1

Introduction

1.1 Statistics

The term *statistics* has several meanings. For example, it can be defined simply as collections of data (information that can be quantified as numbers, qualified as text, or weighed as materials). The term can also denote the theories and techniques used to process data. All the chapters of this textbook demonstrate that statistics is a versatile set of tools for manipulating, organizing, analyzing, and understanding data, which enhances the work of criminal justice practitioners, administrators, and researchers. Written with a capital "S," Statistics is a branch of mathematical science that consists of methods for describing and interpreting data. Criminal justice majors often take a Statistics course to prepare themselves for consuming published research or undertaking their studies. In such classes, they discover that learning Statistics involves acquiring skills that help answer questions about data and solve problems with it. Statistics also reduces large blocks of data so that they are more easily understood, displayed, examined, and applied to achieve a specific goal or answer a particular question.

Statistics with a lowercase "s" are numbers that represent the characteristics of groups. For example, the typical number of years of education in a cohort of Federal Bureau of Investigation (FBI) agents assigned to a field office is a statistic that describes one feature of a subset of people from a larger group—all FBI agents in the country. The number of citizen complaints against police officers in a major US city in 2023 is another such statistic. Statistical facts about the criminal justice system are vast and can be complied with at the local, state, and federal levels. The *Sourcebook of Criminal Justice Statistics* is a commonly used reference volume that contains a wealth of criminal justice statistics of tremendous value to both researchers and practitioners alike. It is only one of many compendiums of criminal justice statistics that can be referenced to garner interesting facts about criminal-justice-related populations and activities.

1.2 Types of Statistics: Descriptive and Inferential

Statistics is divided into two principal areas of study and application. As their names suggest, the branch of Statistics that describes and displays data is called descriptive statistics, and the branch of Statistics that draws inferences from data is called inferential statistics. The purpose of descriptive statistics is to summarize data collection clearly and understandably. Researchers compute statistics or display data to show

where their center lies, how much they vary, and how they are related from one aspect to another in the same dataset. For example, creating a profile of probation officers in an urban probation department could involve counting them and grouping them by gender, age, number of years in the department, type of cases supervised, and so on. Graphs and charts could be created to display these data using this information. This is an example of descriptive statistics.

Do the findings in this probation department (one sample in one agency) generalize to all urban probation departments (a population of such agencies)? This question is answered with inferential statistics, which requires selecting small samples from a population and drawing inferences about the characteristics of the population based on the sample data. For example, suppose that 25% of the probation officers in a national sample of county probation departments are women. Using inferential statistics, the percentage of female probation officers nationwide in all county probation departments can be estimated. Specifically, with inferential statistics, the characteristics of a sample (or subset) of probation departments can be extrapolated to the characteristics of the population of all probation departments. Descriptive statistics describe what the data look like. Inferential statistics make educated guesses about a population based on data from a population sample.

1.3 Basic Statistical Concepts and Terminology

Before specific statistical procedures are introduced, basic statistical concepts and vocabulary are explained and defined below. To understand Statistics, the following terms are critical: *statistic, parameter, population, element, census, sample, variable, and levels of measurement*.

1.3.1 Population, Element, Census, and Sample

A statistic is a number that represents a characteristic of a sample. An individual number is a statistic, and groups of numbers are statistics. A parameter is analogous to a statistic. Statistics describe samples, and parameters describe populations. Population, element, census, and sample are all in statistics textbooks. A population is the complete collection of persons, objects, or events being studied. For example, all the students at a university, all the ships in the US Navy, all the felony cases in the United States filed in 2019, and all the crimes reported in the Uniform Crime Reports in 2020 are populations. Although populations usually consist of large groups, size is irrelevant in defining a population. Populations can be small. For example, to construct the psychological profiles of mass murderers, researchers are unlikely to locate many such individuals in the United States or even around the world. The group of mass murderers is the population of interest, even though it includes a small number of known individuals.

An element is a person, object, or event in a population. Populations consist of elements of varying sizes. Some populations are enormous (e.g., the population of the United States), whereas others are tiny (e.g., justices of the Supreme Court of the United States). For example, one student at a university is an element in the population of all students. Similarly, the USS Constitution is an element in the population of all ships in the US Navy.

A sample is a group of elements chosen from a population. If only 30% of the students at a university were included in a study, they would constitute a sample from the population of all students at that school. If only 1,000 US naval ships were studied, they would constitute a sample of the population of all

vessels in the Navy. A study of only a subset of felony cases filed in 2020 would constitute a sample of the population of all felony cases filed that year.

The greatest extent of information can be obtained by measuring every element in a population—that is, by performing a census. A census is the study of all elements in a population. The US Census Bureau counts all the people in the country by conducting the decennial census (every 10 years). Studying all the students at a university or all the ships in the US Navy would also involve a census. A census is the only approach for gathering information or data from an entire population. On the other hand, conducting a census is very expensive and often logistically impractical or impossible.

Parameters and statistics represent the measured characteristics of populations and samples, respectively. Parameters are always unknown; a population census is required to calculate a parameter's actual value directly and precisely. For this reason, the value of a characteristic of a population is rarely known with certainty because it is costly and onerous to conduct a census. For example, suppose the US Transportation and Safety Administration endeavored to count the number of times people ran a red light in the United States in 2020. The population would be all drivers (licensed or unlicensed) in the country. Information about the number of times officers have ticketed people running red lights is recorded. However, no information is known about the driving infractions of every person (ticketed and unticketed); consequently, the parameter—the number of times people ran a red light in the population in 2020—is unknown. The discrepancy between a sample statistic and its population parameter is called a sampling error, which estimates how close the statistic is to the parameter. The lower the sampling error, the more accurately the statistic approximates the parameter.

1.3.2 Data and Variables

As noted above, data is derived from the word datum, which is defined as an individual piece of information. The plural form, data, is used since researchers and practitioners rarely deal with a single or isolated fact. Data are collections of each piece of information (a datum) about a sample or population. Data can also be defined as the information units collected and analyzed. For example, the FBI collects records of the number and type of criminal offenses that occur annually in the United States, which constitutes the data of the Uniform Crime Reports.

A variable is a characteristic or attribute that differs or varies among all the persons, objects, or events being studied. For example, the racial or ethnic composition of students in an introductory criminal justice course is a variable. The criminal justice field is concerned with an extensive repository of variables that capture the characteristics of people, objects, or events. People include arrestees, inmates, correctional officers, judges, attorneys, police officers, citizen complainants, and court clerks. Objects include police cars, felony cases, jails, prisons, judges' chambers, and gun evidence, and events include crimes, police shootings, convictions, sentences, incarcerations, victimizations, and executions, to name just a few.

In criminal justice research, people, objects, or events are never studied in a general or abstract way. Instead, researchers focus on specific characteristics or attributes known as variables. For example, a study of offenders might include age, race, gender, number of convictions, and type of weapons used in their most recent violent crime. In most studies of criminal offenders, subjects would be of different ages and have committed other offenses or used different weapons. In a survey of only male offenders, gender is a constant, not a variable.

After researchers decide what to study, they examine only the most relevant variables to their investigation. To do so, researchers ask, "What do I want to know most about these particular persons, objects, or events?" The answer to that question generates a list of variables for the study. Investigators examining the backgrounds of probation officers, for example, might be interested in their age, gender, race, or years of education. Those who study the condition of police cars might focus on age, mileage, make, model, and number of accidents.

A study of the characteristics of judges might explore their age, race, gender, political affiliation, district, and number of dispositions rendered but not their marital status, height, weight, or hair color. When conducting a study, researchers only focus on variables most pertinent to their investigation. They also often explore the relationships between variables. For example, a researcher might be interested in whether judges' stated political affiliations are related to the proportion of prison sentences they level for a particular type of offense in which probation is also an option. In summary, different characteristics are called variables; their values differ or vary in a group of people, objects, or events. Most police officers differ in their job satisfaction; most felony cases differ in their lengths of closing times, and so on. These measures are all variables; their values differ for most sample members in a study.

1.3.3 Variable Types

Variables can be classified in several ways, such as the range of values a variable can assume. Some variables can be measured within only a restricted range of values, whereas others can be recorded within an infinite range. These are called discrete and continuous variables, respectively. Discrete variables can take on only a finite number of values that arise from a counting process that uses whole numbers. The number of detainees in a jail is an example of a discrete variable. For instance, a jail can have 134 or 135 prisoners but not 134.5 prisoners. The number of prisoners can only be recorded as a whole number without fractions. Hence, the number of prisoners is a discrete variable.

In contrast, a continuous variable can assume an infinite number of values. An example of a constant variable is the time it takes a police dispatcher to send an ambulance to an accident scene. The time could be measured as 5.20, 20.3, 20.32 seconds, or an even more precise time measurement, such as 20.3271567803 seconds. Other examples of continuous variables are weight, height, and age.

1.3.4 Levels of Variables and Scales

Variables can be discrete or continuous and can be classified into four types based on their measurements. The first and most straightforward variable type is a nominal variable, which merely categorizes or classifies data. The second type is an ordinal variable, which can be measured by the degrees of difference among the data. The third type is an interval variable, which has the same unit of measurement throughout the entire range of data values. Finally, the fourth type is a ratio variable with an absolute zero point, indicating the complete absence of a quantity of data.

1.3.5 Nominal Scales

The values of nominal scales are "names." The measurement of nominal variables shows that only one entity is different from or similar to the others. Nominal variables (sometimes called qualitative variables)

are measured at the lowest level of measurement. One entity measured as a nominal variable has no more or less of a characteristic than another entity measured as a nominal variable. For example, a Ford is neither greater nor lesser than a Toyota as a vehicle type; they are simply different makes of automobiles. In fact, with such variables, only names can be assigned to various categories of the variable. Gender is another example of a nominal variable expressed as male or female. Although a group of police officers can be classified by gender, one officer does not have a "more" gender than another. Each is either a male or a female officer. These types of variables can only be classified into categories. They are called nominal or categorical variables and are measured at the nominal level.

The numbers or values assigned to nominal variables constitute a nominal scale. These scales are useful when data are being tabulated or grouped into categories. Referring to the gender example, "1" can be assigned to men, and "2" can be assigned to women. The numbers on a nominal scale have no mathematical properties. Thus, "2" does not suggest more gender than "1" does. These numbers are simply designations for a category name and are entirely arbitrary. Switching or modifying the numeric codes for men and women would not alter the information about gender if the numbers for men and women differed. Different numbers define membership in distinct categories.

Nominal scale data cannot be meaningfully added, subtracted, multiplied, or divided. For example, it makes no sense to divide the number for women, 2, by the number for men, 1, to compute a fraction. The numbers on a nominal scale can only be counted. For example, in a sample of 120 crime victims, 91 could be men, and 29 could be women. These counts could be converted to percentages, signifying that the group comprised 76% men and 24% women. Criminal justice agency reports contain many examples of nominal variables, such as type of crime, prison location, class of sentence, court case district, and offender gender.

1.3.6 Ordinal Scales

Some variables are measured on a scale that compares the amount or degree of a characteristic. For example, murder is a more serious crime than auto theft. The variable in this case is crime seriousness. Such variables are called quantitative. Quantitative variables carry numerical significance or weight. Important distinctions can be made among these variables, such as showing that one person (or object or event) has more of the defining characteristics of a variable than another.

Suppose a large sample of state police officers was surveyed about their job satisfaction. Some were highly satisfied, some moderately satisfied, and others slightly satisfied. This level of measurement specifies that some officers are more satisfied than others. Furthermore, the findings suggest that those who are moderately satisfied are more satisfied than those who are slightly satisfied. The moderately satisfied group is less satisfied than the highly satisfied group, and the highly satisfied group is more satisfied than all the other officers. These officers' order of job satisfaction could be ranked from highly satisfied to not at all satisfied (descending order) or from not at all satisfied to highly satisfied (ascending order). This variable has an underlying order that is called ordinal.

As illustrated in the example above, a significant limitation of ordinal variables is that they do not convey how much more or less satisfied one officer is compared with another. A highly satisfied officer is more confident than one who is only moderately happy, but precisely how much more is unquantifiable with an ordinal variable. Respondents can be ranked only in terms of "more" or "less." Such variables are measured at the ordinal level, and the numbers assigned to these variables constitute an ordinal scale.

The distinguishing feature of ordinal scales is order. The officers in the survey could be asked to circle a number that best reflects their overall job satisfaction: 0 for "not at all," 1 for "slightly," 2 for "moderately," or 3 for "highly" satisfied. The officers could then be ranked on their job satisfaction using these numbers: 3 is greater than 2, 2 is greater than 1, and 1 is greater than 0, but the values are arbitrary. For example, the values of 10, 20, and 30 could be designated as slightly, moderately, and highly satisfied, respectively. Just as easily, so could the values of 3, 9, and 12. For example, if an officer chooses 3 and another chooses 12, the second officer does not feel exactly 4 times more satisfied than the first. Therefore, the numbers of an ordinal variable cannot be meaningfully added, subtracted, multiplied, or divided.

Although ordinal variables can be assigned values, a more common measurement technique is to list the categories or classes of variables and ask respondents to order them. For example, they could be asked to rank from 1 to 5—least to most serious—the crimes of burglary, robbery, theft, embezzlement, and fraud or to rank from least desired (1) to most desired (9) strategies for lowering crime rates: more police officers, stricter laws, stiffer penalties, broken windows policing, neighborhood watch programs, more prosecutors, additional courts, improved street lighting, and new jails.

1.3.7 Interval Scales

An ordinal variable is quantitative but has limitations. A variable with greater measurement precision than an ordinal variable is known as an interval. Temperature is an example of this type of variable. For instance, if the temperature in one cellblock of a prison is 80 °F and the temperature in another cellblock is 75 °F, the difference is 5 °F. Similarly, if the cellblock temperatures rise to 100 °F from 95 °F, the difference is also 5 °F. In each case, the difference reflects the same magnitude of temperature change, as the thermometer units are spaced at equal intervals throughout the entire range of temperatures. For example, the number of degrees of heat it takes to raise the temperature from 60 to 61 °F is the same number of degrees it takes to raise the temperature from 150 to 151 °F.

Variables with equal intervals among their values are measured on interval scales. Interval-level variables possess a relative rather than an absolute zero point. For example, 0 °F does not mean the absence of temperature; it is merely a relative point on the temperature scale. With a relative zero point, the measurements cannot be converted into ratios. For example, 120 °F is warmer than 30 °F; the former is 90 ° higher. Nevertheless, the former is not four times warmer than the latter, as the value of zero neither marks the absolute beginning point of the scale nor represents the absence of temperature.

1.3.8 Ratio Scales

Variables with equal measurement intervals and a true zero point are measured at the ratio level, and the numbers assigned to these variables constitute a ratio scale. Income is an example of a variable measured at the ratio level. For instance, offender X's income from drug sales is $100,000, and offender Y's is $10,000. Offender X's income is not only $90,000 higher but also 10 times greater than offender Y's income. Income has an absolute zero point; thus, ratios can be constructed from such data.

1.3.9 Likert Scales

A common way to measure attitudinal variables in social science research (e.g., criminology, psychology, or sociology) is the Likert scale, which usually consists of a statement or question and a rating scale that

is divided into equal increments from an extreme negative to an extreme positive opinion, such as the following:

"Capital punishment is a deterrent to violent crime."

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

A participant who chooses a rating of "strongly agree" clearly feels more deeply about the deterrent effects of the death penalty than one who selects a rating of "disagree." Numbers can replace or be coupled with verbal labels, such as listing 1 to 5, in which "1" would be "strongly disagree" and "5" would be "strongly agree." Likert scales do not produce ratio data, as no true zero point exists. Some researchers argue that Likert scales produce interval data, while others state that they produce only ordinal data.

The central issue is whether the scale increments are equally spaced; they appear to be, primarily if numbers represent the different points on the scale. At a closer glance, the psychological properties measured with these scales are at the ordinal level. If two respondents rate an item a "5," the researcher can conclude that they both "strongly agree" with the statement but have no unmistakable evidence that these respondents hold similar degrees of disagreement. They might use the same verbal label to represent different levels of feeling. As mentioned above, replacing or pairing the verbal label with a number from 1 to 5 makes these data appear to be measured on an interval scale because the numbers are equally spaced. Nonetheless, researchers cannot ascertain whether the psychological properties underlying the responses are similarly spaced.

Differences between the different points on the scale might not be equivalent, but they must be so for the measures to be regarded as a product of an interval scale. Is the difference between the ratings of "agree" and "strongly agree" really the same as the difference between the ratings of "neither agree nor disagree" and "agree?" Likert scales only place respondents in an order of magnitude of attitudinal or other differences based on their responses to a survey item, which is similar to an ordinal scale. The preciseness of interval scales remains debatable, but researchers often treat Likert scales as interval-level data and perform parametric tests on these data.

1.3.10 The Hierarchy of Measurement Scales

Criminal justice researchers and practitioners describe people, objects, or events using variables measured at all four levels: nominal, ordinal, interval, or ratio. For example, offenders (people) can be studied based on their gender, the seriousness of their offense, their satisfaction with the court system, or the number of times they have been arrested or convicted. Gender is a nominal variable, the seriousness of the offense is ordinal, satisfaction is interval, and the number of prior convictions is ratio. Jails (an object) can be studied based on their location (city, county, north, south, east, west) (a nominal variable), their level of violence (an ordinal variable), or their average daily population (a ratio variable). Bank robberies (an event) can be studied based on the institution robbed location, e.g. city, county, north, south, east, west (a nominal variable), the degree of threat involved in a robbery (an ordinal variable), or the number of dollars stolen in the incident (a ratio variable).

Similarly, the variable can be measured at distinct levels. For example, illegal drug use can be measured by asking respondents to identify how many times they used illicit drugs in the past month (a ratio variable) or to answer the same question by checking none, 1–4, 5–9, or 10 or more times (an ordinal variable). It can also be measured by asking them, "Did you use any illegal drugs in the past month?" with the response categories "yes" or "no," thereby creating a nominal variable.

In general, variables should be measured at the highest possible level of measurement. The more information available, the more ways cases can be compared. A broader range of statistical analysis can be performed with quantitative rather than qualitative variables. Ordinal- or nominal-level variables can be created from interval- or ratio-level variables, but the scales cannot shift in the opposite direction. Researchers can later combine the ages into categories by collecting data on the actual ages of crime victims. In contrast, if researchers ask respondents to check an age category (e.g., young adult, middle-aged, or senior citizen), they cannot later modify that variable to capture the respondents' actual ages.

The data's level of measurement dictates the calculations that can be done to analyze and present the data. The level of measurement also determines the choice of various statistical procedures (i.e., statistical tests) that can be performed on the data. For example, in a study of the relationship between mental illness and crime, the sample could consist of four racial groups. Suppose an investigator assigns a value of "1" to African Americans, "2" to Whites, "3" to Latinos, and "4" to Asians. Based on the sample data, the investigator then adds the assigned values, divides them by the number of participants, and reports that the mean race/ethnicity is 2.56. Does this suggest that the average race/ethnicity is between Whites and Latinos? Of course not! In this instance, the level of measurement has been improperly used.

Table 1.1 lists each scale with its distinguishing features. As shown, nominal, ordinal, interval, and ratio scales can be arrayed into a hierarchy. The scales are ordered from top to bottom concerning their increasing sophistication. The properties of a scale consist of its distinguishing features and all the features of the preceding scales in the measurement hierarchy. For example, a ratio scale has a true zero-point, equal intervals, order, and categories.

As mentioned above, in the hierarchy of measurement scales, a variable can be expressed in terms of a lower but not a higher scale. For example, the measurements of an interval scale can be expressed in terms of an ordinal or a nominal scale but not a ratio scale. Variables measured at the ordinal level can also be expressed at the nominal level. For example, public opinions about the death penalty can be

Table 1.1 The Four Scales of Measurement with Their Distinguishing Features.

Name	Distinguishing Scale Feature
Nominal	Names or categories
Ordinal	Order
Interval	Equal intervals
Ratio	True zero point

measured on an ordinal scale in which 1 = strongly supports the death penalty, 2 = moderately supports the death penalty, 3 = slightly supports the death penalty, and 4 = does not at all support the death penalty. In addition, people's opinions can be classified on a nominal scale in which 1 = supports the death penalty and 2 = does not support the death penalty.

Variables measured at the nominal level can be expressed only at the nominal level. For example, the race of prisoners in a state prison can be classified as 1 = White, 2 = African American, 3 = Latino, and 4 = Other. This nominal scale can be expressed as only another nominal scale, such as 1 = White and 2 = Not White.

An example of accurate data can help further illuminate the nature of measurement scales—Table 1.2 presents information from the *Sourcebook of Criminal Justice Statistics*. The table contains the 2016 prison populations in 11 Midwestern states. The variable displayed is the inmate population, expressed as a frequency or headcount and measured on a ratio scale. It has an absolute zero point, and the increments are equal among the scale units throughout the range. This variable can also be measured on an ordinal scale. For example, a prison administrator might be interested in studying population levels rather than individual headcounts. In that case, the population could be designated as 1 = high (25,000 or more inmates), 2 = medium (24,999–10,000 inmates), or 3 = low (fewer than 10,000 inmates) and therefore measured on an ordinal scale. Inmate populations could also be measured on a nominal scale as 1 = consistent or 2 = inconsistent with previous trends. Table 1.3 shows the values of the ordinal and nominal scales and the original information about prison populations.

Table 1.2 Presents a sample of prison populations of Midwestern states.

2016 Prison Populations for 11 Midwestern states	
State	**Population**
Illinois	60,800
Indiana	43,200
Iowa	13,400
Kansas	17,200
Minnesota	16,300
Missouri	44,300
Nebraska	8,800
North Dakota	3,100
Ohio	71,000
South Dakota	5,800
Wisconsin	36,500

Source: Adapted from Bureau of Justice Statistics (www.bjs.gov/content/pub/pdf/cpus16.pdf).

Table 1.3 The Data from Table 1.2 Expressed in Terms of Ratio, Ordinal, and Nominal Scales.

State	Population (Ratio)	High/Medium/Low (Ordinal)	Consistent/Non-Consistent (Nominal)
Illinois	60,800	1	2
Indiana	43,200	2	2
Iowa	13,400	3	2
Kansas	17,200	3	2
Minnesota	16,300	3	1
Missouri	44,300	2	1
Nebraska	8,800	3	1
North Dakota	3,100	3	1
Ohio	71,000	1	2
South Dakota	5,100	3	1
Wisconsin	36,500	2	1

1.4 Dependent and Independent Variables

As discussed in this chapter, variables are properties or characteristics of events, objects, or persons that can take on different values (counts or amounts). When social scientists conduct research, they often manipulate variables. In an experiment, the independent variable is the variable varied by the researcher, and the dependent variable is the response measured by the researcher.

An independent variable is the hypothesized cause, whereas the dependent variable is the measured effect. In nonexperimental research, which involves no experimental manipulation, the independent variable is the variable that has a hypothesized relationship with the dependent variable. For example, in research on vaping and lung disease, vaping is the independent variable: behavior in the group under study. When researchers cannot control or manipulate an independent variable, it is technically referred to as a subject or predictor variable (gender, ethnicity, poverty, etc.). Although researchers can never control or manipulate predictor variables, they often treat them as independent variables.

As stated, the dependent variable refers to the effect or outcome in which the researcher is interested (i.e., lung disease because of vaping); the independent variable refers to the presumed cause of the outcome (i.e., vaping). Changes in the independent variable can lead to changes in the dependent variable. Any event or condition can be conceptualized as an independent or a dependent variable. For example, drug use is a hypothesized cause of criminal behavior, but it has also been observed that criminal offending can cause drug use. Hence, drug use is a variable that can be a cause (independent variable) or an effect (dependent variable).

1.5 Practical Application of Statistics

Criminal justice professionals in law enforcement, the courts, and corrections deal with many tasks and challenges that can be tackled with statistics. For example, statistics can assist them in describing the crime in a city, the composition of a police department, overcrowding in a county jail, or the case-processing rate in a criminal court. Professionals can address these issues with basic statistical concepts and procedures.

Counting, classifying, comparing, and organizing information can help present findings to various audiences. Doing so involves the use of statistics. Other problems in criminal justice go beyond simple descriptions and into the areas of explanation and prediction. For example, suppose a police department added 60 officers to its force. What would be the expected impact on relationships with minority residents, crime rates, staff morale, or gas consumption of patrol cars? How many new probation officer positions would be required in the next five years to keep pace with caseload growth or decline? If the number of sentences to prison continues to increase or decrease at its present rate, how many new prison beds must be purchased by the state in five years? Is probation more effective than prison in terms of recidivism? Which community-based corrections programs work better than others? How successful are community policing programs, judicial training, or the public defender's office in achieving their stated goals? Answering these questions involves the statistical procedures presented in this text.

1.6 Introduction to SAS

Statistical Analysis System (SAS) was one of the first software packages developed to analyze data. Today, it is used in many areas of research and industry to generate statistical output. A SAS program has two basic stages: the data step and the procedure or PROC step. The data step allows the user to input data either by hand or from a file and manipulate the data. Once the data step is compiled or run, SAS generates a file called the dataset. The procedures or PROCS are pre-written routines that generate reports, graphs, tables, etc., about the dataset. The programmer can control the output and which variables are of interest through the choice of procedure and statements within the procedure. In the field of criminal justice, SAS can help make better decisions for the safety of citizens, collect, manage, and analyze data efficiently for law enforcement and public safety, gather and manage intelligence, and conduct investigations more effectively.

The SAS code will be highlighted in a boxed area in this text. The examples will demonstrate how to write the code for the examples found in the accompanying chapter. In the first few chapters, the examples will show how to write the data step and the PROC for the desired output. In later chapters, only the necessary process will be shown without the data step unless there is a unique way to input the data. There are several valuable resources for SAS programming. SAS maintains a support website (support.sas.com).

1.7 Summary

Chapter 1 discusses several fundamental concepts that are important in gaining a basic understanding of the field of statistics and introduces subjects covered in subsequent chapters of this textbook. Key terms, such as variable, measurement levels, measurement scales, population, element, census, sample, and descriptive and inferential statistics, were also defined and discussed. These terms will be used throughout this textbook, and their meanings will become more apparent in subsequent problems and exercises. Chapter 2 presents techniques for organizing and displaying data and further illuminates the concepts of variables, measurement scales, and descriptive statistics.

1.8 Exercises

Exercise 1.1 Understanding Terms

Write a brief definition of each of the terms from memory only. Check your accuracy by returning to the relevant sections of the chapter for a refresher.

Statistics: _____
Parameter: _____
Descriptive statistics: _____
Inferential statistics: _____
Population: _____
Sample: _____
Independent variable: _____
Dependent variable: _____
Nominal data: _____
Ordinal data: _____
Interval data: _____
Ratio data: _____

Exercise 1.2 Identification of Variables

Identify five variables for each of the following studies.

1) Male inmates in prisons in the United Kingdom
2) A comparison of criminal justice students in America and Canada
3) Auto theft in the United States
4) Mass shootings in America in any given year
5) Shootings and concealed or open carry laws by state
6) Prison admissions and readmissions
7) A comparison of men and women detainees in urban jails

Exercise 1.3 Recognizing Types of Variables

For the following data sets, identify each variable as nominal, ordinal, interval, or ratio.

Inmate Data

Case Number	Age	Race	Education (Grade)	Previous Commitments	Offense	Years of Sentence	IQ Score
1	22	African American	5	0	Rape	6	90
2	26	African American	6	0	Rape	14	110
3	59	White	14	3	Robbery	25	135
4	27	African American	12	2	Robbery	13	125
5	25	Latino	10	0	Burglary	3	130

Self-reported drug use among college students

Case Number	Gender	Age	Ethnicity Ethnicity	Year in College	Smoke Marijuana	Number of Times per Week
1	Male	18	Caucasian	First-year students	Yes	Over 10
2	Female	20	Caucasian	Junior	Yes	1–5
3	Female	19	African American	Sophomore	Yes	6–10
		19				
4	Male	19	Hispanic	Sophomore	Yes	Over 10
5	Male	21	African American	Senior	Yes	1–5
6	Female	20	Hispanic	Junior	Yes	6–10

Assaults against police officers

Assignment	Circumstance	Weapon Used	Injury Status	Days Off
Patrol	Traffic stop	Handgun	Serious	45
TACT	Drug raid	Shotgun	Slight	5
CID	Homicide investigation	Handgun	Serious	30
Patrol	Domestic disturbance	Knife	None	0
Patrol	Traffic stop	Knife	Moderate	11

Exercise 1.4 Defining Terms (Variables and Constants)

Indicate whether each of the following measures is a variable or a constant.

1) The rated capacity of jails
2) The number of jurors for a felony trial
3) The daily census of a state prison
4) The number of hate crimes filed each year
5) The number of daily meals served to inmates in a state's prison system
6) The number of days in police lock-up
7) Articles in the Bill of Rights

Exercise 1.5 Converting Scales

Using the data sets provided in Exercise 1.2, on pages 1–15, convert the ratio scales from each data set into nominal and ordinal scales. For age, create one set of nominal and ordinal scales that could be applied to both data sets in which age is a variable.

Exercise 1.6 Which Is It? Population, Element, Sample, or Census

_____ All correctional officers in the United Kingdom
_____ Felony courts in X county
_____ One District Attorney
_____ 100 Federal prisons
_____ One maximum security prison
_____ Number of and characteristics of Solicitors in the United Kingdom
_____ Police departments with sworn personnel under 250
_____ Women prison inmates
_____ Characteristics of all Parolees 2020–2024
_____ 500 Secret Service Agents

(NOTE: Due to their nature, each item could be viewed as more than one of the four. Their classification is based on the study's nature.)

1.9 Answers to Exercises

Exercise 1.1 Understanding Terms

Statistics: A numerical quantity to describe a characteristic of a sample.
Parameter: A numerical quantity to describe a characteristic of a population.
Descriptive Statistics: Branch of statistics for summarizing or describing the attributes of a data set.
Inferential Statistics: Drawing inferences or making predictions about a population based on data from a sample.
Population: The total or complete set of cases of interest.
Sample: Any subgroups of cases of interest drawn from a particular population.

Independent Variable: In a study, the presumed cause of an outcome being measured.
Dependent Variable: In a study, the presumed outcome that is affected by changes in the presumed cause or independent variable being measured or manipulated.
Nominal Data: Data that are based on mutually exclusive and exhaustive categories and expressed in words, not numbers.
Ordinal Data: Data with separate categories that place cases (such as people) into rank order.
Interval Data: Data on a continuum that are rank-ordered and separated by equal intervals among units.
Ratio Data: Data on a continuum that are rank-ordered, separated by equal intervals among units, and have an absolute zero point.

Exercise 1.2 Identification of Variables

1) A Study of Male Inmates in Prisons in the United Kingdom: Age, security level, length of sentence, nationality, convicting offense, number of prior incarcerations, etc.
2) A Comparative Study of Criminal Justice Students in America and Canada: Gender, career goals, previous experience, ethnicity, age, grade point average, number of courses required, etc.
3) A Study of Auto Theft in the United States: Number of auto thefts in each state, type of auto stolen, average mileage, age of auto, location of theft, recovery rates, etc.
4) A Study of Mass Shootings in America in Any Given Year: Venue of shooting, perpetrator characteristics, number of victims, characteristics of victims, etc.
5) Research on Prison Readmissions and Recidivism: Time to readmission, reasons for readmission, age, race, prison time before readmissions, etc.
6) A Comparison of Men and Women Detainees in Urban Jails: Age, charges, race, previous arrests, admission status, etc.

Exercise 1.3 Recognizing Types of Variables

1) Inmate Data: Age, ratio; race, nominal; education, interval; previous commitments, ratio; offense, nominal; years of sentence, ratio; I.Q., interval.
2) Self-Reported Drug Use: Gender, nominal; age, ratio; ethnicity, nominal; year in college, ordinal; smoke marijuana, nominal; number of times per week, ratio.
3) Assaults on Police Officers: Assignment, nominal; circumstances, nominal; weapon used; nominal; injury status, ordinal; days off, ratio.

Exercise 1.4 Defining Terms: Variables and Constants

1) Variable
2) Constant
3) Variable
4) Variable
5) Constant
6) Variable
7) Constant

Exercise 1.5 Converting Scales

Age as a Nominal Variable: Young, not young
Age as an Ordinal Variable: Child, teenager, young adult, middle-aged, senior
Previous Commitments as a Nominal Variable: Yes, no
Previous Commitments as an Ordinal Variable: None, fewer than 5, moderate number (6–10); excessive (11+)
Years in Sentence as a Nominal Variable: Sentence, fair, lenient
Years in Sentence as an Ordinal Variable: One year or less; more than one year but less than life; life in prison.

Exercise 1.6 Identification of Data Sets

All correctional officers = population; Felony courts in X county = sample or population; one district attorney = element; 100 federal prisoners = sample; one maximum security prison = element; number and characteristics of Solicitors in the United Kingdom = census; police departments with fewer than 250 sworn personnel = sample or population; women prison inmates = population, characteristics of all parolees 1900–2003 = population or census; 500 MI5 agents = population or sample.

References

Federal Bureau of Investigation. (1990). *Crime in the United States*. Washington, DC: Author.
Hinkle, D.E., Wiersma, W., and Jurs, S.G. (1982). *Basic behavioral statistics*. Boston: Houghton Mifflin Company.
Leedy, P.D. (1985). *Practical research*, 3rd edition. New York: Macmillan Publishing Company.
United States Department of Justice. (2016). *Sourcebook of criminal justice statistics*. Washington, DC: Author.