INTRODUCTION

AND OVERVIEW

LEARNING OBJECTIVES

- Understand the purpose of the book and the structure of the book.

- Review independent, dependent, and extraneous variables and their scales of measurement.

- Review measures of central tendency and variability.

- Review visual representations of data, including the normal distribution.

- Review descriptive and inferential statistical applications of the normal distribution.

The purpose of this book is to provide a hands-on approach for students to understand and apply procedural steps in completing quantitative studies. The book emphasizes a step-by-step guide using research examples for students to move through the hypothesis-testing process for commonly used statistical procedures and research methods. Statistical and research designs are integrated as they are applied to the examples. The structure of each chapter covers the following nine quantitative research procedural steps:

1. A description of a research problem, taking the student through identifying research questions and hypotheses.

2. A method of identifying, classifying, and operationally defining the study variables.

3. A discussion of appropriate research designs.

4. A procedure for conducting an a priori power analysis.

5. A discussion of choosing an appropriate statistic for the problem.

6. A statistical analysis of a data set.

7. A process for conducting data screening and analyses (IBM SPSS) to test null hypotheses.

8. A discussion of interpretation of the statistics.

9. A method of writing the results related to the problem.

The underlying philosophy of the book is to view the quantitative research process from a more holistic and sequential perspective. Concepts are discussed as they are applied during the procedural steps. It is hoped that after completion of the book readers will be better able to plan research and conduct statistical analyses using several commonly used statistical and research designs. The quantitative methodological tools learned by students can actually be applied to their own research, hopefully with less oversight by faculty.

The use of statistical software is an essential tool of researchers. Psychological, educational, social, and behavioral areas of research typically have multifactor or multivariate explanations. Statistical software provides a researcher with sophisticated techniques to analyze the effects and relationships among

many independent variables (factors) and dependent variables (variates) in various combinations all at once and instantly. We will use IBM SPSS statistical software, which has been developed over many decades and is one of the most widely used statistics programs in the world.

Statistical techniques may have more meaning, understandability, and relevance when learned within the context of research. One needs to have an understanding of statistical analyses to consume and construct professional research competently. Knowledge of quantitative research methods is especially important today because of the emphasis on evidence-based practice in psychology (EBPP) to improve clinical work with clients. EBPP refers to using the best available research with clinical expertise in the context of patient characteristics, culture, and preferences (American Psychological Association, 2006).

Ideally, the goal is to help a student achieve self-efficacy in understanding, planning, and conducting actual independent research. Information and skills grow, leading to advanced understanding. We next present a review of foundational information related to research and statistics that will be useful to review prior to completing the chapters that follow.

REVIEW OF FOUNDATIONAL RESEARCH CONCEPTS

A review of foundational concepts related to research and statistics is presented next. Quantitative research involves the interplay among variables after they have been operationalized, allowing a researcher to measure study outcomes. Essential statistical methods used to assess scores of variables include central tendency, variability, and the characteristics of the normal distribution.

Independent, Dependent, and Extraneous Variables

At the core of quantitative research is studying and measuring how variables change. Kerlinger and Pedhazur (1973) stated, "It can be asserted that all the scientist has to work with is variance. If variables do not vary, if they do not have variance, the scientist cannot do his work" (p. 3). Even the father of modern statistics, Sir Ronald Fisher (1973), said, "Yet, from the modern point of view, the study of the causes of variation of any variable phenomenon, from the yield of wheat to the intellect of man, should be begun by the examination and measurement of the variation which presents itself" (p. 3).

An *independent variable (IV)* in a study is the presumed cause variable. In experimental research, the IV is designed and employed to influence some other variable. It is an antecedent condition to an observed resultant behavior. Changes in the independent variable produce changes in the dependent variable.

All variables need to be able to vary. Kerlinger and Lee (2000) identified two types of independent variables: active and attribute. An *active independent variable* is one that is manipulated by the researcher. For example, a researcher designs a study with an IV that has a researcher-specified treatment condition compared to a no-treatment control condition. Other terms used for an active IV are stimulus variable, treatment variable, experimental variable, intervention variable, and *X* variable.

A second type of IV is called an *attribute independent variable*, which is not manipulated but is ready-made or has preexisting values such as gender, age, or ethnocultural grouping. Other terms used are organismic or personological variables.

The terms *classification variable* and *categorical variable* are often used as an IV label. They can be used as either active or attribute types. For example, a manipulated IV that has a treatment condition and a control condition could be called a classification variable. Also, an attribute variable such as gender (male or female) may be referred to as a classification or categorical variable.

A *dependent variable (DV)* is the presumed resulting outcome in research. It is usually observed and measured in response to an IV. We look for changes in a DV caused by an IV. A DV is also referred to as a response variable or a *Y* variable.

An *extraneous variable (EV)* is an unwanted and contaminating variable. An EV acts on a dependent variable like an independent variable does but in a confounding way that confuses an understanding of how the IV is changing the DV. An extraneous variable is undesired noise in a research study. A researcher wants to control an extraneous variable to neutralize its effects.

Variables need to be assigned meaning by specifying activities or operations necessary to measure the variable, which is known as an *operational definition (OD)*. A comprehensive operational definition entails all of the activities and operations that define the variable. For example, an active IV psychotherapy approach might have two conditions (Gestalt therapy and control condition). We can say there are two operational definitions for the IV psychotherapy approach for the sake of brevity. However, each condition has a detailed, comprehensive operational definition that is clearly and fully specified. A brief operational definition of a dependent variable of depressive symptomatology may be scores on

the Beck Depression Inventory (BDI). However, the comprehensive OD would detail key psychometric research used in validating the BDI.

In correlations research, an independent variable is often called a *predictor variable (PV)*, and a dependent variable is called a *criterion variable (CV)*.

Scales of Measurement of Variables

Variables can be assigned scales of measurement. A variable does not have an absolute scale of measurement. The scale of measurement of a variable can change depending on how the variable is being used in different studies and even within the same study. Therefore, there is a research contextual consideration that helps determine the scale of measurement of a variable. The process of thinking through the connection between scales of measurement and variables helps the researcher more clearly see how variables can be measured in a study. Also, the scales of measurement assigned to a variable can be useful in selecting appropriate statistics to use in research.

There are two general classifications of scales of measurement, each having two subcategories; they are *discrete scale* (nominal and ordinal) and *continuous scale* (interval and ratio).

A variable using a *discrete-nominal scale* of measurement has mutually exclusive categories. For example, gender has mutually exclusive categories of male or female, and political affiliation has categories of Republicans, Democrats, or independents. A *discrete-ordinal scale* of measurement variable has ordering along some continuum. It is rank scaled. For example, the order (first, second, third, etc.) in which runners complete a race reflects an ordinal scale.

A *continuous-interval scale* of measurement variable has numerical distances on a scale that are considered approximately equal numerical distances of the attribute being measured. There is no true zero point on the scale; it is considered arbitrary. For example, scores on the Wechsler IQ test are considered interval scaled, and there is an arbitrary zero, but the test does not measure a total absence (true zero) of intelligence. A *continuous-ratio scale* of measurement variable has a true zero point, and the numerical distances on the scale are equal to the attribute being measured. Weight is an example of a ratio-scaled variable. A zero number of pounds is meaningful, and 100 pounds is one-third as heavy as 300 pounds. Other examples of ratio-scaled variables include height, length, and time.

There are times when ordinal-scaled variables such as Likert-type scales are statistically analyzed as a continuous-interval variable (Tabachnick & Fidell,

2007). Howell (2010) states, "We do our best to ensure that our measures relate as closely as possible to what we want to measure, but our results are ultimately only the numbers we obtain and our faith in the relationship between those numbers and the underlying objects or events" (p. 8). A more important gauge of understanding the meaning of scores on dependent variables in a study has to do with their distributions. Measures of central tendency and variability of scores of distributions are discussed next.

REVIEW OF FOUNDATIONAL STATISTICAL INFORMATION

One of the most important tasks of quantitative researchers is to understand the data they are working with. Researchers need to assess their data for issues including dishonest data, cases with atypical scores, and noncompliance with appropriate use of statistical requirements. Also, it is important for researchers to understand the uniqueness of their data sets by examining typical scores, variability among scores, and characteristics and shapes of distributions of scores related to variables in a data set.

Measures of Central Tendency

Measures of central tendency are values that represent typical scores in a distribution or set of scores. We will be using the data in Table 1.1 to demonstrate how to calculate the three most common measures of central tendency: mode, median, and mean.

TABLE 1.1   Values Used to Illustrate Measures of Central Tendency and Variability

$X(N = 6)$	X^2	$(X - \overline{X})$	$(X - \overline{X})^2$
48	2,304	$48 - 49 = -1$	1
43	1,849	$43 - 49 = -6$	36
52	2,704	$52 - 49 = 3$	9
50	2,500	$50 - 49 = 1$	1
48	2,304	$48 - 49 = -1$	1
53	2,809	$53 - 49 = 4$	16
$\Sigma X = 294$	$\Sigma X^2 = 14{,}470$	$\Sigma(X - \overline{X}) = 0$	$\Sigma(X - \overline{X})^2 = 64$
$(\Sigma X)^2 = (294)^2 = 86{,}436$			

Mode (M_o)

The *mode (M_o)* is the score that occurs most often in a set of scores. The mode is the highest point on a graph such as a frequency distribution or a histogram and is referred to as *unimodal*. If there are two scores in a sample that are equally the most frequently occurring, then the distribution is called *bimodal*. The column headed by X (individual score) and $N = 6$ represents six individual scores in the distribution of scores example. The only score that is represented more than once is 48. Thus, the $M_o = 48$ and it is a unimodal distribution.

Median (M_{dn})

The *median (M_{dn})* is a value in the set of which 50 percent of cases fall below and 50 percent above. If the number of a set of ordered scores from low to high is odd, then the score that has half of the other scores below it and half above it is the median. For example, in the set of numbers 4, 5, 7, 8, and 9, the number 7 is the median.

In the set in Table 1.1, the number of scores is even. The six scores ordered are 43, 48, 48, 50, 52, 53. To obtain the median requires calculating the average value between the score at $N/2$ and the score at $(N/2) + 1$. So, $N/2 = 6/2 = 3$ (i.e., the third score) and $(N/2) + 1 = 3 + 1 = 4$ (i.e., the fourth score). The third score in the set is 48 and the fourth score is 50. The average of 48 and 50 is $(48 + 50)/2 = 98/2 = 49$. So, the median of the data set in Table 1.1 is 49.

Mean ($\overline{X}$ or M)

The *mean ($\overline{X}$ or M)* is the sum of individual scores (ΣX) in a data set divided by the number of scores (N). The mean is typically a more precise measure of central tendency than the mode or the median, because the specific value of each score is used to calculate the mean. Also, the mean has the properties of being continuously scaled as an interval or a ratio. The mean is more stable than the mode or median when it is used as a sample measure of central tendency drawn from a population. On the downside, when a sample has extreme scores (skewed), the mean is drawn way from the clustered scores toward the extreme scores. In these situations, the mean may not be the most typical score in a data set. For example, an analysis of salaries of company employees that includes the high salaries of

executives can produce a company employee mean salary that the vast majority of employees are well below. The median might be a better indication of the typical employee salary in the company. Presented next is the calculation of the mean of our sample data in Table 1.1.

$$\overline{X} = \frac{\Sigma X}{N}$$
$$= \frac{294}{6}$$
$$\overline{X} = 49.00$$

where

$\Sigma X =$ sum of the individual scores

$N =$ number of scores

Measures of Variability (Dispersion) of Scores

Foundational to quantitative research is the study of the *measures of variability (dispersion) of scores* in a sample data set. Here we review common measures of range, mean deviation scores, sum of squares, variance, and standard deviation using the data set from Table 1.1.

The *range* of scores in a data set is simply the difference between the highest and lowest scores ($\text{score}_{highest} - \text{score}_{lowest}$). In the example data in Table 1.1, the highest score is 53 and the lowest score is 43; $53 - 43 = 10$, which is the range.

A measure of variability that is used as a component in many statistical formulas that are used in fundamental statistics is called the *total sum of squares*, which is the sum of squared differences of all scores in the data set from their overall mean, $\Sigma(X - \overline{X})^2$.

Variance of the Sample (s^2)

The variance of the sample (s^2) is the *total sum of squares* divided by the number of scores, $\frac{\Sigma(X - \overline{X})^2}{N}$. The symbol of the variance of the population is sigma squared (σ^2). If all the cases are the same value, the variance will equal zero. The larger the variance value, the more the values are spread out in the distribution.

We will be using $N-1$ as the denominator rather than N since the primary focus of this book is to use samples to estimate populations. The use of $N-1$ is a more accurate estimate (*unbiased estimate*) of a population parameter. We do not know what the population mean is, so the sample values have less variability. We use the sample mean instead and with $N-1$ as a compensation for not knowing the population mean. This use of $N-1$ also is referred to as *degrees of freedom (df)*, a practice used in most statistical analyses. As Hays (1963) states, "Thus we say that there are $N-1$ degrees of freedom for a sample variance, reflecting the fact that only $N-1$ deviations are 'free' to be any number, but that given these free values, the last deviation is completely determined" (p. 311). If the goal of studying a sample is to describe the sample and not to estimate a population, then N may be used and not $N-1$. The variance is calculated next using the data from Table 1.1.

$$s^2 = \frac{\Sigma(X - \overline{X})^2}{N-1}$$
$$= \frac{64}{5}$$
$$s^2 = 12.80$$

Standard Deviation of the Sample (*s*)

The *standard deviation of the sample (s)* is the square root of the variance. The symbol for the standard deviation of the population is σ. The standard deviation is a more useful explanatory measure of variability when compared to the variance because it is in the same units as the original data. For example, when presenting the mean and the standard deviation together, they are both in the same metric.

$$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N-1}}$$
$$= \sqrt{\frac{64}{5}}$$
$$= \sqrt{12.80}$$
$$s = 3.58$$

Coefficient of Variation (*C*)

The *coefficient of variation (C)* is a ratio of the standard deviation divided by the mean. The coefficient of variation can be used to compare the variability of different variables as well as the means of the variables. Using the $s = 3.58$ and $\overline{X} = 49$ from the data in Table 1.1, the coefficient of variation is calculated here as a ratio and as a percentage:

$$C = \frac{s}{\overline{X}}$$
$$= \frac{3.58}{49}$$
$$C = .073$$
$$C \times 100 = 7.30\%$$

The coefficient of variation for the DV is 0.073, and converted to a percentage it is 7.30 percent. The standard deviation is approximately 7.30 percent of the mean.

Visual Representations of a Data Set

There are many charts available to assist researchers in more fully understanding their data. Illustrations of a bar chart, histogram, and normal Q-Q plot are presented.

A data set of scores on a dependent variable of depressive symptoms is presented in a *frequency distribution* in Table 1.2. There are scores from 145 participants and there are no missing data. Hence, the data set has complete scores and the values in the Percent and Valid Percent columns are the same. The values in the first column are the scores representing depressive symptoms. For example, the score of 17 has a frequency of nine scores ($9/145 = 6.2\%$) in the data set. The value 17 represents a cumulative percentage of 49.7 percent, which is the closest value to the 50th percentile rank of the distribution of scores.

This data can be shown as a *bar chart* (see Figure 1.1). The horizontal line of the bar chart is called the *abscissa* or *x-axis*, and in this example each number represents a value for depressive symptoms in the data set. The vertical line, also called the *ordinate* or *y-axis*, represents the frequency of scores by participants at each value for depressive symptoms in the data set. For example, you can see the most frequently scored value (16) and the least frequently scored values (5, 7, 9, 35) by participants in the sample.

TABLE 1.2   Frequency Distribution of Scores of Depressive Symptoms

Statistics		
DV		
N	Valid	145
	Missing	0

DV				
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**

|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5	1	0.7	0.7	0.7
	7	1	0.7	0.7	1.4
	8	4	2.8	2.8	4.1
	9	1	0.7	0.7	4.8
	10	7	4.8	4.8	9.7
	11	5	3.4	3.4	13.1
	12	11	7.6	7.6	20.7
	13	6	4.1	4.1	24.8
	14	7	4.8	4.8	29.7
	15	5	3.4	3.4	33.1
	16	15	10.3	10.3	43.4
	17	9	6.2	6.2	49.7
	18	13	9.0	9.0	58.6
	19	2	1.4	1.4	60.0
	20	7	4.8	4.8	64.8
	21	6	4.1	4.1	69.0
	22	4	2.8	2.8	71.7
	23	6	4.1	4.1	75.9
	24	3	2.1	2.1	77.9
	25	4	2.8	2.8	80.7
	26	4	2.8	2.8	83.4
	27	4	2.8	2.8	86.2
	28	3	2.1	2.1	88.3
	29	5	3.4	3.4	91.7
	30	2	1.4	1.4	93.1
	32	7	4.8	4.8	97.9
	33	2	1.4	1.4	99.3
	35	1	0.7	0.7	100.0
	Total	145	100.0	100.0	

FIGURE 1.1   Bar Chart of Scores of Depressive Symptoms
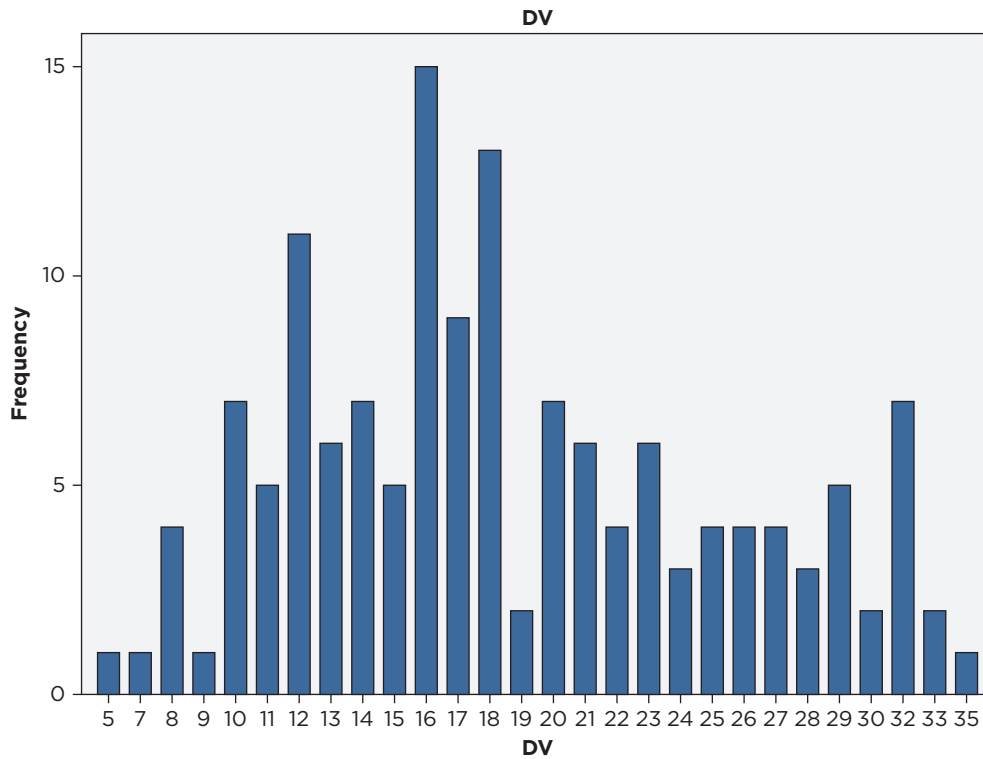
A *histogram* of the data set is presented in Figure 1.2. In a histogram the *x*-axis depicts intervals of scores and the *y*-axis still represents frequencies of scores. However, the scores on the *x*-axis are reported as falling within intervals. In this graph, the intervals are 10-point intervals, so all scores in the data set between 0 and 9 are represented by bars in this interval. Using Table 1.1 for assistance, the scores by frequencies in the data set interval 0−9 are score 5 (frequency = 1 score), score 7 (frequency = 1 score), score 8 (frequency = 4 scores), and score 9 (frequency = 1 score). Therefore, a total of seven scores in the data set are within the range of 0−9. The total numbers of scores within the intervals are interval 0−9 (seven scores), interval 10−19 (80 scores), interval 20−29 (46 scores), and interval 30−39 (12 scores).

A histogram of a data set provides the researcher with a quick visual inspection of the shape of the distribution of scores. For example, we can see

FIGURE 1.2 Histogram of Scores of Depressive Symptoms

Histogram

Mean = 18.73
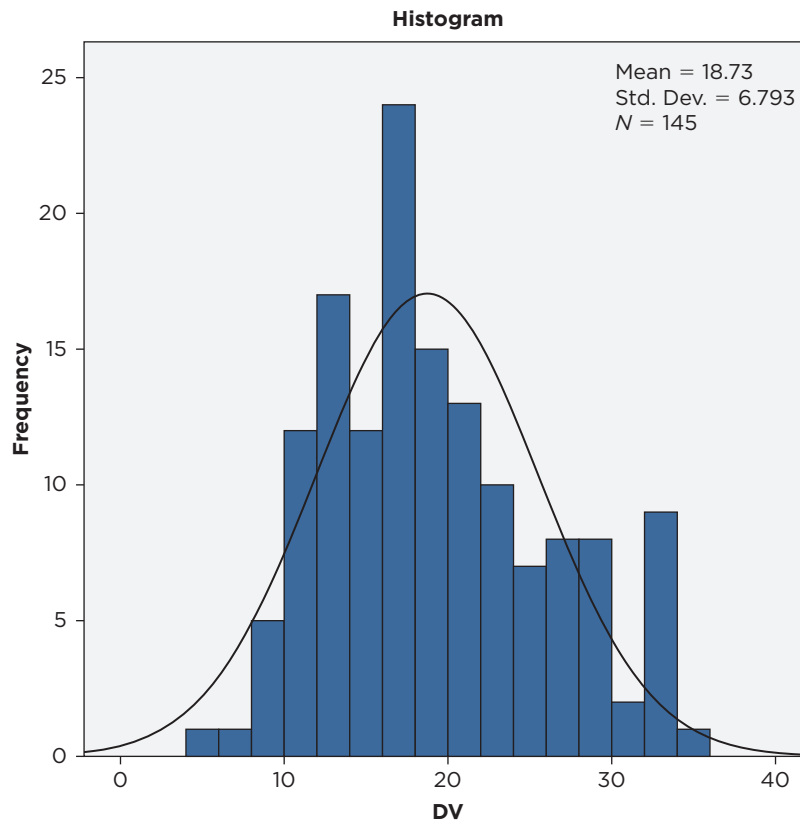Std. Dev. = 6.793
N = 145
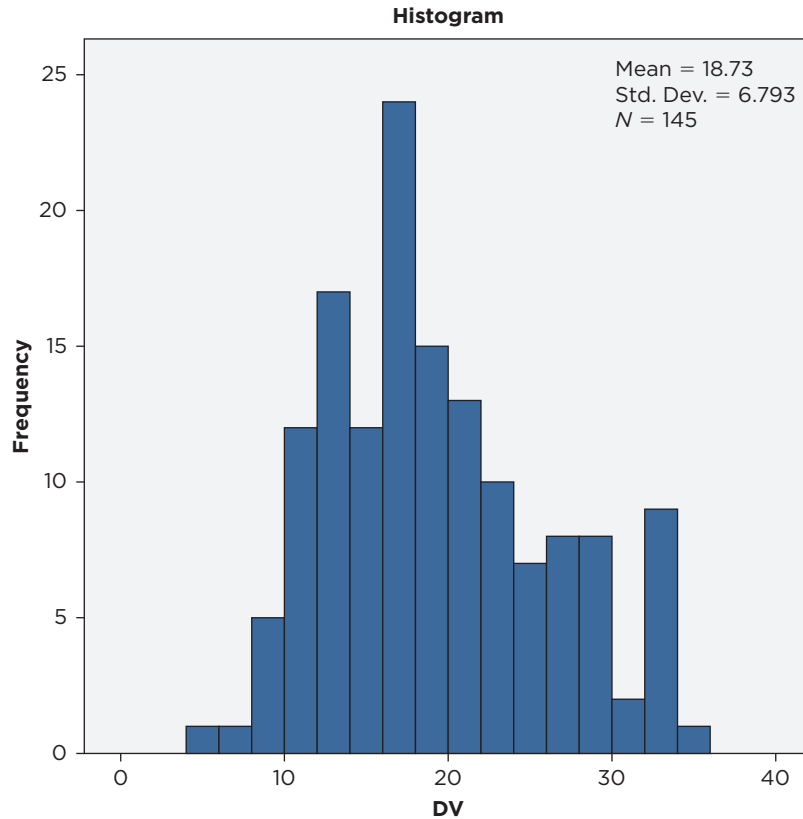
that scores cluster around the mean and there is reasonable *symmetry* (balance) of scores on either side of the mean. Also, there appear to be no extreme scores to the negative side (left) or the positive side (right) of the distribution.

The same histogram of the data set is presented in Figure 1.3 with a normal curve superimposed on the graph. This provides additional information about how well the sample distribution of scores fits a normal curve. If the data set scores were a perfect fit to the normal curve, then the bars would fit fully within the superimposed normal curve.

An example of a useful plot to assess normality of a data set is called a *Q-Q (quantile-quantile) plot* (see Figure 1.4). The Q-Q plot is derived by first subtracting each observed score from the group mean. Then, these *residuals*

Histogram

Mean = 18.73
Std. Dev. = 6.793
N = 145

(differences) are plotted against the expected observed scores if the data are from a normal distribution (Norusis, 2003). Normality exists in the sample distribution of scores if the points on the Q-Q plot fall on or near the straight line. In this example, it appears that the data set shows reasonable normality among the scores.

THE NORMAL DISTRIBUTION

The *normal distribution* (bell curve) is the most studied and widely used curve in the field of probability (Tabak, 2005). Many measurements of human activities have been shown to be normally distributed. A great deal has been discovered about the normal curve over the past 300 years since Abraham de Moivre formulated a

**FIGURE 1.4   Q-Q Plot of Scores of Depressive Symptoms**

Normal Q-Q Plot of DV
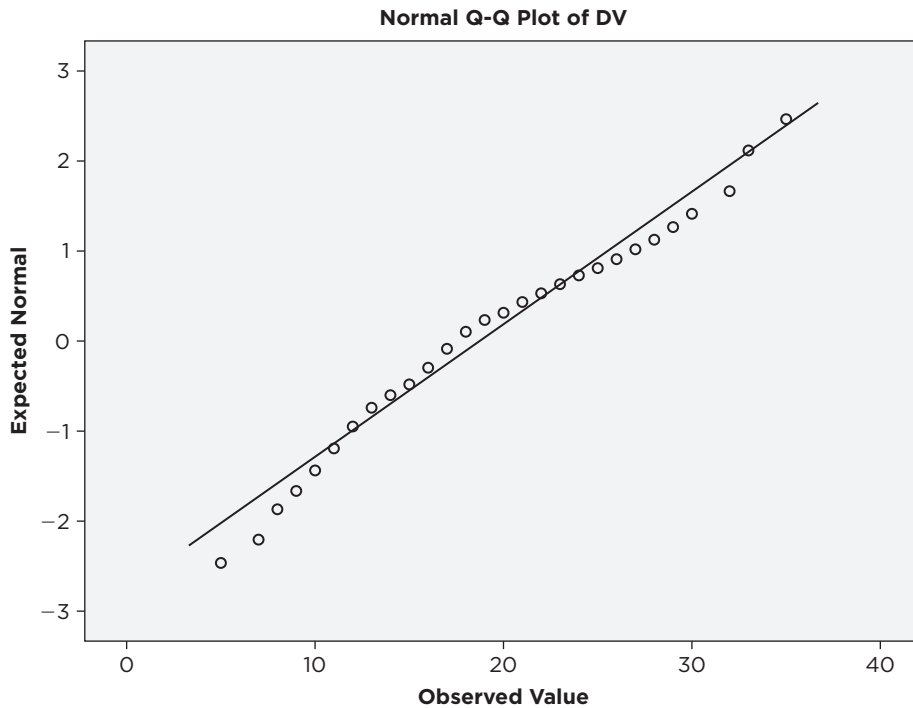
mathematical proof of the normal distribution. The normal distribution has useful properties. If two random variables have a normal distribution, their sum has a normal distribution. In general, all kinds of sums and differences of normal variables have normal distributions. So, many statistics derived from normal variates are themselves normally "distributed" (Salsburg, 2001). The normal distribution has two parameters (constants): the population mean (μ) and the population standard deviation (σ). There are many different normal curves that are based on these two parameters (Snedecor & Cochran, 1967).

Characteristics of the Normal Distribution

The normal distribution has a *peak* (highest point on the curve), *tails* (the extreme left and right points of the curve) and *shoulders* (the left and right sections of the curve between the peak and the tails) (see Figure 1.5). The right side of the

FIGURE 1.5 The Normal Distribution and Standardized Scores
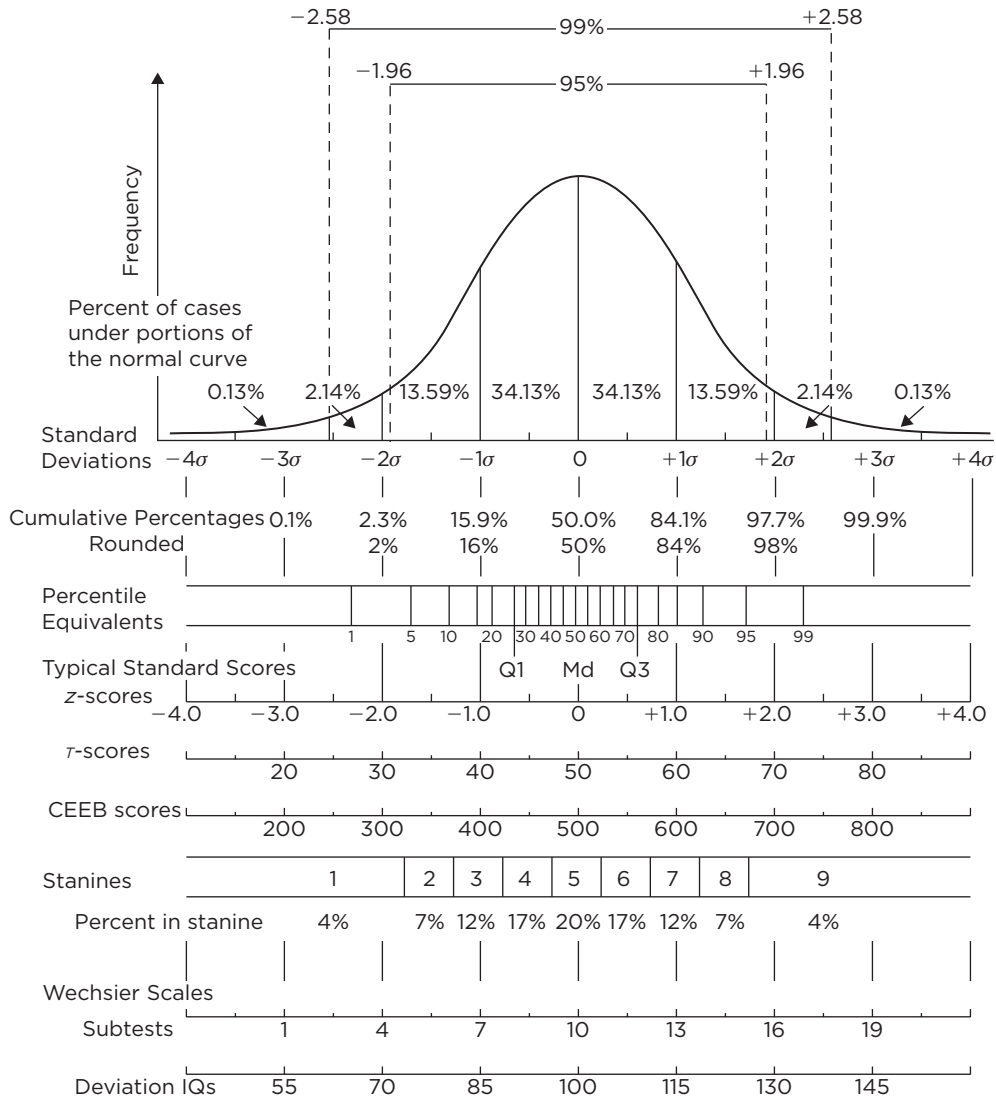
Source: H. D. Seashore, ed., "Methods of Expressing Test Scores," *Test Service Notebook* 148, January 1955. Reprinted from The Psychological Corporation, NCS Pearson, Inc.

curve is also called the *positive side* of the curve, and the left side is referred to as the *negative side*.

The base of the curve is called the abscissa (horizontal axis, *x*-axis) and sections off measurements in standard deviation units of constant percentages such as percentile ranks, *z*-scores, and *T*-scores. The height of the normal curve is called the ordinate (vertical axis, *y*-axis), which represents the percentage of cases under portions of the normal curve.

The area within the normal curve is referred to as a density of 100 percent or a unit of 1.0 for using probability. While the tails on both sides of the normal curve extend to infinity (∞) and never touch the abscissa, 99+ percent of the curve falls within ±3 standard deviations of the curve. Most of the area is in the middle of the curve at the highest point where 68.26 percent is between ±1 standard deviation of the normal curve. The percentage of the area under the curve decreases as the shape of the curve moves toward the tails.

The normal distribution is symmetrical, and each half of the curve is exactly 50 percent density. The mean, median, and mode of the normal curve are the same, as represented by 0 at the midpoint of the curve. Illustrations of using the normal curve in descriptive statistical analyses are discussed next.

Descriptive Statistical Applications of the Normal Distribution

We will assess an individual's measured IQ score compared to the IQ scores of others who are part of a normative sample of individuals whose scores reflect a normal curve. Bob has a measured IQ score on a standardized IQ test that is 80. The population mean of the normative sample is $\mu = 100$, and the standard deviation is $\sigma = 15$.

A *z-score* can be used with this information to compare Bob's score with the normative sample. A *z*-score is a standard score that shows the relative standing of a raw score in a normal distribution. The formula for a *z*-score is $z = X - \mu/\sigma$, where *X* is an individual score, μ is the population mean, and σ is the population standard deviation. The *z*-score of Bob's individual IQ score is $z = 80 - 100/15 = -1.33$. One can visualize where $z = -1.33$ is placed on the normal curve in Figure 1.5.

Next we will find the *percentile rank* of Bob's *z*-score $= -1.33$ (raw score of 80). A percentile rank is the score that indicates what percentage of persons being measured fall equal to or below the particular score. The exact percentages in the normal curve associated with *z*-scores are found using an online statistics calculator.

1. Go to www.danielsoper.com, where there are several free statistics calculators.

2. On the home website > click on *Statistics Calculator* > scroll down and click on the *Normal Distribution* button > click on the *Cumulative Area Under the Normal Curve Calculator.*

3. Type in Bob's *z*-score of −1.33 > click on the *Calculate* button > the *Cumulative area:* is .09175914.

If you take .09175914 times 100 or move the decimal two places to the right, you obtain 9.175914, which is approximately 9.18%. So a *z*-score = −1.33 indicates that equal to or less than 9.18 percent of the norm group obtain an IQ score of 80, when the mean is 100 and the standard deviation is 15. Bob's percentile rank is 9.18.

In another example, Jean scored 105 on the same IQ test. Jean's *z*-score = 105 − 100/15 = +.33. A *z*-score = +.33 is to the right of the center point mean.

1. Go to www.danielsoper.com, where there are several free statistics calculators.

2. On the home website > click on *Statistics Calculator* > scroll down and click on the *Normal Distribution* button > click on the *Cumulative Area Under the Normal Curve Calculator.*

3. Type in Jean's *z*-score of .33 > click on the *Calculate* button > the *Cumulative area:* is .62930002.

If you take .62930002 times 100 or move the decimal place two places to the right, you obtain 62.930002, which is approximately 62.93 percent. So a *z*-score = .33 indicates that equal to or less than 62.93 percent of the norm group obtain an IQ score of 105, when the mean is 100 and the standard deviation is 15. Jean's IQ score of 105 represents a percentile rank of 62.93.

Inferential Statistical Applications of the Normal Distribution

We just showed how the normal curve can be used in a descriptive statistical analysis. The normal curve also plays a key role in *inferential statistics*, which involves inferring information about samples to generalize to populations.

Inferential probability statements made about populations' characteristics (*parameters*) from analysis of sample characteristics (*statistics*) is statistical inference. Parameters are any measurable characteristic of a population, and statistics are any measurable characteristic of a sample. Inferential statistics are estimators in the sense that they estimate parameters. When a random sample is selected from a population with the purpose to understand the population, there is likely to be a difference between the population mean (μ) and the sample mean ($\overline{X}$). The difference between a population parameter and a sample statistic is *sampling error*. If you were able to select an infinite number of sample means from a population, the mean of the sample means would equal the population mean. The infinite number of sample means forms a *sampling distribution of the mean* that is distributed as a normal curve. This is the foundation of one of the most important theorems in statistics. The *central limit theorem* says that "whatever the shape of the frequency distribution of the original population of X's, the frequency distribution of $\overline{X}$ in repeated random samples of size n tends to become normal as n increases" (Snedecor & Cochran, 1967, p. 51). Thus, all that is known about the normal distribution can be applied to the sampling distribution of the mean, including the probability of obtaining a mean by chance and testing hypotheses. The *z*-test is used to test hypotheses when the parameters (μ and σ) are known. A sample mean can be compared to these parameters to see if it belongs to that population. One use of a *z*-test is in situations using test norms where σ and μ are identified in the standardization process.

For example, a researcher wants to determine if a sample of adults who have a learning disability in math would have an average full-scale (FS) IQ score that is different from the normative sample of the Wechsler Adult Intelligence Scale—IV (WAIS-IV). The WAIS-IV norms are: mean = 100 and standard deviation = 15. Since the researcher is not indicating a direction as to whether the average FS IQ score will be higher or lower than the normative sample, it is referred to as *nondirectional* and we can designate the *alternative hypothesis* as H_a: $\mu \neq 100$. The *null hypothesis* is tested, and it is written as H_0: $\mu = 100$. The population mean symbol (μ) is used because we are using the sample mean as an estimator of the population mean in the hypothesis.

A random sample of 50 adults who have a learning disability in math have a mean FS IQ score of 86 on the WAIS-IV. We want to test the null hypothesis

that 50 adults with a learning disability in math are a random sample from a population of adults represented by the standardized norm sample of the WAIS-IV.

A criterion (*critical value*) is identified to compare to a calculated z-test statistic before calculating the z-statistic. We will use a commonly used *alpha level* of $\alpha = .05$. This is a *two-tailed test* since there is no specification that the outcome will show that the sample mean is higher or lower than the norm mean. We will therefore be using both the left (negative) and right (positive) sides of the normal curve to find significance. In this case, if the z-statistic produces a negative value, then we would look in the negative side of the distribution; and if a positive value is found, then we would look in the positive side. Since we are using $\alpha = .05$ and the alternative hypothesis is nondirectional, creating a two-tailed test, we need to distribute half of the alpha (.025) in each tail of the normal curve. We are going to find a z critical value that is located on the abscissa (horizontal axis) of the normal curve where .025 of the curve density falls beyond at either end of the normal curve. We are going to use an online calculator to find the z critical value at .025 in the left tail and right tail of the normal curve using the following directions.

> Go to www.danielsoper.com > click on *Statistics Calculators* > scroll down and click on *Normal Distribution* > click on *Standard Normal Disribution z-score Calculator* > beside *Cumulative probability level:* type 0.025 > click the *Calculate!* button and the answer is z-score: -1.95996398.

Rounded, the z-value is -1.96, corresponding to the $\alpha = .025$ on the negative or left side of the curve. The other half of the $\alpha = .05$ is on the positive or right side of the normal curve since it is a two-tailed test. So, if you followed the same online calculator directions but typed in .975 ($1 - .025$), you would get a rounded $+1.96$ that is a positive value. Thus, the critical value that we will use to compare to the calculated z-statistic is $z_{cv} = \pm 1.96$ ($\alpha = .05$, two-tailed). The formula and calculations of the z-statistic are presented next. The numerator of the formula $\sigma/\sqrt{n}$ is the *standard error of the mean* that is the index reflecting the sampling distribution of the mean.

$$z\text{-statistic} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{86 - 100}{15/\sqrt{50}}$$

$$= \frac{-14}{15/7.07}$$

$$= \frac{-14}{15/7.07}$$

$$= \frac{-14}{2.12}$$

$$z\text{-statistic} = -6.60$$

The z-statistic $= -6.60$ is greater than $z_{cv} = -1.96$, so there is a significant difference in FS IQ between the sample mean (86) of adults who have a learning disability and the mean (100) of the normative sample. The 50 adults with a learning disability in math are not a random sample from a population of adults represented by the standardized norm sample of the WAIS-IV.

One-Sample *t*-Test (Student's *t*-Test)

The z-statistic is not often used since we rarely know the population parameters. A more commonly used statistic that evolved from the z-statistic is used more often and is known as the *one-sample t-test (Student's t-test)*. The originator of the *t*-test was William Gossett, who worked for the Guinness Brewing Company. He used the pseudonym Student when he wrote the seminal scientific article titled "The Probable Error of the Mean," published in the 1908 issue of *Biometrika* (Student, 1908). Gossett had to use the pseudonym Student since, to protect proprietary interests, it was against company policy for employees to publish studies about Guinness.

Gossett developed a new probability distribution called the *t*-distribution that he assumed had an initial normal distribution. In large samples the *t*-distribution is nearly normal, but it is less so when the sample is less than 30. The *t*-distribution works well with all sample sizes as long as statistical assumptions are present. Not

only is the t distribution useful for small samples, but the sample standard deviation and mean can also be used to estimated population parameters.

The one sample t-statistic is very similar to the z-test except that the sample standard deviation (s) is substituted for the known σ and the t-distribution is used to identify probability estimates to compare to obtained statistical values.

For example, a school district wants to assess whether third graders in a particular school perform in reading similarly to all third graders in the school district. The average score on a reading achievement test of third graders in the school district is 105. The sample of 25 third graders from the particular school had an average score of $\overline{X} = 124$ and standard deviation of $s = 13$. The one-sample t-test is used next to test the null hypothesis that H_0: $\mu = 105$ at $\alpha = .01$. The alternative hypothesis is nondirectional, H_a: $\mu \neq 105$. First, we will obtain a critical value using an online calculator to compare to the calculated one-sample t-test. The degrees of freedom $N - 1$ (24) and alpha ($\alpha = .01$) are needed to identify the critical value.

> Go to www.danielsoper.com > click on *Statistics Calculators* > scroll down and click on *t-Distribution* > click on *Student t-Value Calculator* > type in 24 beside *Degrees of freedom:* > click 0.01 next to *Probability level:* > click on *Calculate!* and the answer is *t-value (two-tailed):* ± 2.79693951.

Rounding the t critical value for the two-tailed test is 2.797. Next, the one-sample t-test is calculated.

$$t = \frac{\overline{X} - \mu}{s/\sqrt{N}}$$
$$= \frac{124 - 105}{13/\sqrt{25}}$$
$$= \frac{19}{13/5}$$
$$= \frac{19}{2.6}$$
$$t = 7.31$$

The obtained t-value $= 7.31$ is larger than the critical value $= 2.797$, so we reject the null hypothesis ($p < .01$). We conclude that 25 third graders from a

particular school had a significantly higher average score on reading achievement when compared to the average reading achievement score of all school district third-grade students.

The t-distribution also is used for two commonly used statistics to compare two means for differences, which are the *dependent t-test* and the *independent t-test*.

Dependent t-Test

A comparison is made between two sets of dependent scores when a dependent t-test (paired-sample t-test) is used. The pairs of scores between the two sets are linked together. A dependent t-test is used in the following four situations.

1. A comparison of the pretest and posttest scores of the same participants.

2. A comparison of the scores of one group of participants with another group of participants who are matched on one or more extraneous variables.

3. A comparison of the scores of the same group of participants under two different conditions.

4. A comparison of the scores of naturally occurring correlated pairs, like twins.

We will illustrate the use of the dependent t-test related to research that compares pretest scores with posttest scores on self-compassion from the same 16 participants who received a mindfulness treatment program. Self-compassion is measured by the Self-Compassion Scale (SCS; Neff, 2003), and high scores reflect higher self-compassion. The alternative hypothesis is H_a: $\mu_{preSCS} \neq \mu_{postSCS}$ and the null hypothesis to be tested is H_0: $\mu_{preSCS} = \mu_{postSCS}$.

We start by obtaining a t critical value to compare to an obtained t-value using the online calculator.

> Go to www.danielsoper.com > click on *Statistics Calculators* > scroll down and click on *t-Distribution* > click on *Student t-Value Calculator* > type in 15 beside *Degrees of freedom:* > click 0.05 next to *Probability level:* > click on *Calculate!* and the answer is *t-value (two-tailed): ±2.13144955.*

The rounded value is $t_{CV} = \pm 2.131$ using an $\alpha = .05$ with 15 *df* ($N_{pairs} - 1$, $16 - 1 = 15$). The scores and difference measures used for the analysis are in Table 1.3.

TABLE 1.3   Scores and Difference Measures for Dependent *t* Analysis

ID Number	Pretest	Posttest	Difference (*D*)	D^2
1	65	114	−49	2,401
2	106	120	−14	196
3	45	74	−29	841
4	116	129	−13	169
5	75	105	−30	900
6	46	97	−51	2,601
7	54	95	−41	1,681
8	79	103	−24	576
9	55	109	−54	2,916
10	99	101	−2	4
11	52	94	−42	1,764
12	73	76	−3	9
13	51	77	−26	676
14	88	98	−10	100
15	76	86	−10	100
16	68	88	−20	400
	$\overline{X}_{\text{pre}} = 71.75$	$\overline{X}_{\text{post}} = 97.88$	$\Sigma D = -418$	$\Sigma D^2 = 15{,}334$
		$\overline{D} = 71.75 - 97.88$		
		$\overline{D} = -26.13$		

$$\text{Dependent } t\text{-test} = \frac{\overline{D}}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{N}}{N(N-1)}}}$$

$$= \frac{-26.13}{\sqrt{\dfrac{15{,}334 - \dfrac{(-418)^2}{16}}{16(16-1)}}}$$

$$= \frac{-26.13}{\sqrt{\dfrac{15{,}334 - 10{,}920}{240}}}$$

$$= \frac{-26.13}{\sqrt{18.39}}$$

$$= \frac{-26.13}{4.29}$$

$$t = -6.09$$

The calculated dependent $t = -6.09$ is greater than the $t_{CV} \pm 2.131$, so the null hypothesis is rejected and we conclude that there is a significant difference between the pretest and posttest self-compassion scores. The pretest $\overline{X}_{pre} = 71.75$ and the posttest $\overline{X}_{post} = 97.88$. Since a high score represents higher self-compassion, we conclude that there was a significant gain in self-compassion following the mindfulness treatment.

Independent *t*-Test

The independent *t*-test is used to test if two sample means are significantly different from each other from two independent samples. This is a between group analysis. We are testing whether the two means from independent samples are from different populations. The sample means as estimators of the population parameters based upon the sampling distribution of differences between means.

An independent *t*-test analysis will be demonstrated comparing a randomly assigned group of 16 participants who received a psychotherapy intervention on their changes in thought suppression to a control group ($n = 16$) who received the treatment later. Thought suppression is measured using the White Bear Suppression Inventory (WBSI) (Wegner & Zanakos, 1994), and high scores represent higher perceived thought suppression. The alternative hypothesis is H_a: $\mu_{pschotherapy} \neq \mu_{control}$ and the null hypothesis is H_0: $\mu_{psychotherapy} = \mu_{control}$. The mean and variance of WBSI scores and group sizes for the psychotherapy group of participants were $\overline{X} = 39.75$, $s^2 = 95.67$, $n = 16$, and for the control group $\overline{X} = 53.06$, $s^2 = 130.73$, $n = 16$.

First, we obtain a *t* critical value.

> Go to www.danielsoper.com > click on *Statistics Calculators* > scroll down and click on *t-Distribution* > click on *Student t-Value Calculator* > type in 30 beside *Degrees of freedom:* > click 0.01 next to *Probability level:* > click on *Calculate!* and the answer is *t-value (two-tailed):* ±2.74999566.

The rounded value is $t_{CV} = \pm 2.750$ using an $\alpha = .01$ with 30 *df* ([$n_1 - 1$] + [$n_2 - 1$], [$16 - 1$] + [$16 - 1$] = 30).

$$\text{Independent } t\text{-test} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

$$= \frac{39.75 - 53.06}{\sqrt{\dfrac{95.67}{16} + \dfrac{130.73}{16}}}$$

$$= \frac{-13.31}{\sqrt{5.98 + 8.17}}$$

$$= \frac{-13.31}{3.76}$$

$$t = -3.54$$

The calculated independent $t = -3.54$ is greater than the $t_{\text{CV}} = \pm 2.750$, so we reject the null hypothesis of no differences. The treatment group showed significantly lower thought suppression scores than control group participants who did not receive the treatment ($p < .01$).

We used equal-sized groups in this example of an independent t-test. A pooled variance, $S_p^2 = \dfrac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$, is used if the sizes of groups are unequal and the independent t formula becomes $t = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$.

SUMMARY

This chapter has presented the purpose of the book and information related to the foundations of research and statistics. Several commonly used statistics are covered in the book that are linked to the normal distribution, including: (1) one-way analysis of variance, (2) repeated-measures analysis of variance, (3) factorial analysis of variance, (4) analysis of covariance, and (5) correlation coefficient and multiple regression analysis. Nonparametric statistics are covered later in the book that have less distribution requirements and are referred to as distribution-free statistics. The nonparametric statistics that are covered are Kruskal-Wallis one-way analysis of variance, Mann-Whitney U statistic, Friedman's rank test for k

correlated samples, and the Wilcoxon's matched-pairs signed-ranks test. Each statistic is demonstrated using the hypothesis-testing process that is the subject of the next chapter.

PROBLEM ASSIGNMENT

Review information was presented in this chapter related to measures of central tendency, variability, visual representations of data, and the normal distribution. Moreover, applications of descriptive and inferential statistics of the normal distribution were illustrated. Review problems are presented on the companion website for you to practice on. Use the examples presented in this chapter to guide you as you complete the assignment. Your instructor will evaluate your completed worksheet when it is finished.

KEY TERMS

abscissa

active independent variable

alpha level

alternative hypothesis

attribute independent variable

bar chart

bimodal

categorical variable

central limit theorem

classification variable

coefficient of variation (C)

continuous scale

continuous-interval scale

continuous-ratio scale

criterion variable (CV)

critical value

degrees of freedom (df)

dependent t-test

dependent variable (DV)

discrete scale

discrete-nominal scale

discrete-ordinal scale

extraneous variable (EV)

frequency distribution

histogram

independent t-test

independent variable (IV)

inferential statistics

mean ($\overline{X}$)

mean deviation scores

measures of central tendency

measures of variability (dispersion) of scores

median (M_{dn})

mode (M_o)

negative side

nondirectional

normal distribution

null hypothesis

one-sample t-test (Student's t-test)

operational definition (OD)

ordinate

parameters

peak

percentile rank

positive side

predictor variable (PV)

Q-Q (quantile-quantile) plot

range

residuals

sampling distribution of the mean

sampling error

shoulders

standard deviation of the sample (s)

standard error of the mean

statistics

sum of squares

symmetry

tails

two-tailed test

unbiased estimate

unimodal

variance of the sample (s^2)

x-axis

y-axis

z-score