.5. We will select 3.0 because 3.2 is closer to that number than it is to 3.5. Using 3.0 and .00 yields an area of .4987. In other words, approximately .50 or 50% of scores in the distribution sit between the mean and a raw score of 80. This makes sense because there is a very large gap between the mean (23.54) and 80, so lots of raw scores will lie within this range. Figure 6.8 shows the areas between the mean and each of our three z scores. The area between the mean and z = 3.20 is displayed with a bracket because this area encompasses the area between the mean and z = 1.78.

Figure 6.8 Areas Between the Mean and z
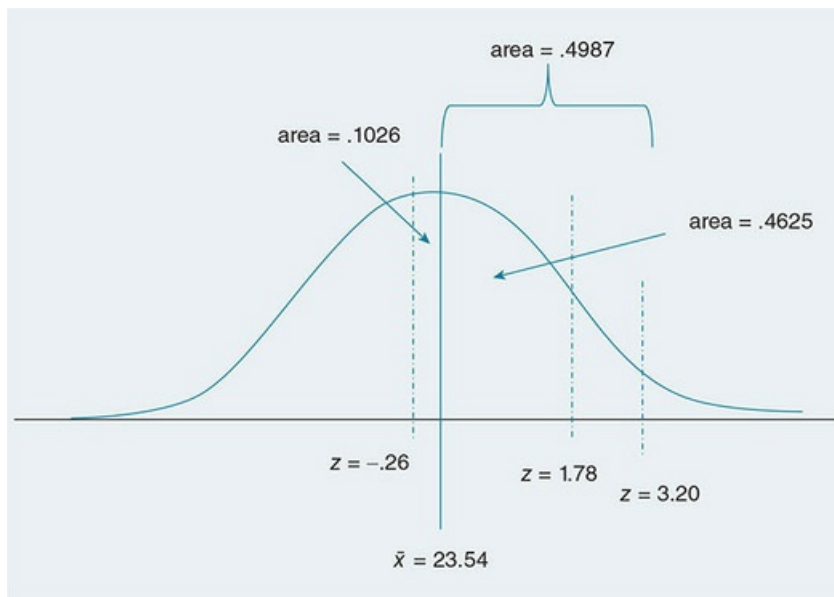
Finding the area between the mean and a z score is informative, but what is generally of interest to criminology and criminal justice researchers is the area *beyond* that z score; that is, researchers usually want to know how big or small a z score would have to be in order to fall into the very tip of either the positive or negative tail of the distribution. To do this, we rely on a simple fact: Because the entire area of the standard normal curve is 1.00, the mean splits the curve exactly in half so that 50% of scores are below it and 50% are above it. As such, we have two pieces of information. First, we know that the total area on each side is .50. Second, for any individual z score, we know the area between the mean and that score. We are looking for the third, unknown number, which is the area beyond that z score. How do you think we can find that number? If you said, "Subtraction," you are right! We subtract the known area from .50 to figure out what's left over.

Let us find the area beyond z for $z = -.26$. The area between the mean and z is .1026 and the total area on the left-hand side is .50, so we use subtraction, as such:

.50 − .1026 = .3974

Thus, .3974 (or approximately 40%, using rounding) of the raw scores in the standard normal distribution for staff in medium-sized jails are less than 19 (alternatively, less than $z = -.26$).

How about for a staff size of 55 officers ($z = 1.78$)? We found that the area between the mean and z is .4625, so

.50 − .4625 = .0375

Approximately .04 (or 4%) of raw scores are greater than 55. This is a small proportion of scores because 55 is so far away from the mean (nearly 2 standard deviations), which leaves very little room for scores exceeding

that number. Figure 6.9 depicts each of these calculated z scores' location on the standard normal curve.

Figure 6.9 Areas Between the Mean and z, Beyond z, and Total Area for Two z Scores
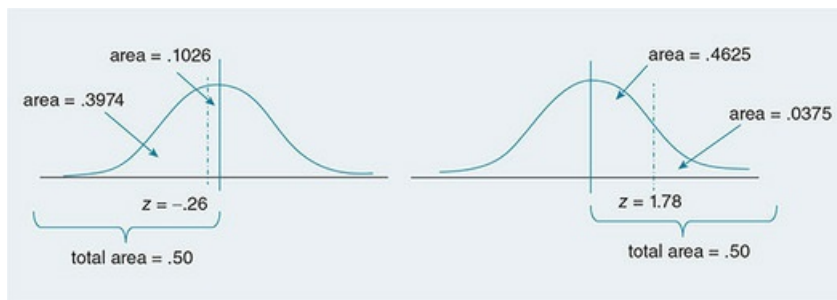
Another handy feature of the standard normal curve is that just as z scores can be used to find areas, areas can likewise be used to find z scores. Basically, you just use the z table backward. We might, for instance, want to know the z scores that lie in the upper 5% (.05) or 1% (.01) of the distribution. These are very unlikely values and are of interest because of their relative rarity. How large would a z score have to be such that only .05 or .01 of scores is above it?

The process of finding these z scores employs similar logic about the area of each side of the curve. First, we know that the z table provides the area between z and the mean; it does not tell you about the area beyond z, so it cannot be used yet. This problem is surmounted using subtraction, as we have already seen. Let us start with the area of .01. The area between the mean and the z score that cuts the distribution at .01 is

$$.50 - .01 = .49$$

Thus, the area between the mean and the z score we are searching for is .49. Since we know this, we can now use the z table.

To do this, scan the areas listed in the body of the z table to find the one that is *closest to* .49. The closest area is .4901. Third, find the z score associated with the identified area. Instead of tracing the two elements of a z score inward to locate the area, as you did before, now start at the area and trace outward along the row and column. The area of .4901 is in the *2.3* row and *.03* column, so the z score is

$$2.3 + .03 = 2.33$$

And that is the answer! We now know that .01 (or 1%) of scores in the standard normal distribution area have z scores greater than 2.33.

Learning Check 6.11

✓

The text uses the example of finding the z scores that are in the top 1% of the standard normal distribution. What z scores are, likewise, in the bottom 1%? Explain how you arrived at your answer. (Hint: You do not need to do any math to answer this question; the calculations have already been done in the text.)

Now let us find the z score associated with an area of .05, but with a twist: We will place this area in the lower (i.e., left-hand) tail of the distribution. Recall that we were working in the upper (right-hand) tail when we did the previous example using an area of .01. The first and second steps of the process are the same no matter which side of the distribution is being analyzed. Subtraction shows that .50 – .05 = .45 of the scores is between the mean and z. Going to the table, you can see that there are actually two areas that are closest to .45. They are .4495 and .4505. The z score for the former is 1.64 and for the latter is 1.65; *however*, since we are on the left (or negative) side of the curve, these z scores are actually –1.64 and –1.65. Be aware of the sign of your z score! Scores on the left side of the standard normal curve are negative, so you have to add a negative sign to the score once you locate it using the table.

Finally, since there are two z scores in this instance, they must be averaged:

$$z_{area=.05} = \frac{-1.64 + (-1.65)}{2} = \frac{-3.29}{2} = -1.645$$

That is the answer! In the standard normal distribution, z scores less than –1.65 have a .05 or less probability of occurring. In other words, these extremely small scores happen 5% of the time or less. That is unlikely indeed.

Learning Check 6.12

✓

There are two points you should be very clear on before leaving this chapter. First, although z scores can be negative, areas are always positive. When you are working on the left side of the standard normal z curve, scores will be negative but areas will not. Second, areas/probabilities and z scores are two ways of expressing the same idea. Large z scores are associated with small probabilities, and small z scores represent large probabilities. The larger the absolute value of a z score is, the smaller the likelihood of observing that score will be. Z scores near z ero (i.e., near the mean of the standard normal curve) are not unlikely or unusual, whereas scores that are very far away from z ero are relatively rare; they are out in the far left or far right tail. Take a moment now to graph the four z values listed in Learning Check 6.10.

Chapter Summary

Probability is the basis of inferential statistics. This chapter introduced two of the major probability distributions: binomial and standard normal. The binomial distribution is for categorical variables that have two potential outcomes (dichotomous or binary variables), and the standard normal curve applies to continuous variables. The binomial probability distribution is constructed using an underlying probability derived from research or theory. This distribution shows the probability associated with each possible outcome, given an overarching probability of success and a predetermined number of cases or trials.

The standard normal curve consists of z scores, which are scores associated with known areas or probabilities. Raw scores can be transformed to z scores using a simple conversion formula that employs the raw score, the mean, and the standard deviation. Because the area under the standard normal curve is a constant 1.00 (i.e., .50 on each side), areas can be added and subtracted, thus allowing probabilities to be determined on the basis of z scores and vice versa.

Both distributions are theoretical, which means that they are constructed on the basis of logic and mathematical theory. They can be contrasted to empirical distributions, which are distributions made from actual, observed raw scores in a sample or population. Empirical distributions are tangible; they can be manipulated and analyzed. Theoretical distributions exist only in the abstract.

Thinking Critically

1. Suppose that in a particular state, female offenders sentenced to prison received a mean sentence length of 30 months ($s = 12$) and male offenders' mean sentence length was 40 months ($s = 10$). If you randomly selected a female prison inmate and found that she had been sentenced to 24 months, would you be surprised? What about if she was serving a 45-month sentence? If you randomly selected a male inmate who had a 60-month sentence, would you be surprised to find this result? What about if his sentence was 31 months? Explain all your answers.

2. Earlier in the chapter, it was shown that in a 52-card deck, any given card has a $\dfrac{1}{52} = .02$ chance of being selected in a random draw. How would this probability change after a card was drawn and not replaced? Calculate the successive changes in probability as one, two, three, four, and five cards are taken from the deck. Does the probability for each remaining card increase or decrease? Why?

Review Problems

1. Eight police officers are being randomly assigned to two-person teams.
   1. Identify the value of N.
   2. Identify the value of r.
   3. How many combinations of r are possible in this scenario? Do the combination by hand first, and then check your answer using the combination function on your calculator.
2. Correctional staff are randomly assigning nine jail inmates to three-person cells.

1. Identify the value of *N*.
   2. Identify the value of *r*.
   3. How many combinations of *r* are possible in this scenario? Do the combination by hand first, and then check your answer using the combination function on your calculator.

3. In a sample of seven parolees, three are rearrested within 2 years of release.
   1. Identify the value of *N*.
   2. Identify the value of *r*.
   3. How many combinations of *r* are possible in this scenario? Do the combination by hand first, and then check your answer using the combination function on your calculator.

4. Out of six persons recently convicted of felonies, five are sentenced to prison.
   1. Identify the value of *N*.
   2. Identify the value of *r*.
   3. How many combinations of *r* are possible in this scenario? Do the combination by hand first, and then check your answer using the combination function on your calculator.

5. Four judges in a sample of eight do not believe that the law provides them with sufficient sanction options when sentencing persons convicted of crimes.
   1. Identify the value of *N*.
   2. Identify the value of *r*.
   3. How many combinations of *r* are possible in this scenario? Do the combination by hand first, and then check your answer using the combination function on your calculator.

6. For each of the following variables, identify the distribution—either binomial or standard normal—that would be the appropriate theoretical probability distribution to represent that variable. Remember that this is based on the variable's level of measurement.
   1. Defendants' completion of a drug court program, measured as *success* or *failure*
   2. The total lifetime number of times someone has been arrested
   3. Crime victims' reporting of their victimization to police, measured as *reported* or *did not report*

7. For each of the following variables, identify the distribution—either binomial or standard normal—that would be the appropriate theoretical probability distribution to represent that variable. Remember that this is based on the variable's level of measurement.
   1. The number of months of probation received by juveniles adjudicated guilty on delinquency charges
   2. City crime rates
   3. Prosecutorial charging decisions, measured as *filed charges* or *did not file charges*

8. According to the UCR, 56% of aggravated assaults reported to police are cleared by arrest. Convert this percentage to a proportion and use it as your value of *p* to do the following:
   1. Compute the binomial probability distribution for a random sample of five aggravated assaults, with *r* defined as the number of assaults that are cleared.
   2. Based on the distribution, what is the outcome (i.e., number of successes) you would *most* expect to see?
   3. Based on the distribution, what is the outcome (i.e., number of successes) you would *least* expect to see?
   4. What is the probability that two or fewer aggravated assaults would be cleared?
   5. What is the probability that three or more would be cleared?

9. According to the BJS, of all criminal charges filed against defendants for domestic violence, 62% are for aggravated assault. Convert this percentage to a proportion and use it as your value of *p* to do the following:
   1. Compute the binomial probability distribution for a random sample of four domestic-violence cases, with *r* defined as the number of cases charged as aggravated assault.
   2. Based on the distribution, what is the outcome (i.e., number of successes) you would *most* expect to see?
   3. Based on the distribution, what is the outcome (i.e., number of successes) you would *least* expect to see?
   4. What is the probability that two or fewer of the charges would be for aggravated assault?
   5. What is the probability that three or more of them would be for assault?

10. According to the UCR, 49% of the hate crimes that were reported to police are racially motivated. Convert this percentage to a proportion and use it as your value of *p* to do the following:
    1. Compute the binomial probability distribution for a random sample of six hate crimes, with *r* defined as the number that are racially motivated.

2. Based on the distribution, what is the outcome (i.e., number of successes) you would *most* expect to see?

3. Based on the distribution, what is the outcome (i.e., number of successes) you would *least* expect to see?

4. What is the probability that two or fewer of the hate crimes were motivated by race?

5. What is the probability that three or more were racially motivated?

11. According to the UCR, 61% of murders are committed with firearms. Convert this percentage to a proportion and use it as your value of p to do the following:

    1. Compute the binomial probability distribution for a random sample of five murders, with r defined as the number that are committed with firearms.

    2. Based on the distribution, what is the outcome (i.e., number of successes) you would *most* expect to see?

    3. Based on the distribution, what is the outcome (i.e., number of successes) you would *least* expect to see?

    4. What is the probability that one or fewer murders were committed with firearms?

    5. What is the probability that four or more murders were committed with firearms?

    The Law Enforcement Management and Administrative Statistics (LEMAS) survey reports that the mean number of municipal police per 1,000 local residents in U.S. cities with populations of 100,000 or more was 1.99 (s = .84). Use this information to answer questions 12 through 15.

12. One department had 2.46 police per 1,000 residents.

    1. Convert this raw score to a z score.
    2. Find the area between the mean and z.
    3. Find the area in the tail of the distribution beyond z.

13. One department had 4.28 police per 1,000 residents.

    1. Convert this raw score to a z score.
    2. Find the area between the mean and z.
    3. Find the area in the tail of the distribution beyond z.

14. One department had 1.51 police per 1,000 residents.

    1. Convert this raw score to a z score.
    2. Find the area between the mean and z
    3. Find the area in the tail of the distribution beyond z.

15. One department had 1.29 officers per 1,000 residents.

    1. Convert this raw score to a z score.
    2. Find the area between the mean and z.
    3. Find the area in the tail of the distribution beyond z.

16. What z scores fall into the upper .15 of the distribution?

17. What z scores fall into the upper .03 of the distribution?

18. What z scores fall into the lower .02 of the distribution?

19. What z scores fall into the lower .10 of the distribution?

20. What z scores fall into the lower .015 of the distribution?

Key Terms

Inferential statistics 134

Probability theory 134

Probability 135

Theoretical prediction 138

Empirical outcome 138

Probability distribution 140

Trial 140

Binomial 140

Binomial probability distribution 140

Binomial coefficient 141

Success 142

Combination 144

Factorial 144

Failure 144

Restricted multiplication rule for independent events 145

Normal curve 148

z score 150

Standard normal curve 151

z table 154

Glossary of Symbols and Abbreviations Introduced in This Chapter

p	The probability of success
N	The total number of trials; the sample size
r	The number of successes over a set of N trials
$\binom{N}{r}$	The combination of "N choose r"; the number of ways for r successes to happen over N trials
q	The probability of failure
z	A score expressed in standard deviation units

Chapter 7 Population, Sample, and Sampling Distributions

A population is the entire universe of objects, people, places, or other units of analysis that a researcher wishes to study. Criminal justice and criminology researchers use all manner of populations. Bouffard (2010), for instance, examined the relationship between men's military service during the Vietnam era and their criminal offending later in life. Rosenfeld and Fornango (2014) assessed how effective order maintenance policing is at reducing robbery and burglary at the neighborhood level. Morris and Worrall (2010) investigated whether prison architectural design influences inmate misconduct. These are three examples of populations—male Vietnam veterans, neighborhoods, prison inmates—that can form the basis for study.

The problem is that populations are usually far too large for researchers to examine directly. There are millions of male military veterans in the United States, thousands of communities nationwide, and approximately 1.5 million prison inmates. Nobody can possibly study any of these populations in its entirety. Samples are thus drawn from populations of interest. Samples are subsets of populations. Morris and Worrall (2010), for example, drew a random sample of 2,500 inmates. This sample, unlike its overarching population, was of manageable size and could be analyzed directly.

Populations and samples give rise to three types of distributions: population, sample, and sampling. A **population distribution** contains all values in the entire population, while a **sample distribution** shows the shape and form of the values in a sample pulled from a population. Population and sample distributions are both empirical. They are made of raw scores derived from actual people or objects. **Sampling distributions**, by contrast, are theoretical arrays of sample statistics. Each of these is discussed in turn in this chapter.

Population distribution: An empirical distribution made of raw scores from a population.

Sample distribution: An empirical distribution made of raw scores from a sample.

Sampling distribution: A theoretical distribution made out of infinite sample statistics.

Empirical Distributions: Population and Sample Distributions

Population and sample distributions are both empirical because they exist in reality; every person or object in the population or sample has a value on a given variable that can be measured and plotted on a graph. To illustrate these two types of distributions, we can use the 2006 Census of Jails (COJ; see Data Sources 3.1). This data set contains information on all jails that house inmates for extended periods either awaiting trial or serving sentences (*N* = 2,949). No sampling was done; every jail in the United States was asked to provide information. This makes the COJ a population data set.

Figure 7.1 is a histogram made in SPSS showing the population distribution for the variable *total number of inmates* , which is a measure of the number of people housed in each facility. Every facility's inmate count was plotted to form this histogram. You should be able to recognize immediately that this distribution has an extreme positive skew.

What might the distribution look like for a sample pulled from this population? The SPSS program can be commanded to select a random sample of cases from a data set, so this function will be used to simulate sampling. A random sample of 50 facilities was pulled from the COJ and the variable *total number of inmates* plotted to form Figure 7.2. The sample distribution looks somewhat similar to the population distribution in that they are both positively skewed; however, you can see that there are clear differences between them. This is because there are only 50 cases in this sample, and 50 is a very small subset of 2,949.

Let's try increasing the sample size to 100. Telling SPSS to randomly select 100 jails and graph their sizes yields the histogram in Figure 7.3. The shape is closer to that of the population, since 100 is a better representation of the full 2,949 than 50 is. Of course, differences linger.

Figure 7.1 Population Distribution for Total Inmates per Correctional Facility (*N* = 2,371)

Mean = 269.63
Std. Dev. = 802.129
N = 2,371
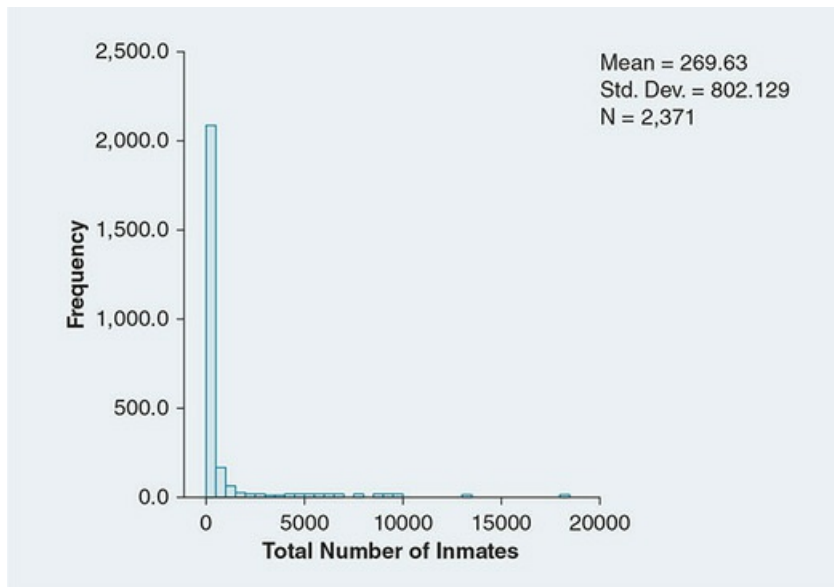
Figure 7.2 Sample Distribution (*N* = 50)
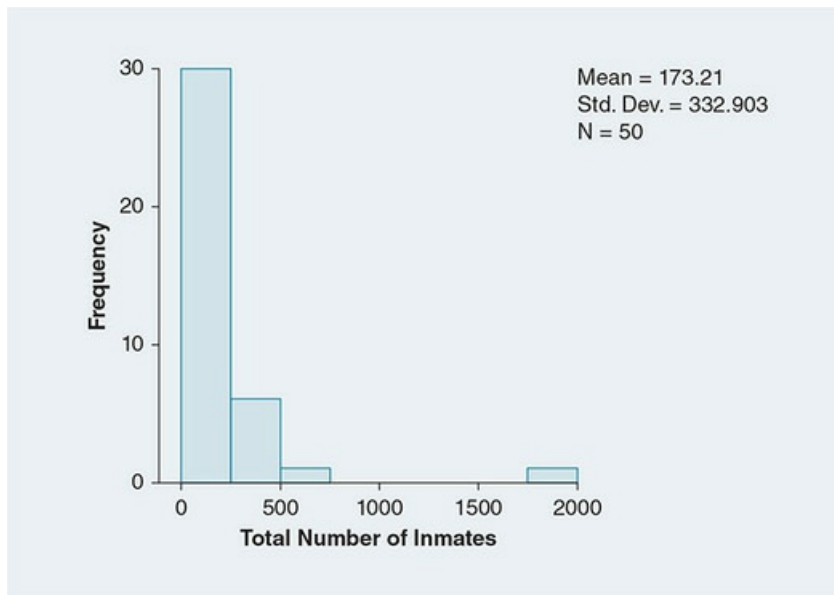
We can try the sampling exercise again and this time pull 500 cases randomly. [Figure 7.3](#) shows the *total inmates* histogram for this third sample. This histogram is a close match to that for the entire population, which is a function of the larger sample size. We have thus demonstrated a fact that will be critical to the understanding of empirical distributions: Among samples drawn at random from populations, larger samples are more-accurate reflections of those populations. The technical term for this is **representativeness**. A representative sample matches the population on key characteristics. A random sample of adults from a population that is 50% female should also be roughly 50% female if it is representative. A sample that was 12% or 70% female would not be representative. Sample size is not the only determinant of how representative a sample will be of its population, but it is an important factor.

> **Representativeness**: How closely the characteristics of a sample match those of the population from which the sample was drawn.

Before we continue, we should stop and consider some new terms and their definitions. First, although we have encountered the word *statistic* several times up to this point, it is necessary now to offer a formal definition of this concept. A **statistic** is a number that describes a sample. This might be a mean, proportion, or standard deviation. The second term is **parameter**. A parameter is just like a statistic except that it describes a population. Populations, like samples, have means, proportions, standard deviations, and so on. Statistics are estimates of parameters. [Table 7.1](#) displays the symbols for some common statistics and their corresponding parameters. The statistic notations for the mean and the standard deviation are familiar from previous chapters, but this is the first time we have considered the population symbols. They are Greek letters. The population mean is the lowercase version of the letter *mu* (μ ; pronounced "mew"), and the standard deviation is a lowercase *sigma* (σ) . The population proportion is a less exciting uppercase *P*. We have previously represented sample proportions with a lowercase *p* , but now that we are differentiating between samples and populations, we are going to change the letter to $\hat{p}$, which is pronounced "p hat."

> **Statistic**: A number that describes a sample that has been drawn from a larger population.
>
> **Parameter**: A number that describes a population from which samples might be drawn.

Figure 7.3 Sample Distribution (N = 100)

Mean = 258.43
Std. Dev. = 522.021
N = 100
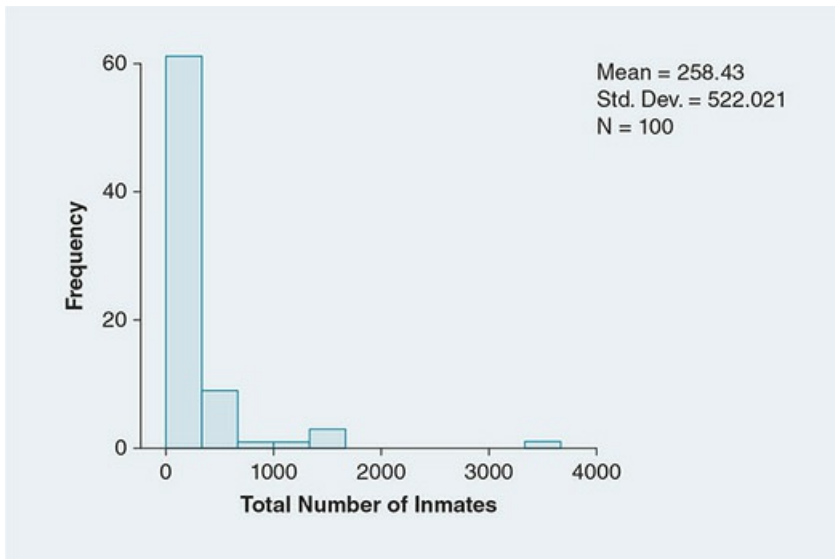
Figure 7.4 Sample Distribution (N = 500)

Mean = 301.95
Std. Dev. = 825.434
N = 500
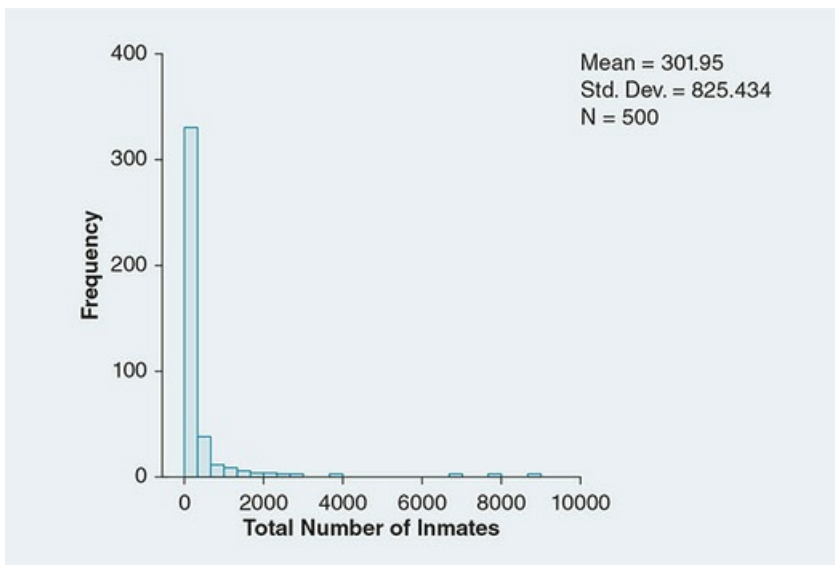
Table 7.1	Symbols for Common Statistics and Their Corresponding Parameters	
	Sample Statistic	*Population Parameter*
Mean	$\bar{x}$	μ
Standard deviation	s	σ
Proportion	$\hat{p}$	P

Often, criminal justice and criminology researchers want to make a statement about a population, but what they actually have in front of them to work with is a sample (because samples are typically more feasible to collect and work with than populations are). This creates a conundrum because sample statistics cannot simply be generalized to population parameters. Statistics are estimates of population parameters, and, moreover, they are estimates that contain error (as demonstrated by Figures 7.1 through 7.4). Population parameters are fixed. This means that they have only one mean and standard deviation on any given measure (e.g., people's ages), one percentage on other measures (e.g., the percentage that is female), and so on. Sample statistics, by contrast, vary from sample to sample because of **sampling error**. Sampling error arises from the fact that multiple (theoretically, infinite) random samples can be drawn from any population. Any given sample that a researcher *actually* draws is only one of a multitude that he or she *could* have drawn. Figure 7.5 depicts this. Every potential sample has its own distribution and set of descriptive statistics. In any given sample, these statistics might be exactly equal to, roughly equal to, or completely different from their corresponding parameters.

Sampling error: The uncertainty introduced into a sample statistic by the fact that any given sample is one of many samples that could have been drawn from that population.

Figure 7.5 Multiple Random Samples Can Be Drawn From a Population
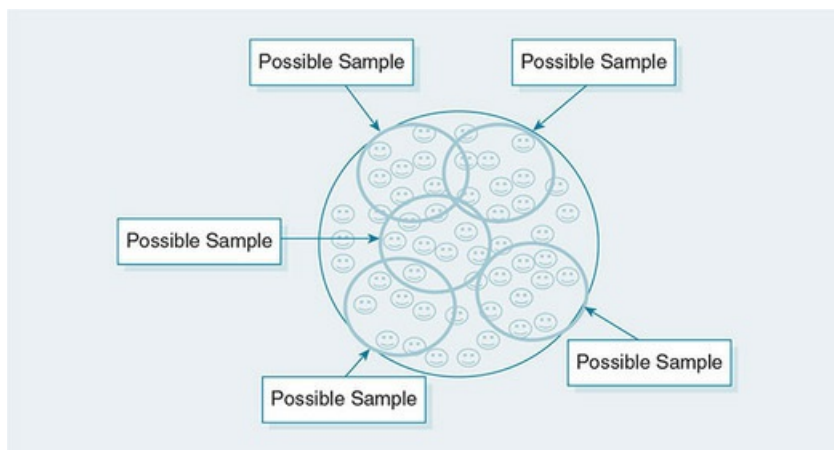
The COJ can be used to illustrate the effects of sampling error. As described earlier, this data set is a

population; therefore, samples can be drawn from it. We will continue using the variable *total number of inmates housed*. The population means is μ = 269.63, and the standard deviation is σ = 802.13. Drawing five random samples of 500 facilities each and computing the mean and the standard deviation of each sample produces the data in Table 7.2.

Table 7.2	Means and Standard Deviations for the Total Inmates in the Population (N = 2,371) and Five Random Samples of N = 500					
	Population	*Sample 1*	*Sample 2*	*Sample 3*	*Sample 4*	*Sample 5*
Mean	269.63	282.80	250.34	200.57	237.07	236.73
Std. Dev.	802.13	910.40	702.14	372.22	652.02	693.40

Figure 7.6 The Relationship Between Samples, Sampling Error, and Populations
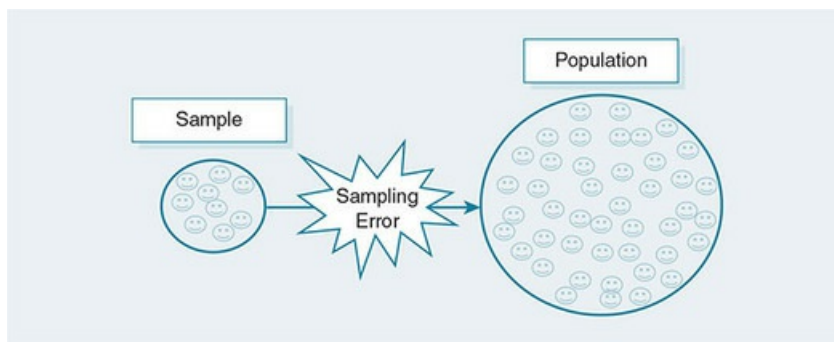
Look how the means and standard deviations vary—this is sampling error! Sample 1's mean is 282.80, while Sample 3's is a very different 200.57. There is a substantial amount of variation among the samples, and each one differs from the true population mean by a smaller or larger amount. In this example, we have the benefit of knowing the true population mean and the standard deviation, but that is usually not the case in criminal justice and criminology research. What researchers generally have is one sample and no direct information about the population as a whole. Imagine, for instance, that you drew Sample 1 in Table 7.2. This sample's mean and standard deviation are reasonable approximations of—though clearly not equivalent to—their corresponding population values, but you would not know that. Now picture Sample 3 being the one you pulled for a particular study. This mean and standard deviation are markedly discrepant from the population parameters, but, again, you would be unaware of that.

This example illustrates the chasm between samples and populations that is created by sampling error and prevents inferences from being made directly. It would be a mistake to draw a sample, compute its mean, and automatically assume that the population mean must be equal or close to the sample mean. As displayed pictorially in Figure 7.5, sampling error prevents direct inference from a sample to the larger population from which it was drawn. What is needed is a bridge between samples and populations so that inferences can be reliably drawn. This bridge is the sampling distribution.

Theoretical Distributions: Sampling Distributions

Sampling distributions, unlike population and sample distributions, are theoretical; that is, they do not exist as empirical realities. We have already worked with theoretical distributions in the form of the binomial and standard normal distributions. Sampling distributions are theoretical because they are based on the notion of multiple (even infinite) samples being drawn from a single population. What sets sampling distributions apart from empirical distributions is that sampling distributions are created not from raw scores but, rather, from sample statistics. These descriptors can be means, proportions, or any other statistic. Imagine plotting the means in Table 7.2 to form a histogram, like Figure 7.7.

Figure 7.7 Histogram of the Five Sample Means in Table 7.2
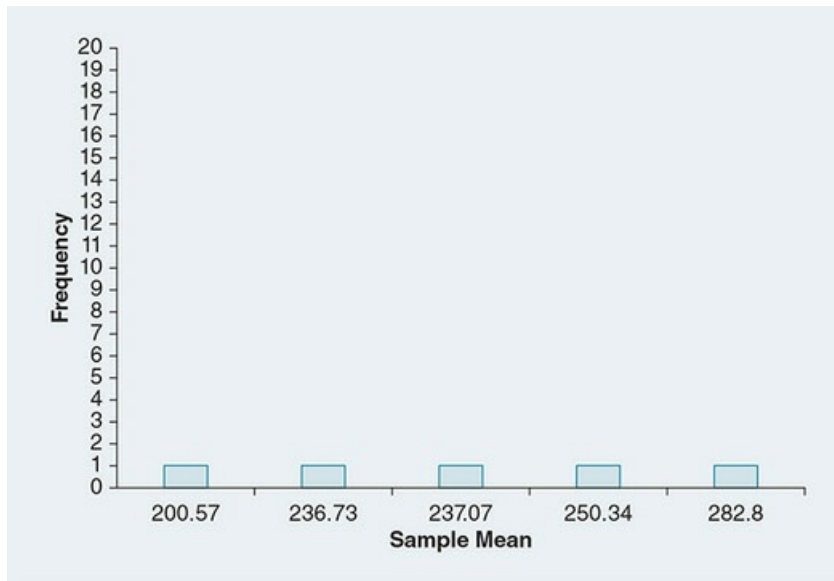
Not terribly impressive, is it? Definitely leaves something to be desired. That is because there are only five samples. Sampling distributions start to take shape only when many samples have been drawn. If we continue the iterative process of drawing a sample, computing and plotting mean, throwing the sample back, and pulling a new sample, the distribution in Figure 7.7 gradually starts looking something like the curve in Figure 7.8.

Now the distribution has some shape! It looks much better. It is, moreover, not just any old shape—it is a normal curve. What you have just seen is the **central limit theorem** (CLT) in action. The CLT states that any time descriptive statistics are computed from an infinite number of large samples, the resulting sampling distribution will be normally distributed. The sampling distribution clusters around the true population mean (here, 269.63), and if you were to compute the mean of the sampling distribution (i.e., the mean of means), the answer you obtained would match the true population mean. Its standard deviation (called the **standard error**) is smaller than the population standard deviation because there is less dispersion, or variability, in means than in raw scores. This produces a narrower distribution, particularly as the size of the samples increases. The mean of the sampling distribution is symbolized as $\mu_{\overline{x}}$ and the standard error is represented as $\sigma_{\overline{x}}$.

The CLT is integral to statistics because of its guarantee that the sampling distribution will be normal when a sample is large. Criminal justice and criminology researchers work with many variables that show signs of skew or kurtosis. The CLT saves the day by ensuring that even skewed or kurtotic variables will produce normal sampling distributions. The inmate count variable demonstrates this. Compare Figure 7.8 to Figure 7.4: Figure 7.4 is derived from a single sample (*N* = 500) and is highly skewed, yet the sampling distribution in Figure 7.8 is normal. That is because even when raw values produce skew, sample statistics will still hover around the population parameter and fall symmetrically on each side of it. Some statistics will be greater than the parameter and some will be smaller, but the majority will be close approximations of (or even precisely equal to) the true population value.

Figure 7.8 Distribution of Sample Means
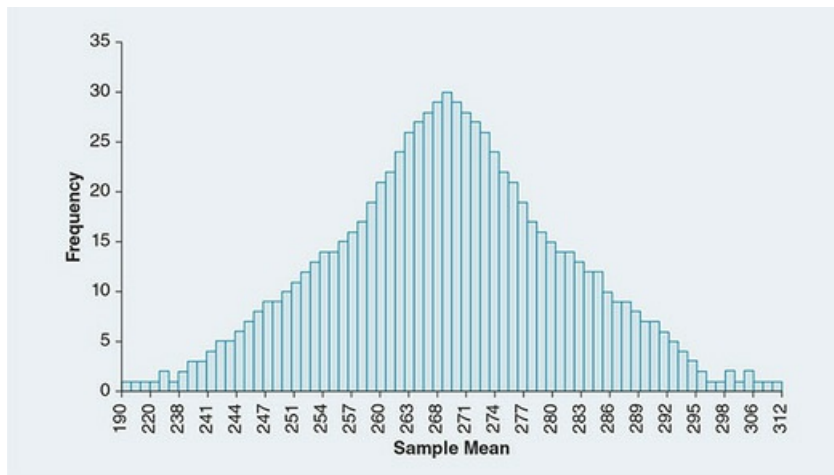
All descriptive statistics have sampling distributions to which the CLT applies. In Chapter 6, you learned that nationwide, .62 (or 62%) of felony defendants are granted pretrial release. Figure 7.9 sketches what the sampling distribution of proportions for the pretrial release variable might look like. You can see that this curve is roughly normal in shape, too, just like the sampling distributions of means in Figure 7.8.

Figure 7.9 Sampling Distribution of Proportions
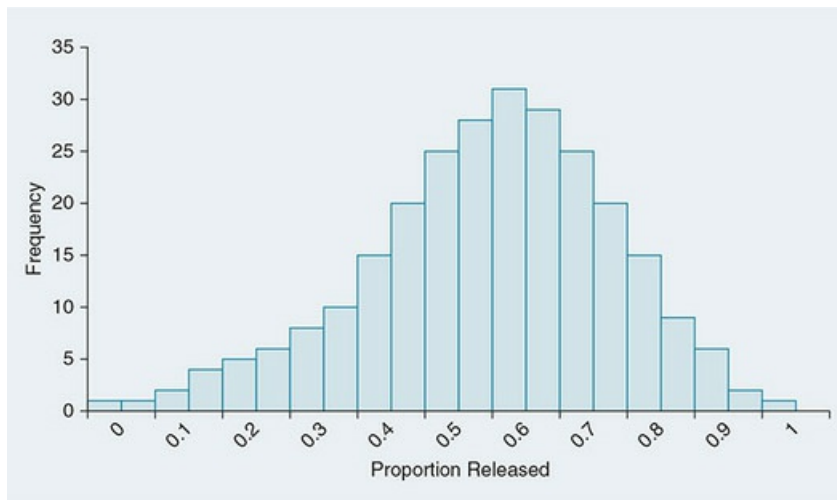
Sample Size and the Sampling Distribution: The z and t Distributions

The key benefit of the sampling distribution being normally distributed is that the standard normal curve can be used. Everything we did in Chapter 6 with respect to using raw scores to find z scores, z scores to find areas, and areas to find z scores can be done with the sampling distribution. The applicability of z , though, is contingent on N being large. "Large" is a vague adjective in statistics because there is no formal rule specifying the dividing line between small and large samples. Generally speaking, large samples are those containing at least 100 cases. When $N \geq 100$, the sampling distribution is assumed to be normally distributed, and the standard normal curve can be used. This requirement reduces the overall usefulness of the z distribution; it turns out that the standard normal curve makes somewhat rigid demands, and many real-world data sets fall short of these expectations.

Although it is generally not advisable to work with samples smaller than 100, there are times when it is unavoidable. In these situations, the z distribution cannot be employed and researchers turn to the ***t*** **distribution** instead. The t distribution—like the z curve—is symmetric, unimodal, and has a constant area of 1.00. The key difference between the two is that t is a family of several different curves rather than one fixed, single curve like z. The t distribution changes shape depending on the size of the sample. When the sample is small, the curve is wide and flattish; as the sample size increases, the t curve becomes more and more normal until it looks identical to the z curve. See Figure 7.10.

> ***t* distribution**: A family of curves whose shapes are determined by the size of the sample. All t curves are unimodal and symmetric and have an area of 1.00.

Figure 7.10 The Family of t Curves
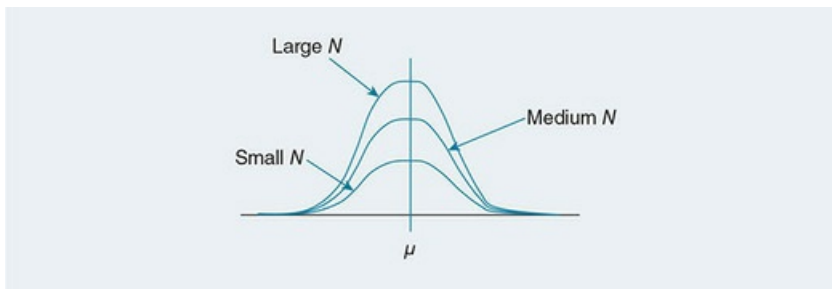
This phenomenon can also be demonstrated using hypothetical examples of random samples from the *Census of State and Federal Adult Correctional Facilities* (CSFACF) data set. Figures 7.11 and 7.12 demonstrate how the t curve would change shape depending on the size of the samples being randomly selected from this population. The curve in Figure 7.11 is based on a sample size of 75, and that in Figure 7.12 is premised on N = 25. The first curve is taller and thinner. It is normal, but contrasting it with Figure 7.8, you can see that Figure 7.11 does not cluster around the population mean as tightly. Scores are more spread out. This trend is even more pronounced in Figure 7.12. This is because there is more variability in smaller samples—it is difficult to get an accurate estimate of the true population parameter when there are only a handful of cases. A sample of N = 75 is not ideal, but it is an improvement over 25.

Figure 7.11 The *t* Distribution for Inmate Population at *N* = 75
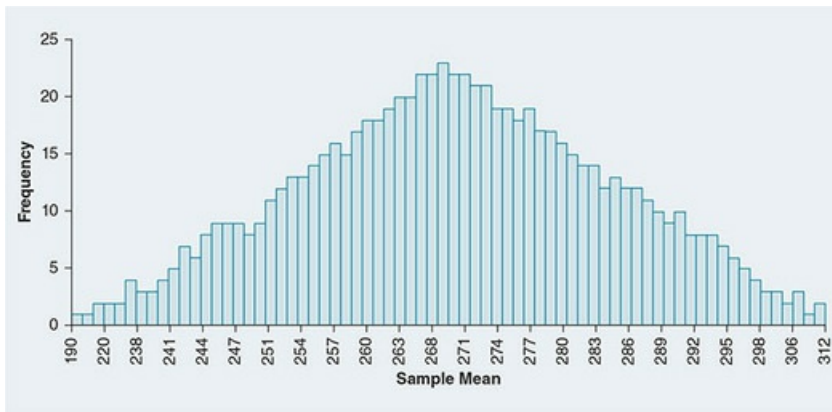
Figure 7.12 The *t* Distribution for Inmate Population at *N* = 25
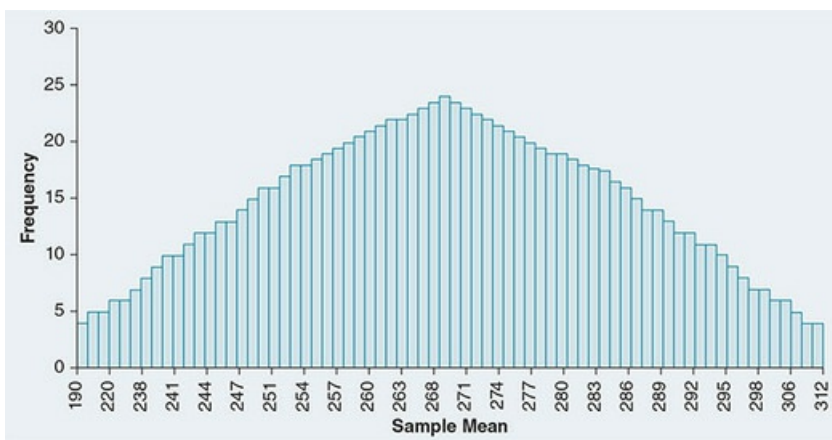
The *t* distribution's flexibility allows it to accommodate samples of various sizes. It is an important theoretical probability distribution because it allows researchers to do much more than they would be able to if *z* were their only option. All else being equal, large samples are better than small ones, and it is always advisable to work with large samples when possible. When a small sample must be used, though, *t* is a trustworthy alternative. We will use both the *z* and *t* distributions in later chapters.

Chapter Summary

Criminal justice and criminology researchers often seek information about populations. Populations, though, are usually too large to analyze directly. Samples are therefore pulled from them and statistical analyses are applied to these samples instead. Population and sample distributions are made of raw scores that have been plotted. These are empirical distributions.

Sampling error, though, introduces an element of uncertainty into sample statistics. Any given sample that is drawn is only one of a multitude of samples that could have been drawn. Because of sampling error, statistics are merely estimates of their corresponding population parameters and cannot be interpreted as matching them exactly. In this way, there is a gap between samples and populations.

The sampling distribution links samples to populations. Sampling distributions are theoretical curves made out of infinite sample statistics. All descriptive statistics have sampling distributions. The CLT ensures that sampling distributions are normal when sample sizes are large (i.e., $N \geq 100$). When this is the case, the *z* distribution can be used. When samples are small, though, the sampling distribution cannot be assumed to be normal. The *t* distribution solves this problem because *t* is a family of curves that change shape depending on sample size. The *t* distribution is more flexible than *z* and must be used any time $N \leq 99$, though it can be used with large samples as well.

1. Suppose you read a research article in which the study authors stated that they had collected a sample of 150 adolescents from a single city and found that 22 of them reported being in a gang. What is the population being studied here? What additional pieces of information would you need to know about this sample in order to decide whether to trust this finding as being truly reflective of that population? What characteristics would the sample need to have in order for it to be trustworthy?

2. In your own words, explain (1) why sample size is important to ensuring that samples are representative of the populations from which they are derived and (2) why the sampling distribution for any given statistic is normally distributed even when the population and sample distributions are skewed.

Review Problems

1. Population distributions are . ..
   1. empirical.
   2. theoretical.
2. Sample distributions are . ..
   1. empirical.
   2. theoretical.
3. Sampling distributions are . ..
   1. empirical.
   2. theoretical.
4. The _____ distribution is made from the raw scores in a sample.
5. The _____ distribution is made from statistics calculated on multiple or infinite samples.
6. The _____ distribution is made from the raw scores in a population.
7. The CLT guarantees that as long as certain conditions are met, a sampling distribution will be . ..
   1. positively skewed.
   2. normally distributed.
   3. negatively skewed.
8. For the CLT's promise of distribution shape to hold true, samples must be . ..
   1. large.
   2. small.
9. When a sample contains 100 or more cases, the correct probability distribution to use is the _____ distribution.
10. When a sample contains 99 or fewer cases, the correct probability distribution to use is the _____ distribution.
11. A researcher gathers a sample of 200 people, asks each one how many times he or she has been arrested, and then plots each person's response. From the list below, select the type of distribution that this researcher has created.
    1. A sample distribution with a large N
    2. A population distribution with a small N
    3. A sampling distribution with a large N
    4. A sample distribution with a small N
12. A researcher gathers a sample of 49 police departments, finds out how many officers were fired for misconduct in each department over a 2-year time span, and plots each department's score. From the list below, select the type of distribution that this researcher has created.
    1. A population distribution with a small N
    2. A population distribution with a large N
    3. A sampling distribution with a small N
    4. A sample distribution with a small N
13. A researcher gathers a sample of 20 cities, calculates each city's mean homicide rate, and plots that mean. Then the researcher puts that sample back into the population and draws a new sample of 20 cities and computes and plots the mean homicide rate. The researcher does this repeatedly. From the list below, select the type of distribution that this researcher has created.
    1. A sampling distribution with a large N

2. A population distribution with a small N

3. A sampling distribution with a small N

4. A population distribution with a large N

14. A researcher has data on each of the nearly 2,000 adult correctional facilities in the United States and uses them to plot the number of inmate-on-inmate assaults that took place inside each prison in a 1-year span. From the list below, select the type of distribution that this researcher has created.

   1. A sample distribution with a large N

   2. A population distribution with a small N

   3. A sampling distribution with a large N

   4. A population distribution with a large N

15. A researcher gathers a sample of 132 people and computes the mean number of times the people in that sample have shoplifted. The researcher then puts this sample back into the population and draws a new sample of 132 people, for whom the researcher computes the mean number of times shoplifted. The researcher does this repeatedly. From the list below, select the type of distribution that this researcher has created.

   1. A sample distribution with a large N

   2. A sampling distribution with a large N

   3. A sample distribution with a small N

   4. A population distribution with a large N

Key Terms

Glossary of Symbols and Abbreviations Introduced in This Chapter

μ	The population mean (mu)
σ	The population standard deviation (sigma)
P	The population proportion
$\hat{p}$	The sample proportion (p hat)
$\mu_{\bar{X}}$	The mean of the sampling distribution
$\sigma_{\bar{X}}$	The standard error
t	A distribution that is a family of curves that can accommodate both small and large sample sizes

Chapter 8 Point Estimates and Confidence Intervals

Learning Objectives

- Explain the problems with point estimates and the need for confidence intervals.
- Explain the trade-off between confidence and precision.
- Identify the key elements of the t curve, and use sample size to select the correct distribution (z or t).
- Use a specified confidence level to find alpha and the critical value of z or t.
- Identify the correct distribution and formula for a given sample statistic and sample size.
- Calculate and interpret confidence intervals for means and proportions with different confidence levels and sample sizes.

Any given sample that is drawn from a population is only one of a multitude of samples that *could have* been drawn. Every sample that is drawn (and every sample that could potentially be drawn) has its own descriptive statistics, such as a mean or proportion. This phenomenon, as you learned in Chapter 7, is called sampling error. The inherent variation in sample statistics such as means and proportions prevents direct inference from a sample to a population. It cannot be assumed that a mean or proportion in a sample is an exact match to the mean or proportion in the population because sometimes sample statistics are very similar to their corresponding population parameters and other times they are quite dissimilar. Flip back to Table 7.2 for an illustration of this concept. Basically, there is always an element of uncertainty in a sample statistic, or what can also be called a **point estimate**.

> **Point estimate**: A sample statistic, such as a mean or proportion.

Fortunately, though, a procedure exists for calculating a range of values within which the parameter of interest is predicted to be. This range stretches out on each side of the point estimate and is called a **confidence interval** (*CI*). The confidence interval acts as a sort of "bubble" that introduces flexibility into the estimate. It is much more likely that an estimate of the value of a population parameter is accurate when the estimate is a range of values rather than one single value.

> **Confidence interval**: A range of values spanning a point estimate that is calculated so as to have a certain probability of containing the population parameter.

Try thinking about it this way: Suppose I guessed that you are originally from Chicago. This is a very precise prediction! Of all the cities and towns in the world, I narrowed my guess down to a single area. Given its precision, though, this prediction is very likely to be wrong; there are more than 7.4 billion people in the world, and only about 2.7 million of them live in Chicago. My point estimate (Chicago) is probably incorrect.

But what if I instead guessed that you are from the state of Illinois? There are 12.9 million people in Illinois, so I have increased my chances of being correct because I have broadened the scope of my estimate. If I went up another step and predicted that you are from the Midwest—without specifying a city or state—I have further increased my probability of being right, since this is a much larger geographic region and contains more than 66 million residents. It is still possible that I am wrong, of course, but I am far more likely to guess your place of origin correctly when I guess a large geographical area, such as a region, than when I guess a much smaller one, such as a city.

This is, conceptually, what a confidence interval does. The interval serves as a buffer zone that allows for greater confidence in the accuracy of a prediction. It also allows us to determine the probability that the prediction we are making is correct. Using distributions—specifically, the z and t probability curves—we can figure out how likely it is that our confidence interval truly does contain the true population parameter. The probability that the interval contains the parameter is called the **level of confidence**.

Level of confidence: The probability that a confidence interval contains the population parameter. Commonly set at 95% or 99%.

The Level of Confidence: The Probability of Being Correct

In the construction of confidence intervals, you get to choose your level of confidence (i.e., the probability that your confidence interval accurately estimates the population parameter). This might sound great at first blush —why not just choose 100% confidence and be done with it, right?—but confidence is actually the classic double-edged sword because there is a trade-off between it and precision. Think back to the Chicago/Illinois/Midwest example. The Chicago guess has a very low probability of being correct (we could say that there is a low level of confidence in this prediction), but it has the benefit of being a very precise estimate because it is just one city. The Illinois guess carries an improvement in confidence because it is a bigger geographical territory; however, because it is bigger, it is also less precise. If I guess that you are from Illinois and I am right, I am still left with many unknown pieces of information about you. I would not know which part of the state you are from, whether you hail from a rural farming community or a large urban center, what the socioeconomic characteristics of your place of origin are, and so on.

The problem gets worse if all I guess is that you are from the Midwest—now I would not even know which state you are from, much less which city! If I want to be 100% sure that I will guess your place of origin correctly, I have to put forth "planet Earth" as my prediction. That is a terrible estimate. If you want greater confidence in your estimate, then, you pay the price of reduced precision and, therefore, a diminished amount of useful information.

Confidence levels are expressed in percentages. Although there is no "right" or "wrong" level of confidence in a technical sense (i.e., there is nothing mathematically preventing you from opting for a 55% or 72% confidence level [*CI*]), 95% and 99% have become conventional in criminal justice and criminology research. Because of the trade-off between confidence and precision, a 99% *CI* has a greater chance than a 95% one of being correct, but the 99% one will be wider and less precise. A 95% *CI* will carry a slightly higher likelihood of error but will yield a more informative estimate. The 99% level would be akin to the Midwest guess in the previous example, whereas the 95% would be like the Chicago guess. You should select your level of confidence by deciding whether it is more important that your estimate be correct or that it be precise.

Confidence levels are set a priori (in advance), which means that you must decide whether you are going to use 95% or 99% before you begin constructing the interval. The reason for this is that the level of confidence affects the calculation of the interval. You will see this when we get to the *CI* formula.

Since we are dealing with probabilities, we have to face the unpleasant reality that our prediction could be incorrect. The flipside of the probability of being right (i.e., your confidence level) is the probability of being wrong. Consider these statements:

100% – 95% = 5%

100% – 99% = 1%

Each of these traditional levels of confidence carries a corresponding probability that a confidence interval does *not* contain the true population parameter. If the 95% level is selected, then there is a 5% chance that the *CI* will not contain the parameter; a 99% level of confidence generates a 1% chance of an inaccurate *CI*.

You will, in all likelihood, never know whether the sample you have in front of you is one of the 95% or 99% that is correct, or whether it is one of the 5% or 1% that is not. There is no way to tell; you just have to compute the confidence interval and hope for the best. This is an intractable problem in statistics because of the reliance on probability.

Three types of confidence intervals will be discussed in this chapter: CI_s for means with large samples ($N \geq$ 100), for means with small samples ($N \leq 99$), and for proportions and percentages. All three types of CI_s are meant to improve the accuracy of point estimates by providing a range of values that most likely contains the true population parameter.

Confidence Intervals for Means With Large Samples

When a sample is of large size ($N \geq 100$), the z distribution (the standard normal curve) can be used to construct a *CI* around a sample mean. Confidence intervals for means with large samples are computed as

$$CI = \bar{x} \pm z_\alpha \left(\frac{s}{\sqrt{N-1}} \right), \qquad \text{Formula 8(1)}$$

where

$\bar{X}$ = the sample mean,

z_α = the z score associated with a given alpha level (i.e., the critical value of z) ,

α = the probability of being wrong (the alpha level),

s = the sample standard deviation, and

N = the sample size.

This formula might appear intimidating, but we can read it left to right and break down the different elements. The starting point is the sample mean. A certain value will be added to and subtracted from the mean to form the interval. That value is the end result of the term on the right side of the ± operator. The z score's subscript α (this is the Greek letter *alpha*) represents the probability that the *CI* does not contain the true population parameter. Recall that every level of confidence carries a certain probability of inaccuracy; this probability of inaccuracy is α or, more formally, the **alpha level**. Alpha is computed as

$$\alpha = 1 - \text{confidence level} \qquad \text{Formula 8(2)}$$

> **Alpha level**: The opposite of the confidence level; that is, the probability that a confidence interval does not contain the true population parameter. Symbolized α.

For 95% and 99% confidence, first convert the percentages to proportions:

$\alpha_{95\%} = 1 - .95 = .05$

$\alpha_{99\%} = 1 - .99 = .01$

Alpha, itself, is not inserted into the *CI* formula; rather, α is used to find the **critical value** of z , and you enter that critical value into the formula (this is the z_α term in Formula 8[1]). The critical value of z is the z score associated with a particular area on the curve. In other words, it is the score beyond which a certain area (this being alpha) is out in the tail. We will see later that the t curve has critical values, too.

> **Critical value**: The value of z or t associated with a given alpha level. Symbolized z_α or t_α .

There is another piece of information that you need to know about *CI*s: They are always **two-tailed**. This is because the normal curve has two halves that are split by the mean, with the result being (pessimistically) that

there are two ways to be wrong with a confidence interval. The first option is for the interval to be wholly above the mean and miss it by being too far out in the positive tail, and the second possibility is that it lands entirely below the mean and misses it by being too far out into the negative side. Since either error is possible —and since we have no control over which type of error we might end up with—we must use both sides of the curve. This is a two-tailed test. In such a test, the alpha level is split in half and placed in each of the two tails of the distribution, as pictured in Figure 8.1. Two-tailed tests, then, actually have *two* critical values. The absolute value of these critical values is the same because the curve is symmetric, but one value is negative and the other is positive. Confidence intervals, by definition, use both of the critical values for any alpha level.

Two-tailed test: A statistical test in which alpha is split in half and placed into both tails of the *z* or *t* distribution.

Figure 8.1 The Alpha Level and Critical Values of *z*
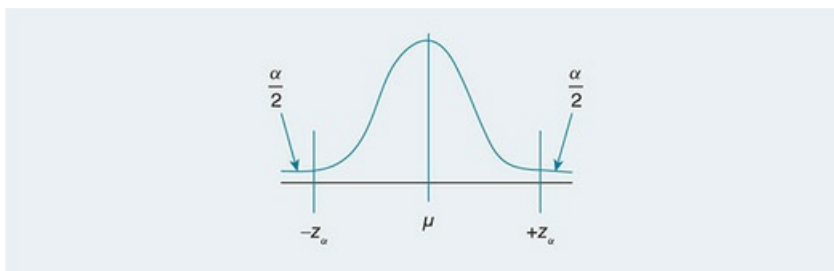
So, what are these critical values? For the standard normal curve, we can find them using the techniques we learned in Chapter 6. Recall that in the *z* -score exercises, we figured out how to find *z* scores using specified areas: Essentially, the *z* table (see Appendix B) is used backward. Let us begin with the 95% confidence interval. First, we have to find α, which is 5%. Second, because this is a two-tailed test, we divide α in half, as shown in Figure 8.1. Doing this shows that the critical value of *z* is going to be the value that leaves 2.5% of cases in each tail of the distribution. Of course, we cannot work with percentages in the *z* table, so we have to convert this to .025. Third, the *z* table is used to find the critical value of *z* . We have done this before, in Chapter 6. What we are asking here is "If the area in the tail is .025, what is *z* ?" The tail must first be subtracted from .50:

.50 − .025 = .4750

Next, go to the *z* table and locate the area that is equal to .4750 or, if there is no area that is exactly .4750, the area that is closest. In this instance, the value we are seeking is actually located in the table. Trace away from .4750 upward along the column and to the left along the row, and record each of the numbers you arrive at. Here,

z = 1.9 + .06 = 1.96

We are not quite done! Remember that there are *two* critical values, not just one. Since the standard normal curve is symmetric and .025 is the area in each tail, these two *z* scores will take on the same absolute value but

will have opposite signs. Thus, $z_{\alpha = .05} = \pm 1.96$. The critical value of z for the top .025 (the positive side of the curve) is 1.96, and the value for the bottom .025 (the negative side) is −1.96. Figure 8.2 displays the two critical values of z for a 95% confidence level.

The same process is used to find z_α for a 99% confidence level. First, alpha is 100% − 99% = 1%. Second, 1% divided in half is .5% and converted to a proportion is .005. Third, .50 − .005 = .4950. This exact value does not appear in the z table, so the z scores associated with the two closest values (.4949 and .4951) must be averaged. These scores are 2.57 and 2.58. Therefore,

$$z_{\sigma = .01} = \frac{2.57 + 2.58}{2} = \frac{5.15}{2} = 2.575$$

Figure 8.2 Critical Values of z for a 95% Level of Confidence
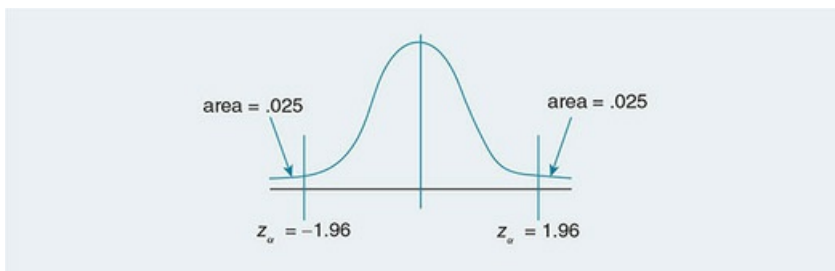
The critical value is 2.58. Again, remember that there are two values, so z is actually ±2.58. This is displayed graphically in Figure 8.3.

Figure 8.3 Critical Values of z for a 99% Level of Confidence
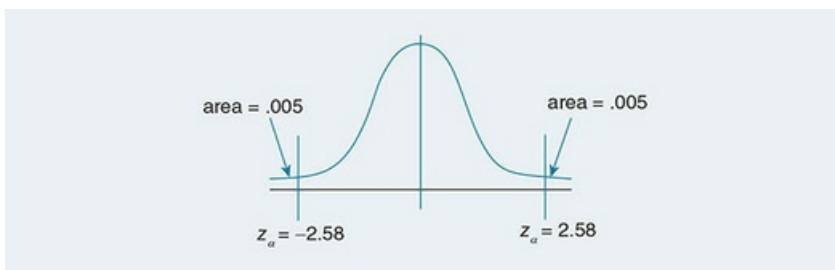
The standard normal curve is handy because it is fixed; therefore, the critical values of z for 95% and 99% confidence intervals will always be ±1.96 and ±2.58, respectively. Later on, you will see that this is not the case when the t distribution is employed; in that situation, the critical value of t will have to be located each time you calculate a confidence interval. For now, though, we can rely on these two critical values.

Now let's consider an example from the 2013 Census of Jails (COJ; see Data Sources 3.1). This data set is ideal for present purposes because it contains population data (i.e., it is a census, not a sample, of jails), so we can compute the actual population mean and then pull a random sample, compute its mean and confidence interval, and see if the computed interval contains the true population value. Remember that researchers are not ordinarily able to access information on the population parameters (and, obviously, if you knew the population mean, there would be no point in attempting to estimate it using a confidence interval). What we are doing here is for demonstration purposes only and is not typical of the reality of most criminal justice and criminology research.

For this example we will use the COJ variable in which each facility reported the number of inmates it currently housed who had been convicted and were awaiting sentencing. The confidence level will be set at 95%.

In the population, the mean number of convicted individuals who have not yet been sentenced is μ = 12.06 (σ = 56.04). Let us use a sample size of N = 183, which is large enough to permit use of the z distribution. The SPSS random-sample generator produces a sample with a mean of $\overline{x} = 18.18$ and a standard deviation of s = 97.52. Since the confidence level is 95%, z_α = ±1.96. Plugging values for N, $\overline{X}$, s, and z_α into Formula 8(1) yields

$$CI = \overline{x} \pm z_\alpha \left(\frac{s}{\sqrt{N-1}} \right) = 18.18 \pm 1.96 \left(\frac{97.52}{\sqrt{183-1}} \right)$$

$$= 18.18 \pm 1.96 \left(\frac{97.52}{13.49} \right)$$

= 18.18 ± 1.96 (7.23)

= 18.18 ± 14.17.

Let's pause for a moment and consider what we have calculated thus far. We have the sample mean (18.18) as the midpoint of the interval, and now we know that the buffer is going to extend 14.17 units above and 14.17 units below the mean. Picture these two extensions of the mean as forming the full width of the interval.

The next step is to compute the lower limit (*LL*) and the upper limit (*UL*) of the interval, which requires the

± operation to be carried out, as such:

LL = 18.18 − 14.17 = 4.01

UL = 18.18 + 14.17 = 32.35

Finally, the full interval can be written out:

95% CI : 4.01 ≤ μ ≤ 32.35

The interpretation of this interval is that there is a 95% chance that the true population mean is 4.01, 32.35, or some number in between. More formally, "There is a 95% chance that the interval 4.01 to 32.35, inclusive, contains the true value of μ." Of course, this also means that there is a 5% chance that the true population mean is *not* in this range. In the present example, we know that μ = 12.06, so we can see that here the confidence interval does indeed contain the population mean (i.e., 12.06 is greater than 4.01 and less than 32.35). This is good! This sample was one of the 95% that produces accurate confidence intervals rather than one of the 5% that does not. Remember, though, that knowing the value of the true population mean is a luxury that researchers generally do not have; ordinarily, there is no way to check the accuracy of a sample-based confidence interval.

Let us repeat the previous example using a 99% confidence level. This set of calculations would proceed as such:

$$CI = \bar{x} \pm z_\sigma \left(\frac{s}{\sqrt{N-1}} \right) = 18.18 \pm 2.58 \left(\frac{97.52}{\sqrt{183-1}} \right)$$
$$= 18.18 \pm 2.58 \left(\frac{97.52}{13.49} \right)$$

= 18.18 ± 2.58 (7.23)

= 18.18 ± 18.65.

And computing the lower and upper limits yields

LL = 18.18 − 18.65 = −.47

UL = 18.18 + 18.65 = 36.83

99% CI : −.47 ≤ μ ≤ 36.83.

There is a 99% chance that the interval –.47 to 36.83, inclusive, contains the population mean μ. (Of course, the negative values are nonsensical because it is impossible for a facility to have fewer than zero unsentenced inmates. From a practical standpoint, the negative values are treated as zero; this *CI* indicates that it is possible the true population mean is zero.) Again, we know that $\mu = 12.06$, so this interval does, indeed, contain the parameter.

Take a look at the difference in width between the 95% and 99% intervals. The 95% interval ranges from 4.01 to 32.35, and the 99% one spans –.47 to 36.83. The 99% interval is much wider than the 95% one. Can you explain the reason for this? If you said that it is because with a 99% confidence level we sacrifice precision in the estimate, you are correct! When we enhanced our confidence, we increased z from 1.96 to 2.58, thus causing the interval to expand. This demonstrates the trade-off between confidence and precision.

For a third example, we will use the Police–Public Contact Survey (PPCS; see Data Sources 2.1). This is not a population: The survey was administered to a random sample of U.S. residents. We must, therefore, use sample statistics to estimate the true population values. Let us consider the ages of respondents who reported having experienced multiple (i.e., more than one) contact with police officers during the past year and construct a 99% *CI* to estimate the mean age for the population. The respondents with multiple contacts ($N = 2{,}304$) had a mean age of 45.23 with a standard deviation of 16.80. The *CI* is

$$CI = 45.23 \pm 2.58\left(\frac{16.80}{\sqrt{2{,}304-1}}\right) = 45.23 \pm 2.58\left(\frac{16.80}{47.99}\right)$$

$= 45.23 \pm 2.58\ (.35)$

$= 45.23 \pm .90$

Now calculate the lower and upper limits:

LL $= 45.23 - .90 = 44.33$

UL $= 45.23 + .90 = 46.13$

Finally, construct the interval:

99% *CI:* $44.33 \le \mu \le 46.13$

We can say with 99% confidence that the interval 44.33 and 46.13, inclusive, contains μ. This is a very precise interval even though we used the 99% level of confidence. The primary reason for this is the sample size. The

larger the sample, the larger the denominator in the $\dfrac{s}{\sqrt{N-1}}$ portion of the *CI* equation will be, thus producing a very small number after the division is complete. Of course, this also depends on the size of the standard deviation (and the quality of the estimate is predicated on the sample being representative of the population), but, generally speaking, larger samples yield narrower confidence intervals compared with smaller samples.

Confidence Intervals for Means With Small Samples

The z distribution can be used to construct confidence intervals when $N \geq 100$ because the sampling distribution of means can be safely assumed to be normal in shape. When $N \leq 99$, though, this assumption breaks down and a different distribution is needed. This alternative distribution is the t curve. The *CI* formula for means with small samples is nearly identical to that for large samples; the only difference is that the critical value of t is used instead of the critical value of z. The formula is

$$CI = \bar{x} \pm t_\alpha \left(\frac{s}{\sqrt{N-1}} \right), \qquad \qquad Formula\ 8(3)$$

where t_α = the critical value of t at a given alpha level.

Research Example 8.1 Do Criminal Trials Retraumatize Victims of Violent Crimes?

Violent crimes frequently leave victims with significant negative psychological and emotional consequences. Some victims develop long-term problems, and many are diagnosed with posttraumatic stress disorder. There is widespread concern about the negative impacts trials can exert on victims already suffering emotional and psychological trauma. During trials, victims have to recount their experiences in open court and undergo cross-examination by defense attorneys. Victims might feel shame, embarrassment, and fear. It is not known with certainty, however, whether trials do indeed have these effects; limited research has been conducted and so it is unclear whether the concerns for victims' wellbeing are empirically founded. Orth and Maercker (2004) addressed this gap in the research. Using a sample of violent-crime victims in Germany, the authors calculated participants' mean levels of posttraumatic stress reactions prior to and then again after their attackers' trials. The researchers subtracted the two means to obtain a measure of change over time. The table summarizes the results.

The results revealed a slight decrease in posttraumatic stress reactions over time, as indicated by the negative difference scores. This decline was very small and was indistinguishable from zero, as indicated by the fact that the 95% *CI*s contained zero (that is, it was possible that the true population differences were zero). These results call into question the common assumption that trials retraumatize victims, since there was no evidence in this study of victims experiencing an increase in emotional distress after trial. More research is needed before firm conclusions can be drawn, but these findings suggest that trials are not significantly painful events for most victims.

Trauma Scale Dimension	Difference between Means	95% Confidence Interval	
		Lower Limit	Upper Limit
Intrusion	−.18	−.51	.16
Avoidance	−.19	−.45	.07
Hyperarousal	−.28	−.64	.08

Source: Adapted from Table 3 in Orth and Maercker (2004).

Source: Adapted from Table 3 in Orth and Maercker (2004).

The critical value of t (i.e., t_α) is found using the t table (see Appendix C). We have not used this table yet, so take a few minutes now to familiarize yourself with it. There are three pieces of information you need in order to locate t_α. The first is the number of tails. As described previously, confidence intervals are always two-tailed, so this is the option that is always used with this type of test. The second determinant of the value of t_α

is the alpha level, the computation of which was shown in Formula 8(2). Finally, finding t_α requires you to first compute the degrees of freedom (df) . With the t distribution, df are related to sample size, as such:

$$df = N - 1 \qquad\qquad\qquad Formula\ 8(4)$$

The df values are located in the rows of the t table. Note that not all of the possible values that df might assume are included on the table. When the number you are looking for is not there, it is customary to use the table df that is closest to the sample-derived df. If the derived df is halfway between two numbers in the table, select the smaller one. For instance, if your sample size were 36, your df would be 36 – 1 = 35. Your options in using the table would be df = 30 and df = 40, and you would elect to use df = 30. The rationale is that using the lower of the two makes for a more conservative estimate (i.e., a wider confidence interval). In statistics, researchers usually err on the side of caution—if you have to choose between inflating your calculated CI and shrinking it, you should opt for inflation because that is in line with custom in criminology and criminal justice research. (The inflation this causes is generally miniscule and has a trivial impact on the CI.)

For an example of CI_s with means and small samples, the Firearm Injury Surveillance Study (FISS; see Data Sources 8.1) will be used. This data set contains information about a sample of patients treated in emergency departments (EDs) for gunshot wounds between the years 1993 and 2013. We will narrow the file down to the most recent year of data available. For the current example, we will analyze the mean age of Hispanic female gunshot victims. There were 40 such victims, and their mean age was 26.18 ($s = 11.55$). We will set confidence at 95%.

Data Sources 8.1 The Firearm Injury Surveillance Study, 1993–2013

The Centers for Disease Control and Prevention (CDC) maintain the National Electronic Injury Surveillance System, of which the FISS is a part. The data are from a nationally representative sample of hospitals stratified by size. Detailed incident data include only those patients who did not die in the ED in which they were treated; those who died prior to arrival are assigned dead on arrival (DOA) status, and those who died after being transferred from the ED to some other hospital unit are coded as transfers (U.S. Department of Health and Human Services, n.d.). This is a limitation of the FISS that circumscribes its usefulness in criminal justice research, but this data set is nonetheless valuable as one of the few major studies that systematically tracks gun-related injuries. Data include patient age, sex, and race, as well as incident characteristics such as whether the shooting was intentional or unintentional and what the relationship was between the shooter and the victim. The most recent version of the FISS was collected in 2013.

The first step we must take to construct the CI is to find t_α. We know that this is a two-tailed test with $\alpha = 1 - .95 = .05$ and $df = 40 - 1 = 39$. You can see that $df = 39$ is not located on the table, so we must use the $df = 40$ row instead. The critical value of t is ± 2.021.

Next, plug all the numbers into Formula 8(2):

$$CI = \bar{x} \pm t_\alpha \left(\frac{s}{\sqrt{N-1}} \right) = 26.18 \pm 2.021 \left(\frac{11.55}{\sqrt{40-1}} \right)$$

$$= 26.18 \pm 2.021 \left(\frac{11.55}{6.24} \right)$$

$= 26.18 \pm 2.021 \ (1.85)$

$= 26.18 \pm 3.74$

Compute the lower and upper limits:

LL = 26.18 − 3.74 = 22.44

UL = 26.18 + 3.74 = 29.92

Finally, assemble the full confidence interval:

95% *CI* : 22.44 ≤ μ ≤ 29.92

There is a 95% chance that the interval 22.44 to 29.92, inclusive, contains the population mean.

Let's try one more example. We will again use the FISS and this time analyze the mean age of males whose firearm injuries occurred in the course of a drug-related crime (the FISS does not contain information about the role gunshot victims played in these transactions). In the 2013 data, there were 124 males shot during drug-related crimes, and their mean age was 30.06 (*s* = 10.08). The confidence level will be 99%.

The three pieces of information needed for finding the critical value of *t* are (1) this is a two-tailed test, (2) alpha is .01 (since 1 − .99 = .01), and (3) the degrees of freedom are 124 − 1 = 123. With these criteria, t_α = ±2.617 (using the *df* = 120 line of the *t* table). Plugging all the numbers into Formula 8(2) yields

$$CI = 30.06 \pm 2.617 \left(\frac{10.08}{\sqrt{124-1}} \right) = 30.06 \pm 2.617 \left(\frac{10.08}{11.09} \right)$$

= 30.06 ± 2.617 (.91)

= 30.06 ± 2.38

The lower and upper limits are

LL = 30.06 − 2.38 = 27.68

UL = 30.06 + 2.38 = 32.44

The interval is

99% *CI:* 27.68 ≤ μ ≤ 32.44

There is a 99% probability that the interval 27.68 to 32.44, inclusive, contains μ .

One more *t* example can be derived from the PPCS. Respondents who had been stopped by police while driving or walking were asked how long that stop lasted. The 208 people who answered the question reported

a mean stop length of 13.30 minutes ($s = 21.19$). We will opt for a 95% confidence interval, and the $df = 208 - 1 = 207$. Looking at the t table, you can see that there is no $df = 207$, so we will have to choose between the $df = 120$ and $df = \infty$ (infinity) values. Sticking to our principle of always opting for the more conservative (i.e., wider) confidence interval, we will select $df = 120$. This makes our critical value ±1.980. Now we can plug in the numbers:

$$CI = 13.30 \pm 1.980\left(\frac{21.19}{\sqrt{208-1}}\right) = 13.30 \pm 1.980\left(\frac{21.19}{14.39}\right)$$

   = 13.30 ± 1.980 (1.47)

   = 13.30 ± 2.91

The lower and upper limits are

   LL = 13.30 − 2.91 = 10.39

   UL = 13.30 + 2.91 = 16.21

The interval is

   99% $CI:$ 10.39 ≤ μ ≤ 16.21

We can be 95% confident in our prediction that the population mean is between 10.39 and 16.21, inclusive.

Learning Check 8.4

The interval calculated in the police stop example spans nearly 6 minutes ($16.21 - 10.39 = 5.82$). Suppose a researcher wants a more precise estimate of μ. Use what you have learned thus far in the chapter to identify two actions that could be taken to shrink the interval and increase precision.

Confidence Intervals With Proportions and Percentages

The principles that guide the construction of confidence intervals around means also apply to confidence intervals around proportions and percentages. There is no difference in the procedure used for proportions versus that for percentages—percentages simply have to be converted to proportions before they are plugged into the *CI* formula. For this reason, we will speak in terms of proportions for the remainder of the discussion.

There are sampling distributions for proportions just as there are for means. Because of sampling error, a sample proportion (symbolized $\hat{p}$, pronounced "*p* hat") cannot be assumed to equal the population proportion (symbolized as an uppercase *P*) , so confidence intervals must be created in order to estimate the population values with a certain level of probability.

> Research Example 8.2 What Factors Influence Repeat Offenders' Completion of a "Driving Under the Influence" Court Program?
>
> Specialized courts are an increasingly popular way for dealing with low-level offenders, especially those who have drug or mental-health problems. The rationale is that these people should not be incarcerated in jail or prison and should instead be allowed to remain in the community and complete one or more treatment programs to help them with their problems. Judges are responsible for supervising the defendants in these courts. Defendants typically appear before the judge every month or two, and the judge praises them when they have done well in their program and reprimands them when they have failed to follow through on an assignment. Usually, charges are dropped when defendants complete the program successfully, and those who drop out or are removed for noncompliance get sentenced to a previously agreed-on penalty (such as a specified jail or probation term).
>
> In recent years, courts dedicated to handling people convicted of driving under the influence (DUI) of drugs or alcohol have appeared as a new incarnation of specialized courts. The success of these courts at reducing recidivism hinges on their ability to keep defendants in the program; defendants who drop out have to be sentenced to jail or probation (which is more expensive) and will not experience the benefits of the treatment regimen. Saum, Hiller, and Nolan (2013) sought to identify the factors associated with treatment completion versus dropout. They gathered records on 141 third-time DUI offenders who went through a DUI court program in Wisconsin. Most of the defendants (114) completed the program, but 27 did not. Overall, the researchers found very few differences between the groups. Having a mental-health problem did not affect the odds of completion, as shown by the wide confidence interval for this variable's predictive capability (.17 to 1.6). The only variable that emerged as significant was the number of days in the jail or work-release sentence; those who had been threatened with more-severe sentences were more likely to drop out. This variable's confidence interval was very small (.95 to 1.00), suggesting that is was a good predictor of program completion versus program dropout. It would appear that more-serious DUI offenders need enhanced supervision and greater incentives to stay in and successfully complete DUI court programs.

How Extensively Do News Media Stories Distort Public Perceptions About Racial Minorities' Criminal Involvement?

Race is inextricably linked with crime in the public's mind. Numerous stereotypes paint blacks and Latinos, in particular, as being responsible for the majority of drug trafficking and violence that takes place in the United States. The media both reflect and fuel these stereotypes, and have a significant influence on the impression the public forms about the "color" of crime. The media do this by choosing which crimes to cover and how to portray the victim and offender in any given violent event. Dixon (2017) sought to examine media (mis-)representation of offender and victim race. Dixon's study added to the existing work on the subject by including Spanish-language television stations, which previous studies had not done. He drew a sample of 117 news programs from the population of programs that aired between 2008 and 2012 in Los Angeles. Each program was coded according to crime type, victim race, perpetrator race, and the race of responding police officers. To figure out how closely news reporting matched real crime characteristics, Dixon compared the news programs to official crime data. The tables show the results, presented as percentages, including the 95% *CI* for each estimate derived from the news programs. The differentials indicate whether people of each race are underrepresented or overrepresented as homicide perpetrators, victims, or police officers.

Race of Homicide Perpetrators

Race	Arrest %	TV Perp. %	95% CI	Differential
Black	27	31	±11	+4
White	10	19	±10	+9
Latino	59	47	±12	−12
Other	3	3	±4	0

Source: Adapted from Table 2 in Dixon (2017).

Race of Homicide Victims

Race	Victim %	TV Victim %	95% CI	Differential
Black	33	38	±12	+5
White	13	35	±12	+22
Latino	50	10	±7	−40
Other	4	18	±10	+14

Source: Adapted from Table 3 in Dixon (2017).

Race of Police Officers

Race	Officer %	TV Officer %	95% CI	Differential
Black	12	9	±4	−3
White	53	73	±6	+20
Latino	30	16	±5	−14
Other	5	5	±3	0

Source: Adapted from Table 4 in Dixon (2017).

Contrary to much previous research, Dixon's (2017) analysis did not find that black perpetrators were significantly overrepresented in news reports; in fact, white perpetrators were overrepresented more often (9 percentage points relative to actual arrest rates) than black perpetrators were (4 points). Interestingly, Latino perpetrators were notably underrepresented (a 12-percentage point differential). A very different pattern manifested in the analyses of victims and police officers. Black victims were slightly

Confidence intervals for proportions employ the z distribution. The normality of the sampling distribution of sample proportions is a bit iffy, but, generally speaking, z is a safe bet as long as the sample is large (i.e., $N \geq$ 100) and contains at least five successes and at least five failures (more on this shortly). The formula for confidence intervals with proportions is

$$CI = \hat{p} \pm z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \qquad \textit{Formula 8(5)}$$

where $\hat{p}$ = the sample proportion.

To illustrate CI_s for proportions with large samples, we will again use the 2013 FISS. This time, we will examine the involvement of handguns in violent altercations. According to the FISS, handguns were the firearm used in 91.30% of gunshots arising from fights in which the type of gun used is known (i.e., excluding the cases in which gun type cannot be determined). The sample size is $N = 173$, which meets the criterion that the sample have at least 100 cases. Now we have to consider whether the requirement that there be at least five successes and at least five failures. Within the context of this problem, we will define "success" as the use of a handgun and "failure" as the use of any other gun type. Handguns were used in 158 instances and other guns in the remaining 15, so this variable meets the criteria regarding sample size being large enough and there being more than five successes and more than five failures.

Confidence will be set at 99%, which means $z_\alpha = \pm 2.58$.

First, the sample percentage needs to be converted to a proportion; dividing 91.30 by 100 and rounding to two decimal places yields .91.

Next, plug the numbers into Formula 8(5) and solve:

$$CI = \hat{p} \pm z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} = .91 \pm 2.58\sqrt{\frac{.91(1-.91)}{173}}$$

$$= .91 \pm 2.58\sqrt{\frac{.91(.09)}{173}} = .91 \pm 2.58\sqrt{\frac{.08}{173}}$$

$$= .91 \pm 2.58\sqrt{.0005}$$

= .91 ± 2.58 (0.02)

= .91 ± .05.

The calculation of the lower and upper limits and the formal statement of the confidence interval proceeds along the same lines for proportions as for means. In the current example,

LL = .91 − .05 = .86

UL = .91 + .05 = .96

and

$99\% CI$: .86 <=P<=.96

We can say with 99% confidence that the interval .86 to .96, inclusive, contains the true population mean P.

Learning Check 8.5

✓

Redo the confidence interval for the analysis of handgun usage in fights, this time using a 95% confidence level. What happened to the width of the interval? Why did this happen?

For a second example, we can again use the FISS and this time analyze the relationship between shooters and victims in arguments. Among those gunshots resulting from arguments ($N = 139$), 28.8% of victims were shot by friends or acquaintances (again, excluding cases in which the victim–shooter relationship was unknown). This variable meets the sample size requirement; since there were 40 friend/acquaintance shootings (successes) and 99 cases involving different relationships (failures), the requirement that there be a minimum of 5 successes and failures each is also met. We can calculate a confidence interval. The proportion is 28.8/100 = .288, which rounds to .29. The confidence level will be set at 95%, making the critical value of z ±1.96. Using Formula 8(4),

$$CI = .29 \pm 1.96\sqrt{\frac{.29(1-.29)}{139}} = .29 \pm 1.96\sqrt{\frac{.29(.71)}{139}}$$

$$= .29 \pm 1.96\sqrt{\frac{.21}{139}} = .29 \pm 1.96\sqrt{.002}$$

$$= .29 \pm 1.96\,(.04)$$

$$= .29 \pm .08.$$

The lower and upper limits are

$LL = .29 - .08 = .21$

$UL = .29 + .08 = .37$

Finally, the confidence interval is

95% CI : $.21 \le P \le .37$

There is a 95% chance that the interval .21 to .37, inclusive, contains the true population proportion P.

> Research Example 8.3 Is There a Relationship Between Unintended Pregnancy and Intimate Partner Violence?
>
> Intimate partner violence (IPV) perpetrated by a man against a female intimate is often associated with not just physical violence but

also a multifaceted web of control in which the woman becomes trapped and isolated. One of the possible consequences is that abused women might not have access to reliable birth control and might therefore be susceptible to unintended pregnancies. These pregnancies, moreover, might worsen the IPV situation because of the emotional and financial burden of pregnancy and childbearing. Martin and Garcia (2011) sought to explore the relationship between IPV and unintended pregnancy in a sample of Latina women in Los Angeles. Latina immigrants might be especially vulnerable to both IPV and unintended pregnancy because of the social isolation faced by those who have not assimilated into mainstream U.S. society. Staff at various prenatal clinics in Los Angeles distributed surveys to their Latina patients. The surveys asked women several questions pertaining to whether their pregnancy had been intentional, whether they had experienced emotional or physical abuse by their partner before or during pregnancy, and their level of assimilation into the United States (as compared to feeling isolated or alienated). They used a statistic called an *odds ratio* . An odds ratio measures the extent to which consideration of certain independent variables (IVs) changes the likelihood that the dependent variable (DV) will occur. An odds ratio of 1.00 means that the IV does not change the likelihood of the DV. Odds ratios greater than 1.00 mean that an IV increases the probability that the DV will occur, and those less than 1.00 indicate that an IV reduces the chances of the DV happening.

In Martin and Garcia's (2011) study, the DVs in the first analysis were physical and emotional abuse and the DV in the second analysis was physical abuse during pregnancy. The researchers found the following odds ratios and confidence intervals.

Prepregnancy physical abuse was not related to the chances that a woman would become pregnant accidentally, as indicated by the odds ratio of .92, which is very close to 1.00. Emotional abuse, surprisingly, significantly reduced the odds of accidental pregnancy (odds ratio = .50). This was an unexpected finding because the researchers predicted that abuse would increase the odds of pregnancy. They theorized that perhaps some emotionally abused women try to get pregnant out of a hope that having a child will improve the domestic situation.

	DV: Change in Odds of Unintended Pregnancy	95% CI
IV: Prepregnancy Physical Abuse	.92	.40 ≤ population odds ≤ 2.16
IV: Prepregnancy Emotional Abuse	.50	.26 ≤ population odds ≤ .97

	DV: Change in Odds of Physical Abuse
IV: Unintended Pregnancy	2.80
95% *CI*	1.01 ≤ population odds ≤ 7.73

Turning to the second analysis, it can be seen that unintended pregnancy substantially increased women's odds of experiencing physical abuse during pregnancy. Contrary to expectations, women's level of acculturation did not alter the odds of abuse or pregnancy, once factors such as a woman's age and level of education were accounted for. The findings indicated that the relationship between IPV and unintended pregnancy is complex and deserving of further study in order to identify risk factors for both.

Why Do Suspects Confess to Police?

The Fifth Amendment to the U.S. Constitution that provides protection from compelled self-incrimination ensures that persons suspected of having committed criminal offenses can refuse to answer questions about the crime during police interrogations. Despite this right, a good number of suspects do talk and do provide incriminating statements. As the wry saying goes, "Everyone has the *right* to remain silent, but not everyone has the *ability* to do so." So what makes suspects confess? Deslauriers-Varin, Beauregard, and Wong (2011) sought to identify some of the contextual factors that make suspects more likely to confess, even when those suspects initially indicated that they wished to remain silent. The researchers obtained a sample of 211 convicted male offenders from a Canadian prison and gathered extensive information about each participant. The researchers analyzed odds ratios. Recall that an odds ratio of 1.00 means that the IV does not alter the odds that the DV will occur. Odds ratios less than 1.00 mean the IV makes the DV less likely, whereas odds ratios greater than 1.00 indicate that the IV makes the DV more likely to happen. The researchers found the following odds ratios and confidence intervals.

The numbers in the following table that are flagged with asterisks are those that are statistically significant, meaning that the IV exerted a noteworthy impact on suspects' decision to remain silent rather than confessing. The initial decision to not confess was the strongest predictor of suspects' ultimate refusal to provide a confession; those who initially resisted confessing were likely to stick to that decision. Criminal history was also related—having only one or two priors was not related to nonconfession, but suspects with three or more prior convictions were substantially more likely to remain silent. This might be because these suspects were concerned about being sentenced harshly as habitual offenders. The presence of an accomplice also made nonconfession more likely, as did a lawyer's advice to not confess. Those accused of drug-related crimes were more likely to not confess, though crime type was not significant for suspects accused of other types of offenses. Finally, the strength of police evidence was a factor in suspects' decisions. Strong police evidence was related to a significant reduction in the odds of nonconfession (i.e., strong evidence resulted in a greater chance that the suspect would confess).

It appeared, then, that there are many factors that impact suspects' choice regarding confession. Most of these factors appear to be out of the control of police interrogators; however, the researchers did not include variables measuring police behavior during interrogation, so there might well be techniques police can use to elicit confessions even from those suspects who are disinclined to offer information. One policy implication from these results involves the importance of the initial decision in the final decision— 79% of offenders stuck with their initial decision concerning whether to confess. Police might, therefore, benefit from focusing their efforts on influencing the initial decision rather than allowing a suspect to formulate a decision first and then applying interrogation tactics.

DV: Suspect Did Not Confess		
Independent Variable	Change in Odds of No Confession	95% CI
Initial Decision: No Confession	25.33*	10.05, 63.82
Criminal History		
1 or 2 Prior Convictions	3.01	.79, 11.42
3+ Prior Convictions	13.29*	3.05, 57.91
Had an Accomplice	2.58*	1.02, 6.52
Police Evidence Is Strong	.23*	.10, .55
Lawyer Advised Nonconfession	3.29*	1.34, 8.07
Crime Was Drug Related	3.55*	1.26, 10.02

Source: Adapted from Table 2 in Deslauriers-Varin et al. (2011).

*Statistically significant.

Source: Adapted from Table 2 in Deslauriers-Varin et al. (2011).

We can also construct a confidence interval using the COJ to estimate the proportion of jail inmates being held for felonies. In a random sample of 177 facilities, 60.7% of inmates were incarcerated for felonies (either awaiting trial or after having been convicted). The sample size is large enough to permit the use of the z distribution, and there are far more than five successes (people confined for felonies) and failures (people held for misdemeanor and other types of offenses). The threshold criteria having been satisfied, we can proceed to the calculations. We will use the 95% confidence level. Dividing 60.7 by 100 and rounding yields a proportion of .61; the computation is

$$CI = .61 \pm 1.96\sqrt{\frac{.61(1-.61)}{177}} = .61 \pm 1.96\sqrt{\frac{.61(.39)}{177}}$$

$$= .61 \pm 1.96\sqrt{\frac{.24}{177}} = .61 \pm 1.96\sqrt{.001}$$

$= .61 \pm 1.96\,(.03)$

$= .61 \pm .06$

The lower and upper limits are

$LL = .61 - .06 = .55$

$UL = .61 + .06 = .67$

Finally, the confidence interval is

95% $CI: .55 \leq P \leq .67$

There is a 95% chance that the interval .55 to .67, inclusive, contains the true population proportion P. Since this is a census, we can check to see if our interval actually does contain the population proportion. The proportion of inmates held for felonies across the entire sample is 60.10%. Not only does the calculated interval contain the mean; it also turns out that the sample mean is nearly identical to the population mean! Of course, this is information researchers typically would not have at their disposal.

Chapter Summary

Confidence intervals are a way for researchers to use sample statistics to form conclusions about the probable values of population parameters. Confidence intervals entail the construction of ranges of values predicted to contain the true population parameter. The researcher sets the level of confidence (probability of correctness) according to her or his judgment about the relative costs of a loss of

confidence versus compromised precision. The conventional confidence levels in criminal justice and criminology research are 95% and 99%. The decision about level of confidence must be made with consideration to the trade-off between confidence and precision: As confidence increases, the quality of the estimate diminishes. All confidence intervals are two-tailed, which means that alpha is divided in half and placed in both tails of the distribution. This creates two critical values. The critical values have the same absolute value, but one is negative and one is positive.

There are two types of CI_s for means: large sample and small sample. Confidence intervals for means with large samples employ the z distribution, whereas those for means with small samples use the t curve. When the t distribution is used, it is necessary to calculate degrees of freedom in order to locate the critical value on the t table.

Confidence intervals can be constructed on the basis of proportions, providing that two criteria are met: First, the sample must contain at least 100 cases. Second, there must be at least five successes and five failures in the sample. These two conditions help ensure the normality of the sampling distribution and, thus, the applicability of the z curve.

Thinking Critically

1. Two researchers are arguing about confidence intervals. One of them contends that precision is more important than the level of confidence, and the other claims that confidence is more important than precision. What would you say to them to help resolve this disagreement? Identify a research scenario in social science or other sciences (medicine, biology, etc.) in which precision would be of utmost concern and one in which confidence would be researchers' main goal. Explain your rationale for each.

2. Suppose you read a research report that claimed to have uncovered a finding with significant implications for policy. The researchers summarize the results of a multicity homicide reduction program that they claim was markedly effective. They support this claim by pointing out that, during the study period, there was a mean reduction of 1.80 homicides per month across the cities in the sample. Along with that mean, they report a 95% confidence interval of $-.40 \leq \mu \leq 4$ (i.e., 1.80 ± 2.20). What is wrong with the researchers' claim that this program was effective? Identify the flaw in their logic and explain why that flaw makes their claim unsupportable.

Review Problems

Answer the following questions with regard to confidence intervals .

1. How many cases must be in a sample for that sample to be considered "large"?
2. "Small" samples are those that have _____ or fewer cases.
3. Which distribution is used with large samples?
4. Which distribution is used with small samples?
5. Why can the distribution that is used with large samples not also be used with small ones?
6. Explain the trade-off between confidence and precision.
7. The Law Enforcement Management and Administrative Statistics (LEMAS) survey asks agencies to report the number of sworn personnel who are designated school resource officers (SROs). In Florida municipal police departments serving populations of 1 million or more, the agencies sampled in LEMAS ($N = 18$) reported a mean of 13.00 SROs ($s = 12.10$). Construct a 95% confidence interval around this sample mean, and interpret the interval in words.
8. The sheriff's offices sampled in LEMAS ($N = 827$) reported that their agencies require new recruits to complete a mean of 599 hours of academy training ($s = 226$). Construct a 99% confidence interval around this mean and interpret the interval in words.
9. The PPCS asks respondents who have been stopped by the police while driving a vehicle how many officers were on the scene during the stop. Female stopped drivers ($N = 2,033$) reported a mean of 1.14 ($s = .36$) officers. Construct a 95% confidence interval around this sample mean, and interpret the interval in words.
10. In the PPCS, respondents who say that they have been stopped by police while driving vehicles are asked to report the reason why they were stopped and the total length of time that the stop took. Among female drivers stopped for illegal use of cellphones while driving ($N = 42$), stops lasted a mean of 9.00 minutes ($s = 5.48$). Construct a 95% confidence interval around this sample mean and interpret the interval in words.
11. The General Social Survey (GSS) contains an item asking respondents about their TV habits. Female respondents ($N = 707$)

said they watched a mean of 3.06 (s = 2.62) hours of TV per day. Construct a 99% confidence interval around this sample value and interpret the interval in words.

12. In a random sample of jails (N = 110) from the COJ, the mean number of male juveniles confined per facility was 1.17 (s = 6.21). Construct a 99% confidence interval around this sample value and interpret the interval in words.

13. In a random sample of prisons (N = 30) from the COJ, the mean number of inmates per security-staff member was 4.34 (s = 3.94). Construct a 95% confidence interval around this sample value and interpret the interval in words.

14. Respondents to the GSS (N = 1,166) worked a mean of 40.27 hours per week (s = 15.54). Construct a 99% confidence interval around this sample value, and interpret the interval in words.

15. The GSS asks respondents whether they keep a gun in their homes. In this survey, 34% of respondents (N = 1,281) said that they have at least one firearm. Construct a 95% confidence interval around this sample value and interpret the interval in words.

16. In the LEMAS survey, 45% of sampled sheriff's offices (N = 823) reported that their agencies' formal mission statements do not include a community-policing component. Construct a 95% confidence interval around this sample value, and interpret the interval in words.

17. According to LEMAS, 31% of municipal law-enforcement agencies (N = 1,967) use computerized statistics to identify high-crime hot spots. Construct a 95% confidence interval around this sample value and interpret the interval in words.

18. The FISS captures information on whether drugs were involved in the altercation that led up to a shooting. In 2013, victims were male in 87.3% of the 142 cases that involved drugs. Construct a 99% confidence interval around this sample value, and interpret the interval in words.

19. The FISS reports that of the 329 female shooting victims whose relationship with their shooter was known, 28.3% were shot by strangers. Construct a 95% confidence interval around this sample value, and interpret the interval in words.

20. The GSS found that 74% of male respondents (N = 573) believe that people suffering from incurable diseases should be permitted to die if that is their choice. Construct a 99% confidence interval around this sample value and interpret the interval in words.

Key Terms

Point estimate 177

Confidence interval 177

Level of confidence 178

Alpha level 180

Critical value 180

Two-tailed test 180

Glossary of Symbols and Abbreviations Introduced in This Chapter

α	The probability that a *CI* does not contain the population parameter
z	The *z* score associated with a given a level
t_α	The *t* score associated with a given a level

Part III Hypothesis Testing

You have now learned about descriptive statistics (Part I) and the theories of probability and distributions that form the foundation of statistics (Part II). Part III brings all of this together to form what most criminal justice and criminology researchers consider to be the high point of statistics: **inferential analyses** or what is also called hypothesis testing. Hypothesis testing involves using a sample to arrive at a conclusion about a population. Samples are vehicles that allow you to make generalizations or predictions about what you believe is happening in the population as a whole. The problem, as we discussed in Chapter 7, is sampling error: It is erroneous to conclude that a sample statistic is an accurate reflection of the population parameter, because the sample is merely one of a multitude of samples that could have been drawn from the population and, therefore, there is an unknown amount of error. Refer back to Figures 7.5 and 7.6 for an illustration.

Inferential analysis: The process of generalizing from a sample to a population; the use of a sample statistic to estimate a population parameter. Also called hypothesis testing.

To account for sampling error, inferential statistics use sampling distributions to make probabilistic predictions about the sample statistic that is being analyzed. The basic strategy is a two-step process: First, a sample statistic is computed. Second, a probability distribution is used to find out whether this statistic has a low or high probability of occurrence. This process should sound very familiar—we already followed these steps when we worked with z scores and areas under the standard normal curve. Hypothesis testing is an expansion of this underlying idea. A sample statistic (such as a mean) can be used to find out whether this value is close to the center of the distribution (i.e., has a high probability of occurrence) or far out in the tail (low probability of occurrence). It is this probability assessment that guides researchers in making decisions and reaching conclusions.

Part III covers some of the bivariate (i.e., involving two variables) inferential tests commonly used in criminology and criminal justice research: chi-square tests of independence, two-population tests for differences between means and between proportions, analyses of variance, and correlations. Part III ends with an introduction to bivariate and multiple regression.

Levels of Measurement
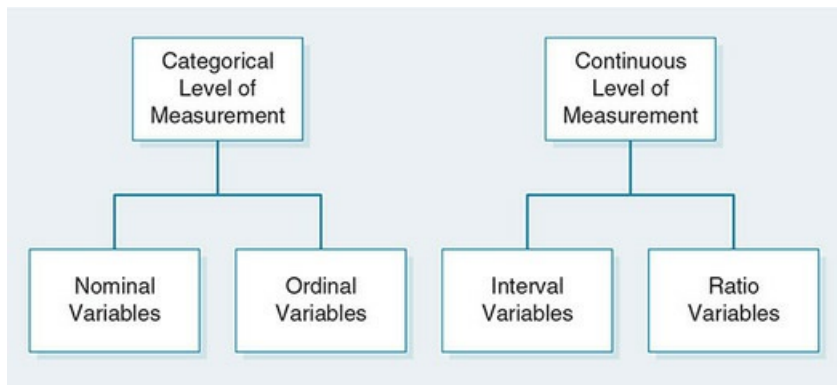
The proper test to use in a given hypothesis-testing situation is determined by the level of measurement of the variables with which you are working. If your memory of levels of measurement has become a bit fuzzy, go back to Chapter 2 now and review this important topic. You will not be able to select the correct analysis unless you can identify your variables' levels of measurement. The figure here is a reproduction of Figure 2.1 showing the levels of measurement. For purposes of hypothesis testing, the most important distinction is that between categorical and continuous variables. Be sure you can accurately identify any given variable's level of measurement before you proceed.

There is no sense denying that you might find many of the concepts presented in the following chapters confusing at first. This is normal! Remember that the process of learning statistics hinges on repetition. Read and reread the chapters, study your lecture notes, and do the chapter review problems at least one time each—things will start to sink in. Terms, formulas, and ideas that initially seemed incomprehensible will gradually solidify in your mind and begin making sense. Remember that most criminology and criminal justice researchers started off in a position just like yours! There was a point when they knew very little about statistics and had to study hard to develop a knowledge base. Commit the time and effort and you will be pleasantly surprised by how well you do.

Learning Check

✓

Take a moment now to test your memory of levels of measurement. Identify the level of measurement of each of the following variables:

1. The survey item that asks, "How many times have you been arrested in your life?" and has respondents write in the answer
2. The survey item that asks, "How many times have you been arrested in your life?" and has respondents circle never, 1–3 times, or 4 or more times
3. Whether a convicted defendant's sentence was jail, probation, or drug treatment
4. Defendants' annual household income, measured as $0–$9,999, $10,000–$19,999, or $20,000 or more
5. The method by which a defendant was convicted, measured as guilty plea, jury trial, or bench trial
6. In a sample of people responding to a survey, attitudes about the court system ranging from "1 = too harsh" to "5 = too lenient."

7. Now you try it! Create four variables, one representing each level of measurement.

Chapter 9 **Hypothesis Testing** A Conceptual Introduction

The purpose of this chapter is to provide a clear conceptual foundation explaining the nature and purpose of hypothesis testing. It is worth developing a solid understanding of the substance of inferential statistics before approaching the specific types of hypothesis tests covered in the chapters that follow. This chapter will help you grasp the overarching logic behind these sorts of tests so that you can approach the "trees" (specific tests) with a clear picture of what the "forest" (underlying conceptual framework) looks like.

By now, you should be very familiar with the idea that researchers are usually interested in populations but, because populations are so large, samples must suffice as substitutes. Researchers draw random samples using a variety of methods. A researcher distributing surveys by mail might use a police department's address database to electronically pull the addresses of 10% of the residences in a particular city. Someone conducting phone surveys might use random-digit dialing to contact 200 respondents. The ultimate goal in statistical research is to generalize from the sample to the population. Hypothesis testing is the process of making this generalization. (Note that throughout the following discussion, we are going to assume that samples are simple and random. When either of these two criteria is not true in a given sample, adjustments sometimes have to be made to the statistics used to analyze them. For present purposes, we are going to assume we are working with simple, random samples.)

What we are really talking about in inferential statistics is the probability of empirical outcomes. There are many (even infinite) random samples that can be drawn from any given population and, therefore, are numerous possible values for sample statistics to take on. A population with a mean of 10, for instance, can produce samples with means of 9, 11, 10, 7, and so on. When we have a sample statistic that we wish to use inferentially, the question asked is, "Out of all the samples and sample statistics possible, what is the probability that I would draw *this* one?"

Learning Check 9.1

Remember the difference between expected and observed or empirical outcomes. Expected outcomes are the results you anticipate seeing on the basis of probability theory. In Chapter 6, you constructed a table of expected outcomes in the context of binomials. You did not use any data; you made this table entirely on the basis of probability calculations. Observed outcomes, by contrast, are what you actually see. These results might or might not mirror expectations. A coin-flip exercise will help refresh your memory. Write down the probability of any given coin flip resulting in heads. Now flip a coin six times and record each outcome; then tally up the total number of heads and tails. Did you see what you expected, on the basis of the underlying probability, or were you surprised at the outcome? Try again with 10 flips. Did the observed outcome match the one you expected?

Sample Statistics and Population Parameters: Sampling Error or True Difference?

Any time a sample statistic is not equal to a population parameter, there are two potential explanations for the difference. (Well, technically there are three, since a mismatch can result from mistakes in the sampling process. For our purposes, though, as mentioned earlier, we are assuming correct research methods and simple, random samples.) First, the inequality could be the product of inherent random fluctuations in sample statistics (i.e., sampling error). In other words, the disparity might simply be a meaningless fluke. If you flipped a fair coin six times, you would expect the coin to land tails side up three times. If, instead, you got four tails, you would not think that there was anything weird happening; the next set of six trials might result in two tails. This is sampling error—variation that is like white noise in the background.

The second possible explanation for the difference is that there is a genuine discrepancy between the sample statistic and the population parameter. In other words, the disparity could represent a bona fide statistical effect. If you flipped a coin 20 times and got 19 tails, you would suspect there was something wrong with the coin because this is an extremely unlikely outcome. Perhaps the coin is weighted on one side, which would mean that it is different from the ordinary quarter or dime you might have in your pocket. Large discrepancies between observed and expected outcomes are sufficiently improbable to lead us to conclude that there is something genuinely unique about the empirical outcome we have in front of us.

When researchers first approach an empirical finding, they do not know which of the two possible explanations accounts for the observed or empirical result. In the earlier coin-flip example, we knew the underlying population probability (.50), and we knew the number of trials that had been conducted (six in the first and 20 in the second). If either of those pieces of information is omitted, then it becomes difficult to make sense of the results. If your friend told you that he flipped a coin and it landed on tails seven times, but he did not tell you the total number of times he flipped it, then you would not know how to interpret his report about seven tails. Similarly, if you did not know that every flip has a .50 probability of tails (and that, by extension, roughly half of a string of flips will be tails), then your friend might say that he flipped a coin 14 times and got seven tails and you would not have a clue as to whether this result is normal or atypical.

In the real world of criminal justice and criminology research, there are missing bits of information that prevent us from being able to readily discriminate between sampling error and true difference. The overarching purpose of hypothesis testing is to determine which of them appears to be the more valid of the two, based on a calculated judgment. This is where probabilities come in. Researchers identify the probability of observing a particular empirical result and then use that probability to make a decision about which explanation seems to be correct.

This is pretty abstract, so an example is in order. We will use the Law Enforcement Management and Administrative Statistics (LEMAS; see [Data Sources 3.2](#)) survey. Suppose we are investigating whether agency type (municipal, county, tribal, or state) affects the ratio of officers to residents within a given jurisdiction. (This is a measure of agency size relative to the size of the population served.) We might predict

that municipal departments have a higher officer-to-resident ratio than county sheriff's offices do. This is because city and town police departments typically serve more-centralized populations, and the concentration is linked with higher crime rates. By contrast, sheriffs' offices generally cover larger territories and populations that are more spread out, with lower crime rates. Using LEMAS, we can calculate the mean ratio of officers per 1,000 residents for each type of agency in the sample. Municipal departments' mean is 2.20, and sheriffs' offices mean is 1.08.

You can see that these means are unequal, and you might be tempted to conclude that our hypothesis has been supported (i.e., that municipal departments are indeed larger than sheriffs' offices, relative to the size of the population served). However, recall that there are *two potential reasons* for this disparity. Their inequality might be meaningless and the differences between the numbers purely the product of chance; in other words, the finding might be a fluke. This is a random sample of agencies, so we cannot rule out the possibility that sampling error manufactured an apparent difference where there actually is none. On the other hand, the means might be unequal because officer-to-resident ratios really do vary across these two types of agencies. In other words, there might be a real difference between the means. These two competing potential explanations can be framed as hypotheses. We will use the LEMAS numbers to guide us through a discussion of these competing hypotheses.

Null and Alternative Hypotheses

There are two hypotheses used to state predictions about whether or not a statistic is an accurate estimate of a parameter. The first is called the **null hypothesis**. The null (symbolized H_0) represents the prediction that the difference between the two samples is merely sampling error or, in other words, chance variation in the data. You can use the word *null* as its own mnemonic device because this word means "nothing." Something that is null is devoid of meaning. In the context of the present example, the null predicts that agency type is not related to officer-to-resident ratio and that the observed difference between the means is just white noise.

> **Null hypothesis**: In an inferential test, the hypothesis predicting that there is no relationship between the independent and dependent variables. Symbolized H_0.

The second possible explanation is that municipal departments really do have a higher ratio than sheriffs' offices do. This prediction is spelled out in the **alternative hypothesis** (symbolized H_1). The alternative hypothesis is sometimes also called the research hypothesis. The alternative or research hypothesis is, essentially, the opposite of the null: The null predicts that there is no relationship between the two variables being examined, and the alternative predicts that they are related.

> **Alternative hypothesis**: In an inferential test, the hypothesis predicting that there is a relationship between the independent and dependent variables. Symbolized H_1. Also referred to as a research hypothesis.

In the context of the present example, the null and alternative hypotheses can be written as

H_0: *Municipal departments and sheriffs' agencies have the same officer-to-resident ratio; that is, the two means are equal, and there is no relationship between agency type and ratio.*
H_1: *Municipal departments have a higher officer-to-resident ratio than sheriffs' agencies do; that is, the means are unequal and there is a relationship between agency type and ratio.*

More common than writing the hypotheses in words is to use symbols to represent the ideas embodied in the longhand versions. Transforming these concepts into such symbols turns the null and alternative hypotheses into

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$,

where μ_1 = the mean ratio in municipal departments, and

μ_2 = the mean ratio in sheriffs' offices.

It might seem strange to write the hypotheses using the symbol for the population mean (μ) rather than the

sample mean ($\overline{X}$), but remember that in inferential statistics, it is the population parameter that is of interest. We use the sample means to make a determination about the population mean(s). Basically, there are two options: There might be one population from which the samples derive (sampling error), or each sample might represent its own population (true difference). If the null is true and there is, in fact, no relationship between agency type and officer-to-resident ratio, then we would conclude that all the agencies come from the same population. If, instead, the alternative is true, then there are actually two populations at play here—municipal and county. Figure 9.1 shows this idea pictorially.

Figure 9.1 One Population or Two?

One Population or Two?

Under the null, all police agencies are part of a single population with one mean.

Under the alternative, there are two populations and the population mean for municipal departments (μ_1) is higher than that for sheriff's offices (μ_2).

μ μ_2 μ_1
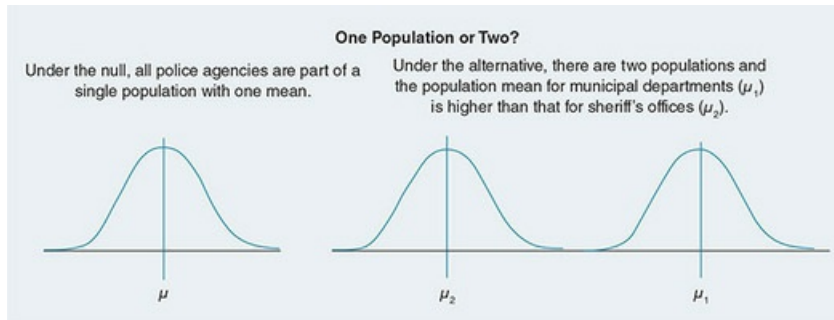
Another reason a researcher might conduct a hypothesis test is to determine if men and women differ in terms of how punitive they feel toward people who have been convicted of crimes. The General Social Survey (GSS; see Data Sources 2.2) asks people whether they favor or oppose capital punishment for people convicted of murder. Among men, 31% oppose capital punishment; this number is 40% among women. A researcher might want to know whether this difference represents a genuine "gender effect" or whether it is merely chance variation. The null and alternative could be set up as such:

> H_0: *Men and women oppose the death penalty equally; the proportions are equal and there is no relationship between gender and death penalty attitudes.*
> H_1: *Men are less likely to oppose the death penalty than women are; the proportions are unequal and there is a relationship between gender and death penalty attitudes.*

Formally stated using symbols, the null and alternative are written as

H_0: $P_1 = P_2$

H_1: $P_1 < P_2$

where P_1 = the proportion of men who oppose capital punishment and

P_2 = the proportion of women who oppose capital punishment.

As with the example regarding police agency type and size, the hypotheses pertaining to gender and death

penalty attitudes are phrased in terms of the population proportion (P) rather than the sample proportion ($\hat{p}$) because we are using a sample estimate to draw an inference about a population parameter. Until we conduct a full hypothesis test, we will not know whether women and men truly do differ in their opposition to capital punishment.

The assumption going into an inferential analysis is that the null is the true state of affairs. In other words, the default assumption is that there is no relationship between the two variables under examination. The goal in conducting the test is to decide whether to retain the null (concluding that there is no relationship) or to reject the null (concluding that there is, in fact, a relationship between the variables). The null can be rejected only if there is solid, compelling evidence that leads you to decide that this hypothesis is inaccurate.

A good analogy to the logic behind hypothesis testing is the presumption of innocence in a criminal trial. At the outset of a trial, the jury must consider the defendant to be legally innocent of the crime of which she or he is accused. The "null" here is innocence and the "alternative" is guilt. If the prosecutor fails to convincingly show guilt, then the innocence assumption stands, and the defendant must be acquitted. If, however, the prosecutor presents sufficient incriminating evidence, then the jury rejects the assumption of innocence and renders a guilty verdict. Of course, the prosecutor does not have to leave the jury with complete certainty that the defendant is guilty; rather, the prosecutor's job is to overcome a reasonable doubt. In other words, the jury can convict when the probability of the defendant's guilt far outweighs the probability of his innocence.

There are good reasons for the null being the default in both criminal trials and scientific research. The presumption of innocence helps prevent wrongful convictions. Likewise, in clinical trials testing new pharmaceutical drugs, it is of utmost importance that a drug be demonstrated to be effective and safe before it is approved and put on the market. Medical researchers err on the side of caution. If there is a .50 probability that a drug will help the people who take it achieve better health, then this might be sufficient to send that drug to market. On the other hand, if a drug has a .60 probability of making people who take it feel better and a .15 probability of killing them, researchers should not let it be released for public consumption.

In criminal justice and criminology research, important questions about theory and policy hang in the balance. Like medical researchers, social scientists err on the side of caution. A criminologist testing a hypothesis about a theory of crime causation must tread carefully. Failing to reject the null hypothesis could lead to erroneous conclusions about the accuracy of the theory; however, the test can be repeated to determine whether there was a mistake in the research design that corrupted the results. Rejecting a null hypothesis that is true, however, could take this criminologist and others doing similar research down a completely wrong path in the study of crime causation. Criminal justice researchers, likewise, often deal with policy questions. They handle matters such as whether or not a particular rehabilitation program reduces recidivism, whether a given policing strategy deters violent crime, and whether school-based prevention programs prevent drug use and gang involvement among youth. For each of these examples, you can see that neither of the two possible mistakes is harmless but that retaining the null (and, possibly, conducting further research into the issue) is safer than leaping to a conclusion that might be incorrect. Erroneously concluding that a rehabilitation program, police strategy, or school-based prevention program works could result in their widespread adoption

across jurisdictions, even though they are ineffective.

In inferential statistics, researchers construct a probability framework based on the assumption of a true null. The question they are trying to answer is, "What is the probability of observing the empirical result that I see in front of me *if the null hypothesis is correct*?" If the probability of the null being true is extremely low, then the null is rejected because it strains the imagination to think that something with such a low likelihood of being correct is the right explanation for an empirical phenomenon. The alternative would, thus, be taken as being the more likely version of reality. If the probability of the null being true is not low, then the null is considered to be a viable explanation for the results, and it is retained.

Think back to the study of police agency type and the ratio of officers per 1,000 local residents. Recall that municipal departments' mean officer-to-resident ratio is 2.20 and sheriff's offices is 1.08. Let us say, for the sake of example, that we determine that the probability of these two means being two parts of the same population is .70. That is a pretty high probability! We would conclude that the null is likely correct and there is only one population. What if we found that the probability was .20? This is a much smaller probability than .70, but it still means that there is a 20% chance that the null is true, which is substantial. A jury in a criminal trial should not convict a defendant if there is a 20% chance he is innocent. If, on the other hand, we found a probability of .01, meaning that there is only a 1% chance that the two samples are from the same population, then rejecting the null would be warranted because it is extremely unlikely to be the true state of affairs (i.e., there is a .99 or 99% chance that the null is false). At a probability of .01, it is highly likely that municipal police departments and county sheriff's agencies are two separate populations with different means.

Of course, as you have probably already figured out, there is always a chance that a researcher's decision regarding whether to reject or retain the null is wrong. We saw when we worked with confidence intervals that the flipside of the probability of being right is the probability of being wrong—any time you make a decision about the null, you are either right or wrong, so the two probabilities sum to 100% or 1.00. There are two types of errors that can be made in this regard. A **Type I error** occurs when a true null is erroneously rejected, whereas a **Type II error** happens when a false null is inaccurately retained. Type I errors are like false positives, and Type II errors are like false negatives. Wrongfully convicting an innocent defendant or approving an ineffective drug for the market is a Type I error, while incorrectly acquitting a guilty defendant or concluding that an effective drug does not work is a Type II error.

Type I error: The erroneous rejection of a true null hypothesis. Symbolized α.

Type II error: The erroneous retention of a false null hypothesis. Symbolized β.

Type I errors are often symbolized using α, which we have seen before. Recall from the confidence interval lesson in [Chapter 8](#) that alpha is the probability of being wrong about the interval containing the true population mean or proportion. The interpretation of alpha in inferential statistics is a little different from confidence intervals because of the involvement of null and alternative hypotheses; however, the underlying logic is the same. The symbol for a Type II error is the uppercase Greek letter beta (β). Researchers can often minimize the probability that they are wrong about a decision, but they can never eliminate it. For this reason,

you should always be circumspect as both a producer and a consumer of statistical information. Never rush haphazardly to conclusions. Any time you or anyone else runs a statistical analysis and makes a decision about the null hypothesis, there is a probability—however minute—that the decision is wrong.

Table 9.1 Type I and Type II Errors		
	. . . and you reject it	*. . . and you retain it*
If the null is true . . .	Type I Error (α)	Correct!
If the null is false . . .	Correct!	Type II Error (β)

There is a trade-off between Type I and Type II error rates. The Type I error rate (α) is set a priori (in advance) of the start of the hypothesis test. A researcher who is worried about making a Type I error could help minimize the chance of this mistake occurring by setting alpha very low, which increases the difficulty of rejecting the null hypothesis. The flipside, however, is that making it harder to reject a true null also makes it hard to reject a false one. By reducing the chance of a Type I error, the researcher has increased the risk of making a Type II error.

The following chapters will cover several different types of hypothesis testing procedures in the bivariate (i.e., two variables) context. The choice between the different tests is made on the basis of each variable's level of measurement. You must identify the levels of measurement of the IVs and DVs and then select the proper test for those measurement types. This book covers four types of bivariate inferential tests. The first is chi-square, which is the statistical procedure used to test for an association between two categorical (nominal or ordinal) variables. If, for example, you had a sample of criminal defendants and wanted to find out whether there was a relationship between the type of crime a defendant was charged with (violent or property) and the disposition method that the defendant chose (guilty plea, jury trial, or bench trial), you would use a chi-square test.

The second type of analysis is a *t* test. The *t* test is used when the DV of interest is continuous (interval or ratio) and the IV is categorical with two classes (such as gender measured as male or female). The *t* test is a test for differences between two means. In the officer-to-resident ratio example, a *t* test would be the analysis we select to determine whether or not to reject the null hypothesis (since we had two types of police agencies and the ratio variable is continuous). The third type of test is the analysis of variance (ANOVA). The ANOVA is an extension of the *t* test and is used when the DV is continuous and the IV is categorical with three or more classes. If we added state police agencies to our ratio analysis, we would use an ANOVA instead of a *t* test. The rationale behind this is that conducting multiple *t* tests is time consuming and cumbersome, and creates statistical problems. The ANOVA streamlines the process by using a single analysis across all classes of the IV.

The final bivariate inferential test we will discuss is correlation. This test is used when both the DV and the IV are continuous. Correlations are tests for linear relationships between two variables. You might predict, for instance, that the level of correctional officer staffing in a prison (i.e., the inmate-to-staff ratio) affects the assault rate—it stands to reason that higher staffing results in better supervision and, thus, fewer assaults. You would test this prediction using a correlation analysis. Table 9.2 is a handy chart that you should study closely and refer to repeatedly throughout the next few chapters.

Table 9.2 Choosing the Appropriate Bivariate Test Based on the Variables' Level of Measurement

The IV is . . .	The DV is . . .	
	Categorical	Continuous
Categorical	Chi-square	t test ANOVA
Continuous	N/A	Correlation

If you choose the wrong test, you will arrive at an incorrect answer. This is true in both hand calculations and SPSS programming. SPSS rarely gives error messages and will usually run analyses even when they are deeply flawed. Remember, GIGO! When garbage is entered into an analysis, the output is also garbage. You must be knowledgeable about the proper use of these statistical techniques, or you risk becoming either a purveyor or a consumer of erroneous results.

Before we leave this chapter and dive into inferential analyses, let's introduce the steps of hypothesis testing. It is useful to outline a framework you can use consistently as you learn the different types of analyses. This lends structure to the learning process and allows you to see the similarities between various techniques. Hypothesis testing is broken down into five steps, as follows:

Step 1. State the null (H_0) and alternative (H_1) hypotheses.
  - The two competing hypotheses that will be tested are laid out.

Step 2. Identify the distribution and compute the degrees of freedom (df) .
  - Each type of statistical analysis uses a certain probability distribution. You have already seen the z and t distributions, and more will be introduced in later chapters. You have encountered the concept of degrees of freedom (df) in the context of the t distribution. Other distributions also require the computation of df.

Step 3. Identify the critical value of the test statistic and state the decision rule.
  - The critical value is based on probability. The critical value is the number that the obtained value (which will be derived in Step 4) must exceed in order for the null to be rejected. The decision rule is an a priori statement formally laying out the criteria that must be met for the null to be rejected. The decision rule is useful because it makes it very clear what must happen in order for the null to be rejected. You will return to the decision rule in Step 5 after computing the critical value in Step 4.

Step 4. Compute the obtained value of the test statistic.
  - This is the analytical heart of the hypothesis test. You will select the appropriate formula, plug in the relevant numbers, and solve. The outcome is the obtained value of the test statistic.

Step 5. Make a decision about the null and state the substantive conclusion.
  - You will revisit your decision rule from Step 3 and decide whether to reject or retain the null based on the comparison between the critical and obtained values of the test statistic. Then you will render a substantive conclusion. Researchers have the responsibility to interpret their statistical findings and draw substantive conclusions that make sense to other researchers and to the public.

Chapter Summary

This chapter provided an overview of the nature, purpose, and logic of hypothesis testing. The goal of statistics in criminology and criminal justice is usually generalization from a sample to a population. This is accomplished by first finding a sample statistic and then determining the probability that that statistic would be observed by chance alone. If the probability of the result being attributable solely to chance is exceedingly low, then the researcher concludes that the finding is not due to chance and is, instead, a genuine effect.

When a researcher has identified two variables that might be related, there are two possible true or correct states of affairs. The first possibility is that the variables are actually not related. This possibility is embodied by the null hypothesis, symbolized H_0. The second possibility is that they are in fact related to one another. This is the alternative hypothesis, H_1, which is sometimes also called a research hypothesis. The null hypothesis is always assumed to be the one that is true, and formal hypothesis testing employing probabilities and a sampling distribution is conducted to determine whether there is sufficient evidence to overrule the null and opt for the alternative instead. A hypothesis test using the five steps outlined in this chapter will ultimately result in the null being either rejected or retained, and the researcher concluding that the variables are, or are not, related to each other.

Thinking Critically

1. You have been summoned for jury duty and decide at the outset of the trial that you will not vote to convict unless you are 99.9% sure of the defendant's guilt. Which type of error does this decision minimize? What does this decision do to the probability that you will make the other type of error? If you reduce your certainty threshold to 75%, how have you altered the chances of making each error?

2. A friend of yours is excited about the results of an evaluation of a new rehabilitation program for adult prison inmates. She compared a sample of released prisoners who completed the program while they were incarcerated to a sample who did not, and she found that 21% of the treatment group was rearrested within one year of release, compared to 36% of the no-treatment group. She asserts that this is definitive proof that the program works. What should you tell your friend? Craft a response using the relevant concepts from this chapter. If you plan to recommend a statistical test, which one in Table 9.2 will you suggest?

Review Problems

1. Suppose a researcher was studying gender differences in sentencing. She found that males sentenced to jail received a mean of 6.45 months and females sentenced to jail received a mean of 5.82 months. Using what you learned in this chapter, describe the two possible reasons for the differences between these two means.

2. Write the symbol for the null hypothesis, and explain what this hypothesis predicts.

3. Write the symbol for the alternative hypothesis, and explain what this hypothesis predicts.

4. You learned in this chapter that the null is assumed to be true unless very compelling evidence suggests that the alternative hypothesis is actually the correct one. Why is this? That is, what is the rationale for the null being the default?

5. Explain what a Type I error is.

6. Explain what a Type II error is.

7. Explain the trade-off between Type I and Type II error rates.

8. Define the word *bivariate*.

9. List and describe each of the five steps for hypothesis tests.

10. If you computed an empirical result, identified the probability of observing that result, and found that the probability was high . ..

    1. would you conclude that this is the product of sampling error, or would you think that it is a true effect?

    2. would you reject or retain the null?

11. If you computed an empirical result, identified the probability of observing that result, and found that the probability was very low . ..

    1. would you conclude that this is the product of sampling error, or would you think that it is a true effect?

    2. would you reject or retain the null?

12. Which inferential statistical analysis would be used if the IV was criminal defendants' ages at sentencing (measured in years) and the DV length of their terms of confinement (measured in months)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation
13. Which inferential statistical analysis would be used if the IV was criminal defendants' gender (measured as male or female) and the DV was the length of their terms of confinement (measured in months)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation
14. Which inferential statistical analysis would be used if the IV was criminal defendants' gender (measured as male or female) and the DV was whether they obtained pretrial release (measured as yes or no)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation
15. Which inferential statistical analysis would be used if the IV was police force size (measured as the number of officers per 1,000 residents) and the DV was crime rates (measured as the number of crimes per 10,000 residents)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation
16. Which inferential statistical analysis would be used if the IV was assault victims' race (measured as white, black, Latino, or other) and the DV was the length of prison terms given to their attackers (measured in months)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation
17. Which inferential statistical analysis would be used if the IV was murder victims' race (measured as white, black, Latino, or other) and the DV was whether the killers were sentenced to death (measured as yes or no)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation
18. Which inferential statistical analysis would be used if the IV was assault victims' gender (measured as male or female) and the DV was the length of prison terms given to their attackers (measured in months)?
    1. Chi-square
    2. *t* test
    3. ANOVA
    4. Correlation

Key Terms

Glossary of Symbols and Abbreviations Introduced in This Chapter	
H_0	The null hypothesis
H_1	The alternative hypothesis
β	The symbol for the Type II error rate

Chapter 10 Hypothesis Testing With Two Categorical Variables Chi-Square

Learning Objectives

- Identify the levels of measurement of variables used with a chi-square test of independence.
- Explain the difference between parametric and nonparametric statistics.
- Conduct a five-step hypothesis test for a contingency table of any size.
- Explain what statistical significance means and how it differs from practical significance.
- Identify the correct measure of association used with a particular chi-square test, and interpret those measures.
- Use SPSS to produce crosstabs tables, chi-square tests, and measures of association.
- Interpret SPSS chi-square output.

The **chi-square test of independence** is used when the independent variable (IV) and dependent variable (DV) are both categorical (nominal or ordinal). The chi-square test is a member of the family of **nonparametric statistics**, which are used when sampling distributions cannot be assumed to be normally distributed, as is the case when a DV is categorical. Chi-square thus sits in contrast to **parametric statistics**, which are used when DVs are continuous (interval or ratio) and sampling distributions are safely assumed to be normal. The *t* test, analysis of variance, and correlation are all parametric. Because they have continuous DVs, they can rely on normal (or at least relatively normal) sampling distributions such as the *t* curve. (There are exceptions to the use of parametric statistics on continuous data, such as when the data are severely skewed, but that is beyond our scope here. For present purposes, we will distinguish the two classes of statistics on the basis of level of measurement.) Before going into the theory and math behind the chi-square statistic, Research Example 10.1 for an illustration of a type of situation in which a criminal justice or criminology researcher would turn to the chi-square test.

> **Chi-square test of independence**: The hypothesis-testing procedure appropriate when both the independent and dependent variables are categorical.
>
> **Nonparametric statistics**: The class of statistical tests used when dependent variables are categorical and the sampling distribution cannot be assumed to approximate normality.
>
> **Parametric statistics**: The class of statistical tests used when variables are continuous and normally distributed and the sampling distribution can be assumed to approximate normality.

A researcher might be interested in finding out whether male and female offenders receive different sentences. In a study like this, *gender* would be nominal (*male or female*) and *sentence* might be measured nominally as well (such as *jail, probation*, or *fine*). The question the researcher would be asking is, "Are these two variables related? In other words, does knowing an offender's gender help me predict what type of sentence he or she receives?" Answering this question requires the use of the chi-square test of independence because both the IV and the DV are categorical.

Conceptual Basis of the Chi-Square Test: Statistical Independence and Dependence

Two variables that are not related to one another are said to possess **statistical independence**. When two variables are related, they have **statistical dependence**. Statistical independence means that knowing which category an object falls into on the IV does not help predict its placement on the DV. Conversely, if two variables are statistically dependent, the independent does have predictive power over the outcome variable.

> **Statistical independence**: The condition in which two variables are not related to one another; that is, knowing what class persons or objects fall into on the independent variable does not help predict which class they will fall into on the dependent variable.
>
> **Statistical dependence**: The condition in which two variables are related to one another; that is, knowing what class persons or objects fall into on the independent variable helps predict which class they will fall into on the dependent variable.

Research Example 10.1 How Do Criminologists' and Criminal Justice Researchers' Attitudes About the Criminal Justice System Compare to the Public's Attitudes?

Many criminal justice and criminology researchers study public opinion about crime and the justice system. What is less commonly known, however, is how these academics themselves perceive crime and society's response to it. Griffin, Pason, Wiecko, and Brace (2016) compared the opinions of a sample of researchers to the opinions expressed by a sample of nonacademics from the general population.

The table displays some of their findings. (The percentages do not sum to 100% because the people responding "I don't know" to each question have been omitted from the table.) Substantial differences emerge between the two groups, with strong majorities of researchers opposing capital punishment and supporting marijuana legalization. The public was more likely to express favorable attitudes toward the death penalty and was more split on the issue of whether marijuana should be legalized.

Question	Public (%)		Researchers (%)	
	Yes	No	Yes	No
Favor the death penalty?	60	35	9.4	83.9
Feel the death penalty is a deterrent?	32	64	2	92.6
Is the death penalty applied fairly?	52	40	11	81.5
Should marijuana be made legal?	58	39	73	12.8

Source: Adapted from Table 1 in Griffin et al. (2016).
Source: Adapted from Table 1 in Griffin et al. (2016).

The researchers ran chi-square tests to determine whether there were differences between researchers and the public. All the tests showed that academics and nonacademics have noticeably divergent opinions about these issues.

In [Research Example 10.1](), the IV was whether survey respondents were either criminology or criminal justice researchers, or members of the public. There were four DVs (each coded yes/no) measuring respondents' attitudes about the death penalty and about marijuana legalization. If these two variables are statistically independent, the two groups (researchers and the public) will be largely similar in their attitudes. In other words, knowing whether someone is or is not a researcher will not help us predict that person's opinion on any

of these four dimensions. If they are statistically dependent, then we would gain predictive power by knowing which of the two groups a given individual is part of. When we want to know whether there is a relationship between two categorical variables, we turn to the chi-square test of independence.

The Chi-Square Test of Independence

Let us work slowly through an example of a chi-square hypothesis test using the five steps described in Chapter 9 and discuss each step in detail along the way. For this example, we will turn to the 2014 General Social Survey (GSS; see Data Sources 2.2) and the issue of gender differences in attitudes about crime and punishment. Theory is somewhat conflicting as to whether women tend to be more forgiving of transgressions and to prefer leniency in punishment or whether they generally prefer harsher penalties out of the belief that offenders pose a threat to community safety. The GSS contains data on the sex of respondents and these persons' attitudes toward the death penalty. The joint frequency distribution is displayed in Table 10.1. We will test for a relationship between gender (the IV) and death-penalty attitudes (the DV). Both of these variables are nominal, so the chi-square test of independence is the correct analysis. Note that we are setting gender as the IV and death-penalty support as the DV. There is no mathematical requirement pertaining to the placement of the variables in the rows versus the columns, but it is customary to place the IV in the rows of the table and the DV in the columns.

You have seen tables like 10.1 before—it is a contingency (or crosstabs) table just like the ones we worked with in Chapter 3! Each cell of the table displays the number of people who fall into particular classes on each variable. (Ignore the superscripts for now; we will come back to these later.) For example, 511 GSS respondents are female and oppose the death penalty, and 752 are male and favor it. We can perform a cursory assessment of the possible relationship between gender and death-penalty support by calculating row percentages for each cell of the table (we use row percentages because the IV is in the rows). Approximately 60% of women favor the death penalty $\left(\left[\frac{778}{1,289} \right] 100 = 60.36 \right)$, and 40% oppose it $\left(\left[\frac{511}{1,289} \right] 100 = 39.64 \right)$. Among men, there is 69% favorability and 31% opposition. Judging by the 8 percentage-point difference between men and women, it appears that there is a relationship between gender and attitudes about capital punishment, with women expressing less support for it. Recall from the previous chapter, however, that it would be erroneous to conclude on the basis of these percentages alone that there is a true difference between men and women in terms of their death-penalty attitudes—we have not yet ruled out the possibility that this discrepancy is the product of chance variation (i.e., sampling error). A formal hypothesis test is required before we reach a conclusion about whether these variables are related.

Table 10.1	GSS Respondents' Sex and Death-Penalty Attitudes		
	Attitude Toward Death Penalty for Persons Convicted of Murder		
Sex	Favor	Oppose	Row Marginal
Female	778[A]	511[B]	1,289
Male	752[C]	338[D]	1,090
Column Marginal	1,530	849	N = 2,379

Figure 10.1 The Chi-Square Probability Distribution, α, χ^2_{crit}, and χ^2_{obt}
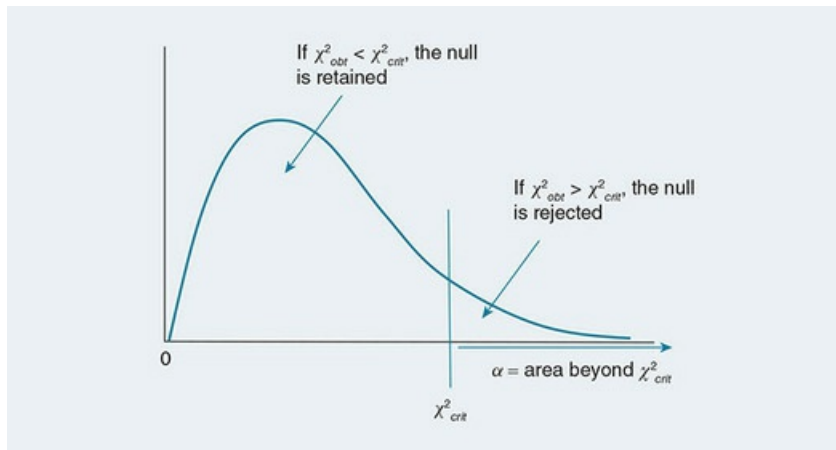
Step 1. State the null (H_0) and alternative (H_1) hypotheses .

The null hypothesis (H_0) in chi-square tests is that there is no relationship between the IV and DV. The chi-square test statistic is χ^2 (χ is the Greek letter *chi* and is pronounced "kye"). A χ^2 value of zero means that the variables are unrelated, so the null is formally written as

$$H_0: \chi^2 = 0$$

The alternative hypothesis (H_1), on the other hand, predicts that there is a relationship. The chi-square statistic gets larger as the overlap or relationship between the IV and DV increases. The chi-square statistic has its own sampling distribution, and the distribution contains only positive values; it is bounded at zero and has no negative side. This is because the statistic is a squared measure and, therefore, cannot take on negative values. As such, the alternative hypothesis is always expressed as

$$H_1: \chi^2 > 0$$

Step 2. Identify the distribution and compute the degrees of freedom .

As mentioned, the χ^2 statistic has its own theoretical probability distribution—it is called the **$\chi 2$ distribution**. The χ^2 table of critical values is located in Appendix D. Like the *t* curve, the χ^2 distribution is a family of differently shaped curves, and each curve's shape is determined by degrees of freedom (*df*) . At small *df* values,

the distribution is extremely nonnormal; as the *df* increases, the distribution gradually normalizes somewhat, but remains markedly different from a normal curve. Unlike the *t* curve, *df* for χ^2 are based not on sample size but, rather, on the size of the crosstabs table (i.e., the number of rows and columns). Looking at Table 10.1, you can see that there are two rows (*female* and *male*) and two columns (*favor* and *oppose*) . The marginals (row and column totals) are not included in the *df* calculation. The formula for degrees of freedom in a χ^2 distribution is

$$df = (r - 1)(c - 1), \qquad\qquad Formula\ 10(1)$$

where

   r = the number of rows, excluding the marginal and

   c = the number of columns, excluding the marginal.

> χ^2 **distribution**: The sampling or probability distribution for chi-square tests. This curve is nonnormal and contains only positive values. Its shape depends on the size of the crosstabs table.

Table 10.1 has two rows and two columns. Inserting these into the formula, the result is

  df = (2 − 1)(2 − 1) = (1)(1) = 1

Step 3. Identify the critical value of the test statistic and state the decision rule .

Remember in Chapter 8 when we used the α (alpha) level to find a particular value of *z* or *t* to plug into a confidence interval formula? We talked about α being the proportion of cases in the distribution that are out in the tail beyond a particular value of *z* or *t*. You learned that the critical value is the number that cuts α off the tail of the distribution. Alpha is the probability that a certain value will fall in the tail beyond the critical value. If α = .05, for instance, then the values of the test statistic that are out in the tail beyond the critical value constitute just 5% of the entire distribution. In other words, these values have a .05 or less probability of occurring if, indeed, there is no relationship between the two variables being analyzed. These values, then, represent observed outcomes that are extremely unlikely if the null hypothesis is true.

The process of finding the critical value of χ^2 (symbolized χ^2_{crit}) employs the same logic as that for finding critical values of *z* or *t*. The value of χ^2_{crit} depends on two considerations: the α level and the *df*. Alpha must be set a priori so that the critical value can be determined before the test is run. Alpha can technically be set at any number, but .05 and .01 are the most commonly used α levels in criminal justice and criminology.

For the present example, we will choose α = .05. Using Appendix D and finding the number at the intersection of α = .05 and *df* = 1, it can be seen that χ^2_{crit} = 3.841. This is the value that cuts .05 (i.e., 5%) of the cases off the tail of the χ^2 distribution. The **obtained value** of χ^2 (symbolized χ^2_{obt}) that is calculated in Step 4 must exceed the critical value in order for the null to be rejected. Figure 10.1 illustrates this concept.

> **Obtained value**: The value of the test statistic arrived at using the mathematical formulas specific to a particular test. The obtained

The decision rule is the a priori statement regarding the action you will take with respect to the null hypothesis based on the results of the statistical analysis that you are going to do in Step 4. The final product of Step 4 will be the obtained value of the test statistic. The null hypothesis will be rejected if the obtained value exceeds the critical value. If $\chi^2_{obt} > \chi^2_{crit}$, then the probability of obtaining this particular χ^2_{obt} value by chance alone is less than .05. Another way to think about it is that the probability of H_0 being true is less than .05. This is unlikely indeed! This would lead us to reject the null in favor of the alternative. The decision rule for the current test is the following: *If* $\chi^2_{obt} > 3.841$, H_0 *will be rejected*.

Step 4. Compute the obtained value of the test statistic .

Now that we know the critical value, it is time to complete the analytical portion of the hypothesis test. Step 4 will culminate in the production of the obtained value, or χ^2_{obt}. In substantive terms, χ^2_{obt} is a measure of the difference between **observed frequencies** (f_o) and **expected frequencies** (f_e). Observed frequencies are the empirical values that appear in the crosstabs table produced from the sample-derived data set. Expected frequencies are the frequencies that *would* appear if the two variables under examination were unrelated to one another. In other words, the expected frequencies are what you would see if the null hypothesis were true. The question is whether observed equals expected (indicating that the null is true and the variables are unrelated) or whether there is marked discrepancy between them (indicating that the null should be rejected because there is a relationship).

Observed frequencies: The empirical results seen in a contingency table derived from sample data. Symbolized f_o.

Expected frequencies: The theoretical results that would be seen if the null were true, that is, if the two variables were, in fact, unrelated. Symbolized f_e.

Let's talk about observed and expected frequencies a little more before moving on. Table 10.2 is a crosstabs table for two hypothetical variables that are totally unrelated to one another. The 100 cases are spread evenly across the four cells of the table. The result is that knowing which class a given case falls into on the IV offers no information about which class that case is in on the DV. For instance, if you were faced with the question, "Who is more likely to fall into category Y on the DV, someone in category A or in category B?" your answer would be that both options are equally likely. The distribution in Table 10.2 illustrates the null hypothesis in a chi-square test—the null predicts that the IV does not help us understand the DV.

Table 10.3 shows a distribution of hypothetical observed frequencies. There is a clear difference between this distribution and that in Table 10.2. In Table 10.2, it is clear that knowing what category a person is in on the IV does help predict their membership in a particular category on the DV. Someone in category A is more likely to be in category Y than in category X, whereas someone in category B is more likely to be in X than in Y. If you had a distribution like that in Table 10.2 and someone asked you to predict whether someone in category A was in X or Y, you would have a 50/50 shot at being right; in other words, you would simply have to guess. You would be wrong half the time (25 out of 50 guesses). On the other hand, if you were looking at Table 10.3 and someone asked you the same question, you would predict the person to be in Y. You would

still be wrong occasionally, but the frequency of incorrect guesses would diminish from 50% to 20% (10 out of 50 guesses).

The chi-square analysis is, therefore, premised on a comparison of the frequencies that are observed in the data and the frequencies that would be expected, theoretically, if there were no relationship between the two variables. If there is minimal difference between observed and expected, then the null will be retained. If the difference is large, the null must be rejected. We already know the observed frequencies, so the first task in Step 4 is to calculate the expected frequencies. This must be done for each cell of the crosstabs table.

Table 10.2 Hypothetical Distribution of Expected Frequencies

Independent Variable	Dependent Variable		Row Marginal
	X	Y	
A	25	25	50
B	25	25	50
Column Marginal	50	50	$N = 100$

Table 10.3 Hypothetical Distribution of Observed Frequencies

Independent Variable	Dependent Variable		Row Marginal
	X	Y	
A	10	40	50
B	40	10	50
Column Marginal	50	50	$N = 100$

The formula for an expected frequency count is

$$f_{e_i} = \frac{rm_i \cdot cm_i}{N},$$

Formula 10(2)

where

f_{e_i} = the expected frequency for cell i,

rm_i = the row marginal of cell i,

cm_i = the column marginal of cell i, and

N = the total sample size.

Since the expected frequency calculations must be done for each cell, it is a good idea to label them as a way to keep track. This is the reason why the numbers in Table 10.1 are accompanied by superscripts. The letters A through D identify the cells. Using Formula 10(2) for each cell,

$$f_{e_A} = \frac{1,289 \cdot 1,530}{2,379} = \frac{1,972,170}{2,379} = 828.99$$

$$f_{e_B} = \frac{1,289 \cdot 849}{2,379} = \frac{1,094,361}{2,379} = 460.01$$

$$f_{e_C} = \frac{1,090 \cdot 1,530}{2,379} = \frac{1,667,700}{2,379} = 701.01$$

$$f_{e_D} = \frac{1,090 \cdot 849}{2,379} = \frac{925,410}{2,379} = 388.99$$

Once the expected frequencies have been calculated, χ^2_{obt} can be computed using the formula

$$\chi^2_{obt} = \sum \frac{\left(f_{o_i} - f_{e_i}\right)^2}{f_{e_i}} \qquad\qquad Formula\ 10(3)$$

where

f_{o_i} = the observed frequency of cell i and

f_{e_i} = the expected frequency of cell i.

Formula 10(3) looks intimidating, but it is actually just a sequence of arithmetic. First, each cell's expected value will be subtracted from its observed frequency. Second, each of these new terms will be squared and divided by the expected frequency. Finally, these terms will be summed. Recall that the uppercase sigma (Σ) is a symbol directing you to sum whatever is to the right of it.

The easiest way to complete the steps for Formula 10(3) is by using a table. We will rearrange the values from Table 10.1 into a format allowing for calculation of χ^2_{obt}. Table 10.4 shows this.

The obtained value of the test statistic is found by summing the final column of the table, as such:

χ^2_{obt} = 3.14 + 5.65 + 3.71 + 6.68 = 19.18

There it is! The obtained value of the test statistic is 19.18.

Step 5. Make a decision about the null and state the substantive conclusion .

It is time to decide whether to retain or reject the null. To do this, revisit the decision rule laid out in Step 3. It was stated that the null would be rejected if the obtained value of the test statistic exceeded 3.841. The obtained value turned out to be 19.18, so $\chi^2_{obt} > \chi^2_{crit}$ and we therefore reject the null. The alternative hypothesis is what we take as being the true state of affairs. The technical term for this is **statistical significance**. A statistically significant result is one in which the obtained value exceeds the critical value and the variables are determined to be statistically related to one another.

Statistical significance: When the obtained value of a test statistic exceeds the critical value and the null is rejected.

Table 10.4 Calculating χ^2_{obt}					
Cell	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\dfrac{(f_o - f_e)^2}{f_e}$
A	778	828.99	$778 - 828.99 = -50.99$	$(-50.99)^2 = 2{,}599.98$	$\dfrac{2{,}599.98}{828.99} = 3.14$
B	511	460.01	$511 - 460.01 = 50.99$	$50.99^2 = 2{,}599.98$	$\dfrac{2{,}599.98}{460.01} = 5.65$
C	752	701.01	$752 - 701.01 = 50.99$	$50.99^2 = 2{,}599.98$	$\dfrac{2{,}599.98}{701.01} = 3.71$
D	338	388.99	$338 - 388.99 = -50.99$	$(-50.99)^2 = 2{,}599.98$	$\dfrac{2{,}599.98}{388.99} = 6.68$
	$N = 2{,}379$	$N = 2{,}379$	$\Sigma = 0$		$\Sigma = 19.18$

The final stage of hypothesis testing is to interpret the results. People who conduct statistical analyses are responsible for communicating their findings in a manner that effectively resonates with their audience, whether it is an audience comprising scholars, practitioners, the public, or the media. It is especially important when discussing statistical findings with lay audiences that clear explanations be provided about what a set of quantitative results actually *means* in a substantive, practical sense. This makes findings accessible to a wide array of audiences who might find criminological results interesting and useful.

In the context of the present example, rejecting the null leads to the conclusion that the IV and the DV are statistically related; that is, there is a statistically significant relationship between gender and death-penalty attitudes. Another way of saying this is that there is a statistically significant difference between men and women in their attitudes toward capital punishment. Note that the chi-square test does not tell us about the precise nature of that difference. Nothing in χ^2_{obt} conveys information about which gender is more supportive or more opposed than the other. This is not a big problem with two-class IVs. Referring back to the percentages reported earlier, we know that a higher percentage of women than men oppose capital punishment, so we can conclude that women are significantly less supportive of the death penalty (40% oppose) compared to men (31% oppose). We will see later, when we use IVs that have more than two classes, that we are not able to so easily identify the location of the difference.

Note, as well, the language used in the conclusion—it is phrased as an association and there is no cause-and-effect assertion being advanced. This is because the relationship that seems to be present in this bivariate analysis could actually be the result of unmeasured omitted variables that are the real driving force behind the gender differences (recall from Chapter 2 that this is the problem of spuriousness and its counterpart, the omitted variable bias). We have not, for instance, measured age, race, political beliefs, or religiosity, all of which might relate to people's beliefs about the effectiveness and morality of capital punishment. If women differ from men systematically on any of these characteristics, then the gender–attitude relationship might be spurious, meaning it is the product of another variable that has not been accounted for in the analysis. It is best to keep your language toned down and to use words like *relationship* and *association* rather than *cause* or *effect*.

Research Example 10.2 Do Victim or Offender Race Influence the Probability That a Homicide Will Be Cleared and That a Case

Will Be Tried as Death-Eligible?

A substantial amount of research has been conducted examining the impact of race on the use of the death penalty. This research shows that among murder defendants, blacks have a higher likelihood of being charged as death-eligible (i.e., the prosecutor files notice that he or she intends to seek the death penalty). The real impact of race, however, is not on the defendants' part but, rather, on the victims': People who kill whites are more likely than people who kill blacks to be prosecuted as death-eligible. Blacks accused of killing whites are the group most likely to face a death sentence, even controlling for relevant legal factors. There are open questions, though, about what happens prior to prosecutors' decisions about whether or not to seek the death penalty. In particular, it is not clear what effect police investigations and clearance rates have in shaping the composition of cases that reach prosecutors' desks. Petersen (2017) used data from Los Angeles County, California, to examine two stages in the justice-system response to homicide: clearance and the decision to seek death. The table shows the racial breakdown of victims and defendants across these categories.

Variables	All Homicides (%)	Cleared Cases (%)	Death-Eligible Cases (%)
Victim Race			
Black	35	36	32
Latino	50	48	38
White	15	16	30
Defendant Race			
Black	–	41	48
Latino	–	46	33
White	–	13	19

Source: Adapted from Table 1 in Petersen (2017).
Source: Adapted from Table 1 in Petersen (2017).

The contingency table shows no overt discrepancies for black victims (i.e., their representation in all three categories remains at roughly one-third), but murders involving Latino victims (which make up 50% of all murders) are slightly less likely to be cleared (48%) and much less likely to be prosecuted as death-eligible (38%). White victims, by contrast, make up only 15% of victims but 30% of victims in death-eligible trials. Looking at defendant race, blacks constitute 41% of the people arrested for homicide and 48% of death-eligible defendants, whites likewise are somewhat overrepresented as defendants (13% versus 19%), while Latinos are markedly underrepresented among defendants (46% compared to 33%). Of course, these relationships are bivariate and do not account for legal factors (i.e., aggravating and mitigating circumstances) that might increase or reduce a prosecutor's inclination to seek the death penalty.

To thoroughly examine the relationship between race and the probability of death-eligible charges being filed, Petersen (2017) estimated a series of predicted probabilities, which are displayed in the figure. These probabilities show the interaction between victim and defendant race and are adjusted to control for case characteristics. The findings mirror previous research showing that black and Latino defendants are more likely to face death-eligible charges when victims are white. White defendants are least likely to face death when victims are Latino and most likely when they are black, but these differences were not statistically significant.

Source: Adapted from Figure 1 in Petersen (2017).
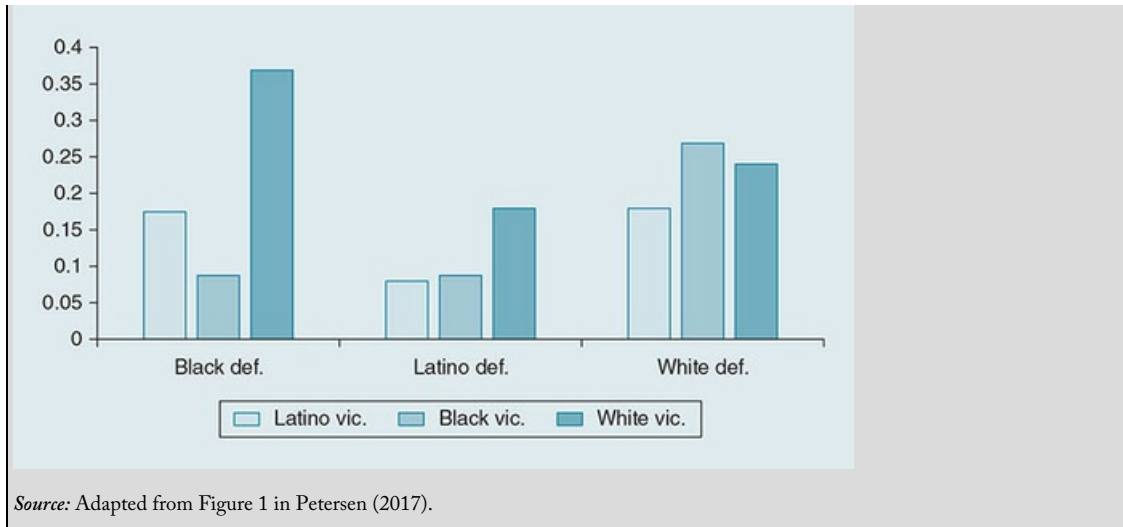
For the second example, let's use the GSS again and this time test for a relationship between education level and death-penalty attitudes. To make it interesting, we will split the data by gender and analyze males and females in two separate tests. We will start with males (see Table 10.5). Using an alpha level of .01, we will test for a relationship between education level (the IV) and death-penalty attitudes (the DV). All five steps will be used.

Step 1. State the null (H0) and alternative (H1) hypotheses .

$H_0 : \chi^2 = 0$

$H_1 : \chi^2 > 0$

Table 10.5 Male Respondents' Education Level and Death-Penalty Attitudes			
	Attitude Toward Death Penalty for Persons Convicted of Murder		
Education	Favor	Oppose	Row Marginal
High school diploma or less	512[A]	193[B]	705
Some college	45[C]	13[D]	58
Bachelor's degree or higher	195[E]	132[F]	327
Column Marginal	752	338	N = 1,090

Step 2. Identify the distribution and compute the degrees of freedom .

The distribution is χ^2 and the $df = (r-1)(c-1) = (3-1)(2-1) = (2)(1) = 2$.

Step 3. Identify the critical value of the test statistic and state the decision rule .

With $\alpha = .01$ and $df = 2$, $\chi^2_{crit} = 9.210$. The decision rule is that *if $\chi^2_{obt} > 9.210$, H_0 will be rejected.*

Step 4. Compute the obtained value of the test statistic .

First, we need to calculate the expected frequencies using Formula 10(3). The frequencies for the first three cells (labeled A, B, and C, left to right) are as follows:

$$f_{e_A} = \frac{705 \cdot 752}{1,090} = \frac{530,160}{1,090} = 486.39$$

$$f_{e_B} = \frac{705 \cdot 338}{1,090} = \frac{238,290}{1,090} = 218.61$$

$$f_{e_C} = \frac{58 \cdot 752}{1,090} = \frac{43,616}{1,090} = 40.01$$

$$f_{e_D} = \frac{58 \cdot 338}{1,090} = \frac{19,604}{1,090} = 17.99$$

$$f_{e_E} = \frac{327 \cdot 752}{1,090} = \frac{245,904}{1,090} = 225.60$$

$$f_{e_F} = \frac{327 \cdot 338}{1,090} = \frac{110,526}{1,090} = 101.4$$

Next, the computational table is used to calculate χ^2_{obt} (Table 10.6). As you can see in the summation cell in the last column, the obtained value of the test statistic is 19.74.

Before moving to Step 5, take note of a couple points about the chi-square calculation table. Both of these features will help you check your math as you work through the computation. First, the expected-frequency column always sums to the sample size. This is because we have not altered the number of cases in the sample: We have merely redistributed them throughout the table. After calculating the expected frequencies, sum them to make sure they add up to N. Second, the column created by subtracting the expected frequencies from the observed frequencies will always sum to zero (or within rounding error of it). The reason for this is, again, that no cases have been added to or removed from the sample. There are some cells that have observed frequencies that are less than expected and others where f_{o_i} is greater than f_{e_i}. In the end, these variations cancel each other out. Always sum both of these columns as you progress through a chi-square calculation. This will help you make sure your math is correct.

Cell	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\dfrac{(f_o - f_e)^2}{f_e}$
Table 10.6 Calculating χ^2_{obt}					
A	512	486.39	25.61	655.87	1.35
B	193	218.61	−25.61	655.87	3.00
C	45	40.01	4.99	24.90	.62
D	13	17.99	−4.99	24.90	1.38
E	195	225.60	−30.60	936.36	4.15
F	132	101.40	30.60	936.36	9.23
	$N = 1{,}090$	$N = 1{,}090$	$\Sigma = .00$		$\Sigma = 19.74$

Step 5. Make a decision about the null and state the substantive conclusion .

The decision rule stated that the null would be rejected if the obtained value exceeded 9.210. Since χ^2_{obt} ended up being greater than the critical value (i.e., 19.74 > 9.210), the null is rejected. There is a statistically significant relationship between education and death-penalty attitudes among male respondents. Calculating row percentages from the data in Table 10.5 shows that approximately 27% of men with high school diplomas or less oppose the death penalty, roughly 22% with some college (no degree) are in opposition, and 40% with a bachelor's degree or higher do not support it. It seems that men with college educations that include at least a bachelor's degree stand out from the other two educational groups in their level of opposition to capital punishment. We are not able to say with certainty, however, whether all three groups are statistically significantly different from the others or whether only one of them stands apart. You can roughly estimate differences using row percentages, but you have to be cautious in your interpretation. The chi-square test tells you only that at least one group is statistically significantly different from at least one other group.

Let's repeat the same analysis for female respondents. Again, we will set alpha at .01 and proceed through the five steps. The data are in Table 10.7.

Step 1. State the null (H0) and alternative (H1) hypotheses .

$H_0: \chi^2 = 0$

$H_1: \chi^2 > 0$

Step 2. Identify the distribution and compute the degrees of freedom .

The distribution is χ^2 and $df = (3 - 1)(2 - 1) = 2$.

Step 3. Identify the critical value of the test statistic and state the decision rule .

With $\alpha = .01$ and $df = 2$, $\chi^2_{crit} = 9.210$. The decision rule is that *if $\chi^2_{obt} > 9.210$, H_0 will be rejected.*

	Attitude Toward Death Penalty for Persons Convicted of Murder		
Education	Favor	Oppose	Row Marginal
High school diploma or less	479[A]	316[B]	795
Some college	85[C]	30[D]	115
Bachelor's degree or higher	214[E]	165[F]	379
Column Marginal	778	511	$N = 1,289$

Table 10.7 Female Respondents' Education Level and Death-Penalty Attitudes

Step 4. Compute the obtained value of the test statistic .

Cell	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\dfrac{(f_o - f_e)^2}{f_e}$
A	479	479.84	−.84	.71	.001
B	316	315.16	.84	.71	.002
C	85	69.41	15.59	243.05	3.50
D	30	45.59	−15.59	243.05	5.33
E	214	228.75	−14.75	217.56	.95
F	165	150.25	14.75	217.56	1.45
	$N = 1{,}289$	$N = 1{,}289$	$\Sigma = .00$		$\Sigma = 11.23$

Table 10.8 Calculating χ^2_{obt}

Step 5. Make a decision about the null and state the substantive conclusion .

The decision rule stated that the null would be rejected if the obtained value exceeded 9.210. Since χ^2_{obt} = 11.23, the null is rejected. There is a statistically significant relationship between education and death-penalty attitudes among female respondents; it appears that women's likelihood of favoring or opposing capital punishment changes with their education level. Another way to phrase this is that there are significant differences between women of varying levels of education. As we did with male respondents, we can use row percentages to gain a sense of the pattern. Opposition to the death penalty is approximately 40%, 26%, and 44% among women with high school diploma or less, some college, or a bachelor's degree or higher, respectively. This is similar to the pattern of opposition seen among men in that those with some college were the most supportive of capital punishment and those with a college degree were the least supportive, but among women the group with a high school diploma or less were nearly as likely as those with college degrees to oppose this penalty.

Learning Check 10.2

✓

One criticism of the chi-square test for independence is that this statistic is sensitive to sample size. The problem lies in the way that χ^2_{obt} is calculated. Sample size can cause a test statistic to be significant or not significant, apart from the actual distribution of observed values. For instance, a crosstabs table with a fairly even distribution of scores might yield a statistically significant χ^2_{obt} if N is large. Similarly, a distribution that looks decidedly uneven (i.e., where there is an apparent relationship between the IV and the DV) can produce nonsignificant χ^2_{obt} if N is small. To see this for yourself, recalculate χ^2_{obt} using the death-penalty and gender data but with a sample size of 88 instead of 2,379. Make a decision about the null hypothesis, recalling that $\chi^2_{crit} = 3.841$.

Sex	Death Penalty		Row Marginal
	Favor	Oppose	
Female	30[A]	19[B]	49
Male	26[C]	13[D]	39
Column Marginal	56	32	$N = 88$

Are you surprised by the results? This demonstrates the importance of being cautious when you interpret statistical significance. Do not leap to hasty conclusions; be aware that there are factors (such as sample size) that can impact the results of a statistical test irrespective of the relationship between the IVs and the DVs.

Measures of Association

The chi-square test alerts you when there is a statistically significant relationship between two variables, but it is silent as to the strength or magnitude of that relationship. We know from the previous two examples, for instance, that gender and education are related to attitudes toward capital punishment, but we do not know the magnitudes of these associations: They could be strong, moderate, or weak. This question is an important one because a trivial relationship—even if statistically significant in a technical sense—is not of much substantive or practical importance. Robust relationships are more meaningful. To illustrate this, suppose an evaluation of a gang-prevention program for youth was declared a success after researchers found a statistically significant difference in gang membership rates among youth who did and did not participate in the program. Digging deeper, however, you learn that 9% of the youth who went through the program ended up joining gangs, compared to 12% of those who did not participate. While any program that keeps kids out of gangs is laudable, a reduction of three percentage points can hardly be considered a resounding success. We would probably want to continue searching for a more effective way to prevent gang involvement. **Measures of association** offer insight into the magnitude of the differences between groups so that we can figure out how strong the overlap is.

> **Measures of association**: Procedures for determining the strength or magnitude of a relationship after a chi-square test has revealed a statistically significant association between two variables.

There are several measures of association, and this chapter covers four of them. The level of measurement of the IV and the DV dictate which measures are appropriate for a given analysis. Measures of association are computed only when the null hypothesis has been rejected—if the null is not rejected and you conclude that there is no relationship between the IV and the DV, then it makes no sense to go on and try to interpret an association you just said does not exist. The following discussion will introduce four tests, and the next section will show you how to use SPSS to compute χ^2_{obt} and accompanying measures of association.

Cramer's V can be used when both of the variables are nominal or when one is ordinal and the other is nominal. It is symmetric, meaning that it always takes on the same value regardless of which variable is posited as the independent and which the dependent. This statistic ranges from 0.00 to 1.00, with higher values indicative of stronger relationships and values closer to 0.00 suggestive of weaker associations. Cramer's V is computed as

$$V = \sqrt{\frac{\chi^2_{obt}}{N \cdot m}} \qquad\qquad Formula\ 10(4)$$

where

χ^2_{obt} = the obtained value of the test statistic,

N = the total sample size, and

m = the smaller of either $(r - 1)$ or $(c - 1)$.

In the first example we saw in this chapter, where we found a statistically significant relationship between gender and death-penalty attitudes, χ^2_{obt} = 19.18, N = 2,379, and there were two rows and two columns, so m = 2 − 1 = 1. Cramer's V is thus

$$V = \sqrt{\frac{19.18}{2,379 \cdot 1}} = \sqrt{.01} = .10$$

This value of V suggests a weak relationship. This demonstrates how statistical significance alone is not indicative of genuine importance or meaning—a relationship might be significant in a technical sense but still very *in*significant in practical terms. This is due in no small part to the chi-square test's sensitivity to sample size, as discussed earlier. Think back to the percentages we calculated for this table. Approximately 60% of women and 69% of men favored the death penalty. This is a difference, to be sure, but it is not striking. If any given respondent were randomly selected out of this sample, there would be a roughly two-thirds likelihood that the person would support capital punishment, irrespective of her or his gender. It is wise, then, to be cautious in interpreting statistically significant results—*statistical* significance does not always translate into *practical* significance.

When both variables under examination are nominal, **lambda** is an option. Like Cramer's V, lambda ranges from 0.00 to 1.00. Unlike Cramer's V, lambda is asymmetric, meaning that it requires that one of the variables be clearly identified as the independent and the other as the dependent. This is because lambda is a proportionate reduction in error measure.

Proportionate reduction in error (PRE) refers to the extent to which knowing a person's or object's placement on an IV helps predict that person's or object's classification on the dependent measure. Referring back to Table 10.1, if you were trying to predict a given individual's attitude toward capital punishment and the only piece of information you had was the frequency distribution of this DV (i.e., you knew that 1,530 people in the sample support capital punishment and 849 oppose it), then your best bet would be to guess the modal category (*mode* = support) because this would produce the fewest prediction errors. There would, though, be a substantial number of these errors—849, to be exact!

Now, suppose that you know a given person's gender or education level, both of which we found to be significantly related to capital punishment attitudes. The next logical inquiry is the extent to which this knowledge improves our accuracy when predicting whether that person opposes or favors the death penalty. In other words, know that we would make 849 errors if we simply guessed the mode for each person in the sample, and now we want to know how many fewer mistakes we would make if we knew each person's gender

or education level. This is the idea behind PRE measures like lambda. Let's do an example using the relationship between education and death penalty attitudes among women.

Lambda is symbolized as λ (the Greek lowercase letter *lambda*) and is calculated as

$$\lambda = \frac{E_1 - E_2}{E_1}, \qquad\qquad Formula\ 10(5)$$

where

$$E_1 = N_{total} - N_{DV\,mode}\ and$$

$$E_2 = \sum_{across\,IV} \left(N_{category\,total} - N_{category\,mode} \right).$$

This equation and its different components looks strange, but, basically, E_1 represents the number of prediction errors made when the IV is ignored (i.e., predictions based entirely on the mode of the DV), and E_2 reflects the number of errors made when the IV is taken into account. Using the education and death-penalty data from Table 10.7, we can first calculate E_1 and E_2:

$E_1 = 1,289 - 778 = 511,$

$E_2 = (795 - 479) + (115 - 85) + (379 - 214) = 511$

and lambda is

$$\lambda = \frac{511 - 494}{511} = \frac{17}{511} = .03$$

Lambda is most easily interpreted by transforming it to a percentage. A lambda of zero shows us that knowing women's education levels does not reduce prediction errors. This makes sense, as you can see in Table 10.7 that "favor" is the mode across all three IV categories. This lack of variation accounts for the overall zero impact of the IV on the DV. Again, we see a statistically significant but substantively trivial association between variables. This makes sense, because we would not expect a single personal characteristic (like education) to have an enormous influence on someone's attitudes. As noted previously, we have not measured women's religiosity, age, social and political views, and other demographic and background factors that are closely linked with people's opinions about capital punishment.

There is a third measure for nominal data that bears brief mention, and that is **phi**. Phi can only be used on 2 × 2 tables (i.e., two rows and two columns) with nominal variables. It is calculated and interpreted just like Cramer's *V* with the exception that phi does not account for the number of rows or columns in the crosstabs table: since it can only be applied to 2 × 2 tables, *m* will always be equal to 1.00. For 2 × 2 tables, Cramer's *V* is identical to phi, but since Cramer's *V* can be used for tables of any size, it is more useful than phi is.

When both variables are ordinal or when one is ordinal and the other is dichotomous (i.e., has two classes), **Goodman and Kruskal's gamma** is an option. Gamma is a PRE measure but is symmetric—unlike lambda—and ranges from −1.00 to +1.00, with zero meaning no relationship, −1.00 indicating a perfect negative relationship (as one variable increases, the other decreases), and 1.00 representing a perfect positive relationship (as one increases, so does the other). Generally speaking, gamma values between 0 and ±.19 are considered weak, between ±.20 and ±.39 moderate, ±.40 to ±.59 strong, and ±.60 to ±1.00 very strong.

Two other measures available when both variables are ordinal are **Kendall's tau$_b$** and **Kendall's tau$_c$**. Both are symmetric. Tau$_b$ is used when the crosstabs table has an equal number of rows and columns, and tau$_c$ is used when they are unequal. Both tau statistics range from −1.00 to +1.00. They measure the extent to which the order of the observations in the IV match the order in the DV; in other words, as cases increase in value on the IV, what happens to their scores on the DV? If their scores on the dependent measure decrease, tau will be negative; if they increase, tau will be positive; and if they do not display a clear pattern (i.e., the two variables have very little dependency), tau will be close to zero. Similar to the tau measures is **Somers' *d***. This measure of association is asymmetric and used when both variables are ordinal. Its range and interpretation mirror those of tau . The calculations of gamma, tau, and *d* are complicated, so we will refrain from doing them by hand and will instead use SPSS to generate these values.

None of the measures of association discussed here is perfect; each has limitations and weaknesses. The best strategy is to examine two or more measures for each analysis and use them to gain a comprehensive picture of the strength of the association. There will likely be variation among them, but the differences should not be wild, and all measures should lean in a particular direction. If they are all weak or are all strong, then you can safely arrive at a conclusion about the level of dependency between the two variables.

SPSS

The SPSS program can be used to generate χ^2_{obt}, determine statistical significance, and produce measures of association. The chi-square analysis is found via the sequence *Analyze → Descriptive Statistics → Crosstabs*. Let us first consider the gender and capital punishment example from earlier in the chapter. Figure 10.2 shows the dialog boxes involved in running this analysis in SPSS. Note that you must check the box labeled *Chi-square* in order to get a chi-square analysis; if you do not check this box, SPSS will merely give you a crosstabs table. This box is opened by clicking *Statistics* in the crosstabs window. By default, SPSS provides only observed frequencies in the crosstabs table. If you want expected frequencies or percentages (row or column), you can go into *Cells* and request them. Since both of these variables are nominal, lambda and Cramer's *V* are the appropriate measures of association. Figure 10.3 shows the output for the chi-square test, and Figure 10.4 displays the measures of association.

The obtained value of the χ^2_{obt} statistic is located on the line labeled *Pearson Chi-Square*. You can see in Figure 10.3 that χ^2_{obt} = 19.182, which is identical to the value we obtained by hand. The output also tells you whether or not the null should be rejected, but it does so in a way that we have not seen before. The SPSS program gives you what is called a *p* **value**. The *p* value tells you the exact probability of the obtained value of the test statistic: The smaller *p* is, the more unlikely the χ^2_{obt} is if the null is true, and, therefore, the probability that the null is, indeed, correct. The *p* value in SPSS χ^2 output is the number located at the intersection of the *Asymp. Sig. (2-sided)* column and the *Pearson Chi-Square* row. Here, *p* = .000. What you do is compare *p* to α. If *p* is less than α, it means that the obtained value of the test statistic exceeded the critical value, and the null is rejected; if *p* is greater than α, the null is retained. Since in this problem α was set at .05, the null hypothesis is rejected because .000 < .05. There is a statistically significant relationship between gender and death-penalty attitudes. Of course, as we have seen, rejection of the null hypothesis is only part of the story because the χ^2 statistic does not offer information about the magnitude or strength of the relationship between the variables. For this, we turn to measures of association.

> *p* **value**: In SPSS output, the probability associated with the obtained value of the test statistic. When *p* < α, the null hypothesis is rejected.

Figure 10.2 Running a Chi-Square Test and Measures of Association in SPSS
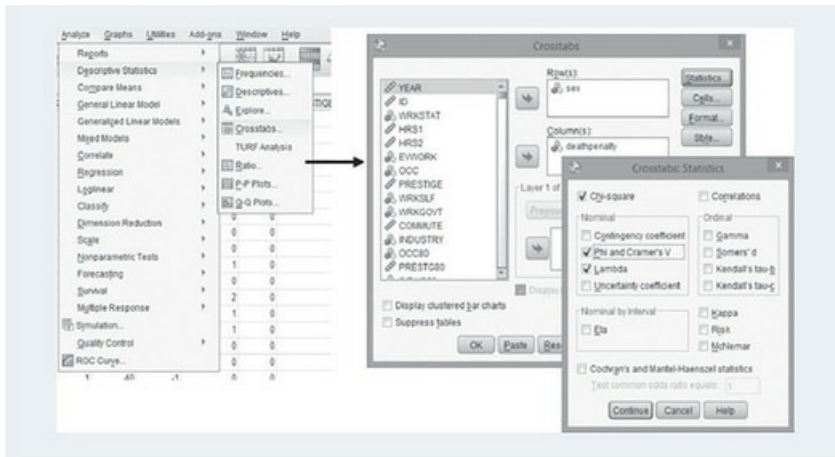
Figure 10.3 Chi-Square Output

Respondent sex* Do you favor or oppose death penalty for murder? Crosstabulation

| | | | Do you favor or oppose death penalty for murder? | | |
			Favor	Oppose	Total
Respondent sex	Female	Count	778	511	1289
		Expected Count	829.0	460.0	1289.0
	Male	Count	752	338	1090
		Expected Count	701.0	389.0	1090.0
Total		Count	1530	849	2379
		Expected Count	1530.0	849.0	2379.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	19.182[a]	1	.000		
Continuity Correction[b]	18.808	1	.000		
Likelihood Ratio	19.276	1	.000		
Fisher's Exact Test				.000	.000
Liner-by-Linear Association	19.174	1	.000		
N of Valid Cases	2379				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 388.99.
b. Computed only for a 2 x 2 table

Figure 10.4 Measures of Association

Directional Measures

				Value	Asymp. Std. Error[a]	Approx. T	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric		.000	.000	[b]	[b]
				.000	.000	[b]	[b]
				.000	.000	[b]	[b]

a. Not assuming the null hypothesis.
b. Cannot be computed because the asymptotic standard error equals zero.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	−.090	.000
	Cramer's V	.090	.000
N of Valid Cases		2379	

Judging by both Cramer's V and lambda, this relationship is very weak. We actually already knew this because we calculated V by hand and arrived at .10, which is within rounding error of the .09 SPSS produces. Lambda is zero, which means that knowing people's gender does not reduce the number of errors made in predicting their death-penalty attitudes. As noted earlier, using multiple tests of association helps provide confidence in the conclusion that the association between these two variables, while statistically significant, is tenuous in a

substantive sense. In other words, knowing someone's gender only marginally improves our ability to predict that person's attitudes about capital punishment.

Chapter Summary

This chapter introduced the chi-square test of independence, which is the hypothesis-testing procedure appropriate when both of the variables under examination are categorical. The key elements of the χ^2 test are observed frequencies and expected frequencies. Observed frequencies are the empirical results seen in the sample, and expected frequencies are those that would appear if the null hypothesis were true and the two variables unrelated. The obtained value of chi-square is a measure of the difference between observed and expected, and comparing χ^2_{obt} to χ^2_{crit} for a set α level allows for a determination of whether the null hypothesis should be retained or rejected.

When the null is retained (i.e., when $\chi^2_{obt} < \chi^2_{crit}$), the substantive conclusion is that the two variables are not related. When the null is rejected (when $\chi^2_{obt} > \chi^2_{crit}$), the conclusion is that there is a relationship between them. *Statistical* significance, though, is only a necessary and not a sufficient condition for *practical* significance. The chi-square statistic does not offer information about the strength of a relationship and how substantively meaningful this association is.

For this, measures of association are turned to when the null has been rejected. Cramer's V, lambda, Goodman and Kruskal's gamma, Kendall's tau$_a$ and tau$_b$, and Somers' d are appropriate in any given situation depending on the variables' levels of measurement and the size of the crosstabs table. SPSS can be used to obtain chi-square tests, p values for determining statistical significance, and measures of association. When $p < \alpha$, the null is rejected, and when $p > \alpha$, it is retained. You should always generate measures of association when you run χ^2 tests yourself, and you should always expect them from other people who run these analyses and present you with the results. Statistical significance is important, but the magnitude of the relationship tells you how meaningful the association is in practical terms.

Thinking Critically

1. Suppose you read a report claiming that children raised in families with low socioeconomic status are less likely to go to college compared to children raised in families with middle and upper income levels. The news story cites college participation rates of 20%, 35%, and 60% among low, middle, and upper socioeconomic statuses, respectively, and explains these differences as meaning that children raised in poor families are less intelligent or less ambitious than those from better-off families. Do you trust this conclusion? Why or why not? If you do not, what more do you need to know about these data before you can make a decision about the findings, and what they mean for the relationship between family income and children's college attendance?

2. Two researchers are arguing about statistical findings. One of them believes that any statistically significant result is important, irrespective of the magnitude of the association between the IVs and the DVs. The other one contends that statistical significance is meaningless if the association is weak. Who is correct? Explain your answer. Offer a hypothetical or real-world example to illustrate your point.

Review Problems

1. A researcher wants to test for a relationship between the number of citizen complaints that a police officer receives and whether that officer commits serious misconduct. He gathers a sample of officers and records the number of complaints that have been lodged against them ($0 - 2$, $3 - 5$, $6+$) and whether they have ever been written up for misconduct (yes or no). Can he use a chi-square to test for a relationship between these two variables? Why or why not?

2. A researcher wishes to test for a relationship between age and criminal offending. She gathers a sample and for each person, she collects his or her age (in years) and whether that person has ever committed a crime (yes or no). Can she use a chi-square to test for a relationship between these two variables? Why or why not?

3. A researcher is interested in finding out whether people who drive vehicles that are in bad condition are more likely than those driving better cars to get pulled over by police. She collects a sample and codes each person's vehicle's condition (good,

fair, poor) and the number of times that person has been pulled over (measured by respondents writing in the correct number). Can she use a chi-square to test for a relationship between these two variables? Why or why not?

4. A researcher is studying the effectiveness of an in-prison treatment program in reducing post-release recidivism. He gathers a sample of recently released prisoners and records, for each person, whether he or she participated in a treatment program while incarcerated (yes or no) and whether that person committed a new crime within 6 months of release (yes or no). Can he use a chi-square to test for a relationship between these two variables? Why or why not?

5. Is a criminal defendant's gender related to the type of sentence she or he receives? A researcher collects data on defendants' gender (male or female) and sentence (jail, probation, fine).
    1. Which of these variables is the IV, and which is the DV?
    2. Identify each variable's level of measurement.
    3. How many rows and columns would the crosstabs table have?

6. Is the value of the goods stolen during a burglary related to the likelihood that the offender will be arrested? A researcher collects data on the value of stolen goods ($299 or less, $300–$599, $600 and more) and on whether the police arrested someone for the offense (yes or no).
    1. Which of these variables is the IV, and which is the DV?
    2. Identify each variable's level of measurement.
    3. How many rows and columns would the crosstabs table have?

7. Is the crime for which a person is convicted related to the length of the prison sentence she or he receives? A research gathers data on crime type (violent, property, drug) and sentence length (18 months or less, 19–30 months, 31 or more months).
    1. Which of these variables is the IV, and which is the DV?
    2. Identify each variable's level of measurement.
    3. How many rows and columns would the crosstabs table have?

8. Is a victim's gender related to whether or not the offender will be convicted for the crime? A researcher collects data on victim gender (male or female) and whether the offender was convicted (yes or no).
    1. Which of these variables is the IV, and which is the DV?
    2. Identify each variable's level of measurement.
    3. How many rows and columns would the crosstabs table have?

9. It might be expected that jails that offer alcohol treatment programs to inmates also offer psychiatric counseling services, since alcohol abuse is frequently a symptom of an underlying psychological problem. The following table displays data from a random sample from the Census of Jails (COJ). With an alpha level of .01, conduct a five-step chi-square hypothesis test to determine whether the two variables are independent.

| Alcohol Treatment | Psychiatric Counseling | | Row Marginal |
	Yes	No	
Yes	54^A	13^B	67
No	9^C	74^D	83
Column marginal	63	87	$N = 150$

10. Is there an association between the circumstances surrounding a violent altercation that results in a shooting and the type of firearm used? The Firearm Injury Surveillance Study (FISS) records whether the shooting arose out of a fight and the type of firearm used to cause the injury (here, handguns vs. rifles and shotguns). With an alpha of .01, conduct a five-step hypothesis test to determine if the variables are independent.

| Fight Involved | Firearm Type | | Row marginal |
	Handgun	Long gun	
Yes	158^A	11^B	169
No	309^C	133^D	442
Column marginal	467	144	$N = 611$

11. Continuing with an examination of gunshots resulting from fights, we can analyze FISS data to determine whether there is a relationship between victims' genders and whether their injuries were the result of fights. With an alpha of .05, conduct a five-step hypothesis test to determine if the variables are independent.

| Gender | Fight Involved | | Row marginal |
	Yes	No	
Male	320ᴬ	1,018ᴮ	1,338
Female	62ᶜ	224ᴰ	286
Column marginal	382	1,242	N = 1,624

12. In the chapter, we saw that there was a statistically significant difference between men and women in terms of their attitudes about capital punishment. We can extend that line of inquiry and find out whether there is a gender difference in general attitudes about crime and punishment. The GSS asks respondents whether they think courts are too harsh, about right, or not harsh enough in dealing with criminal offenders. The following table contains the data. With an alpha level of .05, conduct a five-step chi-square hypothesis test to determine whether the two variables are independent.

| Sex | Attitude Toward Courts | | | Row marginal |
	Too harsh	About right	Not harsh enough	
Male	39ᴬ	42ᴮ	137ᶜ	218
Female	37ᴰ	44ᴱ	164ᶠ	245
Column marginal	76	86	301	N = 463

13. Do men and women differ on their attitudes toward drug laws? The GSS asks respondents to report whether they think marijuana should be legalized. The following table shows the frequencies, by gender, among black respondents. With an alpha level of .05, conduct a five-step chi-square hypothesis test to determine whether the two variables are independent.

| Sex | Should Marijuana Be Legalized? | | Row marginal |
	Yes	No	
Male	50ᴬ	24ᴮ	74
Female	82ᶜ	54ᴰ	136
Column marginal	132	78	N = 210

14. The following table shows the support for marijuana legalization, by race, among male respondents. With an alpha level of .05, conduct a five-step chi-square hypothesis test to determine whether the two variables are independent.

| Race | Should Marijuana Be Legalized? | | Row marginal |
	Yes	No	
Black	53ᴬ	27ᴮ	80
White	336ᶜ	205ᴰ	541
Other	38ᴱ	45ᶠ	83
Column marginal	427	277	N = 704

15. There is some concern that people of lower-income statuses are more likely to come in contact with the police as compared to higher-income individuals. The following table contains Police–Public Contact Survey (PPCS) data on income and police contacts among respondents who were 21 years of age or younger. With an alpha of .01, conduct a five-step hypothesis test

to determine if the variables are independent.

Annual Income	Number of Contacts			Row marginal
	0–2	3–5	6+	
Less than $20,000	315[A]	42[B]	15[C]	372
$20,000–$49,999	196[D]	31[E]	9[F]	236
$50,000 or more	266[G]	32[H]	9[I]	307
Column marginal	777	105	33	$N = 915$

16. One criticism of racial profiling studies is that people's driving frequency is often unaccounted for. This is a problem because, all else being equal, people who spend more time on the road are more likely to get pulled over eventually. The following table contains PPCS data narrowed down to black male respondents. The variables measure driving frequency and whether these respondents had been stopped by police for traffic offenses within the past 12 months. With an alpha of .01, conduct a five-step hypothesis test to determine if the variables are independent.

Driving Frequency	Experienced a Traffic Stop?		Row marginal
	Yes	No	
Almost every day	214[A]	946[B]	1,160
Often	32[C]	238[D]	270
Rarely	6[E]	360[F]	366
Column marginal	252	1,544	$N = 1,796$

17. The companion website (www.sagepub.com/gau) contains the SPSS data file *GSS for Chapter 10.sav*. This is a portion of the 2014 GSS. Two of the variables in this file are *race* and *courts* , which capture respondents' race and their attitudes about courts' harshness, respectively. Run a chi-square analysis to determine if people's attitudes (the DV) vary by race (the IV). Then do the following.
    1. Identify the obtained value of the chi-square statistic.
    2. Make a decision about whether you would reject the null hypothesis of independence at an alpha level of .05 *and* explain how you arrived at that decision.
    3. State the conclusion that you draw from the results of each of these analyses in terms of whether there is a relationship between the two variables.
    4. *If you rejected the null hypothesis* , interpret row percentages and applicable measures of association. How strong is the relationship? Would you say that this is a substantively meaningful relationship?

18. Using *GSS for Chapter 10.sav* (www.sagepub.com/gau), run a chi-square analysis to determine whether there is a relationship between the candidate people voted for in the 2012 presidential election (Barack Obama or Mitt Romney) and their opinions about how well elected officials are doing at controlling crime rates than do the following:
    1. Identify the obtained value of the chi-square statistic.
    2. Make a decision about whether you would reject the null hypothesis of independence at an alpha level of .01 *and* explain how you arrived at that decision.
    3. State the conclusion that you draw from the results of each of these analyses in terms of whether there is a relationship between the two variables.
    4. *If you rejected the null hypothesis* , interpret row percentages and applicable measures of association. How strong is the relationship? Would you say that this is a substantively meaningful relationship?

19. A consistent finding in research on police–community relations is that there are racial differences in attitudes toward police. Although all racial groups express positive views of police overall, the level of support is highest for whites and tends to dwindle among persons of color. The companion website (www.sagepub.com/gau) contains variables from the PPCS (*PPCS*

for Chapter 10.sav) . The sample has been narrowed to males who were stopped by the police while driving a car and were issued a traffic ticket. There are three variables in this data set: *race* , *income* , and *legitimacy*. The legitimacy variable measures whether respondents believed that the officer who pulled them over had a credible reason for doing so. Use SPSS to run a chi-square analysis to determine whether legitimacy judgments (the DV) differ by race (the IV). Based on the variables' level of measurement, select appropriate measures of association. Then do the following:

1. Identify the obtained value of the chi-square statistic.
2. Make a decision about whether you would reject the null hypothesis of independence at an alpha level of .01 *and* explain how you arrived at that decision.
3. State the conclusion that you draw from the results of each of these analyses as to whether or not there is a difference between private and public prisons in terms of offering vocational training.
4. *If you rejected the null hypothesis* , interpret row percentages and applicable measures of association. How strong is the relationship? Would you say that this is a substantively meaningful relationship?

20. Using the *PPCS for Chapter 10.sav* file again (www.sagepub.com/gau), run a chi-square test to determine whether respondents' perceptions of stop legitimacy (the DV) vary across income levels (the IV). Based on the variables' level of measurement, select appropriate measures of association. Then do the following:

1. Identify the obtained value of the chi-square statistic.
2. Make a decision about whether you would reject the null hypothesis of independence at an alpha level of .01 *and* explain how you arrived at that decision.
3. State the conclusion that you draw from the results of each of these analyses in terms of whether there is a difference between private and public prisons in terms of offering vocational training.
4. *If you rejected the null hypothesis* , interpret row percentages and applicable measures of association. How strong is the relationship? Would you say that this is a substantively meaningful relationship?

Key Terms

Glossary of Symbols and Abbreviations Introduced in This Chapter

χ^2	The Greek letter *chi* squared; a symbol for the test of independence and its associated sampling distribution
χ^2_{crit}	The critical value of the chi-square statistic
χ^2_{obt}	The obtained value of the chi-square statistic
f_o	Observed frequencies
f_e	Expected frequencies
p value	The obtained probability on SPSS output that is compared to the alpha level in order to make a decision about the null hypothesis

Chapter 11 Hypothesis Testing With Two Population Means or Proportions

There are many situations in which criminal justice and criminology researchers work with categorical independent variables (IVs) and continuous dependent variables (DVs). They might want to know, for instance, whether male and female police officers differ in the number of arrests they make, whether criminal offenders who have children are given shorter jail sentences compared to those who do not, or whether prisoners who successfully complete a psychological rehabilitation program have a significant reduction in antisocial thinking. Research Example 11.1 illustrates another instance of a categorical IV and a continuous DV.

In Research Example 11.1, Wright, Pratt, and DeLisi (2008) had two groups (multiple homicide offenders [MHOs] and single homicide offenders [SHOs]), each with its own mean and standard deviation; the goal was to find out whether the groups' means differ significantly from one another. A significant difference would indicate that MHOs and SHOs do indeed differ in the variety of crimes they commit, whereas rough equivalency in the means would imply that these two types of homicide offenders are equally diverse in offending. What should the researchers do to find out whether MHOs and SHOs have significantly different diversity indices?

The answer is that they should conduct a two-population test for differences between means, or what is commonly referred to as a *t* **test**. As you probably figured out, these tests rely on the *t* distribution. We will also cover two-population tests for differences between proportions, which are conceptually similar to *t* tests but employ the *z* distribution.

t **test**: The test used with a two-class, categorical independent variable and a continuous dependent variable.

Tests for differences between two means or two proportions are appropriate when the IV is categorical with two classes or groups and the DV is expressed as one mean or proportion per group. Examples of two-class, categorical IVs include gender (*male* or *female*) and political orientation (*liberal* or *conservative*) . Examples of DVs appropriate for two-population tests are the mean number of times people in a sample report that they drove while intoxicated, or the proportion of people in the sample who have been arrested for driving while under the influence of alcohol. The diversity index that Wright et al. (2008) used is continuous, which is why the researchers computed a mean and a standard deviation, and the reason that a *t* test is the appropriate

analytic strategy. In the review problems at the end of the chapter, you will conduct a *t* test to find out whether MHOs and SHOs differ significantly in offending diversity.

Research Example 11.1 Do Multiple Homicide Offenders Specialize in Killing?

Serial killers and mass murderers capture the public's curiosity and imagination. Who can resist some voyeuristic gawking at a killer who periodically snuffs out innocent victims while outwardly appearing to be a regular guy or at the tormented soul whose troubled life ultimately explodes in an episode of wanton slaughter? Popular portrayals of multiple homicide offenders (MHOs) lend the impression that these killers are fundamentally different from more ordinary criminals and from single homicide offenders (SHOs) in that they only commit homicide and lead otherwise crime-free lives. But is this popular conception true?

Wright, Pratt, and DeLisi (2008) decided to find out. They constructed an index measuring diversity of offending within a sample of homicide offenders. This index captured the extent to which homicide offenders committed only homicide and no other crimes versus the extent to which they engaged in various types of illegal acts. The researchers divided the sample into MHOs and SHOs and calculated each group's mean and standard deviation on the diversity index. They found the statistics located in the table.

	MHOs		SHOs	
	$\bar{x}$	s	$\bar{x}$	s
Diversity index	.36	.32	.37	.33

Source: Adapted from Table 1 in Wright et al. (2008).
Source: Adapted from Table 1 in Wright et al. (2008).

Two-Population Tests for Differences Between Means: *t* Tests

There are many situations in which people working in criminal justice and criminology would want to test for differences between two means: Someone might be interested in finding out whether offenders who are sentenced to prison receive significantly different mean sentence lengths depending on whether they are male or female. A municipal police department might implement an innovative new policing strategy and want to know whether the program significantly reduced mean crime rates in the city. These types of studies require *t* tests.

In Chapter 7, you learned the difference between sample, sampling, and population distributions. Recall that sampling distributions are theoretical curves created when multiple or infinite samples are drawn from a single population and a statistic is computed and plotted for each sample. Over time, with repeated drawing, calculating, plotting, throwing back, and drawing again, the distribution of sample statistics builds up and, if the size of each sample is large (meaning $N \geq 100$), the statistics form a normal curve. There are also sampling distributions for differences between means. See Figure 11.1. You have seen in prior chapters how sampling distributions' midpoint is the population mean (symbolized μ) . Sampling distributions for differences between means are similar in that they center on the true population difference, $\mu_1 - \mu_2$. A sampling distribution of differences between means is created by pulling infinite *pairs* of samples, rather than single samples. Imagine drawing two samples, computing both means, subtracting one mean from the other to form a difference score, and then plotting that difference score. Over time, the difference scores build up. If $N \geq$ 100, the sampling distribution of differences between means is normal; if $N \leq 99$, then the distribution is more like "normalish" because it tends to be wide and flat. The *t* distribution, being flexible and able to accommodate various sample sizes, is the probability distribution of choice for tests of differences between means.

There are two general types of *t* tests, one for **independent samples** and one for **dependent samples**. The difference between them pertains to the method used to select the two samples under examination. In independent sampling designs, the selection of cases into one sample in no way affects, or is affected by, the selection of cases into the other sample. If a researcher is interested in the length of prison sentences received by male and female defendants, then that researcher would draw a sample of females and a sample of males. A researcher investigating the effects of judicial selection type on judges' sentencing decisions might draw a sample of judges who were elected and a sample of those who were appointed to their posts. In neither of these instances does the selection of one person into one sample have bearing on the selection of another into the other sample. They are independent because they have no influence on each other.

> **Independent samples**: Pairs of samples in which the selection of people or objects into one sample in no way affected, or was affected by, the selection of people or objects into the other sample.
>
> **Dependent samples**: Pairs of samples in which the selection of people or objects into one sample directly affected, or was directly affected by, the selection of people or objects into the other sample. The most common types are matched pairs and repeated measures.

In dependent-samples designs, by contrast, the two samples are related to each other in some way. The two major types of dependent-samples designs are matched pairs and repeated measures. **Matched-pairs designs** are used when researchers need an experimental group and a control group but are unable to use random assignment to create the groups. They therefore gather a sample from a treatment group and then construct a control group via the deliberate selection of cases that did not receive the treatment but that are similar to the treatment group cases on key characteristics. If the unit of analysis is people, participants in the control group might be matched to the treatment group on race, gender, age, and criminal history.

> **Matched-pairs design**: A research strategy where a second sample is created on the basis of each case's similarity to a case in an existing sample.

Figure 11.1 The Sampling Distribution of Differences Between Means

Figure displays a bell-shaped normal curve. The left side is labeled $\bar{x}_1 < \bar{x}_2$ and the right side is labeled $\bar{x}_1 > \bar{x}_2$. The center point on the horizontal axis is labeled $\bar{\mu}_1 - \bar{\mu}_2$.
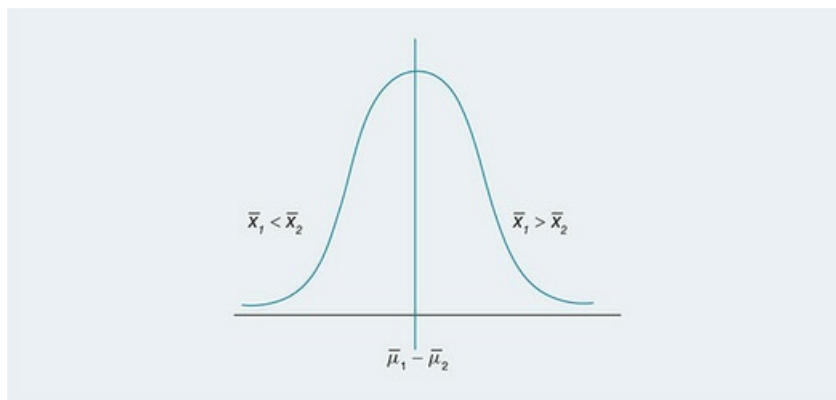
Repeated-measures designs are commonly used to evaluate program impacts. These are before-and-after designs wherein the treatment group is measured prior to the intervention of interest and then again afterward to determine whether the post-intervention scores differ significantly from the pre-intervention scores. In repeated measures, then, the "two" samples are actually the same people or objects measured twice.

> **Repeated-measures design**: A research strategy used to measure the effectiveness of an intervention by comparing two sets of scores (pre and post) from the same sample.

The first step in deciding what kind of t test to use is to figure out whether the samples are independent or dependent. It is sometimes easier to identify dependent designs than independent ones. The biggest clue to look for is a description of the research methods. If the samples were collected by matching individual cases on the basis of similarities between them, or if an intervention was being evaluated by collecting data before and after a particular event, then the samples are dependent and the dependent-samples t test is appropriate. If the methods do not detail a process of matching or of repeated measurement, if all that is said is that two samples were collected or a single sample was divided on the basis of a certain characteristic to form two subsamples, then you are probably dealing with independent samples and should use the independent-samples t test.

There is one more wrinkle. There are two types of independent-samples t tests: **pooled variances** and **separate variances**. The former is used when the two population variances are similar to one another, whereas the latter is used when the variances are significantly disparate. The rationale for having these two options is that when

two samples' variances are similar, they can safely be combined (pooled) into a single estimate of the population variance. When they are markedly unequal, however, they must be mathematically manipulated before being combined. You will not be able to tell merely by looking at two samples' variances whether you should use a pooled-variance or separate-variance approach, but that is fine. In this book, you will always be told which one to use. When we get to SPSS, you will see that this program produces results from both of these tests, along with a criterion to use for deciding between them. (More on this later.) You will, therefore, always be able to figure out which type of test to use.

> **Pooled variances**: The type of *t* test appropriate when the samples are independent and the population variances are equal.
>
> **Separate variances**: The type of *t* test appropriate when the samples are independent and the population variances are unequal.

This has probably all gotten somewhat murky. Luckily, there is a simple set of steps you can follow anytime you encounter a two-population test that will help you determine which type of *t* test to use. This mental sequence is depicted in Figure 11.2.

In this chapter, we will encounter something we have touched on before but have not addressed in detail: **one-tailed tests** versus two-tailed tests. We discussed the *t* distribution in Chapters 7 and 8. The *t* distribution is symmetric and has positive and negative sides. In two-tailed tests, there are two critical values, one positive and one negative. You learned in Chapter 8 that confidence intervals are always two-tailed. In *t* tests, by contrast, some analyses will be two-tailed and some will be one-tailed. Two-tailed tests split alpha (α) in half such that half is in each tail of the distribution. The critical value associated with that α is both positive and negative. One-tailed tests, by contrast, place all α into a single tail. One-tailed tests have α in either the upper (or positive) tail or lower (or negative) tail, depending on the specific question under investigation. The critical value of a one-tailed test is either positive or negative.

> **One-tailed tests**: Hypothesis tests in which the entire alpha is placed in either the upper (positive) or lower (negative) tail such that there is only one critical value of the test statistic. Also called directional tests.

The choice of a one-tailed test versus a two-tailed test is generally made on a case-by-case basis. It depends on whether a researcher has a good reason to believe that the relationship under examination should be positive or negative. Suppose that you are studying an in-prison treatment program that focuses on improving participants' literacy skills. You would measure their reading levels before the program began and then again after it had ended, and would expect to see an increase—you have good reason to predict that post-intervention literacy skills would be greater than pre-intervention ones. In this case, a one-tailed test would be in order (these are also called directional tests, since a prediction is being made about the direction of a relationship). Now suppose you want to know whether that literacy program works better for men or for women. You do not have any particular reason for thinking that it would be more effective for one group than the other, so you set out merely to test for any difference at all, regardless of direction. This would be cause for using a two-tailed test (also called a nondirectional test). Let us work our way through some examples and discuss one-tailed and two-tailed tests as we go.

Figure 11.2 Steps for Deciding Which *t* Test to Use

```
                    ┌─────────────────────┐
                    │ 1. Are the samples  │
                    │    independent?     │
                    └─────────────────────┘
                       │              │
                       ▼              ▼
                    ┌──────┐      ┌──────┐
                    │ Yes  │      │  No  │
                    └──────┘      └──────┘
                       │              │
                       ▼              ▼
              ┌────────────────┐  ┌────────────────┐
              │ Use independent│  │ Use dependent  │
              │ samples t test.│  │ samples t test.│
              └────────────────┘  └────────────────┘
                       │
                       ▼
              ┌────────────────┐
              │ 2. Are the variances
              │    equal?      │
              └────────────────┘
                 │          │
                 ▼          ▼
              ┌──────┐  ┌──────┐
              │ Yes  │  │  No  │
              └──────┘  └──────┘
                 │          │
                 ▼          ▼
          ┌────────────┐  ┌────────────┐
          │ Use pooled-│  │ Use separate-
          │ variances  │  │ variances  │
          │ t test.    │  │ t test.    │
          └────────────┘  └────────────┘
```

1. Are the samples independent?

Yes

No

Use independent samples *t* test.

Use dependent samples *t* test.

2. Are the variances equal?

Yes

No

Use pooled-variances *t* test.

Use separate-variances *t* test.
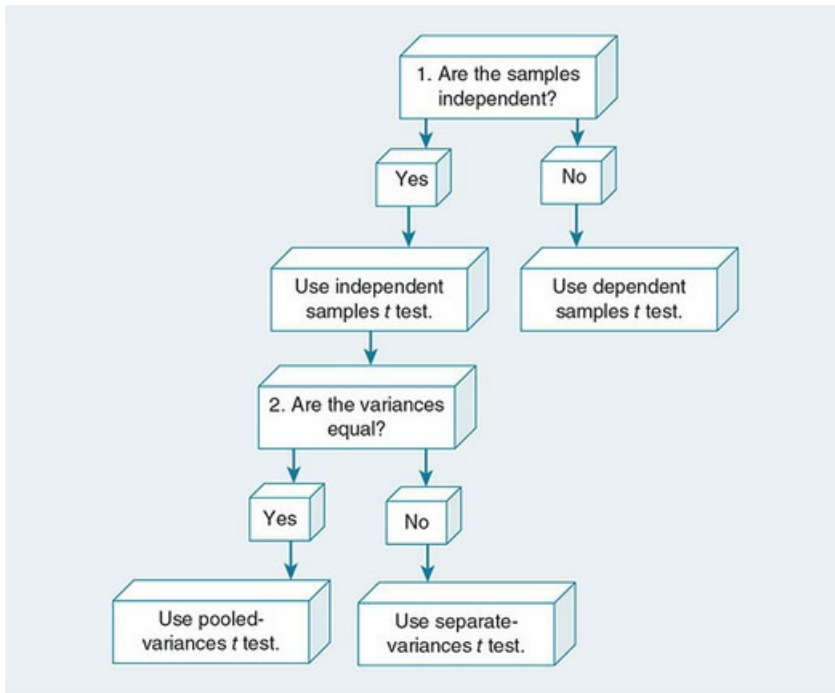
Independent-Samples *t* Tests

A Two-Tailed Test With Equal Population Variances: Transferred Female Juveniles' Ethnicity and Mean Age of Arrest

There is a theoretical and empirical connection between how old people are when they start committing delinquent offenses (this is called the *age of onset*) and their likelihood of continuing law-breaking behavior in adulthood. All else being equal, younger ages of onset are associated with greater risks of adult criminal activity. Using the Juvenile Defendants in Criminal Court (JDCC) data set ([Data Sources 11.1](#)), we can test for a significant difference in the mean age at which juveniles were arrested for the offense that led to them being transferred to adult court. To address questions about gender and ethnicity, the sample is narrowed to females, and we will test for an age difference between Hispanics and non-Hispanic whites in this subsample. Among Hispanic female juveniles in the JDCC sample ($N = 44$), the mean age of arrest was 15.89 years ($s = 1.45$). Among non-Hispanic whites ($N = 31$), the mean age of arrest was 16.57 years ($s = 1.11$). Since the IV is race (*Hispanic; white*) and the DV is age at arrest (*years*) , a *t* test is the proper analytic strategy. We will use an α level of .05, a presumption that the population variances are equal, and the five steps of hypothesis testing.

> Data Sources 11.1 Juvenile Defendants in Criminal Courts
>
> The JDCC is a subset of the Bureau of Justice Statistics' (BJS) State Court Processing series that gathers information on defendants convicted of felonies in large, urban counties. BJS researchers pulled information about juveniles charged with felonies in 40 of these counties in May 1998. Each case was tracked through disposition. Information about the juveniles' demographics, court processes, final dispositions, and sentences was recorded. Due to issues with access to and acquisition of data in some of the counties, the JDCC is a nonprobability sample, and conclusions drawn from it should therefore be interpreted cautiously (BJS, 1998).

Table 11.1   Relevant Numbers

Sample 1: Hispanic Females	Sample 2: Non-Hispanic White Females	Test
$\bar{x}_1 = 15.89$	$\bar{x}_2 = 16.57$	$\alpha = .05$
$s_1 = 1.45$	$s_2 = 1.11$	two-tailed test
$N_1 = 44$	$N_2 = 31$	equal population variances

It is useful in an independent-samples *t* test to first make a table that lays out the relevant pieces of information that you will need for the test. [Table 11.1](#) shows these numbers. It does not matter which sample you designate Sample 1 and which you call Sample 2 as long as you stick with your original designation throughout the course of the hypothesis test. Since it is easy to simply designate the samples in the order in which they appear in the problem, let us call Hispanic females Sample 1 and white females Sample 2.

We will use a two-tailed test because we have no solid theoretical reason for thinking that non-Hispanic whites' mean would be greater than Hispanics' or vice versa. The alternative hypothesis will merely specify a difference (i.e., an inequality) between the means, with no prediction about which one is greater than or less than the other.

***Step 1. State the null* (H$_0$) *and alternative (H$_1$) hypotheses* .**

In *t* tests, the null (*H$_0$*) and alternative (*H$_1$*) are phrased in terms of the population means. Recall that population means are symbolized µ (the Greek letter *mu* , pronounced "mew"). We use the population symbols rather than the sample symbols because the goal is to make a statement about the relationship, or lack thereof, between two variables in the population. The null hypothesis for a *t* test is that the means are equal:

H_0: µ$_1$ = µ$_2$

Equivalence in the means suggests that the IV is not exerting an impact on the DV. Another way of thinking about this is that *H$_0$* predicts that the two samples came from the same population. In the context of the present example, retaining the null would indicate that ethnicity does not affect female juveniles' age of arrest (i.e., that all female juveniles are part of the same population).

The alternative or research hypothesis is that there *is* a significant difference between the population means or, in other words, that there are two separate populations, each with its own mean:

H_1: $\mu_1 \neq \mu_2$

Rejecting the null would lead to the conclusion that the IV does affect the DV; here, it would mean that ethnicity does appear related to age of arrest. Note that this phrasing of the alternative hypothesis is specific to two-tailed tests—the "not equal" sign implies no prediction about the direction of the difference. The alternative hypothesis will be phrased slightly differently for one-tailed or directional tests.

***Step 2. Identify the distribution and compute the degrees of freedom* .**

As mentioned earlier, two-population tests for differences between means employ the *t* distribution. The *t* distribution, you should recall, is a family of curves that changes shape depending on degrees of freedom (*df*) . The *t* distribution is normal at large *df* and gets wider and flatter as *df* declines.

The *df* formula differs across the three types of *t* tests, so you have to identify the proper test before you can compute the *df*. Using the sequence depicted in Figure 11.2, we know (1) that the samples are independent because this is a random sample divided into two groups and (2) that the population variances are equal. This leads us to choose the pooled-variances *t* test. The *df* formula is

$$df = N_1 + N_2 - 2, \qquad\qquad\qquad Formula\ 11(1)$$

where

     N_1 = the size of the first sample and
     N_2 = the size of the second sample.

Pulling the sample sizes from Table 11.1,

df = 44 + 31 – 2 = 73

Step 3. Identify the critical value and state the decision rule .

Three pieces of information are required to find the critical value of *t* (t_{crit}) using the *t* table: the number of tails in the test, the alpha level, and the *df*. The exact *df* value of 73 does not appear on the table, so we use the value that is closest to it, which is 60. With two tails, an α of .05, and 73 degrees of freedom, we see that t_{crit} *is 2.000*.

This is not the end of finding the critical value, though, because we still have to figure out the sign or signs; that is, we need to determine whether t_{crit} is positive, negative, or both. A one-tailed test has only one critical value and it is either positive or negative. In two-tailed tests, there are always two critical values. Their absolute values are the same, but one is negative and one is positive. Figure 11.3 illustrates this.

Given that there are two tails in this current test and, therefore, two critical values, t_{crit} = ±2.000. The decision rule is stated thus: *If t_{obt} is either greater than 2.000 or less than –2.000* , H$_0$ *will be rejected*. The decision rule has to be stated as an "either/or" proposition because of the presence of two critical values. There are, in essence, two ways for the null to be rejected: t_{obt} could be out in the right tail beyond 2.000 or out in the left tail beyond –2.000.

Figure 11.3 The Critical Values for a Two-Tailed Test With α = .05 and *df* = 73

area = .025   area = .025

t_{crit} = –2.000 t_{crit} = –2.000
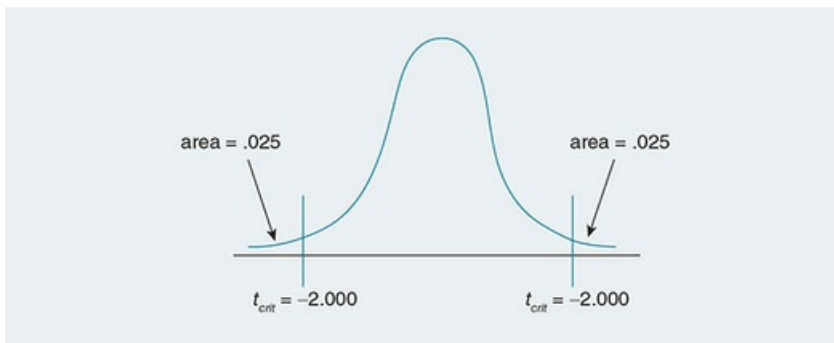
Step 4. Calculate the obtained value of the test statistic .

The formulas for the obtained value of *t* (t_{obt}) vary across the three different types of *t* tests; however, the common thread is to have (1) a measure of the difference between means in the numerator and (2) an estimate of the standard error of the sampling distribution of differences between means in the denominator (remember that the standard error is the standard deviation of a sampling distribution). The estimated standard error is symbolized $\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}$ and the formula for estimating it with pooled variances is

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}} \sqrt{\frac{N_1 + N_2}{N_1 N_2}} \qquad \textit{Formula 11 (2)}$$

This formula might look a bit daunting, but keep in mind that it comprises only sample sizes and standard deviations, both of which are numbers you are accustomed to working with. The most important thing is to

work through the formula carefully. Plug the numbers in correctly, use proper equation-solving techniques (including order of operations), and round correctly. Entering the numbers from our example yields

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(44-1)1.45^2 + (31-1)1.11^2}{44+31-2}} \sqrt{\frac{44+31}{44 \cdot 31}}$$

$$= \sqrt{\frac{(43)2.10 + (30)1.23}{73}} \sqrt{\frac{75}{1,364}} = \sqrt{\frac{90.30 + 36.90}{73}} \sqrt{.05}$$

$$= \sqrt{\frac{127.20}{73}} \sqrt{.05} = \sqrt{1.74} \sqrt{.05}$$

$$= 1.32 \, (.22)$$

$$= .29$$

This is our estimate of the standard error (i.e., the standard deviation of the sampling distribution). Recall that this is not t_{obt}! Be careful. The next step is to plug the standard error into the t_{obt} formula. This formula is

$$t_{obt} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} \qquad \qquad Formula\ 11\,(3)$$

Using our numbers, we perform the calculation:

$$t_{obt} = \frac{15.89 - 16.57}{.29} = \frac{-.68}{.29} = -2.34$$

This is the final answer! The obtained value of t is –2.34. Step 4 is done.

Step 5. Make a decision about the null, and state the substantive conclusion .

We said in the decision rule that if t_{obt} was either greater than 2.000 or less than –2.000, the null would be rejected. So, what will we do? If you said, "Reject the null," you are correct. Since t_{obt} is less than –2.000, the null is rejected. The conclusion is that non-Hispanic white and Hispanic female juveniles transferred to adult court differ significantly in terms of mean age of arrest for their current offense. Another way to think about it is that there is a relationship between ethnicity and mean age at arrest. Looking at the means, it is clear that Hispanic youths were younger, on average (mean = 15.89), than non-Hispanic white youths were (mean = 16.57).

The interpretation of a significant result in a two-tailed test is complicated and must be done carefully. In the present example, we used a two-tailed test because we did not have sufficient theory or prior empirical evidence to make a defensible prediction about the direction of the difference. The fact that we ultimately

found that Hispanics' mean age was lower is not enough to arrive at a conclusion about the reason for this. Never let your data drive your thinking—do not construct a reality around an empirical finding. In an exploratory analysis, it is best to avoid speculating about the larger implications of your results. The more scientifically sound approach is to continue this line of research to uncover potential reasons why Hispanic females' age of arrest might be lower than non-Hispanic white females' and then collect more data to test this prediction.

A One-Tailed Test With Unequal Population Variances: Trial Types and Sentence Lengths

For the second *t*-test example, we will again use the JDCC data on juveniles diverted to adult criminal courts. Let's consider whether the length of time it takes for a juvenile's case to be disposed of is affected by the type of attorney the juvenile has. The prediction will be that juvenile defendants who retain private counsel experience significantly longer time-to-disposition compared to those who use the services of public defenders. This, theoretically, is because private attorneys might file more pretrial motions and spend more time negotiating with the prosecutor and the judge. The sample is narrowed to juveniles charged with violent offenses who were not released pending trial and who were ultimately sentenced to prison. Those with private attorneys ($N = 36$) had a mean of 7.93 months to disposition ($s = 4.53$), whereas those represented by public defenders ($N = 234$) experienced a mean of 6.36 months ($s = 3.66$). We will call the defendants who retained private attorneys Sample 1 and those who were represented by public defenders Sample 2. Using an alpha level of .01 and the assumption that the population variances are unequal, we will conduct a five-step hypothesis test to determine whether persons convicted by juries receive significantly longer prison sentences. Table 11.2 shows the numbers we will need for the analysis.

Step 1. State the null (H_0) *and* (H_1) *alternative hypotheses*.

The null hypothesis is the same as that used above (H_0: $\mu_1 = \mu_2$) and reflects the prediction that the two means do not differ. The alternative hypothesis used in the previous example, however, does not apply in the present context because this time, we are making a prediction about which mean will be greater than the other. The nondirectional sign ($\neq$) must therefore be replaced by a sign that indicates a specific direction. This will either be a greater than (>) or less than (<) sign. We are predicting that jury trials will result in significantly longer mean sentences, so we can conceptualize the hypothesis as *private attorney disposition time > public defender disposition time*. Since defendants represented by private attorneys are Sample 1 and those by public defenders Sample 2, the alternative hypothesis is

H_1: $\mu_1 > \mu_2$

Table 11.2 Relevant Numbers		
Sample 1: Private Attorney	**Sample 2: Public Defender**	**Test**
$\bar{x}_1 = 7.93$	$\bar{x}_2 = 6.36$	$\alpha = .01$
$s_1 = 4.53$	$s_2 = 3.66$	one-tailed test
$N_1 = 36$	$N_2 = 234$	unequal population variances

Step 2. Identify the distribution and compute the degrees of freedom .

The distribution is still t , but the *df* equation for unequal population variances differs sharply from that for equal variances because the situation of unequal variances mandates the use of the separate-variances t test. The *df* formula is obnoxious, but as with the prior formulas we have encountered, you have everything you need to solve it correctly—just take care to plug in the right numbers and use proper order of operations.

Plugging in the correct numbers from the current example,

$$df = \frac{\left(\dfrac{s_1^2}{N_1} + \dfrac{s_2^2}{N_2}\right)^2}{\left(\dfrac{s_1^2}{N_1}\right)^2\left(\dfrac{1}{N_1-1}\right) + \left(\dfrac{s_2^2}{N_2}\right)^2\left(\dfrac{1}{N_2-1}\right)} \qquad \text{Formula 11(4)}$$

$$= \frac{\left(\dfrac{4.53^2}{36} + \dfrac{3.66^2}{234}\right)^2}{\left(\dfrac{4.53^2}{36}\right)^2\left(\dfrac{1}{36-1}\right) + \left(\dfrac{3.66^2}{234}\right)^2\left(\dfrac{1}{234-1}\right)}$$

$$= \frac{\left(\dfrac{20.52}{36} + \dfrac{13.40}{234}\right)^2}{\left(\dfrac{20.52}{36}\right)^2\left(\dfrac{1}{35}\right) + \left(\dfrac{13.40}{234}\right)^2\left(\dfrac{1}{233}\right)}$$

$$= \frac{(.57+.06)^2}{(.57)^2(.03)+(.06)^2(.004)} = \frac{.63^2}{(.32)(.03)+(.004)(.004)}$$

$$= \frac{.40}{.01+.00002} = \frac{.40}{.01}$$

= 40

Step 2 is complete; df = 40. We can use this df to locate the critical value of the test statistic.

Step 3. Identify the critical value and state the decision rule .

With one tail, an α of .01, and 40 degrees of freedom, t_{crit} = 2.423. The sign of the critical value is positive because the alternative hypothesis predicts that $\overline{x}_1 > \overline{x}_2$. Revisit Figure 11.1 for an illustration. When the alternative predicts that $\overline{x}_1$, the critical value will be on the left (negative) side of the distribution and when the alternative is that $\overline{x}_1 > \overline{x}_2$, t_{crit} will be on the right (positive) side. The decision rule is that *if t_{obt} is greater than 2.423* , H$_0$ *will be rejected.*

Step 4. Compute the obtained value of the test statistic .

As before, the first step is to obtain an estimate of the standard error of the sampling distribution. Since the population variances are unequal, the separate variances version of independent-samples t must be used. The standard error formula for the difference between means in a separate-variance t test is

$$\hat{\sigma}_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}} \qquad \text{Formula 11(5)}$$

Plugging in the numbers from the present example yields

$$\hat{\sigma}_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{4.53^2}{36-1} + \frac{3.66^2}{234-1}} = \sqrt{\frac{20.52}{35} + \frac{13.40}{233}} = \sqrt{.59 + .06} = \sqrt{.65} = .81$$

Now, the standard error estimate can be entered into the same t_{obt} formula used with the pooled-variances t test. Using Formula 11(3),

$$t_{obt} = \frac{7.93 - 6.36}{.81} = \frac{1.57}{.81} = 1.94$$

Step 5. Make a decision about the null and state the substantive conclusion .

The decision rule stated that the null would be rejected if t_{obt} exceeded 2.423. Since t_{obt} ended up being 1.94, the null is retained. Juveniles who had privately retained attorneys did not experience a statistically significant increase in the amount of time it took for their cases to be resolved, compared to juveniles who had public defenders. Another way to say this is that there is no relationship between attorney type and disposition time.

Dependent-Samples *t* Tests

The foregoing discussion centered on the situation in which a researcher is working with two independently selected samples; however, as described earlier, there are times when the samples under examination are not independent. The main types of dependent samples are *matched pairs* and *repeated measures*. Dependent samples require a *t* formula different from that used when the study samples are independent because of the manipulation entailed in selecting dependent samples. With dependent-samples *t*, the sample size (*N*) is not the total number of people or objects in the sample but, rather, the number of *pairs* being examined. We will go through an example now to demonstrate the use of this *t* test.

Research Example 11.2 Do Mentally Ill Offenders' Crimes Cost More?

There is ongoing debate about the precise role of mental illness (MI) in offending. Some individuals with MI are unstable or antisocial, but it would be wrong to stigmatize an entire group based on the behavior of a few members. One question that could be asked is whether the crimes of offenders with MI exact a particularly heavy toll on taxpayers relative to offenders who do not have MI. Ostermann and Matejkowski (2014) collected data on all persons released from New Jersey prisons in 2006. The data included the number of times each ex-prisoner was rearrested within a 3-year follow-up period. First, the researchers divided the group according to whether or not each person had received a MI diagnosis and then calculated the average cost of each group's recidivism. The results showed that MI offenders' crimes were nearly three times more expensive compared to non-MI offenders. Next, the authors matched the sample of MI offenders to a subsample of non-MI offenders on the basis of each person's demographic characteristics and offense histories and then recalculated each group's average cost. The results changed dramatically. After the one-to-one matching procedure, the non-MI group's average cost was more than double that of the MI group. It turns out that the initial results—the ones suggesting that MI offenders' crimes are much more expensive—are misleading. It is not good policy to use the mere existence of MI as cause to enhance supervision or restrictions on ex-prisoners. What should be done instead is to focus on the risk factors that are associated with recidivisms among both MI and non-MI offenders. This policy focus would cut costs and create higher levels of social justice.

Dependent-Samples *t* Test: Female Correctional Officers and Institutional Security

The traditionally male-dominated field of correctional security is gradually being opened to women who wish to work in jails and prisons, yet there are lingering concerns regarding how well female correctional officers can maintain order in male institutions. Critics claim that women are not as capable as men when it comes to controlling inmates, which threatens the internal safety and security of the prison environment. Let us test the hypothesis that facilities with relatively small percentages of female security staff will have lower inmate-on-staff assault rates relative to those institutions with high percentages of female security staff because security in the latter will be compromised. We will use data from the Census of Jails (COJ; Data Sources 3.1) and an alpha level of .05. The first sample consists of five jails with below-average percentages of female security staff, and the second sample contains five prisons selected on the basis of each one's similarity to a prison in the first sample (i.e., the second sample's prisons are all male, state-run, maximum-security facilities in Texas with inmate totals similar to those of the first sample). The difference between the samples is that the second sample has above-average percentages of female security staff. Table 11.3 contains the raw data.

Step 1. State the null (N0) *and* (N1) *alternative hypotheses* .

The null, as is always the case with t tests, is $H_0: \mu_1 = \mu_2$. It is being suggested in this problem that low-percentage female jails (i.e., jails with greater percentages of male staff) should have lower assault rates than high-percentage female (low male percentage) jails do; hence, the alternative in words is *low < high* and in formal symbols is $H_1: \mu_1 < \mu_2$.

Table 11.3 Matched Samples of Jails	Assault Rate	
	Low Percentage Female (x_1)	High Percentage Female (x_2)
Pair A	1.30	1.00
Pair B	1.25	.76
Pair C	.86	.99
Pair D	.94	1.10
Pair E	2.26	1.85

Step 2. Identify the distribution and compute the degrees of freedom .

The distribution is t. The degrees of freedom calculation for dependent-samples tests is different from those earlier—for dependent samples, the df is based on the number of pairs rather than on the size of each sample. The df for pairs is

$$df = N_{pairs} - 1 \hspace{4cm} Formula\ 11\,(6)$$

Using the data from Example 3,

$$df = 5 - 1 = 4$$

Step 3. Identify the critical value and state the decision rule .

With an alpha of .05, a one-tailed test, and $df = 4$, the absolute value derived using the t table is 2.132. To determine whether this critical value is positive or negative, refer back to Figure 11.1. The phrasing of our alternative hypothesis makes the critical value negative, so $t_{crit} = -2.132$. The decision rule states that *if* $t_{obt} < -2.132$, H_0 *will be rejected*.

Step 4. Compute the obtained value of the test statistic .

The formulas required to calculate t_{obt} for dependent samples looks quite different from those for independent samples, but the logic is the same. The numerator contains a measure of the difference between means, and the denominator is an estimate of the standard error. Finding the standard error requires two steps. First, the standard deviation of the differences between means (symbolized s_D) is needed. In the independent-samples t tests we conducted earlier, we knew each sample's standard deviation and so we could plug those numbers directly into the standard-error formula. We do not yet know the standard deviation for the matched-pairs sample, however, so we have to obtain that number before tackling the standard error. The standard deviation for matched pairs is computed using the formula

$$s_D = \sqrt{\frac{\sum(x_D - \bar{x}_D)^2}{N_{pairs} - 1}}$$

Formula 11(7)

where

x_D = the difference scores and

$\bar{x}_D$ = the mean of the difference scores.

The standard deviation is then used to find the standard error of the sampling distribution, as follows:

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \frac{s_D}{\sqrt{N_{pairs}}}$$

Formula 11(8)

Finally, the obtained value of the test statistic is calculated as

$$t_{obt} = \frac{\bar{x}_D}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}$$

Formula 11(9)

This computation process is substantially simplified by the use of Table 11.4. This table contains the raw scores and three additional columns. The first column to the right of the raw scores contains the difference scores (x_D), which are computed by subtracting each x_2 score from its corresponding x_1 value. The mean difference score ($\bar{x}_D$) is then needed and is calculated thus:

$$\bar{x}_D = \frac{\sum x_D}{N}$$

Formula 11(10)

The mean difference score is subtracted from each individual difference score (this is the $x_D - \bar{x}_D$ column) to form deviation scores. Finally, each of these scores is squared and the last column summed. The final product of Table 11.4 is the sum of the squared deviation scores, located in the lower-right-hand corner. This number (here, .38) gets entered into the s_D formula and the calculations proceed through t_{obt} as such:

$$s_D = \sqrt{\frac{.38}{5-1}} = \sqrt{.10} = .32 \quad \hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \frac{.32}{\sqrt{5}} = \frac{.32}{2.24} = .14$$

$$t_{obt} = \frac{.18}{.14} = 1.29$$

And Step 4 is (finally) done! The obtained value of t is 1.29.

Table 11.4 Matched Samples of Jails

	Assault Rate by Percentage Female		$x_D = x_1 - x_2$	$x_D - \bar{x}_D$	$(x_D - \bar{x}_D)^2$
	Low Percentage Female (x_1)	High Percentage Female (x_2)			
Pair A	1.30	1.00	1.30 − 1.00 = .30	.30 − .18 = .12	$(12)^2$ = .01
Pair B	1.25	.76	.49	.31	.10
Pair C	.86	.99	−.13	−.31	.10
Pair D	.94	1.10	−.16	−.34	.12
Pair E	2.26	1.85	.41	.23	.05
			$\bar{x}_D = \dfrac{.91}{5} = .18$		$\Sigma = .38$

Step 5. Make a decision about the null and state the substantive conclusion.

The decision rule stated that the null would be rejected if the obtained value was less than –2.132. With t_{obt} = 1.29, the null is retained. The conclusion is that there is no relationship between a jail's level of female staffing and its rate of inmate-on-staff assaults. There is no support for the notion that female correctional officers compromise institutional security.

Research Example 11.3 Do Targeted Interventions Reduce Crime?

The past two decades have seen a shift in criminal justice operations toward an emphasis on multiagency efforts designed to address the specific crime problems plaguing individual neighborhoods and communities. One strategy that has been devised is the "pulling levers" approach that entails identifying high-risk offenders and offering them a choice between tough prosecution and reform. Police and prosecutors explain the penalties these offenders would face if they continue their antisocial behavior, and community social service providers offer treatment and educational opportunities for those who want to put their lives on a better track. In 2007, the Rockford (Illinois) Police Department (RPD) implemented a pulling levers approach in an attempt to reduce violent crime in the city. Corsaro, Brunson, and McGarrell (2013) evaluated the RPD intervention. They measured the mean number of violent crimes that occurred in several areas of the city per month before and after the program. The following table shows these results. The target neighborhood is the one in which the pulling levers strategy was implemented.

Using a statistical analysis, the researchers found that the target neighborhood experienced a significant reduction in nonviolent crime as a result of the pulling levers program. The reduction in violent crime, though, was not statistically significant, so it appeared that the intervention did not affect these types of offenses. The researchers were able to attribute the target area's decline in nonviolent offenses to the pulling levers strategy because crime fell only in the target zone and nowhere else in the city. It thus appears from this analysis that pulling levers is a promising strategy for reducing nonviolent crimes such as drug, property, and nuisance offenses but is perhaps less useful with respect to violent crime.

	Number of Offenses per Month	
Location	Pre-Intervention	Post-Intervention
Target Neighborhood		
Nonviolent	29	22
Violent	21	18
Remainder of City		
Nonviolent	944	859
Violent	567	554
Overall City		
Nonviolent	1,013	881
Violent	588	573

Two-Population Tests for Differences Between Proportions

Two-population tests for differences between proportions follow the same logic as those for differences between means. The IV is still a two-class, categorical measure; however, the DV is a proportion rather than a mean. Differences between proportions have their own sampling distribution, which looks very much like that for differences between means and can be drawn as .

Population proportions are symbolized with the letter P and sample proportions with $\hat{P}$ (pronounced "p hat"). Like confidence intervals for proportions, two-population tests employ the z distribution. Thus, the sample size must be at least 100; these tests cannot be carried out on samples smaller than that. All the other fundamentals for proportion tests are the same as those for tests of differences between means, so we will dive right into an example.

Research Example 11.4 Does the Gender Gap in Offending Rates Differ Between Male and Female Drug Abusers?

It is well known that males have higher rates of criminal involvement than females do, but it is less clear whether this gender differential exists among people who have criminal histories. Kaskela and Pitkänen (2016) examined the gender gap among substance abusers with criminal histories. Using data from a national database in Finland, the researchers compared substance abusers with and without criminal records to determine the magnitude of the differences in offending between men and women in each group. The table displays information about the people in these four groups.

Variable	Criminal Conviction			No Criminal Conviction		
	Men (%)	Women (%)	p	Men (%)	Women (%)	p
Secondary education or higher	29.6	23.2	.010	52.1	48.5	.010
Employed	14.8	9.2	.003	42.8	36.0	<.001
Problem with drugs	77.4	79.6	.354	52.9	55.1	.083
Committed a new crime	76.8	68.5	.001	49.3	35.7	<.001

Several significant differences emerged between men and women in each group, as indicated by the p values of .05 or less. Females of both groups had lower educational attainment and lower employment rates. Men and women did not differ in rates of drug abuse (as opposed to having problems with alcohol only). The gender gap was apparent in both groups, as well, with men significantly more likely than women to have committed new crimes during the study period. Interestingly, the women with criminal histories committed crimes at higher rates (68.5%) than did men without records (49.3%). The gap between the men and women with records was 8.3 percentage points and that between men and women without prior convictions was 13.6 percentage points. These findings suggest that the gender gap narrows among people who have committed criminal offenses in the past and that women with previous run-ins with the law are more likely than men without such histories to engage in future lawbreaking.

Figure 11.4 The Sampling Distribution of Differences Between Proportions
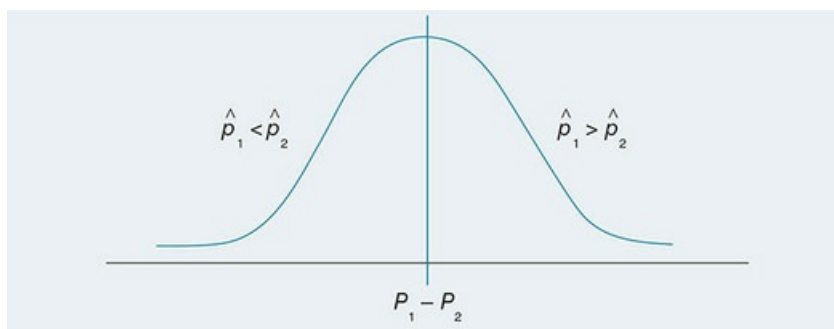
Community-Based Sex-Offender Treatment Programs and Recidivism

Sex offenders are perhaps the most reviled class of criminal offenders, yet imprisoning low-level sex offenders is not necessarily the best option in terms of social and economic policy. These people might be good candidates for community-based treatment, which can be effective at reducing the likelihood of recidivism and is less expensive than imprisonment. Washington State has a program called the Special Sex Offender Sentencing Alternative (SSOSA) that separates low-risk from high-risk sex offenders and sentences the low-risk offenders to community-based supervision and treatment. Participants must maintain standards of good behavior, comply with imposed conditions, and adhere to the required treatment regimen. Transgressors are removed from SSOSA and incarcerated. This program has raised questions about how well the SSOSA program reduces recidivism. It has also generated concerns among critics who fear that allowing convicted sex offenders to remain in the community jeopardizes public safety.

To address the issue of sex offender recidivism and the effectiveness of the SSOSA program, researchers from the Washington State Institute for Public Policy compared recidivism across sex offenders who were sentenced to SSOSA and those who were sentenced to jail or prison instead of to this community-based treatment program. Recidivism was measured as a conviction for a new crime within 5 years of release from prison or the program. Among the SSOSA group ($N = 1,097$), 10% were convicted of new crimes; among the non-SSOSA group that was incarcerated ($N = 2,994$), 30% were reconvicted (Barnoski, 2005). Using an alpha level of .05, let us test the null hypothesis that there is no difference between the population proportions against the alternative hypothesis that the SSOSA group's recidivism rate was significantly lower than the incarceration group's. The use of a directional hypothesis is justified because the SSOSA program is intended to lower recidivism.

Table 11.5 Relevant Numbers		
Sample 1: SSOSA	**Sample 2: Public Defender**	**Test**
$\hat{p} = .10$	$\hat{p} = .30$	$\alpha = .01$
$N_1 = 1,097$	$N_2 = 2,994$	one-tailed test

Step 1. State the null (H_0) and (H_1) alternative hypotheses .

The null hypothesis for a two-population test for differences between proportions represents the prediction that the two proportions do not differ or, in other words, that the two population proportions are equal. The alternative hypothesis can take on the three forms discussed in the preceding section regarding two-population tests for means. Here, calling the SSOSA group Sample 1 and the incarceration group Sample 2, the hypotheses are

H_0: $P_1 = P_2$

H_1: $P_1 < P_2$

Step 2. Identify the distribution and compute the degrees of freedom .

The proper distribution for two-population tests of proportions is the z curve. It is important to note, though, that this distribution can be used only when the samples are large and independent; violation of either of these assumptions can be fatal to a test of this type. You should always have enough knowledge about the research design that produced the data you are working with to determine whether the independence assumption has been met. In the present case, it has, because the SSOSA and incarceration groups were unrelated to each other. Remember that the proportion test can be conducted only when $N \geq 100$; there is an additional requirement that both the proportion of cases that fall into the category of interest ($\hat{p}$) and the proportion that does not (symbolized $\hat{q}$) are both greater than or equal to five. Formally stated,

$$N_{\hat{p}_k} \geq 5 \qquad\qquad\qquad Formula\ 11(11)$$

and

$$N_{\hat{q}_k} \geq 5 \qquad\qquad\qquad Formula\ 11(12)$$

where

$\hat{p}_k$ = the sample proportion for each of Sample 1 and Sample 2 and

$\hat{q}_k = 1 - \hat{p}_k$ for each of Sample 1 and Sample 2.

In the current example, the SSOSA group meets the large sample criterion because $\hat{p} = .10$, which means that $\hat{q} = .90$. Therefore,

1,097(.10) = 109.70

1,097(.90) = 987.30

The incarceration group likewise succeeds, because $\hat{p} = .30$ and $\hat{q} = .70$, so

2,994(.30) = 898.20

2,994(.70) = 2,095.80

The z distribution can be used. There is no need to compute degrees of freedom because they are not applicable to z.

Step 3. Identify the critical value and state the decision rule .

It has been a while since we used the z distribution, but recall that a z value can be found using a known area. Here, alpha is that area. Since $\alpha = .05$ and the test is one-tailed, go to the z table and find the area closest to .50 – .05 = .45. There are actually two areas that fit this description (.4495 and .4505), so

$$z_{crit} = \frac{-1.64 + (-1.65)}{2} = -1.645$$

. The critical value is negative because the alternative hypothesis ($P_1 < P_2$) tells us that we are working on the left (negative side) of the distribution (see Figure 11.5). The decision rule is thus: *If z_{obt} is less than –1.65, H_0 will be rejected.*

Step 4. Compute the obtained value of the test statistic .

For this portion of the test, you need to be very careful about symbols because there are a few that are similar to one another but actually represent very different numbers. Pay close attention to the following:

P = population proportion

$\hat{p}$ = pooled sample proportions as an estimate of the population proportion

$\hat{q}$ = 1.00 – the pooled sample proportions

$\hat{p}_1$ = Sample 1 proportion

$\hat{p}_2$ = Sample 2 proportion

There are a few analytical steps in the buildup to the z_{obt} formula. First, the sample proportions have to be pooled in order to form a single estimate of the proposed population proportion:

$$\hat{p} = \frac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2} \qquad\qquad Formula\ 11(13)$$

The complement of the pooled proportion ($\hat{q}$) is also needed. This is done using the formula

$$\hat{q} = 1.00 - \hat{p} \qquad\qquad Formula\ 11(14)$$

Next, the standard error of the sampling distribution of differences between proportions must be estimated using the formula

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}\hat{q}}\sqrt{\frac{N_1 + N_2}{N_1 N_2}} \qquad\qquad Formula\ 11(15)$$

Finally, the obtained value of the test statistic is calculated as

$$z_{obt} = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}} \qquad\qquad Formula\ 11(16)$$

Now, we will plug in the numbers from the current example and solve the formulas all the way up to z_{obt} :

$$\hat{p} = \frac{1,097(.10)+2,994(.30)}{1,097+2,994} = \frac{109.70+898.20}{4,091} = \frac{1,007.90}{4,091} = .25$$

$$\hat{q} = 1.00 - .25 = .75$$

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{(.25).75)}\sqrt{\frac{1,097+2,994}{1,097\cdot 2,994}} = \sqrt{.19}\sqrt{\frac{4,091}{3,284,418}} = \sqrt{.19}\sqrt{.001} = (.44)(.03) = .01$$

$$z_{obt} = \frac{.10-.30}{.01} = \frac{-.20}{.01} = -20.00$$

Step 5. Make a decision about the null and state the substantive conclusion .

Since z_{obt} = –20.00, which is far less than the critical value of –1.65, the null is rejected. The sex-offender group that went through the SSOSA program had a significantly lower recidivism rate compared to the group sentenced to jail or prison. Of course, we cannot say for sure that this is because the SSOSA program is effective, because the offenders selected to go through this program might be those who were at the lowest risk for reoffending anyway. What does seem clear is that, at the very least, the SSOSA program is not harming public safety, could be reducing recidivism rates, and, therefore, might be a useful cost-savings program.

SPSS

The SPSS program can run all of the t tests discussed in this chapter, though it cannot run tests for differences between proportions, so you would have to use SPSS to derive the proportions and then do the analysis by hand. Independent-samples t tests are located under the *Analyze* menu in the SPSS data screen. In this menu, find *Compare Means* and then *Independent–Samples T Test*. To demonstrate the use of SPSS, some of the examples we did by hand will be replicated. Note that the final answers obtained in SPSS might depart somewhat from those that we calculated by hand because we use only two decimal places and SPSS uses far more than that, so our hand-derived answers have rounding errors.

First, let us run an independent-samples t test using the *ethnicity* and *age at arrest* variables that were under study earlier in this chapter. Figure 11.5 shows how to select these variables for an analysis. The IV is designated as the *Grouping Variable* and the DV goes into the *Test Variable(s)* space. You have to specify the values of the IV, as Figure 11.5 depicts. In the JDCC data set, the ethnicity variable is measured as 1 = Hispanic and 2 = non–Hispanic white. SPSS asks you to define groups in case the IV has more than two categories. If your IV has only two categories, then this step is redundant; if it has three or more, this is where you tell SPSS which two to use in the t test. Click *OK* to obtain the output shown in Figure 11.6.

The first box in Figure 11.6 shows the descriptive statistics for each category. The second box contains the t test results. You will notice that there are two values reported for t_{obt}, along with two p values. This is because SPSS produces results for both pooled- and separate-variances tests. Remember at the beginning of the chapter when we said that SPSS would help you decide which of these tests you should use? Well, this is it! Levene's test for equality of variances is an analysis that SPSS runs automatically. Levene's F statistic is a hypothesis test. The null is that the variances are equal (meaning that they can be pooled). The "F" column displays the obtained value of the F statistic and the "Sig." column shows the p value for F. It is customary to use an alpha level of .05 when determining whether or not to reject the null of equivalence. If $p < .05$, the null is rejected, which means the variances are unequal and the "Equal variances not assumed" row is the one containing the correct t test results. If $p > .05$, the null is retained and the "Equal variances assumed" line is the one to look at. In Figure 11.6, $F = .126$ and $p = .724$, so the null is retained and the variances are equal. The obtained value of t is –2.203, which is close to the value we arrived at by hand (–2.34). The p value is .031, which is less than .05, so judging by this output, the null should be rejected. (Typically, the alpha level .05 is used when interpreting SPSS output. Unless there is good reason to do otherwise, researchers generally consider any results with $p < .05$ to be statistically significant.) There is a relationship between ethnicity and age at arrest or, in other words, there is a significant difference between non-Hispanic white and Hispanic girls in the age at which they were arrested.

Figure 11.5 Running a Dependent-Samples t Test in SPSS
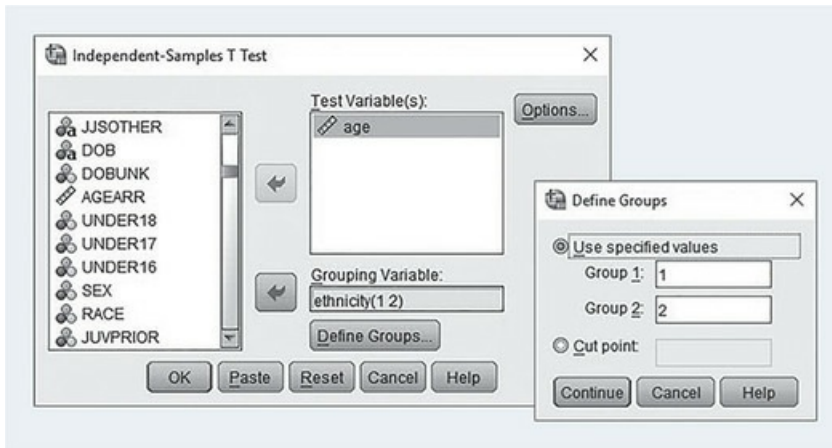
Figure 11.6 SPSS Output for Independent-Samples *t* Test

Group Statistics

	Race and ethnicity	N	Mean	Std. Deviation	Std. Error Mean
AGE AT ARREST	White	31	16.5745	1.10904	.19919
	Hispanic	44	15.8936	1.44616	.21802

Independent-Samples Test

		Levene's Test for Equality of Variances		t test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AGE AT ARREST	Equal variances assumed	.126	.724	-2.203	73	.031	-.68088	.30908	-1.29687	-.06489
	Equal variances not assumed			-2.306	72.421	.024	-.68088	.29531	-1.26951	-.09225

The procedure for dependent-samples *t* tests is a little different. The SPSS file has to be set up so that the rows represent pairs and two columns contain the two sets of scores for each pair. When you go to *Analyze →
Compare Means → Paired-Samples T Test* , you will see a variable list on the left and will need to select the two sets of scores that you wish to compare and move them to the right. Figure 11.7 demonstrates this.

Once you have moved the variables over, click *OK* and output like that in Figure 11.8 will appear.

Figure 11.8 contains the results of the matched sample of male, maximum-security prisons that we used for hand calculations in Example 3. The obtained value of *t* is .964, which differs only slightly from the value of .95 that our hand computations yielded. The *p* value is .390. This is well above the alpha level of .05, so the null is retained. There is no relationship between female staffing and inmate assault rates.

SPSS can be used to run *t* tests. You should use your knowledge of the research design to determine whether the samples are independent or dependent. If they are independent, Levene's *F* statistic is used to determine whether the population variances can be assumed equal. If *F* is not statistically significant (generally at an alpha of .05), then the null of equality is retained and you should use the equal/pooled variances *t* ; if the null is rejected at .05, the unequal/separate variances test is the one to look at. Tests for differences in proportions must be done by hand.

Figure 11.7 Running a Dependent-Samples *t* Test in SPSS
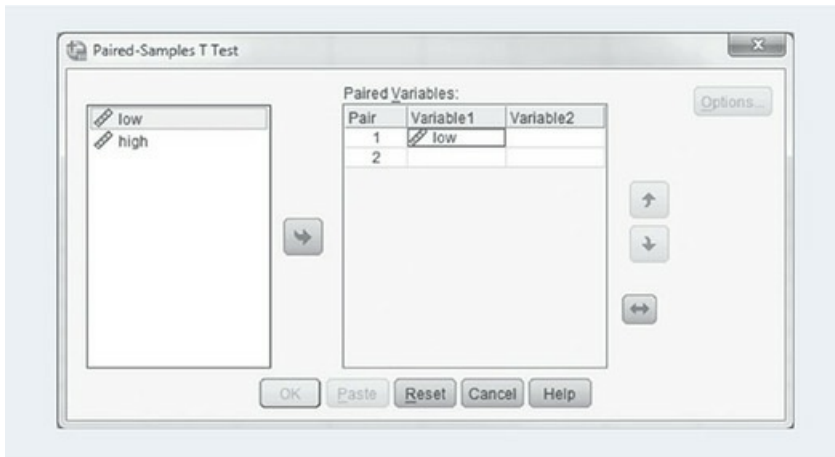
Figure 11.8 SPSS Output for Dependent (Paired) Samples *t* Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Assault rate—low-percentage female staff Assault rate—high-percentage female staff	.42400	.98333	.43976	-.79696	1.64496	.964	4	.390

As always, GIGO! It is your responsibility to make sure that you select the proper test and run that test correctly. SPSS is pretty unforgiving—it will not alert you to errors unless those errors are so serious that the requested analysis simply cannot be run at all. If you make a mistake, such as using dependent samples when you should use independent samples, SPSS will give you a result that resembles the serial killers (mentioned in Research Example 11.1) that opened this chapter: It looks normal on the surface, but it is actually untrustworthy and potentially dangerous.

Chapter Summary

In this chapter, you learned several types of analyses that can be conducted to test for differences between two populations. These can be used to test for differences between means or for differences between proportions; each type of difference has its own sampling distribution. The *t* distribution can be used for means tests and the *z* for proportions tests.

When you approach a hypothesis-testing question involving a two-class, categorical IV and a continuous DV, you first have to ask yourself whether the two samples are independent. If they are not, the dependent-samples *t* test must be used. This would be the case if the sample consists of matched pairs or is a before-and-after design. If the samples are not matched pairs or repeated measures, and are instead independent, the next step is to decide whether the population variances are equal. When you use SPSS, the program automatically produces Levene's test for equality of variances, so you will know whether you should look to the results of the *t* test assuming equality or that in which the assumption of equality has been rejected. If the DV under examination is a proportion, then the two-population tests for differences in proportions should be used. As long as certain criteria are met, the *z* distribution can be employed to determine if two proportions are significantly different from one another.

Thinking Critically

1. Suppose your friend ran a *t* test comparing inmate-on-staff assault rates in prisons with high and low percentages of female staff and found that prisons with higher levels of female staff also had higher assault rates. Does this mean that female correctional staff compromise security? Why or why not? If not, what is the flaw in this conclusion? Refer back to Chapter 2 if needed.

2. A researcher evaluating a policing strategy intended to reduce burglary computes the mean rate of burglary in the city before the strategy was implemented and the mean rate after the program has been in place for six months, then runs an independent-samples *t* test to determine if there is a significant difference. What mistake has this researcher made? How should the researcher collect the data instead, and what kind of *t* test is appropriate?

Review Problems

1. A researcher wants to test the hypothesis that defendants who plead guilty are sentenced more leniently than those who insist on going to trial. The researcher measures the plea decision as *guilty plea or trial* and the sentence as the number of months of incarceration to which defendants were sentenced.
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?

2. A researcher wants to test the hypothesis that defendants who are released on bail are more likely to be convicted than those who are unable to post bail or are denied it. The researcher measures pretrial release as *released or detained* and calculates the proportion within each group that is convicted.
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?

3. A researcher wants to test the hypothesis that male and female judges differ in the severity of the sentences they hand down. The researcher gathers a sample of judges and measures gender as *male or female* and sentence severity as the number of prison sentences they issue in 1 year.
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?

4. A researcher wants to test the hypothesis that police are less likely to arrest a suspect if the suspect has a full-time job. The researcher gathers a sample and measures job status as *employed or unemployed* and records the proportion of each group that was arrested.
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?

5. In Question 1, the researcher studying guilty pleas and sentence severity would use what kind of statistical test?
    1. Test for differences between means
    2. Test for differences between proportions

6. In Question 2, the researcher studying pretrial release and conviction likelihood would use what kind of statistical test?
    1. Test for differences between means
    2. Test for differences between proportions

7. In Question 3, the researcher studying judges' gender and sentencing behavior would use what kind of statistical test?
    1. Test for differences between means
    2. Test for differences between proportions

8. In Question 4, the researcher studying suspect unemployment status and arrest would use what kind of statistical test?
    1. Test for differences between means
    2. Test for differences between proportions

9. Tests for differences between two population means use the ___ distribution, and tests for differences between population

proportions use the ___ distribution.

10. A researcher wishes to find out whether a new drug court program appears to be effective at reducing drug use among participants. He gathers a random sample of drug defendants who are about to enter the drug court's treatment regimen and measures the number of times per month that they use drugs. After the participants finish the program, the researcher again measures their monthly drug use. Which type of *t* test would be appropriate for analyzing the data?
    1. Independent samples, pooled variances
    2. Independent samples, separate variances
    3. Dependent samples

11. A researcher is investigating the relationship between the restrictiveness of gun laws and gun-crime rates. She gathers a sample of states and divides them into two groups: *strict gun laws or lax gun laws*. She then calculates the gun crime rate in each state. She finds that the two groups have unequal variances. Which type of *t* test would be appropriate for analyzing the data?
    1. Independent samples, pooled variances
    2. Independent samples, separate variances
    3. Dependent samples

12. A researcher wishes to test the hypothesis that attorney type affects the severity of the sentence a defendant receives. He gathers a sample of defendants and records attorney type (*publicly funded; privately retained*) and the number of days in the jail or prison sentence. He finds that the groups have equal variances. Which type of *t* test would be appropriate for analyzing the data?
    1. Independent samples, pooled variances
    2. Independent samples, separate variances
    3. Dependent samples

13. In Research Example 11.1, you learned of a study by Wright et al. (2008) in which the researchers set out to determine whether MHOs were diverse in the number and types of crimes they commit or whether, instead, they tend to specialize in killing. The researchers compared MHOs to SHOs for purposes of this study. MHOs (Sample 1; N = 155) had a mean diversity index score of .36 (s = .32) and SHOs (Sample 2; N = 463) had a mean of .37 (s = .33). Using an alpha level of .05, test the hypothesis that the two groups' means are significantly different. Assume equal population variances. Use all five steps.

14. Are juvenile defendants who are released pending adjudication processed more slowly than those who are detained? The JDCC data set records whether a juvenile was released and the number of months it took for that juvenile's case to reach a disposition. The sample will be narrowed to black female youth. Released juveniles (Sample 1) had a mean of 4.30 months to disposition (s = 3.86, N = 123) and detained juveniles (Sample 2) had a mean of 3.57 (s = 4.20, N = 64). Using an alpha level of .01, test the hypothesis that released juveniles' mean time-to-adjudication is significantly greater than detained juveniles' mean. Assume equal population variances. Use all five steps.

15. Do juveniles transferred to adult court get treated more, or less, harshly depending on their age? The JDCC data set contains information about juveniles' age at arrest and the sentence length received by those sent to jail pursuant to conviction. Those male juveniles who were younger than 16 at the time of arrest (N = 85) received a mean of 68.84 days in jail (s = 125.59) and those who were older than 16 at arrest (N = 741) had a jail sentence mean of 95.24 days (s = 146.91). Using an alpha level of .05, test the hypothesis that the over-16 group received a significantly longer mean sentence compared to those younger than 16 at the time of arrest. Assume unequal population variances. Use all five steps.

16. One critique of sentencing research is that it generally focuses on the differences between white defendants and defendants of color and rarely examines differences between minority groups. The JDCC data set can be used to test for differences between the sentences received by black and Hispanic juveniles convicted of property crimes. Black juveniles (Sample 1; N = 229) were sentenced to a mean of 31.09 months of probation (s = 15.42), and Hispanic juveniles (Sample 2; N = 118) received a mean of 40.84 months (s = 16.45). Using an alpha level of .01, test the hypothesis that there is a statistically significant difference between the two group means. Assume unequal population variances. Use all five steps.

17. Do property and drug offenders get sentenced differently? The JDCC data show that among male juveniles convicted and fined, those convicted of drug offenses (Sample 1) had a mean fine of \$674.78 (s = 867.94; N = 160) and that those convicted of property offenses had a mean fine of \$344.91 (s = 251.91; N = 181). Using an alpha level of .05, test the hypothesis that there is a statistically significant difference between the two groups' means. Assume equal population variances. Use all five steps.

18. In Question 16, you tested for a relationship between black and Hispanic juveniles in terms of the length of probation

sentences received. Now let's find out whether there is a between-group difference among those sentenced to community service. The following table shows pairs of youths matched on gender (male) and offense type (weapons). Using an alpha level of .05, test the hypothesis that there is a statistically significant difference between the groups' means. Use all five steps.

| Pair | Community-Service Sentence (Hours) | |
	Black	Hispanic
Pair A	25	30
Pair B	50	40
Pair C	48	50
Pair D	100	110
Pair E	40	30

19. One of the most obvious potential contributors to the problem of assaults against police officers is exposure—all else being equal, jurisdictions wherein officers make more arrests might have elevated rates of officer assaults relative to lower-arrest jurisdictions. The Uniform Crime Reports (UCR) offer state-level information on arrest rates and officer assault rates. The states in the two samples in the following table were selected based on key similarities; that is, they are all in the western region of the country, have similar statewide violent crime rates, and have similar populations. The difference is that the states in the first sample have relatively low arrest rates and those in the second sample have relatively high arrest rates. The table shows the officer assault rate (number of officers assaulted per 1,000 officers) in each pair of states. Using an alpha level of .05, test the hypothesis that there is a statistically significant difference between the groups' means. Use all five steps.

| Pair | Officer Assaults per 1,000 Officers | |
	Low Arrest Rate	High Arrest Rate
Pair A	2.54	2.28
Pair B	4.72	4.08
Pair C	2.75	2.28
Pair D	2.55	2.13
Pair E	3.31	3.38

20. Among juvenile victims of gunshots, are males older than females? The Firearm Injury Surveillance Study (FISS) can be used to derive a set of pairs matched on juvenile status (younger than 18 years), firearm type (handgun), and circumstances of the incident (intentional assault involving drugs). The table lists the ages of each pair. Using an alpha level of .01, test the hypothesis that male victims are significantly older than female victims. Use all five steps.

| Pair | Victim Age | |
	Female	Male
Pair A	12	14
Pair B	17	17
Pair C	15	17
Pair D	16	14
Pair E	16	15
Pair F	13	13

21. The American Bar Association recommends that defendants who obtain pretrial release should have their cases disposed of within 180 days of their first court appearance. One question with regard to the speed of case processing is whether attorney type matters. Some evidence suggests that publicly appointed attorneys move cases faster than their privately retained counterparts do, other evidence points toward the opposite conclusion, and some studies find no difference. The JDCC data set contains information on attorney type, pretrial release, and days to adjudication. The sample consists of juveniles charged with drug felonies who were granted preadjudication release. Among those juveniles represented by public attorneys (Sample 1; $N = 509$), .64 had their cases disposed of in 180 days or less, while .32 of the juveniles who retained private attorneys

334

(Sample 2; N = 73) were adjudicated within 180 days. Using an alpha level of .05, test the hypothesis that there is a significant difference between the proportions. Use all five steps.

22. Are male victims of crime-involved shootings more likely than female victims to be shot by a stranger? The FISS captures data on the circumstances and on victims of shootings that took place in the course of a criminal event. Among males (Sample 1; N = 390), 72% were shot by strangers, and among females (Sample 2; N = 84), 56% were shot by strangers. Using an alpha level of .05, test the hypothesis that males are significantly more likely than females to be shot by strangers. Use all five steps.

23. Do pedestrian stops vary in duration depending on the time of day at which they take place? The data set from the Police–Public Contact Survey (PPCS) *PPCS Independent Samples for Chapter 11.sav* (www.sagepub.com/gau) contains the variables *time* and *minutes* , which measure whether a stop took place during the day or at night and the number of minutes the stop lasted. Run an independent-samples t test to determine whether these two variables are related.
    1. At an alpha level of .05, will you use the results for the pooled/equal variances t or that for separate/unequal variances? How did you make this decision?
    2. What is the obtained value of t ?
    3. Would you reject the null at an alpha level of .01? Why or why not?
    4. What is your substantive conclusion? In other words, are there significant differences between the two IV groups? Are the IV and the DV related?

24. Do juveniles facing charges for violent crimes have more total charges filed against them compared to juveniles charged with property crimes? The JDCC data set contains information on charge type and total charges. The data file *JDCC Independent Samples for Chapter 11.sav* (www.sagepub.com/gau) contains the variables *offense* and *charges*. Run an independent-samples t test to determine whether offense type appears to be related to the total number of charges.
    1. At an alpha level of .05, will you use the results for the pooled/equal variances t or that for separate/unequal variances? How did you make this decision?
    2. What is the obtained value of t ?
    3. Would you reject the null at an alpha level of .05? Why or why not?
    4. What is your substantive conclusion? In other words, are there significant differences between the two IV groups? Are the IV and the DV related?

25. Critics of the use of DNA evidence in criminal trials sometimes argue that DNA usage would clog up courts due to the delay caused by waiting for labs to return test results. Supporters claim, though, that any such delays would be justified by the improvement in the accuracy of felony convictions and acquittals. The Census of State Court Prosecutors is contained in the file *CSCP Independent Samples for Chapter 11.sav*. The variable *DNA* measures whether a prosecutor's office uses DNA evidence in plea negotiations or in criminal trials, and the variable *convictions* displays the number of felony convictions the office obtained within the past year. Run an independent-samples t test.
    1. At an alpha level of .05, will you use the results for the pooled/equal variances t or that for separate/unequal variances? How did you make this decision?
    2. What is the obtained value of t ?
    3. Would you reject the null at an alpha level of .01? Why or why not?
    4. What is your substantive conclusion? In other words, are there significant differences between the two IV groups? Are the IV and DV related?

26. In Question 23, we did not account for PPCS respondents' demographic characteristics. We can ask the same question—whether stop duration varies across day and night—and this time narrow the sample down by gender and race to create a subsample of white males. This strategy removes any potentially confounding effects of gender and race, and allows us to isolate the effects of the day/night variable. Using the data set *PPCS Matched Pairs for Chapter 11.sav* (www.sagepub.com/gau), run a dependent-samples t test using the *day* and *night* variables.
    1. What is the obtained value of t ?
    2. Would you reject the null at an alpha level of .05? Why or why not?
    3. What is your substantive conclusion? In other words, are there significant differences between the two IV groups? Are the IV and the DV related?

Note : Answers to review problems in this and subsequent chapters might vary depending on the number of steps used and whether rounding is employed during the calculations. The answers provided in this book's key were derived using the procedures illustrated in the main text.

Key Terms

t test 247

Independent samples 249

Dependent samples 249

Matched-pairs design 250

Repeated-measures design 250

Pooled variances 250

Separate variances 250

One-tailed tests 251

Glossary of Symbols and Abbreviations Introduced in This Chapter

t_{crit}	The critical value of t
t_{obt}	The obtained value of t
$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}$	The standard error of the sampling distribution of differences between means
x_D	A raw difference score; used with dependent-samples t
$\bar{x}_D$	The mean difference score; used with dependent-samples t
s_D	The standard deviation of the differences; used with dependent-samples t
$\hat{p}$	The sample proportion; used with tests for differences between proportions
$\hat{q}$	The complement of $\hat{p}$; used with tests for differences between proportions
$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}$	The standard error of the sampling distributions of differences between proportions

Chapter 12 Hypothesis Testing With Three or More Population Means Analysis of Variance

In Chapter 11, you learned how to determine whether a two-class categorical variable exerts an impact on a continuous outcome measure: This is a case in which a two-population *t* test for differences between means is appropriate. In many situations, though, a categorical independent variable (IV) has more than two classes. The proper hypothesis-testing technique to use when the IV is categorical with three or more classes and the dependent variable (DV) is continuous is **analysis of variance (ANOVA)**.

> **Analysis of variance (ANOVA)**: The analytic technique appropriate when an independent variable is categorical with three or more classes and a dependent variable is continuous.

As its name suggests, ANOVA is premised on variance. Why do we care about variance when testing for differences between means? Consider the hypothetical distributions displayed in Figure 12.1. The distributions have the same mean but markedly disparate variances—one curve is wide and flat, indicating substantial variance, whereas the other is tall and thin, indicating relatively little variance.

In any analysis of differences between means, the variance associated with each mean must be accounted for. This is what ANOVA does. It combines means and variances into a single test for significant differences between means. A rejected null indicates the presence of a relationship between an IV and a DV.

You might be wondering why, if we have a categorical IV and a continuous DV, we do not just use a series of *t* tests to find out if one or more of the means are different from the others. **Familywise error** is the primary reason that this is not a viable analytic strategy. Every time that you run a *t* test, there is a certain probability that the null is true (i.e., that there is no relationship between the IV and the DV) but will be rejected erroneously. This probability, as we saw in Chapter 9, is alpha, and the mistake is called a Type I error. Alpha (the probability of incorrectly rejecting a true null) attaches to each *t* test, so, in a series of *t* tests, the Type I error rate increases exponentially until the likelihood of mistake reaches an unacceptable level. This is the familywise error rate, and it is the reason that you should not run multiple *t* tests on a single sample.

> **Familywise error**: The increase in the likelihood of a Type I error (i.e., erroneous rejection of a true null hypothesis) that results from running repeated statistical tests on a single sample.

Another problem is that multiple *t* tests get messy. Imagine a categorical IV with classes A, B, C, and D. You

would have to run a separate t test for each combination (AB, AC, AD, BC, BD, CD). That is a lot of t tests! The results would be cumbersome and difficult to interpret.

Figure 12.1 Hypothetical Distributions With the Same Mean and Different Variances
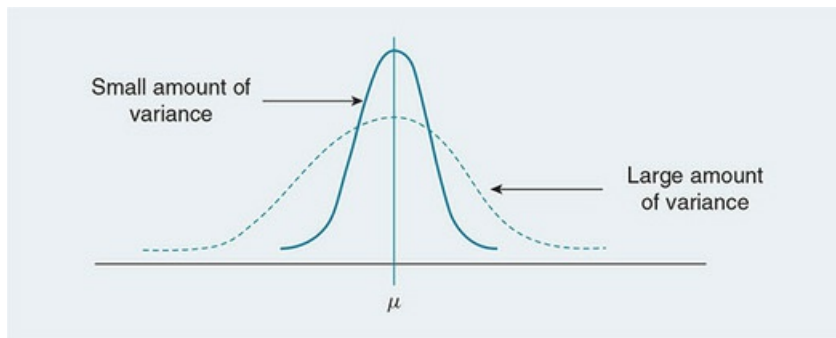
The ANOVA test solves the problems of familywise error and overly complicated output because ANOVA analyzes all classes on the IV simultaneously. One test is all it takes. This simplifies the process and makes for cleaner results.

ANOVA: Different Types of Variances

There are two types of variance analyzed in ANOVA. Both are based on the idea of groups, which are the classes on the IV. If an IV was *political orientation* measured as *liberal, moderate*, or *conservative*, then liberals would be a group, moderates would be a group, and conservatives would be a group. Groups are central to ANOVA.

The first type of variance is **between-group variance**. This is a measure of the similarity among or difference between the groups. It assesses whether groups are markedly different from one another or whether the differences are trivial and meaningless. This is a measure of true group effect. Figure 12.2 illustrates the concept of between-group variance. The groups on the left cluster closely together, while those on the right are distinctly different from one another.

> **Between-group variance**: The extent to each group or classes is similar to or different from the others in a sample. This is a measure of true group effect, or a relationship between the independent and dependent variables.

The second kind of variance is **within-group variance** and measures the extent to which people or objects differ from their fellow group members. Within-group variance is driven by random variations between people or objects and is a measure of error. Figure 12.3 depicts the conceptual idea behind within-group variance. The cases in the group on the left cluster tightly around their group's mean, whereas the cases in the right-hand group are scattered widely around their mean. The left-hand group, then, would be said to have much smaller within-group variability than the right-hand group.

> **Within-group variance**: The amount of diversity that exists among the people or objects in a single group or class. This is a measure of random fluctuation, or error.

The ANOVA test statistic—called the *F* **statistic** because the theoretical probability distribution for ANOVA is the *F* **distribution**—is a ratio that compares the amount of variance between groups to that within groups. When true differences between groups substantially outweigh the random fluctuations present within each group, the *F* statistic will be large and the null hypothesis that there is no IV–DV relationship will be rejected in favor of the alternative hypothesis that there is an association between the two variables. When between-group variance is small relative to within-group variance, the *F* statistic will be small, and the null will be retained.

> *F* **statistic**: The statistic used in ANOVA; a ratio of the amount of between-group variance present in a sample relative to the amount of within-group variance.
>
> *F* **distribution**: The sampling distribution for ANOVA. The distribution is bounded at zero on the left and extends to positive infinity; all values in the *F* distribution are positive.

Figure 12.2 Small and Large Between-Group Variability
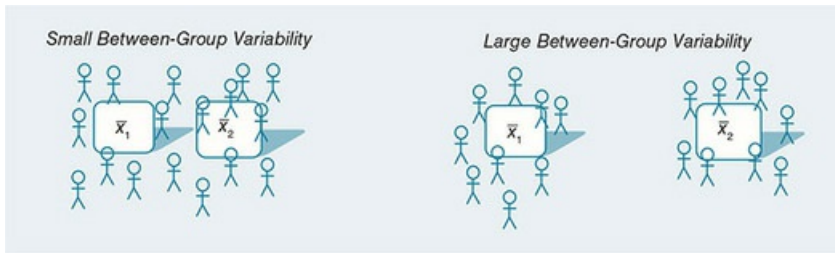
Figure 12.3 Small and Large Within-Group Variability
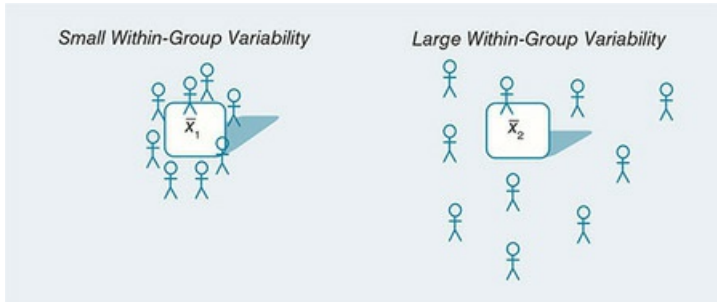
An example might help illustrate the concept behind the *F* statistic. Suppose we wanted to test the effectiveness of a mental-health treatment program on recidivism rates in a sample of probationers. We gather three samples: treatment-program completers, people who started the program and dropped out, and those who did not participate in the program at all. Our DV is the number of times each person is rearrested within 2 years of the end of the probation sentence. There will be some random fluctuations within each group; not everybody is going to have the same recidivism score. This is white noise or, more formally, within-group variance—in any sample of people, places, or objects, there will be variation. What we are attempting to discern is whether the difference *between* the groups outweighs the random variance among the people within each group. If the program is effective, then the treatment-completion group should have significantly lower recidivism scores than the other two groups. The impact of the program should be large relative to the random white noise. We might even expect the dropout group's recidivism to be significantly less than the no-treatment group (though probably not as low as the treatment completers). Figure 12.4 diagrams the two possible sets of results.

Figure 12.4 Recidivism Scores
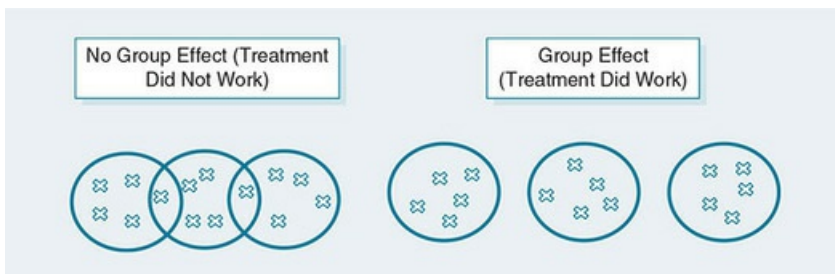
The *x*'s in the figure represent recidivism scores under two possible scenarios: that within-group variance trumps between-group variance, and that between-group variance is stronger than that within groups. The overlap depicted on the left side suggests that the treatment program was ineffective, since it failed to pull one or two groups away from the others. On the right side, the separation between the groups indicates that they are truly different; this implies that the treatment program did work and that those who completed or started

and dropped out are significantly different from each other and from the group that did not participate. An ANOVA test for the left side would yield a small *F* statistic, because the between-group variance is minimal compared to the within-group variance. An ANOVA for the right side, though, would produce a large (statistically significant) *F* because the ratio of between-to-within is high.

The *F* distribution is bounded on the left at zero, meaning it does not have a negative side. As a result, all critical and obtained values of *F* are positive; it is impossible for a correctly calculated *F* to be negative. This is because *F* is based on variance and variance cannot be negative.

Take a moment now to read Research Example 12.1, which describes a situation in which researchers would use ANOVA to test for a difference between groups or, in other words, would attempt to determine whether there is a relationship between a multiple- class IV and a continuous DV.

Franklin and Fearn's (2010) IV (*race* coded as *white; black; Hispanic; Asian*) was a four-class, categorical variable. Their DV (*sentence length*, measured in months) was continuous. ANOVA is the correct bivariate analysis in this situation.

Let's get into an example to see the ANOVA steps and calculations in action. We will use the Juvenile Defendants in Criminal Courts (JDCC; see Data Sources 11.1). We can examine whether attorney type (measured as public defender, assigned counsel, or private attorney) affects the jail sentences received by male youth convicted of weapons offenses. Table 12.1 shows the youths' sentences in months.

Research Example 12.1 Do Asian Defendants Benefit From a "Model Minority" Stereotype?

Numerous studies have found racially based sentencing disparities that are not attributable to differences in defendants' prior records or the severity of their instant offenses. Most such studies have focused on white, black, and Hispanic/Latino defendants. One area of the race-and-sentencing research that has received very little scholarly attention is the effect of race on sentencing among Asians. Franklin and Fearn (2010) set out to determine whether Asian defendants are treated differently from those of other races. They predicted that Asians would be sentenced more leniently due to the stereotype in the United States that Asians are a "model minority," in that they are widely presumed to be an economically, academically, and socially productive group.

To test the hypothesis that Asian defendants are given lighter sentences relative to similarly situated defendants of other races, Franklin and Fearn's (2010) DV was *sentence length*, which was coded as the number of months of incarceration imposed on offenders sentenced to jail or prison. The researchers reported the statistics shown in the table with respect to the mean sentence length across race in this sample.

So, what did the researchers find? It turned out that there were no statistically significant differences between the groups. Franklin and Fearn (2010) retained the null hypothesis that there is no relationship between race and sentencing, and concluded that Asian defendants do not, in fact, receive significantly shorter jail or prison sentences relative to other racial groups once relevant legal factors (e.g., offense type) are taken into account

	Defendant Race				Total
	White	Black	Hispanic	Asian	
Mean Sentence Length (Months)	11.80	17.40	16.50	16.10	15.50

Source: Adapted from Table 1 in Franklin and Fearn (2010).
Source: Adapted from Table 1 in Franklin and Fearn (2010).

We will conduct a five-step hypothesis test to determine whether defendants' conviction histories affect their sentences. Alpha will be set at .01.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

The null hypothesis in ANOVA is very similar to that in t tests. The difference is that now there are more than two means. The null is phrased as

H_0: $\mu_1 = \mu_2 = \mu_3$

The structure of the null is dependent on the number of groups—if there were four groups, there would be a μ_4 as well, and five groups would require the addition of a μ_5.

The alternative hypothesis in ANOVA is a bit different from what we have seen before because the only information offered by this test is whether at least one group is significantly different from at least one other group. The F statistic indicates neither the number of differences nor the specific group or groups that stand out from the others. The alternative hypothesis is, accordingly, rather nondescript. It is phrased as

H_1: *some $\mu_i \neq$ some μ_j*

Table 12.1 Jail Sentences (Months) of Male Juveniles Convicted of Weapons Offenses, by Attorney Type		
Attorney Type		
Public Defender (x_1)	Assigned Counsel (x_2)	Private Attorney (x_3)
1	4	3
2	2	8
9	3	10
3	6	9
6		12
		11
$n_1 = 5$	$n_2 = 4$	$n_3 = 6$

If the null is rejected in an ANOVA test, the only conclusion possible is that at least one group is markedly different from at least one other group—there is no way to tell which group is different or how many between-group differences there are. This is the reason for the existence of **post hoc tests,** which will be covered later in the chapter.

> **Post hoc tests**: Analyses conducted when the null is rejected in ANOVA in order to determine the number and location of differences between groups.

Step 2. Identify the distribution, and compute the degrees of freedom.

As aforementioned, ANOVA relies on the F distribution. This distribution is bounded on the left at zero (meaning it has only positive values) and is a family of curves whose shapes are determined by alpha and the degrees of freedom (df). There are two types of degrees of freedom in ANOVA: between-group (df_B) and within-groups (df_W). They are computed as

$$df_B = k - 1 \qquad \text{Formula 12(1)}$$
$$df_W = N - k, \qquad \text{Formula 12(2)}$$

where

N = the total sample size across all groups and

k = the number of groups.

The total sample size N is derived by summing the number of cases in each group, the latter of which are called *group sample sizes* and are symbolized n_k. In the present example, there are three groups ($k = 3$) and $N = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$. The degrees of freedom are therefore

Table 12.2 Elements of ANOVA			
Sample Sizes	**Means**	**Sums of Squares**	**Mean Squares**
n_k = the sample size of group k; the number of cases in each group	$\bar{x}_k$ = group mean; each group's mean on the DV	SS_B = between-groups sums of squares	MS_B = between-groups mean squares
N = the total sample size across all groups	$\bar{x}_G$ = the grand mean; the mean for the entire sample regardless of group	SS_W = within-groups sums of squares; SS_T = total sums of squares; $SS_B + SS_W = SS_T$	MS_W = within-groups mean squares

$df_B = 3 - 1 = 2$

$df_W = 15 - 3 = 12$

Step 3. Identify the critical value and state the decision rule.

The F distribution is located in Appendix E. There are different distributions for different alpha levels, so take care to ensure that you are looking at the correct one! You will find the between-group df across the top of the table and the within-group df down the side. The critical value is located at the intersection of the proper column and row. With $\alpha = .01$, $df_B = 2$, and $df_W = 12$, $F_{crit} = 6.93$. The decision rule is that *if $F_{obt} > 6.93$, the null will be rejected*. The decision rule in ANOVA is always phrased using a *greater than* inequality because the F distribution contains only positive values, so the critical region is always in the right-hand tail.

Step 4. Compute the obtained value of the test statistic.

Step 4 entails a variety of symbols and abbreviations, all of which are listed and defined in Table 12.2. Stop for

a moment and study this chart. You will need to know these symbols and what they mean in order to understand the concepts and formulas about to come.

You already know that each group has a sample size (n_k) and that the entire sample has a total sample size (N). Each group also has its own mean ($\overline{x}_k$), and the entire sample has a grand mean ($\overline{x}_G$). These sample sizes and means, along with other numbers that will be discussed shortly, are used to calculate the three types of sums of squares. The sums of squares are then used to compute mean squares, which, in turn, are used to derive the obtained value of F. We will first take a look at the formulas for the three types of sums of squares: total (SS_T), between-group (SS_B), and within-group (SS_W).

$$SS_T = \sum_i \sum_k x^2 - \frac{\left(\sum_i \sum_k x\right)^2}{N} \qquad \qquad \textit{Formula 12(3)}$$

where

$\sum_i$ = the sum of all scores i in group k,

$\sum_k$ = the sum of each group total across all groups in the sample,

x = the raw scores, and

N = the total sample size across all groups.

$$SS_B = \sum_k n_k \left(\overline{x}_k - \overline{x}_G\right)^2 \qquad \qquad \textit{Formula 12(4)}$$

where

n_k = the number of cases in group k,

$\overline{x}_k$ = the mean of group k, and

$\overline{x}_G$ = the grand mean across all groups.

$$SS_w = SS_T - SS_B \qquad \qquad \textit{Formula 12(5)}$$

The double summation signs in the SS_T formula are instructions to sum sums. The i subscript denotes individual scores and k signifies groups, so the double sigmas direct you to first sum the scores within each group and to then add up all the group sums to form a single sum representing the entire sample.

Sums of squares are measures of variation. They calculate the amount of variation that exists within and between the groups' raw scores, squared scores, and means. The SS_B formula should look somewhat familiar —in Chapters 4 and 5, we calculated deviation scores by subtracting the sample mean from each raw score. Here, we are going to subtract the grand mean from each group mean. See the connection? This strategy

produces a measure of variation. The sums of squares provide information about the level of variability within each group and between the groups.

The easiest way to compute the sums of squares is to use a table. What we ultimately want from the table are (a) the sums of the raw scores for each group, (b) the sums of each group's *squared* raw scores, and (c) each group's mean. All of these numbers are displayed in .

We also need the grand mean, which is computed by summing all of the raw scores across groups and dividing by the total sample size N, as such:

$$\overline{x}_G = \frac{\sum_i \sum_k x}{N}$$

<div align="right">*Formula 12(6)*</div>

Here,

$$\overline{x}_G = \frac{21+15+53}{5+4+6} = \frac{89}{15} = 5.93$$

Table 12.3 ANOVA Computation Table for *Attorney Type* and *Jail Sentence*

Attorney Type					
Public Defender (x_1)	x_1^2	Assigned Counsel (x_2)	x_2^2	Private Attorney (x_3)	x_3^2
1	1	4	16	3	9
2	4	2	4	8	64
9	81	3	9	10	100
3	9	6	36	9	81
6	36			12	144
				11	121
$n_1 = 5$		$n_2 = 4$		$n_3 = 6$	
$\Sigma x_1 = 21$	$\Sigma x_1^2 = 131$	$\Sigma x_2 = 15$	$\Sigma x_2^2 = 65$	$\Sigma x_3 = 53$	$\Sigma x_3^2 = 519$
$\overline{x}_1 = \frac{21}{5} = 4.20$		$\overline{x}_2 = \frac{15}{4} = 3.75$		$\overline{x}_3 = \frac{53}{6} = 8.83$	

With all of this information, we are ready to compute the three types of sums of squares, as follows. The process begins with SS_T:

$$SS_T = \sum_i \sum_k x^2 - \frac{(\sum_i \sum_k x)^2}{N} = (131+65+519) - \frac{(21+15+53)^2}{15}$$

$$= 715 - \frac{89^2}{15} = 715 - \frac{7,921}{15}$$

= 715 − 528.07

= 186.93

Then it is time for the between-groups sums of squares:

$$SS_B = \sum_k n_k \left(\bar{x}_k - \bar{x}_G \right)^2$$

$= 5(4.20 - 5.93)^2 = 4(3.75 - 5.93)^2 + 6(8.83 - 5.93)^2$

$= 5(-1.73)^2 + 4(-2.18)^2 + 6(2.90)^2$

$= 5(2.99) + 4(4.75) + 6(8.41)$

$= 14.95 + 19.00 + 50.46$

$= 84.41$

Next, we calculate the within-groups sums of squares:

$SS_W = SS_T - SS_B = 186.93 - 84.41 = 102.52$

We now have what we need to compute the mean squares (symbolized *MS*) . Mean squares transform sums of squares (measures of variation) into variances by dividing SS_B and SS_w by their respective degrees of freedom, df_B and df_w . This is a method of standardization. The mean squares formulas are

$$MS_B = \frac{SS_B}{k-1} \qquad\qquad Formula\ 12(7)$$

$$MS_W = \frac{SS_W}{N-k} \qquad\qquad Formula\ 12(8)$$

Plugging in our numbers,

$$MS_B = \frac{84.41}{3-1} = 42.21 \qquad MS_W = \frac{102.52}{15-3} = 8.54$$

We now have what we need to calculate F_{obt} . The *F* statistic is the ratio of between-group variance to within-group variance and is computed as

$$F_{obt} = \frac{MS_B}{MS_W} \qquad\qquad Formula\ 12(9)$$

Inserting the numbers from the present example,

$$F_{obt} = \frac{42.21}{8.54} = 4.94$$

Step 4 is done! F_{obt} = 4.94.

Step 5. Make a decision about the null hypothesis and state the substantive conclusion.

The decision rule stated that if the obtained value exceeded 6.93, the null would be rejected. With an F_{obt} of 4.94, the null is retained. The substantive conclusion is that there is no significant difference between the groups in terms of sentence length received. In other words, male juvenile weapons offenders' jail sentences do not vary as a function of the type of attorney they had. That is, attorney type does not influence jail sentences. This finding makes sense. Research is mixed with regard to whether privately retained attorneys (who cost

349

defendants a lot of money) really are better than publicly funded defense attorneys (who are provided to indigent defendants for free). While there is a popular assumption that privately retained attorneys are better, the reality is that publicly funded attorneys are frequently as or even more skilled than private ones are.

We will go through another ANOVA example. If you are not already using your calculator to work through the steps as you read and make sure you can replicate the results obtained here in the book, start doing so. This is an excellent way to learn the material.

For the second example, we will study handguns and murder rates. Handguns are a prevalent murder weapon and, in some locations, they account for more deaths than all other modalities combined. In criminal justice and criminology researchers' ongoing efforts to learn about violent crime, the question arises as to whether there are geographical differences in handgun-involved murders. Uniform Crime Report (UCR) data can be used to find out whether there are significant regional differences in handgun murder rates (calculated as the number of murders by handgun per 100,000 residents in each state). A random sample of states was drawn, and the selected states were divided by region. <u>Table 12.4</u> contains the data in the format that will be used for computations. Alpha will be set at .05.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: *some $\mu_i \neq$ some μ_j*

Step 2. Identify the distribution and compute the degrees of freedom.

This being an ANOVA, the *F* distribution will be employed. There are four groups, so *k* = 4. The total sample size is *N* = 5 + 5 + 7 + 6 = 23. Using Formulas 12(1) and 12(2), the degrees of freedom are

$df_B = 4 - 1 = 3$

$df_W = 23 - 4 = 19$

Step 3. Identify the critical value and state the decision rule.

With α = .05 and the earlier derived *df* values, F_{crit} = 3.13. The decision rule states that *if* F_{obt} > *3.13* , H_0 *will be rejected*.

Table 12.4 Handgun Murder Rate by Region

Northeast (x_1)	$(x_1)^2$	Midwest (x_2)	$(x_2)^2$	South (x_3)	$(x_3)^2$	West (x_4)	$(x_4)^2$
1.45	2.10	2.12	4.49	4.16	17.31	.14	.02
.30	.09	.10	.01	1.87	3.50	1.09	1.19
.71	.50	1.35	1.82	3.29	10.82	.92	.85
2.17	4.71	.66	.44	2.82	7.95	2.50	6.25
.00	.00	.15	.02	2.52	6.35	1.13	1.28
				2.67	7.13	.19	.04
				1.37	1.88		
$n_1 = 5$		$n_2 = 5$		$n_3 = 7$		$n_4 = 6$	
$\Sigma x_1 = 4.63$ $\Sigma x_1^2 = 7.40$		$\Sigma x_2 = 4.38$ $\Sigma x_2^2 = 6.78$		$\Sigma x_3 = 18.70$ $\Sigma x_3^2 = 54.94$		$\Sigma x_4 = 5.97$ $\Sigma x_4^2 = 9.63$	
$\bar{x}_1 = .93$		$\bar{x}_2 = .88$		$\bar{x}_3 = 2.67$		$\bar{x}_4 = 1.00$	

Step 4. Calculate the obtained value of the test statistic.

We begin by calculating the total sums of squares:

$$SS_T = (7.40 + 6.78 + 54.94 + 9.63) - \frac{(4.63 + 4.38 + 18.70 + 5.97)^2}{23}$$

$$= 78.75 - \frac{33.68^2}{23} = 78.75 - \frac{1,134.34}{23}$$

= 78.75 – 49.32

= 29.43

Before computing the between-groups sums of squares, we need the grand mean:

$$\bar{x}_G = \frac{4.63 + 4.38 + 18.70 + 5.97}{23} = \frac{33.68}{23} = 1.46$$

Now SS_B can be calculated:

$SS_B = 5(93 - 1\,46)^2 + 5(88 - 1\,46)^2 + 7(2\,67 - 1\,46)^2 + 6(1\,00 - 1\,46)^2$

$= 5(-.53)^2 + 5(-.58)^2 + 7(1.21)^2 + 6(-.46)^2$

= 5(.28) + 5(.34) + 7(1.46) + 6(.21)

= 1.40 + 1.70 + 10.22 + 1.26

351

= 14.58

Next, we calculate the within-groups sums of squares:

$$SS_w = 29.43 - 14.58 = 14.85$$

Plugging our numbers into Formulas 12(7) and 12(8) for mean squares gives

$$MS_B = \frac{14.58}{4-1} = 4.86 \quad MS_W = \frac{14.85}{23-4} = .78$$

Finally, using Formula 12(9) to derive F_{obt},

$$F_{obt} = \frac{4.86}{.78} = 6.23$$

This is the obtained value of the test statistic. $F_{obt} = 6.23$, and Step 4 is complete.

Step 5. Make a decision about the null and state the substantive conclusion .

In Step 3, the decision rule stated that if F_{obt} turned out to be greater than 3.13, the null would be rejected. Since F_{obt} ended up being 6.23, the null is indeed rejected. The substantive interpretation is that there is a significant difference across regions in the handgun-murder rate.

Research Example 12.2 Are Juveniles Who Are Transferred to Adult Courts Seen as More Threatening?

Recent decades have seen a shift in juvenile-delinquency policy. There has been an increasing zero tolerance sentiment with respect to juveniles who commit serious offenses. The reaction by most states has been to make it easier for juveniles to be tried as adults, which allows their sentences to be more severe than they would be in juvenile court. The potential problem with this strategy is that there is a prevalent stereotype about juveniles who get transferred or waived to adult court: They are often viewed as vicious, cold-hearted predators. Judges, prosecutors, and jurors might be biased against transferred juveniles, simply because they got transferred. This means that a juvenile and an adult could commit the same offense and yet be treated very differently by the court, potentially even ending up with different sentences.

Tang, Nuñez, and Bourgeois (2009) tested mock jurors' perceptions about the dangerousness of 16-year-olds who were transferred to adult court, 16-year-olds who were kept in the juvenile justice system, and 19-year-olds in adult court. They found that mock jurors rated transferred 16-year-olds as committing more serious crimes, being more dangerous, and having a greater likelihood of chronic offending relative to non-transferred juveniles and to 19-year-olds. The following table shows the means, standard deviations, and F tests.

	16-Year-Olds Transferred	16-Year-Olds Not Transferred	19-Year-Olds	
	Mean (sd)	Mean (sd)	Mean (sd)	F
Serious	7.44 (.24)	5.16 (.23)	5.66 (.24)	26.30
Dangerous	6.92 (.25)	4.76 (.25)	5.47 (.25)	19.45
Chronic	7.18 (.25)	5.74 (.24)	5.87 (.25)	10.22

Source: Adapted from Table 1 in Tang et al. (2009).
Source: Adapted from Table 1 in Tang et al. (2009).

As you can see, all the F statistics were large; the null was rejected for each test. The transferred juveniles' means are higher than the

As another example, we will analyze data from the Firearm Injury Surveillance Study (FISS; Data Sources 8.2) to find out whether victim age varies significantly across the different victim–offender relationships. There are four relationship categories, and a total sample size of 22. Table 12.5 shows the data and calculations of the numbers needed to complete the hypothesis test. We will proceed using the five steps. Alpha will be set at .05.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: *some $\mu_i \neq$ some μ_j*

Step 2. Identify the distribution and compute the degrees of freedom.

This being an ANOVA, the *F* distribution will be employed. There are four groups, so *k* = 4. The total sample size is *N* = 7 + 6 + 4 + 5 = 22. Using Formulas 12(1) and 12(2), the degrees of freedom are

df_B = 4 – 1 = 3

df_W = 22 – 4 = 18

Step 3. Identify the critical value and state the decision rule.

With α =.05 and the earlier derived *df* values, F_{crit} = 3.16. The decision rule states that *if F_{obt} > 3.16* , H_0 *will be rejected.*

Table 12.5 Victim Age and Victim–Offender Relationship

Stranger		Friend/Acquaintance		Intimate Partner		Relative	
(x_1)	$(x_1)^2$	(x_2)	$(x_2)^2$	(x_3)	$(x_3)^2$	(x_4)	$(x_4)^2$
29	841	19	361	45	2,025	23	529
31	961	25	625	36	1,296	22	484
24	576	22	484	39	1,521	19	361
27	729	24	576	40	1,600	18	324
33	1,089	30	900			25	625
28	784	25	625				
34	1,156						
$n_1 = 7$		$n_2 = 6$		$n_3 = 4$		$n_4 = 5$	
$\Sigma x_1 = 206$	$\Sigma x_1^2 = 6{,}136$	$\Sigma x_1 = 145$	$\Sigma x_2^2 = 3{,}571$	$\Sigma x_1 = 160$	$\Sigma x_3^2 = 6{,}442$	$\Sigma x_4^2 = 107$	$\Sigma x_4^2 = 2{,}323$
$\bar{x}_1 = 29.43$		$\bar{x}_2 = 24.17$		$\bar{x}_3 = 40.00$		$\bar{x}_4 = 21.40$	

Step 4. Calculate the obtained value of the test statistic.

The total sums of squares for the data in [Table 12.5](#) is

$$
SS_T = \sum_i \sum_k x^2 - \frac{\left(\sum_i \sum_k x\right)^2}{N}
$$

$$
= (6{,}136 + 3{,}571 + 6{,}442 + 2{,}323) - \frac{(206 + 145 + 160 + 107)^2}{22}
$$

$$
= 18{,}472 - \frac{618.00^2}{22} = 18{,}472 - \frac{381{,}924}{22}
$$

= 18,472 – 17360.18

= 1,111.82

Next, we need the grand mean:

$$
\bar{x}_G = \frac{\sum_i \sum_k x}{N} = \frac{206 + 145 + 160 + 107}{22} = \frac{618.00}{22} = 28.09
$$

Now SS_B can be calculated:

$$
SS_B = \sum_k n_k \left(\bar{x}_k - \bar{x}_G\right)^2
$$

= 7(29 43 – 28 09)^2 + 6(24 17 – 28 09)^2 + 4(40 00 – 28 09)^2 + 5 (21 40 – 28.09)^2

$$=7(1.34)^2 + 6(-3.92)^2 + 4(11.91)^2 + 5(-6.69)^2$$

$$= 7(1.80) + 6(15.37) + 4(141.85) + 5(44.76)$$

$$= 12.60 + 92.22 + 567.40 + 223.80$$

$$= 896.02$$

Next, we calculate the within-groups sums of squares:

$$SS_w = 1,111.82 - 896.02 = 215.80$$

And the mean squares are

$$MS_B = \frac{SS_B}{k-1} = \frac{896.02}{4-1} = 298.67$$

$$MS_W = \frac{SS_W}{N-k} = \frac{215.80}{22-4} = 11.99$$

Finally, F_{obt} is calculated as

$$F_{obt} = \frac{MS_B}{MS_W} = \frac{298.67}{11.99} = 24.91$$

And Step 4 is done. F_{obt} = 24.91.

Step 5. Make a decision about the null and state the substantive conclusion.

In Step 3, the decision rule stated that if F_{obt} turned out to be greater than 3.16, the null would be rejected. Since F_{obt} is 24.91, we reject the null. It appears that victim age does vary across the different victim–offender relationship categories.

After finding a significant F indicating that at least one group stands out from at least one other one, the obvious question is, "Which group or groups are different?" We might want to know which region or regions have a significantly higher or lower rate than the others or which victim–offender relationship or relationships contain significantly younger or older victims. The F statistic is silent with respect to the location and number of differences, so post hoc tests are used to get this information. The next section covers post hoc tests and measures of association that can be used to gauge relationship strength.

When the Null Is Rejected: A Measure of Association and Post Hoc Tests

If the null is not rejected in ANOVA, then the analysis stops because the conclusion is that the IVs and DV are not related. If the null is rejected, however, it is customary to explore the statistically significant results in more detail using measures of association (MAs) and post hoc tests. Measures of association permit an assessment of the strength of the relationship between the IV and the DV, and post hoc tests allow researchers to determine which groups are significantly different from which other ones. The MA that will be discussed here is fairly easy to calculate by hand, but the post hoc tests will be discussed and then demonstrated in the SPSS section, because they are computationally intensive.

Omega squared (ω^2) is an MA for ANOVA that is expressed as the proportion of the total variability in the sample that is due to between-group differences. Omega squared can be left as a proportion or multiplied by 100 to form a percentage. Larger values of ω^2 indicate stronger IV–DV relationships, whereas smaller values signal weaker associations. Omega squared is computed as

$$\omega^2 = \frac{SS_B - (k-1)MS_W}{MS_W + SS_T} \qquad \textit{Formula 12(10)}$$

> **Omega squared**: A measure of association used in ANOVA when the null has been rejected in order to assess the magnitude of the relationship between the independent and dependent variables. This measure shows the proportion of the total variability in the sample that is attributable to between-group differences.

Earlier, we found a statistically significant relationship between region and handgun murder rates. Now we can calculate how strong the relationship is. Using ω^2,

$$\omega^2 = \frac{14.58 - (4-1).78}{.78 + 29.43} = \frac{14.58 - 2.34}{30.21} = \frac{12.24}{30.21} = .41$$

Omega squared shows that 41% of the total variability in the states' handgun-murder rates is a function of regional characteristics. Region appears to be a very important determinate of the prevalence of handgun murders.

We can do the same for the test showing significant differences in victims' ages across four different types of victim–offender relationships. Plugging the relevant numbers into Formula 12(10) yields

$$\omega^2 = \frac{896.02 - (4-1)11.99}{11.99 + 1,111.82} = \frac{896.02 - 35.97}{1,123.81} = \frac{860.05}{1,123.81} = .77$$

This means that 77% of the variability in victims' ages is attributable to the relationship between the victim and the shooter. This points to age being a function of situational characteristics. Younger people are more at risk of firearm injuries in certain types of situations, while older people face greater risk in other circumstances. Of course, we still do not know which group or groups are significantly different from which other group or groups. For this, post hoc tests are needed.

There are many different types of post hoc tests, so two of the most popular ones are presented here. The first is **Tukey's honest significant difference (HSD)**. Tukey's test compares each group to all the others in a series of two-variable hypothesis tests. The null hypothesis in each comparison is that both group means are equal; rejection of the null means that there is a significant difference between them. In this way, Tukey's is conceptually similar to a series of t tests, though the HSD method sidesteps the problem of familywise error.

> **Tukey's honest significant difference**: A widely used post hoc test that identifies the number and location(s) of differences between groups.

Bonferroni is another commonly used test and owes its popularity primarily to the fact that it is fairly conservative. This means that it minimizes Type I error (erroneously rejecting a true null) at the cost of increasing the likelihood of a Type II error (erroneously retaining a false null). The Bonferroni, though, has been criticized for being too conservative. In the end, the best method is to select both Tukey's and Bonferroni in order to garner a holistic picture of your data and make an informed judgment.

> **Bonferroni**: A widely used and relatively conservative post hoc test that identifies the number and location(s) of differences between groups.

The computations of both post hoc tests are complex, so we will not attempt them by hand and will instead demonstrate their use in SPSS.

Learning Check 12.2

✓

Would it be appropriate to compute omega squared and post hoc tests for the ANOVA in the example pertaining to juvenile defendants' attorneys and sentences? Why or why not?

Research Example 12.3 Does Crime Vary Spatially and Temporally in Accordance With Routine Activities Theory?

Crime varies across space and time; in other words, there are places and times it is more (or less) likely to occur. Routine activities theory has emerged as one of the most prominent explanations for this variation. Numerous studies have shown that the characteristics of places can attract or prevent crime and that large-scale patterns of human behavior shape the way crime occurs. For instance, a tavern in which negligent bartenders frequently sell patrons too much alcohol might generate alcohol-related fights, car crashes, and so on. Likewise, when schools let out for summer break, cities experience a rise in the number of unsupervised juveniles, many of whom get into mischief. Most of this research, however, has been conducted in Western nations. De Melo, Pereira, Andresen, and Matias (2017) extended the study of spatial and temporal variation in crime rates to Campinas, Brazil, to find out if crime appears to vary along these two dimensions. They broke crime down by type and ran ANOVAs to test for temporal variation across different units of time (season, month, day of week, hour of day). The table displays the results for the ANOVAs that were statistically significant. (Nonsignificant findings have been omitted.)

Unit of Time	Crime	F value	p Value
Season	Homicide	3.252	.059
Day of week	Homicide	3.821	.009
	Robbery	16.77	.000
	Burglary	3.88	.009
	Theft	39.98	.000
Hour of day	Homicide	3.05	.000
	Rape	3.14	.000
	Robbery	98.02	.000
	Burglary	18.09	.000
	Theft	101.00	.000

Source: Adapted from Table 1 in De Melo et al. (2017).

Source: Adapted from Table 1 in De Melo et al. (2017).

As the table shows, homicide rates vary somewhat across month. Post hoc tests showed that the summer months experienced spikes in homicide, likely because people are outdoors more often when the weather is nice, which increases the risk for violent victimization and interpersonal conflicts. None of the variation across month was statistically significant (which is why there are no rows for this unit of time in the table). There was significant temporal variation across weeks and hours. Post hoc tests revealed interesting findings across crime type. For example, homicides are more likely to occur on weekends (since people are out and about

more during weekends than during weekdays), while burglaries are more likely to happen on weekdays (since people are at work). The variation across hours of the day was also significant for all crime types, but the pattern was different within each one. For instance, crimes of violence were more common in late evenings and into the night, while burglary was most likely to occur during the daytime hours.

SPSS

Let us revisit the question asked in Example 2 regarding whether handgun murder rates vary by region. To run an ANOVA in SPSS, follow the steps depicted in [Figure 12.5](#). Use the *Analyze → Compare Means → One-Way ANOVA* sequence to bring up the dialog box on the left side in [Figure 12.5](#) and then select the variables you want to use. Move the IV to the *Factor* space and the DV to the *Dependent List*. Then click *Post Hoc* and select the *Bonferroni* and *Tukey* tests. Click *Continue* and *OK* to produce the output shown in [Figure 12.6](#).

The first box of the output shows the results of the hypothesis test. You can see the sums of squares, *df* , and mean squares for within groups and between groups. There are also total sums of squares and total degrees of freedom. The number in the *F* column is F_{obt} . Here, you can see that F_{obt} = 6.329. When we did the calculations by hand, we got 6.23. Our hand calculations had some rounding error, but this did not affect the final decision regarding the null because you can also see that the significance value (the *p* value) is .004, which is less than .05, the value at which α was set. The null hypothesis is rejected in the SPSS context just like it was in the hand calculations.

The next box in the output shows the Tukey and Bonferroni post hoc tests. The difference between these tests is in the *p* values in the *Sig.* column. In the present case, those differences are immaterial because the results are the same across both types of tests. Based on the asterisks that flag significant results and the fact that the *p* values associated with the flagged numbers are less than .05, it is apparent that the South is the region that stands out from the others. Its mean is significantly greater than all three of the other regions' means. The Northeast, West, and Midwest do not differ significantly from one another, as evidenced by the fact that all of their *p* values are greater than .05.

Figure 12.5 Running an ANOVA in SPSS
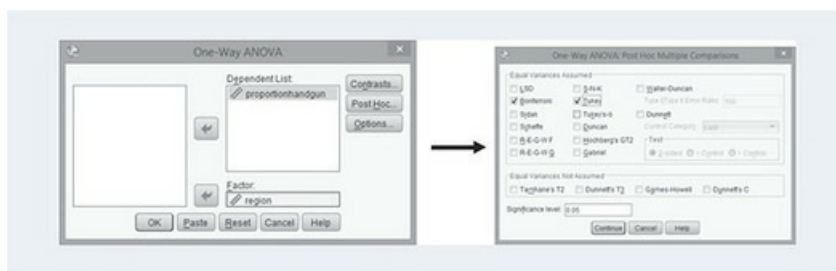
In [Figure 12.7](#), you can see that the F_{obt} SPSS produces (24.719) is nearly identical to the 12.91 we arrived at by hand. Looking at Tukey's and Bonferroni, it appears that the categories "relative" and "friend/acquaintance" are the only ones that do not differ significantly from one another. In the full data set, the mean age of victims shot by relatives is 21.73 and that for the ones shot by friends and acquaintances is 24.05. These means are not significantly different from each other, but they are both distinct from the means for stranger-perpetrated shootings (mean age of 29.58) and intimate-partner shootings (39.12).

Figure 12.6 ANOVA Output

Handgun murders per 100,000

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	14.674	3	4.891	6.329	.004
Within Groups	14.684	19	.773		
Total	29.358	22			

Post hoc Tests

Multiple Comparisons

Dependent Variable: Handgun Murders per 100,000

	(I) Region	(J) Region	Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	Northeast	Midwest	.05037	.55600	1.000	−1.5130	1.6138
		South	−1.74371*	.51475	.015	−3.1911	−.2963
		West	-0.6898	.53233	.999	−1.5658	1.4278
	Midwest	Northeast	−.05037	.55600	1.000	−1.6138	1.5130
		South	−1.79408*	.51475	.012	−3.2415	−.3467
		West	−.11935	.53233	.996	−1.6162	1.3775
	South	Northeast	1.74371*	.51475	.015	.2963	3.1911
		Midwest	1.79408*	.51475	.012	.3467	3.2415
		West	1.67473*	.48909	.014	.2995	3.0500
	west	Northeast	.06898	.53233	.999	−1.4278	1.5658
		Midwest	.11935	.53233	.996	−1.3775	1.6162
		South	−1.67473*	.48909	.014	−3.0500	−.2995
Bonferroni	Northeast	Midwest	.05037	.55600	1.000	−1.5864	1.6872
		South	−1.74371*	.51475	.019	−3.2591	−.2283
		West	−.06899	.53233	1.000	−1.6361	1.4991
	Midwest	Northeast	−05037	.55600	1.000	−1.6872	1.5864
		South	−1.79408*	.51475	.015	−3.3095	−.2787
		West	−.11935	.53233	1.000	−1.6865	1.4478
	South	Northeast	1.74371*	.51475	.019	.2283	3.2591
		Midwest	1.79408*	.51475	.015	.2787	3.3095
		West	1.67473*	.48909	.017	.2349	3.1146
	West	Northeast	.06898	.53233	1.000	−1.4981	1.6361
		Midwest	.11935	.53233	1.000	−1.4478	1.6865
		South	−1.67473*	.48909	.017	−3.1146	−.2349

*The mean difference is significant at the 0.05 level.

We can also use SPSS and the full FISS to reproduce the analysis we did by hand using a sample of cases. Figure 12.7 shows the ANOVA and post hoc tests.

*The mean difference is significant at the 0.05 level.

We can also use SPSS and the full FISS to reproduce the analysis we did by hand using a sample of cases. Figure 12.7 shows the ANOVA and post hoc tests.

Figure 12.7 ANOVA Output

ANOVA

Victim's age

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11795.008	3	3931.669	24.719	.000
Within Groups	163827.511	1030	159.056		
Total	175622.518	1033			

Post HOC Tests

Multiple Comparisons

Dependent Variable: Victim's age

	(I) Offender's relationship to victim	(J) Offender's relationship to victim	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	Stranger	Friend/acq	5.524*	.970	.000	3.03	8.02
		Intimate partner	-9.542*	2.567	.001	-16.15	-2.94
		Relative	7.848*	1.420	.000	4.19	11.50
	Friend/acq	Stranger	-5.524*	.970	.000	-8.02	-3.03
		Intimate partner	-15.066*	2.660	.000	-21.91	-8.22
		Relative	2.323	1.581	.456	-1.75	6.39
	Intimate partner	Stranger	9.542*	2.567	.001	2.94	16.15
		Friend/acq	15.066*	2.660	.000	8.22	21.91
		Relative	17.390*	2.855	.000	10.04	24.74
	Relative	Stranger	-7.848*	1.420	.000	-11.50	-4.19
		Friend/acq	-2.323	1.581	.456	-6.39	1.75
		Intimate partner	-17.390*	2.855	.000	-24.74	-10.04
Bonferroni	Stranger	Friend/acq	5.524*	.970	.000	2.96	8.09
		Intimate partner	-9.542*	2.660	.001	-16.33	-2.76
		Relative	7.848*	1.581	.000	4.10	11.60
	Friend/acq	Stranger	-5.524*	.970	.000	-8.09	-2.96
		Intimate partner	-15.066*	2.660	.001	-22.10	-8.03
		Relative	2.323	1.581	.000	-1.86	6.50
	Intimate partner	Stranger	9.542*	2.567	.000	2.76	16.33
		Friend/acq	15.066*	2.660	.000	8.03	22.10
		Relative	17.390*	2.855	.852	9.84	24.94
	Relative	Stranger	-7.848*	1.420	.001	-11.60	-4.10
		Friend/acq	-2.323	1.581	.000	-6.50	1.86
		Intimate partner	-17.390*	2.855	.000	-24.94	-9.84

*The mean difference is significant at the 0.05 level.

*The mean difference is significant at the 0.05 level.

Chapter Summary

This chapter taught you what to do when you have a categorical IV with three or more classes and a continuous DV. A series of t tests in such a situation is not viable because of the familywise error rate. In an analysis of variance, the researcher conducts multiple between-group comparisons in a single analysis. The F statistic compares between-group variance to within-group variance to determine whether between-group variance (a measure of true effect) substantially outweighs within-group variance (a measure of error). If it does, the null is rejected; if it does not, the null is retained.

The ANOVA F, though, does not indicate the size of the effect, so this chapter introduced you to an MA that allows for a determination of the strength of a relationship. This measure is omega squared (w^2), and it is used only when the null has been rejected—there is no sense in examining the strength of an IV–DV relationship that you just said does not exist! Omega squared is interpreted as the proportion of the variability in the DV that is attributable to the IV. It can be multiplied by 100 to be interpreted as a percentage.

The F statistic also does not offer information about the location or number of differences between groups. When the null is retained, this is not a problem because a retained null means that there are no differences between groups; however, when the null is rejected, it is desirable to gather more information about which group or groups differ from which others. This is the reason for the existence of post hoc tests. This chapter covered Tukey's HSD and Bonferroni, which are two of the most commonly used post hoc tests in criminal justice and criminology research. Bonferroni is a conservative test, meaning that it is more difficult to reject the null

hypothesis of no difference between groups. It is a good idea to run both tests and, if they produce discrepant information, make a reasoned judgment based on your knowledge of the subject matter and data. Together, MAs and post hoc tests can help you glean a comprehensive and informative picture of the relationship between the independent and dependent variables.

Thinking Critically

1. What implications does the relationship between shooting victims' ages and these victims' relationships with their shooters have for efforts to prevent firearm violence? For each of the four categories of victim–offender relationship, consider the mean age of victims and devise a strategy that could be used to reach people of this age group and help them lower their risks of firearm victimization.

2. A researcher is evaluating the effectiveness of a substance abuse treatment program for jail inmates. The researcher categorizes inmates into three groups: those who completed the program, those who started it and dropped out, and those who never participated at all. He follows up with all people in the sample six months after their release from jail and asks them whether or not they have used drugs since being out. He codes drug use as 0 = no and 1 = yes. He plans to analyze the data using an ANOVA. Is this the correct analytic approach? Explain your answer.

Review Problems

1. A researcher wants to know whether judges' gender (measured as *male; female*) affects the severity of sentences they impose on convicted defendants (measured as *months of incarceration*) . Answer the following questions:
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?
    5. What type of hypothesis test should the researcher use?

2. A researcher wants to know whether judges' gender (measured as *male; female*) affects the types of sentences they impose on convicted criminal defendants (measured as *jail; prison; probation; fine; other*) . Answer the following questions:
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?
    5. What type of hypothesis test should the researcher use?

3. A researcher wishes to find out whether arrest deters domestic violence offenders from committing future acts of violence against intimate partners. The researcher measures arrest as *arrest; mediation; separation; no action* and recidivism as *number of arrests for domestic violence within the next 3 years*. Answer the following questions:
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?
    5. What type of hypothesis test should the researcher use?

4. A researcher wishes to find out whether arrest deters domestic violence offenders from committing future acts of violence against intimate partners. The researcher measures arrest as *arrest; mediation; separation; no action* and recidivism as whether these offenders were arrested for domestic violence within the next 2 years (measured as *arrested; not arrested*) . Answer the following questions:
    1. What is the independent variable?
    2. What is the level of measurement of the independent variable?
    3. What is the dependent variable?
    4. What is the level of measurement of the dependent variable?
    5. What type of hypothesis test should the researcher use?

5. A researcher wants to know whether poverty affects crime. The researcher codes neighborhoods as being *lower-class, middle-class* , or *upper-class* and obtains the crime rate for each area (measured as the number of index offenses per 10,000 residents). Answer the following questions:

1. What is the independent variable?
2. What is the level of measurement of the independent variable?
3. What is the dependent variable?
4. What is the level of measurement of the dependent variable?
5. What type of hypothesis test should the researcher use?

6. A researcher wants to know whether the prevalence of liquor-selling establishments (such as bars and convenience stores) in neighborhoods affects crime in those areas. The researcher codes neighborhoods as having *0–1* , *2–3* , *4–5* , or *6 +* liquor-selling establishments. The researcher also obtains the crime rate for each area (measured as the number of index offenses per 10,000 residents). Answer the following questions:
   1. What is the independent variable?
   2. What is the level of measurement of the independent variable?
   3. What is the dependent variable?
   4. What is the level of measurement of the dependent variable?
   5. What type of hypothesis test should the researcher use?

7. Explain within-groups variance and between-groups variance. What does each of these concepts represent or measure?
8. Explain the F statistic in conceptual terms. What does it measure? Under what circumstances will F be small? Large?
9. Explain why the F statistic can never be negative.
10. When the null hypothesis in an ANOVA test is rejected, why are MA and post hoc tests necessary?
11. The Omnibus Crime Control and Safe Streets Act of 1968 requires state and federal courts to report information on all wiretaps sought by and authorized for law enforcement agencies (Duff, 2010). One question of interest to someone studying wiretaps is whether wiretap use varies by crime type; that is, we might want to know whether law enforcement agents use wiretaps with greater frequency in certain types of investigations than in other types. The following table contains data from the U.S. courts website (www.uscourts.gov/Statistics.aspx) on the number of wiretaps sought by law enforcement agencies in a sample of states. The wiretaps are broken down by offense type, meaning that each number in the table represents the number of wiretap authorizations received by a particular state for a particular offense. Using an alpha level of .05, test the null hypothesis of no difference between the group means against the alternative hypothesis that at least one group mean is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Offense Type		
Homicide and Assault (x_1)	Narcotics (x_2)	Racketeering (x_3)
2	25	1
0	1	1
0	4	0
1	2	3
14	21	0
1	3	0
2	12	0
$n_1 = 7$	$n_2 = 7$	$n_3 = 7$

12. Some studies have found that people become more punitive as they age, such that older people, as a group, hold harsher attitudes toward people who commit crimes. The General Social Survey (GSS) asks people for their opinions about courts' handling of criminal defendants. This survey also records respondents' ages. Use the data below and an alpha level of .05 to test the null hypothesis of no difference between the group means against the alternative hypothesis that at least one group mean is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Courts' Handling of Criminal Defendants		
Too Harsh (x_1)	Not Harsh Enough (x_2)	About Right (x_3)
47	50	49
46	56	48
44	52	50
41	46	47
39	49	52
50	47	49
$n_1 = 6$	$n_2 = 6$	$n_3 = 6$

13. In the ongoing effort to reduce police injuries and fatalities resulting from assaults, one issue is the technology of violence against officers or, in other words, the type of implements offenders use when attacking police. Like other social events, weapon use might vary across regions. The UCRs collect information on weapons used in officer assaults. These data can be used to find out whether the percentage of officer assaults committed with firearms varies by region. The following table contains the data. Using an alpha level of .01, test the null of no difference between means against the alternative that at least one region is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Region			
Northeast (x_1)	Midwest (x_2)	South (x_3)	West (x_4)
.76	1.05	2.86	3.55
1.53	5.28	2.41	4.52
2.65	4.92	3.49	3.64
.00	.96	2.12	2.29
.23	1.41	3.39	3.88
$n_1 = 5$	1.50	$n_3 = 5$	4.90
	$n_2 = 6$		.68
			$n_4 = 7$

14. An ongoing source of question and controversy in the criminal court system are the possible advantages that wealthier defendants might have over poorer ones, largely as a result of the fact that the former can pay to hire their own attorneys, whereas the latter must accept the services of court-appointed counsel. There is a common perception that privately retained attorneys are more skilled and dedicated than their publicly appointed counterparts. Let us examine this issue using a sample of property defendants from the JDCC data set. The IV is *attorney type* and the DV is *days to pretrial release* , which measures the number of days between arrest and pretrial release for those rape defendants who were released pending trial. (Those who did not make bail or were denied bail are not included.) Using an alpha level of .05, test the null of no difference between means against the alternative that at least one region is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Attorney Type		
Public Defender (x_1)	Assigned Counsel (x_2)	Private Attorney (x_3)
2	0	0
42	0	0
5	6	0
4	51	1
8	5	3
1	5	24
0	$n_2 = 6$	5
$n_1 = 7$		34
		$n_3 = 8$

15. In , we read about a study that examined whether Asian defendants were sentenced more leniently than offenders of other races. Let us run a similar test using data from the JDCC. The following table contains a sample of juveniles convicted of property offenses and sentenced to probation. The IV is *race* , and the DV is each person's *probation sentence in months*. Using an alpha level of .01, test the null of no difference between means against the alternative that at least one region is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Race			
Asian (x_1)	Black (x_2)	White (x_3)	Other (x_4)
3	12	6	2
9	10	18	6
14	2	3	18
8	6	2	12
24	72	24	$n_3 = 4$
12	$n_2 = 5$	$n_3 = 5$	
$n_1 = 6$			

16. Across police agencies of different types, is there significant variation in the prevalence of bachelor's degrees among sworn personnel? The table contains Law Enforcement Management and Administrative Statistics (LEMAS) data showing a sample of agencies broken down by type. The numbers represent the percentage of sworn personnel that has a bachelor's degree or higher. Using an alpha level of .01, test the null of no difference between means against the alternative that at least one facility type is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Agency Type			
Municipal (x_1)	County (x_2)	State (x_3)	Tribal (x_4)
6.27	1.87	4.00	4.91
6.45	1.90	3.88	4.96
5.89	2.10	4.19	4.80
6.35	1.45	3.94	5.21
6.30	1.78		
6.28			
$n_1 = 6$	$n_2 = 5$	$n_3 = 4$	$n_4 = 4$

17. Let's continue using the LEMAS survey and exploring differences across agencies of varying types. Problem-oriented policing has been an important innovation in the police approach to reducing disorder and crime. This approach encourages officers to investigate ongoing problems, identify their source, and craft creative solutions. The LEMAS survey asks agency top managers whether they encourage patrol officers to engage in problem solving and, if they do, what percentage of their patrol officers are encouraged to do this type of activity. Using an alpha level of .05, test the null of no difference between means against the alternative that at least one agency type is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Agency Type			
Municipal (x_1)	County (x_2)	State (x_3)	Tribal (x_4)
45	36	29	57
44	37	28	58
45	36	31	59
48	34	28	56
39	38	29	57
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$n_4 = 5$

18. Do the number of contacts people have with police officers vary by race? The Police–Public Contact Survey (PPCS) asks respondents to report their race and the total number of face-to-face contacts they have had with officers in the past year. The following table shows the data. Using an alpha level of .05, test the null of no difference between means against the alternative that at least one facility type is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Race		
White (x_1)	Black (x_2)	Other (x_3)
5.13	6.03	1.74
5.00	5.90	1.70
4.98	6.10	2.00
5.20	6.18	1.61
5.10	6.00	$n_3 = 4$
5.15	$n_2 = 5$	
$n_1 = 6$		

19. Are there race differences among juvenile defendants with respect to the length of time it takes them to acquire pretrial release? The data set *JDCC for Chapter 12.sav* (**www.sagepub.com/gau**) can be used to test for whether time-to-release varies by race for juveniles accused of property crimes. The variables are *race* and *days*. Using SPSS, run an ANOVA with *race* as the IV and *days* as the DV. Select the appropriate post hoc tests.

   1. Identify the obtained value of *F*.
   2. Would you reject the null at an alpha of .01? Why or why not?
   3. State your substantive conclusion about whether there is a relationship between race and days to release for juvenile property defendants.
   4. If appropriate, interpret the post hoc tests to identify the location and total number of significant differences.
   5. If appropriate, compute and interpret omega squared.

20. Are juvenile property offenders sentenced differently depending on the file mechanism used to waive them to adult court? The data set *JDCC for Chapter 12.sav* (**www.sagepub.com/gau**) contains the variables *file* and *jail*, which measure the mechanism used to transfer each juvenile to adult court (discretionary, direct file, or statutory) and the number of months in the sentences of those sent to jail on conviction. Using SPSS, run an ANOVA with *file* as the IV and *jail* as the DV. Select the appropriate post hoc tests.

   1. Identify the obtained value of *F*.
   2. Would you reject the null at an alpha of .05? Why or why not?
   3. State your substantive conclusion about whether there is a relationship between attorney type and days to release for juvenile defendants.
   4. If appropriate, interpret the post hoc tests to identify the location and total number of significant differences.
   5. If appropriate, compute and interpret omega squared.

21. The data set *FISS for Chapter 12.sav* (**www.sagepub.com/gau**) contains the FISS variables capturing shooters' intentions

(accident, assault, and police involved) and victims' ages. Using SPSS, run an ANOVA with *intent* as the IV and *age* as the DV. Select the appropriate post hoc tests.

1. Identify the obtained value of *F*.
2. Would you reject the null at an alpha of .05? Why or why not?
3. State your substantive conclusion about whether victim age appears to be related to shooters' intentions.
4. If appropriate, interpret the post hoc tests to identify the location and total number of significant differences.
5. If appropriate, compute and interpret omega squared.

Key Terms

Analysis of variance (ANOVA) 281

Familywise error 281

Between-group variance 282

Within-group variance 282

F statistic 282

F distribution 282

Post hoc tests 286

Omega squared 297

Tukey's honest significant difference (HSD) 298

Bonferroni 298

Glossary of Symbols and Abbreviations Introduced in This Chapter

F	The statistic and sampling distribution for ANOVA
n_k	The sample size of group k
N	The total sample size across all groups
$\bar{x}_k$	The mean of group k
$\bar{x}_G$	The grand mean across all cases in all groups
SS_B	Between-groups sums of squares; a measure of true group effect
SS_W	Within-groups sums of squares; a measure of error
SS_T	Total sums of squares; equal to $SS_B + SS_W$
MS_B	Between-group mean square; the variance between groups and a measure of true group effect
MS_W	Within-group mean square; the variance within groups and a measure of error
ω^2	A measure of association that indicates the proportion of the total variability that is due to between-group differences

Chapter 13 Hypothesis Testing With Two Continuous Variables Correlation

Learning Objectives

- Identify situations in which, based on the levels of measurement of the independent and dependent variables, correlation is appropriate.
- Define positive and negative correlations.
- Use graphs or hypotheses to determine whether a bivariate relationship is positive or negative.
- Explain the difference between linear and nonlinear relationships.
- Explain the *r* statistic conceptually.
- Explain what the null and alternative hypotheses predict about the population correlation.
- Use raw data to solve equations and conduct five-step hypothesis tests.
- Explain the sign, magnitude, and coefficient of determination and use them in the correct situations.
- Use SPSS to run correlation analyses and interpret the output.

Thus far, we have learned the hypothesis tests used when the two variables under examination are both categorical (chi-square) when the independent variable (IV) is categorical and the dependent variable (DV) is a proportion (two-population *z* test for proportions), when the IV is a two-class categorical measure and the DV is continuous (*t* tests), and when the IV is categorical with three or more classes and the DV is continuous (analysis of variance, or ANOVA). In the current chapter, we will address the technique that is proper when both of the variables are continuous. This technique is **Pearson's correlation** (sometimes also called Pearson's *r*) , because it was developed by Karl Pearson, who was instrumental in advancing the field of statistics.

Pearson's correlation: The bivariate statistical analysis used when both independent and dependent variables are continuous.

The question asked in a correlation analysis is, "When the IV increases by one unit, what happens to the DV?" The DV might increase (a **positive correlation**), it might decrease (a **negative correlation**), or it might do nothing at all (no correlation). Figure 13.1 depicts these possibilities.

Positive correlation: When a one-unit increase in the independent variable is associated with an increase in the dependent variable.

Negative correlation: When a one-unit increase in the independent variable is associated with a decrease in the dependent variable.

A positive correlation might be found between variables such as drug use and violence in neighborhoods: Since drug markets often fuel violence, it would be expected that neighborhoods with high levels of drug activity would be more likely to also display elevated rates of violent crime (i.e., as drug activity increases, so does violence). A negative correlation would be anticipated between the amount of collective efficacy in an area and the crime rate. Researchers have found that neighborhoods where residents know one another and are willing to take action to protect their areas from disorderly conditions have lower crime rates. Higher rates of collective efficacy should correspond to lower crime rates because of collective efficacy's protective capacity.

The bivariate associations represented by correlations are **linear relationships**. This means that the amount of change in the DV that is associated with an increase in the IV remains constant across all levels of the IV and is always in the same direction (positive or negative). Linear relationships can be contrasted to nonlinear or curvilinear relationships such as those displayed in Figure 13.2. You can see in this figure how an increase in

the IV is associated with varying changes in the DV. Sometimes the DV increases, sometimes it decreases, and sometimes it does nothing at all. These nonlinear relationships cannot be modeled using correlational analyses.

> **Linear relationship**: A relationship wherein the change in the dependent variable associated with a one-unit increase in the independent variable remains static or constant at all levels of the independent variable.

Figure 13.1 Three Types of Correlations Between a Continuous IV and a Continuous DV
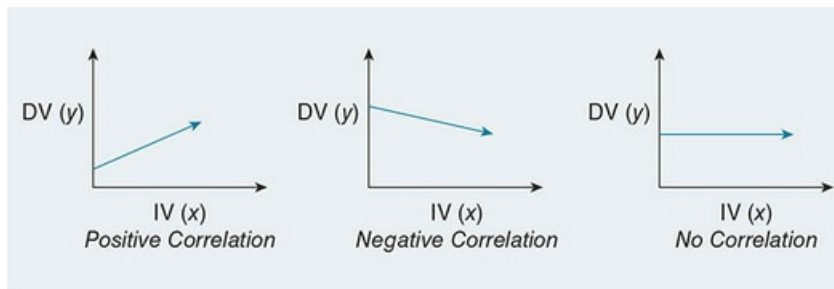

The statistic representing correlations is called the *r* **coefficient.** This coefficient ranges from –1.00 to +1.00. The population correlation coefficient is ρ, which is the Greek letter *rho* (pronounced "row"). A correlation of ±1.00 signals a perfect relationship where a one-unit increase in *x* (the IV) is always associated with exactly the same unit increase of *x*, across all values of both variables. Correlations of zero indicate that there is no relationship between the two variables. Coefficients less than zero signify negative relationships, whereas coefficients greater than zero represent positive relationships. Figure 13.3 depicts the sampling distribution for *r*.

> *r* **coefficient**: The test statistic in a correlation analysis.

Figures 13.4, 13.5, and 13.6 exemplify perfect, strong, and weak positive relationships, respectively. These scatterplots all show that as *y* increases, so does *x*, but you can see how the association breaks down from one figure to the next. Each scatterplot contains what is called a line of best fit—this is the line that minimizes the distance between itself and each value in the data. In other words, no line would come closer to all the data points than this one. The more tightly the data points cluster around the line, the better the line represents the data, and the stronger the *r* coefficient will be. When the data points are scattered, there is a lot of error (i.e., distance between the line and the data points), and the *r* coefficient will be smaller.

Figure 13.2 Examples of Nonlinear Relationships
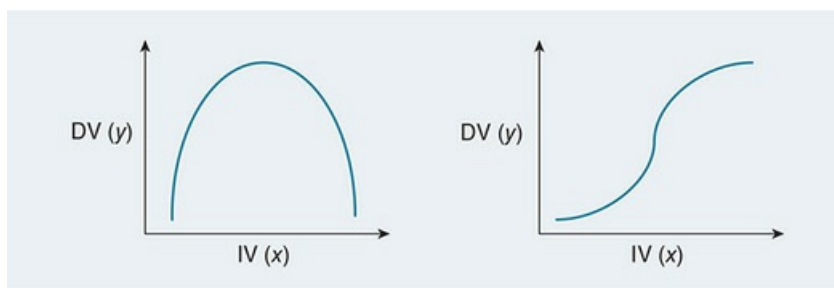
Figure 13.3 The Sampling Distribution of Correlation Coefficients
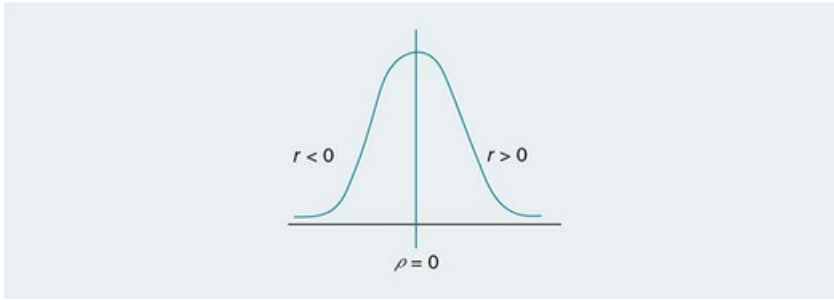
Figure 13.4 A Perfect Linear, Positive Relationship Between *x* and *y*
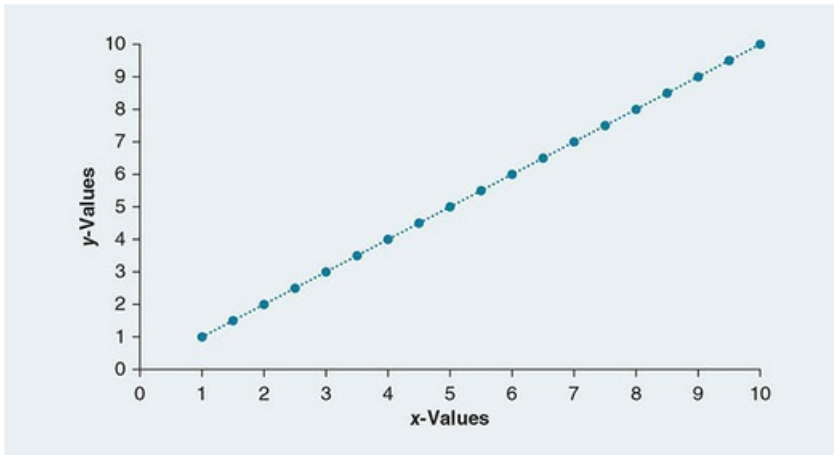
There are no strict rules regarding what constitutes a "strong" or "weak" value of *r*. Researchers use general guidelines to assess magnitudes. In criminal justice and criminology research, values between 0 and ±.29 are generally considered weak, from about ±.30 to ±.49 are moderate, ±.50 to ±.69 are strong, and anything beyond ±.70 is very strong. We will use these general guidelines throughout the chapter when assessing the magnitude of the relationship suggested by a certain *r* value.

Figure 13.5 A Strong Linear, Positive Relationship Between *x* and *y*

Figure 13.6 A Weak Linear, Positive Relationship Between *x* and *y*

As always, it must be remembered that correlation is not causation. A statistically significant correlation between two variables means that there is an empirical association between them (one of the criteria necessary for proving causation, as discussed in Chapter 2), but by itself this is not evidence that the IV causes the DV. There could be another variable that accounts for the DV better than the IV does but that has been omitted from the analysis.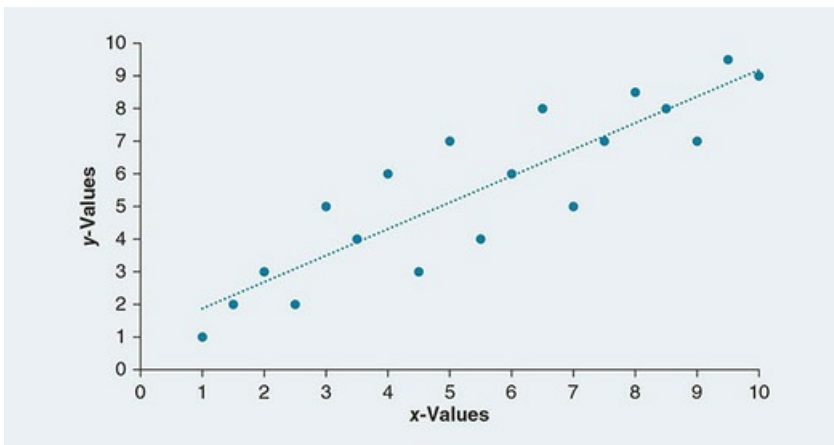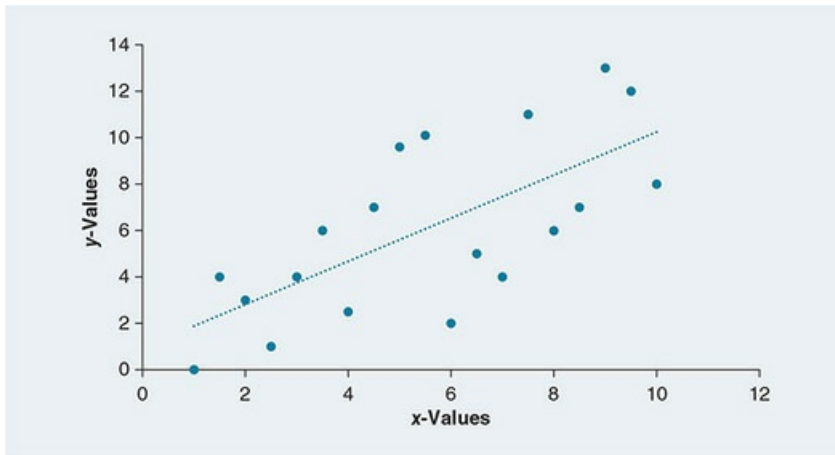 It could also be the case that both the IV and the DV are caused by a third, omitted variable. For instance, crime frequently increases during the summer months, meaning that in any given city, ambient temperature might correlate positively with crime rates. Does this mean heat causes crime? It is possible that hot temperatures make people extra cranky, but a more likely explanation is that crime increases in the summer because people are outdoors and there are more adolescents engaged in unsupervised, unstructured activities. This illustrates the need to think critically about statistically significant relationships between IVs and DVs. Proceed with caution when interpreting correlation coefficients, and keep in mind that there might be more going on than what is captured by these two variables.

Research Example 13.1 Part 1: Is Perceived Risk of Internet Fraud Victimization Related to Online Purchases?

Many researchers have addressed the issue of perceived risk with regard to people's behavioral adaptations. Perceived risk has important consequences at both the individual level and the community level, because people who believe their likelihood of victimization to be high are less likely to connect with their neighbors, less likely to use public spaces in their communities, and more likely to stay indoors to. What has not been addressed with much vigor in the criminal justice and criminology literature is the issue of perceived risk of Internet theft victimization. Given how integral the Internet is to American life and the enormous volume of commerce that takes place online every year, it is important to study the online shopping environment as an arena ripe for theft and fraud.

Reisig, Pratt, and Holtfreter (2009) examined this issue in the context of Internet theft victimization. They used self-report data from a survey administered to a random sample of citizens. Their research question was whether perceived risk of Internet theft victimization would dampen people's tendency to shop online because of the vulnerability created when credit cards are used to make Internet purchases. They also examined whether people's financial impulsivity (the tendency to spend money rather than save it and to possibly spend more than one's income provides for) affected perceived risk. The researchers ran a correlation analysis. Was their hypothesis supported? We will revisit this study later in the chapter to find out.

Correlation analyses employ the t distribution because this probability distribution adequately mirrors the sampling distribution of r at small and large sample sizes (see Figure 13.3). The method for conducting a correlation analysis is to first calculate r and then test for the statistical significance of r by comparing t_{crit} and t_{obt}. Keep this two-step procedure in mind so that you understand the analytic technique in Step 4.

For our first example, we will use the Police–Public Contact Survey (PPCS; see Data Sources 2.1) to find out if there is a relationship between respondents' ages and the number of face-to-face contacts they have had with police officers in the past year. Table 13.1 shows the data for a random sample of seven prisons. We will set alpha at .05.

Table 13.1 Age and Police Contacts	
Age (x)	**Contacts (y)**
23	1
19	0
54	2
68	0
29	2
33	3
45	1
N = 7	*N = 7*

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

In a correlation analysis, the null hypothesis is that there is no correlation between the two variables. The null is phrased in terms of ρ , the population correlation coefficient. This is the Greek letter *rho* (pronounced "roe"). Recall that a correlation coefficient of zero signifies an absence of a relationship between two variables; therefore, the null is

H_0: ρ = 0

Three options are available for the phrasing of the alternative hypothesis. Since correlations use the *t* distribution, these three options are the same as those in *t* tests. There is a two-tailed option (written H_1: $\rho \neq$ 0) that predicts a correlation of unspecified direction. This is the option used when a researcher does not wish to make an a priori prediction about whether the correlation is positive or negative. There are also two one-tailed options. The first predicts that the correlation is negative (H_1: ρ < 0) and the second predicts that it is positive (H_1: ρ > 0).

In the present example, we have no a priori reason for predicting a direction. Past research suggests that young people have more involuntary contacts (such as traffic stops) whereas older people have more voluntary ones (such as calling the police for help). Since the contact variable used here includes all types of experiences, a prediction cannot be made about how age might correlate with contacts; therefore, we will use a two-tailed test. The alternative hypothesis is

H_1: $\rho \neq$ 0

Step 2. Identify the distribution and compute the degrees of freedom.

The *t* distribution is the probability curve used in correlation analyses. Recall that this curve is symmetric and, unlike the χ^2 and *F* distributions, has both a positive side and a negative side. The degrees of freedom (*df*) in correlation are computed as

$$df = N - 2 \qquad\qquad \textit{Formula 13 (1)}$$

In the present example, there are five people in the sample, so

$df = 7 - 2 = 5$

Step 3. Identify the critical value, and state the decision rule.

With a two-tailed test, $\alpha = .05$, and *df* = 5, the value of t_{crit} is 2.571. Since this is a two-tailed test, there are two critical values because half of α is in each tail. This means $t_{crit} = \pm 2.571$. The decision rule is that *if t_{obt} is either > 2.571 or < –2.571* , H_0 *will be rejected*.

Step 4. Compute the obtained value of the test statistic.

There are two parts to Step 4 in correlation analyses: First, the correlation coefficient *r* is calculated, and second, the statistical significance of *r* is tested by plugging *r* into the t_{obt} formula. The formula for *r* looks complex, but we will solve it step by step.

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}} \qquad\qquad \textit{Formula 13(2)}$$

The formula requires several different sums. These sums must be calculated and then entered into the equation. The easiest way to obtain these sums is to use a table. Table 13.2 reproduces the raw data from Table 13.1 and adds three columns to the right that allow us to compute the needed sums.

Table 13.2 Correlation Computation Table for Example 1				
Age (x)	Contacts (y)	xy	x^2	y^2
23	1	23	529	1
19	0	0	361	0
54	2	108	2,916	4
68	0	0	4,624	0
29	2	58	841	4
33	3	99	1,089	9
45	1	45	2,025	1
$\sum x = 271$	$\sum y = 9$	$\sum xy = 333$	$\sum x^2 = 12,385$	$\sum y^2 = 19$

The sums from Table 13.2 can be entered into Formula 13(2):

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}}$$

$$= \frac{7(333) - 271\,9}{\sqrt{\left[7\,12,385 - 271^2\right]\left[7\,19 - 9^2\right]}}$$

$$= \frac{2,331 - 2,439}{\sqrt{[86,695 - 73,441][133 - 81]}} = \frac{-108}{\sqrt{[13,254][52]}}$$

$$= \frac{-108}{\sqrt{689,208}} = \frac{-108}{830.19}$$

$= -.13.$

The first part of Step 4 is thus complete. We now know that $r = -.13$. This is a low value, so we can see already that there is not much of a correlation between the variables, if indeed there is any correlation at all. It is possible this r value is merely a chance finding. Until we either reject or retain the null, we cannot reach any conclusions about r. To make a decision about the null, the following formula is used:

$$t_{obt} = r\sqrt{\frac{N-2}{1-r^2}} \qquad\qquad Formula\ 13\,(3)$$

Plugging our numbers in,

$$t_{obt} = -.13\sqrt{\frac{7-2}{1-(-.13)^2}} = -.13\sqrt{\frac{5}{1-.02}} = -.13\sqrt{\frac{5}{.98}}$$

$$= -.13\sqrt{(5.10)}$$

$= -.13(2.26)$

$= -.29$

Step 4 is complete! We know that $r = -.13$ and $t_{obt} = -.29$.

Learning Check 13.1

The calculated value of *r* and obtained value of *t* will always have the same sign: If *r* is negative, *t* will be negative, and if *r* is positive, *t* will be positive. Can you explain the reason for this? If you need a hint, refer back to Figure 13.3.

Step 5. Make a decision about the null and state the substantive conclusion.

In the decision rule, we said that the null would be rejected if t_{obt} ended up being > 2.571 or < –2.571. Since the t_{obt} we calculated is only –.29, the null must be retained. We conclude that there is no relationship between people's ages and the number of contacts they have with police officers.

For a second example of correlation testing, we turn to the Law Enforcement Management and Administrative Statistics (LEMAS; see Data Sources 3.2) survey. We will also pull from the Uniform Crime Reports (UCR; see Data Sources 1.1). The LEMAS data set contains a variable measuring the ratio of officers per 1,000 citizens, which indicates the size of the police force relative to the local population. From the UCR, we can calculate each city's violent crime rate (per 1,000). We will test the hypothesis that police force size correlates negatively with the crime rate (i.e., cities with larger police departments have less violent crime and vice versa). Alpha will be set at .01. Table 13.3 contains the data and the sums we will need for the *r* formula.

Step 1. State the null (H_0) *and alternative* (H_1) *hypotheses.*

H_0: $\rho = 0$

H_1: $\rho < 0$

Step 2. Identify the distribution and compute the degrees of freedom.

The distribution is *t* , and the *df* are computed using Formula 13(1):

$df = 8 – 2 = 6$

Step 3. Identify the critical value and state the decision rule.

With a one-tailed test, an alpha of .01, and *df* = 6, t_{crit} = –3.143. The critical value is negative because the alternative hypothesis predicts that the correlation is less than zero. The decision rule is that *if* t_{obt} < *–3.143,* H_0 *will be rejected.*

Table 13.3 Police Force Size and Violent Crime Rate				
Officers per 1,000 (x)	Violent Crimes per 1,000 (y)	xy	x²	y²
1.73	11.57	20.02	2.99	133.86
1.61	5.03	8.10	2.59	25.30
1.59	8.07	12.83	2.53	65.12
8.55	.83	7.10	73.10	.69
3.07	1.07	3.28	9.42	1.14
1.91	.30	.57	3.65	.09
.81	.40	.32	.66	.16
2.20	.78	1.72	4.84	0.61
$\Sigma x = 21.47$	$\Sigma y = 28.05$	$\Sigma xy = 53.94$	$\Sigma x^2 = 99.78$	$\Sigma y^2 = 226.97$

Step 4. Compute the obtained value of the test statistic.

Using Formula 13(2) and the sums from Table 13.3,

$$r = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{\left[N\Sigma x^2 - (\Sigma x)^2\right]\left[N\Sigma y^2 - (\Sigma y)^2\right]}}$$

$$= \frac{8(53.94) - (21.47)(28.05)}{\sqrt{\left[8(99.78) - 21.47^2\right]\left[8(226.97) - 28.05^2\right]}}$$

$$= \frac{431.52 - 602.23}{\sqrt{[798.24 - 460.96][1,815.76 - 786.80]}}$$

$$= \frac{-170.71}{\sqrt{[337.28][1,028.96]}} = \frac{-170.71}{\sqrt{347,047.63}} = \frac{-170.71}{589.11}$$

$$= -.29$$

This is a modest *r* value, suggesting that there is an overlap between these two variable, but it is not a substantial amount. Nonetheless, *r* might be statistically significant, so we proceed to finding t_{obt} using Formula 13(3):

$$t_{obt} = -.29\sqrt{\frac{8-2}{1-(-.29)^2}} = -.29\sqrt{\frac{6}{.92}} = -.29\sqrt{6.52}$$

$$= -.29 (2.55)$$

$$= -.74$$

And Step 4 is done! $t_{obt} = -.74$.

Step 5. Make a decision about the null and state the substantive conclusion.

The decision rule stated that the null would be rejected if t_{obt} was less than –2.896. Since –.74 does not meet this requirement, the null must be retained. There is not a statistically significant correlation between the size of a city's police force and the violent crime rate in that area.

Research Example 13.2 Do Prisoners' Criminal Thinking Patterns Predict Misconduct?

Institutional security inside prisons depends in part on officials' ability to accurately classify prisoners as low, medium, or high risk. Reliable classification decisions also benefit inmates so that, for instance, inmates with relatively minimal criminal tendencies are not housed with hardened, dangerous offenders. Static risk factors (such as criminal history) can help officials make accurate decisions. Static risk factors, though, are limited because they capture only general characteristics about inmates and do not tap into those individuals' thoughts and attitudes. Walters (2015) proposed that incorporating dynamic risk factors would improve the accuracy of prison officials' classification decisions. He collected data on all inmates entering a medium-security, men's prison between 2003 and 2010. Each inmate was administered an inventory measuring the extent of his criminal thinking and antisocial attitudes. Walters then collected data on the number of disciplinary infractions each person was found guilty of during the study period. The table shows the correlations between the DVs (infractions) and IVs (criminal thinking and static risk factors).

	Infractions	Criminal Thinking	Criminal History	Age
Infractions	1.00			
Criminal Thinking	.11**	1.00		
Criminal History	.08**	.06*	1.00	
Age	–.18**	–.10**	.07*	1.00

*p < .05; ** p < .001.
Source: Adapted from Table 1 in Walters (2015).

p < .05;

p < .001.
Source: Adapted from Table 1 in Walters (2015).

Walters' analyses showed that, as predicted, criminal thinking was significantly and positively related to prison disciplinary infractions. In other words, inmates scoring higher on this measure of antisocial attitudes also misbehaved more often. This correlation was fairly weak, however, suggesting that although knowing an inmate's score on the criminal thinking inventory helps make accurate classification decisions, this score should be one of many pieces of information used when deciding how to classify an inmate.

Do Good Recruits Make Good Cops?

The purpose of police training academies is to prepare recruits for work on the street; however, it is not known to what extent academy training relates to on-the-job performance. Henson, Reyns, Klahm, and Frank (2010) attempted to determine whether, and to what extent, recruits' academy performance is associated with their later effectiveness as police officers. They used three DVs: the evaluations new officers received from their supervisors, the number of complaints that were lodged against those new officers, and the number of commendations the officers received for exemplary actions. The IVs consisted of various measurements of academy performance. Henson et al. obtained the correlations shown in the table. Note that the authors did not report correlation coefficients that were not statistically significant; the nonsignificant coefficients are replaced with *ns*.

	Evaluation	Complaints	Commendations
Civil Service Exam	ns	−.14**	ns
Physical Agility Rating	.19*	.13*	ns
Overall Academy Score	.12*	ns	.12*

*p <.01; **p <.001.

Source: Adapted from Table 3 in Henson et al. (2010).

p <.01;

p <.001.

Source: Adapted from Table 3 in Henson et al. (2010).

Were the authors' predictions supported? Is academy performance related to the quality of the job those recruits do once they are on the street? The results were mixed. You can see in the table that many of the correlations were statistically significant, but nearly as many were not. Even those results that were statistically significant were weak in magnitude. It would appear from this analysis that academy performance and on-the-job performance are not strongly related. An officer's academy performance is only marginally related to how well he or she carries out the police job.

For a third example, we will draw from the data set Juvenile Defendants in Criminal Courts (JDCC; Data Sources 11.1). This data set contains information on the dollar amount of penalties imposed on juveniles convicted of crimes whose sentence includes paying a fine. A question we might ask is whether there is a correlation between the number of charges a juvenile is prosecuted for and the size of the fine imposed on him or her. It seems logical to expect these variables to correlate positively, such that more charges result in heavier fines. Therefore, this will be a directional test and we will be working on the right-hand (positive) side of the t distribution. Table 13.4 shows the data for a sample of eight juveniles. Alpha will be .05.

Step 1. State the null (H_0) *and alternative* (H_1) *hypotheses.*

H_0: $\rho = 0$

H_1: $\rho > 0$

Table 13.4 Number of Charges and Fine Amount				
Charges (x)	Fine (y)	xy	x^2	y^2
1	120	120	1	14,400
1	250	250	1	62,500
5	1,000	5,000	25	1,000,000
5	750	3,750	25	562,500
2	500	1,000	4	250,000
1	305	305	1	93,025
4	255	1,020	16	65,025
3	800	2,400	9	640,000
$\Sigma x = 22$	$\Sigma y = 3,980$	$\Sigma xy = 13,845$	$\Sigma x^2 = 82$	$\Sigma y^2 = 2,687,450$

Step 2. Identify the distribution and compute the degrees of freedom.

The distribution is t, and the *df* are computed using Formula 13(1):

$df = 8 - 2 = 6$

Step 3. Identify the critical value and state the decision rule.

For a one-tailed test, an alpha of .05, and $df = 6$, the critical value of t is 1.943. The decision rule is that *if* t_{obt} *is greater than 1.943*, H_0 *will be rejected.*

Step 4. Compute the obtained value of the test statistic.

Using Formula 13(2) and the sums from Table 13.4,

$$r = \frac{8(13,845) - 22(3,980)}{\sqrt{\left[8(82) - 22^2\right]\left[8(2,687,450) - 3,980^2\right]}}$$

$$= \frac{110,760 - 87,560}{\sqrt{[656 - 484][21,499,600 - 15,840,400]}}$$

$$= \frac{23,200}{\sqrt{[172][5,659,200]}} = \frac{23,200}{\sqrt{973,382,400}} = \frac{23,200}{31,199.08}$$

$= .74$

The r value is large, indicating a strong positive relationship between the variables. We still need to carry out the calculations for t_{obt}, though, because we have not yet ruled out the possibility that this r is a fluke finding. The t test will tell us that

$$t_{obt} = .74\sqrt{\frac{8-2}{1-(.74)^2}} = .74\sqrt{\frac{6}{.45}} = .74\sqrt{13.33}$$

= .74 (3.65)

= 2.70

Step 5. Make a decision about the null and state the substantive conclusion.

The decision rule stated that the null would be rejected if t_{obt} ended up being greater than 1.943. The value of t_{obt} greatly exceeds 1.943, so we will reject the null. Among juveniles tried in adult courts and sentenced to fines, there is a statistically significant correlation between the number of criminal charges filed against a person and the dollar amount of the fine imposed.

Beyond Statistical Significance: Sign, Magnitude, and Coefficient of Determination

When the null hypothesis is rejected in a correlation hypothesis test, the correlation coefficient r can be examined with respect to its substantive meaning. We have touched on the topic of magnitude versus statistical significance already; note that in all three examples of correlation tests, we made a preliminary assessment of the magnitude of r before moving on to the t_{obt} calculations. In each one, though, we had to do that last step to check for statistical significance before formally interpreting the strength or weakness of r. The reverse of this is also true: Statistical significance is not proof that the variables are strongly correlated. The null can be rejected even when a correlation is of little practical importance. The biggest culprit of misleading significance is sample size: Correlations that are substantively weak can result in rejected nulls simply because the sample is large enough to drive up the value of t_{obt}. When the null is rejected, criminal justice and criminology researchers turn to three interpretive measures to assess the substantive importance of a statistically significant r: *sign*, *magnitude*, and *coefficient of determination*.

The sign of the correlation coefficient indicates whether the correlation between the IV and the DV is negative or positive. Take another look at [Figure 13.1](#) to refresh your memory as to what negative and positive correlations look like. A positive correlation means that a unit increase in the IV is associated with an increase in the DVs, and a negative correlation indicates that as the IV increases, the DV declines.

The magnitude is an evaluation of the strength of the relationship based on the value of r. As noted in the outset of this chapter, there are no rules set in stone for determining whether a given r value is strong, moderate, or weak in magnitude; this judgment is based on a researcher's knowledge of the subject matter. As described previously, a general guideline is that values between 0 and ±.29 are weak, from about ±.30 to ±.49 are moderate, ±.50 to ±.69 are strong, and those beyond ±.70 are very strong.

Third, the coefficient of determination is calculated as the obtained value of r, squared (i.e., r^2). The result is a proportion that can be converted to a percentage and interpreted as the percentage of the variance in the DV that is attributable to the IV. As a percentage, the coefficient ranges from 0 to 100, with higher numbers signifying stronger relationships and numbers closer to zero representing weaker associations.

Let us interpret the sign, magnitude, and coefficient of determination for the correlation coefficient computed in the third example. Since we retained the null in the first and second examples, we cannot apply the three interpretive measures to these r values. In the examination of charges and fines, we found that $r = .74$.

First, the sign of the correlation coefficient is positive, meaning that a greater number of charges is associated with a higher fine amount. Second, the magnitude is quite strong, as .74 exceeds the .70 threshold. Third, the coefficient of determination is $r^2 = .74^2 = .55$. This means that 55% of the variance in the DV (fine amount) can be explained by the IV (number of charges). This is a decent amount of shared variance! Of course, we cannot draw causal conclusions—there are numerous factors that enter into judges' sentencing decisions. Anytime you interpret the outcome of a correlation hypothesis test, keep in mind that statistical significance is

not, by itself, enough to demonstrate a practically significant or substantively meaningful relationship between two variables; moreover, even a strong association does not mean that one variable truly causes the other.
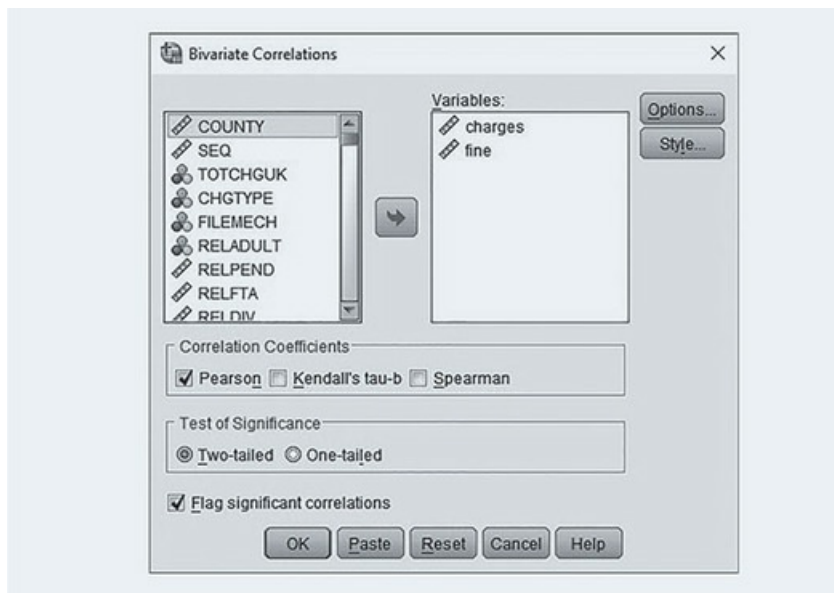
SPSS

Correlations are run in SPSS using the *Analyze → Correlation → Bivariate* sequence. Once the dialog box shown in Figure 13.7 appears, select the variables of interest and move them into the analysis box as shown. In this example, the JDCC data on charges and fines is used. You can see in Figure 13.7 that both of these variables have been moved from the list on the left into the box on the right. After selecting your variables, click *OK*. Figure 13.8 shows the output.

The output in Figure 13.8 is called a correlation matrix, meaning that it is split by what is called a diagonal (here, the cells in the upper-left and lower-right corners of the matrix) and is symmetric on both of the off-diagonal sides. The numbers in the diagonal are always 1.00 because they represent each variable's correlation with itself. The numbers in the off-diagonals are the ones to look at. The number associated with Pearson's correlation is the value of *r*. In Figure 13.8, you can see that $r = .239$, which is much smaller than what we arrived at by hand, but since the SPSS example is employing the entire data set, variation is expected.

The *Sig.* value is, as always, the obtained significance level or *p* value. This number is compared to alpha to determine whether the null will be rejected. If $p < α$, the null is rejected; if $p > α$, the null is retained. Typically, any *p* value less than .05 is considered statistically significant. In Figure 13.8, the *p* value is .000, which is very small and indicates that we would reject the null even if we set α at .001, a very stringent test for statistical significance.

Figure 13.7 Running a Correlation Analysis in SPSS

Correlation matrices can be expanded to include multiple variables. When you do this, SPSS runs a separate analysis for each pair. Let us add juvenile defendants' ages to the matrix. Figure 13.9 shows the output containing all three variables.

Figure 13.8 SPSS Output

Correlations

		Total number of charges	Dollar amount of fine
Total number of charges	Pearson Correlation	1	.239**
	Sig. 92-tailed)		.000
	N	6532	663
Dollar amount of fine	Pearson Correlation	.239**	1
	Sig. (2-tailed)	.000	
	N	663	671

** Correlation is significant at the 0.01 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Learning Check 13.2

✓

In a correlation matrix, the numbers in the diagonal will all be 1.00 because the diagonal represents each variable's correlation with itself. Why is this? Explain why any given variable's correlation with itself is always 1.00.

Part 2: Is Perceived Risk of Internet Fraud Victimization Related to Online Purchases?

Recall that Reisig et al. (2009) predicted that people's perceived likelihood of falling victim to Internet theft would lead to less-frequent engagement in the risky practice of purchasing items online using credit cards. They also thought that financial impulsivity, as an indicator of low self-control, would affect people's perceptions of risk. The following table is an adaptation of the correlation matrix obtained by these researchers.

Were the researchers' hypotheses correct? The results were mixed. On the one hand, they were correct in that the correlations between the IVs and the DV were statistically significant. You can see the significance of these relationships indicated by the asterisks that flag both of these correlations as being statistically significant at an alpha level of .05. Since the null was rejected, it is appropriate to interpret the coefficients. Regarding sign, perceived risk was negatively related to online purchases (i.e., greater perceived risk meant less online purchasing activity), and financial impulsivity was positively related to perceived risk (i.e., financially impulsive people were likely to see themselves as facing an elevated risk of victimization).

The reason the results were mixed, however, is that the correlations—though statistically significant—were not strong. Using the magnitude guidelines provided in this chapter, it can be seen that –.12 and .11 are very weak. The coefficient of determination for each one is $(-.12)^2 = .01$ and $.11^2 = .01$, so only 1% of the variance in online purchases was attributable to perceived risk, and only 1% was due to financial impulsivity, respectively. This illustrates the potential discrepancy between statistical significance and substantive importance: Both of these correlations were statistically significant, but neither meant much in terms of substantive or practical implications. As always, though, it must be remembered that these analyses were bivariate and that the addition of more IVs might alter the IV–DV relationships observed here.

	Perceived Risk	Financial Impulsivity	Online Purchases
Perceived Risk	1.00		
Financial Impulsivity	.11*	1.00	
Online Purchases	–.12*	.05	1.00

Source: Adapted from Table 1 in Reisig et al. (2009).

*$p < .05$.

Source: Adapted from Table 1 in Reisig et al. (2009).

p <.05.

Figure 13.9 SPSS Output

Correlations				
		Total number of charges	Dollar amount of fine	Age at file date
Total number of charges	Pearson Correlation	1	.239**	-.006
	Sig. 2-tailed)		.000	.644
	N	6532	663	6493
Dollar amount of fine	Pearson Correlation	.239**	1	.033
	Sig. (2-tailed)	.000		.390
	N	663	671	671
Age at file date	Pearson Correlation	-.006	.033	1
	Sig. (2-tailed)	.644	.390	
	N	6493	671	7096

It turns out that age bears no relationship with fine severity (or with the number of charges filed, which was not the original purpose of our analysis, but something we can see in the correlation matrix). The correlation between age and fines is .033, and p = .390, which is nowhere close to .05.

Chapter Summary

This chapter introduced Pearson's correlation, the hypothesis-testing technique that is appropriate in studies employing two continuous variables. The correlation coefficient is symbolized r and ranges from –1.00 to +1.00. The value of r represents the amount of linear change in the DV that accompanies a one-unit increase in the IV. Values of r approaching ±1.00 are indicative of strong relationships, whereas values close to zero signify weak correlations. The statistical significance of r is tested using the t distribution because the sampling distribution of r is normal in shape.

When the null hypothesis is rejected in a correlation test, the sign, magnitude, and coefficient of determination should be examined in order to determine the substantive, practical meaning of r. The sign indicates whether the correlation is positive or negative. The magnitude can be assessed using general guidelines and subject-matter expertise. The coefficient of determination is computed by squaring r to produce a proportion (or percentage, when multiplied by 100) that represents the amount of variance in the DV that is attributable to the IV.

Correlation analyses can be run in SPSS. The program provides the calculated value of the correlation coefficient r and its associated significance value (its p value). When p is less than alpha (typically set at .05), the null is rejected and the correlation coefficient can be interpreted for substantive meaning; when p is greater than alpha, the null is retained and r is not interpreted because the conclusion is that the two variables are not correlated.

Thinking Critically

1. At the beginning of the chapter, it was explained that correlations are tests for linear relationships between two variables. Figures 13.4 through 13.6 depicted the straight lines that represent linear relationships. Compare these straight lines to the curvilinear lines pictured in Figure 13.2. Why would it be a problem to run a correlation analysis on two variables whose relationship looked like one of those in Figure 13.2? (Hint: Try drawing a curvilinear line and then a straight line over the top of it.) Do you think the r coefficient would be an accurate reflection of the relationship between these two variables? Explain your answer.

2. In one of the examples, we found a statistically significant, positive relationship between the number of charges filed against juveniles in criminal courts and the severity of the fine imposed (among those sentenced to pay fines). Throughout this book, you have seen examples of the need to consider additional variables before making a decision about the trustworthiness of a bivariate relationship. What more would you want to know about juvenile defendants' cases in order to feel that you have a clear picture of the charge–fine relationship? Identify the variables that you would add to this analysis.

Review Problems

1. A researcher wishes to test the hypothesis that parental incarceration is a risk factor for lifetime incarceration of the adult

children of incarcerated parents. She measures parental incarceration as *had a parent in prison; did not have a parent in prison* and adult children's incarceration as *incarcerated; not incarcerated* .

   1. Identify the IV.
   2. Identify the level of measurement of the IV.
   3. Identify the DV.
   4. Identify the level of measurement of the DV.
   5. What type of inferential analysis should the researcher run to test for a relationship between these two variables?

2. A researcher wishes to test the hypothesis that coercive control can actually increase, rather than reduce, subsequent criminal offending among persons who have been imprisoned. He gathers a sample of people who have spent time in prison and measures coercive control as *number of years spent in prison* and recidivism as *number of times rearrested after release from prison* .

   1. Identify the IV.
   2. Identify the level of measurement of the IV.
   3. Identify the DV.
   4. Identify the level of measurement of the DV.
   5. What type of inferential analysis should the researcher run to test for a relationship between these two variables?

3. A researcher hypothesizes that a new policing strategy involving community meetings designed to educate residents on the importance of self-protective measures will increase the use of such measures. He gathers a sample of local residents and finds out whether they have participated in an educational session (*have participated; have not participated*) . Among each group, he computes the *proportion of residents who now take self-protective measures* .

   1. Identify the IV.
   2. Identify the level of measurement of the IV.
   3. Identify the DV.
   4. Identify the level of measurement of the DV.
   5. What type of inferential analysis should the researcher run to test for a relationship between these two variables?

4. A researcher thinks that there might be a relationship between neighborhood levels of socioeconomic disadvantage and the amount of violent crime that occurs in those neighborhoods. She measures socioeconomic disadvantage as the *percentage of neighborhood residents that live below the poverty line* , and she measures violent crime as the *number of violent offenses reported to police per 1,000 neighborhood residents* .

   1. Identify the IV.
   2. Identify the level of measurement of the IV.
   3. Identify the DV.
   4. Identify the level of measurement of the DV.
   5. What type of inferential analysis should the researcher run to test for a relationship between these two variables?

5. Explain what it means for a relationship to be linear.

6. Explain what it means for a relationship to be nonlinear.

7. Explain the concept of the line of best fit.

8. When a correlation is _____, increases in the IV are associated with increases in the DV.
   1. positive
   2. negative
   3. zero

9. When a correlation is _____, increases in the IV are not associated with any predictable or consistent change in the DV.
   1. positive
   2. negative
   3. zero

10. When a correlation is _____, increases in the IV are associated with decreases in the DV.
   1. positive
   2. negative
   3. zero

11. Is there a correlation between the amount of money states spend on prisons and those states' violent crime rates? There could be a negative correlation insofar as prisons might suppress crime, in which case money invested in incarceration produces a reduction in violence. There could also be a positive correlation, however, if prison expenditures do not reduce crime but, rather, merely reflect the amount of violence present in a state. Morgan, Morgan, and Boba (2010) offer information on the

393

dollars spent per capita on prison expenditures in 2009, and the UCR provide 2009 violent crime data. The following table contains dollars spent per capita and violent crime rates per 1,000 citizens for a random sample of five states. Using an alpha level of .05, test for a correlation between the variables (note that no direction is being specified). Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

State	Prison Dollars per Capita (x)	Violent Crime Rate (y)
Florida	140	6.12
Idaho	135	2.28
Maine	99	1.20
Nebraska	120	2.82
Texas	133	4.91
N = 5		

12. One aspect of deterrence theory predicts that punishment is most effective at preventing crime when would-be criminals know that they stand a high likelihood of being caught and penalized for breaking the law. When certainty breaks down, by contrast, crime rates might increase because offenders feel more confident that they will not be apprehended. The following table contains UCR data on the clearance rate for property crimes and the property crime rate per 100. The prediction is that higher clearance rates are associated with lower property crime rates. Using an alpha of .01, test for a negative correlation between the variables. Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

Region	Percent of Property Crimes Cleared (x)	Property Crime Rate (y)
New England	15.40	2.33
Mid Atlantic	22.80	2.05
East North Central	16.30	2.93
West North Central	19.40	2.84
South Atlantic	19.10	3.50
East South Central	17.60	3.34
West South Central	15.80	3.92
Mountain	16.90	3.10
Pacific	14.50	2.90
N = 9		

13. Are there more police agencies in areas with higher crime rates? We will find out using data from Morgan et al. (2010). The IV is *crimes per square mile* , and the DV is *police agencies per 1,000 square miles*. Using an alpha level of .05, test for a positive correlation between the variables. Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

State	Crime per Sq. Mile (x)	Police Agencies per 1,000 Sq. Miles (y)
Georgia	7.30	9.40
Iowa	1.40	7.20
Kansas	1.30	4.40
Maine	1.00	3.90
Minnesota	1.90	5.30
Missouri	3.50	8.40
Nebraska	.70	3.20
South Carolina	6.90	8.40
Washington State	3.80	3.70
N = 9		

14. In police agencies—much like in other types of public and private organizations—there are concerns over disparities in pay; in particular, lower-ranking officers might feel undercompensated relative to higher-ranking command staff and administrators. Let us investigate whether there is a correlation between the pay of those at the top of the police hierarchy (chiefs) and those at the bottom (officers). We will use a random sample of six agencies from the LEMAS data set. The following table shows the minimum annual salaries of chiefs and officers among the agencies in this sample. (The numbers represent thousands; for instance, 57.5 means $57,500.) Using an alpha level of .05, test for a correlation between the two variables (no direction specified). Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

Agency	Minimum Chief Salary (x)	Minimum Officer Salary (y)
A	57.50	28.40
B	20.40	14.50
C	57.60	31.60
D	70.00	21.50
E	21.70	16.50
F	75.50	33.80
N = 6		

15. It is well known that handguns account for a substantial portion of murders. This has led some people to claim that stricter handgun regulations would help curb the murder rate in the United States. Others, though, say that tougher gun laws would not work because people who are motivated to kill but who cannot obtain a handgun will simply find a different weapon instead. This is called a *substitution effect*. If the substitution effect is operative, then there should be a negative correlation between handgun and knife murders. The following table contains data from a random sample of states. For each state, the handgun and knife murder rates (per 100,000 state residents) are shown. Using an alpha level of .01, test for a negative correlation between the variables. Use all five steps. If appropriate, interpret the sign, magnitude, and coefficient of determination.

State	Handgun Murder Rate (x)	Knife Murder Rate (y)
Nebraska	1.22	.45
Wisconsin	1.15	.39
Arkansas	1.87	.73
North Carolina	2.59	.52
Virginia	1.37	.52
California	2.77	.79
Washington State	1.13	.53
N = 7		

16. Does it take defendants who face multiple charges longer to get through the adjudication process? The following table contains data on a random sample of juveniles from the JDCC data set. The variables are *number of charges* and *months to adjudication;* the latter variable measures the total amount of time that it took for these juveniles to have their cases disposed of. Using an alpha level of .05, test for a positive correlation between the variables. Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

Juvenile	Charges (x)	Months (y)
A	5	7.13
B	3	5.50
C	3	2.47
D	2	1.20
E	4	.47
F	3	.80
G	2	.07
H	1	.00
I	1	.93
J	1	.67
N = 10		

17. Does a juvenile's age correlate with how long it takes for the case to reach disposition? The following table contains a sample of female juveniles from the JDCC. Using an alpha level of .05, test for a correlation between the variables (no direction specified). Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

Female Juvenile	Age (x)	Time (y)
A	17	4
B	14	5
C	16	5
D	17	8
E	13	2
F	15	6
G	16	3
H	13	4
N = 8		

18. Is there a relationship between age and attitudes about crime and punishment? The General Social Survey (GSS) asks respondents whether they oppose or favor the death penalty and whether they think courts are too harsh, about right, or not harsh enough on offenders. These two variables can be summed to form a scale measuring people's general preference for harsher or more-lenient penalties (higher values represent harsher attitudes). The data are displayed in the table. Using an alpha level of .01, test for a correlation between the variables (no direction specified). Use all five steps. If appropriate, interpret the sign, the magnitude, and the coefficient of determination.

Respondent	Age (x)	Sentence Severity (y)
A	18	4
B	46	5
C	63	2
D	19	3
E	22	2
F	35	3
G	39	2
H	74	5
I	68	4
N = 9		

19. Is there a relationship between age and the types of experiences people have with police? The file *PPCS for Chapter 13.sav* contains data from the PPCS. The sample is narrowed to male respondents and contains the variables *age* (respondents' ages in years) and *length* (the number of minutes in a traffic stop, for those respondents who had been stopped in the past year). Run a correlation analysis.
    1. Identify the obtained value of the correlation coefficient *r*.
    2. State whether you would reject the null hypothesis at an alpha of .05 and how you reached that decision.
    3. State the substantive conclusion.
    4. As appropriate, interpret the sign, the magnitude, and the coefficient of determination.
20. Let's revisit the variable measuring attitudes toward criminal punishment. The file *GSS for Chapter 13.sav* (www.sagepub.com/gau) contains data from the GSS, narrowed down to female respondents. The variable *severity* measures how harshly respondents feel toward persons convicted of crimes (higher values indicate greater preferences for severe punishment). The other variables are *children, age* , and *education*. We want to find out whether there are any statistically significant correlations among these variables. Run a correlation analysis using all four variables.
    1. Identify the obtained values of the correlation coefficient *r* for each test.
    2. For each of the *r* values, state whether you would reject the null hypothesis at an alpha of .05 and how you reached

that decision.

3. State the substantive conclusion for each test.

4. As appropriate, interpret the sign, the magnitude, and the coefficient of determination for each test.

Key Terms

Pearson's correlation 312

Positive correlation 312

Negative correlation 312

Linear relationship 312

r coefficient 312

Glossary of Symbols and Abbreviations Introduced in This Chapter

r	The correlation coefficient for a sample
ρ	The correlation coefficient for a population (rho)
r^2	The coefficient of determination

Chapter 14 Introduction to Regression Analysis

Learning Objectives

- Explain the benefits of regression's ability to predict values on the dependent variable (*y*) and explain how regression differs from the tests covered in previous chapters.
- Explain the difference between bivariate and multiple regression, including the importance of controlling for additional independent variables.
- Calculate each element of an ordinary least squares regression model, and compile the elements into an equation predicting *y*.
- Use an ordinary least squares regression equation to find the predicted value of *y* given a certain value of *x*.
- Read and interpret SPSS regression output, including identifying each key element of the equation and forming a conclusion about the statistical significance and substantive strength of relationships.

In Chapter 13, you learned about correlation analysis. Correlation is a statistical procedure for finding two pieces of information: (1) whether two continuous variables are statistically related in a linear fashion and (2) the strength of that relationship. In most criminal justice and criminology research, however, merely finding out whether two variables are correlated is not sufficient. Researchers want to find out whether one of the variables (the independent variable [IV]) can be used to *predict* the other one (the dependent variable [DV]). **Regression analysis** does this. Regression goes a step beyond correlation by allowing researchers to determine how well the IV predicts the DV.

> **Regression analysis**: A technique for modeling linear relationships between one or more independent variables and one dependent variable wherein each independent variable is evaluated on the basis of its ability to accurately predict the values of the dependent variable.

This chapter will discuss two types of regression analyses. **Bivariate regression** employs one IV and one DV. It is similar to bivariate correlation in many respects. **Multiple regression** uses several IVs to predict the DV. The specific type of regression modeling discussed here is **ordinary least squares (OLS) regression**. This is a fundamental form of regression modeling that is used frequently in criminology and criminal justice research. This type of regression produces a line of best fit that is as close to all the data points as it can possibly get. In other words, it minimizes the errors in the prediction of the DV. There are other types of regression, but OLS is the default procedure that is generally employed unless there is good reason to depart from it and use a different technique instead. In OLS, the DV must be continuous and normally distributed. The IVs can be of any level of measurement.

> **Bivariate regression**: A regression analysis that uses one independent and one dependent variable.
>
> **Multiple regression**: A regression analysis that uses two or more independent variables and one dependent variable.
>
> **Ordinary least squares (OLS) regression**: A common procedure for estimating regression equations that minimizes the errors in predicting the dependent variable.

One Independent Variable and One Dependent Variable: Bivariate Regression

Bivariate regression is an extension of bivariate correlations, so we will begin the discussion of OLS using some of the data from Chapter 13. These data were from a sample drawn from the Juvenile Defendants in Criminal Courts (JDCC; see Data Sources 11.1). They measure the number of charges filed against each defendant, and the dollar amount of the fine imposed. (This sample has been narrowed to defendants who were convicted and penalized with fines.) The data from Table 13.4 are reproduced in Table 14.1. They are sorted in ascending order first according to the charges variable (the IV). Sorting them this way prepares the data for easy plotting on a graph.

Table 14.1 Number of Charges and Fine Amount	
Charges (x)	Fine (y)
1	120
1	250
1	305
2	500
3	800
4	255
5	750
5	1,000

Figure 14.1 Scatterplot of Charges and Fine Amount
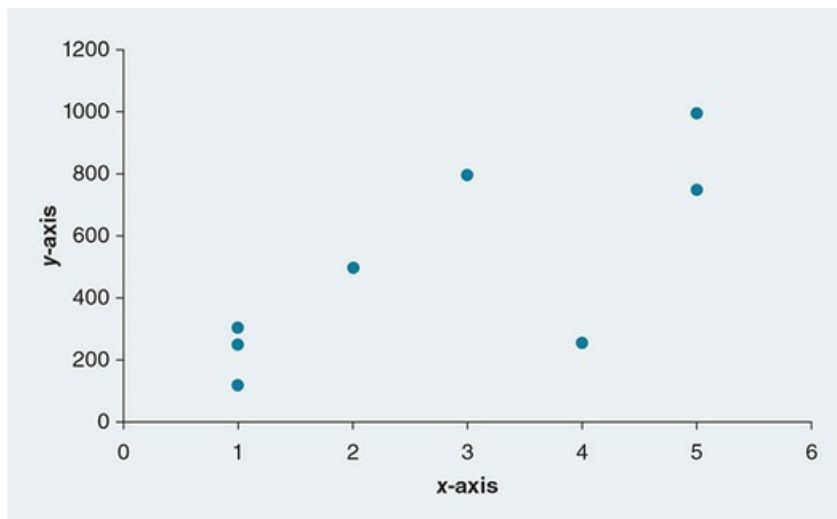
Every x score in Table 14.1 has a corresponding y score. These scores form the graphing coordinates (x, y). These coordinates can be graphed on a scatterplot like that in Figure 14.1. You can find each juvenile on the graph according to that person's x and y scores.

We already know that these two variables are correlated. This means we know they relate to one another insofar as they increase in tandem. Now we want to find out how well charges *predict* fines. In other words, we are looking to estimate what someone's score on the DV would be, given that person's score on the IV. In fact, we want to be able to predict someone's score on the DV even for values not in the data set. For example, what fine would we expect a juvenile defendant with six charges to receive? The value "6" does not occur in the data in Table 14.1, but we can predict it using regression. Correlation does not allow us to do this.

Think about drawing a line in the scatterplot in Figure 14.1. You want the line to come as close as possible to as many of data points as possible. What would this line look like? Would it have a positive or a negative slope? Where would it cross the y (vertical) axis? How steep would it be? Answering these questions will inform us as to how well x predicts y. The line of best fit is the one that comes closer to each of the data points than would any other line that could be drawn. The following formula is that for a line. You might recognize it from prior algebra classes.

$$\hat{y} = a + bx, \qquad\qquad Formula\ 14(1)$$

where

> $\hat{y}$ = the predicted value of y at x,
> a = the y intercept,
> b = the slope coefficient, and
> x = the raw values of the IV.

The predicted values of y —symbolized $\hat{y}$ (pronounced "y hat")—are calculated based on the intercept (a) and a certain slope coefficient (b) that is constant across all values of x. The **intercept** is the point at which the line of best fit crosses the y axis. The intercept is the value of y when x is zero. The **slope** coefficient conveys information about the steepness of that line. For instance, in a random sample of men, suppose the mean number of arrests is 1.2. If you randomly selected one of the men and wanted to make your best guess as to how many times he had been arrested, your guess would be 1.2. This is the estimate that will most minimize the number of times you are wrong. Now suppose you know each man's age. In other words, referencing Formula 14(1), you now have more than a: you have x, too. The question becomes whether knowing x significantly improves your ability to predict any given person's number of arrests, above and beyond the basic

predictive power of the mean. This is what the slope coefficient represents.

Intercept: The point at which the regression line crosses the y axis; also the value of y when $x = 0$.

Slope: The steepness of the regression line and a measure of the change in the dependent variable produced by a one-unit increase in an independent variable.

Once the line given by Formula 14(1) is estimated, you have two different sets of DV scores: the empirical or observed scores (these are the y values), and the predicted values (the $\hat{y}$ scores). If the IV is a good predictor of the DV, these two sets of numbers will be very similar to one another. If the IV does not predict the DV very well, the predicted values will differ substantially from the empirical values. The difference between a predicted value and an empirical value is called a **residual** (also called an error term). Small residuals indicate that x is a useful predictor of y, whereas large residuals suggest that x does not add much to our understanding of y. Knowing men's ages, for example, will improve your ability to predict the number of times they have been arrested, because age is closely linked with criminal offending. Knowing their hair color, on the other hand, will not help predict arrests. The variable *age* will shrink the residuals, but *hair color* will not.

Residual: The difference between a predicted value and an empirical value on a dependent variable.

Ordinary least squares regression provides the formula that minimizes the residuals. If you were to add up all the error terms in a data set after running an OLS model, the total amount of error will be smaller than that you could get using any other regression approach. For this reason, OLS is described with the acronym BLUE: It is the best linear unbiased estimator.

To construct the regression line, b and a must be calculated. In OLS, the slope coefficient b is calculated as

$$b = \frac{N\sum xy - \sum x \sum y}{N\sum x^2 - (\sum x)^2} \qquad \qquad Formula\ 14(2)$$

Once b is known, a can be calculated from the following formula:

$$a = \bar{y} - b\bar{x}, \qquad\qquad \text{Formula 14(3)}$$

where

$\bar{y}$ = the mean of the DV,

b = the slope coefficient, and

$\bar{x}$ = the mean of the IV.

Now we can construct the regression equation for our data. The first step is to calculate the slope coefficient b. In Table 13.4 in Chapter 13, we computed the sums required by Formula 14(2), so these numbers can be pulled from that table and entered into the formula:

$$b = \frac{N\sum xy - \sum x \sum y}{N\sum x^2 - (\sum x)^2} = \frac{8(13,845) - 22(3,980)}{8(82) - 22^2}$$

$$= \frac{110,760 - 87,560}{656 - 484} = \frac{23,200}{172}$$

= 134.88

The slope b is 134.88, meaning that for every one-unit increase in the number of charges against a defendant, the expected fine rises by \$134.88. This is helpful information, but we cannot yet generate predicted values because we do not have a starting point to build from. Finding a will do that. We first need the DV and IV means. These are found using the mean formula with which you are familiar:

$$\bar{y} = \frac{\sum y}{N} = \frac{3,980}{8} = 497.50 \quad \bar{x} = \frac{\sum x}{N} = \frac{22}{8} = 2.75$$

Plugging the means and b into Formula 14(3) yields

$$a = \bar{y} - b\bar{x}$$

a = 497.50 − 134.88(2.75)

= 497.50 − 370.92

= 126.58

Now the entire regression equation can be constructed using the pieces we just computed:

$\hat{Y}$ = 126.58 + 134.88*x*

This equation tells us that the regression line crosses the *y* -axis at 126.58 and that every one-unit increase in *x* produces a 134.88-unit increase in *y*. The equation can be used to predict people's fines on the basis of the number of charges against them. This is accomplished by entering a given value for *x* and solving the equation. For a person with two charges, the predicted fine is

$\hat{y}$ = 126.58 + 134.88*x* = 126.58 + 134.88(2) = 396.34

Looking back at Table 14.1, the person with two charges received a fine of $500. Our predicted value (is roughly $100 off. The residual is 500 − 396.34 = 103.66. Even though we found a strong correlation between this IV and DV in Chapter 13 (recall that *r* = .74), it is clear from the regression residual that there is room for improvement.

Learning Check 14.3

✓

Try calculating the predicted values of *y* for the remaining values in the data set. Compare each predicted value to its corresponding empirical value by calculating residuals. How close are they? What additional variables do you think would reduce the residuals?

Let's try using a value that is not in the data set. Nobody in Table 14.1 had six charges, but we can still use the regression equation to estimate what fine a person with six charges would receive. Plugging in this value of *x* yields

$\hat{y} = 126.58 + 134.88(6) = 935.86$

This regression equation predicts that a juvenile defendant facing six charges will receive a fine of $935.86 (if he or she is penalized with a fine). We cannot compute the residual for this value because we do not have an observed value of *y* against which to compare the predicted value.

Inferential Regression Analysis: Testing for the Significance of *b*

The most common use of regression analysis in criminal justice and criminology research is in the context of hypothesis testing. Just like the correlation coefficient *r*, the slope coefficient *b* does not itself determine whether the null hypothesis should be rejected. The slope coefficient *b* is also unstandardized. This means that it is presented in the DVs' original units. This makes it impossible to figure out whether *b* is "large" or "small." The measurement of the DV can have a dramatic impact on the slope coefficient. For instance, if a researcher trying to predict the length of prison sentences imposed on people convicted of burglary measures the DV ("sentence length") in days and finds that gender (male or female) has a slope of *b* = 10, then the average difference between male and female offenders is only 10 days. By comparison, if sentence length is measured in months, and the researcher finds *b* = 5, then even though this slope is half the size of the previous one, the real-world meaning is substantial. A slope of *b* = 5 means there is a 5-month difference between men and women. This is a big discrepancy!

We will first discuss the procedure for determining whether *b* is statistically significant. If it is not, there is no point converting it to a standardized metric. If it is, then a researcher would take the next step of standardizing it. To figure out whether *b* is significant, a five-step hypothesis test must be conducted. We will conduct this test on the *b* from above using α = .05.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

The null hypothesis in regression is generally that there is no relationship between the IV and the DV and, therefore, that the slope coefficient is zero. The alternative hypothesis is usually two-tailed; one-tailed tests are used only when there is a compelling reason to do so. Here, we will use a two-tailed test because this is the more customary course of action. Two-tailed tests are more conservative than one-tailed tests because they make it harder to reject the null, thereby reducing the risk of a Type I error (refer back to Table 9.1 and the accompanying discussion if you need a refresher). The null and alternative hypotheses are phrased in terms of the population parameters. The default assumption (i.e., the null) is always that *B* = 0. What we are looking for in an inferential test is evidence to lead us to believe that *b* is insignificantly different from zero. In regression, *B* symbolizes the population slope coefficient. The hypotheses are

H_0: *B* = 0

H_1: *B* ≠ 0

Step 2. Identify the distribution and compute the degrees of freedom.

The *t* distribution is the one typically used in regression. When the sample size is large, *z* can be used instead; however, since *t* can accommodate any sample size, it is more efficient to simply use that distribution in all circumstances. In bivariate regression, the degrees of freedom (*df*) are calculated as

$$df = N - 2 \hspace{4cm} Formula\ 14(4)$$

Here,

$$df = 8 - 2 = 6$$

*Step 3. Identify the critical value and state the decision rule.*

With a two-tailed test, α = .05, and *df* = 6, t_{crit} = ±2.447. The decision rule states that *if* t_{obt} *is either < −2.447 or > 2.447* , H_0 *will be rejected.*

Step 4. Compute the obtained value of the test statistic.

The first portion of this step entails calculating *b* and *a* in order to construct the regression equation in Formula 14(1). We have already done this; recall that the regression equation is $\hat{y}$ = 126.58 + 134.88x. Just like all other statistics, *b* has a sampling distribution. See Figure 14.2. The distribution centers on zero because the null predicts that the variables are not related. We need to find out whether *b* is either large enough or small enough to lead us to believe that *B* is actually greater than or less than zero, respectively.

Finding out whether *b* is statistically significant is a two-step process. First, we compute this coefficient's standard error, symbolized SE_b. The standard error is the standard deviation of the sampling distribution depicted in Figure 14.2. The standard error is important because, all else being equal, slope coefficients with larger standard errors are less trustworthy than those with smaller standard errors. A large standard error means that there is substantial uncertainty as to the accuracy of the sample slope coefficient *b* as an estimate of the population slope *B*.

Figure 14.2 The Sampling Distribution of Slope Coefficients
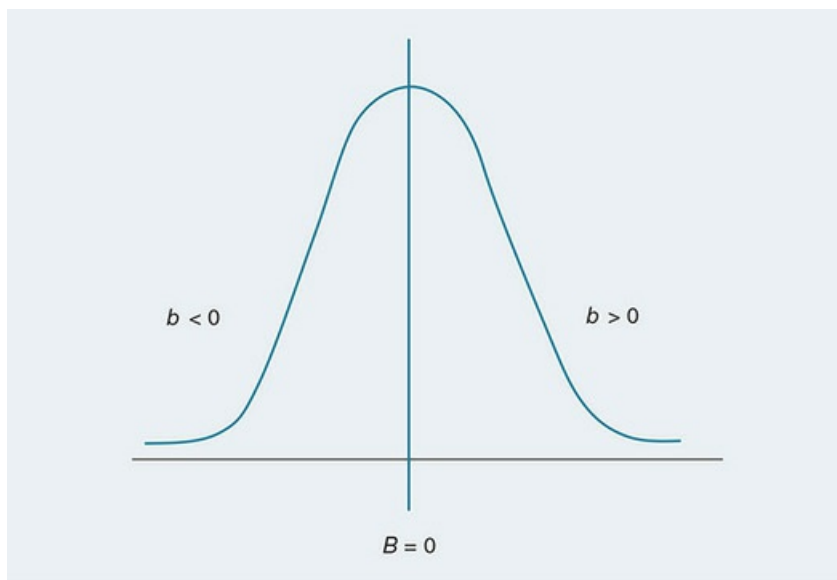
The standard error of the sampling distribution for a given regression coefficient (SE_b) is computed as

$$SE_b = \frac{s_y}{s_x}\sqrt{\frac{1-r^2}{N-2}} \qquad\qquad Formula\ 14(5)$$

where

> s_y = the standard deviation of y
>
> s_x = the standard deviation of x
>
> r = the correlation between x and y

Recall that standard deviations are the mean deviation scores for a particular variable. The standard error of the sampling distribution for a given slope coefficient is a function of the standard deviation of the DV, the standard deviation of the IV in question, and the correlation between the two. This provides a measure of the strength of the association between the two variables that simultaneously accounts for the amount of variance in each. All else being equal, more variance (i.e., larger standard deviations) suggests less confidence in an estimate. Large standard deviations produce a larger standard error, which in turn reduces the chances that a slope coefficient will be found to be statistically significant.

Here, the standard deviation of x (the IV) is 1.75 and the standard deviation of y (the DV) is 317.89. The correlation between these two variables is .74. Plugging these numbers into Formula 14(5) and solving produces

$$SE_b = \frac{317.89}{1.75}\sqrt{\frac{1-.74^2}{8-2}} = 181.65\sqrt{\frac{1-.55}{6}} = 181.65\sqrt{\frac{.45}{6}}$$
$$= 181.65\sqrt{.08}$$

$= 181.65(.28)$

$= 50.86$

This is the standard error of the slope coefficient's sampling distribution. Now SE_b can be entered into the t_{obt} formula, which is

$$t_{obt} = \frac{b}{SE_b} \qquad\qquad Formula\ 14(6)$$

The obtained value of t is the ratio between the slope coefficient and its standard error. Entering our numbers into the equation results in

$$t_{obt} = \frac{134.88}{50.86} = 2.65$$

Step 4 is complete! The obtained value of t is 2.65. We can now make a decision about the statistical

significance of *b*.

Step 5. Make a decision about the null and state the substantive conclusion.

The decision rule stated that the null would be rejected if t_{obt} turned out to be either <−2.447 or >2.447. Since 2.65 is greater than 2.447, the null is rejected. The slope is statistically significant at an alpha of .05. There is a positive relationship between the extent to which a prison is over capacity and the number of major disturbances that occur in that institution. In other words, knowing how many inmates a prison has beyond its rated capacity helps predict the number of major disturbances that a prison will experience in a year.

As with correlation, rejecting the null requires further examination of the IV–DV relationship to determine the strength and quality of that connection. In the context of regression, a rejected null indicates that the IV exerts some level of predictive power over the DV; however, it is desirable to know the magnitude of this predictive capability. The following section describes two techniques for making this assessment.

Beyond Statistical Significance: How Well Does the Independent Variable Perform as a Predictor of the Dependent Variable?

There are two ways to assess model quality. The first is to create a standardized slope coefficient or **beta weight** (symbolized β , the Greek letter *beta*) so the slope coefficient's magnitude can be gauged. The second is to examine the coefficient of determination. Each will be discussed in turn. Remember that these techniques should be used only when the null hypothesis has been rejected: If the null is retained, the analysis stops because the conclusion is that there is no relationship between the IVs and the DVs.

> **Beta weight**: A standardized slope coefficient that ranges from −1.00 to +1.00 and can be interpreted similarly to a correlation so that the magnitude of an IV–DV relationship can be assessed.

Standardized Slope Coefficients: Beta Weights

As noted earlier, the slope coefficient b is unstandardized, which means that it is specific to the units in which the IV and DV are measured. There is no way to "eyeball" an unstandardized slope coefficient and assess its strength because there are no boundaries or benchmarks that can be used with unstandardized statistics—they are specific to whatever metric the DV is measured in. The way to solve this is to standardize b. Beta weights range between 0.00 and ±1.00 and, like correlation coefficients, rely more on guidelines than rules for interpretation of their strength. Generally speaking, betas between 0 and ±.19 are considered weak, from about ±.20 to ±.29 are moderate, ±.30 to ±.39 are strong, and anything beyond ±.40 is very strong. These ranges can vary by topic, though; subject-matter experts must decide whether a beta weight is weak or strong within the customs of their fields of study.

Standardization is accomplished as follows:

$$\beta = b\left(\frac{s_x}{s_y}\right) \hspace{3cm} Formula\ 14(7)$$

We saw in the calculation of SE_b that the The standard deviation of x is 1.75 and the standard deviation of y is 317.89. We already computed b and know that it is 134.88. Plugging these numbers into Formula 14(7), we get

$$\beta = 134.88\left(\frac{1.75}{317.89}\right) = .74$$

In this calculation, rounding would have thrown the final answer off the mark, so the division and multiplication were completed in a single step. The beta weight is .74. If this number seems familiar, it is! The correlation between these two variables is .74. Beta weights will equal correlations (within rounding error) in the bivariate context and can be interpreted the same way. A beta of .74 is very strong.

Learning Check 14.4

✓

You just learned that standardized beta weights are equal to regression coefficients in bivariate regression models. As we will see soon, however, this does not hold true when there is more than one IV. Why do you think this is? If you are not sure of the answer now, continue reading and then come back to this question.

The Quality of Prediction: The Coefficient of Determination

Beta weights help assess the magnitude of the relationship between an IV and a DV, but they do not provide information about how well the IV performs at predicting the DV. This is a substantial limitation because prediction is the heart of regression—it is the reason researchers use this technique. The coefficient of determination addresses the issue of the quality of prediction. It does this by comparing the actual, empirical values of y to the predicted values ($\hat{y}_i$). A close match between these two sets of scores indicates that x does a good job predicting y, whereas a poor correspondence signals that x is not a useful predictor. The coefficient of determination is given by

$$r_{y\hat{y}}^2 \hspace{4cm} Formula\ 14(8)$$

where $r_{y\hat{y}}$ = the correlation between the actual and predicted values of y.

The correlation between the y and $\hat{y}$ and values is computed the same way that correlations between IVs and DVs are and so will not be shown here. In real life, SPSS generates this value for you. The correlation in this example is .74. This makes the coefficient of determination

$$.74^2 = .55$$

This means that 55% of the variance in y can be attributed to the influence of x. In the context of the present example, 55% of the variance in fine amounts is attributable to the number of charges. Again, this value looks familiar—it is the same as the coefficient of determination in Chapter 13! This illustrates the close connection between correlation and regression at the bivariate level. Things get more complicated when additional IVs are added to the model, as we will see next.

Adding More Independent Variables: Multiple Regression

The problem with bivariate regression—indeed, with all bivariate hypothesis tests—is that social phenomena are usually the product of many factors, not just one. There is not just one single reason why a person commits a crime, a police officer uses excessive force, or a prison experiences a riot or other major disturbance. Bivariate analyses risk overlooking variables that might be important predictors of the DV. For instance, in the bivariate context, we could test for whether having a parent incarcerated increases an individual's propensity for crime commission. This is probably a significant factor, but it is certainly not the only one. We can add other factors, such as having experienced violence as a child, suffering from a substance-abuse disorder, and being unemployed, too. Each of these IVs might help improve our ability to understand (i.e., predict) a person's involvement in crime. The use of only one IV virtually guarantees that important predictors have been erroneously excluded and that the results of the analysis are therefore suspect, and it prevents us from conducting comprehensive, in-depth examinations of social phenomena.

Multiple regression is the answer to this problem. Multiple regression is an extension of bivariate regression and takes the form

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \qquad \textit{Formula 14(9)}$$

Revisit Formula 14(1) and compare it to Formula 14(9) to see how 14(9) expands on the original equation by including multiple IVs instead of just one. The subscripts show that each IV has its own slope coefficient. With k IVs in a given study, $\hat{y}$ is the sum of each $b_k x_k$ term and the intercept.

In multiple regression, the relationship between each IV and the DV is assessed while controlling for the effect of the other IV or IVs. The slope coefficients in multiple regression are called **partial slope coefficients** because, for each one, the relationship between the other IVs and the DV has been removed so that each partial slope represents the "pure" relationship between an IV and the DV. Each partial slope coefficient is calculated while holding all other variables in the model at their means, so researchers can see how the DV would change with a one-unit increase in the IV of interest, while holding all other variables constant. The ability to incorporate multiple predictors and to assess each one's unique contribution to is what makes multiple regression so useful.

> **Partial slope coefficient**: A slope coefficient that measures the individual impact of an independent variable on a dependent variable while holding other independent variables constant.

It is interesting and surprising that intelligence outweighed parenting in predicting children's self-control. Intelligence was, in fact, by far the strongest predictor of low self-control: More-intelligent children had more self-control relative to their peers who scored lower on intelligence tests. Paternal low self-control significantly predicted children's low self-control, but the beta was very small. The only other significant variable is sex, with boys displaying higher levels of low self-control compared to girls. The model R^2 = .225, meaning that the entire set of predictors explained 22.5% of the variance in children's self-control. Clearly, childhood intelligence is integral in the development of self-control and, ultimately, in the prevention of delinquency and

crime.

Before getting into more-complex examples, let us work briefly with a hypothetical regression equation containing two IVs, x_1 and x_2. Suppose the line is

$$\hat{y} = 1.00 + .80x_1 + 1.50x_2$$

We can substitute various values for x_1 and x_2 to find . Let's find the predicted value of the DV when $x_1 = 4$ and $x_2 = 2$:

$$\hat{y} = 1.00 + .80(4) + 1.50(2) = 1.00 + 3.20 + 3.00 = 7.20.$$

There it is! If $x_1 = 4$ and $x_2 = 2$, the DV is predicted to be 7.20.

The formulas involved in multiple regression are complex and are rarely used in the typical criminal justice and criminology research setting because of the prevalence of statistical software. We now turn to a discussion of the use of SPSS to obtain and interpret OLS regression output.

Ordinary Least Squares Regression in SPSS

As described earlier, researchers rarely fit regression models by hand. Data sets are typically far too large for this, and the prevalence of user-friendly software programs like SPSS put impressive computing power right at researchers' fingertips. Of course, the flipside of this wide availability of user-friendly interfaces is the potential for them to be used carelessly or incorrectly. People producing research must possess a solid comprehension of the theory and math underlying statistical techniques before they attempt to run analyses in SPSS or other programs. Consumers of research (e.g., police and corrections officials) need to have enough knowledge about statistics to be able to evaluate results, including spotting mistakes when they occur. Consumers can be led astray if they fail to critically examine statistics and if they do not know when to trust empirical findings and when not to. As with the techniques discussed in previous chapters, GIGO applies to regression modeling. Statistical programs will frequently run and produce results even when errors have been made. For instance, SPSS will run an OLS model when the dependent variable is nominal. The results of this test are meaningless and useless, so it is up to producers and consumers to be smart and avoid making these mistakes and being deceived by them if they do occur.

Before discussing the analytical element of running OLS models in SPSS, we should revisit the null and alternative hypotheses. In multiple regression, the null and alternative each apply to every IV. For each IV, the null predicts that the population slope coefficient B_k is zero, and the alternative predicts that it is significantly different from zero. Since the analysis in the current example has two IVs, the null and alternative are

H_0: $B_1 = 0$ and $B_2 = 0$

H_1: B_1 and/or $B_2 \neq 0$

Since each IV has its own null, it is possible for the null to be rejected for one of the variables and not for the other.

To run a regression analysis in SPSS, go to *Analyze → Regression → Linear*. This will produce the dialog box shown in Figure 14.3. Here, the Police–Public Contact Survey (PPCS; see Data Sources 2.1) is being used. The DV is the length of time a vehicle or pedestrian stop lasted. The IVs are characteristics of the respondents. Respondents' sex, age, and race are included. Age is a continuous variable (measured in years), and sex and race are nominal-level variables each coded as a **dummy variable** such that one category is 0 and the other is 1. In this example, 1 = male and 0 = female for the gender variable, and 1 = white and 0 = nonwhite for the race variable. Move the DV and IVs into their proper locations in the right-hand spaces, and then press *OK*. This will produce an output window containing the elements displayed in the following figures.

> **Dummy variable**: A two-category, nominal variable with one class coded as 0 and the other coded as 1.

The first portion of regression output you should look at is the analysis of variance (ANOVA) box. This

might sound odd since we are running a multiple regression analysis, not an ANOVA, but what this box tells you is whether the set of IVs included in the model explains a statistically significant amount of the variance in the DV. If F is not significant (meaning if $p > .05$), then the model is no good. In the event of a nonsignificant F, you should not go on to interpret and assess the remainder of the model. Your analysis is over at that point and what you must do is revisit your data, your hypothesis, or both to find out what went wrong. The problem might be conceptual rather than statistical—the IVs you predicted would impact the DV might not actually do so. There could be an error in your choice of variables to represent theoretical concepts, or there could be a deeper flaw affecting the theory itself. Before you consider possible conceptual issues, check the data to make sure the problem is not caused by a simple coding error.

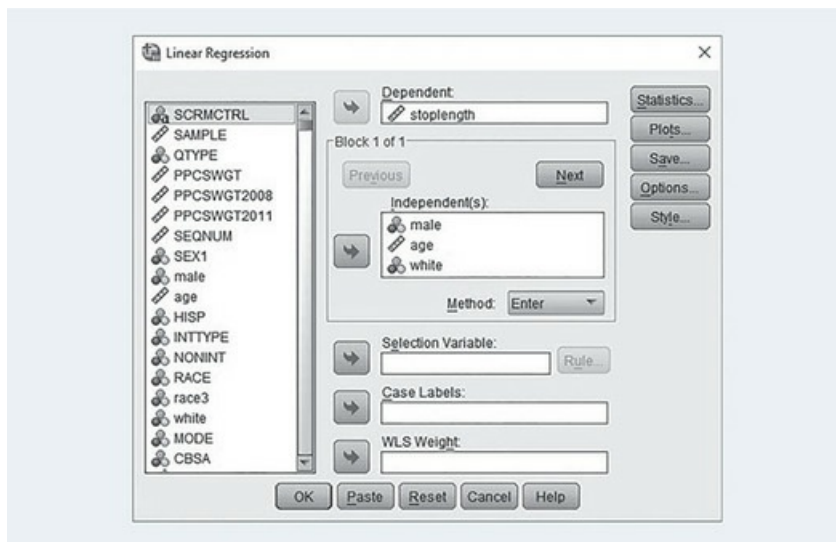**Figure 14.3** Running a Multiple Regression Analysis in SPSS

Figure 14.4 SPSS Regression Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.090ᵃ	.008	.007	9.540

a. Predictors: (Constant), Race (1 = white), Sex (1 = male), Age

ANOVAᵃ

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3181.960	3	1060.653	11.654	.000ᵇ
	Residual	387886.639	4262	91.010		
	Total	391068.599	4265			

a. Dependent Variable: LENGTH OF STOP
b. Predictors: (Constant), Race (1 = white), Sex (1 = male), Age

Coefficientsᵃ

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	12.133	.526		23.078	.000
	Sex (1 = male)	.948	.293	.049	3.240	.001
	Age	−.045	.009	−.075	−4.892	.000
	Race (1 = white)	.135	.391	.005	.344	.731

a. Dependent Variable: LENGTH OF STOP

In [Figure 14.4](#), you can see that F = 11.654 and p = .000, so the amount of the variance in the DV variance that is explained by the IVs is significantly greater than zero. Note that a significant F is not by itself proof that the model is good—this is a necessary but insufficient condition for a high-quality regression model.

Second, look at the "R Square column" in the "Model Summary" box. This is the multiple coefficient of determination and indicates the proportion of the variance in the DV that is explained by all the IVs combined. It is an indication of the overall explanatory power of the model. There are no specific rules for evaluating R square. Generally, values up to .20 are considered fairly low, .21 to .30 are moderate, .31 to .40 are good, and anything beyond .41 is very good. Here, R square is .008, meaning the IVs that we selected explain a trivial .8% of the variance in stop length. There are definitely important variables omitted from this model.

Third, go to the "Coefficients" box at the bottom of the output to see the unstandardized b values, standardized beta weights, and significance test results. The "Unstandardized Coefficients: B" column contains the slope for each variable (the constant is the intercept). The IVs age and sex are statistically significant, but race is not. We know this because the "Sig." values (i.e., p values) for sex and age are both less than .05, but the p value for race is much greater than .05.

Since age and sex are statistically significant, we can interpret their meaning to the model. The unstandardized slope for age is b = −.045. This means that each one-year increase in a person's age is associated with a reduction of .045 minutes in the total length of the stop. Older people's stops are shorter, on average, than younger people's. This makes sense, because younger people are more active in deviant behavior than older people are. Police officers probably take more time with younger drivers and pedestrians they stop to make sure they are not engaged in illegal activity.

Since sex is a dummy variable, the interpretation of b is a bit different from the interpretation of the slope of a continuous predictor. Dummy variables' slopes are comparisons between the two categories. Here, since female is coded 0 and male is coded 1, the slope coefficient b = .948 indicates that males' stops last an average of .948 minutes longer than females' stops. This finding, like that for age, makes sense in light of the fact that males commit more crime than females do. Police might subject males to more scrutiny, which extends the length of the stop.

The Beta column shows the standardized values. The utility of beta weights over b values is the ability to compare the relative strength of each IV. Using the unstandardized slopes results in an apples-to-oranges type of comparison. We are left not knowing whether age or gender is a stronger predictor of stop length. Beta weights answer this question. You can see that β = .049 for sex and β = −.075 for age. Since −.075 represents a stronger relationship than .049 does (it is the absolute value we are examining here), we can conclude that age is the more impactful of the two. Still, −.075 is very small.

> Research Example 14.2 Does Having a Close Black Friend Reduce Whites' Concerns About Crime?
>
> Mears, Mancini, and Stewart (2009) sought to uncover whether whites' concerns about crime as a local and as a national problem were affected by whether or not those whites had at least one close friend who was black. Concern about crime was the DV in this

study. White respondents expressed their attitudes about crime on a 4-point scale where higher values indicated greater concern. The researchers ran an OLS regression model and arrived at the following results with respect to whites' concerns about local crime.

The authors found, contrary to what they had hypothesized, that having a close friend who was black actually *increased* whites' concerns about crime. You can see this in the fact that the slope coefficient for *have a black friend* is statistically significant ($p < .05$) and positive. Age was also related to concern, with older respondents expressing more worry about local crime. Income had a negative slope coefficient such that a one-unit increase in annual income was associated with a .06 reduction in concern. Finally, living in an urban area substantially heightened whites' worry about crime—looking at the beta weights, you can see that *urban* is the strongest IV in the model. The researchers concluded that for whites living in urban areas, having black friends might make the crime problem seem more real and immediate because they are exposed to it vicariously through these friends.

IV	DV: Whites' Concern About Local Crime		
	b	SE_b	β
Have a black friend	.15*	.07	.08
Age	.06**	.02	.10
Sex	−.05	.06	−.03
Education	−.01	.03	−.01
Income	−.06*	.03	−.08
Urban	.48***	.09	.20
Intercept	1.80***	.13	
R^2 = .256			

Source: Adapted from Table 2 in Mears et al. (2009).

*$p < .05$; **$p < .01$; ***$p < .001$.

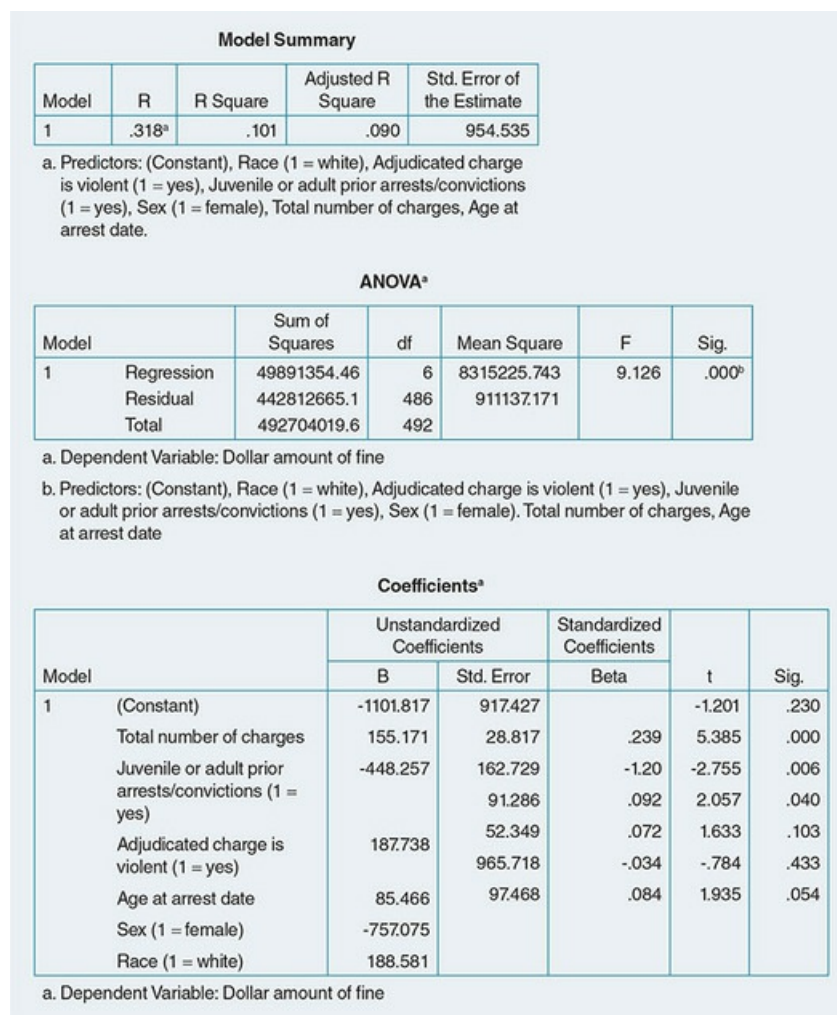*Source:* Adapted from Table 2 in Mears et al. (2009).

p < .05;

p < .01;

p < .001.

For another example, we will turn to the JDCC analysis predicting the dollar amount of fine imposed on juveniles adjudicated delinquent and ordered to pay a fine. We calculated the bivariate equation using "number of charges" as the IV and a small sample of eight juveniles. What happens when we use the full data set and add relevant case characteristics and information about the juvenile defendants? Figure 14.5 shows the OLS output. The variables measuring the number of charges and juveniles' ages at the time of arrest are continuous, and the rest are dummy coded (i.e., 0 and 1). The output shows the coding, so you can see which categories were coded as 0 and which as 1.

The model *F* is 9.126 and statistically significant ($p < .001$). The R square is .101, suggesting the six IVs collectively explain 10% of the variance in fine amount. This is a small percentage and indicates that important variables have been omitted from the model. Turning to the slopes, the variable measuring the total number of charges remains statistically significant ($p < .001$) and is the strongest predictor in the model (β = .239). More charges mean a higher fine. The second-largest predictor is a juvenile's history of prior arrests or convictions (measured as 0 = no priors and 1 = priors). This slope is statistically significant ($p < .01$) and moderate in strength (β = −.120). Interestingly, this slope is negative, meaning that juveniles with prior

records receive smaller fines. This seems backward at first glance but makes sense if we consider that repeat offenders are sentenced to harsher penalties. Juveniles with priors might get smaller fines because they receive other sentences along with those fines (such as probation), whereas first-time offenders are sentenced to only fines and therefore those fines are larger. The final IV that reaches statistical significance at .05 is whether the charge for which a juvenile was adjudicated delinquent was violent. The β of .092 suggests this is a small effect; fines are probably not used as the primary punishment for juveniles adjudicated delinquent on violent charges, so this small effect is not surprising. The race variable fell just shy of the customary .05 alpha level. At $p = .054$, though, most researchers would probably consider it statistically significant since it missed the cut off by only .004. It appears that white juveniles receive larger fines than nonwhite juveniles, but at β = .084, this effect is small.

Figure 14.5 SPSS Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.318[a]	.101	.090	954.535

a. Predictors: (Constant), Race (1 = white), Adjudicated charge is violent (1 = yes), Juvenile or adult prior arrests/convictions (1 = yes), Sex (1 = female), Total number of charges, Age at arrest date.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	49891354.46	6	8315225.743	9.126	.000[b]
	Residual	442812665.1	486	911137.171		
	Total	492704019.6	492			

a. Dependent Variable: Dollar amount of fine

b. Predictors: (Constant), Race (1 = white), Adjudicated charge is violent (1 = yes), Juvenile or adult prior arrests/convictions (1 = yes), Sex (1 = female). Total number of charges, Age at arrest date

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1101.817	917.427		-1.201	.230
	Total number of charges	155.171	28.817	.239	5.385	.000
	Juvenile or adult prior arrests/convictions (1 = yes)	-448.257	162.729	-1.20	-2.755	.006
			91.286	.092	2.057	.040
	Adjudicated charge is violent (1 = yes)	187.738	52.349	.072	1.633	.103
			965.718	-.034	-.784	.433
	Age at arrest date	85.466	97.468	.084	1.935	.054
	Sex (1 = female)	-757.075				
	Race (1 = white)	188.581				

a. Dependent Variable: Dollar amount of fine

The regression equation can be constructed from the output in Figure 14.5 and used to calculate predicted values. The equation is

$$\hat{y} = -1101.82 + 155.17x_{charges} - 448.26x_{priors} + 187.74x_{violent} + 85.47x_{age} - 757.08x_{sex} + 188.58x_{race}$$

Let's compute the predicted value of y (i.e., the predicted fine amount) for a white male juvenile with three charges, no prior record, who was not adjudicated of a violent crime, and was 15 at the time of arrest. Plugging these numbers in,

$\hat{y}$ = –1,101.82 +155.17(3) – 448.26(0) +187.74(0) + 85.47(15) –757.08(0) +188.58(1)

= –1,101.82 + 465.51 + 1,282.05 + 188.58

= 834.32

We would expect a juvenile with these characteristics to receive a fine of \$834.32. Try replicating this analysis, then visit Learning Check 14.6 for additional examples you can use for practice.

Learning Check 14.6

✓

Use the equation $\hat{y} = -1101.82 + 155.17x_{charges} - 448.26x_{priors} + 187.74x_{violent} + 85.47x_{age} - 757.08x_{sex} + 188.58x_{race}$ to calculate the predicted value of y for each of the following juveniles:

1. Nonwhite female age 16 with no priors who had three charges and was adjudicated guilty of a violent offense.
2. White male age 17 with priors who had two charges and was adjudicated guilty of a nonviolent offense.
3. Nonwwhite male age 15 with no priors who had one charge and was adjudicated guilty of a nonviolent offense.
4. White male age 15 with priors who had one charge and was adjudicated guilty of a violent offense.

Research Example 14.3 Do Multiple Homicide Offenders Specialize in Killing?

In Chapter 11, you read about a study by Wright, Pratt, and DeLisi (2008) wherein the researchers examined whether multiple homicide offenders (MHOs) differed significantly from single homicide offenders (SHOs) in terms of the diversity of offending. Diversity was measured as a continuous variable with higher values indicating a greater spectrum of offending. We saw in Chapter 11 that Wright et al. first ran a t test to check for differences between the group means for MHOs and SHOs; that test showed that the difference was not statistically significant. Given that bivariate results are untrustworthy for the reasons discussed in this chapter, Wright et al. ran a multiple OLS regression model. They found the following results.

IV	DV: Diversity Index		
	b	SE$_b$	β
Offender type: SHO	.004	.163	.001
Age	.068*	.008	.355
Age of onset	−.156*	.011	−.551
White	−.375*	.145	−.090
Constant	−1.346		
R^2 = .256			

Source: Adapted from Table 2 in Wright et al. (2008).

*$p < .05$; **$p < .01$.
Source: Adapted from Table 2 in Wright et al. (2008).

$p < .05$; ** $p < .01$.

Offenders' current age, the age at which they started offending, and their race were statistically significant predictors of offending diversity. As foreshadowed by the nonsignificant t test, the coefficient for *Offender type: SHO* was not statistically significant. A dichotomous IV like that used here, where people in the sample were divided into two groups classified as either *MHOs* or *SHOs* , is a dummy variable. The slope is interpreted just as it is with a continuous IV: It is the amount of predicted change in the DV that occurs with a one-unit increase in the IV. Here, you can see that being an SHO increased offending diversity by only .004 of a unit. This is a very trivial change and was not statistically significant. Race is also a dummy variable, because offenders in the sample were classified as either *nonwhite* or *white*. Can you interpret this slope coefficient with respect to what it means about race as a predictor of the diversity index? If you said that white offenders score significantly lower on the diversity index, you are correct.

Wright et al.'s (2008) multiple regression model confirmed that MHOs and SHOs do not differ in terms of offending diversity. This suggests that MHOs do not specialize in killing; to the contrary, they display as much diversity as other types of homicide offenders. The theoretical implication of this finding is that the theories that have been developed to help explain violent offending might be applicable to MHOs because these people are similar to other offenders.

Alternatives to Ordinary Least Squares Regression

This chapter has focused on OLS regression because it is the most basic regression modeling strategy and is generally the starting point for the study of regression. Recall that OLS can be used only when a DV is continuous and (reasonably) normally distributed. As you might have already guessed, researchers often confront DVs that do not meet one or both of these criteria. A DV might be nominal or ordinal, or it could be heavily skewed. It might contain a substantial number of zeros, which throws the OLS calculations off.

Various regression techniques are available for situations when the DV violates the OLS assumptions. We will not go into them in detail, but it is worth knowing the names of at least some of the most common models. Each one is specific to a certain kind of DV.

When a DV is dichotomous (i.e., categorical with two classes), binary logistic regression is used. Binary logistic regression calculates the probability of any given case in the data set falling into the category of interest. The DV might be, for instance, whether a defendant was sentenced to prison. Binary logistic regression would tell the user whether and to what extent each of the IVs (offense type, prior record, and so on) increases the probability of a defendant being sent to prison. Additionally, a severely skewed DV can sometimes be dichotomized (i.e., split in half) and put through binary logistic rather than OLS. Binary logistic regression is very popular. There are two other types of logistic regression that are used for categorical DVs. For a nominal DV with three or more classes, multinomial logistic is available. For an ordinal DV with three or more classes, ordered logistic can be employed. Each of these types of models functions by sequentially comparing pairs of classes.

A technique called Poisson regression is used when a DV is made of count data (sometimes also called event data). Count data are phenomena such as the number of times the people in a sample have been arrested or the number of homicides that occur in a city in a year. Throughout this book, we have treated count data as being part of the continuous level of measurement, and this is often perfectly fine to do, but it can cause problems if done so in the regression realm. Part of the problem is that count data are usually highly skewed, because most of the DVs that are studied are relatively rare events (violent crimes, criminal convictions, and the like). In a data set containing the number of homicides experienced by a sample of cities, the vast majority of cities will cluster at the low end of the distribution, a few will be in the middle, and a small minority will extend out in the tail. Additionally, OLS assumes that the DV can take on any value, including negatives and decimals, but count data can only be positive, whole numbers (integers). For these reasons, OLS is inappropriate for count data. The Poisson distribution is a probability distribution that works well for count data. Poisson regression is, not surprisingly, based on the Poisson probability distribution. This type of regression is common in criminal justice and criminology research, although most data violate the assumptions for using the Poisson distribution and so researchers typically turn to a technique called negative binomial regression. This type of regression works well with count data containing a lot of zeros.

There are also modeling strategies available for times when none of the previously discussed techniques is adequate given the type of hypothesis being tested or the type of data a researcher is working with. A

technique called structural equation modeling (SEM) can be applied when a researcher has multiple DVs rather than just one. A researcher might posit that one variable causes another and that this second variable, in turn, causes a third one. Standard regression can accommodate only one DV, but SEM can handle a more complex causal structure. Multilevel modeling (also called hierarchical linear modeling) is appropriate when data are measured at two units of analysis (such as people nested within neighborhoods or prisoners nested within institutions). Basic regression assumes that all data are of the same unit of analysis, and when that is not the case, the standard-error estimates can be inaccurate, and the significance tests thrown off as a result. Both structural equation modeling and multilevel modeling are based in regression.

Research Example 14.4 Is Police Academy Performance a Predictor of Effectiveness on the Job?

In Chapter 13, you encountered a study by Henson, Reyns, Klahm, and Frank (2010). The researchers sought to determine whether recruits' performance while at the academy significantly influenced their later success as police officers. Henson et al. measured success in three ways: the scores new officers received on the annual evaluations conducted by those officers' supervisors, the number of complaints lodged against these new officers, and the number of commendations they earned. These three variables are the DVs in this study.

You saw in Chapter 13 that the bivariate correlations indicated mixed support for the prediction that academy performance was related to on-the-job performance; however, to fully assess this possible link, the researchers ran an OLS regression model. They obtained the following results.

The results from the three OLS models showed that recruits' civil service exam scores, physical agility exam scores, and overall academy ratings were—with only one exception—unrelated to on-the-job performance. The exception was the positive slope coefficient between overall academy ratings and evaluation scores ($b = .06$; $p < .01$). The demographic variables gender, age, race, and education also bore limited and inconsistent relationships with the three performance measures. These results seem to indicate that the types of information and training that recruits receive are not as clearly and directly related to on-the-job performance as would be ideal. There might be a need for police agencies to revisit their academy procedures to ensure that recruits are receiving training that is current, realistic, and practical in the context in which these recruits will be working once they are out on the street.

| | DDVs | | |
IV	Evaluation b (SE_b)	Complaints b (SE_b)	Commendations b (SE_b)
Civil service exam	−.07 (.02)	−.01 (.01)	.00 (.01)
Overall academy score	.06* (.03)	−.004 (.01)	−.003 (.01)
Physical agility rating	−.01 (.03)	.02 (.02)	.00 (.02)
Gender	−.40 (.25)	−.32* (.12)	−.07 (.14)
Age	−.04 (.02)	−.02* (.01)	.01 (.01)
Race	.13 (.23)	−.10 (.11)	.14 (.13)
Education	−.02 (.29)	−.15 (.14)	.14 (.16)
	R^2 = .048	R^2 = .098	R^2 = .000

Source: Adapted from Table 5 in Henson et al. (2010).

*$p < .01$.

Source: Adapted from Table 5 in Henson et al. (2010).

$p < .01$.

This has been a brief overview of regression and regression-based alternatives to OLS. There are many options available to accommodate hypotheses and data of all types. More information on these and other techniques is available in advanced statistics textbooks and online.

Chapter Summary

This chapter introduced you to the basics of bivariate and multiple regression analysis. Bivariate regression is an extension of bivariate correlation and is useful because correlation allows for only a determination of the association between two variables, but bivariate regression permits an examination of how well the IV acts as a predictor of the DV. When an IV emerges as a statistically significant predictor of the DV, it can then be assessed for magnitude using the standardized beta weight, β , and the coefficient of determination, r^2.

The problem with bivariate analyses of all types, though, is that every phenomenon that is studied in criminal justice and criminology is the result of a combined influence of multiple factors. In bivariate analyses, it is almost certain that one or more important IVs have been omitted. Multiple regression addresses this by allowing for the introduction of several IVs so that each one can be examined while controlling for the others' impacts. In the bivariate regression example conducted in this chapter, age of onset significantly predicted lifetime arrests; however, when education was entered into the regression model, age lost its significance and education emerged as a significant and strong predictor of lifetime arrests. This exemplifies the omitted variable bias: Failing to consider the full gamut of relevant IVs can lead to erroneous results and conclusions.

This also brings us to the end of the book. You made it! You struggled at times, but you stuck with it, and now you have a solid grasp on the fundamentals of the use of statistics in criminal justice and criminology research. You know how to calculate univariate and bivariate statistics and how to conduct hypothesis tests. Just as important, you know how to evaluate the statistics and tests conducted by other people. You know to be critical, ask questions, and always be humble in drawing conclusions because every statistical test contains some level of error, be it the Type I or Type II error rate, omitted variables, or some other source of

uncertainty. Proceed with caution and a skeptical mind when approaching statistics as either a producer or a consumer. Make GIGO a part of your life—when the information being input into the system is deficient, the conclusions are meaningless or possibly even harmful. The bottom line: Question everything!

Thinking Critically

1. People commonly make the mistake of assuming that since multiple regression analyses control for many factors and help rule out spuriousness, they are proof of causation (i.e., proof that the IV of interest causes the DV). Critically examine this misunderstanding about regression. Use the criteria for establishing causation to explain the flaw in this assumption. Refer back to Chapter 2 if you need to.

2. In criminal justice and criminology research (as in most other social sciences), regression models usually produce fairly modest R^2 values. Even explaining a relatively small fraction of the variance in a DV is frequently considered good, even though an R^2 value considered robust within social sciences (such as .40, for instance) leaves the majority of the variance unaccounted for (in this case, 60%). Why do you think this is? That is, why do regression models commonly explain less than half of the variance? (Hint: Think about the DVs that criminology and criminal justice researchers work with.)

Review Problems

1. You learned in this chapter that the key advantage of bivariate regression over correlation is that regression can be used for prediction. Explain this. How is it that regression can be used to predict values not in the data set, but correlation cannot?

2. Identify the two criteria that a DV must meet for OLS regression to be used.

3. Does OLS regression place restrictions on the levels of measurement of the IVs?

4. Explain the advantage of multiple regression over bivariate regression. What does multiple regression do that bivariate regression does not? Why is this important?

5. In a hypothetical example of five prison inmates, let us find out whether prior incarcerations influence in-prison behavior. The following table contains data on the prior incarcerations and the number of disciplinary reports filed against each inmate. Use the data to do the following:

Person	Incarcerations (x)	Disciplinary Reports (y)
A	14	4
B	1	3
C	2	8
D	5	1
E	2	3
N = 5		

1. Calculate the slope coefficient b.

2. Calculate the intercept a. $(\bar{x} = 4.80 \text{ and } \bar{y} = 3.80)$

3. Write out the full regression equation.

4. Calculate the number of disciplinary reports you would expect to be received by a person with
   1. 3 prior incarcerations.
   2. 15 prior incarcerations.

5. Using an alpha level of .05 and two-tailed alternative hypothesis, conduct a five-step hypothesis test to determine whether b is statistically significant. ($s_y = 2.59$; and $s_x = 5.36$; $r = -.09$)

6. If appropriate (i.e., if you rejected the null in Part e), calculate the beta weight.

6. Does the amount of crime in an area predict the level of police presence? The following table displays data on the number of crimes per square mile and the number of police agencies per 1,000 square miles in a sample of states. Use the data to do the following:

State	Crime per Sq. Mile (x)	Police Agencies per 1,000 Sq. Miles (y)
Georgia	7.30	9.40
Iowa	1.40	7.20
Kansas	1.30	4.40
Maine	1.00	3.90
Minnesota	1.90	5.30
Missouri	3.50	8.40
Nebraska	.70	3.20
Pennsylvania	7.60	24.90
South Carolina	6.90	8.40
Washington State	3.80	3.70
N = 10		

1. Calculate the slope coefficient b.

2. Calculate the intercept a. $(\bar{x} = 3.54$ and $\bar{y} = 7.88)$

3. Write out the full regression equation.

4. Calculate how many police agencies per 1,000 square miles you would expect in a state with
    1. 5 crimes per square mile.
    2. 10 crimes per square mile.

5. Using an alpha level of .05 and two-tailed alternative hypothesis, conduct a five-step hypothesis test to determine whether the IV is a significant predictor of the DV. $s_y = 6.39$; and $s_x = 2.76$; $r = -.71$

6. If appropriate (i.e., if you rejected the null in Part e), calculate the beta weight.

7. Research has found that socioeconomic disadvantage is one of the strongest and most consistent predictors of crime. Negative socioeconomic factors such as poverty and unemployment have been shown to profoundly impact crime rates. The following table contains a random sample of states. The IV consists of data from the U.S. Census Bureau on the percentage of adults in the civilian labor force that was unemployed. The DV is UCR-derived violent crime rates per 1,000 persons. Use the data to do the following:

State	Unemployed (x)	Violent Crime Rate (y)
Alaska	8.70	6.33
Arkansas	7.30	5.18
Kentucky	7.60	2.59
Louisiana	7.60	6.20
Montana	5.60	2.54
New Jersey	6.90	3.11
Pennsylvania	6.80	3.81
Virginia	5.40	2.27
Vermont	5.60	1.31
Wyoming	4.50	2.28
N = 10		

1. Calculate the slope coefficient b.
2. Calculate the intercept a. $(\bar{x} = 6.60$ and $\bar{y} = 3.56)$
3. Write out the full regression equation.
4. Calculate how many violent crimes per 1,000 citizens you would expect in a state with
    1. a 4% unemployment rate.
    2. an 8% unemployment rate.
5. Using an alpha level of .05 and two-tailed alternative hypothesis, conduct a five-step hypothesis test to determine whether the IV is a significant predictor of the DV. $s_y = 1.76$; and $s_x = 1.29$; $r = -.79$
6. If appropriate (i.e., if you rejected the null in Part e), calculate the beta weight.

8. Deterrence theory suggests that as the number of crimes that police solve goes up, crime should decrease because would-be offenders are scared by the belief that there is a good chance that they would be caught and punished if they commit an offense. The following table contains regional data from the UCR. The IV is *clearance* and is the percentage of violent crimes that were cleared by arrest or exceptional means in one year. The DV is *violent crime rate* and is the number of violent crimes that occurred in the following year (rate per 1,000 residents). Use the data to do the following.

Region	Clearance (x)	Violent Crime Rate (y)
New England	48.00	2.37
Mid-Atlantic	51.30	3.59
East North Central	35.40	3.83
West North Central	47.70	2.98
South Atlantic	47.30	4.83
East South Central	45.60	4.14
West South Central	43.70	5.32
Mountain	44.50	3.74
Pacific	44.10	3.93
N = 9		

1. Calculate the slope coefficient b .
2. Calculate the intercept a. $(\bar{x} = 45.29$ and $\bar{y} = 3.86)$
3. Write out the full regression equation.
4. Calculate the rate of violent crimes per 1,000 citizens you would expect in a region where
    1. 30% of violent crimes were cleared.

2. 50% of violent crimes were cleared.

5. Using an alpha level of .05 and two-tailed alternative hypothesis, conduct a five-step hypothesis test to determine whether the IV is a significant predictor of the DV. s_Y = .88; and s_X = 4.41; r = −.25

6. If appropriate (i.e., if you rejected the null in Part e), calculate the beta weight.

9. Let us now consider the possible relationship between poverty and crime. The following table contains a random sample of states and the violent crime rate (the DV). The IV is the percentage of families living below the poverty line. Use the table to do the following.

State	Poverty (x)	Violent Crime Rate (y)
Pennsylvania	8.30	3.81
Vermont	6.90	1.31
Florida	9.50	6.12
Hawaii	6.80	2.75
Iowa	7.30	2.79
Missouri	9.80	4.92
Mississippi	9.90	2.81
South Carolina	11.90	6.71
Tennessee	12.20	6.68
Texas	13.20	4.91
N = 10		

1. Calculate the slope coefficient b.
2. Calculate the intercept a . $(\bar{x} = 9.58 \ and \ \bar{y} = 4.28)$
3. Write out the full regression equation.
4. Calculate the rate of violent crimes per 1,000 citizens you would expect in a state with
    1. a 5% poverty rate.
    2. a 10% poverty rate.

5. Using an alpha level of .05 and two-tailed alternative hypothesis, conduct a five-step hypothesis test to determine whether the IV is a significant predictor of the DV. s_Y = 1.87; and s_X = 2.29; r = .78

6. If appropriate (i.e., if you rejected the null in Part e), calculate the beta weight.

10. Is there a relationship between the handgun murder rate and the rate of murders committed with knives? The table shows data from a sample of states. Consider the knife murder rate to be the DV. Use the table to do the following.

State	Handgun Murder Rate (x)	Knife Murder Rate (y)
Nebraska	1.22	.45
Wisconsin	1.15	.39
Arkansas	1.87	.73
North Carolina	2.59	.52
Virginia	1.37	.52
California	2.77	.79
Washington State	1.13	.53
N = 7		

1. Calculate the slope coefficient b.
2. Calculate the intercept a. $(\bar{x} = 1.73 \ and \ \bar{y} = .56)$

3. Write out the full regression equation.
4. Calculate the rate of knife murders per 1,000 citizens you would expect in a state with
    1. a handgun murder rate of 3.00.
    2. a handgun murder rate of 1.75.
5. Using an alpha level of .05 and two-tailed alternative hypothesis, conduct a five-step hypothesis test to determine whether the IV is a significant predictor of the DV. s_y = .15; s_x = .70; r = −.68
6. If appropriate (i.e., if you rejected the null in Part e), calculate the beta weight.

11. The data set *JDCC for Chapter 14.sav* at **www.sagepub.com/gau** contains data from the JDCC survey. The sample has been narrowed to those facing drug charges who were convicted and sentenced to probation. The DV (probation) is the number of months to which convicted juveniles were sentenced. The IVs are the number of charges filed against each defendant (*charges*) and whether the defendant had a prior record of juvenile arrests or convictions (*priors*) . Run a bivariate regression using the DV *probation* and the IV *charges* to determine whether the number of charges significantly predicts the severity of the probation sentence at the bivariate level. Do not include the other variable. Then do the following.
    1. Report the value of the ANOVA *F* and determine whether you would reject the null at an alpha of .05. What does the rejection or retention of *F* mean?
    2. Report the *R* -square value. Using the guidelines provided in the text, interpret the strength of the explanatory power of the IV.
    3. Using the numbers in the output, write out the bivariate regression equation for $\hat{y}$.
    4. Determine whether *b* is statistically significant at α = .05, and explain how you arrived at this decision.
    5. If appropriate, identify the value of the beta weight.
    6. What is your substantive conclusion? Do the number of charges predict probation sentences? Explain your answer.

12. Keeping the same data set and model from Question 11, add *priors* to the model to account for whether or not each defendant had a prior record of arrests or convictions as a juvenile. Recall from the chapter that this is a binary (dummy) variable with "yes" coded as 1 and "no" as zero, so what you will be looking for in the output is a comparison between those with and without priors (i.e., those who had and had not been arrested or convicted before). The slope coefficient will tell you the impact of having a prior record as opposed to not having one (the larger the slope, the stronger the impact).
    1. Report the value of the ANOVA *F* and determine whether you would reject the null at an alpha of .05. What does the rejection or retention of *F* mean?
    2. Report the *R* -square value. Using the guidelines provided in the text, interpret the strength of the explanatory power of the IVs.
    3. Using the numbers in the output, write out the multiple regression equation for $\hat{y}$.
    4. Determine whether each *b* is statistically significant at α = .05, and explain how you arrived at these decisions.
    5. If appropriate, identify the value of the beta weight for each significant IV and compare them. Which one is stronger? Weaker?
    6. What is your substantive conclusion? Do age and/or number of charges seem to be strong predictors of probation sentences? Explain your answer.

13. Using the multiple regression equation you wrote in Question 12, do the following:
    1. Setting *charges* at its mean (2.09), calculate the predicted probation sentence for a juvenile who does not have a prior record (i.e., x_{priors} = 0 in the equation).
    2. Holding *charges* constant at its mean, calculate the predicted probation sentence for a juvenile who does have a prior record (i.e., x_{priors} = 1).
    3. By how much did the predicted sentence change? Was this change an increase or decrease?

14. Using the multiple regression equation you wrote in Question 12, do the following:
    1. Setting *priors* at its mean (.79), calculate the predicted probation sentence for a juvenile facing two charges.
    2. Holding *age* constant at its mean, calculate the predicted probation sentence for a juvenile facing five charges.
    3. By how much did the predicted sentence change? Was this change an increase or a decrease?

15. The data file *Socioeconomics and Violence for Chapter 14.sav* at **www.sagepub.com/gau** contains a sample of states. Violent crime rates (the variable *violentrate*) is the DV. The IVs are the percentage of households receiving food stamps (Supplemental Nutrition Assistance Program, or SNAP; *snap*) ; the percentage of the adult civilian workforce that is unemployed (*unemployed*) ; and the percentage of families that are below the poverty line (*poverty*) . Run a bivariate regression using the DV *violentrate* and the IV *unemployed* to test for a relationship between unemployment and violent crime. Do not include the other variables. Then do the following:

1. Report the value of the ANOVA F and determine whether you would reject the null at an alpha of .05. What does the rejection or retention of F mean?
2. Report the R-square value. Using the guidelines provided in the text, interpret the strength of the explanatory power of the IV.
3. Using the numbers in the output, write out the bivariate regression equation for $\hat{y}$.
4. Determine whether b is statistically significant at $\alpha = .05$, and explain how you arrived at this decision.
5. If appropriate, identify the value of the beta weight.
6. What is your substantive conclusion? Does the unemployment rate seem to be a strong predictor of violent crime? Explain your answer.

16. Using the data file *Socioeconomics and Violence for Chapter 14. sav* , run a multiple regression using all three IVs (*snap, unemployed* , and *poverty*) . Then do the following:
    1. Report the value of the ANOVA F and determine whether you would reject the null at an alpha of .05. What does the rejection or retention of F mean?
    2. Report the R-square value. Using the guidelines provided in the text, interpret the strength of the explanatory power of the IVs.
    3. Using the numbers in the output, write out the multiple regression equation for $\hat{y}$.
    4. Determine whether each b is statistically significant at $\alpha = .05$, and explain how you arrived at these decisions.
    5. Identify the beta weights for each significant IV and compare them. Which one is strongest? Weakest?
    6. What is your substantive conclusion? Do these variables seem to be strong predictors of violent crime rates? Explain your answer.

17. Using the multiple regression equation you wrote in Question 16, find the predicted violent crime rate for a state where 15% of households receive benefits, 10% of people are unemployed, and 16% of families live below the poverty line.

18. Using the multiple regression equation you wrote in Question 16, do the following:
    1. Setting *snap* at its mean (8.25) and *poverty* at its mean (8.94), calculate the predicted violent crime rate for a state with a 4% unemployment rate.
    2. Holding *snap* and *poverty* constant at their means, calculate the predicted violent crime rate for a state with a 10% unemployment rate.
    3. By how much did the predicted rate change? Was this change an increase or decrease?

19. Using the multiple regression equation you wrote in Question 16, do the following:
    1. Setting *unemployment* at its mean (6.77) and *poverty* at its mean (8.94), calculate the predicted violent crime rate for a state where 11% of households receive benefits.
    2. Holding *unemployment* and *poverty* constant at their means, calculate the predicted violent crime rate for a state where 6% of households receive benefits.
    3. By how much did the predicted rate change? Was this change an increase or a decrease?

20. Using the multiple regression equation you wrote in Question 16, do the following:
    1. Setting *unemployment* at its mean (6.77) and *snap* at its mean (8.25), calculate the predicted violent-crime rate for a state where 15% of families live below the poverty line.
    2. Holding *unemployment* and *snap* constant at their means, calculate the predicted violent-crime rate for a state where 20% of families live below the poverty line.
    3. By how much did the predicted rate change? Was this change an increase or decrease?

Key Terms

Regression analysis 336

Bivariate regression 336

Multiple regression 336

Ordinary least squares (OLS) regression 336

Intercept 337

Slope 337

Residual 338

Beta weight 345

Partial slope coefficient 347

Dummy variable 350

Glossary of Symbols and Abbreviations Introduced in This Chapter

y	An empirical value of the dependent variable
$\hat{y}$	A predicted value of the dependent variable
a	The y intercept in a regression equation
b	In a sample, the slope of the regression line
B	In a population, the slope of the regression line
SE_b	The standard error of the slope coefficient
β	Beta weight; a standardized slope coefficient
$r_{y\hat{y}}^2$	The coefficient of determination in a bivariate regression analysis
R^2	The coefficient of determination in a multiple regression analysis

Appendix A. Review of Basic Mathematical Techniques

In order to succeed in this class, you must have a solid understanding of basic arithmetic and algebra. This appendix is designed to help you review and brush up on your math skills.

Section 1: Division

You will be doing a lot of dividing throughout this book. The common division sign "÷" will not be used.

Division will always be presented in fraction format. Instead of "6 ÷ 3," for example, you will see "$\dfrac{6}{3}$." For example,

$$\frac{20}{4} = 5 \qquad \frac{90}{10} = 9$$

Try the following as practice.

1. $\dfrac{6}{2} =$

2. $\dfrac{14}{2} =$

3. $\dfrac{15}{3} =$

4. $\dfrac{9}{3} =$

Section 2: Multiplication

Multiplication is another oft-used technique in this book. The common multiplication sign "×" will not be used, because in statistics, the symbol "x" represents a raw score in a data set. Employing the "×" multiplication sign would therefore introduce confusion. Parentheses will be the most commonly used multiplication sign. Also, when operands or variables are right next to one another, this is an indication that you should use multiplication. Sometimes, too, multiplication will be indicated by a dot between two numbers. For example,

$7(3) = 21$ $7 \cdot 4 = 28$ $10(3)(4) = 120$

Try the following as practice.

1. $3(4) =$
2. $9 \cdot 8 =$
3. $12(2) =$
4. $4 \cdot 5 \cdot 3 =$

Section 3: Order of Operations

Solving equations correctly requires the use of proper order of operations. The correct order is parentheses; exponents; multiplication; division; addition; subtraction. Using any other order could result in erroneous final answers. For example,

$3(5) + 2 = 15 + 2 = 17$

$(4 + 7) - 6 = 11 - 6 = 5$

$\left(\dfrac{8}{2}\right)^2 = (4)^2 = 16$

Try the following as practice.

1. $3 + 2 - 4 =$
2. $4(5) + 7 =$
3. $\dfrac{19 - 4}{5} =$
4. $5 \cdot 6 - \dfrac{16}{4} =$
5. $\left(\dfrac{16}{8}\right)\left(\dfrac{14}{2}\right) =$
6. $2^2 + 3^3 =$
7. $(3 + 2)^2 =$

Section 4: Variables

The formulas in this book require you to plug numbers into equations and solve those equations. You must, therefore, understand the basic tenets of algebra, wherein a formula contains variables, you are told the values of those variables, and you plug the values into the formula. For example,

If $x = 9$ and $y = 7$, then $x + y = 9 + 7 = 16$

If $x = 10$ and $y = 7$, then $xy = 10(7) = 70$

If $x = 2$, $y = 5$, and $z = 8$, then $\left(\dfrac{z}{x}\right)y = \left(\dfrac{8}{2}\right)5 = 4 \cdot 5 = 20$

Try the following as practice.

1. $\dfrac{x}{y}$, where $x = 12$ and $y = 3$
2. xy , where $x = 1$ and $y = 1$
3. $x + y + z$, where $x = 1$, $y = 19$, and $z = 4$
4. $\dfrac{x}{y} + 2$, where $x = 6$ and $y = 3$
5. $\left(\dfrac{x}{6}\right)y + 5$, where $x = 36$ and $y = 11$

Section 5: Negatives

There are several rules with respect to negatives. Negative numbers and positive numbers act differently when they are added, subtracted, multiplied, and divided. Positive numbers get larger as the number line is traced away from zero and toward positive infinity. Negative numbers, by contrast, get smaller as the number line is traced toward negative infinity. Adding a negative number is equivalent to subtracting a positive number. When a positive number is multiplied or divided by a negative number, the final answer is negative. When two negative numbers are multiplied or divided, the answer is positive. For example,

$5 + (-2) = 5 - 2 = 3$

$-5 + (-2) = -5 - 2 = -7$

$-10(9) = -90$ $\dfrac{-90}{-9} = 10$

Try the following as practice.

1. $-3 + (-2) =$
2. $-3 - 4 =$
3. $-5 + 3 =$
4. $3 - 8 =$
5. $(-2)^2 =$
6. $-2^2 =$
7. $(-4)(-5) =$
8. $\dfrac{-9}{3} =$

Section 6: Decimals and Rounding

This book requires you to round. Two decimal places are used here; however, your instructor might require more or fewer, so pay attention to directions. When rounding to two decimal places, you should look at the number in the third decimal place to decide whether you will round up or whether you will truncate. When the number in the third (thousandths) position is 5 or greater, you should round the number in the second (hundredths) position up. When the number in the thousandths position is 4 or less, you should truncate. The diagram below shows these positions pictorially.
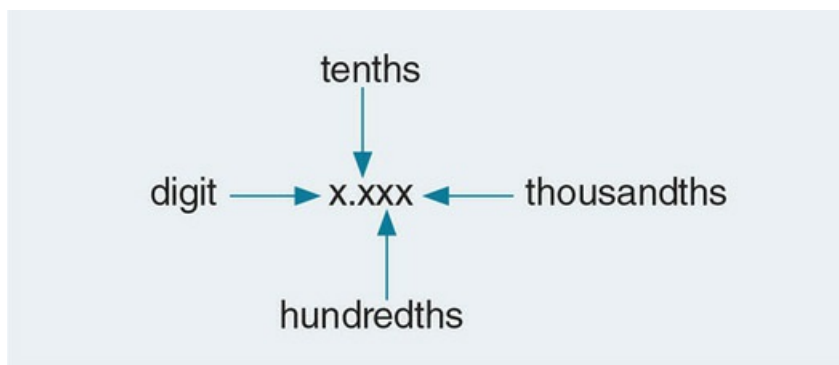
Examples:

.506 rounded to two decimal places = .51 .632 rounded to two decimal places = .63

$$.50 + .70 = 1.20 \quad (.42)(.80) = .336 \approx .34 \quad \sqrt{14} = 3.742 \approx 3.74 \quad \left(\frac{12}{9}\right)5 = (1.33)5 = 6.65$$

Try the following as practice.

1. .50 + .55 =

2. $\dfrac{39}{5}$ =

3. 2.23 − .34 =

4. $\sqrt{12}$

5. 1 − .66 =

6. (.20)(.80) =

7. $\dfrac{1.90}{1.20}$ =

8. $-3\left(\dfrac{19}{8}\right)+10$ =

9. Round this number to two decimal places: .605

10. Round this number to two decimal places: .098

If these operations all looked familiar to you and you were able to do them with little or no difficulty, then you are ready for the course! If you struggled with them, you should speak with your course instructor regarding recommendations and options.

Answers to Appendix A Problems

Section 1

1. 3
2. 7
3. 5
4. 3

Section 2

1. 12
2. 72
3. 24
4. 60

Section 3

1. 1
2. 27
3. 3
4. 26
5. 14
6. 13
7. 25

Section 4

1. 4
2. 1
3. 24
4. 4
5. 71

Section 5

1. −5
2. −7
3. −2
4. −5
5. 4
6. −4
7. 20
8. −3

Section 6

1. 1.05
2. 7.80
3. 1.89
4. 3.46
5. .34
6. .16
7. 1.58
8. 2.88 or 2.86
9. .61
10. .10

Appendix B. Standard Normal (z) Distribution

Area Between the Mean and z

z: tenths and hundredths →	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
z: digits and tenths ↓										
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.5	.4998									
4.0	.4999									

Appendix C. *t* Distribution

df	Level of Significance for One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Appendix D. Chi-Square (χ^2) Distribution

df	.10	.05	.02	.01	.001
			α		
1	2.706	3.841	5.412	6.635	10.827
2	4.605	5.991	7.824	9.210	13.815
3	6.251	7.815	9.837	11.341	16.268
4	7.779	9.488	11.668	13.277	18.465
5	9.236	11.070	13.388	15.086	20.517
6	10.645	12.592	15.033	16.812	22.457
7	12.017	14.067	16.622	18.475	24.322
8	13.362	15.507	18.168	20.090	26.125
9	14.684	16.919	19.679	21.666	27.877
10	15.987	18.307	21.161	23.209	29.588
11	17.275	19.675	22.618	24.725	31.264
12	18.549	21.026	24.054	26.217	32.909
13	19.812	22.362	25.472	27.688	34.528
14	21.064	23.685	26.873	29.141	36.123
15	22.307	24.996	28.259	30.578	37.697
16	23.542	26.296	29.633	32.000	39.252
17	24.769	27.587	30.995	33.409	40.790
18	25.989	28.869	32.346	34.805	42.312
19	27.204	30.144	33.687	36.191	43.820
20	28.412	31.410	35.020	37.566	45.315
21	29.615	32.671	36.343	38.932	46.797
22	30.813	33.924	37.659	40.289	48.268
23	32.007	35.172	38.968	41.638	49.728
24	33.196	36.415	40.270	42.980	51.179
25	34.382	37.652	41.566	44.314	52.620
26	35.563	38.885	42.856	45.642	54.052
27	36.741	40.113	44.140	46.963	55.476
28	37.916	41.337	45.419	48.278	56.893
29	39.087	42.557	46.693	49.588	58.302
30	40.256	43.773	47.962	50.892	59.703

Appendix E. *F* Distribution

						$\alpha = .05$				
						df_B				
df_W	*1*	*2*	*3*	*4*	*5*	*6*	*8*	*12*	*24*	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Source: R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research,* 6th ed. Copyright © R. A. Fisher and F. Yates 1963. Reprinted by permission of Pearson Education Limited.

					$\alpha = .01$					
					df_b					
df_w	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.23	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

Glossary

Alpha level:

The opposite of the confidence level; that is, the probability that a confidence interval does not contain the true population parameter. Symbolized α.

Alternative hypothesis:

In an inferential test, the hypothesis predicting that there is a relationship between the independent and dependent variables. Symbolized H_1. Sometimes referred to as a research hypothesis.

Analysis of variance (ANOVA):

The analytic technique appropriate when an independent variable is categorical with three or more classes and a dependent variable is continuous.

Beta weight:

A standardized slope coefficient that ranges from −1.00 to +1.00 and can be interpreted similarly to a correlation so that the magnitude of an independent variable–dependent variable relationship can be assessed.

Between-group variance:

The extent to which groups or classes in a sample are similar to or different from one another. This is a measure of true group effect, or a relationship between the independent and dependent variables.

Binomial coefficient:

The formula used to calculate the probability for each possible outcome of a trial and to create the binomial probability distribution.

Binomial probability distribution:

A numerical or graphical display showing the probability associated with each possible outcome of a trial.

Binomial:

A trial with exactly two possible outcomes. Also called a dichotomous or binary variable empirical outcome.

Bivariate regression:

A regression analysis that uses one independent variable and one dependent variable.

Bivariate:

Analysis involving two variables. Usually, one is designated the independent variable and the other the dependent variable.

Bonferroni:

A widely used and relatively conservative post hoc test used in ANOVA when the null is rejected as a means of determining the number and location(s) of differences between groups.

Bounding rule:

The rule stating that all proportions range from 0.00 to 1.00.

Categorical variable:

A variable that classifies people or objects into groups. Two types: *nominal* and *ordinal*.

Cell:

The place in a table where a row and a column meet.

Central limit theorem:

The property of the sampling distribution that guarantees that this curve will be normally distributed when infinite samples of large size have been drawn.

χ^2 distribution:

The sampling or probability distribution for chi-square tests. This curve is nonnormal and contains only positive values. Its shape depends on the size of the crosstabs table.]

Chi-square test of independence:

The hypothesis-testing procedure appropriate when both the independent variable and the dependent variables are categorical.

Classes:

The categories or groups within a nominal or ordinal variable.

Combination:

The total number of ways that a success r can occur over N trials.

Confidence interval:

A range of values spanning a point estimate that is calculated so as to have a certain probability of containing the population parameter.

Constant:

A characteristic that describes people, objects, or places and takes on only one value in a sample or population.

Contingency table:

A table showing the overlap between two variables.

Continuous variable:

A variable that numerically measures the presence of a particular characteristic. Two types: *interval* and *ratio*.

Cramer's *V*:

A symmetric measure of association for χ^2 when the variables are nominal or when one is ordinal and the other is nominal. *V* ranges from 0.00 to 1.00 and indicates the strength of the relationship. Higher values represent stronger relationships. Identical to phi in 2 × 2 tables.

Critical value:

The value of z or t associated with a given alpha level. Symbolized z_α or t_α.

Cumulative:

A frequency, proportion, or percentage obtained by adding a given number to all numbers below it.

Dependent samples:

Pairs of samples in which the selection of people or objects into one sample directly affected, or was directly affected by, the selection of people or objects into the other sample. The most common types are matched pairs and repeated measures.

Dependent variable (DV):

The phenomenon that a researcher wishes to study, explain, or predict.

Descriptive research:

Studies done solely for the purpose of describing a particular phenomenon as it occurs in a sample.

Deviation score:

The distance between the mean of a data set and any given raw score in that set.

Dispersion:

The amount of spread or variability among the scores in a distribution.

Dummy variable:

A two-category variable with one class coded as 0 and the other coded as 1.

Empirical outcome:

A numerical result from a sample, such as a mean or frequency. Also called an observed outcome.

Empirical relationship:

The causality requirement holding that the independent and dependent variables possess an observed relationship with one another.

Empirical:

Having the qualities of being measurable, observable, or tangible. Empirical phenomena are detectable with senses such as sight, hearing, or touch.

Evaluation research:

Studies intended to assess the results of programs or interventions for purposes of discovering whether those programs or interventions appear to be effective.

470

Exhaustive:

A property of all levels of measurement whereby the categories or range within a variable capture all possible values.

Expected frequencies:

The theoretical results that would be seen if the null were true—that is, if the two variables were, in fact, unrelated. Symbolized f_e.

Exploratory research:

Studies that address issues that have not been examined much or at all in prior research and that therefore might lack firm theoretical and empirical grounding.

F distribution:

The sampling distribution for ANOVA. The distribution is bounded at zero on the left and extends to positive infinity; all values in the F distribution are positive.

F statistic:

The statistic used in ANOVA; a ratio of the amount of between-group variance present in a sample relative to the amount of within-group variance.

Factorial:

Symbolized $!$, the mathematical function whereby the first number in a sequence is multiplied successively by all numbers below it down to 1.00.

Failure:

Any outcome other than success or the event of interest.

Familywise error:

The increase in the likelihood of a Type I error (i.e., erroneous rejection of a true null hypothesis) that results from running repeated statistical tests on a single sample.

Frequency:

A raw count of the number of times a particular characteristic appears in a data set.

Goodman and Kruskal's gamma:

A symmetric measure of association used when both variables are ordinal or one is ordinal and the other is dichotomous. Ranges from −1.00 to +1.00.

Hypothesis:

A single proposition, deduced from a theory, that must hold true in order for the theory itself to be considered valid.

Independent samples:

Pairs of samples in which the selection of people or objects into one sample in no way affected, or was

affected by, the selection of people or objects into the other sample.

Independent variable (IV):

A factor or characteristic that is used to try to explain or predict a dependent variable.

Inferential analysis:

The process of generalizing from a sample to a population; the use of a sample statistic to estimate a population parameter. Also called hypothesis testing.

Inferential statistics:

The field of statistics in which a descriptive statistic derived from a sample is employed probabilistically to make a generalization or inference about the population from which the sample was drawn.

Intercept:

The point at which the regression line crosses the y-axis; also the value of y when $x = 0$.

Interval variable:

A quantitative variable that numerically measures the extent to which a particular characteristic is present or absent and does not have a true zero point.

Kendall's tau$_b$:

A symmetric measure of association for two ordinal variables when the number of rows and columns in the crosstabs table are equal. Ranges from −1.00 to +1.00.

Kendall's tau$_c$:

A symmetric measure of association for two ordinal variables when the number of rows and columns in the crosstabs table are unequal. Ranges from −1.00 to +1.00.

Kurtosis:

A measure of how much a distribution curve's width departs from normality.

Lambda:

An asymmetric measure of association for χ^2 when the variables are nominal. Lambda ranges from 0.00 to 1.00 and is a proportionate reduction in error measure.

Leptokurtosis:

A measure of how peaked or clustered a distribution is.

Level of confidence:

The probability that a confidence interval contains the population parameter. Commonly set at 95% or 99%.

Level of measurement:

A variable's specific type or classification. Four types: *nominal*, *ordinal*, *interval*, and *ratio*.

Linear relationship:

A relationship wherein the change in the dependent variable associated with a one-unit increase in the independent variable remains static or constant at all levels of the independent variable.

Longitudinal variables:

Variables measured repeatedly over time.

Matched-pairs design:

A research strategy where a second sample is created on the basis of each case's similarity to a case in an existing sample.

Mean:

The arithmetic average of a set of data.

Measures of association:

Procedures for determining the strength or magnitude of a relationship after a chi-square test has revealed a statistically significant association between two variables.

Measures of central tendency:

Descriptive statistics that offer information about where the scores in a particular data set tend to cluster. Examples include the mode, the median, and the mean.

Median:

The score that cuts a distribution exactly in half such that 50% of the scores are above that value and 50% are below it.

Methods:

The procedures used to gather and analyze scientific data.

Midpoint of the magnitudes:

The property of the mean that causes all deviation scores based on the mean to sum to zero.

Mode:

The most frequently occurring category or value in a set of scores.

Multiple regression:

A regression analysis that uses two or more independent variables and one dependent variable.

Mutually exclusive:

A property of all levels of measurement whereby there is no overlap between the categories within a variable.

Negative skew:

A clustering of scores in the right-hand side of a distribution with some relatively small scores that pull the tail toward the negative side of the number line.

Negative correlation:

When a one-unit increase in the independent variable is associated with a decrease in the dependent variable.

Nominal variable:

A classification that places people or objects into different groups according to a particular characteristic that cannot be ranked in terms of quantity.

Nonparametric statistics:

The class of statistical tests used when dependent variables are categorical and the sampling distribution cannot be assumed to approximate normality.

Nonspuriousness:

The causality requirement holding that the relationship between the independent variable and dependent variable not be the product of a third variable that has been erroneously omitted from the analysis.

Normal curve:

A distribution of raw scores from a sample or population that is symmetric and unimodal and has an area of 1.00. Normal curves are expressed in raw units and differ from one another in metrics, means, and standard deviations.

Normal distribution:

A set of scores that clusters in the center and tapers off to the left (negative) and right (positive) sides of the number line.

Null hypothesis:

In an inferential test, the hypothesis predicting that there is no relationship between the independent and dependent variables. Symbolized H_0.

Observed frequencies:

The empirical results seen in a contingency table derived from sample data. Symbolized f_o.

Obtained value:

The value of the test statistic arrived at using the mathematical formulas specific to a particular test. The obtained value is the final product of Step 4 of a hypothesis test.

Omega squared:

A measure of association used in ANOVA when the null has been rejected in order to assess the magnitude of the relationship between the independent and dependent variables. This measure shows the proportion of the total variability in the sample that is attributable to between-group differences.

Omitted variable bias:

An error that occurs as a result of unrecognized spuriousness and a failure to include important third

variables in an analysis, leading to incorrect conclusions about the relationship between the independent and dependent variables.

One-tailed tests:

Hypothesis tests in which the entire alpha is placed in either the upper (positive) or lower (negative) tail such that there is only one critical value of the test statistic. Also called directional tests.

Ordinal variable:

A classification that places people or objects into different groups according to a particular characteristic that can be ranked in terms of quantity.

Ordinary least squares (OLS) regression:

A common procedure for estimating regression equations that minimizes the errors in predicting the dependent variable.

***p* value:**

In SPSS output, the probability associated with the obtained value of the test statistic. When $p < \alpha$, the null hypothesis is rejected.

Parameter:

A number that describes a population from which samples might be drawn.

Parametric statistics:

The class of statistical tests used when dependent variables are continuous and normally distributed and the sampling distribution can be assumed to approximate normality.

Partial slope coefficient:

A slope coefficient that measures the individual impact of an independent variable on a dependent variable while holding other independent variables constant.

Pearson's correlation:

The bivariate statistical analysis used when both independent and dependent variables are continuous.

Percentage:

A standardized form of a frequency that ranges from 0.00 to 100.00.

Phi:

A symmetric measure of association for chi-square with nominal variables and a 2 × 2 table. Identical to Cramer's *V*.

Platykurtosis:

A measure of how flat or spread out a distribution is.

Point estimate:

A sample statistic, such as a mean or proportion.

Pooled variances:

The type of *t* test appropriate when the samples are independent and the population variances are equal.

Population distribution:

An empirical distribution made of raw scores from a population.

Population:

The universe of people, objects, or locations that researchers wish to study. These groups are often very large.

Positive correlation:

When a one-unit increase in the independent variable produces an increase in the dependent variable.

Positive skew:

A clustering of scores in the left-hand side of a distribution with some relatively large scores that pull the tail toward the positive side of the number line.

Post hoc tests:

Analyses conducted when the null is rejected in ANOVA in order to determine the number and location of differences between groups.

Probability distribution:

A table or graph showing the entire set of probabilities associated with every possible empirical outcome.

Probability sampling:

A sampling technique in which all people, objects, or areas in a population have a known chance of being selected into the sample.

Probability theory:

Logical premises that form a set of predictions about the likelihood of certain events or the empirical results that one would expect to see in an infinite set of trials.

Probability:

The likelihood that a certain event will occur.

Proportion:

A standardized form of a frequency that ranges from 0.00 to 1.00.

_r_ coefficient:

The test statistic in a correlation analysis.

Range:

A measure of dispersion for continuous variables that is calculated by subtracting the smallest score from the largest. Symbolized as *R*.

Ratio variable:

A quantitative variable that numerically measures the extent to which a particular characteristic is present or absent and has a true zero point.

Regression analysis:

A technique for modeling linear relationships between one or more independent variables and one dependent variable wherein each independent variable is evaluated on the basis of its ability to accurately predict the values of the dependent variable.

Repeated-measures design:

A research strategy used to measure the effectiveness of an intervention by comparing two sets of scores (pre and post) from the same sample.

Replication:

The repetition of a particular study that is conducted for purposes of determining whether the original study's results hold when new samples or measures are employed.

Representativeness:

How closely the characteristics of a sample match those of the population from which the sample was drawn.

Residual:

The difference between a predicted value and an empirical value on a dependent variable.

Restricted multiplication rule for independent events:

A rule of multiplication that allows the probability that two events will both occur to be calculated as the product of each event's probability of occurrence: That is, $p(A \text{ and } B) = p(A) \cdot p(B)$.

Rule of the complement:

Based on the bounding rule, the rule stating that the proportion of cases that are not in a certain category can be found by subtracting the proportion that are in that category from 1.00.

Sample distribution:

An empirical distribution made of raw scores from a sample.

Sample:

A subset pulled from a population with the goal of ultimately using the people, objects, or places in the sample as a way to generalize to the population.

Sampling distribution:

A theoretical distribution made out of an infinite number of sample statistics.

Sampling error:

The uncertainty introduced into a sample statistic by the fact that any given sample is only one of many

samples that could have been drawn from that population.

Science:

The process of gathering and analyzing data in a systematic and controlled way using procedures that are generally accepted by others in the discipline.

Separate variances:

The type of *t* test appropriate when the samples are independent and the population variances are unequal.

Slope:

The steepness of the regression line and a measure of the change in the dependent variable produced by a one-unit increase in an independent variable.

Somers' *d*:

An asymmetric measure of association for χ^2 when the variables are nominal. Somers' *d* ranges from –1.00 to +1.00.

Standard deviation:

Computed as the square root of the variance, a measure of dispersion that is the mean of the deviation scores. Notated as *s* or *sd*.

Standard error:

The standard deviation of the sampling distribution.

Standard normal curve:

A distribution of *z* scores. The curve is symmetric and unimodal and has a mean of zero, a standard deviation of 1.00, and an area of 1.00.

Statistic:

A number that describes a sample that has been drawn from a larger population.

Statistical dependence:

The condition in which two variables are related to one another; that is, knowing what class persons/objects fall into on the independent variable helps predict which class they will fall into on the dependent variable.

Statistical independence:

The condition in which two variables are not related to one another; that is, knowing what class persons or objects fall into on the independent variable does not help predict which class they will fall into on the dependent variable.

Statistical significance:

When the obtained value of a test statistic exceeds the critical value and the null is rejected.

Success:

   The outcome of interest in a trial.

t **distribution:**

   A family of curves whose shapes are determined by the size of the sample. All *t* curves are unimodal and symmetric and have an area of 1.00.

t **test:**

   The test used with a two-class, categorical independent variable and a continuous dependent variable.

Temporal ordering:

   The causality requirement holding that an independent variable must precede a dependent variable.

Theoretical prediction:

   A prediction, grounded in logic, about whether or not a certain event will occur.

Theory:

   A set of proposed and testable explanations about reality that are bound together by logic and evidence.

Trends:

   Patterns that indicate whether something is increasing, decreasing, or staying the same over time.

Trial:

   An act that has several different possible outcomes.

Tukey's honest significant difference:

   A widely used post hoc test used in ANOVA when the null is rejected as a means of determining the number and location(s) of differences between groups.

Two-tailed test:

   A statistical test in which alpha is split in half and placed into both tails of the *z* or *t* distribution.

Type I error:

   The erroneous rejection of a true null hypothesis. α

Type II error:

   The erroneous retention of a false null hypothesis. β

Unit of analysis:

   The object or target of a research study.

Univariate:

   Involving one variable

Variable:

A characteristic that describes people, objects, or places and takes on multiple values in a sample or population.

Variance:

A measure of dispersion calculated as the mean of the squared deviation scores. Notated as s^2.

Variation ratio:

A measure of dispersion for variables of any level of measurement that is calculated as the proportion of cases located outside the modal category. Symbolized as *VR*.

Within-group variance:

The amount of diversity that exists among the people or objects in a single group or class. This is a measure of random fluctuation, or error.

***z* score:**

A standardized version of a raw score that shows that offers two pieces of information about the raw score: (1) how close it is to the distribution mean and (2) whether it is greater than or less than the mean.

***z* table:**

A table containing a list of *z* scores and the area of the curve that is between the distribution mean and each individual *z* score.

Answers to Learning Checks

Chapter 1

1.1.

1. sample
2. sample
3. population
4. sample

1.2.

1. descriptive
2. evaluation
3. theory testing
4. exploratory

Chapter 1

Chapter 2

2.1. The variable *offense type* could be coded as either a nominal or an ordinal variable because it could be either purely descriptive or it could represent a ranking of severity. An example of a nominal coding scheme is "violent offense, weapons offense, sex offense, property offense." This coding approach does not lend itself to clear ranking in terms of the severity of the crime. An example of ordinal coding is "property offense, violent offense (non-lethal), homicide." This approach allows for a moderate comparison of crime severity across the different categories.

2.2. Zip codes are nominal variables. They are made up of numbers, but these numbers are not meaningful in a statistical sense. Zip codes cannot be added, subtracted, multiplied, or divided. They are merely placeholders designating particular locations. They offer no information about ranking or quantification.

Chapter 3

3.1. A sum greater than 1.00 (for proportions) or 100.00 (for percentages) would suggest that one or more cases in the table or data set got counted twice or that there is an error in the count of the total cases. If the result is less than 1.00 or 100.00, then the opposite has occurred; that is, one or more cases in the sample have been undercounted or the total has been overestimated.

3.2. *State* is a nominal variable, because it comprises categories with no rank order. *Property crime rate* is ratio level, because it is continuous and has a true zero point.

3.3. Rates measure the prevalence of a certain event or characteristic in a population or sample. Percentages break samples or populations down into elements that either do or do not possess a certain characteristic. The calculation of rates requires information about the total population, something not needed for percentages. Rates do not sum to 100 because they are a ratio-level variable that can range from 0 to infinity (theoretically); that is, there is no maximum value that a rate cannot exceed. Percentages are confined to the range 0 to 100.

3.4. In [Table 3.8](#), column percentages would show the percentage of jails that do (and do not) offer GED courses that are small, medium, and large. The column percentages for the jails offering courses would be small = (155/988)100 = 15.69; medium = (315/988)100 = 31.88; and large = (518/988)100 = 52.43. This column sums to 100.00. The column percentages for prisons not providing courses is small = (623/1,383)100 = 45.05; medium = (471/1,383)100 = 34.06; and large = (289/1,383)100 = 20.90. With rounding, this column sums to 100.01.

3.5. Pie charts can be composed only of percentages because the idea behind them is that the entire "pie" represents 100% of the total available options, categories, or cases. The total is then partitioned into its constituent parts, with the "slices" representing percentages. This would not work with rates, because rates are independent numbers that have no maximum values and do not sum to 100% when combined with other rates.

3.6. The histogram in [Figure 3.8](#) and the frequency polygon in [Figure 3.10](#) probably would not benefit substantially from grouping. Their ungrouped distributions display clear shapes; they are distinctly different from the flat, shapeless histogram in [Figure 3.14](#). Grouping might help if a researcher was short on space and wanted to reduce the size of the chart, but it would not meaningfully enhance the interpretability of the data.

3.7. Rates are advantageous compared to raw counts because rates are standardized according to the size of the population. For instance, a police department with 25 officers serving a town of 1,000 residents is much different than a department of 25 officers in a city of 100,000 residents. Without population information, it is impossible to tell whether a 25-officer department is an appropriate size or is far too small. Rates are more informative and useful than raw counts are in this type of situation.

Chapter 4

4.1. The positively skewed distribution should look similar to Figure 4.2, and the negatively skewed one should look like Figure 4.3.

4.2. The modal category is "large" because there are more jails in this group (807) than in either of the other two.

4.3. The modal number of assaults is 1, since the bar associated with this number is higher than any other. The frequency is 136.

4.4. Adding Maryland at 8.59 makes the rank order 1.33^1 1.88^2 1.93^3 4.40^4 5.63^5 5.79^6 6.08^7 7.16^8 8.59^9. The MP =
$$\frac{9+1}{2} = 5$$
. The median is therefore 5.63, since this is the number in position 5 in the ordered list. Adding Maryland moved the median up slightly (from 5.02).

4.5. First, rearrange the categories so that they are in order from *Never* to *Every day or Almost every day*. Next, sum the frequencies until the cumulative sum is equal to or greater than 772.5. Since 457 + 37 + 82 + 217 = 793, we reach the same conclusion as before: *A few days a week* is the modal driving frequency.

4.6. Formula 4(2) would be used to calculate the mean for Table 4.3. The mean property crime rate among this sample of cities is
$$\bar{x} = \frac{130.37 + 297.24 + 38.50 + 325.03 + 239.09 + 149.67}{6} = \frac{1179.90}{6} = 196.65$$

4.7. Formula 4(3) would be used to calculate the mean for Table 2.9. This is because the numbers are arranged so that the left-hand column (*Number Executed*) contains the values in the data set and the right-hand column (*Frequency*) displays the number of times each value occurs. The mean formula with the *f* in the numerator (i.e., Formula 4[3]) is therefore the correct one. Using this formula, the *fx* column and sum would be: 0 + 3 + 4 + 3 + 6 + 7 + 16 = 39. Divided by the sample size (*N* = 50), the mean is .78. Across all 50 states, the mean number of executions per state in 2013 was .78. Since Table 2.9 includes all 50 states, including the 14 that do not authorize the death penalty, there are several more zeroes (14, to be exact!) than there would be if only those 36 states that authorize capital punishment were included in the calculations. These zeros pull the mean down. Similarly, the mean calculated on the data in Table 4.8, which contains only the 36 authorizing states, would be lower if the 14 non-authorizing states were added because there would be more zeros in the numerator and a larger denominator. In Table 2.9, dropping the 14 no-death penalty states from the analysis causes the mean to increase to 39/36 = 1.08. In Table 4.8, adding 14 states with zero executions changes the calculation to 81/50 = 1.62. These numerical outcomes confirm the conceptual predictions.

4.8. The mean cannot be calculated on ordinal data because ordinal variables are made of categories rather than numbers. There is no way to add *A few days a week* to *A few days a month*. The arithmetic required to calculate the mean cannot be done on a categorical variable.

4.9. Subtracting 196.65 from each of the raw scores in Table 4.3 produces the deviation scores −66.28,

100.59, −158.15, 128.38, 42.44, and −46.98. The deviation scores sum to 0.00.

4.10. The histogram in [Figure 4.1](#) is fairly normal, so the mean and the median will be close to one another. There is a slight positive skew, so the mean is probably a bit higher than the median.

Chapter 5

5.1. 45% is male; 4% were not convicted; and 10% received fines.

5.2. This distribution would be leptokurtic because the majority of the values cluster within a small range and only a few values are outside of that range. Thus, the distribution's shape or curve would be tall and thin.

5.3. The x-axis will contain the numbers 0, 0, 1, 1, 2, 3, 10, 19 in that order (or 0 to 19), with a vertical bar at 4.50 representing the mean. Points will be plotted above each number representing its distance from the mean and whether it is above the mean or below it.

5.4. The sample with 10 cases would have a mean of 100/10 = 10, whereas the sample with 50 cases would have a mean of 100/50 = 2. The first sample's mean is much larger than the second sample's mean because in it, the sum of 100 is spread across just 10 cases; in the second sample, by contrast, the 100 is spread across 50 cases. Thus, two samples can have the same sum but very different means. This is why sample size matters.

5.5. The reason is mathematical: Since variances and squared deviation scores are calculated by squaring numbers, and since any number squared is positive (even if the original number was negative), variances and squared deviations are always positive.

5.6. The mean salary is 732,644.99/8 = 91,580.62. The deviation scores are –4,007.62; 47.98; –25,069.56; 5,919.38; 6,419.38; 5,279.71; 5,991.38; and 5,419.38. Squaring and summing the deviation scores produces 813,935,309.50, and dividing by $n-1$ (i.e., 8 – 1 = 7) gives us the variance 813,935,309.50/7 = 116,276,472.78. The standard deviation is the square root of the variance and is 10,783.16.

5.7. The upper and lower limits, in the order they appear in the table, are.85 – 1.23; 58.03 – 83.73; 2.15 – 9.05; and 1.38 – 3.62.

Chapter 6

6.1.

1. The probability of the die landing on 3 is 1/6 = .17. This probability remains the same no matter what the original prediction was.

2. The probability of drawing the ace of hearts is 1/52 = .02. Again, the probability is not dependent on the original prediction.

3. The probability of the die landing on any value *except* 1 is 5/6 = .83. This calculation would apply to any initially predicted outcome.

4. The probability that the drawn card is anything *except* the nine of spades is 51/52 = .98. Again, this would apply to any predicted outcome.

6.2. Homicide = 100.00 − 61.50 = 38.50, so there is a .62 probability of clearance and a .39 probability that no arrest will be made. For rape, 62.20% (or .62) are not cleared. For robbery, the rate of non-clearance is 70.70% (.71). Non-clearance is 46.00% (.46) for aggravated assault, 78.1% (.78) for larceny-theft, and 86.9% (.87) for motor vehicle theft.

6.3. Answers will vary.

6.4. It is in Hospital B that the sex ratio of babies born in a single day would be expected to roughly mirror that in the population. This is because 20 is a larger sample size than 6 and, therefore, is less likely to produce anomalous or atypical results. In Hospital A, it would be relatively easy to have a day wherein 4 (67%), 5 (83%), or even all 6 (100%) babies were of the same gender. In Hospital B, however, it would be highly unlikely to see a result such as 16 of 20 babies being the same sex. Over the course of a year, though, it would be expected that births in both hospitals would be approximately 50% female and 50% male. This is because the sample size in both hospitals would be large enough that variations seen on a day-to-day basis (even fairly large variations) would even out over time.

6.5. N/A

6.6. N/A

6.7. With a clearance probability of p = .62, q = 1.00 − .62 = .38, and N = 6, the resulting probabilities are: p (0) = .003; p (1) = .04; p (2) = .11; p (3) = .24; p (4) = .32; p (5) = .21; and p (6) = .06.

6.8. The four types of deviation from normality include two types of skew (positive and negative) and two types of kurtosis (leptokurtosis and platykurtosis).

6.9. Conceptually, a standard deviation is the mean of the deviation scores; in other words, it is the average distance between the raw data points and the distribution mean. The mean offers information about the location of the center of the distribution, and the standard deviation describes the amount of spread or variability in the scores. Together, the mean and standard deviation offer a more comprehensive picture of the shape of the distribution than either one of these numbers would provide by itself.

6.10.

1. area between = .4162; area beyond = .50 − .4162 = .0838

2. area between = .2422; area beyond = .2578

3. area between = .4931; area beyond = .0069

4. area between = .4990; area beyond = .001

6.11. In the chapter, it was discovered that z scores greater than 2.33 are in the upper 1% of the distribution. The corresponding z score for the bottom 1% is −2.33, and we can say that scores less than −2.33 are in the bottom 1% of the distribution. We know this because the z distribution is symmetric. Whatever numbers and areas are on one side will appear identically (with the opposite sign) on the other side.

6.12.
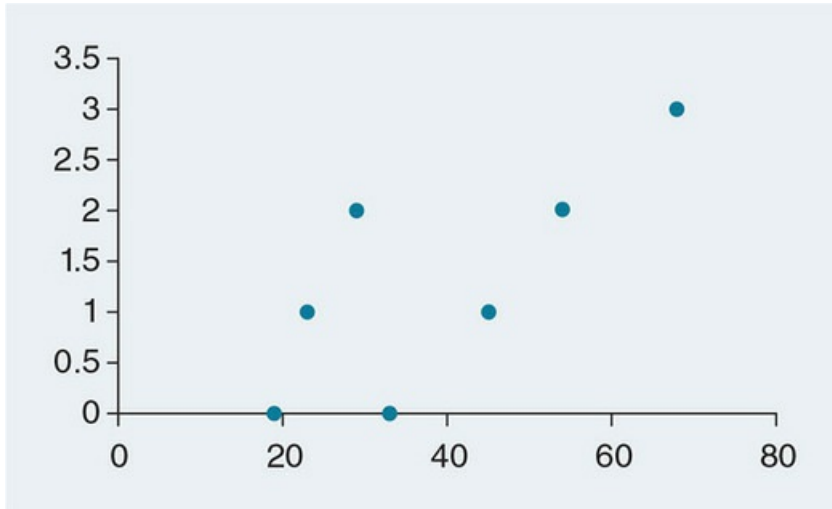
Chapter 7

7.1.

1. The sample mean is noticeably far from the population mean (15-unit difference), so the sample is not representative.
2. The sample proportion is fairly close to the population proportion (.05 units away), suggesting reasonable representativeness.
3. Both sample means are less than their respective population means. While the difference for violent crime is not huge (13 units), that for property crime is substantial (587 units). Therefore, this sample appears to not be representative of the population.

7.2. The sampling distribution takes on a normal shape because most sample statistics cluster around the population parameter. For example, if a mean in a population is 14, and someone pulls sample after sample after sample, computing and plotting each one's mean, then the majority (2/3, in fact) of the sample means will be 14 or close to it. Values significantly far from 14 will occur but infrequently. This means the distribution will peak at 14 and curve downward in the positive and negative directions.

Chapter 8

8.1. The 99% confidence interval (with $z_\alpha = \pm2.58$) will produce a wider confidence interval than the 95% level ($z_\alpha = \pm1.96$). Mathematically, this is because the 99% level's z score is larger and will therefore create a larger interval. Conceptually, this is an illustration of the trade-off between confidence and precision: A higher confidence level means a wider (i.e., less precise) interval.

8.2.

1. $t = \pm2.228$
2. $t = \pm1.725$
3. $t = \pm2.660$

8.3. The reason that critical values of t change depending on sample size is that t is not a single, static curve like z is; instead, t is a family of curves that are taller and more normal shaped at larger sample sizes and smaller and flatter at smaller sample sizes.

8.4. Two options to shrink the size of a confidence interval in order to increase the precision of the estimate are to (1) reduce the confidence level or alpha and (2) increase the sample size.

8.5. The final result of the calculation is $.91 \pm.06$, so the 95% CI : $.85 \le P \le .97$. The interval shrank (became more precise) when the confidence level was reduced. This exemplifies the trade-off between confidence and precision. The probability that the interval is correct fell (from 99% to 95%), but the estimate is more precise.

Part III: Hypothesis Testing

1.
    1. ratio
    2. ordinal
    3. nominal
    4. ordinal
    5. nominal
    6. interval
    7. Answers will vary

Chapter 9

9.1. The probability of a coin flip resulting in heads is .50. Results of flips will vary. The distribution of flips should be closer to a 50/50 split in the 10-flip exercise than in the 6-flip exercise.

9.2. Agency type is the IV and officer-to-resident ratio is the DV. This is because we are predicting that an agency's jurisdiction characteristics (population and land area) will impact the size of its workforce.

Chapter 10

10.1. Expected-frequency final answers should match [Table 10.8](#).

10.2. The expected frequencies, from cell A to cell D, are 31.18, 17.82, 24.82, and 14.18. The obtained value of chi-square is.04 + .08 + .06 + .10 = .28. The obtained value is smaller than the critical value (3.841), so the null is retained. This is the opposite of the conclusion reached in the chapter, where the obtained value exceeded the critical value and the null was rejected. This illustrates the sensitivity of chi-square tests to sample size.

Chapter 11

11.1. Using the sequence 3 + 4/2 results in 5 as an answer, while (3 + 4)/2 is 3.5. This is because the calculator performed the division first when parentheses were not used, but did the addition first and the division second when they were used. Similarly, $-3^2 = -9$, while $(-3)^2 = 9$. Without parentheses, the negative number remains negative; with parentheses, it becomes positive.

Chapter 12

12.1. The final answers to all the elements of the F_{obt} calculations can only be positive because SS_T and SS_B formulas require squaring (hence their names, sums of *squares*) . This eliminates negative signs and ensures that the final answers will be positive.

12.2. In the first example, it was determined that juvenile defendants' attorney types (public, private, or assigned) did not significantly influence the jail sentences that these defendants received on conviction. Since the null was retained, omega squared would not be computed, and post hoc tests would not be examined; all these posttest procedures are appropriate only when the null is rejected.

Chapter 13

13.1. The reason why there will always be correspondence between the sign of the r value and the sign of the t statistic lies in the sampling distribution for r. The negative values of r are located on the left side of the sampling distribution (i.e., the negative side), and the positive values are on the right side (i.e., the positive side). Therefore, r and t will always have the same sign.

13.2. Variables always correlate perfectly with themselves. You can test this by running a correlation between two variables with the same coding. For instance, if each variable is coded 1, 2, 3, 4, 5, the correlation between them will be 1.00. The same holds true if the coding is reversed to be 5, 4, 3, 2, 1. There is a one-to-one correspondence between a variable and itself.

Chapter 14

14.1. Scatterplot:
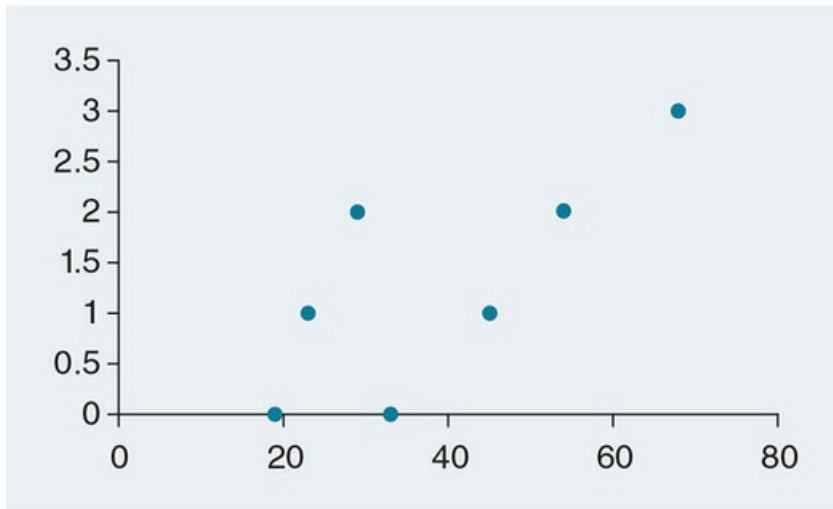
14.2. The numerators are identical. The difference is in the denominators. There are more terms in Formula 13(2) than in Formula 14(2), and information about y appears on the correlation formula but not the formula for the slope coefficient.

14.3. Using the formula $\hat{y} = 126.58 + 134.88x$, the predicted fines are as follows. For $x = 1$, $\hat{y} = 261.46$; for $x = 3$, $\hat{y} = 531.22$; for $x = 4$, $\hat{y} = 666.10$; and for $x = 5$, $\hat{y} = 800.98$. Residuals vary depending on each person's observed y value. For instance, the residual for the person with $x = 1$ and $y = 120$ is −141.46 (120 − 261.46) and that for the person who also had $x = 1$ but whose y value was 305 is 43.54 (305 − 261.46).The residuals could be reduced by adding variables measuring relevant legal characteristics of cases, such as the severity of the charge, whether a juvenile has a prior criminal record, and whether she or he was on probation at the time of the offense.

14.4. The reason is that adding more variables to the model changes the estimation of the relationship between each IV and the DV. For example, a relationship that appears strong at the bivariate level can become weaker when more IVs are added. Multiple regression, unlike bivariate regression, estimates each IV's influence on the DV while controlling for the impact of other variables, so each IV's slope coefficient changes when additional IVs are added to the model.

14.5.

1. $\hat{y} = 7.10$
2. $\hat{y} = 6.7$
3. $\hat{y} = 2.69$
4. $\hat{y} = 40.60$

14.6.

1. $\hat{y} = -1,101.82 + 155.17(3) - 448.26(0) + 187.74(1) + 85.47(16) - 757.08(1) + 188.58(0) = 161.87$
2. $\hat{y} = -1,101.82 + 155.17(2) - 448.26(1) + 187.74(0) + 85.47(17) - 757.08(0) + 188.58(1) = 401.83$
3. $\hat{y} = -1,101.82 + 155.17(1) - 448.26(0) + 187.74(0) + 85.47(15) - 757.08(0) + 188.58(0) = 335.40$
4. $\hat{y} = -1,101.82 + 155.17(1) - 448.26(1) + 187.74(1) + 85.47(15) - 757.08(0) + 188.58(1) = 263.46$

Answers to Review Problems

Chapter 1

1. Science is a systematic and controlled way of gathering information about the world. Methods are integral to science because scientific results are trustworthy only when the procedures used to reach them are considered correct by others in the scientific community. Scientific methods are logical, transparent, and replicable.

3. Samples are subsets of populations. Researchers draw samples because populations are too large to be studied directly. Samples are smaller and therefore more feasible to work with.

5. Hypothesis testing is used to test individual components of theories as a means of determining the validity of those theories. In this type of research, researchers use samples to draw inferences about populations. Evaluation research assesses the effectiveness of a program or intervention. People, places, or objects are measured before and after an intervention, or a control group is used for comparison to the treatment group. Exploratory research delves into new areas of study about which little is known. Hypotheses are generally not possible, because researchers usually do not have theory or prior evidence to guide them. Descriptive research analyzes a sample and provides basic information about those people, places, events, or objects. Usually, inferences are not made to the population and there is no effort to discern the possible impact one variable might have on another.

7. Any three programs or policies; the correctness of students' responses is up to the instructor's judgment. Correct answers will contain elements of random assignment to treatment and control groups, repeated measures, or matched pairs.

9. This would be descriptive research, because that type of research focuses solely on a sample rather than using a sample to draw conclusions about a population.

Chapter 2

1.
    1. education
    2. crime
    3. people

3.
    1. poverty (or income)
    2. violent crime
    3. neighborhoods

5.
    1. money spent on education, health, and welfare
    2. violent crime
    3. countries

7.
    1. police department location (urban or rural)
    2. entry-level pay
    3. police departments

9. The researcher has failed to consider additional variables. A statistical relationship between ice cream and crime is not proof that one causes the other; one or more variables have been erroneously omitted from the analysis. In the present case, the missing variable is probably ambient temperature—both ice cream sales and crime might be higher in warmer months, so the empirical association between them is in fact spurious.

11.
    1. nominal
    2. ratio
    3. nominal
    4. ratio
    5. ratio
    6. ordinal
    7. nominal

13.
    1. The first would produce a ratio variable, the second would create an ordinal variable, and the third would make a nominal variable.
    2. The phrasing that yields a ratio-level variable is best. Researchers who collect data should always use the highest level of measurement possible. Continuous variables can be made into categorical ones later on, but categorical data can never be made continuous.

15.
    1. nominal

2. ratio

17.

1. victim advocacy (presence or absence of witness office)
2. nominal
3. sentencing (months of incarceration imposed)
4. ratio (the sample includes all offenders in each court, not just those sentenced to prison, so there is a zero point)
5. courts

19.

1. homicide rate (homicides per population)
2. ratio
3. handgun ownership (own or not)
4. nominal
5. cities for homicide rates and people for gun ownership

Chapter 3

1.

Charge	f	cf	p	cp	pct	cpct
Violent Offense	13,938	13,938	.25	.25	24.93	24.93
Property Offense	16,241	30,179	.29	.54	29.05	53.98
Drug Offense	18,220	48,399	.33	.87	32.59	86.57
Public-Order Offense	7,504	55,903	.13	1.00	13.42	99.99
	N = 55,903		1.00		99.99	

1.

2. Since this variable is nominal, a pie chart or a bar graph would be appropriate. The pie chart requires percentages be used, whereas the bar graph can be percentages or frequencies. (Frequencies shown here.)

Arrest Charge Among Convicted Felons

3.

Arrest Charge Among Convicted Felons

3.
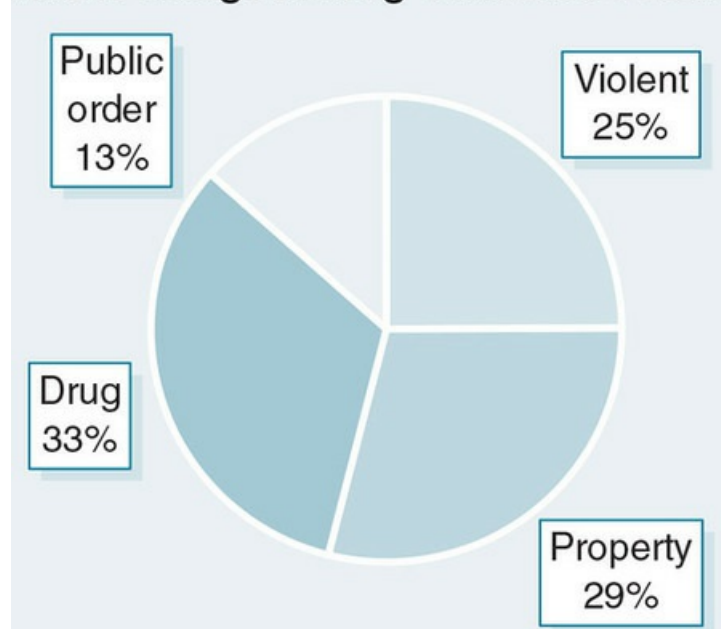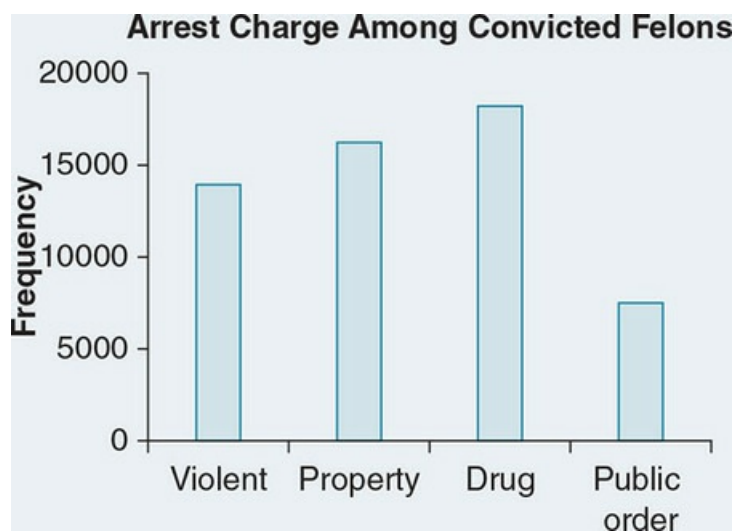
1. The data are ratio, so either a histogram or a frequency polygon would be the correct chart type.
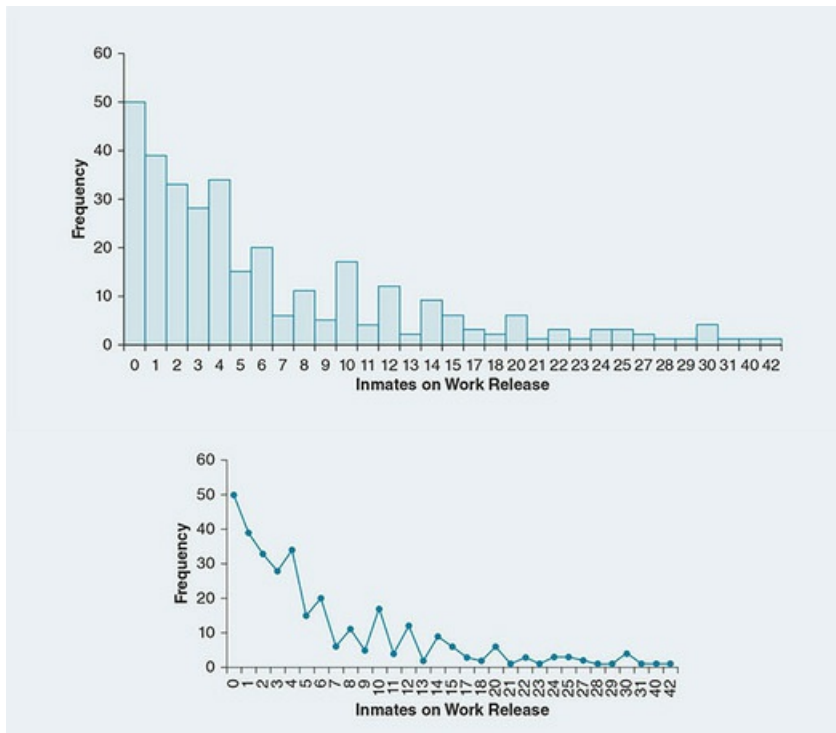
2. The range is 42 – 0 = 42. With 10 intervals, the width of each would be

$$width = \frac{42}{10} = 4.20$$

. Rounding to the nearest whole number, the width is 4.00.

Stated Class Limits	f
0–3	150
4–7	75
8–11	37
12–15	29
16–19	5
20–23	11
24–27	8
28–31	7
32–35	0
36–39	0
40–43	2
	$N = 324$

3. Grouped data are technically ordinal, but the underlying scale is continuous, and so the chart type would be a histogram or frequency polygon.
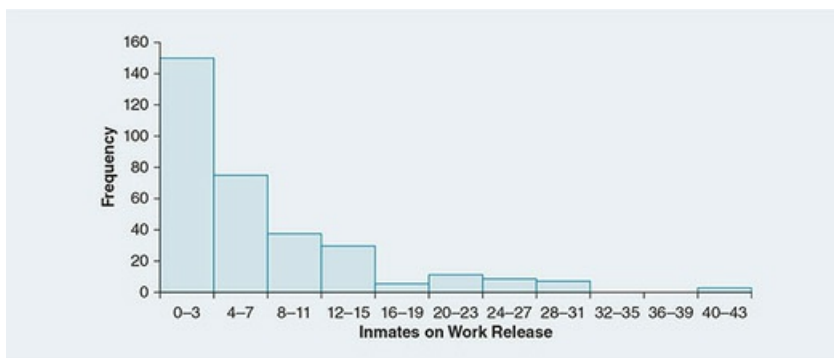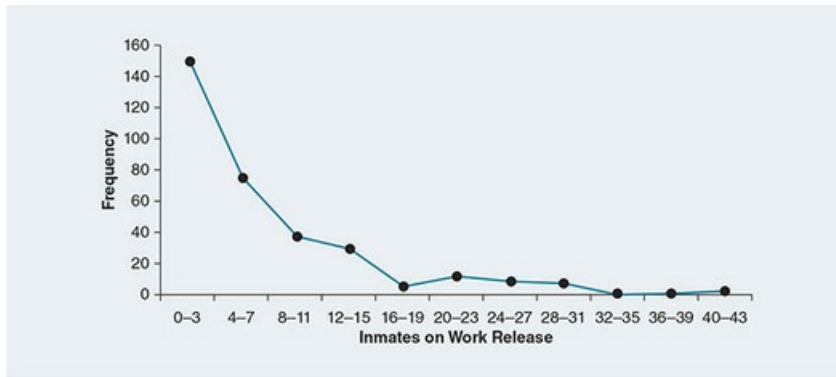
5. The line chart displays an upward trend, meaning that support has been increasing over time.
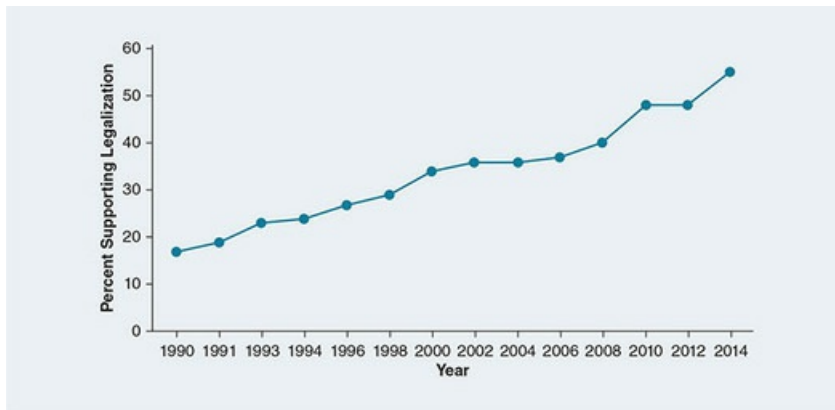
7.

City	Rate per 1,000
Birmingham, AL	17.46
Portland, ME	3.22
San Francisco, CA	7.77
Tampa, FL	6.31
Ann Arbor, MI	1.92
Washington, DC	11.20

1.

2. The variable *city* is nominal, narrowing the available options down to pie charts and bar graphs. Since rates cannot be used for pie charts, a bar graph is correct.
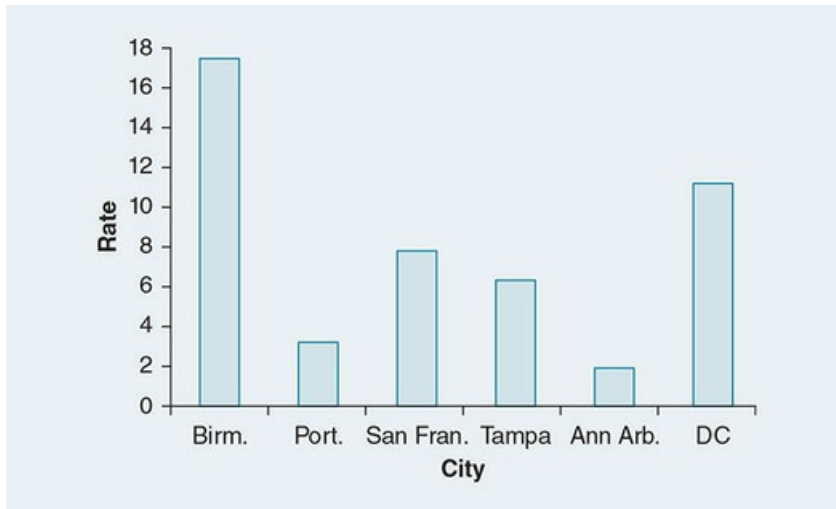
9.
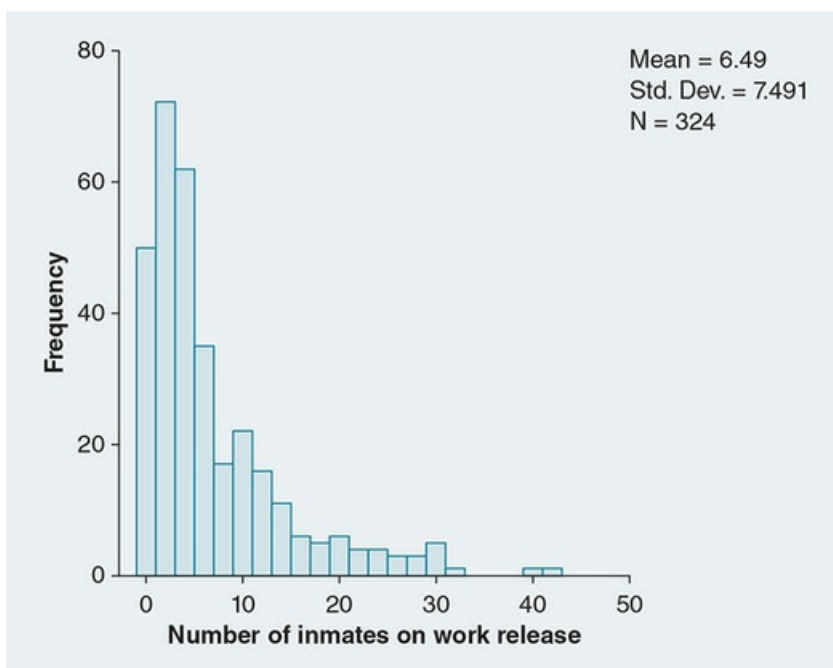
11. SPSS exercise. Using *Transform → Compute* and the command *sworn/population*1000* will produce the correct variable.

13. The variable is categorical, so a pie chart (percentages) or bar chart (percentages or frequencies) can be used.

How are courts at
dealing with criminals?

Legend:
- Not harsh enough
- About right
- Too harsh

How are courts at dealing with criminals?
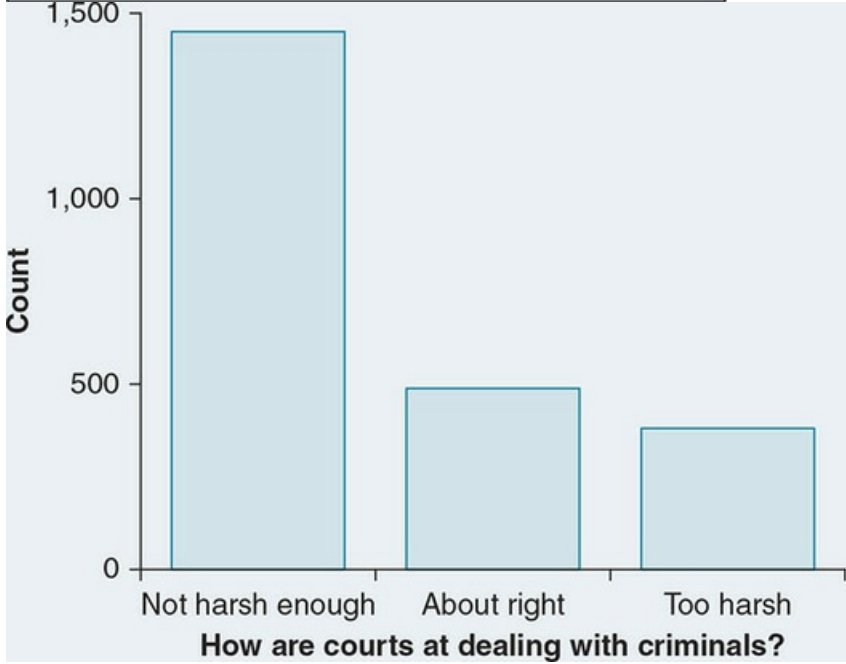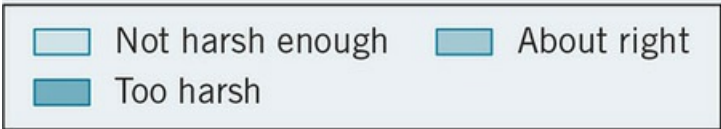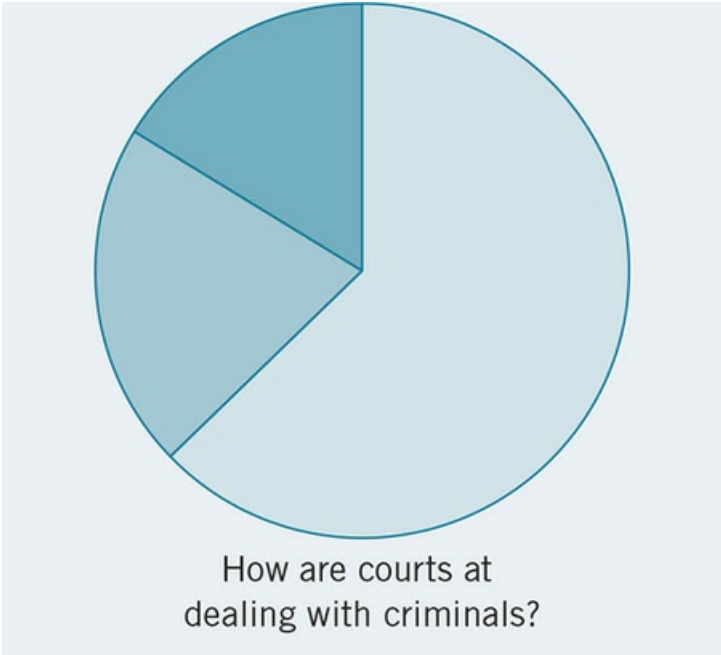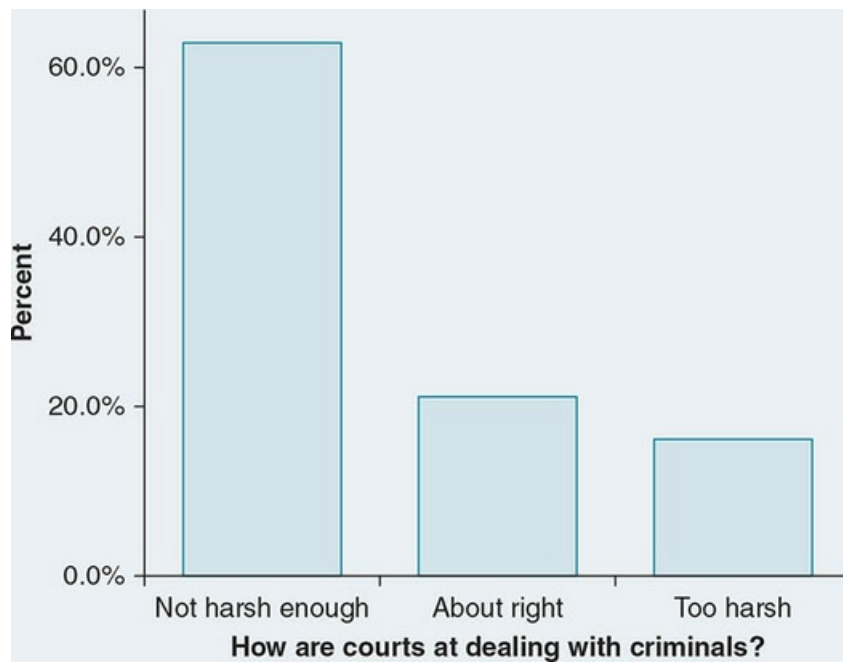
Chapter 4

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer.

1.
    1. ratio
    2. mode, median, mean
3.
    1. nominal
    2. mode
5. The mean is the midpoint of the magnitudes because it is the value that perfectly balances the deviation scores. Deviation scores are produced by subtracting the mean from each raw score; deviation scores measure the distance between raw scores and the mean. They always sum to zero.

7. c

9. a

7. c

11.
    1. nominal; mode
    2. mode = acquaintance
13.
    a.
        1. rank order: 170, 176, 211, 219, 220, 258, 317, 345
        2. $MP = (8 + 1)/2 = 4.5$
        3. $Md = (219 + 220)/2 = 439/2 = 219.50$
    b. mean = (317 + 170 + 211 + 258 + 219 + 345 + 176 + 220)/8 = 1916/8 = 239.50
15.
    1. mean = $((1 \cdot 15) + (2 \cdot 4) + (3 \cdot 2) + (4 \cdot 2) + (5 \cdot 15) + (7 \cdot 1) + (10 \cdot 11) + (15 \cdot 7) + (20 \cdot 1) + (24 \cdot 1) + (30 \cdot 1) + (45 \cdot 2) + (60 \cdot 2) + (120 \cdot 2) + (180 \cdot 1))/70$ = (15 + 8 + 6 + 8 + 75 + 7 + 8 + 110 + 105 + 20 + 24 + 90 + 90 + 120 + 240 + 180)/70 = 1,106/70 = 15.80.
    2. The mean (15.80) is substantially greater than the median (5.00), so this distribution is positively skewed.
17.
    1. $MP = (7 + 1)/2 = 4$, and $Md = 26$
    2. mean = (31 + 26 + 31 + 15 + 13 + 27 + 4)/7 = 147/7 = 21.00
    3. Alaska = 31 – 21.00 = 10.00; Arkansas = 26 – 21.00 = 5.00; Connecticut = 31 – 21.00 = 10.00; Kansas = 15 – 21.00 = –6.00; Montana = 13 – 21.00 = –8.00; South Carolina = 27 – 21.00 = 6.00; Vermont = 4 – 21.00 = –17.00; sum = 0
19. The mean = 19.87, the median = 10.00, and the mode = 10.

Chapter 5

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number.

1. Measures of central tendency offer information about the middle of the distribution (i.e., where scores tend to cluster), but they do not provide a picture of the amount of variability present in the data. Measures of dispersion show whether the data cluster around the mean or whether they are very spread out.

3. two-thirds

5. $VR = 1 - (807/2{,}371) = 1 - .34 = .66$

7. $VR = 1 - (6{,}036/6{,}956) = 1 - .87 = .13$

9.
1. $R = 190 - 48 = 142$
2. mean $= (84 + 50 + 81 + 122 + 190 + 48)/6 = 575/6 = 95.83$
3. variance $= 14{,}300.84/(6 - 1) = 2{,}860.17$
4. $sd = \sqrt{2{,}860.17} = 53.48$

11.
1. $R = 18.70 - 9.90 = 8.80$
2. mean $= (13.70 + 18.70 + 9.90 + 12.40 + 16.20 + 14.00 + 10.50 + 10.40 + 11.00)/9 = 116.80/9 = 12.98$
3. variance $= 71.21/(9 - 1) = 8.90$
4. $sd = \sqrt{8.90} = 2.98$

13.
1. $63.10 - 18.97 = 44.13$ and $63.10 + 18.97 = 82.07$
2. $1.75 - .35 = 1.40$ and $1.75 + .35 = 2.10$
3. $450.62 - 36.48 = 414.14$ and $450.62 + 36.48 = 487.10$

15. The standard deviation is the mean of the deviation scores. It represents the average (i.e., mean) distance between the mean and the individual raw scores in the data set.

17. range = 42; variance = 67.15; standard deviation = 8.19

19. range = 64585; variance = 12294723.87; standard deviation = 3506.38

Chapter 6

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number. Areas from the z table are reported using all four decimal places.

1.

   1. $N = 8$
   2. $r = 2$
   3. 28

3.

   1. $N = 7$
   2. $r = 3$
   3. 35

5.

   1. $N = 8$
   2. $r = 4$
   3. 70

7.

   1. standard normal
   2. standard normal
   3. binomial

9.

   1. With $p = .62$, $q = .38$, and $N = 4$: $p(0) = .02$; $p(1) = .12$; $p(2) = .132$; $p(3) = .36$; $p(4) = .15$
   2. $r = 3$, or 3 of the 4 charges would be for aggravated assault
   3. $r = 0$, or none of the 4 charges would be for aggravated assault
   4. $p(2) + p(1) + p(0) = .32 + .12 + .02 = .46$
   5. $p(3) + p(4) = .36 + .15 = .51$

11.

   1. With $p = .61$, $q = .39$, and $N = 5$: $p(0) = .01$; $p(1) = .07$; $p(2) = .22$; $p(3) = .35$; $p(4) = .27$; $p(5) = .08$
   2. $r = 3$, or 3 of the 5 murders committed with firearms
   3. $r = 0$, or none of the 5 murders committed with firearms
   4. $p(1) + p(0) = .07 + .01 = .08$
   5. $p(4) + p(5) = .27 + .08 = .35$

13.

   1. $z_{4.28} = (4.28 - 1.99)/.84 = 2.29/.84 = 2.73$
   2. area between the mean and $z = 2.73$ is .4968
   3. area in the tail beyond the mean and $z = 2.73$ is $.50 - .4968 = .0032$

15.

   1. $z_{1.29} = (1.29 - 1.99)/.84 = -.70/.84 = -.83$

2. area between the mean and $z = -.83$ is .2967

3. area in the tail beyond the mean and $z = -.83$ is $.50 - .2967 = .2033$

17. $.50 - .03 = .47$. The closest area on the table is .4699, which corresponds to a z score of 1.88. This is the upper tail, so $z = 1.88$.

19. $.50 - .10 = .40$. The closest area on the table is .3997, which corresponds to a z score of 1.28. This is the lower tail, so $z = -1.28$.

Chapter 7

1. a

3. b

5. sampling

7. b

9. z

11. a

13. c

15. b

Chapter 8

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. Calculation steps are identical to those in the text. For numbers close to zero, decimals are extended to the first nonzero number.

1. At least 100.

3. The z distribution.

5. The z distribution (or standard normal curve) is fixed; it cannot change shape to accommodate small samples. Small samples violate the assumption that the scores are perfectly normally distributed, so z cannot be used in these instances.

$$CI = 13.00 \pm 2.110\left(\frac{12.10}{\sqrt{18-1}}\right)$$

$$= 13.00 \pm 2.110\left(\frac{12.10}{4.12}\right)$$

$$= 13.00 \pm 2.110(2.94)$$

7. $df = 17$, $t = \pm 2.110$. $\quad = 13.00 \pm 6.20$

95% *CI*: $6.80 \le \mu \le 19.20$

There is a 95% chance that the interval 6.80 to 19.20, inclusive, contains the population mean.

$$CI = 1.14 \pm 1.96\left(\frac{.36}{\sqrt{2033-1}}\right)$$

$$= 1.14 \pm 1.96\left(\frac{.36}{45.08}\right)$$

$$= 1.14 \pm 1.96(.01)$$

9. $z = 1.96$. $\quad = 1.14 \pm .02$

95% *CI*: $1.12 \le \mu \le 1.16$

There is a 95% chance that the interval 1.11 to 1.17, inclusive, contains the population mean.

$$CI = 3.06 \pm 2.58\left(\frac{2.62}{\sqrt{707-1}}\right)$$

$$= 3.06 \pm 2.58\left(\frac{2.62}{26.57}\right)$$

$$= 3.06 \pm 2.58(.10)$$

11. $z = 2.58$ $\quad = 3.06 \pm .26$

99% *CI*: $2.80 \le \mu \le 3.32$

There is a 99% chance that the interval 2.80 to 3.32, inclusive, contains the population mean.

$$CI = 4.34 \pm 2.045\left(\frac{3.94}{\sqrt{30-1}}\right)$$

13. *df* = 29, *t* = ±2.045

$$= 4.34 \pm 2.045\left(\frac{3.94}{5.39}\right) = 4.34 \pm 2.045(.73) = 4.34 \pm 1.49$$

99% *CI*: $2.85 \le \mu \le 5.83$

There is a 95% chance that the interval 2.85 to 5.83, inclusive, contains the population mean.

$$CI = .34 \pm 1.96\sqrt{\frac{.34(1-.34)}{1281}}$$

$$= .34 \pm 1.96\sqrt{\frac{.22}{1281}}$$

$$= .34 \pm 1.96\sqrt{.0002}$$

$$= .34 \pm 1.96(.01)$$

15. $z = 1.96$ $\quad = .34 \pm .02$

95% *CI*: $.32 \le P \le .36$

There is a 95% chance that the interval .32 to .36, inclusive, contains the population proportion.

$$CI = .31 \pm 1.96\sqrt{\frac{.31(1-.31)}{1967}} = .31 \pm 1.96\sqrt{\frac{.21}{1967}}$$

$$= .31 \pm 1.96\sqrt{.0001}$$

$$= .31 \pm 1.96(.01)$$

17. $z = 1.96$ $\qquad = .31 \pm .02$

95% *CI*: $.29 \leq P \leq .33$

There is a 95% chance that the interval .29 to .33, inclusive, contains the population proportion.

$$CI = .28 \pm 1.96\sqrt{\frac{.28(1-.28)}{329}} = .28 \pm 1.96\sqrt{\frac{.20}{329}}$$

19. $z = 1.96$

$$= .28 \pm 1.96\sqrt{.001}$$

$= .28 \pm 1.96(.03)$

$= .28 \pm .06$

95% *CI*: $.22 \leq P \leq .34$

There is a 95% chance that the interval .22 to .34, inclusive, contains the population proportion.

Chapter 9

1. One possible explanation for the difference is sampling error. There might have been males with unusually long sentences or females with atypically short sentences; these extreme values might have pulled males' mean upward, females' mean downward, or both. The other possible explanation is that men truly are given longer sentences than women, on average. It could be that there is a real difference between the two population means.

3. The symbol is H_1, and it predicts that there is a difference between populations; in other words, it predicts that the observed differences between two or more samples' statistics reflects a genuine difference in the population.

5. A Type I error is the erroneous rejection of a true null (also called false positive). This occurs when a researcher concludes that two or more variables are related when, in fact, they are not.

7. Reducing the likelihood of one type of error increases the probability that the other one will occur. Preventing a Type I error requires increasing the amount of evidence needed to reject the null, which raises the chances that a false null will not be rejected as it should be (Type II error). Preventing a Type II error requires a reduction in the amount of evidence needed to reject the null, thus increasing the probability that a true null will be wrongly rejected (Type I error).

9.

> **Step 1: State the null (N_0) and alternative (N_1) hypotheses.** These are the competing predictions about whether or not there is a true difference between population values.
>
> **Step 2: Identify the distribution and calculate the degrees of freedom.** Each type of statistical test uses a particular sampling distribution, so the correct one (and correct table) must be identified at the outset. With the exception of the z distribution, sampling distributions are families of curves and the degrees of freedom determine the shape of the curve that will be used.
>
> **Step 3: Identify the critical value of the test statistic and state the decision rule.** The critical value is located by using the table associated with the selected distribution. The decision rule states what the obtained value (calculated in Step 4) must be in order for the null to be rejected.
>
> **Step 4: Calculate the obtained value of the test statistic.** This is the mathematical part of the test. Sample means, proportions, standard deviations, and sizes are entered into formulas, which are then solved to produce the obtained value.
>
> **Step 5: Make a decision about the null and state the substantive conclusion.** In this final step, the obtained value is evaluated according to the decision rule (Step 3). If the criteria are met, the null is rejected; if they are not, it is retained. The substantive conclusion is the interpretation of the statistical outcome in the context of the specific variables and samples being analyzed.

11.

1. True effect. Events with low probabilities are atypical. They are unlikely to occur by chance alone.
2. Reject

13. b (categorical IV with two classes and continuous DV)

15. d (continuous IV and continuous DV)

17. a (categorical IV and categorical DV)

Chapter 10

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number. Calculation steps are identical to those in the text; using alternative sequences of steps might result in answers different from those presented here. These differences might or might not alter the final decision regarding the null.

1. Yes. The IV and the DV are both categorical, so chi-square can be used.

3. No. The IV is categorical, but the DV is continuous, so chi-square cannot be used.

5.
    1. The IV is gender and the DV is sentence received.
    2. Both are nominal.
    3. Two rows and three columns

7.
    1. The IV is crime type and the DV is sentence length.
    2. The IV is nominal and the DV is ordinal.
    3. Three rows and three columns

9.

   Step 1: H_0: $\chi^2 = 0$; H_1: $\chi^2 > 0$

   Step 2: χ^2 distribution with $df = (2 - 1)(2 - 1) = 1$

   Step 3: $\chi^2_{crit} = 6.635$. Decision rule: *If* $\chi^2_{obt} > 6.635$, *the null will be rejected.*

   Step 4: Expected frequencies are 28.14 for cell A, 38.86 for B, 34.86 for C, and 48.14 for D. χ^2_{obt} = 23.76 + 17.21 + 19.18 + 13.89 = 74.04

   Step 5: The obtained value is greater than 6.635, so the null is rejected. There is a relationship between whether a jail offers alcohol treatment and whether it offers psychiatric counseling. Row percentages can be used to show that 80.6% of jails that offer alcohol treatment provide counseling, compared to only 10.8% of those that do not offer alcohol treatment. It appears that most jails supply either both of these services or neither of them; relatively few provide only one.

11.

   Step 1: H_0: $\chi^2 = 0$; H_1: $\chi^2 > 0$

   Step 2: χ^2 distribution with $df = (2 - 1)(2 - 1) = 1$

   Step 3: $\chi^2_{crit} = 3.841$. Decision rule: *If* $\chi^2_{obt} > 3.841$, *the null will be rejected.*

   Step 4: Expected frequencies are 314.73 for cell A, 1023.27 for B, 67.27 for C, and 218.73 for D. χ^2_{obt} = .09 + .03 + .41 + .13 = .66

   Step 5: The obtained value is less than 3.841, so the null is retained. There is no relationship between victims' gender and the likelihood that their injuries resulted from fights. Row percentages show that 23.9% of males' injuries occurred during fights, compared to 21.7% of females' injuries. The two percentages are substantively similar to one another, and the small difference between them appears to be a chance finding.

13.

Step 1: H_0: $\chi^2 = 0$; H_1: $\chi^2 > 0$

Step 2: χ^2 distribution with $df = (2 - 1)(2 - 1) = 1$

Step 3: $\chi^2_{crit} = 3.841$. Decision rule: *If $\chi^2_{obt} > 3.841$, the null will be rejected.*

Step 4: Expected frequencies are 46.51 for cell A, 27.49 for B, 85.49 for C, and 50.51 for D. $\chi^2_{obt} = .26 + .44 + .14 + .24 = 1.08$.

Step 5: The obtained value is less than 3.841, so the null is retained. Gender and support for marijuana legalization are statistically independent among black Americans. Looking at row percentages, 67.6% of men and 60.3% of women believe that marijuana should be made legal. There appears to be more support for legalization by men than by women, but this difference is not statistically significant (i.e., appears to be a chance finding).

15.

Step 1: H_0: $\chi^2 = 0$; H_1: $\chi^2 > 0$.

Step 2: χ^2 distribution with $df = (3 - 1)(3 - 1) = 4$

Step 3: $\chi^2_{crit} = 13.277$. Decision rule: *If $\chi^2_{obt} > 13.277$, the null will be rejected.*

Step 4: Expected frequencies are 315.90 for cell A, 42.69 for B, 13.42 for C, 200.41 for D, 27.08 for E, 8.51 for F, 260.70 for G, 35.23 for H, and 11.07 for I. $\chi^2_{obt} = .003 + .01 + .19 + .10 + .57 + .03 + .11 + .30 + .39 = 1.70$

Step 5: The obtained value is less than 13.277, so the null is retained. There is no relationship between annual income and the frequency of contact with police. Row percentages show that 85% of people in the lowest-income category, 83% of those in the middle-income category, and 87% of those in the highest-income group had between zero and two recent contacts. The vast majority of people have very few annual contacts with officers, irrespective of their income.

17.

1. The SPSS output shows $\chi^2_{obt} = 16.125$.
2. The null is rejected at an alpha of .05 because $p = .003$, and .003 is less than .05.
3. Race and attitudes about courts' harshness are statistically dependent.
4. Asking SPSS for row percentages shows the majority of people in all racial groups think courts are not harsh enough, but this percentage is higher among whites (63.0%) than blacks (55.6%) or members of other racial groups (61.8%). Likewise, blacks are more likely than the other two groups to say that the courts are overly harsh on offenders (25.9%). The applicable measures of association are Cramer's V, lambda, gamma, and tau$_c$. All of these values show a fairly weak relationship between these two variables. This makes sense, because people's attitudes about courts' approach to crime control are too complex to be determined by race alone.

19.

1. The SPSS output shows $\chi^2_{obt} = 25.759$.
2. The null is rejected at an alpha of .05 because $p = .003$, and .003 < .05.
3. There is a relationship between race and perceived stop legitimacy. Asking SPSS for row percentages shows that 84.4% of white drivers, 71.3% of black drivers, and 85.5% of drivers of

other races thought the stop was for a legitimate reason. Black drivers appear to stand out from nonblack drivers in that they are less likely to believe their stop was legitimate.

4. The null was rejected, so measures of association can be examined. Since both variables are nominal and there is a clear independent and dependent designation, Cramer's V and lambda are both available. The SPSS output shows that lambda = .000, meaning that the relationship between race and stop legitimacy, while statistically significant, is substantively meaningless. Cramer's V = .107, also signaling a very weak relationship. This makes sense looking at the percentages from Part C. A clear majority of all drivers believed their stop was for a legitimate reason. Black drivers deviated somewhat, but a large majority still endorsed stop legitimacy.

Chapter 11

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number. Calculation steps are identical to those in the text; using alternative sequences of steps might result in answers different from those presented here. These differences might or might not alter the final decision regarding the null.

1.
    1. whether the defendant plead guilty or went to trial
    2. nominal
    3. sentence
    4. ratio (there is no indication that the sample was narrowed only to those who were incarcerated, so theoretically, zeroes are possible)

3.
    1. judge gender
    2. nominal
    3. sentence severity
    4. ratio

5. a

7. a

8. *t; z*

11. b

13.

Step 1: H_0: $\mu_1 = \mu_2$; H_0: $\mu_1 \neq \mu_2$. (*Note:* No direction of the difference was specified, so the alternative is $\neq$.)

Step 2: *t* distribution with *df* = 155 + 463 – 2 = 616

Step 3: t_{crit} = ±1.960 (±1.980 would also be acceptable). The decision rule is that if t_{obt} is greater than 1.960 or less than –1.960, the null will be rejected.

$$\hat{\sigma}_{\bar{x}_1 + \bar{x}_2} = \sqrt{\frac{15.40 + 50.82}{616}} \sqrt{\frac{618}{71765}}$$

$$= \sqrt{.11}\sqrt{.01} = .33(.10) = .03, \text{ and}$$

$$t_{obt} = \frac{-.01}{.03} = -.33$$

Step 4:

Step 5: Since t_{obt} is not greater than 1.960 or less than –1.960, the null is retained. There is no difference between MHOs and SHOs in the diversity of the crimes they commit. In other words, there does not appear to be a relationship between offenders' status as MHOs or SHOs and the variability in their criminal activity.

15.

Step 1: $H_0: \mu_1 = \mu_2$; $H_1: \mu_1 < \mu_2$.

$$df = \frac{(187.77 + 29.17)^2}{35,257.57(.01) + 850.89(.001)} - 2$$

$$= \frac{47,062.96}{352.58 + .85} - 2$$

Step 2: t distribution with $\quad = 131.16$

Step 3: $t_{crit} = -1.658$ (-1.645 would also be acceptable). The decision rule is that if t_{obt} is less than -1.658, the null will be rejected.

Step 4: $\hat{\sigma}_{x_1 - x_2} = 14.73$ and $t_{obt} = -1.79$.

Step 5: Since t_{obt} is less than -1.658, the null is rejected. Juveniles younger than 16 at the time of arrest received significantly shorter mean jail sentences relative to juveniles who were older than 16 at arrest. In other words, there appears to be a relationship between age at arrest and sentence severity for juveniles transferred to adult court.

17.

Step 1: $H_0: \mu_1 = \mu_2$; $H_1: \mu_1 > \mu_2$.

Step 2: t distribution with $df = 160 + 181 - 2 = 339$

Step 3: $t_{crit} = \pm 1.960$ (± 1.980 would also be acceptable). The decision rule is that if t_{obt} is greater than 1.960 or less than -1.960, the null will be rejected.

$$\hat{\sigma}_{\bar{x}_1 + \bar{x}_2} = \sqrt{\frac{119,777,854.56 + 11,422,557}{339}} \sqrt{\frac{341}{28,960}}$$

$$= \sqrt{387,021.86} \sqrt{.01} = 622.11(.10)$$

$$= 62.21, \text{ and}$$

Step 4: $$t_{obt} = \frac{329.87}{62.21} = 5.30$$

Step 5: Since t_{obt} is greater than 1.960, the null is rejected. There is a statistically significant difference between property and drug offenders' mean fines. In other words, there does not appear to be a relationship between crime type and fine amount.

19.

Step 1: $H_0: \mu_1 = \mu_2$; $H_1: \mu_1 \neq \mu_2$

Step 2: t distribution with $df = 5 - 1 = 4$

Step 3: $t_{crit} = \pm 2.776$. The decision rule is that if t_{obt} is greater than 2.776 or less than -2.776, the

null will be rejected.

$$\bar{x}_D = .34; \; s_D = \sqrt{\frac{.30}{4}} = \sqrt{.08} = .28;$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \frac{.28}{\sqrt{5}} = \frac{.28}{2.24} = .13; \; \text{and}$$

$$t_{obt} = \frac{.34}{.13} = 2.62$$

Step 4:

Step 5: Since t_{obt} is not greater than 2.776 or less than –2.776, the null is retained. There is no difference between states with high and low arrest rates in terms of officer assault. In other words, there does not appear to be a relationship between arrest rates and officer assaults.

21.

Step 1: H_0: $P_1 = P_2$; H_1: $P_1 \neq P_2$.

Step 2: z distribution.

Step 3: $z_{crit} = \pm 1.96$ (recall that .50 – .025 = .475). The decision rule is that if z_{obt} is less than –1.96 or greater than 1.96, the null will be rejected.

$$\hat{p} = \frac{325.76 + 23.36}{582} = \frac{349.12}{582} = .60;$$

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{.60(.40)}\sqrt{\frac{582}{37{,}157}}$$

$$= \sqrt{.24}\sqrt{.02} = (.49)(.14) = .07; \; \text{and}$$

$$z_{obt} = \frac{.32}{.07} = 4.57$$

Step 4:

Step 5: Since z_{obt} is greater than 1.96, the null is rejected. There is a significant difference between juveniles represented by public attorneys and those represented by private counsel in terms of the time it takes for their cases to reach disposition. In other words, there appears to be a relationship between attorney type and time-to-disposition among juvenile drug defendants.

23.

1. Equal/pooled variances. Levene's F = 2.810 with a p value of .095. Since .095 > .05, the F statistic is not significant at alpha = .05 (i.e., the null of equal variances is retained).

2. t_{obt} = .977

3. No. The p value for t_{obt} is .330, which well exceeds .01; therefore, the null is retained.

4. There is no statistically significant difference between daytime and nighttime stops in terms of duration. That is, there seems to be no relationship between whether a stop takes place at day or night and the length of time the stop lasts.

25.

1. Unequal/separate variances. Levene's F = 36.062 with a p value of .000. Since .000 < .05, the F statistic is significant at alpha = .05 (i.e., the null of equal variances is rejected).

2. t_{obt} = 8.095.

3. Yes. The p value for t_{obt} is .000, which is less than .01; therefore, the null is retained.

4. There is a statistically significant difference between prosecutors' offices that do and do not use DNA in plea negotiations and trials in the total number of felony convictions obtained each year. In other words, there is a relationship between DNA usage and total felony convictions. (Though one would suspect, of course, that this relationship is spurious and attributable to the fact that larger prosecutors' offices process more cases and are more likely to use DNA as compared to smaller offices.)

Chapter 12

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number. Calculation steps are identical to those in the text; using alternative sequences of steps might result in answers different from those presented here. These differences might or might not alter the final decision regarding the null.

1.
    1. judges' gender
    2. nominal
    3. sentence severity
    4. ratio
    5. independent-samples t test

3.
    1. arrest
    2. nominal
    3. recidivism
    4. ratio
    5. ANOVA

5.
    1. poverty
    2. ordinal
    3. crime rate
    4. ratio
    5. ANOVA

7. Within-groups variance measures the amount of variability present among different members of the same group. This type of variance is akin to white noise: it is the random fluctuations inevitably present in any group of people, objects, or places. Between-groups variance measures the extent to which groups differ from one another. This type of variance conveys information about whether or not there are actual differences between groups.

9. The F statistic can never be negative because it is a measure of variance and variance cannot be negative. Mathematically, variance is a squared measure; all negative numbers are squared during the course of calculations. The final result, then, is always positive.

11.
    Step 1: H_0: $\mu_1 = \mu_2 = \mu_3$ and H_1: some $\mu_i \neq$ some μ_j.
    Step 2: F distribution with $df_B = 3 - 1 = 2$ and $df_W = 21 - 3 = 18$
    Step 3: $F_{crit} = 3.55$ and the decision rule is that if F_{obt} is greater than 3.55, the null will be rejected.

$$ss_T = (206 + 1{,}240 + 11) - \frac{(20 + 68 + 5)^2}{21}$$

$$= 1{,}457 - 411.86$$

Step 4: $\quad = 1{,}045.14$

$SS_B = 7(2.86 - 4.43)^2 + 7(9.71 - 4.43)^2 + 7(.71 - 4.43)^2 = 7(2.46) + 7(27.88) + 7(13.84) = 309.26;$

$SS_w = 1{,}045.14 - 309.26 = 735.88;$

$$MS_B = \frac{309.26}{2} = 154.63$$

$$MS_W = \frac{735.88}{18} = 40.88$$

$$F_{obt} = \frac{154.63}{40.88} = 3.78$$

Step 5: Since F_{obt} is greater than 3.55, the null is rejected. There is a statistically significant difference in the number of wiretaps authorized per crime type. In other words, wiretaps vary significantly across crime types. Since the null was rejected, it is appropriate to examine

$$\omega^2 = \frac{309.26 - 2(40.88)}{40.88 + 1{,}045.14} = .21$$

omega squared:

This means that 21% of the variance in wiretap authorizations is attributable to crime type.

13.

Step 1: H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ and H_1: some $\mu_i \neq$ some μ_j.

Step 2: F distribution with $df_B = 4 - 1 = 3$ and $df_W = 23 - 4 = 19$

Step 3: $F_{crit} = 5.01$ and the decision rule is that if F_{obt} is greater than 5.01, the null will be rejected.

$$ss_T = (9.99 + 58.35 + 42.15 + 91.04)$$

$$- \frac{(5.17 + 15.12 + 14.27 + 23.46)^2}{23}$$

Step 4: $\quad = 201.53 - 146.36 = 55.17;$

$SS_B = 5(1.03 - 2.52)^2 + 6(2.52 - 2.52)^2 + 5(2.85 - 2.52)^2 + 7(3.35 - 2.52)^2 = 5(2.22) + 6(0) + 5(.11) + 7(.69) = 16.48;$

$SS_w = 55.17 - 16.48 = 38.69$

$$MS_B = \frac{16.48}{3} = 5.49$$

$$MS_W = \frac{38.69}{19} = 2.04$$

$$F_{obt} = \frac{5.49}{2.04} = 2.69$$

Step 5: Since F_{obt} is less than 5.01, the null is retained. There are no statistically significant differences between regions in terms of the percentage of officer assaults committed with firearms. In other words, there is no apparent relationship between region and firearm involvement in officer assaults. Since the null was retained, it is not appropriate to calculate omega squared.

15.

Step 1: H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ and H_0: some $\mu_i \neq$ some μ_j.

Step 2: F distribution with $df_B = 4 - 1 = 3$ and $df_W = 20 - 4 = 16$.

Step 3: $F_{crit} = 5.29$ and the decision rule is that if F_{obt} is greater than 5.29, the null will be rejected.

$$SS_T = (1,070 + 5,468 + 949 + 508)$$
$$- \frac{(70 + 102 + 53 + 38)^2}{20}$$
$$= 7,995 - 3,458.45$$

Step 4:
$$= 4,536.55$$

$SS_B = 6(11.67 - 13.15)^2 + 5(20.40 - 13.15)^2 + 5(10.60 - 13.15)^2 + 4(9.50 - 13.15)^2 = 6(2.19) + 5(52.56) + 5(6.50) + 4(13.32) = 361.72$;

$SS_w = 4,536.55 - 361.72 = 4,174.83$

$$MS_B = \frac{361.72}{3} = 120.57$$

$$MS_W = \frac{4,174.83}{16} = 260.93$$

$$F_{obt} = \frac{120.57}{260.93} = .46$$

Step 5: Since F_{obt} is less than 5.29, the null is retained. There are no statistically significant differences between juveniles of different races in the length of probation sentences they receive.

In other words, there is no apparent relationship between race and probation sentences among juvenile property offenders. Since the null was retained, it is not appropriate to calculate omega squared.

17.

Step 1: H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ and H_1: some $\mu_i \neq$ some μ_j.

Step 2: F distribution with $df_B = 4 - 1 = 3$ and $df_W = 20 - 4 = 16$.

Step 3: $F_{crit} = 3.24$ and the decision rule is that if F_{obt} is greater than 3.24, the null will be rejected.

Step 4:

$$SS_T = \left(9,811 + 6,561 + 4,211 + 16,479\right) - \frac{\left(221 + 181 + 145 + 287\right)^2}{20}$$

$$= 37,062 - \frac{834^2}{20} = 37,062 - \frac{695,556}{20}$$

$=37,062 - 34,777.80$

$= 2,284.20$

$SS_B = 5(44.20 - 41.70)^2 + 5(36.20 - 41.70)^2 + 5(29.00 - 41.70)^2 + 5(57.40 - 41.70)^2$

$= 5(2.50)^2 + 5(-5.50)^2 + 5(-12.70)^2 + 5(15.70)^2$

$= 5(6.25) + 5(30.25) + 5(161.29) + 5(246.49)$

$= 31.25 + 151.25 + 806.45 + 1,232.45$

$= 2,221.40$

$SS_W = 2,284.20 - 2,221.40 = 62.80$

$$MS_B = \frac{2,221.40}{4 - 1} = 740.67 \quad MS_W = \frac{62.80}{20 - 4} = 3.93$$

$$F_{obt} = \frac{740.67}{3.93} = 188.47$$

Step 5: Since F_{obt} is greater than 3.24, the null is rejected. There is a statistically significant difference in the percentage of patrol personnel police managers tasked with responsibility for engaging in problem solving, depending on agency type. Since the null was rejected, it is appropriate to examine omega squared:

$$\omega^2 = \frac{2,221.40 - \left(4 - 1\right)3.93}{3.93 + 2,284.20} = \frac{2,209.61}{2,288.13} = .97$$

This means that 97% of the variance in patrol assignment to problem-oriented tasks exists between agencies (i.e., across groups). Agency type is an important predictor of the extent to which top managers allocate patrol personnel to problem solving.

19.
1. F_{obt} = 9.631.
2. Yes. The p value is .000, which is less than .01, so the null is rejected.
3. Among juvenile property defendants, there are significant differences between different racial groups in the amount of time it takes to acquire pretrial release. In other words, there is a relationship between race and time-to-release.
4. Since the null was rejected, post hoc tests can be examined. Tukey and Bonferroni tests show that there is one difference, and it lies between black and white youth. Group means reveal that black youths' mean time-to-release is 40.36 and white youths' is 20.80. Hispanics, with a mean of 30.30, appear to fall in the middle and are not significantly different from either of the other groups.
5. Since the null was rejected, it is correct to calculate omega squared:

$$\omega^2 = \frac{49125.707 - 2(2550.431)}{2550.431 + 1599787.836} = .027$$

Only about 2.7% of the variance in time-to-release is attributable to race. (This means that important variables are missing! Knowing, for instance, juveniles' offense types and prior records would likely improve our understanding of the timing of their release.)

21.
1. F_{obt} = 3.496.
2. The null would be rejected because p = .000, which is less than .05.
3. There is a relationship between shooters' intentions and victims' ages. In other words, victim age varies across assaults, accidents, and officer-involved shootings.
4. Since the null was rejected, post hoc tests can be examined. Tukey and Bonferroni both show that the mean age of people shot by police is significantly different from the mean ages of victims of assaults and accidental shootings. The group means are 28.99, 29.04, and 33.55 for people shot unintentionally, in assaults, and by police, respectively. This shows that people shot by police are significantly older than those shot in other circumstances.
5. It is appropriate to calculate and interpret omega squared. Using the SPSS output,

$$\omega^2 = \frac{1,249.17 - (3-1)178.64}{178.64 + 715,618.38} = \frac{891.89}{715,797.02} = .001$$

This value of omega squared is nearly zero (.01%) and suggests that shooter intent explains a miniscule amount of the variance in victim age.

Chapter 13

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number. Calculation steps are identical to those in the text; using alternative sequences of steps might result in answers different from those presented here. These differences might or might not alter the final decision regarding the null.

1.
    1. parental incarceration
    2. nominal
    3. lifetime incarceration
    4. nominal
    5. chi-square

3.
    1. participation in community meetings
    2. nominal
    3. self-protective measures
    4. ratio
    5. two-population z test for a difference between proportions

5. A linear relationship is one in which a single-unit increase in the independent variable is associated with a constant change in the dependent variable. In other words, the magnitude and the direction of the relationship remain constant across all levels of the independent variable. When graphed, the IV–DV overlap appears as a straight line.

7. The line of best fit is the line that minimizes the distance from that line to each of the raw values in the data set. That is, it is the line that produces the smallest deviation scores (or error). No other line would come closer to all of the data points in the sample.

9. c

11.
   Step 1: H_0: $\rho = 0$ and H_1: $\rho \neq 0$.
   Step 2: t distribution with $df = 5 - 2 = 3$.

$$r = \frac{5(2,274.83) - (627)(17.33)}{\sqrt{[5(79,715) - 627^2][5(76.15) - 17.33^2]}}$$

$$= \frac{11,374.15 - 10,865.91}{\sqrt{[5446 * 80.42]}}$$

$$= \frac{508.24}{661.79} = .77$$

Step 4:

$$t_{obt} = .77\left(\sqrt{\frac{5-2}{1-.77^2}}\right) = .77\left(\sqrt{\frac{3}{.41}}\right) = .77(2.71) = 2.08$$

Step 5: Since t_{obt} is not greater than 3.182, the null is retained. There is no correlation between prison expenditures and violent crime rates. In other words, prison expenditures do not appear to impact violent crime rates. Because the null was retained, it is not appropriate to examine the sign, the magnitude, or the coefficient of determination.

13.

Step 1: H_0: ρ = 0 and H_1: ρ > 0.

Step 2: t distribution with df = 9 – 2 = 7.

Step 3: t_{crit} = 1.895 and the decision rule is that if t_{crit} is greater than 1.895, the null will be rejected.

$$r = \frac{9(202.05) - 27.80(53.90)}{\sqrt{[9(136.34) - 27.80^2][9(367.91) - 53.90^2]}}$$

$$= \frac{1,818.45 - 1,498.42}{\sqrt{[1,227.06 - 772.84][3,311.19 - 2905.21]}}$$

$$= \frac{320.03}{\sqrt{184,404.24}} = .75$$

Step 4:

$$t_{obt} = .75\sqrt{\frac{9-2}{1-.75^2}} = .75\sqrt{\frac{7}{.44}}$$

$$= .75(3.99) = 3$$

Step 5: Since t_{obt} is greater than 1.895, the null is rejected. There is a positive correlation between

crime concentration and concentration of police agencies. In other words, where there is more crime, there are apparently more police agencies. Since the null was rejected, the sign, the magnitude, and the coefficient of determination can be examined. The sign is positive, meaning that a one-unit increase in the IV is associated with an increase in the DV. The magnitude is very strong, judging by the guidelines offered in the text (where values between 0 and ±.29 are weak, from about ±.30 to ±.49 are moderate, ±.50 to ±.69 are strong, and those beyond ±.70 are very strong). The coefficient of determination is $.75^2 = .56$. This means that 56% of the variance in police agencies can be attributed to crime rates.

15.

Step 1: H_0: $\rho = 0$ and H_1: $\rho < 0$.

Step 2: t distribution with $df = 7 - 2 = 5$.

Step 3: $t_{crit} = -3.365$ and the decision rule is that if t_{crit} is less than −3.365, the null will be rejected.

Step 4:

$$r = \frac{7(66.5) - (123.97)(3.93)}{\sqrt{\left[7(12{,}791.57) - (123.97^2)\right]\left[7(2.33) - (3.93)^2\right]}}$$

$$= \frac{465.50 - 487.20}{\sqrt{[74{,}172.40][.89]}}$$

$$= \frac{-21.70}{256.67} = -.08$$

$$t_{obt} = -.08\sqrt{\frac{7-2}{1 - -.08^2}} = .08\sqrt{\frac{5}{.99}} = -.18$$

Step 5: Since t_{obt} is not less than −3.365, the null is retained. There is no correlation between handgun and knife murder rates. In other words, murder handgun rates do not appear to affect knife murder rates. Because the null was retained, it is not appropriate to examine the sign, the magnitude, or the coefficient of determination.

17.

Step 1: H_0: $\rho = 0$ and H_1: $\rho \neq 0$.

Step 2: t distribution with $df = 8 - 2 = 6$.

Step 3: $t_{crit} = \pm 2.447$ and the decision rule is that if t_{crit} is less than −2.447 or greater than 2.447, the null will be rejected.

$$r = \frac{8(570) - 121(37)}{\sqrt{\left[8(1{,}849) - 121^2\right]\left[8(195) - 37^2\right]}}$$

$$= \frac{4{,}560 - 4{,}477}{\sqrt{\left[14{,}792 - 14{,}641\right]\left[1{,}560 - 1{,}369\right]}}$$

$$= \frac{83}{\sqrt{28{,}841}} = .49$$

Step 4:

$$t_{obt} = .49\sqrt{\frac{8-2}{1-.49^2}} = .49\sqrt{\frac{6}{.76}} = .49(2.81) = 1.38$$

Step 5: Since t_{obt} is not greater than 2.447, the null is retained. There is no correlation between age and time-to-disposition among female juveniles. In other words, girls' ages do not appear to affect the time it takes for their cases to reach adjudication. Because the null was retained, it is not appropriate to examine the sign, the magnitude, or the coefficient of determination.

19.

1. For age and contacts, $r = .044$; for age and length, $r = -.113$; for contacts and length, $r = .312$.

2. For age and contacts, the null is retained because .5556 > .05; for age and length, the null is retained because .308 > .05; for contacts and length, the null is rejected because .004 < .05.

3. There is no correlation between age and the number of contacts male Asian respondents had with police in the past year; there is no correlation between age and the duration of traffic stops for those who had been stopped while driving; the number of recent contacts with police is significantly correlated with the duration of traffic stops.

4. The only test for which it is appropriate to examine the sign, the magnitude, and the coefficient of determination is the test for contacts and length. The sign is positive, meaning that a one-unit increase in one of these variables is associated with an increase in the other one. This suggests that those who experience more contacts also have longer encounters with officers. The magnitude is moderate. The coefficient of determination is $.312^2 = .097$, meaning that only 9.7% of the variance in stop length is attributable to number of contacts (or 9.7% of these variables' variance is shared). This suggests that although the variables are statistically related, the relationship is of very modest substantive importance.

Chapter 14

Note: Rounding, where applicable, is to two decimal places in each step of calculations and in the final answer. For numbers close to zero, decimals are extended to the first nonzero number. Calculation steps are identical to those in the text; using alternative sequences of steps might result in answers different from those presented here. These differences might or might not alter the final decision regarding the null. Numbers gleaned from SPSS output are presented using three decimal places.

1. Regression's advantage over correlation is its ability to predict values of the DV (y) using specified values of the IV (x). Because regression fits a line to the data, any value of y can be predicted using any value of x. Values of x can be plugged into the formula for the line of best fit ($\hat{y} = a + bx$, at the bivariate level), and the equation can be solved to produce the predicted value of y at the given value of x.

3. No. The OLS model can accommodate IVs of any level of measurement.

5.

a.
$$b = \frac{5(86) - 24(19)}{5(230) - 24^2} = \frac{430 - 456}{1150 - 576} = -.05$$

b. $a = 3.80 - (-.05)4.80 = 3.80 - (-.24) = 4.04$

c. $\hat{y} = 4.04 - .05x$

d.

    1. For $x = 3$, $\hat{y} = 4.04 - .05(3) = 3.89$

    2. For $x = 15$, $\hat{y} = 4.04 - .05(15) = 3.26$

e.

    Step 1: H_0: B = 0 and H_1: B ≠ 0

    Step 2: t distribution with $df = 5 - 2 = 3$

    Step 3: $t_{crit} = \pm 3.182$ and the decision rule is that if t_{obt} is greater than 3.182 or less than −3.182, the null will be rejected.

    Step 4:
$$SE_b = \frac{\sqrt{\dfrac{26.56}{5-2}}}{\sqrt{114.80}} = \frac{\sqrt{8.85}}{10.71} = \frac{2.97}{10.71} = .28$$

$$t_{obt} = \frac{-.05}{.28} = -.18$$

    Step 5: Since t_{obt} is not less than −3.182, the null is retained. There is not a statistically significant relationship between prisoners' prior incarcerations and their in-prison behavior. That is, prior incarceration history does not appear to significantly predict in-prison behavior.

5f. The null was not rejected, so the beta weight should not be calculated.

7.

$$b = \frac{10(251.13) - 66(35.62)}{10(450.48) - 66^2}$$

$$= \frac{2,511.30 - 2,350.92}{4,504.80 - 4,356} = 1.08$$

a.

b. a = 3.56 – (1.08)6.60 = –3.57

c. $\hat{y}$ = –3.57 + 1.08x

d.

    1. For x = 4, $\hat{y}$ = –3.57 + 1.08(4) = .75

    2. For x = 8, $\hat{y}$ = –3.57 + 1.08(8) = 5.07

e.

    Step 1: H_0: B = 0 and H_1: $B \neq 0$.

    Step 2: t distribution with df = 10 – 2 = 8.

    Step 3: t_{crit} = ±2.306 and the decision rule is that if t_{obt} is greater than 2.306 or less than –2.306, the null will be rejected.

$$SE_b = \frac{\sqrt{\frac{10.58}{10-2}}}{\sqrt{14.88}} = \frac{1.15}{3.86} = .30 \quad t_{obt} = \frac{1.08}{.30} = 3.60$$

Step 4:

Step 5: Since t_{obt} is greater than 2.306, the null is rejected. There is a statistically significant relationship between unemployment rates and violent crime. That is, the concentration of unemployment in a state appears to help predict that state's violent crime rate.

$$\beta = 1.08\left(\frac{1.29}{1.76}\right) = 1.08(.73) = .79$$

f.

9.

$$b = \frac{10(440.07) - 95.80(42.81)}{10(965.02) - 95.80^2}$$

$$= \frac{4,400.70 - 4,101.20}{9,650.20 - 9,177.64} = .63$$

a.

b. a = 4.28 – (.63)9.58 = –1.76

c. $\hat{y}$ = –1.76 + .63x

d.

    1. For x = 5, $\hat{y}$ = –1.76 + .63(5) = 1.39

    2. For x = 10, $\hat{y}$ = –1.76 + .63(10) = 4.54

e.

Step 1: H_0: $B = 0$ and H_1: $B \neq 0$.

Step 2: t distribution with $df = 10 - 2 = 8$.

Step 3: $t_{crit} = \pm 2.306$ and the decision rule is that if t_{obt} is greater than 2.306 or less than -2.306, the null will be rejected.

Step 4:
$$SE_b = \frac{\sqrt{\dfrac{12.65}{10-2}}}{\sqrt{47.26}} = \frac{1.26}{6.87} = .18$$

$$SE_b = \frac{\sqrt{\dfrac{\sum(y_i - \hat{y}_i)^2}{N-2}}}{\sqrt{\sum(x_i - \overline{x})^2}} \qquad t_{obt} = \frac{.63}{.18} = 3.50$$

Step 5: Since t_{obt} is greater than 2.306, the null is rejected. There is a statistically significant relationship between poverty rates and violent crime. That is, the concentration of poverty in a state appears to help predict that state's violent crime rate.

f.
$$\beta = .63\left(\frac{2.29}{1.87}\right) = .63(1.22) = .77$$

11.

1. $F = 19.176$. The p value is .000. Since .000 < .05, the null is rejected. The model is statistically significant, meaning that the IV explains a statistically significant proportion of the variance in the DV. This means that it is appropriate to continue on and examine the slope coefficient. (Recall that an F value that is not statistically significant means that you should not move on and examine the specifics of the model because the model is not useful.)

2. $R^2 = .042$. This means that *charges* explains 4.2% of the variance in *probation*. This is a very small amount of variance explained.

3. $\hat{y} = 23.947 + 1.170x$

4. The slope coefficient's p value is .000, which is less than .05, so b is statistically significant.

5. Since b is statistically significant, it is appropriate to examine the beta weight. Here, beta = .205.

6. The number of charges juvenile drug defendants face is a statistically significant predictor of the length of their probation sentences; however, this relationship is weak, and the number of charges is not a substantively meaningful predictor. Clearly, important variables have been omitted from the model.

13.

1. $\hat{y} = 27.438 + 1.079(2.09) - 4.322(0) = 29.69$

2. $\hat{y} = 27.438 + 1.079(2.09) - 4.322(1) = 25.37$

3. The predicted sentence decreased by 4.322 units—exactly the slope coefficient for the *priors*

variable! The decrease means that juveniles with prior records received probation sentences that were, on average, 4.322 months shorter compared to juveniles without records. This seems counterintuitive, but there are possible explanations (such as judges treating first-time offenders more harshly in order to "teach them a lesson") that we are not able to explore with the present data.

14.
   1. $\hat{y} = 27.438 + 1.079(2) - 4.322(.79) = 26.18$
   2. $\hat{y} = 27.438 + 1.079(5) - 4.322(.79) = 29.42$
   3. The predicted sentence increased by 3.24 units. This means that a juvenile facing five charges should receive, on average, a sentence that is 3.24 months longer than a juvenile facing two charges.

15.
   1. $F = 14.270$. Since $p = .001$, which is less than .05, the F statistic is significant. The IV explains a statistically significant amount of variance in the model, so it is appropriate to continue on and examine the slope coefficient.
   2. $R^2 = .302$. This means that the IV explains 30.2% of the variance in the DV. This is a strong relationship, suggesting that unemployment is a meaningful predictor of violent crime rates.
   3. $\hat{y} = -.410 + .637x$
   4. The p value for the slope coefficient is .001, which is less than .05, so b is statistically significant.
   5. Since b is significant, it is appropriate to examine beta. Beta = .549.
   6. Unemployment rates are statistically significant and substantively meaningful predictors of violent crime. Knowing the unemployment rate in an area significantly improves the ability to predict that area's violent crime rate.

17. $\hat{y} = -1.437 + .648(10) - .231(15) + .320(16) = 6.70$

19.
   1. $\hat{y} = -1.437 + .648(6.77) - .231(11) + .320(8.94) = 3.27$
   2. $\hat{y} = -1.437 + .648(6.77) - .231(6) + .320(8.94) = 4.42$
   3. The rate declined by 1.15 units. This means that states with 11% of the population receiving SNAP benefits should have violent crime rates that are, on average, 1.15 points lower than states where 6% of the population receives SNAP benefits. The negative association between SNAP rates and violent crime might seem backward, but it makes sense if SNAP is viewed as a measure of social support. Research has found that states and countries that provide greater levels of social support to disadvantaged citizens have less violent crime compared to those states or countries that do not provide as much social support.

References

Chapter 1

Bureau of Justice Statistics. (2006). National Crime Victimization Survey, 2004: Codebook. Washington, DC: U.S. Department of Justice.

Corsaro, N., Brunson, R. K., & McGarrell, E. F. (2013). Problem-oriented policing and open-air drug markets: Examining the Rockford pulling levers deterrence strategy. Crime & Delinquency, 59 (7), 1085–1107.

Farrell, A. (2014). Environmental and institutional influences on police agency responses to human trafficking. Police Quarterly , 17 (1), 3–29.

Hughes, C., Warren, P. Y., Stewart, E. A., Tomaskovic-Devey, D., & Mears, D. P. (2017). Racial threat, intergroup contact, and school punishment. Journal of Research in Crime and Delinquency, 54 (5), 583–616.

Kerley, K. R., Hochstetler, A., & Copes, H. (2009). Self-control, prison victimization, and prison infractions. Criminal Justice Review, 34 (4), 553–568.

Muftić, L. R., Bouffard, L. A., & Bouffard, J. A. (2007). An exploratory analysis of victim precipitation among men and women arrested for intimate partner violence. Feminist Criminology, 2 (4), 327–346.

Paoline, E. A. III, Terrill, W., & Ingram, J. R. (2012). Police use of force and officer injuries: Comparing conducted energy devices (CEDs) to hands- and weapons-based tactics. Police Quarterly, 15 (2), 115–136.

White, M. D., Ready, J., Riggs, C., Dawes, D. M., Hinz, A., & Ho, J. D. (2013). An incident-level profile of TASER device deployments in arrest-related deaths. Police Quarterly, 16 (1), 85–112.

Chapter 2

Bouffard, L. A., & Piquero, N. L. (2010). Defiance theory and life course explanations of persistent offending. Crime & Delinquency, 56 (2), 227–252.

Bureau of Justice Statistics. (2011). The Police–Public Contact Survey, 2011. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Dass-Brailsford, P., & Myrick, A. C. (2010). Psychological trauma and substance abuse: The need for an integrated approach. Trauma, Violence, & Abuse, 11 (4), 202–213.

Davis, J. A., & Smith, T. W. (2009). General social surveys, 1972–2008. Chicago: National Opinion Research Center, producer, 2005; Storrs, CT: Roper Center for Public Opinion Research, University of Connecticut.

Gau, J. M., Mosher, C., & Pratt, T. C. (2010). An inquiry into the impact of suspect race on police use of Tasers. Police Quarterly, 13 (1), 27–48.

Graham v. Connor. 490 U.S. 386 (1989).

Haynes, S. H. (2011). The effects of victim-related contextual factors on the criminal justice system. Crime & Delinquency, 57 (2), 298–328.

Kleck, G., & Kovandzic, T. (2009). City-level characteristics and individual handgun ownership: Effects of collective security and homicide. Journal of Contemporary Criminal Justice, 25 (1), 45–66.

Lindsey, A. M., Mears, D. P., Cochran, J. C., Bales, W. D., & Stults, B. J. (2017). In prison and far from home: Spatial distance effects on inmate misconduct. Crime & Delinquency, 63 (9), 1043–1065. Social Justice, 26 (1), 115–138.

Schmidt, J. D., & Sherman, L. W. (1993). Does arrest deter domestic violence? American Behavioral Scientist, 36 (5), 601–609. Crime & Delinquency, 60 (4), 569–591.

Snell, T. L. (2014). Capital punishment in the United States, 2013–Statistical tables (BJS Publication No. 248448). Washington, DC: U.S. Department of Justice.

Terrill, W., & Paoline, E. A. III. (2013). Examining less lethal force policy and the force continuum: Results from a national use-of-force study.

U.S. Department of Justice. (2012). National Crime Victimization Survey, 2012. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Chapter 3

Lane, J., & Fox, K. A. (2013). Fear of property, violent, and gang crime: Examining the shadow of sexual assault thesis among male and female offenders. Criminal Justice and Behavior, 40 (5), 472–496.

Morgan, K. O., Morgan, S., & Boba, R. (2010). Crime: State rankings 2010. Washington, DC: CQ Press.

Nuñez, N., Myers, B., Wilkowski, B. M., & Schweitzer, K. (2017). The impact of angry versus sad victim impact statements on mock jurors' sentencing decisions in a capital trial. Criminal Justice & Behavior, 44 (6), 862–886.

Reaves, B. A. (2013). Felony defendants in large urban counties, 2009—Statistical tables. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

Steffensmeier, D., Zhong, H., Ackerman, J., Schwartz, J., & Agha, S. (2006). Gender gap trends for violent crimes, 1980 to 2003: A UCR–NCVS comparison. Feminist Criminology, 1 (1), 72–98.

Truman, J. L., & Morgan, R. E. (2016). Criminal victimization, 2015 (Publication No. NCJ 250180). Washington, DC: Bureau of Justice Statistics.

Chapter 4

Janssen, H. J., Weerman, F. M., & Eichelsheim, V. I. (2017). Parenting as a protective factor against criminogenic settings? Interaction effects between three aspects of parenting and unstructured socializing in disordered areas. Journal of Research in Crime and Delinquency, 54 (2), 181–207.

Logue, M. A. (2017). The interactive effects of national origin and immigration status on Latino drug traffickers in U.S. federal courts. Race and Justice, 7 (3), 276–296.

Morgan, K. O., Morgan, S., & Boba, R. (2010). Crime: State rankings 2010. Washington, DC: CQ Press.

Piquero, A. R., Sullivan, C. J., & Farrington, D. P. (2010). Assessing differences between short-term, high-rate offenders and long-term, low-rate offenders. Criminal Justice and Behavior, 37 (12), 1309–1329.

Snell, T. L. (2014). Capital punishment, 2013–Statistical tables (Publication No. 248448). Washington, DC: U.S. Department of Justice.

Sorensen, J., & Cunningham, M. D. (2010). Conviction offense and prison violence: A comparative study of murderers and other offenders. Crime & Delinquency, 56 (1), 103–125.

Truman, J. L., & Langton, L. (2015). Criminal victimization, 2014 (BJS Publication No. 248973). Washington, DC: U.S. Department of Justice.

Chapter 5

Copes, H., Kovandzic, T. V., Miller, J. M., & Williamson, L. (2014). The lost cause? Examining the Southern culture of honor through defensive gun use. Crime & Delinquency, 60 (3), 356–378.

Perry, S. W., & Banks, D. (2011). Prosecutors in state courts, 2007—Statistical tables (Publication No. NCJ 234211). Washington, DC: Bureau of Justice Statistics.

Pogarsky, G., & Piquero, A. R. (2003). Can punishment encourage offending? Investigating the "resetting" effect. Journal of Research in Crime and Delinquency, 40 (1), 95–12.

Sydes, M. (2017). Revitalized or disorganized? Unpacking the immigration-crime link in a multiethnic setting. Journal of Research in Crime and Delinquency, 54 (5), 680–714.

Chapter 6

Klinger, D. A. (1995). Policing spousal assault. Journal of Research in Crime and Delinquency, 32 (2), 308–324.

Paoline, E. A. III, Lambert, E. G., & Hogan, N. L. (2006). A calm and happy keeper of the keys: The impact of ACA views, relations with coworkers, and policy views on the job stress and job satisfaction of correctional staff. The Prison Journal, 86 (2), 182–205.

Ray, B., Grommon, E., Buchanan, V., Brown, B., & Watson, D. P. (2017). Access to recovery and recidivism among former prison inmates. International Journal of Offender Therapy and Comparative Criminology, 61 (8), 874–893.

Reaves, B. A. (2013). Felony defendants in large urban counties, 2009—Statistical tables. Washington, DC: Bureau of Justice Statistics.

Chapter 7

Bouffard, L. A. (2010). Period effects in the impact of Vietnam-era military service on crime over the life course. Crime & Delinquency. Advanced online publication. doi:10.1177/0011128710372455

Morris, R. G., & Worrall, J. L. (2010). Prison architecture and inmate misconduct: A multilevel assessment. Crime & Delinquency. Advanced online publication. do:10.1177/0011128709336940

Rosenfeld, R., & Fornango, R. (2014). The impact of police stops on precinct robbery and burglary rates in New York City, 2003 – 2010. Justice Quarterly, 31 (1), 96–122.

Chapter 8

Deslauriers-Varin, N., Beauregard, E., & Wong, J. (2011). Changing their mind about confessing to police: The role of contextual factors in crime confession. Police Quarterly, 14 (5), 5–24.

Dixon, T. L. (2017). Good guys are still always white? Positive change and continued misrepresentation of race and crime on local television news. Communication Research, 44 (6), 775–792.

Martin, K. R., & Garcia, L. (2011). Unintended pregnancy and intimate partner violence before and during pregnancy among Latina women in Los Angeles, California. Journal of Interpersonal Violence, 26 (6), 1157–1175.

Orth, U., & Maercker, A. (2004). Do trials of perpetrators retraumatize crime victims? Journal of Interpersonal Violence , 19 (2), 212–227.

Saum, C. A., Hiller, M. L., & Nolan, B. A. (2013). Predictors of completion of a driving under the influence (DUI) court for repeat offenders. Criminal Justice Review, 38 (2), 207–225.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, & United States Consumer Product Safety Commission. (n.d.). Firearm Injury Surveillance Study, 1993–2010 [computer file]. ICPSR33861-v1. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center

for Injury Prevention and Control [producer]; Ann Arbor MI: Inter-University Consortium for Political and Social Research [distributor].

Chapter 10

Griffin, T., Pason, A., Wiecko, F., & Brace, B. (2016). Comparing criminologists' views on crime and justice issue with those of the general public. Criminal Justice Policy Review. doi:10.1177/0887403416638412

Petersen, N. (2017). Examining the sources of racial bias in potentially capital cases: A case study of police and prosecutorial discretion. Race and Justice, 7 (1), 7–34.

Chapter 11

Barnoski, R. (2005). Sex offender sentencing in Washington State: Recidivism rates. Olympia: Washington State Institute for Public Policy.

Bureau of Justice Statistics. (1998). Juvenile Defendants in Criminal Court, 1998 [data file]. Washington, DC: U.S. Department of Justice.

Corsaro, N., Brunson, R. K., & McGarrell, E. F. (2013). Problem-oriented policing and open-air drug markets: Examining the Rockford pulling levers deterrence strategy. Crime & Delinquency, 59 (7), 1085–1107.

Kaskela, T., & Pitkänen, T. (2016). The effects of gender and previous prison sentence on the risk of committing crime among a substance-abusing population. Crime & Delinquency. doi:10.1177/0011128716682229

Ostermann, M., & Matejkowski, J. (2014). Estimating the impact of mental illness on costs of crimes: A matched samples comparison. Criminal Justice and Behavior, 41 (1), 20–40.

Wright, K. A., Pratt, T. C., & DeLisi, M. (2008). Examining offending specialization in a sample of male multiple homicide offenders. Homicide Studies, 12 (4), 381–398.

Chapter 12

De Melo, S. N., Pereira, D. V. S., Andresen, M. A., & Matias, L. F. (2017). Spatial/temporal variations of crime: A routine activity theory perspective. International Journal of Offender Therapy and Comparative Criminology. doi: 0306624X17703654

Duff, J. C. (2010). Report of the director of the Administrative Office of the United States Courts. Retrieved from www.uscourts.gov/uscourts/Statistics/WiretapReports/2009/2009Wiretaptext.pdf

Franklin, T. W., & Fearn, N. E. (2010). Sentencing Asian offenders in state courts: The influence of a prevalent stereotype. Crime & Delinquency. OnLineFirst.

Tang, C. M., Nuñez, N., & Bourgeois, M. (2009). Effects of trial venue and pretrial bias on the evaluation of juvenile defendants. Criminal Justice Review, 34 (2), 210–225.

Chapter 13

Henson, B., Reyns, B. W., Klahm, C. F., IV, & Frank, J. (2010). Do good recruits make good cops? Problems predicting and measuring academy and street-level success. Police Quarterly, 13 (1), 5–26.

Morgan, K. O., Morgan, S., & Boba, R. (2010). Crime: State rankings 2010. Washington, DC: CQ Press.

Reisig, M. D., Pratt, T. C., & Holtfreter, K. (2009). Perceived risk of Internet theft victimization: Examining the effects of social vulnerability and financial impulsivity. Criminal Justice and Behavior, 36 (4), 369–384.

Walters, G. D. (2015). Criminal thinking as a predictor of prison misconduct and mediator of the static risk–infractions relationship. The Prison Journal, 95 (3), 353–369.

Chapter 14

Henson, B., Reyns, B. W., Klahm, C. F., IV, & Frank, J. (2010). Do good recruits make good cops? Problems predicting and measuring academy and street-level success. Police Quarterly, 13 (1), 5–26.

Mears, D. P., Mancini, C., & Stewart, E. A. (2009). Whites' concern about crime: The effects of interracial contact. Journal of Research in Crime and Delinquency, 46 (4), 524–552.

Petkovsek, M. A., & Boutwell, B. B. (2014). Childhood intelligence and the emergence of low self-control.

Criminal Justice and Behavior. Advanced online publication. doi:10.1177/0093854814537812

Wright, K. A., Pratt, T. C., & DeLisi, M. (2008). Examining offending specialization in a sample of male multiple homicide offenders. Homicide Studies, 12 (4), 381–398.

Index