SEVENTH EDITION

THE PRACTICE OF NURSING RESEARCH

Appraisal, Synthesis, and Generation of Evidence

SUSAN K. **GROVE**
NANCY **BURNS**
JENNIFER R. **GRAY**

ELSEVIER

http://evolve.elsevier.com

evolve http://evolve.elsevier.com/Grove/practice/

21
CHAPTER

Introduction to Statistical Analysis

Data analysis is often considered one of the most exciting steps of the research process. During this phase, you will finally obtain answers to the questions that led to the development of your study. Nevertheless, nurses probably experience greater anxiety about this phase of the research process than any other, as they question issues that range from their knowledge about critically appraising published studies to their ability to conduct research. Critical appraisal of the results section of a quantitative study requires you to be able to (1) identify the statistical procedures used; (2) judge whether these statistical procedures were appropriate for the hypotheses, questions, or objectives of the study and for the data available for analysis; (3) comprehend the discussion of data analysis results; (4) judge whether the author's interpretation of the results is appropriate; and (5) evaluate the clinical importance of the findings (see Chapter 18 for more details on critical appraisal).

As a neophyte researcher performing a quantitative study, you are confronted with many critical decisions related to data analysis that require statistical knowledge. To perform statistical analysis of data from a quantitative study, you need to be able to (1) determine the necessary sample size to power your study adequately; (2) prepare the data for analysis; (3) describe the sample; (4) test the reliability of measures used in the study; (5) perform exploratory analyses of the data; (6) perform analyses guided by the study objectives, questions, or hypotheses; and (7) interpret the results of statistical procedures. We recommend consulting with a statistician or expert researcher early in the research process to help you develop a plan for accomplishing these seven tasks. A statistician is also invaluable in conducting data analysis for a study and interpreting the results.

Critical appraisal of the results of studies and statistical analyses both require an understanding of the statistical theory underlying the process of analysis. This chapter and the following four chapters provide you with the information needed for critical appraisal of the results sections of published studies and for performance of statistical procedures to analyze data in studies and in clinical practice. This chapter introduces the concepts of statistical theory and discusses some of the more pragmatic aspects of quantitative data analysis: the purposes of statistical analysis, the process of performing data analysis, the method for choosing appropriate statistical analysis techniques for a study, and resources for conducting statistical analysis procedures. Chapter 22 explains the use of statistics for descriptive purposes, such as describing the study sample or variables. Chapter 23 focuses on the use of statistics to examine proposed relationships among study variables, such as the relationships among the variables dyspnea, anxiety, and quality of life. Chapter 24 explores the use of statistics for prediction, such as using independent variables of age, gender, cholesterol values, and history of hypertension to predict the dependent variable of cardiac risk level. Chapter 25 guides you in using statistics to determine differences between groups, such as determining the difference in muscle strength and falls (dependent variables) between an experimental or intervention group receiving a strength training program (independent variable) and a comparison group receiving standard care.

Concepts of Statistical Theory

One reason nurses tend to avoid statistics is that many were taught the mathematical mechanics of calculating statistical formulas and were given little or no explanation of the logic behind the analysis procedure or the meaning of the results (Grove, 2007). This mathematical process is usually performed by computer, and information about it offers little assistance to the individuals making statistical decisions or explaining results. We approach data analysis from the perspective of enhancing your understanding of the meaning underlying statistical analysis. You can use this understanding either for critical appraisal of studies or for conducting data analyses.

The ensuing discussion explains some of the concepts commonly used in statistical theory. The logic of statistical theory is embedded within the explanations of these concepts. The concepts presented in this chapter include probability theory, classical hypothesis testing, Type I and Type II errors, statistical power, statistical significance versus clinical importance, inference, samples and populations, descriptive and inferential statistical techniques, measures of central tendency, the normal curve, sampling distributions, symmetry, skewness, modality, kurtosis, variation, confidence intervals, and parametric and nonparametric types of inferential statistical analyses.

Probability Theory

Probability theory addresses statistical analysis as the likelihood of accurately predicting an event or the extent of an effect. Nurse researchers might be interested in the probability of a particular nursing outcome in a particular patient care situation. For example, what is the probability of patients older than 75 years of age with cardiac conditions falling when hospitalized? With probability theory, you could determine how much of the variation in your data could be explained by using a particular statistical analysis. In probability theory, the researcher interprets the meaning of statistical results in light of his or her knowledge of the field of study. A finding that would have little meaning in one field of study might be important in another (Good, 1983; Kerlinger & Lee, 2000). Probability is expressed as a lowercase *p,* with values expressed as percentages or as a decimal value ranging from 0 to 1. For example, if the exact probability is known to be 0.23, it would be expressed as $p = 0.23$. The *p* in statistics is defined as the probability of rejecting the null hypothesis when the null is actually true. Nurse researchers typically consider a $p = 0.05$ value or less to indicate a real effect.

Classical Hypothesis Testing

Classical hypothesis testing refers to the process of testing a hypothesis to infer the reality of an effect. This process starts with the statement of a null hypothesis, which assumes no effect (e.g., no difference between groups, or no relationship between variables). The researcher sets the values of two theoretical probabilities: (1) the probability of rejecting the null hypothesis when it is in fact true (alpha [α], **Type I error**) and (2) the probability of retaining the null hypothesis when it is in fact false (beta [β], **Type II error**). In nursing research, alpha is usually set at 0.05, meaning that the researcher will allow a 5% or lower chance of making a Type I error. The beta is frequently set at 0.20, meaning that the researcher will allow for a 20% or lower chance of making a Type II error.

After conducting the study, the researcher culminates the hypothesis testing process by making a rational decision either to reject or to retain the null hypothesis, based on the statistical results. The following steps outline each of the components of statistical hypothesis testing.

1. State your primary null hypothesis. (Chapter 8 discusses the development of the null hypothesis.)
2. Set your study alpha (Type I error); this is usually $\alpha = 0.05$.
3. Set your study beta (Type II error); this is usually $\beta = 0.20$.
4. Conduct power analyses (Aberson, 2010; Cohen, 1988).
5. Design and conduct your study.
6. Compute the appropriate statistic on your obtained data.
7. Compare your obtained statistic with its corresponding theoretical distribution in the tables provided in the Appendices at the back of this book. For example, if you analyzed your data with a *t*-test, you would compare the *t* value from your study with the critical values of *t* in the table.
8. If your obtained statistic exceeds the critical value in the distribution table, you can reject your null hypothesis. If not, you must accept your null hypothesis. These ideas are discussed in more depth in Chapters 23 through 25 when the results of statistics are presented.

Cox (1958, p. 159) stated, "Significance tests, from this point of view, measure the adequacy of the data to support the qualitative conclusion that there is a true effect in the direction of the apparent difference." Thus, the decision is a judgment and can be in error. The level of statistical significance attained indicates the degree of uncertainty in taking the position that the difference between the two groups is real. Classical hypothesis testing has been largely criticized for such errors in judgments (Cohen, 1994; Loftus 1993). Much emphasis has been placed on researchers providing indicators of effect, rather than just relying on *p* values, specifically, providing the magnitude of the obtained effect (e.g., a difference or relationship) as well as confidence intervals associated with the statistical findings. These additional statistics give consumers of

TABLE 21-1 Type I and Type II Errors

		Decision	
		Reject Null	**Accept Null**
True Population Status	**Null Is True**	Type I Error α	Correct Decision $1 - \alpha$
	Null Is False	Correct Decision $1 - \beta$	Type II Error β

research more information about the phenomenon being studied (Cohen 1994).

Type I and Type II Errors

We choose the probability of making a Type I error when we set alpha, and if we decrease the probability of making a Type I error, we increase the probability of making a Type II error. The relationships between Type I and Type II errors are defined in Table 21-1. Type II error occurs as a result of some degree of overlap between the values of different populations, so in some cases a value with a greater than 5% probability of being within one population may be within the dimensions of another population.

It is impossible to decrease both types of error simultaneously without a corresponding increase in sample size. The researcher needs to decide which risk poses the greatest threat within a specific study. In nursing research, many studies are conducted with small samples and instruments that lack precision and accuracy in the measurement of study variables. Many nursing situations include multiple variables that interact to lead to differences within populations. However, when one is examining only a few of the interacting variables, small differences can be overlooked and could lead to a false conclusion of no differences between the samples. In this case, the risk of a Type II error is a greater concern, and a more lenient level of significance is in order. Nurse researchers usually set the level of significance or $\alpha = 0.05$ for their studies versus a more stringent $\alpha = 0.01$ or 0.001. Setting $\alpha = 0.05$ reduces the risk of a Type II error of indicating study results are not significant when they are.

Statistical Power

Power is the probability that a statistical test will detect an effect when it actually exists. Power is the inverse of Type II error and is calculated as $1 - \beta$. Type II error is the probability of retaining the null hypothesis when it is in fact false. When the researcher sets Type II error at 0.20 before conducting a study, this means that the power of the planned statistic has been set to 0.80. In other words, the statistic will have an 80% chance of detecting an effect if it actually exists.

Reported studies failing to reject the null hypothesis (in which power is unlikely to have been examined) often have a low power level to detect an effect if one exists. Until more recently, the researcher's primary interest was in preventing a Type I error. Therefore, great emphasis was placed on the selection of a level of significance, but little emphasis was placed on power. This point of view is changing as we recognize the seriousness of a Type II error in nursing studies.

As stated in the steps of classical hypothesis testing previously, step 4 is "conducting a power analysis." Power analysis involves determining the required sample size needed to conduct your study after performing steps 1, 2, and 3. Cohen (1988) identified four parameters of power: (1) significance level, (2) sample size, (3) effect size, and (4) power (standard of 0.80). If three of the four are known, the fourth can be calculated by using power analysis formulas. Significance level and sample size are straightforward. Chapter 15 provides a detailed discussion of determining sample size in quantitative studies that includes power analysis. **Effect size** is "the degree to which the phenomenon is present in the population or the degree to which the null hypothesis is false" (Cohen, 1988, pp. 9-10). For example, suppose you were measuring changes in anxiety levels, measured first when the patient is at home and then just before surgery. The effect size would be large if you expected a great change in anxiety. If you expected only a small change in the level of anxiety, the effect size would be small.

Small effect sizes require larger samples to detect these small differences (see Chapter 15 for a detailed discussion of effect size). If the power is too low, it may not be worthwhile conducting the study unless a large sample can be obtained because statistical tests are unlikely to detect differences or relationships that exist. Deciding to conduct a study in these circumstances is costly in time and money, frequently does

TABLE 21-2 Software Applications for Statistical Analysis

Software Application	Website
NCSS (Number Cruncher Statistical System)	www.ncss.com
SPSS (Statistical Packages for the Social Sciences)	www.spss.com
SAS (Statistical Analysis System)	www.sas.com
S+	spotfire.tibco.com
Stata	www.stat.com
JMP	www.jmp.com

not add to the body of nursing knowledge, and can lead to false conclusions. Power analysis can be conducted via hand calculations, computer software, or online calculators and should be performed to determine the sample size necessary for a particular study (Aberson, 2010). Power analysis can be calculated by using the free power analysis software G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) or statistical software such as NCSS, SAS, and SPSS (Table 21-2). In addition, many free sample size calculators are available online that are easy to use and understand. If you have questions, you could consult a statistician.

The power achieved should be reported with the results of the studies, especially studies that fail to reject the null hypothesis (have nonsignificant results). If power is high, it strengthens the meaning of the findings. If power is low, researchers need to address this issue in the discussion of limitations and implications of the study findings. Modifications in the research methodology that resulted from the use of power analysis also need to be reported.

Statistical Significance versus Clinical Importance

The findings of a study can be statistically significant but may not be clinically important. For example, one group of patients might have a body temperature 0.1° F higher than that of another group. Data analysis might indicate that the two groups are statistically significantly different. However, the findings have little or no clinical importance because of the small difference in temperatures between groups. It is often important to know the magnitude of the difference between groups in studies. However, a statistical test that indicates significant differences between groups (e.g., a *t*-test) provides no information on the magnitude of the difference. The extent of the level of significance (0.01 or 0.0001) tells you nothing about the magnitude of the difference between the groups or the relationship between two variables. The magnitude of group differences can best be determined through calculating effect sizes and confidence intervals (see Chapters 22 through 25).

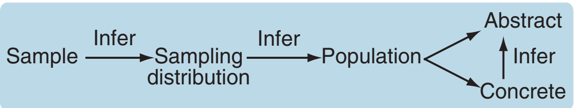
Inference

Statisticians use the terms **inference** and **infer** in a similar way that a researcher uses the term *generalize*. Inference requires the use of inductive reasoning. One infers from a specific case to a general truth, from a part to the whole, from the concrete to the abstract, from the known to the unknown. When using inferential reasoning, you can never prove things; you can never be certain. However, one of the reasons for the rules that have been established with regard to statistical procedures is to increase the probability that inferences are accurate. Inferences are made cautiously and with great care. Researchers use inferences to infer from the sample in their study to the larger population.

Samples and Populations

Use of the terms *statistic* and *parameter* can be confusing because of the various populations referred to in statistical theory. A **statistic**, such as a mean ($\bar{X}$), is a numerical value obtained from a sample. A **parameter** is a true (but unknown) numerical characteristic of a population. For example, μ is the population mean or arithmetic average. The mean of the sampling distribution (mean of samples' means) can also be shown to be equal to μ. A numerical value that is the mean ($\bar{X}$) of the sample is a statistic; a numerical value that is the mean of the population (μ) is a parameter (Barnett, 1982).

Relating a statistic to a parameter requires an inference as one moves from the sample to the sampling distribution and then from the sampling distribution to the population. The population referred to is in one sense real (concrete) and in another sense abstract. These ideas are illustrated as follows:



For example, perhaps you are interested in the cholesterol levels of women in the United States. Your population is women in the United States. You cannot measure the cholesterol level of every woman in the United States; therefore, you select a sample of women from this population. Because you wish your sample to be as representative of the population as possible, you obtain your sample by using random sampling techniques (see Chapter 15). To determine whether the cholesterol levels in your sample are similar to those in the population, you must compare the sample with the population. One strategy would be to compare the mean of your sample with the mean of the entire population. However, it is highly unlikely that you *know* the mean of the entire population; you must make an estimate of the mean of that population. You need to know how good your sample statistics are as estimators of the parameters of the population. First, you make some assumptions. You assume that the mean scores of cholesterol levels from multiple, randomly selected samples of this population would be normally distributed. This assumption implies another assumption: that the cholesterol levels of the population will be distributed according to the theoretical normal curve—that difference scores and standard deviations can be equated to those in the normal curve. The normal curve is discussed later in this chapter.

If you assume that the population in your study is normally distributed, you can also assume that this population can be represented by a normal sampling distribution. You infer from your sample to the sampling distribution, the mathematically developed theoretical population made up of parameters such as the mean of means and the standard error. The parameters of this theoretical population are the measures of the dimensions identified in the sampling distribution. You can infer from the sampling distribution to the population. You have both a concrete population and an abstract population. The concrete population consists of all the individuals who meet your study sample criteria, whereas the abstract population consists of individuals who will meet your sample criteria in the future or the groups addressed theoretically by your framework.

Types of Statistics

There are two major classes of statistics: descriptive statistics and inferential statistics. **Descriptive statistics** are computed to reveal characteristics of the sample and to describe study variables. **Inferential statistics** are computed to draw conclusions and make inferences about the greater population, based on the sample data set. The following sections define the concepts and rationale associated with descriptive and inferential statistics.

Descriptive Statistics

A basic yet important way to begin describing a sample is to create a frequency distribution of the variable or variables being studied. A frequency distribution is a plot of one variable, whereby the *x*-axis consists of the possible values of that variable, and the *y*-axis is the tally of each value. For example, if you assessed a sample for a variable such as pain using a visual analogue scale, and your subjects reported particular values for pain, you could create a frequency distribution as illustrated in Figure 21-1.

Measures of Central Tendency

The measures of central tendency are descriptive statistics. The statistics that represent *measures of central tendency* are the mean, median, and mode. All of these statistics are representations or descriptions of the center or middle of a frequency distribution. The **mean** is the arithmetic average of all of the values of a variable. The **median** is the exact middle value (or the average of the middle two values if there is an even number of observations). The **mode** is the most commonly occurring value in a data set (Grove, 2007; Munro, 2005). It is possible to have more than one mode in a sample, which is discussed later in this chapter. In a normal curve, the mean, median, and mode are equal or approximately equal (see Figure 21-2).

Normal Curve

The theoretical **normal curve** is an expression of statistical theory. It is a theoretical frequency distribution of all *possible* scores (see Figure 21-2). However, no real distribution fits the normal curve exactly. The idea of the normal curve was developed by an 18-year-old mathematician, Gauss, in 1795, who found that data measured repeatedly in many samples from the same population by using scales based on an underlying continuum can be combined into one large sample. From this large sample, one can develop a more accurate representation of the pattern of the curve in that population than is possible with only one sample. In most cases, the curve is similar, regardless of the specific data that have been examined or the population being studied. This theoretical normal curve is symmetrical and unimodal and has continuous values. The mean, median, and mode are equal. The distribution is completely defined by the mean and standard deviation, which are calculated and discussed further in Chapter 22.

Sampling Distributions

The shape of the distribution provides important information about the data. The outline of the distribution shape is obtained by using a histogram. Within this outline, the mean, median, mode, and standard

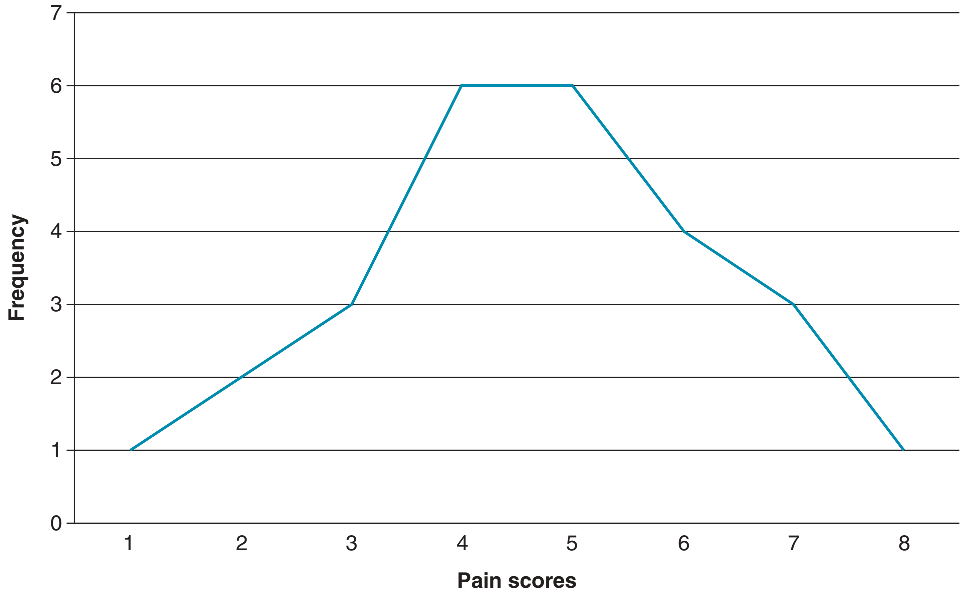


Figure 21-1 Frequency distribution of visual analogue scale pain scores.

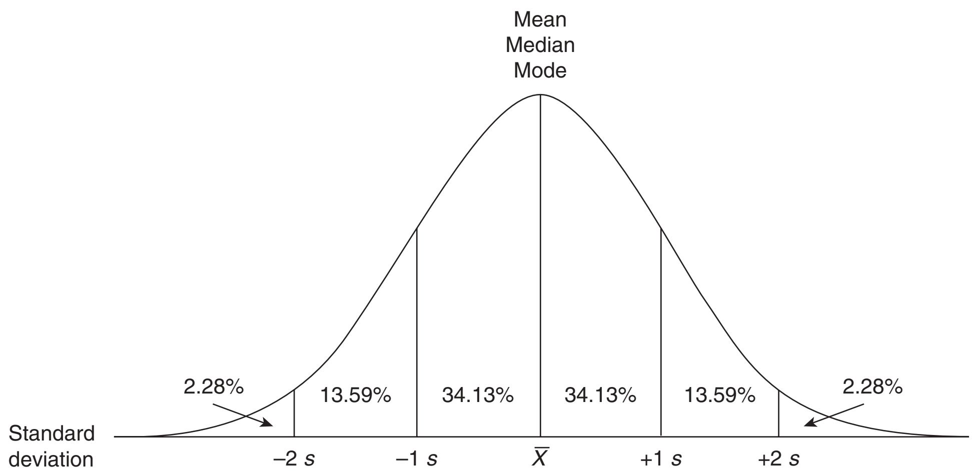


Figure 21-2 Normal curve.

deviation can be graphically illustrated (see Figure 21-2). This visual presentation of combined summary statistics provides insight into the nature of the distribution. As the sample size becomes larger, the shape of the distribution more accurately reflects the shape of the population from which the sample was taken. Even when statistics, such as means, come from a population with a skewed (asymmetrical) distribution, the sampling distribution developed from multiple means obtained from that skewed population tends to fit the pattern of the normal curve. This phenomenon is referred to as the **central limit theorem**.

Symmetry

Several terms are used to describe the shape of the curve (and the nature of a particular distribution). The shape of a curve is usually discussed in terms of symmetry, skewness, modality, and kurtosis. A

symmetrical curve is one in which the left side is a mirror image of the right side (see Figure 21-3). In these curves, the mean, median, and mode are equal and are the dividing point between the left and right sides of the curve.

Skewness

Any curve that is not symmetrical is referred to as **skewed** or **asymmetrical**. Skewness may be exhibited in the curve in various ways. A curve may be **positively skewed**, which means that the largest portion of data is below the mean. For example, data on length of enrollment in hospice are positively skewed. Most people die within the first 3 weeks of enrollment, whereas increasingly smaller numbers survive as time increases. A curve can also be **negatively skewed**, which means that the largest portion of data is above the mean. For example, data on the occurrence of chronic illness by age in a population are negatively skewed, with most chronic illnesses occurring in older age groups. Figure 21-4 includes both a positively skewed distribution and a negatively skewed distribution.

In a *skewed distribution,* the mean, median, and mode are not equal. Skewness interferes with the validity of many statistical analyses; therefore, statistical procedures have been developed to measure the skewness of the distribution of the sample being studied. Few samples are perfectly symmetrical; however, as the deviation from symmetry increases, the seriousness of the impact on statistical analysis increases. In a positively skewed distribution, the mean is greater than the median, which is greater than the mode. In a negatively skewed distribution, the mean is less than the median, which is less than the mode (see Figure 21-4).

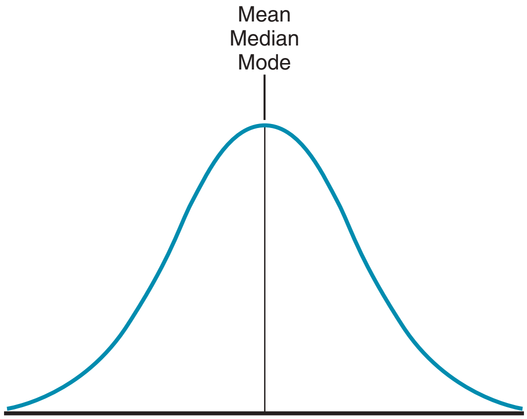


Figure 21-3 Symmetrical curve.

Modality

Another characteristic of distributions is their modality. Most curves found in practice are **unimodal**, which means that they have one mode, and frequencies progressively decline as they move away from the mode. Symmetrical distributions are usually unimodal. However, curves can also be **bimodal** (see Figure 21-5) or multimodal. When you find a bimodal

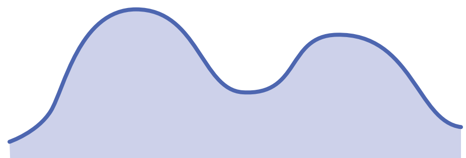


Figure 21-5 Bimodal distribution.

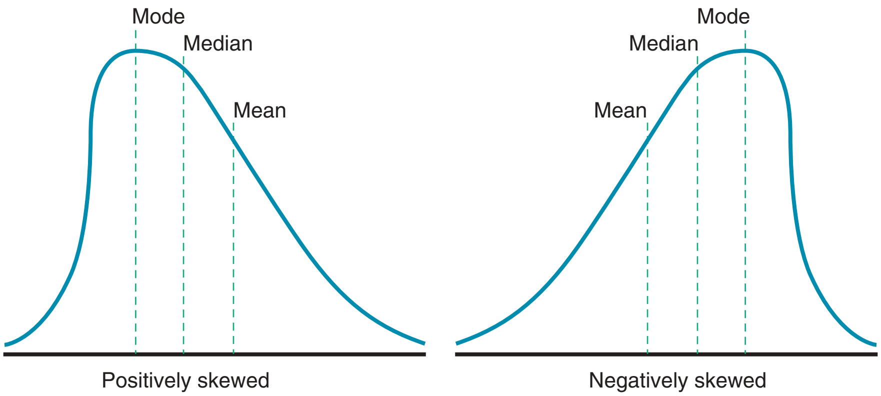


Figure 21-4 Skewness.

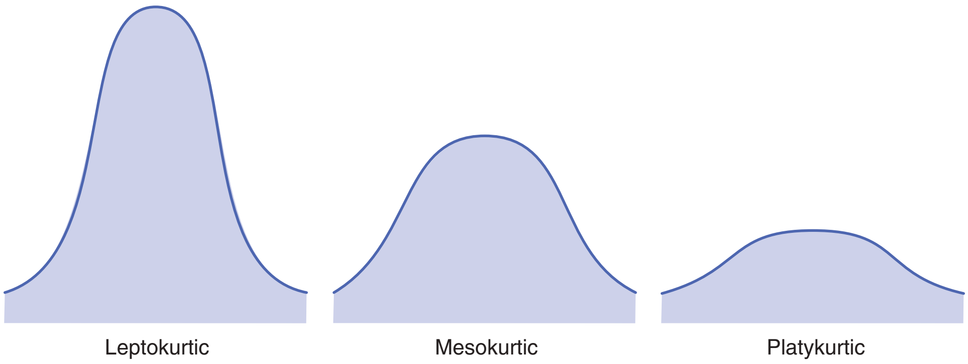


Figure 21-6 Kurtosis.

sample, it usually means that you have not defined your population adequately.

Kurtosis

Another term used to describe the shape of the distribution curve is kurtosis. **Kurtosis** explains the degree of peakedness of the curve, which is related to the spread or variance of scores. An extremely peaked curve is referred to as **leptokurtic**, an intermediate degree of kurtosis is referred to as **mesokurtic**, and a relatively flat curve is referred to as **platykurtic** (see Figure 21-6). Extreme kurtosis can affect the validity of statistical analysis because the scores have little variation in a leptokurtic curve. Many computer programs analyze kurtosis before conducting statistical analyses. A kurtosis of zero indicates that the curve is mesokurtic. Kurtosis values above zero indicate that the curve is leptokurtic, and values below zero that are negative indicate a platykurtic curve (Box, Hunter, & Hunter, 1978).

Tests of Normality

Statistics are computed to obtain an indication of the skewness and kurtosis of a given frequency distribution. The **Shapiro-Wilk's W test** is a formal test of normality that assesses whether the distribution of a variable is skewed or kurtotic or both. This test has the ability to calculate both skewness and kurtosis for a study variable such as pain measured with a visual analog scale. For large samples ($n > 2000$), the **Kolmogorov-Smirnov D test** is an alternative test of normality for large samples.

Variation

The range, standard deviation, and variance are statistics that describe the extent to which the values in the sample vary from one another. The most common of these statistics to be reported in the literature is the standard deviation because of its direct association with the normal curve. If the frequency distribution of any given variable is approximately normal, knowing the standard deviation of that variable allows us to know what percentages of subjects' values on that variable fall between +1 and −1 standard deviation. Referring back to the hypothetical frequency distribution of pain in Figure 21-1, when we calculate a standard deviation, we know that 34.13% of the subjects' pain scores were between the mean pain score and 1 standard deviation above the mean pain score. We also know that 34.13% of the subjects' pain scores were between the mean pain score and 1 standard deviation below the mean. The middle 95.44% of the subjects' scores were between −2 standard deviations and +2 standard deviations.

Confidence Intervals

When the probability of including the value of the parameter within the interval estimate is known, this is referred to as a **confidence interval**. Calculating a confidence interval involves the use of two formulas to identify the upper and lower ends of the interval (see Chapter 22 for calculations). Confidence intervals are usually expressed as "(38.6, 41.4)," with 38.6 being the lower end and 41.4 being the upper end of the interval. Theoretically, we can produce a confidence interval for any parameter of a distribution. It is a generic statistical procedure. Confidence intervals can also be developed around correlation coefficients (Glass & Stanley, 1970). Estimation can be used for a single population or for multiple populations. In **estimation**, we are inferring the value of a parameter from sample data and have no preconceived notion of the value of the parameter. In contrast, in **hypothesis testing**, we have an a priori theory about the value of the parameter or parameters or some combination of parameters. A formula is provided for calculating conference intervals and example confidence intervals are provided for different analysis results in Chapters 22 through 25.

Inferential Statistics

Inferential statistics are computed to draw conclusions and make inferences about the greater population, based on the sample data set. There are two classes of inferential statistics: parametric and nonparametric statistics.

Parametric Statistics

The most commonly used type of statistical analysis is parametric statistics. The analysis is referred to as **parametric statistical analysis** because the findings are inferred to the parameters of a normally distributed population. These approaches to analysis emerged from the work of Fisher (1935) and require meeting the following three assumptions before they can justifiably be used.

1. The sample was drawn from a population for which the variance can be calculated. The distribution is usually expected to be normal or approximately normal (Conover, 1971, Munro, 2005).
2. Because most parametric techniques deal with continuous variables rather than discrete variables, the level of measurement should be at least interval level data or ordinal data with an approximately normal distribution.
3. The data can be treated as random samples (Box et al., 1978).

Nonparametric Statistics

Nonparametric statistical analysis, or distribution-free techniques, can be used in studies that do not meet the first two assumptions of normal distribution and at least interval level data. Most nonparametric techniques are not as powerful as their parametric counterparts (Tanizaki, 1997). In other words, nonparametric techniques are less able to detect differences and have a greater risk of a Type II error if the data meet the assumptions of parametric procedures; this is generally because nonparametric statistics are actually performed on ranks of the original data. When data have been converted into ranks, they inevitably lose accuracy. Because nonparametric statistics have lower statistical power, many researchers choose to submit ordinal data to parametric statistical procedures. If the instrument or measurement procedure yielding ordinal data has been rigorously evaluated, parametric statistics are justified (Fife-Schaw, 1995). For example, researchers often analyze data from a Likert scale with strong reliability and validity as though they are interval level data (see Chapter 17 for a description of Likert scales).

Practical Aspects of Data Analysis

Statistics can be used for a variety of purposes, such as to (1) summarize, (2) explore the meaning of deviations in the data, (3) compare or contrast descriptively, (4) test the proposed relationships in a theoretical model, (5) infer that the findings from the sample are indicative of the entire population, (6) examine causality, (7) predict, or (8) infer from the sample to a theoretical model. These different purposes for data analysis are addressed in Chapters 22 through 25.

The process of quantitative data analysis consists of several stages: (1) preparation of the data for analysis; (2) description of the sample; (3) testing the reliability of measurement; (4) exploratory analysis of the data; (5) confirmatory analysis guided by the hypotheses, questions, or objectives; and (6) post hoc analysis. Statisticians such as Tukey (1977) divided the role of statistics into two parts: exploratory data analysis and confirmatory data analysis. You can perform **exploratory data analysis** to obtain a preliminary indication of the nature of the data and to search the data for hidden structure or models. **Confirmatory data analysis** involves traditional inferential statistics, which you can use to make an inference about a population or a process based on evidence from the study sample.

Although not all of these six stages are reflected in the final published report of the study, they all contribute to the insight you can gain from analyzing the data. Many novice researchers do not plan the details of data analysis until the data are collected and they are confronted with the analysis task. This research technique is poor and often leads to the collection of unusable data or the failure to collect the data needed to answer the research questions. Plans for data analysis need to be made during development of the study methodology. The following section covers the six stages of quantitative data analysis.

Preparing the Data for Analysis

Except in very small studies, computers are almost universally used for data analysis. Use of computers has increased over the last decade with easy-to-use data analysis packages becoming available for personal computers (PCs). When computers are used for analysis, the first step of the process is entering the data into the computer. Table 21-2 lists examples of common statistical packages used for nursing research.

Before entering data into the computer, the computer file that will contain the data needs to be carefully prepared with information from the codebook as described in Chapter 20. The location of each variable in the computer file needs to be identified. Each variable must be labeled in the computer so that the variables involved in a particular analysis are clearly designated on the computer printouts. Develop a systematic plan for data entry that is designed to reduce errors during the entry phase, and enter data during periods when you have few interruptions.

In some cases, data must be reverse-scored before initiating data analysis. Items in scales are often arranged so that sometimes a higher numbered response indicates more of the construct being studied, and sometimes a higher numbered response indicates less of the construct being studied. This arrangement prevents the subject from giving a global response to all items in the scale. To reduce errors, the values on these items need to be entered into the computer exactly as they appear on the data collection form. Values on the items are reversed by computer commands.

Cleaning the Data

Print the data file so that the data can be examined for errors. When the size of the data file allows, you need to cross-check every datum on the printout with the original datum for accuracy. Otherwise, randomly check the accuracy of data points. Correct all errors found in the computer file. Perform a computer analysis of the frequencies of each value of every variable as a second check of the accuracy of the data. Search for values outside the appropriate range of values for that variable. Data that have been scanned into a computer are less likely to have errors but should still be checked.

Identifying Missing Data

Identify all missing data points. Determine whether the information can be obtained and entered into the data file. If a large number of subjects have missing data on specific variables, you need to make a judgment regarding the availability of sufficient data to perform analysis with those variables. In some cases, subjects must be excluded from the analysis because of missing essential data. Missing data can also be imputed via missing data statistical procedures. The rules involving the appropriateness of missing data imputations are complex, and there are many choices of statistical applications. The seminal publication on the subject of missing data imputation was written by Rubin (1976).

Data Transformations

Skewed or non–normally distributed data that do not meet the assumptions of parametric analysis can sometimes be transformed in such a way that the values are distributed closer to the normal curve. Various mathematical operations are used for this purpose. Examples of these operations include squaring each value, calculating the square root of each value, or calculating the logarithm of each value. These operations can allow the researcher to yield a frequency distribution that more closely approximates normality, freeing the researcher to compute parametric statistics.

Data Calculations and Scoring

Sometimes a variable used in the analysis is not collected but calculated from other variables and is referred to as a **calculated variable**. For example, if data are collected on the number of patients on a nursing unit and on the number of nurses on a shift, one might calculate a ratio of nurse to patient for a particular shift. The data are more accurate if this calculation is performed by computer rather than manually. The results can be stored in the data file as a variable rather than being recalculated each time the variable is used in an analysis (Shortliffe & Cimino, 2006).

Data Storage and Documentation

When the data-cleaning process is complete, backups need to be made again; labeled as the complete, cleaned data set; and carefully stored. Data cleaning is a time-consuming process that you will not wish to repeat unnecessarily. If your data are being stored on a PC hard disk drive, be sure to back up the information each time you enter more data. It is wise to keep a second copy of the data filed at a separate, carefully protected site. If your data are being stored on a network, ensure that the network drive is being backed up at least once a day. After data entry, you need to store the original data in secure files for safekeeping. The data files need to be secured as designated by institutional review board policies. This usually includes password protecting computer data files or storing data on encrypted flash drives to which only the research team has access.

Results of data analysis can easily become lost in the mountain of printer paper. Rather than keep paper printouts of statistical output, it is recommended that you make pdf (portable document format) files of each output file and store these files in the same folder as your data sets and reports. There are many free pdf converters available on the Internet for download. A pdf converter allows you to convert any file into a pdf file, which can be read by most computer operating systems. Converting output files into pdf files allows the researcher to transport those files and read them on any computer, even a computer that does not house the statistical software that created the original output file.

All files, including data sets and output files, need to be systematically named to allow easy access later when theses or dissertations are being written or research papers are being prepared for publication. We recommend naming files by time sequence. Name the file by its contents, and at the end of the file name, identify the date (month, day, and year) that the file was created or the analysis was performed. For example, the files named "Rehab Outcomes Data 03-29-12" and "Means and Standard Deviations of Pain Subscales 06-23-12" represent a data file saved on March 29, 2012, and a statistical output file containing means and standard deviations of subscale scores saved on June 23, 2012.

When you are preparing papers that describe your study, the results of each analysis reported in the paper need to be cross-indexed with the output file for reference as needed. As interpretation of the results proceeds and you attempt to link various findings, you may question some of the results. Selected results may not seem to fit with the rest of the results, or they may not seem logical. You may find that you have failed to include necessary statistical information. When rewriting a paper for publication, you may need to report additional results requested by reviewers. The search for a particular output can be time-consuming and frustrating if the printouts have not been carefully organized. It is easy to lose needed results and have to repeat the analysis.

Description of the Sample

After the data have been successfully entered in the computer and stored, researchers start conducting the essential analysis techniques for their studies. The first step is to obtain as complete a picture as possible of the sample. The demographic variables, such as age, gender, and economic status, are analyzed with the appropriate analysis techniques and used to develop the characteristics of the sample. The analysis techniques used in describing the sample are covered in Chapter 22.

Testing the Reliability of Measurement Methods

Examine the reliability of the methods of measurement used in the study. The reliability of observational measures or physiological measures may have been obtained during the data collection phase, but it needs to be noted at this point. Additional examination of the reliability of measurement methods, such as a Likert scale, is possible at this point. If you used a multiple-item Likert scale in data collection, alpha coefficients need to be calculated (Waltz, Strickland, & Lenz, 2010). The value of the coefficient needs to be compared with values obtained for the instrument in previous studies. If the alpha coefficient is unacceptably low (<0.6), you need to determine if you are justified in performing analysis on data from the instrument (see Chapter 16).

Exploratory Analysis of the Data

Examine all the data descriptively, with the intent of becoming as familiar as possible with the nature of the data. You might explore the data by conducting measures of central tendency and dispersion and examining outliers of the data. Neophyte researchers often omit this step and jump immediately into the analyses that were designed to test their hypotheses, questions, or objectives. However, they omit this step at the risk of missing important information in the data and performing analyses that are inappropriate for the data. The researcher needs to examine data on each variable by using measures of central tendency and dispersion. Are the data skewed or normally distributed? What is the nature of the variation in the data? Are there outliers with extreme values that appear different from the rest of the sample that cause the distribution to be skewed? The most valuable insights from a study sometimes come from careful examination of outliers (Tukey, 1977).

In many cases, as a part of exploratory analysis, inferential statistical procedures are used to examine differences and associations within the sample. From an exploratory perspective, these analyses are relevant only to the sample under study. There should be no intent to infer to a population. If group comparisons are made, effect sizes need to be determined for the variables involved in the analyses.

In some nursing studies, the purpose of the study is exploratory analysis. In such studies, it is often found that sample sizes are small, power is low, measurement methods have limited reliability and validity, and the field of study is relatively new. If treatments are tested, the procedure might be approached as a pilot study. The most immediate need is tentative exploration of the phenomena under study. Confirming the findings of these studies requires more rigorously designed studies with much larger samples. Many of these exploratory studies are reported in the literature as confirmatory studies, and attempts are made to infer to larger populations. Because of the unacceptably high risk of a Type II error in these studies, negative findings should be viewed with caution.

Using Tables and Graphs for Exploratory Analysis

Although tables and graphs are commonly thought of as a way of presenting the findings of a study, these tools may be even more useful in helping the researcher to become familiar with the data (see Figure 21-1 of the frequency distribution of visual analogue scale pain scores). Tables and graphs need to illustrate the descriptive analyses being performed, even though they will probably not be included in a research report. These tables and figures are prepared for the sole purpose of helping researchers to identify patterns in their data and interpret exploratory findings, but they are sometimes useful in reporting study results to selected groups (Munro, 2005). Visualizing the data in various ways can greatly increase insight regarding the nature of the data (see Chapter 22).

Confirmatory Analysis

As the name implies, **confirmatory analysis** is performed to confirm expectations regarding the data that are expressed as hypotheses, questions, or objectives. The findings are inferred from the sample to the population. Thus, inferential statistical procedures are used. The design of the study, the methods of measurement, and the sample size must be sufficient for this confirmatory process to be justified. A written analysis plan needs to describe clearly the confirmatory analyses that will be performed to examine each hypothesis, question, or objective.

1. Identify the level of measurement of the data available for analysis with regard to the research objective, question, or hypothesis (see Chapter 16).
2. Select a statistical procedure or procedures appropriate for the level of measurement that will respond to the objective, answer the question, or test the hypothesis.
3. Select the level of significance that you will use to interpret the results, which is usually $\alpha = 0.05$.
4. Choose a one-tailed or two-tailed test if appropriate to your analysis. The extremes of the normal curve are referred to as **tails**. In a **one-tailed test of significance**, the hypothesis is directional, and the extreme statistical values that occur in a single tail of the curve are of interest. In a **two-tailed test of significance**, the hypothesis is nondirectional or null, and the extreme statistical values in both ends of the curve are of interest. Tailedness is discussed in more detail in Chapter 25.
5. Determine the sample size available for the analysis. If several groups will be used in the analysis, identify the size of each group.
6. Evaluate the representativeness of the sample (see Chapter 15).
7. Determine the risk of a Type II error in the analysis by performing a power analysis.
8. Develop dummy tables and graphics to illustrate the methods that you will use to display your results in relation to your hypotheses, questions, or objectives.
9. Determine the degrees of freedom for your analysis. **Degrees of freedom** (*df*) involve the freedom of a score's value to vary given the other existing scores' values. The calculation of *df* varies based on the analysis techniques conducted; Chapters 22 through 25 provide additional information and examples of *df*.
10. Perform the analyses with a computer and rarely manually.
11. Compare the statistical value obtained with the table value by using the level of significance, tailedness of the test, and *df* previously identified.
12. Most analyses are conducted by computer, and the computer printout includes the statistical value obtained by analyzing the data, *p* value, and *df* for each inferential analysis technique.
13. Reexamine the analysis to ensure that the procedure was performed with the appropriate variables and that the statistical procedure was correctly specified in the computer program.
14. Interpret the results of the analysis in terms of the hypothesis, question, or objective.
15. Interpret the results in terms of the framework.

Post Hoc Analysis

Post hoc analyses are commonly performed in studies with more than two groups when the analysis indicates that the groups are significantly different but does not indicate which groups are different. For example, an analysis of variance is conducted to examine the differences among three groups—experimental group, control group, and placebo group—and the groups are found to be significantly different. A post hoc analysis must be performed to determine which of the three groups are significantly different. Post hoc analysis is discussed in more detail in Chapter 25. In other studies, the insights obtained through the planned analyses generate further questions that can be examined with the available data.

Choosing Appropriate Statistical Procedures for a Study

Multiple factors are involved in determining the suitability of a statistical procedure for a particular study. These factors can be related to the nature of the study, the nature of the researcher, and the nature of statistical theory. Specific factors include (1) the purpose of the study; (2) hypotheses, questions, or objectives; (3) design; (4) level of measurement; (5) previous experience in statistical analysis; (6) statistical knowledge level; (7) availability of statistical consultation; (8) financial resources; and (9) access to computers. Use items 1 to 4 to identify statistical procedures that meet the requirements of the study, and narrow your options further through the process of elimination based on items 5 through 9.

The most important factor to examine when choosing a statistical procedure is the study hypothesis. The hypothesis that is clearly stated indicates the statistics needed to test it. An example of a clearly developed hypothesis is, "There is a difference in employment rates between veterans who receive vocational rehabilitation and veterans who are on a wait-list control." This statement tells the researcher that a statistic to determine differences between two groups is appropriate to address this hypothesis.

One approach to selecting an appropriate statistical procedure or judging the appropriateness of an analysis technique is to use a decision tree. A decision tree directs your choices by gradually narrowing your options through the decisions you make. A decision tree that can been helpful in selecting statistical procedures is presented in Figure 21-7.

One disadvantage of decision trees is that if you make an incorrect or uninformed decision (guess), you can be led down a path where you might select an inappropriate statistical procedure for your study. Decision trees are often constrained by space and do not include all the information needed to make an appropriate selection. A more extensive decision tree can be found in *A Guide for Selecting Statistical Techniques for Analyzing Social Science Data* by Andrews, Klem, Davidson, O'Malley, and Rodgers (1981). The following examples of questions designed to guide the selection or evaluation of statistical procedures were extracted from this book:

1. *How many variables does the problem involve?*
2. *How do you want to treat the variables with respect to the scale of measurement?*
3. *What do you want to know about the distribution of the variable?*
4. *Do you want to treat outlying cases differently from others?*
5. *What is the form of the distribution?*
6. *Is a distinction made between a dependent and an independent variable?*
7. *Do you want to test whether the means of the two variables are equal?*
8. *Do you want to treat the relationship between variables as linear?*
9. *How many of the variables are dichotomous?*
10. *Is the dichotomous variable a collapsing of a continuous variable?*
11. *Do you want to treat the ranks of ordered categories as interval scales?*
12. *Do the variables have the same distribution?*
13. *Do you want to treat the ordinal variable as though it were based on an underlying normally distributed interval variable?*
14. *Is the interval variable dependent?*
15. *Do you want a measure of the strength of the relationship between the variables or a test of the statistical significance of differences between groups (Kenny, 1979)?*
16. *Are you willing to assume that an interval-scaled variable is normally distributed in the population?*
17. *Is there more than one dependent variable?*
18. *Do you want to statistically remove the linear effects of one or more covariates from the dependent variable?*
19. *Do you want to treat the relationships among the variables as additive?*
20. *Do you want to analyze patterns existing among variables or among individual cases?*
21. *Do you want to find clusters of variables that are more strongly related to one another than to the remaining variables?*

Each question confronts you with a decision. The decision you make narrows the field of available statistical procedures (see Figure 21-7). Decisions must be made regarding the following:

1. Number of variables (one, two, or more than two)
2. Level of measurement (nominal, ordinal, or interval)
3. Type of variable (independent, dependent, or research)
4. Distribution of variable (normal or non-normal)
5. Type of relationship (linear or nonlinear)

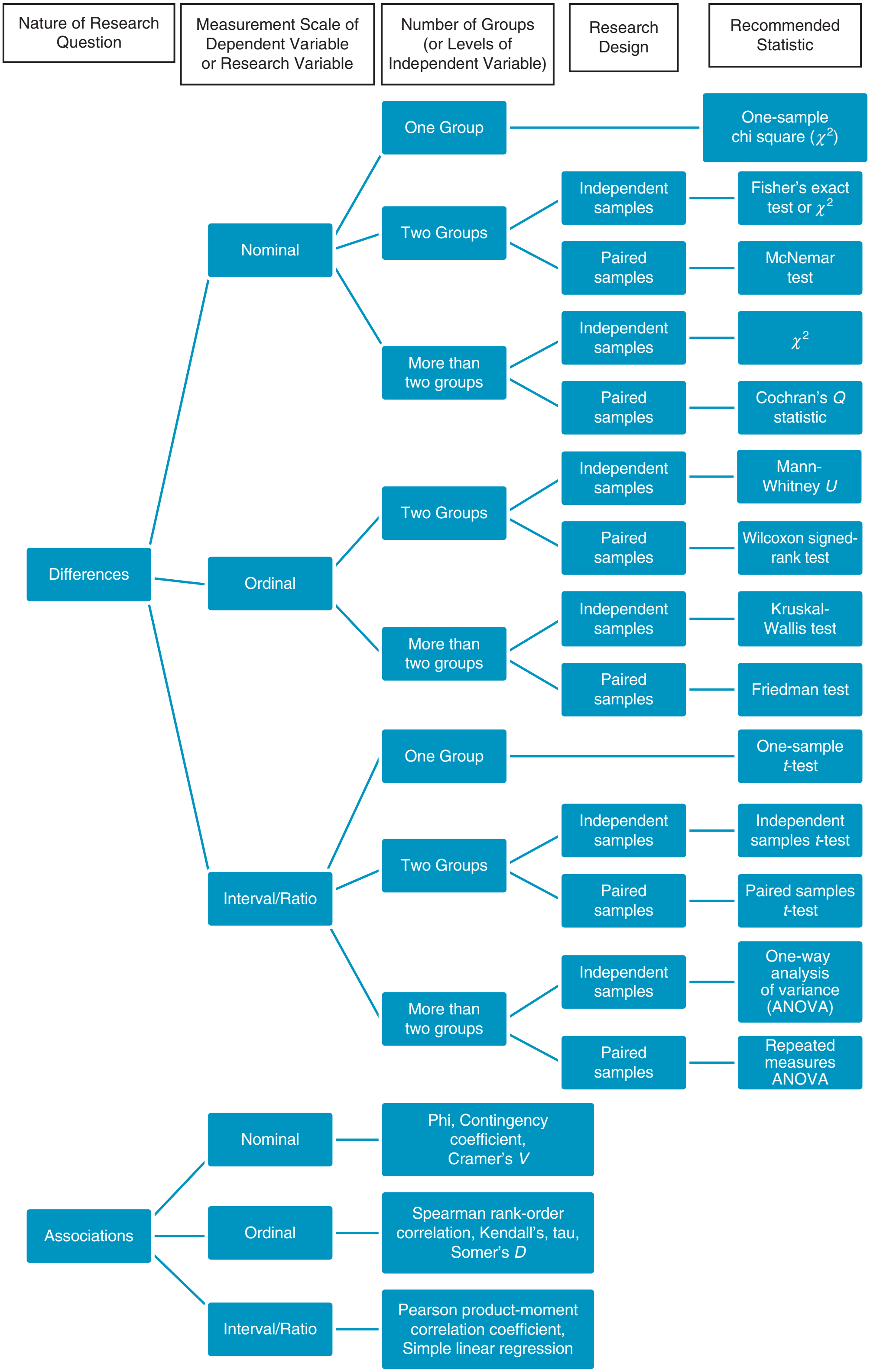


Figure 21-7 Statistical decision tree for selecting an appropriate analysis technique.

6. What you want to measure (strength of relationship or difference between groups)
7. Nature of the groups (equal or unequal in size, matched or unmatched, dependent [paired] or independent)
8. Type of analysis (descriptive, classification, methodological, relational, comparison, predicting outcomes, intervention testing, causal modeling, examining changes across time)

Selecting and evaluating statistical procedures requires that you make many judgments regarding the nature of the data and what you want to know. Knowledge of the statistical procedures and their assumptions is necessary for selecting appropriate procedures. You must weigh the advantages and disadvantages of various statistical options. Access to a statistician can be invaluable in selecting the appropriate procedures.

KEY POINTS

- This chapter introduces you to the concepts of statistical theory and discusses some of the more pragmatic aspects of quantitative data analysis, including the purposes of statistical analysis, the process of performing data analysis, the choice of the appropriate statistical procedures for a study, and resources for statistical analysis.
- Two types of errors can occur when making decisions about the meaning of a value obtained from a statistical test: Type I errors and Type II errors.
- A Type I error occurs when the researcher concludes a significant effect when no significant effect actually exists.
- A Type II error occurs when the researcher concludes no significant effect when an effect actually exists.
- The formal definition of the level of significance, or alpha (α), is the probability of making a Type I error when the null hypothesis is true.
- The p value is the exact value that can be calculated during a statistical computation to indicate the probability of making a Type I error.
- Power is the probability that a statistical test will detect a significant effect when it actually exists.
- Statistics can be used for various purposes, such as to (1) summarize, (2) explore the meaning of deviations in the data, (3) compare or contrast descriptively, (4) test the proposed relationships in a theoretical model, (5) infer that the findings from the sample are indicative of the entire population, (6) examine causality, (7) predict, or (8) infer from the sample to a theoretical model.
- The quantitative data analysis process consists of several stages: (1) preparation of the data for analysis; (2) description of the sample; (3) testing the reliability of measurement; (4) exploratory analysis of the data; (5) confirmatory analysis guided by hypotheses, questions, or objectives; and (6) post hoc analysis.
- A decision tree is provided to assist you in selecting appropriate analysis techniques to use in analyzing study or clinical data.

REFERENCES

Aberson, C. L. (2010). *Applied power analysis for the behavioral sciences*. New York, NY: Routledge Taylor & Francis Group.

Andrews, F. M., Klem, L., Davidson, T. N., O'Malley, P. M., & Rodgers, W. L. (1981). *A guide for selecting statistical techniques for analyzing social science data* (2nd ed.). Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

Barnett, V. (1982). *Comparative statistical inference*. New York, NY: Wiley.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters*. New York, NY: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist*, *49*(12), 997–1003.

Conover, W. J. (1971). *Practical nonparametric statistics*. New York, NY: Wiley.

Cox, D. R. (1958). *Planning of experiments*. New York, NY: Wiley.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fife-Schaw, C. (1995). Levels of measurement. In G. M. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology*. Thousand Oaks, CA: Sage.

Fisher, R. A., Sir. (1935). *The designs of experiments*. New York, NY: Hafner.

Glass, G. V. & Stanley, J. C. (1970). *Statistical methods in education and psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Good, I. J. (1983). *Good thinking: The foundations of probability and its applications*. Minneapolis, MN: University of Minnesota Press.

Grove, S. K. (2007). *Statistics for health care research: A practical workbook*. St. Louis, MO: Saunders.

Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley.

Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). New York, NY: Harcourt Brace.

Loftus, G. R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instrumentation, & Computers*, *25*(2), 250–256.

Munro, B. H. (2005). *Statistical methods for health care research* (5th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Shortliffe, E. H. & Cimino, J. J. (2006). *Biomedical informatics: Computer applications in health care and biomedicine*. New York, NY: Springer Science.

Tanizaki, H. (1997). Power comparison of non-parametric tests: Small-sample properties from Monte Carlo experiments. *Journal of Applied Statistics*, *24*(5), 603–632.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2010). *Measurement in nursing and health research* (4th ed.). New York, NY: Springer Publishing Company.